注意事項：

1. 本試卷共7 題，1-6 每題14分, 第7 題16 分。考試時間： 13:00-17:00 Feb. 18, 2009。

---

1. In multiple linear regression model with n pairs of independent observations $(x_i, y_i), i = 1, \cdots, n$, let
$$Y = X\,\beta + \epsilon$$
where X is an $n \times p$ matrix, β is a $p \times 1$ vector, and ϵ is an $n \times 1$ random vector with $N_n(0, \sigma^2\, I_n)$ distribution, where β and σ^2 are unknown. Let $\hat{\beta}$ and $\hat{Y}$ be the corresponding estimates based on the least-squares. In residual analysis, the following problems are to be studied .
   (i) Find the distribution of the vector of residuals $e = Y - \hat{Y}$.
   (ii) Let $H = X(X^TX)^{-1}X^T$, $h_{ii} = x_i^T(X^TX)^{-1}x_i$ be the ith diagonal element of H. Illustrate how to determine whether a point is a leverage point based on h_{ii}.
   (iii) Let $\hat{\beta}_{(i)}$ and $\hat{Y}_{(i)}$ and $e_{(i)}$, $i = 1, \cdots, n$, be the corresponding deleted estimates and residuals for fitting the regression model after deletion of the ith observation. Show that
$$e_{(i)} = e_i/(1 - h_{ii}).$$
$$\text{Note that} \quad [X_{(i)}^T X_{(i)}]^{-1} = (X^TX)^{-1} + \frac{(X^TX)^{-1}x_i x_i^T(X^TX)^{-1}}{1 - h_{ii}}.$$
   (iv) Give the definitions of Cook's D, $DEBETAS$ and $DEFITS$ for measure of influence of observations. Explain their statistical meanings behind these statistics.

2. Consider a binary response regression model with n pairs of independent observations $(x_i, y_i), i = 1, \cdots, n$, where
$$P(y_i = 1) = p_i = 1 - P(y_i = 0),\; 0 < p_i < 1,$$
and
$$E(y_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$
depends on x_i, and $\beta^T = (\beta_0, \beta_1)$ is the unknown parameter vector.
   (i) Write down the logit transformation function $\eta(x)$ for the logistic regression model.
   (ii) Write down the corresponding log-likelihood function for the n observations, and give the equations for finding the maximum likelihood estimate (MLE) of β.
   (iii) Let $\hat{\beta}$ be the MLE of β, give the form of the large-sample approximate covariance matrix $\text{Var}(\hat{\beta})$.
   (iv) Define what are the "odds" and "odds ratio" in this model, and explain how do the odds ratio related to the unknown parameter β_1.

3. The quality department of a weaving company would like to understand the effects of the following three factors to the dye color of a blend fabric. The three factors are the cycle times (3 levels), temperatures (2 levels) and operators (3 persons). The first two factors are considered to have fixed effects and the third one has random effect. The response value is the evaluation score after comparing the dye color of the blend fabric with the standard color.
   (i) Explain how to determine whether a factor has a fixed effect or random effect.
   (ii) Write out an appropriate model for the above experiment and the corresponding ANOVA table including the formula of computing the sum of squares for each effect and the EMS.
   (iii) Illustrate what would be the proper testing statistics whether the effects are significant or not? (Assume that the three way interactions do not exist).

4. A 2^{5-2} design of experiment has been used to investigate the effect of A = condensation temperature, B = amount of material 1, C = solvent volume, D = condensation time, and E = amount of material 2 on yield. The results obtained are as follows:
$$e = 23.2 \quad ad = 16.9 \quad cd = 23.8 \quad bde = 16.8$$
$$ab = 15.5 \quad bc = 16.2 \quad ace = 23.4 \quad abcde = 18.1$$
   (i) Find the design generators used in this experiment, and explain what is the resolution of this design and the meaning of it.

(ii) Write down the complete defining relation and the aliases of the main effects and two-way interactions for this design .

(iii) Estimate the main effects and prepare an analysis of variance table, and explain which interaction terms may be used as the error.

5. Let $X_1, \cdots, X_n$ be a random sample from the Normal distribution $N_p(\mu, \Sigma)$. For $n > p$, let $\bar{X}_n = \sum_{i=1}^{n} X_i/n$.

   (i) Find the characteristic function of X_1.

   (ii) Let $Z_k = (\sqrt{\frac{k}{k+1}})(X_{k+1} - \bar{X}_k), k = 1, 2, \cdots, n-1$. Show that if $\Sigma = \sigma^2 I$, $Z_1, Z_2, \cdots, Z_{n-1}, \bar{X}_n$ are independent.

6. Assume that the two p-dimensional random samples $y_{11}, y_{12}, \cdots, y_{1n_1}$ and $y_{21}, y_{22}, \cdots, y_{2n_1}$ from the two normal populations to be compared have the same covariance matrix Σ, but distinct mean vectors μ_1 and μ_2. In discriminant analysis, the discriminant function is the linear combination $z = a^T y$ of these p variables that maximizes the distance between the two (transformed) group mean vectors

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{s_z^2}.$$

   (i) Write down the criterion in terms of a, and the corresponding sample means $\bar{y}_1, \bar{y}_2$ as well as S_{pl} the pooled sample covariance matrix of the two samples .

   (ii) Find the best discriminant function.

7. Assume $X_1, X_2, \cdots, X_n$ is a random sample of size n from $N_p(\mu, \Sigma)$ distribution. We are interested in testing the hypothesis that Σ is of the special form as

$$H_0 : \Sigma = \sigma^2 \, I_p \quad \text{versus} \quad H_1 : \Sigma \neq \sigma^2 \, I_p$$

   where σ^2 is not specified.

   (i) Find the maximum likelihood estimate of σ^2 under H_0.

   (ii) Find the likelihood ratio test statistic for testing H_0, express it in a more explicit form.

   (iii) Give the asymptotic null distribution for the test statistic.