IPS9 in R: Inference for Proportions (Chapter 8)

Bonnie Lin and Nicholas Horton (nhorton@amherst.edu)

January 19, 2019

Introduction and background

These documents are intended to help describe how to undertake analyses introduced as examples in the Ninth Edition of *Introduction to the Practice of Statistics* (2017) by Moore, McCabe, and Craig.

More information about the book can be found here. The data used in these documents can be found under Data Sets in the Student Site. This file as well as the associated R Markdown reproducible analysis source file used to create it can be found at https://nhorton.people.amherst.edu/ips9/.

This work leverages initiatives undertaken by Project MOSAIC (http://www.mosaic-web.org), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the `mosaic` package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the mosaic package vignettes (http://cran.r-project.org/web/packages/mosaic). A paper describing the mosaic approach was published in the *R Journal*: https://journal.r-project.org/archive/2017/RJ-2017-024.

Chapter 8: Inference for Proportions

This file replicates the analyses from Chapter 8: Inference for Proportions.

First, load the packages that will be needed for this document:

```
library(mosaic)
library(readr)
```

Section 8.1: Inference for a single proportion

```
ROBOT <- read_csv("https://nhorton.people.amherst.edu/ips9/data/chapter08/EG08-01ROBOT.csv")
ROBOT
```

```
## # A tibble: 2 x 2
##   Decrease Count
##   <chr>    <dbl>
## 1 Yes        910
## 2 No         986
```

```
### Example 8.2, page 487
prop.test(910, 910 + 986,
          alternative="two.sided",
          conf.level=0.95)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  910 out of 910 + 986
## X-squared = 2.9668, df = 1, p-value = 0.08499
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
```

```
##  0.4572704 0.5027272
## sample estimates:
##         p
## 0.4799578
```

By default, the `read_csv()` function will output the types of columns, as we see above. To improve readability for future coding, we will suppress the "Parsed with column specification" message by adding `message = FALSE` at the top of the code chunks.

```
### Example 8.5 and 8.6, page 492-493
prop.test(13, 20, correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  13 out of 20
## X-squared = 1.8, df = 1, p-value = 0.1797
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.4328543 0.8188082
## sample estimates:
##    p
## 0.65
```

Here we can replicate the results in the book using the `prop.test` function with the `correct = FALSE` option to turn off continuity correction. A more appropriate approach given the small sample size would be to use exact inference and the `binom.test()` function.

```
binom.test(13, 20)
```

```
##
##
##
## data:  13 out of 20
## number of successes = 13, number of trials = 20, p-value = 0.2632
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4078115 0.8460908
## sample estimates:
## probability of success
##                   0.65
```

```
INSTAG <- read_csv("https://nhorton.people.amherst.edu/ips9/data/chapter08/EG08-11INSTAG.csv")
### Example 8.11, page 507
INSTAG_prop <- INSTAG %>%
  group_by(Sex) %>%
  mutate(Prop_by_Sex = Count/sum(Count)) %>%
  ungroup() %>%
  filter(User == "Yes")
INSTAG_prop
```

```
## # A tibble: 2 x 4
##   Sex    User  Count Prop_by_Sex
##   <chr>  <chr> <dbl>       <dbl>
## 1 1Women Yes     328       0.611
## 2 2Men   Yes     234       0.440
```

```
### Example 8.15, page 513
prop.test(x = c(328, 234), n = c(537, 532), correct = FALSE)
```

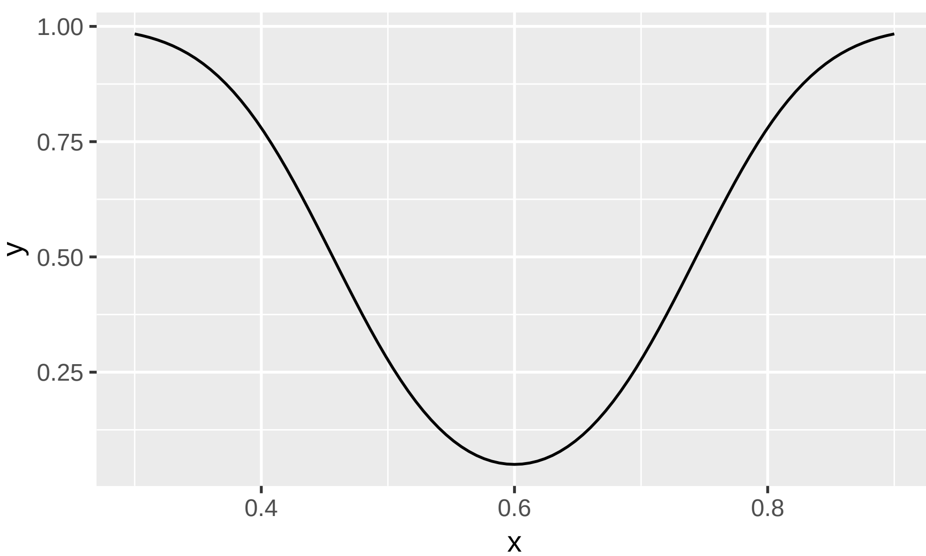
```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(328, 234) out of c(537, 532)
## X-squared = 31.323, df = 1, p-value = 2.185e-08
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1119627 0.2299396
## sample estimates:
##    prop 1    prop 2
## 0.6108007 0.4398496
```

```
### Example 8.16, page 515
zstar = qnorm(.975) # for 95% confidence interval
p = 0.5 # planned proportion estimate
moe = 0.1 # margin of error
p*(zstar/moe)^2
```

```
## [1] 192.0729
```

```
### Example 8.17, page 516
### Power curve for two proportions
calculatePower <- function(altmean = x, prop1 = 0.4, prop2 = 0.6, n = 97, alpha = 0.05){
  range <- 1-alpha/2
  se <- sqrt((prop1*(1-prop2)/n) + (prop2*(1-prop1)/n))
  bound1 <- qnorm(range, mean = 0.6, se) #sig lvl in null dist
  bound2 <- qnorm(alpha/2, mean = 0.6, se)
  1 - pnorm(bound1, altmean, se) + pnorm(bound2, altmean, se) #calc area past sig lvl
}

gf_function(fun = calculatePower, xlim = c(0.3, 0.9))
```



Section 8.2: Comparing two proportions