Chapter 5 Solutions

5.1 True. If observations are randomly selected from the populations, we can generalize our results to those populations.

5.2 False. The samples must be representative of the population in order to generalize the results to the population. The only way to ensure this is to have random selection.

5.3 False. Transformations can only help improve the situation if there are problems with normality or equal variances.

5.4 True. This method has a larger family-wise error rate, but it has a smaller chance of missing actual differences that exist. We are comfortable stating that differences we find are not due to chance, because the ANOVA F-test was significant.

5.5 (c). ANOVA is an extension of the two-sample t-test in which we are trying to compare the means of several populations.

5.6 (b). In fact, since we want the model to be unbiased, we expect that some of the errors are positive and others are negative.

5.7 (d). We use the same plots in the ANOVA situation as we used with the regression model.

5.8 (c). If the group standard deviations are not the same, sometimes a transformation may result in a data set where the group standard deviations are more similar.

5.9 Randomization protects against bias, permits conclusions about cause, and justifies using a probability model.

5.10 a. If the groups have very different standard deviations, and a transformation might help to equalize them.

b. This plot helps us find which reexpression technique might be useful. If a straight line fits the graph, compute the slope and consider a power transformation using $1 -$ slope as the power (where power $= 0$ means take logs).

5.11 a. The factor of interest (ethnicity) cannot be assigned. The study is observational, and inference about cause is not justified.

b. Inference from sample to population requires that the samples be random, that is, chosen from the populations of interest using random numbers. This sample is a random sample so, yes, we can infer about the population.

5.12 a. Answers will vary. We give three possibilities. **One:** The factor of interest is observational. Because favorite music could not be randomly assigned, inference about cause is not justified here. **Two:** (Basically, this is a concrete instance of reason one above.) It is not possible to rule out that any causal relationship goes in the other direction: Teens who speed prefer heavy metal. **Three:** (This is not a particular instance of reason one above.) The response is not the actual frequency of speeding, but the *reported* frequency, which might reflect actual behavior, but might not. There is no way to use the data to rule out the possibility, for example, that teens who prefer heavy metal tend to exaggerate how often they speed.

b. The samples were randomly chosen, so generalizing from sample to population is justified. We can conclude that for the teens at the high school in question, those who prefer heavy metal tend to report a higher frequency of driving over 80 mph.

5.13

- Categorical, with many categories. ANOVA since there are more than two categories.
- Categorical, with only two categories—Male and Female. Two-sample t-test since there are only two categories.
- Categorical, with five categories. ANOVA since there are more than two categories.
- Categorical, with three categories. ANOVA since there are more than two categories.
- Quantitative.
- Quantitative.
- Quantitative.
- Quantitative.

5.14 a. ANOVA is exactly the right analysis to use *because* there are four groups. ANOVA is designed to analyze the differences between means for at least two populations.

b. The response variable is the age of the cars, which is quantitative.

c. As long as the conditions are met, the different sample sizes are okay.

d. Since the data were generated through the use of a random sample, you can generalize to the larger population.

5.15 a. The explanatory variable is type of font, and there are four different fonts being used. The response variable is the final exam score, which is being used as a measure of student performance.

b. This was a randomized experiment because the treatments (fonts used on the exam) were randomly assigned to the subjects (students in the course).

c. The subjects were randomly assigned to treatments, which makes the observations independent of each other.

5.16 a. The explanatory variable is type of dog food, and there are three different types of dog food being used. The response variable is the average number of hours of sleep per 24 hours, which is being used as a measure of energy.

b. This was a randomized experiment because the treatments (type of dog food) were randomly assigned to the subjects (Border Collies).

c. The subjects were randomly assigned to treatments, which makes the observations independent of each other.

5.17 a. The 40 students in the class.

b. The 4 different fonts.

c. A design is balanced if all treatments are assigned to the same number of units. Since we have 40 units and 4 treatments, randomly assign 10 students to each font.

5.18 a. The 45 Border Collies available for the experiment.

b. The 3 different dog foods.

c. A design is balanced if all treatments are assigned to the same number of units. Since we have 45 units and 3 treatments, randomly assign 15 Border Collies to each type of dog food.

5.19 The degrees of freedom for type of font is $K - 1 = 4 - 1 = 3$. The degrees of freedom for error is $n - K = 40 - 4 = 36$. The total degrees of freedom is $n - 1 = 40 - 1 = 39$.

5.20 The degrees of freedom for type of dog food is $K - 1 = 3 - 1 = 2$. The degrees of freedom for error is $n - K = 45 - 3 = 42$. The total degrees of freedom is $n - 1 = 45 - 1 = 44$.

5.21 a. If the four groups all have the same mean score, there is a 0.003 probability of collecting sample data in which the sample mean scores are as or more different than those displayed by this sample.

b. We have only shown that there is some difference among the four fonts, not that they are all different. It could be that only one is different from the other three.

c. We cannot make any conclusion about specific differences that exist, only that at least one difference exists.

d. We do have evidence that at least one difference exists.

e. Since the treatments were randomly assigned to subjects, we can make a cause-and-effect conclusion.

f. Conclusions from this analysis can only be generalized to the population of students like those in this instructor's class.

5.22 a. Not possible. If all residuals were positive, that would mean that all observations in the group had response values higher than the mean response for the group.

b. Possible. If one observation in the group has either a very high or very low outlier, it could pull the group mean so far toward it that all other values are on the other side of the mean.

c. Possible. Your score may be just slightly higher than the group mean, whereas mine is quite far below the group mean.

d. Possible. Your score is closer to the mean in your group than mine is to the mean in my group.

5.23 a. The null hypothesis is that the average amount of schooling that American adults have had is the same for people of all three political viewpoints. In symbols, this is $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$, where group 1 represents those who are liberal, group 2 represents those who are moderate, and group 3 represents those who are conservative.

b. We need to have the actual values of the amount of schooling for all people in each group in order to conduct the ANOVA.

c. We need to compute the residuals in order to assess whether the conditions are met for ANOVA.

5.24 a. The null hypothesis is that the average size of turtles is the same in all three states. In symbols, this is $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$, where group 1 represents those in Nebraska, group 2 represents those in Oklahoma, and group 3 represents those in Texas.

b. We need to have the actual values of the size of the turtles for all turtles in each group in order to conduct the ANOVA.

c. We need to compute the residuals in order to assess whether the conditions are met for ANOVA.

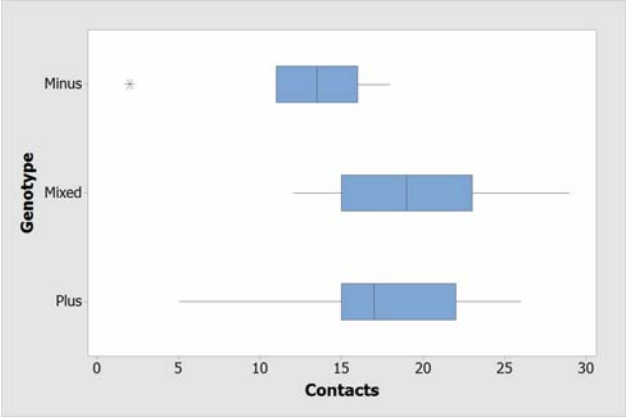
5.25 a. Answers will vary. One possible answer is a common standard deviation of 0.1. With a standard deviation of this size, there will be very little overlap between observations in the three groups. This would suggest that the means are, in fact, different.

b. Answers will vary. One possible answer is a common standard deviation of 3.0. Since the difference between the largest mean and the smallest mean is only 1.4, with a standard deviation of 3.0 in each group, there will be lots of overlap between observations in the three groups which makes it harder to conclude that there is a true difference in the population means.

5.26 a. Answers will vary. One possible answer is a common standard deviation of 0.2. With a standard deviation of this size, there will be very little overlap between observations in the three groups. This would suggest that the means are, in fact, different.

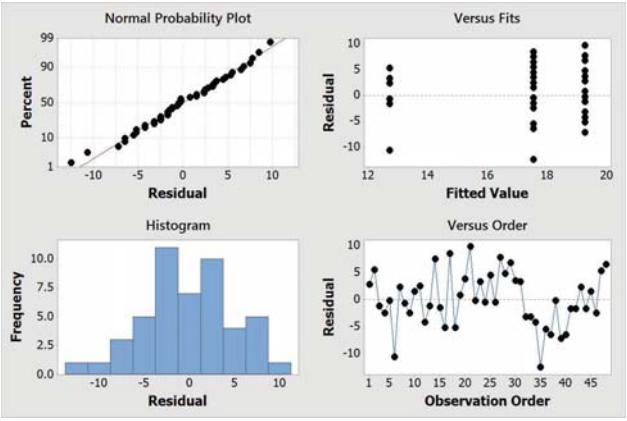
b. Answers will vary. One possible answer is a common standard deviation of 8.0. Since the difference between the largest mean and the smallest mean is only 4.1, with a standard deviation of 8.0 in each group, there will be lots of overlap between observations in the three groups, which makes it harder to conclude that there is a true difference in the population means.

5.27 a. The boxplots suggest that the Mixed and Plus groups are similar, but the Minus group seems to have fewer contacts. This is also seen when comparing the means of the three groups.

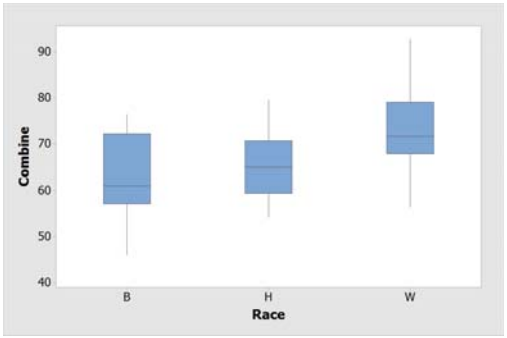


Variable	Genotype	N	N*	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Contacts	Minus	10	0	12.70	4.52	2.00	11.00	13.50	16.00	18.00
	Mixed	19	0	19.26	4.79	12.00	15.00	19.00	23.00	29.00
	Plus	19	0	17.53	5.46	5.00	15.00	17.00	22.00	26.00

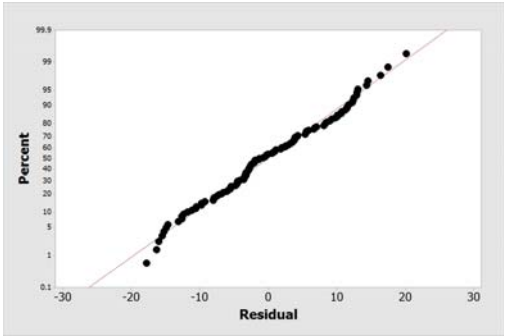
b. The residuals plot shows that the normality condition is met. The standard deviations from part (a) are close enough that the equal variances condition is met. And the mice should be independent of each other.



5.28 a. The following are the boxplots of the combined scores, broken down by the race of the individual taking the exam. It appears that there may be a difference between the scores of the various races. Specifically, it appears that whites generally did a little bit better, that Hispanics had scores in the middle, and that blacks had the lowest scores, though the difference between Hispanics and blacks does not appear to be large and might turn out to be statistically insignificant. The spreads of the three samples appear to be similar so continuing by conducting an ANOVA analysis would likely be the next step.



b. To check the conditions necessary for conducting an ANOVA we begin by checking the normality of the residuals. Following is a normal plot of the residuals in this case. While there is a small amount of curvature to this plot, there is not enough to be worried about.



Next we check whether the data are consistent with the idea that the variances for the three groups are the same. The residuals versus fitted values plot suggests that the data are comparable with the equal variances requirement. Also the ratio of the largest standard deviation to the smallest is $\frac{8.83}{7.14} = 1.24$ which is much less than our suggested cutoff of 2. Finally we need to evaluate both the independence of the errors and the idea that they come from a population with mean 0. In this case, there is no reason to believe that any one individual's exam score is related to any other individual's score, so the independence of the errors seems reasonable. We also have no reason to believe that there is any bias in these scores other than the one that we are modeling, so a mean error of 0 also seems reasonable.

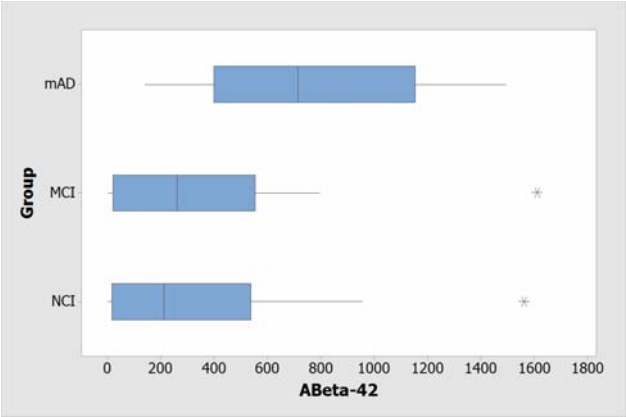
5.29 The dotplot suggests that there may be very different amounts of variability between the three different types of county. In particular, it appears that small counties have very little variability with respect to child poverty rates as do medium and large counties. Therefore, we are concerned about the equal variances condition.

5.30 The distribution of values for the low concentration group appear to be skewed to the left. There might be some concern about normality.

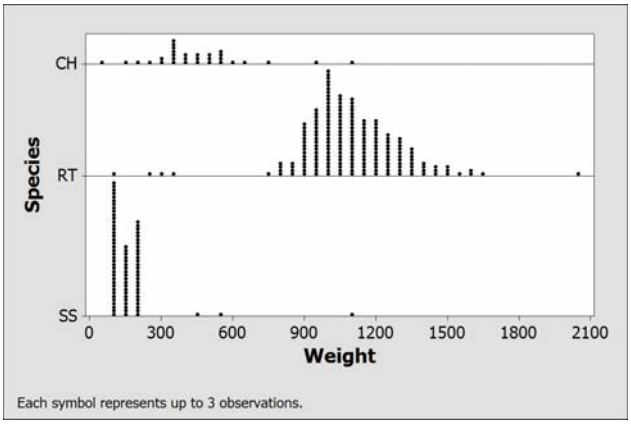
5.31 a.

