Chapter 8 Solutions

8.1 False. The null hypothesis is that the population variances are the same. If the P-value is small, we will reject this null hypothesis and conclude that at least one of the variances is different from the others.

8.2 The condition that needs to be satisfied for the two-way ANOVA is that the treatment group variances are the same. The treatment groups are combinations of levels of the two factors. In other words, the student should have only run one Levene's test using all treatment groups instead of two different tests, one for the data grouped by the levels of Factor A and the other for the data grouped by the levels of Factor B.

8.3 Computer output gives the following result.

```
Levene's Test (Any Continuous Distribution)
Test statistic = 1.12, $P$-value = 0.341
```

Since the P-value is reasonably large, we do not reject the null hypothesis. The data are consistent with the populations having the same standard deviation.

8.4 Computer output gives the following result.

```
Levene's Test (Any Continuous Distribution)
Test statistic = 0.73, p-value = 0.393
```

Since the P-value is reasonably large, we do not reject the null hypothesis. The data are consistent with the populations having the same standard deviation.

8.5 Computer output gives the following result.

```
Levene's Test (Any Continuous Distribution)
Test statistic = 1.22, p-value = 0.297
```

Since the P-value is reasonably large, we do not reject the null hypothesis. The data are consistent with the populations having the same standard deviation.

8.6 Here is some computer output for doing Levene's test for both factors.

```
Levene's Test (Any Continuous Distribution)
Test statistic = 27.44, p-value = 0.000
```

The P-value is small enough that we reject the null hypothesis. We have significant evidence that at least one of the four populations (males and females in each of the two provinces) has a different standard deviation.

8.7 Computer output gives the following result:

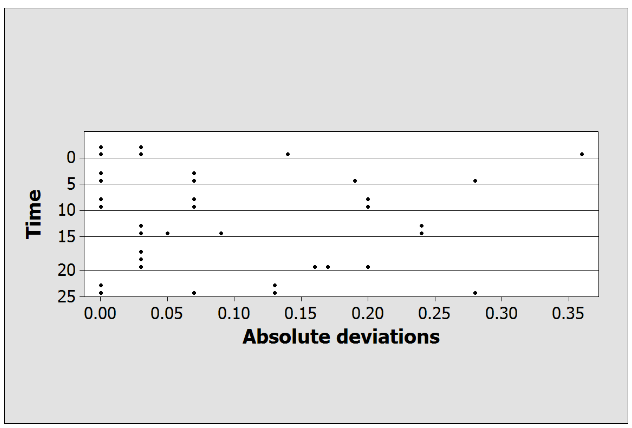
Method	Statistic	P-Value
Levene	7.17	0.007

Since the P-value is quite small, we reject the null hypothesis. The data are not consistent with the populations having the same standard deviation.

8.8 a. The medians for the six time periods are given as follows:

Variable	Time	N	Median
Percent	0	6	0.5000
	5	6	0.4000
	10	6	0.1330
	15	6	0.2405
	20	6	0.2335
	25	6	0.1330

The dotplot is given below.



There does not appear to be much difference in the absolute deviations in the 6 different time periods. They all have about the same spread and center.

b. The ANOVA table is given below.

One-way ANOVA: deviations versus Time

Source	DF	SS	MS	F	P
Time	5	0.0020	0.0004	0.04	0.999
Error	30	0.3378	0.0113		
Total	35	0.3397			

The F-statistic is 0.04 and the P-value is 0.999.

c. Levene's test gives the following output.

```
Levene's Test (Any Continuous Distribution)
Test statistic = 0.04, p-value = 0.999
```

Note that the test statistic and P-value are identical to what we computed in part (b). We fail to reject the null hypothesis. The data are consistent with the 6 time periods having the same standard deviation.

8.9 The multiplier of the standard error. This is the part of the confidence interval that governs the confidence level.

8.10 The method produces intervals that are quite wide in comparison to the other methods, because the Bonferroni multiplier is larger than other methods. This means that the Bonferroni intervals are not as precise as the other intervals.

8.11 a. The confidence intervals are given in the output as follows.

```
Grouping Information Using Tukey Method
Participant   N    Mean  Grouping
JW           24  4.9030  A
DR           24  4.0470    B
DJ           24  4.0225    B
RL           24  3.9875    B
AR           24  3.7798    B
BK           24  3.5544    B
MF           24  3.5026    B
TS           24  2.4689      C
Means that do not share a letter are significantly different.

Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of Participant
Individual confidence level = 99.75%

Participant = AR subtracted from:
Participant    Lower   Center    Upper
BK           -1.0499  -0.2254   0.5991
DJ           -0.5817   0.2427   1.0672
DR           -0.5573   0.2672   1.0917
JW            0.2988   1.1233   1.9477
MF           -1.1017  -0.2772   0.5473
RL           -0.6167   0.2078   1.0323
TS           -2.1354  -1.3109  -0.4864
```

```
Participant = BK subtracted from:
Participant    Lower   Center    Upper
DJ           -0.3564   0.4681   1.2926
DR           -0.3319   0.4926   1.3171
JW            0.5242   1.3486   2.1731
MF           -0.8763  -0.0518   0.7727
RL           -0.3913   0.4332   1.2576
TS           -1.9100  -1.0855  -0.2610

Participant = DJ subtracted from:
Participant    Lower   Center    Upper
DR           -0.8000   0.0245   0.8489
JW            0.0560   0.8805   1.7050
MF           -1.3444  -0.5200   0.3045
RL           -0.8594  -0.0350   0.7895
TS           -2.3781  -1.5537  -0.7292

Participant = DR subtracted from:
Participant    Lower   Center    Upper
JW            0.0316   0.8561   1.6805
MF           -1.3689  -0.5444   0.2801
RL           -0.8839  -0.0594   0.7651
TS           -2.4026  -1.5781  -0.7536

Participant = JW subtracted from:
Participant    Lower   Center    Upper
MF           -2.2249  -1.4005  -0.5760
RL           -1.7400  -0.9155  -0.0910
TS           -3.2587  -2.4342  -1.6097

Participant = MF subtracted from:
Participant    Lower   Center    Upper
RL           -0.3395   0.4850   1.3095
TS           -1.8582  -1.0337  -0.2092

Participant = RL subtracted from:
Participant    Lower   Center    Upper
TS           -2.3432  -1.5187  -0.6942
```

The results are that JW takes significantly longer than all of the others and TS takes a significantly shorter time than all others.

b. Answers may vary. One possible answer is Fisher's LSD if you are only exploring to see where any differences might occur. Another possible answer is Tukey if you want to control the experimentwise error. You might even say Bonferroni if the consequences of mistakenly identifying a difference that wasn't one is a serious error.

8.12 The computer output is given as follows.

```
Grouping Information Using Bonferroni Method and 95.0% Confidence

MomRace     N   Mean  Grouping
hispanic  164  118.5  A
white     906  117.9  A
other      48  117.1  A B
black     332  110.6    B

Means that do not share a letter are significantly different.

Bonferroni 95.0% Simultaneous Confidence Intervals
Response Variable BirthWeightOz
All Pairwise Comparisons among Levels of MomRace

MomRace = black  subtracted from:
MomRace    Lower  Center  Upper  ----+---------+---------+---------+--
hispanic   2.374   7.955  13.54                 (------*------)
other     -2.447   6.583  15.61           (-----------*-----------)
white      3.557   7.309  11.06                  (----*----)
                                 ----+---------+---------+---------+--
                                  -8.0       0.0       8.0      16.0

MomRace = hispanic  subtracted from:
MomRace   Lower  Center  Upper  ----+---------+---------+---------+--
other    -10.97  -1.372  8.223  (-----------*-----------)
white     -5.61  -0.646  4.316         (-----*-----)
                                ----+---------+---------+---------+--
                                 -8.0       0.0       8.0      16.0

MomRace = other  subtracted from:
MomRace   Lower  Center  Upper  ----+---------+---------+---------+--
white    -7.934  0.7261  9.387      (----------*----------)
                                ----+---------+---------+---------+--
                                 -8.0       0.0       8.0      16.0
```

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

8.13 a. The computer output is given as follows.

```
Grouping Information Using Bonferroni Method and 95.0% Confidence

Overwt    N   Mean  Grouping
2       204  153.2  A
1       109  144.4    B
0       187  136.3      C

Means that do not share a letter are significantly different.

Bonferroni 95.0% Simultaneous Confidence Intervals
Response Variable SystolicBP
All Pairwise Comparisons among Levels of Overwt

Overwt = 0  subtracted from:
Overwt    Lower  Center  Upper    +---------+---------+---------+------
1        0.2254   8.051  15.88    (-----------*----------)
2       10.2910  16.866  23.44                   (--------*--------)
                                  +---------+---------+---------+------
                                0.0       7.0      14.0      21.0

Overwt = 1  subtracted from:
Overwt  Lower  Center  Upper    +---------+---------+---------+------
2       1.109   8.814  16.52      (----------*----------)
                                +---------+---------+---------+------
                              0.0       7.0      14.0      21.0
```

All three weight groups are significantly different from one another.

b. The computer output is given below.

```
Grouping Information Using Tukey Method and 95.0% Confidence

Overwt    N   Mean  Grouping
2       204  153.2  A
1       109  144.4    B
0       187  136.3      C

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

```
Tukey 95.0% Simultaneous Confidence Intervals
Response Variable SystolicBP
All Pairwise Comparisons among Levels of Overwt

Overwt = 0  subtracted from:
Overwt    Lower  Center  Upper  ---------+---------+---------+-------
1        0.4260   8.051  15.68  (----------*---------)
2       10.4596  16.866  23.27                (--------*--------)
                                ---------+---------+---------+-------
                                       7.0      14.0      21.0

Overwt = 1  subtracted from:
Overwt  Lower  Center  Upper  ---------+---------+---------+-------
2       1.307   8.814  16.32   (----------*---------)
                              ---------+---------+---------+-------
                                     7.0      14.0      21.0
```

All three weight groups are significantly different from one another.

c. They give the same results.

8.14 a. The computer output is given as follows.

```
Grouping Information Using Fisher Method

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
```

Time 0 has significantly more larvae that metamorphosed than all other time periods. Time 5 has significantly more than time 25. All others are not significantly different from one another.

b. The computer output is given as follows.

```
Grouping Information Using Tukey Method

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

Means that do not share a letter are significantly different.


Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of Time

Individual confidence level = 99.51%


Time =  0 subtracted from:

Time    Lower   Center    Upper    -+---------+---------+---------+--------
 5    -0.4455  -0.1902   0.0652          (--------*-------)
10    -0.6133  -0.3580  -0.1027     (-------*--------)
15    -0.6077  -0.3523  -0.0970     (-------*--------)
20    -0.5718  -0.3165  -0.0612      (-------*--------)
25    -0.6223  -0.3670  -0.1117    (--------*-------)
                                   -+---------+---------+---------+--------
                                 -0.60     -0.30      0.00      0.30


Time =  5 subtracted from:

Time    Lower   Center   Upper    -+---------+---------+---------+--------
10    -0.4232  -0.1678  0.0875           (-------*--------)
15    -0.4175  -0.1622  0.0932           (--------*-------)
20    -0.3817  -0.1263  0.1290            (--------*-------)
25    -0.4322  -0.1768  0.0785           (-------*--------)
                                  -+---------+---------+---------+--------
                                -0.60     -0.30      0.00      0.30
```

```
Time = 10 subtracted from:

Time    Lower   Center   Upper    -+---------+---------+---------+--------
15    -0.2497   0.0057  0.2610                (-------*--------)
20    -0.2138   0.0415  0.2968                 (-------*--------)
25    -0.2643  -0.0090  0.2463               (--------*-------)
                                  -+---------+---------+---------+--------
                                -0.60     -0.30      0.00      0.30


Time = 15 subtracted from:

Time    Lower   Center   Upper    -+---------+---------+---------+--------
20    -0.2195   0.0358  0.2912                 (-------*--------)
25    -0.2700  -0.0147  0.2407               (--------*-------)
                                  -+---------+---------+---------+--------
                                -0.60     -0.30      0.00      0.30


Time = 20 subtracted from:

