DOWNLOAD PDF WHAT IS PRESENT: THE PROBLEM OF MEASUREMENT

Chapter 1 : Why GDP fails as a measure of well-being - CBS News

What five points did Howitt () make regarding the measurement of crime? 1) None of the measures is useless, as each present a different outlook on the extent of crime. 2) Different measures may be useful for different purposes.

Jones In Brief First New Concepts Statement in 15 Years Statements of financial accounting concepts are issued from time to time as part of a conceptual framework for use by FASB in developing standards for financial accounting and reporting. The framework helps provide consistency and a sense of purpose and direction for new standards. In June , FASB issued a proposed concepts statement on using cash flow information in accounting measurements. After considering the comments received, FASB changed two of the fundamental conclusions contained in the ED and decided to issue a new draft. A March 31, , revision, with a comment deadline of August 1, , took the position that the sole purpose of using present value measurements at initial recognition and for fresh starts is for determination of fair value. The earlier proposal had given other purposes. The proposal, in addition to limiting the purpose of present value measurements at initial recognition, presents the concepts that would apply when calculating present values where there is a range of estimated cash flows. The ED also speaks of the need to adjust cash flows to reflect the uncertainty of the estimates and, in the case of liabilities, the credit standing of the issuer. FASB would like to issue a final statement in early Its ultimate impact rests with the subsequent standards that are developed using the concepts. As of this writing, FASB is expected to issue either a final concept statement or a revised exposure draft in the first quarter of If a final document is issued, it would be the first new SFAC since , and FASB would be the first standards setter to incorporate present value techniques into its conceptual framework. The argument behind the use of present value techniques is that more relevant information is produced due to factoring in the uncertainties and risks associated with the amount and timing of cash flows. This form of accounting measurement is designed to capture the economic substance of a set of cash flows in a manner similar to that of how the market behaves. In other words, present value techniques attempt to measure assets or liabilities at their fair value. Opponents of the use of present value in accounting measurements argue that its application results in a decrease in the reliability of accounting information. A present value calculation requires numerous estimates regarding the timing and amount of future cash flows, interest rates, and economic conditions. The use of these estimates is seen as a threat to the reliability of accounting information either through differing opinions as to future conditions decreased verifiability of estimates , or through the introduction of bias in estimates decreased neutrality. In the exposure draft, FASB has taken the position that the use of present value techniques should result in more relevant information being provided to financial statement users without sufficiently damaging the reliability of the information. FASB has identified fair value as the sole objective when present value techniques are applied in measuring assets or liabilities at the time of initial recognition or for fresh start measurements. This is the fundamental change that FASB made to its proposal. That earlier draft conceptualized that present value as a measurement technique could have other objectives for initial recognition purposes. In adopting fair value as the measurement attribute, FASB is attempting to "capture the elements that taken together would comprise a market price if one existed. As set forth in an alternative view section of the exposure draft, it can be argued that market-based assumptions are not relevant if an entity is engaged in the acquisition or settlement of nonfinancial assets or liabilities in a noncurrent transaction. In such a situation, an alternative approach, such as entity-specific value the fair value unique to the entity based on entity-specific assumptions of cash flows or risk premium or cost accumulation may be more relevant and appropriate. Given the above measurement objective, accounting measurements employing present value techniques should contain an adjustment that reflects the market risks and uncertainties related to the asset or liability being measured. Because the market demands a premium in dealing with risks and uncertainties, any measurement attempting to simulate a market price must do so as well. Additionally, FASB advocates that the credit standing of the entity obligated to pay should always be considered when measuring a liability. Expected Cash Flow Approach. Traditionally, the risk adjustment is accomplished by adjusting the discount factor interest rate. The greater the risk, the higher the interest rate used to discount the cash flows. However, FASB appears to favor incorporating risk and uncertainty by developing a probability-weighted cash flow estimate that utilizes a wider range of information and then discounting these expected cash flows by the risk-free interest rate. The use of probabilities in this expected cash flow approach, while injecting a measure of imprecision, is expected to have several advantages over using a single estimate of cash flows. First, a broader set of information is utilized in the measurement process because a wide range of possible outcomes is considered rather than a single, best estimate. Second, the use of probabilities should make any assumptions about the uncertainties of the cash flow stream explicit, whereas the traditional approach buries these assumptions within the interest rate. Third, a reliable estimate of the market risk premium may often not be available, making the risk-free interest rate the most reliably determined value. Finally, the use of probability-weighted cash flows allows for an explicit adjustment for uncertainties with regard to the timing of cash flows, in addition to uncertainties with regard to the amount of cash flows. While both approaches are conceptually consistent, FASB has left the choice as to the method of risk adjustment up to either the entity or to the context of a specific standard. Unique Aspects of Liabilities. FASB concluded that the concepts developed for the use of present value measurement in initial recognition of assets also apply to liabilities. Liabilities, however, create the following unique problems: For example, the proceeds of a loan are the price that the lender is willing to pay and represent fair value for initial measurement. This is normally reflected in the market rate of interest that the borrower has to pay. This would be true whether the liability was for bonds, notes payable, or warranties. Illustrations Two examples are presented below to illustrate how the proposed concept statement may affect accounting measurement. It should be noted at this point that because the concepts do not establish GAAP, the issuance of a final statement of FASB will not mandate the accounting treatment shown below. Instead, these illustrations are merely applications of the ideas found in the exposure draft. The fair value of the services is not readily determinable, and the note is not readily marketable. It is made up of two elements: The present value of the note and the imputed fair value of the landscaping services is computed as follows:

Chapter 2 : Observer effect (physics) - Wikipedia

Read "Present Status of the Problem of Measurement of X-Ray Intensity and Quality, Radiology" on DeepDyve, the largest online rental service for scholarly research with thousands of academic publications available at your fingertips.

