**Winter, 2015 Biometry Quiz #2**

1. Data on the catch-per-unit-effort (CPE; number of fish per net per hour) of yellow perch (*Perca flavescens*) captured with gill nets in a Midwestern lake were obtained during midday at three randomly selected sites sampled at five depths during two months. The researcher’s goal was to determine if the mean CPE of Yellow Perch differed between months, among depths, or among combinations of depths and months. The data from this study were entered into R and analyzed with the commands at the end of the quiz handout. Answer the questions below **with the fullest amount of detail that you can provide – be specific and refer to results where appropriate** (you may want to label figures and tables on the output).
   1. **[6 pts]** Assess all assumptions (with the exception of independence) on the original scale.

For the next three questions, state whether you will interpret on the original or transformed scale.

* 1. **[4 pts]** What type(s) of effects are **and** are not evident in these results?
  2. **[6 pts]** For the effect(s) that you identified in (b), which group means are **and** are not statistically different? For the statistically different groups, specifically state which group mean is larger.
  3. **[4 pts]** Modifya graphic to illustrate your results from question (c) (*make sure to explicitly identify which graph(s) you marked*).
  4. **[4 pts]** For the largest difference in group means, specifically describe how the two group means compare (on the original scale, if possible).

1. **[6 pts]** Thoroughly explain why one experiment where two factors are simultaneously manipulated is “better” than two separate experiments where one factor at a time is manipulated. Demonstrate your points with an illustrative example, assume that there are 84 individuals available for experimentation and that one factor has three levels and the other factor has four levels.
2. Marqus *et al.* (2000) examined the resource abundance and insect herbivore density on woody fabaceous plants in the Sonoran Desert. In one part of their study they recorded the number of leaf herbivore species on and the leaf biomass of several woody fabaceous legumes. Their goal for this analysis was to predict the number of leaf herbivore species on a legume from the leaf biomass of the legume. The data from this study were entered into R and analyzed with the commands on the separate sheet. Answer the questions below **with the fullest amount of detail that you can provide – be specific and provide appropriate evidence for your answer** (you may want to label figures and tables on the output).
3. **[8 pts]** Choose one of the four sets of results to answer the questions below. Fully explain why you chose to use those results.
4. **[3 pts]** Is there a significant relationship evident in your results?
5. **[5 pts]** Fully interpret the slope in your results.
6. **[6 pts]** Fully interpret the two predictions in your results.
7. **[4 pts XC]** How does a one unit change of your explanatory variable effect your response variable *on the original scale*?
8. **[6 pts]** H.C. Watson demonstrated the species-area relationship for Britain’s vascular flora in 1859 (which has since been accepted for many other fauna from many other locations). Watson’s relationship describes how the number of species (N) relates to area (A), using the formula N=*a*A*b* where *a* and *b* are constants. From Watson’s formula, it is possible to estimate the amount of species that may become extinct or threatened when the area available to them is reduced by habitat destruction (due, for instance, to warming). However, these calculations cannot be made unless *a* and *b* have been established from empirical data on N and A. Explain how one would estimate *a* and *b* by applying a linearized model to data on N and A. In your explanation, you should
   1. demonstrate (*algebraically show each step of your work*) how you linearized this function,
   2. explicitly identify the response and explanatory variables of your linearized function, and
   3. explain how *a* and *b* could be determined if your linearized function was applied to data on N and A (*i.e., how do these parameters relate to the intercept and slope of the linearized function*).

***R RESULTS FOR QUESTION #1***

**> d75 <- read.table("box7\_5.txt",header=TRUE)**

**> d75$Month <- factor(d75$Month,levels=c("Jun","Aug"))**

**> d75$Depth <- factor(d75$Depth)**

**> d75$comb <- d75$Depth:d75$Month**

**> str(d75)**

'data.frame': 30 obs. of 4 variables:

$ Month: Factor w/ 2 levels "Jun","Aug": 1 1 1 1 1 1 1 1 1 1 ...

