Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series/ Un Réseau stratégique de connaissances Changements de population et parcours de vie Document de travail

Volume 3 | Issue 3 Article 1

March 2015

Error in Demographic and Other Quantitative Data and Analyses

Thomas K. Burch Prof
Univ. of Victoria, Canada, tkburch@uvic.ca

Error in Demographic and Other Quantitative Data and Analyses*

Thomas K. Burch
University of Victoria
[Revised 15 March 2015]

* This is a rough draft version of comments at a Population Research Group [Univ. of Victoria] Seminar, 28 November 2013. Criticisms and suggestions welcome.

Introduction

Today I would like to talk about errors in statistical data. My remarks are in two parts. The first part deals with the proposition that the statistical data we consume, analyse and produce contain more error from more sources than we sometimes recognise. The second asks: How can we better deal with these errors? In thinking of subtitles for the talk, things like 'catalogue of horrors' and 'counsel of despair' came to mind. But I shall try to be positive.

The topic of error in statistics can become highly technical, but for the most part this is not a technical presentation. It deals with common sense and experience; much of it is anecdotal.

Deming on Errors in Survey

I begin with a classic article entitled 'On errors in surveys,' by W. Edwards Deming [1944]. Deming is best known as one of the founders of modern statistical quality control in industry. His ideas were adopted by the Japanese auto industry some decades ago, and are widely credited with enabling the Japanese to produce higher quality automobiles than Detroit. Deming was trained in electrical engineering, with MS and Ph.D. degrees in mathematics and physics. He studies statistics with R.A Fisher and Jerzy Neyman, and worked with Walter A. Shewhart an earlier pioneer in industrial quality control.

Beginning in the 1930's Deming worked on surveys for various U.S. government agencies. During this period he and other government statisticians [Tepping, Geoffry, Hansen, F.F. Stephan, Hurwitz, Madow] developed much of the theory and practice of complex sampling design. The 1944 article was based on this experience. Like my presentation, his article is non-technical; there are no formulas, equations or statistical tables.

As an aside, it is worth noting that Deming's paper is 70 years old. This poses the question whether all of his concerns are still relevant. We have 70 more years of experience with surveys and other forms of data collection. And we have new tools, not least the computer, and new statistical techniques. But to me, his ideas are as relevant today as they were then, possibly even more so given the commercialisation of survey and other statistical research and such ‘inventions’ as ‘big data.’ As I see it, we are inundated with shoddy data and naïve or even deliberately faulty analysis. Data and statistical analyses often become just another form of propaganda or advertising.

A recent wholesale critique of social science data, especially economic data, is by a Spanish university economist, Phillip Bagus: ‘Morgenstern’s forgotten contribution: a stab to the heart of modern economics [2011]. Bagus revisits Morgenstern’s 1950 monograph [2nd edition in 1963] *On the Accuracy of Economic Observations.* Oskar Morgenstern was a Princeton economist, co-inventor of game theory, with the mathematician John Von Neumann. He notes that, in contrast to physics, there is no estimate of statistical error within economics. He continues: ‘The problem of error in economics observations is still a widely neglected problem. The various sources of error that come into play in the social science suggest that the error in economic observations is substantial. As the error might be substantial, this paper argues that the usefulness of econometrics becomes questionable’ [Abstract, p.540]. His conclusion may be extreme, but the problems he addresses are real.

Many of Bagus’s concerns are echoed in a recent and much cited work on income and wealth inequality – *Capital in the Twenty-First Century*, by Thomas Piketty [2014] a French economist. The book is based largely on macro-economic data – long time-series for many nations. But the author regularly discusses the weaknesses of the data before using it. And he seldom relies on one data set alone to reach his conclusions. He triangulates, using data as only one part of a well-developed argument.

Both of these works suggest that Deming’s concerns about error are as relevant as ever, even in economics, arguably the strongest social science discipline.

Deming begins: ‘There are thirteen different factors that affect the usefulness of surveys. The chief aim of this article is to point out **the need for directing effort toward all of them when planning a survey**, and the futility of concentrating on only one or two of them’ [p.359, emphasis added]. He notes that the thirteen factors are not always distinguishable, and that ‘there are other ways of classifying them.’ I don’t know whether he chose thirteen because it is often considered an unlucky number, but it is an appropriate choice.