Time    Lower   Center   Upper    -+---------+---------+---------+--------
25    -0.3058  -0.0505  0.2048              (-------*--------)
                                  -+---------+---------+---------+--------
                                -0.60     -0.30      0.00      0.30
```

Time 0 is significantly different from times 10, 15, 20, and 25. All others are not significantly different from one another.

c. They are different. Fisher's LSD found more differences. Specifically, it found a difference between time 5 and time 25. This is consistent with the idea that the amount of larvae that would metamorphose would decrease as time went on.

8.15 a. The hypotheses are
$H_0 : \frac{1}{2}(\mu_D + \mu_E) = \frac{1}{2}(\mu_D + \mu_E)$
$H_a : \frac{1}{2}(\mu_D + \mu_E) \neq \frac{1}{2}(\mu_D + \mu_E)$

b. The comparison is $\frac{1}{2}\mu_D + \frac{1}{2}\mu_E - \frac{1}{2}\mu_F - \frac{1}{2}\mu_G$. We estimate it using $\frac{1}{2}\bar{y}_D + \frac{1}{2}\bar{y}_E - \frac{1}{2}\bar{y}_F - \frac{1}{2}\bar{y}_G$.

8.16 Contrast. We want to compare TS with the average of the other players.

8.17 a. $H_0 : \frac{1}{2}(\mu_{1p} + \mu_{8p}) = \mu_N$
$H_0 : \frac{1}{2}(\mu_{1p} + \mu_{8p}) \neq \mu_N$

b. $\frac{1}{2}(\mu_{1p}) + \frac{1}{2}(\mu_{8p}) - \mu_N$
The means for all five groups are given in the following output.

Variable	Treatment	N	Mean
Longevity	1 pregnant	25	64.80
	1 virgin	25	56.76
	8 pregnant	25	63.36
	8 virgin	25	38.72
	none	25	63.56

Putting the relevant sample means into the contrast gives us

$$\frac{1}{2}(64.8) + \frac{1}{2}(63.36) - 63.56 = 0.52$$

Therefore, the estimated value of the contrast is 0.52.

c. The standard error for the contrast is $\sqrt{MSE \sum_{i=1}^{k} \frac{c_i^2}{n_i}}$. Given below is the ANOVA table for this analysis, which gives us the relevant MSE.

One-way ANOVA: Longevity versus Treatment

Source	DF	SS	MS	F	P
Treatment	4	11939	2985	13.61	0.000
Error	120	26314	219		
Total	124	38253			

Using $MSE = 219$ in the preceding equation, we get $\sqrt{219\left(\frac{\left(\frac{1}{2}\right)^2}{25} + \frac{\left(\frac{1}{2}\right)^2}{25} + \frac{(-1)^2}{25}\right)} = 3.625$.

d. To complete the test, we compute the test statistic $t = \frac{\text{estimated value}-\text{hypothesized value}}{\text{standard error}}$. In this case, we get $t = \frac{0.52-0}{3.625} = 0.143$. Using a t-distribution with 120 degrees of freedom, we find a P-value of 0.8865. We fail to reject the null hypothesis. We do not have significant evidence that living with pregnant females results in a different life span than living alone for male fruit flies.

8.18 a. $H_0 : \frac{1}{2}(\mu_{ov} + \mu_{ob}) = \mu_N$
$H_0 : \frac{1}{2}(\mu_{ov} + \mu_{ob}) \neq \mu_N$

b. $\frac{1}{2}(\mu_{ov}) + \frac{1}{2}(\mu_{ob}) - \mu_N$
The means for all three groups are given in the output below.

Variable	Overwt	N	Mean
SystolicBP	0	187	136.32
	1	109	144.37
	2	204	153.18

Putting the relevant sample means into the contrast gives us

$$\frac{1}{2}(144.37) + \frac{1}{2}(153.18) - 136.32 = 12.455$$

Therefore, the estimated value of the contrast is 12.455.

c. The standard error for the contrast is $\sqrt{MSE\sum_{i=1}^{k}\frac{c_i^2}{n_i}}$. Given below is the ANOVA table for this analysis, which gives us the relevant MSE.

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

Using $MSE = 731$ in the above equation, we get $\sqrt{731\left(\frac{\left(\frac{1}{2}\right)^2}{109} + \frac{\left(\frac{1}{2}\right)^2}{204} + \frac{(-1)^2}{187}\right)} = 2.546$.

d. To complete the test, we compute the test statistic $t = \frac{\text{estimated value–hypothesized value}}{\text{standard error}}$. In this case, we get $t = \frac{12.455-0}{2.546} = 4.89$. Using a t-distribution with 497 degrees of freedom, we find a P-value of approximately 0. We have significant evidence that the average systolic blood pressure is different for those of normal weight than for those who are either overweight or obese. It appears that those who have normal weight have a lower systolic blood pressure.

8.19 a. $H_0 : \frac{1}{2}(\mu_s + \mu_m) = \mu_l$
$H_0 : \frac{1}{2}(\mu_s + \mu_m) \neq \mu_l$

b. $\frac{1}{2}(\mu_s) + \frac{1}{2}(\mu_m) - \mu_l$
The means for all three groups are given in the output below.

Variable	Size	N	Mean
Noise	1	12	824.17
	2	12	833.75
	3	12	772.50

Putting the relevant sample means into the contrast gives us

$$\frac{1}{2}(824.17) + \frac{1}{2}(833.75) - 772.50 = 56.46$$

Therefore, the estimated value of the contrast is 56.46.

c. The standard error for the contrast is $\sqrt{MSE\sum_{i=1}^{k}\frac{c_i^2}{n_i}}$. The following is the ANOVA table for this analysis, which gives us the relevant MSE.

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

Using $MSE = 116$ in the above equation, we get $\sqrt{116\left(\frac{\left(\frac{1}{2}\right)^2}{12}+\frac{\left(\frac{1}{2}\right)^2}{12}+\frac{(-1)^2}{12}\right)} = 3.808$.

d. To complete the test, we compute the test statistic $t = \frac{\text{estimated value}-\text{hypothesized value}}{\text{standard error}}$. In this case, we get $t = \frac{56.46-0}{3.808} = 14.83$. Using a t-distribution with 33 degrees of freedom, we find a P-value of approximately 0. We have significant evidence that the amount of noise made by large cars is different from the amount of noise made by small and medium cars taken together. It appears that large cars are not as noisy as small- and medium-sized cars.

8.20 a. $H_0 : \frac{1}{2}\left(\mu_{breast}+\mu_o\right) = \frac{1}{3}\left(\mu_s+\mu_c+\mu_{bronchus}\right)$
$H_0 : \frac{1}{2}\left(\mu_{breast}+\mu_o\right) \neq \frac{1}{3}\left(\mu_s+\mu_c+\mu_{bronchus}\right)$

b. $\frac{1}{2}\left(\mu_{breast}+\mu_o\right) - \frac{1}{3}\left(\mu_s+\mu_c+\mu_{bronchus}\right)$
The means for all five groups are given in the output that follows.

Variable	Organ	N	Mean
Survival	Breast	11	1396
	Bronchus	17	211.6
	Colon	17	457
	Ovary	6	884
	Stomach	13	286.0

Putting the relevant sample means into the contrast gives us

$$\frac{1}{2}(1396)+\frac{1}{2}(884)-\left(\frac{1}{3}(286)+\frac{1}{3}(457)+\frac{1}{3}(211.6)\right) = 821.8$$

Therefore, the estimated value of the contrast is 821.8.

c. The standard error for the contrast is $\sqrt{MSE\sum_{i=1}^{k}\frac{c_i^2}{n_i}}$. The following is the ANOVA table for this analysis, which gives us the relevant MSE.

```
Source  DF        SS       MS     F      P
Organ    4  11535761  2883940  6.43  0.000
Error   59  26448144   448274
Total   63  37983905
```

Using $MSE = 448{,}274$ in the above equation, we get

$$\sqrt{448{,}274\left(\frac{\left(\frac{1}{2}\right)^2}{11}+\frac{\left(\frac{1}{2}\right)^2}{6}+\frac{\left(\frac{1}{3}\right)^2}{13}+\frac{\left(\frac{1}{3}\right)^2}{17}+\frac{\left(\frac{1}{3}\right)^2}{17}\right)} = 196.36$$

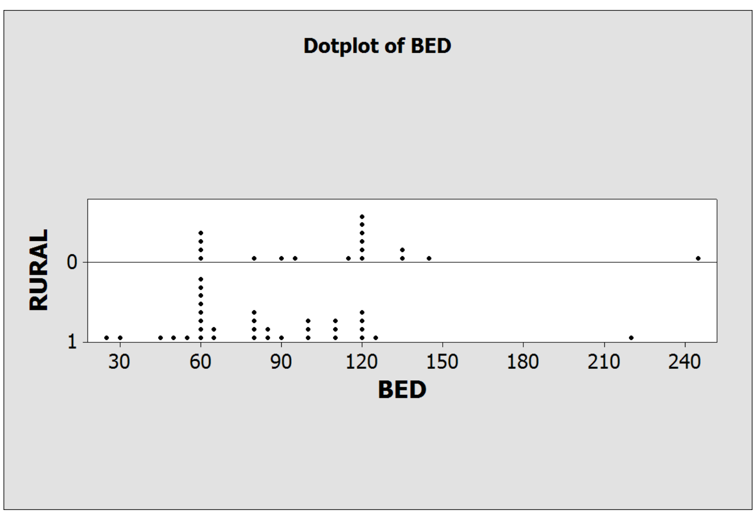
.

d. To complete the test, we compute the test statistic $t = \frac{\text{estimated value} - \text{hypothesized value}}{\text{standard error}}$. In this case, we get $t = \frac{821.8-0}{196.36} = 4.19$. Using a t-distribution with 59 degrees of freedom, we find a P-value of approximately 0. Reject the null hypothesis. We have significant evidence that the survival time for patients with breast or ovary cancers is different from patients with stomach, bronchus, or colon cancers. It appears that patients with breast or ovary cancers have longer survival times.

8.21 False. This statement implies that there are no conditions required for the Wilcoxon-Mann-Whitney test. In fact, the residuals still need to be independent and random. Further, the residuals do need to come from a continuous distribution with median 0.

8.22 False. The nonparametric tests are concerned with population medians, while the parametric tests compare population means.

8.23 a. The normality condition is violated. The dotplot shows that both groups have outliers.

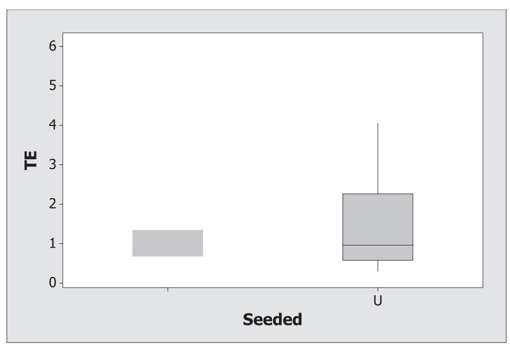


b. `Mann-Whitney Test and CI: Bed_0, Bed_1`

```
        N  Median
Bed_0  18  120.00
Bed_1  34   80.00


Point estimate for ETA1-ETA2 is 26.50
95.1 Percent CI for ETA1-ETA2 is (3.99,52.00)
W = 613.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0089
The test is significant at 0.0086 (adjusted for ties)
```

The P-value is less than 0.01. We conclude that the median number of beds are significantly different.

8.24 a. The normality condition is violated. The boxplot shows that both the seeded and unseeded clouds have outliers and the distribution of rainfall from the unseeded clouds is quite skewed to the right.

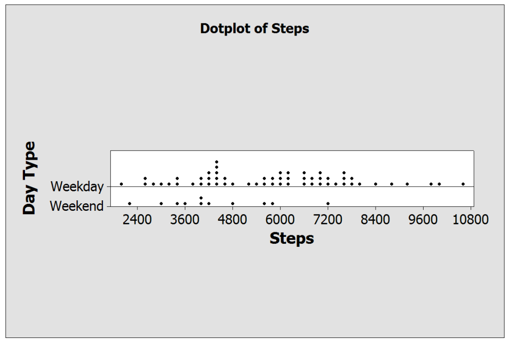


b. `Mann-Whitney Test and CI: TE_S, TE_U`

```
       N  Median
TE_S  14   0.980
TE_U  14   0.960


Point estimate for ETA1-ETA2 is -0.030
95.4 Percent CI for ETA1-ETA2 is (-0.840,0.419)
W = 199.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.8904
The test is significant at 0.8903 (adjusted for ties)
```

The P-value is 0.8904, which is greater than 0.05. We cannot conclude that the median amounts of rainfall are significantly different.

8.25 a. Either test would be ok. Both groups have distributions that look reasonably normal.



b. `Two-sample T for Moderate`

```
Day Type   N  Mean  StDev  SE Mean
Weekday   57  2618   1783      236
Weekend   11  1104    795      240


Difference = mu (Weekday) - mu (Weekend)
Estimate for difference:  1515
95% CI for difference:  (830, 2199)
T-Test of difference = 0 (vs not =): T-Value = 4.50  P-Value = 0.000  DF = 33
```

The P-value is essentially 0. We conclude that the mean number of steps is significantly different on weekdays than it is on weekend days.

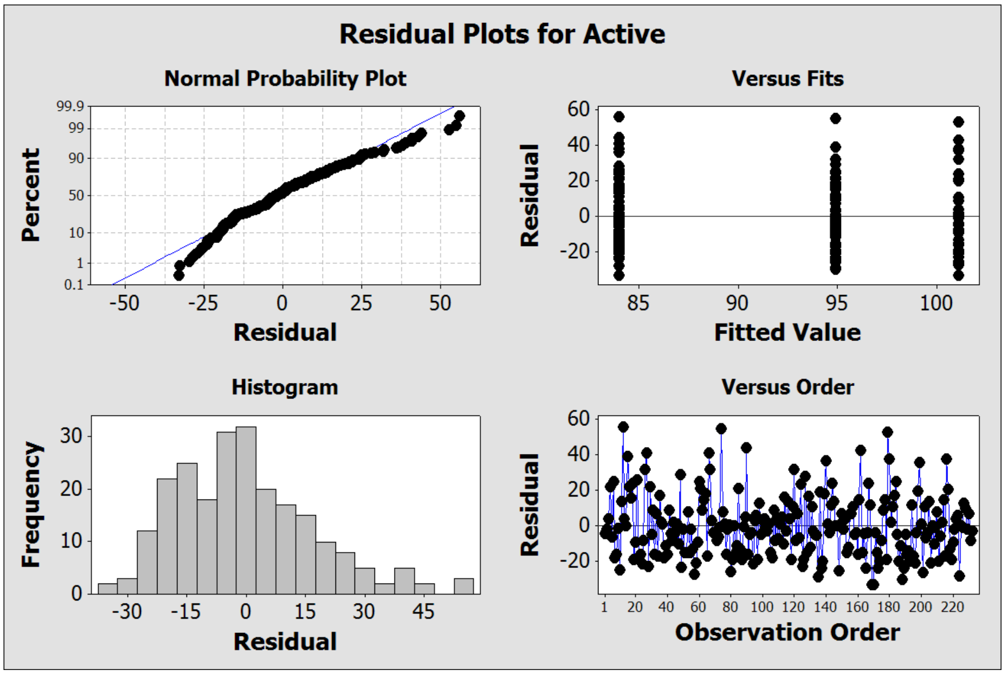
c. `Mann-Whitney Test and CI: Moderate_Weekday, Moderate_Weekend`

```
                   N  Median
Moderate_Weekday  57  2660.0
Moderate_Weekend  11  1263.0


Point estimate for ETA1-ETA2 is 1361.0
95.1 Percent CI for ETA1-ETA2 is (283.0,2613.9)
W = 2125.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0085
The test is significant at 0.0085 (adjusted for ties)
```

The P-value is very small. We conclude that the median number of steps is significantly different on weekdays than it is on weekend days.

d. Both tests give essentially the same conclusion. Both have P-values that are quite small and since the distributions are symmetric, the mean and the median are comparable measures of center.

8.26 a. Normality of residuals is violated.

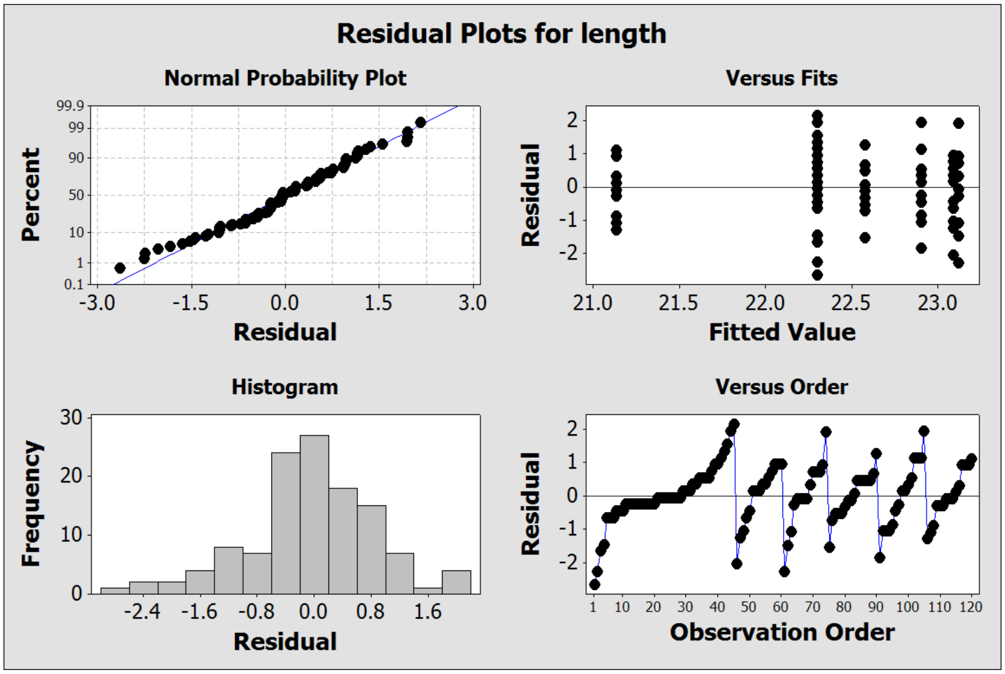


b. `Kruskal-Wallis Test on Active`

```
Exercise    N  Median  Ave Rank      Z
1          41   97.00     146.4   3.14
2          91   95.00     132.8   2.97
3         100   81.50      89.4  -5.35
Overall   232             116.5

