# Contents

# Chapter 1

# Introduction

In many situations, data are only available in a grouped form. Typical continuous variables such as income, age, test scores and many more are for various reasons classified into a few class intervals. The implication is that the usual statistical techniques employed for continuous variables can no longer be applied in the usual sense. Often when researchers are confronted with grouped data, the underlying continuous nature of the variable is ignored and the data do not comply to the requirements of the statistical tests applied.

The maximum likelihood (ML) estimation procedure of *Matthews and Crowther (1995)* will be utilized to fit a continuous distribution to a grouped data set. This grouped data set may be a single frequency distribution or various frequency distributions that arise from a cross classification of several factors in a multifactor design. It will also be shown how to fit a bivariate normal distribution to a two-way contingency table where the two underlying continuous variables are jointly normally distributed.

This thesis is organized in three different parts, each playing a vital role in the explanation of analysing grouped data with the ML estimation of *Matthews and Crowther*. All the examples, applications and simulations are done with the SAS procedure IML, listed in the Appendix.

# Part I

The ML estimation procedure of *Matthews and Crowther* is formulated. This procedure plays an integral role and is implemented in all three parts of the thesis. In Part I the exponential distribution is fitted to a grouped data set to explain the technique. Two different formulations of the constraints are employed in the ML estimation procedure and provide identical results. The justification of the method is further motivated by a simulation study. Similar to the exponential distribution, the estimation of the normal distribution is also explained in detail. Part I is summarized in Chapter 5 where a general method is outlined to fit continuous distributions to a grouped data set. Distributions such as the Weibull, the log-logistic and the Pareto distributions can be fitted very effectively by formulating the vector of constraints in terms of a linear model.

# Part II

In Part II it is explained how to model a grouped response variable in a multifactor design. This multifactor design arise from a cross classification of the various factors or independent variables to be analysed. The cross classification of the factors results in a total of $T$ cells, each containing a frequency distribution. Distribution fitting is done simultaneously to each of the $T$ cells of the multifactor design. Distribution fitting is also done under the additional constraints that the parameters of the underlying continuous distributions satisfy a certain structure or design. The effect of the factors on the grouped response variable may be evaluated from this fitted design. Applications of a single-factor and a two-factor model are considered to demonstrate the versatility of the technique.

# Part III

A two-way contingency table where the two variables have an underlying bivariate normal distribution is considered. The estimation of the bivariate normal distribution reveals the complete underlying continuous structure between the two variables. The ML estimate of the correlation coefficient $\rho$ is used to great effect to describe the relationship between the two variables. Apart from an application a simulation study is also provided to support the method proposed.

# Part I

# Fitting distributions to grouped data

# Chapter 2

# The ML estimation procedure

In this chapter the ML estimation procedure of *Matthews and Crowther (1995)* is presented. This procedure is employed to find the ML estimates in the statistical analysis of grouped data. The formulation and explanation of the ML estimation procedure described in this chapter will be used throughout the thesis.

## 2.1 Formulation

Consider a total of $n$ observations tabulated in a frequency distribution with $k$ classes.

**Table 2.1:** General formulation of a frequency distribution.

| Class Interval | Frequency |
|---|---|
| $(-\infty, x_1)$ | $f_1$ |
| $[x_1, x_2)$ | $f_2$ |
| $\vdots$ | $\vdots$ |
| $[x_{k-2}, x_{k-1})$ | $f_{k-1}$ |
| $[x_{k-1}, \infty)$ | $f_k$ |

The observations in Table 2.1 originate from a continuous distribution and information concerning the distribution is now only available in grouped form. In Table 2.1 the first and last intervals of the frequency distribution may be open ended class intervals.

Denote the vector of the first $(k-1)$ frequencies in Table 2.1 by

$$\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{k-1} \end{pmatrix} \tag{2.1}$$

with corresponding vector of upper class boundaries

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \end{pmatrix} . \tag{2.2}$$

It is assumed that the vector $\mathbf{f}$ is a random vector with some discrete distribution such as Poisson, multinomial or product multinomial. Assume multinomial sampling and define

$$\mathbf{p}_0 = \frac{1}{n}\mathbf{f} \tag{2.3}$$

as the vector of relative frequencies. Let $\boldsymbol{\pi}_0$ denote the vector of probabilities, where the $i$-th element of $\boldsymbol{\pi}_0$ is the probability that an observation falls in the $i$-th class interval. Hence, the expected value of $\mathbf{p}_0$ is

$$\mathrm{E}(\mathbf{p}_0) = \boldsymbol{\pi}_0 \tag{2.4}$$

with covariance matrix

$$\mathrm{Cov}(\mathbf{p}_0) = \frac{1}{n}\left(\mathrm{diag}\left[\boldsymbol{\pi}_0\right] - \boldsymbol{\pi}_0\boldsymbol{\pi}_0'\right) = \mathbf{V}_0 \tag{2.5}$$

where $\mathrm{diag}\left[\boldsymbol{\pi}_0\right]$ is a diagonal matrix with the elements of $\boldsymbol{\pi}_0$ on the diagonal.

The vector of cumulative relative frequencies is denoted by

$$\mathbf{p} = \mathbf{C}\mathbf{p}_0 \tag{2.6}$$

where $\mathbf{C}$ is a triangular matrix such that

$$\mathbf{C} : (k-1) \times (k-1) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} . \tag{2.7}$$

The expected value of $\mathbf{p}$ is

$$\begin{aligned} \mathrm{E}(\mathbf{p}) &= \mathbf{C}\boldsymbol{\pi}_0 \\ &= \boldsymbol{\pi} \end{aligned} \tag{2.8}$$

with covariance matrix

$$\begin{aligned} \mathrm{Cov}(\mathbf{p}) &= \mathbf{C}\mathbf{V}_0\mathbf{C}' \\ &= \mathbf{C}\left\{\frac{1}{n}(\mathrm{diag}\,[\boldsymbol{\pi}_0] - \boldsymbol{\pi}_0\boldsymbol{\pi}_0')\right\}\mathbf{C}' \\ &= \frac{1}{n}\left\{\mathbf{C}\,\mathrm{diag}\left[\mathbf{C}^{-1}\boldsymbol{\pi}\right]\mathbf{C}' - \boldsymbol{\pi}\boldsymbol{\pi}'\right\} \\ &= \mathbf{V} . \end{aligned} \tag{2.9}$$

## 2.2 Estimation

The frequency vector $\mathbf{f}$ is distributed according to a multinomial distribution and consequently belongs to the exponential class. Since the vector of cumulative relative frequencies is a one-to-one transformation of $\mathbf{f}$, the random vector $\mathbf{p}$ may be implemented in the ML estimation procedure of *Matthews and Crowther (1995)* presented in Proposition 1. Utilizing the ML estimation, it is possible to find the ML estimate of $\boldsymbol{\pi}$, under the restriction that $\boldsymbol{\pi}$ satisfies the constraints defined in the ML estimation procedure.

The basic foundation of this research are given in the following two propositions. The proofs are given in Matthews and Crowther (1995).

**Proposition 1** *(ML estimation procedure)*
*Consider a random vector of cumulative relative frequencies* $\mathbf{p}$*, which may be considered as a non-singular (one-to-one) transformation of the canonical vector of observations, belonging to the exponential family, with*

$$\mathrm{E}(\mathbf{p}) = \boldsymbol{\pi} \quad \textit{and} \ \mathrm{Cov}(\mathbf{p}) = \mathbf{V} \ .$$

*The observed* $\mathbf{p}$ *is the unrestricted ML estimate of* $\boldsymbol{\pi}$ *and the covariance matrix* $\mathbf{V}$ *may be a function of* $\boldsymbol{\pi}$*. Let* $\mathbf{g}(\boldsymbol{\pi})$ *be a continuous vector valued function of* $\boldsymbol{\pi}$*, for which the first order partial derivatives,*

$$\mathbf{G}_{\pi} = \frac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \tag{2.10}$$

*with respect to* $\boldsymbol{\pi}$ *exist. The ML estimate of* $\boldsymbol{\pi}$*, subject to the constraints* $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$ *is obtained iteratively from*

$$\widehat{\boldsymbol{\pi}} = \mathbf{p} - (\mathbf{G}_{\pi}\mathbf{V})' \left(\mathbf{G}_{p}\mathbf{V}\mathbf{G}_{\pi}'\right)^{*} \mathbf{g}(\mathbf{p}) \tag{2.11}$$

*where* $\mathbf{G}_{p} = \left.\dfrac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right|_{\pi = p}$ *and* $\left(\mathbf{G}_{p}\mathbf{V}\mathbf{G}_{\pi}'\right)^{*}$ *is a generalized inverse of* $\left(\mathbf{G}_{p}\mathbf{V}\mathbf{G}_{\pi}'\right)$*.*

The iterative procedure implies a double iteration over $\mathbf{p}$ and $\boldsymbol{\pi}$. The procedure starts with the unrestricted ML estimate of $\boldsymbol{\pi}$, as the starting value for both $\mathbf{p}$ and $\boldsymbol{\pi}$. Convergence is first obtained over $\mathbf{p}$ using (2.11). The converged value of $\mathbf{p}$ is then used as the next value of $\boldsymbol{\pi}$, with convergence over $\mathbf{p}$ starting again at the observed $\mathbf{p}$. In this procedure $\mathbf{V}$ is recalculated for each new value of $\boldsymbol{\pi}$ in the iterative procedure. Convergence over $\boldsymbol{\pi}$ ultimately leads to $\widehat{\boldsymbol{\pi}}$, the restricted ML estimate of $\boldsymbol{\pi}$ .

**Proposition 2** *The asymptotic covariance matrix of* $\widehat{\boldsymbol{\pi}}$*, under* $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$*, is*

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\pi}}\right) \approxeq \mathbf{V} - (\mathbf{G}_{\pi}\mathbf{V})' \left(\mathbf{G}_{\pi}\mathbf{V}\mathbf{G}_{\pi}'\right)^{*} (\mathbf{G}_{\pi}\mathbf{V}) \tag{2.12}$$

*which is estimated by replacing* $\boldsymbol{\pi}$ *by* $\widehat{\boldsymbol{\pi}}$*.*

In *Matthews and Crowther (1995)* it is assumed that the restrictions are linearly independent, but in *Matthews and Crowther (1998)*, it is shown that if the restrictions are linearly dependent, it leads to the generalized inverse, $\left(\mathbf{G}_{\pi}\mathbf{V}\mathbf{G}_{\pi}'\right)^{*}$, to be introduced in (2.11) and (2.12).

The objective is now to find the ML estimate of $\boldsymbol{\pi}$, under the constraints that the cumulative relative frequencies $\boldsymbol{\pi}$, equal the cumulative distribution curve, $F(\mathbf{x})$ at the upper class boundaries $\mathbf{x}$. This implies that the ML estimate of $\boldsymbol{\pi}$ is to be obtained under the restriction

$$F(\mathbf{x}) = \boldsymbol{\pi} \tag{2.13}$$

which means that the vector of constraints in (2.11) may be formulated as

$$\mathbf{g}(\boldsymbol{\pi}) = F(\mathbf{x}) - \boldsymbol{\pi} = \mathbf{0} \ . \tag{2.14}$$

The set of constraints in Proposition 1 is essentially not unique and may be dependent. Any function say $\mathbf{g}_1(\boldsymbol{\pi})$, that implies the same constraints on $\boldsymbol{\pi}$ as $\mathbf{g}(\boldsymbol{\pi})$, may be used and will provide the same results. The objective now is to choose $\mathbf{g}(\boldsymbol{\pi})$ in such a way to simplify the calculation of derivatives and to streamline the estimation process. In some instances it is possible to find the ML estimate of $\boldsymbol{\pi}$ under constraints, by making use of traditional methods, but the procedure suggested in Proposition 1 provides an elegant and straightforward method for obtaining the ML estimates.

## 2.3 Goodness of fit

In order to test the deviation of the observed probabilities $\mathbf{p}$ from the restricted ML estimates $\widehat{\boldsymbol{\pi}}$, imposed by the constraints $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$, it is convenient to formulate and test the null hypothesis

$$\mathbf{H}_0 : \mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$$

by some goodness of fit statistic like the Pearson $\chi^2$-statistic

$$\boldsymbol{\chi}^2 = \sum_{i=1}^{k} \frac{(p_i - \widehat{\pi}_i)^2}{\widehat{\pi}_i} \tag{2.15}$$

or the Wald statistic

$$\mathbf{W} = \mathbf{g}(\mathbf{p})'(\mathbf{G}_p \mathbf{V} \mathbf{G}_p')^* \mathbf{g}(\mathbf{p}) \ . \tag{2.16}$$

Both the Pearson and the Wald statistic have a $\chi^2$-distribution with $r$ degrees of freedom, where $r$ is equal to the number of linear independent constraints in $\mathbf{g}(\boldsymbol{\pi})$.

Another useful measure, is the measure of discrepancy

$$\mathbf{D} = \mathbf{W}/n \tag{2.17}$$

which will provide more conservative results for large sample sizes. As a rule of thumb the observed and expected frequencies are considered to not deviate significantly from each other if the discrepancy is less than 0.05.

# Chapter 3

# The exponential distribution

To illustrate the underlying methodology of fitting a distribution via the ML estimation process described in Proposition 1, it will be shown how to fit an exponential distribution to the frequency data in Table 2.1.

The probability density function (pdf) of an exponential random variable with expected value $\mu$ is given by

$$f(x;\mu) = \frac{1}{\mu} e^{-x/\mu} \tag{3.1}$$

and the cumulative distribution function (cdf) is given by

$$F(x;\mu) = 1 - e^{-x/\mu} \ . \tag{3.2}$$

To fit an exponential distribution it is required (see 2.13) that

$$\mathbf{1} - \exp(-\theta \mathbf{x}) = \boldsymbol{\pi} \tag{3.3}$$

where $\mathbf{1} : (k-1) \times 1$ is a vector of ones, $\mathbf{x}$ is the vector of upper class boundaries and $\theta = \mu^{-1}$.

From this requirement (3.3) two alternative ways of performing the estimation procedure are described. In Sections 3.1 and 3.2 it will be shown that although the specifications of the two sets of constraints, $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$, seem completely different, the final results obtained are identical.

## 3.1 Direct set of constraints

A direct set of constraints in (2.11) follows from (3.3) with

$$\mathbf{g}(\boldsymbol{\pi})=\{\mathbf{1}-\exp(-\theta\mathbf{x})\}-\boldsymbol{\pi}\ . \tag{3.4}$$

The parameter $\theta$ is expressed in terms of the cumulative probabilities in (3.3) and hence

$$\theta=-\frac{\mathbf{x}'\ln(\mathbf{1}-\boldsymbol{\pi})}{\mathbf{x}'\mathbf{x}}\ . \tag{3.5}$$

The chain rule for matrix differentiation is employed in (3.6) to obtain the following matrix of partial derivatives

$$\begin{aligned}
\mathbf{G}_\pi &= \frac{\partial\mathbf{g}(\boldsymbol{\pi})}{\partial\boldsymbol{\pi}} \\
&= \frac{\partial\left(\{\mathbf{1}-\exp(-\theta\mathbf{x})\}-\boldsymbol{\pi}\right)}{\partial\boldsymbol{\pi}} \\
&= -\frac{\partial\exp(-\theta\mathbf{x})}{\partial\boldsymbol{\pi}}-\frac{\partial\boldsymbol{\pi}}{\partial\boldsymbol{\pi}} \\
&= -\frac{\partial\exp(-\theta\mathbf{x})}{\partial\theta}\cdot\frac{\partial\theta}{\partial\boldsymbol{\pi}}-\mathbf{I} \qquad (3.6)\\
&= -\frac{\partial\begin{pmatrix}\exp(-\theta x_1)\\ \exp(-\theta x_2)\\ \vdots\\ \exp(-\theta x_{k-1})\end{pmatrix}}{\partial\theta}\cdot\left(-\frac{\mathbf{x}'}{\mathbf{x}'\mathbf{x}}\right)\cdot\frac{\partial\begin{pmatrix}\ln(1-\pi_1)\\ \ln(1-\pi_2)\\ \vdots\\ \ln(1-\pi_{k-1})\end{pmatrix}}{\partial\boldsymbol{\pi}}-\mathbf{I} \\
&= \begin{pmatrix}\exp(-\theta x_1)\cdot x_1\\ \exp(-\theta x_2)\cdot x_2\\ \vdots\\ \exp(-\theta x_{k-1})\cdot x_{k-1}\end{pmatrix}\cdot\left(-\frac{\mathbf{x}'}{\mathbf{x}'\mathbf{x}}\right)\cdot\operatorname{diag}\begin{bmatrix}-(1-\pi_1)^{-1}\\ -(1-\pi_2)^{-1}\\ \vdots\\ -(1-\pi_{k-1})^{-1}\end{bmatrix}-\mathbf{I} \\
&= -\left(\operatorname{diag}\left[\exp(-\theta\mathbf{x})\right]\right)\cdot\mathbf{P}_x\cdot\mathbf{D}_\pi-\mathbf{I} \qquad (3.7)
\end{aligned}$$

where

$$\mathbf{P}_x=\mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}' \tag{3.8}$$

is the projection matrix of $\mathbf{x}$ and

$$\mathbf{D}_\pi = \frac{\partial \ln(\mathbf{1}-\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} = -\left(\text{diag}\left[\mathbf{1}-\boldsymbol{\pi}\right]\right)^{-1} . \tag{3.9}$$

The estimation procedure in Proposition 1 utilizes a double iteration over $\boldsymbol{\pi}$ and $\mathbf{p}$ starting with the observed vector of cumulative relative frequencies as the initial values for both convergencies over $\boldsymbol{\pi}$ and $\mathbf{p}$. The iterative procedure may be summarised as follows:

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

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

$\mathbf{P}_x = \mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'$

DO OVER $\boldsymbol{\pi}$

$\boldsymbol{\pi} = \mathbf{p}$

$\mathbf{V} = \frac{1}{n}\left\{\mathbf{C}\,\text{diag}\left[\mathbf{C}^{-1}\boldsymbol{\pi}\right]\mathbf{C}' - \boldsymbol{\pi}\boldsymbol{\pi}'\right\}$

$\theta_\pi = -\dfrac{\mathbf{x}'\ln(\mathbf{1}-\boldsymbol{\pi})}{\mathbf{x}'\mathbf{x}}$

$\mathbf{D}_\pi = -\left(\text{diag}\left[\mathbf{1}-\boldsymbol{\pi}\right]\right)^{-1}$

$\mathbf{G}_\pi = -\left(\text{diag}\left[\exp\left(-\theta_\pi \mathbf{x}\right)\right]\right)\cdot \mathbf{P}_x \cdot \mathbf{D}_\pi - \mathbf{I}$

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

DO OVER $\mathbf{p}$

$\theta_p = -\dfrac{\mathbf{x}'\ln(\mathbf{1}-\mathbf{p})}{\mathbf{x}'\mathbf{x}}$

$\mathbf{D}_\pi = -\left(\text{diag}\left[\mathbf{1}-\boldsymbol{\pi}\right]\right)^{-1}$

$\mathbf{G}_p = -\left(\text{diag}\left[\exp\left(-\theta_p \mathbf{x}\right)\right]\right)\cdot \mathbf{P}_x \cdot \mathbf{D}_p - \mathbf{I}$

$\mathbf{g}(\mathbf{p}) = \left\{\mathbf{1} - \exp\left(-\theta_p \mathbf{x}\right)\right\} - \mathbf{p}$

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

END

END

From the above it follows that convergence is first obtained over $\mathbf{p}$ where the parameter $\theta_p$, the vector of constraints $\mathbf{g}(\mathbf{p})$ and the matrix of partial derivatives $\mathbf{G}_p$ are all functions of $\mathbf{p}$. Convergence over $\mathbf{p}$ leads to the next value of $\boldsymbol{\pi}$ with convergence over $\mathbf{p}$ starting again at the observed vector of cumulative relative frequencies namely $\mathbf{p}^{\dagger}$. The values of $\mathbf{V}$, $\theta_\pi$ and $\mathbf{G}_\pi$ are recalculated for every value of $\boldsymbol{\pi}$ when iterating over $\boldsymbol{\pi}$. Convergence over $\boldsymbol{\pi}$ ultimately leads to $\widehat{\boldsymbol{\pi}}$, the restricted ML estimate of $\boldsymbol{\pi}$ under $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$ with corresponding ML estimator

$$\widehat{\theta} = -\frac{\mathbf{x}' \ln(\mathbf{1} - \widehat{\boldsymbol{\pi}})}{\mathbf{x}'\mathbf{x}} \tag{3.10}$$

and hence the ML estimator for the exponential distribution

$$\widehat{\mu} = \frac{1}{\widehat{\theta}} = -\left(\frac{\mathbf{x}' \ln(\mathbf{1} - \widehat{\boldsymbol{\pi}})}{\mathbf{x}'\mathbf{x}}\right)^{-1} \tag{3.11}$$

follows. The iterative process is illustrated in Example 3.1.

**Example 3.1**

Consider $n = 100$ observations simulated from an exponential distribution with expected value $\mu = \theta^{-1} = 50$. The grouped data set is shown in Table 3.1.

**Table 3.1:** Simulated data set from an exponential distribution.

| Class interval | Frequency |
|---|---|
| $[0, 12.5)$ | 17 |
| $[12.5, 25)$ | 14 |
| $[25, 50)$ | 31 |
| $[50, 100)$ | 26 |
| $[100, \infty)$ | 12 |

Table 3.2 shows the various values of $\boldsymbol{\pi}$ and $\mathbf{p}$ in the double iteration process, with corresponding values for $\mu = \theta^{-1}$. The results in Table 3.2 can be calculated directly, or can be obtained using the SAS program *EXP1.SAS* listed in Appendix A.1.

**Table 3.2:** Double iteration process.

| Iteration over $\boldsymbol{\pi}$ | | Iteration over $\mathbf{p}$ | | | |
|---|---|---|---|---|---|
| | | $j=1$ | $j=2$ | $j=3$ | $j=4$ |
| $i=1$ | $\begin{pmatrix} 0.1700 \\ 0.3100 \\ 0.6200 \\ 0.8800 \end{pmatrix}$ $\mu_\pi = 48.83$ | $\begin{pmatrix} 0.1700 \\ 0.3100 \\ 0.6200 \\ 0.8800 \end{pmatrix}$ $\mu_p = 48.83$ | $\begin{pmatrix} 0.2373 \\ 0.4184 \\ 0.6620 \\ 0.8862 \end{pmatrix}$ $\mu_p = 46.03$ | $\begin{pmatrix} 0.2380 \\ 0.4194 \\ 0.6629 \\ 0.8863 \end{pmatrix}$ $\mu_p = 45.99$ | $\begin{pmatrix} 0.2380 \\ 0.4194 \\ 0.6629 \\ 0.8863 \end{pmatrix}$ $\mu_p = 45.99$ |
| $i=2$ | $\begin{pmatrix} 0.2380 \\ 0.4194 \\ 0.6629 \\ 0.8863 \end{pmatrix}$ $\mu_\pi = 45.99$ | $\begin{pmatrix} 0.1700 \\ 0.3100 \\ 0.6200 \\ 0.8800 \end{pmatrix}$ $\mu_p = 48.83$ | $\begin{pmatrix} 0.2137 \\ 0.3820 \\ 0.6186 \\ 0.8563 \end{pmatrix}$ $\mu_p = 51.63$ | $\begin{pmatrix} 0.2147 \\ 0.3833 \\ 0.6197 \\ 0.8553 \end{pmatrix}$ $\mu_p = 51.72$ | $\begin{pmatrix} 0.2147 \\ 0.3833 \\ 0.6197 \\ 0.8553 \end{pmatrix}$ $\mu_p = 51.72$ |
| $i=3$ | $\begin{pmatrix} 0.2147 \\ 0.3833 \\ 0.6197 \\ 0.8553 \end{pmatrix}$ $\mu_\pi = 51.72$ | $\begin{pmatrix} 0.1700 \\ 0.3100 \\ 0.6200 \\ 0.8800 \end{pmatrix}$ $\mu_p = 48.83$ | $\begin{pmatrix} 0.2143 \\ 0.3829 \\ 0.6196 \\ 0.8570 \end{pmatrix}$ $\mu_p = 51.49$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_p = 51.57$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_p = 51.57$ |
| $i=4$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_\pi = 51.57$ | $\begin{pmatrix} 0.1700 \\ 0.3100 \\ 0.6200 \\ 0.8800 \end{pmatrix}$ $\mu_p = 48.83$ | $\begin{pmatrix} 0.2143 \\ 0.3828 \\ 0.6196 \\ 0.8570 \end{pmatrix}$ $\mu_p = 51.49$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_p = 51.58$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_p = 51.58$ |
| $i=5$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_\pi = 51.58$ | $\begin{pmatrix} 0.1700 \\ 0.3100 \\ 0.6200 \\ 0.8800 \end{pmatrix}$ $\mu_p = 48.83$ | $\begin{pmatrix} 0.2143 \\ 0.3828 \\ 0.6196 \\ 0.8570 \end{pmatrix}$ $\mu_p = 51.49$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_p = 51.58$ | $\begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$ $\mu_p = 51.58$ |

The procedure starts with the unrestricted ML estimate of $\boldsymbol{\pi}$

$$\boldsymbol{\pi} = \mathbf{p} = \begin{pmatrix} 0.1700 \\ 0.3100 \\ 0.6200 \\ 0.8800 \end{pmatrix}$$

(the observed vector of cumulative relative frequencies) and after convergence the restricted ML estimate of $\boldsymbol{\pi}$

$$\widehat{\boldsymbol{\pi}} = \begin{pmatrix} 0.2152 \\ 0.3841 \\ 0.6207 \\ 0.8561 \end{pmatrix}$$

is obtained. The elements of $\widehat{\boldsymbol{\pi}}$ follow a cumulative exponential curve at the upper class boundaries and hence the ML estimate

$$\widehat{\mu} = -\left(\frac{\mathbf{x}' \ln(\mathbf{1} - \widehat{\boldsymbol{\pi}})}{\mathbf{x}'\mathbf{x}}\right)^{-1} = 51.58$$

follows. The estimated exponential distribution is therefore

$$f(x) = \frac{1}{51.58} \exp(-\frac{x}{51.58})$$

and is shown in Figure 3.1, together with the observed frequency distribution (blue line) and estimated frequency distribution (red line).

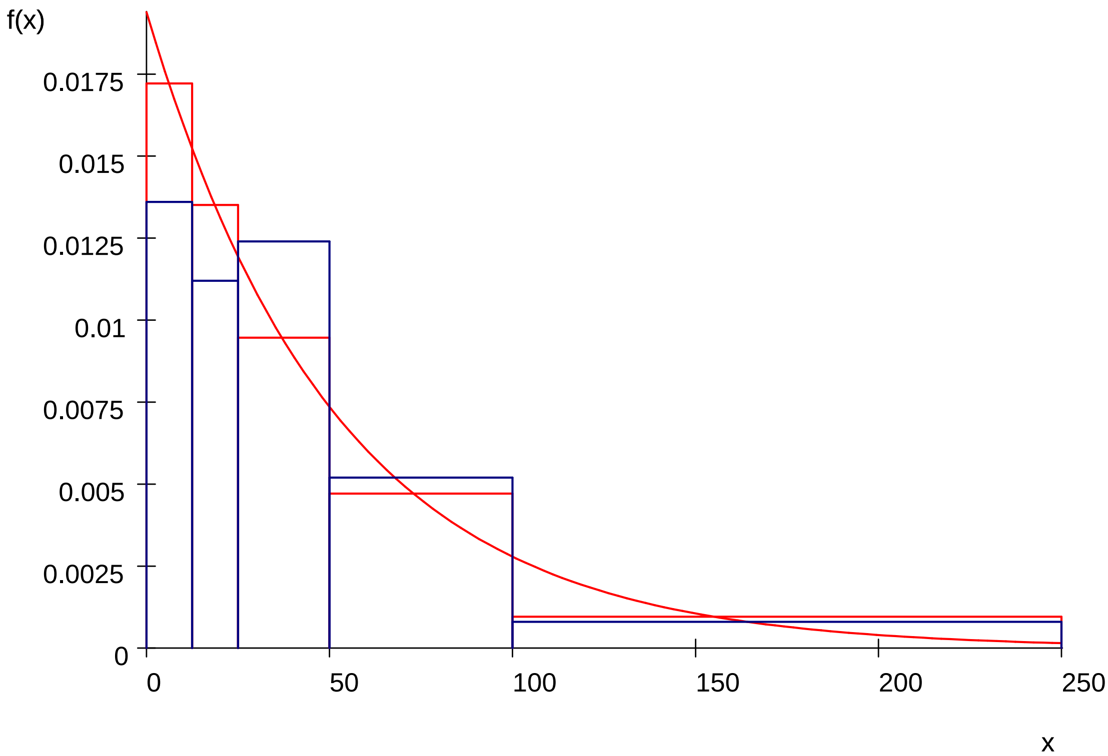



**Figure 3.1:** The estimated exponential distribution with the observed and estimated frequency distribution.

## 3.2 Constraints in terms of a linear model

An alternative formulation of the vector of constraints may be developed. The linear model

$$\ln(\mathbf{1} - \boldsymbol{\pi}) = -\theta \mathbf{x} \tag{3.12}$$

follows from the requirement (3.3) implying that $\ln(\mathbf{1} - \boldsymbol{\pi})$ is a scalar multiple of the upper class boundaries, $\mathbf{x}$. Or equivalently, $\ln(\mathbf{1} - \boldsymbol{\pi})$ is in the vector space generated by $\mathbf{x}$.

Since $\mathbf{Q}_x = \mathbf{I} - \mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'$ is the projection matrix of the vector space orthogonal to $\mathbf{x}$, the vector of constraints, $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$, may now be expressed in terms of a new $\mathbf{g}(\boldsymbol{\pi})$ namely

$$\mathbf{g}(\boldsymbol{\pi}) = \mathbf{Q}_x \ln(\mathbf{1} - \boldsymbol{\pi}) \ . \tag{3.13}$$

The rationale behind the constraints (3.13) is that $\ln(\mathbf{1}-\boldsymbol{\pi})$ is an element of the vector space of $\mathbf{x}$ if and only if $\ln(\mathbf{1}-\boldsymbol{\pi})$ is orthogonal to the error space of $\mathbf{x}$ (i.e. vector space orthogonal to $\mathbf{x}$) in which case $\mathbf{Q}_x \ln(\mathbf{1}-\boldsymbol{\pi}) = \mathbf{0}$. The vector of constraints (3.13) consists out of $(k-2)$ linear independent functions, since

$$\begin{aligned}
\operatorname{rank}(\mathbf{Q}_x) &= \operatorname{rank}(\mathbf{I}) - \operatorname{rank}\left(\mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\right) \\
&= (k-1) - \operatorname{rank}(\mathbf{x}) \\
&= (k-1) - 1
\end{aligned}$$

The matrix of partial derivatives is now much simpler than the previous formulation (3.7) with

$$\begin{aligned}
\mathbf{G}_\pi &= \frac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\
&= \frac{\partial}{\partial \boldsymbol{\pi}} \left\{\mathbf{Q}_x \ln(\mathbf{1}-\boldsymbol{\pi})\right\} && (3.14) \\
&= \mathbf{Q}_x \mathbf{D}_\pi && (3.15)
\end{aligned}$$

where $\mathbf{D}_\pi = -\left(\operatorname{diag}\left[\mathbf{1}-\boldsymbol{\pi}\right]\right)^{-1}$ (previously derived in (3.9)).

The restricted ML estimate of $\boldsymbol{\pi}$ namely $\widehat{\boldsymbol{\pi}}$ is obtained after convergence of the iterative procedure and leads to the ML estimators

$$\widehat{\theta} = -\frac{\mathbf{x}' \ln(\mathbf{1}-\widehat{\boldsymbol{\pi}})}{\mathbf{x}'\mathbf{x}}$$

and

$$\widehat{\mu} = \frac{1}{\widehat{\theta}} \ .$$

Using the multivariate delta theorem (see *Bishop, Fienberg and Holland (1975) p.492*) the asymptotic variance of $\widehat{\theta}$ follows

$$\begin{aligned}
\operatorname{Var}(\widehat{\theta}) &\approxeq \left\{\frac{\partial \theta}{\partial \boldsymbol{\pi}}\right\} \operatorname{Cov}(\widehat{\boldsymbol{\pi}}) \left\{\frac{\partial \theta}{\partial \boldsymbol{\pi}}\right\}' \\
&= \left\{\frac{\mathbf{x}'}{\mathbf{x}'\mathbf{x}} \mathbf{D}_\pi\right\} \operatorname{Cov}(\widehat{\boldsymbol{\pi}}) \left\{\frac{\mathbf{x}'}{\mathbf{x}'\mathbf{x}} \mathbf{D}_\pi\right\}' && (3.16)
\end{aligned}$$

where $\operatorname{Cov}(\widehat{\boldsymbol{\pi}})$ is given in Proposition 2 (2.12).

Applying the multivariate delta theorem again it follows that

$$\begin{aligned} \mathrm{Var}(\widehat{\mu}) &\approxeq \left\{\frac{\partial \mu}{\partial \theta}\right\}^2 \mathrm{Var}(\widehat{\theta}) \\ &= \frac{1}{\theta^4}\,\mathrm{Var}(\widehat{\theta}) \end{aligned} \tag{3.17}$$

and hence

$$\widehat{\mu} \approxeq \mathrm{N}\left(\mu, \frac{1}{\theta^4}\,\mathrm{Var}(\widehat{\theta})\right) . \tag{3.18}$$

**Example 3.2**

In this example the estimation of the exponential distribution to the simulated frequency distribution in Table 3.1 is revisited. The vector of constraints (3.13) is now formulated in terms of a linear model. The results are exactly the same as in the previous formulation (3.4), although the intermediate iterations differ. The restricted ML estimate of $\boldsymbol{\pi}$ is tabulated in Table 3.3.

The restricted and unrestricted ML estimate of $(-\ln(\mathbf{1}-\boldsymbol{\pi}))$ are tabulated in Table 3.3.

**Table 3.3:** The restricted and unrestricted ML estimates.

| Upper class boundaries | Unrestricted MLE | | Restricted MLE | |
|---|---|---|---|---|
| | $\mathbf{p}$ | $-\ln(\mathbf{1}-\mathbf{p})$ | $\widehat{\boldsymbol{\pi}}$ | $-\ln(\mathbf{1}-\widehat{\boldsymbol{\pi}})$ |
| 12.5 | 0.1700 | 0.18633 | 0.21522 | 0.24235 |
| 25 | 0.3100 | 0.37106 | 0.38412 | 0.48471 |
| 50 | 0.6200 | 0.96758 | 0.62069 | 0.96941 |
| 100 | 0.8800 | 2.1203 | 0.85613 | 1.9388 |

According to (3.12) the plot of $\ln(\mathbf{1}-\widehat{\boldsymbol{\pi}})$ against $\mathbf{x}$ should follow a straight line. In Figure 3.2 the unrestricted ML estimates are indicated with blue dots, while the restricted ML estimates are indicated with red circles. The circles follow the straight line

$$y = 0.019388x$$

implying that $\widehat{\theta} = 0.019388$ and consequently $\widehat{\mu} = 0.019388^{-1} = 51.578$.

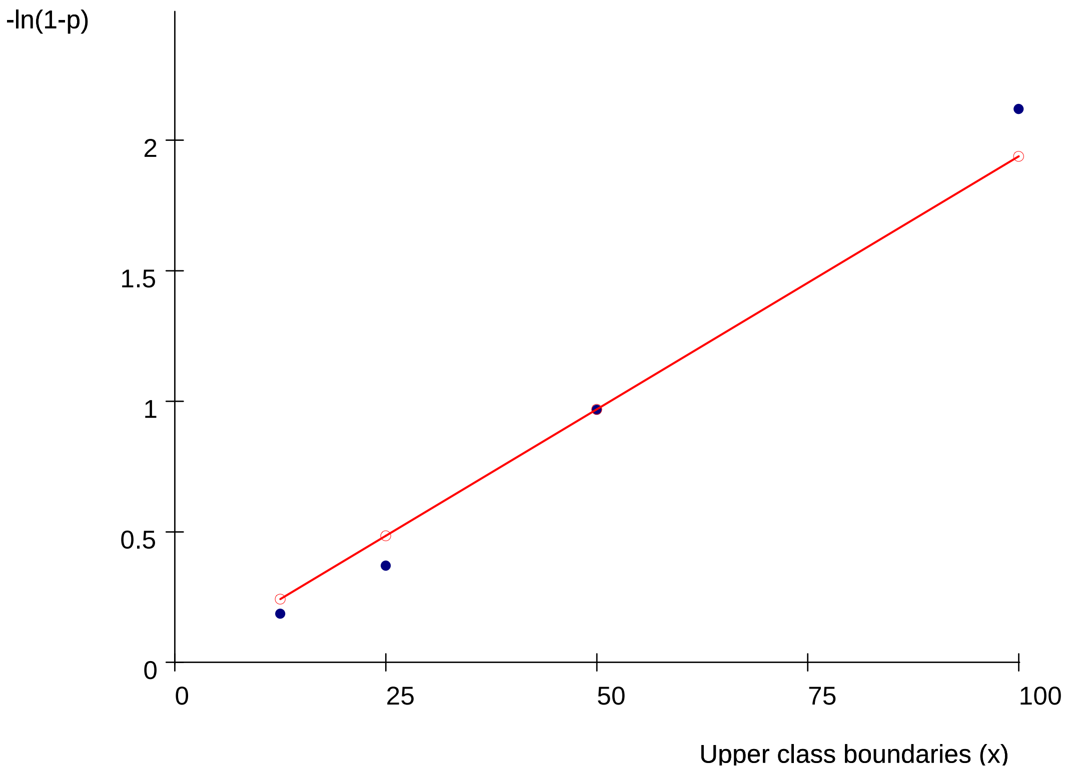



**Figure 3.2:** The values of $-\ln(\mathbf{1}-\widehat{\boldsymbol{\pi}})$ follow a straight line.

Other relevant statistics are summarised in Table 3.4.

**Table 3.4:** ML estimates and goodness of fit statistics.

| MLE | | Goodness of fit | | | |
|---|---|---|---|---|---|
| Estimate | Std. error | Statistic | Value | df | prob |
| $\widehat{\mu} = 51.578$ | $\widehat{\sigma}_{\widehat{\mu}} = 5.654$ | Pearson | $4.376$ | $3$ | $0.2236$ |
| | | Wald | $4.295$ | $3$ | $0.2313$ |

As can be expected, the Pearson and Wald statistics indicate an adequate fit.

