MEASURING INEQUALITY OF HAPPINESS IN NATIONS
In search for proper statistics

Wim Kalmijn and Ruut Veenhoven

ABSTRACT

Comparative research on happiness typically focuses on the level of happiness in nations, which is measured using the mean. There have also been attempts to compare inequality of happiness in nations and this is measured using the standard deviation. There is doubt about the appropriateness of that latter statistic and some prefer to use the statistics currently used to compare income inequality in nations, in particular the Gini coefficient. In this paper, we review the descriptive statistics that can be used to quantify inequality of happiness in nations. This review involves five steps: (1) we consider how happiness nations is assessed, (2) next we list the statistics of dispersion and considers their underlying assumptions; (3) we construct hypothetical distributions that cover our notion of inequality; (4) we define criteria of performance and (5) we check how well the step-2 statistics meet the step-4 demands when applied to the step-3 hypothetical distributions We then applied the best performing statistics to real distributions of happiness in nations. Of the nine statistics considered, five failed this empirical test. One of the failed statistics is the Gini coefficient. Its malfunction was foreseen on theoretical grounds: the Gini coefficient assumes a ratio level of measurement, while happiness measures can at best be treated at the interval level. The Gini coefficient has been designed for application to 'capacity' variables such as income rather than to 'intensity' variables such as happiness. Four statistics proved to be satisfactory; these were (1) the standard deviation, (2) the mean absolute difference, (3) the mean pair difference and (4) the interquartile range. Since all four statistics performed about equally well, there is no reason to discontinue the use of the standard deviation when quantifying inequality of happiness in nations.

Keywords: Social inequality, happiness, inequality in quality-of-life, measures of dispersion, standard deviation, Gini coefficient.

1   INTRODUCTION

Happiness is an increasingly common theme in cross-national research. The first comparative study on happiness was carried out by Cantril (1965) in 1960, and involved representative samples for 14 nations. 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. There have also been large-scale cross-national comparisons of happiness among university students by Michalos (1991) and Diener et al. (1995). The findings of all this research are gathered in the World Database of Happi-

ness (Veenhoven 2004). To date, the section on distributional findings of this database lists the results of 2702 surveys among general population samples in 116 nations and provides time series for more than 20 years for some 15 nations.

Level of happiness in nations

 Such data are commonly used to identify the societal determinants of happiness, and in particular to appraise the impact of disputed matters such as economic growth, state welfare and political freedom. Analyses of this kind have been reported by e.g. Diener (1995), Inglehart (1990, 2000), Graham & Pettinato (2001) and Veenhoven (1984, 2000). In addition to these correlational analyses, there are also trend studies, many of which deal with the question of whether economic growth makes us any happier in the end e.g. Easterlin (1974), Lane (2000), Hagerty &Veenhoven (2003).

Most of this research is inspired by the utilitarian quest for the 'greatest happiness of the greatest number' and therefore focuses on the level of happiness in nations. This level is quantified using statistics with a 'central tendency' or '' 'location' and in particular a mean score for such items.

Inequality of happiness in nations

The cross-national differences in happiness can also be considered from an egalitarian perspective. Here the focus is not on the level of happiness in the country but on the inequality in happiness; that is: on the degree to which citizens in a country differ in the enjoyment of their life[1]. This calls for another kind of statistics; instead of central tendency statistics we need to use statistics that describe the 'dispersion' of the distribution.

The happiness research group at Erasmus University has explored this issue of inequality of happiness in nations. The first study concerned the impact of the 1980 economic recession on inequality of happiness in West-European countries and found no systematic effect (Chin Hon Foei 1989). The second investigation considered dispersion of happiness in 28 nations and looked at the pattern of concomitants. This study found systematic differences (Veenhoven 1990). Later studies used data on inequality of happiness in nations to test theories of happiness (Veenhoven & Ehrhardt 1995) and to check for the possible equalizing effects of a welfare state (Veenhoven 2000b).

All these analyses measured inequality of happiness in nations 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 2702 surveys that involved questions on happiness can be found, and where the standard deviation is also used for nation rankings and time trends[2].

The issue

There are qualms about the use of standard deviation for characterizing inequality of happiness. Investigators in the neighboring field of income inequality tend to prefer the Gini coefficient and some of them argue that a standard deviation is not independent of the mean. This criticism has been voiced at conferences and in a review of a paper submitted to a scholarly journal.

The main reason for looking at inequality of happiness in nations is in the possibility that this may reveal differences across nations other than those observed for the level of happiness. If so, this would mean that an egalitarian policy aimed at reducing differences in happiness would differ from a utilitarian policy aimed at producing a higher average level of happiness. To check this possibility we need measures for the

general happiness level of a nation or nations and for inequality of the same, which are mutually independent, at least ideally.

It is worth taking a fresh look at the statistics for inequality, since the comparative research on this matter is very much focused on inequality of income, and the statistics commonly used for quantifying this variable may not be the best for charting inequality in other matters. 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.

Plan of this paper
In this paper we take a look at of the statistics that are in use for quantifying inequality of happiness in nations and investigate their merits in a systematic way. We start by considering how happiness in nations is assessed (Section 2). Next we list the statistics of dispersion that apply to this methodology and briefly discuss their characteristics (Section 3). After a number of methodological considerations (Section 4) we specify what we understand as inequality and on this basis we construct series of hypothetical distributions that vary in degree of inequality (Section 5). We then define criteria for performance and subsequently assess how well the various statistics meet these demands when applied to the constructed distributions. The statistics that pass the test are then applied to real distributions of happiness in nations (Section 6). Our conclusions are presented in Section 7.

2   MEASUREMENT OF HAPPINESS

Happiness is typically measured by self-report and cross-national studies on happiness mostly used 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, for instance 'very happy', 'fairly happy', 'not too happy' and 'unhappy'. Happiness is also rated on pictorial scales using smilies and on graphical scales such as Cantril's (1965) 'ladder scale'. Whatever the scale used, the respondent has to select one out of a limited number of discrete ratings. By definition, people who select different ratings are unequally happy. People with the same rating are considered to be equally happy; happiness differences between such people may exist, but are not measurable and are ignored in this context.

Inequality of happiness in a society includes two elements: (1) the relative numbers of subjects that select a different happiness rating and (2) the magnitude of each of these differences. These judgments have to be averaged for society as a whole

This *actual* inequality in happiness is not necessarily the same as the inequality citizens *believe* to exist. A person may think that inequality is great and rising in the country, while the differences are in fact small and diminishing. In this paper we focus on actual inequality of happiness in nations.

