STUDENTS' JUDGMENT OF THE VALIDITY OF SOCIETAL STATISTICAL GENERALIZATION

Hanan Innabi, NCHRD. Jordan

ABSTRACT

One of the main ideas in Statistics is the relationship between the sample statistic and the population parameter and how this relationship varies according to the sample properties such as size and variety. Students' understanding of this relationship can help them to be critical thinkers with regard to the daily issues that are related to statistical reasoning in addition to helping them to learn statistics in a meaningful way. This study investigated the factors that 11th grade students take into consideration when judging the validity of a given statistical generalization, in particular the sample size and variety. This study also investigated whether judgements seem to be influenced by the students' social beliefs and experiences. The results showed that most of the students do not take the sample size and variety into consideration when judging the validity of a given generalization. A relatively high percentage of students based their judgement on the personal beliefs regardless of the properties of the selected sample. This study identified some pre teaching misconceptions, that students have with regard to 'sampling'.

The Problem

A statistical generalization is a statement made about the population from the knowledge we shall obtain from the sample. As in all inductive inferences, we cannot establish that the statistical generalization is true with absolute certainty. Our concern usually is how likely it is that the conclusion is valid. The crucial feature that determines the strength of a statistical generalization is the representativeness of the sample. 'Sample is representative' means that the features of the population that concern us are reflected accurately in features of the sample (Salmon, 1984). Usually, it is not easy to tell whether a sample is representative, but two criteria are important to note: 1. The sample is large enough. 2. The sample is varied enough.

The use of statistical generalizations appears not only in scientific research, where researchers generalize the results they find in the sample to the population, but also in daily social issues, where many conclusions we form or judge are based on a sample of behaviours or observations (Nisbett, & Ross, 1980). Statistical generalizations is an aspect of critical thinking (Ennis, 1985), which is now the goal of most educational policies in the world.

Personal perspective and personal narrow experiences lead individuals to be biased in their judgements (Shaughnessy, 1992). The belief bias effect arises when subjects evaluate the validity of an argument on the basis of whether or not its conclusions conform to their prior beliefs, rather than on the basis of whether it is logically entailed by the premises. Truth status can be assessed on the basis of what the subject can retrieve from his experience or, as in Tversky and Kahneman terms, on the basis of the availability of relevant information (Evans, 1989). Falk (1989) for example found that subjects thought their own coincidences were more surprising than coincidences that happened to others. Self-coincident events are more likely to be remembered or recalled than similar coincidences of others. Nisbett and Ross (1980) maintained that people fail to apply necessary statistical principles to a very wide range of social judgements. They claimed that people often make overconfident judgements about others based on small and unreliable amounts of information; they are often insensitive to the possibility that their samples of information about people may be highly biased.

This study tries to provide some knowledge about how students think when they judge the validity of a given statistical generalization (that related to their personal life) before studying the

sampling techniques at schools. The benefit of this knowledge is related to the idea that taking into consideration student's interests, ideas, conceptions and misconceptions about the related content and process prior to teaching is an important base of meaningful learning.
In particular, this study tries to answer the following two questions:

1. Do 11th grade students who have not studied sampling techniques take the factors of sample size and sample bias (variety) into consideration when judging the validity of a given statistical generalization?
2. To what extent do the cultural beliefs and experiences of these students affect their judgement about the validity of a given statistical generalization?

In this research we deal with the term 'bias' from the perception of the 'selection bias': the extent that the sample is varied in order to capture the variety present in the population.

Theoretical Background

The law of large numbers says that if a random sample is large enough, the relative frequencies of outcomes in the sample have a very high probability of being a close approximation of those in the population (Pollatsek et al, 1984). According to Evans (1989), in making judgements, people have little appreciation of the law of large numbers. For most people, the similarity of a sample statistic to a population parameter does not depend on the size of the sample. The small samples as well as large samples have a high probability of looking like the population. Tversky and Kahneman (1971, 1972) explained this by the representativeness heuristic. They insisted that the idea that sampling variance decreases in proportion to sample size is apparently not part of man's repertoire of intuitions.
According to Landwehr (1989), most people would not recognise a statistically significant difference between proper samples, also people do not realise that a carefully drawn sample of a few hundred subjects can tell much about a very large population. Schrage, (1983) showed that the effect of sample size on probability and variation is not a factor for people who are statistically naïve.

