Stat 630 - 01 Lab 9

Chris Kalra

11/1/2018

Exercise 1

```
library(openintro)
```

```
## Please visit openintro.org for free statistics materials
```

```
##
## Attaching package: 'openintro'
```

```
## The following objects are masked from 'package:datasets':
##
##     cars, trees
```

```
data('ncbirths')
matureweight=subset(ncbirths$gained, ncbirths$mature=='mature mom')
nonmatureweight=subset(ncbirths$gained, ncbirths$mature!='mature mom')
t.test(matureweight, nonmatureweight)
```

```
##
##  Welch Two Sample t-test
##
## data:  matureweight and nonmatureweight
## t = -1.3765, df = 175.34, p-value = 0.1704
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.3071463  0.7676886
## sample estimates:
## mean of x mean of y
##  28.79070  30.56043
```

With a p-value of .1704 and a confidence interval that includes 0, we conclude that weight gained by mothers is not dependent on whether or not she is a mature mom. We need to assume that the samples are random and that the data are normal (which is assumed true due the CLT, since both of our samples are large)

Exercise 2

By clicking the data set "ncbirths" and scrolling down, one can see that the age cutoff for young versus mature mom is 35. That is, an age under 35 assigns a mother to the "young mom" category, and an age of 35 or over assigns a mother to the "mature mom" category

Exercise 3

```
agediff=ncbirths$fage - ncbirths$mage
t.test(agediff, conf.level=.9)
```

```
##
##  One Sample t-test
##
## data:  agediff
## t = 17.673, df = 828, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  2.405432 2.899755
## sample estimates:
## mean of x
##  2.652593
```

We are 90% confident that, on average, the father is between 2.405 and 2.8998 years older than the mother

Exercise 4

Question of interest: Does aother's age does affect whether or not her baby is born prematurely (a "premie")?
Null hypothesis: A mother's age does affect whether or not her baby is born prematurely (a "premie")
Alternative hypothesis: A mother's age does not affect whether or not her baby is born prematurely (a "premie")

```
AgeAndPremie=subset(ncbirths$mage, ncbirths$premie=='premie', na.rm=TRUE)
AgeAndNonpremie=subset(ncbirths$mage, ncbirths$premie=='full term', na.rm=TRUE)
t.test(AgeAndNonpremie, AgeAndPremie)
```

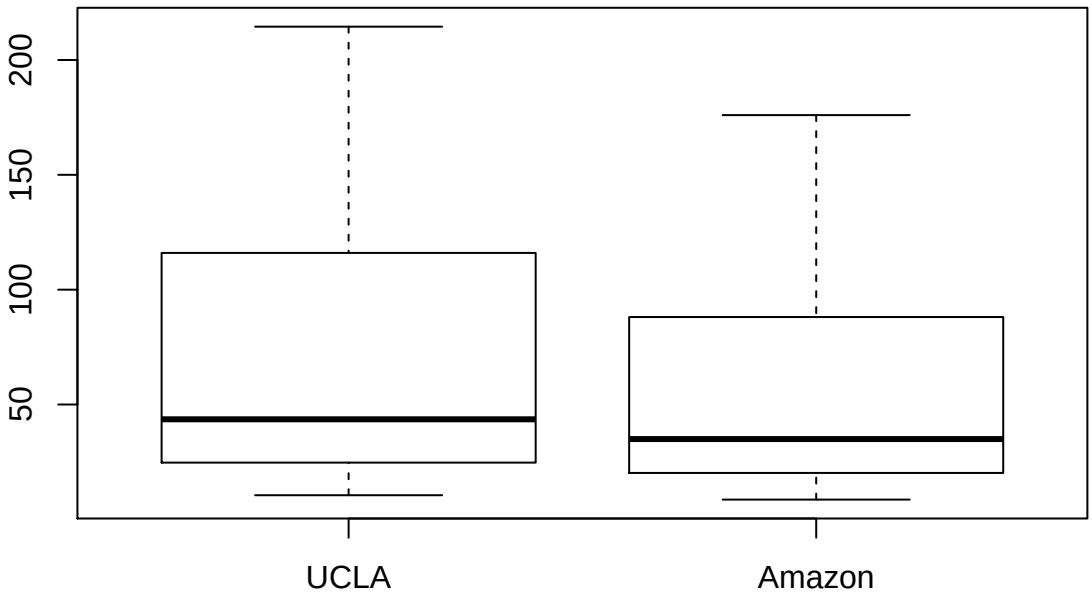
```
##
##  Welch Two Sample t-test
##
## data:  AgeAndNonpremie and AgeAndPremie
## t = 0.21913, df = 201.9, p-value = 0.8268
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.999804  1.249804
## sample estimates:
## mean of x mean of y
##    27.000    26.875
```

```
#Alternatively, we could do >t.test(ncbirths$mage ~ ncbirths$premie) #instead
```