Here is his list:

1. Variability in response [respondents may give different answers on different occasions]
2. Differences in kind [e.g., personal interview vs. phone] and degree of canvass [long or short interviews or questionnaires; check-off questions or deeper questioning with probes and follow-ups to answers]
3. Biases and differences arising from the interviewer
4. Biases of the auspices
5. Imperfections in design of the questionnaire and tabulation plans
6. Changes in universe before tabulations are available
7. Biases from non-response [including omissions]

8. Biases from late reports
9. Biases from unrepresentative selection of date or period covered
10. Biases from unrepresentative selection of respondents
11. Sampling errors and biases
12. Processing errors
13. Errors in interpretation

A discussion of each of the thirteen sources of error would take too long. So I shall focus on a few concrete examples, each of which illustrates one or another of the sources. Note that 'unrepresentative selection of respondents' [#10] and 'sampling errors and biases' [#11] are only two of the thirteen factors.

At the outset, it should be said that in all but trivial cases [for example, counting the number of persons in this room], collecting accurate statistical data is not easy, despite widespread opinion to the contrary. It is difficult enough when one is trying hard to produce good data. It's all the harder if the persons providing, collecting or publishing the data don't really care about data or statistics, and don't know much about them.[1] Then there is the case where sponsors of a survey have a vested interest in certain results, and pay to get them. One can lie with numbers as well as with words, and people and organisations often do. It is worth re-reading *How to Lie with Statistics* [Huff, 1954] from time to time. In the non-scientific, commercial arenas, worthless data and statistical reports are commonplace. Government agencies such as Statistics Canada and the US Bureau of the Census try hard to produce high quality data and generally do. But, as we shall see later, even they can fall short. High-profile polls are often among the worst offenders.

The Literary Digest

My first example is perhaps the biggest polling fiasco in the history of polling. *The Literary Digest*, a successful American magazine, had conducted polls that correctly predicted the winner in each presidential election between 1916 and 1932 [much of this material is from Dennis Deturck, www.math.upenn.edu/~deturck/m170/wk4/lecture/case1.html]. As the 1936 election approached, they predicted that Alfred Landon, Republican governor of Kansas, would beat Franklin D. Roosevelt, running for his second term. Their polling method was that of a short mail-out/mail-back sample ballot.

The prediction was that Landon would beat Roosevelt 57% to 43%. The result: Roosevelt beat Landon 62% to 38%. The 19% error in their prediction is thought to be the largest ever in a major public opinion poll.

[1] During the Vietnam War, Robert McNamara pushed combat officers to provide statistics on 'body counts,' 'kill ratios,' and other things. Many later admitted that they had simply made up the numbers. On the battlefield they had better things to do.

There were two major problems with the poll. The first arose from their identification of the target population, or 'sampling frame.' The list of recipients was compiled from telephone directories, club membership lists, lists of magazine subscribers, etc., all sources tilted towards persons of higher socio-economic status. Note this was near the end of the Great Depression and telephones were still something of a luxury. Thus, even a perfect sample of this population would be biased.

But their sampling was far from perfect. Approximately 10 million questionnaires were sent out, but only about 2.4 million responded, for a response rate of under 25%. The magazine appeared to believe that the large sample would guarantee good results, and they apparently were thorough in their analysis: '…the incoming tide of marked ballots…will be triple-checked, verified, five times cross-classified, and totaled. When the last figure has been totted and checked, if past experience is a criterion, the country will know to within a fraction of 1 percent the actual popular vote of forty million [voters].' *The Literary Digest* ceased publication a few years later.

One object lesson is that the largest sample in the world cannot correct for bias in the definition of the sampled population. A large sample from a biased frame is a biased sample – it hits the wrong target with great reliability. A corollary is that a smaller sample, which allows resources to be devoted to other sources of error, may well be superior. The fledging pollster George Gallup, working with a sample of about 50,000, correctly predicted the outcome of the election.

A second object lesson is that self-selection in any form – in volunteering for a panel or survey, in returning a questionnaire, in non-response to particular questions, etc. – is typically a significant source of bias.

Although survey techniques and statistical methods have improved greatly over the intervening 80 years, not a few contemporary surveys are still characterised by biased sampling frames and various forms of self-selection, notably in mailed questionnaires and online polls.[2]

The Case of the Teen-Age Widows

In 1953 the US Bureau of the Census published a special report on marital status. Table 5, entitled 'Single Years of Age: Persons 14 Years Old and Over by Marital Status and Sex.' An excerpt from this table contains the following data for males:

[2] For a recent discussion of state-of-the-art surveys, see Weisberg, 2005.