Variable	Group	N	Mean	StDev
ABeta-42	mAD	17	761.3	426.7
	MCI	21	341.0	406.4
	NCI	19	336.3	435.6

b. Parallel boxplots show that the MCI and NCI groups have skewed distributions, so the normality condition for ANOVA is not met.



5.32 a. The dotplots are shown below. There is right-skewness in the weights of the sharp-shinned hawks. There are outliers in all three groups, which suggests concerns about normality and the weights for the sharp-shinned hawks have much less variability than the other two groups.



b. The standard deviations are

Variable	Species	StDev
Weight	CH	162.0
	RT	189.2
	SS	80.65

The ratio of largest standard deviation to smallest standard deviation is $\frac{189.2}{80.65} = 2.35$, and these are large samples, so we have concerns about the equal variances condition.

5.33 a. The degrees of freedom for Occupation can be found by subtracting the degrees of freedom for Error from the degrees of freedom for Total $DF = 972 - 968 = 4$. The sum of squares for Occupation can be found by subtracting the sum of squares for Error from the sum of squares for Total. $SSGroups = 206,147 - 195,149 = 10,998$. The F can be found by dividing the mean square for Occupation by the mean square for Error. $F = \frac{2749}{202} = 13.609$.

b. Since the degrees of freedom for Occupation is 4 and the degrees of freedom for treatments is the number of groups minus 1, the number of different occupations considered is 5.

c. Since the P-value is so small, we reject the null hypothesis. We have strong evidence that people in at least one of the occupations studied have a different average life span than people in at least one other occupation.

5.34 a. The degrees of freedom for aphid/plant combination can be found by subtracting the degrees of freedom for Error from the degrees of freedom for Total $DF = 51 - 46 = 5$. The sum of squares for aphid/plant combination can be found by subtracting the sum of squares for Error from the sum of squares for Total. $SSGroups = 64.77 - 39.87 = 24.90$. The F can be found by dividing the mean square for aphid/plant combination by the mean square for Error. $F = \frac{4.9807}{0.8667} = 5.75$.

b. Since the degrees of freedom for aphid/plant combination is 5 and the degrees of freedom for treatments is the number of groups minus 1, the number of different occupations considered is 6.

c. Since the P-value is so small, we reject the null hypothesis. We have strong evidence that mean amount of honeydew produced by aphids is different for at least one aphid/plant combination.

5.35 a.

```
One-way ANOVA: meth labs versus type

Source  DF     SS     MS     F
type     2   37.51  18.755  5.101
Error    9   33.09   3.677
Total   11   70.60
```

b. The MS for county type measures the amount of variability between the means of the three county types.

c. Using an F-distribution with 2 and 9 degrees of freedom, the P-value is 0.033.

d. The null hypothesis is that the three sizes of counties have the same average number of meth labs. In symbols, this is $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$. The alternative hypothesis is that at least one type of county has a different average number of meth labs than the other two sizes of county. With a P-value of 0.033, we reject the null hypothesis and conclude that at least one type of county has a different average number of meth labs.

5.36 a.

```
One-way ANOVA: Score versus Coarseness

Source       DF        SS     MS         F
Coarseness    1     10609   10609      24.3
Error        14      6113   436.642
Total        15     16722
```

b. The MS for coarseness measures the amount of variability in mean palatability score for different coarseness levels.

c. Using an F distribution with 1 and 14 degrees of freedom, the P-value is 0.0002.

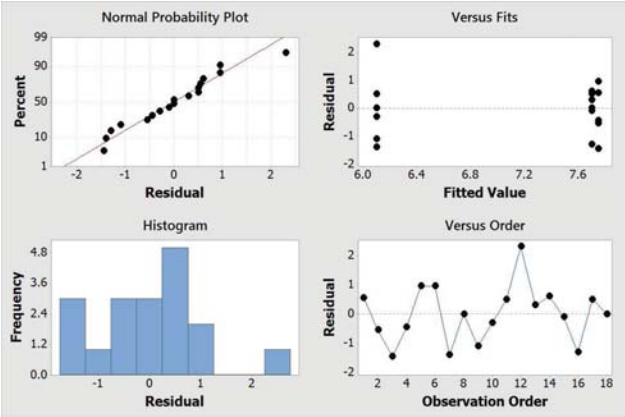
d. The null hypothesis is that the two levels of coarseness have the same average palatability score. In symbols, this is $H_0 : \alpha_1 = \alpha_2 = 0$. The alternative hypothesis is that the two levels of coarseness have different average palatability scores. With a P-value of 0.0002, we reject the null hypothesis and conclude that the mean palatability scores are different for the two levels of coarseness.

5.37 The null hypothesis is that the mean number of contacts is the same for all three types of mice. The ANOVA table given below gives a P-value of 0.006. We have significant evidence that the mean number of contacts is different for at least one of the three types of mice.

```
Analysis of Variance

Source    DF  Adj SS  Adj MS  F-Value  P-Value
Genotype   2   285.4  142.70     5.66    0.006
Error     45  1134.5   25.21
Total     47  1419.9
```

5.38 a. The hypotheses are $H_0 : \alpha_a = \alpha_f = \alpha_v = 0$ vs. H_a : at least one $\alpha_k \neq 0$ where (a) stands for the meniscus arrow treatment group, (f) stands for the FasT-Fix treatment group, and (v) stands for the vertical suture treatment group, and the means are the mean values of the stiffness level.

b. The graphs of the residuals are given below. The residuals are reasonably normally distributed and show similar variability. The actual standard deviations range from 0.693 (FasT-Fix) to 1.327 (meniscus arrow). These are close enough that an ANOVA procedure is acceptable.



c. The ANOVA table is given below. Since the P-value of 0.022 is relatively small, we reject the null hypothesis. It appears that there is at least one treatment that has a different amount of stiffness in comparison to the other two treatments.

```
Analysis of Variance

Source  DF  Adj SS  Adj MS  F-Value  P-Value
Method   2   10.57   5.285     4.98    0.022
Error   15   15.91   1.061
Total   17   26.48
```

5.39 a. The null hypothesis is that the mean combined exam score is the same for the populations of white, black, and Hispanic firefighters in New Haven, Connecticut. The alternative is

that at least one group of firefighters has a different mean combined exam score. In symbols we should write

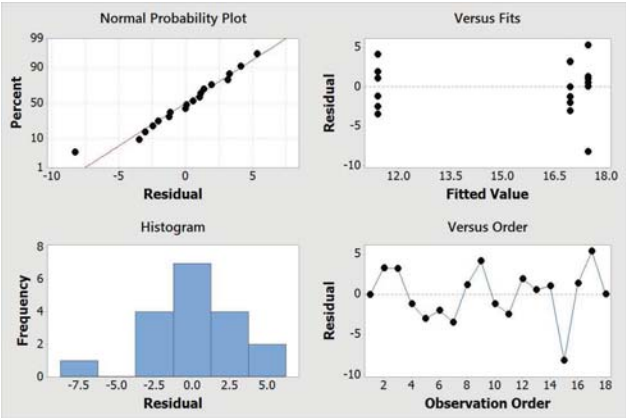
$H_0 : \alpha_B = \alpha_H = \alpha_W = 0$
H_a : at least one $\alpha_k = 0$

b. The ANOVA suggests that there is a difference for the mean combined exam score for at least one of the three groups of firefighters. The F-statistic is 13.60 (with 2 and 115 df), which results in a P-value near 0 (Minitab reports less than 0.001). Looking at the boxplots from Exercise 5.28 part (a) it would appear that white firefighters have a higher mean score than the other two groups. Without further testing, we cannot tell whether there is a statistically significant difference between blacks and Hispanics. (The output for this ANOVA follows.)

Source	DF	SS	MS	F-Value	P-Value
Race	2	1972	985.83	13.60	0.000
Error	115	8339	72.51		
Total	117	10311			

5.40 a. The hypotheses are $H_0 : \alpha_a = \alpha_f = \alpha_v = 0$ vs. H_a : at least one $\alpha_k = 0$ where (a) stands for the meniscus arrow treatment group, (f) stands for the FasT-Fix treatment group, and (v) stands for the vertical suture treatment group, and the means are the mean values of the displacement level.

b. The graphs of the residuals are given below. The residuals are reasonably normally distributed and show similar variability. The actual standard deviations range from 2.67 (vertical suture) to 4.47 (FasT-Fix). These are close enough that an ANOVA procedure is acceptable.



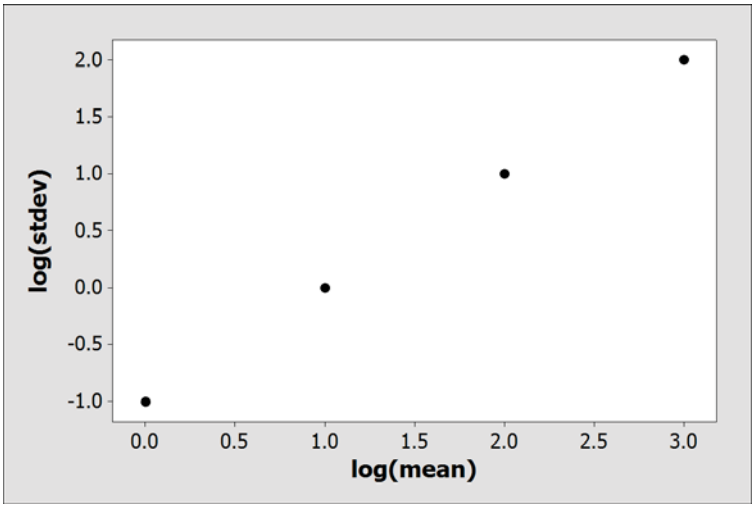
c. The ANOVA table is given below. Since the P-value of 0.014 is relatively small, we reject the null hypothesis. It appears that there is at least one treatment that has a different amount of displacement in comparison to the other two treatments.

```
Analysis of Variance

Source  DF  Adj SS  Adj MS  F-Value  P-Value
Method   2   136.5   68.26     5.78    0.014
Error   15   177.0   11.80
Total   17   313.6
```

5.41 a.

Group	Mean	St. Dev.
A	1.0	0.1
B	10.0	1.0
C	100.0	10.0
D	1000.0	100.0

b. $S_{max}/S_{min} = 100/0.1 = 1000$. This is much larger than 2, so a transformation is called for.



The slope of the line is equal to 1.

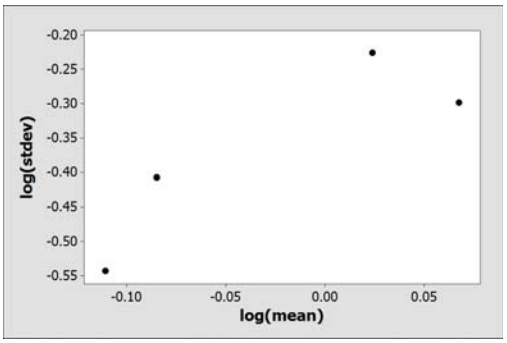
c. $p = 1 - slope = 1 - 1 = 0$. This suggests that the best transformation might be the logarithm.

d.

Group	Mean	St. Dev.
A	−0.0015	0.0436
B	0.9985	0.0436
C	1.9985	0.0436
D	2.9985	0.0436

e. $S_{max}/S_{min} = 0.0436/0.0436 = 1$. Now the standard deviations are all exactly the same, and we have met the equal standard deviations condition.

5.42 a. Here is a plot of log(s) versus log(ave) for the four groups.

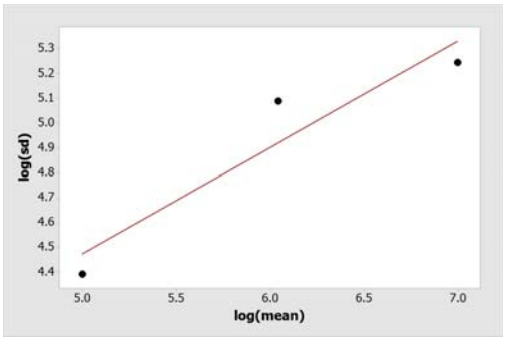


Yes, these points fit fairly close to a straight line.

b. The slope appears to be approximately 1.4.

c. $p = 1 - 1.4 = -0.4$. This suggests the reciprocal of the square root or perhaps, the reciprocal of the third root as an appropriate transformation.

5.43 a. Here is a plot of log(sd) versus log(mean) for the three groups.



Yes, these points fit fairly close to a straight line.

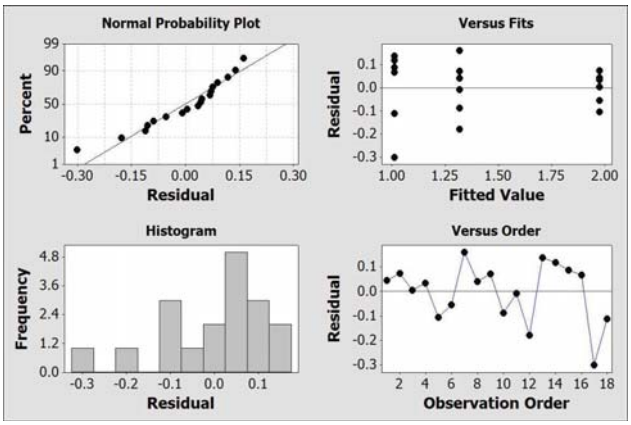
b. The slope is appears to be approximately 0.5.

c. $p = 1 - 0.5 = 0.5$. This suggests the square root as an appropriate transformation.

