**STAT 501 – Homework 8 – Covering Lesson 9 – Due Date Oct 25**

**Instructions**: Use Word to type your answers within this document. Then, submit your answers in the appropriate dropbox in ANGEL by the due date. The point distribution is located next to each question. If there are multiple parts, then the points are divided equally over the subparts.

1. **(6X4= 24 points)** Use the “Solution Concentrations” dataset in the Lesson 9 folder. Let *y* = concentration of a chemical solution and *x* = time (hours) since preparation of the solution. There are *n* = 15 observations with 3 observations at each of five different times (unique times are 1, 3, 5, 7, and 9 hours).
   1. The first two columns of the dataset give *y* and *x*. Graph *y* versus *x*. What are the noteworthy features of this plot? For instance, what is the direction of the association, are the data linear or curvilinear, is there any non-constant variance issue?
   2. Use statistical software to fit a straight-line model to the data using *y* = concentration and *x* = time. As part of doing the regression, request a graph of residuals. [In Minitab, use Stat > Regression > Regression; in the dialog box for setting up the regression, click Graphs to request the residual plot.] Write a brief interpretation of the “residuals versus fits” plot. Describe the model difficulties a residual plot with the pattern that you see indicates.
   3. Refer back to the graphs for the previous two parts. Explain why it would be best to first try transforming the *y*-variable in this case, rather than the *x*-variable.
   4. The third column in the dataset gives the square root of *y* (the square root of the concentration). Using the square root of *y* as the response variable and time as the *x*-variable, plot the square root of *y* versus *x*, fit a regression model, and examine the plot of “residuals versus fits.” Describe your findings.
   5. The fourth column in the dataset gives ln(*y*) (the natural logarithm of y). Using ln(*y*) as the response variable and time as the *x*-variable, plot ln(*y*) versus *x*, fit a regression model, and examine the plot of “residuals versus fits.” Describe your findings.
   6. Of the models examined in the previous parts of this problem, which is the “best?” Briefly explain.
2. **(5X4=20 points)** The “Diet” dataset is a simulated dataset containing *y* = calories, *x1* = carbohydrate (carb) and *x2*= fat intake for eleven meals. It is believed that both carb and fat have a linear effect on calorie amount generated but that there is a possible interaction effect also.
   1. Use Minitab to estimate the regression equation

E(y) = β0 + β1 x1 + β2 x2 + β3 x1 x2

[The command sequence (v17) is Stat 🡪 Regression 🡪 Regression 🡪 Fit regression model 🡪 select “Model” tab 🡪 Select carb and fat under the “Predictors” box and click “Add” for “Interactions through order 2.” You should be able to see “carb\*fat” added to the list of “Terms in the model.”]

What is the fitted regression equation to predict the calorie amount using this model?

* 1. Test whether β3 in the above model is significant at the 5% significance level. What does β3 signify in the context of this problem?
  2. Derive prediction equations to compute the calorie amount based on the fat intake for a meal consisting of:

1. 20 carb units
2. 30 carb units
3. 40 carb units
4. 50 carb units

*[Hint: Use the equation from part (a), plug in each value of carb, and simplify.]*

* 1. Overlay the four regression equations in part (c) on the same graph.

[One way to do this in Minitab is to use Calc > Calculator to calculate four new variables representing the prediction equations for 20, 30, 40, and 50 carb units. Then select Graph > Scatterplot (with Connect Line) and select four pairs of Y and X variables:

* + 1. Y = prediction equation for 20 carb units and X = fat
    2. Y = prediction equation for 30 carb units and X = fat
    3. Y = prediction equation for 40 carb units and X = fat
    4. Y = prediction equation for 50 carb units and X = fat

Finally click Multiple Graphs and select “Overlaid on the same graph.”]

* 1. Use your answers to the previous parts to comment on the interaction effect between carb and fat intake on the amount of calories generated.

*[Hint: Compare the slopes of the prediction equations in part (c).]*

1. **(4X8 =32 points)** The “Transformations” dataset is a simulated dataset containing four pairs of variables (*y1*, *x1*), (*y2*, *x2*), (*y3*, *x3*), and (*y4*, *x4*). Each pair of variables is best modeled using linear regression with one of the following four models:
2. Response variable *y*, predictor variable ln(*x*) [natural logarithm]
3. Response variable ln(*y*), predictor variable *x*
4. Response variable ln(*y*), predictor variable ln(*x*)
5. Response variable *y*, predictor variables *x* and *x2*

Your task is to determine which model goes with each pair of variables. You should use each model exactly once. Include Minitab plots and output to support your conclusions.

1. **(6X4= 24 points)** Open the “Chemical Reaction Data” dataset in the Lesson 9 folder. Measurements concerning the rate of velocity of a chemical reaction for different concentrations of a substrate were taken. We are interested in characterizing the relationship between concentration *(x)* and velocity *(y)*.
   1. Fit a quadratic regression model and report the estimated regression equation. [Remember to use the Model button in the Regression dialog in Minitab v17.]
   2. Perform a statistical test of the significance of the quadratic term. Specifically, report the test statistic, d.f., *p*-value, and your conclusion in the context of the problem.
   3. Visually assess a plot of the residuals versus the fitted values and comment on any violations to the regression assumptions.
   4. Create a new variable equal to the natural logarithm of concentration. [In Minitab, use Calc > Calculator, enter a new variable name (such as lnConc) in the Store result in variable box, and in the Expression box enter ln(Concentration).] Provide a plot of velocity versus the natural logarithm of concentration. Also, fit and report the estimated equation for the following regression model:

*Veli*= β0 + β1 ln(*Conci*) + β2 (ln(*Conci*))2 + ε*i*

* 1. Perform a statistical test of the significance of the quadratic term in the previous model. Again, report the test statistic, d.f., p-value, and your conclusion in the context of the problem.
  2. Using the model in part (d), what is a 95% prediction interval for velocity at a concentration level of *e*–3? Interpret this interval in the context of the problem.