Question 2.5 (8 pts)

a. 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. However, the within group independence is not obvious with the information given.

b. 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).
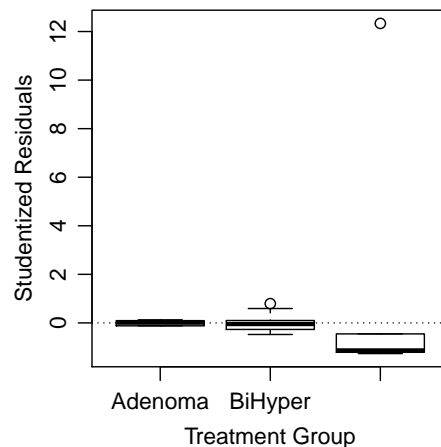

Figure 1. Residual plot from the one-way ANOVA of UTH levels by syndrome type.

c. 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 2).
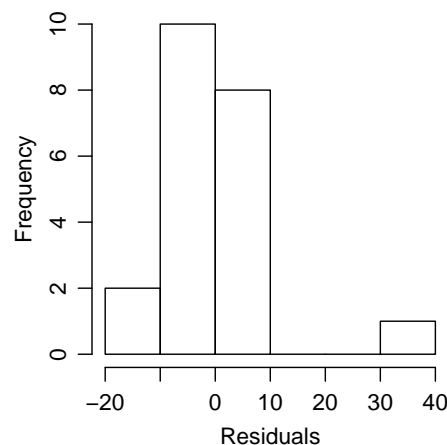

Figure 2. Histograms of residuals from the one-way ANOVA of UTH levels by syndrome type.

d. Observation 20 appears to be a significant outlier (p<0.00005). This was also evident on the residual plot (Figure 1) and histogram of residuals (Figure 2).

R commands

```
> library(NCStats)
> library(multcomp)
> # the next four lines is just an alternative way to enter the data
> #   using read.table() is easier and more efficient
> cause <- c(rep("Adenoma",6),rep("BiHyper",10),rep("Carcinoma",5))
> cause <- factor(cause,levels= c("Adenoma","BiHyper","Carcinoma"))
> uth <- c(3.1,3,1.9,3.8,4.1,1.9,8.3,3.8,3.9,7.8,9.1,15.4,7.7,6.5,5.7,13.6,
  10.2,9.2,9.6,53.8,15.8)
> d <- data.frame(uth,cause)
> lm1 <- lm(uth~cause,data=3)
> leveneTest(lm1)
> residPlot(lm1,main="")
> adTest(lm1$residuals)
> hist(lm1$residuals,xlab="Residuals",main="")
> outlierTest(lm1)
```

Question 2.6 (8 pts)

a. 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
```

b. It appears that the mean UTH for the adenoma group is significantly different from the mean UTH for both the bilateral hyperplasia (p=0.0120; Table 2) and carcinoma (p=0.0020; Table 2) groups and that the mean UTH for the bilateral hyperplasia and carcinoma groups are NOT statistically different (p=0.2499; Table 2). These results are shown visually in Figure 3.

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 0.011957544
Carcinoma - Adenoma = 0 8.233333   1.993920 4.129218 0.001960041
Carcinoma - BiHyper = 0 3.020000   1.827458 1.652568 0.249929534
```

c. The mean of the bilateral hyperplasia group is between 1.13 and 9.29 units greater than the mean for the adenoma group (Table 3). The mean for the carcinoma group is between 3.13 and 13.34 units greater than the mean for the adenoma group (Table 3). The means 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  1.131740  9.294926
Carcinoma - Adenoma 8.233333  3.131342 13.335325
Carcinoma - BiHyper 3.020000 -1.656052  7.696052
```
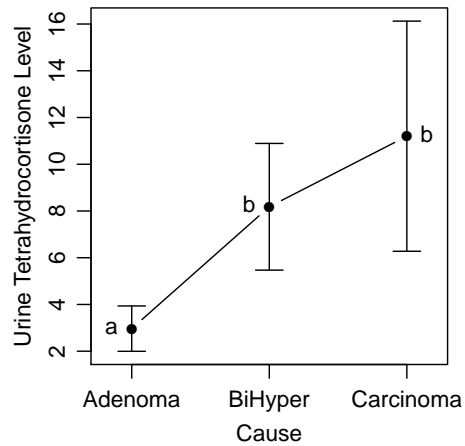
Figure 3. Plot of mean (with 95% CI) UTH level by syndrome type with the 20th individual removed. Different letters indicate means that are significantly different.

R commands

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