Chapter 7 Solutions

7.1   a. The degrees of freedom for each factor is the number of levels for that factor minus 1. In this case, that is $4 - 1 = 3$.

    b. The degrees of freedom for each factor is the number of levels for that factor minus 1. In this case, that is $5 - 1 = 4$.

    c. For a model with more than one observation per cell, the degrees of freedom for the interaction term is $(K - 1)(J - 1)$, which in this case is $(4 - 1)(5 - 1) = 12$.

    d. For a model with more than one observation per cell, the degrees of freedom for the residual term is $KJ(c - 1)$, which in this case is $4(5)(3 - 1) = 40$.

7.2   a. The degrees of freedom for each factor is the number of levels for that factor minus 1. In this case, that is $3 - 1 = 2$.

    b. The degrees of freedom for each factor is the number of levels for that factor minus 1. In this case, that is $3 - 1 = 2$.

    c. For a model with more than one observation per cell, the degrees of freedom for the interaction term is $(K - 1)(J - 1)$, which in this case is $(3 - 1)(3 - 1) = 4$.

    d. For a model with more than one observation per cell, the degrees of freedom for the residual term is $KJ(c - 1)$, which in this case is $3(3)(3 - 1) = 18$.

7.3 c. The lines on the plot will be approximately parallel. They will, however, not necessarily be the same.

7.4 a. The lines on the plot will be nonparallel.

7.5   a. large

    b. less

7.6   a. we cannot tell

7.7 Yes. Both Republicans and Democrats score higher on Soul than Heart, but the difference between the two is greater for the Democrats than for the Republicans. An interaction plot is given below that illustrates this relationship.
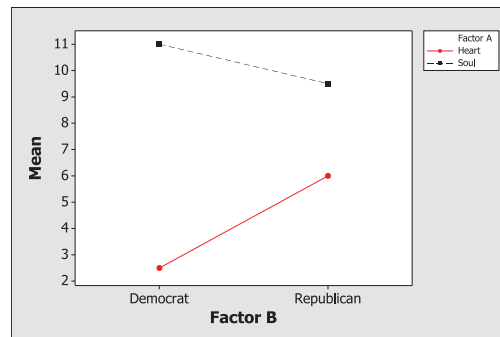
7.8 No. For both males and females, the average scores for sweat and tears are the same and the average score for blood is smaller. There is also the same amount of difference between the average scores for blood and sweat for both sexes. An interaction plot is given below that illustrates this relationship.
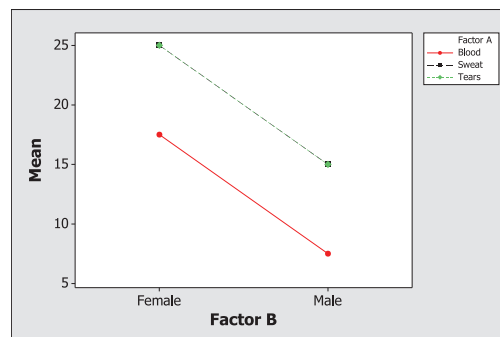
7.9 True. If interaction is present, the effect of one main effect will be different for different levels of another main effect.

7.10 True. The errors must still be independent, normally distributed, have mean 0, and have the same variance.

7.11 The difference between Yes and No for A is 10 when B is "No" and 10 when B is "Yes." Those two differences are the same, so no interaction is present.

7.12 The difference between Yes and No for A is 12 when B is "No" and 4 when B is "Yes." Those two differences differ, so an interaction is present.

7.13   a. The log transformation shrinks all values greater than 1 back toward zero. So all of the means and all of the magnitudes of deviations from the means will get smaller. Thus all SSs will be smaller for the log transformed data.

   b. The degrees of freedom depend on the number of groups and sample size within each cell, but not the scale chosen for the response variable. So dfs will be the same as for the original data.

   c. Put parts (a) and (b) together. SSs get smaller, dfs don't change, and $MS = SS/df$ so the MSs get smaller after the log transformation.

   d. The mean squares for each main effect and the interaction all get smaller, but so does the MSE. We can't tell the direction of the change when computing each of the F-ratios after dividing by the MSE. We would need to do both ANOVAs.

   e. Since the dfs don't change with the log transformation and we can't tell which way the F-ratios go (part (d)), we also can't tell how the P-values might change. Again, we would need to do both ANOVAs.

7.14 We find each SS by summing the squares in the box for that term. The magnitudes are clearly smallest for the residuals, slightly larger for the interaction, bigger still for Factor B, and the largest for Factor A. Thus we clearly have $SSE < SSAB < SSB < SSA$. Each of the main effects and the interaction have just 1 df, so $SS = MS$ for those terms. The residuals will have 4 df (sample size is 8 and we lose 1 df for the grand average and each of the other terms). $MSE = SSE/4$ will make the smallest SS even smaller as an MS. Thus we have $MSE < MSAB < MSB < MSA$.

7.15 It is clear that summing squares gives $SSB < SSA < SSAB$, and each has 1 df, so we must have the same order for $MSB < MSA < MSAB$. Where does MSE fit? SSE is given as 180 (saving the time to sum the squares yourself) and it has 4 df (since sample size is 8 and we've already lost one df for the grand average and each of the three other terms). Thus we have $MSE = 180/4 = 45$. This will fit between $SSB = MSB = 8\left(2^2\right) = 32$ and $SSA = MSA = 8\left(4^2\right) = 128$. Thus we have $MSB < MSE < MSA < MSAB$.

7.16 It is clear that summing squares gives $SSAB < SSA < SSB$, and each has 1 df, so we must have the same order for $MSAB < MSA < MSB$. Where does MSE fit? SSE is given as 60 (saving the time to sum the squares yourself) and it has 4 df (since sample size is 8 and we've already lost one df for the grand average and each of the three other terms). Thus we have $MSE = 60/4 = 15$. This will fit between $SSAB = MSAB = 8\left(1^2\right) = 8$ and $SSA = MSA = 8\left(2^2\right) = 32$. Thus we have $MSAB < MSE < MSA < MSB$.

7.17 The comparison between #1 and #3 is easy since both have $SSE = 60$ and thus the same $MSE_1 = MSE_3 = 60/4 = 15$. So the comparative size of the F-ratios will depend only on the relative magnitude of the effects, which are larger for all three terms in decomposition #1 than for decomposition #3.

So where does decomposition #2 fit? Its $SSE = 180$ is three times larger than #1 or #3, as is the $MSE_2 = 180/4 = 45$. For it to have a bigger F-ratio, the corresponding squared effect would need to be more than three times bigger than the squared effect in the other decomposition. No chance for that happening with the main effects of decomposition #1, which are actually bigger in magnitude than #2 (10 compared to 4 for A and 5 compared to 2 for B). The squared AB interaction effect for #2, $5^2 = 25$, is larger than for #1, $4^2=16$, but not 3 times larger. So the F-ratios for all three terms will be larger for decomposition #1 than for decomposition #2. When

comparing squared effects for #2 to #3, we see that #2 is more than three times larger for Factor A ($4^2 = 16$ compared to $2^2 = 4$) and for the AB interaction ($5^2 = 25$ compared to $1^2 = 1$), but not for Factor B (which is actually smaller for #2 than #3).