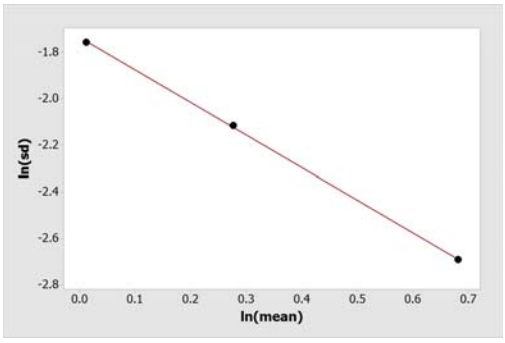
5.44 a. We are interested in seeing if the amount of fenthion is different at different time periods. This means that *fenthion* is the response variable and *time* is the explanatory variable.

b. The samples of olive oil are random and the amount of fenthion measured in individual samples should be independent. To check the normality and equal variances conditions, we look at plots of the residuals. The linear trend in the normal probability plot suggests that the normality condition is okay, but we might have an issue with the equal variances condition.

In fact, the standard deviation at time 0 is 0.0677 and the standard deviation at time 4 is 0.1727, giving a ratio of $\frac{0.1727}{0.0677} = 2.55$. This is larger than our rule of thumb so we will not do the analysis on this data.



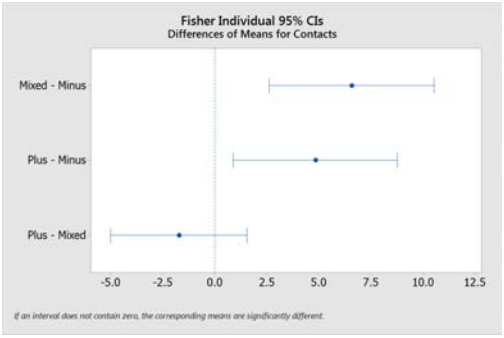
c. Here is a plot of log(sd) versus log(mean) for the three groups.



Yes, these points fit strongly on a straight line. The slope appears to be approximately -1.4, so $p = 1 - (-1.4) = 2.4$. This suggests the square as an appropriate transformation.

5.45 The computer output for Fisher's LSD is given as follows along with a graph showing the confidence intervals.

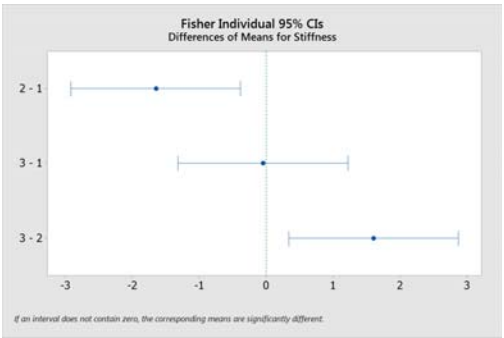
```
Grouping Information Using the Fisher LSD Method and 95% Confidence

Genotype   N   Mean  Grouping
Mixed     19  19.26  A
Plus      19  17.53  A
Minus     10  12.70    B

Means that do not share a letter are significantly different.
```



This analysis shows that while both Mixed and Plus mice are different from Minus mice, there is no significant difference between the mean number of contacts for Mixed and Plus mice.

5.46 The computer output for Fisher's LSD is give below along with a graph showing the confidence intervals.

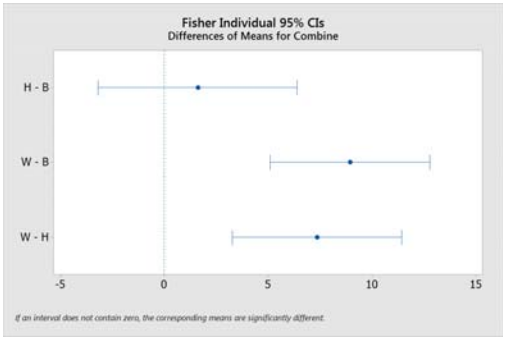
```
Grouping Information Using the Fisher LSD Method and 95% Confidence

Method  N   Mean  Grouping
1       6  7.750  A
3       6  7.700  A
2       6  6.100    B

Means that do not share a letter are significantly different.


Fisher Individual Tests for Differences of Means

Difference  Difference       SE of                              Adjusted
of Levels     of Means  Difference       95% CI       T-Value   P-Value
2 - 1           -1.650       0.595  (-2.918, -0.382)    -2.77     0.014
3 - 1           -0.050       0.595  (-1.318,  1.218)    -0.08     0.934
3 - 2            1.600       0.595  ( 0.332,  2.868)     2.69     0.017
```



This analysis shows that while both vertical sutures and FasT-Fixes are different from meniscus arrows, there is no significant difference between the vertical suture and FasT-Fixes. Since a higher value is better here, what we have shown is that FasT-Fixes, while not alone at being the best method, are, based on this experiment, indistinguishable from vertical sutures. Just based on this study, doctors should only consider meniscus arrows if there is a medical reason why neither of the other two methods can be performed. Finally, this shows that the FasT-Fix is just as good as the vertical suture, so if it is cheaper, or otherwise preferable, it is a reasonable option.

5.47 The following output shows that, in fact, whites have a significantly higher mean score than both of the other groups, but there is no significant difference between blacks and Hispanics.

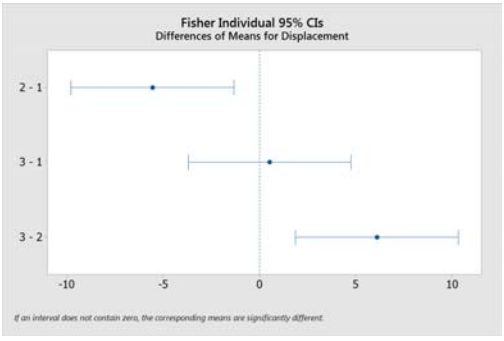
```
Grouping Information Using Fisher Method

Race   N    Mean  Grouping
W     68  72.678  A
H     23  65.337    B
B     27  63.736    B

Means that do not share a letter are significantly different.
```



5.48 The computer output for Fisher's LSD is given as follows along with a graph showing the confidence intervals.

```
Grouping Information Using the Fisher LSD Method and 95% Confidence

Method  N   Mean  Grouping
3       6  17.47  A
1       6  16.95  A
2       6  11.38    B

Means that do not share a letter are significantly different.
```

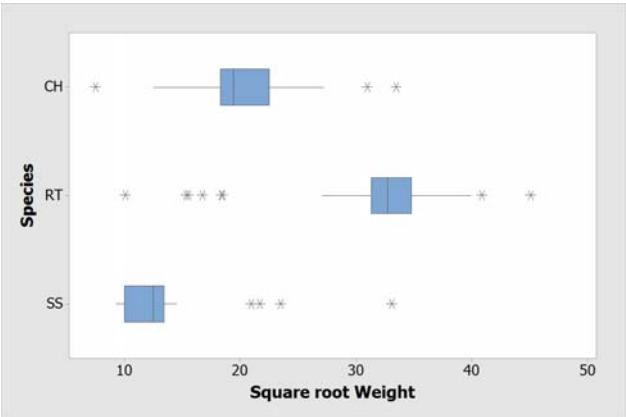
```
Fisher Individual Tests for Differences of Means

Difference  Difference       SE of                                  Adjusted
of Levels     of Means  Difference       95% CI      T-Value   P-Value
2 - 1           -5.57        1.98  (-9.79, -1.34)    -2.81     0.013
3 - 1            0.52        1.98  (-3.71,  4.74)     0.26     0.798
3 - 2            6.08        1.98  ( 1.86, 10.31)     3.07     0.008
```



This analysis shows that while both vertical sutures and FasT-Fixes are different from meniscus arrows, there is no significant difference between the vertical suture and FasT-Fixes. In this case lower values are better, but the meniscus arrow values are artificially low because they failed sooner. The conclusion that we can draw here is that while the FasT-Fix is not significantly better than the vertical sutures, it is also not significantly worse. Basically, this shows that the FasT-Fix is just as good as the vertical suture, so if it is cheaper, or otherwise preferable, it is a reasonable option.

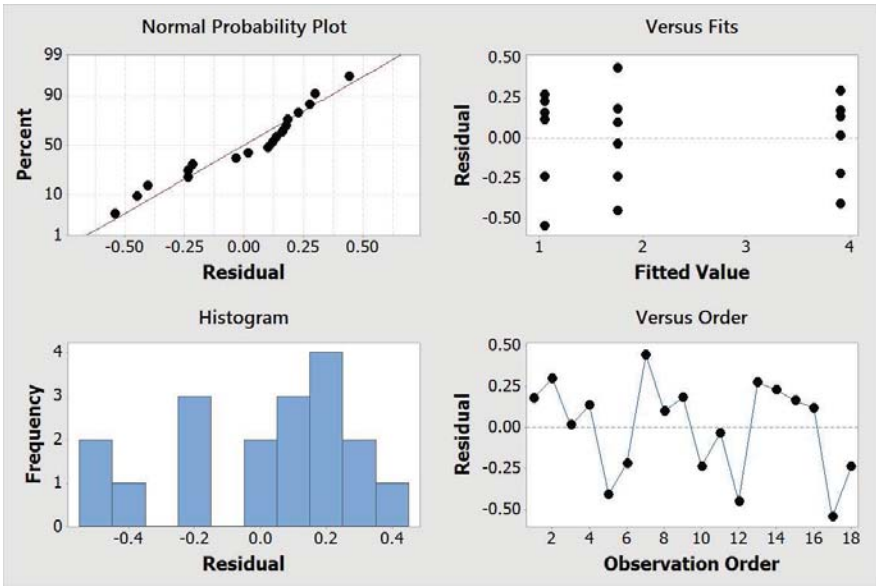
5.49 a. In square root scale, the sample means are 20.144, 32.940, and 11.922. The sample SDs are 3.861, 3.063, 2.422.

b. The parallel boxplots show that weights tend to be the highest for the Red Tailed Hawks, somewhat smaller for the Cooper's Hawks, and the smallest for the Sharp-Shinned Hawks.



c. The boxplots show that while the standard deviations are now fairly similar, all three distributions are still skewed with a large number of outliers, so the normality condition is not met.

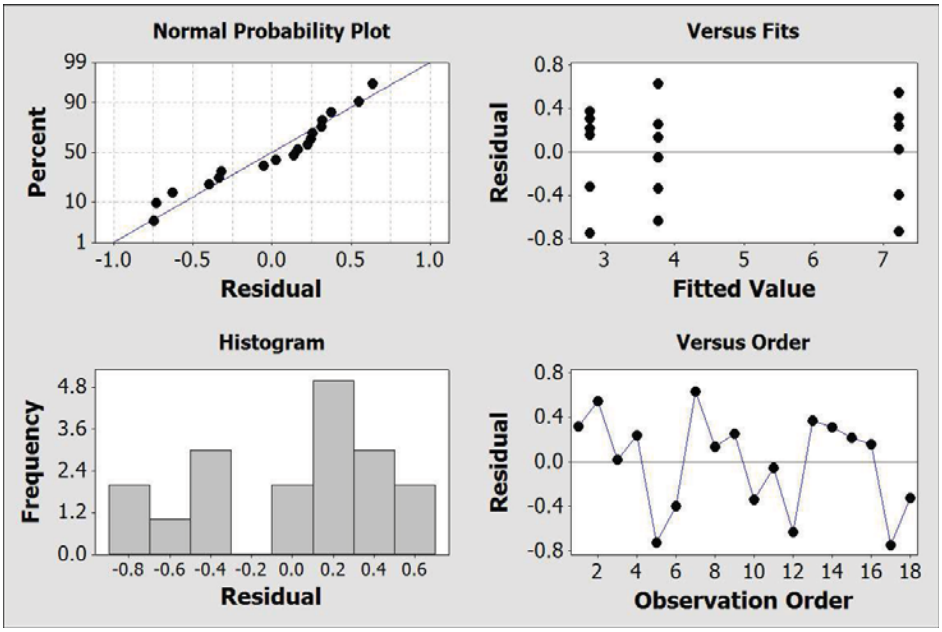
5.50 a. As stated earlier, the samples are random and the exponential amounts of fenthion should be independent. To check the normality and equal variances conditions, we look at plots of the residuals. The plots given below suggest that both conditions hold.



Now we continue with the analysis. The null hypothesis is that the mean square amount of fenthion is the same for all three time periods. The ANOVA table given below gives a P-value of approximately 0. We have significant evidence that the mean square amount of fenthion is different for at least one of the three time periods.

```
Analysis of Variance
Source  DF  Adj SS   Adj MS  F-Value  P-Value
Time     2  26.582  13.2912  145.54   0.000
Error   15  1.370    0.0913
Total   17  27.952
```

b. As stated earlier, the samples are random and the exponential amounts of fenthion should be independent. To check the normality and equal variances conditions, we look at plots of the residuals. The plots given below suggest that both conditions hold.



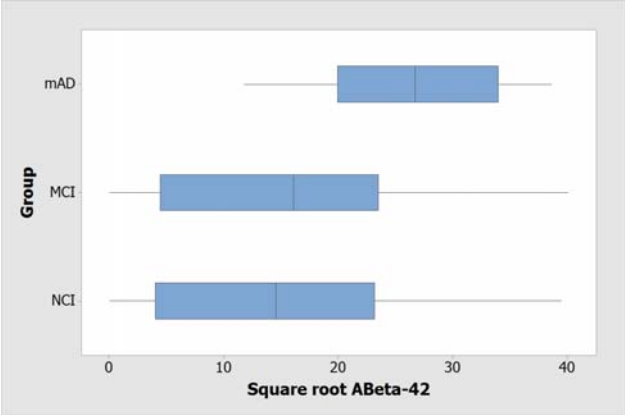
Now we continue with the analysis. The null hypothesis is that the mean exponential amount of fenthion is the same for all three time periods. The ANOVA table given below gives a P-value of approximately 0. We have significant evidence that the mean exponential amount of fenthion is different for at least one of the three time periods.

```
One-way ANOVA: exponential fenthion versus Time

Source  DF      SS      MS       F      P
Time     2  65.244  32.622  155.95  0.000
Error   15   3.138   0.209
Total   17  68.381
```

c. Both transformations are similar in nature. Pick the one that the client feels most comfortable with.