Overview Modern philosophical discussions about measurementâ€”spanning from the late nineteenth century to the present dayâ€”may be divided into several strands of scholarship. These strands reflect different perspectives on the nature of measurement and the conditions that make measurement possible and reliable. The main strands are mathematical theories of measurement, operationalism, conventionalism, realism, information-theoretic accounts and model-based accounts. These strands of scholarship do not, for the most part, constitute directly competing views. Instead, they are best understood as highlighting different and complementary aspects of measurement. The following is a very rough overview of these perspectives: Mathematical theories of measurement view measurement as the mapping of qualitative empirical relations to relations among numbers or other mathematical entities. Information-theoretic accounts view measurement as the gathering and interpretation of information about a system. These perspectives are in principle consistent with each other. While mathematical theories of measurement deal with the mathematical foundations of measurement scales, operationalism and conventionalism are primarily concerned with the semantics of quantity terms, realism is concerned with the metaphysical status of measurable quantities, and information-theoretic and model-based accounts are concerned with the epistemological aspects of measuring. Nonetheless, the subject domain is not as neatly divided as the list above suggests. Issues concerning the metaphysics, epistemology, semantics and mathematical foundations of measurement are interconnected and often bear on one another. Hence, for example, operationalists and conventionalists have often adopted anti-realist views, and proponents of model-based accounts have argued against the prevailing empiricist interpretation of mathematical theories of measurement. These subtleties will become clear in the following discussion. The list of strands of scholarship is neither exclusive nor exhaustive. It reflects the historical trajectory of the philosophical discussion thus far, rather than any principled distinction among different levels of analysis of measurement. Some philosophical works on measurement belong to more than one strand, while many other works do not squarely fit either. This is especially the case since the early s, when measurement returned to the forefront of philosophical discussion after several decades of relative neglect. The last section of this entry will be dedicated to surveying some of these developments. A Brief History Although the philosophy of measurement formed as a distinct area of inquiry only during the second half of the nineteenth century, fundamental concepts of measurement such as magnitude and quantity have been discussed since antiquity. Two magnitudes have a common measure when they are both whole multiples of some magnitude, and are incommensurable otherwise Book X, def. The discovery of incommensurable magnitudes allowed Euclid and his contemporaries to develop the notion of a ratio of magnitudes. Ratios can be either rational or irrational, and therefore the concept of ratio is more general than that of measure Michell , ; Grattan-Guinness Aristotle distinguished between quantities and qualities. Aristotle did not clearly specify whether degrees of qualities such as paleness correspond to distinct qualities, or whether the same quality, paleness, was capable of different intensities. This topic was at the center of an ongoing debate in the thirteenth and fourteenth centuries Jung This theory was later refined by Nicole Oresme, who used geometrical figures to represent changes in the intensity of qualities such as velocity Clagett ; Sylla These developments made possible the formulation of quantitative laws of motion during the sixteenth and seventeenth centuries Grant The concept of qualitative intensity was further developed by Leibniz and Kant. Leibniz argued that this principle applies not only to changes in extended magnitudes such as length and duration, but also to intensities of representational states of consciousness, such as sounds Jorgensen ; Diehl An example is length: For Kant, the possibility of such synthesis was grounded in the forms of intuition, namely space and time. Intensive magnitudes, like warmth or colors, also come in continuous degrees, but their apprehension takes place in an instant rather than through a successive synthesis of parts. Scientific developments during the nineteenth century challenged the distinction between extensive and intensive magnitudes. Thermodynamics and wave optics showed that differences in temperature and hue corresponded to differences in spatio-temporal magnitudes such as velocity and wavelength. Electrical magnitudes such as resistance and conductance were shown to be capable of addition and division despite not being extensive in the Kantian sense, i. For example, 60 is twice 30, but one would be mistaken in thinking that an object measured at 60 degrees Celsius is twice as hot as an object at 30 degrees Celsius. This is because the zero point of the Celsius scale is arbitrary and does not correspond to an absence of temperature. When subjects are asked to rank on a scale from 1 to 7 how strongly they agree with a given statement, there is no prima facie reason to think that the intervals between 5 and 6 and between 6 and 7 correspond to equal increments of strength of opinion. These examples suggest that not all of the mathematical relations among numbers used in measurement are empirically significant, and that different kinds of measurement scale convey different kinds of empirically significant information. The study of measurement scales and the empirical information they convey is the main concern of mathematical theories of measurement. A key insight of measurement theory is that the empirically significant aspects of a given mathematical structure are those that mirror relevant relations among the objects being measured. This mirroring, or mapping, of relations between objects and mathematical entities constitutes a measurement scale. As will be clarified below, measurement scales are usually thought of as isomorphisms or homomorphisms between objects and mathematical entities. Other than these broad goals and claims, measurement theory is a highly heterogeneous body of scholarship. It includes works that span from the late nineteenth century to the present day and endorse a wide array of views on the ontology, epistemology and semantics of measurement. Two main differences among mathematical theories of measurement are especially worth mentioning. These relata may be understood in at least four different ways: This issue will be especially relevant to the discussion of realist accounts of measurement Section 5. Second, different measurement theorists have taken different stands on the kind of empirical evidence that is required to establish mappings between objects and numbers. As a result, measurement theorists have come to disagree about the necessary conditions for establishing the measurability of attributes, and specifically about whether psychological attributes are measurable. Debates about measurability have been highly fruitful for the development of measurement theory, and the following subsections will introduce some of these debates and the central concepts developed therein. Although accounts of measurement varied, the consensus was that measurement is a method of assigning numbers to magnitudes. For example, Helmholtz Bertrand Russell similarly stated that measurement is any method by which a unique and reciprocal correspondence is established between all or some of the magnitudes of a kind and all or some of the numbers, integral, rational or real. Defining measurement as numerical assignment raises the question: Moreover, the end-to-end concatenation of rigid rods shares structural featuresâ€"such as associativity and commutativityâ€"with the mathematical operation of addition. A similar situation holds for the measurement of weight with an equal-arms balance. Here deflection of the arms provides ordering among weights and the heaping of weights on one pan constitutes concatenation. Early measurement theorists formulated axioms that describe these qualitative empirical structures, and used these axioms to prove theorems about the adequacy of assigning numbers to magnitudes that exhibit such structures. Specifically, they proved that ordering and concatenation are together sufficient for the construction of an additive numerical representation of the relevant magnitudes. An additive representation is one in which addition is empirically meaningful, and hence also multiplication, division etc. A hallmark of such magnitudes is that it is possible to generate them by concatenating a standard sequence of equal units, as in the example of a series of equally spaced marks on a ruler. Although they viewed additivity as the hallmark of measurement, most early measurement theorists acknowledged that additivity is not necessary for measuring. Examples are temperature, which may be measured by determining the volume of a mercury column, and density, which may be measured as the ratio of mass and volume. Nonetheless, it is important to note that the two distinctions are based on significantly different criteria of measurability. As discussed in Section 2 , the extensive-intensive distinction focused on the intrinsic structure of the quantity in question, i. The fundamental-derived distinction, by contrast, focuses on the properties of measurement operations. A fundamentally measurable magnitude is one for which a fundamental measurement operation has been found. Consequently, fundamentality is not an intrinsic property of a magnitude: Moreover, in fundamental measurement the numerical assignment need not mirror the structure of spatio-temporal parts. Electrical resistance, for example, can be fundamentally measured by connecting resistors in a series Campbell This is considered a fundamental measurement operation because it has a shared structure with numerical addition, even though objects with equal resistance are not generally equal in size. The distinction between fundamental and derived measurement was revised by subsequent authors. Fundamental measurement requires ordering and concatenation operations satisfying the same conditions specified by Campbell. Associative measurement procedures are based on a correlation of two ordering relationships, e. Derived measurement procedures consist in the determination of the value of a constant in a physical law. The constant may be local, as in the determination of the specific density of water from mass and volume, or universal, as in the determination of the Newtonian gravitational constant from force, mass and distance. Duncan Luce and John Tukey in their work on conjoint measurement, which will be discussed in Section 3. A complementary line of inquiry within measurement theory concerns the classification of measurement scales. Stevens , distinguished among four types of scales: Nominal scales represent objects as belonging to classes that have no particular order, e. Ordinal scales represent order but no further algebraic structure. For example, the Mohs scale of mineral hardness represents minerals with numbers ranging from 1 softest to 10 hardest , but there is no empirical significance to equality among intervals or ratios of those numbers. The Kelvin scale, by contrast, is a ratio scale, as are the familiar scales representing mass in kilograms, length in meters and duration in seconds. Stevens later refined this classification and distinguished between linear and logarithmic interval scales As Stevens notes, scale types are individuated by the families of transformations they can undergo without loss of empirical information. Empirical relations represented on ratio scales, for example, are invariant under multiplication by a positive number, e. Linear interval scales allow both multiplication by a positive number and a constant shift, e. Ordinal scales admit of any transformation function as long as it is monotonic and increasing, and nominal scales admit of any one-to-one substitution. Absolute scales admit of no transformation other than identity. Two issues were especially contested. Several physicists, including Campbell, argued that classification and ordering operations did not provide a sufficiently rich structure to warrant the use of numbers, and hence should not count as measurement operations. The second contested issue was whether a concatenation operation had to be found for a magnitude before it could be fundamentally measured on a ratio scale. The debate became especially heated when it re-ignited a longer controversy surrounding the measurability of intensities of sensation. It is to this debate we now turn. These differences were assumed to be equal increments of intensity of sensation. This law in turn provides a method for indirectly measuring the intensity of sensation by measuring the intensity of the stimulus, and hence, Fechner argued, provides justification for measuring intensities of sensation on the real numbers. Those objecting to the measurability of sensation, such as Campbell, stressed the necessity of an empirical concatenation operation for fundamental measurement.