3 STATISTICS OF INEQUALITY

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

3.1 Range

The *theoretical* range of a distribution is the difference between the highest and the lowest *possible* rating, in the above example the difference between 10 and 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.

3.2 Average deviation from the mean

For any observation, the deviation from the mean may be positive, negative or zero, 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. However, this *mean absolute deviation* is a rather obsolete measure, since its mathematical tractability and its relationship to distribution model parameters is quite complicated (Kendall and Stuart, 1977: 44).

3.3 Variance and its square root, the standard deviation

The standard solution for the problem of getting rid of the algebraic sign is squaring 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 former is that it is expressed in the same unit as the basic observations.

3.4 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.

3.5 Interquartile range

The interquartile range (or interquartile distance) is the difference between the third and the first quartile.

When applying this statistic, the assumption is made that happiness should be considered as a continuous variable, which is measured using a discrete rating scale: someone selects a rating of 6, if they consider the label of that rating as a better qualification of their happiness feeling than the adjacent ones, even if the label of rating 6 does not fit perfectly. In this way, rating 6 covers a 'range' that falls symmetrically about that number, say from 5.50 to 6.50.

If a large part of the distribution is concentrated in one or both most extreme ratings, a problem may occur. This is caused by the fact that on a [1; 10] scale the ratings 1 and 10 represent half the 'interval width' of the other: [1.00; 1.50] and [9.50; 10.00] respectively. Therefore, one may expect an anomalous behavior for this statistic in cases when the majority of the subjects selects one of both extreme ratings, i.c. either 1 or 10. However, such extreme distributions have not been observed in the case of happiness in nations.

3.6   Mean pair distance

The mean pair distance 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 $\frac{1}{2}N(N-1)$ 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 Section 2.

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 still an underlying assumption.

The computation required for this statistic is relatively tedious, but with today's computers, this is not a serious drawback.

3.7   Gini coefficient

The Gini coefficient, also referred to as *Gini's concentration ratio*, (Gini 1912) is currently used to characterize inequality in income distributions, see also Theil (1967: 121 –128). For any income distribution, one can construct a *Lorenz curve*. In principle, each point of this curve corresponds with 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, more precisely: a broken line, through the points (0,0) and (1,1), but generally all other points will be expected to be found below the diagonal through (0,0) and (1,1). The closer the Lorenz curve is to that diagonal, the less the income inequality of the sample or population concerned.

The *Gini coefficient* (*G*) 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 $\frac{1}{2}$. Clearly $G = 0$ 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 $0 < G < 1$ and G 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 G is bounded to the interval [0; 1] is certainly an advantage when comparing different values of inequality.

Assumed distribution model
Various distribution models have been proposed for income distribution. The most widely applied are the 'Pareto distribution' and the 'log-normal 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 mathematical distribution model.

It appears that in both cases mentioned above, this relationship is a relatively simple one and it is this finding that 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 their model parameters, and this may make their applicability to such situations less attractive.

Moreover, happiness distributions are essentially different from income distributions, since the measured happiness is bounded on both the low and the high values. This makes Pareto and log-normal distributions invalid models for happiness distributions.

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 can be obtained by dividing the mean pair distance (with unity power) by $2mN/(N–1)$; for large values of N, this divisor equals approximately twice the mean happiness rating. For $m = 0$, $G := 0$.

3.8   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:

$$T = \ln(N) + \sum_{i=1}^{k} \left(\frac{i \, n_i}{Nm} \right) \ln \left(\frac{i}{Nm} \right)$$

where $\ln(N)$ indicates the natural logarithm of the sample size N, n_i is the absolute frequency of the ratings with value i ($i = 1(1)k$ [3]; $\Sigma n_i = N$), m is the mean happiness value and Σ denotes a summation over all k possible ratings i (Sen, 1997: 34 - 36). If all respondents report the same rating (complete equality), $T = 0$.

.

3.9   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 percentage in 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) $100 \cdot (k-1)/k$ in the case of a uniform distribution (Series 5, Table 2d below).

4.   METHODOLOGICAL CONSIDERATIONS

As noted in the introduction, this paper was prompted by the claims that measures used for assessing inequality of income in nations should also be used for measuring inequality in happiness. This criticism assumes that these inequalities are of the same kind. Yet we see fundamental differences between the two, which has consequences for the applicability of inequality statistics.

4.1   Distribution of numbers or of quantities?

When talking about 'distributions' we can easily overlook subtle differences in kinds of distributions. One of these differences is that between distributions of 'numbers' and of 'quantities'.

In the case of happiness, we are dealing with 'number distributions'. Taking the happiness rating 3, we associate this with the relative frequency of the *number of subjects* that selected the response option 3. Similarly, the cumulative relative frequency is the relative number of subjects selecting a rating ≤ 3.

When applied to incomes, one mostly considers distributions of quantities. The relative frequency for an income class number 3 can be defined as (a) the *fraction of subjects* with an income between the boundaries of class 3, just as for case of happiness, but also as (b) the fraction of the *total income* of the population that is earned by all subjects in class 3 together. In the latter case - the income distribution in a more narrow sense - the total income of the population is considered to be a quantity that can be distributed over the people in the various income classes. The latter choice gives some more emphasis on the distribution of a total income rather than the former, which considers the statistical distribution of the individuals for their incomes.

In this discussion on inequality of happiness in nations, we will avoid the interpretation under (b). Unless stated otherwise, we speak of distributions as a 'number distribution'. The reasons are explained below in Section 4.2.

4.2   Variable of 'capacity' or 'intensity' ?

Incomes can be measured by *counting* the number of Euros earned by a subject or a set of subjects over a period. Annual incomes of individuals can be aggregated, e.g. to a household income or to a national annual income. Incomes can be transferred from one person to someone else, e.g. in the case of taxation within one family, but also e.g. when a plumber appoints an assistant.

Theoretically an approach similar to the second one given in Section 4.1 for income distributions can be applied to happiness, provided one is prepared to assume the existence of something like a total amount or 'quantity' of happiness, consisting of a kind of 'happiness molecules'. Thinking along this line, Ng (1996: 1-28) assumes that everybody acquires a number of such units ('utils') of happiness, but also of unhappiness. A subject rating of his or her happiness is supposed to count the number of such 'pleasure utils', collected over a period, as a measure of his or her happiness. In the same way, counting the number of 'pain utils' over the same period results in an amount of unhappiness. In this approach, unhappiness is just negative happiness. The net sum of the happiness and unhappiness utils quantifies the subject's average happiness over that period. According to Ng, in this approach not only cardinal measurement is justified, but even at the ratio level of happiness measurement.