H = 29.80  DF = 2  P = 0.000
H = 29.83  DF = 2  P = 0.000  (adjusted for ties)
```

The P-value is approximately 0. We conclude that the group medians of the *Active* variable are significantly different for the different levels of student activity.

8.27 a. Equal variance condition is violated.

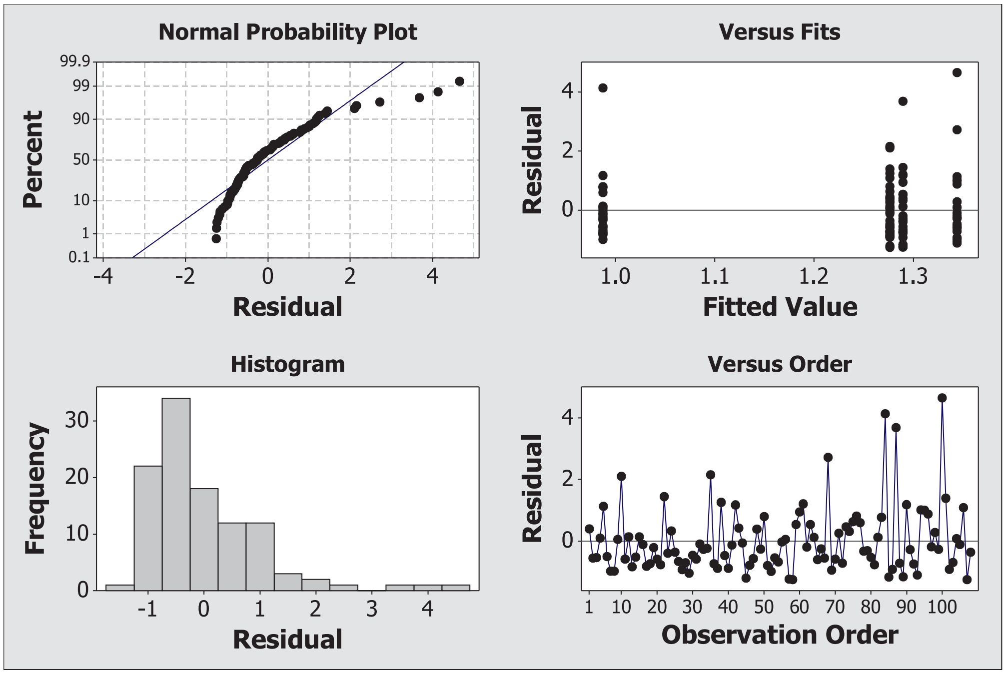


b. `Kruskal-Wallis Test on Length`

```
Bird              N  Median  Ave Rank      Z
hedge_sparrow    14   23.05      80.9   2.33
mdw_pippit       45   22.25      54.8  -1.40
robin            16   22.55      64.7   0.52
tree_pippit      15   23.25      82.8   2.65
wagtail          15   23.05      72.6   1.44
wren             15   21.05      19.8  -4.85
Overall         120              60.5

H = 34.80  DF = 5  P = 0.000
H = 35.04  DF = 5  P = 0.000  (adjusted for ties)
```

The P-value is approximately 0. We conclude that the group medians of the egg lengths are significantly different for the different types of bird nest.

8.28 a. Normality of residuals is violated.

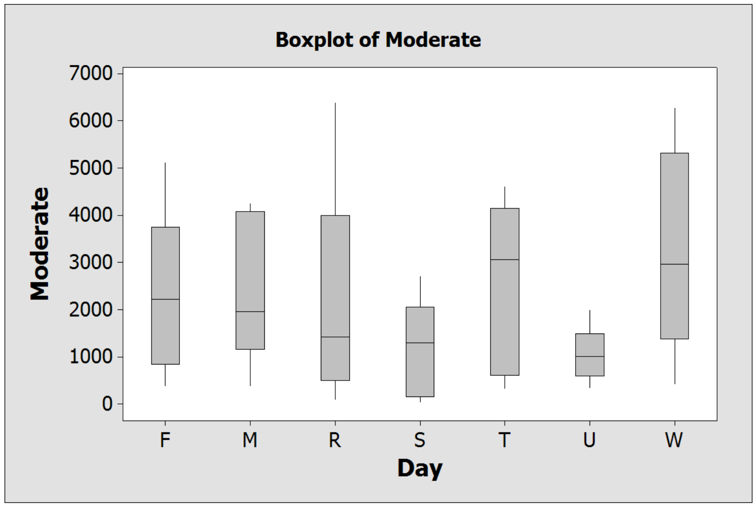


b. `Kruskal-Wallis Test: TE versus Season`

```
Kruskal-Wallis Test on TE

Season      N  Median  Ave Rank      Z
Autumn     24  0.9700      55.8   0.23
Spring     32  1.1950      58.4   0.84
Summer     24  0.7550      45.1  -1.67
Winter     28  0.9800      57.0   0.49
Overall   108              54.5

H = 2.90  DF = 3  P = 0.408
H = 2.90  DF = 3  P = 0.408  (adjusted for ties)
```

The P-value is 0.408, which is substantially larger than 0.05. We cannot reject the null hypothesis. We do not have enough information to conclude that the seasons differ with respect to the amount of rainfall per cloud.

8.29 a. While the distributions appear to be symmetric and bell-shaped, they do not have similar variances, so the Kruskal-Wallis test would be more appropriate.



b. `Kruskal-Wallis Test: Moderate versus Day`

```
Kruskal-Wallis Test on Moderate

Day        N  Median  Ave Rank      Z
F         12    2231      34.8   0.05
M         11    1972      37.8   0.60
R         10    1428      31.6  -0.50
S          5    1301      19.8  -1.73
T         10    3071      37.9   0.58
U          6    1018      20.3  -1.84
W         14    2970      42.7   1.74
Overall   68              34.5

H = 9.06  DF = 6  P = 0.170
H = 9.06  DF = 6  P = 0.170  (adjusted for ties)
```

The P-value is greater than 0.05. We cannot conclude that the median number of steps per day is different for the different days of the week.

c. Even though the medians for Saturday (S) and Sunday (U) are the smallest of the seven days, the variability of steps taken within each day of the week is pretty large. The values for the Saturday and Sunday medians are not unusually small, considering the amount of variability within each day. However, when all weekdays are combined and both weekend days are combined, the difference between the two medians is bigger in comparison to the amount of variability of the individual measurements.

8.30 The Minitab output below matches, up to slight roundoff differences, the R output provided in the text. The value of the test statistic is 14.95 and the P-value is 0.005.

```
Kruskal-Wallis Test: Survival versus Organ

Kruskal-Wallis Test on Survival

Organ      N  Median  Ave Rank      Z
Breast    11  1166.0      47.0   2.84
Bronchus  17   155.0      23.3  -2.37
Colon     17   372.0      35.9   0.88
Ovary      6   406.0      40.2   1.06
Stomach   13   124.0      24.2  -1.79
Overall   64              32.5

H = 14.95  DF = 4  P = 0.005
H = 14.95  DF = 4  P = 0.005  (adjusted for ties)
```

8.31 It is possible to list the set of triples that have an average of 32 or more, or, equivalently, a total of 96 or more. There are only three of them: 23/33/40, 25/33/40, and 26/33/40. So the probability is $\frac{3}{35} = 0.086$.

8.32 a. The P-value is 1.0, as large as any P-value can ever get. Conclusion: There is no evidence that temperature has an effect on O-ring failure.

b. The P-value is $0.1\% + 0.8\% + 0.1\% = 1\%$, or 0.01. Conclusion: There is strong evidence that O-ring failures are more frequent when the temperature is below 65 degrees.

8.33 a. There are $\binom{7}{4} = 35$ ways to choose a subset of four from a set of seven. If the choice is made purely at random, these 35 ways are equally likely. If the treatment has no effect, the chance of getting a subset with an average of 16 or more is $\frac{1}{35} = 0.03$ because only one subset has an average that large.

b. The P-value is $\frac{2}{35} = 0.06$.

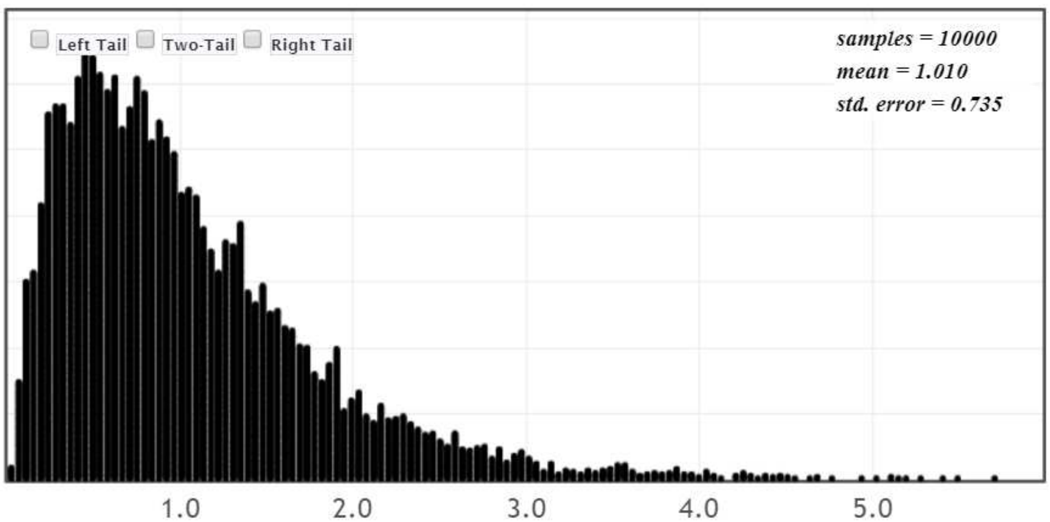
c. Both tests use the same randomization, which assigns probability $\frac{1}{35}$ to each of the 35 subsets of size four. However, the proportion in (a) only count subsets with an average as large or larger than 16 (as in a one-sided test). The test based on the F-statistic is two-sided—we count subsets with an unusually high average *or* an unusually low average, in the present case, one subset of each kind.

d. The proportion from (a) makes more sense here because the research alternative hypothesis is one-sided, $H_a : \mu_T > \mu_C$, that the chimps exposed to the demo chimps will crack more nuts on average. The randomization F-test has an alternative that $H_a : \mu_T \neq \mu_C$, that the difference is large in *either* direction.

e. A t-test to compare two means can be either one-sided (find the P-value from one tail) or two-sided (use extremes in both tails to find the P-value). The alternative hypothesis for an

F-test is always two-sided (a large difference in either direction), even though we only look in the upper tail in the distribution to find a P-value (since the deviations are all squared when computing the F-statistic). When comparing just two groups, a two-sided t-test is equivalent to the ANOVA F-test

8.34 The graph below shows F-statistics for 10,000 samples generated under a null hypothesis of no difference in fruit fly lifetimes depending on the number of potential mates. The largest statistic in these 10,000 simulations is $F = 5.69$, well below the $F = 13.61$ observed for the actual sample. $F = 13.61$ is a very extreme statistic, very unlikely to occur by random chance alone. The estimated P-value from this randomization distribution is zero. This finding is consistent with the very small P-value from the original ANOVA table.



8.35 In one set of 10,000 simulations we found:

Only 1 F-statistic for $B12$ that was greater than 60.331, so the randomization P-value= 0.0001.

573 F-statistics for *Antibiotics* beyond 5.297, so the randomization P-value= 0.0573.

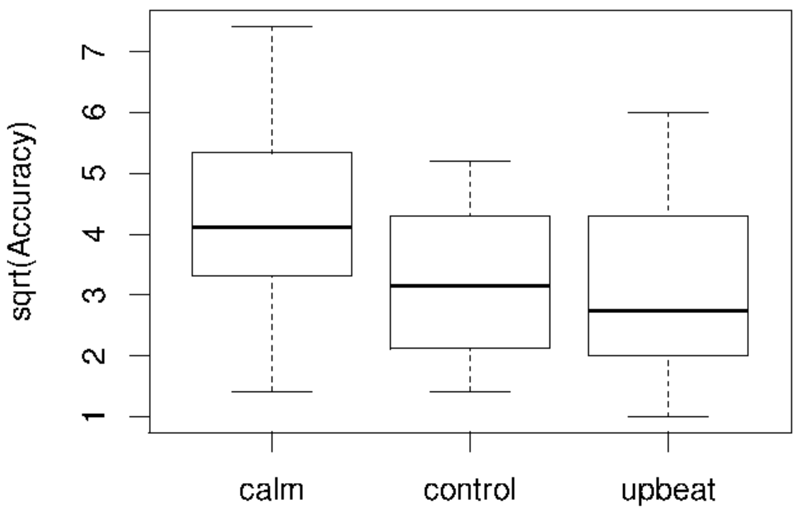
8 F-statistics for *B12*Antibiotics* beyond 47.669, so the randomization P-value= 0.0008.

These give much the same conclusions as we found in the original ANOVA table.

8.36 There is one within-subjects variable: *Medicine* (placebo, drug A, drug B). There are no between-subjects variables.

8.37 There are two within-subjects variables: word style (typed versus handwritten) and sound (music versus silence). These factors are crossed. There are no between-subjects variables.

8.38 a. Here are the side-by-side boxplots.



Here are the means and standard deviations for *sqrt*(*Accuracy*) within each group.

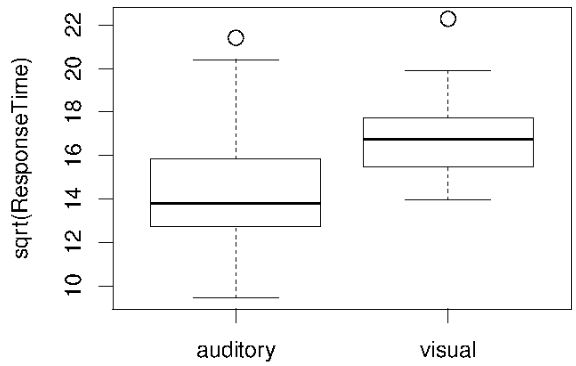
Music	Mean	SD
calm	4.32	1.63
control	3.20	1.26
upbeat	3.12	1.42

Variability is similar in all three groups, but the *Accuracy* is somewhat worse when listening to calm music.

b. The null hypothesis is that there is no difference in average accuracy across types of music. In symbols
$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ where α_i is the effect of being in *Music* group i, or
$H_0 : \mu_1 = \mu_2 = \mu_3$ where μ_i is the mean *sqrt*(*Accuracy*) for *Music* group i.

c. The boxplots provide some evidence that accuracy is worse when listening to calm music.

8.39 a. Here are the side-by-side boxplots.



Here are the means and standard deviations for $sqrt(ResponseTime)$ for the two types of stimuli.

Stimulus	Mean	SD
auditory	14.43	2.81
visual	16.91	1.70

The response times tend to be faster for the auditory stimulus, but slightly more variable than for the visual stimulus.

b. The null hypothesis is that there is no difference in average (sqrt) response times between the two kinds of stimulus. In symbols
$H_0 : \alpha_1 = \alpha_2 = 0$ where α_1 is the effect of the auditory stimulus and α_2 is for visual, or
$H_0 : \mu_a = \mu_v$ where μ is the mean $sqrt(ResponseTime)$ for the respective stimuli.

c. The boxplots provide some evidence that response times are faster with the auditory stimulus.

8.40 Here is some computer output for doing a repeated measures ANOVA on the **MusicTime** data.

