

HOMEWORK 6

一、证明题

设 $X = (X_1, \dots, X_m)^\top$ 是 m 维随机变量, 协方差矩阵为 $\text{cov}(X) \stackrel{\text{def}}{=} \Sigma$. 设矩阵 Σ 的特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$; 这 m 个特征值对应的单位特征向量为 $\alpha_1, \dots, \alpha_m$ (特征向量矩阵为 $(\alpha_1, \alpha_2, \dots, \alpha_m) \in \mathbb{R}^{m \times m}$, 其中 $\alpha_k = (\alpha_{1k}, \alpha_{2k}, \dots, \alpha_{mk})^\top$ 为列向量, 若用 Y_k 表示 X 的第 k 个主成分, 且 $Y = (Y_1, \dots, Y_m)^\top$, 证明:

1. $\text{cov}(Y) = \text{diag}(\lambda_1, \dots, \lambda_m)$;
2. $\sum_i \lambda_i = \sum \sigma_{ii} (\sigma_{ii} = \text{var}(X_i))$.
3. 因子负荷量为

$$\rho(Y_k, X_i) = \text{cor}(Y_k, X_i) = \frac{\sqrt{\lambda_k} \alpha_{ik}}{\sqrt{\sigma_{ii}}}$$

4. $\sum_{i=1}^m \sigma_{ii} \rho(Y_k, X_i)^2 = \lambda_k$
5. $\sum_k \rho^2(Y_k, X_i) = 1$

二、五大联赛球员数据分析

编程语言可以使用 R/python 等一切能满足题目要求的语言。具体任务见“五大联赛球员数据分析”任务文档。

最后以 HTML 的形式提交报告。报告中需包括题目内容中涉及的代码和相关文字解释、结果分析。(提示: R 语言可用 Rmarkdown/Quarto 输出分析报告; Python 可用 Jupyter 输出 HTML 报告; 也自己进行格式调整后输出 HTML 格式的报告)

提交时间: 5 月 29 日, 晚 20:00 之前。请预留一定的时间, 迟交作业扣 3 分, 作业抄袭 0 分。