Professor Notes on 1-Way ANOVA HW 2

- When asked to assess differences among "all" group means then you must use the overall p-value found in the ANOVA table. The p-values found in the coefficients table do not answer this question because those p-values do not compare all pairs of groups.
- Don't say "the groups differ" or the "the mean differs among groups." Be explicit ... say "the mean length of cuckoo eggs differs among the three host species nests."
- Don't say "first group" or "second group" when those groups have specific names. For example, say "the intercept represents the mean length of cuckoo eggs in the tree pipet group."
- You should be moving away from using the words "intercept" and "slope." You should now be replacing these with words like "the sample mean of the first group" and "the difference in sample means between the second group and the first group."

Raising Young Cuckoos

1. The overall sample size is n=44 and the number of groups is I=3.
2. The coefficients from fitting the one-way ANOVA to the cuckoo data are in Table 1. Interpretations of the coefficients are below the table.

Table 1. Coefficients from the one-way ANOVA fit of egg length by bird species.

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     23.09000    0.26189  88.165   <2e-16
speciesSparrow   0.03143    0.37693   0.083    0.934
speciesWagtail -0.18667    0.37037  -0.504    0.617
---
Residual standard error: 1.014 on 41 degrees of freedom
Multiple R-squared: 0.009624,Adjusted R-squared: -0.03869
F-statistic: 0.1992 on 2 and 41 DF,  p-value: 0.8202
```

   - The "Intercept" coefficient is the sample mean cuckoo egg length in the pipet (reference group) nests. Thus, the mean length of cuckoo eggs in the pipet nests is 23.09 mm.
   - The "speciesSparrow" estimate is the difference in the sample mean cuckoo egg length in the sparrow and pipet nests. In this case, the mean cuckoo egg length in the sparrow nests is 0.03 mm longer than the mean cuckoo egg length in the pipet nests.
   - The "speciesWagtail" estimate is the difference in the sample mean cuckoo egg length in the wagtail and pipet nests. In this case, the mean cuckoo egg length in the wagtial nests is 0.19 mm *shorter* than the mean cuckoo egg length in the pipet nests.
3. The interpretations of the p-values in Table 1 are below.
   - The "intercept" p-value is used to determine if the mean cuckoo egg length in the pipet (reference group) nests differs from zero. In this case, $p < 0.00005$ suggests that the mean length of cuckoo eggs in pipet nests is different from zero.
   - The "speciesSparrow" p-value is used to determine if the difference in the mean cuckoo egg length in the sparrow and pipet nests is different from zero. In this case, $p = 0.9340$ suggests that this difference is not different than zero, suggesting that there is no difference in the mean cuckoo egg length between the sparrow and pipet nests.
   - The "speciesWagtail" p-value is used to determine if the difference in the mean cuckoo egg length in the wagtail and pipet nests is different from zero. In this case, $p = 0.6170$ suggests that this difference in not different than zero suggesting that there is no difference in the mean cuckoo egg length between the wagtail and pipet nests.

4. The ANOVA table for the one-way ANOVA is in Table 2. Demonstrating how each degrees-of-freedom is calculated is below the table.

Table 2. Analysis of variance table for the egg length by bird species.

```
          Df Sum Sq Mean Sq F value Pr(>F)
species    2  0.410 0.20496  0.1992 0.8202
Residuals 41 42.182 1.02883
```

- The $df_{species}$, or df_{Among}, is one less than the number of groups (host bird species; I =3) in the analysis.
- The $df_{Residuals}$, or df_{Within}, is the total number of cuckoo eggs (n =44) minus the total number of groups (I =3) in the analysis.
- The df_{Total} is not shown in Table 2 but it is the total number of cuckoo eggs (n =44) minus 1.

5. Interpretations of the mean-squares in Table 2 are below.

- The $MS_{species}$, or MS_{Among}, is the variability that is explained by using separate means (full model) that was not explained when only one mean (simple model) was used to represent all three groups. It is also the variability of the group means.
- The $MS_{Residuals}$, or MS_{Within}, is the variability in cuckoo egg length around the mean cuckoo egg length of each host bird species group or the variability unexplained by using a separate mean for each group (i.e., the full model).
- The MS_{Total} is not shown in Table 2, however it represents the variability in cuckoo egg length around the grand mean or the variability unexplained by using only one mean (simple model) to represent all three groups.

6. The F test statistic is the ratio of the variability explained by the full model that was not explained by the simple model to the variability of the simple model. Alternatively, the F test statistic is the ratio of the variability of group means to the variability of individuals within the groups.

7. No, the overall p-value ($p = 0.8202$) (Table 2) is greater than 0.05. There is no statistical difference in mean cuckoo length among the three species of host birds (Figure 1).

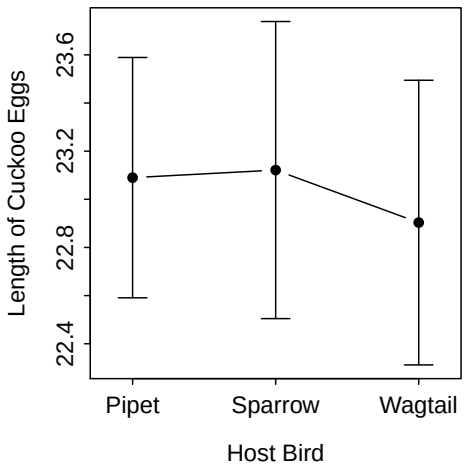

Figure 1. Plot of mean (with 95% CI) length of Cuckoo eggs by host bird type. No differences among means were detected.

R Appendix

```
library(NCStats)
setwd("c:/biometry/")
cuc <- read.csv("Cuckoos.csv")
lm1 <- lm(length~species,data=cuc)
anova(lm1)
summary(lm1)
fitPlot(lm1,xlab="Host Bird",ylab="Length of Cuckoo Eggs")
```