MAE5763 - Modelos Lineares Generalizados - Resolução da Lista 2

Guilherme Marthe - 8661962

3/11/2020

1 Exercício 1

No enunciado é apresentado o seguinte modelo

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Onde $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ e σ^2 é conhecido. É pedido mostrar as equivalências entre as estatisticas de Wald, Razão de verossimilhanças e Escore quando se testa $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$. Como sugerido, vamos deixar todas elas em função da estimativa $\hat{\beta}$.

1.1 Estatística de Wald

Quando a estatística de Wald testa somente um parâmetro, como é o nosso caso, ela tem a seguinte forma

$$\xi_W = \frac{\left(\hat{\beta} - \beta^0\right)^2}{\hat{\text{Var}}(\hat{\beta})}$$

Para a hipótese solicitada $\beta^0 = 0$. Abaixo calculamos a variância de $\hat{\beta}$ a partir da matriz modelo X. Seja $\theta = (\alpha, \beta)^T$ o vetor de parâmetros do modelo em questão. A matriz de variância e covariância de θ é calculada por:

$$\text{Var}(\hat{\theta}) = \sigma^2 \left(X^T X\right)^{-1}$$

Onde a matriz modelo X é

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

E

$$\begin{aligned} X^T X &= \begin{bmatrix} 1 & ... & 1 \\ x_1 & ... & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \\ &= \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix} \end{aligned}$$

Uma vez que $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. O inverso é de $X^T X$ é

$$(X^T X)^{-1} = \frac{1}{nS_{xx}} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

Onde $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$ E $\text{Var}(\hat{\theta}) = \frac{\sigma^2}{nS_{xx}} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$. Então, $\text{Var}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$.

Concluímos então que a estatistica de Wald é

$$\xi_W = \frac{\left(\hat{\beta}\right)^2}{\hat{\text{Var}}(\hat{\beta})} = \frac{\beta^2 S_{xx}}{\sigma^2}$$

Onde $\text{Var}(\hat{\beta}) = \hat{\text{Var}}(\hat{\beta})$, uma vez que σ^2 é conhecido.

##Estatística de razão de vessossimilhanças

A estatística de razão de verrossimilhanças é calculada a partir do valor da função de verossimilhanças do modelo estimado sob H_1 e sob H_0. A expressão dessa estatística é

$$\xi_{RV} = 2\{L(\hat{\theta}) - L(\theta^0)\}$$

Todavia, o modelo linear que estamos avaliando é um caso de um MLG, sendo a distribuição normal a distribuição pertencente à família exponencial de interesse e a ligação canônica que é a identidade. Assim a estatistica do teste de razão de verossimilhanças pode ser escrita como uma diferença de desvios:

$$\xi_{RV} = \phi \left[D(y; \mu^0) - D(y; \hat{\mu}) \right]$$

Onde, em termos dos parâmetros do modelo proposto

$$\begin{aligned} \hat{\mu}_i &= \hat{\alpha} + \hat{\beta} x_i \\ &= \bar{y} - \hat{\beta}\bar{x} + \hat{\beta} x_i \\ &= \bar{y} + \hat{\beta}(x_i - \bar{x}) \end{aligned} \tag{1}$$

No caso normal, os desvios são:

$$\xi_{RV} = \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} (y_i - \hat{\mu}_i^0)^2 - \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 \right)$$

Vamos manipulá-lo para deixa-lo em função das estimativas de β:

$$\begin{aligned} \xi_{RV} &= \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} (y_i - \hat{\mu}_i^0)^2 - \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 \right) \\ &= \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} y_i^2 - 2\hat{\mu}_i^0 y_i + (\mu_i^0)^2 - y_i^2 - \hat{\mu}_i^2 + 2\hat{\mu}_i y_i \right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(\mu_i^0)^2 - \hat{\mu}_i^2 - 2y_i(\hat{\mu}_i^0 - \hat{\mu}_i)\right) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}((\hat{\mu}_i^0 + \hat{\mu}_i)(\hat{\mu}_i^0 - \hat{\mu}_i) - 2y_i(\hat{\mu}_i^0 - \hat{\mu}_i)\right) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}((\hat{\mu}_i^0 + \hat{\mu}_i) - 2y_i)(\hat{\mu}_i^0 - \hat{\mu}_i)\right)
\end{aligned}
$$

