FACTORS CONSIDERED BY SECONDARY STUDENTS WHEN JUDGING THE VALIDITY OF A GIVEN STATISTICAL GENERALIZATION

Hanan Innabi
UAEU, United Arab Emirates
hinnabi@uaeu.ac.ae

This study investigated the factors that 12th grade students in United Arab Emirates take into consideration when judging the validity of a given statistical generalization, in particular, in terms of the sample size and sample selection bias. The sample consisted of 360 students who had not studied sampling yet. Results show that a small percentage of the students take the sample size and selection bias into consideration properly. Many students based their judgment on their personal beliefs regardless of the properties of the selected sample. This study identified some pre- teaching misconceptions that students have with regard to 'sampling.' Such misconceptions are 'any sample represents the population,' and, 'any sample does not represent the population.'

INTRODUCTION

The relationship between sample and population depends on the concept that the sample as a part of the population can be examined in order to obtain a generalization true of the population, or what is called a *statistical generalization*.

As in all inductive inferences, we cannot establish that the statistical generalization is true with absolute certainty. Our concern, usually, is how likely is it that the conclusion is valid? The crucial feature that determines the strength of a statistical generalization is the representativeness of the sample. In another words, to what extent are the features of the population that concern us reflected accurately in features of the sample (Salmon, 2002). Usually, it is not easy to tell whether a sample is representative. However, two criteria are considered noteworthy: 1) the sample is large enough 2) the sample is varied enough. In some cases, a very small sample can support a strong generalization; in others, a very large sample is required. The real question is whether the sample is large enough to capture, or represent, the variety present in the population.

This study tries to provide some knowledge of how secondary students think when they judge the validity of a given statistical generalization. In particular, this study tries to answer the following question: What factors do secondary grade students in UAE who have not studied sampling methods take into consideration when judging the validity of a given statistical generalization?

Living in an uncertain world requires the ability to reason statistically. Many pervasive and persistent errors have been found in people's reasoning about uncertainty (Konold, 1994). Numerous studies have showed the weaknesses in people's reasoning in those social situations which demand statistical reasoning (Kuhn, 1991; Konold, 1989; Kahneman and Tversky, 1972, 1973, 1979; Nisbett and Ross, 1980). Personal perspective and individual narrow experiences lead individuals to be biased in their judgments (Nisbett, Krantz, and Jepson, 1993; Evans, 1989; Falk, 1989; Shaughnessy, 1992).

Many studies have shown that the effect of sample size on probability and variation is not a factor for people who are statistically naïve (Schrage, 1983; Kahneman and Tversky, 1973, 1979; Tversky and Kahneman, 1973; Well, Pollatsek, and Boyce, 1990; Innabi, 1999). The view that statistically naïve people ignore sample size has been modified by subsequent research. A number of studies have shown that subjects may take account of sample size if the form of the problem is modified or when the variable is manipulated in alternative tasks (Evans and Dusoir, 1977; Nisbett *et al*., 1993; Bar-Hillel, 1979; Cosmides and Toody, 1996).

Rubin, Bruce, and Tenny (1994) showed that students have inconsistent models of the relationship between samples and populations. Their answers in different problem settings fall in varying amounts under the influence of intuitions about sample representativeness or sample variability.

INSTRUMENT

A written problem was designed to explore the factors which influence students when judging the validity of statistical generalizations. It presented information about a sample and the relevant population. Then, a conclusion about the population was presented based on a sample statistic. Students were expected to rate the validity of the generalization as *valid conclusion* or *not valid* conclusion or *cannot judge*. Students were also requested to offer all the reasons justifying their selection.

The aim of this study was to investigate the factors which students take into consideration when they judge the validity of a given statistical generalization, in particular the sample size factor and the sample bias factor (In this research we deal with the term *bias* from the perception of the *selection bias*). Thus, the problem in the instrument had five alternate forms differing according to information about the given sample.

The content of this problem was about a student in the University of UAE who was interested in the number of visits that students make to the University library during term time. Therefore, he picked a sample of University students and asked them 'How many times do you visit the University library every week during term time?'. The average number of weekly visits in this sample was three, so he concluded that the average number of visits UAE University students make to the library during term is approximately three visits per week (consider that the information that was given by the sample's students is reliable). There were five versions to this problem. The only difference among these versions is the nature of the sample which was selected in each one. These samples are: 1) Large/biased sample: He selected a sample of 600 students randomly from those who were entering the library entrance. 2) Large/not biased sample: He selected a sample of 600 male and female students randomly from different scientific and humanities colleges in different years of study. 3) Small/biased sample: He selected a sample of 6 students randomly from those who were entering the library entrance. 4) Small/not biased sample: He selected a sample of 6 male and female students randomly from different scientific and humanities colleges in different years of study. 5) No information about the sample size or sample bias: He selected a sample of university students.