For a 95% confidence interval for $\mu$

$$\widehat{\mu} \pm 1.96\,(\widehat{\sigma}_{\widehat{\mu}})$$

the margin of error is $1.96\,(5.654) = 11.082$, resulting in the confidence interval

$$(40.496\ ,\ 62.660)\ .$$

The SAS program *EXP2.SAS* listed in Appendix A.2 estimates the exponential distribution utilising the vector of constraints as a linear model.

## 3.3 Simulation study

In this study 1000 samples were simulated from an exponential distribution with expected value $\mu = 50$. Each sample consisted of 100 observations and were classified into the 5 class intervals of Table 3.1. Since the data was simulated from an exponential distribution with expected value $\mu = 50$ the true population value for $\boldsymbol{\pi}$ follows from

$$\boldsymbol{\pi} = \mathbf{1} - \exp\left(-\frac{\mathbf{x}}{50}\right) = \begin{pmatrix} 0.2212 \\ 0.3935 \\ 0.6321 \\ 0.8647 \end{pmatrix}$$

which implies that the standard error for $\widehat{\mu}$ is

$$\begin{aligned} \sigma_{\widehat{\mu}} &\approxeq \sqrt{50^4 \operatorname{Var}(\widehat{\theta})} \\ &= 5.458 \end{aligned}$$

($\operatorname{Var}(\widehat{\theta})$ derived in (3.16)). This compares well with the standard deviation of 5 of the mean of an ungrouped sample of 100 observations from this exponential distribution.

The ML estimate for $\mu$ as well as its estimated standard error were calculated for each of the 1000 generated frequency distributions. The true theoretical values as well as the descriptive statistics for the ML estimates are summarised in Table 3.5.

**Table 3.5:** Simulation results for the exponential distribution.

| MLE | Theoretical Value | Mean | Std. deviation | $P_5$ | Median | $P_{95}$ |
|---|---|---|---|---|---|---|
| $\widehat{\mu}$ | 50 | 50.127 | 5.727 | 41.381 | 49.676 | 59.956 |
| $\widehat{\sigma}_{\widehat{\mu}}$ | 5.458 | 5.487 | 0.716 | 4.421 | 5.418 | 6.732 |

From Table 3.5 it follows that the mean and median of the ML estimates are relatively close to the theoretical values. Further it is known that approximately 90% of the $\widehat{\mu}$-values should be within $1.645$ standard deviations from $\mu = 50$ i.e. $1.645\sigma_{\widehat{\mu}} = 8.978$. This is in accordance with the fifth and the ninety-fifth percentile of the $\widehat{\mu}$-values tabulated in Table 3.5. The standard deviation of the $\widehat{\mu}$-values is also quite close to the standard error $\sigma_{\widehat{\mu}}$.

In Table 3.6 the percentiles of the estimated 1000 Pearson and Wald statistics are tabulated. The critical values of a $\chi^2$-distribution with 3 degrees of freedom is also shown in Table 3.6.

**Table 3.6:** Percentiles of the Pearson and Wald statistic.

| | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|
| | $P_5$ | $P_{10}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{90}$ | $P_{95}$ |
| Pearson | 0.4370 | 0.6794 | 1.2829 | 2.5617 | 4.3586 | 6.5765 | 8.1324 |
| Wald | 0.3529 | 0.6299 | 1.2751 | 2.5533 | 4.2654 | 6.4586 | 8.0703 |
| | Critical values of a $\chi^2$-distribution with 3 degrees of freedom. | | | | | | |
| | $\chi^2_{0.05}$ | $\chi^2_{0.10}$ | $\chi^2_{0.25}$ | $\chi^2_{0.50}$ | $\chi^2_{0.75}$ | $\chi^2_{0.90}$ | $\chi^2_{0.95}$ |
| $\chi^2\,(3)$ | 0.3518 | 0.5844 | 1.2125 | 2.366 | 4.1083 | 6.2514 | 7.8147 |

From Table 3.6 it is clear that the empirical percentiles of the Pearson and Wald statistics correspond very well to the theoretical percentiles of a $\chi^2$-distribution with 3 degrees of freedom.

The simulation study was done with the SAS program *EXPSIM.SAS* listed in Appendix A.3.

# Chapter 4

# The normal distribution

Analogous to the exponential distribution described in Chapter 3 the normal distribution with pdf

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right\} \tag{4.1}$$

will be fitted to grouped data using a direct set of constraints and also constraints specified in terms of a linear model. The mean and variance of the normal distribution are $\mu$ and $\sigma^2$ respectively.

By means of standardisation, $z=\frac{x-\mu}{\sigma}$, the standard normal distribution with pdf

$$\phi(z)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}z^{2}\right\} \tag{4.2}$$

is obtained. The cdf of the standard normal distribution is denoted by $\boldsymbol{\Phi}(z)$.

To fit a normal distribution to the frequency data in Table 2.1 it is required that

$$\boldsymbol{\Phi}\left(\frac{\mathbf{x}-\mu\mathbf{1}}{\sigma}\right)=\boldsymbol{\pi} \tag{4.3}$$

where $\boldsymbol{\Phi}(\cdot)$ is the (vector valued) cdf of the standard normal distribution, $\mathbf{1}$ is the $(k-1)$ vector of ones and $\mathbf{x}$ is the vector of upper class boundaries defined in (2.2).

## 4.1 Direct set of constraints

To fit a normal distribution to grouped data a direct set of constraints, $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$, with

$$\mathbf{g}(\boldsymbol{\pi}) = \boldsymbol{\Phi}(\mathbf{z}) - \boldsymbol{\pi} \tag{4.4}$$

follows from (4.3). The vector of standardised upper class boundaries in (4.4) is a function of the parameters to be estimated namely

$$\begin{aligned} \mathbf{z} &= \left(\frac{\mathbf{x} - \mu\mathbf{1}}{\sigma}\right) \\ &= \begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma} \\ \frac{\mu}{\sigma} \end{pmatrix} \\ &= \mathbf{X}\boldsymbol{\alpha} \end{aligned} \tag{4.5}$$

with

$$\mathbf{X} = \begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix} \tag{4.6}$$

and

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma} \\ \frac{\mu}{\sigma} \end{pmatrix} \tag{4.7}$$

the vector of so-called natural parameters.

Under normality (see 4.3)

$$\begin{aligned} \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) &= \left(\frac{\mathbf{x} - \mu\mathbf{1}}{\sigma}\right) \\ &= \mathbf{X}\boldsymbol{\alpha} \end{aligned} \tag{4.8}$$

which leads to the expression

$$\boldsymbol{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \ . \tag{4.9}$$

The parameters of the normal distribution are now specified in terms of the cumulative relative frequencies $\boldsymbol{\pi}$. Hence, from (4.5) and (4.9) the standardised upper class boundaries may be expressed as

$$\mathbf{z} = \mathbf{P}_X\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) \tag{4.10}$$

where

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{4.11}$$

is the projection matrix generated by the columns of $\mathbf{X}$. This implies that, under normality the vector $\mathbf{z}$ is the projection of $\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})$ on the vector space of $\mathbf{X}$.

From the chain rule for matrix differentiation, employed in (4.12), it follows that the matrix of partial derivatives is

$$
\begin{aligned}
\mathbf{G}_\pi &= \frac{\partial \mathrm{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\
&= \frac{\partial \boldsymbol{\Phi}(\mathbf{z})}{\partial \boldsymbol{\pi}} - \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\pi}} \\
&= \frac{\partial \boldsymbol{\Phi}(\mathbf{z})}{\partial \mathbf{z}} \cdot \frac{\partial \mathbf{z}}{\partial \boldsymbol{\pi}} - \mathbf{I} && (4.12) \\
&= \operatorname{diag}[\boldsymbol{\phi}(\mathbf{z})] \cdot \mathbf{P}_X \cdot \mathbf{D}_\pi - \mathbf{I} && (4.13)
\end{aligned}
$$

where

$$\mathbf{D}_\pi = \frac{\partial \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} . \qquad (4.14)$$

To solve (4.14) set $\boldsymbol{\nu} = \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})$ then $\boldsymbol{\Phi}(\boldsymbol{\nu}) = \boldsymbol{\pi}$ and hence

$$
\begin{aligned}
\mathbf{D}_\pi &= \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\pi}} \\
&= \left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\nu}}\right)^{-1} \\
&= \left(\frac{\partial \boldsymbol{\Phi}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right)^{-1} \\
&= (\operatorname{diag}[\boldsymbol{\phi}(\boldsymbol{\nu})])^{-1} \\
&= \left(\operatorname{diag}\left[\boldsymbol{\phi}\left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})\right)\right]\right)^{-1} && (4.15)
\end{aligned}
$$

with $\boldsymbol{\phi}(\cdot)$ the vector valued pdf of the standard normal distribution.

The vector of constraints (4.4) and the matrix of partial derivatives (4.13) may be implemented in the ML estimation procedure, where the restricted ML estimate $\widehat{\boldsymbol{\pi}}$ is obtained iteratively in a double iterative procedure.

The iterative procedure may be summarized as follows:

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

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

$\mathbf{P}_X = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$

DO OVER $\boldsymbol{\pi}$

$\boldsymbol{\pi} = \mathbf{p}$

$\mathbf{V} = \frac{1}{n}\left\{\mathbf{C}\,\mathrm{diag}\left[\mathbf{C}^{-1}\boldsymbol{\pi}\right]\mathbf{C}' - \boldsymbol{\pi}\boldsymbol{\pi}'\right\}$

$\mathbf{D}_{\pi} = \left(\mathrm{diag}\left[\boldsymbol{\phi}\left(\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\pi}\right)\right)\right]\right)^{-1}$

$\mathbf{z}_{\pi} = \mathbf{P}_X\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\pi}\right)$

$\mathbf{G}_{\pi} = \mathrm{diag}\left[\boldsymbol{\phi}\left(\mathbf{z}_{\pi}\right)\right]\cdot\mathbf{P}_X\cdot\mathbf{D}_{\pi} - \mathbf{I}$

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

DO OVER $\mathbf{p}$

$\mathbf{D}_{p} = \left(\mathrm{diag}\left[\boldsymbol{\phi}\left(\boldsymbol{\Phi}^{-1}\left(\mathbf{p}\right)\right)\right]\right)^{-1}$

$\mathbf{z}_{p} = \mathbf{P}_X\boldsymbol{\Phi}^{-1}\left(\mathbf{p}\right)$

$\mathbf{G}_{p} = \mathrm{diag}\left[\boldsymbol{\phi}\left(\mathbf{z}_{p}\right)\right]\cdot\mathbf{P}_X\cdot\mathbf{D}_{p} - \mathbf{I}$

$\mathbf{g}(\mathbf{p}) = \boldsymbol{\Phi}\left(\mathbf{z}_{p}\right) - \mathbf{p}$

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

END

END

For convergence over $\mathbf{p}$ the vector of upper class boundaries $\mathbf{z}_p$, the matrix of partial derivatives $\mathbf{G}_p$ and the vector of constraints $\mathbf{g}(\mathbf{p})$ are all functions of $\mathbf{p}$. Utilizing

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

convergence is obtained over $\mathbf{p}$ resulting in a new value for $\boldsymbol{\pi}$. For convergence over $\boldsymbol{\pi}$ the covariance matrix $\mathbf{V}$, vector of upper class boundaries $\mathbf{z}_{\pi}$ and the matrix of partial derivatives $\mathbf{G}_{\pi}$ are all functions of $\boldsymbol{\pi}$. Convergence over $\boldsymbol{\pi}$ leads to the restricted ML estimate $\widehat{\boldsymbol{\pi}}$ which follows a cumulative normal distribution curve at the upper class boundaries $\mathbf{x}$. From the restricted ML estimate $\widehat{\boldsymbol{\pi}}$ the ML estimator

$$\widehat{\boldsymbol{\alpha}} = \begin{pmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \end{pmatrix} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\boldsymbol{\Phi}^{-1}\left(\widehat{\boldsymbol{\pi}}\right) \tag{4.16}$$

follows and consequently the ML estimators for the normal distribution are

$$\widehat{\mu} = \frac{\widehat{\alpha}_2}{\widehat{\alpha}_1} \tag{4.17}$$

and

$$\widehat{\sigma} = \frac{1}{\widehat{\alpha}_1}\ . \tag{4.18}$$

See (4.7) for the formulation of the parameters.

**Example 4.1**

The normal distribution will now be fitted to 100 observations simulated from a normal population with mean 58 and standard deviation 15. The data is shown in Table 4.1.

**Table 4.1:** Simulated data set from a normal distribution.

| Class Interval | Frequency |
|---|---|
| $[0, 40)$ | 9 |
| $[40, 50)$ | 26 |
| $[50, 60)$ | 24 |
| $[60, 75)$ | 27 |
| $[75, 100)$ | 14 |

The various values for $\mathbf{p}$ and $\boldsymbol{\pi}$ in the double iteration process are calculated with the SAS program *NORM1.SAS* (listed in Appendix A.4) and tabulated in Table 4.2. The corresponding values for $\mu$ and $\sigma$ are also listed in Table 4.2.

**Table 4.2:** Double iteration process.

| Iteration over $\boldsymbol{\pi}$ | | Iteration over $\mathbf{p}$ | | | |
|---|---|---|---|---|---|
| | | $j=1$ | $j=2$ | $j=3$ | $j=4$ |
| $i=1$ | $\begin{pmatrix} 0.0900 \\ 0.3500 \\ 0.5900 \\ 0.8600 \end{pmatrix}$ $\mu_\pi = 57.79$ $\sigma_\pi = 14.76$ | $\begin{pmatrix} 0.0900 \\ 0.3500 \\ 0.5900 \\ 0.8600 \end{pmatrix}$ $\mu_p = 57.79$ $\sigma_p = 14.76$ | $\begin{pmatrix} 0.0950 \\ 0.2721 \\ 0.5375 \\ 0.8734 \end{pmatrix}$ $\mu_p = 58.68$ $\sigma_p = 14.27$ | $\begin{pmatrix} 0.0951 \\ 0.2713 \\ 0.5367 \\ 0.8736 \end{pmatrix}$ $\mu_p = 58.69$ $\sigma_p = 14.26$ | $\begin{pmatrix} 0.0951 \\ 0.2713 \\ 0.5367 \\ 0.8736 \end{pmatrix}$ $\mu_p = 58.69$ $\sigma_p = 14.26$ |
| $i=2$ | $\begin{pmatrix} 0.0951 \\ 0.2713 \\ 0.5367 \\ 0.8736 \end{pmatrix}$ $\mu_\pi = 58.69$ $\sigma_\pi = 14.26$ | $\begin{pmatrix} 0.0900 \\ 0.3500 \\ 0.5900 \\ 0.8600 \end{pmatrix}$ $\mu_p = 57.79$ $\sigma_p = 14.76$ | $\begin{pmatrix} 0.1196 \\ 0.3094 \\ 0.5670 \\ 0.8791 \end{pmatrix}$ $\mu_p = 57.50$ $\sigma_p = 14.92$ | $\begin{pmatrix} 0.1206 \\ 0.3078 \\ 0.5667 \\ 0.8796 \end{pmatrix}$ $\mu_p = 57.49$ $\sigma_p = 14.92$ | $\begin{pmatrix} 0.1206 \\ 0.3078 \\ 0.5667 \\ 0.8796 \end{pmatrix}$ $\mu_p = 57.49$ $\sigma_p = 14.92$ |
| $i=3$ | $\begin{pmatrix} 0.1206 \\ 0.3078 \\ 0.5667 \\ 0.8796 \end{pmatrix}$ $\mu_\pi = 57.49$ $\sigma_\pi = 14.92$ | $\begin{pmatrix} 0.0900 \\ 0.3500 \\ 0.5900 \\ 0.8600 \end{pmatrix}$ $\mu_p = 57.79$ $\sigma_p = 14.76$ | $\begin{pmatrix} 0.1188 \\ 0.3084 \\ 0.5666 \\ 0.8794 \end{pmatrix}$ $\mu_p = 57.52$ $\sigma_p = 14.88$ | $\begin{pmatrix} 0.1197 \\ 0.3068 \\ 0.5663 \\ 0.8799 \end{pmatrix}$ $\mu_p = 57.52$ $\sigma_p = 14.89$ | $\begin{pmatrix} 0.1197 \\ 0.3068 \\ 0.5663 \\ 0.8799 \end{pmatrix}$ $\mu_p = 57.52$ $\sigma_p = 14.89$ |
| $i=4$ | $\begin{pmatrix} 0.1197 \\ 0.3068 \\ 0.5663 \\ 0.8799 \end{pmatrix}$ $\mu_\pi = 57.52$ $\sigma_\pi = 14.89$ | $\begin{pmatrix} 0.0900 \\ 0.3500 \\ 0.5900 \\ 0.8600 \end{pmatrix}$ $\mu_p = 57.79$ $\sigma_p = 14.76$ | $\begin{pmatrix} 0.1188 \\ 0.3084 \\ 0.5666 \\ 0.8794 \end{pmatrix}$ $\mu_p = 57.52$ $\sigma_p = 14.89$ | $\begin{pmatrix} 0.1197 \\ 0.3068 \\ 0.5663 \\ 0.8799 \end{pmatrix}$ $\mu_p = 57.52$ $\sigma_p = 14.89$ | $\begin{pmatrix} 0.1197 \\ 0.3068 \\ 0.5663 \\ 0.8799 \end{pmatrix}$ $\mu_p = 57.52$ $\sigma_p = 14.89$ |

From Table 4.2 it can be seen that the ML procedure converges extremely fast. The procedure starts off with the unrestricted ML estimate for $\boldsymbol{\pi}$ (observed cumulative relative frequencies)

$$\boldsymbol{\pi} = \mathbf{p} = \begin{pmatrix} 0.0900 \\ 0.3500 \\ 0.5900 \\ 0.8600 \end{pmatrix}$$

and converges ultimately to the restricted ML estimate of $\boldsymbol{\pi}$

$$\widehat{\boldsymbol{\pi}} = \begin{pmatrix} 0.1197 \\ 0.3068 \\ 0.5663 \\ 0.8799 \end{pmatrix} .$$

The elements of $\widehat{\boldsymbol{\pi}}$ follow a cumulative normal distribution curve at the upper class boundaries of $\mathbf{x}$ and hence the ML estimates of the natural parameters follows from (4.16) with

$$\widehat{\boldsymbol{\alpha}} = \begin{pmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} 0.06717 \\ 3.86338 \end{pmatrix} .$$

From (4.17) and (4.18) the ML estimates for the normal distribution are

$$\widehat{\mu} = 57.52 \quad \text{and} \quad \widehat{\sigma} = 14.89 .$$

The estimated normal distribution is shown in Figure 4.1, together with the observed frequency distribution (blue line) and estimated frequency distribution (red line).

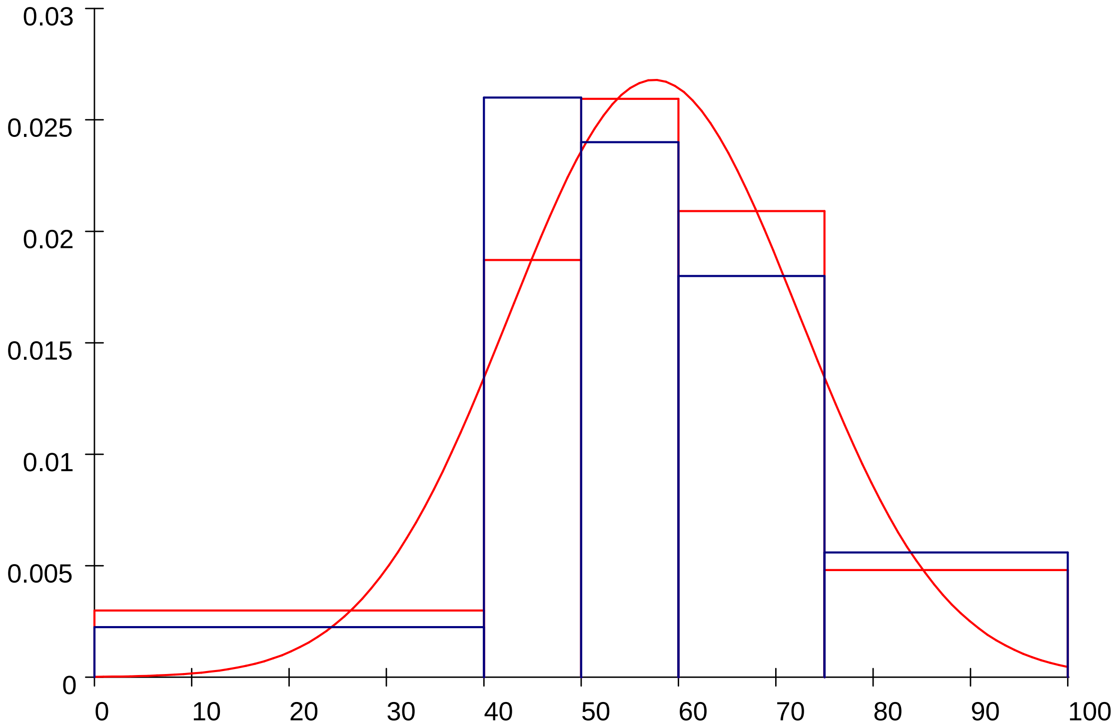



**Figure 4.1:** The estimated normal distribution with the observed and estimated frequency distribution.

## 4.2 Constraints in terms of a linear model

In the previous section a normal distribution was fitted to a grouped data set utilizing a direct set of constraints. In this section the constraints will be formulated in terms of a linear model.

From (4.3) it is possible to formulate the linear model

$$\begin{aligned} \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) &= \left(\frac{\mathbf{x}-\mu\mathbf{1}}{\sigma}\right) \\ &= \mathbf{X}\boldsymbol{\alpha} \end{aligned} \tag{4.19}$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix} \tag{4.20}$$

is the design matrix and

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma} \\ \frac{\mu}{\sigma} \end{pmatrix} \tag{4.21}$$

is the vector of natural parameters.

The linear model (4.19) implies the vector of constraints

$$\mathbf{g}(\boldsymbol{\pi}) = \mathbf{Q}_X \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}) = \mathbf{0} \tag{4.22}$$

to be imposed in the ML estimation procedure, where

$$\mathbf{Q}_X = \mathbf{I} - \mathbf{P}_X \tag{4.23}$$

is the projection matrix orthogonal to $\mathbf{X}$ and $\mathbf{P}_X$ is previously defined in (4.11). According to (4.22) the vector of cumulative probabilities will be fitted such that $\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})$ is orthogonal to the error space of $\mathbf{X}$ or equivalently such that $\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})$ is in the vector space of $\mathbf{X}$.

The matrix of partial derivatives follows

$$\begin{aligned} \mathbf{G}_\pi &= \frac{\partial \mathbf{Q}_X \boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= \mathbf{Q}_X \mathbf{D}_\pi \end{aligned} \tag{4.24}$$

where $\mathbf{D}_\pi = \left(\text{diag}\left[\phi\left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})\right)\right]\right)^{-1}$ is already derived in (4.15).

Employing the vector of constraints (4.22) and the matrix of partial derivatives (4.24) in the ML estimation procedure the restricted ML estimate, $\widehat{\boldsymbol{\pi}}$, is obtained. It follows from (4.19) that the ML estimator of $\boldsymbol{\alpha}$ is

$$\widehat{\boldsymbol{\alpha}} = \begin{pmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \end{pmatrix} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \boldsymbol{\Phi}^{-1}\left(\widehat{\boldsymbol{\pi}}\right) \tag{4.25}$$

with asymptotic covariance matrix

$$\begin{aligned} \text{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) &\approxeq \left(\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\pi}}\right) \text{Cov}(\widehat{\boldsymbol{\pi}}) \left(\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\pi}}\right)' \\ &= \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\} \text{Cov}(\widehat{\boldsymbol{\pi}}) \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\}' . \end{aligned} \tag{4.26}$$

The ML estimators

$$\widehat{\mu} = \frac{\widehat{\alpha}_2}{\widehat{\alpha}_1} \quad \text{and} \quad \widehat{\sigma} = \frac{1}{\widehat{\alpha}_1} \tag{4.27}$$

follows from (4.25) and (4.21).

Let

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} \frac{\alpha_2}{\alpha_1} \\ \frac{1}{\alpha_1} \end{pmatrix} \tag{4.28}$$

denote the vector of original parameters for the normal distribution. To find the asymptotic distribution for the ML estimate $\widehat{\boldsymbol{\beta}}$, the multivariate $\delta$-theorem is once again implemented and hence

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &\approx \mathrm{N}\left(\boldsymbol{\beta}, \mathrm{Cov}(\widehat{\boldsymbol{\beta}})\right) & (4.29) \\ &= \mathrm{N}\left(\begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \mathbf{B}\,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\mathbf{B}'\right) & (4.30) \end{aligned}$$

where

$$\begin{aligned} \mathbf{B} &= \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}} \\ &= \frac{\partial \begin{pmatrix} \mu \\ \sigma \end{pmatrix}}{\partial \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}} = \begin{pmatrix} -\frac{\alpha_2}{\alpha_1^2} & \frac{1}{\alpha_1} \\ -\frac{1}{\alpha_1^2} & 0 \end{pmatrix} . & (4.31) \end{aligned}$$

**Example 4.2**

The normal distribution will now be fitted to the frequency distribution tabulated in Table 4.1, now employing the vector of constraints as a linear model (4.22). By making use of the SAS program *NORM2.SAS* in Appendix A.5, the ML estimation procedure yields exactly the same restricted ML estimate for $\boldsymbol{\pi}$, as in Example 4.1, namely

$$\widehat{\boldsymbol{\pi}} = \begin{pmatrix} 0.1197 \\ 0.3068 \\ 0.5663 \\ 0.8799 \end{pmatrix}$$

although the intermediate iterations differ. The elements of

$$\boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}}) = \begin{pmatrix} -1.17652 \\ -0.50480 \\ 0.16691 \\ 1.17448 \end{pmatrix}$$

are the estimates of the inverse normal probabilities (standardised upper class boundaries) and

$$\begin{aligned} \mathbf{P}_X &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \begin{pmatrix} 0.64486 & 0.40187 & 0.15888 & -0.20561 \\ 0.40187 & 0.30841 & 0.21495 & 0.07477 \\ 0.15888 & 0.21495 & 0.27103 & 0.35514 \\ -0.20561 & 0.07477 & 0.35514 & 0.77570 \end{pmatrix} \end{aligned}$$

is the projection matrix generated by the columns of $\mathbf{X}$. Multiplying these two matrices lead to

$$\mathbf{P}_X\boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}}) = \boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}})$$

which means that $\boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}})$ is in the vector space of $\mathbf{X}$ and consequently $\boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}})$ is a linear combination of the columns of $\mathbf{X}$ in (4.20). It is also clear that

$$\mathbf{Q}_X\boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}}) = \mathbf{0}$$

indicating that $\boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}})$ is orthogonal to the error space of $\mathbf{X}$. (See 4.22 and 4.23.)

The ML estimates and goodness of fit statistics are summarized in Table 4.3

**Table 4.3:** ML estimates and goodness of fit statistics for the normal distribution.

| MLE | | Goodness of fit | | | |
|---|---|---|---|---|---|
| Estimate | Std. error | Statistic | Value | df | prob |
| $\widehat{\mu} = 57.515$ | $\widehat{\sigma}_{\widehat{\mu}} = 1.556$ | Pearson | 4.654 | 2 | 0.0976 |
| $\widehat{\sigma} = 14.887$ | $\widehat{\sigma}_{\widehat{\sigma}} = 1.327$ | Wald | 4.855 | 2 | 0.1455 |

According to the goodness of fit statistics summarized in Table 4.3, the null hypothesis of an adequate fit is not rejected at a 5% level of significance. The adequate fit is further illustrated in Figure 4.1.

The estimated standard errors $\widehat{\sigma}_{\widehat{\mu}}$ and $\widehat{\sigma}_{\widehat{\sigma}}$ in Table 4.3 follows from the estimated covariance matrix

$$\widehat{\text{Cov}}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\text{Cov}}\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma} \end{pmatrix} = \begin{pmatrix} 2.4219 & 0.0353 \\ 0.0353 & 1.7622 \end{pmatrix}$$

which is estimated by substituting the restricted ML estimate $\widehat{\boldsymbol{\pi}}$ in $\text{Cov}\left(\widehat{\boldsymbol{\pi}}\right)$.

The 95% confidence intervals for $\mu$ and $\sigma$ are tabulated in Table 4.4.

**Table 4.4:** 95% confidence intervals for $\mu$ and $\sigma$.

| Parameter | Margin of error | Confidence interval |
|---|---|---|
| $\mu$ | $1.96\,(1.556) = 3.049$ | $(54.951, 61.049)$ |
| $\sigma$ | $1.96\,(1.327) = 2.601$ | $(12.286, 17.488)$ |

From the confidence intervals reported in Table 4.4 the population parameters $\mu$ and $\sigma$ do not differ significantly from the theoretical values 58 and 15.

## 4.3 Simulation study

Similar to the simulation study done for the exponential distribution in the previous chapter, 1000 samples were simulated, each containing 100 observations. These samples were all simulated from a normal population with mean $\mu = 58$ and standard deviation $\sigma = 15$. The descriptive statistics for the 1000 sample means and sample standard deviations of the ungrouped data sets are summarised in Table 4.5.

**Table 4.5:** Descriptive statistics for sample statistics of ungrouped data sets.

| Statistic | Mean | Std. deviation | $P_5$ | Median | $P_{95}$ |
|---|---|---|---|---|---|
| $\bar{x}$ | 57.993 | 1.489 | 55.582 | 57.919 | 60.446 |
| $s$ | 14.902 | 1.078 | 13.244 | 14.881 | 16.673 |

Evaluating the sample statistics for the ungrouped data sets, the mean and median are very close to the theoretical values. The standard deviation of $\bar{x}$ is close to the standard error of $\bar{x}$, i.e.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5 \ .$$

The 1000 simulated data sets were all classified into the same set of class intervals as that of Table 4.1. The normal distribution was fitted to each of the 1000 generated frequency distributions and the descriptive statistics for the ML estimates are tabulated in Table 4.6.

**Table 4.6:** Simulation results for the normal distribution.

| MLE | Theoretical Value | Mean | Std. deviation | $P_5$ | Median | $P_{95}$ |
|---|---|---|---|---|---|---|
| $\widehat{\mu}$ | 58.000 | 57.993 | 1.548 | 55.512 | 57.945 | 60.598 |
| $\widehat{\sigma}_{\widehat{\mu}}$ | 1.569 | 1.562 | 0.146 | 1.343 | 1.550 | 1.826 |
| $\widehat{\sigma}$ | 15.000 | 14.915 | 1.384 | 12.797 | 14.823 | 17.376 |
| $\widehat{\sigma}_{\widehat{\sigma}}$ | 1.341 | 1.340 | 0.171 | 1.091 | 1.320 | 1.653 |

In the case of a normal distribution with $\mu = 58$ and $\sigma = 15$ the theoretical value for $\boldsymbol{\pi}$ is

$$\boldsymbol{\pi} = \boldsymbol{\Phi}\left(\frac{\mathbf{x} - 58\,(\mathbf{1})}{15}\right) = \boldsymbol{\Phi}\begin{pmatrix} -1.2000 \\ -0.5333 \\ 0.1333 \\ 1.1333 \end{pmatrix} = \begin{pmatrix} 0.11507 \\ 0.29690 \\ 0.55304 \\ 0.87146 \end{pmatrix}$$

leading to the asymptotic covariance matrix

$$\text{Cov}\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma} \end{pmatrix} \approx \begin{pmatrix} 2.46085 & 0.05201 \\ 0.05201 & 1.79748 \end{pmatrix}$$

and yielding the standard errors $\sigma_{\widehat{\mu}} = 1.569$ and $\sigma_{\widehat{\sigma}} = 1.341$ tabulated in Table 4.6. In view of the fact that the standard error for a random sample from a $\mathrm{N}(58, 15^2)$ distribution is $\frac{15}{\sqrt{100}} = 1.5$, not much accuracy has been lost by using a grouped sample in the estimation of $\mu$. As is evident from Table 4.6 the mean and median of each of the ML estimates compare extremely well with the theoretical values (approximate in the case of $\sigma_{\widehat{\mu}}$ and $\sigma_{\widehat{\sigma}}$). It is also interesting to note that standard deviations for $\widehat{\mu}$ and $\widehat{\sigma}$ are close to the standard errors $\sigma_{\widehat{\mu}}$ and $\sigma_{\widehat{\sigma}}$. To evaluate the fifth and the ninety fifth percentiles the margin of error for the 90% confidence intervals are summarised in Table 4.7.

**Table 4.7:** 90% margin of error for the ML estimators of the normal distribution.

| Estimate | Std. Error | Margin of Error |
|---|---|---|
| $\widehat{\mu}$ | $\sigma_{\widehat{\mu}}$ | $1.645\sigma_{\widehat{\mu}} = 2.581$ |
| $\widehat{\sigma}$ | $\sigma_{\widehat{\sigma}}$ | $1.645\sigma_{\widehat{\sigma}} = 2.206$ |

It is known that approximately 90% of the $\widehat{\mu}$-values should be in the interval $(55.419, 60.581)$, while 90% of the $\widehat{\sigma}$-values should be in the interval $(12.794, 17.206)$. This compares well with the simulated values in Table 4.6.

The goodness of fit statistics were calculated for each of the 1000 fitted normal distributions. From Table 4.8 it follows that the Pearson and Wald statistics correspond very well to that of a $\chi^2$-distribution with 2 degrees of freedom.

**Table 4.8:** Percentiles of the Pearson and Wald statistic.

| | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|
| | $P_5$ | $P_{10}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{90}$ | $P_{95}$ |
| Pearson | 0.1291 | 0.2355 | 0.5945 | 1.3728 | 2.7147 | 4.6345 | 5.8393 |
| Wald | 0.1066 | 0.2054 | 0.5925 | 1.3742 | 2.7591 | 4.6721 | 6.1128 |
| | Percentiles of a $\chi^2$-distribution with 2 degrees of freedom. | | | | | | |
| | $\chi^2_{0.05}$ | $\chi^2_{0.10}$ | $\chi^2_{0.25}$ | $\chi^2_{0.50}$ | $\chi^2_{0.75}$ | $\chi^2_{0.90}$ | $\chi^2_{0.95}$ |
| $\chi^2(2)$ | 0.1026 | 0.2107 | 0.5754 | 1.3863 | 2.7726 | 4.6052 | 5.9915 |

# Chapter 5

# The Weibull, log-logistic and Pareto distributions

In this chapter it will be shown how to fit the Weibull, log-logistic and Pareto distributions to a grouped data set. Estimation will be done by constructing the vector of constraints in terms of a linear model. This method is preferred due to the simplicity and the overall generalization of the technique. This generalization is outlined in 3 easy steps where the estimation of the exponential and normal distributions are also considered.

## 5.1 The Weibull distribution

The pdf of the Weibull distribution is

$$f(x;\kappa,\theta) = \frac{\kappa}{\theta^{\kappa}} x^{\kappa-1} \exp\left[-\left(\frac{x}{\theta}\right)^{\kappa}\right] \tag{5.1}$$

with cdf

$$F(x;\kappa,\theta) = 1 - \exp\left[-\left(\frac{x}{\theta}\right)^{\kappa}\right] \ . \tag{5.2}$$

The parameter $\kappa$ is a shape parameter with $\theta$ the so-called scale parameter. The three basic shapes of the Weibull distribution are illustrated in Figure 5.1.

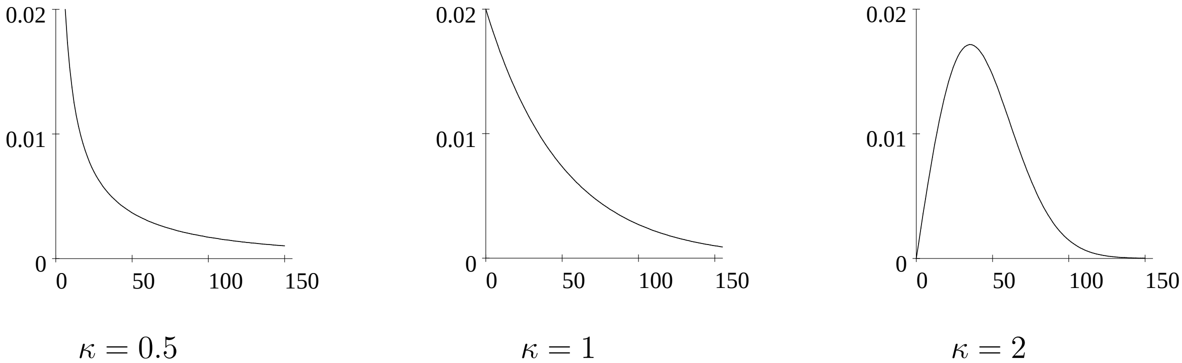



**Figure 5.1:** Weibull distributions with $\theta = 50$.

The mean and variance of the Weibull distribution are

$$\mu = \theta \left[ \Gamma \left( 1 + \frac{1}{\kappa} \right) \right] \tag{5.3}$$

and

$$\sigma^2 = \theta^2 \left[ \Gamma \left( 1 + \frac{2}{\kappa} \right) - \Gamma^2 \left( 1 + \frac{1}{\kappa} \right) \right] \tag{5.4}$$

respectively.

To fit a Weibull distribution it is required that

$$\boldsymbol{\pi} = \mathbf{1} - \exp \left[ - \left( \frac{\mathbf{x}}{\theta} \right)^{\kappa} \right] \tag{5.5}$$

which implies that

$$\ln \left( \mathbf{1} - \boldsymbol{\pi} \right) = - \left( \frac{\mathbf{x}}{\theta} \right)^{\kappa} . \tag{5.6}$$

Taking the natural logarithm of (5.6) yields the linear model

$$\begin{aligned} \ln \left[ - \ln \left( \mathbf{1} - \boldsymbol{\pi} \right) \right] &= \kappa \ln \mathbf{x} - \left( \kappa \ln \theta \right) \mathbf{1} \\ &= \begin{pmatrix} \ln \mathbf{x} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \kappa \\ \kappa \ln \theta \end{pmatrix} \\ &= \mathbf{X} \boldsymbol{\alpha} \end{aligned} \tag{5.7}$$

where

$$\mathbf{X} = \begin{pmatrix} \ln \mathbf{x} & -\mathbf{1} \end{pmatrix} \tag{5.8}$$

is the design matrix and

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \kappa \\ \kappa \ln \theta \end{pmatrix} \tag{5.9}$$

is the vector of natural parameters.