To summarize the orders of the F-ratios:
Factor A: #3 < #2 < #1
Factor B: #2 < #3 < #1
Interaction AB: #3 < #2 < #1

7.18 Comparison, blocking (the sections of the garden are blocks), crossing (each combination of fertilizer and water was used), randomization (via the coin tossing). There was no replication: Each of the four conditions (combinations of fertilizer and water) was measured only once within a block.

7.19 Comparison, replication (each combination of fertilizer and water level was used 10 times), crossing (each combination of fertilizer and water was used), randomization (via drawing numbers 1–4). There was no blocking.

7.20 For each species, the control data tend to be higher than the flooded data. For each condition, the river species data tend to be higher than the European species data.

7.21 When Fe is high, Zn high is better than Zn low. But when Fe is low, Zn high is worse than Zn low. The best combination is Fe high and Zn high.

7.22 The lines are close to parallel, so interaction is not a worry here.

7.23 The lines are not parallel, so there is evidence of interaction.

7.24 $1 - \text{slope} = 0.986$, almost exactly one. There is no need to transform.

7.25 $1 - \text{slope} = 1.099$, almost exactly one. There is no need to transform.

7.26   a. If each subject is a unit, we randomly assign five men and five women to each machine. The design is a two-way factorial. Sex is an observational factor with two levels. Machine is an experimental factor, also with two levels.

   b. For the block design, each subject uses both machines, in an order chosen randomly, with a separate randomization for each subject. This design has three factors: the same two factors of interest as before, plus a nuisance factor, subject.

   c. Each subject is either male or female. Sex goes with the larger units. Machine (treadmill or rower) is assigned to time slots, the smaller units.

7.27   a. There are two factors of interest, one observational factor, sex of server, and one exper-imental factor, Face (yes/no). For Face, the experimental unit is a table: Each table got a check either with or without a happy face. For Sex of Server, the observational unit is the server, and *there is only one unit of each kind.* A two-way factorial design would have several servers of each sex.

b. There is one df for Male/Female. Because there is only one server of each sex, the Male/Female difference and the difference between servers are one and the same: There is no way to tell how much of any observed difference is due to each source. For interaction, $df = 0$.

c. The design *should* have used more than one server of each sex. The data from the actual design do not allow conclusions about male/female differences or about possible interaction. (The only possible conclusions would be based on a pair of two-sample t-tests, one for each server.)

d. For each of the six servers, randomly assign "face" or "no face" to a sequence of tables. The resulting design has units of two sizes, servers (large unit) and tables (small unit), with each larger unit corresponding to a block of smaller units.

7.28 Following is some output for fitting the two-way main effects only model to the **SugarEthanol** data.

```
Response: Ethanol
               Df Sum Sq Mean Sq F value    Pr(>F)
Sugar           1 1806.2 1806.25  15.666 0.002241
factor(Oxygen)  3 1125.5  375.17   3.254 0.063576
Residuals      11 1268.2  115.30
```

Here is some output when the interaction term is included.

```
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
```

We see that the P-values for both main effects are smaller when the interaction term is dropped to give main effects only. A primary reason for this is the drop in the MSE (from 135.87 to 115.30), which occurred even though the SSE went up by 181.25 (the amount of variability explained by the interaction term in the bigger model). But moving the 3 df for the interaction to increase the error df from 8 to 11 was more than enough to compensate for the relatively small increase in SSE.

The overall conclusion is not much different. We still have strong evidence of a difference due to the type of sugar and only marginal evidence (but now significant at a 10% level) for a difference due to oxygen concentration.

7.29   a. The study could have been done as a two-way factorial with one observational factor, Diagnosis (ADHD or control), and one experimental factor, Noise (present or absent). Half of the ADHD subjects would be assigned at random to the noise condition, the other half to no noise. The control subjects would be assigned similarly, half to noise, half to no noise.

   b. Yes, because the two factors of interest are crossed, the design makes it possible to detect interaction.

   c. The block design makes it easier to detect interaction because each subject is tested under both conditions, noise and no noise, instead of just one or the other. For the factorial design (no blocks), difference between subjects are part of the residual error. For the block design, these differences between subjects are split off from residual error. (For more on this, see Example 8.15.)

7.30   a. The plots show that the average time to burn 200 calories for each machine depends on sex.
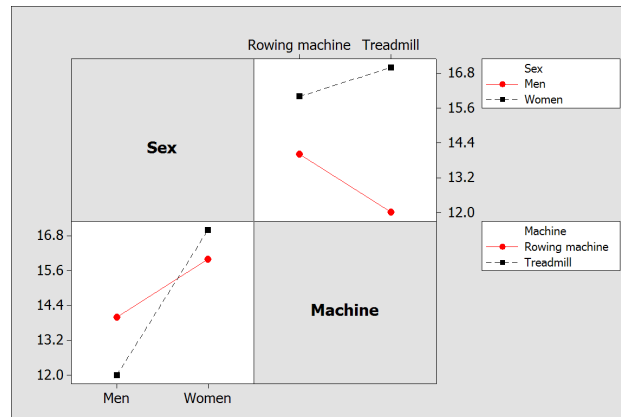
   b. No, it is not appropriate to use a two-way ANOVA model in this case, because the subjects themselves become another factor.

   c. Yes, it is appropriate to use a two-way ANOVA model in this case because we have two and only two factors (sex and type of machine) and all possible combinations of the levels of these two factors are observed.

7.31 There is an interaction. The percent of males and females who report having been drunk at least twice is the same for the Northern countries. But in the Eastern and Continental countries, there is a difference. In both the Eastern and Continental countries, a smaller percentage of males than females report having been drunk at least twice. This difference is most pronounced in the Eastern countries.

