

36-401 Homework #6 Due 10/19/00

1. (10 pts) Do RwG problem 3-15 on page 104. The data is in salamander.dat. **Turn in** the table of coefficients and p-values, the graph, **and** an interpretation of the results in terms of the *substantive meaning* of the variables. The graph should mark the points separately for lunged vs lungless as well as showing the two regression lines. An Splus code example for this type of graphing is given in spruce.q.
2. (40 pts) Re-examine the wages dataset, now including the race and sector categorical variables (be sure to convert them to factors). Examine the univariate distributions of the two new variables. Re-examine the bivariate distributions, now including race and sector. Build a model using all of the cases to predict $\log_{10}(\text{wage})$. Consider adding any interactions that would make sense. Perform appropriate model selection according to the BIC criterion. (To include categorical variables in `sum.step()`, you need to create new $X_1 * X_2$ columns in your data.frame; if the categorical variables have $R > 2$ levels, you need to create $R - 1$ new variables in your data.frame.)

Turn in:

- (10 pts) A log of the work you performed, E.g. “I looked at a histogram of sector”, “Variable selection using the BIC criterion was performed”, “A — by — interaction was tried, but found not to be significant”.
- (10 pts) A description of your final model or models including the prediction equation with numeric coefficients (with an appropriate number of decimal places) and a paragraph giving your interpretation of what the model tells us about the U.S. wage structure.
- (10 pts) A multi-panel plot or plots that can be used for assumption checking of your best model. Include a few sentences with your interpretation of the plots.
- (10 pts) A representative conditional effects plot of your choice. Pick one continuous covariate for the x-axis. Put “predicted wage” on the y-axis. Plot a small number of curves each representing a few well chosen values for the other covariates in your model. Label the curves with the values for those other covariates.