LINDA LEE HO

ANÁLISE DE CONTAGENS MULTIVARIADAS

Tese apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Doutor em Engenharia.

SÃO PAULO
1995

LINDA LEE HO

ANÁLISE DE CONTAGENS MULTIVARIADAS

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para obtenção
do título de Doutor em Engenharia.

Área de Concentração:
Engenharia de Produção.

Orientador:
Prof. Dr. Pedro Luiz O. Costa Neto.

Coorientador:
Prof. Dr. Julio da Motta Singer.

São Paulo
1995

Pad. Nº 234392l
29/02/96.
OK

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA

TERMO DE JULGAMENTO
DE
DEFESA DE TESE DE DOUTORAMENTO

Aos 15 *dias do mês de* setembro *de* 1995, *às* 16:30 *horas,* no Departamento de Engenharia de Produção *da Escola Politécnica da Universidade de São Paulo, presente a Comissão Julgadora, integrada pelos Senhores Professores* Drs. Júlio da Motta Singer, Co-Orientador da candidata, Pedro Rodrigues Bueno Neto, Clóvis de Araújo Peres, Cicília Yuko Wada e Dalton Francisco de Andrade Nº Func. Co-Or. 504998 *iniciou-se a Defesa de Tese de Doutoramento do Senhor* **LINDA LEE HO** Nº USP 1429952

Título da Tese "Análise de Contagens Multivariadas"

Concluída a argüição, procedeu-se ao julgamento na forma regulamentar, tendo a Comissão Julgadora atribuído ao candidato as seguintes notas:

*Prof.Dr.Júlio da Motta Singer	(9,5)	(nove e meio)
Prof.Dr.Pedro Rodrigues Bueno Neto	(95)	(nove e meio)
Prof.Dr.Clóvis de Araújo Peres	(9.5)	(nove e meio)
Profa.Dra.Cicília Yuko Wada	(9.5)	(nove e meio)
Prof.Dr.Dalton Francisco de Andrade	(9.5)	(nove e meio)

Para constar, é lavrado o presente termo, que vai assinado pela Comissão Julgadora e pelo Secretário da Seção de Pós-Graduação.

São Paulo, 15 *de* setembro *de* 1995.

Presidente [signature]

[signatures]

Secretário Mara Fátima de Jesus Luz Sanches [signature]

Observações Orientador: Prof. Dr. Pedro Luiz de Oliveira Costa Neto
Nº Func. Or. 127302

Homologada pela C.P.G. em reunião realizada a 18/09/1995

BC

FICHA CATALOGRÁFICA

Ho, Linda Lee
Análise de contagens multivariadas. São Paulo, 1995.
109p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Contagem multivariada 2. Regressão Poisson I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

AGRADECIMENTOS

Tive o privilégio de realizar este trabalho sob a orientação dos professores Pedro Luiz de O. Costa Neto e Julio da Motta Singer. Quero agradecê-los pela orientação precisa, pela paciência e atenção com que sempre me atenderam, pelas minuciosas leituras e pelas sugestões dadas. Meus agradecimentos também para:

- os Professores Clóvis de Araújo Peres e Pedro Rodrigues Bueno Neto pelas sugestões dadas no Exame de Qualificação,
- a Escola Politécnica pela oportunidade do desenvolvimento da pesquisa,
- o Departamento de Engenharia de Produção e o Instituto de Matemática e Estatística pela disponibilização de equipamentos, software e hardware,
- as amigas Sílvia e Cris da Biblioteca Central da Escola Politécnica pela revisão nas referências bibliográficas,
- a amiga Vera da Biblioteca do Departamento da Engenharia de Produção, pelo esforço em conseguir as referências necessárias à realização deste trabalho,
- ao amigo Carlos Hugo, por conceder um conjunto de dados sem o qual não seria possível apresentar um exemplo real,
- aos amigos do Departamento de Engenharia de Produção da EPUSP e do Departamento de Estatística do IME, pelo incentivo continuado durante o desenvolvimento do tema.

Sumário

Lista de Tabelas

Lista de abreviaturas

AIC - Critério de Informação de Akaike

DZ - Double zero

EP - Even Point

FGM - Função Geradora dos Momentos

FGP - Função Geradora de Probabilidade

MM - Métodos dos Momentos

MQG - Mínimos Quadrados Generalizados

MQQ - Mínimo Qui-quadrado

MV - Máxima Verossimilhança

Resumo

Este trabalho apresenta uma análise estatística de contagens multivariadas proveniente de várias populações através de modelos de regressão . Foram considerados casos onde os vetores respostas obedeçam às distribuições Poisson Multivariada e Poisson Log-Normal Multivariada. Esta distribuição admite correlação de ambos sinais entre componentes do vetor resposta, enquanto que as distribuições mais usuais para dados de contagens (como a Poisson Multivariada) admitem apenas correlação positiva entre as componentes do vetor resposta. São discutidos métodos de estimação e testes de hipóteses sobre os parâmetros do modelo para o caso bivariado. Estes modelos de regressão foram aplicados a um conjunto de dados referentes a contagens de dois tipos de defeitos em 100 gramas de fibras têxteis de quatro máquinas craqueadeiras, sendo duas de um fabricante e as outras de um segundo fabricante. Os resultados obtidos nos diferentes modelos de regressão foram comparados. Para estudar o comportamento das estimativas dos parâmetros de uma distribuição Poisson Log-Normal, amostras foram simuladas segundo esta distribuição .

Abstract

Regression models are presented to analyse multivariate counts from many populations. Due to the random vector characteristics, we consider two classes of probability models: Multivariate Poisson distribution and Multivariate Poisson Log-Normal distribution. The last distribution admits negative and positive correlations between two components of a random vector under study, while other distributions (as Multivariate Poisson) admit only positive correlation. Estimation methods and test of hypotheses on the parameters in bivariate case are discussed. The proposed techniques are illustrated by numerical examples, considering counts of two types of defects in 100g of textile fibers produced by four machines, two from one manufacturer and the other two from another one. The results from different regression models are compared.

The empirical distribution of Poisson Log-Normal parameter estimations are studied by simulated samples.

Capítulo 1

Apresentação

1.1 Introdução

Em função da atual situação econômica e com a necessidade de tornarem seus produtos cada vez mais competitivos, as empresas, de modo geral, estão efetuando controle cada vez maior sobre eles. Para um determinado produto, geralmente são controlados vários aspectos (variáveis) e para verificar se o processo de fabricação está sob controle estatístico, as empresas utilizam entre outros instrumentos, gráficos de controle da média e da amplitude de cada variável individualmente. Se os dados observados de uma determinada variável estiverem dentro dos limites estabelecidos, o processo está sob controle estatístico. Ao usar este procedimento, supõe-se que as variáveis são independentes, o que às vezes não é verdade e nestes casos as empresas podem estar tomando decisões erradas.

Em casos de variáveis que obedeçam a uma distribuição Normal Multivariada, métodos que avaliam duas ou mais variáveis simultaneamente são relativamente comuns. Entre esses métodos, alguns utilizam a estatística T^2 de Hotelling [ALT (1986)]. De modo geral, esses métodos estatísticos multivariados são desenvolvidos para variáveis contínuas e muitas vezes sua utilização pode ser inadequada quando as variáveis resultam de contagens.

Alguns exemplos de contagens multivariadas discutidos na literatura são:

- número de acidentes sofridos por ferroviários ingleses em 2 períodos distintos [AR-

BOUS; KERRICH (1951)].

- número de bactérias de vários tipos em salas esterilizadas [AITCHISON; HO (1989)].
- número de defeitos na superfície e número de defeitos internos em lentes [AITCHISON; HO (1989)].

Para modelar este tipo de dados existe uma vasta classe de distribuições de contagens multivariadas; entre elas destacam-se as distribuições Poisson Multivariada e Poisson Composta Multivariada.

M'KENDRICK (1926) apresentou uma primeira versão da distribuição Poisson Bivariada obtida a partir de equações diferenciais. Outros autores como MARITZ (1952), HAMDAN; AL-BAYYATI (1969) ou MARSHALL; OLKIN (1985) apresentaram outras maneiras de definir a distribuição Poisson Bivariada e KAWAMURA (1979) as generalizou para o caso de dimensão maior que dois. HOLGATE (1964) e LOUKAS et al. (1986) apresentaram diferentes métodos de estimação dos parâmetros dessa distribuição.

ARBOUS; KERRICH (1951) utilizaram distribuições Poisson Compostas para modelar contagens multivariadas. KOCHERLAKOTA (1988) apresentou um tratamento unificado para esta classe de distribuições, considerando o caso bivariado e várias opções de composição de variáveis contínuas ou discretas univariadas com a distribuição Poisson. A distribuição Poisson Multivariada e as distribuições Poisson Compostas Multivariadas citadas em KOCHERLAKOTA (1988) apresentam uma característica em comum: admitem apenas correlação positiva entre as componentes. Contagens multivariadas que apresentam correlações negativas entre as componentes não podem ser modeladas por nenhuma das duas distribuições citadas. AITCHISON; HO (1989) apresentam um caso particular de uma distribuição Poisson Composta Multivariada, compondo uma variável Log-Normal Multivariada com distribuições Poisson independentes, denominando-a distribuição Poisson Log-Normal Multivariada, aplicando-a a alguns conjuntos de dados e fazendo comparações dos resultados obtidos sob esta distribuição com aqueles obtidos a partir da distribuição Poisson Bivariada. A principal vantagem da distribuição Poisson Log-Normal Multivariada sobre as anteriores é admitir correlações tanto positivas quanto negativas. Apesar desta vantagem, apresenta o inconveniente de não ter uma forma explícita para sua função de probabilidade, requerendo o uso de aproximações polinomiais para cálculo das probabilidades. Além disto, as estimativas de máxima verossimilhança dos seus parâmetros só podem ser obtidos através de métodos iterativos.

Os trabalhos anteriormente mencionados referem-se à análise estatística de contagens multivariadas provenientes de uma única população. Para ilustrar o tipo de problema que será enfocado neste trabalho, considere-se o seguinte exemplo de contagens multivariadas envolvendo várias populações.

Exemplo 1.1 - Nas **Tabelas 1.1 a 1.4** estão resumidos os dados fornecidos por uma indústria têxtil referentes a contagens de dois tipos de defeitos (D_1 e D_2) por 100 g de fibras têxteis em quatro máquinas craqueadeiras, aqui denominadas de A_1, A_2, B_1 e B_2, com A_j, j = 1, 2 produzidas pelo fabricante A e as demais pelo fabricante B.

Tabela 1.1 - Frequência de amostras com defeitos do tipo D_1 e D_2 da Máquina A_1.

Nº de defeitos do tipo D_1	Nº de defeitos do tipo D_2 0	1	2	3	4	5	6	7	...	Total
0	13	26	22	16	3	2	.	.	...	82
1	3	9	12	7	5	2	1	.	...	39
2	.	2	1	.	.	1	.	.	...	4
3	.	1	.	2	.	.	.	1	...	4
4	.	.	1	.	.	.	.	.	...	1
...	...	...	...	...	...	...	...	...	...	...
Total	16	38	38	25	8	5	1	1	...	130

Tabela 1.2 - Frequência de amostras com defeitos do tipo D_1 e D_2 da Máquina A_2.

Nº de defeitos do tipo D_1	Nº de defeitos do tipo D_2 0	1	2	3	4	5	6	7	...	Total
0	15	14	11	10	3	1	.	.	...	54
1	2	12	6	7	5	3	.	1	...	36
2	.	1	2	1	.	2	.	.	...	6
3	.	.	.	.	1	1	.	1	...	3
4	.	.	.	1	.	.	.	.	...	1
...	...	...	...	...	...	...	...	...	...	...
Total	17	27	19	19	9	7	0	2	...	100

Tabela 1.3 - Frequência de amostras com defeitos do tipo D_1 e D_2 da Máquina B_1.

Nº de defeitos do tipo D_1	Nº de defeitos do tipo D_2							
	0	1	2	3	4	5	...	Total
0	11	35	21	13	5	1	...	86
1	1	5	3	1	1	.	...	11
2	.	1	.	2	.	.	...	3
...	...	...	...	...	...	...	...	...
Total	12	41	24	16	6	1	...	100

Tabela 1.4 - Frequência de amostras com defeitos do tipo D_1 e D_2 da Máquina B_2.

Nº de defeitos do tipo D_1	Nº de defeitos do tipo D_2							
	0	1	2	3	4	5	...	Total
0	20	35	17	7	4	1	...	84
1	4	7	3	2	.	.	...	16
...	...	...	...	...	...	...	...	...
Total	24	42	20	9	4	1	...	100

No **Exemplo 1.1**, suponha que haja interesse em avaliar se o número de defeitos do tipo D_1 está associado ao número de defeitos do tipo D_2. Caso haja associação, pode haver interesse em verificar se ela é positiva ou negativa. A associação é positiva se o número de defeitos do tipo D_1 tende a crescer com o aumento do número de defeitos do tipo D_2 e é negativa se o número de defeitos do tipo D_1 tende a crescer com a diminuição do número de defeitos do tipo D_2.

Quando n elementos são classificados segundo os atributos A e B, com respectivamente r e s categorias de respostas possíveis e considerados como uma amostra aleatória simples de uma população multinomial com probabilidades p_{ij}, $i = 1, \ldots, r$ e $j = 1, \ldots, s$, $\sum_{i,j} p_{ij} = 1$, um teste bastante conhecido para avaliar a associação entre estes atributos é baseado na estatística Qui-quadrado de Pearson. Porém, para verificar se há associação entre os números de defeitos do tipo D_1 e D_2, este teste pode ser inadequado por vários motivos: a presença de caselas com frequência igual a zero é relativamente alta, requerendo redução da dimensão da tabela para aplicá-lo, acarretando consequentemente, uma perda de informação. Além disso, em muitas situações também há interesse em se fazerem inferências sobre o total de contagens (n), o que não é possível sob o modelo multinomial, sob o qual n é considerado fixo.

A motivação em avaliar esta associação está vinculada ao fato de que com o conhecimento do comportamento de um dos tipos de defeitos é possível prever o comportamento do outro tipo de defeito. Atualmente, com um mercado altamente competitivo, as empresas de um modo geral, procuram meios alternativos para obter uma redução de custo e/ou de tempo de produção. A constatação de uma associação como a descrita acima poderia resultar uma simplificação do processo, e consequentemente uma economia de recursos, uma vez que não haveria necessidade de coletar dados dos dois tipos de defeitos; nesse caso, bastaria que se monitorasse apenas um deles, provavelmente o de menor custo e/ou de maior facilidade.

Paralelamente, pode haver interesse em avaliar se a distribuição das respostas é homogênea nas diferentes máquinas e/ou nos diferentes fabricantes, em termos de associação e/ou em termos de taxas médias de defeitos. O uso de uma análise de variância multivariada neste tipo de dados pode ser inadequada, uma vez que ela é baseada na suposição de uma distribuição Normal Multivariada para o vetor resposta e numa mesma estrutura de covariância em todas as populações. A análise enfocada neste trabalho também poderia gerar uma economia de recursos, pois poderia fornecer subsídios para um planejamento amostral mais adequado. Por exemplo, retirando uma amostra das máquinas de cada fabricante ao invés de coletar dados de todas elas.

1.2 Objetivos e estrutura do trabalho

Utilizando modelos de regressão, trabalhos foram desenvolvidos para análise estatística de contagens univariadas provenientes de mais de uma população; entre eles, FROME et al. (1973) consideraram um modelo de Regressão Poisson para descrever a variável resposta (no caso uma variável univariada com distribuição Poisson) em função de um conjunto de variáveis preditoras. FROME et al. (1973) e KOCH et al. (1986) apresentaram alguns métodos de estimação dos parâmetros; entre eles destacam-se: máxima verossimilhança (**MV**), mínimos quadrados generalizados (**MQG**) e mínimo qui-quadrado (**MQQ**).

Este trabalho enfocará a análise estatística de contagens multivariadas provenientes de várias populações, dando maiores detalhes para o caso de contagens bivariadas. Serão também utilizados modelos de regressão para descrever as relações entre as contagens multivariadas e as várias populações, apresentando um método de estimação e testes de hipóteses sobre os parâmetros do modelo de regressão.

Na Seção 1.3 estão descritas algumas das principais distribuições utilizadas para modelar contagens multivariadas com as respectivas propriedades: Poisson Multivariada e Poisson Composta Multivariada. No Capítulo 2, estão apresentados modelos de regressão para o caso de a distribuição geradora dos dados ser Poisson Log-Normal Multivariada e Poisson Multivariada. A escolha da primeira distribuição baseia-se na possibilidade de esta distribuição admitir correlações de ambos os sinais. Estão discutidos métodos de estimação e testes de hipótese sobre os parâmetros dos modelos. Além disto, estão descritos os problemas computacionais envolvidos na estimação de seus parâmetros. No Capítulo 3 estão apresentados exemplos de análise de contagens multivariadas utilizando os dados do **Exemplo 1.1.** Os resultados apresentados foram obtidos considerando as distribuições de probabilidade Poisson Log-Normal Bivariada, Poisson Bivariada e Poisson independentes como modelos do mecanismo de geração dos dados. No Capítulo 4 estão apresentadas as distribuições empíricas das estimativas dos parâmetros de distribuições Poisson Log-Normal Bivariada obtidas a partir de amostras simuladas. Finalmente no último capítulo estão algumas sugestões para futuras pesquisas.

1.3 Alguns modelos para contagens multivariadas

Nesta seção estão apresentados dois modelos probabilísticos para contagens multivariadas: distribuição Poisson Multivariada e distribuição Poisson Composta Multivariada. Para defini-los são necessárias algumas distribuições correlatas: Bernoulli Multivariada e Binomial Multivariada.

Considere a situação onde há interesse em estudar duas variáveis simultaneamente; por exemplo, número de defeitos mecânicos e número de defeitos elétricos num certo período em determinado equipamento.

Sejam

$$Z_1 = \begin{cases} 1 & \text{se ocorreu defeito elétrico} \\ 0 & \text{caso contrário} \end{cases}$$

e

$$Z_2 = \begin{cases} 1 & \text{se ocorreu defeito mecânico} \\ 0 & \text{caso contrário} \end{cases}$$

variáveis aleatórias com distribuição de Bernoulli.

O resultado de uma única observação simultânea das duas variáveis pode ser disposto na seguinte tabela:

Defeito elétrico	Defeito mecânico ocorreu	não ocorreu	
ocorreu	Z_1=1,Z_2=1	Z_1=1,Z_2=0	Z_1=1
não ocorreu	Z_1=0,Z_2=1	Z_1=0,Z_2=0	Z_1=0
	Z_2=1	Z_2=0	

Neste caso, o vetor aleatório $\mathbf{Z}=(Z_1, Z_2)'$ pode assumir quatro valores $(1,1)$, $(1,0)$, $(0,1)$ e $(0,0)$ com probabilidades p_{11}, p_{10}, p_{01} e p_{00} respectivamente, $\sum_{i,j} p_{ij} = 1$. Nestas condições o vetor $\mathbf{Z}$ tem distribuição Bernoulli Bivariada, cuja função de probabilidade é

$$P(\mathbf{Z}) = P(Z_1 = i_1, Z_2 = i_2) = p_{i_1 i_2} \tag{1.1}$$

onde i_j assume valor 1 ou 0, j=1, 2. A expressão (1.1) reescrita segundo a notação de MARSHALL; OLKIN (1985) e outros autores é

$$P(\mathbf{Z}) = p_{11}^{u} p_{10}^{Z_1 - u} p_{01}^{Z_2 - u} p_{00}^{1 - Z_1 - Z_2 - u} \tag{1.2}$$

onde $u = min(Z_1, Z_2)$.

Suponha que esta variável bidimensional tenha sido observada n vezes e sejam:

- U_{11} o número de vezes em que ocorreram os dois tipos de defeitos,
- U_{10} o número de vezes em que ocorreu apenas o defeito elétrico,
- U_{01} o número de vezes em que ocorreu apenas o defeito mecânico,
- U_{00} o número de vezes em que não ocorreram defeitos,
- $Y_1=U_{11}+U_{10}$ o número total de ocorrências de defeito elétrico,
- $Y_2=U_{11}+U_{01}$ o número total de ocorrências de defeito mecânico.

Nestas condições, o vetor aleatório bidimensional $\mathbf{Y}=(Y_1, Y_2)'$ tem distribuição Binomial Bivariada (MARSHALL; OLKIN, 1985), cuja função de probabilidade é

$$P(\mathbf{Y}) = \sum_{\mathbf{U}\in\mathbf{A}} \frac{n!}{U_{11}!U_{01}!U_{10}!U_{00}!} p_{11}^{U_{11}} p_{01}^{U_{01}} p_{10}^{U_{10}} p_{00}^{U_{00}} \tag{1.3}$$

onde $\mathbf{U}=(U_{00}, U_{01}, U_{10}, U_{11})'$ e $\mathbf{A}=\{\mathbf{U} \in \mathbf{N}^4\text{: } U_{11}+U_{10}=Y_1 \text{ e } U_{11}+U_{01}=Y_2\}$.

Utilizando a notação de MARSHALL; OLKIN (1985), a expressão (1.3) pode ser reescrita como

$$P(Y_1 = y_1, Y_2 = y_2) =$$
$$= \sum_{u=0}^{min(y_1,y_2)} \binom{n}{u, y_1 - u, y_2 - u, n - y_1 - y_2 - u} p_{11}^{u} p_{10}^{y_1-u} p_{01}^{y_2-u} p_{00}^{n-y_1-y_2-u}. \tag{1.4}$$

A seguir estão, respectivamente, as expressões das funções de probabilidade da distribuição Bernoulli Multivariada e da distribuição Binomial Multivariada (KAWAMURA, 1979) para uma dimensão K

$$P(\mathbf{Z}) = P(Z_1 = i_1, \ldots, Z_k = i_K) = p_{i_1 i_2 \ldots i_K} \tag{1.5}$$

onde $\mathbf{Z}=(Z_1, Z_2, \ldots, Z_K)'$ representa um vetor de dimensão K, assumindo 2^K valores possíveis, $(1,1,1,\ldots,1)$, $(1,0,0,\ldots,0)$, ..., $(0,0,0,\ldots,1)$, ..., $(0,0,0,\ldots,0)$ com probabilidades $p_{111\ldots1}$, $p_{100\ldots0}$, ..., $p_{000\ldots1}$, ..., $p_{000\ldots0}$, respectivamente, $\sum\limits_{i_1 i_2 \ldots i_K} p_{i_1 i_2 \ldots i_K} = 1$ e

$$P(\mathbf{Y}) = \sum_{\mathbf{U}\in\mathbf{A}} \frac{n!}{\prod_{\mathbf{i}} U_{\mathbf{i}}!} \prod_{\mathbf{i}} p_{\mathbf{i}}^{U_{\mathbf{i}}} \tag{1.6}$$

onde $\mathbf{Y}=(Y_1, Y_2, \ldots, Y_K)'$ representa um vetor aleatório de dimensão K; $\mathbf{U} = (U_{11\ldots1}, \ldots, U_{00\ldots1}, \ldots, U_{00\ldots0})'$; $\mathbf{A} = \{\mathbf{U} \in \mathbf{N}^{2^K}\text{: } \sum_{i_1=1} U_{\mathbf{i}}=Y_1, \sum_{i_2=1} U_{\mathbf{i}}=Y_2, \ldots, \sum_{i_K=1} U_{\mathbf{i}}=Y_K\}$;

o índice $\mathbf{i}=(i_1, i_2, \ldots, i_K)'$ representa um vetor de dimensão K, onde cada elemento i_j é igual a 1 ou 0; $U_{\mathbf{i}}$ é igual ao número de vezes em que ocorreu o vetor $\mathbf{Z}=(i_1, i_2, \ldots, i_K)'$ em n observações.

1.3.1 Distribuição Poisson Multivariada

O caso bivariado desta distribuição foi considerado por HOLGATE (1964) e, posteriormente, o caso multivariado por KAWAMURA (1979). Existem várias formas de definir esta distribuição e uma delas é semelhante ao caso univariado, como limite da distribuição Binomial Multivariada (KAWAMURA, 1979). No caso bivariado, HAMDAN; AL-BAYYATI (1969) calcularam o limite da expressão (1.4) considerando $n \to \infty$, $p_{11} \to 0$, $p_{01} \to 0$, $p_{10} \to 0$, com $np_{11} \to \lambda_{11}$, $np_{01} \to \lambda_{01}$ e $np_{10} \to \lambda_{10}$, onde λ_{10}, λ_{01} e λ_{11} são constantes, utilizando

$$\lim_{n\to\infty} \left[1 - \frac{(\lambda_{10} + \lambda_{01} + \lambda_{11})}{n}\right]^{(n-y_1-y_2+u)} = \exp\{-(\lambda_{10} + \lambda_{01} + \lambda_{11})\}$$

para obter a função de probabilidade da distribuição Poisson Bivariada:

$$P(Y_1 = y_1, Y_2 = y_2) =$$

$$= \begin{cases} \sum_{u=0}^{min(y_1,y_2)} \exp\{-(\lambda_{10} + \lambda_{01} + \lambda_{11})\} \frac{\lambda_{11}^u}{u!} \frac{\lambda_{10}^{(y_1-u)}}{(y_1-u)!} \frac{\lambda_{01}^{(y_2-u)}}{(y_2-u)!} & \text{se } y_1 \geq 0, y_2 \geq 0 \\ 0 & \text{caso contrário.} \end{cases} \tag{1.7}$$

A função de probabilidade da distribuição Poisson Multivariada para dimensão K é obtida similarmente, calculando-se o limite da expressão (1.6) para $n \to \infty$ e $p_{\mathbf{i}} \to 0$ com $np_{\mathbf{i}} \to \lambda_{\mathbf{i}}$, $\lambda_{\mathbf{i}}$ constante. O resultado deste limite é dado pela expressão

$$P(\mathbf{Y}) = \begin{cases} \sum_{\mathbf{U}\in\mathbf{A}} \frac{\prod_i \lambda_i^{U_i} \exp\{-\sum_i \lambda_i\}}{\prod_i U_i!} & \text{se } Y_1 \geq 0, \ldots, Y_K \geq 0 \\ 0 & \text{caso contrário.} \end{cases} \tag{1.8}$$

Uma outra maneira de definir a distribuição Poisson Multivariada é através de somas de distribuições Poisson univariadas independentes (KAWAMURA, 1979). No caso bivariado, o vetor $\mathbf{Y}=(Y_1, Y_2)'$ é definido a partir de $Y_1 = U_{11} + U_{10}$ e $Y_2 = U_{11} + U_{01}$ onde U_{10}, U_{01}, U_{11} são variáveis com distribuições de Poisson univariadas independentes de parâmetros λ_{10}, λ_{01}, λ_{11}, respectivamente, conforme MARSHALL; OLKIN (1985) e KAWAMURA (1973). Desse modo

$$\begin{aligned} P(Y_1 = y_1, Y_2 = y_2) &= P(U_{10} + U_{11} = y_1, U_{01} + U_{11} = y_2) = \\ &= \sum_{u=0}^{min(y_1,y_2)} P(U_{11} = u, U_{10} = y_1 - u, U_{01} = y_2 - u). \end{aligned} \tag{1.9}$$

Como U_{11}, U_{10} e U_{01} são independentes, então $P(Y_1 = y_1, Y_2 = y_2)$ é dada pela expressão (1.7). Este fato permite simular uma distribuição Poisson Bivariada a partir de distribuições Poisson univariadas independentes.

Generalizando esta definição para uma dimensão K $>$ 2, sejam $U_{1000\ldots0}$, $U_{0100\ldots0}$, $U_{0010\ldots0}$, ..., $U_{0000\ldots1}$, $U_{1100\ldots0}$, $U_{1010\ldots0}$, ..., $U_{1111\ldots1}$ variáveis Poisson univariadas independentes de parâmetros $\lambda_{1000\ldots0}$, $\lambda_{0100\ldots0}$, $\lambda_{0010\ldots0}$, ..., $\lambda_{0000\ldots1}$, $\lambda_{1100\ldots0}$, $\lambda_{1010\ldots0}$, ..., $\lambda_{1111\ldots1}$, respectivamente e sejam os componentes do vetor $\mathbf{Y}$ dados por $Y_1 = \sum_{i_1=1} U_{\mathbf{i}}$, $Y_2 = \sum_{i_2=1} U_{\mathbf{i}}$, ..., $Y_K = \sum_{i_K=1} U_{\mathbf{i}}$. Então o vetor $\mathbf{Y}$ segue uma distribuição Poisson Multivariada cuja função de probabilidade está expressa em (1.8).