```
Error: Subject
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 19   66.7      3.51

Error: Subject:Music
          Df Sum Sq  Mean Sq  F value   Pr(>F)
Music      2   18.0     9.02     6.52   0.0037
Residuals 38   52.5     1.38
```

The P-value for the *Music* factor (0.0037) provides strong evidence for a difference in mean $sqrt(Accuracy)$ depending on the type of music (or silence).

Although you may use different software to get similar results, the R command to obtain this output is

```
summary(aov(sqrt(Accuracy)~Music+Error(Subject/Music),data=MusicTime))
```

8.41 Here is some computer output for doing a repeated measures ANOVA on the **AudioVisual** data.

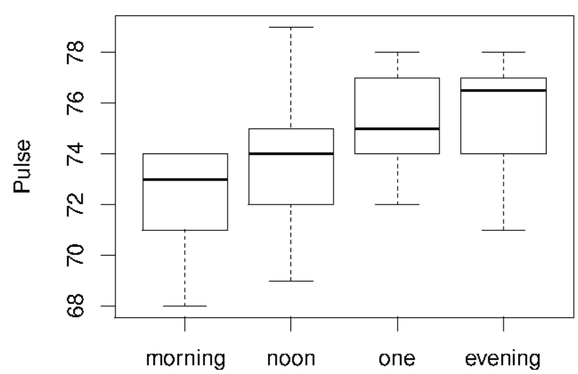
```
Error: Subject
          Df  Sum Sq Mean Sq F value Pr(>F)
Residuals 35     295   8.42

Error: Subject:Stimulus
          Df Sum Sq  Mean Sq  F value   Pr(>F)
Stimulus   1  111.1    111.1     46.7  6.2e-08
Residuals 35   83.2      2.4
```

The P-value for the *Stimulus* factor (6.2×10^{-8}) is tiny, providing strong evidence for a difference in mean $sqrt(ResponseTime)$ between the auditory and visual stimulus. From the cell means, we see that the mean is smaller (faster response) for auditory stimuli.

Although you may use different software to get similar results, the R command to obtain this output is

```
summary(aov(sqrt(Response)~Stimulus+Error(Subject/Stimulus),data=AudioVisual))
```

8.42 Here is a set of side-by-side boxplots to compare the pulse rates for different times of the day.



The boxplot suggests that pulse rates might tend to increase as the day goes along.

Here are the means and standard deviations for *Pulse* for at each *Time* of the day.

Time	Mean	SD
morning	72.23	1.93
noon	74.15	2.56
one	75.38	1.44
evening	75.69	2.13

Variability is relatively consistent among the four groups (no cell SD is more than twice another). As with the boxplots, the mean pulse increases as the day does along.

We use a repeated measures ANOVA to test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ where the μ_i represent the mean pulse at each of the respective times of the day.

Here is some computer output for fitting this model.

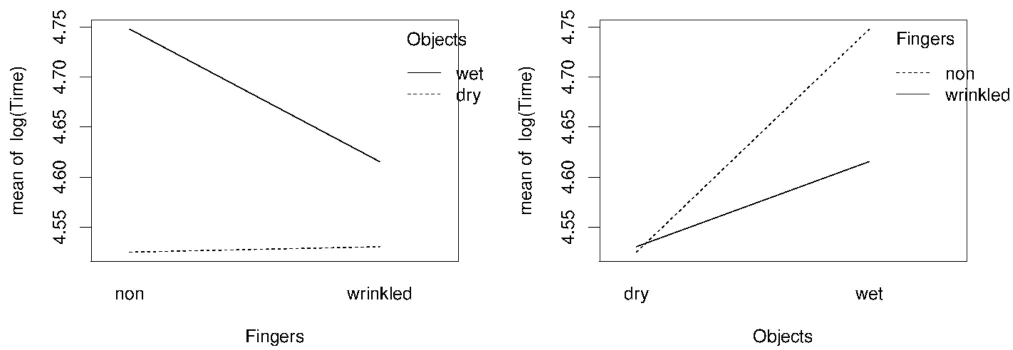
```
Error: Day
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 25    104    4.14

Error: Day:Time
          Df Sum Sq Mean Sq F value  Pr(>F)
Time       3    192    64.1    15.1 8.6e-08
Residuals 75    318     4.2
```

The tiny P-value for the *Time* factor provides very strong evidence that this student's average pulse rate is different at different times of the day. From the boxplots and cell means, it would appear that mean pulse rate increases as the day goes along.

8.43 a. Here is a table of cell means for $log(Time)$, along with row and column means.

		Fingers		
		nonwrinkled	wrinkled	Row mean
Objects	dry	4.53	4.53	4.53
	wet	4.75	4.62	4.68
	Column mean	4.64	4.57	

b. Two versions of the interaction graph are possible.



c. It looks like participants go faster with dry objects and whether fingers are wrinkled or not makes little difference. However, with wet objects, although the time tends to be longer, the increase is quite a bit larger for nonwrinkled fingers than it is for wrinkled fingers.

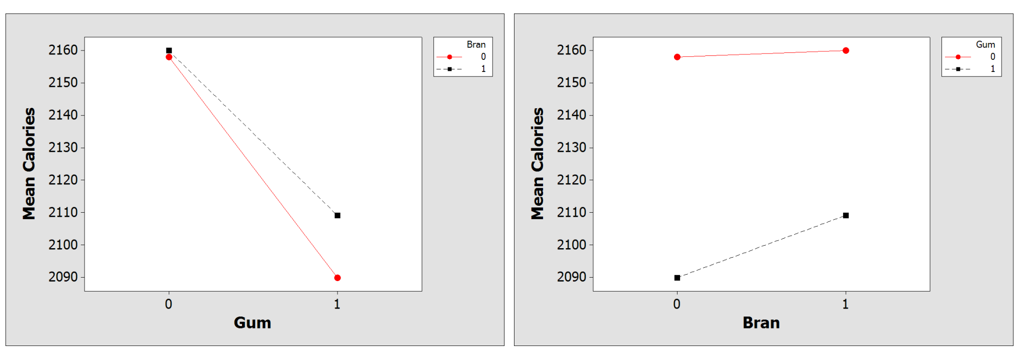
8.44 a. The two factors are Bran (no/yes) and Gum (no/yes). Here is a table showing how each cell in the 2 × 2 table corresponds to each of the cracker types.

	Bran	
Gum	No	Yes
No	Control	Bran
Yes	Gum	Combo

b. Here is a table of cell means, along with row and column means.

	Bran		
Gum	No	Yes	Ave
No	2158.0	2160.0	2159.0
Yes	2090.0	2109.0	2099.5
	2124.0	2134.60	

c. Two versions of the interaction graph are possible.



d. For this particular dataset, the graph on the right is easier to interpret. We assume that the purpose of the fiber is to reduce the number of digested calories, which means that lower values are better. Thus the best of the four treatments is the gum only, with no bran. The main effect of gum is substantial: On average, gum reduced the number of digested calories by 60. On balance, bran actually made things worse, increasing the number of calories a modest amount, about 10 calories on average. However, the data show a possible interaction: When bran was added to the control diet, the effect was negligible, but when bran was added to gum, the number of digested calories went up by 20. However, we should still do a test to see if this difference is more than we'd expect by chance variation.

8.45 Here is some output for fitting the repeated measures model.

```
Error: Participant
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 19   2.12   0.112

Error: Participant:Fingers
          Df Sum Sq Mean Sq F value Pr(>F)
Fingers    1 0.0805  0.0805    6.72  0.018 *
Residuals 19 0.2278  0.0120
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Error: Participant:Objects
          Df Sum Sq Mean Sq F value  Pr(>F)
Objects    1  0.474   0.474      69 9.5e-08 ***
Residuals 19  0.130   0.007
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Error: Participant:Fingers:Objects
                Df Sum Sq Mean Sq F value  Pr(>F)
Fingers:Objects  1 0.0949  0.0949    74.4 5.4e-08 ***
Residuals       19 0.0243  0.0013
```

a. There is evidence (P-value $= 0.018$) of a difference between the two *Fingers* conditions.

b. There is strong evidence (P-value < 0.0001) of a difference between the two *Objects* conditions.

c. There is strong evidence (P-value < 0.0001) of an interaction between *Fingers* and *Objects*.

8.46 Here is some output for fitting the repeated measures model.

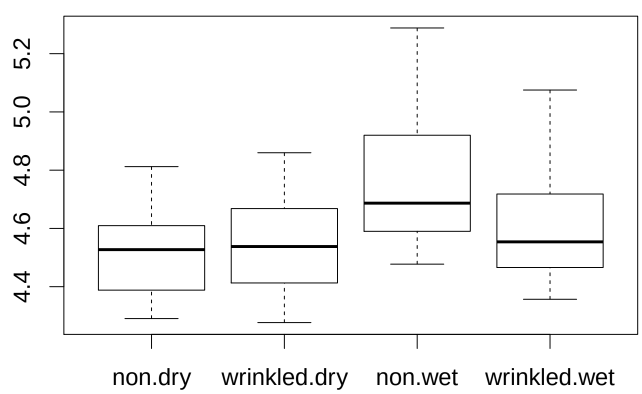
```
Error: Subj
          Df  Sum Sq Mean Sq F value Pr(>F)
Residuals 11 2323146  211195

Error: Subj:Bran
          Df Sum Sq Mean Sq F value Pr(>F)
Bran       1   1338    1338   0.02   0.89
Residuals 11 765234   69567

Error: Subj:Gum
          Df Sum Sq Mean Sq F value Pr(>F)
Gum        1  42407   42407   1.04    0.33
Residuals 11 449740   40885

Error: Subj:Bran:Gum
          Df Sum Sq Mean Sq F value Pr(>F)
Bran:Gum   1    894     894    0.06   0.81
Residuals 11 162974   14816
```

a. There is not convincing evidence (P-value = 0.89) of a difference due to Bran.

b. There is not convincing evidence (P-value = 0.33) of a difference due to Gum.

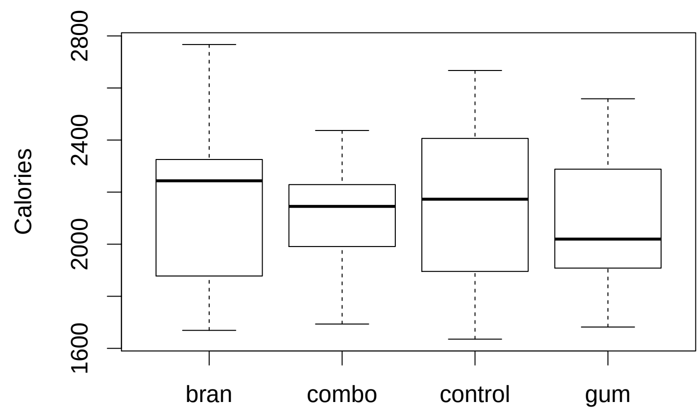
c. There is not convincing evidence(P-value = 0.81) of an interaction between *Bran* and *Gum*.

8.47 Here are side-by-side boxplots for each cell in the **Wrinkle** data.



The story in the boxplots is similar to what we observed in the interaction plot—not much difference between wrinkled and nonwrinkled fingers when handling dry objects, but a big increase in time when nonwrinkled fingers handle wet objects. The effect is confirmed by the small P-value for the interaction term in the ANOVA.

8.48 Here are side-by-side boxplots for each cell in the **CrackerFiber** data.



Although the median is slightly higher for Bran alone and lowest for Gum alone, there is a lot of variability and overlap between the distributions in all four cells. This is what is missing in the interaction plot which focuses on just how the centers differ. While that plot is useful for interpreting the nature of an interaction when present, the ANOVA table is needed to help us determine if the interaction is statistically significant.

8.49 a. $Y = \mu + \alpha_k + \epsilon$, where α_k is the treatment effect for $k = 1, 2, 3$.

b. $Y = \beta_0 + \beta_1 lemon + \beta_2 paper + \epsilon$ where $lemon = 0$ if no lemon juice was present, $lemon = 1$ if lemon juice was used, $paper = 0$ if no paper towel was present, $paper = 1$ if paper towel was used.

c. β_0 represents the mean shelf life for control strawberries. β_1 represents the change in mean effect on shelf life of adding lemon juice. β_2 represents the change in mean effect on shelf life of storing the strawberries on a paper towel.

8.50 a. $Y = \mu + \alpha_k + \epsilon$, where α_k is the color effect for $k = 1, 2, 3, 4$.

b. $Y = \beta_0 + \beta_1 Ind_E + \beta_2 Ind_F + \beta_3 Ind_G + \epsilon$, where $Ind_E = 1$ if the diamond was color E, $Ind_F = 1$ if the diamond was color F, $Ind_G = 1$ if the diamond was color G.

c. β_0 represents the mean number of carats for D-colored diamonds. β_1 represents the change in mean number of carats for E-colored diamonds. β_2 represents the change in mean number of carats for an F-colored diamond. β_3 represents the change in mean number of carats for a G-colored diamond.

8.51 a. Since there are four categories to the *Momrace* variable, we need three indicator variables in the model.

b. (Answers will vary depending on which indicator is left out of the model.) The regression equation is given below.

```
BirthWeightOz = 117.15 - 6.58 MomRace_black + 1.37 MomRace_hispanic
                + 0.73 MomRace_white
```

This equation suggests that children of mothers of "other races" have mean birth weight of 117.15 ounces ($\hat{\beta}_0$). Children of black mothers have mean birth weight 6.58 ounces less than those of mothers of "other races." In other words, their mean birth weight is $117.15 - 6.58 = 110.57$ ounces ($\hat{\beta}_0 + \hat{\beta}_1$). Children of Hispanic mothers have mean birth weight 1.37 ounces more than those of mothers of "other races." They have a mean birth rate of $117.15 + 1.37 = 118.52$ ounces ($\hat{\beta}_0 + \hat{\beta}_2$). And children of white mothers have mean birth weight 0.73 ounces more than those of mothers of "other races." They have a mean birth rate of $117.15 + 0.73 = 117.87$ ounces ($\hat{\beta}_0 + \hat{\beta}_3$). Note that all of these combinations of estimated coefficients are the same as the estimated group effects in the ANOVA model.

c. The ANOVA table produced by the regression model is given below.

Source	DF	SS	MS	F	P
Regression	3	14002.4	4667.5	9.53	0.000
Residual Error	1446	708331.7	489.9		
Total	1449	722334.1			

And the ANOVA table produced by the ANOVA model is given below.

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

Note that, while they call the sources of variation different things, the numerical values are identical. In the regression case, this means that at least one group has a different mean weight than the "other races" (this is because we set up the "other races" as being the null group.).

8.52 a. The ANOVA table is given below.

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			

b. (Answers will vary depending on which indicator is left out of the model.) The regression equation is given by:

```
exponential fenthion = 7.22 - 3.46 Time_3 - 4.44 Time_4
```

Note that the base model is for time 0. This suggests that the average amount of the exponential of fenthion at time 0 is 7.22. The mean decreases by 3.46 at time 3 and by 4.44 at time 4 (in comparison to time 0).

c. If we look at $\hat{\beta}_0$, $\hat{\beta}_0 + \hat{\beta}_1$, and $\hat{\beta}_0 + \hat{\beta}_2$, we will get the estimated group effects from the ANOVA model.

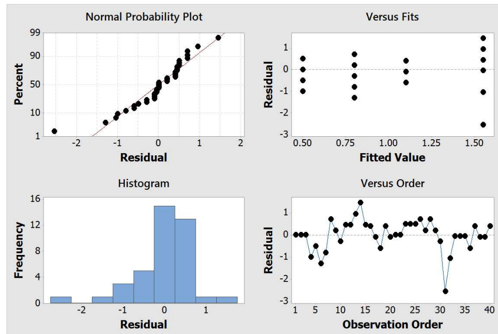
d. The regression output is given as follows.

