Chapter 2

Linear Model Theory

2.1 Estimation in Linear Models

2.1.1 Maximum Likelihood Estimation

– Let us consider the estimation of unknown parameters of the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}). \tag{2.1}$$

– Since $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, the likelihood function $L(\boldsymbol{\beta}, \sigma^2|\mathbf{y})$ and log-likelihood function $l(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \log\left(L(\boldsymbol{\beta}, \sigma^2|\mathbf{y})\right)$have the forms

$$L(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^n|\sigma^2\mathbf{I}|}} \cdot \exp\left(-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right), \tag{2.2a}$$

$$l(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}. \tag{2.2b}$$

– The maximum likelihood estimators $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ of the parameter vector $\boldsymbol{\beta}$ and $\hat{\sigma}^2$ of the scalar $\sigma^2 > 0$ can be obtained by finding the maximum of the log-likelihood function $l(\boldsymbol{\beta}, \sigma^2|\mathbf{y})$ with respect to $\boldsymbol{\beta}, \sigma^2$ by the process of differentiation.

– The maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are the solutions to the maximization problem

$$\hat{\boldsymbol{\beta}}, \hat{\sigma}^2 = \arg\max_{\boldsymbol{\beta},\sigma^2} l(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \arg\max_{\boldsymbol{\beta},\sigma^2}\left(-\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right). \tag{2.3}$$

– The local maximum of $l(\boldsymbol{\beta}, \sigma^2|\mathbf{y})$ with respect to $\boldsymbol{\beta}$ occurs at a point, where the gradient vector (partial derivates)

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial l}{\partial \beta_0} \\ \frac{\partial l}{\partial \beta_1} \\ \vdots \\ \frac{\partial l}{\partial \beta_p} \end{pmatrix}$$

is vanishing

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{\partial \left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}} = \frac{\partial \left(-\frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}} \\ &= \left(-\frac{1}{2\sigma^2}\right) \frac{\partial \left(-2\mathbf{y}'\mathbf{X}\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}} + \left(-\frac{1}{2\sigma^2}\right) \frac{\partial \left(\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}} \\ &= \left(-\frac{1}{2\sigma^2}\right)(-2\mathbf{X}'\mathbf{y}) + \left(-\frac{1}{2\sigma^2}\right)(2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{1}{\sigma^2}\left(\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \mathbf{X}'\mathbf{y}\right) = \mathbf{0}. \end{aligned} \tag{2.4}$$

– Since $\frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{0}$ for any $\sigma^2 > 0$ if and only if the *normal equations*

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \tag{2.5}$$

holds for $\boldsymbol{\beta}$.

– The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is the solution to the normal equations

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}, \tag{2.6}$$

i.e.,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{2.7}$$

– Thus, the maximum of the log-likelihood function $l(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \log\left(L(\boldsymbol{\beta}, \sigma^2|\mathbf{y})\right)$ with respect to $\boldsymbol{\beta}$ is

$$\begin{aligned} l(\hat{\boldsymbol{\beta}}, \sigma^2|\mathbf{y}) &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{2\sigma^2} \\ &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})}{2\sigma^2} \\ &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{2\sigma^2}, \end{aligned} \tag{2.8}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

– The partial derivate $\frac{\partial l}{\partial \sigma^2}$ has the form

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2}\frac{2\pi}{(2\pi\sigma^2)} + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2(\sigma^2)^2} = -\frac{n}{2\sigma^2} + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2(\sigma^2)^2} \tag{2.9}$$

– In equation (2.9), replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, and setting $\frac{\partial l}{\partial \sigma^2} = 0$, we obtain equivalent equalities

$$-\frac{n}{2\sigma^2} + \frac{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{2(\sigma^2)^2} = 0, \tag{2.10a}$$

$$\frac{-n\sigma^2 + (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{2(\sigma^2)^2} = 0, \tag{2.10b}$$

$$-n\sigma^2 + (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = 0. \tag{2.10c}$$

