

Chapter 7

Normal distributions

In this chapter it will be shown how to fit normal distributions simultaneously to the T cells of a multifactor design. Under equality of variances a multifactor model is discussed to explain the influence of the factors of the multifactor design. An application of a single factor model is presented to illustrate the theory.

7.1 Estimation of distributions

To fit normal distributions simultaneously to the T cells of any multifactor design it is required that

$$\Phi(\mathbf{z}) = \pi \quad (7.1)$$

where

$$\mathbf{z} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_T \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{x} - \mu_1 \mathbf{1}}{\sigma_1} \\ \frac{\mathbf{x} - \mu_2 \mathbf{1}}{\sigma_2} \\ \vdots \\ \frac{\mathbf{x} - \mu_T \mathbf{1}}{\sigma_T} \end{pmatrix} \quad (7.2)$$

is the concatenated vector of standardised upper class boundaries and

$$\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \vdots \\ \boldsymbol{\pi}_T \end{pmatrix} \quad (7.3)$$

is the concatenated vector of cumulative relative frequencies.

Taking the inverse normal function from (7.1) leads to the linear model

$$\begin{aligned} \Phi^{-1}(\boldsymbol{\pi}) &= \begin{pmatrix} \Phi^{-1}(\boldsymbol{\pi}_1) \\ \Phi^{-1}(\boldsymbol{\pi}_2) \\ \vdots \\ \Phi^{-1}(\boldsymbol{\pi}_T) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\mathbf{x} - \mu_1 \mathbf{1}}{\sigma_1} \\ \frac{\mathbf{x} - \mu_2 \mathbf{1}}{\sigma_2} \\ \vdots \\ \frac{\mathbf{x} - \mu_T \mathbf{1}}{\sigma_T} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{X}\boldsymbol{\alpha}_1 \\ \mathbf{X}\boldsymbol{\alpha}_2 \\ \vdots \\ \mathbf{X}\boldsymbol{\alpha}_T \end{pmatrix} = (\mathbf{I}_T \otimes \mathbf{X}) \boldsymbol{\alpha} \end{aligned} \quad (7.4)$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix} \quad (7.5)$$

is the design matrix for normality within each cell and

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_T \end{pmatrix} \quad (7.6)$$

is the concatenated vector of natural parameters with

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_t} \\ \frac{\mu_t}{\sigma_t} \end{pmatrix} \quad t = 1 \dots T \quad (7.7)$$

the natural parameters for the t -th cell.

From (7.4) the vector of constraints for normality, $\mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) = \mathbf{0}$, follows where

$$\begin{aligned} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) &= \begin{pmatrix} \mathbf{Q}_X \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_1) \\ \mathbf{Q}_X \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_2) \\ \vdots \\ \mathbf{Q}_X \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_T) \end{pmatrix} \\ &= (\mathbf{I}_T \otimes \mathbf{Q}_X) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \end{aligned} \quad (7.8)$$

and

$$\begin{aligned} \mathbf{G}_{\text{nor}}(\boldsymbol{\pi}) &= \frac{\partial \mathbf{g}_{\text{nor}}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= \frac{\partial}{\partial \boldsymbol{\pi}} \{ (\mathbf{I}_T \otimes \mathbf{Q}_X) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \} \\ &= (\mathbf{I}_T \otimes \mathbf{Q}_X) \cdot \mathbf{D}_{\boldsymbol{\pi}} \end{aligned} \quad (7.9)$$

with $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ the projection matrix orthogonal to \mathbf{X} and $\mathbf{D}_{\boldsymbol{\pi}} = \frac{\partial \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}$.

To solve $\mathbf{D}_{\boldsymbol{\pi}} = \frac{\partial \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}$ set $\boldsymbol{\nu} = \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})$ then $\boldsymbol{\Phi}(\boldsymbol{\nu}) = \boldsymbol{\pi}$ and hence

$$\begin{aligned} \mathbf{D}_{\boldsymbol{\pi}} &= \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\pi}} \\ &= \left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\nu}} \right)^{-1} \\ &= \left(\frac{\partial \boldsymbol{\Phi}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right)^{-1} \\ &= (\text{diag}[\boldsymbol{\phi}(\boldsymbol{\nu})])^{-1} \\ &= (\text{diag}[\boldsymbol{\phi}(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}))])^{-1} . \end{aligned} \quad (7.10)$$

Employing the maximum likelihood procedure in Proposition 1 with vector of constraints

$$\begin{aligned}\mathbf{g}(\boldsymbol{\pi}) &= \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) \\ &= (\mathbf{I}_T \otimes \mathbf{Q}_X) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})\end{aligned}\quad (7.11)$$

and matrix of partial derivatives

$$\begin{aligned}\mathbf{G}_\pi &= \mathbf{G}_{\text{nor}}(\boldsymbol{\pi}) \\ &= (\mathbf{I}_T \otimes \mathbf{Q}_X) \cdot \mathbf{D}_\pi\end{aligned}\quad (7.12)$$

the restricted ML estimate $\hat{\boldsymbol{\pi}}$ follows, with asymptotic covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\pi}}) \cong \mathbf{V} - (\mathbf{G}_\pi \mathbf{V})' (\mathbf{G}_\pi \mathbf{V} \mathbf{G}_\pi')^* (\mathbf{G}_\pi \mathbf{V}) .$$

For each of the T subpopulations, the vector of restricted cumulative relative frequencies $\hat{\boldsymbol{\pi}}_t$ for $t = 1, 2, \dots, T$ follow a cumulative normal distribution curve at the upper class boundaries of \mathbf{x} . Each $\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\pi}}_t)$ for $t = 1, 2, \dots, T$ is a linear combination of the columns of \mathbf{X} characterising a specific fitted normal distribution with its own set of parameter estimates.

The ML estimate of $\boldsymbol{\alpha}$ follows from (7.4)

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \cdot \boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\pi}})\quad (7.13)$$

which consists of two sets of estimators namely

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_1 &= \begin{pmatrix} \hat{\alpha}_{11} \\ \hat{\alpha}_{12} \\ \vdots \\ \hat{\alpha}_{1T} \end{pmatrix} = \begin{pmatrix} 1/\hat{\sigma}_1 \\ 1/\hat{\sigma}_2 \\ \vdots \\ 1/\hat{\sigma}_T \end{pmatrix} \\ &= \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \cdot \boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\pi}})\end{aligned}\quad (7.14)$$

and

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_2 &= \begin{pmatrix} \hat{\alpha}_{21} \\ \hat{\alpha}_{22} \\ \vdots \\ \hat{\alpha}_{2T} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_1/\hat{\sigma}_1 \\ \hat{\mu}_2/\hat{\sigma}_2 \\ \vdots \\ \hat{\mu}_T/\hat{\sigma}_T \end{pmatrix} \\ &= \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \cdot \boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\pi}}) .\end{aligned}\quad (7.15)$$

Note: In (7.14) $[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']_1$ is the first row of the matrix $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and in (7.15) $[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']_2$ is the second row of the matrix $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

It follows that

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\alpha}}) &\approx \left(\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\pi}}\right) \text{Cov}(\hat{\boldsymbol{\pi}}) \left(\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\pi}}\right)' \\ &= \left\{ \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \mathbf{D}_\pi \right\} \text{Cov}(\hat{\boldsymbol{\pi}}) \left\{ \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \mathbf{D}_\pi \right\}' . \end{aligned} \quad (7.16)$$

The ML estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are obtained from

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_T \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_{21}/\hat{\alpha}_{11} \\ \hat{\alpha}_{22}/\hat{\alpha}_{12} \\ \vdots \\ \hat{\alpha}_{2T}/\hat{\alpha}_{1T} \end{pmatrix} = \frac{\hat{\boldsymbol{\alpha}}_2}{\hat{\boldsymbol{\alpha}}_1} \quad (7.17)$$

and

$$\hat{\boldsymbol{\sigma}} = \begin{pmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \vdots \\ \hat{\sigma}_T \end{pmatrix} = \begin{pmatrix} 1/\hat{\alpha}_{11} \\ 1/\hat{\alpha}_{12} \\ \vdots \\ 1/\hat{\alpha}_{1T} \end{pmatrix} = \frac{1}{\hat{\boldsymbol{\alpha}}_1} . \quad (7.18)$$

Note: An element wise division for $\frac{\hat{\boldsymbol{\alpha}}_2}{\hat{\boldsymbol{\alpha}}_1}$ and $\frac{1}{\hat{\boldsymbol{\alpha}}_1}$ are understood in (7.17) and (7.18).

Let

$$\begin{aligned} \boldsymbol{\beta} &= \left(\boldsymbol{\mu} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) + \left(\boldsymbol{\sigma} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \\ &= \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{pmatrix} \end{aligned} \quad (7.19)$$

be the concatenated vector of original parameters with

$$\boldsymbol{\beta}_t = \begin{pmatrix} \mu_t \\ \sigma_t \end{pmatrix} = \begin{pmatrix} \alpha_{2t}/\alpha_{1t} \\ 1/\alpha_{1t} \end{pmatrix} \quad t = 1, 2, \dots, T. \quad (7.20)$$

Hence

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}}) &\approx \left(\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}} \right) \text{Cov}(\hat{\boldsymbol{\alpha}}) \left(\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}} \right)' \\ &= \mathbf{B} \text{Cov}(\hat{\boldsymbol{\alpha}}) \mathbf{B}' \end{aligned} \quad (7.21)$$

where

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{B}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}_T \end{pmatrix} \quad (7.22)$$

with

$$\begin{aligned} \mathbf{B}_t &= \left(\frac{\partial \boldsymbol{\beta}_t}{\partial \boldsymbol{\alpha}_t} \right) \\ &= \begin{pmatrix} -\frac{\alpha_{2t}}{\alpha_{1t}^2} & \frac{1}{\alpha_{1t}} \\ \frac{1}{\alpha_{1t}^2} & 0 \end{pmatrix} \quad t = 1, 2, \dots, T \end{aligned} \quad (7.23)$$

the partial derivatives for the t -th cell.

In terms of Kronecker products the matrix \mathbf{B} in (7.22) can be calculated from

$$\mathbf{B} = \begin{pmatrix} -\frac{\alpha_2}{\alpha_1^2} \\ \frac{1}{\alpha_1^2} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{\alpha_1} \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{\alpha_1^2} \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (7.24)$$

Consequently it follows that the asymptotic covariance matrices for $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\sigma}}$ are

$$\text{Cov}(\hat{\boldsymbol{\mu}}) \approx \mathbf{B}_\mu \text{Cov}(\hat{\boldsymbol{\alpha}}) \mathbf{B}_\mu' \quad (7.25)$$

where

$$\mathbf{B}_\mu = \begin{pmatrix} \frac{\alpha_2}{\alpha_1^2} \\ \frac{1}{\alpha_1^2} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{\alpha_1} \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \quad (7.26)$$

and

$$\text{Cov}(\hat{\sigma}) \cong \mathbf{B}_\sigma \text{Cov}(\hat{\alpha})\mathbf{B}'_\sigma \quad (7.27)$$

where

$$\mathbf{B}_\sigma = \left(-\frac{\mathbf{1}}{\alpha_1^2} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right) . \quad (7.28)$$

7.2 Equality of variances

Equality of variances

$$\begin{pmatrix} \sigma_1 - \sigma_2 \\ \sigma_1 - \sigma_3 \\ \vdots \\ \sigma_1 - \sigma_T \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (7.29)$$

is expressed in terms of matrix notation as

$$\mathbf{H}\alpha_1 = \mathbf{0} \quad (7.30)$$

where

$$\begin{aligned} \mathbf{H} &= \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{1}_{(T-1)} & -\mathbf{I}_{(T-1)} \end{pmatrix} \end{aligned} \quad (7.31)$$

is a matrix of contrasts and

$$\begin{aligned} \alpha_1 &= \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1T} \end{pmatrix} = \begin{pmatrix} \sigma_1^{-1} \\ \sigma_2^{-1} \\ \vdots \\ \sigma_T^{-1} \end{pmatrix} \\ &= \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \cdot \Phi^{-1}(\boldsymbol{\pi}) \end{aligned} \quad (7.32)$$

is a subset of the vector of natural parameters α formulated in (7.6) and (7.7).

Hence, the vector of constraints for equality of variances is $\mathbf{g}_{\text{var}}(\boldsymbol{\pi}) = \mathbf{0}$, with

$$\mathbf{g}_{\text{var}}(\boldsymbol{\pi}) = \mathbf{H} \cdot \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \quad (7.33)$$

and matrix of partial derivatives

$$\begin{aligned} \mathbf{G}_{\text{var}}(\boldsymbol{\pi}) &= \frac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= \mathbf{H} \cdot \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \cdot \mathbf{D}_{\boldsymbol{\pi}} . \end{aligned} \quad (7.34)$$

($\mathbf{D}_{\boldsymbol{\pi}}$ previously derived in (7.10).)

The restricted ML estimate of $\boldsymbol{\pi}$ follows by implementing

$$\begin{aligned} \mathbf{g}(\boldsymbol{\pi}) &= \begin{pmatrix} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{var}}(\boldsymbol{\pi}) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_T \otimes \mathbf{Q}_X \\ \mathbf{H} \cdot \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \end{pmatrix} \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \end{aligned} \quad (7.35)$$

and

$$\begin{aligned} \mathbf{G}_{\boldsymbol{\pi}} &= \begin{pmatrix} \mathbf{G}_{\text{nor}}(\boldsymbol{\pi}) \\ \mathbf{G}_{\text{var}}(\boldsymbol{\pi}) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_T \otimes \mathbf{Q}_X \\ \mathbf{H} \cdot \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \end{pmatrix} \cdot \mathbf{D}_{\boldsymbol{\pi}} \end{aligned} \quad (7.36)$$

in the ML estimation procedure.

The restricted ML estimate $\hat{\boldsymbol{\pi}}$ is now estimated such that:

1. $\hat{\boldsymbol{\pi}}_t$, ($t = 1, 2, \dots, T$) follows a cumulative normal distribution curve at the upper boundaries of \mathbf{x} and
2. the fitted normal distributions have equal variances over the T cells.

7.3 Multifactor model

To explain the effect of the factors on the grouped response variable, a linear model may be formulated on the cells of the multifactor design. Since a normal distribution is fitted to each cell, the mean μ , of the fitted normal distribution will be used as a representative measure for each cell.

Formulate the linear model

$$\boldsymbol{\mu} = \mathbf{Y}\boldsymbol{\gamma} \quad (7.37)$$

where \mathbf{Y} is the matrix specifying a specific design and $\boldsymbol{\gamma}$ is the vector of parameters.

Suppose e.g. that there exists a linear relationship between the dependent variable and one of the explanatory variables, the model becomes

$$\boldsymbol{\mu} = \begin{pmatrix} 1 & y_1 \\ 1 & y_2 \\ \vdots & \vdots \\ 1 & y_T \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} \quad (7.38)$$

where (y_1, y_1, \dots, y_T) are the corresponding values of one of the factors in the design.

Model (7.38) implies that $\boldsymbol{\mu}$ is a linear combination of the columns of \mathbf{Y} . Therefore, the linear model (7.38) on the treatment means implies the constraints

$$\mathbf{g}_{\text{mod}}(\boldsymbol{\mu}) = \mathbf{Q}_Y \boldsymbol{\mu} = \mathbf{0} \quad (7.39)$$

where $\mathbf{Q}_Y = \mathbf{I} - \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'$ is the projection matrix orthogonal to the columns of \mathbf{Y} .

Under equality of variances it follows from (7.15) that

$$\frac{1}{\sigma} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_T \end{pmatrix} = \begin{pmatrix} \frac{\mu_1}{\sigma} \\ \frac{\mu_2}{\sigma} \\ \vdots \\ \frac{\mu_T}{\sigma} \end{pmatrix} = \begin{pmatrix} \alpha_{21} \\ \alpha_{22} \\ \vdots \\ \alpha_{2T} \end{pmatrix} = \boldsymbol{\alpha}_2$$

leading to an equivalent formulation of the vector of constraints

$$\mathbf{g}_{\text{mod}}(\boldsymbol{\mu}) = \mathbf{Q}_Y \boldsymbol{\alpha}_2 = \mathbf{Q}_Y \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \quad (7.40)$$

which is expressed in terms of the so-called standardised means. The matrix of partial derivatives is

$$\mathbf{G}_{\text{mod}}(\boldsymbol{\pi}) = \mathbf{Q}_Y \frac{\partial \boldsymbol{\alpha}_2}{\partial \boldsymbol{\pi}} = \mathbf{Q}_Y \left(\mathbf{I}_T \otimes [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']_2 \right) \cdot \mathbf{D}_{\boldsymbol{\pi}}. \quad (7.41)$$

Utilizing the maximum likelihood procedure with

$$\begin{aligned} \mathbf{g}(\boldsymbol{\pi}) &= \begin{pmatrix} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{var}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{mod}}(\boldsymbol{\pi}) \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{I}_T \otimes \mathbf{Q}_X) \\ \mathbf{H} \cdot (\mathbf{I}_T \otimes [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']_1) \\ \mathbf{Q}_Y (\mathbf{I}_T \otimes [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']_2) \end{pmatrix} \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \end{aligned} \quad (7.42)$$

and

$$\begin{aligned} \mathbf{G}_{\boldsymbol{\pi}} &= \begin{pmatrix} \mathbf{G}_{\text{nor}}(\boldsymbol{\pi}) \\ \mathbf{G}_{\text{var}}(\boldsymbol{\pi}) \\ \mathbf{G}_{\text{mod}}(\boldsymbol{\pi}) \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{I}_T \otimes \mathbf{Q}_X) \\ \mathbf{H} \cdot (\mathbf{I}_T \otimes [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']_1) \\ \mathbf{Q}_Y (\mathbf{I}_T \otimes [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']_2) \end{pmatrix} \cdot \mathbf{D}_{\boldsymbol{\pi}} \end{aligned} \quad (7.43)$$

leads to the restricted ML estimate of $\boldsymbol{\pi}$ with the following properties:

1. $\hat{\boldsymbol{\pi}}_t$ for $t = 1, 2, \dots, T$ follows a cumulative normal distribution curve at the upper boundaries of \mathbf{x}
2. the fitted normal distributions have equal variances
3. the ML estimate $\hat{\boldsymbol{\mu}}$ satisfy the multifactor design in (7.36)

It is now possible to evaluate the effect of the factor(s) by means of the ML estimate

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Y}'\mathbf{Y})^* \mathbf{Y}' \hat{\boldsymbol{\mu}} \quad (7.44)$$

with asymptotic covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\gamma}}) = \{(\mathbf{Y}'\mathbf{Y})^* \mathbf{Y}'\} \text{Cov}(\hat{\boldsymbol{\mu}}) \{(\mathbf{Y}'\mathbf{Y})^* \mathbf{Y}'\}' . \quad (7.45)$$

7.4 Application: Single-factor model

A total of 898 students who were enrolled for a first year Statistics course at the University of Pretoria were included in this investigation. The students were all enrolled for Statistics (STATS) for the first time and obtained at least an E symbol for Grade 12 Mathematics (MATHS) on the higher grade. The aim of this study is to investigate the effect of achievement in MATHS on the performance of STATS. The STATS exam paper counted out of 108 marks and the results were classified into a total of 5 categories to illustrate the technique. The data is summarised in Table 7.1.

Table 7.1: Data set of 898 first year students.

MATHS	STATS					Total
	[0 – 40)	[40 – 50)	[50 – 60)	[60 – 75)	[75 – 108]	
A	0	4	19	53	84	160
B	3	17	35	65	19	139
C	24	44	56	68	19	211
D	43	57	82	48	6	236
E	59	53	26	13	1	152
Total	129	175	218	247	129	898

Take

$$\mathbf{x} = \begin{pmatrix} 39.5 \\ 49.5 \\ 59.5 \\ 74.4 \end{pmatrix} \quad (7.46)$$

as the vector of upper class boundaries. Since the exam mark is treated as a continuous variable and recorded to the nearest integer, the upper class boundaries in \mathbf{x} are taken half-way between the gaps of the respective class intervals. The performance in STATS will now be evaluated over the 5 levels of MATHS, specifying the 5 cells of the single-factor design. A total of 4 models will be fitted with the SAS program *FACTOR1* listed in Appendix B1 to explain the effect of MATHS on the grouped variable STATS.

7.4.1 Model 1: Unequal variances

It is assumed that the STATS mark is normally distributed for each level of MATHS. Therefore, normal distributions are fitted simultaneously to the 5 levels of MATHS, i.e. the 5 levels of the single-factor design. Normality within each cell is estimated such that $\Phi^{-1}(\pi_t)$ for $t = 1, 2, \dots, 5$ is a linear combination of

$$\begin{aligned} \mathbf{X} &= \begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix} \\ &= \begin{pmatrix} 39.5 & -1 \\ 49.5 & -1 \\ 59.5 & -1 \\ 74.4 & -1 \end{pmatrix} \end{aligned} \quad (7.47)$$

or equivalently such that $\Phi^{-1}(\pi_t)$ is orthogonal to

$$\mathbf{Q}_X = \mathbf{I}_4 - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' . \quad (7.48)$$

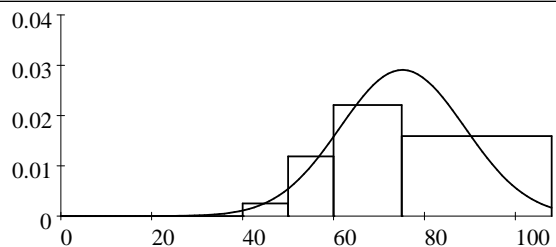
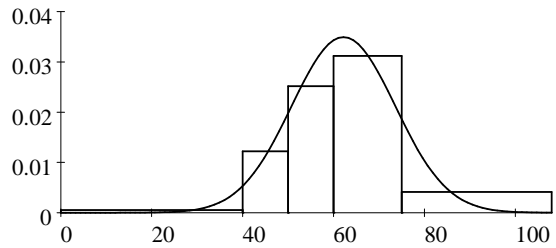
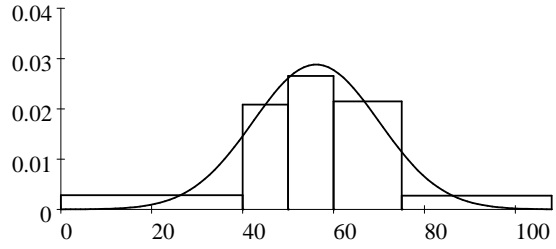
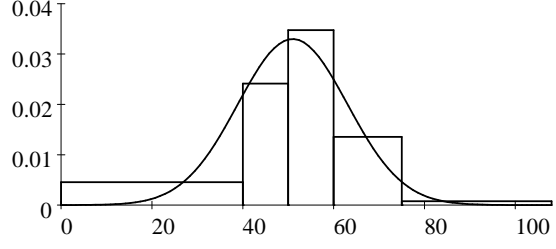
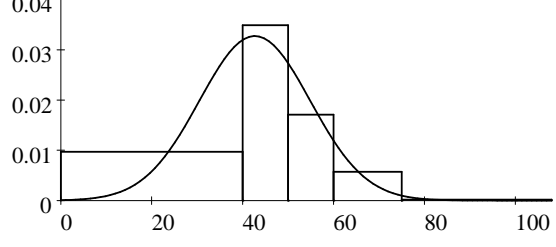
Since $\text{rank}(\mathbf{Q}_X) = 2$ the vector of constraints $\mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) = \mathbf{0}$, with

$$\begin{aligned} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) &= \begin{pmatrix} \mathbf{Q}_X \Phi^{-1}(\pi_1) \\ \mathbf{Q}_X \Phi^{-1}(\pi_2) \\ \mathbf{Q}_X \Phi^{-1}(\pi_3) \\ \mathbf{Q}_X \Phi^{-1}(\pi_4) \\ \mathbf{Q}_X \Phi^{-1}(\pi_5) \end{pmatrix} \\ &= (\mathbf{I}_5 \otimes \mathbf{Q}_X) \cdot \Phi^{-1}(\boldsymbol{\pi}) \end{aligned} \quad (7.49)$$

consists out of 10 linear independent functions.

Utilizing the ML estimation procedure, the restricted ML estimate for $\boldsymbol{\pi}$ is obtained leading to the ML estimates for the fitted normal distributions summarised in Table 7.2.

Table 7.2: ML estimates for model with unequal variances.

MATHS	STATS	n	$\hat{\mu}$ ($\hat{\sigma}_{\hat{\mu}}$)	$\hat{\sigma}$ ($\hat{\sigma}_{\hat{\sigma}}$)	$\hat{\tau}^M$ ($\hat{\sigma}_{\hat{\tau}^M}$)
A		160	75.2 (1.38)	13.7 (1.30)	17.7 (1.17)
B		139	62.2 (1.03)	11.4 (0.82)	4.8 (0.94)
C		211	56.1 (1.00)	13.8 (0.83)	-1.3 (0.91)
D		236	51.0 (0.83)	12.1 (0.70)	-6.5 (0.81)
E		152	42.7 (1.13)	12.2 (1.05)	-14.7 (1.00)
$\hat{\tau}_0$ ($\hat{\sigma}_{\hat{\tau}_0}$)					57.4 (0.49)

A definite positive monotone trend in STATS over the levels of MATHS is evident from Table 7.2. The $\hat{\mu}$ -values range from 42.7 for an E-symbol in MATHS, up to 75.2 for an A-symbol in MATHS. There is a slight variation with regard to the $\hat{\sigma}$ -values, revealing that students with a B symbol in MATHS had the smallest variation in STATS. According to the goodness of fit statistics tabulated in Table 7.3 the model fitted the data extremely well. The degrees of freedom in Table 7.3 follows from the number of linear independent constraints in (7.49).

Table 7.3: Goodness of fit statistics for model with unequal variances.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
1	10	7.059	0.7199	6.356	0.7845

The mean in the i -th level of MATHS may be expressed in terms of the single factor model

$$\mu_i = \tau_0 + \tau_i^M \quad i = 1, 2, \dots, 5 \quad (7.50)$$

where

$$\begin{aligned} \tau_0 &= \text{overall mean} \\ \tau_i^M &= \text{effect for the } i\text{-th level of MATHS} \quad i = 1, 2, \dots, 5 \end{aligned}$$

In matrix notation (7.50) leads to

$$\boldsymbol{\mu} = \mathbf{L}\boldsymbol{\lambda}$$

where

$$\mathbf{L} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{pmatrix} : 5 \times 5 \quad (7.51)$$

and λ denotes the vector of estimable parameters

$$\lambda = \begin{pmatrix} \tau_0 \\ \tau_1^M \\ \tau_2^M \\ \tau_3^M \\ \tau_4^M \end{pmatrix} \quad (7.52)$$

with the last parameter $\tau_5^M = -\sum_{i=1}^4 \tau_i^M$, the effect for an E symbol for MATHS, omitted.

From the restricted ML estimate $\hat{\pi}$, the ML estimate of λ is

$$\hat{\lambda} = (\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\hat{\mu} \quad (7.53)$$

with asymptotic covariance matrix

$$\text{Cov}(\hat{\lambda}) \cong \{(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\} \text{Cov}(\hat{\mu}) \{((\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}')'\} . \quad (7.54)$$

The full set of ML estimates in (7.50) is obtained from

$$\hat{\tau} = \mathbf{S}\hat{\lambda} \quad (7.55)$$

where

$$\hat{\tau} = \begin{pmatrix} \hat{\tau}_0 \\ \hat{\tau}_1^M \\ \hat{\tau}_2^M \\ \hat{\tau}_3^M \\ \hat{\tau}_4^M \\ \hat{\tau}_5^M \end{pmatrix} \quad (7.56)$$

and

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & -1 & -1 & -1 \end{pmatrix} : 6 \times 5 . \quad (7.57)$$

The asymptotic covariance matrix for $\hat{\tau}$ follows from

$$\text{Cov}(\hat{\tau}) \cong \mathbf{S} \text{Cov}(\hat{\lambda}) \mathbf{S}' \quad (7.58)$$

From the effects for the single factor model ($\hat{\tau}$ -values) listed in Table 7.2 it can be concluded that the average STATS mark for students with an A symbol in MATHS is 17.7 higher than the overall average of $\hat{\tau}_0 = 57.4$. The $\hat{\tau}^M$ -values drop substantially over the categories of MATHS indicating the strong effect of MATHS on STATS. The average STATS mark for C-symbol students is significantly lower than the overall average on the 10% level of significance, since the p-value is

$$\begin{aligned} \Phi\left(\frac{\hat{\tau}_3^M}{\hat{\sigma}_{\hat{\tau}_3^M}}\right) &= \Phi\left(\frac{-1.3}{0.91}\right) \\ &= \Phi(-1.428) \\ &= 0.08 . \end{aligned}$$

In SAS the matrices \mathbf{L} (7.51) and \mathbf{S} (7.57) may be programmed as:

- $\mathbf{L} = \boxed{\text{J}(5,1,1) \parallel \text{DESIGNF}(\text{CUSUM}(\text{J}(5,1,1)))}$
- $\mathbf{S} = \boxed{\text{BLOCK}(1, \text{DESIGNF}(\text{CUSUM}(\text{J}(5,1,1))))}$

where 5 is the number of levels for the single factor MATHS.

7.4.2 Model 2: Equal variances

From Table 7.2 it is clear that the standard deviations of the normal distributions stayed fairly stable over the levels of MATHS, implying that the additional constraints of equal variances $\mathbf{g}_{\text{var}}(\boldsymbol{\pi}) = \mathbf{0}$, with

$$\begin{aligned}\mathbf{g}_{\text{var}}(\boldsymbol{\pi}) &= \mathbf{H}\boldsymbol{\alpha}_1 \\ &= \mathbf{H} \cdot \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})\end{aligned}\quad (7.59)$$

where

$$\mathbf{H} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\alpha}_1 = \boldsymbol{\sigma}^{-1} = \begin{pmatrix} 1/\sigma_1 \\ 1/\sigma_2 \\ 1/\sigma_3 \\ 1/\sigma_4 \\ 1/\sigma_5 \end{pmatrix}$$

are feasible.

