Question 1

(i) Good idea since a log transform linearizes the data.

(ii) 
\[ t = \frac{-0.213}{0.0806} = -2.643, \]
\[ .01 < P < 0.05, \text{ since, from tables, } t_{df=28,0.975} = 2.048, \ t_{df=28,0.995} = 2.76 \]

The F-statistic has 3 and 28 df.

(iii) The test is highly sig \((F = 72, P \approx 0)\). At least one of the coefficients is not qual to zero.

(iv) \(H_{\logdisp} : \beta_{\logdisp} = 0\) given that loghp and logwt are already in the model.

\[ t = -0.77 \text{ is not significant } (P = 0.45). \]

\(H_{\loghp} : \beta_{\logdisp} = 0\) given that logdisp and logwt are already in the model.

\[ t = -2.643 \text{ is highly significant, } (P = 0.013). \]

\(H_{\logdisp} : \beta_{\logdisp} = 0\) given that loghp and logwt are already in the model.

\[ t = -3.4, \text{ is significant, } (0.01 < P < 0.05, \text{ but close to 1\% level}). \]

(v) There is a relationship between logmpg and at least one of the three variables. However, since logdisp is not significant, this variable gives little extra information about logmpg once loghp and logwt are fitted and can be omitted from the model.

(vi) The statement is NOT true. The test shows that once loghp ang logwt are fitted there is not enough additional information provided by logdisp to warrant it’s inclusion in the model, (from the graph we can see that logdisp is correlated with both of these predictors). However it is also evident from the graph that there is most likely a linear relationship between logmpg and logdisp.

Question 2  \((These \ solutions \ are \ for \ the \ modified \ version \ of \ the \ 2001 \ paper \ on \ the \ STAT200 \ web \ page.)\)

(i) There is evidence that not all three slopes are equal, \((F = 8.67, P = 0.001313)\).

(ii) (a) Four cylinder cars:

\[ \logmpg = 40.87 - 0.135 \text{ disp} \]

Six cylinder cars:

\[ \logmpg = (40.87196 - 21.78997) + (-0.13514 + 0.1387150)\text{disp} \]

\[ = 19.08 + 0.0036\text{disp} \]
Eight cylinder cars:
\[
\log\text{mpg} = (40.87196 - 18.83916) + (-0.13514 + 0.11551)\text{disp}
\]
\[
\log\text{mpg} = 22.03 - 0.0196\text{disp}
\]

(b) \(H: \beta_{4\text{cyl}} - \beta_{6\text{cyl}} = 0\)
The test of this hypothesis is highly significant, \((t = 3.817, P = 0.0008)\).

\(H: \beta_{4\text{cyl}} - \beta_{8\text{cyl}} = 0\)
The test of this hypothesis is highly significant, \((t = 3.909, P = 0.0006)\).

(c) 5% CI for \(\beta_{4\text{cyl}}:\)
\[
-0.13514 \pm 2.06 \times 0.02791 \quad \text{where } t_{26,0.975} = 2.06
\]
That is the 95% CI is given by \((-0.193, -0.078)\)

(d) The CI for the four cylinder cars does not contain zero nor does it overlap with the CI’s from the 6 and 8 cylinder cars. Hence the slope of 4 cylinder cars is unlikely to be zero and is different from the slope for both 6 and 8 cylinder cars. In addition there is almost complete overlap between the CI’s for 6 and 8 cylinder cars so it is unlikely the slopes for 6 and 8 cylinder cars are different.

(iii) Four cylinder cars:
\[
\log(\text{mpg}) = 40.87 - 0.135 \times 150 = 20.60
\]
From the graph 150 is a borderline displacement falling between 4 and 6 cylinder cars and is probably a reasonable estimate as is 19.6 for 6 cylinder cars. However 150 is well outside the observed displacements for 8 cylinder cars and represents extrapolation. The prediction at a displacement of 150 cannot be trusted for 8 cylinder cars.

(iv) Residual Plot:
There are possible outliers at the top of the graph. Other wise the points seem to have a reasonably random scatter and there is little evidence that the equal variance assumption has been violated. It also appears the model fits reasonably well except perhaps for possible outliers.

