Lab 6

Code

Kevin White

4/13/2023

Exercise 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? Explain your reasoning.

It appears that the these percentages are sample statistics as the data is teken from polls not the entire population

Exercise 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?

With “with more than 50,000 men and women selected from 57 countries” We must assume that the data is independent as we don’t have over 10% of the population. And with people from smaller countrys probobly excuded we can assume this data may have been different if those from other contrys where included in the polls

Exercise 3:

What does each row of Table 6 correspond to? What does each row of atheism correspond to?

Hide

```
data(atheism)
```

From the Rows in table 6, we can see: - the Sample size of each sample from each country - The presentage of country who claim to be religious - The presentage of country who claim to be non religions - The presentage of country who claim to be atheist - The presentage of country who did no respond or refused to answer the question

From the rows in “atheism” - Nationality of an individual - religious status of an individual - the year the individual was asked

Exercise 4:

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?

Hide

```
us12 <- atheism %>%
  filter(nationality == "United States", year == "2012")

prop.table(table(as.factor(as.character(us12$nationality)), us12$response))
```

```
##
##                   atheist non-atheist
##   United States 0.0499002   0.9500998
```

This information kind of aggrees with the other data, the 5% athiest matches, but the non-atheist is not broken down into other groups like table 6 of the report is/

Exercise 5:

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?

We must assume that the poll is independent as we dont have more than 10% of the population

We can also assume a normal distrabution because there are more than 10 responses for each option

Exercise 6:

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?

Hide

```
us12 %>%
  summarize(N = n(), atheist = sum(response == "atheist")) %>%
  mutate(p_hat = atheist / N,
         se = sqrt(p_hat * (1 - p_hat) / N),
         me = qnorm(0.975) * se,
         lower = p_hat - me,
         upper = p_hat + me)
```

```
## # A tibble: 1 × 7
##       N atheist  p_hat      se     me  lower  upper
##   <int>   <int>  <dbl>   <dbl>  <dbl>  <dbl>  <dbl>
## 1  1002      50 0.0499 0.00688 0.0135 0.0364 0.0634
```

Th margin of error is 0.01348187

Exercise 7:

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, and interpet the interval in context of the data. It may be helpful to create new data sets for each of the two countries first, and then use these data sets to construct the confidence intervals.

Hide

```
afghanistan12 <- atheism %>%
  filter(nationality == "Afghanistan", year == "2012")

prop.table(table(as.factor(as.character(afghanistan12$nationality)), afghanistan12$respons
         e))
```

```
##
##               atheist non-atheist
##   Afghanistan       0           1
```

Hide

```
afghanistan12 %>%
  summarize(N = n(), atheist = sum(response == "atheist")) %>%
  mutate(p_hat = atheist / N,
         se = sqrt(p_hat * (1 - p_hat) / N),
         me = qnorm(0.975) * se,
         lower = p_hat - me,
         upper = p_hat + me)
```

```
## # A tibble: 1 × 7
##       N atheist p_hat    se    me lower upper
##   <int>   <int> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  1031       0     0     0     0     0     0
```

Afghanistan has 0% atheists with an apparent margin of error 0 Our inferences can not be met here because we don't have at least 10 of athist and non-athiest

Hide

```
poland12 <- atheism %>%
  filter(nationality == "Poland", year == "2012")

prop.table(table(as.factor(as.character(poland12$nationality)), poland12$response))
```

```
##
##             atheist non-atheist
##   Poland 0.04952381  0.95047619
```

Hide

```
poland12 %>%
  summarize(N = n(), atheist = sum(response == "atheist")) %>%
  mutate(p_hat = atheist / N,
         se = sqrt(p_hat * (1 - p_hat) / N),
         me = qnorm(0.975) * se,
         lower = p_hat - me,
         upper = p_hat + me)
```

```
## # A tibble: 1 × 7
##       N atheist  p_hat      se     me  lower  upper
##   <int>   <int>  <dbl>   <dbl>  <dbl>  <dbl>  <dbl>
## 1   525      26 0.0495 0.00947 0.0186 0.0310 0.0681
```

Poland has 5% atheists with an apparent margin of error 0.01855864

Our inferences can be met as we have less than 10% of the population and at least 10 atheist and non-atheists.

