

- (A) probit 模型的预测概率一般更优 (B) 逻辑回归的预测概率排序一般更优
 (C) 两个模型的预测概率差不多 (D) 以上都不对

7) (5分)对 $2 \times 2 \times 2$ 列联表数据, 假设模型 $P(Y = 1|X = x, Z = k) = \text{expit}(\alpha_k + \beta_k x)$ 成立, 则对 (1) $\beta_1 = -1, \beta_2 = 1$ (2) $\beta_1 = \beta_2 = 0$ (3) $\beta_1 = -1, \beta_2 = 0$ (4) $\beta_1 = \beta_2 = 1$ 这四个情形, 直觉上似然比检验功效的顺序应该是 ()

- (A) (1)<(2)<(3)<(4) (B) (1)<(3)<(2)<(4) (C) (2)<(1)<(3)<(4) (D) (2)<(3)<(1)<(4)

2. 解答题 (共 75 分)

1) (8分) 据统计, 美国佛罗里达州的死亡率高于美国平均水平, 但该州作为热门度假、养老胜地, 人们趋之若鹜。请运用统计学相关知识解释这一看似矛盾的现象。

2) (17分) 设 Y_1, \dots, Y_n 是来自二项分布总体 $B(n, p)$ 的简单样本, 试先构造 $\text{logit}(p)$ 的 $(1 - \alpha)$ Wald 置信区间, 进而得到 p 的 $(1 - \alpha)$ 置信区间, 所得结果相对于直接基于 p 的 MLE 的正态性构造的置信区间有何优势?

3) (25分) 记 $(X_1, Y_1), \dots, (X_n, Y_n)$ 为二元总体 (X, Y) ($X, Y = 0, 1$) 的简单样本, 相应的配对列联表数据记为 $n_{jk} = \#\{i: X_i = j, Y_i = k\}$ ($j, k = 0, 1$)。记 $\pi_{jk} = P(X=j, Y=k)$ ($j, k = 0, 1$), $q = \pi_{10}/(\pi_{10} + \pi_{01})$ 。考虑假设检验问题 $H_0: q = 1/2 \leftrightarrow H_1: q \neq 1/2$ 。

- (i) 写出观测数据 $(n_{00}, n_{01}, n_{10}, n_{11})$ 的似然函数 (参数包含 q 和若干冗余参数);
- (ii) 推导 Wald 检验统计量 (不是得分检验统计量), 其零极限分布是什么?
- (iii) 写出相应检验的 P 值表达式, 并推导其零极限分布。

4) (25分) 某研究者为分析企业创新专利申请数量的影响因素, 随机选取国内 n 家制造业企业连续 T 年的平衡面板数据进行实证研究。记 y_{it} 为第 i 家企业在第 t 年申请的专利数 (响应变量), z_{it} 为 p 维解释变量 (包含企业当年研发投入强度、企业规模、资产负债率等)。假设 $y_{it} \sim \text{Poisson}(\lambda_{it})$, $\lambda_{it} = \exp(\alpha_i + \beta' z_{it})$, $i = 1, \dots, n, t = 1, \dots, T$ 。

- (i) 应用条件似然法构造 β 的 MLE $\hat{\beta}$ 满足的方程;
- (ii) 求 $\hat{\beta}$ 的渐近分布。