```
The regression equation is
exponential fenthion = 7.20 - 1.12 Time


Predictor      Coef   SE Coef       T      P
Constant     7.1999    0.1791   40.20  0.000
Time       -1.11957   0.06204  -18.05  0.000

S = 0.447391   R-Sq = 95.3%   R-Sq(adj) = 95.0%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  65.179  65.179  325.64  0.000
Residual Error  16   3.203   0.200
Total           17  68.381
```

We don't show the graphs here, but the conditions for regression are met in this case. We conclude that there is a linear relationship between time and the exponential amount of fenthion present in the olive oil.

e. We prefer to use the regression analysis in this case because there is more to the relationship than just a difference among the groups; there is actually a linear relationship.

8.53 a. (Answers will vary depending on which indicator is left out of the model.) The conditions for regression are met as evidenced by the following plots.



The regression equation is given as follows.

Regression Equation

`diff = 0.500 + 0.300 Ultra_10 + 1.050 Oil_10 - 0.750 inter`

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	0.500	0.228	2.19	0.035
Ultra_10	0.300	0.323	0.93	0.358
Oil_10	1.050	0.323	3.26	0.002
inter	-0.750	0.456	-1.64	0.109

Model Summary

S	R-sq	R-sq(adj)
0.721207	24.32%	18.02%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	6.019	2.0062	3.86	0.017
Error	36	18.725	0.5201		
Total	39	24.744			

The overall model is significant as is the constant term and the oil main effect. The ultrasound main effect and the interaction terms are not significant.

b. The y-intercept of 0.500 suggests that we should expect a difference of 0.5 ml of oil between control samples and treated samples (control samples having more oil) when the treatment is 5 ml of oil and 5 minutes of ultrasound and this value is significantly different from 0. The positive coefficient for the oil main effect implies that as we go from 5 ml of oil to 10 ml of oil, we increase the difference in oil remaining between the treatment and control samples by 1.05 ml on average. This number, too, is significantly different from 0. The positive coefficient for the ultrasound main effect implies that as we go from 5 minutes of ultrasound to 10 minutes of ultrasound, we increase the difference in oil remaining between the treatment and control samples by 0.3 ml. However, this value is not significantly different from 0. Finally, if we both increase the amount of oil from 5 to 10 ml and increase the amount of ultrasound from 5 to 10 minutes, we need to subtract 0.75 ml from the total amount of oil difference we expect. This value, though, like the main effect for ultrasound is not significant.

c. Both analyses find the overall model significant. Both result in the same R^2 value. The ANOVA model looks at just the oil alone, just the amount of ultrasound alone and the interaction between the two variables. The only significant factor is oil. The regression model takes uses 5 ml of oil and 5 minutes of ultrasound as the baseline and then uses 10 ml of oil, 10 minutes of ultrasound and their interaction as the variables. Only the constant term and the oil variable are significant.

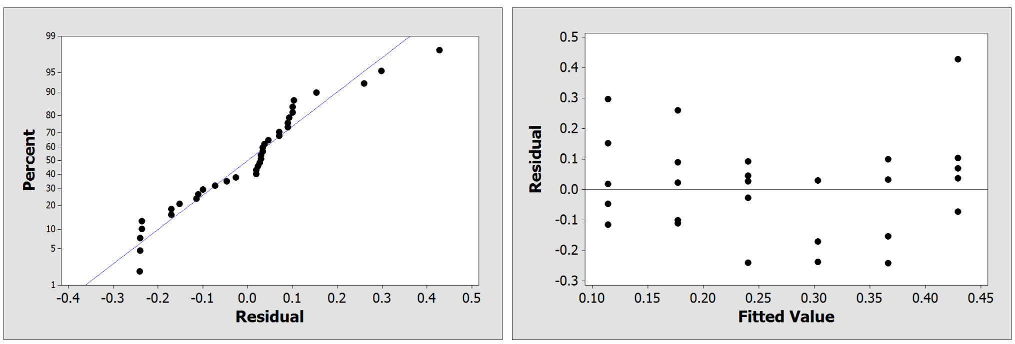
8.54 a. (Answers will vary depending on which indicator is left out of the model.) Using time 0 as the base time, the regression output is given as follows.

```
The regression equation is
Percent = 0.536 - 0.190 Time_5 - 0.358 Time_10 - 0.352 Time_15 - 0.317 Time_20
          - 0.367 Time_25


Predictor      Coef  SE Coef      T      P
Constant    0.53567  0.05938   9.02  0.000
Time_5     -0.19017  0.08397  -2.26  0.031
Time_10    -0.35800  0.08397  -4.26  0.000
Time_15    -0.35233  0.08397  -4.20  0.000
Time_20    -0.31650  0.08397  -3.77  0.001
Time_25    -0.36700  0.08397  -4.37  0.000


S = 0.145446   R-Sq = 49.9%   R-Sq(adj) = 41.5%


Analysis of Variance

Source          DF       SS       MS     F      P
Regression       5  0.63091  0.12618  5.96  0.001
Residual Error  30  0.63464  0.02115
Total           35  1.26554
```

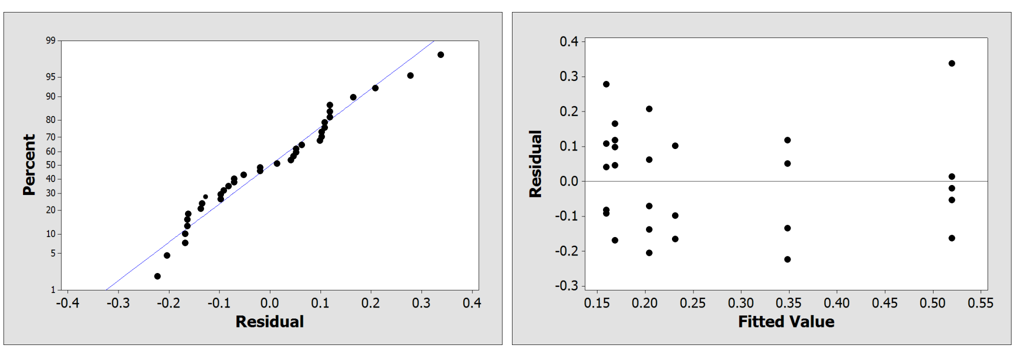
The y-intercept gives the estimated value of the mean percent of sea slugs that metamorphosed at time 0. $\hat{\beta}_1$ gives the difference between the estimated mean percent that metamorphosed at time 0 and the estimated mean percent that metamorphosed at time 5. Each slope coefficient after that in the model gives the difference between the estimated mean percent that metamorphosed at time 0 and the estimated percent that metamorphosed at a different time period. Since all of the estimated coefficients are negative, this suggests that the mean percent that metamorphosed was highest at time 0.

b. The overall model explains a significant amount of variation and each of the coefficients, individually, are also statistically significant. The mean percent of sea slugs that metamorphosed at time 0 is significantly different from 0, and the mean percent of sea slugs that metamorphosed at all other time periods are significantly different from the mean percent that metamorphosed at time 0.

c. We start by checking the conditions. Independence and randomness have been met. The following are the normal probability plot and the residual plot.



The normal probability plot is very roughly linear, but there is clearly curvature in the residual plot. So we fit the quadratic model. In this case, the conditions are met and we proceed (graphs given below).



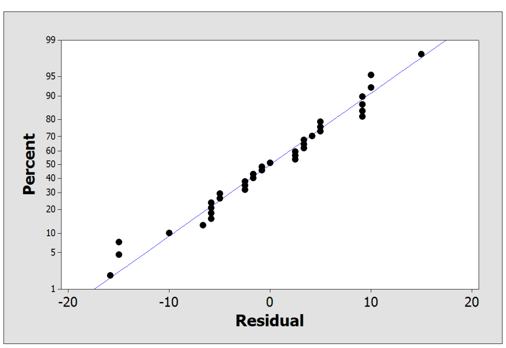
The regression output is given below.

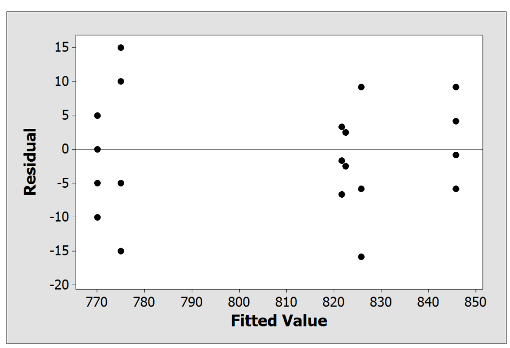
```
  The regression equation is
Percent = 0.519 - 0.0396 Time + 0.00108 time squared


Predictor           Coef    SE Coef      T      P
Constant         0.51946    0.05327   9.75  0.000
Time            -0.03964    0.01002  -3.96  0.000
time squared   0.0010807  0.0003848   2.81  0.008


S = 0.143972   R-Sq = 46.0%   R-Sq(adj) = 42.7%


Analysis of Variance

Source          DF       SS       MS      F      P
Regression       2  0.58152  0.29076  14.03  0.000
Residual Error  33  0.68403  0.02073
Total           35  1.26554
```

The overall model explains a significant amount of variation and each of the coefficients, individually, are also statistically significant. The adjusted R^2 value is 0.427, which means that the model explains about 43% of the variation in the mean amount of sea slugs that metamorphosed.

d. We prefer to use the regression analysis in this case because there is more to the relationship than just a difference among the groups; there is actually a quadratic relationship.

8.55 a. Since this was a randomized experiment, the observations should be independent of each other. The normal probability plot and residual plot are given as follows and show no concerning patterns.





The ANOVA table is given below.

Source	DF	SS	MS	F	P
Size	2	26051.4	13025.7	199.12	0.000
Type	1	1056.3	1056.3	16.15	0.000
Interaction	2	804.2	402.1	6.15	0.006
Error	30	1962.5	65.4		
Total	35	29874.3			

```
S = 8.088   R-Sq = 93.43%   R-Sq(adj) = 92.34%
```

We conclude that both size and type are significant, as is the interaction.

b. The normal probability and residual plots are identical to those produced in part (a). The output from the regression analysis is given below.

```
    The regression equation is
Noise = 770 + 52.5 Size_1 + 51.7 Size_2 + 5.00 Type_1 - 1.67 Size1Type1
        + 19.2 Size2Type1
```

Predictor	Coef	SE Coef	T	P
Constant	770.000	3.302	233.20	0.000
Size_1	52.500	4.670	11.24	0.000
Size_2	51.667	4.670	11.06	0.000
Type_1	5.000	4.670	1.07	0.293
Size1Type1	-1.667	6.604	-0.25	0.802
Size2Type1	19.167	6.604	2.90	0.007

```
S = 8.08806   R-Sq = 93.4%   R-Sq(adj) = 92.3%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	27911.8	5582.4	85.34	0.000
Residual Error	30	1962.5	65.4		
Total	35	29874.3			

770 represents the estimated mean noise level for large cars with the new filter. If we stay with the new filter and move to medium-sized cars, we add 51.667 to the mean noise level.

If we stay with the new filter and move to small cars (from large), we add 52.5 to the noise level. If we stay with large cars and move to the standard filter, we add 5 to the noise level. If we move to medium-sized cars and the standard filter, we add 51.667, 5.000, and 19.167 to the value for large cars with the new filter. If we move to small cars and the standard filter, we add 52.5 and 5.000, and subtract 1.667 to the value for large cars with the new filter.

c. The regression analysis says that medium and small cars are significantly different from large cars in terms of noise. Filter type is not significant for large cars or small cars, but is significant for medium-sized cars.

d. Both analyses find the overall model significant. Both result in the same R^2 value. The ANOVA model looks at just the size alone, just the type of filter alone and the interaction between the two variables. All of these are significant. The regression model takes each of the five other groups of vehicles and compares them to the large vehicles with the new filter.

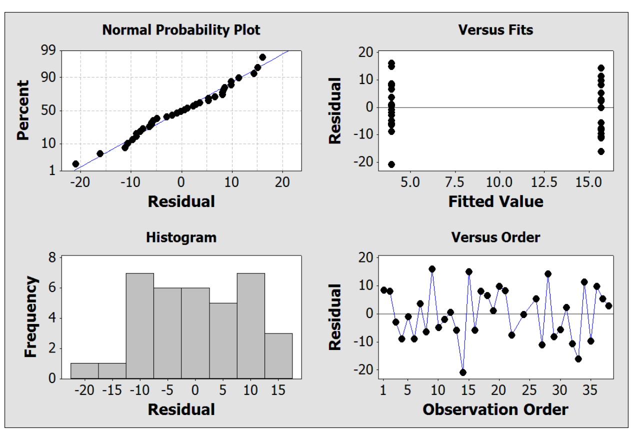
8.56 a. The residuals from the ANOVA model must be independent and random. They must be normally distributed and have equal variances.

b. The relationship between the explanatory and response variables must be linear. The residuals from the linear regression model must be independent and random. They must be normally distributed and have equal variances.

c. The ANOVA conditions must be satisfied when looking at the relationship between the response variable and the categorical factor. The linear regression conditions must be satisfied when modeling the relationship between the response variable and the covariate. Finally, there must be no interaction between the covariate and the factor of interest.

8.57 c. When we transform data, some conditions that had been met before may now not be met. We need to recheck everything before proceeding with the analysis.

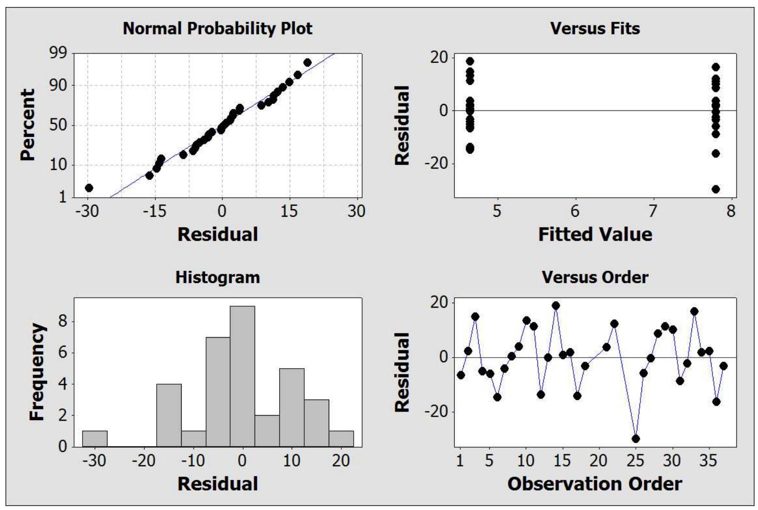
8.58 a. We start by checking the conditions. This was a randomized experiment, so the observations should be independent of each other. The plots of the residuals shown below indicate that there are no issues with either normality or constant variance of the residuals. The conditions are met so we can proceed with the ANOVA.



The ANOVA table that follows indicates that we have significant evidence that the initial average weight loss is different for the incentive group than for the control group.

Source	DF	SS	MS	F	P
Group	1	1239.9	1239.9	14.48	0.001
Error	34	2911.1	85.6		
Total	35	4151.0			

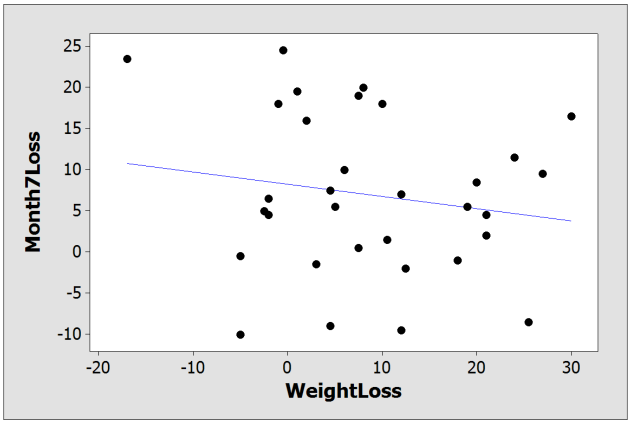
b. We start by checking the conditions. This was a randomized experiment so the observations should be independent of each other. The plots of the residuals shown below indicate that there are no issues with either normality or constant variance of the residuals. The conditions are met so we can proceed with the ANOVA.



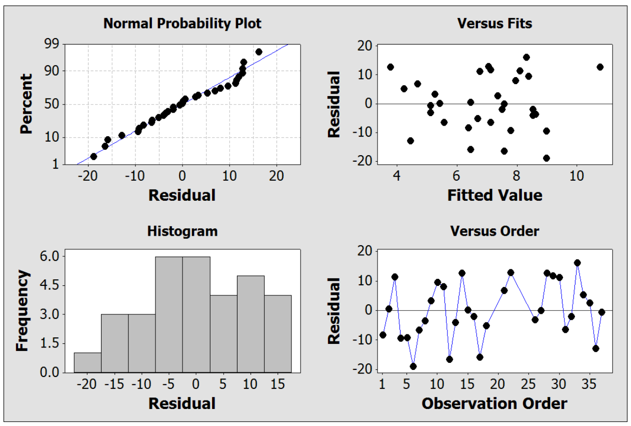
The ANOVA table below indicates that the final average weight loss is not significantly different for the incentive group than for the control group.

