Lab 07

Jyosna Philip

2024-01-09

# INTRODUCTION

We aim for the examination of data dispersion within a singular group and the rigorous comparison of variances between two distinct populations. These can be done using chi squared test and f test respectively.

# TESTING VARIANCE OF SINGLE POPULATION

To test the variance of single population, chi square test is used. The library required for this is imported.

library(EnvStats)

##   
## Attaching package: 'EnvStats'

## The following objects are masked from 'package:stats':  
##   
## predict, predict.lm

# Importing Dataset

# Dataset Description

This dataset contains online shopping data,and details regarding the use of discount coupons.

online\_shopping <- read.csv("C:/Users/jyosn/Downloads/archive (2)/online\_shopping.csv")  
online\_shopping<- na.omit(online\_shopping)  
head(online\_shopping)

## X CustomerID Gender Location Tenure\_Months Transaction\_ID Transaction\_Date  
## 1 0 17850 M Chicago 12 16679 2019-01-01  
## 2 1 17850 M Chicago 12 16680 2019-01-01  
## 3 2 17850 M Chicago 12 16696 2019-01-01  
## 4 3 17850 M Chicago 12 16699 2019-01-01  
## 5 4 17850 M Chicago 12 16700 2019-01-01  
## 6 5 17850 M Chicago 12 16701 2019-01-01  
## Product\_SKU Product\_Description  
## 1 GGOENEBJ079499 Nest Learning Thermostat 3rd Gen-USA - Stainless Steel  
## 2 GGOENEBJ079499 Nest Learning Thermostat 3rd Gen-USA - Stainless Steel  
## 3 GGOENEBQ078999 Nest Cam Outdoor Security Camera - USA  
## 4 GGOENEBQ079099 Nest Protect Smoke + CO White Battery Alarm-USA  
## 5 GGOENEBJ079499 Nest Learning Thermostat 3rd Gen-USA - Stainless Steel  
## 6 GGOENEBJ079499 Nest Learning Thermostat 3rd Gen-USA - Stainless Steel  
## Product\_Category Quantity Avg\_Price Delivery\_Charges Coupon\_Status GST  
## 1 Nest-USA 1 153.71 6.5 Used 0.1  
## 2 Nest-USA 1 153.71 6.5 Used 0.1  
## 3 Nest-USA 2 122.77 6.5 Not Used 0.1  
## 4 Nest-USA 1 81.50 6.5 Clicked 0.1  
## 5 Nest-USA 1 153.71 6.5 Clicked 0.1  
## 6 Nest-USA 1 153.71 6.5 Clicked 0.1  
## Date Offline\_Spend Online\_Spend Month Coupon\_Code Discount\_pct  
## 1 1/1/2019 4500 2424.5 1 ELEC10 10  
## 2 1/1/2019 4500 2424.5 1 ELEC10 10  
## 3 1/1/2019 4500 2424.5 1 ELEC10 10  
## 4 1/1/2019 4500 2424.5 1 ELEC10 10  
## 5 1/1/2019 4500 2424.5 1 ELEC10 10  
## 6 1/1/2019 4500 2424.5 1 ELEC10 10

tail(online\_shopping)

## X CustomerID Gender Location Tenure\_Months Transaction\_ID  
## 52919 52918 15002 M Chicago 41 23800  
## 52920 52919 13155 F California 8 22504  
## 52921 52920 18077 M Chicago 34 24250  
## 52922 52921 16085 M California 15 39991  
## 52923 52922 16085 M California 15 39991  
## 52924 52923 13659 F Chicago 8 39998  
## Transaction\_Date Product\_SKU Product\_Description  
## 52919 2019-03-22 GGOEGGCX056299 Gift Card - $25.00  
## 52920 2019-03-10 GGOEGGCX056399 Gift Card - $250.00  
## 52921 2019-03-28 GGOEGGCX056299 Gift Card - $25.00  
## 52922 2019-10-06 GGOEGOCD078399 Google Leather Perforated Journal  
## 52923 2019-10-06 GGOEGOCR078499 Google Spiral Leather Journal  
## 52924 2019-10-06 GGOEGOCC077999 Google Spiral Journal with Pen  
## Product\_Category Quantity Avg\_Price Delivery\_Charges Coupon\_Status  
## 52919 Gift Cards 1 25.00 0.0 Clicked  
## 52920 Gift Cards 1 250.00 0.0 Clicked  
## 52921 Gift Cards 1 25.00 0.0 Used  
## 52922 Notebooks & Journals 1 10.80 6.0 Clicked  
## 52923 Notebooks & Journals 1 9.60 6.0 Used  
## 52924 Notebooks & Journals 1 5.59 6.5 Not Used  
## GST Date Offline\_Spend Online\_Spend Month Coupon\_Code Discount\_pct  
## 52919 0.05 3/22/2019 2000 2523.41 3 GC30 30  
## 52920 0.05 3/10/2019 2500 1294.22 3 GC30 30  
## 52921 0.05 3/28/2019 2000 1066.12 3 GC30 30  
## 52922 0.05 10/6/2019 3000 2230.76 10 NJ10 10  
## 52923 0.05 10/6/2019 3000 2230.76 10 NJ10 10  
## 52924 0.05 10/6/2019 3000 2230.76 10 NJ10 10