5.51 a. In square root scale, the sample means are 26.4, 14.7, and 14.2. The sample SDs are 8.2, 11.5, and 11.9.

b. The parallel boxplots show that Abeta levels tend to be higher in the mAD group than in the other two groups.



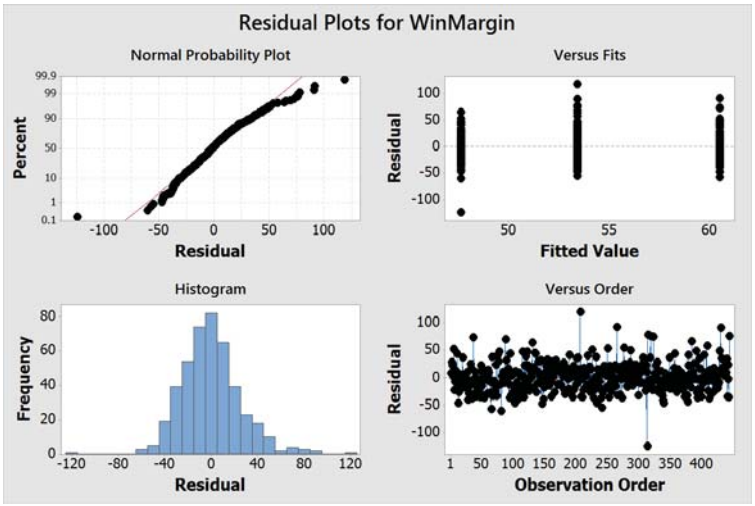
c. Parallel boxplots of the transformed data show reasonable symmetry for each of the groups. The normality condition for ANOVA is now met.

d. `Analysis of Variance`

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Group	2	1702	850.9	7.36	0.001
Error	54	6240	115.5		
Total	56	7942			

The ANOVA F-statistic is 7.36 and the P-value is 0.001. There is strong evidence for the alternative hypothesis that Abeta is related to group membership.

5.52 a. While this is not a random sample of games, they are games played at different times, so the winning margin should be independent from one game to the next. The following graphs of residuals show that both the equal variance and normality conditions are met.

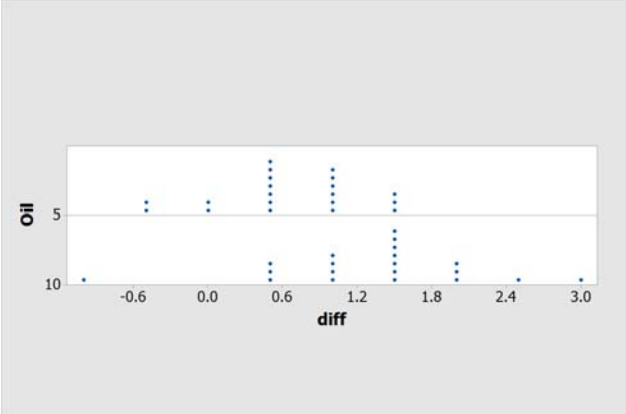


b. The ANOVA table follows:

```
Analysis of Variance
Source DF Adj SS Adj MS F-Value P-Value
Blanks 2 9514 4757.2 6.99 0.001
Error 441 300202 680.7
Total 443 309717
```

The F-statistic is 6.99, which translates into a P-value of 0.001. This is small enough that we feel comfortable rejecting the null hypothesis of no difference in winning margin for the three different groups. It appears that the amount of the winning margin is related to the number of blank tiles that a player receives.

5.53 a. The dotplot and summary statistics are given below.



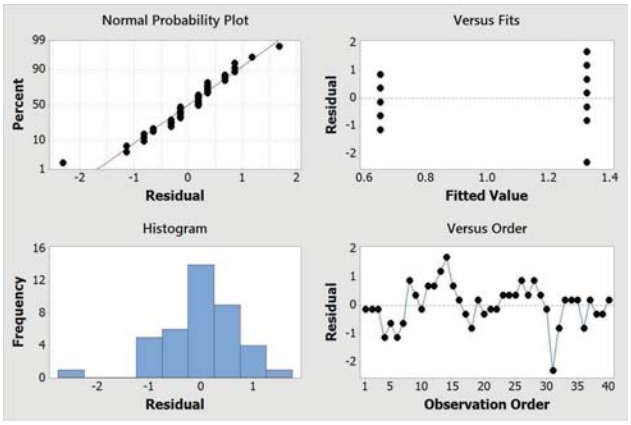
Variable	Oil	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
diff	5	20	0	0.650	0.131	0.587	-0.500	0.500	0.500	1.000	1.500
	10	20	0	1.325	0.189	0.847	-1.000	1.000	1.500	1.875	3.000

From the dotplot it appears that there is a difference in the difference in amount of oil deapsorbed between treated and control samples based on how much oil was in the sample to begin with. The samples with 10 ml of oil in them to begin with showed a bigger difference between treated samples and control samples than the samples with 5 ml of oil in them. The descriptive statistics agree with this, giving a mean difference of 0.65 ml for the samples with 5 ml of oil in them and a mean difference of 1.325 for the samples with 10 ml of oil in them.

b. The output from the 2 sample t-test gives a P-value of 0.006. This is strong evidence against the null hypothesis that the two population means are the same. We therefore reject H_0 and conclude that there is a significant difference between the mean amounts of difference in oil deapsorbed on treated samples and in control samples.

```
Two-sample T for diff

Oil   N   Mean  StDev  SE Mean
 5   20  0.650  0.587     0.13
10   20  1.325  0.847     0.19


Difference =  ( 5) -  (10)
Estimate for difference:  -0.675
95\% CI for difference:  (-1.144, -0.206)
T-Test of difference = 0 (vs ): T-Value = -2.93  P-Value = 0.006  DF = 33
```

c. We first check the conditions for running an ANOVA model by graphing the residuals (see below). The residuals appear to be approximately normally distributed and have approximately the same variance.



We now run the ANOVA model with the resulting output given below.

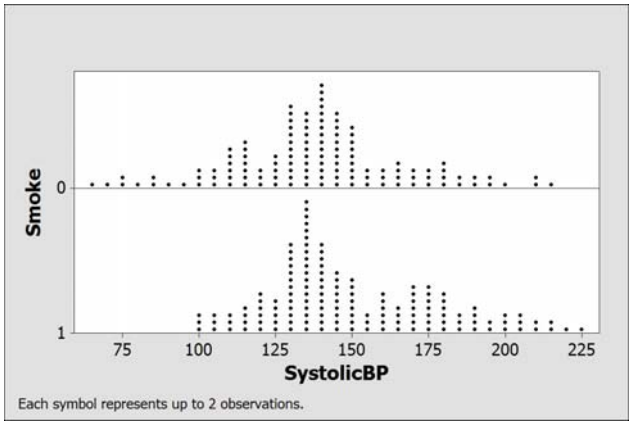
```
Analysis of Variance

Source  DF  Adj SS  Adj MS  F-Value  P-Value
Oil      1   4.556  4.5563     8.58    0.006
Error   38  20.188  0.5313
Total   39  24.744
```

The P-value in the ANOVA table is quite small: 0.006. We reject the null hypothesis and conclude that there is a difference in the means of the response variable based on how much oil was present in the sample.

d. The conclusions to parts (b) and (c) are the same. In essence, both the two-sample t-test and the ANOVA table are testing the same hypotheses because there were only two groups

in this experiment. Note that the two-sample t-test does not assume equal variances, but the sample standard deviations are similar enough that the P-values end up being the same for both tests.

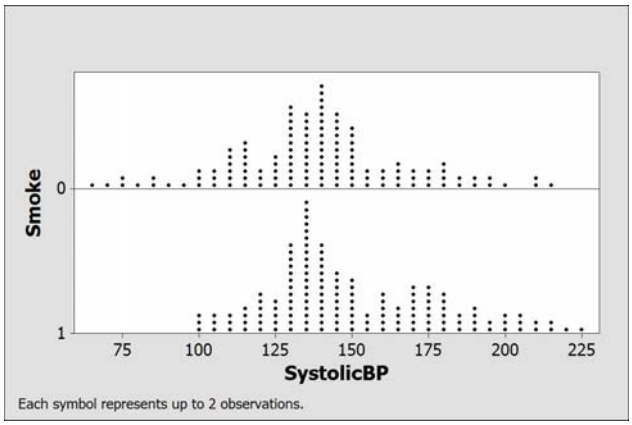
5.54 a. The dotplots suggest that there is at least a small difference between the systolic blood pressure of smokers and nonsmokers.



The means for the two groups are different enough, with respect to the standard deviations, that we suspect the difference will be statistically significant.

Variable	Smoke	N	N*	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
SystolicBP	0	234	0	139.18	27.50	67.00	125.00	138.00	150.25	215.00
	1	266	0	150.03	27.49	100.00	132.00	142.50	169.50	224.00

b. The observations are random and therefore likely independent. The dotplots in part (a) show that there is no strong skewness nor are there outliers and the standard deviations are very similar.

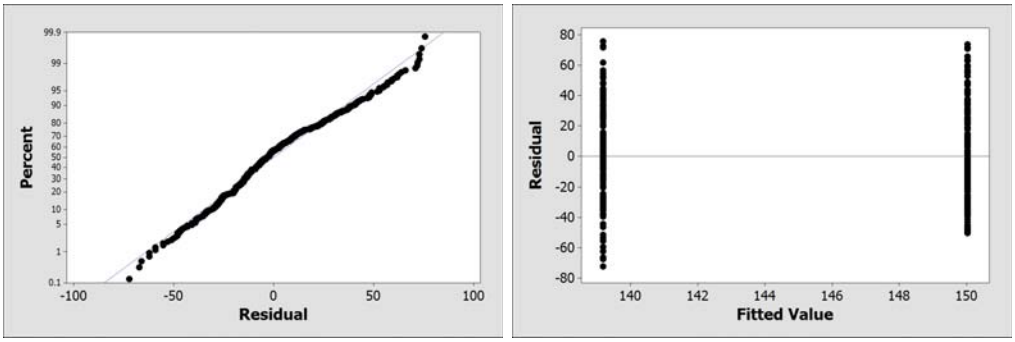


The t-test gives the following results:

```
     Two-sample T for SystolicBP

Smoke    N   Mean  StDev  SE Mean
0      234  139.2   27.5      1.8
1      266  150.0   27.5      1.7

Difference = mu (0) - mu (1)
Estimate for difference:  -10.84
95\% CI for difference:  (-15.68, -6.00)
T-Test of difference = 0 (vs not =): T-Value = -4.40  P-Value = 0.000  DF = 498
```

The output reports that the P-value is approximately 0. Therefore, we have significant evidence that there is a difference in the systolic blood pressure between smokers and nonsmokers.

c. The observations are random and therefore likely independent. The normal probability plot of the residuals is roughly linear, so the normality condition is met and the residuals versus fits plot indicates that the equal variances condition is met.



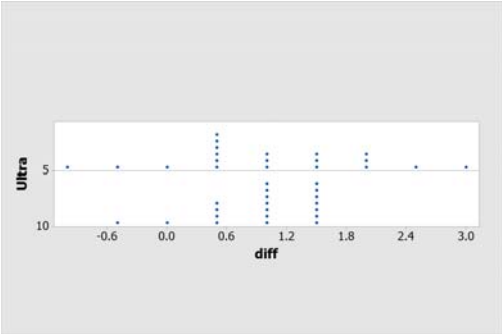
The ANOVA table for this analysis is given below.

```
One-way ANOVA: SystolicBP versus Smoke

Source   DF      SS     MS      F      P
Smoke     1   14635  14635  19.36  0.000
Error   498  376440    756
Total   499  391075
```

The null hypothesis is that smokers and nonsmokers have the same average systolic blood pressure. Since the P-value is approximately 0, we have significant evidence that the mean systolic blood pressure for smokers and nonsmokers is different.

d. The P-values and conclusions are the same. The F-statistic is 19.36, which is the same thing as $(-4.4)^2 = 19.36$, the t-statistic squared.

5.55 a. The dotplots and descriptive statistics are given below:



Variable	Ultra	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
diff	5	20	0	1.025	0.222	0.993	-1.000	0.500	1.000	1.875	3.000
	10	20	0	0.950	0.125	0.560	-0.500	0.500	1.000	1.500	1.500

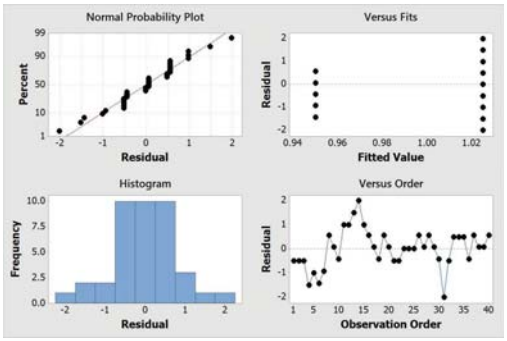
It does not appear that there is much difference in the response variable here based on how long the sand samples are exposed to the ultrasound. The means are very close (1.025 and 0.95) and the overlap in the dotplot is large.

b. Output for the t-test is given below.

```
Two-sample T for diff

Ultra   N   Mean  StDev  SE Mean
 5     20  1.025  0.993     0.22
10     20  0.950  0.560     0.13


Difference =  ( 5) -  (10)
Estimate for difference:  0.075
95\% CI for difference:  (-0.446, 0.596)
T-Test of difference = 0 (vs ): T-Value = 0.29  P-Value = 0.771  DF = 29
```

The P-value for this test is 0.771, which is quite large. We do not have enough evidence to reject the null hypothesis. That is, we do not have enough evidence to suggest that there is a difference between the difference in the amount of oil deapsorbed in treated samples and control samples, depending on the length of time that ultrasound was used.

c. We first check the conditions for running an ANOVA model by graphing the residuals (graphs follow). The residuals appear to be approximately normally distributed and have approximately the same variance.



