Stats with Sparrows - 5

Julia Schroeder

5

Hypothesis testing

What do you remember about hypothesis testing? I am sure there are some concepts in your brains about p-values and 0.05?

p-value — the probability of observing this particular result (and more extreme results) when the null hypothesis is true (i.e. there is no effect).

Ok, let's try this out. But first, - you know, housekeeping:

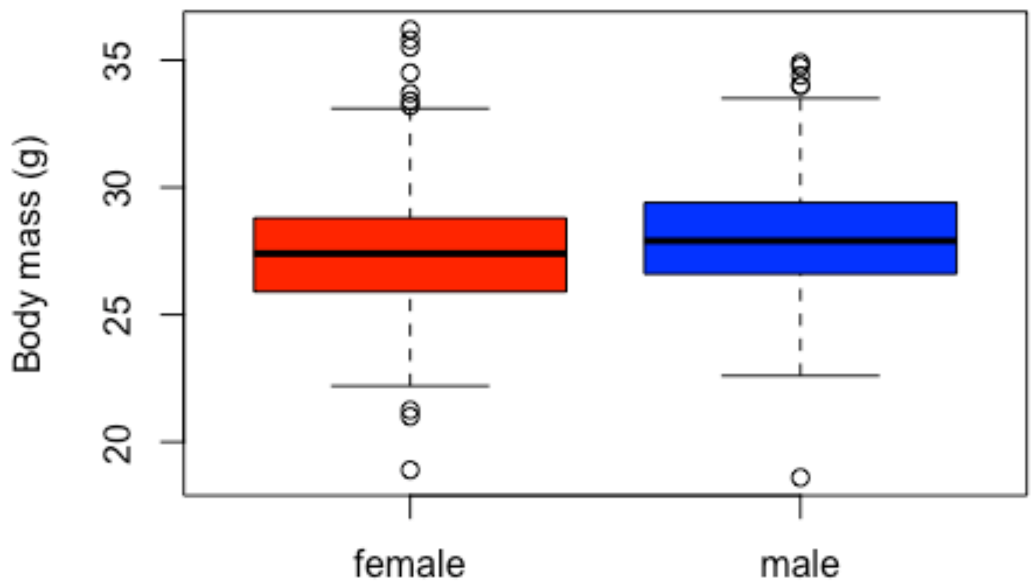
```
rm(list=ls())

setwd("~/Box Sync/Teaching/IntroStats")

d<-read.table("SparrowSize.txt", header=TRUE)
```

We will use it to test for a difference in female and male body mass in house sparrows. Let's see what we can gleam from a boxplot:

```
boxplot(d$Mass~d$Sex.1, col = c("red", "blue"), ylab="Body mass (g)")
```



It looks like males are slightly heavier than females, but how can we tell if that difference means something? We first have to find out our null hypothesis. In this case, it is that the DIFFERENCE between males and females in body mass is different from zero. If this difference is positive, it means males have larger body mass, if it is negative, it means males have lower body mass than females.

What test can we use to test for this?

```
t.test1 <- t.test(d$Mass~d$Sex.1)
t.test1

##
##  Welch Two Sample t-test
##
## data:  d$Mass by d$Sex.1
## t = -5.5654, df = 1682.9, p-value = 3.039e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7669117 -0.3672162
## sample estimates:
## mean in group female   mean in group male
##             27.46852             28.03558
```

R is quite helpful here in that it tells us the alternative hypothesis: that the difference between the two groups is not equal to 0. That is what we found, because the t-value is large, given the degrees of freedom (1683). This means that with a veyr high probability, males are heavier than females. So should we accept the alternative hypothesis then?

Yes, but wait, does it actually make biological sense? The p-value is super small, so most people would be very excited.But I want to teach you to not too much focus on the p-values, and instead look at 95%CIs, and the parameter estimates. The male mean is 28.0g, the female mean is 27.5g. So the difference is about half a gram. Howver, that's the mean difference. But can we say something about the precisision of this effect? Look at the output, it actually gives us a 95CI of the difference between males and females. 95% of the differences between males and females fall between -0.77g and -0.37g (males heavier). This gives us a good indication about how important this difference is in biology.

5% of the times, however, the difference will be outside of this interval. That's a type 1 error. There is a 5% chance that this data is actually not representing the real world, and that the difference between the sexes is actully 0.

Large datasets are more likely to pick up on small effect sizes (remember the square root law). Let's see if we would reduce our dataset to the 50 first rows, could we still detect a difference between male and female body mass?

```
d1<-as.data.frame(head(d, 50))
length(d1$Mass)
```

```
## [1] 50
```

```
t.test2 <- t.test(d1$Mass~d1$Sex)
t.test2
```

```
##
##  Welch Two Sample t-test
##
## data:  d1$Mass by d1$Sex
## t = 0.34955, df = 7.4242, p-value = 0.7364
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.997304  2.699685
## sample estimates:
## mean in group 0 mean in group 1
##        29.34286        28.99167
```

Oh. All of the sudden, we find no difference! Just because we took a smaller dataset. The lesson here is that with large datasets, you are more likely to encounter a statistically significant effect, but whether or not this effect is actually meaningful, is not something you can understand by looking at the p-value. You need to look at the absolute effect size. In the first t.test, the difference was really small. 0.3g more or less is not a lot. It is actually less than what a bird fluctuates over a day.

Excercises:

Test if wing length in 2001 differs from the grand-total mean Test if male and female wing length differ in 2001 Test if male and female wing length differ in the full dataset Report in a table (on whiteboard)

Test if male and female tarsus differs in the full dataset Report in text (on whiteboard)