**STAT 441/541 Statistical Methods II**

**Handout: Multiple Regression Part 2**

**Ag Data Example**

**NOTE: Part 2 extends our analysis in Part 1 and uses the same scenario and dataset**

**The dataset is Excel file: Ag Data.xlsx**

A soil scientist wants to relate the daily evaporation from the soil to soil temperature, air temperature, relative humidity, and total wind. The scientist collects data at a number of locations in Texas on the variables maximum, minimum, and average soil temperature ; maximum, minimum, and average air temperature ; maximum, minimum, and average relative humidity ; and total wind . The response is the daily amount of evaporation from the soil .

A subject matter expert recommended looking at independent variables related to air temperature and relative humidity by fitting the following model:

(a) Assess the goodness of fit of the regression model. Paste the summary table (below the coefficients table) from R output. Looking at the summary table:

(i) What is the model standard deviation?

(ii) Give the value and interpretation for Adjusted R-squared:

(b) Assess multicollinearity of the independent variables. Paste the VIF values from R output:

(i) Give the VIF value for each independent variable and its interpretation:

(ii) Should we be concerned about multicollinearity? Justify your answer.

(c) Predict new *y* values using the estimated multiple regression model when x6=90 and x9=70:

(i) Paste and interpret the confidence interval from R output:

(ii) Paste and interpret the prediction interval from R output:

(d) Check assumptions for regression analysis. The assumptions are:

The model has been properly specified

The variance of the errors is for all observations

The errors are independent

The errors are normally distributed and there are no outliers

(i) Paste and interpret the Residuals vs Fitted plot from R output:

(ii) Paste and interpret the Normal Q-Q plot from R output:

(iii) Paste and interpret the Scale-Location plot from R output:

(iv) Paste and interpret the Residuals vs Leverage plot from R output:

(v) Paste and interpret the histogram of residuals from R output:

(vi) Paste and interpret the boxplot of residuals from R output:

(vii) Paste the Shapiro-Wilk test from R output and perform a hypothesis test at a significance level of :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

(viii) Paste the Breusch-Pagan test from R output and perform a hypothesis test at a significance level of :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

(ix) Are there any potential outliers identified by the influence.measures function? If so, which observations are flagged with an asterisk and why has it been flagged?

(x) Have the assumptions been met? Justify your answer.

(e) What is your overall conclusion about this multiple regression model?

(f) How would you proceed with the analysis?