DOWNLOAD PDF WHAT IS PRESENT: THE PROBLEM OF MEASUREMENT

Chapter 3 : Measurement problem - Wikipedia

The measurement problem in quantum mechanics is the problem of how (or whether) wave function collapse occurs. The inability to observe this process directly has given rise to different interpretations of quantum mechanics and poses a key set of questions that each interpretation must answer.

The discount rate element of the NPV formula is a way to account for this. A rational investor would not be willing to postpone payment. A company may determine the discount rate using the expected return of other projects with a similar level of risk, or the cost of borrowing money needed to finance the project. The managers feel that buying the equipment or investing in the stock market are similar risks. NPV of the initial investment Because the equipment is paid for up front, this is the first cash flow included in the calculation. Identify the number of periods t The equipment is expected to generate monthly cash flow and last for 5-years which means there will be 60 cash flows and 60 periods included in the calculation. However, because the equipment generates a monthly stream of cash flows, the annual discount rate needs to be turned into a periodic or monthly rate. Using the following formula, we find that the periodic rate is 0. NPV of future cash flows Assume the monthly cash flows are earned at the end of the month, with the first payment arriving exactly one month after the equipment has been purchased. This is a future payment, so it needs to be adjusted for the time value of money. An investor can perform this calculation easily with a spreadsheet or calculator. To illustrate the concept, the first 5 payments are discounted in the table below. The calculation could be more complicated if the equipment was expected to have any value left at the end of its life, but in this example, it is assumed to be worthless. That formula can be simplified to the following calculation: In this case, the NPV is positive; the equipment should be purchased. If the present value of these cash flows had been negative because the discount rate was larger, or the net cash flows were smaller, the investment would have been avoided. Estimated factors include investment costs, discount rate, and projected returns. The payback method calculates how long it will take for the original investment to be repaid. For this reason, payback periods calculated for longer investments have a greater potential for inaccuracy. Moreover, the payback period is strictly limited to the amount of time required to earn back initial investment costs. Comparisons using payback periods do not account for the long-term profitability of alternative investments. Net Present Value vs. This method is used to compare projects with different lifespans or amount of required capital. Although the IRR is useful, it is usually considered inferior to NPV because it makes too many assumptions about reinvestment risk and capital allocation. It accounts for the time value of money and can be used to compare investment alternatives that are similar. The NPV relies on a discount rate of return that may be derived from the cost of the capital required to make the investment, and any project or investment with a negative NPV should be avoided. An important drawback of using an NPV analysis is that it makes assumptions about future events that may not be reliable.

Chapter 4 : Definitions, Uses, Data Types, and Levels of Measurement

The ratio scale of measurement is the most informative scale. It is an interval scale with the additional property that its zero position indicates the absence of the quantity being measured.

We need now to accept that physics is essentially indeterministic; that particles may be somewhere without being at any particular place; that they may have energy and momentum without having any particular value for them; and a host more non-classical oddities. Most of these ideas are simply unfamiliar conceptions and, in the end, the best thing is just to get used to the idea that world depicted by quantum theory is very different from the world delivered by our raw senses. There are other problems in quantum theory that should not be accommodated with this forgiving attitude. This chapter will develop the one that it most prominent and has proven most intractable: It depends on the fact that a quantum system can evolve in time in two ways. One way, you will recall from the last chapter, is Schroedinger evolution, in which the wave of the system propagates in the familiar manner of waves. The other way a quantum system can evolve in time is through the "collapse of the wave packet" that arises when we perform a measurement: When will a wave packet undergo Schroedinger evolution or collapse? Earlier, we saw that there is only a rule of thumb to guide us. Schroedinger evolution arises when matter waves are left to themselves or when they interact with just a few others. Measurement arises when a matter wave interacts with a macroscopic measuring device. That means that a matter wave interacting with a photographic plate collapses. Sometimes it is said that the last collapse does not happen until an intelligent human agent actually looks at the plate. That last claim is extremely strange. Are we supposed to believe that human intelligence enters into the time evolution of fundamental particles in the same way as perturbing fields? The lack of a precise principle to decide which evolution will arise has created a constellation of puzzles known at the "measurement problem. The radioactive element Neptunium NP is extremely unstable. It will undergo radioactive decay quite quickly. It has a "half life" of 53 minutes. That means that if we start with a lump of NP and wait 53 minutes we will have only half a lump left, near enough, and lot of radioactive decay products. Now, individual atoms of NP are governed by Schroedinger evolution; the probabilities only enter when we measure to see if the atom has decayed or not. So over 53 minutes the atom evolves into a half: The collapse into one or other of these components only arises when we take a measurement. That may happen when we use a Geiger counter to check for radioactive decay products. If we find them, then the atom collapses into the decayed component. Otherwise it collapsed into the undecayed component. So far everything seems reasonable. What Schroedinger realized was that there was quite some arbitrariness in our division between Schroedinger evolution and wave collapse. It was quite possible for that one collapse to be magnified. The decay products of the one decaying atom might trigger the collapse of others. So instead of having just one atom entering into a superposition over 53 minutes, we might have very many atoms all coupled together entering the superposed state after 53 minutes. The cat paradox arises when we push this process of amplification to an extreme. Instead of coupling the one atom of NP to a collection of other radioactive atoms, we couple it to the atoms of cat. The coupling is simple, although cruel. A Geiger counter is set up to sense the decay of the atom. If it decays, the Geiger counter will trigger the opening of a can of poison. The atom, Geiger counter, poison and cat are all enclosed in a box. We then wait 53 minutes. In that time, the atom evolves into a superposition of undecayed and decayed atom. With it, the poison evolves into a superposition of released and unreleased poison; and the cat into a superposition of live cat and dead cat. So the cat is neither alive nor dead. The evolution, as far as the cat is concerned, is something like this: What finally decides whether the cat is alive or dead is our observation. After 53 minutes we open the box and observe, that is, "measure," the life state of the cat. There is a widespread sense that there is something wrong with a theory that allows observation to play such an important role. Most people have an instinctive sense that the fact of life or death for the cat is not decided merely by our observation. After 53 minutes, the cat is definitely just one of alive or dead; whether we look in the box does not change that circumstance in any way. This instinctive reaction is surely correct. However having it really only sharpens the problem. It does not solve it. In the last two decades especially, there has been a huge amount of work devoted to finding variations to standard quantum theory or just new ways to think about the same theory that avoid this problem. There is no consensus on which approach is the correct one or even if some sort of repair is needed. Einstein, Realism and the Quantum Theory. University of Chicago Press, , p. If the decay occurs, the counter tube fires and, by means of a relay, sets a little hammer into motion that shatters a small bottle of prussic acid. When the entire system has been left alone for an hour, one would say that the cat is still alive provided no atom has decayed in the meantime. The first atomic decay would have poisoned it. This prevents us from continuing naively to give credence to a "fuzzy model" as a picture of reality. In itself this contains nothing unclear or contradictory. There is a difference between a blurred or unsharply taken photograph and a shot of clouds and mist. However as the new quantum theory solidified in the late s and thereafter became standard physics, Einstein increasingly found himself playing a rather different role, a critic of the new. He had no doubt about the great successes of quantum theory in exploring atomic phenomena and accommodating the results of experiments. His concern, however, was that the theory was only a provisional stopping point on the path to a better theory. We shall see in a coming chapter how Einstein elaborated these worries. He concentrated on the idea that the quantum wave was not a complete description of reality, but, in some way, merely described averages. The best known expression of these worries came in a paper Einstein co-authored with Boris Podolsky and Nathan Rosen, known universally as the "EPR" paper. Rosen, "Can quantum-mechanical description of physical reality be considered complete? Received March 25, ; published May 15, In the aftermath of this paper, Einstein and Schroedinger exchanged letters in which they aired their common concerns about quantum theory. In that correspondence, Einstein put to Schroedinger what we now see is an early version of the cat paradox. He outlined a "crude macroscopic example" in a letter to Schroedinger of August 8, In principle this can quite easily be represented quantum-mechanically. But, according to your equation, after the course of a year this is no longer the case at all. In a letter back to Einstein of August 19, , he characterized his newly-conceived cat thought experiment as "very similar to your exploding powder keg. In , Einstein assembled his reactions to them. He took the example of the decay of a radioactive atom: The radioactive process consists in the emission of a comparatively light particle. For the sake of simplicity we neglect the motion of the residual atom after the disintegration process. Rather than considering a system which comprises only a radioactive atom and its process of transformation , one considers a system which includes also the means for ascertaining the radioactive transformation â€" for example, a Geiger-counter with automatic registration-mechanism. Let this latter include a registration-strip, moved by a clockwork, upon which a mark is made by tripping the counter. True, from the point of view of quantum mechanics this total system is very complex and its configuration space is of very high dimension. But there is in principle no objection to treating this entire system from the standpoint of quantum mechanics. Here too the theory determines the probability of each configuration of all its co-ordinates for every time instant. If one considers all configurations of the coordinates, for a time large compared with the average decay time of the radioactive atom, there will be at most one such registration-mark on the paper strip. To each coordinate configuration corresponds a definite position of the mark on the paper strip. But, inasmuch as the theory yields only the relative probability of the thinkable co-ordinate-configurations, it also offers only relative probabilities for the positions of the mark on the paper strip, but no definite location for this mark. Yet the quantum mechanical formalism yields no single mark, but many marks, weighted probabilistically. Einstein continued to explain that he regarded the standard quantum account of the mark as one for which "there is hardly likely to be anyone who would be inclined to consider it seriously. Such an interpretation is certainly by no means absurd from a purely logical standpoint, yet there is hardly likely to be anyone who would be inclined to consider it seriously. For, in the macroscopic sphere it simply is considered certain that one must adhere to the program of a realistic description in space and time; whereas in the sphere of microscopic situations one is more readily inclined to give up, or at least to modify, this program. Generally speaking, most of those responses fall into five groups. Accept the standard account. This response essentially urges that the standard treatment is adequate. It is intelligible only in so far as it repeats the rule of thumb for deciding when quantum systems evolve according to Schroedinger evolution: This rule of thumb is adequate, perhaps, for practical physics. There is no doubt that a massive detector in a particle accelerator will engage in a measurement process. But a rule of thumb is less than the unambiguous rule that univocal theorizing requires. If you have a real interest in the foundations of quantum theory, you should not rest easy with this response. Sometimes the standard treatments try to go further and assure us that all is well with things just the way they are.

