```
## Warning:  package 'TH.data' was built under R version 3.5.2
```

Professor Notes on 1-Way ANOVA Homework 4

- 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.
- On the third question, 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). Make sure you refer to these results as being transformed (i.e., "mean natural log iron levels").
- The fourth question should include a plot of means with appropriate significance letters. Note that when using Tukey's method that there should be letters on each point. With Dunnett's method it may be appropriate to leave one point without a letter, but this should be explained in the figure label. Make sure you refer to these results as being transformed (i.e., "mean natural log iron levels").
- 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. Also note that when you back-transform from the log scale that the difference in means becomes a ratio of means such that the first group is that multiple of the second group.
- Make sure to note how concise the answer key is. Work to get your answers this concise.

Iron and Mining

1. The individuals are likely independent but this is not abundantly clear. The measurements are from 120 unique rivers, so there is not multiple measurements on the same river. However, some of the rivers are likely in the same watershed and would share characteristics (e.g., geogological, other land use, etc.) based on that. The data are likely independent enough for our purposes. The variances appear to be equal (Levene's test $p = 0.108$), though the residual plot suggests several outliers (Figure 1-Right). The residuals are strongly not normal (Anderson-Darling $p < 0.0005$) and appear strongly right-skewed (Figure 1-Left). Finally, there is evidence for significant outliers (outlier test $p < 0.0005$). Thus, the assumptions for a one-way ANOVA have NOT been met.
2. The iron levels were transformed to the natural log scale. On this scale, the variances appear to be equal (Levene's $p = 0.472$), the residuals appear to be normal (Anderson-Darling $p < 0.0005$), and no significant outliers are present ($p < 0.0005$; Figure 2). Thus, these data will be analyzed on the natural log scale.

3. The mean natural log of iron levels differs among the three mine types ($p < 0.0005$; Table 1).

Table 1. Analysis of variance table for the natural log of iron levels by mine type.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
use	2	90.091	45.045	21.743	9.35e-09
Residuals	117	242.392	2.072		

4. It appears that the mean natural log of iron levels for the abandoned mines is significantly greater than that for the unmined ($p < 0.0005$) and reclaimed mines ($p < 0.0005$), but the mean natural log of

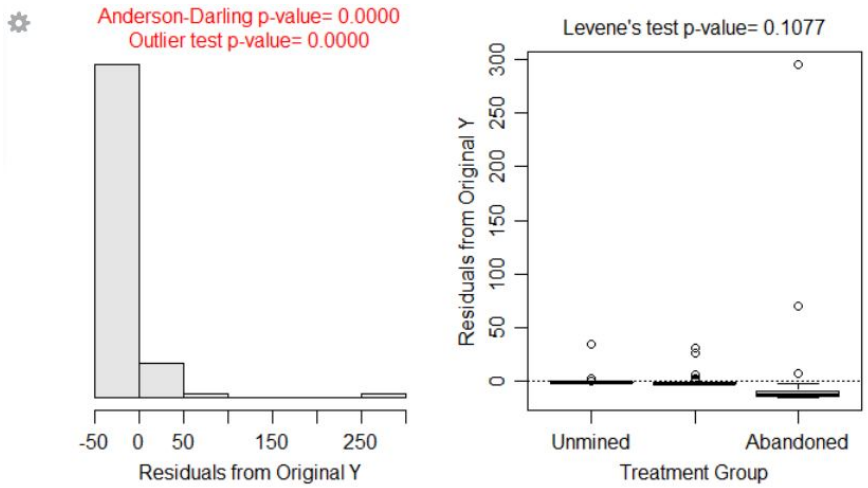

Figure 1. Histogram of residuals (left) and boxplot of residuals by treatment group (right) for the iron mining study.

