《多元统计分析》课后作业

姓名： 李倩倩

学号： 2024017349

班级： 统计 24-1 班

中国石油大学（北京）克拉玛依校区文理学院数学与统计系

Friday 17th April, 2026

作业要求

1. 可以和其他同学讨论作业当中的问题，但应当自己独立完成作业
2. 计算、证明等要有过程，要有主要步骤的说明
3. 请将计算、绘图所用的 R 代码以及生成的结果和图像一并添加在作业文件当中
4. 请使用 LaTeX编辑并生成 PDF 格式的文件，第 X 周作业文件命名方式：学号-姓名-X.pdf
5. 评分标准：每一问得分 $\in \{2, 1, 0\}$
   - 2: 按时完成并上交作业，且答案基本正确
   - 1: 按时完成并上交作业，且答案部分正确
   - 0: 答案完全错误，或者迟交作业(规定时间72小时之后)
6. 请将完成的 PDF 格式的作业文件发送至邮箱：xiaolei@cup.edu.cn
7. 每位同学可以有一次迟交作业的机会，但不得晚于规定时间三日之后
8. 第 5 周作业截止时间：2026年4月17日24:00

目录

Chapter 1

第 5 周作业

第 5 周作业截止时间：2026年4月17日24:00

第 5 周作业完成时间：Friday 17th April, 2026 22:35

1. [2 分] 假设 $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}\,,\,\Sigma)$，其中

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}\,, \qquad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.1}$$

令

$$\mathcal{A} = (1\,,\,1)\,, \qquad \mathcal{B} = (1\,,\,-1) \tag{1.2}$$

证明 $\mathcal{A}\,\boldsymbol{X}$ 与 $\mathcal{B}\,\boldsymbol{X}$ 相互独立.

【证明】 由于 $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}\,,\,\Sigma)$，故

$$\begin{pmatrix} \mathcal{A}\,\boldsymbol{X} \\ \mathcal{B}\,\boldsymbol{X} \end{pmatrix} = \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} \boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \boldsymbol{X}$$

是 $\boldsymbol{X}$ 的线性变换，其仍服从二元正态分布. 计算其协方差：

$$\mathrm{Cov}\,(\mathcal{A}\,\boldsymbol{X}\,,\,\mathcal{B}\,\boldsymbol{X}) = \mathcal{A}\,\Sigma\,\mathcal{B}^{\mathrm{T}} = (1\,,\,1)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \times 1 + 1 \times (-1) = 0\,.$$

对于联合正态分布的随机变量，协方差为零等价于相互独立，故 $\mathcal{A}\,\boldsymbol{X}$ 与 $\mathcal{B}\,\boldsymbol{X}$ 相互独立. □

2. 假设

$$\boldsymbol{X} \sim N_2\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right), \quad (\boldsymbol{Y}\,|\,\boldsymbol{X}\,) \sim N_2\left(\begin{pmatrix} X_1 \\ X_1 + X_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \tag{1.3}$$

(a) [2 分] 确定 $Y_2\,|\,Y_1$ 的分布.

【解】 记 $C = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$，则 $\boldsymbol{Y}\,|\boldsymbol{X} \sim N_2\,(C\boldsymbol{X}\,,\,I_2)$，即 $\boldsymbol{Y} = C\boldsymbol{X} + \boldsymbol{\varepsilon}$，其中 $\boldsymbol{\varepsilon} \sim N_2\,(\boldsymbol{0}\,,\,I_2)$ 且

与 $\boldsymbol{X}$ 独立. 从而 $\boldsymbol{Y}$ 服从二元正态分布，其

$$\mathbb{E}(\boldsymbol{Y}) = C\,\mathbb{E}(\boldsymbol{X}) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

$$\mathrm{Var}(\boldsymbol{Y}) = C\,\mathrm{Var}(\boldsymbol{X})\,C^{\mathrm{T}} + I_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + I_2 = \begin{pmatrix} 2 & 3 \\ 3 & 6 \end{pmatrix} + I_2 = \begin{pmatrix} 3 & 3 \\ 3 & 7 \end{pmatrix}.$$

