Lab 7

**3) What does each row of Table 6 correspond to? What does each row of atheism correspond to?**

Each row of atheism corresponds to an observation, or a person who responded to the survey that asked if they are religious or not.

**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?**

table(us12$response)

atheist non-atheist

50 952

> 50/1002 = 0.0499002 = 4.99%

Yes, this calculation agrees with the Table’s percentage, because both say the proportion of atheists in the US is 5%.

**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?**

The data needs to be independent, where the sample size is less than 10% of the total population and from a random sample, which are true. Also, we are looking at the success-failure, where the success and failure are at least 10. This will be true, because 1002 observations in the US has 50 successes and 952 failures.

**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?**

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

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 )

The margin of error is Z\* \* SE = 1.96 \* 0.0069 = 0.0135

Also, p hat – lower value = .0499 - .0364 = 0.0135

**7) Using the inference 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 inference function to construct the confidence intervals.**

ind12 <- subset(atheism, nationality == "India" & year == "2012")

fra12 <- subset(atheism, nationality == "France" & year == "2012")

inference(ind12$response, est = "proportion", type = "ci", method = "theoretical", success = "atheist")

Single proportion -- success: atheist

Summary statistics: p\_hat = 0.0302 ; n = 1092

Check conditions: number of successes = 33 ; number of failures = 1059

Standard error = 0.0052

95 % Confidence interval = ( 0.0201 , 0.0404 )

Margin of error = 0.0302 - 0.0201 = 0.0101

inference(fra12$response, est = "proportion", type = "ci", method = "theoretical", success = "atheist")

Single proportion -- success: atheist

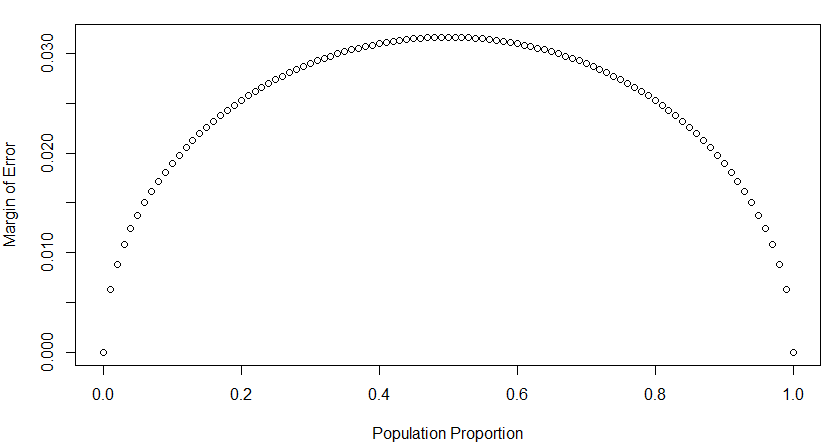
Summary statistics: p\_hat = 0.2873 ; n = 1688

Check conditions: number of successes = 485 ; number of failures = 1203

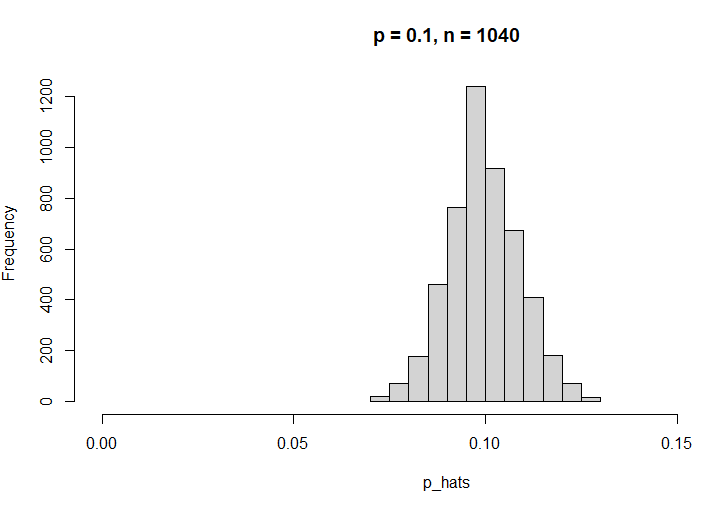
Standard error = 0.011

95 % Confidence interval = ( 0.2657 , 0.3089 )

Margin of error = 0.2873 - 0.2657 = 0.0216

**8) Describe the relationship between p and me.**

The formula gives a parabola to simulate the population proportion vs margin of error. This means that margin of error peaks at a p value that is about half the population it is describing. The p value increases the margin of error as it increases until a certain point is reached, then decreases it. It is smallest when p is 0 or 1. If there are 0 atheists or all people are atheists, then there can’t be any error.

