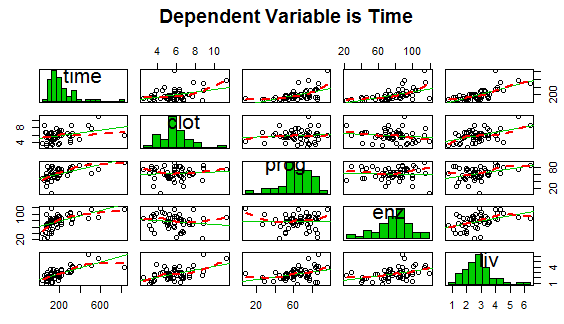
**James Young Homework 5**

**1. State the model**

,

**2. Produce ONE scatterplot of all variables against the dependent variable. Make y is the first entry in scatterplot in top left corner. Explain in one sentence the visual assessment of linearity of each variable with respect to the dependent variable**

All 4 independent variables seem to have some

week positive correlation, however “liv” seems

most linear visually while the others seem visually

less linear at this point.

**3. Fit the multiple regression using R and provide**

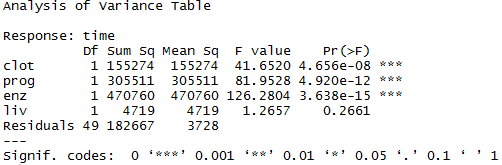
**the snip of the coefficients**



**4. Fit the estimated regression function (use the ‘hat’ function) using your results from step 3**

= -621.597550 +33.163828x+4.271860x+4.125738x+14.091563x

**5. Use the F value in step 3 to test if all β n is equal to zero. State the hypothesis, use the R output and find the F test statistic, and provide the decision rule, and conclusion. There should only be one hypothesis test.**

H0: = = = 0

H1: at least one 0

The critical F value for 1 DF in num. and 49 DF

in denom. is [4.03,4.06]. The F value of clot,

prog, and enz is > [4.03,4.06]. The F value of liv

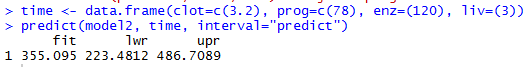
is < [4.03,4.06]. Therefore, we reject H0 and

conclude the first 3 independent variables are significant and contribute to the predictive power of the model.

**6. In 1-2 sentences interpret each coefficient**

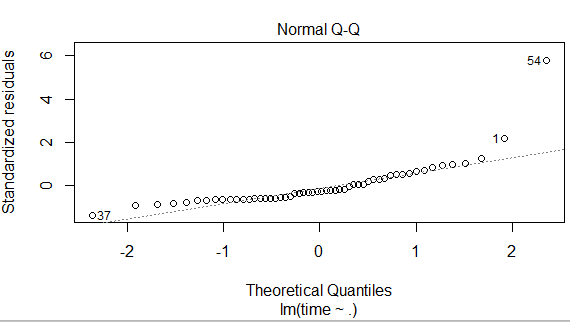
“Clot”, “prog”, and “enz” are all significant based upon F tests meaning these independent variables play an important role in the value of response variable “time”. The independent variable “liv” was seen to be insignificant based upon F tests and therefore is not as influential in determining the value of response variable “time”.

**7. Use the code below and provide snip of the prediction intervals and predict when clot=3.2,prog=78,enz=120,liv=3. State in one sentence what the prediction means.**



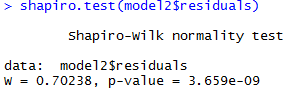
We can predict with 95% confidence the patients survival time will be between 223.4812 and 486.7089 with a true mean time of 355.095.

**8. Provide the only the qqplot with the qqline. Examine the residuals distribution from the qqplot and comment on the distribution**

The qqline plot appears normal in the bulk but

the tails deviate slightly especially point 1 and

34 near quantile 2.

**9. Perform the Shapiro-Wilks Test to test the normality. You need hypothesis test, R snip with the result and the decision rule and conclusion**

Shapiro-Wilk Test (with )

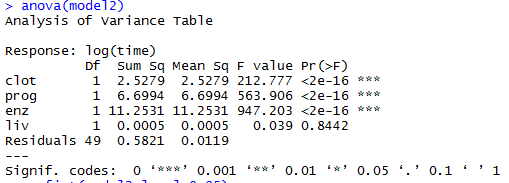
H0 : All residuals came from normal distribution

H1: Any residuals did not come from a normal distribution

Decision: Calculated p-value = 0.70238. 0.70238 > 0.01 therefore

we fail to reject H0 and conclude the residuals are from a normal distribution.

**10. Refit the regression using the log and provide a snip of the coefficients from the R output**



**11. Interpret each of the coefficients**

Independent variable “clot”, “prog”, and “enz”

had lower p-values than the original data, but

the significance code is still the same. “liv”

actually has a higher p-value now and is still

not significant.

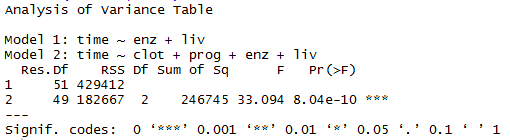
**12. Compare this model to the first model, which is a better fit and why? Use the 2 R or any other tool to assess which is a better fit.**



log

normal

Looking at the two models, the log model has a better adjusted R-squared so it is a better fit.

**13. Conduct a partial F test to test if whether env and liv are significant after taking clot and prog into consideration. You will have a full and reduced model. State the hypothesis, calculate the test statistic, and provide the decision rule, and conclusion.**

H0:

H1: at least one of the above 0

The calculated F value is 33.094 which correlates to

a p-value well below 0.05, therefore we reject H0

and conclude “clot” and “prog” do have significant

information once “enz” and “liv” are taken into consideration.

R code Appendix