Age	Marital Status	
	Widowed	Divorced
14	1,470	1,470
15	1,380	840
16	750	690
17	840	630
18	720	960
19	570	1,740
20	1,260	4,170
21	1,530	7,170
22	1,920	10,770
23	1,860	13,200
24	2,310	15,150

What do we make of these data? One of my professors in graduate school, Frederick F. Stephan looked at the above data and thought to himself 'These data are completely incorrect; they're nonsense.'

Stephan, coincidentally, had been a colleague of W. Edwards Deming at the Bureau of the Census. He was one of the group mentioned earlier working to develop modern complex sampling designs. And he was one of the most critical social scientists I have ever encountered – in the good sense of the word. He questioned everything, and left no easy assumptions unchallenged.

When he saw the above data, he rejected it as error. He found it impossible to accept that there would be a substantial number of widowed and divorced 14 year old males. [As an aside, some states in the US allowed marriage as early as 14, especially for females.] Even if that number were correct, it was impossible to accept that from age 14 that number would tend to decline, such that there were more males at age 14 than at higher ages – until ages 21 and 19 respectively. Finally, it was impossible to believe that there would be more 14 year-old males than females in these two categories [data not shown], since females married earlier than males.

Stephan teamed with Ansley J. Coale, who already had done a lot of work evaluating the accuracy of US censuses. He had shown, for example, that age reporting in the census was more accurate [less age-heaping] if the question asked was date of birth rather than age [both had been used in the past].

In 1950, census data were still processed using IBM cards. Data from the census schedules was manually punched into cards which were then run through sorters and other IBM machines. Coale and Stephan eventually traced the error back to one defective card-punch machine; it began entering data one column to the right of the first column, thereby shifting all the rest of the data one column to the right. As a result, head of household data became 'race'; race became 'sex,' so that White became male and Negro female, etc. Marital status would now be determined by the second digit of age. The erroneously designated teen-agers were really middle-aged males.

Overall, the number of cards affected was small; they estimate that the percentage of cards affected for various sub-categories of population ranged between 0.01% to 0.5%. Who wouldn't be happy with errors of that magnitude? The problem is that the errors dumped cases into cells whose frequencies should have been zero, or very close to zero – for example, 14 year-old male widowers. Their lesson: '…users must scrutinize numbers in small cells of census tabulations with special care' [p. 346]. Recall that Statistics Canada often supresses data based on small cell frequencies.

The punching operation was checked for quality control on a sample basis. Coale and Stephan note that a 100% check might have caught this particular error, but that would have doubled the cost of punching. In their view such a reallocation of funds might well have caused other, more important deficiencies in the census. They – and Deming – would argue for a balanced approach to the use of resources, taking into account all possible sources of error, not just one or two.

A reaction to this story might be that with the replacement of IBM machines by computers, this kind of error is now much less likely. I would comment that in computer programming, a common error – common enough to give it a special name – is the 'one-off error.' This can take many forms. A loop is ended too soon or too late because of confusion in the use of signs: =, > or =, >, etc. Or, there is confusion with regard to the indices in arrays. Many physical scientists begin counting at zero. In statistics we begin counting at 1. Mathematics software often use the zero convention. In one of my favourite programs, MathCad, and in programming languages such as C++, the default is zero; if I want to begin counting at 1, I must change it. Although card-punch machines are no longer with us, one-off errors still are. The TV rating company Nielsen recently revealed that 'due to software errors' the ratings assigned two prime-time programs had been interchanged.

Stephan's aphorism based on this work was: 'If the data look wrong, there's a good chance they are.' Another came up when I was presenting a dissertation proposal at a brown-bag – a study of Latin American fertility, using routine vital statistics and census data. Someone remarked that Latin American birth and death registration data were notoriously poor. I replied, 'Well, they're better than nothing' – to which Stephan remarked 'Unless they're completely misleading.'

US and Canadian Labour-Force Data

For a more up-to-date example, let's stay with first-rate government statistical agencies.

