Thota, Sunil Raj - Hypothesis Testing with R.R

# Intermediate Analytics  
# ALY 6015  
# Module 1 - Hypothesis Testing with R  
# 01/30/2021  
# Sunil Raj Thota  
# NUID: 001099670

# Get and set the working directories  
getwd()

## [1] "G:/NEU/Coursework/2021 Q1 Winter/ALY 6015 IA/Discussions & Assignments"

setwd('G:/NEU/Coursework/2021 Q1 Winter/ALY 6015 IA/Discussions & Assignments')  
getwd()

## [1] "G:/NEU/Coursework/2021 Q1 Winter/ALY 6015 IA/Discussions & Assignments"

# Installed the above packages into the workspace  
install.packages("datasets")  
install.packages("plyr")  
install.packages("dplyr")  
install.packages("tidyr")  
install.packages("MASS")  
  
# Loaded the below libraries into the workspace  
library(plyr)

library(dplyr)

library(tidyr)

library(MASS)

require(datasets)

# Part A

data(chem) # Load the Chem Data set into the Environment  
View(chem) # To View the Chem Data set  
str(chem) # To observe the structure of the Data set

## num [1:24] 2.9 3.1 3.4 3.4 3.7 3.7 2.8 2.5 2.4 2.4 ...

head(chem) # It shows first few rows in the Data set

## [1] 2.9 3.1 3.4 3.4 3.7 3.7

tail(chem) # It shows last few rows in the Data set

## [1] 3.4 2.2 3.5 3.6 3.7 3.7

summary(chem) # Provides the Descriptive Stats of the Chem Data set

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.200 2.775 3.385 4.280 3.700 28.950

# Part B

tTest <- t.test(chem,  
 alternative = "greater",  
 mu = 1)  
tTest

##   
## One Sample t-test  
##   
## data: chem  
## t = 3.0337, df = 23, p-value = 0.002952  
## alternative hypothesis: true mean is greater than 1  
## 95 percent confidence interval:  
## 2.427162 Inf  
## sample estimates:  
## mean of x   
## 4.280417

# Part C

data(cats) # Load the Cats Data set into the Environment  
View(cats) # To View the Cats Data set  
str(cats) # To observe the structure of the Data set

## 'data.frame': 144 obs. of 3 variables:  
## $ Sex: Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Bwt: num 2 2 2 2.1 2.1 2.1 2.1 2.1 2.1 2.1 ...  
## $ Hwt: num 7 7.4 9.5 7.2 7.3 7.6 8.1 8.2 8.3 8.5 ...

head(cats) # It shows first few rows in the Data set

## Sex Bwt Hwt  
## 1 F 2.0 7.0  
## 2 F 2.0 7.4  
## 3 F 2.0 9.5  
## 4 F 2.1 7.2  
## 5 F 2.1 7.3  
## 6 F 2.1 7.6

tail(cats) # It shows last few rows in the Data set

## Sex Bwt Hwt  
## 139 M 3.6 15.0  
## 140 M 3.7 11.0  
## 141 M 3.8 14.8  
## 142 M 3.8 16.8  
## 143 M 3.9 14.4  
## 144 M 3.9 20.5

summary(cats) # Provides the Descriptive Stats of the Cats Data set

## Sex Bwt Hwt   
## F:47 Min. :2.000 Min. : 6.30   
## M:97 1st Qu.:2.300 1st Qu.: 8.95   
## Median :2.700 Median :10.10   
## Mean :2.724 Mean :10.63   
## 3rd Qu.:3.025 3rd Qu.:12.12   
## Max. :3.900 Max. :20.50

maleData <- subset(cats,  
 subset = (cats$Sex == "M"))  
View(maleData)  
str(maleData)

## 'data.frame': 97 obs. of 3 variables:  
## $ Sex: Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Bwt: num 2 2 2.1 2.2 2.2 2.2 2.2 2.2 2.2 2.2 ...  
## $ Hwt: num 6.5 6.5 10.1 7.2 7.6 7.9 8.5 9.1 9.6 9.6 ...

summary(maleData)

## Sex Bwt Hwt   
## F: 0 Min. :2.0 Min. : 6.50   
## M:97 1st Qu.:2.5 1st Qu.: 9.40   
## Median :2.9 Median :11.40   
## Mean :2.9 Mean :11.32   
## 3rd Qu.:3.2 3rd Qu.:12.80   
## Max. :3.9 Max. :20.50

femaleData <- subset(cats,  
 subset = (cats$Sex == "F"))  
View(femaleData)  
str(femaleData)

## 'data.frame': 47 obs. of 3 variables:  
## $ Sex: Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Bwt: num 2 2 2 2.1 2.1 2.1 2.1 2.1 2.1 2.1 ...  
## $ Hwt: num 7 7.4 9.5 7.2 7.3 7.6 8.1 8.2 8.3 8.5 ...

summary(femaleData)

## Sex Bwt Hwt   
## F:47 Min. :2.00 Min. : 6.300   
## M: 0 1st Qu.:2.15 1st Qu.: 8.350   
## Median :2.30 Median : 9.100   
## Mean :2.36 Mean : 9.202   
## 3rd Qu.:2.50 3rd Qu.:10.100   
## Max. :3.00 Max. :13.000

tTestCats <- t.test(maleData$Bwt,  
 femaleData$Bwt,  
 var.equal = FALSE)  
tTestCats