Source	DF	SS	MS	F	P
Group	1	82	82	0.69	0.413
Error	31	3680	119		
Total	32	3762			

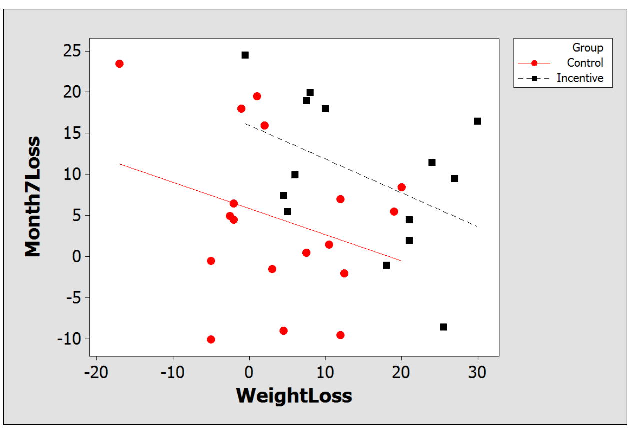
c. We start by checking conditions one more time. We already know that the conditions required by the ANOVA have been met. As seen in the scatterplot below, there does appear to be a weak negative linear relationship between the initial weight loss and the final weight loss.



And the normality and constant variance conditions also hold as evidenced by the graphs below.



Finally, we check to see if both groups have similar slopes. The scatterplot below suggests that the estimated regression lines are close enough to having the same slope that this condition is met.

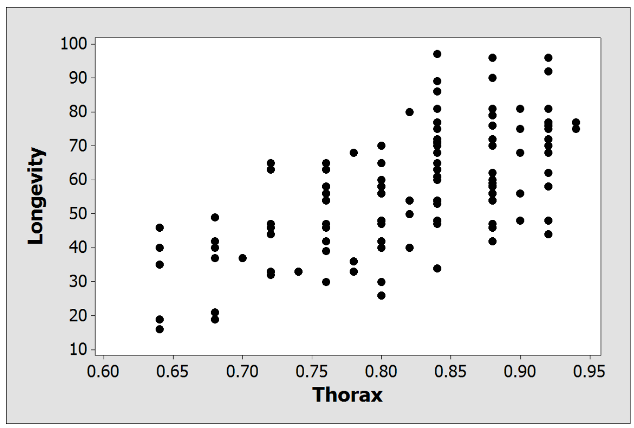


All conditions have been met. The ANOVA table is given below.

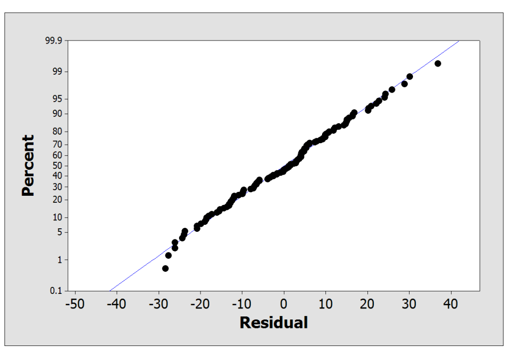
Source	DF	Seq SS	Adj SS	Adj MS	F	P
WeightLoss	1	82.72	365.70	365.70	4.49	0.043
Group	1	503.32	503.32	503.32	6.18	0.019
Error	29	2362.63	2362.63	81.47		
Total	31	2948.68				

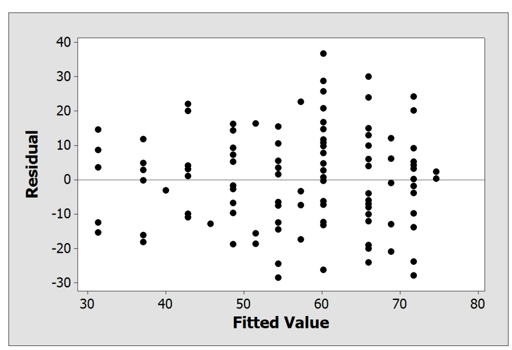
We now conclude that the incentive group did have a significantly different weight loss at the final check-in than the control group, when we took the initial amount of weight loss into consideration.

8.59 a. We checked the conditions for the ANOVA model in Chapter 5 and they were met. We now proceed to check the conditions for the linear regression model with *longevity* as the response and *Thorax* as the explanatory variable. The scatterplot below indicates that there is, indeed, a linear, positive relationship between the two variables.

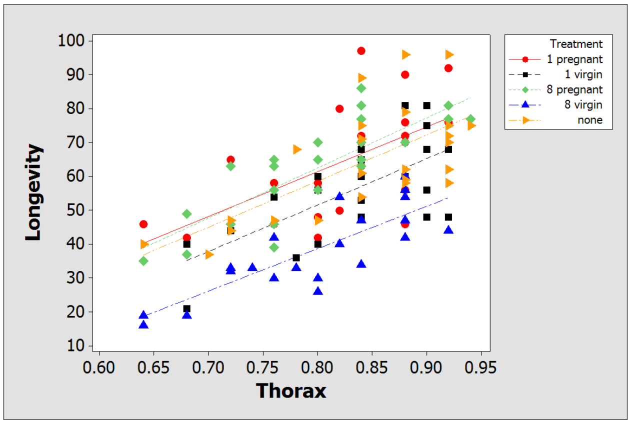


Next, we check the residuals from the regression with *longevity* as the response and *Thorax* as the explanatory variable. The following are the normal probability and residual versus fits plots. The normal probability plot is fine. There appears to be a little bit more variability for larger fitted values, but it is not enough to worry about.





Finally, we check the scatterplot of *Longevity* versus *Thorax* to see if the slopes appear to be about the same for each of the five treatment groups. The graph below shows that this condition has been met.



All of the conditions have been met. The ANOVA table for this analysis is given below.

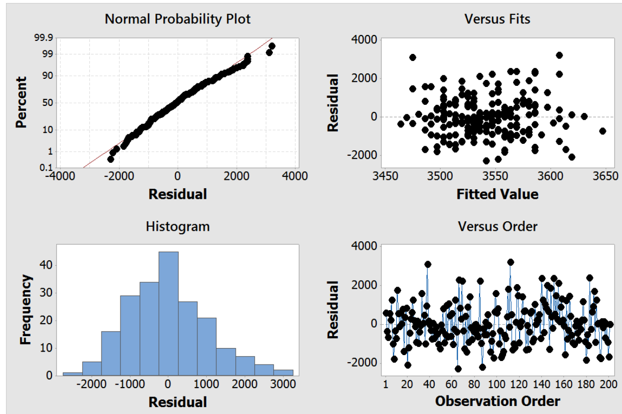
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Thorax	1	15496.6	13168.9	13168.9	119.22	0.000
Treatment	4	9611.5	9611.5	2402.9	21.75	0.000
Error	119	13144.7	13144.7	110.5		
Total	124	38252.8				

`S = 10.5100   R-Sq = 65.64%   R-Sq(adj) = 64.19%`

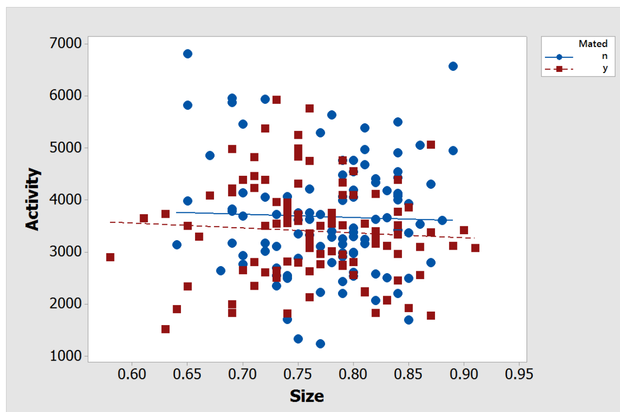
What we learn is that both the treatment and the size of the thorax are significant in explaining the longevity of male fruit flies such as these.

b. We prefer the ANCOVA model for this dataset. Back in Chapter 5, the ANOVA model only accounted for 28.9% of the variation in the fruit flies' life spans. In the model above, we have accounted for 64.19% of the variation. And both variables are highly significant.

8.60 a. In a previous exercise we verified that the ANOVA conditions were met for the one-way ANOVA model. Next we need to check the linear regression conditions between *Activity* and *Size*. The four plots below suggest that the normality, linearity, and constant variance conditions are all met.



Next we check to make sure that there is no interaction between Size and Mated. The scatterplot below shows two parallel lines, which means this condition is also met.



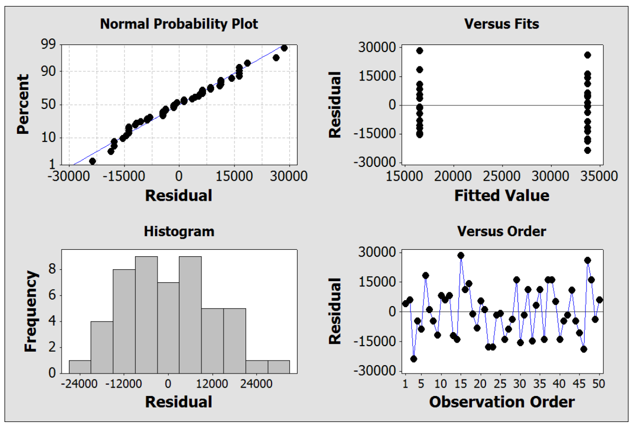
We have now shown that all conditions have been met, so we run the ANCOVA model. The output is given below.

```
Analysis of Variance for activity, using Adjusted SS for Tests

Source    DF     Seq SS     Adj SS   Adj MS     F      P
Size       1     243089     501847   501847  0.47  0.496
Mated      1    4192988    4192988  4192988  3.89  0.050
Error    198  213257939  213257939  1077060
Total    200  217694016


S = 1037.82   R-Sq = 2.04%   R-Sq(adj) = 1.05%
```

Mated is still the only significant variable and the significance is weak. The R-squared (adjusted) is merely 1.05%.

b. In this case neither model is really useful. None of the variables is strongly significant and the R-squared (adjusted) is just too small to indicate a useful model.

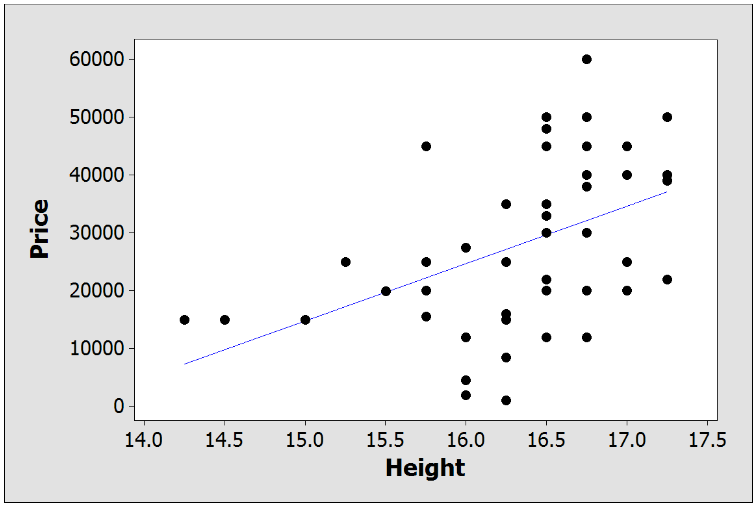
8.61 a. First, we check the conditions. We don't know how the horses were chosen other than that they came from Internet listings. We will have to assume that the horse prices are independent of each other. The residual plots that follow show that both the normality and constant variance conditions hold. All conditions are met.



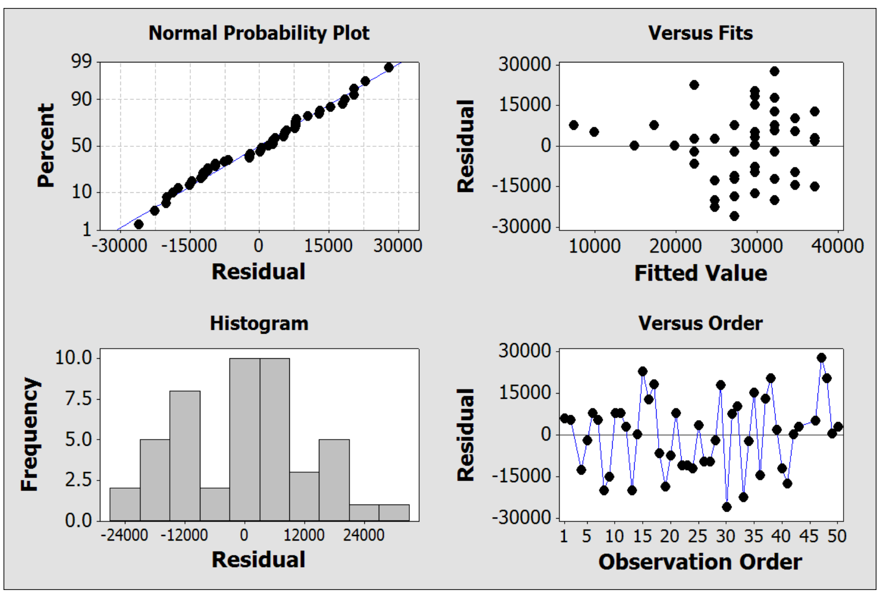
The ANOVA table given below shows that we have significant evidence of a difference in price between the two sexes of horses. Also, the ANOVA model accounts for 30.97% of the variability in selling price.

```
Source  DF           SS          MS      F      P
Sex      1   3560407500  3560407500  22.98  0.000
Error   48   7435532500   154906927
Total   49  10995940000

S = 12446   R-Sq = 32.38%   R-Sq(adj) = 30.97%
```

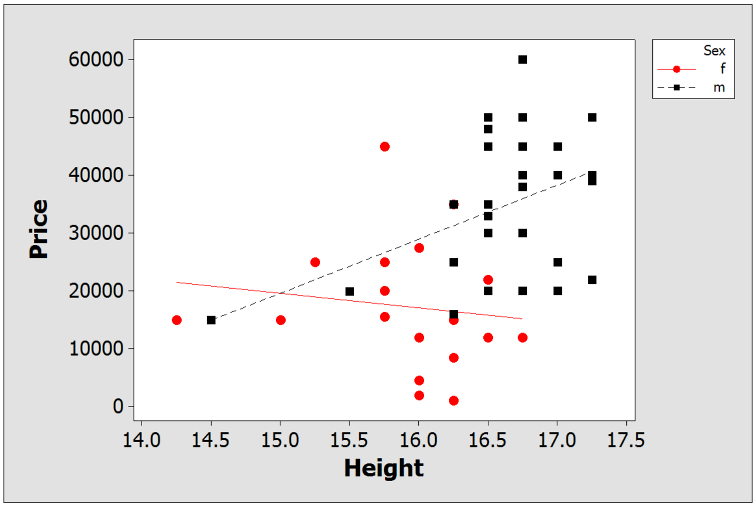
b. We now check the extra conditions necessary for ANCOVA. First, we look at a scatterplot of the data. The plot given below shows that there is a moderate positive relationship between *Height* and *Price* of the horses.



Next, we check the residual from the linear regression. The plots below show that the normal condition is met and, except for a few of the lowest priced horses, the constant variance condition is met.