Note: Since the rows of \mathbf{H} are all orthogonal to the vector of ones, an equivalent formulation of the vector of constraints may be constructed with

$$\mathbf{g}_{\text{var}}(\boldsymbol{\pi}) = \mathbf{Q}_{\mathbf{H}}\boldsymbol{\alpha}_1$$

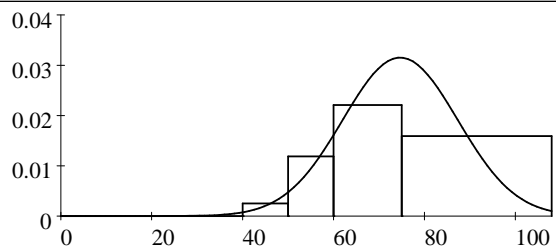
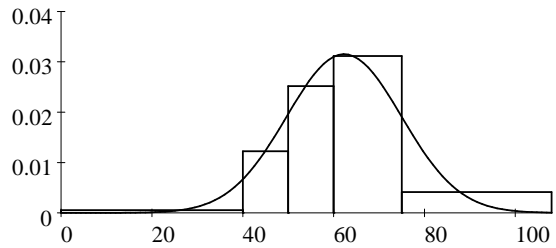
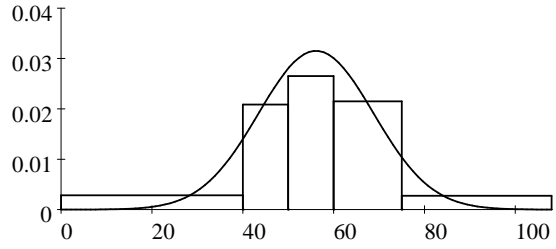
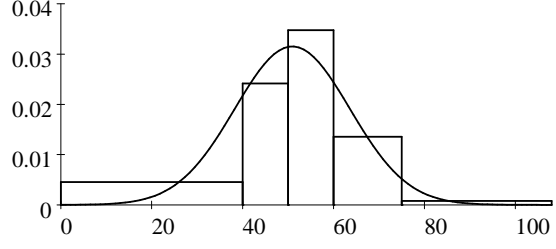
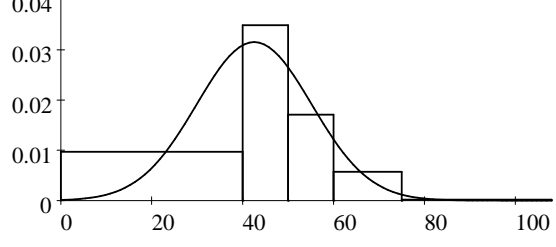
where $\mathbf{Q}_{\mathbf{H}} = \mathbf{I}_5 - \frac{1}{5}\mathbf{1}\mathbf{1}'$, is the projection matrix orthogonal to the vector of ones.

After employing the ML procedure with the vector of constraints

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{var}}(\boldsymbol{\pi}) \end{pmatrix} = \mathbf{0} \quad (7.60)$$

the restricted ML estimate $\hat{\boldsymbol{\pi}}$ was obtained and the results for Model 2 are summarised in Table 7.4.

Table 7.4: ML estimates for model with equal variances.

MATHS	STATS	n	$\hat{\mu}$ ($\hat{\sigma}_{\hat{\mu}}$)	$\hat{\sigma}$ ($\hat{\sigma}_{\hat{\sigma}}$)	$\hat{\tau}^M$ ($\hat{\sigma}_{\hat{\tau}^M}$)
A		160	74.7 (1.15)	12.7 (0.40)	17.3 (1.01)
B		139	62.3 (1.13)	12.7 (0.40)	5.0 (0.99)
C		211	56.1 (0.91)	12.7 (0.40)	-1.2 (0.85)
D		236	50.9 (0.87)	12.7 (0.40)	-6.4 (0.82)
E		152	42.5 (1.13)	12.7 (0.40)	-14.8 (0.99)
$\hat{\tau}_0$ ($\hat{\sigma}_{\hat{\tau}_0}$)					57.3 (0.47)

No substantial changes with regard to the $\hat{\mu}$ -values were obtained from that of Model 1, with the $\hat{\sigma}$ -values now estimated constant with $\hat{\sigma} = 12.7$. The values of the goodness of fit statistics in Table 7.5 increased somewhat from that of Model 1, but still provided a satisfactory fit.

Table 7.5: Goodness of fit statistics for model with equal variances.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
2	14	13.218	0.5094	12.374	0.5763

The degrees of freedom for this model is 14, since an additional 4 constraints were imposed in (7.59) for equality of variances.

7.4.3 Model 3: Ordinal factor

Due to the very strong monotone trend in STATS over the categories of MATHS, MATHS will now be incorporated as an ordinal factor in the ML estimation process. The single factor model on the levels of MATHS is

$$\boldsymbol{\mu} = \mathbf{Y}_3 \boldsymbol{\gamma}_3 \tag{7.61}$$

where

$$\mathbf{Y}_3 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \\ 1 & -1 \\ 1 & -2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma}_3 = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} .$$

The complete set of vector of constraints for Model 3 is

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{var}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{mod3}}(\boldsymbol{\pi}) \end{pmatrix} = \mathbf{0} \tag{7.62}$$

where

$$\begin{aligned} \mathbf{g}_{\text{mod3}}(\boldsymbol{\pi}) &= \mathbf{Q}_{Y_3} \boldsymbol{\alpha}_2 \\ &= \mathbf{Q}_{Y_3} \cdot \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \end{aligned} \quad (7.63)$$

and $\mathbf{Q}_{Y_3} = \mathbf{I}_5 - \mathbf{Y}_3 (\mathbf{Y}_3' \mathbf{Y}_3)^{-1} \mathbf{Y}_3'$.

Note: The vector of constraints in (7.63) is formulated in terms of $\boldsymbol{\alpha}_2$, since $\boldsymbol{\alpha}_2$ is a scalar multiple of $\boldsymbol{\mu}$ in (7.61) under equality of variances.

Utilizing the ML estimation procedure with the vector of constraints (7.62) the restricted ML estimate $\hat{\boldsymbol{\pi}}$ is estimated such that the vector $\hat{\boldsymbol{\mu}}$ is a linear combination of \mathbf{Y}_3 . (See Table 7.6.)

The ML estimate for $\boldsymbol{\gamma}_3$ is

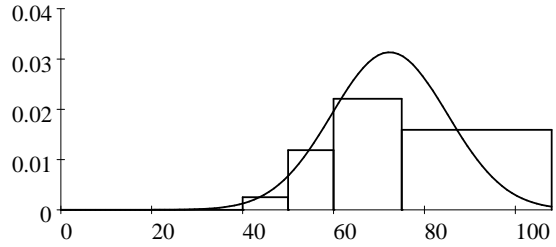
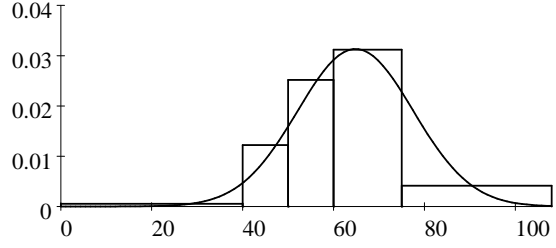
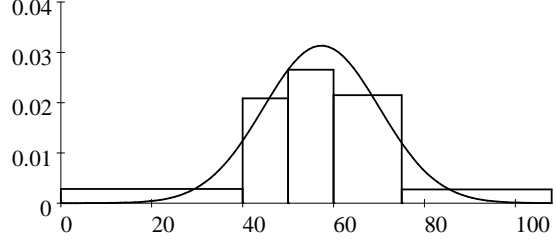
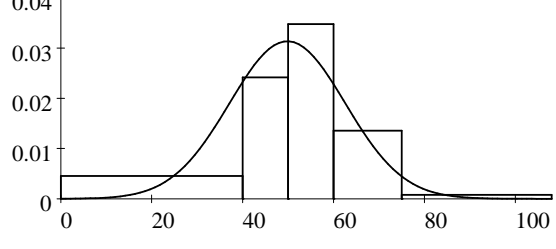
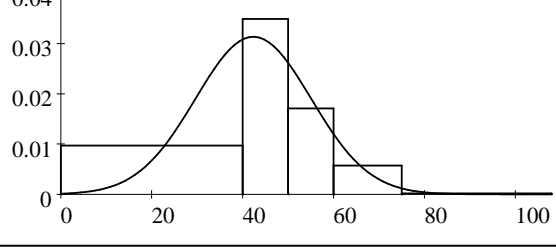
$$\begin{aligned} \hat{\boldsymbol{\gamma}}_3 &= (\mathbf{Y}_3' \mathbf{Y}_3)^{-1} \mathbf{Y}_3' \hat{\boldsymbol{\mu}} \\ &= \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} \\ &= \begin{pmatrix} 57.3 \\ 7.5 \end{pmatrix} \end{aligned}$$

indicating that the estimated average STATS mark for students with a C symbol for maths is 57.3 and that every increase of one symbol in MATHS implies an estimated increase of 7.5 in STATS. (See Table 7.6.) The standard errors of $\hat{\boldsymbol{\gamma}}_3$

$$\begin{aligned} \hat{\sigma}_{\hat{\boldsymbol{\gamma}}_3} &= \begin{pmatrix} \hat{\sigma}_{\hat{\gamma}_1} \\ \hat{\sigma}_{\hat{\gamma}_2} \end{pmatrix} \\ &= \begin{pmatrix} 0.4563 \\ 0.3521 \end{pmatrix} \end{aligned}$$

enable the construction of confidence intervals and the testing of relevant hypotheses.

Table 7.6: ML estimates for model with an ordinal factor.

MATHS	STATS	n	$\hat{\mu}$ ($\hat{\sigma}_{\hat{\mu}}$)	$\hat{\sigma}$ ($\hat{\sigma}_{\hat{\sigma}}$)	$\hat{\tau}^M$ ($\hat{\sigma}_{\hat{\tau}^M}$)
A		160	72.3 (0.87)	12.7 (0.40)	15.0 (0.70)
B		139	64.8 (0.60)	12.7 (0.40)	7.5 (0.35)
C		211	57.3 (0.46)	12.7 (0.40)	0.0 (0.00)
D		236	49.8 (0.55)	12.7 (0.40)	-7.5 (0.35)
E		152	42.4 (0.81)	12.7 (0.40)	-15.0 (0.70)
$\hat{\tau}_0$ ($\hat{\sigma}_{\hat{\tau}_0}$)					57.3 (0.46)

The value of the Pearson and Wald statistic in Table 7.7 increased substantially from that of the previous model indicating a weaker fit.

Table 7.7: Goodness of fit statistics for model with an ordinal factor.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
3	17	25.150	0.0914	24.388	0.1093

Since $\text{rank}(\mathbf{Q}_{Y_3}) = 3$, an additional 3 linear independent constraints are included in the vector of constraints leading to 17 degrees of freedom for Model 3.

7.4.4 Model 4: Regression model

Since the original scale of measurement for MATHS was done on an interval scale, the following class midpoints were taken as representative values for the five levels of MATHS.

MATHS	A	B	C	D	E
Class Midpoint	90	75	65	55	45

The implication of this is that the "distances" between the MATHS categories are not the same as in the case of Model 3.

The linear model measuring a linear trend in MATHS is

$$\boldsymbol{\mu} = \mathbf{Y}_4 \boldsymbol{\gamma}_4$$

where

$$\mathbf{Y}_4 = \begin{pmatrix} 1 & 90 \\ 1 & 75 \\ 1 & 65 \\ 1 & 55 \\ 1 & 45 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma}_4 = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} .$$

The complete set of vector of constraints for Model 4 is

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{var}}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{mod4}}(\boldsymbol{\pi}) \end{pmatrix} = \mathbf{0} \quad (7.64)$$

where

$$\begin{aligned} \mathbf{g}_{\text{mod4}}(\boldsymbol{\pi}) &= \mathbf{Q}_{\mathbf{Y}_4} \boldsymbol{\alpha}_2 \\ &= \mathbf{Q}_{\mathbf{Y}_4} \cdot \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \cdot \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \end{aligned}$$

and $\mathbf{Q}_{\mathbf{Y}_4} = \mathbf{I}_5 - \mathbf{Y}_4 (\mathbf{Y}'_4 \mathbf{Y}_4)^{-1} \mathbf{Y}'_4$.

The ML estimation procedure with vector of constraints (7.64) yields the ML estimate

$$\begin{aligned} \hat{\boldsymbol{\gamma}}_4 &= (\mathbf{Y}'_4 \mathbf{Y}_4)^{-1} \mathbf{Y}'_4 \hat{\boldsymbol{\mu}} \\ &= \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} \\ &= \begin{pmatrix} 12.2 \\ 0.68 \end{pmatrix} \end{aligned}$$

suggesting a slope of 0.68 for STATS on MATHS. This means that an increase of one mark in MATHS will lead to an estimated increase of 0.68 marks in STATS. From the vector of standard errors

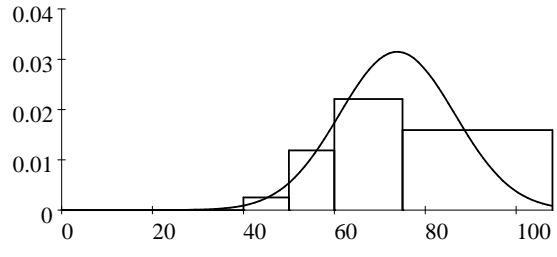
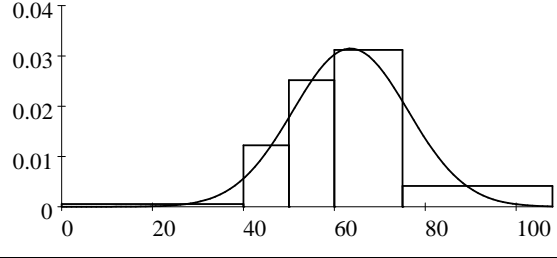
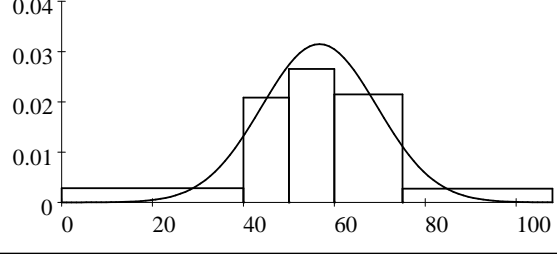
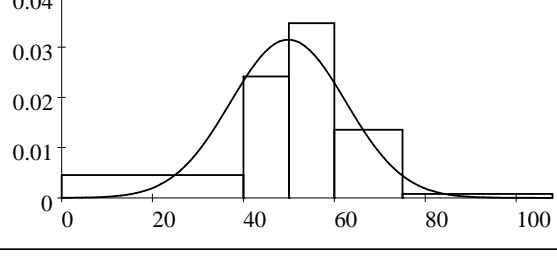
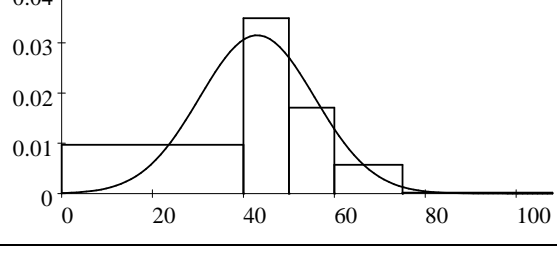
$$\begin{aligned} \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\gamma}}_4} &= \begin{pmatrix} \hat{\sigma}_{\hat{\gamma}_1} \\ \hat{\sigma}_{\hat{\gamma}_2} \end{pmatrix} \\ &= \begin{pmatrix} 2.108 \\ 0.0319 \end{pmatrix} \end{aligned}$$

this increase is significant, since

$$\begin{aligned} \frac{\hat{\gamma}_2}{\hat{\sigma}_{\hat{\gamma}_2}} &= \frac{0.68}{0.0319} \\ &= 21.317 . \end{aligned}$$

See Table 7.8 for the complete set of the ML estimates.

Table 7.8: ML estimates for regression model.

Maths	Stats	n	$\hat{\mu}$ ($\hat{\sigma}_{\hat{\mu}}$)	$\hat{\sigma}$ ($\hat{\sigma}_{\hat{\sigma}}$)	$\hat{\tau}^M$ ($\hat{\sigma}_{\hat{\tau}^M}$)
A		160	73.8 (0.93)	12.7 (0.40)	16.4 (0.77)
B		139	63.6 (0.56)	12.7 (0.40)	6.2 (0.29)
C		211	56.7 (0.45)	12.7 (0.40)	-0.7 (0.03)
D		236	49.9 (0.55)	12.7 (0.40)	-7.5 (0.35)
E		152	43.0 (0.77)	12.7 (0.40)	-14.4 (0.67)
$\hat{\tau}_0$ ($\hat{\sigma}_{\hat{\tau}_0}$)					57.4 (0.46)

According to the goodness of fit statistics tabulated in Table 7.9, this model showed a substantial better fit than the previous model where MATHS was modelled on an ordinal scale.

Table 7.9: Goodness of fit statistics for regression model.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
4	17	16.813	0.4671	16.010	0.5168

Chapter 8

Log-logistic distributions

In the case where the grouped response vector has a positive skew distribution, the log-logistic distribution may be fitted very effectively to the T frequency distributions of a multifactor design. Due to the skewness of the response variable, the median of the fitted log-logistic distributions will be used as a representative measure for each of the T frequency distributions.

From the cdf of the log-logistic distribution

$$F(x; \kappa, \theta) = \frac{e^{\theta x^{\kappa}}}{1 + e^{\theta x^{\kappa}}}$$

the median ν is obtained from

$$\frac{e^{\theta \nu^{\kappa}}}{1 + e^{\theta \nu^{\kappa}}} = 0.5$$

leading to

$$\nu = \exp\left(-\frac{\theta}{\kappa}\right). \quad (8.1)$$

In the multifactor model the medians will be employed in a linear model to determine the effect of the explanatory variables or so-called factors on the grouped response variable.

8.1 Estimation of distributions

Analogous to Section 5.2, where a log-logistic curve was fitted to a single frequency distribution, the log-logistic curve may be fitted simultaneously to the T cells of a multifactor design using

$$\begin{aligned}
 \ln \left(\frac{\pi}{\mathbf{1} - \pi} \right) &= \begin{pmatrix} \ln \left(\frac{\pi_1}{\mathbf{1} - \pi_1} \right) \\ \ln \left(\frac{\pi_2}{\mathbf{1} - \pi_2} \right) \\ \vdots \\ \ln \left(\frac{\pi_T}{\mathbf{1} - \pi_T} \right) \end{pmatrix} \\
 &= \begin{pmatrix} \kappa_1 \ln \mathbf{x} + \theta_1 \mathbf{1} \\ \kappa_2 \ln \mathbf{x} + \theta_2 \mathbf{1} \\ \vdots \\ \kappa_T \ln \mathbf{x} + \theta_T \mathbf{1} \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{X}\alpha_1 \\ \mathbf{X}\alpha_2 \\ \vdots \\ \mathbf{X}\alpha_T \end{pmatrix} \\
 &= (\mathbf{I}_T \otimes \mathbf{X}) \boldsymbol{\alpha}
 \end{aligned} \tag{8.2}$$

where

$$\mathbf{X} = \begin{pmatrix} \ln \mathbf{x} & \mathbf{1} \end{pmatrix} \tag{8.3}$$

is the design matrix for a log-logistic distribution and

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_T \end{pmatrix} \quad \text{where} \quad \alpha_t = \begin{pmatrix} \kappa_t \\ \theta_t \end{pmatrix}, \quad t = 1 \dots T \tag{8.4}$$

is the concatenated vector of parameters.

The linear model (8.2) suggests the vector of constraints

$$\mathbf{g}_{\log}(\boldsymbol{\pi}) = \mathbf{0}$$

where

$$\begin{aligned} \mathbf{g}_{\log}(\boldsymbol{\pi}) &= \begin{pmatrix} \mathbf{Q}_X \ln \left(\frac{\pi_1}{\mathbf{1} - \boldsymbol{\pi}_1} \right) \\ \mathbf{Q}_X \ln \left(\frac{\pi_2}{\mathbf{1} - \boldsymbol{\pi}_2} \right) \\ \vdots \\ \mathbf{Q}_X \ln \left(\frac{\pi_T}{\mathbf{1} - \boldsymbol{\pi}_T} \right) \end{pmatrix} \\ &= (\mathbf{I}_T \otimes \mathbf{Q}_X) \cdot \ln \left(\frac{\boldsymbol{\pi}}{\mathbf{1} - \boldsymbol{\pi}} \right) \end{aligned} \quad (8.5)$$

with $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ the projection matrix orthogonal to the columns of \mathbf{X} given in (8.3).

The matrix of partial derivatives is

$$\begin{aligned} \mathbf{G}_{\log}(\boldsymbol{\pi}) &= \frac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= (\mathbf{I}_T \otimes \mathbf{Q}_X) \cdot \mathbf{D}_{\boldsymbol{\pi}} \end{aligned} \quad (8.6)$$

where

$$\begin{aligned} \mathbf{D}_{\boldsymbol{\pi}} &= \frac{\partial}{\partial \boldsymbol{\pi}} \ln \left(\frac{\boldsymbol{\pi}}{\mathbf{1} - \boldsymbol{\pi}} \right) \\ &= \frac{\partial}{\partial \boldsymbol{\pi}} \{ \ln(\boldsymbol{\pi}) - \ln(\mathbf{1} - \boldsymbol{\pi}) \} \\ &= \{ \text{diag}(\boldsymbol{\pi}) \}^{-1} + \{ \text{diag}(\mathbf{1} - \boldsymbol{\pi}) \}^{-1} . \end{aligned} \quad (8.7)$$

Employing the maximum likelihood procedure with

$$\mathbf{g}(\boldsymbol{\pi}) = \mathbf{g}_{\log}(\boldsymbol{\pi}) \quad \text{and} \quad \mathbf{G}_{\boldsymbol{\pi}} = \mathbf{G}_{\log}(\boldsymbol{\pi}) \quad (8.8)$$

the restricted ML estimate of $\boldsymbol{\pi}$ follows with asymptotic covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\pi}}) \cong \mathbf{V} - (\mathbf{G}_{\boldsymbol{\pi}} \mathbf{V})' (\mathbf{G}_{\boldsymbol{\pi}} \mathbf{V} \mathbf{G}_{\boldsymbol{\pi}}')^* (\mathbf{G}_{\boldsymbol{\pi}} \mathbf{V}) . \quad (8.9)$$

From the restricted ML estimator $\hat{\pi}$, it is possible to obtain the ML estimator of α

$$\hat{\alpha} = \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \cdot \ln \left(\frac{\hat{\pi}}{\mathbf{1} - \hat{\pi}} \right) \quad (8.10)$$

which consists out of two sets of estimators namely

$$\hat{\kappa} = \begin{pmatrix} \hat{\kappa}_1 \\ \hat{\kappa}_2 \\ \vdots \\ \hat{\kappa}_T \end{pmatrix} = \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \cdot \ln \left(\frac{\hat{\pi}}{\mathbf{1} - \hat{\pi}} \right) \quad (8.11)$$

and

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_T \end{pmatrix} = \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \cdot \ln \left(\frac{\hat{\pi}}{\mathbf{1} - \hat{\pi}} \right) . \quad (8.12)$$

The asymptotic covariance matrix of $\hat{\alpha}$ is

$$\text{Cov}(\hat{\alpha}) \cong \left\{ \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{D}_\pi \right\} \text{Cov}(\hat{\pi}) \left\{ \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{D}_\pi \right\}' \quad (8.13)$$

with \mathbf{D}_π given in (8.7).

The asymptotic standard errors of $\hat{\kappa}$ and $\hat{\theta}$ can be calculated directly from

$$\text{Cov}(\hat{\kappa}) \cong \left\{ \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \mathbf{D}_\pi \right\} \text{Cov}(\hat{\pi}) \left\{ \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_1 \right) \mathbf{D}_\pi \right\}' \quad (8.14)$$

and

$$\text{Cov}(\hat{\theta}) \cong \left\{ \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \mathbf{D}_\pi \right\} \text{Cov}(\hat{\pi}) \left\{ \left(\mathbf{I}_T \otimes \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]_2 \right) \mathbf{D}_\pi \right\}' . \quad (8.15)$$

8.2 Multifactor model

In the case where log-logistic distributions are fitted simultaneously to a grouped positive skew response variable in a multifactor design, the median (8.1) will be used as a representative measure for each cell. The medians of the fitted log-logistic distributions will be employed in a linear model to evaluate the effect of the explanatory variables on the response variable over the T cells of the multifactor design.

The concatenated vector of medians for the T cells in the multifactor design is

$$\boldsymbol{\nu} = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_T \end{pmatrix} = \exp \left(\begin{pmatrix} \begin{pmatrix} -\theta_1 \\ \kappa_1 \end{pmatrix} \\ \begin{pmatrix} -\theta_2 \\ \kappa_2 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} -\theta_T \\ \kappa_T \end{pmatrix} \end{pmatrix} \right) = \exp \left(-\frac{\boldsymbol{\theta}}{\boldsymbol{\kappa}} \right). \quad (8.16)$$

Let

$$\boldsymbol{\nu} = \mathbf{Y}\boldsymbol{\gamma} \quad (8.17)$$

specify the the multifactor model. The objective is to estimate $\boldsymbol{\pi}$ such that $\boldsymbol{\nu}$ is in the vector space generated by the columns of \mathbf{Y} implying the vector of constraints

$$\mathbf{g}_{\text{mod}}(\boldsymbol{\pi}) = \mathbf{Q}_Y \boldsymbol{\nu} = \mathbf{0} \quad (8.18)$$

with $\mathbf{Q}_Y = \mathbf{I} - \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'$ the projection matrix orthogonal to the columns of \mathbf{Y} . Implementing the chain rule the matrix of partial derivatives

$$\begin{aligned} \mathbf{G}_{\text{mod}}(\boldsymbol{\pi}) &= \frac{\partial \mathbf{Q}_Y \boldsymbol{\nu}}{\partial \boldsymbol{\pi}} \\ &= \mathbf{Q}_Y \cdot \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\alpha}} \cdot \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\pi}} \\ &= \mathbf{Q}_Y \cdot \mathbf{A} \cdot \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{D}_{\boldsymbol{\pi}} \end{aligned} \quad (8.19)$$

follows, where

$$\mathbf{A} = \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\alpha}} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_T \end{pmatrix} \quad (8.20)$$

and

$$\begin{aligned} \mathbf{A}_t &= \frac{\partial \boldsymbol{\nu}_t}{\partial \boldsymbol{\alpha}_t} \\ &= \frac{\partial \exp\left(-\frac{\theta_t}{\kappa_t}\right)}{\partial \begin{pmatrix} \kappa_t \\ \theta_t \end{pmatrix}} \\ &= \begin{pmatrix} \frac{\theta_t}{\kappa_t^2} \exp\left(-\frac{\theta_t}{\kappa_t}\right) & -\frac{1}{\kappa_t} \exp\left(-\frac{\theta_t}{\kappa_t}\right) \end{pmatrix}, \quad t = 1 \cdots T. \end{aligned} \quad (8.21)$$

To compute \mathbf{A} in (8.20) define the two vectors

$$\mathbf{a}_\kappa = \begin{pmatrix} \frac{\theta_1}{\kappa_1^2} \exp\left(-\frac{\theta_1}{\kappa_1}\right) \\ \frac{\theta_2}{\kappa_2^2} \exp\left(-\frac{\theta_2}{\kappa_2}\right) \\ \vdots \\ \frac{\theta_T}{\kappa_T^2} \exp\left(-\frac{\theta_T}{\kappa_T}\right) \end{pmatrix} = \frac{\boldsymbol{\theta}}{\boldsymbol{\kappa}^2} \exp\left(-\frac{\boldsymbol{\theta}}{\boldsymbol{\kappa}}\right) \quad (8.22)$$

and

$$\mathbf{a}_\theta = \begin{pmatrix} -\frac{1}{\kappa_1} \exp\left(-\frac{\theta_1}{\kappa_1}\right) \\ -\frac{1}{\kappa_2} \exp\left(-\frac{\theta_2}{\kappa_2}\right) \\ \vdots \\ -\frac{1}{\kappa_T} \exp\left(-\frac{\theta_T}{\kappa_T}\right) \end{pmatrix} = -\frac{\mathbf{1}}{\boldsymbol{\kappa}} \exp\left(-\frac{\boldsymbol{\theta}}{\boldsymbol{\kappa}}\right). \quad (8.23)$$

Using (8.22) and (8.23) the matrix \mathbf{A} may be calculated from

$$\mathbf{A} = \left(\text{diag}[\mathbf{a}_\kappa] \otimes \begin{pmatrix} 1 & 0 \end{pmatrix} \right) + \left(\text{diag}[\mathbf{a}_\theta] \otimes \begin{pmatrix} 0 & 1 \end{pmatrix} \right). \quad (8.24)$$

Employing the ML estimation procedure with

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_{\log}(\boldsymbol{\pi}) \\ \mathbf{G}_{\text{mod}}(\boldsymbol{\pi}) \end{pmatrix} \quad \text{and} \quad \mathbf{G}_{\boldsymbol{\pi}} = \begin{pmatrix} \mathbf{G}_{\log}(\boldsymbol{\pi}) \\ \mathbf{G}_{\text{mod}}(\boldsymbol{\pi}) \end{pmatrix} \quad (8.25)$$

leads to the restricted ML estimate of $\boldsymbol{\pi}$ such that:

1. the elements of $\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\pi}}_2, \dots, \hat{\boldsymbol{\pi}}_T$ follow T log-logistic curves at the upper boundaries of \mathbf{x} and
2. the ML estimate

$$\hat{\boldsymbol{\nu}} = \exp \left(-\frac{\hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\kappa}}} \right)$$

is a linear combination of \mathbf{Y} in (8.17).