7.32   a. The two factors are sex, which is observational (we can't assign this) and has two levels, and happy face, which is experimental and also has two levels
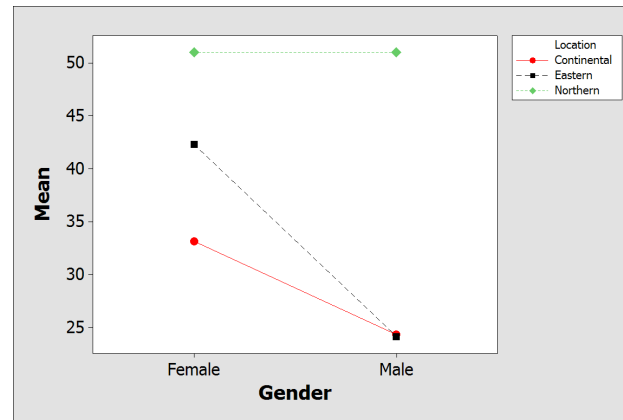
b. There is an interaction present. For the male, the average percentage tip was higher without the happy face, while for the female, the average percentage tip was higher with the happy face.
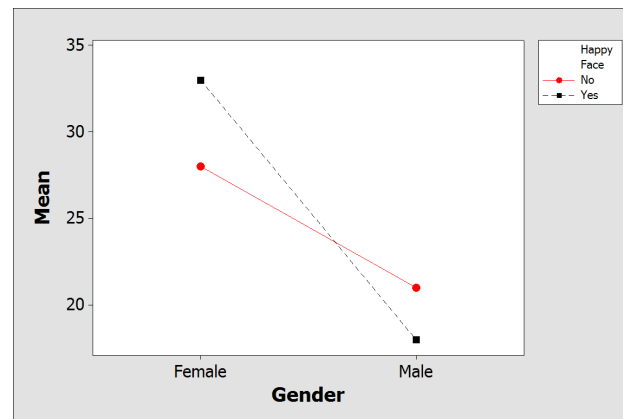

7.33 The table follows. The df for both Face and Gender is 1 since each factor has two levels. The interaction df is the df for Face multiplied by the df for Gender. Finally, the df for Residuals is $KJ(c-1) = 2(2)(24) = 96$. Next, we compute the SS. Since both Face, Gender, and the Interaction have 1 df, the SS and the MS are the same within each line of the table. This gives SS for Gender of 2,500 and MS for Interaction of 400. Knowing that the MS for residual is 100 and the df is 96, we find that the SS is 9,600 $(SS = MS \cdot df)$. Finally, we get the SS for Face by subtracting the SS for Gender, Interaction, and Residuals from the Total SS. This gives $25,415 - 2,500 - 400 - 9,600 = 12,915$. To get the F-values for Face, Gender, and Interaction, we divide the appropriate MS by the MS residuals. Using 1 and 96 degrees of freedom, we find that the P-values for Face and Gender are approximately 0 and the P-value for the interaction is 0.048. All three are significant.

Source	df	SS	MS	F
Face (Yes/No)	1	12,915	12,915	129.15
Gender (M/F)	1	2,500	2,500	25.0
Interaction	1	400	400	4.0
Residuals	96	9,600	100	
Total	99	25,415		

7.34   a. The ANOVA table is given below:

```
Source        DF        SS        MS        F        P
Province       1   32275.2   32275.2   349.82   0.000
Sex            1   11020.8   11020.8   119.45   0.000
Interaction    1    5320.0    5320.0    57.66   0.000
Error        476   43916.7      92.3
Total        479   92532.8
```

The table suggests that both province and sex are significant as is the interaction. Looking at the interaction plot that follows we see that the average scores in Nairobi are lower than the average scores in Pwani, men have lower scores than women within a province, and the difference between the men and women is more pronounced in Nairobi.
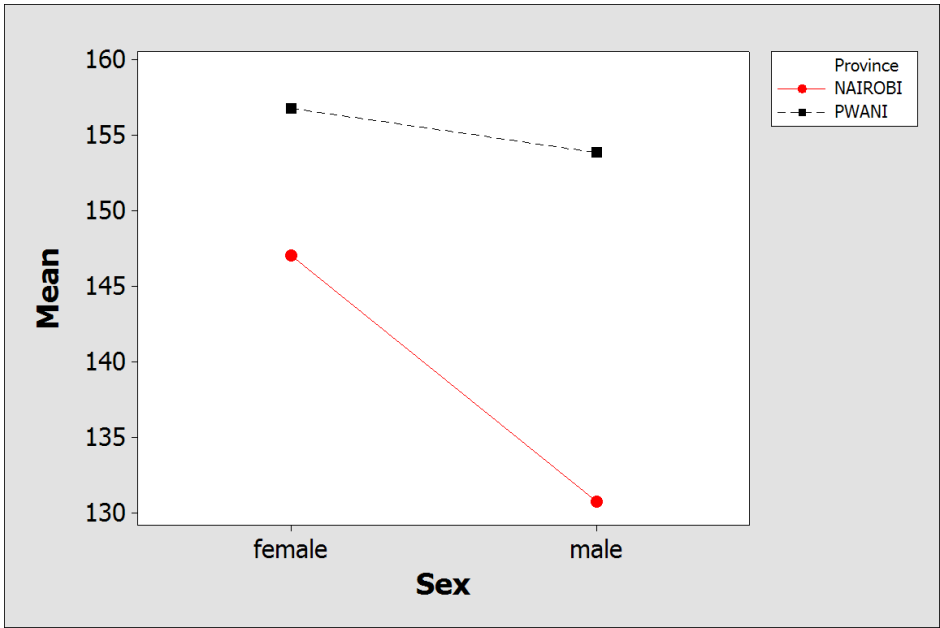
b. Yes, both look reasonable. The normal plot of the residuals and the residuals versus fits graphs are given below. The standard deviations for the four groups are

```
       Results for Province = NAIROBI

       Variable          Sex      StDev
       Attitude Score    female    8.06
                         male     13.97


       Results for Province = PWANI

       Variable          Sex      StDev
       Attitude Score    female    6.44
                         male      8.22
```
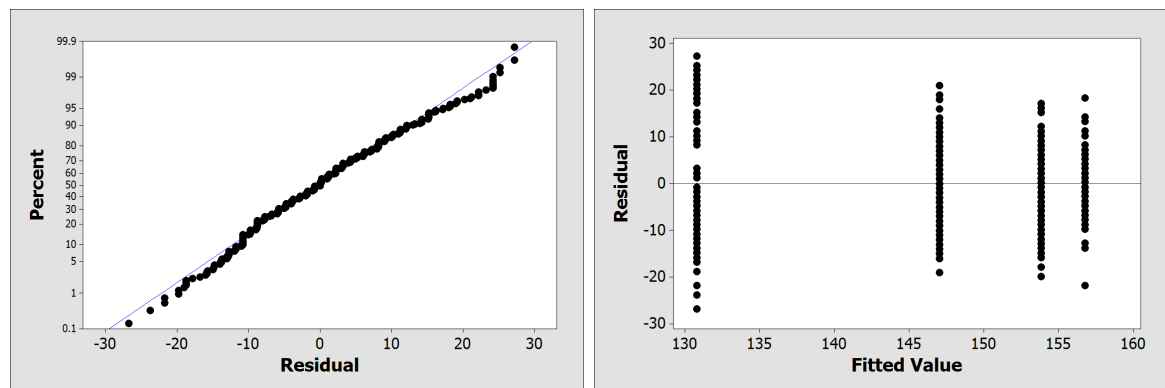
The smallest standard deviation is 6.44 and the largest is 13.97, which is just a little larger than our rule of thumb of being less than twice the size of the smallest standard deviation.


c. A balanced design means that there are an equal number of observations for each combination of levels of the two factors. The problem here is that each school is in only one of the provinces. If School A is in Nairobi, then all observations coming from School A will be observations from Nairobi. We will have no observations that are both from School A and from Pwani.

