Home of sparrow

STA 106-project1

Erin K. Melcon

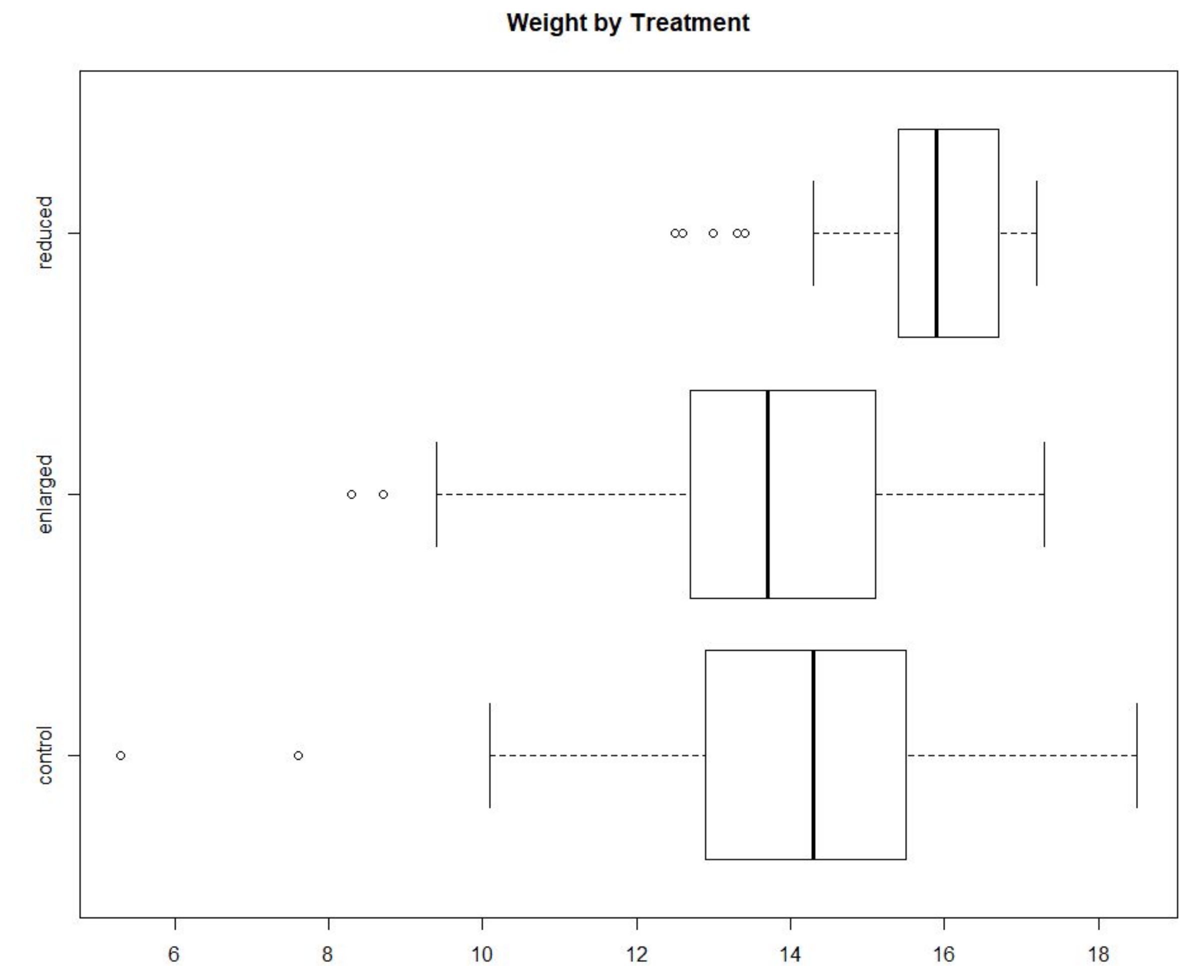
Yutian Yang
Mengwei Hu

STA106 Project1
Erin K. Melcon
Yutian Yang
Mengwei Hu
04/28/19

Section I: Introduction

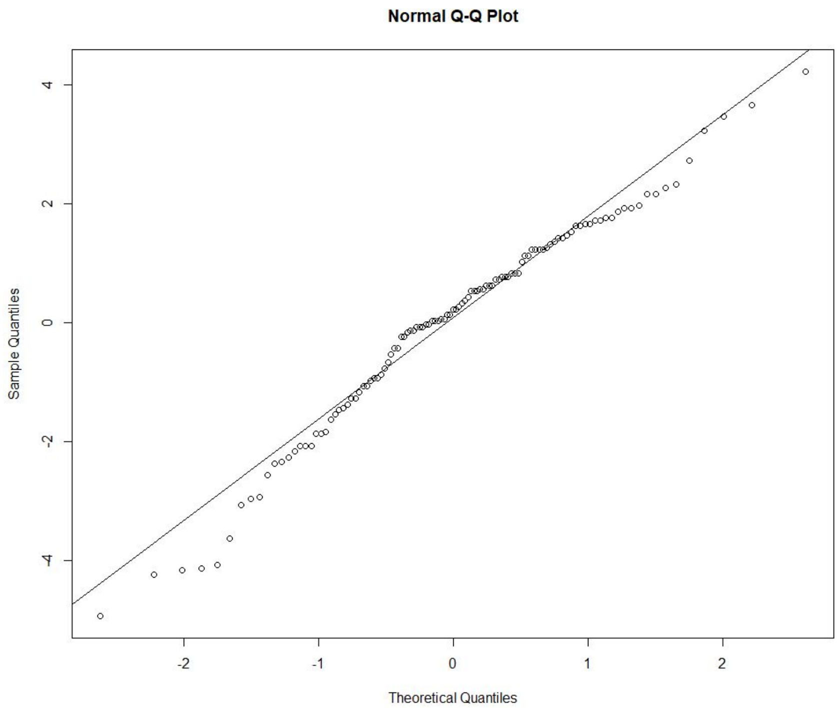
Sparrow is widely recognized in urban and rural areas today. House sparrow is easily seen by human and has often used to represent the common. They usually nest by buildings under the roof, which makes other birds less likely to take over their nests due to the indirect influence of human. Sparrow has a controversial effect on agriculture. Around the 1950s in China, they were regarded as a significant pest because they eat and destroy the grain. However, after attempts of controlling the scale of sparrows, grains then faced