**9) Describe the sampling distribution of sample proportions at n=1040n=1040 and p=0.1p=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 data appears nearly normal, with a center at 0.10 p hat value, a spread that ranges from ~0.075 to 0.125, and a solid, symmetric shape.

**10) Repeat 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 p^? How does p affect the sampling distribution?**

P <- 0.1

n <- 400

p\_hats <- rep(0, 5000)

for(i in 1:5000){

samp <- sample(c("atheist", "non\_atheist"), n, replace = TRUE, prob =

c(p, 1-p))

p\_hats[i] <- sum(samp == "atheist")/n

}

hist(p\_hats, main = "p = 0.1, n = 400", xlim = c(0, 0.18))

p <- 0.02

n <- 1040

p\_hats <- rep(0, 5000)

for(i in 1:5000){

samp <- sample(c("atheist", "non\_atheist"), n, replace = TRUE, prob =

c(p, 1-p))

p\_hats[i] <- sum(samp == "atheist")/n

}

hist(p\_hats, main = "p = 0.02, n = 1040", xlim = c(0, 0.18))

p <- 0.02

n <- 400

p\_hats <- rep(0, 5000)

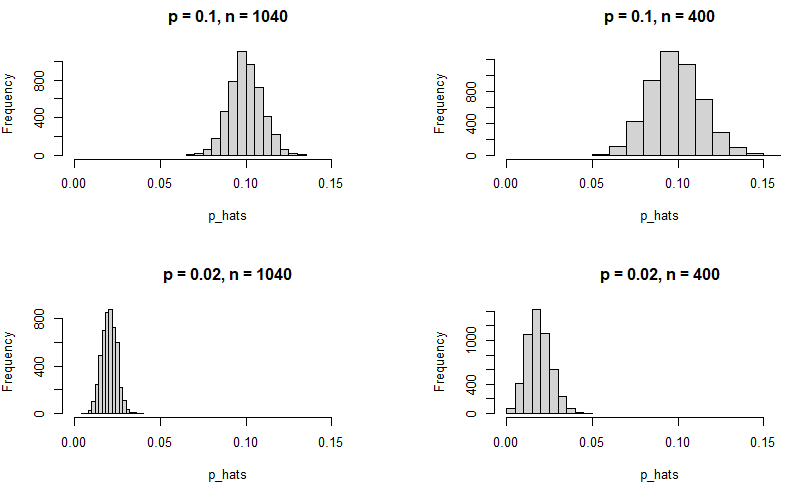
for(i in 1:5000){

samp <- sample(c("atheist", "non\_atheist"), n, replace = TRUE, prob =

c(p, 1-p))

p\_hats[i] <- sum(samp == "atheist")/n

}

hist(p\_hats, main = "p = 0.02, n = 400", xlim = c(0, 0.18))

Based on these plots, n appears to create more spread when the sample size gets smaller, because there is a larger standard error and spread when the sample size decreases. In addition, the graphs with smaller sample sizes appear a little more left-skewed and right-skewed, respectively. When p gets smaller, the p hat centers around the p value to the left of a larger p value’s histogram. This makes sense because p is the true success population proportion.

**11) 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?**

Yes, it would be sensible because both countries would have a random sample that is less than 10% of the population of that country, which would satisfy independence. In addition, the success-failure condition would be met if n\*p and n\*(1-p) are both greater than 10, which is true for Australia (which checks both). However, Ecuador has a success value of 8 (0.02\*400), so I would be skeptical of proceeding for inference for Ecuador.

**ON YOUR OWN**

1. **Answer the following two questions using the inference function. As always, write out the hypotheses for any tests you conduct and outline the status of the conditions for inference.**

**a. Is there convincing evidence that Spain has seen a change in its atheism index between 2005 and 2012?  
Hint: Create a new data set for respondents from Spain. Form confidence intervals for the true proportion of athiests in both years, and determine whether they overlap.**

Ho: No convincing evidence that Spain has seen a change in atheism between 2005 -2012.

p2005 = p2012

Ha: Convincing evidence that Spain has seen a change in atheism between 2005 -2012.

p2005 =/= p2012

spain12 <- subset(atheism, nationality == "Spain" & year == "2012")

spain05 <- subset(atheism, nationality == "Spain" & year == "2005")

inference(spain12$response, est = "proportion", type = "ci", method = "theoretical", success = "atheist")

Single proportion -- success: atheist

Summary statistics: p\_hat = 0.09 ; n = 1145

Check conditions: number of successes = 103 ; number of failures = 1042

Standard error = 0.0085

95 % Confidence interval = ( 0.0734 , 0.1065 )

inference(spain05$response, est = "proportion", type = "ci", method = "theoretical", success = "atheist")