We now run the ANOVA model with the resulting output given below.

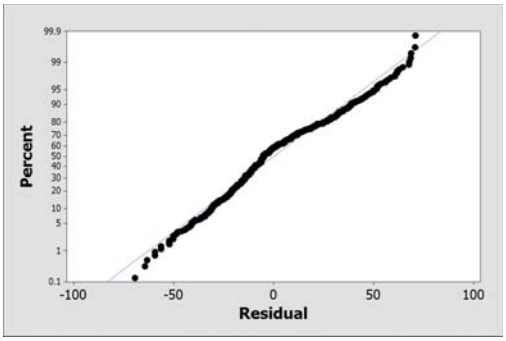
```
Analysis of Variance

Source  DF   Adj SS   Adj MS  F-Value  P-Value
Ultra    1   0.0563  0.05625     0.09    0.770
Error   38  24.6875  0.64967
Total   39  24.7437
```

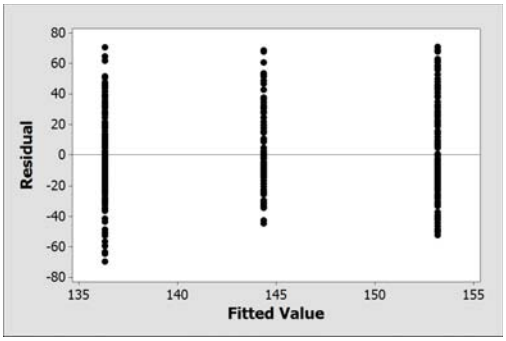
The P-value for this ANOVA model is 0.770. We do not have enough evidence to reject the null hypothesis. That is, we do not have enough evidence to suggest that there is a difference between the difference in the amount of oil deapsorbed in treated samples and control samples, depending on the length of time that ultrasound was used.

d. The conclusions to parts (b) and (c) are the same. In essence, both the two-sample t-test and the ANOVA table are testing the same hypotheses because there were only two groups in this experiment. Note that the two-sample t-test does not assume equal variances, but the sample standard deviations are similar enough that the P-values end up being very similar for both tests.

5.56 a. If we do three tests, we need to concern ourselves with the family-wise error rate as well as the individual error rate. If we perform three two-sample tests (necessary to compare three groups) and each has a 5% individual error rate, there is a higher risk that *at least one* of the three tests is wrong.

b. First, we check conditions. The dataset consisted of randomly selected individuals, so the observations should be independent of each other. The normal probability plot of the residuals and the residual plot are given below. They indicate that both the normality and the equal variances condition hold since the normal probability plot is roughly linear and the variability is approximately the same for all three fitted values.





The ANOVA table is given below.

```
One-way ANOVA: SystolicBP versus Overwt

Source   DF      SS     MS      F      P
Overwt    2   27801  13900  19.02  0.000
Error   497  363274    731
Total   499  391075
```

Since the P-value is approximately 0, we have significant evidence that the mean systolic blood pressure for people of at least one weight category is different from the mean systolic blood pressure for people of at least one other weight category.

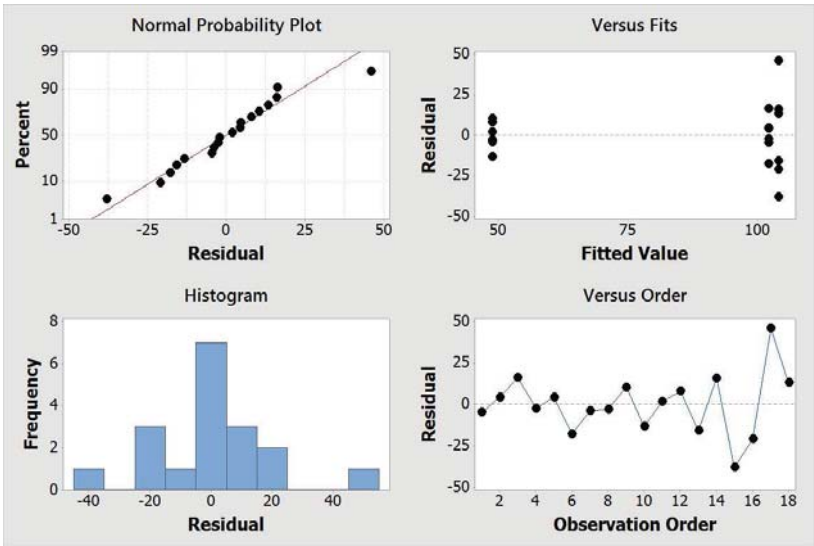
c. From the following results, we have significant evidence that all three groups have significantly different mean systolic blood pressures. In fact, the more overweight the group is, the higher the mean systolic blood pressure is.

```
Grouping Information Using Fisher Method

Overwt    N    Mean  Grouping
2       204  153.18  A
1       109  144.37    B
0       187  136.32      C

Means that do not share a letter are significantly different.


Fisher 95\% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Overwt

Simultaneous confidence level = 87.90%
```

```
Overwt = 0 subtracted from:

Overwt  Lower  Center  Upper  -----+---------+---------+---------+----
1        1.65    8.05  14.45                  (-----*-----)
2       11.49   16.87  22.24                            (-----*----)
                              -----+---------+---------+---------+----
                                 -10         0        10        20


Overwt = 1 subtracted from:

Overwt  Lower  Center  Upper  -----+---------+---------+---------+----
2        2.51    8.81  15.12                    (-----*-----)
                              -----+---------+---------+---------+----
                                 -10         0        10        20
```

5.57 a. The hypotheses are $H_0 : \alpha_a = \alpha_f = \alpha_v = 0$ vs. H_a : at least one $\alpha_k \neq 0$ where (a) stands for the meniscus arrow treatment group, (f) stands for the FasT-Fix treatment group, and (v) stands for the vertical suture treatment group, and the means are the mean values of the load at failure.

b. The graphs of the residuals follow. The residuals are reasonably normally distributed but we worry about the equality of the variances. The actual standard deviations range from 8.69 (meniscus arrow) to 30.6 (FasT-Fix). The second standard deviation is well over twice the size of the first standard deviation.

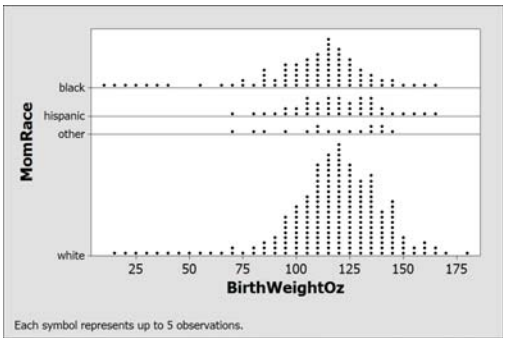


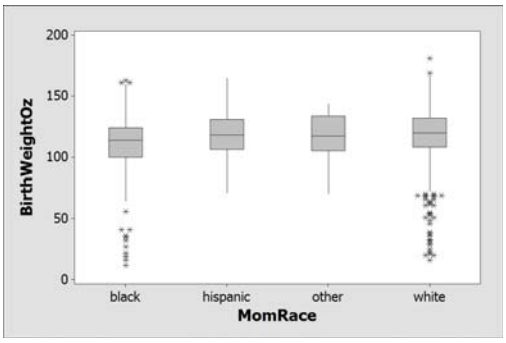
5.58 a. According to the ANOVA table, the P-value is approximately 0. The null hypothesis being tested with this table is that the mean salary of both sexes is the same. Since the P-value is so small, we have significant evidence that the mean salaries for the two sexes are different.

b. According to the computer output given, $R^2 = 0.0429$. This says that about 4.3% of the variation in salary is explained by grouping the salaries according to sex. This is very small. Clearly, there are many other variables that explain differences in salaries.

c. The conclusions from the ANOVA are clearly violated. It appears that neither the normality nor equal variances condition are met. What's more, there is at least one large outlier and ANOVA is not robust to lack of normality when outliers are present.

5.59 a. Dotplots and boxplots of the distributions are given below. All four distributions appear to have similar centers. The distributions for whites and blacks are skewed to the left with many low outliers. The distributions for Hispanics and others do not have outliers and have distributions that are reasonably symmetric.



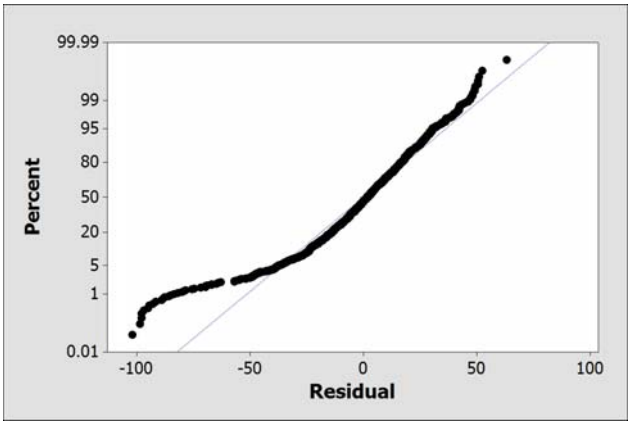


b. The statistics are given below.

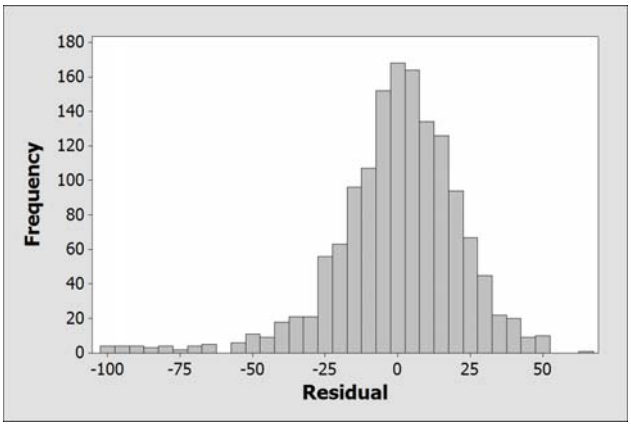
Variable	MomRace	N	Mean	StDev
BirthWeightOz	black	332	110.56	23.40
	hispanic	164	118.52	18.17
	other	48	117.15	17.60
	white	906	117.87	22.52

c. Looking at the plots from part (a), we see that there is a lot of variability in each group. And in fact, that variability makes it seem as though there is not much difference in the centers of the groups. There is far too much overlap between the observations in the groups to reasonably conclude that the centers are different. We see this from part (b) as well. There we see that the biggest difference in the means is between the Hispanic babies and the black babies. The difference is $118.52 - 110.56 = 7.96$ ounces. But this is very small compared to the standard deviations in the groups, which are all around 20 ounces.

5.60 a. The sample was random and the birth weights ought to be independent of each other. From the summary statistics in part (b) of the previous problem, the ratio of the standard deviations is less than 2. Following is the normal probability plot of the residuals.



The data obviously deviate from a normal distribution. However, looking at a histogram of the residuals, we see that the distribution is mound shaped and not horribly skewed. For such a large sample size, this is acceptable since the ANOVA procedure is robust to lack of normality.



b. The ANOVA table follows.

```
One-way ANOVA: BirthWeightOz versus MomRace

Source     DF      SS    MS     F      P
MomRace     3   14002  4667  9.53  0.000
Error    1446  708332   490
Total    1449  722334
```

The null hypothesis is that the mean birth weight for all children of mothers belonging to these four racial groups is the same. In this case, the P-value is approximately 0 so we have significant evidence that the mean birth weights are not the same for all four racial groups.

5.61 a. The computer output for Fisher's LSD using 95% confidence is given as follows.

```
Grouping Information Using Fisher Method

MomRace     N    Mean  Grouping
hispanic  164  118.52  A
white     906  117.87  A
other      48  117.15  A B
black     332  110.56    B

Means that do not share a letter are significantly different.

Fisher 95\% Individual Confidence Intervals
All Pairwise Comparisons among Levels of MomRace
Simultaneous confidence level = 79.73%

MomRace = black subtracted from:

MomRace   Lower  Center  Upper  -----+---------+---------+---------+----
hispanic   3.81    7.96  12.10                      (------*------)
other     -0.12    6.58  13.29                (----------*----------)
white      4.52    7.31  10.09                       (---*----)
                               -----+---------+---------+---------+----
                                 -6.0       0.0       6.0      12.0

MomRace = hispanic subtracted from:

MomRace  Lower  Center  Upper  -----+---------+---------+---------+----
other    -8.50   -1.37   5.75   (-----------*-----------)
white    -4.33   -0.65   3.04          (-----*-----)
                              -----+---------+---------+---------+----
                                -6.0       0.0       6.0      12.0

MomRace = other subtracted from:

MomRace  Lower  Center  Upper  -----+---------+---------+---------+----
white    -5.70    0.73   7.16       (----------*----------)
                              -----+---------+---------+---------+----
                                -6.0       0.0       6.0      12.0
```

We conclude that black mothers have babies with a significantly smaller mean birth weight than white or Hispanic mothers.

b. Answers may vary. We have concluded that babies from black mothers have an average birth weight that is about 7–8 ounces smaller than babies from white and Hispanic mothers. For the typical baby, a difference of 7–8 ounces is not going to make much difference. This size difference will only be important for very small babies. But our test is about the mean, not about very small babies. So in the end, this difference is probably not practically important.

5.62 We test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ where μ_i represents the mean log commute time at the respective cities. The following is some computer output for the ANOVA table.

```
             Df Sum Sq Mean Sq F value   Pr(>F)
City          3   27.8   9.283   17.36 3.98e-11
Residuals  1996 1067.4   0.535
```

The large F-statistic (17.36) and small P-value provide strong evidence to reject this null hypothesis and conclude that there is a difference in mean log commute times between at least two of these cities.

5.63 We test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ where μ_i represents the mean of the square root of commute times at the respective cities. Here is some computer output for the ANOVA table.

```
              Df Sum Sq Mean Sq F value   Pr(>F)
City           3    186   62.04   20.99 2.17e-13
Residuals   1996   5899    2.96
```

The large F-statistic (20.99) and small P-value provide strong evidence to reject this null hypothesis and conclude that there is a difference in means of the square roots of commute times between at least two of these cities.