7.35   a. The slope appears to be approximately 0.5. This means that $p = 1 - 0.5 = 0.5$, which suggests that the appropriate transformation is the square root.

b. The slope appears to be approximately 0.8. This means that $p = 1 - 0.8 = 0.2$, which suggests that the appropriate transformation is the fifth root.

7.36 The slope appears to be approximately 0.6. This means that $p = 1 - 0.6 = 0.4$. This suggests trying either the square root or the third root.

**7.37**   a. The dotplots are given as follows:
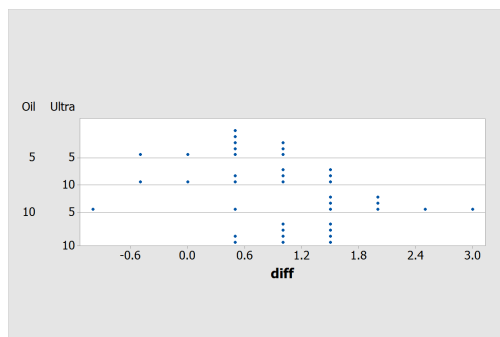
In this set of dotplots, it appears that three groups are all quite similar (both groups with 5 ml of oil and the group with 10 ml of oil and 10 minutes of ultrasound). However, the other group (that with 10 ml of oil and 5 minutes of ultrasound) appears to be different. More specifically, it appears that there is a bigger difference between treated and control samples under those conditions. This suggests that perhaps neither *Oil*, nor *Ultra* will be significant by themselves, but the interaction between them might be.

b. The interaction plot is given below:
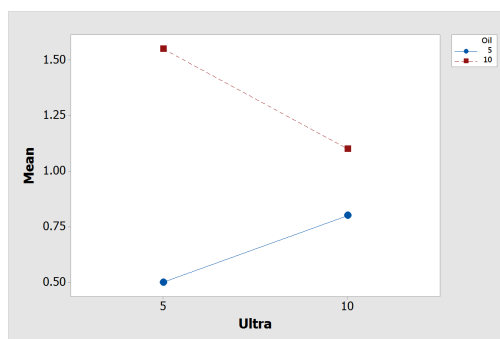
This plot does suggest that there might be an interaction. That is, it might be that for samples with lesser amounts of oil, longer ultrasound exposure is better, while for samples with more oil, less exposure to ultrasound is better.

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.
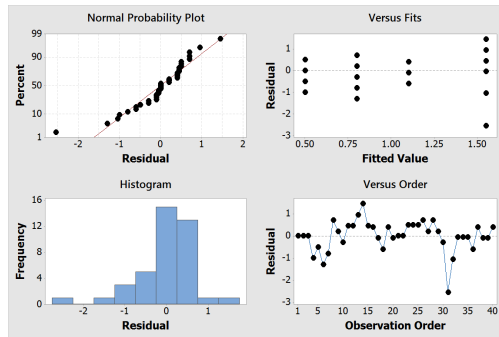
The ANOVA output is given below:

```
Analysis of Variance
```

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Ultra	1	0.0563	0.05625	0.11	0.744
Oil	1	4.5562	4.55625	8.76	0.005
Ultra*Oil	1	1.4062	1.40625	2.70	0.109
Error	36	18.7250	0.52014		
Total	39	24.7437			

The conclusions we can draw from this ANOVA table are that, as expected, the amount of time exposed to ultrasound is not significant. Somewhat surprisingly, neither is the two-way interaction. However, the main effect for the amount of oil in the sample is significant. The difference in how much oil gets deapsorbed using any ultrasound at all (compared to using no ultrasound) is different for different amounts of oil in the sample.

7.38   a. The boxplot below shows that there is a small difference between these two samples with scores on the Captain's exam being slightly higher, but it is not obvious that this difference will be significant.
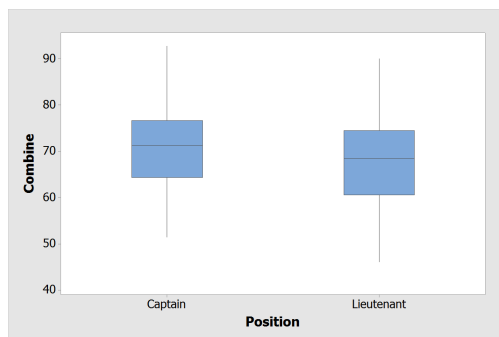

b. The interaction plot below shows the possibility that there is an interaction. Hispanics seem to have the most difference between the two types of exams, whites have somewhat less difference, and blacks do not seem to have any difference at all between the two exams.
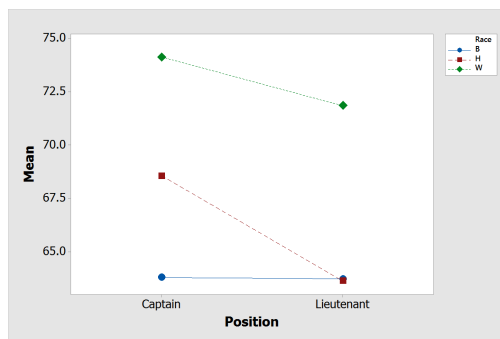
c. The first condition that we check is normality. The normal plot given below of the residuals is perfectly acceptable.
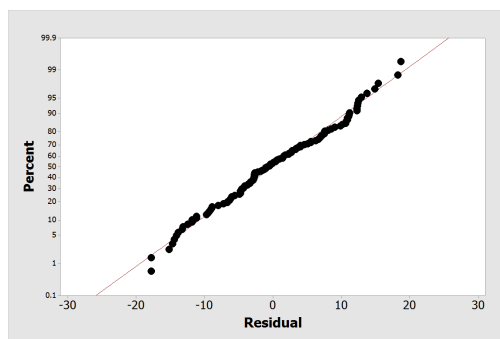
Next we check the data for consistency with the equality of variances condition. The residuals versus fitted values plot given below seems ok, and when we check the ratio of largest standard deviation to smallest we get $\frac{9.15}{5.77} = 1.59$ which is small enough.
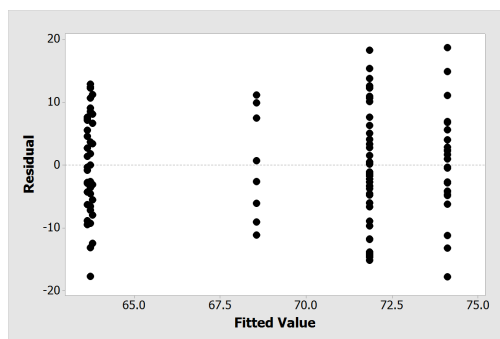
Finally, we already assessed the independence and mean 0 conditions in this data in a previous exercise and determined that the data satisfied these conditions.

