Hw4 R

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?"

```
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: sample statistics

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: They need to assume that the sample method used is representative of the entire world's population.With 51927 observations we can say that this is less than 10% of the world's population. Comparing countries and religiosity there will be a minimum sample size requirements for interference for smaller countries with religious minorities.

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: Table 6 corresponds to the survey results of each country. Each row of atheism corresponds each country's percentage of convinced atheists.

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:

```
canada12 <- filter(atheism, nationality == "Canada", year == "2012")
tally(~ response, data=canada12, format = "proportion")
```

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

The percentages match from with the table 6 data on Canada even though 8.98% is rounded off to 9%

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: The range of intervals that the atheists fall into is defined by these values. This means that we are 95% confident that the sample intervals are accurate and will align with the actual population mean.

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: The individuals selected for the survey need to be from a random sample. Here, since the sample is less than 10% of the total population, this condition is met. Observations need to be independent which we are going to assume is met.The number of observations is 50 which is greater than 10 (required amount). We are also going to assume that the observations are a nearly normal distribution.

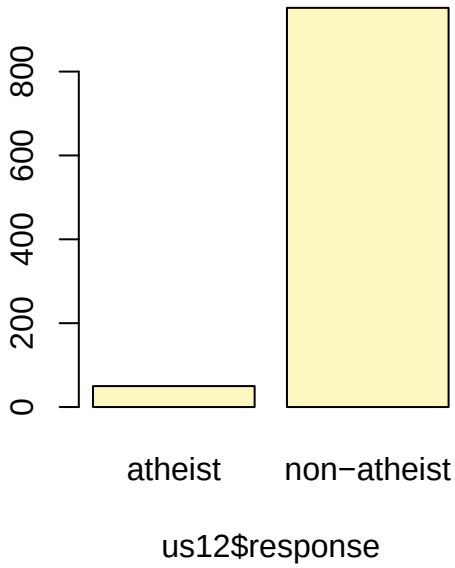
Although formal confidence intervals don't show up in the report, suggestions of inference appear at the bottom of page 7: "In general, the error margin for surveys of this kind is ± 3-5% at 95% confidence".

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:

```
inference(us12$response, est = "proportion", type = "ci", method = "theoretical", success = "atheist")
```

```
## Warning: package 'openintro' was built under R version 4.3.3

## Warning: package 'airports' was built under R version 4.3.3

## Warning: package 'cherryblossom' was built under R version 4.3.3

## Warning: package 'usdata' was built under R version 4.3.3

## Warning: package 'BHH2' was built under R version 4.3.3

## Single proportion -- success: atheist
## Summary statistics:
```




```
## p_hat = 0.0499 ;  n = 1002
## Check conditions: number of successes = 50 ; number of failures = 952
## Standard error = 0.0069
## 95 % Confidence interval = ( 0.0364 , 0.0634 )
```

confidence interval = (0.0364, 0.0634) margin of error = (0.0634 - 0.0364) / 2 = 0.0135

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 confidence interval would be narrower for 90% than 95% because the higher the level of confidence, higher the range of possible values as it has a greater interval.