# Question 2.5

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-Left).
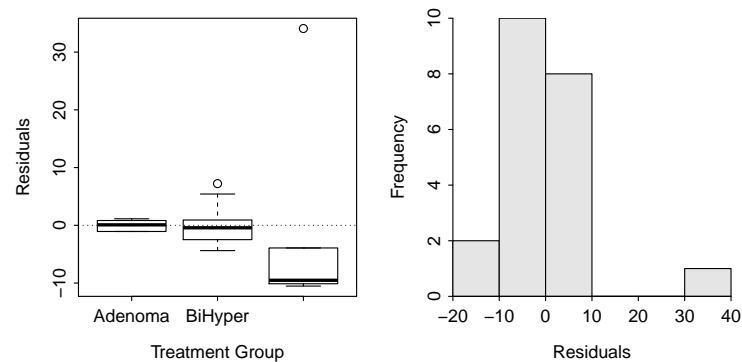


Figure 1. Residual plot (left) and histogram of residuals (right) from the one-way ANOVA of UTH levels by syndrome type.

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 1-Right).

d. 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).

R commands

```
> library(NCStats)
> # the next four lines is just an alternative way to enter the data
> #   using read.table() is easier and more efficient
> d <- data.frame(cause=factor(c(rep("Adenoma",6),rep("BiHyper",10),rep("Carcinoma",5)),
                         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) )
> lm1 <- lm(uth~cause,data=3)
> levenesTest(lm1)
> residPlot(lm1)
> adTest(lm1$residuals)
> outlierTest(lm1)
```