d. The Minitab output for this data is given below. Based on this analysis it appears that only race is significant in the model. The R-square (adjusted) is 17.62% for the two-way model

with interaction, while the R-square (adjusted) was 17.72% for the one-way model. Because nothing in the two-way model was significant other than race, and because the R-square (adjusted) actually went down for the two-way model, the recommendation is to use the one-way model.

```
Analysis of Variance
Source              DF      Adj SS     Adj MS     F-Value    P-Value
 Race               2       1749.4     874.71      12.05      0.000
 Position           1        121.7     121.73       1.68      0.198
 Race*Position      2         64.0      32.00       0.44      0.645
Error              112       8131.3     72.60
Total              117      10310.8


Model Summary
      S      R-sq      R-sq(adj)    R-sq(pred)
 8.52061    21.14%      17.62%        12.76%
```

7.39   a. The points at the far right have the largest fitted value (cell mean) and largest variability (cell standard deviation). This corresponds to the female birds getting the hormone (female, yes).

   b. The variability of the residuals increases as the fitted values increase. This indicates a problem with the condition that the errors have constant variance.

   c. The standard deviation in the "female, yes" cell (7.51) is much larger (more than 3 times larger) that the standard deviations for either of the cells where no hormone is given (2.12 and 2.49.)

   d. To reduce the extra variability associated with larger values, we should transform down. Thus a square root would be preferred over squaring the responses. Note: Other "down" transformations (such as log) might also be useful.

7.40   a. Most of the residuals along the line stay between -5 and $+5$, but the two atypical residuals are around -10 and $+8$. Since the "female, yes" group has a much larger standard deviation (7.51) than the other three cells, there's a good chance the more extreme residuals are from that cell. We can confirm this by looking at the residuals versus fitted values plot where the residuals with the largest magnitude are from the group at the far right which has a fitted value (cell mean) over 30. This matches the "female, yes" group (cell mean = 31.08).

   b. Although the normal quantile plot has individual points at both ends that stray a bit from the line, the other 18 points look fine. This does not show a strong need for a transformation to address an obvious departure from normality.

7.41   a. The point on the line has the next to smallest $log(SD)$ and next to smallest $log(average)$. The cell we need has the same properties for its mean and standard deviation. That would be

the "female, no" cell (mean $= 14.88$, SD $= 2.49$). Taking logs, we have $(log(14.88), log(2.49)) =$ $(2.70, 0.91)$. This looks like a good match for the point in question.

b. The slope of the line is 1.05, so the power we should use for the transformation is $1 - 1.05 =$ -0.05 or about zero. That corresponds to a log transformation.

c. The earlier plot in the section used log_{10}, while the plot for this exercise used a natural log. You can check this finding each type of log for the values in the cell means table. Note that, despite the different scales, the general appearance of the plot, and slope of the fitted line to help determine the transformation, are identical.

7.42 For both intervals and the effect sizes, we find the standard deviation with

$$SD = \sqrt{MSE} = \sqrt{206.6} = 14.37$$

Also, a 95% t-value with 56 df is 2.00.

When doing the comparison conditional on *meat*, the sample sizes are 20 and 20, so the standard error is

$$SE_{Meat} = 14.37\sqrt{\frac{1}{20} + \frac{1}{20}} = 4.54$$

and the interval comparing high to low quality for meat is

$$(85.90 - 83.90) \pm 2.00(4.54) = 2.00 \pm 9.08 = -7.08 \text{ to } 11.08$$

We are 95% sure that the mean weight gain for high quality meat could be between about 7 grams less and about 11 grams more than for low quality meat.

When doing the comparison conditional on *cereal*, the sample sizes are 10 and 10, so the standard error is

$$SE_{Cereal} = 14.37\sqrt{\frac{1}{10} + \frac{1}{10}} = 6.43$$

and the interval comparing high to low quality for cereal is

$$(99.75 - 79.05) \pm 2.00(6.43) = 20.70 \pm 12.86 = 7.84 \text{ to } 33.56$$

We are 95% sure that the mean weight gain for high quality cereal could be somewhere between 7.8 and 33.6 grams more than for low quality cereal.

Note that the interval is wider for cereal, since those sample sizes are smaller, but it fails to include zero; providing good evidence that quality matters when using cereal to feed rats. On the other hand, the interval for meat includes zero, indicating that quality might not be important at all when feeding rats meat-based protein.

a. For the effect sizes we divide the difference in the cell means by the overall $SD = \sqrt{MSE} = 14.37$.

High versus low quality for meat: $(85.90 - 83.90)/14.37 = 0.14$ or 14%.

High versus low quality for cereal: $(99.75 - 79.05)/14.37 = 1.44$ or 144%.

We see almost no effect on quality for meat protein, but a more substantial effect (in favor of high quality) for cereal protein.

7.43 We begin our analysis by looking at boxplots of the four groups of flies (shown below).
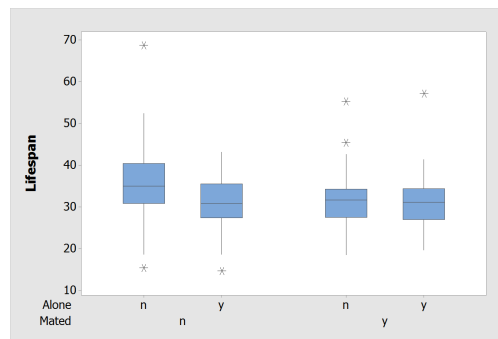

This graph suggests that there may be a difference in the average life span between those who were given mating opportunities and those who were not, though that difference may not be very large. Similarly there may be a small difference in the average life span between those who were living with a male competitor and those who were living alone. It also appears that there might be an interaction, in that the effect of living alone is more pronounced for those who were not given mating opportunities.

To take a different look at the interaction, we created an interaction plot (shown as follows).
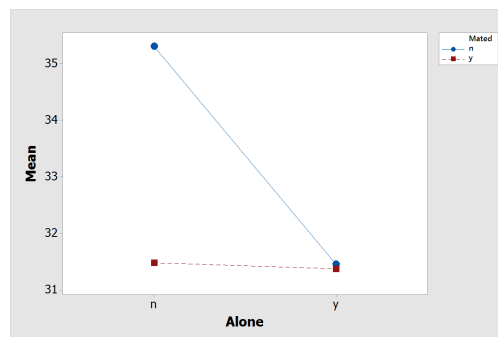
This plot verifies the ideas that we already saw in the boxplots.

Because it appears that there may be an interaction between the two variables, we begin by fitting a two-way nonadditive ANOVA model. The following four graphs show that the normality and constant variance conditions are met.
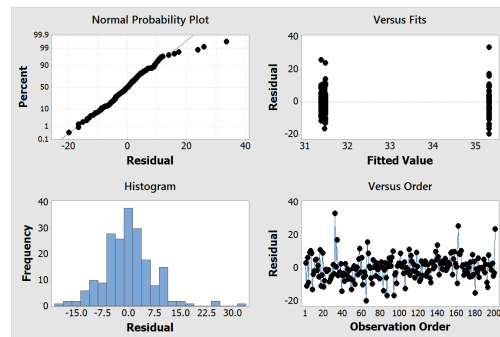
Because the flies were randomly allocated to treatments we are comfortable with the independence and zero mean conditions.

