SEIS 631

Dan Ward

Assignment 7

**Q1) In the first paragraph, several key findings are reported. Do these percentages appear to be sample statistics (derived from the data sample) or population parameters?** Sample statistics

**Q2) The title of the report is “Global Index of Religiosity and Atheism”. To generalize the report’s findings to the global human population, what must we assume about the sampling method? Does that seem like a reasonable assumption?** We must assume that the sampling is random and that the 50,000 men and women from 57 countries sampled are representative of the total population. If the 57 countries were representative of the overall population and not biased, then yes, it is reasonable.

**Q3) What does each row of Table 6 correspond to? What does each row of atheism correspond to?** Each row in Table 6 corresponds to the overall results of a country. Each row in the atheism file corresponds to an individual’s response.

**Q4) Using the command below, create a new dataframe called us12 that contains only the rows in atheism associated with respondents to the 2012 survey from the United States. Next, calculate the proportion of atheist responses. Does it agree with the percentage in Table 6? If not, why?** Yes, it does, this is the data that the table was built on. 50/1002 = 4.99% the table reported 5%.

**Q5) Write out the conditions for inference to construct a 95% confidence interval for the proportion of atheists in the United States in 2012. Are you confident all conditions are met?**

Conditions: 1) Independent Observations 2) At least 10 observed successes and 10 observed failures. Assuming the observations were independent, we did have over 10 observed successes and over 10 observed successes so I am confident that all conditions are met.

**Q6) Based on the R output, what is the margin of error for the estimate of the proportion of the proportion of atheists in US in 2012?**

ME=(0.06518465-0.03805375)/2 = .013565

**Q7) Using the prop.test function, calculate confidence intervals for the proportion of atheists in 2012 in two other countries of your choice, and report the associated margins of error. Be sure to note whether the conditions for inference are met. It may be helpful to create new data sets for each of the two countries first, and then use these data sets in the prop.test function to construct the confidence intervals.**

**Japan:**

Independence is assumed met and 372 atheists and 840 religious so sample size is met. Both conditions have been met.

Confidence interval is (0.2816, 0.3335) with Margin of Error = (.3335-.2816)/2=.02595

**Spain:**

Independence is assumed met and 103 atheists and 1042 religious so sample size is met. Both conditions have been met.

Confidence interval is (0.0747, 0.1079) with Margin of Error = (.1079-.0747)/2=.0166

**Q8) Describe the relationship between p and me.** The ME relationship with p forms an arc with ME starting at 0 and approaching .03 as p approaches .5, at which point ME decreases in the same arc to 0 as p approaches 1.0.

**Q9) Describe the sampling distribution of sample proportions at n = 1040 and p = 0.1. Be sure to note the center, spread, and shape. Hint: Remember that R has functions such as mean to calculate summary statistics.** The mean of the values is 0.09969 with a median of 0.09904. The histogram is a very normal distribution with a spread of 0.07019 to 0.12981.

**Q10) Replicate the above simulation three more times but with modified sample sizes and proportions: for n = 400 and p = 0.1, n = 1040 and p = 0.02, and n = 400 and p = 0.02. Plot all four histograms together by running the par(mfrow = c(2,2)) command before creating the histograms. You may need to expand the plot window to accommodate the larger two-by-two plot. Describe the three new sampling distributions. Based on these limited plots, how does n appear to affect the distribution of the sample proportion? How does p affect the sampling distribution?**

For n = 400 and p = 0.1: The mean is .1002, with a median of .1000. With a spread of 0.055 to 0.16. Again with a normal distribution.

For n = 1040 and p = 0.02: The mean is 0.01994, with a median of .02019. With a spread of 0.00769 to 0.03846. With a normal distribution

For n = 400 and p = 0.02: The mean is .02005, with a median of .02000. With a spread of 0.0025 to 0.05. With a slightly right skewed normal distribution.

Based on these plots increasing n reduces the spread and creates a tighter normal distribution. Decreasing p significantly reduces the size of p-hat

**Q11) If you refer to Table 6, you’ll find that Australia has a sample proportion of 0.1 on a sample size of 1040, and that Ecuador has a sample proportion of 0.02 on 400 subjects. Let’s suppose for this exercise that these point estimates are actually the truth. Then given the shape of their respective sampling distributions, do you think it is sensible to proceed with inference and report margin of errors, as the reports does?** For Australia yes, it meets the conditions of independence and it meets the condition of at least 10 successes/Failures. For Ecuador it meets the condition of independence but it fails the condition requirement of 10 as with only 400 sampled a 2% proportion yields only 8.