Group Lab Assignment

Peter Greene, Courtney Jones

Dec 11, 2018

#### Introduction

We were curious to see if after graduating from Biola University with an undergraduate degree, would we be able to afford a house off of a starting salary. After looking at how taxes have increased across the state of California we began to research other states housing prices and taxes. We decided to see a comparison between West Coast and East Coast states and chose Tennessee as the state to compare with California. After comparing both states, we noticed that there were major differences in incomes and taxes. Testing what it would be like to be a single person upon graduation, we chose the taxes for a single income. With a federal tax of 22% across the nation and a state tax in California at 8% and Tennessee at 0%, we started to plot out our estimations. Our hypothesis was that a house in California would be more expensive than Tennessee but that earnings would also be more in California than in Tennessee. Upon observing the data collected on housing, it became very obvious that a house in California would be more. Through calculations we found that a student straight out of school could afford a $163,000 house in La Mirada and a $147,000 house in Knoxville only using up 25% of their take home pay on a 15 year loan. We were curious if that amount would fall into a confidence interval of 95% in the observed cities. Unfortunately, we discovered that a student on a starting salary will not be able to afford and only use up 25% of their take home pay on a 15 year loan in either city. —

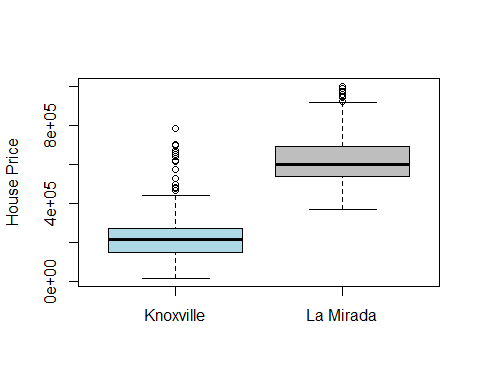
#### Methods:

The boxplots shown in the outcome of our data prove very differential data.

The mean list price of a house in La Mirada is significantly larger than the mean listing price in Knoxville. The standard deviation of these list prices also differs by around $11,000, while the sample sizes were virtually the same. The boxplot below shows that the listing prices of California houses is much greater than those in Tennessee, which will help contribute to our hypothesis by giving us a good view of which state will be more practical to find reasonable housing prices.

|  |  |  |  |
| --- | --- | --- | --- |
| Subdivision | Mean of List Price | Standard Deviation of List Price | Sample Size |
| La Mirada | *$599,000* | *$133,020* | *324* |
| Knoxville | *$215,000* | *$122,337* | *342* |

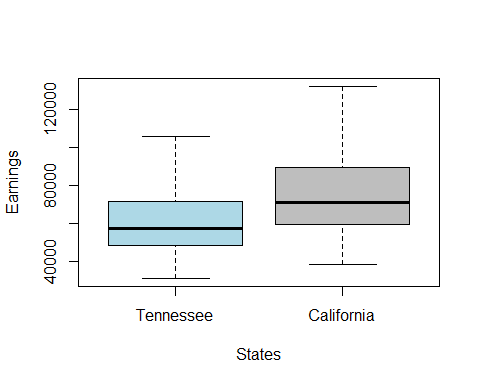
boxplot(Houses\_Knoxville$PRICE, Houses\_La\_Mirada\_csv$PRICE, title = "A Boxplot for Housing", names = c("Knoxville", "La Mirada"), col = c("lightblue", "gray"), ylab = "House Price")



This boxplot shows the mean degree earning in both La Mirada and Knoxville. The mean degree earning of someone just out of college is much larger in La Mirada than Knoxville. The standard deviation of their earnings differs by around $4,000, while their sample size is the exact same at 54. The boxplot shown is able to give us the information on the average degree earnings, which helps get us to our hypothesis by seeing how much income we will have to spend on a house.