– The maximum likelihood estimator $\hat{\sigma}^2$ is the solution

$$\hat{\sigma}^2 = \frac{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{n} = \frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{n}. \tag{2.11}$$

– Thus, the maximum of the log-likelihood function

$$\begin{aligned} l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2|\mathbf{y}) &= -\frac{n}{2}\log\left(2\pi\hat{\sigma}^2\right) - \frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{2\hat{\sigma}^2} \\ &= -\frac{n}{2}\log\left(2\pi\hat{\sigma}^2\right) - \frac{n}{2}. \end{aligned} \tag{2.12}$$

– Since it can be shown that the Hessian matrix (second partial derivates) is a negative definite matrix (in Löwner sense)

$$\begin{pmatrix} \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2} \\ \frac{\partial^2 l}{\partial \boldsymbol{\beta}' \partial \sigma^2} & \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix}_{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2} = \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}(\mathbf{X}'\mathbf{X}) & \frac{\left(\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}'\mathbf{y}\right)}{(\hat{\sigma}^2)^2} \\ \frac{\left(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X} - \mathbf{y}'\mathbf{X}\right)}{(\hat{\sigma}^2)^2} & -\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\hat{\sigma}^2)^3} \end{pmatrix} \leq_L \mathbf{0}, \tag{2.13}$$

the solutions $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2$ are the global maximum of the log-likelihood function $l(\boldsymbol{\beta}, \sigma^2 | \mathbf{y})$.

– The maximum likelihood estimator $\hat{\sigma}^2$ is a bias estimator, i.e., $\mathrm{E}(\hat{\sigma}^2) = \frac{n-(p+1)}{n}\sigma^2$, and hence the variance parameter σ^2 is usually estimated by the unbiased estimator

$$\tilde{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - (p+1)} = \frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{n - (p+1)}, \tag{2.14}$$

which is also called as the restricted maximum likelihood estimator.

– Note that the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ minimizes the quadratic form $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, and thus is also called as the ordinary least squares estimator:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2. \tag{2.15}$$

– For any given values of the explanatory variables $\mathbf{x}_*$, the maximum likelihood estimator for the expected value $\mu_* = \mathbf{x}_*'\boldsymbol{\beta}$ is

$$\hat{\mu}_* = \mathbf{x}_*'\hat{\boldsymbol{\beta}} = \mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.16}$$

– The fitted values and the raw residuals of the linear model are

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}\mathbf{y}, \tag{2.17}$$

$$\mathbf{e} = \mathbf{y} - \hat{\boldsymbol{\mu}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{M}\mathbf{y}, \tag{2.18}$$

where $\mathbf{P} = \mathbf{P}_{\mathrm{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{M} = \mathbf{I}-\mathbf{P} = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ are orthogonal projectors onto the column space $\mathcal{C}(\mathbf{X})$ and onto the complement space $\mathcal{C}(\mathbf{X})^{\perp}$, respectively.

Example 2.1.

Let us consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}).$$

Then the expected value of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \mathrm{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \mathrm{E}(\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}, \tag{2.19}$$

and the covariance matrix is

$$\begin{aligned} \mathrm{Cov}(\hat{\boldsymbol{\beta}}) &= \mathrm{Cov}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \mathrm{Cov}(\mathbf{y}) \cdot \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{2.20}$$

2.1.2 Best Linear Unbiased Estimation

– Note that the maximum likelihood estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the linear estimator $\hat{\boldsymbol{\beta}} = \mathbf{G}\mathbf{y}$, where $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

– As a linear transformation of $\mathbf{y}$, based on Theorem 1.1 (a) and Example 2.1, the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ follows the normal distribution

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{2.21}$$

– Thus also holds

$$\mathbf{x}_*'\hat{\boldsymbol{\beta}} \sim N\left(\mathbf{x}_*'\boldsymbol{\beta}, \sigma^2\mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*\right). \tag{2.22}$$

– More generally, a linear estimator $\mathbf{G}\mathbf{y}$ of $\boldsymbol{\beta}$ is said to be *linear unbiased estimator* if (and only if)

$$\mathrm{E}(\mathbf{G}\mathbf{y}) = \boldsymbol{\beta} \qquad \text{for all } \boldsymbol{\beta} \in \mathbb{R}_{p+1}, \tag{2.23}$$

i.e., the expected estimation error is zero.

