

学号:

姓名:

[25分] (1) 对于多元线性回归模型 $Y = X\beta + \epsilon$, 其中 $Y \in R^{n \times 1}$, $X \in R^{n \times p}$, $\beta \in R^{p \times 1}$, $\epsilon \in R^{n \times 1}$. 请写出高斯-马尔科夫定理, 证明最小二乘估计是BLUE. (2) 岭回归估计和Lasso估计是在什么情况下提出的? 其定义和统计思想是什么? 如何选择其中的调整参数? (3) 当 $n^{-1}X^T X = I_p$ 时, 记最小二乘估计是 $\hat{\beta}_{ls}$, 请写出岭回归估计和Lasso估计的表达式. (4) 给出上述两个估计量偏差和方差的表达式, 并与最小二乘估计进行比较.

2. [25分] 设 Y 是 q 维随机变量且服从指数分布族

$$f(y; \theta, \phi) = \exp \left\{ \frac{y^T \theta - b(\theta)}{\psi} + c(y, \psi) \right\},$$

其中 θ 是自然参数, $\psi > 0$ 是扩散参数, $b(\cdot)$ 和 $c(\cdot, \cdot)$ 是确定的函数. 假设 Y 在给定协变量 X 的条件下服从上述指数分布族, 且 X 通过线性预报 $X\beta$ 来影响 Y . (1) 当 Y 自于正态分布、指数分布、二项分布、泊松分布时, 分别给出它们的典则连接函数. (2) 请推导 β 的估计方程, 并给出指数分布和二项分布下具体的形式.

3. [15分] 某医院病人在手术后的寿命结果如下, 请给出生存函数的估计.

区间	死亡数	删失数	区间	死亡数	删失数
(0, 2]	166	5	(2, 4]	76	12
(4, 6]	27	14	(6, 8]	5	12
(8, 10]	5	10	(10, 12]	4	15
(12, 14]	8	13	(14, ∞]	28	0

4. [10分] 某医院对10位病人手术后的生存时间进行了调查, 记录如下(单位: 年):

11, 11⁺, 7⁺, 7, 3, 3⁺, 16, 15⁺, 20, 4,

请给出生存函数的估计.

5. [25分] 在临床试验中, 得到两组数据如下:

1) 治疗组: 9, 13, 13⁺, 18, 23, 28⁺, 31, 34, 45⁺, 48, 161⁺,

12) 控制组: 5, 5, 8, 8, 12, 16⁺, 23, 27, 30, 33, 43, 45.

(1) 请画出治疗组和控制组的生存函数曲线; (2) 设两组数据的分布函数为 $S_1(t)$ 和 $S_2(t)$, 请给出两种不同的方法检验假设 $H_0: S_1(t) = S_2(t)$ 对任意 $t > 0$ 是否成立.