作业 **5**

1. 简单随机样本 $y_1,...,y_n$ iid $\sim(\mu,\sigma^2)$ 是最简单 (不含自变量) 的线性模型, 这是因为该模型可写为:

$$y_i=\mu+\epsilon_i, i=1,...,n, \quad \epsilon_i, i=1,...,n \ \ \text{iid} \ \sim(0,\sigma^2)$$

其中 $\mu$ 是未知参数。使得误差平方和 $\sum_{i=1}^n(y_i-\mu)^2$ 最小的 $\mu$ 称为最小二乘估计，记为 $\hat{\mu}$ 。求 $\hat{\mu}$ 及其方差。$y_i$ 的拟合值 $\hat{y}_i$ 和残差 $e_i$ 应如何定义?

2. 假设随机样本数据 $(x_i,y_i), i=1,...,n$ 满足下述过原点的线性回归模型（无截距项)

$$y_i=bx_i+\epsilon_i, E(\epsilon_i)=0, \mathrm{var}(\epsilon_i)=\sigma^2, \text{且 } x_i \text{ 与 } \epsilon_i \text{ 独立}, i=1,...,n.$$

通过极小化误差平方和 $\sum_{i=1}^n(y_i-bx_i)^2$，求未知参数 $b$ 的 LS 估计 $\widehat{b}$, 并求其方差。

> 第 3-5 题基于简单线性回归模型：假设独立样本 $(x_i,y_i), i=1,...,n$ 满足下述模型
>
> $$y_i=a+bx_i+\epsilon_i, \epsilon_i\sim(0,\sigma^2), \text{且 } x_i \text{ 与 } \epsilon_i \text{ 独立}, i=1,...,n.$$
>
> 未知参数 $a,b,\sigma^2$ 的 LS 估计分别记为 $\hat{a},\hat{b},\hat{c}$。

3. 证明给定 $x_1,...,x_n$ 条件下

$$var(\hat{a})=\sigma^2/n+\bar{x}^2\sigma^2/s_{xx}, \quad var(\hat{b})=\sigma^2/s_{xx}, \quad cov(\hat{a},\hat{b})=-\bar{x}\sigma^2/s_{xx}$$

何时 $\hat{a},\hat{b}$ 不相关?

4. 定义最小二乘得到的残差 $e_i=y_i-\hat{y}_i=y_i-\hat{a}-\hat{b}x_i$, 证明

$$E(e_i)=0, \quad var(e_i|\mathbf{x})=\left(1-\frac{1}{n}-\frac{(x_i-\bar{x})^2}{s_{xx}}\right)\sigma^2,$$

由此证明 $\hat{\sigma}^2$ 是 $\sigma^2$ 的无偏估计。

5. 假设 $x_i=0$ 或 $1, i=1,...,n$, 记 $m=\sum x_i$ 为 1 的个数。当 $m$ 等于或近似等于 $n/2$ 时，称为是均衡设计 (balanced design)，否则是不均衡的 (unbalanced)。基于 $b$ 的 LS 估计的精确性 (方差大小)，讨论均衡设计的优越性。