SOCY2200 Statistics **Instructor: Natasha Sarkisian**

Answers to Assignment 4. Confidence Intervals.

1. You are interested in estimating the mean number of TV shows watched by BC students during a typical one month period. A random sample of n=500 students fill out daily reports. The sample mean is found to be $\overline{X}$ = 27 with a sample standard deviation of s=4.1. Set the 95% confidence interval for the mean. Interpret the interval in words; that is, make a sentence about the estimate rather than simply reporting the two numbers that form the interval.

   1. Set level of confidence at CL = 95%
   2. Obtain corresponding z. Since we want to use CL=95%, we want the middle 95% of the area under the curve, which means we want a z that excludes .025 in each tail. Half of needed 95% will be in the positive z half of the curve shown in our table, so we must divide 95 by 2. 95/2 = 47.50. Enter the Table B1 with 47.50 and read the corresponding z to be z=1.96 (the closest entry to 47.50).
   3. Set the interval: $\overline{X} \pm z^*\sigma_{\overline{X}}$ = 27 ± 1.96 (4.1/sqrt(500)) = 27 ± 1.96*(4.1/22.36) = 27 ± 0.36, which means the interval extends from 26.64 to 27.36.
   4. Set the confidence interval: Probability(26.64 ≤ μ ≤ 27.36) = .95
   5. In words: Based on the results from a random sample of 500 students, we conclude that there is a 95% chance that the true mean number of TV shows watched by BC students during a typical month is between 26.64 and 27.36 shows.

2. Do the same problem (#1) again assuming the sample size is only n=20. Describe how the results differ and why.

   1. Set CL at .95.
   2. Obtain t. In Table B2, we look at the column in the two-tailed part of the table that corresponds to .05, since 1-.95=.05. We read down the column marked as .05 to the row with the degrees of freedom equal to 20-1 = 19. The entry is t=2.093.
   3. Calculate the confidence interval limits: $\overline{X} \pm t^*\sigma_{\overline{X}}$ = 27 ± 2.093*(4.1/sqrt(20)) = 27 ± 1.92 which means the interval extends from 25.08 to 28.92.
   4. Set the confidence interval: Probability(25.08≤ μ ≤ 28.92) = .95
   5. In words: Based on the results from this smaller random sample with n=20, we conclude that there is a 95% chance that the mean number of TV shows watched by BC students during a typical month is between 25.08 and 28.92 shows.
   6. Comparison: When the sample size is smaller, the confidence interval for the same level of confidence is wider – its width is only 0.72 when the sample size is 500 versus 3.84 when the sample size is 20. This is because our estimate of the mean is less precise when the sample is smaller, so we need to construct a wider interval to maintain the same level of confidence.

3. You are interested in estimating the mean number of parties attended by BC students in a month. A random sample of 16 students is selected and you find the sample mean to be $\overline{X}$ = 7.2 with a sample standard deviation of s=1.6. Set the 90% confidence interval for the mean. Interpret the interval in words.

   1. Set CL at .90.
   2. Obtain t. In Table B2, we look at the column in the two-tailed part of the table that corresponds to .10, since 1-.90=.10. We read down the column marked as .10 to the row with the degrees of freedom equal to 16-1 = 15. The entry is t=1.753.
   3. Calculate the confidence interval limits: $\overline{X} \pm t^*\sigma_{\overline{X}}$ = 9.2 ± 1.753*(1.6/sqrt(16)) = 7.2 ± 0.7 which means the interval extends from 6.5 to 7.9.
   4. Set the confidence interval: Probability(6.5≤ μ ≤ 7.9) = .90
   5. In words: Based on the results from this random sample of 16 students, we are 90% confident that the true mean number of parties attended by BC students in a month is between 6.5 and 7.9 parties.

4. Using the General Social Survey 2012 data (gss2012.dta), calculate the mean for the number of hours respondents watched TV per day (tvhours) as well as the 90% confidence interval for this mean. Write down your results here, both as a formal statement with the associated probability and in words (remember that since this is a survey based on a representative sample of U.S. population, we can make statements about that population). Don't forget to open a Stata log file (.log) and submit it with your assignment on Canvas.

   The sample mean is: __3.09____ ______hours________
   number *units*
   (units of measurement rather than units of observation; e.g., years are the units of measurement for age)

   Confidence interval: Probability(2.96 ≤ μ ≤ 3.22)=.90

   In words: Based on this national representative sample of 1298 individuals, we are 90% confident that the actual average number of hours per day that Americans watched TV in 2012 was between 2.96 and 3.22 hours per day.

Stata output:

```
. mean tvhours, level(90)

Mean estimation                    Number of obs    =    1298

--------------------------------------------------------------
             |       Mean   Std. Err.     [90% Conf. Interval]
-------------+------------------------------------------------
     tvhours |   3.088598   .0795248      2.957698    3.219498
--------------------------------------------------------------
```

5. A public opinion poll asked a random sample of n=1600 respondents how they think current Presidential policies will affect inflation. It found that 60% think they will increase inflation.
(a) What is the maximum margin of error for this poll (expressed as %)?
(b) What is the actual margin of error for the proportion thinking they will increase inflation (expressed as %)?
(c) Construct the 95% confidence interval for this proportion (expressed as %); make a probability statement and interpret in words.
(d) What additional information should we acquire about the poll to better understand the potential for bias in these poll results?

