作业 **6**

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 μ 是未知参数。使得误差平方和 $\sum_{i=1}^n(y_i-\mu)^2$ 最小的 μ 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)

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

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 3-6 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,σ^2 的 LS 估计分别记为 $\hat{a},\hat{b},\hat{\sigma}^2$。
>
> 记号：$\mathbf{x}=(x_1,...,x_n)^\top$， $s_{ab}=\sum(a_i-\bar{a})(b_i-\bar{b})$, $s_{aa}=\sum(a_i-\bar{a})^2$。

3. 证明给定 $\mathbf{x}=(x_1,...,x_n)^\top$ 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. 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 估计的方差，讨论均衡设计的优越性。

5. 定义最小二乘得到的残差 $e_i=y_i-\hat{y}_i=y_i-\hat{a}-\hat{b}x_i$, 残差平方和定义为 $RSS=\sum_{i=1}^n e_i^2$。证明 $e_i=(\epsilon_i-\bar{\epsilon})-(x_i-\bar{x})s_{x\epsilon}/s_{xx}$, 并利用该表达证明

$$RSS=s_{yy}-s_{xy}^2/s_{xx}=s_{\epsilon\epsilon}-s_{x\epsilon}^2/s_{xx}.$$

6. 定义点 (x_i,y_i) 与 $(\bar{x},\bar{y})$ 决定的直线的斜率为 $k_i=\frac{y_i-\bar{y}}{x_i-\bar{x}}$, 当 $x_i=\bar{x}$ 时定义 $k_i=0$，$i=1,...,n$。则 b 的最小二乘估计 $\hat{b}=s_{xy}/s_{xx}$ 可表示成所有斜率 $k_i, i=1,...,n$ 的加权和

$$\hat{b}=\frac{1}{s_{xx}}\sum(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{s_{xx}}\sum(x_i-\bar{x})^2\left(\frac{y_i-\bar{y}}{x_i-\bar{x}}\right)=\sum w_{0i}k_i$$

其中权重 $w_{0i}=(x_i-\bar{x})^2/s_{xx}$, $\sum w_{0i}=1$ 。对任何序列 $w_1,...,w_n$ (只依赖于 $x_1,...,x_n$), $w_i\geq 0, \sum w_i=1$, 定义 b 的一个估计

$$\widetilde{b}=\sum w_ik_i=\sum w_i\left(\frac{y_i-\bar{y}}{x_i-\bar{x}}\right).$$

(a) 证明 $E(\widetilde{b})=b$.

(b) 求 $var(\widetilde{b}|\mathbf{x})$.

(c) 证明 $var(\widetilde{b}|\mathbf{x})\geq\sigma^2/s_{xx}=var(\hat{b}|\mathbf{x})$.