QUANTIFICATION OF HAPPINESS INEQUALITY

Wim Kalmijn

QUANTIFICATION
OF
HAPPINESS INEQUALITY

Kwantificering van ongelijkheid in geluk

Wim Kalmijn

Cover design: © 2010 Joop van Opijnen and Wim Kalmijn
Lay-out: Van Opijnen Vormgeving, Schiedam NL
www.jopoinvorm.nl
Font: Frutiger / Cambria math
Printed in The Netherlands by Ipskamp Drukkers B.V., Enschede

ISBN 978-90-9025604-7
NUR 916

QUANTIFICATION OF HAPPINESS INEQUALITY

Kwantificering van ongelijkheid in geluk

P r o e f s c h r i f t

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 2 december 2010 om 15.30 uur door

Wilhelm Marinus Kalmijn
geboren te Overschie

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof.dr. R. Veenhoven
Overige leden:	Prof.dr. P.A. Dykstra
	Prof.dr. P. Theuns
	Prof.dr. M.A. Zwanenburg
Copromotor:	Dr. L.R. Arends

Opgedragen aan Marian Kalmijn-van der Lecq

Keywords

Distribution
Inequality
Happiness
Life satisfaction
Measurement
Parameter estimation
Quantification
Scale interval approach

Contents

Chapter 6

THE SCALE INTERVAL APPROACH I PRINCIPLE, DESCRIPTION AND METHODOLOGY

Chapter 7

THE SCALE INTERVAL APPROACH II. EXPERIENCES WITH 100 CASES

Summary

A summary in Dutch is given at *page 235**.*

Happiness is considered to be an important aspect of human life and this is reflected in a growing interest of social sciences during the past decennia. Happiness research is only possible if happiness can be measured and quantified. The measurement of happiness, more specifically the way observation results are further processed, is discussed in this dissertation, which is intended to be a methodological contribution to happiness research. Happiness in this context is defined as "the degree to which an individual judges the overall quality of his/her life-as-a-whole favorably".

Traditionally, this happiness is measured by simply asking the person to rate it. A frequently used method is to ask a closed question, e.g. "Taking all together, how happy would you say you are ?" and to offer a limited number (3 – 7) of response categories, one of which has to be ticked, e.g. "pretty happy". In particular such happiness questions with textual response categories, shortly referred to as "verbal scales", form the object of the investigation presented here.

Investigators of happiness are not just interested in individual happiness scores, they are also interested in happiness in communities. In this dissertation we shall refer to "nations", but our findings are equally applicable to other defined collectivities. Not all individuals are equally happy. With respect to this happiness inequality, we have to distinguish between within-nations and between-nations inequality. To collect information on both, social scientists examine samples from the population or a sub-population, e.g. from all adult citizens. This sampling should be done at random, but in reality this never happens. Nevertheless, we assume that all samples discussed in this study can be considered as if sampling has been at random, since information on the happiness distribution within the population represented by that sample can only be obtained under that assumption.

Aspects of inequality

The inequality within a sample is expressed in the frequency distribution of the response categories. Usually this distribution is characterized by two statistics. One concerns the central value, about which the individual happiness

ratings are spread; the other one pertains to the dispersion, that is the within-sample happiness inequality.

Sociologists are interested in both elements, because it is assumed that a society performs better as the central value of the happiness distribution is larger, but also as there is less happiness inequality within the population. This latter extension is introduced by the "egalitarians", while the "utilitarians" are only interested in the general happiness level.

An adequate view on the happiness situation in some nation requires us to be able to quantify both aspects. It is, however, not sufficient for happiness investigators to know estimates of descriptive parameters of the statistical happiness distribution, they also seek to quantify the association with other societal variables that may act as conditions of happiness. This research also requires quantification of happiness inequality. The Happiness Research Department of Erasmus University Rotterdam in The Netherlands investigates such relationships. Within the context of this research several methodological questions have arisen in the last decennium . Some of these questions, together with the corresponding answers and/or solutions, are discussed in this dissertation.

Problems with measuring happiness inequality

The core activity of inferential Statistics is to obtain conclusions about populations on the basis of results from sample measurements. This objective forms also the central issue of this dissertation on happiness measurement. The latter involves a number of specific problems, in particular when verbal scales are applied. The first action in such cases is to convert alphanumeric text into numbers. The usual procedure is to code each response category with a cipher. The next step is that these ciphers are viewed as cardinal numbers, i.e. as numbers to which the application of the usual basic arithmetical operations is admissible. A sample average value can be calculated on this basis as a weighted average of the code numbers to serve as a measure of the central tendency, together with e.g. the standard deviation accordingly as a measure of the within-sample dispersion. Equidistance of the category positions on the scale of measurement had been introduced by the coding implicitly. After this coding, the text of the categories is ignored completely. Questions such as whether e.g. the French "heureux" has exactly the same meaning as the British and/or the American "happy" cease to be a problem. All data obtained using any four point scale are treated in exactly the same way. However, in practice the number of categories varies usually between 3 and 7, so an additional treatment is required to enable comparison of results obtained by using scales with different lengths. Traditionally, this is achieved by proportionally stretching the different primary scales to one common secondary 0 to 10 scale. The value 0 is assigned to the most unhappy category, irrespective of its original label text, and the most happy one receives the value 10. All other categories are located equidistantly in between according to their position in the order of the primary scale of measurement, and a sample average and standard deviation are obtained through this linear transformation, both on a 0 to 10 scale. Generalization of these values also makes them applicable to the population.

One more problem in this approach is the limited number of response categories on the scale; in practice, seven is the maximum value. This means that happiness, as measured in the sample, can adopt not more than seven different values. Generalizing the sample results in a direct way implies that this is also declared to apply to the distribution of happiness within the nation. Happiness as measured in the sample is a necessarily discrete variable, but happiness in a nation is viewed in the same way, although it seems more obvious that a continuous variable on the 0 to 10 interval would be much more appropriate.

Alternatives to the traditional approach

Many methodological objections can be raised against the traditional procedure as it has been described above. Most of them could presumably be met by discontinuing the application of verbal scales and replacing these with numerical e.g. 1 to 10 scales. Such a policy decision would, however, imply that most published studies on happiness would become inaccessible for meta-analyses and for trend studies; this would concern thousands of studies. Therefore, in this thesis we developed an alternative approach to the radical switch towards numerical scales and we examined how we can meet, as much as possible, the above methodological objections against the traditional procedures.

Plan of this dissertation

The first chapter is an introduction in which we describe the context from which this dissertation emanated. Special attention is paid to the three core concepts of its title: happiness, inequality and quantification.

Chapter 2 starts with a description of how happiness usually is measured. Special attention is given to a number of general methodological aspects, in particular to happiness as a variable. In this respect, there are essential differences between happiness and income and these are specified into more detail. A more fundamental view on ‘ inequality’ as a concept and on its quantification is introduced In chapter 3. Inequality within a sample is described in terms of a binary relation, defined on the set of all individual response category selections in a sample. This enables us to calculate not only a minimum (zero) value of the inequality, but also a maximum. The latter is realized if the respondents of half the sample select the most happy category and those of the other half the most unhappy alternative. Inequality in some sample as a whole can be quantified then as the percentage of that maximum attainable value. Inequality within a continuous distribution in a population can be defined in a related way by using differentials. The practical elaboration is generally rather complicated, at least for nonlinear probability density functions. However, numerical integration may enable one to design a picture of the inequality as a function of it parameters as is demonstrated for a beta probability distribution.

The findings of chapter 3 are applied in chapter 4, in which we deal with the question of which measures are apt to characterize the happiness inequality in a sample from a nation. Nine candidates are applied to five series of hypothetical distributions, with increasing inequality within each series. Assessment is done against a number of previously selected criteria. Eventually It appears that the standard deviation and the mean (absolute) distance between all possible pairs of respondents within a sample are the preferred statistics. The mean absolute deviation from the mean and the interquartile range are a second choice as they perform slightly less well. The coefficient of variation, the Gini coefficient and Theil’s entropy measure definitely fall below standard for happiness as it is usually measured. The range and the percentage outside the modal category are also rejected as candidates. In happiness research, the standard deviation was the usual measure of inequality within a sample, and the conclusion is that there is no reason to discontinue this practice.

The above mentioned antagonism between utilitarians and egalitarians can be solved by constructing a measure that honours both views. The Inequality-Adjusted Happiness (IAH) is presented in the fifth chapter as to be such a measure; this IAH is a linear combination of the average happiness and the standard deviation within a sample. The theoretical maximum value of this index equals 100 in an ideal society, against close to zero for the most miserable variety. If one assigns equal weights to both above views, then between 2000 and 2010 IAH-values of 140 nations range from 19 to 79, so the discriminative power of the index seems to be sufficient.

The last two chapters are focused on the way sample results are translated into characteristics describing the happiness distribution in the population, i.e. the distribution of the happiness values over all inhabitants of that nation. The usual method has various methodological shortcomings. "Thurstone values" have been introduced in 1993 by Veenhoven c.s. as a first correction and further improvement was expected from the "Happiness Scale Interval" approach. This approach is the first that accepts and takes into account the idea that e.g. "pretty happy" does not represent a single point on the happiness scale of measurement , rather it represents an interval of contiguous happiness values. A group of judges is requested to identify on a line segment from 0 to 10 the position of the boundary between e.g. "not too happy" and "pretty happy" as adjacent response options on the same verbal scale. The native judges perform this task on verbal scales that have been translated into their mother tongue. Their opinions on the position of these boundaries are the basis of the conversion of sample data from a nation with the same language into information on the happiness distribution in that population in a later stage. In principle, this method can even be applied to observations from the past to improve these estimates. Three models with divergent validity have been developed and one of them is recommended for future application. As far as known, the innovative element of this approach is that it is the first considering and treating happiness as a continuous random variable and also the first that directly measures the cumulative frequency of the happiness distribution, at least a number of points of it.

Chapter 6 is devoted to the methodological problems of current practice, to the principles of the scale interval approach and to the various models for further processing happiness observations on this basis. The methodological advantages of the recommended approach as compared to the current practice are a better validity of the results and the possibility to quantify the inaccuracy of the estimates of the parameters of the happiness distribution both within and between nations.

An preliminary evaluation of this method, based on the analysis of one hundred of the first cases, is described in the seventh and final chapter. A case in this context is defined as one of the 3 – 10 different happiness questions, that have been judged by the same group of judges within a short time interval. The results of this explorative study suggest good perspectives. The precision of estimates is determined predominantly by the number of judges per case. Good instruction of the local co-investigator and of the judges is important and is expected to contribute to the reduction of the percentage, 11 % until now, of judgements to be rejected. It is remarkable that the reasons for rejection are found predominantly at the unhappy end of the scale of measurement. There are good reasons to assume that the cause of a number of anomalies lies in ill-designed combinations of the lead question and the formulation of the corresponding alternative response options. In principle, the scale interval approach includes in this way an interesting opportunity to discriminate between well usable happiness measures and less apt specimens.

✧ ✧ ✧

Chapter 1

Introduction

1.1 Context

Immediately the clock strikes mid-night each new year, the very first priority for everybody is to wish those present a "Happy New Year" whether we know them or not. Although the way this is done may differ slightly in different places, this universal practice demonstrates to what extent in our culture happiness is considered to be an important matter. We want to wish to our co-citizens happiness, at least at that specific moment. The question of how effective such wishes are is a different issue and one that will not be discussed here.

It is not very difficult to find various other examples to the one given above to demonstrate that happiness is an important value in modern society, and it is not surprising that this social development is reflected in an increased interest in social sciences of this phenomenon. Although happiness has been given attention by philosophers for more than two millennia, e.g. as εὐδαιμονία in "Ethica" by Aristoteles (384 – 323 BC), the level of interest has increased very rapidly during the last decennia.

Happiness as an object of science has a relatively long tradition. Buijs (2007, Chapter 6) describes how in the seventeenth century , but especially in the eighteenth, several recommendations were made to investigate the subject using the same techniques as had been recently developed in the physical sciences and by applying mathematical methods. This principle was advocated by prominent philosophers, including e.g. Descartes and Locke. Happiness was considered to be property that could be measured, at least in principle, as was postulated by e.g. R. Lucas. J.-B. Merian proposed to reserve the name "psychometrics" for a future more exact science of happiness on the basis of measurements, which were considered as not yet possible at that time (1766). However, the number of people that successfully accepted such an invitation remained extremely modest. The nature of the most contributions were deductive and no description of an operational empirical procedure has been found before the twentieth century.

The earliest measurements more or less related to happiness, that we have found, have been reported by Webb (1915) in his thesis. The author does not measure happiness or some directly related property of subjects, but the judgment of their tendency to be "cheerful" (as opposed to being depressed and low-spirited), expressed by other members of their peer group within a boarding-school. A similar approach is given by Washburn et al. (1925), who do not apply peer rating only, but also request a self report by the girls judged in the test.

The first large scale survey that included measuring happiness was conducted in 1946 (Easterlin, 1974) by the American Institute of Public Opinion (AIPO) using a sample of 3.151 respondents from the USA population. This test was entirely based on self report.

In the last fifty years, happiness has become an increasingly common theme in cross-national research. The first comparative study on happiness was carried out by Cantril (1965) in 1960. Since then, items on happiness have been adopted in the core questionnaires of several international survey programs, such as the Euro-barometer (since 1973), the World Value Survey (since 1980) and lately the European Welfare Survey.

The present dissertation emanates from the happiness research as it is being developed by Veenhoven and his co-workers at the Faculty of Social Sciences of the Erasmus University Rotterdam in The Netherlands. At (ir)regular occasions, this research has given and is still giving rise to methodological questions, in particular questions connected with the quantification of happiness inequality both within and between nations and other collectivities. The application of Statistics is unavoidable and vital when attempting to answer such questions and to solve underlying problems in the field of happiness studies. The simple reason is that happiness is always measured at the individual level in samples, while sociologists are only interested in the populations that are (assumed to be) represented by these samples. Therefore, the core subject of this dissertation is to bridge this gap, i.e. to find ways to convert the sample information into valid and valuable information about the population with respect to happiness. Within this context, such information can only be described in terms of probability and of statistical distributions and their parameters. A number of those separate background contributions to this happiness research have been collected in a more coherent way in this dissertation, where some of them will be elaborated in more depth and detail than is considered acceptable for publication in current scholarly journals.

1.2 Happiness

A central activity of the above mentioned happiness research group at Erasmus University Rotterdam is the maintenance of the "World Database of Happiness, a continuous register of scientific research on subjective appreciation of life", further abbreviated WDH (Veenhoven, 2010).

The findings of all research on happiness are gathered in this WDH. The section "Happiness in Nations" on distributional findings of this database covers the results of 4704 surveys among general population samples in 225 nations and provides time-series for more than 20 years for some 15 nations (data at October 1, 2009).

Within this context, happiness is defined as "the degree to which an individual judges the overall quality of his/her life-as-a-whole favorably". In view of the context of the present study, it is obvious that the same description of happiness as a concept should be adopted. For a more detailed discussion, the reader is referred to chapter 2 of the 'Introductory text to the collection of "Measures of Happiness" in the WDH[1].

Psychologists investigate the happiness concept typically at the level of individuals, trying to explain why one person enjoys life more than someone else (e.g. Diener et al. 1999). Sociologists focus rather on happiness in collectivities, such as nations (e.g. Veenhoven, 1999). The most common questions in studies on happiness in nations are (i) how happy citizens typically are and (ii) why people enjoy life more in one nation than in another. Most happiness research examines the association between happiness and its conditions.

1.3 Inequality

With respect to happiness, inequality is described in the literature in various ways, which can be collected broadly in two subsets.

The first one concerns variables, more precisely the values these can adopt. In some specified respect, objects are either equal or they are not. 'Unequal' and 'different' are synonymous terms in this case. If two respondents give the same

[1] *Chapter 2: Concept of Happiness: Available at:*
http://worlddatabaseofhappiness.eur.nl/hap_quer/introtext_measures2.pdf

answer to the same question when asked to report their degree of happiness, then and only then is their happiness equal by definition. Mathematics uses two symbols for this relation between the values of such variables: "=" and "≠". This neutral interpretation of inequality can be referred to as the 'logical' interpretation.

The second subset of inequality connotations can be described as 'sociological' and applies to nations or to other societies. Inequality in such communities concerns differences between the positions of its citizens within this society, especially from a socio-economic point of view, and happiness is one of the essential dimensions taken into account. This second approach is linked to what is known as the discussion between "utilitarians" and "egalitarians". In section 1.5 we shall pay attention to happiness in this perspective.

Since our intention is this dissertation to be a primarily methodological one, our focus will be on the 'neutral' logical interpretation, although this will not exclude that tools will be provided that are valuable when applied to the sociological approach to inequality in happiness studies.

From a methodological point of view, inequality is the core concept in the empirical science of happiness and its relationship with anything that influences this state of mind, either actually or potentially. Happiness may vary and this gives rise to inequality at three levels: (i) intrapersonal over time, (ii) interpersonal within some society or a part of it, and (iii) between such societies/communities. As a consequence, happiness may also vary within a society over time.

If there was no intrapersonal variability, it would be impossible to exert any influence on the happiness of individuals; this would have serious psychotherapeutic and political implications. If there was no variability of happiness at any of the three above levels, it would be virtually impossible to measure happiness at all. It would even be most doubtful whether we would be able to recognize the phenomenon happiness anyhow.

The fact that it appears to be possible to measure happiness in the sense of the intensity of subjective well-being and life satisfaction to some extent at all three levels (see e.g. Veenhoven, 1984 and 2002a) proves the existence of happiness variability.

1.4 Quantification

The quantification of happiness inequality is a key element of happiness research, since it can considered to be a necessary condition for it. There are several types of questions where quantification is relevant. The first concerns the quantification of the average happiness in a nation. We want to know, and to describe, whether on average inhabitants of nation XX are happier than those of nation YY, and to what extent. Calculating the inaccuracy of these estimated average values and of their difference requires the knowledge of the happiness inequality within these nations, at least an estimate of it. The second type of questions refers to inequality as such and is raised in the context of the above mentioned debate between utilitarians and egalitarians.

Moreover, as long as one is not capable of quantifying happiness, it is very difficult, if not impossible, to establish its relationship with any of its conditions. An investigation of how happiness has developed over time within a nation and the comparison of two or more nations for their happiness levels at specific points in time also requires to be able to quantify happiness.

Quantification of happiness does not only deal with the direct measurement of happiness using a sample consisting of subjects from a nation, it also includes the conversion of the sample results to valid and valuable happiness information about the nation as a whole, the central study object of this dissertation.

As has been pointed out already, in the vast majority of all happiness studies scientists investigate the relationship between happiness and variables other than happiness; these latter variables are usually referred to as "correlates". A number of them are considered to be conditions. In this publication, however, we shall confine ourselves to the quantification of happiness inequality per se, which obviously is most relevant to support the correlational studies, but we shall do so without dealing with such studies as such. A few inevitable exceptions will be found in sections 6.7 and 6.8.

1.5 Utilitarianism and egalitarianism on happiness

We have already mentioned the controversy between utilitarianism and egalitarianism with respect to happiness in section 1.3. For some time, there has been a discussion about the quality of society and calls for social reform. Over the last centuries, political philosophers have brought a system into being where the debate centres on distinguishing standards for evaluating the quality of society.

Different standards

One of the standards for a good society is that its citizens should be happy. This principle is central to 'utilitarian' moral philosophy, more precisely, to 'rule-utilitarianism', which holds that policy makers should aim at devising rules to create a society that provides "the greatest happiness for the greatest number of citizens" (Bentham, 1789). This criterion is put into practice in empirical happiness research, in particular in studies where average happiness across nations is compared and an attempt is made to identify the societal characteristics behind the observed differences (Veenhoven 1997, 2004).

Another standard used to evaluate the quality of a society is the degree of inequality among its citizens. This principle is central to a tradition of 'egalitarian' moral philosophy, which holds that policy makers should try to reduce inequality as much as possible. This criterion is also applied in empirical social research, mostly in cross-national comparisons of equal rights and income inequality. These principles can come into conflict. The promotion of happiness may be at the cost of social equality, and in this context a standard objection against utilitarianism is that it legitimizes the repression of a minority. Likewise, social equality can be obtained to the detriment of happiness, the failed communist experiment has shown this to be the case. Since there is broad support for both principles, policy makers must look for options that satisfy each of the above tenets.

Obviously, both views on what is the best quality of a society have consequences for the measurement of happiness. Whereas the primary interest of egalitarians concerns the inequality within a nation, utilitarians are more focused on inequality between nations. Both interests deserve our attention.

1.6 Some terminological remarks

(a) In this context and in view of our choice in favour of the logical interpretation of inequality, we prefer to deal with 'inequality' rather than with 'equality'. The reason is that inequality may exist in gradations and can be measured in this case, at least in principle. This does not apply to 'equality', which in this approach is basically a 'zero-inequality'. This is illustrated best by George Orwell's (1945) famous exception: "All animals are equal, but some animals are more equal than others."

in the sociological approach of inequality, we see less reasons to systematically avoid the term 'equality', which in that context can be considered to be a concept complementary to inequality and the latter one is defined then in a less strict way than is done in the methodological sense.

(b) In this thesis we will describe the object of our investigations as "happiness" in a general sense of subjective well-being, that is self-reported using ordinal scales with a small number of response categories, as will be demonstrated in section 2.1. Most of our findings, however, may apply mutatis mutandis to other phenomena that are measured in a similar way, such as e.g. disparity in self-esteem among pupils in schools. For reasons of readability, we will describe our findings only in terms of "happiness" without making reference to this extension.

(c) For the same reason, we shall deal with happiness in "nations" without the restriction of this term to a strictly defined geographic/political meaning in agreement with international law. Our findings equally concern other demographic groups/societies/communities, even is this enlargement is not added verbatim. This is also done to avoid confusion problems with respect to the term "population", which has both a demographical and a statistical meaning. In principle, within this publication, the term "population" is used generally in its statistical sense, i.e. as the set of all happiness values within a nation rather than the set of their bearers.

(d) People may believe that inequality is great and rising in their country, while the existing differences are in fact small and diminishing. In this dissertation, we ignore this believed inequality and only deal with actual inequality of happiness.

(e) Finally, we shall describe the behaviour of individuals as that of males, although everything equally applies to women, and all the text should be read accordingly. Politically correct formulations as "he or she", "(s)he", "his or her" etc. have been avoided merely in favour of readability, while all other possible reasons are strictly excluded; this is done in agreement with the "Promotiereglement EUR" (Rules for Promotion Erasmus University Rotterdam NL).

1.7 Plan of this dissertation

The structure of this dissertation is depicted schematically in Fig.1.1.

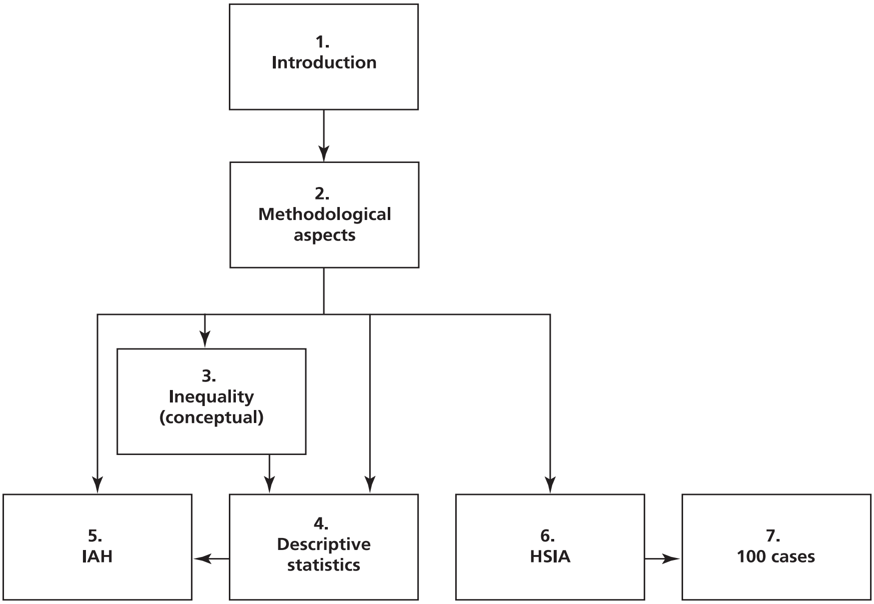


Fig. 1.1 Relationship between the seven chapters of this dissertation

Chapter 2 starts with a description of the way happiness is usually measured at the individual level. The individual responses in a sample are collected and should be converted into information on the happiness distribution of the nation represented by that sample. This chapter is devoted to a number of methodological issues and problems that are (to be) encountered in this process.

A conceptual approach of inequality is presented in chapter 3. The first part deals with inequality of measured happiness within a sample. Inequality is described as a binary relation on the set of all individual responses. The total amount of inequality has a minimum (zero) value when all respondents report the same happiness rating. What is more interesting is that there is also a maximum value for this inequality in a sample. They are related to both the mean (absolute) pair distance and the standard deviation of the frequency distribution of the happiness ratings in the sample. The results of this approach will be used in chapter 4. In the second part of chapter 3, a method for the quantification of inequality of the distribution of a continuous variable in relation to its probability density function is described. The method is illustrated for some simple linear probability densities, and is also applied to the beta distribution, since the latter will be introduced in chapter 6 as a possible model for a continuous latent happiness variable.

We shall examine a number of current descriptive statistics for their aptness to characterize the dispersion in the sample of the results of happiness measurement in chapter 4. To this end, a number of criteria are formulated and applied to the statistics obtained by application to a number of hypothetical frequency distributions. This enables us to demonstrate that there is no reason to discontinue using the standard deviation as the measure of happiness inequality within a nation, since for happiness distributions it appears to be highly superior over e.g. the Gini coefficient. Some special attention will be paid to the reasons why this coefficient, which is very popular among economists, fails in this respect.

As has been pointed out in section 1.5, utilitarians and egalitarians have different views on what is the most desirable happiness situation of a nation. We introduce the concept of the Inequality Adjusted Happiness (IAH) in chapter 5, proposing a new index of societal performance on the basis of both the average happiness value and the dispersion in the sample, which also allows us to give different weights to both views on happiness.

Methods that (cl)aim to solve a number of the methodological problems of measuring happiness, mentioned in chapter 2, are discussed in chapter 6. In particular, we will pay attention to what is known as the “Happiness Scale Interval” approach (Veenhoven, 2009). In this project, each of the members of a panel is presented with a verbal happiness measurement scale and a continuum ranging from what they see as the most happy situation they can conceive to the most unhappy at the other end of the scale.

The instruction is to partition the scale according to the number of categories, each with their own label, and to identify the positions of the boundaries between the categories on that scale as they judge these. We then develop a methodology to process the observational results, first those of the panel of judges, and subsequently those of a sample from the same nation. This method appears to enable us to bridge the gap between happiness as it is measured in a sample on one side and the distribution of happiness within the population on the other.

An explorative evaluation of 100 of the first cases, in which this scale interval method has been applied to 52 different happiness questions and to 20 existing cross-national happiness studies, is given in chapter 7. Recommendations have been formulated on the basis of these findings, not only with respect to similar studies in future, but also for further methodological studies on the measurement of happiness both within and between 'nations'.

The dissertation is completed with eight appendices on various issues occurring in the above mentioned chapters.

✧ ✧ ✧

Chapter 2

Methodological Aspects of Happiness Measurement

Parts of this chapter have been published in:
Kalmijn, W.M. & Veenhoven, R. (2005)
Measuring Inequality of Happiness in Nations. In Search for Proper Statistics.
Journal of Happiness Studies, **6**, (4), 357 – 396
ISSN 1389-4978 (Print) 1573-7780 (Online)
DOI 10.1007/s10902-005-8855-7

2.1 Introduction

In this chapter we start with a description of the dominant method for measuring happiness in cross-national studies, with a focus on the application of questions with verbal response categories. We decided not to re-investigate the validity of this approach, but in section 2.3. we list the main assumptions we made for our own study.

Special attention will be paid to the properties of happiness as a variable, which also depends on the view on the nature of happiness. As a contribution to clarify this, we shall compare happiness to income. The major differences between the two require a serious analysis before methods of investigation into one of them are applied to the other one. For now, we have adopted the view to consider happiness as mainly an intensity variable, but we allow for developments in a different direction. The view on happiness is directly linked to the design of the questions used to measure it and a wrong match is expected to give rise to serious problems in the measurement of happiness.

Application of the above mentioned method of measuring happiness gives rise to two major problems. One is that in this way happiness is measured at a level of measurement that is essentially ordinal, i.e. nonmetric. The application of the desired arithmetical operations, however, requires the data to be at the metric level of measurement; the problem of how to bridge this gap is known as the *cardinalization problem*.

The other one is the very large number of combinations of a question and all admitted responses to that specific question. This number, several hundreds, makes it hardly possible to compare and to combine the findings obtained from different studies. The current way-out for these two problems will be described, and in chapter 6 it will be subjected to a critical analysis and more valid alternatives will be presented.

Another methodological problem, which crops up in chapter 4, concerns the (in)dependency of location statistics of a sample happiness distribution, such as the average value, and the statistics that are used to quantify the inequality of that distribution. A distinction is made between various forms of dependency; these should be encountered in different ways.

2.2 The measurement of happiness

Different methods are in use for measuring happiness. Sometimes the happiness of individuals is assessed indirectly, e.g. by content analysis of interviews or diaries. In other situations the hedonic level is judged by external observers. The vast majority of happiness information, however, is obtained as the responses of individuals to questions on their own happiness. These questions may be either single questions or multiple ones, the single questions being the most frequently applied, at least in cross-national studies.

A typical example of such a frequently used question is: “Taking all things together, how would you say things are these days - would you say you are ... ?” The respondent will then be asked to make a choice out of e.g. four possible ratings:

- ☐ “unhappy” (R_1)
- ☐ “not too happy” (R_2)
- ☐ “pretty happy” (R_3)
- ☐ “very happy” (R_4)

In this example, happiness is rated by the respondent on a four step verbal rating scale. We use the term “verbal” for written textual responses, as it is used in the World Database of Happiness (WDH), introduced in section 1.2.

In this dissertation, the possible ratings are referred to as ‘categories’ and their descriptions as e.g. “pretty happy” as ‘labels’. The term ‘categories’ stems from the name “the method of successive categories”, as is in use for the above method of measurement among psychometricians; see e.g. Guildford (1954, Ch. 10).

In this thesis, we confine ourselves to single closed questions with multiple choice answers as in the above example. More specifically, we focus on questions with verbal responses, although some of our findings will be also applicable to other responses, such as pictorial ones. Within this specific context, we define an "item" as a combination of one question and a specified number of mutually exclusive response categories, all with a completely defined content. In principle, measurements with verbal responses in different languages will be considered to be manifestations of the same item if their translation into English is identical.

With this reference to English, a link is made to the WDH, where (US) English is the standard language. One of the collections of the WDH deals with "measures of happiness". This concept includes the application of what we have called "items" above, but it is a wider concept, since it also covers other methods used to assess happiness besides self-report to single questions. In other words: our "items" are a subset of the "measures of happiness" in the WDH.

All measurements of happiness that have been reported in at least one publication and fit the definition of happiness as described in section 1.2 are to be gathered in the collection of 'Happiness Measures' [2] of the WDH and ordered by their measurement code.

> *This code is unique for each measurement of happiness and contains information on the contents of the measurement in a standard notation; the item in this section is coded O-HL/c/sq/v/4/g. Since the precise meaning of the code is not relevant in this context, it is not described here; the code only acts as a reference to the List of "Measures of Happiness" in the WDH, where a complete description and explanation is available under "Classification"* [2].

The number of measures of happiness is still increasing. The studies of which the findings have been entered in the WDH include 685 such measures, 296 of which are single verbal questions (June 2010). Note that occurring in a publication and being a measure of happiness, life satisfaction etc. are the only criteria for registration, so inclusion in the Item Bank as such cannot be considered as a quality label of the aptness of that measure to assess quality of life.

[2] *Available at: http://www.worlddatabaseofhappiness.eur.nl/hap_quer/hqi_fp.htm*

In most items, the respondent has to select one out of a limited number of discrete ratings. In the above example, the four possible responses are denoted as R_1, R_2, R_3 and R_4 respectively. In general we shall use the symbol R_j for the j-th response, being a member of a set of k possible alternatives, written as $\{R_j \mid j = 1(1)\ k\}$; in the above example $k = 4$.

In this example, R_1 corresponds to the most unhappy situation and R_k to the happiest one. This is the most frequently occurring choice; in this chapter, we will assume that this choice has been made. In case of a scale with R_1 as the happiest situation, a simple reversion of the order of the code numbers will enable the application of the methods to be described in this dissertation.

In a survey, happiness questions of the above type are presented to members of a sample from a population, e.g. some nation, to obtain information about the happiness situation in that population as a whole. The happiness distribution of such a nation is defined as the probability distribution of the individual happiness values of all members of that nation. This distribution is unknown, but it has to be estimated from the frequency distribution of the individual happiness values in the sample that is assumed to represent that population. The central issue in this dissertation is how to convert this sample information into valid and useful information on the population that is represented by the sample in which the measurements were performed.

The basic results in this type of investigations are the counted absolute frequencies $\{n_j\}$ at which members of that sample with size N select one out of the k alternatives $\{R_j \mid j=1(1)k\}$. Respondents who report "Don't know" or who do not make a choice are ignored in this context.

From these absolute frequencies, we can compute the k relative frequencies $\{f_j := n_j / N\}$ and the k cumulative relative frequencies $\{F_j \mid j=1(1)k\}$, which in the above example are defined as:

$F_1 := f_1$

$F_2 := f_1 + f_2$

$F_3 := f_1 + f_2 + f_3$, and

$F_4 := f_1 + f_2 + f_3 + f_4$ (:=1)

while the symbol ":=" means "is defined as". In general $F_j := \sum_{i=1}^{j} f_i$

In this way, the total basis information can be summarized as $\{N; F_j \mid j=1(1)k\}$ under the condition $0 \leq F_1 \leq F_2 \leq \ldots \leq F_{k-1} \leq F_k := 1$.

An example is given in Table 2.1

Table 2.1. Example of a frequency distribution for the happiness distribution (sample of size 200; using a four point rating scale).

code j	response category label	frequency absolute n_j	relative f_j	cumulative F_j
1	"unhappy"	40	0,20	0,20
2	"not too happy"	60	0,30	0,50
3	"pretty happy"	80	0,40	0,90
4	"very happy"	20	0,10	1,00
	TOTAL	N = 200	1,00	

In Fig 2.1 the frequency distribution f_j and the cumulative frequency F_j of the distribution in table2.1 have been plotted against the code number (j) of the response category. Note that for j=2 according to table2.1 F_j =0,5 and not 0,2.

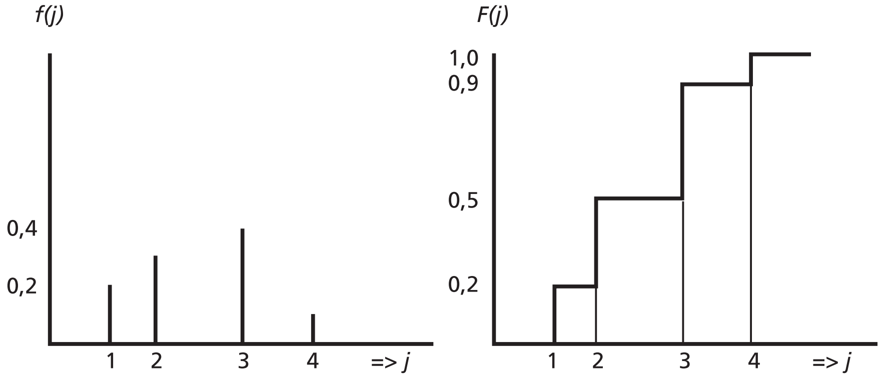


Fig 2.1. Frequency distribution (left) and cumulative frequency distribution (right) for happiness measurement of a sample using a four-point rating scale. In these diagrams, the scale ratings are not equidistant. The upper value of F(j) applies at all four steps in the right-hand diagram,

The usual method for processing measurement results, such as those shown above, is a simple straight-forward four-step procedure, which can be summarized as follows.

Step 1: transform the text of the labels into numbers by coding, replacing each of the k labels with its position in the list of categories as its code, i.e. its rank order number j in the left hand column of table 2.1;

Step 2: calculate the average value as the weighted average of the *k* code numbers, using their relative frequencies *f* as weights;

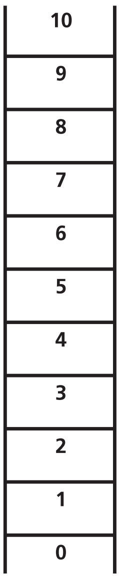
Step 3: calculate the sample standard deviation *s* in the usual way;

Step 4: report the statistics obtained in step 2 and 3 as estimates of the corresponding population parameters.

For the above example, the estimates of the population mean and the population standard deviation value are 2,4 and 0,92 respectively. We will refer to this method as the "*traditional method*". In section 2.5, we will consider some of the methodological problems of this approach. In chapter 6 we will go into this in more depth and also discuss alternatives to the traditional method.

Different scales of --measurement

The above way of measuring happiness is one of the very many scales of measurements that have been applied ever in this area. Another example is the adapted version of Cantril's self-anchoring ladder rating of life (Cantril,1946; Kilpatric & Cantril 1960); see Fig. 2.2 .



The respondent is presented with the question: "Here is a picture of a ladder. The '10' at the top of the ladder means the best possible life you can imagine. The '0' at the bottom of the ladder means the worst possible life you can imagine. On which place of the ladder is your life as a whole? Please mark the number that best corresponds with how you feel about your life now.".

In this case an 11-point pictorial scale is presented to the respondent.

Fig 2.2. Cantril's Ladder Scale

2.3 Assumptions on the measurement of happiness

In this thesis, we make a number of assumptions, which are summarized as follows:

- all respondents are fairly familiar with concepts as happiness, life satisfaction etc., and their associations are sufficiently close to those of the researcher who conducts the study
- respondents are reasonably well capable to judge their own happiness situation
- participants are presented with a clear and unambiguous instruction
- respondents are prepared to perform their tasks in a serious manner
- respondents are physically, mentally, culturally and linguistically capable to understand the instructions and to perform their task accordingly
- respondents perform their task independently of each other
- responses have a perfect repeatability, i.e., in the hypothetical situation that a respondent is presented with the same item within a short time interval, say the same day, and any 'memory effect' is completely eliminated, the same happiness category will be selected
- the composition of the sample is in a good agreement with that of a random sample from the same target population

Most of these assumptions are debatable to some extent, but this debate is beyond the scope of this thesis and in this, we will accept the assumptions unconditionally. For a discussion on their plausibility, the reader is referred to e.g. Veenhoven (1984 and 2002a).

The above list of assumptions is extended with one conditional assumption, which will be discussed in section 7.5:

- in general, respondents are capable of discriminating between gradations of happiness on the basis of the labels of ordered alternative categories

The term "in general" is included, since this assumption does not necessarily apply to ill-constructed items.

2.4 Happiness as a variable

2.4.1 Views on the nature of happiness and satisfaction

Although a fundamental discussion on this topic is beyond the scope of this dissertation, we cannot avoid the need to make a few comments on the basis of our findings that are relevant in this context. A crucial point is the answer to the question how one considers an unhappy person. Is he a person with a relatively low happiness intensity ? Or is he somebody in which the balance between happiness and something like ‘anti-happiness’ is in favour of the latter ? The latter approach assumes the existence of some autonomous entity that acts as a counterpart of happiness/satisfaction; what is measured eventually is the net result of both. See Fig. 2.3

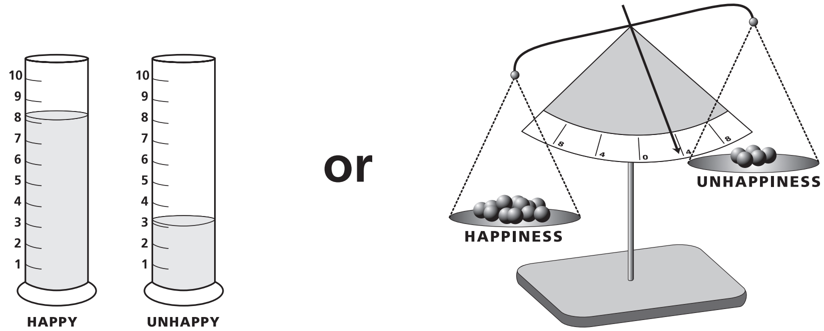


Fig. 2.3 Which model for happiness ?

In order to clarify this problem, we shall consider, in section 2.6, happiness in contrast to income as a variable with a quite different nature.

2.4.2 Intensity versus extensity variables

In the physical sciences a distinction can be made between *intensity* variables and *extensity* variables; the latter class is also referred to as *capacity variables*. Examples of extensity variables are mass, weight, volume, amount of heat, electrical charge etc.; pressure, temperature and density are intensity variables.

An example may clarify this distinction. Suppose one has two vessels. Let the first one contain 1 kg (≈ 1 litre) of water at a temperature of 40°C and with a density of (almost) 1000 kg/m^3. The second is filled with 3 kg of water at 20°C, having almost the same density and a volume of 3 litre. If both liquids are poured into a third container, we get 1 + 3 = 4 kg, with a total volume of 1+3 = 4 litre. However, measurement of the temperature will not result in 40+20 = 60 °C, nor will the new density be about 1000 +1000 = 2000 kg/m^3. The values of these variables of the mixture are 25 °C and 1000 kg/m^3 respectively. In principle, values of *extensity variables* may be largely added in the case of combination of quantities. In such cases, for *intensity variables*, one may generally expect a result that is close to the weighted average value, unless ‘interaction phenomena’ occur such as contraction and/or chemical reactions. The above considerations can also be applied to variables in the behavioural sciences, such as happiness and incomes.

Incomes can be measured by *counting* the number of euros earned by an individual or a set of individuals over a period. The incomes of individuals can be aggregated, e.g. to give a household income or a national income. In principle, incomes can be transferred from one person to someone else, e.g. in the case of taxation within one family, but also, for example, when a plumber appoints an assistant. So incomes have all characteristics of extensity variables.

In contrast, we consider measures of happiness to be intensity variables, at least in the way it is measured as described in section 2.2. In our view, rating your happiness is not just counting your blessings. Somebody who is unhappy has a low happiness level. Being able to ‘transfer happiness’ from one person to an other is something difficult to imagine. This difference has the consequence that operations with respect to income cannot always be transferred to happiness situations automatically.

If Peter has an annual income of € 30.000 and his wife Anna € 40.000, then the family income is € 70.000/year. But if the happiness of Peter and Anna are 3 and 6 respectively on a [0, 10] Cantril scale, it is not very sensible to say that they are very happy family with happiness rating 3+6=9, nor that Anna is twice as happy as Peter is.

Theoretically an approach similar to the second one given above for income distributions could be applied to happiness, provided one is prepared to assume the existence of something like a ‘happiness pie’ and of amounts or ‘quantities’ of happiness, consisting of a kind of ‘happiness molecules’, if not a priori, then at least a posteriori.

As an example, Yew-Kwang Ng (1996: 1 - 28), thinking along this line, assumes that everybody acquires a number of such units (‘utils’) of happiness, but also of unhappiness. A subject rating his happiness is assumed to count the number of such ‘pleasure utils’, collected over some period, as a measure of his happiness. In the same way, counting the number of ‘pain utils’ over the same period results in an ‘amount of unhappiness’, in such a way that one util of pleasure is neutralized exactly by just one util of pleasure. In his approach, unhappiness is just negative happiness or anti-happiness. The net algebraic sum of the happiness and unhappiness utils quantifies the subject’s average happiness over that period. However, the claim of Yew-Kwang Ng that this ‘net happiness’ is a variable at the ratio level of measurement, has to be considered incorrect.

2.4.3 Unipolar versus bipolar happiness scales

There is a direct impact of the view taken on happiness on the scale type to be selected for its measurement. If happiness is viewed as an intensity variable, a unipolar scale type is appropriate. There is only one pole, that of happiness. The item in section 2.2 is an example of such a scale. The higher the code number of a category, the larger the intensity of happiness/satisfaction as it is experienced by the respondent. For such scales, the lowest value represents the minimum intensity and a very frequently chosen label for this category includes the words “not at all”.

Measure O-HL/g/sq/v/7/a, however, is a typical example of a bipolar scale. The lead question is: “If you were to consider your life in general, how happy or unhappy would you say you are, on the whole?” and the seven categories are:

7 = ”completely happy” 1 = ”completely unhappy”
6 = ”very happy” 2 = ”very unhappy”
5 = ”fairly happy” 3 = ”fairly unhappy”
4 = ”neither happy, nor unhappy”

There are two poles: a happy and an unhappy one; the anchor point in the middle of the scale obviously represents the equilibrium between them. In principle, such a scale should be symmetric with respect to this neutral category and if a 7-points bipolar scale is chosen, its more usual form for application in other situations than with respect to happiness is [+3, +2, +1, 0, -1, -2, -3]. When applied to the measurement of happiness, however, all these scale point values are augmented by 4. In this way negative and zero ratings are avoided. This has both calculation advantages and psychological advantages, if one assumes that some people are reluctant to report negative responses.

The unipolar approach seems to fit best to the view on happiness as an intensity variable. If a respondent has adopted this view and subsequently is presented with a bipolar scale, he may feel responding to this scale is a difficult task, and problems in its execution are not unlikely. We will discuss this problem in chapter 7 in more detail. Yet a view based upon a (in)balance between “pleasures and pains” (Bentham) suggests the application of a bipolar scale is more appropriate. In their choice, authors either consciously or unconsciously, but anyhow at least implicitly adopt one of the two views on happiness as dominant. For meta-analysis, their disagreement is most unpleasant, but apparently unavoidable as long as no consensus on this has been reached. We do not exclude the possibility that the choices for happiness and for life-satisfaction-in-the-narrower-sense eventually may be different. We do not even reject the option that both views are true, but at the same time incomplete. Then the situation will determine which one is the more useful one. However, as has been put forward already in the beginning of section 2.4.1, a further discussion on this is beyond the scope of this study.

2.4.4 Level of measurement

The measurement practice described in section 2.2. has consequences for what is referred to as the level of measurement. Happiness is measured as a discrete variable, which is self-reported as one of a small number of response categories. Categories of discrete variables are either unordered or ordered. If they are not ordered, the level of measurement is referred to as "nominal". At the ordinal level of measurement, the categories are ordered by definition.

In some cases the additional assumption may be justified that the categories have either equal or unequal but known mutual distances on some underlying metric scale; such cases are sometimes referred to as "pseudo-metric". The terms "nominal" and "ordinal level of measurement" and the underlying principles stem from Stevens (1946) and are fundamental in our considerations.
Schematically:

variables and their level of measurement

variable	discrete		discrete or continuous	
level of measurement	nonmetric		metric	
	nominal	ordinal	interval	ratio
numbers	ordinal code numbers		cardinal numbers	
calculation operations	not admissible		admissible	

In the above scheme, variables at the metric level can be either discrete or continuous. Happiness measurements however, are always discrete, even when they have been 'upgraded' from the ordinal to the interval level. The continuous variant of happiness occurs only in models in which a latent happiness variable is postulated as a continuous one, such as in section 6.4.

2.5 Problems with the application of verbal scales of measurement

We described the usual way for measuring happiness in section 2.2. In the further processing of the measurements two major problems arise. Since in practice the solutions to both problems are interconnected, we shall discuss them jointly.

The first problem is that we just gave only one example, but for verbal questions, this is a choice of hundreds of alternatives. The 7-point item in section 2.4.3 is an example of one of these.
These alternatives differ in numerous aspects, such as

- the object of study, e.g. whether the subject is asked to rate his happiness or his life-satisfaction-in-a-narrower-sense
- the period concerned by the question (if specified)
- the precise formulation of the lead question
- the number of categories
- the labels of the categories in case of verbal categories and the corresponding position in the ordered list of categories.

This heterogeneity hampers the comparison and the combination of results, since different results are in general obtained by applying different items.

The other problem is caused by the fact that happiness is measured by verbal items at a level of measurement that is essentially ordinal; see section 2.4. As has been pointed out in section 2.2, the traditional method for e.g. calculation of average values, starts by converting text to numbers. This is done by coding in the most simple way: each label is replaced by its position on the rating scale. The next necessary step is to convert these code numbers into metric numbers, to make admissible the arithmetical operations that are required for the calculation of average values, standard deviations etc. The problem how to convert ordinal numbers into cardinal numbers is known as the *cardinalization problem*. In the traditional method, the solution of this problem is to simply declare the distinction between ordinal and cardinal numbers non-existent, either due to mathematical ignorance or by denying it as long as no alternative solution is available.

In chapter 6, we shall discuss the consequences of this ‘solution’, and present alternative ways of dealing with this problems with their pros and cons. In the present chapter, we confine ourselves to a description of the traditional solution. The result of the above procedure is a rating scale with k equidistant points {1(1)k}, so with a width = k–1; this equidistance is actually introduced by the coding. In this way happiness is considered and measured as a discrete variable, which can adopt only a limited number (k) of different values, which number has been chosen by the investigator. The responses are reported as ratings, which are coded as numbers, R_j being reported as a “rating = j”. Having adopted this solution, an average happiness within the sample can be calculated by weighing each of the k ratings with the corresponding counted relative frequency. The result is a, generally broken, rational number in the interval [1,k]. The standard deviation of the measured distribution can be calculated in a similar way.

The solution of the many-items-problem is to convert the ratings and the corresponding statistics on all different ‘primary’ rating scales to a single common ‘secondary’ scale, for which a [0, 10] scale is the conventional choice, with “0” as the most unhappy and “10” as the most happy situation. This conversion is established by a linear transformation of the primary scale and the corresponding rescaling procedure is known under various synonymous terms such as “linear stretching”, “direct stretching” or “direct rescaling”. The procedure assumes a solution of the cardinalization problem including equidistance of the category ratings. This procedure is described in Appendix B, and will be discussed in chapter 6.

The result of this transformation in the case of the item in section 2.2. would be that the category “unhappy” is replaced with the rating 0,00, and the three other categories with the ratings 3,33, 6,67 and 10,00 respectively.
Note that after this transformation, individual happiness is still measured on a 4-point scale, albeit with different ratings, but which not is modified into an 11-point scale.

2.6 Happiness versus income

Over the last two decennia, economists have taken an increasing interest in happiness. This has also had methodological consequences, not only with respect to the choice of the preferred statistics and statistical techniques, but also for the view on happiness and even the jargon that is used, e.g. the expression “concave function” is sometimes replaced with “diminished return” in a situation without any return. It is not surprising that economists are naturally inclined to consider happiness, at least as it is measured, primarily as something with identical or almost identical properties to income, properties they are familiar with.

However, not every economist appears to be sufficiently conscious of the existence of various fundamental and partly interdependent differences between happiness and income and their distributions in methodological respect. Therefore, these differences deserve some special and explicit attention, which is given in this section.

In section 2.4.1, we discussed Peter and Anna’s incomes and their happiness and the differences between them to make it clear that income should be considered an extensity variable against happiness as an intensity variable, at least in our view. In relation to this, we have rejected the idea of the existence of some ’happiness pie’ to be distributed over all citizens of a nation without doing the same for an ‘income pie’. Additionally, this view implies that happiness cannot be transferred to someone else, whereas there are certainly opportunities to do this with respect to incomes.

In principle, an income can be measured by counting the number of euro’s, so it has a natural zero, but no natural maximum value, at least not in theory. This makes income a variable at the ratio level of measurement.

Whereas in all income distribution models the income variable is defined on the $[0, \infty)$ interval, happiness is defined on a closed interval, either [0, 10] or $[1,k \mid k \in \mathbb{N}]$.

Even if the happiness value is considered to be metric, this is at best at the interval level of measurement. The ‘lower end’ of the scale corresponds to the most miserable situation a respondent can imagine, but this cannot be considered objectively as an absolute zero.

Since the level of measurement determines which statistics are admissible at that level and which are not, this implies that statistics for income distributions are not necessarily applicable to happiness distributions, and the reverse is equally true.

The next difference is that incomes are expressed as a combination of a number and a unit, e.g. 30.000 euro/year. Therefore, incomes can be compared only if both components are taken into account, which can introduce the need for some way to scale standardization. Happiness values, however are expressed as numbers without a unit and the common secondary [0, 10] scale ensures standardization in an entirely different manner. In contrast, due to the use of units, incomes are measured objectively, at least nominal incomes. Happiness values, however, are subjective as the respondents 'observation' of the scale end points is completely subjective: there is no way to compare the happiest situation Peter can conceive with that of Anna and subjective measurement of happiness only pretends to estimate the position between the individually perceived end-of-the-scale happiness situations.

Summarizing gives the following list of - interconnected - differences:

INCOME	**(MEASURED) HAPPINESS**
objectively measurable	subjectively rating
(counting)	(selecting from alternatives)
level of measurement: ratio	ordinal or interval
(almost) continuous	measured as discrete
number x unit (e.g. €/year)	dimensionless number only
transferable (to some extent)	not transferable
aggregation meaningful concept	aggregation meaningless
model: income $\in [0, \infty]$	happiness $\in [1, k]$ or $\in [0, 10]$
natural zero	"0" is not a natural zero
extensity variable	intensity variable
	(albeit not uncontested)

One more aspect is that among economists there is a much stronger association between "distribution" and "to distribute something", in the sense of partitioning an amount of something over a number of subjects or parties, the 'pie concept'. In this way, annual world copper production is distributed over nations and annual tax return is distributed over the various government departments of that nation. As a consequence, the term "income distribution" in economics has at least two different interpretations, which should not be confused. The cumulative frequency distribution may be defined for an income of e.g. € 30.000 / year as either

(1) the proportion of inhabitants whose income ≤ € 30.000 /year, or
(2) the proportion of the national income that is earned by all inhabitants together with an income ≤ € 30.000 /year each.

We denote these two cumulative frequencies for income = I as $F(I)$ and $\Phi(I)$ respectively. Clearly, $\Phi(I) < F(I)$. The two proportions can be linked by calculating both $F(I)$ and $\Phi(I)$ for a not too small number of different income values I. Plotting Φ (·) as the ordinate against F(·) for all these points and connecting these points results in the *Lorenz curve*, which plays an important role in the quantification of income inequality. Some economists recommend also applying this method to happiness inequality. In this case and applied to the happiness distribution given in section 2.2, the two cumulative distributions as defined above, albeit for happiness now, would result in four points in Fig. 2. 4, one for each category. Their coordinates in rhe Lorenz curve diagram are presented in table 2.2. One of these four points, the point (1,1), is trivial.