|  |  |  |  |
| --- | --- | --- | --- |
| City | Mean of Degree Earnings | Standard Deviation of Degree Earnings | Sample Size |
| La Mirada | *$74,694* | *$21,092* | *54* |
| Knoxville | *$60,502* | *$17,018* | *54* |

boxplot(Biola\_Degrees\_and\_earnings$`Knox Pay`,Biola\_Degrees\_and\_earnings$`LA Pay`, title = "A Boxplot For Earnings", names = c("Tennessee", "California"), col = c("lightblue", "gray"), ylab = "Earnings", xlab = "States")



Our Hypothesis was that the earnings and the housing would be much more in California than in Knoxville. Our hypothesis was correct when we found a P-Value of 0.0001 in earnings and a P-Value of 2.2e-16 in housing comparisons. Both of which are less than alpha 0.05. Thus we rejected the H0 and accepted the HA that earnings and housing prices are more in California.

Taking the mean earnings of BIOLA degrees, we then subtracted the income federal tax of 22% and 8% from the California income for the state tax of. With that amount of take home pay we found what 25% of that number would be and divided it by 12 to discover what the monthly pay would be. Using those numbers we discovered the price amount able to be afforded by a starting salary.

taxes <- 60502 \* .22  
Subtract\_taxes <- 60502 - 13310.44  
Find\_25Percent\_Take\_Home <- 47191.56\*.25  
Monthly\_Take\_Home\_Knox <- 11797.89/12  
983.15

## [1] 983.15

The 25% of a monthly Take home pay on a $60,502 salary in Knoxville is $983.15. With this Salary one can afford a $147,000 house. Will this house price be in an interval of the data with a 95% confidence?

t.test(Knox\_Mira\_Houses$Knox\_PRICE)

##   
## One Sample t-test  
##   
## data: Knox\_Mira\_Houses$Knox\_PRICE  
## t = 34.927, df = 341, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 218039.2 244062.8  
## sample estimates:  
## mean of x   
## 231051

With the interval between $218,039 and $244,062 we would not be able to afford the average house in Knoxville

taxes <- 74694 \* .30  
Subtract\_taxes <- 74694 - 22408  
Find\_25Percent\_Take\_Home <- 52286 \* .25  
Monthly\_Take\_Home\_LaMirada <- 13071.5/12  
1089.29

## [1] 1089.29

The 25% of a monthly take home pay on a $74,694 salary in California is $1089.29. With this salary one can afford a $163,000 house. Will this house price be in an interval of the data with a 95% confidence?

t.test(Knox\_Mira\_Houses$Mira\_PRICE)

##   
## One Sample t-test  
##   
## data: Knox\_Mira\_Houses$Mira\_PRICE  
## t = 85.172, df = 323, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 614882.9 643960.2  
## sample estimates:  
## mean of x   
## 629421.6

With the interval between $614,882 and 643,960 we would not be able to afford the average house in La Mirada.

#### Conclusions/Discussions:

In conclusion, it would be significantly more difficult to afford a house at a reasonable price in La Mirada, California, than it would be in Knoxville, Tennessee. Although after analyzing deeper we came to realize that it would be very unlikely that you could afford a house, on a starting salary, in either state. However saddening this may sound, this is off of only 25% of the students take home pay on a 15 year house loan. The numbers would change using a 30 year loan. Also, a starting salary is just that, starting. A raise will soon happen and hopefully the student will have saved a good down payment at the time of the raise, which will make the house much more affordable.

#### ——————— Appendices ———————

library(mosaic)  
library(readxl)  
Biola\_Degrees\_and\_earnings <- read\_excel("C:/Users/Peter Greene/Desktop/Biola Fall 2018/Prob and Stat/Group Project/Biola Degrees and earnings.xlsx")  
Houses\_Knoxville <- read\_excel("C:/Users/Peter Greene/Desktop/Biola Fall 2018/Prob and Stat/Group Project/Houses Knoxville.xlsx")  
Houses\_La\_Mirada\_csv <- read\_excel("C:/Users/Peter Greene/Desktop/Biola Fall 2018/Prob and Stat/Group Project/Houses La Mirada.csv.xlsx")  
Knox\_Mira\_Houses <- read\_excel("C:/Users/Peter Greene/Desktop/Biola Fall 2018/Prob and Stat/Group Project/Knox-Mira-Houses.xlsx")  
summary(Knox\_Mira\_Houses)

