**Statistical Analysis (I): Estimation & Testing**

**Homework Assignment 2**

**PART A**

1. a) Your friend claims that the average house price in this area is above $150K. Do you agree? Briefly explain what the *p-values* in these cases mean?
2. He also claims that the average living area is more than 1800 Sq. Ft. Do you agree with this? (Use a 5% significance level for both.). Briefly explain what the *p-values* in these cases mean?

**Solution:**

Let be the mean house prices in Dollars and be the mean living area in Sq.ft

Consider the first hypothesis

Since the population variance is not known, our test statistic is:

The sample information we have is:

163862.1, under null hypothesis = 150000, s=67651.56, n =1047

The calculated t=6.63, the p-value=1-pt(t,n-1)=1-pt(6.63,1046)= 2.684541e-11.

The p-value (2.684541e-11) is less than our (0.05), so we reject the null and agree with friend’s claim.

**1(b) Solution:**

The next hypothesis is

Since the population variance is not known, our test statistics is:

The sample information we have is:

1807.303, under null hypothesis = 1800, s= 641.4609, n =1047

The calculated t= 0.3683, the p-value=1-pt(t,n-1)=1-pt(0.3683,1046)= 0.356

The p-value (0.356) is greater than our (0.05), so we fail to reject the null and dis-agree with friend’s claim.

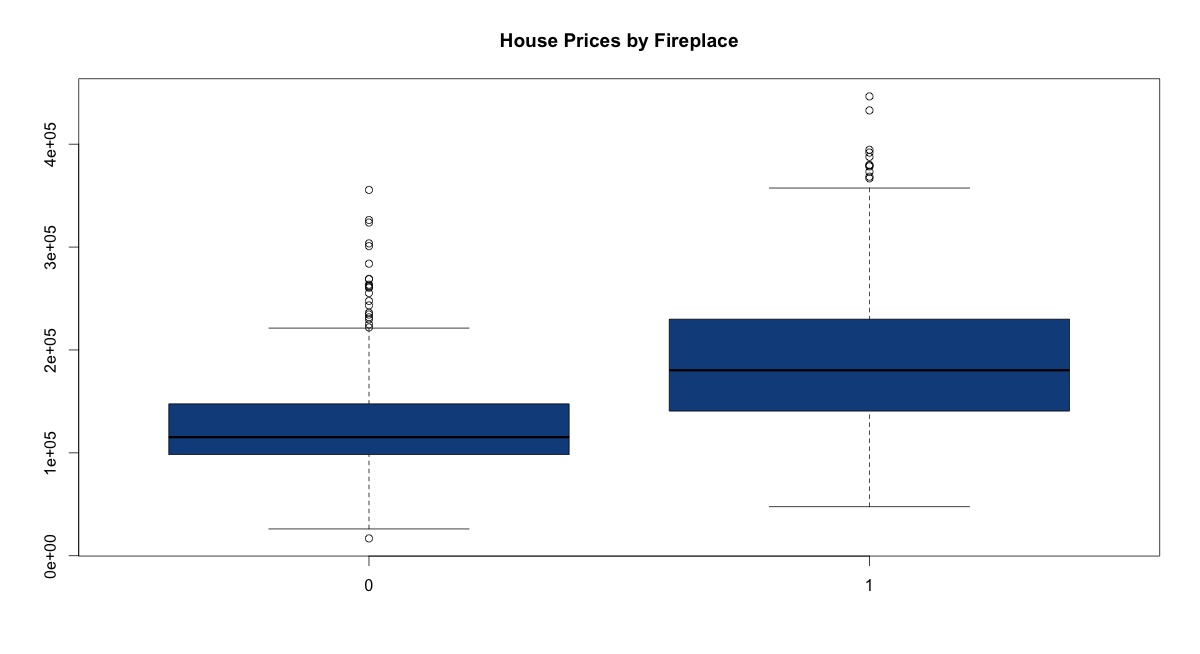
Q2. Are the home prices higher for houses with fireplaces as compared to those without?

a) Create side-by-side box plots of the house prices of the two groups and comment them.

b) Formulate an appropriate hypothesis and test it in order to check the above claim.Assume that the population standard deviations of house prices in the two groups are equal.