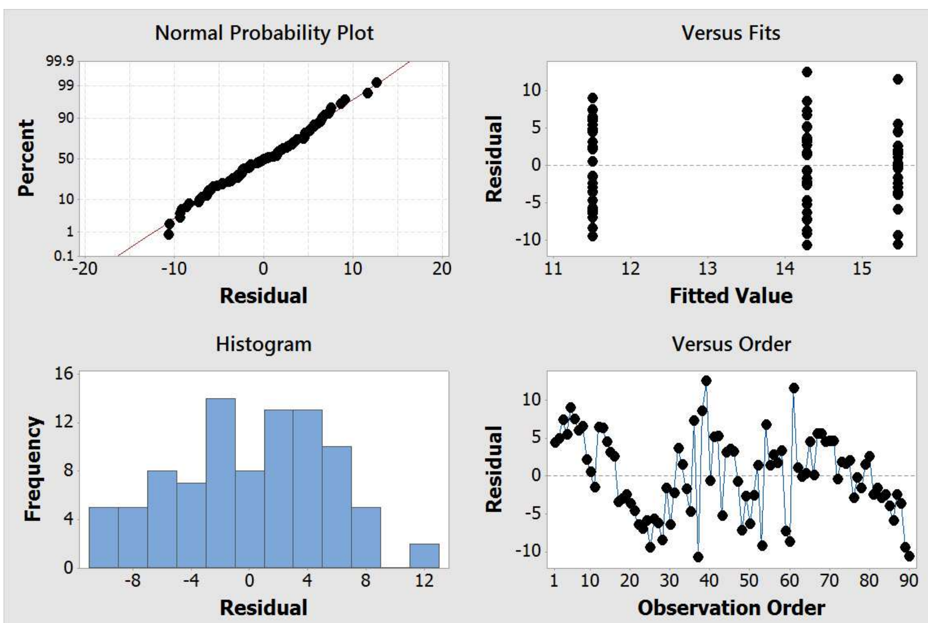


Finally, we check to see if the slopes for the regression lines for each sex are the same. The plot given below shows that they are not. In fact, for females there is a negative relationship between *Price* and *Height*, whereas there is a positive relationship for males. We should not use an ANCOVA model for this dataset.



c. We should use either the ANOVA model or a multiple regression model that fits two lines because the ANCOVA model is not appropriate for this dataset.

8.62 a. We start by checking conditions. First, we will hope that the prices set on the cars are independent. Since each car is different (age, mileage, etc.), this seems like a reasonable assumption. Next, we check the residual plots. The plots that follow suggest that the data meets both the normality and the constant variance conditions. So the conditions are met.

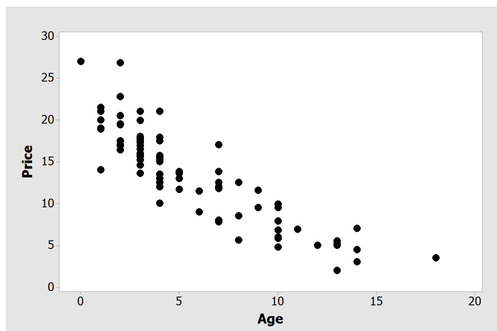


The ANOVA table below shows that there is a significant difference in price based on the model of car. This ANOVA model accounts for about 26% of the variation in the asking price of the cars.

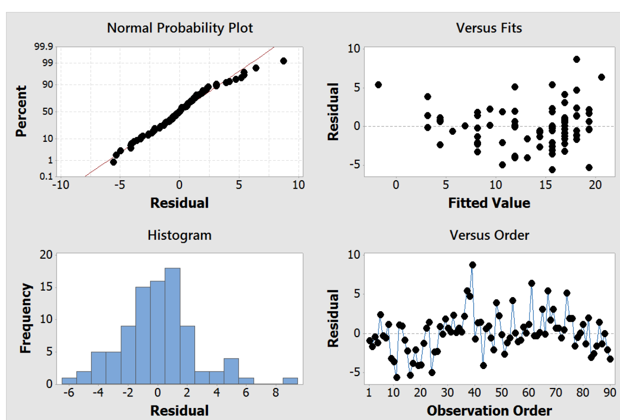
Source	DF	Adj SS	Adj MS	F-Value	P-Value
CarType	2	247.8	123.91	4.32	0.016
Error	87	2493.5	28.66		
Total	89	2741.3			

S	R-sq	R-sq(adj)	R-sq(pred)
5.35354	9.04%	6.95%	2.66%

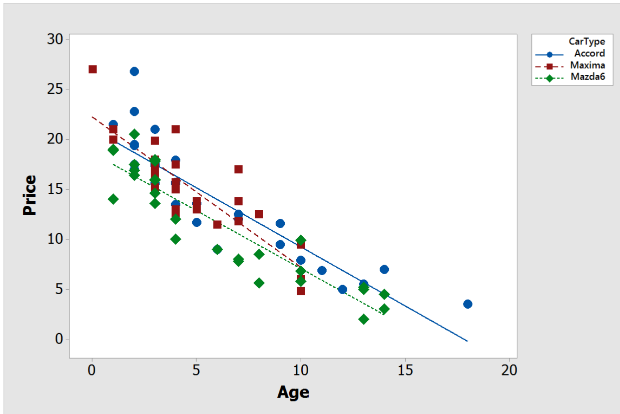
b. The conditions for ANOVA have been met. Now we check the conditions for regression and the interaction between age and manufacturer. We begin by looking at a scatterplot of *Price* versus *Age*. The plot given below shows that there is a negative relationship between the two, but it appears to have some curvature to it.



We will continue with caution and check the residual plots for the regression fit. These plots are given below. The normal probability plot shows a reasonable straight line pattern. The residual versus fitted values plot shows the same curvature that we saw in the original scatterplot, due mostly to some high priced (for their age) outliers.



Finally, we check to see if there is an interaction between the age of the car and the manufacturer of the car. The scatterplot below shows that there may be an interaction, with Porsche having a shallower slope than the other two.



All in all, we are hesitant to use the ANCOVA model. The actual ANOVA table from this model is given below.

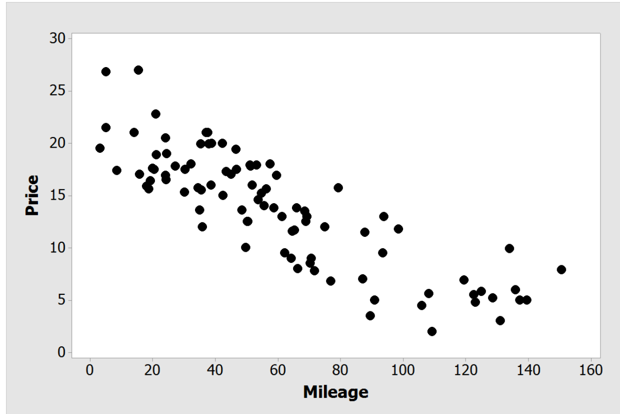
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Age	1	1990.46	1990.46	340.32	0.000
CarType	2	87.42	43.71	7.47	0.001
Error	86	503.00	5.85		
Total	89	2741.26			

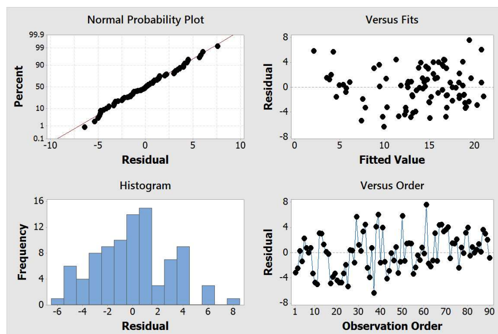
S	R-sq	R-sq(adj)	R-sq(pred)
2.41843	81.65%	81.01%	79.79%

The conclusion we would draw is that both *Age* and *CarType* are significant when predicting the price of a car. And this model does a better job than the ANOVA using *CarType* alone because $R^2(adj) = 81.01\%$, a value that is much higher than 6.95%. But we are reluctant to use this model given all of the issues with the conditions that we found.

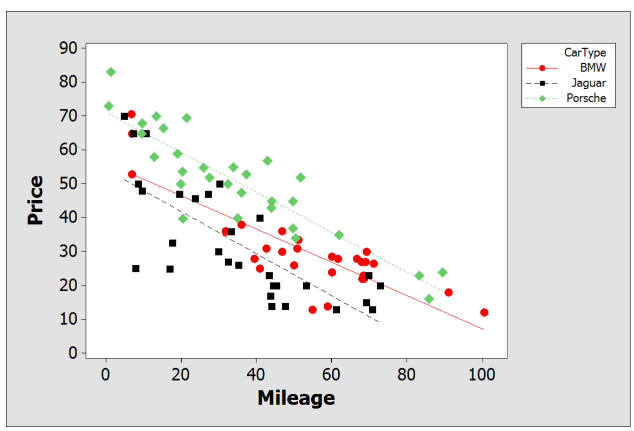
c. The conditions for ANOVA have been met. Now we check the conditions for regression and the interaction between mileage and manufacturer. We begin by looking at a scatterplot of *Price* versus *Mileage*. The plot given below shows that there is a negative relationship between the two, and it is much more linear than the relationship between *Price* and *Age*.



We will continue by checking the residual plots for the regression fit. These plots are given below. The normal probability plot shows that the normal condition is met. The residual versus fitted values plot shows the a couple of outliers, but in general the constant variance condition is met.



Finally, we check to see if there is an interaction between the mileage of the car and the manufacturer of the car. The scatterplot below shows that this condition is met. All conditions are now met for the ANCOVA model.



The actual ANOVA table from this model is given below.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Mileage	1	1812.94	1812.94	229.11	0.000
CarType	2	116.95	58.47	7.39	0.001
Error	86	680.51	7.91		
Total	89	2741.26			

S	R-sq	R-sq(adj)	R-sq(pred)
2.81299	75.18%	74.31%	72.63%

The conclusion we draw is that both *Mileage* and *CarType* are significant when predicting the price of a car. And this model does a better job than the ANOVA using *CarType* alone because $R^2(adj) = 74.31\%$, a value that is much higher than 6.95%.

d. We would use the ANCOVA model using *Mileage* as the covariate. This model meets all of the conditions and explains much of the variation in the asking price of the car.

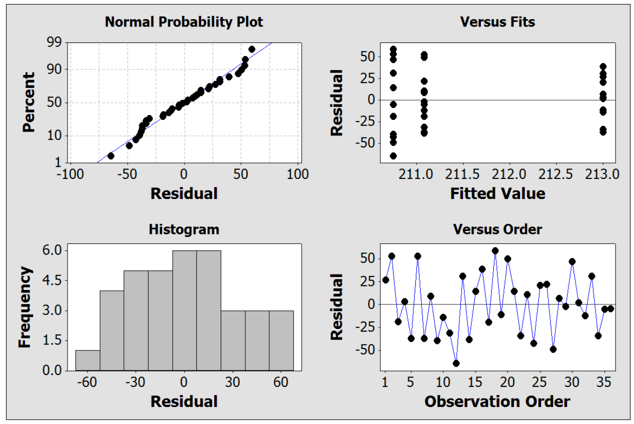
8.63 Here is some output for fitting the regression model (using *Oxygen*) as a quantitative predictor.

```
Response: Ethanol
          Df Sum Sq Mean Sq F value    Pr(>F)
Sugar      1 1806.2 1806.25  18.292 0.0009011 ***
Oxygen     1 1110.0 1110.05  11.242 0.0051924 **
Residuals 13 1283.7   98.75
```

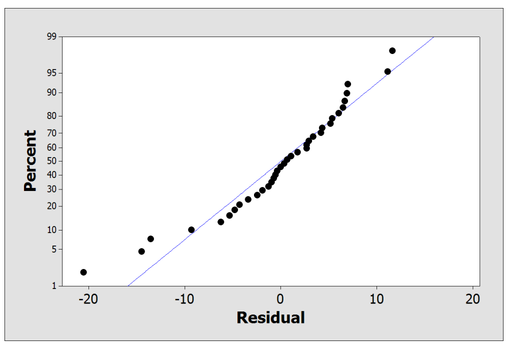
Here is some output for the two-way ANOVA with interaction (treating *Oxygen* as categorical).

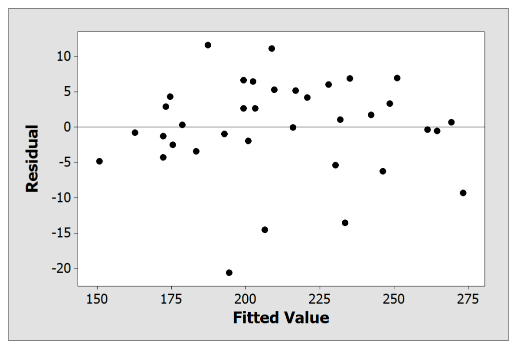
```
Response: Ethanol
                     Df  Sum Sq Mean Sq F value   Pr(>F)
Sugar                 1 1806.25 1806.25 13.2935 0.006532
factor(Oxygen)        3 1125.50  375.17  2.7611 0.111472
Sugar:factor(Oxygen)  3  181.25   60.42  0.4446 0.727690
Residuals             8 1087.00  135.87
```

8.64 a. First, we check conditions. We start with the conditions for ANOVA. Products were randomly assigned to the types of display, so the observations should be independent of each other. The residual plots given below suggest that both the normality and constant variation conditions are met.

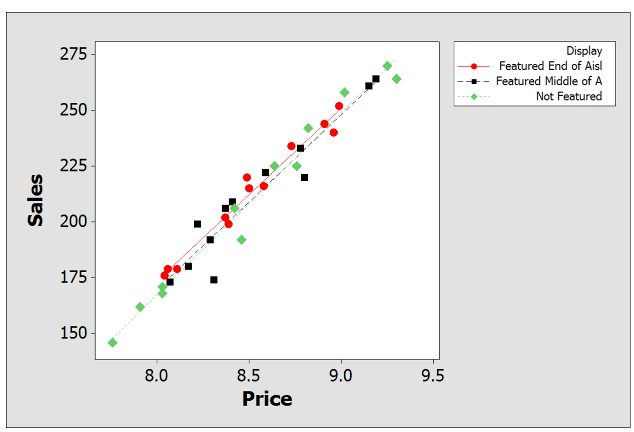


Next, we check the regression conditions. The scatterplot below shows that there is a strong, positive, linear relationship between *Price* and *Sales*. The normal probability plot shows a small amount of curvature, but not enough to worry about. The residuals versus fits graph shows some low outliers (points with large negative residuals).





Finally, the following scatterplot shows that there is no interaction between the type of display and the price of the products.

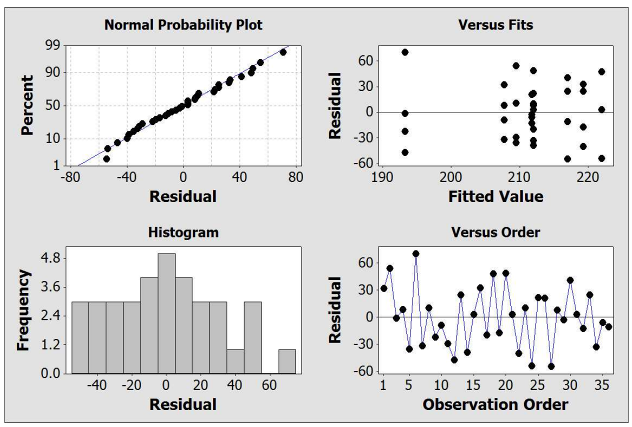


The ANOVA table is given below and it suggests that the type of display is not significant in determining the amount sales. Price is, however, very significant.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Price	1	36718	36796	36796	768.87	0.000
Display	2	113	113	56	1.18	0.321
Error	32	1531	1531	48		
Total	35	38363				

S = 6.91785   R-Sq = 96.01%   R-Sq(adj) = 95.63%

b. Again, we have independent observations. The residual plots are given below and show that the residuals meet the normal and constant variance conditions.