Single proportion -- success: atheist

Summary statistics: p\_hat = 0.1003 ; n = 1146

Check conditions: number of successes = 115 ; number of failures = 1031

Standard error = 0.0089

95 % Confidence interval = ( 0.083 , 0.1177 )

**There is a decrease in the range of the confidence interval for Spain in 2012 vs. 2005. Based on the intervals, Spain has seen a decrease in people calling themselves atheists from 2005-2012, since the intervals go from (0.083 - 0.1177) to (0.0734 - 0.1065). Both p hat values are greater than 0.05 and are included in the other year’s confidence interval. The data overlap, so it is hard to conclude whether there is a real change for the atheism proportion for the population of Spain, so we would fail to reject a null hypothesis that there is no change in people calling themselves atheists in Spain, because the differences don’t appear convincing.**

**b. Is there convincing evidence that the United States has seen a change in its atheism index between 2005 and 2012?**

Ho: No convincing evidence that USA has seen a change in atheism between 2005 -2012.

p2005 = p2012

Ha: Convincing evidence that USA has seen a change in atheism between 2005 -2012.

p2005 =/= p2012

us05 <- subset(atheism, nationality == "United States" & year == "2005")

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

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 )

inference(us05$response, est = "proportion", type = "ci", method = "theoretical", success = "atheist")

Single proportion -- success: atheist

Summary statistics: p\_hat = 0.01 ; n = 1002

Check conditions: number of successes = 10 ; number of failures = 992

Standard error = 0.0031

95 % Confidence interval = ( 0.0038 , 0.0161 )

There is convincing evidence that the number of atheists in the US increased between 2005 to 2012, because the parameters of the confidence intervals do not overlap, as 2005 is (0.0038 to 0.0161) and 2012 is (0.0364 to 0.0634). The p hat values of 2005 and 2012 are .0499 and .01 which both are less than 0.05. Neither p hat value is in the parameter of the other year’s confidence interval. This is convincing evidence to reject the null hypothesis, and that the United States has seen a change in its atheism index between these two years.

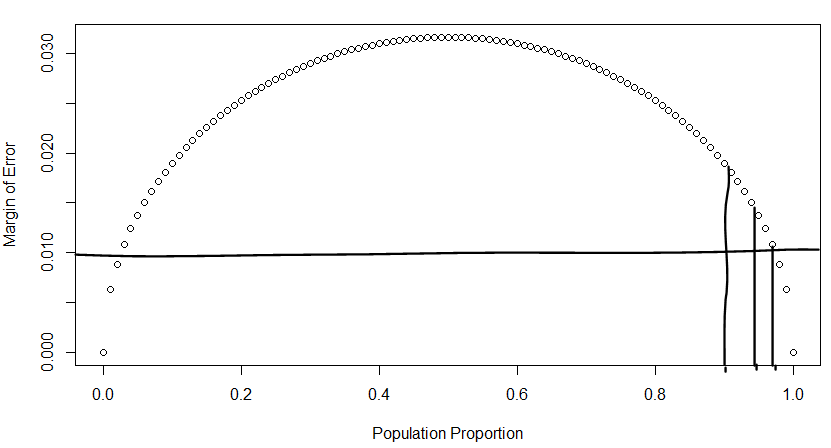
1. **If in fact there has been no change in the atheism index in the countries listed in Table 4, in how many of those countries would you expect to detect a change (at a significance level of 0.05) simply by chance?  
   Hint: Look in the textbook index under Type 1 error.**

A Type 1 error is made when we reject a null hypothesis when it is in fact true. This is when we say the p value is less than 0.05, when it is supposed to be greater than 0.05.

There are 39 countries in Table 4, and saying we make an error by chance is saying at a significance level of 0.05:

39 countries \* 0.05 = about 2 countries would detect change by chance.

1. **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?  
   Hint: Refer to your plot of the relationship between p and margin of error. Do not use the data set to answer this question.**

To get a margin of error less than or equal to 1%, we would need a p value that is very close to 1, so it can represent as close to the total population as possible. This means we need a p value that is between about 0.975 – 1.0 of the population.

If we had to calculate the number of people we need with no knowledge of p, however, we need to look at the margin of error formula, which is ME = critical value \* sqrt((p\*(1-p))/n). If we have no knowledge of p, we would assume the case that p is 0.5 because we know that’s the worst case for margin of error, so if we asked enough people, the p value wouldn’t really matter for margin of error.

We get: 0.01 = 1.96 \* sqrt((0.5\*0.5)/n)

Solving for n, we get 9604 people to ask.