Table 2.2 Construction of Lorenz curve for happiness distribution

category		**cumulative distribution**		
J	f_j	jf_j	$F(j)$	$\Phi(j)$
1	0,2	0,2x1=0,2	0,2	0,2/2,4=0,08
2	0,3	0,3x2=0,6	0,5	0,8/2,4=0,33
3	0,4	0,4x3=1,2	0,9	2,0/2,4=0,83
4	0,1	0,1x4=0,4	1,0	2,4/2,4=1,00
Sum	1,0	2,4		

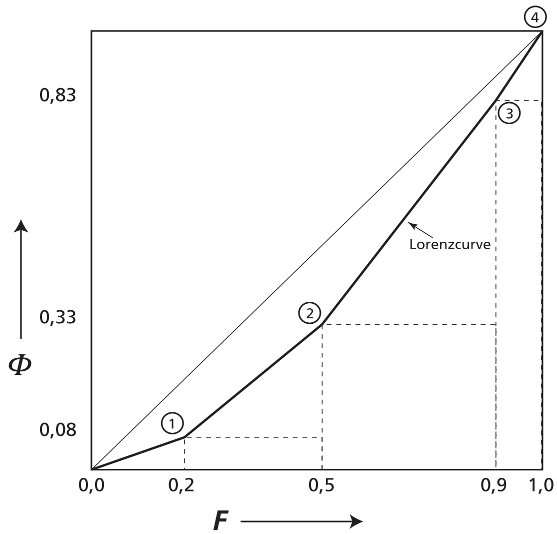


Fig. 2.4. Lorenz curve for the happiness distribution of table 2.2.

If one connects the consecutive points by straight lines, the resulting broken line may be considered as the Lorenz curve for this particular case. In section 4.7 we shall discuss to what extent this approach is admissible from methodological point of view.

As regards happiness in this thesis, we will in general use the term "distribution of happiness" only according to $F(\cdot)$ as given above. The cumulative frequency for the happiness rating "3", denoted $F(3)$, is defined as the proportion of the sample members who have reported a happiness rating ≤ 3. The distribution function $\Phi(\cdot)$ is introduced only incidentally in the discussion and on these occasions it will always be announced explicitly, as for example in section 4.7.

2.7 Happiness as a continuous variable

Until now in this chapter, happiness has been described a discrete variable which can adopt only a small number (k) of distinct values. Quite incidentally, scientists suggest in publications that in their view happiness could be considered to be a continuous rather than to be a discrete variable. This continuous happiness variable is measured using a discrete rating scale in such a way that a respondent selects a rating of e.g. "*j*" if he considers the label of that rating as a better qualification of his happiness feeling than the adjacent ones, even if the label of rating "*j*" also does not fit perfectly. In this approach. A rating "*j*" actually covers some 'range' on the continuous happiness scale around that number, while the exact position within this range for this person is completely unknown. However, even if authors express this view, either explicitly or implicitly, they allow for this concept only to a limited extent.

In the above context, the most frequently observed approach is the one in which a rating with code j also covers all happiness values that are closer to the value j than to any other rating. The expression "closer than" implies a metric view on the happiness scale. In practice, this means that a rating j in the case of an equidistant scale is considered to cover all happiness values between $j - ½$ and $j + ½$. When this view is also applied to both terminal ratings "1" and "k", this implies that the effective scale length is extended by one unit: $[1, k] \rightarrow [½, k + ½]$.

Whereas an adequate graphical representation of the discrete distribution is as a bar chart, the continuous distribution as described above should be represented by a histogram (Fig. 2.5 for the example of section 2.2). In this histogram the height of each component is equal to a corresponding relative frequency, making the total area of all *k* components equals unity.

Fig. 2.5. Graphical representation of the happiness distribution of table 2.1 as a bar chart and as a histogram in case of a discrete c.q. continuous distribution.

An example of this approach is found in Ventegodt (1995). We illustrate the consequences for the linear transformation of the scales of measurement in his study in our Appendix B. A second example of this approach is that of Langford (2006), who translates the above views into consequences for the calculation of quartiles and the interquartile range as a measure for happiness inequality (see section 4.2.5).

The conversion of discrete into continuous has no consequences for the estimation of the mean happiness in the population, at least not in the variant as described above. More precisely: the sample average value that is obtained according to the traditional method is an unbiased estimator of the mean value if the above continuous model is adopted. Presumably this is generally assumed unconsciously or intuitively, but a more formal way demonstration is given in Appendix F.1 (Eq. [F.7]). Appendix F also contains a demonstration of how a similar statement with respect to the standard deviation would be invalid. Even in the case where the notion of continuity is observable, until now we have not find a reported standard deviation other than obtained using the traditional method. The continuity of happiness is a concept that will receive extensive attention in chapter 6.

2.8 Dependency between inequality statistics and the mean value

Various descriptive statistics will be assessed in chapter 4 for their aptness to quantify the inequality of happiness as measured in a sample. One of the criteria is the dependency between the candidate inequality statistic and the mean happiness value. Whether or not this exists is usually an empirical issue, but whenever it is found, the question arises whether the way happiness is measured is responsible for this finding or is there an intrinsic dependency, i.e. caused by socio-economic conditions which influence both the level of happiness and its inequality. In this context, we feel the need to have a closer methodological look to different kinds of (in)dependency and to distinguish these in relation to happiness. The views will be illustrated below for the situation where the standard deviation is used as a statistical measure of dispersion.

Stochastic dependency

First, due to the sampling process, there may be a stochastic dependency between the two statistics: when they have been computed from the same set of observations, they will have a simultaneous statistical distribution, one of the characteristics of which is a covariance between both statistics. This kind of dependency deals with the situation in which different samples are taken from the same population.

If a random sample consists of observations from a normally distributed population, the statistics will have a zero covariance and will be stochastically independent. (Cramér, 1946: 382). In this case, 'independency' means that, when due to sampling errors an accidentally higher average value is obtained, this does not give rise to a systematically higher or lower value of the estimated standard deviation.

If the distribution is skewed, a stochastic dependency exists, in the sense that for positively skewed distributions, 'skew to the right', higher mean values correspond to systematically higher values of the standard deviation (Keeping, 1962: 110). This may be expected to occur at low mean values in view of the existence of a lower boundary for the mean value.

Structural dependency

A different type of dependency is the one that arises from the way happiness is measured. This type of dependency will be referred to here as 'structural dependency'. Since happiness is measured on a rating scale with both a minimum and a maximum possible rating, the variance (s^2) and the standard deviation (s) have theoretical maximum values that are dependent on the mean value m.

In Appendix A, it is demonstrated that:

[2.1] $0 \leq s \leq \sqrt{(h - m) \cdot (m - u)} \leq \frac{1}{2} \, | h - u |$,

where *h* and *u* are the ratings corresponding to the highest and the lowest possible degree of happiness on the scale of measurement, either without or after linear scale transformation. The largest possible value of *s* is reached at a mean value at the middle of the rating scale and diminishes to zero at both ends of the scale. The maximum values of the variance and the standard deviation for various values of the mean value *m* in case of a scale with lowest and highest possible ratings of 0 and 10 respectively and which scale is denoted as a [0, 10] scale, are listed in table 2.3 as an example.

Table 2.3 Maximum values of the sample variance and standard deviation, for different mean values on a [0, 10] scale

	Mean	Maximum variance	Maximum standard deviation
Minimum	0	0	0,00
	1	9	3,00
	2	16	4,00
	3	21	4,58
	4	24	4,90
Top	5	25	5,00
	6	24	4,90
	7	21	4,58
	8	16	4,00
	9	9	3,00
Maximum	10	0	0,00

Apparently, in this case for 2 ≤ mean ≤ 8, the theoretical maximum value of the standard deviation varies between 4 and 5. Although both the mean and the standard deviation are bounded statistics, in practice, these boundaries are (i) only modestly dependent on the value of *m*, and (ii) fairly remote from almost all empirical values of the statistic in nations studies (1,4 – 2,9 on this scale).

A more serious limitation of the scope for the standard deviation may occur at its lower bound. This effect depends on the value of *k*, as is demonstrated in Appendix A. In the case of a four point scale, the minimum attainable standard deviation can reach the value 1,67 under unfavourable conditions, which is much closer to and even above the values reported on a number of nations in which the happiness inequality is rather modest. It might be alluring to label these effects as 'systematic errors' or' biases', but this would be incorrect. The limitation of the scope for the standard deviation is inherent to the way happiness is measured. If the average value equals its maximum value, the only value the standard deviation can adopt equals zero, but it also should. Any other value would be erroneous. This conclusion, however, does not eliminate the existence of this form of structural dependency as a phenomenon.

A second kind of structural dependency may occur (or not) when there is some proportionality between the location and the dispersion parameter, say between *m* and *s*, but also when the location statistic is present in the definition of the dispersion statistic, which is e.g. the case with the coefficient of variation *s/m*. If *m* and *s* are independent, *m* and *s/m* are clearly not, since *m* occurs in both statistics *m* and *s/m*. If however, *s* is proportional to *m*, *m* and *s* are dependent, but in this case, *m* and *s/m* are not. Selecting *s/m* as an appropriate dispersion statistic would eliminate the dependency between the location and the dispersion parameters in this particular case.

In practice, the latter kind of structural dependency between *m* and *s* is known from measurement errors in various technical measurements, but until now, no indications have been obtained that such a structural dependency occurs in happiness measurement.

Intrinsic dependency

A third type of dependency has already been mentioned, it is the one that arises from the 'fact' that in a society there may be spontaneous or forced socio-economic mechanisms that act to make the level of happiness influence its inequality or the reverse. It is also conceivable that such mechanisms exert their influence upon both the average happiness and the happiness inequality, albeit not necessarily in the same way. This type of dependency will be referred to here as 'intrinsic dependency'. If present, it may give rise to substantial correlations between the mean value and the standard deviation when comparing a number of nations. When Ott (2005) reports a correlation coefficient between the mean happiness level and a within-nation standard deviation of -0,66 for a set of 80 nations, this value seems too large to declare structural dependency exclusively accountable for it. Obviously in this case (also) intrinsic dependency is to be assumed. It is evident that this intrinsic dependency has to be ignored when judging the aptness of the inequality statistic that is used.

✧ ✧ ✧

Chapter 3

Inequality. A Conceptual Approach

Parts of this chapter have been published in:
Kalmijn, W.M. & Arends, L.R. (2010)
Measures of Inequality: Application to Happiness in Nations.
Social Indicators Research, **99** (1), 147-161.
ISSN 0303-8300 (Print) 1573-0921 (Online)
DOI 10.1007/s11205-010-9573-z

ABBREVIATIONS USED IN THIS CHAPTER:

CII	Inequality Index for a continuous variable (section 3.4)
CIV	Inequality Value for a continuous variable (section 3.4)
DII	Inequality Index for a metric discrete variable (section 3.3)
MIV	mid-interval value(s) (section 3.3)
MPD	mean pair distance (section 3.3)
NII	Inequality Index for a nominal variable (section 3.2)
p.d.f.	probability density function (section 3.4)

3.1 Introduction

As has been pointed out in section 1.3, inequality is a core element in happiness research. Quantification of inequality, especially that of happiness inequality, deserves a more fundamental treatise and in this chapter, we will develop such a conceptual approach.

In the present chapter a notion of inequality is proposed which has a clear minimum and maximum value and methods for quantification of that kind of inequality are developed. These methods can be used to gauge the descriptive statistics that are available in standard statistical programmes. This will be discussed in chapter 4, where we will assess a number of current descriptive statistics for inequality of a distribution. In theory the indices developed in the present chapter can also be used as alternative inequality statistics, but we definitely do not recommended their application in daily happiness measurement.

We have described In chapter 2 how happiness in nations is commonly measured. This is typically done in 'survey studies' and the relevant aspects of this technique are that happiness is assessed using self-reports and using samples taken from the adult population from a nation. Happiness is measured as a discrete variable, which is self-reported as one of a small number (k) of response categories.

The k possible response categories are coded and n_j is the counted number of respondents selecting the response category coded with rank order number j, where j=1(1)k. Response "don't know" and non-response are ignored, so the effective total sample size is:

$$[3.1] \qquad N := \sum_{j=1}^{k} n_j$$

This practice has consequences for the level of measurement. Categories of discrete variables are either unordered or ordered. If they are not ordered, the level of measurement is usually referred to as "nominal"; we will consider this level of measurement specifically in section 3.2.

At the ordinal level of measurement, the categories are ordered by definition. In some cases the additional assumption may be justified that the categories have either equal or unequal, but known mutual distances on some underlying metric scale; such scales are sometimes referred to as 'pseudo-metric'. The terms "nominal" and "ordinal level of measurement" and the underlying principles stem from Stevens (1946), see section 2.4, and they are fundamental in our considerations.

Happiness is always measured as a discrete variable, but in the conversion of the sample findings to happiness information about the population represented by that sample sometimes a latent variable is postulated, which is mapped onto the discrete scale of measurement. If this latent variable is continuous, the corresponding level of measurement is necessarily metric. A method is developed on the basis of such a model in Chapter 6. The continuous nature of such happiness variables requires a special method for quantification of its inequality, which is developed in section 3.4. The general result is then applied to five specific models of continuous distributions.

3.2 Inequality at the nominal level of measurement

Although happiness is seldom measured at the nominal level, we will start in this section with some views on inequality at that level of measurement. This is done mainly to introduce some concepts considered fundamental to our approach. We shall describe happiness ratings in a sample in terms of sets and inequality in terms of relations between the elements of such sets, and we will introduce this approach without the complication of ordering. In addition to this didactic reason, this approach is adopted to demonstrate that, in this respect, there is an essential difference between nominal and ordinal situations.

At the nominal level of measurement of some specified variable, e.g. happiness, two respondents within the same sample are either equal or unequal with respect to that variable as they select the same or different ratings if the same [1, *k*] happiness scale is presented to them.

An obvious measure for the inequality can be obtained by describing the happiness ratings of all *N* sample members in mathematical terms as elements of a set and by considering inequality to be a binary relation on that set, i.e. between any pair of these elements. This relation "is unequal to" is symmetric, but neither reflexive nor transitive. In this approach, and in the case of a nominal level of measurement, for each of these N^2 pairs, the inequality relation is either true or false, depending on whether the two selected happiness ratings are different or identical. This inequality relation of any pair can be represented as an indicator variable 0/1, where "FALSE" → 0 and "TRUE " → 1; the outcome is referred to as the inequality value of that pair. Their sum of all N^2 pairs is defined as the total inequality of that sample.

Objections may be raised against the way pairs have been counted resulting in N^2 pairs rather than in the possibly expected 1/2N(N-1) actually different pairs. Therefore we have to define explicitly what is defined a pair in this context. Consider as an example the set {A, B} with N=2 elements only; now N^2 = 4 binary relations can be identified, not only A–B, but also B–A and even A–A and B–B. The third and the fourth pair may be labelled "improper pairs" and could be ignored, since the inequality relation is antireflexive. The second relation (B–A) can also be ignored, albeit for a different reason. Since the relation is symmetric, the relation B–A gives the same contribution to the total inequality as A–B does already. Nevertheless we prefer to count all four pairs as pairs, since (a) this makes the mathematics more convenient for larger values of k, (b) the improper pairs will never be counted as unequal ones and therefore they will not contribute to inequality. Our choice doubles the value of the total inequality, but, as we shall compare the total inequality to its maximum value, the choice will not affect their ratio, which will be used as an inequality indicator. Hence, in this chapter, the total number of pairs is adopted to be N^2 and not the binomial coefficient $\binom{N}{2} := ½N(N-1)$.

We will illustrate our approach by an example with *N* = 8 and *k* = 4: eight ratings, each of which corresponding to one row and to one column, all ratings on a 4-point scale

n_j	2		1	3			2	
j	1	1	2	3	3	3	4	4
1	0	0	1	1	1	1	1	1
1	0	0	1	1	1	1	1	1
2	1	1	0	1	1	1	1	1
3	1	1	1	0	0	0	1	1
3	1	1	1	0	0	0	1	1
3	1	1	1	0	0	0	1	1
4	1	1	1	1	1	1	0	0
4	1	1	1	1	1	1	0	0

Each shaded cell on the main diagonal corresponds to one improper pair. Each cell above this diagonal corresponds to one and only one cell below that diagonal with the same content, so all cells together above the main diagonal contribute to the total inequality by the same amount as all the cells together below this diagonal. Each of all *k* blocks, containing n_j^2 cells and including n_j shaded cells on the main diagonal, has a zero contribution to the total inequality, irrespective of the way the pairs are counted.

We then count the number of 'unequal pairs', i.e. pairs with inequality value =1 as a measure of the total 'amount of inequality' in the set. This statistic will be denoted *S* and equals the sum of the inequality values of all N^2 pairs. In our above example *S* = 46.

The group consisting of all subjects that respond in favour of the same response R_j, has the size n_j, and delivers a contribution to the total inequality that equals $\sum_{i \neq j} n_i = N - n_j$ since:

[3.1] $N := \sum_{j=1}^{k} n_j$

The value of *S* is obtained as the difference between the total number of pairs and the total number of 'equal pairs', resulting in

[3.2] $S := N^2 - \sum_{j=1}^{k} n_j^2$

This result is also made clear by considering the above scheme. In our example $S = 8^2 - (2^2+1^2+3^2+2^2) = 64 - 18 = 46$, the value of which has already been found.

In view of the constraint [3.1], the maximum value of *S* can be found by applying Lagrange's method of undetermined multipliers, i.e. by putting the partial derivatives of:

[3.3] $$FS := S - 2\lambda\left[N - \sum_{j=1}^{k} n_j\right]$$

with respect to all $\{ n_j | j = 1(1)k \}$ equal to zero. For reasons of convenience, we write the multiplier this time as "$- 2\lambda$". The result is

[3.4] $$\frac{\partial(FS)}{\partial n_j} = -2n_j + 2\lambda = 0 \Longrightarrow n_j = \lambda \Longrightarrow n_j = N/k \qquad \forall\, j = 1(1)k$$

The extreme value of *FS*, and therefore also that of *S*, is a maximum since

[3.5] $$\frac{\partial^{\,2}(FS)}{\partial n_j^2} = -2 < 0 \;\;\Longrightarrow\;\; S_{max} = N^2 - \frac{N^2}{k} = \frac{k-1}{k} N^2 \;.$$

The maximum inequality is reached if the observed frequencies are all equal or almost equal.

The addition "almost" refers to the case that k is not a divisor of N.
If N= 40 and k=7, S is maximized in case of a sample $\{n_1, n_2,\ldots,n_7\}$ *=*$\{5,5,6,6,6,6,6\}$ *or some permutation of these frequencies.*

This result enables us to define an index number. We will call it the *"Nominal Inequality Index"*, and denote it *NII*, defining it as a number rounded to integer values:

[3.6] $$NII := \frac{S}{S_{max}} \cdot 100 \text{ , so } 0 \le NII \le 100.$$

Combining of [3.2], [3.5] and [3.6] results in:

[3.7] $$NII = \frac{N^2 - \sum_1^k n_j^2}{N^2} \cdot \frac{k}{k-1} \cdot 100$$

To readers who do not consider equality as a zero-inequality, but as a complementary concept to inequality, the value of 100 – NII might be an option to serve as an indicator for the 'degree of equality', but in our view this is not a recommended practice.

3.3 Inequality at the ordinal and at the discrete metric level of measurement

Contrary to measurements at the nominal level, inequality relations in the ordinal case can be distinguished as either "<" or ">". This is the situation as it occurs in section 2.2 with four ordered categories. The order in such situations is always assumed to be unambiguous.

3.3.1 Assumed equidistance

First we consider the case in which the various ratings are assumed to be equidistant. We shall do so for the item that has been introduced in section 2.2, with k=4 response categories: 1=unhappy, 2=not too happy, 3=pretty happy and 4=very happy. Equidistance means that e.g. the difference between "very happy" and "not too happy" is equal to that between "pretty happy" and "unhappy", whereas both these differences are twice that between "pretty happy" and "not too happy". Under these assumptions, the ordinal numbers of the ratings {1, 2, 3, 4} can be treated as if they were cardinal. This approach will be referred to as the "pseudo-metric" one. In this case, the arithmetical operations which are required for the calculation of average values, standard deviations and that of various other statistics, are admissible.

An obvious way to quantify the total inequality is to apply the procedure that was adopted in section 3.2, but to give the inequality value of each pair a weight according to the distance of the ratings of both members on the happiness scale. A suitable value for this weight is the absolute value of the difference of the ratings. In the above example, a pair consisting of the ratings of an unhappy and a pretty happy person contributes to the total amount of inequality with a weight |1-3| = 2. Along this line, the joint contribution of all individuals with the same rating j to the total amount S of inequality can be written as:

[3.8] $$S(j) = n_j \sum_{i=1}^{k} |j - i| n_i$$

and the total amount of inequality within the sample of effective size N is:

[3.9] $$S := \sum_{j=1}^{k} \sum_{i=1}^{k} |j - i| n_j n_i$$

The maximum value of S can be found by putting the partial derivatives of:

[3.10] $FS := S + 2\lambda[N - \sum_{j=1}^{k} n_j]$

with respect to each n_j separately equal to zero, adopting "2λ" for the multiplier this time.

Table 3.1. Individual contributions to the total inequality S

	n(1)	n(2)	- - -	n(j)	- - -	n(k-1)	n(k)
n (1)	0	1.n (2)	- - -	(j-1).n (j)	- - -	(k-2).n (k-1)	(k-1).n (k)
n (2)	1.n (1)	0	- - -	(j-2).n (j)	- - -	(k-3).n (k-1)	(k-2).n (k)
- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
n (j)	(j-1).n (1)	(j-2).n (2)	- - -	0	- - -	(k-1-j).n (k-1)	(k-j).n (k)
- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
n (k-1)	(k-2).n (1)	(k-3).n (2)	- - -	(k-1-j).n (j)	- - -	0	1.n (k)
n (k)	(k-1).n (1)	(k-2).n (2)	- - -	(k-j).n (j)	- - -	1.n (k-1)	0

In Table 3.1, the j-th row corresponds to the respondents of category j. The sum of all cells in that row is the contribution of a single individual in that category. Multiplication by the frequency, denoted before the left-hand column, results in the total contribution *S(j)* [3.8] of the j-th category. After this multiplication, the total amount of inequality is obtained as the sum S of all k x k cells within the rectangle.

One has to be aware of the fact that, for the differentiation of *FS* with respect to n_j, *after the multiplication* with the n_j , (i) terms with n_j occur in the shaded j-th row and the j-th column only, (ii) the sums of the cells in that column and that row are equal, so their joint contribution to S can be replaced with twice that of the j-th row, (iii) the result of the partial differentiation with respect to n_j can be found in the shaded j-th row *within the rectangle* but for the value of λ, so that the row sum of each row within the rectangle equals λ, and (iv) after multiplication of the shaded row sum by n_j, the sum of all row sums up to the total amount of inequality $\lambda \sum n_j = \lambda N = S_{max}, \Longrightarrow \lambda = S_{max}/N$. Hence:

[3.11] $$\frac{\partial(FS)}{\partial n_j} = 2\sum_{i=1}^{k} |j - i|\, n_i - 2\lambda = 0 \qquad \forall\, j = 1(1)k$$

which can also be written as:

[3.12] $$\sum_{i=1}^{k} |j-i|\, n_i = \lambda \qquad \forall\, j = 1(1)k$$

or as:

[3.13] $$(j-1)n_1 + (k-j)n_k + \sum_{i=2}^{k-1} |j-i|\, n_i = \lambda \qquad \forall\, j = 1(1)k$$

As the reader can verify by substitution in [3.13], the solution of the *k* equations:

[3.12] $$\sum_{i=1}^{k} |j-i|\, n_i = \lambda \qquad \forall\, j = 1(1)k$$

is simply $n_1 = n_k = ½N$ and $n_j = 0$ for $j = 2(1)\ k-1$. Apparently, the inequality is maximal if the sample members are distributed equally over both terminal categories, leaving empty the other *k*-2 categories.
From [3.9], the corresponding maximum value of *S* is found to be:

[3.14] $$S_{max} = \tfrac{1}{2}(k-1)N \ .$$

Now we can define a *discrete inequality index DII* for this situation in a way analogous to the *NII* in [3.6] for the nominal case, by substitution of the results of [3.9] and [3.14]:

[3.15] $$DII := \frac{S}{S_{max}} \cdot 100 = \frac{2\sum_{j=1}^{k}\sum_{i=1}^{k}|j-i|n_i n_j}{(k-1)N^2} \cdot 100$$

In eq. [3.9], one may consider raising the difference $|j - i|$ to some power >1 if more weight is assigned to the distance, or <1 in case of less weight. As long as there is no evidence for such a choice, we maintain the unity exponent value. There is, however, a quite different reason to consider an exponent = 2, since $|j - i|^2 = (j - i)^2$ and in this way one gets rid of the absolute values.
In that case eq. [3.9] is to be replaced with a similar statistic, denoted $S^{(2)}$:

[3.16] $$S^{(2)} := \sum_{j=1}^{k}\sum_{i=1}^{k}(j-i)^2 n_i n_j$$

[3.17] $$S^{(2)} = \sum\sum j^2 n_i n_j + \sum\sum i^2 n_i n_j - 2\sum\sum ij n_i n_j$$
$$= \left[\sum_j j^2 n_j\right]\left[\sum_i n_i\right] + \left[\sum_i i^2 n_i\right]\left[\sum_j n_j\right] - 2\left[\sum_i i n_i\right]\left[\sum_j j n_j\right]$$
$$= 2N\sum_j j^2 n_j - 2\left[\sum_j j n_j\right]^2 = 2N(N-1)s^2$$

where s^2 is the sample variance. The maximum value of the sample standard deviation s = ½(k-1), as is demonstrated in Appendix A, so:

[3.18] $$S^{(2)}_{max} = \tfrac{1}{2}(k-1)^2 N(N-1)$$

and this value is also realized when ½N respondents select the response "1" and all other ½N select the response "k".

This result is not surprising, but there may be good reasons to prove conjectures like this one, even if they are very plausible. Such a reason could be the different findings in the nominal and the ordinal case. In case of a nominal scale, no terminal categories can be identified. Each of the k categories has 'equal rights' to be considered as such and in section 3.2, it has been proven that the total inequality is maximal if the frequencies in all categories are equal or almost equal. This finding demonstrates that problems with variables at the ordinal level of measurement cannot always solved by treating the variable as nominal.

3.3.2 Estimated distance between response options

We will describe in chapter 6 a method proposed by Veenhoven (2009), in which the positions on a scale of the k-1 boundaries between the k categories are determined empirically, rejecting the equidistance assumption in this way. From these boundary values, the mid-interval values (MIV) $\{m_j \mid j=1(1)k\}$ of the k categories are obtained. Since the positions of k-2 intermediate intervals do not occur in the formulae [3.14] and [3.18] for S_{max}, the obvious conclusion is that in the case of the MIV-approach (i) the maximum inequality is obtained again by the equipartition of the N sample members over both terminal categories and (ii) the formulae [3.9], [3.14] and [3.15] are still applicable by simply replacing the ordinal numbers of the categories with the corresponding MIV.

3.3.3 *Ordinal level of measurement*

The former of these above two conclusions even applies to the truly ordinal situation, but the latter will not, since nothing is known about the magnitude of the distances between the positions of the categories; all we know is their algebraic signs. No suitable statistic *OII* (ordinal inequality index) has been proposed yet as an indicator for the amount of inequality at this level of measurement.

To some readers, it might be alluring to solve this problem by taking into account the number of intermediate categories for each pair and to augment this number by unity; however, the results of this approach will turn out to be identical to those in the pseudo-metric case.

3.3.4 *Relationship with mean pair distance and standard deviation*

From the equation:

[3.9] $$S := \sum_{j=1}^{k} \sum_{i=1}^{k} |j - i| n_j n_i$$

in section 3.3.1, it will be clear that in the metric discrete case, the total amount of inequality *S* is proportional to the mean pair distance, abbreviated MPD, also known as the "mean absolute distance". This statistic is obtained by dividing *S* by *N*(*N*-1), i.e. ignoring the improper pairs. We will apply this finding in chapter 4.

A similar comment follows from [3.17] with respect to the sample variance and the standard deviation.

3.4 Inequality in case of continuous distributions

Although happiness is always measured as a discrete observable variable, metric or not, there is a good reason to pay attention to the continuous case, more specifically to the beta distribution. This distribution is proposed in chapter 6 as a model for the probability distribution of an unobservable latent continuous happiness variable, which is mapped onto a discrete ordinal scale of happiness measurement. Such latent variables are assumed to be random variables with a continuous probability density function (p.d.f.), denoted *g*(*x*), on the domain {*x*} of the random variable *X*. This domain of *g*(*x*) may be either finite or infinite, but in view of the fact that in case of happiness-related variables the domain is always finite, only that class of probability distributions will be dealt with. Without loss of generality, we confine ourselves more specifically to distributions on the domain [0, 1], since by linear transformation

of a random variable in any other finite domain, a distribution on the [0, 1] domain can always be obtained.
The p.d.f. $g(x|\theta)$ has p parameters ($p \geq 0$). If $p > 1$, θ is a parameter vector with dimension p. The value of θ is estimated on the basis of the observations, given the structure of $g(x|\theta)$, which has to be chosen by the researcher.
By definition in our situation:

[3.19] $$\int_{-\infty}^{\infty} g(x|\theta)dx = 1$$

which is the continuous equivalent of [3.1] in the discrete case.
If in the case of discrete distribution $k \to \infty$ and $N \to \infty$ at the same time, the discrete distribution develops towards a continuous one, and then the amount of inequality S in the continuous situation is defined in a way that is very similar to the approach in the discrete case. The total distribution is partitioned in differentials, each of which acts as the equivalent of an individual in the discrete situation.

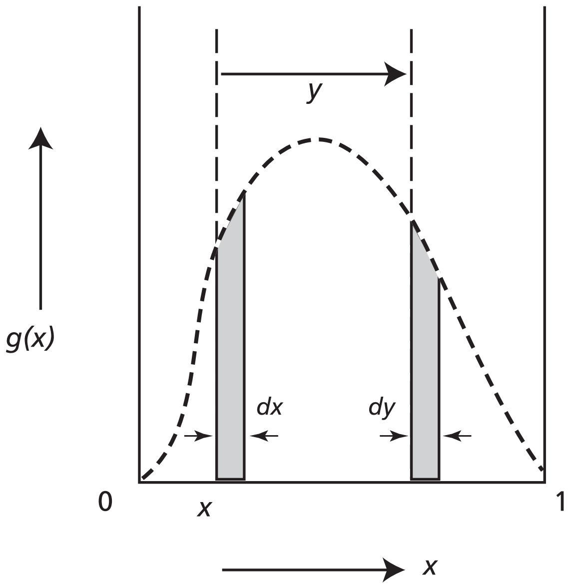


Fig. 3.1. Inequality contribution of the pair g(x)dx and g(x+y)dy in case of a continuous distribution with density g(x)

Consider the part of the distribution between the values *x* and *x*+d*x* with area *g*(*x*)d*x* and a second part of the distribution, at a distance *y* from *x*, so between *x*+*y* and *x*+*y*+d*y* with area *g*(*x*+*y*)d*y*. For a given value of *x* ($0 \leq x \leq 1$), $-x \leq y \leq 1-x$. It should be noted that *g*(*x*+*y*) is the value of the p.d.f. *g*(*x*) of the random variable *X* for *X*=*x*+*y*.

The contribution of this 'pair' to the total amount of inequality can be defined as the product of the two area's and their absolute distance, i.e.:

[3.20] $$dS = [g(x)dx] \cdot [g(x+y)dy] \cdot |y| \ .$$

Now the total amount of inequality can be written as the double integral:

[3.21]

$$S = \int_{x=0}^{1} \int_{y=-x}^{1-x} g(x)g(x+y)|y|dydx \ = \int_{0}^{1} \left[\int_{0}^{1-x} g(x+y)ydy - \int_{-x}^{0} g(x+y)ydy \right] g(x)dx$$

This total amount of inequality *S* will be referred to as the *Continuous Inequality Value (CIV)*. Contrary to the *DII* in section 3.3, this *CIV* is *not* an index number. If, however, this *CIV* is divided by its maximum attainable value and is multiplied by 100, a statistic is obtained, which is referred to as the *Continuous Inequality Index (CII)*, an index in complete analogy to *DII* for the discrete distributions. Just as *DII*, both *CIV* and *CII* have been developed for this approach only and not to extend the standard list of current dispersion descriptive statistics. For a specified type of the p.d.f. the value of *S* depends on the value of the parameter θ.

Contrary to the discrete cases in the previous section, the maximum inequality is obtained for the value of the vector θ that maximizes S(θ), but now without a term $\lambda \left[1 - \int_0^1 g(x|\theta)dx\right]$, since the value of the integral in this term is independent of θ according to [3.19]. This value of θ is obtained by putting the derivative of *S* with respect to θ equal to zero and solving that equation for θ. If $p \geq 2$, θ is a vector and *S* should be differentiated partially with respect to each element of θ separately; θ can then be solved from these *p* equations.

We will demonstrate this approach below, in practice for five cases, all on the interval [0, 1], two of which are relatively simple.

CASE I : the simplest one, the uniform distribution (*p* = 0);

CASE II : the asymmetric triangular distribution with *p* =1;

CASE III : the unimodal triangular distribution with *p* = 1;

CASE IV : the split triangular distribution with *p* = 1;

CASE V : the standard beta distribution (*p* = 2).

We will also calculate the expectation or mean value μ, the variance σ^2 the skewness γ_1 and the kurtosis γ_2 for each of the distributions, where γ_1 and γ_2 are defined here as:

[3.22] $$\gamma_1 := \frac{E(x-\mu)^3}{[\sigma^2]^{3/2}}$$

and:

[3.23] $$\gamma_2 := \frac{E(x-\mu)^4}{[\sigma^2]^2}$$

respectively, and the operator *E* denotes the expected value. This is done to examine how inequality of a continuous random variable is related to the several distribution moments, especially the even ones.

CASE I : *the uniform distribution of a random variable X on* [0, 1].

[3.24] $$g(x) = \begin{cases} 1 & x \in [0,1] \subset \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

Since $\int_{-\infty}^{\infty} g(x)dx = \int_0^1 dx = 1$, g(x) = 1 is a p.d.f.

This p.d.f. has no parameters at all, so the single value of *S* is its maximum automatically. It can be calculated from [3.21] and [3.24], resulting in:

[3.25] $$S = \int_0^1 \left[\int_0^{1-x} y dy - \int_{-x}^{0} y dy \right] dx = \frac{1}{3}$$

For this distribution $\mu = 1/2, \quad \sigma^2 = 1/12, \ \gamma_1 = 0 \ \text{ and } \ \gamma_2 = 9/5$

CASE II : *the asymmetric triangular distribution on* [0, θ]; Fig. 3.2a.

In this case, which is the simplest skew distribution, the p.d.f. of *X* is

[3.26] $$g(x) = \begin{cases} 2\theta^{-2}(\theta - x) & x \in [0,\theta] \subseteq]0,1] \subset \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

Since $\int_0^1 g(x|\theta)dx = \int_0^\theta 2\theta^{-2}(\theta - x)dx = 1$, $g(x)$ is a p.d.f.

In this case, *p* = 1 and the value of *S* depends on the value of the parameter θ:

[3.27]

$$S(\theta) = \theta^{-4} \int_0^\theta \left[\int_0^{\theta - x} 2(\theta - x - y)y\,dy - \int_{-x}^{0} 2(\theta - x - y)y\,dy \right] 2(\theta - x)dx \; = \frac{4}{15}\theta$$

Hence 0 ≤ S(θ) ≤ 4/15 and the extreme values are reached for θ = 0 and θ =1 respectively. For this distribution
$\mu = (1+\theta)/3$, $\sigma^2 = \theta^2/18$, $\gamma_1 = (2/5)\sqrt{2} \approx 0{,}57$ and $\gamma_2 = 12/5 = 2{,}4$

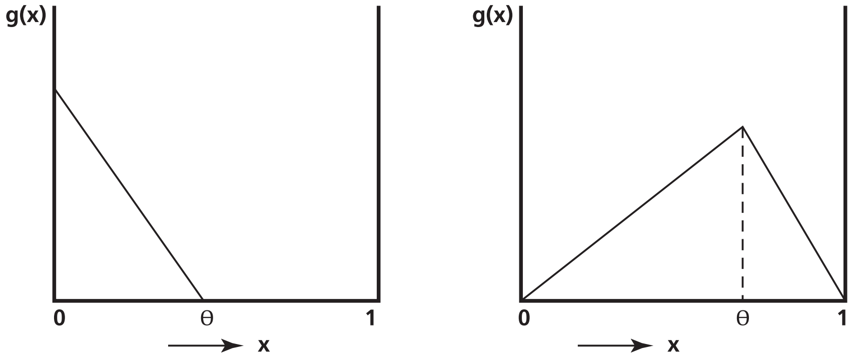


Fig. 3.2 Triangular distributions (a) CASE II (b) CASE III

CASE III : *the unimodal triangular distribution on* [0, 1]. Fig. 3.2b.

[3.28] $$g(x) = \begin{cases} 2\theta^{-1}x & x \in [0,\theta] \subseteq [0,1] \subset \mathbb{R} \\ 2(1-\theta)^{-1}(1-x) & x \in [\theta,1] \subseteq [0,1] \subset \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

Check for [3.19]:

[3.29] $$\int_0^\theta 2\theta^{-1}x\,dx + \int_\theta^1 2(1-\theta)^{-1}(1-x)\,dx = \theta + (1-\theta) = 1$$

Because in this case $g(x)$ is not defined in a unique way over the complete interval (0, 1), we have to calculate S as the sum of two components, one for each of two subdomains separately. These two components are:

[3.30] $$S_1 = \int_0^\theta \left[\int_0^{\theta-x} 2\theta^{-1}(x+y)y\,dy + \int_{\theta-x}^{1-x} 2(1-\theta)^{-1}(1-x-y)y\,dy - \int_{-x}^{0} 2\theta^{-1}(x+y)y\,dy\right] 2\theta^{-1}x dx$$

[3.31]
$$S_2 = \int_\theta^1 \left[\int_0^{\theta-x} 2(1-\theta)^{-1}(1-x-y)y\,dy + \int_{-x}^{\theta-x} 2\theta^{-1}(x+y)y\,dy - \int_{\theta-x}^{0} 2(1-\theta)^{-1}(1-x-y)y\,dy\right] 2(1-\theta)^{-1}(1-x)dx$$

[3.32]
$$S(\theta) = S_1 + S_2 = \left[\frac{1}{3}\theta(1-\theta) + \frac{4}{15}\theta^3\right] + \left[\frac{1}{3}\theta(1-\theta) + \frac{4}{15}(1-\theta)^3\right] = \frac{4}{15} - \frac{2}{15}\theta(1-\theta)$$

The maximum value of S equals 4/15 and is obtained at θ either 0 or 1, just as in the identical case II for $\theta = 1$; the minimum value (7/30) corresponds to $\theta = 1/2$. The difference between the maximum and the minimum value is rather modest: 0,267 – 0,233 = 0, 033.
The moments of this distribution are:

[3.33] $$\mu = (1+\theta)/3$$

[3.34] $$\sigma^2 = [1-\theta(1-\theta)]/18$$

[3.35] $$\gamma_1 = \frac{(1-2\theta)(2-\theta)(1+\theta)}{[1-\theta(1-\theta)]^{3/2}} \cdot \frac{\sqrt{2}}{5}$$

[3.36] $$\gamma_2 = 12/5 = 2{,}4$$

For θ = 0(0,5)1, γ_1 adopts the values +0,57, 0 and −0,57 respectively.
These results are in complete agreement with Kotz & Van Dorp (2004).

CASE IV (Fig. 3.3): *the symmetric split triangular distribution on* [0, 1]:

[3.37] $$g(x) := \begin{cases} \theta^{-2}(\theta - x) & x \in [0, \theta] \subseteq \left[0, \frac{1}{2}\right] \subset \mathbb{R} \\ \theta^{-2}(x - 1 + \theta) & x \in [1 - \theta, 1] \subseteq \left[\frac{1}{2}, 1\right] \subset \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

Check for [3.19]:

Since $\int_0^1 g(x|\theta)dx = \int_0^\theta \theta^{-2}(\theta - x)dx + \int_{1-\theta}^1 \theta^{-2}(x - 1 + \theta)dx = \frac{1}{2} + \frac{1}{2} = 1$,

$g(x)$ is a p.d.f.

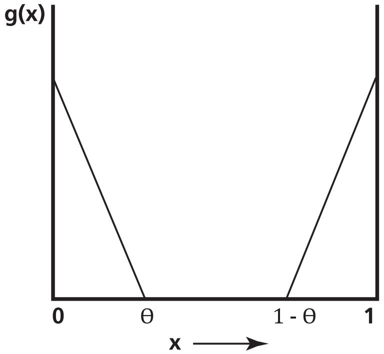


Fig. 3.3 Split distibution

Just as in case III, we have to calculate *S* as the sum of two nontrivial components, one for each of two subdomains separately. These two components are:

[3.38] $$S_1 = \theta^{-2} \int_0^\theta \left[\int_0^{\theta - x} (\theta - x - y)ydy + \int_{1-\theta-x}^{\theta - x} (x + y - 1 + \theta)ydy - \int_{-x}^{0} (\theta - x - y)ydy \right] (\theta - x)dx$$

[3.39]
$$S_2 = \theta^{-2} \int_{1-\theta}^1 \left[\int_0^{1-x}(x + y - 1 + \theta)ydy - \int_{-x}^{\theta - x}(\theta - x - y)ydy - \int_{1-\theta-x}^{0}(x + y - 1 + \theta)ydy \right] (x - 1 + \theta)dx$$

Because p=1, the maximum inequality is obtained by finding the maximum value of:

[3.40] $$S(\theta) := S_1 + S_2 = \left(\frac{1}{4} - \frac{1}{10}\theta\right) + \left(\frac{1}{4} - \frac{1}{10}\theta\right) = \frac{1}{2} - \frac{1}{5}\theta$$

This maximum value equals 1⁄2 and is obtained at $\theta \downarrow 0$. A larger value of θ reduces the inequality of the distribution, with 2/5 as its minimum value at $\theta = ½$.

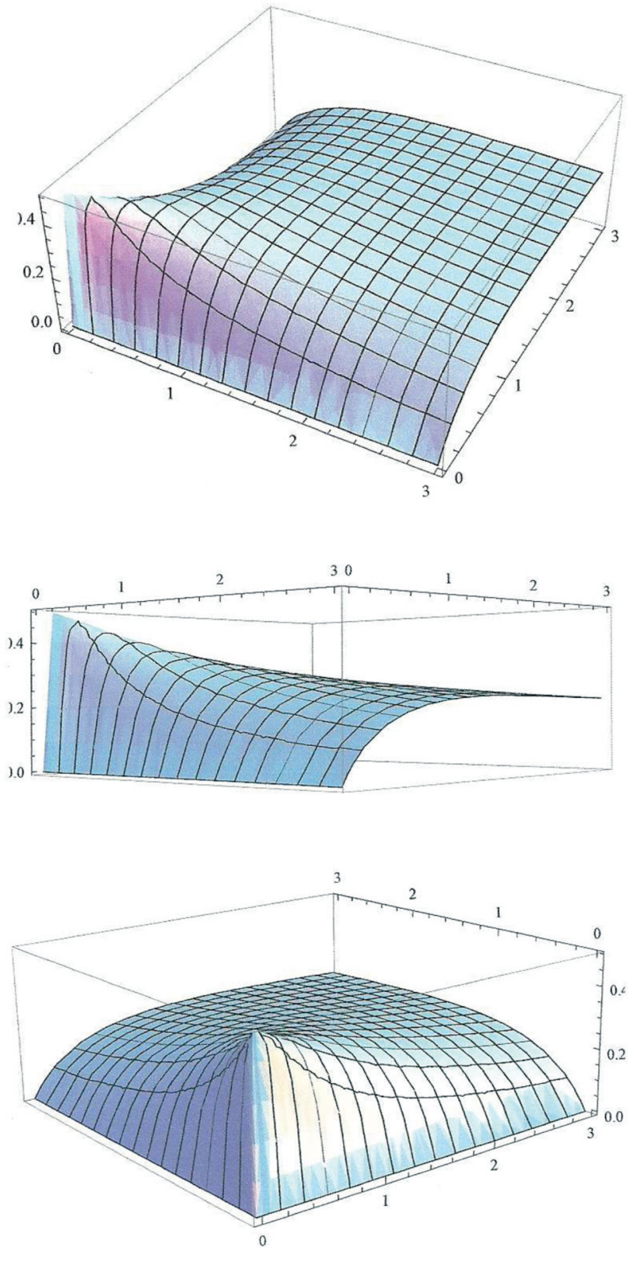


Fig. 3.4 a,b and c. Continuous Inequality Value (CIV) S (α, β) of a standard beta distribution. Plot for 0 < α , $\beta \leq$ 3 (Prepared by Dr R.J. Stroeker)

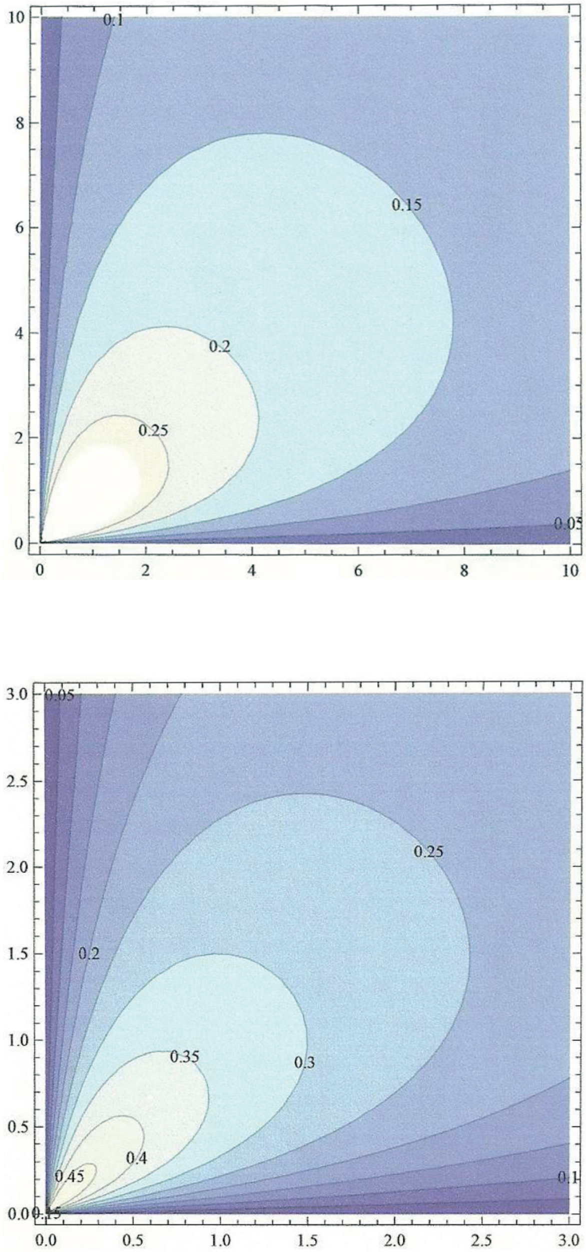


Fig. 3.5 a and b. Contour plots of the Continuous Inequality Value
S (α, β) *= 0,10(0,05)0,45 for a standard beta distribution with parameters*
0 $\leq \alpha, \beta$ *= (10 ; 3) (Prepared by Dr. R.J. Stroeker)*

The moments of this distribution are:

[3.41] $\mu = ½$

[3.42] $\sigma^2 = (3 - 4\theta + 2\theta^2)/12$

[3.43] $\gamma_1 = 0$

[3.44] $\gamma_2 = \left(\frac{3}{5}\right)\frac{15-40\theta+60\theta^2-48\theta^3+16\theta^4}{[(3-4\theta+2\theta^2)/12]^2}$

For $\theta = 0$, $\sigma^2 = ¼$ and $\gamma_2 = 1$. If θ increases to ½, the value of σ^2 decreases to 1/8 and that of γ_2 increases to 4/3.

CASE V : *The standard beta distribution.*

In this case, the random variable *X* has a p.d.f. with two shape parameters α and β:

[3.45] $g(x) := \begin{cases} [B(\propto,\beta)]^{-1}x^{\alpha-1}(1-x)^{\beta-1} & x \in [0,1] \subset \mathbb{R} \quad \alpha,\beta \in \mathbb{R}^+ \\ 0 & \text{otherwise} \end{cases}$

where

[3.46] $B(\alpha,\beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \qquad t \in [0,1] \subset \mathbb{R} \quad \alpha,\beta \in \mathbb{R}^+$

is the complete beta function with parameters α and β.
The total amount of inequality in this case can be written as:

[3.47] $S(\alpha,\beta) := [B(\alpha,\beta]^{-1}(I_1 - I_2)$,

where

[3.48] $I_1 := \int_{x=0}^{1} x^{\alpha-1}(1-x)^{\beta-1}\left[\int_{y=0}^{1-x} y(x+y)^{\alpha-1}(1-x-y)^{\beta-1}dy\right]dx$

and

[3.49] $I_2 := \int_{x=0}^{1} x^{\alpha-1}(1-x)^{\beta-1}\left[\int_{y=-x}^{0} y(x+y)^{\alpha-1}(1-x-y)^{\beta-1}dy\right]dx$

As long as no simple analytical expression for *S* (α, β) is available, its value has to be obtained by numerical integration. A very informative result is given in Fig. 3.4 in which three three-dimensional plots of *S* (α, β) are depicted from different perspectives. The same function has also been represented by two contour plots, one for $0 < \alpha$, $\beta \leq 10$ and a second one as an enlargement of the former one , but this time for $0 < \alpha, \beta \leq 3$. In the figures 3.4 and 3.5, the parameters α and β have been replaced with p and q respectively for typographic reasons.

Note that:

(a) The beta distribution with $\alpha = \beta = 1$ is identical to case I

(b) The beta distribution with $\alpha = 1, \beta = 2$ is identical to case II with $\theta = 1$

(c) S is symmetrical in its parameters, i.e. $S(p,q) = S(q,p)$.

This is completely confirmed by the computations used to devise the contour plots. (Fig. 3.5).

The above case IV was selected as a distribution with a linear p.d.f., but with a shape that showed some similarity to that of a beta distribution with small values of both α and β; for reasons of convenience both shape parameters were set equal here. Considering both p.d.f. gave rise to the conjecture, that the inequality *CIV* of a beta distribution (i) has a maximum value 1⁄2 and (ii) increases as the value of the sum of α and β decreases. The above contour plot confirms these expectations reasonably well and makes visual the extent to which the inequality depends on the values of the distribution shape parameters. It appears however, that for very small values of only one of the parameters, the inequality *CIV* shows an unexpected and rather steep descent. The most obvious explanation of this phenomenon is found by considering the variance of the beta distribution:

[3.50] $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ for e.g. $\alpha \ll \beta$ and $\alpha \ll 1$.

In that case:

$$\sigma^2 \approx \frac{\alpha}{\beta(\beta+1)} \Longrightarrow \lim_{\alpha\to 0} \sigma^2 = 0 \Longrightarrow \lim_{\alpha\to 0} \sigma = 0\ .$$

Since both *CIV* and the standard deviation are known to be a good measure of the inequality of a distribution, this explanation seems to be well acceptable. Moreover, it also makes clear why, in the example, the descent is steeper as the value of β is closer to zero.

As long as in happiness distributions no values of the shape parameters are to be expected close to zero, there seems no reason to worry about this finding in practice. This expectation will be verified in chapter 7.

In table 3.3. the findings of the five cases are summarized.

Table 3.3 Values of the CIV in 5 cases together with the corresponding values of the variance σ^2, the ratio CIV/σ and the coefficient of kurtosis γ_2.
NOTE: "maximum" ("minimum") refers to the situation in which the CIV adopts its maximum (minimum) value.

Case		CIV	σ^2	CIV / σ	γ_2
I		1/3	1/12	1,15	9/5
II	maximum	4/15	1/18	1,13	12/5
	minimum	0	0		12/5
III	maximum	4/15	1/18	1,13	12/5
	minimum	7/30	1/24	1,14	12/5
IV	maximum	1/2	1/4	1,00	1
	minimum	2/5	1/8	1,13	4/3
V	maximum	≈1/2	1/4	≈1	1
	minimum	0	0		3

From this table it follows:

(i) whenever possible, we calculated the ratio of *CIV* and the standard deviation, both considered as valid measures of the inequality of the distribution. In five rows we found almost the same value (about 1,14) and the other two the value 1. The fact that *CIV* is based on *MPD*, the mean pair distance (*MPD*), and that for a normal distribution the ratio of the values of the expected *MPD* and of the standard deviation equals 1,13 (Owen, 1962), suggest the hypothesis that this value applies to a wider family of distributions, in particular to unimodal ones, at least approximately. The maximum situations of IV and V are so far remote from unimodal distributions, that it is not surprising that this results in different kurtosis values

(ii) in cases II and III, the value of the parameter (θ) does not influence the kurtosis, but in IV and V a decreasing *CIV* gives rise to a smaller variance, as is to be expected, but also to an increasing coefficient of kurtosis. Apparently the variance is not always the only determinant of the inequality as defined in [3.21], at least not in cases IV and V

(iii) in the situations of cases IV and V in which *CIV* attains its maximum value, not only both *CIV* values are (almost) equal, but the same holds for the variance, the coefficient of kurtosis and for the ratio *CIV* / σ. This supports the idea of a certain similarity of both distributions if the parameter values approach zero.

The above result of the continuous uniform distribution can be compared to the discrete continuous distribution of a random variable *X* with

[3.51] $\pi_j := \text{Prob}\{X = j\} = k^{-1} \quad j = 1(1)k$

with a variance $\sigma^2 = (k^2 - 1)/12$.

For a correct comparison, the standard deviation has to be adjusted for the fact that the domain range of *X* equals *k*-1, giving:

[3.52] $$\sigma_{adj} = \frac{1}{k-1}\sqrt{\frac{k^2-1}{12}} = \sqrt{\frac{1}{12} \cdot \frac{k+1}{k-1}}$$

For $k \to \infty$, its value approaches $1/\sqrt{12} = 0{,}29$ which is in perfect agreement with the above variance of the continuous case (I).

