T-Test Materials

Brier Gallihugh, M.S.

2024-07-11

Table of contents

An Introduction to T-tests 1
Assumptions of T-tests 1
Normality of Residuals 1
Graphical Depiction of Normality of Residuals 2
Statistical Depiction of Normality of Residuals 3
Homogeneity of Variance 3
Graphical Depiction of Homogeneity of Variance 3
Statistical Depiction of Homogeneity of Variance 4
Running a T-test 4

An Introduction to T-tests

Assumptions of T-tests

Normality of Residuals

```
library(tidyverse)

data <- starwars %>%
  filter(sex == "male" | sex == "female")

model <- lm(height ~ sex, data = data)

residuals <- data.frame(res = residuals(model))

problem <- residuals %>% filter(res > 2.5 | res < -2.5)

nrow(problem)/nrow(data)
```

Line 4

Filtering for male the female using the `filter()` function

Line 6

Running a linear regression (that is a t-test here) to get residuals

Line 8

Calculate residuals for the observations

Line 10

Find potentially problematic observations

```
[1] 0.8421053
```

Graphical Depiction of Normality of Residuals

```
residual_graph <- ggplot(residuals,aes(x = res)) +
  geom_histogram(aes(y=after_stat(density))) +
  stat_function(fun = dnorm,
                args = list(mean = mean(residuals$res),
                            sd = sd(residuals$res)),
                col = "blue",
                linewidth = 1) +
  theme_classic()

print(residual_graph)
```

Line 1

We are plotting the residuals here. We give ggplot a `geom` (i.e., histogram)

Line 2

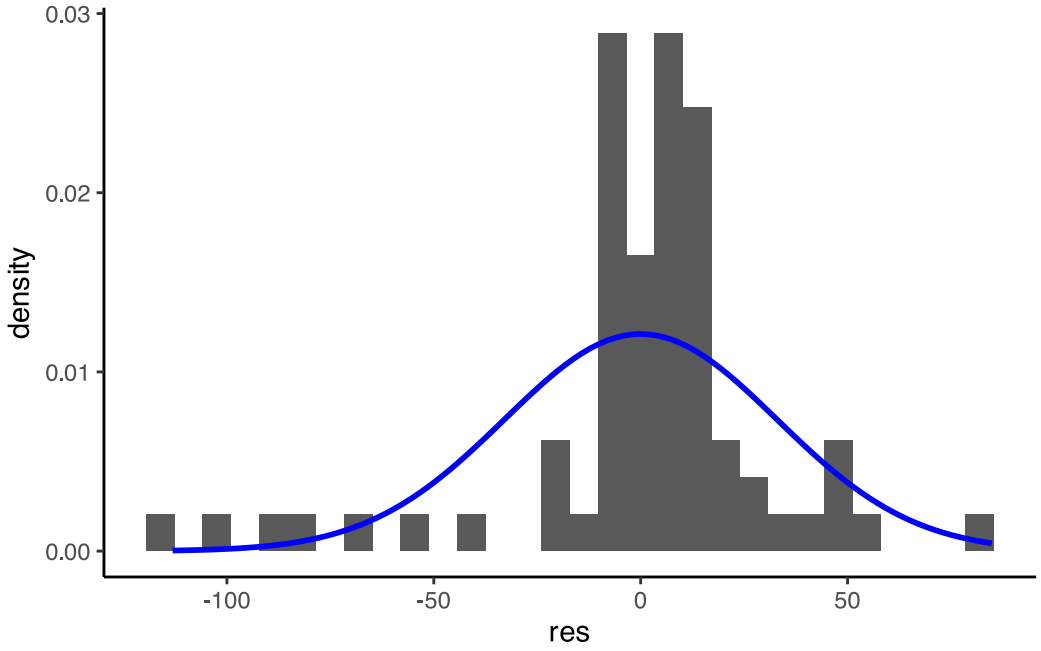
We also give some other arguments like a density distribution.