– It is easy to confirm that (2.23) is equivalent to $\mathbf{G}\mathbf{X} = \mathbf{I}$.

– A linear unbiased estimator $\mathbf{G}\mathbf{y}$ is said to the *best linear unbiased estimator*, BLUE, of $\boldsymbol{\beta}$, if the Löwner ordering $\mathrm{Cov}(\mathbf{G}\mathbf{y}) \leq \mathrm{Cov}(\mathbf{F}\mathbf{y})$ holds for every linear unbiased estimator $\mathbf{F}\mathbf{y}$ of $\boldsymbol{\beta}$.

Theorem 2.1 (The fundamental equations of the BLUE)**.** The linear estimator $\mathbf{G}\mathbf{y}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$ if and only if the matrix $\mathbf{G}$ satisfies the equations

$$\mathbf{G}(\mathbf{X} : \mathbf{M}) = (\mathbf{I} : \mathbf{0}). \tag{2.24}$$

Proof. See

Rao, C.R. (1973). Representations of best linear unbiased estimators in the Gauss–Markoff model with a singular dispersion matrix. *J. Multivariate Anal.*, **3**, 276–292. □

– It can be shown that the maximum likelihood estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ under normality is also the best linear unbiased estimator of $\boldsymbol{\beta}$.

– The linear estimator $\mathbf{G}\mathbf{y}$ is the best linear unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$ if and only if the matrix $\mathbf{G}$ satisfies the equations

$$\mathbf{G}(\mathbf{X} : \mathbf{M}) = (\mathbf{X} : \mathbf{0}). \tag{2.25}$$

– The linear estimator $\mathbf{G}\mathbf{y}$ is the best linear unbiased estimator of $\mathbf{K}'\boldsymbol{\beta}$ if and only if the matrix $\mathbf{G}$ satisfies the equations

$$\mathbf{G}(\mathbf{X} : \mathbf{M}) = (\mathbf{K}' : \mathbf{0}). \tag{2.26}$$

– Consider the partitioned linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}),$$

where $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$. The linear estimator $\mathbf{Gy}$ is the best linear unbiased estimator of $\boldsymbol{\beta}_2$ if and only if the matrix $\mathbf{G}$ satisfies the equations

$$\mathbf{G}(\mathbf{X}_1 : \mathbf{X}_2 : \mathbf{M}) = (\mathbf{0} : \mathbf{I} : \mathbf{0}). \tag{2.27}$$

Example 2.2.

Let us show that in the partitioned linear model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, the linear estimator

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y} \tag{2.28}$$

is the best linear unbiased estimator of $\boldsymbol{\beta}_2$, where $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$.

In showing, we need a property that for the orthogonal projector $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ the following decomposition holds

$$\begin{aligned}\mathbf{P}_\mathbf{X} = \mathbf{P}_{(\mathbf{X}_1:\mathbf{X}_2)} &= \mathbf{P}_{\mathbf{X}_1} + \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2} \\ &= \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' + \mathbf{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1,\end{aligned} \tag{2.29}$$

since $\mathbf{M}_1\mathbf{M}_1 = \mathbf{M}_1$, see Matrix Tricks for Linear Statistical Models : Our Personal Top Twenty, (2011), Chapter 8.

