#Q1.

1.

nfl1=data.frame(nfl1)

str(nfl1)

summary(nfl1)

attach(nfl1)

names(nfl1)

library(Hmisc)

# plots for all varibales:

model1= lm(y~x1 , data=nfl1)

model1

summary(model1)

plot(model1)

model3= lm(y~x3 , data=nfl1)

model3

summary(model3)

plot(model3)

model4 =lm(y~x4 , data=nfl1)

model4

summary(model4)

plot(model4)

model5 =lm(y~x5 , data=nfl1)

model5

summary(model5)

plot(model5)

model6 = lm(y~x6 ,data=nfl1)

model6

summary(model6)

plot(model6)

model7 = lm(y~x7 ,data=nfl1)

model7

summary(model7)

plot(model7)

model8 = lm(y~x8 ,data=nfl1)

model8

summary(model8)

plot(model8)

model9 = lm(y~x9 ,data=nfl1)

model9

summary(model9)

plot(model9)

Insert the table here

**Conclusion: based on above table I can say that x1,x7 and x8 are the most significant variables**

#2. Analysis of variance using the 3 most relevant IV.

attach(nfl1)

combined\_data = data.frame(cbind(x1,x7,x8))

str(combined\_data)

stack\_data = stack(combined\_data)

stack\_data

Anova\_results = aov(values~ind, data=stack\_data)

Anova\_results

summary(Anova\_results)

#RESULT == F(2, 81) = 417, p<2e-16(\*\*\*)

Result:

summary(Anova\_results)

Df Sum Sq Mean Sq F value Pr(>F)

ind 2 78601450 39300725 417.4 <2e-16 \*\*\*

Residuals 81 7626284 94152

---

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

**Conclusion: based on above analysis of variance I can say that**

3. *t* statistics for testing the hypotheses for the three coefficients:

1. help(t.test)

t.test(y,x1,paired=TRUE)

data: y and x1

t = -28.435, df = 27, p-value < 2.2e-16

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

95 percent confidence interval:

-2246.937 -1944.492

sample estimates: mean of the differences -2095.714

2. help(t.test)

t.test(y,x7,paired=TRUE)

data: y and x7

t = -19.651, df = 27, p-value < 2.2e-16

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

95 percent confidence interval:

-48.21564 -39.09865

sample estimates: mean of the differences -43.65714

3.data: y and x8

t = -30.621, df = 27, p-value < 2.2e-16

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

95 percent confidence interval:

-2236.067 -1955.218

sample estimates: mean of the differences -2095.643

conclusion: In all the three cases we reject the null hypothesis since p value is less that 0.05 and the mean of all the variables is not zero.

Q5:

Xbar +or – t \* standard error.

58.25