STAT 614 TEST 3 LINEAR and MULTIPLE LINEAR REGRESSION

Linear Regression (1 – 9)

**Attending Class** The following data represent the number of days absent and the final grade for a sample of college students in a general education course at a large state university

Numberofabscences FinalGrade,

0, 89.2,

1, 86.4,

2, 83.5,

3, 81.1,

4, 78.2,

5, 73.9,

6, 64.3,

7, 71.8,

8, 65.5,

9, 66.2

# 1) Using R code to create a data table and assign the data table the variable

#  **FG**

# 2) The Numberofabscences is the x variable and Final Grade is the y variable. Which

# variable therefore is the explanatory variable ?

# 3) Use R code (ggplot method) to create a scatterplot for the data.

# 4) How would you rate the linear relationship of the data points in the scatter plot?

# weak strong very strong perfect

# 5) Use and show R code to produce the slope and the intercept of the

# Linear Regression Model, and then write(type) the equation indicating the

# response variable, the intercept, the explanatory variable and its slope

# coefficient.

# 6) Use and show R code that will produce the summary table for your

# model

#7) In the summary table;

# a) Find the value that estimates the coefficient of the slope of the population.

# b) Find the value that gives a typical distance between the population

# slope coefficient and a sampled slope coefficient.

# c) Can we reject the null hypothesis that the true slope coefficient of

# the population is 0 ? Justify your answer (two or three sentences)

# d) What proportion of the variability in Final Grade is explained by you model?

# e) Find the correlation coefficient.

# f) Is the full model significant? Justify your answer. (one or two sentences)

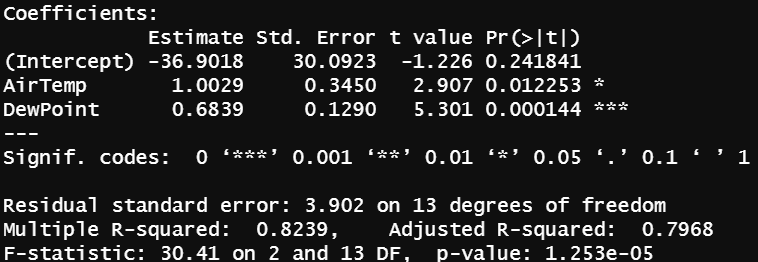
#8) Find a 95% confidence interval for the population slope coefficient. (Show all of your steps

# as demonstrated in class)

#9) Find the residual for the Final Grade value of 81.1 for 3 absences in the table. Is the

# Final Grade value of 81.1 for 3 absences above or below average? (Show all of your work)

Multiple Linear Regression (10-18)

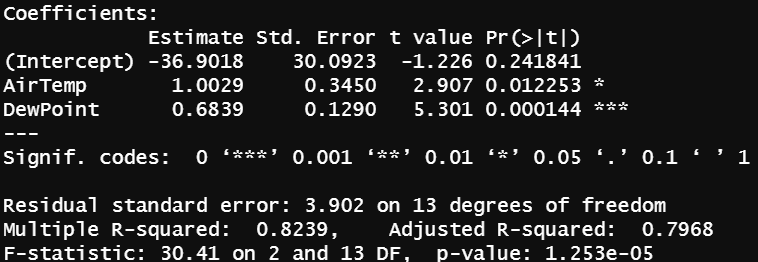


The summary table above give results for the model **HeatIndex = 1.0029AirTemp + 0.06839 DewPoint**

10) Is the overall model significant? Justify your answer.

11) Interpret the slope for **AirTemp** for this multiple linear regression model.

12) Is each individual explanatory variable’s coefficient significant?



13) What proportion of the variability of HeatIndex is explained by only explanatory variables that impact the response variable?

14) Explain what the summary table value 3.902 tells us about the model. (Three or four sentences)

15) If you wanted to introduce an interaction term to the model, which expression below would be used?

a) AirTemp/DewPoint

b) AirTemp + DewPoint

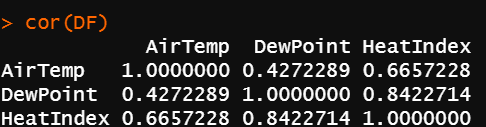
c) AirTemp – Dewpoint

d) AirTemp\*Dewpoint

e) AirTemp\*AirTemp

16) Is it true that for any Multiple Linear Regression Model, the interaction term is always significant ?

17) Does the correlation matrix below indicate any multicollinearity problems involving the explanatory variables ? Justify your answer.



18) Using **five or six sentences** evaluate your model by using results of the summary table. Indicate if you would use the model to predict **HeatIndex**