即 $\boldsymbol{Y} \sim N_2\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 & 3 \\ 3 & 7 \end{pmatrix}\right)$. 由二元正态的条件分布公式:

$$\mathbb{E}\,(Y_2\,|Y_1\,) = 3 + \frac{3}{3}(Y_1 - 1) = Y_1 + 2\,,$$

$$\mathrm{Var}\,(Y_2\,|Y_1\,) = 7 - \frac{3^2}{3} = 4\,.$$

故 $Y_2\,|Y_1 \sim N_1\,(Y_1 + 2\,,\,4)$.

(b) [2 分] 确定 $\boldsymbol{W} = \boldsymbol{X} - \boldsymbol{Y}$ 的分布.

【解】 $\boldsymbol{W} = \boldsymbol{X} - \boldsymbol{Y} = \boldsymbol{X} - C\boldsymbol{X} - \boldsymbol{\varepsilon} = (I_2 - C)\boldsymbol{X} - \boldsymbol{\varepsilon}$，其中 $I_2 - C = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}$. 由于 $\boldsymbol{X}$ 与 $\boldsymbol{\varepsilon}$ 独立且均为正态，$\boldsymbol{W}$ 服从二元正态分布，且

$$\mathbb{E}(\boldsymbol{W}) = \mathbb{E}(\boldsymbol{X}) - \mathbb{E}(\boldsymbol{Y}) = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

$$\begin{aligned} \mathrm{Var}(\boldsymbol{W}) &= (I_2 - C)\,\mathrm{Var}(\boldsymbol{X})\,(I_2 - C)^{\mathrm{T}} + I_2 \\ &= \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} + I_2 = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} + I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}. \end{aligned}$$

故 $\boldsymbol{W} \sim N_2\left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\right)$.

3. 假设

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N_3(\boldsymbol{\mu}\,,\,\Sigma) \tag{1.4}$$

若已知

$$Y\,|\,Z \sim N_1(-Z\,,\,1) \tag{1.5}$$

$$\mu_{Z|Y} = -\frac{1}{3} - \frac{1}{3}Y \tag{1.6}$$

$$(X\,|\,Y\,,\,Z) \sim N_1(2+2Y+3Z\,,\,1) \tag{1.7}$$

(a) [2 分] 计算 $\boldsymbol{\mu}$ 和 Σ.

【解】 记 $\boldsymbol{\mu} = (\mu_X, \mu_Y, \mu_Z)^{\mathrm{T}}$，$\Sigma = (\sigma_{ij})$.

由 $Y\,|Z \sim N_1(-Z\,,\,1)$，由正态条件分布公式得

$$\frac{\sigma_{YZ}}{\sigma_{ZZ}} = -1\,, \quad \mu_Y + \frac{\sigma_{YZ}}{\sigma_{ZZ}}(0-\mu_Z) = 0\,, \quad \sigma_{YY} - \frac{\sigma_{YZ}^2}{\sigma_{ZZ}} = 1\,.$$

即 $\sigma_{YZ} = -\sigma_{ZZ}$，$\mu_Y = -\mu_Z$，$\sigma_{YY} - \sigma_{ZZ} = 1$.

由 $\mu_{Z|Y} = -\frac{1}{3} - \frac{1}{3}Y$，得

$$\frac{\sigma_{YZ}}{\sigma_{YY}} = -\frac{1}{3}\,, \quad \mu_Z + \frac{\sigma_{YZ}}{\sigma_{YY}}(0-\mu_Y) = -\frac{1}{3}\,.$$

即 $\sigma_{YZ} = -\sigma_{YY}/3$，$\mu_Z + \mu_Y/3 = -1/3$.

