# Chapter 2
# The Two-Sample t-test, Regression, and ANOVA: Making Connections

## Activity Solutions

**1.** Units: each student
Population: set of all students at this college who would be willing to be part of the study
Explanatory variable: type of game (standard or with a color distracter)
Response variable: the completion time (in seconds)

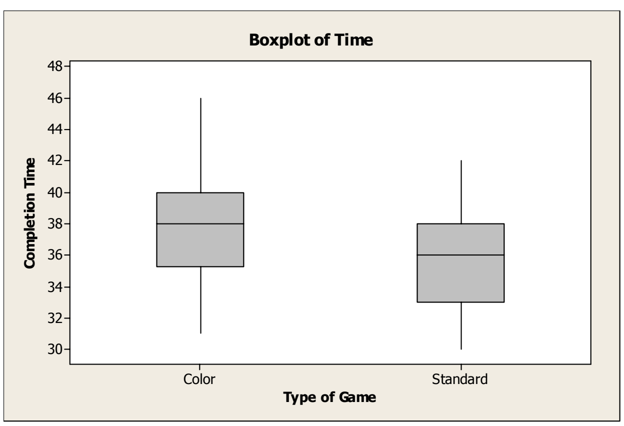
**2.** This study is an experiment, since students were randomly allocated to one of the two types of games.

**3.** Null hypothesis: $H_0$: $\mu_1 = \mu_2$ (the mean completion time for the standard game is equal to the mean completion time for the color distracter game)

Alternative hypothesis: $H_a$: $\mu_1 \neq \mu_2$ (the mean completion time for the standard game is not equal to the mean completion time for the color distracter game)

Note: Some students might choose a one-sided alternative to test whether the color distracter lengthens average completion time. We use the more conservative two-sided test because it is more comparable to the F-test using ANOVA.

**4.**




Using the boxplot or Figure 2.2 in the text, we see the color group appears to have a higher mean while the variances appear to be equivalent. There do not appear to be any unusual observations.

```
Variable     Type      Mean     StDev
Time         Color     38.100  3.655
             Standard  35.550  3.395
```

**5.** $\mu_1 = \dfrac{(y_{11}+y_{12}+y_{13})}{3} = \dfrac{(15+17+16)}{3} = \dfrac{48}{3} = 16$

$\mu_2 = \dfrac{(y_{21}+y_{22}+y_{23})}{3} = \dfrac{(11+9+10)}{3} = \dfrac{30}{3} = 10$

$\varepsilon_{11} = y_{11} - \mu_1 = 15 - 16 = -1$

$\varepsilon_{13} = y_{13} - \mu_1 = 16 - 16 = 0$

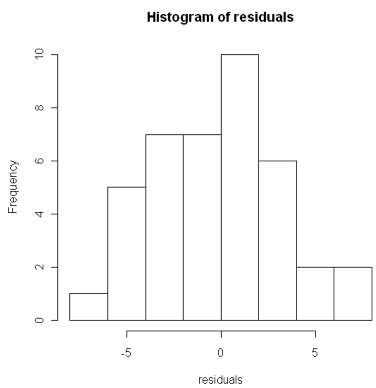
$\varepsilon_{21} = y_{21} - \mu_2 = 11 - 10 = 1$

**6.** The sample sizes are both $n_1 = n_2 = 20$.
$y_{1,12} = 37$ and $y_{2,12} = 37$

$\hat{\varepsilon}_{1,12} = 37 - 38.1 = -1.1$

$\hat{\varepsilon}_{2,12} = 37 - 35.55 = 1.45$

**7.**

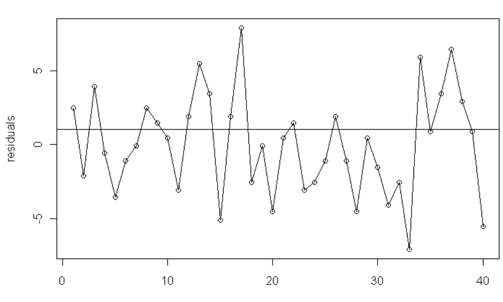



Based on the histogram, there is not strong evidence to suggest the residuals are not normally distributed.

**8.** 1.08 < 2 or 1.16 < 4

Thus we have no reason to reject the hypothesis that $\sigma_1 = \sigma_2$.

**9.**



There is no pattern in the residuals, so there is nothing to suggest that the observations are not independent.

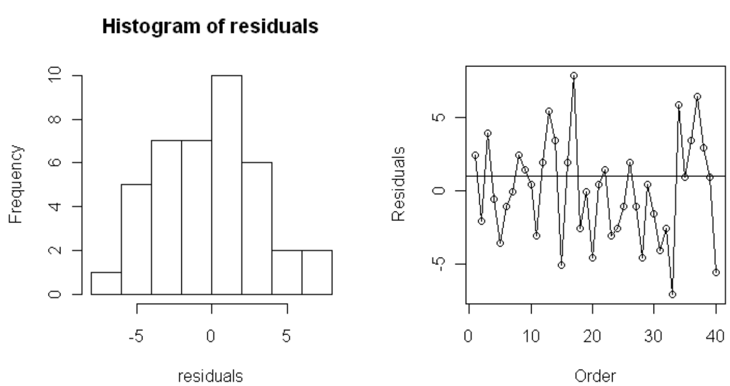
**10.** The test statistic and (two-sided) p-value are 2.2862 and.02791, respectively. If the mean completion times for the two types of games were actually the same, then we would only observe a difference this large in the sample mean completion times in 2.79% of experiments like this one. This seems quite rare to be a difference simply due to the random allocation process alone. This allows us to conclude that the completion times do vary because of presence or absence of color distraction. The confidence interval for the difference between the two means is (0.292, 4.808).

**11.** Time = 35.6 + 2.55X where X = 1 represents the color group.

**12.** The t-statistic and p-value are 2.286 and.0279, respectively. Based on the p-value, there is sufficient evidence to conclude that $\beta_1$ is significantly different from 0. The 95% confidence interval for $\beta_1$ is (0.292, 4.808). These are exactly the same values found in Question 10 using the two-sample pooled-variance t-test. This is not surprising, since $\beta_1 = \mu_1 - \mu_2$ is the difference in the average completion times. Testing Ho: $\beta_1 = 0$ is equivalent to testing Ho: $\mu_1 = \mu_2$.

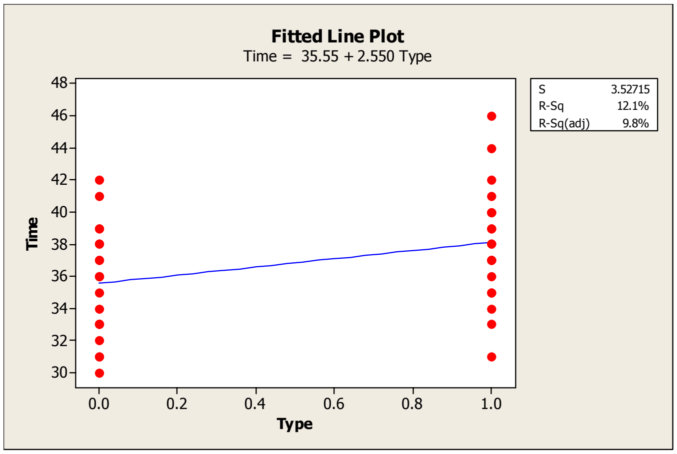
**13.** Time = 35.6 - 2.55X where X = 1 represents the standard group. The choice of indicator value does not affect the conclusions.

**14.**




The results of the informal test of equal variances are identical to the one conducted in Question 8. The residual plots are identical to those that were created for the two-sample t-test.

**15.**




If the color game is used instead of the standard game, the expected mean completion time will increase by 2.55. The y intercept is 35.55, this is the expected mean completion time when X = 0 (the standard game is played).

**16.** Since $\mu_1 = \mu + \alpha_1$, and $\mu_2 = \mu + \alpha_2$, $\alpha_1 = \mu_1 - \mu$, and $\alpha_2 = \mu_2 - \mu$ ($\mu$ is always the same). Therefore, $\mu_1 = \mu_2$ is no different than saying $\alpha_1 = \alpha_2 = 0$. If the two population means are the same, then they both have the same "effect" and so there would just be the one "grand mean" to represent both population means. That is $\alpha_i = 0$ for both groups (i=1, 2).

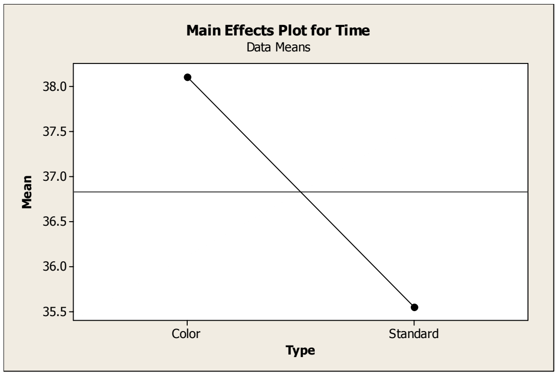
**17.** $y_{1,3} = \mu + \alpha_1 + \varepsilon_{1,3}$ and $y_{2,20} = \mu + \alpha_2 + \varepsilon_{2,20}$

**18.** $\mu$ represents an overall (background, benchmark) level of response common to both groups. Thus, there is no need to differentiate between groups.

**19.** $\bar{y}_{\bullet\bullet} = 36.825$, $\bar{y}_{1\bullet} = 38.1$, and $\bar{y}_{2\bullet} = 35.5$

**20.** Estimate for effect size for color distracter is 38.1 – 36.825 = 1.275. Estimate for effect size for standard is 35.55 – 36.825 = -1.275.