Chapter 5 : Present Value-Based Measurements and Fair Value

The lack of a precise principle to decide which evolution will arise has created a constellation of puzzles known at the "measurement problem." The best known example is "Schroedinger's cat," a thought experiment devised by Erwin Schroedinger in

In general we write: Your teacher has a particular fondness for this symbol since the first computer he had much access to had that nickname. There are three important rules for using the summation operator: Since multiplication distributes over addition, the sum of a constant times a set of numbers is the same as the constant times the sum of the set of numbers. Joe got scores of and for his verbal and quantitative SAT scores whereas Jim got scores of and , respectively. In addition to the operations of addition, subtraction, multiplication, and division, several other arithmetic operators often appear. Exponentiation and absolute value are two such. Also, various symbols of inclusion parentheses, brackets, braces, vincula are used. When the square root symbol surd and symbol of inclusion, in recent history a vinculum, but historically parentheses is used, we general although not quite always mean only the positive square root. The absolute value operator indicates the distance always non-negative a number is from the origin zero. The symbol used is a vertical line on either side of the operand. There is a proscribed order for arithmetic operations to be performed. Some calculators are algebraic and handle this appropriately, others do not. Parentheses and other symbols of inclusion are used to modify the normal order of operations. We say these symbols of inclusion have the highest priority or precidence. Exponentiation is done next. There is confusion when exponents are stacked which we will not deal with here except to say computer scientists tend to do it from left to right while mathematicians know that is wrong. Multiplication and Division are done next, in order, from left to right. Addition and Subtraction are done next, in order, from left to right. Precision The distinction between accuracy and precision, reviewed in Numbers lesson 9 , is very important. This ties in with significant figures, and proper rounding of results. I have several major concerns regarding significant digits. There needs to be sufficient not to few. Slide rule accuracy or three significant digits has a long-standing precident in science. We are not doing science here so two may suffice, but rarely one. There should not be too many significant digits. Generally, more than 5 is probably a joke, especially in the "softer" sciences. Care must be taken so that a primary statistics such as variance is not incorrectly derived from a secondary statistic such as standard deviation in such a way that accuracy is lost. We will discuss this more in textbook Chapter 3. A mean and standard deviation or mean and margin of error should be given to compatible precision. There are proper rules, but they are difficult to explain to the general public. Thus every statistics book gives its own heuristic. Uses and Abuses of Statistics Most of the time, samples are used to infer something draw conclusions about the population. If an experiment or study was done cautiously and results were interpreted without bias , then the conclusions would be accurate. However, occasionally the conclusions are inaccurate or inaccurately portrayed for the following reasons: Sample is too small. Even a large sample may not represent the population. Unauthorized personnel are giving wrong information that the public will take as truth. A possibility is a company sponsoring a statistics research to prove that their company is better. Visual aids may be correct, but emphasize different aspects. Often a chart will use a symbol which is both twice as long and twice as high to represent something twice as much. The area, in this case however, is four times as much! Precise statisitics or parameters may incorrectly convey a sense of high accuracy. Misleading or unclear percentages are often used. Statistics are often abused. Many examples could be added, even books have been written but it will be more instructive and fun to find them on your own. Types of Data A dictionary defines data as facts or figures from which conclusions may be drawn. Thus, technically, it is a collective, or plural noun. Some recent dictionaries acknowledge popular usage of the word data with a singular verb. My mother and step-mother were both English teachers, so clearly no offense is intended above. Datum is the singular form of the noun data. Data can be classified as either numeric or nonnumeric. Specific terms are used as follows: Qualitative data are nonnumeric. Qualitative data are often termed catagorical data. Some books use the terms individual and variable to reference the objects and characteristics described by a set of data. They also stress the importance of exact definitions of these variables, including what units they are recorded in. The reason the data were collected is also important. Quantitative data are numeric. Quantitative data are further classified as either discrete or continuous. Discrete data are numeric data that have a finite number of possible values. Another classic is the spin or electric charge of a single electron. Quantum Mechanics, the field of physics which deals with the very small, is much concerned with discrete values. When data represent counts, they are discrete. An example might be how many students were absent on a given day. Counts are usually considered exact and integer. Continuous data have infinite possibilities: The real numbers are continuous with no gaps or interruptions. Physically measureable quantities of length, volume, time, mass, etc. At the physical level microscopically , especially for mass, this may not be true, but for normal life situations is a valid assumption. The structure and nature of data will greatly affect our choice of analysis method. By structure we are referring to the fact that, for example, the data might be pairs of measurements. Consider the legend of Galileo dropping weights from the leaning tower of Pisa. The times for each item would be paired with the mass and surface area of the item. Something which Galileo clearly did was measure the time it took a pendulum to swing with various amplitudes. Galileo Galilei is considered a founder of the experimental method. Levels of Measurement The experimental scientific method depends on physically measuring things. The concept of measurement has been developed in conjunction with the concepts of numbers and units of measurement. Statisticians categorize measurements according to levels. Each level corresponds to how this measurement can be treated mathematically. Nominal data have no order and thus only gives names or labels to various categories. Ordinal data have order, but the interval between measurements is not meaningful. Interval data have meaningful intervals between measurements, but there is no true starting point zero. Ratio data have the highest level of measurement. Ratios between measurements as well as intervals are meaningful because there is a starting point zero. Nominal comes from the Latin root nomen meaning name. Nomenclature, nominative, and nominee are related words. Gender is something you are born with, whereas sex is something you should get a license for. Colors To most people, the colors: To an electronics student familiar with color-coded resistors, this data is in ascending order and thus represents at least ordinal data. To a physicist, the colors: Temperatures What level of measurement a temperature is depends on which temperature scale is used. Only Kelvin and Rankine have true zeroes starting point and ratios can be found. Celsius and Fahrenheit are interval data; certainly order is important and intervals are meaningful. Rankine has the same size degree as Fahrenheit but is rarely used. To interconvert Fahrenheit and Celsius, see Numbers lesson Note that since , the use of the degree symbol on tempertures Kelvin is no longer proper. Although ordinal data should not be used for calculations, it is not uncommon to find averages formed from data collected which represented Strongly Disagree, Also, averages of nominal data zip codes, social security numbers is rather meaningless!