In the physical sciences, a distinction is made sometimes between *capacity variables* and *intensity variables*. An example may clarify this distinction. Suppose one has two vessels. The first one contains 3 liters (i.e. 3 kg) of water at a temperature of 40 °C and with a density of (almost) 1000 kg/m^3. The second is filled with 4 liter of alcohol at 30°C, having a density of 790 kg/m^3 and a mass of 3.16 kg   If both liquids are poured into a third container, we get a mixture with a volume of almost 3 + 4 liter = 7 liter, with a total mass of 3+3.16 kg = 6.16 kg. However, measurement of the temperature will not result in 40+30 = 70 °C, nor will the new density be about 1790 kg/m^3.

Variables such as volume and mass are referred to as *capacity variables*; their values may be largely added in cases of combination of quantities. Temperature and density are examples of *intensity variables*. When systems are combined, one may generally expect a result for intensity variables that is close to the weighted average value, albeit only provided one ignores 'interaction situations', in casu situations in which phenomena like contraction and/or chemical reactions occur.

The fact that one can measure annual incomes by *counting* the number of Euros implies that an annual income can be considered to be a capacity variable. In contrast, we consider measures of happiness to be intensity variables. In our view, rating your happiness is not just counting your blessings. Some one who is unhappy has a low happiness level. Being able to 'transfer happiness' is something that is difficult to imagine. This difference has the consequence that operations with respect to income cannot always be automatically transferred to happiness situations.

A related aspect is that annual incomes have a natural zero, but do not have a theoretical maximum value; whereas we see happiness as a bounded variable (cf. Section 2) and in this respect we differ from Ng (1966). Our conclusions and recommendations should be judged in this context.

Above, in Section 4.1, we introduced two possible ways of defining a happiness distribution. Although happiness distributions in the sense of the second type are mathematically possible, they were not used as models for the happiness studies discussed here. The Gini coefficient, however, was introduced, based on of the Lorenz curve. When applied to a happiness distribution, the abscissa of a point on the Lorenz curve represents the relative cumulative frequency of the happiness distribution in the first definition, the number distribution, for some happiness rating, whereas the same statistic for the same happiness rating, but for the second type of distribution, is plotted as the corresponding ordinate. Apparently, the Gini coefficient includes both types of distribution and this makes it debatable as to how useful it is for our purposes.

The application of Theil's inequality to happiness also assumes the existence of some 'total amount of happiness', which has been distributed over all members of the society in question. In the equation in Section 3.8, the ratio i/Nm is considered to be the relative 'share of the total amount of happiness' of an individual that selects a rating i.

4.3   Level of measurement of happiness

The way of measuring happiness described in Section 2 is performed at the *ordinal level of measurement* as defined by Stevens (1946). The responses are converted using a monotonous transformation to a numerical scale with nonnegative integers as possible ratings. Usually, these ordinal numbers are treated as *if they were* cardinal numbers, which implies that happiness ratings are postulated to be equidistant. For justification see e.g. VanPraag (1991), Ng (1996, 1997), VanPraag & Ferrer-i-Carbonell (2004: 319).

As soon as this assumption has been made, the rating numbers can be treated, as if they would have been obtained at the metric level of measurement, at least at the interval level of measurement. This allows the application of mathematical operations, like addition, subtraction etc., which in turn is a condition for calculating statistics as averages[4] and standard deviations as admissible operations. The application of statistical methods that do not assume equidistance, such as 'ordered probits', appears to produce almost identical results (Frey & Stutzer 2002). In this way, happiness ratings are treated as measurements, obtained at the *interval level of measurement.* The measurement is not made at the ratio level, since there is no natural 'zero level' of happiness and on the basis of these ratings one cannot say, "John is twice as happy as Peter". Again this conclusion does not fit in the approach of Ng and others, who assume that happiness is measured at the ratio level of measurement, in a similar manner to income.

An important property of income is that it is not expressed as a number only, it is expressed as a combination of a number and a *unit*, e.g. as $14,000 per year, it has a '*dimension*' in the sense in which physicists use this term, in casu $/year. This means that some statistics used for income distributions, like the average value and the standard deviation, also have this dimension, irrespective of whether one is dealing with a distribution of numbers or of quantities. Some of the inequality statistics enumerated in Section 3 have this dimension, but others are dimensionless. The latter class includes the coefficient of variation, the Gini coefficient and Theil's inequality measure. Changing the dimension of the variable, in casu to the monthly income 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.

One has to bear in mind that each of the different statistics listed in Section 3 requires a minimum level of measurement of the happiness ratings as follows:

Level of measurement	Statistic of inequality
Ratio level only	Coefficient of variation
	Gini coefficient
	Theil's entropy measure
At least interval level	Standard deviation
	Mean absolute difference
	Mean pair distance
	Interquartile range
	Range
All levels	% outside mode

4.4   Dependency of inequality statistics on the mean value

As noted in the introductory section, we need a measure of dispersion that is independent of, or at least not strongly dependent on, the central tendency, e.g. the mean. However, in this context, it is necessary to distinguish between various kinds of dependency. This is illustrated below for the situation where the standard deviation is used as a statistical test 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 covariance between the statistics.

If a random sample consists of observations from a normally distributed population, the statistics will have zero covariance and will be stochastically independent. (Cramér, 1946: 382). If the distribution is skewed, there is stochastic dependency, in the sense that, for positively skewed distributions, 'skewed to the right', higher mean values correspond with 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.

This kind of dependency deals with the situation in which different samples from the same population are taken. In this case, 'independency' means that, when due to sampling errors (accidentally) higher average value is obtained, this does not give rise to a systematically higher or lower value of the estimated standard deviation.

Structural dependency
A different type of dependency is that which 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 (Section 2), the variance (s^2) and the standard deviation (s) have theoretical maximum values that depend on the mean value m. See Appendix A, where it is demonstrated that:

$$0 \leq s \leq \sqrt{(h - m) \cdot (m - u)} \leq \tfrac{1}{2} \cdot |h - u|,$$

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 for 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 on a scale ranging from 0 to 10 [5] are listed in Table 1 below as an example.

Apparently, in this case for $2 \leq m \leq 8$, 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 (a) only modestly dependent on the value of m, and (b) fairly remote from almost all empirical values of the statistic in nations studies (1.3 – 2.5 on this scale).

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. If e.g. 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 case.
     In practice, the latter kind of structural dependency between m and s is known from measurement errors in some technical measurements, such as that of pressure, temperature, and electric flow rate etc. Until now, no indications have been obtained that such a dependency occurs in happiness measurement.

