INTERMEDIATE ANALYTICS

ALY6015, FALL 2019

MODULE 2 ASSIGNMENT

HYPOTHESIS TESTING

SUBMITTED BY: SHIVANI ADSAR

NUID: 001399374

CRN: 71933

SUBMITTED TO: PROF. LI, TENGLONG

DATE: 12/11/2019

Introduction

The assignment aims at conducting hypothesis tests on various datasets by using one sample T-tests, two-sample T-tests and F-test.

Analysis

A: Loading the MASS Library

We have used "library(MASS)" to install the MASS library package in R Studio.

B: One-Sample T-Test

We have used the "chem" dataset to test the hypothesis that the flour production company is producing whole meal flour with greater than 1 part per million copper in it. We have performed the One-sample t-test to decide the hypothesis. Considering,

Ho: true mean of flour and copper is lesser than and equal to 1

Ha: true mean of flour and copper is greater than 1

Using, t.test(chem, mu=1, alternative = "greater") for testing the hypothesis, the results are as below:

Observations: Since, the p-value (0.002952) is lesser than the significance level (0.05), we will be rejecting the null hypothesis of the true mean of flour and copper being lesser than and equal to 1. It can be concluded that, the company produces flour with greater than 1 part per million of copper.

```
        One Sample t-test

data:  chem
t = 3.0337, df = 23, p-value = 0.002952
alternative hypothesis: true mean is greater than 1
95 percent confidence interval:
 2.427162      Inf
sample estimates:
mean of x
 4.280417
```

Fig.1: Output of One-Sample T-Test

C: Two-Sample T-Test

We have used the "cats" dataset to perform hypothesis testing if the male and female cats have the same body weight.

In order to perform the hypothesis, we have used the subset function to separate the vectors.

male <-subset(cats, subset=(cats$Sex=="M"))

female <-subset(cats, subset=(cats$Sex=="F"))

We have performed the Two-sample t-test to decide the hypothesis. Considering,

Ho: Means of body weights are equal

Ha: Means of body weights are not equal

Using, t.test(male$Bwt, female$Bwt, alternative = "two.sided") for testing the hypothesis, results are as below:

Observations: Since, the p-value is lesser than the significance level (0.05), we will be rejecting the null hypothesis of the means of body weights of male and female cats being equal. It is concluded that, the body weights of male and female cats are not equal.

```
        Welch Two Sample t-test

data:  male$Bwt and female$Bwt
t = 8.7095, df = 136.84, p-value = 8.831e-15
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.4177242 0.6631268
sample estimates:
mean of x mean of y
 2.900000  2.359574
```

Fig.2: Output of Two-Sample T-Test

D: Paired T-Test

We have used the “shoes” dataset to perform hypothesis testing if material A wore better than material B.

We have performed the Paired t-test to decide the hypothesis. Considering,

Ho: μA-μB < =0

Ha: μA-μB > 0

Using, t.test(shoes$A,shoes$B, paired=TRUE, alternative = "greater") for testing the hypothesis, the results are as below:

Observations: Since, the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis of the difference in the means of Shoe A and Shoe B lesser than and equal to 0.

```
        Paired t-test

data:  shoes$A and shoes$B
t = -3.3489, df = 9, p-value = 0.9957
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.6344264        Inf
sample estimates:
mean of the differences
                 -0.41
```

Fig.3: Output of Paired T-Test

E: Test of Equal or Given Proportions

We have used the “bacteria” dataset to perform hypothesis testing if the drug treatment had a significant effect of the presence of the bacteria compared with the placebo. We have performed the Test of Equal Proportions to decide the hypothesis. Considering,

Ho: P < 0.5

Ha: P > 0.5

Using, prop.test(x=x,n=n,p=0.5,alternative='greater') for testing the hypothesis, the results are as below:

Observations: Since, the p-value is lesser than the significance level (0.05), we will reject the null hypothesis of the P value being lesser than 0.5. It can be concluded that the drug treatment has a significant effect on placebo.

```
        1-sample proportions test with continuity correction

data:  x out of n, null probability 0.5
X-squared = 80.405, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.7545481 1.0000000
sample estimates:
        p
0.8045455
```

Fig.4: Output of Test of Equal Proportions

F: F-test

We have used the “cats” dataset to perform hypothesis testing for the variance of the body weight in male and female cats. We have performed the F-Test to decide the hypothesis. Considering,

Ho: V1/V2 equal to 1

Ha: V1/V2 not equal to 1

Using, var.test(male$Bwt, female$Bwt) for testing the hypothesis, the results are as below:

Observations: Since, the p-value is lesser than the significance level (0.05), we will reject the null hypothesis, that the ratio of variances is equal to 1. It can be concluded that variances in the body weight of male and female cats is not equal.

```
        F test to compare two variances

data:  male$Bwt and female$Bwt
F = 2.9112, num df = 96, denom df = 46, p-value = 0.0001157
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.723106 4.703057
sample estimates:
ratio of variances
          2.911196
```

Fig.5: Output of F-Test

Conclusion

The hypothesis tests have been carried out on some of the datasets, and the analysis is used to decide the hypothesis.

References

1. Maindonald, J. H., & Braun, J. (2010). Data analysis and graphics using R: an example-based approach. Cambridge: Cambridge University Press.
2. (n.d.). Retrieved from https://homepage.divms.uiowa.edu/~luke/classes/STAT4580/histdens.html#density-plot-basics.