1a) Simply wanting to compare (as opposed to making a decision) is best done through a confidence interval. Consistency would be measured by variance or standard deviation. So, confidence interval for ratio of variances or ratio of standard deviations:

\[
\left( \frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1, n_2-1, \alpha}}, \frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1, n_2-1, 1-\alpha}} \right)
\]

\[
\alpha \text{ wasn't specified, say } \alpha = 0.05, \text{ so 95\% CI}
\]

\[
\left( \frac{2.8550^2}{3.2564^2} \frac{1}{2.27}, \frac{2.8550^2}{3.2564^2} \frac{1}{2.27} \right)
\]

\[
F_{\text{low}, d_1, d_2} = \frac{1}{F_{\text{up}, d_1, d_2}} = 2.27, 125, 4244
\]

\[
(0.3458, 1.7813)
\]

We are 95\% confident that the variance of the improvement scores of honey is between 0.3458 and 1.7813 times the variance of the improvement scores of OM.

1c) We need two iid normal random samples that are independent of each other. The random allocation helps for independence and same distribution (there is nothing that helps lack of randomization). We check normality using two Q-Q plots, one for each sample. This interval is not robust at all to non-normality and shouldn’t be used if the Q-Q plot isn’t very close to the line.
2. In a 20-meter by 50-meter test plot, each of 1000 square-meter cells was checked to see whether they contained a rare plant species, Carolina Bog-mint. After two months, each cells was checked for evidence of rooting by wild hogs. Results appear in the table below.

<table>
<thead>
<tr>
<th>Carolina Bog-mint</th>
<th>Rooting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Absent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

a. Treat the cells with Carolina Bog-mint as one population and the cells without Carolina Bog-mint as another population. What would be a sensible hypothesis for the researchers to test? Write appropriate null and alternative hypotheses.

*Answer:* I think a two-sided hypothesis is more reasonable here, but I could see that a student might think that the alternative should show the bog-mint being more affected. Ho: \( p_1 = p_2 \), HA: \( p_1 \neq p_2 \), where \( p_1 \) is the probability of cells with Carolina Bog-mint being rooted, and \( p_2 \) is the probability of cells with no Carolina Bog-mint being rooted (Make sure they define the parameters).

b. Construct an appropriate Z test statistic to test your hypothesis in (a) and compute its p-value. State your conclusion in terms of your hypotheses in (a) assuming \( \alpha = 0.01 \).

*Answer:* \( \hat{p}_1 = 20/120 = 0.167 \) and \( \hat{p}_2 = 230/880 = 0.261 \). \( Z = \frac{0.167 - 0.261}{\sqrt{0.167(1-0.167)/120 + 0.261(1-0.261)/880}} = -2.55 \). Two-sided p-value is \( 2 \times 0.0108 = 0.0216 \). If \( \alpha = 0.01 \), we fail to reject the null hypothesis of no difference in the proportion of rooting in the two types of cells.

c. If we look at the table as a 2-by-2 contingency table, the chi-square test of population homogeneity is \( X^2 = 5.0505 \), \( p = 0.0246 \). Compare to your results in (b).

*Answer:* Though the p-values are different, the conclusions would be the same for \( \alpha = 0.01 \) (fail to reject null) or \( \alpha = 0.05 \) (reject null).

d. In (c), we assumed that the row totals were fixed for our test of population homogeneity. Why might that assumption be incorrect? What other hypothesis could we test? Would results differ? Why or why not?

*Answer:* The number of cells with Carolina Bog-mint is actually random, not fixed. A test of independence of the pair of variables could be used instead. The test statistic for this test is identical to the test statistics for population homogeneity.
3. In the plot below, the total amount of wine consumed (in millions of liters)

   a. Even though the data appears linear, researchers often recommend “transforming both sides” for scatterplots showing this pattern. Explain why they might do so.