Intrinsic dependency
A third type of dependency is the one that arises from the fact that in a society there may be spontaneous or forced *socio-economic mechanisms* acting in a society that 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 com-

paring a number of nations. When Ott (2005) reports a correlation coefficient of -0.66 for a set of 80 countries, this is obviously a case of intrinsic dependency. Clearly, the inequality statistic cannot be made accountable for his observation.

Consequences

This distinction makes clear that the initial objective, to select an inequality statistic that is not, or only weakly dependent on the mean happiness, needs a more precise formulation. When judging the performance of a statistic, only the stochastic and the structural dependency should be taken into account. Therefore, in this context intrinsic dependency has to be ignored completely.

Having refined the dependency concept in this context, 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 fairly weak. Only at very extreme mean values, which usually also give rise to strong skewness, might this dependency be a problem that would deserve more attention. As long as such societies are not found, we can ignore this issue.

Consequently, it is the presence of the mean value in the denominator of a happiness inequality statistic that introduces its a dependency on the mean, which in this context is most undesirable.

4.5   Consequences for measuring inequality in happiness

Since we see happiness as an intensity variable, we foresee problems with statistics that assume capacity 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.

5   CASES OF INEQUALITY

The goal of this investigation was to assess the performance of the statistics reviewed in Section 3. For this purpose, we needed examples of more and less unequal distributions, and we made the choice to use constructed distributions. It could be argued that such distributions are artificial and that therefore the conclusions do not necessarily apply to real situations. Yet constructed distributions have known properties and this enables us to understand why statistics perform as they do.

We devised a number of artificial happiness distributions, most of them with the number of possible different ratings (k) = 10 and all with the effective sample size (N) = 100. Next, we ranked these distributions from least to most unequal. Decision rules were applied in this ranking. These decision rules can be seen as an explicit account of our notion of inequality.

5.1   Notions of inequality

The term 'inequality'[6] is used with slightly different meanings. At its most basic, it refers to the proportion of people who are unequally happy, irrespective of the size of the difference and of the mean level of happiness. We call this 'nominal inequality'. In a wider sense, the term also takes into account the (absolute) sizes of the differences. We call this 'weighted inequality'. Sometimes, the term also denotes inequity in happiness. In this use of the word, inequality is combined with level of happiness. We refer to this as 'pseudo inequality'.

Nominal inequality
If the happiness of individuals is measured at the nominal level of measurement, e.g. as either happy or unhappy, their measured happiness can only be either equal or unequal; this is why we have adopted the term 'nominal inequality'. However, the concept of nominal inequality is still applicable at other levels of measurement. If the happiness of individuals is unequal, the magnitude of the difference between the ratings is ignored in this case. Although the individual differences have no gradations, this does not hold for the inequality of the distribution in a sample.

Suppose that happiness is rated on a [1; 10] scale in a sample. The largest possible inequality occurs when each of the ten possible ratings has been selected by just 10 % of all subjects. Then for each subject, 90 % of the total sample has selected a rating that is unequal to his/hers. So on an average, the inequality statistic defined above, is also 90 %, which is apparently the highest possible value.
If however, 95 % of the subjects have selected one out of the ten possible ratings, and 5 % all selected a second one, for each subject of the majority only 5 % selected an unequal rating and for each subject of the minority 95 % did so. In this case, the weighted average value is only 9.5 %, which implies a much lower degree of inequality than in the case of the uniform distribution.

When dealing with inequality of this kind, we judge the degree of inequality on the basis of the following rules.

- Zero-inequality occurs when everybody selects the same rating.
- The highest inequality at this level is found in the case of the uniform distribution.
- 'Flat' distributions have a higher inequality than that with a pronounced top.
- 'Wide flat' distributions are more unequal than 'narrower flat' ones;
- If there is more than one pronounced top, inequality is higher as there are more tops and as they differ less in height.

Weighted inequality
When used in this sense, the term 'inequality' refers also to the magnitude of differences between individuals. A difference between score 4 and score 7 on the 10-step happiness scale is then three times as unequal as a difference between score 4 and score 5. This kind of inequality is reflected in the mean pair distance statistic in Section 3.6. We also take this kind of inequality into account and that brings us to one more rule for judging inequality in distributions.

- If there are two or more tops, inequality increases with their distance.

Pseudo inequalities

Sometimes the word inequality is also used to denote relative deprivation, for example when one deems a country more unequal when 80% of its inhabitants are unhappy (e.g. rating 4) and 20% are happy (e.g. rating 8) then in the reversed case of 80% happy and 20% unhappy. In this use of the inequality, level and dispersion are mixed up. Here equality is confused with equity.

Since we are looking for a measure of inequality that is independent of the mean (cf. Section 1), we will stay away from this notion. In this context, we consider the 80% - 20% distribution as equally unequal as a 20% - 80% distribution.

Another case of pseudo inequality occurs when a distribution of 10% unhappy and 90% happy is considered as more unequal than a 50% - 50% distribution with the same difference between the two ratings, because being unhappy in the face of a happy majority would be more difficult to bear than the reverse situation. This may be true in the case of income, but this reasoning does not apply for happiness because such effects, if any, are already included in the overall evaluation of life. In fact, in this case the 10 - 90 distribution is less unequal than the 50 - 50 one, since the average percentages of subjects that select a different rate are 18 and 50 respectively.

It is important to stress that these rules stem from our view on inequality and follow our earlier choice to consider happiness as an intensity variable. What we consider here as 'equally unequal' will not necessarily apply to an approach in which a total amount of happiness is assumed to be distributed over a number of individuals. This is how we see inequality in happiness; other views may lead to different conclusions.

5.2   Hypothetical distributions

We started with an attempt to devise a series of hypothetical distributions that vary from most to least possible inequality. This appeared to be difficult; since the above rules denote different aspects of inequality one cannot combine all these notions in a single series of distributions. We ended up with six series of increasing inequality, each of which highlights one of the above aspects of inequality. These distributions are presented in the upper parts of the Tables 2a, 2b, 2c and 2d.

All the distributions are denoted with a letter and sometimes a number. Distributions we consider as 'equally unequal' are denoted using the same letter and different digits. Although the aptness of some statistics is, a priori, debatable on the basis of the discussions in Sections 3 and 4, we have applied all of them, except the range, to the constructed distributions to demonstrate their behavior in this context.

Series 1

The first series 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 uni-modal (A and B) to normal (C) and then flattens (D and E) and turns next into bi-modal (F, G and H). The series ends up 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 considered to be maximally unequal.

