Mathematics teachers make statistical inference based on the distribution of sampled values

Kai-Lin Yang

National Taiwan Normal University, Taiwan

The study presented in this paper illustrated part-characteristics underlying mathematics teachers' alternative understanding or misunderstanding of confidence interval (CI) - related concepts. Firstly we developed assessment instruments to explore teachers' understanding of CI-related concepts. We found that mathematics teachers knew mean, variance and some properties of a random variable, what a CI for a proportion measures, and the relationships between sample sizes, confidence level and the width of a CI for a proportion. In addition, they were able to calculate a CI for a proportion. However, they did not integrate their understanding of a random variable appropriately when estimating a parameter for a random variable, and did not transfer their understanding of CIs for a proportion to CIs for a mean. One critical characteristic underlying their thoughts was that they made statistical inference mainly based on the distribution of sampled values.

Keywords: confidence interval, mathematics teacher, statistics

Introduction

Due to the relevance and importance of statistical inference, many countries include a basic study of statistical inference in their curricula of high schools (around 16- to 18-year-olds). Taiwan senior-high-school mathematics curriculum has included many statistical topics, e.g. median and quartiles, mean and standard deviation, populations and samples, sampling and statistical inference (Ministry of Education, Taiwan 2008). Under this statistical curriculum which emphasizes less on simulations and data analysis, we found most of the mathematics teachers hold negative attitudes towards statistical learning and teaching (Yang 2011).

Teachers may share the same misconceptions as the students. For example, teachers encounter conceptual obstacles as they attempt to conduct, or make sense of, hypothesis testing (Liu and Thompson 2009). In a review paper, Sotos et al. (2007) find few researchers have studied misconceptions concerning CIs which are one fundamental concept of statistical inference. The purpose of this study specifically focuses on misunderstanding of CI-related concepts held by mathematics teachers.

Meaning of confidence intervals

Gardner and Altman (1986) summarized that CIs present a range of values, on the basis of the sample data, in which the population value for such a difference may lie. A CI produces a move from a single value estimate, such as the sample mean, to a range of values that are considered to be plausible for the population. The width of a CI based on a sample statistic depends partly on its standard error, which counts on both the standard deviation and the sample size. It relies on the degree of "confidence" that we want to associate with the resulting interval as well.

Suppose that in a study a random sample of 400 twelfth graders from ten senior high schools was selected to estimate the average national college entrance exams (NCEE) score of all twelfth graders in Taiwan, and the average for the sample was found to be 300 with a sample standard deviation of 60. A 95% CI for the average NCEE score of all twelfth graders in Taiwan is from 294 to 306. Put simply, this means that there is 95% confidence rather than chance that the indicated range includes the 'population' mean. In a statistical sense, the CI means that if a series of identical studies were carried out repeatedly on different samples from the same populations, and there is about 95% chance that the population mean would be included among these CIs for the population mean.

After learning CI-related concepts, mathematics teachers are supposed to have an understanding of what CIs and confidence levels do not say. Referring to Moore (1995), we cannot claim that we have 95% probability that a proportion lies within this CI because no randomness remains after we draw one particular sample and construct one particular level CI from it. The true proportion either is or is not between the CI. Thus, probability is replaced by confidence. We can claim that we have 95% confidence that a proportion lies within this CI which is calculated by a method that gives correct results in about 95% of a large number of repeated random samples of the same sample size.

Misconception of confidence intervals

Literature from research on statistical cognition and application suggests statistical inference concepts are commonly misunderstood by students and even misinterpreted by researchers. Cumming, Williams and Fidler (2004) found a deep misconception concerning the question "What is the probability that the next replication mean will fall within the original 95% CI?" An internet investigation in which researchers were asked to answer this question suggested that a majority of the researchers held a misconception that the original 1-α% CI will capture about 1-α% of replication means of the samples. This misconception is consistent with the law of small numbers intuition of understanding sampling variability (Sotos et al. 2007). After searching the studies related to misconceptions of the CI concept, Sotos et al. found that only one study concerning undergraduate psychology and ecology students' understanding of CIs related to the p-value concept, Fidler (2006). Students' misconceptions of CIs detected by Fidler include (1) CIs are a range of plausible values for the sample means; (2) CIs are a range of individual scores; (3) CIs are a range of individual scores within one standard deviation; (4) The width of a CI increases as the sample size increases; (5) The width of a CI is not affected by sample size; and (6) A 90% CI is wider than a 95% CI for the same data. The misconceptions (1), (2) and (3) are related to what a CI measure. The misconceptions (4), (5) and (6) are related to relationships between confidence level, width and sample size. The highest percentage of students, 73%, held misconception (6). Misconceptions (1) to (6) are less related to the interpretation of a confidence level.

