Mustafa Abdullayev

Inferential Statistics With R

AY6015 – Intermediate Analytics

Term: Winter 2020 - A

Instructor: Joe Manseau

01/20/2020

# Introduction

This is a Microsoft Word Report accompanying R Script. It contains my R code, outputs, my comments, and findings. My main aim is to utilize R and its statistical methods to carry out hypotheses testing. In my analysis, I used the cabbage dataset. This is built-in datasets in R. This data is collected from cabbage farm trials. Further explanations can be found [here](https://cran.r-project.org/web/packages/MASS/MASS.pdf). Data was cleaned, organized and ready for analysis. It has 4 variables and 60 observations. The main variable I am interested in is Cult: Factor giving the cultivar of cabbage. There are 2 cults: c39 and c52. I will carry out two different statistical tests. Firstly, I will conduct a T-test to see if the cultivar of cabbage has an effect on the head weight. Secondly, I will carry out an F-test to see if the variation of Ascorbic acid content is different between cultivars. Since I also provided R script with all the codes and comments, I removed some of the codes and comments from my report (such as package loading). It is due to keep my report brief, succinct and to the point.

# Effect of Cult on Head Weight.

This is the first part of my analysis. Here, I am curious if 2 differents cults of cabbage have the same mean head weight. The default choice is that they have the same weight since we do not know anything to the contrary yet. In order to test that, I will carry out a two-sample T-test. I will use the cabbage dataset. In this case, my alternative hypothesis is that mean head weight is different between different cultivars. More formally :

H0 : mean head weights are the same: M1 = M2

HA : mean head weights are not the same: M1 != M2

My test is two-sided because there is no direction. I am only concerned if there is an effect. Moreover, since this test is not extremely sensitive to errors (such as a drug test), using a 0.05 significance level is reasonable.

The P-value for this test (details below) is 0.004972, which is extremely small (smaller than the significance level). So, getting the current result, under the assumption that the null hypothesis is true, is extremely unlikely. Therefore, I can reject the null hypothesis, stating that the true difference in means is equal to zero, with 95% confidence. So there is evidence that a cultivar of cabbage has a significant effect on the head weight of cabbage.

Additionally, we can see that 95% confidence interval for true mean is [0.1971677, 1.0561656]. This interval does not contain 0. So, without any other test, with 95% confidence, I can conclude that average head weight for cult c39 is bigger than c52.

c39 <- cabbages[**which**(cabbages**$**Cult **==** "c39"),]  
c52 <- cabbages[**which**(cabbages**$**Cult **==** "c52"),]  
   
*# T-test : Do 2 different cultivars of Cabbages have same mean Head weights ?*  
*# H0 : they are same : M1 = M2*  
*# HA : they are not same : M1 != M2*  
*# significance = 0.05 (alpha)*   
*# my test is two sided*  
  
**t.test**(c39**$**HeadWt,c52**$**HeadWt,alternative = "two.sided")   
## Welch Two Sample t-test  
  
## data: c39$HeadWt and c52$HeadWt  
## t = 2.9209, df = 57.777, p-value = 0.004972  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.1971677 1.0561656  
## sample estimates:mean of x mean of y are 2.906667 and 2.280000  
*### my p-value is extremely small (0.004972).*  
*###I can reject the null hypothesis stating that the head weight of two different cultivars of cabbages are the same,with the 95% confidence*

# Effect of Cult on the variance of Ascorbic acid content.

This is the second part of my analysis. Here, I want to investigate the effect of the cult on the variation of ascorbic acid content. More precisely, I want to test the hypothesis if the cultivar of cabbage influences the variance of ascorbic acid content. The default choice is that they have the same variance since we do not know anything to the contrary yet. In order to test that, I will carry out an F-test. I will use the cabbage dataset. In this case, my alternative hypothesis is that the variation of ascorbic acid content in two different cultivars of cabbage is not the same. More formally :

H0 : the ratio of variances is equal to 1: V1 = V2

HA : the ratio of variances is not equal to 1: V1 != V2

My test is two-sided because there is no direction. I am only concerned if there is an effect. Moreover, since this test is not extremely sensitive to errors (such as a drug test), using a 0.05 significance level is reasonable.

The P-value for this test (details below) is 0.3613, which is not small (bigger than the significance level). So, getting the current result, under the assumption that the null hypothesis is true, is likely. Therefore, I am unable to reject the null hypothesis, stating that the ratio of variances is equal to 1, with 95% confidence. So, there is no enough evidence to assume that cultivar of cabbage has a significant effect on the amount of ascorbic acid variation.

*# F-Test : are the variances of Ascorbic acid content same for two different cultivars of the cabbage?*  
*# H0 : they are same : v1 = v2*  
*# H1 : they are not same : v1 != v2*  
*# significance level = 0.05 (95% Confidence)*  
*# my test is two sided*  
  
**var.test**(c39**$**VitC,c52**$**VitC,alternative = "two.sided")   
## F test to compare two variances  
  
## data: c39$VitC and c52$VitC  
## F = 0.70977, num df = 29, denom df = 29, p-value = 0.3613  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.3378266 1.4912287  
## sample estimates:  
## ratio of variances 0.7097723  
*### p-value of the test is 0.3613 which is bigger than my alpha (0.05).*  
*### So, with 95% confidence, I am unable to reject my null hypothesis.*  
*### I accept default choice, H0, stating that variance in the content of Ascorbic acid for 2 different cultivars of cabbage is statistically same.*

# 

# Conclusion

To conclude, I carried out 2 statistical tests with the aid of R. I utilized cabbage datasets. This was a built-in dataset provided by R. My main aim was to utilize these datasets in order to prove/disprove some hypotheses. First, I proved that I have evidence to assume that the cultivar of cabbage has a significant effect on its head weight. On the other hand, I did not find any significant effect of the cultivar when it comes to the variation on the amount of ascorbic content.

# Reference

Ripley, B. (2019, December 20). Support Functions and Datasets for Venables and Ripley's MASS [R package MASS version 7.3-51.5]. Retrieved from [Package MASS](https://cran.r-project.org/web/packages/MASS)

# 