1. (a) $H_0: \mu = 12$ vs $H_1: \mu > 12$.
(b) The student should take a large number of samples, say 1000, of size 15 from the distribution, $N(\mu = 12.2, \sigma = 0.3)$. He should compute the mean for each sample and then count the number of samples that have mean $> 12$. Suppose the number is $k$. Then the estimated power is $k/1000$.
(c) $\text{Power} = \Pr(\bar{X} > 12.1|\mu = 12.2) = 0.9017$
(d) 56

2. (a) A completely randomised design.
(b) Analysis of Variance for yield

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>4</td>
<td>41.63</td>
<td>10.4075</td>
<td>9.537</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>16.37</td>
<td>1.091</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>58.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$H_0: \mu_1 = \mu_2 = \ldots = \mu_5$ vs $H_1: \text{not } H_0$.
$F = 9.537$. Null distribution is $F_{4,15}$, and $5.564 < F_{4,15}(0.99) < 6.226$.
Therefore $P$-value < 0.01.
We reject $H_0$ and conclude that there is a difference between the treatment means.
(c) 
<table>
<thead>
<tr>
<th>3</th>
<th>2</th>
<th>1</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH</td>
<td>AL</td>
<td>D</td>
<td>BH</td>
<td>BL</td>
</tr>
<tr>
<td>mean</td>
<td>30.8</td>
<td>31.9</td>
<td>34.0</td>
<td>34.3</td>
</tr>
</tbody>
</table>

(d) $\psi = \frac{1}{2}(\mu_{AL} + \mu_{AH}) - \frac{1}{2}(\mu_{BL} + \mu_{BH})$.
$\hat{\psi} = -2.98 \text{ and } se(\hat{\psi}) = 0.5223$.
95% confidence interval: $(-4.063, -1.837)$.
Since the confidence interval does not contain zero, there is a significant difference between mean response of light source A and that for B at the 0.05 level.

3. (a) Temperature with 3 levels: 24°C, 17°C and natural.
   Sex with 2 levels: male and female.
(b) There is a difference in survival rates of the male and female flies. There is not much difference between the survival rates for different temperatures. There appears to be no
interaction.

(c) $H_0$: no interaction vs $H_1$: there is interaction.

\[ F = 2.752. \quad F_{2.24}(0.95) < 3.403. \]

Therefore retain $H_0$ and conclude there is no interaction, ie. the additive model is appropriate.

(d) Additive ANOVA:

(e)

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Factor} & \text{DF} & \text{SS} & \text{MS} & \text{F} \\
\hline
\text{Cage} & 2 & 50.99 & 25.4 & 1.459 \\
\text{Sex} & 1 & 1778.7 & 1778.7 & \\
\text{Error} & 26 & 452.6 & 17.41 & \\
\hline
\text{Total} & 29 & 2282.2 & & \\
\hline
\end{array}
\]

(f) $H_0$: temperature does not affect survival rates

$H_1$: temperature does affect survival rates.

\[ F = 1.459 \text{ and } 3.403 < F_{2.26}(0.95) < 3.316. \]

Therefore retain $H_0$ and conclude that temperature does not have a significant effect on survival rates.

4. (a) The deeper the soil, the less the water content. Depth is very highly significant as a factor on soil water content as the $t$–value is $-6.67$ with a corresponding $P$–value of $\approx 0.000$.

(b) Predicted value at 100 cm: 0.264.

95% Prediction interval: (0.2341, 0.2938).

(c) $H_0: \mu(x) = \alpha + \beta x$ vs $H_1: \mu(x) = \text{arbitrary. (one-way model)}$

\[ \bar{y} = 0.287. \]

Between SS = \[ \sum_{i=1}^{4} n_i(\bar{y}_i - \bar{y})^2 = \ldots = 0.007864. \]

Within SS = total SS – between SS = \ldots = 0.00214.

\[ F = \frac{(0.0023931 - 0.00214)/(14 - 12)}{0.00214/12} = 0.7096 < F_{2,12}(0.95) = 3.885. \]

Therefore retain $H_0$, ie. simple linear regression is adequate.

5. (a) $-0.0523$. For a given speed of movement, the longer the prey, the less time it takes to catch it. On average, an increase of 1000 twips (10 units) will result in a decrease in the time to catch it of half a second.

(b) Model is $\mu = \alpha + \beta_1 \text{length} + \beta_2 \text{speed}$

$H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$.

\[ T = 6.409 > t_{16}(0.999) = 3.686. \]

Reject $H_0$ and conclude speed is significant in evading being caught.
(c) This is testing the fitted model, $H_1: \mu = \alpha + \beta_1 \text{length} + \beta_2 \text{speed}$, against the null model, $H_0: \mu(\text{length, speed}) = \mu$. This is a test of the utility of the model.

6. (a) Forward selection:

1. $x_3$ is to be considered: $F = \ldots = 28.45 > F_{in} = 4$. Therefore $x_3$ is in.
2. $x_1, x_3$ is next to be considered: $F = \ldots = 17.60 > F_{in} = 4$. Therefore $x_1$ is in.
3. $x_1, x_2, x_3$ is next to be considered: $F = \ldots = 1.689 < F_{in} = 4$. Therefore $x_2$ is not in.

Selected model is $\mu(x_1, x_2, x_3, x_4) = \alpha + \beta_1 x_1 + \beta_3 x_3 = \alpha + \beta_1 \text{length} + \beta_2 \text{speed}$.

(b) $R - \text{sq} = 0.8219$, adj $R - \text{sq} = 0.7997$, $s = 0.0786$ and $C_p = 3.408$.

(c) The number of variables is the same in the two models.

The model in this question has larger $R$-sq and adj $R$-sq (which is good) and a smaller $s$ (also good). Its $C_p$ is considerably smaller than that for the model in Question 5 and is close to $p$ (the number of parameters, $p = 3$). Hence, on all counts, this is a much better model.

7. (a) $y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + e_{ijk}$, where etcetera

(b) $\hat{\sigma}^2 = 0.004779$, $\hat{\sigma}^2_{\alpha B} = 0.005055$, $\hat{\sigma}^2_B = 0.033036$.

(c) $H_0: \sigma^2_B = 0 \; \text{vs} \; H_1: \sigma^2_B \neq 0$.

$F = 14.31 > F_{3,6}(0.99) = 9.780$.

Therefore reject $H_0$ at 0.01 level of significance and conclude that location has a significant effect on the mean weight of wethers.

8. (a) i. $\mu(i, x) = \alpha + \beta x$.

ii. not relevant.

(b) i. $\mu(i, x) = \mu + \alpha_i + \beta x + \beta_i x$, where $\sum \alpha_i = 0$ and $\sum \beta_i = 0$.

This model says that both diet and initial weight have an effect on weight gain and that there is interaction between diet and initial weight.

ii. 49.25.