3.5 Application of the findings in section 3.4 to the measurement of happiness

In the previous section, we assumed the domain of the random variable *X* to be $[0,1] \subseteq \mathbb{R}$,

For the distribution of happiness in e.g. nations, it is usual to adopt $[0,10] \subseteq \mathbb{R}$ as the domain. In that case, formula [3.21] is to be replaced with:

[3.53] $$S(\theta) = \int\limits_{x=0}^{10} \left[\int\limits_{0}^{10-x} g^*(x+y)y dy - \int\limits_{-x}^{0} g^*(x+y)y dy \right] g^*(x) dx$$

where *g*(*x*) has been replaced with $g^*(x) := (1/10) \cdot g(x)$ the Jacobian determinant (1/10) being required in order to replace [3.19] with:

[3.54] $$\int\limits_{0}^{10} g^*(x|\theta) dx = 1$$

and *x* and *y* have been adjusted accordingly.

As a consequence, this linear transformation of the random variable will not affect the numerical value of the inequality measure $S(\theta)$ of its probability distribution, in other words: $S(\theta)$ is invariant under linear transformation of the random variable.

3.6 Conclusions

A set-theoretic approach to inequality as a relation on the set of the responses of all members of a sample from a population produces a number of additional inequality statistics. These statistics can be used for computing the maximum possible degrees of inequality and for ranking different happiness distributions according to increasing inequality. This applies to both discrete and continuous happiness variables, albeit in slightly different ways.

In the discrete situation, happiness is measured by applying a measurement scale with *k* ordered categories. In this situation, the inequality of the distribution can adopt a minimum (zero) value, but also a maximum. The latter situation occurs if all ½*N* sample members select the lowest possible rating and the other ½*N* the highest possible one. This finding even applies to the truly ordinal case, i.e. if the distances between the ratings are unknown.

For the distribution of a variable at the nominal level of measurement, we defined a measure for its happiness inequality, which is referred to as the *Nominal Inequality Index* (*NII*) and for the discrete metric case the *Discrete Inequality Index* (*DII*) of that distribution. The latter statistic is equal to the mean pair distance, but for a factor which contains the sample size only. Our intention is definitely not to add this indices to the list of current dispersion measures for daily use, but just to use this measure for selecting the most appropriate measures from that list.

In case of a continuous distribution, it is possible to define a statistic, called *Continuous Inequality Value* (*CIV*), given the p.d.f. of the probability distribution. For the standard beta distribution with parameters α and β it appears that $CIV < 0{,}5$ and that *CIV* decreases as α and/or β increase. In a contour plot, a more quantitative picture is given for $CIV(\alpha, \beta)$.

The results of section 3.3 will be applied in chapter 4 to the judgment of various descriptive statistics for the quantification of happiness inequality within nations.

✧ ✧ ✧

Chapter 4

Descriptive Statistics for Happiness Inequality in Samples

Parts of this chapter have been published in:
Kalmijn, W.M. & Veenhoven, R. (2005)
Measuring Inequality of Happiness in Nations. In Search for Proper Statistics.
Journal of Happiness Studies, **6**, (4), 357 – 396
ISSN 1389-4978 (Print) 1573-7780 (Online)
DOI 10.1007/s10902-005-8855-7

ABBREVIATIONS USED IN THIS CHAPTER:

DII	Inequality Index for a metric discrete variable (section 4.2)
GC	Gini coefficient (section 4.2.7)
MPD	mean pair distance (section 4.2.6)
Prob	Probability (section 4.7).

4.1 Introduction

We have described the measurement of happiness in samples, resulting in some happiness distribution over various response categories in section 2.2. Such distributions have a number of characteristics, the two most important of which are one a statistic for the location and two, a statistic for the dispersion. In section 1.5, we discussed the antagonism between utilitarians and egalitarians. The interest of utilitarians is focused completely on the location statistic, since it characterizes the central value of a distribution. The dispersion statistic reflects the inequality within a distribution and hence this is also considered to be important from the egalitarian point of view.

Inequality research has been extended beyond income differences, because income disparities have become less relevant in affluent societies and because other inequality issues have begun to appear on the political agenda, in particular inequalities in health and in social contacts. The happiness research group at Erasmus University Rotterdam has explored the issue of inequality of happiness in nations. Results have been published by Chin Hon Foei (1980), Veenhoven (1990), Veenhoven & Ehrhardt (1995). and Veenhoven (2000).

In these studies inequality of happiness in nations was measured using the standard deviation as the descriptive statistic. This statistic is also used in the World Database of Happiness, where, in the section 'Distributional findings in nations', standard deviations for all 4336 surveys that involved questions on happiness can be found, and where the standard deviation is also used for nation rankings and time trends.

The motive to reconsider this common practice came forward with the involvement of economists in happiness research, as has been described in section 2.6, with their own favourite methods and statistics. One of such statistics is the "Gini coefficient", which is commonly used as a measure of income inequality. In an examination of the (cor)relation between income inequality and happiness inequality, some feel it obvious, if not inevitable to apply the same statistic for both inequalities, and the Gini coefficient was considered to be the most serious candidate for this role. A submitted draft publication from the above mentioned Happiness research group, based on a different approach, was refused for this reason. Our methodological doubts about the validity of this opinion were the direct motivation to undertake the investigation described in this chapter and partly also already in chapter 2.

Our aim in this chapter is to select the most appropriate statistic to quantify the inequality of happiness in a sample. We begin this chapter with an inventory of current descriptive dispersion statistics as candidates (section 4.2). In section 4.3, we will select the criteria for the assessment of their aptness for this task. Section 4.4 covers some methodological remarks. The candidate descriptive dispersion statistics will be tested by applying these to some series of hypothetical distributions in section 4.5. In this context, we prefer application to hypothetical distributions over to 'natural' distributions, because (i) they meet all our assumptions by definition, (ii) we have control over their inequality and (iii) causes of possible problems are generally clear at once.

How the various candidate perform with respect to the hypothetical distributions, and additionally to a small set of related natural distributions as a check, is described in section 4.6. Since one of the findings was that the Gini coefficient fails as a measure of happiness inequality, we discuss possible explanations in section 4.7. The chapter is completed with conclusions and recommendations in section 4.8.

4.2 Statistics of inequality

There are many statistics in use for quantifying the dispersion in distributions, e.g. Kendall and Stuart, (1977: 42 - 52). The following candidates may apply to the measurement of happiness and will be considered in turn:

- the range (4.2.1.)
- the average deviation from the mean (4.2.2)
- the variance and its square root, the standard deviation (4.2.3)
- the relative standard deviation (4.2.4.)
- the interquartile range (4.2.5)
- the mean pair distance (4.2.6)
- the Gini coefficient (4.2.7)
- Theil's measure of 'entropy' (4.2.8)
- the percentage outside modus (4.2.9)

Besides these inequality statistics for practical use, we have defined four indices of inequality in chapter 3 :

- *CII* = Inequality Index for a continuous variable (section 3.4)
- *CIV* = Inequality Value for a continuous variable (section 3.4)
- *DII* = Inequality Index for a metric discrete variable (section 3.3)
- *NII* =Inequality Index for a nominal variable (section 3.2).

These indices have not been introduced as an extension of the above list of current inequality statistics, but their function is to be used to gauge the latter for their validity. In this context, we shall apply in particular the index DII for this purpose, since happiness is always measured in samples as a discrete variable.

4.2.1 Range

The theoretical range of a distribution is the difference between the highest and the lowest possible rating, in the case of the application of a [1, *k*] scale the difference $k-1$. The *actual* range is the difference between the highest and the lowest *selected* rating on the scale. This latter difference is useful as a dispersion measure for small samples, say for $N \leq 10$. In large-scale happiness surveys, the actual range will mostly concur with the theoretical range and for this reason we will not consider this statistic in any more detail.

4.2.2 Average deviation from the mean

The deviation from the mean may be positive, negative or zero for any observation, but the average deviation of all observations will be zero, as follows from the definition of the mean. If we ignore the algebraic sign of the deviations, we can compute an average that is not zero and does indicate the degree of dispersion in the distribution. This mean absolute deviation, however, is a rather obsolete measure, since its mathematical tractability and its relationship to distribution model parameters is quite complicated (Kendall and Stuart, 1977: 44).

4.2.3 Variance and standard deviation

The standard solution for the problem of getting rid of the algebraic sign is to square the individual differences from the mean. The average squared difference is the *variance* and its square root is the *standard deviation* of a distribution. The advantage of the latter over the variance is that it is expressed in the same unit and on the same scale as the basic observations.

4.2.4 Relative standard deviation

The relative standard deviation, also called '*coefficient of variation*', is the ratio of the standard deviation (s) and the mean value (m), usually expressed as a percentage.

4.2.5 Interquartile range

The interquartile range, or interquartile distance, is the difference between the third and the first quartile. The calculation of these quartiles is not always a simple nor unique process as it also depends on the definition of the quartiles. These problems have been reviewed by Langford (2006).

Assuming a discrete distribution, the observations are listed in ascending order and then the set of observations is partitioned into four almost equal parts. In the example given in table 2.1 the result is a partition into four quarters of 50 observations each, as is specified in table 4.1.

If k is not a divisor of N, the sizes of the four quarters are not exactly equal, but only approximately. For such cases Langford (o.c.) proposes a solution in his paper.

Table 4.1. Frequencies of ratings for calculation of the quartiles of the distribution of table 2.1

Rating j =	1	2	3	k=4	sum	j_{max}
First quarter	40	10	--	--	50	2
Second quarter	--	50	--	--	50	2
Third quarter	--	--	50	--	50	3
Fourth quarter	--	--	30	20	50	
Sum n_j	40	60	80	20	N=200	

The ratings in the right-hand column are the first quartile (2), the second quarter or median (2) and the third quartile (3) respectively. Although in this situation the quartiles are defined, their difference is not, since subtraction of quartiles as ordinal numbers is not an admissible arithmetic operation.

In this approach, the interquartile range is undefined and hence nonexistent, so unusable as a measure for the inequality of a sample. As integers, quartiles can adopt only a very limited number of values, so they can discriminate between nations for their happiness only very rarely. Moreover, their value is sometimes very sensitive to one or two observations and in other cases hardly sensitive to substantial differences. If in the above example one of the respondents changes his choice from "not too happy" to "pretty happy", the value of the median suddenly changes from 2 into 3 , whereas the change of 30 respondents from "unhappy" into "not too happy " has no consequences at all. Such findings are highly uncomfortable.

The solution proposed by Langford is to abandon the strictly discrete approach and to replace it with a continuous distribution according to the method described in section 2.7. The quartiles are found by partitioning the total area of the histogram (Fig, 2.1) into four parts with equal area's by three vertical lines. The intersections of these lines with the happiness axis are the values of the quartiles. As the reader may verify, their values in our example are 1,67, 2,50 and 3,13 respectively, resulting in an interquartile range value of 1,46. In this approach the effects of the above 1 vs. 30 "changing respondents" are that in the first case, the median does not change from 2 into 3, but from 2,50 into 2,51, whereas in the second case the interquartile range is reduced substantially from 1,46 to 1,18. Within the context of this chapter we shall follow Langford's approach, with the extension that, if in extreme situations quartiles are obtained outside the interval [1, k], their values are replaced with the nearest value within that interval. In real situations, this is not expected to happen, but this situation cannot always be excluded in hypothetical distributions.

In their paper on this subject, Kalmijn & Veenhoven (2005) had already applied the above method, but without the extension of the continuous distribution outside the [1, k] interval. Therefore, in extreme distributions, values of the inter-quartile range values were obtained that are different from the ones given in this chapter. However, these differences do not impact the conclusions.

4.2.6 Mean pair distance

The mean pair distance, abbreviated MPD, in a sample is obtained as follows. For any possible combination of two out of *N* subjects in the sample, the (absolute) difference of their ratings is determined. For example, if the ratings of A and B are 4 and 7 respectively, the difference for this pair equals $|4 - 7| = 3$.

The average value of the absolute differences of all possible ½*N*(*N*–1) pairs is reported as an indicator for inequality in the distribution. The fact that this measure takes into account all possible observable differences means that it intuitively fits very well with the inequality concept as it is described in chapter 3. One might raise the rating difference to some power, larger (smaller) than unity, if one assumes that the subjectively *perceived* difference by A and B in the above case of the three rating units is more (less) than three times that of one single rating difference. Even a zero power might be an option, if one assumes that subjects are only capable of observing the existence and the algebraic sign of a happiness difference, but are unable to estimate its magnitude.

It will require, however, much more research to produce arguments for the appropriate value of the power in all these cases. Moreover, in the case of subjectively perceived difference, one has to demonstrate that in the above case both A and B perceive the difference of three units as (almost) equal. As long as no further information on this subject is available, we shall use the method only with ‘objective’ differences and unity power, although one must remain aware of the fact that in this situation the equidistance of the ratings is yet an underlying assumption. The computation required for this statistic is relatively tedious, but with today’s computers, this is not a serious drawback.

4.2.7 Gini coefficient

The Gini coefficient, also referred to as *Gini's concentration ratio*, (Gini 1912) is frequently used to characterize inequality in income distributions, see also Theil (1967: 121 –128). Any income distribution can be represented by a *Lorenz curve*, as has been described in section 2.6. In principle, each point of this curve corresponds to some value *I* of the income distribution. The abscissa of this point is the relative number of people in a population that have an income up to and including the value *I*; the corresponding ordinate is the sum of their incomes divided by the sum of all incomes over the population. The Lorenz curve will be a curve through the points (0,0) and (1,1), but generally all other points of the curve will be expected to be found below the diagonal through (0,0) and (1,1); see Fig. 4.1. The closer the Lorenz curve is to that diagonal, the smaller the income inequality of the sample concerned.

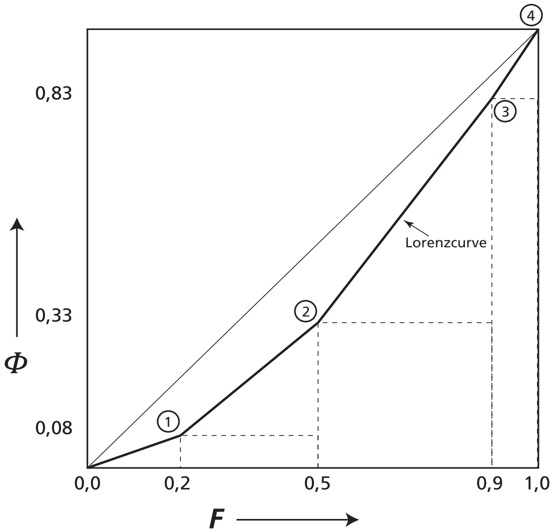


Fig. 4.1 Lorenz curve for the happiness distribution of table 2.1

The *Gini coefficient* (*GC*) is defined as the ratio between the area between the Lorenz curve and the above diagonal and the area of the complete triangle below that diagonal, the latter being equal to ½. Clearly *GC* = 0 in the case of complete income equality, whereas *G* = 1 refers to the situation where one single individual earns all the money, leaving nothing for all other *N*–1 ones. In practice 0 < *GC* < 1 and *GC* is one of the many possible income inequality indicators, which at first sight makes the Gini coefficient at least a candidate for indicating happiness inequality. The fact that *GC* is bounded to the interval

[0, 1) is certainly an advantage when comparing different values of inequality and is presumably responsible for at least a part of its popularity. This also implies that *GC* is dimensionless, i.e. it does not matter in what currency the incomes are expressed.

Various distribution models have been proposed for income distribution. The most widely applied are the 'Pareto distribution' and the 'lognormal income distribution'. Both distributions are defined on the interval [0, ∞) and are positively skewed, due to the existence of a theoretical (zero) minimum income and the non-existence of a theoretical maximum value. A theoretical relationship can be derived between the distribution model parameters and the Gini coefficient for any specified mathematical distribution model. It appears that in both cases mentioned above this relationship is a relatively simple one, and it is this finding that also makes the Gini coefficient attractive when determining measures of income distribution.

However, as long as there is no simple mathematical model available to describe the happiness distribution, Gini coefficients cannot be related to its model parameters, and this may make their applicability to such situations less attractive. As has been pointed out in section 2.6, happiness distributions are essentially different from income distributions, including the ones mentioned above.

The Gini coefficient is related to the mean pair distance (Sen, 1997: 29-34). It can be proven from the latter that the Gini coefficient, the sample mean happiness m, the sample size N and the mean pair distance, abbreviated as *MPD* (with unity power) are connected by the relationship

[4.1] $$GC = \frac{N-1}{2mN} MPD \approx \frac{MPD}{2m} \text{ for large values of } N$$

For the theoretical case $m = 0$, $GC := 0$.

The above relationship is proven In Appendix C for a happiness distribution on the [1, k] interval under some assumptions. These are that the scale points are equidistant and that it is admissible to obtain a Lorenz curve by connecting the $k + 1$ points as has been done in Fig. 2.4. Within the context of our comparative study in this chapter, we shall adopt the Gini values as calculated from [4.1]. In section 4.7 we will reconsider the Gini coefficient, including the above assumptions.

4.2.8 Theil's measure of 'entropy'

The entropy measure proposed by Theil (1967: 91 - 96) arises from the application of a thermodynamic concept, 'entropy' to information theory. It is a non-negative inequality measure, which for this case can be written as:

[4.2] $$T := \ln(N) + \sum_{j=1}^{k} \left(\frac{jn_j}{Nm}\right) \ln\left(\frac{j}{Nm}\right)$$

where ln(*N*) indicates the natural logarithm of the sample size *N*, n_j is the absolute frequency of the ratings with value *j* (*j* =1(1)*k*; $\Sigma n_j = N$), *m* is the mean happiness value and Σ denotes a summation over all *k* possible ratings "*j*" (Sen, 1997: 34 - 36). If all respondents report the same rating (complete equality), $T := 0$.

The application of Theil's inequality to happiness assumes the existence of some 'total amount of happiness', which has been distributed over all members of the society in question. In Eq. [4.2], the ratio j/Nm is considered to be the relative 'share of the total amount of happiness' of an individual that selects a rating "*j* " while jn_j/Nm is the share of all individuals together who do so.

4.2.9 Percentage outside modus

The percentage outside the modal rating is simply defined as the difference between 100 % and the percentage of the ratings in the modal one; a mode of a distribution is defined as a value of the variable for which the relative frequency has a local maximum value (Kendall & Stuart, 1977: 40). If the distribution is not unimodal, the mode with the highest percentage has to be selected. If this choice turns out to be ambiguous, it does not matter which 'highest' mode is adopted. Clearly this statistic has a zero value in the case of complete equality. Its theoretical maximum value is (almost) 100·(k–1)/k in the case of a uniform distribution (Series 5, table 4.6 below).

4.3 Criteria of judgement

When assessing the various candidate statistics for their aptness as measures of the happiness inequality in samples, we have to specify the criteria to be applied. The following eight, most of which may be felt obvious, were selected:

1: Single finite number as result.

A usable statistic should express the degree of inequality value in a single finite number, either in combination with a unit or not, and should do so for any conceivable distribution of happiness.

2: Interval level of measurement

In section 2.2 we have described how happiness is measured at the ordinal level of measurement according to Stevens (1956) , but how the observed happiness is treated as if it were metric. Hence, a statistic should be admissible for application to the distribution of variables measured at the interval level of measurement. In our view, happiness is not measured at the ratio level, since there is no natural 'zero level' of happiness.

3: Independence of scale range.

In view of the exuberant variation in scales used in happiness research, the requirement of comparability implies that the candidate operates in a way independent of the number of possible ratings on the scale of measurement or at worst, weakly dependent.

4: Independence of sample size.

Sample sizes tend to differ across nations and samples are typically larger in larger nations, albeit not for statistical reasons. Hence the values of the statistics must be independent of sample size, at least where large samples are concerned.

5: Independence of the mean.

The ongoing discussion between egalitarians and utilitarians regarding happiness calls for a measure of inequality that is independent of the mean. Hence a useful statistic of dispersion should be fully independent of the average value or at least only weakly dependent on it.

6: Equal values for equally unequal distributions.

A basic requirement of any statistic is that it yields equal values for distributions considered 'equally unequal'. Inequality statistics should be invariant under operations such as 'translation' and 'reflection' of the happiness distribution along the happiness scale.

> *If in a distribution of some variable, e.g. happiness, all individual ratings are augmented by the same amount (d), the complete distribution will be 'translated' along a horizontal axis over a distance d. This operation will change the value of some statistics, e.g. the average value, whereas others like the standard deviation and the skewness are unaffected. The latter are said to be 'invariant under translation'. 'Invariant under reflection' means that in the case of reflection to a vertical mirror line, the statistic of the image has the same value as that of the original. Reflection of the distribution will change the algebraic sign of the skewness, but will not affect the standard deviation.*

7: Differentiation between more and less unequal distributions.
To be a usable statistic, a candidate must enable the user to distinguish between distributions we consider to be 'unequally unequal'.

8: Sensitive to degree of inequality.
Finally, a usable statistic reflects the degrees of inequality in a distribution, and the values it generates must fit our notion of what is more, or less, equal.

This list of criteria is not exhaustive. It can be extended with others. So it may be attractive if an inequality statistic has a simple relationship with the distribution model parameters, but this requires that such a mathematical model is available.

4.4 Methodological remarks

In chapter 2, we have presented a number of general methodological views on happiness and its measurement. In this present section we shall pay specifically methodological attention to the statistics that have been nominated for the quantification of happiness inequality in samples.

Some of the inequality statistics enumerated in section 4.3 have a unit as described in section 2.6, whereas others are essentially dimensionless. The latter class includes the coefficient of variation, the Gini coefficient and Theil's inequality measure. Conversion of the dimension of the variable, e.g. of the monthly income from USD into EUR has no influence on these statistics. The introduction of the Euro in 2002 as a new currency reduced the *numerical* value of all Dutch incomes by a factor 2,2. This affected income distribution in the Netherlands, but did not affect income inequality, nor was the numerical value of the Gini coefficient influenced by this event.

It is clear that the coefficient of variation (the ratio s/m) is dimensionless, since both m and s are always expressed in the dimension of the variable of the distribution. When calculating the Gini coefficient, this statistic is made dimensionless in a related way, if the main pair distance is divided by approximately twice the mean value (section 3.7). The result gives a form of 'standardization', in this case to a bounded interval [0; 1].

Happiness ratings, however, are of an essentially different nature. They are already dimensionless, due to their origin as ordinal numbers. The distinction between statistics that have a dimension and those that are dimensionless is clearly visible in cases where the variable has some dimension. The same distinction exists for happiness inequality statistics, but here it is invisible in the values. In the latter case, the 'standardization' is obtained in a completely different way: by using a fixed left and right-hand boundary for the happiness rating. Division by the mean value is not required to achieve this. Instead, the possibility to compare values for the inequality of different societies requires a linear transformation of the ratings to a common scale with a standardized length.

With respect to dependency, cf. section 2.8, our conclusion is that, theoretically, the standard deviation is dependent on the value of the mean happiness rating, but that in most practical situations this type of dependency is not very strong, at least not at higher values of the standard deviation. Moreover, this dependency is caused by the way happiness is measured in practice rather than by the choice of the inequality statistic, so this type of structural dependency is expected to occur at all candidates and should be ignored in the selection procedure. The presence of the mean value in the denominator of a happiness inequality statistic, however, introduces a dependency on the mean, which in this context is most undesirable.

Since we see happiness as an intensity variable, we may foresee problems with statistics that assume extensity variables, and in particular with the Gini coefficient, Theil's inequality measure and the coefficient of variation. These problems will emerge as a structural dependency of the value of inequality statistic on the value of the mean. Moreover, these problem statistics require that the happiness measurements be obtained at the ratio level of measurement, which in our approach they are not.

4.5 Hypothetical distributions

We started with an attempt to devise one single series of hypothetical distributions that vary from most to least possible inequality. This appeared to be difficult, since inequality may be caused in different ways and one cannot combine all these notions in one single series of distributions. We ended up with six series of increasing inequality, each of which highlights a different way of inequality. These distributions are presented in the upper parts of the tables 4.2, 4.3, 4.4 and 4.5, in which each column corresponds to the rating in the upper row.

The ratings of all the happiness distributions are postulated to be equidistant and all scales to be metric. Most scales are chosen as [1, 10] ; a few were [1, 5] scales. Each distribution is denoted with a letter and sometimes a digit additionally. Distributions we consider as 'equally unequal' a priori, have been denoted using the same letter and different digits. Although the aptness of some statistics is, even a priori, debatable on the basis of the discussions in sections 4.3 and 4.4, we have applied all of them, except the range, to the hypothetical distributions to demonstrate their behaviour in this context.

Each row in *the upper part* of each of the tables 4.2 – 4.6 correponds to one of the distributions and each column to one of the *k* ratings. The number in a "cell" is the absolute frequency of the rating. The ratings are given in the upper row and the codes of the distributions in the left hand column, whereas the DII value of each distribution can be found in the right hand column. The numerical values of the various statistics are given in the *lower part* of the table. In the bottom row, the rank correlation coeficient tau-B between the values of the statistics and the DII values are listed.

Series 1

The first series of 12 hypothetical distributions (table 4.2) depicts increasing 'segregation' in distributions. At the start in distribution A1 and A2 there is no inequality, since everybody is equally happy. Moving from A to K, inequality increases gradually and the shape of the distribution changes from unimodal (A and B) through normal (D) and then turns bimodal (E, F and H). The series ends with distribution K, where half the sample has the lowest and the other half has the highest possible happiness rating (complete 50/50 split with maximum difference), a distribution which is maximally unequal according to section 3.3.

The distributions D1 and D2 are included because they are (almost) normal, but with different mean values. Just as the uniform distribution G, their exact position within Series 1 is determined by their *DII*-value in the right-hand column.

Their approximate normality follows from the consideration of their skewness and kurtosis. Just like normal distributions, the distributions C1 and C2 are symmetric and have zero skewness. In this case, the coefficient of kurtosis equals 2,99 for both distributions. For normal distributions, this coefficient of kurtosis has the value 3 (Kendall and Stuart, 1977: 88), so the normality is approximated very well.

Table 4.2 Series 1.

	1	2	3	4	5	6	7	8	9	10	*DN*
A1	**0**	**0**	**0**	**0**	**0**	**100**	**0**	**0**	**0**	**0**	**0**
A2	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**100**	**0**
B1	**0**	**0**	**0**	**0**	**50**	**50**	**0**	**0**	**0**	**0**	**11**
B2	**50**	**50**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**11**
D	**20**	**20**	**20**	**20**	**20**	**0**	**0**	**0**	**0**	**0**	**36**
C1	**0**	**3**	**6**	**14**	**27**	**27**	**14**	**6**	**3**	**0**	**37**
C2	**3**	**6**	**14**	**27**	**27**	**14**	**6**	**3**	**0**	**0**	**37**
F	**0**	**10**	**10**	**20**	**10**	**10**	**20**	**10**	**10**	**0**	**55**
H	**0**	**0**	**50**	**0**	**0**	**0**	**0**	**50**	**0**	**0**	**56**
E	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**73**
G	**10**	**10**	**20**	**10**	**0**	**0**	**10**	**20**	**10**	**10**	**77**
K	**50**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**50**	**100**

	mean	st.dev.	CV%	GINI	Theil	Int.R.	MAD	MPD	POM
A1	**6,00**	**0,00**	**0**	**0,00**	**0,00**	**0,50**	**0,00**	**0,00**	**0**
A2	**10,00**	**0,00**	**0**	**0,00**	**0,00**	**0,25**	**0,00**	**0,00**	**0**
B1	**5,50**	**0,50**	**9**	**0,05**	**0,004**	**1,00**	**0,50**	**0,50**	**50**
B2	**1,50**	**0,50**	**33**	**0,17**	**0,06**	**1,00**	**0,50**	**0,50**	**50**
D	**3,00**	**1,41**	**47**	**0,27**	**0,12**	**2,50**	**1,20**	**1,60**	**80**
C1	**5,50**	**1,50**	**27**	**0,15**	**0,04**	**1,85**	**1,20**	**1,65**	**73**
C2	**4,50**	**1,50**	**33**	**0,18**	**0,06**	**1,85**	**1,20**	**1,65**	**73**
F	**5,50**	**2,16**	**39**	**0,22**	**0,08**	**3,50**	**1,90**	**2,46**	**80**
H	**5,50**	**2,50**	**46**	**0,23**	**0,11**	**5,00**	**2,50**	**2,50**	**50**
E	**5,50**	**2,87**	**52**	**0,30**	**0,15**	**5,00**	**2,50**	**3,30**	**90**
G	**5,50**	**3,07**	**56**	**0,31**	**0,17**	**5,50**	**2,67**	**3,46**	**80**
K	**5,50**	**4,50**	**82**	**0,41**	**0,39**	**9,00**	**4,50**	**4,50**	**50**
tauB		**1,00**	**0,82**	**0,83**	**0,82**	**0,92**	**0,98**	**1,00**	**0,49**

Series 2 (table 4.3)

Series 2 was constructed to detect a possible dependency on the mean. Again the series starts with zero inequality situation (A) and moves on towards the greatest possible inequality in situation K. The trick is that the means differ in distributions that are equally unequal: A 1,2,3; B1, 2; L1, 2; M1, 2 and N1, 2.

Table 4.3 Series 2

	1	2	3	4	5	6	7	8	9	10	*DII*
A1	0	0	0	0	0	100	0	0	0	0	0
A2	0	0	0	0	0	0	0	0	0	100	0
A3	0	0	100	0	0	0	0	0	0	0	0
B1	0	0	0	0	50	50	0	0	0	0	11
B2	50	50	0	0	0	0	0	0	0	0	11
L1	0	0	0	0	50	0	50	0	0	0	22
L2	50	0	50	0	0	0	0	0	0	0	22
M1	50	0	0	0	0	50	0	0	0	0	56
M2	0	0	50	0	0	0	0	50	0	0	56
N1	0	50	0	0	0	0	0	50	0	0	67
N2	0	0	0	50	0	0	0	0	0	50	67
K	50	0	0	0	0	0	0	0	0	50	100

	mean	st.dev.	CV%	GINI	POM	Theil	Int.R.	MAD	MPD
A1	6,00	0,00	0	0,00	0	0,00	0,50	0,00	0,00
A2	10,00	0,00	0	0,00	0	0,00	0,25	0,00	0,00
A3	3,00	0,00	0	0,00	0	0,00	0,50	0,00	0,00
B1	5,50	0,50	9	0,05	50	0,004	1,00	0,50	0,50
B2	1,50	0,50	33	0,17	50	0,06	1,00	0,50	0,50
L1	6,00	1,00	17	0,08	50	0,014	2,00	1,00	1,00
L2	2,00	1,00	50	0,25	50	0,13	2,00	1,00	1,00
M1	3,00	2,00	67	0,42	50	0,28	5,00	2,50	2,50
M2	5,00	2,00	40	0,25	50	0,11	5,00	2,50	2,50
N1	5,00	3,00	60	0,30	50	0,19	6,00	3,00	3,00
N2	7,00	3,00	43	0,21	50	0,09	6,00	3,00	3,00
K	5,50	4,50	82	0,41	50	0,39	9,00	4,50	4,50
tauB		1,00	0,80	0,75	0,44	0,77	0,98	1,00	1,00

Series 3 (table 4.4)

In this case, only the lowest and the highest possible ratings are used, but at different frequencies. Again the degree of inequality increases from A to K. The two distributions S1 and S2 are 'equally unequal' and enable us to establish whether the various statistics confirm this.

Table 4.4 series 3

	1	2	3	4	5	6	7	8	9	10	*DII*
A4	0	0	0	0	0	0	0	0	0	100	0
P	5	0	0	0	0	0	0	0	0	95	19
Q	10	0	0	0	0	0	0	0	0	90	36
R	15	0	0	0	0	0	0	0	0	85	51
S1	20	0	0	0	0	0	0	0	0	80	64
S2	80	0	0	0	0	0	0	0	0	20	64
T	25	0	0	0	0	0	0	0	0	75	75
U	30	0	0	0	0	0	0	0	0	70	84
V	35	0	0	0	0	0	0	0	0	65	91
W	40	0	0	0	0	0	0	0	0	60	96
X	45	0	0	0	0	0	0	0	0	55	99
K	50	0	0	0	0	0	0	0	0	50	100

	mean	st.dev.	CV%	GINI	Theil	Int.R.	MAD	MPD	POM
A4	10,00	0,00	0	0,00	0,00	0,25	0,00	0,00	0
P	9,55	1,96	21	0,05	0,03	0,26	0,86	0,86	5
Q	9,10	2,70	30	0,09	0,07	0,28	1,62	1,62	10
R	8,65	3,21	37	0,13	0,11	0,29	2,30	2,30	15
S1	8,20	3,60	44	0,18	0,14	0,31	2,88	2,88	20
S2	2,80	3,60	129	0,51	0,62	0,31	2,88	2,88	20
T	7,75	3,90	50	0,22	0,18	8,33	3,38	3,38	25
U	7,30	4,12	56	0,26	0,22	8,41	3,78	3,78	30
V	6,85	4,29	63	0,30	0,26	8,45	4,10	4,10	35
W	6,40	4,41	69	0,34	0,30	8,48	4,32	4,32	40
X	5,95	4,48	75	0,37	0,35	8,50	4,46	4,46	45
K	5,50	4,50	82	0,41	0,39	8,50	4,50	4,50	50
tauB	---	1,00	0,87	0,87	0,87	1,00	1,00	1,00	1,00

Series 4 (table 4.5)

This series depicts another case of equally unequal distributions that differ in mean value. In this case we compare two triangular distributions only. The skewness of Y1 is negative ('to the left') and that of Y2 is positive ('to the right') to the same extent. For both Y1 and Y2 *DII* =59.

Table 4.5 Series 4. Influence of skewness

	1	2	3	4	5	6	7	8	9	10
Y1	1	3	5	7	9	11	13	15	17	19
Y2	19	17	15	13	11	9	7	5	3	1

	mean	st.dev.	CV%	GINI	Theil	Int.R.	MAD	MPD	POM
Y1	7,15	2,35	33	0,18	0,06	3,65	1,97	2,63	81
Y2	3,85	2,35	61	0,34	0,19	3,65	1,97	2,63	81

Series 5 (table 4.6)

Lastly Series 5 was constructed to assess dependency on scale range. We compared two uniform distributions with different numbers of possible ratings, a ten-point scale for E and a five-point scale for Z. Comparison of the inequality statistics is only possible after a linear scale transformation of the ratings onto a [1; 10] scale: {1; 2; 3; 4; 5} => {1; 3,25; 5,50; 7,75; 10}. The procedure of this transformation is described in Appendix B.

Table 4.6 Series 5. Influence of scale length

	1	2	3	4	5	6	7	8	9	10
Z	20	20	20	20	20					
E	10	10	10	10	10	10	10	10	10	10

	mean	st.dev.	CV%	GINI	Theil	Int.R.	MAD	MPD	POM
Z	3,00	1,41	47	0,27	0,12	2,50	1,14	1,60	80
ZZ	5,50	**3,19**	**58**	**0,33**	0,18	**5,00**	**2,57**	**3,60**	**80**
E	5,50	**2,87**	**52**	**0,30**	0,15	**5,00**	**2,50**	**3,30**	**90**

ZZ := Z corrected allowing for the transformation [1, 5] => [1, 10]

4.6 How the statistics perform

We formulated eight evaluation criteria (section 4.3) to assess the usefulness of the nine statistics described in section 4.2 for quantifying inequality in happiness in nations. These criteria are listed below with a review of how well the nine statistics performed for each criterion. The evaluation is based on the values the statistics yield when applied to the hypothetical distributions constructed in section 4.5. These values are presented in the lower parts of tables 4.2 to 4.6.

Criterion 1: Single finite number as result

All statistics meet the condition that it should express the degree of inequality value in a single finite number for any conceived distribution of happiness.

Criterion 2: Interval level of measurement

Candidates should be applicable to the distribution of variables measured at the interval level of measurement.

Not every inequality statistic is applicable at any level of measurement. The following restrictions apply to the statistics enumerated in section 4.3.

LEVEL OF MEASUREMENT	STATISTIC OF INEQUALITY
Ratio level only	Coefficient of variation Gini coefficient Theil's entropy measure
Ratio and interval level	Standard deviation Mean absolute difference Mean pair distance Interquartile range Range
All levels	% outside mode

The coefficient of variation, the Gini coefficient and Theil's measure require a higher level of measurement; that is, the ratio level. Since happiness measurements do not meet this condition, this disqualifies these three statistics.

Criterion 3: Independence of scale range

The application of a candidate statistic should be independent of the number of possible ratings on the scale of measurement or at worst weakly dependent. This requirement was tested in Series 5. As can be seen from table 4.6 after linear scale transformation of the ratings, all the statistics show an influence of the number of possible ratings. The mean absolute deviation shows the smallest relative difference, whereas the largest is found for Theil's measure. In the other cases the difference between k = 5 and k = 10 was less than 10 %, except for the Gini coefficient and the percentage outside the mode. For nonuniform distributions larger differences are to be expected for the percentage outside the mode.

Criterion 4: Independence of sample size

The values of the statistics must be independent of sample size, at least where large samples are concerned. Ignoring effects of factors like $(N-1)/N$, there are no reasons to expect any influence from sample size, except possibly for Theil's measure in view of the term $\ln(N)$.

Criterion 5: Independence of the mean

A useful measure of inequality should be fully independent of the average value or at least only weakly dependent on it, ignoring intrinsic dependence as has been mentioned in section 2.8.

We have announced in section 4.3 that problems could be expected with three statistics: the coefficient of variation, the Gini coefficient and Theil's inequality measure. This expectation was tested by comparing the values found for distributions that are equally unequal, but differ in central tendency. These distributions are denoted by the same letter and a different number, e.g. in table 4.2 the distributions A1 and A2, B1 and B2 and D1 and D2. Going through these columns we can see that four statistics yield identical values: standard deviation, mean absolute deviation, mean pair distance and the % outside the modus. The same pattern emerges in the tables 4.3, 4.4 and 4.5. The other statistics produce different values. These are: the coefficient of variation, the Gini coefficient, Theil's entropy measure and for extreme distributions also the interquartile range. The nature of this dependency is not stochastic, since this variant of dependency has not been introduced; to demonstrate its existence would require a number of large random samples taken simultaneously from the same population. This structural dependency was not established for the standard deviation, the mean absolute deviation, the mean pair distance and the percentage outside the mode, but some stochastic dependency cannot be excluded.

The conclusion is that, as a statistic of inequality, the Gini coefficient may possibly have a sound conceptual basis in its relationship with the mean pair distance, but that the need to divide by almost twice the mean happiness value (Eq. [4.1]) makes it an inadequate test for inequality of happiness.

Criterion 6: Equal values for equally unequal distributions

A basic requirement of any statistic is that it yields equal values for distributions considered 'equally unequal'. The comparison of values produced for distributions denoted with the same character also shows whether statistics yield the same values for distributions, which we consider to be equally unequal. Series 4 on table 4.5 is particularly instructive and disqualifies the coefficient of variation, the Gini coefficient and Theil's entropy measure.

Criterion 7: Differentiation between more and less unequal distributions

A usable statistic distinguishes between distributions that we consider to be 'unequally unequal'. The indicator should also be sensitive to different ways of inequality. A glance at the tables 4.2, 4.3 and in particular 4.4 shows us that most statistics can be used to differentiate between distributions that differ in degree of inequality, that is, distributions denoted with a different character. Yet not all the statistics perform equally well on this criterion.

The percentage outside the mode fails to pick up several differences. One can see in table 4.2 that this statistic yields the same values for the distributions B, H and K and also identical values for the distributions D, F and G. Likewise, in table 2b we see identical values for the distributions B, L, M, N and K.
In table 4.2, the mean absolute difference does not show a difference in inequality between the distributions C and D, nor between E and H.
Irregularities in Series 2 and 3 with respect to the coefficient of variation, the Gini coefficient and Theil's measure arise from mean value differences.

Criterion 8: Sensitivity to the degree of inequality

To be usable, a statistic must finally reflect the degrees of inequality in a distribution, and the values it generates must fit in our notion of what is more, or less, equal.

The degree to which this requirement is met was judged by considering the degree to which a statistic yielded higher values for distributions that are considered more unequal on the basis of the Discrete Inequality Index *DII*. In the tables 4.2 – 4.6, the distributions have been ranked according to increasing, more precisely non-increasing *DII*-values. A look at the tables 4.2, 4.3 and 4.4 shows that the values tend to get higher if we move from low to high *DII* values. Yet one can also see that the increase is not equally consistent in all cases. The degree of consistency in the succession can be quantified by computing Kendall's tau-B rank order coefficient (Kendall, 1962[3]: 4). These tau-B's are presented at the bottom of the tables 4.2, 4.3, 4.4 and 4.5. Table 4.2 shows that most statistics reflect the degree of inequality fairly to very consistently. For the standard deviation and the mean pair distance, the rank correlation is perfect, and for the mean absolute difference it is almost perfect. The other tau-B's vary between 0,82 and 0,92 and the mutual differences are modest. Only the percentage outside the mode is a poor performer here.

In series 2 and 3, the rank correlation is perfect for the standard deviation, the mean absolute deviation and the mean pair distance, while for series 3 this also holds for the interquartile distance and exceptionally even for the percentage outside the modus this time.

Overall evaluation

The performance of the nine statistics is summarized in table 4.7. This overview clearly shows that five of the nine statistics considered can be disqualified as a means for quantifying inequality of happiness in nations. These inapt measures are: (1) coefficient of variation, (2) the Gini coefficient, (3) Theil's entropy measure, (4) percentage outside the mode and (5) the range. The interquartile range has in general a good performance, but not always a very good one.

Table 4.7 Summary performance of nine descriptive statistics for happiness inequality in samples

	I numerical result	**II** interval level	**III** dependent on *k*-value	**IV** indepen-dent of *N*	**V** dependent on mean	**VI** 'equal for equal'	**VII** 'unequal for unequal'	**VIII** sensitive to inequality
Standard deviation	OK	OK	Weakly	OK	OK	OK	OK	OK
Mean absolute deviation	OK	OK	Weakly	OK	OK	OK	OK	OK
Mean pair distance	OK	OK	Weakly	OK	OK	Ok	OK	OK
Interquartile range	OK	OK	Weakly	OK	OK	OK	OK	OK
Coefficient of variation	OK	**FAILS**	Weakly	OK	**FAILS**	**FAILS**	OK	**FAILS**
Gini coefficient	OK	OK	Weakly	OK	**FAILS**	**FAILS**	OK	**FAILS**
Theil's measure	OK	OK	**FAILS**	??	**FAILS**	**FAILS**	**FAILS**	**FAILS**
Perc. outside mode	OK	OK	**FAILS**	OK	OK	OK	**FAILS**	**FAILS**
Range	OK	OK	OK	OK	OK	OK	**FAILS**	**FAILS**

The three remaining statistics meet the demands required for determining inequality of happiness in nations. Appropriate measures are (1) standard deviation, (2) mean absolute difference, and, (3) mean pair distance. There is no clear winner among these three suitable statistics; all perform about equally well. For the mean pair distance, this finding is not surprising, since the distributions within one series have been ranked according to their *DII* value, which in turn has been defined on the basis of the mean paired distance in the sample (section 3.3). If all distances were squared, this would result in an index in which the standard deviation plays a similar role (section 3.3.1), so the assessment result on the standard deviation is also not unexpected.

Performance on real distributions

One may wonder how the statistics perform when applied to real, non-hypothetical distributions. Obviously, this question is meaningful only for the statistics that are considered acceptable on the basis of our earlier findings. To this end, we selected the happiness distributions of eight Eastern European countries all for the same period (1999-2000), all from response to the same question “All things considered, how satisfied are you with your life as-a-whole in these days?” and all using a 10-step numerical rating scale. The distribution data were taken from the World Database of Happiness, Section Nations.

The acceptable inequality statistics were computed for each distribution and the nations have been ranked according to the standard deviations, to give table 4.8.

Table 4.8 Inequality statistics from the same distribution for each of 8 different countries

Nation	Year	Standard deviation	Mean absolute distance	Mean pair distance	Inter-quartile range
Romania	1999	2,77	2,39	3,17	4,77
Bulgaria	1999	2,65	2,24	3,03	4,32
Ukraine	1999	2,59	2,16	2,94	4,01
Russia	1999	2,57	2,14	2,92	4,03
Poland	1999	2,53	2,10	2,86	3,49
Hungary	1999	2,42	1,99	2,74	3,34
Moldova	2000	2,32	1,86	2,61	3,11
Belarus	2000	2,21	1,78	2,50	3,28

The ranking of the eight nations for both the mean absolute distance and the mean pair distance, is identical to the ranking obtained on the basis of the standard deviation. The ranking of the interquartile range shows two inversions (shaded) with respect to those of the other three statistics.

The above results support the conclusion that no advance in understanding is to be expected from switching from using the standard deviation of a happiness sample data set to some other inequality statistic.

4.7 The failing of the Gini coefficient to quantify happiness inequality

We have introduced the Gini coefficient (*GC*) as a potential indicator of the quantification of the happiness inequality within a sample (section 4.2.7). Its value has been calculated using a simple formula, which is based on the relationship with the mean pair distance *MPD* and the sample average happiness *m*. This formula $GC \approx MPD/2m$ for not too small samples has been derived as Eq. [C.9] in Appendix C.

The Gini coefficient has been disqualified in the section 4.6 as a measure of happiness inequality in samples, mainly on the basis of its unsatisfactory rank correlation with the discrete inequality index *DII*. This result is related to the fact that the *GC* assumes that happiness is measured at the ratio level of measurement. Moreover, in view of the way the *GC* value has been obtained, it becomes clear that, since *MPD* is independent of *m*, *GC* must depend on the mean, which is confirmed by the observations.

Eq. [C.9] has been derived on the basis of the assumption that there is a line that acts as a Lorenz curve. Without a Lorenz curve, the *GC* is nonexistent by definition, however, for the distribution of income as a continuous variable, the number of different incomes is sufficiently large to justify the assumption that a Lorenz curve exists, but for a discrete distribution of the happiness variable *H*, there are only $k-1$ nontrivial points. In Appendix C, the problem is 'solved' by drawing straight lines between consecutive happiness points, however, in the case of a discrete distribution there are no more than $k+1$ points, including the point (0,0), since

$$F(2{,}87) := \text{Prob}\,\{H \leq 2{,}87\} = \text{Prob}\{H \leq 2\} = F(2)$$

and similarly $\Phi(2{,}87) = \Phi(2)$, which also applies to any other decimal; the cumulative distribution functions $F(\cdot)$ and $\Phi(\cdot)$ are used as defined the Eq. [C.1] and [C.2] in Appendix C respectively. Since there are no more points than the $k-1$ happiness ratings, the Lorenz curve is nonexistent in the case of a discrete distribution and so is the *GC*.

Some may argue that the 'continuisation' way-out as described in section 2.7 can be assumed to justify the broken-line solution, but this is not correct. If we assume the situation is as shown in Fig. 2.5 (right hand), then between the happiness values H = 1,5 and 2,5 the happiness distribution is assumed to be uniformly continuous.

In that case, anywhere on that interval:

[4.3] $dF = f_2 dh$, and

[4.4] $d\Phi \sim hdh = c_2 hdh$, with c_2 = a proportionality constant.

From these two equations it follows that

[4.5] $$\frac{d\Phi}{dF} = ch,$$

with $c := c_2/f_2$ as a constant. In view of [4.4], this means that a Lorenz curve over the *happiness* interval [1,5, 2,5] is not a straight line, but a part of a convex parabola. A *GC* in this model may be no longer nonexistent, but a value computed from Eq. [C.9] will be negatively biased.

This way-out, however, does not really solve all further problems, since the GC has been devised as a measure of income inequality and not to quantify inequality of happiness. Some numerical examples will demonstrate this. If e.g. in a sample of 100 persons, one person selects a rating 4 on an equidistant [1, 4] scale and all others a rating 1, the Gini coefficient is not at all close to unity, as in the case of a similar income distribution, but now the coefficient = 0,03 only ! Moreover, if in that situation the inequality is maximal, i.e. when half the sample selects a rating 1 and the other half a rating 4, the Gini coefficient = 0,33, which actual upper limit value is independent of the sample size, but slightly decreases as the number of response categories increases. Clearly, this limit remains far below the theoretical upper limit which is equal to unity.

Inequality of happiness and inequality of income appear to be very different matters.

We have paid extensive attention to the differences between happiness and income as variables in section 2.6. Although in theory, the Gini coefficient looked like a rather poor candidate for the quantification of happiness inequality, it was included in our assessment, partly for the sake of completeness, but mainly to prove our expectations empirically. The above considerations and findings demonstrate this unambiguously and make also clear why the Gini coefficient is entirely inapt to serve as an indicator of happiness inequality within nations.

4.8 Conclusion

Of the nine statistics considered here, five are not suitable for quantifying inequality of happiness in nations, both for theoretical and empirical reasons. These are the coefficient of variation, the Gini coefficient, Theil's entropy measure, the percentage outside the mode and the range.

The standard deviation is the most commonly used statistic for measuring inequality of happiness in nations and in tests it performs equally as well as two other statistics of disparity; i.e. the mean absolute difference and the mean pair distance. The interquartile range falls between both classes: often, it performs very well, but in general not as well as the other three preferred statistics. Hence our conclusion is that there is no reason to discontinue the use of the standard deviation.

The above conclusions and recommendations hold within the context of approaching the measurement of happiness as essentially an intensity variable rather than as an extensity variable on the basis of an 'amount of happiness', that can be distributed over the members of a nation. The study described in this chapter has been triggered by the recommendations of economists that the Gini coefficient as an inequality measure is superior over the standard deviation as a measure of happiness inequality. Our research has delivered overwhelming evidence for the contrary, thus our recommendation is to ignore the Gini coefficient as a measure of happiness inequality of samples.

✧ ✧ ✧

Chapter 5

Inequality-adjusted happiness in nations

Parts of this chapter have been published in:
Veenhoven, R. & Kalmijn, W.M. (2005)
Inequality-Adjusted Happiness in Nations. Egalitarianism and Utilitarianism Married in a New Index of Societal Performance.
Journal of Happiness Studies, vol. 6 (4), 412 – 455.
1389-4978 (Print) 1573-7780 (Online)
DOI 10.1007/s10902-005-8857-5

5.1 Background

We have pointed out in section 1.5, that, according to the utilitarian creed, the quality of a society should be judged using the degree of happiness of its members, the best society being the one that provides the greatest happiness for the greatest number. Following the egalitarian principle, the quality of a society should rather be judged by the disparity in happiness among citizens, a society being better if differences in happiness are smaller.

This calls for the use of appropriate social indicators; policy makers must know what interventions are most likely to serve both principles. This requires a measure that marries happiness level and inequality in the research arena.

A similar problem exists in public health. One guiding principle in this field is to preserve life for as long as possible and performance on that criterion is commonly measured using average life expectancy. Yet another moral lead is to promote good health, which is typically measured using surveys of self-reported disabilities. These goals can also come into conflict, since longevity can come at the cost of good health. People can be kept alive, but with a poor quality of life, reflected in their having to deal with bad health for too long. Good health can, in some cases, come at the cost of longevity if its maintenance requires therapies that shorten life. How to find a balance between a short and healthy life and a prolonged but unhealthy life? Policy makers in this field needed an outcome measure that reflects an acceptable mix of these aims.

In response the World Health Organization proposed a combined measure, called Disability-Adjusted Life Years; abbreviated to DALY's, which was used for the first time as an outcome criterion in a worldwide comparison of national healthcare systems (WHO 2002).

Likewise, the Human Development Index has been adjusted for inequality. A 'Gender related Development Index' (GDI) was proposed in the Human Development Report of 1995 and a correction for poverty was introduced in the 1997 report (UNDP 1998). Further Hicks (1997) proposed a variant of the HDI that adjusts for inequalities in education and longevity in nations.

As for happiness, performance on these standards can be measured using cross-national surveys, where degree of happiness is measured using the mean response to a question about happiness and disparity is expressed as the standard deviation. In this chapter we marry these measures together in an index of 'Inequality-Adjusted Happiness' (IAH) that gives a weight to either view. Although the method is elaborated into more detail for the case of equal weights, it is applicable to any choice in this respect. The measure is a linear combination of the average happiness value and the standard deviation and it is expressed as an integer number on a 0 to 100 scale.

5.2 Options for combination

As indicated above, we measure happiness and inequality in nations using responses to questions about happiness to be found in general population surveys. The degree of happiness in nations is measured using the average, and inequality in happiness using the standard deviation: How can these pieces of information best be combined?

The possible configurations of the average and the standard deviation of the responses to the item on happiness in a nation are depicted in Figure 5.1. The average or mean is denoted by the symbol 'm' and is plotted on the horizontal axis, and varies between u, the rating corresponding to the most unhappy conceivable situation, and h for the most happy one. We assume that $u < h$, so $u \leq m \leq h$. The standard deviation is denoted by the symbol 's' and is plotted vertically. All the theoretically possible combinations of the mean and the standard deviation lie within this semicircle or at its circumference. We have presented a formal derivation of this diagram in Appendix A.

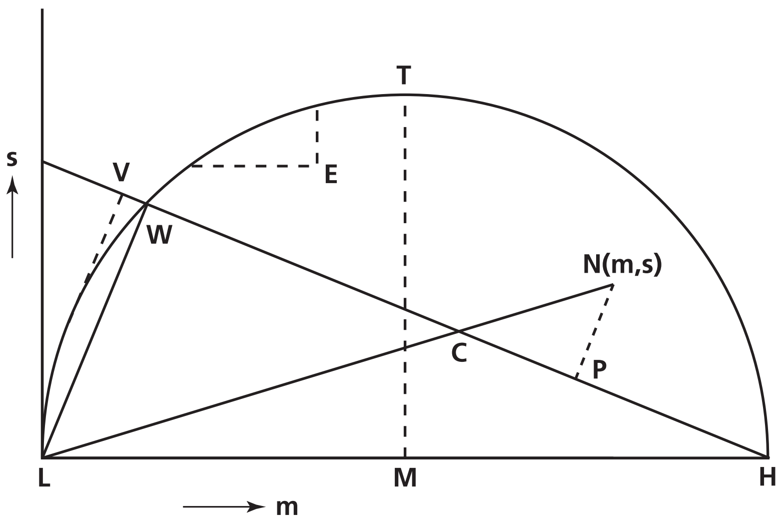


Figure 5. 1 Projection of a nation *N* on WH as an IAH-axis

Mathematically, the problem is to map the points in this two-dimensional vector space onto a one-dimensional subspace. The positions in the latter space must reflect the degree to which societies meet these values. Utilitarians and egalitarians will agree that no better society is conceivable than the one that is represented by point H, albeit for different reasons. Yet they will disagree about the worst possible society. Egalitarians will select point T and for utilitarians this is the point L. If one selects some point E inside the semicircle, both views agree on the fact that any other point for which the mean is smaller and at the same time the standard deviation is larger than that of E represents a society that is worse than E. Their arguments are different, but they agree on the conclusion. Therefore, any 'compromise' between both principles on what is the worst conceivable society is like must be represented by a point on the circumference of the semicircle, somewhere between T and L. The exact position of this point W depends on the weights that are assigned to both views.

