BACS-hw7.R

2022-04-27

setwd("C:/Users/eason/Desktop/清大 BACS/資料/")  
m1 <- read.csv("pls-media1.csv", header = TRUE)  
m2 <- read.csv("pls-media2.csv", header = TRUE)  
m3 <- read.csv("pls-media3.csv", header = TRUE)  
m4 <- read.csv("pls-media4.csv", header = TRUE)  
  
#Question 1  
#a  
experiment <- rbind(m1, m2, m3, m4)  
exp\_groups <- split(experiment$INTEND.0, experiment$media)  
media\_mean <- sapply(split(experiment$INTEND.0, experiment$media), mean)  
round(media\_mean, 2)

## 1 2 3 4   
## 4.81 3.95 4.72 4.89

#b  
#Descriptive Statistics  
media\_length <- sapply(split(experiment$INTEND.0, experiment$media), length)  
media\_sd <- sapply(split(experiment$INTEND.0, experiment$media), sd)  
media\_se <- media\_sd / sqrt(media\_length)  
  
media\_length

## 1 2 3 4   
## 42 38 40 46

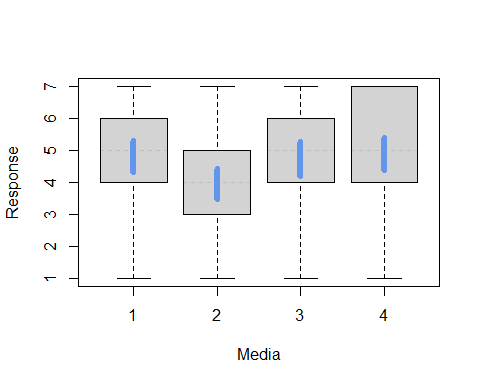
media\_sd

## 1 2 3 4   
## 1.641506 1.523640 1.753933 1.816324

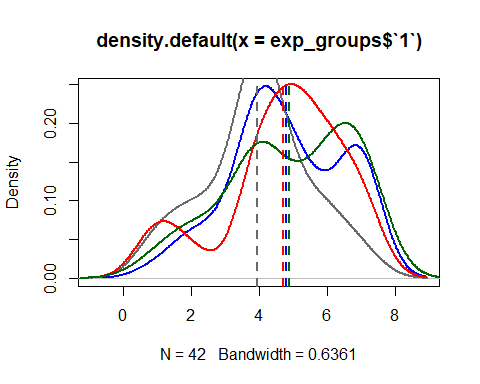
media\_se

## 1 2 3 4   
## 0.2532898 0.2471671 0.2773212 0.2678024

#Boxplot  
boxplot(experiment$INTEND.0 ~ experiment$media,  
 xlab="Media", ylab="Response",   
 medcol="gray", medlty="dashed", medlwd=1)  
segments( 1:4, media\_mean - 1.96\*media\_se,   
 1:4, media\_mean + 1.96\*media\_se,   
 lwd=6, col="cornflowerblue")



#Density Plots  
plot(density(exp\_groups$`1`), col="blue", lwd=2)  
lines(density(exp\_groups$`2`), col="dimgrey", lwd=2)  
lines(density(exp\_groups$`3`), col="red", lwd=2)  
lines(density(exp\_groups$`4`), col="darkgreen", lwd=2)  
abline(v =media\_mean, lty="dashed", col=c("blue", "dimgrey", "red", "darkgreen"), lwd=2)



#c  
  
  
  
#Question 2  
#a  
  
#b  
all\_media <- experiment$INTEND.0  
media\_grand\_mean <- mean(all\_media)  
n\_total <- length(all\_media)   
k <- length(exp\_groups)  
n\_total