It came to light recently that temporary foreign workers had been hired to work in a cafeteria on a Native reserve in Saskatchewan. The unemployment rate on the reserve was estimated to be 30-40%. But the unemployment rate in the larger geographic area containing the reserve was only 4.5%, below the threshold required for the hiring of temporary foreign workers. Part of the explanation is that the monthly labour-force survey does not include Native reserves [The 23 Jan. 2015 edition of the *Globe and Mail* featured this story on its front page, with a banner headline]. Given estimates of Natives on reserves of 300-350 thousand, this exclusion would have little effect on national figures. But at the local level, it can make a big difference. In the case at hand, it could have raised the unemployment rate above the threshold.

It is not entirely Statistics Canada's fault. Some reserves refuse to co-operate with the census and other government surveys. Others are in remote locations. But it's worth thinking about before taking a monthly unemployment rate as absolute truth.

In the US the Bureau of Labor Statistics also collects monthly data on employment. A survey of individuals produces unemployment figures and classifications of workers as full-time/part-time, 'long-term' unemployed, 'discouraged worker,' etc. The official definition of 'full-time worker' is surprising: anyone working over 35 hours per week. It doesn't matter whether the 35 hours are amassed in one job or 4. So, a person who works 40 hours in three different jobs, with no benefits [health-care, retirement, paid holiday, etc.] is classified as a full-time worker – not what most of us would think of as a real job.

When an interviewer brought this to the attention of Paul Krugman [Princeton economist and winner of the Nobel Economics Award], his response was 'No, I didn't know that.'

To add to the confusion, in a labour-force survey of businesses rather than individuals, it is the business themselves who define 'full-time' – each is free to choose its own definition. Does the US really know how many Americans have real full-time jobs?

I note in passing that the US has the highest incarceration rate in the world, not counting Seychelles. By some estimates, if all prisoners were included as being in the labour market, the US unemployment rate would be as much as two percentage points higher.

Governments are not above falsifying data or fudging. A *Globe and Mail* column for 17 Dec. 2014 is headlined 'China's GDP revision a new way to be wrong.' The author, John Foley, comments: 'China's economic output just increased by 3 per cent, or roughly $275-billion [US], with just the flick of a pen. The statistical revision is designed to improve reported GDP rather than just increase it. Since China's economic aggregates are of diminishing use, the result is just a new way to be misled' [p.B8].

The European debt crisis of 2009 on was related in no small measure to the practice of some EU members, notably Greece, of hiding debt through 'securitisation' deals with major international banks. Their published financial accounts were designed to mislead.

Opinion Polls and Other Commerical/Political Surveys

We started with the non-scientific *Literary Digest* poll. Let me conclude my concrete examples with some more recent polls, most of them of local interest.

Victoria's Vital Signs: For the ninth year, the Victoria Foundation has conducted 'Greater Victoria's Annual Checkup,' on 'community wellbeing.' It is difficult to find a clear statement of survey methodology, or, indeed, whether there were one or two different surveys.

On p.6, the report contains two brief paragraphs on 'How the Survey Was Done.' In May, 'invitations to participate were mailed to **15,841** [emphasis in original] randomly selected households, representing **10% of all households** [emphasis in original] in the region, and divided proportionately across all municipal areas.' One person age 18 and over was invited to respond for each household. The results were weighted by age, sex and geographic location using 2011 census data. Earlier, on p.6 it is noted that 2,239 people completed the survey, for a response rate of 14%. The report then comments 'The high number of respondents creates a statistically accurate snapshot of wellbeing in Greater Victoria….' Note the equation of large sample with accurate data.

The report also mentions from time to time a 'Vital Signs Survey,' with indications that 1,725 people [1,449 plus 276 youth] participated in an online survey, a different number from the one noted above. Page42 provides some details. The survey was done online, with a 'self-selected' respondent population, and 'data are accurate to within +/-3%....' It then notes that 'Most market research involves much smaller samples of around 400 and a margin of error of plus or minus 5%....'

Amalgamation Yes: Amalgamation Yes describes itself as a grass-roots organisation promoting amalgamation in the Capital Regional District of Victoria, BC. Their large website contains a section entitled 'Research and Studies' and a sub-section entitled 'Polls.' It features an Angus Reid poll done for Amalgamation Yes, but contains excerpts from others.