Agora vamos substituir os parâmetros da média usando (1) e o fato de que $\hat{\mu}_i^0 \overset{H_0}{=} \hat{\alpha}^0 \overset{H_0}{=} \bar{y}$:

$$
\begin{aligned}
\xi_{RV} &= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}((\hat{\mu}_i^0 + \hat{\mu}_i) - 2y_i)(\hat{\mu}_i^0 - \hat{\mu}_i)\right) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(2\bar{y} + \hat{\beta}(x_i - \bar{x}) - 2y_i)(-\hat{\beta}(x_i - \bar{x}))\right) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(\hat{\beta}(x_i - \bar{x}) - 2(y_i - \bar{y}))(-\hat{\beta}(x_i - \bar{x}))\right) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}2\hat{\beta}(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}^2(x_i - \bar{x})^2\right) \\
&= \frac{1}{\sigma^2}2\hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{\sigma^2}\left[2\hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2\right] \\
&= \frac{1}{\sigma^2}\left[2\hat{\beta}S_{xy} - \hat{\beta}^2 S_{xx}\right] \\
&= \frac{1}{\sigma^2}\left[2\hat{\beta}S_{xy}\frac{S_{xx}}{S_{xx}} - \hat{\beta}^2 S_{xx}\right] \\
&= \frac{1}{\sigma^2}\left[2\hat{\beta}^2 S_{xx} - \hat{\beta}^2 S_{xx}\right] \\
&= \frac{\hat{\beta}^2 S_{xx}}{\sigma^2}\left[2 - 1\right] \\
&= \frac{\hat{\beta}^2 S_{xx}}{\sigma^2} = \xi_W
\end{aligned}
$$

Então, de fato as estatísticas de Wald e de Razão de verossimilhanças são iguais.

1.2 Estatística de Escore

A estatística de escore não depende de se calcular os valores na função de verossimilhança, apenas sob a função escore, que é a derivada da função de verossimilhança. Para o caso de teste uniparamétrico a estatística tem a seguinte forma:

$$
\xi_{SR} = \left[\hat{U}_\beta(\beta^0)\right]^2 \cdot \hat{\text{Var}}_0(\hat{\beta})
$$

As funções escore para os parâmetros do modelo $\theta = (\alpha, \beta)^T$ podem ser obtidas por meio de forma matricial $U_\theta = \left[U_\alpha, U_\beta\right]^T = \sigma^{-2}X^T(y - \mu)$. Apresentação do desenvolvimento dessas expressões a seguir:

$$
\begin{aligned}
U_\theta &= \sigma^{-2} \begin{bmatrix} 1 & ... & 1 \\ x_1 & ... & x_n \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ \vdots \\ y_n - \mu_n \end{bmatrix} \\
&= \sigma^{-2} \begin{bmatrix} \sum_{i=1}^n y_i - \mu_i \\ \sum_{i=1}^n x_i(y_i - \mu_i) \end{bmatrix}
\end{aligned}
$$

Como explicado, só precisamos usar $U_\beta = \sigma^{-2} \sum_{i=1}^n x_i(y_i - \mu_i)$. É importante notar o seguinte fato:

$$
\begin{aligned}
S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^n x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x}\bar{y} \\
&= \sum_{i=1}^n x_i y_i - x_i \bar{y} - \sum_{i=1}^n y_i \bar{x} - \bar{x}\bar{y} \\
&= \sum_{i=1}^n x_i y_i - x_i \bar{y} + \bar{x} \sum_{i=1}^n y_i - \bar{y} \\
&= \sum_{i=1}^n x_i y_i - x_i \bar{y} + 0 \\
&= \sum_{i=1}^n x_i (y_i - \bar{y})
\end{aligned}
$$

Uma vez que a soma dos desvios das observações com relação à media é zero. Assim, sob H_0, $\hat{\mu}_i \overset{H_0}{=} \alpha^0 \overset{H_0}{=} \bar{y}$ e então:

$$
\begin{aligned}
\hat{U}_\beta(\beta^0) &= \sigma^{-2} \sum_{i=1}^n x_i (y_i - \mu_i) \\
&= \sigma^{-2} \sum_{i=1}^n x_i (y_i - \hat{\mu}) \\
&= \sigma^{-2} S_{xy}
\end{aligned}
$$

A variância assintótica de β é a mesma que já apresentamos, e não há alterações nela, uma vez que não depende de parâmetros do modelo e σ^2 é conhecido. Assim $\hat{\text{Var}}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$. Assim a estatística de escore é dada por

$$
\begin{aligned}
\xi_{SR} &= \left[\hat{U}_\beta(\beta^0)\right]^2 \cdot \hat{\text{Var}}_0(\hat{\beta}) \\
&= \sigma^{-4} S_{xy}^2 \frac{\sigma^2}{S_{xx}} \\
&= \sigma^{-4} S_{xy}^2 \frac{\sigma^2 S_{xx}}{S_{xx} S_{xx}} \\
&= \sigma^{-2} \hat{\beta}^2 S_{xx}
\end{aligned}
$$

$$= \frac{\hat{\beta}^2 S_{xx}}{\sigma^2} = \xi_W = \xi_{RV}$$

Mostrando a equivalência requisitada.