Let us now write the linear estimator as $\hat{\boldsymbol{\beta}}_2 = \mathbf{G}\mathbf{y} = (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1\mathbf{y}$, where $\mathbf{G} = (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1$. Hence, based on (2.27), the estimator $\hat{\boldsymbol{\beta}}_2$ is the BLUE of $\boldsymbol{\beta}$ if and only if the equations

$$(\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1(\mathbf{X}_1 : \mathbf{X}_2 : \mathbf{M}) = (\mathbf{0} : \mathbf{I} : \mathbf{0})$$

holds. Since $\mathrm{M}_1\mathbf{X}_1 = \mathbf{0}$, then clearly $(\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_1 = \mathbf{0}$. Also straightforwardly $(\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2 = \mathbf{I}$. And finally, we also have that

$$\begin{aligned}
(\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1\mathbf{M} &= (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \\
&= (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1} + \mathbf{P}_{\mathrm{M}_1\mathbf{X}_2}) \\
&= (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1(\mathrm{M}_1 - \mathbf{P}_{\mathrm{M}_1\mathbf{X}_2}) \\
&= (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1 - (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1 \\
&= (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1 - (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1 = \mathbf{0}. \qquad (2.30)
\end{aligned}$$

Thus $\mathbf{G} = (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1$ satisfies the fundamental BLUE equations of (2.27), and hence $\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathrm{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathrm{M}_1\mathbf{y}$ is the BLUE of $\boldsymbol{\beta}$.

2.2 Hypothesis Testing in Linear Models

2.2.1 Wald Test Statistic

– In linear model $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, consider general linear hypotheses

$$H_0 : \mathbf{K}'\boldsymbol{\beta} = \mathbf{0},$$
$$H_1 : \mathbf{K}'\boldsymbol{\beta} \neq \mathbf{0}, \qquad \mathbf{K}' \in \mathbb{R}^{q,(p+1)}.$$

– If the null hypothesis H_0 holds, then

$$\mathbf{K}'\hat{\boldsymbol{\beta}} \sim N(\mathbf{0}, \sigma^2\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}) \tag{2.31}$$

and furthermore

$$Q = \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{\sigma^2} \sim \chi^2_q, \tag{2.32}$$

where χ^2_q denotes the central χ^2-distribution with q degrees freedom.

– If H_0 holds, the Wald test statistic is following the distribution

$$W = \frac{Q/q}{\frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{\sigma^2}/n-(p+1)} = \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'(\tilde{\sigma}^2\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{q}$$
$$= \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'\left(\widetilde{\mathrm{Cov}}(\mathbf{K}'\hat{\boldsymbol{\beta}})\right)^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{q} \sim F_{q,n-(p+1)}, \tag{2.33}$$

where $F_{q,n-(p+1)}$ denotes the central F-distribution with degrees freedoms q and $n-(p+1)$.

– In Wald test, the p-value is obtained as the probability $p = P(F_{q,n-(p+1)} > W_{obs})$, where W_{obs} is the calculated observed value of the Wald test statistic.

– If the matrix $\mathbf{K}'$ is chosen as containing the given values of the explanatory variables $\mathbf{x}_*$, then

$$\hat{\mu}_* = \mathbf{x}_*'\hat{\boldsymbol{\beta}} = \mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \sim N(\mathbf{x}_*'\boldsymbol{\beta}, \sigma^2\mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*). \tag{2.34}$$

– Hence the following also hold:

$$\frac{\mathbf{x}_*'\hat{\boldsymbol{\beta}} - \mathbf{x}_*'\boldsymbol{\beta}}{\sqrt{\sigma^2\mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*}} \sim N(0,1), \quad \frac{\frac{\mathbf{x}_*'\hat{\boldsymbol{\beta}} - \mathbf{x}_*'\boldsymbol{\beta}}{\sqrt{\sigma^2\mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*}}}{\sqrt{\frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{\sigma^2}/n-(p+1)}} \sim t_{n-(p+1)}, \quad \frac{\mathbf{x}_*'\hat{\boldsymbol{\beta}} - \mathbf{x}_*'\boldsymbol{\beta}}{\sqrt{\tilde{\sigma}^2\mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*}} \sim t_{n-(p+1)}.$$

– Thus the $100(1-\alpha)\%$ confidence interval for the expected value $\mu_* = \mathbf{x}_*'\boldsymbol{\beta}$ has the lower- and the upperbounds as

$$\left[\mathbf{x}_*'\hat{\boldsymbol{\beta}} - t_{\alpha/2}\sqrt{\tilde{\sigma}^2\mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*}, \mathbf{x}_*'\hat{\boldsymbol{\beta}} + t_{\alpha/2}\sqrt{\tilde{\sigma}^2\mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*}\right]. \tag{2.35}$$

2.2.2 Likelihood Ratio Test

– In a partitioned linear model $\mathbf{y} \sim N(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \sigma^2\mathbf{I})$, let us test the significance of the parameters $\boldsymbol{\beta}_2$.

