Lab 07

Nilanjana Dey

2024-01-09

#INTRODUCTION:

This dataset contains all physical and clinical parameters to determine PCOS and infertility related issues. There are a total of 44 parameters.

This data is collected from 10 different hospitals across Kerala, India.

The unit used is feet to cm

Blood pressure entered as systolic and diastolic separately

RBS means Random glucose test

Beta-HCG cases are mentioned as Case I and II

Blood Group indications: A+ = 11, A- = 12, B+ = 13, B- = 14, O+ =15, O- = 16, AB+ =17, AB- = 18

library(EnvStats)

## Warning: package 'EnvStats' was built under R version 4.2.3

##   
## Attaching package: 'EnvStats'

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

## LOADING DATASET

library(readxl)

## Warning: package 'readxl' was built under R version 4.2.3

PCOS <- read\_excel("C:/Users/NILANJANA/Downloads/jhkjfsfljbj skbjbfjjf.xlsx",   
 sheet = "Full\_new")

## New names:  
## • `` -> `...45`

View(PCOS)

SINGLE SAMPLE VARIANCE TEST

# HERE WE CHOOSE HAEMOGLOBIN LEVEL AS OUR TARGET VARIABLE

target\_var=PCOS$`Hb(g/dl)`

# A SAMPLE OF 200 OBSERVATIONS IS DRAWN FROM THE SELECTED TARGET VARIABLE  
sample\_single=sample(target\_var,200,replace=TRUE)

HYPOTHESIS TESTING:

H0 : sigma.squared = 1 vs H1 : sigma.squared != 1

# BOTH TAILED TEST  
varTest(sample\_single, alternative = "two.sided", conf.level = 0.95,   
 sigma.squared = 1, data.name = NULL)

##   
## Results of Hypothesis Test  
## --------------------------  
##   
## Null Hypothesis: variance = 1  
##   
## Alternative Hypothesis: True variance is not equal to 1  
##   
## Test Name: Chi-Squared Test on Variance  
##   
## Estimated Parameter(s): variance = 0.7785166  
##   
## Data: sample\_single  
##   
## Test Statistic: Chi-Squared = 154.9248  
##   
## Test Statistic Parameter: df = 199  
##   
## P-value: 0.01811339  
##   
## 95% Confidence Interval: LCL = 0.6456285  
## UCL = 0.9573531

INTERPRETATION : We conclude that the population variance is not equal to the assumed variance since p-value < 0.05.

## TEST FOR EQUALITY OF TWO POPULATION VARIANCES

# We are creating 2 populations depending on whether a female has PCOS or not.   
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

p1=filter(PCOS,`PCOS (Y/N)` ==0)

library(dplyr)  
p2=filter(PCOS,`PCOS (Y/N)` ==1)

## We choose Weight as our target variable for testing the equality of two population variances.

# A sample of 150 weights is drawn from 1st population where all females have PCOS  
s1=sample(p1$`Weight (Kg)`,150,replace=TRUE)  
s1

## [1] 70.0 52.0 49.0 58.0 67.0 57.0 57.0 54.0 50.0 59.0 45.0 60.0 70.0 54.7 53.2  
## [16] 48.0 56.0 59.0 55.0 42.0 57.0 64.0 50.0 60.0 59.0 64.4 43.0 58.0 75.0 52.0  
## [31] 45.0 51.0 60.0 71.0 68.0 78.0 55.0 53.4 75.0 65.0 61.0 63.0 53.5 52.0 52.0  
## [46] 63.0 50.0 50.0 46.0 50.0 57.0 67.0 48.0 50.0 45.0 68.0 53.0 55.0 56.0 53.0  
## [61] 47.0 60.0 85.0 65.0 43.7 46.0 42.0 72.0 57.6 65.0 62.0 50.0 53.0 63.0 66.0  
## [76] 57.0 40.0 48.0 45.0 54.0 56.0 55.7 75.0 52.0 53.0 56.0 48.0 54.0 65.0 80.5  
## [91] 52.0 56.0 56.0 67.0 54.0 58.0 48.0 58.0 58.9 71.0 48.0 71.0 53.0 55.0 55.5  
## [106] 68.0 50.0 52.0 67.0 46.0 69.0 76.9 35.0 66.3 56.0 56.0 56.0 48.0 58.5 54.0  
## [121] 55.0 50.0 71.8 65.0 68.0 85.0 76.9 65.0 66.0 56.0 62.2 60.0 62.5 50.0 71.0  
## [136] 65.0 56.0 64.2 48.0 51.0 70.0 60.0 45.0 56.0 72.0 52.0 56.0 49.3 53.0 52.0

# A sample of 70 weights is drawn from 2nd population where all females dont have PCOS  
s2=sample(p2$`Weight (Kg)`,70,replace=TRUE)  
s2

## [1] 50.0 60.0 61.0 50.0 50.0 58.0 66.0 88.0 55.0 89.0 52.0 62.0  
## [13] 62.0 63.0 79.0 70.0 66.0 65.0 31.0 69.0 68.0 53.5 72.0 64.0  
## [25] 54.0 73.5 79.0 74.0 53.0 83.0 60.0 59.0 64.0 56.4 74.0 88.0  
## [37] 54.0 89.0 60.0 63.0 104.0 50.0 76.0 65.0 36.0 60.0 89.0 83.0  
## [49] 71.0 42.0 53.6 66.0 54.0 74.0 60.0 76.0 72.0 79.0 82.0 65.0  
## [61] 45.0 54.0 60.0 50.0 85.0 56.0 85.0 72.0 83.0 62.7

# HYPOTHESIS TESTING

H0: sigma.squared 1 = sigma.squared 2 vs H1: sigma.squared 1 != sigma.squared 2

# BOTH TAILED TEST  
var.test(s1,s2, alternative = "two.sided", conf.level = 0.95)

##   
## F test to compare two variances  
##   
## data: s1 and s2  
## F = 0.44774, num df = 149, denom df = 69, p-value = 4.788e-05  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.2937114 0.6619699  
## sample estimates:  
## ratio of variances   
## 0.4477421

Interpretation : We conclude that the variances of both the populations are equal since the null hypothesis gets accepted as the p-value < 0.05.

Conclusion : The variance of weights of females having PCOS is not equal to the variance of Weights of females not having PCOS.This indicates that the distribution of weights for both the populations from their mean weights is not equal.