   Answer: Using the same transformation for both sides often preserves the apparent linear relationship in the plot, while reducing the effect of large values with high leverage. It can also reduce fan-shaped heteroscedasticity. Detail lost at the lower end of the scale is more apparent.

   b. The natural log for the response variable, total consumption, was regressed on the natural log for the explanatory variable, state population. The regression coefficients and their standard errors appear below.

   Interpreting the slope parameter estimate (the coefficient for lpop).

   Answer: For each unit increase in the log of State Population in millions, the mean of the log of wine consumption in millions of liters increases by 1.023.

   | Parameter | Coefficient | Standard Error | t Value | Pr > |t| |
   |-----------|-------------|----------------|---------|-------|-----------|
   | Intercept | 2.012412923 | 0.12212321     | 16.48   | <.0001 |
   | lpop      | 1.022943981 | 0.07290330     | 14.03   | <.0001 |

   c. Construct a 95% confidence interval for the slope parameter and interpret the confidence interval.

   Answer: t(.975,48)=2.0106. CI is (0.876, 1.169). For each unit increase in the log of State Population in millions, the mean of the log of wine consumption in millions of liters increases by 0.876 to 1.023.

   d. If you exponentiate the slope parameter (i.e., take its antilog), how would you interpret it?

   Answer: exp(1.0229)=2.781. For each unit increase in the log of State Population in millions, the mean wine consumption in millions of liters increases by a factor of 2.781.

   e. The residual plot for the regression is given below (lconsume represents log consumption). Comment on any assumptions that can be checked using this plot.

   Answer: The up and down spread from left to right looks fairly constant, it is a bit wider spread in the middle – but we would expect some of that since there are more observations their than at the ends. The assumption of equal variances of the errors is thus approximately met. Checking that the mean of the errors is zero could be off some as it appears slightly higher for the low values than in the middle. It isn’t too egregious though. Normality and independence aren’t checked from this plot.
4a)

\[
\begin{array}{cccc}
\text{Model} & 4 & 3679.6 & \frac{3679.6}{4} = 919.9 & \frac{919.9}{99.09} = 9.28 \\
\text{Error} & 45 & 8138.5 & \frac{-3679.6 + 4458.9}{45} = 99.09 \\
\text{Total} & 49 & 8138.5 & \\
\end{array}
\]

b) \[ \sigma^2 = \sqrt{\text{MSE}} = \sqrt{99.09} = 9.547 \]

c) \[ R^2 = \frac{\text{SS}_{\text{model}}}{\text{SS}_{\text{total}}} = \frac{3679.6}{8138.5} = 0.452 \]

45.2% of the variability in the charging time of the batteries is explained by the brand.

d) With a p-value of 0.00001474 < 0.05, we reject the null hypothesis that brands A-D have the same average charging time (by using the ANOVA F-test). This doesn't tell us which brand charges quickest though! Hus's MCB could identify which charges fastest (and which were significantly slower than it). Tukey's HSD would allow each brand to be compared to each other brand (this would allow us to account for price in the decision too, for example).
5) a) You would look at the residual vs. individual predictor plots for the model without any squared term. A curved pattern would indicate a transformation might be needed.

b) Multicollinearity causes Type I and III tests to differ (one is conditioned on all variables entered above it, three on all the other variables). In the presence of multicollinearity, only the final type I and final type III tests will agree.

Type I: with a p-value of $0.01 \leq \alpha < 0.05$, we reject the null hypothesis that $B_4 = 0$ given that $X_1, X_2, \ldots, X_3$ were already in the model. That is, $X_4$ is a significant predictor of the variation in $Y$ left unexplained by $X_1, X_2$, and $X_3$.

