Solution of the Exercise

Data.

$$X = \begin{bmatrix} 126 & 78 \\ 128 & 80 \\ 128 & 82 \\ 130 & 82 \\ 130 & 84 \\ 132 & 86 \end{bmatrix} \qquad (n = 6,\ \text{systolic} = \text{col}_1,\ \text{diastolic} = \text{col}_2)$$

(a) Center the data. Column means:

$$\bar{x}_{\text{sys}} = \frac{126 + 128 + 128 + 130 + 130 + 132}{6} = 129, \qquad \bar{x}_{\text{dia}} = \frac{78 + 80 + 82 + 82 + 84 + 86}{6} = 82.$$

Centered matrix $X_c = X - \mathbf{1}\bar{x}^\top$:

$$X_c = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

(b) Sample covariance matrix.

$$C = \frac{1}{n-1} X_c^\top X_c = \frac{1}{5} \begin{bmatrix} 22 & 28 \\ 28 & 40 \end{bmatrix} = \begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix}.$$

So Var(systolic) $= 4.4$, Var(diastolic) $= 8.0$, Cov $= 5.6$. Correlation:

$$\rho = \frac{5.6}{\sqrt{4.4 \cdot 8.0}} \approx 0.944 \quad \text{(strong positive)}.$$

(c) Eigenvalues and eigenvectors of C. Solve $\det(C - \lambda I) = 0$. For $C = \begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix}$:

$$\lambda_{1,2} = \frac{\operatorname{tr} C \pm \sqrt{(\operatorname{tr} C)^2 - 4 \det C}}{2} = \frac{12.4 \pm \sqrt{12.4^2 - 4 \cdot 3.84}}{2} = \frac{12.4 \pm \sqrt{138.4}}{2} \approx \{\, 12.0822,\ 0.3178 \,\}.$$

(Checks: $\lambda_1 + \lambda_2 = \operatorname{tr} C = 12.4$, $\lambda_1 \lambda_2 = \det C = 3.84$.)

Exact form: multiplying C by 5 gives $\tilde{C} = \begin{bmatrix} 22 & 28 \\ 28 & 40 \end{bmatrix}$ with eigenvalues $\mu = 31 \pm \sqrt{865}$, hence

$$\lambda = \frac{\mu}{5} = \frac{31 \pm \sqrt{865}}{5}.$$

Unit eigenvectors (columns of V):

$$v_1 \approx \begin{bmatrix} 0.5891 \\ 0.8081 \end{bmatrix} \quad (\text{for } \lambda_1 \approx 12.0822), \qquad v_2 \approx \begin{bmatrix} -0.8081 \\ 0.5891 \end{bmatrix} \quad (\text{for } \lambda_2 \approx 0.3178).$$

(Any overall sign is acceptable.)

(d) Interpretation (PCA view).

- **Variance explained.** Total variance $= \operatorname{tr} C = 12.4$.

$$\text{Explained by PC1} = \frac{12.0822}{12.4} \approx 97.44\%, \qquad \text{PC2} \approx 2.56\%.$$

  So the data are essentially one–dimensional.

- **Principal direction.** $v_1 \propto (0.589, 0.808)$ has *both* positive entries: the main variation is that systolic and diastolic pressures increase together (consistent with $\rho \approx 0.94$). Diastolic has slightly higher loading (0.808).

- **Secondary direction.** v_2 has opposite signs: tiny residual trade-off pattern where one increases while the other decreases ($\approx 2.6\%$ variance).

- **Methodology.** Center $\to$ covariance $\to$ eigendecompose. The largest eigenpair (λ_1, v_1) gives the dominant pattern and its variance.