1. (a) (i) infinitely many

(ii) \( X \) is a \( n \times 6 \) matrix \( \Rightarrow \) rank \( X \leq 6 \)
\( X \) is not of full rank (21 solution to normal equations), i.e. \( \text{rank}(X) \leq 6 \)

(iii) \( \hat{\beta}_3 \) \( (\hat{\beta}_3 = 25 \text{ or } 36, \text{ not unique}) \)
\( (\hat{\beta}_3 + \hat{\beta}_4) \) [estimate = 43 \text{ or } 43] OK
\( (\hat{\beta}_3 - \hat{\beta}_4) \) [estimate = 7 \text{ or } 29, \text{ not unique}]

(iv) A. \( \beta_3 = 0 \) (say) \( \text{or } (\hat{\beta}_3 - \hat{\beta}_4) = 0 \) (say) \( \text{not estimable} \)
B. \( \beta_3 + \beta_4 = 0 \) (say), \( \beta_3 + \beta_4 \text{ is estimable} \)

(b) (i) \( t = \frac{\hat{\beta} - \beta}{\text{se} (\hat{\beta})} \)
\( \text{se} (\hat{\beta}) = \sqrt{\text{var}(\hat{\beta}) + \text{var}(\beta) - 2 \text{ cor}(\hat{\beta}, \beta)} \)
\( \text{Could be obtained directly from disp.} (\text{lm}(y - x)) \)
\( \text{or from disp.} \cdot (\text{lm}(y - x)) \)

(ii) \( \text{deviance (} \text{lm}(y - x) \text{)} = d\hat{\beta}_1 \) on \( (n-1) \) d.f.
\( \frac{\text{deviance (} \text{lm}(y - I(1+x) - 1) \text{)} = d(\hat{\beta}_0)}{d(\hat{\beta}_1)/(n-1)} \)
Q4.2 (a) Using the raw data, a plot of (standardised) residuals versus fitted values showed "fanning" - non-constant variance (of the errors). After a log transformation (of the response variable) the corresponding plot appeared to be OK - constant error variance.

(b) Test for interaction:
\[
F = \frac{(10.57 - 7.34)^2}{7.34/2.3} = 1.27 \quad \text{not sig}
\]

(c) Block: \[
F = \frac{(14.24 - 10.57)^2}{10.57/31} = 5.38^* \]
Substance: \[
F = \frac{(77.77 - 10.57)^2}{10.57/31} = 49.27^{***}
\]

Substance
\[
\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
0 & -1.376 & -2.453 & -3.58 & -3.701
\end{array}
\]

No sig. diff. between S3 & S4, or between S4 and S5; all other differences are sig. sig.
Effects, relative to S1, are given above.

(1) Desirable: \{ better understanding \}

Check claim interpolation

Substance as a factor is equivalent to fitting a quadratic in \( \text{Flw} \). If simpler model, linear or quadratic, (or maybe cubic) is OK then it is worth pursuing.

If claim is true, then would expect some of the higher order terms to be NS.
Q. 3. (a) (i)  

<table>
<thead>
<tr>
<th>Model</th>
<th>Parametric Form</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$</td>
<td>47</td>
</tr>
<tr>
<td>2</td>
<td>$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>$y_{ij} = \mu + \alpha_i + \beta_2 x_{ij} + \epsilon_{ij} + \epsilon_{ij}$</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>$y_{ij} = \mu + \alpha_i + \beta_0 x_{ij} + \beta_1 x_{ij} + \epsilon_{ij}$</td>
<td>42</td>
</tr>
</tbody>
</table>

(ii) Models 2, 4 and 5.  
(iii) Models 2 and 3 (only).  
(iv) Test for interaction: 

$$F = \frac{(45.792 - 37.8540)/2}{37.8548/42} = 4.403$$  

Interaction is sig. no need for any further test.

(b) (i) Model 5  
(ii) $\beta_2$ (only)  
(iii) $H_0: \alpha_1 = \alpha_2$  

$$t = \frac{544.9737 - 347.9560}{143.15735} = 1.376 \text{ (ns)}$$  

$H_0: \beta_1 = \beta_3: \quad t = \frac{1.8226 - 1.4136}{0.950177} = 0.430 \text{ (ns)}$

(iv) Fit $\ln(y \times \text{treat2.f} \times d)$ where treat2.f is a factor with two levels: $[1, \text{for treatment 1} \text{ and } 2, \text{for treatment 2}]$

$$F^* = \frac{(5.00431 - 37.8548)/2}{37.8548/42} = 6.761 \text{ sig.}$$
Leverage - how "unusual" is the (set of) explanatory variable(s) for each observation?

Standardized residuals - \[ \frac{e_i}{\sqrt{\hat{\sigma}^2 (1 - h_{ii})}} \] - residuals with constant variance, and variance \( n \) - allow for leverage (Ans)

Cooks' distance - effect of each observation on \( \hat{\beta} \)
[CD \leq 0.5 probably OK, CD > 1.0 not OK]

Observation 15: very unusual test statistic for leverage leading to a large influence on \( \hat{\beta} \), even though the point does not have an especially large standardized residual (12.5 \times 1.0).