Shaughnessy thinks that the media plays a role in the unwarranted confidence in small samples among people as he said 'In fact, television advertisements play off this misconception all the time, with phrases like two out of three doctors say'. (Shaughnessy, 1992, p 478)

The picture 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, 1979,1982; Olson ,1976; Cosmides and Toody, 1996) . Bar-Hillel (1982), for example found that over 80% of her 72 subjects correctly chose the larger sample and 4% chose the smaller sample in the following problem:

> Two pollsters are conducting a survey to estimate the proportion of voters who intend to vote YES on a certain referendum. Firm A is surveying a sample of 400 individuals. Firm B is surveying a sample of 1000 individuals. Whose estimate would you be more confident in accepting?
> Firm A's____ Firm B's____ About the same____
> (Bar-Hillel, 1982, p. 79)

Well, Pollatsek, and Boyce (1990) conducted a series of experiments. Different versions of the problems were presented to undergraduate students who had not previously taken a college statistics course. In one version, called the "accuracy" version, students were just asked which would be closer to the population average, the average in a large sample or the average in a small sample. In the "tail" version of this problem, students were asked to estimate how likely it was that the sample average was a certain distance from the population average. Students tended to do well on the accuracy version, but poorly on the tail version. Their results suggested that naive subjects' appreciation for the law of large numbers often does not result from in-depth understanding of the relation between sample size and variability.

An early indication of people's insensitivity to considerations of randomness versus bias in sample selection came from a study by Nisbett and Borgida (1975). Two groups had been asked to make predictions of the behaviour of participants in Psychology experiments. One group was given no sampling information and the other group was assured of random selection. Predictions of the two groups were, despite sampling considerations, nearly the same.

Rubin et al (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 match the population or sample does not match the population.

We noticed that the research studied the effect of sample size on people's judgement can be categorized either with studies which used 'complex (difficult) problems' (such as in the Kahneman and Tversky studies and Well et al, 1990 in the tail version) or studies which used 'simple problems' (such as Evans and Dusoir, 1977., Bar-Hillel, 1982., Well et al, 1990 in the accuracy version). The format of the problems in the 'simple problems' usually contained two samples, one is bigger than the other sample and the subjects had been asked to compare the two given samples.

In the present research the problem that have been used can be classified with the 'simple problems' but the approach that had been followed was different. We tried to capture whether the students take 'sample size' into account but we did not try to direct students to compare two different sizes. We did not ask the students to look to sample size, we just gave one sample size and described the way this sample had been selected (or sometimes we did not mention anything about the sample) and put a conclusion depending on this sample and asked the students to judge the validity of this conclusion.

Method

Paper and pencil test was given to 600 students from 20 secondary schools in Amman. The students had not yet studied the sampling techniques content of the curriculum at school. The problem of the instrument was used to explore the factors that students take into consideration when judging the validity of the given statistical generalization. This problem contained information about a selected sample and a generalization (based on the sample result) about a population. The students were asked to judge the validity of the given conclusion as "valid conclusion" or "not valid conclusion" or " cannot judge". They were also asked to explain in detail all the reasons for their judgement.

To explore the influence of beliefs and stereotyped thinking on the students' ability to judge validity of statistical generalization, the problem in the instrument was concerned with students' cultural beliefs and experiences. The content of this problem was about a mother of an 11th grade student who noticed that her son is very worried about the new 'Tawgehee' examination. She wanted to know if most of the 11th grade students in the Kingdom of Jordan are worried. So she selected a sample from 11th grade students and asked them about their worries. She found that 80% of the selected students were worried. The mother concluded that 80% of 11th grade students in Jordan are worried about the new examination system.

This problem is labelled as a 'belief' problem because the new examination system is an issue in Jordan these days. There is a lot of discussion by parents, teachers, students, and educators about it. The sample of this study was from 11th grade students for whom this new examination takes place next year. For those students this issue is related to their educational life, and each student has his/her attitude and opinion about it based on their expectations.

Since 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 variety factor, the problem of the instrument had five alternate forms differing according to information about the given sample. The five alternate forms presented information about:

(Uno) Large/biased sample.(The mother co-operated with the Ministry of Education to ask 500 students from a school for creative students).