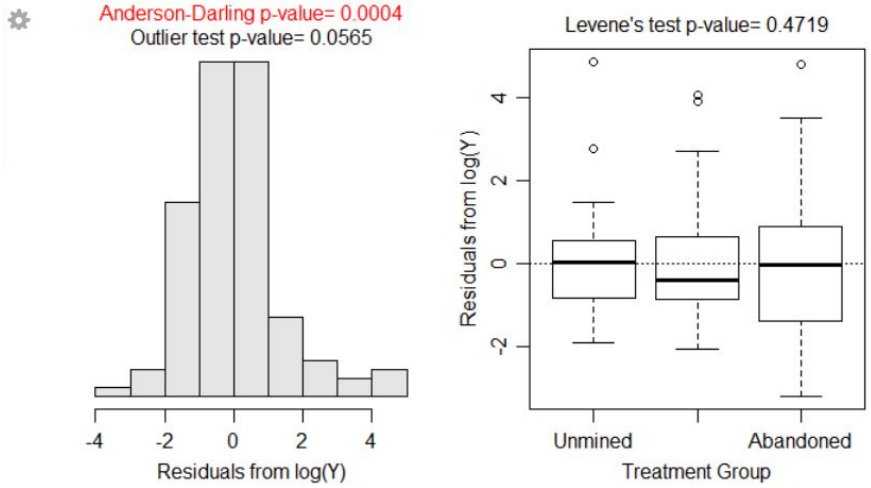

Figure 2. Histogram of residuals (left) and boxplot of residuals by treatment group (right) for the log transformed iron mining study.

iron levels does not differ between the unmined and reclaimed mines sites ($p = 0.069$; Table 2). These results are shown visually in Figure 3.

Table 2. Tukey's multiple comparison results for the natural log of iron levels by mine site type.

	Estimate	Std. Error	t value	p value
Reclaimed - Unmined = 0	0.6987774	0.3125378	2.235817	6.919437e-02
Abandoned - Unmined = 0	2.1989605	0.3397823	6.471674	5.172320e-09
Abandoned - Reclaimed = 0	1.5001831	0.3226359	4.649772	2.280144e-05

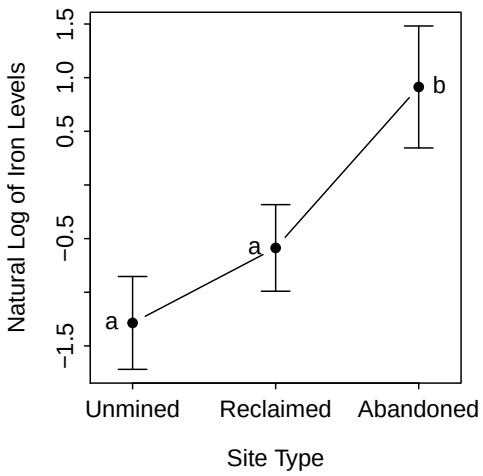

Figure 3. Plot of mean (with 95% CI) natural log of iron levels by site type. Different letters indicate means that are significantly different.

5. The mean iron level for abandoned mines sites is between 4.02 and 20.20 times greater than the mean iron levels for the unmined sites (Table 3).

   Table 3. Back-transformed Tukey's confidence interval results for the ratio of mean iron levels between pairs of site types.

	Estimate	lwr	upr
Reclaimed - Unmined	2.011292	0.9577963	4.223545
Abandoned - Unmined	9.015637	4.0244588	20.196929
Abandoned - Reclaimed	4.482510	2.0840540	9.641254

R Appendix

```
library(NCStats)
setwd("c:/biometry/")
d <- read.csv("AcidMineDrainage.csv")
d$use <- factor(d$use,levels=c("Unmined","Reclaimed","Abandoned"))

lm1 <- lm(FE~use,data=d)
transChooser(lm1)

d$logFE <- log(d$FE)
lm2 <- lm(logFE~use,data=d)
anova(lm2)
```

```r
mc2 <- glht(lm2,mcp(use="Tukey"))
summary(mc2)
fitPlot(lm2,xlab="Mine Site Type",ylab="Natural Log Iron Level")
addSigLetters(lm2,c("a","a","b"),pos=c(2,2,4))
exp(confint(mc2)$confint)
```