|  |  |  |  |
| --- | --- | --- | --- |
| DNSC 6213-10 Statistics for Analytics II | Lab 2 | GWlogoBlue | Student ID**:**  **G****27279521** |

**(This assignment is to be completed individually. While you may interact with others to gain mastery of the generic course content, all work associated with this assignment must be strictly and exclusively yours, including creating the R code and determining the correct answers/completing the template; the only collaboration permitted is with the instructor.)**

Daxguard, Inc., a pharmaceutical manufacturer, wishes to know about the effect of its new hypertension drug, Tensure, on men and women of different ages. A random sample of people suffering from hypertension is selected, and each individual in the trial is prescribed Tensure for 10 weeks. At the end of the 10-week trial, the difference in systolic blood pressure is measured. The data are contained in a file named Daxguard.txt, which has a header record. The data in this file consist of three variables: Age, Gender, and Sysdiff (reduction in systolic blood pressure: higher values reflect greater reduction in blood pressure). Based on these data, and using the framework described in class, please use R to complete the following template. ***Unless otherwise noted, all results should be accurate to at least 4 decimal places.*** Use an  level of .05:

1. Among *males and females combined*, can we be reasonably certain that Age is related to Sysdiff? (Yes/No) No  
   Why/why not? the multiple r squared is 0.04272 which is only 4.272% has related between Age and Sysdiff among males and females combined. and P value was 0.601 which is bigger than 0.05 which mean we failed to reject null hypothesis  
   If so, does Sysdiff increase or decrease with increases in Age? Age is increased by 0.01236 when Sysdiff increased
2. Among *males*, can we be reasonably certain that Age is related to Sysdiff? (Yes/No) Yes  
   Why/why not? the r squared is 0.6439 which is 64.39% has related between Age and Sysdiff among males. p value is less than 0.05 which mean we reject the null hypothesis.  
   If so, does Sysdiff increase or decrease with increases in Age? Age is incrased by 0.50252 when Sysdiff increased
3. Among *females*, can we be reasonably certain that Age is related to Sysdiff? (Yes/No) Yes  
   Why/why not? the r squared is 0.6295 which is 62.95% has related between Age and Sysdiff among females. p value is less than 0.05 which mean we reject the null hypothesis.  
   If so, does Sysdiff increase or decrease with increases in Age? When Sysidff Incrased, age is decreaed by 0.46563.
4. If your answer to Question 1 was different than your answer to Question 2 and/or Question 3, how do you explain this difference? From Non-Additive Analysis of Covariance Model, there was one intersection between Genders in the graph.
5. What is the simple effect of Age among males? (Please provide a value.) 0.50252  
   Please interpret this simple effect. the value is comes from the slope of the male
6. What is the simple effect of being female among those people at the average Age level? (Please provide a value.) 3.484316  
   Please interpret this simple effect. So before start the question, I had to understand the question first. The question itself is asking me the simple effect of being female is the simple effect of gender where the issue of which gender gets subtracted from which is made explicit. So we need to check the difference between same level of age with different Gender.
7. If you were asked to provide a main effect of Age, what value would you provide? (Please provide a value.) 0.018445  
   How did you derive this value? (-0.46563 + 0.50252) / 2  
   Please interpret this main effect. From the summary of male, I was able to see the slope is 0.50252 while female slope was -0.46563. So from ANACOVA interaction, we could calculate main effect of age from average of two slope. which is (-0.46563 + 0.50252) / 2. From the calculation, the result of it is 0.018445.
8. If you were asked to provide a main effect of Gender, what value would you provide? (Please provide a value.) 3.484316  
   How did you derive this value? 25.59772 - 22.11341  
   Please interpret this main effect. Unlike question 6, We need the average difference value of Age for this question. So I created female\_49.51 and male\_49.51 which is the predict(each gender in average age which is 49.51). Once I got all the values at age of 49.51 for male and female, need to subtract these two value.
9. What is the expected difference in Sysdiff between a male who is 30 years old and a female who is 25 years old? (subtract the female Sysdiff from the male Sysdiff)-24.70115

Paste your full R script immediately below this line:

library(dplyr)

#A02

df <- daxguard

summary(df)

typeof(df$Age)

typeof(df$Gender)

typeof(df$Sysdiff)

df$Gender <- as.factor(df$Gender)

# Question 1

total.lm <- lm(Sysdiff~Age , data = df)

summary(total.lm)

# Question 2

male <- subset(df, Gender == 'M')

male.lm <- lm(Sysdiff~Age, data = male)

summary(male.lm)

# Quesiton 3

female <- subset(df, Gender =='F')

female.lm <- lm(Sysdiff~Age, data = female)

summary(female.lm)

# Question 4

plot(1, type = "n",

xlab = "age", ylab = "Sysdiff",

xlim = c(-10,100), ylim = c(-10,100))

abline(female.lm, lty = 1)

abline(male.lm, lty = 2)

legend("bottomright", c("F","M"), lty=c(1,2))

# Question 5

summary(male.lm)

# Question 6

total.lm <- lm(Sysdiff~Age + Gender + Age\*Gender, data = df)

summary(total.lm)

summary(df) # to check the average of age

# Age mean is 49.51

predict(total.lm, data.frame(Gender ="F", Age = 49.51)) - predict(total.lm, data.frame(Gender ="M", Age = 49.51))

# Question 7

# Main effect of Age = Average of slope

summary(male.lm) #0.50252

summary(female.lm)# -0.46563

(-0.46563 + 0.50252) / 2

# Question 8

female\_49.51 <- predict(total.lm, data.frame(Gender ="F", Age = 49.51))

female\_49.51

male\_49.51 <- predict(total.lm, data.frame(Gender ="M", Age = 49.51))

male\_49.51

female\_49.51 - male\_49.51

# Question 9

#Subtract Sysdiff of Male,30 to Sysdiff of Female,25

predict(total.lm, data.frame(Gender = "M", Age = 30)) -

predict(total.lm, data.frame(Gender = "F", Age = 25))