Chapter 8: The Completely Randomized Design

8.1 a. Yes, the mean for Device A is considerably (relative to the standard deviations) smaller than the mean for Device D.

b. $H_o : \mu_A = \mu_B = \mu_C = \mu_D$ versus H_a : Difference in $\mu' s$

Reject H_o if $F \geq F_{.05,3,20} = 3.10$

$SSW = 5[(.1767)^2 + (.2091)^2 + (.1532)^2 + (.2492)^2] = 0.8026$

$\bar{y}_{..} = 0.0826 \Rightarrow$

$SSB = 6[(-0.1605 - .0826)^2 + (0.0947 - .0826)^2 + (0.1227 - .0826)^2 + (0.2735 - .0826)^2]$
$= 0.5838 \Rightarrow$

$F = \frac{.5838/3}{.8026/20} = 4.85 > 3.10 \Rightarrow$

Reject H_o and conclude there is significant difference among the mean difference in pH readings for the four devices.

c. p-value $= P(F_{3,20} \geq 4.85) \Rightarrow$ p-value $= 0.0107$

d. The data must be independently selected random samples from normal populations having the same value for σ.

e. Suppose the devices are more accurate at higher levels of pH in the soil, and if by chance all soil samples with high levels of pH are assigned to a particular device, then that device may be evaluated as more accurate based just on the chance selection of soil samples and not on a true comparison with the other devices.

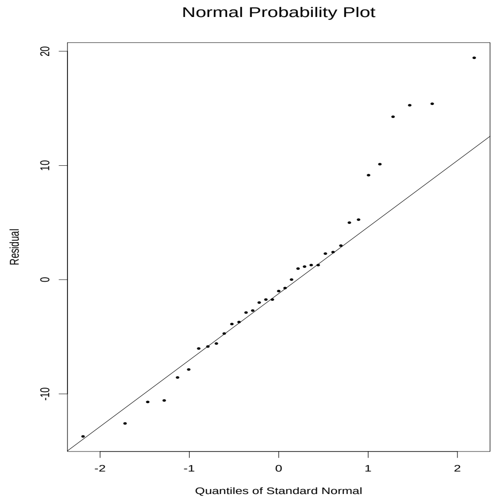
8.2 a. Yes, because the box for the LowTar Brand is completely below all the other boxes.

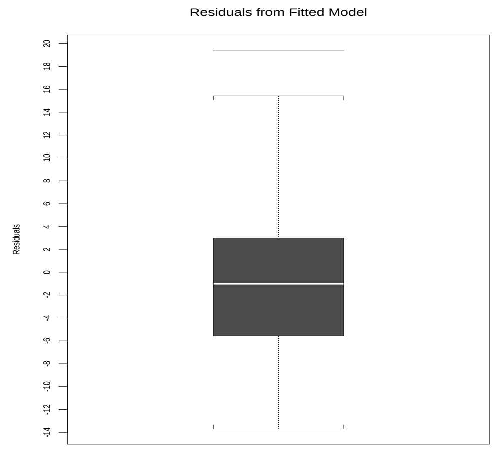
b. Yes, because the F-test has an extremely low p-value (less than 0.001).

c. p-value < 0.001

d. We would be stating that the brands have different average tar content when in fact they have the same level of tar content. Thus, the manufacturer's claim about producing a brand having a lower average tar content would be false. Consumers may change brands and not be receiving the benefit of a lower tar content.

8.3 a. The box plot and normal probability plot are given here:





Based on the box plots and normal probability plots it appears that the distributions of several of the varities have a nonnormal distribution. The Levine's test yields L = 2.381 with p-value = 0.74. Thus, there is not a significant difference in the variety variances.

b. The AOV table is given here:

Source	df	SS	MS	F	p-value
Variety	4	1096.7	274.2	3.73	0.014
Error	30	2205.4	73.5		
Total	34	3302.1			

Reject H_o if $F \geq 2.69 \Rightarrow$

Since $F = 3.73 > 2.69$, reject H_o and conclude there is a significant difference in the mean yield of five varieties.