由 $\sigma_{YZ} = -\sigma_{ZZ} = -\sigma_{YY}/3$，得 $\sigma_{YY} = 3\sigma_{ZZ}$. 代入 $\sigma_{YY} - \sigma_{ZZ} = 1$：$2\sigma_{ZZ} = 1$，故

$$\sigma_{ZZ} = \frac{1}{2}\,, \quad \sigma_{YY} = \frac{3}{2}\,, \quad \sigma_{YZ} = -\frac{1}{2}\,.$$

结合 $\mu_Y = -\mu_Z$ 与 $\mu_Z + \mu_Y/3 = -1/3$，得 $\mu_Z - \mu_Z/3 = -1/3$，即 $\mu_Z = -1/2$，$\mu_Y = 1/2$.

由 $(X\,|Y\,,\,Z) \sim N_1(2+2Y+3Z\,,\,1)$，设 $(\sigma_{XY}\,,\,\sigma_{XZ}) = (a\,,\,b)$，则

$$(a\,,\,b)\Sigma_{(Y,Z)}^{-1} = (2\,,\,3)\,,$$

其中 $\Sigma_{(Y,Z)} = \begin{pmatrix} 3/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}$，$|\Sigma_{(Y,Z)}| = 1/2$，$\Sigma_{(Y,Z)}^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$. 故

$$(a\,,\,b) = (2\,,\,3)\Sigma_{(Y,Z)} = \left(2\cdot\tfrac{3}{2} + 3\cdot(-\tfrac{1}{2})\,,\,2\cdot(-\tfrac{1}{2}) + 3\cdot\tfrac{1}{2}\right) = \left(\tfrac{3}{2}\,,\,\tfrac{1}{2}\right)\,.$$

常数项：$2 = \mu_X - (2,3)(\mu_Y,\mu_Z)^{\mathrm{T}} = \mu_X - (1-3/2) = \mu_X + 1/2$，得 $\mu_X = 3/2$. 条件方差：

$$1 = \sigma_{XX} - (2,3)(\sigma_{XY},\sigma_{XZ})^{\mathrm{T}} = \sigma_{XX} - (2\cdot\tfrac{3}{2} + 3\cdot\tfrac{1}{2}) = \sigma_{XX} - \tfrac{9}{2}\,, \quad \sigma_{XX} = \frac{11}{2}\,.$$

综上

$$\boldsymbol{\mu}=\begin{pmatrix}3/2\\1/2\\-1/2\end{pmatrix},\qquad \Sigma=\begin{pmatrix}11/2 & 3/2 & 1/2\\3/2 & 3/2 & -1/2\\1/2 & -1/2 & 1/2\end{pmatrix}.$$

(b) [2 分] 确定 $X\,|\,Y$ 的分布.

【解】

$$\mathbb{E}(X\,|Y)=\mu_X+\frac{\sigma_{XY}}{\sigma_{YY}}(Y-\mu_Y)=\frac{3}{2}+\frac{3/2}{3/2}\left(Y-\frac{1}{2}\right)=1+Y\,,$$

$$\mathrm{Var}(X\,|Y)=\sigma_{XX}-\frac{\sigma_{XY}^2}{\sigma_{YY}}=\frac{11}{2}-\frac{(3/2)^2}{3/2}=\frac{11}{2}-\frac{3}{2}=4\,.$$

故 $X\,|Y\sim N_1(1+Y\,,4)$.

(c) [2 分] 确定 $X\,|\,Y+Z$ 的分布.

【解】 令 $W=Y+Z$. 由联合正态的线性性，$(X,W)^{\mathrm{T}}$ 服从二元正态，且

$$\mathbb{E}(W)=\frac{1}{2}+\left(-\frac{1}{2}\right)=0\,,$$

$$\mathrm{Var}(W)=\sigma_{YY}+\sigma_{ZZ}+2\sigma_{YZ}=\frac{3}{2}+\frac{1}{2}-1=1\,,$$

$$\mathrm{Cov}(X,W)=\sigma_{XY}+\sigma_{XZ}=\frac{3}{2}+\frac{1}{2}=2\,.$$

故

$$\mathbb{E}(X\,|W)=\frac{3}{2}+\frac{2}{1}(W-0)=\frac{3}{2}+2(Y+Z)\,,$$

$$\mathrm{Var}(X\,|W)=\frac{11}{2}-\frac{2^2}{1}=\frac{3}{2}\,.$$

即 $X\,|Y+Z\sim N_1\left(\frac{3}{2}+2(Y+Z)\,,\frac{3}{2}\right)$.