Chapter 6 : Measurement in Science (Stanford Encyclopedia of Philosophy)

Measurement is an integral part of modern science as well as of engineering, commerce, and daily life. Measurement is often considered a hallmark of the scientific enterprise and a privileged source of knowledge relative to qualitative modes of inquiry. [].

Designing a research project takes time, skill and knowledge. With Qualtrics survey software , we make the survey creation process easier, but still you may feel overwhelmed with the scope of your research project. Here are 5 common errors in the research process: Population Specification This type of error occurs when the researcher selects an inappropriate population or universe from which to obtain data. Packaged goods manufacturers often conduct surveys of housewives, because they are easier to contact, and it is assumed they decide what is to be purchased and also do the actual purchasing. In this situation there often is population specification error. The husband may purchase a significant share of the packaged goods, and have significant direct and indirect influence over what is bought. For this reason, excluding husbands from samples may yield results targeted to the wrong audience. Sampling Sampling error occurs when a probability sampling method is used to select a sample, but the resulting sample is not representative of the population concern. Unfortunately, some element of sampling error is unavoidable. This is accounted for in confidence intervals, assuming a probability sampling method is used. Suppose that we collected a random sample of people from the general U. This sample would not be representative of the general adult population and would influence the data. The entertainment preferences of females would hold more weight, preventing accurate extrapolation to the US general adult population. Sampling error is affected by the homogeneity of the population being studied and sampled from and by the size of the sample. Selection Selection error is the sampling error for a sample selected by a nonprobability method. Interviewers conducting a mall intercept study have a natural tendency to select those respondents who are the most accessible and agreeable whenever there is latitude to do so. Such samples often comprise friends and associates who bear some degree of resemblance in characteristics to those of the desired population. Non-responsive Nonresponse error can exist when an obtained sample differs from the original selected sample. In telephone surveys, some respondents are inaccessible because they are not at home for the initial call or call-backs. Others have moved or are away from home for the period of the survey. Not-at-home respondents are typically younger with no small children, and have a much higher proportion of working wives than households with someone at home. People who have moved or are away for the survey period have a higher geographic mobility than the average of the population. Thus, most surveys can anticipate errors from non-contact of respondents. Online surveys seek to avoid this error through e-mail distribution, thus eliminating not-at-home respondents. Measurement Measurement error is generated by the measurement process itself, and represents the difference between the information generated and the information wanted by the researcher. A retail store would like to assess customer feedback from at-the-counter purchases. The survey is developed but fails to target those who purchase in the store. Instead, results are skewed by customers who bought items online. Best Practices for Research Sample.

Chapter 7 : Levels of Measurement

Level of measurement refers to the particular way that a variable is measured within scientific research, and scale of measurement refers to the particular tool that a researcher uses to sort the data in an organized way, depending on the level of measurement that they have selected. Choosing the.

With Safari, you learn the way you learn best. Get unlimited access to videos, live online training, learning paths, books, tutorials, and more. Basic Concepts of Measurement Before you can use statistics to analyze a problem, you must convert information about the problem into data. That is, you must establish or adopt a system of assigning values, most often numbers, to the objects or concepts that are central to the problem in question. This is not an esoteric process but something people do every day. For instance, when you buy something at the store, the price you pay is a measurement: Similarly, when you step on the bathroom scale in the morning, the number you see is a measurement of your body weight. Depending on where you live, this number may be expressed in either pounds or kilograms, but the principle of assigning a number to a physical quantity weight holds true in either case. Data need not be inherently numeric to be useful in an analysis. For instance, the categories male and female are commonly used in both science and everyday life to classify people, and there is nothing inherently numeric about these two categories. Similarly, we often speak of the colors of objects in broad classes such as red and blue, and there is nothing inherently numeric about these categories either. This kind of thinking in categories is a completely ordinary, everyday experience, and we are seldom bothered by the fact that different categories may be applied in different situations. For instance, an artist might differentiate among colors such as carmine, crimson, and garnet, whereas a layperson would be satisfied to refer to all of them as red. The point is that the level of detail used in a system of classification should be appropriate, based on the reasons for making the classification and the uses to which the information will be put. Measurement Measurement is the process of systematically assigning numbers to objects and their properties to facilitate the use of mathematics in studying and describing objects and their relationships. Some types of measurement are fairly concrete: Note that the particular system of measurement used is not as important as the fact that we apply a consistent set of rules: Although any system of units may seem arbitrary try defending feet and inches to someone who grew up with the metric system! Measurement is not limited to physical qualities such as height and weight. Tests to measure abstract constructs such as intelligence or scholastic aptitude are commonly used in education and psychology, and the field of psychometrics is largely concerned with the development and refinement of methods to study these types of constructs. Although you can test the accuracy of one scale by comparing results with those obtained from another scale known to be accurate, and you can see the obvious use of knowing the weight of an object, the situation is more complex if you are interested in measuring a construct such as intelligence. These issues are particularly relevant to the social sciences and education, where a great deal of research focuses on just such abstract concepts. Levels of Measurement Statisticians commonly distinguish four types or levels of measurement, and the same terms can refer to data measured at each level. The levels of measurement differ both in terms of the meaning of the numbers used in the measurement system and in the types of statistical procedures that can be applied appropriately to data measured at each level. Nominal Data With nominal data, as the name implies, the numbers function as a name or label and do not have numeric meaning. For instance, you might create a variable for gender, which takes the value 1 if the person is male and 0 if the person is female. The 0 and 1 have no numeric meaning but function simply as labels in the same way that you might record the values as M or F. However, researchers often prefer numeric coding systems for several reasons. First, it can simplify analyzing the data because some statistical packages will not accept nonnumeric values for use in certain procedures. Hence, any data coded nonnumerically would have to be recoded before analysis. Second, coding with numbers bypasses some issues in data entry, such as the conflict between upper- and lowercase letters to a computer, M is a different value than m, but a person doing data entry might treat the two characters as equivalent. Nominal data is not limited to two categories. For instance, if you were studying the relationship between years of experience and salary in baseball players, you might classify the players according to their primary position by using the traditional system whereby 1 is assigned to the pitchers, 2 to the catchers, 3 to first basemen, and so on. Consider the example of coding gender so 0 signifies a female and 1 signifies a male. Is there some quality of gender-ness of which men have more than women? Clearly not, and the coding scheme would work as well if women were coded as 1 and men as 0. The same principle applies in the baseball example: The numbers are merely a convenient way to label subjects in the study, and the most important point is that every position is assigned a distinct value. Another name for nominal data is categorical data, referring to the fact that the measurements place objects into categories male or female, catcher or first baseman rather than measuring some intrinsic quality in them. Ordinal Data Ordinal data refers to data that has some meaningful order, so that higher values represent more of some characteristic than lower values. For instance, in medical practice, burns are commonly described by their degree, which describes the amount of tissue damage caused by the burn. A first-degree burn is characterized by redness of the skin, minor pain, and damage to the epidermis outer layer of skin only. A second-degree burn includes blistering and involves the superficial layer of the dermis the layer of skin between the epidermis and the subcutaneous tissues , and a third-degree burn extends through the dermis and is characterized by charring of the skin and possibly destruction of nerve endings. These categories may be ranked in a logical order: However, there is no metric analogous to a ruler or scale to quantify how great the distance between categories is, nor is it possible to determine whether the difference between first- and second-degree burns is the same as the difference between second- and third-degree burns. Many ordinal scales involve ranks. For instance, candidates applying for a job may be ranked by the personnel department in order of desirability as a new hire. This ranking tells you who is the preferred candidate, the second most preferred, and so on, but does not tell you whether the first and second candidates are in fact very similar to each other or the first-ranked candidate is much more preferable than the second. You could also rank countries of the world in order of their population, creating a meaningful order without saying anything about whether, say, the difference between the 30th and 31st countries was similar to that between the 31st and 32nd countries. The numbers used for measurement with ordinal data carry more meaning than those used in nominal data, and many statistical techniques have been developed to make full use of the information carried in the ordering while not assuming any further properties of the scales. For instance, it is appropriate to calculate the median central value of ordinal data but not the mean because it assumes equal intervals and requires division, which requires ratio-level data. Interval Data Interval data has a meaningful order and has the quality of equal intervals between measurements, representing equal changes in the quantity of whatever is being measured. The most common example of the interval level of measurement is the Fahrenheit temperature scale. If you describe temperature using the Fahrenheit scale, the difference between 10 degrees and 25 degrees a difference of 15 degrees represents the same amount of temperature change as the difference between 60 and 75 degrees. Addition and subtraction are appropriate with interval scales because a difference of 10 degrees represents the same amount of change in temperature over the entire scale. However, the Fahrenheit scale has no natural zero point because 0 on the Fahrenheit scale does not represent an absence of temperature but simply a location relative to other temperatures. Multiplication and division are not appropriate with interval data: Ratio Data Ratio data has all the qualities of interval data meaningful order, equal intervals and a natural zero point. Many physical measurements are ratio data: It should be noted that although many physical measurements are ratio-level, most psychological measurements are ordinal. This is particularly true of measures of value or preference, which are often measured by a Likert scale. For instance, a person might be presented with a statement e. These choices are sometimes assigned numbers e. Not from the point of view of a statistician, but sometimes you do have to go with what the boss wants rather than what you believe to be true in absolute terms. Continuous and Discrete Data Another important distinction is that between continuous and discrete data. Continuous data can take any value or any value within a range. Most data measured by interval and ratio scales, other than that based on counting, is continuous: In the course of data analysis and model building, researchers sometimes recode continuous data in categories or larger units. For instance, weight may be recorded in pounds but analyzed in pound increments, or age recorded in years but analyzed in terms of the categories of 0â€”17, 18â€”65, and over Various rules of thumb have been proposed. For instance, some researchers say that when a variable has

