Tutorial Notes 6 of MATH3424

1 Summary of course material

1. Qualitative variables:

   To represent a set of categories of K, we need $K-1$ indicator variables.

2. Interaction variables (multiplicative / interaction effects)

3. Analysis of separate regression equations for two groups of the data:

   (a) Each group has a separate regression model.

   (b) The models have the same intercept but different slopes.

   (c) The models have the same slope but different intercepts.

4. ANOVA by multiple Linear Regression

5. Seasonality

2 Questions

2.1 Assignment 2 Problem 4

Test the hypothesis $H_0 : \beta_1 = \beta_3 = 0.5$ in the following model: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \varepsilon$.

① Define $X' = X_1 + X_3$, Ho becomes:

X $H_0' : \beta' = 0.5$

where $\boxed{Y = \beta_0 + \beta' X' + \varepsilon}$, $\hat{\beta}'$, $s.e.(\hat{\beta}')$, $t = \dfrac{\hat{\beta}' - 0.5}{s.e.(\hat{\beta}')}$

Check if $t > t_{n-2, \alpha/2}$

$H_0' : \beta' = 0.5$ v.s. $H_1' : \beta' \neq 0.5$

$\Leftrightarrow$ $H_0' : Y = \beta_0 + 0.5(X_1 + X_3) + \varepsilon$ v.s. $H_1' : Y = \beta_0 + \beta'(X_1 + X_3) + \varepsilon$

Original $H_0 : Y = \beta_0 + 0.5(X_1 + X_3) + \varepsilon$ v.s. $H_1' : Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \varepsilon$ ✓

(which one is "harder" to reject ?)

Reduced $Y = \beta_0 + 0.5(X_1 + X_3) + \varepsilon$

Full $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \varepsilon$

$Y' = Y - 0.5(X_1 + X_3)$, $\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} Y_i'$

$$SSE(RM) = \sum_{i=1}^{n} \left(Y_i - (\hat{\beta}_0 + 0.5(X_1 + X_3)) \right)^2$$

$$F = \frac{[SSE(RM) - SSE(FM)] / \cancel{(p+1-k)} \; 2}{SSE(FM) / (n-p-1) \quad (n-3)}$$

2.2

Three catalysts are used in a chemical process. The following are yield data from the process:

	Catalyst		
	1	2	3
	79.5	81.5	78.1
	82.0	82.3	80.2
	80.6	81.4	81.5
	84.9	79.5	83.0
	81.0	83.0	82.1
Mean	81.6000	81.5400	80.9800
Variance	4.2050	1.7230	3.6270

Given that the overall sample variance is 2.8135.

(a) Write down a one-way classification model (a model in terms of population means of catalysts) for the analysis of the above data set. Define the variables clearly.

(b) Write down a regression model (a model in terms of indicator variables) for the analysis of the above data set. Define the variables clearly.

(c) Estimate the unknown parameters in part (b).

(d) Hence or otherwise, estimate the unknown parameters in part (a).

(e) Test all population means of catalysts are equal at $\alpha = 0.05$. Write down the test statistic, the critical value and your conclusion clearly.

(a) $Y_{ij} = \mu_j + \varepsilon_{ij}$, $i = 1, \cdots, 5$, $j = 1, 2, 3$

μ_j is the population mean of catalyst j.

(b)
$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_5 \\ Y_6 \\ \vdots \\ Y_{10} \\ Y_{11} \\ \vdots \\ Y_{15} \end{pmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{group 1} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{group 2} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{group 3} \end{array}
\qquad
X = \begin{bmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} 1 \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} 2 \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} 3 \end{array}$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon$$

$n = 15$

Catalyst	x_{i1}	x_{i2}
1	1	0
2	0	1
3	0	0

$$\hat{\beta} = (X'X)^{-1} X'Y$$

$$X'X = \begin{bmatrix} n & \frac{n}{3} & \frac{n}{3} \\ \frac{n}{3} & \frac{n}{3} & 0 \\ \frac{n}{3} & 0 & \frac{n}{3} \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} \frac{3}{n} & -\frac{3}{n} & -\frac{3}{n} \\ -\frac{3}{n} & \frac{6}{n} & \frac{3}{n} \\ -\frac{3}{n} & \frac{3}{n} & \frac{6}{n} \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum_{i=1}^{15} Y_i \\ \sum_{i=1}^{5} Y_i \\ \sum_{i=6}^{10} Y_i \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \left(\sum_{i=1}^{15} Y_i - \sum_{i=1}^{5} Y_i - \sum_{i=6}^{10} Y_i\right) \Big/ n/3 \\ \left(-\sum_{i=1}^{15} Y_i + 2\sum_{i=1}^{5} Y_i + \sum_{i=6}^{10} Y_i\right) \Big/ n/3 \\ \left(-\sum_{i=1}^{15} Y_i + \sum_{i=1}^{5} Y_i + 2\sum_{i=6}^{10} Y_i\right) \Big/ n/3 \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \bar{Y}_3 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

$$\begin{array}{ccc} 1 & 2 & 3 \\ \mu_1 & \mu_2 & \mu_3 \end{array}$$

(a) $Y_{ij} = \mu_j + \varepsilon_{ij}$

(b) $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$

population mean

$$\boxed{\begin{array}{cccc} \text{Catelyst} & 1 & 2 & 3 \\ & \beta_0+\beta_1 & \beta_0+\beta_2 & \beta_0 \end{array}}$$

(d) $\hat{\mu}_1 = \hat{\beta}_0 + \hat{\beta}_1 = \dfrac{\sum_{i=1}^{5} Y_i}{n/3} = \bar{Y}_1$

$\hat{\mu}_2 = \hat{\beta}_0 + \hat{\beta}_2 = \dfrac{\sum_{i=6}^{10} Y_i}{n/3} = \bar{Y}_2$

$\hat{\mu}_3 = \hat{\beta}_0 = \bar{Y}_3$

(e) Test:

$H_0: \mu_1 = \mu_2 = \mu_3$ v.s. H_1: at least one of them not equal to others.

$\Longleftrightarrow$ $H_0: Y_i = \beta_0 + \varepsilon_i$ $\hat{Y}_i = \bar{Y}$

$H_1: Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$

$$F = \frac{SSR/2}{SSE/12} = \underline{0.1835} < F_{(2,12;0.05)}$$

$$SST = \sum_{i=1}^{15} (Y_i - \bar{Y})^2 = 39.3893$$

$$SSE = \underline{\sum_{i=1}^{5} (Y_i - \bar{Y}_1)^2} + \underline{\sum_{i=6}^{10} (Y_i - \bar{Y}_2)^2} + \underline{\sum_{i=11}^{15} (Y_i - \bar{Y}_3)^2}$$

$$= 38.22$$

$$\underline{SSR} = SST - SSE = 1.1693$$

ANOVA:

$$F = \frac{SS_{treatments} / (I-1) \rightarrow 2}{\underline{SS_{Error}} / \underline{(n-I)} \rightarrow 12}$$

I: number of groups.

$= 3$