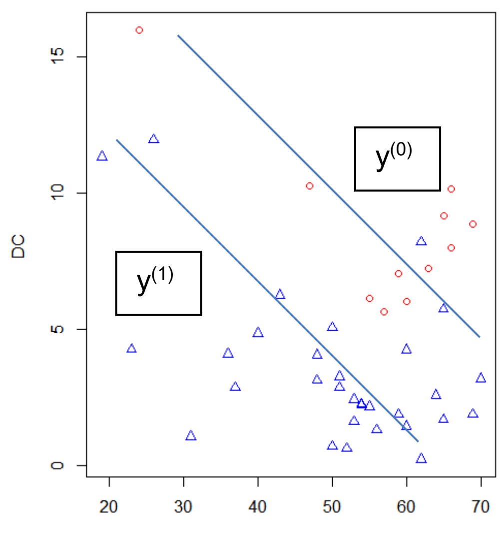
MATH 4310 **Name:** Solution

Test #3

Problem 1. For subjects undergoing stem cell transplants, dendritic cells (DCs) are critical to the generation of immunologic tumor responses. A study is conducted on 44 subjects who underwent a medical intervention and the outcome variable is the concentration of **DC cells**. One of the independent variables is the **age** of the subject (AGE), and the second independent variable is the **mobilization method** which is either Method 0 or Method 1. Data are stored in test31.csv



a) Write the regression equation using the multiple linear regression with dummy variables (dummy variable x_2).
b) Write the regression equations when the dummy variable takes the values x_2 =0 and x_2=1.
c) Sketch the regression lines
d) Perform the hypothesis test to validate the regression model, that is, test the hypothesis H_0: $\beta\_1 = 0$ vs H_1: $\beta\_1 \neq 0$ using significance level $\alpha = 0.05$.
e) Compute the 95\% confidence interval of the regression coefficient β_1.

<u>SOLUTION</u>

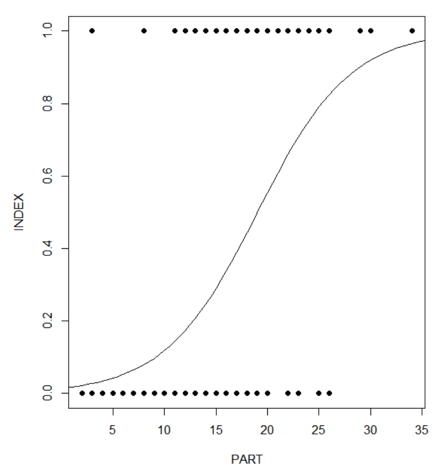
```
> Prob1 <- read.csv("C:/Users/test3_1.csv")
> x1 <- Prob1$AGE
> x2 <- Prob1$METHOD
> y <- Prob1$DC
> relation <- lm(y~x1+x2, data = Prob1)
> print(summary(relation))
Call:
lm(formula = y ~ x1 + x2, data = Prob1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7075 -1.4134 -0.5344  1.4642  6.0338

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.23907    1.64913   9.241 1.41e-11 ***
x1          -0.11585    0.02597  -4.461 6.23e-05 ***
x2          -5.88035    0.84596  -6.951 1.93e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.347 on 41 degrees of freedom
Multiple R-squared:  0.5768,   Adjusted R-squared:  0.5562
F-statistic: 27.94 on 2 and 41 DF,  p-value: 2.206e-08
```

```
> confint(relation, level=0.95)
                  2.5 %      97.5 %
(Intercept) 11.9085761 18.5695634
x1          -0.1682921 -0.0634029
x2          -7.5887930 -4.1719009
```

SOLUTION Prob 1

a) **$y = 15.239 - 0.116\ x1\ - 5.880\ x2$**
b) **@ x2=0: $y^{(0)} = 15.239 - 0.116\ x1$**
   @ x2=1: $y^{(1)} = 9.359 - 0.116\ x1$
c) **sketch is on the page above**
d) **Test hypothesis H_0: beta1 = 0 vs H_1: beta1 ≠ 0**
   We reject H_0 since p-value = 6.23e-05 < 0.05
e) **C.I. = -0.116 ± t(0.95;41)*0.026; t(0.95;41)=2.020**
   C.I = (-0.168,-0.063)