The vector of constraints

$$\mathbf{g}(\boldsymbol{\pi}) = \mathbf{Q}_X \ln \left[ -\ln \left( \mathbf{1} - \boldsymbol{\pi} \right) \right] = \mathbf{0} \tag{5.10}$$

follows from (5.7) where $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the projection matrix orthogonal to $\mathbf{X}$. The matrix of partial derivatives becomes

$$\begin{aligned} \mathbf{G}_\pi &= \frac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= \frac{\partial \left\{ \mathbf{Q}_X \ln \left[ -\ln \left( \mathbf{1} - \boldsymbol{\pi} \right) \right] \right\}}{\partial \boldsymbol{\pi}} \\ &= \mathbf{Q}_X \mathbf{D}_\pi \end{aligned} \tag{5.11}$$

where

$$\begin{aligned} \mathbf{D}_\pi &= \frac{\partial \ln \left[ -\ln \left( \mathbf{1} - \boldsymbol{\pi} \right) \right]}{\partial \boldsymbol{\pi}} \\ &= \left\{ \operatorname{diag} \left[ -\ln \left( \mathbf{1} - \boldsymbol{\pi} \right) \right] \right\}^{-1} \frac{\partial}{\partial \boldsymbol{\pi}} \left\{ -\ln \left( \mathbf{1} - \boldsymbol{\pi} \right) \right\} \\ &= -\left\{ \operatorname{diag} \left[ \ln \left( \mathbf{1} - \boldsymbol{\pi} \right) \right] \right\}^{-1} \left\{ \operatorname{diag} \left[ \mathbf{1} - \boldsymbol{\pi} \right] \right\}^{-1} . \end{aligned} \tag{5.12}$$

The restricted ML estimate $\widehat{\boldsymbol{\pi}}$ is estimated such that $\ln \left[ -\ln \left( \mathbf{1} - \widehat{\boldsymbol{\pi}} \right) \right]$ is a linear combination of $\mathbf{X}$ leading to the ML estimator

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \ln \left[ -\ln \left( \mathbf{1} - \widehat{\boldsymbol{\pi}} \right) \right] \tag{5.13}$$

with asymptotic covariance matrix

$$\operatorname{Cov}(\widehat{\boldsymbol{\alpha}}) \approxeq \left\{ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi \right\} \operatorname{Cov}(\widehat{\boldsymbol{\pi}}) \left\{ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi \right\}' . \tag{5.14}$$

The parameters of the Weibull distribution are

$$\boldsymbol{\beta} = \begin{pmatrix} \kappa \\ \theta \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \exp\left( \frac{\alpha_2}{\alpha_1} \right) \end{pmatrix} . \tag{5.15}$$

Hence, the ML estimator for $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\kappa} \\ \widehat{\theta} \end{pmatrix} = \begin{pmatrix} \widehat{\alpha}_1 \\ \exp\left(\frac{\widehat{\alpha}_2}{\widehat{\alpha}_1}\right) \end{pmatrix} \tag{5.16}$$

with asymptotic covariance matrix

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) \approxeq \mathbf{B}\,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\mathbf{B}' \tag{5.17}$$

where

$$\begin{aligned} \mathbf{B} &= \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}} \\ &= \frac{\partial \begin{pmatrix} \kappa \\ \theta \end{pmatrix}}{\partial \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}} \\ &= \begin{pmatrix} 1 & 0 \\ -\frac{\alpha_2}{\alpha_1^2}\exp\left(\frac{\alpha_2}{\alpha_1}\right) & \frac{1}{\alpha_1}\exp\left(\frac{\alpha_2}{\alpha_1}\right) \end{pmatrix} . \end{aligned} \tag{5.18}$$

According to the multivariate delta theorem the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ is

$$\widehat{\boldsymbol{\beta}} \approxeq \mathrm{N}\left(\boldsymbol{\beta}, \mathbf{B}\,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\mathbf{B}'\right) .$$

## 5.2 The log-logistic distribution

The log-logistic distribution is defined in a manner analogous to the definition of the lognormal distribution. If $\log(x)$ follows a logistic distribution then $x$ is said to follow a log-logistic distribution.

The pdf of the log-logistic distribution is

$$f(x;\kappa,\theta) = \frac{e^{\theta}\kappa x^{\kappa-1}}{(1+e^{\theta}x^{\kappa})^2} \tag{5.19}$$

with cdf

$$F(x;\kappa,\theta) = \frac{e^{\theta}x^{\kappa}}{1+e^{\theta}x^{\kappa}} . \tag{5.20}$$

Setting $F(x;\kappa,\theta)=\pi$ it follows that

$$\frac{e^{\theta}x^{\kappa}}{1+e^{\theta}x^{\kappa}}=\pi$$

and therefore

$$\begin{aligned}\frac{\pi}{1-\pi} &= \frac{\left(e^{\theta}x^{\kappa}\right)/\left(1+e^{\theta}x^{\kappa}\right)}{\left(1+e^{\theta}x^{\kappa}-e^{\theta}x^{\kappa}\right)/\left(1+e^{\theta}x^{\kappa}\right)} \\ &= e^{\theta}x^{\kappa}\,. \end{aligned} \tag{5.21}$$

The mean and variance are given by

$$\mu=\exp\left(-\frac{\theta}{\kappa}\right)\left[\Gamma\left(1+\frac{1}{\kappa}\right)\Gamma\left(1-\frac{1}{\kappa}\right)\right] \tag{5.22}$$

and

$$\sigma^2=\exp\left(-\frac{2\theta}{\kappa}\right)\left[\Gamma\left(1+\frac{2}{\kappa}\right)\Gamma\left(1-\frac{2}{\kappa}\right)-\Gamma^2\left(1+\frac{1}{\kappa}\right)\Gamma^2\left(1-\frac{1}{\kappa}\right)\right] \tag{5.23}$$

respectively.

Implementing $\boldsymbol{\pi}=F(\mathbf{x})$, it follows from (5.21) that

$$\frac{\boldsymbol{\pi}}{\mathbf{1}-\boldsymbol{\pi}}=e^{\theta}\mathbf{x}^{\kappa}$$

resulting in the linear model

$$\begin{aligned}\ln\left(\frac{\boldsymbol{\pi}}{\mathbf{1}-\boldsymbol{\pi}}\right) &= \kappa\ln\mathbf{x}+\theta\mathbf{1} \\ &= \begin{pmatrix}\ln\mathbf{x} & \mathbf{1}\end{pmatrix}\begin{pmatrix}\kappa \\ \theta\end{pmatrix} \\ &= \mathbf{X}\boldsymbol{\alpha} \end{aligned} \tag{5.24}$$

where

$$\mathbf{X}=\begin{pmatrix}\ln\mathbf{x} & \mathbf{1}\end{pmatrix} \tag{5.25}$$

and

$$\boldsymbol{\alpha}=\begin{pmatrix}\kappa \\ \theta\end{pmatrix}\,. \tag{5.26}$$

The constraints formulated in terms of a linear model is

$$\mathbf{g}\left(\boldsymbol{\pi}\right) = \mathbf{Q}_X \ln\left(\frac{\boldsymbol{\pi}}{\mathbf{1}-\boldsymbol{\pi}}\right) = \mathbf{0} \tag{5.27}$$

with matrix of partial derivatives

$$\begin{aligned}\mathbf{G}_\pi &= \frac{\partial \mathbf{g}\left(\boldsymbol{\pi}\right)}{\partial \boldsymbol{\pi}} \\ &= \frac{\partial \mathbf{Q}_X \ln\left(\frac{\boldsymbol{\pi}}{\mathbf{1}-\boldsymbol{\pi}}\right)}{\partial \boldsymbol{\pi}} \\ &= \mathbf{Q}_X \mathbf{D}_\pi\end{aligned}$$

where $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ and

$$\begin{aligned}\mathbf{D}_\pi &= \frac{\partial}{\partial \boldsymbol{\pi}}\left\{\ln\left(\frac{\boldsymbol{\pi}}{\mathbf{1}-\boldsymbol{\pi}}\right)\right\} \\ &= \frac{\partial}{\partial \boldsymbol{\pi}}\left\{\ln\left(\boldsymbol{\pi}\right) - \ln\left(\mathbf{1}-\boldsymbol{\pi}\right)\right\} \\ &= \left\{\operatorname{diag}\left[\boldsymbol{\pi}\right]\right\}^{-1} + \left\{\operatorname{diag}\left[\mathbf{1}-\boldsymbol{\pi}\right]\right\}^{-1} \ .\end{aligned} \tag{5.28}$$

In the ML estimation procedure $\widehat{\boldsymbol{\pi}}$ is estimated such that $\ln\left(\frac{\widehat{\boldsymbol{\pi}}}{\mathbf{1}-\widehat{\boldsymbol{\pi}}}\right)$ is in the vector space of $\mathbf{X}$. The ML estimator $\widehat{\boldsymbol{\alpha}}$ follows from the linear model (5.24)

$$\widehat{\boldsymbol{\alpha}} = \begin{pmatrix}\widehat{\kappa} \\ \widehat{\theta}\end{pmatrix} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \ln\left(\frac{\widehat{\boldsymbol{\pi}}}{\mathbf{1}-\widehat{\boldsymbol{\pi}}}\right) \tag{5.29}$$

with asymptotic covariance matrix

$$\operatorname{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) = \operatorname{Cov}\begin{pmatrix}\widehat{\kappa} \\ \widehat{\theta}\end{pmatrix} = \left\{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}_\pi\right\}\operatorname{Cov}\left(\widehat{\boldsymbol{\alpha}}\right)\left\{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}_\pi\right\}' \tag{5.30}$$

where $\mathbf{D}_\pi$ is derived in (5.28). As in the case of the Weibull distribution, the ML estimators of the log-logistic are approximately normally distributed.

## 5.3 The Pareto distribution

The Pareto distribution has been successfully used to model the income of a population (Johnson & Kotz (1970)). The pdf and cdf of the Pareto distribution are

$$f(x,\kappa,\theta) = \kappa\theta^{\kappa}x^{-(\kappa+1)} \tag{5.31}$$

and

$$F(x) = 1 - \left(\frac{x}{\theta}\right)^{-\kappa} \tag{5.32}$$

for $x > \theta,\ \theta > 0$ and $\kappa > 0$ .

The mean and variance for the Pareto distribution are given by

$$\mu = \frac{\kappa\theta}{\kappa - 1} \qquad \kappa > 1 \tag{5.33}$$

and

$$\sigma^2 = \frac{\kappa\theta^2}{(\kappa-1)^2(\kappa-2)} \qquad \kappa > 2 \tag{5.34}$$

respectively.

To fit a Pareto distribution it is required that

$$\boldsymbol{\pi} = \mathbf{1} - \left(\frac{\mathbf{x}}{\theta}\right)^{-\kappa} . \tag{5.35}$$

Taking the natural logarithm of (5.35) leads to

$$\begin{aligned}
\ln(\mathbf{1}-\boldsymbol{\pi}) &= -\kappa \ln\left(\frac{\mathbf{x}}{\theta}\right) \\
&= -\kappa(\ln\mathbf{x} - \ln\theta) \\
&= \begin{pmatrix} -\ln\mathbf{x} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \kappa \\ \kappa\ln\theta \end{pmatrix} \\
&= \mathbf{X}\boldsymbol{\alpha}
\end{aligned} \tag{5.36}$$

where

$$\mathbf{X} = \begin{pmatrix} -\ln\mathbf{x} & \mathbf{1} \end{pmatrix} \tag{5.37}$$

is the design matrix and

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \kappa \\ \kappa\ln\theta \end{pmatrix} \tag{5.38}$$

is the vector of natural parameters.

Hence, the vector of constraints may be written as

$$\mathbf{g}(\boldsymbol{\pi}) = \mathbf{Q}_X \ln(\mathbf{1} - \boldsymbol{\pi}) = \mathbf{0} \tag{5.39}$$

where $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. This implies that the restricted ML estimate $\widehat{\boldsymbol{\pi}}$ will be fitted such that $\ln(\mathbf{1} - \boldsymbol{\pi})$ is orthogonal to the error space of $\mathbf{X}$ with matrix of partial derivatives

$$\begin{aligned} \mathbf{G}_\pi &= \mathbf{Q}_X \frac{\partial \mathbf{g}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= \frac{\partial \mathbf{Q}_X \ln(\mathbf{1} - \boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= \mathbf{Q}_X \mathbf{D}_\pi \end{aligned}$$

where

$$\begin{aligned} \mathbf{D}_\pi &= \frac{\partial \ln(\mathbf{1} - \boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \\ &= -\{\text{diag}[\mathbf{1} - \boldsymbol{\pi}]\}^{-1} . \end{aligned} \tag{5.40}$$

The ML estimator for $\boldsymbol{\alpha}$ follows

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \ln(\mathbf{1} - \widehat{\boldsymbol{\pi}}) \tag{5.41}$$

with asymptotic covariance matrix

$$\text{Cov}(\widehat{\boldsymbol{\alpha}}) = \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\} \text{Cov}(\widehat{\boldsymbol{\pi}}) \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\}' . \tag{5.42}$$

Define the vector of parameters for the Pareto distribution

$$\boldsymbol{\beta} = \begin{pmatrix} \kappa \\ \theta \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \exp\left(\dfrac{\alpha_2}{\alpha_1}\right) \end{pmatrix} . \tag{5.43}$$

(The parameterization follows from (5.38).)

Therefore the ML estimates for $\kappa$ and $\theta$ are

$$\widehat{\kappa} = \widehat{\alpha}_1 \quad \text{and} \quad \widehat{\theta} = \exp\left(\frac{\widehat{\alpha}_2}{\widehat{\alpha}_1}\right)$$

implying that

$$\widehat{\boldsymbol{\beta}} \approxeq \mathrm{N}\left(\begin{pmatrix} \kappa \\ \theta \end{pmatrix}, \mathbf{B}\,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\mathbf{B}'\right)$$

where

$$\mathbf{B} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}} = \begin{pmatrix} 1 & 0 \\ -\frac{\alpha_2}{\alpha_1^2} \cdot \exp\left(\frac{\alpha_2}{\alpha_1}\right) & \frac{1}{\alpha_1} \cdot \exp\left(\frac{\alpha_2}{\alpha_1}\right) \end{pmatrix}.$$

## 5.4 Generalization

In this section a short summary of fitting the distributions, tabulated in Table 5.1 will be given.

**Table 5.1:** Characteristics of distributions considered.

| | PDF and CDF | Mean and Variance |
|---|---|---|
| Exponential | $f(x;\mu) = \frac{1}{\mu}e^{-x/\mu}$ <br> $F(x;\mu) = 1 - e^{-x/\mu}$ | $\mu$ <br> $\sigma^2 = \mu^2$ |
| Normal | $f(x;\mu,\sigma^2) = \boldsymbol{\phi}\left(\frac{x-\mu}{\sigma}\right)$ <br> $F(x;\mu,\sigma^2) = \boldsymbol{\Phi}\left(\frac{x-\mu}{\sigma}\right)$ | $\mu$ <br> $\sigma^2$ |
| Weibull | $f(x;\kappa,\theta) = \frac{\kappa}{\theta^\kappa}x^{\kappa-1}\exp\left[-\left(\frac{x}{\theta}\right)^\kappa\right]$ <br> $F(x;\kappa,\theta) = 1 - \exp\left[-\left(\frac{x}{\theta}\right)^\kappa\right]$ | $\mu = \theta\left[\Gamma\left(1+\frac{1}{\kappa}\right)\right]$ <br> $\sigma^2 = \theta^2\left[\Gamma\left(1+\frac{2}{\kappa}\right) - \Gamma^2\left(1+\frac{1}{\kappa}\right)\right]$ |
| Log-logistic | $f(x;\kappa,\theta) = \dfrac{e^\theta \kappa x^{\kappa-1}}{(1+e^\theta x^\kappa)^2}$ <br> $F(x;\kappa,\theta) = \dfrac{e^\theta x^\kappa}{1+e^\theta x^\kappa}$ | $\mu = \exp\left(-\frac{\theta}{\kappa}\right)\left[\Gamma\left(1+\frac{1}{\kappa}\right)\Gamma\left(1-\frac{1}{\kappa}\right)\right]$ <br> $\sigma^2 = \exp\left(-\frac{2\theta}{\kappa}\right)\left[\Gamma\left(1+\frac{2}{\kappa}\right)\Gamma\left(1-\frac{2}{\kappa}\right)\right.$ <br> $\left.-\Gamma^2\left(1+\frac{1}{\kappa}\right)\Gamma^2\left(1-\frac{1}{\kappa}\right)\right]$ |
| Pareto | $f(x;\kappa,\theta) = \kappa\theta^\kappa x^{-(\kappa+1)}$ <br> $F(x;\kappa,\theta) = 1 - \left(\frac{x}{\theta}\right)^{-\kappa}$ | $\mu = \dfrac{\kappa\theta}{\kappa-1}$ <br> $\sigma^2 = \dfrac{\kappa\theta^2}{(\kappa-1)^2(\kappa-2)}$ |

In the case of the distributions $F(x;\boldsymbol{\beta})$, specified in Table 5.1, the requirement

$$F(\mathbf{x};\boldsymbol{\beta}) = \boldsymbol{\pi} \tag{5.44}$$

where $F(\mathbf{x};\boldsymbol{\beta})$ denotes the distribution function at the upper class boundaries $\mathbf{x}$ with parameter vector $\boldsymbol{\beta}$, may be transformed into the linear model

$$\mathbf{h}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\alpha} \tag{5.45}$$

which implies that the estimation procedure may be performed in the three steps outlined below.

**Step 1:** The vector of constraints is given by

$$\mathbf{g}(\boldsymbol{\pi}) = \mathbf{Q}_X\mathbf{h}(\boldsymbol{\pi}) = \mathbf{0} \tag{5.46}$$

with matrix of partial derivatives

$$\mathbf{G}_\pi = \mathbf{Q}_X\mathbf{D}_\pi \tag{5.47}$$

where $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{D}_\pi = \dfrac{\partial \mathbf{h}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}$.

**Step 2:** The ML estimate of $\boldsymbol{\alpha}$ follows as

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{h}(\widehat{\boldsymbol{\pi}}) \tag{5.48}$$

with asymptotic covariance matrix

$$\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}) \approxeq \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\} \mathrm{Cov}(\widehat{\boldsymbol{\pi}}) \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\}' . \tag{5.49}$$

**Step 3:** The ML estimates of the original parameters namely $\widehat{\boldsymbol{\beta}}$, are obtained from $\widehat{\boldsymbol{\alpha}}$ with

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}\right) \approxeq \mathbf{B}\,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\,\mathbf{B}' \tag{5.50}$$

where $\mathbf{B} = \dfrac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}}$. From the multivariate delta theorem, it follows that

$$\widehat{\boldsymbol{\beta}} \approxeq \mathrm{N}\left(\boldsymbol{\beta}, \mathbf{B}\,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\mathbf{B}'\right) . \tag{5.51}$$

To fit the various continuous distributions in Table 5.1 to grouped data by means of the three steps listed above, a summary of the constraints and derivatives are given in Table 5.2(A) & Table 5.2(B).

**Table 5.2(A):** Constraints

| | $\boldsymbol{\beta}$ | $\mathbf{h}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\alpha}$: $\mathbf{h}(\boldsymbol{\pi})$ | $\mathbf{X}$ | $\boldsymbol{\alpha}$ |
|---|---|---|---|---|
| Exponential | $\mu = \frac{1}{\alpha}$ | $\ln(\mathbf{1} - \boldsymbol{\pi})$ | $(-\mathbf{x})$ | $\frac{1}{\mu}$ |
| Normal | $\begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} \frac{\alpha_2}{\alpha_1} \\ \frac{1}{\alpha_1} \end{pmatrix}$ | $\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi})$ | $\begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix}$ | $\begin{pmatrix} \frac{1}{\sigma} \\ \frac{\mu}{\sigma} \end{pmatrix}$ |
| Weibull | $\begin{pmatrix} \kappa \\ \theta \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ e^{\frac{\alpha_2}{\alpha_1}} \end{pmatrix}$ | $\ln[-\ln(\mathbf{1} - \boldsymbol{\pi})]$ | $\begin{pmatrix} \ln\mathbf{x} & -\mathbf{1} \end{pmatrix}$ | $\begin{pmatrix} \kappa \\ \kappa\ln\theta \end{pmatrix}$ |
| Log-logistic | $\begin{pmatrix} \kappa \\ \theta \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ | $\ln\left(\frac{\boldsymbol{\pi}}{\mathbf{1} - \boldsymbol{\pi}}\right)$ | $\begin{pmatrix} \ln\mathbf{x} & \mathbf{1} \end{pmatrix}$ | $\begin{pmatrix} \kappa \\ \theta \end{pmatrix}$ |
| Pareto | $\begin{pmatrix} \kappa \\ \theta \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ e^{\frac{\alpha_2}{\alpha_1}} \end{pmatrix}$ | $\ln(\mathbf{1} - \boldsymbol{\pi})$ | $\begin{pmatrix} -\ln\mathbf{x} & \mathbf{1} \end{pmatrix}$ | $\begin{pmatrix} \kappa \\ \kappa\ln\theta \end{pmatrix}$ |

**Table 5.2(B):** Derivatives

| | $\mathbf{D} = \frac{\partial\mathbf{h}(\boldsymbol{\pi})}{\partial\boldsymbol{\pi}}$ | $\mathbf{B} = \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\alpha}}$ |
|---|---|---|
| Exponential | $-(\text{diag}[\mathbf{1} - \boldsymbol{\pi}])^{-1}$ | $-\frac{1}{\alpha^2}$ |
| Normal | $(\text{diag}[\boldsymbol{\phi}(\boldsymbol{\Phi}^{-1}(\boldsymbol{\pi}))])^{-1}$ | $\begin{pmatrix} -\frac{\alpha_2}{\alpha_1^2} & \frac{1}{\alpha_1} \\ -\frac{1}{\alpha_1^2} & 0 \end{pmatrix}$ |
| Weibull | $-(\text{diag}[\ln(\mathbf{1} - \boldsymbol{\pi})])^{-1}(\text{diag}[\mathbf{1} - \boldsymbol{\pi}])^{-1}$ | $\begin{pmatrix} 1 & 0 \\ -\frac{\alpha_2}{\alpha_1^2} \cdot e^{\frac{\alpha_2}{\alpha_1}} & \frac{1}{\alpha_1} \cdot e^{\frac{\alpha_2}{\alpha_1}} \end{pmatrix}$ |
| Log-logistic | $(\text{diag}[\boldsymbol{\pi}])^{-1} + (\text{diag}[\mathbf{1} - \boldsymbol{\pi}])^{-1}$ | $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ |
| Pareto | $-(\text{diag}[\mathbf{1} - \boldsymbol{\pi}])^{-1}$ | $\begin{pmatrix} 1 & 0 \\ -\frac{\alpha_2}{\alpha_1^2} \cdot e^{\frac{\alpha_2}{\alpha_1}} & \frac{1}{\alpha_1} \cdot e^{\frac{\alpha_2}{\alpha_1}} \end{pmatrix}$ |

**Example 5.1**

A typical example was taken from a data set with $n = 206$ insurance policies. The annual income (in R1000) of the policy holders is reported in Table 5.3.

**Table 5.3:** Income of a group of insurance policy holders.

| Income (in R1000) | $[0, 40)$ | $[40, 75)$ | $[75, 125)$ | $[125, 175)$ | $[175, \infty)$ |
|---|---|---|---|---|---|
| Frequency | 9 | 37 | 67 | 63 | 30 |

For this example the normal, Weibull and log-logistic distributions are fitted and the results are given in Table 5.4.

**Table 5.4:** Estimates of parameters and test statistics

| | MLE | | | | | Wald | | | Discrep- |
|---|---|---|---|---|---|---|---|---|---|
| | $\widehat{\boldsymbol{\beta}}$ | Estimate | Std. Error | $\widehat{\mu}$ | $\widehat{\sigma}$ | Statistic | df | prob | ancy |
| Normal | $\widehat{\mu}$<br>$\widehat{\sigma}$ | 118.4<br>51.4 | 3.7604<br>3.0834 | 118.4 | 51.4 | 3.980 | 2 | 0.1367 | 0.019 |
| Weibull | $\widehat{\kappa}$<br>$\widehat{\theta}$ | 2.4647<br>134.44 | 0.1675<br>4.2552 | 119.2 | 51.7 | 1.293 | 2 | 0.5240 | 0.006 |
| Log-logistic | $\widehat{\kappa}$<br>$\widehat{\theta}$ | 3.3337<br>$-15.710$ | 0.2293<br>1.0883 | 129.7 | 88.0 | 8.731 | 2 | 0.0127 | 0.042 |

According to the Wald statistic the Weibull distribution provided the best fit, followed by the normal distribution. The distributions are illustrated in Figure 5.2. In constructing the histogram, it is assumed that the income of all the policy holders in the sample is less than R500 000. The distributions were all fitted with the SAS program *FIT.SAS* listed in Appendix A.

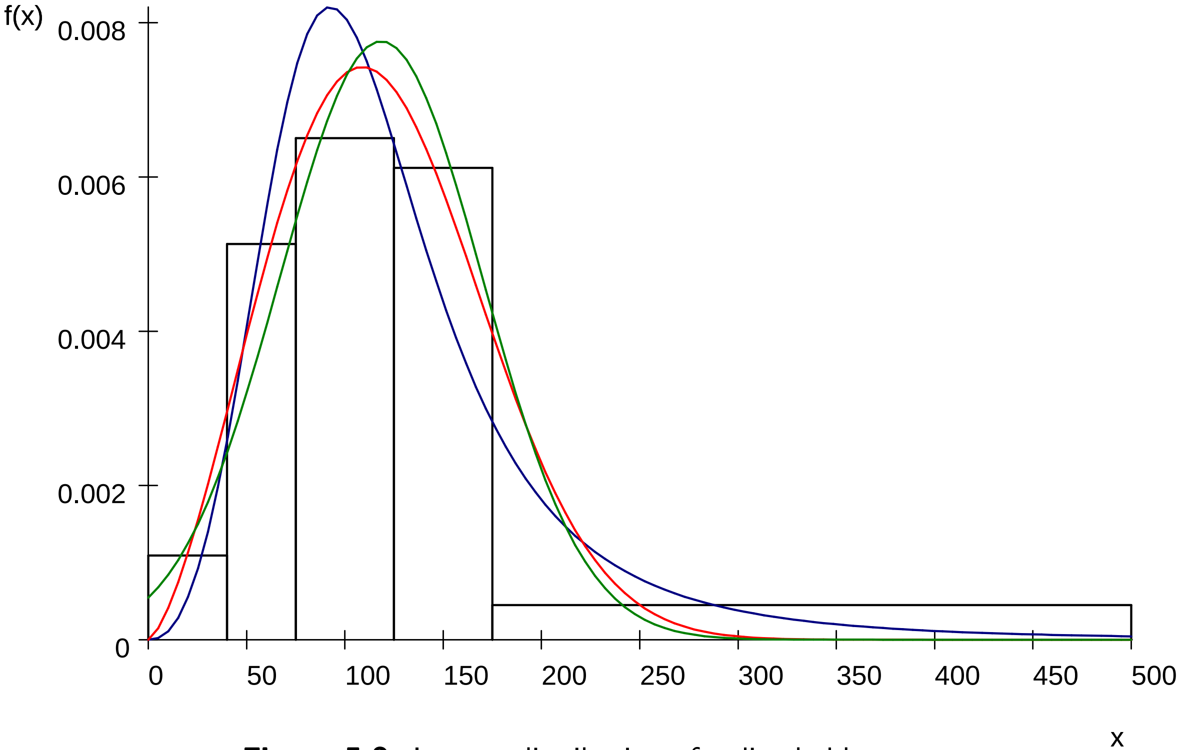



**Figure 5.2:** Income distribution of policy holders.

| | | |
|---|---|---|
| Normal: *Green* | Weibull: *Red* | Log-logistic: *Blue* |

# Part II

# Linear models for grouped data

# Chapter 6

# Multifactor design

Consider any single-factor or multifactor design resulting in a cross classification of $T$ different cells to be analysed. The response vector in each cell is a frequency distribution of an underlying continuous response variable, categorised in $k$ class intervals. The focus is to model the behavior of this grouped response variable over the $T$ cells to evaluate the effect of the explanatory variables on the dependent variable. The basic formulation of the grouped response variable, to be modeled over the $T$ cells of the multifactor design is summarised in Table 6.1.

**Table 6.1:** Grouped data in a multifactor design.

| | Class interval | | | | |
|---|---|---|---|---|---|
| Cells | $(-\infty, x_1)$ | $[x_1, x_2)$ | $\cdots$ | $[x_{k-2}, x_{k-1})$ | $[x_{k-1}, \infty)$ |
| 1 | $f_{11}$ | $f_{12}$ | $\cdots$ | $f_{1,k-1}$ | $f_{1k}$ |
| 2 | $f_{21}$ | $f_{22}$ | $\cdots$ | $f_{2,k-1}$ | $f_{2k}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\cdots$ | $\vdots$ | $\vdots$ |
| $T$ | $f_{T1}$ | $f_{T2}$ | $\cdots$ | $f_{T,k-1}$ | $f_{Tk}$ |

## 6.1 Formulation

Considering the frequencies tabulated in Table 6.1, let

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1,k-1} \\ f_{21} & f_{22} & \cdots & f_{2,k-1} \\ \vdots & \vdots & \cdots & \vdots \\ f_{T1} & f_{T2} & \cdots & f_{T,k-1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}'_1 \\ \mathbf{f}'_2 \\ \vdots \\ \mathbf{f}'_T \end{pmatrix} : T \times (k-1) \tag{6.1}$$

be the matrix where the rows of $\mathbf{F}$ denote the $T$ cells of the multifactor design and the columns of $\mathbf{F}$ denote the first $(k-1)$ class intervals of the grouped response variable. Similarly to the estimation of distribution functions done in Part I, only the first $(k-1)$ class intervals need to be considered for each cell.

Define

$$\text{vec}(\mathbf{F}) = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_T \end{pmatrix} : T(k-1) \times 1 \tag{6.2}$$

as the so-called concatenated frequency vector where the $T$ **rows** of $\mathbf{F}$ in (6.1) are stacked row by row in a single column vector. The frequency vector for the $t$-th cell in (6.2) is

$$\mathbf{f}_t = \begin{pmatrix} f_{t1} \\ f_{t2} \\ \vdots \\ f_{t,k-1} \end{pmatrix} \qquad t = 1, 2 \cdots T \tag{6.3}$$

and consists of the first $(k-1)$ frequencies with corresponding vector of upper class boundaries

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \end{pmatrix} . \tag{6.4}$$

**Note:** The definition of $\text{vec}(\mathbf{F})$ (6.2) differs from the standard definition where the **columns** of $\mathbf{F}$ (6.1) are stacked as a single column vector. (See *Muirhead (1972) (p.17)*). However, by stacking the rows below each other coincides with the definition of the COLVEC function in SAS which is used extensively in this thesis for the computer programming of applications of grouped data in a multifactor design.

It is assumed that the vector $\mathbf{f}$ is a product multinomial vector with fixed subtotals

$$\mathbf{n} = \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_T \end{pmatrix} \tag{6.5}$$

allocated to each of the $T$ cells.

Define

$$\mathbf{p}_0 = \begin{pmatrix} \mathbf{p}_{01} \\ \mathbf{p}_{02} \\ \vdots \\ \mathbf{p}_{0T} \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1}\mathbf{f}_1 \\ \frac{1}{n_2}\mathbf{f}_2 \\ \vdots \\ \frac{1}{n_T}\mathbf{f}_T \end{pmatrix} = \left((\text{diag}\,(\mathbf{n}))^{-1} \otimes \mathbf{I}_{k-1}\right) \cdot \mathbf{f} \tag{6.6}$$

as the concatenated vector of relative frequencies for the $T$ cells. Hence, let

$$\text{E}(\mathbf{p}_0) = \begin{pmatrix} \boldsymbol{\pi}_{01} \\ \boldsymbol{\pi}_{02} \\ \vdots \\ \boldsymbol{\pi}_{0T} \end{pmatrix} = \boldsymbol{\pi}_0$$

then

$$\text{Cov}(\mathbf{p}_0) = \begin{pmatrix} \mathbf{V}_{01} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{02} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_{0T} \end{pmatrix} = \mathbf{V}_0 \tag{6.7}$$

where

$$\mathrm{Cov}\left(\mathbf{p}_{0t}\right)=\frac{1}{n_t}\left(\mathrm{diag}\left(\boldsymbol{\pi}_{0t}\right)-\boldsymbol{\pi}_{0t}\boldsymbol{\pi}_{0t}^{\prime}\right)=\mathbf{V}_{0t} \qquad ,\ t=1,\cdots,T \tag{6.8}$$

is the covariance matrix for the vector of relative frequencies for the $t$-th cell.

Following (6.7) and (6.8) the covariance matrix of $\mathbf{p}_0$ may be expressed in terms of Kronecker products

$$\mathbf{V}_0=\left\{\left(\mathrm{diag}\left[\mathbf{n}\right]\right)^{-1}\otimes\mathbf{I}_{k-1}\right\}\cdot\left\{\mathrm{diag}\left[\boldsymbol{\pi}_0\right]-\mathrm{diag}\left[\boldsymbol{\pi}_0\right]\left(\mathbf{I}_T\otimes\left(\mathbf{1}_{k-1}\mathbf{1}_{k-1}^{\prime}\right)\right)\mathrm{diag}\left[\boldsymbol{\pi}_0\right]\right\} \tag{6.9}$$

where $\mathbf{1}_{k-1}$ is a $(k-1)$ vector of ones.

Define the concatenated vector of cumulative relative frequencies

$$\mathbf{p}=\begin{pmatrix}\mathbf{p}_1\\ \mathbf{p}_2\\ \vdots\\ \mathbf{p}_T\end{pmatrix}=\begin{pmatrix}\mathbf{Cp}_{01}\\ \mathbf{Cp}_{02}\\ \vdots\\ \mathbf{Cp}_{0T}\end{pmatrix}=\left(\mathbf{I}_T\otimes\mathbf{C}\right)\mathbf{p}_0 \tag{6.10}$$

where

$$\mathbf{C}=\begin{pmatrix}1&0&\cdots&0\\ 1&1&\cdots&0\\ \vdots&\vdots&\ddots&\vdots\\ 1&1&\cdots&1\end{pmatrix}:(k-1)\times(k-1)\ . \tag{6.11}$$

In (6.10) $\mathbf{p}_t=\mathbf{Cp}_{0t}$ for $t=1,2,\cdots,T$ is the cumulative relative frequency vector for the $t$-th cell in the multifactor design.

The random vector $\mathbf{p}$ consists of the cumulative relative frequencies from $T$ independent multinomial populations, therefore let

$$\mathrm{E}(\mathbf{p})=\begin{pmatrix}\boldsymbol{\pi}_1\\ \boldsymbol{\pi}_2\\ \vdots\\ \boldsymbol{\pi}_T\end{pmatrix}=\boldsymbol{\pi} \tag{6.12}$$

where

$$\mathrm{E}(\mathbf{p}_t)=\boldsymbol{\pi}_t \qquad ,\ t=1,\cdots,T$$

is the expected value for the vector of cumulative relative frequencies for the $t$-th cell and

$$\mathrm{Cov}(\mathbf{p}) = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_T \end{pmatrix} = \mathbf{V} \tag{6.13}$$

where

$$\begin{aligned} \mathrm{Cov}(\mathbf{p}_t) &= \frac{1}{n_t}\left\{\mathbf{C}\,\mathrm{diag}\left(\mathbf{C}^{-1}\boldsymbol{\pi}_t\right)\mathbf{C}' - \boldsymbol{\pi}_t\boldsymbol{\pi}_t'\right\} \\ &= \mathbf{V}_t \qquad , \; t = 1, \cdots, T \end{aligned} \tag{6.14}$$

is the covariance matrix for the vector of cumulative relative frequencies for the $t$-th cell.

From (6.10) it follows that the covariance matrix of $\mathbf{p}$ may also be expressed by

$$\mathbf{V} = (\mathbf{I}_T \otimes \mathbf{C})\,\mathbf{V}_0\,(\mathbf{I}_T \otimes \mathbf{C})' \tag{6.15}$$

where $\mathbf{V}_0$ is the covariance matrix of $\mathbf{p}_0$ in (6.9).

**Note:** For simplicity the class boundaries $\mathbf{x}$ are assumed to be constant over the different cells. The extension to the situation where this is not the case, can be done in a straight forward way.

## 6.2 Estimation

The ML estimation procedure entails that distribution fitting be done under the restriction that the cumulative relative frequencies equal the cumulative distribution curve at the upper class boundaries, for every cell in the multifactor design, i.e.

$$\begin{pmatrix} F_1(\mathbf{x}, \boldsymbol{\beta}_1) \\ F_2(\mathbf{x}, \boldsymbol{\beta}_2) \\ \vdots \\ F_T(\mathbf{x}, \boldsymbol{\beta}_T) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \vdots \\ \boldsymbol{\pi}_T \end{pmatrix} \tag{6.16}$$

with

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_T \end{pmatrix} \tag{6.17}$$

the concatenated vector of original parameters to be estimated.

Utilizing the ML estimation procedure, the vector of constraints to be imposed is

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{pmatrix} F_1(\mathbf{x}, \boldsymbol{\beta}_1) \\ F_2(\mathbf{x}, \boldsymbol{\beta}_2) \\ \vdots \\ F_T(\mathbf{x}, \boldsymbol{\beta}_T) \end{pmatrix} - \begin{pmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \vdots \\ \boldsymbol{\pi}_T \end{pmatrix} = \mathbf{0}. \tag{6.18}$$

In the case where (6.16) may be transformed into the linear model

$$\mathbf{h}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{X}\boldsymbol{\alpha}_1 \\ \mathbf{X}\boldsymbol{\alpha}_2 \\ \vdots \\ \mathbf{X}\boldsymbol{\alpha}_2 \end{pmatrix} = (\mathbf{I}_T \otimes \mathbf{X})\,\boldsymbol{\alpha} \tag{6.19}$$

with

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_T \end{pmatrix} \tag{6.20}$$

a simultaneous distribution fitting for the $T$ frequency distributions is outlined in the following three steps.