even greater destruction by the insect pest. From this, we can see sparrows are related to human tightly. To know more about sparrows, we are interested in how the sparrows' nests are related to their sizes. We will construct a hypothesis test to find if there are significant differences between the weight of sparrows by separating them into groups based on the size of their nests. In addition to that, we will construct three confidence interval to find out the specific results of the relationship between the sparrow weights and the size of the nests.

Section II: Summary

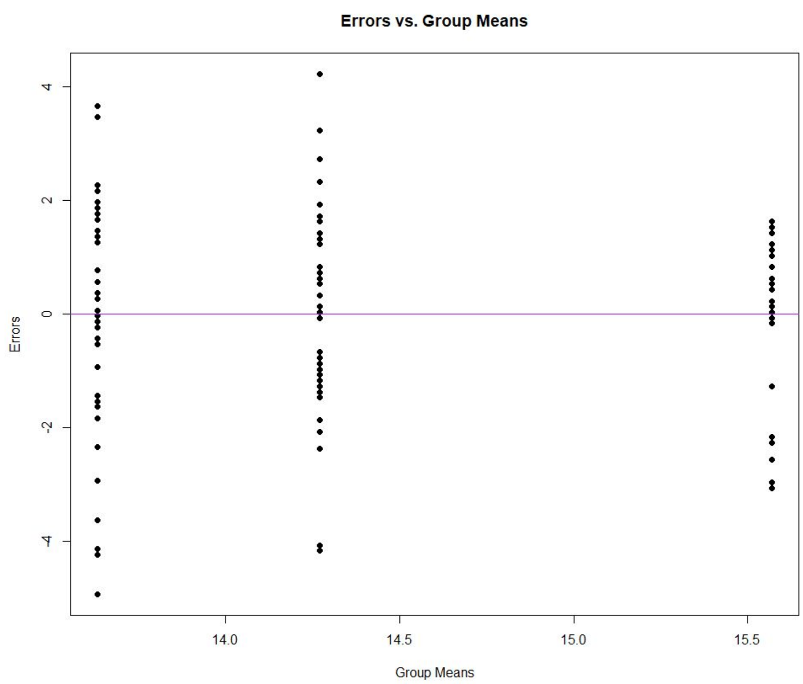


	Enlarged	Control	Reduced	Overall
Mean	13.51556	13.92444	15.56923	14.13448
Standard derivation	2.103996	2.419631	1.459252	2.242596
Size	45	45	26	116

