Homework #6

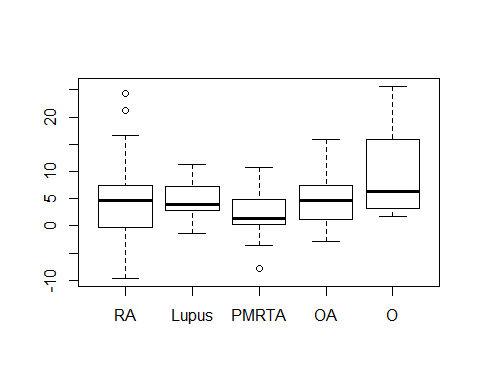
Glenn Clapp

8.2.2 Patients suffering from rheumatic diseases or osteoporosis often suffer critical losses in bone mineral density (BMD). Alendronate is one medication prescribed to build or prevent further loss of BMD. Holcomb and Rothenber looked at 96 women taking alendronate to determine if a difference existed in the mean percent change in BMD among five different primary diagnosis classifications. Group 1 patients were diagnosed with rheumatoid arthritis (RA). Group 2 patients were a mixed collection of patients with diseases including lupus, Wegener’s granulomatosis and polyarteritis, and other vasculitic diseases (LUPUS). Group 3 patients had polymyalgia rheumatica or temporal arthritis (PMRTA). Group 4 patients had osteoarthritis (OA) and group 5 patients had osteoporosis (O) with no other rheumatic diseases identified in the medical record. Changes in BMD are shown in the following table.

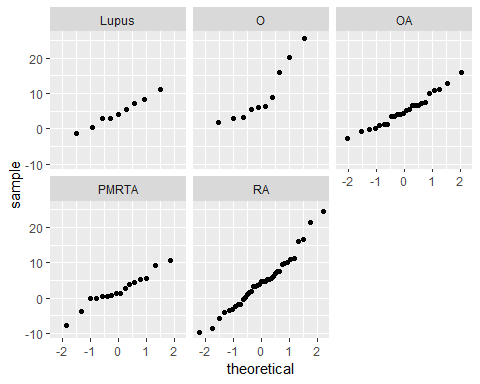
Go through the ten steps of analysis of variance hypothesis testing to see if you can conclude that there is a difference among population means. Let = 0.05 for each test. Use Tukey’s HSD procedure to test for significant differences among individual pairs of means (if appropriate). Use the same value for the F test. Construct a dot plot and sid-by-side boxplots of the data.

1. Description of data
2. Assumptions: The four sets of data constitute independent simple random samples from the four indicated populations. We assume that the four populations of measurements are normally distributed with equal variances.
3. Hypotheses: , .
4. Test statistic
5. Distribution of test statistic
6. Decision rule 7.Calculation of test statistic
7. Statistical decision 9.Conclusion
8. p-value

BMD <- read.csv('BMD\_data.csv')  
BMD\_stacked <- read.csv('BMD\_data - stacked.csv')  
boxplot(BMD)



gf\_qq(~BMD|Disorder, data=BMD\_stacked)



then examine the levene’s test for equal variances

with(BMD\_stacked, leveneTest(BMD, Disorder, center=mean))

## Levene's Test for Homogeneity of Variance (center = mean)  
## Df F value Pr(>F)   
## group 4 2.2802 0.06667 .  
## 91   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Because of this failure of the Levene test, we have sufficient evidence to reject the null hypothesis (of that test) that the variances of the samples are equal. However, as an exercise, we move forward with the ANOVA.

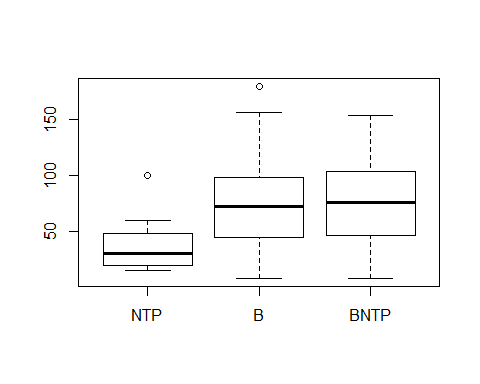
AnovaModel.1 <- lm(BMD ~ Disorder, data=BMD\_stacked)  
anova(AnovaModel.1)

## Analysis of Variance Table  
##   
## Response: BMD  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Disorder 4 355.5 88.864 2.2772 0.06697 .  
## Residuals 91 3551.1 39.024   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

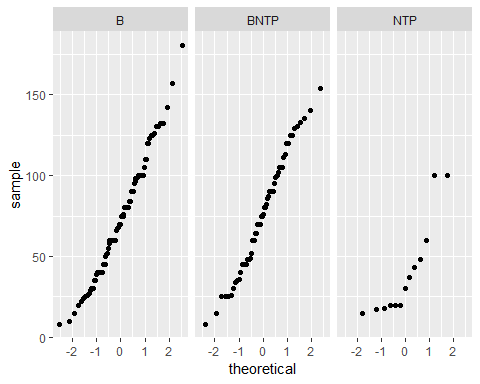
There is insufficient evidence to reject the null hypothesis. So, we conclude that the mean bone mineral density change for each sample is equal.