(Due) Large/not biased sample.(The mother co-operated with the Ministry of Education to ask 500 students from different schools in different areas in the Kingdom).

(Tre) Small/biased sample.(The mother asked five students from her son's class)

(Quattro) Small/not biased sample.(The mother asked five students each from a different school in the Kingdom).

(Cinque) No information about the sample is given. (The mother selected students from the 11th grade).

Two stages of coding proccess were carried out; in the first stage each statement in the student's written explanations was coded. By the end of this stage and as a result of it, five categories of students' explanations were generated which led to the second stage of the coding process.

Results

Students' explanations of their judgements about the validity of the given generalization can be described as follows:

Sample size: The explanations that took the 'sample size' into consideration regardless of the 'sample bias' were determined. Table (1) shows the percentage of students who took the sample size into consideration according to the situation of the given sample in the problem (i.e. whether it was a small sample or large sample or no information about the sample was given). The pattern in Table (1) shows that the students did better (in taking the sample size into consideration) when information about the sample was given compared to when there was no information about the sample. Also, it can be noticed that when information was given about the sample, the performance of students differed according to the sample situation (if it was small or large). When the sample was small, students did better in taking the size factor into consideration. In general the percentage of students who mentioned the sample size in their explanations about their judgement was 25%.

Table (1)

Percentage of students who took the sample size into consideration according to the given sample

No information about the sample	Large sample	Small sample	Total
6	17	43	25

Sample bias: The explanations that took the 'sample bias' into consideration regardless of the 'sample size' were also determined. The general percentage of students who took the sample bias into consideration was 26%. Table (2) shows the percentage of students who took the sample bias into consideration according to the situation of the sample that was given to the students in the problem (i.e. biased sample, not biased sample, and no information about the sample).

Table (2)
Percentage of students who took the sample bias into consideration according to the given sample

No information about the sample	Not biased sample	Biased sample	Total
8	21	41	26

Its clear that students did better in taking the bias factor into consideration when information about the sample was given compared to when there was no information about the sample.

Bias Factor as a Positive Factor :
We meant by 'the student took the bias factor into consideration' that the student mentioned in their written explanation that the sample was biased (when it was biased) or was not biased (when it was not biased).

When the students' written explanations about their judgements were analyzed, we found that some students could just not take the bias factor into consideration, but also considered the biased sample as a 'good' sample to represent the population. These students considered the bias factor as a positive factor providing a 'valid' conclusion. The percentage of students who considered the bias factor (The sample from the high achievement level to represent the students in Jordan) as a positive factor was 4%.

Adequate explanations: The explanations that took the size and bias factors together into consideration in a proper way were catergorized as adequate explanation. The percentage of students who provided an adequate explanation was 28%.

Insufficient statistical explanations: The student's answer was categorized into this category when it contained a correct statistical statement, but not enough to support the judgement. For example, in the case of large and biased sample some students judged the conclusion as 'valid' and explained their judgement by saying that the sample was large. Also in the case of small and not biased samples, some students judged the conclusion as 'valid' because the sample was not biased. The percentage of students who provided insufficient explanations was 9%.

Inadequate statistical explanations: The student's answer was categorized into this category when it provided a statistical explanation that was not suitable or correct. Such explanations as, 'the conclusion is valid (in small and biased sample) because the mother did not depend just on her son's opinion she depended on more than one opinion' or 'the conclusion is valid because any part represents the whole' or 'the conclusion is not valid because she should take all the 11^{th} grade students in Jordan'. The inadequate explanations that were found in the written explanations were divided into the two following categorise: (1) Any sample will represent the population (7%). (2)Any sample will not represent the population (17%).

Personal explanations: Many explanations of students' judgements have reflected their personal beliefs in a direct way. Most of these explanations reflected a fear of the examination. For example, some students said that the conclusion is valid because they think that most students in Jordan are worried or because they themselves are worried. Also, some students said that the conclusion is not valid because they think that the percentage of the worried students is much higher or much lower than the percentage in the given conclusion (which is 80%). Such explanations have been considered as 'personal explanations'.