**Step 1:** The restricted ML estimate $\widehat{\boldsymbol{\pi}}$ is obtained by implementing the vector of constraints, $\mathbf{g}(\boldsymbol{\pi}) = \mathbf{0}$, with

$$\mathbf{g}(\boldsymbol{\pi}) = (\mathbf{I}_T \otimes \mathbf{Q}_X)\,\mathbf{h}(\boldsymbol{\pi}) \tag{6.21}$$

and matrix of partial derivatives

$$\mathbf{G}_\pi = (\mathbf{I}_T \otimes \mathbf{Q}_X)\,\mathbf{D}_\pi \tag{6.22}$$

where $\mathbf{Q}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{D}_\pi = \dfrac{\partial \mathbf{h}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}$ in the ML estimation process.

**Step 2:** The ML estimate of $\boldsymbol{\alpha}$ follows as

$$\widehat{\boldsymbol{\alpha}} = \left(\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\right)\mathbf{h}(\widehat{\boldsymbol{\pi}}) \tag{6.23}$$

with asymptotic covariance matrix

$$\mathrm{Cov}\,(\widehat{\boldsymbol{\alpha}}) \approxeq \left\{\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\}\mathrm{Cov}\,(\widehat{\boldsymbol{\pi}})\left\{\mathbf{I}_T \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_\pi\right\}' \ . \tag{6.24}$$

**Step 3:** The ML estimates of the original parameters namely $\widehat{\boldsymbol{\beta}}$, are obtained from $\widehat{\boldsymbol{\alpha}}$ with

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}\right) \approxeq \mathbf{B}\,\mathrm{Cov}\,(\widehat{\boldsymbol{\alpha}})\,\mathbf{B}' \tag{6.25}$$

where $\mathbf{B} = \dfrac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}}$. From the multivariate delta theorem, it follows that

$$\widehat{\boldsymbol{\beta}} \approxeq \mathrm{N}\left(\boldsymbol{\beta}, \mathbf{B}\,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\mathbf{B}'\right) \ . \tag{6.26}$$

It follows from (6.23) that each of the $T$ estimated distribution functions will have its own set of parameter estimates characterising the shape and locality of the distribution. Certain parameter structures may now be defined which may be incorporated to evaluate the effect of the factor(s) on the respons variable in any multiway design.

# 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

$$\mathbf{\Phi}\left(\mathbf{z}\right)=\boldsymbol{\pi} \tag{7.1}$$

where

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

is the concatenated vector of standardised upper class boundaries and

$$\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \vdots \\ \boldsymbol{\pi}_T \end{pmatrix} \tag{7.3}$$

is the concatenated vector of cumulative relative frequencies.

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

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

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{x} & -\mathbf{1} \end{pmatrix} \tag{7.5}$$

is the design matrix for normality within each cell and

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_T \end{pmatrix} \tag{7.6}$$

is the concatenated vector of natural parameters with

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma_t} \\ \dfrac{\mu_t}{\sigma_t} \end{pmatrix} \quad t = 1 \cdots T \tag{7.7}$$

the natural parameters for the $t$-th cell.

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

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

and

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

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

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

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

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

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

and matrix of partial derivatives

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

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

$$\text{Cov}\left(\widehat{\boldsymbol{\pi}}\right) \approxeq \mathbf{V} - \left(\mathbf{G}_\pi \mathbf{V}\right)' \left(\mathbf{G}_\pi \mathbf{V} \mathbf{G}_\pi'\right)^* \left(\mathbf{G}_\pi \mathbf{V}\right) \ .$$

For each of the $T$ subpopulations, the vector of restricted cumulative relative frequencies $\widehat{\boldsymbol{\pi}}_t$ for $t = 1, 2, \cdots, T$ follow a cumulative normal distribution curve at the upper class boundaries of $\mathbf{x}$. Each $\boldsymbol{\Phi}^{-1}(\widehat{\boldsymbol{\pi}}_\mathbf{t})$ for $t = 1, 2, \cdots, T$ 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)

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

which consists of two sets of estimators namely

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

and

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

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

It follows that

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

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

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

and

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

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

Let

$$\begin{aligned}\boldsymbol{\beta} &= \left(\boldsymbol{\mu}\otimes\begin{pmatrix}1\\0\end{pmatrix}\right) + \left(\boldsymbol{\sigma}\otimes\begin{pmatrix}0\\1\end{pmatrix}\right) \\ &= \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\\ \vdots\\ \boldsymbol{\beta}_T\end{pmatrix}\end{aligned} \tag{7.19}$$

be the concatenated vector of original parameters with

$$\boldsymbol{\beta}_t = \begin{pmatrix} \mu_t \\ \sigma_t \end{pmatrix} = \begin{pmatrix} \alpha_{2t}/\alpha_{1t} \\ 1/\alpha_{1t} \end{pmatrix} \qquad t = 1, 2, \cdots T \ . \tag{7.20}$$

Hence

$$\begin{aligned} \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}\right) &\approxeq \left(\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}}\right) \mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) \left(\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}}\right)' \\ &= \mathbf{B}\,\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right)\mathbf{B}' \end{aligned} \tag{7.21}$$

where

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

with

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

the partial derivatives for the $t$-th cell.

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

$$\mathbf{B} = \left(-\frac{\boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_1^2} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right) + \left(\frac{\mathbf{1}}{\boldsymbol{\alpha}_1} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right) + \left(-\frac{\mathbf{1}}{\boldsymbol{\alpha}_1^2} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right) . \tag{7.24}$$

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

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\mu}}\right) \approxeq \mathbf{B}_\mu\,\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right)\mathbf{B}_\mu' \tag{7.25}$$

where

$$\mathbf{B}_\mu = \left(\frac{\boldsymbol{\alpha}_2}{\boldsymbol{\alpha}_1^2} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}\right) + \left(\frac{\mathbf{1}}{\boldsymbol{\alpha}_1} \otimes \begin{pmatrix} 0 & 1 \end{pmatrix}\right) \tag{7.26}$$

and

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\sigma}}\right) \approxeq \mathbf{B}_{\sigma}\,\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right)\mathbf{B}_{\sigma}' \tag{7.27}$$

where

$$\mathbf{B}_{\sigma} = \left(-\frac{\mathbf{1}}{\boldsymbol{\alpha}_1^2} \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}\right) . \tag{7.28}$$

## 7.2 Equality of variances

Equality of variances

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

is expressed in terms of matrix notation as

$$\mathbf{H}\boldsymbol{\alpha}_1 = \mathbf{0} \tag{7.30}$$

where

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

is a matrix of contrasts and

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

is a subset of the vector of natural parameters $\boldsymbol{\alpha}$ 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

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

and matrix of partial derivatives

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

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

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

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

and

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

in the ML estimation procedure.

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

1. $\widehat{\boldsymbol{\pi}}_t$, $(t = 1, 2, \cdots, T)$ 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 $\mu$, of the fitted normal distribution will be used as a representative measure for each cell.

Formulate the linear model

$$\boldsymbol{\mu} = \mathbf{Y}\boldsymbol{\gamma} \tag{7.37}$$

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

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

where $(y_1, y_1, \cdots, y_T)$ 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

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

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

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

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

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

Utilizing the maximum likelihood procedure with

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

and

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

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

1. $\widehat{\boldsymbol{\pi}}_t$ for $t = 1, 2, \cdots, T$ 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 $\widehat{\boldsymbol{\mu}}$ 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

$$\widehat{\gamma} = (\mathbf{Y}'\mathbf{Y})^* \, \mathbf{Y}'\widehat{\boldsymbol{\mu}} \tag{7.44}$$

with asymptotic covariance matrix

$$\mathrm{Cov}\,(\widehat{\gamma}) = \left\{(\mathbf{Y}'\mathbf{Y})^* \, \mathbf{Y}'\right\} \mathrm{Cov}\,(\widehat{\boldsymbol{\mu}}\ ) \left\{(\mathbf{Y}'\mathbf{Y})^* \, \mathbf{Y}'\right\}' . \tag{7.45}$$

## 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

$$\mathbf{x} = \begin{pmatrix} 39.5 \\ 49.5 \\ 59.5 \\ 74.4 \end{pmatrix} \tag{7.46}$$

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, specifing 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 $\mathbf{\Phi}^{-1}(\boldsymbol{\pi}_t)$ for $t = 1, 2, \cdots 5$ is a linear combination of

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

or equivalently such that $\mathbf{\Phi}^{-1}(\boldsymbol{\pi}_t)$ is orthogonal to

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

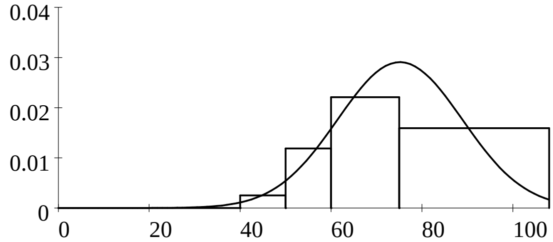
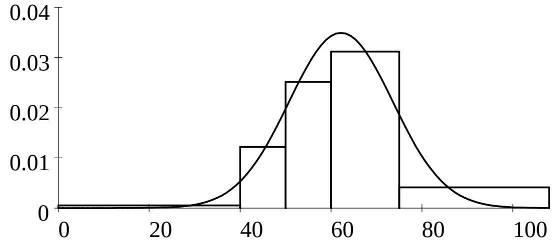
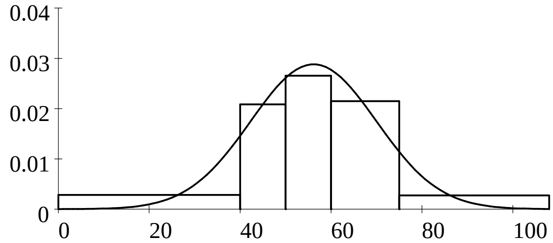
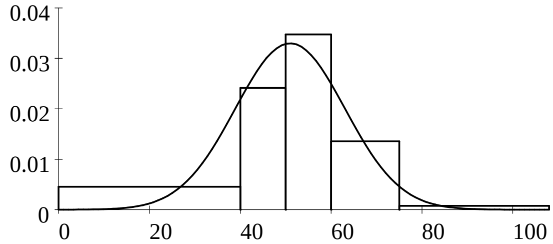
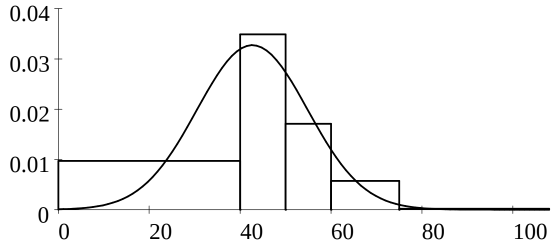
Since $\mathrm{rank}\left(\mathbf{Q}_X\right) = 2$ the vector of constraints $\mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) = \mathbf{0}$, with

$$\begin{aligned} \mathbf{g}_{\text{nor}}(\boldsymbol{\pi}) &= \begin{pmatrix} \mathbf{Q}_X\mathbf{\Phi}^{-1}(\boldsymbol{\pi}_1) \\ \mathbf{Q}_X\mathbf{\Phi}^{-1}(\boldsymbol{\pi}_2) \\ \mathbf{Q}_X\mathbf{\Phi}^{-1}(\boldsymbol{\pi}_3) \\ \mathbf{Q}_X\mathbf{\Phi}^{-1}(\boldsymbol{\pi}_4) \\ \mathbf{Q}_X\mathbf{\Phi}^{-1}(\boldsymbol{\pi}_5) \end{pmatrix} \\ &= \left(\mathbf{I}_5 \otimes \mathbf{Q}_X\right) \cdot \mathbf{\Phi}^{-1}(\boldsymbol{\pi}) \end{aligned} \tag{7.49}$$

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 | $n$ | $\widehat{\mu}$ $(\widehat{\sigma}_{\widehat{\mu}})$ | $\widehat{\sigma}$ $(\widehat{\sigma}_{\widehat{\sigma}})$ | $\widehat{\tau}^M$ $(\widehat{\sigma}_{\widehat{\tau}^M})$ |
|---|---|---|---|---|---|
| **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) |
| $\widehat{\tau}_0$ $(\widehat{\sigma}_{\widehat{\tau}_0})$ | | | | | 57.4 (0.49) |

A definite positive monotone trend in STATS over the levels of MATHS is evident from Table 7.2. The $\widehat{\mu}$ -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 $\widehat{\sigma}$ -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

$$\mu_i = \tau_0 + \tau_i^M \qquad i = 1, 2, \cdots 5 \tag{7.50}$$

where

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

In matrix notation (7.50) leads to

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

where

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

and $\boldsymbol{\lambda}$ denotes the vector of estimable parameters

$$\boldsymbol{\lambda} = \begin{pmatrix} \tau_0 \\ \tau_1^M \\ \tau_2^M \\ \tau_3^M \\ \tau_4^M \end{pmatrix} \tag{7.52}$$

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

From the restricted ML estimate $\widehat{\boldsymbol{\pi}}$, the ML estimate of $\boldsymbol{\lambda}$ is

$$\widehat{\boldsymbol{\lambda}} = (\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\widehat{\boldsymbol{\mu}} \tag{7.53}$$

with asymptotic covariance matrix

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\lambda}}\right) \approxeq \left\{(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\right\} \mathrm{Cov}\left(\widehat{\boldsymbol{\mu}}\right) \left\{((\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\right\}' \ . \tag{7.54}$$

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

$$\widehat{\boldsymbol{\tau}} = \mathbf{S}\widehat{\boldsymbol{\lambda}} \tag{7.55}$$

where

$$\widehat{\boldsymbol{\tau}} = \begin{pmatrix} \widehat{\tau}_0 \\ \widehat{\tau}_1^M \\ \widehat{\tau}_2^M \\ \widehat{\tau}_3^M \\ \widehat{\tau}_4^M \\ \widehat{\tau}_5^M \end{pmatrix} \tag{7.56}$$

and

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

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

$$\text{Cov}\left(\widehat{\boldsymbol{\tau}}\right) \approxeq \mathbf{S}\,\text{Cov}\left(\widehat{\boldsymbol{\lambda}}\right)\mathbf{S}^{'} \tag{7.58}$$

From the effects for the single factor model ($\widehat{\tau}$-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 $\widehat{\tau}_0 = 57.4$. The $\widehat{\tau}^M$-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

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

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

- $\mathbf{L} =$ `J(5,1,1) || DESIGNF(CUSUM(J(5,1,1)))`
- $\mathbf{S} =$ `BLOCK(1 , DESIGNF(CUSUM(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 $\mathbf{g}_{\mathsf{var}}(\boldsymbol{\pi}) = \mathbf{0}$, with

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

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

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

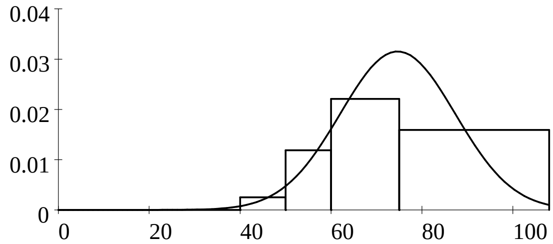
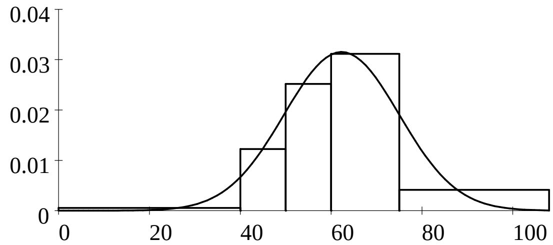
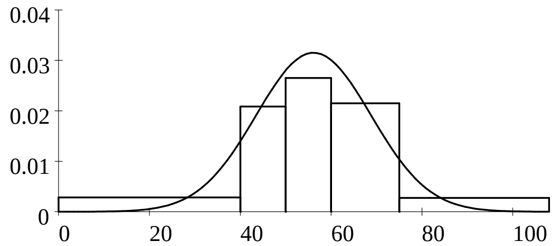
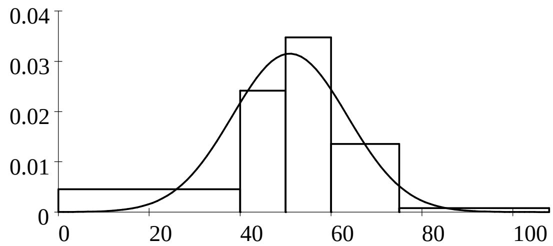
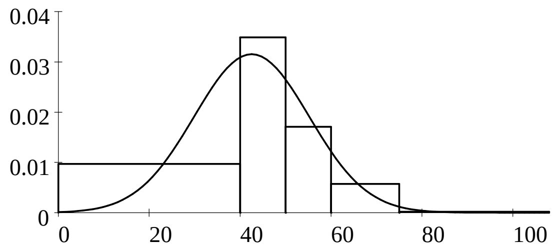
where $\mathbf{Q}_{\mathbf{H}} = \mathbf{I}_5 - \frac{1}{5}\mathbf{1}\mathbf{1}'$, is the projection matrix orthogonal to the vector of ones.

After employing the ML procedure with the vector of constraints

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

the restricted ML estimate $\widehat{\boldsymbol{\pi}}$ 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 | $n$ | $\widehat{\mu}$ $(\widehat{\sigma}_{\widehat{\mu}})$ | $\widehat{\sigma}$ $(\widehat{\sigma}_{\widehat{\sigma}})$ | $\widehat{\tau}^M$ $(\widehat{\sigma}_{\widehat{\tau}^M})$ |
|---|---|---|---|---|---|
| **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) |
| $\widehat{\tau}_0$ $(\widehat{\sigma}_{\widehat{\tau}_0})$ | | | | | 57.3 (0.47) |

No substantial changes with regard to the $\widehat{\mu}$-values were obtained from that of Model 1, with the $\widehat{\sigma}$-values now estimated constant with $\widehat{\sigma} = 12.7$. 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 $\widehat{\boldsymbol{\pi}}$ is estimated such that the vector $\widehat{\boldsymbol{\mu}}$ is a linear combination of $\mathbf{Y}_3$. (See Table 7.6.)

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

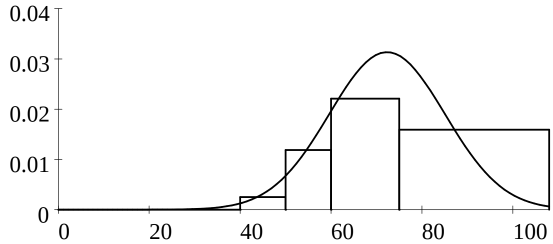
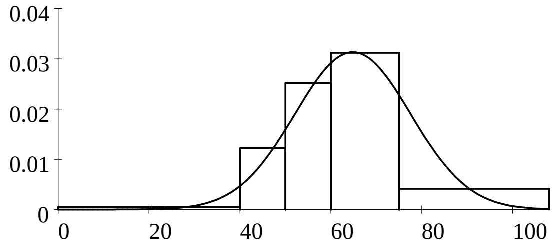
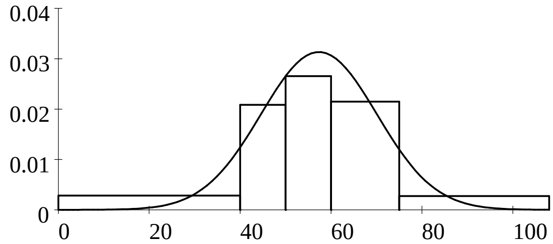
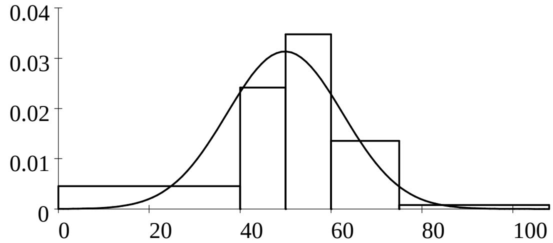
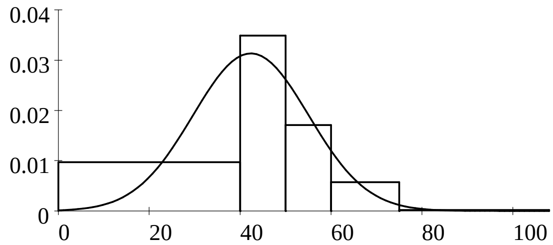
$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_3 &= (\mathbf{Y}_3'\mathbf{Y}_3)^{-1}\mathbf{Y}_3'\widehat{\boldsymbol{\mu}} \\ &= \begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \end{pmatrix} \\ &= \begin{pmatrix} 57.3 \\ 7.5 \end{pmatrix} \end{aligned}$$

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 $\widehat{\boldsymbol{\gamma}}_3$

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

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

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

<table>
<tr><th>MATHS</th><th>STATS</th><th>$n$</th><th>$\widehat{\mu}$<br>($\widehat{\sigma}_{\widehat{\mu}}$)</th><th>$\widehat{\sigma}$<br>($\widehat{\sigma}_{\widehat{\sigma}}$)</th><th>$\widehat{\tau}^M$<br>($\widehat{\sigma}_{\widehat{\tau}^M}$)</th></tr>
<tr><td>A</td><td>
</td><td>160</td><td>72.3<br>(0.87)</td><td>12.7<br>(0.40)</td><td>15.0<br>(0.70)</td></tr>
<tr><td>B</td><td>
</td><td>139</td><td>64.8<br>(0.60)</td><td>12.7<br>(0.40)</td><td>7.5<br>(0.35)</td></tr>
<tr><td>C</td><td>
</td><td>211</td><td>57.3<br>(0.46)</td><td>12.7<br>(0.40)</td><td>0.0<br>(0.00)</td></tr>
<tr><td>D</td><td>
</td><td>236</td><td>49.8<br>(0.55)</td><td>12.7<br>(0.40)</td><td>−7.5<br>(0.35)</td></tr>
<tr><td>E</td><td>
</td><td>152</td><td>42.4<br>(0.81)</td><td>12.7<br>(0.40)</td><td>−15.0<br>(0.70)</td></tr>
<tr><td>$\widehat{\tau}_0$<br>($\widehat{\sigma}_{\widehat{\tau}_0}$)</td><td></td><td></td><td></td><td></td><td>57.3<br>(0.46)</td></tr>
</table>

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

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

where

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

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

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

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

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

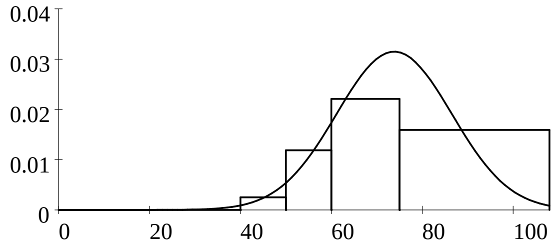
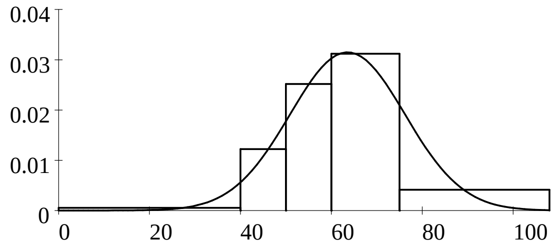
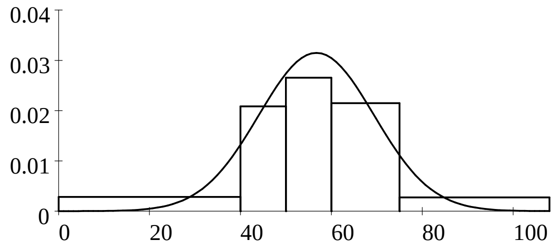
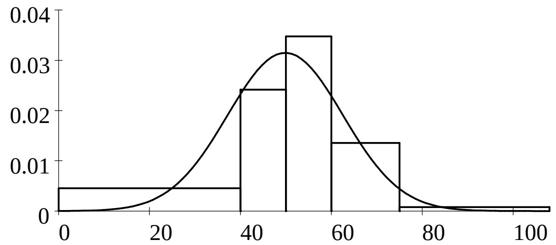
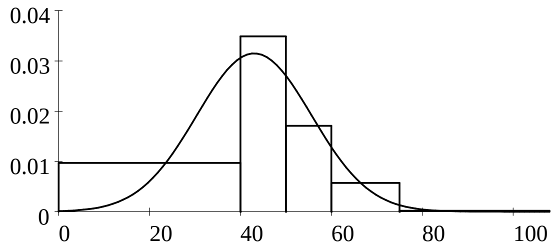
$$\begin{aligned} \widehat{\sigma}_{\widehat{\gamma}_4} &= \begin{pmatrix} \widehat{\sigma}_{\widehat{\gamma}_1} \\ \widehat{\sigma}_{\widehat{\gamma}_2} \end{pmatrix} \\ &= \begin{pmatrix} 2.108 \\ 0.0319 \end{pmatrix} \end{aligned}$$

this increase is significant, since

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

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

**Table 7.8:** ML estimates for regression model.

| **Maths** | **Stats** | $n$ | $\widehat{\mu}$ $(\widehat{\sigma}_{\widehat{\mu}})$ | $\widehat{\sigma}$ $(\widehat{\sigma}_{\widehat{\sigma}})$ | $\widehat{\tau}^M$ $(\widehat{\sigma}_{\widehat{\tau}^M})$ |
|---|---|---|---|---|---|
| **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) |
| $\widehat{\tau}_0$ $(\widehat{\sigma}_{\widehat{\tau}_0})$ | | | | | 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 $\nu$ is obtained from

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

leading to

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

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

$$\begin{aligned}
\ln\left(\frac{\boldsymbol{\pi}}{\mathbf{1}-\boldsymbol{\pi}}\right) &= \begin{pmatrix} \ln\left(\frac{\boldsymbol{\pi}_1}{\mathbf{1}-\boldsymbol{\pi}_1}\right) \\ \ln\left(\frac{\boldsymbol{\pi}_2}{\mathbf{1}-\boldsymbol{\pi}_2}\right) \\ \vdots \\ \ln\left(\frac{\boldsymbol{\pi}_T}{\mathbf{1}-\boldsymbol{\pi}_T}\right) \end{pmatrix} \\
&= \begin{pmatrix} \kappa_1 \ln \mathbf{x} + \theta_1 \mathbf{1} \\ \kappa_2 \ln \mathbf{x} + \theta_2 \mathbf{1} \\ \vdots \\ \kappa_T \ln \mathbf{x} + \theta_T \mathbf{1} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{X}\boldsymbol{\alpha}_1 \\ \mathbf{X}\boldsymbol{\alpha}_2 \\ \vdots \\ \mathbf{X}\boldsymbol{\alpha}_T \end{pmatrix} \\
&= (\mathbf{I}_T \otimes \mathbf{X})\,\boldsymbol{\alpha} \qquad (8.2)
\end{aligned}$$

where

$$\mathbf{X} = \begin{pmatrix} \ln \mathbf{x} & \mathbf{1} \end{pmatrix} \qquad (8.3)$$

is the design matrix for a log-logistic distribution and

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_T \end{pmatrix} \quad \text{where} \quad \boldsymbol{\alpha}_t = \begin{pmatrix} \kappa_t \\ \theta_t \end{pmatrix} \quad , \; t = 1 \cdots T \qquad (8.4)$$

is the concatenated vector of parameters.

The linear model (8.2) suggests the vector of constraints

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

where

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

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

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

where

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

Employing the maximum likelihood procedure with

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

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

$$\mathrm{Cov}(\widehat{\boldsymbol{\pi}}) \approxeq \mathbf{V} - (\mathbf{G}_\pi \mathbf{V})' (\mathbf{G}_\pi \mathbf{V} \mathbf{G}_\pi')^* (\mathbf{G}_\pi \mathbf{V}) \ . \qquad (8.9)$$

From the restricted ML estimator $\widehat{\boldsymbol{\pi}}$, it is possible to obtain the ML estimator of $\boldsymbol{\alpha}$

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

which consists out of two sets of estimators namely

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

and

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

The asymptotic covariance matrix of $\widehat{\boldsymbol{\alpha}}$ is

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) \approxeq \left\{\left(\mathbf{I}_T \otimes \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\mathbf{D}_\pi\right\} \mathrm{Cov}\left(\widehat{\boldsymbol{\pi}}\right) \left\{\left(\mathbf{I}_T \otimes \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\mathbf{D}_\pi\right\}' \tag{8.13}$$

with $\mathbf{D}_\pi$ given in (8.7).

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

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\kappa}}\right) \approxeq \left\{\left(\mathbf{I}_T \otimes \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]_1\right)\mathbf{D}_\pi\right\} \mathrm{Cov}\left(\widehat{\boldsymbol{\pi}}\right) \left\{\left(\mathbf{I}_T \otimes \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]_1\right)\mathbf{D}_\pi\right\}' \tag{8.14}$$

and

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\theta}}\right) \approxeq \left\{\left(\mathbf{I}_T \otimes \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]_2\right)\mathbf{D}_\pi\right\} \mathrm{Cov}\left(\widehat{\boldsymbol{\pi}}\right) \left\{\left(\mathbf{I}_T \otimes \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]_2\right)\mathbf{D}_\pi\right\}' . \tag{8.15}$$

## 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

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

Let

$$\boldsymbol{\nu} = \mathbf{Y}\gamma \tag{8.17}$$

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

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

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

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

follows, where

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

and

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

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

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

and

$$\mathbf{a}_\theta = \begin{pmatrix} -\frac{1}{\kappa_1} \exp\left(-\frac{\theta_t}{\kappa_t}\right) \\ -\frac{1}{\kappa_2} \exp\left(-\frac{\theta_t}{\kappa_t}\right) \\ \vdots \\ -\frac{1}{\kappa_T} \exp\left(-\frac{\theta_t}{\kappa_t}\right) \end{pmatrix} = -\frac{\mathbf{1}}{\boldsymbol{\kappa}} \exp\left(-\frac{\boldsymbol{\theta}}{\boldsymbol{\kappa}}\right) \; . \tag{8.23}$$

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

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

Employing the ML estimation procedure with

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

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

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

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

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

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

$$\begin{aligned} \text{Cov}\left(\widehat{\boldsymbol{\nu}}\right) &\approxeq \left\{\frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\alpha}}\right\} \text{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) \left\{\frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\alpha}}\right\}' \\ &= \mathbf{A}\,\text{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) \mathbf{A}'\,. \end{aligned} \tag{8.26}$$

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

$$\widehat{\gamma} = \left(\mathbf{Y}'\mathbf{Y}\right)^{-1} \mathbf{Y}'\widehat{\boldsymbol{\nu}} \tag{8.27}$$

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

$$\text{Cov}\left(\widehat{\gamma}\right) \approxeq \left\{\left(\mathbf{Y}'\mathbf{Y}\right)^{-1} \mathbf{Y}'\right\} \text{Cov}\left(\widehat{\boldsymbol{\nu}}\right) \left\{\left(\mathbf{Y}'\mathbf{Y}\right)^{-1} \mathbf{Y}'\right\}'\,. \tag{8.28}$$

## 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.

| | PRODUCT | | | |
|---|---|---|---|---|
| AGE | I | II | III | Total |
| 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

$$\ln\left(\frac{\boldsymbol{\pi}_t}{\mathbf{1}-\boldsymbol{\pi}_t}\right) \quad , \ t = 1, 2, \cdots, 12$$

is in the column space of

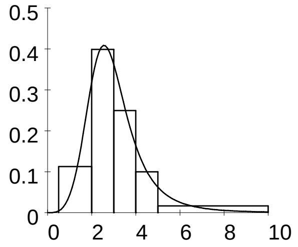
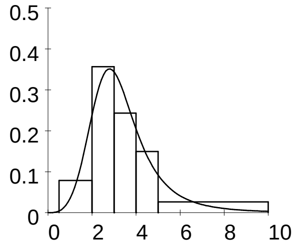
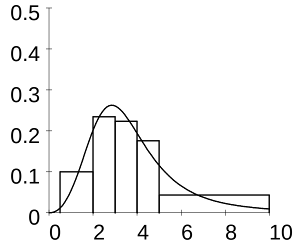
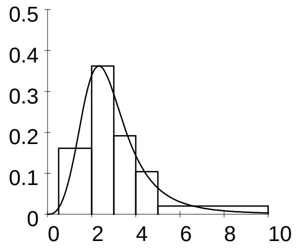
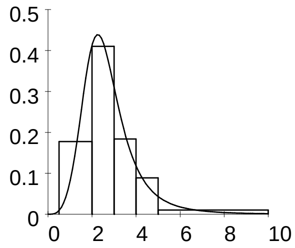
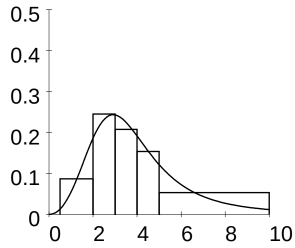
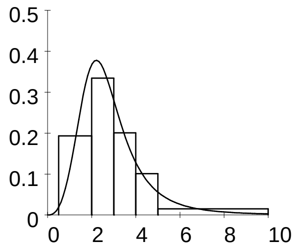
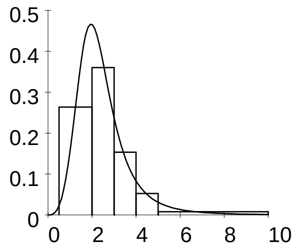
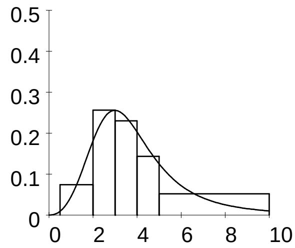
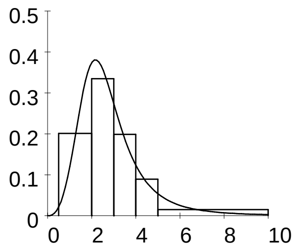
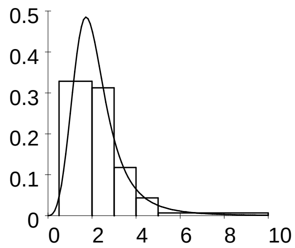
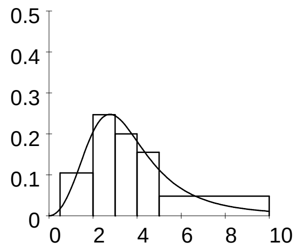
$$\mathbf{X} = \begin{pmatrix} \ln \mathbf{x} & \mathbf{1} \end{pmatrix} = \begin{pmatrix} \ln 2 & 1 \\ \ln 3 & 1 \\ \ln 4 & 1 \\ \ln 5 & 1 \end{pmatrix} . \quad (8.30)$$

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

$$\mathbf{g}_{\text{log}}(\boldsymbol{\pi}) = \begin{pmatrix} \mathbf{Q}_X \ln\left(\frac{\boldsymbol{\pi}_1}{\mathbf{1}-\boldsymbol{\pi}_1}\right) \\ \mathbf{Q}_X \ln\left(\frac{\boldsymbol{\pi}_2}{\mathbf{1}-\boldsymbol{\pi}_2}\right) \\ \vdots \\ \mathbf{Q}_X \ln\left(\frac{\boldsymbol{\pi}_{12}}{\mathbf{1}-\boldsymbol{\pi}_{12}}\right) \end{pmatrix} \quad (8.31)$$

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

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 $(\widehat{\mu})$ and medians $(\widehat{\nu})$ in Table 8.3 it is clear that Product III is the most expensive product. The standard deviations $(\widehat{\sigma})$ 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:

$\nu_{ij}^{AP}$ : median in $(ij)$-th cell

$\bar{\nu}_i^A = \frac{1}{3}\sum_{j=1}^{3} \nu_{ij}^{AP}$ : average median for $i$-th level of AGE

$\bar{\nu}_j^P = \frac{1}{4}\sum_{i=1}^{4} \nu_{ij}^{AP}$ : average median for $j$-th level of PRODUCT

$\bar{\nu} = \frac{1}{12}\sum_{i=1}^{4}\sum_{j=1}^{3} \nu_{ij}^{AP} = \frac{1}{4}\sum_{i=1}^{4} \bar{\nu}_i^A = \frac{1}{3}\sum_{j=1}^{3} \bar{\nu}_j^P$ : overall average median

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

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

where

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

Since

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

it follows for the main effects that

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

and for the interaction effects that

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

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

$$
\boldsymbol{\nu} = \mathbf{Z}\boldsymbol{\lambda}
$$

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

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

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

with corresponding vectors of ones

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

The saturated model (8.36) may therefore be partitioned as

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

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

**Table 8.5:** Partitioning of the saturated model.

| **Submatrices** | **Parameters** |
|---|---|
| $\mathbf{1} = \mathbf{1}_A \otimes \mathbf{1}_P : (12 \times 1)$ | $\tau_0$ : 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 $\boldsymbol{\lambda}$ is

$$\widehat{\boldsymbol{\lambda}}= \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\widehat{\boldsymbol{\nu}} \tag{8.40}$$

with asymptotic covariance matrix

$$\text{Cov}\left(\widehat{\boldsymbol{\lambda}}\right) \approxeq \left\{\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right\}\text{Cov}\left(\widehat{\boldsymbol{\nu}}\right)\left\{\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right\}' . \tag{8.41}$$

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

$$\widehat{\boldsymbol{\tau}} = \mathbf{S}\widehat{\boldsymbol{\lambda}} \tag{8.42}$$

where

$$\begin{aligned} \mathbf{S} &= \text{Block}\begin{pmatrix}1 & \mathbf{D}_A & \mathbf{D}_P & \mathbf{D}_A \otimes \mathbf{D}_P\end{pmatrix} \\ &= \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_A & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_A \otimes \mathbf{D}_P \end{pmatrix} \end{aligned} \tag{8.43}$$

and

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

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

$$\widehat{\boldsymbol{\tau}}^A = \begin{pmatrix} \widehat{\tau}_1^A \\ \widehat{\tau}_2^A \\ \widehat{\tau}_3^A \\ \widehat{\tau}_4^A \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\tau}}^B = \begin{pmatrix} \widehat{\tau}_1^P \\ \widehat{\tau}_2^P \\ \widehat{\tau}_3^P \end{pmatrix} \tag{8.45}$$

with the interaction effects included in

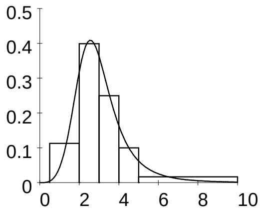
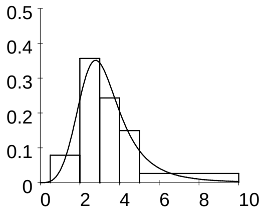
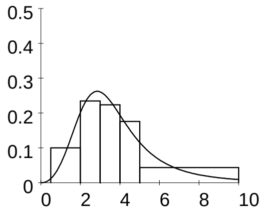
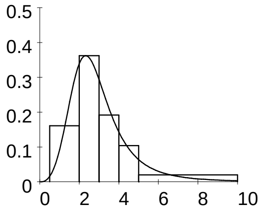
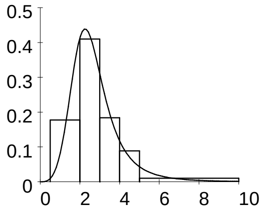
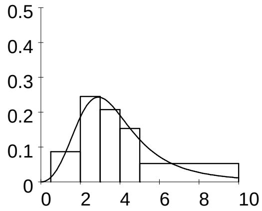
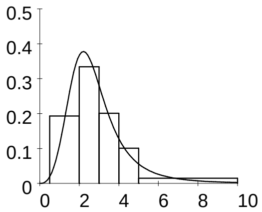
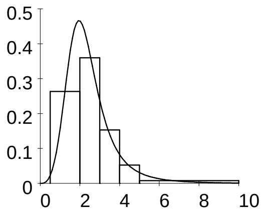
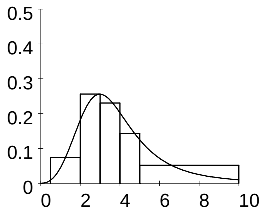
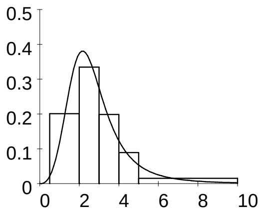
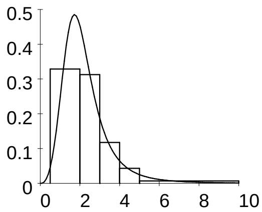
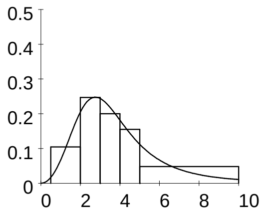
$$\widehat{\boldsymbol{\tau}}^{AP} = \begin{pmatrix} \widehat{\tau}_{11}^{AP} \\ \widehat{\tau}_{12}^{AP} \\ \widehat{\tau}_{13}^{AP} \\ \widehat{\tau}_{21}^{AP} \\ \widehat{\tau}_{22}^{AP} \\ \widehat{\tau}_{23}^{AP} \\ \widehat{\tau}_{31}^{AP} \\ \widehat{\tau}_{32}^{AP} \\ \widehat{\tau}_{33}^{AP} \\ \widehat{\tau}_{41}^{AP} \\ \widehat{\tau}_{42}^{AP} \\ \widehat{\tau}_{43}^{AP} \end{pmatrix} . \tag{8.46}$$

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

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\tau}}\right) \approxeq \mathbf{S}\,\mathrm{Cov}\left(\widehat{\boldsymbol{\lambda}}\right)\mathbf{S}' \ . \tag{8.47}$$

A complete summary of all the effects $(\widehat{\tau})$ with standard errors $(\widehat{\sigma}_{\widehat{\tau}})$ 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 $\widehat{\tau}^{AP}$-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 I | PRODUCT II | PRODUCT III | $\widehat{\tau}^A$ <br> $\widehat{\sigma}_{\widehat{\tau}^A}$ |
|---|---|---|---|---|
| **20-29** |  <br> $\widehat{\tau}^{AP} = -0.08$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.038$ |  <br> $\widehat{\tau}^{AP} = 0.41$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.045$ |  <br> $\widehat{\tau}^{AP} = -0.32$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.053$ | 0.25 <br> 0.032 |
| **30-39** |  <br> $\widehat{\tau}^{AP} = -0.01$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.034$ |  <br> $\widehat{\tau}^{AP} = 0.00$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.034$ |  <br> $\widehat{\tau}^{AP} = 0.02$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.044$ | 0.06 <br> 0.027 |
| **40-49** |  <br> $\widehat{\tau}^{AP} = 0.00$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.036$ |  <br> $\widehat{\tau}^{AP} = -0.16$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.036$ |  <br> $\widehat{\tau}^{AP} = 0.16$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.047$ | $-0.09$ <br> 0.028 |
| **50-59** |  <br> $\widehat{\tau}^{AP} = 0.10$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.044$ |  <br> $\widehat{\tau}^{AP} = -0.24$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.043$ |  <br> $\widehat{\tau}^{AP} = 0.15$ <br> $\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.055$ | $-0.22$ <br> 0.034 |
| $\widehat{\tau}^P$ <br> $\widehat{\sigma}_{\widehat{\tau}^P}$ | $-0.22$ <br> 0.022 | $-0.41$ <br> 0.023 | 0.63 <br> 0.029 | $\widehat{\tau}_0 = 2.89$ <br> $\widehat{\sigma}_{\widehat{\tau}_0} = 0.018$ |

### 8.3.2 Model 2: No interaction model

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

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

In matrix notation the medians are to be fitted such that

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

where

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

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

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

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

**Note:** The vector of constraints

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

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 fiited 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

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

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

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

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 $\mathbf{g}_{\text{mod2}}(\boldsymbol{\pi}) = \mathbf{0}$. See (8.50) and (8.51).

