The taste of mature cheese is related to the concentration of several chemicals in the final product. In a study of cheddar cheese from the Latrobe Valley samples of cheese were subjected to chemical analysis and subjected to taste test. Taste is the response variable. For acetic acid (acetic) and hydrogen sulphide (H2S) natural logs were taken of the concentrations while the concentration of lactic acid (lactic) was untransformed.

(Source: Moore & McCabe, 1999).

(a) Examine the pairs plot. What does this graph show?

![Cheese: Pairs Plot](image)

(b) The multiple linear regression model with taste as the response variable and acetic, H2S and lactic as the explanatory variables is fitted.

| Term     | Estimate | Std. Error | t value | Pr(>|t|) |
|----------|----------|------------|---------|----------|
| (Intercept) | -28.8768 | 19.7354 | -1.463 | 0.15540 |
| acetic   | 0.3277   | 4.4598 | 0.073 | 0.94198 |
| H2S      | 3.9118   | 1.2484 | 3.133 | 0.00425 |
| lactic   | 19.6705  | 8.6291 | 2.280 | 0.03108 |

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-Squared: 0.6518, Adjusted R-squared: 0.6116
F-statistic: 16.22 on 3 and 26 DF, p-value: 3.81e-06
(i) Test the hypothesis, $H: \beta_1 = \beta_2 = \beta_3 = 0$. Make sure you write down a suitable alternative.

(ii) Write down and test appropriate hypotheses for the three partial regression coefficients, $\beta_1$, $\beta_2$ and $\beta_3$.

(iii) Explain carefully what the tests in parts (i) and (ii) imply. (Be sure to explain if it appears all three explanatory variables are required in the model and how you come to your conclusion.)

(c) The model containing only the explanatory variables H2S and lactic is fitted.

|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | -27.592  | 8.982      | -3.072  | 0.00481  |
| H2S         | 3.946    | 1.136      | 3.475   | 0.00174  |
| lactic      | 19.887   | 7.959      | 2.499   | 0.01885  |

Residual standard error: 9.942 on 27 degrees of freedom
Multiple R-Squared: 0.6517,        Adjusted R-squared: 0.6259
F-statistic: 25.26 on 2 and 27 DF,  p-value: 6.551e-07

(i) Find a 99% CI for the 3 parameters, $(\beta_0, \beta_1, \beta_2)$, in this model.

(ii) Predict the mean value of taste when H2S=5 and lactic=1.5.
Solution

1. (a) The plots show there is a positive linear relationship between taste and the predictors, although the relationship appears weaker between taste and acetic. It also appears there is a relationship between the predictor variables (i.e., they appear correlated).

(b) (i) H: $\beta_1 = \beta_2 = \beta_3 = 0$

   Alt: Not all the $\beta$’s are equal to zero.

   The test is highly significant, ($F = 16.22, P = 3.81 \times 10^{-6}$). This implies it is very unlikely that all the $\beta$’s are zero and that there is a relationship between taste and at least one of the variables acetic, H2S and lactic.

(ii) – H$_1$: $\beta_1 = 0$ given that the variables H2S and lactic have been fitted in the model.

   The t value is not significant, ($t = 0.073, P = 0.94$). This implies that $\beta_1 = 0$ given the other two variables have already been fitted.

   – H$_2$: $\beta_2 = 0$ given that the variables acetic and lactic have been fitted in the model.

   The t value is highly significant, ($t = 3.133, P = 0.004$). This implies that $\beta_2 \neq 0$ given the other two variables have already been fitted.

   – H$_3$: $\beta_3 = 0$ given that the variables H2S and acetic have been fitted in the model.

   The t value is significant, ($t = 2.28, P = 0.03$). This implies that $\beta_3 \neq 0$ given the other two variables have already been fitted.

(iii) We saw that a relationship exists between the response variable and at least one of the predictor variables. Each of the tests looks at the additional information supplied by that variable after the other two variables are fitted. Thus, it appears that since the $t$ test for acetic is not significant the additional information supplied by acetic acid about taste is not sufficient to warrant its inclusion in the model provided that the variables H2S and lactic have already been included.

   However, since both the $t$ tests for H2S and lactic are significant, it appears that both provide sufficient additional information to warrant their inclusion after the other variables have been fitted.

   It appears that only the variables H2S and lactic should be retained in the model.
(c) (i) 99% confidence interval for the parameters.
   We use the equation
   \[ \hat{\beta}_j \pm t_{df=27,0.005} \times se(\hat{\beta}_j) \]
   for \( j = 0, 1, 2 \). You should be able to do this using a calculator (and you
   may need to do so in the exam). The estimates of the \( \beta \)'s and the standard
   error of the \( \beta \)'s is contained in the \( R \) output but the value of \( t \) would be
   given in the exam.
   The 95% CI for the intercept is:
   \[ -27.592 \pm 2.77 \times 8.982 = -27.592 \pm 24.88 = ( -52.5 - 2.7 ) \]

(ii) Find predicted value by substituting the values of \( H_2S=5 \) and \( \text{Lactic} = 1.5 \)
   in to the final model:
   \[ \text{taste} = -27.59 + 3.946 \times 5 + 19.887 \times 1.5. \]