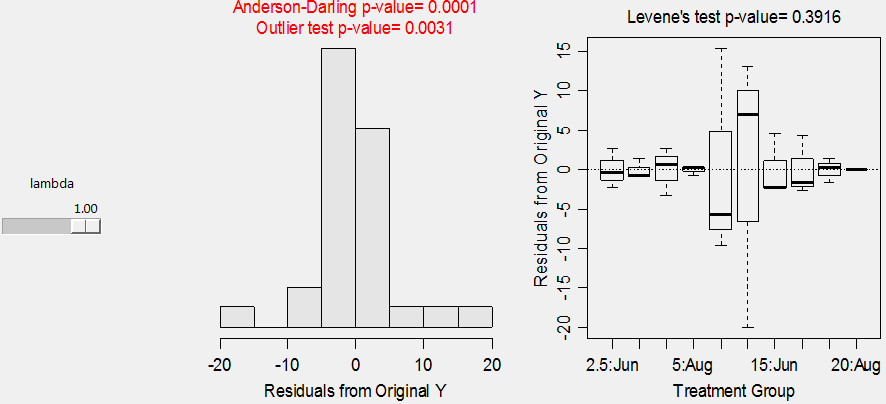
$ Depth: Factor w/ 5 levels "2.5","5","10",..: 1 1 1 2 2 2 3 3 3 4 ...

$ CPE : int 2 4 7 6 10 12 8 12 33 10 ...

$ comb : Factor w/ 10 levels "2.5:Jun","2.5:Aug",..: 1 1 1 3 3 3 5 5 5 7 ...

**> lm1 <- lm(CPE~Depth\*Month,data=d75)**

**> transChooser(lm1)**



**> anova(lm1)**

Df Sum Sq Mean Sq F value Pr(>F)

Depth 4 2249.8 562.45 10.4222 9.954e-05

Month 1 9.6 9.63 0.1785 0.6772

Depth:Month 4 400.2 100.05 1.8539 0.1582

Residuals 20 1079.3 53.97

Total 29 3739.0

**> mc1a <- glht(lm1,mcp(Depth="Tukey"))**

**> confint(mc1a)**

Estimate lwr upr

5 - 2.5 == 0 5.0000 -12.9485 22.9485

10 - 2.5 == 0 13.3333 -4.6152 31.2818

15 - 2.5 == 0 8.0000 -9.9485 25.9485

20 - 2.5 == 0 -2.6667 -20.6152 15.2818

10 - 5 == 0 8.3333 -9.6152 26.2818

15 - 5 == 0 3.0000 -14.9485 20.9485

20 - 5 == 0 -7.6667 -25.6152 10.2818

15 - 10 == 0 -5.3333 -23.2818 12.6152

20 - 10 == 0 -16.0000 -33.9485 1.9485

20 - 15 == 0 -10.6667 -28.6152 7.2818

**> mc1b <- glht(lm1,mcp(Month="Tukey"))**

**> confint(mc1b)**

Estimate lwr upr

Aug - Jun == 0 -3.6667 -16.1786 8.8453

**> lm2 <- lm(CPE~comb,data=d75)**

**> mc2 <- glht(lm2,mcp(comb="Tukey"))**

**> glhtSig(mc2)**

[1] "10:Aug - 2.5:Jun" "10:Aug - 2.5:Aug" "10:Aug - 5:Jun" "10:Aug - 5:Aug"

"15:Aug - 10:Aug" "20:Jun - 10:Aug" "20:Aug - 10:Aug"

