solution-2.R

wickhamc

Thu Oct 15 10:33:06 2015

library(ggplot2)
cdc <- read.csv(url("http://stat511.cwick.co.nz/homeworks/cdc.csv"))

# 1

## a

cdc$wt\_diff <- with(cdc, weight - wtdesire)

*(0 pts)*

## b

males = subset(cdc, gender == "m")
females = subset(cdc, gender == "f")

mean(males$weight)

## [1] 190.0305

sd(males$weight)

## [1] 38.01829

length(males$weight)

## [1] 492

mean(males$wtdesire)

## [1] 179.2012

sd(males$wtdesire)

## [1] 25.80331

length(males$wtdesire)

## [1] 492

mean(females$weight)

## [1] 152.0551

sd(females$weight)

## [1] 34.71281

length(females$weight)

## [1] 508

mean(females$wtdesire)

## [1] 134.752

sd(females$wtdesire)

## [1] 19.34982

length(females$wtdesire)

## [1] 508

*(2 pts)*

## c

qplot(wt\_diff, data = females, geom = "histogram", binwidth = 1)



It appears about one quarter of females are happy with their weight (the difference between weight and desired weight is zero). Of those that aren't most desire to be lighter (the difference is positive). The majority of females fall between wanting to weigh 10 pounds heavier and 50 pounds lighter. There seems to be a tendancy for the difference to be a multiple of five, probably reflecting that most people reported their weight and desired weight to the nearest five pounds. There are a few females reporting wanting to lose over 100 pounds.

*(1pt plot, 1pt description)*

## d

# qplot(weight,data=cdc,geom="histogram", binwidth = 5) + facet\_wrap(~gender,nrow=2)
qplot(weight,data=cdc,geom="histogram", binwidth = 1) + facet\_wrap(~gender,nrow=2)



As to be expected on average males in this sample are heavier than females in this sample. Apart from the different centers, both weight distributions have similar spreads are similar shapes: unimodal and roughly symmetric. There seem to be unusually many females reporting a weight of 150lbs and males reporting 160lbs.

*(1pt plot, 1pt description)*

# 2

t.test(females$wt\_diff)

##
## One Sample t-test
##
## data: females$wt\_diff
## t = 15.702, df = 507, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 15.13822 19.46808
## sample estimates:
## mean of x
## 17.30315

We have convincing evidence that the population mean difference in weight and weight desired for females is not equal to zero (paired t-test p-value < 2.2e-16) We estimate that a female's weight is on average 17 lbs more than her desired weight. With 95% confidence we estimate that a female weight is on average between 15 and 19 lbs greater than her desired weight.

*Or equivalently worded in terms of desired weight loss or gain.*

*(1 pt doing correct test, 3pts for summary)*