For versions one, three, and four, the expected answer to the closed question is *the conclusion is not valid*. For version two the expected answer is *the conclusion is valid*. For version five where there is no information about the selected sample the expected answer is *I can not judge the conclusion*. It was expected that students provide a proper statistical explanation for their judgments on the given conclusion such as *the conclusion is not valid because the sample is small' or 'the conclusion is valid because the sample is big enough and varied*. Notice that this problem tried to capture whether the students take sample size and bias into account without directing them to compare two different samples as the previous research usually did.

PROCEDURES

The instrument was given to 360 12th grade students in the science stream from 12 secondary schools (6 female schools and 6 male schools) chosen randomly in the city of Alain in the United Arab Emirates. The students had not studied sampling techniques at school. Six copies of each form were given in each of the schools.

The student's explanation (answer) was coded into four codes as follows: 1) the sample size code: If the student's answer took into consideration the size of the sample as a factor to judge the validity of the conclusion, she/he was given the code 1, otherwise 0. 2) the sample bias code: If the students' answer took into consideration the sample bias as a factor to judge the validity of the conclusion she/he was given the code 1, otherwise 0. 3) the other factor codes (two variables): Any other factors rather than the size and bias the students presented in their explanations, were written down and given a code from 1 to *n*. Since only a few of the students put more than two *other* reasons for their judgment, it was decided to put two variables for the *other factors*.

The data were entered to the computer using the *SPSS* packages. The actual case number became 338 after we deleted some improper cases.

To get a clearer view of the students' answers and to summarize the data, it was necessary to look at the answers as a whole within each of the five forms (large\bias, large\not bias, small\bias, small\not bias, no information about the sample) and according to each judgment (valid, not valid, cannot judge). Students' answers were coded again by one or more of the following categories:

1) Adequate statistical explanation: when the students explained the *expected* answer in the closed question by a *proper* statistical explanation using the sample size and bias factors, their explanation was considered as *adequate statistical explanation*. For example, in the case of large and biased sample, the answers that judged the conclusion as *not valid* because the sample is biased, was considered as *adequate explanation*.
2) Insufficient statistical explanation: Some explanations that students provided were not enough to support their judgment. For example in the case of the large and biased sample the response which mentioned that the conclusion is *valid* because the sample is large was considered an *insufficient answer*.
3) Personal belief explanation: This category contained the explanations that reflected students' personal opinion or experience about the subject of attending the libraries. For example the following explanations have been considered as *personal explanations*: the conclusion is *valid* because: 'the number of visits is a reasonable number,' 'there are negative attitudes towards the library,' 'youth do not worry about reading,' 'I go to the library' or the conclusion is *not valid* because: 'I have a sister in the university who has never entered the library,' 'I think that the average of the number of visits should be more than three,' or *I cannot judge* the conclusion because: 'I am not in the university,' 'I do not know the university students' need for a library,' 'I have no idea.'
4) Inadequate statistical explanation: this category contained the responses which contained an inadequate statistical explanation. Such explanations as, the conclusion is *valid* because: 'any sample of the students represents the whole students in the University/ any part represent the whole,' 'because he selected from the students who were entering the library,' 'his selection from those who go to the library makes the conclusion stronger' or the conclusion is *not valid* because: 'he should take all the students in the University,' 'any part does not represent the whole,' 'if he had selected any other 600 students he would have found another result.'

RESULTS AND DISCUSSION

Results of this study showed that students considered five factors when judging the validity of a given statistical generalization. These factors are: 1) sample size, 2) sample bias, 3) any sample represents the population, 4) any sample does not represent the population, 5) personal experiences and expectations.

The percentage of students who mentioned the sample size factor did not exceed 34%. Assuming that students wrote all of the reasons that led them to their judgements, we can say that two-thirds of the students could not see the sample size as a factor that effects the validity of the statistical generalizations. A similar statement can be made about the sample bias factor, only 11% of the students took the sample bias into consideration. i.e., correctly said that the sample is biased (or not biased). The above results support the idea that the sample characteristics are apparently not part of man's repertoire of intuitive ideas (Kahneman and Tversky, 1972; Evans, 1989).