– The considered hypotheses are thus

$$\begin{aligned} H_0 &: \text{Model } \boldsymbol{\mu} = \mathbf{X}_1\boldsymbol{\beta}_1 \text{ holds,} \\ H_1 &: \text{Model } \boldsymbol{\mu} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 \text{ holds.} \end{aligned}$$

– If σ^2 is assumed to be known, the likelihood ratio statistic (minus two times natural logarithm of the ratio of the maximum likelihoods) has the form

$$\begin{aligned} LR &= -2 \cdot \log \left(\frac{\max_{\boldsymbol{\beta}_1} L_{H_0}(\boldsymbol{\beta}_1|\mathbf{y})}{\max_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2} L_{H_1}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2|\mathbf{y})} \right) = 2 \cdot l_{H_1}(\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2|\mathbf{y}) - 2 \cdot l_{H_0}(\hat{\boldsymbol{\beta}}_1|\mathbf{y}) \\ &= 2\left(-\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(\mathbf{y}'\mathbf{M}\mathbf{y})}{2\sigma^2}\right) - 2\left(-\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(\mathbf{y}'\mathbf{M}_1\mathbf{y})}{2\sigma^2}\right) \\ &= \frac{\mathbf{y}'\mathbf{M}_1\mathbf{y} - \mathbf{y}'\mathbf{M}\mathbf{y}}{\sigma^2} = \frac{\mathbf{y}'(\mathbf{M}_1 - \mathbf{M})\mathbf{y}}{\sigma^2}, \end{aligned} \tag{2.36}$$

where

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{I} - \mathbf{P}_{\mathbf{X}_1} = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1', \\ \mathbf{M} &= \mathbf{I} - \mathbf{P}_{\mathbf{X}} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \end{aligned}$$

– Since

$$\mathbf{M}_1 - \mathbf{M} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}_1}) - (\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = \mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}_1}$$
$$= (\mathbf{P}_{\mathbf{X}_1} + \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}) - \mathbf{P}_{\mathbf{X}_1} = \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}, \tag{2.37}$$

the statistic LR has the expression

$$LR = \frac{\mathbf{y}'\mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{y}}{\sigma^2}. \tag{2.38}$$

– If H_0 holds, then $\sigma^2 \cdot LR \sim \chi_q^2$ asymptotically.

– If H_0 holds, the likelihood ratio type test statistic is following the distribution

$$LRF = \frac{LR/q}{\tilde{\sigma}^2/\sigma^2} = \frac{\mathbf{y}'\mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{y}/q}{\tilde{\sigma}^2} \sim F_{q,n-(p+1)}. \tag{2.39}$$

– In likelihood ratio test, the p-value is obtained as the probability $p = P(F_{q,n-(p+1)} > LRF_{obs})$, where LRF_{obs} is the calculated observed value of the likelihood ratio test statistic.

2.3 Prediction and Predictive Effect Size in Linear Models

2.3.1 BLUP and Prediction Interval

– In case of linear models, consider the prediction of the new observation Y_f (observable in future) with given values of the explanatory variables $\mathbf{x}_f$:

$$\begin{pmatrix} \mathbf{y} \\ Y_f \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu} \\ \mu_f \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}' & 1 \end{pmatrix}\right], \tag{2.40a}$$

$$\begin{pmatrix} \boldsymbol{\mu} \\ \mu_f \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{x}_f'\boldsymbol{\beta} \end{pmatrix}. \tag{2.40b}$$