The distributions C1 and C2 are included because they are (almost) normal, but with different mean values, although their exact position within Series 1 is somewhat arbitrary and can be debated[7].

Series 2
Series 2 is constructed to detect a possible dependency on the mean. The series starts again 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).

Series 3
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.

Series 4
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').

Series 5 and 6
Lastly Series 5 and 6 were 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 has been described by Veenhoven & Kalmijn (2005, Appendix A).

6   HOW THE STATISTICS PERFORM

We formulated eight evaluation criteria[8] to assess the usefulness of these nine statistics 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 5. These values are presented in the lower parts of Tables 2a-d.

6.1   Criterion 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

Performance:
All the statistics meet this condition.

6.2   Criterion 2: Interval level of measurement

We have seen in Section 2 that happiness is measured at the ordinal level, but can be treated as a pseudo-metric. Hence, a statistic should be applicable to the distribution of variables measured at the interval level of measurement.

Performance
Not all candidates require the measurement be made at the interval level of measurement. 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 those three statistics.

In principle, quartiles of a distribution can be measured without making the assumption that happiness is measured at the metric level. Treating the ratings as strictly ordinal variables enables the calculation of all quartiles as integer (ordinal) numbers. However, although in this situation the quartiles are defined, their difference is not, since subtraction of quartiles as ordinal numbers is not an admissible mathematical operation.

6.3   Criterion 3: Independence of scale range

There is much variation in scales used in happiness research and it is helpful when scores on different scales can be compared, if necessary after linear scale transformation. Hence the third criterion is that the statistic used be independent of the number of possible ratings on the scale of measurement or at worst weakly dependent.

Performance:
This requirement was tested in Series 5. In Table 2d one can see that, 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 one is found for Theil's measure. In all other cases the difference between $k = 5$ and $k = 10$ did not exceed 15 %. For non-uniform distributions larger differences are to be expected for the percentage outside the mode.

6.4   Criterion 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.

Performance:
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 $\ln(N)$.

6.5   Criterion 5: Independence of the mean

As noted in the introduction, 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.

Performance:

We announced that problems could be expected with three statistics: the coefficient of variation, the Gini coefficient and Theil's inequality measure in Section 4.5.
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 2a the distributions A1 and A2, B1 and B2 and C2 and C2. Going through these columns we can see that four statistics yield identical values: standard deviation, mean absolute deviation, mean pair distance and the 5 outside the modus. The same pattern emerges in the Tables 2b, 2c and 2d. The other statistics produce different values. These are: the coefficient of variation, the Gini coefficient, Theil's entropy measure and for extreme distributions 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 has 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 makes it an inadequate statistic for inequality of happiness.

6.6   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'. Inequality statistics should be invariant under operations such as 'translation' and 'reflection' of the happiness distribution along the happiness scale[9].

Performance:

The above test of independence of the mean (Section 6.5) involved a check on this criterion. 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 2d is particularly instructive and disqualifies the coefficient of variation and the Gini coefficient.

6.7   Criterion 7: Differentiation between more and less unequal distributions

Obviously a usable statistic must also distinguish between distributions we consider to be 'unequally unequal'. The indicator should also be sensitive to different concepts of inequality.

Performance:

A glance at the Tables 2a, 2b and 2c 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 2a that this statistic yielded the same values for the distributions D, F and G and also identical values for the distributions H and K. Likewise, in Table 2b we see identical values for the distributions B, L, M, N and K.

The mean absolute difference does not show a difference in inequality between the distributions C and D in Table 2a and nor between E and H.

Irregularities in Series 2 with respect to the Gini coefficient and Theil's measure arise from mean value differences[10].

6.8   Criterion 8: Sensitive to 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 our notion of what is more or less, unequal.

Performance:

The degree to which this requirement is met was judged by considering the degree to which a statistic yielded higher values for distributions we considered more unequal. A look at the Tables 2a, 2b and 2c shows that the values tend to get higher if we move from the left (A) to the right (K). 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 right of the Tables 2a, 2b and 2c.

Table 2a shows that most statistics reflect the degree of inequality fairly consistently, tau-b's vary between 0.8 and 0.9 and the mutual differences are modest. Examination of the values of the statistics suggests that all of them rank distribution E as more unequal than F, the first one in which bi-modality is introduced. If one interchanges the positions of the two distributions, this would give an interesting improvement in the rank correlation. The same holds for interchanging the distributions G and H, apparently the concentration in H gives rise to a lower value. Application of the above re-orderings gives into tau-B values of 0.90 for both the standard deviation and the mean pair distance and a tau-B value of 0.97 for the mean absolute deviation. However, as has been pointed out before, this statistic does not detect inequality differences between C and D or between or E and H in Series 1.

In the Series 2 and 3 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.

6.9   Overall evaluation

The performance of the nine statistics is summarized in Table 3. This overview clearly shows that five of the nine statistics considered can be disqualified as a means for testing 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.

Four statistics meet the demands required for determining inequality of happiness in nations. Appropriate measures are (1) standard deviation, (2) mean absolute difference, (3) mean pair distance and (4) interquartile range. There is no clear winner among the four suitable statistics; all perform about equally well.

6.10   Performance on real distributions

One may wonder how the statistics perform when applied to non-artificial distributions. As a matter of fact this is meaningful only for the statistics that were acceptable on the basis of our earlier findings. To this end, we selected the happiness distributions of eight Eastern European countries in 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 rating scale. The distribution data were taken from the World Database of Happiness, Section Nations.

The acceptable inequality statistics have been computed for each distribution and the nations have been ranked according to the standard deviations, resulting in Table 4.

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

The above picture supports the conclusion that no advance in understanding is to be expected from switching from using the standard deviation to some other inequality statistic.

7   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 three other statistics of disparity; i.e. the mean absolute difference, the mean pair distance and the interquartile range. Hence 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 a capacity variable on the basis of an 'amount of happiness', that can be distributed over the members of the society under study.

REFERENCES.

Buchanan, W. (1953),
How Nations See Each Other. A Study in Public Opinion,
University of Illinois Press, Urbana, USA

Cantril, H. (1965),
The Pattern of Human Concern,
Rutgers University Press, New Brunswick, New Jersey, USA

Chin-Hon-Foei, S. (1989),
Life Satisfaction in the EC Countries, 1975-1984,
Veenhoven, R., Ed.: "Did the Crisis Really Hurt?" Universitaire Pers Rotterdam, 1989,
Netherlands, 24 - 43

Cramér, H., (1946),
Mathematical Methods of Statistics,
Princeton University Press, Princeton, USA.