1. **Solution:**

The boxplots:



From the above box-plots we can see that the median house prices are higher for houses with a fireplace.

**Solution (b):**

Let be the mean house prices in Dollars for houses **with** a fireplace and be the mean house prices in Dollars for houses **without** a fireplace.

Our Hypothesis:

Assuming that the population variances are equal, our test statistics is:

where

The sample information we have is:

, ,

We have = 60089.52.

The calculated t= 16.758, the p-value=1-pt(t,1047-2)=1-pt(16.758,1045)= 0

The p-value (0) is less than our (0.05), so we reject the null and conclude that the houses with fireplaces have higher average prices than those without fireplaces.

Q3. Any house aged more than 30 years is considered an “old” house. Your friend

claims that old houses have larger lot sizes than new houses. Do you agree? Explain.

Use a significance level of 5% for your test. Historical data suggests that old houses

include some very large and some very small lot sizes but new houses are more

homogeneous in their lot sizes.

**Solution:**

Let be the mean lot size for old houses and be the mean lot size for new houses

Our Hypothesis:.

Assuming that the population variances are different, our test statistics is:

The sample information we have is:

, ,

The calculated t= ()/ sqrt()

t= -0.5888

degrees of freedom=



= (0.7249)2 /302 = 0.00174

= (0.7986)2/745 = 0.0008560563

df= 6.73951E-06/ 1.1E-08 = 610.2713 =approx. 610.

The p-value=1-pt(t,610)=1-pt(-0.5888,610)= 0.7219

The p-value (0.7219) is greater than our (0.05), so we fail to reject the null and maintain out status quo that lot sizes of old houses are less than those of the new ones.

Q4. Based on the evidence available here, would you be willing to claim that fireplaces

have become more fashionable? For simplicity, its OK to compare only “new” houses and

“old” houses. Use a significance level of 5% for your test. Use a significance level of 5% for your test.

**Solution:**

Let be the proportion of old houses with fireplaces and be the proportion of new houses with fireplaces.

Our hypothesis now becomes a two-sided one:

Test statistic is:



The sample information we have is:

no = 302 and nn = 745

po =135/302 =0.447, pn =486/745=0.652

Sample pooled proportion is: = 0.593

Therefore,

z = = = -6.12

P-value = P (P1-P2 < -0.205;

= pnorm(-6.12)

= 4.678768e-10

Since p-value < α-value, hence we reject H0.

We can say that the proportion of old houses with fire places are more than those for new ones.

**PART B**

1. Suppose that houses with 1-2 bedrooms are considered to be “Small Houses”, those with 3-4 are “Medium Houses” and 5-6 as “Big Houses”. Can we conclude that the prices of Small, Medium and Big houses are significantly different, at 1% level of significance? (2marks)

**Solution:**

Let be the mean price for small houses, be the mean price for medium houses and be the mean price for large houses.

We want to test the following:

**Null hypothesis:**

i.e. the mean prices of small, medium and large houses are the same

**Alternative hypothesis:**

i.e. not all s are equal

We perform a one-way ANOVA of Prices on Bedrooms to test the above hypothesis. The results from the test are:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| aov.model<-aov(Price~factor(Bedrooms)) | | | | | | |
| summary(aov.model) | | | | | | |
|  | **Df** | **Sum Sq** | **Mean Sq** | **F value** | **Pr(>F)** |  |
| **factor(Bedrooms)** | 2 | 4.84E+11 | 2.42E+11 | 58.71 | <2e-16 | \*\*\* |
| **Residuals** | 1044 | 4.30E+12 | 4.12E+09 |  |  |  |
|  |  |  |  |  |  |  |
| Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1 | | | | | | |

We see that the p-value is less than our level of significance. We reject the null hypothesis that the mean prices of small, medium and large houses are the same.

The contents of the ANOVA table can be interpreted easily as per the classroom sessions.