JOHNSON; KOTZ (1969) apresentaram um caso particular de distribuição Poisson Multivariada considerando apenas $K+1$ variáveis Poisson independentes: $U_{1000\ldots0}$, $U_{0100\ldots0}$,

$U_{0010...0}$, ..., $U_{0000...1}$, $U_{1111...1}$ de parâmetros $(\lambda_{1000...0} - \lambda_{1111...1})$, $(\lambda_{0100...0} - \lambda_{1111...1})$, $(\lambda_{0010...0} - \lambda_{1111...1})$, ..., $(\lambda_{0000...1} - \lambda_{1111...1})$, $\lambda_{1111...1}$, respectivamente e o vetor $\mathbf{Y}$ com a mesma definição anterior. A distribuição Poisson Multivariada assim definida apresenta a propriedade de equicovariância entre dois componentes quaisquer do vetor $\mathbf{Y}$; a covariância comum é dada por $\lambda_{1111...1}$.

HOLGATE (1964) considera as variáveis aleatórias U_{10}, U_{01}, U_{11} com distribuições Poisson independentes de parâmetros $(\lambda_{10} - \lambda_{11})$, $(\lambda_{01} - \lambda_{11})$, λ_{11}, respectivamente. Essa reparametrização é bastante útil na dedução dos estimadores de **MV** dos parâmetros correspondentes. A expressão (1.7) reescrita segundo esta reparametrização é

$$P(Y_1 = y_1, Y_2 = y_2) =$$

$$= \sum_{u=0}^{min(y_1,y_2)} \exp\{-(\lambda_{10} + \lambda_{01} - \lambda_{11})\} \frac{\lambda_{11}^u}{u!} \frac{(\lambda_{10}-\lambda_{11})^{(y_1-u)}}{(y_1-u)!} \frac{(\lambda_{01}-\lambda_{11})^{(y_2-u)}}{(y_2-u)!}. \tag{1.10}$$

A seguir estão resumidas algumas propriedades desta distribuição. O vetor de médias da distribuição Poisson Multivariada é

$$E(\mathbf{Y}) = \begin{pmatrix} \sum_{i_1=1} \lambda_{\mathbf{i}} \\ \sum_{i_2=1} \lambda_{\mathbf{i}} \\ \cdots \\ \sum_{i_K=1} \lambda_{\mathbf{i}} \end{pmatrix}. \tag{1.11}$$

No caso bivariado, temos:

$$E(\mathbf{Y}) = \begin{pmatrix} \lambda_{10} + \lambda_{11} \\ \lambda_{01} + \lambda_{11} \end{pmatrix}. \tag{1.12}$$

A matriz de covariância da distribuição Poisson Multivariada é

$$Var(\mathbf{Y}) = \begin{pmatrix} \sum_{i_1=1} \lambda_{\mathbf{i}} & \sum_{i_1=i_2=1} \lambda_{\mathbf{i}} & \cdots & \sum_{i_1=i_K=1} \lambda_{\mathbf{i}} \\ \sum_{i_1=i_2=1} \lambda_{\mathbf{i}} & \sum_{i_2=1} \lambda_{\mathbf{i}} & \cdots & \sum_{i_2=i_K=1} \lambda_{\mathbf{i}} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{i_1=i_K=1} \lambda_{\mathbf{i}} & \sum_{i_2=i_K=1} \lambda_{\mathbf{i}} & \cdots & \sum_{i_K=1} \lambda_{\mathbf{i}} \end{pmatrix}. \tag{1.13}$$

No caso bivariado, a matriz correspondente é

$$Var(\mathbf{Y}) = \begin{pmatrix} \lambda_{10} + \lambda_{11} & \lambda_{11} \\ \lambda_{11} & \lambda_{01} + \lambda_{11} \end{pmatrix}. \tag{1.14}$$

O coeficiente de correlação linear entre Y_j e Y_k é dada por

$$Corr(Y_j, Y_k) = \frac{\sum_{i_j=i_k=1} \lambda_{\mathbf{i}}}{\sqrt{(\sum_{i_j=1} \lambda_{\mathbf{i}})(\sum_{i_k=1} \lambda_{\mathbf{i}})}} \tag{1.15}$$

e no caso bivariado, temos

$$Corr(Y_1, Y_2) = \frac{\lambda_{11}}{\sqrt{(\lambda_{10} + \lambda_{11})(\lambda_{01} + \lambda_{11})}}. \tag{1.16}$$

A covariância entre Y_j e Y_k deve satisfazer à desigualdade

$$0 \leq \sum_{i_j=i_k=1} \lambda_{\mathbf{i}} \leq min\left(\sum_{i_j=1} \lambda_{\mathbf{i}}, \sum_{i_k=1} \lambda_{\mathbf{i}}\right). \tag{1.17}$$

No caso bivariado, temos (HOLGATE, 1964):

$$0 \leq \lambda_{11} \leq min(\lambda_{10}, \lambda_{01}). \tag{1.18}$$

Como decorrência de (1.17), o coeficiente de correlação linear entre (Y_j e Y_k) satisfaz a desigualdade:

$$0 \leq Corr(Y_j, Y_k) \leq min\left(\sqrt{\frac{\sum_{i_j=1} \lambda_{\mathbf{i}}}{\sum_{i_k=1} \lambda_{\mathbf{i}}}}, \sqrt{\frac{\sum_{i_k=1} \lambda_{\mathbf{i}}}{\sum_{i_j=1} \lambda_{\mathbf{i}}}}\right). \tag{1.19}$$

No caso bivariado, a desigualdade correspondente é (HOLGATE, 1964):

$$0 \leq Corr(Y_1, Y_2) \leq min\left(\sqrt{\lambda_{10}/\lambda_{01}}, \sqrt{\lambda_{01}/\lambda_{10}}\right). \tag{1.20}$$

Conforme as expressões (1.19) e (1.20), o coeficiente de correlação linear assume apenas valores positivos.

A função geradora de momentos (**FGM**) é dada por:

$$FGM(s_1, s_2, \ldots, s_K) = \exp\{-\sum_{\mathbf{i}} \lambda_{\mathbf{i}} + \sum_{\mathbf{i}} \lambda_{\mathbf{i}} \mathbf{s}^{\mathbf{i}}\} \tag{1.21}$$

onde $\mathbf{s}^{\mathbf{i}} = s_1^{i_1} s_2^{i_2} \ldots s_K^{i_K}$. No caso bivariado, a **FGM** correspondente (KAWAMURA, 1973) é

$$FGM(s_1, s_2) = \exp\{-\lambda_{10} - \lambda_{01} - \lambda_{11} + \lambda_{10}s_1 + \lambda_{01}s_2 + \lambda_{11}s_1s_2\}. \tag{1.22}$$

Marginalmente Y_j é uma variável com distribuição Poisson de parâmetro igual a $\sum_{i_j=1} \lambda_{\mathbf{i}}$. No caso bivariado, as distribuições marginais de Y_1 e Y_2 são distribuições Poisson de parâmetros $(\lambda_{10} + \lambda_{11})$ e $(\lambda_{01} + \lambda_{11})$, respectivamente (KAWAMURA, 1973).

Conforme KAWAMURA (1987), valem as seguintes relações de recorrência:

$$Y_1 P(\mathbf{Y}) = \sum_{i_1=1} \lambda_{\mathbf{i}} P(\mathbf{Y} - \mathbf{i}), \tag{1.23}$$

$$Y_2 P(\mathbf{Y}) = \sum_{i_2=1} \lambda_{\mathbf{i}} P(\mathbf{Y} - \mathbf{i}), \tag{1.24}$$

$$\ldots,$$

$$Y_K P(\mathbf{Y}) = \sum_{i_K=1} \lambda_{\mathbf{i}} P(\mathbf{Y} - \mathbf{i}) \tag{1.25}$$

e

$$P(\mathbf{Y} = \mathbf{0}) = \exp\{-\sum_{\mathbf{i}} \lambda_{\mathbf{i}}\}. \tag{1.26}$$

No caso bivariado, as relações de recorrência são (KAWAMURA, 1973):

$$\begin{aligned} y_1 P(Y_1 = y_1, Y_2 = y_2) &= \lambda_{10} P(Y_1 = y_1 - 1, Y_2 = y_2) + \\ &+ \lambda_{11} P(Y_1 = y_1 - 1, Y_2 = y_2 - 1), \end{aligned} \quad (1.27)$$

$$\begin{aligned} y_2 P(Y_1 = y_1, Y_2 = y_2) &= \lambda_{01} P(Y_1 = y_1, Y_2 = y_2 - 1) + \\ &+ \lambda_{11} P(Y_1 = y_1 - 1, Y_2 = y_2 - 1) \end{aligned} \quad (1.28)$$

e

$$P(Y_1 = 0, Y_2 = 0) = \exp -(\lambda_{10} + \lambda_{01} + \lambda_{11}). \quad (1.29)$$

Com as relações de recorrência (1.23), (1.24) e (1.25) ou (1.27) e (1.28), é possível obter a distribuição de probabilidade a partir da probabilidade de um ponto $\mathbf{Y} \neq \mathbf{0}$.

Para obter os estimadores de máxima verossimilhança (**MV**) no caso bivariado, HOLGATE (1964) derivou a expressão (1.10) em relação aos parâmetros λ_{10}, λ_{01} e λ_{11}. As expressões apresentadas a seguir referem-se a uma observação de contagens bivariadas $\mathbf{Y}_i$, sendo suprimido, para efeito de simplificação de notação, o índice i

$$\frac{\partial}{\partial \lambda_{10}} P(\mathbf{Y}) = -P(Y_1, Y_2) + P(Y_1 - 1, Y_2), \quad (1.30)$$

$$\frac{\partial}{\partial \lambda_{01}} P(\mathbf{Y}) = -P(Y_1, Y_2) + P(Y_1, Y_2 - 1) \quad (1.31)$$

e

$$\begin{aligned} \frac{\partial}{\partial \lambda_{11}} P(\mathbf{Y}) &= P(Y_1, Y_2) - P(Y_1 - 1, Y_2) - \\ &- P(Y_1, Y_2 - 1) + P(Y_1 - 1, Y_2 - 1). \end{aligned} \quad (1.32)$$

Utilizando as expressões (1.30), (1.31) e (1.32), as relações de recorrência (1.27) e (1.28) e considerando uma amostra de tamanho n, as equações de verossimilhança

$$\sum_{i=1}^{n} \frac{1}{P(\mathbf{Y}_i)} \frac{\partial P(\mathbf{Y}_i)}{\partial \lambda_{10}} = \sum_{i=1}^{n} \frac{1}{P(\mathbf{Y}_i)} \frac{\partial P(\mathbf{Y}_i)}{\partial \lambda_{01}} = \sum_{i=1}^{n} \frac{1}{P(\mathbf{Y}_i)} \frac{\partial P(\mathbf{Y}_i)}{\partial \lambda_{11}} = 0 \qquad (1.33)$$

equivalem a

$$\frac{\overline{Y_1}}{(\lambda_{10} - \lambda_{11})} - \frac{\lambda_{11}}{(\lambda_{10} - \lambda_{11})}\overline{R} - 1 = 0, \qquad (1.34)$$

$$\frac{\overline{Y_2}}{(\lambda_{01} - \lambda_{11})} - \frac{\lambda_{11}}{(\lambda_{01} - \lambda_{11})}\overline{R} - 1 = 0 \qquad (1.35)$$

e

$$\frac{\overline{Y_1}}{(\lambda_{10} - \lambda_{11})} + \frac{\overline{Y_2}}{(\lambda_{01} - \lambda_{11})} - \left(1 + \frac{\lambda_{11}}{(\lambda_{10} - \lambda_{11})} + \frac{\lambda_{11}}{(\lambda_{01} - \lambda_{11})}\right)\overline{R} - 1 = 0 \qquad (1.36)$$

onde

$$\overline{R} = \sum_{i=1}^{n} \frac{R(\mathbf{Y}_i)}{n}$$

e

$$R(\mathbf{Y}_i) = \frac{P(Y_{i1} - 1, Y_{i2} - 1)}{P(Y_{i1}, Y_{i2})}.$$

Eliminando $\overline{R}$ das expressões (1.34), (1.35) e (1.36), chega-se às soluções:

$$\hat{\lambda}_{10} = \overline{Y_1} \qquad (1.37)$$

e

$$\hat{\lambda}_{01} = \overline{Y_2}. \qquad (1.38)$$

Substituindo λ_{10} e λ_{01} pelos seus estimadores (1.37) e (1.38), respectivamente, em qualquer uma das expressões (1.34), (1.35) ou (1.36), obtém-se a equação $\overline{R} = 1$, cuja solução $\hat{\lambda}_{11}$ em termos de λ_{11} pode ser obtida através de métodos iterativos.

A matriz de covariância dos estimadores de **MV** é apresentada por HOLGATE (1964) e corresponde a

$$Var(\ \hat{\lambda}_{10},\ \ \hat{\lambda}_{01},\ \ \hat{\lambda}_{11}\ ) = \frac{1}{n}\begin{pmatrix} \lambda_{10} & \lambda_{11} & \lambda_{11} \\ \lambda_{11} & \lambda_{01} & \lambda_{11} \\ \lambda_{11} & \lambda_{11} & \lambda^* \end{pmatrix} \tag{1.39}$$

onde

$$\lambda^* = \frac{\lambda_{11}^2(\lambda_{10} + \lambda_{01} - 2\lambda_{11}) - \lambda_{11}^2 + (\lambda_{10} - 2\lambda_{11})(\lambda_{01} - 2\lambda_{11})}{(\lambda_{10}\lambda_{01} - \lambda_{11}^2)(Q - 1) - (\lambda_{10} + \lambda_{01} - 2\lambda_{11})}$$

e

$$Q = \sum_{Y_1,Y_2=1}^{\infty} \frac{P^2(Y_1 - 1, Y_2 - 1)}{P(Y_1, Y_2)}.$$

A covariância entre (Y_1, Y_2) é igual a λ_{11} e seu estimador pelo método dos momentos (**MM**) é dado pela covariância amostral S_{12}. Este estimador é consistente, pois

$$\lim_{n\to\infty} E(S_{12}) = \lim_{n\to\infty}(n - 1)\lambda_{11}/n = \lambda_{11} \tag{1.40}$$

e

$$Var(S_{12}) = \frac{\lambda_{10}\lambda_{01} + \lambda_{11} + \lambda_{11}^2}{n} + O(n^{-2}). \tag{1.41}$$

HOLGATE (1964) sugeriu um estimador alternativo para λ_{11} denominado "Double-Zero " (**DZ**). Ele é obtido a partir das proporções esperadas e observadas no ponto (0,0). A probabilidade de $(Y_1 = 0, Y_2 = 0)$ é igual a $\exp\{-(\lambda_{10} + \lambda_{01} - \lambda_{11})\}$; seja ϕ a proporção observada do evento $(Y_1 = 0, Y_2 = 0)$ numa amostra de n elementos de uma distribuição Poisson Bivariada. Então fazendo $\exp\{-(\lambda_{10} + \lambda_{01} - \lambda_{11})\} = \phi$ e, substituindo λ_{10} e λ_{01} por seus estimadores de **MV**, define-se um estimador alternativo para λ_{11}, dado por

$$\tilde{\lambda}_{11} = \overline{Y_1} + \overline{Y_2} + \ln \phi. \tag{1.42}$$

HOLGATE (1964) mostrou que

$$E(\tilde{\lambda}_{11}) = \lambda_{11} + \frac{\exp\{\lambda_{10} + \lambda_{01} - \lambda_{11}\} - 1}{n} \tag{1.43}$$

e

$$Var(\tilde{\lambda}_{11}) = \frac{\exp\{\lambda_{10} + \lambda_{01} - \lambda_{11}\} - 1}{n}. \tag{1.44}$$

LOUKAS et al. (1986) sugeriram um outro estimador para λ_{11} denominado "Even-Point " (**EP**), cuja definição está a seguir.

A função geradora de probabilidade (**FGP**) da distribuição Poisson Bivariada (KAWAMURA, 1973) é igual a

$$\begin{aligned} FGP(u,v) &= \sum_{i=0,j=0}^{\infty} u^i v^j P(Y_1 = i, Y_2 = j) = \\ &= \exp\{(\lambda_{10} - \lambda_{11})(u-1) + (\lambda_{01} - \lambda_{11})(v-1) + \\ &\quad + \lambda_{11}(uv - 1)\} \end{aligned} \tag{1.45}$$

com u e $v \in R$. Calculando a **FGP** nos pontos (1, 1) e $(-1, -1)$ obtêm-se, respectivamente,

$$FGP(1,1) = 1 = \sum_{i=0,j=0}^{\infty} P(Y_1 = i, Y_2 = j) \tag{1.46}$$

e

$$\begin{aligned} FGP(-1,-1) &= \exp\{-2(\lambda_{10} + \lambda_{01} - 2\lambda_{11})\} = \\ &= \sum_{i=0,j=0}^{\infty} P(Y_1 = 2i, Y_2 = 2j) + P(Y_1 = 2i+1, Y_2 = 2j+1) - \end{aligned}$$

$$-P(Y_1 = 2i, Y_2 = 2j+1) - P(Y_1 = 2i+1, Y_2 = 2j). \tag{1.47}$$

Somando (1.46) e (1.47), resulta que para o evento A={ Y_1 e Y_2 são pares ou Y_1 e Y_2 são ímpares},

$$P(A) = \frac{1 + \exp\{-2(\lambda_{10} + \lambda_{01} - 2\lambda_{11})\}}{2}. \tag{1.48}$$

Seja δ a proporção observada da ocorrência do evento A numa amostra de tamanho n. Igualando (1.48) a δ, obtém-se o estimador **EP** do parâmetro λ_{11}, dado por

$$\lambda_{11}^* = \frac{\overline{Y_1} + \overline{Y_2}}{2} + \frac{\ln(2\delta - 1)}{4} \quad \text{se} \quad \delta > n/2. \tag{1.49}$$

A esperança e a variância de (1.49) são respectivamente

$$E(\lambda_{11}^*) = \lambda_{11} + \frac{1}{8n} - \frac{\exp\{4(\lambda_{10} + \lambda_{01} - 2\lambda_{11})\}}{8n} \tag{1.50}$$

e

$$Var(\lambda_{11}^*) = \frac{\exp\{4(\lambda_{10} + \lambda_{01} - 2\lambda_{11})\} - 1}{16n}. \tag{1.51}$$

Para os casos em que λ_{01} é igual a λ_{10}, uma comparação entre as eficiências dos estimadores **DZ**, **EP** e **MM** em relação ao estimador de **MV**, em termos de variância generalizada, está em KOCHERLAKOTA; KOCHERLAKOTA (1992).

No caso da distribuição Poisson Bivariada, uma hipótese de interesse é $H_0 : \lambda_{11} = 0$, que corresponde à independência entre Y_1 e Y_2. Considerando o **Exemplo 1.1**, isto corresponde a avaliar se o número de defeitos do tipo D_1 não está associado ao número de defeitos do tipo D_2. Algumas estatísticas para testar esta hipótese estão em KOCHERLAKOTA; KOCHERLAKOTA (1992).

1.3.2 Distribuição Poisson Composta Multivariada

Considere a situação em que na expressão (1.8) $\lambda_i = k_i\tau$ onde k_i é uma constante positiva e τ representa uma variável aleatória univariada discreta ou contínua com uma função de probabilidade $P(\tau)$ (para o caso discreto) ou uma função densidade $g(\tau)$ (para o caso contínuo) e com uma função geradora de momentos $M(t)$. Nestas condições, a distribuição de $\mathbf{Y}$ é Poisson Composta Multivariada. Maiores detalhes sobre o caso trivariado pode ser encontrado em (LIANG, 1989).

Nesta seção estão descritos alguns resultados desta classe de distribuições, dando maiores detalhes para o caso bivariado com $(\lambda_{10} - \lambda_{11}) = k_{10}\tau$, $(\lambda_{01} - \lambda_{11}) = k_{01}\tau$ e $\lambda_{11} = k_{11}\tau$.

KOCHERLAKOTA (1988) apresentou um tratamento unificado para a distribuição Poisson Composta Bivariada, cujos principais resultados estão descritos a seguir.

A função de probabilidade de $\mathbf{Y}$ é dada por

$$P(Y_1 = y_1, Y_2 = y_2) =$$
$$= \sum_{u=0}^{min(y_1,y_2)} \binom{y_1}{u} \binom{y_2}{u} u! M^{(y_1+y_2-u)}(\gamma)\delta^u \frac{(\lambda_{10}-\lambda_{11})^{y_1}}{y_1!} \frac{(\lambda_{01}-\lambda_{11})^{y_2}}{y_2!} \tag{1.52}$$

onde

$$\gamma = -(\lambda_{10} + \lambda_{01} - 2\lambda_{11}),$$

$$\delta = \frac{\lambda_{11}}{(\lambda_{10} - \lambda_{11})(\lambda_{01} - \lambda_{11})}$$

e $M^{(j)}(t)$ representa a j-ésima derivada da função geradora de momentos $M(t)$.

O momento fatorial conjunto da distribuição Poisson Composta Bivariada é dado por

$$\mu_{[i,j]} = (\lambda_{10})^i (\lambda_{01})^j \sum_{u=0}^{min(i,j)} \frac{i!j!\delta^u \mu'_{i+j-u}}{(i-u)!(j-u)!u!} \tag{1.53}$$

onde

$$\delta = \frac{\lambda_{11}}{\lambda_{10}\lambda_{01}},$$

$$\mu_{[i,j]} = E(Y_1^{[i]}Y_2^{[j]}),$$

$Y_1^{[i]}$ representa $Y_1(Y_1-1)\dots(Y_1-i+1)$ e μ_i' é o i-ésimo momento de τ.

De (1.53), obtém-se o coeficiente de correlação linear entre Y_1 e Y_2 dado por

$$Corr(Y_1, Y_2) = \frac{\lambda_{10}\lambda_{01}\sigma^2 + \mu_1'\lambda_{11}}{\sqrt{[\lambda_{10}^2\sigma^2 + \mu_1'\lambda_{10}][\lambda_{01}^2\sigma^2 + \mu_1'\lambda_{01}]}} \tag{1.54}$$

onde σ^2 representa a variância de τ. Conforme (1.54), esta classe de distribuições também só admite correlações positivas, pois λ_{10}, λ_{01}, λ_{11} e σ^2 assumem somente valores positivos e $\mu_1' = \mathrm{E}(\tau)$ assume valores positivos no contexto da análise de contagens.

As expressões (1.52), (1.53) e (1.54) correspondentes aos casos em que a variável aleatória τ obedece às distribuições Gama, Poisson, Normal ou Inversa Gaussiana podem ser encontradas em KOCHERLAKOTA (1988). Ainda segundo KOCHERLAKOTA (1988), valem as seguintes relações de recorrência para ($Y_1 \geq 1$, $Y_2 \geq 1$):

- se $y_1 \geq y_2$

$$\begin{aligned} P(\mathbf{Y}) = {} & \frac{(\lambda_{10}-\lambda_{11})}{(\lambda_{01}-\lambda_{11})}\frac{(y_1+1)}{y_2}P(\mathbf{Y}+\mathbf{E_1}-\mathbf{E_2}) + \\ & +(y_1-y_2+1)\frac{\lambda_{11}}{(\lambda_{10}-\lambda_{11})}\frac{1}{y_2}P(\mathbf{Y}-\mathbf{E_2}) \end{aligned} \tag{1.55}$$

- se $y_1 \leq y_2$

$$\begin{aligned} P(\mathbf{Y}) = {} & \frac{(\lambda_{10}-\lambda_{11})}{(\lambda_{01}-\lambda_{11})}\frac{(y_2+1)}{y_1}P(\mathbf{Y}-\mathbf{E_1}+\mathbf{E_2}) + \\ & +(y_2-y_1+1)\frac{\lambda_{11}}{(\lambda_{01}-\lambda_{11})}\frac{1}{y_1}P(\mathbf{Y}-\mathbf{E_1}) \end{aligned} \tag{1.56}$$

com $\mathbf{E_1} = (1,0)'$ e $\mathbf{E_2} = (0,1)'$.

Um trabalho relacionado com a estimação dos parâmetros desta distribuição quando τ obedece a uma distribuição Normal encontra-se em PAPAGEOURGIOU et al. (1983), onde são comparados em termos de variância generalizada, os estimadores obtidos segundo os métodos **EP**, **DZ** e **MM** em relação aos obtidos segundo o método **MV**.

Para maiores detalhes sobre esta classe de distribuições, ver KOCHELAKOTA (1988), LIANG (1989) e KOCHERLAKOTA; KOCHERLAKOTA (1992).

Os exemplos de distribuição Poisson Composta Bivariada citados em KOCHERLAKOTA (1988) são casos onde os parâmetros da distribuição Poisson Bivariada seguem uma mesma distribuição univariada. Na seção seguinte está apresentado um caso onde os parâmetros de distribuições Poisson independentes obedecem a uma distribuição multivariada.

1.3.3 Distribuição Poisson Log-Normal Multivariada

Esta distribuição foi sugerida por AITCHISON; HO (1989). Sejam Y_j, $j = 1, \ldots, K$, variáveis aleatórias independentes com distribuição Poisson de parâmetros $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_K)'$ aleatórios obedecendo a uma distribuição Log-Normal Multivariada, cuja função densidade é

$$g(\boldsymbol{\lambda} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) =$$
$$= (2\pi)^{-K/2}(\lambda_1 \ldots \lambda_K)^{-1}(\det \boldsymbol{\Sigma})^{-1/2} \exp\{-1/2[\ln(\boldsymbol{\lambda}) - \boldsymbol{\mu}]'\boldsymbol{\Sigma}^{-1}[\ln(\boldsymbol{\lambda}) - \boldsymbol{\mu}]\} \tag{1.57}$$

onde

$$\boldsymbol{\mu} = (\ \mu_1,\ \ \mu_2,\ \ \ldots,\ \ \mu_K\ )'\ ,\ \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1K} \\ \sigma_{12} & \sigma_{22} & \ldots & \sigma_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \ldots & \sigma_{KK} \end{pmatrix}$$

são constantes e $\mathbf{ln}(\boldsymbol{\lambda})$ representa um vetor cujos elementos são os logaritmos dos elementos de $\boldsymbol{\lambda}$. A função de probabilidade da distribuição Poisson Log-Normal Multivariada é dada por

$$P(\mathbf{Y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{R_+^K} \prod_{j=1}^{K} f(Y_j \mid \lambda_j) g(\boldsymbol{\lambda} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\lambda} \tag{1.58}$$

onde $f(Y_j \mid \lambda_j)$ é a função de probabilidade da distribuição Poisson de parâmetro λ_j. Segundo AITCHISON; HO (1989), pode-se mostrar que

$$E(Y_j) = \exp\{\mu_j + \frac{1}{2}\sigma_{jj}\} = \alpha_j \tag{1.59}$$

e

$$Var(Y_j) = \alpha_j + \alpha_j^2(\exp\{\sigma_{jj}\} - 1). \tag{1.60}$$

Das expressões (1.59) e (1.60), observa-se que $Var(Y_j) \geq E(Y_j)$, de modo que as distribuições marginais podem ter sobredispersão em relação à distribuição Poisson.

Além disso, a covariância e o coeficiente de correlação linear entre Y_i e Y_j são , respectivamente

$$Cov(Y_i, Y_j) = \alpha_i \alpha_j (\exp\{\sigma_{ij}\} - 1) \tag{1.61}$$

e

$$Corr(Y_i, Y_j) = \frac{\exp\{\sigma_{ij}\} - 1}{[(\exp\{\sigma_{ii}\} - 1 + \alpha_i^{-1})(\exp\{\sigma_{jj}\} - 1 + \alpha_j^{-1})]^{1/2}}. \tag{1.62}$$

De (1.61) e (1.62), constata-se que esta distribuição admite covariâncias ou correlações tanto positivas quanto negativas.

O coeficiente de correlação linear entre λ_i e λ_j é dado por

$$Corr(\lambda_i, \lambda_j) = \frac{\exp\{\sigma_{ij}\} - 1}{\{(\exp\{\sigma_{ii}\} - 1)(\exp\{\sigma_{jj}\} - 1)\}^{1/2}}. \tag{1.63}$$

Comparando (1.62) e (1.63), observa-se que

$$| \, Corr(Y_i, Y_j) \, | < | \, Corr(\lambda_i, \lambda_j) \, |, \tag{1.64}$$

de modo que a amplitude do coeficiente de correlação linear entre duas variáveis Poisson Log-Normal está sempre limitada pela amplitude do coeficiente de correlação linear entre as duas variáveis Log-Normais subjacentes e é diretamente proporcional a α_i e α_j.