Diener, E., Shao, L., Smith, H.L. & Suh, E.M. (1995),
National Differences in Reported Subjective wellbeing: Why do they occur?
Social Indicators Research, 1995, Vol. 34, 7 - 32

Diener, E. & Suh, E.M. (2000),
Culture and Subjective Wellbeing,
MIT press, Cambridge USA.

Diener, E. (2000),
Subjective Wellbeing. The Science of Happiness and a Proposal for a National Index,
The American Psychologist, 2000, Vol. 55, 34 - 43

Easterlin, R.A. (1974),
Does Economic Growth Improve the Human Lot? Some Empirical Evidence,
David, P.A. & Melvin, W.R. (Eds.) "Nations and Households in Economic Growth", Academic Press, New York, USA, 89 – 12.

Ferrer-i-Carbonell, A. & Frijters, P. (2004),
How Important is Methodology for the Estimates of Determinants of Happiness?
Economic Journal (in press)

Frey, B.S. & Stutzer, A. (2002),
Happiness and Economics; How the Economy and Institutions Affect Wellbeing,
Princeton University Press, Princeton, USA

Gallup, G.P. (1976),
Human Needs and Satisfactions: A Global Survey,
Public Opinion Quarterly, Vol. 40, 459 - 467

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, Vol. 3, part 2, 1-158

Graham, C. & Pettinato, S. (2001),
Frustrated Achievers: Winners, Losers, and Subjective Wellbeing in New Market Economies,
Brookings Institution, Centre on Social and Economic Dynamics Working Paper Series no.21,
Washington D.C., USA

Hagerty, M. & Veenhoven, R. (2003),
Wealth and Happiness Revisited; Growing Wealth of Nations Does Go with Greater Happiness,
Social Indicators Research, 2003, Vol. 64, 1-27

Inglehart, R. (1990),
Culture Shift in Advanced Industrial Society,
Princeton University Press, Princeton, USA

Inglehart, R. & Klingeman, H.D. (2000),
Genes, Culture, Democracy and Happiness,
Diener, E. & Suh, E.M. (Eds.) "Culture and Subjective wellbeing", MIT Press, Cambridge,
USA, 165 - 184

Keeping, E.S. (1962),
Introduction to Statistical Inference,
VanNostrand Reinhold, New York, USA.

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.

Michalos, A.C. (1991),
Global Report on Student Wellbeing, Volume I: Life Satisfaction and Happiness,
Springer, New York, USA

Ng, Y-K (1996),
Happiness Surveys: Some Comparability Issues and an Exploratory Survey Based on Just Perceivable Increments,
Social Indicators Research, Vol. 38, 1 – 27

Ng, Y-K (1997),
A Case for Happiness, Cardinalism, and Interpersonal Comparability,
The Economic Journal, Vol. 107, 1848 – 1858

Orwell, G (1945[1]),
Animal Farm,
Secker and Warburg Ltd, London, UK.

Ott, J. (2005)
Level and Inequality of Happiness in Nations. Does Greater Happiness of a Greater Number Imply Greater Inequality in Happiness?
Journal of Happiness Studies, Vol. 6, pp. 357-396

Saris, W., Bunting, B., Scherpenzeel, A.C. & Veenhoven, R. (Eds.) (1996),
A Comparative Study of Satisfaction with Life in Europe,
Eötvös University Press, Budapest, Hungary.

Sen, A. (1997²),
On Economic Inequality,
Clarendon Press, Oxford, UK

Stevens, S.S. (1946),
On the Theory of Scales of Measurement,
Science, vol. 103, 670 - 680.

Theil, H, (1967),
Economics and Information Theory
North-Holland Publishing Cy, Amsterdam, The Netherlands.

VanPraag, B.M.S. (1991),
Ordinal and Cardinal Utility: an Integration of the Two Dimensions in the Welfare Concept
Journal of Econometrics, vol. 50, pp. 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, The Netherlands

Veenhoven, R. (1990),
Inequality in Happiness, Inequality in Countries Compared between Countries,
Paper 12th Work Congress of Sociology, Madrid, Spain

Veenhoven, R. (1993),
Happiness in Nations,
Studies in Socio-Cultural Transformation, nr. 2, RISBO, Erasmus University Rotterdam, Netherlands

Veenhoven, R. & Ehrhardt, J. (1995),
The Cross-national Pattern of Happiness; Test of Predictions Implied in Three Theories of Happiness,
Social Indicators Research, 1995, Vol. 34, 33 - 68

Veenhoven, R. (2000),
Happiness in Society,
In: Allmendinger, J. (Ed.): *"Gute Gesellschaft? Verhandlung des 30. Kongresses der Deutschen Gesellschaft für Soziologie",* Leske & Budrich, Opladen, Germany, 1265 - 1314

Veenhoven, R. (2000b),
Well-being in the Welfare state: Level not Higher, Distribution not More Equitable,
Journal of Comparative Policy Analysis: Research and Practice, 2000, Vol. 2, 91 - 125

Veenhoven, R. (2002),
Die Rückkehr der Ungleichheit in die moderne Gesellschaft? Die Verteilung der Le-
benszufriedenheit in den EU-Ländern von 1973 bis 1996 (Return of inequality in mod-
ern 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, 2002, Opladen, Germany, ISBN 3-8100-3368-5, pp. 273-294

Veenhoven, R. & Kalmijn, W.M. (2005),
Inequality-adjusted Happiness in Nations,
Journal of Happiness Studies, Vol. 6, pp 421-455

Table 1
Maximum values of the sample variance and standard deviation,
For different mean values on an [0; 10] scale

	Mean	Maximal variance	Maximal 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

Table 2a
Hypothetical distributions, Series 1

Distribut		EQUAL										UNEQUAL		
		A1	A2	B1	B2	C1	C2	D	E	F	G	H	K	
Happy	10	0	100	0	0	3	0	0	10	0	10	0	50	
	9	0	0	0	0	6	0	0	10	10	10	0	0	
	8	0	0	0	0	14	3	0	10	10	20	50	0	
	7	0	0	0	0	27	6	0	10	20	10	0	0	
	6	100	0	50	0	27	14	0	10	10	0	0	0	
	5	0	0	50	0	14	27	20	10	10	0	0	0	
	4	0	0	0	0	6	27	20	10	20	10	0	0	
	3	0	0	0	0	3	14	20	10	10	20	50	0	
	2	0	0	0	50	0	6	20	10	10	10	0	0	
Unhappy	1	0	0	0	50	0	3	20	10	0	10	0	50	

