t-test in R

Pranav Gopalkrisha, 1940223

2020-08-21

setwd("~/Downloads")  
cs2m = read.csv("cs2m.csv")  
grades = read.csv("grades.csv")  
View(cs2m)  
View(grades)

**DESCRIPTION**

t-test is a method of testing the equality of the population means of two random samples. It is relevant in cases where the population is unknown and needs to be inferred from the sample. It is relevant in cases where the populations need not be known, yet their equality needs to be checked.

**ASSUMPTIONS**

1. Samples are independent

2. Populations have the same variance

3. Populations follow normal distribution

\_\_\_

Testing assumptions

1. Use your sense

2. F-test

3. Shapiro-Wilk test

**HYPOTHESES**

Hypotheses that are tested in t-test are

1. H\_0: Difference between the two population means is 0

2. H\_1: Difference between the two population means is not 0 ========================

**QUESTION 1**

Test whether there is significant difference between mean scores of quiz1 and quiz3

quiz1 = grades$quiz1  
quiz3 = grades$quiz3

Testing for assumptions…

1. Are they independent? There is no reason to state that quiz1 score affects quiz3 score Hence, without further information, we may conclude they are independent

\_\_\_

2. Are their variances equal (using F-test)?

H\_0 for F-test: Difference of variances is 0

H\_1 for F-test: Difference of variances is not 0

var.test(quiz1, quiz3, conf.level = 0.95)

##   
## F test to compare two variances  
##   
## data: quiz1 and quiz3  
## F = 1.1556, num df = 104, denom df = 104, p-value = 0.4623  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.7851229 1.7007616  
## sample estimates:  
## ratio of variances   
## 1.155555

The p-value of the F-test is 0.4623 > 0.05. Hence, we accept the null hypothesis. Hence, we conclude that the variances of quiz1 and quiz3 are equal.

\_\_\_

3. Do their populations follow normal distribution?

H\_0 for Shapiro-Wilk test: Population follows normal distribution

H\_1 for Shapiro-Wilk test: Population does not follow normal distribution

shapiro.test(quiz1)

##   
## Shapiro-Wilk normality test  
##   
## data: quiz1  
## W = 0.88123, p-value = 1.178e-07

p = 1.178e-07 < 0.05

shapiro.test(quiz3)

##   
## Shapiro-Wilk normality test  
##   
## data: quiz3  
## W = 0.83191, p-value = 1.414e-09

p = 1.414e-09 < 0.05 In both cases, p < 0.05. Hence, we reject the null hypothesis that their populations do not follow normal distribution

————

The t-test

t.test(quiz1, quiz3, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: quiz1 and quiz3  
## t = -1.5552, df = 208, p-value = 0.1214  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.1662001 0.1376287  
## sample estimates:  
## mean of x mean of y   
## 7.466667 7.980952

The p-value = 0.1214 > 0.05 Hence we accept the null hypothesis. Hence, we can conclude that the mean test scores between quiz1 and quiz2 are not significantly different. Now, both sample populations may not follow normal distribution, but since the t-test is robust, the results of the t-test may be worth considering. (In fact, their means are within 0.6 units of each other)

========================

**QUESTION 2**

Test whether there is significance difference between Blood Pressure of women across Drug Reaction

BP = cs2m$BP  
DrugR = cs2m$DrugR

Testing for assumptions…

1. Are they independent? Blood pressure and drug reaction may or may not be independent, since we have insufficient information about the nature of the drug Hence, we assume them as independent.

\_\_\_

2. Are their variances equal (using F-test)?

H\_0 for F-test: Difference of variances is 0

H\_1 for F-test: Difference of variances is not 0

var.test(BP, DrugR, conf.level = 0.95)

##   
## F test to compare two variances  
##   
## data: BP and DrugR  
## F = 2018.2, num df = 29, denom df = 29, p-value < 2.2e-16  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 960.6027 4240.2764  
## sample estimates:  
## ratio of variances   
## 2018.222

The p-value of the F-test < 2.2e-16 < 0.05. Hence, we reject the null hypothesis. Hence, we conclude that the variances of BP and DrugR are unequal.

