HYPOTHESES TESTING

"This document provides an information to houses sold in City of Melbourne . It includes all the details of the different types of houses sold in different years . We have used the variable name as 'housing.dataset' in order to read our csv file and store in the dataframe . We will use different variables for Hypotheses Testing."

Introduction

The Housing dataset consist of eight attributes almost 35,000 houses in the city of melbourne . Few data's having incorrect values and missing values . This is the summary of the houses sold in city of Melbourne . Housing dataset after removing missing values and incorrect values .

- Removed NA values
- Removed values having empty landsize and building area
- Removed values where building area exceeding land area
- Removed values of building less than eleven meters(minimum building size is 11 meters).REFERENCE
- Removed incorrect values where date of the house was sold before the house was build

---

Problem-1

Two types of houses namely townhouse and duplex are compared for their prices. Two independent samples of 500 houses were selected. Summary results were as follows:

	townhouse	duplex
sample size	500	500
sample mean	909920.4	613756.8
sample variance	159962788568	57199360507

Using a 0.05 significance level, is there evidence that the two types of houses differ in their mean price?

Solution

1st step

Null Hypotheses-> H_0 : $u_1 = u_2$

Alternate Hypotheses-> H_1 : $u_1 != u_2$

2nd step

The null hypothesis of the two-tailed test about population proportion can be expressed as follows:

$p = p_0$ ‘ where p_0 is a hypothesized value of the true population proportion p.

3rd step

At .05 significance level

4th step

If z > 1.96, We reject null hypotheses.

If z < -1.96, We reject null hypotheses.

5th step

Let us calculate the test statistics

```
z=(909920.4-613756.8)/sqrt((159962788568/100)+(57199360507/100))
print(z)
```

```
## [1] 6.355352
```

6th step

Reject H_0, we have sufficient evidence to conclude u_1 != u_2

i.e. the mean prices of two different types of houses differ

Confidence interval

```
a=(909920.4-613756.8)+(1.96)*sqrt((159962788568/100)+(57199360507/100))
b=(909920.4-613756.8)-(1.96)*sqrt((159962788568/100)+(57199360507/100))
```

204826.3 to 387500.9

We are 95% confident that the mean price of two different types of houses is between 204826.3 and 387500.9. These evidence shows us to reject the null hyotheses.

Problem2

The real estate agent claims that the mean of landsize of houses in city of melbourne is equal or more than 607 meters. In a sample of 30 houses, it was found that the mean of those samples is 321.6 meters. Assume the standard deviation is 1128.851. At .05 significance level,can we reject the null hypotheses that the mean of landsize is lesser than 321.6 meters?

Solution

We can state the hypotheses as:

Null Hypotheses-> H_0 : u >= 607

Alternate Hypotheses-> H_1 : u < 607

2nd step

The null hypothesis of the lower tail test of the population mean can be expressed as follows:

u >= u_0 where u_0 is a hypothesized lower bound of the true population mean u.

3rd step

At .05 significance level

4th step

If z < -2.58, We reject null hypotheses.

5th step

Let us calculate the test statistics

```
t=(321.6-607)/(1128.851/sqrt(30))
print(t)
```

```
## [1] -1.384771
```

6th step

we fail to reject H_0 Null hypotheses, we have sufficient evidence to conclude H_0: u >= 607)

critical value

We then compute the critical value at .05 significance level.

```
alpha=0.05
t.alpha=qt(1-alpha,df=30-1)
print(-t.alpha)
```

```
## [1] -1.699127
```

The test statistic -1.384771 is greater than the critical value of -1.699127. Hence, at .05 significance level, we fail to reject the null hypotheses.

Problem3

The mean of number of car parking in the year 2016 was 1.634558. In the sample of 40 houses in the year of 2017, the mean of car parking is 1.45. At the standard deviation 0.9412223. And at significance level at .05, can we reject the null hypotheses that the mean number of car parking does not differ from last year?

Solution

We can state the hypotheses as:

Null Hypotheses-> H_0 : u = 1.634558

Alternate Hypotheses-> H_1 : u != 1.634558

2nd step

The null hypothesis of the two-tailed test about population proportion can be expressed as

follows:

$p = p_0$

where p_0 is a hypothesized value of the true population proportion p.

3rd step

At .05 significance level

4th step

If $z > 1.96$ and $z < -1.96$, We reject null hypotheses.

5th step

Let us calculate the test statistics

```
z=(1.45-1.634558)/(0.9412223/sqrt(40))
print(z)
```

```
## [1] -1.24014
```

Critical interval

```
alpha=0.05
z.half.alpha=qnorm(1-(alpha/2))
c(-z.half.alpha,z.half.alpha)
```

```
## [1] -1.959964  1.959964
```

The test statistic -1.24014 lies between the critical values -1.959964 and 1.959964. Hence,

at .05 significance level, we fail to reject the null hypothesis that the mean number of car parking does not differ from last year.

Alternative Solution

```r
pval=2*pnorm(z)
pval
```

```
## [1] 0.2149237
```

Instead of using the critical value, we apply the pnorm function to compute the two-tailed

p-value of the test statistic. It doubles the lower tail p-value as the sample mean is less

than the hypothesized value. Since it turns out to be greater than the .05 significance

level, we fail to reject the null hypothesis.

Problem4

The price of the house on the city of melbourne have a mean of 1152201 or lesser, In a sample of 37 houses, It is found that the prices of houses are 1579838, Assume that sd is 693401.4,At 0.05 significance level ,can we reject null hypotheses?

Solution

We can state the hypotheses as:

Null Hypotheses-> H_0 : u <= 1152201

Alternate Hypotheses-> H_1 : u > 1152201

2nd step

The null hypothesis of the upper tail test of the population mean can be expressed as

follows:

u <= u_0

where u_0 is a hypothesized upper bound of the true population mean u.

3rd step

At .05 significance level

4th step

If z > 2.58, We reject null hypotheses.

5th step

Let us calculate the test statistics

```
t=(1579838-1152201)/(693401.4/sqrt(37))
print(t)
```

```
## [1] 3.751383
```

6th step

Reject H_0, we have sufficient evidence to conclude H_0: u > 607)

i.e. the mean price is greater than population mean.

critical value

We then compute the critical value at .05 significance level.

```
alpha=0.05
z.alpha=qnorm(1-alpha)
print(z.alpha)
```

```
## [1] 1.644854
```

The test statistic 3.751383 is greater than the critical value of 1.644854. Hence, at .05 significance level, we reject the null hypotheses.