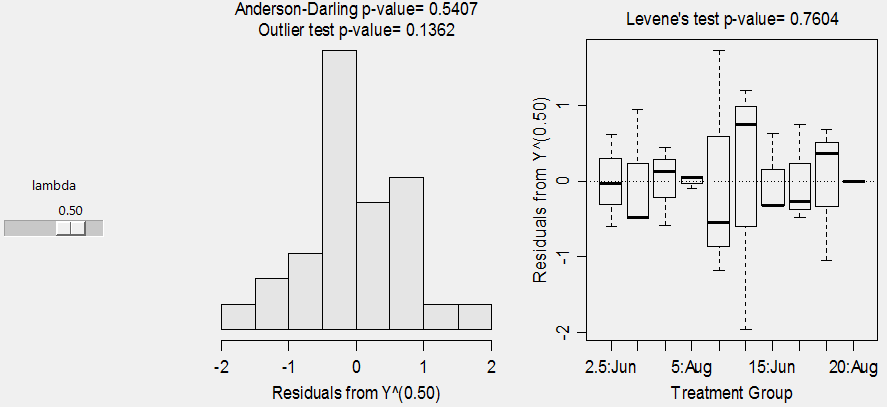
**> fitPlot(lm1,which="Depth",ylim=c(0,45)) # LEFT**

**> fitPlot(lm1,which="Month",ylim=c(0,45)) # CENTER**

**> fitPlot(lm1,ylim=c(0,45),interval=FALSE) # RIGHT**



**> transChooser(lm1)**



**> d75$tCPE <- d75$CPE^(0.5)**

**> lm3 <- lm(tCPE~Depth\*Month,data=d75)**

**> anova(lm3)**

Df Sum Sq Mean Sq F value Pr(>F)

Depth 4 68.388 17.0970 20.8113 6.743e-07 \*\*\*

Month 1 0.890 0.8896 1.0828 0.31048

Depth:Month 4 8.891 2.2228 2.7056 0.05969 .

Residuals 20 16.430 0.8215

Total 29 94.599

**> mc3a <- glht(lm3,mcp(Depth="Tukey"))**

**> confint(mc3a)**

Estimate lwr upr

5 - 2.5 == 0 1.0053 -1.2091 3.2197

10 - 2.5 == 0 1.9924 -0.2220 4.2068

15 - 2.5 == 0 1.4626 -0.7518 3.6770

20 - 2.5 == 0 -0.9712 -3.1856 1.2432

10 - 5 == 0 0.9871 -1.2273 3.2015

15 - 5 == 0 0.4573 -1.7571 2.6717

20 - 5 == 0 -1.9765 -4.1909 0.2379

15 - 10 == 0 -0.5298 -2.7442 1.6846

20 - 10 == 0 -2.9636 -5.1780 -0.7492

20 - 15 == 0 -2.4338 -4.6482 -0.2194

**> mc3b <- glht(lm3,mcp(Month="Tukey"))**

**> confint(mc3b)**

Estimate lwr upr

Aug - Jun == 0 -1.548584 -3.092313 -0.004854

**> lm4 <- lm(CPE~comb,data=d75)**

**> mc4 <- glht(lm4,mcp(comb="Tukey"))**

**> glhtSig(mc4)**

[1] "10:Aug - 2.5:Jun" "10:Aug - 2.5:Aug" "10:Aug - 5:Jun" "10:Aug - 5:Aug"

"15:Aug - 10:Aug" "20:Jun - 10:Aug” "20:Aug - 10:Aug"

**> fitPlot(lm3,which="Depth",ylim=c(0,7)) # LEFT**

**> fitPlot(lm3,which="Month",ylim=c(0,7)) # CENTER**

**> fitPlot(lm3,ylim=c(0,7),interval=FALSE) # RIGHT**



***R RESULTS FOR QUESTION #3***

**> lb <- read.csv("LeafBiomass.csv")**

**> lb$logbiomass <- log(lb$biomass)**

**> lb$logspecies <- log(lb$species)**

**> str(lb)**

'data.frame': 34 obs. of 4 variables:

$ biomass : int 1 1 1 3 2 2 3 9 21 40 ...