**Table 8.8:** Descriptive statistics for the no interaction model.

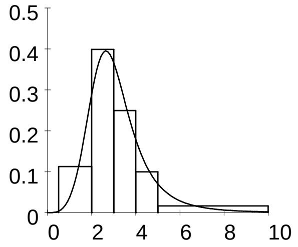
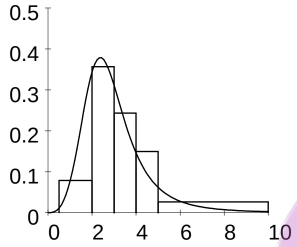
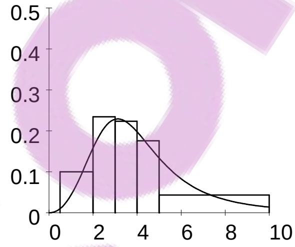
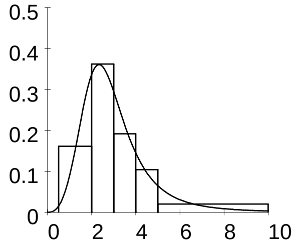
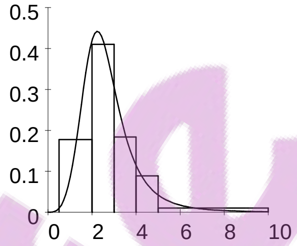
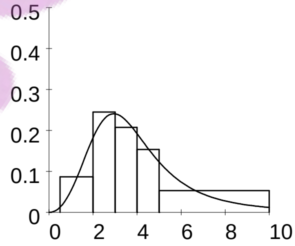
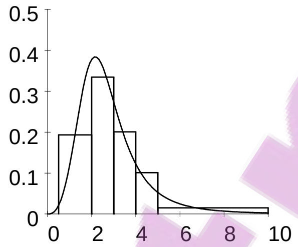
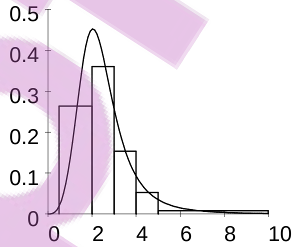
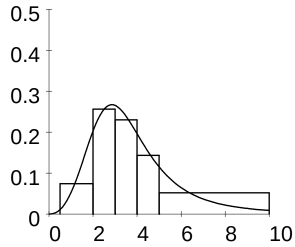
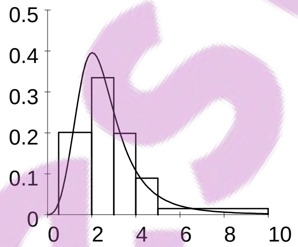
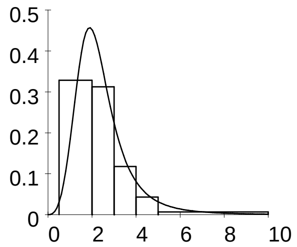
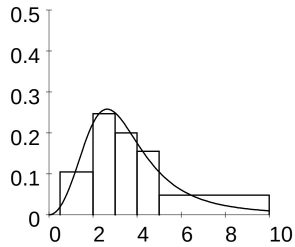
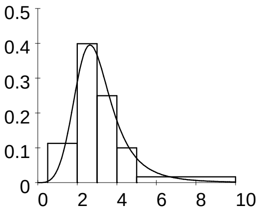
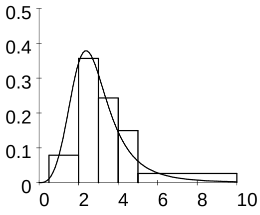
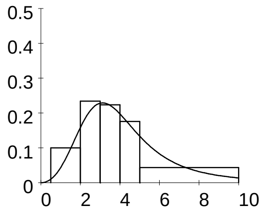
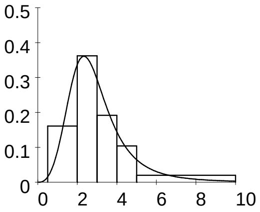
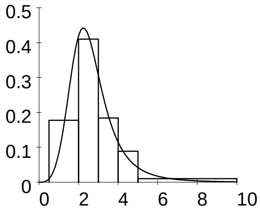
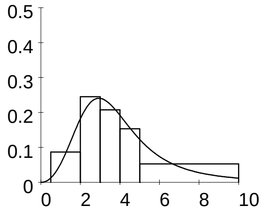
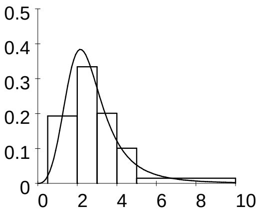
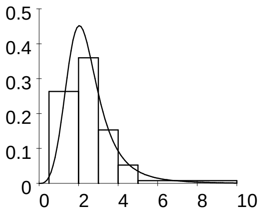
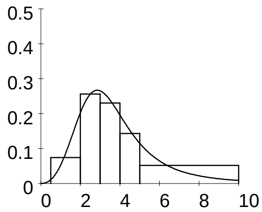
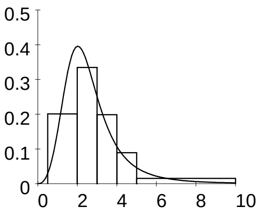
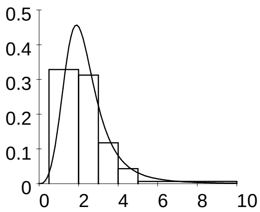
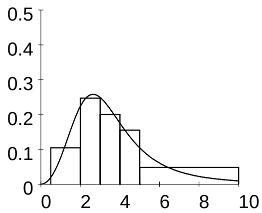
| **AGE** | **PRODUCT** | | |
|---|---|---|---|
| | **P1** | **P2** | **P3** |
| **20-29** | <br>$\widehat{\mu} = 3.199$ $\widehat{\nu} = 2.935$<br>$\widehat{\sigma} = 1.476$ $\widehat{\sigma}_{\widehat{\nu}} = 0.033$ | <br>$\widehat{\mu} = 3.053$ $\widehat{\nu} = 2.727$<br>$\widehat{\sigma} = 1.673$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

**Table 8.9:** Effects for the no interaction model.

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

### 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

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

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

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

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

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

Model (8.53) implies

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

to be implemented in the vector of constraints

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

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

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

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

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

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

respectively and hence the effect for Product III is

$$\widehat{\tau}_3^P = -\left(\widehat{\gamma}_3 + \widehat{\gamma}_4\right) = -\left(-0.22774 - 0.44003\right) = 0.66777 \tag{8.57}$$

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

The estimated two-factor model is

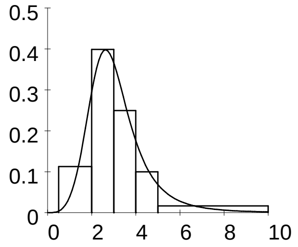
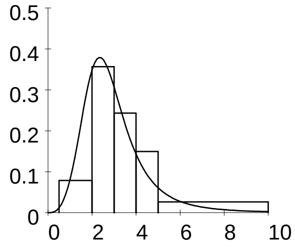
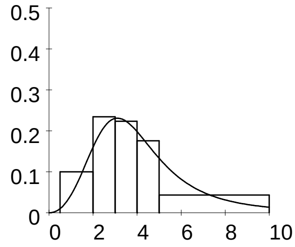
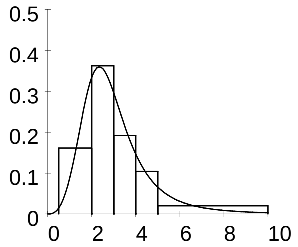
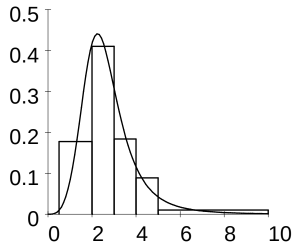
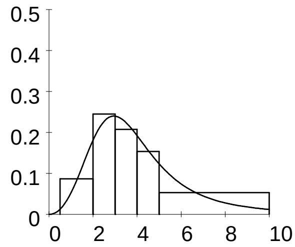
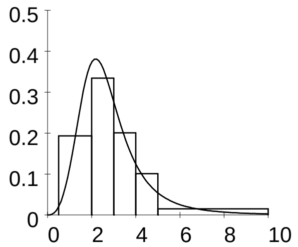
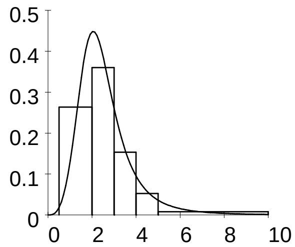
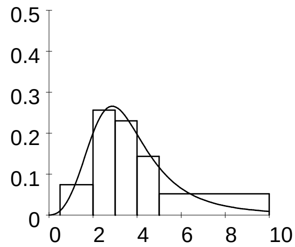
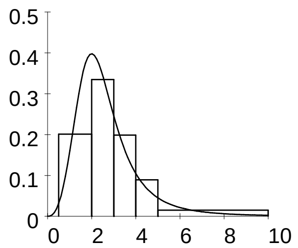
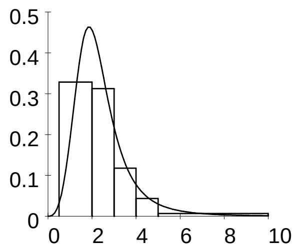
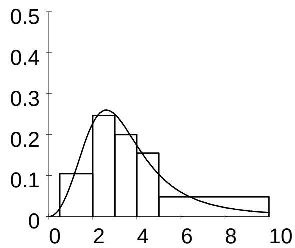
$$\widehat{\nu}_{ij} = \left(3.5897 + \widehat{\tau}_j^P\right) - 0.01817 z_i^A \quad , \ i = 1,2,3,4 \text{ and } j = 1,2,3 \tag{8.58}$$

where

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

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

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

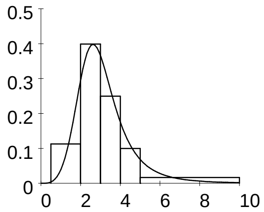
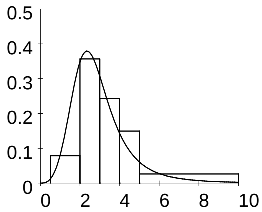
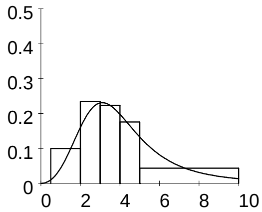
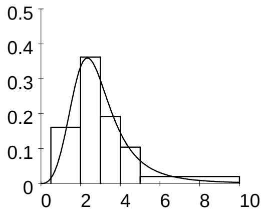
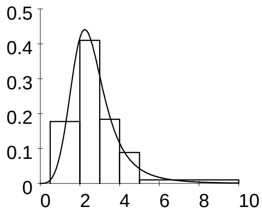
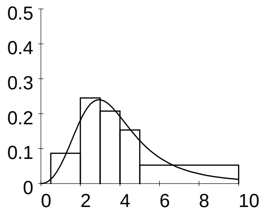
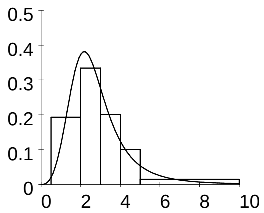
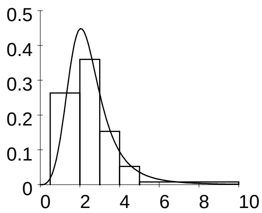
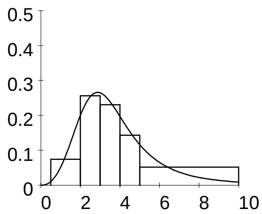
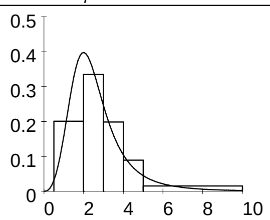
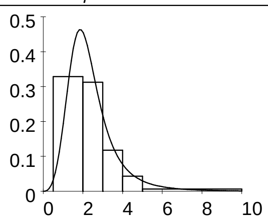
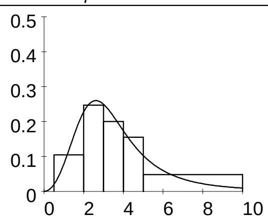
In Table 8.13 the effects for AGE reveal the same pattern as that of an ordinal variable. This follows since the distances between the class midpoints are equal. The effects of AGE show a constant drop of R18 per AGE category. Since all the interaction effects $\left(\widehat{\tau}^{AP}\right)$ are zero 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 I | PRODUCT II | PRODUCT III | $\hat{\tau}^{A}$ / $\hat{\sigma}_{\hat{\tau}^{A}}$ |
|---|---|---|---|---|
| **20-29** |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ | 0.273 0.022 |
| **30-39** |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ | 0.091 0.007 |
| **40-49** |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ | −0.091 0.007 |
| **50-59** |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ |  $\hat{\tau}^{AP} = 0$ $\hat{\sigma}_{\hat{\tau}^{AP}} = 0$ | −0.273 0.022 |
| $\hat{\tau}^{P}$ / $\hat{\sigma}_{\hat{\tau}^{P}}$ | −0.228 0.021 | −0.440 0.021 | 0.668 0.028 | 2.872 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

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

where $\widetilde{\mathbf{z}}^A$ 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 $\mathbf{Q}_4 = \mathbf{I} - \mathbf{Y}_4\left(\mathbf{Y}_4'\mathbf{Y}_4\right)^{-1}\mathbf{Y}_4'$ 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 $\widehat{\gamma}_4$ is

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

implying that the overall trend in PREMIUM over AGE is

$$\widehat{\nu}_{i\cdot} = \widehat{\gamma}_1 + \widehat{\gamma}_2 z_i^A = 3.4879 - 0.01532 z_i^A \ . \tag{8.62}$$

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

**PRODUCT I:**

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

**PRODUCT II:**

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

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

$$-\left(\widehat{\gamma}_3 + \widehat{\gamma}_4\right) = -\left(-0.39227 + 0.33708\right) = 0.05519$$

and the effect on the overall slope (8.62) is

$$-\left(\widehat{\gamma}_5 + \widehat{\gamma}_6\right) = -\left(0.00447 - 0.01963\right) = 0.01516$$

leading to the regression line

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

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 $\widehat{\tau}^A$-values is cancelled out by the interaction effects for PRODUCT III, the $\widehat{\tau}^{AP}$-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.

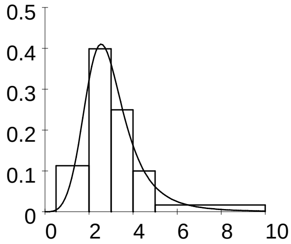
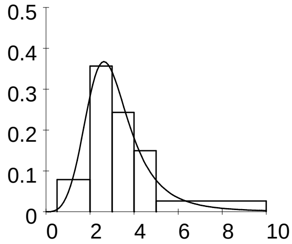
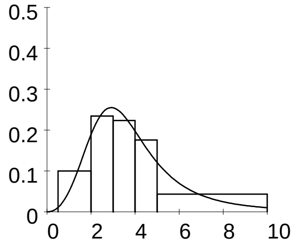
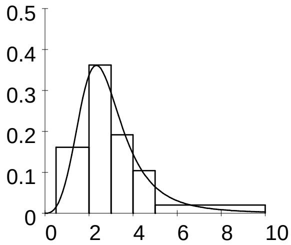
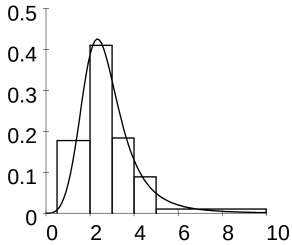
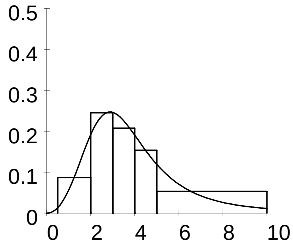
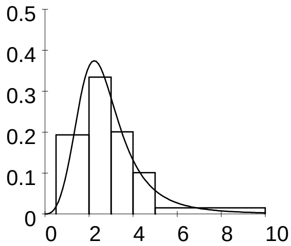
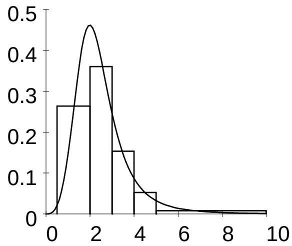
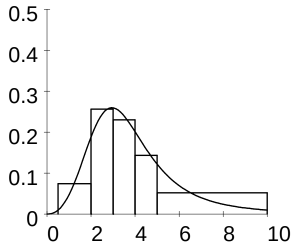
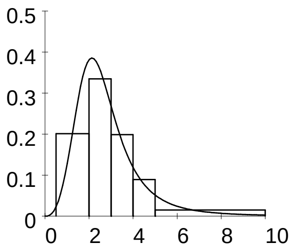
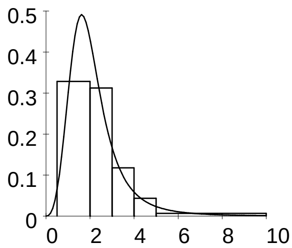
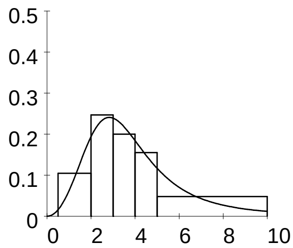
| AGE | PRODUCT | | |
|---|---|---|---|
| | **P1** | **P2** | **P3** |
| **20-29** | <br>$\widehat{\mu} = 3.084$ $\widehat{\nu} = 2.830$<br>$\widehat{\sigma} = 1.422$ $\widehat{\sigma}_{\widehat{\nu}} = 0.033$ | <br>$\widehat{\mu} = 3.279$ $\widehat{\nu} = 2.969$<br>$\widehat{\sigma} = 1.654$ $\widehat{\sigma}_{\widehat{\nu}}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

**Table 8.17:** Effects for regression model with interaction.

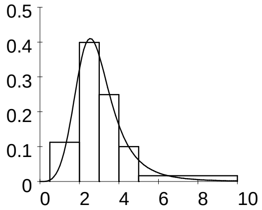
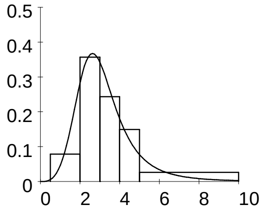
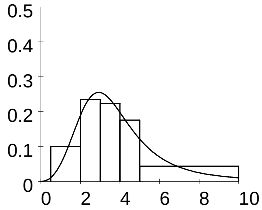
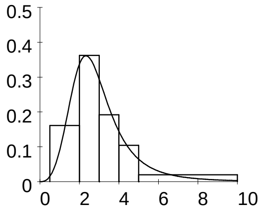
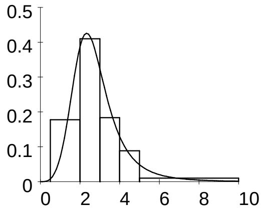
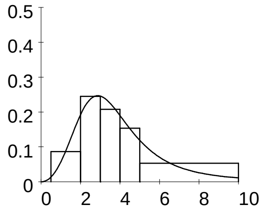
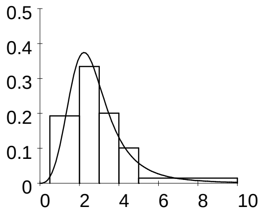
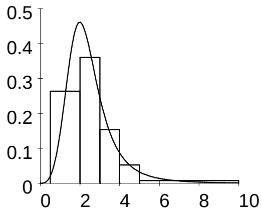
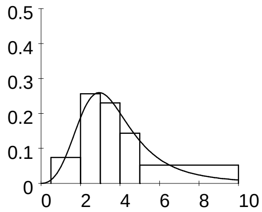
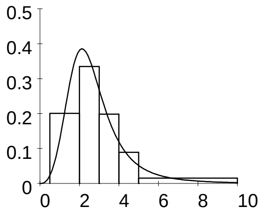
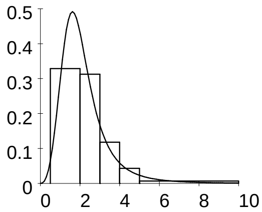
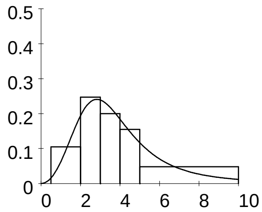
| AGE | PRODUCT I | PRODUCT II | PRODUCT III | $\widehat{\tau}^A$<br>$\widehat{\sigma}_{\widehat{\tau}^A}$ |
|---|---|---|---|---|
| **20-29** | <br>$\widehat{\tau}^{AP} = -0.067$<br>$\widehat{\sigma}_{\widehat{\tau}^{AP}} = 0.031$ | <br>$\widehat{\tau}^{AP} = 0.294$<br>$\widehat{\sigma}_{\widehat{\tau}^{AP}}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

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

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

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

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

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

with

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

the matrix of frequencies listed in Table 9.1.

Define

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

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

$$\mathbf{p}_0 = \frac{1}{n}\mathbf{f} \tag{9.4}$$

denote the vector of relative frequencies. Hence

$$\text{E}\,(\mathbf{p}_0) = \boldsymbol{\pi}_0 \tag{9.5}$$

and

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

Define the matrix of relative frequencies

$$\mathbf{P}_0 = \frac{1}{n}\mathbf{F} \tag{9.7}$$

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

$$\mathbf{P} = \mathbf{C}_I \cdot \mathbf{P}_0 \cdot \mathbf{C}_J \tag{9.8}$$

where

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

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

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

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

$$\mathbf{p} = \mathbf{C}\mathbf{p}_0 \tag{9.11}$$

with

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

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

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

and

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

## 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 $i = 1, 2, \cdots, I$ and $j = 1, 2, \cdots, J$ where $\pi_{ij}$ 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.

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

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

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

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

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

## 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 $\mu_x$ and standard deviation $\sigma_x$, i.e.

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

The marginal distribution of $y$ is normal with mean $\mu_y$ and standard deviation $\sigma_y$, i.e.

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

## 10.3 Standard bivariate normal distribution

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

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

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

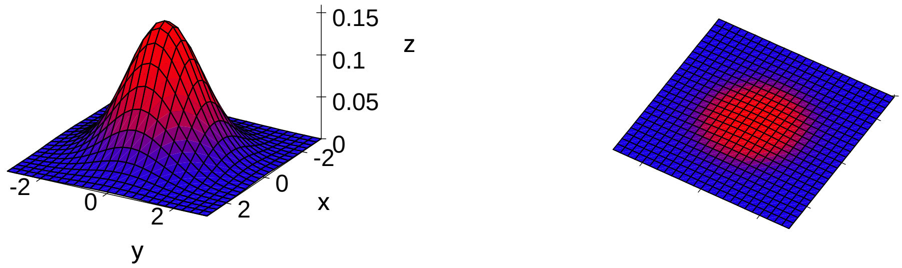
$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{10.5}$$

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

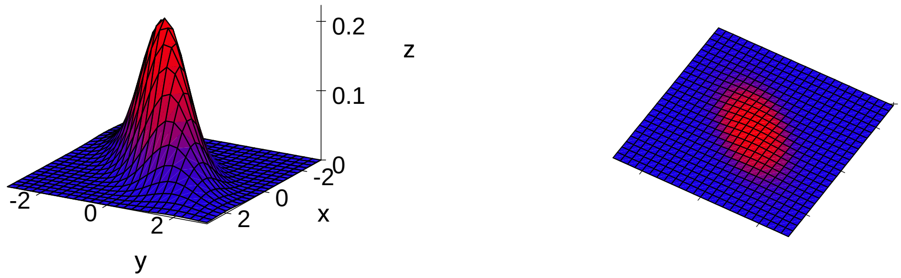
The standard bivariate normal curve is displayed in Table 10.1 to illustrate the effect of the correlation coefficient $\rho$.

**Table 10.1:** The effect of the correlation coefficient $\rho$.

**Case 1:** $\rho = 0$




**Case 2:** $\rho = 0.7$




**Case 3:** $\rho = -0.7$

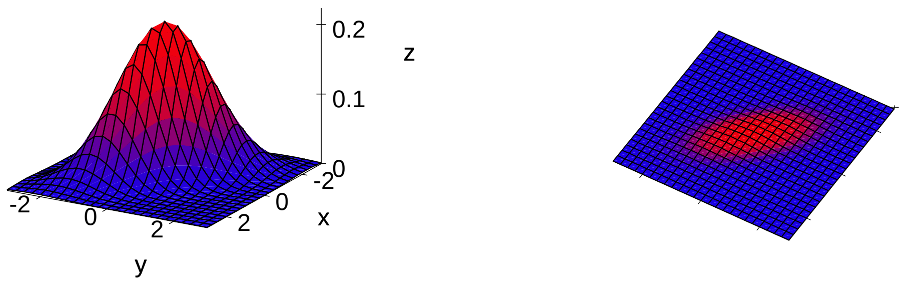


Table 10.1 is summarised as follows:

**Case 1:** $\rho = 0$

The contour curves are circles, indicating no relationship between $x$ and $y$. For all other values of $\rho$ 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

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

where

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

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

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

where

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

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

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

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** | |
|---|---|---|
| $Q_1$ | $-\infty < z_x < 0$ | $-\infty < z_y < 0$ |
| $Q_2$ | $-\infty < z_x < 0$ | $0 \leq z_y < \infty$ |
| $Q_3$ | $0 \leq z_x < \infty$ | $-\infty < z_y < 0$ |
| $Q_4$ | $0 \leq z_x < \infty$ | $0 \leq z_y < \infty$ |

Define

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

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

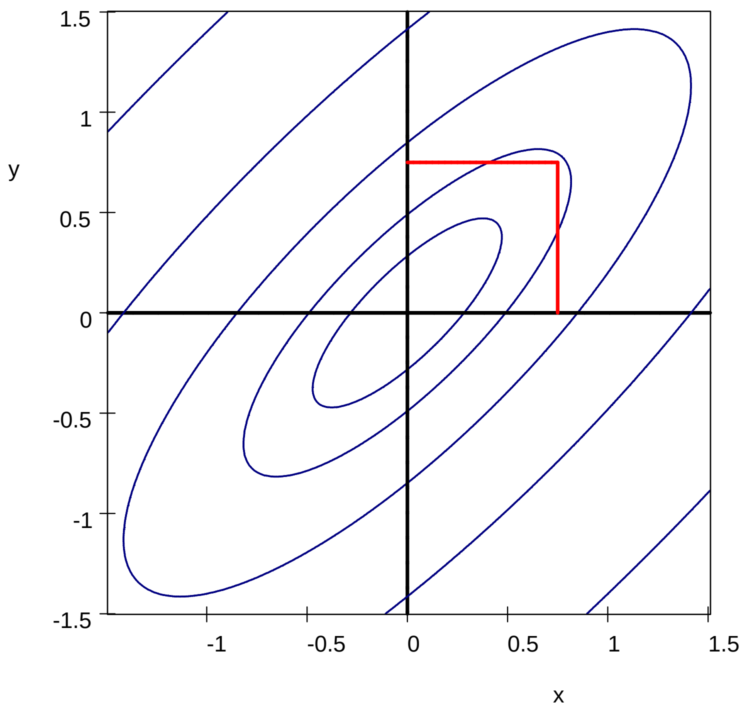
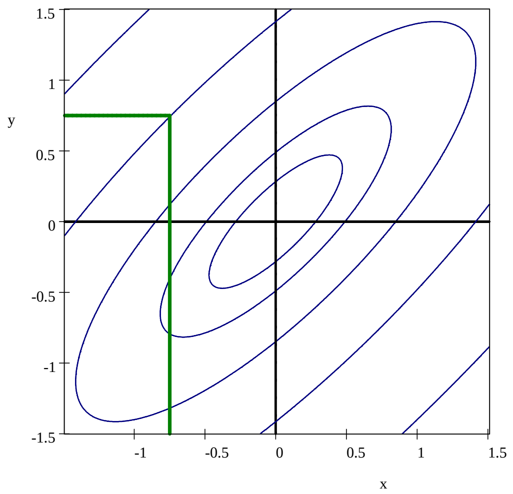



**Figure 10.1:** Integration region of $\Phi_0(a, b; \rho)$

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.

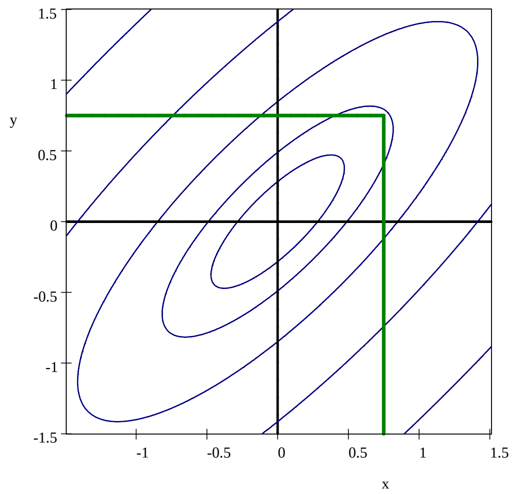
**Table 10.3:** Bivariate normal probabilities in the four quadrants.

**Quadrant 2:** $(a < 0 \text{ and } b \geq 0)$



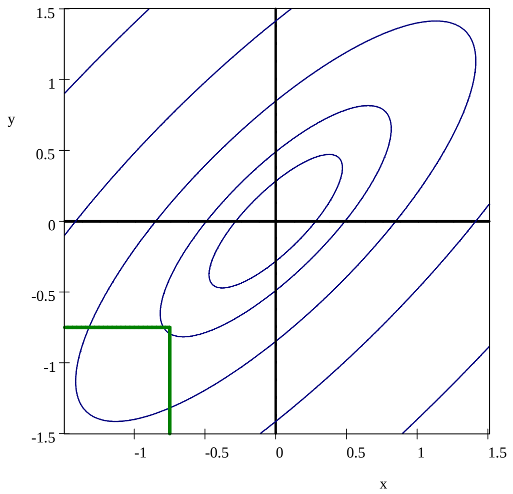

$$\begin{aligned}\Phi(a,b;\rho) &= \Phi_0(\infty,\infty;\rho) - \Phi_0(-a,\infty;\rho)\\ &+ \Phi_0(\infty,b;-\rho) - \Phi_0(-a,b;-\rho)\end{aligned}$$

**Quadrant 4:** $(a \geq 0 \text{ and } b \geq 0)$



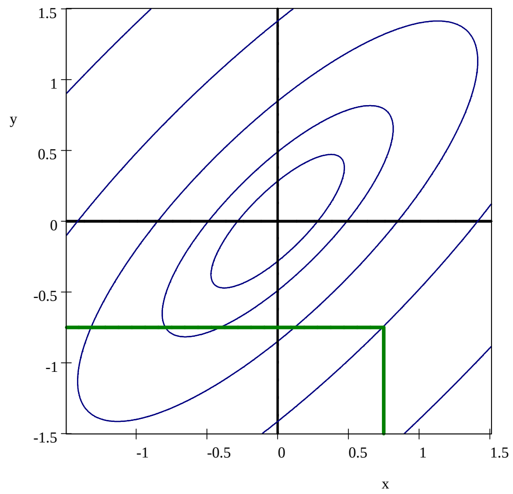

$$\begin{aligned}\Phi(a,b;\rho) &= \Phi_0(\infty,\infty;\rho) + \Phi_0(a,\infty;-\rho)\\ &+ \Phi_0(\infty,b;-\rho) + \Phi_0(a,b;\rho)\end{aligned}$$

**Quadrant 1:** $(a < 0 \text{ and } b < 0)$




$$\begin{aligned}\Phi(a,b;\rho) &= \Phi_0(\infty,\infty;\rho) - \Phi_0(-a,\infty;\rho)\\ &- \Phi_0(\infty,-b;\rho) + \Phi_0(-a,-b;\rho)\end{aligned}$$

**Quadrant 3:** $(a \geq 0 \text{ and } b < 0)$




$$\begin{aligned}\Phi(a,b;\rho) &= \Phi_0(\infty,\infty;\rho) + \Phi_0(a,\infty;-\rho)\\ &- \Phi_0(\infty,-b;\rho) - \Phi_0(a,-b;-\rho)\end{aligned}$$

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**

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

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

**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*

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

*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*

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

*and therefore*

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

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

■

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,$\rho$) or by making use of the SAS program *Phi.SAS* also listed in the Appendix.

### 10.5.2 Calculation of $\rho$

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

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

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

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

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

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 $\rho$

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

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 $\rho$ consider Table 10.4 and Table 10.5.

