2448030\_MDS272\_L6.R

Kevin

2024-11-17

##Q1  
# Data  
library(readxl)  
college1=read\_excel("C:\\Users\\kmgs4\\Downloads\\College1.xlsx")  
college2=read\_excel("C:\\Users\\kmgs4\\Downloads\\College2.xlsx")  
head(college1)

## # A tibble: 6 × 1  
## WeeklyExpenditure\_College1  
## <dbl>  
## 1 853  
## 2 577  
## 3 1125  
## 4 511  
## 5 1797  
## 6 1379

head(college2)

## # A tibble: 6 × 1  
## WeeklyExpenditure\_College2  
## <dbl>  
## 1 1727  
## 2 722  
## 3 748  
## 4 367  
## 5 891  
## 6 797

# means  
mean\_college1=mean(college1$WeeklyExpenditure\_College1);mean\_college1

## [1] 1139.695

mean\_college2=mean(college2$WeeklyExpenditure\_College2);mean\_college2

## [1] 1201.767

# Given data  
sigma1=35  
sigma2=25  
n1=200  
n2=150  
# Z-test  
z\_cal=(mean\_college1 - mean\_college2) / sqrt((sigma1^2 / n1) + (sigma2^2 / n2));z\_cal

## [1] -19.34864

library(BSDA)

## Warning: package 'BSDA' was built under R version 4.4.2

## Loading required package: lattice

##   
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':  
##   
## Orange

test1 = z.test(college1,college2,alternative="two.sided",mu=0,sigma.x=sigma1,sigma.y=sigma2,conf.level=0.99);test1

##   
## Two-sample z-Test  
##   
## data: college1 and college2  
## z = -19.349, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 99 percent confidence interval:  
## -70.33509 -53.80824  
## sample estimates:  
## mean of x mean of y   
## 1139.695 1201.767

test5 = z.test(college1,college2,alternative="two.sided",mu=0,sigma.x=sigma1,sigma.y=sigma2,conf.level=0.95);test5

##   
## Two-sample z-Test  
##   
## data: college1 and college2  
## z = -19.349, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -68.35935 -55.78398  
## sample estimates:  
## mean of x mean of y   
## 1139.695 1201.767

test10 = z.test(college1,college2,alternative="two.sided",mu=0,sigma.x=sigma1,sigma.y=sigma2,conf.level=0.90);test10

##   
## Two-sample z-Test  
##   
## data: college1 and college2  
## z = -19.349, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 90 percent confidence interval:  
## -67.34846 -56.79487  
## sample estimates:  
## mean of x mean of y   
## 1139.695 1201.767

# Compare with critical values for 1%, 5%, and 10% levels  
critical\_values\_1= c(qnorm(0.01/ 2), qnorm(1 -0.01 / 2));critical\_values\_1

## [1] -2.575829 2.575829

critical\_values\_5=c(qnorm(0.05/ 2), qnorm(1 -0.05 / 2));critical\_values\_5

## [1] -1.959964 1.959964

critical\_values\_10=c(qnorm(0.1/ 2), qnorm(1 -0.1 / 2));critical\_values\_10

## [1] -1.644854 1.644854

list(z\_cal,critical\_values\_1,critical\_values\_5,critical\_values\_10)

## [[1]]  
## [1] -19.34864  
##   
## [[2]]  
## [1] -2.575829 2.575829  
##   
## [[3]]  
## [1] -1.959964 1.959964  
##   
## [[4]]  
## [1] -1.644854 1.644854

##Q2  
# Given data  
test\_eq = z.test(college1, college2, alternative="two.sided", mu=0, sigma.x = 30, sigma.y = 30, conf.level = 0.95);test\_eq

##   
## Two-sample z-Test  
##   
## data: college1 and college2  
## z = -19.156, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -68.42268 -55.72066  
## sample estimates:  
## mean of x mean of y   
## 1139.695 1201.767

##Q3  
# Load data  
library(readxl)  
cosmetics=read\_excel("C:\\Users\\kmgs4\\Downloads\\cosmetics dataset.xlsx")  
head(cosmetics)