\_\_\_

3. Do their populations follow normal distribution?

H\_0 for Shapiro-Wilk test: Population follows normal distribution

H\_1 for Shapiro-Wilk test: Population does not follow normal distribution

shapiro.test(BP)

##   
## Shapiro-Wilk normality test  
##   
## data: BP  
## W = 0.9462, p-value = 0.1337

p = 0.1337 > 0.05

shapiro.test(DrugR)

##   
## Shapiro-Wilk normality test  
##   
## data: DrugR  
## W = 0.63824, p-value = 2.211e-07

p = 2.211e-07 < 0.05 Hence, we accept the null hypothesis for BP, but reject it for DrugR Hence, only BP’s population can be concluded to follow normal distribution.

————

The t-test

t.test(BP, DrugR, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: BP and DrugR  
## t = 30.4, df = 29.029, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 118.3006 135.3660  
## sample estimates:  
## mean of x mean of y   
## 127.3333 0.5000

The p-value < 2.2e-16 < 0.05 Hence we reject the null hypothesis. Hence, we can say that there is significant difference between blood pressure and drug reaction. From this, we may conclude that these two measures are unrelated, or weakly related. Now, one sample population may not follow normal distribution, but since the t-test is robust, the results of the t-test may be worth considering.

========================

**QUESTION 3**

Test whether Drug Reaction and Anxiety Levels were significantly associated

(Significantly associated => their populations are similar)

DrugR = cs2m$DrugR  
AnxtyLH = cs2m$AnxtyLH

Testing for assumptions…

1. Are they independent? Anxiety level and drug reaction may or may not be independent, since we have insufficient information about the nature of the drug Hence, we assume them as independent.

\_\_\_

2. Are their variances equal (using F-test)?

H\_0 for F-test: Difference of variances is 0

H\_1 for F-test: Difference of variances is not 0

var.test(DrugR, AnxtyLH, conf.level = 0.95)

##   
## F test to compare two variances  
##   
## data: DrugR and AnxtyLH  
## F = 1.0045, num df = 29, denom df = 29, p-value = 0.9905  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.4780896 2.1103753  
## sample estimates:  
## ratio of variances   
## 1.004464

The p-value of the F-test is 0.9905 > 0.05. Hence, we accept the null hypothesis. Hence, we conclude that the variances of DrugR are AnxtyLH equal. (In fact, the ratio of their variances is very close to 1)

\_\_\_

3. Do their populations follow normal distribution?

H\_0 for Shapiro-Wilk test: Population follows normal distribution

H\_1 for Shapiro-Wilk test: Population does not follow normal distribution

shapiro.test(DrugR)

##   
## Shapiro-Wilk normality test  
##   
## data: DrugR  
## W = 0.63824, p-value = 2.211e-07

p = 2.211e-07 < 0.05

shapiro.test(AnxtyLH)

##   
## Shapiro-Wilk normality test  
##   
## data: AnxtyLH  
## W = 0.63662, p-value = 2.107e-07

p = 2.107e-07 < 0.05 For both cases, p < 0.05, hence we reject the null hypothesis for both samples. Hence, none of the populations follow normal distribution.

————

The t-test

t.test(DrugR, AnxtyLH, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: DrugR and AnxtyLH  
## t = 0.25414, df = 58, p-value = 0.8003  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.2292129 0.2958795  
## sample estimates:  
## mean of x mean of y   
## 0.5000000 0.4666667

The p-value = 0.8003 > 0.05

Hence we accept the null hypothesis. Hence, we can say that there is no significant difference between drug reaction and anxiety level. From this, we may conclude that these two measures are significantly associated. Now, both sample populations may not follow normal distribution, but since the t-test is robust, the results of the t-test may be worth considering.