Solution Exercise 6.3

1. We must test whether the mean values differ significantly from each other. We find Mahalanobis' distance (section 5.1.3)

$$\begin{aligned} D^2 &= [-20\ 30]\begin{bmatrix} 25 & 40 \\ 40 & 100 \end{bmatrix}^{-1}\begin{bmatrix} -20 \\ 30 \end{bmatrix} \\ &= \tfrac{1}{900}[-20\ 30]\begin{bmatrix} 100 & -40 \\ -40 & 25 \end{bmatrix}\begin{bmatrix} -20 \\ 30 \end{bmatrix} \\ &= 122.78 \end{aligned}$$

The test statistic becomes (T. 5.12)

$$\frac{106+81-2-1}{2(106+81-2)} \times \frac{106\times 81}{106+81} \times 122.78 = 2803$$

Since this value is much larger than

$$F(2{,}184)_{0.999} = 7.17$$

we strongly reject the hypothesis that the two mean values are equal.

2. Using theorems 5.2 and 5.8 we must determine c so that

$$P\left\{\frac{f_1(\boldsymbol{X})}{f_2(\boldsymbol{X})} < c|\pi_1\right\} = P\left\{\frac{f_1(\boldsymbol{X})}{f_2(\boldsymbol{X})} \geq c|\pi_2\right\}$$

or

$$P\left\{N\left(\frac{1}{2}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2, \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2\right) < \log c\right\} = P\left\{N\left(-\frac{1}{2}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2, \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2\right) \geq \log c\right\}$$

which – when we assume that the true parameters in the two distributions are equal to the estimated values – is equivalent to

$$P\{N(61.39, 122.78) < \log c\} = P\{N(-61.39, 122.78) \geq \log c\}$$

or

$$P\left\{N(0,1) < \frac{\log c - 61.39}{\sqrt{122.78}}\right\} = P\left\{N(0,1) \geq \frac{\log c + 61.39}{\sqrt{122.78}}\right\}$$

i.e.

$$P\left\{N(0,1) < \frac{\log c - 61.39}{11.08}\right\} = P\left\{N(0,1) \geq \frac{\log c + 61.39}{11.08}\right\}$$

From this it is easily seen that $\log c$ must be equal to 0. (This could also be seen from the previous equations, but is probably more straightforward from this one). Therefore the decision region for population 1 becomes

$$R_1 = \{\boldsymbol{x}|\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \tfrac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \tfrac{1}{2}\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 \geq 0\}$$

$$= \{\boldsymbol{x} | (x_1 \;\; x_2)^{\boldsymbol{T}} \begin{bmatrix} \frac{1}{9} & -\frac{2}{45} \\ -\frac{2}{45} & \frac{1}{36} \end{bmatrix} \begin{bmatrix} -20 \\ 30 \end{bmatrix} - \frac{1}{2}[380 \;\; 120] \begin{bmatrix} \frac{1}{9} & -\frac{2}{45} \\ -\frac{2}{45} & \frac{1}{36} \end{bmatrix} \begin{bmatrix} 380 \\ 120 \end{bmatrix} + \frac{1}{2}[400 \;\; 90] \begin{bmatrix} \frac{1}{9} & -\frac{2}{45} \\ -\frac{2}{45} & \frac{1}{36} \end{bmatrix} \begin{bmatrix} 400 \\ 90 \end{bmatrix} \geq 0\}$$

$$= \{\boldsymbol{x} | (x_1 \;\; x_2)^{\boldsymbol{T}} \begin{bmatrix} -\frac{32}{9} \\ \frac{31}{18} \end{bmatrix} - 6195.55 + 7401.39 \geq 0\}$$

$$=\{\boldsymbol{x} | -3.56x_1 + 1.72x_2 + 1205.84 \geq 0\}$$

3. Inserting $x_1 = 390$ and $x_2 = 100$ gives

$$-3.56 \times 390 + 1.72 \times 100 + 1205.84 = -10.56 < 0$$

Therefore the observation is not included in R_1 and we choose the second population, i.e. the patient has disease B.

■