The asymptotic covariance matrix of $\hat{\boldsymbol{\nu}}$ is

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\nu}}) &\cong \left\{ \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\alpha}} \right\} \text{Cov}(\hat{\boldsymbol{\alpha}}) \left\{ \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\alpha}} \right\}' \\ &= \mathbf{A} \text{Cov}(\hat{\boldsymbol{\alpha}}) \mathbf{A}' . \end{aligned} \quad (8.26)$$

The effect of the factors for the multifactor design can be explained from the ML estimate

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\hat{\boldsymbol{\nu}} \quad (8.27)$$

and for the purpose of statistical inference, the standard errors are obtained from the asymptotic covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\gamma}}) \cong \left\{ (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}' \right\} \text{Cov}(\hat{\boldsymbol{\nu}}) \left\{ (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}' \right\}' . \quad (8.28)$$

8.3 Application: Two-factor model

The premiums of 8334 policyholders in the short-term insurance are classified into the 5 categories listed in Table 8.1.

Table 8.1: Frequency distribution of PREMIUM.

PREMIUM	Frequency
R51-R200	1920
R201-R300	2726
R301-R400	1677
R401-R500	930
R500-R1000	1081

The objective is to explain the effect of the age of the policyholder (AGE) and the type of product (PRODUCT) on the PREMIUM of the policy. The variable AGE is classified into 4 categories, while PRODUCT consists out of three types of insurance policies. A cross classification of these two factors result in a total of 12 cells summarised in Table 8.2.

Table 8.2: Contingency table of AGE and PRODUCT.

AGE	PRODUCT			Total
	I	II	III	
20-29	930	415	461	1806
30-39	1105	800	1017	2922
40-49	832	764	656	2252
50-59	448	416	490	1354
Total	3315	2395	2624	8334

The 12 cells in Table 8.2 are to be modeled in a two-factor design. Due to the positive skew nature of PREMIUM a log-logistic curve will be fitted to the frequency distribution of PREMIUM in each

of the 12 cells. The variable PREMIUM is modeled in hundreds of rands, which implies that the vector of upper class boundaries is

$$\mathbf{x} = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}. \quad (8.29)$$

(See Table 8.1.) The median of the fitted log-logistic curves will be modeled over the 12 cells to investigate the effect of the two factors AGE and PRODUCT on PREMIUM. This will be described in a total of 4 models. The results for all 4 models were all obtained from the SAS program *FACTOR2* listed in Appendix B2.

8.3.1 Model 1: Saturated model

A log-logistic curve is fitted to every cell in the two-factor design, such that

$$\ln \left(\frac{\pi_t}{\mathbf{1} - \pi_t} \right), \quad t = 1, 2, \dots, 12$$

is in the column space of

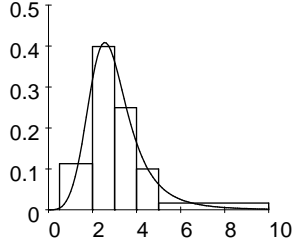
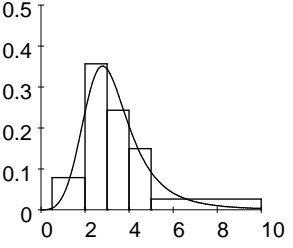
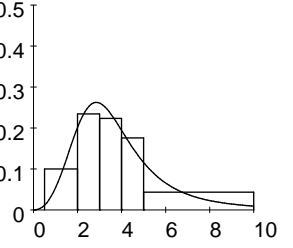
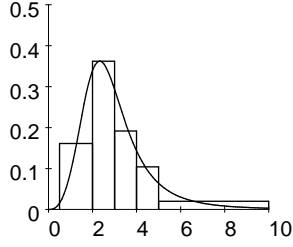
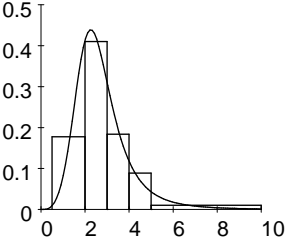
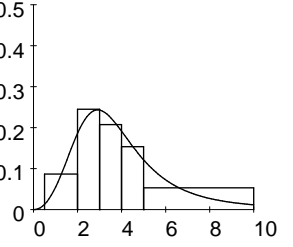
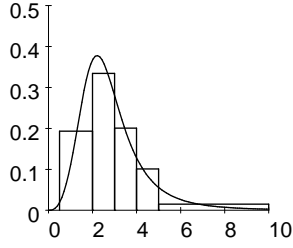
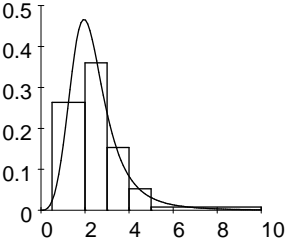
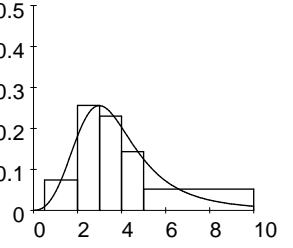
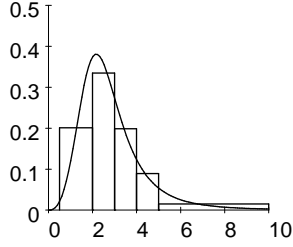
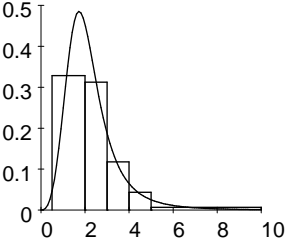
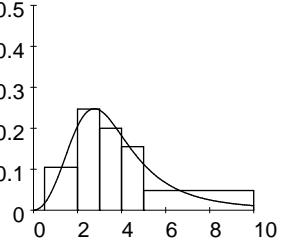
$$\mathbf{X} = \begin{pmatrix} \ln \mathbf{x} & \mathbf{1} \end{pmatrix} = \begin{pmatrix} \ln 2 & 1 \\ \ln 3 & 1 \\ \ln 4 & 1 \\ \ln 5 & 1 \end{pmatrix}. \quad (8.30)$$

Implementing the vector of constraints $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{g}_{\log}(\boldsymbol{\pi}) = \mathbf{0}$ with

$$\mathbf{g}_{\log}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{Q}_X \ln \left(\frac{\pi_1}{\mathbf{1} - \pi_1} \right) \\ \mathbf{Q}_X \ln \left(\frac{\pi_2}{\mathbf{1} - \pi_2} \right) \\ \vdots \\ \mathbf{Q}_X \ln \left(\frac{\pi_{12}}{\mathbf{1} - \pi_{12}} \right) \end{pmatrix} \quad (8.31)$$

where $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, in the ML estimation procedure, a total of 12 log-logistic distributions are fitted simultaneously to the frequency distributions of the two-factor design listed in Table 8.3.

Table 8.3: Descriptive statistics for the saturated model.

AGE	PRODUCT		
	I	II	III
20-29	 $\hat{\mu} = 3.097$ $\hat{\nu} = 2.842$ $\hat{\sigma} = 1.427$ $\hat{\sigma}_{\hat{\nu}} = 0.038$	 $\hat{\mu} = 3.462$ $\hat{\nu} = 3.143$ $\hat{\sigma} = 1.715$ $\hat{\sigma}_{\hat{\nu}} = 0.066$	 $\hat{\mu} = 4.031$ $\hat{\nu} = 3.447$ $\hat{\sigma} = 2.786$ $\hat{\sigma}_{\hat{\nu}} = 0.086$
30-39	 $\hat{\mu} = 3.080$ $\hat{\nu} = 2.712$ $\hat{\sigma} = 1.831$ $\hat{\sigma}_{\hat{\nu}} = 0.040$	 $\hat{\mu} = 2.790$ $\hat{\nu} = 2.538$ $\hat{\sigma} = 1.365$ $\hat{\sigma}_{\hat{\nu}} = 0.039$	 $\hat{\mu} = 4.260$ $\hat{\nu} = 3.588$ $\hat{\sigma} = 3.167$ $\hat{\sigma}_{\hat{\nu}} = 0.063$
40-49	 $\hat{\mu} = 2.941$ $\hat{\nu} = 2.584$ $\hat{\sigma} = 1.768$ $\hat{\sigma}_{\hat{\nu}} = 0.045$	 $\hat{\mu} = 2.496$ $\hat{\nu} = 2.235$ $\hat{\sigma} = 1.349$ $\hat{\sigma}_{\hat{\nu}} = 0.039$	 $\hat{\mu} = 4.173$ $\hat{\nu} = 3.588$ $\hat{\sigma} = 2.806$ $\hat{\sigma}_{\hat{\nu}} = 0.074$
50-59	 $\hat{\mu} = 2.903$ $\hat{\nu} = 2.544$ $\hat{\sigma} = 1.768$ $\hat{\sigma}_{\hat{\nu}} = 0.061$	 $\hat{\mu} = 2.295$ $\hat{\nu} = 2.019$ $\hat{\sigma} = 1.372$ $\hat{\sigma}_{\hat{\nu}} = 0.054$	 $\hat{\mu} = 4.131$ $\hat{\nu} = 3.443$ $\hat{\sigma} = 3.223$ $\hat{\sigma}_{\hat{\nu}} = 0.090$

The log-logistic curves tabulated in Table 8.3 provide an excellent fit for PREMIUM. This is further motivated by the goodness of fit statistics reported in Table 8.4. The degrees of freedom follows from the 24 linear independent constraints in (8.31).

Table 8.4: Goodness of fit statistics for the saturated model.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
1	24	30.799	0.1597	30.266	0.1761

Evaluating the means ($\hat{\mu}$) and medians ($\hat{\nu}$) in Table 8.3 it is clear that Product III is the most expensive product. The standard deviations ($\hat{\sigma}$) indicate that the variation in PREMIUM is the highest for Product III which can also be seen from the some-what flatter log-logistic curves displayed in Table 8.3. Product II portrays the most drastic drop in PREMIUM over the categories of AGE indicating a possible interaction between AGE and PRODUCT.

Define the following functions of the medians:

$$\begin{aligned} \nu_{ij}^{AP} & : \text{median in } (ij)\text{-th cell} \\ \bar{\nu}_i^A &= \frac{1}{3} \sum_{j=1}^3 \nu_{ij}^{AP} : \text{average median for } i\text{-th level of AGE} \\ \bar{\nu}_j^P &= \frac{1}{4} \sum_{i=1}^4 \nu_{ij}^{AP} : \text{average median for } j\text{-th level of PRODUCT} \\ \bar{\nu} &= \frac{1}{12} \sum_{i=1}^4 \sum_{j=1}^3 \nu_{ij}^{AP} = \frac{1}{4} \sum_{i=1}^4 \bar{\nu}_i^A = \frac{1}{3} \sum_{j=1}^3 \bar{\nu}_j^P : \text{overall average median} \end{aligned}$$

The median of the (ij) -th cell may be expressed by the two-factor model

$$\nu_{ij} = \tau_0 + \tau_i^A + \tau_j^P + \tau_{ij}^{AP} \quad , \quad i = 1, 2, 3, 4 \quad \text{and} \quad j = 1, 2, 3 \quad (8.32)$$

where

$$\begin{aligned} \tau_0 &= \bar{\nu} && : \text{ overall median} \\ \tau_i^A &= \bar{\nu}_i^A - \tau_0 && : \text{ effect for the } i\text{-th level of AGE} \\ &= \bar{\nu}_i^A - \bar{\nu} \\ \tau_j^P &= \bar{\nu}_j^P - \tau_0 && : \text{ effect for the } j\text{-th level of PRODUCT} \\ &= \bar{\nu}_j^P - \bar{\nu} \\ \tau_{ij}^{AP} &= \nu_{ij}^{AP} - (\tau_0 + \tau_i^A + \tau_j^P) && : \text{ interaction effect for the } i\text{-th level of AGE} \\ &= \nu_{ij}^{AP} - \bar{\nu}_i^A - \bar{\nu}_j^P + \bar{\nu} && : \text{ and } j\text{-th level of PRODUCT} \end{aligned}$$

Since

$$\sum_{i=1}^4 \tau_i^A = \sum_{j=1}^3 \tau_j^P = \sum_{i=1}^4 \tau_{ij}^{AP} = \sum_{j=1}^3 \tau_{ij}^{AP} = 0 \quad (8.33)$$

it follows for the main effects that

$$\tau_4^A = - \sum_{i=1}^3 \tau_i^A \quad \text{and} \quad \tau_3^P = - \sum_{j=1}^2 \tau_j^P \quad (8.34)$$

and for the interaction effects that

$$\tau_{4j}^{AP} = - \sum_{i=1}^3 \tau_{ij}^{AP} \quad \text{and} \quad \tau_{i3}^{AP} = - \sum_{j=1}^2 \tau_{ij}^{AP} . \quad (8.35)$$

In matrix notation, the saturated model (8.32) may be written as

$$\begin{pmatrix} \nu_{11} \\ \nu_{12} \\ \nu_{13} \\ \nu_{21} \\ \nu_{22} \\ \nu_{23} \\ \nu_{31} \\ \nu_{32} \\ \nu_{33} \\ \nu_{41} \\ \nu_{42} \\ \nu_{43} \end{pmatrix} = \mathbf{Z}\boldsymbol{\lambda} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \tau_0 \\ \tau_1^A \\ \tau_2^A \\ \tau_3^A \\ \tau_1^P \\ \tau_2^P \\ \tau_{11}^{AP} \\ \tau_{12}^{AP} \\ \tau_{21}^{AP} \\ \tau_{22}^{AP} \\ \tau_{31}^{AP} \\ \tau_{32}^{AP} \end{pmatrix} \quad (8.36)$$

where $\mathbf{Z} : (12 \times 12)$ is the design matrix and $\boldsymbol{\lambda} : (12 \times 1)$ consists out of the estimable parameters.

Since AGE has 4 levels and PRODUCT has 3 levels define the design matrices

$$\mathbf{D}_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix} \quad \text{and} \quad \mathbf{D}_P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \quad (8.37)$$

with corresponding vectors of ones

$$\mathbf{1}_A = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{1}_P = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \quad (8.38)$$

The saturated model (8.36) may therefore be partitioned as

$$\begin{aligned} \boldsymbol{\nu} &= \mathbf{Z}\boldsymbol{\lambda} \\ &= \begin{pmatrix} \mathbf{1} & \mathbf{Z}_A & \mathbf{Z}_P & \mathbf{Z}_{AP} \end{pmatrix} \begin{pmatrix} \tau_0 \\ \boldsymbol{\lambda}^A \\ \boldsymbol{\lambda}^P \\ \boldsymbol{\lambda}^{AP} \end{pmatrix} \end{aligned} \quad (8.39)$$

with a description of the submatrices and parameters listed in Table 8.5.

Table 8.5: Partitioning of the saturated model.

Submatrices	Parameters
$\mathbf{1} = \mathbf{1}_A \otimes \mathbf{1}_P : (12 \times 1)$	τ_0 : overall median
$\mathbf{Z}_A = \mathbf{D}_A \otimes \mathbf{1}_P : (12 \times 3)$	$\boldsymbol{\lambda}^A : \begin{pmatrix} \tau_1^A \\ \tau_2^A \\ \tau_3^A \end{pmatrix} =$ effects for AGE
$\mathbf{Z}_P = \mathbf{1}_A \otimes \mathbf{D}_P : (12 \times 2)$	$\boldsymbol{\lambda}^P : \begin{pmatrix} \tau_1^P \\ \tau_2^P \end{pmatrix} =$ effects for PRODUCT
$\mathbf{Z}_{AP} = \mathbf{Z}_A \odot \mathbf{Z}_P : (12 \times 6)$	$\boldsymbol{\lambda}^{AP} : \begin{pmatrix} \tau_{11}^{AP} \\ \tau_{12}^{AP} \\ \tau_{21}^{AP} \\ \tau_{22}^{AP} \\ \tau_{31}^{AP} \\ \tau_{32}^{AP} \end{pmatrix} =$ interaction effects for AGE and PRODUCT

Note: The operator \odot in Table 8.5 performs a direct product on all rows of \mathbf{Z}_A and \mathbf{Z}_P . The result has the same number of rows as \mathbf{Z}_A and \mathbf{Z}_P and the number of columns is equal to the product of the number of columns of \mathbf{Z}_A and \mathbf{Z}_P . See (8.36).

The ML estimate for λ is

$$\hat{\lambda} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\hat{\nu} \quad (8.40)$$

with asymptotic covariance matrix

$$\text{Cov}(\hat{\lambda}) \cong \left\{ (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \right\} \text{Cov}(\hat{\nu}) \left\{ (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \right\}' . \quad (8.41)$$

The complete set of effects for the two-factor design may be obtained from

$$\hat{\tau} = \mathbf{S}\hat{\lambda} \quad (8.42)$$

where

$$\begin{aligned} \mathbf{S} &= \text{Block} \left(1 \quad \mathbf{D}_A \quad \mathbf{D}_P \quad \mathbf{D}_A \otimes \mathbf{D}_P \right) \\ &= \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_A & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_A \otimes \mathbf{D}_P \end{pmatrix} \end{aligned} \quad (8.43)$$

and

$$\hat{\tau} = \begin{pmatrix} \hat{\tau}_0 \\ \hat{\tau}^A \\ \hat{\tau}^B \\ \hat{\tau}^{AP} \end{pmatrix} : (20 \times 1) \quad (8.44)$$

consists out of all the effects for the two-factor model. In (8.44) the main effects are

$$\hat{\tau}^A = \begin{pmatrix} \hat{\tau}_1^A \\ \hat{\tau}_2^A \\ \hat{\tau}_3^A \\ \hat{\tau}_4^A \end{pmatrix} \quad \text{and} \quad \hat{\tau}^B = \begin{pmatrix} \hat{\tau}_1^B \\ \hat{\tau}_2^B \\ \hat{\tau}_3^B \end{pmatrix} \quad (8.45)$$

with the interaction effects included in

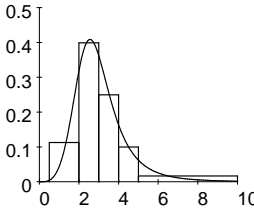
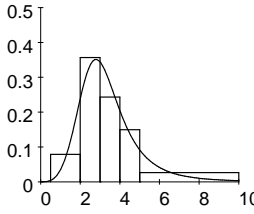
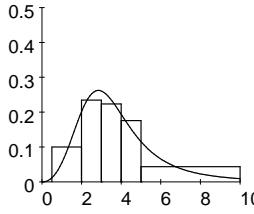
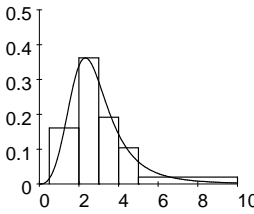
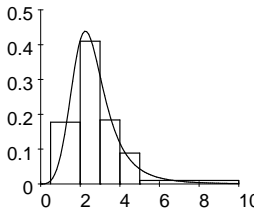
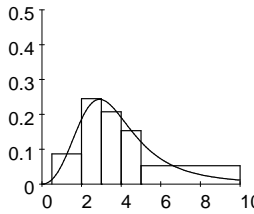
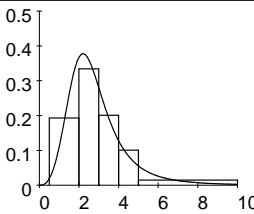
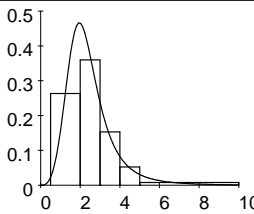
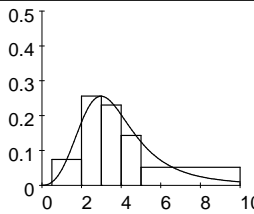
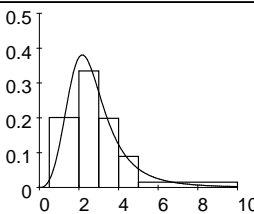
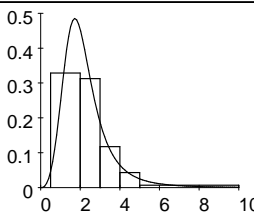
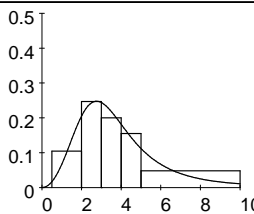
$$\hat{\tau}^{AP} = \begin{pmatrix} \hat{\tau}_{11}^{AP} \\ \hat{\tau}_{12}^{AP} \\ \hat{\tau}_{13}^{AP} \\ \hat{\tau}_{21}^{AP} \\ \hat{\tau}_{22}^{AP} \\ \hat{\tau}_{23}^{AP} \\ \hat{\tau}_{31}^{AP} \\ \hat{\tau}_{32}^{AP} \\ \hat{\tau}_{33}^{AP} \\ \hat{\tau}_{41}^{AP} \\ \hat{\tau}_{42}^{AP} \\ \hat{\tau}_{43}^{AP} \end{pmatrix}. \quad (8.46)$$

The asymptotic standard errors for $\hat{\tau}$ are calculated from

$$\text{Cov}(\hat{\tau}) \cong \mathbf{S} \text{Cov}(\hat{\lambda}) \mathbf{S}' . \quad (8.47)$$

A complete summary of all the effects ($\hat{\tau}$) with standard errors ($\hat{\sigma}_{\hat{\tau}}$) is given in Table 8.5. The overall median is R289. Investigating the main effects a decreasing monotone trend in PREMIUM over the categories of AGE is evident. Starting with a premium of R25 above the overall median for the youngest policyholders and dropping down to a premium of R22 below the overall median for the oldest policyholders. PRODUCT III is the most expensive product with a PREMIUM of R63 above the overall median. The premiums for PRODUCT I and PRODUCT II are both below average with premiums of R22 and R41 below the overall median respectively. The interaction effects, i.e. the $\hat{\tau}^{AP}$ -values, show a very clear interaction structure between AGE and PRODUCT. Apart from the overall decreasing effect in the PREMIUM over the categories of AGE, the PREMIUM drops even more drastically over the AGE categories for PRODUCT II. This is contrasted with PRODUCT III, which is a relatively cheaper policy for the younger policyholders. All the standard errors are included which enable the testing of certain hypotheses and the construction of confidence intervals.

Table 8.6: Effects for the saturated model.

AGE	PRODUCT			$\hat{\tau}^A$ $\hat{\sigma}_{\hat{\tau}^A}$
	I	II	III	
20-29	 $\hat{\tau}^{AP} = -0.08$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.038$	 $\hat{\tau}^{AP} = 0.41$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.045$	 $\hat{\tau}^{AP} = -0.32$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.053$	0.25 0.032
30-39	 $\hat{\tau}^{AP} = -0.01$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.034$	 $\hat{\tau}^{AP} = 0.00$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.034$	 $\hat{\tau}^{AP} = 0.02$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.044$	0.06 0.027
40-49	 $\hat{\tau}^{AP} = 0.00$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.036$	 $\hat{\tau}^{AP} = -0.16$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.036$	 $\hat{\tau}^{AP} = 0.16$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.047$	-0.09 0.028
50-59	 $\hat{\tau}^{AP} = 0.10$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.044$	 $\hat{\tau}^{AP} = -0.24$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.043$	 $\hat{\tau}^{AP} = 0.15$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.055$	-0.22 0.034
$\hat{\tau}^P$ $\hat{\sigma}_{\hat{\tau}^P}$	-0.22 0.022	-0.41 0.023	0.63 0.029	$\hat{\tau}_0 = 2.89$ $\hat{\sigma}_{\hat{\tau}_0} = 0.018$

8.3.2 Model 2: No interaction model

In the case of no interaction between AGE and PRODUCT the two-factor model is

$$\nu_{ij} = \tau_0 + \tau_i^A + \tau_j^P \quad , \quad i = 1, 2, 3, 4 \quad \text{and} \quad j = 1, 2, 3. \quad (8.48)$$

In matrix notation the medians are to be fitted such that

$$\begin{aligned} \boldsymbol{\nu} &= \mathbf{Y}_2 \boldsymbol{\gamma}_2 \\ &= \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 & -1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{pmatrix} \\ &= \left(\mathbf{1} \quad \mathbf{Z}_A \quad \mathbf{Z}_P \right) \begin{pmatrix} \tau_0 \\ \boldsymbol{\lambda}^A \\ \boldsymbol{\lambda}^P \end{pmatrix} \end{aligned} \quad (8.49)$$

where

$$\begin{aligned} \tau_0 = \gamma_1 & \quad : \quad \text{overall median} \\ \boldsymbol{\lambda}^A = \begin{pmatrix} \tau_1^A \\ \tau_2^A \\ \tau_3^A \end{pmatrix} = \begin{pmatrix} \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} & \quad : \quad \text{effects for AGE} \\ \boldsymbol{\lambda}^P = \begin{pmatrix} \tau_1^P \\ \tau_2^P \end{pmatrix} = \begin{pmatrix} \gamma_5 \\ \gamma_6 \end{pmatrix} & \quad : \quad \text{effects for PRODUCT} \end{aligned}$$

See Table 8.5 for an explanation of the submatrices $\mathbf{1}$, \mathbf{Z}_A and \mathbf{Z}_P in (8.49).

It follows that $\boldsymbol{\pi}$ is to be estimated such that $\boldsymbol{\nu}$ is in the column space of \mathbf{Y}_2 and therefore implies the constraints

$$\mathbf{g}_{\text{mod2}}(\boldsymbol{\pi}) = \mathbf{Q}_{Y_2}\boldsymbol{\nu} = \mathbf{0} \quad (8.50)$$

where $\mathbf{Q}_{Y_2} = \mathbf{I} - \mathbf{Y}_2(\mathbf{Y}_2'\mathbf{Y}_2)^{-1}\mathbf{Y}_2'$.

Note: The vector of constraints

$$\mathbf{g}_{\text{mod2}}(\boldsymbol{\pi}) = \mathbf{Z}'_{AP}\boldsymbol{\nu} = \mathbf{0} \quad (8.51)$$

with \mathbf{Z}_{AP} also defined in Table 8.5 is simply a reformulation of (8.50) and will provide exactly the same results. This follows since the columns of \mathbf{Z}_{AP} generate the orthogonal vector space of \mathbf{Y}_2 or simply because the model is to be fitted such that all the interaction effects in $\boldsymbol{\lambda}^{AP}$ (see Table 8.5) are zero.

The no interaction model is obtained by employing the vector of constraints

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_{\log}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{mod2}}(\boldsymbol{\pi}) \end{pmatrix} = \mathbf{0}$$

in the ML estimation procedure. The ML estimate of $\boldsymbol{\gamma}_2$ in (8.49) is

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Y}_2'\mathbf{Y}_2)^{-1}\mathbf{Y}_2'\hat{\boldsymbol{\nu}} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \\ \hat{\gamma}_5 \\ \hat{\gamma}_6 \end{pmatrix} = \begin{pmatrix} \hat{\tau}_0 \\ \hat{\tau}_1^A \\ \hat{\tau}_2^A \\ \hat{\tau}_3^A \\ \hat{\tau}_1^P \\ \hat{\tau}_2^P \end{pmatrix} = \begin{pmatrix} 2.8775 \\ 0.2879 \\ 0.0761 \\ -0.1160 \\ -0.2305 \\ -0.4380 \end{pmatrix}$$

containing the effects for the no interaction model.

The fitted log-logistic curves under the constraints of no interaction between AGE and PRODUCT are displayed in Table 8.8 and Table 8.9. In Table 8.8 the estimated medians proportionately reflect the row and column effects tabulated in Table 8.9. The strong negative linear trend in PREMIUM over

the AGE categories is evident, with PRODUCT III the most expensive product. All the interaction effects in Table 8.9 are now equal to zero.

From the goodness of fit statistics tabulated in Table 8.7, Model 2 shows a substantial drop in fit from that of Model 1. (See Table 8.4.) This is due to the clear interaction pattern seen in Model 1 where the saturated model was fitted. However, by calculating the measure of discrepancy the fit is still satisfactory, since $\mathbf{D} = 0.015 < 0.05$.

Table 8.7: Goodness of fit statistics for no interaction model.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
2	30	124.8	<0.0001	125.7	<0.0001

The degrees of freedom for Model 2 is 30, since an additional 6 linear independent constraints are included in $\mathbf{g}_{\text{mod}2}(\boldsymbol{\pi}) = \mathbf{0}$. See (8.50) and (8.51).

Table 8.8: Descriptive statistics for the no interaction model.