**21.**




**22.** $y_{2,20} - \bar{y}_{2.} = 30 - 35.55 = -5.55$

**23.** Analysis of Variance for Time

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Type | 1 | 65.03 | 65.03 | 5.23 | 0.028 |
| Error | 38 | 472.75 | 12.44 | | |
| Total | 39 | 537.77 | | | |

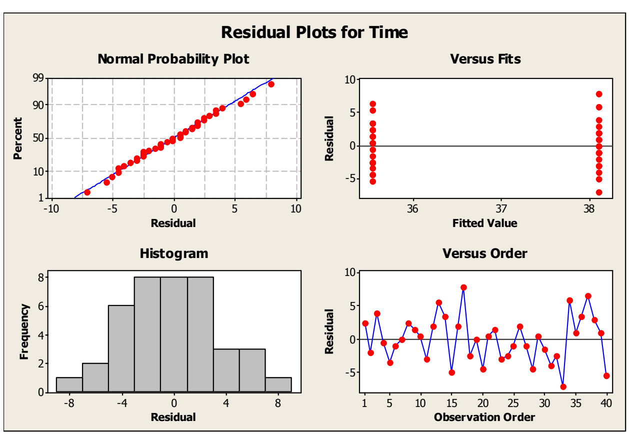
$F = 5.23$, $p = 0.028$.

The small p-value, 0.028, suggests that $\alpha_1$ and $\alpha_2$ are significantly different. This allows us to conclude that the reaction times do vary because of presence or absence of color distraction.

**24.** The p-value is identical to that found for the two-sample t-test and regression model.

**25.** SQRT{5.226758} = 2.286.This value (2.286) is identical to the t-statistic for testing that the regression slope is zero and the two-sample t-statistic.

**26.**



The spread in the residuals appears similar for the two groups. There is no trend or cyclical behavior by time order, although the residuals seem to have a slight increase in variability for later observations. But that does not seem pronounced. The graphs do not give enough evidence to reject the hypothesis that the residuals come from a symmetric, mound-shaped (normal) population.

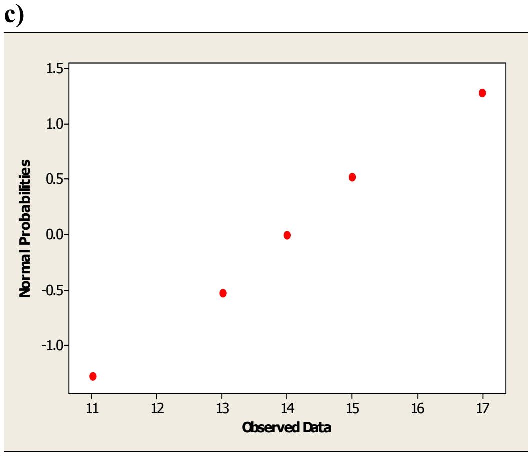
**27.** The mean responses (and thus the random error terms) of all three models are identical. All three models describe two populations with the same variances, but possibly different means. In each model, the assumptions about the random errors are the same (normal with mean zero and variance $\sigma^2$). Thus, the p-value must also be identical for all three models. It is not obvious why the square of a t-distributed statistic should have an F-distribution (a proof of that requires a bit of statistical theory), but when the model assumptions are the same, it is comforting that either test statistic provides the same p-value.

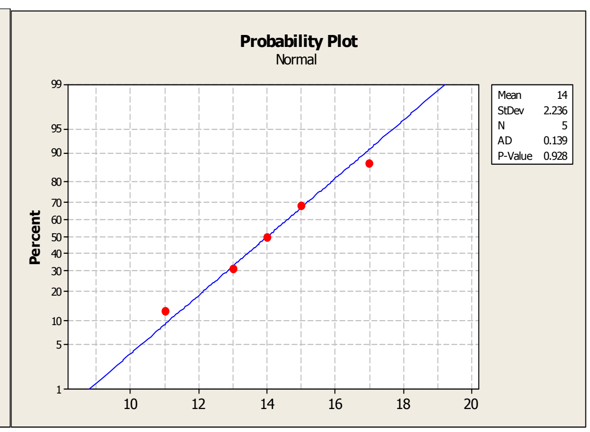
## Extended Activity Solutions

**28.**

| **a)** | **b)** | normal percentiles |
|---|---|---|
| 11 | 0.1 | -1.28155 |
| 13 | 0.3 | -0.52440 |
| 14 | 0.5 | 0.00000 |
| 15 | 0.7 | 0.52440 |
| 17 | 0.9 | 1.28155 |

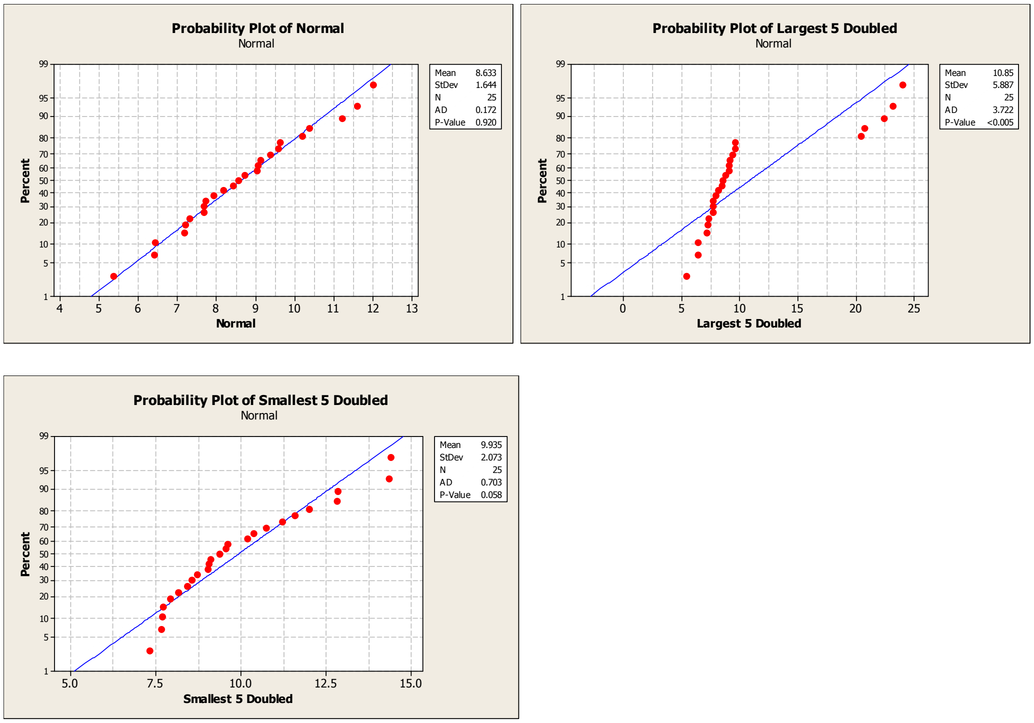
**c)**




**d-f)** The normal probability plot evaluates the shape of the distribution, not the mean or the variance. Multiplying, dividing, adding or subtracting will change the mean and variance, but not the shape. Except that the x-axis labels, the probability plots for **(d) – (f)** are identical to Part (c).

**29. a)**



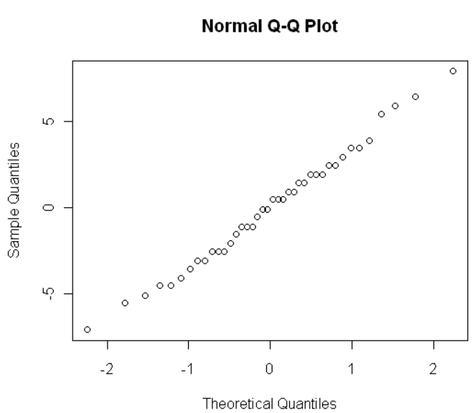

**b)** The five largest points are moved even farther to the right. The probability plot no longer looks like a straight line.

**c)** On the left, the observed probability plot seems curved down.

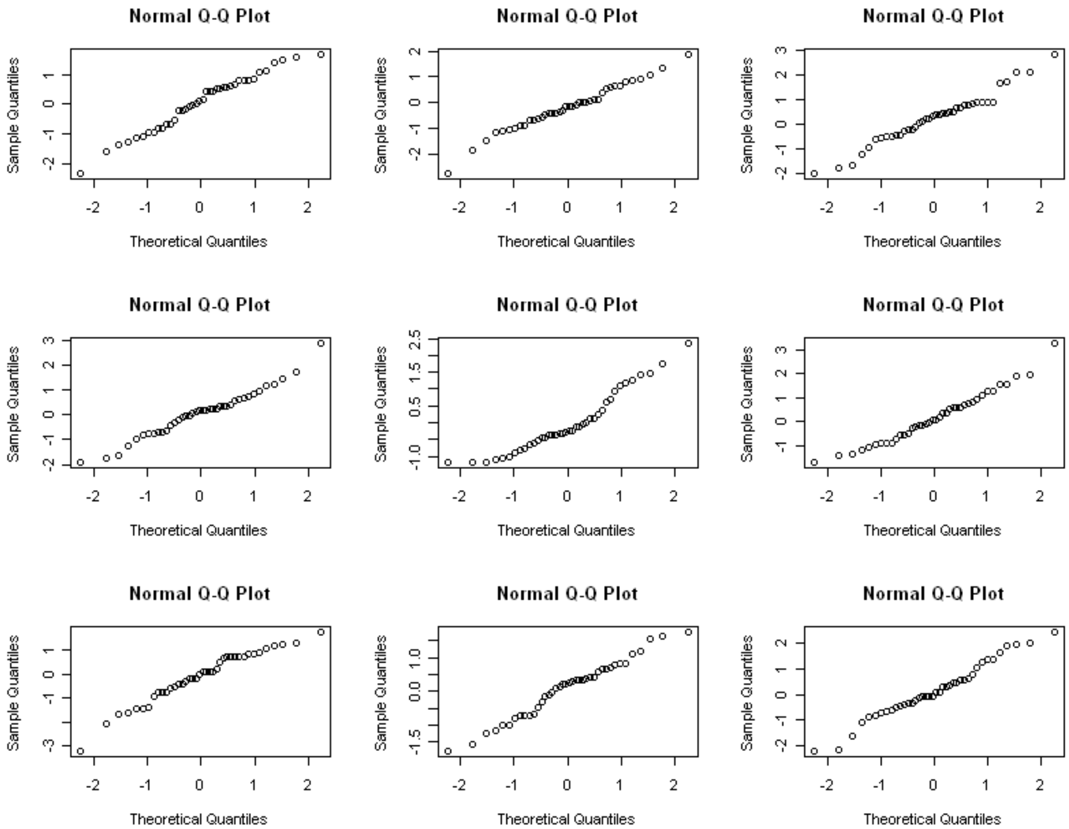
**d)** The plot would be curved down on the left (as in Part (c)) and curved up on the right.