Common good-bad dimension

An obvious choice for the one-dimensional space we are looking for is a straight line through H and W, in such a way that good societies will be mapped close to H and bad ones will be to be found nearer to W. The point N in the *m*-*s* diagram with abscissa = *m* and ordinate = *s* represents some society with this average value and standard deviation respectively. Its projection onto the line HW can be made in various ways. We will consider two of them: the orthogonal projection and the central projection.

Orthogonal projection

If for N the image is chosen that is the nearest to N, the result is the orthogonal projection of the point N onto HW, i.e. the point of intersection of the line HW and a straight line through N and perpendicular to HW. In Fig. 5.1, this intersection is denoted P. Now we define the inequality-adjusted happiness (IAH) as:

[5.1] $IAH_o := (\underline{ZP} / \underline{ZH}) \times 100,$

where Z is some zero point, which will be defined later, and *ZP* denotes the length of the (straight) line segment ZP. The subscript 'o' means that the projection is orthogonal.

Central projection

In the case of central projection, one has to select a centre of projection outside HW. Now the point N is connected to this centre by a straight line, and its point of intersection with HW, denoted C, is the central projection of N with respect to the centre. For this centre, we made a choice in favour of point L. In this case, inequality-adjusted happiness will be defined as:

[5.2] $IAH_c := (\underline{ZC} / \underline{ZH}) \times 100.$

Scale properties

Different options are available for the point Z. One is the point W. This means that in that case the projection W corresponds to an *IAH*-value of 0 and the *IAH*-value of H is 100. However, there is one disadvantage, at least in theory: all points in the semicircle segment LW will be projected to the left of W, resulting in a negative *IAH*-value.

In Appendix D, it is shown that, in the case of equal weights to the views of egalitarians and utilitarians, this situation occurs only when on a scale with 0 and 10 as lowest and highest score respectively, the average happiness of a nation is less than 1,46 and at the same time the ratio of the standard deviation and the mean exceeds the value 2,41. Until now, we have found no nation for which this outcome has been reported, so the above objection appears to be merely theoretical, however, as more weight will be given to the strictly egalitarian view, W approaches T and the size of the segment LW will increase; in this situation negative index values may eventually become a reality.

Therefore, if one wishes to avoid negative index values, including theoretically possible ones, one has to establish which possible projection is maximally remote from H and to select this point as Z. For the orthogonal projection, this is the point V, being the point of intersection of HW and the tangent to the semicircle that is perpendicular to HW; for central projection, it is the point of intersection of HW and the left-hand vertical tangent.

In both cases, the scale value on the index-scale is different from the one on the "short axis", where W is selected as a zero. When a choice is made in favour of a longer axis, a relatively smaller segment of the scale is used for real situations. Moreover, the calculation of *IAH* is slightly more complicated.

Computation

The value of *IAH* can be calculated for each of these variants from *u*, *h*, *m* and *s* using a formula that is derived in appendix D.

Inspection of the formulae [D.6] and [D.7] for IAH_c in Appendix D shows that, in the case $u = 0$, this statistic is a monotonically increasing function of *m/s* (the ratio of the mean and the standard deviation) only. This means that, in this case, a ranking of societies according to their IAH_c - values is identical to the one on the basis of their mean/standard deviation ratios. Veenhoven (2003a, b) has used this ratio as a measure of inequality-adjusted happiness.

An advantage of using IAH_c over that ratio is that it results in an index scale that ranges from 0 to100, which makes the comparison of nations for happiness somewhat easier. Moreover, in contrast to the ratio mean/standard deviation, IAH_c is also a meaningful statistic in the case of a scale with $u \neq 0$, including 'reversed scales'.

The main advantage of IAH_c (and IAH_o) is that their values are basically independent of the underlying measuring scale and are at least insensitive to linear scale transformation; linear transformation of scores to a secondary rating scale is a procedure that has been described in Appendix B.

A serious problem with respect to the central projection arises when a relatively large weight is assigned to the utilitarian view. In this case, the

IAH-axis approaches the *m*-axis and the paradoxical result is a small *IAH*-value for almost all societies, which makes distinction between them very difficult. Index-values will, eventually, hardly depend on the mean value, which is just the only criterion of the utilitarian pure-sang.

In the case where full weight is given to the utilitarian view, projection is even impossible, since the centre of projection is no longer outside the projection axis and all societies will have a zero *IAH*-value. When the orthogonal projection is selected, which is essentially a linear transformation of the (*m, s*) vector onto a one-dimensional subspace (*IAH*), these problems do not occur.

When most or even all the weight is given to the strictly egalitarian view, in case of both central and orthogonal projection, societies with equal standard deviations get unequal *IAH*-values, which may differ substantially; in this case they are ranked according to their mean values. The only exception is when $s = 0$ and the mean $m > u$: in the case of central projection $IAH = 100$, irrespective of *m*, whereas in the case of orthogonal projection, $IAH < 100$ and increases with the value of *m*.

In the case of a choice in favour of a strictly utilitarian view, orthogonal projection will give a projection onto the *m*-axis, and this is to be considered to be a sound result. In the case of a zero weight to the utilitarian view, different situations with different *m*-values, but all with zero standard deviation, are mapped in a way that seems acceptable from both points of view.

The ratio m/s is easily recognized as the reciprocal of s/m, a statistic that is often called the "relative standard deviation" or the "coefficient of variation" and is usually reported as a percentage. This statistic is a measure for the dispersion in a distribution. As such it is defined only if the variable is measured at the ratio level of measurement. However, happiness is measured at best at the interval level. At first glance, one might conclude that, if the coefficient of variation is not defined and hence does not exist, its reciprocal value cannot exist. This conclusion is not correct.
The condition that the variable is to be measured on the ratio scale arises from the fact that it should have a natural zero. The problem is not that m occurs in the ratio s/m, but that m occurs just in its denominator. For its reciprocal ratio, this problem does not exist for s, since s is defined in such a way that it is nonnegative and has a natural zero, and in a way can be considered to be a variable at the ratio level. Hence the fact that s/m is not defined is not an argument in itself against the use of m/s in this index of Inequality-Adjusted Happiness.

5.3 Our choice

The above considerations leave us with three problems:

(1) How do we weigh the utilitarian and the egalitarian approach?
(2) Do we project orthogonally or centrally?
(3) Do we express the combined index on a short or a long scale?

We made the following choices:

Equal weights

We opted for a combination that gives equal weight to the utilitarian and egalitarian principles. Though this choice may be arbitrary, it is a clear one and no less arbitrary than any other choice. In terms of Figure 5.1, this means that we locate point W half way between T and L on the semicircle circumference

Orthogonal projection

Central projection might be an obvious choice, since it can be easily interpreted as related to the ratio of the mean to the standard deviation and fits earlier use of the ratio of the mean and the standard deviation as a measure of Inequality-Adjusted Happiness (Veenhoven 2003a, 2003b). As we have seen, however, this projection method gives rise to problems, which become more serious as more weight is given to the strictly utilitarian view on happiness. It could be argued that these objections are mainly theoretical and can be ignored as being practically irrelevant for two reasons. One, we have already made a decision in favour of equal weights. Two, the problems with very small or even zero standard deviations can arise only at a very small number of distinct mean values (section 5.2). Such values of the standard deviation are all well below the ones that have found for nations until now, since none of the 140 nations listed in the WHD[1] shows a standard deviation below 1,4 on a [0, 10] scale of measurement.

These problems do not occur in the case of orthogonal projection, thus for reasons of generality, we prefer to select the orthogonal projection method.

Long scale

Finally we opted for the long scale option, because this excludes the possibility of negative values under all circumstances.

[1] *Available at: http://www.worlddatabaseofhappiness.eur.nl/hap_nat/nat_fp.php*

Formula

This variant of Inequality-Adjusted Happiness index can be computed using the following formula, the derivation of which is explained in Appendix A, where it follows from [D.5] and $b := |h - u|$:

[5.3] $$IAH_o = 96{,}0(|m-u| - 0{,}414 \cdot s)/(|h-u|) + 3{,}96,$$

where *m* is the mean score on an indicator of happiness in a society, *u* and *h* are ratings that correspond to the most unhappy and happy situations respectively, and s is the standard deviation of the distribution of the happiness ratings. Rounding of *IAH*-values to integers is recommended.

> *From this formula, it follows that for m = u (then s = 0), $IAH_o = 3{,}96 \approx 4$. The reason why in this case $IAH_o > 0$ is that the choice of the worst possible society is a compromise between two views: a society with $IAH_o = 2$ is less attractive than one with $IAH_o = 4$, but only from a utilitarian point of view.*

5.4 Differences across nations

We can now proceed to consider the actual scores on this index. To do this we used the following item O-SLW/c/sq/n/10/a , that has been used in 140 nations, mainly in the World Value Surveys[2].

"Taking all together, how satisfied or dissatisfied are you currently with your life as a whole?"

1	2	3	4	5	6	7	8	9	10
Dissatisfied									Satisfied

Average values and standard deviations obtained in general population surveys using this item were taken from the World Database of Happiness; section 'Distributional Findings in Nations' (Veenhoven 2010). These data were combined using the above formula. The resulting IAH-scores for 140 countries have already been referred to in section 5.3. Fifteen illustrative cases are presented in table 5.1.

[2] *The discrete 0-10 version of this item is applied in the Gallup World Poll and is coded as O-SLW/c/sq/n/11/a.*

Table 5.1 Inequality-Adjusted Happiness in nations in the 1990s
Scores on 0-100 IAH-index; some illustrative cases

BOTTOM 5:		MIDDLE:		TOP 5:	
Burundi	25	South Korea	58	Denmark	78
Benin	22	China	53	Switzerland	74
Zimbabwe	21	South Africa	50	Finland	73
Togo	20	Iran	50	Canada	73
Tanzania	19	India	48	The Netherlands	72

Data: World Database of Happiness, Collection of Happiness in Nations, Rank report_Inequality Ajustedhappiness[3]

Pattern of differences

Costa Rica scores best with 79 points on the IAH-scale, but this score is the result of only one survey; therefore it was not included in this table, and Tanzania scores the worst with only 19 points. These extremes illustrate that we still have a long way to go to achieve the best possible society, which would score 100, but also that we are in most cases well above the theoretically worst possible score of about 0.

The differences make sense at first glance. It will be no surprise that countries like Denmark, Switzerland and The Netherlands perform well, since they have the reputation of being livable and egalitarian. It will be no surprise either to find African countries such as Tanzania and Zimbabwe at the bottom, since life is quite miserable in these countries and inequalities widespread.

The actual variation on this scale is 59 points and the cases are well spread over this range. Presumably , this range may even broaden somewhat when more data on less happy countries become available.

We plotted the standard deviations of happiness against the average value in various nations and incorporated the IAH-axis in that scattergram (Fig. 5.2). The pattern that appears illustrates that the main variability between the countries is more or less in the same direction as that of the IAH-axis and that projection onto the IAH-axis provides a good discrimination between the societies that are more and less successful in meeting utilitarian and egalitarian demands simultaneously.

From Fig. 5.2 it can also be seen that there is more divergence in level and inequality of happiness in the left top part of the scattergram than in the right

[3] *Available at: http://www.worlddatabaseofhappiness.eur.nl/hap_nat/nat_fp.php*

bottom area. Looking closely, we can see a cluster of Latin American nations where high happiness level goes with high inequality. There is also a cluster of former communist countries characterized by a low level of happiness and high inequality.

The confluence of high level and low inequality in the right bottom area of Figure 5.2 would seem the logical result of the fact that the scope of the standard deviation is reduced as the mean is closer to the extremes of the scale. Yet this is probably not the whole story, because comparison with the maximally possible values shows that there is still scope for variation.

> *Computation based on Appendix A gives maximum possible values for the standard deviation of 4,9, 4,6 and 4,0 for the average values 6, 7 and 8 respectively on a [0, 10] scale.*

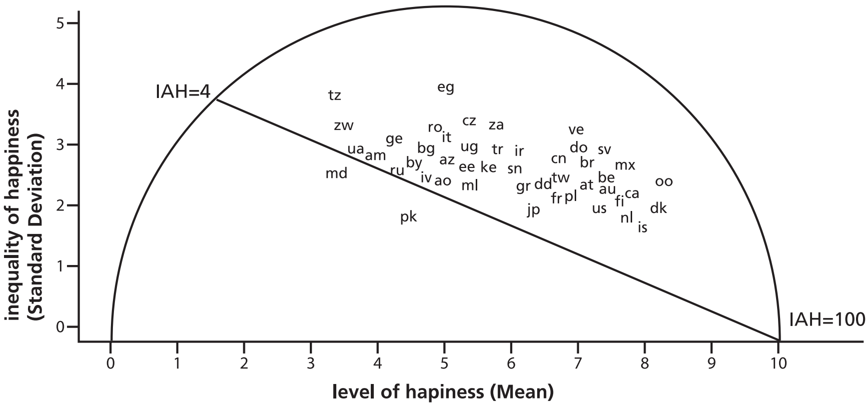


Figure 5.2 Plot of level and inequality of happiness in a selection of 140 nations in the 1990s

Trend Over Time

To be useful for policy evaluation, the IAH must also reflect change over time. Does it? The trends over the last 30 years in the United States and the European Union are increasing: Inequality-Adjusted Happiness rose 3 points in both the USA and in the eight first member states of the European Union. These eight nations, participating in the EuroBarometer survey since 1973, are: Belgium, France, West Germany, Great Britain, Ireland, Italy, Luxembourg and The Netherlands. The above EU IAH increase is a weighted average, weighted by population size.

There are also time series of at least 30 years for Japan and for the European nations separately.

Table 5.3 Trend Inequality-Adjusted Happiness in 11 rich nations
Change 1973-2004 in points on 0-100 scale

	Nation	*IAH*
Significant increase:	Italy	+11
	Denmark	+5
	USA	+3
	Luxembourg	+3
	France	+3
No significant change:	Ireland	+2
	Great Britain	+1
	The Netherlands	+1
	(West) Germany	+1
Significant decrease:	Japan	-4
	Belgium	-6

Data: World Database of Happiness, Distributional Findings in Nations, Trend Report 2005/4c

Inequality-Adjusted Happiness has increased in most developed countries; Italy and Denmark have witnessed particular great gains in IAH. Yet IAH has declined in Belgium and Japan.

The data for the USA presented here should be regarded as a minimum estimate. This trend is based on the responses to a 3-step question on happiness. Responses to the 11-step ladder rating of 'Best-Worst possible life (fig. 2.2) show an increase of about 7 IAH-points over the period 1973-2004.

In most cases the rise in IAH is due to a simultaneous rise in the average level of happiness and a decrease in differences in happiness. In Japan average happiness stagnated while inequality of happiness increased slightly. In Germany the IAH rose until unification in 1990, then it dropped as a result of a slight drop in average happiness and a coincident widening of differences in happiness. Both developments may be due to the temporary costs of the unification and in particular to the massive migration that took place in the country.

Correlation with nation characteristics

One can get a more systematic view on the IAH differences by considering the correlations with quantifiable nation characteristics. Such an investigation is necessary to assess the performance of the IAH in its application. However, following the statement made in section 1.4, this is beyond the scope of this study. The reader who is interested in this assessment is referred to Veenhoven and Kalmijn (2005).

5.5 Discussion

What are the strengths and weaknesses of this new social indicator? Below we will first consider its technical qualities and next its use in the policy process.

Meaning of IAH

This measure of Inequality-Adjusted Happiness is meant to indicate how well a society meets the demands of utilitarian and egalitarian ideology. It does so by adjusting average happiness for inequality in happiness. There are advantages and disadvantages to this combination. The main advantage is that this index conveys a broader meaning than each of its constituents does separately; it provides information about the degree to which both demands are met and warns against attainment of one at the cost of the other. A disadvantage is that a same score can represent different situations, especially in the medium range: an IAH-score of 50 can result from the combination of low average - high inequality, but also from the combination high average - low inequality. Any projection of a vector space onto one with a lower dimensionality gives rise to loss of information and is justified only if this loss is relatively small. The proposal of the IAH-index is an attempt to minimize this loss in a controlled way.

This combination of level and inequality of happiness seems easy to understand and makes more sense than currently used indicators of societal performance such as the Human Development Index (UNDP 2000) and the Index of Social Progress (Estes 1984).

Discriminating power of IAH

We have shown above, that this measure differentiates well among contemporary societies. The scores vary from 19 to 78 on this 0 to 100 scale.

The trend analysis presented in section 5.4 also showed that this measure of Inequality-Adjusted Happiness is sensitive to change over time. The pattern of change observed in 11 rich countries over the last 30 years is not very consistent, but in general, it signifies some social progress.

Data availability

This social indicator is based on responses to questions about happiness in samples of the general population in different nations. At this moment, such data are available for 140 nations and cover about two-third of the world's population. The variation among these nations is sufficiently great to reveal the relationship of *IAH* with societal organization (cf. section 5.4). As yet *IAH* cannot be computed for all the nations of the present world, in particular not for nations in the Middle East and for many nations in Africa. Hopefully this will change in the coming decennia.

As yet time series on happiness are only available for a handful of rich nations and cover no more than 15 to 40 years However increasingly longer time series are emerging from various periodical survey programs, such as the Euro-barometer, the European Welfare Survey, The International Social Survey Program and the World Value Survey.

Policy use

This measure of Inequality-Adjusted Happiness is helpful for policy makers who are trying to raise the average level of happiness in their country while minimizing inequality of happiness. Firstly, observations using this measure can give them information about how far things have to move to reach the ideal situation. In the list of *IAH* values in the WDH[4], they can see the gap between current score of their country and the theoretical maximum of 100.

Table 5.1 also gives information on the gap between what is realistically possible and what country scores best, which is currently (2009) Denmark with a score of 78. Policy makers can also see in the above mentioned listing how their country performs in comparison to other nations and can assess whether they are doing better or worse than similar countries. Lastly, knowing a countries IAH may help policy makers to find ways to improve the performance of their country.

Better than the mere mean?

Is this index of Inequality-Adjusted Happiness more useful for the policy process than just using a simple average happiness as a measure? It depends,

[4] *World Database of Happiness, Rank report Inequality of Happiness, Internet: worlddatabaseofhappiness.eur.nl/hap_nat/findingreports/RankReport_InequalityHappiness.php*

both on the countries under consideration and on the purpose that the measure is used for.

To our knowledge about the most happy nations of the present day *IAH* does not deliver much additional information, since a high level of happiness is typically accompanied with low inequality and therefore produces similar *IAH* scores. Yet among the not-so-happy nations there is less confluence of average and dispersion and is *IAH* therefore a more informative measure. If used for assessing how well the country is doing, *IAH* provides additional information, in particular for the not-so-happy nations. The more mean and standard deviation diverge, the more useful this combination measure.

Moreover, this coincidence may be specific for this set of nations at this present time. It is conceivable that we will get into situations where utilitarian and egalitarian principles dictate different policies and where this index can be used to help to identify workable compromises. The use of this *IAH*-index is that it provides an evidence base for discussions about the best ways to combine the principles of utilitarianism and egalitarianism. It helps to identify the policy directions that do so. Egalitarians will not be convinced by data on average happiness alone.

Public appeal

For the same reason, the *IAH*-index is likely to have considerable public appeal. People have reservations about 'mere' utilitarianism and this principle will be better accepted when combined with egalitarianism, even if this combination is not of real consequence.

5.6 Conclusion

The degree to which a society meets the principles of utilitarianism and egalitarianism simultaneously can be measured using a linear combination of the level and dispersion of happiness. This measure can be expressed as a number on a 0 to 100 scale and is called 'Inequality-Adjusted Happiness', abbreviated as IAH. This measure can be applied to nations and shows good differentiation at this level, both when compared across borders and over time. Scores on this index show how well a country is doing and correlations of *IAH* with societal characteristics indicate ways for policy makers to improve performance.

✧ ✧ ✧

Chapter 6

The Scale-Interval Approach I
Principle, description and methodology.

Parts of this chapter have been published in:
Kalmijn, W.M., Arends, L.R. & Veenhoven, R. (2010)
Measures of Inequality: Application to Happiness in Nations
Social Indicators Research (in press)
DOI 10.1007/s11205-010-9688-2

ABBREVIATIONS USED IN THIS CHAPTER:

df	degree(s) of freedom (section 6.6)
HSIA	Happiness Scale Interval Approach (section 6.3)
HSIS	Happiness Scale Interval Study (section 6.3)
MIV	Mid-Interval Value(s) (section 6.3)
p.d.f.	probability density function (section 6.6)
Prob	probability (section 6.6)
WDH	World Database of Happiness (section 6.1)

6.1 Estimation of the mean happiness level in a nation. The common practice.

Happiness is generally measured by self-report and cross-national studies on happiness mostly use single questions. In section 2.2 we described the measurement method for a typical example of a closed question with four alternative verbal response categories. The question of this item is "Taking all things together, how would you say things are these days - would you say you are ... ?". The four possible response categories are: "unhappy", "not too happy", "pretty happy" and "very happy".

We will use the same example in this chapter to demonstrate the problems arising in the further processing of the observed frequency distribution and some solutions for these. Two such problems have been put forward in section 2.5 which make the method as described in section 2.2 inadequate for application to verbal response items.

One is the cardinalization problem: verbal categories fall in the ordinal level of measurement class and the application of elementary arithmetical operations such as addition and multiplication requires the data to consist of cardinal numbers.

The other problem concerns the (in)comparability of results due the overwhelming variety in all the items ever applied when attempting to quantify happiness and happiness inequality, and which have since been included in the WDH.

The standard solution for the first problems is to code the labels of the categories from text into integer code numbers, which are still ordinal, and to subsequently treat these code numbers as if they were cardinal. A consequence of this procedure is that the categories are automatically made equidistant. This solution enables one to solve the second problem, which is done by direct stretching, i.e. by a linear transformation of all scale points to a common [0, 10] scale, in such a way that the most happy category always gets the value "10" and the least happy the value "0". The four scale values of the above example are then transformed into {0,00; 3,33; 6,67; 10,00}. This procedure is described in Appendix B.

Having 'solved' the above two problems in this way, it is possible to estimate the mean happiness level in a nation. Common practice is to calculate the sample average on the [0, 10] scale by weighing each of the four secondary scale values with the corresponding observed relative frequency and to use the weighted average as an estimate of the mean happiness level of the nation from which the sample had been drawn. A similar procedure is followed to estimate the within-nation standard deviation.

6.2 Thurstone values

The common practice for measuring happiness and their underlying assumptions as has been described in 6.1 are not uncontested, but as long as no suitable alternatives are available, this has hardly any consequences. At least two fundamental elements of this approach are to be criticized.

One, this common practice discards the labels of the response categories immediately after the coding. No distinction is made between e.g. "unhappy", "very unhappy", "completely unhappy" and "not too happy" . Any category which is the most unhappy one among all the *k* response categories of the scale is given the position "0" on the secondary [0, 10] scale, irrespective of the text of its label.

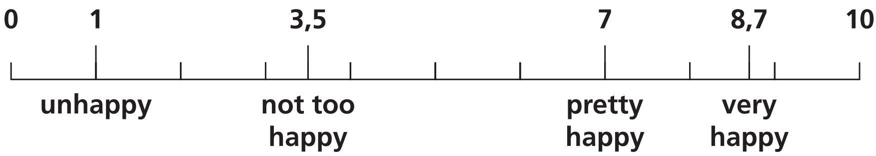
Two, from scientific point of view, it is naïve to consider the difference between ordinal and cardinal numbers as a mathematically interesting idiosyncrasy, but without any consequence for practical scientific research, at least in social sciences. Since this view is the basis for the equidistance assumption, the criticism concerns this assumption as well. If happiness is measured using a numerical scale, which is usually presented in a more or less pictorial way and includes 7-11 response categories, the equidistance assumption seems reasonably well acceptable. This also applies to the stretching of a scale [1,10] or even [1, 7] to [0, 10]. A type of scales, known as the "Best-Worst Ladder Scales", meets reasonably well all the underlying assumptions for direct rescaling.

Some justification is given by e.g. Van Praag (1991), Yew-Kwang Ng (1996, 1997), Van Praag & Ferrer-i-Carbonell (2004: 319). Direct stretching is a practice that is applied in the WDH to numerical scales with at least 7 scale points. This, however, does by no means justify the application of this method to verbal scales with, generally, only 3, 4 or 5, and incidentally 6 or 7 scale points and in which the equidistance is introduced by the coding procedure. Yet this approach is applied to 3- and 4-point scales, without any methodological scruples, in e.g. the US General Social Survey (GSS).

There is no basis on which to assume equidistance between the pair "not too happy"-"pretty happy" and the pair "pretty happy"-"very happy" on a [0, 10] scale; and if this equidistance existed, it should vanish if the label "very happy" is replaced with "extremely happy". Moreover, it is doubtful that the stretching of an 1-2-3 scale towards an 0-5-10 scale is as proportional as is assumed to be, that it will actually result in reliable estimates.

We shall refer to this rather naïve common practice for verbal items with relatively small values of *k* as the "traditional approach".

A possible alternative to the common practice – measuring happiness as a discrete variable using equidistant ratings ranging from 0 to 10 – might be the use of "verbal analogue scales". In this method, all the members of a panel are requested to place *k* marks on a line segment of about 10 cm, one mark for each of the possible response categories. They may be asked e.g. “Please place a mark on this line, at the position of which you feel is the most appropriate for the judgment ‘pretty happy’, irrespective of your personal happiness judgment”. One end (10) of the line represents the most happy conceivable situation of the respondent personally and the other end (0) the most unhappy conceivable one. Then for each category, the average position of all those given by the panel members is adopted as the transformed position of that category on the [0, 10] scale, giving e.g. the fictitious result:



Jones and Thurstone (1955) describe a method in which they presented 51 verbal qualifications to a panel of 905 respondents, who were requested to select the most appropriate appreciation rating on a 9-point Likert scale for each qualification separately. As a result, the 51 qualifications could be mapped on a common interval scale.

Veenhoven (1993) and 12 co-workers rated, on a metric [0, 10] scale, the degree of happiness denoted by verbal labels of 29 commonly used survey questions. Their ratings were based on the English version of these questions and made in the context of other response options. Some illustrative findings are presented in table 6.1. As one can see, the context makes a difference indeed. Yet these differences were deemed to be too small and the estimates too uncertain: the between experts standard deviation was 0,8 on an average. Therefore, the average value, i.e. averaged over the different items involved, was used for computing ‘transformed’ means. These averages are still in use in the World Database of Happiness and their use badly needs to be reconsidered. In WDH, this method is referred to as the ‘Thurstone transformation’ and the values as those listed in the right hand column of table 6.1 as “Thurstone values”. Although these Thurstone values have been established for one specific language (English), it is current practice in the WDH also to apply these to items in other languages.

Table 6.1 Average rating of 12 experts on 4 happiness intensity labels in different contexts

Label	Context			Average
	k=3	*k*=4	*k*=5	
"very happy"	9,2	9,3	9,4	9,3
"fairly happy"	6,5	na	6,9	6,7
"not very happy"	3,5	3,9	3,4	3,7
"not at all happy"	1,1	1,0	na	1,0

Summarized from: Veenhoven (1993: 110); "na" = not available.

A similar study was conducted by Bartram and Yelding (1973) among 166 adult London ITV watchers. A number of their qualifications overlapped the Thurstone values; the absolute differences of the numerical values range from 0,1 to 0,7, which is in good agreement with the inaccuracy of these numbers. However, it should be noted that the procedure according to Veenhoven was not completely identical to that of Jones and Thurstone, nor to that of Bartram and Yelding, since Veenhoven engaged 12 experts in his study vs. the 905 non-experts of Jones and Thurstone and the 166 of Bartram and Yelding.

6.3 The Happiness Scale Interval Study (HSIS)

6.3.1 The problem

The Thurstone method offers a solution to two of the problems inherent in the common practice when applied to verbal rating scale items as described in section 6.1: the equidistance problem and the problem of completely neglecting all the labels of the response categories in the further analysis. Nevertheless, the method has at least three week points:

(i) the Thurstone values are based on the judgements of experts and it is unclear to what extent their joint decisions reflect the views of non-experts in a sample from some nation.

(ii) the Thurstone values have been localized by Dutch experts on the basis of the English version of the items as they are collected in the WDH. The assumption is that translation into the local language of any nation will give rise to identical feelings associated with the same item. On an average, the meaning of "gelukkig" to a Dutchman is assumed to be exactly the same as that of "happy" to a US inhabitant or to any other English speaking individual, irrespective of his nation. But how valid is this assumption ? Its only justification is that it has not (yet) been investigated;

(iii) the Thurstone values do not take into account the phrasing of the lead question, nor the number and the labels of the alternative response categories and their position in the order of the total set of categories.

6.3.2 Principle of a solution

Veenhoven (2009), in order to counter these problems, has started his International "Happiness Scale Interval Study" (HSIS). The International Happiness Scale Interval Approach (further abbreviated HSIA) is designed to obtain better estimates of mean happiness and of happiness inequality and it differs in several aspects from the Thurstone approach. Rather than picking a few frequently used items, this study has the intention to cover all the questions that have ever been used in surveys of the general population in nations and used verbal response options. Currently, there are 117 such survey items, about half of which has been applied in more than one language. An essential difference between the HSIA and all previously applied methods, is that the response options to these questions are no longer linked to a single value on the happiness scale, instead each category is related to an interval of happiness values, and the aim is to estimate the boundaries between the various contiguous intervals.

These estimates are made within the context of the full question, including all other response options. This operation is performed for each of the selected language variants. Judgements are made by native speakers. As yet some 2000 judges have participated in this project and the intention is that in the end about 10.000 will have been involved. A last difference between the approach of Veenhoven and that of others is that a new technique is applied for establishing the boundaries between the category intervals (section 6.3.3).

6.3.3 Boundary localization device

The degree of happiness denoted by verbal response options is rated using a web based 'Scale Interval Recorder' developed by Veenhoven & Hermus (2005). A screenshot of this instrument is presented on Figure 6.1.

On the left half of the screen is a question on happiness that has been used in one or more national surveys, in this case a standard item in the

American General Social Survey. On the right half of the screen is a vertical bar with numerical values ranging from 10 at the top to 0 at the bottom, these extremes being labeled respectively as 'best possible' and 'worst possible'. On the bar are slides, which the judge can move up or down. In this case of a 3-step question, there are two slides which can be used to partition the bar into three intervals. To the right of these intervals are the verbal response options and these words move with the slides in such a way that they are always in the middle of the interval. At the start, the slides are in the middle of the bar. The task of the judge is to move the slides and partition the bar into three intervals that best correspond to the words on the right. In this case the label 'very happy' is judged to denote the interval between 10 and 8, its mid-interval value, abbreviated MIV being 9. The label 'pretty happy' is seen to apply to the interval between 8 and 6, with MIV = 7 and lastly the option ' not too happy' is seen to cover the interval between 6 and zero, with MIV=3.

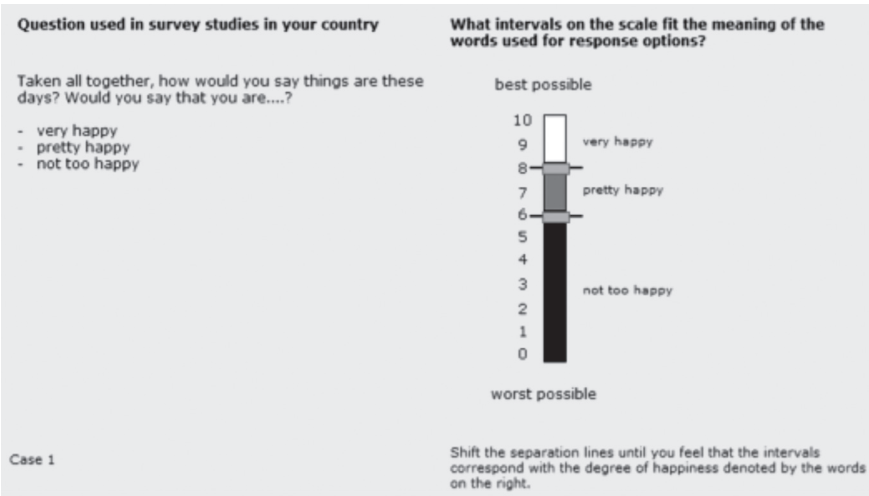


Figure 6.1 Happiness Scale Interval recorder

This task is performed by university students who are recruited by local co-investigators in different countries. The aim is to have each item judged by at least 200 native speakers per session, but as yet the number is mostly lower. Judges rate no more than ten different questions in a session, which takes about 15 - 20 minutes and is done behind a PC, independently of each other. Some of the co-investigators have rewarded their judges for this job, but most called for voluntary participation. Further details of this study are described in section 7.2.

6.3.4 The merits of this innovative approach

The HSIA does not pretend to solve all problems concerning measurement of happiness nor those of measuring life satisfaction etc., but it (cl)aims to do so especially the three problems mentioned in section 6.3.1 The initial objective of Veenhoven with this study can be described as providing a means to replace the Thurstone values with the MIV as a more valid alternative, leaving the method intact that was applied to estimate the happiness mean value of a nation and the corresponding within-nation standard deviation. This better validity concerns all three weaknesses.

However, in this chapter we shall demonstrate that, and how, the Happiness Scale Interval Approach (HSIA), albeit unintentionally, results in a method to solve a problem that until now generally was ignored, if recognized at all, but which is fundamental in the conversion of sample results to information on the happiness distribution in the population that is represented by this sample. This problem concerns the validity of the approach in which happiness is measured as a discrete variable in its relationship to happiness as a psychological concept. The respondent has to make a forced choice out of a limited number of alternatives. However, if we consider happiness as the intensity of something in a subject's personal situation, it becomes obvious to look for a continuous variable rather than to stick to a discrete one. If we managed to construct some variable that is related to happiness as measured above, and that is simultaneously continuous, this would improve the validity of the estimates of the parameters of the population happiness distribution, at least in this respect.

In our view, the most innovative element of the HSIA is that it facilitates a method to solve this problem and to bridge the gap between the measurement of happiness in a sample and the happiness distribution of happiness in the population. The link between the two is found in the cumulative distribution of both happiness distributions, a concept which is neglected all too frequently in applied Statistics. The HSIA is the first approach that allows the application of models in which the empirical cumulative frequency distribution can be compared in the cut points directly to the assumed distribution of a continuous happiness variable at the population level.

6.4 The model underlying the Happiness Scale Interval Approach

The model underlying the Happiness Scale Interval Study postulates the existence of a variable, here denoted *H*, that, in this application, expresses the intensity of the feelings of happiness of a respondent. In this description, we shall deal with the application of the model to the measurement of happiness, but it is equally applicable to the measurement of life satisfaction or some other related subjective self-judgment of the respondent's hedonic situation.

The following properties are assigned to this variable *H*:

I. *H* is postulated to be a variable, measured at the metric level of measurement and expressed as a real number in the closed interval [0, 10].

II. the value $H = 0$ represents the respondent's subjectively worst conceivable situation with respect to his happiness, whereas $H = 10$ represents the subjectively best conceivable situation. This choice excludes the possibility of any *H* -value outside the [0, 10] interval.

III. *H* is an intensity variable and is a strictly increasing continuous function of the happiness intensity as experienced by the respondent: if a person at the moment t_2 feels happier than at the moment t_1, then $h_2 > h_1$, where h_1 and h_2 are the *H*-values at t_1 and t_2 respectively.

IV. the variable *H* is a latent variable. It is unobservable as such, but can be mapped by the respondent onto a set of *k* different verbal, numerical or pictorial observable ordered categories (ratings) $\{R_j \mid j = 1(1)\ k\}$, *k* being a natural number, usually $k \leq 12$. The order of the categories is assumed to be unambiguous.

V. the interval [0, 10] can be partitioned into *k* contiguous (sub)intervals, each of which being defined as the subset of *H*-values that are mapped to the same image. All these (sub)intervals are right-hand closed half open intervals, except the closed interval including the value $H = 0$.

VI. the above mapping is monotonous, while the interval with the largest *H*-values is mapped as the happiest category R_k.

VII. the variable *H* is a random variable; within a population, it has a probability distribution: different individuals in that population will have a happiness which is represented by generally different *H*-values. In general, different populations will have different probability distributions of *H*. These are of the same type, but have different values of the parameters.

VIII. except for H=0 and H=10, the H-values of the interval boundaries are subjective, since the interpretation of the possible responses is also subjective. This applies in particular to verbal descriptions of the categories (the labels), which may have a strong cultural component. Not only the language/nation combination will influence their interpretation, but also conditions as social class, age etc; moreover the emotional value of terms may shift over time. Therefore, when linking H-values to response categories, especially the verbal ones, some degree of variability in the results is to be expected.

As an example, we consider the next situation

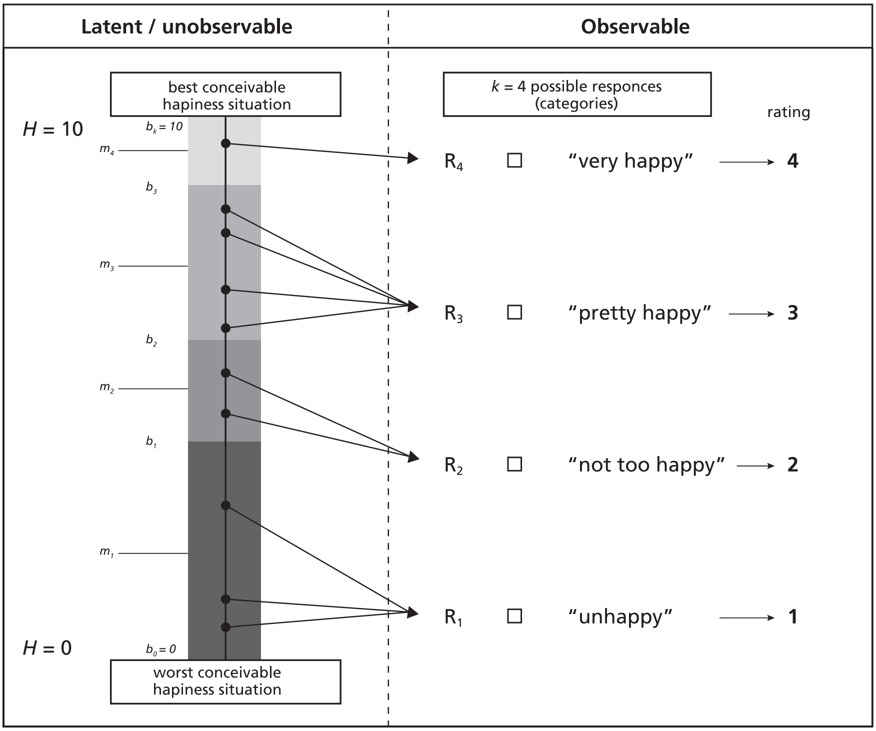


Fig. 6.2 Schematic representation of the model for the happiness measurement.

In this model, there is a one-to-one correspondence between each of the k presented different categories R_j and one of the intervals of {h}. The upper boundary of the j-th interval will be denoted b_j and this half-open interval $(b_{j-1}, b_j]$, with j =1(1)k, $b_o = 0$ and $b_k = 10$. For convenience reasons, the set

$\{(b_{j\text{-}1}\ ,\ b_j\ ]\ ;\ j=1(1)k\}$ is assumed to include also the closed interval $[b_o, b_1]$. The values $\{b_j\ ;\ j = 1(1)k\text{-}1\}$ are also referred to as 'cut points'; however, this term is usually extended to include also the values $b_o = 0$ and $b_k = 10$. We shall use the terms "boundary values" and "cut points" synonymously. In this way, there is also a one-to-one relation between each category R_j and the MIV of the j-th interval, which is defined as:

[6.1] $m_j := ½(b_{j\text{-}1} + b_j)$.

6.5 Further assumptions of the model

In an ideal world, there would be complete consensus about the *H*-values of all interval boundaries. Under VIII in the previous section, however, we have pointed out why individual opinions on the same boundary can be expected to differ. Each panel member is requested to report the value of *H* at which, in his personal opinion, a shift ought to be made towards a "more happy judgment category". If for a certain shift we plot the cumulative proportion of respondents that has made that shift as the ordinate against the corresponding *H*-values as abscissa, a monotonically increasing more or less S-shaped curve is obtained which often is called a "sigmoïd" (section 7.3.3, Fig 7.1).

In this procedure the basic assumption is that every respondent with $R = R_j$ will report this on the basis of his happiness feeling in the interval $(b_{j\text{-}1}\ ,\ b_j\ ]$. In the further process, the average boundary values are applied. In this way, the number of respondents is overestimated by those respondents that have a personal b_j -value above the average value and at the same time have a personal *H*-value between these two. In the same way, the frequency is underestimated by respondents with a relatively low personal b_j -value and an *H*-value in between the personal and the average panel value of b_j.

As long as the distribution of the personal b_j -values around their average value is symmetrical, at least approximately, these two effects may be expected to compensate to a large extent. A similar situation exists with respect to the lower boundary of the interval. The (a)symmetry of these distributions can be judged on the basis of the skewness of the distribution of the individual b_j-values in the panel about their average value (not to be confused with the skewness of the happiness distribution as a whole).

In the HSIS, two identical phrasings, but within different items, are judged separately and independently within each item. This practice was also not applied for the determination of the Thurstone values nor to similar other approaches. The proposed practice is justified in the comparison of the MIV of the judgment “very satisfied” within two different items of the WDH as an example. Item coded O-SLW/c/sq/v/5/p raises the question “All things together, how satisfied are you with your life as-a-whole these days ?” with five response categories: completely satisfied/very satisfied/satisfied/not very satisfied/not at all satisfied. Item O-SLS/c/sq/v/3/a asks “How satisfied are you with the way you are getting on now ?” with three response categories: very satisfied/all right/not at all. On a [0, 10] scale, the MIV of “very satisfied” for these different questions with different alternatives were 7,6 and 8,9 respectively, which demonstrates that the other categories and their phrasings should not be ignored.

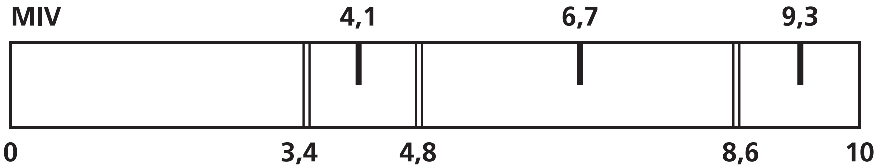
Intuitively, one might expect that the average result of all respondents in the determination of the Thurstone and related values, whether or not done by experts, is a good estimate for the MIV as defined in the HSIA. The answer to the question whether this expectation is correct is negative, at least in general. The reason is that the k MIV are not mutually independent. As will be demonstrated in Appendix F.3, the set of MIV for any item has to satisfy a condition, which for $k = 4$ and a [0, 10] scale can be written as

$$m_4 - m_3 + m_2 - m_1 = ½ \text{ x}10 = 5.$$

In general, after substitution of (the positions of) some set of four marks in this equation, the result will not be true and in this case these four average positions cannot be considered to be a set of unbiased estimates of the MIV. In the case of modest departures from this condition, some adjustment procedure of the marks position may be a ‘solution’ to deliver a more or less valid estimation of the MIV. In practice, however, it appears that it is rather exceptional when acceptable results are obtained along these lines.

Consequently, generally speaking, Thurstone values cannot be considered to be pseudo-MIV, since usually they do not satisfy our criterion that their ‘alternating sum’ equals the value 5. This is easily demonstrated for the scale example in section 2.1. The Thurstone values of the four responses in the WDH have been agreed to be {0,6 ; 4,1; 6,7; 9,3}.

Since 9,3 – 6,7 + 4,1 – 0,6 = 6,1 ≠ 5, the set of Thurstone values of this item clearly does not satisfy our MIV criterion, in this particular example, not even approximately. This can also be demonstrated in the graphical representation below.



Suppose that all Thurstone values are MIV, and that at least the largest three of them are correct. Then the boundary values are {0; 3,4; 4,8; 8,6; 10}, so the smallest Thurstone value in this case should be 1,7 and not 0,6.

6.6 Conversion of the sample data into information about the population happiness distribution

The happiness distribution of a community is defined as the probability distribution of the individual *H*-values within that community. This population probability distribution is unknown, but it can be estimated from the frequency distribution of the individual *H*-values in the sample from that population. The average value and the standard deviation can be estimated from the corresponding frequency distribution parameters of the *k* responses $\{R_j\}$ in the sample that is assumed to represent the community of the study.

If the variable *H* is assumed to be a random variable, it will have a cumulative distribution function, denoted as $G(h)$:= Probability $\{H \leq h\}$. $G(h)$ is a monotonically nondecreasing function of *h* with $G(-\infty) = 0$ and $G(\infty) = 1$. In the case where *H* is assumed to be a discrete random variable, $G(h)$ is a step function with *k* steps, one at each value *h* that *H* can adopt; the size of the *j*-th step is Prob$\{H = h_j\}$. If however *H* is assumed to be continuous, $G(h)$ is a continuous function. Then we define:

[6.2] $$g(h) \coloneqq \frac{dG(h)}{dh},$$

provided it exists; this derivative is called the probability density function (p.d.f.) of *H*. Whether or not $g(h)$ exists depends on the further assumptions made on $G(h)$.

The conversion of the sample results into information on the happiness distribution in the population can be made on the basis of different models. At least five such models are available, two of which have been mentioned already in the sections 6.1 and 6.2 respectively, and three different models, all based on the HSIA. The models III, IV and V are depicted in Fig. 6.3.

Model I: The traditional model (section 6.1).

The various categories are coded with integer numbers, which are declared to be cardinal numbers, followed by a direct stretching, corresponding to a discrete polytomous distribution model. The *k* observed frequencies are considered to be estimates of the corresponding values of the population parameters $\{\pi_j \mid \; j = 1(1)k, \quad \sum_k j = 1\}$, with:

[6.3] π_j:=Prob{H=10(j-1)/(k-1)} j=1(1)k

where the *k* possible happiness values {0, 10/(*k*-1), 2x10/(*k*-1), ...10} are error-free and equidistant by definition. The sample average happiness value is an unbiased estimator of the population mean happiness value and the sample standard deviation is an (almost) unbiased estimator of the population standard deviation, at least for sufficiently large samples.

When the item described in section 2.1 is used, the sample consists of four kinds of respondents: unhappy respondents, not-too-happy respondents, pretty happy and very pretty respondents. It is assumed that, this time, the nation consists four kinds of citizens: unhappy citizens, not-too-happy citizens, pretty happy citizens and very happy citizens in almost the same proportion as in the sample. In other words: the population happiness distribution is polytomous discrete and the number of species is chosen by the researcher. This model fits into the “traditional” method.

Model II The Thurstone model on the basis of Thurstone values (section 6.2)

This model is very similar to the model I. The only difference is that the sample happiness values: {0, 10/(*k*-1), 2x10/(*k*-1), ...10} are replaced with the appropriate Thurstone values, if available, and which are in general not equidistant.

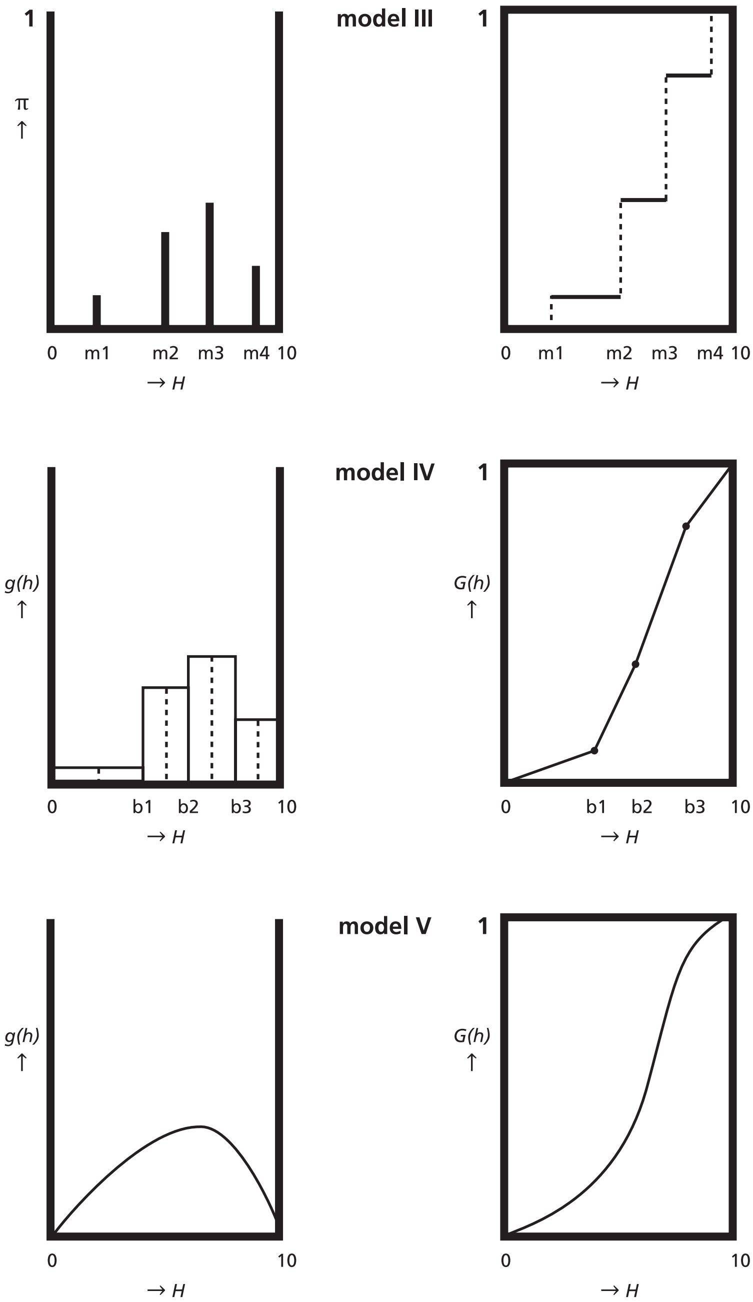


Fig 6.3 Probability values and densities (left) and cumulative probabilities (right) for h∈ [0, 10] in three models: III (discrete distribution), IV (semi-continuous distribution) and V (beta distribution), all on the basis of a four-point rating scale.

Model III: The polytomous Veenhoven model.

Under the model described in section 6.4 and 6.5 , it is assumed that each respondent with a happiness feeling corresponding to any *H*-value in the interval (b_{j-1} , b_j] will respond as R_j. All we can measure, however, is the number of respondents with R_j , but it is unknown which *H*-value in the interval (b_{j-1} , b_j] belongs to each of them. Therefore, we have to make assumptions on the unknown distribution of *H* over [0, 10], more precisely, over each of the *k* intervals ⊂ [0, 10].

One way-out could be to locate all respondents in the middle of the interval and to use the MIV as an estimate of the *H*-value of all of them. This is what the Veenhoven model does. So eventually, the model is equivalent to the previous one, the only difference being the replacement of Thurstone Values with MIV.

The sample average value obtained in this way is an unbiased estimator of the population mean happiness value, but the standard deviation is a biased estimator of the corresponding population statistic, since it ignores the measurement error in the estimated cut point position.

This approach yet considers happiness to be a discretely distributed variable, albeit with 'adjusted' ratings. The number of parameters of this model depends on the value of *k* and equals 2*k*-2, *k*-1 for the positions of the cut points and *k*-1 for the probabilities {p_j | j = 1(1)*k*; Σ *p* =1} where p_j := the estimated probability that an individual, 'selected' at random from the population, will report R_j.

In this model, the cumulative probability distribution *G*(*h*) := Prob{*H* ≤ *h*} is a step function with a step of size p_j at *H*= b_j, where at each step the value of *G*(*h*) is the higher one.

Model IV: the 'semi-continuous' model.

A second alternative is to assume that all *H*-values in an interval are equally likely, i.e. to assume a uniform distribution of *H* over each of the *k* intervals separately. In this case, consecutive points in the cumulative distribution plot with co-ordinates (b_{j-1}, $G(b_{j-1})$) and (b_j , $G(b_j)$) are connected by straight lines, making *G*(*h*) a broken line with kinks in all cut points where *H* = b_j. At these *H*-values, *G*(*h*) is not differentiable, so *g*(*h*) does not exist there.

Consequently, in this approach the density function *g*(*h*) is a step function with steps in all *H* = b_j and horizontal lines at different elevations in between. In other words: at each cut point, the probability density is changing stepwise to remain constant until the step at the next boundary.

As long as no explanation can be raised for such steps at a number of points, the number of which is selected by the investigator, such a model is not very satisfactory. A sufficiently realistic model should at least satisfy the condition that its p.d.f. is continuous over the complete interval (0, 10). Since this condition is not met here, although the random variable *H* is continuous, the model is referred to as 'semi-continuous'. Just like model III, model IV has $2k$-2 parameters. As long as no better alternative is available, we have to accept this model. The consequences of this assumption for the estimation of the population mean and variance are derived in Appendix F.1, including those for the precision of these estimators.

In model IV, the standard deviation deserves some special attention, since it is a multifunctional statistic. One of its functions is to act as an indicator of the happiness inequality, in this case of the within-nation inequality. Another function is to act as a measure of the precision of some other statistic, e.g. the estimated population mean, or even of its own precision; this second function is relevant for the construction of confidence intervals. The reason why we make this distinction is that the variance may be the sum of two or more components, as is the case in model IV. Here, the variance has three main components. One is introduced in the construction phase, because all estimated cut points are stochastic: they are average values of the opinions of a number of accidentally participating judges. The second and third variance components are both introduced in the application phase, the second being the usual variability between persons that select different response categories, and the third one for taking into account that there is also variability between respondents who choose in favour of the same category choice, but who do so on the basis of different *H*-values within the same interval (Appendix F, Eq. [F.10]). The latter component is excluded in the Veenhoven model. For the first role of the standard deviation, the quantification of the within-nation inequality, we only need the second and the third component, since the same case has been applied for all respondents , so the uncertainty with respect to the cut point position is identical for all sample members and hence can be ignored. However, for the 95 % confidence interval for the true, but unknown mean happiness value and for comparisons between nations, the first variance component should be included, so in its second function, we need a larger standard deviation than is appropriate in the context of its first function. In other words, one always has to make absolutely clear which standard deviation is referred to.