It was noticed that more students were taking the factor (size) into consideration in the situations where this factor was small than the situations when this factor was large. The percentages of students who took the *size* and *bias* into consideration when no information about the sample was given were 28%, 12% respectively and when the sample was small/bias these percentages were 59% and 18% and when the sample was small/not biased these percentages were 48%, 8% Whilst in the situation where the sample was large enough/not biased, the percentages were 22% and 9% and when the sample was large/biased, the percentages were 14%, 7%. These results remind us of previous research which showed that the form of the provided problem (framing of problem instructions) affects naïve subjects in taking the sample size into their judgments and prediction.

The analysis of students' responses showed (see Table 1) that not all the students who took the sample size and bias into consideration provided an adequate answer. Just one fifth of the students used the two factors in an adequate way in their judgment. The results showed that around half of the students presented an inadequate statistical explanation and 6% of the students presented an insufficient explanation.

Table 1: Numbers of the students according to their explanations' categories

	Number of students	Adequate	Inadequate	Insufficient	Personal	Adequate+ Personal	Insufficient + Personal	Inadequate + Personal	No explanation
Large/ biased	Valid 17	0	1	7	6	0	2	0	1
	Not valid 39	4	19	0	11	1	1	1	2
	Cant judge 14	0	5	0	4	1	1	1	2
Small/ biased	Valid 8	0	5	0	2	0	0	1	0
	Not valid 54	33	11	0	4	2	0	3	1
	Cant judge 6	0	3	0	1	0	0	0	2
Large/ not biased	Valid 19	6	1	8	3	0	0	0	1
	Not valid 31	0	21	0	3	0	0	5	2
	Cant judge 19	0	14	0	4	0	0	0	1
Small/ not biased	Valid 6	0	2	1	1	0	1	0	1
	Not valid 41	21	13	0	3	2	0	0	2
	Cant judge 19	0	10	0	6	0	0	0	3
No information	Valid 9	0	4	0	3	0	0	0	2
	Not valid 39	0	27	0	9	0	0	1	2
	Cant judge 17	5	5	0	3	1	0	1	2
Total	338 100%	69 20%	141 42%	16 5%	63 19%	7 2%	5 1%	13 4%	24 7%

The analysis of students' explanations revealed the following misconceptions related to the relationship between sample and population:

1. Some students (3%) did not realise that the sample which was clearly biased did not represent the population. They looked at the bias in the *biased sample* as a factor that made the conclusion valid. This observation seems to agree with what Kahneman and Tversky mentioned about the representativeness heuristic.
2. Some students used information that was not sufficient to support their judgement. Some of them took only one factor of sample properties (size or bias) into consideration in supporting their judgement of the validity of the conclusion and forgot about the other.
3. Any sample represents the whole: some students (4%) considered that any sample irrespective of its size and bias was a good representative of the population i.e., could not see the differences (variability) among students and could not see that different sample selections made any difference in representing the population. It seems that those students

believed that any part can represent the whole without any understanding of the error between the *sample statistic* and the *population parameter* and how the sample properties can affect this error.

4. Any sample will not represent the whole: some students (17%) insisted that 'any sample will not represent the population.' This may indicate that those students could see that the *sample statistic* differs from the *population parameter*, but they could not see that this difference can be reduced through manipulation of sample properties. Students in this category were able to see the variation among the elements of the sample to the extent of leading them not to *believe* in sampling. In other words they have the belief that no *true* knowledge about any population can be obtained through sampling.

It appeared that students' personal expectations about the population studied affected their judgement so that if the given conclusion matched their expectation, the conclusion was judged as 'valid,' otherwise 'not valid.' The results showed that 26% of the students provided a personal explanation for their judgement, 19% of the students provided just personal explanations, whilst 7% provided adequate or inadequate or insufficient explanation in addition to the personal explanation.

Part of the effect of the personal beliefs on statistical judgement may can be explained by the *availability* heuristic suggested by Kahneman and Tversky. Some students in responding to the problem, may used easily availabe information reflecting their first idea or impression about this issue. More investegation is needed here.

As mentioned above, some students (7%) provided in addition to the *personal* explanations, other *statistical* explanations. It may be the case that those students provided one of these explanations to support the other one. In another words, these students provided *personal* explanations in order to support their *statistical* explanations or vice versa (i.e., provided *statistical* explanations to support their *personal* beliefs). More investigation is needed here to understand this point.

