

Stat 328 Lab #6 Summer 2000

1. In your textbook on page 475, there is a data set with a 0-1 response variable, $y =$ hiring status. Enter that data set into JMP. Then enter or compute $y' = 1 - y$ and make y' a JMP nominal variable. (Make x_3 a JMP continuous variable.) Use the JMP facility for logistic regression to fit the model

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

to these data, where $p =$ the probability of being hired $=$ the probability y is 1 $=$ the probability that the nominal variable y' takes its "first value" (namely, $y' = 0$).

(a) Compare the printout you get from JMP to the one on page 484 of the textbook. In particular, compare the estimates of the β 's and their standard errors.

(b) Add 95% confidence intervals for the β 's to the JMP report (under the check mark in JMP-IN). Is it clear from these data that each of the variables x_1 , x_2 and x_3 have an impact on the probability of being hired in this scenario?

(c) Save the "prediction formula" to the JMP data table, add 2 rows to the table and evaluate the fitted/estimated probability of being hired for $x_1 = 5$, $x_2 = 5$ and $x_3 = 0$ and for $x_1 = 5$, $x_2 = 5$ and $x_3 = 1$.

(d) First for females and then for males, plot in the (x_1, x_2) -plane the line where the fitted model says there is a .5 probability of being hired. (Note that the combinations of x_1 and x_2 that have fitted probability of being hired are "above and to the right" of these lines.)

2. Vardeman gave you real control charting data sets for fractions of broken tablets (in a table labeled Table 3.3) and for leaks detected in some radiator inspections (in a table labeled Table 3.4). (He also passed out different simulated data sets and JMP printouts for them made from truly stable processes with characteristics somewhat like the real data sets.) Enter the real data sets into JMP.

(a) Use JMP's control chart facility to decide if there is evidence in the data of process change over the study periods. (For the first of these, make a p -chart and for the second, make a u -chart. Don't give JMP values for p and λ .)

(b) Verify by hand that the center lines on the charts in (a) are at respectively

$$\hat{p}_{\text{Pooled}} = \frac{\text{total defective}}{\text{total inspected}}$$

and at

$$\hat{\lambda}_{\text{Pooled}} = \frac{\text{total defects}}{\text{total inspected}}$$

and that control limits are as laid out on the Stat 328 formula sheets (once the above estimates are substituted for the parameters p and λ).

3. Start with a blank JMP data table, add 200 rows and fill Column 1 with simulated independent standard normal values. (Highlight the whole Column 1 heading, then in the "Cols" menu click on the "Column Info" item to bring up a dialogue box where you should choose "formula." That will bring up the calculator and you can choose "Normal" under the "Random" menu. Clicking "evaluate" will then put the simulated values into Column 1.)

(a) Add a column of row numbers (a "time" variable) to the data table and plot the simulated values against row number/time. (Use the "overlay plots" option in the "graph" menu.) This is meant to illustrate the appearance of observations from a stable process producing measurements.

(b) If we thought of the plotted points in (a) as "averages of samples of size $n = 1$," what would be "standards given" ($\mu = 0, \sigma = 1$) Shewhart control limits for them? Are there points outside control limits/evidence of process change in your plot from (a)?

(c) Run the "Fit Y by X" (using the row number variable as X) JMP procedure on these "data." Try smoothing these data with several different splines (including, say, $\lambda = 10$ and $\lambda = 10,000$) and with several different polynomials (including, say, a line and a 5th degree polynomial) and note that there is (as expected) no real trend in these data.

(d) Make up "lag 1," "lag 2," and "lag 3" versions of the series of simulated independent standard normal values. (Make up new columns using "formulas" where the formulas are the column 1 name you use, with subscripts " $i - 1$," " $i - 2$," and " $i - 3$." Compute correlations among the original series and these 3 lagged series using the JMP "Correlations" (or "Multivariate") facility. Are these correlations large?

(e) Use the JMP "Fit Model" function and regress the simulated values in column 1 against their own lagged versions. Save the predicted values in the data table and plot both the original values and the predicted values on the same set of axes using the "Overlay Plots" option. How well can one do predicting the simulated values from previous simulated values?

4. As in Chapter 16 of the JMP manual, load the JMP time series data set "seriesal.jmp" into the data table. (This has a time series " C_n ," and 8 lagged versions of the time series in it.)

(a) Plot the series C_n against a time variable. Note that this doesn't look like the "stable process" series in exercise 3 above. There is "pattern" here to be described and possibly exploited to make forecasts.

(b) Fit (using "Fit Y by X") a spline with $\lambda = 10,000$ and a 5th order polynomial to this series. Note that these are nearly the same. Save the predicted values and residuals from one of these fits. Plot the residuals against time. Note that there is no real "long term" trend left in these residuals (try fitting polynomials or splines with large λ). But there *is perhaps* still some pattern here that could be modeled. Make lag 1, lag 2 and lag 3 versions of your residual series. Compute the correlations among these. How do these compare to the correlations you saw in 3 (d)?

(c) Regress the residuals from part (b) on their lag 1, lag 2 and lag 3 versions (do a MLR) and save the fitted values. Add those to the predicted values you saved in part (b). Then use the overlay plot facility to plot the original series, the series of predictions from (b) and these new predictions against time. Note that the new predictions capture back some of the jaggedness smoothed out earlier, but that at

those times that the original series "changes direction" even the new predictions of necessity "lag" behind the movement of the original series.

5. As in Chapter 16 of the JMP manual, load the JMP time series data set "seriesg.jmp" into a JMP data table.

(a) Plot the Passengers time series against time. Note the nonlinear trend, the increasing amplitude of the annual pattern, and the clear regularity of that annual pattern. To put this problem on a scale where the trend is hopefully closer to linear and the amplitude of the annual pattern might be closer to constant, take the common (base 10) log of the passenger counts. Plot those against time and note that this plot looks more like a single yearly pattern shifted up by a constant amount each year.

(b) One way to try to capture the shape of the annual pattern (in the logpassengers) is through the use of 11 dummy variables. Rather than making those up by hand, we can simply use the JMP "season" nominal variable that gives month numbers. Fit a MLR model to the log passenger values with a "time" term and using the dummies that JMP automatically creates when you use the nominal variable "season." Save the prediction formula. (Doing this instead of saving only the predicted values will allow JMP to make predictions/forecasts beyond the rows in the data table where there are observations.) Raise 10 to the power of the prediction of logpassengers. Plot on one set of axes logpassengers and predicted logpassengers against time and on another set passengers and predicted passengers against time. Note that you can indeed forecast sensibly in time with this simply MLR analysis.

(c) Is there any correlation structure left in the logpassenger residuals after taking out the linear time trend and the seasonal effects? (Compute several lags of the residuals and look at correlations.) Note that you could follow up here with the kind of regression of residuals on lagged residuals and improved predictions investigated in part (c) of the previous problem in this context as well, but it isn't required that you do so for purposes of completing this lab.