R code:

# Import Data

setwd("C:/Users/XuWei/OneDrive/BU - Boston University/04 - 2022 Spring Semester/MET CS 555 A2 Data Analysis and Visualization with R/Lab - Session 3 (Group work)")

bp <- read.csv("Lab2Data\_CouplesAge.csv", header = TRUE, fileEncoding = "UTF-8-BOM")

head(bp)

wifeAge <- bp$Age\_of\_Wife

husbandAge <- bp$Age\_of\_Husband

(1). There is a claim that there is a correlation between age of wives and husbands. Examine that claim based on the 10 samples provided with 95% confidence.

R code:

t.test(wifeAge, husbandAge, conf.level = 0.95)

R output:

Welch Two Sample t-test

data: wifeAge and husbandAge

t = 0.17555, df = 12.674, p-value = 0.8634

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

95 percent confidence interval:

-6.803354 8.003354

sample estimates:

mean of x mean of y

29.5 28.9

Alphal level = 1- confidence level = 1 – 0.95 = 0.05

1. Set up the hypotheses and select the alpha level
   1. H0: there is no correlation between age of wives and husbands.
   2. H1: there is correlation between age of wives and husbands.
   3. alpha = 0.05
2. Select the appropriate test statistic
   1. t = 0.17555
3. State the decision rule
   1. Reject H0 if p-value < alpha
4. Compute the test static and the associated p-value
   1. P-value = 0.8634
5. State your conclusion
   1. Since p-value > alpha, fail to reject H0
   2. There is no significant evident concluded that there is correlation between age of wives and husbands.

(2). Build a linear model to predict age of a husband as a function of the wife.

R Code:

# Build a linear model to predict age of a husband as a function of the wife.

p <- as.data.frame(53)

colnames(p) <- "wifeAge"

# Calculating the correlation between Age and Blood pressure

cor(bp$Age\_of\_Husband, bp$Age\_of\_Wife)

# Creating a Linear regression model

model <- lm(husbandAge ~ wifeAge, data = bp)

# Summary of our linear regression model

summary(model)

predict(model, newdata = p)

R Output:

> # Summary of our linear regression model

> summary(model)

Call:

lm(formula = husbandAge ~ wifeAge, data = bp)

Residuals:

Min 1Q Median 3Q Max

-7.5985 -2.7769 -0.7889 4.0762 5.2111

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 27.96376 5.04017 5.548 0.000542 \*\*\*

wifeAge 0.03174 0.16293 0.195 0.850418

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.796 on 8 degrees of freedom

Multiple R-squared: 0.00472, Adjusted R-squared: -0.1197

F-statistic: 0.03794 on 1 and 8 DF, p-value: 0.8504

> source("C:/Users/XuWei/OneDrive/BU - Boston University/04 - 2022 Spring Semester/MET CS 555 A2 Data Analysis and Visualization with R/Lab - Session 3 (Group work)/Lab - Session 3 (Group wrok).R")

> predict(model, newdata = p)

1

29.64582