2 Exercício 2

O modelo apresentado pelo enunciado é o seguinte:

$$Y_{ij}|x_j \overset{\text{ind}}{\sim} \text{FE}(\mu_{ij}, \phi)$$

Com a parte sistemática dada por

$$g(\mu_{ij}) = \eta_{ij} = \alpha_i + \beta x_j$$

Com $i = 1, 2$ e $j = 1, \dots, r$. Note que esse modelo equivale à um modelo de retas paralelas com a mesma inclinação β. Pede-se para encontrar a estatística do teste de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$. Ou seja, iremos testar se as retas paralelas correspondentes aos grupos $i = 1$ e $i = 2$ são horizontais ou têm uma inclinação não nula.

A matriz modelo deste modelo é a seguinte. Note que as primeiras r linhas correspondem ao modelo quando $i = 1$ e as últimas r linhas quando $i = 2$.

$$X = \begin{bmatrix} 1 & 0 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_r \\ 0 & 1 & x_1 \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_r \end{bmatrix}_{2r \times 3}$$

Definamos os seguinte vetores para desenvolvermos a estatística de score:

$$\theta = \begin{bmatrix} \beta \\ \alpha_1 \\ \alpha_2 \\ \phi \end{bmatrix}, \quad \theta^0 = \begin{bmatrix} 0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \phi_0 \end{bmatrix}$$

Onde θ corresponde ao vetor de parâmetros do modelo completo e θ^0 é o vetor de parâmetros proposto em H_0. Assim, estatística de escore é a seguinte:

$$\xi_{SR} = \left[\hat{U}_\beta^0(\theta^0)\right]^T \hat{\text{Var}}_0(\hat{\beta})\, \hat{U}_\beta^0(\theta^0)$$

$\hat{U}_\beta^0(\theta^0)$ é a função score associada ao parâmetro β do modelo proposto, calculada com as estimativas do modelo sob H_0. $\hat{\text{Var}}_0(\hat{\beta})$ é a variância assintótica de $\hat{\beta}$ estimada sob H_0.

2.1 Variância assintótica de $\hat{\beta}$

Iniciando pelo componente $\text{Var}(\hat{\beta})$. No caso dos MLGs em que se testam os parâmetros em modelos encaixados, como no nosso caso, podemos decompor a variância assintótica de $\hat{\beta}$ da sequinte maneira:

$$\text{Var}(\hat{\beta}) = \phi^{-1}(R^T W R)^{-1}$$

Com $R = X_1 - X_2 C$ e $C = (X_2^T W X_2) X_2^T W X_1$. X_1 e X_2 são partições da matriz modelo X. Concretamente, elas têm a seguinte forma:

$$X_1 = \begin{bmatrix} x_1 \\ \vdots \\ x_r \\ x_1 \\ \vdots \\ x_r \end{bmatrix}_{2r \times 1} \qquad X_2 = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}_{2r \times 2}$$

Note que C equivale ao vetor de parâmetros provindas da regressão de X_2 em relação às colunas de X_1. O cálculo de C é desenvolvido à seguir:

$$\begin{aligned} X_2^T W &= \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \end{bmatrix} \cdot \text{diag}\{\omega_{11}, \cdots, \omega_{1r}, \omega_{21}, \cdots, \omega_{2r}\} \\ &= \begin{bmatrix} \omega_{11} & \cdots & \omega_{1r} & 0 & \dots & 0 \\ 0 & \dots & 0 & \omega_{21} & \cdots & \omega_{2r} \end{bmatrix} \\ X_2^T W X_2 &= \begin{bmatrix} \omega_{11} & \dots & \omega_{1r} & 0 & \dots & 0 \\ 0 & \dots & 0 & \omega_{21} & \dots & \omega_{2r} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^r w_{1j} & 0 \\ 0 & \sum_{j=1}^r w_{2j} \end{bmatrix} \end{aligned}$$