A função de probabilidade (1.58) não apresenta uma expressão explícita, e para cálculos envolvendo esta função, uma reparametrização é muito útil. Nesse sentido, note que, correspondendo à matriz $\boldsymbol{\Sigma}$, existe uma única matriz triangular

$$\mathbf{T} = \begin{pmatrix} \tau_{11} & 0 & 0 & \dots & 0 \\ \tau_{21} & \tau_{22} & 0 & \dots & 0 \\ \tau_{31} & \tau_{32} & \tau_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \tau_{K1} & \tau_{K2} & \tau_{K3} & \dots & \tau_{KK} \end{pmatrix}$$

tal que $\boldsymbol{\Sigma} = \mathbf{TT'}$. Desse modo, qualquer hipótese referente a $\boldsymbol{\Sigma}$ pode ser igualmente expressa em termos de $\mathbf{T}$.

No caso bivariado, quando σ_{21} for igual a zero, τ_{21} é igual a zero, o que corresponde ao caso em que Y_1 e Y_2 têm distribuições Poisson Log-Normais univariadas independentes. No caso de $\sigma_{11} = \sigma_{21} = \sigma_{22} = 0$, temos $\tau_{11} = \tau_{21} = \tau_{22} = 0$, o que corresponde ao caso em que Y_1 e Y_2 seguem distribuições Poisson univariadas independentes. Quando $\tau_{22} = 0$, temos a situação em que a correlação entre λ_1 e λ_2 é perfeita com $\sigma_{21} = \pm\sqrt{\sigma_{11}\sigma_{22}}$. Neste caso, Y_1 e Y_2 apresentam uma correlação máxima e seguem distribuições Poisson de parâmetros λ e $\alpha\lambda^\beta$, respectivamente, onde λ tem distribuição Log-Normal univariada de parâmetros (μ e σ^2).

Considere a seguinte transformação:

$$\ln(\boldsymbol{\lambda}) - \boldsymbol{\mu} = \mathbf{T}\mathbf{v}\sqrt{2} \tag{1.65}$$

onde $\mathbf{v}=(v_1, v_2, \dots, v_K)'$, que implica

$$\boldsymbol{\lambda} = \exp(\boldsymbol{\mu} + \mathbf{T}\mathbf{v}\sqrt{2}) \tag{1.66}$$

onde $\exp(\boldsymbol{\mu}+\mathbf{T}\mathbf{v}\sqrt{2})$ representa um vetor cujos elementos são exponenciais dos elementos de $(\boldsymbol{\mu}+\mathbf{T}\mathbf{v}\sqrt{2})$. Substituindo (1.66) em (1.58), obtém-se a função de probabilidade escrita em termos de $\boldsymbol{\mu}$ e $\mathbf{T}$:

$$P(\mathbf{Y} \mid \boldsymbol{\mu}, \mathbf{T}) = \pi^{-K/2} \int_{R^K} H(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{T}, \mathbf{v}) \exp(-\mathbf{v}'\mathbf{v}) d\mathbf{v} \tag{1.67}$$

com

$$\begin{aligned} H(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{T}, \mathbf{v}) \; &= \; \exp\{\mathbf{Y}'(\boldsymbol{\mu} + \mathbf{T}\mathbf{v}\sqrt{2}) - \mathbf{1}'_K \exp(\boldsymbol{\mu} + \mathbf{T}\mathbf{v}\sqrt{2}) - \\ & \quad -\mathbf{1}'_K \ln[\boldsymbol{\Gamma}(\mathbf{Y} + \mathbf{1}_K)]\} \end{aligned} \tag{1.68}$$

onde $\boldsymbol{\Gamma}(\mathbf{Y}+\mathbf{1}_K)$ representa um vetor cujos elementos são valores de funções gama calculadas nos elementos de $(\mathbf{Y}+\mathbf{1}_K)$ e $\mathbf{1}_K$ representa um vetor K-dimensional, cujos elementos são iguais a 1.

Desse modo, reescrevendo a expressão (1.58) como (1.67), pode-se obter aproximadamente a probabilidade de $\mathbf{Y}$ através de métodos de integração numérica. Maiores detalhes sobre este assunto serão dados na Seção 2.4.

O número de parâmetros da distribuição Poisson Log-Normal é $\frac{1}{2}K(K+3)$. Os estimadores de **MV** de $\boldsymbol{\mu}$ e $\mathbf{T}$ são obtidos iterativamente. Para isso, pode-se considerar o método de Newton-Raphson, que será visto com maiores detalhes na Seção 2.2.

Capítulo 2

Regressão Poisson

2.1 Introdução

Com relação ao **Exemplo 1.1**, suponha que haja interesse em estudar como varia o número médio de defeitos dos tipos D_1 e D_2 relativamente às diferentes máquinas e/ou fabricantes. Isto pode ser feito através de um modelo de Regressão

$$E(\mathbf{Y}_i) = f(\mathbf{X}_i, \boldsymbol{\beta}) \qquad (2.1)$$

onde $\mathbf{Y}_i$ é uma observação de $\mathbf{Y}$, $i = 1, \ldots, n$, $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{ti})'$ é um vetor com t variáveis preditoras e $\boldsymbol{\beta}=(\beta_1, \beta_2, \ldots, \beta_t)'$ é um vetor de parâmetros. A função $f(\mathbf{X}_i, \boldsymbol{\beta})$ relaciona o valor esperado da variável resposta (dependente) com as variáveis preditoras (independentes).

Modelos de regressão envolvendo contagens univariadas foram considerados por vários autores. FROME et al. (1973) consideraram um modelo de Regressão Poisson univariada. KOCH et al. (1986) utilizaram um modelo log-linear, isto é, $f(\mathbf{X}_i, \boldsymbol{\beta}) = \exp(\mathbf{X}_i'\boldsymbol{\beta})$ para analisar um conjunto de dados referentes a contagens de um tipo de bactéria em vários laboratórios e apresentaram um procedimento para obter as estimativas de **MV** dos parâmetros $\boldsymbol{\beta}$, tendo como estimador inicial o estimador de mínimos quadrados generalizados (**MQG**)

$$\hat{\boldsymbol{\beta}}^{(0)} = \left(\mathbf{X}'diag[\hat{\boldsymbol{\mu}}^{(0)}]\mathbf{X}\right)^{-1}\mathbf{X}'diag[\hat{\boldsymbol{\mu}}^{(0)}]\left[\ln(diag[\mathrm{N}]^{-1}\hat{\boldsymbol{\mu}}^{(0)})\right] \tag{2.2}$$

com

$$\hat{\boldsymbol{\mu}}^{(0)} = \begin{cases} \mathbf{y} = (y_1, y_2, \ldots, y_n)' & \text{se } y_i > 0, i = 1, \ldots, n \\ (\mathbf{y} + \mathbf{1}_n) & \text{caso contrário,} \end{cases}$$

onde $\mathbf{y}$ é uma amostra de contagens univariadas,

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1t} \\ x_{21} & x_{22} & \ldots & x_{2t} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{2n} & \ldots & x_{nt} \end{pmatrix}$$

é uma matriz de variáveis independentes e $diag[\mathrm{N}]$ é uma matriz diagonal tendo na diagonal principal os elementos do vetor $\mathrm{N}=(N_1, N_2, \ldots, N_n)'$ onde N_i corresponde a uma medida de exposição associada à i-ésima observação (população em risco, por exemplo). As estimativas dos parâmetros são obtidas a partir de iterações da expressão

$$\hat{\boldsymbol{\beta}}^{(w)} = \hat{\boldsymbol{\beta}}^{(w-1)} + \left[\widehat{Var}(\hat{\boldsymbol{\beta}}^{(w-1)})\right]\mathbf{X}'\left(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(w-1)}\right) \tag{2.3}$$

onde $\widehat{Var}(\hat{\boldsymbol{\beta}}^{(w)}) = \left(\mathbf{X}'diag[\hat{\boldsymbol{\mu}}^{(w)}]\mathbf{X}\right)^{-1}$ é uma estimativa da matriz de covariância assintótica de $\hat{\boldsymbol{\beta}}$ na w-ésima iteração e $\hat{\boldsymbol{\mu}}^{(w)} = diag[\mathrm{N}][\exp(\mathbf{X}\hat{\boldsymbol{\beta}}^{(w)})]$ é um vetor contendo os valores estimados de $\mathbf{y}$ na w-ésima iteração, $w = 1, 2, \ldots$.

Neste capítulo serão apresentados modelos de regressão para descrever as relações entre as contagens multivariadas provenientes de várias populações, considerando como resposta um vetor aleatório que obedece à distribuição Poisson Log-Normal Multivariada ou à distribuição Poisson Multivariada. No restante desta seção estão apresentados alguns exemplos de matrizes de especificação $\mathbf{X}_i$ para os dados do **Exemplo 1.1** e as alternativas para testes de hipóteses sobre os parâmetros dos modelos de regressão. Nas Seções 2.2 e 2.3 estão discutidos, respectivamente a estimação dos parâmetros dos modelos de Regressão

Poisson Log-Normal Bivariada e Regressão Poisson Bivariada através do método de **MV**; na Seção 2.4 estão descritos alguns problemas computacionais envolvidos no procedimento de estimação.

Considere uma amostra aleatória $\mathbf{Y}_1$, $\mathbf{Y}_2$, ..., $\mathbf{Y}_n$ do vetor resposta que obedece a uma distribuição Poisson Log-Normal Multivariada. Então

$$E\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \cdots \\ \mathbf{Y}_n \end{pmatrix} = \mathbf{exp}\begin{pmatrix} \mathbf{X}_1\boldsymbol{\beta} + (1/2)vecdiag(\boldsymbol{\Sigma}_1) \\ \mathbf{X}_2\boldsymbol{\beta} + (1/2)vecdiag(\boldsymbol{\Sigma}_2) \\ \cdots \\ \mathbf{X}_n\boldsymbol{\beta} + (1/2)vecdiag(\boldsymbol{\Sigma}_n) \end{pmatrix} \tag{2.4}$$

que pode ser reescrito como

$$E(\mathbf{Y}) = \mathbf{exp}\{\mathbf{X}\boldsymbol{\beta} + \frac{1}{2}\mathbf{vecdiag}(\boldsymbol{\Sigma})\} \tag{2.5}$$

onde $\mathbf{Y} = (\mathbf{Y}_1,\ \mathbf{Y}_2,\ \ldots,\ \mathbf{Y}_n)'$, $\mathbf{X} = (\mathbf{X}_1,\ \mathbf{X}_2,\ \ldots,\ \mathbf{X}_n)'$ é a matriz de especificação do modelo, $\mathbf{vecdiag}(\boldsymbol{\Sigma}) = (vecdiag(\boldsymbol{\Sigma}_1), vecdiag(\boldsymbol{\Sigma}_2), \ldots, vecdiag(\boldsymbol{\Sigma}_n))'$ e $vecdiag(\boldsymbol{\Sigma}_i)$ é um vetor de dimensão $\mathrm{K} \geq 2$, composto pelos elementos da diagonal da matriz $\boldsymbol{\Sigma}_i$. Os parâmetros a serem estimados são $\boldsymbol{\beta}$ e $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, n$.

Exemplo 2.1 - Considerem apenas os dados da **Tabela 1.1**, referentes ao número de defeitos dos tipos D_1 e D_2 da craqueadeira A_1 e suponha $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}$. Os parâmetros a serem estimados são $\boldsymbol{\beta}_{A_1}$=(β_{1A_1}, $\beta_{2A_1})'$ e $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$. Aqui o parâmetro β_{iA_1} está relacionado ao número esperado de defeitos do tipo D_i, i=1, 2, na craqueadeira A_1 e a matriz

$$\mathbf{X} = \mathbf{1}_n \bigotimes \mathbf{I}_2$$

onde $\mathbf{I}_2$ é uma matriz identidade de dimensão 2, n é igual ao tamanho da amostra obtida na máquina A_1 e $\otimes$ representa o produto direto entre duas matrizes tal que, se $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ então

$$\mathbf{C} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1j}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2j}\mathbf{B} \\ \dots & \dots & \dots & \dots \\ a_{k1}\mathbf{B} & a_{k2}\mathbf{B} & \dots & a_{kj}\mathbf{B} \end{pmatrix}.$$

Suponha que haja interesse em testar se existe associação entre o número de defeitos do tipo D_1 e o número de defeitos do tipo D_2 na máquina A_1. Isto corresponde a testar H_0: σ_{12}=0. O interesse em estudar associação positiva ou negativa pode ser expresso através das hipóteses alternativas H_a: $\sigma_{12} > 0$ ou $\sigma_{12} < 0$, respectivamente.

Exemplo 2.2 - Considere agora os dados das **Tabelas 1.1** a **1.4** e suponha novamente $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}$. Os parâmetros a serem estimados são $\boldsymbol{\Sigma}$ e $\boldsymbol{\beta}=\left(\boldsymbol{\beta}'_{A_1}, \ \boldsymbol{\beta}'_{A_2}, \ \boldsymbol{\beta}'_{B_1}, \ \boldsymbol{\beta}'_{B_2} \right)'$ com $\boldsymbol{\beta}_{j_k}=(\beta_{1j_k}, \beta_{2j_k})'$, onde o parâmetro β_{ij_k} está associado ao número médio de defeitos do tipo D_i na máquina j_k com i=1, 2; j= A, B e k = 1, 2. Neste exemplo

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_{A_1}} \otimes \mathbf{X}_{A_1} \\ \mathbf{1}_{n_{A_2}} \otimes \mathbf{X}_{A_2} \\ \mathbf{1}_{n_{B_1}} \otimes \mathbf{X}_{B_1} \\ \mathbf{1}_{n_{B_2}} \otimes \mathbf{X}_{B_2} \end{pmatrix}$$

com

$$\mathbf{X}_{A_1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{X}_{A_2} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{X}_{B_1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{X}_{B_2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

onde n_{j_k} é o tamanho da amostra da máquina j_k.

Suponha que haja interesse em testar se o número médio de defeitos dos dois tipos é igual nas quatro máquinas. Isto pode ser formulado através da hipótese

$$H_0: \ \boldsymbol{\beta}_{A_1} = \boldsymbol{\beta}_{A_2} = \boldsymbol{\beta}_{B_1} = \boldsymbol{\beta}_{B_2}. \tag{2.6}$$

Alternativamente, pode haver interesse em testar se a magnitude da associação entre o número de defeitos dos dois tipos não varia entre as quatro máquinas. Isto equivale a testar a hipótese

$$H_0: Cov(Y_{1A_1}, Y_{2A_1}) = \ldots = Cov(Y_{1B_2}, Y_{2B_2}), \tag{2.7}$$

onde Y_{ij_k} representa o número de defeitos do tipo D_i na máquina j_k. Lembrando que $Cov(Y_{1j_k}, Y_{2j_k}) = \alpha_{1j_k}\alpha_{2j_k}[\exp(\sigma_{12}) - 1]$, onde $\alpha_{ij_k} = E(Y_{ij_k})$, a hipótese (2.7) pode ser reescrita como

$$H_0: \alpha_{1A_1}\alpha_{2A_1} = \alpha_{1A_2}\alpha_{2A_2} = \alpha_{1B_1}\alpha_{2B_1} = \alpha_{1B_2}\alpha_{2B_2}. \tag{2.8}$$

Substituindo $\alpha_{ij_k}=E(Y_{ij_k})$ por $\exp(\beta_{ij_k} + \frac{1}{2}\sigma_{ii})$, a hipótese (2.8) expressa em termos dos parâmetros $\boldsymbol{\beta}$ é dada por

$$H_0: \mathbf{1}_2'\boldsymbol{\beta}_{A_1} = \mathbf{1}_2'\boldsymbol{\beta}_{A_2} = \mathbf{1}_2'\boldsymbol{\beta}_{B_1} = \mathbf{1}_2'\boldsymbol{\beta}_{B_2}. \tag{2.9}$$

Note que a não rejeição da hipótese (2.6) implica a não rejeição da hipótese (2.9), porém a recíproca não é verdadeira.

Exemplo 2.3 - Ainda considerando os dados das **Tabelas 1.1 a 1.4**, suponha que haja interesse em comparar os diferentes fabricantes e/ou diferentes máquinas de um fabricante com relação ao número de defeitos dos dois tipos. Nesse caso, tanto os parâmetros quanto a matriz **X** têm estrutura semelhante à do **Exemplo 2.2** com

$$\mathbf{X}_{A_1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}; \mathbf{X}_{A_2} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix};$$

$$\mathbf{X}_{B_1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}; \mathbf{X}_{B_2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Note que o parâmetro $\boldsymbol{\beta}_{j_2}$, j=A, B, está associado à diferença entre as máquinas do fabricante j quanto ao número médio de defeitos.

Suponha que haja interesse em saber se o número médio de defeitos dos dois tipos é igual nas máquinas de fabricantes diferentes; isto corresponde a testar

$$H_0 : \begin{cases} \boldsymbol{\beta}_{A_1} = \boldsymbol{\beta}_{B_1} \\ \boldsymbol{\beta}_{A_2} = \boldsymbol{\beta}_{B_2}. \end{cases} \tag{2.10}$$

Testar se o número médio de defeitos é igual entre as máquinas dos dois fabricantes corresponde a testar a hipótese

$$H_0 : \begin{cases} \boldsymbol{\beta}_{A_2} = \mathbf{0} \\ \boldsymbol{\beta}_{B_2} = \mathbf{0}. \end{cases} \tag{2.11}$$

A hipótese (2.7) considerando a situação do **Exemplo 2.3**, reescrita em termos de $\boldsymbol{\beta}$, corresponde a

$$H_0 : \begin{cases} \mathbf{1}_2'\boldsymbol{\beta}_{A_2} & = & \mathbf{0} \\ \mathbf{1}_2'\boldsymbol{\beta}_{B_2} & = & \mathbf{0} \\ \mathbf{1}_2'\boldsymbol{\beta}_{A_1} & = & \mathbf{1}_2'\boldsymbol{\beta}_{B_1}. \end{cases} \tag{2.12}$$

De modo semelhante, podem-se construir modelos de Regressão Poisson Multivariada. Considere uma amostra aleatória $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ proveniente de uma distribuição Poisson Multivariada; utilizando um modelo log-linear, temos

$$E\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \cdots \\ \mathbf{Y}_n \end{pmatrix} = \exp\begin{pmatrix} \mathbf{X}_1\boldsymbol{\beta} \\ \mathbf{X}_2\boldsymbol{\beta} \\ \cdots \\ \mathbf{X}_n\boldsymbol{\beta} \end{pmatrix} \tag{2.13}$$

que pode ser reescrito como

$$E(\mathbf{Y}) = \exp\{\mathbf{X}\boldsymbol{\beta}\} \tag{2.14}$$

onde $\mathbf{Y} = (\mathbf{Y}_1,\ \mathbf{Y}_2,\ \ldots,\ \mathbf{Y}_n)'$, $\mathbf{X} = (\mathbf{X}_1,\ \mathbf{X}_2,\ \ldots,\ \mathbf{X}_n)'$ é a matriz de especificação do modelo e $Cov(Y_{ij}, Y_{ik}) = \theta_{ijk}^2$, $j \neq k$, $i = 1, \ldots, n$.

Exemplo 2.4 - Considere os dados da **Tabela 1.1** como no **Exemplo 2.1** e suponha $Cov(Y_{i1}, Y_{i2}) = \theta^2$, $i = 1, \ldots, n$. Então os parâmetros a serem estimados são $\boldsymbol{\beta}$ e θ. Neste caso, a matriz de especificação $\mathbf{X}$ tem estrutura semelhante à do **Exemplo 2.1** e testar se existe associação entre as frequências dos dois tipos de defeitos equivale a testar a hipótese $H_0 : \theta = 0$.

Exemplo 2.5 - Considere agora os dados das **Tabelas 1.1** a **1.4** como no **Exemplo 2.2** e suponha novamente $Cov(Y_{i1}, Y_{i2}) = \theta^2$, $i = 1, \ldots, n$. Neste caso, a matriz $\mathbf{X}$ também tem estrutura idêntica à do **Exemplo 2.2**. A hipótese de igualdade do número médio de defeitos nas quatro máquinas pode ser expressa como (2.6).

Outros exemplos de modelos de Regressão Poisson Log-Normal Multivariada e Poisson Multivariada podem ser construídos envolvendo observações concomitantes do vetor resposta e de variáveis preditoras, uma vez que as colunas da matriz $\mathbf{X}$ podem conter variáveis contínuas.

Nas Seções 2.2 e 2.3, será considerada a estimação dos parâmetros dos modelos (2.5) e (2.14) respectivamente, através do método de **MV** que geralmente é obtida iterativamente. Para isso pode-se considerar o método de Newton-Raphson (THISTED, 1988) que consiste em iterar a expressão

$$\widehat{\Theta}^{(w)} = \widehat{\Theta}^{(w-1)} - \left(\frac{\partial^2}{\partial \Theta \partial \Theta'} \log L_n(\Theta) \mid_{\Theta = \widehat{\Theta}^{(w-1)}} \right)^{-1} \left(\frac{\partial}{\partial \Theta} \log L_n(\Theta) \mid_{\Theta = \widehat{\Theta}^{(w-1)}} \right) \quad (2.15)$$

onde Θ é o vetor de parâmetros de dimensão r a ser estimado do modelo; $\Theta^{(0)}$ é um valor inicial fixo e $L_n(\Theta) = \prod_{i=1}^{n} P(\mathbf{Y}_i \mid \Theta)$ é a função de verossimilhança correspondente a uma amostra de tamanho n, com $w = 1, 2, \ldots$. O processo iterativo continua até que esteja satisfeito algum critério de parada; por exemplo,

$$D = \max_{1 \leq j \leq r} \mid \widehat{\theta}_j^{(w)} - \widehat{\theta}_j^{(w-1)} \mid < \epsilon,$$

onde $\widehat{\theta}_j^{(w)}$ e $\widehat{\theta}_j^{(w-1)}$ correspondem ao j-ésimo elemento de $\widehat{\Theta}^{(w)}$ e $\widehat{\Theta}^{(w-1)}$, respectivamente e $\epsilon > 0$ é um valor escolhido convenientemente.

A primeira derivada e a derivada parcial de segunda ordem de $\log L_n(\Theta)$ são, respectivamente

$$\frac{\partial}{\partial \Theta} \log L_n(\Theta) = \sum_{i=1}^{n} \frac{1}{P(\mathbf{Y}_i \mid \Theta)} \frac{\partial}{\partial \Theta} P(\mathbf{Y}_i \mid \Theta) \quad (2.16)$$

e

$$\begin{aligned} \frac{\partial^2}{\partial \Theta \partial \Theta'} \log L_n(\Theta) &= \sum_{i=1}^{n} \frac{1}{P(\mathbf{Y}_i \mid \Theta)} \frac{\partial^2}{\partial \Theta \partial \Theta'} P(\mathbf{Y}_i \mid \Theta) - \\ &- \frac{1}{P(\mathbf{Y}_i \mid \Theta)^2} \left(\frac{\partial}{\partial \Theta} P(\mathbf{Y}_i \mid \Theta) \right) \left(\frac{\partial}{\partial \Theta'} P(\mathbf{Y}_i \mid \Theta) \right). \end{aligned} \quad (2.17)$$

Assintoticamente, $\sqrt{n}(\widehat{\Theta} - \Theta)$ tem distribuição Normal $\left(\mathbf{0}, \mathbf{I}_{\Theta}^{-1}\right)$, onde $\mathbf{I}_{\Theta}$ é a Matriz de Informação de Fisher, dada por

$$\mathbf{I}_{\Theta} = E \left[\frac{\partial}{\partial \Theta} \log P(\mathbf{Y}, \Theta) \right] \left[\frac{\partial}{\partial \Theta} \log P(\mathbf{Y}, \Theta) \right]'. \quad (2.18)$$

Obtidos os estimadores de Θ, geralmente deseja-se fazer inferências sobre os parâmetros através de testes de hipótese do tipo

$$H_0 : \mathbf{R}\Theta = \mathbf{0} \tag{2.19}$$

onde $\mathbf{R}$ é uma matriz de constantes devidamente escolhida.

Para testar a hipótese acima, pode-se empregar, por exemplo, a estatística de Wald

$$\mathbf{W} = \widehat{\Theta}'\mathbf{R}'\left[\mathbf{R}\widehat{Var}(\widehat{\Theta})\mathbf{R}'\right]^{-1}\mathbf{R}\widehat{\Theta} \tag{2.20}$$

onde $\widehat{Var}(\widehat{\Theta})$ é um estimador consistente de $Var(\widehat{\Theta})$; sob (2.19), $\mathbf{W}$ segue assintoticamente uma distribuição $\chi^2_{(q)}$ onde q é igual ao número de linhas de $\mathbf{R}$.

2.2 Estimação em modelos de Regressão Poisson Log-Normal Bivariada

Nesta seção será considerada a estimação dos parâmetros de modelos de Regressão Poisson Log-Normal Bivariada. A função de probabilidade correspondente ao vetor $\mathbf{Y}_i$=(Y_{i1}, Y_{i2}, ..., Y_{iK})' oriunda de uma distribuição Poisson Log-Normal Multivariada é dada por

$$P(\mathbf{Y}_i \mid \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}_i) = \int_{R_+^K} \prod_{l=1}^{K} f(Y_{il} \mid \lambda_l) g(\boldsymbol{\lambda} \mid \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}_i) d\boldsymbol{\lambda} \tag{2.21}$$

com $g(\boldsymbol{\lambda} \mid \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ e $f(Y_{il} \mid \lambda_l)$ expressas conforme (1.57) e (1.58), respectivamente. A reparametrização

$$\ln(\boldsymbol{\lambda}) - \mathbf{X}_i\boldsymbol{\beta} = \mathbf{T}_i\mathbf{v}\sqrt{2} \tag{2.22}$$

implica

$$\boldsymbol{\lambda} = \exp\{\mathbf{X}_i\boldsymbol{\beta} + \mathbf{T}_i\mathbf{v}\sqrt{2}\} \tag{2.23}$$

onde $\mathbf{v} = (v_1, v_2, \ldots, v_K)'$ e $\mathbf{T}_i$ é uma matriz triangular tal que $\boldsymbol{\Sigma}_i = \mathbf{T}_i\mathbf{T}_i'$. Substituindo (2.23) em (2.21), obtém-se a função de probabilidade expressa em termos de $\boldsymbol{\beta}$ e $\mathbf{T}_i$:

$$P(\mathbf{Y}_i \mid \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{T}_i) = \pi^{-K/2} \int_{R^K} H(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{T}_i, \mathbf{v}) \exp(-\mathbf{v}'\mathbf{v}) d\mathbf{v} \tag{2.24}$$

onde

$$\begin{aligned} H(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{T}_i, \mathbf{v}) &= \exp\{\mathbf{Y}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{T}_i\mathbf{v}\sqrt{2}) - \mathbf{1}_K' \exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{T}_i\mathbf{v}\sqrt{2}) - \\ &\quad -\mathbf{1}_K' \ln[\Gamma(\mathbf{Y}_i + \mathbf{1}_K)]\}. \end{aligned} \tag{2.25}$$

No caso bivariado ($K = 2$) com t populações e supondo uma amostra de tamanho n e $\mathbf{T}_1 = \mathbf{T}_2 = \ldots = \mathbf{T}_n = \mathbf{T}$, os parâmetros a serem estimados são $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_t')'$, onde $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j})'$, $j = 1, \ldots, t$ e

$$\mathbf{T} = \begin{pmatrix} \tau_{11} & 0 \\ \tau_{21} & \tau_{22} \end{pmatrix},$$

cujos elementos para efeito de simplicação, foram dispostos num único vetor $\boldsymbol{\Theta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_t', \tau_{11}, \tau_{21}, \tau_{22})$.

As expressões doravante apresentadas referem-se a uma observação de contagens bivariadas $\mathbf{Y}_i$, sendo suprimido, para efeito de simplificação de notação, o índice i.