**e)** The plot would be S-shaped.

**30. a)**

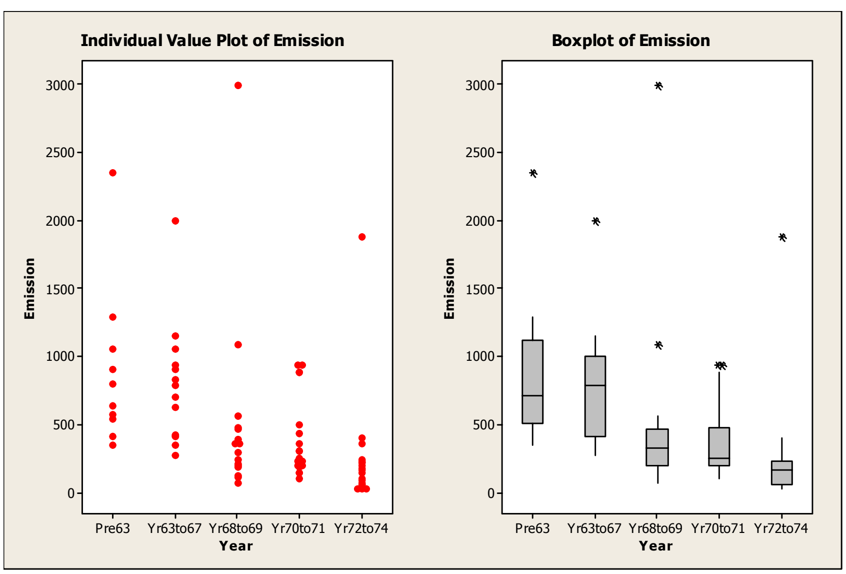


**b-c)**




The plot in Part (a) does resemble the nine plots displayed.

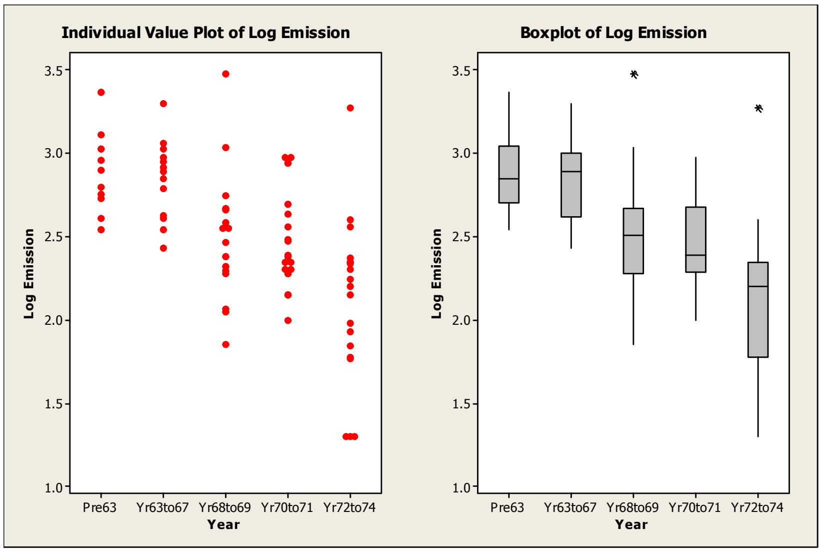
**31. a)**



| Variable | Year | Mean | StDev |
|---|---|---|---|
| Emission | Pre63 | 891 | 592 |
| | Yr63to67 | 801 | 455 |
| | Yr68to69 | 506 | 708 |
| | Yr70to71 | 381.4 | 287.9 |
| | Yr72to74 | 244.1 | 410.8 |

The data do not look consistent with data from a normal population within each group. The data appear to be skewed right within each group with at least one outlier.

**b)**




| Variable | Year | Mean | StDev |
|---|---|---|---|
| Log Emission | Pre63 | 2.8810 | 0.2476 |
| | Yr63to67 | 2.8437 | 0.2399 |
| | Yr68to69 | 2.4995 | 0.3935 |
| | Yr70to71 | 2.4804 | 0.2943 |
| | Yr72to74 | 2.101 | 0.495 |

The data are no longer right skewed within each group, but it is still questionable as to whether the data within each group are consistent with normal populations.

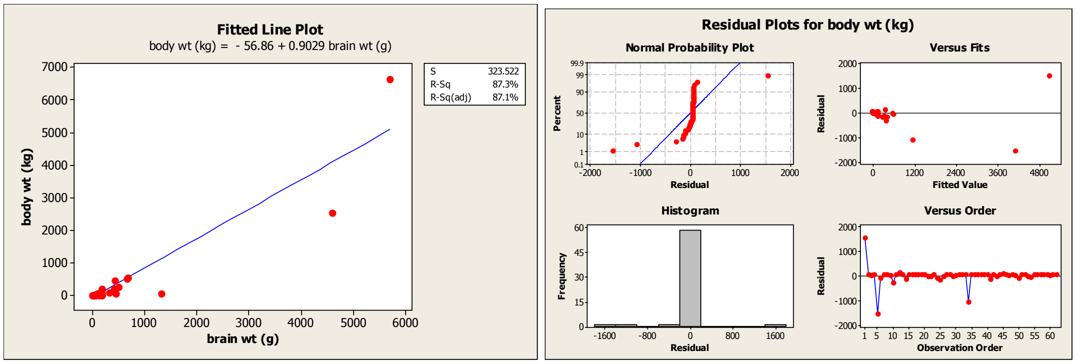
**c)** The ANOVA test applied to the log-transformed data yield the following results:

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Year | 4 | 6.016 | 1.504 | 11.42 | 0.000 |
| Error | 73 | 9.615 | 0.132 | | |
| Total | 77 | 15.631 | | | |

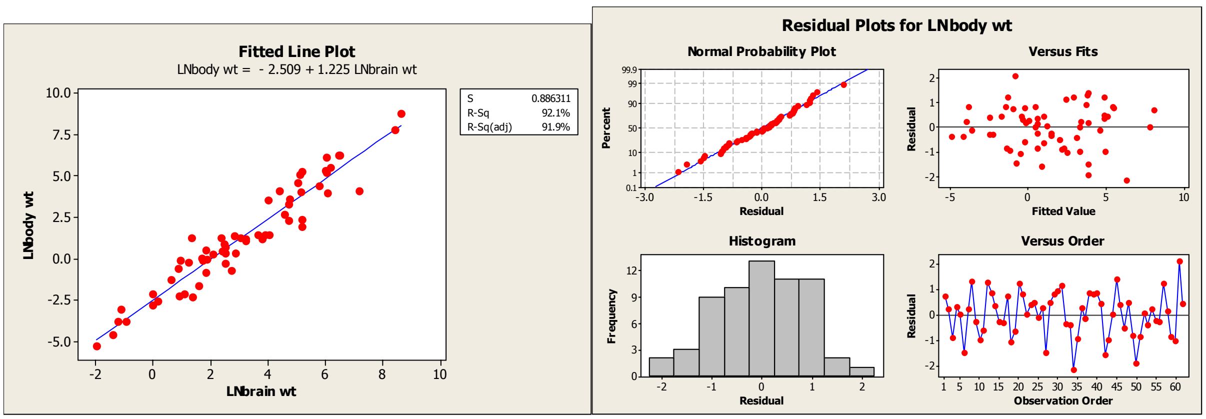
So we reject the null hypothesis that the means of the log-transformed emission levels are identical across year, and conclude that the mean log-levels vary for at least two time periods.

Note that the solutions use log base 10 ($\log_{10}x$), not the natural log($\log_e x = \ln x$). No matter what log transformation is used the F-tests and p-values will be the same since there is a linear relationship: for any base b, $\log_b x = \log_{10}x/\log_{10}b = \log_e x/\log_e b$.

**32. a)**

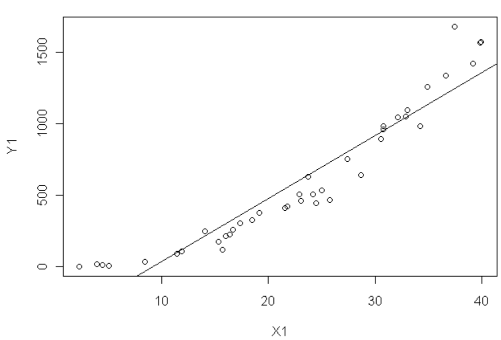



**b)** Answers will vary; The natural logarithm (ln) transformation of both variables is used:

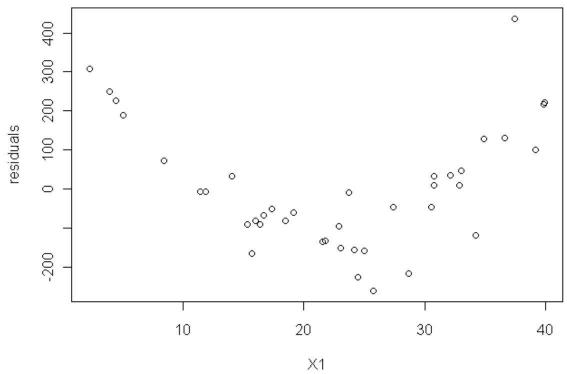



**33.** Answers will vary. Only the solution using the first set of x and y variables (X1 and Y1) is shown here.

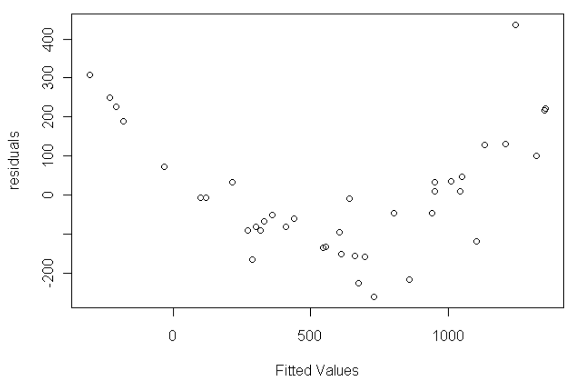
**a)** Scatterplot of X1 versus Y1:



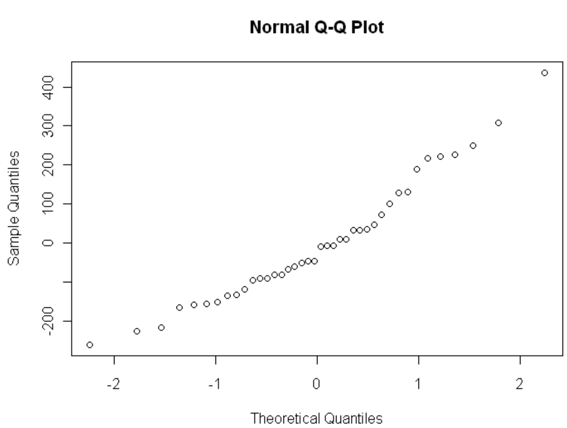
Plot of the residuals versus X1:




Plot of the residuals versus the predicted values:

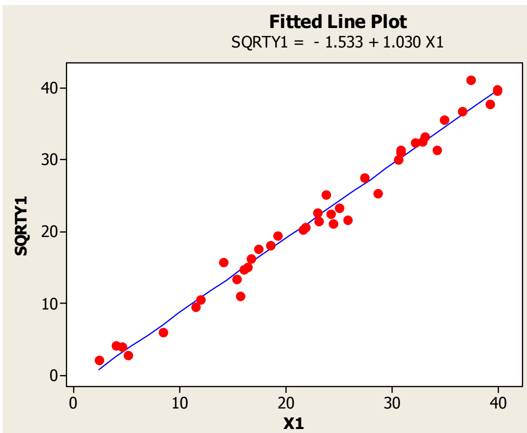


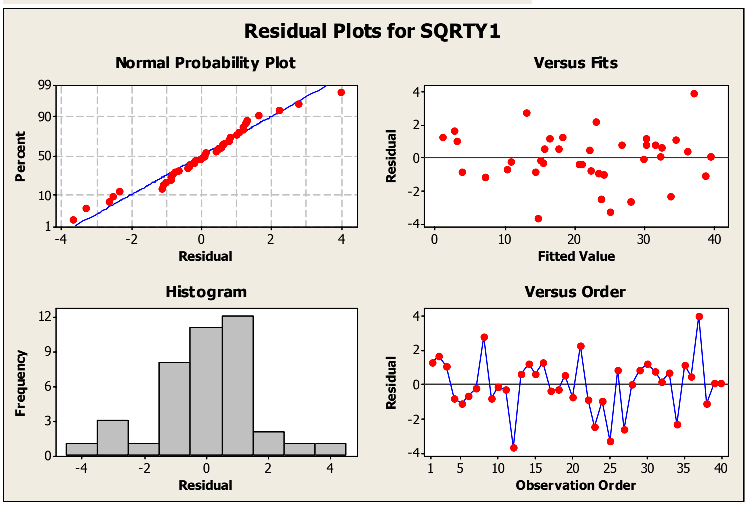

Construct a normal probability plot of the residuals:



Although the normal probability plot indicates that the residuals are consistent with observations from a normal population, the plot of the residuals versus the explanatory/fitted values indicate that a line may not be appropriate for modeling the data.

**c)** Based on the plot of the residuals versus the explanatory variable, we'll try a square root transformation of the response variable. The XY plot of the residuals plots are shown:




The plot of the residuals versus the fitted values appears to be a random scatter, and normal probability plot indicates that the residuals appear normally distributed.

**34.** t = 2.2862 and the p-value =.0279

**35.** The estimated standard deviation of the random errors is 1.1154, and the test statistic for the null hypothesis that $\beta_1 = 0$ is 2.286. The p-value =.0279.

**36.** The test statistic and p-value in Questions 34 and 35 are identical.

**37.** $SS_{Type} = 65.02$ , $SS_{Error} = 472.75$ , $MS_{Type} = 65.02$ , $MSE = 12.44$

**38.** Sample variance of completion times is 13.789, so the total sum of squares is $SS_{Total} = 39 \times 13.789 = 537.775$ .

$SS_{Total} = SS_{Type} + SS_{Error}$ = 65.02 + 472.75 = 537.77

**39.** F-statistic = 5.2268 and p-value = 0.0279

**40.** $\hat{\beta}_1 \pm t_{38}^{*} SE_{\hat{\beta}_1}$ , where $SE_{\hat{\beta}_1}$ is the standard error of $\hat{\beta}_1$ . The 95% confidence interval for $\beta_1$ is (0.292, 4.808).

## Exercise Solutions

**E.1** With a larger sample size, the measure of variation shown in the denominator of the t-statistics is $s_p*((1/n_1)+(1/n_2))^{(1/2)}$. The denominator will be smaller if $n_1= n_2=100$ than if $n_1= n_2=16$, $[(2/100)^{(1/2)}$ is smaller than $(2/16)^{(1/2)}]$. In both cases, since $s_1=s_2=10$, $s_p=10$ regardless of the sample size. Thus, a larger sample size results in a larger t-statistic and a smaller p-value.

**E.2** If the hypothesis test results in a small p-value, we can reject the null hypothesis and conclude that the slope is not zero. However, this does not mean that the regression model provides a good estimate of the response value for a given value of $x_i$. For example, there may be an outlier for which the value of the explanatory variable is a poor estimate. The R^2 value is a much better measure of how well the predicted values fit the observed values. Furthermore, the regression test might not have been appropriate (and the p-value may not be accurate) if the data did not meet the assumptions of the least squares regression model.

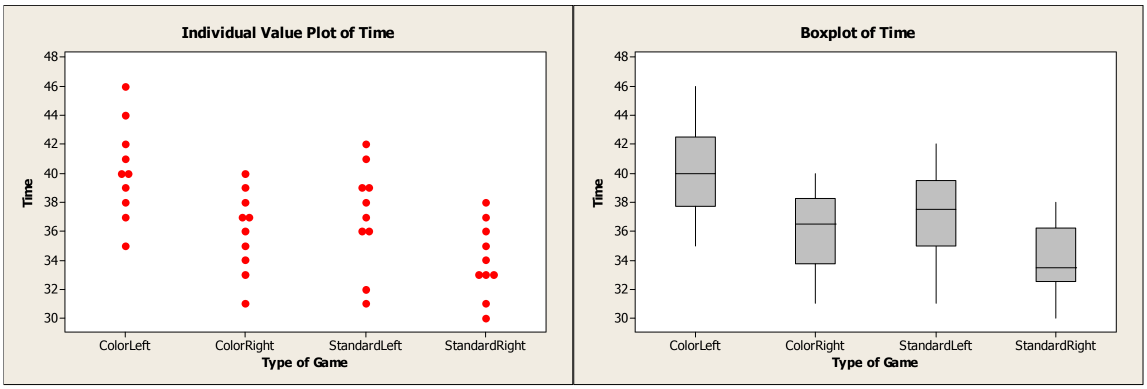
**E.3** There are no assumptions about the error terms or distribution that need to be satisfied in order to calculate $b_0$ and $b_1$ in a simple linear regression model. The values $b_0$ and $b_1$ can be calculated for any data set with distinct x and y values. Model assumptions are needed to determine whether these estimators are actually estimating anything meaningful (as is done with hypothesis tests and confidence intervals).

**E.4** It is not appropriate to have the observed responses in the model without an error term (unless all the observed data fall perfectly along a straight line). The regression model predicts a value of y. These predictions form a straight line based upon our estimate of the intercept and slope coefficient with $b_0$, and $b_1$. Thus, the second equation is appropriate.

**E.5** With only two levels being compared in an F-test, the overall grand mean is found by taking the mean of the two group means, and as the mean is the value that is midway between the two values, the grand mean is midway equidistant from the two group means. Therefore, the two effect values will be equal and opposite.

**E.6** **a)** Looking at box plots of the data, we can see that the Color Left group has the largest median, while the Standard Right group has the smallest median, and that all groups are roughly symmetric. The individual value plot shows that the Standard Left and the Color Left groups have observations (31 and 46 respectively) that are relatively far from other points, but probably not far enough to be considered outliers.

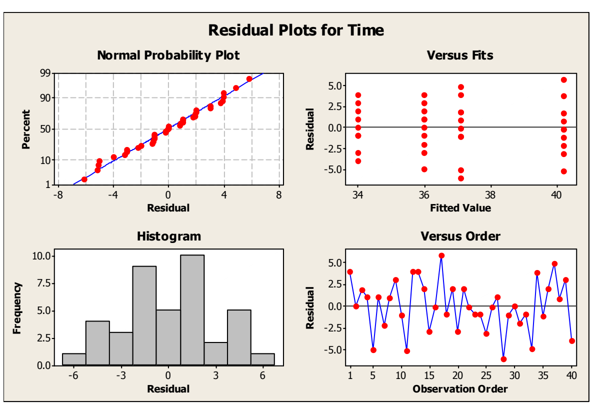
The spreads for all groups are fairly similar. In general, the center of the right hand groups is lower than the left hand groups.




**b)** The ANOVA test shows an F-statistic of 7.18 with a corresponding p-value of 0.001. Based on this, we would reject the null hypothesis that all of the group means are the same, in favor of the alternative that at least one of the group means is different from the others.

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|---|---|---|---|---|---|---|
| Type2 | 3 | 201.275 | 201.275 | 67.092 | 7.18 | 0.001 |
| Error | 36 | 336.500 | 336.500 | 9.347 | | |
| Total | 39 | 537.775 | | | | |

**c)** The histogram of the individual value plot has two peaks, but with a small data set such as this, normal probability plots are more reliable. The residuals in the normal probability plot form a fairly straight line. There do not appear to be any patterns in the residual plots which would provide evidence that the model assumptions are not met.