Lines 3-8

Here we are basically providing what is needed to draw a normal distribution given the data using the `stat_function()` function. The `col` and `linewidth` arguments simply change the colr and size of the normal curve. The `theme_classic()` just changes some aesthetic things. I personally prefer this theme for all ggplot2 graphs

Line 10

`print` will show us the graph output



Statistical Depiction of Normality of Residuals

We can also test the assumption statistically using the `shapiro.test()` function here

```
shapiro.test(residuals$res)
```

```
    Shapiro-Wilk normality test

data:  residuals$res
W = 0.84834, p-value = 5.13e-07
```

Homogeneity of Variance

Homogeneity of variance is important even for a basic t-test. Below is how we might go about testing this assumption.

Graphical Depiction of Homogeneity of Variance

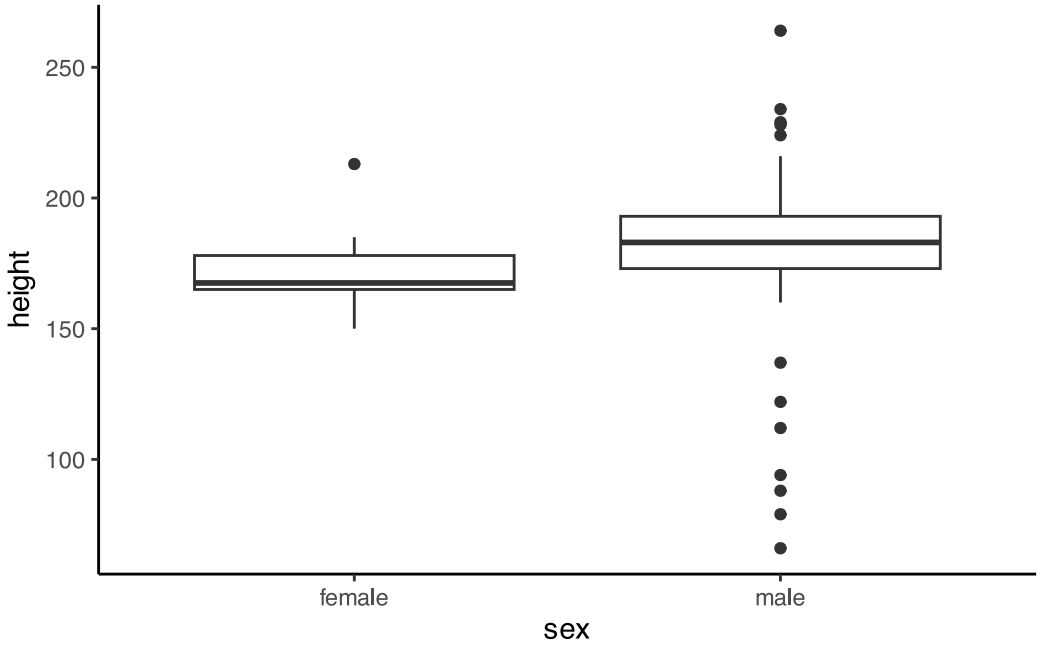
```
variance_boxplot <- ggplot(data,aes(x = sex,
                                    y = height)) +
  geom_boxplot() +
  theme_classic()

print(variance_boxplot)
```

Lines 1-2

Graphically we can represent this as a boxplot with the group variable as the `x` and the outcome as the `y`. We see this here

Lines 3-4

We again provide a `geom` for ggplot2 to use and provide a `theme()` choice here



Statistical Depiction of Homogeneity of Variance

We can also test the assumption using the Bartlett test. This can be shown below

```
bartlett.test(height ~ sex,data)
```

```
    Bartlett test of homogeneity of variances

data:  height by sex
Bartlett's K-squared = 9.6316, df = 1, p-value = 0.001913
```

Running a T-test

```
t.test(height ~ sex, data = data)
```

Line 1

The `t.test()` function will take a DV and IV argument as well as the dataframe used. We can see this here

```
    Welch Two Sample t-test

data:  height by sex
t = -1.1817, df = 48.471, p-value = 0.2431
alternative hypothesis: true difference in means between group female and group
male is not equal to 0
95 percent confidence interval:
 -20.396341   5.293584
sample estimates:
mean in group female   mean in group male
            171.5714             179.1228
```