# MAST30025: Linear Statistical Models

## Week 9 Lab

1. Recall Question 5 from the Week 8 lab. In a manufacturing plant, filters are used to remove pollutants. We are interested in comparing the lifespan of 5 different types of filters. Six filters of each type are tested, and the time to failure in hours is given in the dataset `filters` (on the website, in `csv` format).

   (a) Is $\mu - \tau_1 + \tau_5$ estimable?

   (b) Is $\tau_1 - \frac{1}{2}\tau_3 - \frac{1}{2}\tau_4$ estimable?

   (c) In the week 8 lab you were asked to find two solutions to the normal equations. Verify that they produce the same estimate of $\tau_4 - \tau_5$.

   (d) Do your two solutions produce the same estimate of $2\mu + \tau_1$?

   (e) Write down the quantities corresponding to: (i) the lifespan of type 1 filters; (ii) the difference between the lifespans of type 2 and type 3 filters; (iii) the amount by which type 4 filters outlive the average filter; (iv) the expected total time to failure of a set of filters containing one of each type.

   Verify directly that all of these quantities are estimable, and estimate them.

   (f) Fit a `lm` model using `contr.treatment` contrasts (the default). This gives estimates of $\mu_1, \mu_2 - \mu_1, \dots, \mu_5 - \mu_1$. Use these to estimate $\bar{\mu}, \mu_1 - \bar{\mu}, \dots, \mu_5 - \bar{\mu}$. Check your answers by fitting a `contr.sum` model.

2. According to the Gauss-Markov theorem, the estimator for $\mathbf{t}^T\boldsymbol{\beta}$ with the lowest variance is $\mathbf{t}^T\mathbf{b}$. Assuming that $\mathbf{t}^T\boldsymbol{\beta}$ is estimable, show that this variance is $\sigma^2\mathbf{t}^T(X^TX)^c\mathbf{t}$.

3. For the one-way classification model, with $n_i$ observations in group $i$, show that

$$SS_{Reg} := \hat{\mathbf{y}}^T\hat{\mathbf{y}} = \mathbf{y}^TX(X^TX)^cX^T\mathbf{y} = \sum_{i=1}^{k}(\bar{y}_i)^2 n_i.$$

4. Consider the one-way classification model with 3 levels ($k = 3$). Find all estimable quantities of the form $\sum_{i=1}^{3} a_i\tau_i$.

5. Consider the two-way classification model

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}.$$

   Suppose that you have at least one sample from each combination of factor levels.

   Treatment contrasts for the first factor are defined here as $\sum_i a_i\tau_i$, where $\sum_i a_i = 0$. Show that these treatment contrasts are estimable.

PROBLEM 1

(a) $\mu - \tau_1 + \tau_5$ is estimable, to check use the theorem

$t^T (X^TX)^c (X^TX) = t$.

(e) (i) $\mu + \tau_1$

(ii) $\tau_2 - \tau_3$

(iii) $\tau_4 - \frac{1}{5}\sum_{i=1}^{5} \tau_i$

(iv) $(\mu+\tau_1) + (\mu+\tau_2) + \dots + (\mu+\tau_5)$

② Variance$(t^T b)$ $= t^T \, \mathrm{Var}(b) \, t$.

$= t^T \, \mathrm{var}(\underbrace{(X^TX)^c}_{A} X^T y) \, t$.

$= t^T \left( (X^TX)^c X^T \sigma^2 I \, (X(X^TX)^c) \right) t$

$= \underline{t^T (X^TX)^c \, t \, \sigma^2 I}$

$A A^c A = A$

$A^c A A^c = A^c$?