EXPLAINING THE LAW OF LARGE NUMBERS

Sue Finch, School of Psychological Science, La Trobe University, Australia

The study examined 40 students' explanations of a statement referring to decreases in variance with sample size. Content analysis coded 22 explanation features. Four groups of subjects were defined. Five undergraduates were unsure of the relationship described in the statement. A second group reversed the statement; a diversity-based view of within sample variability was proposed to explain their responses. Those with the best understanding gave multiple meanings to the statement. The final group described variance as decreasing with sample size - partly explained by changes in a distribution's centre or extremes. A diversity of interpretations of the statement was thus identified; the implications for good teaching of the LLN are discussed.

Concepts of variability and uncertainty are fundamental to all users of statistics. Understanding variability means knowing that variations in data can arise from both sampling variability and deterministic causes. The Law of Large Numbers (LLN) describes sampling variability. Understanding the LLN is based on knowing how to construct a sampling distribution and how to quantify variability in samples and in sampling distributions. Importantly, the LLN specifies the relationship between sample size and the variability of sample statistics (the size-variance relation).

Many psychologists' reliance on the 'Law of Small Numbers' was reflected in a willingness to replicate with small samples, suggesting an insensitivity to this relationship. Further evidence was preference for a substantive 'explanation' of a difference between a large sample with a significant result and a small sample without one (Tversky and Kahneman, 1971). Students asked to construct sampling distributions for different sample sizes were insensitive to sampling variability, and in an applied problem comparing extremes of distributions based on a large versus small samples, around half indicated size did not matter (Kahneman and Tversky, 1972).

These two studies stimulated many more. Insensitivity to sample size can depend on problem features such as complexity (Evans and Dusoir, 1977), reference to the centre or extremes of a distribution, and type of judgment required (Bar-Hillel, 1979), construct reliability and validity, domain of application, and reference to statistical features (Nisbett, Krantz, Jepson and Kunda, 1983). Subject factors included quantitative skills (Reagan, 1989), statistical and disciplinary training (Fong and Nisbett, 1991; Lehman,

Lempert and Nisbett, 1988), and understanding about variability in the phenomenon under study (Nisbett et al., 1983).

Previous research has assessed statistical reasoning by response choices and explanations of responses. Knowing the LLN most often meant recognising that results from large samples are more stable; rarely has understanding of the size-variance relation been examined directly. One study of students found that asking about a distribution's centre encouraged thinking about the similarity of large samples to populations; reversals of the LLN in response to questions about the tails of a distribution were associated with a focus on the extremes (Well, Pollatsek and Boyce, 1990). Further, knowing about the difference between distributions of scores and means, and experience with sampling distributions did not ensure good understanding of the LLN. Indeed, Well et al. (1990) concluded: "people appear to understand that averages of larger samples are more likely to resemble the population average but do not understand the implications ... for the variability of the sample average" (p.310).

The present study aimed to explore understanding of the size-variance relation by asking undergraduate and postgraduate subjects to explain a statement referring to decreases in variance with sample size. Understanding relations has been described in other domains as a basis for deeper understanding (Chi and VanLehn, 1991).

METHOD

Twenty third year psychology students (UGs) and 20 psychology postgraduates (PGs) participated. The study was part of a larger project; relevant aspects are described.

Subjects read a description of a research study and rated statements about the methodology (4 point scale, true to false) including: 'Use of a large sample meant smaller variance in the results'. The statement was based on earlier studies of similar subjects' articulations of the sample size-variance relation. It is ambiguous; the LLN and standard error are not explicitly mentioned. But by explaining their ratings, subjects can make explicit their understanding of the terms 'variance' and 'results' as well as their judgment of the relationship involved. The statement will be referred to as the relation statement (RS).

Subjects discussed their ratings with an interviewer, explained the reasons for their rating, and verbalised the kinds of things they were thinking about in rating the RS. Subjects' explanations were tape recorded and transcribed verbatim.

Coding: Twenty-two explanation features were coded, described here in 6 groups. (1) Distribution mentioned: *sample, population, distribution*. (2) Measure: *variable* measured (scores), *characters* or types of people in general. (3) Statistic: *variance* (or sd), other measure of *spread, variability* (variation, deviation, colloquial term), *results* (unspecified), *error* (chance), an *estimate*. (4) Relationship between large samples and variability: *not related, increases* or *decreases*; also references to *relative changes*. (5) Other distributional features of large samples: a greater proportion *closer to the mean*, reduced *outliers*, greater *coverage* of types and a *normal curve*. (6) Gains in *stability, confidence* (knowledge), or *accuracy* with size.

Some subjects stated that RS could have a few meanings: 'variance won't change (with) how big or small your sample is, what will change is your understanding and surety of what that variance is .... larger sample gives you a clearer understanding.' For this subject, two explanations were coded: *no relation* of size to *variance* and size gives gains in *confidence*. Implicitly, the first explanation interprets the RS in terms of the sample variance. In the analysis, different meanings were coded as separate *explanations* with 22 binary variables corresponding to the 22 features above. A hierarchical, divisive clustering method was used for data reduction. Coding and analysis details are available from the author.

RESULTS AND DISCUSSION

Five UGs gave no explanation after encouragement to think carefully about the RS. Some were unsure 'which way it should go'. The other 35 subjects gave 59 explanations. First, 5 clusters of explanations are described; features not discussed in the descriptions were generally not mentioned by subjects giving those explanations.

Increasing cluster: All 6 explanations referred to (i) *increases* in *variance* or *variation*, and (ii) *characters* or *coverage* (features rarely referred to with *decreases*). No other explanations referred to *increases* in variance with sample size. These explanations appear to focus on the diversity of possible kinds of people within a sample, rather than referencing the RS to the size-variance relation. A large sample may seem to have the potential to include more different kinds of people and variability is seen to increase and, indeed, the sample range does increase with size.