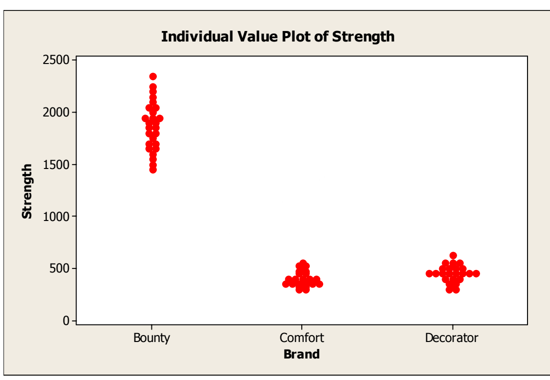


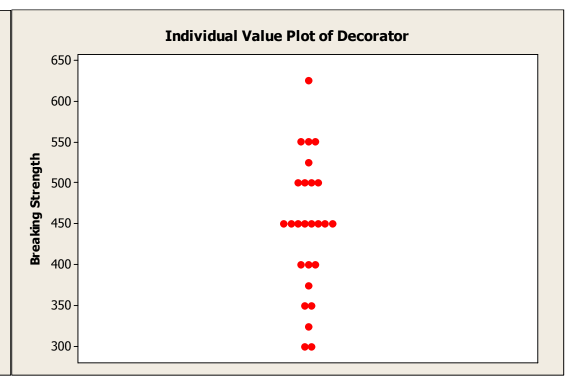

**d)** Answers will vary. The skill level of the subjects and the order in which students played the game could influence the results. It is possible that if all the students chosen were right-handed, then the results will be biased towards one direction; it does seem to be the case that the right-handed group has a lower mean than the other groups.

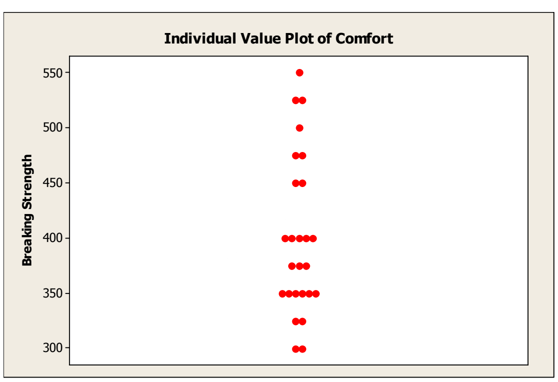
**e)** Based on the results of the ANOVA test, there is a difference among the four group means, and since this experiment was done with random allocation, we can say that this was caused by the type of game. Since data were collected as a random sample of college students who agreed to participate in the experiment, we could say that these results would hold for the population of students who agreed to participate. If there were no difference between these students and the rest of the student body, then these results could hold for the entire college, but as this experiment was only done at this particular college, we cannot generalize these results to other colleges or to the entire population.

**E.7** **a)** The explanatory variable is the brand of paper towels; the observational units are the sheets of paper towels used; and the response variable is the breaking strength of the paper towel, as measured in grams. The null hypothesis is that the mean breaking strength for the three brands is the same, and the alternative hypothesis is that the mean breaking strength of at least one of the brands is different from the others. $H_0$: $\mu_1=\mu_2=\mu_3$, $H_a$: at least one of the means is different from the others.

**b)** Looking at individual value plots, the center of the bounty plot is much higher than that of the other plots, though it also has a greater spread than the other plots. Tcomfort brand has the lowest center, and a similar spread to the decorator brand. Some may say that comfort brand is slightly skewed to the right or that a  possible outlier may exist in the Decorator brand of 625 grams, but both of these are so slight they are unlikely to be an issue in the analysis.



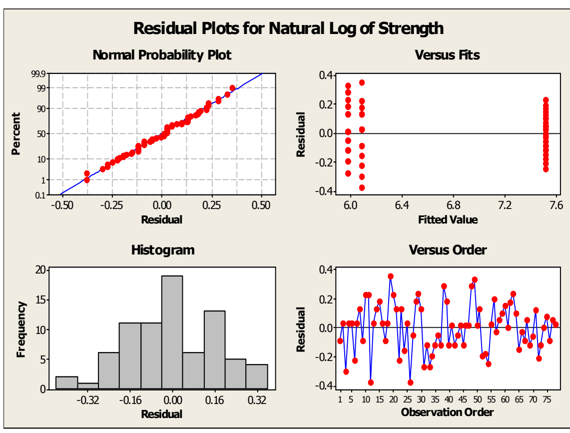

**c)** The results of the ANOVA are that there at least one of the means is different from the others, with a p-value of <0.001.

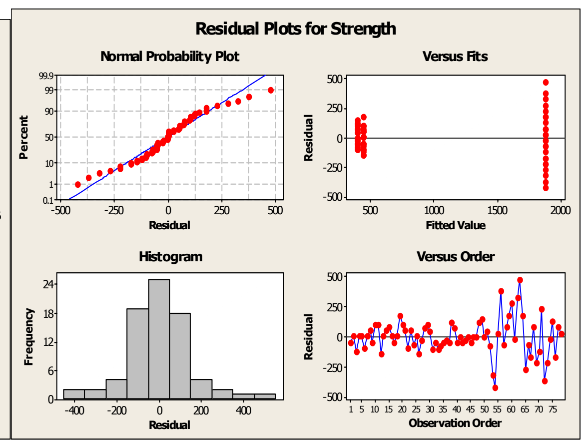
| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|---|---|---|---|---|---|---|
| Brand | 2 | 36349151 | 36349151 | 18174575 | 809.64 | 0.000 |
| Error | 75 | 1683582 | 1683582 | 22448 | | |
| Total | 77 | 38032732 | | | | |

**d)** The equal variances assumption is violated in this study as the standard deviation of the Bounty brand is 235.4, while the other two brands is 71.6 and 82.4 for Comfort and Decorator respectively. After the natural log transformation the ANOVA test gives us, a p-value of less than 0.001. So the ANOVA test still shows that a significant difference exists among the brands.

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|---|---|---|---|---|---|---|
| Brand | 2 | 38.898 | 38.898 | 19.449 | 704.57 | 0.000 |
| Error | 75 | 2.070 | 2.070 | 0.028 | | |
| Total | 77 | 40.968 | | | | |

**e)** The group standard deviations are much closer after the transformation than before; after the transformation they are: 0.1267, 0.1736, 0.1914 for Bounty, Comfort, and Decorator respectively. Comparing the residual graphs, the versus order graph shows a pattern before the transformation, in that all the residuals from the Bounty group are much larger than that from any other group, but after the transformation these residuals appear more random. The normal probability plot before the transformation is somewhat curved at the tails, while after the transformation this goes away. The transformation data should be used to state a conclusion because the data better fits the model assumptions.

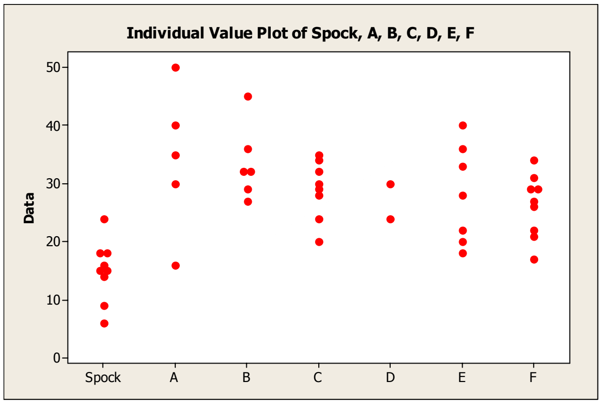



**f)** If the students had selected one roll of each brand of paper towels, and then randomly selected sheets from that roll, the results would only hold for those rolls of paper towels and not the greater population of paper towels, as it is possible that those particular paper towel rolls were different in some way from the general population of those brands.

**g)** The results would hold for the entire population of these brands of paper towels from which these samples were selected. However, caution is still needed. For example, there may be different production facilities or distribution centers for a brand of paper towel for different cities or regions, and the same brand of paper towel produced in another plant may differ. If a single distribution center supplied all those stores for which students selected a particular brand, the results would hold for that center. However, the results would not hold for the entire population of paper towels for each brand unless the paper towels they randomly choose were representative of the entire population.

**E.8** **a)** Looking at individual value plots and box plots, it appears that the center for Spock's Judge is lower than that of the other judges, and Judge A appears to have a higher center than the other judges do. The spread of Judges A and E is greater than that of the other judges, although the unequal sample sizes (especially the very small sample size for Judge D) make this difficult to compare.



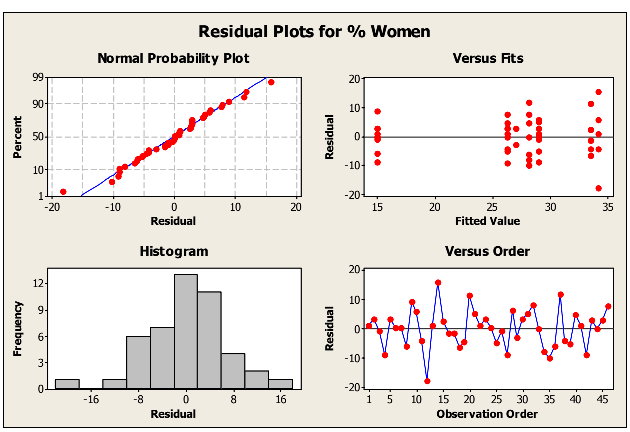

**b)** The ANOVA test shows there was a difference between the group means. A p-value of less than 0.001 indicates that the null hypothesis that all the group means are equal should be rejected.

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|---|---|---|---|---|---|---|
| Judge | 6 | 1849.72 | 1849.72 | 308.29 | 6.28 | 0.000 |
| Error | 39 | 1914.71 | 1914.71 | 49.10 | | |
| Total | 45 | 3764.43 | | | | |

**c)** We should ask the defense counsel how it collected the data. If the data was a random sample of other judges in that area, and if all recent venire data was collected or not. If it was not a random sample of judges in the Boston area, and didn't include data on all recent venires, then the data could be biased and would not be acceptable to use in court.