For the sake of this clarity, it is proposed to reserve the label "Veenhoven – Kalmijn" statistics for those that have been calculated on the basis of the semi-continuous model (IV). The Veenhoven – Kalmijn mean value can be computed directly from the MIV, weighed by the observed frequencies. A 95 % confidence interval for this population parameter is based on the uncertainties of both the construction and the application phase. The V–K standard deviation only includes the variability from the application phase and is a valid measure of the happiness inequality within the nation.

Model V: the beta distribution as continuous model

Because model IV is not satisfactory in all respects, there is at least one alternative to be considered. This is known as the beta distribution, which has a continuous density function in a closed interval with finite boundaries (see e.g. Kendall & Stuart, 1977; 35 and 46). As applied to our situation, it is defined by:

[6.4] $$dG(h) = [10 \cdot B(\alpha,\beta)]^{-1}\, h^{\alpha-1}(10-h)^{\beta-1}dh,$$

where B(α, β) is the complete beta function with parameters α and β ; this function is defined in Appendix H.2. This model of the beta distribution has only two parameters, α and β , which are positive real numbers; they are usually referred to as the two shape parameters of the distribution. This number of parameters is considerably smaller than in the models III and IV, because in this model, there are no categories at all in the population distribution. The density function $g(h)$ is continuous over the complete domain, finite and positive for all $h \in (0, 10)$ and zero outside the interval [0, 10]. All relevant properties and other information on the beta distribution have been collected in Appendix H. Most of this information can be found in various textbooks on Calculus and Statistics and/or in other public sources (e.g. Gupta & Nadarajah, 2004),

In applying this distribution as the model, the empirical frequency information, available as $\{F_j \mid j=1(1)\,k\}$, is compared to the corresponding values of $G(b_j)$, minimizing the differences between F and G jointly. The value of G is dependent on both α and β for all $\{b_j \mid j=1(1)k\text{-}1\}$. The comparison of F and G is possible and meaningful only at k-1 values of $H = b_j$ ($j = 1(1)k$-1), since the equations $F(0) = G(0) = 0$ and $F(10) = G(10) = 1$ are trivial. The situation can be considered to be one with a screen before the cumulative distribution function $G(h)$, which is observable only through one of the k-1 very narrow windows at

$H = b_j$ (j=1(1)k-1). The two model parameters $\{\alpha, \beta\}$ are to be estimated from these k-1 comparisons, leaving k-3 degrees of freedom (df).

For k = 3, there is always a unique solution with a perfect fit.

For k = 2, the number of solutions for this underdetermined situation is infinite.

For $k \geq 4$, we have an overdetermined situation and in general there will be no perfectly fitting distribution, so we have to look for the 'best fitting' solution.

If one has found this distribution, it would be possible to a apply a 'goodness-of-fit test' (see e.g. Cramér, 1974; 416-424). K. Pearson has proposed a test statistic for such situations, which is based on the multinomial distribution of N respondents over k possible responses and which is defined as

[6.5] $$\sum_{j=1}^{k} \frac{\left(n_j - \mathcal{E}n_j | H_o\right)^2}{\mathcal{E}n_j | H_o}$$

where $\mathcal{E}n_j|H_o$:= the expected value of n_j under the null hypothesis H_o that the estimated distribution is a perfect representation of the actual distribution in the population. Under this H_o and under some additional conditions, Pearson's statistic is approximately distributed as chi-square with in our case k-3 degrees of freedom (df). These conditions are that k >3, that N is not too small and that responses with conditional expectation $\mathcal{E}n_j|H_o \leq 5$ are 'pooled' with an adjacent response, which is obviously done at the cost of the number of df due to the effective reduction of k.

The application of such a test in other than comparative situations is debatable from the point of view of standard statistical test theory. Moreover, one has to be aware of the fact that H_o as it is defined above, can be specified as $F(h) = G(h)$ for all $h \in [0, 10]$. In this way, H_o is an element of a wider class H_o^+, which is the null-hypothesis that is actually tested above and which can be specified as $F(b_j) = G(b_j)$ for j =0(1)k, so for k+1 H-values only, two of which are trivial and k-1 are not. Rejection of H_o^+ implies rejection of H_o , but not the reverse. For all these reasons, we do not recommend the application of Pearson's goodness-of-fit test as long as we do not have a more valid alternative for the distribution of H.

To judge the goodness of fit, one might compute the value of $|F(b_j)\text{-}G(b_j)|$ for $j = 0(1)k$; its maximum value is known as Kolmogorov-Smirnov's statistic (Goodman; 1954). If $k > 3$, the numerical value of the statistic, ignoring its statistical significance, may be considered to be a suitable indicator for examining the fit of the beta distribution. So, for the distribution of *H*, the beta distribution seems to be the most valid, applying the standard beta distribution to the random variable H/10.

The two parameters of the beta distribution cannot be interpreted directly as a location and a dispersion parameter as is the case for e.g. the normal distribution. From the relationship between α, β, μ and σ^2 (Appendix H.5), the mean μ and the variance σ^2 of the distribution of *H* can be estimated by direct substitution of the estimates of the shape parameters α and β in:

[H.7] $$\mu = \frac{\alpha}{\alpha + \beta}$$

and

[H.9] $$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

In general, the values of the estimates obtained in this way will not be identical to those of the corresponding sample statistics. They may, however, be more valid, because they allow for the assumption of a continuous random variable *H* with a continuous p.d.f. over (0, 10).

The beta distribution also enables one to compute a potentially useful statistic which, just like the mean value, should be considered to be a location statistic for the happiness distribution within some society, but as a measure of inequality in a comparative study of nations, especially in relationship to other characteristics. It is the "percentage happy", which is defined in this context as the percentage of the society for which the happiness, expressed as the *H*-value, is closer to their most happy situation than to the most unhappy one, i.e. for which $H > 5$. In the above notation, this proposed statistic is defined as the estimate of $[1 - G(5)]\cdot 100\%$, and can be computed on the basis of the estimates of the parameters α and β. Since the value of this statistic is influenced by both the mean value and the variance of the distribution, it may be considered to be a possible alternative for 'Inequality-adjusted happiness' as has been announced already in section 5.5.

6.7 Application and merits of the model

The application of the method is a two-phase process. The first step is the scale construction phase, carried out by the judges in a panel as has been described by Veenhoven (2009). The second phase is application of the method to characterize the happiness of a population by a sample of subjects using this scale. Note that we use the term 'panel' for the scale construction phase and 'sample' for the application phase as a contribution to strengthen the distinction, and the separation, of these two phases.

In the HSIS, the judges in the construction phase have to identify their personal opinions with respect to the k-1 cut points $\{b_j \mid j = 1\ (1)\ k\text{-}1\}$, bearing in mind that $b_0 = 0$ and $b_k = 10$ are fixed. The values of the k-1 boundaries of a given item have to be estimated as the average values reported by n panel members separately and independently. Each of the judges has to specify the above mapping by indicating the b-values he feels to separate the consecutive categories, ignoring his personal happiness self-judgment.

In the second stage, the outcomes of the first stage are applied to the observed frequencies of the various categories as counted in a sample of N subjects from the relevant population. The central tendency and dispersion are assessed from these results, the former expressed in the mean and the latter in the standard deviation. These statistics are used to compute estimates of the parameters of the distribution of the variable H in the population represented by the study sample. Obviously both stages will contribute to the eventual inaccuracy of these estimates.

We have to emphasize that the application phase of the method described in this chapter is only applicable to samples for which the 'complete' empirical sample cumulative distribution $\{F_j \mid j=1(1)k\}$ is known, albeit for k happiness values only. Knowledge of the average value and the standard deviation of the sample happiness only is insufficient, at least for verbal scales.

The main possible merits of the above approach, some of which are potential, can be summarized as follows.

(a) Improvement of the validity of the method in that sense that the proposed approach considers happiness no longer as a discretely distributed variable, but allows for its continuous nature. In this way, the method described in this paper is no doubt closer to reality and is to be considered more relevant for social scientists than previously conventional methods were. Moreover, as compared to the method of direct rescaling, the criticism raised to the latter method does not apply to the results obtained according to the HSIA. This especially includes the objections against the controversial treating of ordinal ratings as if they were cardinal, since in the proposed approach equidistance between the ratings is no longer assumed.

(b) A consequence could also be an improvement of our correlational findings, at least in the validity perspective. Moreover, it is conceivable, at least theoretically, that this improvement of the validity of the happiness measurement may also result in higher numerical values of the association measures between happiness and conditions of happiness. Such an expectation would be based on the assumption that associations that are really present, may be blurred by the fact that happiness is measured in a suboptimal way rather than due to the fact that the associations are intrinsically insufficiently strong.

(c) Meta-analytical studies are almost always hampered by the problem that different findings that need to be combined arise from the application of different WDH items. It is to be expected that the results obtained according to the HSIA will be more reliable than those obtained according to previously current methods, so the method may seriously enlarge our meta-analytical opportunities. Similar considerations can be applied to the investigation of trends of happiness in nations or other societies.

(d) Finally, the method enables the opportunity to optimize the set of questions. Items with a relative large skipping rate, with a large interval width inequality and/or in which a relatively poor consensus about the positions of the boundaries has been observed within panels and/or between panels from different nations, are less suitable than those without these problems. All these observations could be good reasons to discontinue the application of such items, although still a number of studies will remain where they have been applied in the past. In this way, the present approach may contribute to the standardization and improving the quality of measuring happiness.

In chapter 7, we shall evaluate the application of this approach to a number of verbal scales and test to what extent the underlying assumptions and the model can be corroborated or not.

6.8 Alternative solutions for the problems encountered in this chapter

In the HSIA, the properties of the scale of measurement are measured externally. i.e. in a separate study. Other studies estimate these distances 'internally', i.e. within the context of the correlational studies on the basis of e.g. ordered probits or related statistical or econometrical techniques, which are mentioned here for the sake of completeness only. Their application is confined to the investigation of the association between happiness and one or more other condition variables at the metric level of measurement and such associations are outside the scope of this study.

A common element of these alternatives is that they circumvent the estimation of the happiness distribution as such, since the user is not interested in that distribution, but in the significance of the measure of association with the correlate only, avoiding in this way the problems inherent in common practice. Developments are moving towards a situation in which numerical scales are selected more frequently at the cost of verbal scales. Nevertheless, the WDH has an inheritance of thousands of studies based on the use of verbal items which are still of value for comparative studies and for meta-analysis.

We present two serious alternatives for examining the association between happiness and dichotomous correlates. These may be not always be the most efficient ones, but this may be compensated by the advantage that they are quite sound from a methodological point of view. In some situations, these methods may be an option for re-analysis of an existing data set.

First alternative: direct comparison along identical scales

In the case where the aim of a study is to compare two nations or to investigate the influence of some dichotomous correlate, e.g. male/female, employed/unemployed, urban/rural, on happiness, the usual approach is to run a comparative study and to evaluate the results using Student's two-sample test; in this way one considers only the central values to be relevant. This test requires that the positions of the *k* ratings are known, at least as estimates.

There is, however an alternative available, which should at least be considered, since it has some interesting properties, the most interesting being that it simply circumvents the cardinalization problem. This procedure requires the application of the same item and in the same language/culture at both levels of the correlate. Let this item be the one given in section 2.1. A suitable presentation of the results obtained at the ordinal level of measurement is as follows.

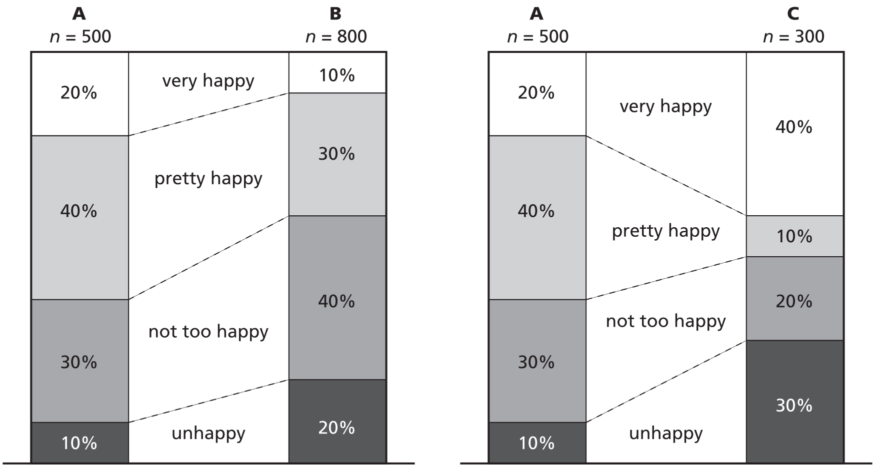


*Fig.6.4 Visual comparison of distributions with the same 4-point scale.
Left: A happier than B, right: C more inequality than A.*

This pictorial presentation in Fig 6.4 includes two sets of two ordered 'stack diagrams', which act as the linear equivalents of pie diagrams for variables at the nominal level of measurement; each of both is essentially a projection of the relevant cumulative frequency line onto the common probability axis. The percentages in the components are the relative shares of the relevant total frequency distribution, summing up to 100% in each bar. The method has at least two advantages. One, it is applicable to measurement at the ordinal level; no assumptions are to be made on the positions and on the mutual distances of the central values of the *k* categories. Moreover, no information is lost, since the complete distribution is mapped. The reader can judge the difference between the samples on the basis of e.g. the percentage "very happy", but also of the percentages "very happy" and "pretty happy" together. In case B all boundaries between the categories are below the corresponding ones in A, it is immediately clear that A is a happier society than B.

A test is available to those who feel urged to establish some statistical significance of the happiness difference. This test is known as the *Kolmogorov-Smirnov two-sample test*. If a *k*-point scale has been applied, there exist *k*-1 unknown, but common happiness values at which the difference of the cumulative frequencies $F_A - F_B$ is known from the observations. Let *D* be defined as the maximum absolute value of these *k*-1 differences; in the above diagram, it is the largest of all distances between corresponding boundaries in both bars, expressed in their common length as a unit. Then under the null hypothesis that both samples are random samples from populations with identical, albeit unknown distributions and for a two-sided application, the statistic $4D^2N_AN_B/(N_A+ N_B)$ has approximately a chi-square distribution with 2 df, provided both sample sizes N_A and N_B are not too small, say at least 40 each (Goodman, 1954). For smaller samples, the reader is referred to e.g. Siegel (1956, p. 127-136, 278-279). The power of the test is only slightly less than that of Student's t-test (Siegel, 1956, p.136).

The null hypothesis of the Kolmogorov-Smirnov test is that $\mathcal{E}F_A = \mathcal{E}F_B$, more precisely that F_A and F_B have equal expectations for k-1 non-trivial happiness values only, so the test is also sensitive to differences other than the central values of the two distributions as is demonstrated in Fig. 6.4 right. Let the happiness distribution of a third sample C be e.g. 30 - 20 – 10 – 40 (%). In this situation, a difference between A and C is found that is caused by internal inequality differences between A and C, which may be also an interesting finding. Such a difference may be significant, but it is neither significantly positive, nor significantly negative, since it concerns the distributions as a whole and not simply their central values only. Hence, for a correct judgment visual inspection of both distributions is always inevitable. It is not possible to express the magnitude of the difference in a number otherwise than as the value of Kolmogorov-Smirnov's D, either as a fraction or as a percentage, in our example as D = 0,20 or as D = 20% (absolute), which number does not necessarily characterize the difference between the central values only. Some researchers will consider this to be a less attractive aspect of this method.

Second alternative: Dichotomization of the happiness distributions.

The above situation can also be analyzed by reduction of the number of categories to k'= 2 for both distributions, which are required to be estimated on the same scale of measurement. In the above situation, the obvious combination seems to be "happy" := "very happy" + "pretty happy" and consequently "unhappy" := "not too happy" + "unhappy". In other situations the problem may recommend different optimal combinations.

The advantage of this approach is that is fully parametric. The percentage "happy" in A is P_A = 60 % and in B it is P_B = 40 %, so the difference is quantifiable. A 95% confidence interval for the true but unknown percentage difference $P_A – P_B$ can be obtained as:

[6.6] $$(P_A - P_B) \pm 2\sqrt{\frac{P_A(100 - P_A)}{N_A} + \frac{P_B(100 - P_B)}{N_B}},$$

where N_A and N_B are the sample sizes. If this confidence interval contains positive (negative) values only, the difference $P_A - P_B$ is positively (nega ti-vely) statistically significant at the 5 % level. Inserting the observations of Fig.2 and rounding to integer percents gives as a result the 95% confidence interval [15%, 25%], so this difference is significantly positive at the 5% level.

Just as in the previous approach, this one also circumvents the cardinality problem. Its main disadvantage is a rather small loss of information: the position of the boundary between "very happy" and "pretty happy" in both distributions is not taken into account and the same holds for the boundary between "not too happy" and "unhappy". An additional point is that the appropriate way to carry out the dichotomization has to be determined on the basis of the context and objective of the study and may be somewhat arbitrary unless this is sufficiently clear a priori. Nevertheless, for scales with small *k*, this method is to be considered to be a most serious alternative in case of a comparative studies with a dichotomous correlate, in particular if the primary scale has a small *k*-value.

✧ ✧ ✧

Chapter 7

The Scale Interval Approach II. Experiences with 100 cases.

ABBREVIATIONS USED IN THIS CHAPTER:

cpp	cut point position (section 7.2)
HSIA	Happiness Scale Interval Approach (section 7.1)
HSIS	Happiness Scale Interval Study (section 7.2)
MIV	Mid-Interval Value(s) (section 7.2)
MLE	Maximum Likelihood Estimator(s) (section 7.4)
WDH	World Database of Happiness (section 7.2)

Alpha-2 ISO nation codes have been listed in table 7.1 (section 7.2).

7.1 Introduction

In chapter 6, we have described the Happiness Scale Interval Approach (further abbreviated HSIA) as a method to obtain (more) valid and useful estimates of population mean happiness value and standard deviation, the latter being used as a measure of the within-population happiness inequality.

In the present chapter we analyze 100 of the first cases and consider our experiences with this approach. This will allow us to test whether or not the method works. What are the problems that arise with the application of the HSIA ? How can these problems be resolved? What further research is needed ?

It should be emphasized that this study is exploratory, thus there are no a priori formulated hypotheses to be tested. Moreover, the available data cannot be considered to be a random sample of any population. As a consequence, the reader should not expect any statistical significance statement to be made in this chapter.

Construction and Application Phase

What is essential to understand about this approach is that it is a two-step method and that a clear and consequent distinction between the two steps has to be maintained. The two steps are the scale construction phase and the application phase.

In the construction phase, the objective is to determine, as precisely as possible, the positions of the *k* -1 cut points between the *k* categories of the relevant scale. The judges are expected to express their opinion on the meaning of the category labels and their consequences for the positions of the cut points on the scale, They are required consciously to exclude any shading related to their own level of happiness.

In the subsequent application phase, these judgements are applied in the analysis of results obtained in general population surveys using the same item in the same language. Respondents in the sample are presented with the item question and categories only and not with the panel results from the construction phase. This time, the separate measurement of the happiness of all individual respondents results in a cumulative frequency happiness distribution, which has been obtained from a sample of the population to be studied. The distribution is processed on the basis of the results of the construction phase, and in this way information is produced on the happiness situation of that population.

Note that in this chapter, we reserved the term “panel” for the set of “judges” involved in the construction phase, whereas “sample” and “respondents” always refer to participants in the application phase.

Plan of this chapter

We will describe in section 7.2 which data has been collected from a number of sessions and which calculations have been performed on that data. This section deals with the construction phase only. The main findings of this phase are collected in section 7.3.

The contents of section 7.4 concerns some examples of application of the HSIA to existing and already reported survey observations. Application of the method requires that one knows the complete observed cumulative happiness distribution of the sample representing the nation in question. As a rule, this basic material is not reported in detail at the level we need for this study and this has limited our choice of data sets. We applied the method to a series of 20 studies measuring the happiness of the Dutch population over the period 1980 – 2008 and which meet all our requirements. The latter also implies that items have been applied that were included in the construction phase as described in section 7.2. The main findings of the sections 7.3 and 7.4 will be discussed in section 7.5,. Conclusions and recommendations, including those for further research on the basis of this study, are presented in section 7.6.

7.2 Available data

7.2.1 Nomenclature

Following the definition given in section 2.2 , we define, in this specific context, an "item" as a combination of one question and a specified number (*k*) of mutually exclusive response categories, all with a completely defined content. In principle, items with verbal responses in different languages will be considered to be manifestations of the same item if their translation in (US) English is identical. The further terms will be explained by applying them to a specific sample.

In the WDH-survey of this project, the term "study" is used to describe a session for which a number of judges have been invited to participate; we will use the terms "study" and "session" as synonymous in this context. For the study coded "spanish5", a number of participants were recruited by the University of Monterrey (Mexico); eventually 57 of them were registered as judges. During this session, 5 different items , all in the local language (i.c. Spanish), were presented subsequently to the judges. The first item (O-SLL/u/sq/v/4/b[1], with 4 categories, so with 3 cut points), was judged by 54 judges, while for the other items, the numbers of judges were 52, 51, 51 and 50 respectively. Each of them delivered a judgement, being a set of (his opinion on) the three cut point positions, further abbreviated "cpp", of the same item by the same judge. All these cut points are reported as numbers on the closed interval [0, 10] and rounded on 0,1.

Not all judgements appeared to be usable. For the above first item, 4 judgements were skipped afterwards. The remaining set of 54 – 4 = 50 judgements includes 50x3 =150 cpp and is called a case: so a is case a set of accepted judgements, including all cpp of the same item as delivered by the judges in the same session. This set of judges who have delivered an accepted judgement on this case is referred to as a "panel"; the panel size of the first case was 50 and for the other cases in this study, the panel sizes were 50, 49, 48 and 49 respectively.

A session included up to ten cases and its total duration did not exceed 20 minutes. No information is available on the additional time spent on instruction of the judges before a session.

[1] *Full detail about this item is available at: http://worlddatabaseofhappiness.eur.nl/ha-_cor/desc_hind.php?/ind=597*

7.2.2 Available data

This chapter deals with the first 100 cases obtained in the HSIS. In all these cases the items include verbal response categories only with *k* = 3, 4, 5 or 7. The basic observations have been contributed by 12 institutes listed in Table 7.1. As can be seen in Table 7.2, the total data set covers 14 sessions.

Table 7.1. Participating institutes

NATION *(Alpha-2 Nation codes according to ISO 3166-1)*		INSTITUTE(S)	CASES
CL	Chile	Universidad Catolica del Norte, Antofagasta	17
DE	Germany	University of Koblenz	4
HR	Croatia	VO PILAR Institute Social Studies, Zagreb	2
HU	Hungary	Eötvös Logrand University, Budapest	7
JP	Japan	Kobe University	4
MX	Mexico	University of Monterrey	5
NL	The Netherlands	Erasmus University Rotterdam	15
RO	Romania	University of Oradea	8
RU	Russian Federation	High School of Economics St. Petersburg Branch	10
US	USA	University of Michigan, University of Notre Dame and Texas A & M University	28

Table 7.2. Distribution of the 100 cases over the numbers of categories per item

Nation	Language/code		#	categories per items (*k*)			
	ISO 639-1	WDH listing	cases	3	4	5	7
HR	hr	Croatian1	2	-	2	-	-
NL	nl	Dutch3 and 4	15	5	3	6	1
US	en	English2 and 4	28	10	11	6	1
DE	ge	German1	4	1	-	2	1
HU	hu	Hungarian2	7	1	5	-	1
JP	ja	Japanese1	4	-	3	1	-
RO	ro	Romanian1	8	-	4	3	1
RU	ru	Russian1	10	-	7	2	1
CL	es	Spanish2, 3 and 4	17	4	10	1	2
MX	es	Spanish5	5	1	3	1	-
TOTAL			100	22	48	22	8

The listing of the sessions in the WDH survey of this project is made in terms of language, so Dutch3 and Dutch4 are two different sessions with different groups of Dutch judges.

The total number of different items involved in this study is substantially smaller than that of the cases, because 22 items were presented in more than one session, usually in different nations. One of the 100 cases was rejected according to the skipping rules as defined in section 7.2.3. This case has been ignored in most of our further analyses. The distribution of all 52 items over the remaining 99 cases is as follows:

# identical items	1	2	3	4	5	6	7	8	9	sum
# cases of identical items	30	12	5	1	2	-	1	-	1	52

This survey expresses that the study in this chapter includes e.g. 12 sets of 2 cases each, where both cases cover the same item. One item (WDH coded O-SLL/u/sq/v/4/b) was included in 9 different panel sessions in 6 different nations. Of these 52 items, 23 asked for the degree of happiness and 28 for the degree of satisfaction with life in a narrower sense; in one item, the focus was on mood.

7.2.3 Screening of panel judgements and of cases

Inspection of the data within a case at the level of individual judges revealed that some judges had clearly misunderstood the instructions, whether intentionally or not. As an example: incidentally a judge had placed all slides close to each other at either the upper or the lower boundary, in this way leaving almost the full scale to one terminal category. Obviously such a contribution of this judge has to be ignored.

In order to skip such judgements in an objective way, we devised a set of skipping rules. The choices of these rules are inevitably debatable as were any other one, but their existence guarantees the exclusion of arbitrary application and subjectivity. Our skipping rules for judgements are:

(i) a judgement is skipped if two or more adjacent categories each have a width that is less than 110 % of the minimum width that is technically possible;

(ii) additionally a judgement is skipped if one of the categories covers
- 85 % or more of the total scale in case of a 3-point scale
- 80 % or more of the total scale in case of a 4-point scale

- 70 % or more of the total scale in case of a 5-point scale
- 60 % or more of the total scale in case of a 6-point scale
- 50 % or more of the total scale in case of a scale of more than 6 points.

For the case as a whole, the rules are:
(iii) a case result is rejected if more than one third of all judgements is skipped and/or the number of non-rejected judgements is less than 15
(iv) a case result is declared doubtful if more than 20 % of all judgements is skipped, but not more than one third.

The expression "technically possible" concerns the fact that the slides have a certain extension; the recorded value is that of the middle of the slide. Therefore, even if two slides are positioned exactly close to each other, a width of the interval between the two is recorded that corresponds to some distance, which in our equipment amounts up to 0,5 unit on the [0, 10] scale. However, at both terminal categories only one slide is involved and recording is established in such a way that a slide in close contact with the end point of the scale results in a boundary value of either 0,0 or 10,0. The 110 % is selected in favour of a tolerance of 10 %.

A judge may want to express that in his opinion a category has a zero width and therefore should be represented on the scale by a single 'point value' rather than by an interval; he can establish this by positioning the two slides close to each other or, for a terminal category, one slide close to the relevant end point of the scale. If e.g. a respondent with a good sense for logics assigns only the value 10,0 to the interval for "completely happy" since in his view any *H*-value <10,0 is by definition not completely, then this can be judged only as fully correct. His judgement should not be ignored, also because similar respondents are to be expected in the sample during the application phase. A similar reasoning is to be applied to a judgement, in which two slides are located close to another in about the middle of the scale for the interval "neither satisfied, nor dissatisfied".

Therefore, the above skipping rules accept the possibility of a choice in favour of one or more zero-width category intervals, but not unlimited. So the rules skip judgements in which two or even more adjacent zero-width intervals are reported, because this would imply that these labels are all considered to have an identical meaning. Skipping rule (ii) prevents the situation that too many intervals are simultaneously relatively small. Note that we do not skip judges, only judgements.

In order to detect some of the possible causes for rejection of observations, the judges were usually asked to report additionally:

- whether or not the language of the item as presented was their mother tongue,
- how seriously they had performed the task (1-4, 1 = very seriously; 4 = not at all seriously), and/or
- how difficult they had experienced the task to be (1-4; 1 = very difficult; 4 = not at all difficult).

Not all the judges replied to these questions and in most cases not all three questions were presented to the panel members.

7.2.4 Calculations on the basis of the observed data

For each case and for each of the *k* –1 boundaries between categories within that case, the *n* observed values, *n* being the number of accepted judgements, can be considered to be elements of a statistical distribution. The following statistics have been computed for each of these distributions:

(a) the average value
(b) the median value
(c) the variance and the standard deviation with *n* –1 degrees of freedom
(d) the skewness, expressed as $\propto_3 := m_3/m_2^{3/2}$, where m_i := *i*-th central sample moment
(e) the kurtosis, expressed as $\propto_4 := m_4/m_2^2 - 3$
(f) the number of judges that assigned a zero-width to the first category
(g) the same for the *k*–th category

Moreover we computed from the observations of all cut points together

(h) the correlation coefficients between all different cut points, each on the basis of the *n* pairs of individual observed positions

For each case, we computed at the individual level

(i) all *k* mid-interval values (MIV). The MIV of a category is defined as the average value of the *H*-values of the cut points at both ends of that category interval.

Within each case and for each category separately, the *n* MIV of that category were considered to be random variables with a statistical distribution in the same way as was done for the cut points, so the same statistics can be computed for each of these MIV distributions, sc.

(j) the average value
(k) the median value
(l) the variance and the standard deviation with $n-1$ degrees of freedom
(m) the skewness
(n) the kurtosis
(o) the number of judges that assigned a zero-width to the first category
(p) the same for the k–th category.

Finally, from the observations of all cut points together, we computed

(q) the correlation coefficients between all different MIV, each on the basis of the n pairs of individual calculated MIV.

7.3 Main findings concerning the construction phase

7.3.1 Zero-width terminal categories

For some terminal categories, the (English translation of the) description may invite a judge to assign only the scale end value to that category, resulting in a zero-width. In the present data set, such descriptions include: “completely (dis)satisfied/satisfying”, “extremely happy”, “extremely dissatisfied”, “fully (dis)satisfied”, “completely (un)happy”, “not at all happy/satisfied”, “totally not happy”, “all the time” and “none of the time”. Labels with “extraordinary” were also included.

Such descriptions apply to six of the seven accepted cases with k =7, and a considerable part of the panel made a choice in favour of a zero-width; for the various cases the percentages varied between 0 and 64 % with an average value of 17 % against 2 % for the situation in which not one of the above labels was presented. The results of the German case in particular had a very large contribution to this phenomenon, with 36 % at the upper and even 64 % at the lower end of the scale. The corresponding labels of this case were “völlig glücklich” and “völlig unglücklich” respectively, with “völlig” as a commonly usued translation of “completely”.

The 92 cases with k = 3, 4 and 5 were considered jointly; two of these were ignored at the unhappy end of the scale in this context, because the terminal category was a combination of two original categories, “not very happy/not at all happy” and “not too happy/not happy at all”. Our findings on this are summarized in table 7.3.

Table 7.3 Average reported zero-width terminal categories for k =3, 4 or 5.

scale end	suggested	# cases	% reported
upper (*H*=10)	yes	12	10
	no	80	8
lower (*H*=0)	yes	46	16
	no	44	12

KEY: *In 46 out of 46+44=90 cases, the description of the category at the lower end of the scale may suggest a zero-width as an option. In 16 % of all these situations, a judge actually reported this width (averaged over all judges and over all 46 cases), against 84 % in which non-zero-widths have been reported.*

From table 7.3 it follows that, contrary to our expectations, the above descriptions barely stimulate a choice in favour of a zero-width of a terminal category in the case of verbal scales with k = 3, 4 or 5, at least in general. However, the data in table 7.3 include the accepted judgements only. Among the rejected judgements, visual inspection suggests at first glance considerably more zero-widths.

To verify this, we also considered the rejected the judgements of 24 cases, 6 of each "row" in table 7.3 and selected at random. The results are given in table 7.4. In the 6 cases in the first row, the label of the highest category suggested to report a zero width. These 6 cases included 19 rejected judgements in total, in 3 of which the judges reported a zero-width upper interval (16 %) against 11 % of the accepted judgements. Since the latter percentage is based on a sample of size 6 only, it is not necessarily equal to the corresponding value (10 %) in table 7.3. A similar inspection was carried out with the 3x6 cases of the other rows. The difference between the accepted and the rejected judgements with respect to the percentage of zero-width among them is shown in table 7.4.

Table 7.4 . Average percentage reported zero-width terminal categories among accepted and rejected judgements in a sample of 24 cases.

scale end	suggested ?	# cases	% zero-width reported accepted	rejected
upper (*H* =10)	yes	6	11	16
	no	6	15	84
lower (*H* =0)	yes	6	25	76
	no	6	9	65

The percentages of zero-width responses among the rejected judgements turn out to be much larger than the corresponding values among the accepted judgements. In particular at the lower scale end, in the vast majority of all rejected judgements, a zero width for the most unhappy category was registered. The skipping rules in section 7.2.3 do not per se give rise to rejection of a judgement on the basis of a zero-width interval, but if an adjacent interval is also given the technically minimum width, which is 0,5 on a [0, 10] scale, this combination was a sufficient reason to reject this judgement. Inspection of rejected judgements suggested a frequent occurrence of this combination, in particular at the lower end of the scale. All six cases of the third row in table 7.4 together included 37 rejected judgements in total, in 21 (57%) of which this combination at the lower scale end was found. The same combination, also at the lower scale end, was observed in even 43 of 66 (65%) rejected judgements in one of the six cases in the second row. This phenomenon deserves more attention.

The zero-width in the terminal intervals may responsible, at least in part, for the finding that, on an average, these intervals are shorter than the other intervals. When the intervals are ordered from low to high happiness values, the average interval lengths for the 3-point scales are 3,2 – 4,4 – 2,4 respectively. For the 4-point scales, these average lengths are 1,7 – 3,0 – 3,5 – 1,8 and for the 5-point scales 1,7 – 2,1 – 2,2 – 2,6 – 1,4. For all three *k*-values, the interval of the second highest happiness is the widest on an average. This is not the case for *k* =7, where the average lengths are 0,6 – 1,6 – 1,7 – 1,7 – 2,0 – 1,8 – 0,7. The pattern, however, is the same. Only a limited value may be assigned to these findings, since the items have been selected and by no means at random; moreover several items were identical for different cases, i.e. in different nations/languages, so they are dependent to some extent.

This applies in particular to the results of *k* =7, since there are 8 cases for 3 items only.

7.3.2 Screening and skipping of judgements

The result of the screening was that only one of the cases had to be discarded and that seven of the remaining 99 cases were qualified as doubtful. However, in the specific context of this chapter, we have not ignored these eight cases, since the aim of the present study was to obtain information about what happens if this method is applied rather than to produce "revised scale information" about a small number of items. The 1+7 cases were submitted by the participants in Chile, Germany and Hungary. In all of them, we first considered the response to the additional questions.

The first one asked whether the response question as presented was in the same language as the mother tongue of the judge. In our 8 cases the percentage of negative answers to this question varied between 0 and 6 among the skipped judgements against 1 to 4 for the accepted ones. The second additional question was how seriously has the judges accomplished their task. The percentage that said they had had done so either very or quite seriously ranged from 85 to 96 among the accepted judgements against 93 to 100 for the skipped ones, except for two Hungarian cases with 79 and 81 %, but in which 10 and 11 % of the judges respectively had given no reply at all. In the three Chile cases, the judges were also asked how difficult the task was experienced. The percentages of those who judged the task either as not very or as not at all difficult ranged from 84 to 85 for the accepted against 86 – 89 for the skipped of judgements in a panel of size 33. The above observations do not reveal any clear difference between the accepted and the skipped judgements with respect to one or more the above two or three areas.

Most sessions included 5 – 10 cases, in principle all with the same judges. We also considered the possibility that during such a session some tiredness (in 15-20 minutes) and/or out of sheer cussedness developed, resulting in a gradually increasing percentage of skipped judgements within such a session. However, again no clear pattern emerged to support this hypothesis.

A more detailed analysis of the skipped judgements of the eight doubtful/rejectable cases possibly could bring more clarity on the question of why judgements are skipped and the consequences of this for the case as a whole. In three cases with three-point scales, in total 25 out of in total 99 judgements, were skipped. In all cases, the reason was that the middle category covered more than 85% of the scale, leaving less than 15% for the other two; its label was either “fairly happy” or “pretty happy”. The category “very happy” was given a zero width by 4 judges and 11 assigned it to the least happy category, whereas the other 10 judges did so to both terminal categories, leaving the whole scale to the middle category.

Only one case with a four-point scale was declared doubtful with 27 skipped judgements out of 116 in total. Not less than 20 judges declared both the adjacent categories “not at all happy” and “not very happy” to be zero-width categories. At the other end of the scale only one judge did the same to both “very happy” and “quite happy”. Of the 6 other skipped judgements one category covered more than 80 % of the total scale.

Two cases with five-point scales were declared doubtful. Both cases had been submitted by the same German panel with size 200. The item in the first case (O-HL/c/sq/v/5/n) was designed in a less usual way. The question is “When you consider your life-as-a-whole now, would you say you are ...?” and the ordered response categories are “fairly unhappy”/”rather unhappy”/”rather happy”/ “fairly happy”/”very happy”. Of the 59 skipped judgements in this case, 32 judges had assigned a zero width to both adjacent categories “fairly unhappy” and “rather unhappy”, in German translation “ziemlich unglücklich” and “eher unglücklich”. Of these 32 judges, 14 also gave a zero width to the adjacent category “rather happy”. However, at the upper end of the scale none of the judges gave a zero width to both adjacent categories “very happy” and “fairly happy”. Four judges gave a zero width to four categories, leaving all the scale to “fairly happy”. A similar result was obtained with the other doubtful five-point scale case, although this item includes a perfectly symmetric bipolar response scale. In total 43 of the 66 skipped judgements included both categories at the lower end of the scale as zero-width, whereas 16 of them even included the neutral middle category “neither happy nor unhappy”. A similar behaviour at the upper end of the scale was demonstrated by 6 and 3 judges respectively. Six judges left the total scale to one category only, albeit not always the same one.

The two problematic cases with a seven-point scale concern the same item (O-HL/g/sq/v/7/a), also with a perfectly symmetric bipolar response scale. One case was delivered by the German panel, mentioned under the five-point scales, this time with 61 skipped judgements out of 200. Zero width to the two (three) categories at the lower scale end was given by 48 (32) judges. At the upper end of the scale the corresponding numbers are 7 (3) respectively. Three judges reduced the scale to one category only.

A similar result was obtained from the other (Hungarian) panel with 42 skipped judgements out of 116. Zero width to the two (three) categories at the lower scale end has been given by 32 (19) judges. At the upper end of the scale these numbers are 4 (2) respectively. Only one judge reduced the scale to one category: "completely unhappy". His own ?

A most remarkable observation is that one of the other cases with a seven-point scale item had no skipped judgements at all and even no judge that assigns a zero width to any of the terminal categories. From our Russian local contact, we learned that the judges had been told that they were completely free to locate the slides in the positions they felt were adequate to indicate the change of the level of happiness. The Russian judges were instructed using a demonstration with a scale without a zero-width category as an example, but they were not told explicitly that each category should be given a non-zero width. This finding may underline the importance of adequate and sufficiently detailed instruction as the start of the judgement session.

7.3.3 Estimated cut point positions and their standard deviations

In the ideal world, there would be a complete consensus among all members of a panel on the happiness value at which one switches from e.g. "pretty happy" to the next category "very happy". In reality however, different panel members report different *H*-values for the same cut point. If one plots against *H* the percentage of the panel members that have made the switch, a S- shaped curve is obtained(Fig.7.1).

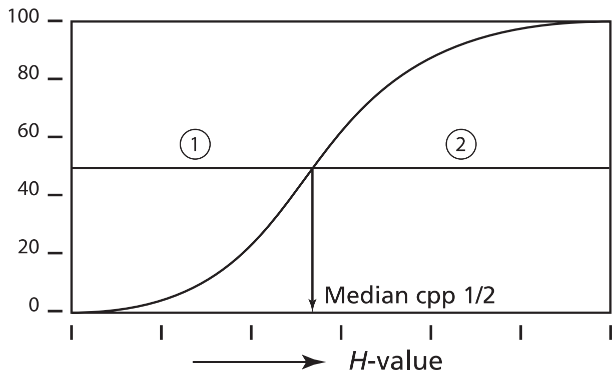


Fig. 7.1 Percentage of panel 'switching' from rating 1 to rating 2

If the individual switch points are normally distributed about some value, the curve is a cumulative normal distribution function, where both the average and the median value coincide with the *H*-value at which the ordinate has the value 50%.

This *H*-value can be estimated then as the average value of that cut point, but also by using the probit method. Whether serious departures from normality are present or not can be investigated in various ways. Application of normal plots appeared to be an insensitive method. Since the variable is bounded at two sides, a skewness towards the middle of the scale could be expected and this was found. The distributions of the observed cut points of the terminal categories were clearly skewed: at the lower end of the scale the skewness was positive in 82 cases against negative for 18. At the upper side of the scale, these numbers were 11 and 89 respectively. This makes the application of the probit method less appropriate.

On an average, the difference between the median and the average value did not exceed the value 0,2 absolutely, except the value -0,21 for the 'lowest' cut point of $k = 4$. In view of the standard deviations of the distributions within the panel, we decided to select the average cut point as the estimate of the cut point value. The ranges of the observed cut points are listed in table 7.5.

Table 7.5. Range of observed cut points for k =3,4,5,7 on a [0, 10] scale.

k	#	1/2	2/3	3/4	4/5	5/6	6/7
3	22	2,0-4,6	6,5-8,2	-	-	-	-
4	48	0,7-3,0	3,9-6,1	7,6-8,8	-	-	-
5	22	0,8-3,8	2,6-5,5	4,9-7,2	8,0-9,1	-	-
7	8	0,3-1,1	1,7-2,7	3,3-4,5	5,1-6,0	7,1-7,8	8,9-9,4

KEY: *for the 48 (#) cases of a 4-pointsscale (k =4), the boundary between the upper category (4) and the adjacent one (3), denoted 3/4, ranges from 7,6 to 8,8.*

From the cut points, the MIV can be computed directly. The ranges of these MIV are summarized in table 7.6 in the same way as was done in table 7.5 for the cut points.

Table 7.6 Range of MIV as calculated from cut points for k =3,4,5,7 on a [0, 10] scale.

k	#	1	2	3	4	5	6	7
3	22	0,8-2,3	4,8-6,2	8,3-9,1	-	-	-	-
4	48	0,3-1,5	2,3-4,6	5,8-7,4	8,8-9,4	-	-	-
5	22	0,4-1,9	1,7-4,7	3,9-6,4	6,6-8,2	9,0-9,6	-	-
7	8	0,2-0,6	1,1-1,8	2,6-3,6	4,4-5,2	6,1-6,8	8,0-8,6	9,5-9,7

The variability between the panel members within a case is expressed in the standard deviation; the distribution of the standard deviations is given in table 7.7. Their average values of all cases with the same *k* –value for each cut point separately have been tabulated in table 7.8.

Table 7.7. Frequency distribution of the within cases standard deviation of observed cut points for k =3,4,5,7 on a [0, 10] scale.

k	#	$s \le 0,5$	$0,5 < s \le 1,0$	$1,0 < s \le 1,5$	$s > 1,5$
3	22x2	-	2	19	23
4	48x3	-	36	65	43
5	22x4	3	39	42	4
7	8x6	4	34	10	-
SUM	324	7 (2%)	111 (34%)	136 (42%)	70 (22%)

KEY: *The 22 cases with k =3 together have 22x2 = 44 cut points. Of the 44 corresponding standard deviations, 23 exceed the value 1,5.*

Table 7.8. Average standard deviations between observed cut points for k =3,4,5,7 on a [0, 10] scale.

k	#	1/2	2/3	3/4	4/5	5/6	6/7
3	22	1,6	1,4	-	-	-	-
4	48	1,2	1,4	1,2	-	-	-
5	22	1,1	1,1	1,0	0,9	-	-
7	7	0,7	0,9	0,9	0,9	0,9	0,7

The general pattern is that the standard deviation has a tendency to decrease with an increasing number of categories and to increase with the distance to the nearest scale end, which is not surprising. These standard deviations are a measure of the (lack of) consensus among the panel members about the location of relevant cut point. Division by the square root of the panel size results in the standard error of the mean cut point value as obtained from this study.

7.3.4 Correlation coefficients

It is obvious to expect that estimates of different cut points within the same item are positively correlated, at least adjacent ones. If for example, a judge locates the boundary between "not too happy" and "pretty happy" at a relatively large *H*-value, he reduces the scope for the boundary between "pretty happy" and "very happy" by removing in particular relatively small *H*-values, so a relatively large *H*-value for the latter cut point is to be expected. In a similar way, the scope for the boundary between "unhappy" and "not too happy" is enlarged in this case, allowing relatively large *H*-values in particular. Hence an upward tendency is to be expected at both sides, resulting in a positive correlation coefficient.

The observations confirm this expectation. Of the 403 non-trivial observed correlation coefficients between the various cut points, 354 were positive and 49 negative. If we confine ourselves to the 219 correlation coefficients between the upper and the lower cut points of the same category, 218 of these were found to be positive, and only one was negative (-0,08). Obviously these positive correlation coefficients have an unfavourable effect on the precision of the estimated MIV. The MIV are expected to be predominantly positively correlated as is demonstrated in Appendix F, section F.4. This prediction is also confirmed by our observations. Of the 721 non-trivial correlation coefficients between MIV-estimates, 659 are positive and 62 are negative. The 318 correlation coefficients between the MIV of adjacent categories are even all positive.

7.3.5 The same item in different nations/languages

Ideally for each item, the cut points should be identical in all nations, even after translation of the labels into the relevant language. To what extent this corresponds to reality can be found by judging the scales of the same item in different nations. Our data set includes five items with four or more cases and these enable a good opportunity for comparison. The results are depicted on pages 158 and 159, both for the MIV and the cut points. The rejected case is not included ; the two doubtful cases are marked with [?].

The ranges of the MIV are collected in table 7.9, one column for each item. When judging the findings in this table, it should be borne in mind that, for theoretical reasons, the range increases systematically with an increasing number of cases.

Table 7.9 Ranges of MIV of cases of the same item (5 items)

# cases	9	7	4	5	5
category	k = 4	4	4	5	7
7	-	-	-	-	0,1
6	-	-	-	-	0,5
5	-	-	-	0,4	0,7
4	0,6	0,4	0,5	0,9	0,9
3	1,4	0,7	0,9	1,1	1,0
2	1,3	1,4	1,6	1,2	0,6
1(low)	0,7	0,9	1,0	0,6	0,1

KEY: *Of the five cases with* ***k*** *= 7, the largest MIV of the middle category (nr 4) is 5,1 and smallest one is 4,2, so the tabulated range of the five MIV is 5,1 – 4,2 = 0,9*

Inspection of table 7.9 and of the survey on pages 158 and 159 gives rise to the following observations:

(a) Within one item, the corresponding MIV of the different nations are spread over a range of ≥ 1 *H*-unit in 8 out of 24 ranges;
(b) On an average, the ranges of the 7-point scales are smaller than the ranges of the other four items;
(c) The ranges increase towards the middle of the scale;
(d) Ranges within a scale are not symmetric with respect to the middle: categories in the lower part of the scale have larger ranges than their 'antipodes' in the upper part;
(e) Dutch MIV are in general larger than the corresponding US values, in particular at the lower part of the scale, where the Dutch categories are 'wider' than the corresponding US ones.

7.3.6 The same item in the same language, but in different nations

Our data set includes five items presented in two different nations, Chili and Mexico, which both use Spanish as their standard language.

Table 7.10 Cut points of the same item in two Spanish speaking nations

Item code	nation	cut points			
		1/2	2/3	3/ 4	4/5
O-SLL/u/sq/v/4/b	CL-1	2,0	5,3	8,8	-
	CL-2	1,5	4,8	8,8	-
	MX	2,1	5,6	8,8	-
O-SLS/c/sq/v/3/a	CL	3,1	8,0	-	-
	MX	4,0	8,2	-	-
O-SLu/g/sq/v/4/b	CL-1	1,2	4,8	8,6	-
	CL-2	1,7	5,1	8,7	-
	MX	3,0	6,1	8,6	-
O-SLu/g/sq/v/4/b	CL-1	1,3	4,7	8,4	-
	CL-2	0,9	4,4	8,6	-
	MX	2,4	5,7	8,6	-
O-SLW/c/sq/v/5/g	CL	0,9	3,5	5,9	8,8
	MX	2,0	4,1	6,4	8,9

The data in table 7.10 suggests that the corresponding cut point estimates at the upper end of the scale are not sensitive to the nations difference, but that this observation does not apply to the lower end of the scale. This effect deserves further investigation, also for other languages in different nations, e.g., English. Note that in table 7.10, except for the upper cut points, the Mexican are always located at larger happiness values than those from Chile.

7.4 Main findings on the application phase

7.4.1 Applications of some results obtained in the construction phase

In order get a first impression of how the method works in practice, we selected 20 existing studies within one nation (The Netherlands), that had been analyzed previously on the basis of one of the cases included in our set of 100 and in the same (Dutch) language. The observed data of these studies were re-analyzed by subjecting the happiness frequency distributions to an analysis, this time on the basis of the findings in the construction phase of that case. The 20 studies included five different items, so five cases.

COMPARISON OF THE SAME ITEM IN DIFFERENT NATIONS/LANGUAGES

FIRST ITEM: Measure code: O-SLL/u/sq/v/4/b — 4 response categories

DOCUMENTATION: http://www.worlddatabaseofhappiness.eur.nl/scalestudy/itemreports/osll_u_4b.html

Lead question (translated or verbatim):

On the whole, how satisfied are you with the life you lead?

This item is a standard in the studies of EUROBAROMETER

		CL-1 MIV	CL-1 cutpoint	CL-2 MIV	CL-2 cutpoint	HR MIV	HR cutpoint	MX MIV	MX cutpoint	NL MIV	NL cutpoint	US-1 MIV	US-1 cutpoint	US-2 MIV	US-2 cutpoint	US-3 MIV	US-3 cutpoint	HU MIV	HU cutpoint
very satisfied	4	9,4		9,4		9,1		9,4		9,1		8,9		9,0		9,0		8,8	
			8,8		8,8		8,3		8,8		8,1		7,9		8,0		8,1		7,6
fairly satisfied	3	7,0		6,8		6,4		7,2		6,7		6,0		6,2		6,3		5,8	
			5,3		4,8		4,5		5,6		5,2		4,2		4,4		4,6		4,0
not very satisfied	2	3,6		3,1		2,9		3,9		3,9		2,8		3,1		3,3		2,6	
			2,0		1,5		1,2		2,1		2,5		1,3		1,7		2,0		1,2
not at all satisfied	1	1,0		0,7		0,6		1,1		1,3		0,7		0,9		1,0		0,6	
		spanish2-1		spanish4-1		croatian1-2		spanish5-1		dutch3-7		english2b-10		english4a-3		english4b-3		hungarian2b-6	

SECOND ITEM: Measure code: O-HL/u/sq/v/4/a — 4 response categories

DOCUMENTATION: http://www.worlddatabaseofhappiness.eur.nl/scalestudy/itemreports/ohl_u_4a.html

Lead question (translated or verbatim):

Taking all things together, would you say you are.....?

This item is a standard in the studies of WORLD VALUE SURVEY

		CL MIV	CL cutpoint	HR MIV	HR cutpoint	RU MIV	RU cutpoint	RO MIV	RO cutpoint	NL MIV	NL cutpoint	US MIV	US cutpoint	HU [?] MIV	HU [?] cutpoint
very happy	4	9,3		9,2		9,0		8,9		9,1		9,1		9,0	
			8,6		8,4		8,1		7,7		8,2		8,2		8,1
quite happy	3	6,5		6,5		6,8		6,1		6,8		6,3		6,1	
			4,4		4,6		5,4		4,5		5,3		4,4		4,1
not very happy	2	2,6		2,9		4,0		3,1		3,9		2,8		2,7	
			0,8		1,2		2,7		1,6		2,5		1,1		1,2
not at all happy	1	0,4		0,6		1,3		0,8		1,2		0,6		0,6	
		spanish2-6		russian1-1		russian1-1		romanian1-1		dutch3-4		english2b-7		hungarian2b-4	

THIRD ITEM: Measure code: O-HL/c/sq/v/4/g — 4 response categories

DOCUMENTATION: http://www.worlddatabaseofhappiness.eur.nl/scalestudy/itemreports/ohl_c_4g.html

Lead question (translated or verbatim):

Taking all things together, how would you say things are these days - would you say you are..?

		CL-1 MIV	CL-1 cutpoint	CL-2 MIV	CL-2 cutpoint	HU MIV	HU cutpoint	RU MIV	RU cutpoint
very happy	4	9,4		9,2		9,0		8,9	
			8,7		8,5		8,0		7,8
pretty happy	3	6,6		6,2		6,1		6,5	
			4,4		3,9		4,2		5,2
not too happy	2	2,7		2,3		2,8		3,9	
			0,9		0,7		1,4		2,6
very unhappy	1	0,5		0,3		0,7		1,3	
		spanish2-7		spanish3-3		hungarian2b-1		russian1-3	

COMPARISON OF THE SAME ITEM IN DIFFERENT NATIONS/LANGUAGES

FOURTH ITEM: Measure code: O-HL/g/sq/v/7/a — 7 response categories

DOCUMENTATION: http://www.worlddatabaseofhappiness.eur.nl/scalestudy/itemreports/ohl_g_7a.html

Lead question (translated or verbatim):

If you were to consider your life in general, how happy or unhappy would you say you are, on the whole?

Item in ISSP 2002

		US MIV	US cutpoint	NL MIV	NL cutpoint	CL-1 MIV	CL-1 cutpoint	CL-2 MIV	CL-2 cutpoint	DE [?] MIV	DE [?] cutpoint
completely happy	**7**	9,7		9,7		9,6		9,7		9,6	
			9,3		9,4		9,1		9,4		9,1
very happy	**6**	8,5		8,6		8,1		8,4		8,2	
			7,7		7,8		7,1		7,4		7,3
fairly happy	**5**	6,6		6,8		6,1		6,4		6,2	
			5,5		5,8		5,2		5,3		5,1
neither happy nor unhappy	**4**	4,6		5,1		4,4		4,4		4,2	
			3,7		4,5		3,6		3,5		3,3
fairly unhappy	**3**	2,8		3,6		2,6		2,6		2,6	
			2,0		2,7		1,7		1,8		1,8
very unhappy	**2**	1,2		1,7		1,1		1,1		1,2	
			0,4		0,6		0,4		0,3		0,5
completely unhappy	**1**	0,2		0,3		0,2		0,2		0,3	
code in WDH		english2b-4		dutch3-3		spanish2-3		spanish3-5		german1-5	

FIFTH ITEM: Measure code: O-SLL/u/sq/v/5/a — 5 response categories

DOCUMENTATION: http://www.worlddatabaseofhappiness.eur.nl/scalestudy/itemreports/osll_u_5a.html

Lead question (translated or verbatim):

"I am very happy with the life I lead".