Lembrando que

$$\frac{\partial}{\partial \boldsymbol{\Theta}} P(\mathbf{Y}) = \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y}) \\ \frac{\partial}{\partial \mathbf{T}^*} P(\mathbf{Y}) \end{bmatrix}, \tag{2.26}$$

com $\mathbf{T}^* = (\tau_{11}, \tau_{21}, \tau_{22})'$,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y}) &= \frac{\partial}{\partial \boldsymbol{\beta}} \pi^{-1} \int_{R^2} H(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{T}, \mathbf{v}) \exp\{\mathbf{v}'\mathbf{v}\} d\mathbf{v} = \\ &= \pi^{-1} \int_{R^2} H(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{T}, \mathbf{v}) \exp\{-\mathbf{v}'\mathbf{v}\} \times \\ &\quad \times \left[\mathbf{X}'\mathbf{Y} - \mathbf{X}'\exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\mathbf{v}\sqrt{2})\right] d\mathbf{v} \end{aligned} \tag{2.27}$$

que após manipulações algébricas pode ser escrita na forma

$$\frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y}) = \mathbf{X}' \left[\mathbf{Y} P(\mathbf{Y}) - (\mathbf{Y} + \mathbf{1}_2) \# \mathbf{A} \right] \tag{2.28}$$

onde # é um operador tal que, se **Z**=**R**#**S**, então o elemento (i,j) de **Z** é igual ao produto do elemento (i,j) da matriz **R** pelo elemento (i,j) da matriz **S** com

$$\mathbf{A} = \begin{bmatrix} P(\mathbf{Y} + \mathbf{E_1}) \\ P(\mathbf{Y} + \mathbf{E_2}) \end{bmatrix},$$

onde $\mathbf{E_1}=(1,0)'$, $\mathbf{E_2}=(0,1)'$ e

$$\begin{aligned} \frac{\partial}{\partial \mathbf{T}^*} P(\mathbf{Y}) &= \frac{\partial}{\partial \mathbf{T}^*} \pi^{-1} \int_{R^2} H(\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{T}, \mathbf{v}) \exp(-\mathbf{v}'\mathbf{v}) d\mathbf{v} = \\ &= \pi^{-1} \int_{R^2} H(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{T}, \mathbf{v}) \exp(-\mathbf{v}'\mathbf{v}) \mathbf{b}(\mathbf{v}) d\mathbf{v} \end{aligned} \tag{2.29}$$

com

$$\mathbf{b}(\mathbf{v}) = \begin{bmatrix} \left(\mathbf{E_1}'\mathbf{Y} - \exp\{\mathbf{E_1}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\mathbf{v}\sqrt{2})\}\right) \mathbf{E_1}'\mathbf{v}\sqrt{2} \\ \left(\mathbf{E_2}'\mathbf{Y} - \exp\{\mathbf{E_2}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\mathbf{v}\sqrt{2})\}\right) \mathbf{E_1}'\mathbf{v}\sqrt{2} \\ \left(\mathbf{E_2}'\mathbf{Y} - \exp\{\mathbf{E_2}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\mathbf{v}\sqrt{2})\}\right) \mathbf{E_2}'\mathbf{v}\sqrt{2} \end{bmatrix}. \tag{2.30}$$

Integrando por partes, a expressão (2.29) pode ser escrita na forma

$$\frac{\partial}{\partial \mathbf{T}^*} P(\mathbf{Y}) = \begin{bmatrix} \mathbf{E_1}'\mathbf{T}'\mathbf{M}\mathbf{E_1} \\ \mathbf{E_1}'\mathbf{T}'\mathbf{M}'\mathbf{E_2} \\ \mathbf{E_2}'\mathbf{T}'\mathbf{M}\mathbf{E_2} \end{bmatrix} \tag{2.31}$$

onde

$$\mathbf{M} = \mathbf{Y}\mathbf{Y}'P(\mathbf{Y}) - \mathbf{Y}(\mathbf{Y}' + \mathbf{1}_2')\#\mathbf{A}' - \mathbf{A}\#(\mathbf{Y} + \mathbf{1}_2)\mathbf{Y}' - diag\left[(\mathbf{Y} + \mathbf{1}_2)\#\mathbf{A}\right] + \mathbf{B} \tag{2.32}$$

e

$$\mathbf{B} = \begin{bmatrix} (Y_1+1)(Y_1+2)P(\mathbf{Y}+2\mathbf{E_1}) & (Y_1+1)(Y_2+1)P(\mathbf{Y}+\mathbf{1}_2) \\ (Y_1+1)(Y_2+1)P(\mathbf{Y}+\mathbf{1}_2) & (Y_2+1)(Y_2+2)P(\mathbf{Y}+2\mathbf{E_2}) \end{bmatrix}. \quad (2.33)$$

As derivadas parciais de segunda ordem de (2.24) em relação a $\boldsymbol{\Theta}$ são

$$\frac{\partial^2}{\partial\boldsymbol{\Theta}'\partial\boldsymbol{\Theta}}P(\mathbf{Y}) = \begin{bmatrix} \frac{\partial^2}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}}P(\mathbf{Y}) & \frac{\partial^2}{\partial\boldsymbol{\beta}'\partial\mathbf{T}^*}P(\mathbf{Y}) \\ \frac{\partial^2}{\partial\mathbf{T}^{*\prime}\partial\boldsymbol{\beta}}P(\mathbf{Y}) & \frac{\partial^2}{\partial\mathbf{T}^{*\prime}\partial\mathbf{T}^*}P(\mathbf{Y}) \end{bmatrix} \quad (2.34)$$

onde

$$\begin{aligned} &\frac{\partial^2}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}}P(\mathbf{Y}) = \\ &= \pi^{-1}\int_{R^2} H(\mathbf{Y},\mathbf{X}\boldsymbol{\beta},\mathbf{T},\mathbf{v})\exp\{-\mathbf{v}'\mathbf{v}\}\{\left[\mathbf{X}'\mathbf{Y}-\mathbf{X}'\mathbf{exp}(\mathbf{X}\boldsymbol{\beta}+\mathbf{T}\mathbf{v}\sqrt{2})\right]\times \\ &\times\left[\mathbf{X}'\mathbf{Y}-\mathbf{X}'\mathbf{exp}(\mathbf{X}\boldsymbol{\beta}+\mathbf{T}\mathbf{v}\sqrt{2})\right]'-\mathbf{X}'diag\left[\mathbf{exp}(\mathbf{X}\boldsymbol{\beta}+\mathbf{T}\mathbf{v}\sqrt{2})\right]\mathbf{X}\}d\mathbf{v} \end{aligned} \quad (2.35)$$

que após manipulações algébricas pode ser escrita como

$$\begin{aligned} \frac{\partial^2}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}}P(\mathbf{Y}) &= \mathbf{X}'\{\mathbf{Y}\mathbf{Y}'P(\mathbf{Y})-\mathbf{Y}(\mathbf{Y}'+\mathbf{1}_2')\#\mathbf{A}'-\mathbf{A}\#(\mathbf{Y}+\mathbf{1}_2)\mathbf{Y}'- \\ &-diag[(\mathbf{Y}+\mathbf{1}_2)\#\mathbf{A}]+\mathbf{B}\}\mathbf{X}. \end{aligned} \quad (2{:}36)$$

$$\frac{\partial^2}{\partial\boldsymbol{\beta}'\partial\mathbf{T}^*}P(\mathbf{Y}) = \begin{bmatrix} \tau_{11}\mathbf{C}+\tau_{21}\mathbf{D} \\ \tau_{11}\mathbf{D}+\tau_{21}\mathbf{E} \\ \tau_{22}\mathbf{E} \end{bmatrix} \quad (2.37)$$

onde

$$\begin{aligned}\mathbf{C} &= Y_1^2\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y})-(2Y_1+1)(Y_1+1)\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{E_1})+\\ &+(Y_1+1)(Y_1+2)\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+2\mathbf{E_1}),\end{aligned} \tag{2.38}$$

$$\begin{aligned}\mathbf{D} &= Y_1Y_2\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y})-Y_1(Y_2+1)\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{E_2})-Y_2(Y_1+1)\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{E_1})+\\ &+(Y_1+1)(Y_2+1)\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{1}_2)\end{aligned} \tag{2.39}$$

e

$$\begin{aligned}\mathbf{E} &= Y_2^2\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y})-(2Y_2+1)(Y_2+1)\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{E_2})+\\ &+(Y_2+1)(Y_2+2)\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+2\mathbf{E_2}).\end{aligned} \tag{2.40}$$

Os valores de

$$\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{E_1}),\ \frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+2\mathbf{E_1}),\ \frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{E_2}),$$

$$\frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+2\mathbf{E_2})\ \text{e}\ \frac{\partial}{\partial\boldsymbol{\beta}'}P(\mathbf{Y}+\mathbf{1}_2)$$

de (2.38), (2.39) e (2.40) são obtidos aplicando-se a expressão (2.28) adequadamente.

A segunda derivada de P($\mathbf{Y}$) com relação a $\mathbf{T}^*$ é igual a

$$\frac{\partial^2}{\partial\mathbf{T}^{*\prime}\partial\mathbf{T}^*}P(\mathbf{Y})=\begin{bmatrix}\tau_{11}\mathbf{F}+\tau_{21}\mathbf{G}\\ \tau_{11}\mathbf{G}+\tau_{21}\mathbf{H}\\ \tau_{22}\mathbf{H}\end{bmatrix}+\begin{bmatrix}\mathbf{E_1}'\mathbf{M} & 0\\ \mathbf{E_2}'\mathbf{M} & 0\\ \mathbf{0}_2 & \mathbf{M}_{2,2}\end{bmatrix} \tag{2.41}$$

onde $\mathbf{M}_{i,j}$ é o elemento (i,j) da matriz $\mathbf{M}$, $\mathbf{0_2}$ é um vetor de dimensão 2 cujos elementos são iguais a zero,

$$\begin{aligned}
\mathbf{F} &= Y_1^2 \frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}) - (2Y_1+1)(Y_1+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{E_1}) + \\
&+ (Y_1+1)(Y_1+2)\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+2\mathbf{E_1}),
\end{aligned} \tag{2.42}$$

$$\begin{aligned}
\mathbf{G} &= Y_1 Y_2 \frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}) - Y_1(Y_2+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{E_2}) - \\
&- Y_2(Y_1+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{E_1}) + \\
&+ (Y_1+1)(Y_2+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{1}_2)
\end{aligned} \tag{2.43}$$

e

$$\begin{aligned}
\mathbf{H} &= Y_2^2 \frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}) - (2Y_2+1)(Y_2+1)\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{E_2}) + \\
&+ (Y_2+1)(Y_2+2)\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+2\mathbf{E_2}).
\end{aligned} \tag{2.44}$$

Os valores de

$$\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{E_1}),\ \frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+2\mathbf{E_1}),\ \frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{E_2}),$$

$$\frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+2\mathbf{E_2}) \text{ e } \frac{\partial}{\partial \mathbf{T}^{*\prime}} P(\mathbf{Y}+\mathbf{1}_2)$$

de (2.42), (2.43) e (2.44) são obtidos aplicando-se a expressão (2.31) adequadamente.

Considerando a matriz

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} \otimes \mathbf{X}_1 \\ \mathbf{1}_{n_2} \otimes \mathbf{X}_2 \\ \cdots \\ \mathbf{1}_{n_t} \otimes \mathbf{X}_t \end{pmatrix} \tag{2.45}$$

onde

$$\mathbf{X}_j = \left[\begin{array}{cccccccc} \mathbf{Z}_1 & \mathbf{Z}_2 & \ldots & \mathbf{Z}_{j-1} & \mathbf{I}_2 & \mathbf{Z}_{j+1} & \ldots & \mathbf{Z}_t \end{array} \right]$$

j=1, 2, ..., t; $\mathbf{Z}_k = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $k = 1, 2, \ldots, j-1, j+1, \ldots, t$ e supondo os elementos de $\mathbf{T}^*$ conhecidos, a Matriz de Informação de Fisher de $\Theta^* = (\beta_1', \beta_2', \ldots, \beta_t')'$ é dada por

$$\mathbf{I}_{\Theta^*} = \begin{pmatrix} \mathbf{L}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_2 & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{L}_t \end{pmatrix} \tag{2.46}$$

onde

$$\mathbf{L}_j = \begin{pmatrix} \alpha_{1j} - \alpha_{1j}^2 \exp(\sigma_{11j}) + A_j^* & -\alpha_{1j}\alpha_{2j} \exp(\sigma_{12j}) + B_j^* \\ -\alpha_{1j}\alpha_{2j} \exp(\sigma_{12j}) + B_j^* & \alpha_{2j} - \alpha_{2j}^2 \exp(\sigma_{22j}) + C_j^* \end{pmatrix}, \tag{2.47}$$

lembrando que $\alpha_{ij} = E(Y_{ij})$, com

$$A_j^* = \sum_{Y_{1j}=0}^{\infty} \sum_{Y_{2j}=0}^{\infty} (Y_{1j}+1)^2 \frac{P(\mathbf{Y}_j + \mathbf{E_1})^2}{P(\mathbf{Y}_j)},$$

$$B_j^* = \sum_{Y_{1j}=0}^{\infty} \sum_{Y_{2j}=0}^{\infty} (Y_{1j}+1)(Y_{2j}+1) \frac{P(\mathbf{Y}_j + \mathbf{E_1})P(\mathbf{Y}_j + \mathbf{E_2})}{P(\mathbf{Y}_j)},$$

$$C_j^* = \sum_{Y_{1j}=0}^{\infty} \sum_{Y_{2j}=0}^{\infty} (Y_{2j}+1)^2 \frac{P(\mathbf{Y}_j + \mathbf{E_2})^2}{P(\mathbf{Y}_j)}.$$

Também pode-se observar que

$$Var(\widehat{\Theta^*}) = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{V}_t \end{pmatrix} \tag{2.48}$$

com

$$\mathbf{V}_j = \frac{1}{D_j^*}\begin{pmatrix} \alpha_{2j} - \alpha_{2j}^2 \exp(\sigma_{22j}) + C_j^* & \alpha_{1j}\alpha_{2j}\exp(\sigma_{12j}) - B_j^* \\ \alpha_{1j}\alpha_{2j}\exp(\sigma_{12j}) - B_j^* & \alpha_{1j} - \alpha_{1j}^2 \exp(\sigma_{11j}) + A_j^* \end{pmatrix} \tag{2.49}$$

onde D_j^* é igual ao determinante de $\mathbf{L}_j$, $j = 1, \ldots, t$.

2.3 Estimação em modelos de Regressão Poisson Bivariada

Nesta seção será considerada a estimação dos parâmetros do modelo de Regressão Poisson Bivariada pelo método **MV**.

A função de probabilidade de um vetor $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iK})'$ segundo uma distribuição Poisson Multivariada é dada pela expressão (1.8). Para o caso bivariado ($K = 2$), considere a reparametrização

$$\begin{pmatrix} \ln(\boldsymbol{\lambda}_i^*) \\ \lambda_{11_i} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta} \\ \theta_i^2 \end{pmatrix} \tag{2.50}$$

que implica

$$\begin{pmatrix} \boldsymbol{\lambda}_i^* \\ \sqrt{\lambda_{11_i}} \end{pmatrix} = \begin{pmatrix} \exp(\mathbf{X}_i\boldsymbol{\beta}) \\ \theta_i \end{pmatrix} \tag{2.51}$$

com $\boldsymbol{\lambda}_i^* = (\lambda_{10i}, \lambda_{01i})'$, $i = 1, \ldots, n$. Para efeito de simplificação de notação, os parâmetros de uma distribuição Poisson Bivariada λ_{10}, λ_{01} e λ_{11} serão doravante denotados λ_1, λ_2 e λ respectivamente. A expressão (1.10) correspondente à função de probabilidade de $\mathbf{Y}_i$=$(Y_{i1}, Y_{i2})'$ reescrita em termos de $\boldsymbol{\beta}$ e θ_i resulta em

$$\begin{aligned} P(Y_{i1}, Y_{i2}) &= |\det\left(\frac{\partial \boldsymbol{\Lambda}_i}{\partial \boldsymbol{\Theta}_i}\right)| \sum_{u=0}^{min(Y_{i1},Y_{i2})} \exp\left[-\mathbf{1}_2'\exp(\mathbf{X}_i\boldsymbol{\beta}) + \theta_i^2\right] \times \\ &\times \frac{[\mathbf{E_1}'(\exp\mathbf{X}_i\boldsymbol{\beta}) - \theta_i^2]^{(Y_{i1}-u)}}{(Y_{i1}-u)!} \frac{[\mathbf{E_2}'(\exp\mathbf{X}_i\boldsymbol{\beta}) - \theta_i^2]^{(Y_{i2}-u)}}{(Y_{i2}-u)!} \frac{\theta_i^{2u}}{u!} \end{aligned} \tag{2.52}$$

com $\boldsymbol{\Theta}_i = (\boldsymbol{\beta}', \theta_i)'$, $\boldsymbol{\Lambda}_i = (\boldsymbol{\lambda}_i'^{*}, \lambda_i)'$.

Para o caso bivariado ($K = 2$) com t populações e supondo uma amostra de tamanho n e $\theta_1 = \ldots = \theta_n = \theta$, os parâmetros a serem estimados foram dispostos num único vetor $\boldsymbol{\Theta} = (\boldsymbol{\beta}', \theta)'$ com $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_t')'$. As expressões referentes à primeira derivada e à derivada parcial de segundo ordem de (2.52), necessárias para a obtenção das estimativas dos parâmetros pelo método de Newton-Raphson e apresentadas a seguir, referem-se a uma observação de contagens bivariadas, sendo suprimido, para efeito de simplificação de notação, o índice i. Lembrando que

$$\frac{\partial}{\partial \boldsymbol{\Theta}} P(\mathbf{Y}) = \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y}) \\ \frac{\partial}{\partial \theta} P(\mathbf{Y}) \end{bmatrix}, \tag{2.53}$$

onde

$$\frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y}) = \mathbf{X}' diag \left(\frac{\partial}{\partial \boldsymbol{\lambda}^*} P(\mathbf{Y}) \right) \mathbf{exp}(\mathbf{X}\boldsymbol{\beta}) \tag{2.54}$$

com $\frac{\partial}{\partial \boldsymbol{\lambda}^*} P(\mathbf{Y})$ expressas em (1.30) e (1.31) e

$$\frac{\partial}{\partial \theta} P(\mathbf{Y}) = \frac{\partial}{\partial \lambda} P(\mathbf{Y}) \frac{\partial \lambda}{\partial \theta} \tag{2.55}$$

onde $\frac{\partial}{\partial \lambda} P(\mathbf{Y})$ está expressa conforme (1.32) e substituindo esses resultados em (2.55) resulta

$$\frac{\partial}{\partial \theta} P(\mathbf{Y}) = [P(\mathbf{Y}) - P(\mathbf{Y} - \mathbf{E_1}) - P(\mathbf{Y} - \mathbf{E_2}) + P(\mathbf{Y} - \mathbf{1}_2)]\, 2\theta. \tag{2.56}$$

As derivadas parciais de segunda ordem de (2.52) em relação a $\boldsymbol{\Theta}$ são

$$\frac{\partial^2}{\partial \boldsymbol{\Theta}' \partial \boldsymbol{\Theta}} P(\mathbf{Y}) = \begin{bmatrix} \frac{\partial^2}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} P(\mathbf{Y}) & \frac{\partial^2}{\partial \boldsymbol{\beta}' \partial \theta} P(\mathbf{Y}) \\ \frac{\partial^2}{\partial \theta \partial \boldsymbol{\beta}} P(\mathbf{Y}) & \frac{\partial^2}{\partial^2 \theta} P(\mathbf{Y}) \end{bmatrix} \tag{2.57}$$

onde

i)

$$\frac{\partial^2}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} P(\mathbf{Y}) \quad = \quad \mathbf{X}'\mathbf{L}\mathbf{X} \tag{2.58}$$

com $\mathbf{L} = diag\left[\frac{\partial}{\partial \boldsymbol{\lambda}^*} P(\mathbf{Y})\right] diag\,[\exp(\mathbf{X}\boldsymbol{\beta})] + \frac{\partial^2}{\partial \boldsymbol{\lambda}^{*\prime} \partial \boldsymbol{\lambda}^*} P(\mathbf{Y})\,[\exp(\mathbf{X}\boldsymbol{\beta})]\,[\exp(\mathbf{X}\boldsymbol{\beta})]'$,

ii)

$$\frac{\partial^2}{\partial \boldsymbol{\beta}' \partial \theta} P(\mathbf{Y}) = 2\theta \mathbf{S} \tag{2.59}$$

com $\mathbf{S} = \left[\frac{\partial}{\partial \boldsymbol{\beta}'} P(\mathbf{Y}) - \frac{\partial}{\partial \boldsymbol{\beta}'} P(\mathbf{Y} - \mathbf{E}_1) - \frac{\partial}{\partial \boldsymbol{\beta}'} P(\mathbf{Y} - \mathbf{E}_2) + \frac{\partial}{\partial \boldsymbol{\beta}'} P(\mathbf{Y} - \mathbf{1}_2)\right]$.

Os valores de

$$\frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y} - \mathbf{E}_1), \frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y} - \mathbf{E}_2) \text{ e } \frac{\partial}{\partial \boldsymbol{\beta}} P(\mathbf{Y} - \mathbf{1}_2)$$

de (2.59) podem ser obtidos, aplicando-se a expressão (2.54) adequadamente.

iii)

$$\begin{aligned} \frac{\partial^2}{\partial^2 \theta} P(\mathbf{Y}) \quad &= \quad \left[\frac{\partial}{\partial \lambda} P(\mathbf{Y}) - \frac{\partial}{\partial \lambda} P(\mathbf{Y} - \mathbf{E}_1) - \frac{\partial}{\partial \lambda} P(\mathbf{Y} - \mathbf{E}_2) + \frac{\partial}{\partial \lambda} P(\mathbf{Y} - \mathbf{1}_2)\right] 4\theta^2 + \\ &\quad + 2\,[P(\mathbf{Y}) - P(\mathbf{Y} - \mathbf{E}_1) - P(\mathbf{Y} - \mathbf{E}_2) + P(\mathbf{Y} - \mathbf{1}_2)]\,. \end{aligned} \tag{2.60}$$

Os valores de

$$\frac{\partial}{\partial \lambda} P(\mathbf{Y} - \mathbf{E}_1), \frac{\partial}{\partial \lambda} P(\mathbf{Y} - \mathbf{E}_2) \text{ e } \frac{\partial}{\partial \lambda} P(\mathbf{Y} - \mathbf{1}_2)$$

de (2.60) podem ser obtidos, aplicando-se a expressão (1.32) adequadamente.

Considerando a matriz de especificação $\mathbf{X}$ idêntica a (2.45) e supondo θ conhecido, resulta que $\Theta^* = \ln(\Lambda^*)$, com $\Theta^* = (\beta'_1, \beta'_2, \ldots, \beta'_t)'$ e $\Lambda^* = (\lambda'^*_1, \lambda'^*_2, \ldots, \lambda'^*_t)'$. Pode-se observar que

$$Var(\widehat{\Theta^*}) = \frac{\partial \Theta^*}{\partial \Lambda^*} \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A}_t \end{bmatrix} \frac{\partial \Theta^{*\prime}}{\partial \Lambda^*} \qquad (2.61)$$

com $\mathbf{A}_j = \frac{1}{n_j}\begin{bmatrix} \lambda_{1j} & \lambda \\ \lambda & \lambda_{2j} \end{bmatrix}$ (HOLGATE, 1964). A expressão (2.61), após manipulações algébricas, pode ser escrita como

$$Var(\widehat{\Theta^*}) = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{B}_t \end{bmatrix} \qquad (2.62)$$

com $\mathbf{B}_j = \frac{1}{n_j}\begin{bmatrix} \lambda_{1j}^{-1} & \frac{\lambda}{\lambda_{1j}\lambda_{2j}} \\ \frac{\lambda}{\lambda_{1j}\lambda_{2j}} & \lambda_{2j}^{-1} \end{bmatrix}$.

2.4 Aspectos computacionais

A obtenção de estimadores de **MV** consiste em achar um vetor $\widehat{\Theta}$ tal que $L_n(\Theta)$ seja máxima, que geralmente equivale a obter a solução do sistema de equações

$$\frac{\partial}{\partial \Theta} \log L_n(\Theta) = \mathbf{0}.$$

Uma das maneiras de verificar se a solução encontrada é um ponto de sela, de mínimo ou de máximo é analisar $\partial^2/\partial\Theta\partial\Theta'\{\log L_n(\boldsymbol{\theta})\}$, que neste último caso não deve ser uma matriz negativa definida.

Outras formas de busca podem ser utilizadas para procurar a solução do sistema de equações, como calcular o valor de $\partial/\partial\boldsymbol{\Theta}\{\log L_n(\boldsymbol{\theta})\}$ (THISTED, 1988) para valores de θ_i igualmente espaçados, $\theta_i \in \boldsymbol{\Theta}$ mantendo outros componentes θ_j, $j \neq i$ constantes. Em casos vetoriais, a busca se torna bastante trabalhosa devido ao grande número de combinações de busca.

Um esquema iterativo para obtenção do estimador de **MV**, consta basicamente de três componentes:

- um método para decidir o valor inicial,
- um método para obter a próxima iteração a partir da anterior
- um método para decidir a parada do processo iterativo.

O sucesso do processo iterativo depende muito da escolha adequada do valor inicial. Ele deve ser escolhido de modo a garantir que seu valor esteja suficientemente próximo da solução, pois se estiver afastado, o processo iterativo pode divergir. É recomendável usar estimadores consistentes (os estimadores de **MM**, por exemplo) como valores iniciais para $\boldsymbol{\beta}$ e $\boldsymbol{\Sigma}$. Ajustes devem ser feitos quando o valor inicial de $\boldsymbol{\Sigma}$ não for uma matriz positiva-definida. No caso da Regressão Poisson Log-Normal Multivariada, este ajuste é necessário quando ocorrerem casos de subdispersão, isto é, quando a variância amostral for menor que a média amostral. Nesta situação, AITCHISON; HO (1989) sugerem estabelecer um número positivo pequeno como valor inicial para σ_{ii} e tomar σ_{ij} e σ_{ji} iguais a zero. Mesmo não ocorrendo a subdispersão, a estimativa inicial de $\boldsymbol{\Sigma}$ pode não ser positiva-definida. Nestes casos, AITCHISON; HO (1989) sugerem usar o ajuste de Levenberg-Marquadt (THISTED, 1988), que consiste em adicionar constantes aos elementos da diagonal da estimativa inicial de $\boldsymbol{\Sigma}$. Outro ajuste é necessário quando o valor inicial de $\partial^2/\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}'\{\log L_n(\boldsymbol{\theta})\}$ corresponder a uma matriz negativa definida, podendo-se também utilizar o ajuste de Levenberg-Marquardt, nesse caso.

Atenção especial deve ser dada ao cálculo de $\partial^2/\partial\Theta\partial\Theta'\{\log L_n(\theta)\}$, pois essa matriz está relacionada à estimativa da variância assintótica das estimativas e qualquer erro no seu cálculo pode gerar imprecisões na obtenção do erro padrão das estimativas.

Muitas vezes, a única alternativa é iniciar com um valor arbitrário e observar o progresso das iterações, esperando que seja possível ajustar o valor inicial ou o método iterativo para obter convergência.

Para verificar a estabilidade da solução, THISTED (1988) sugere começar o processo iterativo em diferentes pontos iniciais e verificar se há convergência para o mesmo resultado.

Conforme (2.15), a w-ésima iteração do procedimento de Newton-Raphson é determinada pelo vetor

$$\mathbf{P} = \left(\frac{\partial^2}{\partial\Theta\partial\Theta'}\log L_n(\Theta)\mid_{\Theta=\widehat{\Theta}^{(w-1)}}\right)^{-1}\left(\frac{\partial}{\partial\Theta}\log L_n(\Theta)\mid_{\Theta=\widehat{\Theta}^{(w-1)}}\right)$$

que é conhecido como Passo de Newton ($\mathbf{P}$). Muitas vezes $\mathbf{P}$ pode ser *inadmissível* ou seja, conduzir a pontos distantes da solução. Uma sugestão dada por THISTED (1988) para evitar este tipo de problema é concretizar a iteração utilizando $k\mathbf{P}$, com $k < 1$ e avaliar o comportamento do processo iterativo.

