|  |  |  |
| --- | --- | --- |
| Assignment 1 | August 12  15338673 | |
| Paul-Willem Janse van Rensburg | | Multivariate Statistical Analysis |

1. Multivariate paired comparison hypothesis test: Write an R program that uses a multivariate dataset and level of significance as arguments to do a hypothesis test as described on p.275. The output of the program must contain the following: (i)- test statistic, (ii) critical value, (iii) p-value and (iv) a conclusion. Use the program and the Effluent data (Table 6.1) to test the hypothesis  at . Do you obtain the same answers as Example 6.1?

library(data.table)

library(here)

library(MASS)

hypothesis\_test <- function(df, significance){

dbar <- colMeans(df)

sd <- cov(df)

n <- nrow(df)

p <- ncol(df)

t\_square <- n\*dbar%\*%solve(sd)%\*%dbar

crit <- ((n-1)\*p/(n-p))\*qf(1-significance,p,n-p)

p\_val <- 1-pf(t\_square, p, n-p)

if(t\_square > crit){

print(paste0('T^2 = ',round(t\_square, 4),' > ', round(crit, 4),

' We thus reject the null hypothesis and conclude that there is a nonzero mean difference between the measurements with significance level alpha = '

, significance))

}else{

print(paste0('T^2 = ',round(t\_square, 4),' < ', round(crit, 4),

' We thus do not reject the null hypothesis and conclude that there is not a significant difference between the measurements at significance level alpha = '

, significance))

}

print(paste0('p-value = ', p\_val))

}

table\_6\_1 <- as.data.frame(fread(file = here('/assignment\_1/T6-1.dat')))

d1 <- table\_6\_1$V1 - table\_6\_1$V3

d2 <- table\_6\_1$V2 - table\_6\_1$V4

d <- data.frame(d1 = d1, d2 = d2)

hypothesis\_test(d, 0.05)

T^2 = 13.6393 > 9.4589 We thus reject the null hypothesis and conclude that there is a nonzero mean difference between the measurements with significance level alpha = 0.05

p-value = 0.00188652536222111

1. Confidence intevals for  (p. 276): Consider the following 2 intervals

-confidence interval

.

Bonferroni confidence interval

.

Write an R program that gives these intervals as output. The program must receive themultivariate data set, vector **a** and the level of significance as arguments. Use the program and calculate 95% confidence intervals for the Effluent data (Table 6.1). Do you obtain the same answers as Example 6.1?

confidence\_interval <- function(df, a, significance){

dbar <- colMeans(df)

sd <- cov(df)

n <- nrow(df)

p <- ncol(df)

t\_conf\_int <- sqrt((((n-1)\*p/(n-p))\*qf(1-significance,p,n-p))\*(a%\*%sd%\*%a/n))

bonf\_conf\_int <- qt(1-significance/2\*p, n-1)\*sqrt((a%\*%sd%\*%a/n))

return(list('t\_conf\_int' = c(a%\*%dbar - t\_conf\_int, a%\*%dbar + t\_conf\_int),

'bonf\_conf\_int' = c(a%\*%dbar - bonf\_conf\_int, a%\*%dbar + bonf\_conf\_int)))

}

confidence\_interval(d, c(1,0),0.05)

$t\_conf\_int

[1] -22.45327 3.72600

$bonf\_conf\_int

[1] -17.077580 -1.649692

1. Confidence region for the mean,: Consider the following expression for the  confidence region for the mean (p. 276):

.

Use this expression and write an R program to construct a confidence region for any two dimensional dataset. Thus, your R program should use any two dimensional dataset and level of significance as arguments to draw an ellipse. Use this program and draw an ellipse for the Effluent data (Table 6.1).

1. Multivariate repeated measures design: Write an R program that uses a multivariate dataset, matrix **C** and level of significance as arguments to do a hypothesis test as described on p.280. The output of the program must contain the following: (i)- test statistic, (ii) critical value, (iii) p-value (iv) a conclusion and (v) the confidence intervals obtained using the following formula on p. 281

.

Use this program and the Sleeping-Dog data (Table 6.2) and repeat Example 6.2 at .

1. Use R to do the following exercises:
2. Exercise 6.1
3. Exercise 6.2
4. Exercise 6.8
5. Exercise 6.23 (Write your own R program for Box’s test)
6. Use PROC GLM of SAS to do the following:
7. Repeat Examples 6.9 and 6.13
8. Exercise 6.33 (a), (b) and (c). What are your conclusions?