Exercise 8:

Describe the relationship between p and me.

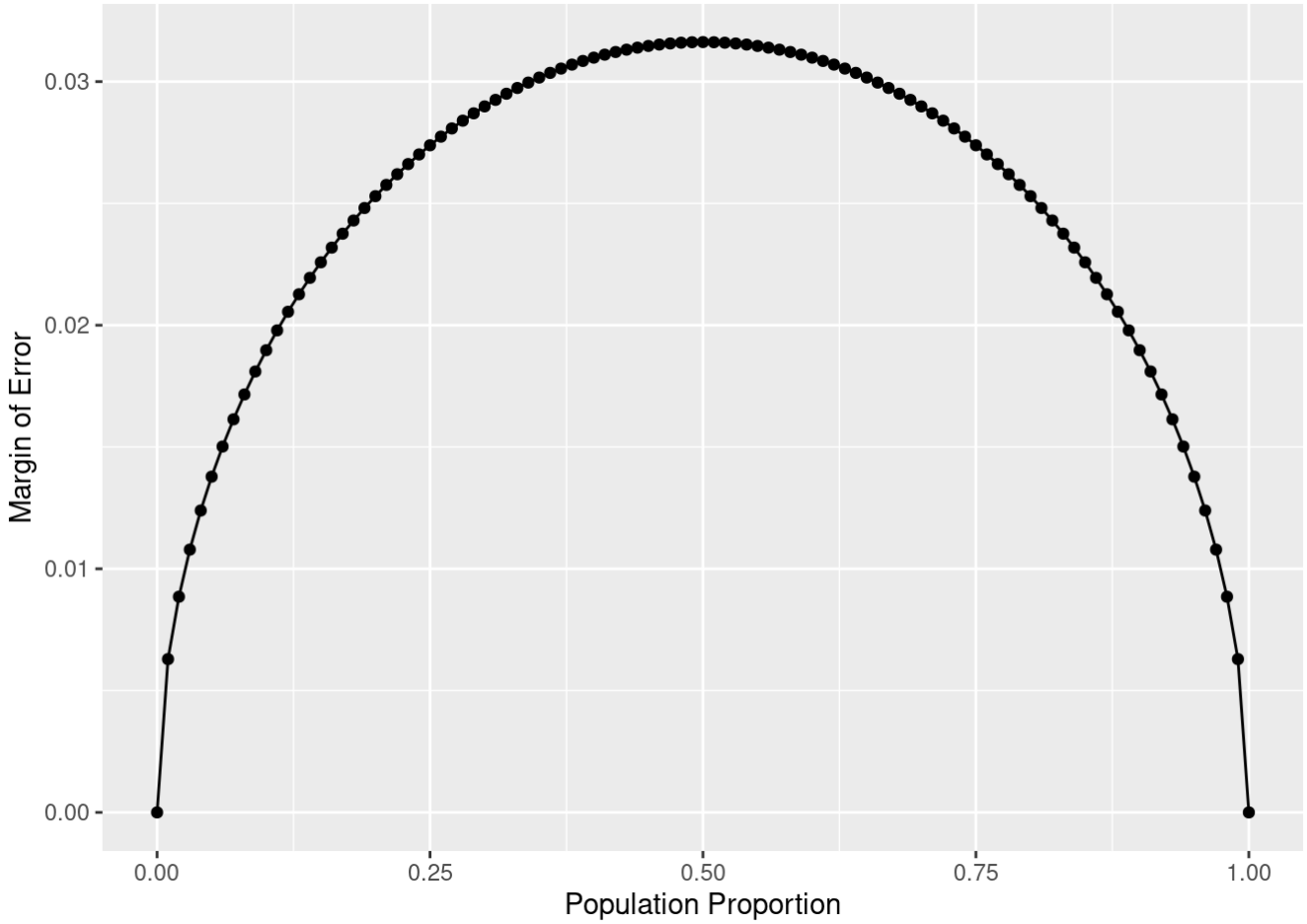
Hide

```
n <- 1000

p <- seq(from = 0, to = 1, by = 0.01)
me <- 2 * sqrt(p * (1 - p)/n)

dd <- data.frame(p = p, me = me)
qplot(x = p, y = me, data = dd,
      ylab = "Margin of Error",
      xlab = "Population Proportion") +
  geom_line()
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
```