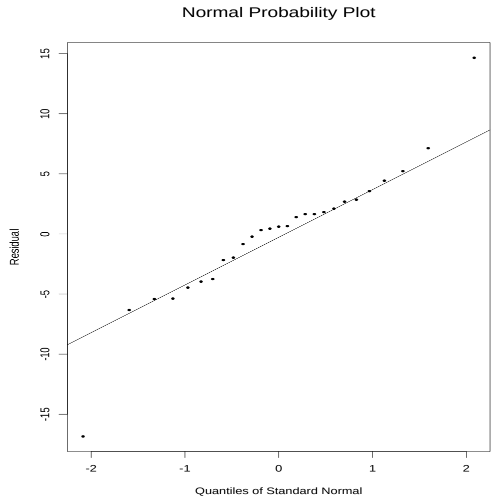
c. The Kruskal-Wallis yields $H' = 10.01$ with p-value = 0.040. Thus, reject H_o and conclude there is a significant difference in the distributions for the yields of the five varieties.

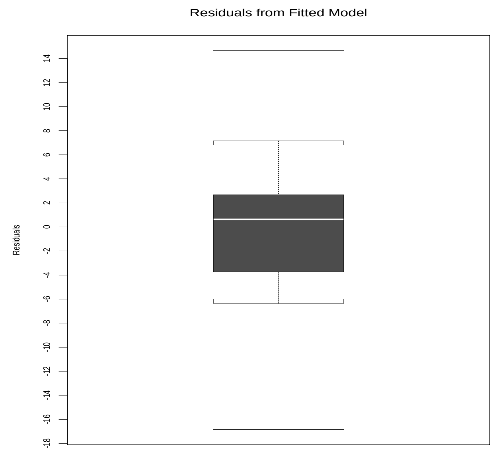
d. The only difference in the conclusions is that the p-value for the F-test is somewhat smaller than the p-value for the Kruskal-Wallis test.

8.4 The F-test is testing for a difference in the mean yields of the five varieties; whereas the Kruskal-Wallis test is testing whether there is a difference in the distribution of yields for the five varieties.

8.5 a. The Kruskal-Wallis yields $H = 21.32 > 9.21$ with $df = 2 \Rightarrow$ p-value < 0.001. Thus, reject H_o and conclude there is a significant difference in the distributions of deviations for the three suppliers.

b. The box plots and normal probability plot are given here:





The box plots and normal probability plots of the residuals indicate that the normality condition may be violated. The Levine test yields $L = 3.89$ with $0.025 <$ p-value $<$ 0.05 . Thus, there is significant evidence that the equal variance condition is also violated

c. The AOV table is given here:

Source	df	SS	MS	F	p-value
Supplier	2	10723.8	5361.9	161.09	0.000
Error	24	798.9	33.3		
Total	26	11522.7			

Reject H_o if $F \geq 3.40$

$F = 161.09 > 3.40$, reject H_o and conclude there is a significant difference in the mean deviations of the three suppliers.

d. 95% C.I. on μ_A : $189.23 \pm (2.064)(5.77)/\sqrt{9} \Rightarrow (185.26, 193.20)$

95% C.I. on μ_B : $156.28 \pm (2.064)(5.77)/\sqrt{9} \Rightarrow (152.31, 160.25)$

95% C.I. on μ_C : $203.94 \pm (2.064)(5.77)/\sqrt{9} \Rightarrow (199.97, 207.91)$

Since the upper bound on the mean for supplier B is more than 20 units less than the lower bound on the mean for suppliers A and B, there appears to be a practical difference in the three suppliers. However, because the normality and equal variance assumptions may not be valid, the C.I.'s may not be accurate.

8.6 For Tukey's W: $q_{.05}(3, 12) = 3.77 \Rightarrow \quad W = (3.77)\sqrt{\frac{.00815}{5}} = .1522 \Rightarrow$

Pairs Not Significantly Different: $(V2, V3)$

8.7 The box plots indicate the distribution of the residuals is slightly right skewed. This is confirmed with an examination of the normal probability plot. The Hartley test yields

$F_{max} = 2.35 < 7.11$ using an $\alpha = 0.05$ test. Thus, the conditions needed to run the ANOVA F-test appear to be satisfied. From the output, $F = 15.68$, with p-value $< 0.0001 < 0.05$. Thus, we reject H_o and conclude there is significant evidence of a difference in the average weight loss obtained using the five different agents.

