**Winter, 2015 Biometry Quiz #1**

1. Dalling *et al.* (1998) examined the dispersal patterns and seed bank dynamics of the pioneer tree *Miconia argentae* on Barro Colorado Island, Panama. In one part of their study they recorded the density of seeds (number per m2) in the top 3 cm of the soil at 0, 5, 10, and 20 m away from the crown of four *Miconia argentae* trees. In other words, for each tree, the density of seeds at 0, 5, 10, and 20 m away from the tree was recorded. The author’s goal was to describe how the mean density of seeds changed (if at all) as one moved away from the tree. The data from this study were entered into R and analyzed with the commands and 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).
2. **[8 pts]** Fully assess ALL assumptions appropriate to this analysis, **on the original scale only**.

*Answers to b-e should refer to either the original or transformed results. Questions b-d will refer to “density” but this may be interpreted as “transformed density” if you choose to use the transformed scale. Either way, you should be very precise with your language. Your answer to e should be made on the original scale (i.e., density).*

1. **[3 pts]** What specific conclusion about seed “density” and distance from the tree can be made from the results in the ANOVA table?
2. **[3 pts]** What is the difference in **sample** mean seed “density” between the 0 and 5 m distances? Be sure to explicitly state which distance has a higher “density.”
3. **[6 pts]** On the schematic below, manually construct a means plot (i.e., a fitPlot() but without confidence intervals) from the provided results. Include letters by each mean that show which treatments are statistically different. Make sure to fully label the axes.

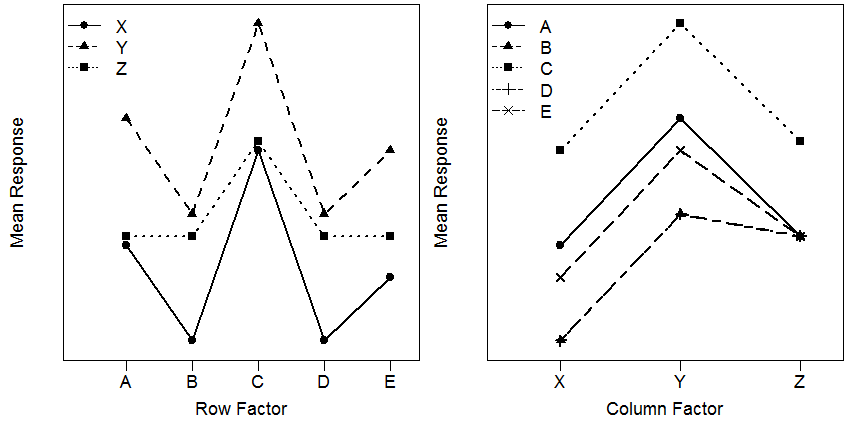


1. **[3 pts]** Specifically interpret the confidence interval for the distances with the most different mean seed density.
2. **[2 pts]** Briefly summarize the findings of this portion of the author’s study (i.e., what is the “take-home message” from these results).
3. **[18 pts]** Vanderlan and Robinson (2008) examined the effectiveness of riparian wetlands in improving water quality in Patroon Creek, a tributary to the Hudson River, in urban Albany, NY. In one part of their study they recorded the concentrations of chloride ion concentrations in water samples collected from each of four positions in the wetlands (labeled as A-D, with A being furthest from and D being the closest to Patroon Creek) and at each of two sections of the watershed (upstream or downstream). The author’s goal was to determine if the mean chloride ion concentration was affected by position in the wetland (POS), section of the watershed (SEC), or the interaction between both factors.
   1. The two-way ANOVA table for these data is shown below. Fill in the missing results. Note that there is no overall among source row.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **df** | **SS** | **MS** | **F** | **p** |
| POS | 3 | 482.3 | 160.8 | 0.145 | 0.932 |
| SEC | 1 | 6658.6 | 6658.6 | 5.992 | 0.024 |
| POS:SEC | 3 | 1552.3 | 517.4 | 0.466 | 0.709 |
| Residuals | 19 | 21113.1 | 1111.2 |  |  |
| Total | 26 | 29806.3 | 1146.4 |  |  |

* 1. What effects are AND are not evident from these results.
  2. What is the variance among individuals ignoring group membership?
  3. What is the variance among individuals within each group?
  4. How many total water samples were used in this analysis?

1. **[6 pts]** The pair of plots below are interaction plots for the same data. Assuming that these are ideal data with no sampling variability, determine which effects are evident. Provide a short explanation for each of your answers. Make sure to address each possible effect.



1. **[16 pts]** Answer FOUR of the six questions below (a-f).
2. Completely compare and contrast the meanings of MSWithin, MSTotal, and MSAmong. *Your statements should be general but you may refer to a specific instance as an example.*
3. Completely compare and contrast the concepts of a “full” and a “simple” model. *Your statements should be general but you may refer to a specific instance as an example.*
4. Define experiment-wise and comparison-wise error rates. Identify the specific and relative sizes of each error rate.
5. Describe when a Tukey’s HSD and when a Dunnet’s procedure would be appropriate to use. For the Dunnet’s situation describe why the Dunnet’s method would be “better” than Tukey’s method in the same situation.
6. 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. If you decided to demonstrate your points with an illustrative example, assume that there are 30 individuals available for experimentation and that one factor has two levels and the other factor has three levels.
7. Mathematically show that the difference in two means of a log-transformed variable becomes a RATIO of two means on the original scale*.* *You do not need complete sentences for this question but you should mathematically show each step in the proof.*