Relating to the personal beliefs explanations, the analysis of students' written explanations was used to divide the students into the four following groups:

1. Students who just provided personal explanations to support their judgement (33%).
2. Students who provided 'adequate statistical' explanations in addition to personal explanations to support their judgement (multi-explanations) (5%).
3. Students who provided 'inadequate statistical' explanations in addition to personal explanations to support their judgement (multi-explanations) (5%).
4. Students who provided 'insufficient statistical' explanations in addition to personal explanations to support their judgement (multi-explanations) (3%).

It can be noticed that most of the students who provided personal explanations provided them as separate explanations and not as multi-explanation with another sort of explanation. We may say that a relatively high percentage of students (one third), just depended on their personal beliefs to judge the validity of the given conclusion.

One of the interesting results found in this research was when some students (18 out of 120) explained their judgement 'valid' or 'not valid' in the situation where no information was given to them about the sample size and sample variety. Those students used the factors size and bias (even though there was no information about them) to explain their judgement. For example, some students said that the conclusion (which based on an unknown sample) is valid because 'the sample has a proper size' or 'the sample is not biased'.

Conclusions

Assuming that students wrote all of the reasons that led them to their judgements, we can say that around three quarters of the students could not see the sample 'size' as a factor that effected the validity of the statistical generalizations. A similar statement can be made about the sample 'bias' factor, the percentage was 26% of the students who took the sample 'bias' into consideration.

The above results support the idea that the sample characteristics are apparently not part of man's repertoire of intuitive ideas (Tversky and Kahneman, 1972,1974, and Evans, 1989).

Considering the results found in the situation where no information about the sample was provided -i.e. when the format of the question was 'a sample has been selected'. We found that the percentages of students who were aware of the sample size or sample variety were less than 10%, whilst more students took the factors 'sample size' and 'sample variety' into consideration when the information about the size of the selected sample or about the selection method were given in the problem. One can conclude based on the above that when some information about the sample was given, more students tend to take into consideration either the sample size factor or the sample bias. Maybe we can indicate from this result that students' understanding of the size and bias concepts is an inert understanding which need activation; if we activate students' minds by giving them information about sample size and sample bias they start to use these concepts.

On the other hand, more students took the sample size or bias into consideration when it was 'negative' -i.e. small or biased than when the factor was positive -i.e. large enough or not biased. In the problem the percentages of students who took the 'size' and 'variety' into consideration when no information about the sample was given were 6%, 8% respectively and when the sample was 'small' and 'biased' these percentages were 43% and 41%. Whilst in the situation where the sample was 'large enough' and 'not biased', the percentages were 17% and 21%.
Notice that even when 'negative' information about the sample (i.e. 'small' or 'biased' sample), the percentage of students who took these 'negative' factors into consideration did not exceed 43%.

With regard to which factor (size or bias), students took more into consideration when they wanted to judge the validity of a given statistical generalization, results showed that there was no difference between the two factors.

In this research an analysis was done to investigate how students take both factors; sample size and sample variety together into consideration. The results of this analysis showed that, less than one third of the students took the two factors in an adequate way in their judgement. This may lead us to conclude that most of the students could not take the two factors; sample size and bias together into consideration in an adequate way.

The analysis of students' explanations disclosed the following misconceptions about the relationship between the sample and the population:

1. Some students (4%) 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. Those students who considered the conclusion as valid based their judgement on the fact that the selected sample is the 'best' sample to represent the population. This observation seems to agree with what Kahneman and Tversky mentioned about the representativeness heuristic.
2. Some students (9%) 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. For example when the sample was '500 high achievement students, many students judged the conclusion as 'valid' because the sample size was large enough without realizing the 'bias' factor in the sample selected.
3. Some students (7%) considered that any sample irrespective of its size and bias was a good representative of the population i.e. 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. Many students insisted that 'any sample will not represent the population'. The percentage of students who said that was 17% . 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.

We can relate the above two points with what Rubin, Bruce, and Tenny (1990) called sampling representativeness (sample gives everything) and sampling variability (sample does not give anything). This research agrees with their results of the existence of the two patterns of intuition reasoning among students. However there was a difference between the findings of this research and their research. In this research, more students said that any sample will not represent the population (sample variability) than who said 'any sample will represent the population' (sampling representativeness). Whilst in Rubin, Bruce, and Tenny study there was no clear pattern of the difference between sampling variability and sampling representativeness intuition.