10 or more categories or, alternatively, 16 or more categories , it can safely be analyzed as continuous. This is a decision to be made based on the context, informed by the usual standards and practices of your particular discipline and the type of analysis proposed. Discrete variables can take on only particular values, and there are clear boundaries between those values. As the old joke goes, you can have 2 children or 3 children but not 2. In fact, any variable based on counting is discrete, whether you are counting the number of books purchased in a year or the number of prenatal care visits made during a pregnancy. Data measured on the nominal scale is always discrete, as is binary and rank-ordered data. Operationalization People just starting out in a field of study often think that the difficulties of research rest primarily in statistical analysis, so they focus their efforts on learning mathematical formulas and computer programming techniques to carry out statistical calculations. However, one major problem in research has very little to do with either mathematics or statistics and everything to do with knowing your field of study and thinking carefully through practical problems of measurement. This is the problem of operationalization, which means the process of specifying how a concept will be defined and measured. Operationalization is always necessary when a quality of interest cannot be measured directly. An obvious example is intelligence. There is no way to measure intelligence directly, so in the place of such a direct measurement, we accept something that we can measure, such as the score on an IQ test. For a third example, suppose you wish to measure the amount of physical activity performed by individual subjects in a study. Because many of the qualities studied in the social sciences are abstract, operationalization is a common topic of discussion in those fields. However, it is applicable to many other fields as well. For instance, the ultimate goals of the medical profession include reducing mortality death and reducing the burden of disease and suffering. Mortality is easily verified and quantified but is frequently too blunt an instrument to be useful since it is a thankfully rare outcome for most diseases. Examples of operationalization of burden of disease include measurement of viral levels in the bloodstream for patients with AIDS and measurement of tumor size for people with cancer. Decreased levels of suffering or improved quality of life may be operationalized as a higher self-reported health state, a higher score on a survey instrument designed to measure quality of life, an improved mood state as measured through a personal interview, or reduction in the amount of morphine requested for pain relief. Some argue that measurement of even physical quantities such as length require operationalization because there are different ways to measure even concrete properties such as length. A ruler might be the appropriate instrument in some circumstances, a micrometer in others. Even if you concede this point, it seems clear that the problem of operationalization is much greater in the human sciences, when the objects or qualities of interest often cannot be measured directly. Proxy Measurement The term proxy measurement refers to the process of substituting one measurement for another. Although deciding on proxy measurements can be considered as a subclass of operationalization, this book will consider it as a separate topic. The most common use of proxy measurement is that of substituting a measurement that is inexpensive and easily obtainable for a different measurement that would be more difficult or costly, if not impossible, to collect. For a simple example of proxy measurement, consider some of the methods police officers use to evaluate the sobriety of individuals while in the field. Instead, the officer might rely on observable signs associated with drunkenness, simple field tests that are believed to correlate well with blood alcohol content, a breath alcohol test, or all of these. Observational signs of alcohol intoxication include breath smelling of alcohol, slurred speech, and flushed skin. Field tests used to evaluate alcohol intoxication quickly generally require the subjects to perform tasks such as standing on one leg or tracking a moving object with their eyes.

Chapter 8 : Basic Concepts of Measurement - Statistics in a Nutshell, 2nd Edition [Book]

Ratio: Ratio data have the highest level of measurement. Ratios between measurements as well as intervals are meaningful because there is a starting point (zero). Ratios between measurements as well as intervals are meaningful because there is a starting point (zero).

