Demos_HW7 part 2

Michael Demos

April 6, 2017

In August of 2012, news outlets ranging from the Washington Post to the Huffington Post ran a story about the rise of atheism in America. The source for the story was a poll that asked people, "Irrespective of whether you attend a place of worship or not, would you say you are a religious person, not a religious person or a convinced atheist?"The full press release for the poll, conducted by WIN-Gallup International, is found

http://www.wingia.com/web/files/richeditor/filemanager/Global_INDEX_of_Religiosity_and_Atheism_ PR__6.pdf

```
library(mosaic)
```

Preliminary Questions

1. 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*?

SOLUTION:

These percentages appear to be sample statistics. This is because they are based on the poll–which is clearly not a census; therefore, the actual population of the world was not included in this poll. In other words, the subjects that were polled only make up a sample of the global population.

2. 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?

SOLUTION:

To generalize the report's findings to the entire global population, we must assume that the sampling method used was a census. This is an unreasonable assumption because it is impossible to poll each and every person amongst the entire human population. Therefore, we cannot generalize the report's findings.

The data

Turn your attention to Table 6 (pages 15 and 16), which reports the sample size and response percentages for all 57 countries. While this is a useful format to summarize the data, we will base our analysis on the original data set of individual responses to the survey. Load this data set into R with the following command.

```
load(url("http://www.openintro.org/stat/data/atheism.RData"))
```

3. What does each row of Table 6 correspond to? What does each row of `atheism` correspond to?

SOLUTION:

Each row of Table 6 corresponds to one of the 57 polled countries and their respective response percentages for their samples. Each row of 'atheism' corresponds to a specific response of an individual respondant, categorized by country and year of their response.

To investigate the link between these two ways of organizing this data, take a look at the estimated proportion of atheists in the United States. Towards the bottom of Table 6, we see that this is 5%. We can check this number using the `atheism` data by running the commands below. Make sure you understand what each of the commands below does after running it.

```
us12 <- filter(atheism, nationality == "United States", year == "2012")
tally(~ response, data=us12, format = "proportion")
```

```
## response
##     atheist non-atheist
##   0.0499002   0.9500998
```

4. Using a similar series of commands, confirm the calculation of the proportion of atheist responses in our neighboring country of Canada. Does it agree with the percentage of 9% in Table 6?

SOLUTION:

```
ca12 <- filter(atheism, nationality == "Canada", year == "2012")
tally(~ response, data=ca12, format = "proportion")
```

```
## response
##      atheist non-atheist
##   0.08982036  0.91017964
```

Our calculation of the proportion of atheist responses in Canada, 0.08982, agrees with the percentage of 9% in Table 6, when rounded up.

Inference on proportions

The table 6 provides *statistics*, that is, calculations made from the sample of 51,927 people. What we'd like, though, is insight into the population *parameters*. You answer the question, "What proportion of people in your sample reported being atheists?" with a statistic; while the question "What proportion of people on earth would report being atheists" is answered with an estimate of the parameter.

A confidence interval

Here is how we'd compute a 95% confidence interval for the proportion of atheists in the United States in 2012.

```
confint(prop.test(~response, data=us12,conf.level=0.95))
```

```
##           p      lower      upper level
## 1 0.0499002 0.03761982 0.06574456  0.95
```

5. Interpret this confidence interval in the context of the problem.

SOLUTION:

If we took all possible samples of size n = 1002 and took all possible 95% confidence intervals, 95% of the intervals would contain the true percentage of Americans who claim to be atheists.

6. 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?

SOLUTION:

Conditions: random sample, large sample size. I am confident that the sample size is large. I am not confident that the sample is random–it is never stated that the samples are random in the report.

Although formal confidence intervals don't show up in the report, suggestions of inference appear at the bottom of page 7: "In general, the error margin for surveys of this kind is ± 3-5% at 95% confidence".

7. 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?

SOLUTION:

Based on the R output of p = .0499002, the margin of error is calculated to be plus or minus 1.348%.

8. Calculate a 90% confidence interval for the proportion of atheists in the United States in 2012. Does it make sense that this confidence interval would be wider or narrower than the 95% confidence interval we already calculated?

SOLUTION:

```
confint(prop.test(~response, data=us12,conf.level=0.90))
```

```
##           p      lower      upper level
## 1 0.0499002 0.03930392 0.06302676   0.9
```

It makes sense that this interval would be narrower than the 95% confidence interval because we sacrificed 5% of 'confidence' for a narrower (more accurate) interval.