4. 已知

$$Z\sim N_1(0\,,1) \tag{1.8}$$

$$Y\,|\,Z\sim N_1(1+Z\,,1) \tag{1.9}$$

$$(X\,|\,Y\,,Z)\sim N_1(1-Y\,,1) \tag{1.10}$$

(a) [2 分] 确定 $\begin{pmatrix}X\\Y\\Z\end{pmatrix}$ 的分布.

【解】 由条件分布的层级结构可知 $(X\,,Y\,,Z)^{\mathrm{T}}$ 服从三元正态分布. 利用全期望、全方差公式

计算：

$$\begin{aligned}
\mathbb{E}(Z) &= 0\,, \quad \mathrm{Var}(Z) = 1\,, \\
\mathbb{E}(Y) &= \mathbb{E}[\mathbb{E}(Y\,|Z\,)] = \mathbb{E}(1+Z) = 1\,, \\
\mathrm{Var}(Y) &= \mathbb{E}[\mathrm{Var}(Y\,|Z\,)] + \mathrm{Var}[\mathbb{E}(Y\,|Z\,)] = 1 + \mathrm{Var}(1+Z) = 1+1 = 2\,, \\
\mathrm{Cov}(Y,Z) &= \mathrm{Cov}[\mathbb{E}(Y\,|Z\,), Z] = \mathrm{Cov}(1+Z, Z) = 1\,, \\
\mathbb{E}(X) &= \mathbb{E}[\mathbb{E}(X\,|Y,Z\,)] = \mathbb{E}(1-Y) = 0\,, \\
\mathrm{Var}(X) &= 1 + \mathrm{Var}(1-Y) = 1+2 = 3\,, \\
\mathrm{Cov}(X,Y) &= \mathrm{Cov}(1-Y, Y) = -\mathrm{Var}(Y) = -2\,, \\
\mathrm{Cov}(X,Z) &= \mathrm{Cov}(1-Y, Z) = -\mathrm{Cov}(Y,Z) = -1\,.
\end{aligned}$$

故

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 & -2 & -1 \\ -2 & 2 & 1 \\ -1 & 1 & 1 \end{pmatrix}\right).$$

(b) [2 分] 确定 $(Y\,|\,X\,,\,Z\,)$ 的分布.

【解】 记 $\Sigma_{(X,Z)} = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$，$|\Sigma_{(X,Z)}| = 2$，$\Sigma_{(X,Z)}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$. 又 $\Sigma_{Y,(X,Z)} = (-2\,,\,1)$，故

$$\begin{aligned}
\Sigma_{Y,(X,Z)}\Sigma_{(X,Z)}^{-1} &= \frac{1}{2}(-2,1)\begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix} = \frac{1}{2}(-1\,,\,1) = \left(-\frac{1}{2}\,,\,\frac{1}{2}\right), \\
\mathbb{E}(Y\,|X,Z\,) &= 1 + \left(-\frac{1}{2}\,,\,\frac{1}{2}\right)\begin{pmatrix} X-0 \\ Z-0 \end{pmatrix} = 1 - \frac{X}{2} + \frac{Z}{2}\,, \\
\mathrm{Var}(Y\,|X,Z\,) &= 2 - \left(-\frac{1}{2}\,,\,\frac{1}{2}\right)\begin{pmatrix} -2 \\ 1 \end{pmatrix} = 2 - \frac{3}{2} = \frac{1}{2}\,.
\end{aligned}$$

故 $Y\,|X,Z \sim N_1\left(1 - \frac{X}{2} + \frac{Z}{2}\,,\,\frac{1}{2}\right)$.

