Biometry Homework #4

ANCOVA

Auriel Fournier

April 28, 2014

**Problem Overview**

*Background*

I want to look at the impact of amount of food and age on duckling weight.

*Statement of Hypothesis*

Null – there is no difference between age and duckling weight, when corrected for meal amount.

*Variables*

The response variable is meal weight of ducklings in grams. The predictor variables are age and meal amount. Age is a continuous variable, meal amount is a categorical variable.

*Statement of Assumptions*

1 – Independence

I assume that the study was designed in such a way that all samples are independent.

2 – Normality

I assume that the response variable fits a normal distribution and will examine this with histograms. If it does not a transformation will be used to improve the fit.

3 – Homogeneous Variance

I assume that the variance between variables and groups within those variables is consistent. If not I will transform the data to help better fit this assumption.

4 – Linear relationship between variables

I assume that there is a linear relationship between age and meal amount. This will be visually examined.

5 – Within group regression slopes relating Y and X are equal

I assume that the slope of the lines for each group (meal amount) will be the same, this will be tested for by including an interaction term in a second ANCOVA.

**Methods**

Standard protocols were used to collect these data.

|  |  |
| --- | --- |
| **Figure 1 – Examination of variances for the four levels of meal amount (colored) and age (black) against duckling density.**    **Figure 2 – Visual assessment of the linearity of the relationship between age (X) and weight (Y)** | **Results**  *Verifications of Assumptions*  I examined the assumptions of normality and homogeneity of variance using histograms and box plots. The variances were over a similar spread (Figure 1). The histogram showed a skewed distribution, so a variety of transformations were done on the data and a square root transformation was found to have the best correction. Because of small sample size this does not meet a true normal distribution. I am assuming that data was collected independently as part of the study design.  I graphed the data points and drew connecting lines for the age data by meal amount to see if |

|  |  |
| --- | --- |
| **Table 1 – One-Way ANOVA on age vs meal amount**    **Table 2 – Linear Regression on Age and Duck Weight**    **Table 3 – One Way ANOVA on Meal Amount and Duck Weight**    **Table 4 - ANCOVA on Duck Weight by Age & Meal Amount**      **Figure 3 – Residuals and Normal Q-Q plot from ANCOVA** | there was a linear relationship (Figure 2). There appears to be a linear relationship between age and duckling weight and each category of meal amount has a similar slope and a linear relationship.  The assumption of group regression slopes being equal is examined by running a one way ANOVA on meal amount and duck weight. The results are summarized in Table 1 and show that there is not an interaction between them. This means that there are similar slopes between the different categories.  *Analysis/Results*  I first ran a one-way ANOVA and a linear regression to see how much of the variation was explained by each variable on its own. The first model looked at age and was highly significant, the second model looked at meal amount and was not (Tables 2 and 3). This is correct, I do not want our categorical variable to be significant, only our continuous variable. Then I ran the ANCOVA which found all levels of meal amount and age to be significant predictors of duckling weight (Table 4). It has an R2 value of .96, which is higher than the one- way ANOVA. It shows a positive relationship between age and duckling weight and each level of meal and duckling weight.  The residuals from the ANCOVA show that the residuals are well distributed, while the Normal Q-Q plot shows that the data was not normally distributed, since the ends of the line do not follow the line.  **Discussion**  I found the model that included both age and meal amount explained more of the variation in duckling density then age or meal amount alone, allowing us to reject our null hypothesis that there was no interaction between these two variables and duckling weight. These results are logical since individuals who are fed more will be larger as they age. |

**Citations**

None

**Appendix 1 – R Markdown**

I did my analysis in R. Included with this document was an html file generated by R Markdown which shows all of my code and the outputs from running it.

Also provided was the R Markdown file and .csv file which will allow you to replicate my analysis.