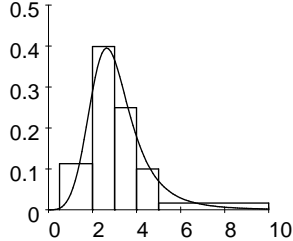
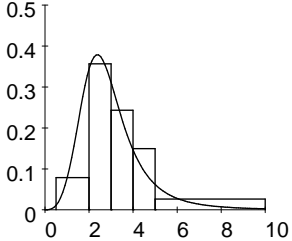
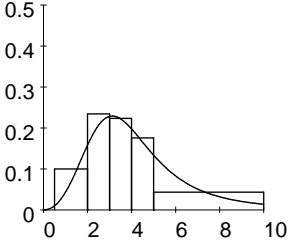
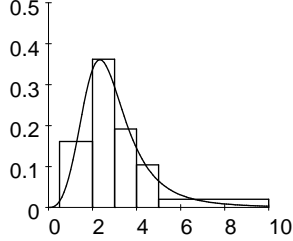
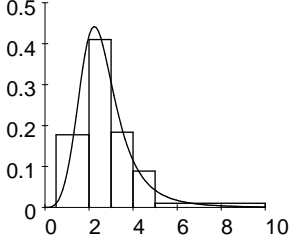
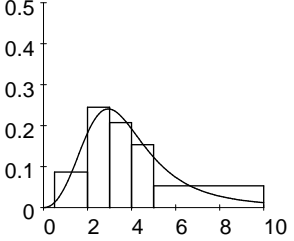
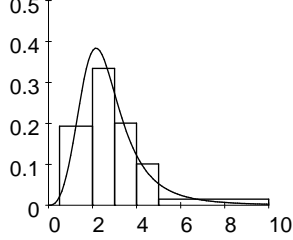
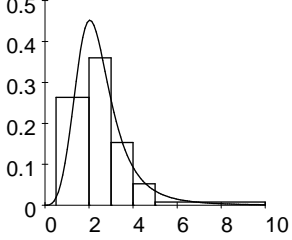
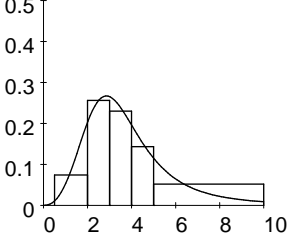
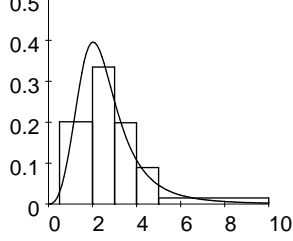
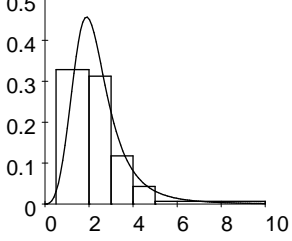
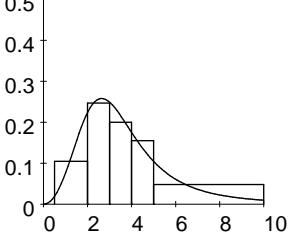
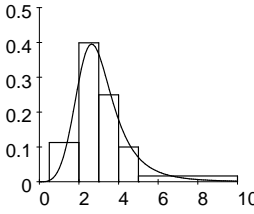
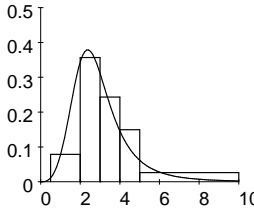
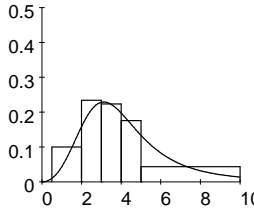
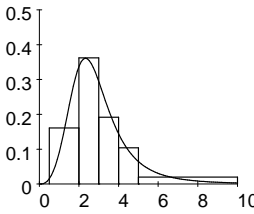
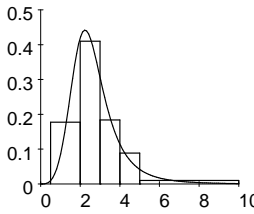
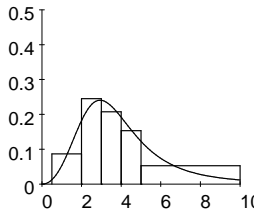
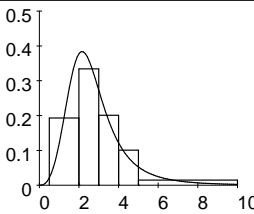
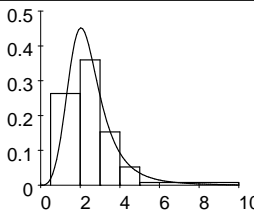
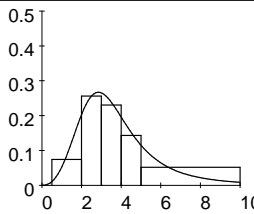
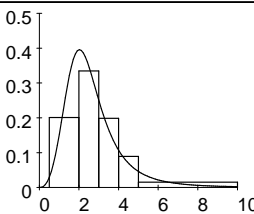
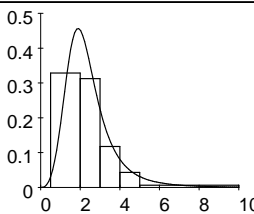
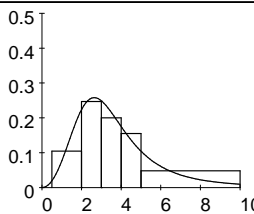
AGE	PRODUCT		
	P1	P2	P3
20-29	 $\hat{\mu} = 3.199$ $\hat{\nu} = 2.935$ $\hat{\sigma} = 1.476$ $\hat{\sigma}_{\hat{\nu}} = 0.033$	 $\hat{\mu} = 3.053$ $\hat{\nu} = 2.727$ $\hat{\sigma} = 1.673$ $\hat{\sigma}_{\hat{\nu}} = 0.039$	 $\hat{\mu} = 4.539$ $\hat{\nu} = 3.834$ $\hat{\sigma} = 3.328$ $\hat{\sigma}_{\hat{\nu}} = 0.050$
30-39	 $\hat{\mu} = 3.092$ $\hat{\nu} = 2.723$ $\hat{\sigma} = 1.837$ $\hat{\sigma}_{\hat{\nu}} = 0.031$	 $\hat{\mu} = 2.766$ $\hat{\nu} = 2.516$ $\hat{\sigma} = 1.357$ $\hat{\sigma}_{\hat{\nu}} = 0.031$	 $\hat{\mu} = 4.302$ $\hat{\nu} = 3.622$ $\hat{\sigma} = 3.204$ $\hat{\sigma}_{\hat{\nu}} = 0.043$
40-49	 $\hat{\mu} = 2.885$ $\hat{\nu} = 2.531$ $\hat{\sigma} = 1.749$ $\hat{\sigma}_{\hat{\nu}} = 0.033$	 $\hat{\mu} = 2.590$ $\hat{\nu} = 2.323$ $\hat{\sigma} = 1.384$ $\hat{\sigma}_{\hat{\nu}} = 0.031$	 $\hat{\mu} = 3.991$ $\hat{\nu} = 3.430$ $\hat{\sigma} = 2.691$ $\hat{\sigma}_{\hat{\nu}} = 0.045$
50-59	 $\hat{\mu} = 2.755$ $\hat{\nu} = 2.399$ $\hat{\sigma} = 1.740$ $\hat{\sigma}_{\hat{\nu}} = 0.041$	 $\hat{\mu} = 2.475$ $\hat{\nu} = 2.192$ $\hat{\sigma} = 1.427$ $\hat{\sigma}_{\hat{\nu}} = 0.040$	 $\hat{\mu} = 3.962$ $\hat{\nu} = 3.298$ $\hat{\sigma} = 3.109$ $\hat{\sigma}_{\hat{\nu}} = 0.050$

Table 8.9: Effects for the no interaction model.

AGE	PRODUCT			$\hat{\tau}^P$ $\hat{\sigma}_{\hat{\tau}^P}$
	I	II	III	
20-29	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	0.29 0.028
30-39	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	0.08 0.024
40-49	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	-0.12 0.025
50-59	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	-0.25 0.030
$\hat{\tau}^A$ $\hat{\sigma}_{\hat{\tau}^A}$	-0.23 0.021	-0.44 0.021	0.67 0.028	$\hat{\tau}_0 = 2.88$ $\hat{\sigma}_{\hat{\tau}_0} = 0.017$

8.3.3 Model 3: Regression model with no interaction

The decreasing monotone trend in PREMIUM over the categories of AGE can be modeled more effectively by incorporating AGE as a so-called covariate. Instead of the 3 dummy variables used in

$$\mathbf{Z}_A = \mathbf{D}_A \otimes \mathbf{1}_P = \left(\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} : 12 \times 3$$

the effect of AGE on PREMIUM can be modeled with the single covariate

$$\tilde{\mathbf{z}}^A = \mathbf{z}^A \otimes \mathbf{1}_P = \left(\left(\begin{pmatrix} 24.5 \\ 34.5 \\ 44.5 \\ 54.5 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 24.5 \\ 24.5 \\ 24.5 \\ 34.5 \\ 34.5 \\ 34.5 \\ 44.5 \\ 44.5 \\ 44.5 \\ 44.5 \\ 54.5 \\ 54.5 \\ 54.5 \end{pmatrix} : 12 \times 1 \quad (8.52)$$

where $\mathbf{z}^A = \begin{pmatrix} 24.5 & 34.5 & 44.5 & 54.5 \end{pmatrix}'$ represents the vector of class midpoints for AGE.

The model to be fitted is

$$\begin{aligned}
 \boldsymbol{\nu} &= \mathbf{Y}_3 \boldsymbol{\gamma}_3 \\
 &= \begin{pmatrix} 1 & 24.5 & 1 & 0 \\ 1 & 24.5 & 0 & 1 \\ 1 & 24.5 & -1 & -1 \\ 1 & 34.5 & 1 & 0 \\ 1 & 34.5 & 0 & 1 \\ 1 & 34.5 & -1 & -1 \\ 1 & 44.5 & 1 & 0 \\ 1 & 44.5 & 0 & 1 \\ 1 & 44.5 & -1 & -1 \\ 1 & 54.5 & 1 & 0 \\ 1 & 54.5 & 0 & 1 \\ 1 & 54.5 & -1 & -1 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{1} & \tilde{\mathbf{z}}^A & \mathbf{Z}_P \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} \tag{8.53}
 \end{aligned}$$

Model (8.53) implies

$$\mathbf{g}_{\text{mod3}}(\boldsymbol{\pi}) = \mathbf{Q}_{Y_3} \boldsymbol{\nu} = \mathbf{0}$$

to be implemented in the vector of constraints

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_{\log}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{mod3}}(\boldsymbol{\pi}) \end{pmatrix} \tag{8.54}$$

where $\mathbf{Q}_{Y_3} = \mathbf{I} - \mathbf{Y}_3 (\mathbf{Y}_3' \mathbf{Y}_3)^{-1} \mathbf{Y}_3'$. Since $\text{rank}(\mathbf{Y}_3) = 4$ a total of 8 linear independent constraints are included in $\mathbf{g}_{\text{mod3}}(\boldsymbol{\pi}) = \mathbf{0}$. The total number of linear independent constraints in (8.54) are equal to 32.

After employing the ML estimation procedure the restricted ML estimate $\hat{\pi}$, yields the ML estimate

$$\hat{\gamma}_3 = (\mathbf{Y}'_3 \mathbf{Y}_3)^{-1} \mathbf{Y}'_3 \hat{\nu} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \end{pmatrix} = \begin{pmatrix} 3.5897 \\ -0.01817 \\ -0.22774 \\ -0.44003 \end{pmatrix}. \quad (8.55)$$

It follows from (8.55) that the effects for Product II and Product III are

$$\hat{\tau}_1^P = \hat{\gamma}_3 = -0.22774 \quad \text{and} \quad \hat{\tau}_2^P = \hat{\gamma}_4 = -0.44003 \quad (8.56)$$

respectively and hence the effect for Product III is

$$\hat{\tau}_3^P = -(\hat{\gamma}_3 + \hat{\gamma}_4) = -(-0.22774 - 0.44003) = 0.66777 \quad (8.57)$$

meaning that the estimated median for Product III is R66.78 above the overall median.

The estimated two-factor model is

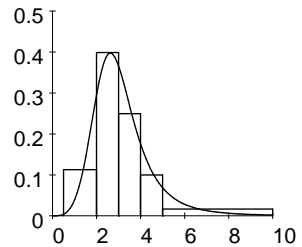
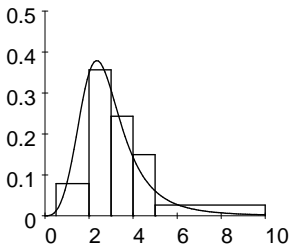
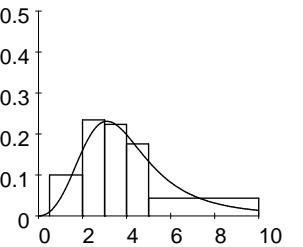
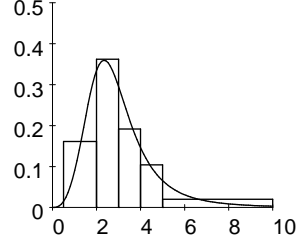
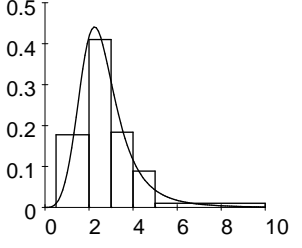
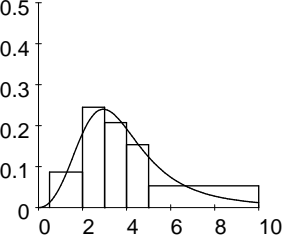
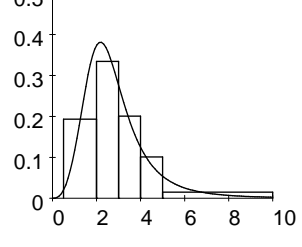
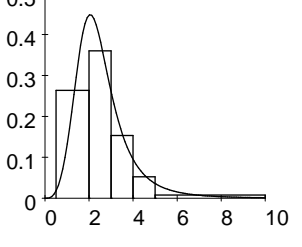
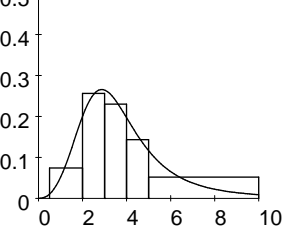
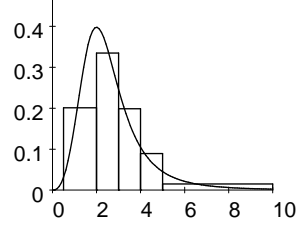
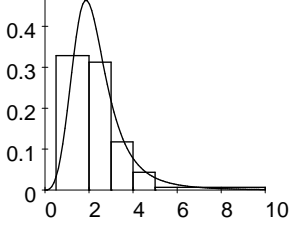
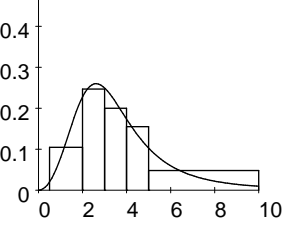
$$\hat{\nu}_{ij} = (3.5897 + \hat{\tau}_j^P) - 0.01817z_i^A, \quad i = 1, 2, 3, 4 \text{ and } j = 1, 2, 3 \quad (8.58)$$

where

$$\begin{aligned} \hat{\nu}_{ij} &= \text{estimated premium in the } ij\text{-th category} \\ z_i^A &= \text{the class midpoint for the } i\text{-th category for AGE} \\ \hat{\tau}_j^P &= \text{effect for the } j\text{-th category for PRODUCT} \end{aligned}$$

According to (8.58) the PREMIUM drops with R1.82 per year, or equivalently the PREMIUM drops with R18.17 per age category of 10 years. This rate of change in PREMIUM over AGE is the same for all three products, since no interaction between AGE and PRODUCT was assumed. See the estimated medians in Table 8.10.

Table 8.10: Descriptive statistics for no interaction regression model.

AGE	PRODUCT		
	I	II	III
20-29	 $\hat{\mu} = 3.179$ $\hat{\nu} = 2.917$ $\hat{\sigma} = 1.465$ $\hat{\sigma}_{\hat{\nu}} = 0.028$	 $\hat{\mu} = 3.034$ $\hat{\nu} = 2.705$ $\hat{\sigma} = 1.682$ $\hat{\sigma}_{\hat{\nu}} = 0.033$	 $\hat{\mu} = 4.508$ $\hat{\nu} = 3.812$ $\hat{\sigma} = 3.286$ $\hat{\sigma}_{\hat{\nu}} = 0.045$
30-39	 $\hat{\mu} = 3.106$ $\hat{\nu} = 2.735$ $\hat{\sigma} = 1.844$ $\hat{\sigma}_{\hat{\nu}} = 0.022$	 $\hat{\mu} = 2.774$ $\hat{\nu} = 2.523$ $\hat{\sigma} = 1.359$ $\hat{\sigma}_{\hat{\nu}} = 0.024$	 $\hat{\mu} = 4.313$ $\hat{\nu} = 3.631$ $\hat{\sigma} = 3.214$ $\hat{\sigma}_{\hat{\nu}} = 0.039$
40-49	 $\hat{\mu} = 2.908$ $\hat{\nu} = 2.554$ $\hat{\sigma} = 1.756$ $\hat{\sigma}_{\hat{\nu}} = 0.025$	 $\hat{\mu} = 2.610$ $\hat{\nu} = 2.341$ $\hat{\sigma} = 1.394$ $\hat{\sigma}_{\hat{\nu}} = 0.024$	 $\hat{\mu} = 4.012$ $\hat{\nu} = 3.449$ $\hat{\sigma} = 2.701$ $\hat{\sigma}_{\hat{\nu}} = 0.039$
50-59	 $\hat{\mu} = 2.730$ $\hat{\nu} = 2.372$ $\hat{\sigma} = 1.743$ $\hat{\sigma}_{\hat{\nu}} = 0.035$	 $\hat{\mu} = 2.439$ $\hat{\nu} = 2.160$ $\hat{\sigma} = 1.405$ $\hat{\sigma}_{\hat{\nu}} = 0.031$	 $\hat{\mu} = 3.928$ $\hat{\nu} = 3.267$ $\hat{\sigma} = 3.093$ $\hat{\sigma}_{\hat{\nu}} = 0.044$

From (8.58) the regression lines for each of the three products may be constructed. These regression lines have the same slope with different intercepts and are tabulated in Table 8.11. The regression lines reported in Table 8.11 agree with the estimated medians reported in Table 8.10.

Table 8.11: Estimated regression lines for regression model with no interaction.

PRODUCT	$\hat{\nu}_{ij}$
I	$3.36196 - 0.01817z_i^A$
II	$3.14967 - 0.01817z_i^A$
III	$4.25747 - 0.01817z_i^A$

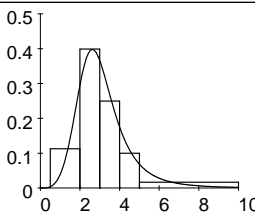
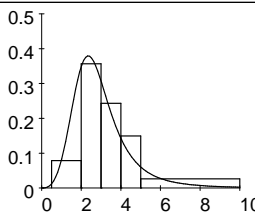
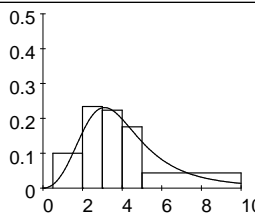
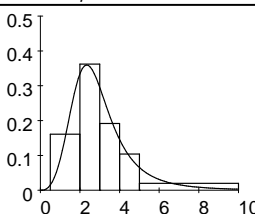
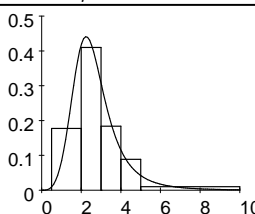
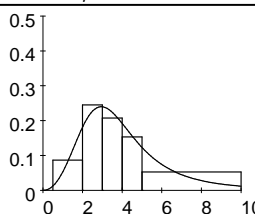
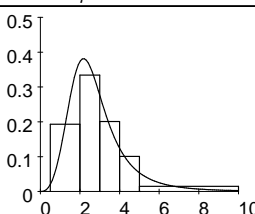
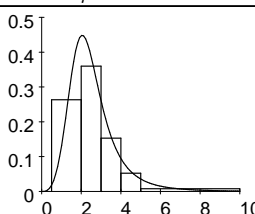
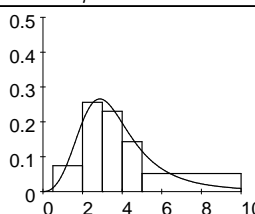
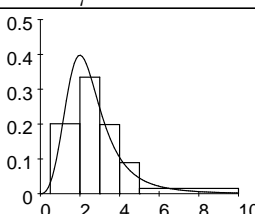
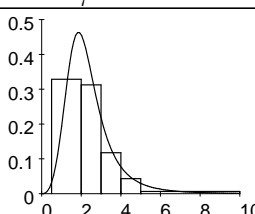
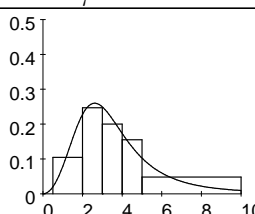
In Table 8.13 the effects for AGE reveal the same pattern as that of an ordinal variable. This follows since the distances between the class midpoints are equal. The effects of AGE show a constant drop of R18 per AGE category. Since all the interaction effects ($\hat{\tau}^{AP}$) are zero the medians in Table 8.10 proportionately reflect the row and column effects in Table 8.13.

Comparing the goodness of fit statistics of Model 3 (see Table 8.12) with that of Model 2 (see Table 8.7), the fit for the two models stayed practically the same. This motivates that the inclusion of AGE as a covariate in the model is doing practically just as good as the three dummy variables in the previous model, emphasizing the solid linear trend in PREMIUM over AGE.

Table 8.12: Goodness of fit statistics for regression model with no interaction.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
3	32	126.0	<0.0001	126.8	<0.0001

Table 8.13: Effects for no interaction regression model.

AGE	PRODUCT			$\hat{\tau}^A$
	I	II	III	$\hat{\sigma}_{\hat{\tau}^A}$
20-29	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	0.273 0.022
30-39	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	0.091 0.007
40-49	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	-0.091 0.007
50-59	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	 <p>$\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$</p>	-0.273 0.022
$\hat{\tau}^P$	-0.228	-0.440	0.668	2.872
$\hat{\sigma}_{\hat{\tau}^P}$	0.021	0.021	0.028	0.017

8.3.4 Model 4: Regression model with interaction

Since the PREMIUM of the three products do not change at the same rate over the categories of AGE, different slopes for each PRODUCT will be introduced leading to the model

$$\begin{aligned}
 \boldsymbol{\nu} &= \mathbf{Y}_4 \boldsymbol{\gamma}_4 \\
 &= \begin{pmatrix} 1 & 24.5 & 1 & 0 & 24.5 & 0 \\ 1 & 24.5 & 0 & 1 & 0 & 24.5 \\ 1 & 24.5 & -1 & -1 & -24.5 & -24.5 \\ 1 & 34.5 & 1 & 0 & 34.5 & 0 \\ 1 & 34.5 & 0 & 1 & 0 & 34.5 \\ 1 & 34.5 & -1 & -1 & -34.5 & -34.5 \\ 1 & 44.5 & 1 & 0 & 44.5 & 0 \\ 1 & 44.5 & 0 & 1 & 0 & 44.5 \\ 1 & 44.5 & -1 & -1 & -44.5 & -44.5 \\ 1 & 54.5 & 1 & 0 & 54.5 & 0 \\ 1 & 54.5 & 0 & 1 & 0 & 54.5 \\ 1 & 54.5 & -1 & -1 & -54.5 & -54.5 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{pmatrix} \\
 &= \left(\mathbf{1} \quad \tilde{\mathbf{z}}^A \quad \mathbf{Z}_P \quad (\tilde{\mathbf{z}}^A \odot \mathbf{Z}_P) \right) \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{pmatrix} \tag{8.59}
 \end{aligned}$$

where $\tilde{\mathbf{z}}^A$ is defined in (8.52) and \mathbf{Z}_P is previously defined in Table 8.5. The vector of constraints to be imposed in the ML estimation procedure is

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_{\log}(\boldsymbol{\pi}) \\ \mathbf{g}_{\text{mod4}}(\boldsymbol{\pi}) \end{pmatrix} = \mathbf{0} \tag{8.60}$$

where $\mathbf{g}_{\text{mod4}}(\boldsymbol{\pi}) = \mathbf{Q}_4 \boldsymbol{\nu}$ with $\mathbf{Q}_4 = \mathbf{I} - \mathbf{Y}_4 (\mathbf{Y}_4' \mathbf{Y}_4)^{-1} \mathbf{Y}_4'$ the projection matrix orthogonal to \mathbf{Y}_4 . A total of 6 linear independent constraints are included in $\mathbf{g}_{\text{mod4}}(\boldsymbol{\pi})$ bringing the total number of

linear independent constraints in $\mathbf{g}(\boldsymbol{\pi})$ to 30.

Employing the ML estimation procedure with the vector of constraints (8.60) the ML estimate for $\hat{\boldsymbol{\gamma}}_4$ is

$$\hat{\boldsymbol{\gamma}}_4 = (\mathbf{Y}'_4 \mathbf{Y}_4)^{-1} \mathbf{Y}'_4 \hat{\boldsymbol{\nu}} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \\ \hat{\gamma}_5 \\ \hat{\gamma}_6 \end{pmatrix} = \begin{pmatrix} 3.4879 \\ -0.01532 \\ -0.39227 \\ 0.33708 \\ 0.00447 \\ -0.01963 \end{pmatrix} \quad (8.61)$$

implying that the overall trend in PREMIUM over AGE is

$$\hat{\nu}_i = \hat{\gamma}_1 + \hat{\gamma}_2 z_i^A = 3.4879 - 0.01532 z_i^A. \quad (8.62)$$

Due to the interaction that exists between AGE and PRODUCT, the three regression equations for PREMIUM are as follows:

PRODUCT I:

$$\begin{aligned} \hat{\nu}_{i1} &= (3.4879 + \hat{\gamma}_3) + (-0.01532 + \hat{\gamma}_5) z_i^A \\ &= (3.4879 - 0.39227) + (-0.01532 + 0.00447) z_i^A \\ &= 3.0956 - 0.01085 z_i^A \end{aligned} \quad (8.63)$$

PRODUCT II:

$$\begin{aligned} \hat{\nu}_{i2} &= (3.4879 + \hat{\gamma}_4) + (-0.01532 + \hat{\gamma}_6) z_i^A \\ &= (3.4879 + 0.33708) + (-0.01532 - 0.01963) z_i^A \\ &= 3.8250 - 0.03496 z_i^A \end{aligned} \quad (8.64)$$

PRODUCT III: For PRODUCT III the effect on the overall intercept (8.62) is

$$-(\hat{\gamma}_3 + \hat{\gamma}_4) = -(-0.39227 + 0.33708) = 0.05519$$

and the effect on the overall slope (8.62) is

$$-(\hat{\gamma}_5 + \hat{\gamma}_6) = -(0.00447 - 0.01963) = 0.01516$$

leading to the regression line

$$\begin{aligned}\hat{\nu}_{i3} &= (3.4879 + 0.05519) + (-0.01532 + 0.01516) z_i^A \\ &= 3.5431 - 0.00016z_i^A\end{aligned}\tag{8.65}$$

See Table 8.16 where all the estimated medians are tabulated. For each product the estimated medians follow an unique trend over AGE. For PRODUCT I the premium drops with an estimated R1.09 per year, while for PRODUCT II the premium drops with an estimated R3.50 per year. For PRODUCT III no real trend over AGE is evident with a slope that is practically equal to zero.

Investigating the effects in Table 8.17, the marginal and the partial trend over AGE may be examined. Overall, the PREMIUM starts with R23 above the overall median of R288.30 and drops down linearly, with an estimated R15.30 per age category, to R23 below the overall median. It is interesting to note that this overall drop in PREMIUM seen by the $\hat{\tau}^A$ -values is cancelled out by the interaction effects for PRODUCT III, the $\hat{\tau}^{AP}$ -values, implying no trend over AGE for PRODUCT III. For PRODUCT II the effect of AGE on PREMIUM is rather drastic. Starting with R29.40 above the marginal effect for the youngest policy holders and dropping to R29.40 below the marginal effects for the oldest policy holders.

According to Table 8.15 the fit of Model 4 is much better than that of Model 5 indicating different trends in PREMIUM over AGE for the three products. This satisfactory fit further explained in Table 8.18 where the observed and expected frequencies are reported.

Table 8.15: Goodness of fit statistics for regression model with interaction.

		Pearson		Wald	
Model	df	Statistic	p-value	Statistic	p-value
4	30	49.9	0.0127	50.0	0.0122

Table 8.16: Descriptive statistics for regression model with interaction.

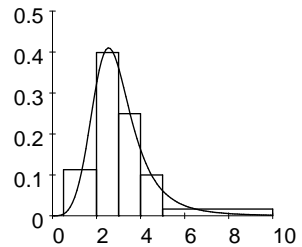
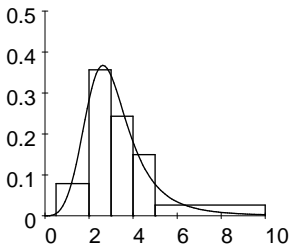
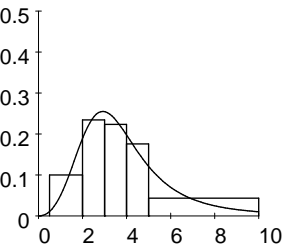
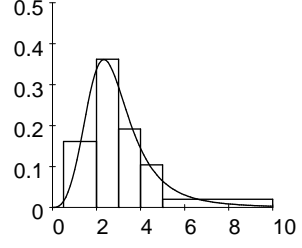
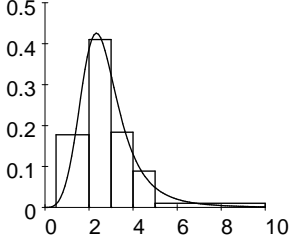
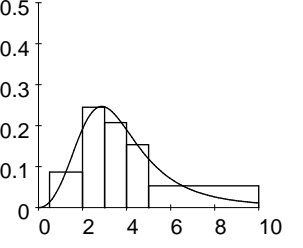
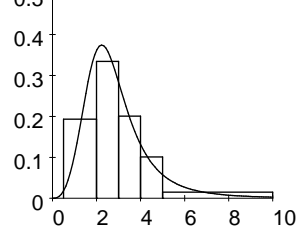
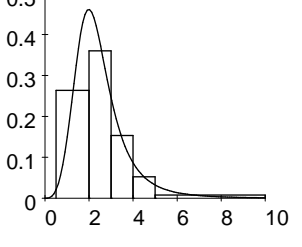
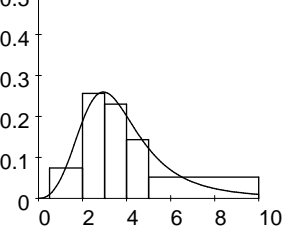
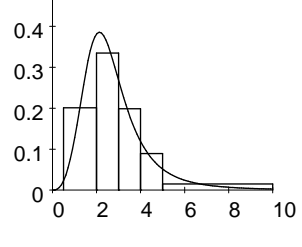
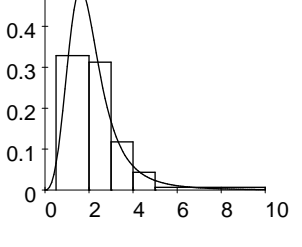
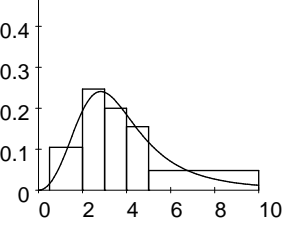
AGE	PRODUCT		
	P1	P2	P3
20-29	 $\hat{\mu} = 3.084$ $\hat{\nu} = 2.830$ $\hat{\sigma} = 1.422$ $\hat{\sigma}_{\hat{\nu}} = 0.033$	 $\hat{\mu} = 3.279$ $\hat{\nu} = 2.969$ $\hat{\sigma} = 1.654$ $\hat{\sigma}_{\hat{\nu}} = 0.046$	 $\hat{\mu} = 4.142$ $\hat{\nu} = 3.539$ $\hat{\sigma} = 2.874$ $\hat{\sigma}_{\hat{\nu}} = 0.068$
30-39	 $\hat{\mu} = 3.090$ $\hat{\nu} = 2.721$ $\hat{\sigma} = 1.836$ $\hat{\sigma}_{\hat{\nu}} = 0.022$	 $\hat{\mu} = 2.879$ $\hat{\nu} = 2.619$ $\hat{\sigma} = 1.406$ $\hat{\sigma}_{\hat{\nu}} = 0.027$	 $\hat{\mu} = 4.199$ $\hat{\nu} = 3.538$ $\hat{\sigma} = 3.117$ $\hat{\sigma}_{\hat{\nu}} = 0.042$
40-49	 $\hat{\mu} = 2.972$ $\hat{\nu} = 2.613$ $\hat{\sigma} = 1.782$ $\hat{\sigma}_{\hat{\nu}} = 0.028$	 $\hat{\mu} = 2.532$ $\hat{\nu} = 2.269$ $\hat{\sigma} = 1.359$ $\hat{\sigma}_{\hat{\nu}} = 0.026$	 $\hat{\mu} = 4.111$ $\hat{\nu} = 3.536$ $\hat{\sigma} = 2.761$ $\hat{\sigma}_{\hat{\nu}} = 0.044$
50-59	 $\hat{\mu} = 2.860$ $\hat{\nu} = 2.504$ $\hat{\sigma} = 1.753$ $\hat{\sigma}_{\hat{\nu}} = 0.045$	 $\hat{\mu} = 2.210$ $\hat{\nu} = 1.920$ $\hat{\sigma} = 1.411$ $\hat{\sigma}_{\hat{\nu}} = 0.043$	 $\hat{\mu} = 4.246$ $\hat{\nu} = 3.534$ $\hat{\sigma} = 3.332$ $\hat{\sigma}_{\hat{\nu}} = 0.072$

Table 8.17: Effects for regression model with interaction.