(c) [2 分] 确定 $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1+Z \\ 1-Y \end{pmatrix}$ 的分布.

【解】 $(U,V)^{\mathrm{T}}$ 是 $(X,Y,Z)^{\mathrm{T}}$ 的仿射变换，故服从二元正态. 计算

$$\begin{aligned}
\mathbb{E}(U) &= 1+0=1\,, \quad \mathbb{E}(V)=1-1=0\,,\\
\mathrm{Var}(U) &= \mathrm{Var}(Z)=1\,, \quad \mathrm{Var}(V)=\mathrm{Var}(Y)=2\,,\\
\mathrm{Cov}(U,V) &= \mathrm{Cov}(Z,-Y)=-1\,.
\end{aligned}$$

故 $\begin{pmatrix} U \\ V \end{pmatrix} \sim N_2\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\right)$.

(d) [2 分] 计算 $\mathbb{E}\,(Y\,|\,U=2)$.

【解】 由 $U=1+Z$，$U=2$ 等价于 $Z=1$. 故

$$\mathbb{E}(Y\,|U=2)=\mathbb{E}(Y\,|Z=1)=1+Z\Big|_{Z=1}=2\,.$$

5. 已知

$$\boldsymbol{X} \sim N_3\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{pmatrix}\right) \tag{1.11}$$

(a) [2 分] 利用 X_1 与 X_2 的一个线性函数，求 X_3 的最佳线性逼近.

【解】 最佳线性逼近即条件期望 $\mathbb{E}(X_3\,|X_1,X_2)$. 记

$$\Sigma_{(1,2)}=\begin{pmatrix} 11 & -6 \\ -6 & 10 \end{pmatrix}, \quad |\Sigma_{(1,2)}|=74\,, \quad \Sigma_{(1,2)}^{-1}=\frac{1}{74}\begin{pmatrix} 10 & 6 \\ 6 & 11 \end{pmatrix}, \quad \Sigma_{3,(1,2)}=(2\,,\,-4)\,.$$

回归系数：

$$\Sigma_{3,(1,2)}\Sigma_{(1,2)}^{-1}=\frac{1}{74}(2,-4)\begin{pmatrix} 10 & 6 \\ 6 & 11 \end{pmatrix}=\frac{1}{74}(-4\,,\,-32)=\left(-\frac{2}{37}\,,\,-\frac{16}{37}\right)\,.$$

故

$$\hat{X}_3=3-\frac{2}{37}(X_1-1)-\frac{16}{37}(X_2-2)=\frac{145-2X_1-16X_2}{37}\,.$$

(b) [2 分] 计算 X_3 与 $(X_1\,,\,X_2)$ 的多重相关系数.

【解】 多重相关系数的平方为

$$R^2=\frac{\Sigma_{3,(1,2)}\Sigma_{(1,2)}^{-1}\Sigma_{(1,2),3}}{\sigma_{33}}=\frac{1}{6}\cdot\left(-\frac{2}{37}\,,\,-\frac{16}{37}\right)\begin{pmatrix} 2 \\ -4 \end{pmatrix}=\frac{1}{6}\cdot\frac{-4+64}{37}=\frac{60}{6\cdot 37}=\frac{10}{37}\,.$$

故 $R_{3\cdot 12}=\sqrt{10/37}\approx 0.5199$.

(c) [2 分] 令 $Z_1 = X_2 - X_3$，$Z_2 = X_2 + X_3$，如果 $(Z_3 \,|\, Z_1\,,\, Z_2) \sim N_1\,(Z_1 + Z_2\,,\,10)$，确定 $\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}$ 的分布.