## Knox\_PRICE Mira\_PRICE   
## Min. : 18000 Min. :368888   
## 1st Qu.:149750 1st Qu.:539000   
## Median :215000 Median :599000   
## Mean :231051 Mean :629422   
## 3rd Qu.:274907 3rd Qu.:688275   
## Max. :782500 Max. :999000   
## NA's :18

summary(Houses\_Knoxville)

## PROPERTY TYPE ADDRESS CITY   
## Length:342 Length:342 Length:342   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## STATE ZIP PRICE BEDS   
## Length:342 Min. :37721 Min. : 18000 Min. :1.000   
## Class :character 1st Qu.:37914 1st Qu.:149750 1st Qu.:3.000   
## Mode :character Median :37919 Median :215000 Median :3.000   
## Mean :37908 Mean :231051 Mean :3.237   
## 3rd Qu.:37922 3rd Qu.:274907 3rd Qu.:4.000   
## Max. :37938 Max. :782500 Max. :6.000   
##   
## BATHS SQUARE FEET LOT SIZE YEAR BUILT   
## Min. :1.000 Min. : 576 Min. : 6098 Min. :1899   
## 1st Qu.:2.000 1st Qu.:1373 1st Qu.: 17859 1st Qu.:1954   
## Median :2.500 Median :1998 Median : 27878 Median :1979   
## Mean :2.273 Mean :2073 Mean : 58013 Mean :1977   
## 3rd Qu.:2.500 3rd Qu.:2587 3rd Qu.: 54776 3rd Qu.:2004   
## Max. :5.000 Max. :6230 Max. :827640 Max. :2019   
## NA's :236   
## $/SQUARE FEET   
## Min. : 18.0   
## 1st Qu.: 97.0   
## Median :111.0   
## Mean :111.6   
## 3rd Qu.:126.0   
## Max. :355.0   
##

summary(Houses\_La\_Mirada\_csv)

## PROPERTY TYPE ADDRESS CITY   
## Length:324 Length:324 Length:324   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## STATE ZIP PRICE BEDS   
## Length:324 Min. :90240 Min. :368888 Min. :2.000   
## Class :character 1st Qu.:90620 1st Qu.:539000 1st Qu.:3.000   
## Mode :character Median :90650 Median :599000 Median :3.000   
## Mean :91247 Mean :629422 Mean :3.395   
## 3rd Qu.:92801 3rd Qu.:688275 3rd Qu.:4.000   
## Max. :92835 Max. :999000 Max. :6.000   
## BATHS SQUARE FEET LOT SIZE YEAR BUILT   
## Min. :1.000 Min. : 734 Min. : 1500 Min. :1909   
## 1st Qu.:2.000 1st Qu.:1260 1st Qu.: 5648 1st Qu.:1953   
## Median :2.000 Median :1477 Median : 6204 Median :1956   
## Mean :2.059 Mean :1606 Mean : 7187 Mean :1958   
## 3rd Qu.:2.000 3rd Qu.:1851 3rd Qu.: 7350 3rd Qu.:1961   
## Max. :5.500 Max. :3976 Max. :87405 Max. :2018   
## $/SQUARE FEET   
## Min. :240.0   
## 1st Qu.:358.8   
## Median :404.5   
## Mean :409.1   
## 3rd Qu.:457.0   
## Max. :682.0

sd(Houses\_Knoxville$PRICE)

## [1] 122337

sd(Houses\_La\_Mirada\_csv$PRICE)

## [1] 133020.2

summary(Biola\_Degrees\_and\_earnings)

## Degree LA Pay Knox Pay   
## Length:54 Min. : 38220 Min. : 31068   
## Class :character 1st Qu.: 59648 1st Qu.: 48258   
## Mode :character Median : 71170 Median : 57071   
## Mean : 74694 Mean : 60502   
## 3rd Qu.: 89068 3rd Qu.: 71424   
## Max. :131956 Max. :105816