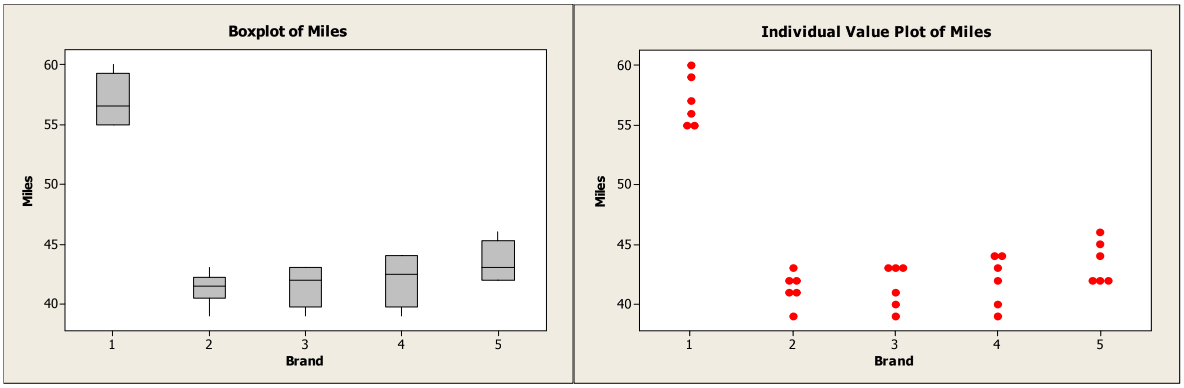
**d)** Based on the ANOVA test, at least one of the group means is significantly different from the others. However, it doesn't indicate which one is different from the others. While the graph indicates that Spock's Judge has a different mean from the others, the p-value only relates to the null hypothesis (the comparison of the equality of all group means).

The ANOVA test may not be valid, as the variances are not roughly equal for the different judges, as Judge A has a standard deviation of 12.58 which is twice that of most of the other judges. In addition, the normal probability plot shows the residuals forming a straight line albeit with some outliers.



It may be more appropriate to simply conduct a test based on two groups: Spock's Judge and all other judges.

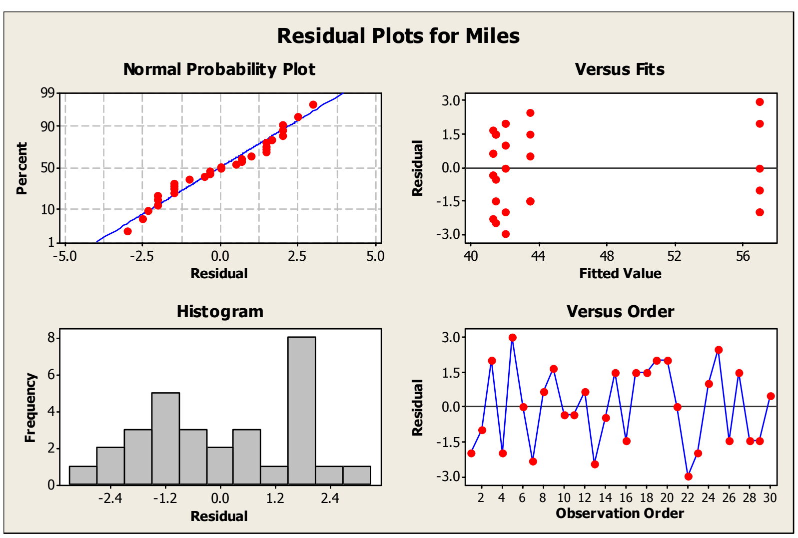
**E.9** **a)** The center of brand 1 is higher than any of the others, while the center of brand 2 is the lowest. There may be some skewness in the brands, but it is difficult to be certain with such small data sets. The spread of the five brands appear roughly similar; there don't seem to be any clear outliers.




**b)** The ANOVA test shows that a significant difference exists between the group means, based on a p-value of less than 0.001.

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|---|---|---|---|---|---|---|
| Brand | 4 | 1085.53 | 1085.53 | 271.38 | 80.45 | 0.000 |
| Error | 25 | 84.33 | 84.33 | 3.37 | | |
| Total | 29 | 1169.87 | | | | |

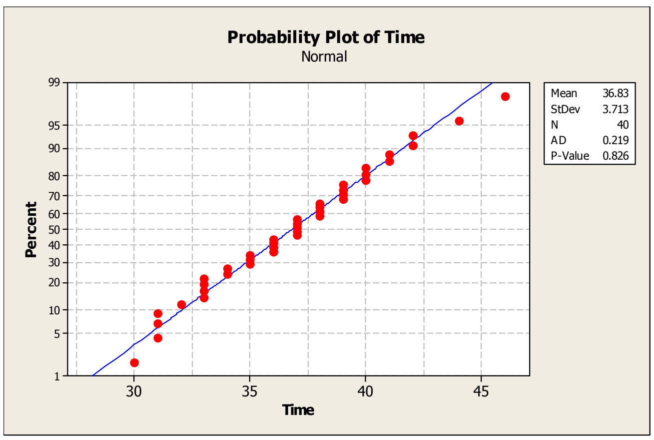
**c)** The group standard deviations are 2.098, 1.366, 1.761, 2.098, and 1.761, respectively, for the five groups. Therefore, the standard deviations are roughly equal. The residual versus order plot doesn't have any practical meaning in this example, because the true order is not given. The data are only listed by each brand. The histogram of the residuals is non-normal, however, the normal probability plot appears to be somewhat normal, but is slightly s-shaped. So, it would appear that the normality assumption may not be met.



**d)** Although we could technically treat ‘brand’ as a quantitative explanatory variable, there is no reason to expect a linear relationship between brands. Forcing the group means to form a straight line (as in simple linear regression) would cause the predicted values for each group to be much less accurate.

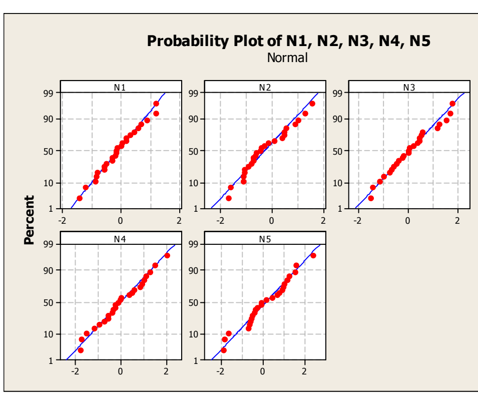
**E.10 a)** Looking at the residual plots from the data, the histogram doesn’t appear to be normal, however that could be a result of binning. Looking at the normal probability plot, the data appear to be roughly normal, though the residuals seem to fade a little at the end points. Since time is listed by the nearest second (responses appear somewhat grouped), there are clusters of points in the probability plot.

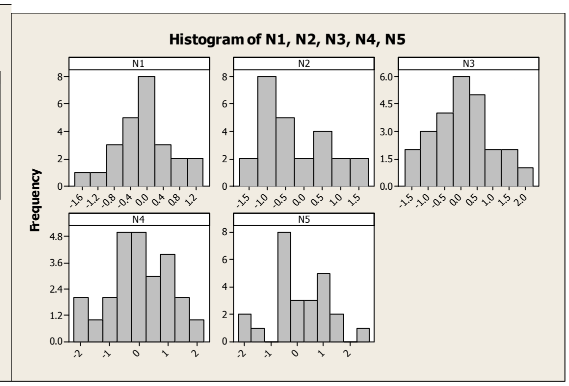
**b)**



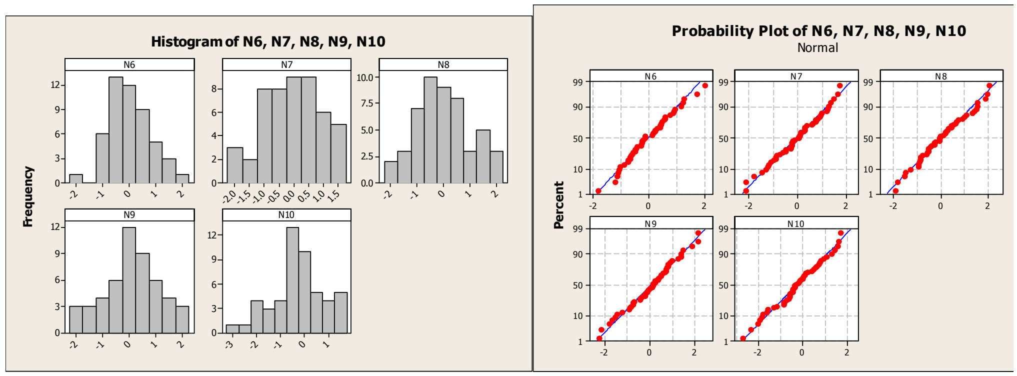

**c)** Residuals should be used to test normality instead of the observed data. The observed responses could be clustered at certain points, because each group could have a different mean.

**E.11 a)**



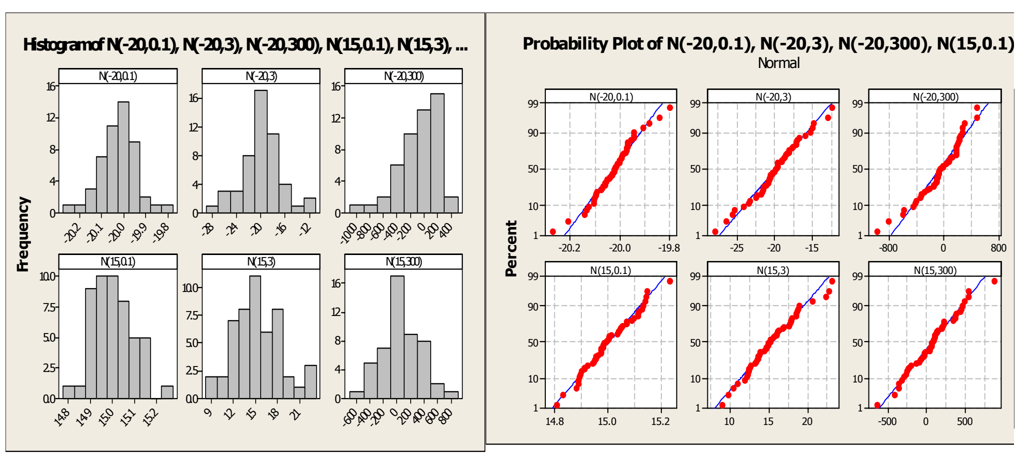



**b)**

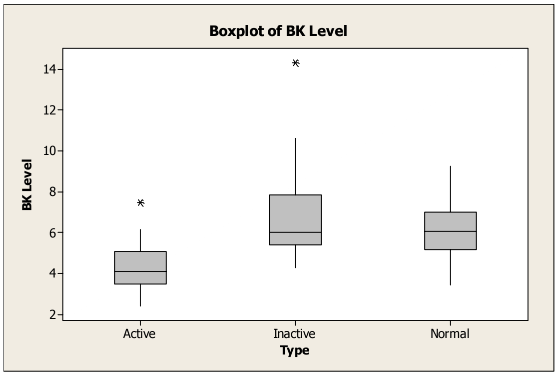



**c)** In general, larger sample sizes improve our ability to determine whether a sample is truly from a normal distribution. With larger sample sizes, the data tend to appear more normal. However, note that in Part (b), even with a sample size of 50 from a normal distribution, it is possible for the histogram to look skewed.