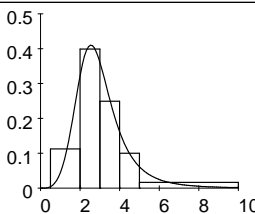
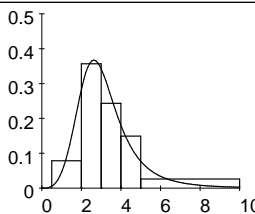
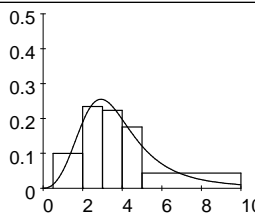
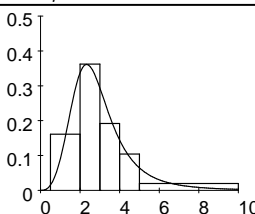
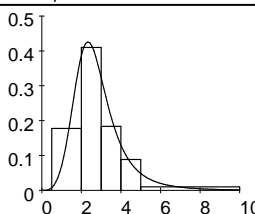
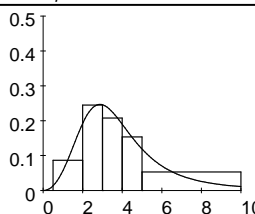
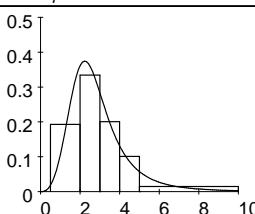
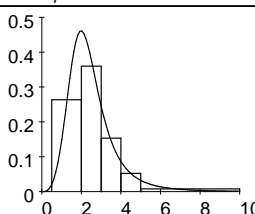
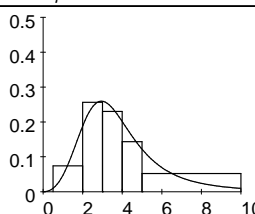
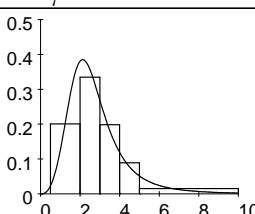
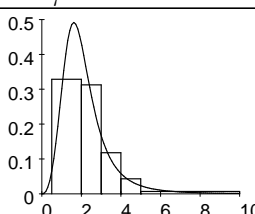
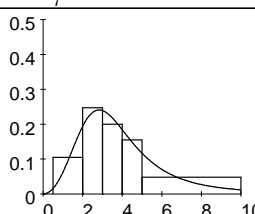
AGE	PRODUCT			$\hat{\tau}^A$ $\hat{\sigma}_{\hat{\tau}^A}$
	I	II	III	
20-29	 $\hat{\tau}^{AP} = -0.067$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.031$	 $\hat{\tau}^{AP} = 0.294$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.034$	 $\hat{\tau}^{AP} = -0.227$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.042$	0.230 0.026
30-39	 $\hat{\tau}^{AP} = -0.022$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.010$	 $\hat{\tau}^{AP} = 0.098$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.011$	 $\hat{\tau}^{AP} = -0.076$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.014$	0.077 0.009
40-49	 $\hat{\tau}^{AP} = 0.022$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.010$	 $\hat{\tau}^{AP} = -0.098$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.011$	 $\hat{\tau}^{AP} = 0.076$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.014$	-0.077 0.009
50-59	 $\hat{\tau}^{AP} = 0.067$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.031$	 $\hat{\tau}^{AP} = -0.294$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.034$	 $\hat{\tau}^{AP} = 0.227$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0.042$	-0.230 0.026
$\hat{\tau}_P$	-0.216	-0.438	0.654	2.883
$\hat{\sigma}_{\hat{\tau}_P}$	0.021	0.022	0.028	0.017

Table 8.18: Observed and expected frequencies for regression model with interaction.

AGE	PRODUCT								
	I			II			III		
20-29	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>
	R51-R200	157	166	R51-R200	49	68	R51-R200	69	61
	R201-R300	371	359	R201-R300	148	144	R201-R300	108	108
	R301-R400	232	239	R301-R400	101	109	R301-R400	103	107
	R401-R500	93	96	R401-R500	62	51	R401-R500	81	73
	R500+	77	70	R500+	55	44	R500+	100	112
30-39	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>
	R51-R200	267	271	R51-R200	213	194	R51-R200	132	145
	R201-R300	400	378	R201-R300	328	317	R201-R300	249	235
	R301-R400	212	237	R301-R400	147	174	R301-R400	211	226
	R401-R500	115	109	R401-R500	71	66	R401-R500	156	155
	R500+	111	109	R500+	41	49	R500+	269	256
40-49	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>
	R51-R200	241	229	R51-R200	302	289	R51-R200	73	84
	R201-R300	278	289	R201-R300	275	283	R201-R300	168	155
	R301-R400	167	167	R301-R400	117	117	R301-R400	151	155
	R401-R500	84	74	R401-R500	40	42	R401-R500	94	105
	R500+	62	72	R500+	30	33	R500+	170	156
50-59	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>	Premium	<i>f</i>	<i>m</i>
	R51-R200	135	139	R51-R200	205	223	R51-R200	77	73
	R201-R300	150	155	R201-R300	130	120	R201-R300	121	112
	R301-R400	89	83	R301-R400	49	43	R301-R400	98	106
	R401-R500	40	36	R401-R500	18	16	R401-R500	76	73
	R500+	34	35	R500+	14	15	R500+	118	126

Part III

Bivariate normal distribution

Chapter 9

Bivariate grouped data

Consider a bivariate data set with n observations classified in a two-way contingency table with I rows and J columns. The frequencies of the IJ cells are denoted by f_{ij} in Table 9.1.

Table 9.1: Contingency table with I rows and J columns.

Y	X				
	$(-\infty, y_1]$	$(y_1, y_2]$	\cdots	$(y_{J-2}, y_{J-1}]$	$(y_{J-1}, y_J]$
$(-\infty, x_1]$	f_{11}	f_{12}	\cdots	$f_{1,J-1}$	f_{1J}
$(x_1, x_2]$	f_{21}	f_{22}	\cdots	$f_{2,J-1}$	f_{2J}
\vdots	\vdots	\vdots	\cdots	\vdots	\vdots
$(x_{I-2}, x_{I-1}]$	$f_{I-1,1}$	$f_{I-1,2}$	\cdots	$f_{I-1,J-1}$	$f_{I-1,J}$
$(x_{I-1}, x_I]$	f_{I1}	f_{I2}	\cdots	$f_{I,J-1}$	f_{IJ}

The objective is to fit a bivariate distribution curve to the two-way grouped data set in Table 9.1.

9.1 Formulation

The vectors of upper class boundaries are

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{I-1} \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{J-1} \end{pmatrix} \quad (9.1)$$

with

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1,J-1} & f_{1J} \\ f_{21} & f_{22} & \cdots & f_{2,J-1} & f_{2J} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ f_{I1} & f_{I2} & \cdots & f_{I-1,J-1} & f_{I-1,J} \\ f_{I1} & f_{I2} & \cdots & f_{I,J-1} & f_{IJ} \end{pmatrix} \quad (9.2)$$

the matrix of frequencies listed in Table 9.1.

Define

$$\mathbf{f} = \text{vec}(\mathbf{F}) \quad (9.3)$$

as the column vector where the elements of \mathbf{F} are stacked row by row below each other. It is assumed that \mathbf{f} has a multinomial distribution *i.e.*

$$\mathbf{f} \sim \text{mult}(n, \boldsymbol{\pi}_0) .$$

Let

$$\mathbf{p}_0 = \frac{1}{n} \mathbf{f} \quad (9.4)$$

denote the vector of relative frequencies. Hence

$$E(\mathbf{p}_0) = \boldsymbol{\pi}_0 \quad (9.5)$$

and

$$\begin{aligned} \text{Cov}(\mathbf{p}_0) &= \frac{1}{n} (\text{diag}(\boldsymbol{\pi}_0) - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0') \\ &= \mathbf{V}_0 . \end{aligned} \quad (9.6)$$

Define the matrix of relative frequencies

$$\mathbf{P}_0 = \frac{1}{n}\mathbf{F} \quad (9.7)$$

where \mathbf{F} is given in (9.2). The matrix with cumulative relative frequencies may be obtained from

$$\mathbf{P} = \mathbf{C}_I \cdot \mathbf{P}_0 \cdot \mathbf{C}_J \quad (9.8)$$

where

$$\mathbf{C}_I : (I \times I) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{C}_J : (J \times J) = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (9.9)$$

From *Muirhead (1982) (p.74)* it follows that

$$\begin{aligned} \text{vec}(\mathbf{P}) &= \text{vec}(\mathbf{C}_I \cdot \mathbf{P}_0 \cdot \mathbf{C}_J) \\ &= (\mathbf{C}'_J \otimes \mathbf{C}_I) \text{vec}(\mathbf{P}_0) \\ &= (\mathbf{C}'_J \otimes \mathbf{C}_I) \mathbf{p}_0 \end{aligned} \quad (9.10)$$

From (9.10) the random vector of cumulative relative frequencies is

$$\mathbf{p} = \mathbf{C}\mathbf{p}_0 \quad (9.11)$$

with

$$\mathbf{C} = (\mathbf{C}'_J \otimes \mathbf{C}_I). \quad (9.12)$$

The expected value and covariance matrix of the random vector \mathbf{p} is

$$\begin{aligned} E(\mathbf{p}) &= E(\mathbf{C}\mathbf{p}_0) \\ &= \mathbf{C}\boldsymbol{\pi}_0 \\ &= \boldsymbol{\pi} \end{aligned} \quad (9.13)$$

and

$$\begin{aligned}
 \text{Cov}(\mathbf{p}) &= \text{Cov}(\mathbf{C}\mathbf{p}_0) \\
 &= \frac{1}{n} \mathbf{C} \{ \text{diag}(\boldsymbol{\pi}_0) - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0' \} \mathbf{C}' \\
 &= \frac{1}{n} \{ \mathbf{C} \text{diag}(\mathbf{C}^{-1} \boldsymbol{\pi}) \mathbf{C}' - \boldsymbol{\pi} \boldsymbol{\pi}' \} \\
 &= \mathbf{V}.
 \end{aligned} \tag{9.14}$$

9.2 Estimation

Estimation of the bivariate distribution curve $F(x, y)$, is obtained such that

$$P(X \leq x_i, Y \leq y_j) = \pi_{ij} \tag{9.15}$$

for $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$ where π_{ij} is the expected cumulative relative frequency in (9.13). The complete set of expected cumulative relative frequencies is given in Table 9.2.

Table 9.2: Expected cumulative relative frequencies for a bivariate grouped data set.

X	Y				
	$(-\infty, y_1]$	$(y_1, y_2]$	\dots	$(y_{J-2}, y_{J-1}]$	$(y_{J-1}, y_J]$
$(-\infty, x_1]$	π_{11}	π_{12}	\dots	$\pi_{1,J-1}$	π_{1J}
$(x_1, x_2]$	π_{21}	π_{22}	\dots	$\pi_{2,J-1}$	π_{2J}
\vdots	\vdots	\vdots	\dots	\vdots	\vdots
$(x_{I-2}, x_{I-1}]$	$\pi_{I-1,1}$	$\pi_{I-1,2}$	\dots	$\pi_{I-1,J-1}$	$\pi_{I-1,J}$
$(x_{I-1}, x_I]$	π_{I1}	π_{I2}	\dots	$\pi_{I,J-1}$	π_{IJ}

Imposing the restriction (9.15) in the ML estimation procedure, leads to the ML estimate of $\boldsymbol{\pi}$ under constraints, that will satisfy the characteristics of the specified bivariate continuous distribution.

Chapter 10

The bivariate normal distribution

In this chapter a few of the basic concepts of the bivariate normal distribution will be discussed. These concepts are of importance in the estimation of the bivariate normal distribution to a two-way contingency table. It will also be shown how to calculate bivariate normal probabilities by making use of a series of gamma functions. The one-to-one relationship between the correlation coefficient and the bivariate normal probabilities is explained in detail since it plays a major role in the estimation of the bivariate normal distribution discussed in the next chapter.

10.1 Joint distribution

The bivariate normal distribution with pdf

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\} \quad (10.1)$$

where $-\infty < \mu_x, \mu_y < \infty$, $0 < \sigma_x, \sigma_y < \infty$ and $-1 \leq \rho \leq 1$ is to be fitted to the two-way contingency table in Table 9.1. The pdf of the bivariate normal distribution involves 5 parameters and a special notation for this joint distribution is

$$(x, y) \sim \text{BVN}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho) .$$

10.2 Marginal distributions

When x and y are jointly normally distributed, each of the two marginal distributions by itself is normally distributed. The marginal distribution of x is normal with mean μ_x and standard deviation σ_x , i.e.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \cdot \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu_x}{\sigma_x} \right)^2 \right\}. \quad (10.2)$$

The marginal distribution of y is normal with mean μ_y and standard deviation σ_y , i.e.

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \cdot \exp \left\{ -\frac{1}{2} \left(\frac{y - \mu_y}{\sigma_y} \right)^2 \right\}. \quad (10.3)$$

10.3 Standard bivariate normal distribution

By making use of standardisation it is possible to obtain the standard bivariate normal distribution

$$f(z_x, z_y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2] \right\} \quad (10.4)$$

where $z_x = \left(\frac{x - \mu_x}{\sigma_x} \right)$ and $z_y = \left(\frac{y - \mu_y}{\sigma_y} \right)$. In this case

$$(z_x, z_y) \sim \text{BVN}(0, 0, 1, 1, \rho)$$

with

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (10.5)$$

where $\sigma_{xy} = \text{Cov}(x, y)$, the only parameter determining the shape of the bivariate normal distribution.

The standard bivariate normal curve is displayed in Table 10.1 to illustrate the effect of the correlation coefficient ρ .

Table 10.1: The effect of the correlation coefficient ρ .

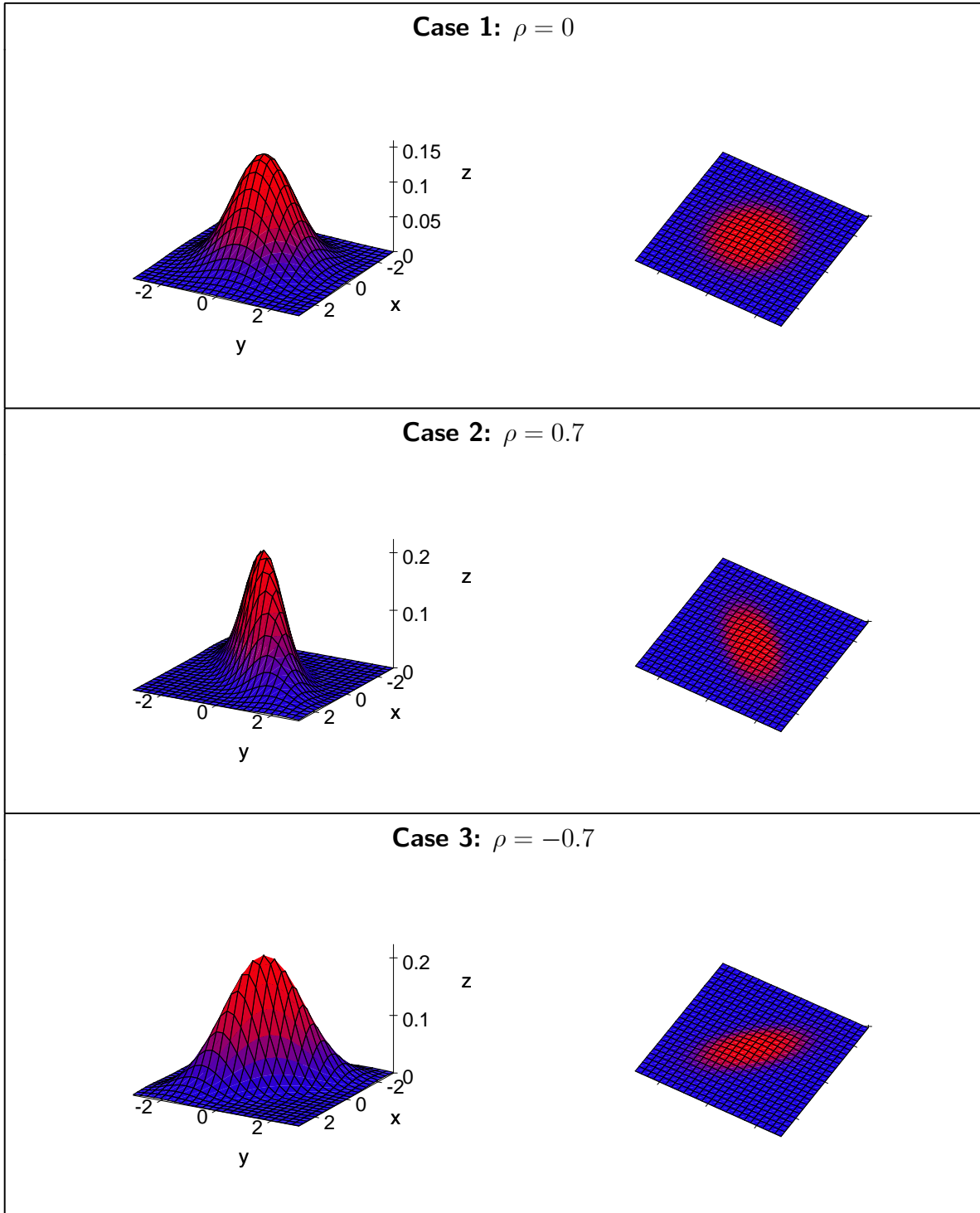


Table 10.1 is summarised as follows:

Case 1: $\rho = 0$

The contour curves are circles, indicating no relationship between x and y . For all other values of ρ the contour curves are ellipses.

Case 2: $\rho = 0.7$

When x and y are positively related so that $\rho > 0$, the principal axis has a positive slope, implying that the surface tends to run along a line with a positive slope. It is clear that high x values are related with high y values and visa versa.

Case 3: $\rho = -0.7$

When x and y are negatively related, $\rho < 0$, the principal axis has a negative slope and the surface runs along a line with a negative slope.

10.4 Conditional distributions

The density function of the conditional distribution of x for any given value of y is

$$f(x|y) = \frac{f(x, y)}{f(y)}$$

where $f(x, y)$ is the joint density function of x and y and $f(y)$ is the marginal density function of y . When x and y are jointly normally distributed the conditional pdf of x for any given y is

$$f(x|y) = \frac{1}{\sqrt{2\pi}\sigma_{x|y}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu_{x|y}}{\sigma_{x|y}} \right)^2 \right] \quad (10.6)$$

where

$$\begin{aligned} \mu_{x|y} &= \mu_x + \left(\rho \frac{\sigma_x}{\sigma_y} \right) (y - \mu_y) \\ \sigma_{x|y}^2 &= \sigma_x^2 (1 - \rho^2) \end{aligned}$$

The parameter $\alpha_{x|y} = \mu_x - \left(\rho \frac{\sigma_x}{\sigma_y} \right) \mu_y$ is the intercept of the line of regression of x on y and the parameter $\beta_{x|y} = \rho \frac{\sigma_x}{\sigma_y}$ is the slope of this line.

The conditional distribution of y for any given x follows similarly with

$$f(y|x) = \frac{1}{\sqrt{2\pi}\sigma_{y|x}} \exp \left[-\frac{1}{2} \left(\frac{y - \mu_{y|x}}{\sigma_{y|x}} \right)^2 \right] \quad (10.7)$$

where

$$\begin{aligned} \mu_{y|x} &= \mu_y + \left(\rho \frac{\sigma_y}{\sigma_x} \right) (x - \mu_x) \\ \sigma_{y|x}^2 &= \sigma_y^2 (1 - \rho^2) \end{aligned}$$

The parameter $\alpha_{y|x} = \mu_y - \left(\rho \frac{\sigma_y}{\sigma_x} \right) \mu_x$ is the intercept of the line of regression of y on x and the parameter $\beta_{y|x} = \rho \frac{\sigma_y}{\sigma_x}$ is the slope of this line.

10.5 Bivariate normal probabilities

10.5.1 Calculation of bivariate normal probabilities

The probability

$$\Phi(a, b; \rho) = \int_{-\infty}^b \int_{-\infty}^a \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left(-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2] \right) dz_x dz_y \quad (10.8)$$

corresponds to the volume under the surface of the standard bivariate normal distribution over the region $-\infty < z_x \leq a$ and $-\infty < z_y \leq b$. The lines $z_x = 0$ and $z_y = 0$ divide the domain in 4 so-called quadrants. See Table 10.2.

Table 10.2: The four quadrants of the bivariate normal distribution.

Quadrant	Region
Q_1	$-\infty < z_x < 0 \quad -\infty < z_y < 0$
Q_2	$-\infty < z_x < 0 \quad 0 \leq z_y < \infty$
Q_3	$0 \leq z_x < \infty \quad -\infty < z_y < 0$
Q_4	$0 \leq z_x < \infty \quad 0 \leq z_y < \infty$

Define

$$\Phi_0(a, b; \rho) = \int_0^b \int_0^a \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2]\right) dz_x dz_y \quad (10.9)$$

as the integral where integration of the standard bivariate normal distribution takes place in the positive quadrant, Q_4 . See Figure 10.1.

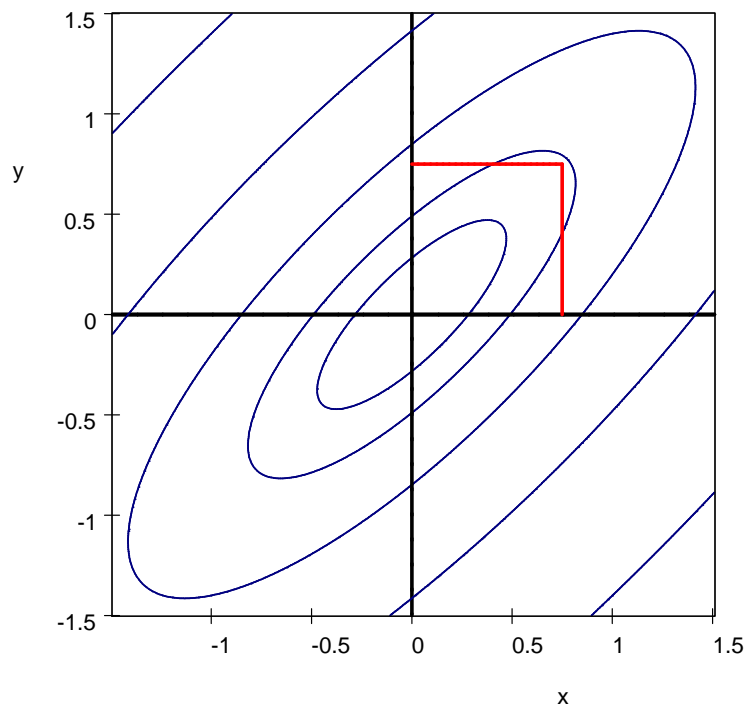
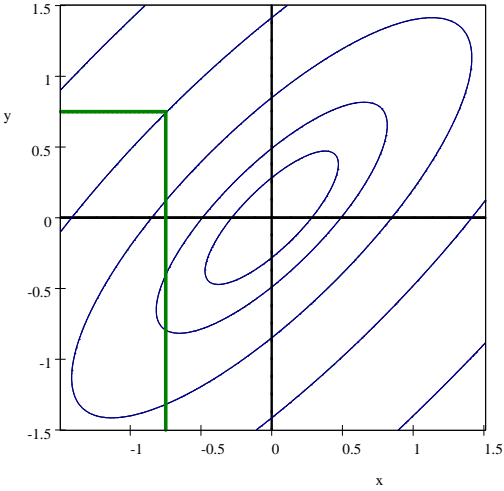
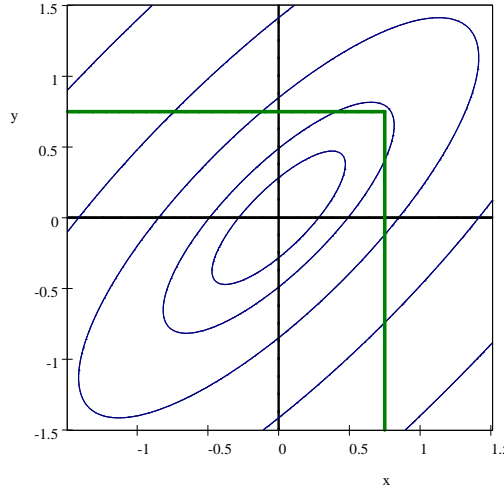
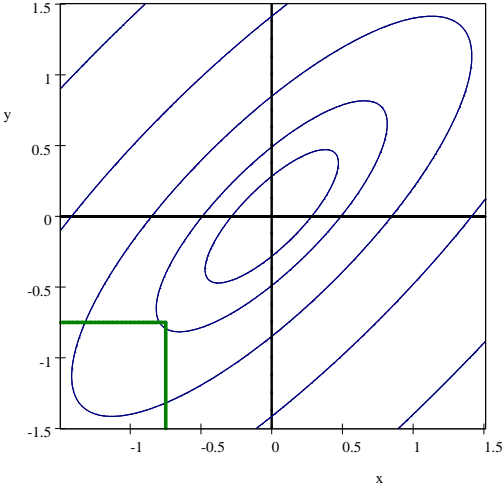
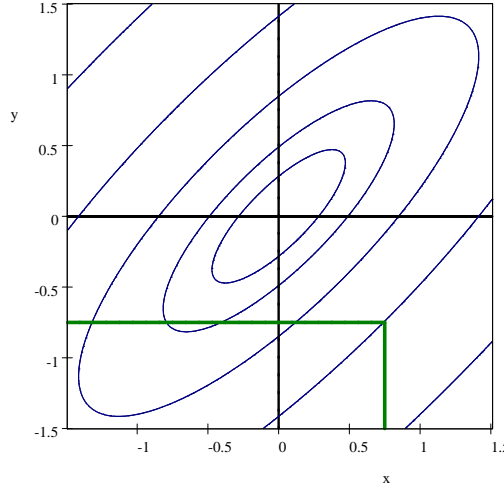


Figure 10.1: Integration region of $\Phi_0(a, b; \rho)$

Due to the symmetry of the bivariate normal distribution, any bivariate normal probability $\Phi(a, b; \rho)$ in (10.8) can be calculated as a linear combination of $\Phi_0(a, b; \rho)$ -values in (10.9), summarised in Table 10.3.

Table 10.3: Bivariate normal probabilities in the four quadrants.

<p>Quadrant 2: ($a < 0$ and $b \geq 0$)</p>  <p>$\Phi(a, b; \rho) = \Phi_0(\infty, \infty; \rho) - \Phi_0(-a, \infty; \rho)$ $+ \Phi_0(\infty, b; -\rho) - \Phi_0(-a, b; -\rho)$</p>	<p>Quadrant 4: ($a \geq 0$ and $b \geq 0$)</p>  <p>$\Phi(a, b; \rho) = \Phi_0(\infty, \infty; \rho) + \Phi_0(a, \infty; -\rho)$ $+ \Phi_0(\infty, b; -\rho) + \Phi_0(a, b; \rho)$</p>
<p>Quadrant 1: ($a < 0$ and $b < 0$)</p>  <p>$\Phi(a, b; \rho) = \Phi_0(\infty, \infty; \rho) - \Phi_0(-a, \infty; \rho)$ $- \Phi_0(\infty, -b; \rho) + \Phi_0(-a, -b; \rho)$</p>	<p>Quadrant 3: ($a \geq 0$ and $b < 0$)</p>  <p>$\Phi(a, b; \rho) = \Phi_0(\infty, \infty; \rho) + \Phi_0(a, \infty; -\rho)$ $- \Phi_0(\infty, -b; \rho) - \Phi_0(a, -b; -\rho)$</p>

In Algorithm 1 it will be shown how to calculate the bivariate normal probability in the positive Quadrant Q_4 , as a series of gamma functions.