**Table 10.4:** Relationship between $\rho$ and the four volumes of the bivariate normal distribution.

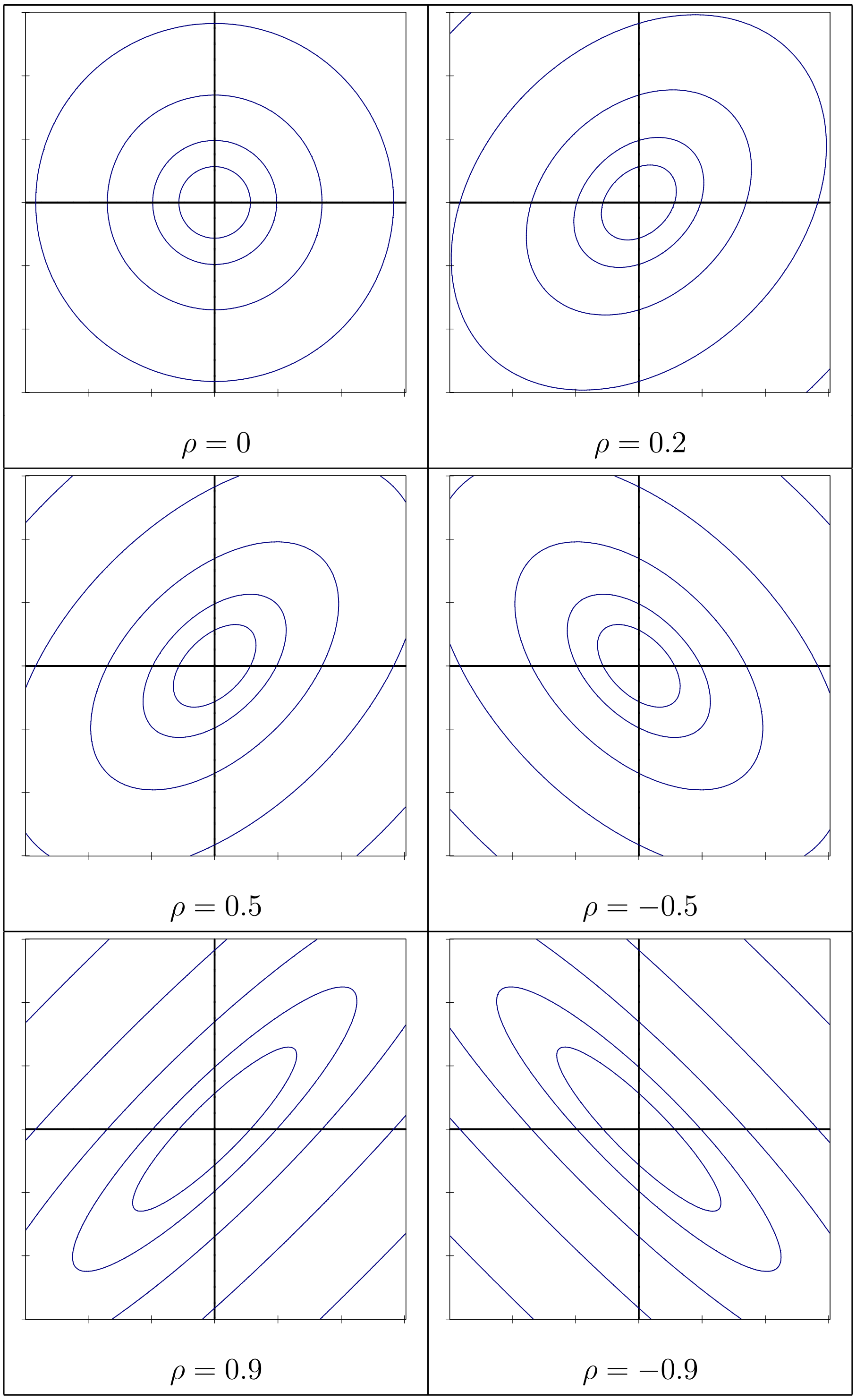
| $(\text{VOL}\,1 + \text{VOL}\,4)$ | $(\text{VOL}\,2 + \text{VOL}\,3)$ | $\rho$ | | |
|---|---|---|---|---|
| 1 | 0 | $\sin(\pi/2)$ | = | 1 |
| 0.9 | 0.1 | $\sin(\pi/2(0.8))$ | = | 0.951 06 |
| 0.8 | 0.2 | $\sin(\pi/2(0.6))$ | = | 0.809 02 |
| 0.7 | 0.3 | $\sin(\pi/2(0.4))$ | = | 0.587 79 |
| 0.6 | 0.4 | $\sin(\pi/2(0.2))$ | = | 0.309 02 |
| 0.5 | 0.5 | $\sin(\pi/2(0))$ | = | 0.0 |
| 0.4 | 0.6 | $\sin(\pi/2(-0.2))$ | = | −0.309 02 |
| 0.3 | 0.7 | $\sin(\pi/2(-0.4))$ | = | −0.587 79 |
| 0.2 | 0.8 | $\sin(\pi/2(-0.6))$ | = | −0.809 02 |
| 0.1 | 0.9 | $\sin(\pi/2(-0.8))$ | = | −0.951 06 |
| 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 $\rho$.



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

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

with

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

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

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

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

**Table 11.1:** Bivariate normal probabilities.

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

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.

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

where

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

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

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

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 $\mathbf{x}$:

$$\begin{aligned} \mathbf{\Phi}_x &= \boldsymbol{\pi}_x \\ \begin{pmatrix} \Phi_{1J} \\ \Phi_{2J} \\ \vdots \\ \Phi_{I-1,J} \end{pmatrix} &= \begin{pmatrix} \pi_{1,J} \\ \pi_{2,J} \\ \vdots \\ \pi_{I-1,J} \end{pmatrix} \end{aligned} \tag{11.9}$$

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

2. Marginal distribution of $\mathbf{y}$:

$$\begin{aligned} \mathbf{\Phi}_y &= \boldsymbol{\pi}_y \\ \begin{pmatrix} \Phi_{I1} & \Phi_{I2} & \cdots & \Phi_{I,J-1} \end{pmatrix}' &= \begin{pmatrix} \pi_{I1} & \pi_{I2} & \cdots & \pi_{I,J-1} \end{pmatrix}' \end{aligned} \tag{11.10}$$

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

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

$$\begin{aligned} \boldsymbol{\Phi}_{xy} &= \boldsymbol{\pi}_{xy} \\ \text{vec}\begin{pmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1,J-1} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2,J-1} \\ \vdots & \vdots & \cdots & \vdots \\ \Phi_{I-1,1} & \Phi_{I-1,2} & \cdots & \Phi_{I-1,J-1} \end{pmatrix} &= \text{vec}\begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1,J-1} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2,J-1} \\ \vdots & \vdots & \cdots & \vdots \\ \pi_{I-1,1} & \pi_{I-1,2} & \cdots & \pi_{I-1,J-1} \end{pmatrix} \end{aligned} \tag{11.11}$$

(First $(I-1)(J-1)$ elements of $\boldsymbol{\Phi}$ (11.7) and $\boldsymbol{\Pi}$ (11.8).)
In $\boldsymbol{\Phi}_{xy}$ 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.

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

where $-\infty < \mu_x,\ \mu_y < \infty$, $0 < \sigma_x,\ \sigma_y < \infty$ and $-1 < \rho < 1$. The parameters $\mu_x$ and $\sigma_x$ are functions of the marginal distribution of $x$, while the parameters $\mu_y$ and $\sigma_y$ are functions of the marginal distribution of $y$. The parameter $\rho$ 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

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

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

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

which leads to

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

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 $\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\pi}_x\right)$ on the vector space of $\mathbf{X}$ i.e.

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

where

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

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

$$\boldsymbol{\pi}_y = \begin{pmatrix} \pi_{I1} & \pi_{I2} & \cdots & \pi_{I,J-1} \end{pmatrix}' \tag{11.17}$$

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

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

which leads to

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

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

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

where

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

### 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

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

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

$$\boldsymbol{\pi}_0 = \mathbf{C}^{-1}\boldsymbol{\pi} \ . \tag{11.23}$$

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

The expressions for the 4 volumes are as follows:

$$\mathrm{VOL}\,1 = \mathbf{v}_1'\boldsymbol{\pi}_0 = \mathbf{v}_1'\mathbf{C}^{-1}\boldsymbol{\pi} \tag{11.24}$$
$$\mathrm{VOL}\,2 = \mathbf{v}_2'\boldsymbol{\pi}_0 = \mathbf{v}_2'\mathbf{C}^{-1}\boldsymbol{\pi} \tag{11.25}$$
$$\mathrm{VOL}\,3 = \mathbf{v}_3'\boldsymbol{\pi}_0 = \mathbf{v}_3'\mathbf{C}^{-1}\boldsymbol{\pi} \tag{11.26}$$
$$\mathrm{VOL}\,4 = \mathbf{v}_4'\boldsymbol{\pi}_0 = \mathbf{v}_4'\mathbf{C}^{-1}\boldsymbol{\pi} \tag{11.27}$$

where

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

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 $\rho$.

This implies that

$$\sum_{q=1}^{4} \mathbf{v}_q = 1 \tag{11.29}$$

and following from (11.22) it is now possible to express $\rho$ as

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

## 11.3 Vector of constraints

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

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

consists out of three sets of constraints.

### 11.3.1 Marginal distribution of $\mathbf{x}$

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

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

$$x \sim \mathrm{N}\left(\mu_x, \sigma_x^2\right) \ .$$

### 11.3.2 Marginal distribution of $\mathbf{y}$

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

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

$$y \sim \mathrm{N}\left(\mu_y, \sigma_y^2\right) \ .$$

### 11.3.3 Joint distribution of $\mathbf{x}$ and $\mathbf{y}$

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

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 $\mathbf{x}$ and $\mathbf{y}$. The bivariate normal distribution to be fitted is such that

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

The elements of

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

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 $\boldsymbol{\pi}$

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

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

### 11.4.1 Marginal distribution of x

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

where

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

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

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

where

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

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

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

## 11.4.2 Marginal distribution of $\mathbf{y}$

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

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

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

where

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

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

From the chain rule for matrix differentiation it follows that

$$\begin{aligned}
\frac{\partial \mathbf{g}_{xy}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} &= \frac{\partial \boldsymbol{\Phi}_{xy}}{\partial \boldsymbol{\pi}} - \frac{\partial \boldsymbol{\pi}_{xy}}{\partial \boldsymbol{\pi}} \\
&= \frac{\partial \boldsymbol{\Phi}_{xy}}{\partial \begin{pmatrix} \mathbf{z}_x \\ \mathbf{z}_y \\ \rho \end{pmatrix}} \cdot \frac{\partial \begin{pmatrix} \mathbf{z}_x \\ \mathbf{z}_y \\ \rho \end{pmatrix}}{\partial \boldsymbol{\pi}} - \frac{\partial \boldsymbol{\pi}_{xy}}{\partial \boldsymbol{\pi}} \\
&= \begin{pmatrix} \underset{(1)}{\frac{\partial \boldsymbol{\Phi}_{xy}}{\partial \mathbf{z}_x}} & \underset{(2)}{\frac{\partial \boldsymbol{\Phi}_{xy}}{\partial \mathbf{z}_y}} & \underset{(3)}{\frac{\partial \boldsymbol{\Phi}_{xy}}{\partial \rho}} \end{pmatrix} \cdot \begin{pmatrix} \underset{(4)}{\frac{\partial \mathbf{z}_x}{\partial \boldsymbol{\pi}}} \\ \underset{(5)}{\frac{\partial \mathbf{z}_y}{\partial \boldsymbol{\pi}}} \\ \underset{(6)}{\frac{\partial \rho}{\partial \boldsymbol{\pi}}} \end{pmatrix} - \mathbf{I}_{xy} \qquad (11.45)
\end{aligned}$$

where

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

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

1.

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

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

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

It now follows that

$$\begin{aligned}
\frac{\partial \mathbf{\Phi}_{xy}}{\partial \mathbf{z}_x} &= \left(\frac{\partial \mathbf{\Phi}_{xy}}{\partial z_{x_1}}, \frac{\partial \mathbf{\Phi}_{xy}}{\partial z_{x_2}}, \cdots, \frac{\partial \mathbf{\Phi}_{xy}}{\partial z_{x_{I-1}}}\right) \\
&= \left(\operatorname{vec}\left(\mathbf{E}_1 \mathbf{\Delta}_x\right), \operatorname{vec}\left(\mathbf{E}_2 \mathbf{\Delta}_x\right), \cdots, \operatorname{vec}\left(\mathbf{E}_{I-1} \mathbf{\Delta}_x\right)\right) \qquad (11.47)
\end{aligned}$$

where

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

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

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

2. Likewise

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

and therefore it follows that

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

where

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

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

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

3.

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

Simplification of the derivative above leads to

$$\begin{aligned}
&\frac{\partial}{\partial\rho}\left(-\frac{1}{2(1-\rho^2)}\left[z_1^2-2\rho z_1z_2+z_2^2\right]\right)\\
&=\frac{z_1z_2}{(1-\rho^2)}-\frac{\rho}{(1-\rho^2)^2}\left[z_1^2-2\rho z_1z_2+z_2^2\right]\\
&=-\frac{\rho}{(1-\rho^2)^2}z_1^2+\frac{(1-\rho^2)+2\rho^2}{(1-\rho^2)^2}z_1z_2-\frac{\rho}{(1-\rho^2)^2}z_2^2
\end{aligned}$$

and therefore

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

where

$$\widetilde{\Psi}\left(z_{x_i},z_{y_j},k,l;\rho\right)=\int_{-\infty}^{z_{y_j}}\int_{-\infty}^{z_{x_i}}\frac{1}{2\pi\sqrt{1-\rho^2}}\underbrace{z_1^kz_2^l}\exp\left(-\frac{1}{2(1-\rho^2)}\left[z_1^2-2\rho z_1z_2+z_2^2\right]\right)dz_1dz_2 \quad (11.54)$$

Define the integral

$$\widetilde{\Psi}_0\left(z_{x_i},z_{y_j},k,l;\rho\right)=\int_0^{z_{y_j}}\int_0^{z_{x_i}}\frac{1}{2\pi\sqrt{1-\rho^2}}\underbrace{z_1^kz_2^l}\exp\left(-\frac{1}{2(1-\rho^2)}\left[z_1^2-2\rho z_1z_2+z_2^2\right]\right)dz_1dz_2 \quad (11.55)$$

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

**Quadrant 1:** $\left(z_{x_i}<0\ ,\ z_{y_j}<0\right)$

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

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

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

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

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

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

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

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

**Algorithm 2**

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

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

**Proof.** *Since*

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

*it follows that*

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

*from (10.12) it follows that*

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

■

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

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

See (11.38).

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

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

See (11.43).

6. From (11.30) it follows that

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

## 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 $\boldsymbol{\pi}$

$\boldsymbol{\pi} = \mathbf{p}$

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

Calculate $\mathbf{z}_{x\_\pi}$, $\mathbf{z}_{y\_\pi}$ and $\rho_{\pi}$ from $\boldsymbol{\pi}$.

Calculate $\mathbf{G}_{\pi}$ (as a function of $\boldsymbol{\pi}$)

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

DO OVER $\mathbf{p}$

Calculate $\mathbf{z}_{x\_p}$, $\mathbf{z}_{y\_p}$ and $\rho_p$ from $\mathbf{p}$.

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

$$\mathbf{g}(\mathbf{p}) = \begin{pmatrix} \boldsymbol{\Phi}\left(\mathbf{z}_{x\_p}\right) \\ \boldsymbol{\Phi}\left(\mathbf{z}_{y\_p}\right) \\ \text{vec}\left(\boldsymbol{\Phi}\left(\mathbf{z}_{x\_p}, \mathbf{z}_{y\_p}, \rho_p\right)\right) \end{pmatrix} - \begin{pmatrix} \mathbf{p}_x \\ \mathbf{p}_y \\ \mathbf{p}_{xy} \end{pmatrix}$$

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

END

END

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

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

where the vectors of standardised upper class boundaries are calculated from

$$\mathbf{z}_{x\_p} = \mathbf{P}_X\boldsymbol{\Phi}^{-1}\left(\mathbf{p}_x\right) \quad \text{and} \quad \mathbf{z}_{y\_p} = \mathbf{P}_Y\boldsymbol{\Phi}^{-1}\left(\mathbf{p}_y\right) \tag{11.64}$$

projecting $\mathbf{\Phi}^{-1}(\mathbf{p}_x)$ and $\mathbf{\Phi}^{-1}(\mathbf{p}_y)$ 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

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

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 $\rho_p$. The calculation of $\rho_p$ will therefore be done iteratively, starting at a value say, $\rho = 0$, untill iteration over (11.65) leads to a unique estimate for $\rho_p$. (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 $\boldsymbol{\pi}$.

For convergence over $\boldsymbol{\pi}$ the covariance matrix $\mathbf{V}$ and the matrix of partial derivatives $\mathbf{G}_\pi$ are all functions of $\boldsymbol{\pi}$. Convergence over $\boldsymbol{\pi}$ leads to the restricted ML estimate of $\boldsymbol{\pi}$, i.e. $\widehat{\boldsymbol{\pi}}$ , 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 $\widehat{\boldsymbol{\pi}}$, discussed in the previous section. In matrix notation $\widehat{\boldsymbol{\pi}}$ can be represented as

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

where $\widehat{\pi}_{ij}$ 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 $\widehat{\boldsymbol{\pi}}$ is

$$\text{Cov}\left(\widehat{\boldsymbol{\pi}}\right) \approxeq \mathbf{V} - \left(\mathbf{G}_\pi \mathbf{V}\right)' \left(\mathbf{G}_\pi \mathbf{V} \mathbf{G}_\pi'\right)^* \left(\mathbf{G}_\pi \mathbf{V}\right) \ .$$

### 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 $\widehat{\boldsymbol{\pi}}$ with

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

and

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

where

$$\widehat{\boldsymbol{\pi}}_x = \begin{pmatrix} \widehat{\pi}_{1J} \\ \widehat{\pi}_{2J} \\ \vdots \\ \widehat{\pi}_{I-1,J} \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\pi}}_y = \begin{pmatrix} \widehat{\pi}_{I1} & \widehat{\pi}_{I2} & \cdots & \widehat{\pi}_{I,J-1} \end{pmatrix}' . \tag{11.69}$$

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

The corresponding covariance matrices are

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

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

where

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

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 $\mu_x,\ \mu_y,\ \sigma_x,\ \sigma_y$ and $\rho$ with their standard errors are all functions of the restricted ML estimate $\widehat{\boldsymbol{\pi}}$ and will be discussed briefly. The ML estimates for the $\mu$'s and $\sigma$'s follow from (11.67) and (11.68) and according to the multivariate delta theorem

$$\widehat{\boldsymbol{\beta}}_x = \begin{pmatrix} \widehat{\mu}_x \\ \widehat{\sigma}_x \end{pmatrix} \approxeq \mathrm{N}\left( \begin{pmatrix} \mu_x \\ \sigma_x \end{pmatrix}, \mathbf{B}_x \,\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_x\right) \mathbf{B}_x' \right) \tag{11.72}$$

and

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

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

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

The only parameter that remains is $\rho$ and is estimated from

$$\begin{aligned} \widehat{\rho} &= \sin\left(\frac{\pi}{2}\left[\left(\widehat{\mathrm{VOL1}} + \widehat{\mathrm{VOL4}}\right) - \left(\widehat{\mathrm{VOL2}} + \widehat{\mathrm{VOL3}}\right)\right]\right) \\ &= \sin\left(\frac{\pi}{2}\left[(\mathbf{v}_1' + \mathbf{v}_4') - (\mathbf{v}_2' + \mathbf{v}_3')\right]\mathbf{C}^{-1}\widehat{\boldsymbol{\pi}}\right) \quad . \end{aligned} \tag{11.74}$$

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 $\widehat{\rho}$ is to be solved iteratively over (11.74) beginning at any starting point, say $\widehat{\rho} = 0$ untill convergence leads to the unique ML estimate for $\rho$. The variance of $\widehat{\rho}$ follows

$$\mathrm{Var}\left(\widehat{\rho}\right) = \left(\frac{\partial \rho}{\partial \mathbf{p}}\right) \mathbf{V} \left(\frac{\partial \rho}{\partial \mathbf{p}}\right)' \tag{11.75}$$

where

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

## 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 $\chi^2$ statistic

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

and the Wald statistic

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

is

$$\text{df} = IJ - I - J + 2\ . \tag{11.77}$$

In (11.76) $p_{ij}$ for $i = 1, 2, \cdots, I$ and $j = 1, 2, \cdots, J$ 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

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

The elements of $\mathbf{P}$ are also referred to as the unrestricted ML estimates of $\boldsymbol{\pi}$. The elements of $\widehat{\boldsymbol{\Pi}}$ 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** $(y)$ **0-49** | **50-59** | **60-74** | **75-100** | **Total** |
|---|---|---|---|---|---|
| **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 $\chi^2$ test of independence, $\chi^2 = 326$ $(\mathrm{df} = 9,$ $p$ 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

$$r = 0.5495 \tag{12.1}$$

with an estimated regression line of

$$\widehat{y} = 25.8 + 0.5187x \ . \tag{12.2}$$

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

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

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 $\widehat{\boldsymbol{\pi}}$ will be given. The estimates obtained from $\widehat{\boldsymbol{\pi}}$ are the ML estimates for the bivariate normal distribution. This follows since the vector $\widehat{\boldsymbol{\pi}}$ 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

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

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

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

#### 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 $\mathbf{x}$.

| $\mathbf{p}_x$ | $\widehat{\boldsymbol{\alpha}}_x$ | $\widehat{\mu}_x$ | $\widehat{\sigma}_x$ | $\widehat{\mathbf{z}}_x$ |
|---|---|---|---|---|
| $\begin{pmatrix} 0.31635 \\ 0.59920 \\ 0.78552 \end{pmatrix}$ | $\begin{pmatrix} 0.06345 \\ 4.22132 \end{pmatrix}$ | 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 $\mathbf{\Phi}^{-1}(\mathbf{p}_x)$ into the vector space of $\mathbf{X}$

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

and is employed in the vector of constraints

$$\mathbf{g}_x(\pi) = \mathbf{\Phi}\left(\widehat{\mathbf{z}}_x\right) - \boldsymbol{\pi}_x = \mathbf{0} \; . \tag{12.9}$$

**In SAS IML:** $\boxed{\mathbf{\Phi}\left(z_x\right) = \text{PROBNORM}(z_x)}$

The unrestricted estimate for the vector of natural parameters

$$\widehat{\boldsymbol{\alpha}}_x = \begin{pmatrix} \dfrac{1}{\widehat{\sigma}_x} \\ \dfrac{\widehat{\mu}_x}{\widehat{\sigma}_x} \end{pmatrix} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Phi}^{-1}\left(\mathbf{p}_x\right) \tag{12.10}$$

leads to the unrestricted estimates for $\widehat{\mu}_x$ and $\widehat{\sigma}_x$ 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 $\mathbf{y}$.

| $\mathbf{p}_y$ | $\widehat{\boldsymbol{\alpha}}_y$ | $\widehat{\mu}_y$ | $\widehat{\sigma}_y$ | $\widehat{\mathbf{z}}_y$ |
|---|---|---|---|---|
| $\begin{pmatrix} 0.24129 \\ 0.53217 \\ 0.79491 \end{pmatrix}$ | $\begin{pmatrix} 0.06021 \\ 3.61002 \end{pmatrix}$ | 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

$$\widehat{\mathbf{z}}_y = \mathbf{P}_y\mathbf{\Phi}^{-1}\left(\widehat{\boldsymbol{\pi}}_y\right) \tag{12.11}$$

is employed in the vector of constraints

$$\mathbf{g}_y(\pi) = \mathbf{\Phi}\left(\mathbf{z}_y\right) - \boldsymbol{\pi}_y = \mathbf{0} \; . \tag{12.12}$$

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

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 $(\widehat{z}_x, \widehat{z}_y) = (0, 0)$ 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.

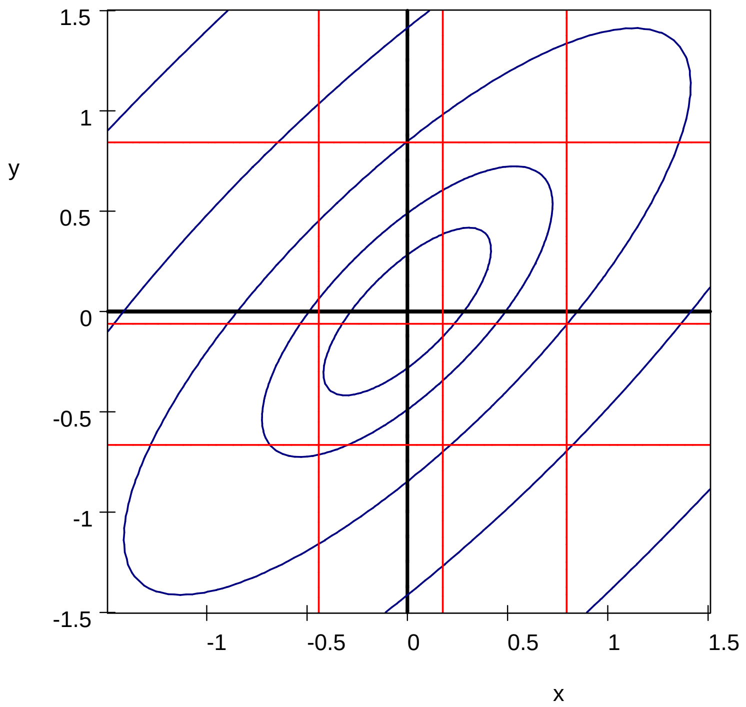



**Figure 12.1:** Contour diagram of the bivariate normal distribution with the four quadrants and the standardised upper class boundaries.

The ML estimator for $\rho$ is obtained from

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

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

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

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 $\widehat{\rho}$. Since $\widehat{\rho}$ is to be estimated, the value of $\widehat{\rho}$ is obtained iteratively over (12.13), starting at any value between -1 and 1. In Table 12.4 various starting points for $\widehat{\rho}$ were being used, all leading to the same unique unrestricted estimate for $\rho$. (Convergence criterion = 1e-10.)

**Table 12.4:** Unrestricted estimate for $\rho$ obtained iteratively

| Starting point $\widehat{\rho} = -0.5$ | | Starting point $\widehat{\rho} = 0$ | | Starting point $\widehat{\rho} = 0.5$ | |
|---|---|---|---|---|---|
| 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 $\rho$ is now estimated from the observed frequencies (unrestricted ML estimate for $\boldsymbol{\pi}$).

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

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

From Table 12.5 it follows that

$$\begin{aligned}\widehat{\rho} &= \sin\left(\frac{\pi}{2}\left[(0.3979789 + 0.3392913) - (0.1177174 + 0.1450123)\right]\right)\\ &= \sin\left(\frac{\pi}{2}\left[0.737\,27 - 0.262\,73\right]\right)\\ &= \sin\left(\frac{\pi}{2}\left[0.474\,54\right]\right)\\ &= 0.678\,27 \end{aligned} \tag{12.15}$$

indicating a positive relationship between MATHS and STATS.

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

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

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

### 12.1.2 ML estimates

After convergence of the ML estimation procedure the restricted ML estimate for $\boldsymbol{\pi}$ in matrix notation is

$$\widehat{\mathbf{\Pi}} = \begin{pmatrix} 0.17170 & 0.25637 & 0.31180 & 0.31869 \\ 0.22874 & 0.38802 & 0.53638 & 0.56766 \\ 0.25077 & 0.45896 & 0.70922 & 0.79166 \\ 0.25569 & 0.48298 & 0.81013 & 1.00000 \end{pmatrix} \tag{12.16}$$

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

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

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 | $p$-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

$$\mathbf{D} = \frac{\mathbf{W}}{n} = \frac{44.994}{746} = 0.06 \tag{12.18}$$

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 $\mathbf{x}$.

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

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

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

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

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

$$\widehat{\sigma}_{\widehat{\mu}_x} = 0.62047 \quad \text{and} \quad \widehat{\sigma}_{\widehat{\sigma}_x} = 0.67075 \tag{12.20}$$

and therefore a 95% confidence interval for $\mu_x$ is

$$(65.628, 68.061) \ .$$

### Marginal distribution of STATS

The ML estimates obtained from the marginal distribution of $\mathbf{y}$ are tabulated in Table 12.8.

**Table 12.8:** ML estimates for the marginal distribution of $\mathbf{y}$.

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

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

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

$$\widehat{\boldsymbol{\Phi}}_y = \boldsymbol{\Phi}\left(\widehat{\mathbf{z}}_y\right) = \boldsymbol{\Phi}\begin{pmatrix} -0.65670 \\ -0.04266 \\ 0.87839 \end{pmatrix} = \begin{pmatrix} 0.25569 \\ 0.48298 \\ 0.81013 \end{pmatrix} = \widehat{\boldsymbol{\pi}}_y \ . \tag{12.21}$$

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

$$\widehat{\sigma}_{\widehat{\mu}_y} = 0.63940 \quad \text{and} \quad \widehat{\sigma}_{\widehat{\sigma}_y} = 0.64606 \ . \tag{12.22}$$

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.

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

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

The ML estimate for $\rho$ is estimated by adding the appropriate relative frequencies under constraints

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

(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 |
|---|---|
| $Q_1 : z_x < 0,\ z_y < 0$ | $\widehat{\text{VOL1}} = 0.366415$ |
| $Q_2 : z_x < 0,\ z_y > 0$ | $\widehat{\text{VOL2}} = 0.133585$ |
| $Q_3 : z_x > 0,\ z_y < 0$ | $\widehat{\text{VOL3}} = 0.133585$ |
| $Q_4 : z_x > 0,\ z_y > 0$ | $\widehat{\text{VOL4}} = 0.366415$ |

The ML estimate for $\rho$ is

$$\begin{aligned}\widehat{\rho} &= \sin\left(\frac{\pi}{2}\left[2\left(0.366415\right)-2\left(0.133585\right)\right]\right) \\ &= \sin\left(\frac{\pi}{2}\left[0.732\,83-0.267\,17\right]\right) \\ &= \sin\left(\frac{\pi}{2}\left[0.465\,66\right]\right) \\ &= 0.667\,95 \end{aligned} \tag{12.24}$$

with a standard error of

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

Since

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

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

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

where

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

is the intercept and

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

is the slope, yielding the regression equation

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

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.

# Chapter 13

# Simulation study

The purpose of this sumulation study is to prove that a bivariate normal distribution can be fitted accurately to a two-way contingency table by employing the ML estimation procedure presented in Part III of this thesis. A total of 1000 samples were simulated from a bivariate normal distribution such that

$$(x, y) \sim \text{BVN}\left(11, 48, 3^2, 8^2, -0.7\right) \ .$$

Each of the data sets consisted of 1000 observations and the descriptive statistics for the sample statistics are listed in Table 13.1. From Table 13.1 it can be concluded that the sample statistics of the simulated data sets correspond very well to the theoretical values.

**Table 13.1:** Descriptive statistics for the sample statistics.

| Stat | Mean | Std.dev | $P_{05}$ | Median | $P_{95}$ |
|---|---|---|---|---|---|
| $\overline{x}$ | 11.008 | 0.0957 | 10.849 | 11.008 | 11.157 |
| $s_x$ | 2.9972 | 0.0655 | 2.887 | 2.998 | 3.110 |
| $\overline{y}$ | 47.978 | 0.2620 | 47.550 | 47.970 | 48.403 |
| $s_y$ | 7.9952 | 0.1765 | 7.703 | 7.994 | 8.291 |
| $r$ | $-0.6999$ | 0.0163 | $-0.7273$ | $-0.7000$ | $-0.6734$ |

The next step will be to cross tabulate each of the bivariate data sets into a two-way contingency table and to fit a bivariate normal distribution to each of the 1000 bivariate grouped data sets. This simulation study was done with of the SAS program *BVNSIM.SAS* listed in the Appendix C4.

## 13.1 Theoretical distribution

The simulated data sets were all categorised in a two-way contingency table, with the following upper class boundaries

$$\mathbf{x} = \begin{pmatrix} 8 \\ 10 \\ 12 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 45 \\ 50 \\ 55 \end{pmatrix} .$$

The first and last class intervals, for both variables were treated as open ended class intervals and the frequency distribution for the theoretical distribution is given in Table 13.2.

**Table 13.2:** Theoretical frequency distribution for $\mathrm{BVN}\,(11, 48, 3^2, 8^2, -0.7)$ distribution.

| X | Y | | | | |
|---|---|---|---|---|---|
| | $(-\infty, 45)$ | $[45, 50)$ | $[50, 55)$ | $[55, \infty)$ | **Total** |
| $(-\infty, 8)$ | 4.722 | 18.436 | 41.402 | 94.095 | 158.655 |
| $[8, 10)$ | 27.011 | 55.569 | 69.373 | 58.832 | 210.786 |
| $[10, 12)$ | 79.095 | 86.607 | 65.581 | 29.834 | 261.117 |
| $[12, \infty)$ | 243.002 | 84.264 | 34.150 | 8.026 | 369.441 |
| **Total** | 353.830 | 244.876 | 210.507 | 190.787 | 1000 |

The cumulative relative frequencies for the theoretical distribution, expressed in terms of matrix notation, is

$$\mathbf{\Pi} = \begin{pmatrix} 0.00472 & 0.02316 & 0.06456 & 0.15866 \\ 0.03173 & 0.10574 & 0.21651 & 0.36944 \\ 0.11083 & 0.27144 & 0.44780 & 0.63056 \\ 0.35383 & 0.59871 & 0.80921 & 1.00000 \end{pmatrix} . \tag{13.1}$$

The ML estimators of the 5 parameters of the bivariate normal distribution are all asymptotically normally distributed with standard errors functions of (13.1). The standard errors and percentiles of the ML estimators are listed in Table 13.3.

**Table 13.3:** Theoretical values for the ML estimators of the bivariate normal distribution.

| ML estimate | Standard error | Margin of error | Percentiles | | |
|---|---|---|---|---|---|
| | | | $P_{05}$ | Median | $P_{95}$ |
| $\widehat{\mu}_x$ | $\sigma_{\widehat{\mu}_x} = 0.1054$ | $1.645\sigma_{\widehat{\mu}_x} = 0.1733$ | $10.827$ | $11$ | $11.173$ |
| $\widehat{\sigma}_x$ | $\sigma_{\widehat{\sigma}_x} = 0.1123$ | $1.645\sigma_{\widehat{\sigma}_x} = 0.18466$ | $2.8153$ | $3$ | $3.1733$ |
| $\widehat{\mu}_y$ | $\sigma_{\widehat{\mu}_y} = 0.2788$ | $1.645\sigma_{\widehat{\mu}_y} = 0.45854$ | $47.541$ | $48$ | $48.459$ |
| $\widehat{\sigma}_y$ | $\sigma_{\widehat{\sigma}_y} = 0.3065$ | $1.645\sigma_{\widehat{\sigma}_y} = 0.50415$ | $7.4958$ | $8$ | $8.1733$ |
| $\widehat{\rho}$ | $\sigma_{\widehat{\rho}} = 0.021085$ | $1.645\sigma_{\widehat{\rho}} = 0.03468$ | $-0.7347$ | $-0.7$ | $-0.6653$ |

The descriptive statistics for the ML estimates of the 1000 fitted bivariate normal distributions are summarised in Table 13.4.

**Table 13.4:** Simulation results of 1000 fitted bivariate normal distributions.

| MLE | Theoretical Value | Mean | Std.dev | $P_{05}$ | Median | $P_{95}$ |
|---|---|---|---|---|---|---|
| $\widehat{\mu}_x$ | $11$ | $11.010$ | $0.1042$ | $10.842$ | $11.008$ | $11.178$ |
| $\widehat{\sigma}_{\widehat{\mu}_x}$ | $0.1054$ | $0.1055$ | $0.0045$ | $0.0980$ | $0.1055$ | $0.1130$ |
| $\widehat{\sigma}_x$ | $3$ | $3.0007$ | $0.1166$ | $2.8063$ | $3.0006$ | $3.1978$ |
| $\widehat{\sigma}_{\widehat{\sigma}_x}$ | $0.1123$ | $0.1125$ | $0.0066$ | $0.1017$ | $0.1124$ | $0.1238$ |
| $\widehat{\mu}_y$ | $48$ | $47.973$ | $0.2829$ | $47.503$ | $47.971$ | $48.426$ |
| $\widehat{\sigma}_{\widehat{\mu}_y}$ | $0.2788$ | $0.2788$ | $0.0121$ | $0.2590$ | $0.2785$ | $0.2996$ |
| $\widehat{\sigma}_y$ | $8$ | $7.9938$ | $0.3203$ | $7.4700$ | $7.9914$ | $8.5373$ |
| $\widehat{\sigma}_{\widehat{\sigma}_y}$ | $0.3065$ | $0.3066$ | $0.0187$ | $0.2763$ | $0.3062$ | $0.3387$ |
| $\widehat{\rho}$ | $-0.7$ | $-0.7006$ | $0.0243$ | $-0.7421$ | $-0.7002$ | $-0.6604$ |
| $\widehat{\sigma}_{\widehat{\rho}}$ | $0.021085$ | $0.0211$ | $0.0013$ | $0.0189$ | $0.0211$ | $0.0231$ |

It is evident from Table 13.4, that the mean for all the ML estimates are remarkably close to the theoretical values. It is also interesting to note that the standard deviation of the 5 ML estimates $\widehat{\mu}_x$, $\widehat{\sigma}_x$, $\widehat{\mu}_y$, $\widehat{\sigma}_y$ and $\widehat{\rho}$ are very close to the mean of its standard errors. E.g. the standard deviation of the $\widehat{\mu}_x$-values is $0.1042$ and the mean of the $\widehat{\sigma}_{\widehat{\mu}_x}$-values is $0.1055$. The percentiles of the ML estimates in the simulation study (see Table 13.4) correspond extremely well to that of the theoretical distribution given in Table 13.3.

A comparison between the descriptive statistics of the sample statistics of the ungrouped bivariate data sets in Table 13.1 with that of the descriptive statistics of the ML estimates of the grouped data sets tabulated in Table 13.4 shows are very close to each other. This motivates that not too much accuracy is being lost with a grouped data set, when analysed correctly.

The Wald and Pearson goodness of fit statistics were calculated for each of the 1000 estimated bivariate normal distributions. The percentiles of these two statistics are tabulated in Table 13.5 and agrees with a $\chi^2$-distribution with 10 degrees of freedom.

**Table 13.5:** Percentiles of the Pearson and Wald statistic.

| | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|
| | $P_5$ | $P_{10}$ | $P_{25}$ | $P_{50}$ | $P_{75}$ | $P_{90}$ | $P_{95}$ |
| Pearson | 3.8374 | 4.8481 | 7.1377 | 9.8363 | 13.3152 | 16.7631 | 18.9273 |
| Wald | 4.0572 | 5.2029 | 7.6182 | 10.6859 | 14.6539 | 19.3063 | 23.5933 |
| | Percentiles of a $\chi^2$-distribution with 10 degrees of freedom. | | | | | | |
| | $\chi^2_{0.05}$ | $\chi^2_{0.10}$ | $\chi^2_{0.25}$ | $\chi^2_{0.50}$ | $\chi^2_{0.75}$ | $\chi^2_{0.90}$ | $\chi^2_{0.95}$ |
| $\chi^2\,(10)$ | 3.9403 | 4.8652 | 6.7372 | 9.3418 | 12.5489 | 15.9872 | 18.3070 |

It can therefore be concluded that the empirical and theoretical distributions of the Pearson and Wald statistics correspond to each other.