– The maximum likelihood predictor (the best linear unbiased predictor) for the new observation is

$$\hat{Y}_f = \hat{\mu}_f = \mathbf{x}_f'\hat{\boldsymbol{\beta}} = \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.41}$$

with prediction error $e_f = Y_f - \hat{Y}_f$ following normal distribution

$$Y_f - \hat{Y}_f \sim N\left(0, \sigma^2(1 + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f)\right). \tag{2.42}$$

– Thus the $100(1-\alpha)\%$ prediction interval for the new observation Y_f has the the lower and the upper points

$$\left[\mathbf{x}_f'\hat{\boldsymbol{\beta}} - t_{\alpha/2}\sqrt{\tilde{\sigma}^2(1 + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f)}, \mathbf{x}_f'\hat{\boldsymbol{\beta}} + t_{\alpha/2}\sqrt{\tilde{\sigma}^2(1 + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f)}\right]. \tag{2.43}$$

Example 2.3.

Let us show that prediction error $e_f = Y_f - \hat{Y}_f$ has the variance

$$\mathrm{Var}(e_f) = \sigma^2(1 + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f). \tag{2.44}$$

The prediction error $e_f = Y_f - \hat{Y}_f$ can be written as

$$Y_f - \hat{Y}_f = (-\mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' : 1)\begin{pmatrix}\mathbf{y}\\ Y_f\end{pmatrix} = \mathbf{b}'\mathbf{z}, \tag{2.45}$$

where

$$\mathbf{b} = \begin{pmatrix}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f\\ 1\end{pmatrix}, \qquad \mathbf{z} = \begin{pmatrix}\mathbf{y}\\ Y_f\end{pmatrix}.$$

Denoting

$$\boldsymbol{\Sigma} = \sigma^2\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0}' & 1\end{pmatrix},$$

the prediction error $e_f = Y_f - \hat{Y}_f$ has the variance

$$
\begin{aligned}
\mathrm{Var}(Y_f - \hat{Y}_f) &= \mathrm{Var}(\mathbf{b}'\mathbf{z}) = \mathbf{b}'\boldsymbol{\Sigma}\mathbf{b} \\
&= (-\mathbf{x}'_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' : 1)\left[\sigma^2\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0}' & 1\end{pmatrix}\right]\begin{pmatrix}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f\\ 1\end{pmatrix} \\
&= \sigma^2(-\mathbf{x}'_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' : 1)\begin{pmatrix}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f\\ 1\end{pmatrix} \\
&= \sigma^2(\mathbf{x}'_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f + 1) = \sigma^2(1 + \mathbf{x}'_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f). \qquad (2.46)
\end{aligned}
$$

2.3.2 Predictive Effect Size and d-value

- Consider evaluating the effect on the response variable Y when the values of explanatory variables are changed from the set of values $\mathbf{x}_{1f}$ to the values $\mathbf{x}_{2f}$.
- Let the random variables Y_{1f} and Y_{2f} be unobserved values of the response variable Y given the explanatory values $\mathbf{x}_{1f}$ and $\mathbf{x}_{2f}$, respectively.
- Let us measure the effect size by predicting the difference $Y_{2f} - Y_{1f}$.
- In linear models, the maximum likelihood predictor for the difference $Y_{2f} - Y_{1f}$ is

$$\hat{Y}_{2f} - \hat{Y}_{1f} = (\mathbf{x}_{2f} - \mathbf{x}_{1f})'\hat{\boldsymbol{\beta}} = (\mathbf{x}_{2f} - \mathbf{x}_{1f})'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.47}$$

  with the $100(1-\alpha)\%$ prediction interval for the difference $Y_{2f} - Y_{1f}$ being

$$\left[(\mathbf{x}_{2f} - \mathbf{x}_{1f})'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2}\sqrt{\tilde{\sigma}^2(2 + (\mathbf{x}_{2f} - \mathbf{x}_{1f})'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x}_{2f} - \mathbf{x}_{1f}))}\right] \tag{2.48}$$

- If the $100(1-\alpha)\%$ prediction interval does not contain the value **zero**, we may make inference that there are significant predictive effect size difference between the values $\mathbf{x}_{1f}$ and $\mathbf{x}_{2f}$ at $\alpha\%$ significance level.