Null cluster: Fourteen explanations included all explanations referring to there being *no relationship* between sample size and some other feature, primarily (i) *variance*,

(ii) *variance* in the *variable* studied, or (iii) *variance*/*variation* in the *population*. This is the only cluster with consistent references to the *variable* studied or the *population*. Here two types of explanations can be described: one simply stating that variance is not related to size, and an elaboration based on the *variable* or *population*. The elaborations suggests subjects recognise (i) population variance as a fixed quantity, and (ii) the RS is not about the distribution of the measured variable.

Variance cluster: Six of the 7 explanations refer to larger samples as improving *confidence*, *accuracy* or *stability* of the *variance*. These explanations recognise large sample results as more reliable but it is not clear if variance means sample variance (as an estimator of population variance) or standard error.

Distributions cluster: The 9 explanations are relatively less consistent but have some common characteristics. With large samples there is either more of the distribution *close to the centre* or a reduced role for *outliers*, in some cases with respect to *characters* or to the *sample*. None of the explanations describes relations between sample size and statistics. Rather, a distribution and/or aspects of shape are the focus. Mostly the explanations are vague. Concern with the *sample* per se and *characteristics* rather than sample statistics suggests the RS statement was interpreted with respect to the sample rather than the LLN.

Decreasing cluster: The 23 explanations refer to *decreases* in *variance*, *variation* or another measure of *spread*. Fifteen explanations elaborate that decreases related to *estimates*, *error*, distributional shape (more *close to mean*, *a normal distribution*) or that *changes were relative*. The elaborations suggest 3 emergent understandings: that some kind of error variance reduces with size; that asymptotic distribution shape relates to changes in variance; and, that the RS applies to estimates. An ordering can be suggested for all 23 explanations: (i) 8 restatements suggest little understanding of the LLN; (ii) reference to error is ambiguous but may reflect an intuitive notion of the standard error and, similarly, reference to shape begins to recognise relationships between distributional shape and other parameters; (iii) explanations referencing estimates most closely related to the LLN.

Next is an account of subjects' responses referring to explanation clusters and to UG/PG group. Twenty-two subjects gave explanations unique to a cluster, 11 gave responses from 2 clusters, and 2 gave explanations from 3 clusters. Most Increasing cluster explanations occurred without explanations from other clusters. In contrast, the

Null cluster combined with all other clusters and only once occurred alone. The Variance cluster never occurred alone, combining with Null and Distribution clusters.

Considering the patterns of co-occurrence of explanation types, what can be said about understanding the LLN? There were few explanations precisely resembling the LLN, but four kinds of interpretations of the RS can be characterised.

Group 1: Five UGs gave no explanation. With little intuition of the relationship described, decreases in variance seem as plausible as increases.

Group 2: Five subjects (3 PG) reversed the RS. Their increasing relation explanations include a diversity-view of within sample variance that easily leads to the inference that larger samples have greater variance. Generally, the reversals occurred without other kinds of explanation. Unlike some other subjects, this group had a single interpretation of the RS; the RS prompted no reference to the LLN.

Group 3: Twelve subjects (10 PG) mentioned null relations; 11 provided two or more explanations. Most additional explanations were from the Variance or Decreasing cluster. Group 3 shows signs of developing a deeper understanding, being able to give multiple interpretations of the RS: some specified that the score or population variance was not relevant to the RS but indicated that variance could decrease in some other sense or that reliability gains came from large samples.

Group 4: All but one of the 18 subjects (7 PG) gave explanations from the Distributions or Decreasing cluster; most (13) gave one explanation. Many of the explanations were restatements of the RS or referred to distributional shape (close to the mean, outliers, or normal distribution). Group 4's understanding appears to fall between that of Groups 2 and 3. It considers an appropriate relational form but explains it in a relatively unelaborated way.

These results can guide better teaching of the LLN. *Variability* clearly has many meanings and referents that need to be clarified in good teaching. Distribution and measure referents are important for precisely describing the LLN, but distributional referents were not often used by subjects. Reversals of the relation described by the RS (Group 2) were linked a diversity-based view of sample variance that neglects frequency (cf. Well et al., 1990). These students need to learn to distinguish between a heuristic idea of variability as spread and frequency based measures of variance. Group 3's multiple interpretations provide a basis for good understanding of the LLN; they have begun to distinguish distributional referents and their notions of stability of statistics can be built

on to develop an understanding of unbiased estimation. Group 4's thinking about changes in a distribution in terms of the observations about the mean or outliers provides a basis for inferring the variance will decrease with size. Implicitly they take some account of frequency - more in the centre, less in the tails in large samples; even with an intuitive notion of variability appropriate inferences can be made. These are intuitions to be made more precise. The uncertainty felt by 5 UGs is unsurprising in light of the different interpretations. The variability in the language of variability is not to be underestimated.

REFERENCES

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

Chi, M.and VanLehn, K. (1991). The content of physics self explanations, *Journal of the Learning Sciences*, *1*, 69-105.

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

Fong, G. and Nisbett, R. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, *120*, 34-45.

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

Lehman, D., Lempert, R. and Nisbett, R. (1988). The effects of graduate training on reasoning. *American Psychologist*, *43*, 431-442.

Nisbett, R., Krantz, D., Jepson, C., and Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339-363.

Reagan, R. (1989). Variations on a seminal demonstration of people's insensitivity to sample size. *Organisational Behaviour and Human Decision Processes*, *43*, 52-57.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105-110.

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