8.8 a. Fisher's LSD: Pairs Not Significantly Different: (4,1), (2,3)

b. Tukey's W: Pairs Not Significantly Different: (4,1), (4,2), (1,2), (2,3), (3,S)

8.9 a. Based on the box plots and the normal probability plot, the condition of normality of the population distributions appears to be satisfied. The Hartley test yields:

$F_{max} = \frac{(1.0670)^2}{(0.8452)^2} = 1.59 < 14.5$ (value from Table 12 with $\alpha = .01$) $\Rightarrow$ There is not significant evidence of a difference in the 4 population variances.

b. From the ANOVA table, we have p-value < 0.001. Thus, there is significant evidence that the mean ratings differ for the four groups.

c. 95% C.I. on $\mu_I : 8.3125 \pm 2.048\frac{\sqrt{0.9763}}{\sqrt{8}} = (7.6, 9.0)$

95% C.I. on $\mu_{II} : 6.4375 \pm 2.048\frac{\sqrt{0.9763}}{\sqrt{8}} = (5.7, 7.1)$

95% C.I. on $\mu_{III} : 4.0000 \pm 2.048\frac{\sqrt{0.9763}}{\sqrt{8}} = (3.3, 4.7)$

95% C.I. on $\mu_{IV} : 2.5000 \pm 2.048\frac{\sqrt{0.9763}}{\sqrt{8}} = (1.8, 3.2)$

d. The C.I.'s are the same as those given in the output.

8.10 For Tukey's W: $q_{.05}(4, 28) = 3.87 \Rightarrow \quad W = (3.87)\sqrt{\frac{.953}{8}} = 1.336 \Rightarrow$

All pairs are significantly different.

8.11 a. The Kruskal-Wallis test yields: $H' = 26.62$ with df = 3 $\Rightarrow$ p-value $< 0.0001 \Rightarrow$

Thus, there is significant evidence that the distribution of ratings differ for the four groups.

b. The two procedures yield equivalent conclusions.

8.12 a. All middle-grade students from a wide variety of academic abilities and backgrounds.

b. Sixth-grade students with academic ability and background similar to the students in the study.

c. One; one class per treatment group

d. Only sixth-grade students were chosen, all students had similar ability and background, only one sample for each treatment, only one teacher

e. Randomly assign 30 classes to each of the three groups. Each group's classes should vary in grade level. The student's in the classes should be randomly selected from a population of classes which vary in ability and background. Randomly select 30 teachers with each teacher teaching one class from each group. There a number of other factors that should be considered.

8.13 a. F $= \frac{4020.0/3}{881.9/36} = 54.70$ with df = 3,36 $\Rightarrow$ p-value $< 0.0001 < 0.05 \Rightarrow$

There is significant evidence of a difference in the average leaf size under the four growing conditions.

b. 95% C.I. on $\mu_A : 23.37 \pm 2.028 \frac{\sqrt{881.9/36}}{\sqrt{10}} = (20.20, 26.54)$

95% C.I. on $\mu_B : 8.58 \pm 2.028 \frac{\sqrt{881.9/36}}{\sqrt{10}} = (5.41, 11.75)$

95% C.I. on $\mu_C : 14.93 \pm 2.028 \frac{\sqrt{881.9/36}}{\sqrt{10}} = (11.76, 18.10)$

95% C.I. on $\mu_D : 35.35 \pm 2.028 \frac{\sqrt{881.9/36}}{\sqrt{10}} = (32.18, 38.52)$

The C.I. for the mean leaf size for Condition D implies that the mean is much larger for Condition D than for the other three conditions.

c. $F = \frac{18.08/3}{103.17/36} = 2.10$ with df = 3,36 $\Rightarrow$ p-value = 0.1174 $\Rightarrow$

There is not significant evidence of a difference in the average nicotine content under the four growing conditions.

d. From the given data, it is not possible to conclude that the four growing conditions produce different average nicotine content.

e. No. If the testimony was supported by this experiment, then the test conducted in part (c) would have had the opposite conclusion.

8.14 The plots indicate that for both leaf size and nicotine content the normality and equal variance conditions have not been violated.

8.15 a. Because the data appears to be nonnormal in distribution, the Levene test will be used.

