Q1 a) Let β_1 be the slope in the linear relationship between 'Alive' and 'NT3'. We are testing

$H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$

The test statistic is

$$t = \frac{0.639 - 0}{0.1805} = 3.54$$

We compare the test statistic with the t_6-distribution.

$\mathbb{P}(T_6 \geq 3.143) = 0.01$ and $\mathbb{P}(T_6 \geq 3.707) = 0.005$

p-value $= 2 \times \min\{\mathbb{P}(T_6 \geq 3.54), \mathbb{P}(T_6 \leq 3.54)\}$

which is between 0.01 and 0.02. This is moderate evidence against the null hypothesis, suggesting a linear association between 'Alive' and 'NT3'.

b) The sum of the residuals is zero, since

$$\hat{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

and

$$\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

(This is one of the equations we solve to get the least squares estimates $\hat{\beta}_0$, and $\hat{\beta}_1$. See page 130 of notes.)

c) We need an estimate of the mean percentage of alive cells for NT3 concentration 25g/ml and the standard error.

estimated mean at 25g/ml of NT3 $= 55.04 + 0.639 \times 25 = 71.015$.

The standard error of this mean is computed by taking the square root of

$$\begin{bmatrix} 1 & 25 \end{bmatrix} \begin{bmatrix} 11.397 & -0.4884 \\ -0.4884 & 0.0326 \end{bmatrix} \begin{bmatrix} 1 \\ 25 \end{bmatrix}$$

≈ 7.3520 s.e.(estimate) $= \sqrt{7.3520} = 2.7115$

The 95% confidence interval is formed by

estimate $\pm$ (critical-value) $\times$ s.e.(estimate)

$= 71.015 \pm t_{6;0.975} \times 2.7115$

$= 71.015 \pm 2.447 \times 2.7115$

$\approx 71.015 \pm 6.6349$ (per cent)

We are 95% confident that the true mean percentage of Alive cells at NT3 concentration of 25ng/ml is between 64.38% and 77.65%

d) The assumptions of the linear regression model are

- linearity of mean in explanatory variable
- constant variance
- normality
- independence

Departures in the first three assumptions can be detected with these plots. There is no obvious trend or change in the spread of residuals in the residuals v fitted values plot, so first two assumptions are ok. The normal probability plot appears straight which is consistent with the normality assumption.

Q2 a) Let ρ be the correlation between CAA concentration and initial biofilm thickness.

Test $H_0: \rho = 0$ against $H_1: \rho \neq 0$.

Test statistic $\dfrac{r-0}{\sqrt{(1-r^2)/(n-2)}} = \dfrac{-0.3508}{\sqrt{\dfrac{1-(-0.3508)^2}{20-2}}} = -1.5893$

Under the null hypothesis, the test statistic has a t_{n-2} distribution

p-value $= 2 \times \min\{\mathbb{P}(T_{18} \geqslant -1.5893), \mathbb{P}(T_{18} \leqslant -1.5893)\}$

$\mathbb{P}(T_{18} \leqslant -1.5893) = \mathbb{P}(T_{18} \geqslant 1.5893)$ by symmetry

$\mathbb{P}(T_{18} \geqslant 1.330) = 0.1$ and $\mathbb{P}(T_{18} \geqslant 1.734) = 0.05$

So the p-value is between 0.1 and 0.2. This is inconclusive evidence against the null, suggesting no correlation between CAA concentrations and initial biofilm thickness.

b) Units are $\mu m / (\mu g/ml)$

c) estimated mean reduction in thickness is

$$5.002 + 0.02956 \times 75 = 7.2190 \ \mu m$$

d) Let β_1 be the true slope in the linear relationship between biofilm thickness reduction and CAA concentration. We test