For example, an April 2013 online poll associated with the Black Press reports that 74% of respondents favour amalgamation to reduce the number of municipal governments. The results are preceded by a disclaimer: 'The web poll is informal, not scientific. It represents the opinions of site visitors who voluntarily participate. Black Press is not responsible for the statistical accuracy of opinions expressed here.' Sample size is 126.

A 2014 poll attributed to Survey Monkey is claims 45.45% of respondents [note the use of two decimal places, perhaps to suggest accuracy] favour the province forcing amalgamation. There are no respondents from five municipalities in the CRD [Sooke, Colwood, Metchosin, Highlands or View Royal]. Sample size is 22! A simple random sample of this size would have a probable error of +/-21%. This is one of dozens of sites offering to do 'web polls.' One advertises 'Make your own poll in minutes. Get instant results.' Some will pay for responses: 'Earn $25-50 per day completing online surveys in your spare time.'

The largest space is given to the poll by Angus Reid, generally considered a reputable polling organisation. In July 2014, they surveyed 441 respondents from the 13 CRD communities. 84% are found to favour amalgamation, with only 12% opposed. The report contains the following statement regarding error: 'A probability sample of this size would carry a margin of error of +-4.7 percent 19 times out of 20.' Note that they do not claim to have a probability sample. Translation: 'If we had a probability sample – but we don't -- the error would be +-4.7%.' Clearly this is misleading; I would say it borders on the fraudulent.

In 2007 Angus Reid switched from telephone polling to online polling. In effect they establish panels of respondents containing 'enough panelists from every important demographic…to guarantee that samples are representative of the entire adult population' [see 16 page 'white paper' on Angus Reid website]. In traditional sampling terminology, these would be quota samples. In any given poll, a random sample is drawn from the panel. I could find no information regarding response rates of those chosen from the panel. Some panelists agree in advance to participate. And an incentive of $1 to $5 is provided for each time they participate. But results are weighted, which suggests that response is not 100%.

Angus Reid justifies the use of online polls by noting that four out of five Canadians have access to the internet. This of course means that one out of five – a fifth of the population – is totally excluded from their polls. They also point to a track record of successful electoral forecasts in Canada, the US and the UK, better, they claim, than any other major polling organisation. Since persons who actually vote are far from being a random sample of the adult population, I don't see this as a compelling argument for the representativeness of their non-electoral polls. Angus Reid, along with all the other pollsters, gave the May 2013 BC provincial race for Premier to the NDP; the Liberals won.

Manulife on Debt in Retirement: In *The Times Colonist* [Victoria] recently, there's an interesting case where for once the headline is probably more accurate than the figures quoted in the story: 'Some Canadians struggle to balance retirement, debt.' The text goes on to give specific figures, based on an online poll conducted by Research House for Manulife. The final paragraph quotes the polling industry's professional body, the Marketing Research and Intelligence Association, to the effect that online polls cannot be assigned a margin of error, and that the sample may not be representative. I would say that Manulife wasted its money, except perhaps from a purely public relations perspective. They wrap their statements in the mantle of 'research.'

Some Concluding Thoughts

What can we learn from all of this? Are there some broad principles or rules-of-thumb to guide our day-to-day work? I would suggest the following:

1] Be skeptical of reported data, especially from interested parties, but also from objective scientific organisations and researchers. We all make mistakes even if we are not all liars. Recall F.F. Stephan's aphorism: 'If the data look wrong, there's a good chance they are.'

2] Accept that all our data and calculations are approximations to the real world. As empiricists we tend to think our data give us Truth. Social theorists tend to think their theories are Truth. We are both wrong. Scientific theories, models and data are simply tools that may or may not help us understand small portions of a vast and complex world. None is literally and absolutely true. Incompleteness and error are inevitable.

3] Recognise at the same time that for most our purposes we don't need high levels of precision. Morgenstern quotes Norbert Wiener as saying 'Economics is a one or two digit science,' a statement that applies equally to sociology and demography. What substantive theory or hypothesis would rise or fall depending on what happens in the second decimal place? Would it make a real difference if this month's unemployment rate were to be 6.78 rather than 6.84? They're both around 6.8, which is very close to 7%. And as we know already, neither is correct. Would my model of the determinants of fertility be falsified, if my measure of the total fertility rate were 1.786 rather than 1.689?

