

1 Linear models & estimable functions

Let $X \in \mathbb{R}^{n \times p}$ be the *model matrix* of a certain linear model $Y = X\beta + e$, where \mathbb{R}^p is the *parameter space*, and \mathbb{R}^n is the *observation space* (or *measurement space*) of the model.

Then, X can be regarded as a linear map from the parameter space \mathbb{R}^p into the measurement space \mathbb{R}^n , i.e. $X : \mathbb{R}^p \rightarrow \mathbb{R}^n : \beta \mapsto X \cdot \beta$. The image $\text{image}(X) = X(\mathbb{R}^p) \subset \mathbb{R}^n$ of $X : \mathbb{R}^p \rightarrow \mathbb{R}^n$ is the *estimation space*, which of course is a subspace of the measurement space \mathbb{R}^n .

Remark 1.1 Let two vector spaces V and W be given, with V^* and W^* being their respective dual spaces. Then, every linear map $f : V \rightarrow W$ naturally induces its **dual map** $f^* : W^* \rightarrow V^* : \rho \mapsto \rho \circ f$. More explicitly, given $\rho \in W^*$, $f^*(\rho) \in V^*$ is the functional on V defined as follows:

$$V \rightarrow \mathbb{R} : \mathbf{v} \mapsto (\rho \circ f)(\mathbf{v}) = \rho(f(\mathbf{v}))$$

Remark 1.2 In particular, the linear map:

$$\mathbb{R}^p \xrightarrow{X} \mathbb{R}^n : \beta \mapsto X\beta$$

induces

$$(\mathbb{R}^n)^* \xrightarrow{X^*} (\mathbb{R}^p)^* : \rho^T \mapsto (\beta \mapsto \rho^T X \beta),$$

Definition 1.3 (Estimable functionals on parameter space)

The set of estimable functionals on the parameter space \mathbb{R}^p is, by definition, the image $\text{image}(X^*) \subset (\mathbb{R}^p)^*$. In other words, a function $\lambda^T \in (\mathbb{R}^p)^* = \mathbb{R}^{1 \times p}$ is an estimable functional on \mathbb{R}^p if and only if $\lambda = X^T \rho$, for some $\rho \in \mathbb{R}^n$.

Lemma 1.4 Two functionals $\rho_1, \rho_2 \in (\mathbb{R}^n)^*$ induce, via $\mathbb{R}^p \xrightarrow{X} \mathbb{R}^n$, the same estimable functional in $(\mathbb{R}^p)^*$ if and only if $\rho_1 - \rho_2 \in \ker(X^T)$.

PROOF

$$\begin{aligned} & \text{Two functionals } \rho_1, \rho_2 \in (\mathbb{R}^n)^* \text{ induce the same functional in } (X(\mathbb{R}^p))^* \\ \iff & \rho_1^T X \beta = \rho_2^T X \beta, \text{ for each } \beta \in \mathbb{R}^p \\ \iff & (\rho_1^T X - \rho_2^T X) \beta = \mathbf{0}, \text{ for each } \beta \in \mathbb{R}^p \\ \iff & \rho_1^T X - \rho_2^T X = \mathbf{0} \in \mathbb{R}^{1 \times p} = (\mathbb{R}^p)^* \\ \iff & X^T(\rho_1 - \rho_2) = \mathbf{0} \\ \iff & \rho_1 - \rho_2 \in \ker(X^T) \end{aligned}$$

□

2 Fundamental Theorem of Testing in Linear Models: Testing a linear model against a reduced model

Theorem 2.1

Let $X \in \mathbb{R}^{n \times p}$ and $X_0 \in \mathbb{R}^{n \times q}$ be given, with $\text{Col}(X_0) \subset \text{Col}(X) \subset \mathbb{R}^n$. Let P_X and P_{X_0} denote the projection map from \mathbb{R}^n onto $\text{Col}(X)$ and $\text{Col}(X_0)$, respectively. Let $Y : \Omega \rightarrow \mathbb{R}^n$ be \mathbb{R}^n -valued random variables defined on a probably space Ω . Then, the following statements are true:

1. If there exists $\beta \in \mathbb{R}^p$ such that $Y - X\beta \sim N(0, \sigma^2 I_{n \times n})$, then

$$\begin{aligned} \frac{\|(P_X - P_{X_0})Y\|^2 / \text{rank}(P_X - P_{X_0})}{\|(\text{id}_{\mathbb{R}^n} - P_X)Y\|^2 / \text{rank}(\text{id}_{\mathbb{R}^n} - P_X)} &= \frac{Y^T(P_X - P_{X_0})Y / \text{rank}(P_X - P_{X_0})}{Y^T(\text{id}_{\mathbb{R}^n} - P_X)Y / \text{rank}(\text{id}_{\mathbb{R}^n} - P_X)} \\ &\sim F\left(\text{rank}(P_X - P_{X_0}), \text{rank}(P_X - P_{X_0}); \frac{1}{2\sigma^2} \|(P_X - P_{X_0})X\beta\|^2\right) \end{aligned}$$

2. If there exists $\gamma \in \mathbb{R}^q$ such that $Y - X_0\gamma \sim N(0, \sigma^2 I_{n \times n})$, then

$$\begin{aligned} \frac{\|(P_X - P_{X_0})Y\|^2 / \text{rank}(P_X - P_{X_0})}{\|(\text{id}_{\mathbb{R}^n} - P_X)Y\|^2 / \text{rank}(\text{id}_{\mathbb{R}^n} - P_X)} &= \frac{Y^T(P_X - P_{X_0})Y / \text{rank}(P_X - P_{X_0})}{Y^T(\text{id}_{\mathbb{R}^n} - P_X)Y / \text{rank}(\text{id}_{\mathbb{R}^n} - P_X)} \\ &\sim F(\text{rank}(P_X - P_{X_0}), \text{rank}(P_X - P_{X_0}); 0) \end{aligned}$$

Remark 2.2 (Intuitive idea behind testing a linear model against a reduced model)

Assume the “full model” $Y = X\beta + e$, with $e \sim N(0, \sigma^2 I_{n \times n})$, is true. Then, $\|(\text{id}_{\mathbb{R}^n} - P_X)Y\|^2$ is the squared norm of the error vector between the observed vector Y and its full-model estimate $P_X Y$. Next, observe that:

$$P_X Y = P_{X_0} Y + (P_X - P_{X_0}) Y.$$

Hence, $\|(P_X - P_{X_0})Y\|^2$ is the squared norm of the error vector between the full-model estimate $P_X Y$ of Y and the reduced-model estimate $P_{X_0} Y$ of Y . Thus, intuitively, the ratio

$$\frac{\|(P_X - P_{X_0})Y\|^2 / \text{rank}(P_X - P_{X_0})}{\|(\text{id}_{\mathbb{R}^n} - P_X)Y\|^2 / \text{rank}(\text{id}_{\mathbb{R}^n} - P_X)}$$

is the ratio of the dimension-adjusted full-to-reduced-model error to the dimension-adjusted observed-to-full-model error. If this ratio of dimension-adjusted errors turns out to be “large”, then we reject the reduced model in favour of the full model.

3 Special Case: Testing full one-way ANOVA model against the reduced model of the vanishing of all contrasts (i.e. no treatment effects)

Remark 3.1

By choosing to test the full model against only those reduced models whose corresponding estimation spaces are defined by the vanishing of contrasts, one in effect chooses to only test the full model against those reduced models whose corresponding estimation spaces contain the vector $(1, 1, \dots, 1) \in \mathbb{R}^n$.

Note that the vector $(1, 1, \dots, 1) \in \mathbb{R}^n$ in the measurement space \mathbb{R}^n spans the estimation subspace in \mathbb{R}^n corresponding to the “grand-mean” reduced model $Y = \mu + e$, with $\mu \in \mathbb{R}^n$ being a fixed grand-mean vector, and $e \sim N(0, \sigma^2 I_{n \times n})$.

4 Special Case: Testing full one-way ANOVA model against the reduced model of the vanishing of a given contrast

5 Motivating Example

GOAL: To evaluate the effect of smoking on heart rate.

- Suppose the collection \mathcal{U} of all human beings has been divided into four overlapping groups: \mathcal{U}_1 non-smokers, \mathcal{U}_2 light smokers, \mathcal{U}_3 moderate smokers, and \mathcal{U}_4 heavy smokers, i.e.

$$\mathcal{U} = \bigsqcup_{j=1}^4 \mathcal{U}_j$$

- In each group, we take a random sample of 6 individuals, and measure their heart rates after three minutes of resting.
- Let μ_j be the population mean heart rate of \mathcal{U}_j , $j = 1, \dots, 4$, respectively. We would like to use the collected data to test:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

versus

$$H_1 : \text{not all of } \mu_1, \mu_2, \mu_3, \text{ and } \mu_4 \text{ are equal.}$$