# Target Variable

Our target variable is Avg\_Price.

# POPULATION

population=online\_shopping$Avg\_Price

pop\_var=var(population)  
pop\_var

## [1] 4116.076

#Sample

s=sample(population,150,replace=FALSE)

# CHI-SQUARE TEST FOR SINGLE MEAN

H0: Variance of Avg\_Price of all transactions 4000 H1: VAriance of Avg\_Price of all transactions !=4000

varTest(s,alternative = "two.sided",sigma.squared = 4000)

##   
## Results of Hypothesis Test  
## --------------------------  
##   
## Null Hypothesis: variance = 4000  
##   
## Alternative Hypothesis: True variance is not equal to 4000  
##   
## Test Name: Chi-Squared Test on Variance  
##   
## Estimated Parameter(s): variance = 4370.657  
##   
## Data: s  
##   
## Test Statistic: Chi-Squared = 162.807  
##   
## Test Statistic Parameter: df = 149  
##   
## P-value: 0.4152391  
##   
## 95% Confidence Interval: LCL = 3526.118  
## UCL = 5561.393

Since p-value is 0.38 and thus greater than 0.05, we fail to reject H0. Hence variance of the population, that is , variance of avg\_price of all transactions is 4000.

# CONCLUSION

We have carried out chi-square test for test of variance for single population. Our p-value came out greater than 0.05 and thus we can conclude that the variance of avg\_price of all transactions is 4000.

# TESTING EQUALITY OF VARIANCES

Here our target variables are avg\_price of male and females respectively. # GETTING POPULATIONS

popMale=subset(online\_shopping,Gender=='M')  
popMale1=popMale$Avg\_Price  
popFemale=subset(online\_shopping,Gender=='F')  
popFemale1=popFemale$Avg\_Price

# GETTING SAMPLES

sMale=sample(popMale1,150,replace=TRUE)  
sFemale=sample(popFemale1,150,replace=TRUE)

## COMPARING VARIANCES OF TWO POPULATION

TWO-SIDED TEST H0: Variances of Avg\_Price spend by Males and Females are equal H1: Variances of Avg\_Price spend by Males and Females are not equal

var.test(sMale,sFemale,alternative = "two.sided",conf.level = 0.95)

##   
## F test to compare two variances  
##   
## data: sMale and sFemale  
## F = 1.9952, num df = 149, denom df = 149, p-value = 3.073e-05  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 1.445369 2.754110  
## sample estimates:  
## ratio of variances   
## 1.99517

p-value is less than 0.05, Hence we reject H0. The variances are not equal.

RIGHT-SIDED HYPOTHESIS TEST H0: Variances of Avg\_Price spend by Males and Females are equal H1: Variances of Avg\_Price spend by Males is greater than Females

var.test(sMale,sFemale,alternative = "greater",conf.level = 0.95)

##   
## F test to compare two variances  
##   
## data: sMale and sFemale  
## F = 1.9952, num df = 149, denom df = 149, p-value = 1.537e-05  
## alternative hypothesis: true ratio of variances is greater than 1  
## 95 percent confidence interval:  
## 1.522516 Inf  
## sample estimates:  
## ratio of variances   
## 1.99517

p-value is 0.016<0.05, Hence we reject H0. The variance of avg\_price spend by males is significantly greater than that of females.

# CONCLUSION

We used F-Test to compare the variance of two population,here, it was to compare the variance of avg\_price spent by males and females. During right-tailed hypothesis test, our p-value was less than 0.05. Hence we can conclude that the variance of avg\_price spend by males is significantly greater than that of females.