No caso de um parâmetro, THISTED (1988) cita dois critérios de parada: absoluto e relativo. O primeiro é mais conveniente quando a solução está mais próxima do valor zero, já que o segundo critério poderia gerar problemas numéricos. O critério relativo é mais satisfatório quando a solução é um valor grande. No caso de um parâmetro, é possível que $\mid \theta^{(w)} - \theta^{(w-1)} \mid$ seja pequeno embora o valor de $\partial/\partial\theta\{\log L_n(\theta)\mid_{\theta=\widehat{\theta}^{(w)}}\}$ não esteja próximo de zero. Na prática, a situação inversa é mais comum, ou seja, o valor de $\partial/\partial\theta\{\log L_n(\theta)\mid_{\theta=\widehat{\theta}^{(w)}}\}$ está próximo de zero e a diferença $\mid \theta^{(w)} - \theta^{(w-1)} \mid$ não satisfaz o critério de convergência. É aconselhável acompanhar a sequência das iterações, verificando se $\theta^{(w)}$ está suficientemente próximo de $\theta^{(w-1)}$ e se $\partial/\partial\theta\{\log L_n(\theta)\mid_{\theta=\widehat{\theta}^{(w)}}\}$ está próximo de zero, para decidir se a convergência foi alcançada.

No caso de estimação de um vetor de parâmetros, THISTED (1988) sugere definir um critério de parada absoluto ou relativo utilizando alguma medida derivada da norma do vetor $\mathbf{d}=\Theta^{(w)} - \Theta^{(w-1)}$.

Segundo THISTED (1988), um processo iterativo é dito não convergente quando depois de um especificado número de iterações, o critério de parada não foi alcançado. No caso de o processo iterativo divergir, isto é facilmente verificável; basta comparar os valores de $\partial/\partial\theta\{\log L_n(\theta^{(w)})\}$ e $\partial/\partial\theta\{\log L_n(\theta^{(0)}\}$. Maiores dificuldades são encontradas nos casos onde ocorrem oscilações ou ciclos. Isto pode ocorrer quando a terceira derivada do $\log L_n(\theta)$ for mal comportada perto da solução.

Além dos aspectos envolvidos no esquema iterativo para cálculo dos estimadores de **MV**, um outro igualmente importante deve ser levado em conta neste trabalho. Refere-se ao cálculo das probabilidades, uma vez que a função de probabilidade da distribuição Poisson Log-Normal Multivariada não apresenta uma expressão explícita, sendo necessário utilizar métodos de integração numérica na sua obtenção.

Segundo NAYLOR; SMITH (1982), integrais da forma

$$\int_{-\infty}^{+\infty} f(v)\exp\{-v^2\}dv \tag{2.63}$$

podem ser aproximadas por

$$\int_{-\infty}^{+\infty} f(v)\exp\{-v^2\}dv = \sum_{i=1}^{n} \alpha_i^{(n)} f(v_i^{(n)}) + R_n \tag{2.64}$$

onde α_i é um número de Christoffel, v_i são as raízes dos polinômios de Hermite e

$$R_n = \frac{\pi^{1/2} f(\epsilon)^{(2n)}}{2^n(2n)(2n-1)\dots(n+2)(n+1)}$$

para algum ϵ, - $\infty < \epsilon < \infty$.

As integrais múltiplas como aquela da expressão (2.24) podem ser aproximadas por

$$\int_{R^K} H(\mathbf{Y},\mathbf{X},\boldsymbol{\beta},\mathbf{T},\mathbf{v})\exp\{(-\mathbf{v}'\mathbf{v})\}d\mathbf{v} =$$

$$= \textstyle\sum_{iK=1}^{n} \cdots \sum_{i1=1}^{n} \alpha_{iK}\dots\alpha_{i1} H(\mathbf{Y},\mathbf{X},\boldsymbol{\beta},\mathbf{T},v_{i1}^{(1)},\dots,v_{iK}^{(K)}) \tag{2.65}$$

onde α_{ij} são os números de Christoffel e $v_{ij}^{(j)}$ são as raízes dos polinômios de Hermite, $j = 1, \ldots, K$. Segundo AITCHISON; HO (1989), boas aproximações são obtidas com polinômios de grau 5.

Capítulo 3

Análise dos exemplos

3.1 Introdução

Neste capítulo estão apresentadas algumas aplicações dos métodos descritos anteriormente aos dados do **Exemplo 1.1** correspondentes à comparação das quatro craqueadeiras com relação à frequência de defeitos de cada tipo e à associação entre as frequências de defeitos dos dois tipos. Como a procedência de duas máquinas é de um fabricante (A) e as outras duas de um outro fabricante (B), neste contexto, há interesse em estudar a homogeneidade das distribuições de respostas entre as máquinas de cada um dos fabricantes bem como entre as máquinas dos dois fabricantes.

Dadas as características das respostas, foi analisada a adequação de distribuições Poisson independentes, distribuições Poisson Bivariada e distribuições Poisson Log-Normal Bivariada como modelos probabilísticos para representação dos dados através de testes de aderência baseados na estatística Qui-quadrado de Pearson, cujos resultados estão apresentados na **Tabela 3.1**.

Tabela 3.1 - Níveis descritivos referentes a testes de aderência aplicados aos dados do **Exemplo 1.1**.

	Poisson		
Máquina	Bivariada	Log-Normal Bivariada	independentes
A_1	0.8904	0.1998	0.2821
A_2	0.1050	0.1654	0.0328
B_1	0.2890	0.0170	0.6092
B_2	0.7231	0.4153	0.8274

Os resultados da **Tabela 3.1** não indicam com segurança uma distribuição de probabilidade mais adequada aos dados. Qualquer uma das três distribuições consideradas pode ser utilizada como modelo probabilístico para os dados da máquina A_1, ao passo que com relação à máquina A_2, duas das três distribuições podem ser consideradas (apenas distribuições Poisson independentes não poderiam ser utilizadas). Análises semelhantes podem ser feitas com relação às máquinas B_1 e B_2. Uma das possíveis causas destes resultados pode ser o fato de os dados estarem agrupados ou ao fato de haver grande número de caselas com frequência zero.

Algumas recomendações baseadas nos exemplos apresentados em AITCHISON; HO (1989) podem ser usadas como guia na escolha entre as distribuições Poisson Log-Normal Bivariada e Poisson Bivariada numa análise estatística de contagens. É recomendável utilizar uma distribuição Poisson Log-Normal Bivariada em casos onde há sobredispersão amostral moderada (torno de 30%) em cada contagem e um coeficiente de correlação amostral moderado (em torno de 0.25) assim como em casos com sobredispersão alta (superior a 70 %) em cada contagem e um coeficiente de correlação amostral negativo moderado (em torno de -0.20). No entanto, é recomendável utilizar uma distribuição Poisson Bivariada em casos com subdispersão amostral e/ou uma sobredispersão amostral baixa em cada contagem e um coeficiente de correlação amostral positivo em torno de zero.

Pelas recomendações citadas, não fica evidente qual deve ser a distribuição de probabilidade mais adequada aos dados do **Exemplo 1.1**, visto não ocorrer subdispersão amostral ou sobredispersão amostral em todas as contagens (ver **Tabela 3.2**). Nesse contexto, foram considerados modelos de Regressão Poisson Log-Normal Bivariada e modelos de Regressão Poisson Bivariada para tentar responder às questões de interesse. Esses modelos podem ser respectivamente expressos nas formas (2.5) e (2.14).

Para ajustar esses modelos foi utilizado o método de **MV** através das iterações de Newton-Raphson descrito em (2.15) e implementado através de uma subrotina desenvolvida na linguagem IML do SAS (ver Apêndices A e B).

Para simplificar a notação, será utilizado um único índice para representar as observações correspondentes às diferentes máquinas e fabricantes. Nesse sentido, os índices $i = 1$ a $i = 130$ correspondem às observações da máquina 1 do fabricante A; $i = 131$ a $i = 230$ correspondem às observações da máquina 2 do fabricante A; $i = 231$ a $i = 330$ correspondem às observações da máquina 1 do fabricante B e $i = 331$ a $i = 430$ correspondem às observações da máquina 2 do fabricante B.

Na Seção 3.2 estão apresentados os detalhes sobre os modelos de Regressão Poisson Log-Normal Bivariada. Na Seção 3.3 os dados foram reanalisados segundo os modelos de Regressão Poisson Bivariada. Neste caso, também foram obtidas as estimativas da covariância pelo Método de **MV**, Método dos Momentos (**MM**), Double-Zero (**DZ**) e Even-Point (**EP**). Na Seção 3.4 estão comparados os ajustes dos diferentes modelos discutidos nas Seções 3.2 e 3.3 através do Critério de Informação de Akaike (**AIC**) [KENDALL; STUART; ORD (1983)].

3.2 Utilização de modelos de Regressão Poisson Log-Normal Bivariada

Nesta seção estão apresentados os resultados obtidos com os modelos de Regressão Poisson Log-Normal Bivariada considerando os dados do **Exemplo 1.1**.

Inicialmente será considerado um modelo da forma (2.5), com n=430,

$$\mathbf{X}_1 = \ldots = \mathbf{X}_{130} = \mathbf{X}_{A_1},$$

$$\mathbf{X}_{131} = \ldots = \mathbf{X}_{230} = \mathbf{X}_{A_2},$$

$$\mathbf{X}_{231} = \ldots = \mathbf{X}_{330} = \mathbf{X}_{B_1},$$

$$\mathbf{X}_{331} = \ldots = \mathbf{X}_{430} = \mathbf{X}_{B_2}$$

onde

$$\mathbf{X}_{A_1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{X}_{A_2} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{X}_{B_1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{X}_{B_2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

e

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= \ldots = \boldsymbol{\Sigma}_{130} = \boldsymbol{\Sigma}_{A_1}, \\ \boldsymbol{\Sigma}_{131} &= \ldots = \boldsymbol{\Sigma}_{230} = \boldsymbol{\Sigma}_{A_2}, \\ \boldsymbol{\Sigma}_{231} &= \ldots = \boldsymbol{\Sigma}_{330} = \boldsymbol{\Sigma}_{B_1}, \\ \boldsymbol{\Sigma}_{331} &= \ldots = \boldsymbol{\Sigma}_{430} = \boldsymbol{\Sigma}_{B_2}. \end{aligned} \tag{3.1}$$

Os valores iniciais para o algoritmo utilizado na obtenção dos estimadores de **MV** foram calculados a partir de

$$\boldsymbol{\Sigma}_{jk}^{(0)} = \log\left(\frac{\mathbf{S}_{jk} - diag(\overline{\mathbf{Y}}_{jk})}{\overline{\mathbf{Y}}_{jk}\overline{\mathbf{Y}}_{jk}'} + \mathbf{1}_2\mathbf{1}_2'\right) \tag{3.2}$$

j=A, B, k=1, 2, onde $\mathbf{S}_{jk}$, $\overline{\mathbf{Y}}_{jk}$ são respectivamente a matriz de covariância amostral e o vetor de médias amostrais correspondentes à máquina j_k. Nos casos de subdispersão amostral (máquinas A_1, B_1 e B_2, (ver **Tabela 3.2**)), a utilização de (3.2) para obtenção do valor inicial para $\boldsymbol{\Sigma}_{jk}$ fica prejudicada. Nestes casos, conforme sugestão de AITCHISON; HO (1989), foram estabelecidos valores pequenos e positivos (0.01) como valores iniciais para $\sigma_{ii_{jk}}$, e valores nulos para $\sigma_{is_{jk}}$ e $\sigma_{si_{jk}}$, $i = s = 1, 2$. Com relação à máquina A_2, utilizou-se o ajuste de Levenberg-Marquadt [AITCHISON; HO (1989)], pois a expressão (3.2) gerou uma matriz negativa-definida.

Com os valores iniciais para $\boldsymbol{\Sigma}_{jk}$ definidos, obtiveram-se os valores iniciais

$$\boldsymbol{\beta}_{jk}^{(0)} = \log\left(\overline{\mathbf{Y}}_{jk}\right) - (1/2)vecdiag\left(\boldsymbol{\Sigma}_{jk}^{(0)}\right) \tag{3.3}$$

j=A, B e k=1, 2. Na **Tabela 3.2** estão as médias amostrais e suas respectivas variâncias e covariâncias observadas, utilizadas para obter os valores iniciais de $\boldsymbol{\beta}$ e $\boldsymbol{\Sigma}$ no processo iterativo.

Tabela 3.2 - Números médios de defeitos observados e correspondentes variâncias e covariâncias observadas.

	Máquina			
	A_1	A_2	B_1	B_2
Nº médio de defeitos do tipo D_1	0.4846	0.6100	0.1700	0.1600
Variância amostral	0.5928	0.6645	0.2031	0.1358
Nº médio de defeitos do tipo D_2	1.9615	2.0700	1.6600	1.300
Variância amostral	1.8977	2.6516	1.2770	1.2424
Covariância amostral	0.2668	0.5730	0.0382	−0.0181

Na **Tabela 3.3** estão resumidos os valores iniciais utilizados no processo iterativo.

Tabela 3.3 - Valores iniciais utilizados no processo iterativo para estimação de $\boldsymbol{\beta}_{jk}$ e $\boldsymbol{\Sigma}_{jk}$ sob o Modelo (3.1).

	Máquinas			
Parâmetro	A_1	A_2	B_1	B_2
$\beta_1^{(0)}$	−0.9138	−0.6959	−2.1539	−1.8376
$\beta_2^{(0)}$	0.6687	0.5307	0.5018	0.2574
$\sigma_{11}^{(0)}$	0.3788	0.4032	0.7638	0.0100
$\sigma_{22}^{(0)}$	0.0100	0.3937	0.0100	0.0100
$\sigma_{12}^{(0)}$	0.0000	0.3742	0.0000	0.0000

As estimativas dos parâmetros com seus respectivos erros padrões estimados (obtidos de 2.18) estão indicadas na **Tabela 3.4**.

Tabela 3.4 - Estimativa dos parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\Sigma}$ e os respectivos erros padrões sob o Modelo (3.1).

	Máquinas			
Parâmetros	A_1	A_2	B_1	B_2
β_1	−0.9630 (0.1969)	−0.7260 (0.1762)	−2.2137 (0.5721)	−1.8326 (0.2500)
β_2	0.6474 (0.0694)	0.6302 (0.0925)	0.5055 (0.0781)	0.2624 (0.0877)
σ_{11}	0.4744 (0.3057)	0.4954 (0.2093)	0.8892 (1.1208)	0.0000 ($< 10^{-4}$)
σ_{12}	0.1577 (0.0894)	0.3214 (0.0907)	0.0481 (0.1315)	0.0000 ($< 10^{-4}$)
σ_{22}	0.0524 (0.0455)	0.2085 (0.0831)	0.0026 (0.0136)	0.0000 ($< 10^{-4}$)

Pode-se observar que algumas estimativas de $\sigma_{12_{jk}}$, j=A, B , k=1, 2, são muito próximas de zero (com erros padrões menores que 10^{-4}). Especificamente, as estimativas dos componentes da matriz $\boldsymbol{\Sigma}$ para máquina B_2 são todas iguais a zero, o que vem sugerir que as frequências de defeitos dos tipos D_1 e D_2 são variáveis independentes com distribuição Poisson, sendo esta suposição confirmada através do teste de aderência (ver **Tabela 3.1**).

Considere agora a hipótese

$$H_0 : \begin{cases} \boldsymbol{\Sigma}_{A_1} & = & \boldsymbol{\Sigma}_{A_2} \\ \\ \boldsymbol{\Sigma}_{B_1} & = & \boldsymbol{\Sigma}_{B_2} \end{cases} \tag{3.4}$$

que sugere que o modelo (3.1) pode ser reduzido a um outro modelo da forma (2.5), com as matrizes de especificação idênticas ao modelo inicial, porém com

$$\begin{aligned} \boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_{230} = \boldsymbol{\Sigma}_A, \\ \boldsymbol{\Sigma}_{231} = \ldots = \boldsymbol{\Sigma}_{430} = \boldsymbol{\Sigma}_B. \end{aligned} \tag{3.5}$$

O nível descritivo referente ao teste da hipótese (3.4) através da estatística de Wald da forma (2.20) é $p = 0.7417$, sendo plausível considerar esta hipótese. Diante desse resultado e utilizando procedimento análogo àqueles empregados na análise do modelo inicial (3.1), pode-se ainda testar a hipótese

$$H_0 : \boldsymbol{\Sigma}_A = \boldsymbol{\Sigma}_B. \tag{3.6}$$

O nível descritivo do teste da hipótese (3.6) é $p = 0.0547$, indicando evidências contrárias à simplificação adicional ao modelo (3.5).

As estimativas dos parâmetros do modelo reduzido (3.5) foram obtidas utilizando procedimento análogo ao empregado no ajuste do modelo inicial. Sob este modelo, as hipóteses de igualdade das taxas médias de defeitos e de homogeneidade de associação das frequências dos dois tipos de defeitos nas várias máquinas podem ser respectivamente expressas como

$$H_0 : \boldsymbol{\beta}_{A1} + \frac{1}{2} vecdiag(\boldsymbol{\Sigma}_A) = \ldots = \boldsymbol{\beta}_{B2} + \frac{1}{2} vecdiag(\boldsymbol{\Sigma}_B) \tag{3.7}$$

e

$$H_0 : \begin{cases} \sigma_{12A} = \sigma_{12B} \\ \mathbf{1}_2' \left[\boldsymbol{\beta}_{A1} + \frac{1}{2} vecdiag(\boldsymbol{\Sigma}_A)\right] = \ldots = \mathbf{1}_2' \left[\boldsymbol{\beta}_{B2} + \frac{1}{2} vecdiag(\boldsymbol{\Sigma}_B)\right]. \end{cases} \tag{3.8}$$

Os níveis descritivos dos testes das hipóteses (3.7) e (3.8) através da estatística de Wald (2.20) são menores que 0.0001, indicando evidências contrárias à validade das mesmas. Como as máquinas são de diferentes procedências, há interesse em estudar a homogeneidade das taxas médias de defeitos para as máquinas de um mesmo fabricante. Esta hipótese pode ser expressa como

$$H_0 : \boldsymbol{\beta}_{j1} = \boldsymbol{\beta}_{j2} \tag{3.9}$$

j= A, B. O nível descritivo referente ao teste da hipótese (3.9) através da estatística de Wald é $p = 0.2287$, indicando que as taxas médias de defeitos podem ser consideradas iguais para as máquinas de um mesmo fabricante. A hipótese de homogeneidade de associação das frequências dos dois tipos de defeitos em máquinas de um mesmo fabricante pode ser expressa como

$$H_0 : \mathbf{1}_2' \boldsymbol{\beta}_{j1} = \mathbf{1}_2' \boldsymbol{\beta}_{j2} \tag{3.10}$$

$j = A, B$. No entanto, sob o Modelo (3.5), a hipótese de igualdade das médias nas máquinas j_1 e j_2, $j = A, B$, implica a homogeneidade de associação entre as frequências dos dois tipos de defeitos nas mesmas máquinas (ver **Exemplo 2.2**). Assim, o resultado do teste da hipótese (3.9) indica que a associação entre as frequências dos dois tipos de defeitos são homogêneas para as máquinas de cada um dos fabricantes.

Desta forma, o modelo proposto (3.5) pode ser reduzido a um modelo da forma (2.5) com

$$\mathbf{X}_1 = \ldots = \mathbf{X}_{230} = \mathbf{X}_A,$$
$$\mathbf{X}_{231} = \ldots = \mathbf{X}_{430} = \mathbf{X}_B$$

onde

$$\mathbf{X}_A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \text{ e } \mathbf{X}_B = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

e mantendo

$$\begin{aligned} \boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_{230} = \boldsymbol{\Sigma}_A \\ \boldsymbol{\Sigma}_{231} = \ldots = \boldsymbol{\Sigma}_{430} = \boldsymbol{\Sigma}_B. \end{aligned} \tag{3.11}$$

As estimativas dos parâmetros com os respectivos erros padrões estimados (obtidos de 2.18) sob o Modelo (3.11) estão indicadas na **Tabela 3.5**.

Tabela 3.5 - Estimativas dos parâmetros com os respectivos erros padrões estimados sob o Modelo (3.11).

	Parâmetro				
Fabricante	β_1	β_2	σ_{11}	σ_{21}	σ_{22}
A	−0.8454	0.6417	0.4572	0.2265	0.1122
	(0.1314)	(0.0560)	(0.1912)	(0.0715)	(0.1260)
B	−1.8579	0.3919	0.1122	0.0053	0.0002
	(0.3382)	(0.0582)	(0.5782)	(0.0553)	(0.0048)

Essencialmente esse modelo reduzido indica que as taxas médias de defeitos e a associação entre as frequências de defeitos dos dois tipos são homogêneas para as máquinas de um mesmo fabricante.

Além da comparação do número de defeitos entre máquinas e fabricantes, outro enfoque de interesse é testar se a associação entre as frequências dos dois tipos de defeitos é nula. Esta hipótese, sob o modelo reduzido (3.11), corresponde a

$$H_0 : \sigma_{12j} = 0$$

$j = A, B$. Como foi constatada a heterogeneidade da distribuição das respostas entre os fabricantes nas análises anteriores, esta hipótese será testada separadamente para cada fabricante. Os níveis descritivos correspondentes são $p = 0.0008$ e $p = 0.4622$, respectivamente para os fabricantes A e B, sugerindo que as frequências dos dois tipos de defeitos apresentam uma associação positiva nas máquinas do fabricante A, enquanto que nas máquinas do fabricante B, não existe associação entre as frequências dos dois tipos de defeitos.

As estimativas das taxas médias de defeitos e dos coeficientes de correlação entre as frequências dos dois tipos de defeitos com os respectivos erros padrões sob o Modelo (3.11) estão na **Tabela 3.6**.

Tabela 3.6 - Estimativa das taxas médias de defeitos e dos coeficientes de correlação entre as frequências dos dois tipos de defeitos com os respectivos erros padrões sob o Modelo (3.11).

	Máquinas do Fabricante	
	A	B
Nº médio de defeitos do tipo D_1	0.5397 (0.0562)	0.1650 (0.0290)
Nº médio de defeitos do tipo D_2	2.0093 (0.1172)	1.4800 (0.0938)
Coeficiente de correlação entre as frequências dos defeitos do tipo D_1 e D_2	0.2076 (0.0422)	0.0026 (0.0284)

3.3 Utilização de modelos de Regressão Poisson Bivariada

Nesta seção, os dados do **Exemplo 1.1** foram reanalisados considerando modelos de Regressão Poisson Bivariada. Inicialmente, é considerado um modelo da forma (2.14) com matrizes de especificação $\mathbf{X}_{A_1}$, $\mathbf{X}_{A_2}$, $\mathbf{X}_{B_1}$ e $\mathbf{X}_{B_2}$ idênticas às do Modelo (3.1) e

$$\begin{aligned} \theta_1 = \ldots = \theta_{130} = \theta_{A_1}, \\ \theta_{131} = \ldots = \theta_{230} = \theta_{A_2}, \\ \theta_{231} = \ldots = \theta_{330} = \theta_{B_1}, \\ \theta_{331} = \ldots = \theta_{430} = \theta_{B_2}. \end{aligned} \tag{3.12}$$

Os valores iniciais para o algoritmo utilizado na obtenção dos estimadores de **MV** foram calculados a partir de

$$\boldsymbol{\beta}_{jk}^{(0)} = \ln\left(\overline{\mathbf{Y}}_{jk}\right) \tag{3.13}$$

e

$$\theta_{jk}^{(0)} = \sqrt{C_{jk}} \tag{3.14}$$

onde $\overline{\mathbf{Y}}_{jk}$ e C_{jk} são respectivamente, o vetor das médias amostrais e a covariância amostral correspondente à máquina j_k, $j = A, B$ e $k = 1, 2$. Quanto $C_{jk} < 0$ (ver **Tabela 3.2**), estabeleceram-se valores pequenos e positivos (0.01) como valores iniciais para θ_{jk}.

As estimativas dos parâmetros com seus erros padrões estimados (obtidos de 2.18) estão indicadas na **Tabela 3.7**.

Tabela 3.7- Estimativas dos parâmetros com os respectivos erros padrões estimados sob o Modelo (3.12).

	Máquina			
Parâmetro	A_1	A_2	B_1	B_2
β_1	-0.7244 (0.1260)	-0.4943 (0.1280)	-1.7720 (0.2425)	-1.8326 (0.2500)
β_2	0.6737 (0.0626)	0.7275 (0.0695)	0.5068 (0.0774)	0.2624 (0.0877)
θ	0.4565 (0.0856)	0.6618 (0.0721)	0.2081 (0.1437)	0.0000 (0.2404)

Além da estimativa de **MV** da covariância, outras estimativas deste parâmetro foram obtidas pelo Método dos Momentos (**MM**) (1.40), Double-Zero (**DZ**) (1.42) e Even-Point (**EP**) (1.49) cujos resultados estão apresentados na **Tabela 3.8** com os respectivos erros padrões estimados.

Tabela 3.8 - Estimativas da covariância e correspondentes erros padrões estimados.

	Estimativas da covariância segundo os métodos de			
Máquina	**MV**	**MM**	**DZ**	**EP**
A_1	0.02084 (0.0782)	0.2668 (0.0996)	0.1435 (0.2631)	***
A_2	0.4381 (0.0955)	0.5730 (0.1471)	0.7829* (0.2380)	0.8099* (0.2069)
B_1	0.0433 (0.0598)	0.0382 (0.0567)	-0.3773** (0.2844)	***
B_2	0.0000 ($< 10^4$)	-0.0181** (0.0436)	-0.1494** (0.2000)	***

Examinando a **Tabela 3.8** constata-se que os erros padrões estimados das estimativas de **MV** das covariâncias são menores que os erros padrões estimados das estimativas obtidas por outros métodos (exceto para a covariância correspondente à máquina B_1, através do método dos Momentos). Isto mostra que nesta aplicação, a estimativa de **MV** da covariância parece ser mais eficiente. Além disto, observou-se que algumas estimativas da covariância são inadequadas. Mais especificamente,

* não atende aos limites estabelecidos : $0 \leq \text{Cov}(Y_1,Y_2) \leq \min [\ \text{E}(Y_1)\ ,\ \text{E}(Y_2)\ ]$ conforme (1.18).

** por definição, a $\text{Cov}(Y_1,Y_2)$ corresponde ao parâmetro de uma distribuição de Poisson univariada, não podendo assumir valores negativos (ver Seção 1.2).

*** por definição, o estimador **EP** só pode ser obtido se à proporção observada do evento A= (Y_1 e Y_2 são pares ou Y_1 e Y_2 são ímpares) for maior que 0.5 conforme (1.49). Os casos assinalados não satisfazem esta restrição, impossibilitando a obtenção da estimativa da covariância pelo método **EP**.

As hipóteses de igualdade das taxas médias de defeitos e de homogeneidade de associação nas quatro máquinas podem ser respectivamente expressas como:

$$H_0 : \boldsymbol{\beta}_{A_1} = \ldots = \boldsymbol{\beta}_{B_2} \tag{3.15}$$

e

$$H_0 : \theta_{A1} = \theta_{A2} = \theta_{B1} = \theta_{B2}. \tag{3.16}$$

Os níveis descritivos dos testes das hipóteses (3.15) e (3.16) são respectivamente $p < 0.0001$ e $p = 0.0033$, indicando evidências contrárias às duas hipóteses testadas. Como as máquinas são de fabricantes diferentes, há interesse em testar as hipóteses de igualdade das taxas médias de defeitos e de homogeneidade de associação entre as frequências dos dois tipos das respostas para as máquinas de cada um dos fabricantes. Estas hipóteses podem ser respectivamente expressas como

$$H_0 : \begin{cases} \boldsymbol{\beta}_{A1} = \boldsymbol{\beta}_{A2} \\ \boldsymbol{\beta}_{B1} = \boldsymbol{\beta}_{B2} \end{cases} \tag{3.17}$$

e

$$H_0 : \begin{cases} \theta_{A1} = \theta_{A2} \\ \theta_{B1} = \theta_{B2}. \end{cases} \tag{3.18}$$

Os níveis descritivos referentes aos testes das hipóteses (3.17) e (3.18) são respectivamente $p = 0.1959$ e $p = 0.1421$, indicando homogeneidade na distribuição das respostas para as máquinas de cada um dos fabricantes. Desta forma, o modelo proposto pode ser reduzido a um outro modelo da forma (2.14), com

$$\mathbf{X}_1 = \ldots = \mathbf{X}_{230} = \mathbf{X}_A,$$
$$\mathbf{X}_{231} = \ldots = \mathbf{X}_{430} = \mathbf{X}_B,$$

onde

$$\mathbf{X}_A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \text{ e } \mathbf{X}_B = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

e

$$\begin{aligned} \theta_1 = \ldots = \theta_{230} = \theta_A, \\ \theta_{231} = \ldots = \theta_{430} = \theta_B. \end{aligned} \tag{3.19}$$

Em síntese, este modelo reduzido evidencia que as taxas médias de defeitos e a associação entre as frequências de defeitos dos dois tipos são homogêneas para as máquinas de um mesmo fabricante.