# Part IV

# Chapter 14

# Résumé

The main objective of this research is to provide a theoretical foundation for analysing grouped data, taking the underlying continuous nature of the variable(s) into account. Statistical techniques have been developed and applied extensively for continuous data, but the analysis for grouped data has been somewhat neglected. This creates numerous problems especially in the social and economic disciplines, where variables are grouped for various reasons. Due to a lack for the appropriate statistical techniques to evaluate grouped data, researchers are often tempted to ignore the underlying continuous nature of the data and employ e.g. the class midpoint values as an alternative. This leads to an oversimplification of the problem and valuable information in the data is being ignored.

The first part of the thesis demonstrates how to fit a continuous distribution to a grouped data set. By implementing the ML estimation procedure of *Matthews and Crowther* (1995: *A maximum likelihood estimation procedure when modelling in terms of constraints.* South African Statistical Journal, 29, 29-51) the ML estimates of the parameters are obtained. The standard errors of the ML estimates are derived from the multivariate delta theorem. It is interesting to note that not much accuracy has been lost by grouping the data, justifying that statistical inference may be done effectively from a grouped data set. The main concern of this part of the thesis was to foster the basic principles. The examples and distributions discussed are merely used to illustrate and explain the philosophy from basic principles. The fit of various other continuous distributions, not mentioned in the thesis, such as the gamma distribution and the lognormal distribution can also be done using the same approach.

The second part of the thesis concentrates on the analysis of generalised linear models where the response variable is presented in grouped format. A cross classification of the independent variables leads to various so-called cells, each containing a frequency distribution of the response variable. Due to the nature of the response variable the usual analysis of variance and covariance models etc. can no longer be applied in the usual sense. A completely new approach, where a specified underlying continuous distribution for the grouped variable is fitted to each cell in the multifactor design is introduced. Certain measures such as the average, median or even any other percentile of the fitted distributions are modelled to explain the influence of the independent variables on the response variable. This evaluation may be done by means of a saturated model where no additional constraints are employed in the ML estimation procedure or by means of any other model where certain structures with regard to the independent variables are incorporated. The main objective is ultimately to provide a satisfactory model that describes the data as effectively as possible, revealing the various trends in the data. Employing the multivariate delta theorem, the standard errors for the ML estimates are calculated, enabling testing of relevant hypotheses. The goodness of fit of the model is evaluated with the Pearson and Wald statistics.

Two applications of multi-factor models are presented. In the first application normal distributions are fitted to the cells in a single factor design. The behavior of the mean of the fitted normal distributions revealed the effect of the single independent variable. Various models are employed to explain the versatility of the technique. Apart from the single factor model a two factor model was employed for data from short term insurance. The positive skewness of the grouped response variable suggested that a log-logistic distribution is to be fitted to the data. The median of the log-logistic distributions was modelled in a two factor model to explain the effect of the independent variable on the response variable. It is also illustrated how to incorporate a grouped independent variable as a covariate or regressor in the model. In the past where researchers might have been restricted to tabulations and graphical representations it is now shown that the possibilities of modelling a grouped response variable in a generalised model are in principle unlimited. The application of a three factor model or any higher order model follows similarly. A typical example pursue from the population census data where the grouped variable income can be explained utilising independent variables such as gender, province, population group, age, education level, occupation, etc.

A final intriguing contribution, given in the third part, is the fit of a bivariate normal distribution to a two-way contingency table. In the case where the underlying distribution of two grouped response variables are jointly normally distributed it is often required to investigate the association between two variables. Traditionally, classical measures such as kappa and McNemar were employed, but are limited in the sense that the complete bivariate structure between the two variables are not revealed. Since all five parameters are estimated, statistical inferences are possible with regard to the marginal as well as the partial distributions. The estimation of the parameter $\rho$, the correlation coefficient, explains the relationship between the two variables. The calculation of $\rho$ is done by implementing *Sheppard's theorem on median dichotomy (1898)*, which is based on the volumes of the four quadrants of the bivariate normal distribution. It is shown that the calculation of the correlation coefficient, using the standard regression techniques, could lead to incorrect results due to the fact that the required conditions are not met. The method proposed is motivated by a simulation study.

Although various aspects of modelling grouped data are addressed in this thesis, this forms the basic building blocks for the beginning of a completely new and promising field of research with unlimited possibilities and exciting applications to be analysed.

# References


**Agresti, A.** (1984) *Analysis of ordinal categorical data.* New York: Wiley.

**Agresti, A.** (1984) *Categorical data analysis.* New York: Wiley.

**Bishop, Y.V.V., Fienberg S.E. and Holland, P.W.** (1975) *Discrete multivariate analysis.* Cambridge, MA: MIT Press.

**Central Statistical Service, Republic of South Africa** (1985). *Survey of household expenditure*. report No. 01-11-02(1985).

**Elderton, W.P. and Johnson, N.L.** (1969) *Systems of frequency curves*. Cambridge University Press.

**Johnson, N.L. and Kotz, S.** (1970) *Continuous univariate distributions-1. Distributions in statistics*. John Wiley & Sons

**Joubert, H.M. and Crowther, N.A.S.** (1992) Fitting of distributions in the case of survey data. *HSRC Report WS-50*.

**Kendall, M.G. and Stuart, A.** (1958) *The advanced theory of statistics.* Griffin.

**Matthews, G.B. and Crowther, N.A.S.** (1995) A maximum likelihood estimation procedure when modelling in terms of constraints. *S.A Statist. J.*, **29**, *29-51*.

**Matthews, G.B. and Crowther, N.A.S.** (1997) A maximum likelihood estimation procedure when modelling categorical data in terms of cross-product ratios. *S.A Statist. J.*, **31**, *161-184*.

**Matthews, G.B. and Crowther, N.A.S.** (1998) A maximum likelihood estimation procedure for the generalized linear model with restrictions. *S.A Statist. J.*, **32**, *119-144*.

**Muirhead, R.J.** (1982) *Aspects of Multivariate statistical theory.* John Wiley & Sons

**Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W.** (1996) *Applied linear statistical models.* Mc Graw-Hill.

**Pearson, K.** (1924). On the mean error of frequency distributions. *Biometrika*, **16**, *198-200*

# Part V

# Appendix

# Appendix A

# SAS programs: Part I

## A.1 EXP1.SAS

```
proc iml worksize= 60;
f={17,14,31,26,12}; n=f[+];
x={12.5,25,50,100};
C={1 0 0 0,
   1 1 0 0,
   1 1 1 0,
   1 1 1 1};
CI=inv(C);
k=nrow(f); k1=k-1;
v1=J(k1,1,1);
Px=x*inv(x‘*x)*x‘;
p=C*f[1:k1]/n;
i=0; p0=p; diff1=1;
do while (diff1 > 1e-9);
  i=i+1; pi=p; p=p0;
  V=(C*diag(CI*pi)*C‘-pi*pi‘)/n;
  thetapi=-(x‘*log(v1-pi))/(x‘*x); mupi=1/thetapi;
  Dpi=-inv(diag(v1-pi));
```

```
  Gpi=-diag(exp(-thetapi*x))*Px*Dpi - I(k1);
  j=0; diff=1;
  do while (diff > 1e-9);
    j=j+1; pv=p;
    thetap=-(x`*log(v1-p))/(x`*x); mup=1/thetap;
    Dp=-inv(diag(v1-p));
    Gp=-diag(exp(-thetap*x))*Px*Dp - I(k1);
    g=(v1-exp(-thetap*x))-p;
    print i j g pi p thetapi mupi thetap mup;
    p=p-(Gpi*V)`*ginv(Gp*V*Gpi`)*g;
    diff=sqrt((p-pv)`*(p-pv));
  end;
  diff1=sqrt((p-pi)`*(p-pi));
end;
```

## A.2 EXP2.SAS

```
proc iml worksize= 60;
f={17,14,31,26,12}; n=f[+];
x={12.5,25,50,100};
k=nrow(f); k1=k-1;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
p=C*f[1:k1]/n;
v1=J(k1,1,1);
Q=I(k1)-x*inv(x`*x)*x`;
i=0; p0=p; diff1=1;
do while (diff1 > 1e-9);
    i=i+1; pi=p; p=p0;
    Dpi=inv(diag(pi-v1));
    Gpi=Q*Dpi;
    V=(C*diag(CI*pi)*C`-pi*pi`)/n;
    j=0; diff=1;
```

```
    do while (diff > 1e-9);
        j=j+1; pv=p;
        Dp=inv(diag(p-v1));
        Gp=Q*Dp;
        g=Q*log(v1-p);
        print i j p pi;
        p=p-(Gpi*V)‘*ginv(Gp*V*Gpi‘)*g;
        diff=sqrt((p-pv)‘*(p-pv));
        if i=1 & j=1 then do;
            Wald=g‘*ginv(Gp*V*Gp‘)*g;
            GpV=Gp*V;
            df=trace(GpV*ginv(GpV‘*GpV)*GpV‘);
            discr=wald/n;
        end;
    end;
    diff1=sqrt((p-pi)‘*(p-pi));
end;
Cov_pi=V-(Gpi*V)‘*ginv(Gpi*V*Gpi‘)*(Gpi*V);
theta=-(x‘*log(v1-pi))/(x‘*x);
Var_theta=((x‘*Dpi)/(x‘*x))*Cov_pi*((x‘*Dpi)/(x‘*x))‘;
mu=1/theta;
SE_mu=sqrt(Var_theta/(theta**4));

e=(CI*pi*n)//(n-(CI*pi*n)[+]);
Pearson=(((f-e)##2)/e)[+];
P_pvalue=1-probchi(Pearson,df);
W_pvalue=1-probchi(Wald,df);
print mu SE_mu Pearson P_pvalue Wald W_pvalue df;
```

## A.3 EXPSIM.SAS

```
proc iml;
rep=1000; n=100; theta0=50;
matrix=J(rep,4,0);
x={12.5,25,50,100};
xl=0//x;
xu=x//250;
mid=(xl+xu)/2;
mlb=J(n,1,1)@xl‘;
mub=J(n,1,1)@xu‘;
k=nrow(xu); k1=k-1;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
v1=J(k1,1,1);
Q=I(k1)-x*inv(x‘*x)*x‘;
do r=1 to rep;
  y=theta0#ranexp(J(n,1,r));
  my=y@J(k,1,1)‘;
  t=((my>mlb)=(my<=mub));
  f=t[+,]‘;
  p=C*f[1:k1]/n;
  i=0; p0=p; diff1=1;
  do while (diff1 > 1e-9);
    i=i+1; pi=p; p=p0;
    Dpi=inv(diag(pi-v1));
    Gpi=Q*Dpi;
    V=(C*diag(CI*pi)*C‘-pi*pi‘)/n;
    j=0; diff=1;
    do while (diff > 1e-9);
      j=j+1; pv=p;
      Dp=inv(diag(p-v1));
      Gp=Q*Dp;
      g=Q*log(v1-p);
```

```
      p=p-(Gpi*V)`*ginv(Gp*V*Gpi`)*g;
      diff=sqrt((p-pv)`*(p-pv));
      if i=1 & j=1 then do;
        Wald=g`*ginv(Gp*V*Gp`)*g;
        GpV=Gp*V;
        df=trace(GpV*ginv(GpV`*GpV)*GpV`);
        pvalue=1-probchi(Wald,df);
        discr=wald/n;
      end;
    end;
    diff1=sqrt((p-pi)`*(p-pi));
  end;
  theta=-(x`*log(v1-pi))/(x`*x);
  mu=1/theta;
  Cov_pi=V-(Gpi*V)`*ginv(Gpi*V*Gpi`)*(Gpi*V);
  Var_theta=((x`*Dpi)/(x`*x))*Cov_pi*((x`*Dpi)/(x`*x))`;
  SE_mu=sqrt(Var_theta/(theta**4));
  e=(CI*pi*n)//(n-(CI*pi*n)[+]);
  Pearson=(((f-e)##2)/e)[+];
  matrix[r,1]=mu;
  matrix[r,2]=SE_mu;
  matrix[r,3]=Pearson;
  matrix[r,4]=Wald;
end;
create d from matrix[colname={'mu' 'SE_mu' 'Pearson' 'Wald'}];
append from matrix;

proc means data=d n mean std p5 p50 p95;
var mu SE_mu wald;
run;

proc univariate data=d normal plot;
var Pearson;
output out=pp pctlpts=5 10 25 50 75 90 95 pctlpre=pp;
```

```
run;

proc univariate data=d normal plot;
var Wald;
output out=pw pctlpts=5 10 25 50 75 90 95 pctlpre=pw;
run;

proc print data=pp;
run;
proc print data=pw;
run;
```

## A.4 NORM1.SAS

```
proc iml worksize= 60;
f={9,26,24,27,14}; n=f[+];
x={40,50,60,75};
k=nrow(f); k1=k-1;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
v1=J(k1,1,1);
XD=x||J(k1,1,-1);
XXX=inv(XD`*XD)*XD`;
Px=XD*inv(XD`*XD)*XD`;
p=C*f[1:k1]/n;
i=0; p0=p; diff1=1;
do while (diff1 > 1e-9);
  i=i+1; pi=p; p=p0;
  V=(C*diag(CI*pi)*C`-pi*pi`)/n;
  alphapi=XXX*probit(pi);
  mupi=alphapi[2]/alphapi[1]; sigmapi=1/alphapi[1];
  zpi=XD*alphapi;
  Dpi=inv(diag(pdf('normal',probit(pi))));
```

```
  Gpi=diag(pdf('normal',zpi))*Px*Dpi - I(k1);
  j=0; diff=1;
  do while (diff > 1e-9);
    j=j+1; pv=p;
    alphap=XXX*probit(p);
    mup=alphap[2]/alphap[1]; sigmap=1/alphap[1];
    zp=XD*alphap;
    Dp=inv(diag(pdf('normal',probit(p))));
    Gp=diag(pdf('normal',zp))*Px*Dp - I(k1);
    g=probnorm(zp)-p;
    print alphap i j g pi[format=6.4] p[format=6.4] mupi sigmapi mup sigmap;
    p=p-(Gpi*V)`*ginv(Gp*V*Gpi`)*g;
    diff=sqrt((p-pv)`*(p-pv));
  end;
  diff1=sqrt((p-pi)`*(p-pi));
end;
```

## A.5 NORM2.SAS

```
proc iml worksize= 60;
f={9,26,24,27,14}; n=f[+];
x={40,50,60,75};
n=f[+];
k=nrow(f); k1=k-1;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
p=C*f[1:k1]/n;
v1=J(k1,1,1);
XD=x||J(k1,1,-1);
XXX=inv(XD`*XD)*XD`;
Px=XD*inv(XD`*XD)*XD`;
Q=I(k1)-Px;
```

```
*** Theoretical value ***;
*p=(probnorm((x-58)/15));
***;

i=0; p0=p; diff1=1;
do while (diff1 > 1e-9);
  i=i+1; pi=p; p=p0;
  Dpi=inv(diag(pdf('normal',probit(pi))));
  Gpi=Q*Dpi;
  V=(C*diag(CI*pi)*C`-pi*pi`)/n;
  j=0; diff=1;
  do while (diff > 1e-9);
    j=j+1; pv=p;
    Dp=inv(diag(pdf('normal',probit(p))));
    Gp=Q*Dp;
    g=Q*probit(p);
    p=p-(Gpi*V)`*ginv(Gp*V*Gpi`)*g;
    print i j p pi;
    diff=sqrt((p-pv)`*(p-pv));
    if i=1 & j=1 then do;
      Wald=g`*ginv(Gp*V*Gp`)*g;
      GpV=Gp*V;
      df=trace(GpV*ginv(GpV`*GpV)*GpV`);
      discr=wald/n;
    end;
  end;
  diff1=sqrt((p-pi)`*(p-pi));
end;
Cov_pi=V-(Gpi*V)`*ginv(Gpi*V*Gpi`)*(Gpi*V);
alpha=XXX*probit(pi);
Cov_alpha=(XXX*Dpi)*Cov_pi*(XXX*Dpi)`;
mu=alpha[2]/alpha[1]; sigma=1/alpha[1];
print mu sigma;
beta=mu//sigma;
```

```
B=J(2,2,0);
B[1,1]=-alpha[2]/((alpha[1])**2);
B[1,2]=1/(alpha[1]);
B[2,1]=-1/((alpha[1])**2);
Cov_beta=B*Cov_alpha*B‘;
SE_beta=sqrt(diag(Cov_beta));

e=(CI*pi*n)//(n-(CI*pi*n)[+]);
Pearson=(((f-e)##2)/e)[+];
P_pvalue=1-probchi(Pearson,df);
W_pvalue=1-probchi(Wald,df);
print beta Cov_beta SE_beta, mu sigma, Pearson P_pvalue Wald W_pvalue df;

probitp=probit(p);
Pprobitp=Px*probitp;
Qprobitp=Q*probitp;
print probitp [format=9.7] Pprobitp[format=9.7] Qprobitp[format=9.7];
```

## A.6 NORMSIM.SAS

```
proc iml worksize= 60;
rep=1000; n=100; mu0=58; sigma0=15; x={40,50,60,75};
matrix=J(rep,8,0);
k1=nrow(x); k=k1+1;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
v1=J(k1,1,1);
XD=x||J(k1,1,-1);
XXX=inv(XD‘*XD)*XD‘;
Px=XD*inv(XD‘*XD)*XD‘;
Q=I(k1)-Px;
xl=0//x; xu=x//100;
mlb=J(n,1,1)@xl‘; mub=J(n,1,1)@xu‘;
```

```
start data;
  sp=mu0*J(n,1,1)+sigma0*rannor(J(n,1,r));
  xbar=sp[+]/n;
  xstd=sqrt(sp`*sp/n-xbar**2);
  sss=sp@J(1,k,1);
  t=(sss>mlb)=(sss<=mub);
  f=t[+,]`;
  p=C*f[1:k1]/n;
finish;

start fit;
  i=0; p0=p; diff1=1;
  do while (diff1 > 1e-9);
    i=i+1; pi=p; p=p0;
    Dpi=inv(diag(pdf('normal',probit(pi))));
    Gpi=Q*Dpi;
    V=(C*diag(CI*pi)*C`-pi*pi`)/n;
    j=0; diff=1;
    do while (diff > 1e-9);
      j=j+1; pv=p;
      Dp=inv(diag(pdf('normal',probit(p))));
      Gp=Q*Dp;
      g=Q*probit(p);
      p=p-(Gpi*V)`*ginv(Gp*V*Gpi`)*g;
      diff=sqrt((p-pv)`*(p-pv));
      if i=1 & j=1 then do;
        Wald=g`*ginv(Gp*V*Gp`)*g;
        GpV=Gp*V;
        df=trace(GpV*ginv(GpV`*GpV)*GpV`);
        discr=wald/n;
      end;
    end;
    diff1=sqrt((p-pi)`*(p-pi));
```

```
  end;
  Cov_pi=V-(Gpi*V)`*ginv(Gpi*V*Gpi`)*(Gpi*V);
  alpha=XXX*probit(pi);
  Cov_alpha=(XXX*Dpi)*Cov_pi*(XXX*Dpi)`;
  mu=alpha[2]/alpha[1]; sigma=1/alpha[1];
  beta=mu//sigma;
  B=J(2,2,0);
  B[1,1]=-alpha[2]/((alpha[1])**2);
  B[1,2]=1/(alpha[1]);
  B[2,1]=-1/((alpha[1])**2);
  Cov_beta=B*Cov_alpha*B`;
  SE_beta=diag(sqrt(diag(Cov_beta)));

  e=(CI*pi*n)//(n-(CI*pi*n)[+]);
  Pearson=(((f-e)##2)/e)[+];
  P_pvalue=1-probchi(Pearson,df);
  W_pvalue=1-probchi(Wald,df);

  matrix[r,1]=xbar;
  matrix[r,2]=xstd;
  matrix[r,3]=mu;
  matrix[r,4]=(SE_beta[1,1]);
  matrix[r,5]=sigma;
  matrix[r,6]=(SE_beta[2,2]);
  matrix[r,7]=Pearson;
  matrix[r,8]=Wald;
finish;

do r=1 to rep;
  run data;
  run fit;
end;

create d from
```

```
  matrix[colname={'xbar' 'xstd' 'mu' 'SE_mu' 'sigma' 'SE_sigma' 'Pearson' 'Wald'}];
append from matrix;

proc means data=d maxdec=3 n mean std p5 p50 p95;
var xbar xstd mu SE_mu sigma SE_sigma;
run;

proc means data=d maxdec=4 p5 p10 p25 p50 p75 p90 p95;
var Pearson Wald;
run;
```

## A.7 FIT.SAS

```
proc iml worksize= 60;
*************************;
*    Exponential ='E'   *;
*    Normal      ='N'   *;
*    Weibull     ='W'   *;
*    Log-logistic='L'   *;
*    Pareto      ='P'   *;
*************************;

*===>; distr='W';
*===>; f={9,37,67,63,30}; x={40,75,125,175}; x=x-0.5;

n=f[+];
k=nrow(f); k1=k-1;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
v1=J(k1,1,1);
p=C*f[1:k1]/n;

start X;
```

```
    if distr='E' then XD=-x;
    if distr='N' then XD=x||(-v1);
    if distr='W' then XD=log(x)||(-v1);
    if distr='L' then XD=log(x)||v1;
    if distr='P' then XD=-log(x)||v1;
finish;

start h;
    if distr='E' then h=log(v1-p);
    if distr='N' then h=probit(p);
    if distr='W' then h=log(-log(v1-p));
    if distr='L' then h=log(p/(v1-p));
    if distr='P' then h=log(v1-p);
finish;

start D(Dp,p) global(distr,v1);
    if distr='E' then Dp=inv(diag(p-v1));
    if distr='N' then Dp=inv(diag(pdf('normal',probit(p))));
    if distr='W' then Dp=-inv(diag(log(v1-p)))*inv(diag(v1-p));
    if distr='L' then Dp=inv(diag(p))+inv(diag(v1-p));
    if distr='P' then Dp=-inv(diag(v1-p));
finish;

start beta;
    if distr='E' then beta=1/alpha;
    if distr='N' then do;
        beta[1]=alpha[2]/alpha[1];
        beta[2]=1/alpha[1];
    end;
    if (distr='W' | distr='P') then do;
        beta[1]=alpha[1];
        beta[2]=exp(alpha[2]/alpha[1]);
    end;
    if distr='L' then beta=alpha;
```

```
finish;

start B;
      if distr='E' then B=-1/(alpha**2);
      if distr='N' then do;
            B[1,1]=-alpha[2]/((alpha[1])**2);
            B[1,2]=1/(alpha[1]);
            B[2,1]=-1/((alpha[1])**2);
      end;
      if (distr='W' | distr='P') then do;
            B[1,1]=1;
            B[2,1]=-(alpha[2])/((alpha[1])**2)*exp((alpha[2])/(alpha[1]));
            B[2,2]=inv(alpha[1])*exp((alpha[2])/(alpha[1]));
      end;
      if distr='L' then B=I(nrow(alpha));
finish;

start wald;
      Wald=g`*ginv(Gp*V*Gp`)*g;
      GpV=Gp*V;
      df=trace(GpV*ginv(GpV`*GpV)*GpV`);
finish;

start mu;
      if distr='E' then mu=beta;
      if distr='N' then mu=beta[1];
      if distr='W' then mu=beta[2]*(gamma(1+1/beta[1]));
      if distr='L' then mu=exp(-beta[2]/beta[1])
                    *gamma(1+1/beta[1])*gamma(1-1/beta[1]);
      if distr='P' then mu=(beta[1]*beta[2])/(beta[1]-1);
finish;

start sigma;
      if distr='E' then sigma=beta;
```

```
      if distr='N' then sigma=beta[2];
      if distr='W' then sigma=sqrt(beta[2]**2
                        *(gamma(1+2/beta[1])-(gamma(1+1/beta[1]))**2));
      if distr='L' then sigma=sqrt(exp(-2*beta[2]/beta[1])
                        *gamma(1+2/beta[1])*gamma(1-2/beta[1])
                        - (gamma(1+1/beta[1])*gamma(1-1/beta[1]))**2));
      if distr='P' then sigma=sqrt((beta[1]*beta[2]**2)
                        /((beta[1]-1)**2*(beta[1]-2)));
finish;

run X;
Q=I(k1)-XD*inv(XD`*XD)*XD`;

i=0; p0=p; diff1=1;
do while (diff1 > 1e-9);
      i=i+1; pi=p; p=p0;
      run D(Dpi,pi);
      Gpi=Q*Dpi;
      V=(C*diag(CI*pi)*C`-pi*pi`)/n;
      j=0; diff=1;
      do while (diff > 1e-9);
            j=j+1; pv=p;
            run D(Dp,p);
            run h;
            Gp=Q*Dp;
            g=Q*h;
            print i j p pi g;
            p=p-(Gpi*V)`*ginv(Gp*V*Gpi`)*g;
            diff=sqrt((p-pv)`*(p-pv));
            if i=1 & j=1 then run wald;
      end;
      diff1=sqrt((p-pi)`*(p-pi));
end;
Cov_pi=V-(Gpi*V)`*ginv(Gpi*V*Gpi`)*(Gpi*V);
```

```
alpha=inv(XD`*XD)*XD`*h;
Cov_alpha=(inv(XD`*XD)*XD`*Dpi)*Cov_pi*(inv(XD`*XD)*XD`*Dpi)`;
SE_alpha=sqrt(diag(Cov_alpha)*J(nrow(alpha),1,1));
print alpha Cov_alpha SE_alpha;

beta=J(nrow(alpha),1,0); run beta;
B=J(nrow(alpha),nrow(alpha),0); run B;
Cov_beta=B*Cov_alpha*B`;
SE_beta=sqrt(diag(Cov_beta)*J(nrow(beta),1,1));
print beta Cov_beta SE_beta;

run mu; run sigma;
print mu sigma;

e=(CI*pi*n)//(n-(CI*pi*n)[+]);
Pearson=(((f-e)##2)/e)[+];
P_pvalue=1-probchi(Pearson,df);
W_pvalue=1-probchi(Wald,df);
discr=wald/n;
print Pearson P_pvalue Wald W_pvalue df discr;
```

# Appendix B

# SAS programs: Part II

## B.1 FACTOR1.SAS

```
data d;
set phdabc.wisk;
if jaar=2003 & vlak=1 & wisk in('A','B','C','D','E') & 0<=finaal<=108;
maths=wisk;
if 0<=eksamen<40   then stats=40;
if 40<=eksamen<50  then stats=50;
if 50<=eksamen<60  then stats=60;
if 60<=eksamen<75  then stats=75;
if 75<=eksamen<=108 then stats=108;
keep maths stats;
run;

proc freq data=d noprint;
tables maths / out=factor1;
tables stats / out=class;
tables maths*stats / out=freq;
run;
```

```
*** Start: Empty cells ***;
data t;
maths='A'; stats=39; count=0;
output;
run;
data freq; set freq t;
run;
proc sort data=freq;
by maths stats;
run;
*** Finish: Empty cells ***;

proc transpose data=freq out=freq prefix=c;
by maths;
var count;
run;

proc iml worksize=200 symsize=2000;
use freq; read all var{c1 c2 c3 c4 c5} into freq;
use class; read all var{stats} into class;
use factor1; read all var{maths} into factor1;

n=freq[+];
nt=nrow(freq);
k=nrow(freq); k1=k-1;
x=class[1:k1]; x=x-0.5;
nn=freq[,+];
f=colvec(freq[,1:k1]);  f=f<>0.0001;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
v1=J(k1,1,1);
po=inv(diag(nn)@I(k1))*f;
p=(I(nt)@C)*po;
print freq factor1 class x;
```

```
XD=x||-v1;
XXX=inv(XD`*XD)*XD`; XXX1=XXX[1,]; XXX2=XXX[2,];
Px=XD*inv(XD`*XD)*XD`;
Q=I(k1)-Px;
nor=(I(nt)@Q);

H=J(nt-1,1,1)||-I(nt-1);
var=H*(I(nt)@XXX1);

nfac1=nrow(factor1);
*Yar={2,1,0,-1,-2};                          *<=== Factor A: ordinal ***;
Yar={90,75,65,55,45};                       *<=== Factor A: linear   ***;
YD=J(nt,1,1)||Yar;
YYY=inv(YD`*YD)*YD`;
Qr=I(nt)-YD*inv(YD`*YD)*YD`;
reg=Qr*(I(nt)@XXX2);

*ZD=nor;                                    *<=== Model 1;
*ZD=nor//var;                               *<=== Model 2;
ZD=nor//var//reg;                          *<=== Model 3-4;

i=0; p0=p; diff1=1;
do while (diff1 > 1e-9);
        i=i+1; pi=p; p=p0;
        Dpi=inv(diag(pdf('normal',probit(pi))));
        Gpi=ZD*Dpi;
    pio=(I(nt)@CI)*pi;
        Vo=inv(diag(nn)@I(k1))*(diag(pio)
           -(diag(pio))*(I(nt)@(v1*v1`))*(diag(pio)));
        V=(I(nt)@C)*Vo*(I(nt)@C)`;
        j=0; diff=1;
        do while (diff > 1e-9);
                j=j+1; pv=p;
```

```
                Dp=inv(diag(pdf('normal',probit(p))));
                Gp=ZD*Dp;
                hp=probit(p);
                g=ZD*hp;
*                print i j p pi g;
                p=p-(Gpi*V)`*ginv(Gp*V*Gpi`)*g;
                diff=sqrt((p-pv)`*(p-pv));
                if i=1 & j=1 then do;
                        Wald=g`*ginv(Gp*V*Gp`)*g;
                        GpV=Gp*V;
                        df=trace(GpV*ginv(GpV`*GpV)*GpV`);
                        discr=wald/n;
                end;
        end;
        diff1=sqrt((p-pi)`*(p-pi));
end;
Cov_pi=V-(Gpi*V)`*ginv(Gpi*V*Gpi`)*(Gpi*V);

alpha=(I(nt)@XXX)*hp;
alpha1=(I(nt)@XXX1)*hp;
alpha2=(I(nt)@XXX2)*hp;
Cov_alpha=((I(nt)@XXX)*Dpi)*Cov_pi*((I(nt)@XXX)*Dpi)`;

mu=alpha2/alpha1;
sigma=1/alpha1;
beta=(mu@{1,0})+(sigma@{0,1});
B11=-alpha2/(alpha1#alpha1);
B12=1/alpha1;
B21=-1/(alpha1#alpha1);
I11=J(2,2,0);I12=J(2,2,0);I21=J(2,2,0);
I11[1,1]=1;I12[1,2]=1;I21[2,1]=1;
B=(diag(B11)@I11)+(diag(B12)@I12)+(diag(B21)@I21);
Cov_beta=B*Cov_alpha*B`;
B1=(diag(B11)@{1 0})+(diag(B12)@{0 1});
```

```
B2=(diag(B21)@{1 0});
Cov_mu=B1*Cov_alpha*B1`;
Cov_sigma=B2*Cov_alpha*B2`;
SE_mu=sqrt(diag(Cov_mu)*J(nrow(mu),1,1));
SE_sigma=sqrt(diag(Cov_sigma)*J(nrow(sigma),1,1));
print mu SE_mu, sigma SE_sigma;

gamma=YYY*mu;
Cov_gamma=YYY*Cov_mu*YYY`;
SE_gamma=sqrt(diag(Cov_mu)*J(nrow(gamma),1,1));
print gamma SE_gamma;

Za=designf(cusum(J(nfac1,1,1)));

LD=J(nt,1,1)||Za;
LLL=inv(LD`*LD)*LD`;
lambda=LLL*mu;
lambda=choose(abs(lambda)<1e-9,0,lambda);
Cov_lambda=LLL*Cov_mu*LLL`;
Cov_lambda=choose(abs(Cov_lambda)<1e-9,0,Cov_lambda);
SE_lambda=sqrt(diag(Cov_lambda)*J(nrow(lambda),1,1));
print lambda SE_lambda;

TTT=block(1,Za);
tau=TTT*lambda;
Cov_tau=TTT*Cov_lambda*TTT`;
SE_tau=sqrt(diag(Cov_tau)*J(nrow(tau),1,1));
print tau SE_tau;

count=cusum(1//nfac1);
tau0=tau[count[1]:count[1]];     SE_tau0=SE_tau[count[1]:count[1]];
tau1=tau[count[1]+1:count[2]];   SE_tau1=SE_tau[count[1]+1:count[2]];
print tau0 SE_tau0, tau1 SE_tau1;
```

```
piom=(shape(pio,nt));
exp1=piom#(repeat(nn,1,k1));
exp2=nn-exp1[,+];
exp=exp1||exp2;

Pearson=(((freq-exp)##2)/exp)[+];
P_pvalue=1-probchi(Pearson,df);
W_pvalue=1-probchi(Wald,df);
print freq exp, Pearson P_pvalue Wald W_pvalue df;
```

## B.2 FACTOR2.SAS

```
proc freq data=phdabc.sbib noprint;
tables product / out=product;
tables agegrp / out=agegrp;
tables agec / out=agec;
tables premium / out=class;
tables agegrp*product*premium / out=b;
run;

proc transpose data=b out=freq prefix=c;
by agegrp product;
var count;
run;

proc iml worksize=200 symsize=2000;
use freq; read all var{c1 c2 c3 c4 c5} into freq;
use class; read all var{premium} into class;
use agegrp; read all var{agegrp} into factor1;
use product; read all var{product} into factor2;
print freq factor1 factor2; print class;
```

```
n=freq[+];
nt=nrow(freq);
k=nrow(class); k1=k-1;
x=class[1:k1];
nn=freq[,+];
f=colvec(freq[,1:k1]);  f=f<>0.0001;
C=J(k1,1,1)@cusum(J(1,k1,1))<=J(1,k1,1)@cusum(J(k1,1,1));
CI=inv(C);
v1=J(k1,1,1);
po=inv(diag(nn)@I(k1))*f;
p=(I(nt)@C)*po;

XD=log(x)||v1;
XXX=inv(XD`*XD)*XD`; XXX1=XXX[1,]; XXX2=XXX[2,];
Px=XD*inv(XD`*XD)*XD`;
Qx=I(k1)-Px;

nfac1=nrow(factor1);
nfac2=nrow(factor2);
print n nt k x, f po p ;

*Y1=designf(cusum(J(nfac1,1,1))@J(nfac2,1,1));  *<=== Factor A: dummy;
Y1={24.5,34.5,44.5,54.5}@J(nfac2,1,1);          *<=== Factor A: linear;
Y2=designf(J(nfac1,1,1)@cusum(J(nfac2,1,1)));
Y12=hdir(Y1,Y2);
*YD=J(nt,1,1)||Y1||Y2;            *<=== Only main effects;
YD=J(nt,1,1)||Y1||Y2||Y12;        *<=== Main effects with interaction;
Py=YD*inv(YD`*YD)*YD`;
Qy=I(nt)-Py;

start GGG(p,g,GG) global(nt,v1,Qx,XXX,XXX1,XXX2,Qy,h,D,kappa,theta,nu,A,Y12);
  h=log(p/((J(nt,1,1)@v1)-p));
  D=inv(diag(p))+inv(diag((J(nt,1,1)@v1)-p));
```

```
  glog=(I(nt)@Qx)*h;
  GGlog=(I(nt)@Qx)*D;

  kappa=(I(nt)@XXX1)*h;
  theta=(I(nt)@XXX2)*h;
  nu=exp(-theta/kappa);
   A1=nu#(theta/(kappa#kappa));
   A2=nu#(-1/kappa);
   A=diag(A1)@{1 0} + diag(A2)@{0 1};
  greg=Qy*nu;
   GGreg=Qy*A*(I(nt)@XXX)*D;

*  g=glog;          *<=== Model 1;
*  GG=GGlog;        *<=== Model 1;
  g=glog//greg;    *<=== Model 2-4;
  GG=GGlog//GGreg; *<=== Model 2-4;
finish;

i=0; p0=p; diff1=1;
do while (diff1 > 1e-9);
  i=i+1; pi=p; p=p0;
    pio=(I(nt)@CI)*pi;
  Vo=inv(diag(nn)@I(k1))*(diag(pio)- (diag(pio))*(I(nt)@(v1*v1`))*(diag(pio))`);
  V=(I(nt)@C)*Vo*(I(nt)@C)`;
  run GGG(pi,gpi,GGpi);
  j=0; diff=1;
  do while (diff > 1e-9);
    j=j+1; pv=p;
    run GGG(p,gp,GGp);
    print i j p pi gp;
    p=p-(GGpi*V)`*ginv(GGp*V*GGpi`)*gp;
    diff=sqrt((p-pv)`*(p-pv));
    if i=1 & j=1 then do;
      Wald=gp`*ginv(GGp*V*GGp`)*gp;
```

```
      GpV=GGp*V;
      df=trace(GpV*ginv(GpV`*GpV)*GpV`);
      discr=wald/n;
    end;
  end;
  diff1=sqrt((p-pi)`*(p-pi));
end;
nnm=shape(nn,nfac1);
thetam=shape(theta,nfac1);
kappam=shape(kappa,nfac1);

Cov_pi=V-(GGpi*V)`*ginv(GGpi*V*GGpi`)*(GGpi*V);
Cov_alpha=((I(nt)@XXX)*D)*Cov_pi*((I(nt)@XXX)*D)`;

mu=exp(-theta/kappa)#gamma(J(nt,1,1)+1/kappa)#gamma(J(nt,1,1)-1/kappa);
sigma=sqrt(exp(-2*theta/kappa)
  #(gamma(J(nt,1,1)+2/kappa)#gamma(J(nt,1,1)-2/kappa)
  -(gamma(J(nt,1,1)+1/kappa)#gamma(J(nt,1,1)-1/kappa))##2));
mum=shape(mu,nfac1);
sigmam=shape(sigma,nfac1);
print mum sigmam;

Cov_nu=A*Cov_alpha*A`;
SE_nu=sqrt(diag(Cov_nu)*J(nrow(nu),1,1));
num=shape(nu,nfac1); SE_num=shape(SE_nu,nfac1);
print num SE_num;

YYY=inv(YD`*YD)*YD`;
gamma=YYY*nu;
Cov_gamma=YYY*Cov_nu*YYY`;
SE_gamma=sqrt(diag(Cov_gamma)*J(nrow(gamma),1,1));
print gamma SE_gamma;

D1=designf(cusum(J(nfac2,1,1)));
```

```
DDD=block(1,1,D1,D1);
delta=DDD*gamma;
Cov_delta=DDD*Cov_gamma*DDD`;
SE_delta=sqrt(diag(Cov_delta)*J(nrow(delta),1,1));
print delta SE_delta;

Za=designf(cusum(J(nfac1,1,1))@J(nfac2,1,1));  *<=== Factor A: dummy;
Zb=designf(J(nfac1,1,1)@cusum(J(nfac2,1,1)));
Zab=hdir(Za,Zb);
LD=J(nt,1,1)||Za||Zb||Zab;           *<== saturated model;
LLL=inv(LD`*LD)*LD`;
lambda=LLL*nu;
lambda=choose(abs(lambda)<1e-9,0,lambda);
Cov_lambda=LLL*Cov_nu*LLL`;
Cov_lambda=choose(abs(Cov_lambda)<1e-9,0,Cov_lambda);
print LD lambda;

S1=designf(cusum(J(nfac1,1,1)));
S2=designf(cusum(J(nfac2,1,1)));
S12=S1@S2;
S=block(1,S1,S2,S12);
tau=S*lambda;
Cov_tau=S*Cov_lambda*S`;
SE_tau=sqrt(diag(Cov_tau)*J(nrow(tau),1,1));
print tau SE_tau;

count=cusum(1//nfac1//nfac2//(nfac1*nfac2));
tau0=tau[1:1];                      SE_tau0=SE_tau[1:1];
tau1=tau[count[1]+1:count[2]];   SE_tau1=SE_tau[count[1]+1:count[2]];
tau2=tau[count[2]+1:count[3]];   SE_tau2=SE_tau[count[2]+1:count[3]];
tau12=tau[count[3]+1:count[4]]; SE_tau12=SE_tau[count[3]+1:count[4]];
tau12m=shape(tau12,nfac1);       SE_tau12m=shape(SE_tau12,nfac1);
print tau0 tau1 tau2 tau12m, SE_tau0 SE_tau1 SE_tau2 SE_tau12m;
```

```
piom=(shape(pio,nt));
exp1=piom#(repeat(nn,1,k1));
exp2=nn-exp1[,+];
exp=exp1||exp2;

Pearson=(((freq-exp)##2)/exp)[+];
P_pvalue=1-probchi(Pearson,df);
W_pvalue=1-probchi(Wald,df);
print freq exp, Pearson P_pvalue Wald W_pvalue df;

*** Start: Graph ***;
*** Eerste fig: 4.5 en 3.5cm - Tweede fig: 4 en 3cm;
xl=0.5//class[1:k1];
xu=class;
width=xu-xl;
print xl xu x width;
```