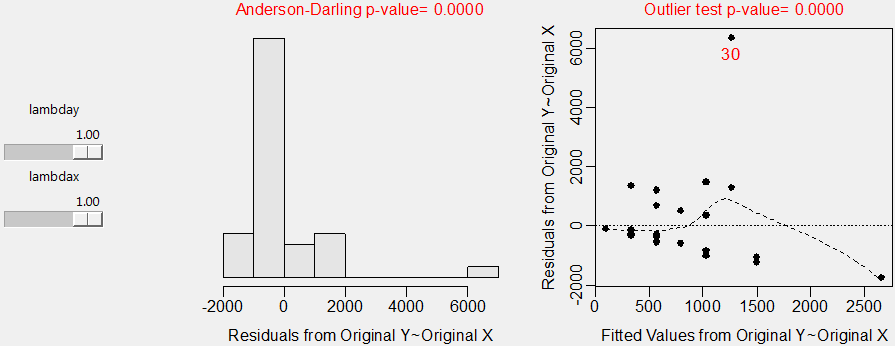
$ species : int 2 2 1 1 2 2 2 2 2 2 ...

$ logbiomass: num 0 0 0 1.099 0.693 ...

$ logspecies: num 0.693 0.693 0 0 0.693 ...

**> SLR1 <- lm(biomass~species,data=lb)**

**> transChooser(SLR1)**



**> summary(SLR1)**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -132.1 444.4 -0.297 0.7682

species 232.7 107.4 2.167 0.0378

Residual standard error: 1352 on 32 degrees of freedom

Multiple R-squared: 0.1279, Adjusted R-squared: 0.1007

F-statistic: 4.694 on 1 and 32 DF, p-value: 0.03782

**> confint(SLR1)**

2.5 % 97.5 %

(Intercept) -1037.27108 773.0795

species 13.92893 451.4920

**> predict(SLR1,data.frame(species=5),interval="prediction")**

fit lwr upr

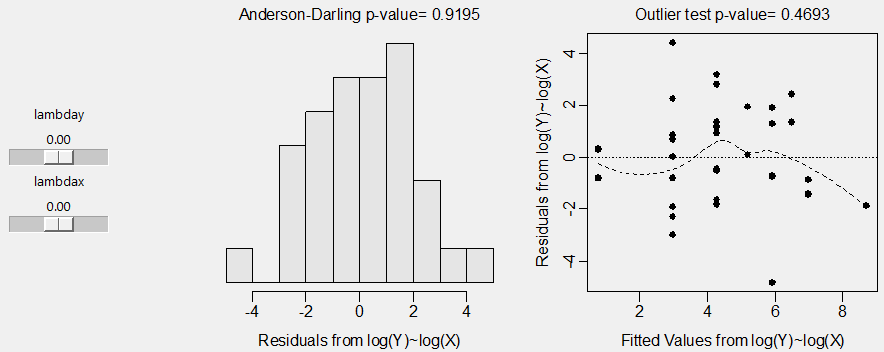
1 1031.457 -1781.347 3844.26

**> predict(SLR1,data.frame(species=5),interval="confidence")**

fit lwr upr

1 1031.457 459.9579 1602.955

**> transChooser(SLR1)**



**> SLR2 <- lm(logbiomass~logspecies,data=lb)**

**> summary(SLR2)**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.7878 0.8165 0.965 0.342

logspecies 3.1907 0.6620 4.820 3.35e-

Residual standard error: 2.066 on 32 degrees of freedom

Multiple R-squared: 0.4206, Adjusted R-squared: 0.4025

F-statistic: 23.23 on 1 and 32 DF, p-value: 3.351e-05

**> confint(SLR2)**

2.5 % 97.5 %

(Intercept) -0.8753913 2.450988

logspecies 1.8422974 4.539096

**> predict(SLR2,data.frame(logspecies=log(5)),interval="prediction")**

fit lwr upr

1 5.923027 1.599815 10.24624

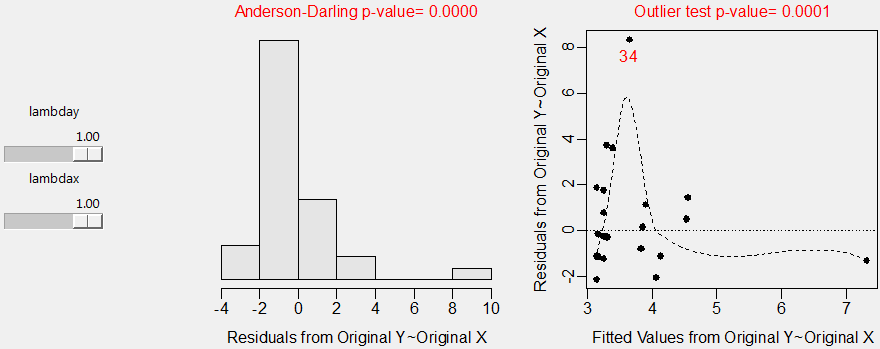