As estimativas dos parâmetros com os respectivos erros padrões obtidos sob o Modelo (3.19) estão apresentadas na **Tabela 3.9**.

Tabela 3.9 - Estimativas dos parâmetros e os erros padrões correspondentes sob o Modelo (3.19).

	Parâmetro		
Fabricante	β_1	β_2	θ
A	-0.6178	0.6975	0.5528
	(0.0898)	(0.0465)	(0.0550)
B	-1.8018	0.39204	0.1128
	(0.1741)	(0.0581)	(0.1718)

O interesse em avaliar se a associação entre as frequências dos dois tipos de defeitos é nula equivale a hipótese

$$H_0 : \theta_j = 0$$

$j = A, B$. Como as análises anteriores indicaram heterogeneidade da distribuição de respostas entre os fabricantes, esta hipótese foi testada para cada fabricante separadamente, observando-se os níveis descritivos $p < 0.0001$ e $p = 0.2557$, respectivamente para os fabricantes A e B. Estes resultados sugerem que as frequências dos dois tipos de defeitos apresentam associação positiva apenas nas máquinas do fabricante A.

As estimativas das taxas médias de defeitos e dos coeficientes de correlação entre as frequências dos dois tipos de defeitos com os respectivos erros padrões sob o Modelo (3.19) estão na **Tabela 3.10**.

Tabela 3.10 - Estimativa das taxas médias de defeitos e dos coeficientes de correlação entre as frequências dos dois tipos de defeitos com os respectivos erros padrões sob o Modelo (3.19).

	Máquinas do Fabricante	
	A	B
Nº médio de defeitos do tipo D_1	0.5391 (0.0484)	0.1650 (0.0287)
Nº médio de defeitos do tipo D_2	2.0087 (0.0934)	1.4800 (0.0860)
Coeficiente de correlação entre as frequências dos defeitos do tipo D_1 e D_2	0.2937 (0.0519)	0.0257 (0.0228)

3.4 Comparação dos modelos de regressão

Sob certos aspectos, os resultados obtidos sob os modelos de Regressão Poisson Log-Normal Bivariada e de Regressão Poisson Bivariada são coerentes. Os diferentes modelos de regressão indicam homogeneidade das respostas para máquinas do mesmo fabricante tanto no que diz respeito à associação das frequências dos dois tipos de defeitos quanto às taxas médias de defeitos. Além disso, constatou-se que existe associação positiva entre as frequências dos dois tipos de defeitos apenas para as máquinas do fabricante A.

Embora as distribuições Poisson Bivariada e Poisson Log-Normal Bivariada sejam diferentes classes de distribuições, elas podem ser vistas como distribuições Multinomiais com muitas categorias de respostas correspondendo às diferentes combinações de contagens bivariadas. Sob este aspecto, o ajuste dos diferentes modelos de regressão pode ser comparado através do Critério de Informação de Akaike (**AIC**) [KENDALL; STUART; ORD (1983)] dado por

$$\textbf{AIC} = -2\text{V} + 2\text{p}$$

onde V é o logaritmo da função de verossimilhança e p é o número de parâmetros do modelo de regressão. Os valores do Critério de Informação de Akaike correspondentes aos vários modelos de regressão considerados neste trabalho estão na **Tabela 3.11**.

Tabela 3.11 - Valores do Critério de Informação de Akaike (**AIC**).

Modelos de Regressão		p	V	**AIC**
Poisson Log-Normal Bivariada	(3.1)	20	−1003.19	2046.38
	(3.5)	14	−1004.93	2037.87
	(3.11)	10	−1007.78	2035.55
Poisson Bivariada	(3.12)	12	−1020.91	2065.83
	(3.19)	06	−1005.33	2022.66

Pelo **AIC** observa-se que o modelo (3.5) tem ajuste semelhante ao modelo (3.11). Considerando todos os modelos de regressão, o melhor corresponde ao de Regressão Poisson Bivariada (3.19) apresentando menor **AIC**.

No exemplo prático apresentado, não houve evidência forte de que o modelo Poisson Log-Normal Bivariada fosse mais adequado que o modelo Poisson Bivariado, visto que as frequências de defeitos do **Exemplo 1.1** não apresentam coeficientes de correlações amostrais positivos moderados ou sobredispersão amostral moderada ou forte. No entanto, utilizando uma amostra de tamanho 50 gerada sob uma distribuição Poisson Log-Normal Bivariada com covariância entre as variáveis igual a -0.0952, cujos dados estão resumidos na **Tabela 3.12**, pode-se observar que neste caso, o modelo Poisson Log-Normal Bivariado é mais adequado que o modelo Poisson Bivariado conforme discussão a seguir. Na **Tabela 3.13** estão as estimativas dos parâmetros sob as duas distribuições mencionadas.

Tabela 3.12 - Frequências das observações de uma amostra simulada.

Y_2 \ Y_1	0	1	2	3	4	5	...	Total
0	3	9	3	2	1	1	...	19
1	7	7	2	1	1	.	...	18
2	2	3	.	.	1	.	...	6
3	2	1	1	.	.	.	...	4
4	2	.	1	.	.	.	...	3
...	...	...	...	...	...	...	...	...
Total	16	20	7	3	3	1	...	50

Tabela 3.13 - Estimativas dos parâmetros e os correspondentes erros padrões utilizando os dados da **Tabela 3.12**.

		Estimativas sob uma distribuição	
Parâmetro	Valor Observado	Poisson Log-Normal Bivariada	Poisson Bivariada
$E(Y_1)$=1.0000	1.2000	1.2003 (0.1776)	1.2000 (0.1549)
$E(Y_2)$=1.0000	1.0800	1.0815 (0.2353)	1.0800 (0.1470)
$Var(Y_1)$=1.2214	1.5510	1.5746 (0.4621)	1.2000 (0.1549)
$Cov(Y_1, Y_2)$=−0.0952	−0.3243	−0.2799 (0.5341)	0.0000 ($< 10^{-4}$)
$Var(Y_2)$=1.2214	1.3812	1.4217 (0.3043)	1.0800 (0.1470)

A amostra simulada apresenta sobredispersões amostrais de 29% e 28% para as contagens e um coeficiente de correlação amostral de -0.22. Sob as distribuições Poisson Log-Normal Bivariada e Poisson Bivariada, os logaritmos da função de verossimilhança são respectivamente -141.62 e -144.31, e os valores do Critério de Informação de Akaike são respectivamente 293.25 e 294.63. Conforme a **Tabela 3.13**, as estimativas das médias são bastante semelhantes com erro padrão menor sob a distribuição Poisson Bivariada. As estimativas das variâncias sob uma distribuição Poisson Log-Normal Bivariada apresentam discrepâncias inferiores a 1.9% em relação aos valores observados embora apresentem erros padrões maiores e as obtidas sob a outra distribuição apresentam erros padrões menores, porém estão subestimadas com discrepâncias superiores a 5% em relação aos valores observados. A estimativa da covariância sob uma distribuição Poisson Log-Normal Bivariada apresentou uma discrepância de 13% em relação ao valor observado e sob a outra distribuição, o método iterativo forneceu estimativa nula para este parâmetro com 100% de discrepância. As observações acima levam a concluir que o modelo de distribuição Poisson Log-Normal Bivariada é mais adequado neste exemplo.

Capítulo 4

Distribuição empírica das estimativas dos parâmetros da Poisson Log-Normal Bivariada através de amostras simuladas

4.1 Introdução

Amostras de distribuições Poisson Log-Normal Bivariadas foram simuladas para estudar o comportamento das distribuições empíricas das estimativas dos parâmetros correspondentes. Neste capítulo estão descritos esses resultados.

4.2 Geração das amostras

Inicialmente foram geradas 1000 amostras de tamanho 50 e 1000 amostras de tamanho 100 segundo uma distribuição Poisson Log-Normal Bivariada com parâmetros

$$\boldsymbol{\mu}= (-0.1, -0.1)' \text{ e } \boldsymbol{\Sigma} = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}.$$

Com estes parâmetros, o valor esperado e a matriz de covariância de uma observação $\mathbf{Y}$ são respectivamente

$$\mathrm{E}(\mathbf{Y})= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ e } Var(\mathbf{Y}) = \begin{pmatrix} 1.2214 & 0.1052 \\ 0.1052 & 1.2214 \end{pmatrix}.$$

Outras 1000 amostras de tamanhos 50 e 1000 amostras de tamanhos 100 foram geradas segundo uma distribuição Poisson Log-Normal Bivariada com parâmetro $\boldsymbol{\mu}$ igual àquele utilizado na simulação anterior e $\boldsymbol{\Sigma} = \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 0.2 \end{pmatrix}$. O valor esperado e a matriz de covariância de uma observação $\mathbf{Y}$ gerada com estes parâmetros são, respectivamente

$$\mathrm{E}(\mathbf{Y}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ e } Var(\mathbf{Y}) = \begin{pmatrix} 1.2214 & -0.0952 \\ -0.0952 & 1.2214 \end{pmatrix}.$$

Esse modelo de probabilidade admite covariância negativa entre os dois componentes de $\mathbf{Y}_i$. O interesse em gerar amostras a partir de uma distribuição em que essa covariância é negativa está centrado no fato de outras distribuições usualmente consideradas para modelar dados do tipo daqueles enfocados neste trabalho admitirem apenas covariâncias positivas entre as componentes da variável resposta. Esse é o caso da distribuição Poisson Multivariada e das distribuições Poisson Compostas Bivariadas citadas em KOCHERLAKOTA (1988).

Para gerar um vetor $\mathbf{Y}_i=(Y_{1i}, Y_{2i})'$ com distribuição Poisson Log Normal Bivariada de parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ foi utilizado o seguinte esquema:

1) Gerar Z_{1i} e Z_{2i} variáveis aleatórias independentes segundo uma distribuição Normal (0, 1).

2) Obter $\mathbf{Z}^*_i=(Z^*_{1i}, Z^*_{2i})' = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}_i$ com $\mathbf{Z}_i=(Z_{i1},Z_{i2})'$, $\mathbf{A} = \mathbf{T}\sqrt{diag(\mathbf{v})}$ onde $\mathbf{v}$ é um vetor composto pelos autovalores da matriz $\boldsymbol{\Sigma}$ e $\mathbf{T}$ é uma matriz cujas colunas são os autovetores de $\boldsymbol{\Sigma}$; neste caso $\mathbf{Z}^*_i$ obedece a uma distribuição Normal Bivariada de parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$.

3) Obter $\lambda_{1i} = \exp(Z^*_{1i})$ e $\lambda_{2i} = \exp(Z^*_{2i})$.

4) Gerar $(Y_{1i}, Y_{2i})'$ variáveis aleatórias independentes segundo uma distribuição Poisson de parâmetros respectivamente iguais a λ_{1i} e λ_{2i}.

As funções RANPOI e NORMAL do pacote estatístico SAS foram utilizadas para gerar amostras aleatórias sob as distribuições Poisson e Normal, respectivamente (ver Apêndice C).

4.3 Pontos iniciais utilizados no processo iterativo

Para cada amostra simulada, calculou-se a matriz $\boldsymbol{\Sigma}^{(0)}$ conforme (3.2). Nos casos em que a expressão (3.2) gerou uma matriz negativa-definida foi utilizado o ajuste de Levenberg-Marquadt [AITCHISON; HO (1989)]. Em caso de subdispersão, foi utilizada a sugestão de AITCHISON; HO (1989), em que se consideram valores pequenos e positivos (por exemplo: 0.01) para σ_{ii}, e valores nulos para σ_{is} e σ_{si}, $i = s = 1,2$. As matrizes $\boldsymbol{\Sigma}_{kj}$, cujos elementos são derivados dos elementos da matriz $\boldsymbol{\Sigma}^{(0)}$, foram utilizadas como valores iniciais no processo iterativo. Para que as matrizes $\boldsymbol{\Sigma}_{kj}$ fossem positivas definidas, alguns cuidados foram tomados, tais como, manter o produto dos elementos da diagonal igual ao produto dos mesmos elementos da matriz $\boldsymbol{\Sigma}^{(0)}$ e os elementos fora da diagonal ($\sigma_{12_{kj}} = \sigma_{21_{kj}}$) limitados a $\sqrt{\sigma_{11}^{(0)}\sigma_{22}^{(0)}}$ de modo que as matrizes tivessem determinantes positivas. Essas matrizes compõem uma malha regular limitada em torno de $\boldsymbol{\Sigma}^{(0)}$; seus elementos são iguais a

$$
\begin{aligned}
\sigma_{11_{kj}} &= a_k \sigma_{11}^{(0)} \\
\sigma_{12_{kj}} &= \sigma_{21_{kj}} = \sigma_{12}^{(0)}/2 + b_j \\
\sigma_{22_{kj}} &= a_k^{-1} \sigma_{22}^{(0)}
\end{aligned}
\tag{4.1}
$$

onde a_k e b_j são constantes, com $a_k \in$ A $= \{2/4, 3/4, 1, 5/4, 6/4, 4/6, 4/5, 4/3, 2\}$, $k = 1, \ldots, 9$, $b_j \in$ B $= \{0, \frac{(1-\epsilon)}{3}dd, \frac{2(1-\epsilon)}{3}dd, (1-\epsilon)dd\ \}$ com $\epsilon > 0$ e $dd = \sqrt{\sigma_{11}^{(0)}\sigma_{22}^{(0)}} - \sigma_{12}^{(0)}$, $j = 1, \ldots, 4$. Para simplificação de notação, as matrizes $\boldsymbol{\Sigma}_{kj}$ foram enumeradas, utilizando um único índice da seguinte forma: $\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_1^*$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_2^*$, $\boldsymbol{\Sigma}_{13} = \boldsymbol{\Sigma}_3^*$, $\boldsymbol{\Sigma}_{14} = \boldsymbol{\Sigma}_4^*$, $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_5^*$,

$\Sigma_{22} = \Sigma_6^*, \ldots, \Sigma_{kj} = \Sigma_m^*, \ldots, \Sigma_{94} = \Sigma_{36}^*$, com $m = j + 4(k-1)$. Tendo a matriz inicial Σ_m^*, obteve-se o valor inicial μ_m^* aplicando (3.3) adequadamente.

As matrizes Σ_{pq}, cujos elementos são constantes, foram utilizadas como valores iniciais no processo iterativo em amostras onde não houve convergência com os valores iniciais calculados conforme (4.1). Como as amostras foram geradas com parâmetros $\Sigma = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$ ou $\Sigma = \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 0.2 \end{pmatrix}$, valores arbitrários em tornos destes valores foram escolhidos para compor as matrizes constantes Σ_{pq} de modo a formar uma malha limitada em torno dos parâmetros Σ mencionados; seus elementos são iguais a

$$\begin{aligned} \sigma_{11_{pq}} &= \sigma_{22_{pq}} = c_p \\ \sigma_{12_{pq}} &= \sigma_{21_{pq}} = c_p d_q \end{aligned} \tag{4.2}$$

onde c_p e d_q são constantes, $c_p \in$ C ={0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6}, $p = 1, \ldots, 8$, e $d_q \in$ D ={-1/2, -1/4, 0, 1/4, 1/2}, $q = 1, \ldots, 5$. Novamente, para efeito de simplificação de notação, as matrizes Σ_{pq} foram enumeradas, utilizando um único índice da seguinte forma $\Sigma_{11} = \Sigma_1^f$, $\Sigma_{12} = \Sigma_2^f$, $\Sigma_{13} = \Sigma_3^f$, $\Sigma_{14} = \Sigma_4^f$, $\Sigma_{15} = \Sigma_5^f =$, $\Sigma_{21} = \Sigma_6^f, \ldots,$ $\Sigma_{pq} = \Sigma_r^f =, \ldots, \Sigma_{85} = \Sigma_{40}^f$, com $r = q + 5(p-1)$. O valor inicial μ_r^f também foi obtido, aplicando (3.3) adequadamente.

A obtenção das estimativas dos parâmetros nas amostras geradas obedeceu ao seguinte esquema. Para uma dada amostra i:

Passo 1: Calcular os valores iniciais Σ_{im}^* e μ_{im}^* conforme (4.1) e (3.3), respectivamente, $m = 1, \ldots, 36$.

Passo 2: Calcular os valores iniciais Σ_{ir}^* e μ_{ir}^* conforme (4.2) e (3.3), respectivamente, $r = 1, \ldots, 40$.

Passo 3: Inicializar o processo com $m = 1$.

Passo 4: Aplicar o procedimento iterativo de Newton-Raphson (2.15) para obter as estimativas de **MV** dos parâmetros com valores iniciais $\boldsymbol{\Sigma}^*_{im}$ e $\boldsymbol{\mu}^*_{im}$. Se houver convergência segundo o critério de parada estabelecido (ver Seção 2.4) parar; caso contrário fazer $m = m + 1$.

Passo 5: Se $m > 36$ ir para o Passo 6; caso contrário voltar para o Passo 4.

Passo 6: Inicializar o processo com $r = 1$.

Passo 7: Aplicar o procedimento iterativo de Newton-Raphson (2.15) para obter as estimativas de **MV** dos parâmetros com os valores iniciais $\boldsymbol{\Sigma}^*_{ir}$ e $\boldsymbol{\mu}^*_{ir}$. Se houver convergência segundo o critério de parada (ver Seção 2.4), parar, caso contrário fazer $r = r + 1$.

Passo 8: Se $r > 40$, parar; caso contrário voltar para o Passo 7.

O critério de parada utilizado na simulação foi o critério absoluto (ver Seção 2.4). O processo iterativo foi interrompido quando a maior diferença absoluta entre as estimativas dos componentes dos parâmetros em sucessivas iterações era menor que 10^{-5}. A estabilidade das estimativas dos parâmetros foi verificada através de uma subamostra de tamanho 20. Utilizaram-se diferentes pontos iniciais, todos próximos da estimativa fornecida pelo procedimento de Newton-Raphson e observaram-se convergências para o mesmo resultado.

Aplicado este esquema, observou-se que em apenas cerca de 80% das amostras simuladas houve convergência do processo iterativo. Essa informação está detalhada na **Tabela 4.1**.

Tabela 4.1 - Porcentagem de amostras onde houve convergência.

Covariância	Tamanho da amostra 50	100
positiva	81.5	84.4
negativa	85.2	86.6

As possíveis causas de não convergência em algumas amostras simuladas podem estar relacionadas com os tamanhos das amostras utilizados nas simulações, com a aproximação utilizada para calcular as probabilidades ou com o método iterativo para obter as estimativas dos parâmetros. Esse problema é frequente na obtenção de estimativas de parâmetros multivariados através de métodos iterativos, como menciona THISTED (1988).

4.4 Distribuição empírica

Nas **Tabelas 4.2 e 4.3** estão apresentadas algumas estatísticas descritivas como médias e desvios padrões amostrais das distribuições empíricas das estimativas dos parâmetros obtidas através de amostras simuladas.

Tabela 4.2 - Médias e desvios padrões amostrais das estimativas dos parâmetros através de amostras simuladas - caso de covariância positiva (ver Seção 4.2).

Parâmetro	Tamanho de amostra 50	100
μ_1=−0.1000	−0.1196 (0.1907)	−0.0967 (0.1353)
μ_2=−0.1000	−0.1105 (0.1922)	−0.1046 (0.1347)
σ_{11} =0.2000	0.2009 (0.2072)	0.1971 (0.1671)
σ_{12}=0.1000	0.0652 (0.1487)	0.0689 (0.1105)
σ_{22}=0.2000	0.2072 (0.2100)	0.1816 (0.1494)
$E(Y_1)$=1.000	0.9946 (0.1641)	1.0082 (0.1133)
$E(Y_2)$=1.0000	1.0064 (0.1630)	0.9928 (0.1132)
$Var(Y_1)$=1.2214	1.2613 (0.4822)	1.2576 (0.3480)
$Cov(Y_1, Y_2)$=0.1052	0.0836 (0.1801)	0.0799 (0.1269)
$Var(Y_2)$=1.2214	1.2818 (0.4473)	1.2100 (0.3092)

Tabela 4.3 - Médias e desvios padrões amostrais das estimativas dos parâmetros através de amostras simuladas - caso de covariância negativa (ver Seção 4.2).

	Tamanho de amostra	
Parâmetro	50	100
μ_1=−0.1000	−0.1173	−0.1031
	(0.1892)	(0.1317)
μ_2=−0.1000	−0.1102	−0.1040
	(0.1872)	(0.1315)
σ_{11}=0.2000	0.2232	0.2003
	(0.2344)	(0.1660)
σ_{12}=−0.1000	−0.0723	−0.0839
	(0.1366)	(0.1093)
σ_{22}=0.2000	0.1968	0.2000
	(0.2099)	(0.1466)
$E(Y_1)$=1.0000	1.0069	1.0031
	(0.1588)	(0.1093)
$E(Y_2)$=1.0000	1.0004	1.0023
	(0.1549)	(0.1113)
$Var(Y_1)$=1.2214	1.3210	1.2540
	(0.5544)	(0.3426)
$Cov(Y_1, Y_2)$=−0.0952	−0.0630	−0.0753
	(0.1293)	(0.1040)
$Var(Y_2)$=1.2214	1.2615	1.2444
	(0.4960)	(0.2878)

Examinado as **Tabelas 4.2 e 4.3**, observou-se que os desvios padrões das estimativas com amostras de tamanho 100 são menores que os desvios padrões das estimativas com amostras de tamanho 50. As médias amostrais das estimativas de $\boldsymbol{\mu}$ com amostras de tamanho 100 são mais próximas do valor esperado, com discrepâncias entre 3% e 5%

em relação a este valor enquanto que as obtidas com amostras menores apresentaram discrepâncias entre 10% e 20%.

O mesmo comportamento foi observado com relação às médias amostrais das estimativas das variâncias, com discrepâncias entre 1% e 3% em amostras de tamanho 100 e entre 3% a 8% em amostras menores. No entanto, pode-se notar que são casos de superestimação deste parâmetro, independentemente do tamanho da amostra.

As médias amostrais das estimativas das médias estão próximas do valor esperado com discrepâncias inferiores a 0.8% em relação a este valor, independentemente do tamanho da amostra. No entanto, as médias amostrais das estimativas de σ_{12} e das covariâncias não estão próximas dos valores esperados, apresentando discrepâncias entre 15% e 35% em relação aos respectivos valores esperados, observando-se subestimação para o caso da covariância positiva e superestimação para o caso da covariância negativa.

Testes de aderência da hipótese de normalidade foram efetuados com as distribuições empíricas das estimativas dos parâmetros, utilizando o método de Shapiro-Wilks (ROYSTON, 1988) através do procedimento UNIVARIATE do pacote estatístico SAS, cujos resultados estão resumidos na **Tabela 4.4**.

Tabela 4.4 - Níveis descritivos referente ao teste de aderência da hipótese de normalidade das distribuições empíricas das estimativas dos parâmetros.

Parâmetro	Covariância positiva Tamanho de amostra 50	 100	Covariância negativa Tamanho da amostra 50	 100
μ_1	0.0262	0.0004	0.0006	0.0001
μ_2	0.0001	0.0001	0.0001	0.0350
σ_{11}	0.0000	0.0000	0.0000	0.0000
σ_{12}	0.0001	0.9998	0.0001	0.9874
σ_{22}	0.0000	0.0000	0.0000	0.0000
$E(Y_1)$	0.2005	0.0060	0.1647	0.6684
$E(Y_2)$	0.0731	0.7221	0.2210	0.4941
$Var(Y_1)$	0.0000	0.0000	0.0000	0.0000
$Cov(Y_1, Y_2)$	0.0000	0.0000	0.0164	0.5618
$Var(Y_2)$	0.0000	0.0000	0.0000	0.0000

Assintoticamente, a distribuição das estimativas de **MV** dos parâmetros tem distribuição Normal cujo parâmetro da variância está relacionada com a expressão (2.18). Analisando a **Tabela 4.4**, observa-se que apenas algumas distribuições empíricas das estimativas dos parâmetros obedecem a distribuições normais. Entre essas, a maioria está relacionada com a estimação da média. As possíveis causas para o fato de as distribuições empíricas não seguirem distribuições normais podem estar relacionadas com os tamanhos de amostras utilizados nas simulações, com a aproximação empregada para calcular as probabilidades e com o método iterativo utilizado para obter as estimativas dos parâmetros.

Capítulo 5

Conclusões e sugestões

Este trabalho apresentou uma análise estatística de contagens multivariadas provenientes de várias populações através de modelos de regressão supondo que as contagens obedeçam a algumas distribuições de probabilidade como Poisson Multivariada e Poisson Log-Normal Multivariada. O interesse nesta última distribuição está relacionado com o fato de admitir correlações de ambos os sinais entre as contagens; este não é o caso da distribuição Poisson Multivariada e nem das distribuições Poisson Compostas Multivariadas citadas em KOCHERLAKOTA (1988) que admitem apenas correlações positivas.

No Capítulo 3, foram apresentadas algumas aplicações de Regressão Poisson Log-Normal Bivariada e de Regressão Poisson Bivariada utilizando os dados de contagens de dois tipos de defeitos em 100g de fibras têxteis produzidas por 4 máquinas craqueadeiras, sendo duas de um fabricante e as outras de um segundo fabricante. Os diferentes modelos de regressão apontam resultados coerentes como heterogeneidade da distribuição de respostas entre as máquinas e entre os fabricantes e homogeneidade da distribuição de respostas em máquinas de um mesmo fabricante. Estes resultados podem fornecer subsídios para planejar uma amostra mais adequada, por exemplo, estratificando a população das máquinas, em dois estratos: máquinas do fabricante A e máquinas do fabricante B. A constatação de uma associação positiva entre as frequências dos dois tipos de defeitos em máquinas do fabricante A poderia gerar uma economia de recursos, uma vez que o conhecimento do comportamento de um dos tipos de defeitos, poderia ser utilizado para prever o comportamento do outro, não sendo necessário coletar os dados dos dois tipos de defeitos. Os resultados dos diferentes modelos de regressão foram comparados através do Critério de Informação de Akaike que sugere como melhor modelo entre os propostos, um dos modelos de Regressão Poisson Bivariada. Este resultado está de acordo com as recomendações citadas no Capítulo 3, visto que há casos de subdispersões em 3 máquinas e taxas de sobredispersão inferiores a 30% nas contagens.

Ainda no Capítulo 3, um outro exemplo utilizando os dados de uma amostra simulada de tamanho 50 sob uma distribuição Poisson Log-Normal Bivariada com correlação negativa entre as contagens foi considerado. Os dados apresentam um coeficiente de correlação amostral de -0.22 e uma sobredispersão amostral em torno de 30% em cada contagem; o Critério de Informação de Akaike sugere que um modelo Poisson Log-Normal Bivariado parece ser mais adequado neste caso. Este resultado também é coerente com as recomendações citadas no Capítulo 3.

No Capítulo 4, amostras foram simuladas segundo distribuições Poisson Log-Normal Bivariadas para estudar o comportamento da distribuição empírica das estimativas dos parâmetros correspondentes. Foram utilizados dois conjuntos de valores iniciais, o primeiro calculado a partir de momentos amostrais e o segundo com valores fixos. Em estudos preliminares, a sugestão de AITCHISON; HO (1988) de atribuir um valor inicial nulo para σ_{ij} em caso de subdispersão mostrou-se eficiente apenas nos casos de coeficientes de correlação entre as contagens próximos de zero (menores que 0.01). Em outras situações, houve necessidade de adotar valores iniciais não nulos para σ_{ij} com a finalidade de obter as estimativas dos parâmetros pelo método de Newton-Raphson. Estas considerações foram utilizadas na construção dos valores iniciais. No final de todo processo, as estimativas dos parâmetros foram obtidas em torno de 80% das amostras simuladas. Apesar de esta ser uma simulação restrita, constatou-se que valores iniciais derivados de momentos amostrais nem sempre são eficientes para obter as estimativas dos parâmetros no processo iterativo de Newton-Raphson. Observou-se também que as distribuições empíricas das estimativas das médias de uma distribuição Poisson Log-Normal Bivariada obedecem a uma distribuição Normal. As possíveis causas da não normalidade nas distribuições empíricas das estimativas dos outros parâmetros podem estar relacionadas com os tamanhos das amostras utilizados na simulação, com a aproximação utilizada para calcular as probabilidades e com o método iterativo empregado para obter as estimativas dos parâmetros.