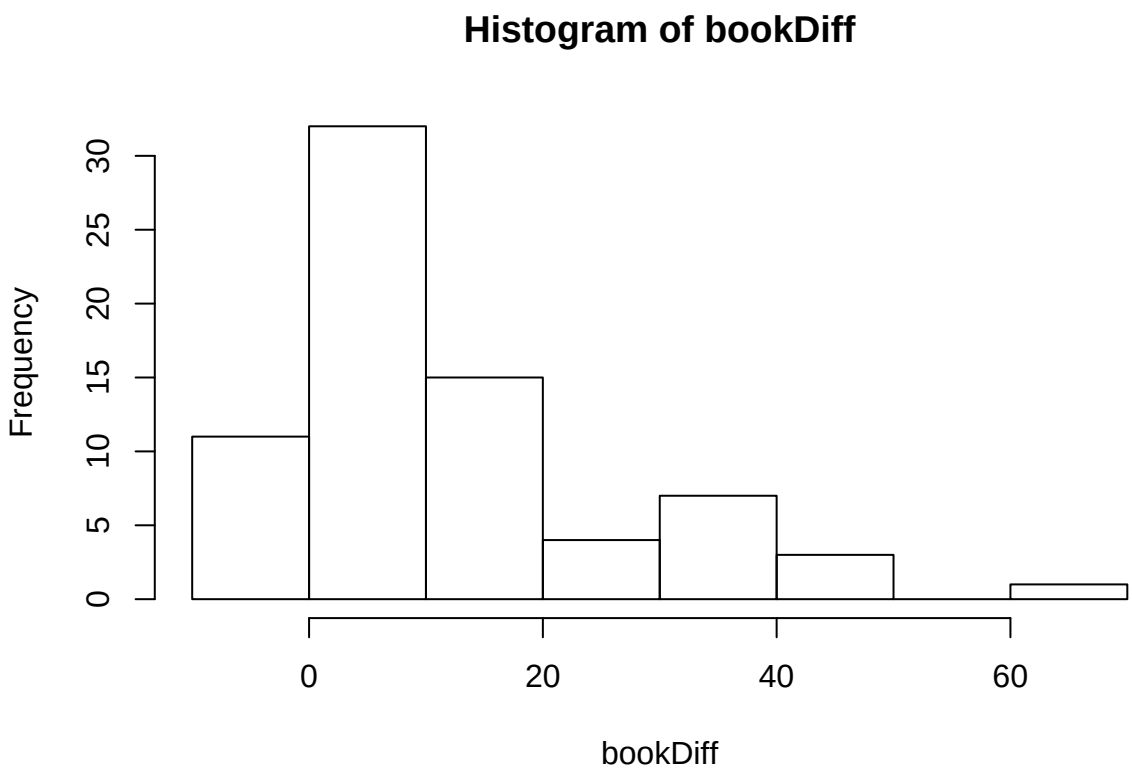
With a p-value of .8268 and a confidence interval that includes 0, we conclude that a mother's age does affect whether or not her baby is born prematurely (a "premie")

Exercise 5

```
data('textbooks')
#a)
boxplot(textbooks$uclaNew, textbooks$amazNew, names=c('UCLA', 'Amazon'))
```




```
#b)
bookDiff=(textbooks$uclaNew - textbooks$amazNew)
hist(bookDiff)
```




```
mean(bookDiff)
```

```
## [1] 12.76164
```

UCLA is, on average, $12.76 more expensive than Amazon for the same new textbook

```
#c)
t.test(bookDiff, alternative=c('greater'))
```

```
##
##  One Sample t-test
##
## data:  bookDiff
## t = 7.6488, df = 72, p-value = 3.464e-11
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  9.981505      Inf
## sample estimates:
## mean of x
##  12.76164
```

With a p-value $<.0001$, a t-value of 7.65 on 72 degrees freedom, and a confidence interval for the differences having a lower bound of 9.98, we reject the null hypothesis (that the book prices are equal), and thus conclude that the UCLA textbook prices are higher than the Amazon textbook prices for the same book when new

We assume that the textbooks were chosen at random, and that the differences are normally distributed (which is satisfied by the CLT, as we have a large enough sample size for the CLT to apply)

Note that the output above matches the output if we were to perform a paired t-test instead

```
t.test(textbooks$uclaNew, textbooks$amazNew, paired = TRUE, alternative = c('greater'))
```