**Assignment 4:**

**Due at 11:00pm on the 26th of November through Moodle.**

**Provide all the code (.R file) and all results from R along with your interpretation in a document.**

**Hypothesis Testing:**  
Load in the survey dataset from the MASS package in R.

1. Suppose it is up to you to determine if there is a difference between the male and female students who study at the University of Adelaide. Compare these groups across the following variables ‘Wr.Hnd’, ‘W.Hnd’, ‘Pulse‘, ‘Clap’, ‘Exer’ ‘Smoke’ and ‘Age’ using graphical displays or tables. Make sure to interpret each graph or table.

Wr.Hnd vs Sex: Numerical Continuous Vs CategoricChart, box and whisker chart

Description automatically generated

The boxplot shows that male generally have a bigger writing hand span that the female with a few overlap.

Pulse vs Sex: Numerical Continuous Vs Categorical

Chart, box and whisker chart

Description automatically generated

There’s no big difference between male and female pulse but female have a slightly higher pulse median than male.

Age vs Sex: Numerical Continuous Vs Categorical

Chart, box and whisker chart

Description automatically generated

The age of the male and look similar spread out, with outliers around the same years

Clap vs Sex: Categorical Vs Categorical

Chart, bar chart

Description automatically generated

The hand place on top when clapping are close in all three categories but with right hand being the largest over the other 2 categories.

Exer vs Sex: Categorical Vs Categorical

Chart, bar chart

Description automatically generated

Male student exercises more frequently than female and more male student do not exercise that female.

Smoke vs Sex: Categorical Vs Categorical

Chart, bar chart

Description automatically generated

Majority of the sample does not smoke.

W.Hnd vs Sex: Categorical Vs Categorical

Chart, bar chart

Description automatically generated

Majority of the sample write with their right hands.

1. From Q1, choose one numerical variable to test if there is a significant difference between the male and females’ students and explain this choice. Make sure in your answer to explain the null and alternative hypotheses, any assumptions needed and if they are met, results and interpretation of the results. Conclude your findings.

Writing hand span was choosing as it is said that male tends to have bigger hands than females.

#H0: mu\_male= mu\_female

#H0: mu\_male - mu\_female=0

#mu\_male is true mean writing hand span for male

#mu\_female is true mean writing hand span for female

#H1: mu\_male not equal mu\_female

#H1: mu\_male - mu\_female != 0

mwr\_hnd= Wr.Hnd[Sex=="Male"]

mwr\_hnd

fwr\_hnd= Wr.Hnd[Sex=="Female"]

fwr\_hnd

#Happy with the assumptions of normality and equal var.

mfwr\_hnd=boxplot(mwr\_hnd,fwr\_hnd,names=c("Male","Female"),ylab= "Centimetres",col=4:2, main = paste("Boxplot of Male Vs Female Writing Hand Span"))

t.test(mwr\_hnd, fwr\_hnd)

Welch Two Sample t-test

data: mwr\_hnd and fwr\_hnd

t = 10.619, df = 215.27, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

1.747753 2.544482

sample estimates:

mean of x mean of y

19.74188 17.59576

#pvalue=2.2e-16 < 0.05, therefore reject H0 and conclude

#that there is a signficant difference between true mean male writing hand span

#and true mean female writing hand span.

#95% CI tells us that 95% the true population mean difference for

#male and female writing hand span lie within 19.74188cm and 17.59576cm

1. From Q1, choose one categorical variable to test if there is a significant difference between the male and females’ students and explain this choice. Make sure in your answer to explain the null and alternative hypotheses, any assumptions needed and if they are met, results and interpretation of the results. Conclude your findings.

#H0: p\_male = p\_females, p\_males is population proportion

# Of males smoking habit and p\_females is population

#Proportion of females smoking habit

#H0: sex and smoking habit are independent of each other

#H1: sex and smoking habit are dependent of each other

> smoke\_chisq$expected

Smoke Female Male

Heavy 5.523404 5.476596

Never 94.400000 93.600000

Occas 9.540426 9.459574

Regul 8.536170 8.463830

> smoke\_chisq

Pearson's Chi-squared test

data: mf\_smoke\_table

X-squared = 3.5536, df = 3, p-value = 0.3139

> prop.test(mf\_smoke\_table)

4-sample test for equality of proportions without continuity correction

data: mf\_smoke\_table

X-squared = 3.5536, df = 3, p-value = 0.3139

alternative hypothesis: two.sided

sample estimates:

prop 1 prop 2 prop 3 prop 4

0.4545455 0.5265957 0.4736842 0.2941176

smoke\_chisq<-chisq.test(mf\_smoke\_table)

smoke\_chisq$expected

#happy with assumptions as exp >=5

smoke\_chisq

#p-value=0.3139 >0.05, so fail to reject H0 that sex and

# smoking habit are independent

1. Suppose it is up to you to determine if there is a difference between the students who exercise (‘Exer) at the University of Adelaide.

Compare the 3 exercise groups across the following variables ‘Wr.Hnd’, ‘Pulse‘, ‘Height’, and ‘Age’ using graphical displays. Make sure to interpret each graph.

#Exer VS Wr.Hnd

Chart, box and whisker chart

Description automatically generated

Writing hand span have no impact on Exercise as all the boxplots overlap.

#Exer VS Height

Chart, box and whisker chart

Description automatically generated

Student heights have no impact on Exercise as the boxplot overlap significantly.

#Exer VS Age

Chart

Description automatically generated with medium confidence

Age have no impact on Exercise as the boxplot overlap significantly.

#Exer VS Pulse

Chart, box and whisker chart

Description automatically generated

Student who exercises often have a lower beats per min pulse median than the other two categories who have very similar medians.

1. Choose one of these numerical variables from Q4 to test if there is a significant difference between the 3 exercise groups and explain this choice. Make sure in your answer to explain the null and alternative hypotheses, any assumptions needed and if they are met, results and interpretation of the results. Conclude your findings.

#H0: Exercises habit and Pulse are independent of each other

#H1: Exercises habit and Pulse are dependent of each other

Exer\_Pulse\_table=table(Pulse,Exer)

Exer\_Pulse\_table

Exer\_Pulse\_table\_chisq<-chisq.test(Exer\_Pulse\_table)

Exer\_Pulse\_table\_chisq$expected

#happy with assumptions as exp >=5

Pearson's Chi-squared test

data: Exer\_Pulse\_table

X-squared = 86.927, df = 84, p-value = 0.3918

#p-value=0.3918 >0.05, so fail to reject H0 that Exercises habit and

Pulse are independent