Stat873 **HW08**

1. In one-way model $Y = J\mu + \epsilon$, $Y \in R^n$, $J = \begin{pmatrix} 1_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1_{n_p} \end{pmatrix} \in R^{n\times p}$, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \in R^p$ and $\epsilon \sim N(0, \sigma^2 I_n)$.

   (1) Show that $\mathrm{R}(1_n) \subset \mathcal{R}(J)$.

   $J1_p = 1_n \Longrightarrow 1_n \in \mathcal{R}(J) \Longrightarrow \mathcal{R}(1_n) \subset \mathcal{R}(J)$.

   (2) Find H_0 under which the model is reduced to $Y = 1_n\gamma + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_n)$.

   Let $H_0 : \mu_1 = \cdots = \mu_p$. Under H_0, $\mu = 1_p\mu_1$ and $J\mu = J1_p\mu_1 = 1_n\mu_1$.
   Thus the model reduced by H_0 is $Y = 1_n\mu_1 + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_n)$.

2. In 1 $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} \in R^n$ where $Y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix} \in R^{n_i}$, $i = 1, ..., p$. Let $n_1 + \cdots + n_p = n$, $\overline{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, $\overline{y} = \frac{1}{n}\left(n_1\overline{y}_1 + \cdots + n_p\overline{y}_p\right)$ and $\mathrm{CSS}_i = \sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2$, $i = 1, ..., p$. Express the followings only using y_{ij}, $\overline{y}_i$, $\overline{y}$, CSS_i and summations.

   (1) SSE

   $\mathrm{SSE} = Y'(I_n - JJ^+)Y = \sum_{i=1}^p Y_i'(I_{n_i} - 1_{n_i}1_{n_i}^+)Y_i = \sum_{i=1}^p \mathrm{CSS}_i$.

   (2) C.SSTO

   C.SSTO$= Y'(I_n - 1_n1_n^+)Y = \sum_{i=1}^p \sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2$.

   (3) SSM

   SSM$= \sum_{i=1}^p \sum_{j=1}^{n_i}(\overline{y}_i - \overline{y})^2 = \sum_{i=1}^p n_i(\overline{y}_i - \overline{y})^2$.

3. Data file T6-10.dat contains four variables y, x_1, x_2 and Type. Consider regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, $\epsilon \sim N(0, \sigma^2)$.

   (1) With $x_1 = 1$ and $x_2 = -1$, find a 90% confidence interval for $E(y)$.

   When $x_1 = 1$ and $x_2 = -1$, the 90% confidence interval for $E(y)$ is
   $\widehat{y} \pm t_{0.05}(n-3)S_{\widehat{y}} = 10.3184 \pm t_{0.05}(56)\,1.2788 = (8.18,\ 12.46)$.

   (2) With $x_1 = 0.5$ and $x_2 = -0.5$, find a 90% prediction interval for y.

   When $x_1 = 0.5$ and $x_2 = -0.5$, the 90% prediction interval for y is
   $\widehat{y} \pm t_{0.05}(n-3)S_{\widehat{y}-y} = 10.0745 \pm t_{0.05}(56)\sqrt{1.2795^2 + 14.61779} = (3.33,\ 16.81)$.

(3) Report your test on $H_0: 2\beta_0 - \beta_2 = 8$.

$H_0: 2\beta_0 - \beta_2 = 8$ vs $H_a: 2\beta_0 - \beta_2 \neq 8$

Test statistics: $F = \frac{(l'\widehat{\beta}-8)\left[l'(X'X)^{-1}l\right]^{-1}(l'\widehat{\beta}-8)}{MSE}$ where $l' = (2,\ 0,\ -1)$

p-value: $P(F(1,\ n-3) > F_{ob})$

$F_{ob} = \frac{297.41405}{14.61779} = 20.35$

p-value: $P(F(1,56) > 20.35) < 0.0001$

Reject H_0.