As population proportion reaches 50% the margin of error increases from 0, and from 50% as the population proportion increases to 1.0 the margin of error decreases to 0. This creates a quadratic form.

Exercise 9:

Describe the sampling distribution of sample proportions at n=300 and p=0.1. Be sure to note the center, spread, and shape.

There is a symmetrical proportion centered around .1 with a spread of about .05 to the left and right

Exercise 10:

Keep n constant and change p . How does the shape, center, and spread of the sampling distribution vary as p changes. You might want to adjust min and max for the x -axis for a better view of the distribution.

Everything seams to be the same except what the figure is centerd around, there is still and even distrabution that spreads about .05 to the left and right of the shape

Exercise 11:

Now also change n. How does n appear to affect the distribution of p^?

The larger the sample size n, the more of a normal distribution the figure takes, as n gets smaller depending on p, there may be a tail to the data as it is scque left and right. However, when you add more data, this gets rid of the scque

Exercise 12:

If you refer to Table 6, you'll find that Australia has a sample proportion of 0.1 in 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. Construct their sampling distributions by using these values as inputs in the app. Do you think it is sensible to proceed with inference and report margin of errors, as the report does?

For Australia with a sample proportion of .1 there is at at least 10 people for atheist, and 10 for non-atheist, this is a good data set to assume normal distribution. However with Ecuador, with a sample population of 400 this means only 8 people identified as atheist this is not enough data to assume normal distribution

Exercise 13:

Is there convincing evidence that Spain has seen a change in its atheism index between 2005 and 2012? As always, write out the hypotheses for any tests you conduct and outline the status of the conditions for inference. If you find a significant difference, also quantify this difference with a confidence interval.

Ho: Spain 2005 = Spain 2012

Ha: Spain 2005 != Spain 2012

Hide

```
spain05 <- atheism %>%
  filter(nationality == "Spain", year == "2005")

prop.table(table(as.factor(as.character(spain05$nationality)), spain05$response))
```

```
##
##          atheist non-atheist
##   Spain 0.100349    0.899651
```

Hide

```
spain05 %>%
  summarize(N = n(), atheist = sum(response == "atheist")) %>%
  mutate(p_hat = atheist / N,
         se = sqrt(p_hat * (1 - p_hat) / N),
         me = qnorm(0.975) * se,
         lower = p_hat - me,
         upper = p_hat + me)
```

```
## # A tibble: 1 × 7
##       N atheist p_hat      se     me  lower upper
##   <int>   <int> <dbl>   <dbl>  <dbl>  <dbl> <dbl>
## 1  1146     115 0.100 0.00888 0.0174 0.0830 0.118
```

Hide

```
spain12 <- atheism %>%
  filter(nationality == "Spain", year == "2012")

prop.table(table(as.factor(as.character(spain12$nationality)), spain12$response))
```

```
##
##            atheist non-atheist
##   Spain 0.08995633  0.91004367
```

