Laboratory #8.
Applying the General Linear Model - ANCOVA

The purpose of this lab is to increase your skill and facility in applying the general linear model to "data situations" where there are at least two explanatory variables. Based on what you learned from Lab 7, you will be asked to construct a model to analyze a data situation concerning growth rates of ducklings in relation to experimental diets (exercise 14.12 from Sokal and Rohlf 1995, 16.11 in S&R 2012).

Once you have completed the lab write-up, you should have

    - ability to translate a data situation into a model-based analytic format

    - ability to use a statistical package to execute any model you write

    - greater facility in applying the general linear model to data situations with more than one explanatory variable.

At this point make sure that you have the following data set:
    DuckWeights   srex1412.dat

It is available on the course website: www.mun.ca/biology/schneider/b4605/Data/Labs

Laboratory #8. Applying the General Linear Model

Analyze the DuckWeight data in exercise 14.12 from Sokal and Rohlf (1995).   SREX1412.dat
Follow the steps in the generic recipe for the execution of the general linear model.

The response variable is: _____

The explanatory variable (categorical) is : _____

The explanatory variable (regression variable) is: _____

This is an analysis of covariance –ANCOVA-- because there is at least one categorical variable
and one regression variable. ANCOVA has two uses. One is to compare slopes. The interaction
sum of squares measures the heterogeneity in slopes. The other use is statistical control via the
ratio scale variable. Duck weight it included in the analysis to control statistically for age effects.
The interaction term is not the focus of the analysis, and so not included in the model. However,
if there is substantial heterogeneity of slopes, this variability will appear in the residual,
potentially reducing the power of the analysis be increasing the error MS.

Write your model here.

_____

Note: In any statistical package you must define whether each explanatory
                variable is categorical (ANOVA type) or a regression variable.
          Some packages assume all variables are categorical (*e.g.* Minitab).
                For these packages you must declare the regression variables
                (MTB > covariate X1)
          Some packages assume all variables are regression variables (*e.g.,* SAS)
                For these packages you must declare the ANOVA variables (SAS > Class X2)
          R/S+ define each of your variables as being a factor or ratio scale variable when
                the data are imported. Make sure that R has correctly classified both of
                your explanatory variables.
Note:
          You will find a substantial bowl in the residual vs fit plot.
          The 1995 text recommends a square root transformation to correct the problem.
          Transform the data by taking the square root of the response variable,
                repeat the analysis, and plot the residuals vs fits.
          Did this correct the problem of a bowl ?
          The response variable is not readily interpretable – grams$^{1/2}$ = ?
          Try a more interpretable transformation – grams$^{1/3}$ ~ length
          Then try a more general model a power law – Grams = α Age$^{\beta}$
                To execute regress the log Weight against the log of Age.
          Finally, try a quadratic term. Analyze the <u>untransformed</u> response variable in relation
          to fishmeal diet ration in 4 categories and to age as a quadratic relation: Age + Age^2.

Write-up for this lab.

Report results of analysis of the DuckWeights data I(Sokal and Rohlf 1995).
Using the generic recipe for data analysis with the GLM, show steps 1-3a for the following sequence of models.

> wt = f(age, meal)   age is regression variable
> sqrt(wt) = f(age, meal)
> ln(wt) = f(ln(age), meal)
> wt = f(age, age^2, meal)
> $wt^{1/3}$ = f(age, meal)

Compare the transformations with respect to reduction of the initially prominent bowl in the residual vs fit plot.   Give a reason for your choice of best model with respect to the straight line assumption. Then for the model you chose complete steps 3b, 4, 7, and 9 of the generic recipe for general linear model analysis. Your write up should include
        - an evaluation of the residual vs fit plot for each model (step 3).
        - the reason for your choice of final model to reduce the bowl.
        - a plot to evaluate the normality assumption for your final model.
        - the ANOVA table for your final model.
        -a comment on the use of regression for statistical control with your final model
                -how much did it reduce the unexplained variance (SSres/SStotal) ?
                -did it increase the weight of evidence (LR = (SSres/SStotal)$^{-n/2}$) ?
                -did the use of statistical control change the decision?

1.   Construct model.
                List response and explanatory variables
                Assign symbols, state units and type of measurement scale for each variable.
                Write out statistical model.
2.   Execute model and show the plot of residuals against fitted values.
3a.   Evaluate the model, using residuals.
                Is straight line inappropriate ? If not, revise the model (back to step 1).
3b.   If n small, evaluate assumptions for using chisquare, t, or F distribution.
                        residuals homogeneous ? (residual versus fit plot).
                        residuals normal ? (histogram of residuals, quantile or normal score plot).
                If not met, empirical distribution (by randomization) may be necessary.
4.   Partition df and SS according to preferred model only. Calculate likelihood ratio for
        full (omnibus) model and for fixed effect (treatment).
7.   ANOVA: Show the ANOVA table with p-value for treatment effect.
9.   Declare decision about the treatment effect. No decision about the regression
        variable is needed, it is present in the model for statistical control.
                If $p < \alpha$ then reject H_o, report effect sizes (means), and interpret the main effects.
                If $p \geq \alpha$ then cannot reject H_o. Report the ANOVA table with explained
                variance for the fixed effect (meal).
                In either case report the test statistic with degrees of freedom and Type I error.

Extra: Run your analysis with an interaction term to investigate whether this reduces the error mean square.