【解】 由线性变换，$(Z_1, Z_2)^{\mathrm{T}}$ 服从二元正态. 计算

$$
\begin{aligned}
\mathbb{E}(Z_1) &= 2 - 3 = -1\,, \quad \mathbb{E}(Z_2) = 2 + 3 = 5\,,\\
\mathrm{Var}(Z_1) &= \sigma_{22} + \sigma_{33} - 2\sigma_{23} = 10 + 6 - 2(-4) = 24\,,\\
\mathrm{Var}(Z_2) &= \sigma_{22} + \sigma_{33} + 2\sigma_{23} = 10 + 6 + 2(-4) = 8\,,\\
\mathrm{Cov}(Z_1, Z_2) &= \sigma_{22} - \sigma_{33} = 10 - 6 = 4\,.
\end{aligned}
$$

又由 $Z_3 \,|Z_1, Z_2 \sim N_1(Z_1 + Z_2\,,\,10)$，可写 $Z_3 = Z_1 + Z_2 + \varepsilon$，其中 $\varepsilon \sim N(0, 10)$ 与 (Z_1, Z_2) 独立. 从而

$$
\begin{aligned}
\mathbb{E}(Z_3) &= -1 + 5 = 4\,,\\
\mathrm{Var}(Z_3) &= \mathrm{Var}(Z_1) + \mathrm{Var}(Z_2) + 2\mathrm{Cov}(Z_1, Z_2) + 10 = 24 + 8 + 8 + 10 = 50\,,\\
\mathrm{Cov}(Z_3, Z_1) &= \mathrm{Var}(Z_1) + \mathrm{Cov}(Z_1, Z_2) = 24 + 4 = 28\,,\\
\mathrm{Cov}(Z_3, Z_2) &= \mathrm{Cov}(Z_1, Z_2) + \mathrm{Var}(Z_2) = 4 + 8 = 12\,.
\end{aligned}
$$

故

$$
\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} -1 \\ 5 \\ 4 \end{pmatrix}, \begin{pmatrix} 24 & 4 & 28 \\ 4 & 8 & 12 \\ 28 & 12 & 50 \end{pmatrix} \right).
$$

6. 假设 $(X\,,\,Y\,,\,Z)^{\mathrm{T}}$ 服从三维正态分布，且

$$(Y \,|\, Z\,) \sim N_1(2Z\,,\,24) \tag{1.12}$$

$$(Z \,|\, X\,) \sim N_1(2X + 3\,,\,14) \tag{1.13}$$

$$X \sim N_1(1\,,\,4) \tag{1.14}$$

$$\rho_{_{XY}} = 0.5 \tag{1.15}$$

(a) [2 分] 确定 $(X\,,\,Y\,,\,Z)^{\mathrm{T}}$ 的分布.

【解】 由题设 $\mu_X = 1$，$\sigma_{XX} = 4$.

由 $Z\,|X \sim N_1(2X+3\,,\,14)$：

$$\mathbb{E}(Z) = 2\mu_X + 3 = 5\,,$$
$$\mathrm{Cov}(X,Z) = 2\sigma_{XX} = 8\,,$$
$$\mathrm{Var}(Z) = 14 + 4\sigma_{XX} = 14 + 16 = 30\,.$$

由 $Y\,|Z \sim N_1(2Z\,,\,24)$：

$$\mathbb{E}(Y) = 2\mathbb{E}(Z) = 10\,,$$
$$\mathrm{Cov}(Y,Z) = 2\sigma_{ZZ} = 60\,,$$
$$\mathrm{Var}(Y) = 24 + 4\sigma_{ZZ} = 24 + 120 = 144\,.$$

由 $\rho_{XY} = 0.5$：$\sigma_{XY} = 0.5\sqrt{4\times 144} = 12$. 综上

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N_3\left(\begin{pmatrix} 1 \\ 10 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 12 & 8 \\ 12 & 144 & 60 \\ 8 & 60 & 30 \end{pmatrix}\right).$$

(b) [2 分] 对于给定的 Z 值，计算 X 与 Y 的偏相关系数.