I recall one text on demographic methods as saying that method A of calculating q_x is better than method B because it is more accurate in the fifth decimal place. As another example, there is a substantial literature on techniques for calculating the stable population model, with claims that one is superior because the resulting intrinsic growth rate r is 'more accurate' in the third or fourth decimal place. But the stable model is an abstract one-sex model that assumes unchanging are-specific fertility and over time, with no migration. What does 1.134% tell us that 1.0% does not? In both of the above examples, the basic demographic data could easily contain errors of 4 or 5%, if not larger. And conventional procedures would use data in five-year age intervals, which would often introduce more error. Only one or two digits are justified; only one or two digits are needed.

The biological ecologists Puccia and Levins, writing about 'loop analysis' [path analysis without numerical coefficients, only signs] comment that in interventions to save a lake from pollution, the most one can hope for from models and analysis is that one is pushing the system in the right direction, that is, not making things even worse. Precise numbers are not meaningful. The system is too complex, and too many variables are unmeasured or unmeasurable.

There are situations in which precision matters, of course. Against one other candidate, I win or lose an election if my percentage of votes is 50.001 or 49.999. But these seldom occur in demography or other social science.

4] Recognise the phenomenon known as the propagation of errors. There is a very technical branch of applied mathematics devoted to this. All we need to know is that initial errors travel through our algorithms in ways not always easy to predict. Many take consolation in the view that 'errors cancel one another out.' This may or may not be true, but we shouldn't count on it. Consider the following simple examples, with A and B as numbers and a and b their relative errors:

Division: aA/bB: the errors cancel yielding the correct result ONLY if a = b
Subtraction: aA – bB: if b = a, then the result is a(A-B), incorrect.

The results will differ considerably depending not only on whether a = b or not, but whether one is great than 1 [e.g., 1.03] while the other is less than 1 [e.g., 0.95].

Even in these simple calculations, cancellation is probably the exception, not the rule. Imagine what might happen when one is dealing with complex algorithms such as a life table or a stable population model.

Most importantly, we need to know that the errors in our results may not be proportionate to errors in our input data; the error may escalate. Morgenstern gives a dramatic example using a simple set of two simultaneous equations in two variables. The equations are identical except for a difference small difference in the fifth decimal place of the y exponent. The solutions have opposite sign and differ by about 200,000.

Morsgenstern's equations:
$x – y = 1$
$x – 1.00001\ y = 0$
solution: x = 100,001, y = 100,000
$x – y = 1$
$x – 0.99999\ y = 0$
solution: x = -99,999, y = -100,000

The case of the teen-age widows is a prime example. Another is standard demographic technique for calculating what are known as residual estimates of net internal migration. For age groups alive at two censuses, the population of an area is 'survived' from time 1 to time two using some estimate of mortality [typically either life-table or census survival ratios]. The difference between the projected and enumerated populations at time 2 is attributed to net migration. But the number of net migrants is almost always much smaller [by orders of magnitude] than the populations being differenced. Small relative errors in one or both population figures results in large relative errors in estimated net migration. In the early literature on this subject, researchers pointed to errors or 40% or more, or estimates that had the wrong sign – showing net migration in instead of net migration out [Truesdell, 1938 warned of calculations of this sort, which he termed 'residual relationships'].

5] It follows that we should be generous in rounding our results. With computers it is easy to carry as many digits as we like during our calculations. Rounding need only occur in the final result[s].

6] It is useful to try to get some idea of the impact of possible errors by means of some sort of 'interval arithmetic.' In calculating a crude birth rate, for example, we might assume that 97% of births were registered and 96% of population enumerated, and see how much difference that makes to the resulting CBR. Different plausible patterns of assumed errors might point to a CBR in the range 11 to 13 per 1,000. Simple calculations of this sort are easy to do on an ordinary spreadsheet. Some mathematical software [Mathematica, Matlab, Maple] facilitates more complex calculations, with input numbers expressed as ranges. The most user-friendly software for this purpose – now no longer available – was called FuziCalc. It was a full-featured spreadsheet that worked with fuzzy numbers, and, again with input expressed as a range and a distribution, with high, low and most likely figures.

7] Adopt better methods of approximation. Rather than relying on data in five-year intervals, for example, and using the mid-point of the age group in other calculations [these will almost certainly be incorrect for the youngest and oldest age intervals], we can transform such discrete data into continuous form using software such as TableCurve or Scientist. These facilitate finding a function that best fits the discrete data at hand [an 'approximating function']. Further calculation can be done using calculus [integrals, derivatives, etc.] rather than finite sums [See Burch, 1995 for an example relating to numbers of kin].