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 the percentage of the students who used their 'personal' beliefs to judge the validity of the given conclusion was (46%). We may indicate from this that students' educational life affected their judgement.

Part of the effect of the personal beliefs on statistical judgement can be explained by the 'availability' heuristic suggested by Tversky and Kahneman (1973). Students in their responses, used the information easily available to their minds of what they themselves or their friends think about this issue.

The results showed that some students provided in addition to the 'personal' explanations, other 'statistical' explanations (adequate or inadequate or insufficient explanations). It can indicate that those students provided one of these explanations to support the other one. In other 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).

In the situation where there was no information about the sample, we found explanations that used the sample size or variety to judge the validity of the given conclusion. Some students gave those sample properties where in fact there was no information about either the sample size or the method of selection. This may lead us to say that students used 'statistical language' i.e. used the sample size and variety (even if it is not given) to support their personal beliefs. In other words if the given conclusion is matched with what they believed or expected the 'given unknown sample' is 'large enough' and 'not biased' according to them. If the conclusion does not match what they expect, the 'given unknown sample' is 'small' and 'biased'.

Implementations

This research disclosed some misconceptions that students have before starting formally learning sampling techniques at schools. Among the misconceptions, identified are: any sample can represent the population; no sample regardless of its size and variety can represent the population; a conclusion is valid (or not valid) because personal experience supports it; the bigger the sample regardless of its variety the more valid the conclusion.

It is hoped that the misconceptions that this research disclosed before teaching statistics with regard to the students' understanding of the relationship between 'sample' and 'population' will be useful to improve the statistical curriculum and teaching methods.

One approach that can be followed to probe students' errors is provided by Shaughnesy (1993). Shaughnesy's work suggests that we should include examples of misuses and abuses of statistics in our classes on probability and statistics, and encourage our students to rebut them with correct analysis. He suggests using the problems that have been used in the research as tools to probe students' statistical reasoning errors

One application of this research regarding Shaughnessy's suggestion is the possibility of using problems like 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 judgements and to clarify how beliefs and conceptions can affect decisions under uncertainty.

It is hoped that teaching 'sampling' 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.

REFERENCES

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

Bar-Hillel, M. (1982). Studies of representativeness. In Kahneman, D., Slovic, P., Tversky, A (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge University press.

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

Ennis, R. (1985). A logical basis for measuring critical thinking. Educational Leadership. Vol. 43. No. 2.

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

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

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

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

Landwehr, J. (1989). A reaction to alternative conceptions of probability. In: Alternative conceptions of probability: implications for research, teaching, and curriculum. (Chair) Garfield, J. Eleventh annual meeting of the north American chapter of the International Group for the Psychology of Mathematics Education, New Jersey.

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

Nisbett, R., Ross, L. (1980). Human inference: Strategies and shortcomings of social Judgement. Prentice-Hall, Inc, Englewood.

Olson, C.(1976). Some apparent violations of the representativeness heuristic in human judgement. Journal of Experimental Psychology: Human Perception and Performance. Vol. 2. pp. 599-608.

Pollatatsek, A., Konold, C., Well, A., Lima, S. (1984). Beliefs underlying random sampling. Memory & Cognition. Vol. 12. No. 4. pp. 395-401.

Rubin, A., Bruce, B., Tenney, Y. (199). Learning about sampling: Trouble of the core of statistics.

Salmon, L. (1984). An introduction to logic and critical thinking. Harcourt Brace Jovanovich, New York.
Schrage, G. (1983). Interpretation of stochastic models. In: Decision making under uncertainty. Amsterdam: North-Holland.
Shaughnessy, J. (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). National Council of Teachers of Mathematics. Macmillan publishing company. New York.
Shaughnessy, J. (1993). Probability and Statistics. The Mathematics teacher. Vol. 86. No 3.
Tversky, A., Kahneman, D. (1971). Belief in the law of small numbers. Psychological Bulletin, Vol. 76, No 2, pp. 105-110.
Tversky, A., Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, Vol. 5, pp. 207-232.
Tversky, A., Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. In Kahneman, D., Slovic, P., Tversky, A (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge University press, (1982).
Well, A., Pollatsek, A., Boyce, S. (1990). Understanding the effects of sample size on the variability of the mean. Organizational behavior and human decision processes. Vol. 47. pp. 289-312.