Define and distinguish among nominal, ordinal, interval, and ratio scales Identify a scale type Discuss the type of scale used in psychological measurement Give examples of errors that can be made by failing to understand the proper use of measurement scales Types of Scales Before we can conduct a statistical analysis, we need to measure our dependent variable. Exactly how the measurement is carried out depends on the type of variable involved in the analysis. Different types are measured differently. To measure the time taken to respond to a stimulus, you might use a stop watch. A rating scale is more appropriate in this case with labels like "very favorable," "somewhat favorable," etc. For a dependent variable such as "favorite color," you can simply note the color-word like "red" that the subject offers. Although procedures for measurement differ in many ways, they can be classified using a few fundamental categories. In a given category, all of the procedures share some properties that are important for you to know about. The categories are called "scale types," or just "scales," and are described in this section. Nominal scales When measuring using a nominal scale, one simply names or categorizes responses. Gender, handedness, favorite color, and religion are examples of variables measured on a nominal scale. The essential point about nominal scales is that they do not imply any ordering among the responses. For example, when classifying people according to their favorite color, there is no sense in which green is placed "ahead of" blue. Responses are merely categorized. Nominal scales embody the lowest level of measurement. This is what distinguishes ordinal from nominal scales. Unlike nominal scales, ordinal scales allow comparisons of the degree to which two subjects possess the dependent variable. For example, our satisfaction ordering makes it meaningful to assert that one person is more satisfied than another with their microwave ovens. On the other hand, ordinal scales fail to capture important information that will be present in the other scales we examine. In particular, the difference between two levels of an ordinal scale cannot be assumed to be the same as the difference between two other levels. In our satisfaction scale, for example, the difference between the responses "very dissatisfied" and "somewhat dissatisfied" is probably not equivalent to the difference between "somewhat dissatisfied" and "somewhat satisfied. Statisticians express this point by saying that the differences between adjacent scale values do not necessarily represent equal intervals on the underlying scale giving rise to the measurements. In our case, the underlying scale is the true feeling of satisfaction, which we are trying to measure. What if the researcher had measured satisfaction by asking consumers to indicate their level of satisfaction by choosing a number from one to four? Would the difference between the responses of one and two necessarily reflect the same difference in satisfaction as the difference between the responses two and three? The answer is No. Changing the response format to numbers does not change the meaning of the scale. We still are in no position to assert that the mental step from 1 to 2 for example is the same as the mental step from 3 to 4. Interval scales Interval scales are numerical scales in which intervals have the same interpretation throughout. As an example, consider the Fahrenheit scale of temperature. The difference between 30 degrees and 40 degrees represents the same temperature difference as the difference between 80 degrees and 90 degrees. This is because each degree interval has the same physical meaning in terms of the kinetic energy of molecules. Interval scales are not perfect, however. In particular, they do not have a true zero point even if one of the scaled values happens to carry the name "zero. Zero degrees Fahrenheit does not represent the complete absence of temperature the absence of any molecular kinetic energy. In reality, the label "zero" is applied to its temperature for quite accidental reasons connected to the history of temperature measurement. Since an interval scale has no true zero point, it does not make sense to compute ratios of temperatures. For example, there is no sense in which the ratio of 40 to 20 degrees Fahrenheit is the same as the ratio of to 50 degrees; no interesting physical property is preserved across the two ratios. After all, if the "zero" label were applied at the temperature that Fahrenheit happens to label as 10 degrees, the two ratios would instead be 30 to 10 and 90 to 40, no longer the same! For this reason, it does not make sense to say that 80 degrees is "twice as hot" as 40 degrees. Such a claim would depend on an arbitrary decision about where to "start" the temperature scale, namely, what temperature to call zero whereas the claim is intended to make a more fundamental assertion about the underlying physical reality. Ratio scales The ratio scale of measurement is the most informative scale. It is an interval scale with the additional property that its zero position indicates the absence of the quantity being measured. You can think of a ratio scale as the three earlier scales rolled up in one. Like a nominal scale, it provides a name or category for each object the numbers serve as labels. Like an ordinal scale, the objects are ordered in terms of the ordering of the numbers. Like an interval scale, the same difference at two places on the scale has the same meaning. And in addition, the same ratio at two places on the scale also carries the same meaning. The Fahrenheit scale for temperature has an arbitrary zero point and is therefore not a ratio scale. However, zero on the Kelvin scale is absolute zero. This makes the Kelvin scale a ratio scale. For example, if one temperature is twice as high as another as measured on the Kelvin scale, then it has twice the kinetic energy of the other temperature. Another example of a ratio scale is the amount of money you have in your pocket right now 25 cents, 55 cents, etc. Money is measured on a ratio scale because, in addition to having the properties of an interval scale, it has a true zero point: Since money has a true zero point, it makes sense to say that someone with 50 cents has twice as much money as someone with 25 cents or that Bill Gates has a million times more money than you do. What level of measurement is used for psychological variables? Rating scales are used frequently in psychological research. For example, experimental subjects may be asked to rate their level of pain, how much they like a consumer product, their attitudes about capital punishment, their confidence in an answer to a test question. Typically these ratings are made on a 5-point or a 7-point scale. These scales are ordinal scales since there is no assurance that a given difference represents the same thing across the range of the scale. For example, there is no way to be sure that a treatment that reduces pain from a rated pain level of 3 to a rated pain level of 2 represents the same level of relief as a treatment that reduces pain from a rated pain level of 7 to a rated pain level of 6. In memory experiments, the dependent variable is often the number of items correctly recalled. What scale of measurement is this? You could reasonably argue that it is a ratio scale. First, there is a true zero point: Moreover, a difference of one represents a difference of one item recalled across the entire scale. It is certainly valid to say that someone who recalled 12 items recalled twice as many items as someone who recalled only 6 items. But number-of-items recalled is a more complicated case than it appears at first. Consider the following example in which subjects are asked to remember as many items as possible from a list of Assume that a there are 5 easy items and 5 difficult items, b half of the subjects are able to recall all the easy items and different numbers of difficult items, while c the other half of the subjects are unable to recall any of the difficult items but they do remember different numbers of easy items. Some sample data are shown below.

Chapter 9 : Improving Your Test Questions

84 Part 2 / Basic Tools of Research: Sampling, Measurement, Distributions, and Descriptive Statistics Chapter 7 Measurement M easurement is the foundation of scientific inquiry.

