**Stat 587 – Lab 9**

**Goals:** In this lab, we will learn how to fit a linear regression. Specifically, we will see how to:

 fit the regression

 test hypotheses or construct confidence intervals for regression parameters,

 make predictions at new observations

 construct a confidence interval for the line (mean Y at given X)

 construct a prediction interval for observations (single Y at given X)

 analyze a subset of the data

 do the ANOVA lack of fit test

We will use the meat.txt data set. This is the meat pH vs time data set used as a class example. Read the data set. You will have two columns: time and pH. Create a logtime variable (see lab 5 if you don't remember how to do this).

There are two ways to fit a simple linear regression: Analyze / Fit Y by X and Analyze / Fit model. Fit Fit Y by X only fits models with one X variable. Fit model fits models with one or many X variables. They provide many of the same results, but some things are only available one way or the other. I'll tell you when you need to know which to use.

1. Use Analyze / Fit Y by X / Fit line to fit a straight-line regression. Y is pH, X is log time. If you don't remember how to use Fit Y by X, you put pH into the Y, Response box and log time into the X, Factor box, then click OK. After you see the plot of the data, click the red triangle by 'Bivariate Fit' and choose Fit line. You should see a results box looking like this:



1. Working from the top, you have:

a plot of the data, with the fitted line

The fitted equation (below linear fit)

Various summary statistics about the fit.

 The most useful (to me) is the Root Mean Square Error.

 This is the estimated sd for observations around the line

The ANOVA lack of fit test (closed by default). Click the grey triangle to open it

The ANOVA table for the regression (labeled Analysis of Variance)

 This provides a test of slope = 0

The estimated parameters, their se's and tests of parameter = 0

1. Various useful information can be obtained by clicking in the correct places:

95% Confidence intervals for the regression parameters:

 Right click inside the Parameter Estimates table, choose Columns,

 then Lower 95% and Upper 95%. Those values are added to the table

Overlay 95% confidence intervals for observations:

 Left click the red triangle by Linear Fit (below the plot, not above it), then Confid Curves Fit

Overlay 95% prediction intervals for observations:

 Left click the red triangle by Linear Fit (below the plot, not above it), then Confid Curves Individ

Residual vs predicted value plot:

 Left click the red triangle by Linear Fit (below the plot, not above it), then Plot Residuals

Save Predicted values and related info for current observations

 Left click the red triangle by Linear Fit (below the plot, not above it), then:

 Predictions: Save Predicted - to get predicted values added to the data table

 Confidence intervals: Mean Confidence Limit formula

 Prediction intervals: Indiv Confidence Limit formula

Change 95% to something else (for all but regression parameters)

 Click the red triangle by Linear Fit (below the plot, not above it), then Set Alpha level

I don't know how to get the se for a predicted value from Fit Y by X.

 You can get it from Fit Model (see below).

1. Predictions and related information for new X values

The concept is that you get predictions for new X values by adding the X value you want to the data table. **Do not add a Y value (it should be a missing value).** That’s so that you don't change the fitted regression.

All the Save Predicted values information (predictions, conf. ints, pred ints) are interactive.

 JMP explicitly indicates this when it says 'something formula'. It also works for predictions.

To get predictions for new X values, set up what you want (predictions, conf. ints, pred its) for the current observations, then

 Enter a new X value in the time column and hit enter,

 you will see the log time, predicted Y, confidence interval, prediction interval, appear

1. Using Fit model to fit a simple linear regression

Analyze / Fit Model, put pH in the Y box and log time in the Construct Model Effects box

click Run

The output is similar to that from Fit Y by X and is labeled in the same way. The Lack of Fit test and residual vs predicted value plot are shown by default.

1. Additional information is obtained from Fit Model in a slightly different way. Look for the red triangle at the very top, by Response pH. The Save Columns option provides more choices than you got in Fit Y by X. You will see some options that are 'something Values' and others that are 'the same thing Formula'. I suggest the Formula version because that allows you to add X values to the table and get updated information. If you have a very large sample size, you might prefer the “value” version because it will be faster.

Predictions: Save Columns / Prediction Formula

Se of the Prediction: Save Columns / Std Err Pred Formula

Confidence intervals: Mean Confidence Limit Formula

Prediction Intervals: Indiv Confidence Limit Formula

All of these are “formula” versions so if you add new X values to the data table, you get info.

As far as I can tell, the intervals are only 95% intervals.

Some other possible columns are “value” versions. Two useful ones are:

Se of a predicted observation: Save Columns / Std Error of Individual

Se of a predicted mean (the line), non-formula version: Save Columns / Std Error of Predicted

If the table does not update when you add a new X value, you probably chose the 'Value' version, not the 'Formula' version.

You can always use the 'Value' versions, but you must provide all your new X values, then chose what you want displayed.