Statistics

		A1	A2	B1	B2	C1	C2	D	E	F	G	H	K	
Mean		6	10	5.5	1.5	6.5	4.5	3	5.5	5.5	.5	5.5	5.5	
														Tau-B
Standard deviation		0.00	0.00	0.50	0.50	1.50	1.50	1.41	2.87	2.16	3.07	2.50	4.50	0.85
Mean absolute deviation		0.00	0.00	0.50	0.50	1.20	1.20	1.20	2.50	1.90	2.67	2.50	4.50	0.90
Coefficient of variation (%)		0	0	9	33	23	33	47	52	39	56	45	82	0.82
Interquartile range		0.50	0.25	1.00	0.75	1.85	1.85	2.50	5.00	3.50	5.50	5.00	8.50	0.88
Gini coefficient		0.00	0.00	0.05	0.17	0.13	0.18	0.27	0.30	0.22	0.31	0.23	0.41	0.83
Theil's inequality measure		0.00	0.00	0.004	0.06	0.03	0.06	0.12	0.15	0.08	0.17	0.11	0.39	0.82
Mean pair distance		0.00	0.00	0.51	0.51	1.67	1.67	1.62	3.33	2.48	3.49	2.53	4.55	0.84
% Outside mode		0	0	50	50	73	73	80	90	80	80	50	50	0.44

Table 2b
Hypothetical distributions, Series 2

Distributions		EQUAL											UNEQUAL	
		A1	A2	A3	B1	B2	L1	L2	M1	M2	N1	N2	K	
Happy	10	0	100	0	0	0	0	0	0	0	0	50	50	
	9	0	0	0	0	0	0	0	0	0	0	0	0	
	8	0	0	0	0	0	0	0	0	50	50	0	0	
	7	0	0	0	0	0	50	0	0	0	0	0	0	
	6	100	0	0	50	0	0	0	50	0	0	0	0	
	5	0	0	0	50	0	50	0	0	0	0	0	0	
	4	0	0	0	0	0	0	0	0	0	0	50	0	
	3	0	0	100	0	0	0	50	0	50	0	0	0	
	2	0	0	0	0	50	0	0	0	0	50	0	0	
Unhappy	1	0	0	0	0	50	0	50	50	0	0	0	50	
Statistics														
Mean		6,0	10,0	3,0	5,5	1,5	6,0	2,0	3,5	5,5	5,0	7,0	5,5	
														Tau-B
Standard deviation		0.00	0.00	0.00	0.50	0.50	1.00	1.00	2,50	2.50	3.00	3.00	4.50	1.00
Mean absolute deviation		0.00	0.00	0.00	0.50	0.50	1.00	1.00	2.50	2.50	3.00	3.00	4.50	1.00
Coefficient of variation (%)		0	0	0	9	33	17	50	71	46	60	43	82	0.75
Interquartile range		0.50	0.25	0.50	1.00	0.75	2.00	2.00	4.75	5.00	6.00	5.75	8.50	0.93
Gini coefficient		0.00	0.00	0.00	0.05	0.17	0.08	0.25	0.36	0.23	0.30	0.21	0.41	0.75
Theil's inequality measure		0.00	0.00	0.00	0.004	0.06	0.01	0.13	0.28	0.11	0.19	0.09	0.39	0.75
Mean pair distance		0.00	0.00	0.00	0.51	0.51	1.01	1.01	2.53	2.53	3.03	3.03	4.55	1.00
% Outside mode		0	0	0	50	50	50	50	50	50	50	50	50	----

Table 2c
Hypothetical distributions, Series 3

Distributions		EQUAL											UN- EQUAL	
		A4	P	Q	R	S1	S2	T	U	V	W	X	K	
Happy	**10**	100	95	90	86	80	20	75	70	65	60	55	50	
	9	0	0	0	0	0	0	0	0	0	0	0	0	
	8	0	0	0	0	0	0	0	0	0	0	0	0	
	7	0	0	0	0	0	0	0	0	0	0	0	0	
	6	0	0	0	0	0	0	0	0	0	0	0	0	
	5	0	0	0	0	0	0	0	0	0	0	0	0	
	4	0	0	0	0	0	0	0	0	0	0	0	0	
	3	0	0	0	0	0	0	0	0	0	0	0	0	
	2	0	0	0	0	0	0	0	0	0	0	0	0	
Unhappy	**1**	0	5	10	15	20	80	25	30	35	40	45	50	
Statistics														
Mean		10.00	9.55	9.10	8.65	8.20	2.80	7.75	7.20	6.85	6.40	5.95	5.50	
														Tau-B
Standard deviation		0.00	1.96	2.70	3.21	3.60	3.60	3.90	4.12	4.29	4.41	4.48	4.50	1.00
Mean absolute deviation		0.00	0.86	1.64	2.32	2.91	2.91	3.41	3.82	4.14	4.36	4.50	4.55	1.00
Coefficient of variation (%)			20	30	37	44	129	50	56	63	69	75	82	0.82
Interquartile range		0.25	0.26	0.28	0.29	0.31	0.31	8.33	8.41	8.45	8.48	8.50	8.50	1.00
Gini coefficient		0.00	0.04	0.09	0.13	0.18	0.51	0.22	0.26	0.30	0.34	0.37	0.41	0.82
Theil's inequality measure		0.00	0.03	0.07	0.11	0.14	0.62	0.18	0.22	0.26	0.30	0.34	0.39	0.82
Mean pair distance		0.00	0.86	1.64	2.32	2.91	2.91	3.41	3.82	4.14	4.36	4.50	4.55	1.00
% Outside mode		0	5	10	15	20	20	25	30	35	40	45	50	1.00

Table 2d
Hypothetical distributions, Series 4, 5 and 6

Distributions		**SERIES 4**		**SERIES 5**			**SERIES 6**		
		Y1	**Y2**	**E**	**Z**		**E**	**Zcorr**	
Happy	**10**	19	1	10			10	20	**10**
	9	17	3	10			10		
	8	15	5	10			10	20	**7.75**
	7	13	7	10			10		
	6	11	9	10			10	20	**5.5**
	5	9	11	10	20	**5**	10		
	4	7	13	10	20	**4**	10		
	3	5	15	10	20	**3**	10	20	**3.25**
	2	3	17	10	20	**2**	10		
Unhappy	**1**	1	19	10	20	**1**	10	20	**1**