【解】 由公式

$$\rho_{XY\cdot Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{(1-\rho_{XZ}^2)(1-\rho_{YZ}^2)}}\,.$$

计算

$$\rho_{XZ} = \frac{8}{\sqrt{4\times 30}} = \frac{4}{\sqrt{30}}\,, \quad \rho_{XZ}^2 = \frac{8}{15}\,, \quad 1-\rho_{XZ}^2 = \frac{7}{15}\,,$$
$$\rho_{YZ} = \frac{60}{\sqrt{144\times 30}} = \frac{5}{\sqrt{30}}\,, \quad \rho_{YZ}^2 = \frac{5}{6}\,, \quad 1-\rho_{YZ}^2 = \frac{1}{6}\,,$$
$$\rho_{XZ}\rho_{YZ} = \frac{20}{30} = \frac{2}{3}\,, \quad \rho_{XY} - \rho_{XZ}\rho_{YZ} = \frac{1}{2} - \frac{2}{3} = -\frac{1}{6}\,.$$

故

$$\rho_{XY\cdot Z} = \frac{-1/6}{\sqrt{(7/15)(1/6)}} = -\frac{1}{6}\sqrt{\frac{90}{7}} = -\sqrt{\frac{5}{14}} = -\frac{\sqrt{70}}{14} \approx -0.5976\,.$$

(c) [2 分] 你认为利用 Y 和 Z 的一个线性函数逼近 X 是否合理?

【解】 由于 $(X,Y,Z)^{\mathrm{T}}$ 服从联合正态分布，X 关于 (Y,Z) 的条件期望 $\mathbb{E}(X\,|Y,Z)$ 本身就是 (Y,Z) 的线性函数，因此用线性函数逼近 X 在均方误差意义下是最优的. 计算 X 与 (Y,Z) 的多重相关系数：

$$\Sigma_{(Y,Z)} = \begin{pmatrix} 144 & 60 \\ 60 & 30 \end{pmatrix}, \quad |\Sigma_{(Y,Z)}| = 720\,, \quad \Sigma_{(Y,Z)}^{-1} = \frac{1}{720}\begin{pmatrix} 30 & -60 \\ -60 & 144 \end{pmatrix}.$$

$$\Sigma_{X,(Y,Z)}\Sigma_{(Y,Z)}^{-1} = \frac{1}{720}(12,8)\begin{pmatrix}30 & -60\\ -60 & 144\end{pmatrix} = \frac{1}{720}(-120\,,\,432) = \left(-\frac{1}{6}\,,\,\frac{3}{5}\right),$$

$$R^2_{X\cdot(Y,Z)} = \frac{1}{4}\left(-\frac{1}{6}\,,\,\frac{3}{5}\right)\begin{pmatrix}12\\ 8\end{pmatrix} = \frac{1}{4}\left(-2+\frac{24}{5}\right) = \frac{14/5}{4} = \frac{7}{10}\,.$$

$R \approx 0.837$，说明 (Y,Z) 能解释 X 约 70% 的方差，线性逼近

$$\hat{X} = 1 - \frac{1}{6}(Y-10) + \frac{3}{5}(Z-5) = -\frac{1}{3} - \frac{Y}{6} + \frac{3Z}{5}$$

的效果较好，合理.

7. 设

$$\boldsymbol{X} \sim N_4\left(\begin{pmatrix}1\\ 2\\ 3\\ 4\end{pmatrix},\ \begin{pmatrix}4 & 1 & 2 & 4\\ 1 & 4 & 2 & 1\\ 2 & 2 & 16 & 1\\ 4 & 1 & 1 & 9\end{pmatrix}\right) \tag{1.16}$$

(a) [2 分] 给出用 $(X_1\,,\,X_4)$ 的一个函数对 X_2 的最佳线性逼近，并解释逼近的效果.

