Multiple choice problems

1. $\operatorname{Cov}\left(\frac{1}{2}\underline{X}_1 + \frac{1}{2}\underline{X}_2 + \frac{1}{2}\underline{X}_3 + \frac{1}{2}\underline{X}_4,\ \underline{X}_1 + \underline{X}_2 + \underline{X}_3 - 3\underline{X}_4\right)$

$= 3 \cdot \frac{1}{2} \operatorname{Cov} \underline{X}_1 - \frac{3}{2} \operatorname{Cov} \underline{X}_4 = 0$

We have used the rules $\operatorname{Cov}(\underline{X} + \underline{Y}, \underline{Z}) = \operatorname{Cov}(\underline{X}, \underline{Z}) + \operatorname{Cov}(\underline{Y}, \underline{Z})$ and $\operatorname{Cov}(A\underline{X}, B\underline{Y}) = A \operatorname{Cov}(\underline{X}, \underline{Y}) B^T$ (A and B are numbers here).

(Remember that the $\underline{X}_i$ are independent, so $\operatorname{Cov}(\underline{X}_i, \underline{X}_j) = 0$ for $i \neq j$.)

2. Let Σ = covariance matrix.

$$\det(\lambda I - \Sigma) = \begin{vmatrix} \lambda-1 & 2 & 0 \\ 2 & \lambda-5 & 0 \\ 0 & 0 & \lambda-2 \end{vmatrix} = (\lambda-2)(\lambda^2 - 6\lambda + 1)$$

$$\lambda = 2 \quad \text{el.} \quad \lambda = \frac{6 \pm \sqrt{36-4}}{2} = \frac{6 \pm 4\sqrt{2}}{2} = 3 \pm 2\sqrt{2}$$

Second eigenvalue: 2

Solve $(2I - \Sigma)\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \underline{0}$:

$$\begin{aligned} x + 2y &= 0 \\ 2x - 3y &= 0 \end{aligned} \qquad x = y = 0$$

Unit eigenvector = $PC2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (or $\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$).

3. $$\begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\operatorname{Cov} \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11}-\sigma_{12} & \sigma_{12}-\sigma_{13} \\ \sigma_{12}-\sigma_{22} & \sigma_{22}-\sigma_{23} \\ \sigma_{13}-\sigma_{23} & \sigma_{23}-\sigma_{33} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11}-2\sigma_{12}+\sigma_{22} & \sigma_{12}-\sigma_{13}-\sigma_{22}+\sigma_{23} \\ \sigma_{12}-\sigma_{13}-\sigma_{22}+\sigma_{23} & \sigma_{22}-2\sigma_{23}+\sigma_{33} \end{bmatrix}$$

4. Easiest: Trace formula:

$$E(\underline{X}^T A \underline{X}) = \operatorname{tr}(A \underbrace{\operatorname{Cov} \underline{X}}_{=I}) + (E\underline{X})^T A \underbrace{(E\underline{X})}_{=0} = \operatorname{tr} A = -1+1+2 = 2$$

5. $R^2_{adj} = 1 - \frac{n-1}{n-p}\frac{SSE}{SST} = 1 - \frac{n-1}{n-p}(1-R^2)$, $n=10$, $p=2$

$$1-R^2 = \frac{n-p}{n-1}(1-R^2_{adj})$$

$$R^2 = 1 - \frac{n-p}{n-1}(1-R^2_{adj}) = 1 - \frac{8}{9}(1-0{,}6777) = 0.7135$$

6. $\binom{10}{2}$ tests, significance level $0{,}05/\binom{10}{2} = 0.0011$.

7. Perhaps easiest: Effect coding like in two-level factorial design, with design matrix $\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$.

In general: $\hat{\beta}_j = \frac{1}{n}\left(\sum_{x_{ij}=1} y_i - \sum_{x_{ij}=-1} y_i\right)$.

$\hat{\beta}_0 = \frac{1}{4}(-1+0+3+5) = \frac{7}{4}$, $\hat{\beta}_1 = \frac{1}{4}(1-0+3+5) = \frac{9}{4}$.

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \frac{7}{4} + \frac{9}{4}$, $-\frac{1}{2}$ $(x=-1)$ or 4 $(x=1)$.

$$SSE = (-1+\tfrac{1}{2})^2 + (0+\tfrac{1}{2})^2 + (3-4)^2 + (5-4)^2$$
$$= 0{,}25 + 0{,}25 + 1 + 1 = 2{,}5$$

8. $1 = ABCE$ (first generator)

$= BCDF = ADEF$ (second generator and product with first)

Resolution: IV

Written answer problems

1. $$\begin{bmatrix} \underline{X} \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \underline{\mu}_x \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_x & \underline{\sigma}_{xY} \\ \underline{\sigma}_{xY}^T & \sigma_Y^2 \end{bmatrix} \right)$$

$$Y \mid \underline{X} = \underline{x} \sim N(\mu_Y + \underline{\sigma}_{xY}^T \Sigma_x^{-1}(\underline{x} - \underline{\mu}_x),\ \sigma_Y^2 - \underline{\sigma}_{xY}^T \Sigma_x^{-1} \underline{\sigma}_{xY})$$

So $$E(Y \mid \underline{X} = \underline{x}) = \mu_Y + \underline{\sigma}_{xY}^T \Sigma_x^{-1}(\underline{x} - \underline{\mu}_x)$$

$$= \underbrace{\mu_Y - \underline{\sigma}_{xY}^T \Sigma_x^{-1} \underline{\mu}_x}_{\text{intercept}} + \underbrace{(\underline{\sigma}_{xY}^T \Sigma_x^{-1})}_{1 \times k} \underbrace{\underline{x}}_{k \times 1},$$

which is the expected value in a mult. lin. regr. model having intercept $\mu_Y - \underline{\sigma}_{xY}^T \Sigma_x^{-1} \underline{\mu}_x$ and other coefficients $\underline{\sigma}_{xY}^T \Sigma_x^{-1}$.

2a. Residual plot and Q-Q plot OK.

$$SSE = (105-8) \cdot 0.2579^2 = 6.452$$

A tennis player will on average have $0.27 \cdot 10^{12}$ more blood cells than a basketball player of the same

sex, height and weight.

b. $H_0: \beta_{gym} = \beta_{scu} = \beta_{sml} = \beta_{ten} = 0$ (coefficients of levels of sport)

H_1: at least one $\neq 0$.

In $C\beta = \underline{0}$ notation: $r = \text{rank } C = 4$ ($\beta_{gym} = 0, \ldots$).

SSE of reduced model:

$$SSE_0 = (105-4) \cdot 0.2696^2 = 7.3411$$

$$F = \frac{(SSE_0 - SSE)/r}{SSE/(n-p)} = \frac{(7.3411 - 6.452)/4}{6.452/(105-8)} = 3.3443$$

$F \sim F_{4,\, 105-8}$ if H_0 is true. Reject for large values.

$P(F \geq 3.3443) < 0.05$ since $P(F \geq 2.465) = 0.05$. Reject H_0 (keep original model).

From leaps package output: $C_p = 8$ for full model, $C_p = 13$ for reduced model. Supports to prefer full model.