Statistics

	Y1	Y2	E	Z	E	Zcorr
Mean	7.15	3.85	5.50	3.00	5.50	5.50
Standard deviation	2.35	2.35	2.87	1.41	2.87	3.18
Mean absolute deviation	1.97	1.97	2.50	1.20	2.50	2.57
Coefficient of variation (%)	33	61	52	47	52	58
Interquartile range	3.65	3.65	5.00	2.50	5.00	5.63
Gini coefficient	0.18	0.34	0.30	0.27	0.30	0.33
Theil's inequality measure	0.06	0.19	0.15	0.12	0.15	0.19
Mean pair distance	2.66	2.66	3.33	1.62	3.33	3.64
% Outside mode	81	81	90	80	90	80

Table 3
Summary scheme: performance of statistics of dispersion on requirements for measuring inequality of happiness in nations

	Criteria							
	1 Numerical result	**2** Interval level	**3** Independence of scale range	**4** Independence of N	**5** Independence of mean	**6** 'Equal for equal'	**7** 'Unequal for unequal'	**8** Sensitive to degree of inequality
Statistics of dispersion								
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	**FAILS**	Weakly	OK	**FAILS**	**FAILS**	OK	**FAILS**
Theil's measure	OK	**FAILS**	**FAILS**	??	**FAILS**	**FAILS**	**FAILS**	**FAILS**
% Outside mode	OK	OK	**FAILS**	OK	OK	OK	**FAILS**	**FAILS**
Range	OK	OK	OK	OK	OK	OK	**FAILS**	**FAILS**

Table 4
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	*4.02*
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	*3.13*
Belarus	2000	2.21	1.78	2.50	*3.28*

APPENDIX A

Maximum and minimum values of the standard deviation at different mean values

Consider a fictitious population consisting of people who are either extremely happy, happiness rating $= h$ on some rating scale, or extremely unhappy, rating $= u,$ in proportions of α ($0 \leq \alpha \leq 1$) and $(1-\alpha)$ respectively. No in between happiness ratings have been selected by anyone.

In this case the mean happiness value

$$m = \alpha \cdot h + (1-\alpha) \cdot u = u + \alpha \cdot (h-u), \qquad (1)$$

and the variance

$$var = \alpha \cdot (h-m)^2 + (1-\alpha) \cdot (u-m)^2. \qquad (2)$$

From (1) it follows that

$$\alpha = (m-u)/(h-u),$$

and its substitution into (2) results in

$$var = (h-m) \cdot (m-u) \qquad (3)$$

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 happy people select a happiness rating between h and m, which requires that at the same time a, not necessarily equal, number of 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,

$$(h-m) \cdot (m-u) = max(s^2) = [max(s)]^2 \qquad (4)$$

As the reader can easily verify, this relationship can also be written as

$$[m - \tfrac{1}{2}(h+u)]^2 + [max(s)]^2 = [\tfrac{1}{2}(h-u)]^2 \qquad (5)$$

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 $\tfrac{1}{2} |h-u|,$ so the maximum value of the theoretical maximum standard deviation is just equal to $\tfrac{1}{2} |h-u|$. 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 with one of the ratings on the original rating scale.

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.

Summarizing:

$$0 \leq s \leq \sqrt{(h-m) \cdot (m-u)} \leq \tfrac{1}{2} |h-u|. \qquad (6)$$

Although a point inside the semicircle or at its circumference can represent any society, the reverse is not true. There are two reasons for this.

The first one is 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 m is not a continuous variable, but it is a discrete variable. If we assume $u \leq h$, then $m = u\,(1/N)\,h$ and so $\alpha = (m–u)/(h–u)$ and s are also discrete numbers too.

In order to illustrate this, we consider a sample of 4 subjects, rating their happiness on a $[1;3]$ scale. The 15 different possible outcomes are listed in Table 5

Table 5.

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}) amounts up to $N(k − 1) +1$ and this is finite, whereas the number of points of a true semicircle is not. Moreover, not for each of those m-values, the maximum value of s reaches its 'semicircle value' $\sqrt{(h–m)\cdot(m–u)}$, as is computed in the right hand column and which is a boundary value only.

Table 3 also reveals that there is a second reason, which arises from the fact that k is a relatively small integer. Hence, situations in which $s = 0$ can occur for those k values of m only, that correspond to one of the ratings of the original 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, such a zone is bounded by a semicircle with, in the case of an original rating scale has, a diameter equal to unity and a maximum height of 0.5. In case of transformed scales, these diameters need to be adjusted accordingly, although the number of 'empty zones' will keep the value k−1.

Although the above considerations are not of great practical interest, they make clear that the statement that "any point inside the semicircle or at its boundary represents a possible society" is not correct. For the line $s = 0$, it is not even approximately true for large samples.

NOTES

[1] In the sociology of inequality the focus is often on differences between social categories, such as white- and blue-collar workers. In this approach, the focus is on differences among individual citizens in a country.

[2] World Database of Happiness, Distributional Findings in Nations, Finding Reports.
www2.eur.nl/fsw/research/happiness/hap_nat/findingreports/list_of_reports.htm)

[3] $i = j(1)k$ means that the value of the rating i varies between j and k with steps equal to unity, so i can adopt one of the values $\{j; j+1; j+2; ... ; k-2; k-1; k\}$

[4] In this paper we use the terms 'average' and 'mean' synonymously.

[5] Happiness is commonly rated on a numerical scale ranging from 0 (least happy) to 10 (most happy). This scale is denoted as [0 ; 10].

[6] We speak about the measurement of 'inequality' and not the measurement of 'equality'. The reason is that the former concept exists in gradations and can be quantified, e.g. with the standard deviation. This does not apply to equality, which 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".

[7] 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 kurtosis can be expressed as:
$$N\left[\Sigma\, n_i(i-m)^4\right]/\left[\Sigma\, n_i(i-m)^2\right]^2 ,$$

where m = the mean happiness value, n_i = the absolute frequency of the rating i and $N = \Sigma n_i$. Summation is over all different values of i, so over all k different ratings. For normal distributions, the above kurtosis has the value 3 (Kendall and Stuart, 1977: 88), for the distribution C1 and C2 the value 2.99 is found, so the normality is approximated very well.

[8] 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. As has been pointed out in Section 3.7, such valid models have not been proposed until now for happiness distributions.

[9] If in a distribution of some variable, e.g. happiness, all individual ratings are augmented by the same amount *(d)*, the complete histogram 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 histogram will change the algebraic sign of the skewness, but will not affect the standard deviation.

[10] The values of these statistics form in irregular series, which depends on the order in which two equally unequal distributions are introduced in the table. If e.g. the order of L1 and L2 is reversed, a different pattern trend will occur. The same holds for the pairs M1/M2 and N1/N2.