HW1

Mauro Gonzalez

2025-09-06

Q1

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

mean(hprice$homeprice)
```

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

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

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

```
#This means that the average home price in the dataset is $94,411.
#However, there is a wide range of home prices:
#over 1.5 billion (1583110349)
```

Q2

```
t.test(hprice$homeprice)
```

```
##
##  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
```

```
#The confidence interval is [90062,98760.14]
#Meaning that a vast majority of home prices will fall in between that range.
```

Q3

```r
mean(hprice$homeprice [hprice$ajwtr==1])
```

```
## [1] 111243
```

```r
std.error(hprice$homeprice [hprice$ajwtr==1])
```

```
## [1] 4655.883
```

Q4

```r
msacoast <- mean(hprice$homeprice [hprice$ajwtr==1])
msainland <- mean(hprice$homeprice [hprice$ajwtr==0])
msacoast - msainland
```

```
## [1] 28854.07
```

```r
#This question requires taking the mean price of coastline-adjacenet
#and inland MSA home prices, and then subtracting one from the other.
#I hypothesize that coastal MSA home prices will be higher.
#This hypothesis is proven true:
#subtracting inland prices from coastal prices gives a $28,854.07 difference.
```

Q5

```r
cor(hprice$homeprice, hprice$ypc, method=c("pearson"))
```

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

Q6

```r
cor(hprice$homeprice, hprice$ypc)
```

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

```r
#There is a strong positive correlation between home price and per captia income.
```

Q7

```r
#There is an effect of per capita income on home sale price.
#In a correlation, x's effect on y is equal to y's effect on x
#(in this case, with a high coefficient score)
```

Q8

```r
shapiro.test(hprice$homeprice)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  hprice$homeprice
## W = 0.7327, p-value < 2.2e-16
```

```
#This would not change my responses.
#While the data is evidently not normally distributed, this does not
#necessarily change any previous answers.
#However it does mean that measures like "mean" and "variance" are less useful than median.
```