Algorithm 1

$$\begin{aligned}\Phi_0(a, b; \rho) &= \int_0^b \int_0^a \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2]\right) dz_x dz_y \\ &= \sum_{i=0}^{\infty} \frac{2\rho\sqrt{1-\rho^2}}{4\pi i!} \Gamma^2\left(\frac{i+1}{2}\right) \cdot G\left(\frac{a^2}{2(1-\rho^2)}, \frac{i+1}{2}\right) \cdot G\left(\frac{b^2}{2(1-\rho^2)}, \frac{i+1}{2}\right) \\ &\text{for } a, b \geq 0\end{aligned}\tag{10.10}$$

where $G(x, \kappa) = \int_0^x \frac{1}{\Gamma(\kappa)} t^{\kappa-1} e^{-t} dt$ is the gamma distribution with shape parameter κ .

Proof. Since

$$\exp\left(\frac{\rho z_x z_y}{1-\rho^2}\right) = \sum_{i=0}^{\infty} \frac{\left(\frac{\rho z_x z_y}{1-\rho^2}\right)^i}{i!}\tag{10.11}$$

it follows that

$$\begin{aligned}\Phi_0(a, b; \rho) &= \int_0^b \int_0^a \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2]\right) dz_x dz_y \\ &= \sum_{i=0}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(\frac{\rho}{1-\rho^2}\right)^i \frac{1}{i!} \\ &\quad \cdot \int_0^a \exp\left(-\frac{z_x}{2(1-\rho^2)}\right) z_x^i dz_x \cdot \int_0^b \exp\left(-\frac{z_y}{2(1-\rho^2)}\right) z_y^i dz_y\end{aligned}$$

Substitution of $s = \frac{z^2}{2(1-\rho^2)}$ in $\int_0^c e^{-\frac{z^2}{2(1-\rho^2)}} z^i dz$ yields

$$\begin{aligned}\int_0^c e^{-\frac{z^2}{2(1-\rho^2)}} z^i dz &= \int_0^{\frac{c^2}{2(1-\rho^2)}} e^{-s} (2s)^{\frac{i-1}{2}} (1-\rho^2)^{\frac{i+1}{2}} ds \\ &= 2^{\frac{i-1}{2}} (1-\rho^2)^{\frac{i+1}{2}} \Gamma\left(\frac{i+1}{2}\right) \int_0^{\frac{c^2}{2(1-\rho^2)}} \frac{1}{\Gamma\left(\frac{i+1}{2}\right)} s^{\left(\frac{i+1}{2}-1\right)} e^{-s} ds \\ &= 2^{\frac{i-1}{2}} (1-\rho^2)^{\frac{i+1}{2}} \Gamma\left(\frac{i+1}{2}\right) \cdot G\left(\frac{c^2}{2(1-\rho^2)}, \frac{i+1}{2}\right)\end{aligned}\tag{10.12}$$

and therefore

$$\begin{aligned}\Phi_0(a, b; \rho) &= \sum_{i=0}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(\frac{\rho}{1-\rho^2}\right)^i \frac{1}{i!} \cdot 2^{i-1} (1-\rho^2)^{i+1} \Gamma^2\left(\frac{i+1}{2}\right) \\ &\quad \cdot G\left(\frac{a^2}{2(1-\rho^2)}, \frac{i+1}{2}\right) \cdot G\left(\frac{b^2}{2(1-\rho^2)}, \frac{i+1}{2}\right) \\ &= \sum_{i=0}^{\infty} \frac{(2\rho)^i \sqrt{1-\rho^2}}{4\pi i!} \Gamma^2\left(\frac{i+1}{2}\right) \cdot G\left(\frac{a^2}{2(1-\rho^2)}, \frac{i+1}{2}\right) \cdot G\left(\frac{b^2}{2(1-\rho^2)}, \frac{i+1}{2}\right)\end{aligned}$$

■

The probability $\Phi_0(a, b; \rho)$ can be calculated by making use of the SAS program *Phi0.SAS* listed in the Appendix. The probability $\Phi(a, b; \rho)$ can be obtained by making use of the SAS function `PROBBNRM(a,b, ρ)` or by making use of the SAS program *Phi.SAS* also listed in the Appendix.

10.5.2 Calculation of ρ

Integration over each of the four quadrants tabulated in Table 10.2 leads to the definition of the following four probabilities or so-called volumes

$$\text{VOL 1} = \iint_{Q_1} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2]\right\} dz_x dz_y \quad (10.13)$$

$$\text{VOL 2} = \iint_{Q_2} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2]\right\} dz_x dz_y \quad (10.14)$$

$$\text{VOL 3} = \iint_{Q_3} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2]\right\} dz_x dz_y \quad (10.15)$$

$$\text{VOL 4} = \iint_{Q_4} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)} [z_x^2 - 2\rho z_x z_y + z_y^2]\right\} dz_x dz_y \quad (10.16)$$

The probability or the total volume of the positive quadrant Q_4 may be expressed in terms of the correlation coefficient

$$\frac{\arcsin \rho}{2\pi} = \text{VOL 4} - \frac{1}{4} \quad (10.17)$$

which is referred to as *Sheppard's theorem on median dichotomy (1898)*. (See *Kendall and Stuart (1958) p.351*). Due to the symmetry of the bivariate normal distribution i.e.

$$\text{VOL 1} = \text{VOL 4} \quad \text{and} \quad \text{VOL 2} = \text{VOL 3}$$

and the property

$$\text{VOL 1} + \text{VOL 2} + \text{VOL 3} + \text{VOL 4} = 1$$

it follows that

$$\frac{2}{\pi} \arcsin \rho = (\text{VOL 1} + \text{VOL 4}) - (\text{VOL 2} + \text{VOL 3})$$

which leads to the expression of ρ

$$\rho = \sin \left(\frac{\pi}{2} [(\text{VOL 1} + \text{VOL 4}) - (\text{VOL 2} + \text{VOL 3})] \right) . \quad (10.18)$$

As an illustration of the one-to-one relationship between the volumes of the respective quadrants of the bivariate normal distribution and the correlation coefficient ρ consider Table 10.4 and Table 10.5.

Table 10.4: Relationship between ρ and the four volumes of the bivariate normal distribution.

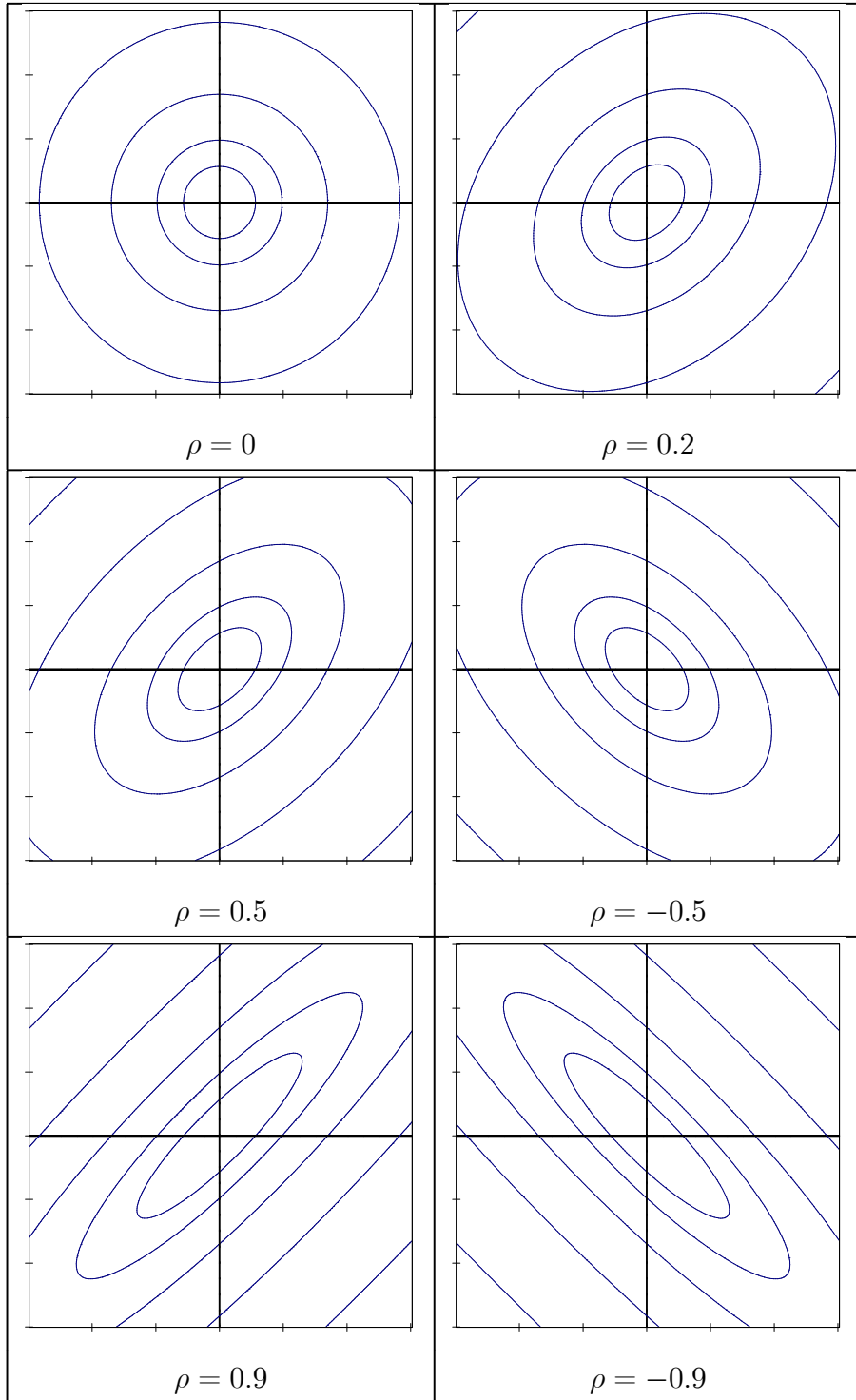
(VOL 1 + VOL 4)	(VOL 2 + VOL 3)	ρ
1	0	$\sin (\pi / 2) = 1$
0.9	0.1	$\sin (\pi / 2(0.8)) = 0.95106$
0.8	0.2	$\sin (\pi / 2(0.6)) = 0.80902$
0.7	0.3	$\sin (\pi / 2(0.4)) = 0.58779$
0.6	0.4	$\sin (\pi / 2(0.2)) = 0.30902$
0.5	0.5	$\sin (\pi / 2(0)) = 0.0$
0.4	0.6	$\sin (\pi / 2(-0.2)) = -0.30902$
0.3	0.7	$\sin (\pi / 2(-0.4)) = -0.58779$
0.2	0.8	$\sin (\pi / 2(-0.6)) = -0.80902$
0.1	0.9	$\sin (\pi / 2(-0.8)) = -0.95106$
0	1	$\sin (-\pi / 2) = -1$

In the case where $\rho = 0$

$$\text{VOL 1} = \text{VOL 2} = \text{VOL 3} = \text{VOL 4} = 0.25 ,$$

resulting in an even distribution of the volumes over the four quadrants.

Table 10.5: Contours of bivariate normal distribution with ρ .



For a slight positive relationship of $\rho = 0.2$ (see Table 10.5), the volumes of the positive and negative quadrants are slightly higher than the two mixed quadrants. Comparing $\rho = 0.5$ with $\rho = -0.5$ it is clear that the two graphs are mirror images of each other. Further, it is also clear that a stronger positive relationship is associated with higher volumes in the positive and negative quadrants, while a stronger negative relationship is associated with higher volumes in the two mixed quadrants. (See Table 10.5.)

Chapter 11

Estimating the bivariate normal distribution

In this chapter the estimation procedure to fit a bivariate normal distribution (10.1) to the two-way contingency table in Table 9.1 is described.

11.1 Bivariate normal probabilities

After standardising the vector of upper class boundaries \mathbf{x} in (9.1), the vector of standardised upper class boundaries is

$$\begin{aligned}
 \mathbf{z}_x &= \frac{\mathbf{x} - \mu_x \mathbf{1}}{\sigma_x} \\
 &= \left(\mathbf{x} \quad -\mathbf{1} \right) \begin{pmatrix} \frac{1}{\sigma_x} \\ \frac{\mu_x}{\sigma_x} \\ \frac{\mu_x}{\sigma_x} \\ \sigma_x \end{pmatrix} \\
 &= \mathbf{X} \boldsymbol{\alpha}_x
 \end{aligned} \tag{11.1}$$

with

$$\mathbf{X} = \left(\mathbf{x} \quad -\mathbf{1} \right) \quad \text{and} \quad \boldsymbol{\alpha}_x = \begin{pmatrix} \frac{1}{\sigma_x} \\ \frac{\mu_x}{\sigma_x} \\ \frac{\mu_x}{\sigma_x} \\ \sigma_x \end{pmatrix}. \tag{11.2}$$

Similarly it follows from standardising \mathbf{y} in (9.1) that

$$\begin{aligned} \mathbf{z}_y &= \frac{\mathbf{y} - \mu_y \mathbf{1}}{\sigma_y} \\ &= \begin{pmatrix} \mathbf{y} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_y} \\ \frac{\mu_y}{\sigma_y} \end{pmatrix} . \\ &= \mathbf{Y}\boldsymbol{\alpha}_y \end{aligned} \tag{11.3}$$

with

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y} & -\mathbf{1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\alpha}_y = \begin{pmatrix} \frac{1}{\sigma_y} \\ \frac{\mu_y}{\sigma_y} \end{pmatrix} . \tag{11.4}$$

The vectors $\boldsymbol{\alpha}_x$ in (11.2) and $\boldsymbol{\alpha}_y$ in (11.3) are referred to as the vectors of so-called natural parameters.

The bivariate normal probabilities

$$\Phi_{ij} = F(z_{x_i}, z_{x_j}) = P(Z_x \leq z_{x_i}, Z_y \leq z_{y_j}) \tag{11.5}$$

with corresponding standardised upper class boundaries are tabulated in Table 11.1.

Table 11.1: Bivariate normal probabilities.

	z_{y_1}	z_{y_2}	\cdots	$z_{y_{J-1}}$	z_{y_J}
z_{x_1}	Φ_{11}	Φ_{12}	\cdots	$\Phi_{1,J-1}$	Φ_{1J}
z_{x_2}	Φ_{21}	Φ_{22}	\cdots	$\Phi_{2,J-1}$	Φ_{2J}
\vdots	\vdots	\vdots	\cdots	\vdots	\vdots
$z_{x_{(I-1)}}$	$\Phi_{I-1,1}$	$\Phi_{I-1,2}$	\cdots	$\Phi_{I-1,J-1}$	$\Phi_{I-1,J}$
z_{x_I}	Φ_{I1}	Φ_{I2}	\cdots	$\Phi_{I,J-1}$	Φ_{IJ}

To fit a bivariate normal distribution to the contingency table in Table 9.1 it is required that the bivariate normal probabilities should equal the corresponding cumulative relative frequencies i.e.

$$[\boldsymbol{\Phi}]_{ij} = [\boldsymbol{\Pi}]_{ij} \quad \text{for } i = 1, 2, \dots, I \text{ and } j = 1, 2, \dots, J \tag{11.6}$$

where

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1,J-1} & \Phi_{1J} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2,J-1} & \Phi_{2J} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \Phi_{I-1,1} & \Phi_{I-1,2} & \cdots & \Phi_{I-1,J-1} & \Phi_{I-1,J} \\ \Phi_{I1} & \Phi_{I2} & \cdots & \Phi_{I,J-1} & \Phi_{IJ} \end{pmatrix} \quad (11.7)$$

is the matrix with bivariate normal probabilities defined in (11.5) and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1,J-1} & \pi_{1,J} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2,J-1} & \pi_{2,J} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \pi_{I-1,1} & \pi_{I-1,2} & \cdots & \pi_{I-1,J-1} & \pi_{I-1,J} \\ \pi_{I1} & \pi_{I2} & \cdots & \pi_{I,J-1} & \pi_{IJ} \end{pmatrix} \quad (11.8)$$

is the corresponding matrix with expected cumulative relative frequencies defined in (9.15).

It follows from (11.6), that the following three conditions must hold:

1. Marginal distribution of \mathbf{x} :

$$\Phi_{\mathbf{x}} = \pi_{\mathbf{x}} \quad \begin{pmatrix} \Phi_{1J} \\ \Phi_{2J} \\ \vdots \\ \Phi_{I-1,J} \end{pmatrix} = \begin{pmatrix} \pi_{1,J} \\ \pi_{2,J} \\ \vdots \\ \pi_{I-1,J} \end{pmatrix} \quad (11.9)$$

(First $(I - 1)$ elements of last columns of Φ (11.7) and Π (11.8).)

2. Marginal distribution of \mathbf{y} :

$$\Phi_{\mathbf{y}} = \pi_{\mathbf{y}} \quad \left(\Phi_{I1} \ \Phi_{I2} \ \cdots \ \Phi_{I,J-1} \right)' = \left(\pi_{I1} \ \pi_{I2} \ \cdots \ \pi_{I,J-1} \right)' \quad (11.10)$$

(First $(J - 1)$ elements of last rows of Φ (11.7) and Π (11.8).)

3. Joint distribution of \mathbf{x} and \mathbf{y} :

$$\mathbf{\Phi}_{xy} = \boldsymbol{\pi}_{xy}$$

$$\text{vec} \begin{pmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1,J-1} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2,J-1} \\ \vdots & \vdots & \cdots & \vdots \\ \Phi_{I-1,1} & \Phi_{I-1,2} & \cdots & \Phi_{I-1,J-1} \end{pmatrix} = \text{vec} \begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1,J-1} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2,J-1} \\ \vdots & \vdots & \cdots & \vdots \\ \pi_{I-1,1} & \pi_{I-1,2} & \cdots & \pi_{I-1,J-1} \end{pmatrix} \quad (11.11)$$

(First $(I - 1)(J - 1)$ elements of $\mathbf{\Phi}$ (11.7) and $\mathbf{\Pi}$ (11.8).)

In $\mathbf{\Phi}_{xy}$ and $\boldsymbol{\pi}_{xy}$ the elements of the joint bivariate probabilities and the elements of the joint cumulative relative frequencies are stacked row by row as a single column vector.

11.2 Parameters

The bivariate normal distribution depends on five parameters i.e.

$$(x, y) \sim \text{BVN}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$$

where $-\infty < \mu_x, \mu_y < \infty$, $0 < \sigma_x, \sigma_y < \infty$ and $-1 < \rho < 1$. The parameters μ_x and σ_x are functions of the marginal distribution of x , while the parameters μ_y and σ_y are functions of the marginal distribution of y . The parameter ρ is a function of the joint distribution of x and y .

11.2.1 Marginal distribution of \mathbf{x}

From the properties of the bivariate normal distribution it follows that the marginal cumulative relative frequencies

$$\boldsymbol{\pi}_x = \begin{pmatrix} \pi_{1,J} \\ \pi_{2,J} \\ \vdots \\ \pi_{I-1,J} \end{pmatrix} \quad (11.12)$$

follow a cumulative $N(\mu_x, \sigma_x^2)$ distribution curve at the upper class boundaries of \mathbf{x} and hence

$$\begin{aligned}\Phi^{-1}(\boldsymbol{\pi}_x) &= \mathbf{z}_x \\ &= \mathbf{X}\boldsymbol{\alpha}_x\end{aligned}\quad (11.13)$$

which leads to

$$\boldsymbol{\alpha}_x = \begin{pmatrix} \frac{1}{\sigma_x} \\ \frac{\mu_x}{\sigma_x} \\ \sigma_x \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\Phi^{-1}(\boldsymbol{\pi}_x) . \quad (11.14)$$

Under normality (11.13), the standardised upper class boundaries \mathbf{z}_x , is a function of the natural parameters $\boldsymbol{\alpha}_x$. By substituting (11.14) in (11.13) it follows that \mathbf{z}_x is the projection of $\Phi^{-1}(\boldsymbol{\pi}_x)$ on the vector space of \mathbf{X} i.e.

$$\mathbf{z}_x = \mathbf{P}_X \Phi^{-1}(\boldsymbol{\pi}_x) \quad (11.15)$$

where

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \quad (11.16)$$

is the projection matrix of the vector space generated by the columns of \mathbf{X} .

11.2.2 Marginal distribution of \mathbf{y}

The cumulative relative frequencies

$$\boldsymbol{\pi}_y = \left(\pi_{I1} \quad \pi_{I2} \quad \cdots \quad \pi_{I,J-1} \right)' \quad (11.17)$$

follow a cumulative $N(\mu_y, \sigma_y^2)$ distribution curve at the upper class boundaries of \mathbf{y} and hence

$$\begin{aligned}\Phi^{-1}(\boldsymbol{\pi}_y) &= \mathbf{z}_y \\ &= \mathbf{Y}\boldsymbol{\alpha}_y\end{aligned}\quad (11.18)$$

which leads to

$$\boldsymbol{\alpha}_y = \begin{pmatrix} \frac{1}{\sigma_y} \\ \frac{\mu_y}{\sigma_y} \\ \sigma_y \end{pmatrix} = (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\Phi^{-1}(\boldsymbol{\pi}_y) . \quad (11.19)$$

Under normality (11.18), the standardised upper class boundaries \mathbf{z}_y , is a function of the natural parameters α_y . By substituting (11.19) in (11.18) it follows that \mathbf{z}_y is the projection of $\Phi^{-1}(\boldsymbol{\pi}_y)$ on the vector space of \mathbf{Y} i.e.

$$\mathbf{z}_y = \mathbf{P}_Y \Phi^{-1}(\boldsymbol{\pi}_y) \quad (11.20)$$

where

$$\mathbf{P}_Y = \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}' \quad (11.21)$$

11.2.3 Joint distribution of \mathbf{x} and \mathbf{y}

The one-to-one relationship between the correlation coefficient and the volumes of the four quadrants of the bivariate normal distribution

$$\rho = \sin\left(\frac{\pi}{2}[(\text{VOL } 1 + \text{VOL } 4) - (\text{VOL } 2 + \text{VOL } 3)]\right) \quad (11.22)$$

is explained in the previous chapter. The four quadrants of the bivariate normal distribution are denoted by Q_1, Q_2, Q_3 and Q_4 and by adding the relative frequencies in the 4 quadrants it is possible to calculate the volume for each quadrant. In matrix notation the vector of relative frequencies is

$$\boldsymbol{\pi}_0 = \mathbf{C}^{-1}\boldsymbol{\pi} \quad (11.23)$$

(See (9.12) for an explanation of the matrix \mathbf{C} .)

The expressions for the 4 volumes are as follows:

$$\text{VOL } 1 = \mathbf{v}'_1\boldsymbol{\pi}_0 = \mathbf{v}'_1\mathbf{C}^{-1}\boldsymbol{\pi} \quad (11.24)$$

$$\text{VOL } 2 = \mathbf{v}'_2\boldsymbol{\pi}_0 = \mathbf{v}'_2\mathbf{C}^{-1}\boldsymbol{\pi} \quad (11.25)$$

$$\text{VOL } 3 = \mathbf{v}'_3\boldsymbol{\pi}_0 = \mathbf{v}'_3\mathbf{C}^{-1}\boldsymbol{\pi} \quad (11.26)$$

$$\text{VOL } 4 = \mathbf{v}'_4\boldsymbol{\pi}_0 = \mathbf{v}'_4\mathbf{C}^{-1}\boldsymbol{\pi} \quad (11.27)$$

where

$$\mathbf{v}_q = \text{vec}(\mathbf{V}_q) \quad \text{for } q = \{1, 2, 3, 4\} \quad (11.28)$$

and \mathbf{V}_q is an $(I \times J)$ indicator matrix such that:

1. $[\mathbf{V}_q]_{ij} = 1$ if the (i, j) -th cell $\in Q_q$ for $q = \{1, 2, 3, 4\}$
2. $[\mathbf{V}_q]_{ij} = 0$ if the (i, j) -th cell $\notin Q_q$ for $q = \{1, 2, 3, 4\}$
3. Cells containing the lines $z_x = 0$ or $z_y = 0$, i.e. belonging to more than one quadrant, should be allocated proportionately to the standard bivariate normal distribution, depending on the value of ρ .

This implies that

$$\sum_{q=1}^4 \mathbf{v}_q = 1 \quad (11.29)$$

and following from (11.22) it is now possible to express ρ as

$$\rho = \sin \left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1} \boldsymbol{\pi} \right) . \quad (11.30)$$

11.3 Vector of constraints

The vector of constraints, $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$, with

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_x(\boldsymbol{\pi}) \\ \mathbf{g}_y(\boldsymbol{\pi}) \\ \mathbf{g}_{xy}(\boldsymbol{\pi}) \end{pmatrix} \quad (11.31)$$

consists out of three sets of constraints.

11.3.1 Marginal distribution of x

$$\begin{aligned}
 \mathbf{g}_x(\boldsymbol{\pi}) &= \boldsymbol{\Phi}_x - \boldsymbol{\pi}_x & (11.32) \\
 &= \boldsymbol{\Phi}(\mathbf{z}_x) - \boldsymbol{\pi}_x \\
 &= \begin{pmatrix} \Phi_{1J} \\ \Phi_{2J} \\ \vdots \\ \Phi_{I-1,J} \end{pmatrix} - \begin{pmatrix} \pi_{1J} \\ \pi_{2J} \\ \vdots \\ \pi_{I-1,J} \end{pmatrix}
 \end{aligned}$$

The $(I - 1)$ constraints in $\mathbf{g}_x(\boldsymbol{\pi})$ refer to the marginal cumulative relative frequencies $\boldsymbol{\pi}_x$, that has to follow a cumulative normal distribution curve at the standardised upper class boundaries \mathbf{x} . This follows from the properties of the bivariate normal distribution, since the marginal distribution of x is

$$x \sim N(\mu_x, \sigma_x^2) .$$

11.3.2 Marginal distribution of y

$$\begin{aligned}
 \mathbf{g}_y(\boldsymbol{\pi}) &= \boldsymbol{\Phi}_y - \boldsymbol{\pi}_y & (11.33) \\
 &= \boldsymbol{\Phi}(\mathbf{z}_y) - \boldsymbol{\pi}_y \\
 &= \begin{pmatrix} \Phi_{I1} \\ \Phi_{I,2} \\ \vdots \\ \Phi_{I,J-1} \end{pmatrix} - \begin{pmatrix} \pi_{I1} \\ \pi_{I,2} \\ \vdots \\ \pi_{I,J-1} \end{pmatrix}
 \end{aligned}$$

The $(J - 1)$ constraints in $\mathbf{g}_y(\boldsymbol{\pi})$ refer to the marginal cumulative relative frequencies $\boldsymbol{\pi}_y$, that has to follow a cumulative normal distribution curve at the upper class boundaries \mathbf{y} . This follows since the marginal distribution of y is

$$y \sim N(\mu_y, \sigma_y^2) .$$

11.3.3 Joint distribution of x and y

$$\mathbf{g}_{xy}(\boldsymbol{\pi}) = \boldsymbol{\Phi}_{xy} - \boldsymbol{\pi}_{xy} \quad (11.34)$$

The $(I - 1)(J - 1)$ constraints in $\mathbf{g}_{xy}(\boldsymbol{\pi})$ refer to the joint cumulative relative frequencies $\boldsymbol{\pi}_{xy}$, that has to follow a cumulative bivariate normal distribution curve at the intersections of the upper class boundaries x and y . The bivariate normal distribution to be fitted is such that

$$(x, y) \sim \text{BVN}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho) .$$

The elements of

$$\begin{aligned} \boldsymbol{\Phi}_{xy} &= \text{vec}(\boldsymbol{\Phi}(\mathbf{z}_x, \mathbf{z}'_y)) \\ &= \text{vec} \begin{pmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1,J-1} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2,J-1} \\ \vdots & \vdots & \cdots & \vdots \\ \Phi_{I-1,1} & \Phi_{I-1,2} & \cdots & \Phi_{I-1,J-1} \end{pmatrix} \end{aligned}$$

are the cumulative probabilities from the standard bivariate normal distribution at the intersections of the class boundaries \mathbf{z}_x and \mathbf{z}_y stacked row by row below each other as a single column vector and the elements of

$$\boldsymbol{\pi}_{xy} = \text{vec} \begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1,J-1} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2,J-1} \\ \vdots & \vdots & \cdots & \vdots \\ \pi_{I-1,1} & \pi_{I-1,2} & \cdots & \pi_{I-1,J-1} \end{pmatrix}$$

are the cumulative relative frequencies, also stacked row by row below each other.

11.4 Matrix of Partial Derivatives

As in the case of the vector of constraints, the matrix of partial derivatives of $\mathbf{g}(\boldsymbol{\pi})$ with respect to $\boldsymbol{\pi}$

$$\mathbf{G}_{\boldsymbol{\pi}} = \frac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} = \begin{pmatrix} \frac{\partial \mathbf{g}_x(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ \frac{\partial \mathbf{g}_y(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ \frac{\partial \mathbf{g}_{xy}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \end{pmatrix} \quad (11.35)$$

also consists out of three sets and will be derived below.