Referred to Garfield and Ben-Zvi (2008), other misconceptions include (1) a confidence level refers to the chance that the CI includes the sample mean; (2) a confidence level refers to the chance that the population mean will be between the upper and lower limits of the CI; (3) a confidence level refers to the percentage of data included in the CI; (4) a wider CI means less confidence; (5) a narrow CI is always better regardless of confidence level. In sum, we have found that misconceptions of CIs are related to the covering meaning of CIs, factors influencing the width of a CI, and the meaning of confidence levels.

Method

Two confidence intervals are calculated for two samples from a given population. Assume the two samples have the same standard deviation and that the confidence level is fixed. Compared to the smaller sample, the confidence interval for the larger sample will be:

(a) Narrower (correct)

(b) Wider

(c) The same width

(d) It depends on the confidence level

Figure 1. One item for assessing understanding of CIs in Stone et al.'s study

Suppose there is a population of test scores on a large, standardized exam for which the mean and standard deviation are unknown. Two different random samples of 50 data values are taken from the population. One sample has a larger sample standard deviation (SD) than the other. Each of the samples is used to construct a 95% confidence interval. How do you think these two confidence intervals would compare?

(a) The two samples would produce identical values for the lower and upper bounds of the two confidence intervals.

(b) The confidence interval based on the sample with the larger standard deviation would be wider. (correct)

(c) The confidence interval based on the sample with the smaller standard deviation would be wider.

(d) The two confidence intervals would have the same width because they are both 95% intervals.

Figure 2. One item for assessing understanding of CIs in ARTIST

Although there are some instruments for measuring concepts related to statistical reasoning (e.g. Garfield 2003), statistical inference (e.g. delMas et al. 2007) and statistical understanding (e.g. Stone et al. 2009), few instruments are specifically focused on the CI concept. And 'knowing-that' knowledge is different from 'knowing-to' knowledge (see Mason and Spence 1999). An exemplary item for measuring knowing-that knowledge is shown in figure 1, and an exemplary item for measuring knowing-to knowledge (Stone et al. 2009) is shown in figure 2 (ARTIST 2011). However, the item in figure 2 requires to be improved because alternative (a) and (d) can be inferred by each other.

In order to investigate mathematics teachers' alternative understanding and misunderstanding of CIs, we firstly develop comparatively complete and reliable instruments for assessing CI-related concepts and use them as a tool for collecting teachers' thoughts. CI-related concepts are referred to binomial, normal, sampling distributions and CIs.

Twenty four in-service mathematics teachers were taking a graduate course in introduction to statistics. After CIs were taught by a statistician, they were asked to answer the test items individually and then to discuss their answers in class. We did not intervene in their discussion except asking them to explain their thoughts more clearly. After their discussion, the applied statistician would tell them correct answers and ask them how they think about the answers. Six teachers were conveniently selected to interview for further clarifying their understanding during the discussion. Three types of data were obtained: videotapes of the discussion, written materials, and audiotapes of the interviews.

The data analysis did not aim at examining the extent to which teachers hold misunderstanding but investigating mathematics teachers' thoughts of CI-related concepts. Thus, we decided to explore qualitative data without a pre-determined theoretical or descriptive framework (Yin 1994). The analysis was guided by questions concerning teachers' incorrect answers and their thinking about their incorrect answers.

Result and Discussion

In this paper, we revealed one underlying reason which resulted in mathematics teachers' misunderstanding of CIs. That is, some teachers made statistical inference mainly based on distributions of sampled values.

The first type of thoughts was the confusion between the distribution of samples and the distribution of a random variable. The item, in figure 3, asked teachers to find the maximum likelihood estimator of the population parameter, the proportion of red balls to all balls. Some teachers who were able to correctly understand three general properties of a random variable misunderstood the most likely estimate for the proportion as the mode of these sample proportions and answered (A). These general properties referred to the distribution of a random variable, the expected value of a random variable, and a sample of a random variable.

> The number of red and blue balls in one bag was one hundred. Jon randomly drew twenty balls in this bag, and recorded the number of red balls. After repeating this act ten times, the recorded numbers were 2, 6, 4, 6, 10, 6, 18, 4, 8, 6. Which number of red balls in this bag is most likely?
>
> (A)30 (B)35 (correct) (C)60 (D)70

Figure 3. An item related to binomial distributions.

The mathematics teachers did not integrate the properties into estimating a parameter of a random variable. We agreed that the mode of the sample proportions could be considered as an alternative method to estimate population proportion in some cases. However, we were also concerned whether teachers distinguished sample proportions from the population proportion, and then identified the observations as the estimates from ten samples instead of ten data which were representative of part of the population.