**> predict(SLR2,data.frame(logspecies=log(5)),interval="confidence")**

fit lwr upr

1 5.923027 4.936909 6.909145

**> SLR3 <- lm(species~biomass,data=lb)**

**> transChooser(SLR3)**



**> summary(SLR3)**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.1505175 0.3969917 7.936 4.67e-09

biomass 0.0005497 0.0002537 2.167 0.0378

Residual standard error: 2.078 on 32 degrees of freedom

Multiple R-squared: 0.1279, Adjusted R-squared: 0.1007

F-statistic: 4.694 on 1 and 32 DF, p-value: 0.03782

**> confint(SLR3)**

2.5 % 97.5 %

(Intercept) 2.341872e+00 3.959163120

biomass 3.290428e-05 0.001066558

**> predict(SLR3,data.frame(biomass=1000),interval="prediction")**

fit lwr upr

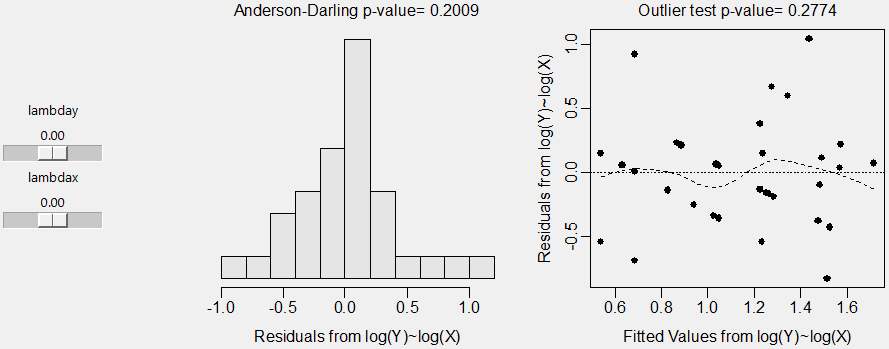
1 3.700249 -0.5975918 7.998089

**> predict(SLR3,data.frame(biomass=1000),interval="confidence")**

fit lwr upr

1 3.700249 2.956734 4.443764

**> transChooser(SLR3)**



**> SLR4 <- lm(logspecies~logbiomass,data=lb)**

**> summary(SLR4)**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.53995 0.13869 3.893 0.000472

logbiomass 0.13183 0.02735 4.820 3.35e-05

Residual standard error: 0.42 on 32 degrees of freedom

Multiple R-squared: 0.4206, Adjusted R-squared: 0.4025

F-statistic: 23.23 on 1 and 32 DF, p-value: 3.351e-05

**> confint(SLR4)**

2.5 % 97.5 %

(Intercept) 0.25744313 0.8224658

logbiomass 0.07611759 0.1875403

**> predict(SLR4,data.frame(logbiomass=log(1000)),interval="prediction")**

fit lwr upr

1 1.450597 0.5707452 2.330448

**> predict(SLR4,data.frame(logbiomass=log(1000)),interval="confidence")**

fit lwr upr

1 1.450597 1.245412 1.655782