

Classification with Two Multivariate Normal Populations

- Assume $f_1(x)$ and $f_2(x)$ are multivariate normal densities, with mean-covariance as (μ_1, Σ_1) and (μ_2, Σ_2) , respectively, and $\Sigma_1 = \Sigma_2 = \Sigma$.
- The allocation rule that minimizes the ECM is: we allocate x_0 to π_1 if and only if

$$(\mu_1 - \mu_2)^T \Sigma^{-1} x_0 - \frac{1}{2} (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2) \geq \ln \left[\frac{c(1|2)}{c(2|1)} \left(\frac{p_2}{p_1} \right) \right]$$

109

Unknown Mean and Covariance

- When population quantities μ_1 , μ_2 , and Σ are unknown, population parameters can be replaced with their sample counterparts.
- Suppose we have n_1 (p -dimension) observations from π_1 and n_2 observations from π_2 , with $n_1 + n_2 - 2 \geq p$. Allocate x_0 to π_1 if and only if

$$(\bar{x}_1 - \bar{x}_2)^T S_{pooled}^{-1} x_0 - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T S_{pooled}^{-1} (\bar{x}_1 + \bar{x}_2) \geq \ln \left[\frac{c(1|2)}{c(2|1)} \left(\frac{p_2}{p_1} \right) \right]$$

110

Example 11.3

In a study of hemophilia A carriers, blood samples were collected for two groups of women and measurements on two variables $X_1 = \log_{10}(\text{AHF activity})$ and $X_2 = \log_{10}(\text{AHF-like antigen})$. The first group of $n_1 = 30$ women were selected from a population of women who did not carry the hemophilia gene. The second group of $n_2 = 22$ women was selected from known hemophilia A carriers. Bivariate normal distribution seem to fit the data fairly well.

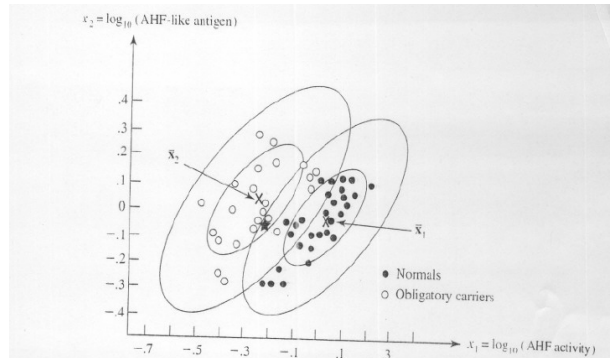


Figure 11.4 Scatter plots of $[\log_{10}(\text{AHF activity}), \log_{10}(\text{AHF-like antigen})]$ for the normal group and obligatory hemophilia A carriers.