# Appendix C

# SAS Programs: Part III

## C.1 Phi0.SAS

```
proc iml;
*===>;a=1; b=2; rho=0.5;

pi=(gamma(0.5))**2;
diff=1; Phi0=0; i=0;
do while (diff>1e-8);
  vorige=Phi0;
  Phi0=Phi0 + ((2*rho)##i*sqrt(1-rho##2))/(4*pi*gamma(i+1))*gamma((i+1)/2)##2
      * probgam((a##2/(2*(1-rho##2))),(i+1)/2)
      * probgam((b##2/(2*(1-rho##2))),(i+1)/2);
  i=i+1;
  diff=abs(Phi0-vorige);
end;
check=probbnrm(a,b,rho)-probbnrm(a,0,rho)-probbnrm(0,b,rho)+probbnrm(0,0,rho);
print Phi0 check;
```

## C.2 Phi.SAS

```
proc iml;
*===>; a=1; b=2; rho=0.5;

pi=(gamma(0.5))**2;

start Phi0(Phi0,a,b,rho) global(pi);
diff=1; Phi0=0; i=0;
do while (diff>1e-8);
  vorige=Phi0;
  Phi0=Phi0 + ((2*rho)##i *sqrt(1-rho##2))/(4*pi*gamma(i+1))*gamma((i+1)/2)##2
     * probgam((a##2/(2*(1-rho##2))),(i+1)/2)
     * probgam((b##2/(2*(1-rho##2))),(i+1)/2);
  i=i+1;
  diff=abs(Phi0-vorige);
end;
finish;

if a<0 & b<0 then do;
  run Phi0(Phi01,10,10,rho);
  run Phi0(Phi02,-a,10,rho);
  run Phi0(Phi03,10,-b,rho);
  run Phi0(Phi04,-a,-b,rho);
  Phi=Phi01-Phi02-Phi03+Phi04;
end;
if a<0 & b>=0 then do;
  run Phi0(Phi01,10,10,rho);
  run Phi0(Phi02,-a,10,rho);
  run Phi0(Phi03,10,b,-rho);
  run Phi0(Phi04,-a,b,-rho);
  Phi=Phi01-Phi02+Phi03-Phi04;
end;
```

```
if a>=0 & b<0 then do;
  run Phi0(Phi01,10,10,rho);
  run Phi0(Phi02,a,10,-rho);
  run Phi0(Phi03,10,-b,rho);
  run Phi0(Phi04,a,-b,-rho);
  Phi=Phi01+Phi02-Phi03-Phi04;
end;
if a>=0 & b>=0 then do;
  run Phi0(Phi01,10,10,rho);
  run Phi0(Phi02,a,10,-rho);
  run Phi0(Phi03,10,b,-rho);
  run Phi0(Phi04,a,b,rho);
  Phi=Phi01+Phi02+Phi03+Phi04;
end;
check=probbnrm(a,b,rho);
print Phi check;
```

## C.3 BVN.SAS

```
proc iml;
pie=gamma(0.5)##2;

freq={106  90  35  5 ,
      57   73  59  22,
      15   40  57  27,
       2   14  45  99};

x={59.5,69.5,79.5};
y={49.5,59.5,74.5};

n=freq[+];
nfr=freq[,+];
nfc=freq[+,];
nr=nrow(freq); nr1=nr-1; Er=J(nr,1,1); Er1=J(nr1,1,1);
nc=ncol(freq); nc1=nc-1; Ec=J(nc,1,1); Ec1=J(nc1,1,1);
rc=nr*nc;
Cr=J(nr,1,1)@cusum(J(1,nr,1))<=J(1,nr,1)@cusum(J(nr,1,1));
Cc=J(nc,1,1)@cusum(J(1,nc,1))<=J(1,nc,1)@cusum(J(nc,1,1));
C=Cr@Cc; CI=inv(C);
fxy=colvec(freq);
p=C*fxy/n;

XD=x||J(nr1,1,-1);
XXX=inv(XD`*XD)*XD`;
PmX=XD*inv(XD`*XD)*XD`;

YD=y||J(nc1,1,-1);
YYY=inv(YD`*YD)*YD`;
PmY=YD*inv(YD`*YD)*YD`;

IV=cusum(j(rc,1,1)); IM=shape(IV,nr);
```

```
xx=IM[1:nr1,nc]; yy=IM[nr,1:nc1]; xy=IM[1:nr1,1:nc1];
Gmx=J(nr1,rc,0); Gmy=J(nc1,rc,0); Gmxy=J(nr1*nc1,rc,0);
ij=0;
do i=1 to nr1; Gmx[i,xx[i]]=1; end;
do j=1 to nc1; Gmy[j,yy[j]]=1; end;
do i=1 to nr1; do j=1 to nc1;
  ij=ij+1;
  Gmxy[ij,xy[i,j]]=1;
end; end;

start F0(F0,z1,z2,rho,k,l) global(pie);
i=1; diff2=1;
F0= 2**((k+l)/2) * (1-rho**2)**((k+l+1)/2) / (4*pie)
  * gamma((k+1)/2) * gamma((l+1)/2)
  * probgam((z1**2/(2*(1-rho**2))),(k+1)/2)
  * probgam((z2**2/(2*(1-rho**2))),(l+1)/2);
do while (diff2>1e-9);
  vF0=F0;
  F0= F0+2**((k+l)/2)*(1-rho**2)**((k+l+1)/2) / (4*pie)*(2*rho)**i
    * gamma((i+k+1)/2) * gamma((i+l+1)/2) / gamma(i+1)
    * probgam((z1**2/(2*(1-rho**2))),(i+k+1)/2)
    * probgam((z2**2/(2*(1-rho**2))),(i+l+1)/2);
  diff2=abs(vF0-F0);
  i=i+1;
end;
finish;

start F (F,rho,zy,zx,k,l);
  if zx<0 & zy<0 then do;
    run F0(F1,10,10,rho,k,l);
    run F0(F2,-zx,10,rho,k,l);
    run F0(F3,10,-zy,rho,k,l);
    run F0(F4,-zx,-zy,rho,k,l);
    F=F1-F2-F3+F4;
```

```
  end;
  if zx<0 & zy>=0 then do;
    run F0(F1,10,10,rho,k,l);
    run F0(F2,-zx,10,rho,k,l);
    run F0(F3,10,zy,-rho,k,l);  F3=F3*(-1)**k;
    run F0(F4,-zx,zy,-rho,k,l);  F4=F4*(-1)**k;
    F=F1-F2+F3-F4;
  end;
  if zx>=0 & zy<0 then do;
    run F0(F1,10,10,rho,k,l);
    run F0(F2,zx,10,-rho,k,l);  F2=F2*(-1)**l;
    run F0(F3,10,-zy,rho,k,l);
    run F0(F4,zx,-zy,-rho,k,l);  F4=F4*(-1)**l;
    F=F1+F2-F3-F4;
  end;
  if zx>=0 & zy>=0 then do;
    run F0(F1,10,10,rho,k,l);
    run F0(F2,zx,10,-rho,k,l);  F2=F2*(-1)**l;
    run F0(F3,10,zy,-rho,k,l);  F3=F3*(-1)**k;
    run F0(F4,zx,zy,rho,k,l);
    F=F1+F2+F3+F4;
  end;
finish;

start prob (pp,x1,x2,y1,y2,rho);
  pp=probbnrm(x2,y2,rho)-probbnrm(x2,y1,rho)
    -probbnrm(x1,y2,rho)+probbnrm(x1,y1,rho);
finish;

start volume(p,rho,vv,zx,zy,II,JJ) global(nr,nc,nr1,nc1,Er,Ec,CI,pie);
  zx1=-10//zx; zx2=zx//10;
  zy1=-10//zy; zy2=zy//10;

  run prob(ppIJ,zx[II-1],zx[II],zy[JJ-1],zy[JJ],rho);
```

```
run prob(ppIJ1,zx[II-1],0,zy[JJ-1],0,rho);
run prob(ppIJ2,zx[II-1],0,0,zy[JJ],rho);
run prob(ppIJ3,0,zx[II],zy[JJ-1],0,rho);
run prob(ppIJ4,0,zx[II],0,zy[JJ],rho);

run prob(ppI,((zx[II-1])*Ec),((zx[II])*Ec),zy1,zy2,rho);
run prob(ppI1,((zx[II-1])*Ec),(0*Ec),zy1,zy2,rho);
run prob(ppI2,(0*Ec),((zx[II])*Ec),zy1,zy2,rho);
run prob(ppJ,zx1,zx2,((zy[JJ-1])*Er),((zy[JJ])*Er),rho);
run prob(ppJ1,zx1,zx2,((zy[JJ-1])*Er),(0*Er),rho);
run prob(ppJ2,zx1,zx2,(0*Er),((zy[JJ])*Er),rho);

volc1=J(nr,nc,0);volc2=J(nr,nc,0);volc3=J(nr,nc,0);volc4=J(nr,nc,0);

volc1[1:II-1,1:JJ-1]=1;
volc2[1:II-1,JJ+1:nc]=1;
volc3[II+1:nr,1:JJ-1]=1;
volc4[II+1:nr,JJ+1:nc]=1;

volc1[II,1:JJ-1]=(ppI1[1:JJ-1]/ppI[1:JJ-1])`;
volc2[II,JJ+1:nc]=(ppI1[JJ+1:nc]/ppI[JJ+1:nc])`;
volc3[II,1:JJ-1]=(ppI2[1:JJ-1]/ppI[1:JJ-1])`;
volc4[II,JJ+1:nc]=(ppI2[JJ+1:nc]/ppI[JJ+1:nc])`;

volc1[1:II-1,JJ]=(ppJ1[1:II-1]/ppJ[1:II-1]);
volc2[1:II-1,JJ]=(ppJ2[1:II-1]/ppJ[1:II-1]);
volc3[II+1:nr,JJ]=(ppJ1[II+1:nr]/ppJ[II+1:nr]);
volc4[II+1:nr,JJ]=(ppJ2[II+1:nr]/ppJ[II+1:nr]);

volc1[II,JJ]=ppIJ1/ppIJ;
volc2[II,JJ]=ppIJ2/ppIJ;
volc3[II,JJ]=ppIJ3/ppIJ;
volc4[II,JJ]=ppIJ4/ppIJ;
```

```
  v1=colvec(volc1);
  v2=colvec(volc2);
  v3=colvec(volc3);
  v4=colvec(volc4);
  vv=(v1+v4)-(v2+v3);

  vol1=v1`*CI*p;
  vol2=v2`*CI*p;
  vol3=v3`*CI*p;
  vol4=v4`*CI*p;
finish;

start rho (p,rhop,vvp,zxp,zyp,IIp,JJp,drdp) global(pie,CI);
  rhop=0; diffr=1;
  do while (diffr > 1e-10);
    rhov=rhop;
    run volume(p,rhop,vvp,zxp,zyp,IIp,JJp);
    rhop=sin(pie/2*(vvp`*CI*p));
    diffr=sqrt((rhop-rhov)**2);
    drdp=cos(pie/2*(vvp`*CI*p))*pie/2*vvp`*CI;
  end;
finish;

start GGxy(p,rho,zx,zy,Dx,Dy,drdp,GGxy) global(nr1,nc1,rc,Pmx,Pmy,Gmx,Gmy,Gmxy);
  ZZx=zx@J(1,nc1,1);
  ZZy=zy`@J(nr1,1,1);
  dFdzxm=diag(pdf('normal',zx))*probnorm((ZZy-rho*ZZx)/sqrt(1-rho**2));
  dFdzym=probnorm((ZZx-rho*ZZy)/sqrt(1-rho**2))*diag(pdf('normal',zy));
  do i=1 to nr1;
    EEr=J(nr1,nr1,0);
    EEr[i,i]=1;
    tyd=colvec(EEr*dFdzxm);
    if i=1 then dFdzx=tyd;
    else dFdzx=dFdzx||colvec(tyd);
```

```
  end;
  do j=1 to nc1;
    EEc=J(nc1,nc1,0);
    EEc[j,j]=1;
    tyd=colvec(dFdzym*EEc);
    if j=1 then dFdzy=colvec(tyd);
    else dFdzy=dFdzy||colvec(tyd);
  end;
  dFdr=J(nr1,nc1,0); rho2=rho**2;
  do i=1 to nr1; do j=1 to nc1;
    run F(F00,rho,zx[i],zy[j],0,0);
    run F(F20,rho,zx[i],zy[j],2,0);
    run F(F11,rho,zx[i],zy[j],1,1);
    run F(F02,rho,zx[i],zy[j],0,2);
    dFdr[i,j]=rho/(1-rho2)*F00 - rho/((1-rho2)**2)*F20
        + (1+rho2)/((1-rho2)**2)*F11 - rho/((1-rho2)**2)*F02;
  end; end;
  dFdr=colvec(dFdr);
  dzxdp=Pmx*Dx*Gmx;
  dzydp=Pmy*Dy*Gmy;
  GGxy=(dFdzx||dFdzy||dFdr)*(dzxdp//dzydp//drdp) - Gmxy;
finish;

start marginal(px,alphaxp,muxp,sigmaxp,zxp,nr,IIp,Dxp,Gmx,XD,XXX,Pmx,GGxp);
  alphaxp=XXX*probit(px);
  zxp=XD*alphaxp;
  muxp=alphaxp[2]/alphaxp[1];
  sigmaxp=1/alphaxp[1];
  do IIp=1 to nr until (zxp[IIp]>=0); end;
  Dxp=inv(diag(pdf('normal',probit(px))));
  GGxp=diag(pdf('normal',zxp))*Pmx*Dxp*Gmx - Gmx;
finish;

i=0; p0=p; diff1=1;
```

```
do while (diff1 > 1e-8);
  i=i+1; pi=p; p=p0;
  V=(C*diag(CI*pi)*C`-pi*pi`)/n;
  matrixpi=shape(pi,nr);
  pix=matrixpi[1:nr1,nc];
  piy=matrixpi[nr,1:nc1]`;
  pixy=colvec(matrixpi[1:nr1,1:nc1]);
  run marginal(pix,alphaxpi,muxpi,sigmaxpi,zxpi,nr,IIpi,Dxpi,Gmx,XD,XXX,Pmx,GGxpi);
  run marginal(piy,alphaypi,muypi,sigmaypi,zypi,nc,JJpi,Dypi,Gmy,YD,YYY,Pmy,GGypi);
  run rho(pi,rhopi,vvpi,zxpi,zypi,IIpi,JJpi,drdpi);
  run GGxy(pi,rhopi,zxpi,zypi,Dxpi,Dypi,drdpi,GGxypi);
  GGpi=GGxpi//GGypi//GGxypi;
  j=0; diff=1;
  do while (diff > 1e-8);
    j=j+1; pv=p;
    matrixp=shape(p,nr);
    px=matrixp[1:nr1,nc];
    py=matrixp[nr,1:nc1]`;
    pxy=colvec(matrixp[1:nr1,1:nc1]);
    run marginal(px,alphaxp,muxp,sigmaxp,zxp,nr,IIp,Dxp,Gmx,XD,XXX,Pmx,GGxp);
    run marginal(py,alphayp,muyp,sigmayp,zyp,nc,JJp,Dyp,Gmy,YD,YYY,Pmy,GGyp);
    run rho(p,rhop,vvp,zxp,zyp,IIp,JJp,drdp);
    run GGxy(p,rhop,zxp,zyp,Dxp,Dyp,drdp,GGxyp);
    GGp=GGxp//GGyp//GGxyp;
    gx=probnorm(zxp)-px;
    gy=probnorm(zyp)-py;
    gxy=probbnrm(zxp@Ec1,Er1@zyp,rhop)-pxy;
    g=gx//gy//gxy;
    print i j g pi p, matrixp zxp zyp,
              rhopi muxpi sigmaxpi muypi sigmaypi,
              rhop muxp sigmaxp muyp sigmayp;
    p=p-(GGpi*V)`*ginv(GGp*V*GGpi`)*g;
    if i=1 & j=1 then do;
      Wald=g`*ginv(GGp*V*GGp`)*g;
```

```
      GGpV=GGp*V;
      df=trace(GGpV*ginv(GGpV`*GGpV)*GGpV`);
      pvalue=1-probchi(Wald,df);
      discr=wald/n;
      Cov_rho=drdpi*V*drdpi`;
      SE_rho=sqrt(Cov_rho);
    end;
    diff=sqrt((p-pv)`*(p-pv));
  end;
  diff1=sqrt((p-pi)`*(p-pi));
end;
mux=muxp; sigmax=sigmaxp;
muy=muyp; sigmay=sigmayp;
rho=rhop;

Cov_pi=V-(GGpi*V)`*ginv(GGpi*V*GGpi`)*(GGpi*V);

alphax=alphaxp;
Cov_alphax=(XXX*Dxpi*Gmx)*Cov_pi*(XXX*Dxpi*Gmx)`;
Ax=J(2,2,0);
Ax[1,1]=-alphax[2]/((alphax[1])**2);
Ax[1,2]=1/(alphax[1]);
Ax[2,1]=-1/((alphax[1])**2);
Cov_musigx=Ax*Cov_alphax*Ax`;
SE_mux=sqrt(Cov_musigx[1,1]);
SE_sigmax=sqrt(Cov_musigx[2,2]);

alphay=alphayp;
Cov_alphay=(YYY*Dypi*Gmy)*Cov_pi*(YYY*Dypi*Gmy)`;
Ay=J(2,2,0);
Ay[1,1]=-alphay[2]/((alphay[1])**2);
Ay[1,2]=1/(alphay[1]);
Ay[2,1]=-1/((alphay[1])**2);
Cov_musigy=Ay*Cov_alphay*Ay`;
```

```
SE_muy=sqrt(Cov_musigy[1,1]);
SE_sigmay=sqrt(Cov_musigy[2,2]);

print mux SE_mux sigmax SE_sigmax,
      muy SE_muy sigmay SE_sigmay,
      rho SE_rho;

t_rho=rho/SE_rho;
p_rho=(1-probnorm(t_rho))*2;
print t_rho p_rho;

alpha_xy=mux-muy*rho*sigmax/sigmay;
beta_xy=rho*sigmax/sigmay;
alpha_yx=muy-mux*rho*sigmay/sigmax;
beta_yx=rho*sigmay/sigmax;
print alpha_xy beta_xy alpha_yx beta_yx;

exp=shape((CI*pi*n),nr);
pearson=(((freq-exp)##2)/exp)[+];
print n freq exp Pearson Wald;
```

## C.4 BVNSIM.SAS

```
proc iml worksize=30000 symsize=30000;
pie=gamma(0.5)##2;

n=1000;
rep=10;
number=1;
x={8,10,12}; y={45,50,55};
mu={11,48};
sig={9  -16.8,
     -16.8  64};
call eigen(L,H,SIG);
sig12=H*diag(sqrt(L))*H‘;

nr=nrow(x)+1; nr1=nr-1; Er=J(nr,1,1); Er1=J(nr1,1,1);
nc=nrow(y)+1; nc1=nc-1; Ec=J(nc,1,1); Ec1=J(nc1,1,1);
rc=nr*nc;
Cr=J(nr,1,1)@cusum(J(1,nr,1))<=J(1,nr,1)@cusum(J(nr,1,1));
Cc=J(nc,1,1)@cusum(J(1,nc,1))<=J(1,nc,1)@cusum(J(nc,1,1));
C=Cr@Cc; CI=inv(C);

XD=x||J(nr1,1,-1);
XXX=inv(XD‘*XD)*XD‘;
PmX=XD*inv(XD‘*XD)*XD‘;

YD=y||J(nc1,1,-1);
YYY=inv(YD‘*YD)*YD‘;
PmY=YD*inv(YD‘*YD)*YD‘;

IV=cusum(j(rc,1,1)); IM=shape(IV,nr);
xx=IM[1:nr1,nc]; yy=IM[nr,1:nc1]; xy=IM[1:nr1,1:nc1];
Gmx=J(nr1,rc,0); Gmy=J(nc1,rc,0); Gmxy=J(nr1*nc1,rc,0);
ij=0;
```

```
do i=1 to nr1; Gmx[i,xx[i]]=1; end;
do j=1 to nc1; Gmy[j,yy[j]]=1; end;
do i=1 to nr1; do j=1 to nc1;
    ij=ij+1;
    Gmxy[ij,xy[i,j]]=1;
end; end;

/*
*** Begin: Theoretical values ***;
poprho=sig[1,2]/sqrt(sig[1,1]*sig[2,2]);
popzx=((x-mu[1,1])/sqrt(sig[1,1]))//10;
popzy=((y-mu[2,1])/sqrt(sig[2,2]))//10;
poppi=probbnrm((popzx)@J(nc,1,1),J(nr,1,1)@(popzy),poprho);
freq=shape(CI*poppi*n,nr);
fxy=colvec(freq);
p=C*fxy/freq[+];
*** End: Theoretical values ***;
*/

start data;
sp=sig12*rannor(J(2,n,number))+ mu*J(1,n,1);

smu=sp[,+]/n;
ssig=sp*sp`/n-smu*smu`;
D=inv(sqrt(diag(ssig)));
scor=D*ssig*D;
smux=smu[1]; smuy=smu[2];
ssigx=sqrt(ssig[1,1]); ssigy=sqrt(ssig[2,2]);
srho=ssig[1,2]/sqrt(ssig[1,1]*ssig[2,2]);

sp=sp`;
spx=sp[,1];
spy=sp[,2];
f=j(nr,nc,0);
```

```
do k=1 to n;
    t=j(nr,nc,0);
    do III=1 to nr1 until (spx[k] <= x[III]); end;
    do JJJ=1 to nc1 until (spy[k] <= y[JJJ]); end;
    t[III,JJJ]=1;
    f=f+t;
end;
freq=f;
freq=freq<>1e-6;
fxy=colvec(freq);
p=C*fxy/freq[+];
finish;

start F0(F0,z1,z2,rho,k,l) global(pie);
i=1; diff2=1;
F0= 2**((k+l)/2) * (1-rho**2)**((k+l+1)/2) / (4*pie)
    * gamma((k+1)/2) * gamma((l+1)/2)
    * probgam((z1**2/(2*(1-rho**2))),(k+1)/2)
    * probgam((z2**2/(2*(1-rho**2))),(l+1)/2);
do while (diff2>1e-9);
    vF0=F0;
    F0= F0+2**((k+l)/2) *(1-rho**2)**((k+l+1)/2) / (4*pie) * (2*rho)**i
        * gamma((i+k+1)/2) * gamma((i+l+1)/2) / gamma(i+1)
        * probgam((z1**2/(2*(1-rho**2))),(i+k+1)/2)
        * probgam((z2**2/(2*(1-rho**2))),(i+l+1)/2);
    diff2=abs(vF0-F0);
    i=i+1;
end;
finish;

start F (F,rho,zy,zx,k,l);
    if zx<0 & zy<0 then do;
        run F0(F1,10,10,rho,k,l);
        run F0(F2,-zx,10,rho,k,l);
```

```
        run F0(F3,10,-zy,rho,k,l);
        run F0(F4,-zx,-zy,rho,k,l);
        F=F1-F2-F3+F4;
    end;
    if zx<0 & zy>=0 then do;
        run F0(F1,10,10,rho,k,l);
        run F0(F2,-zx,10,rho,k,l);
        run F0(F3,10,zy,-rho,k,l);     F3=F3*(-1)**k;
        run F0(F4,-zx,zy,-rho,k,l);     F4=F4*(-1)**k;
        F=F1-F2+F3-F4;
    end;
    if zx>=0 & zy<0 then do;
        run F0(F1,10,10,rho,k,l);
        run F0(F2,zx,10,-rho,k,l);     F2=F2*(-1)**l;
        run F0(F3,10,-zy,rho,k,l);
        run F0(F4,zx,-zy,-rho,k,l);     F4=F4*(-1)**l;
        F=F1+F2-F3-F4;
    end;
    if zx>=0 & zy>=0 then do;
        run F0(F1,10,10,rho,k,l);
        run F0(F2,zx,10,-rho,k,l);     F2=F2*(-1)**l;
        run F0(F3,10,zy,-rho,k,l);     F3=F3*(-1)**k;
        run F0(F4,zx,zy,rho,k,l);
        F=F1+F2+F3+F4;
    end;
finish;

start prob (pp,x1,x2,y1,y2,rho);
    pp=probbnrm(x2,y2,rho)-probbnrm(x2,y1,rho)
      -probbnrm(x1,y2,rho)+probbnrm(x1,y1,rho);
finish;

start volume(p,rho,vv,zx,zy,II,JJ) global(nr,nc,nr1,nc1,Er,Ec,CI,pie);
    zx1=-10//zx; zx2=zx//10;
```

```
zy1=-10//zy; zy2=zy//10;

run prob(ppIJ,zx[II-1],zx[II],zy[JJ-1],zy[JJ],rho);
run prob(ppIJ1,zx[II-1],0,zy[JJ-1],0,rho);
run prob(ppIJ2,zx[II-1],0,0,zy[JJ],rho);
run prob(ppIJ3,0,zx[II],zy[JJ-1],0,rho);
run prob(ppIJ4,0,zx[II],0,zy[JJ],rho);

run prob(ppI,((zx[II-1])*Ec),((zx[II])*Ec),zy1,zy2,rho);
run prob(ppI1,((zx[II-1])*Ec),(0*Ec),zy1,zy2,rho);
run prob(ppI2,(0*Ec),((zx[II])*Ec),zy1,zy2,rho);
run prob(ppJ,zx1,zx2,((zy[JJ-1])*Er),((zy[JJ])*Er),rho);
run prob(ppJ1,zx1,zx2,((zy[JJ-1])*Er),(0*Er),rho);
run prob(ppJ2,zx1,zx2,(0*Er),((zy[JJ])*Er),rho);

volc1=J(nr,nc,0);volc2=J(nr,nc,0);volc3=J(nr,nc,0);volc4=J(nr,nc,0);

volc1[1:II-1,1:JJ-1]=1;
volc2[1:II-1,JJ+1:nc]=1;
volc3[II+1:nr,1:JJ-1]=1;
volc4[II+1:nr,JJ+1:nc]=1;

volc1[II,1:JJ-1]=(ppI1[1:JJ-1]/ppI[1:JJ-1])`;
volc2[II,JJ+1:nc]=(ppI1[JJ+1:nc]/ppI[JJ+1:nc])`;
volc3[II,1:JJ-1]=(ppI2[1:JJ-1]/ppI[1:JJ-1])`;
volc4[II,JJ+1:nc]=(ppI2[JJ+1:nc]/ppI[JJ+1:nc])`;

volc1[1:II-1,JJ]=(ppJ1[1:II-1]/ppJ[1:II-1]);
volc2[1:II-1,JJ]=(ppJ2[1:II-1]/ppJ[1:II-1]);
volc3[II+1:nr,JJ]=(ppJ1[II+1:nr]/ppJ[II+1:nr]);
volc4[II+1:nr,JJ]=(ppJ2[II+1:nr]/ppJ[II+1:nr]);

volc1[II,JJ]=ppIJ1/ppIJ;
volc2[II,JJ]=ppIJ2/ppIJ;
```

```
    volc3[II,JJ]=ppIJ3/ppIJ;
    volc4[II,JJ]=ppIJ4/ppIJ;

    v1=colvec(volc1);
    v2=colvec(volc2);
    v3=colvec(volc3);
    v4=colvec(volc4);
    vv=(v1+v4)-(v2+v3);

    vol1=v1`*CI*p;
    vol2=v2`*CI*p;
    vol3=v3`*CI*p;
    vol4=v4`*CI*p;
finish;

start rho (p,rhop,vvp,zxp,zyp,IIp,JJp,drdp) global(pie,CI);
    i=0;
    rhop=0; diffr=1;
    do while ((diffr > 1e-10) & (i<100));
    i=i+1;
        rhov=rhop;
        run volume(p,rhop,vvp,zxp,zyp,IIp,JJp);
        rhop=sin(pie/2*(vvp`*CI*p));
        diffr=sqrt((rhop-rhov)**2);
        drdp=cos(pie/2*(vvp`*CI*p))*pie/2*vvp`*CI;
    end;
finish;

start GGxy(p,rho,zx,zy,Dx,Dy,drdp,GGxy) global(nr1,nc1,rc,Pmx,Pmy,Gmx,Gmy,Gmxy);
    ZZx=zx@J(1,nc1,1);
    ZZy=zy`@J(nr1,1,1);
    dFdzxm=diag(pdf('normal',zx))*probnorm((ZZy-rho*ZZx)/sqrt(1-rho**2));
    dFdzym=probnorm((ZZx-rho*ZZy)/sqrt(1-rho**2))*diag(pdf('normal',zy));
    do i=1 to nr1;
```

```
        EEr=J(nr1,nr1,0);
        EEr[i,i]=1;
        tyd=colvec(EEr*dFdzxm);
        if i=1 then dFdzx=tyd;
        else dFdzx=dFdzx||colvec(tyd);
    end;
    do j=1 to nc1;
        EEc=J(nc1,nc1,0);
        EEc[j,j]=1;
        tyd=colvec(dFdzym*EEc);
        if j=1 then dFdzy=colvec(tyd);
        else dFdzy=dFdzy||colvec(tyd);
    end;
    dFdr=J(nr1,nc1,0); rho2=rho**2;
    do i=1 to nr1; do j=1 to nc1;
        run F(F00,rho,zx[i],zy[j],0,0);
        run F(F20,rho,zx[i],zy[j],2,0);
        run F(F11,rho,zx[i],zy[j],1,1);
        run F(F02,rho,zx[i],zy[j],0,2);
        dFdr[i,j]=rho/(1-rho2)*F00 - rho/((1-rho2)**2)*F20
                + (1+rho2)/((1-rho2)**2)*F11 - rho/((1-rho2)**2)*F02;
    end; end;
    dFdr=colvec(dFdr);
    dzxdp=Pmx*Dx*Gmx;
    dzydp=Pmy*Dy*Gmy;
    GGxy=(dFdzx||dFdzy||dFdr)*(dzxdp//dzydp//drdp) - Gmxy;
finish;

start marginal(px,alphaxp,muxp,sigmaxp,zxp,nr,IIp,Dxp,Gmx,XD,XXX,Pmx,GGxp);
    alphaxp=XXX*probit(px);
    zxp=XD*alphaxp;
    muxp=alphaxp[2]/alphaxp[1];
    sigmaxp=1/alphaxp[1];
    do IIp=1 to nr until (zxp[IIp]>=0); end;
```

```
    Dxp=inv(diag(pdf('normal',probit(px))));
    GGxp=diag(pdf('normal',zxp))*Pmx*Dxp*Gmx - Gmx;
finish;

start fit;
i=0; p0=p; diff1=1;
do while (diff1 > 1e-8);
    i=i+1; pi=p; p=p0;
    V=(C*diag(CI*pi)*C`-pi*pi`)/n;
    matrixpi=shape(pi,nr);
    pix=matrixpi[1:nr1,nc];
    piy=matrixpi[nr,1:nc1]`;
    pixy=colvec(matrixpi[1:nr1,1:nc1]);
    run marginal(pix,alphaxpi,muxpi,sigmaxpi,zxpi,nr,IIpi,Dxpi,Gmx,XD,XXX,Pmx,GGxpi);
    run marginal(piy,alphaypi,muypi,sigmaypi,zypi,nc,JJpi,Dypi,Gmy,YD,YYY,Pmy,GGypi);
    run rho(pi,rhopi,vvpi,zxpi,zypi,IIpi,JJpi,drdpi);
    run GGxy(pi,rhopi,zxpi,zypi,Dxpi,Dypi,drdpi,GGxypi);
    GGpi=GGxpi//GGypi//GGxypi;
    j=0; diff=1;
    do while (diff > 1e-8);
        j=j+1; pv=p;
        matrixp=shape(p,nr);
        px=matrixp[1:nr1,nc];
        py=matrixp[nr,1:nc1]`;
        pxy=colvec(matrixp[1:nr1,1:nc1]);
        run marginal(px,alphaxp,muxp,sigmaxp,zxp,nr,IIp,Dxp,Gmx,XD,XXX,Pmx,GGxp);
        run marginal(py,alphayp,muyp,sigmayp,zyp,nc,JJp,Dyp,Gmy,YD,YYY,Pmy,GGyp);
        run rho(p,rhop,vvp,zxp,zyp,IIp,JJp,drdp);
        run GGxy(p,rhop,zxp,zyp,Dxp,Dyp,drdp,GGxyp);
        GGp=GGxp//GGyp//GGxyp;
        gx=probnorm(zxp)-px;
        gy=probnorm(zyp)-py;
        gxy=probbnrm(zxp@Ec1,Er1@zyp,rhop)-pxy;
        g=gx//gy//gxy;
```

```
        *print r i j g pi p,
              rhopi muxpi sigmaxpi muypi sigmaypi,
              rhop muxp sigmaxp muyp sigmayp;
        p=p-(GGpi*V)`*ginv(GGp*V*GGpi`)*g;
        if i=1 & j=1 then do;
            Wald=g`*ginv(GGp*V*GGp`)*g;
            GGpV=GGp*V;
            df=trace(GGpV*ginv(GGpV`*GGpV)*GGpV`);
            pvalue=1-probchi(Wald,df);
            discr=wald/n;
            Cov_rho=drdpi*V*drdpi`;
            SE_rho=sqrt(Cov_rho);
        end;
        diff=sqrt((p-pv)`*(p-pv));
    end;
    diff1=sqrt((p-pi)`*(p-pi));
end;
mux=muxp; sigmax=sigmaxp;
muy=muyp; sigmay=sigmayp;
rho=rhop;

Cov_pi=V-(GGpi*V)`*ginv(GGpi*V*GGpi`)*(GGpi*V);

alpha_xy=mux-muy*rho*sigmax/sigmay;
beta_xy=rho*sigmax/sigmay;
alpha_yx=muy-mux*rho*sigmay/sigmax;
beta_yx=rho*sigmay/sigmax;

alphax=alphaxp;
Cov_alphax=(XXX*Dxpi*Gmx)*Cov_pi*(XXX*Dxpi*Gmx)`;
Ax=J(2,2,0);
Ax[1,1]=-alphax[2]/((alphax[1])**2);
Ax[1,2]=1/(alphax[1]);
Ax[2,1]=-1/((alphax[1])**2);
```

```
Cov_musigx=Ax*Cov_alphax*Ax`;
SE_mux=sqrt(Cov_musigx[1,1]);
SE_sigmax=sqrt(Cov_musigx[2,2]);

alphay=alphayp;
Cov_alphay=(YYY*Dypi*Gmy)*Cov_pi*(YYY*Dypi*Gmy)`;
Ay=J(2,2,0);
Ay[1,1]=-alphay[2]/((alphay[1])**2);
Ay[1,2]=1/(alphay[1]);
Ay[2,1]=-1/((alphay[1])**2);
Cov_musigy=Ay*Cov_alphay*Ay`;
SE_muy=sqrt(Cov_musigy[1,1]);
SE_sigmay=sqrt(Cov_musigy[2,2]);
exp=shape((CI*pi*n),nr);
Pearson=(((freq-exp)##2)/exp)[+];

print r freq exp matrixp,
      mux SE_mux sigmax SE_sigmax,
      muy SE_muy sigmay SE_sigmay,
      rho SE_rho;
finish;

start write;
stats[r,1]=number;
stats[r,2]=i;
stats[r,3]=j;
stats[r,4]=smux;
stats[r,5]=ssigx;
stats[r,6]=smuy;
stats[r,7]=ssigy;
stats[r,8]=srho;
stats[r,9]=mux;
stats[r,10]=SE_mux;
stats[r,11]=sigmax;
```

```
stats[r,12]=SE_sigmax;
stats[r,13]=muy;
stats[r,14]=SE_muy;
stats[r,15]=sigmay;
stats[r,16]=SE_sigmay;
stats[r,17]=rho;
stats[r,18]=SE_rho;
stats[r,19]=Pearson;
stats[r,20]=Wald;
finish;

stats=J(rep,20,0);
do r=1 to rep;
     run data;
    run fit;
     run write;
    number=number+1;
end;

create a from stats [colname={'number' 'i' 'j' 'smux' 'ssigx' 'smuy' 'ssigy'
'srho' 'mux' 'SE_mux' 'sigmax' 'SE_sigmax' 'muy' 'SE_muy' 'sigmay' 'SE_sigmay'
'rho' 'SE_rho' 'Pearson' 'Wald'}];
append from stats;

quit;

proc means data=a mean std p5 p50 p95;
run;
```