Uma sugestão para futuras pesquisas é aplicar estes modelos de regressão em contagens multivariadas de dimensões maiores que dois, com ou sem observações concomitantes de covariáveis. Nestes casos, o leque de hipóteses de interesse tende a aumentar muito. Outra possibilidade é utilizar outros métodos iterativos na obtenção dos estimadores de **MV**. Entre eles, há métodos semelhantes ao Newton-Raphson (THISTED, 1988) como o método Newton discretizado, o método da Secante Generalizada ou o método Quasi-Newton. Entre os algoritmos utilizados nos métodos Quasi-Newton, dois são bastante conhecidos: o algoritmo de Davidon-Fletcher-Powell e o de Broyden-Fletcher-Goldfarb-Shanno (THISTED, 1988). Além disso, outras aproximações para o cálculo das probabilidades de uma distribuição Poisson Log-Normal Multivariada podem ser empregadas. Entre as aproximações, existem as propostas por NAYLOR; SMITH (1982), CROUCH; SPIELGMAN (1990) e LIU; PIERCE (1994).

Pode-se também planejar uma simulação mais abrangente com vários valores de $\boldsymbol{\mu}$, matrizes $\boldsymbol{\Sigma}$ e outros tamanhos de amostras para estudar melhor o comportamento assintótico das estimativas dos parâmetros da distribuição Poisson Log-Normal Multivariada.

Bibliografia

[1] AITCHISON, J.; HO, C.H. The multivariate Poisson log-normal distribution. **Biometrika**, v.76, n.4, p.643-53, 1989.

[2] AITCHISON, J. A general class of distributions on the simplex. **Journal of the Royal Statistical Society**. Series B. v.47, n.1, p.136-46, 1985.

[3] ALT, F. Multivariate quality control. In: Kotz, S.; Johnson, N. L. **Encyclopedia of statistical sciences**, New York, John Wiley, 1986. v.6, p.110-22.

[4] ALT, F.; SMITH, N. D. Multivariate process control. In: Krishnaiah, P. R.; Rao, C.R. **Handbook of statistics**, The Netherlands, Elsevier, 1988. v.7, p.333-51.

[5] ARBOUS, A.G.; KERRICH, J.E. Accident statistics and the concept of accident-proneness. **Biometrics**, v.7, n.4, p.340-432, 1951.

[6] BULMER, M.G. On fitting the Poisson lognormal distribution to species-abundance data. **Biometrics**, v.30, n.1, p.101-10, 1974.

[7] CAMPELL, J.T. The Poisson correlation fuction. **Proceedings of the Edinburgh Mathematical Society**. Series 2. v.4, p.18-26, 1934.

[8] CHEN, C.F. Score tests for regression models. **Journal of the American Statistical Association**, v.78, n.381, p.158-61, 1983.

[9] CROUCH, E. A. C.; SPIELGELMAN, D. The evaluation of integrals of the form $\int f(t)\exp(-t^2)dt$: application to logistic normal models. **Journal of the American Statistical Association**, v.85, n.410, p.464-9, 1990.

[10] EDWARDS, C.B.; GURLAND, J. A class of distributions applicable to accidents. **Journal of the American Statistical Association**, v.56, n.295, p.503-17, 1961.

[11] FROME, E.L.; KUTNER, M.H.; BEAUCHAMP, J.J. Regression analysis of Poisson-distributed data. **Journal of the American Statistical Association**, v.68 , n.344, p.935-40, 1973.

[12] GRIFFITHS, R.C.; MILNE, R.K.; WOOD, R. Aspects of correlation in bivariate Poisson distributions and processes. **Australian Journal of Statistics**, v.21, n.3, p.238-55, 1979.

[13] HAMDAN, M.A.; JENSEN, D.R. A bivariate binomial distributions and some applications. **Australian Journal of Statistics**, v.18, n.3, p.163-9, 1976.

[14] HAMDAN, M.A.; AL-BAYYATI, H.A. A note on the bivariate Poisson distribution. **The American Statistician**, v.23, n.4, p.32-3, 1969.

[15] HOLGATE, P. Estimation for the bivariate Poisson distribution. **Biometrika**, v.51, part 1-2, p.241-5, 1964.

[16] HUDSON, W.N.; TUCKER, H.G. Limit theorems for the multivariate binomial distribution. **Journal of Multivariate Analysis**, v.18, n.1, p.32-45, 1986.

[17] JOHNSON, N.L.; KOTZ, S. **Distributions in statistics: discrete distributions.** Boston, Houghton Mifflin, 1969.

[18] JOHNSON, N.L.; KOTZ, S. **Distributions in statistics: continuous multivariate distributions 2.** New York, John Wiley, 1972.

[19] KAWAMURA, K. Calculation of density for the multivariate distribution. **Kodai Mathematical Journal**, v.10, n.2, p.231-41, 1987.

[20] KAWAMURA, K. A note on the recurrent relations for the bivariate distribution. **Kodai Mathematical Journal**, v.8, n.1, p.70-8, 1986.

[21] KAWAMURA, K. Direct calculation of maximum likelihood estimador for the bivariate Poisson distribution. **Kodai Mathematical Journal**, v.7, n.2, p.211-21, 1984.

[22] KAWAMURA, K. The structure of multivariate Poisson distribution. **Kodai Mathematical Journal**, v.2, n.3, p.337-45, 1979.

[23] KAWAMURA, K. The structure of trivariate Poisson distribution. **Kodai Mathematical Seminar Reports**, v.28, n.1, p.1-8, 1976.

[24] KAWAMURA, K. The structure of bivariate Poisson distribution. **Kodai Mathematical Seminar Reports**, v.25, n.2, p.246-56, 1973.

[25] KENDALL, M.G.; STUART, A.; ORD, J.K. **The advanced theory of statistics: design and analysis, and time-series.** 4. ed. London, Charles Griffin, 1983. v.3.

[26] KENNEDY Jr. W.J.; GENTLE, J.E. **Statistical computing.** New York, Marcel Dekker, 1980.

[27] KOCH, G.G.; ATKINSON, S.S.; STOKES, M.E. Poisson Regression. In: Kotz, S.; Johnson, N.L. **Encyclopedia of statistical sciences.** New York, John Wiley, 1986. v.7, p.32-41.

[28] KOCHERLAKOTA, S. On the compounded bivariate distribution: a unified treatment. **Annals of Institute of Mathematical Statistics**, v.40, n.1, p.61-76,

1988.

[29] KOCHERLAKOTA, S.; KOCHERLAKOTA K. **Bivariate discrete distributions.** New York, Marcel Dekker, 1992.

[30] LIANG, W.G. The structure of compounded trivariate Poisson distribution. **Kodai Mathematical Journal**, v.12, n.1, p.30-40, 1989.

[31] LIU, Q.; PIERCE, D. A. A note on Gauss-Hermite quadrature. **Biometrika**, v.81, n.3, p.624-9, 1994.

[32] LOUKAS, S.; KEMP, C.D.; PAPAGEOURGIOU H. Even-Point estimation for the bivariate Poisson distribution. **Biometrika**, v.73, n.1, p.222-3, 1986.

[33] MARITZ, J.S. Note on a certain family of discrete distributions. **Biometrika**, v.39, part 1-2, p.196-8, 1952.

[34] MARQUARDT, D.W. An algorithm for least-squares estimation of nonlinear parameters. **Journal of Society Industrial Applied Mathematics**, v.2, n.2, p.431-41, 1963.

[35] MARSHALL, A.W.; OLKIN I. A family of bivariate distributions generated by the bivariate Bernoulli distribution. **Journal of the American Statistical Association**, v.80, n.390, p.332-38, 1985.

[36] M'KENDRICK, A.G. Applications of mathematics to medical problems. **Proceedings of the Edinburgh Mathematical Society**. Series 1. v.44, p.98-130, 1926.

[37] NAYLOR, J.C.; SMITH A.F.M. Applications of a method for the efficient computation of posterior distributions. **Applied Statistics**, v.31, n.3, p.214-25, 1982.

[38] PAPAGEOURGIOU, H.; KEMP, C.D.; LOUKAS, S. Some methods of estimation for the bivariate Hermite distribution. **Biometrika**, v.70, n.2, p.479-84, 1983.

[39] PATEL, Y.C. Even-Point estimation and moment estimation in Hermite distribution. **Biometrics**, n.4, v.32, p.865-73, 1976.

[40] RAO, C.R.; CHAKRAVARTI, I.M. Some small sample tests of significance for a Poisson distribution. **Biometrics**, v.12, n.3, p.264-82, 1956.

[41] ROYSTON, J.P. Shapiro-Wilks statistics. In: Kotz, S.; Johnson, N.L. **Encyclopedia of statistical sciences.** New York, John Wiley, 1988. v.8, p.430-1.

[42] SALZER, H.E.; ZUCKER, R.; CAPUANO, R. Table of the zeroes and weight factors of the first twenty Hermite polynomials. **Journal of Research National Bureau Standards**, v.48, n.2, p.111-6, 1952.

[43] SAS/IML User's Guide. Release 6.03 Edition. Cary, SAS Institute, 1988.

[44] SHABAN, S.A. Poisson log-normal distributions. In: Crow, E.L.; Shimizu K. **Log-normal distributions: theory and applications.** New York, Marcel Dekker, 1988. p.195-210.

[45] SUBRAHMANIAM, K. A test for intrinsic correlation in the theory of accident proneness. **Journal the Royal of Statistical Society** . Series B. v.28, n.1, p.180-9, 1966.

[46] THISTED, R.A. **Elements of statistical computing: numerical computation.** New York, Chapman and Hall, 1988.

Apêndice A

Programa para calcular as estimativas de MV dos parâmetros de uma distribuição Poisson Log-Normal Bivariada

Este programa foi escrito através da linguagem IML do pacote estatístico SAS. Ele é utilizado para calcular as estimativas de MV dos parâmetros de uma distribuição Poisson Log-Normal Bivariada através do método de Newton-Raphson. Esta listagem corresponde à obtenção da estimativa dos parâmetros do modelo Regressão de Poisson Log-Normal Bivariada (3.1). As estimativas para outros modelos de regressão assim como as estimativas dos parâmetros das amostras simuladas podem ser obtidas, com algumas alterações apropriadas.

```
proc iml worksize=250; reset nolog linesize=132 pagesize=60;
*--------------------------;
*entrada  de dados;
*dimensao de Z deve k x 3;
* onde 1. coluna = Y1;
* 2. coluna = Y2;
* 3. coluna = frequencia;
*--------------------------;
*==================================;
* dados da maquina A1;
*==================================;
Za={ 0 0 13, 0 1 26, 0 2 22, 0 3 16, 0 4 3, 0 5 2,
```

```
1 0 3, 1 1 9, 1 2 12, 1 3 7,1 4 5,1 5 2,1 6 1,
 2 1 2, 2 2 1, 2 5 1, 3 1 1, 3 3 2, 3 7 1,  4 2 1 };
*=================================;
* dados da maquina A2;
*=================================;
Zb={ 0 0 15, 0 1 14, 0 2 11, 0 3 10, 0 4 3, 0 5 1,
1 0 2, 1 1 12, 1 2 6, 1 3 7, 1 4 5, 1 5 3, 1 7 1,
 2 1 1, 2 2 2, 2 3 1, 2 5 2, 3 4 1, 3 5 1, 3 7 1,
 4 3 1 };
*===================================;
* dados da maquina B1;
*===================================;
Zc={ 0 0 11, 0 1 35, 0 2 21, 0 3 13, 0 4 5, 0 5 1,
1 0 1, 1 1 5, 1 2 3, 1 3 1, 1 4 1,  2 1 1, 2 3 2 };
*=====================================;
* dados da maquina B2;
*=====================================;
Zd={ 0 0 20, 0 1 35, 0 2 17, 0 3 7, 0 4 4, 0 5 1,
1 0 4, 1 1 7, 1 2 3, 1 3 2 };
*==========================================================;
* PARAMETROS E NOMES DAS MATRIZES UTILIZADAS NO PROGRAMA;
*==========================================================;
*valores iniciais;
*dimensao de np e p0 deve ser de beta+3;
*DIMENSAO DE X DEVE SER 2 x BETA;
* np = igual ao vetor dos parametros na iteracao i+1;
* p0 = igual ao vetor dos parametros na iteracao i;
*x = matriz de planejamento;
*    de dimensao 2 x beta;
*varcov =  matriz de covariancia amostral;
*sigma = matriz parametro sigma ;
*yb = vetor das medias amostrais;
*k= dimensao da matriz sigma;
*vp = valor pequeno para iniciar sigma;
*r=tamanho do vetor do parametro(beta + T);
*b=tamanho do vetor beta;
* s= numero de colunas de sigma ;
* n= tamanho da amostra;
* t = numero de pares distintos dos valores de Y;
*n1= maior valor observado de Y1;
*n2= maior valor observado de Y2;
*fdp =matriz da probabilidade conjunta;
*pdbeta= primeira derivada em relacao a beta;
*pdt= primeira derivada em relacao a T;
*sdbeta=segunda derivda em relacao a beta para qq matriz X;
*m=segunda derivada em relacao a beta quando X igual a identidade;
```

```
*sdp = segunda derivada parcial de beta e t;
*sdt= segunda derivada de t;
*-------------------------------------;
 l={1, 1};
i1={ 1, 0};
i2={0,1};
zero={0};
b=8; *==========> alterar convenientemente;
 r=B+3;
n1=zero;
n2=zero;
np = repeat(zero,1,r);
p0= repeat (zero,1,r);
cc=repeat(zero,1,r);
*-----------------------------------------------;
* INDIQUE AS MATRIZES DE PLANEJAMENTO;
*-----------------------------------------------;
*===============> alterar convenientemente;
XA={1 0 0 0 0 0 0 0, 0 1 0 0 0 0 0 0 };
Xb={0 0 1 0 0 0 0 0, 0 0 0 1 0 0 0 0 };
Xc={0 0 0 0 1 0 0 0, 0 0 0 0 0 1 0 0 };
Xd={0 0 0 0 0 0 1 0, 0 0 0 0 0 0 0 1 };
*-----------------------------------------------;
varcova= { 0 0, 0 0};
varcovb= { 0 0, 0 0};
varcovc= { 0 0, 0 0};
varcovd= { 0 0, 0 0};
S1= { 0 0 , 0 0 };
sigma1 = { 0 0 , 0 0 };
sigma2 = { 0 0 , 0 0 };
sigma3 = { 0 0 , 0 0 };
sigma4 = { 0 0 , 0 0 };
yba = { 0,0};
ybb = { 0,0};
ybc = { 0,0};
ybd = { 0,0};
pdbeta= repeat (zero,1,b);
pdz1=repeat(zero,1,b);
pdzz1=repeat(zero,1,b);
pdz2=repeat(zero,1,b);
pdzz2=repeat(zero,1,b);
pdz=repeat(zero,1,b);
sdbeta=repeat(zero,b,b);
sdbetaz1=repeat(zero,b,b);
sdbetzz1=repeat(zero,b,b);
sdbetaz2=repeat(zero,b,b);
```

```
sdbetzz2=repeat(zero,b,b);
sdbetaz=repeat(zero,b,b);
m=repeat(zero,b,b);
mz1=repeat(zero,b,b);
mzz1=repeat(zero,b,b);
mz2=repeat(zero,b,b);
mzz2=repeat(zero,b,b);
mz=repeat(zero,b,b);
pdt={0 0 0};
PDTZ1={0 0 0 };
PDTZZ1={0 0 0};
PDTZ2={0 0 0};
PDTZZ2={0 0 0};
PDTZ={0 0 0};
s=ncol(sigma1);
vp={0.01};
*---------------------------------------------;
* calculo dos valores  iniciais dos parametros;
*---------------------------------------------;
*------------------------------------------------------------;
*macro para calculo das matrizes de variancia e covariancia;
*e medias amostrais ;
*============================================================;
start inicial ( z,  varcov,yb);
n =sum(z[,3]);
y=(z[,1]||z[,2])`;
t=ncol(y);
do k=1 to t;
yb = yb + y[,k]#z[k,3];
end;
yb=yb/n;
do k= 1 to t;
varcov = varcov+( (y[,k]-yb)*(y[,k]-yb)`#z[k,3])/(n-1);
end;
finish;
*----------------------------;
*rodando para amostra;
*----------------------------;
run inicial (za,varcova,yba);
run inicial (zb,varcovb,ybb);
run inicial (zc,varcovc,ybc);
run inicial (zd,varcovd,ybd);
*-----------------------------------------;
* impressao dos resultados;
* media , variancias e covariancias amostrais;
*------------------------------------------;
```

```
print yba; print varcova;
print ybb; print varcovb;
print ybc; print varcovc;
print ybd; print varcovd;
*--------------------------------------------------------;
+++++++++++++ valor inicial de sigma +++++++++++++++++;
*--------------------------------------------------------;
*------------------------------------------------------   ;
* a matriz  inicial de sigma deve ser p.d.                ;
* verificar o problema da subdispersao                    ;
* Se houve  subdispersao, iniciar com                     ;
* valores pequenos positivos para sigma(ii) e             ;
* zero para sigma(ij).                                    ;
* verificar se a matriz e  positiva definia,              ;
* senao  usar o ajuste de Marquardt                       ;
* *-------------------------------------------------------;
*========================================================;
* VARCOVA E YBA resumem as informacoes das 4 maquinas     ;
*========================================================;
start sigma(varcov,yb,sigma)global(vp);
s=ncol(varcov);
D=varcov-diag(YB);
print d;
*--------------------------------------------------;
* verificar se ha  subdispersao                    ;
*--------------------------------------------------;
do i=1 to s-1;
     do j=i+1 to s;
          if vecdiag(d)[i] < 0  then  do ;
          sigma[i,j]= 0.0 ;
          sigma[i,i]= vp;
          sigma[j,i]=sigma[i,j];
          end;
          else do;
          sigma[i,i] = log( D[i,i]/(yb[i]##2) + 1);
          sigma[i,j] = log( D[i,j]/(yb[i]#yb[j]) + 1);
          sigma[j,i]= sigma[i,j];
          end;
      end;
 end;
do i=s to s;
             if vecdiag(d)[s] < 0 then do;
             sigma[,s]  =0.0;
             sigma[s,] = 0.0;
             sigma[s,s] = vp;
             end;
```

```
             else do;
             sigma[s,s] = log( D[s,s]/(yb[s]##2) + 1);
             end;
end;
print sigma;
*-----------------------------------------------------;
* AJUSTE DE MARQUARDT                                 ;
*----------------------------------------------------;
if det(sigma) < 0 then do;
   a=abs(min(eigval(sigma)));
   SIGMA=sigma+ 1.1#DIAG(a||a);
   end;
PRINT SIGMA;
finish;
run sigma(varcova,yba,sigma1);
run sigma(varcovb,ybb,sigma2);
run sigma(varcovc,ybc,sigma3);
run sigma(varcovd,ybd,sigma4);
*=====================================================;
*==========> alterar convenientemente quando envolver;
* parametro sigma comum para varias populacoes;
*
sigma1=(sigma1+sigma2+sigma3+sigma4)/4;
print sigma1;
*=====================================================;
*---------------------------------------------------;
* decomposicao de cholesky para a matriz sigma;
*---------------------------------------------------;
T1=ROOT(SIGMA1)`;
print t1;
*-------------------------------------;
*calculo dos valores iniciais de beta;
*-----------------------------------;
betaa= (log(yba)-  0.5#vecdiag(sigma1));
betab= (log(ybb)-  0.5#vecdiag(sigma1));
betac= (log(ybc)-  0.5#vecdiag(sigma1));
betad= (log(ybd)-  0.5#vecdiag(sigma1));
*===========>alterar convenientemente;
beta1=betaa//betab//betac//betad;
PRINT BETA1;
*-----------------------------------------------;
*calculo da funcao densidade                    ;
* atraves de aproximacoes, usando               ;
*  polinomios de hermite                        ;
*  de grau 5                                    ;
* valores das raizes e constantes dos polimonios ;
```

```
*-----------------------------------------------;
v1 = { [5] 0.0 [5] -0.958572464613819 [5] 0.958572464613819
     [5] -2.020182870456086 [5] 2.020182870456086};
j1 = {0.0, -0.958572464613819, 0.958572464613819,
     -2.020182870456086, 2.020182870456086};
v2= repeat(j1,5,1)';
v=v1 // v2;
alfa1= { [5] 0.9453087204829 [10] 0.3936193231522
        [10] 0.01995324205905};
j2= { 0.9453087204829, 0.3936193231522, 0.3936193231522,
    0.01995324205905,0.01995324205905 };
alfa2= repeat(j2, 5,1)';
alfafa12=alfa1#alfa2;
*---------------------------------------------------;
*macro para calcular funcao de probabilidade ;
*---------------------------------------------------;
start fdp(x,beta,t,z,fdp)global(alfafa12,v,l,n1,n2,zero);
n1=max(z[,1]);
n2=max(z[,2]);
fdp=repeat(zero, n1+5 ,n2+5);
xbeta=repeat(x*beta,1,25);
cte=xbeta+t*v#sqrt(2);
expcte=exp(cte);
   do i = 0 to n1+4;if i = 0 then fati=1;
   else fati=fati*i;
        do j = 0 to n2+4; if j = 0 then fatj=1;
        else fatj=fatj*j;
        fatij= fati*fatj;
        y=i||j;
        h=(alfafa12#(exp(y*cte -
        l'*expcte))/(arcos(-1)))/fatij;
*-------------------------;
*p=probabilidade do vetor ;
*-------------------------;
         p=sum(h);
         fdp[i+1,j+1]=p;
   end; end;
finish;
*----------------------------------------------------;
* macro para calcular a primeira derivada  em relacao a beta;
* e a t;  e a derivada de segunda ordem em relacao a;
* beta                                               ;
*----------------------------------------------------;
start pdbeta(y1,y2,x,t,pdbeta,pdt,sdbeta,m,fdp);
y=y1//y2;
p=fdp[y1+1,y2+1];
```

```
A=fdp[y1+2,y2+1]//fdp[y1+1,y2+2];
B=((y1+1)#(y1+2)#fdp[y1+3,y2+1]||(y1+1)#(y2+1)#fdp[y1+2,y2+2
   ])// ((y1+1)#(y2+1)#fdp[y1+2,y2+2]||(y2+1) #
   (y2+2)#fdp[y1+1,y2+3]);
pdbeta=(x'*(Y#p - (y+1)#A))';
sdbeta=X'*( Y*Y'#FDP[Y1+1,Y2+1] - Y*(Y'+1)#A'- A#(Y+1)*Y' -
       DIAG((Y+1)#A) +B)*X;
m=( Y*Y'#FDP[Y1+1,Y2+1] - Y*(Y'+1)#A'- A#(Y+1)*Y' -
   DIAG((Y+1)#A) +B);
PDT=(T[,1]'*M[,1]//
    T[,1]'*M[2,]'//
    T[,2]'*M[,2])';
finish;
*-------------------------------------------------------------;
*---------------PROCESSO ITERATIVO -----------------------;
*-------------------------------------------------------------;
do iter=1 to 50;
passo= 1;
eps=0.00000000000001;
print iter passo eps;
*zera as matrizes para cada passo;
*------------------------------------------------------;
*lnv=soma acum do ln da fc de verossimilhanca;
*lnpd=primeira derivada do log verossimilhanca;
*lnsd=segunda derivada do log verossimilhanca;
*---------------------------------------------;
lnv=zero;
lnpd=repeat(zero,1,r);
lnsd=repeat(zero,r,r);
P0=BETA1'||t1[1,1]||t1[2,1]||t1[2,2];
s1=t1*t1';
P1=BETA1'||s1[1,1]||s1[2,1]||s1[2,2];
*-----------------------------------------------------;
* monta o procedimento de newton-raphson;
* para cada uma das maquinas;
* entrar com os parametros adequados para rodar;
*-----------------------------------------------;
*++++++++++++++++++++++++++++++++++++++++++++++++++++++;
* amostra da maquina a;
* za,xa;
* alterar os dados de entrada nas linhas de comando assinaladas convenientemente;
*++++++++++++++++++++++++++++++++++++++++++++++++++++++;
*==========;
*MAQUINA A1;
*==========;
run fdp(xa,beta1,t1,za,fdp); *===> trocar xa e za;
```

```
nt=nrow(za); *=====> trocar  za;
do k=1 to nt;
y=(za[,1]||za[,2])'; *=====> trocar  za;
y1=y[1,k];
y2=y[2,k];
p=fdp[y1+1,y2+1];
lnv=lnv+za[k,3]#log(p); *=============> trocar za;
run pdbeta(y1,y2,xa,t1,pdbeta,pdt,sdbeta,m,fdp); *==> trocar xa;
lnpd=lnpd+ za[k,3]#(pdbeta||pdt)/p; *===> trocar za;
run pdbeta(y1+1,y2,xa,t1,pdz1,pdtz1,sdbetaz1,mz1,fdp);*=>trocar xa;
run pdbeta(y1+2,y2,xa,t1,pdzz1,pdtzz1,sdbetzz1,mzz1,fdp);*=>trocar xa;
run pdbeta(y1,y2+1,xa,t1,pdz2,pdtz2,sdbetz2,mz2,fdp);*=>trocar xa;
run pdbeta(y1,y2+2,xa,t1,pdzz2,pdtzz2,sdbetzz2,mzz2,fdp);*=>trocar xa;
run pdbeta(y1+1,y2+1,xa,t1,pdz,pdtz,sdbetaz,mz,fdp);*=>trocar xa;
C=Y1##2#PDBETA- (2#Y1+1)#(Y1+1)#PDZ1 +(Y1+1)#(Y1+2)#PDZZ1;
D=Y1#Y2#PDBETA - Y1#(Y2+1)#PDZ2 - Y2#(Y1+1)#PDZ1
  +(Y1+1)#(Y2+1)#PDZ;
E=Y2##2#PDBETA - (2#y2+1)#(Y2+1)#PDZ2 +(Y2+1)#(Y2+2)#PDZZ2;
SDP=(T1[1,1]#C + T1[2,1]#D)//
    (T1[1,1]#D +T1[2,1]#E)//
    (T1[2,2]#E);
F=Y1##2#PDT- (2#Y1+1)#(Y1+1)#PDTZ1 +(Y1+1)#(Y1+2)#PDTZZ1;
G=Y1#Y2#PDT - Y1#(Y2+1)#PDTZ2 - Y2#(Y1+1)#PDTZ1
  +(Y1+1)#(Y2+1)#PDTZ;
H=Y2##2#PDT - (2#y2+1)#(Y2+1)#PDtZ2 +(Y2+1)#(Y2+2)#PDTZZ2;
SDT=(T1[1,1]#F + T1[2,1]#G + (M[1,]||0))//
    (T1[1,1]#G +T1[2,1]#H +(M[2,]||0))//
    (T1[2,2]#H + (0||0||M[2,2]));
*-----------------------------------------------------------;
* CALCULAR A PRIMEIRA E A SEGUNDA DERIVADA;
* DO LOG DA FUNCAO DE VEROSSIMILHANCA;
*-----------------------------------------------------------;
  *=====> trocar za;
LNSD=LNSD+Za[K,3]#(-
     ((PDBETA||PDT)/P)'*((PDBETA||PDT)/P) +
     ((SDbeta||sdp')//(sdp||sdt))/p );
end;
*==========;
*MAQUINA A2;
*==========;
run fdp(xb,beta1,t1,zb,fdp); *===> trocar xa e za;
nt=nrow(zb); *=====> trocar  za;
do k=1 to nt;
y=(zb[,1]||zb[,2])'; *=====> trocar  za;
y1=y[1,k];
y2=y[2,k];
```

```
p=fdp[y1+1,y2+1];
lnv=lnv+zb[k,3]#log(p); *=============> trocar za;
run pdbeta(y1,y2,xb,t1,pdbeta,pdt,sdbeta,m,fdp); *==> trocar xa;
lnpd=lnpd+ zb[k,3]#(pdbeta||pdt)/p;  *===> trocar za;
run pdbeta(y1+1,y2,xb,t1,pdz1,pdtz1,sdbetaz1,mz1,fdp);*=>trocar xa;
 run pdbeta(y1+2,y2,xb,t1,pdzz1,pdtzz1,sdbetzz1,mzz1,fdp);*=> trocar xa;
run pdbeta(y1,y2+1,xb,t1,pdz2,pdtz2,sdbetz2,mz2,fdp);*=>trocar xa;
run pdbeta(y1,y2+2,xb,t1,pdzz2,pdtzz2,sdbetzz2,mzz2,fdp);*=>trocar xa;
run pdbeta(y1+1,y2+1,xb,t1,pdz,pdtz,sdbetaz,mz,fdp);*=>trocar xa;
C=Y1##2#PDBETA- (2#Y1+1)#(Y1+1)#PDZ1 +(Y1+1)#(Y1+2)#PDZZ1;
D=Y1#Y2#PDBETA - Y1#(Y2+1)#PDZ2 - Y2#(Y1+1)#PDZ1
  +(Y1+1)#(Y2+1)#PDZ;
E=Y2##2#PDBETA - (2#y2+1)#(Y2+1)#PDZ2 +(Y2+1)#(Y2+2)#PDZZ2;
SDP=(T1[1,1]#C + T1[2,1]#D)//
    (T1[1,1]#D +T1[2,1]#E)//
    (T1[2,2]#E);
F=Y1##2#PDT- (2#Y1+1)#(Y1+1)#PDTZ1 +(Y1+1)#(Y1+2)#PDTZZ1;
G=Y1#Y2#PDT - Y1#(Y2+1)#PDTZ2 - Y2#(Y1+1)#PDTZ1
  +(Y1+1)#(Y2+1)#PDTZ;
H=Y2##2#PDT - (2#y2+1)#(Y2+1)#PDtZ2 +(Y2+1)#(Y2+2)#PDTZZ2;
SDT=(T1[1,1]#F + T1[2,1]#G + (M[1,]||0))//
    (T1[1,1]#G +T1[2,1]#H +(M[2,]||0))//
    (T1[2,2]#H + (0||0||M[2,2]));
*-----------------------------------------------------------;
* CALCULAR A PRIMEIRA E A SEGUNDA DERIVADA;
* DO LOG DA VEROSSIMILHANCA;
*-----------------------------------------------------------;
    *====> trocar za;
LNSD=LNSD+Zb[K,3]#(-
     ((PDBETA||PDT)/P)`*((PDBETA||PDT)/P) +
     ((SDbeta||sdp`)//(sdp||sdt))/p );
end;
*==========;
*MAQUINA B1;
*==========;
run fdp(xc,beta1,t1,zc,fdp); *===> trocar xa e za;
nt=nrow(zc); *=====> trocar  za;
do k=1 to nt;
y=(zc[,1]||zc[,2])`; *=====> trocar  za;
y1=y[1,k];
y2=y[2,k];
p=fdp[y1+1,y2+1];
lnv=lnv+zc[k,3]#log(p); *=============> trocar za;
run pdbeta(y1,y2,xc,t1,pdbeta,pdt,sdbeta,m,fdp);*==> trocar xa;
lnpd=lnpd+ zc[k,3]#(pdbeta||pdt)/p; *===> trocar za;
run pdbeta(y1+1,y2,xc,t1,pdz1,pdtz1,sdbetaz1,mz1,fdp);*>trocar xa;
```

```
run pdbeta(y1+2,y2,xc,t1,pdzz1,pdtzz1,sdbetzz1,mzz1,fdp);*=>trocar xa;
run pdbeta(y1,y2+1,xc,t1,pdz2,pdtz2,sdbetz2,mz2,fdp);*=>trocar xa;
run pdbeta(y1,y2+2,xc,t1,pdzz2,pdtzz2,sdbetzz2,mzz2,fdp);*=>trocar xa;
run pdbeta(y1+1,y2+1,xc,t1,pdz,pdtz,sdbetaz,mz,fdp);*=>trocar xa;
C=Y1##2#PDBETA- (2#Y1+1)#(Y1+1)#PDZ1 +(Y1+1)#(Y1+2)#PDZZ1;
D=Y1#Y2#PDBETA - Y1#(Y2+1)#PDZ2 - Y2#(Y1+1)#PDZ1
  +(Y1+1)#(Y2+1)#PDZ;
E=Y2##2#PDBETA - (2#y2+1)#(Y2+1)#PDZ2 +(Y2+1)#(Y2+2)#PDZZ2;
SDP=(T1[1,1]#C + T1[2,1]#D)//
    (T1[1,1]#D +T1[2,1]#E)//
    (T1[2,2]#E);
F=Y1##2#PDT- (2#Y1+1)#(Y1+1)#PDTZ1 +(Y1+1)#(Y1+2)#PDTZZ1;
G=Y1#Y2#PDT - Y1#(Y2+1)#PDTZ2 - Y2#(Y1+1)#PDTZ1
  +(Y1+1)#(Y2+1)#PDTZ;
H=Y2##2#PDT - (2#y2+1)#(Y2+1)#PDtZ2 +(Y2+1)#(Y2+2)#PDTZZ2;
SDT=(T1[1,1]#F + T1[2,1]#G + (M[1,]||0))//
    (T1[1,1]#G +T1[2,1]#H +(M[2,]||0))//
    (T1[2,2]#H + (0||0||M[2,2]));
*-----------------------------------------------------------;
* CALCULAR A PRIMEIRA E A SEGUNDA DERIVADA;
* DO LOG DA VEROSSIMILHANCA;
*-----------------------------------------------------------;
  *=====> trocar za;
LNSD=LNSD+Zc[K,3]#(-
     ((PDBETA||PDT)/P)‘*((PDBETA||PDT)/P) +
     ((SDbeta||sdp‘)//(sdp||sdt))/p );
end;
*==========;
*MAQUINA B2;
*==========;
run fdp(xd,beta1,t1,zd,fdp); *===> trocar xa e za;
nt=nrow(zd); *=====> trocar  za;
do k=1 to nt;
y=(zd[,1]||zd[,2])‘; *=====> trocar  za;
y1=y[1,k];
y2=y[2,k];
p=fdp[y1+1,y2+1];
lnv=lnv+zd[k,3]#log(p); *============> trocar za;
run pdbeta(y1,y2,xd,t1,pdbeta,pdt,sdbeta,m,fdp);*=> trocar xa;
lnpd=lnpd+ zd[k,3]#(pdbeta||pdt)/p;  *===> trocar za;
run pdbeta(y1+1,y2,xd,t1,pdz1,pdtz1,sdbetaz1,mz1,fdp);*=>trocar xa;
run pdbeta(y1+2,y2,xd,t1,pdzz1,pdtzz1,sdbetzz1,mzz1,fdp);*=> trocar xa;
run pdbeta(y1,y2+1,xd,t1,pdz2,pdtz2,sdbetz2,mz2,fdp);*=>trocar xa;
run pdbeta(y1,y2+2,xd,t1,pdzz2,pdtzz2,sdbetzz2,mzz2,fdp);*=>trocar xa;
run pdbeta(y1+1,y2+1,xd,t1,pdz,pdtz,sdbetaz,mz,fdp);*=>trocar xa;
C=Y1##2#PDBETA- (2#Y1+1)#(Y1+1)#PDZ1 +(Y1+1)#(Y1+2)#PDZZ1;
```

```
D=Y1#Y2#PDBETA - Y1#(Y2+1)#PDZ2 - Y2#(Y1+1)#PDZ1
  +(Y1+1)#(Y2+1)#PDZ;
E=Y2##2#PDBETA - (2#y2+1)#(Y2+1)#PDZ2 +(Y2+1)#(Y2+2)#PDZZ2;
SDP=(T1[1,1]#C + T1[2,1]#D)//
    (T1[1,1]#D +T1[2,1]#E)//
    (T1[2,2]#E);
F=Y1##2#PDT- (2#Y1+1)#(Y1+1)#PDTZ1 +(Y1+1)#(Y1+2)#PDTZZ1;
G=Y1#Y2#PDT - Y1#(Y2+1)#PDTZ2 - Y2#(Y1+1)#PDTZ1
  +(Y1+1)#(Y2+1)#PDTZ;
H=Y2##2#PDT - (2#y2+1)#(Y2+1)#PDtZ2 +(Y2+1)#(Y2+2)#PDTZZ2;
SDT=(T1[1,1]#F + T1[2,1]#G + (M[1,]||0))//
    (T1[1,1]#G +T1[2,1]#H +(M[2,]||0))//
    (T1[2,2]#H + (0||0||M[2,2]));
*-----------------------------------------------------------;
* CALCULAR A PRIMEIRA E A SEGUNDA DERIVADA;
* DO LOG DA VEROSSIMILHANCA;
*-----------------------------------------------------------;
    *====> trocar za;
LNSD=LNSD+Zd[K,3]#(-
     ((PDBETA||PDT)/P)`*((PDBETA||PDT)/P) +
     ((SDbeta||sdp`)//(sdp||sdt))/p );
end;
determ=det(lnsd);
np=p0-passo#(lnpd*inv(lnsd));
*------------------------------;
*varassin=variancia assintotica de p0;
*------------------------------;
varassin=inv(-lnsd);
*-------------------------------------------------;
* matriz tr = matriz de transf de t para sigma;
* dimensao de tr deve ser de rXr;
* onde as r-3 linhas estao relacionadas com a varassint de beta;
* as 3 ultimas estao relacionadas com a varassint de sigma;
*=====================================================;

  tta=repeat(zero,b,3);
  tt=(2#t1[1,1]||0||0)//
     (t1[2,1]||t1[1,1]||0)//
     (0||2#t1[2,1]||2#t1[2,2]);
  tr=(I(b)||tta)//
     (tta`||tt);
*--------------------------------;
* varas=variancia assintotica de;
* das estimativas de beta e sigma;
*--------------------------------;
varas=tr*varassin*tr`;
```

```
*--------------------------------------------------------;
* DESIGNACAO DE NOVOS VALORES DE BETA E T;
* BETA = SAO OS R-3 PRIMEIROS VALORES DE NP    ;
* T = SAO OS 3 ULTIMOS VALORES DE NP;
*-------------------------------------------------------;
BETA1=NP[1:b];
T1= (NP[b+1]||0)//(NP[b+2]||NP[b+3]);
s2=t1*t1`;
np1=np[1:b]`||
    s2[1,1]||s2[2,1]||s2[2,2];
*---------------------------------------------;
* criterio de convergencia;
*---------------------------------------------;
     do i= 1 to r;
if np[i] ^=0 then cc[i]=(np1[i]-p1[i])/np1[i];
else do;
          if np1[i] =0 & p1[i] = 0 then cc[i] =0;
          else do ; if np1[i]=0 & p1[i] ^=0 then cc[i]=1;
               end;
          end;
     end;
if max(abs(cc)) < eps then stop;
end;
*----------------------------------------------------;
*impressao dos resultados;
*----------------------------------------------------;
print lnsd;  print determ;
PRINT VARASSIN;
print varas;
print lnv;
print lnpd;
print p0; print np;
print p1; print np1;
*==================================;
* CALCULO DAS ESTIMATIVAS DE E(Y) E;
* VAR(Y);
*=================================;
BETA0=P0[1:b];
T0= (p0[b+1]||0)//(p0[b+2]||p0[b+3]);
SIGMA0=T0*T0`;
 JK= I(b/2) @ (SIGMA0);
EY=exp(beta0+1/2#vecdiag(jk));
VarY=diag(EY)+ (EY*EY`)#(exp(jk)-1);
print ey;
print vary;
*======================================;
```

```
* calculo para obter as estimativas da variancia;
* assintotica de EY e VARY ;
* montando a matriz das derivadas;
*========================================;
JL=repeat(I(2),b/2,1);
ZB= repeat(zero,b,1);
JM=DIAG(EY) + 2#DIAG( VARY - DIAG(EY) );
JN=repeat(zero,b/2,b);
JO=VECDIAG( (1/2)#DIAG(EY) + (EY)*(EY')#(2#EXP(JK)-1) ) # J1;
jp=REPEAT(ZERO, B/2,3);
do I= 1 to b/2;
JN[i,2#i-1]=Vary[2#i-1,2*i];
JN[i,2#i]=JN[i,2#i-1];
JP[i,1]=Vary[2#i-1,2#i];
JP[i,2]=JP[i,1];
JP[i,3]=exp( beta0[2#i-1] +beta0[2#i] + 1/2#(SIGMA0[1,1] +
        SIGMA0[2,2] + SIGMA0[1,2]) );
end;
*====================================;
* MTR= MATRIZ DE TRANSFORMACAO;
*====================================;
MTR=( DIAG (EY) || ( ( (1/2) #EY#JL) ||ZB) ) //
     ( ( JM//JN ) || ( (JO||ZB)//JP) );
*=======================================================;
* ONDE ( DIAG (EY) = DERIVADA DE EY EM RELACAO A BETA;
*  ( ( (1/2) #EY#JL) ||ZB) ) = DERIVADA DE EY EM RELACAO A SIGMA;
* ( JM//JN )= DERIVADA DE VARY EM RELACAO A BETA;
* ( (JO||ZB)//JP) = DERIVADA DE VARY EM RELACAO A SIGMA;
*=====================================================;
VAREST=MTR*VARAS*MTR';
*VAREST= VARIANCIA ESTIMATIDA DA EY E VARY;
PRINT VAREST;
```