Suppose the collected data are as follows:

non-smokers $j = 1$	light smokers $j = 2$	moderate smokers $j = 3$	heavy smokers $j = 4$
69	55	66	91
52	60	81	72
71	78	70	81
58	58	77	67
59	62	57	95
65	66	79	84

6 Sums of Squares and Their Relevant Properties

$$\begin{aligned}
 SS_{\text{inter-level}} &:= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 = \sum_{j=1}^k n_j (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 \\
 SS_{\text{intra-level}} &:= \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\cdot j})^2 \\
 SS_{\text{total}} &:= \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2
 \end{aligned}$$

Proposition 6.1

$$SS_{\text{total}} = SS_{\text{inter-level}} + SS_{\text{intra-level}}$$

Proposition 6.2

Suppose:

1. Y_{ij} , for $j = 1, \dots, k$ and $i = 1, \dots, n_j$, are independent random variables.
2. there exist $\sigma^2 > 0$ and $\mu_1, \dots, \mu_k \in \mathbb{R}$ such that $Y_{ij} \sim N(\mu_j, \sigma^2)$, for each $j = 1, \dots, k$, and each $i = 1, \dots, n_j$.

Then, regardless of whether H_0 ($\mu_1 = \mu_2 = \mu_3 = \mu_4$) is true, we have:

- $\frac{SS_{\text{intra-level}}}{\sigma^2}$ has a χ^2 distribution with $(n - k)$ degrees of freedom. In particular, $E\left[\frac{SS_{\text{intra-level}}}{\sigma^2}\right] = n - k$.
- $SS_{\text{intra-level}}$ and $SS_{\text{inter-level}}$ are independent random variables.

Proposition 6.3

Suppose:

1. Y_{ij} , for $j = 1, \dots, k$ and $i = 1, \dots, n_j$, are independent random variables.
2. there exist $\sigma^2 > 0$ and $\mu_1, \dots, \mu_k \in \mathbb{R}$ such that $Y_{ij} \sim N(\mu_j, \sigma^2)$, for each $j = 1, \dots, k$, and each $i = 1, \dots, n_j$.

Then,

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \implies \frac{SS_{\text{inter-level}}}{\sigma^2} \sim \chi_{k-1}^2$$

Corollary 6.4

Suppose:

1. Y_{ij} , for $j = 1, \dots, k$ and $i = 1, \dots, n_j$, are independent random variables.
2. there exist $\sigma^2 > 0$ and $\mu_1, \dots, \mu_k \in \mathbb{R}$ such that $Y_{ij} \sim N(\mu_j, \sigma^2)$, for each $j = 1, \dots, k$, and each $i = 1, \dots, n_j$.

Then,

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \implies F := \frac{SS_{\text{inter-level}}/(k-1)}{SS_{\text{intra-level}}/(n-k)} \sim \mathcal{F}_{n-k}^{k-1}$$

Lemma 6.5

Suppose:

1. Y_{ij} , for $j = 1, \dots, k$ and $i = 1, \dots, n_j$, are independent random variables.
2. there exist $\sigma^2 > 0$ and $\mu_1, \dots, \mu_k \in \mathbb{R}$ such that $Y_{ij} \sim N(\mu_j, \sigma^2)$, for each $j = 1, \dots, k$, and each $i = 1, \dots, n_j$.

Then,

$$E[SS_{\text{inter-level}}] = (k-1)\sigma^2 + \sum_{j=1}^k n_j(\mu_j - \mu)^2,$$

where $\mu := \frac{1}{n} \sum_{j=1}^k \mu_j$, and $n = n_1 + n_2 + \dots + n_k$. In particular,

$$\frac{E[SS_{\text{inter-level}}/(k-1)]}{E[SS_{\text{intra-level}}/(n-k)]} = \frac{\sigma^2 + \frac{1}{k-1} \sum_{j=1}^k n_j(\mu_j - \mu)^2}{\sigma^2} \geq 1$$

7 Experimental Design of Univariate One-way ANOVA

Let $n, N \in \mathbb{N}$, with $n \leq N$. Let $\mathcal{U} = \{1, 2, \dots, N\}$, which represents the finite population, or universe, of N elements.

A Technical Results

Theorem A.1

Let Y_1, Y_2, \dots, Y_n be independent identically distributed random variables with common distribution $N(\mu, \sigma^2)$. Then,

- the following two random variables are independent:

$$\bar{Y} := \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{and} \quad S^2 := \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

- the random variable

$$\frac{1}{\sigma^2} \cdot (n-1) S^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

has a χ^2 distribution with $(n-1)$ degrees of freedom.