Lab 5: Normal Distribution

Yasmine Pascal

02/18/2022

## Instructions

This document is the template you will use to type up your responses to the lab exercises that you can find on Canvas.

To complete the lab type your BRIEF answers and the R code in the spaces provided below, according to these guidelines:

* In general, if an exercise is numbered you will submit the R code that you used to answer the question in the gray area between the three ` marks.
* If you need to submit an answer in addition to the code type this between the line that starts #### and the beginning of the gray area for the code.
* Do not ever type the View() command in a lab report. This works only in RStudio on your computer, so don’t include that command in your lab report or you will not be able to compile the document, a process called *knitting*.
* To produce a document that you can submit on Canvas just click on the Knit button in the upper left quadrant of RStudio.
* We recommend pressing the Knit button after entering each answer: it’s very common to get errors, and they’re easier to fix if you can easily isolate the code that caused the error.
* The Knit button will create two files: one is a .Rmd file (R Markdown) and the other is an .docx file. You will need to submit both these documents for the lab assignment on Canvas.

## Preliminaries

#### Exercise 1

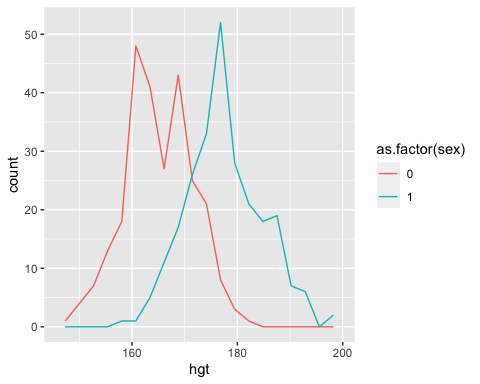
library(tidyverse)  
library(mosaic)  
library(ggformula)  
library(openintro)

## The Data

#### Exercise 2

The average height of a man is 177.7 cm while the average height of a woman is 164.9 cm, with a standard deviation of 7.2cm and 6.5cm respectively. In this group, the shortest woman was 147.2 cm while the shortest man was 157.2 cm and the tallest man was 198.1 cm while the tallest woman was 164.9 cm. The histograms are both unimodal and are roughly symmetrical, as they trail off in both directions.

gf\_freqpoly(~hgt, color= ~as.factor(sex), data=bdims, bins = 20)



favstats( ~hgt| as.factor(sex), data= bdims)

## as.factor(sex) min Q1 median Q3 max mean sd n missing  
## 1 0 147.2 160.0 164.5 169.50 182.9 164.8723 6.544602 260 0  
## 2 1 157.2 172.9 177.8 182.65 198.1 177.7453 7.183629 247 0

#### Exercise 3

mdims<- filter(bdims, sex==1)  
fdims<- filter(bdims, sex== 0)

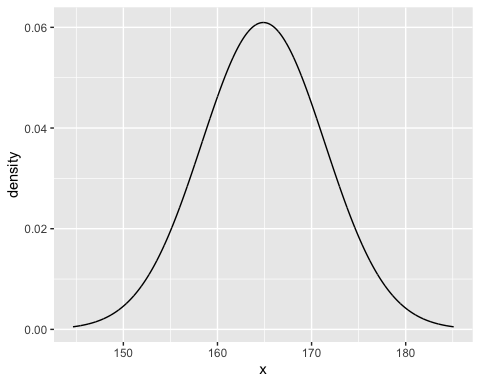
## Evaluating the Normal Distribution

#### Exercise 4:

fhgtmean <- mean(~hgt, data= fdims)  
fhgtsd <- sd(~hgt, data= fdims)

#### Exercise 5:

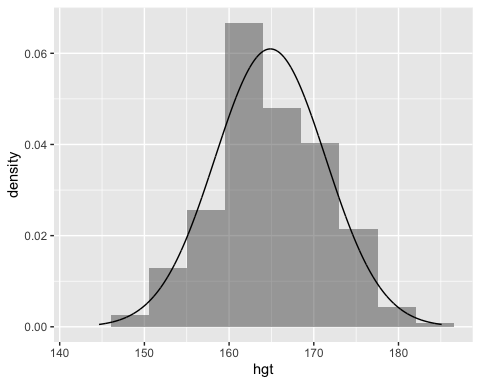
gf\_dist("norm", mean = fhgtmean, sd = fhgtsd)



#### Exercise 6:

Even though the peak of the normal distribution and the histograms are not in the same place, the histogram roughly follows a normal distribution

gf\_dhistogram(~hgt, data= fdims, bins= 10) %>% gf\_dist("norm", mean = fhgtmean, sd = fhgtsd)

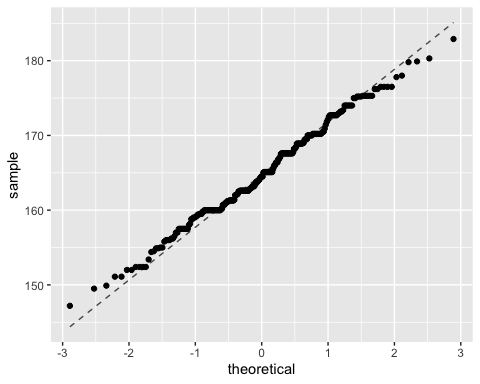


## A Q-Q plot: the standard method

#### Exercise 7:

The points mainly follow the line and the points of deviation occurs at both tails of the data.

gf\_qq(~hgt, data=fdims) %>%   
gf\_qqline()



## Computing normal probabilities

#### Exercise 8:

The proportion of women who are taller than 182cm is .0038

prop(~(hgt>182), data= fdims, success= "TRUE")

## prop\_TRUE   
## 0.003846154

#### Exercise 9:

The probability that a randomly selected woman is above 182cm is .004. This value is reasonably close to the empirical value I computed in the previous exercise.

1-xpnorm(182, mean=fhgtmean, sd= fhgtsd)

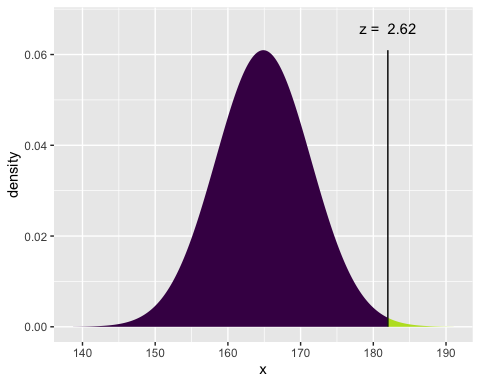
##

## If X ~ N(164.9, 6.545), then

## P(X <= 182) = P(Z <= 2.617) = 0.9956

## P(X > 182) = P(Z > 2.617) = 0.004434

##



## [1] 0.004434387