11.4.1 Marginal distribution of \mathbf{x}

$$\begin{aligned} \frac{\partial \mathbf{g}_x(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} &= \frac{\partial \Phi_x}{\partial \boldsymbol{\pi}} - \frac{\partial \boldsymbol{\pi}_x}{\partial \boldsymbol{\pi}} \\ &= \frac{\partial \Phi(\mathbf{z}_x)}{\partial \boldsymbol{\pi}} - \mathbf{I}_x \end{aligned} \quad (11.36)$$

where

$$\mathbf{I}_x = \frac{\partial \boldsymbol{\pi}_x}{\partial \boldsymbol{\pi}} : (I - 1) \times IJ. \quad (11.37)$$

Since $\mathbf{z}_x = \mathbf{X}\boldsymbol{\alpha}_x$ with $\boldsymbol{\alpha}_x = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\Phi^{-1}(\boldsymbol{\pi}_x)$ it follows from the chain rule for matrix differentiation

$$\begin{aligned} \frac{\partial \Phi(\mathbf{z}_x)}{\partial \boldsymbol{\pi}} &= \frac{\partial \Phi(\mathbf{z}_x)}{\partial \mathbf{z}_x} \cdot \frac{\partial \mathbf{z}_x}{\partial \boldsymbol{\alpha}_x} \cdot \frac{\partial \boldsymbol{\alpha}_x}{\partial \boldsymbol{\pi}_x} \cdot \frac{\partial \boldsymbol{\pi}_x}{\partial \boldsymbol{\pi}} \\ &= \text{diag}[\phi(\mathbf{z}_x)] \cdot \mathbf{P}_X \cdot \mathbf{D}_x \cdot \mathbf{I}_x \end{aligned} \quad (11.38)$$

where

$$\mathbf{D}_x = \frac{\partial \Phi^{-1}(\boldsymbol{\pi}_x)}{\partial \boldsymbol{\pi}_x}. \quad (11.39)$$

To solve (11.39) set $\boldsymbol{\nu} = \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_x)$ then $\boldsymbol{\Phi}(\boldsymbol{\nu}) = \boldsymbol{\pi}_x$ and hence

$$\begin{aligned}
 \mathbf{D}_x &= \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\pi}_x} \\
 &= \left(\frac{\partial \boldsymbol{\pi}_x}{\partial \boldsymbol{\nu}} \right)^{-1} \\
 &= \left(\frac{\partial \boldsymbol{\Phi}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right)^{-1} \\
 &= (\text{diag}[\boldsymbol{\phi}(\boldsymbol{\nu})])^{-1} \\
 &= (\text{diag}[\boldsymbol{\phi}(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_x))])^{-1}.
 \end{aligned} \tag{11.40}$$

11.4.2 Marginal distribution of y

$$\begin{aligned}
 \frac{\partial \mathbf{g}_y(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} &= \frac{\partial \boldsymbol{\Phi}_y}{\partial \boldsymbol{\pi}} - \frac{\partial \boldsymbol{\pi}_y}{\partial \boldsymbol{\pi}} \\
 &= \frac{\partial \boldsymbol{\Phi}(\mathbf{z}_y)}{\partial \boldsymbol{\pi}} - \mathbf{I}_y
 \end{aligned} \tag{11.41}$$

where

$$\mathbf{I}_y = \frac{\partial \boldsymbol{\pi}_y}{\partial \boldsymbol{\pi}} : (J-1) \times IJ. \tag{11.42}$$

Since $\mathbf{z}_y = \mathbf{X}\boldsymbol{\alpha}_y$ and $\boldsymbol{\alpha}_y = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_y)$ it follows from the chain rule for matrix differentiation

$$\begin{aligned}
 \frac{\partial \boldsymbol{\Phi}(\mathbf{z}_y)}{\partial \boldsymbol{\pi}} &= \frac{\partial \boldsymbol{\Phi}(\mathbf{z}_y)}{\partial \mathbf{z}_y} \cdot \frac{\partial \mathbf{z}_y}{\partial \boldsymbol{\alpha}_y} \cdot \frac{\partial \boldsymbol{\alpha}_y}{\partial \boldsymbol{\pi}_y} \cdot \frac{\partial \boldsymbol{\pi}_y}{\partial \boldsymbol{\pi}} \\
 &= \text{diag}[\boldsymbol{\phi}(\mathbf{z}_y)] \cdot \mathbf{P}_Y \cdot \mathbf{D}_y \cdot \mathbf{I}_y
 \end{aligned} \tag{11.43}$$

where

$$\begin{aligned}
 \mathbf{D}_y &= \frac{\partial \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_y)}{\partial \boldsymbol{\pi}_y} \\
 &= \left(\text{diag}[\boldsymbol{\phi}(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_y))] \right)^{-1}.
 \end{aligned} \tag{11.44}$$

11.4.3 Joint distribution of \mathbf{x} and \mathbf{y}

From the chain rule for matrix differentiation it follows that

$$\begin{aligned}
 \frac{\partial \mathbf{g}_{xy}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} &= \frac{\partial \Phi_{xy}}{\partial \boldsymbol{\pi}} - \frac{\partial \boldsymbol{\pi}_{xy}}{\partial \boldsymbol{\pi}} \\
 &= \frac{\partial \Phi_{xy}}{\partial \begin{pmatrix} \mathbf{z}_x \\ \mathbf{z}_y \\ \rho \end{pmatrix}} \cdot \frac{\partial \begin{pmatrix} \mathbf{z}_x \\ \mathbf{z}_y \\ \rho \end{pmatrix}}{\partial \boldsymbol{\pi}} - \frac{\partial \boldsymbol{\pi}_{xy}}{\partial \boldsymbol{\pi}} \\
 &= \begin{pmatrix} \frac{\partial \Phi_{xy}}{\partial \mathbf{z}_x} & \frac{\partial \Phi_{xy}}{\partial \mathbf{z}_y} & \frac{\partial \Phi_{xy}}{\partial \rho} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \mathbf{z}_x}{\partial \boldsymbol{\pi}} \\ \frac{\partial \mathbf{z}_y}{\partial \boldsymbol{\pi}} \\ \frac{\partial \rho}{\partial \boldsymbol{\pi}} \end{pmatrix} - \mathbf{I}_{xy} \quad (11.45)
 \end{aligned}$$

where

$$\mathbf{I}_{xy} = \frac{\partial \boldsymbol{\pi}_{xy}}{\partial \boldsymbol{\pi}} : (I - 1)(J - 1) \times IJ. \quad (11.46)$$

A total of 6 derivatives that are labeled in (11.45), are simplified in (1) to (6) below.

1.

$$\begin{aligned}
& \frac{\partial}{\partial z_{x_i}} F(z_{x_i}, z_{y_j}) \\
&= \frac{\partial}{\partial z_{x_i}} \int_{-\infty}^{z_{y_j}} \int_{-\infty}^{z_{x_i}} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right\} dz_1 dz_2 \\
&= \int_{-\infty}^{z_{y_j}} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}(z_{x_i}^2 - 2\rho z_{x_i} z_2 + z_2^2)\right\} dz_2 \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{(z_{x_i}^2 - \rho^2 z_{x_i}^2)}{2(1-\rho^2)}\right\} \int_{-\infty}^{z_{y_j}} \exp\left\{-\frac{1}{2(1-\rho^2)}(z_2 - \rho z_{x_i})^2\right\} dz_2 \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2}z_{x_i}^2\right\} \int_{-\infty}^{z_{y_j}} \exp\left\{-\frac{1}{2}\left(\frac{z_2 - \rho z_{x_i}}{\sqrt{1-\rho^2}}\right)^2\right\} dz_2
\end{aligned}$$

Set $w = \left(\frac{z_2 - \rho z_{x_i}}{\sqrt{1-\rho^2}}\right)$ then

$$dw = \frac{1}{\sqrt{1-\rho^2}} dz_2$$

and consequently

$$\begin{aligned}
\frac{\partial}{\partial z_{x_i}} F(z_{x_i}, z_{y_j}) &= \frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2}z_{x_i}^2\right\} \int_{-\infty}^{\frac{z_{y_j} - \rho z_{x_i}}{\sqrt{1-\rho^2}}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}w^2\right\} dw \\
&= \phi(z_{x_i}) \Phi\left(\frac{z_{y_j} - \rho z_{x_i}}{\sqrt{1-\rho^2}}\right)
\end{aligned}$$

It now follows that

$$\begin{aligned}
\frac{\partial \Phi_{xy}}{\partial \mathbf{z}_x} &= \left(\frac{\partial \Phi_{xy}}{\partial z_{x_1}}, \frac{\partial \Phi_{xy}}{\partial z_{x_2}}, \dots, \frac{\partial \Phi_{xy}}{\partial z_{x_{I-1}}}\right) \\
&= (\text{vec}(\mathbf{E}_1 \Delta_x), \text{vec}(\mathbf{E}_2 \Delta_x), \dots, \text{vec}(\mathbf{E}_{I-1} \Delta_x))
\end{aligned} \tag{11.47}$$

where

$$\Delta_x = \text{diag}(\phi(\mathbf{z}_x)) \cdot \Phi\left(\frac{(\mathbf{z}'_y \otimes \mathbf{1}_{I-1}) - \rho(\mathbf{z}_x \otimes \mathbf{1}'_{J-1})}{\sqrt{1-\rho^2}}\right) \tag{11.48}$$

and $\mathbf{E}_i : (I-1 \times I-1)$, $i = \{1 \dots I-1\}$ is a matrix such that

$$\begin{aligned}
[\mathbf{E}_i]_{rs} &= 1 \quad \text{if } i = r = s \\
[\mathbf{E}_i]_{rs} &= 0 \quad \text{elsewhere.}
\end{aligned} \tag{11.49}$$

2. Likewise

$$\frac{\partial}{\partial z_{y_j}} F(z_{x_i}, z_{y_j}) = \phi(z_{y_j}) \Phi \left(\frac{z_{x_i} - \rho z_{y_j}}{\sqrt{1 - \rho^2}} \right)$$

and therefore it follows that

$$\begin{aligned} \frac{\partial \Phi_{xy}}{\partial \mathbf{z}_y} &= \left(\frac{\partial \Phi_{xy}}{\partial z_{y_1}}, \frac{\partial \Phi_{xy}}{\partial z_{y_2}}, \dots, \frac{\partial \Phi_{xy}}{\partial z_{y_{J-1}}} \right) \\ &= (\text{vec}(\Delta_y \mathbf{E}_1), \text{vec}(\Delta_y \mathbf{E}_2), \dots, \text{vec}(\Delta_y \mathbf{E}_{J-1})) \end{aligned} \quad (11.50)$$

where

$$\Delta_y = \Phi \left(\frac{(\mathbf{z}_x \otimes \mathbf{1}'_{J-1}) - \rho (\mathbf{z}'_y \otimes \mathbf{1}_{I-1})}{\sqrt{1 - \rho^2}} \right) \cdot \text{diag}(\phi(\mathbf{z}_y)) \quad (11.51)$$

and $\mathbf{E}_j : (J - 1 \times J - 1)$, $j = \{1, \dots, J - 1\}$ is a matrix such that

$$\begin{aligned} [\mathbf{E}_j]_{vw} &= 1 \quad \text{if } j = v = w \\ [\mathbf{E}_j]_{vw} &= 0 \quad \text{elsewhere.} \end{aligned} \quad (11.52)$$

3.

$$\begin{aligned} &\frac{\partial F(z_{x_i}, z_{y_j})}{\partial \rho} \\ &= \frac{\partial}{\partial \rho} \int_{-\infty}^{z_{y_j}} \int_{-\infty}^{z_{x_i}} \frac{1}{2\pi \sqrt{1 - \rho^2}} \cdot \exp \left\{ -\frac{1}{2(1 - \rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \right\} dz_1 dz_2 \\ &= \int_{-\infty}^{z_{y_j}} \int_{-\infty}^{z_{x_i}} \frac{\partial}{\partial \rho} \left\{ \frac{1}{2\pi \sqrt{1 - \rho^2}} \right\} \cdot \exp \left\{ -\frac{1}{2(1 - \rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \right\} dz_1 dz_2 + \\ &\quad \int_{-\infty}^{z_{y_j}} \int_{-\infty}^{z_{x_i}} \frac{1}{2\pi \sqrt{1 - \rho^2}} \cdot \frac{\partial}{\partial \rho} \exp \left\{ -\frac{1}{2(1 - \rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \right\} dz_1 dz_2 \\ &= \frac{\rho}{1 - \rho^2} \int_{-\infty}^{z_{y_j}} \int_{-\infty}^{z_{x_i}} \frac{1}{2\pi \sqrt{1 - \rho^2}} \cdot \exp \left\{ -\frac{1}{2(1 - \rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \right\} dz_1 dz_2 + \\ &\quad \int_{-\infty}^{z_{y_j}} \int_{-\infty}^{z_{x_i}} \frac{1}{2\pi \sqrt{1 - \rho^2}} \cdot \exp \left\{ -\frac{1}{2(1 - \rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \right\} \cdot \\ &\quad \underbrace{\frac{\partial}{\partial \rho} \left\{ -\frac{1}{2(1 - \rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \right\}}_{dz_1 dz_2} \end{aligned}$$

Simplification of the derivative above leads to

$$\begin{aligned} & \frac{\partial}{\partial \rho} \left(-\frac{1}{2(1-\rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \right) \\ &= \frac{z_1 z_2}{(1-\rho^2)} - \frac{\rho}{(1-\rho^2)^2} [z_1^2 - 2\rho z_1 z_2 + z_2^2] \\ &= -\frac{\rho}{(1-\rho^2)^2} z_1^2 + \frac{(1-\rho^2) + 2\rho^2}{(1-\rho^2)^2} z_1 z_2 - \frac{\rho}{(1-\rho^2)^2} z_2^2 \end{aligned}$$

and therefore

$$\begin{aligned} \frac{\partial F(z_{x_i}, z_{y_j})}{\partial \rho} &= \frac{\rho}{1-\rho^2} \tilde{\Psi}(z_{x_i}, z_{y_j}, 0, 0; \rho) - \frac{\rho}{(1-\rho^2)^2} \tilde{\Psi}(z_{x_i}, z_{y_j}, 2, 0; \rho) + \\ & \frac{1+\rho^2}{(1-\rho^2)^2} \tilde{\Psi}(z_{x_i}, z_{y_j}, 1, 1; \rho) - \frac{\rho}{(1-\rho^2)^2} \tilde{\Psi}(z_{x_i}, z_{y_j}, 0, 2; \rho) \end{aligned} \quad (11.53)$$

where

$$\tilde{\Psi}(z_{x_i}, z_{y_j}, k, l; \rho) = \int_{-\infty}^{z_{y_j}} \int_{-\infty}^{z_{x_i}} \frac{1}{2\pi\sqrt{1-\rho^2}} \underbrace{z_1^k z_2^l}_{z_1^k z_2^l} \exp\left(-\frac{1}{2(1-\rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2]\right) dz_1 dz_2 \quad (11.54)$$

Define the integral

$$\tilde{\Psi}_0(z_{x_i}, z_{y_j}, k, l; \rho) = \int_0^{z_{y_j}} \int_0^{z_{x_i}} \frac{1}{2\pi\sqrt{1-\rho^2}} \underbrace{z_1^k z_2^l}_{z_1^k z_2^l} \exp\left(-\frac{1}{2(1-\rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2]\right) dz_1 dz_2 \quad (11.55)$$

where integration takes place from the origin. Depending on the specific location of (z_{x_i}, z_{y_j}) , $\tilde{\Psi}(z_{x_i}, z_{y_j}, k, l; \rho)$ (11.54) can be expressed in terms of $\tilde{\Psi}_0(z_{x_i}, z_{y_j}, k, l; \rho)$ (11.55) as follows:

Quadrant 1: $(z_{x_i} < 0, z_{y_j} < 0)$

$$\begin{aligned} \tilde{\Psi}(z_{x_i}, z_{y_j}, k, l; \rho) &= \tilde{\Psi}_0(\infty, \infty, k, l; \rho) - \tilde{\Psi}_0(-z_{x_i}, \infty, k, l; \rho) - \\ & \tilde{\Psi}_0(\infty, -z_{y_j}, k, l; \rho) + \tilde{\Psi}_0(-z_{x_i}, -z_{y_j}, k, l; \rho) \end{aligned} \quad (11.56)$$

Quadrant 2: $(z_{x_i} < 0, z_{y_j} \geq 0)$

$$\begin{aligned} \tilde{\Psi}(z_{x_i}, z_{y_j}, k, l; \rho) &= \tilde{\Psi}_0(\infty, \infty, k, l; \rho) - \tilde{\Psi}_0(-z_{x_i}, \infty, k, l; \rho) + \\ & (-1)^k \tilde{\Psi}_0(\infty, z_{y_j}, k, l; -\rho) - (-1)^k \tilde{\Psi}_0(-z_{x_i}, z_{y_j}, k, l; -\rho) \end{aligned} \quad (11.57)$$

Quadrant 3: ($z_{x_i} \geq 0, z_{y_j} < 0$)

$$\begin{aligned} \tilde{\Psi}(z_{x_i}, z_{y_j}, k, l; \rho) &= \tilde{\Psi}_0(\infty, \infty, k, l; \rho) + (-1)^l \tilde{\Psi}_0(z_{x_i}, \infty, k, l; -\rho) - \\ &\tilde{\Psi}_0(\infty, -z_{y_j}, k, l; \rho) - (-1)^l \tilde{\Psi}_0(z_{x_i}, -z_{y_j}, k, l; -\rho) \end{aligned} \quad (11.58)$$

Quadrant 4: ($z_{x_i} \geq 0, z_{y_j} \geq 0$)

$$\begin{aligned} \tilde{\Psi}(z_{x_i}, z_{y_j}, k, l; \rho) &= \tilde{\Psi}_0(\infty, \infty; \rho, k, l) + (-1)^l \tilde{\Psi}_0(z_{x_i}, \infty, -\rho, k, l) + \\ &(-1)^k \tilde{\Psi}_0(\infty, z_{y_j}, k, l; -\rho) + \tilde{\Psi}_0(z_{x_i}, z_{y_j}, k, l; \rho) \end{aligned} \quad (11.59)$$

The integral $\tilde{\Psi}_0(z_{x_i}, z_{y_j}, k, l; \rho)$ is expressed as a series of gamma functions in Algorithm 2.

Algorithm 2

$$\begin{aligned} \Psi_0(z_{x_i}, z_{y_j}, k, l; \rho) &= \int_0^{z_{y_j}} \int_0^{z_{x_i}} \frac{1}{2\pi\sqrt{1-\rho^2}} \underbrace{z_1^k z_2^l}_{z_1^k z_2^l} \exp\left(-\frac{1}{2(1-\rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2]\right) dz_1 dz_2 \\ &= \frac{2^{\frac{k+l}{2}} (1-\rho^2)^{\frac{k+l+1}{2}}}{4\pi} \sum_{i=0}^{\infty} \left\{ \frac{(2\rho)^i}{i!} \Gamma\left(\frac{i+k+1}{2}\right) \Gamma\left(\frac{i+l+1}{2}\right) \right. \\ &\quad \left. \cdot G\left(\frac{z_{x_i}^2}{2(1-\rho^2)}, \frac{i+k+1}{2}\right) \cdot G\left(\frac{z_{y_j}^2}{2(1-\rho^2)}, \frac{i+l+1}{2}\right) \right\} \end{aligned} \quad (11.60)$$

where $G(x; \kappa) = \int_0^x \frac{1}{\Gamma(\kappa)} t^{\kappa-1} e^{-t} dt$ is the gamma distribution with shape parameter κ .

Proof. Since

$$\exp\left(\frac{\rho z_1 z_2}{(1-\rho^2)}\right) = \frac{\sum_{i=0}^{\infty} \left(\frac{\rho z_1 z_2}{1-\rho^2}\right)^i}{i!}$$

it follows that

$$\begin{aligned} \tilde{\Psi}_0(z_{x_i}, z_{y_j}, k, l; \rho) &= \int_0^{z_{y_j}} \int_0^{z_{x_i}} \frac{1}{2\pi\sqrt{1-\rho^2}} \underbrace{z_1^k z_2^l}_{z_1^k z_2^l} \exp\left(-\frac{1}{2(1-\rho^2)} [z_1^2 - 2\rho z_1 z_2 + z_2^2]\right) dz_1 dz_2 \\ &= \sum_{i=0}^{\infty} \frac{1}{2\pi i! \sqrt{1-\rho^2}} \left(\frac{\rho}{1-\rho^2}\right)^i \\ &\quad \cdot \int_0^{z_{x_i}} \exp\left(-\frac{z_1^2}{2(1-\rho^2)}\right) z_1^{i+k} dz_1 \cdot \int_0^{z_{y_j}} \exp\left(-\frac{z_2^2}{2(1-\rho^2)}\right) z_2^{i+l} dz_2 \end{aligned}$$

from (10.12) it follows that

$$\begin{aligned}
\tilde{\Psi}_0(z_{x_i}, z_{y_j}, k, l; \rho) &= \sum_{i=0}^{\infty} \frac{1}{2\pi i! \sqrt{1-\rho^2}} \left(\frac{\rho}{1-\rho^2}\right)^i \\
&\quad \cdot 2^{\frac{i+k-1}{2}} (1-\rho^2)^{\frac{i+k+1}{2}} \Gamma\left(\frac{i+k+1}{2}\right) G\left(\frac{z_{x_i}^2}{2(1-\rho^2)}, \frac{i+k+1}{2}\right) \\
&\quad \cdot 2^{\frac{i+l-1}{2}} (1-\rho^2)^{\frac{i+l+1}{2}} \Gamma\left(\frac{i+l+1}{2}\right) G\left(\frac{z_{y_j}^2}{2(1-\rho^2)}, \frac{i+l+1}{2}\right) \\
&= \frac{2^{\frac{k+l}{2}} (1-\rho^2)^{\frac{k+l+1}{2}}}{4\pi} \sum_{i=0}^{\infty} \left\{ \frac{(2\rho)^i}{i!} \Gamma\left(\frac{i+k+1}{2}\right) \Gamma\left(\frac{i+l+1}{2}\right) \right. \\
&\quad \left. \cdot G\left(\frac{z_{x_i}^2}{2(1-\rho^2)}, \frac{i+k+1}{2}\right) \cdot G\left(\frac{z_{y_j}^2}{2(1-\rho^2)}, \frac{i+l+1}{2}\right) \right\}
\end{aligned}$$

■

4. Since $\mathbf{z}_x = \mathbf{X}\alpha_x$ and $\alpha_x = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\Phi^{-1}(\boldsymbol{\pi}_x)$ it follows that

$$\begin{aligned}
\frac{\partial \mathbf{z}_x}{\partial \boldsymbol{\pi}} &= \frac{\partial \mathbf{z}_x}{\partial \alpha_x} \cdot \frac{\partial \alpha_x}{\partial \boldsymbol{\pi}_x} \cdot \frac{\partial \boldsymbol{\pi}_x}{\partial \boldsymbol{\pi}} \\
&= \mathbf{P}_X \cdot \mathbf{D}_x \cdot \mathbf{I}_x
\end{aligned} \tag{11.61}$$

See (11.38).

5. Similarly as in 4 above, $\mathbf{z}_y = \mathbf{Y}\alpha_y$ and $\alpha_y = (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\Phi^{-1}(\boldsymbol{\pi}_y)$ and therefore

$$\begin{aligned}
\frac{\partial \mathbf{z}_y}{\partial \boldsymbol{\pi}} &= \frac{\partial \mathbf{z}_y}{\partial \alpha_y} \cdot \frac{\partial \alpha_y}{\partial \boldsymbol{\pi}_y} \cdot \frac{\partial \boldsymbol{\pi}_y}{\partial \boldsymbol{\pi}} \\
&= \mathbf{P}_Y \cdot \mathbf{D}_y \cdot \mathbf{I}_y
\end{aligned} \tag{11.62}$$

See (11.43).

6. From (11.30) it follows that

$$\begin{aligned}
\frac{\partial \rho}{\partial \boldsymbol{\pi}} &= \frac{\partial}{\partial \boldsymbol{\pi}} \left\{ \sin\left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1} \boldsymbol{\pi}\right) \right\} \\
&= \cos\left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1} \boldsymbol{\pi}\right) \cdot \left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1}\right) .
\end{aligned}$$

11.5 Iterative procedure

A very short outline of the iterative procedure is as follows and will be discussed briefly.

\mathbf{p}^\dagger = observed cumulative relative frequencies

$\mathbf{p} = \mathbf{p}^\dagger$

DO OVER π

$\pi = \mathbf{p}$

Calculate $\mathbf{V} = \text{Cov}(\pi)$

Calculate \mathbf{z}_{x_π} , \mathbf{z}_{y_π} and ρ_π from π .

Calculate \mathbf{G}_π (as a function of π)

$\mathbf{p} = \mathbf{p}^\dagger$

DO OVER \mathbf{p}

Calculate \mathbf{z}_{x_p} , \mathbf{z}_{y_p} and ρ_p from \mathbf{p} .

Calculate \mathbf{G}_p (as a function of \mathbf{p})

$$\mathbf{g}(\mathbf{p}) = \begin{pmatrix} \Phi(\mathbf{z}_{x_p}) \\ \Phi(\mathbf{z}_{y_p}) \\ \text{vec}(\Phi(\mathbf{z}_{x_p}, \mathbf{z}_{y_p}, \rho_p)) \end{pmatrix} - \begin{pmatrix} \mathbf{p}_x \\ \mathbf{p}_y \\ \mathbf{p}_{xy} \end{pmatrix}$$

$$\mathbf{p} = \mathbf{p} - (\mathbf{G}_\pi \mathbf{V})' (\mathbf{G}_\pi \mathbf{V} \mathbf{G}_p)^* \mathbf{g}(\mathbf{p})$$

END

END

The procedure starts off with the unrestricted vector of cumulative relative frequencies. Convergence is first obtained over \mathbf{p} utilizing

$$\mathbf{p} = \mathbf{p} - (\mathbf{G}_\pi \mathbf{V})' (\mathbf{G}_\pi \mathbf{V} \mathbf{G}_p)^* \mathbf{g}(\mathbf{p}) \quad (11.63)$$

where the vectors of standardised upper class boundaries are calculated from

$$\mathbf{z}_{x_p} = \mathbf{P}_X \Phi^{-1}(\mathbf{p}_x) \quad \text{and} \quad \mathbf{z}_{y_p} = \mathbf{P}_Y \Phi^{-1}(\mathbf{p}_y) \quad (11.64)$$

projecting $\Phi^{-1}(\mathbf{p}_x)$ and $\Phi^{-1}(\mathbf{p}_y)$ into the respective vector spaces of \mathbf{X} and \mathbf{Y} . These standardised upper class boundaries divide the cells of the contingency table into 4 so-called quadrants leading to an estimate for

$$\rho_p = \sin\left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1}\mathbf{p}\right) . \quad (11.65)$$

Care should be taken to the cells belonging to more than one quadrant, since allocation of the relative frequencies \mathbf{p} should be done proportionately to the bivariate distribution, thus depending on the value of ρ_p . The calculation of ρ_p will therefore be done iteratively, starting at a value say, $\rho = 0$, until iteration over (11.65) leads to a unique estimate for ρ_p . (Explained in detail in the next chapter.) The vector of constraints $\mathbf{g}(\mathbf{p})$ and the matrix of partial derivatives \mathbf{G}_p are now all functions of \mathbf{p} and convergence over \mathbf{p} ultimately leads to a new value for π .

For convergence over π the covariance matrix \mathbf{V} and the matrix of partial derivatives \mathbf{G}_π are all functions of π . Convergence over π leads to the restricted ML estimate of π , i.e. $\hat{\pi}$, that satisfies all the properties of the bivariate normal distribution.

11.6 ML estimates

The ML estimates of the bivariate normal distribution can be obtained from the restricted ML estimate $\hat{\pi}$, discussed in the previous section. In matrix notation $\hat{\pi}$ can be represented as

$$\hat{\Pi} = \begin{pmatrix} \hat{\pi}_{11} & \hat{\pi}_{12} & \cdots & \hat{\pi}_{1,J-1} & \hat{\pi}_{1J} \\ \hat{\pi}_{21} & \hat{\pi}_{22} & \cdots & \hat{\pi}_{2,J-1} & \hat{\pi}_{2J} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \hat{\pi}_{I-1,1} & \hat{\pi}_{I-1,2} & \cdots & \hat{\pi}_{I-1,J-1} & \hat{\pi}_{I-1,J} \\ \hat{\pi}_{I1} & \hat{\pi}_{I2} & \cdots & \hat{\pi}_{I,J-1} & \hat{\pi}_{IJ} \end{pmatrix} \quad (11.66)$$

where $\hat{\pi}_{ij}$ corresponds to the restricted ML estimate of the cumulative relative frequency for the i -th row and the j -th column of the two-way contingency table. The asymptotic covariance matrix of $\hat{\pi}$ is

$$\text{Cov}(\hat{\pi}) \approx \mathbf{V} - (\mathbf{G}_\pi \mathbf{V})' (\mathbf{G}_\pi \mathbf{V} \mathbf{G}'_\pi)^* (\mathbf{G}_\pi \mathbf{V}) .$$

11.6.1 ML estimates of the natural parameters

The ML estimates of the vectors of natural parameters are functions of the restricted ML estimate $\hat{\boldsymbol{\pi}}$ with

$$\hat{\boldsymbol{\alpha}}_x = \begin{pmatrix} \hat{\alpha}_{x1} \\ \hat{\alpha}_{x2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\hat{\sigma}_x} \\ \frac{\hat{\mu}_x}{\hat{\sigma}_x} \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\pi}}_x) \quad (11.67)$$

and

$$\hat{\boldsymbol{\alpha}}_y = \begin{pmatrix} \hat{\alpha}_{y1} \\ \hat{\alpha}_{y2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\hat{\sigma}_y} \\ \frac{\hat{\mu}_y}{\hat{\sigma}_y} \end{pmatrix} = (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\pi}}_y) \quad (11.68)$$

where

$$\hat{\boldsymbol{\pi}}_x = \begin{pmatrix} \hat{\pi}_{1J} \\ \hat{\pi}_{2J} \\ \vdots \\ \hat{\pi}_{I-1,J} \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{\pi}}_y = \left(\hat{\pi}_{I1} \quad \hat{\pi}_{I2} \quad \cdots \quad \hat{\pi}_{I,J-1} \right)' . \quad (11.69)$$

See the last column and row of $\hat{\boldsymbol{\Pi}}$ (11.66).