## [1] 166

k

## [1] 4

sstr <- sum(media\_length)\*((media\_mean - media\_grand\_mean)^2)  
df\_mstr <- k - 1  
mstr <- sstr/df\_mstr  
  
sse <- sum((media\_length - 1)\*(media\_sd^2))  
df\_mse <- n\_total - k  
mse <- sse/df\_mse  
  
f\_value <- mstr/mse  
f\_value

## 1 2 3 4   
## 0.7338318 8.5822573 0.2356619 1.4781264

p\_value <- pf(f\_value, df\_mstr, df\_mse, lower.tail = FALSE)  
p\_value

## 1 2 3 4   
## 5.332505e-01 2.537585e-05 8.714106e-01 2.224810e-01

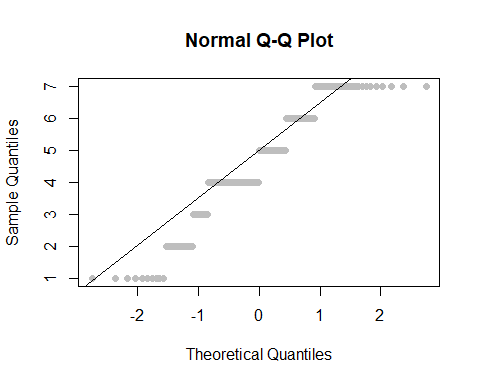
#c  
ftest\_aov <- aov(experiment$INTEND.0 ~ factor(experiment$media))  
summary(ftest\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(experiment$media) 3 22.5 7.508 2.617 0.0529 .  
## Residuals 162 464.8 2.869   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#d 要懂怎麼看結果  
TukeyHSD(ftest\_aov)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = experiment$INTEND.0 ~ factor(experiment$media))  
##   
## $`factor(experiment$media)`  
## diff lwr upr p adj  
## 2-1 -0.86215539 -1.84660332 0.1222925 0.1085727  
## 3-1 -0.08452381 -1.05596494 0.8869173 0.9959223  
## 4-1 0.08178054 -0.85664966 1.0202107 0.9959032  
## 3-2 0.77763158 -0.21843807 1.7737012 0.1825044  
## 4-2 0.94393593 -0.01996662 1.9078385 0.0573229  
## 4-3 0.16630435 -0.78431033 1.1169190 0.9687417

#e  
qqnorm(experiment$INTEND.0, pch=19, col="gray")  
qqline(experiment$INTEND.0)



f\_test <- oneway.test(experiment$INTEND.0 ~ factor(experiment$media), var.equal = TRUE)  
f\_test

##   
## One-way analysis of means  
##   
## data: experiment$INTEND.0 and factor(experiment$media)  
## F = 2.6167, num df = 3, denom df = 162, p-value = 0.05289

qf(df1=4-1, df2=nrow(experiment)-4, p=c(0.95, 0.99))

## [1] 2.660406 3.904807

qf

## function (p, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)   
## {  
## if (missing(ncp))   
## .Call(C\_qf, p, df1, df2, lower.tail, log.p)  
## else .Call(C\_qnf, p, df1, df2, ncp, lower.tail, log.p)  
## }  
## <bytecode: 0x000000001510e1c8>  
## <environment: namespace:stats>

#Question 3  
#a  
  
#b 先把他整理成一個表格, 有點好奇是不是一定要apply才可以多種分開, rank\_msq那邊看一下  
all\_ranks <- rank(experiment$INTEND.0)  
group\_ranks <- split(all\_ranks, experiment$media)  
  
group\_ranksums <- sapply(group\_ranks, sum)  
group\_n <- sapply(group\_ranks, length)  
group\_rankmeans <- sapply(group\_ranks, mean)  
  
N <- nrow(experiment)  
groups <- data.frame(ranksum = group\_ranksums,  
 n = group\_n,  
 rankmeans = group\_rankmeans)  
ngroups <- nrow(groups)  
  
rank\_msq <- function(strategy){  
 strategy['ranksum']^2/ strategy['n']  
}  
  
H <- 12/(N\*(N+1))\*sum(apply(groups, 1, rank\_msq)) - 3\*(N+1)  
kw\_p <- 1 - pchisq(H, df=ngroups-1)  
  
#c  
kruskal.test(experiment$INTEND.0 ~ experiment$media, data = experiment)

##   
## Kruskal-Wallis rank sum test  
##   
## data: experiment$INTEND.0 by experiment$media  
## Kruskal-Wallis chi-squared = 8.8283, df = 3, p-value = 0.03166

#d 要懂結果有甚麼意思 找不到老師要的FSA套件，於是跳過執行，僅列出code  
#install.packages(FSA)  
#library(FSA)  
#dunnTest(experiment$INTEND.0 ~ experiment$media, data = experiment)