8] We need to be careful about relying solely on one data set or one survey. Replication is required before an empirical finding is secure. One of the problems with many polls, incidentally, is that they do not provide sufficient detail on methodology such that someone else could try to reproduce their results. Methodology is often considered proprietary, a business secret.

9] Closely related to replication is the notion of triangulation, coming at a research question or hypothesis from several different angles. Two early studies on differential mortality by sex provide a good example of the value of triangulation and the perils of looking at only one data set.

Francis Madigan [1957] carried out an imaginative study comparing the mortality experience of Catholic religious teaching orders of men and women – brothers and sisters. He had large samples, with excellent documentation on age, age at death, age at entry to the order, education, etc. The subject populations were selected for reasonably good health and intelligence, led similar life styles, and engaged in similar work. It seemed like a perfect natural experiment. Madigan reasons that if a mortality differential was found in this population, he could conclude that it was due largely to biological rather than to behavioural, socio-cultural or environmental factors.

In the event, a substantial differential was observed. Madigan's study was clever and carried out with great care. In retrospect, it had one notable flaw. He actually mentions it in passing but failed to grasp its potential importance. 'It must be admitted that the Brothers are more likely to smoke and to take an occasional drink' [p.204]. It must be noted in fairness that the first US Surgeon General's report on smoking and health was in 1964. Madigan published in 1957.

Preston's research [1970] could benefit from the steadily accumulating evidence that smoking is a serious health risk. But even so he used a much broader approach than Madigan, in effect triangulating. Preston looked at time trends in mortality by sex, available data on tobacco use from various sources, sex differentials in cause of death, and various other data. In a statistical study of international mortality trends, he found correlations with earlier mortality declines and factors like stress and mobility. But, he notes, '…the correlations are weak and become trivial when account is taken of the presence of the physical factor examined, cigarette smoking' [p.18].

Madigan relied on one source of data and got it wrong. Preston relied on several different kinds of data and a many-stranded argument, looking more closely at possible causal systems, and got it right.

Deming is quoted as saying: 'Information is not knowledge.' Statistical description and analysis without substantive theory is a risky business.[3]
10] When all is said and done, however, we can't give up on quantitative data, however. Most important questions in social science – and in life -- are matters of more or less, not either/or.

*** *** ***

---

[3] I was once giving a talk to an Epidemiology Department on 'Computer Modelling of Theory.' I had barely started when one of the faculty said 'You should know at the outset that we don't do theory. We only do statistical studies of risk factors.' Fortunately, he didn't speak for all in the discipline.

Bibliography

Bagus, Philipp. 2011. Morgenstern's forgotten contribution: A stab to the heart of modern economics. *American Journal of Economics and Sociology* 70[2]:540-62.

Burch, Thomas K. 1995. Estimating the Goodman, Keyfitz and Pullum kinship equations: an alternative procedure. *Mathematical Population Studies* 5[2]:161-70.

Coale, Ansley J., and Stephan, Frederick F. 1962. The case of the Indians and the teen-age widows. *Journal of the American Statistical Association* 57[298]:338-47.

Deming, W. Edwards. 1944. On errors in surveys. *American Sociological Review* 9[4]:359-69.

Huff, Darrell. 1954. *How to Lie With Statistics*. New York: W.W. Norton Co.

Madigan, Francis C. 1957. Are sex mortality differential biologically caused? *Milbank Memorial Fund Quarterly* 35[2]:202-23

Morsgenstern, Oskar. 1950, 1963 [2nd ed]. *On The Accuracy of Economic Observations.* Princeton, NJ: Princeton University Press.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Cambridge, MA: Harvard University Press.

Preston, Samuel H. 1970. An international comparison of excessive adult mortality. *Population Studies* 24[1]:5-20.

Truesdell, Leon E. 1938. Residual relationships and velocity of change as pitfalls in the field of statistical forecasting. *Journal of the American Statistical Association* 33[202]:373-79.

U.S. Bureau of the Census, *U.S. Census of Population: 1950.* Vol. IV. *Special Reports,* Part 2, Chapter D, Marital Status. U.S. Government Printing Office, Washington, D.C., 1953.

Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.

**** **** ****