5.64 Here is the ANOVA table from the earlier exercise comparing natural logs of commute times for the four cities. The tiny P-value indicates that significant differences exist in some of the means.

```
             Df Sum Sq Mean Sq F value   Pr(>F)
City          3   27.8   9.283   17.36 3.98e-11
Residuals  1996 1067.4   0.535
```

Since all of the sample sizes are the same ($n_i = 500$) we compute Fisher's LSD, using the $MSE = 0.535$ from the ANOVA table and a t^*-value of 1.961 for 95% confidence with 1996 df.

$$LSD = 1.961\sqrt{0.535}\sqrt{1/500 + 1/500} = 0.091$$

Here are the means of the square roots of the commute times for the four cities, ordered from smallest to largest.

```
Minneapolis   Boston    Houston    Washington
   2.941       3.031     3.042      3.262
```

The only pairs of cities with mean log commute times within $LSD = 0.091$ of each other are Minneapolis versus Boston (barely) and Boston versus Houston. Thus we find that the mean for Minneapolis is less than both Houston and Washington; and the mean for Washington is more than the other three cities.

5.65 Here is the ANOVA table from the earlier exercise comparing square roots of commute times for the four cities. The tiny P-value indicates that significant differences exist in some of the means.

```
               Df Sum Sq Mean Sq F value   Pr(>F)
City            3    186   62.04   20.99 2.17e-13
Residuals    1996   5899    2.96
```

Since all of the sample sizes are the same ($n_i = 500$) we compute Fisher's LSD, using the $MSE = 2.96$ from the ANOVA table and a $t*$-value of 1.961 for 95% confidence with 1996 df.

$$LSD = 1.961\sqrt{2.96}\sqrt{1/500 + 1/500} = 0.213$$

Here are the means of the square roots of the commute times for the four cities, ordered from smallest to largest.

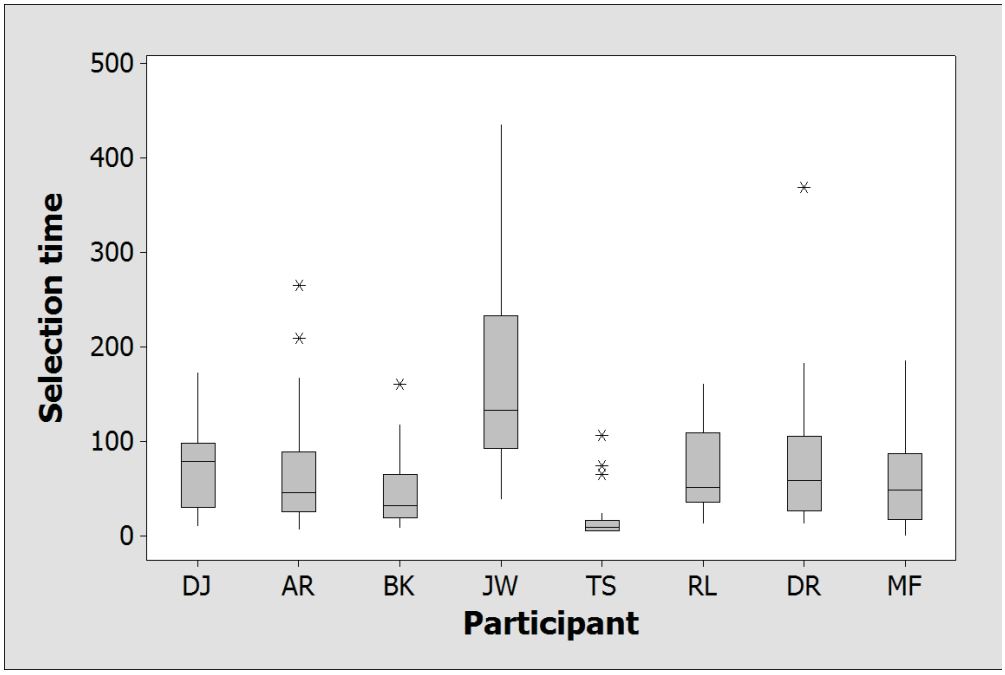
```
Minneapolis   Houston   Boston    Washington
   4.489       4.730     4.759      5.321
```

The only pair of cities with mean square root of commute times within $LSD = 0.213$ of each other are Houston and Boston. Thus we find that the mean for Minneapolis is less than the other three cities and the mean for Washington is more than the other three cities.

5.66 a. The boxplots show that most of the distributions are skewed to the right (obviously, some decisions are harder and take a little longer to make). Most participants take about the same amount of time to decide, but JW is quite a bit slower and TS is a bit faster than the rest of the participants. Descriptive statistics and boxplots are given as follows.

```
Descriptive Statistics: DJ, AR, BK, JW, TS, RL, DR, MF

          Total
Variable  Count   Mean  StDev  Minimum     Q1  Median      Q3  Maximum
DJ           24  69.63  41.62    11.00  30.75   79.50   99.00   174.00
AR           24   68.3   66.9      7.0   26.0    46.5    89.8    266.0
BK           24  47.96  39.25     9.00  20.25   33.00   65.75   161.00
```

JW		24	163.9	104.2	39.0	93.5	134.0	233.8	436.0
TS		24	19.33	25.83	5.00	6.00	9.50	17.00	107.00
RL		24	67.13	44.55	13.00	36.25	51.50	109.75	162.00
DR		24	80.1	75.8	13.0	27.3	59.5	106.5	369.0
MF		24	63.8	56.0	1.0	18.3	49.0	87.5	187.0



b. The ANOVA table is given below. The test statistic is 10.89, and the P-value is approximately 0. This leads to the conclusion that at least one of the participants has a different mean selection time than the other participants.

One-way ANOVA: DJ, AR, BK, JW, TS, RL, DR, MF

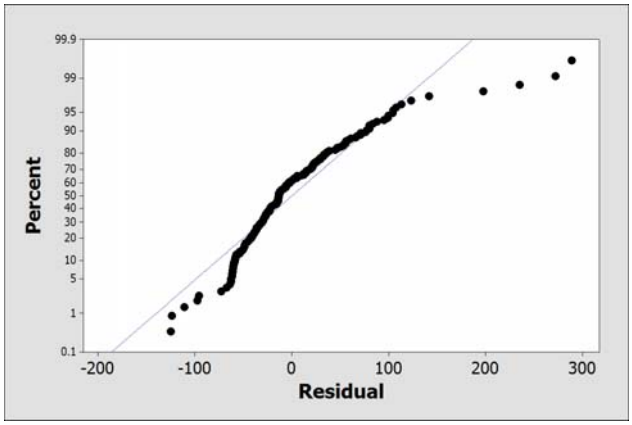
Source	DF	SS	MS	F	P
Factor	7	287196	41028	10.89	0.000
Error	184	693126	3767		
Total	191	980322			

c. The computer output for Fisher's LSD follows. What we learn is that JW takes significantly longer to make his/her selections. TS takes a significantly shorter amount of time than everyone except BK.

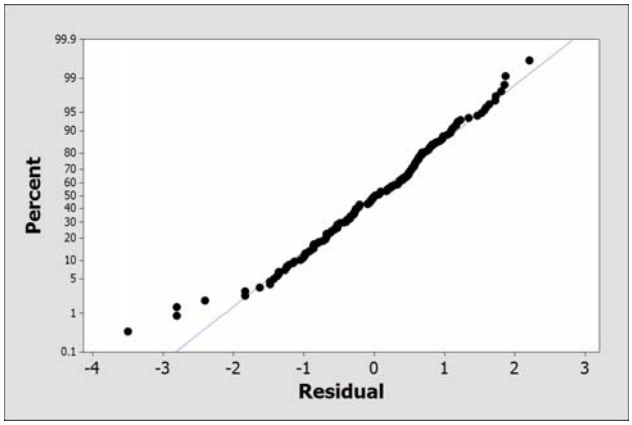
```
Grouping Information Using Fisher Method

Participant   N    Mean  Grouping
JW           24  163.88  A
DR           24   80.13    B
DJ           24   69.63    B
AR           24   68.29    B
RL           24   67.13    B
MF           24   63.83    B
BK           24   47.96    B C
TS           24   19.33      C

Means that do not share a letter are significantly different.

Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Participant
Simultaneous confidence level = 49.70%

Participant = AR subtracted from:

Participant   Lower  Center   Upper  --------+---------+---------+---------+-
BK           -55.29  -20.33   14.62              (---*--)
DJ           -33.62    1.33   36.29                 (--*---)
DR           -23.12   11.83   46.79                  (--*---)
JW            60.63   95.58  130.54                          (---*--)
MF           -39.41   -4.46   30.50                (---*--)
RL           -36.12   -1.17   33.79                (---*--)
TS           -83.91  -48.96  -14.00           (--*---)
                                    --------+---------+---------+---------+-
                                         -100         0       100       200

Participant = BK subtracted from:

Participant   Lower  Center   Upper  --------+---------+---------+---------+-
DJ           -13.29   21.67   56.62                   (--*---)
DR            -2.79   32.17   67.12                    (--*---)
JW            80.96  115.92  150.87                            (---*--)
MF           -19.08   15.88   50.83                 (---*--)
RL           -15.79   19.17   54.12                 (---*--)
TS           -63.58  -28.63    6.33              (--*---)
                                    --------+---------+---------+---------+-
                                         -100         0       100       200
```

```
Participant = DJ subtracted from:

Participant   Lower  Center   Upper  --------+---------+---------+---------+-
DR           -24.46   10.50   45.46                 (--*---)
JW            59.29   94.25  129.21                         (--*---)
MF           -40.75   -5.79   29.16              (--*---)
RL           -37.46   -2.50   32.46              (---*--)
TS           -85.25  -50.29  -15.34          (---*--)
                                    --------+---------+---------+---------+-
                                        -100         0       100       200
Participant = DR subtracted from:

Participant   Lower  Center   Upper  --------+---------+---------+---------+-
JW            48.79   83.75  118.71                       (--*---)
MF           -51.25  -16.29   18.66             (--*---)
RL           -47.96  -13.00   21.96             (---*--)
TS           -95.75  -60.79  -25.84         (---*--)
                                    --------+---------+---------+---------+-
                                        -100         0       100       200
Participant = JW subtracted from:

Participant    Lower   Center    Upper --------+---------+---------+---------+-
MF           -135.00  -100.04   -65.09       (--*--)
RL           -131.71   -96.75   -61.79       (--*---)
TS           -179.50  -144.54  -109.59  (---*--)
                                   --------+---------+---------+---------+-
                                       -100         0       100       200

Participant = MF subtracted from:

Participant   Lower  Center  Upper  --------+---------+---------+---------+-
RL           -31.66    3.29  38.25               (--*---)
TS           -79.46  -44.50  -9.54          (---*--)
                                   --------+---------+---------+---------+-
                                       -100         0       100       200
Participant = RL subtracted from:

Participant   Lower  Center   Upper  --------+---------+---------+---------+-
TS           -82.75  -47.79  -12.84          (--*---)
                                    --------+---------+---------+---------+-
                                        -100         0       100       200
```

5.67 a. The normal probability plot follows. Clearly, the residuals do not come from a population with a normal distribution, as can be seen by the unusually large values in the upper tail; therefore, the ANOVA model is not appropriate for this data.



b. The normal probability plot looks considerably better using the natural log of the selection times. Even though there are still some problems in the lower tail, we will proceed with ANOVA.



So we continue with the ANOVA model. The ANOVA table is given below.

```
One-way ANOVA: lnSelectionTime versus Participant

Source        DF       SS      MS      F      P
Participant    7   78.750  11.250  12.99  0.000
Error        184  159.371   0.866
Total        191  238.121
```

The test statistic is 12.99, and the P-value is approximately 0. We conclude that there is a significant difference in the natural log of the selection times of at least one participant.

5.68 We use Fisher's LSD to discover what differences exist. The following output below shows more differences than when we used the raw selection times. JW still has a significantly longer time than all others. Now TS has a significantly shorter time than all others. And DR and MF have significantly different times than each other.