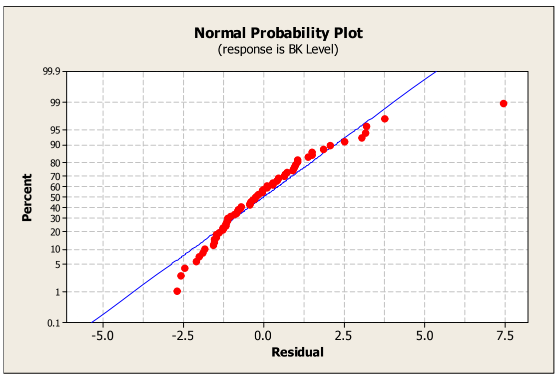
**d)** The mean and standard deviation do not affect the normal probability plots because they are still samples pulled from a normal distribution regardless of the mean and standard deviation. To be considered normal does not require that a population has a specific mean and standard deviation, but rather an appropriate shape. Note that there is a linear relationship between a standard normal and any other normal distribution with any general mean and variance. For a normal random variable Y with mean $\mu$ and standard deviation $\sigma$, $Y = \mu + \sigma X$, where X is assumed to be a random variable from the standard normal distribution.



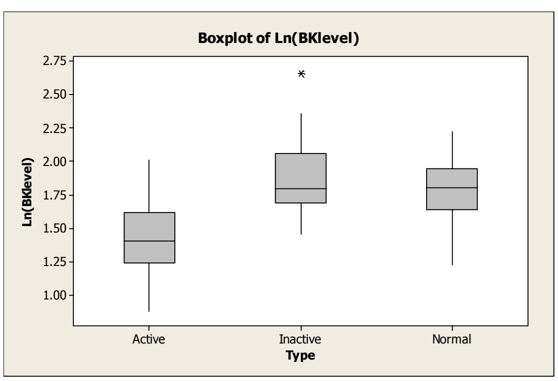

**E.12 a)** Looking at box plots of the data, there are clear outliers for the Active and Inactive groups, and both groups are right-skewed. The standard deviation for the inactive group is larger: 2.19, versus 1.318 and 1.392, for the active and normal groups respectively. The mean of the Active group is slightly lower than that of the other groups: 4.306, versus 6.857, 6.095 for the Inactive and normal groups respectively. The data do not appear to follow a normal distribution for the inactive or active groups.



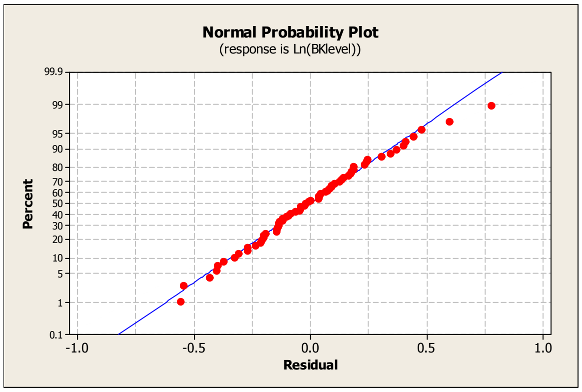

**b)** Looking at the normal probability plot, there is one large outlier, and the error terms on the whole do not seem to fit the line very well; there is a general curve to the data. The residuals do not seem to follow a normal distribution.




**c)** After transforming the data, the standard deviations are closer (0.2951, 0.2809, and 0.2357 for the active, inactive, and normal groups, respectively). The means for the three groups are also closer than before. Looking at the boxplots of the data, the Active group appears to be more normal than before, but the Inactive group is still right-skewed, appears to have an outlier, and is possibly non-normal despite the transformation.



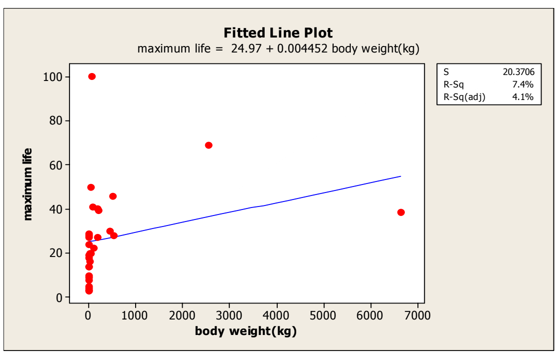

**d)** After the transformation, the error terms look more normal than before; though there are a few points that are not on the line, it seems to be roughly normal.

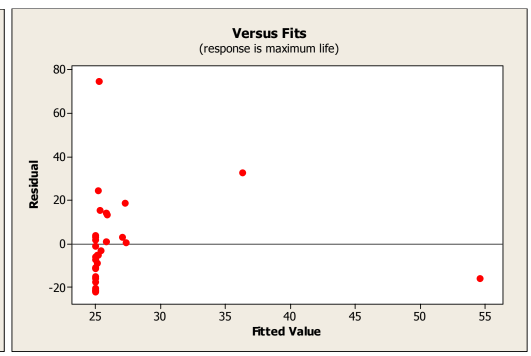


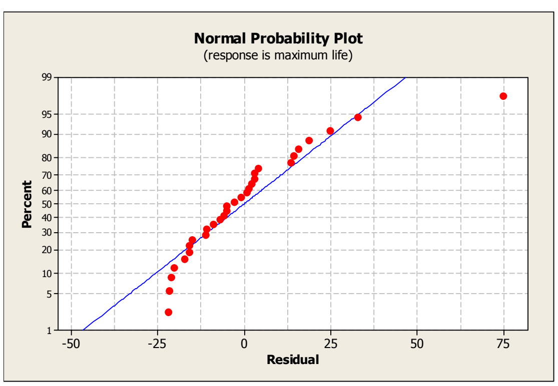

**e)** The ANOVA test gave a p-value of less than 0.001, which means we should reject the null hypothesis, and that at least one of the group means is different from the others.

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|---|---|---|---|---|---|---|
| Type | 2 | 2.2526 | 2.2526 | 1.1263 | 15.44 | 0.000 |
| Error | 62 | 4.5238 | 4.5238 | 0.0730 | | |
| Total | 64 | 6.7765 | | | | |

**E.13 a)**

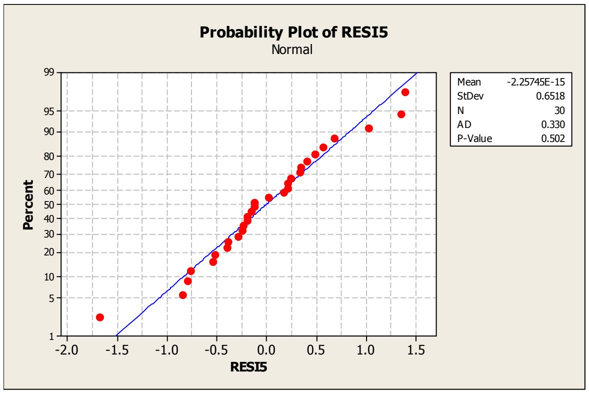


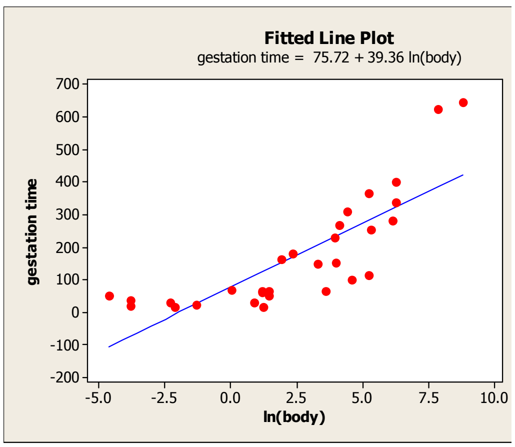



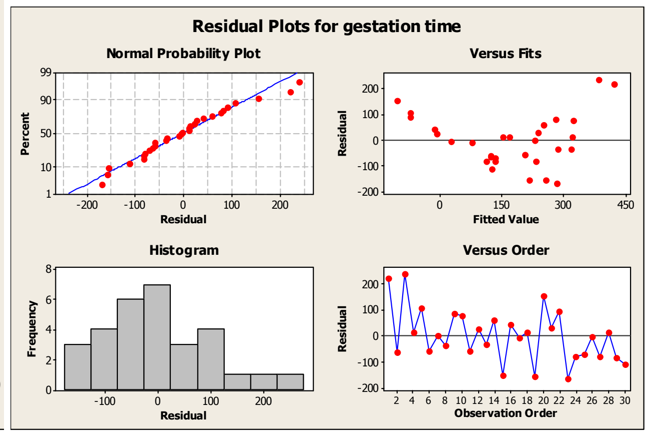

**b)** Answers will vary. Ln(life) = 2.56 + 0.169 LN(Body)

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 2.5602 | 0.1485 | 17.24 | 0.000 |
| LN(Body) | 0.16896 | 0.03527 | 4.79 | 0.000 |

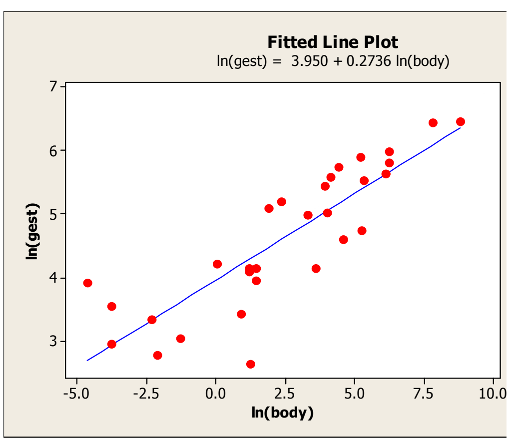


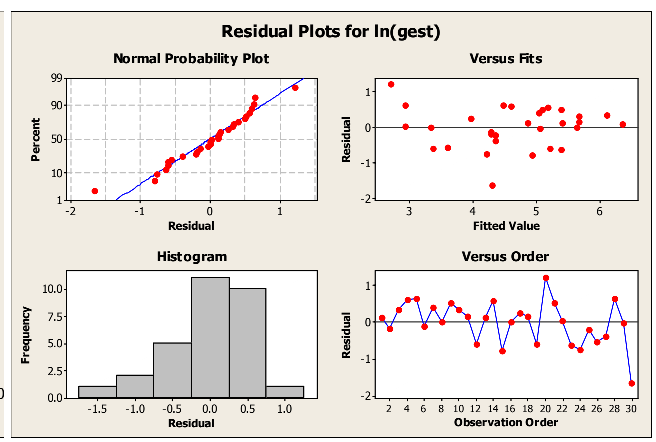

**c)** Selected models may vary. The model gestation time = 75.7 + 39.4 LN(Body)with an R-Sq = 64.7 is an improvement over the original data, however there is some curvature in the scatterplot and residuals.



