

回归分析题目回忆版

1. 设有

$$Y = X\beta + \xi, \quad \xi \sim N(0, \sigma^2 I_n).$$

设 A 为 $p \times n$ 矩阵, 且满足 $(AX)^{-1}$ 存在, 令

$$\tilde{\beta} = (AX)^{-1}AY.$$

(a) 证明 $\tilde{\beta}$ 为 β 的无偏估计。

(b) 证明 $\text{MSE}(\hat{y}^*) > \text{MSE}(\hat{y})$, 其中 $\hat{y}^* = x'\tilde{\beta}$ 。

2. 设带惩罚项的目标函数为

$$\|y - X\beta\|^2 + \lambda\|\beta\|^2, \quad \lambda > 0.$$

(a) 推出 β 此时的估计。

(b) 对如何确定 λ 做简要说明。

3. 设有参数 y, x_1, x_2, \dots, x_p , 其中

$$x_i = (x_{i1}, x_{i2}, \dots, x_{in}), \quad y = (y_1, y_2, \dots, y_n).$$

描述引入和剔除变量的过程, 并给出相应的统计量。

4. 有 m 个分组, 每个分组有 n_i 个事件, r_i 为发生次数, x_i 为参数或协变量。题中提示:

$$\hat{v}_i = \frac{1}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}.$$

(a) 写出 Logistic (logit) 模型。

(b) 若 $z_i = 1$ 和 $z_i = 0$, 该如何进行修正。

5. 从网络购买 $j = 1, 2, 3$ 样物品, 商家称总重量 (三件) 为 W 。现分别称重得 y_1, y_2, y_3 。设 ξ_i 为每次称重的误差, 互不相关, 且

$$E(\xi_i) = 0, \quad \text{Var}(\xi_i) = \sigma^2.$$

设真实重量分别为 $\theta_1, \theta_2, \theta_3$ 。

(a) 构造模型。

(b) 做假设检验, 判断是否显著; $\hat{\sigma}^2$ 的无偏估计是什么。

6. 设药品的品类为因素 A ，表格中数字 y_{ij} 代表药效生效时间。原稿旁注：数字不一定对。

药品类别	观测值
A (药名不清)	0.0, 1.0, 0.18, 0.23
新药品	0.195, 0.212
阿斯匹林	0.260, 0.275, 0.196

- (a) 构造单因素方差分析模型。
 (b) 求 $\mu_2 - \mu_3$ 的 95% 置信区间，并检验新药和阿斯匹林之间药效是否显著。

7. 设有

$$y = X\beta + \xi, \quad \xi \sim N(0, \sigma^2 I).$$

- (a) 求 $\hat{\beta}_{OLS}$ 。
 (b) 判断 $\hat{\beta}_{OLS}$ 是否无偏。
 (c) 写出 A 最优准则、 D 最优准则、 E 最优准则。