Problem 2. A study explores the relationship between the Hospital Anxiety and Depression **index** (INDEX) and **participation** in a post-surgery rehabilitation program (PART=1, if participated, and PART=0, if not). We wish to predict the likelihood of participation to the program if we know the patient Hospital Anxiety and Depression index. Data are stored in test32.csv



a) Compute and report the logistic regression equation.
b) Test the null hypothesis H0: β1 = 0 vs H1: β1 ≠ 0 at significance level 0.05
c) Compute the odds ratio.
d) Compute the 95% confidence interval for odds ratio
e) Can you conclude the odds that a woman with a high index score will participate are higher that the odds that a woman with a low index score will participate in a rehabilitation program? Explain.

SOLUTION

```
> Prob2 <- read.csv("C:/Users/test3_2.csv")
> INDEX <- Prob2$INDEX    #Y
> PART <- Prob2$PART      #X
> logit_mod <- glm(PART~INDEX, family="binomial", data = Prob2)
> summary(logit_mod)

Call:
glm(formula = PART ~ INDEX, family = "binomial", data = pro2)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8676  -0.7524  -0.4772   0.7558   2.6802

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.23095    0.65608  -6.449 1.13e-10 ***
INDEX        0.22241    0.03806   5.844 5.11e-09 ***

> (coefficients(logit_mod))
(Intercept)       INDEX
 0.01453861  1.24908442

> exp(confint(logit_mod, level=0.95))

                  2.5 %    97.5 %
(Intercept) 0.003646643 0.0483399
INDEX       1.164784203 1.3532355
```

SOLUTION Prob2

a) **ln(p/(1-p)) = -4.231 + 0.222 x1.**
b) **Since the p-value is 5.11e-09, we reject H0 and accept the alternative hypothesis that beta1 is different from 0**
c) **odds ratio = 1.249**
d) **CI = (1.165, 1.353)**
e) **Since the p-value=5.11e-09, we rejected the null hypothesis. This can also be seen by the CI of the odds ratio above not containing the value 1. Hence, we conclude that the odds that a woman with a high index score will participate are higher that the odds that a woman with a low index score will participate in a rehabilitation program.**

Problem 3. A pharmaceutical company administered three different vaccines (type A, B, C) to 6 individuals each and measured the antibody presence in their blood after a week. Results are stored in the file test33.csv

(i) Apply an appropriate non-parametric method to test the alternative hypothesis that the antibody responses are different at level of significance 0.05.

(ii) If the test is significant, run a post-hoc analysis and analyze the outcome.

SOLUTION

Part (i): We apply the Kruskal–Wallis test

```
> Data$Type = factor(Data$Type,levels=unique(Data$Type))
> kruskal.test(Antibodies ~ Type,data = Data)

        Kruskal-Wallis rank sum test

data:  Antibodies by Type
Kruskal-Wallis chi-squared = 8.2222, df = 2, p-value = 0.01639
```

Conclusion: Since the p-value is less than 0.05, we accept the alternative hypothesis that the Antibody values are different in the 3 Types

Part (ii): Since the Kruskal–Wallis test is significant, we run the Dunn test as post-hoc test

```
 > dunnTest(Antibodies ~ Type, data=Data,method="bh")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Benjamini-Hochberg method.

  Comparison          Z      P.unadj       P.adj
1      A - B  2.8118380 0.004925931 0.01477779
2      A - C  0.9192547 0.357962356 0.35796236
3      B - C -1.8925832 0.058413313 0.08761997
```

SOLUTION Prob3

(i) Since the p-value = 0.01639 is less than 0.05, we accept the alternative hypothesis

(ii) Since only the p-value of the A-B comparison is less than 0.05, we conclude that only the difference between the antibodies associated with Types A and B are significantly different.