		NL-1 MIV	NL-1 cutpoint	NL-2 MIV	NL-2 cutpoint	NL-3 MIV	NL-3 cutpoint	US-1 MIV	US-1 cutpoint	US-2 MIV	US-2 cutpoint
strongly agree	**5**	9,2		9,4		9,3		9,0		9,2	
			8,5		8,7		8,6		8,0		8,4
agree	**4**	7,2		7,5		7,4		6,6		7,1	
			5,9		6,2		6,2		5,3		5,8
neither agree, nor disagree	**3**	5,2		5,4		5,4		4,3		4,7	
			4,4		4,5		4,5		3,3		3,6
disagree	**2**	3,4		3,3		3,4		2,2		2,6	
			2,3		2,1		2,3		1,1		1,6
strongly disagree	**1**	1,2		1,0		1,2		0,6		0,8	
code in WDH		dutch3-5		dutch4-3		dutch4-7		english4a-4		english4b-4	

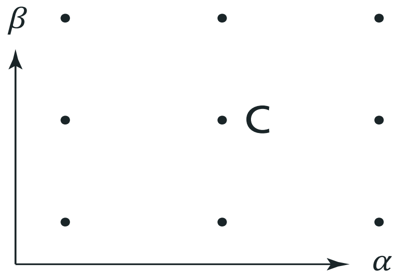
From the previous study, as recorded in the WDH, section Happiness in Nations, we already knew the complete relative frequency distribution of the sample. This enabled the calculation of the average value and the standard deviation of the distribution applying five different models (see section 6.6):

I. the traditional method, i.e. by treating the code numbers of the categories as if they were cardinal numbers, followed by a linear scale transformation of a [1, *k*] scale onto a [0, 10] scale, according to the procedure as described in Appendix B.
II. the same source provided estimates of the population mean and standard deviation on the bases of the 'Thurstone transformation' of the ratings, as described in section 6.2.
III. the Veenhoven model, which also assumes a polytomous discrete population distribution, but in which the code numbers of the categories are replaced this time with their mid-interval values (MIV).
IV. the semi-continuous model, assuming a continuous latent happiness variable, distributed uniformly within each category, resulting in the same mean estimate as model III, but in a higher standard deviation estimate.
V. a continuous distribution of the latent happiness variable, which is described as a beta distribution on the interval [0, 10] with two shape parameters α and β.

Applying the semi-continuous model IV, the estimated mean value and standard deviation were calculated following the procedures described in Appendix F.1, We constructed on this basis, 95% confidence intervals (further denoted CI95) for the true but unknown population mean happiness value. Furthermore we approximated the maximum likelihood estimators (further abbreviated MLE) of the shape parameters α and β in model V.

7.4.2 Estimation of the shape parameters of the beta distribution in the fully continuous model

The fully continuous model is based on the beta distribution of happiness (model V). The shape parameters α and β of this distribution have to be estimated as maximum likelihood estimators (MLE). Approximated values were obtained by using the following heuristic procedure. In the α - β plane, we plot nine points, each of them representing a beta distribution on the interval [0, 10], being its abscissa and its ordinate, as follows :



The location of the points is selected in such a way that the target (α, β) combination is expected to be found within this 'square', assuming that, within this area, the function log $L(\alpha, \beta)$ has only one single maximum and no saddle point. In order to realize this, we select the moment estimators of α and β as obtained by substitution of the values of the sample average and variance in the expressions in [H.20] and [H.21] in Appendix H as the coordinates of the initial centre C. The initial dimensions of the square are adopted arbitrarily as 2,56 x 2,56 whenever possible within the limits of the shape parameters.
Now for each of these nine distributions the value is calculated for:

[7.1] $$\log L(\alpha,\beta) = -N\log[B(\alpha,\beta)] + \sum_j n_j \left[\log \int_{b(j-1)}^{b(j)} h^{\alpha-1}(10-h)^{\beta-1}dh\right]$$

ignoring the constant term - N log(10).

There are two possible outcomes: either the value of the above expression for the distribution C_1 is larger than any of the other eight, or it is not. If the maximum value for the expression [7.1] does not correspond to C_1, but to one of the other eight distributions, that one is selected as the new centre C_2 for the next iteration step, without reducing the dimensions of the square this time. Otherwise, the next iteration step is made by maintaining C_1 as the centre, so C_2= C_1, but now by halving the dimensions of the square. In this way the procedure is applied in 20 iteration steps; it was assumed that this number of steps is sufficient.

It appeared that in the last five steps, none of the estimates changed by more than 1 %, which was considered sufficiently accurate for this purpose. The values obtained in this way show considerable dispersion: among the 20 Dutch recalculated surveys, the estimates $\{\hat{\alpha}\}$ varied between 4,2 and 10,6 and even in studies within the same case differences of several units were observed. The estimates $\{\hat{\beta}\}$ are smaller and these range from 1,2 to 2,9. Since both parameters are rather sensitive to changes in the population mean μ, this is less alarming than it looks. Moreover, it should be borne in mind that μ is related to the ratio α/β only and not by their separate values. All twenty density curves are unimodal and skewed to the left; the density equals zero at both $H = 0$ and $H = 10$.

7.4.3 Confidence intervals in the semi-continuous model

Inspection of Table 7.11 learns that the 95% confidence intervals for the true, but unknown mean *H*-value, further denoted CI95, under the semi-continuous model (IV) are rather wide: 0,4 – 0,7. This width is based on the standard error of the mean. The latter is built up from two components; the first is introduced in the construction phase and the second one in the application phase.

The size of the first component is rather constant within the same item, since all studies apply that same MIV, albeit with slightly different weights due to different samples. The size of this variance component depends on the actual panel size, which for these cases varies between 28 and 32. For the five items the following average variance components have been found:

A	B	C	D	E
0,024	0,009	0,006	0,027	0,008

The smaller this component is after adjusting for substantial panel size differences, the more consensus within the panel on the position of the boundaries between the categories and hence on that of the MIV. For a judgement of the usefulness of a scale on this basis, e.g. that of the scale in case D , one has also to take the number of categories into account.

The component introduced in the application phase depends on the sample size, more specifically, it is proportional to N^{-1}. These variance components are on an average

A	B	C	D	E
0,004	0,005	0,006	0,001	0,002

The sample sizes are about 1.000 – 1.500, except for D with substantially larger samples. Clearly, in general the first component is mainly determining the size of the CI95, although for the 4-point scales the two are in a better balance. But the general recommendation for improving the precision of the estimated mean happiness value in the population is to enlarge the panel size rather than the sample size.

All twenty estimates according to model V are slightly larger than those on the basis of model IV, but each one is well within the corresponding CI95. Within each item, all differences between the estimates from both models

Table 7.11 Estimated mean values and standard deviations 1981-2008 in The Netherlands

Case	Year	estimated mean in different models				st.dev		CI95
		I Traditional	II Thurstone	III and IV Veenhoven-Kalmijn	V-IV difference beta distribution	I	IV	IV CI for true unknown population mean happiness value.
O-HL/c/sq/v/3/ab: Taking all things together, how would you say things are these days? Would you say you are…? (1) not too happy; (2) pretty happy; (3) very happy								
A	1982	6,9	7,6	7,1	+0,18	3,1	2,1	[6,7; 7,4]
	1983	7,1	7,7	7,2	+0,17	2,9	2,0	[6,8; 7,5]
	1984	6,9	7,6	7,1	+0,17	2,9	2,0	[6,8; 7,4]
	1985	6,7	7,5	6,9	+0,17	2,9	2,0	[6,6; 7,3]
	1986	7,0	7,6	7,1	+0,16	2,9	2,0	[6,8; 7,5]
O-HL/u/sq/v/4/a: Taking all things together, would you say you are.....? (1) not at all happy; (2) not very happy; (3) quite happy; (4) very happy.								
B	1981	7,7	7,7	7,5	+0,16	1,7	1,5	[7,2; 7,7]
	1990	8,0	8,0	7,7	+0,16	2,2	1,8	[7,5; 7,9]
	2006	7,9	7,9	7,6	+0,15	2,0	1,7	[7,3; 7,8]
	2006	7,9	7,9	7,6	+0,14	2,0	1,7	[7,4; 7,9]
O-SLL/u/sq/v/4/b: On the whole, how satisfied are you with the life you lead? (1) not at all satisfied; (2) not very satisfied; (3) fairly satisfied; (4) very satisfied.								
C	1981	7,8	7,5	7,5	+0,13	2,3	1,9	[7,3; 7,7]
	2000	7,9	7,6	7,5	+0,13	2,1	1,7	[7,3; 7,8]
	2006	8,1	7,7	7,7	+0,16	2,1	1,7	[7,5; 7,9]
	2007	8,3	7,8	7,8	+0,16	1,9	1,5	[7,6; 8,0]
	2008	8,2	7,8	7,8	+0,15	2,0	1,7	[7,6; 8,0]
O-SLL/c/sq/v/5/d: How satisfied are you with the life you currently lead? (1) not so satisfied; (2) fairly satisfied; (3) satisfied; (4) very satisfied; (5) extraordinary satisfied.								
D	1980	5,7	8,3	6,7	+0,09	2,5	1,9	[6,4; 7,1]
	1997	5,8	8,5	6,8	+0,06	2,2	1,7	[6,5; 7,1]
	2000	5,9	8,5	6,8	+0,14	2,2	1,6	[6,5; 7,2]
	2002	5,8	8,5	6,8	+0,07	2,2	1,7	[6,5; 7,2]
	2004	5,8	8,5	6,7	+0,13	2,2	1,7	[6,4; 7,1]
O-HL/g/sq/v/7/a: If you were to consider your life in general, how happy or unhappy would you say you are, on the whole? (1) completely unhappy; (2) very unhappy; (3) fairly unhappy; (4) neither happy, nor unhappy; (5) fairly happy; (6) very happy; (7) completely happy.								
E	2002	7,1	7,3	7,3	+0,06	1,4	1,4	[7,1; 7,5]

are relatively small, and also quite stable, whereas they decrease as the scale includes more categories: from 0,17 on an average for a 3-point scale to 0,06 for a 7-point scale.

For each of the cases A – D, the estimated mean values according to the semi-continuous model within the same case are quite close to each other: this range ≤ 0,3, which is much less than the range 7,8 – 6,7 = 1,1 for all twenty estimates together. This finding was confirmed in a simple one-way analysis of variance. In addition, we considered two more potential causes of the variability among the twenty estimated values:

(a) period:
1980 – 1990: range 6,7 – 7,5 vs
1997 – 2008: range 6,8 – 7,8;

(b) happiness vs. life satisfaction in a narrower sense:
happiness (cases A,B and E) : 6,9 – 7,7 and
life satisfaction (cases C and D): 6,7 – 7,8.

Clearly, there is no indication that these variables are serious candidates for explaining the differences between the twenty Dutch studies. The same applies to differences of composition of the panels, since all five cases have been obtained within one session, so almost from the same panel.

Since all five estimated averages in case D are smaller than any of the 15 estimates of the four other cases, the cause of the difference is supposed to be found in some particularity of this item. The fact that four of the five categories of this scale have a 'positive' label may be responsible for this problem, which has not been investigated into more detail.

The within-item variability, expressed as the range ≤ 0,3, is also much smaller than the CI95 width. The reason for this difference is that the CI95 is based on a standard error that has a relatively large 'sample error component', albeit this time due to the panel. All these surveys studies, however, apply the same case, in which the same panel has been involved, so this does not provide an explanation of the 'anomaly'.

Point estimates obtained according to the traditional method, i.e. by treating the code numbers of the *k* categories as if they were cardinal, followed by a linear transformation of a [1, *k*] to a [0, 10] scale, are sometimes covered by the CI95 (cases A and E), but not always. For case C, they are always above the upper confidence limit, whereas for case D, all estimates were below the lower confidence limit of model IV. In case B they are close to the upper confidence limit. Within each case, the mean happiness estimates cover a wider range than those obtained according to model IV. The estimated standard deviations obtained according to the traditional method are always substantially larger than those found using the semi-continuous model, which is not surprising.

Estimates on the basis of Thurstone model (II) show a different pattern. After rounding to 0,1, none of the 20 estimates is smaller than the corresponding point estimates according to model IV. In cases C and E both estimates are always almost identical, but in the three other cases, the Thurstone method delivered larger estimates. In particular the estimates of case D are considerably larger and these differences deserve special attention, which is an additional reason to label this item as suspect and to suspend its application until the problems have been clarified.

7.4.4 Comparison to estimates obtained using a numerical scale

An important objective of the HSIS is the wish to make estimates obtained from items with verbal responses comparable to those made on the basis of numerical scales. Whereas the number of categories of verbal items is never more than seven, this is in practice the minimum for numerical scales. The most frequently used numerical scale is the [1, 10] scale where happiness is measured as a discrete variable and the ratings are presented and assumed to be equidistant, both before and after direct stretching to a 10-point (!) [0, 10] scale. Since several surveys in The Netherlands have been conducted using such scales, we can compare the estimates to those listed in our table 7.11.in the same year. (Veenhoven 2010)

The questions on the numerical scales were identical (C) or similar (B, D, E) to the lead questions of the corresponding verbal scale as specified in table 7.11. The questions of the discrete numerical scales were:

B: “Taking all together, how happy would you say you are ?” ; [1, 10] scale

C: “On the whole, how satisfied are you with the life you lead ?” [1, 10] or [0, 10] scale

D: “On the whole, how satisfied are you with the life you lead ?” [1, 10] or [0, 10] scale

E: “How happy are you on the whole ?” [0, 10] scale.

The results are collected in Table 7.12

Table 7.12. Comparison of estimated means from verbal and numerical scales in The Netherlands (1981-2008). All mean values on a [0, 10] scale. In the right hand column the primary numerical scale of measurement.

		Veenhoven-Kalmijn		Numerical scale[2]	
Case	Year	Mean	CI 95	Mean	Scale
B	2006	7,6	[7,3;7,8]	7,8	[1, 10]
	2006	7,6	[7,3; 7,9]	7,8	[1, 10]
C	1981	7,5	[7,2; 7,7]	7,4	[1, 10]
	2000	7,5	[7,2; 7,8]		
	2006	7,7	[7,5; 7,9]	7,5	[0, 10]
	2007	7,8	[7,6; 8,0]	7,7	[0, 10]
	2008	7,8	[7,5; 8,0]	7,8	[1, 10]
D	1980	6,7	[6,4; 7,1]	7,4	[1, 10]
	1997	6,8	[6,5; 7,1]	7,6	[1, 10]
	1999			7,6	[1, 10]
	2000	6,8	[6,6; 7,2]	7,6	[1, 10]
	2002	6,8	[6,6; 7,2]	7,7	[0, 10]
	2004	6,7	[6,4; 7,1]	7,6	[0, 10]
	2007			7,5	[0, 10]
	2008			7,7	[0, 10]
E	2002	7,3	[7,1; 7,5]	7,8	[0, 10]

The estimated mean on the basis of a numerical scale is sometimes within the CI95 of the verbal scales according to Veenhoven-Kalmijn (B and C) , but not for cases D and E. This is one more reason to discard at least the suspect case D. Moreover, the V-K estimate in case E does not fit very well in the total happiness pattern of The Netherlands in that period. If larger panels are applied, the CI95 will be narrower and will cover the 'numerical scale mean' less frequently. It is clear that this comparison deserves more future research effort, but at least one of the candidates for causing these anomalies is the structure of the verbal item, and in particular the labels of the categories and their positions in the order .

It has already been noted that the CI95 in table 7.11 are rather wide. One might compare this result to e.g. the results from a Dutch survey in 2002 with

[2] *Full text of questions and observed responses taken from 'Country report Netherlands' in the WDH (Veenhoven 2010).*

a [0, 10] numerical scale and a sample size of 2.360. In this case we obtained a CI 95 = [7,80; 7,91], which is substantially narrower. This, however, is not an entirely fair comparison, since the latter result is based on a discrete model. All discrete happiness models deliver more precise estimators for the population mean happiness than related models that assume a continuous happiness variable. This is always attained at the cost of the validity, since in discrete happiness models one variance component is structurally ignored, being the variability between subjects that select the same rating, but who are only approximately equally happy (section 7.4.3); however, this explains only a very minor part of the discrepancy. Actually, this finding supports the hypothesis that numerical scales are superior to verbal ones, even if the best model is applied in the latter case.

One may wonder whether, and if so, how an approach similar to the HSIA can also be applied to numerical scales. The answer to this question requires one to bear in mind that the three essential elements of the HSIA are:

(i) happiness in the population is not only considered to be a continuous variable, but also treated as such

(ii) measured happiness is considered to be an element of an interval and not a single value

(iii) in the cut points of the intervals, the cumulative frequency distribution value is known directly from the observed frequencies in the sample.

This means that such an approach is not applicable as long as discrete scales of measurements are applied. However, if one would be prepared to consider an e.g. [1, 10] scale to be a semi-continuous one according to the principle as described in section 2.7 (see Fig. 2.5), a solution similar to the HISA is conceivable. An additional assumption should be the equidistance, which for $k \leq 7$ seems reasonably well acceptable. This as is done in the WDH.
In this approach, we consider a rating j representing the half-open interval $(j - ½, j + ½]$. The total scale is then a [½, 10+½] scale with length just equal to 10. So in this particular case, the transformation of ratings should not be done by stretching, but simply by subtracting the value 0,5. Now the k–1 = 9 cut points are {1½, 2½, ..., 9½} before and {1, 2, ..., 9} after this transformation onto a [0, 10] scale. Obviously these cut point values after transformation should be applied in this approach, and is the calculation of the 'adjusted' standard deviation the appropriate one.

The same assumptions can also be made for [1, k] numerical scales with $7 \leq k < 10$, but the procedure is slightly more complicated, since in this case not only subtraction is required, but also subsequently proportional stretching. As the reader can verify, the position of the j-th cut point on the [0, 10] scale simply equals $10j/k$, whereas the j-th MIV has the position $(10j-5)/k$ for j=1(1)k.

Contrary to the application to verbal scales, these cut points and MIV can be considered to be free of random errors, which improves the precision of the estimated mean happiness of the population compared to that obtained by the application of a verbal scale.

7.4.5 Sensitivity

The question may arise as how sensitive estimates are to variations in the average position of the cut points as reported by the judges. In order to get an impression of this, we considered the first of the twenty recalculated surveys in table 7.11 (item O-HL/c/sq/v/3/ab ; survey 1982). We modified the position of the cut point between “very happy” and “pretty happy” artificially from 7,73 to 7,23 and to 6,73 by subtracting 0,5 and 1,0 respectively from all individual judgements on this cpp. On the basis of this hypothetical input into the construction phase, we re-estimated the population mean happiness value according to the semi-continuous model IV. This estimate was reduced from 7,04 to 6,83 and 6,59 respectively. We expected the impact on the estimated standard deviation to be very modest and we found a reduction of 2,10 to 2,05 and 2,02 respectively. This result is indicative of the sensitivity to cut point variations, as it is based on one single application only, but the preliminary message is that variability in the position of the cut points has an impact on the estimated mean happiness that should not be neglected without further investigation.

7.5 Discussion

7.5.1 Views on the nature of happiness and satisfaction

In section 2.4 we have raised the question of what are our views on happiness and their consequences for scales of measurement, in particular the distinction between unipolar and bipolar scales of measurement. If one considers happiness to be an intensity variable with high happiness values at the upper end and lower values near to the bottom, a unipolar scale type is the obvious choice. There is only one pole, that of happiness. Items A, B, C and D in table 7.11 are all examples of such a scale. The higher the code number of a category, the larger the intensity of happiness/satisfaction as it is experienced by the judge.

Case E, however, is a typical example of a bipolar scale. There are two poles: a happy and an unhappy one; the anchor point in the middle of the scale obviously represents the equilibrium between both. Such a scale is usually symmetric with respect to that neutral category and if a 7-points bipolar scale is chosen, its usual form for application in other situations than with respect to happiness is [-3, -2, -1, 0, 1, +2, +3]. As has been pointed out in section 2.4.3, all these scale point values are augmented by 4, when applied to measurement of happiness.

Of the rejected case and the seven doubtful ones in the HSIA construction phase, all four scales with k = 3 or 4 were unipolar, and all the other four, with k = 5 or 7, were bipolar

The unipolar approach seems to fit best to the view of happiness as an intensity variable. If a judge has adopted this view and subsequently is presented with a bipolar scale, this may cause the judge to feel it is a difficult task to locate the cut points appropriately, and problems with the execution of this task are not unlikely. On the other hand, understanding labels like "completely (un)happy" in case of bipolar scales is a very difficult job and it is not unlikely that for many judges assigning a zero-width to a terminal category will be seen as the only way-out of their confusion.

Most items in this study apply a scale of the unipolar type (34 of the 52 items) against 18 of the bipolar. The distribution over the various k-values is given in table 7.13.

Table 7.13 Unipolar and bipolar scales in the 52 items in this study

Scale type	k = 3	4	5	7	total
Unipolar scale	14	14	6	-	34
Bipolar scale	-	8	7	3	18

Apparently, in this data set there is a clear shift from unipolar to bipolar as the number of categories increases. If we consider the items with k = 4, there are 19 cases with 8 different items in which the respondent in the application phase is asked to report his happiness; 6 out of these 8 items apply a unipolar scale against 2 where a bipolar is applied. Asking for satisfaction with life in a narrower sense was asked in 29 cases with 14 different items, of which 8 use unipolar against 6 bipolar scales. For *k* =5, these numbers were 2 and 2 for the 4 happiness items and 4 and 4 respectively for the 8 items on life satisfaction in the narrower sense; the item on the mood was ignored in this respect. For a correct interpretation of these frequencies, one has to bear in mind that the 52 cases in this study can in no way be considered to be a random sample from all items registered in the Item Bank of the WDH.

7.5.2 Skipped judgements and acceptability of cases

The analysis on the basis of the additional questions (mother tongue, seriousness and difficulty) of the skipped judgements did not reveal a clear difference between them and the accepted judgements, so the causes of the problems are to be found at a higher level than at the individual. Several hypotheses can be formulated as potential explanations.

(i) The panel members, at least most of them, might be relatively happy persons. As a consequence, they are then less able to make adequate judgements on situations as these are more remote from their own.

(ii) Judges did not, or not correctly, understand the instructions, so the briefing was inadequate, not only to the judges, but may be also to the person who had to run the local session;

(iii) Some scales are confusing the judges because of the use of labels that were felt to be almost identical synonyms or to have an order within the scale that did not fit well with the common meanings in the language/culture in question. Confused judges may develop irrational behaviour when performing their task, resulting in their judgement being skipped by the coordinating investigator.

(iv) Confusion may also be expected if panel members are presented with a unipolar scale, while a bipolar scale fits much better in their feeling on happiness or on life satisfaction, or just the reverse.

In this context, the most important and clear finding was that problems arise predominantly at the lower part of the scale. This is taken into account by the first hypothesis, but leaves us with the question why the Hungarian cases do not give a better performance, although the Hungarian people are not known to be the most happy people in the world. If the explanation holds, the Hungarian judges are much happier than the average Hungarian and this nonrepresentativity makes their contribution at least questionable. Anyhow, the until now neglected, but certainly possible relation between the judgement of the judge in the construction phase and his personal happiness as a confounding element between the two HSIS phases deserves future investigation.

The second option is supported not only by the observation that in a relatively large Russian panel none of the judges had to be skipped, but also by the fact that in the HSIS results have been incidentally submitted from a panel consisting of only one single judge.

An example of (iii) is the case with item O-HL/c/sq/v/5/n, mentioned in section 7.4. The scale of this item is more or less bipolar, but asymmetric with both the labels “fairly (un)happy” and ”rather (un)happy” within one scale. A second example will be discussed later on as case D in the context of an application study with 20 Dutch results (section 7.5.4). Together with (iv), these examples demonstrate that the HSIS may contribute to a separation of the better from the worse scales, but the development of good objective criteria is not an easy job, even if one succeeds in confining oneself to rational elements only. Of course, this operation is not very meaningful as long as there is not more clarity and consensus on the views on the nature of happiness.

The above exposition demonstrates that more research is necessary on this point and that without further investigation no progress in this respect can be expected.

7.5.3 Different cases with the same item

From the examples in which the same item was applied in different nations/languages (pages 158 and 159), it is clear that it is meaningful to evaluate verbal scales in different cultures/language separately, since differences of more than one *H*-unit on a [0, 10] scale are by no means exceptional.

The above mentioned survey demonstrates clearly than in the application the HSIS results obtained from the same nation/language are to be used. In this perspective, results of both analysis carried out according to the traditional method and the application of Thurstone values are most questionable, due to their universal nature of the position of the categories on a [0, 10] happiness scale.

7.5.4 Comparison with estimates obtained along other lines

From the comparison of the 20 Dutch studies, it appeared that the estimates on the basis of HSIS delivered rather stable estimates of the mean happiness value, albeit with the exception of one item as compared to the other four. With the exception of this item, most differences of the results on the basis of the HSIS were modest, none exceeding 0,6 unit on a [0, 10] scale; they did not demonstrate a clear systematic pattern.

The exception mentioned above concerns the case labelled D with code O-SLL/c/sq/v/5/d. The most plausible explanation of the discrepancies of these and the various other estimates of the population mean happiness value may be found in a combination of two conditions.

One is the structure of the item, more specifically the unusual location on the scale of the category "fairly satisfied", positioned as the category next to the least satisfied one. In the HSIS this results in a MIV of 4,6 on a [0, 10] scale. The other condition is the fact that the Thurstone values have been defined as context-free numbers, i.e. that the values are independent not only of the number and the labels of the other $k-1$ categories, but also of their positions within the scale. For "fairly satisfied" the Thurstone value 6,5 has been adopted as the standard in the WDH. In the traditional method, the text of the label is in no way taken into account; the position of the second out of five categories is always exactly 2,500 000. The above mentioned finding is an argument against the way Thurstone values have been selected in general, but it may also be a good reason to discontinue the use of this specific item in future happiness studies. Generally speaking the CI95 for the true, but unknown population mean happiness value were rather wide and this makes it desirable to increase the number of judges, if possible to at least 100, but a panel size of 200 is to be preferred.

A more serious problem seems to be the composition of the panels, which predominantly consisted of undergraduate students. We have not investigated to what extent their terminological opinions are in line with those of a random sample from the target population in the application phase. However, with such a random sample (more) problems may be expected with judges who do not completely understand the instructions. See also section 7.5.6.

7.5.5 Bias and precision of the population parameter estimate. Model validity

One of the most important properties of an estimator of a statistical distribution parameter is its bias. This bias is the difference between the expected value of the estimator and its target value, usually the true value. The usual definition of the bias B of $\hat{\theta}$ as an estimator of a parameter θ is $B := \mathcal{E}\hat{\theta} - \theta$ where $\mathcal{E}\hat{\theta}$ is the expected value of $\hat{\theta}$. An estimator is unbiased if

B = 0 and it is called positively biased if $B > 0$. From this definition it follows that a discussion on a bias of an estimator is only meaningful in relationship to the associated distribution model and its definition in statistical terms.

Within the context of the HSIA, we have described three models for happiness as a random variable, i.e. as a variable with a probability distribution. The first one is the polytomous discrete model, in this chapter also referred to as model III or as the "Veenhoven model". In this model, the happiness variable can adopt only a small number (k) of discrete values. This does not only apply to the measured happiness in the sample, but equally to the population that is represented by that sample. In the context of the HSIS, one can decide to accept e.g. the k MIV as the H-values of the polytomous distribution and the j-th observed relative frequency f_j as the unbiased estimator of the corresponding probability π_j of the population probability distribution with $\sum \pi_j = 1$.

Within this model, the sample average value is an unbiased estimator of the population mean value. Theoretically, the sample standard deviation as an estimator of the population standard deviation is slightly biased, but for sufficiently large samples, say > 100, this bias can be ignored in practice.

In terms of similarity of the model and the subject of investigation, the validity of any discrete distribution as a model is debatable. A distribution in which a nation consists of say just four kinds of people with respect to their happiness does not fit in any view on happiness in nations and its frequency distribution, in particular not if it is borne in mind that the number of types, i.c. four, is just an accidental decision made by the researcher of that specific study.

The model V assumes a beta distribution for the latent happiness variable and is no doubt the most valid one of the three. There is, however, a serious problem with respect to the assessment of the inaccuracy of the estimator of the population mean, which in practice we are unable to estimate. The unacceptable consequence is that it is impossible to report a CI95 for the population mean happiness under this model. There are several reasons for this. The first is the property of MLE that they are always consistent, but not necessarily unbiased. A more serious problem is that the estimation of the variance of MLE requires the calculation of the second derivative of the beta function with respect to its parameters, which is quite a difficult mathematical problem. Finally, even if the MLE of the shape parameters α and β were unbiased, this would not necessarily imply that $\hat{\mu} = \hat{\alpha}/(\hat{\alpha} + \hat{\beta})$ can also be assumed to be unbiased.

Given the fact that the model V estimators of the population are only slightly higher than those found on the basis of model IV, and that they fall well within the CI95 for the true but unknown population mean happiness value , we recommend that the semi-continuous model (IV) is applied for the estimation of the population happiness distribution parameters , in spite of its lower degree of validity when compared to model V.

7.5.6 Results from other, but similar studies

Several publications have been issued on studies in which attempts have been made to link adverbs such as e.g. "pretty" to values on a scale. Examples are Mosier (1941), Jones & Thurstone (1955), Myers & Warner (1968), Bartram & Yelding (1973), Wildt & Mazis (1978), Voss et al. (1996), Smith et al. (2005) and Worcester & Burns (1975). Some of these authors have additionally investigated the effect of translations of labels into other languages.

Generally speaking, the authors ignore the influence of the exact question formulation and that of other categories present on the scale; Worcester & Burns (1975) are a positive exception. Moreover, the above authors do not think in terms of intervals; their approach is based on the Visual Approach Scales, more like that of the Thurstone values; see section 6.2. So for the present investigation, the value of their findings is rather limited. Nevertheless, we shall refer to some of their findings which are relevant for our own research Myers & Warner (1968) report a study in which 126 US judges were asked to assign an integer number to each of 50 labels, varying from 1 := "the worst thing I could say about a product", up to 21 := "the best thing I could say about a product". The panel consisted of four rating groups: housewives, business executives, graduate business students and undergraduate business students. In table 7.14, for 4 out of these 50 labels, the average values and the within-group standard deviation value is shown, both after a direct rescaling of a [1, 21] to a [0,10] scale for comparison reasons. These observations demonstrate that the composition of a sample, and hence that of a panel influences the result to an extent that is not negligible.

Table 7.14. Means and standard deviations of four rating subgroups according to Myers and Warner (1968) after transformation to a [0, 10] scale.

Label	subgroup	house-wives	bus. exec.	grad. stud.	under-grad.	range
"very good"	average	7,2	7,9	8,0	7,9	0,8
	st. deviation	1,4	1,3	1,1	0,7	
"quite good"	average	6,7	6,3	7,4	7,3	1,0
	st. deviation	1,4	1,5	1,0	1,0	
"good"	average	6,7	6,4	6,9	6,8	0,5
	st. deviation	1,0	1,6	1,1	1,0	
"fairly good"	average	5,5	5,5	5,7	6,1	0,6
	st. deviation	1,2	1,9	1,1	1,1	

Since the instructions in the above study were different from the ones we presented to the HSIS judges, no simple comparison is admissible as for the absolute values for the different labels. However, the within-group standard deviations, which quantify the (lack of) consensus within a subgroup, are well in line with the values obtained in HSIS. The right hand column gives the ranges, rounded to 0,1 between the average values of the four subgroups. This makes clear that in the study of Myers and Warner, different subgroups within the panel report different rating levels on the same label; this applies to all four labels in our selection, but also to the vast majority of the other 46 labels in their study. So there are good arguments for the hypothesis that the composition of a panel may result in a serious sample bias. Therefore an investigation into what extent the HSIS judges represent the corresponding national population is strongly recommended.

Smith et al. (2005) investigated the same problem, albeit along a different line. Part of their study concerned adverbs in relation to the adjectives "important" and "unimportant", which looks like quite similar to happy/unhappy. They found the following order of increasing importance, where e.g. "not at all" in this context means "not at all important":

not at all < not < not very < not too < neither IMPORTANT,
nor UNIMPORTANT < fairly < pretty < IMPORTANT (without adverb)
< quite < very < definitely < completely < exceptionally < extremely.

For "happy", there are no reasons to expect an order with many 'inversions'. This means that scales with such inversions are to be considered at least as candidates for rejection. No such scales have been included in our set of 100 cases.

Although their study concerns numerical and visual analogue scales only and no verbal ones, Mazaheri & Theuns (2008) report a number of findings, of which those on uni/bipolar scales may be interesting in view of the problems discussed in section 7.5.1. Students of their University were requested to report their own life satisfaction, but additionally also their life dissatisfaction, both on one of 12 different scales. As for the first response, they found that the application of a [-5, 5] scale resulted in remarkably fewer responses in the [-5, -1] part of the scale than on the same question in the corresponding part [0, 4] of a [0, 10] scale. As a result the latter delivered lower average values than the [-5, 5] scales. The (transformed) average score on the [-5, 5] scale was 7,5 and no clear difference between unipolar and bipolar scales could be detected. The average values of the [0, 10] scales were smaller, 0,4 points of a [0, 10] scale on a unipolar scale and 0,7 points on a bipolar scale, both on an average. The results of the dissatisfaction measurement, however, revealed a remarkable difference. When applying unipolar scales of measurement, dissatisfaction appears to be measured simply as the complement of satisfaction. Their sum is almost exactly equal to 10 on a [0, 10] scale; the difference always being <0,3. In the case a bipolar scale was presented, the sum exceeded at least the value 11,3. The combination of our findings and those of Mazaheri and Theuns make clear that more research in necessary, in particular on the nature of satisfaction and dissatisfaction. Is dissatisfaction more than "lack of satisfaction" only and if so, what more and under which condition is that found to be manifest ?

7. 6 Conclusions and recommendations

7.6.1 Conclusions

For verbal scales with categories to measure happiness in nations, the HSIA and the subsequent application of the semi-continuous model (IV) provides estimators of the true but unknown mean happiness value and the within-nation standard deviation, which are superior over statistics obtained by current methods like the traditional direct rescaling and on the basis of Thurstone values, at least qualitatively, i.e. in terms of content validity. This approach considers happiness no longer as a discretely distributed variable, but allows for its continuous nature. In this way, the method described in this thesis is no doubt closer to the reality and is to be considered more relevant for social scientists than conventional methods are.

Moreover, as compared to the method of direct rescaling, the criticism of the latter method does not apply to the results obtained according to the HISA method. This especially includes the objections against completely ignoring of the labels of the categories and the controversial treatment of ordinal ratings as if they were cardinal, since in the proposed approach, no equidistance between the ratings is assumed. This is particularly in the interest of meta-analytical studies, since the proposed method will be more valid than one that operates according to previously current methods and hence will facilitate a better comparison.

An additional advantage over the discrete models is that the 'scale values' are no longer considered – and treated – as error-free, and their inaccuracy due to their stochastic nature is fully taken into account in the calculation of the standard deviation. Some scientists may regret the negative impact of this approach to the precision of estimates, but better validity has a price, at least in this case.

Although the fully continuous model on the basis of a beta distribution of a latent happiness variable(V) has an even better validity in the sense of a better correspondence to what is assumed to be happiness, we recommended that investigators apply the semi-continuous model (IV), since the latter model enables them to estimate the inaccuracy and hence to construct 95 % confidence intervals for the true, but unknown mean happiness in the population that is represented by the sample. The estimated differences between the results on the basis of both models are of the order of 0,1 point on a [0, 10] scale, which is sufficiently small compared to the width of the confidence intervals, at least at current panel sizes. The observed differences between panels in different nations for the same item exceed 1 point on a [0, 10] happiness scale in about one third of the cases included in this study. Such differences correspond to differences in the estimated population mean happiness value of the order of some tenth's in a [0, 10] happiness scale. This finding justifies an approach that allows for differences between nations/languages/cultures.

It might be alluring to compare the participating nations for specific differences. Although this would be certainly interesting, the number of cases for most nations in this data set was considered to be too small for a meaningful investigation on this subject. The results support the idea that it might be possible to distinguish between adequate and less adequate items, which may in future give a basis for continuing or discontinuing the various items in this field of research.

From the data described in this study, about 11 % of all judgements were skipped; this percentage is judged as higher than acceptable. No indications have been found that. Individual judges are responsible for this level. The problems seem to have been caused by either the inadequate instruction regarding how to make the judgement and how the instructions were to be understood, or by the construction of the specific item. However, if due to a high percentage of skipped judgements, a case is rejected, this is justified irrespective of which was the true cause.

Our collection included seven cases that were judged as "doubtful" and the question arises what should be done with such cases in future sessions. In principle, there are two possibilities. One is that such a doubtful case is an incident, i.e. that cases with the same item, but in other nations/languages/ sessions do not give rise to doubts. If so, we recommend to repeat the case in another session, if possible. Additionally, there may be a reason to reconsider and to improve the instructions. If, however, this item also gives rise to doubtful cases in other nations/sessions, then the obvious explanation is that the item should be considered to be ill-constructed.

The weakest points in the present set-up are (i) the instructions for judging,(ii) the number of judges per session and (iii) the anti-random composition of the judge panels. The study of Myers and Warner (1968) corroborates the conjecture that the composition of a panel influences the results to an extent that does not justify one to ignore this effect.

7.6.2 Recommendations for conducting future HSIS sessions

The results of 100 of the first cases of the HSIS justify the idea to expand this study to other nations and to other verbal items for measuring happiness and life satisfaction. The main additional recommendation is to ensure a good, i.e. uniform, clear, unambiguous and sufficiently detailed instruction for both the judges and, separately, for the person who is conducting the local session(s). An English version should be prepared of this instruction sheets, but also a translation that is checked thoroughly by at least two persons who are (preferably native) speakers of the language of the translated version. The check covers the briefing of the judges and of the local coordinator, but also the items involved. Moreover, to reduce the inaccuracy of the estimates, it is recommended that the judge panel size increases should increase to at least 100 judges, but preferably to 200.

Until now, we have asked all judges in each session whether the (local) language in which the items were presented, was their mother tongue or not. What should be done in case of a negative answer ? If it is given by a student from abroad who comes to visit the university for one or for a few years, we suggest that his judgements should be destroyed. If, however, the judge is a student who has lived (almost) all his life in that area, but has immigrated parents, no doubt his judgement should be included. The justification is that a sample from that nation will also include such respondents. If this view is adopted, this question should be replaced with e.g., "Have you lived permanently in this country for the past ten years ? " and the researcher should act according to the answer. However, in view of the poor results in elucidating causes of problems with 'difficult cases', a more obvious alternative would be to discontinue to raising such questions.

7.6.3 Recommendations for future methodological research

For future studies, it is desirable to find out to what extent the personal happiness situation of a judge influences his responses in the construction phase. Until now, we have assumed that there is no such confounding, but the only justification for this assumption is that it has not yet been investigated.

A second question to be answered is to what extent the actual composition of the various panels causes a serious bias with respect to the interpretation of the category labels.

A secondary objective of the HSIS operations is to make a screening between adequate and less or even not at all adequate items. This requires a further and fundamental investigation into the nature if happiness and life satisfaction as we measure it and what we would like to measure, i.e. how the respondent experiences this. More specifically: Do we continue to measure happiness just as an intensity variable, as has been pointed out in chapter 2 and by Kalmijn and Veenhoven (2005) ? Or would a representation as an (un)balance between happiness and unhappiness be more appropriate , at least in some cases ? If so, in which ones ?

In this data set, we have found that the majority of the problems with judgements occur at the lower end of the scale of measurement; insufficient instruction cannot explain this. This finding deserves further research; the explanations given in the present chapter are highly speculative and require further examination, including a possible connection with the immediately preceding paragraph.

The first results do not suggest a perfect concordance between the results obtained according to the HSIA and those in the same year and nation, but on the basis of numerical scale. This issue deserves more research, in which especially the ‘quality’ of the verbal items certainly should be taken into account. The outcomes of such a study could have inevitable consequences for the question which items are considered to be adequate and which are not. However, this is not the only criterion. As is demonstrated in sections 7.4 and 7.5, there may be also other ‘constructional’ properties of an item which can make it less appropriate. Comparison of the results from different items within the same nation, as has been described in section 7.4, may be helpful to this end.

One requirement could be that the width of the various intervals of the item should not be too diverse. If this diversity is sufficiently large, this can give rise to e.g. bimodality in the observed happiness distribution which is rashly, but incorrectly interpreted as a ‘split-happiness’ situation in the relevant nation. A possible criterion could be the ratio of the widths of the widest and the narrows interval. A critical value equal 2 could be an choice, but this is as arbitrary as any other value would be.

Eventually, this research should result in a set of criteria, on the basis of which the future choice of items can be standardized to a relatively small number and at least in a list of items that are recommended to be excluded from application to future happiness studies.

This reduction would be a major contribution to future research, in particular to meta-analytical studies. Editorial Boards of scholarly journals that consider themselves as leaders in this field could play an important role, especially if guidelines for authors include instructions on this.

✧ ✧ ✧

APPENDICES

A. Upper and lower bounds of the standard deviation.
Calculation of the maximum and minimum possible values of the standard deviation at different mean happiness values.

B. Direct rescaling.
Linear transformation of scores onto another rating scale.

C. Gini Coefficient and Mean Pair Distance.
How and under what assumptions can the relationship between the Gini-coefficient and the Mean Pair Distance can be established ?

D. Formulae for Inequality-Adjusted Happiness.
Derivation of the formulae for the computation of the IAH values.

E. The calculation of the parameter estimates according to the HSIA.
Description of an EXCEL computer programme for the calculation of the parameter estimates according to the scale interval approach.

F. Parameter estimation according to Happiness Scale Interval Approach.
Formulae are derived and collected, that are needed for the parameter estimation according to the Happiness Scale Interval Approach .

G. Rounding observations and statistics.
Rules are given for rounding parameter estimates in presentations in agreement with their precision.

H. Properties of the beta distribution.
Summary of properties of the beta distribution which are relevant for its application to happiness distribution models.

Appendix A
Upper and lower bounds of the standard deviation

Let the happiness in a nation be distributed on a domain $[u,h] \subset \mathbb{R}$, and let the mean happiness value m ($u \le m \le h$) be known. The question then arises which values the within-sample standard deviation s can adopt in this situation.

In order to answer this question, we consider a fictitious situation in which all inhabitants of a nation are either extremely happy (happiness rating = h on the chosen rating scale) or extremely unhappy (rating = u) in proportions of a ($0 \le a \le 1$) and $(1 - a)$ respectively. No happiness ratings in between have been selected by anyone this time.

In this case, the mean happiness value is:

[A.1] $m = a \cdot h + (1 - a) \cdot u = u + a \cdot (h - u)$,

and the variance is

[A.2] $\text{var} = a \cdot (h - m)^2 + (1 - a) \cdot (u - m)^2$.

From [A.1] it follows that

$a = (m - u) / (h - u)$,

and its substitution into [A.2] results in

[A.3] $\text{var} = (h - m) \cdot (m - u)$

For a given value of m, this value is the maximum attainable value for the variance. Any other situation, but with the same value of m, can be realized only if one or more extremely happy people selects a happiness rating between h and m, which requires that at the same time a, not necessarily equal, number of extremely unhappy people have to shift towards a rating that is closer to m. The result of this process is necessarily a smaller value of the variance, hence:

[A.4] $(h - m) \cdot (m - u) = \max(s^2) = [\max(s)]^2$

As the reader can verify, this relationship can also be written as:

[A.5] $[m - ½(h + u)]^2 + [\max(s)]^2 = [½(h - u)]^2$

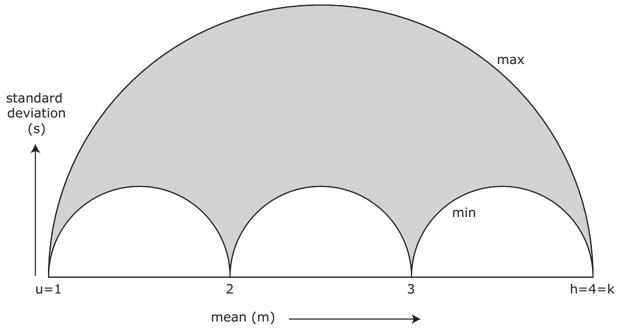


Fig A.1 Upper and lower bounds of the sample standard deviation

Therefore, plotting the theoretical maximum standard deviation max(s) against m results in a semicircle with a centre at the middle of the rating scale at the m-axis and a radius of ½ ($h - u$), so the theoretically maximum value of the standard deviation is equal to ½ ($h - u$). See fig. A.1.

Any real or fictitious happiness distribution can be represented by a point with co-ordinates (m, s), situated within this semicircle or at its boundary.

Summarising:

[A.6] $0 \leq s \leq \sqrt{(h - m) \cdot (m - u)} \leq \frac{1}{2}(h - u)$.

The minimum value of the standard deviation is equal to zero, which value is obtained if all people select the same happiness rating. Obviously, this value is attainable only at those values of m that correspond to one of the ratings on the original rating scale.

Between two consecutive values of m for which $s = 0$, inevitably $s > 0$, and there is an 'empty zone' in the m-s-diagram. For reasons that are similar to the above ones, such a zone is bounded by a semicircle with, in the case of an original rating scale, a diameter equal to unity and a maximum height of 0,5. In the case of transformed scales, these diameters need to be adjusted accordingly, although the number of 'empty zones'. will remain the value k–1.

The lower bound of the in the case of a discrete and equidistant [1, k] scale after linear transformation to a [0, 10] scale can be calculated as:

[A.7] $$s_{\min}^2 = \left[m \bmod \left(\frac{10}{k-1}\right)\right]\left[\frac{10}{k-1} - m \bmod \left(\frac{10}{k-1}\right)\right]$$

where *m* is the average value and *m* mod (*p*) is the remainder in the division of *m* by *p*, which can also be written as *m* mod (*p*) := *m*/*p* – INTEGER(*m*/*p*).

The value of its square root s_{min} has been tabulated in table A.1 for *m* = 0(0,5)10 and for *k* =3,4,5. The values for *k* = 11, i.e. the case of a discrete [0, 10] scale, have been added for comparison reasons. In the case of non-equidistant ratings, the semicircles upon the *m*-axis have unequal diameters, the values of which have to replace the value 10/(*k*-1) in the above equation.

Table A.1 Minimum and maximum standard deviation in the case of an equidistant k-point scale for average values m = 0(0,5)10 and k = 3,4,5,11

	standard deviation				
average value	|-------------- minimum			--------------|	maximum
m	*k*=3	*k*=4	*k*=5	*k*=11	any *k*
0,00	0,00	0,00	0,00	0,00	0,00
0,50	1,50	1,19	1,00	0,50	2,18
1,00	2,00	1,53	1,22	0,00	3,00
1,50	2,29	1,66	1,22	0,50	3,57
2,00	2,45	1,63	1,00	0,00	4,00
2,50	2,50	1,44	0,00	0,50	4,33
3,00	2,45	1,00	1,00	0,00	4,58
3,50	2,29	0,73	1,22	0,50	4,77
4,00	2,00	1,33	1,22	0,00	4,90
4,50	1,50	1,59	1,00	0,50	4,97
5,00	0,00	1,67	0,00	0,00	5,00
5,50	1,50	1,59	1,00	0,50	4,97
6,00	2,00	1,33	1,22	0,00	4,90
6,50	2,29	0,73	1,22	0,50	4,77
7,00	2,45	1,00	1,00	0,00	4,58
7,50	2,50	1,44	0,00	0,50	4,33
8,00	2,45	1,63	1,00	0,00	4,00
8,50	2,29	1,66	1,22	0,50	3,57
9,00	2,00	1,53	1,22	0,00	3,00
9,50	1,50	1,19	1,00	0,50	2,18
10,0	0,00	0,00	0,00	0,00	0,00

Although any nation and any sample from it can be represented by a point inside the semicircle or at its circumference, the reverse is not true. There are two reasons for this. The first one, due to the k–1 semicircles, has been described above. The second arises from the fact that the sample sizes are always finite numbers. Let N be the size of a sample in which all subjects rate their happiness on a [1, k] scale. The mean happiness score in a sample is not a continuous, but a discrete variable. If we assume $u \leq h$, then $m = u$ (1/N) h and so $a = (m–u)/(h–u)$ and s are discrete numbers too.

To illustrate this, we will consider a sample of 4 subjects only, rating their happiness on a [1, 3] scale. The 15 different possible outcomes are listed in Table A.2

Table A.2 All possible outcomes in case of N=4 and k=3

ratings	m	s	max(s)
1111	1,00	0,00	0,00
1112	1,25	0,43	0,66
1122	1,50	0,50	0,87
1113	1,50	0,87	0,87
1222	1,75	0,43	0,97
1123	1,75	0,83	0,97
2222	2,00	0,00	1,00
1223	2,00	0,71	1,00
1133	2,00	1,00	1,00
2223	2,25	0,43	0,97
1233	2,25	0,83	0,97
2233	2,50	0,50	0,87
1333	2,50	0,87	0,87
2333	2,75	0,43	0,66
3333	3,00	0,00	0,00

Consequently, the number of different points (m, s_{max}) is a finite number, whereas the number of points of a true semicircle is not. Moreover, for some of those m-values, the standard deviation s reaches its 'semicircle value' $\sqrt{(h–m)\cdot(m–u)}$, as is computed in the right hand column, but for
m = 1,25(0,50)2,75 the maximum value of s is smaller than the semicircle value in the right hand column. The above considerations make clear that the statement that "any point inside the semicircle or at its boundary represents a possible nation" is not correct. For the line $s = 0$, it is not even approximately true for large samples.

Appendix B
Direct Rescaling
(Linear transformation of scores onto another rating scale)

Happiness is typically measured by self-report and cross-national studies on happiness mostly use single questions. An example of a commonly used item is presented below:

"Taking all together, how satisfied or dissatisfied are you currently with your life as a whole?"

1	2	3	4	5	6	7	8	9	10
Dissatisfied									Satisfied

In this case, happiness is rated on a 10-step numerical scale. Other items use verbal rating scales, e.g., the 4-step rating scale:

'very happy', 'fairly happy', 'not too happy' and 'unhappy'.

Happiness may be also rated on pictorial scales using smilies and other graphical scales. Whatever the scale used, the respondent has to select one out of a limited number of discrete ratings, which is recorded eventually as a number, in the above scales one of the numbers from the sets {1 (1) 10} and {1 (1) 4} or e.g. {0 (1) 3} respectively.

To compare results obtained using different scales, the results of the primary numerical scale are commonly subjected to a linear transformation onto a common 'secondary' scale. We shall give the formulae to be used for this transformation below.

Let r_1 = the rating on the primary scale,
h_1 = the rating on the primary scale for the most happy situation, and
u_1 = the rating on the primary scale for the most unhappy situation.

In the above first example $u_1 = 1$ and $h_1 = 10$.
The ratings after transformation are denoted r_2, h_2 and u_2 respectively.

In most studies $h > u$ is chosen, so $u \leq r \leq h$. Some researchers, however, prefer $u > h$ and in the latter case $h \leq r \leq u$.

The three underlying assumptions for the linear transformation of happiness ratings are:

(a) the possible ratings of the primary scale are considered to be observations at the 'metric' level of measurement, so it is admissible to apply the common arithmetic operations,

(b) $u_1 \rightarrow u_2$, and

(c) $h_1 \rightarrow h_2$.

The last two assumptions mean that the extreme possible ratings of the primary and the secondary scale are assumed to correspond perfectly to the same verbal or pictorial description label.

The situation in which $h_1 > u_1$ and $h_2 > u_2$ can be represented as follows:

u_1		r_1		h_1
□	------------------------------------	□	-----------------	□
u_2		r_2		h_2

From the proportionality

[B.1] $$\frac{r_1 - u_1}{r_2 - u_2} = \frac{h_1 - u_1}{h_2 - u_2},$$

it follows for the linear transformation, that

[B.2] $$r_1 \rightarrow r_2 = u_2 + \frac{h_2 - u_2}{h_1 - u_1} \cdot (r_1 - u_1).$$

As the reader may verify, this formula also holds in the case $h_1 < u_1$ and/or $h_2 < u_2$.

The formula [B.2] can also be applied to the linear transformation of **mean values** *m*:

[B.3] $$m_1 \rightarrow m_2 = u_2 + \frac{h_2 - u_2}{h_1 - u_1} \cdot (m_1 - u_1).$$

For the corresponding **standard deviation** *s*, the transformation formula is

[B.4] $$s_1 \rightarrow s_2 = \left|\frac{h_2 - u_2}{h_1 - u_1}\right| \cdot s_1 .$$

This is based on the fact that, when X is a random variable and a and c are constants, then

[B.5] $\text{var}\{aX+c\} = a^2 \text{ var}\{X\}$, so

[B.6] $s\{aX+c\} = a \cdot s\{X\}$.

Example:
Consider the transformation of m_1 = 2,15 and s_1 = 0,64 as the results of measurements obtained using the above 4-step rating scale

1	2	3	4
'very happy',	'fairly happy',	'not too happy',	'unhappy'.

Note that *in this example* the primary scale is a reversed scale !

We want to transform those statistics onto an [0, 10] scale, so with $u_2 = 0$ and $h_2 = 10$; this is the conventional secondary scale in studies of happiness in nations. In this case the corresponding transformation formulae are:

[B.7] $m_1 \rightarrow m_2 = \dfrac{10}{h_1 - u_1} \cdot (m_1 - u_1)$ and

[B.8] $s_1 \rightarrow s_2 = \dfrac{10 \cdot s_1}{|h_1 - u_1|}$.

Inserting $h_1 = 1$, $u_1 = 4$, $m_1 = 2{,}15$ and $s_1 = 0{,}64$ respectively results in the values $m_2 = 6{,}17$ and $s_2 = 2{,}13$ for the corresponding statistics on the [0;10] scale.

Continuous distribution

It was described in section 2.7 how incidentally the happiness distribution is considered to be a continuous one. In the most frequently occurring approach, this implies that the scale is extended at both end points (Fig. 2.4).