The two-way ANOVA table from fitting this model is

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

Source        DF     Seq SS      Adj SS   Adj MS     F       P
Mated          1     212.06      191.99   191.99   3.61   0.059
Alone          1     198.94      197.23   197.23   3.71   0.056
Mated*Alone    1     176.76      176.76   176.76   3.33   0.070
Error        197   10471.83    10471.83    53.16
Total        200   11059.58


S = 7.29085    R-Sq = 5.31%    R-Sq(adj) = 3.87%
```

None of the variables is strongly significant, but there is weak evidence that both Mated and Alone are related to life span of these flies. There is even weaker evidence of an interaction. In fact, the overall R-squared (adjusted) is only 3.87%.

In an effort to get a better model, we try dropping the interaction from the model. The conditions are still met (see the four plots below).
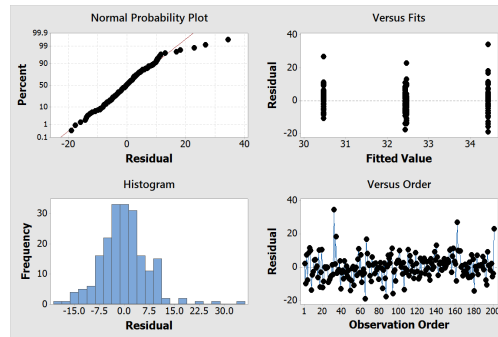
The ANOVA results are given below:

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

Source   DF    Seq SS     Adj SS  Adj MS     F      P
Mated     1    212.06     193.67  193.67  3.60  0.059
Alone     1    198.94     198.94  198.94  3.70  0.056
Error   198  10648.58   10648.58   53.78
Total   200  11059.58


S = 7.33353   R-Sq = 3.72%   R-Sq(adj) = 2.74%
```

We notice here that not only did the P-values not change, but the R-squared (adjusted) went down to a measly 2.74%. This model is not better than the first one.

The final conclusion is that while the variables Mated and Alone may have a weak association with life span of these male fruit flies, neither the two-way additive nor the two-way nonadditive model is worth using since they only account for less than 4% of the variation.

7.44 We begin our analysis by looking at boxplots of the four groups of flies (shown below).
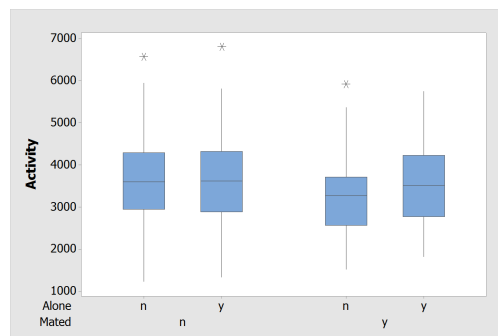
This graph suggests that there may be a difference in the average activity level between those who were given mating opportunities and those who were not, though that difference is small. It does

not appear that there is a difference in the average activity level between those who were living with a male competitor and those who were living alone. It does appears that there might be an interaction, in that the effect of living alone is more pronounced for those who were given mating opportunities.

To take a different look at the interaction, we created an interaction plot which follows.
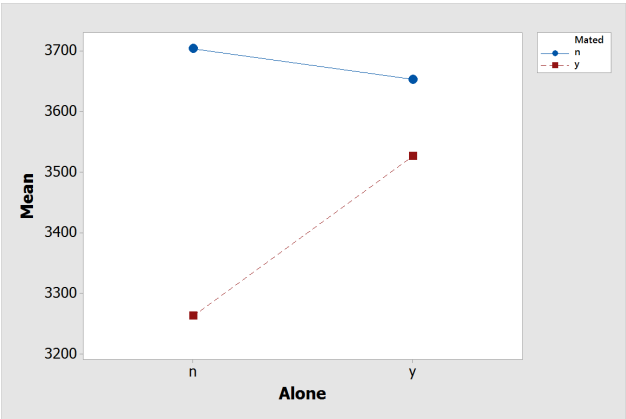
This plot verifies the ideas that we already saw in the boxplots, though it might suggest a small association between activity level and whether the fly lived alone or not.

Because it appears that there may be an interaction between the two variables, we begin by fitting a two-way nonadditive ANOVA model. The following four graphs show that the normality and constant variance conditions are met.
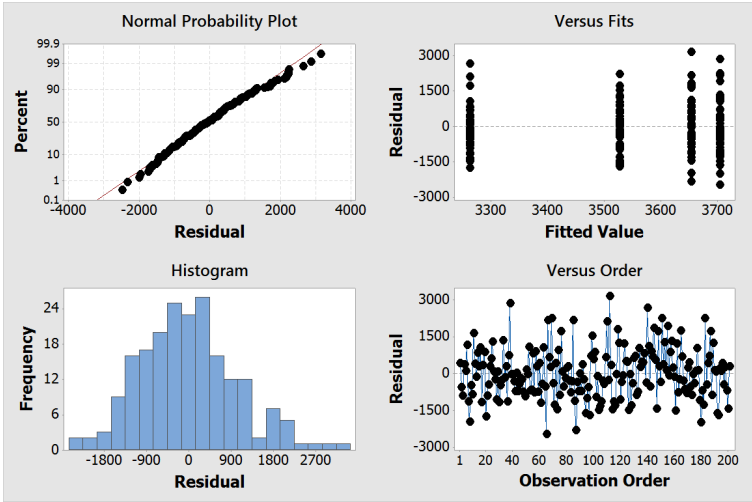
Because the flies were randomly allocated to treatments we are comfortable with the independence and zero mean conditions.

The two-way ANOVA table from fitting this model is

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

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Mated	1	3934230	4040198	4040198	3.76	0.054
Alone	1	563739	571349	571349	0.53	0.467
Mated*Alone	1	1236508	1236508	1236508	1.15	0.285
Error	197	211959539	211959539	1075937		
Total	200	217694016				

```
S = 1037.27   R-Sq = 2.63%   R-Sq(adj) = 1.15%
```

None of the variables is strongly significant, but there is weak evidence that *Mated* is related to activity level of these flies. However, the overall R-squared (adjusted) is only 1.15%. This suggests that the model is not particularly useful. We will try a simpler one-way ANOVA model with just *Mated* to see if this is any better.

The independence and mean 0 conditions are still met and the four plots given below show that the data satisfies the normal and constant variance conditions.
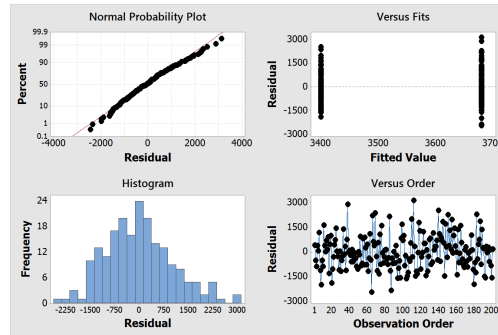
The ANOVA table from software is given below.