– Note that

$$T = \frac{(\mathbf{x}_{2f} - \mathbf{x}_{1f})'\hat{\boldsymbol{\beta}} - (Y_{2f} - Y_{1f})}{\sqrt{\tilde{\sigma}^2(2 + (\mathbf{x}_{2f} - \mathbf{x}_{1f})'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x}_{2f} - \mathbf{x}_{1f}))}} \sim t_{n-(p+1)}. \tag{2.49}$$

– Inference made by use of prediction interval for the difference $Y_{2f} - Y_{1f}$ is equivalent for testing predictive hypothesis

$$H_0 : y_{1f} = y_{2f},$$
$$H_1 : y_{1f} \neq y_{2f}.$$

– The test statistic for the predictive hypothesis is

$$T = \frac{(\mathbf{x}_{2f} - \mathbf{x}_{1f})'\hat{\boldsymbol{\beta}}}{\sqrt{\tilde{\sigma}^2(2 + (\mathbf{x}_{2f} - \mathbf{x}_{1f})'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x}_{2f} - \mathbf{x}_{1f}))}} \tag{2.50}$$

– Similarly as p-value is calculated in testing of parameters, the so-called d-value is obtained for the predictive hypothesis as

$$d = 2 \times P(t_{n-(p+1)} > |T_{obs}|). \tag{2.51}$$

Assignment 2.1.

In a rat study, it was investigated how much weight gains in rats occur within 85 days when they are exposed to different amounts of recombinant bovine growth hormone (rBGH). The full dataset can be found in file ratsRBGH.txt.

```
      gender rbGH weightgain
1          1    1     274.99
2          1    1     289.67
3          1    1     346.40
4          1    1     344.32
5          1    1     364.63
6          1    2     478.62

.
.
59         2    6     164.93
60         2    6     177.85

Description: Weight gains in rats over 85 day period in 6 treatment
conditions of recombinant bovine growth hormone (rbGH):
1=Control (0 mg/kg per day)
2=Subcutaneous Injection (1.0 mg/kg per day)
3=Oral Glavage (0.1 mg/kg per day)
4=Oral Glavage (0.5 mg/kg per day)
5=Oral Glavage (5 mg/kg per day)
6=Oral Glavage (50 mg/kg per day)

Gender: 1=Male, 2=Female

Source: J.C. Juskevich and C.G. Guyer (1990). "Bovine Growth Hormone: Human
Food Safety Evaluation," Science, Vol.249,#4971,pp875-884.
```

Denote explanatory variables as X_1=rbGH and X_2=gender. Consider modelling the response variable Y=weightgain by following two different models:

$$\begin{aligned} \mathcal{M}_{1|2}: \quad & Y_i = \beta_0 + \beta_j + \alpha_h + \varepsilon_i, \\ \mathcal{M}_{12}: \quad & Y_i = \beta_0 + \beta_j + \alpha_h + \gamma_{jh} + \varepsilon_i, \end{aligned}$$

where in each model the random error term ε_i is assumed to follow normal distribution $\varepsilon_i \sim N(0, \sigma^2)$.

Consider the following hypotheses

$$\begin{aligned} H_0 &: \text{Model } \mathcal{M}_{1|2} \text{ is the true model,} \\ H_1 &: \text{Model } \mathcal{M}_{12} \text{ is the true model.} \end{aligned}$$

Use the Wald statistic

$$\begin{aligned} W &= \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'(\tilde{\sigma}^2\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{q} \\ &= \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'\left(\widetilde{\mathrm{Cov}}(\mathbf{K}'\hat{\boldsymbol{\beta}})\right)^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{q} \sim F_{q,n-(p+1)}, \end{aligned}$$

to test the above hypotheses.

Assignment 2.2.