In the records check studies, the samples included known crime victims selected from police records. The studies were done in Washington, D. A key objective of these early studies was to determine the best length for the reporting period for a survey, balancing the need to increase the number of crime reports with the need to reduce memory errors. Some of these studies were inspired by a conference described in Biderman, that brought cognitive psychologists and survey researchers together to examine the memory issues raised by the NCS. Unfortunately, some of the most intriguing findings from the resulting experiments were never published and are buried in hard-to-find memoranda. For several reasons, the NCVS results are widely used as benchmarks to which statistics from other surveys on crime and crime victimization are compared. Measuring Crime and Crime Victimization: Measurement Problems in Criminal Justice Research: The National Academies Press. Results of the first interview, which is necessarily unbounded, are discarded. The initial interview is done face to face to ensure maximum coverage of the population; if necessary, subsequent interviews are also conducted in person. Examples of Measurement Problems Despite these impressive design features and the large body of methodological work that shaped it, the NCVS is not without its critics. Two recent controversies illustrate the problems of the NCVS and of crime surveys more generally. One controversy centers on the number of incidents of defensive gun use in the United States; the other concerns the number of incidents of rape. In both cases, seemingly similar surveys yield widely discrepant results; the ensuing methodological controversies point to unresolved issues in how to collect data on crime, gun use, and crime victimization in surveys. Defensive Gun Use In , McDowall and Wiersema published an estimate of the number of incidents over a four-year period in which potential crime victims had used guns to protect themselves during an actual or attempted crime. Their estimate was based on data from the NCVS, which gathers information about several classes of crimeâ€"rape, assault, burglary, personal and household larceny, and car theft. The key estimates McDowall and Wiersema presented were that between and there were some , incidents of defensive gun use in the United States, roughly 65, per year. Although big numbers, they pale by comparison with the total number of crimes reported during the same periodâ€"guns were used defensively in fewer than one in victimizations reported in the NCVS; moreover, criminal offenders were armed about 10 times more often than their victims. These are just the sort of statistics dear to gun control advocates. McDowall and Wiersema note, however, that their estimates of defensive gun use differ markedly from those based on an earlier survey by Kleck ; see also Kleck and Gertz, These numbers were derived from a national telephone survey of 1, registered voters who were asked: The two surveys covered different populations the civilian noninstitutional population in the NCVS versus registered voters with a telephone in the Kleck survey , interviewed respondents by different methods in-person versus telephone , covered different recall periods six months in NCVS versus five years in the Kleck study , and asked their respondents markedly different questions. Still, the difference between 65, incidents a year and some , is quite dramatic and would seem to demand a less mundane explanation than one involving routine methodological differences. A later telephone survey by Kleck and Gertz yielded an even higher estimateâ€"2. McDowall and Wiersema cite two other possible explanations of the differences between the results of the NCVS and the earlier Kleck study. The later Kleck and Gertz estimates also rest on a similarly small base of positive reportsâ€"66 out of nearly 5, completed interviews. Even a few mistaken respondents could have a large impact on the results. The NCVS excludes preemptive use of firearms e. It is possible that much of the disparity between the NCVS estimates and those derived from the two Kleck studies reflects the broader net cast in the latter surveys. Respondents were asked both sets of questionsâ€"both written to cover a one-year recall periodâ€"and the experiment varied which ones came first in the interview. The sample included 3, respondents, selected from a list of likely gun owners. Overall, the Kleck items yielded three times more reports of defensive gun use than the NCVS-style items. What was particularly interesting in the results was that the two sets of items appeared to yield virtually nonoverlapping sets of incidents; of the 89 reports of defensive gun use, only 9 were mentioned in response to both sets of items. Prevalence of Rape An even more disparate set of figures surrounds the issue of the number of women in the United States who have been the victim of attempted or completed rapes. Once again, the studies from which the estimates are drawn differ in many crucial particularsâ€"they sample different populations, ask different questions that are based on different definitions of rape, conduct data collection via different methods, and cover different recall periods. As with the estimates of defensive gun use, what is surprising is not that the estimates differ from each other but that they differ so widely. Several studies converge on the estimate that about one-quarter of American women have been victims of completed or attempted rape at some time in their lives see, for example, Koss, Most of these figures do not accord well with the rape estimates from the NCVS; the NCVS covers a more limited periodâ€"six monthsâ€"and does not produce estimates of lifetime victimization. The NCVS figure translates into fewer than , victims since the same person may have experienced multiple victimizations. Lynch explores a number of differences between the two studies, including: Despite the methodological differences between the two surveys, the difference between the two estimates is probably not significant. The estimates from both surveys have large standard errors approximately , for the NWS estimate and approximately 32, for the NCVS , and the standard error of the difference is on the order of , One major difference between the NCVS and most of the other surveys assessing the frequency of rape involves the basic strategy used to elicit reports about rapes and other crimes. The NCVS begins with a battery of yes-no items designed to prompt reports about a broad array of completed or attempted crimes. Once the respondent completes these initial screening items, further questions gather more detailed information about each incident; the final classification of an incident in the NCVS reflects these detailed reports rather than the answers to the initial screening questions. Most of the other surveys on rape differ from this procedure in two key waysâ€"first, they ask multiple screening questions specifically crafted to elicit reports about rape and, second, they omit the detailed follow-up questions. For example, a survey by Koss, Gidycz, and Wisniewski included five items designed to elicit reports of attempted or completed rape. The items are quite specific. There is little doubt that including multiple concrete items will clarify the exact concepts involved and prompt fuller recall. Multiple items provide more memory cues and probably trigger more attempts at retrieval; both the added cues and the added time on task are likely to improve recall Bradburn and Sudman, ; Burton and Blair, ; Cannell et al. The NCVS is a general-purpose crime survey, and its probes cover a broad array of crimes. The NWS and the Koss surveys use much more detailed probes that focus on a narrower range of crimes. At the same time, the absence of detailed information about each incident could easily lead to classification errors. A study by Fisher and Cullen included both yes-no screening items of the type used by Koss and colleagues, the NWS, and many other studies of rape and the more detailed questions about each incident featured by the NCVS. They compared responses to the screening questions with the final classifications of the incidents based on the detailed reports. There were twice as many positive answers to the rape screening questions as there were incidents ultimately classified as rapes based on the detailed reports. The rape screening items also captured many incidents involving some other type of sexual victimization. In addition, some incidents classified as rapes on the basis of the detailed information were initially elicited by screening items designed to tap other forms of sexual victimization. The results suggest that, even when the wording of screening items is quite explicit, respondents can still misclassify incidents. Factors Affecting Reporting in Crime Surveys Many surveys on sensitive subjects adopt methods primarily designed to reduce underreportingâ€"that is, the omission of events that should, in principle, be reported. And it is certainly plausible that women would be reluctant to report extremely painful and personal incidents such as attempted or completed rapes. There are also reasons to believe that crime surveys, like other surveys that depend on recall, may be prone to errors in the opposite direction as well. Because crime is a relatively rare event, most respondents are not in the position to omit eligible incidents; they do not have any to report. The vast majority of respondents can only overreport defensive gun use, rapes, or crime victimization more generally. In his discussion of the controversy over estimates of defensive gun use, Hemenway makes the same point. All survey questions are prone to errors, including essentially random reporting errors. For the moment, let us accept the view that 1 percent of all adults used a gun to defend themselves against a crime over the past year. If the sample accurately reflects this underlying distribution, then only 1 percent of respondents are in the position to underreport defensive gun use; the remaining 99 percent can only overreport it. Even if we suppose that an underreport is, say, 10 times more likely than an overreport, the overwhelming majority of errors will still be in the direction of overreporting. It is not hard to imagine an error rate of the magnitude of 1 in 40 arising from respondent inattention, misunderstanding of the questions, interviewer errors in recording the answers, and other essentially random factors. Even the simplest survey itemsâ€"for instance, those asking about sex and ageâ€" yield less than perfectly reliable answers. Random errors can, in the aggregate, yield systematic biases when most of the respondents are in the position to make errors in only one direction. Aside from sheer unreliability, though, reporting in crime surveys may be affected by several systematic factors that can introduce additional distortions of their own. We focus on two of these systematic factors here. First, we address the potentially sensitive nature of the questions on many crime surveys and the impact of the mode of data collection on the answers to such questions. This is followed by an examination of the effects of the context in which survey items are presented, including the physical setting of the interview, the perceived purpose and sponsorship of the study, and prior questions in the interview. Page 18 Share Cite Suggested Citation: For example, Koss lists 20 surveys on sexual victimization of women; only 4 all of them involving local samples from individual communities appear to use self-administered questionnaires. The remainder rely on interviewers to collect the data, in either face-to-face or telephone interviews. The NCVS uses both methods; the initial interview is done face to face, but later interviews are, to the extent possible, done by telephone. The last decade has seen dramatic changes in the methods used to collect survey data, including the introduction of several new methods of computerized self-administration. With ACASI a computer simultaneously displays the item on screen and plays a recording of it to the survey respondent via earphones. The respondent enters an answer directly into the computer using the keypad. Two trends have spurred the development and rapid adoption of new methods of computerized self-administration of surveys. First, various technological changesâ€"such as the introduction of lighter, more powerful laptop computers, development of the World Wide Web, widespread adoption of e-mail, and improvements in sound card technologyâ€"have made the new methods possible. Second, the need for survey data on sensitive topics, such as illicit drug use and sexual behaviors related to the spread of AIDS, has made the new methods highly desirable, since they combine the privacy of self-administration with the power and flexibility of computer administration. Widespread interest in the new methods has spurred survey methodologists to reexamine the value of self-administration for collecting survey data on sensitive topics. Gains from Self-Administration There is strong evidence to support the value of self-administration for eliciting reports about sensitive behaviors. For example, if 6 percent of respondents report using cocaine during the previous year under self-administration but only 4 percent report using cocaine under interviewer administration, the ratio would be 1: The data are from two of the largest mode comparisons done to date: The ratios range from a little over 1: Tourangeau, Rips, and Rasinski They also summarize the evidence that self-administration improves reporting about other sensitive topics, including sexual partners, abortion, smoking, and church attendance. Mode, Privacy, and the Presence of Third Parties It is natural to think that at least some of the gains from self-administration result from the reduced risk of disclosure to other household mem Page 20 Share Cite Suggested Citation: Interviews are often conducted under less than ideal conditions, and, although most survey organizations train their interviewers to try to find private settings for the interviews, other household members are often present. For example, Silver and colleagues examined the proportion of interviews done for the American National Election Studies ANES in which other household members were present. The proportion varied somewhat from one survey to the next, but roughly half of all interviews conducted between and were done in the presence of another household member Silver et al. Similarly, Martin and colleagues noted that some 58 percent of NCS interviews were conducted within earshot of someone other than the interviewer and respondent for a more recent estimate, see Coker and Stasny,