$L = 0.84$ with $df = 2, 17 \Rightarrow$ p-value = 0.4489 $\Rightarrow$

There is not significant evidence of a difference in population variances.

b. No, because the distributions appear to be not normally distributed.

c. The data was transformed using the logarithmic transformation.

The data is still somewhat skewed right for Machine C but the ANOVA will be conducted anyways.

$F = \frac{9.848/2}{11.964/17} = 7.00$ with df = 2,17 $\Rightarrow$ p-value = 0.0061 $\Rightarrow$

There is significant evidence of a difference in the mean diameters of the three machines.

d. The analysis using the original data yield p-value = 0.094 which would not support a difference in the mean diameters.

e. She could have selected an equal number of observations from each machine. Equal sample sizes has a moderating influence on the unequal variances effects on the F-test.

8.16 The Kruskal-Wallis test yields identical results for the transformed and original data because the transformation was strictly increasing which maintains the order of the data after the transformation has been performed.

$H = 9.89$ with df = 2 $\Rightarrow$ p-value = 0.0071.

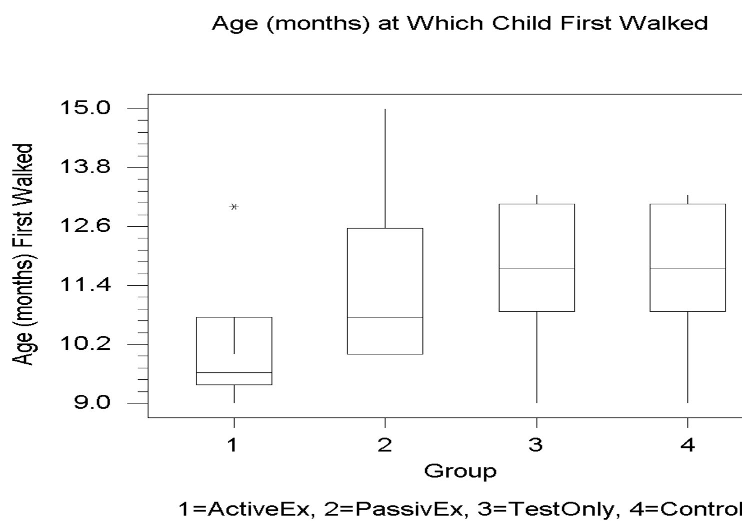
Thus, our conclusion is the same as was reached using the transformed data.

8.17 a. The summary statistics for the four groups are given here.

```
Descriptive Statistics: ActiveEx, PassivEx, TestOnly, Control
```

Variable	N	Mean	Median	TrMean	StDev	SE Mean
ActiveEx	6	10.125	9.625	10.125	1.447	0.591
PassivEx	6	11.375	10.750	11.375	1.896	0.774
TestOnly	6	11.708	11.750	11.708	1.520	0.621
Control	5	12.350	12.000	12.350	0.962	0.430

b. Box plots are given here:



c. The Control group has the largest mean age. The four groups have similar levels of varibility in ages with the Control groups standard deviation somewhat smaller than the other three groups. The box plots are not very informative because of the very small sample sizes in each of the four groups.