Q-Q Plot:
Approximates a straight line reasonably well except perhaps for a couple of outliers. The assumption of a normal distribution appears reasonable.
Question 3

(i) Analysis of Variance Table

Response: cs

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>treat</td>
<td>(3)</td>
<td>37.58</td>
<td>(12.527)</td>
<td>(12.3)</td>
<td>2e-04</td>
</tr>
<tr>
<td>Residuals</td>
<td>16</td>
<td>16.30</td>
<td>(1.019)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(ii) H: \( \mu_C = \mu_L = \mu_E = \mu_{LE} \), (that is all four treatments are equally effective in reducing stress.)

The \( F \)-test is highly significant, \( (F = 12.3, P \approx 0) \). It appears all four treatments are not equally effective.

(iii) Contrasts:

<table>
<thead>
<tr>
<th></th>
<th>(a) Levor</th>
<th>(b) Epin</th>
<th>(c) Levor:Epin</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>E</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>LE</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

(iv) For Levor and Epin contrasts: \((-1) \times (-1) + 1 \times (-1) + (-1) \times 1 + 1 \times 1 = 0\)

hence these two contrasts are orthogonal.

Similarly we can show that the Levor and Levor:Epin contrasts are orthogonal as are the Epin and Levor:Epin contrasts.

(v) H: \( \psi_{Epin} = 0 \), (no effect of levorphanol when averaged over all levels of epinephrine).

\( F = 12.6 \) is highly significant \( (P = 0.0027) \).

H: \( \psi_{Levor} = 0 \), (no effect of epinephrine when averaged over all levels of levorphanol).

\( F = 18.2 \) is highly significant \( (P = 0.0006) \).

H: \( \psi_{Levor:Epin} = 0 \), (the effect of levorphanol is the same whether epinephrine is present or absent).

\( F = 6.05 \) is significant \( (P = 0.026) \).

(vi) (Since the interaction is significant we only need to interpret the interaction for which we look at the interaction table of means.)

The effect of levorphanol is not the same when epinephrine is present as when it is absent \( (P = 0.026) \). From the interaction table it appears that levorphanol reduces
the cortical sterone level in both cases but the reduction is greater when epinephrine is present. (Ephineprine appears to have the effect of actually increasing the cortical sterone level.) The best result appears to be when levorpanol is used alone.

**Question 4**

(i) Completed Table:

<table>
<thead>
<tr>
<th>Df</th>
<th>Mean Sq</th>
<th>Expected Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>fertilizer (2)</td>
<td>12.15514</td>
<td>$\sigma^2 + 3\sigma_{fert:loc}^2 + 12Q_1$</td>
<td>78</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>location (3)</td>
<td>4.60598</td>
<td>$\sigma^2 + 3\sigma_{fert:loc}^2 + 9\sigma_{loc}^2$</td>
<td>29</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>fertilizer:location (6)</td>
<td>0.15544</td>
<td>$\sigma^2 + 3\sigma_{fert:loc}^2$</td>
<td>2.68</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Residuals</td>
<td>24</td>
<td>0.05793</td>
<td>$\sigma^2$</td>
<td></td>
</tr>
</tbody>
</table>

(ii) H:\(fert: \mu_1 + \mu_2 = \mu_3\)

Test is highly significant, \((F = 78, P < 0.001)\).

H:\(location: \sigma_{loc}^2 = 0\)

Test is highly significant, \((F = 29, P < 0.001)\). H:\(fert:loc: \sigma_{fert:loc}^2 = 0\)

Test is significant, \((F = 2.68, P < 0.05)\).

(iii) Components of Variance:

\[
\hat{\sigma}^2 = 0.058
\]
\[
\hat{\sigma}_{fert:loc}^2 = \frac{0.15544 - 0.05793}{3} = 0.033
\]
\[
\hat{\sigma}_{loc}^2 = \frac{4.60598 - 0.15544}{9} = 0.495
\]

(iv) \(\text{Var}(y) = \hat{\sigma}^2 + \hat{\sigma}_{loc}^2 + \hat{\sigma}_{fert:loc}^2 = 0.058 + 0.498 + 0.033 = 0.586\)

(v) There is a significant linear trend \((P = 3.32 \times 10^{-5})\). Once the linear trend is removed there is also a quadratic trend \((P = 1.65 \times 10^{-11})\). There is an overall decline although the best yield is at the second level of fertilizer which explains the quadratic nature of the trend.

(vi) Conclusion: It would appear that the second level of fertilizer leads to the best yield. We also see that a major source of variability in yields is the different locations with random variation and additional variation due to the interaction of fertilizers and location only representing a small amount of the overall variation.