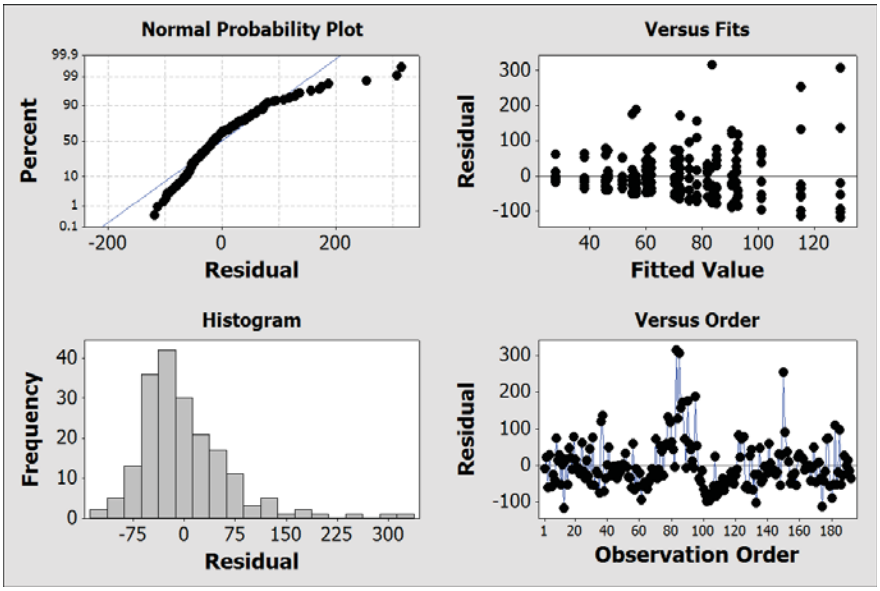
```
Grouping Information Using Fisher Method
Participant   N    Mean  Grouping
JW           24  4.9030  A
DR           24  4.0470    B
DJ           24  4.0225    B C
RL           24  3.9875    B C
AR           24  3.7798    B C
BK           24  3.5544    B C
MF           24  3.5026      C
TS           24  2.4689        D

Means that do not share a letter are significantly different.
Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Participant
Simultaneous confidence level = 49.70%

Participant = AR subtracted from:
Participant    Lower   Center    Upper
BK           -0.7554  -0.2254   0.3047
DJ           -0.2873   0.2427   0.7728
DR           -0.2629   0.2672   0.7973
JW            0.5932   1.1233   1.6533
MF           -0.8073  -0.2772   0.2529
RL           -0.3223   0.2078   0.7378
TS           -1.8410  -1.3109  -0.7809
Participant     +---------+---------+---------+---------
BK                           (--*---)
DJ                              (---*--)
DR                              (---*--)
JW                                    (--*---)
MF                           (--*---)
RL                              (--*---)
TS                    (--*---)
              +---------+---------+---------+---------
             -3.0      -1.5       0.0       1.5
```

```
Participant = BK subtracted from:

Participant    Lower   Center    Upper
DJ           -0.0619   0.4681   0.9982
DR           -0.0375   0.4926   1.0226
JW            0.8186   1.3486   1.8787
MF           -0.5819  -0.0518   0.4782
RL           -0.0969   0.4332   0.9632
TS           -1.6156  -1.0855  -0.5555

Participant     +---------+---------+---------+---------
DJ                                 (--*---)
DR                                 (--*---)
JW                                      (---*---)
MF                               (---*--)
RL                                  (---*--)
TS                         (---*--)
                +---------+---------+---------+---------
              -3.0       -1.5       0.0       1.5


Participant = DJ subtracted from:

Participant    Lower   Center    Upper
DR           -0.5056   0.0245   0.5545
JW            0.3505   0.8805   1.4106
MF           -1.0500  -0.5200   0.0101
RL           -0.5650  -0.0350   0.4951
TS           -2.0837  -1.5537  -1.0236

Participant     +---------+---------+---------+---------
DR                                  (--*---)
JW                                       (---*--)
MF                             (---*--)
RL                                (---*--)
TS                       (---*--)
                +---------+---------+---------+---------
              -3.0       -1.5       0.0       1.5
```

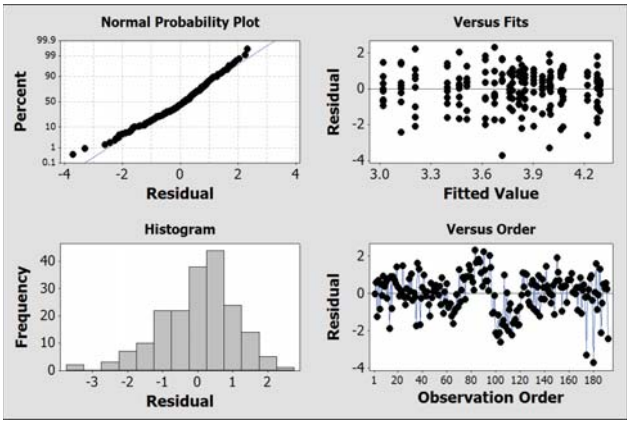
```
Participant = DR subtracted from:

Participant    Lower   Center    Upper
JW            0.3260   0.8561   1.3861
MF           -1.0745  -0.5444  -0.0143
RL           -0.5895  -0.0594   0.4706
TS           -2.1082  -1.5781  -1.0481

Participant     +---------+---------+---------+---------
JW                                    (---*--)
MF                             (--*---)
RL                                (---*--)
TS                      (--*---)
                +---------+---------+---------+---------
              -3.0       -1.5       0.0       1.5


Participant = JW subtracted from:

Participant    Lower   Center    Upper
MF           -1.9305  -1.4005  -0.8704
RL           -1.4455  -0.9155  -0.3854
TS           -2.9642  -2.4342  -1.9041

Participant     +---------+---------+---------+---------
MF                     (---*--)
RL                        (---*--)
TS              (---*--)
                +---------+---------+---------+---------
              -3.0       -1.5       0.0       1.5

Participant = MF subtracted from:

Participant    Lower   Center    Upper
RL           -0.0451   0.4850   1.0150
TS           -1.5638  -1.0337  -0.5037

Participant     +---------+---------+---------+---------
RL                               (--*---)
TS                         (--*---)
                +---------+---------+---------+---------
              -3.0       -1.5       0.0       1.5
```

```
Participant = RL subtracted from:

Participant    Lower   Center    Upper
TS           -2.0487  -1.5187  -0.9886

Participant     +---------+---------+---------+---------
TS                    (---*--)
                +---------+---------+---------+---------
              -3.0       -1.5       0.0       1.5
```

5.69 We begin by checking conditions. The curvature and large values in the upper tail, seen in the normal plot, show that the data do not meet the normal condition. We see this also in the histogram of the residuals, though the residuals do form a mound shape, they are quite right-skewed. The fan or funnel shape in the residual plot suggests increasing variability, so the equal variance condition is also not met.



We next try taking the natural log of the selection times. Now the graphs of the residuals show that all conditions are reasonably met.

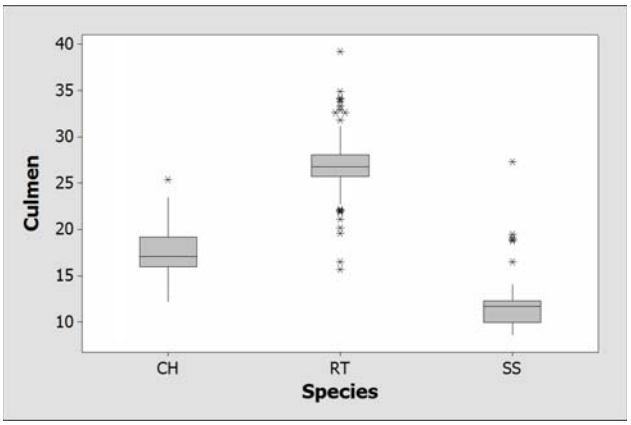


Now we continue with the ANOVA table.

```
One-way ANOVA: lnSelectionTime versus Round

Source   DF      SS    MS     F      P
Round    23   23.82  1.04  0.81  0.713
Error   168  214.30  1.28
Total   191  238.12
```

Here, we see that the test statistic is $F = 0.81$ and the P-value is 0.713. We do not have enough evidence to reject the null hypothesis that the mean selection times are the same for all rounds.

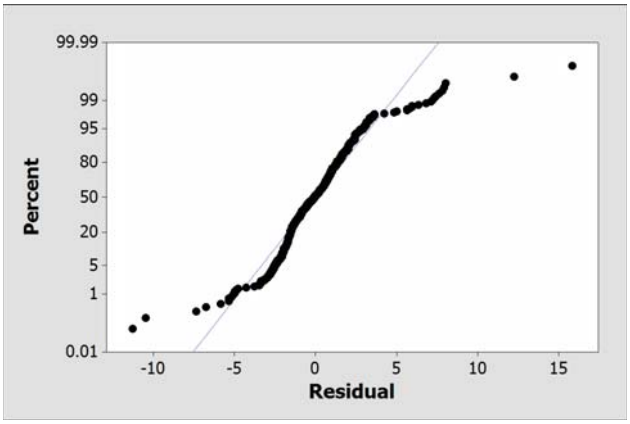
5.70 We begin by examining the data itself. Does it appear that there is a difference in the mean length of the culmen for the three different species? The following boxplot suggests that there is a difference. It looks like the red-tailed hawks have the longest mean culmen. It is less clear whether we will find significant evidence of a difference between the Cooper's hawks and the sharp-shinned hawks.



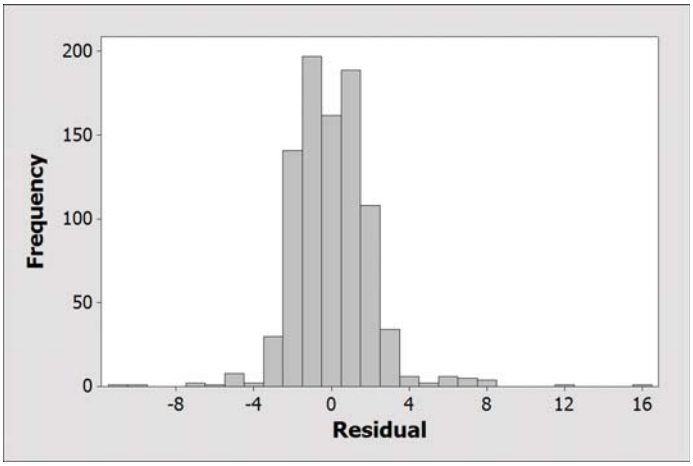
We next check the conditions necessary to use an ANOVA model. We have no reason to believe that the culmen lengths of those birds observed are not independent of each other. The standard deviations of the three groups are given below.

Variable	Species	N	StDev
Culmen	CH	70	2.393
	RT	573	2.050
	SS	258	1.906

Clearly, the ratio of the largest standard deviation to the smallest standard deviation is less than 2, so the equal variances condition is met. Finally, we produce a normal probability plot of the residuals.



The graph suggests that the normal distribution is not a good fit for the residuals. But ANOVA is robust to non-normality as long as there are no large outliers. The following histogram suggests that the distribution is mound-shaped and does not have large outliers.



Since the conditions have been met, we compute the ANOVA table. It is given below.

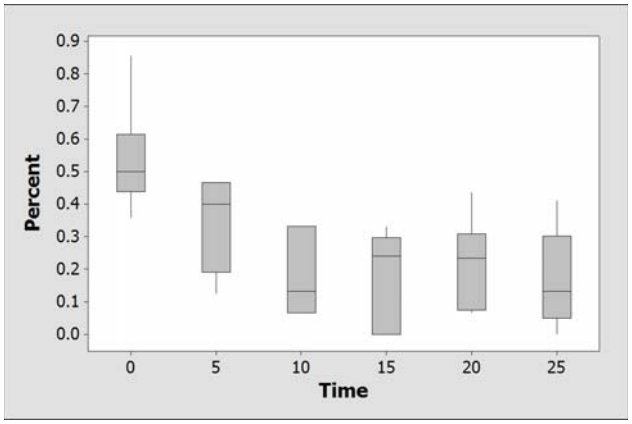
```
One-way ANOVA: Culmen versus Species

Source    DF        SS        MS        F      P
Species    2  44020.11  22010.05  5294.43  0.000
Error    898   3733.18      4.16
Total    900  47753.29
```

Since the P-value is approximately 0, we have strong evidence that there is a difference between the mean culmen length of at least two of the hawk species. To figure out exactly which differences exist, we proceed with Fisher's LSD.

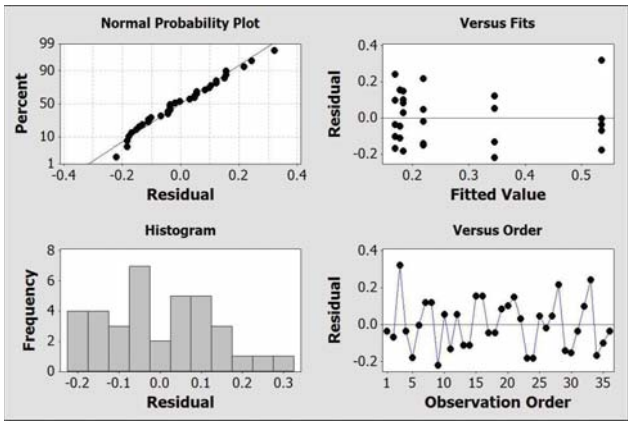
```
Grouping Information Using Fisher Method

Species    N    Mean  Grouping
RT       573  26.966  A
CH        70  17.566    B
SS       258  11.481      C

Means that do not share a letter are significantly different.

Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Species

Simultaneous confidence level = 87.84%

Species = CH subtracted from:

Species   Lower  Center   Upper      +---------+---------+---------+---------
RT        8.894   9.400   9.907                                    (*
SS       -6.624  -6.085  -5.546                  *)
                                    +---------+---------+---------+---------
                                 -16.0      -8.0       0.0       8.0

Species = RT subtracted from:

Species    Lower   Center    Upper
SS       -15.785  -15.485  -15.185

Species      +---------+---------+---------+---------
SS           (*
             +---------+---------+---------+---------
          -16.0      -8.0       0.0       8.0
```

In fact, we find that all three species has its own unique mean culmen length.

5.71 We begin by examining the data itself. Does it appear that there is a difference in the mean percent that metamorphosed over the six different time periods? The boxplot that follows suggests that there is a difference. It looks like the largest percent metamorphosed at time 0, followed by time 5. The other four time periods (10, 15, 20, and 25) look like they have similar percents of larvae that metamorphosed.



Next, we check conditions. Since this was a randomized experiment, the observations should be independent of each other. The residual plots given below indicate that both the normal and equal variance conditions hold. Specifically, the normal probability plot shows only a very slight departure from a linear trend in the lower tail, and the residual versus fits plot has an unstructured band of points.