Source	DF	SS	MS	F	P
Mated	1	3934230	3934230	3.66	0.057
Error	199	213759786	1074170		
Total	200	217694016			

```
S = 1036   R-Sq = 1.81%   R-Sq(adj) = 1.31%
```

Mated is still only marginally significant and the R-squared (adjusted) is still an abysmal 1.31%.

The final conclusion is that while the variable *Mated* may have a weak association with the activity level of these male fruit flies, the model is not worth using since it only accounts for about 1% of the variation.

7.45 We start our investigation by looking at a plot of the difference in oil deapsorbed between treated and control samples, taking all three explanatory variables into consideration. The following boxplots breaks the data down by the eight different treatment combinations:
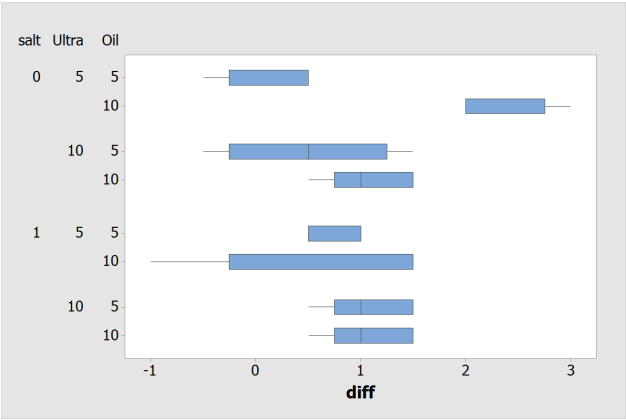
It appears from the boxplots that there are some differences between the eight groups. In particular, amount of oil seems to make a difference in both ultrasound groups when using distilled water, but does not make much difference in salt water for either ultrasound group. Also, amount of time exposed to ultrasound does seem to make a difference for 5 ml of oil, across salt categories. There is enough potentially going on here that we start with a full three-way ANOVA model.

The residual plots from this model suggest that the conditions are met. The residuals look reasonably normally distributed and have similar variances.
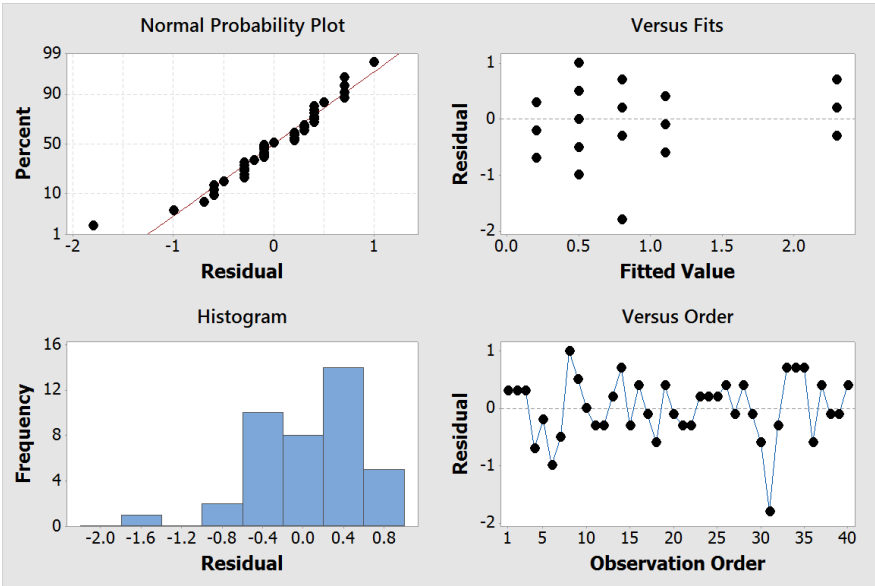

The full ANOVA table is given below:

```
Analysis of Variance

Source           DF    Adj SS    Adj MS   F-Value   P-Value
  Ultra           1    0.0563   0.05625      0.16     0.692
  Oil             1    4.5562   4.55625     12.90     0.001
  salt            1    0.0562   0.05625      0.16     0.692
  Ultra*Oil       1    1.4062   1.40625      3.98     0.055
  Ultra*salt      1    1.4062   1.40625      3.98     0.055
  Oil*salt        1    4.5562   4.55625     12.90     0.001
  Ultra*Oil*salt  1    1.4062   1.40625      3.98     0.055
Error            32   11.3000   0.35313
Total            39   24.7437
```

What we see is that Oil is the only significant main effect, but all of the interaction effects are (or nearly are) significant. So we produce an interaction plot matrix to see just what we can conclude.
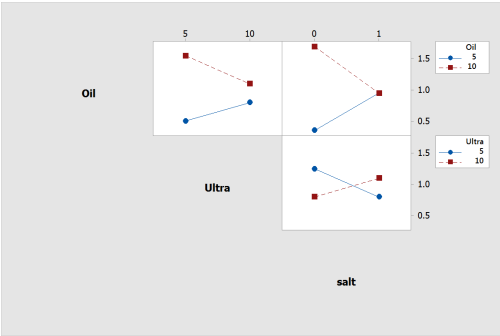

What we see is as follows: When there is more oil in the sample, less ultrasound is better and freshwater is better. When there is less oil in the sample, more ultrasound is better and salt water is better. And we see that if the sample has freshwater, less ultrasound is better and when the water is salty, more ultrasound is better.

Looking at one more graph, this time a three-way interaction graph (the combination of the two-way interaction graphs for each level of salt).
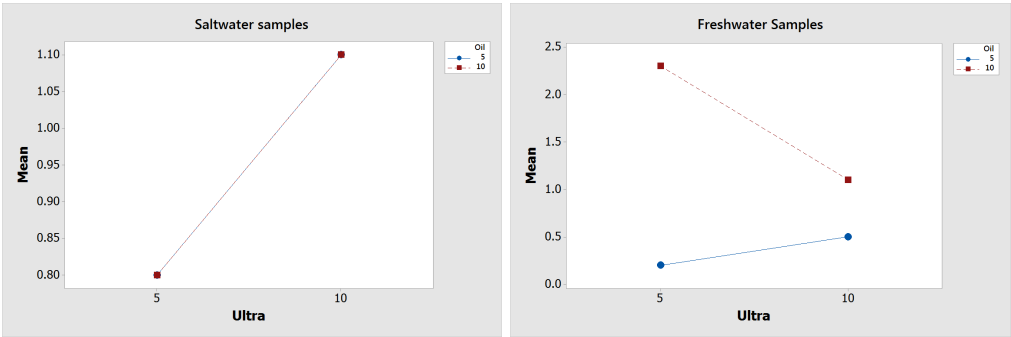
What this graph tells us is that for saltwater, there is very little difference between amount of oil or amount of ultrasound. But for freshwater, the best combination is 10 ml of oil and 5 minutes of ultrasound.