The obvious consequences of the rescaling are that in this case two of the three underlying assumptions for the linear transformation of happiness ratings have to be modified from

(b) $u_1 \rightarrow u_2$, and

(c) $h_1 \rightarrow h_2$

to:

(b) $u_1 - ½ \rightarrow u_2$, and

(c) $h_1 + ½ \rightarrow h_2$ respectively in case $u_1 < h_1$.

In the case of a reversed primary scale ($u_1 > h_1$) the assumptions (b) and (c) need to be modified accordingly.

An example can be found in Ventegodt (1995) , who applies a five-point primary scale and a [0, 100] secondary one, resulting in the transformation
{1, 2, 3, 4, 5} → {10, 30, 50, 70, 90}.

Appendix C

The relationship between the Gini coefficient and the mean pair distance

Let $\{ j = 1(1)\ k \in \mathbb{N}\}$ be the k ratings on an equidistant happiness rating scale with absolute frequencies $\{n_j\}$ in a sample with size $N = \sum n_j$. The average value is $m := \sum j n_j / N$. For the j-th happiness values, we define two cumulative distribution functions as mentioned in section 2.6 :

[C.1] $F(j) := \frac{\sum_{i=1}^{j} n_i}{\sum_{i=1}^{k} n_i}$, and

[C.2] $\Phi(j) := \frac{\sum_{i=1}^{j} i n_i}{\sum_{i=1}^{k} i n_i}$, while

[C.3] $0 \leq F(j) \leq \Phi(j) \leq 1$ for $j = 1(1)k$.

Each of the happiness values $\{j\}$ can be represented as a point in a diagram with abscissa $F(j)$ and ordinate $\Phi(j)$. In Fig. C.1 P and Q are these points for the happiness values $j-1$ and j respectively, whereas the point (1,1) represents the k-th happiness value. In order to obtain a Lorenz curve we connect the point (0,0) and consecutive 'happiness points' by straight line segments, which together form the Lorenz 'curve' as a broken line. In section 4.9 we discuss the question whether this is admissible, but for the moment we assume that this is justified.

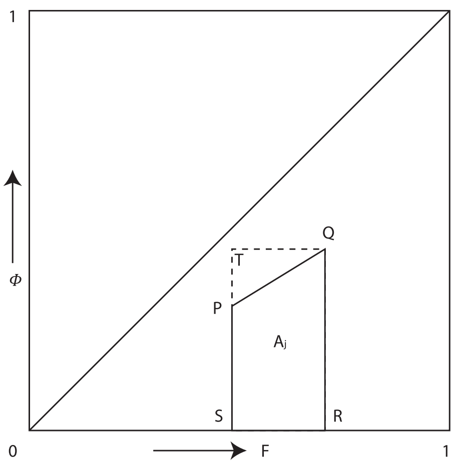


Fig. C.1 Two points P and Q of the Lorenz curve

The Gini coefficient, denoted *GC*, is defined in section 4.2.7 as the ratio of the area between the Lorenz curve and the diagonal over that curve and the total area below that diagonal, the latter area being equal to ½, so the area below the Lorenz 'curve' equals ½ (1– GC); in our case it is the sum of *k* trapezoids.

We will now consider the area A_j of the *j*-th trapezoid below the Lorenz curve between the two points P and Q corresponding to the happiness value $i = j - 1$ and $i = j$ respectively as the difference between a rectangle QRST and a triangle PQT with a common base $\underline{QT} = \underline{RS} = n_j / N$.

$$A_j = (N^{-1}n_j)\left(\sum_{i=1}^{j} in_i - \frac{1}{2}jn_j\right)\left(\sum_{i=1}^{k} in_i\right)^{-1} = \frac{1}{2}(N^2m)^{-1}\left(2n_j\sum_{i=1}^{j} in_i - jn_j^2\right), \text{ so}$$

[C.4] $$(N^2m)(1-G) = \sum_{j=1}^{k} 2N^2mA_j = \sum_{j=1}^{k}\left(2n_j\sum_{i=1}^{j} in_i - jn_j^2\right)$$

For *k=4* the right hand member of [C.4] can be represented schematically as :

$+2n_1n_1 - n_1^2$			
$+2n_2n_1$	$+4n_2n_2 - 2n_2^2$		
$+2n_3n_1$	$+4n_3n_2$	$+6n_3n_3 - n_3^2$	
$+2n_4n_1$	$+4n_4n_2$	$+6n_4n_3$	$+8n_4n_4 - 4n_4^2$

An equivalent, but more convenient representation, that we shall denote S_1, is

$+1n_1^2$	$+1n_1n_2$	$+1n_1n_3$	$+1n_1n_4$ +
$+1n_2n_1$	$+2n_2^2$	$+2n_2n_3$	$+2n_2n_4$ +
$+1n_3n_1$	$+2n_3n_2$	$+3n_32$	$+3n_3n_4$ +
$+1n_4n_1$	$+2n_4n_2$	$+3n_4n_3$	$+4n_4^2$

The mean pair distance *MPD* is defined as :

[C.5]

$$MPD := [N(N-1)]^{-1}\sum_{i=1}^{k}\sum_{i=1}^{k}|i-j|n_in_j \;\Rightarrow\; [N(N-1)]MPD = \sum_{i=1}^{k}\sum_{i=1}^{k}|i-j|n_in_j$$

The right-hand double sum in [C.5] can be represented in a way similar to that of S_1:

$+0.n_1^2$	$+1n_1n_2$	$+2n_1n_3$	$+3n_1n_4$ +
$+1n_2n_1$	$+0.n_2^2$	$+1n_2n_3$	$+2n_2n_4$ +
$+2n_3n_1$	$+1n_3n_2$	$+0.n_3^2$	$+1n_3n_4$ +
$+3n_4n_1$	$+2n_4n_2$	$+1n_4n_3$	$+0.n_4^2$ +

Now we define half its value as

[C.6] $S_2 := ½N(N\text{-}1)\cdot MPD$,

and represent it by dropping all terms below the main diagonal as

$+0.n_1^2$	$+1n_1n_2$	$+2n_1n_3$	$+3n_1n_4$+
$+0.n_2n_1$	$+0.n_2^2$	$+1n_2n_3$	$+2n_2n_4$+
$+0.n_3n_1$	$+0.n_3n_2$	$+0.n_3^2$	$+1n_3n_4$+
$+0,n_4n_1$	$+0.n_4n_2$	$+0.\ n_4n_3$	$+0.n_4^2$

The sum

[C.7] $S := S_1+S_2$

can be obtained by addition of the corresponding terms, giving

$+1n_1^2$	$+2n_1n_2$	$+3n_1n_3$	$+4n_1n_4$+
$+1n_2n_1$	$+2n_2^2$	$+3n_2n_3$	$+4n_2n_4$+
$+1n_3n_1$	$+2n_3n_2$	$+3n_3^2$	$+4n_3n_4$+
$+1n_4n_1$	$+2n_4n_2$	$+3n_4n_3$	$+4n_4^2$.

Columnwise summation of this polynomial and generalizing for k results in :

[C.8] $S := S_1+S_2=1Nn_1+2Nn_2+ \dots +kNn_k=N(Nm)=N^2m$.

From $S_2=S–S_1=N^2m–N^2m(1–G) =N^2mG$, and

[C.6] $S_2=½N(N– 1)\cdot MPD$,

it follows the relationship between the Gini coefficient GC and the mean pair distance *MPD*, provided an equidistant rating scale has been applied:

[C.9] $GC := \left(\frac{N-1}{N}\right)\left(\frac{1}{2m}\right)MPD \approx \frac{MPD}{2m}$ for larger samples.

Appendix D
Computation of the Inequality-Adjusted Happiness Index

We will define the exact position of the most miserable society as has been defined in section 5.2 in the *m-s*-diagram. To this end, we have to quantify the weights given to the both views on happiness.

Let these be w_E and w_U for the egalitarian and utilitarian view respectively, where $0 \leq (w_E , w_U) \leq 1$ and $w_E + w_U = 1$.

The point W, that represents this society in Fig. D.1, can be obtained by partitioning the arc TL into two parts according to these weights. Since arc (TL) = π/2 (expressed in radians ; π/2 radians = 90°), arc (LW) = $w_E \cdot \pi/2$ and arc (WT) = $w_U \cdot \pi/2$. The angle LMW equals $w_E \cdot \pi/2$ and the angle WHL (denoted φ) has half its value = $w_E \cdot \pi/4$.

If the assumption $w_E = w_U = 0{,}5$ is made, then $\varphi = \pi/8$ (i.e. 22°30′). This value of φ has been adopted throughout this appendix, whenever a numerical value has been substituted. For unequal weights, the value of φ in the various formulae has to be adjusted accordingly.

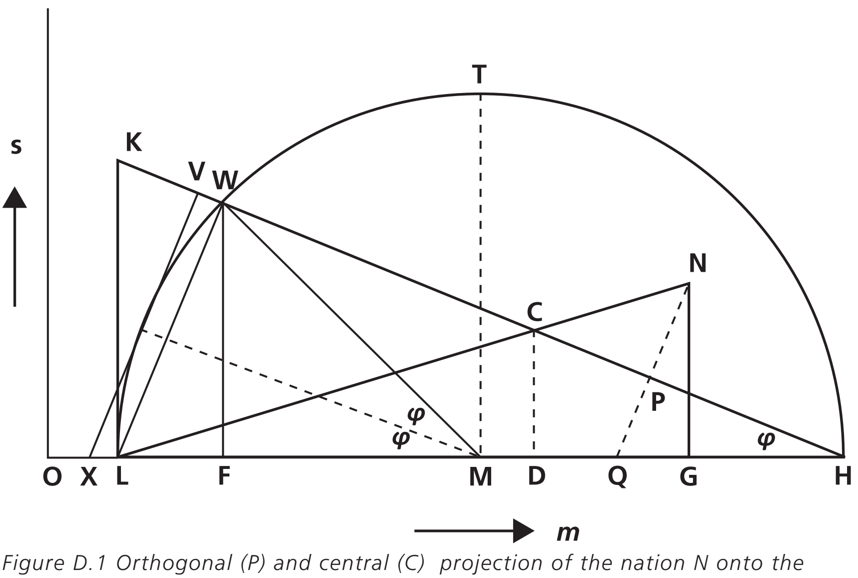


Figure D.1 Orthogonal (P) and central (C) projection of the nation N onto the IAH-axis WH.

We define $b := |h - u|$, where h is the rating corresponding to the most happy situation and u that of the unhappiest situation. For a $[1, k]$ k-point rating scale $b = k - 1$. Moreover we apply the notation $\underline{LH}$ for the length of the line segment LH, so $\underline{LH} = \underline{HL} = b$ and $u = \underline{OL}$, where O is the origin of the m-s-diagram.

Orthogonal projection

If in the case of the orthogonal projection the point W is chosen as the one in which the Inequality-Adjusted Happiness *IAH* has a zero value, the calculation of *IAH* is very simple.

The cartesian coordinates of W in the above m-s-diagram are:

[D.1] $m_W = u + \underline{LF} = u + \underline{LM} - \underline{FM} = u + ½\, b(1 - \cos 2\varphi)$, and

[D.2] $s_W = \underline{WF} = ½\, b\sin(2\varphi)$.

For $u = 0$, $b = 10$ and $\varphi = \pi/8$, $m_W = 1{,}46$ and $s_w/m_w = 2{,}41$. Therefore, in this case for all points (m, s) inside the semicircle segment LW $m < m_w = 1{,}46$ and $s/m > s_w/m_w = 2{,}41$.

Since in Fig. D.1 WL and NQ are parallel (both are perpendicular to HW), $IAH_o = (\underline{WP} / \underline{WH})$ x100 = $(\underline{LQ} / \underline{LH})$ x100. If the coordinates of N are m and s respectively, $\underline{LQ} = \underline{LG} - \underline{QG} = (m - u) - s.\tan\varphi$, so

[D.3] $IAH_o = 100\,(m - u - s\cdot\tan\varphi)/b$,

where $\tan\varphi = 0{,}414$ for "equal weights". The index "o" in IAH_o indicates that the projection is orthogonal, whereas IAH_c will indicate that central projection has been applied.

If, however, one sticks to the condition that, for any theoretically possible (m, s) combination $0 \leq IAH \leq 100$, it is the point V that corresponds to $IAH_o = 0$. In this case $IAH_o = (\underline{VP} / \underline{VH})$ x 100 = $(\underline{XQ} / \underline{XH})$ x 100, X being the point of intersection of the m-axis and the tangent through V to the semicircle. Now

$\underline{XQ} = \underline{XL} + \underline{LQ} = \underline{VW}/\cos\varphi + (m - u - s\cdot\tan\varphi) = (\underline{LM} - \underline{LM}\cos\varphi)/\cos\varphi + (m - u - s\cdot\tan\varphi)$

$$\underline{XQ} = \frac{1}{2}b\left(\frac{1}{\cos\varphi} - 1\right) + (m - u - s\cdot\tan\varphi)$$, and

$$\underline{XH} = \frac{1}{2}b\left(\frac{1}{\cos\varphi} + 1\right)$$

Therefore

[D.4] $$IAH_o = 100\cdot\frac{m - u - s\tan\varphi + \frac{1}{2}b\left(\frac{1}{\cos\varphi} - 1\right)}{\frac{1}{2}b\left(\frac{1}{\cos\varphi} + 1\right)}$$

Substitution of $\cos\varphi = 0{,}924$ and $\tan\varphi = 0{,}414$ results in

[D.5] $IAH_o = 96{,}0(m - u - 0{,}414 \cdot s)/b + 3{,}96$.

As one might have expected, IAH_o is obtained by linear transformation of *m* and *s*, irrespective of the choice of the zero point. Note that in this context the term "linear transformation" has a meaning that is not entirely identical to the one in the case of "linear transformation of happiness scores" as used in Appendix B.

Central projection

For the central projection, the coordinates of C (m_C, s_C) as point of intersection of HW and NL follow from :

$b = \underline{LD} + \underline{DH}$,

$$\frac{\underline{LD}}{\underline{CD}} = \frac{\underline{LG}}{\underline{NG}} = \frac{m_N - u}{s_N}$$

$\underline{CD} = s_C$,

$\underline{DH} = s_C / \tan\varphi$, and

$$\frac{m_C - u}{s_C} = \frac{m_N - u}{s_N} \quad .$$

The result is :

[D.6] $$s_C = \frac{b}{\frac{m_N}{s_N} - \frac{u}{s_N} + \frac{1}{\tan\varphi}}$$

In this formula, it is assumed that $s_N > 0$.

In the case where W is selected as the point with $IAH_c = 0$,

$IAH_c = (\underline{WC} / \underline{WH}) \times 100 = 100 \times (\underline{WF} - \underline{CD}) / \underline{WF}$

$$IAH_c = 100 \cdot \frac{\frac{1}{2} b \sin 2\varphi - \frac{b}{\frac{m_N}{s_N} - \frac{u}{s_N} + \frac{1}{\tan\varphi}}}{\frac{1}{2} b \sin 2\varphi}$$

[D.7] $$IAH_c = 100 - \frac{200}{\left(\frac{m_N}{s_N} - \frac{u}{s_N} + \frac{1}{\tan\varphi}\right) \cdot \sin 2\varphi}$$

Substitution of $\tan\varphi = 0{,}414$ and $\sin(2\varphi) = 0{,}707$ gives :

[D.8] $$IAH_c = 100 - \frac{283}{\frac{m_N}{s_N} - \frac{u}{s_N} + 2{,}41} \qquad s_N > 0$$

$$= 100 \qquad s_N = 0$$

If however, it is required that $0 \leq IAH \leq 100$, it is the point K that corresponds to $IAH_c = 0$, K being the intersection point of the *IAH*-axis and the vertical tangent to the semicircle.

In this case :

IAH_c = ($\underline{KC}$ / $\underline{KH}$) x 100 = [($\underline{KL}$ – $\underline{CD}$)/$\underline{KL}$)] x 100.

$$IAH_c = 100 \cdot \frac{b\tan\varphi - \frac{b}{\frac{m_N}{s_N} - \frac{u}{s_N} + \frac{1}{\tan\varphi}}}{b\tan\varphi}$$, so

[D.9] $$IAH_c = 100 - \frac{100}{1 - \frac{u\tan\varphi}{s_N} + \frac{m_N}{s_N} \cdot \tan\varphi}$$

Substitution of $\tan\varphi = 0{,}414$ gives :

[D.10] $$IAH_c = 100 - \frac{100}{1 - 0{,}414\frac{u}{s_N} + 0{,}414\frac{m_N}{s_N}}$$

In the case $u = 0$, IAH_c is a monotonically increasing function of the ratio *m/s* only. Comparison of the formulae [D.3], [D.5], [D.8] and [D.10] reveals that *b* occurs in the formulae for IAH_o, but not in those for IAH_c.
However, the suggestion that the value of the former one is dependent on the number of possible ratings of the happiness measuring scale, whereas the latter is not, is false. In the formulae for IAH_o, *b* acts as a scaling factor for both *m – u*

and *s*. In the case of central projection, there is an 'internal scaling', since both ($m - u$) and *s* are measured on the same scale and only their *ratio* occurs in the formulae for the index.

NOTE: some researchers prefer to apply a 'reversed scale', i.e. a scale at which the most happy situation corresponds to the lowest ranking number *h*; in that case $h \leq m \leq u$.

If one also wants to include these cases, in the formulae [D.3] to [D.10] inclusive, the difference $m_N – u$ must be replaced with its absolute value $| m_N – u |$. For formula [D.10] this generalization results in

[D.11] $$IAH_c = 100 - \frac{100}{1 + 0{,}414\frac{|m_N - u|}{s_N}} .$$

Appendix E
The calculation of the parameter estimates according to the HSIA

We have devised an EXCEL computer programme for the calculation of the parameter estimates in the context of the Happiness Scale Interval Study, which is briefly described en demonstrated in this appendix. The programme exists in five variaties, one for each value of $k = 3(1)7$. A workbook is prepared for each case separately. For each application of this case, a new copy of this workbook is used. The first worksheet is identical for each application using the same case, the other three are different.

In this appendix we included a copy of a part of a workbook, so note the difference between the workbook as described and the parts of it that are depicted in this appendix. We shall demonstrate the programme with an example, using case B, in table 7.11 and applied to the first of the four surveys (1981 in The Netherlands).

The workbook consists of four worksheets. The first sheet has been designed for the data entry in the construction phase and the second is destined for the application phase and returns all final results as the output. The other two worksheets of the programme are used to perform the various calculations. These two have no input cells and are rather extensive; therefore they are not depicted in this appendix. Two different copies of the first worksheet and one of the second worksheet have been included in this appendix.

All information about the case, the language, the series, the happiness measure etc. is stored in the upper part of the first worksheet. The bottom part is filled with the observational data, where each row contains the total judgement of one judge about the cut points of that specific case. This first copy of the first worksheet is depicted as E1 of this appendix.

The programme tests for each row/judgement whether or not the skipping rules as defined in section 7.2.3 apply. In the example, two judgements have been indicated for skipping (Worksheet E.1). The skipped judgements are transferred manually to the very bottom part of this worksheet. They remain visible, but are ignored in further calculations. If the percentage of skipped judgements justifies, a message about the case in total is delivered. The final result after skipping is presented in the second copy of the first worksheet, which is printed as Worksheet.E.2. Although the total number of judgments per case can adopt much larger values than 30 as in this example, we have removed most of the empty rows here for trivial reasons. The programme computes all statistics listed in section 7.2.4 and the results are stored in one of the two calculation worksheets (not depicted here).

When the results of the construction phase are applied in a survey to a sample of happiness ratings from a nation, at least in the same language and using the same item, the sample results have to be entered into the application worksheet. The data to be entered are (a) the frequency distribution in the sample, either as absolute or as relative frequencies per category, (b) the effective sample size, i.e. ignoring all "don't know/no answer" responses and (c) the source of the above data for documentation purposes. A copy of this second worksheet is depicted in this appendix as E3.

On the basis of this input and the results of the construction phase, the programme computes the various statistics that act as estimates of the relevant population parameters. This is done according to the procedures and formulae specified and derived in section 7.4.2 and in Appendix F sections F.1 and F.4.

In the application phase, the programme computes:

(a) the estimated mean population happiness value on a [0, 10] scale as Veenhoven-Kalmijn statistic according to section 6.6

(b) 95 % confidence limits for the population mean hapiness

(c) the rounding interval for this estimate according to the rounding rules in Appendix G (Rounding observations and statistics)

(d) the Veenhoven-Kalmijn (i.e. adjusted) within-population standard deviation

(e) the 'unadjusted standard deviation', to be used to quantify the within-nation inequality if the happiness distribution within that nation is considered as discrete polytomous (with k =5)

(f) estimates of both shape parameters of the best fitting beta distribution in the fully continuous model

(g) the estimated mean, standard deviation and skewness of this beta distribution
(h) for comparison reasons: the estimated population mean and standard deviation according to the traditional method as described in e.g. section 6.1 together with a warning not to apply this method in this case.

For more detailed information on the programme, the reader is referred to: http://www.worlddatabaseofhappiness.eur.nl/scalestudy/scale_fp.htm

Worksheet E1

INPUT

ESTIMATED MEAN VALUE AND 'ADJUSTED' STANDARD DEVIATION OF A POPULATION

© W.M. Kalmijn 2010-04-29

Nation: NL - The Netherlands

Judges public: Students Erasmus University Rotterrdam NL

Year: 2006

Language: Dutch

Number of participating judges (max. 400) = 32
(before correction for invalid judgements)

WDH Measure of happiness code: O-HL/u/sq/v/4/a

HSIA case code: dutch3; case 4

Number of possible ratings (k) = 4

Identification number of first judge: 1101

Internal code number SCP-: 105

Lead question (verbatim): Hoe gelukkig bent u, alles bijeengenomen ?

Translation into (US) English: Taking all things together, would you say you are.....?

# 'zero width'		Response category label (verbatim):	Translation into (US) English:	0
1	4	heel gelukkig	very happy	**4**
	3	tamelijk gelukkig	quite happy	**3**
	2	niet zo gelukkig	not very happy	**2**
4	1	helemaal niet gelukkig	not at all happy.	**1**

accepted judgements

COMMENTS:

==

ESTIMATED CUT POINT POSITION VALUES.

Accepted judgements:

n = 32   100 % accepted, k= 4 h-max= 10 d = 0,50 (minimum distance)

After skipping 0 **declared invalid judgements !**

		Test for native	Test for seriousl	difficulty	**and**	1: boundary between subintervals 4 and 3	2: boundary between subintervals 3 and 2	3: boundary between subintervals 2 and 1		
	1	1	1			7,8	5,4	1,9	0	0
	2	1	1			7,9	3,0	1,1	0	0
	3	1	1			7,9	4,0	1,7	0	0
	4	1	1			7,9	5,5	1,9	0	0
	5	1	1			7,9	5,9	4,0	0	0
	6	1	1			7,8	4,9	2,0	0	0
	7	1	1			6,9	5,4	4,0	0	0
	8	1	1			7,9	5,5	4,0	0	0
	9	1	1			8,5	4,4	0,4	0	0
	10	1	1			8,9	4,9	0,0	0	0
	11	1	1			10,0	7,9	4,8	0	0
	12	1	1			8,5	6,3	2,2	0	0
SKIP !!	**13**	2	1			4,3	0,5	0,0	1	1
	14	1	1			8,5	3,8	1,1	0	0
	15	1	1			9,3	6,9	3,8	0	0
	16	1	1			8,4	5,9	4,1	0	0
	17	1	1			8,8	4,9	2,4	0	0
	18	1	1			7,9	4,9	1,9	0	0
	19	1	2			7,7	5,4	2,9	0	0
	20	1	2			7,7	5,2	2,4	0	0
	21	1	2			8,4	6,4	3,0	0	0
SKIP !!	**22**	2	1			6,0	0,5	0,0	1	1
	23	1	2			8,0	6,0	3,0	0	0
	24	1	2			6,6	5,6	3,9	0	0
	25	1	2			8,4	5,0	3,1	0	0
	26	1	2			7,9	3,9	0,4	0	0
	27	1	2			7,9	5,9	2,0	0	0
	28	1	2			9,0	5,9	5,0	0	0
	29	1	2			7,8	5,5	1,9	0	0
	30	1	1			8,4	7,0	5,1	0	0
	31	1	2			8,9	3,7	0,0	0	0
	32	1	2			9,7	4,2	0,5	0	0

SKIPPED OBSERVATIONS: copied

INPUT after skipping

ESTIMATED MEAN VALUE AND 'ADJUSTED' STANDARD DEVIATION OF A POPULATION

© W.M. Kalmijn 2010-04-29

Nation: NL - The Netherlands

Language: Dutch

WDH Measure of happiness code: O-HL/u/sq/v/4/a

Number of possible ratings (k) = 4

Judges public: Students Erasmus University Rotterrdam NL

Year: 2006

Number of participating judges (max. 400) = 32
(before correction for invalid judgements)

HSIA case code: dutch3; case 4

Identification number of first judge: 1101

Internal code number SCP-: 105

Lead question (verbatim):	Translation into (US) English:
Hoe gelukkig bent u, alles bijeengenomen ?	Taking all things together, would you say you are.....?

# 'zero width'		Response category label (verbatim):	Translation into (US) English:	0
1	4	heel gelukkig	very happy	**4**
	3	tamelijk gelukkig	quite happy	**3**
	2	niet zo gelukkig	not very happy	**2**
2	1	helemaal niet gelukkig	not at all happy.	**1**

accepted judgements

COMMENTS:

==

ESTIMATED CUT POINT POSITION VALUES.

Accepted judgements:

n = 30   94 % accepted, k= 4 h-max= 10 d = 0,50 (minimum distance)

After skipping 2 **declared invalid judgements !**

	Test for native	Test for seriousl	Test for difficulty	and	1 4 3	2 boundary between subintervals 3 2	3 2 1		
1	1	1			7,8	5,4	1,9	0	0
2	1	1			7,9	3,0	1,1	0	0
3	1	1			7,9	4,0	1,7	0	0
4	1	1			7,9	5,5	1,9	0	0
5	1	1			7,9	5,9	4,0	0	0
6	1	1			7,8	4,9	2,0	0	0
7	1	1			6,9	5,4	4,0	0	0
8	1	1			7,9	5,5	4,0	0	0
9	1	1			8,5	4,4	0,4	0	0
10	1	1			8,9	4,9	0,0	0	0
11	1	1			10,0	7,9	4,8	0	0
12	1	1			8,5	6,3	2,2	0	0
13	1	1			8,5	3,8	1,1	0	0
14	1	1			9,3	6,9	3,8	0	0
15	1	1			8,4	5,9	4,1	0	0
16	1	1			8,8	4,9	2,4	0	0
17	1	1			7,9	4,9	1,9	0	0
18	1	2			7,7	5,4	2,9	0	0
19	1	2			7,7	5,2	2,4	0	0
20	1	2			8,4	6,4	3,0	0	0
21	1	2			8,0	6,0	3,0	0	0
22	1	2			6,6	5,6	3,9	0	0
23	1	2			8,4	5,0	3,1	0	0
24	1	2			7,9	3,9	0,4	0	0
25	1	2			7,9	5,9	2,0	0	0
26	1	2			9,0	5,9	5,0	0	0
27	1	2			7,8	5,5	1,9	0	0
28	1	1			8,4	7,0	5,1	0	0
29	1	2			8,9	3,7	0,0	0	0
30	1	2			9,7	4,2	0,5	0	0
								0	0

SKIPPED OBSERVATIONS:

									copied
SKIP !!	1	2	1		4,3	0,5	0,0	1	1
SKIP !!	2	2	1		6,0	0,5	0,0	1	1

OUTPUT

ESTIMATED MEAN VALUE AND 'ADJUSTED' STANDARD DEVIATION OF A POPULATION

This programme computes the "Veenhoven-Kalmijn statistics", i.e. the estimated mean value and standard deviation of the happiness distribution in a population based on the assumption of a homogeneous distribution of the happiness within each of the 4 below intervals separately.
Moreover, the parameters of the 'best fitting' beta distribution are estimated.
For comparison reasons only, the programme also calculates the estimated mean value and standard deviation of the population on a [0, 10] scale according to the traditional method.

APPLICATION (including reference) **© W.M. Kalmijn 2010-04-29**

WVS-1; 1981.

Measure code: O-HL/u/sq/v/4/a Number of different possible ratings: (categories) 4 SCP- 105

Lead question (verbatim):	Translation into (US) English:	dutch3; case 4
Hoe gelukkig bent u, alles bijeengenomen ?	Taking all things together, would you say you are.....?	

	Response category label (verbatim):	Translation into (US) English:	# Accepted judgements (*n*):
4	heel gelukkig	very happy	30
3	tamelijk gelukkig	quite happy	Nation of the judges:
2	niet zo gelukkig	not very happy	NL - The Netherlands
1	helemaal niet gelukkig	not at all happy.	

Effective sample size (*N*) 1.502

ENTER:"**1**" in case of absolute or "**2**" in case of relative frequencies: 2

		observed frequencies						contribution to		
	Response category label (translated)	absolute	percentage	#	lower boundary	mid value	upper boundary	average	variance unadj.	variance adj.
4	very happy		32,3	485	8,24	9,12	10,00	2,94	0,89	0,98
3	quite happy		65,4	982	5,31	6,77	8,24	4,43	0,30	0,77
2	not very		2,2	33	2,48	3,90	5,31	0,09	0,28	0,29
1	not at all		0,2	3	0,00	1,24	2,48	0,00	0,08	0,08
	SUM	1.504	100,1							
		1.504								

		2 decimals
Rounding interval for the mean	0,01	
Upper 95% confidence limit for the mean		7,68
average value in sample	7,5	7,46
Lower 95% confidence limit for the mean		7,23
'adjusted' variance	2,1	2,12
'adjusted' standard deviation	1,5	1,46
cf 'unadjusted' standard deviation	1,2	1,25
		df
construction component	0,011	29
application component	0,0014	>999
variance of the mean	0,012	37
standard error of the mean	0,110	1502,5

COMMENTS / CALCULATIONS:

PARAMETERS OF THE 'BEST FITTING' BETA DISTRIBUTIOI

		on that basis estimated		2 decimals
estimated alpha (α)	10,56	distribution mean	7,6	7,62
estimated beta (β)	3,29	standard deviation	1,1	1,10
		skewness (alpha-3)	-0,60	(skew to the left)

For comparison reasons only: **PARAMETER ESTIMATES ACCORDING TO THE TRADITIONAL APPROACH:**

(transformation in case of a 4-point scale is strongly dissuaded)

Average value = 7,66 eviation = 1,71

Appendix F
Parameter estimation Happiness Scale Interval Approach

Formulae have been derived and collected in this appendix, that are needed for the parameter estimation according to the Happiness Scale Interval Approach.

Contents of this Appendix

F.1. Mean and variance of the distribution of H-values

Two options are available for the probability distribution of the random variable H over the closed interval $[0, 10] \subset \mathbb{R}$, one in which this distribution is assumed to be discrete and the other one where it is continuous.

Discrete case

In the discrete case, the variable H can adopt only a discrete natural number ($k \in \mathbb{N}$) of different values $\{h_j \mid j = 1(1)k\}$ and will do so with relative frequencies $\{f_j \mid j = 1(1)k\}$, while :

[F.1] $$\sum_{j=1}^{k} f_j = 1$$

For this case, the expected value of the random variable H, usually denoted μ, is defined as :

[F.2] $$\mu := \mathcal{E}\{H\} := \sum_{j=1}^{k} h_j f_j$$

The variance of the random variable H, usually denoted σ^2, is defined as

[F.3] $$\sigma^2 := var\,\{H\} := \mathcal{E}(H-\mu)^2 = \mathcal{E}\{H^2\} - [\mathcal{E}\{H\}]^2$$

In this particular case

[F.4] $$\sigma^2 := \mathcal{E}\{H^2\} - [\mathcal{E}\{H\}]^2 = \sum_j h_j^2 f_j - \left[\sum_j h_j f_j\right]^2$$

If we assume the distribution of H to be (i) continuous and (ii) uniform over each interval ($b_{j\text{-}1}$, b_j], then the probability density in the j-th interval

[F.5] $$g(h) = (b_j - b_{j-1})^{-1} f_j \qquad \text{for } 0 \leq b_{j-1} \leq h \leq b_j < 10$$

Now the expected (or mean) value of H equals

[F.6] $$\mathcal{E}\{H\} := \int_0^{10} hg(h)dh = \sum_j \int_{b(j-1)}^{b(j)} hg(h)dh$$

This function can be written as

[F.7] $$\mathcal{E}\{H\} = \sum_j \int_{b(j-1)}^{b(j)} \left(b_j - b_{j-1}\right)^{-1} f_j\, hdh =$$

$$= \sum_j \frac{1}{2}\left(b_j - b_{j-1}\right)^{-1}\left(b_j^2 - b_{j-1}^2\right) f_j = \sum_j \frac{1}{2}\left(b_j + b_{j-1}\right) f_j = \sum_j m_j f_j$$

just as in the discrete case, but with the mid-interval value $m_j := ½(b_{j\text{-}1} + b_j)$ instead of h_j .

The continuous case

The variance in the continuous case can be found from

[F.8] $$\mathcal{E}\{H^2\} := \int_0^{10} h^2\, g(h)dh = \sum_j \int_{b(j-1)}^{b(j)} h^2\, g(h)dh$$

$$= \sum_j \int_{b(j-1)}^{b(j)} \left(b_j - b_{j-1}\right)^{-1} f_j h^2 dh =$$

$$= \sum_j \frac{1}{3}\left(b_j - b_{j-1}\right)^{-1}\left(b_j^3 - b_{j-1}^3\right)f_j = \sum_j \frac{1}{3}\left(b_j^2 + b_j b_{j-1} + b_{j-1}^2\right)f_j =$$

$$= \sum_j \frac{1}{3}\left(b_j^2 + 2b_j b_{j-1} + b_{j-1}^2 - b_j b_{j-1}\right)f_j = \frac{4}{3}\sum_j m_j^2 f_j - \frac{1}{3}\sum_j b_j b_{j-1} f_j$$

[F.9] $$var\{H\} := \mathcal{E}\{H^2\} - [\mathcal{E}\{H\}]^2 = \sum_j m_j^2 f_j - \left[\sum_j m_j f_j\right]^2 + \frac{1}{3}\sum_j \left(m_j^2 - b_j b_{j-1}\right)f_j$$

The last term is the difference between the variance in the 'discrete' and the 'continuous approach'. It is necessarily positive, since it can also be written as the sum of a number of squares including at least one nonzero one :

[F.10] $$\frac{1}{3}\sum_j \frac{1}{4}\left(b_j^2 + 2b_j b_{j-1} + b_{j-1}^2 - 4b_j b_{j-1}\right)f_j =$$

$$= \frac{1}{3}\sum_j [\frac{1}{4}(b_j^2 - 2b_j b_{j-1} + b_{j-1}^2)]f_j = \frac{1}{12}\sum_j \left(b_j - b_{j-1}\right)^2 f_j > 0$$

so the uniform distribution approach necessarily results in a higher variance. This is not surprising, since in this case variability within the different ratings is added to the variability between these ratings.

F.2. Comparison of the estimates from different methods

In this section, we shall compare the estimates of the mean and the variance as obtained according to the three different procedures discussed in section 6.4:

(c) In model III (cf section 6.5) , it is assumed that all respondents giving the same response R_j are equally happy and have the same *H*-value, for which the MIV of the *j*-th interval is the obvious one to be selected. These *k* responses are the only ones available, not only for the sample members, but also in the population as a whole. In other words, the population probability distribution of *H* is assumed to be discrete with only *k* possible values, just as in the 'traditional' approach.

(d) The variable *H* is assumed to be continuous and has a distribution which is uniform over each of the *k* intervals as is described in section F.1 of this Appendix.

(e) The variable *H* is assumed to be a continuous variable with a beta distribution. Estimates of the two model parameters α and β are calculated.

Subsequently, estimates of the mean and the variance of the distribution are calculated on the basis of these estimates of α and β.

An important property of any estimator is whether it is biased or not. If θ is a parameter or a function of one or more parameters of a probability distribution of some random variable, and is estimated by a statistic $\hat{\theta}$, then the bias of $\hat{\theta}$ with expectation $E(\hat{\theta})$ is defined as the difference $E(\hat{\theta})-\theta$, where θ is either a scalar or a vector, and θ will be accordingly. It should be emphasized that a bias is defined only if the distribution of the statistic is known and that it depends on which type of probability distribution is adopted for the random variable. Hence the same statistic, which is an unbiased estimator in model III and/or IV is not necessarily unbiased in e.g. model V.

Model III: the discrete polytomous approach

In the traditional approach, it is very unusual to specify the probability distribution in the population explicitly. Implicitly, the situation in the population is assumed to be structurally identical to that of the sample, but with a larger size only. This means that this probability distribution is assumed to be a discrete polytomous distribution with 2*k* parameters, *k* for the probabilities and *k* for the mid-interval values, 2(*k*-1) of which parameters being independent. The parameters are estimated as the *k* relative frequencies in the sample. In that case the sample mean is an unbiased estimator of the mean happiness of the population probability distribution. The second moment about the mean of the sample is made an unbiased estimator of the population variance by the application of Bessel's correction, i.e. by replacing the denominator *n* with *n*–1. Its square root underestimates the value of the population standard deviation systematically, but since this estimator is consistent, usually the sample size is sufficiently large to neglect this negative bias.

Model IV: the semi-continuous approach

If the model IV is adopted, the sample average value is an unbiased estimator of the population mean value, just as under model III, since for each interval, the choice between III and IV does not affect its contribution (see section 6.1). This conclusion does not only hold under the assumption of a uniform distribution, but for any model that assumes a distribution of *H* that within each of the *k* intervals is symmetrical with respect to its MIV. Asymmetry in the distributions within intervals, gives rise to a different estimate of the mean; this is the case if a beta distribution is chosen as in the model V; see below. However, in model IV, the population variance σ^2 is smaller than that according to model V, as will be demonstrated below.

Model V: the beta distribution

The beta distribution has been adopted as a model for a happiness variable with a fully continuous distribution. The skewness of a beta distribution is proportional to the difference $\beta - \alpha$ of the parameters of this distribution (see Appendix H.5). In this section, we confine ourselves to the case $\alpha, \beta > 1$, i.e. to unimodal distributions with $g(0) = g(10) = 0$.

If $\alpha > \beta$, the distribution is "skewed to the left" (negative skewness). In this case, for a beta distribution on the interval [0, 10]

[F.11] 5 < *mean* < *median* < *mode* < 10

The mean of a semi-continuous variable H on the interval [0, 10] is defined as

[F.12] $$\mu := \mathcal{E}\{H\} := \int_0^{10} hg(h)dh = \sum_j \int_{b(j-1)}^{b(j)} hg(h)dh$$

so the contribution of the interval $(h, h+dh]$ to the mean is proportional to both h and $g(h)$.

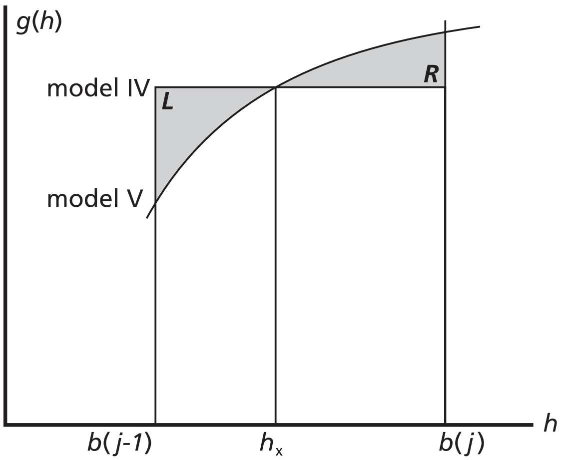


Fig. F.1. Bias of estimators in model IV and V

Consider, in Fig.F.1, an interval $(b_{j-1}, b_j]$ with $g'(h) > 0$ for all $h \in (b_{j-1}, bj]$, where $g'(h)$ denotes the first derivative of $g(h)$ with respect to h. Under the models IV and V, the area below $g(h)$, which equals Prob$\{b_{j-1} < H \leq b_j\}$, has the same numerical value. Let h_x be the H-value at which the density $g(h)$ has the same value in both models IV and V. Although the two shaded areas *L* and *R* have equal sizes and represent equal proportions of the total hapiness distribution, the contribution to the mean value *m* of *R* in model V is larger than that of *L* in model IV. This will apply to all intervals in which all *H*-values are smaller than the mode. In a similar way, in intervals with all *H*-values larger than the modal one, $g(h) < 0$, but now the effect will act the other way round and this will neutralize the former effect, at least partially. However, when the skewness of the distribution of *H* is negative, the majority of the total distribution, both in terms of interval length and area below the *g*(*h*) curve, will be found at *H*-values less than the mode; hence the former effect will dominate the latter one. So as a net effect, the estimate of the mean value in the case of model V is to be expected to be larger than the one obtained by the application of the model III and IV.

For relatively unhappy communities, in which H has a positively skew distribution,

[F.13] $0 < mode < median < mean < 5$,

and the opposite effect is to be expected, i.e. application of the beta distribution model will result in smaller estimates of the mean value than those obtained from the models III and IV.

Eventually, the consequence is that in the case of the application of model V, the average happiness values of the various nations will cover a wider range.

It has already been explained in Section F.1 that for the variance the application of model IV results in higher estimates than that of model III, the difference being quantified in [F.10]. For a continuous random variable *H* on the interval [0, 10]

[F.14] $$\sigma^2 := \mathcal{E}(h-\mu)^2 = \int_0^{10} (h-\mu)^2\, g(h)dh = \sum_j \int_{b(j-1)}^{b(j)} (h-\mu)^2 g(h)dh$$

We pointed out above that, in the case the model V is adopted, for intervals where $g'(h) > 0$ for all values of *H*, the majority of the distribution in such an interval is found in its right-hand part $(h_x, b_j]$. Unless this interval covers the mean value *m*, this means that – as compared to the model IV – , generally a larger part of this area will be closer to the mean value and therefore will have a smaller contribution to the total variance. A similar situation occurs where $g'(h) < 0$, but this time with the same effect. So eventually the application of model V is expected to result in smaller variance estimates than those found in model IV.

If these conclusions are confirmed by empirical results, the combination of a wider range of the mean hapiness values with smaller standard deviations and standard errors would improve the discriminating power of the measurement process by application of the beta distribution model.

Estimates of the parameters of the beta distribution can be obtained as maximum likelihood estimators (MLE). The likelihood function for this situation can be written as :

[F.15]

$$L(\propto, \beta) = \prod_j \left[Prob\{b_{j-1} < H \leq b_j | \alpha, \beta\}\right]^{n(j)} = [10 \cdot B(\propto, \beta)]^{-N} \prod \left[\int_{b(j-1)}^{b(j)} h^{\propto-1}(10-h)^{\beta-1} dh \right]^{n(j)}$$

and has to be maximized by putting :

[F.16] $$\frac{\partial \log(L)}{\partial \propto} = \frac{\partial \log(L)}{\partial \beta} = 0$$

These MLE of α and β can be been obtained by using iterative procedures using numerical analysis computer programmes. Obviously, this approach is only meaningful if k>3.

Although MLE may be biased, they are always consistent, so in the case of large samples their bias is negligible. However even if the estimates of α and β were unbiased, this does not imply that the estimates of μ and σ^2, as nonlinear functions of these are unbiased or at least also consistent as well. This may justify a further investigation of the bias in the estimates of μ and σ^2, but under the conditions of the measurement of happiness in nations or in similar communities.

A more quantitative approach to the above subject is reported by Tamhane et al. (2002) based on their simulation studies. The authors solve the problems of an ordinal scale along similar lines as presented here. Tamhane et al. also adopt the standard beta distribution for a latent response variable, but one major difference is their proposal to apply ‘internal’ estimates for the cut points, i.e. the cut point values are obtained as MLE from the data collected in the experiments in which the scales are applied. As a dispersion parameter, they introduce a statistic η^2 to obtain estimators of a location and a dispersion parameter that are mutually independent. This η^2 is defined as the ratio of the variance σ^2 and $\mu(1-\mu)$. The denominator is the upper bound of the variance of any random variable with a closed interval [0, 1] as its domain, as is demonstrated in Appendix A. In case of a standard beta distribution $\eta^2 = (1+ \alpha + \beta)^{-1} < 1$. The preferred estimation method of Tamhane et al. is based on statistical efficiency rather than on systematic errors only.

The translation of the main relevant findings of these authors to our situation can be summarized as follows: If the variable *H* has a (non-standard) beta distribution on the interval [0, 10] and the population mean and variance are estimated from the MIV according to the above model III, the results appear to be biased. For $\mu > 5$, the estimated mean values were negatively biased, i.e. the Bias(μ) := E{μ} - $\mu < 0$. For $k = 5$, $N = 30$ and $\mu = 7$, the bias in the mean value is about 0,1 à 0,2, depending on the value of the of the standard deviation; for $N = 60$, the bias is about halved and for $k = 10$, it is doubled. For this bias in case $\mu = 9$, the pattern is somewhat diffuse, but until now, no nations have been found with this high average value. The bias in η^2 seems too capricious to enable a clear simple and conclusion about a bias in the standard deviation.

F.3. Position of "marks" as potential estimators of MIV-values

In some studies, respondents have been asked to connect each of a set of say *k* qualifications with the most adequate point on the [0,10] *H*-line. The question arises whether or not such points can be considered as a (personal) MIV in our model.

We shall denote the set of the model interval boundaries as a vector ***b***, which for typographical reasons will be written as a row vector:

[F.17] $\boldsymbol{b} := [\{b_j | b_{j-1} \leq b_j, j = 0(1)k\}]$,

In this notation b_j is the upper boundary of the *j*-th interval of a continuous $[b_0, b_k]$ scale and dim($\boldsymbol{b}$)=k+1.

For a [0,10] scale $b_0 = 0$ and $b_k = 10$, so the vector ***b*** can be written then as

[F.18] $\boldsymbol{b} := [0, b_1, ..., b_{k-1}, 10]$,

In the same way, we define a vector ***m*** of the MIV with dimension *k*:

[F.19] $\boldsymbol{m} := [\{m_j | j = 1(1)k\}]$

and for the *h*-values connected with the *k* categories:

[F.20] $\boldsymbol{r} := [\{rj| \; r_{j-1} \leq r_j, \; j = 1(1)k\}]$,
r_j being the image of the mapping of the category R_j into $\{h\}$, i.e. the position of the *r*-th mark in the interval [0, 10] as selected by the respondent.
The vectors ***b*** and ***m*** are interconnected by the relation
[F.21] $m_j := ½(b_{j-1} + b_j) \quad j = 1(1)k$.

The question may arise whether there is always a unique relation between the vectors ***b*** and ***r***. The answer is negative, as can be demonstrated by a simple example, using a three-step rating scale. The vector $\boldsymbol{r}_1$ = [3, 7, 9] can act as the ***m*** vector connected to ***b*** = [0, 6, 8, 10], but for the vector $\boldsymbol{r}_2$ = [3, 6, 9] no such ***b*** vector can be found with an ***m*** vector which is equal to $\boldsymbol{r}_2$.

In order to devise a tractable criterion, we define a vector ***c*** with dimension *k*:
[F.22] $\boldsymbol{c} := [\{c_j\} = (-1)^{j+k} \mid j = 1(1)k\}]$
and the scalar product of the vectors ***c*** and ***m***:
[F.23] $S_a = \boldsymbol{c} \, . \, \boldsymbol{m}$
This linear combination of the MIV can also be written as the 'alternating sum'
[F.24] $S_a = m_k - m_{k-1} + m_{k-2} - m_{k-3} + \ldots \pm m_1$.

As the reader can verify easily, substitution of
[F.21] $m_j := ½(b_{j-1} + b_j) \quad j = 1(1)k$
results in
[F.25] $S_a = ½ \, [b_k - (-1)^k \, b_o]$,
irrespective of the values of ***m***, so for a [0, 10] scale, $S_a = 5$ for any ***m*** vector; even the dimension of this vector does not matter.

This condition explains why in the above example the vector $\boldsymbol{r}_2$ = [3, 6, 9] with an alternating sum value 6 cannot act as the ***m*** vector of any conceivable ***b*** vector in case of a [0,10] scale. Obviously the above conclusion also applies to the Thurstone values.

The question may rise whether or not the scale interval approach is also applicable to situations in which the positions of the ratings on a discrete [1, *k*] are available on the basis of either observations or decisions. In section F.5, we shall deal with this question.

F.4. Covariances between mid-interval values

Since $S_a = 5$ for any ***m*** vector, defined in [F.19], the $\{m_j\}$ are linearly dependent, i.e. every MIV can be written as a linear combination of all other ones and a constant (i.c. $b_k = 10$). Then they are also stochastically dependent on each other, since the variance of S_a can be written as

[F.26] $var\{S_a\} = \sum_j \sum_{j'} c_{jj'} .\ cov\left\{m_j,\ m_{j'}\right\}$ $j, j' = 1(1)k$ with

[F.27] $c_j = (-1)^{j+k}$ and $c_{jj'} := c_j c_{j'} = (1)^{j+k}(1)^{j'+k} = (-1)^{j+j'}$.

This double sum [F.26] consists of k variance terms with $j = j'$ and $c_{jj'} = +1$. If all $\{m_j\}$ were mutually stochastically independent, all remaining $k(k\text{-}1)$ terms in this double sum would vanish. Since $S_a = ½b_k$ has a zero variance, this would require that also all m_j have a zero variance.

For the majority of the $k^2\text{-}k$ 'true' covariances, $c_{jj'} = -1$. If k is even, $½k^2$ true covariances have $c_{jj'} = -1$ against $½k^2\text{-}k$ with $c_{jj'} = +1$; for odd k, these numbers are $½(k^2\text{-}1)$ and $½(k\text{-}1)^2$ respectively. It is obvious to expect that adjacent m_j values, which all have $c_{jj'} = -1$, have positive covariances: if a panel member selects a relatively low value for m_2, the remaining scope for m_1 will be limited and a relatively small value for the latter is to be expected in that case. Moreover, it is not surprising that the highest positive correlations are found between adjacent m-values.

In the **scale construction phase**, each judge of a panel of size n has to indicate the positions of the interval boundaries $\{b_j \mid j = 1(1)\ k\text{-}1\}$. From these boundaries, the individual mean interval values are calculated. In this context, the following notation has been used:

β_j := true but unknown H-value of the upper boundary of the j-th interval; $j = 1(1)k$

b_{jp} := estimate of β_j by the p-th judge; $p = 1(1)n$

$b_{j.}$:= $n^{-1} \sum_p b_{jp}$:= estimate of b_j

a_{jp} := individual deviation of the p-th panel member from b_j,

$\{a_{jp}\}$ distributed $N(0,\ \sigma_j^2)$ and $cov\{a_{jp},\ a_{jp'}\} = 0$ for $p \neq p'$.

The average value of the $b_{jp} = \beta_j + a_{jp}$,

[F.28] $b_{j.} = \beta_j + n^{-1} \sum_p a_{jp}$,

is an unbiased estimator of b_j, since all $\{a_{jp}\}$ have a zero expectation, and has a variance

[F.29] $var\left(b_{j.}\right) = n^{-1}\sigma_j^2$.

We have to bear in mind that for the calculation of the j-th MIV $m_j := \frac{1}{2}\left(b_{j-1} + b_j\right)$ the interval boundary values are not independent, so its variance has to be estimated as

[F.30] $var\{m_j\} = \frac{1}{4}n^{-1}\left[s_{j-1}^2 + s_j^2 + 2cov\{b_{j-1,}, b_j\}\right]$

where s_j^2 is the estimator $\hat{\sigma}_j^2$ of σ_j^2 .

In the **scale application phase**, each of the *N* subjects in the sample has to select one out of the *k* possible ratings $\{R_j \mid j=1(1)k\}$. The following notation has been adopted:

h^* := target value, i.e. the mean happiness in the study population

h_i := individual *H*-value of the *i*-th sample subject $i = 1(1)N$

u_i := $h_i - h^*$= individual deviation from h* of the *i*-th panel subject with $\{u_i\}$ distributed $N(0, \sigma_o^2)$ and $cov\{u_i, u_{i'}\} = 0$ for $i \neq i'$.

f_j := relative frequency of the subjects selecting rating R_j while $\sum_{j=1}^{k} f_j = 1$,

$m_{j\cdot}$:= average value of m_j as obtained in the scale construction phase.

So the model for h_i is :

[F.31] $h_i = h^* + u_i$

and its average value :

[F.32] $h. = h^* + N^{-1}\sum u_i$.

An unbiased estimator of $h^* = h. - N^{-1} \sum u_i$ can be computed as :

[F.33] $h. = \sum_j f_j m_j.$

The variance of the estimated h^* equals :

[F.34] $var\{\hat{h}^*\} = \sum_j \sum_{j'} f_j f_{j'} cov\{m_{j\cdot}, m_{j'\cdot}\} + N^{-1}\sigma_0^2$

The (co)variances in this expression are known from the scale construction phase. The number of degrees of freedom of this variance estimator can be approximated by Satterthwaite's rule (1946), which in this case results in

[F.35] $$df \approx \frac{\left[\sum_j \sum_{j'} f_j f_{j'} cov\{m_{j\cdot}, m_{j'\cdot}\} + N^{-1}\sigma_0^2\right]^2}{(k-1)^{-1}\left[\sum_j \sum_{j'} f_j f_{j'} cov\{m_{j\cdot}, m_{j'\cdot}\}\right]^2 + (N-1)^{-1}[N^{-1}\sigma_0^2]^2}$$

The variability between respondents about the meaning of the various response categories and the boundaries between them on the *H*-scale does not only occur in the construction phase, it also occurs in the application phase.

It may happen that the subjective *H*-sensation of a respondent is situated between the positions of his boundary between two consecutive categories at one hand and the average position of the same boundary as found from the panel results at the other. In this case, the response will be either too high or too low with respect to the relevant MIV. The probability of this event is highest near to the average boundary values between the intervals and relatively small in the mid of the intervals. In the calculations, this variability is not taken into account for large samples, because in this case one may expect that positive and negative differences between the above mentioned positions of the same boundary will occur at almost equal frequencies and largely neutralize each other, so ignoring this will not introduce a bias.

The consequence is that in the above model, the term u_i must be considered to be a combination of two effects. The first one expresses the true happiness inequality, some people are happier than others, whereas the second is caused by the fact that individuals do not assign equal meanings to the same qualification. Hence the estimator of the above model parameter σ^2 also has these two components, which cannot be separated in a simple way.