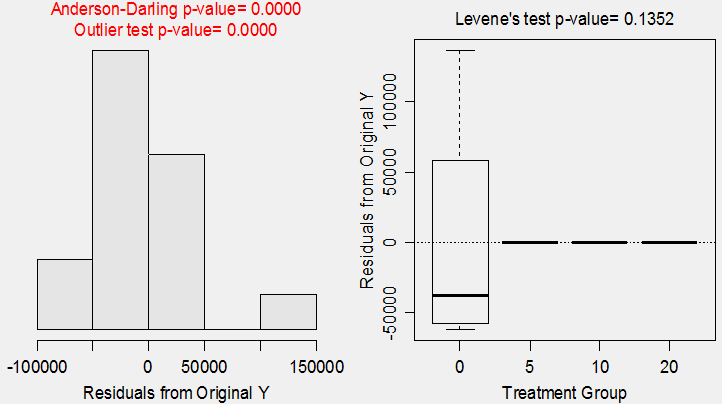
***R RESULTS FOR FIRST QUESTION***

**> d <- read.csv("SoilSeed.csv")**

**> d$dist <- factor(d$dist)**

**> lm1 <- lm(dens~dist,data=d)**

**> transChooser(lm1)**



**> anova(lm1)**

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

dist 3 2.9680e+10 9893308544 4.571 0.02344

Residuals 12 2.5972e+10 2164352866

Total 15 5.5652e+10

**> summary(lm1)**

Coefficients:

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

(Intercept) 99708 23261 4.286 0.00106

dist5 -99017 32897 -3.010 0.01086

dist10 -99678 32897 -3.030 0.01047

dist20 -99695 32897 -3.031 0.01046

---

Residual standard error: 46520 on 12 degrees of freedom

Multiple R-squared: 0.5333, Adjusted R-squared: 0.4166

F-statistic: 4.571 on 3 and 12 DF, p-value: 0.02344

**> mc1 <- glht(lm1,mcp(dist="Tukey"))**

**> summary(mc1)**

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

5 - 0 == 0 -99017.00 32896.45 -3.010 0.0465

10 - 0 == 0 -99677.58 32896.45 -3.030 0.0450

20 - 0 == 0 -99694.75 32896.45 -3.031 0.0451

10 - 5 == 0 -660.57 32896.45 -0.020 1.0000

20 - 5 == 0 -677.75 32896.45 -0.021 1.0000

20 - 10 == 0 -17.17 32896.45 -0.001 1.0000

**> confint(mc1)**

Estimate lwr upr

5 - 0 == 0 -99017.0034 -196634.8822 -1399.1245

10 - 0 == 0 -99677.5783 -197295.4572 -2059.6995

20 - 0 == 0 -99694.7494 -197312.6283 -2076.8705

10 - 5 == 0 -660.5750 -98278.4539 96957.3039

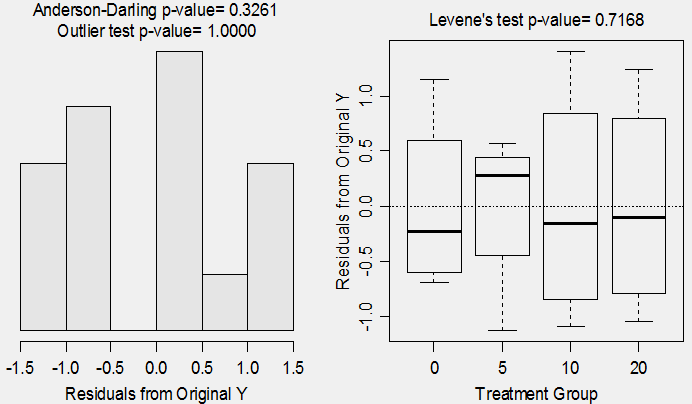
20 - 5 == 0 -677.7461 -98295.6249 96940.1328

20 - 10 == 0 -17.1711 -97635.0500 97600.7078

**> d$logdens <- log(d$dens)**

**> lm2 <- lm(logdens~dist,data=d)**

**> transChooser(lm2)**



**> anova(lm2)**

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

dist 3 202.727 67.576 78.313 3.781e-08

Residuals 12 10.355 0.863

Total 15 213.081

**> summary(lm2)**

Coefficients:

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

(Intercept) 11.2282 0.4645 24.175 1.51e-11

dist5 -4.8571 0.6568 -7.395 8.33e-06

dist10 -8.2642 0.6568 -12.582 2.85e-08

dist20 -9.0189 0.6568 -13.731 1.06e-08

---

Residual standard error: 0.9289 on 12 degrees of freedom

Multiple R-squared: 0.9514, Adjusted R-squared: 0.9393

F-statistic: 78.31 on 3 and 12 DF, p-value: 3.781e-08

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

**> summary(mc2)**

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

5 - 0 == 0 -4.8571 0.6568 -7.395 < 0.001

10 - 0 == 0 -8.2642 0.6568 -12.582 < 0.001

20 - 0 == 0 -9.0189 0.6568 -13.731 < 0.001

10 - 5 == 0 -3.4070 0.6568 -5.187 0.00105

20 - 5 == 0 -4.1618 0.6568 -6.336 < 0.001

20 - 10 == 0 -0.7547 0.6568 -1.149 0.66792

**> confint(mc2)**

Estimate lwr upr

5 - 0 == 0 -4.8571 -6.8058 -2.9084

10 - 0 == 0 -8.2642 -10.2129 -6.3154

20 - 0 == 0 -9.0189 -10.9676 -7.0702

10 - 5 == 0 -3.4070 -5.3558 -1.4583

20 - 5 == 0 -4.1618 -6.1105 -2.2131

20 - 10 == 0 -0.7547 -2.7035 1.1940