## # A tibble: 6 × 11  
## Label Brand Name Price Rank Ingredients Combination Dry Normal Oily  
## <chr> <chr> <chr> <dbl> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Moisturizer LA M… Crèm… 175 4.1 Algae (Sea… 1 1 1 1  
## 2 Moisturizer SK-II Faci… 179 4.1 Galactomyc… 1 1 1 1  
## 3 Moisturizer DRUN… Prot… 68 4.4 Water, Dic… 1 1 1 1  
## 4 Moisturizer LA M… The … 175 3.8 Algae (Sea… 1 1 1 1  
## 5 Moisturizer IT C… Your… 38 4.1 Water, Sna… 1 1 1 1  
## 6 Moisturizer TATC… The … 68 4.2 Water, Sac… 1 0 1 1  
## # ℹ 1 more variable: Sensitive <dbl>

sample\_data = sample(cosmetics$Price, size = 55, replace = FALSE)  
sample\_data

## [1] 14 45 60 42 76 42 99 55 35 18 7 42 48 34 85 28 90 38 34  
## [20] 51 25 38 55 115 65 15 7 39 24 24 60 3 62 32 8 30 52 90  
## [39] 28 34 149 48 40 18 270 122 24 25 72 7 20 34 28 20 44

population\_mean=mean(cosmetics$Price);population\_mean

## [1] 55.58424

population\_sd=sd(cosmetics$Price);population\_sd

## [1] 45.01443

#Z-TEST  
test\_cosmetics = z.test(sample\_data,y = NULL, alternative = "two.sided", mu = population\_mean, sigma.x = population\_sd, sigma.y = NULL, conf.level = 0.95);test\_cosmetics

##   
## One-sample z-Test  
##   
## data: sample\_data  
## z = -1.1597, p-value = 0.2462  
## alternative hypothesis: true mean is not equal to 55.58424  
## 95 percent confidence interval:  
## 36.64897 60.44193  
## sample estimates:  
## mean of x   
## 48.54545

##Q4  
# Data  
raw\_milk = c(65.5, 65.8, 68.1, 67.9, 66.6, 66.2, 65.7, 67.8, 65.4, 67.5, 66.8, 65.2, 67.8)  
pasteurized\_milk = c(72.5, 77.1, 74.8, 73.9, 76.5, 74.3, 77.7)  
# F-test using function  
f\_test=var.test(raw\_milk, pasteurized\_milk, alternative = "two.sided", conf.level = 0.98);f\_test

##   
## F test to compare two variances  
##   
## data: raw\_milk and pasteurized\_milk  
## F = 0.32292, num df = 12, denom df = 6, p-value = 0.09053  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 98 percent confidence interval:  
## 0.04183797 1.55665637  
## sample estimates:  
## ratio of variances   
## 0.3229193

# F-test using test statistic  
sd\_raw = sd(raw\_milk)  
sd\_past = sd(pasteurized\_milk)  
n1 = length(raw\_milk);n1

## [1] 13

n2 = length(pasteurized\_milk);n2

## [1] 7

f\_calc = ((sd\_raw\*2))/((sd\_past\*2));f\_calc

## [1] 0.5682599

# Degrees of freedom  
df1=length(raw\_milk) - 1; df1

## [1] 12

df2=length(pasteurized\_milk) - 1;df2

## [1] 6

# F critical value at 0.01 (two-tailed) significance level  
f\_critical=qf(1-(0.02/2),df1,df2);f\_critical

## [1] 7.718333

list(f\_calc,f\_test,f\_critical)

## [[1]]  
## [1] 0.5682599  
##   
## [[2]]  
##   
## F test to compare two variances  
##   
## data: raw\_milk and pasteurized\_milk  
## F = 0.32292, num df = 12, denom df = 6, p-value = 0.09053  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 98 percent confidence interval:  
## 0.04183797 1.55665637  
## sample estimates:  
## ratio of variances   
## 0.3229193   
##   
##   
## [[3]]  
## [1] 7.718333