So we continue to the ANOVA table.

```
One-way ANOVA: Percent versus Time

Source  DF      SS      MS     F      P
Time     5  0.6309  0.1262  5.96  0.001
Error   30  0.6346  0.0212
Total   35  1.2655
```

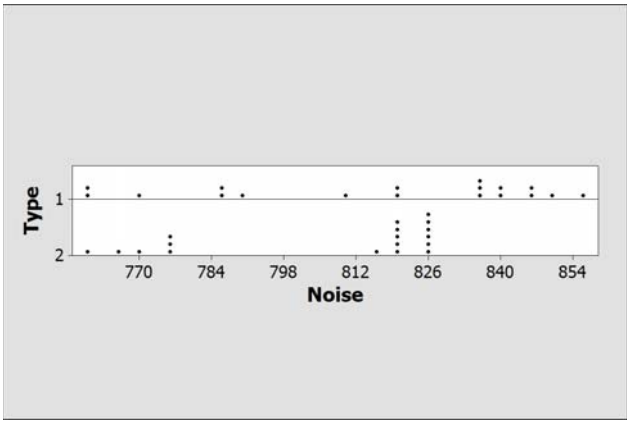
Since the P-value is given as 0.001, we have significant evidence to suggest that the mean percent of larvae that metamorphosed in water was different for at least one time period. We continue our analysis with Fisher's LSD to determine exactly which differences exist.

```
Time  N    Mean  Grouping
 0    6  0.5357  A
 5    6  0.3455    B
20    6  0.2192    B C
15    6  0.1833    B C
10    6  0.1777    B C
25    6  0.1687      C

Means that do not share a letter are significantly different.


Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Time

Simultaneous confidence level = 65.64%


Time =  0 subtracted from:

Time    Lower   Center    Upper   --+---------+---------+---------+-------
 5    -0.3617  -0.1902  -0.0187           (-----*------)
10    -0.5295  -0.3580  -0.1865    (------*------)
15    -0.5238  -0.3523  -0.1808    (------*------)
20    -0.4880  -0.3165  -0.1450     (------*------)
25    -0.5385  -0.3670  -0.1955   (------*------)
                                  --+---------+---------+---------+-------
                                -0.50     -0.25      0.00      0.25


Time =  5 subtracted from:

Time    Lower   Center    Upper   --+---------+---------+---------+-------
10    -0.3393  -0.1678   0.0037           (------*------)
15    -0.3337  -0.1622   0.0093            (------*-----)
20    -0.2978  -0.1263   0.0452              (------*------)
25    -0.3483  -0.1768  -0.0053           (------*------)
                                  --+---------+---------+---------+-------
                                -0.50     -0.25      0.00      0.25
```

```
Time = 10 subtracted from:

Time    Lower   Center   Upper   --+---------+---------+---------+-------
15    -0.1658   0.0057  0.1772                (------*------)
20    -0.1300   0.0415  0.2130                  (------*------)
25    -0.1805  -0.0090  0.1625                (------*-----)
                                 --+---------+---------+---------+-------
                                -0.50     -0.25      0.00      0.25


Time = 15 subtracted from:

Time    Lower   Center   Upper   --+---------+---------+---------+-------
20    -0.1357   0.0358  0.2073                  (-----*------)
25    -0.1862  -0.0147  0.1568                (-----*------)
                                 --+---------+---------+---------+-------
                                -0.50     -0.25      0.00      0.25


Time = 20 subtracted from:

Time    Lower   Center   Upper   --+---------+---------+---------+-------
25    -0.2220  -0.0505  0.1210             (------*------)
                                 --+---------+---------+---------+-------
                                -0.50     -0.25      0.00      0.25
```

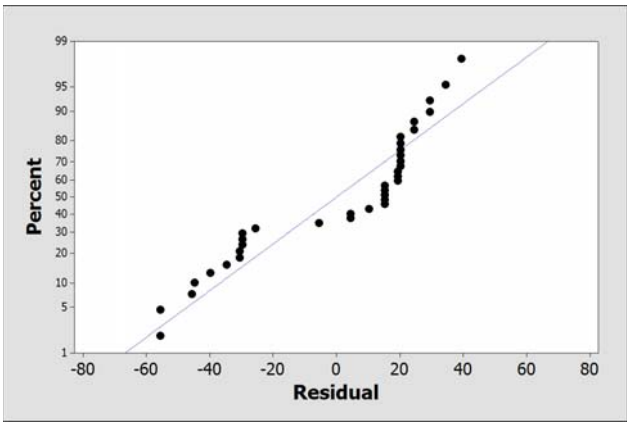
As we suspected, time 0 had a significantly higher mean percent of larvae that metamorphosed. It is significantly higher than all other time periods. Time 5 is also significantly higher than time 25. We note that, while we have shown that there is a difference in the mean percent of larvae metamorphosing at the different time periods, we have not proven the scientific reason why that is happening. Scientists, apparently, did not measure the level of chemicals in this water, they just supposed that the chemicals would disperse as the seawater came in.

5.72 We begin by looking at a dotplot of the data.



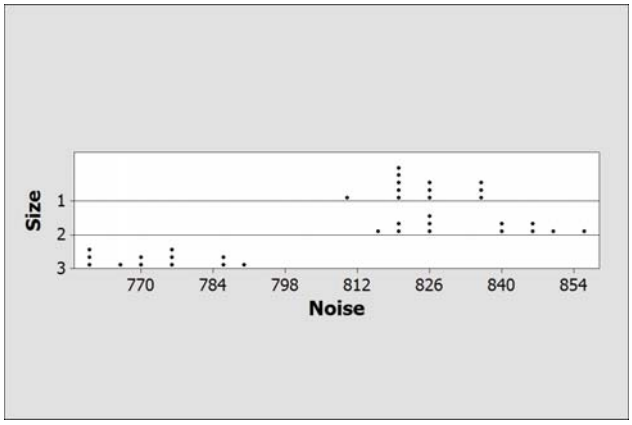
This graph suggests that there might be a difference in the mean noise level between the two filters, but there is a lot of overlap between the distributions of the noise level for the two filters. It is not obvious whether any difference will be significant or not. If there is a significant difference, it appears that the new filter is quieter.

Next, we check the conditions for using an ANOVA model. The data come from a randomized experiment so the noise levels should be independent of each other. The standard deviations of the two groups are 32.22 for group 1 and 25.64 for group 2. Their ratio is quite a bit smaller than 2, so we are comfortable with the equal variance condition. Finally, we produce a normal probability plot of the residuals.

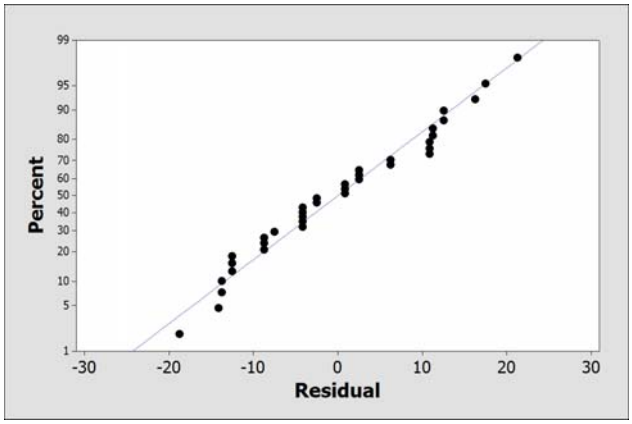


This graph does not appear to support normality of the residuals. The use of the ANOVA model is not appropriate in this case. Also, the standard transformations (natural log, square root, square) do not improve the normality.

5.73 a. A dotplot of the noise level for the three different sizes indicates that the large cars have much less noise, but that the small and medium sized cars have about the same level of noise.



Next, we check conditions. The cars were presumably chosen at random so the noise levels should be independent. The smallest standard deviation was 7.64 and the largest was 13.51, so we meet the equal variance condition. The normal probability plot follows, and it indicates that a normal distribution is plausible for the residuals. Conditions are met and we can continue with an ANOVA model.



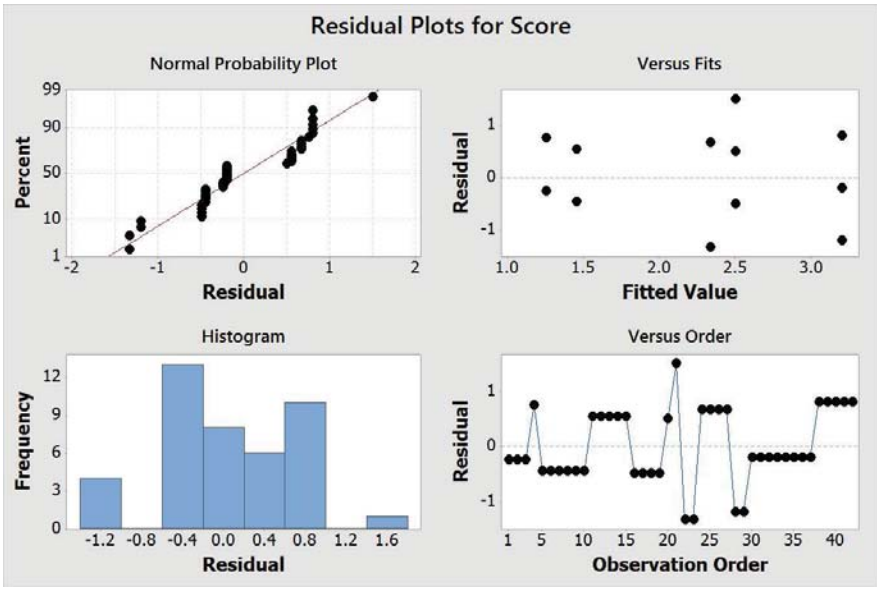

```
One-way ANOVA: Noise versus Size

Source  DF     SS     MS       F      P
Size     2  26051  13026  112.44  0.000
Error   33   3823    116
Total   35  29874
```

Since the P-value is approximately 0, we have significant evidence that the mean noise level for at least one size of car is different. Using Fisher's LSD, we discover that all three sizes of cars have significantly different mean noise levels. The medium-sized cars are the loudest, followed by the small cars and then the large cars.

```
Grouping Information Using Fisher Method

Size   N    Mean  Grouping
2     12  833.75  A
1     12  824.17    B
3     12  772.50      C

Means that do not share a letter are significantly different.


Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Size

Simultaneous confidence level = 88.02%


Size = 1 subtracted from:

Size   Lower  Center   Upper  --------+---------+---------+---------+-
2       0.64    9.58   18.52                   (-*--)
3     -60.61  -51.67  -42.73     (-*-)
                              --------+---------+---------+---------+-
                                    -40         0        40        80


Size = 2 subtracted from:

Size   Lower  Center   Upper  --------+---------+---------+---------+-
3     -70.19  -61.25  -52.31  (--*-)
                              --------+---------+---------+---------+-
                                    -40         0        40        80
```

b. Even if we had been able to use an ANOVA model in the previous exercise, the conclusions would have necessarily been different. When comparing the noise levels between the two filters, we were using data from a randomized experiment. This means that, if we had been able to find a significant difference, we would have been able to make a cause-and-effect conclusion. In the analysis of noise level by size, we did not have data from a randomized experiment, but rather a random sample. This means that, while we can generalize our findings to the larger populations, we cannot conclude that the differences we saw were due to cause-and-effect.

5.74 a. Because the session notes were presented to the raters in a random order, we have met the condition of independence for this dataset. To check for equal variances and normality, we rely on the following residual plots.

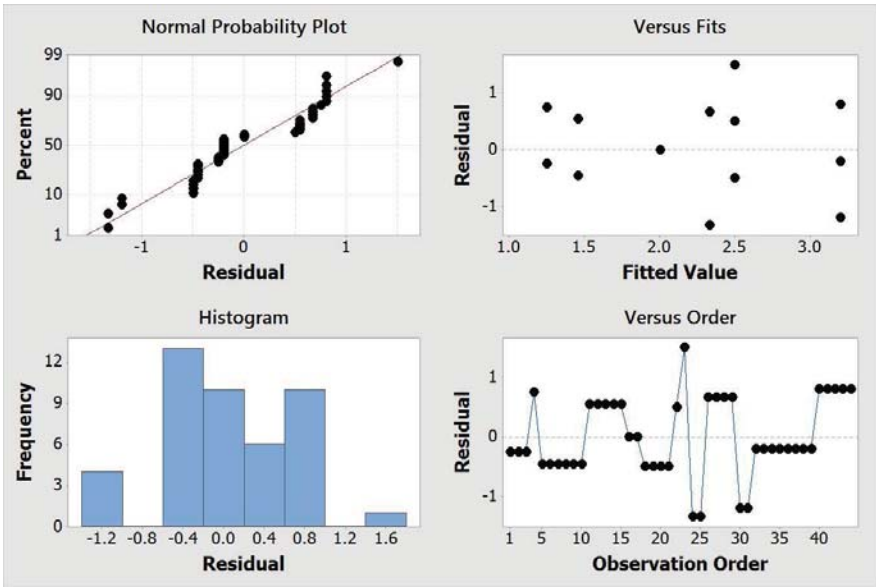


These plots show us that both conditions are reasonably met. Next we compute the ANOVA table.

```
Analysis of Variance
Source  DF  Adj SS  Adj MS  F-Value P-Value
Group   4   24.62   6.1557  12.17   0.000
Error   37  18.71   0.5057
Total   41  43.33
```

Here we see that the F-value is large and the P-value is very small (≈ 0) leading us to reject the null hypothesis. In other words we have reason to believe that the true mean scores are different for the different time periods. But how much of a difference is there? The researchers suspected that there would be improvement over time. So how much of a difference is there comparing time VI to time I? To answer that question we compute the effect size. The average score for time period VI is 3.2 and the average score for time period I is 1.25, so the effect size is $\frac{3.2-1.25}{0.711} = 2.74$. This translates into a 274% increase in score, which is huge. It appears that the therapy is working.

b. Once again we check for normality and equal variance. They rely on the following residual plots.



The normality still is not a concern, but some might question the equal variance condition. We do not worry about this, because as noted earlier, there are only 5 possible response values and so, when we only have two responses, it is not unlikely that they would be the same. The variability in the other 5 groups is similar enough (and small enough) that we feel comfortable continuing. The ANOVA table is:

```
Analysis of Variance
Source  DF  Adj SS  Adj MS  F-Value P-Value
Group   5   24.83   4.9670  10.09   0.000
Error   38  18.71   0.4924
Total   43  43.55
```

Once again we see that the F-value is large and the P-value is very small (≈ 0) leading us to reject the null hypothesis. The effect size in comparing time period VI to time period I is 2.78, remarkably similar to the analysis in part (a).