The research group was interested in finding out how vitamin C affects the length of the teeth of the guinea pigs. In the study, a total of 60 guinea pigs were divided into three equal groups, with 0.5 milligrams of vitamin C per day administered to guinea pigs in the first group, 1 mg of vitamin C per day, and 2 mg of vitamin C per day. In addition, each of these three groups was divided into two so that half of the group's guinea pigs received vitamin C in the juice (OJ) and half as crystals (VC). The teeth of the guinea pig were measured prior to the beginning of the experimentation and subsequent measurements were made, see file teethpig.txt.

	growth	dose	level
1	4.2	VC	0.5
2	11.5	VC	0.5
3	7.3	VC	0.5
4	5.8	VC	0.5
.			
59	29.4	OJ	2.0
60	23.0	OJ	2.0

Let us denote the explanatory variables as $X_1 = \mathsf{level}$ and $X_2 = \mathsf{dose}$, and let us consider the following linear covariance model

$$\mathcal{M}_{12}: \quad Y_i = \beta_0 + \beta_1 x_{i1} + \alpha_j + \gamma_j x_{i1} + \varepsilon_i.$$

Consider the predictive effect size in case of dose changed from juice to crystals with the level hold on $X_1 = 0.5$. Construct appropriate 80 % prediction interval.

Assignment 2.3.

Let us assume $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. Show that the maximum likelihood estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is also the BLUE of $\boldsymbol{\beta}$. Consider then the estimator $\mathbf{X}\hat{\boldsymbol{\beta}}$. Calculate the expected value $\mathrm{E}(\mathbf{X}\hat{\boldsymbol{\beta}})$ and the covariance matrix $\mathrm{Cov}(\mathbf{X}\hat{\boldsymbol{\beta}})$. What distribution residuals $\mathbf{X}\hat{\boldsymbol{\beta}}$ are following?

Assignment 2.4.

In a rat study, it was investigated how much weight gains in rats occur within 85 days when they are exposed to different amounts of recombinant bovine growth hormone (rBGH). The full dataset can be found in file ratsRBGH.txt.

```
     gender rbGH weightgain
1         1    1     274.99
2         1    1     289.67
3         1    1     346.40
4         1    1     344.32
5         1    1     364.63
6         1    2     478.62

.
.
59        2    6     164.93
60        2    6     177.85

Description: Weight gains in rats over 85 day period in 6 treatment
conditions of recombinant bovine growth hormone (rbGH):
1=Control (0 mg/kg per day)
2=Subcutaneous Injection (1.0 mg/kg per day)
3=Oral Glavage (0.1 mg/kg per day)
4=Oral Glavage (0.5 mg/kg per day)
5=Oral Glavage (5 mg/kg per day)
6=Oral Glavage (50 mg/kg per day)

Gender: 1=Male, 2=Female

Source: J.C. Juskevich and C.G. Guyer (1990). "Bovine Growth Hormone: Human
Food Safety Evaluation," Science, Vol.249,#4971,pp875-884.
```

Denote explanatory variables as X_1=rbGH and X_2=gender. Consider modelling the response variable Y=weightgain by the following model:

$$\mathcal{M}_{1|2}: \quad Y_i = \beta_0 + \beta_j + \alpha_h + \varepsilon_i,$$

where the random error term ε_i is assumed to follow normal distribution $\varepsilon_i \sim N(0, \sigma^2)$.

(a) Test all pairwise average differences

$$H_0: \left(\frac{\mu_{1h} + \mu_{2h} + \cdots + \mu_{kh}}{k}\right) - \left(\frac{\mu_{1h_*} + \mu_{2h_*} + \cdots + \mu_{kh_*}}{k}\right) = 0, \qquad h \neq h_*,$$

i.e., test whether there is average differences on means in the levels of X_2=gender variable.

(b) Test all pairwise predictive average differences

$$H_0: \left(\frac{Y_{i1h} + Y_{i2h} + \cdots + Y_{ikh}}{k}\right) - \left(\frac{Y_{i1h_*} + Y_{i2h_*} + \cdots + Y_{ikh_*}}{k}\right) = 0, \qquad h \neq h_*,$$

i.e., test whether there is average differences on random variables in the levels of X_2=gender variable.