Assignment 3 – Solutions

1.   Suppose that a poll of 1,200 American adults is conducted and it is found that

46% of respondents support the Democratic stimulus package.

(a) What is the margin of error for this poll?

±2.82%

(b) What is the 95% confidence interval?

[43.18%, 48.82%]

(c) What is the 99% confidence interval?

[42.29%, 49.71%]

(d) What does this margin of error mean?

This margin of error is the half-width of the poll's 95% confidence interval.

(e) What does the 95% confidence interval mean?

There is a 95% probability that the interval contains the 'true' parameter, which in

this case is the true degree of support for the Democratic stimulus package.

(f) How large would the sample size have to be to guarantee a margin of error no larger

than 2 percent?

The sample size would have to contain around 2386 respondents if calculated

based on the proportion p=.46 from this problem, and 2401 respondents if

calculated from the maximum possible MoE when p=.5. Both are derived from

the equation MoE=1.96*Sqrt(p(1-p))/(Sqrt(N)).

(g) Do you agree with the following statement? "Being 99% certain is always better than being 95% certain, so I always prefer a 99% confidence interval to a 95% confidence interval"? Why or why not?

> No. The range of values within a 99% confidence interval are bigger than the range of values within a 95% confidence interval. In some cases, a 99% confidence interval is actually less informative than the 95% confidence interval, so we often settle for a little less in order to focus more tightly on the true parameter value.

2.　　In a survey of 10,000 American adults, it is found that the average respondent has owned 3.4 cars in his lifetime, with a standard deviation of 2.1.

(i) Does it make sense to talk about the margin of error of the average number of cars owned? Why or why not?

> It does not make sense to talk about the margin of error of the average number of cars owned because the margin of error is determined for an estimated proportion of the population, not a mean.

(ii) Form a 90% confidence interval for the average number of cars owned in the population.

> [3.36, 3.43]

(iii) Interpret the 90% confidence interval.

> There is a 90% probability that the 'true' average number of cars owned in the population lies within the interval [3.36, 3.43].

(iv) Suppose that we wanted the half width of the confidence interval to be 0.1. How large would the sample size have to be? Can this not be determined?

> We cannot determine how big the sample size would need to be before conducting the survey since we would have to know Xbar and s, but we only know these quantities after we have actually fielded the survey (at which point it is too late to select a different sample size).

3. Suppose that two public opinion polls are conducted, each with a sample size of 1,000. In the first poll, conducted last week, it is found that 64% of respondents approve of President Obama. In the second poll, conducted this week, it is found that 62% of respondents approve of President Obama. Did the population proportion of American adults who approve of President Obama decrease? Justify your result.

> The population proportion of American adults who approve of President Obama did not necessarily decrease or change. Using the QuickCalc spreadsheet for hypothesis testing between proportions of two populations we calculate a z-stat of .9265, which is less than 1.96. We accept the null hypothesis that the two proportions are equal.

4. Suppose that the population average SAT scores of Stony Brook students is 1220. Suppose that there are 6 students registered for POL 351 and these students have a sample mean SAT score of 1250 with a standard deviation of 90. Test the null hypothesis that the average PSC 203 student has an SAT score of 1220 against the alternative that the average is different from 1220, using the $\alpha = .05$ significance level. What does this result mean? Would your answer change if you used a 10% significance level?

a. Using the QuickCalc spreadsheet for hypothesis testing a population mean we get a resulting Z-statistic of 1.63. This has an absolute value less than the critical value of 1.96, and thus with a .05 significance level we accept the null hypothesis that the average PSC 203 student has an SAT score of 1220.

b. If we used a 10% significance level, the Z-statistic of 1.63 has an absolute value less than the critical value of 1.64, and thus with a .10 significance level we still accept the null hypothesis that the average PSC 203 student has an SAT score of 1220. The answer does not change.

Unfortunately, neither of these answers are reliable because n<30, and the confidence intervals for the hypothesis test are unreliable.

5. Consider the 2001 Israeli Election Study data set. Were Israeli Arabs less likely to vote for Benjamin Netanyahu in the 1999 election? Consider the variables B20 and JEWARAB in your analysis (use the 5% significance level).

The resulting Z-statistic from the hypothesis test for two population proportions is 22.66, which has an absolute value greater than the critical value of 1.96, and thus with a .05 significance level we can safely reject the null hypothesis that Israeli Arabs are equally likely to vote for Benjamin Netanyahu in the 1999 election as Jewish Israelis.

6. Consider the 2004 American National Election Study data file available on the course webpage. The variable "ID" has voters place themselves on a 1-7 ideology scale.

(i) According to this scale, are males in the population more conservative than females in

the population? Justify your answer with an appropriate hypothesis test (using the 5%

significance level). (5 points)

Using SPSS, we may obtain the mean ideology score, standard deviation and

sample size for both males and females. Select Analyze – Compare Means –

Means and select 'ID' as the dependent variable and 'Female' as the independent

variable, being sure to display standard deviation and sample size.

This reveals that males have a mean ideology score of 4.41 and females have a

mean ideology score of 4.11. It would appear that males are more conservative

than females. We can test this hypothesis by plugging the previously determined

values into the QuickCalc spreadsheet for the hypothesis test for the difference

between two population means (independent samples). The z-score is 2.72, which

is larger than the 5% significance level critical value of 1.96. Thus, we may reject

the null hypothesis that the means of males and females are equal and conclude

that males really are more conservative than females in the population.

(ii) According to this scale, are non-Blacks in the population more conservative than

Blacks in the population? Justify your answer with an appropriate hypothesis test (use the

1% significance level). (5 points)

Using SPSS in the manner described in part i, we may obtain the mean ideology

score, standard deviation and sample size for both Blacks and non-Blacks. We

find that Blacks have a mean ideology score of 4.16 and non-Blacks have a mean

ideology score of 4.27. It would appear that non-Blacks are more conservative

than Blacks.

We can test this hypothesis by plugging the previously determined values into the QuickCalc spreadsheet for the hypothesis test for the difference between two population means (independent samples). The z-score is .7213, which is smaller than the 1% significance level critical value of 2.57. Thus, we may not reject the null hypothesis that the means of Blacks and non-Blacks are equal and we cannot conclude that non-Blacks really are more conservative than Blacks in the population.