

Math 484: Lab 3 Monday, Oct. 17

A) *Variable selection with SAS: problem 3.30*: Copy and paste the *SAS* program for this problem from (<http://parker.ad.siu.edu/Olive/lreghw.txt>) into the *SAS* editor. Then perform the menu commands “Run>Submit” to obtain about 15 pages of output. Do not print out the output.

The key *SAS* code is shown below.

```
proc reg data=fitness;
  model Oxygen=Age Weight RunTime RunPulse RestPulse MaxPulse;
  output out =a p = pred r = resid;
  model Oxygen=Age Weight RunTime RunPulse RestPulse MaxPulse
    / selection=forward;
  model Oxygen=Age Weight RunTime RunPulse RestPulse MaxPulse
    / selection=backward;
  model Oxygen=Age Weight RunTime RunPulse RestPulse MaxPulse
    / selection=cp best = 10;
run;

proc rsquare cp data = fitness;
model Oxygen=Age Weight RunTime RunPulse RestPulse MaxPulse;

proc plot data = a;
  plot resid*(pred);
  plot Oxygen*pred;

proc reg data=fitness;
  model Oxygen=Age RunTime RunPulse MaxPulse;
  output out =sub p = pred r = resid;

proc plot data = sub;
  plot resid*(pred);
  plot Oxygen*pred;
run;
```

The data is from SAS Institute (1985, p. 695-704, 717-718). Aerobic fitness is being measured by the ability to consume oxygen. The response  $Y = \text{Oxygen}$  (uptake rate) is expensive to measure, and it is hoped that the OLS  $\hat{Y}$  can be used instead. The variables are *Age* in years, *Weight* in kg, *RunTime* = time in minutes to run 1.5 miles, *RunPulse* = heart rate when  $Y$  is measured, *RestPulse* = heart rate while running and *MaxPulse* = maximum heart rate recorded while running.

The *selection* commands do forward selection, backward elimination and all subset selection where the best ten models with the lowest  $C_p$  are recorded. The *proc rsquare* command also does all subsets regression with the  $C_p$  criterion.

The plots give the response and residual plots for the full model and the submodel that used *Age*, *RunTime*, *RunPulse*, *MaxPulse* and a constant, the minimum  $C_p$  model.

B) *Variable Selection with Minitab: 3.31*: Get the data set *prof.mtb* as described in Problem 2.15. The data is described in McKenzie and Goldman (1999, p. ED-22-ED-23). Assign the response variable to be *instrucr* (the instructor rating from course evaluations) and the predictors to be *interest* in the course, *manner* of the instructor, and *course* = rating of the course.

a) To get residual and response plots you need to store the residuals and fitted values. Use the menu commands “Stat>Regression>Regression” to get the regression window. Put *instrucr* in the **Response** and *interest*, *manner* and *course* in the **Predictors** boxes. The click on **Storage**. From the resulting window click on **Fits** and **Residuals**. Then click on **OK** twice.

b) To get a response plot, use the commands “Graph>Plot,” (double click) place *instrucr* in the **Y** box, and *Fits1* in the **X** box. Then click on **OK**. Print the plot by clicking on the graph and then clicking on the printer icon.

c) To make a residual plot, use the menu commands “Graph>Plot” to get a window. Place “Resid1” in the **Y** box and “Fits1” in the **X** box. Then click on **OK**. Print the plot by clicking on the graph and then clicking on the printer icon.

d) To perform all subsets regression, use the menu commands “Stat>Regression>Best Subsets” to get the regression window. Put *instrucr* in the **Response** and *interest*, *manner* and *course* in the **Free predictors** boxes. The submodel that uses *manner* and *course* looks good and could be checked with response and residual plots.

e) To possibly perform stepwise selection, use the menu commands “Stat>Regression>Stepwise” to get the regression window. Put *instrucr* in the **Response** and *interest*, *manner* and *course* in the **Free predictors** boxes.

C) In *Arc* enter the menu commands “File>Load>Data>ARCG” and open the file *wool.lsp*. From the menu “Wool” select “Transform” and double click on *Cycles*. Select the log transformation. Use the menu commands “Graph&Fit>Fit linear LS” to obtain a dialog window. Use *Amp*, *Len* and *Load* as the predictors and *Cycles* as the response. Repeat with  $\log(\text{cycles})$  as the response. Use the menu commands “Graph&Fit>Plot of” to create a dialog window. Double click on L1:Fit-Values and double click on L2:Fit-Values. Note that the plot is linear with very high correlation.