7.46 After having analyzed the data for the oral examination for promotion scores for a group of 118 firefighters, I find that the mean scores differ both across racial lines and between the two levels of promotion being sought, but there is no interaction between the two variables. To come to these conclusions I started by looking at an interaction plot for the oral exam. The plot (shown as follows)
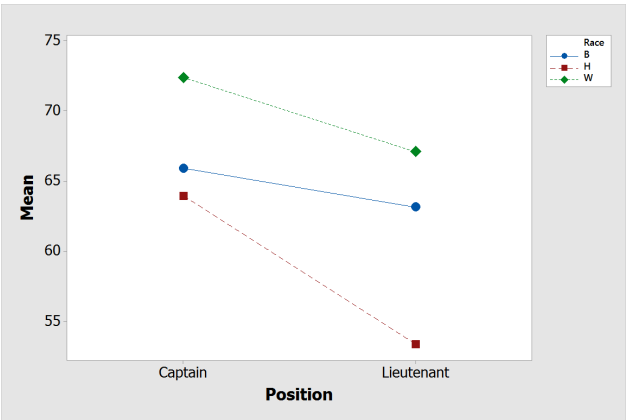
suggests that there might be an interaction. It looks like Hispanics do much better on the oral exam for Captain than they do for Lieutenant. Blacks and whites have less of a difference between the two exams.

Because the interaction plot shows a possible interaction, I began with an ANOVA model with interaction. The result is given below and shows that race (P-value $= 0.001$) and position (P-value $= 0.014$) are significant, but the interaction is not (P-value $= 0.510$).

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

Source          DF    Seq SS    Adj SS   Adj MS     F      P
Race             2    2553.4    2030.9   1015.5   7.88   0.001
Position         1     892.8     801.0    801.0   6.22   0.014
Race*Position    2     174.8     174.8     87.4   0.68   0.510
Error          112   14432.6   14432.6    128.9
Total          117   18053.6


S = 11.3518    R-Sq = 20.06%    R-Sq(adj) = 16.49%
```

Since the interaction was not significant, we reran the model without it, getting the following output.

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

Source      DF    Seq SS    Adj SS   Adj MS     F       P
Race         2    2553.4    2475.4   1237.7   9.66   0.000
Position     1     892.8     892.8    892.8   6.97   0.009
Error      114   14607.4   14607.4    128.1
Total      117   18053.6


S = 11.3197    R-Sq = 19.09%    R-Sq(adj) = 16.96%
```

In this model both variables are even more significant (P-values of < 0.001 and 0.009) and the R-square (adjusted) rose slightly to 16.96%. The following group of graphs shows that the conditions for the two-way ANOVA have been met.
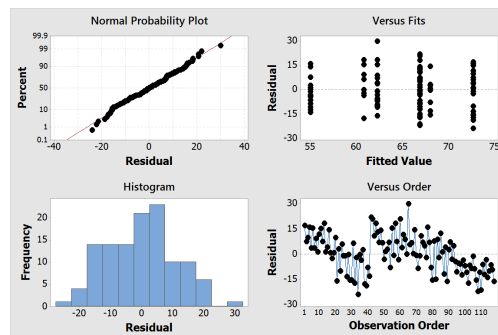

Taken together, here are the final results. It appears that whites do better generally than blacks and Hispanics on this oral exam. It also appears that people taking the Captain's exam generally do better on the oral portion than those taking the Lieutenant's exam.

7.47 Here are some diagnostic plots for the **FatRats** data.

Residuals versus fitted values plot and normal quantile plot
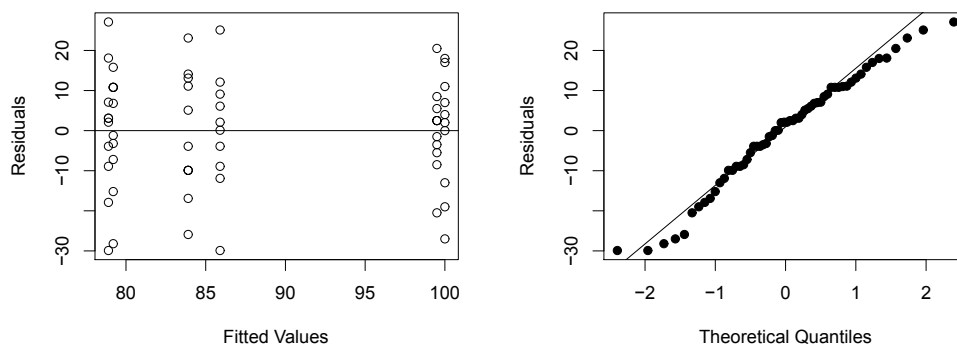
Tukey plot and log(SD) versus log(average) plot
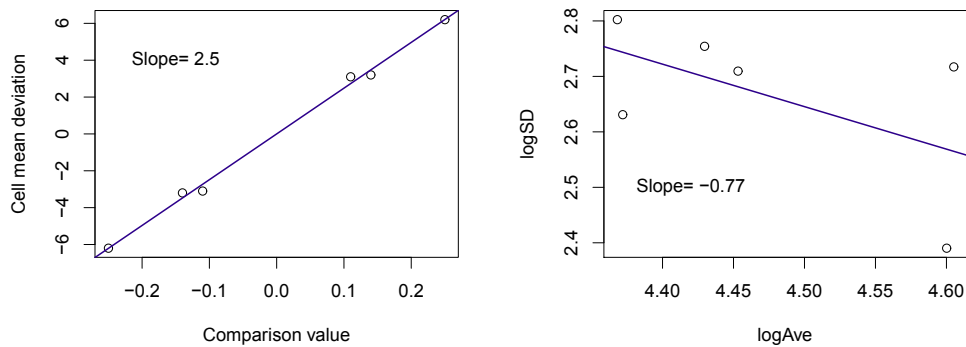
The residuals versus fits plot looks good, roughly equal scatter above and below the zero line and similar variability across all six groups. No issues that suggest transforming here.

The normal quantile plot also looks good. Just a touch of wiggle away from the line at the low and high ends, but nothing serious to raise concerns about a clear lack of normality in the residuals.

The Tukey plot shows a strong linear trend with a slope of 2.5. This suggests a transformation with power $1 - 2.5 = -1.5$. So perhaps a reciprocal (power$= -1$) or slightly higher negative power.

The log(SD) versus log(average) plot has a slope of -0.77, suggesting a power transformation of $1 - (-0.77) = 1.77$. So perhaps squaring the weight gains.

Where does all this leave us? Two plots show no substantial problems, one suggests transforming down and another suggests transforming up. The fit is much weaker in the log(SD) versus log(average) plot, so we might discount its suggestion somewhat. But a transformation that makes one of the latter two plots better is likely to make the other worse. This suggest a couple of courses of possible action.

1. Try a transformation, such as a reciprocal, suggested by the Tukey ploy and check what happens to the other results.

2. Don't worry about doing any transformation. The standard ANOVA conditions on the residuals look pretty good in the first two plots, and it's generally easier to interpret results in the original scale.