The second type of thoughts was the confusion between the distribution of a random variable and the distribution of a sample statistic. The item, in figure 4, asked teachers to identify a 95% level CI for a population mean, the mean height of female college students. Some teachers ignored the sample size and used the standard deviation to estimate the bounds of the 95% level CI. Rossman and Chance (2004) mentioned that this error may come from learners' forgetfulness of dividing the sample standard deviation by the sample size. On the contrary, we found they failed to transfer their understanding of CIs for a proportion to CIs for a mean, and mistook the standard deviation as the standard error, although being able to correctly calculate a confidence interval for a proportion and understand relationships between "standard errors and confidence level", "sampling errors and the length of CIs", and "sample sizes and sampling errors" (or CIs, confidence levels). This implied that they needed to integrate procedural and descriptive knowledge into discriminating between "standard deviations and standard errors".

In a university, twenty five female students were randomly selected. We found that their average height was 160 cm, and the standard deviation was 10 cm. Which statement is correct based on the data? (maybe more than one correct statement)
(A) (156 cm, 164 cm) is a 95% confidence interval for the average height of all the female students in this university. (correct)
(B) (140 cm, 180 cm) is a 95% confidence interval for the average height of all the female students in this university.
(C) About 95% of all the female students' heights are within (156 cm, 164 cm).
(D) If one randomly selects twenty five female students in this university, the probability that their average height will be within (156 cm, 164 cm) is 0.95.

Figure 4. An item related to interpretation of confidence intervals and levels

Reflective Remarks

A coherent understanding of statistical inference entails integrating ideas of sample representativeness and sampling variability to reason about population parameters (Rubin, Bruce and Tenney 1991). In contrast, mathematics teachers may tend to focus on one sample data or sampled values that just produce the particular outcome, and assess probabilities by the propensities of the particular sample or samples at hand. Accordingly, it is suggested that teacher educators could notice how to shift teachers' attention to the discrimination between the distribution of a random variable where a sample comes from and that of a sample statistic which constitutes the base of statistical inference by relating their focus on the distribution of sampled values at hand to the theoretical distribution of a sample statistic.

Acknowledgement

This study arose from research supported by the National Science Council of Taiwan (NSC 98-2511-S-003-009 –M).

References

Cumming, G., J. Williams, and F. Fidler. 2004. Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics* 3: 299–311.

delMas, R. C., J. Garfield, A. Ooms, and B. Chance. 2007. Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal* 6 (2): 28–58.

Fidler, F. 2006. Should psychology abandon p-values and teach CIs instead? Evidence-based reforms in statistics education. *Proceedings of the seventh international conference on teaching statistics*. http://www.stat.auckland.ac.nz/~iase/publications/17/5E4_FIDL.pdf (accessed October 1, 2011).

Gardner, M. J., and D. G. Altman. 1986. Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal* 292: 746–750.

Garfield, J. 2003. Assessing statistical reasoning. *Statistics Education Research Journal* 2 (1): 22–38.

Garfield, J., and D. Ben-Zvi. 2008. *Developing students' statistical reasoning: Connecting research and teaching*. New York: Springer.

Liu, Y., and P. W. Thompson. 2009. Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies* 4 (2): 126–138.

Mason, J., and M. Spence. 1999. Beyond mere knowledge of mathematics: The importance of knowing-to act in the moment. *Educational Studies in Mathematics* 38: 135–161.

Ministry of Education, Taiwan. 2008. *2010 senior high school mathematics curriculum guideline*. http://mathcenter.ck.tp.edu.tw/MCenter/Center/CourseOutline.aspx (accessed October 16, 2011).

Moore, D. S. 1995. *The Basic Practice of Statistics.* New York: W. H. Freeman.

Regents of the University of Minnesota. 2006. *Assessment Resource Tools for Improving Statistical Thinking* (ATRIST). https://app.gen.umn.edu/artist/resources.html. (accessed September 19, 2011).

Rossman, A. J., and B. Chance. 2004. Anticipating and addressing student misconceptions (paper presented at ARTIST Roundtable Conference, Appleton, WI). http://www.rossmanchance.com/artist/proceedings/rossman.pdf

Rubin, A., B. Bruce, and Y. Tenney. 1991. Learning about Sampling: Trouble at the Core of Statistics. *Proceedings of the Third International Conference on Teaching Statistics,* ed. D. Vere-Jones, 1: 314-319. ISI Publications in Statistical Education, Dunedin, New Zealand.

Sotos, A. E. C., S. Vanhoof, W. Van den Noortgate, and P. Onghena. 2007. Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review* 2: 98–113.

Stone, A, T. Reed-Rhoads, T. J. Murphy, and P. K. Imbrie. 2009. Use of item response theory to facilitate concept inventory development. *Proceedings of the Research in Engineering Education Symposium*. http://rees2009.pbworks.com/f/rees2009_submission_72.pdf (accessed October 1, 2011).

Yang, K. L. 2011. Exploring Taiwan Senior High School Teachers' Conceptions about Statistics Education (paper presented at The Third Asian Conference on Education, Osaka, Japan, 27–30, October, 2011).

Yin, R. K. 1994. *Case study research: Design and methods*. 2nd ed. Beverly Hills, California: Sage.