Overall we collected 116 samples, 45 sparrows are in the control group (Normal size nest), 45 sparrows are in the enlarged group (nest larger than normal), and 26 sparrows are in the reduced group (nest smaller than normal). We find out the overall sample mean of the weights of the sparrows is 14.13448 grams. The mean of the control group is 13.92444 grams, the mean of the enlarged group is 13.51556 grams, and the mean of the reduced group is 15.56923 grams. The overall standard deviation is 2.242596 grams. The standard derivation of the control group is 2.419631 grams, the standard derivation of the enlarged group is 2.103996 grams, the standard derivation of the control group is 1.459252 grams.



Section III: Diagnostic s



Our null hypothesis is that sparrows come from different sizes of the nest have same weight mean. The alternative hypothesis is that at least one group weight mean is different from others. Comparing the mean of the groups, we find out enlarged group mean of weight is the smallest, and the reduced group mean of weight is the largest (13.51556<13.92444<15.56923). We can see a trend that the heavier sparrows tend to choose smaller nests, and the lighter sparrows tend to choose smaller nests. By calculating the approximate confidence interval of the group that has the smallest sample mean(13.51556-2*2.103996<enlarged true mean<13.51556-2*2.103996), we found out the reduced group mean(15.56923) is in this interval(9.307568<enlarged true mean<17.723552). This means these three group means do not have significant differences. Our assumption is sparrow weight is not related to the size of the nests, and we fail to reject the null hypothesis. After removing outliers, we processed the Shapiro-Wilks test and get p-value = 0.0371, which is larger than alpha = 0.01. We fail to reject the null hypothesis, which means our data is normal. We then run the Brown-Forsythe test and get p-value equals to 0.1785612, which is larger than any significant level, therefore, our data has constant variance.

Section IV: Analysis

We then run the data into the ANOVA model and get F statistics equal to 9.4231, and p-value equals to 0.000167. We then find the power is 0.855733. We calculate each multiplier and get Tukey multiplier equals to 2.973358, Scheffe multiplier equals to 3.097758, and Bonferroni multiplier equals to 2.998884. We decide to pick Tukey multiplier to calculate the simultaneous confidence interval because it has the smallest value.

99% Tukey confidence intervals				
	Lower Bound	Upper Bound	Decision to Hypothesis test	p-value
μ1-μ2 (control - enlarge)	-0.91706	1.73484	Fail to reject H0	0.630776
μ1-μ3 (control - reduced)	-3.19416	-0.09542	Reject H0	0.005765
μ2-μ3 (enlarged - reduced)	-3.60305	-0.5043	Reject H0	0.000412

This is the probability that we conclude that at least one group weight mean is different from others, and in reality, the null hypothesis that sparrows come from different sizes of the nest have same weight mean is false.

Section V: Interpretation

The 99% confidence interval for μ1-μ2 (control - enlarge) is (-0.91706, 1.73484). We could be 99% confident that the mean of the weight difference of the sparrow in the control nest is different from the weight of the sparrow in the reduced nest by -0.91706 to 1.73484. The value 0 is in this interval, which illustrates that there is no significant difference between the sparrow weight of control group and enlarged group. So we fail to reject the null hypothesis.

The 99% confidence interval for μ1-μ3 (control - reduced) is (-3.19416, -0.09542). We could be 99% confident that the mean of the weight of the sparrow in the control nest is smaller than the weight of the sparrow in the reduced nest by 0.09542 to 3.19416. And 0 is not in this interval. So there is a significant difference between the true sparrow weight mean of the control group and reduced group. We reject the null hypothesis.

The 99% confidence interval for μ2-μ3 (enlarged - reduced) is (-3.60305, -0.5043). We could be 99% confident that the mean of the weight of the sparrow in the reduced nest is smaller than the weight of the sparrow in the enlarged nest by 0.5043 to 3.60305. The interval also does not contain 0, and this illustrates that there is a significant difference between the true mean of enlarged and reduced group. And we reject null hypothesis.