【解】

$$\Sigma_{(1,4)} = \begin{pmatrix}4 & 4\\ 4 & 9\end{pmatrix},\quad |\Sigma_{(1,4)}| = 20\,,\quad \Sigma_{(1,4)}^{-1} = \frac{1}{20}\begin{pmatrix}9 & -4\\ -4 & 4\end{pmatrix},\quad \Sigma_{2,(1,4)} = (1,1)\,.$$

回归系数：

$$\Sigma_{2,(1,4)}\Sigma_{(1,4)}^{-1} = \frac{1}{20}(1,1)\begin{pmatrix}9 & -4\\ -4 & 4\end{pmatrix} = \frac{1}{20}(5\,,\,0) = \left(\frac{1}{4}\,,\,0\right).$$

故最佳线性逼近为

$$\hat{X}_2 = 2 + \frac{1}{4}(X_1-1) + 0\cdot(X_4-4) = \frac{7}{4} + \frac{1}{4}X_1\,.$$

即有效变量只有 X_1. 多重相关系数的平方为

$$R^2 = \frac{1}{\sigma_{22}}\left(\frac{1}{4}\,,\,0\right)\begin{pmatrix}1\\ 1\end{pmatrix} = \frac{1/4}{4} = \frac{1}{16} = 0.0625\,,\quad R \approx 0.25\,.$$

条件方差为 $\mathrm{Var}(X_2\,|X_1,X_4) = 4 - 1/4 = 15/4 = 3.75$. 相对 $\sigma_{22} = 4$ 仅下降 6.25%，逼近效果较差，说明 (X_1,X_4) 对 X_2 的解释能力很弱.

(b) [2 分] 给出用 $(X_1\,,\,X_3\,,\,X_4)$ 的一个函数对 X_2 的最佳线性逼近，与 (a) 的结果进行对比.

【解】

$$\Sigma_{(1,3,4)} = \begin{pmatrix}4 & 2 & 4\\ 2 & 16 & 1\\ 4 & 1 & 9\end{pmatrix},\quad \Sigma_{2,(1,3,4)} = (1,2,1)\,.$$

设 $\boldsymbol{v}=(v_1,v_2,v_3)^{\mathrm{T}}=\Sigma_{(1,3,4)}^{-1}(1,2,1)^{\mathrm{T}}$，即求解

$$\begin{cases} 4v_1+2v_2+4v_3=1 \\ 2v_1+16v_2+v_3=2 \\ 4v_1+v_2+9v_3=1 \end{cases}$$

计算 $|\Sigma_{(1,3,4)}|=4(144-1)-2(18-4)+4(2-64)=572-28-248=296$. 由 Cramer 法则：

$$v_1=\frac{1}{296}\det\begin{pmatrix}1&2&4\\2&16&1\\1&1&9\end{pmatrix}=\frac{143-34-56}{296}=\frac{53}{296}\,,$$

$$v_2=\frac{1}{296}\det\begin{pmatrix}4&1&4\\2&2&1\\4&1&9\end{pmatrix}=\frac{68-14-24}{296}=\frac{30}{296}=\frac{15}{148}\,,$$

$$v_3=\frac{1}{296}\det\begin{pmatrix}4&2&1\\2&16&2\\4&1&1\end{pmatrix}=\frac{56+12-62}{296}=\frac{6}{296}=\frac{3}{148}\,.$$

故最佳线性逼近为

$$\hat{X}_2=2+\frac{53}{296}(X_1-1)+\frac{30}{296}(X_3-3)+\frac{6}{296}(X_4-4)=\frac{425+53X_1+30X_3+6X_4}{296}\,.$$

多重相关系数的平方为

$$R^2=\frac{1}{4}(1,2,1)\boldsymbol{v}=\frac{1}{4}\cdot\frac{53+60+6}{296}=\frac{119}{1184}\approx 0.1005\,,\quad R\approx 0.317\,.$$

条件方差为 $\mathrm{Var}(X_2\,|X_1,X_3,X_4)=4-119/296=1065/296\approx 3.598$.

对比： 相比 (a)，加入 X_3 后 R^2 从 $1/16\approx 0.0625$ 提升至 ≈ 0.1005，解释能力有所改进，但整体 R^2 仍较小，说明 (X_1,X_3,X_4) 联合对 X_2 的线性逼近效果仍然较差，X_2 大部分变异无法由其余变量的线性组合解释.