MAST30025: Linear Statistical Models

Week 11 Lab

1. We study the effect of various breeds and diets on the milk yield of cows. A study is conducted on 9 cows and the following data obtained:

	Diet		
Breed	1	2	3
1	18.8	16.7	19.8
	21.2		23.9
2	22.3	15.9	21.8
		19.2	

(a) Express this as a two-factor model with no interaction in matrix form.

(b) Express this as a two-factor model with interaction in matrix form.

(c) Express the hypothesis that there is no interaction in terms of your parameters. Eliminate any redundancies.

(d) Input this data into R. Plot an interaction plot between breed and diet.

(e) Test for the presence of interaction.

(f) What is the degrees of freedom used for the interaction test?

(g) From the interaction model, what is the estimated amount of milk produced from breed 2 and diet 3?

(h) Fit an additive model. What is the estimated amount of milk produced from breed 2 and diet 3 now?

(i) Test the hypothesis (under the additive model) that the 2nd and 3rd diets are equivalent in terms of milk produced.

(j) Find a 95% confidence interval, under the additive model, for the amount of milk produced from breed 2 and diet 3. Use both matrix calculations and the `estimable` function from the `gmodels` package.

(k) Find the same confidence interval under the interaction model.

(l) Why is the second interval wider than the first?

2. We study the growth of peas when fed three different types of fertilizer. A study is conducted where the samples are divided into 6 "blocks", corresponding to different plots of land. The data is stored in the `npk` data frame in R. This data frame contains 5 variables:

   - block: label of the block of the sample
   - N: indicator (0/1) for the application of nitrogen
   - P: indicator (0/1) for the application of phosphate
   - K: indicator (0/1) for the application of potassium
   - yield: yield of peas in pounds/plot

   (a) Fit an additive linear model with all variables; then repeat without the block variables. Does the fitted model change? Are the block variables significant?

   (b) Fit a model with the fertilizer variables and all pairwise interaction terms. Are the interaction terms significant?

   (c) Perform variable selection using stepwise selection with AIC, starting from the model with no interaction terms (but considering them for inclusion). What do you find?

   (d) What is the best treatment for peas, according to your final model? Find a 95% confidence interval for the yield of this treatment.

3. Suppose that $\mathbf{y} \sim MVN(\mu\mathbf{1}, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}.$$

   For what values of ρ are the sample mean and sample variance independent?

4. In the one-way classification model, show that any linear combination of $\bar{y}_1 - \bar{y}_\cdot, \ldots, \bar{y}_k - \bar{y}_\cdot$ can be written as a linear combination of $\bar{y}_1, \ldots, \bar{y}_k$. Does the converse hold?