

Stat 328 Lab #4 Summer 2000

On the Stat 328 Web Page you will find a fairly large data set titled "Wisconsin Restaurant Data." There are 181 cases of responses made on the 1980 Wisconsin Restaurant Survey conducted by the University of Wisconsin Small Business Development Center. Recorded are values for:

"outlook"	a business outlook rating (1 = very unfavorable to 7 = very favorable)
"sales"	gross 1979 sales (\$1000)
"newcap"	new capital invested in 1979 (\$1000)
"value"	estimated market value of the business (\$1000)
"costgood"	cost of goods sold as a percentage of the sales
"wages"	wages as a percentage of sales
"ads"	advertising as a percentage of sales
"typefood"	a code for type of food sold (1 = fast food, 2 = supper club, 3 = other)
"seats"	number of seats in the dining area
"owner"	a code for type of ownership (1 = sole proprietorship, 2 = partnership, 3 = corporation)
"FTemploy"	number of full time employees
"PTemploy"	number of part time employees
"size"	a code summarizing the number of employees (1 = 1 to 9.5 employees, 2 = 10 to 20 employees, 3 = over 20 employees, where a part time employee counts as .5)

Download the data set and read it into JMP.

Your fundamental task is to build an effective model for "sales" in terms of the other variables. You have at your disposal JMP, all you've learned about MLR, and your good business sense. You may:

- transform any variable(s)
- make up new predictors from the ones above (for example, allowing for interactions that you expect on the basis of what you know about business, or combining variables into a single one ... for example replacing all of the last 3 predictors with $FTemploy + .5(PTemploy)$ looks to me like it might be sensible)

Your search should consider issues such as parsimony/simplicity of the final model you settle on, the possibility that a few unusual cases dominate the fitting of the model and should really be eliminated from the fitting, goodness of fit criteria like R^2 , $PRESS$, MSE and C_p , sensible-looking residuals, etc. Be sure that if you use the qualitative variables above in your analysis, you handle them in a sensible manner (for example, you surely don't want to treat the codes for "typefood" as if they were values of a measured variable).

When you have settled on a model you'd be willing to use in a report to the Governor of Wisconsin (about variables related to revenue for restaurants in the state) compute and plot residuals against all of the predictors and against \hat{y} , and normal plot them as means of model checking. Then make 95% prediction limits for all the different sets of predictors you retain in your data set. (If you end up modeling a transform of y , compute the limits on the transformed scale and then "untransform" them to get limits on the original scale.) How do your predictions agree with the original y 's? Write a paragraph or two interpreting your model (to the extent that is possible) to the proverbial "layman." (What variables are the best predictors of revenue, etc.)

There is (obviously) no "right" answer here. Do something sensible and document and defend it.