Type III: with a p-value $0.10 \leq \alpha < 0.05$ we fail to reject $B_4 = 0$ given that $X_1, X_2, X_3, X_5, \ldots, X_5$ are all in the model. $X_4$ is not a significant predictor of the variation in $Y$ after accounting for all the other variables.
c) The model containing $X_4$ with $CP = \rho + 1$ and having the fewest variables is the one containing $X_2$, $X_4$, $X_5$, and $X_2^2$. ($CP = 3.1534 < 4 + 1 = 5$) which contains both $X_2$ and $X_2^2$ too. The best three variable model with $X_4$ has $CP = 6.1149 > 3 + 1 = 4$.

d) DFFITs identifies influential points. The typical cut-off is typically $2 \sqrt{\frac{k+1}{n}} = 2 \sqrt{\frac{5+1}{50}} \approx 0.748$

<table>
<thead>
<tr>
<th>Obs</th>
<th>DFFITs</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2.9478</td>
</tr>
<tr>
<td>9</td>
<td>1.7729</td>
</tr>
<tr>
<td>32</td>
<td>1.7798</td>
</tr>
<tr>
<td>42</td>
<td>1.8533</td>
</tr>
</tbody>
</table>

Note that DFFITs won't detect pairs of influential points that have similar X values, and can change when any single point is removed.

This obs has

$X_1 = 8.53$ $X_2 = 9.85$ $X_3 = 7.62$ $X_4 = -1.27$ $X_5 = -1.84$

In general, it is always acceptable to run the analysis with and without the outlier and report both results.

As this observation has very extreme predictor values, it may also be possible to restrict the range of predictor values used in the regression.
6. Engineers extracted two cores from a bridge deck for each combination of its three spans and three positions (right side of lane, middle of lane, left side of line). The compression strength of the cores was measured in megapascals. A plot appears below as a visual aide.

a. Write down an interaction model for this two-factor experiment, listing all assumptions.

\[ Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}, \]

where \( i = 1, 2, 3; j = 1, 2, 3; k = 1, 2. \) \( \epsilon_{ijk} \) are iid \( \text{Normal}(0, \sigma^2) \). They should put constraints on the parameters too.

b. Results from Type I tests are included below. Discuss test results. How would you expect these results to compare to the Type III tests?

Answer: There is no evidence of interaction, which is somewhat consistent with the plot from (a). Both main effects are significant. I would expect the results to be identical since the design is orthogonal.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>2</td>
<td>19.30777778</td>
<td>9.6538889</td>
<td>12.96</td>
<td>0.0022</td>
</tr>
<tr>
<td>Span</td>
<td>2</td>
<td>109.2044444</td>
<td>54.6022222</td>
<td>73.29</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Position*Span</td>
<td>4</td>
<td>2.5322222</td>
<td>0.6330556</td>
<td>0.85</td>
<td>0.5283</td>
</tr>
</tbody>
</table>

c. Based on the results above, is it permissible to study multiple comparisons for the main effects of Position and Span? Why or why not?

Answer: Given that the interaction term is not significant, studying multiple comparisons is permissible. We have the option of refitting the model after dropping the interaction term, but that would be data-snooping. Note that if we did so, the numerator for the F test would be unchanged due to orthogonality.

d. Suppose you wanted to test whether cores from the middle of the lane had a greater compressive strength than the average of cores from the two sides. Write down a hypothesis to test this and explain in general terms how you would test this hypothesis.

Answer: \( H_0: \mu_{\text{middle.}} - (\mu_{\text{right.}} + \mu_{\text{left.}})/2 = 0 \) vs. \( H_o: \mu_{\text{center.}} - (\mu_{\text{right.}} + \mu_{\text{left.}})/2 \neq 0. \) We could test with a CONTRAST statement in PROC GLM. Other options are ESTIMATE in PROC GLM, or LSMESTIMATE in PROC GLIMMIX. Here are results:

contrast 'middle vs edges' position -0.5 1.0 -0.5;

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>middle vs edges</td>
<td>1</td>
<td>15.34027778</td>
<td>15.34027778</td>
<td>20.59</td>
<td>0.0014</td>
</tr>
</tbody>
</table>