IPS9 in R: Introduction to Inference (Chapter 6)

Shukry Zablah (szablah20@amherst.edu) and Nicholas Horton (nhorton@amherst.edu)

July 18, 2018

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 6: Introduction to Inference

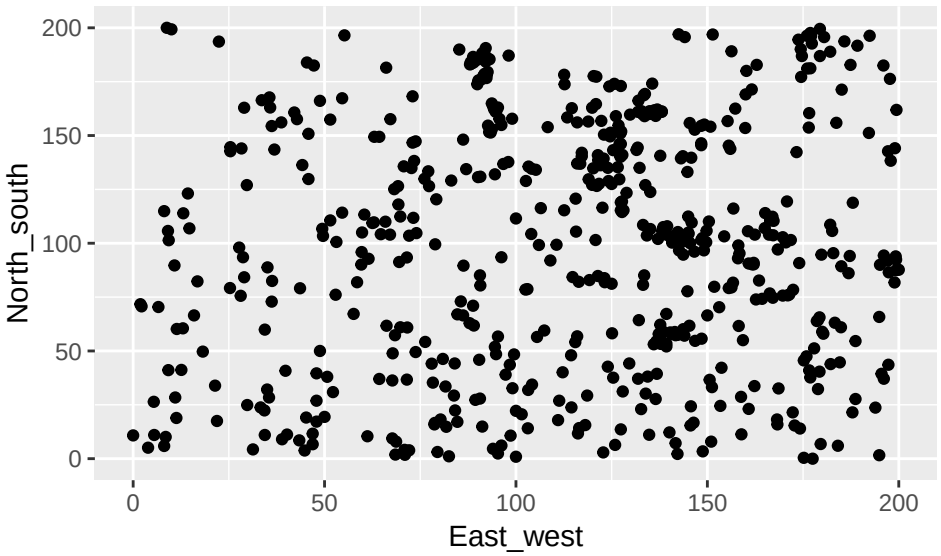
This file replicates the analyses from Chapter 6: Introduction to Inference.

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

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

Section 6.1: Estimating with Confidence

```
Wade <- read_csv("https://nhorton.people.amherst.edu/ips9/data/chapter06/EG06-01WADE.csv")
Wade %>%
  mutate(North_south = ns,
         East_west = ew) %>%
  gf_point(North_south ~ East_west, data = Wade)
```



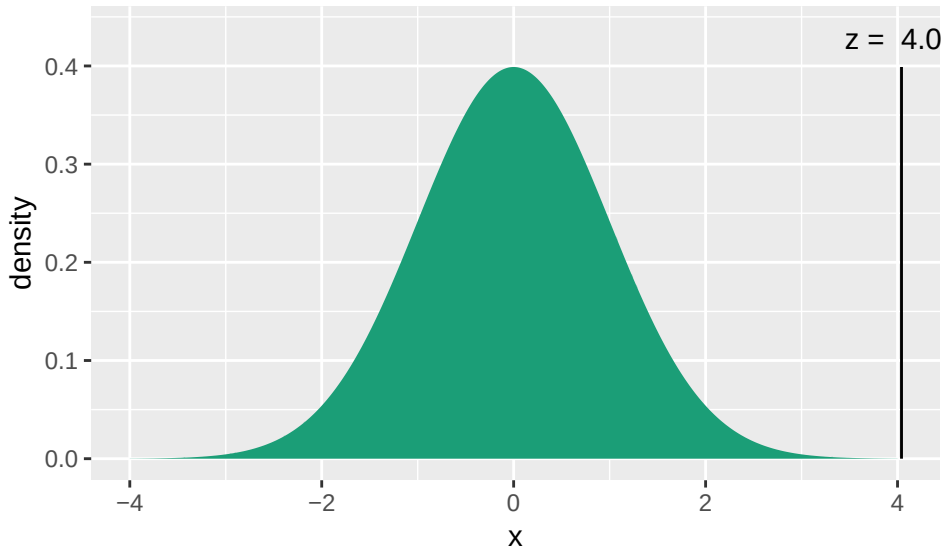
XX "One approach to the analysis of these data indicates that a pattern as clustered or more clustered than this would only happen 4% of the time if the data were at random" XX I'm unsure how to replicate this.

Section 6.2: Tests of Significance

Look at Ex6.17 in page 375. We will recreate the calculation as a function of the vector of observations, the population mean, and the known standard deviation. We then visualize it and get the probability of getting a z score equal or greater.

```
z.test <- function(x, mu, sd){
   z = (mean(x) - mu) / (sd / sqrt(length(x)))
   return(z)
}
z <- z.test(x = c(15.84, 15.33, 15.58), mu = 15, sd = 0.25)
xpnorm(z)
```

```
##
## If X ~ N(0, 1), then
##  P(X <= 4.041) = P(Z <= 4.041) = 1
##  P(X >  4.041) = P(Z >  4.041) = 2.656e-05
##
```




```
## [1] 0.9999734
```

```
z.confint <- function(x, mu, sd) {
  err <- qnorm(0.995)*sd/sqrt(length(x))
  lower <- mean(x) - err
  upper <- mean(x) + err
  return(c(lower, upper))
}

z.confint(x = c(15.84,15.33,15.58), mu = 15, sd = 0.25)
```

```
## [1] 15.21154 15.95512
```

Section 6.3: Use and abuse of tests

Section 6.4: Power and inference as a decision