$H_0: \beta_1 = 0$ against $H_1: \beta_1 \neq 0$

Test statistic

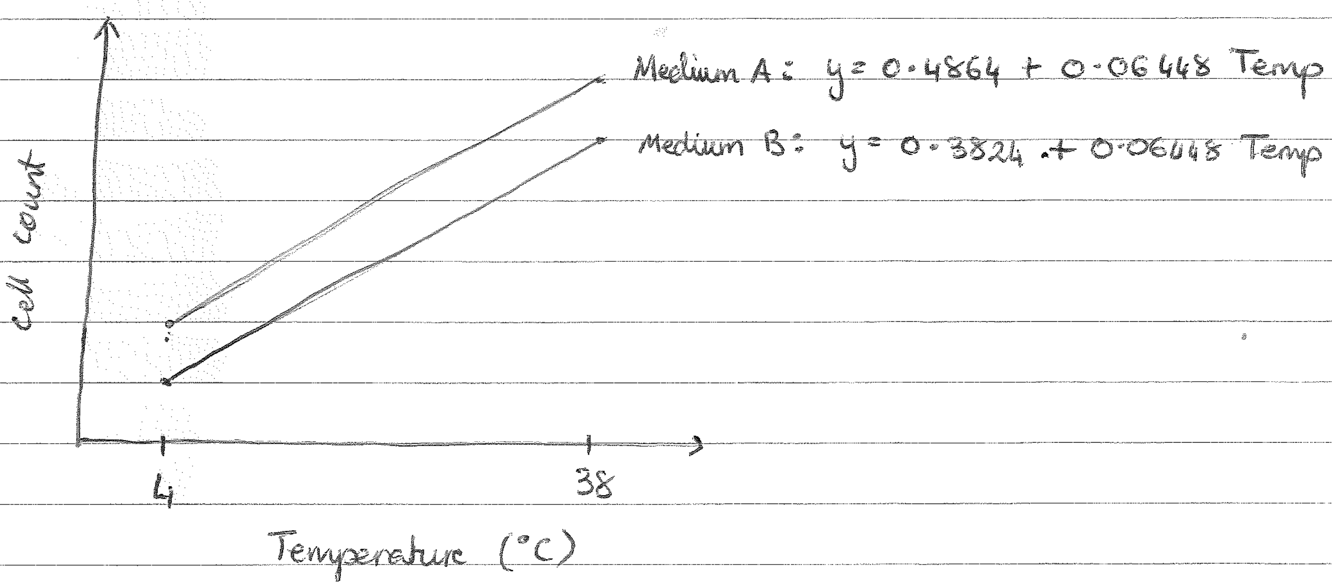
$$t = \frac{0.02956}{0.01910} = 1.5476$$

p-value $= 2 \times \min\{\mathbb{P}(T_{18} \geq 1.5476), \mathbb{P}(T_{18} \leq 1.5476)\}$

As $\mathbb{P}(T_{18} \geq 1.330) = 0.1$ and $\mathbb{P}(T_{18} \geq 1.734) = 0.05$, the p-value is between 0.1 and 0.2. There is inconclusive evidence of a linear association between CAA concentrations and the reduction in biofilm thickness.

Q3 a) The error degrees of freedom is $23 = n - p$. The number of terms in the linear regression p is 3 (intercept, temp, medium B). So the number of observations / experimental trials used in the analysis is $23 + 3 = \underline{\underline{26}}$.

b)



c) We are testing whether the coefficient for temperature in the linear relationship is zero. Denote this by β_1. We test

$H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$

Test statistic $t = \frac{0.064485}{0.001613} = 39.978$

We compare this to the t_{23}-distribution

p-value $= 2 \times \min\{\mathbb{P}(T_{23} \geq 39.978), \mathbb{P}(T_{23} \leq 39.978)\}$

$\mathbb{P}(T_{23} \geq 4.415) = 0.0001$

So the p-value is less than 0.0002. This is very strong evidence against the null hypothesis, suggesting a linear association between cell count and temperature.

d) Confidence interval: estimate ± (critical value) × s.e.(estimate)

$-0.104045 \pm t_{23;0.95} \times 0.040080$

$t_{23;0.95} = 1.714$ (from tables). So

$-0.104045 \pm 1.714 \times 0.040080$

$= -0.104045 \pm 0.0687$

We are 90% confidence the true coefficient for Medium B is between -0.1727 and -0.0353.

e) $\left(\begin{array}{c}\text{Estimated mean cell count} \\ \text{at } 30°C \text{ in Medium A}\end{array}\right) = 0.4864 + 0.064485 \times 30$

$= 2.4209$ ($\times 10^6$ cells).

f) We have the estimated mean for medium A at 30°c from part (e). We now need the standard error. The estimate for σ is 0·1267 (Root Mean Square Error in output). We need to compute $x(X^TX)^{-1}x^T$ where $x = [1 \quad 30 \quad 0]$.

$$[1 \quad 30 \quad 0]\begin{bmatrix} 0.1223 & 2.143e\text{-}05 & 5.334e\text{-}04 \\ 2.143e\text{-}05 & 1.6207e\text{-}04 & 1.186e\text{-}05 \\ 5.334e\text{-}04 & 1.186e\text{-}05 & 0.1001 \end{bmatrix}\begin{bmatrix} 1 \\ 30 \\ 0 \end{bmatrix}$$

$$= [0.1229 \quad 0.0049 \quad 0.0009]\begin{bmatrix} 1 \\ 30 \\ 0 \end{bmatrix} = 0.2694$$

The standard error for the estimated mean ~~for~~ medium A at 30°C is $0.1267 \times \sqrt{0.2694} = 0.0658$.