Hide

```
spain12 %>%
  summarize(N = n(), atheist = sum(response == "atheist")) %>%
  mutate(p_hat = atheist / N,
         se = sqrt(p_hat * (1 - p_hat) / N),
         me = qnorm(0.975) * se,
         lower = p_hat - me,
         upper = p_hat + me)
```

```
## # A tibble: 1 × 7
##       N atheist  p_hat      se     me  lower upper
##   <int>   <int>  <dbl>   <dbl>  <dbl>  <dbl> <dbl>
## 1  1145     103 0.0900 0.00846 0.0166 0.0734 0.107
```

2005 Confidence interval = (0.084, 0.118) 05 p-hat = 0.10

2012 Confidence interval = (0.073, 0.107) 12 p-hat = 0.09

There is not sufficient evidence to reject the null hypothesis as the Confidence intervals overlap by quite a bit, and p-hat for 12 is withint he confidence intarval for 05

Exercise 14:

Is there convincing evidence that the US has seen a change in its atheism index between 2005 and 2012? As always, write out the hypotheses for any tests you conduct and outline the status of the conditions for inference. If you find a significant difference, also quantify this difference with a confidence interval.

Hide

```
us05 <- atheism %>%
  filter(nationality == "United States", year == "2005")

prop.table(table(as.factor(as.character(us05$nationality)), us05$response))
```

```
##
##                    atheist non-atheist
##   United States 0.00998004  0.99001996
```

Hide

```
us05 %>%
  summarize(N = n(), atheist = sum(response == "atheist")) %>%
  mutate(p_hat = atheist / N,
         se = sqrt(p_hat * (1 - p_hat) / N),
         me = qnorm(0.975) * se,
         lower = p_hat - me,
         upper = p_hat + me)
```

```
## # A tibble: 1 × 7
##       N atheist   p_hat      se      me   lower  upper
##   <int>   <int>   <dbl>   <dbl>   <dbl>   <dbl>  <dbl>
## 1  1002      10 0.00998 0.00314 0.00615 0.00383 0.0161
```

Hide

```
us12 <- atheism %>%
  filter(nationality == "United States", year == "2012")

prop.table(table(as.factor(as.character(us12$nationality)), us12$response))
```

```
##
##                   atheist non-atheist
##   United States 0.0499002   0.9500998
```

Hide

```
us12 %>%
  summarize(N = n(), atheist = sum(response == "atheist")) %>%
  mutate(p_hat = atheist / N,
         se = sqrt(p_hat * (1 - p_hat) / N),
         me = qnorm(0.975) * se,
         lower = p_hat - me,
         upper = p_hat + me)
```

```
## # A tibble: 1 × 7
##       N atheist  p_hat      se     me  lower  upper
##   <int>   <int>  <dbl>   <dbl>  <dbl>  <dbl>  <dbl>
## 1  1002      50 0.0499 0.00688 0.0135 0.0364 0.0634
```

2005 Confidence interval = (0.0062, 0.0038) 05 p-hat = 0.0099

2012 Confidence interval = (0.0364, 0.0634) 12 p-hat = 0.0499

Their is sufficient evidence to reject the null hypothesis because the p-hat for 2012 is not within the confidence interval for 2005. This is supported even more by the lack of overlap between he confidence intervals of 12 and 05

Exercise 15:

If in fact there has been no change in the atheism index in the countries listed in Table 4, in how many of those countries would you expect to detect a change (at a significance level of 0.05) simply by chance?

if there was no recorded change for all 39 country that participated in both 2005 and 2012. With a 0.05 significance level we could get a type 1 error or false positive 5% of the time. This would mean that 5% of 39 or about 2 country would still show having a significant change.

Exercise 16:

Suppose you're hired by the local government to estimate the proportion of residents that attend a religious service on a weekly basis. According to the guidelines, the estimate must have a margin of error no greater than 1% with 95% confidence. You have no idea what to expect for p. How many people would you have to sample to ensure that you are within the guidelines?

(1.96/0.01)^2 * 1/2*(1 - 1/2)= 9604