(a) 1/sqrt(n)=1/sqrt(1600)=1/40=.025
The maximum margin of error is 2.5%.
(b) 1.96*sqrt(.6*.4/1600)=0.024
The actual margin of error for this proportion is 2.4%.
(c) The confidence interval has two limits -- .6-.024 and .6+0.024, so
Probability(57.6%≤ p≤ 62.4%)=.95
Based on the data from a random sample of 1600 respondents, we are 95% confident that the true population proportion of people who think that the current Presidential policies will increase inflation is between 57.6% and 62.4%.
(d) To understand the potential for bias, we would need to know how the sample was selected, what level of nonresponse there was, whether any demographic weighting was attempted when calculating the results, and how exactly the question was worded.

6. The following graph describes the average number of authors for published journal articles that are using four different types of methods – quantitative, qualitative, triangulation, and other. Based on the error bars in the graph, what can we conclude (with 95% confidence) about the differences between the average number of authors in these four types of articles? In your response, make sure to compare each pair (six comparisons total) and state for each whether, on the basis of this figure, we can be 95% confident that those two types of articles are different in terms of the average number of authors.

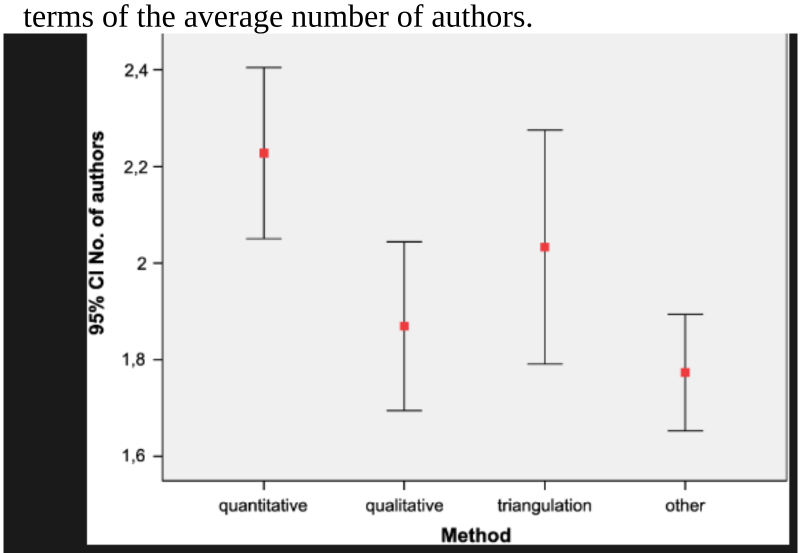

Figure 3. Error bar graph for research methods (95% confidence interval, n = 695).

Based on this figure, we are 95% confident that the average number of authors in articles using quantitative methodology is different from the average number of authors in articles using qualitative or "other" methods. To make other comparisons, we divide error bars in half to also see if standard errors overlap:

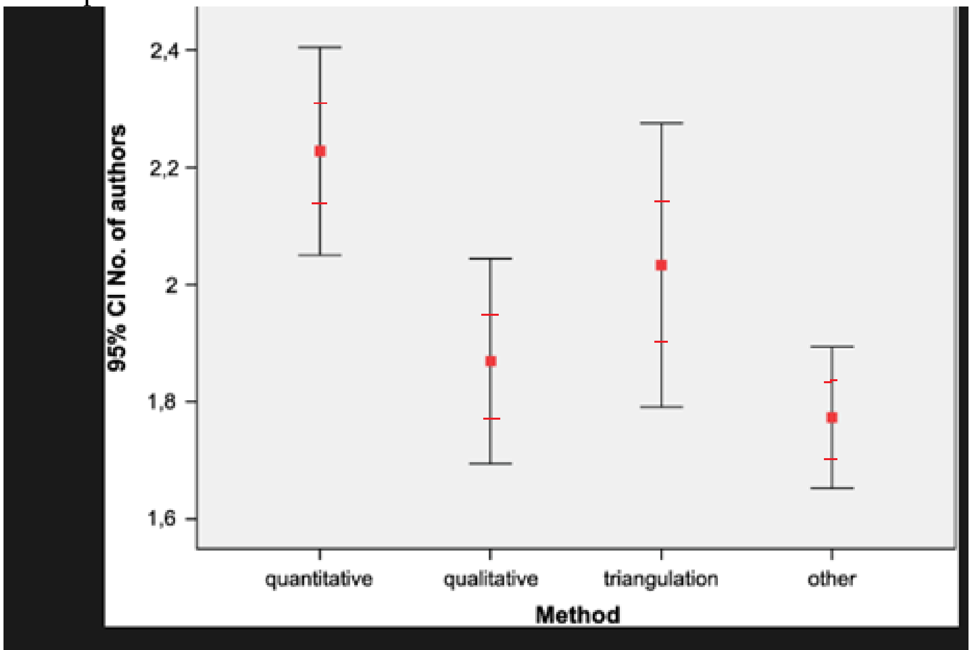

Figure 3. Error bar graph for research methods (95% confidence interval, n = 695).

We find no significant difference between the average number of authors in articles using quantitative methodology and those using triangulation. We also find no significant differences between qualitative and triangulation-based articles, or qualitative and "other method" articles. With regard to the comparison between triangulation and "other" articles, further testing to evaluate whether these groups are different is needed.

So the only two comparisons where we can be 95% certain about differences are quantitative vs qualitative and quantitative vs "other."