GEOG606: Assignment #2

Anika Cartas

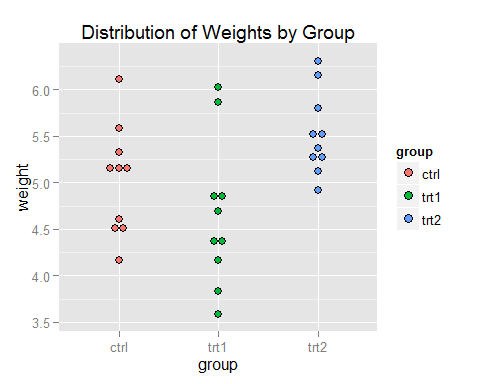
Monday, March 02, 2015

We use R’s built-in dataset, PlantGrowth (you could use help(PlantGrowth) for details about this dataset). There are three groups of plant weight data from an experiment consisting of 2 treatments (trt1 and trt2) and 1 control group (ctr1).

**Use two testing methods (not limited to the methods covered in class) to test if and how the two treatments are different from the control.**

library(ggplot2)  
p <- PlantGrowth  
  
#Preview the data set  
ggplot(p, aes(x = group, y = weight, fill=group), ) + geom\_dotplot(binaxis = "y", stackdir = "center") + ggtitle("Distribution of Weights by Group")

## stat\_bindot: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.



### 1. Shapiro-Wilk Test

Test for normality of datasets. If p <=0.05, we can state with 95% confidence the data does not fit the normal distribution. If p > 0.05, we can state that no significant departure from normality was found.

sw.ctrl = shapiro.test(p[p$group=="ctrl","weight"])  
sw.trt1 = shapiro.test(p[p$group=="trt1","weight"])  
sw.trt2 = shapiro.test(p[p$group=="trt2","weight"])  
  
cat("P values: ctrl =",sw.ctrl[["p.value"]],"trt1 =",sw.trt1[["p.value"]],"trt2 =",sw.trt2[["p.value"]])

## P values: ctrl = 0.7474734 trt1 = 0.451944 trt2 = 0.5642519

All three groups return p > 0.05, so we don't have to rule out normal distribution for the data.

### 2. Wilcoxon Signed-Rank Test

Use to decide whether the population distributions of two samples are identical without assuming a normal distribution. The null-hypothesis is that the populations of the two treatment samples are identical.

Actually, the help file seems to indicate that setting paired=TRUE and including two values means this test will be set to test #3 performed below.

wilcox.test(p[p$group=="trt1","weight"],p[p$group=="trt2","weight"],paired=TRUE,conf.level=0.95)

##   
## Wilcoxon signed rank test  
##   
## data: p[p$group == "trt1", "weight"] and p[p$group == "trt2", "weight"]  
## V = 5, p-value = 0.01953  
## alternative hypothesis: true location shift is not equal to 0

Our p value is < 0.05, so we can with 95% confidence say that the populations of the treatments are *not* identical.

### 3. Mann-Whitney-Wilcoxon Test

Use to decide if the population distributions of our two samples and the control group are identical without assuming a normal distribution.

wc.ct1 <- wilcox.test(weight ~ group, data=p[p$group!="trt2",]) #ctrl vs trt1

## Warning in wilcox.test.default(x = c(4.17, 5.58, 5.18, 6.11, 4.5, 4.61, :  
## cannot compute exact p-value with ties

wc.ct2 <- wilcox.test(weight ~ group, data=p[p$group!="trt1",]) #ctrl vs trt2  
wc.t1t2 <- wilcox.test(weight ~ group, data=p[p$group!="ctrl",]) #trt1 vs trt2  
  
cat("P values: Ctrl~Trt1 =",wc.ct1[["p.value"]],"Ctrl~Trt2 =",wc.ct2[["p.value"]],"Trt1~Trt2 =",wc.t1t2[["p.value"]])

## P values: Ctrl~Trt1 = 0.1985958 Ctrl~Trt2 = 0.06301284 Trt1~Trt2 = 0.008930698

I performed this test on all 3 combinations, just to check. Based on the p-values, we can only say that the two treatments come from different populations, and not make any statement about their relationship to the control group.

### 4. Kruskal-Wallis Test

Use to test the null hypothesis that the population distributions of our samples are identical. Since we already know from the previous tests that the two treatments do not come from the same population, I only need to compare each treatment with the control.

To verify my previous results, I will also compare the two treatments with each other.

k.ct1 <- kruskal.test(weight ~ group, data=p[p$group!="trt2",]) #ctrl vs trt1  
k.ct2 <- kruskal.test(weight ~ group, data=p[p$group!="trt1",]) #ctrl vs trt1  
k.t1t2 <- kruskal.test(weight ~ group, data=p[p$group!="ctrl",]) #ctrl vs trt1  
  
cat("P values: Ctrl~Trt1 =",k.ct1[["p.value"]],"Ctrl~Trt2 =",k.ct2[["p.value"]],"Trt1~Trt2 =",k.t1t2[["p.value"]])

## P values: Ctrl~Trt1 = 0.1857113 Ctrl~Trt2 = 0.05878172 Trt1~Trt2 = 0.0101652

Based on the resulting p-values, we cannot rule out that trt1 and ctrl come from the same population. For trt2 and ctrl, our p-value is just barely above 0.05 so again we could not say with 95% confidence that the two populations are not identical. However, we can verify our conclusions from the previous statements that trt1 and trt2 come from different populations, as the p value is <0.05.