1. Let

$$\begin{aligned} P_0(x_i) &= 1 \\ P_1(x_i) &= \frac{x_i - \bar{x}}{d} \\ P_2(x_i) &= (\frac{x_i - \bar{x}}{d})^2 - (\frac{n^2-1}{12}) \end{aligned}$$

(a) Show that $P_0(x), P_1(x)$ and $P_2(x)$ are the orthogonal polynomial of degree 0,1 and 2 with respect to equally spaced data points, $x_1, \cdots, x_n$, where $\bar{x} = \sum_{i=1}^{n} x_i/n$ and $d = x_{i+1} - x_i$, $i = 1, 2, \cdots, n-1$. (10 points)

(b) Give the following data and the values of orthogonal polynomial for $n = 5$:

$y:$	2	3	1	-1	2
$x:$	101	102	103	104	105

$P_1(x_i)$	-2	-1	0	1	2
$P_2(x_i)$	2	-1	-2	-1	2

Find the fitted equation of a quadratic-degree orthogonal polynomial model. (10 points)

2. For a multiple linear regression model, suppose that the true model is

$$y = X_p\beta_p + X_r\beta_r + \varepsilon.$$

Consider the subset model

$$y = X_p\beta_p + \varepsilon.$$

(a) Find the biase of least-squares estimators of β_p from subset model. (10 points)

(b) Show that the expectation of $\hat{\sigma}^2$ from subset model is

$$E(\hat{\sigma}^2) = \sigma^2 + \frac{\beta_r' X_r' [I - X_p(X_p'X_p)^{-1}X_p'] X_r\beta_r}{n-p}.$$

(10 points)

3. Consider the single-factor fixed-effects analysis of variance with 3 treatments:

$$y_{ij} = \mu\tau_i + \varepsilon_{ij}, i = 1, 2, 3, j = 1, 2, \cdots, n.$$

We can use two indicator variables to rewrite the model as

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \varepsilon_{ij}, i = 1, 2, 3, j = 1, 2, \cdots, n,$$

where $x_{1j} = 1$ if the observation j is from treatment 1 and 0 otherwise; and $x_{2j} = 1$ if observation j is from treatment 2 and 0 otherwise.

(a) Find the least-squares estimators of β_i, $i = 1, 2, 3$. (10 points)

(b) Show that the regression sum of squares of full model is

$$SS(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{3} \frac{y_{i.}^2}{n},$$

where $y_{i.} = \sum_{j=1}^{n} y_{ij}, i = 1, 2, 3.$ (10 points)

4. Let y_i denote n observations at points, $x_i, i = 1, 2, \cdots, n$. Suppose that y_i can be formed into two groups and for each group, the observations may be fitted by a simple linear regression:

$$\begin{aligned} y_i &= \beta_{01} + \beta_{11} x_i + \varepsilon_i, i = 1, \cdots, n_1, \\ y_i &= \beta_{02} + \beta_{12} x_i + \varepsilon_i, i = n_1 + 1, \cdots, n. \end{aligned}$$

Use the indicator variable to one linear regression model to do standard statistical analysis for testing $H_0 : \beta_{01} = \beta_{02}$ and $\beta_{11} = \beta_{12}$ (i.e. two regression lines are identical) and $H_0 : \beta_{11} = \beta_{12}$ (i.e. two regression lines are parallel), and also write down the corresponding statistic for each test. (15 points)

5. Write down the purpose of each term: (15 points)

(a) Cook's D, $D_i(X'X, pMS_{Res}) = \frac{(\hat{\beta_{(i)}} - \hat{\beta})'(X'X)(\hat{\beta_{(i)}} - \hat{\beta})}{pMS_{Res}}$

(b) $DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j\,(i)}}{\sqrt{S_{(i)}^2 C_{jj}}}$

(c) $DFFITS_i = \frac{\hat{y}_i - \hat{y_{(i)}}}{\sqrt{S_{(i)}^2 h_{ii}}}$

6. Suppose that the variance of the observation, σ^2, is functionally related to the mean, $E(y)$. Find the variance-stabilizing transformations if (a) $\sigma^2 \propto E(y)$ and (b) $\sigma^2 \propto E(y)^2$. (10 points)