The corresponding covariance matrices are

$$\text{Cov}(\hat{\boldsymbol{\alpha}}_x) = \left\{ (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{D}_x\mathbf{I}_x \right\} \text{Cov}(\hat{\boldsymbol{\pi}}) \left\{ (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{D}_x\mathbf{I}_x \right\}' \quad (11.70)$$

$$\text{Cov}(\hat{\boldsymbol{\alpha}}_y) = \left\{ (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{D}_y\mathbf{I}_y \right\} \text{Cov}(\hat{\boldsymbol{\pi}}) \left\{ (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{D}_y\mathbf{I}_y \right\}' \quad (11.71)$$

where

$$\mathbf{D}_x = \left(\text{diag} \left[\phi \left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_x) \right) \right] \right)^{-1}, \quad \mathbf{D}_y = \left(\text{diag} \left[\phi \left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}_y) \right) \right] \right)^{-1}$$

and

$$\mathbf{I}_x = \frac{\partial \boldsymbol{\pi}_x}{\partial \boldsymbol{\pi}}, \quad \mathbf{I}_y = \frac{\partial \boldsymbol{\pi}_y}{\partial \boldsymbol{\pi}} .$$

11.6.2 ML estimates of the original parameters

The ML estimates of the original parameters namely μ_x , μ_y , σ_x , σ_y and ρ with their standard errors are all functions of the restricted ML estimate $\hat{\boldsymbol{\pi}}$ and will be discussed briefly. The ML estimates for the μ 's and σ 's follow from (11.67) and (11.68) and according to the multivariate delta theorem

$$\hat{\boldsymbol{\beta}}_x = \begin{pmatrix} \hat{\mu}_x \\ \hat{\sigma}_x \end{pmatrix} \cong N \left(\begin{pmatrix} \mu_x \\ \sigma_x \end{pmatrix}, \mathbf{B}_x \text{Cov}(\hat{\boldsymbol{\alpha}}_x) \mathbf{B}'_x \right) \quad (11.72)$$

and

$$\hat{\boldsymbol{\beta}}_y = \begin{pmatrix} \hat{\mu}_y \\ \hat{\sigma}_y \end{pmatrix} \cong N \left(\begin{pmatrix} \mu_y \\ \sigma_y \end{pmatrix}, \mathbf{B}_y \text{Cov}(\hat{\boldsymbol{\alpha}}_y) \mathbf{B}'_y \right). \quad (11.73)$$

The matrices of derivatives in (11.72) and (11.73) are

$$\mathbf{B}_x = \frac{\partial \boldsymbol{\beta}_x}{\partial \boldsymbol{\alpha}_x} = \begin{pmatrix} -\frac{\alpha_{x2}}{\alpha_{x1}^2} & \frac{1}{\alpha_{x1}} \\ -\frac{1}{\alpha_{x1}^2} & 0 \end{pmatrix}$$

and

$$\mathbf{B}_y = \frac{\partial \boldsymbol{\beta}_y}{\partial \boldsymbol{\alpha}_y} = \begin{pmatrix} -\frac{\alpha_{y2}}{\alpha_{y1}^2} & \frac{1}{\alpha_{y1}} \\ -\frac{1}{\alpha_{y1}^2} & 0 \end{pmatrix}.$$

The only parameter that remains is ρ and is estimated from

$$\begin{aligned} \hat{\rho} &= \sin \left(\frac{\pi}{2} \left[(\widehat{\text{VOL}}1 + \widehat{\text{VOL}}4) - (\widehat{\text{VOL}}2 + \widehat{\text{VOL}}3) \right] \right) \\ &= \sin \left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1} \hat{\boldsymbol{\pi}} \right). \end{aligned} \quad (11.74)$$

In (11.74) the restricted ML estimates of the relative frequencies of the 4 quadrants are simply added to obtain the ML estimates for the 4 so-called volumes. For the cells belonging to more than one quadrant, the relative frequencies are added proportionately to the fitted bivariate normal distribution. This requires that $\hat{\rho}$ is to be solved iteratively over (11.74) beginning at any starting point, say $\hat{\rho} = 0$ until convergence leads to the unique ML estimate for ρ . The variance of $\hat{\rho}$ follows

$$\text{Var}(\hat{\rho}) = \left(\frac{\partial \rho}{\partial \mathbf{p}} \right) \mathbf{V} \left(\frac{\partial \rho}{\partial \mathbf{p}} \right)' \quad (11.75)$$

where

$$\frac{\partial \rho}{\partial \mathbf{p}} = \cos \left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1} \mathbf{p} \right) \cdot \left(\frac{\pi}{2} [(\mathbf{v}'_1 + \mathbf{v}'_4) - (\mathbf{v}'_2 + \mathbf{v}'_3)] \mathbf{C}^{-1} \right).$$

11.7 Goodness of fit

Since the vector of constraints in

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{g}_x(\boldsymbol{\pi}) \\ \mathbf{g}_y(\boldsymbol{\pi}) \\ \mathbf{g}_{xy}(\boldsymbol{\pi}) \end{pmatrix} = \mathbf{0}$$

consists out of $(I - 1) + (J - 1) + (I - 2)(J - 2)$ linear independent constraints, the degrees of freedom for the Pearson χ^2 statistic

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}} \quad (11.76)$$

and the Wald statistic

$$\mathbf{W} = \mathbf{g}(\mathbf{p})' (\mathbf{G}_p \mathbf{V} \mathbf{G}_p')^{-1} \mathbf{g}(\mathbf{p})$$

is

$$df = IJ - I - J + 2. \quad (11.77)$$

In (11.76) p_{ij} for $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$ is the observed cumulative relative frequency in the (i, j) -th cell (see (9.11)) and in matrix notation the observed cumulative relative frequencies may be represented as

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1,J-1} & p_{1J} \\ p_{21} & p_{22} & \cdots & p_{2,J-1} & p_{2J} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ p_{I-1,1} & p_{I-1,2} & \cdots & p_{I-1,J-1} & p_{I-1,J} \\ p_{I1} & p_{I2} & \cdots & p_{I,J-1} & p_{IJ} \end{pmatrix}. \quad (11.78)$$

The elements of \mathbf{P} are also referred to as the unrestricted ML estimates of $\boldsymbol{\pi}$. The elements of $\hat{\boldsymbol{\Pi}}$ in (11.66) are the restricted ML estimates of $\boldsymbol{\pi}$ obtained from the ML estimation procedure and satisfies the properties of the bivariate normal distribution.

Chapter 12

Application

The association between Grade 12 Mathematics (MATHS) and first year Statistics (STATS) is investigated. First year students who had Mathematics on HG and who were enrolled for Statistics for the first time in 2004 were included in the sample. The results are shown in Table 12.1.

Table 12.1: Two-way contingency table of 746 first year students, row percentages in brackets.

MATHS (x)	STATS (y)				Total
	0-49	50-59	60-74	75-100	
0-59	106 (44.92%)	90 (38.14%)	35 (14.83%)	5 (2.12%)	236
60-69	57 (27.01%)	73 (34.60%)	59 (27.96%)	22 (10.43%)	211
70-79	15 (10.79%)	40 (28.78%)	57 (41.01%)	27 (19.42%)	139
80-100	2 (1.25%)	14 (8.75%)	45 (28.13%)	99 (61.88%)	160
Total	180 (24.13%)	217 (29.09%)	196 (26.27%)	153 (20.51%)	746 (100%)

The row percentages in Table 12.1 reveal a definite interaction structure between MATHS and STATS. Low MATHS marks correspond with low STATS marks and vice versa, identifying a positive correlation between the two variables. The Pearson χ^2 test of independence, $\chi^2 = 326$ ($df = 9$, p value < 0.001), shows a very strong association between the two variables.

Traditionally researchers might have been tempted to use the class midpoint as an estimate for the values within a particular class interval. By using this approach the sample correlation coefficient is

$$r = 0.5495 \quad (12.1)$$

with an estimated regression line of

$$\hat{y} = 25.8 + 0.5187x . \quad (12.2)$$

Since we are dealing with a bivariate grouped data set, the basic assumptions for applying these statistical techniques are not met and the results obtained in (12.1) and (12.2) might be incorrect.

In this chapter a bivariate normal distribution will be fitted to the data in Table 12.1. It is justified to assume that MATHS (x) and STATS (y) are jointly normally distributed and therefore the estimation of the correlation structure between these two variables may be done more effectively by fitting a bivariate normal distribution. By doing this, the complete underlying bivariate continuous structure between the two variables will be taken into account.

12.1 ML estimation procedure

The vectors of upper class boundaries are

$$\mathbf{x} = \begin{pmatrix} 59.5 \\ 69.5 \\ 79.5 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 49.5 \\ 59.5 \\ 74.5 \end{pmatrix} \quad (12.3)$$

respectively.

The projection matrix for

$$\mathbf{X} = \begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix} = \begin{pmatrix} 59.5 & -1 \\ 69.5 & -1 \\ 79.5 & -1 \end{pmatrix}$$

is

$$\mathbf{P}_X = \begin{pmatrix} 0.83333 & 0.33333 & -0.16667 \\ 0.33333 & 0.33333 & 0.33333 \\ -0.16667 & 0.33333 & 0.83333 \end{pmatrix} \quad (12.4)$$

and the projection matrix for

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y} & -\mathbf{1} \end{pmatrix} = \begin{pmatrix} 49.5 & -1 \\ 59.5 & -1 \\ 74.5 & -1 \end{pmatrix}$$

is

$$\mathbf{P}_Y = \begin{pmatrix} 0.76316 & 0.39474 & -0.15789 \\ 0.39474 & 0.34211 & 0.26316 \\ -0.15789 & 0.26316 & 0.89474 \end{pmatrix}. \quad (12.5)$$

These two projection matrices play a major role in the estimation of the bivariate normal distribution, since the standardised upper class boundaries are estimated such that \mathbf{z}_x is in the vector space generated by \mathbf{X} and \mathbf{z}_y is in the vector space generated by \mathbf{Y} .

A step by step explanation of the results during the iterative procedure will be presented to give more insight into the ML estimation procedure.

- Firstly, the estimates for the unrestricted ML estimate \mathbf{p} will be given. The vector \mathbf{p} is the observed vector of cumulative relative frequencies and is used as the starting point for the iterative ML estimation procedure.
- Secondly the estimates for the restricted ML estimate $\hat{\boldsymbol{\pi}}$ will be given. The estimates obtained from $\hat{\boldsymbol{\pi}}$ are the ML estimates for the bivariate normal distribution. This follows since the vector $\hat{\boldsymbol{\pi}}$ is the ML estimate of $\boldsymbol{\pi}$ under the constraints (11.31), obtained iteratively from the ML estimation procedure.

12.1.1 Unrestricted estimates

The observed frequencies are elements of

$$\mathbf{F} = \begin{pmatrix} 106 & 90 & 35 & 5 \\ 57 & 73 & 59 & 22 \\ 15 & 40 & 57 & 27 \\ 2 & 14 & 45 & 99 \end{pmatrix} \quad (12.6)$$

and the matrix with unrestricted (observed) cumulative relative frequencies is

$$\mathbf{P} = \begin{pmatrix} 0.14209 & 0.26273 & 0.30965 & 0.31635 \\ 0.21850 & 0.43700 & 0.56300 & 0.59920 \\ 0.23861 & 0.51072 & 0.71314 & 0.78552 \\ 0.24129 & 0.53217 & 0.79491 & 1.00000 \end{pmatrix}. \quad (12.7)$$

Marginal distribution of MATHS

The unrestricted estimates for the marginal distribution of MATHS are tabulated in Table 12.2 and will be discussed briefly.

Table 12.2: Unrestricted estimates obtained from the marginal distribution of \mathbf{x} .

\mathbf{p}_x	$\hat{\boldsymbol{\alpha}}_x$	$\hat{\mu}_x$	$\hat{\sigma}_x$	$\hat{\mathbf{z}}_x$
$\begin{pmatrix} 0.31635 \\ 0.59920 \\ 0.78552 \end{pmatrix}$	$\begin{pmatrix} 0.06345 \\ 4.22132 \end{pmatrix}$	66.535079	15.76167	$\begin{pmatrix} -0.44634 \\ 0.18811 \\ 0.82256 \end{pmatrix}$

Note: The elements of \mathbf{p}_x are elements contained in the last column of \mathbf{P} (12.7).

Since the marginal distribution for MATHS has to follow a normal distribution, the vector of standardised upper class boundaries for \mathbf{x} follows by projecting $\Phi^{-1}(\mathbf{p}_x)$ into the vector space of \mathbf{X}

$$\hat{\mathbf{z}}_x = \mathbf{P}_X \Phi^{-1}(\mathbf{p}_x) \quad (12.8)$$

and is employed in the vector of constraints

$$\mathbf{g}_x(\boldsymbol{\pi}) = \boldsymbol{\Phi}(\hat{\mathbf{z}}_x) - \boldsymbol{\pi}_x = \mathbf{0} . \quad (12.9)$$

In SAS IML: $\boldsymbol{\Phi}(z_x) = \text{PROBNORM}(z_x)$

The unrestricted estimate for the vector of natural parameters

$$\hat{\boldsymbol{\alpha}}_x = \begin{pmatrix} 1 \\ \hat{\sigma}_x \\ \hat{\mu}_x \\ \hat{\sigma}_x \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Phi}^{-1}(\mathbf{p}_x) \quad (12.10)$$

leads to the unrestricted estimates for $\hat{\mu}_x$ and $\hat{\sigma}_x$ indicating that the average mark for MATHS is 66.5 with a standard deviation of 15.8.

Marginal distribution of STATS

The unrestricted estimates for the marginal distribution of STATS are tabulated in Table 12.3.

Table 12.3: Unrestricted estimates obtained from the marginal distribution of \mathbf{y} .

\mathbf{p}_y	$\hat{\boldsymbol{\alpha}}_y$	$\hat{\mu}_y$	$\hat{\sigma}_y$	$\hat{\mathbf{z}}_y$
$\begin{pmatrix} 0.24129 \\ 0.53217 \\ 0.79491 \end{pmatrix}$	$\begin{pmatrix} 0.06021 \\ 3.61002 \end{pmatrix}$	60.04601	16.63317	$\begin{pmatrix} -0.63404 \\ -0.03283 \\ 0.86899 \end{pmatrix}$

Note: The elements of \mathbf{p}_y are elements contained in the last row of \mathbf{P} (12.7).

Following the same rationale for the standardised upper class boundaries for \mathbf{y} , the vector of standardised upper class boundaries

$$\hat{\mathbf{z}}_y = \mathbf{P}_y \boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\pi}}_y) \quad (12.11)$$

is employed in the vector of constraints

$$\mathbf{g}_y(\boldsymbol{\pi}) = \boldsymbol{\Phi}(\mathbf{z}_y) - \boldsymbol{\pi}_y = \mathbf{0} . \quad (12.12)$$

In SAS IML: $\Phi(z_y) = \text{PROBNORM}(z_y)$

At this initial step of the iterative procedure it follows from Table 12.3 that the average mark for STATS is 60.5, with a standard deviation of 16.6.

Joint distribution of MATHS and STATS

From the estimates of the standardised upper class boundaries (see Table 12.2 and Table 12.3) it follows that the origin $(\hat{z}_x, \hat{z}_y) = (0, 0)$ is located in the second class interval for MATHS and the third class interval for STATS. In Figure 12.1 a contour diagram of the bivariate normal distribution with the four quadrants and the standardised upper class boundaries is shown.

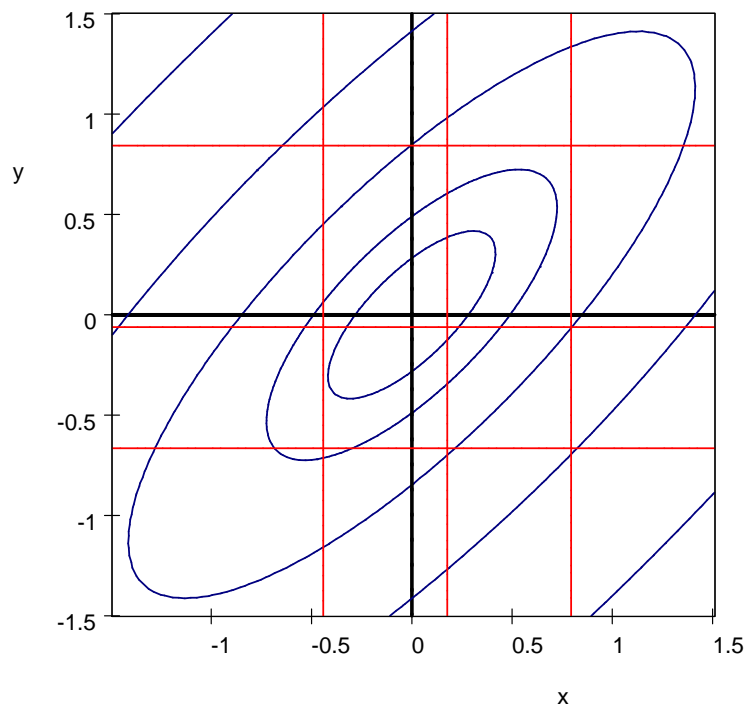


Figure 12.1: Contour diagram of the bivariate normal distribution with the four quadrants and the standardised upper class boundaries.

The ML estimator for ρ is obtained from

$$\hat{\rho} = \sin \left(\frac{\pi}{2} \left[\left(\widehat{\text{VOL1}} + \widehat{\text{VOL4}} \right) - \left(\widehat{\text{VOL2}} + \widehat{\text{VOL3}} \right) \right] \right) . \quad (12.13)$$

The volumes are estimated by the total of the observed relative frequencies located in each of the four quadrants. In matrix notation the observed relative frequencies are

$$\frac{1}{746} \mathbf{F} = \begin{pmatrix} 0.14209 & 0.12064 & 0.04692 & 0.00670 \\ 0.07641 & 0.09786 & 0.07909 & 0.02949 \\ 0.02011 & 0.05362 & 0.07641 & 0.03619 \\ 0.00268 & 0.01877 & 0.06032 & 0.13271 \end{pmatrix} . \quad (12.14)$$

For those cells situated in only one quadrant, the relative frequencies can simply be added, but for cells situated in more than one quadrant, allocation has to be done proportionately to the bivariate normal distribution, thus depending on the value of $\hat{\rho}$. Since $\hat{\rho}$ is to be estimated, the value of $\hat{\rho}$ is obtained iteratively over (12.13), starting at any value between -1 and 1. In Table 12.4 various starting points for $\hat{\rho}$ were being used, all leading to the same unique unrestricted estimate for ρ . (Convergence criterion = 1e-10.)

Table 12.4: Unrestricted estimate for ρ obtained iteratively

Starting point $\hat{\rho} = -0.5$	Starting point $\hat{\rho} = 0$	Starting point $\hat{\rho} = 0.5$
1. 0.6128852	1. 0.6383751	1. 0.6616935
2. 0.6708946	2. 0.6735298	2. 0.6761977
3. 0.6773286	3. 0.6776614	3. 0.6780025
4. 0.6781484	4. 0.6781915	4. 0.6782358
5. 0.6782547	5. 0.6782603	5. 0.6782661
6. 0.6782685	6. 0.6782692	6. 0.6782700
7. 0.6782703	7. 0.6782704	7. 0.6782705
8. 0.6782705	8. 0.6782706	8. 0.6782706
9. 0.6782706	9. 0.6782706	9. 0.6782706
10. 0.6782706	10. 0.6782706	10. 0.6782706
11. 0.6782706	18. 0.6782706	11. 0.6782706

Evaluating the estimates for the 4 volumes of the bivariate normal distribution in Table 12.5, it is clear that the property of symmetry has not been met and ρ is now estimated from the observed frequencies (unrestricted ML estimate for π).

Table 12.5: Unrestricted estimates for the volumes of the four quadrants

Quadrant	Unrestricted estimates for VOL
$Q_1 : z_x < 0, z_y < 0$	$\widehat{VOL1} = 0.3979789$
$Q_2 : z_x < 0, z_y > 0$	$\widehat{VOL2} = 0.1177174$
$Q_3 : z_x > 0, z_y < 0$	$\widehat{VOL3} = 0.1450123$
$Q_4 : z_x > 0, z_y > 0$	$\widehat{VOL4} = 0.3392913$

From Table 12.5 it follows that

$$\begin{aligned}
 \hat{\rho} &= \sin\left(\frac{\pi}{2} [(0.3979789 + 0.3392913) - (0.1177174 + 0.1450123)]\right) \\
 &= \sin\left(\frac{\pi}{2} [0.73727 - 0.26273]\right) \\
 &= \sin\left(\frac{\pi}{2} [0.47454]\right) \\
 &= 0.67827
 \end{aligned} \tag{12.15}$$

indicating a positive relationship between MATHS and STATS.

This estimate for ρ is now being used in the vector of constraints $\mathbf{g}_{xy}(\pi) = \mathbf{0}$ where

$$\begin{aligned}
 \mathbf{g}_{xy}(\pi) &= \Phi_{xy} - \mathbf{p}_{xy} \\
 &= \Phi((\hat{\mathbf{z}}_x \otimes \mathbf{1}_4), (\mathbf{1}_4 \otimes \hat{\mathbf{z}}_y), \hat{\rho}) - \text{vec} \begin{pmatrix} 0.14209 & 0.26273 & 0.30965 \\ 0.21850 & 0.43700 & 0.56300 \\ 0.23861 & 0.51072 & 0.71314 \end{pmatrix}
 \end{aligned}$$

In SAS IML: $\Phi(z_x, z_y, \rho) = \text{PROBBNRM}(z_x, z_y, \rho)$

12.1.2 ML estimates

After convergence of the ML estimation procedure the restricted ML estimate for π in matrix notation is

$$\hat{\Pi} = \begin{pmatrix} 0.17170 & 0.25637 & 0.31180 & 0.31869 \\ 0.22874 & 0.38802 & 0.53638 & 0.56766 \\ 0.25077 & 0.45896 & 0.70922 & 0.79166 \\ 0.25569 & 0.48298 & 0.81013 & 1.00000 \end{pmatrix} \quad (12.16)$$

and possesses all the properties of the bivariate normal distribution. The matrix of expected frequencies is

$$\mathbf{M} = \begin{pmatrix} 128.0903 & 63.1637 & 41.3464 & 5.1455 \\ 42.5489 & 55.6571 & 69.3300 & 18.1950 \\ 16.4324 & 36.4943 & 76.0140 & 38.1641 \\ 3.6702 & 14.2495 & 57.3636 & 80.1349 \end{pmatrix} \quad (12.17)$$

and according to the Pearson and Wald statistics tabulated in Table 12.6, the bivariate normal distribution did not provide an extremely good fit.

Table 12.6: Goodness of fit statistics

Statistic	Value	df	<i>p</i> -value
Pearson	45.191	10	2.0089E-6
Wald	44.994	10	2.1799E-6

However, taking into account the rather large sample size, the measure of discrepancy

$$\mathbf{D} = \frac{\mathbf{W}}{n} = \frac{44.994}{746} = 0.06 \quad (12.18)$$

is only just higher than the cut off value of 0.05, suggesting that the fit is not too poor. This is further motivated by comparing the observed frequencies in \mathbf{F} (12.6) with the expected frequencies in \mathbf{M} (12.17).

Marginal distribution of MATHS

The ML estimates obtained from the marginal distribution of \mathbf{x} are tabulated in Table 12.7.

Table 12.7: ML estimates for the marginal distribution of \mathbf{x} .

$\hat{\boldsymbol{\pi}}_x$	$\hat{\boldsymbol{\alpha}}_x$	$\hat{\mu}_x$	$\hat{\sigma}_x$	$\hat{\mathbf{z}}_x$
$\begin{pmatrix} 0.31869 \\ 0.56766 \\ 0.79166 \end{pmatrix}$	$\begin{pmatrix} 0.06418 \\ 4.28995 \end{pmatrix}$	66.84445	15.58162	$\begin{pmatrix} -0.47135 \\ 0.17043 \\ 0.81221 \end{pmatrix}$

Note: The elements of $\hat{\boldsymbol{\pi}}_x$ are elements contained in the last column row of $\hat{\boldsymbol{\Pi}}$ (12.16).

The marginal cumulative relative frequencies $\hat{\boldsymbol{\pi}}_x$ follow a cumulative normal distribution at the upper class boundaries \mathbf{x} and therefore

$$\hat{\boldsymbol{\Phi}}_x = \boldsymbol{\Phi}(\hat{\mathbf{z}}_x) = \boldsymbol{\Phi} \begin{pmatrix} -0.47135 \\ 0.17043 \\ 0.81221 \end{pmatrix} = \begin{pmatrix} 0.31869 \\ 0.56766 \\ 0.79166 \end{pmatrix} = \hat{\boldsymbol{\pi}}_x . \quad (12.19)$$

The estimated standard errors for $\hat{\mu}_x$ and $\hat{\sigma}_x$ are

$$\hat{\sigma}_{\hat{\mu}_x} = 0.62047 \quad \text{and} \quad \hat{\sigma}_{\hat{\sigma}_x} = 0.67075 \quad (12.20)$$

and therefore a 95% confidence interval for μ_x is

$$(65.628, 68.061) .$$

Marginal distribution of STATS

The ML estimates obtained from the marginal distribution of y are tabulated in Table 12.8.

Table 12.8: ML estimates for the marginal distribution of y .

$\hat{\pi}_y$	$\hat{\alpha}_y$	$\hat{\mu}_y$	$\hat{\sigma}_y$	\hat{z}_y
$\begin{pmatrix} 0.25569 \\ 0.48298 \\ 0.81013 \end{pmatrix}$	$\begin{pmatrix} 0.06140 \\ 3.69619 \end{pmatrix}$	60.19482	16.28563	$\begin{pmatrix} -0.65670 \\ -0.04266 \\ 0.87839 \end{pmatrix}$

Note: The elements of $\hat{\pi}_y$ are elements contained in the last row of $\hat{\Pi}$ (12.16).

Similarly to the marginal distribution of x , it follows that the marginal cumulative relative frequencies $\hat{\pi}_y$ follow a cumulative normal distribution at the upper class boundaries of y

$$\hat{\Phi}_y = \Phi(\hat{z}_y) = \Phi \begin{pmatrix} -0.65670 \\ -0.04266 \\ 0.87839 \end{pmatrix} = \begin{pmatrix} 0.25569 \\ 0.48298 \\ 0.81013 \end{pmatrix} = \hat{\pi}_y . \quad (12.21)$$

The estimated standard errors for $\hat{\mu}_y$ and $\hat{\sigma}_y$ are

$$\hat{\sigma}_{\hat{\mu}_y} = 0.63940 \quad \text{and} \quad \hat{\sigma}_{\hat{\sigma}_y} = 0.64606 . \quad (12.22)$$

and may be used for inferential purposes.

Joint distribution of MATHS and STATS

The joint cumulative relative frequencies at the intersections of the standardised upper class boundaries are equal to the probabilities of the bivariate normal distribution i.e.

$$\widehat{\Phi}_{xy} = \text{vec}(\Phi(\widehat{\mathbf{z}}_x, \widehat{\mathbf{z}}'_y)) = \text{vec} \begin{pmatrix} 0.17170 & 0.25637 & 0.31180 \\ 0.22874 & 0.38802 & 0.53638 \\ 0.25077 & 0.45896 & 0.70922 \end{pmatrix} = \widehat{\boldsymbol{\pi}}_{xy} .$$

Note: The elements of $\widehat{\boldsymbol{\pi}}_{xy}$ are the first $(I - 1)(J - 1)$ elements contained in $\widehat{\boldsymbol{\Pi}}$ (12.16).

The ML estimate for ρ is estimated by adding the appropriate relative frequencies under constraints

$$\frac{1}{746} \mathbf{M} = \begin{pmatrix} 0.17170 & 0.08467 & 0.05542 & 0.00690 \\ 0.05704 & 0.07461 & 0.09294 & 0.02439 \\ 0.02203 & 0.04892 & 0.10190 & 0.05116 \\ 0.00492 & 0.01910 & 0.07690 & 0.10742 \end{pmatrix} \quad (12.23)$$

(see (12.17)). The symmetrical nature of the fitted bivariate normal distribution is portrayed by Table 12.9.

Table 12.9: ML estimates for the volumes of the four quadrants

Quadrant	ML estimates for VOL
$Q_1 : z_x < 0, z_y < 0$	$\widehat{\text{VOL}}1 = 0.366415$
$Q_2 : z_x < 0, z_y > 0$	$\widehat{\text{VOL}}2 = 0.133585$
$Q_3 : z_x > 0, z_y < 0$	$\widehat{\text{VOL}}3 = 0.133585$
$Q_4 : z_x > 0, z_y > 0$	$\widehat{\text{VOL}}4 = 0.366415$

The ML estimate for ρ is

$$\begin{aligned}\hat{\rho} &= \sin\left(\frac{\pi}{2}[2(0.366415) - 2(0.133585)]\right) \\ &= \sin\left(\frac{\pi}{2}[0.73283 - 0.26717]\right) \\ &= \sin\left(\frac{\pi}{2}[0.46566]\right) \\ &= 0.66795\end{aligned}\tag{12.24}$$

with a standard error of

$$\hat{\sigma}_{\hat{\rho}} = 0.0303 .\tag{12.25}$$

Since

$$t = \frac{\hat{\rho}}{\hat{\sigma}_{\hat{\rho}}} = 22\tag{12.26}$$

the null hypothesis of $H_0 : \rho = 0$ is rejected, indicating a significant association between MATHS and STATS.

The estimated regression line of STATS (y) on MATHS (x) is

$$\hat{y}_{y|x} = \hat{\alpha}_{y|x} + \hat{\beta}_{y|x}x$$

where

$$\begin{aligned}\hat{\alpha}_{y|x} &= \hat{\mu}_y - \left(\hat{\rho}\frac{\hat{\sigma}_y}{\hat{\sigma}_x}\right)\hat{\mu}_x \\ &= 12.528\end{aligned}$$

is the intercept and

$$\begin{aligned}\hat{\beta}_{y|x} &= \hat{\rho}\frac{\hat{\sigma}_y}{\hat{\sigma}_x} \\ &= 0.6981\end{aligned}$$

is the slope, yielding the regression equation

$$\hat{y}_{y|x} = 13.5 + 0.70x .\tag{12.27}$$

According to this regression line it is clear that for every increase of 1% in MATHS, the STATS mark increases with 0.7%. The estimated correlation coefficient and regression equation for the

fitted bivariate normal distribution, differ substantially from that where the class midpoint values were used as an estimate for the values within a class interval emphasizing the importance of the technique. Compare with (12.1) and (12.2).

All the results for this application were obtained from the SAS program *BVN.SAS* listed in Appendix C3.