**Null hypothesis significance testing**

Given a data set of self-reports of height and weight (use the Height and Weight Data.csv file), use a *t* test to assess the null hypothesis that the mean weight of male respondents is equal to the mean weight of female respondents.

H0:

If you have not yet learned to use *t* tests (in the Tests of Differences chapter):

> file = read.csv("Height and Weight Data.csv")

> wtGender = with(file, data.frame(weight, sex)) # weight is the dependent variable, sex the grouping

> wtGender = na.omit(sentence)

> shapiro.test(wtGender$weight) # a test of normality

Shapiro-Wilk normality test

data: wtGender$weight

W = 0.87074, p-value = 4.919e-12

We have no reason to reject the null hypothesis, of normality of the dependent variable

> summary(wtGender)

weight sex

Min. : 39.0 F :112

1st Qu. : 55.0 M: 88

Median : 63.0

Mean : 65.8

3rd Qu.: 74.0

Max. :166.0

> wtGender$weight; sex = wtGender$sex # making for easier coding for tests

> options(scipen = 999, digits = 3) # removes scientific notation

There is rather an imbalance between the number of males and females in the samples. So instead of using the usual t test (otherwise known as Student's t test) – which would be

t.test(weight ~ sex, paired=FALSE, var.equal=TRUE) # dependent variable attached by ~ to factor

- we'll use Welch's *t* test:

> t.test(weight ~ sex, var.equal=FALSE)

Welch Two Sample t-test

data: weight by sex

t = -10, df = 200, p-value <0.0000000000000002

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-21.4 -14.6

sample estimates:

mean in group F mean in group M

57.9 75.9

Show your analysis of these results. You may also want to use the describeBy function, as in the exercise for the previous chapter (descriptive statistics).

ANSWER