Overall, the enlarged size of the nest does not affect the weight of sparrows. However, the reduced size of nests attracts heavy sparrows more.

VI. Conclusion

By analyzing the sparrows' weight of different-size-nests, we find out that the reduced group average weight is the largest, and there is no difference between the large nest and normal nest which creates influence on the weight of sparrows. So we conclude that the larger size of the nest will likely not have effects on the weight of the sparrow, but the heavier sparrows tend to choose small nests.

```r
sparrow <- read.csv("C:/Users/charl/Downloads/sparrow.csv")
boxplot(Weight ~ Treatment, data = sparrow, main = "Weight by Treatment", horizontal = TRUE)
aggregate(Weight ~ Treatment, data = sparrow, mean)
aggregate(Weight ~ Treatment, data = sparrow, sd)
aggregate(Weight ~ Treatment, data = sparrow, length)
mean(sparrow$Weight)
sd(sparrow$Weight)
give.me.CI = function(ybar,ni,ci,MSE,multiplier){
  if(sum(ci) != 0 & sum(ci !=0 ) != 1){
    return("Error - you did not input a valid contrast")
  } else if(length(ci) != length(ni)){
    return("Error - not enough contrasts given")
  }
  else{
    estimate = sum(ybar*ci)
    SE = sqrt(MSE*sum(ci^2/ni))
    CI = estimate + c(-1,1)*multiplier*SE
    result = c(estimate,CI)
    names(result) = c("Estimate","Lower Bound","Upper Bound")
    return(result)
  }
}
group.means =  by(sparrow$Weight,sparrow$Treatment,mean)
group.nis = by(sparrow$Weight,sparrow$Treatment,length)
the.model = lm(Weight ~ Treatment, data = sparrow)
anova.table = anova(the.model)
MSE = anova.table[2,3]
nt = sum(group.nis)
a = length(group.means)
alpha = 0.05
the.model = lm(Weight ~ Treatment,data = sparrow)
sparrow$ei = the.model$residuals
nt = nrow(sparrow) #Calculates the total sample size
a = length(unique(sparrow$Treatment)) #Calculates the value of a
SSE = sum(sparrow$ei^2) #Sums and squares the errors (finds SSE)
MSE = SSE/(nt-a) #Finds MSE
eij.star = the.model$residuals/sqrt(MSE)
alpha = 0.01
t.cutoff= qt(1-alpha, nt-a)
CO.eij = which(abs(eij.star) > t.cutoff)
CO.eij

outliers = CO.eij
new.data = sparrow[-outliers,]
new.model = lm(Weight ~ Treatment,data = new.data)
qqnorm(new.model$residuals)
qqline(new.model$residuals)
ei = new.model$residuals
the.SWtest = shapiro.test(ei)
the.SWtest
plot(new.model$fitted.values, new.model$residuals, main = "Errors vs. Group Means",xlab = "Group Means",yl
ab = "Errors",pch = 19)
abline(h = 0,col = "purple")
library(car)
```

```
the.BFtest = leveneTest(ei~ Treatment, data=sparrow, center=median)
p.val = the.BFtest[[3]][1]
p.val
Tuk = qtukey(1-alpha,a,nt-a)/sqrt(2)
Tuk
Scheffe = sqrt((a-1)*qf(1-alpha, a-1, nt-a))
Scheffe
g=3
Bonferroni = qt(1-alpha/(2*g),nt-a)
Bonferroni
library(asbio)
tukeyCI(sparrow$Weight, sparrow$Treatment, conf.level = 1-alpha) #Tukey
give.me.power = function(ybar,ni,MSE,alpha){
  a = length(ybar) # Finds a
  nt = sum(ni) #Finds the overall sample size
  overall.mean = sum(ni*ybar)/nt # Finds the overall mean
  phi = (1/sqrt(MSE))*sqrt( sum(ni*(ybar - overall.mean)^2)/a) #Finds the books value of phi
  phi.star = a *phi^2 #Finds the value of phi we will use for R
  Fc = qf(1-alpha,a-1,nt-a) #The critical value of F, use in R's function
  power = 1 - pf(Fc, a-1, nt-a, phi.star)# The power, calculated using a non-central F
  return(power)
}
the.power = give.me.power(group.means,group.nis,MSE,0.01)
the.power
```