L22 Linear model: Regression and one-way ANOVA

1. Linear model

   (1) Linear function
   $y = f(x)$ is a linear function if $f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$.
   If $y \in R^m$ and $x \in R^n$, then
   $$y = f(x) \text{ is a linear function } \iff \exists\, A \in R^{m\times n} \text{ such that } y = f(x) = Ax.$$
   Pf: $\Leftarrow$**:** $f(\alpha u + \beta v) = A(\alpha u + \beta v) = \alpha Au + \beta Av = \alpha f(u) + \beta f(v)$.
   $\Rightarrow$**:** Let $A_i = f(e_i) \in R^m$ where $e_i \in R^n$ is the ith column of I_n, $i = 1, ..., n$.
   $$\begin{aligned}\text{Then } f(x) &= f(x_1e_1 + \cdots + x_ne_n) = x_1f(e_1) + \cdots + x_nf(e_n) \\ &= x_1A_1 + \cdots + x_nA_n = Ax.\end{aligned}.$$

   (2) Linear model
   If in a model for random variable y, $E(y)$ is a linear function of unknown parameter vector $\beta \in R^p$, then this model is called a linear model. So a linear model for y with parameter vector $\beta \in R^p$ is
   $$y = (m_1, .., m_p)\beta + \epsilon \text{ with } E(\epsilon) = 0.$$

   (3) Samples from a linear model
   Suppose $y_1, ..., y_n$ is a random sample from the above linear model where y_i is observed when $(m_1, ..., m_p) = (m_{i1}, ..., m_{in})$. Let $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $M = \begin{pmatrix} m_{11} & \cdots & m_{1p} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{np} \end{pmatrix}$. Then $Y = M\beta + \epsilon$ with $E(\epsilon) = 0 \in R^n$ characterizes a linear model based on sample.

   Comments: Statistical inference is based on sample $Y = M\beta + \epsilon$.
   $E(Y) = M\beta \in L(M)$, a linear space in R^n.
   Matrix M characterizes the model, i.e., different M gives different models.

2. Linear regression model is a linear model

   (1) Regression model
   Linear regression $y = \beta_0 + \beta_1x_1 + \cdots + \beta_{p-1}x_{p-1} + \epsilon$ with $E(\epsilon) = 0$ is a linear model where
   $$E(y) = \beta_0 + \beta_1x_1 + \cdots + \beta_{p-1}x_{p-1} = (1,\, x_1, ...x_{p-1})\beta \text{ with } \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$
   is called the regression function.

   Comments: Regression model represents infinite number of populations,
   Caustion: $E(y)$ is a function of $(1,\, x_1, ..., x_p)$ and is also a function of $\beta \in R^p$.

   (2) Samples from a regression model
   Suppose y_i is observed when $(1,\, x_1, ..., x_{p-1})$ is $(1,\, x_{i1}, ..., x_{i\,p-1})$, $i = 1, ..., n$.
   Let $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in R^n$, $X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1\,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n\,p-1} \end{pmatrix} \in R^{n\times p}$ and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$ with $E(\epsilon) = 0_n$. Then $Y = X\beta + \epsilon$. So $E(Y) = X\beta$ lines in $\mathcal{R}(X)$, a linear space in R^n.

3. One-way ANOVA is a linear model

(1) One-way ANOVA model
In an experiment response y is affected by p levels of a single factor.
Suppose under the ith level, called the ith treatment, $E(y) = \mu_i$.
Let d_i be the indicator for the ith treatment, i.e., $d_i = \begin{cases} 1 & i\text{th treatment is applied} \\ 0 & \text{otherwise} \end{cases}$
and $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$. The one-way ANOVA below is a linear model.

$$y = d_1\mu_1 + \cdots + d_p\mu_p + \epsilon = (d_1, .., d_p)\mu + \epsilon \text{ with } E(\epsilon) = 0.$$

Comments: This model represents p populations since $(d_1, .., d_p)$ with one $d_i = 1$ and rest $d_j = 0$ has p different values and $(d_1, ..., d_p)\mu$ has p values: $\mu_1, ..., \mu_p$.

(2) Samples from one-way ANOVA
Suppose y_i is observed on an experiment unit when $(d_1, .., d_p) = (d_{i1}, ..., d_{ip})$, $i = 1, ..., n$.
Let $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in R^n$, $D = \begin{pmatrix} d_{11} & \cdots & d_{1p} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{np} \end{pmatrix} \in R^{n\times p}$ and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$ with $E(\epsilon) = 0$.
Then $Y = D\mu + \epsilon$. So $E(Y) = D\mu \in L(D) = R^p$.

Comment: Matrix D specifies how many and which experiment units are subject to which treatment. Hence it is called the design matrix.

Ex1: $D = \begin{pmatrix} 1_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1_{n_p} \end{pmatrix}$ is often denoted by J.

(3) Reparameterization

Let $\mu_. = \frac{\mu_1+\cdots+\mu_p}{p} = \frac{1_p'}{p}\mu = 1_p^+\mu$, $\alpha_i = \mu_i - \mu_i$ and $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} = (I_p - 1_p 1_p^+)\mu$. Then

$$\theta = \begin{pmatrix} \mu_. \\ \alpha \end{pmatrix} = \begin{pmatrix} 1_p^+ \\ I_p - 1_p 1_p^+ \end{pmatrix} \mu = (1_p,\, I_p - 1_p 1_p^+)^+ \mu \in R^{p+1} \text{ and } \mu = (1_p,\, I_p)\theta \in R^p.$$

Here $\alpha_1, ..., \alpha_p$ are called the factor effects. Clearly $\mu_i = \mu_j \Longleftrightarrow \alpha_i = \alpha_j$.
When using θ there is a restriction $\alpha_1 + \cdots + \alpha_p = 0 \Longleftrightarrow (0, 1_p')\theta = 0$.

(4) A model with restriction
By the reparameterization in (3), for data vector Y

$$Y = J\mu + \epsilon \text{ with } E(\epsilon) = 0 \quad \Longleftrightarrow \quad \begin{array}{l} Y = J(1_p, I_p)\theta + \epsilon \text{ with } E(\epsilon) = 0 \\ \text{under the restriction } (0,\, 1_p')\theta = 0. \end{array}$$

Ex2: $(0,\, 1_p')\theta = 0 \Longleftrightarrow \theta \in \mathcal{N}((0,\, 1_p'))$. So

$$\begin{aligned} E(Y) = J\,(1_p,\, I_p)\theta &\in J\,(1_p,\, I_p)\,\mathcal{N}((0, 1_p')) = J\,(1_p,\, I_p)\,\mathcal{N}\left(\begin{pmatrix} 0 & 0 \\ 0 & 1_p 1_p^+ \end{pmatrix}\right) \\ &= J\,(1_p,\, I_p)\,\mathcal{R}\left(\begin{pmatrix} 1 & 0 \\ 0 & I_p - 1_p 1_p^+ \end{pmatrix}\right) = J\,\mathcal{R}\left((1_p,\, I_p - 1_p 1_p^+)\right) \\ &= J\,\mathcal{R}(I_p) = \mathcal{R}(J). \end{aligned}$$