

Professor Notes on 1-Way ANOVA Homework 3

- For the independence assumption, be clear that you have thought through what it means to be independent within or among groups. It is not adequate to just say “the groups are independent.” Also note that there is nothing in the data that speaks to independence. Assessing independence is purely a thought process.
- Note that it is hard to assess equal variances and normality with these data because of the outlier.
- On the first question of the second part, you must use the p-value from the ANOVA table. That is the p-value that assesses whether all group means are equal or not. The Tukey’s multiple comparisons are only used to assess difference in paired means AFTER it has been determined that there is a difference in means. The linear models coefficients table is not appropriate for answering that question (you should get out of the habit of using `summary()` on your `lm()` results).
- The second question should include a plot of means with appropriate significance letters.
- In the last questions, make sure to clearly indicate which group is greater (or lesser). Don’t just say that you are 95% confident that the difference is between such-and-such.
- Make sure to note how concise the answer key is. Work to get your answers this concise.

Cushing’s Syndrome I

1. The individuals appear to be independent among groups (i.e., an individual cannot possibly be in more than one group as that person cannot have two reasons for the syndrome). They are likely independent within groups as there is no indication of any connection between individuals within a group (for example, there is no indication that they are related biologically or were paired by the researchers). However, the within group independence is not obvious with the information given.
2. The Levene’s test suggests that the variances are equal ($p = 0.2181$). The residual plot is not of much help because individual 20 appears to be such a large outlier (Figure 1-Left).

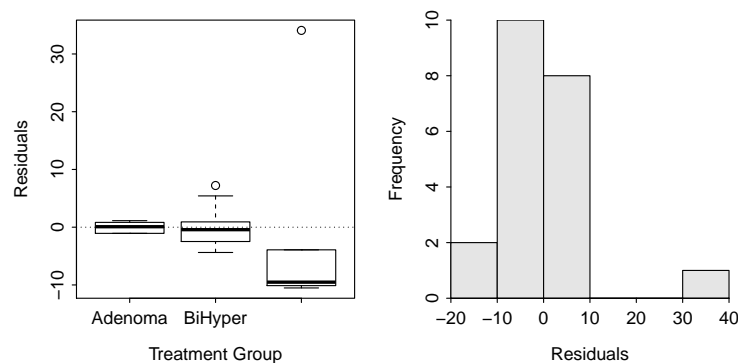


Figure 1. Residual plot (left) and histogram of residuals (right) from the one-way ANOVA of UTH levels by syndrome type.

3. The Anderson-Darling normality test strongly suggests that the residuals are not normally distributed ($p < 0.00005$). The histogram of the residuals is not of much help because of the very large outlier (Figure 1-Right).
4. Observation 20 appears to be a significant outlier ($p < 0.00005$). This was also evident on the residual plot and histogram of residuals (Figure 1).

Cushing's Syndrome II

1. The mean UTH is significantly different among the three groups ($p = 0.0017$; Table 1).

Table 1. Analysis of variance table for the UTH values by syndrome type with the 20th individual removed.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cause	2	181.12	90.561	9.4911	0.001706
Residuals	17	162.21	9.542		

2. It appears that the mean UTH for the adenoma group is significantly different from the mean UTH for both the bilateral hyperplasia ($p = 0.0118$; Table 2) and carcinoma ($p = 0.0018$; Table 2) groups and that the mean UTH for the bilateral hyperplasia and carcinoma groups are NOT statistically different ($p = 0.2497$; Table 2). These results are shown visually in Figure 2.

Table 2. Tukey's multiple comparison results for the UTH values by syndrome type with the 20th individual removed.

	Estimate	Std. Error	t value	p value
BiHyper - Adenoma = 0	5.213333	1.595136	3.268268	0.011832838
Carcinoma - Adenoma = 0	8.233333	1.993920	4.129218	0.001836852
Carcinoma - BiHyper = 0	3.020000	1.827458	1.652568	0.249674918

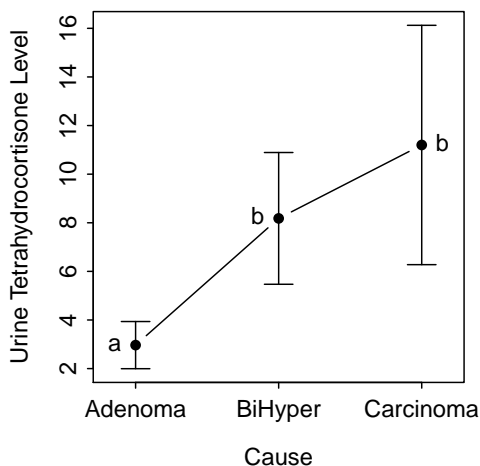


Figure 2. Plot of mean (with 95% CI) UTH level by syndrome type with the 20th individual removed. Different letters indicate means that are significantly different.

3. The mean UTH of the bilateral hyperplasia group is between 1.13 and 9.29 units greater than the mean UTH for the adenoma group (Table 3). The mean UTH for the carcinoma group is between 3.13 and 13.34 units greater than the mean UTH for the adenoma group (Table 3). The mean UTH for the bilateral hyperplasia and carcinoma groups are not statistically different.

Table 3. Tukey's confidence interval results for the difference in mean UTH values by syndrome type with the 20th individual removed.

	Estimate	lwr	upr
BiHyper - Adenoma	5.213333	1.131739	9.294927
Carcinoma - Adenoma	8.233333	3.131341	13.335326
Carcinoma - BiHyper	3.020000	-1.656053	7.696053

R Appendix

```
library(NCStats)
setwd("c:/biometry/")
d <- read.csv("Cushings.csv")

# First question
lm1 <- lm(uth~cause,data=3)
levenesTest(lm1)
residPlot(lm1)
adTest(lm1$residuals)
outlierTest(lm1)

# Second question
d2 <- d[-20,]
lm2 <- lm(uth~cause,data=d2)
anova(lm2)
mc2 <- glht(lm2,mcp(cause="Tukey"))
summary(mc2)
confint(mc2)
fitPlot(lm2,xlab="Cause",ylab="Urine Tetrahydrocortisone Level")
addSigLetters(lm2,c("a","b","b"),pos=c(2,2,4))
```