Project 1

Your names here

2023-09-05

- Note: for each question, be sure to include both R code and output that is pertinent to your answer(s).
- Use hprice in the faraway package.
- The data include 324 observations from 36 US metropolitan statistical areas (MSAs) over 9 years from 1986-1994: 36 x 9 = 324
- Assume that the MSAs in the data are a simple random sample of the population of MSAs in the US. See https://www2.census.gov/geo/maps/metroarea/us_wall/Mar2020/CBSA_WallMap_Mar2020.pdf for MSAs
- Refer to the R manual for faraway (on Canvas - see Module 1) for background information about this dataset as well as variable definitions.
- The housing sale price is the outcome variable of interest. Because the dataset has a natural log transformed price variable, narsp, we will recode this to create a variable called "homeprice" by transforming narsp back to the dollar unit for an easier interpretation as follows:

```
hprice$homeprice <- exp(hprice$narsp)*1000
```

1. What are the mean and the variance of homeprice? What do they mean?

```
mean(hprice$homeprice)
```

```
## [1] 94411.42
```

```
var(hprice$homeprice)
```

```
## [1] 1583110349
```

#the mean is the average housing sale price #the variance is the squared standard deviation and shows the spread from the mean

2. Construct a 95% confidence interval of the average homeprice. What does the confidence interval imply?

```
t.test(hprice$homeprice, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  hprice$homeprice
## t = 42.711, df = 323, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  90062.70 98760.14
## sample estimates:
## mean of x
##  94411.42
```

We can estimate that the true mean homeprice falls between 90062.70 and 98760.14 95% of the time

3. Estimate the average homeprice by whether the MSA was adjacent to a coastline, noted in the variable ajwtr, and the standard errors.

```
library(dplyr)
std.error <-function(x) sd(x)/sqrt(length(x))
hprice%>%
  group_by(ajwtr)%>%
  summarize(m=mean(homeprice),
            stderror=std.error(homeprice))
```

```
## # A tibble: 2 x 3
##   ajwtr       m stderror
##   <fct>   <dbl>    <dbl>
## 1 0      82389.    1229.
## 2 1     111243.    4656.
```

4. Test the difference in homeprice between coastline MSAs and non-coastline MSAs. Clearly state the formula for the null hypothesis, the test method, and your rationale for selecting the method. What do you conclude about the hypothesis?

```
#use t test to determine if mean homeprice in coastline and non-coastline MSAs is the same
#null- mean homeprice of coastal cities= mean homeprice of non-coastal cities

#test equal variance
#null- Variance of homeprice in coastal cities= variance of homeprice in non costal cities
var.test(homeprice ~ ajwtr, hprice, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  homeprice by ajwtr
## F = 0.097496, num df = 188, denom df = 134, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.07088604 0.13297389
```

```
## sample estimates:
## ratio of variances
##         0.09749617
```

```
#reject null

#use t-test
#null- mean homeprice of costal cities= mean homeprice of non-coastal cities
t.test(homeprice ~ ajwtr, hprice, var.equal=FALSE, conf.int = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  homeprice by ajwtr
## t = -5.9922, df = 152.79, p-value = 1.43e-08
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -38367.19 -19340.96
## sample estimates:
## mean in group 0 mean in group 1
##        82388.89       111242.96
```

```
#reject the null
```

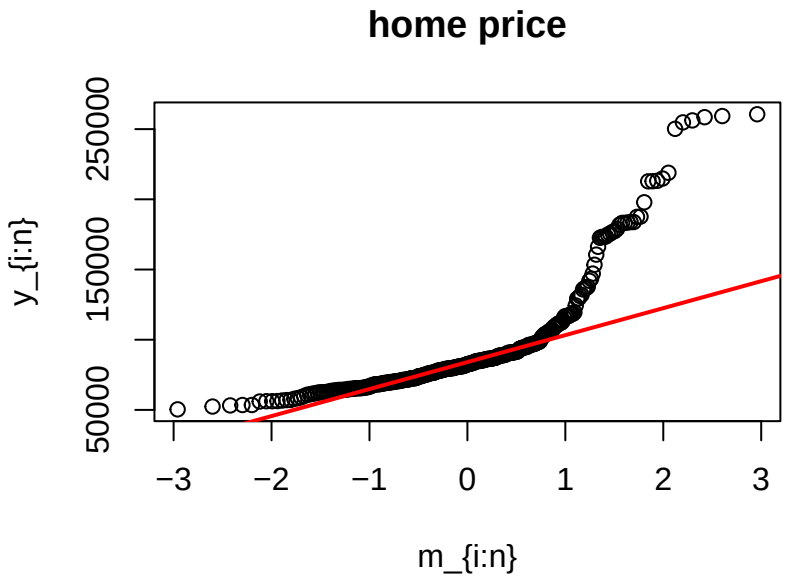
5. Estimate the Pearson correlation coefficient between homeprice and per capita income of the MSA for a given year, noted in the variable ypc.

```
cor(hprice$homeprice, hprice$ypc, method = 'pearson')
```

```
## [1] 0.7437474
```

6. Test whether the correlation coefficient between homeprice and ypc is 0, or not. Clearly state the null hypothesis being tested and include the formula.
7. Can you say that per capita income has an effect on the home sales price using the results from #6? Why or why not?
8. Test the normality of homeprice. Would this change your responses to questions 1-7? Why or why not?

```
library(ggplot2)
qqnorm(hprice$homeprice, main="home price", ylab="y_{i:n}", xlab="m_{i:n}")
qqline(hprice$homeprice, col="red",lwd=2)
```



#homeprice does not follow a normal distribution, therefore answers 1-7 are not accurate. The sample means are innacurate to the true mean