Under some assumptions, however, it is possible to obtain an impression of the relative contribution of both components to the total variance. The first assumption is that the components are not or almost not dependent on each other. Furthermore, the construction phase may produce some information in general terms about the second component. In this case not only the study sample, but also the panel could be considered as a random sample from the study population, the average variance of the MIV would be an indication of the size of this component. We use the term 'average', since in general the various m_j appear to have different variances and one cannot predict a priori which of the *k* presented categories will be selected by a respondent in the sample. In this way a rough estimate of the true inequality standard deviation can be obtained as the square root of the difference between the sample variance and the mean variance from the construction phase as pointed out above, provided the above assumptions are not violated too strongly.

F.5 The interval approach on the basis of marks

In the end of section F.3, we raised the question of whether the interval approach could be applied if the position of the scale ratings is either decided or estimated as marks. In the present section, we will demonstrate that, in principle, the answer is affirmative and our approach is quite similar to the one described in chapter 6.

In this case, we request the judges to place, for each category, a mark on a line segment from 0 to 10 on the position they see as the most appropriate one. The average positions of these marks, denoted $\{M_j\,;\, j=1(1)k\}$ and their variances and covariances are known then on the basis of all individual judgements. Returning to our example in section 2.2, let us suppose that, hypothetically, the average positions of the marks have been judged to be:

M_4 = 9 for "very happy",
M_4 = 7 for "pretty happy",
M_4 = 4 for "not too happy" and
M_4 = 2 for "unhappy".

The underlying assumptions of our further approach are that (i) all members of a sample within a survey in which this item is applied, will share the views of the judges, at least on an average, and (ii) they will select the category with the label that is closest to their own happiness feeling in such a way, that if somebody's choice between "very happy" and "pretty happy" is made in favour of the former, it is allowed to conclude that his happiness on the underlying latent scale has a value of at least the "mid-mark value" 8.

In this way an interval of happiness values around each mark M_j is obtained. The mid-mark values act as the boundary values of these intervals $\{b_j\}$ as defined in section 6.4, so :

[F.36] $b_j := ½\,(M_j + M_{j+1})$ $\quad j=1(1)k-1$

together with $b_o=0$ and $b_k=10$.

Now for each of the values $H = b_j$, the observed cumulative frequency is known as :

[F.37] $F(b_j) = F_j$, $\quad j=1(1)k$

which is an estimator of $G(b_j)$.

Since no estimates of other points of G are known, we have to make a third assumption on G, for which the same models are available as those mentioned in section F.2. Connecting the available estimated points of G, i.e. the points $\{(b_j.\ F_j)\}$, with straight line segments implies a choice in favour of the semi-continuous model (IV) with uniform distributions over each interval between the consecutive cut points {b_j}. In this case the mid-interval values {m_j} do not necessarily coincide with the corresponding mark position, but are to be calculated as

[F.38] $m_j = {}^1/_2\,[b_{j-1} + b_j] = {}^1/_4\,M_{j-1} + {}^1/_2\,M_j + {}^1/_4\,M_{j+1}$

[F.39] $m_1 = {}^1/_4\,M_1 + {}^1/_4\,M_2$ and

[F.40] $m_k = {}^1/_4\,M_{k-1} + {}^1/_4\,M_k + 5$.

As the reader may verify, the mid-interval values in our hypothetical example are {1,5, 4,25, 6,75, 9}, so the intervals are 'asymmetric around the marks'.

In this model, an unbiased estimate of the population mean happiness is obtained as the weighted average of these {m_j}, each weighted by its corresponding relative frequency in the sample f_j.

The calculation of the different standard deviations is not identical to the one according to the method as developed in section F.4, but formulae can be derived along similar lines and according to similar principles. The actual derivation, however, is considered beyond the scope of this study.

Appendix G
Rounding observations and statistics

G.1 Basic principles of rounding

The way observations and statistics are presented in publications should reflect their precision. The more precise a number, the larger the number of decimal places used is justified. In this way, there is a difference between the three rounded numbers 5, 5,0 and 5,00.

The Dutch Standardization Institute has defined rules for this rounding procedure (Nederlands Normalisatie Instituut, 1967). The basic underlying principle is that the effect of rounding should always be less than 1% of the standard error of the observation or estimate. The instructions use the term "standard deviation", but from the context one has to conclude that "standard error" is in better agreement with the conventional terminology.

This appendix summarizes the relevant current rounding rules, with special reference to statistics for happiness of nations. The rules concern what is called the "rounding interval". This rounding interval, denoted in this appendix as "*a*", is defined as the smallest possible positive difference between two rounded values of the same statistic as reported in a study report. If a set of average happiness values is reported in two decimal places, the rounding interval *a* = 0,01.

We have formulated in section G.4 some simple rules of thumb for larger samples, say >100, which are only approximately in agreement with the rounding rules, but the application of which might be an improvement of current practices in at least some research institutes in this field. This has been done on the basis of the above rounding rules

G.2 The six basic rules for rounding

1. Rounding numbers should be done only after all computations have been completed.
2. Admissible values for the rounding interval are all integer (positive, zero and negative) powers of 10 only. In other words, they all belong to the series { ...; 0,001 ; 0,01 ; 0,1 ; 1 ; 10; 100; 1000; ...}
3. For the rounding interval *a*, the maximum value should be selected from the above series that does not exceed half the standard error of the observation or of the statistic.

So: $a \leq ½$ (standard error) $< 10a$.

Example: if some statistic has a standard error of 0,062, then the value *a* = 0,01 has to be selected, since :

$$0{,}01 \leq 0{,}031 < 0{,}1.$$

4. If more than one decimal is to be dropped, rounding is done in one single step.
5. Numbers are rounded to the nearest rounded value. If the choice is not unambiguous, the last digit after rounding should be even.
   So for *a*= 0,01:

   4,663 → 4,66;
   3,187 → 3,19;
   5,325 → 5,32;
   5,335 → 5,34.
6. The rounding interval to be applied to confidence limits is the same as that to be used for the corresponding point estimates.

A consequence of the first rule above is that rounded numbers do not always follow the arithmetic rules for exact numbers. If e.g. 18,3 + 54,3 + 27,4 = 100,0 and a =1, then after rounding 18 + 54 + 27 = 100 and not 99 ! In such situations, some people seek to 'adjust' one or more numbers e.g. by rounding up 27,4 to 28, but this optical 'solution' is essentially incorrect and should be avoided under all circumstances.

G.3 Rounding interval values in happiness studies

Symbols:

a = rounding interval
s = estimated standard deviation between individual happiness ratings
df = degrees of freedom of estimated standard deviation
n = number of pairs in a correlational analysis
N = sample size

Statistic	standard error	*a* =	while
average values	$s/\sqrt{N}$	0,1	$N \leq 25\ s^2$
		0,01	$25\ s^2 < N \leq 2.500\ s^2$
standard deviation	$s/\sqrt{(2\text{xdf})}$	0,1	$\text{df} \leq 12,5\ s^2$
		0,01	$12,5\ s^2 < \text{df} \leq 1250\ s^2$
percentage (%)	$\sqrt{(\%\text{x}(100\text{-}\%)/N)}$	1	$N \leq 0,25\text{x}\%\text{x}(100\text{-}\%)$
		0,1	$N \leq 25\text{x}\%\text{x}(100\text{-}\%)$
		0,01	$N \leq 2500\text{x}\%\text{x}(100\text{-}\%)$
correlation coefficient	$\approx 1/\sqrt{(n\text{-}1)} \approx 1/\sqrt{n}$	0,1	$n \leq 25$
		0,01	$25 < n \leq 2.500$

G.4 Some practical rounding rules of thumb

For practical use in studies on happiness in nations, the following rounding rules of thumb are recommended. They are approximately correct only and are based on a number of assumptions:

- average values are reported on the basis of a [0, 10] or a [1, 10] scale.
- standard deviations are of the order of magnitude of $s = 2$ on that scale.
- sample sizes N: $100 \leq N \leq 5.000$
- for the percentage of e.g. unhappy people 10% ≤ % ≤ 90%.

Rules of thumb for samples of size > 100:

- average happiness values,
- confidence limits for the mean happiness value,
- standard deviations and
- correlation coefficients

are all reported to **two decimal places**, e.g. 6,48.

Percentages of e.g. (un)happy people are reported as **integers** if N < 600 and rounded to **one decimal place** otherwise e.g. 18,3 %.

Appendix H

Properties of the beta distribution

(Survey of relevant formulae concerning the beta distribution).

H.1 The beta function (B)

The complete beta function with (shape) parameters α and β is defined as

[H.1] $$B(\propto,\beta) := \int_0^1 t^{\propto-1}(1-t)^{\beta-1}dt \quad t \in [0,1] \subset \mathbb{R}\ , \quad \alpha,\beta \in \mathbb{R}^+$$

and the *incomplete beta function* of the argument *x* and with parameters α and β as :

[H.2] $$B(x;\ \propto,\beta) := \int_0^x t^{\propto-1}(1-t)^{\beta-1}dt \quad ; \quad x,t \in [0,1] \subset \mathbb{R} \quad ; \quad \alpha,\beta \in \mathbb{R}^+$$

Some particular values of $B(\alpha,\beta)$ relevant in this context are those with α and/or β = 1 or 2:

$B(1,1) = 1$ $\qquad$ $B(2,2) = 1/6$

$B(1,2) = B(2,1) = ½$

$B(\alpha,1) = \alpha^{-1}$ $\qquad$ $B(1,\beta) = \beta^{-1}$

$B(\alpha,2) = [\alpha(\alpha+1)]^{-1}$ $\qquad$ $B(2,\beta) = [\beta(\beta+1)]^{-1}$

H.2 The standard beta distribution

The random variable *X* has a standard beta distribution if it has a probability density function :

[H.3]

$$g(x) := \begin{cases} [B(\propto,\beta)]^{-1}x^{\alpha-1}(1-x)^{\beta-1} & for\ X = x \in [0,1] \subset \mathbb{R}, \quad \alpha,\beta \in \mathbb{R}^+ \\ 0 & \text{otherwise} \end{cases}$$

The (cumulative) distribution function of this *X* is the ratio of the incomplete and the complete beta function:

[H.4] $$G(x) := \text{Prob}\ \{X \le x\} = \int_{-\infty}^x g(t)dt = [B(\propto,\beta)]^{-1}\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$$

H.3 The non-standard beta distribution

The random variable *Y* has a non-standard beta distribution if it has a probability density function :

[H.5]

$$g^*(y) := \begin{cases} [B[\alpha,\beta)]^{-1}(D-C)^{-1}(y-C)^{\alpha-1}(D-y)^{\beta-1} & \text{for } \; Y = y \in [C,D] \subset \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

This distribution can be converted into the standard beta distribution (for which C = 0 and D = 1) by the linear transformation

[H.6] $$y \to x = \frac{y-C}{D-C} \;\Leftrightarrow\; y = C + (D-C)x.$$

The factor $(D - C)^{-1}$ is the Jacobian determinant $|\frac{dx}{dy}|$, which is required for normalization reasons, i.e. to establish

$G^*(y) = 0$ for $y \leq C$ and $G^*(y) = 1$ for $y \geq D$

H.4 Moments of the standard beta distribution of a random variable *X*

The mean value of *X* is determined completely by the ratio α / β :

[H.7] $$\mu := \mu_1' := \mathcal{E}X = \frac{\alpha}{\alpha+\beta} = \frac{\alpha/\beta}{\alpha/\beta+1} \;\Leftrightarrow\; \frac{\alpha}{\beta} = \frac{\mu}{1-\mu},$$

where $\mathcal{E}X$ is the expectation of *X*.

The first three central moments of this distribution are:

[H.8] $\mu_1 := 0$,

[H.9] $$\sigma^2 := \mu_2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$ and

[H.10] $$\mu_3 = \frac{2\alpha\beta(\beta-\alpha)}{(\alpha+\beta)^3(\alpha+\beta+1)(\alpha+\beta+2)}$$

The skewness of a distribution is often defined as μ_3, but sometimes as the "coefficient of skewness", being the ratio $\mu_3 / \mu_2^{3/2}$, for which various symbols are current such as γ_1 and β_3.

It can be computed as :

[H.11] $$\gamma_1 = \frac{2(\beta-\alpha)}{\alpha+\beta+2}\sqrt{\frac{\alpha+\beta+1}{\alpha\beta}}$$

The skewness is always positive if $\mu < 0,5$ and negative if $\mu > 0,5$.

In a similar way, a coefficient of kurtosis, is defined as the ratio :

[H.12] $$\gamma_2 := \mu_4/\mu_2^2 = \frac{3(\alpha+\beta+1)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}[2(\alpha+\beta)^2 + \alpha\beta(\alpha+\beta-6)]$$

where μ_4 is the fourth central moment of the distribution.

Some limits:

[H.13] $$\lim_{\alpha,\beta\downarrow 0} \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2}$$

[H.14] $$\lim_{\alpha,\beta\downarrow 0} \gamma_2 = \alpha/\beta + \beta/\alpha - 1$$

[H.15] $$\lim_{\alpha,\beta\to\infty} \sigma^2 = 0$$

and

[H.16] $$\lim_{\alpha,\beta\to\infty} \gamma_2 = 3$$

H.5 Moment estimators of the parameters

The distribution parameter α and the moments μ and σ^2 are related by :

[H.17] $$\sigma^2 = \frac{\mu(1-\mu)}{\alpha/\mu + 1}$$

Consequently :

[H.18] $$\propto = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]$$

[H.19] $$\beta = (1-\mu)\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]$$

On this basis, the moment estimators $\hat{\alpha}$ and $\hat{\beta}$ can be obtained as :

[H.20] $$\hat{\propto} = \bar{X}\left[\frac{\bar{X}(1-\bar{X})}{m_2} - 1\right]$$ and

[H.21] $$\hat{\beta} = (1-\bar{X})\left[\frac{\bar{X}(1-\bar{X})}{m_2} - 1\right],$$

where $\overline{X}$ is the sample average value and m_2 is the second central moment of the sample, i.e. the sample variance, but with the denominator n and not n-1

H.6 The shape of the probability density function $g(x)$

For $\alpha = \beta = 1$, the uniform distribution on the interval [0, 1] is obtained with $g(x) = 1$ and $G(x) = x$.

For $\alpha, \beta < 1$, the distribution is U-shaped.

For $\beta < 1$ and $\alpha \geq 1$, $g(x)$ is a monotonically increasing function (for $\alpha > 2$ J-shaped).

For $\alpha < 1$ and $\beta \geq 1$, $g(x)$ is a monotonically decreasing function (for $\beta > 2$ inverse J-shaped).

For $\alpha, \beta > 1$, the distribution is unimodal with

[H.22] $$Mode = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{(\alpha - 1)}{(\alpha - 1) + (\beta - 1)} \in (0,1) \subset \mathbb{R}$$

The situation at and around $X = 0$ is mainly determined by the value of α:

[H.23] $G(0) = 0$

[H.24] $$g(0) := \lim_{x \downarrow 0} g(x) = \begin{cases} \infty & \text{for } \alpha < 1 \\ \beta & \text{for } \alpha = 1 \\ 0 & \text{for } \alpha > 1 \end{cases}$$

[H.25]

$$g'(0) := \lim_{x \downarrow 0} g'(x) = \lim_{x \downarrow 0} [B(\alpha,\beta)]^{-1} (\alpha - 1) x^{\alpha - 2} = \begin{cases} -\infty & \text{for } \alpha < 1 \\ \infty & \text{for } 1 \leq \alpha < 2 \\ \beta(\beta + 1) & \text{for } \alpha = 2 \\ 0 & \text{for } \alpha > 2 \end{cases}$$

In the same way, the situation at and around $X = 1$ is mainly determined by the value of β:

[H.26] $G(1) = 1$

[H.27] $$g(1) := \lim_{x \uparrow 1} g(x) = \begin{cases} \infty & \text{for } \beta < 1 \\ \alpha & \text{for } \beta = 1 \\ 0 & \text{for } \beta > 1 \end{cases}$$

and

[H.28]

$$g'(1) := \lim_{x \uparrow 1} g'(x) = \lim_{x \uparrow 1} [B(\alpha, \beta)]^{-1} (\beta - 1)(1 - x)^{\beta - 2} = \begin{cases} -\infty & \text{for } \quad \beta < 1 \\ \infty & \text{for } 1 \leq \beta < 2 \\ \alpha(\alpha + 1) & \text{for } \quad \beta = 2 \\ 0 & \text{for } \quad \beta > 2 \end{cases}$$

REFERENCES

Bartram, P. & Yelding, D. (1973),
The Development of an Empirical Method of Selecting Phrases Used in Verbal Rating Scales: A Report on a Recent Experiment,
Journal of the Market Research Society, **15** (2), 151 – 156.

Bentham, J. (1789),
Introduction to the Principles of Morals and Legislation,
London (UK)

Buijs, P., (2007),
De eeuw van het geluk (The Age of Happiness),
Verloren, Hilversum (NL).
ISBN 978-90-6550-999-4.

Cantril, H. (1946),
The Intensity of an Attitude,
Journal of Abnormal Social Psychology, **41**, 129–135.

Cantril, H. (1965),
The Pattern of Human Concern,
Rutgers University Press, New Brunswick, New Jersey (US)

Chin-Hon-Foei, S. (1989),
Life Satisfaction in the EC Countries, 1975-1984, in
Veenhoven, R., Ed.: *"Did the Crisis Really Hurt?"*
Universitaire Pers Rotterdam, (NL), 24–43.

Cramér, H., (1946),
Mathematical Methods of Statistics,
Princeton University Press, Princeton, (US)
ISBN 0-691-08004-6

Diener, E., Suh, E.M., Lucas, R.E. & Smith, H.L. (1999)
Subjective Well-Being: Three Decades of Progress
Psychological Bulletin, **125** (2), 276 – 302

Easterlin, R.A., (1974, p. 109),
Does Economic Growth Improve the Human Lot? Some Empirical Evidence, in
David, P.A. & Melvin, W.R. (Eds.),
'Nations and Households in Economic Growth';
Academic Press, New York (US), pp. 89 –125.

Estes, R. (1984)
The Social Progress of Nations
Preager, New York, (US)

Ferrer-i-Carbonell, A. & Frijters, P. (2004)
How Important is Methodology for the Estimates of the Determinants of Happiness?
Economic Journal, **114** (497), 641–659.

Gini, C. (1912),
Variabilità e Mutabilità, contributo allo studio delle distribuzioni e relazoni statistiche (Variability and Instability, a contribution to the study of statistical distributions and relationships),
Studi Economico-Giuridici dell' Univ. di Cagliari, **3** (2), 1–158.

Goodman, L.A., (1954),
Kolmogorov-Smirnov Tests for Psychological Research,
Psychological Bulletin, **51**, 160–168.

Guildford, J.P., (1936[1], 1954[2]),
Psychometric Methods,
McGraw-Hill Book Company, New York/Toronto/London (US)

Gupta, A.K. & Nadarajah, S. (Eds), (2004),
Handbook of Beta Distribution and Its Applications,
Marcel Dekker, New York (US).

Hicks, H.A. (1997),
The Inequality-Adjusted Human Development Index: A Constructive Proposal,
World Development, **25**, 1283-1298

Jones, L.V. & Thurstone, L.L. (1955),
The Psychophysics of Semantics: An Experimental investigation,
The Journal of Applied Psychology, **39** (1), 31–36.

Kalmijn, W.M. & Arends, L.R. (2010),
Measures of Inequality: Application to Happiness in Nations,
Social Indicators Research, **99** (1), 147-161
ISSN 0303-8300 (print), 1573-0921 (on line).In press.
DOI 10.1007/s11205-010-9573-z

Kalmijn, W.M., Arends, L.R. & Veenhoven, R. (2010),
Happiness Scale Interval Study. Methodological Considerations.
Social Indicators Research In press
DOI 10.1007/s11205-010-9688-2

Kalmijn, W.M. & Veenhoven, R. (2005),
Measuring Inequality of Happiness in Nations. In Search for Proper Statistics.
Journal of Happiness Studies, **6**, 357–396

Keeping, E.S. (1962),
Introduction to Statistical Inference,
Van Nostrand Reinhold, New York (US).

Kendall, M.G. (1948, 1962^3),
Rank Correlation Methods,
C. Griffin, London, (UK).

Kendall, M.G. & Stuart, A. (1977^4),
The Advanced Theory of Statistics, vol. I,
Ch. Griffin & Cy. Ltd, London/High Wycombe, (UK).

Kilpatric, F.P. & Cantril H. (1960),
Self-anchoring Scale: A Measure of Individuals' Unique Reality World,
Journal of Individual Psychology, **16**, 158–173.

Langford, E. (2006)
Quartiles in Elementary Statistics
Journal of Statistics Education, **14**, (3).

Kotz, S. & Dorp, J.R. van (2004, p. 9 -10)
Beyond Beta, Other Continuous Families of Distributions with Bounded Support and Applications,
World Scientific Press, Singapore.

Mazaheri, M. & Theuns, P. (2009)
Effects of Varying Response Formats on Self-ratings of Life Satisfaction,
Social Indicators Research,
ISSN 0303-8300 (print), 1573-0921 (on line)
DOI 10.1007/s11295-008-9263-2

Mosier, C.I. (1941),
A Psychometric Study of Meaning
J. of Social Psychology **13**, 123–140.

Myers, J.H & Warner, W.G. (1968),
Semantic Properties of Selected Evaluation Adjectives,
J. Marketing Res. **5**, 409–412.

Nederlands Normalisatie Instituut (1967),
Afronden van waarnemingen (Rounding off observations), in
Receptbladen voor de statistische verwerking van waarnemingen
(Instructions for statistical treatment of series of observations),
Nederlandse Norm NEN 1047, Ch. 2.1
Nederlands Normalisatie Instituut, Rijswijk ZH (NL)

Orwell, G (1945[1]),
Animal Farm,
Secker and Warburg Ltd, London, (UK).

Ott, J (2005),
Level and Equality of Happiness in Nations; Does Happiness of a Greater Number Imply Greater Inequality in Happiness?
Journal of Happiness Studies, **6** (4), pp. 397–420

Owen, D.B. (1962),
Handbook of Statistical Tables,
Addison-Wesley Publishing Company Inc, Reading Mass. (US), p. 140

Praag, B.M.S. van (1991),
Ordinal and Cardinal Utility: an Integration of the Two Dimensions in the Welfare Concept.
Journal of Econometrics, **50**, 69-89

Praag, B.M.S. van & Ferrer-i-Carbonell, A. (2004),
Happiness Quantified, a Satisfaction Calculus Approach,
Oxford University Press, Oxford, (UK).

Satterthwaite, F.E., (1946),
An Approximate Distribution of Estimates of Variance Components,
Biometrics Bulletin, **2**, 110–114.

Sen, A. (1997[2]),
On Economic Inequality,
Clarendon Press, Oxford (UK).

Siegel, S. (1956),
Nonparametric Statistics for the Behavioral Sciences,
McGraw-Hill Book Company, New York/Toronto/London (US).

Smith, T.W. et al., (2005),
Methods for Assessing and Calibrating Response Scales Across Countries and Languages,
Comparative Sociology, **4** (3-4), 365–415

Stevens, S.S. (1946),
On the Theory of Scales of Measurement
Science, **103**, 670 – 680.

Tamhane, A.C., Ankenman, B.C., Yang, Y. (2002),
The Beta Distribution as a Latent Response Model for Ordinal Data (I); Estimation of Location and Dispersion Parameters,
Journal of Statistical Computation and Simulation, **72** (6), 473-494.

Theil, H, (1967),
Economics and Information Theory,
North-Holland Publishing Cy, Amsterdam (NL).

Thurstone, L.L. (1927),
Psychophysical Analysis,
American Journal of Psychology, **38**, 368–389.

UNDP (2000)
Human Development Report 1998
UNDP United Nations Development Program

VanPraag, B.M.S., (1991),
Ordinal and Cardinal Utility: an Integration of the Two Dimensions in the Welfare Concept.
Journal of Econometrics, **50**, 69–89

VanPraag, B.M.S. & Ferrer-i-Carbonell, A., (2004),
Happiness Quantified, a Satisfaction Calculus Approach,
Oxford University Press, Oxford, (UK).

Veenhoven, R. (1984),
Conditions of Happiness,
Kluwer Academic, Dordrecht (NL).

Veenhoven, R. (1990),
Inequality in Happiness, Inequality in Countries Compared between Countries,
Paper 12th Work Congress of Sociology, Madrid (ES).

Veenhoven, R. (1993),
Happiness in Nations, Studies in Socio-Cultural Transformation, nr. 2,
RISBO, Erasmus University Rotterdam, (NL).

Veenhoven, R. (1997),
Le bonheur du plus grand nombre comme but des politiques sociales,
Revue Québécoisede Psychologie, **18**, 29-74.

Veenhoven, R. (1999),
Quality-of-life in Individualistic Society: A comparison in 43 nations in the early 1990's ,
Social Indicators Research, **48**, 157 –186.

Veenhoven, R. (2000),
Well-being in the Welfare State: Level not Higher, Distribution not More Equitable,
Journal of Comparative Policy Analysis: Research and Practice, **2**, 91–125.

Veenhoven, R. (2002a),
Het grootste geluk voor het grootste aantal. Geluk als richtsnoer voor beleid.
(Inaugural lecture Erasmus University Rotterdam),
Sociale Wetenschappen, nr. 4 pp. 1–43.
ISSN 0037-8077

Veenhoven, R. (2002b),
Die Rückkehr der Ungleichheit in die moderne Gesellschaft? Die Verteilung der Lebenszufriedenheit in den EU-Ländern von 1973 bis 1996 (*Return of inequality in modern Society? Dispersion of life-satisfaction in EU-nations* 1973 -1996), in
Wolfgang Glatzer, Roland Habich & Karl-Ulrich Maier (Hrgs),
'Socialer Wandel und Gesellschaftliche Dauerbeobachtung. Festschrift für Wolfgang Zapf',
Leske+Bundrich, Opladen, (DE).
ISBN 3-8100-3368-5, pp. 273-294.

Veenhoven, R. (2003a),
Equality-adjusted Happiness in Nations
Paper presented at the conference of the International Society for Quality of life Studies (ISQOLS) Frankfurt (DE).

Veenhoven, R. (2003b),
Equality-Adjusted Happiness in 61 Nations in the 1990s; How Well Nations Combine High Level and Small Difference in Happiness
World Database of Happiness (www.eur.nl/fsw/research/happiness),
Happiness in Nations, Rank Report 2003-4

Veenhoven, R. (2004),
Veiligheid en geluk (Security and happiness)
In: E.R. Muller (ed.) *'Veiligheid',*
Kluwer, Alphen aan de Rijn (NL), 153–188

Veenhoven, R. (2005),
Return of Inequality in Modern Society? Trend in Dispersion of Life Satisfaction in EU Nations 1973-1996,
Journal of Happiness Studies, **6**, (4), 457– 487.

Veenhoven, E. & Hermus, P. (2006)
Scale Interval Recorder. Tool for Assessing Relative Weights of Verbal Response Options on Survey Questions. Web survey program.
Erasmus University Rotterdam, Department of Social Sciences & Risbo Contract Research, The Netherlands.

Veenhoven, R. (2009),
International Scale Interval Study: Improving the Comparability of Responses to Survey Questions about Happiness. In:
Moller, V. & Huschka, D. (Eds.) *'Quality of Life and the Millennium Challenge: Advances in Quality-of-Life Studies, Theory and Research',*
Social Indicators Research Series 35, 45–58.
Springer, e-ISBN 978-1-4020-8569-7.

Veenhoven, R. (2010)
World Database of Happiness (WDH),
Continuous Register of Research on Subjective Appreciation of Life,
Erasmus University Rotterdam (NL).
Available at: http://www.worlddatabaseofhappiness.eur.nl

Veenhoven, R. & Ehrhardt, J. (1995),
The Cross-national Pattern of Happiness; Test of Predictions Implied in Three Theories of Happiness,
Social Indicators Research, **34**, 33– 68.

Veenhoven, R. & Kalmijn, W.M. (2005),
Inequality-Adjusted Happiness in Nations. Egalitarianism and Utilitarianism Married Together in a New Index of Societal Performance,
Journal of Happiness Studies, **6** (4), 421–455.

Ventegodt, S. (1995),
Liskvalitet i Denmark (Quality of Life in Denmark. Results of a Population Survey),
Forskningcentrets Forlag, København (DK), 64–67.

Voss, K.E. et al. (1996),
An Exploration of the Comparability of Semantic Adjectives in Three Languages; A Magnitude Estimation Approach,
International Marketing Review, **13** (5), 44–58.

Washburn, M.F., Harding, L. Simons, H. & Tomlinson, D. (1925)
Further Experiments on Directed Recalls as a Test of Cheerful and Depressed Temperaments.
American Journal of Psychology, **36**, 454–456.
See also:
http://www1.eur.nl/fsw/happiness/hap_cor/desc_study.php?studyid=148

Webb, E. (1918),
Character and Intelligence (diss)
Cambridge University Press, Cambridge (UK).
See also:
http://www1.eur.nl/fsw/happiness/hap_cor/desc_study.php?studyid=121

WHO (2001)
The World Health Report 2000
Geneva (CH).

Wildt, A.R. & Mazis, M.B. (1978),
Determinants of Scale Response: Label Versus Position,
Journal of Marketing Research **15** (2), 261–267.

Yew-Kwang Ng (1996),
Happiness Surveys: Some Comparability Issues and an Exploratory Survey Based on Just Perceivable Increments,
Social Indicators Research, **38**, 1 – 27.

Yew-Kwang Ng (1997),
A Case for Happiness, Cardinalism, and Interpersonal Comparability,
The Economic Journal, **107**, 1848 – 1858.

Samenvatting

Geluk geldt als belangrijk in het leven van mensen. Deze zienswijze werkt al enkele decennia lang door in een groeiende belangstelling voor het geluk vanuit de wetenschap. Onderzoek van geluk is echter alleen mogelijk als geluk meetbaar en kwantificeerbaar is. Dit proefschrift gaat over de meting van geluk en daarbij vooral over de wijze waarop de uitkomsten van die metingen verder verwerkt worden. Als zodanig levert het een methodologische bijdrage aan dat wetenschappelijk onderzoek. Onder geluk wordt in dit verband verstaan de mate van subjectieve voldoening van het individu met het eigen bestaan.

Dit geluk wordt doorgaans gemeten door het aan de persoon zelf te vragen. Een veel voorkomende vorm is die van een gesloten vraag, bijvoorbeeld "Hoe gelukkig voelt u zich al met al?" . De proefpersoon krijgt dan een beperkt aantal (3 tot 7) verbale antwoordcategorieën aangeboden, waarvan er één moet worden aangekruist, bijvoorbeeld "tamelijk gelukkig". Het zijn in het bijzonder de geluksvragen met zulke verbale antwoordcategorieën, kortweg aangeduid als "verbale schalen", die het voorwerp van deze studie zijn.

Nu zijn geluksonderzoekers niet alleen geïnteresseerd in individuele geluksscores, maar ook in geluk in grotere samenlevingsverbanden, bijvoorbeeld in landen. In dit proefschrift wordt vrijwel steeds kortheidshalve de aanduiding "naties" gebruikt, maar de beschouwingen kunnen even goed op andere gedefinieerde collectiva toegepast worden.

Niet alle mensen zijn even gelukkig. Bij deze ongelijkheid van geluk wordt onderscheid gemaakt tussen ongelijkheid binnen naties en die tussen naties. Om daarover meer te weten te komen wordt er gewerkt met steekproeven uit de bevolking of een deel daarvan, bijvoorbeeld alleen de volwassen burgers. Zulke steekproeven behoren aselect getrokken te worden. In de praktijk gebeurt dat eigenlijk nooit, maar wij nemen in onze beschouwingen aan dat er gewerkt is met steekproeven die daar wel voor door kunnen gaan. Alleen onder die aanname mag uit de steekproefresultaten informatie afgeleid worden over het geluk in de natie die door de steekproef gerepresenteerd wordt. En daar is het uiteindelijk om begonnen.

Aspecten van ongelijkheid

De ongelijkheid binnen de steekproef wordt samengevat in de frequentieverdeling van de gezamenlijke antwoordcategorieën. Deze verdeling wordt doorgaans gekarakteriseerd door twee statistische grootheden. De ene geeft de centrale waarde aan, waar omheen de individuele geluksscores gespreid zijn, de andere karakteriseert die spreiding, de geluksongelijkheid in de steekproef. Beide elementen zijn sociologisch interessant, omdat aangenomen wordt dat een samenleving het beter doet naarmate de centrale waarde groter is, maar ook naarmate de ongelijkheid daarin kleiner is. Die laatste toevoeging is afkomstig uit de kring der zogenoemde egalitaristen, terwijl de utilitaristen alleen naar het algemene geluksniveau kijken.

Voor een goed beeld van de gelukssituatie in een natie is het op zich al noodzakelijk om deze beide aspecten te kunnen kwantificeren. Geluksonderzoekers zijn echter niet tevreden met deze beschrijvende parameters van de statistische verdeling van het geluk, ze zoeken ook naar de samenhang hiervan met andere karakteristieken van samenlevingen die dat geluk zouden kunnen beïnvloeden. Ook voor het onderzoek van zo'n samenhang is kwantificering van het geluk noodzakelijk. De afdeling "Geluksonderzoek" van de Erasmus Universiteit Rotterdam doet dergelijk onderzoek. In het kader daarvan zijn de afgelopen jaren een aantal vragen en problemen van methodologische aard aan de orde gekomen. Dit proefschrift is een neerslag van een aantal van deze vragen en van antwoorden de daarop gegeven zijn.

Methodologische problemen bij meting van ongelijkheid in geluk

De kerntaak van de inferentiële statistiek is het trekken van conclusies over populaties op basis van gegevens die verkregen zijn door meting in steekproeven. Die opgave staat ook in dit proefschrift over geluksmeting centraal. Het meten van geluk brengt een aantal bijzondere problemen met zich mee, met name wanneer daarbij gebruik gemaakt wordt van verbale schalen. In die gevallen maakt het kwantificeren van geluk en van geluksverschillen het allereerst nodig tekst van de antwoorden om te zetten in getallen.

Gebruikelijk is om de diverse antwoordcategorieën te coderen met een cijfer dat de rangorde van de daarmee corresponderende categorie op de meetschaal aangeeft. Deze cijfercodes worden vervolgens aangezien voor kardinale getallen, waarmee rekenkundige bewerkingen uitgevoerd mogen worden. Dat gebeurt dan ook en op deze manier wordt het steekproefgemiddelde als gewogen gemiddelde van de cijfercodes berekend als geschatte maat voor de centrale waarde en daarna bijvoorbeeld de standaardafwijking als maat voor de ongelijkheid binnen de steekproef. Bij die codering wordt impliciet equidistantie van de categorieën geïntroduceerd.

Na het coderen worden de teksten verder volledig genegeerd. Evenzo wordt voorbij gegaan aan de mogelijkheid dat bijvoorbeeld "heureux" in het Frans niet noodzakelijk precies hetzelfde betekent als "gelukkig" in onze taal. Zodoende worden alle resultaten verkregen met zogenoemde vierpuntsschalen op precies dezelfde wijze verwerkt . Bovendien varieert het aantal categorieën ook nog, en wel van 3 tot 7, wat een aanvullende bewerking nodig maakt om tot vergelijkbare uitkomsten te kunnen komen. Gebruikelijk is om dit te bereiken door alle schalen gelijkmatig op te rekken tot een gemeenschappelijke schaal van 0 tot 10: de meest ongelukkige categorie krijgt daarop de waarde 0, ongeacht de bijbehorende tekst, en de meest gelukkige de waarde 10. Alle overige categorieën krijgen daartussen nieuwe equidistante posities toegewezen overeenkomstig hun volgnummers op de primaire meetschaal. Via deze lineaire schaaltransformatie komt men dan tot een steekproefgemiddelde en een standaardafwijking, beide op een 0 tot 10 schaal en deze uitkomsten worden gegeneraliseerd door ze ook van toepassing te verklaren op de desbetreffende natie.

Er is echter nog een ander probleem bij deze werkwijze en dat is het beperkte aantal antwoordcategorieën, maximaal zeven. Dat betekent dat het gemeten geluk in de steekproef nooit meer dan zeven verschillende waarden kan aannemen. Bij de generalisatie wordt datzelfde ook van toepassing verklaard op de geluksverdeling in de natie. Qua geluk bestaat de natie dan uit (maximaal) zeven verschillende soorten inwoners. Nu is het gemeten geluk noodzakelijkerwijs een discrete variabele, maar het geluk zelf wordt eveneens zo beschouwd, hoewel het veel meer voor de hand ligt om daarvoor te denken aan een continue variabele op het interval van 0 tot 10.

Doel van deze studie

Tegen de hierboven beschreven traditionele werkwijze zijn tal van bezwaren van methodologische aard in te brengen. Vele daarvan zouden wellicht te ondervangen zijn door het werken met verbale schalen te beëindigen en voortaan alleen te werken met numerieke schalen van bijvoorbeeld 1 tot 10. Een dergelijke beslissing zou echter betekenen dat alle gepubliceerde geluksonderzoeken waarin verbale schalen zijn toegepast, – en dat aantal loopt in de duizenden – niet meer bruikbaar zouden zijn voor meta-analyses en voor trendonderzoekingen. Daarom wordt in dit proefschrift een andere aanpak ontwikkeld en wordt nagegaan op welke wijze zo goed mogelijk tegemoet gekomen kan worden aan de methodologische bedenkingen tegen de traditioneel gevolgde werkwijze en toch de uitkomsten te kunnen blijven gebruiken, ook al zijn ze met behulp van verbale schalen verkregen.

Aanpak

Het eerste hoofdstuk beschrijft de context waarbinnen dit proefschrift tot stand gekomen is. Daarnaast worden de drie kernbegrippen in de titel ervan nader belicht: geluk, ongelijkheid en kwantificering. Na deze inleiding wordt in hoofdstuk 2 uiteengezet hoe het geluk doorgaans gemeten wordt. Aandacht wordt besteed aan een aantal gezichtspunten van algemene methodologische aard, met name aan geluk als variabele. Gewezen wordt op essentiële verschillen in dit opzicht tussen geluk enerzijds en inkomen anderzijds, twee zaken waarvan de onderlinge samenhang een veelvuldig gekozen voorwerp van onderzoek is.

In hoofdstuk 3 wordt op meer fundamentele wijze ingegaan op het begrip "ongelijkheid" en op de kwantificering daarvan. Ongelijkheid binnen een steekproef wordt daarin beschreven als een binaire relatie, gedefinieerd op de verzameling van individuele categoriekeuzes in een steekproef. Dat stelt in staat om niet alleen de minimale waarde (nul) van de ongelijkheid te bepalen, maar ook de maximale waarde, welke bereikt wordt als de helft van de respondenten kiest voor het meest gelukkige en de andere helft voor het minst gelukkige alternatief. De ongelijkheid binnen een steekproef als geheel is dan te kwantificeren als het percentage van dat maximum.

Ongelijkheid binnen een continue verdeling in een populatie blijkt op een verwante manier gedefinieerd te kunnen worden door gebruik te maken van differentialen. Voor niet-lineaire kansdichtheden wordt de uitvoering al snel heel gecompliceerd. Numerieke integratie maakt het echter in beginsel toch mogelijk om een beeld van de ongelijkheid te ontwerpen, hetgeen gedemonstreerd wordt voor de standaard beta waarschijnlijkheidsverdeling.

De bevindingen van hoofdstuk 3 worden toegepast in hoofdstuk 4. Dit gaat over de vraag welke maten geschikt zijn om de ongelijkheid van geluk binnen een steekproef uit een natie te karakteriseren. Negen kandidaten voor deze maat worden toegepast op vijf reeksen van hypothetische verdelingen met binnen elke reeks oplopende ongelijkheid. De beoordeling geschiedt op een aantal vooraf gekozen criteria. Uiteindelijk blijkt dat de standaardafwijking en de gemiddelde (absolute) afstand tussen alle mogelijke paren van de respondenten in de steekproef de meest geschikte maten te zijn.

De gemiddelde absolute afstand tot het gemiddelde en de kwartielafstand zijn een goede tweede keus, zij doen het allebei net iets minder goed. Voor het geluk, zoals dat gemeten pleegt te worden, blijven de variatiecoëfficiënt, de Gini-coëfficient en Theil's entropie-index onmiskenbaar beneden de maat. De spreidingsbreedte en het percentage buiten de modale categorie komen evenmin in aanmerking. De conclusie is dat in het geluksonderzoek de standaardafwijking in het verleden terecht als indicator voor de ongelijkheid gebruikt is en dat er geen reden is om voortaan een andere koers te gaan varen.

De eerdergenoemde tegenstelling tussen de opvattingen van de utilitaristen en de egalitaristen zou opgelost kunnen worden door een maat te construeren die aan beide zienswijzen recht doet. In het vijfde hoofdstuk wordt deze maat aangeboden in de vorm van de "Inequality-Adjusted Happiness" (IAH) als een lineaire combinatie van het gemiddelde en de standaardafwijking in de steekproef. Deze index heeft de waarde 100 als theoretisch maximum in de ideale samenleving en de waarde dichtbij nul voor de meest miserabele variant. Geeft men aan beide zienswijzen gelijke gewichten, dan worden tussen 2000 en 2010 voor 140 naties IAH-waarden gevonden, die variëren van 19 tot 79. Het onderscheidend vermogen van de index lijkt daarmee ruim voldoende.

In de laatste twee hoofdstukken (6 en 7) wordt vooral ingegaan op de wijze waarop steekproefuitkomsten vertaald worden in karakteristieken die de geluksverdeling in de populatie beschrijven, d.w.z. de verdeling van de brgelukswaarden over alle inwoners van eenzelfde natie.

Zoals gezegd wordt de gebruikelijke methode gekenmerkt door verscheidene methodologische tekortkomingen. Een eerste correctie kwam tot stand in 1993 door Veenhoven c.s. in de vorm van de zogenoemde "Thurstone-waarden". Een verdere verbetering wordt verwacht van de "Happiness Scale Interval" benadering. Daarbij wordt voor het eerst methodisch uitgegaan van het idee dat "tamelijk gelukkig" niet voorgesteld wordt door een enkel punt op de meetschaal, maar door een interval van een groot aantal elkaar opvolgende gelukswaarden. Aan een groep beoordelaars wordt gevraagd aan te geven bij welke gelukswaarde op een lijn van 0 tot 10 naar hun mening de grens ligt tussen bijvoorbeeld “niet zo gelukkig” en “tamelijk gelukkig” als twee aangrenzende antwoordmogelijkheden op dezelfde verbale schaal in hun moedertaal. Op basis van hun opvattingen daaromtrent kunnen dan in een later stadium steekproefgegevens worden verwerkt tot gegevens betreffende de geluksverdeling in de populatie, mits van dezelfde natie, in elk geval in dezelfde taal. Ook op steekproefresultaten uit het verleden kan deze aanpak in principe alsnog worden toegepast.

Drie modellen met uiteenlopende validiteit zijn hiervoor ontwikkeld en één daarvan wordt aanbevolen voor verdere toepassing. Het innovatieve van deze aanpak is dat, voor zover bekend, deze het voor het eerst mogelijk maakt om het geluk als continue variabele te behandelen en bovendien daarvan de cumulatieve frequentieverdeling rechtstreeks te schatten, althans een aantal punten daarvan.

Hoofdstuk 6 is gewijd aan de problemen die aan de heersende praktijk kleven, vervolgens aan de principes van bovengenoemde schaal-interval benadering en ten slotte aan de diverse modellen die op deze basis ontwikkeld zijn voor de verwerking van waarnemingsresultaten die bij de geluksmeting verkregen zijn. De methodologische voordelen van de aanbevolen werkwijze ten opzichte van de conventionele aanpak zijn een betere validiteit van de resultaten en de mogelijkheid om de onnauwkeurigheid te kwantificeren van de verkregen schattingen van de parameters van de geluksverdelingen, zowel binnen als tussen naties.

In het slothoofdstuk worden honderd van de eerste “cases” geëvalueerd. Een case is in dit verband gedefinieerd als één van de 3-10 verschillende geluksvragen die behandeld worden door dezelfde groep beoordelaars binnen een beperkt tijdsverloop. De resultaten van deze verkenning bieden goede perspectieven. De precisie van de uitkomsten wordt vooral bepaald door het aantal beoordelaars per case. Wel blijkt het belang van goede instructie; dit geldt voor zowel de beoordelaars zelf als de proefleider ter plaatse. Voor een deel kan daarmee wellicht de uitval (11 %) verminderd worden. Opvallend is dat de redenen om beoordelingen af te keuren vooral aan de ‘ongelukskant’ van de geluksschaal optreden. Er zijn goede redenen om van een aantal anomalieën de oorzaak te zoeken in een verkeerd opgezette combinatie van vraag en antwoordcategorieën. Op deze wijze kan de schaal-interval methode in beginsel de mogelijkheid bieden om goed bruikbare geluksvragen te onderscheiden van de minder deugdelijke exemplaren.

Dankbetuiging

(Acknowledgements)

Bij de voltooiing van mijn proefschrift wil ik graag een woord van dank richten aan alle mensen die op een of andere wijze hebben bijgedragen aan mijn promotie. Een klein aantal hunner wil ik hieronder in het bijzonder noemen.

Allereerst mijn promotor Ruut Veenhoven. Hij was degene die mij in 2008 op het spoor met bestemming promotie heeft gezet. De inhoud van dit proefschrift is de neerslag van mijn bijdrage aan een jarenlange samenwerking met veel discussies. Ik besef terdege dat die niet altijd eenvoudig moet zijn geweest met een eigenzinnige bèta, wiens denkbeelden ook niet altijd spoorden met de in dit onderzoeksgebied gangbare. Gelukkig hebben de persoonlijke verhoudingen daar niet onder geleden. De (deel)tijd die ik als vrijwilliger ondersteunend meegewerkt heb aan het geluksonderzoek heeft dan ook onmiskenbaar bijgedragen aan mijn eigen geluk en wordt met deze promotie hopelijk niet afgesloten.

Aan de gesprekken met mijn co-promotor Lidia Arends bewaar ik heel plezierige herinneringen. Zij waren altijd bemoedigend en gaven mij het vertrouwen dat ik met mijn soms onorthodoxe – of moet ik juist zeggen orthodoxe ? – benadering inhoudelijk toch op de goede weg zat. Het was voor mij geruststellend te weten dat er vanuit de statistiek kritisch naar mijn bevindingen gekeken werd. Daarnaast betoonde Lidia zich zeer vasthoudend om onze gezamenlijke artikelen daadwerkelijk in de wetenschappelijke vakpers gepubliceerd te krijgen, en met succes. Haar bijdrage aan het promotietraject heb ik dan ook zeer gewaardeerd.

Niet alle problemen kon ik op eigen kracht oplossen. De fraaie oplossingen van één daarvan, namelijk die op bladzijden 68 en 69, zijn vervaardigd door Dr. Roel Stroeker en ik ben hem daarvoor zeer erkentelijk, evenals Professor Harm Bart voor zijn bemiddeling daarbij.

Ofschoon mijn medevrijwilligers van de"Onderzoeksgroep Geluk" met hun talloze gesprekken zeker ook bijgedragen hebben aan mijn meningsvorming, zullen zij ongetwijfeld begrijpen waarom ik op deze plaats Joop Mulder apart bedank. Hij heeft tal van computertechnische problemen voor mij opgelost en heeft mij bijgestaan bij de verwerking van veel gegevens tot iets waarmee ik daadwerkelijk aan de slag kon. Al met al te veel om op te noemen. Hartelijk dank, Joop, voor al deze hulp.

I am also indebted to Miranda Aldham-Breary for her contribution to this dissertation, which was an increase of its quality through a substantial shift towards “Right English”. Miranda, many thanks for your linguistic advices on this.

De vormgeving van een proefschrift als dit brengt de nodige problemen met zich mee. Gelukkig heb ik daarvoor een beroep kunnen doen op de professionele bekwaamheden van Joop van Opijnen. Het was vanwege de vele formules en vergelijkingen voor hem geen gemakkelijke opdracht, maar zijn inspanningen hebben geleid tot een resultaat waar wij allebei trots op mogen zijn. Slechts wie mijn aangeleverde ‘manuscripten’ heeft kunnen vergelijken met het eindresultaat zal ten volle kunnen begrijpen hoe erkentelijk ik Joop voor zijn bijdrage ben.

Het proefschrift zelf moge dan zonder hen tot stand gekomen zijn, aan de promotieplechtigheid zullen onze beide zonen Reinier en Guido wel een bijdrage leveren. Zij zullen hun vader als diens paranimfen ter zijde staan en het verheugt me dat ze de uitnodiging daartoe met graagte aanvaard hebben.

Het laatste dankwoord is bestemd voor Marian, de vrouw waarmee ik nu al bijna vijftig jaar lief en leed deel. Dat geldt zeker ook voor het promotietraject in de afgelopen twee jaar. Het stemt me tot grote dankbaarheid dat dit mogelijk was en dat ik dit proefschrift aan haar kan opdragen.

Wim Kalmijn

Curriculum Vitae

De schrijver van dit proefschrift werd geboren op 28 juni 1934 te Overschie, tot aan de annexatie in 1941 een randgemeente van Rotterdam. Vanaf 1945 bezocht hij het Marnix-Gymnasium te Rotterdam en behaalde daar in 1951 het bèta-diploma. Vervolgens studeerde hij scheikundige technologie aan de toenmalige Technische Hogeschool, thans Technische Universiteit te Delft. Van 1954 tot 1960 was hij als student-assistent verbonden aan het laboratorium voor kristalkunde van de afdeling Mijnbouwkunde. In 1960 behaalde hij het diploma scheikundig ingenieur. Tijdens het laatste gedeelte van zijn studie was hij tevens werkzaam als leraar natuur- en scheikunde, eerst aan het Johan de Wit Gymnasium in Dordrecht en daarna aan het Gemeentelyceum in Vlaardingen.

Van 1960 tot 1990 was hij in dienst van het Unilever Research Laboratorium te Vlaardingen. Hij hield zich achtereenvolgens bezig met fysisch-chemisch onderzoek, personeelszaken en opleiding, statistiek en proefopzetten en tenslotte met informatiebeheer. In de periode 1966 -1967 onderbrak hij zijn werkzaamheden gedurende ruim een half jaar voor een opleiding tot statisticus door het Mathematisch Centrum te Amsterdam. Van 1969 tot 1981 leidde hij de sectie Statistiek van genoemd laboratorium. In de laatste vier jaar van deze periode was hij tevens secretaris van de Vereniging voor Statistiek en van de Stichting Opleidingen Statistiek.

Tussen 1992 en 1995 doceerde hij statistiek, informatiekunde en informatiesystemen aan studenten van de Stichting Opleiding Sociale Arbeid, destijds een onderdeel van de Hogeschool Haarlem.

Na een oproep in het jaarprogramma 1998-1989 van HOVO (Hoger Onderwijs Voor Ouderen) van de Erasmus Universiteit Rotterdam trad hij als vrijwilliger in deeltijd toe tot de “Onderzoeksgroep Geluk” onder leiding van Prof. dr. Ruut Veenhoven en behorend tot de Faculteit der Sociale Wetenschappen van de Erasmus Universiteit. De inhoud van dit proefschrift is de neerslag van een deel van de werkzaamheden die in dit kader zijn verricht ter ondersteuning van genoemd geluksonderzoek. De meeste hoofdstukken zijn bewerkingen van een aantal publicaties in sociaal-wetenschappelijke tijdschriften.

Wetenschappelijke publicaties vóór 1998:

- Kalmijn, W.M., *Confidence Intervals for Batch Properties Based on Both Sampling and Analytical Variance*, Statistica Neerlandica **24** (1970), 133 -141.
- Kalmijn, W.M., *Statistiek bij het opzetten van proeven*, Voedingsmiddelentechnologie, **7** (nr. 50, 11 dec. 1974), 20-23.
- Kalmijn, W.M., *Simultaneous Confidence Regions for Repeat Preference Testing*, Applied Statistics , **25** (1976), 117-122.

What are the three major contributions of this dissertation to Happiness Research?

I The understanding how and why happiness and income are essentially two different kinds of variables, each of which with has its own appropriate inequality measures. Standard deviation is shown to be an adequate inequality measure for happiness distribution in samples, and e.g. the Gini coefficient is definitely not. The reverse applies to income inequality. This knowledge enables us to prevent or to remove inappropriate barriers in publications on these subjects.

II A measure for judging the happiness in a nation allowing for both an utilitarian and an egalitarian view to happiness. This "Inequality-Adjusted Happiness" index is a linear combination of the mean happiness in a nation and the inequality within this nation, expressed in the standard deviation. The use of this statistic may also circumvent barriers, this time in discussions with policy makers on how to improve the performance of their nation with respect to its happiness situation.

III A better method for converting the happiness measurements in a sample, using 'verbal questions', into more valid and valuable information on the happiness distribution in the population represented by this sample. This method concerns inequality both within and between nations.
Various serious methodological objections are raised against the commonly applied procedures. Therefore, an alternative approach has been developed on the basis of the view of happiness as a continuous variable and of partitioning the two-sided bounded happiness scale into a small number of contiguous intervals, each corresponding to one of the response categories and with empirically determined boundaries between these intervals.
As far as known, this approach is the first to consider and treat happiness systematically as a continuous variable. Moreover it is the first method in which the cumulative happiness probability distribution in the population is estimated directly, at least some points of this distribution curve.
This approach also enables us to subject previous happiness information to a re-analysis. Moreover it improves the possibility to compare, and sometimes to combine, results from different nations. Finally, the method offers the opportunity, at least in principle, to separate the useful from the ill-constructed items among the too numerous happiness measures used in this field.

The interest of social scientists in happiness as a research object has increased over the last four decennia. Happiness is defined in this context as the extent to which an individual judges his satisfaction with his own life as a whole. Individuals are different in this judgement. Happiness research focuses in particular on conditions that are related to happiness and which may be responsible for these differences, or at least for a part of these. Most of these conditions are societal and therefore the interest of sociologists concerns the statistical distribution of happiness within and between nations or other societies rather than individual happiness.

Happiness research requires happiness to be measured and to be quantified. By definition this measurement is always made at the individual level and the usual way to do so is by simply asking the individual to rate his own happiness. The conversion of individual responses to this question in a sample into valid and useful information on the happiness distribution in the population represented by this sample is a complicated statistical problem. In his dissertation we describe and evaluate various methods to bridge that gap, providing a methodological contribution to happiness research. For a more detailed specification of this contribution PTO.

Wim Kalmijn (1934) is a part-time member of the Happiness Research Group of the Erasmus University Rotterdam (The Netherlands) and is more specifically involved in statistical and other methodological problems.