The 95% confidence interval is

$$2.4209 \pm t_{23;0.975} \times 0.0658$$
$$= 2.4209 \pm 2.069 \times 0.0658$$
$$= 2.4209 \pm 0.1361 \quad (\times 10^6 \text{ cells})$$

Q4 a) We assume pemax $\sim N(\beta_0 + \beta_1 \text{weight}, \sigma^2)$ and that the individuals are independent. Note that this includes
- assumption of linearity of mean in explanatory variable
- assumption of constant variance
- assumption of Normality.

b) $(A) = \dfrac{\text{Estimate}}{\text{tStat}} = \dfrac{63.5456}{5.003} = 12.7015$ (From rearranging $\text{tstat} = \dfrac{\text{Estimate}}{\text{SE}}$)

(C) Error degrees of freedom $= n - 2 = 25 - 2 = 23$

(B) $\text{p-value} = 2 \times \min\{\mathbb{P}(T_{23} \geqslant 3{\cdot}944), \mathbb{P}(T_{23} \leqslant 3{\cdot}944)\}$

As $\mathbb{P}(T_{23} \geqslant 3{\cdot}768) = 0{\cdot}0005$ and $\mathbb{P}(T_{23} \geqslant 4{\cdot}415) = 0{\cdot}0001$, the p-value is between $0{\cdot}0002$ and $0{\cdot}001$

c) estimated mean pemax $= 63{\cdot}5456 + 1{\cdot}1867 \times 50 = 122{\cdot}8806$ (units)

d) $R^2 = r^2 = 0{\cdot}635^2 = 0{\cdot}4032$.

e) Let β_1 denote the coefficient of weight in the linear relationship between pemax and weight.

Test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$

The test statistic is $3{\cdot}944$ and p-value between $0{\cdot}0002$ and $0{\cdot}001$ (from part b). This is strong evidence against the null hypothesis, suggesting a linear association between weight and pemax.

f) The normal probability plot is relatively straight indicating normal distribution for residuals. The residual vs fitted value shows some fanning / increasing variability as fitted values suggesting the assumption of constant variance is violated. There is no clear trend in the residual plot so linearity assumption ~~maybe~~ is ok.

Q5 a) Error degrees of freedom $= n-2 = 15$ so there were 17 soil samples

b) $P_p = \beta_0 + \beta_1 P_i + U$, where $U \sim N(0, \sigma^2)$

c) The estimates of intercept and slope are 62·57 and 1·23, respectively. The estimated mean availability of phosphorous at 17 ppm of inorganic phosphorous is

$$62.57 + 1.23 \times 17 = 83.48 \text{ (ppm)}$$

d) $$\text{tStat} = \frac{\text{Estimate}}{\text{SE}} = \frac{1.2291}{0.3058} = 4.019294 \quad (A)$$

$$\text{p-value} = 2 \times \min\{\mathbb{P}(T_{15} \geq 4.01924), \mathbb{P}(T_{15} \leq 4.01926)\} = (B).$$

As $\mathbb{P}(T_{15} \geq 3.733) = 0.001$ and $\mathbb{P}(T_{15} \geq 4.073) = 0.0005$, the p-value is between 0.001 and 0.002.

e) Let β_1 be the true slope in the linear relationship between inorganic phosphorous and available phosphorous.

Test $H_0: \beta_1 = 1$ against $H_1: \beta_1 \neq 1$

Test statistic $$t = \frac{\text{estimate} - \text{hypothesised}}{\text{s.e.(estimate)}}$$

$$= \frac{1.2291 - 1}{0.3058} = 0.74918$$

$$\text{p-value} = 2 \times \min\{\mathbb{P}(T_{15} \geq 0.74918), \mathbb{P}(T_{15} \leq 0.74918)\}$$

As $\mathbb{P}(T_{15} \geq 0.691) = 0.25$ and $\mathbb{P}(T_{15} \geq 1.341) = 0.1$, the p-value is between 0.2 and 0.5.

There is no evidence against the null hypothesis, suggesting β_1 is equal to one.

f) residual = observation − estimated mean

$$= 51 - (62.5694 + 1.2291 \times 12.6)$$

$$= 51 - 78.0561 = -27.0561.$$

g) Normal probability plot of residuals to check the assumption of normality. Plot residuals against fitted values to check constant variance and linearity (residuals should display no trend.

h) The matrix is symmetric so $(E) = 0.2763$

$$(D) = (\text{SE of intercept})^2 = 4.4519^2 = 19.81941$$

$$(F) = (\text{SE of slope})^2 = 0.3058^2 = 0.093514$$