##   
## Welch Two Sample t-test  
##   
## data: maleData$Bwt and femaleData$Bwt  
## t = 8.7095, df = 136.84, p-value = 8.831e-15  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.4177242 0.6631268  
## sample estimates:  
## mean of x mean of y   
## 2.900000 2.359574

# Part D

data(shoes) # Load the Shoes Data set into the Environment  
View(shoes) # To View the Shoes Data set  
str(shoes) # To observe the structure of the Data set

## List of 2  
## $ A: num [1:10] 13.2 8.2 10.9 14.3 10.7 6.6 9.5 10.8 8.8 13.3  
## $ B: num [1:10] 14 8.8 11.2 14.2 11.8 6.4 9.8 11.3 9.3 13.6

head(shoes) # It shows first few rows in the Data set

## $A  
## [1] 13.2 8.2 10.9 14.3 10.7 6.6 9.5 10.8 8.8 13.3  
##   
## $B  
## [1] 14.0 8.8 11.2 14.2 11.8 6.4 9.8 11.3 9.3 13.6

tail(shoes) # It shows last few rows in the Data set

## $A  
## [1] 13.2 8.2 10.9 14.3 10.7 6.6 9.5 10.8 8.8 13.3  
##   
## $B  
## [1] 14.0 8.8 11.2 14.2 11.8 6.4 9.8 11.3 9.3 13.6

summary(shoes) # Provides the Descriptive Stats of the Shoes Data set

## Length Class Mode   
## A 10 -none- numeric  
## B 10 -none- numeric

tTestShoes <-  
 t.test(shoes$A,  
 shoes$B,  
 paired = TRUE,  
 alternative = "less")  
tTestShoes

##   
## Paired t-test  
##   
## data: shoes$A and shoes$B  
## t = -3.3489, df = 9, p-value = 0.004269  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -0.1855736  
## sample estimates:  
## mean of the differences   
## -0.41

# Part E

data(bacteria) # Load the Bacteria Data set into the Environment  
View(bacteria) # To View the Bacteria Data set  
str(bacteria) # To observe the structure of the Data set

## 'data.frame': 220 obs. of 6 variables:  
## $ y : Factor w/ 2 levels "n","y": 2 2 2 2 2 2 1 2 2 2 ...  
## $ ap : Factor w/ 2 levels "a","p": 2 2 2 2 1 1 1 1 1 1 ...  
## $ hilo: Factor w/ 2 levels "hi","lo": 1 1 1 1 1 1 1 1 2 2 ...  
## $ week: int 0 2 4 11 0 2 6 11 0 2 ...  
## $ ID : Factor w/ 50 levels "X01","X02","X03",..: 1 1 1 1 2 2 2 2 3 3 ...  
## $ trt : Factor w/ 3 levels "placebo","drug",..: 1 1 1 1 3 3 3 3 2 2 ...

head(bacteria) # It shows first few rows in the Data set

## y ap hilo week ID trt  
## 1 y p hi 0 X01 placebo  
## 2 y p hi 2 X01 placebo  
## 3 y p hi 4 X01 placebo  
## 4 y p hi 11 X01 placebo  
## 5 y a hi 0 X02 drug+  
## 6 y a hi 2 X02 drug+

tail(bacteria) # It shows last few rows in the Data set

## y ap hilo week ID trt  
## 215 n a hi 11 Z24 drug+  
## 216 y a hi 0 Z26 drug+  
## 217 y a hi 2 Z26 drug+  
## 218 y a hi 4 Z26 drug+  
## 219 n a hi 6 Z26 drug+  
## 220 y a hi 11 Z26 drug+

summary(bacteria) # Provides the Descriptive Stats of the Bacteria Data set

## y ap hilo week ID trt   
## n: 43 a:124 hi:122 Min. : 0.000 X03 : 5 placebo:96   
## y:177 p: 96 lo: 98 1st Qu.: 2.000 X04 : 5 drug :62   
## Median : 4.000 X05 : 5 drug+ :62   
## Mean : 4.455 X07 : 5   
## 3rd Qu.: 6.000 X08 : 5   
## Max. :11.000 X09 : 5   
## (Other):190

tableData <- table(bacteria$y, bacteria$ap)  
propTestBacteria <- prop.test(table(bacteria$y, bacteria$ap),  
 conf.level = 0.95,  
 alternative = "two.sided")  
propTestBacteria

##   
## 2-sample test for equality of proportions with continuity correction  
##   
## data: table(bacteria$y, bacteria$ap)  
## X-squared = 4.6109, df = 1, p-value = 0.03177  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## 0.02813119 0.36288182  
## sample estimates:  
## prop 1 prop 2   
## 0.7209302 0.5254237

# Part F

data(cats) # Load the Cats Data set into the Environment  
View(cats) # To View the Cats Data set  
  
varTestCats <- var.test(maleData$Bwt, femaleData$Bwt)  
varTestCats

##   
## F test to compare two variances  
##   
## data: maleData$Bwt and femaleData$Bwt  
## F = 2.9112, num df = 96, denom df = 46, p-value = 0.0001157  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 1.723106 4.703057  
## sample estimates:  
## ratio of variances   
## 2.911196