F-test

1940223, Pranav

2020-10-23

# F TEST

#

F test is a test to check the equality in the variances of two populations, given the samples from these populations. The hypotheses of these tests are

# H\_0: The variances of these populations are equal  
# H\_1: The variances of these populations are unequal (with some type of inequality)

F test’s main assumption is that the populations follow normal distributions.

#

# QUESTION 1

#

Generate/ enter 2 samples (remember to use rnorm if generating) and test for equality of variances.

#

The samples are

q1\_data1 = rnorm(30, mean = 5, sd = 6)  
q1\_data2 = rnorm(20, mean = 3, sd = 8)  
q1\_data1

## [1] 0.4673815 4.9991241 5.1122205 8.2767770 6.9540109 7.4337828  
## [7] 7.7414484 13.9075975 5.1262200 -11.9210025 4.5885723 10.7657810  
## [13] -4.0396630 7.4855961 -5.7439017 -3.6942477 10.2194806 5.9689796  
## [19] 9.6176602 12.5124449 8.0945125 3.1909360 -0.7795413 -6.9688227  
## [25] 9.6545703 2.2812336 1.9176037 -3.2693824 15.3089096 1.5264136

q1\_data2

## [1] 3.2104760 -6.2108843 -0.7448245 -6.5394222 -0.0497775 4.9193598  
## [7] 15.7931501 -6.0313037 -1.0302884 2.5639172 3.8116916 8.6104510  
## [13] 9.1023040 4.5462616 -6.1786410 1.7888361 9.1034060 -10.6660163  
## [19] 17.9784201 13.8664155

We know that the populations are derived from a normal distribution. Nevertheless, let us confirm it with the Shapiro-Wilk test…

shapiro.test(q1\_data1)

##   
## Shapiro-Wilk normality test  
##   
## data: q1\_data1  
## W = 0.96778, p-value = 0.4803

shapiro.test(q1\_data2)

##   
## Shapiro-Wilk normality test  
##   
## data: q1\_data2  
## W = 0.96704, p-value = 0.6915

For both cases, p > 0.05. Hence, we may conclude that the samples follow normal distribution (amazing revelations here).

#

Hyptheses are

# H\_0: Population variances are equal  
# H\_0: Population variances are unequal  
#

Now for the F-test…

var.test(q1\_data1, q1\_data2)

##   
## F test to compare two variances  
##   
## data: q1\_data1 and q1\_data2  
## F = 0.67778, num df = 29, denom df = 19, p-value = 0.3367  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.282179 1.512308  
## sample estimates:  
## ratio of variances   
## 0.6777778

p > 0.05. Hence, we accept H\_0. Hence, we may conclude that the samples are drawn from populations with equal variances.

#

# QUESTION 2

#

Use any in-built data set, choose any 2 variables and test for equality of variances (or you can import a data set and do this question)

#

Data set chosen: handdFootLength.csv

q2\_data = read.csv("~/Downloads/handFootLength.csv")  
head(q2\_data, 5)

## Person Hand.Length Foot.Length  
## 1 1 10 15  
## 2 2 12 14  
## 3 3 9 13  
## 4 4 8 13  
## 5 5 7 11

The samples are

q2\_data1 = q2\_data$Hand.Length  
q2\_data2 = q2\_data$Foot.Length

To test whether the samples are drawn from normal distributions, we use the Shapiro-Wilk test…

shapiro.test(q2\_data1)

##   
## Shapiro-Wilk normality test  
##   
## data: q2\_data1  
## W = 0.94994, p-value = 0.6678

shapiro.test(q2\_data2)

##   
## Shapiro-Wilk normality test  
##   
## data: q2\_data2  
## W = 0.90866, p-value = 0.2719

In both cases, p > 0.05. Hence, we may conclude that both samples are drawn from normal distributions.

#

Hyptheses are

# H\_0: Population hand and foot lengths have equal variance  
# H\_0: Population hand and foot lengths have unequal variance  
#

Now for the F-test…

var.test(q2\_data1, q2\_data2)

##   
## F test to compare two variances  
##   
## data: q2\_data1 and q2\_data2  
## F = 0.96, num df = 9, denom df = 9, p-value = 0.9525  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.2384504 3.8649544  
## sample estimates:  
## ratio of variances   
## 0.96

p = 0.9525 > 0.05. Hence, we may accept H\_0. Hence, we may conclude that population hand and foot lengths have equal variances.