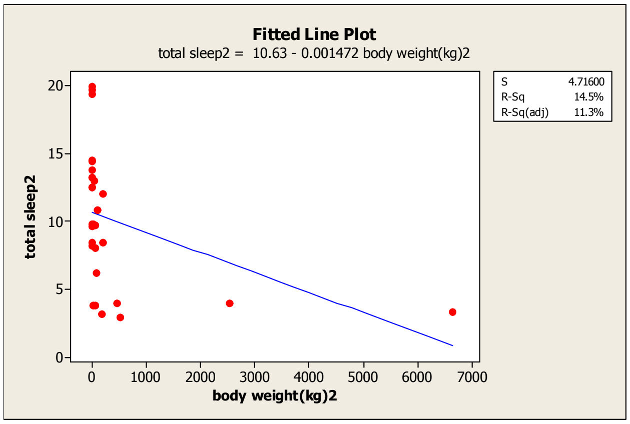



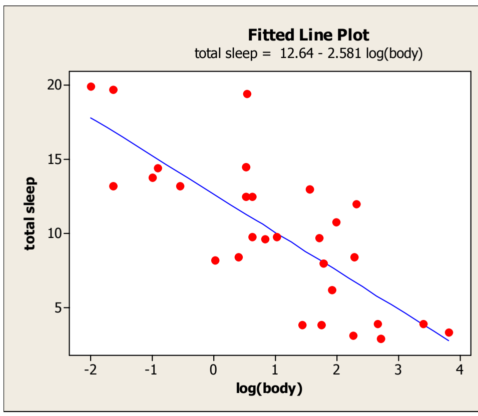

While the model may still not fit well when ln(body) is small, using ln(gest) = 3.950 + 0.2736 ln(body)with an R-Sq = 73.2% is likely to be a better model.

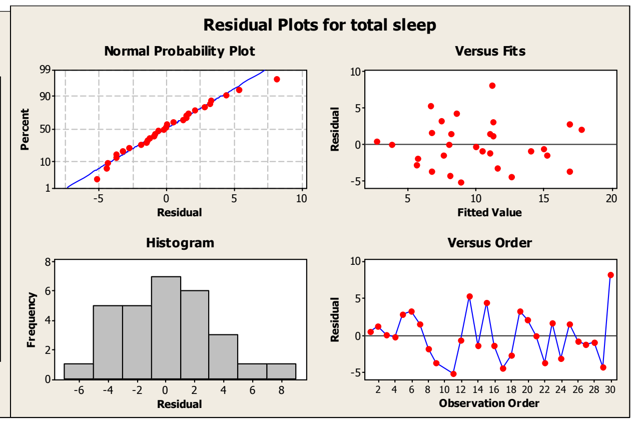



**d)** Here, we have removed the data point for which there was no data on sleep time. One solution is total sleep = 12.64 - 2.581 log(body) with an R-Sq = 60.5%:




**E.14 a)** Fail to reject the null hypothesis at the 0.05 significance level. Based on this, there does not appear to be a difference between the means of the two groups. In other words, the data are consistent with the null hypothesis that there is no difference between the two group means.

```
Two-sample T for ColorLeft vs StandardLeft (assuming equal variances)
T-Value = 2.04  P-Value = 0.057  DF = 18
```

**b)** The p-value was 0.006 so we would reject the null hypothesis, as the difference in the group means is significant at the 0.05 significance level. Based on this, there is a difference in the group means.

```
Two-sample T for ColorLeft vs ColorRight(assuming equal variances)
T-Value = 3.10 P-Value = 0.006 DF = 18
```

**c)** Fail to reject the null hypothesis, as the difference in the means is not significant at the 0.05 significance level.

```
Two-sample T for StandardLeft vs ColorRight (assuming equal variances)
T-Value = 0.77 P-Value = 0.450 DF = 18
```

**d)** The other three possible t-tests for differences in mean completion time would be between color right and standard right; between standard left and standard right; and between standard right and color left.

**e)** If each of these tests used an alpha level of 0.05, then the probability that at least one of the tests will inappropriately reject the null hypothesis is 1-0.95^6 = 26.5%. Note, however, that the solution assumes that each test is independent, but it is likely these tests are dependent (it is reasonable to assume that the test of $H_0$: colorleft vs. colorright is not independent of the test $H_0$: standardleft vs. standardright). The calculation assuming independence is, at best, an upper bound on the probability.

**f)** If we use the Bonferroni method with an overall significance level of 0.10, then the individual critical value is 0.10/6=0.016667. If we test for an overall (familywise) comparison, then none of the conclusions in Parts (a) – (c) will change. The only test for which we received a p-value that was significant remains significant with a critical value of 0.016667, so using this critical value wouldn't change our conclusions.

**E.15 a)**

| Transformation | None | SQRT | LN | Log base 10 | Reciprocal |
|---|---|---|---|---|---|
| p-value | 0.181 | 0.174 | 0.168 | 0.168 | 0.158 |
| Confidence interval lower | -2.967 | -0.716 | -0.733 | -0.3181 | -0.0406 |
| Confidence interval upper | 0.603 | 0.141 | 0.141 | 0.0612 | 0.2218 |

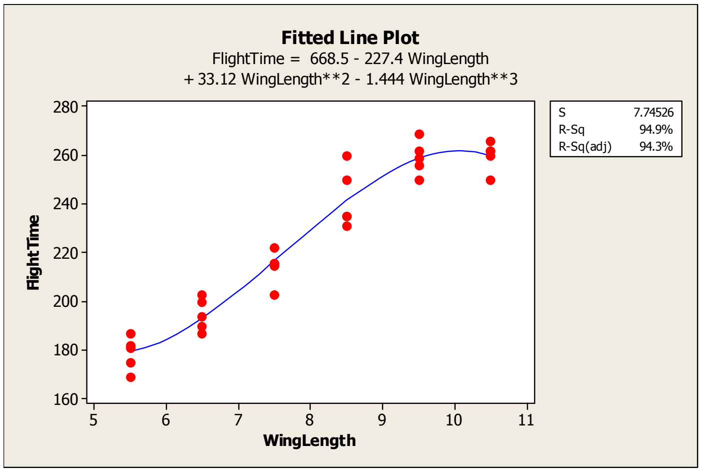
**b)**

| Transformation | SQRT | LN | Log base 10 | Reciprocal |
|---|---|---|---|---|
| Back transformed | $X^2$ | $e^X$ | $10^X$ | 1/X |
| Confidence interval lower | 0.512655 | 0.4804 | 0.480728 | -24.6305 |
| Confidence interval upper | 0.19881 | 1.15142 | 1.15133 | 4.5085 |

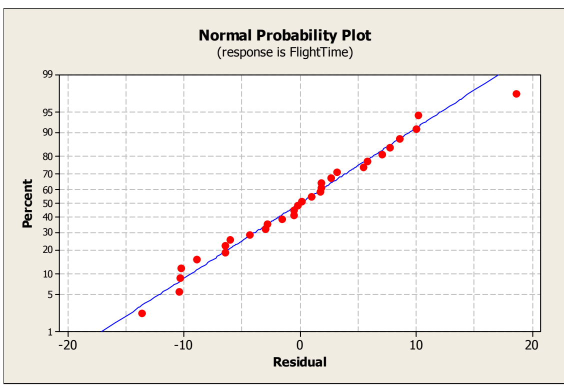
These back transformed confidence intervals tell us different things. The back transformed confidence interval of the square root transformation is meaningless, as the confidence interval no longer contains zero and now has a lower bound greater than the upper bound. The reciprocal back transformation is not very useful, either, as the bounds are too large, unrealistically so. Moreover, the log back transformations now tell us something about the ratio between the two means instead of the difference between them, though this can still be interpreted and is somewhat useful.

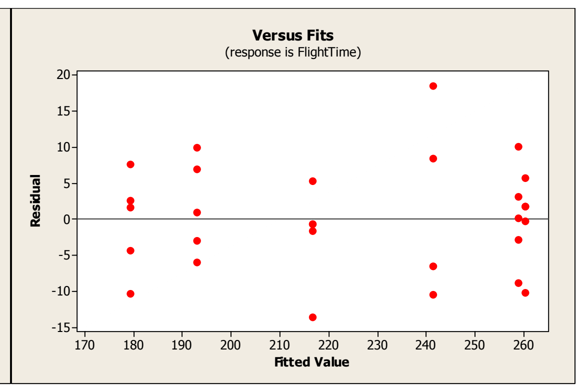
**E.16 a)** Transformations are unlikely to alleviate the Exercise.

**b)** The model obtained is Flight time=668.5-227.4(wing length)+33.12(wing length)$^2$-1.444(wing length)$^3$.




**c)** The data appear normal when we look at the normal probability plot, but when we look at residuals v versus fits, although the residuals no longer have such a pronounced curved pattern as before, the spread of the groups does not appear to be equal. The variances may be considered roughly equal.




**d)** Plotting a graph of the line, visual inspection shows that a root occurs at about x=10. In order to find the specific value that will maximize flight time, we need to take the derivative of the polynomial function and set it to zero to find the maximum value. From this, we obtain two values: 5.2042, and 10.0866 (one of these corresponds to the minimum and the other to the maximum). Thus, if the model is correct, 10.0866 is the wing length which will maximize flight time.