8.2.4 Gold et al. investigated the effectiveness on smoking cessation of a nicotine patch, bupropion SR, or both, when co-administerd with cognitive-behavioral therapy. Consecutive consenting patients (n=164) assigned themselves to one of three treatments according to personal preference: nicotine patch (NTP, n=13), bupropion SR (B; n = 92), and bupropion SR plus nicotine patch (BNTP,n = 59). At their first smoking cessation class, patients estimated the number of packs of cigarettees they currently smoked per day and the numbers of years they smoked. The “pack years” is the average number of packs the subject smoked per day multiplied by the number of years the subject had smoked. The results are shown in the following table.

PY <- read.csv('Pack\_Years.csv')  
PY\_stacked <- read.csv('Pack\_Years - Stacked.csv')  
boxplot(PY)



gf\_qq(~Pack\_Years|Cessation\_technique, data=PY\_stacked)



then examine the levene’s test for equal variances

with(PY\_stacked, leveneTest(Pack\_Years, Cessation\_technique, center=mean))

## Levene's Test for Homogeneity of Variance (center = mean)  
## Df F value Pr(>F)  
## group 2 0.6902 0.503  
## 161

There is not sufficient evidence to reject the null hypothesis, so the variances of the samples is considered equal.

AnovaModel.2 <- lm(Pack\_Years ~ Cessation\_technique, data=PY\_stacked)  
anova(AnovaModel.2)

## Analysis of Variance Table  
##   
## Response: Pack\_Years  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Cessation\_technique 2 14490 7244.8 5.8779 0.003437 \*\*  
## Residuals 161 198442 1232.6   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

There is sufficient evidence to reject the null hypothesis in this case. We conclude that at least one of the sample means is not equal to the others.

Now do a post-hoc analysis.

.Pairs <- glht(AnovaModel.2, linfct = mcp(Cessation\_technique = "Tukey"))  
cat('pairwise tests:\n')

## pairwise tests:

summary(.Pairs)

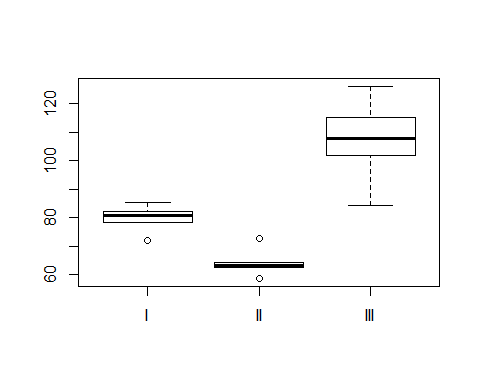
##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: lm(formula = Pack\_Years ~ Cessation\_technique, data = PY\_stacked)  
##   
## Linear Hypotheses:  
## Estimate Std. Error t value Pr(>|t|)   
## BNTP - B == 0 2.862 5.856 0.489 0.87266   
## NTP - B == 0 -33.319 10.402 -3.203 0.00455 \*\*  
## NTP - BNTP == 0 -36.181 10.757 -3.364 0.00277 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

Based on these post-hoc tests, we conclude that they nicotine patch participants accumulated a different average number of pack years than either of the other cessation techniques.

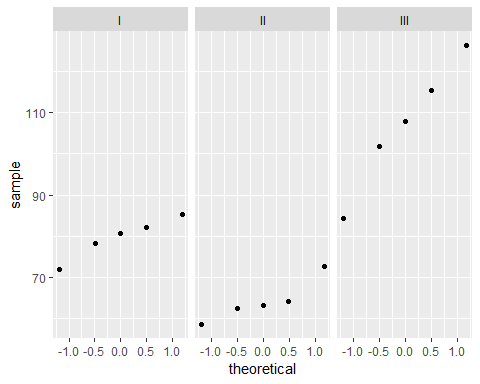
For the following exercises, perform the Kruskal-Wallis test at the indicated level of significance and determine the p value:

13.8.2 The following are outpatient charges (-$100) made to patients for a certain surgical procedure by samples of hospitals located in three different areas of the country: 1. Data 2. Assumptions: The samples are independent random samples from their respective populations. The measureemnt scale employed is at least ordinal. The distributions of the values in the sampled populations are identical except for the possibility that one or more of the populations are composed of values that tend to be larger than those of the other populations. 3. Hypotheses: , .Let = 0.01 4. Test statistic: H = (12/n(n+1))\*sum((R^2/n)-3(n+1)) 5. Distribution of test statistic. Critical values of H for various sample sizes and levelsare given in Appendix Table N. 6. Decision rule: The null hypothesis will be rejected if the computed value of H is so large that the probability of obtaining a value that large or larger when H\_0 is true is equal to or less than the chosen significance level, . 7. Calculation of test statistic: When the three samples are combined into a single series and ranked, the table of ranks shown in Table 13.8.2 may be constructed. The null hypothesis implies that the observations in the three samples constitute a single sample size 15 from a single population. If this is true, we could expect the ranks to be well distributed among the three groups. Consequently, we would expect the total sum of ranks to be divided among the three groups in proportion to group size. 8. Statistical decision 9. Conclusion 10. p-value

OUT <- read.csv('Outpatient\_charges.csv')  
OUT\_stacked <- read.csv('Outpatient\_charges - stacked.csv')  
boxplot(OUT)



gf\_qq(~Charge|Area, data=OUT\_stacked)



favstats(Charge~Area, data=OUT\_stacked)

## Area min Q1 median Q3 max mean sd n missing  
## 1 I 71.94 78.15 80.75 82.05 85.40 79.658 5.042972 5 0  
## 2 II 58.63 62.50 63.24 64.20 72.70 64.254 5.173778 5 0  
## 3 III 84.21 101.76 107.74 115.30 126.15 107.032 15.676389 5 0

kruskal.test(Charge ~ Area, data=OUT\_stacked)

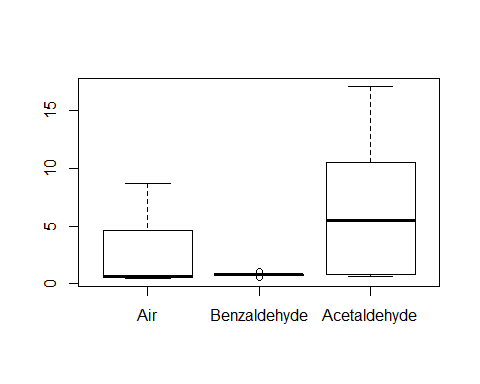
##   
## Kruskal-Wallis rank sum test  
##   
## data: Charge by Area  
## Kruskal-Wallis chi-squared = 11.52, df = 2, p-value = 0.003151

There is sufficient evidence to reject the null hypothesis in this case that the means are equal.

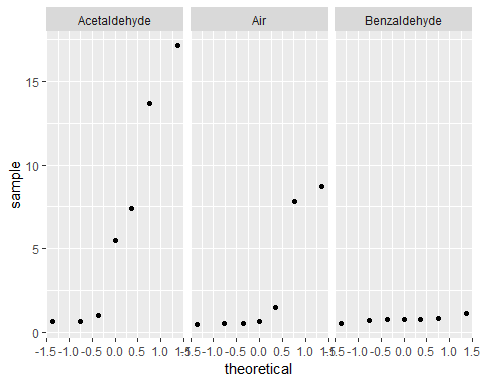
## Post HOC tests## non-parametric non-paired.

13.8.4 Refer to Example 13.8.1. Another variable of interest to Lacroix et al. was the number of alveolar cells in three groups of subjects exposed to air, benzaldeghyde, or acetaldehyde. The following table gives the information for six guinea pigs in each of the three treatment groups.

CELL <- read.csv('Alveolar\_cells.csv')  
CELL\_stacked <- read.csv('Alveolar\_cells - stacked.csv')  
boxplot(CELL)



gf\_qq(~Cells|Gas, data=CELL\_stacked)



favstats(Cells~Gas, data=CELL\_stacked)

## Gas min Q1 median Q3 max mean sd n  
## 1 Acetaldehyde 0.65 0.820 5.48 10.560 17.11 6.5714286 6.6450845 7  
## 2 Air 0.48 0.550 0.65 4.655 8.72 2.8942857 3.6918417 7  
## 3 Benzaldehyde 0.56 0.755 0.81 0.820 1.11 0.8042857 0.1628759 7  
## missing  
## 1 0  
## 2 0  
## 3 0

kruskal.test(Cells ~ Gas, data=CELL\_stacked)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Cells by Gas  
## Kruskal-Wallis chi-squared = 2.7342, df = 2, p-value = 0.2548

Fail to reject the null.