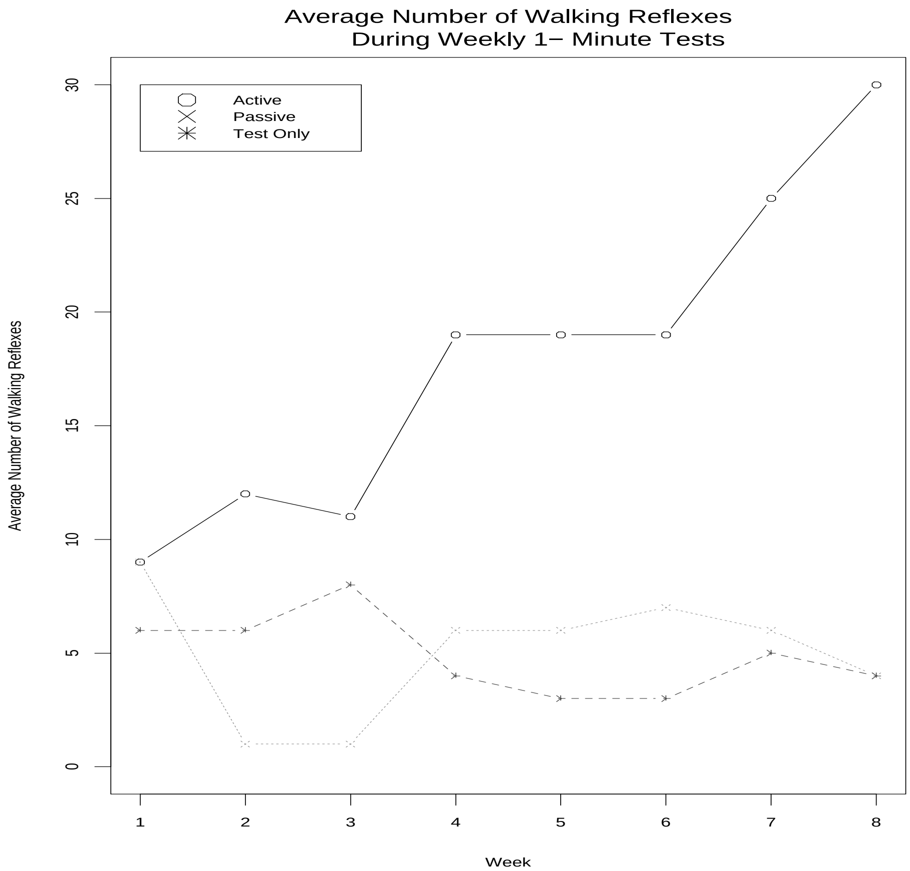
d. The F-test from the Minitab AOV has a p-value of 0.129 which indicates that there is not significant evidence in the data that the four groups have different mean ages.

e. The Minitab output for the LSD and Tukey procedure are given here.

```
Tukey's pairwise comparisons

    Family error rate = 0.0500
Individual error rate = 0.0111

Critical value = 3.98

Intervals for (column level mean) - (row level mean)

          ActiveEx    Control     PassivEx

Control       -4.809
               0.359

PassivEx      -3.714      -1.609
               1.214       3.559

TestOnly      -4.047      -1.942      -2.797
               0.881       3.226       2.131


Fisher's pairwise comparisons

Family error rate = 0.191
Individual error rate = 0.0500

Critical value = 2.093

Intervals for (column level mean) - (row level mean)

            ActiveEx     Control    PassivEx

Control       -4.147
              -0.303

PassivEx      -3.082      -0.947
               0.582       2.897

TestOnly      -3.416      -1.280      -2.166
               0.249       2.564       1.499
```

In Tukey's procedures, the 6 pairwise comparisons intervals all contain 0. Therefore, Tukey's procedure agrees with the conclusion from the F-test that there is not significant evidence of a difference in the four group means. Fisher's pairwise comparisons procedure has the interval for comparing the mean of the Control with the mean of ActiveEx group not containing 0 but with the other 5 pairwise comparisons intervals all containing 0. This would seem to contradict the F-test and Tukey conclusions. However, if we use the Protected Fisher LSD procedure, then no further comparisons would have been made once the conclusion of the F-test indicated no significant evidence of a difference. Thus, all three procedures, F-test, Tukey and Protected Fisher LSD, have reached the same conclusion: No significant evidence at the .05 level of a difference in the means of the four groups.

f. A plot of the average number of walking reflexes observed is given here for the three groups.



The average number of walking reflexes for the active exercise group increases dramatically over the 8 weeks whereas the values for the Passive and Test Only groups remains relatively constant.

8.18 a. Box plots are given here:

The box plots are all relatively symmetric with about the same level of spread. There appears to be some differences in their medians. This is confirmed by an examination of the summary statistics: Median, StDev, Q1 and Q3.

b. There is no strong indication from the box plots that the assumptions of equal variance and normality appear to be violated.

c. The assumption of normality of the residuals does not appear to be violated based on the normal probability plot. The data values are mostly close to the straight line with the exception of the largest values for the residuals.

d. The Levene's test yields $F = 0.416,\ \ df = 4, 65 \Rightarrow$ p-value $= 0.797$. Thus, there is not significant evidence in the data that the treatment variances are different.

e. The mean rating from the applicant using crutches was significantly higher than the mean rating for the applicant who was hard of hearing. No other pairs of rating means were found to be significantly different. Thus there is very little evidence that the various types of physical handicaps produce different mean empathy ratings.