(b) Summary which gave p-value \( \geq 0.05 \). Not inconsistent since test in summary allows for contribution of \( x_2 \) and \( x_3 \) whereas test in ANOVA does not.

(c) High leverage caused by unusual value of \( x_1 \).

(d) Estimate:
\[
2.97991 - 2 \times 0.92921 + 0.5 \times 1.8274
\]
\[
= 2.032
\]

\[
P_i : \text{est. } \pm \frac{0.975}{22} \sqrt{0.1911^2 + \text{var}(\text{fit})}
\]
\[
= 2.032 \pm 0.274 \times 0.2377 = 2.032 \pm 0.493 = (1.539, 2.525)
\]

\[
\text{var}(\text{fit}) = 0.00149 + 4 \times 0.00295 + 0.01029
\]
\[
= 0.0199865
\]
Q5. (a) (ii) Interaction plots for each pair of factors.

(ii) Source  df  SS  MS  F  P
  Cleanser  3  362  121  38.38  0.001
  L (Lines)  3  63802  21267  6745.48  0.001
  T (Time)  1  6229  6229  1975.72  0.001
  CL  9  12  1.33  1.33  0.705
  CT  3  4  1.33  1.33  0.705
  LT  3  529  176.33  4.79  0.001
  CLT  9  6  0.67  0.67  0.705
  Residual  288  908  3.15
  Total  319

[Residual (revised): 309  930  3.09]

(iii) Why? To determine which levels do differ significantly.

For Lines * Time, need to consider differences between
Lines within each Time, and Time within each Line, using
L x T means averaged over replicates and Cleaners.

Cell entries are means of 40 observations.

Hence se(diff between any pair of means) =

\[ \sqrt{\frac{3.15}{40 + 40}} = 0.397 \]

For Lines (within each Time) \[ \text{LSD}_{0.05} = 0.397 \times \frac{0.788}{\sqrt{40}} = 0.724 \]

For Time (within each Line) \[ \text{LSD}_{0.05} = 0.397 \times \frac{0.788}{\sqrt{40}} = 0.724 \]

For Cleaners, use overall Cleaner means with

\[ \text{LSD}_{0.05} = 0.724 \times \frac{0.788}{\sqrt{40}} = 0.724 \]

(b) Source  df  SS  MS  F  P
  C  3  362  121  21.94  0.001
  L  3  63802  21267  3856.99  0.001
  CL  9  12  1.33  1.33  0.705
  Samples(CL) 114  794  5.514
  Samples 159  64970
  T  1  6229  6229  786.8  0.001
  CT  3  4  1.33  1.68  0.05
  LT  3  529  176.33  4.79  0.001
  CLT  9  6  0.67  0.67  0.705
  Residual 114  114  0.79
  Total 319
(a) Subject random
   Time fixed

(ii) \( D \times T, S \times T, S(D) \)

(iii) Source | df | \( F \)  
\[ \begin{array}{ccc}
\text{D} & 2 & 0/1 \\
S(D) & 18 & 0/0 \\
T_{D \times T} & 1 & 0/0 \\
T_{S(S(D))} & 18 & 0/0 \\
\text{Resid} & & \\
\text{Total} & 41 & \end{array} \]

(iv) \( D, T \) and \( D \times T \)

- \( D \times T \) since expect no (positive) diff between times for the placebo
- but expect different for drugs A and B.

\[ \begin{array}{c}
\text{Initial} \\
\text{Final} \\
\text{A} \\
\text{B} \\
\text{P} \\
\end{array} \]

(b) \[ y_{ij} = \mu + \alpha_i + \beta_i x_{ij} + e_{ij} \]

\[ (y_{ij} - x_{ij}) = \mu + \alpha_i + \beta_i x_{ij} + e_{ij} \]

\[ \beta_i = (\beta_i + 1) \]
Q 7. (a) Used when there are incomplete blocks. Used to facilitate the use of blocking in order to improve precision.

Enable greater precision to be obtained on effects of interest by sacrificing information on effects of less (or no) interest.

(b) i) Confound ABCD, ABE, CDE

\[
\begin{align*}
& a \ b \ ab \ cd \ abcd \ ac \ bd \ acde \ bde \\
& a \ b \ ab \ cd \ abcd \ ac \ bd \ acde \ bde \\
& a \ abc \ d \ abd \ ae \ be \ acde \ bcode \\
& a \ abc \ d \ abd \ ae \ be \ acde \ bcode
\end{align*}
\]

Analysis: either how to assume that some effects are negligible (e.g. 3+ factor interaction) or use a (half) normal plot.

\[
\begin{array}{c|c|c}
\text{Source} & \text{df} & \text{Total} \\
\hline
\text{Blocks} & 7 & \text{Total} 63 \\
\text{Main} & 5 & \\
2-f. & 10 & \\
3-f. & 10 & \\
4-f. & 5 & \\
5-f. & 1 & \\
\text{Residual} & 25 & \\
\hline
\end{array}
\]

Confound BCDE, ABC and ADE in the extra replicate (extra 4 blocks)

Analysis: no need to assume effects negligible.