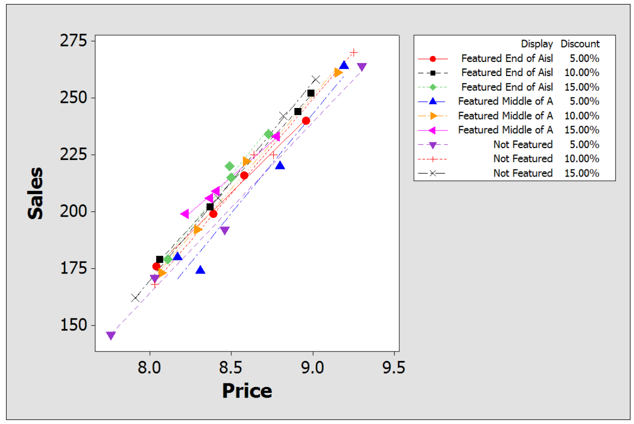


The ANOVA table shows that neither of the explanatory variables nor the interaction are significant.

Source	DF	SS	MS	F	P
Discount	2	1288.4	644.19	0.48	0.623
Display	2	35.4	17.69	0.01	0.987
Interaction	4	884.8	221.19	0.17	0.954
Error	27	36154.0	1339.04		
Total	35	38362.6			

S = 36.59   R-Sq = 5.76%   R-Sq(adj) = 0.00%

c. The conditions for the two-way ANOVA model were checked in part (b) and were found to have been met. The scatterplot below suggests that the slopes for all of the combinations of the two categorical variables are reasonably similar. The conditions are met for ANCOVA in this setting.



d. The ANOVA table for this ANCOVA model is given below. We conclude that, after taking *Price* into consideration, *Discount* is now significant. Both the *Display* and the interaction terms continue to be not significant. This model is also better in that it explains 97.73% of the variation, compared to essentially none of the variation in the two-way ANOVA model.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Price	1	36718.5	35505.7	35505.7	1423.91	0.000
Display	2	112.7	113.1	56.6	2.27	0.124
Discount	2	799.7	799.9	399.9	16.04	0.000
Display*Discount	4	83.4	83.4	20.9	0.84	0.514
Error	26	648.3	648.3	24.9		
Total	35	38362.6				

S = 4.99353 R-Sq = 98.31% R-Sq(adj) = 97.73%

Since the interaction and *Display* were not significant, we removed them and came to our final model. This is a simple model (only two explanatory variables), yet it has an adjusted R^2 of 0.9759.

Analysis of Variance for Sales, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Price	1	36718	36230	36230	1372.84	0.000
Discount	2	800	800	400	15.15	0.000
Error	32	844	844	26		
Total	35	38363				

S = 5.13714 R-Sq = 97.80% R-Sq(adj) = 97.59%

Chapter 8 Online Solutions

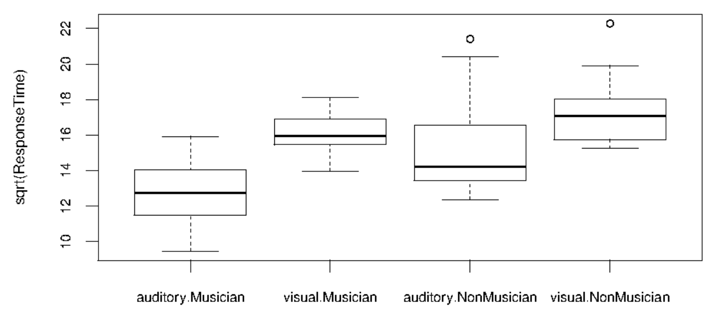
8.65 There is one within-subjects variable: *time* (week0, week2, week4, and week8). There is one between-subjects variable: *group* (drug or placebo).

8.66 There is one within-subjects variable: *eye* (right and left eye measured for each subject). There are two between-subjects variables: *age* (young or old) and *sex* (male or female).

8.67 There is one within-subjects variable: *time* (before and after). There are two between-subjects variables: *sex* (male or female) and *medicine* (drug or placebo).

8.68 There are two within-subjects variables: *drug form* (tablet and capsule) and *day* (each subject was measured on day 1 and day 2). There is one between-subjects variable: *protocol* (A or B).

8.69 a. Here are side-by-side boxplots for $sqrt(ResponseTime)$.



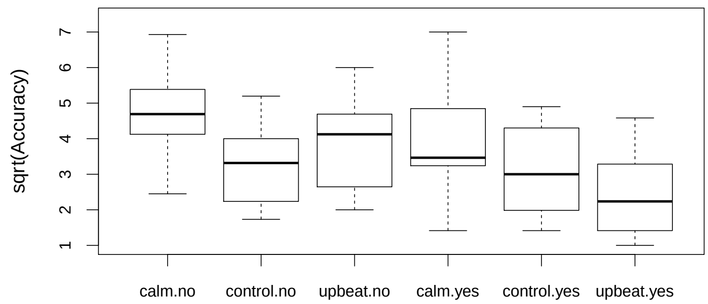
Here are some summary statistics to compare $sqrt(ResponseTime)$ across the four combinations of *Group* and *Stimulus*.

Stimulus:Group	Mean	Std. deviation
auditory:Musician	12.70	2.19
visual:Musician	16.13	1.16
auditory:NonMusician	15.40	2.69
visual:NonMusician	17.35	1.81

Response times tend to be lower for the auditory stimulus and also lower for musicians.

b. The boxplots hint at a possible interaction, but the graphical evidence is not clear. The $visual - auditory$ difference is a bit greater for musicians than nonmusicians. Responses are faster for auditory stimuli than for visual stimuli and this difference is more pronounced among musicians than among nonmusicians.

8.70 a. Here are side-by-side boxplots for *sqrt*(*Accuracy*).



Here are some summary statistics to compare *sqrt*(*Accuracy*) across the six combinations of *Music* and *MusicBg*.

Music:MusicBg	Mean	Std. deviation
calm:no	4.86	1.60
control:no	3.27	1.23
upbeat:no	3.91	1.26
calm:yes	3.88	1.59
control:yes	3.15	1.34
upbeat:yes	2.48	1.25

Accuracy tends to be worst for the calm condition. Musicians tend to have better accuracy than nonmusicians in each music condition.

b. The best combination (lowest average accuracy score) is for upbeat, yes.

8.71 Here is some computer output for running the repeated measures ANOVA.

```
Error: Subject
          Df  Sum Sq Mean Sq F value Pr(>F)
Group    1    64.1   64.1    9.45 0.0041  **
Residuals 34  230.7     6.8

Error: Subject:Stimulus
               Df Sum Sq  Mean Sq  F value   Pr(>F)
Stimulus        1  111.1    111.1     50.98  3e-08 ***
Stimulus:Group  1    9.1      9.1      4.19  0.048 *
Residuals      34   74.1      2.2
```

a. There is very strong evidence (P-value=3×10^{-8}) for a difference between the two types of *Stimulus*.

b. There is very strong evidence (P-value=0.0041) of a *Group* effect, that is, a difference between musicians and nonmusicians.

c. There is some evidence (P-value=0.048) of interaction between *Stimulus* and *Group*.

8.72 Here is some computer output for running the repeated measures ANOVA.

```
Error: Subject
          Df  Sum Sq Mean Sq F value Pr(>F)
MusicBg    1   10.5     10.5   3.37  0.083
Residuals 18   56.2      3.12

Error: Subject:Music
              Df Sum Sq  Mean Sq  F value   Pr(>F)
Music          2   18.0      9.02     6.74   0.0033 **
Music:MusicBg  2    4.4       2.2     1.65   0.2068
Residuals     36   48.1       1.34
```

a. There is very strong evidence (P-value $= 0.0033$) for a difference among the three *Music* conditions.

b. There is only weak evidence (P-value $= 0.083$) of a *MusicBg* effect.

c. There is no convincing evidence (P-value $= 0.2068$) of interaction between *Music* and *MusicBg*.

8.73 a. Here is the repeated measures ANOVA for the original data.

```
Error: Day
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 25  103.6   4.145

Error: Day:Time
          Df Sum Sq Mean Sq F value   Pr(>F)
Time       3  192.4   64.14   15.12 8.63e-08
Residuals 75  318.1    4.24
```

The relevant F-statistic for comparison to randomization samples is $F = 15.12$.

b. In one set of 5000 randomizations, the largest F-statistic for the *Pulse* factor was $F = 7.67$. Thus none of the simulated values were larger than the $F = 15.12$ from the original sample, so the randomization P-value ≈ 0.

c. This very small P-value gives strong evidence that the average pulse rate changes between different time periods in the day.

8.74 a. Here is the repeated measures ANOVA for the original data.

```
Error: Subject
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 19   66.7   3.511

Error: Subject:Music
          Df Sum Sq Mean Sq F value  Pr(>F)
Music      2  18.03   9.016    6.52 0.00368 **
Residuals 38  52.55   1.383
```

The relevant F-statistic for comparison to randomization samples is $F = 6.52$.

b. In one set of randomizations, we find 20 of the 5000 simulations give an F-statistic as large as $F = 6.52$ from the original sample. This produces a randomization P-value $= 20/5000 = 0.004$. This is close to the P-value from the F-test in the ANOVA table. The randomization P-value will vary slightly for other sets of 5000 randomizations.

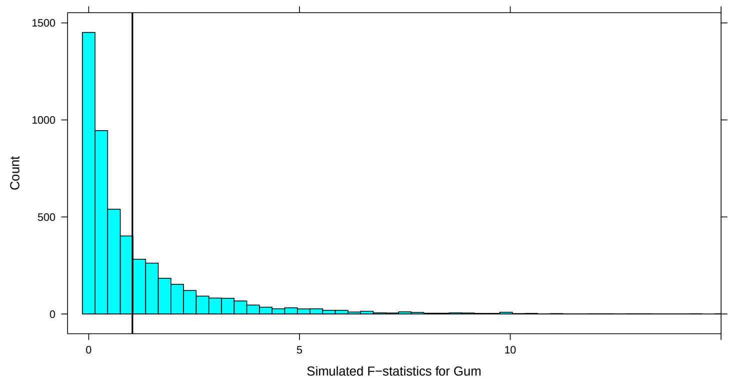
c. This very small P-value gives strong evidence that the average accuracy at timing 45 seconds is different depending on the type of music playing.

8.75 a. Each of the 12 subjects (A–L) have four values, the calories digested for crackers with each combination of *Bran* and *Gum*. To create randomization samples, we can scramble each subject's set of four values and reassign them to the four types of crackers for that subject.

b. To assess the main effect for *Gum*, run the repeated measures ANOVA for each randomization sample and record the F-statistic for the *Gum* main effect

c. To get a randomization P-value for the *Gum* main effect, find the proportion of randomization samples that give an F-statistic (as in part b) that are greater than or equal to the $F = 1.037$ that was observed for the original sample.

d. Here is a histogram of *Gum* F-statistics for 5000 randomization samples.



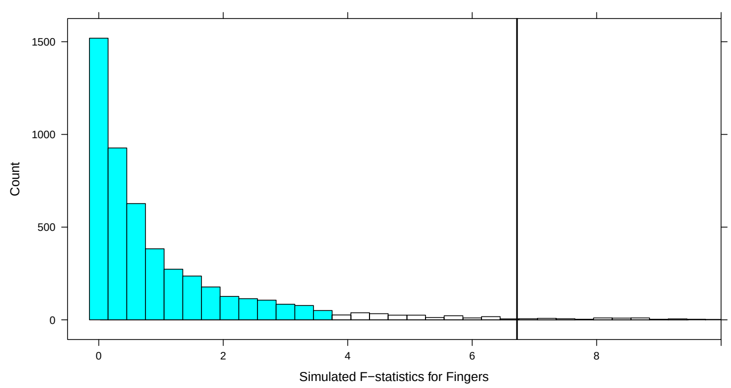
1693 of these values are at or beyond $F = 1.307$, so the randomization P-value is $1693/5000 = 0.3386$. This is relatively close to the P-value (0.33) obtained using the F-distribution in the original ANOVA. Answers will vary slightly for other sets of 5000 randomizations.

8.76 a. Each of the 20 participants (p1–p20) have four values, the $log(Time)$ to complete the task with each combination of *Fingers* and *Objects*. To create randomization samples we can scramble each subject's set of four values and reassign them to the four treatments.

b. To assess the main effect for *Fingers*, run the repeated measures ANOVA for each randomization sample and record the F-statistic for the *Fingers* main effect

c. To get a randomization P-value for the *Fingers* main effect, find the proportion of randomization samples that give an F-statistic (as in part b) that are greater than or equal to the $F = 6.72$ that was observed for the original sample.

d. Here is a histogram of *Fingers* F-statistics for 5000 randomization samples.



In these randomizations, 88 of the values are at or beyond $F = 6.72$, so the randomization P-value is $88/5000 = 0.0176$. This is very close to the P-value (0.018) obtained using the F-distribution in the original ANOVA. Answers will vary slightly for other sets of 5000 randomizations.

8.77 a. First for each *Subject*, randomly scramble values of $sqrt(ResponseTime)$ for the auditory and visual *Stimulus* conditions. Then randomly assign 18 of the subjects to the musicians groups and leave the other 18 as nonmusicians.

b. Here is the repeated measures ANOVA for the original data.

```
Error: Subject
          Df  Sum Sq Mean Sq F value Pr(>F)
Group    1    64.1   64.1    9.45 0.0041  **
Residuals 34  230.7     6.8

Error: Subject:Stimulus
               Df Sum Sq  Mean Sq  F value   Pr(>F)
Stimulus        1  111.1    111.1     50.98  3e-08 ***
Stimulus:Group  1    9.1      9.1      4.19  0.048 *
Residuals      34   74.1      2.2
```

The relevant F-statistic for the interaction term to compare to randomization samples is $F = 4.19$.

c. In one set of randomizations, we find that 238 of the 5000 simulations give a F-statistic for the interaction that is at least as big as the $F = 4.19$ value from the original sample. This produces a randomization P-value $= 238/5000 = 0.0476$. This is close to the P-value from the F-test in the ANOVA table. The randomization P-value will vary slightly for other sets of 5000 randomizations.

d. This is a relatively small P-value, so we have somewhat convincing evidence that there is interaction between *Stimulus* and *Group* in this situation.

8.78 a. First for each *Subject*, randomly scramble the three $sqrt(Accuracy)$ values and assign to the calm, control, and upbeat *Music* conditions. Then randomly assign 10 of the subjects to have a music background and leave the other 10 to have no music background, i.e., randomly scramble the values of *MusicBg* among the 20 subjects.

b. Here is the repeated measures ANOVA for the original data.

```
Error: Subject
          Df  Sum Sq Mean Sq F value Pr(>F)
MusicBg    1   10.5     10.5   3.37  0.083
Residuals 18   56.2      3.12
```

```
Error: Subject:Music
              Df  Sum Sq  Mean Sq  F value   Pr(>F)
Music          2   18.0      9.02     6.74   0.0033 **
Music:MusicBg  2    4.4       2.2     1.65   0.2068
Residuals     36   48.1       1.34
```

The relevant F-statistic for the interaction term to compare to randomization samples is $F = 1.65$.

c. In one set of randomizations, we find that 1020 of the 5000 simulations give an F-statistic for the interaction that is at least as big as the $F = 1.65$ value from the original sample. This produces a randomization P-value $= 1020/5000 = 0.2040$. This is close to the P-value from the F-test in the ANOVA table. The randomization P-value will vary slightly for other sets of 5000 randomizations.

d. This is not a small P-value, so we have no convincing evidence that there is interaction between *Music* and *MusicBg* in this situation.