Apêndice B

Programa para calcular as estimativas de MV dos parâmetros de uma distribuição Poisson Bivariada

Este programa foi escrito através da linguagem IML do pacote estatístico SAS. Ele foi utilizado para calcular as estimativas de MV de parâmetros de uma distribuição Poisson Bivariada através do método de Newton-Raphson. Esta listagem corresponde à obtenção da estimativa dos parâmetros do modelo de Regressão Poisson Bivariada (3.12). Os resultados dos outros exemplos podem ser obtidos com algumas alterações apropriadas.

```
proc iml worksize=250; reset nolog;
options linesize=132 pagesize=500;
*=====dados da maquina A1=================;

za={ 0 0 13, 0 1 26, 0 2 22, 0 3 16, 0 4 3,
0 5 2, 1 0 3, 1 1  9, 1 2 12, 1 3 7, 1 4 5,
1 5 2, 1 6 1, 2 1 2, 2 2 1, 2 5 1, 3 1 1 ,3 3 2,  3 7 1, 4 2 1};

*=======dados da maquina A2=============;
zb={ 0 0 15, 0 1 14, 0 2 11, 0 3 10, 0 4 3,
0 5 1, 1 0 2, 1 1 12, 1 2 6, 1 3 7, 1 4 5,
1 5 3, 1 7 1, 2 1 1, 2 2 2, 2 3 1, 2 5 2 ,3 4 1, 3 5 1, 3 7 1, 4 3 1};

*======dados da maquina B1==============;
zc={ 0 0 11, 0 1 35, 0 2 21, 0 3 13, 0 4 5,
```

```
0 5 1, 1 0 1, 1 1 5 , 1 2 3, 1 3 1, 1 4 1,
             2 1 1,  2 3 2                              };

*==========dados da maquina B2===========;
zd={ 0 0 20, 0 1 35, 0 2 17, 0 3 7 , 0 4 4,
0 5 1, 1 0 4, 1 1 7 , 1 2 3, 1 3 2};
*  ;

zero ={ 0};
r= 3;
zero4=repeat(zero,2,2);
para=repeat(zero, 1, r);
npara=repeat(zero, 1,r);

*==macro para calcular media, variancia e covariancia amostrais===;
*================================================================;
start inicial (z, varcov, yb);

n=sum(z[,3]);
y=(z[,1]||z[,2])`;
t=ncol(y);

do k=1 to t;
yb = yb + y[,k]#z[k,3];
end;

yb=yb/n;

do k=1 to t;
varcov=varcov + ( (y[,k] - yb)*(y[,k] - yb)`#z[k,3] )/(n-1);
end;
finish;

varcova=repeat(zero, 2,2); yba= repeat(zero,2, 1);
run inicial (za, varcova, yba);
cov=varcova[1,2];
print varcova; print yba;
*-----------------------------------------------;
* macro para calcular a ftp;
*-----------------------------------------------;
*l = covariancia l1 = lambda1 l2 = lambda2;

start fdp( fdp, z, l, l1, l2  )
global(zero                    );

n1=max(z[,1]);
```

```
n2=max(z[,2]);

fdp=repeat(zero, n1+3,n2+3);

*Atencao -------------------------------;

*===========================================;
* a matriz fdp foi dimensionada de modo a;
*ter duas primeiras linhas e duas primeiras;
*colunas iguais a zero, para facilitar as;
*derivadas futuras, a probabilidade dos pontos;
*comeca na terceira linha, e coluna , equivale o;
*ponto zero,zero;
*================================================;

fdp[3,3]= exp(-l1-l2+l);

do i = 4 to n2+3;
fdp[3,i]=(l2-l)#fdp[3,i-1]/(i-3);
end;

do j = 4 to n1+3;
fdp[j,3]=(l1-l)#fdp[j-1,3]/(j-3);
end;

do i=4 to n1+3;
do j=4 to n2+3;
fdp[i,j]= ((l1-l)#fdp[i-1,j]+l#fdp[i-1,j-1])/(i-3);
end;
end;

finish;

*--------------------------------------------;
*macro para calcular a primeira derivada ;
*--------------------------------------------;

*pdl1=primeira derivada em relacao a beta1;
*pdl2=primeira derivada em relacao a beta2;
*pdl=primeira derivada em relacao a theta;

start pd  ( y1,y2,fdp, pdl1,pdl2,pdl) global (l1, l2, l);
b1=log(l1);
b2=log(l2);
b=sqrt(l);
```

```
pdl1= (fdp[y1+2,y2+3] - fdp[y1+3,y2+3])#(exp(b1));
pdl2= (fdp[y1+3,y2+2] - fdp[y1+3,y2+3])#(exp(b2));
pdl = (fdp[y1+3,y2+3] - fdp[y1+2,y2+3] - fdp[y1+3,y2+2]
      + fdp[y1+2,y2+2])#2#b;

finish;

*rodando para as amostras;
*========================;

l1=yba[1]; l2=yba[2]; l=cov;
b1=log(l1); b2=log(l2);
if l <= 0 then l=0.0001;
 b=sqrt(l);
para=log(yba)`||b;

niter=50;
do iter= 1 to niter;   print iter;
passo = 1;
eps= 0.00001;

lnv =zero;
lnpd= repeat(zero, 1, r);
lnsd = repeat(zero, r,r);

*para amostra da maquina A;
*=========================;

run fdp( fdp, za, l  , l1    , l2    );

nt= nrow(za);

      do k = 1 to nt;
      y=(za[,1]||za[,2])` ;
      y1=y[1,k];
      y2=y[2,k];
      p = fdp[y1+3,y2+3];

      lnv=lnv+za[k,3]#log(p);


      run pd ( y1, y2, fdp, l1y, l2y, ly);
      lnpd=lnpd + za[k,3]#(l1y||l2y||ly)/p;


*montando a segunda derivada;
```

```
*============================;
          run pd (y1-1,y2,fdp,l1z1, l2z1, lz1);
          run pd (y1, y2-1, fdp, l1z2, l2z2, lz2);
          run pd (y1-1, y2-1,fdp, l1zz, l2zz, lzz);


          sdl1 = (l1z1- l1y)#(exp(b1)) + l1y;
          sdl1l2= (l2z1 - l2y)#(exp(b1));
          sdl1l= (lz1 - ly)#(exp(b1)) ;

          sdl2l1=(l1z2 - l1y)#(exp(b2));
          sdl2 = (l2z2 - l2y)#(exp(b2)) +l2y;
          sdl2l= (lz2 - ly)#(exp(b2)) ;


          sdll1= (l1y - l1z1 - l1z2 + l1zz)#2#b;
          sdll2 = (l2y - l2z1 - l2z2 + l2zz)#2#b;
          sdl = (ly - lz1 - lz2 + lzz)#2#b+ly/b;

          sd=(sdl1||sdl1l2||sdl1l)//
             (sdl2l1||sdl2||sdl2l)//
             (sdll1||sdll2||sdl);

         lnsd=lnsd+za[k,3]#(
              sd/p -  ((l1y||l2y||ly)/p)`*((l1y||l2y||ly)/p) );

         end;

npara= para - passo#(lnpd*inv(lnsd));

print npara; print para;
print lnv;
varassin=inv(-lnsd);

*=======================================;
*designacao dos novos valores;
*=======================================;

l1=exp(npara[1,1]);
l2=exp(npara[1,2]);
l=(npara[1,3])#(npara[1,3]);

b1=npara[1,1];
b2=npara[1,2];
b=npara[1,3];
```

```
cc= (npara-para);

para=npara[1,1]||npara[1,2]||npara[1,3];
if max(abs(cc)) < eps then stop;
end;

print lnv;
print lnpd;
print lnsd;
print varassin;
```

Apêndice C

Programa para gerar amostras sob uma distribuição Poisson Log-Normal Bivariada

Este programa gera 25 amostras aleatórias de tamanho 50 de uma distribuição Poisson Log-Normal de parâmetros $\boldsymbol{\mu}=(-0.1,-0.1)'$ e $\boldsymbol{\Sigma}=\begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 0.2 \end{pmatrix}$ com $\boldsymbol{\Sigma}=\mathbf{AA}'$, onde $\mathbf{A}=\begin{pmatrix} 0.3872983 & 0.2236068 \\ -0.3872983 & 0.2236068 \end{pmatrix}$

```
options linesize=132 pagesize=500;
data a;
*---------------------------------------;
* parametros para simulacao das amostras;
*nr=numero de amostras aleatorias;
*ta=tamanho da amostra;
*z1 e z2 sao variaveis aleatorias normais
* padronizadas independentes;
*mi1 e mi2 = valores dos parametros;
* a11, a12,a21 e a22 = elementos da ;
*                      matriz A ;
* y1 e y2 = variaveis poisson lognormal;
*---------------------------------------;

nr=25 ;ta=50;
   do j=1 to nr;
       do i=1 to ta;
       z1=normal(0);
       z2=normal(0);
       a11= 0.3872983;a12=0.2236068;
       a21=-0.3872983;a22=0.2236068;
```

```
        mi1= -.1 ;mi2=-.1 ;
        x1=a11*z1+a12*z2+mi1;
        x2=a21*z1+a22*z2+mi2;
        l1=exp(x1);
        l2=exp(x2);
        y1=ranpoi(0,l1);
        y2=ranpoi(0,l2); output;
        end;
     end;
  proc freq; tables j*y1*y2/out=sai norow nocol nopercent noprint;
  data c; set sai; keep y1 y2 count j;
  proc freq; tables j/out=s noprint ;data d;set s; put count @;
*-----------------------------------------------------;
* count = igual ao numero de pares distintos de y1 e y2
          nas nr amostras;
*----------------------------------------------------;
16 14 14 14 19 20 19 16 16 18 16 16 12 17 11 14 21 15 17 17
 14 19 16 14 16
  data c3; set c; put y1 @;
*-------------;
*valores de y1;
*-------------;
0 0 0 0 1 1 1 1 1 2 2 2 2 3 4 5 0 0 0 0 1 1 1 1 2 2 2 3 4 5 0 0 0 0
 1 1 1 1 2 2 2 3 4 4 0 0 0 0 0 1 1 1 1 2 2 2 3 3 0 0 0 0 0 1 1 1
1 2 2 2 2 2 3 3 4 4 6 0 0 0 0 1 1 1 1 1 1 2 2 2 2 2 3 4 4 5 7 0 0 0
 0 0 1 1 1 1 2 2 2 2 3 3 4 4 4 5 0 0 0 0 1 1 1 1 1 2 2 2 3 3 4 5
0 0 0 0 0 1 1 1 1 2 2 2 3 3 4 5 0 0 0 0 0 1 1 1 1 2 2 2 3 3 3 3 4 5
 0 0 0 0 1 1 1 1 2 2 2 3 3 3 5 6 0 0 0 0 0 0 1 1 1 2 2 2 2 2 3 3
0 0 0 0 1 1 1 2 2 2 3 4 0 0 0 0 0 1 1 1 1 1 2 2 2 3 3 3 4 0 0 0 0 1
 1 1 2 2 3 3 0 0 0 0 0 1 1 1 2 2 2 2 2 3 0 0 0 0 0 1 1 1 1 1 2 2
2 2 2 3 3 3 4 4 4 0 0 0 0 1 1 1 1 2 2 3 3 3 4 7 0 0 0 0 0 1 1 1 1 2
 2 2 3 3 4 5 5 0 0 0 0 0 1 1 1 2 2 2 3 3 3 4 4 5 0 0 0 0 1 1 1 1
2 2 2 3 4 4 0 0 0 0 1 1 1 1 1 1 2 2 2 2 3 3 3 4 5 0 0 0 0 1 1 1 1 1
 2 2 2 3 3 3 4 0 0 0 0 0 1 1 1 1 2 2 2 3 5 0 0 0 0 1 1 1 1 2 2 2
2 3 3 5 7
  data c1; set c; put y2 @;
*-------------------;
*valores de y2;
*-------------------;
0 1 2 4 0 1 2 3 4 0 1 2 6 1 0 1 0 1 2 3 0 1 2 5 0 1 2 1 1 2 0 1 2 5
 0 1 2 3 0 1 2 1 0 1 0 1 2 3 6 0 1 2 3 0 1 2 0 1 0 1 2 3 4 0 1 2
3 0 1 2 3 4 0 1 0 3 0 0 1 2 3 0 1 2 3 4 8 0 1 2 3 4 4 0 2 0 0 0 1 2
 3 4 0 1 2 3 0 1 3 4 0 1 0 1 2 0 0 1 2 4 0 1 2 3 4 0 2 3 0 1 1 1
0 1 2 3 4 0 1 2 3 0 1 2 0 1 1 1 0 1 2 3 5 0 1 2 3 0 1 2 0 1 2 4 0 0
 0 1 2 3 0 1 2 3 0 1 2 1 2 3 1 0 0 1 2 3 5 6 0 1 3 0 1 2 3 4 1 2
0 1 2 3 0 1 2 0 1 2 1 0 0 1 2 3 4 0 1 2 3 4 0 1 2 0 1 2 0 0 1 2 3 0
```

```
 1 2 0 1 2 3 0 1 2 3 4 0 1 2 0 1 2 3 4 1 0 1 2 3 4 0 1 2 3 5 0 1
2 3 4 0 1 2 0 2 3 0 1 2 3 0 1 2 3 0 1 0 1 2 0 1 0 1 2 3 4 0 1 2 3 0
 1 2 0 1 3 0 1 0 1 2 3 4 0 1 2 0 1 2 0 1 2 0 1 0 0 1 2 3 0 1 2 3
0 1 2 1 0 2 0 1 2 3 0 1 2 3 4 5 0 1 2 3 0 2 3 2 1 0 1 2 3 0 1 2 3 4
 0 1 3 0 1 3 2 0 1 2 3 4 0 1 2 4 0 1 3 0 2 0 1 2 3 0 1 2 3 0 1 2
3 0 5 0 2
  data c2; set c; put count @;
*----------------------------------------------;
*frequencia do par y1 e y2;
*----------------------------------------------;
1 15 3 1 5 5 2 1 1 3 4 2 1 4 1 1 5 6 4 1 12 7 4 1 4 1 1 2 1 1 9 8 2
 1 6 5 1 2 4 7 2 1 1 1 7 6 5 1 1 8 4 5 2 5 2 2 1 1 7 4 4 2 1 2 6
2 1 5 4 2 1 1 2 3 1 1 1 6 7 3 1 6 6 3 2 1 1 2 2 2 2 1 1 1 1 1 1 3 7
 2 2 2 9 7 3 1 3 2 1 1 2 1 1 1 1 1 5 8 3 1 10 4 6 1 1 2 2 2 1 2 1
1 7 7 4 1 1 9 4 3 2 2 2 1 3 1 2 1 5 9 3 2 1 2 6 1 2 1 5 3 4 1 1 1 1
 2 7 10 4 1 4 8 2 3 2 1 3 1 1 1 1 1 6 4 1 1 1 1 9 10 1 5 4 3 1 1
1 1 8 6 3 1 7 5 2 11 3 1 2 1 11 8 1 1 1 7 3 2 1 1 3 1 1 3 4 1 1 7 9
 5 3 10 5 5 3 1 1 1 7 10 1 1 1 10 3 2 3 5 2 3 1 1 10 4 2 5 1 5 4
1 1 1 3 3 1 2 1 1 1 1 1 1 1 9 6 2 3 10 7 1 1 1 4 1 1 2 1 1 7 6 1 7
 1 8 3 2 3 2 1 2 3 1 1 1 1 9 4 4 1 1 5 9 4 1 2 1 4 1 1 1 1 1 5 9 4
2 6 6 3 1 3 3 4 2 1 1 6 5 2 3 10 8 1 1 1 1 1 1 1 1 3 2 1 1 1 10 5 2
 3 10 2 4 2 1 3 3 1 1 1 1 1 11 8 4 1 2 7 5 1 2 2 4 1 1 1 5 4 4 4
5 6 6 1 2 3 2 3 2 1 1 1
  run;
```