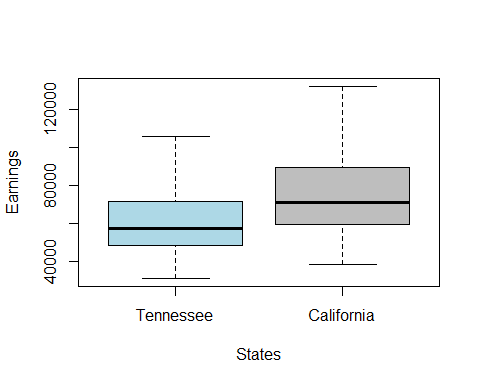
sd(Biola\_Degrees\_and\_earnings$`LA Pay`)

## [1] 21092.69

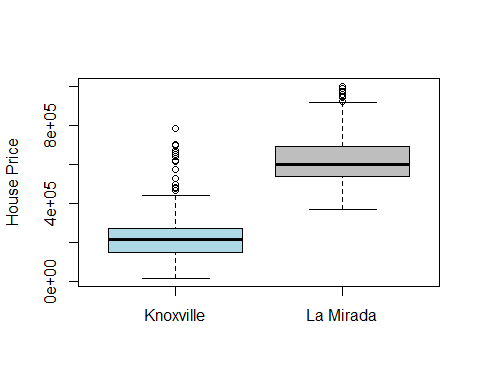
sd(Biola\_Degrees\_and\_earnings$`Knox Pay`)

## [1] 17018.1

boxplot(Biola\_Degrees\_and\_earnings$`Knox Pay`,Biola\_Degrees\_and\_earnings$`LA Pay`, title = "A Boxplot For Earnings", names = c("Tennessee", "California"), col = c("lightblue", "gray"), ylab = "Earnings", xlab = "States")



boxplot(Houses\_Knoxville$PRICE, Houses\_La\_Mirada\_csv$PRICE, title = "A Boxplot for Housing", names = c("Knoxville", "La Mirada"), col = c("lightblue", "gray"), ylab = "House Price")



attach(Biola\_Degrees\_and\_earnings)  
t.test(`LA Pay`,`Knox Pay`,var.equal = TRUE, alternative = "greater")

##   
## Two Sample t-test  
##   
## data: LA Pay and Knox Pay  
## t = 3.848, df = 106, p-value = 0.0001019  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 8072.035 Inf  
## sample estimates:  
## mean of x mean of y   
## 74694.07 60502.15

attach(Knox\_Mira\_Houses)  
t.test(Mira\_PRICE,Knox\_PRICE)

##   
## Welch Two Sample t-test  
##   
## data: Mira\_PRICE and Knox\_PRICE  
## t = 40.165, df = 651.69, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 378894.8 417846.3  
## sample estimates:  
## mean of x mean of y   
## 629421.6 231051.0

60502 \* .22

## [1] 13310.44

60502 - 13310.44

## [1] 47191.56

47191.56\*.25

## [1] 11797.89

11797.89/12

## [1] 983.1575

t.test(Knox\_Mira\_Houses$Knox\_PRICE)

##   
## One Sample t-test  
##   
## data: Knox\_Mira\_Houses$Knox\_PRICE  
## t = 34.927, df = 341, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 218039.2 244062.8  
## sample estimates:  
## mean of x   
## 231051

74694 \* .30

## [1] 22408.2

74694 - 22408

## [1] 52286

52286 \* .25

## [1] 13071.5

13071.5/12

## [1] 1089.292

t.test(Knox\_Mira\_Houses$Mira\_PRICE)

##   
## One Sample t-test  
##   
## data: Knox\_Mira\_Houses$Mira\_PRICE  
## t = 85.172, df = 323, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 614882.9 643960.2  
## sample estimates:  
## mean of x   
## 629421.6