It is hoped that the above misconceptions and misunderstanding that students have before starting to formally learn sampling techniques at schools will be considered in textbooks and in teaching. One technique that could be used to change students' misconceptions is to confront students with examples and situations that lead them to see their misconceptions and motivate them to change them (Shaughnesy, 1993). For example when we teach students the condition that the sample is a subset of the population, it is not enough to give counter examples. We also have to give examples of samples contained in the population which do not represent the population. An application is the possibility of using problems similar to the problem that have been used in the instrument of this research as material in classes to focus students' attention on errors being made in formulating judgments and to clarify how beliefs and conceptions can affect decisions when in doubt. An effective way to make students realize how easy it is for response bias to affect the results of a survey would be to assign the students to do an experiment to demonstrate the extent of the response bias. For example, have them ask the same question about the library, but in two different locations (the library and the cafeteria, for example).

It is hoped that teaching *sampling* in UAE will be done in a way that aims to help students believe in sampling as a scientific technique that helps us make conclusions about a population, to understand that any conclusion based on sample results involves a degree of uncertainty and to realize that the validity of a statistical generalization is dependent on the properties of both the sample and the population (see for example, Watson, 2000; Phung, 2005; Lawson, Schwiers, Doellman, Grady, and Kelnhofer, 2003). We do not want students to be only able to define the terms *sample* and *population* and to calculate how many samples one can get of size n from a population with size N and to spell out the methods of sampling without real understanding and being unable to reason critically when for example reading a statistical generalization in a newspaper.

REFERENCES

Bar-Hillel, M. (1979). The role of sample size in sample evaluation. *Organisational Behavior and Human Performance,* 24, 245-257.

Cosmides, L. and Toody, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition,* 58, 1-73.

Evans, J. (1989). *Bias in human reasoning: causes and consequences*. London: Lawrence Erlbaum..

Evans, J. and Dusoir, A. (1977). Proportionality and sample size as factors in statistical judgments. *Acta Psychologica,* 41, 129-137.

Falk, R. (1989). Judgement of coincidences: Mine versus yours. *American Journal of Psychology,* 102, 477-493.

Innabi, H. (1999). *Students' Understanding of the Relationship Between Sample and Population.* Unpublished Ph.D. dissertation, University of Nottingham, England.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology,* 3, 430-454.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review,* 80, 237-251.

Kahneman, D. and Tversky, A. (1979). The simulation heuristic. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction,* 6, 59-98.

Konold, C. (1994). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.), *Radical Constructivism in Mathematics Education*. Dordrecht: Kluwer.

Kuhn, D. (1991). *The Skills of Argument*. Cambridge: Cambridge University Press.

Lawson, T. J., Schwiers, M., Doellman, M., Grady, G., and Kelnhofer, R. (2003). Enhancing students' ability to use statistical reasoning with everyday problems. *Teaching of Psychology,* 30(2), 107-110.

Nisbett, R., Krantz, D., and Jepson, C. (1993). The use of statistical heuristics in everyday inductive reasoning. In Nisbett, R. (Ed.), *Rules For Reasoning*. New Jersey: Lawrence Erlbaum.

Nisbett, R. and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgement*. Englewood: Prentice-Hall, Inc.

Phung, N. (2005). Public opinion polls, chicken soup and sample size. *Teaching Statistics,* 27(3), 89-92.

Rubin, A. D., Bruce, B. C., and Tenney, Y. (1991). Learning about sampling: Trouble of the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics*, Dunedin, New Zealand. Voorborg, The Netherlands: International Statistical Institute.

Salmon, M. (2002). *Introduction to Logic and Critical Thinking* (4^{th} edition). Belmont, CA: Wadsworth.

Schrage, G. (1983). Interpretation of stochastic models. In R. Scholz (Ed.), *Decision Making Under Uncertainty.* Amsterdam: North-Holland Publishing.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning,* (pp. 465-494). New York: Macmillan.

Shaughnessy, J. M. (1993). Probability and statistics. *The Mathematics Teacher,* 86(3), 244-248.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology,* 5, 207-232.

Watson, J. M. (2000). Statistics in context. *Mathematics Teacher,* 93(1), 54-58.

Well, A., Pollatsek, A., and Boyce, S. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behavior and Human Decision Processes,* 47, 289-312.