E invertendo a matriz do resultado anterior

$$(X_2^T W X_2)^{-1} = \begin{bmatrix} \frac{1}{\sum_{j=1}^r w_{1j}} & 0 \\ 0 & \frac{1}{\sum_{j=1}^r w_{2j}} \end{bmatrix}$$

O segundo componente de C

$$X_2^T W X_2 = \begin{bmatrix} \omega_{11} & \dots & \omega_{1r} & 0 & \dots & 0 \\ 0 & \dots & 0 & \omega_{21} & \dots & \omega_{2r} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_r \\ x_1 \\ \vdots \\ x_r \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^r x_j w_{1j} \\ \sum_{j=1}^r x_j w_{2j} \end{bmatrix}$$

Por fim

$$C = \begin{bmatrix} \frac{1}{\sum_{j=1}^r w_{1j}} & 0 \\ 0 & \frac{1}{\sum_{j=1}^r w_{2j}} \end{bmatrix} \cdot \begin{bmatrix} \sum_{j=1}^r x_j w_{1j} \\ \sum_{j=1}^r x_j w_{2j} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{j=1}^r x_j w_{1j}}{\sum_{j=1}^r w_{1j}} \\ \frac{\sum_{j=1}^r x_j w_{2j}}{\sum_{j=1}^r w_{2j}} \end{bmatrix}$$

$$\begin{aligned} R &= X_1 - X_2 C \\ &= \begin{bmatrix} x_1 \\ \vdots \\ x_r \\ x_1 \\ \vdots \\ x_r \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sum_{j=1}^r x_j w_{1j}}{\sum_{j=1}^r w_{1j}} \\ \frac{\sum_{j=1}^r x_j w_{2j}}{\sum_{j=1}^r w_{2j}} \end{bmatrix} \\ &= \begin{bmatrix} x_1 - \frac{\sum_{j=1}^r x_j w_{1j}}{\sum_{j=1}^r w_{1j}} \\ \vdots \\ x_r - \frac{\sum_{j=1}^r x_j w_{1j}}{\sum_{j=1}^r w_{1j}} \\ x_1 - \frac{\sum_{j=1}^r x_j w_{2j}}{\sum_{j=1}^r w_{2j}} \\ \vdots \\ x_r - \frac{\sum_{j=1}^r x_j w_{2j}}{\sum_{j=1}^r w_{2j}} \end{bmatrix} \end{aligned}$$

Para o cálculo da variância assintótica $\text{Var}(\hat{\beta})$ sob H_0, $\text{Var}_0(\hat{\beta})$, escrevemos os componentes respeitando as restrições testadas, $\text{Var}_0(\hat{\beta}) = \phi_0^{-1}(R_0^T W_0 R_0)^{-1}$. Nessas condições, a parte sistemática, os pesos, e os elementos de R podem ser escritos como:

$$\eta_{ij} = \alpha_i + \beta x_j \overset{H_0:\beta=0}{\rightarrow} \eta_i = \alpha_i^0$$

$$\omega_{ij} \overset{H_0:\beta=0}{\rightarrow} \omega_i^0$$

$$R = \left(x_j - \frac{\sum_{j=1}^r x_j w_{2j}}{\sum_{j=1}^r w_{2j}} \right)_{2r\times 1} \overset{H_0:\beta=0}{\rightarrow} R_0 = (x_j - \bar{x})_{2r\times 1}$$

$$W = \text{diag}\left\{\omega_{11}, \cdots, \omega_{1r}, \omega_{21}, \cdots, \omega_{2r}\right\} \overset{H_0:\beta=0}{\rightarrow} W_0 = \text{diag}\left\{\omega_1^0, \cdots, \omega_1^0, \omega_2^0, \cdots, \omega_2^0\right\}$$

E a variância assintótica é desenvolvida a partir dos seguintes componentes:

$$\begin{aligned} R_0^T W_0 &= \begin{bmatrix} x_1 - \bar{x} & \dots & x_r - \bar{x} & x_1 - \bar{x} & \dots & x_r - \bar{x} \end{bmatrix} \cdot \text{diag}\left\{\omega_1^0, \cdots, \omega_1^0, \omega_2^0, \cdots, \omega_2^0\right\} \\ &= \begin{bmatrix} \omega_1^0(x_1 - \bar{x}) & \dots & \omega_1^0(x_r - \bar{x}) & \omega_2^0(x_1 - \bar{x}) & \dots & \omega_2^0(x_r - \bar{x}) \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
R_0^T W_0 R_0 &= \begin{bmatrix} \omega_1^0(x_1-\bar{x}) & \dots & \omega_1^0(x_r-\bar{x}) & \omega_2^0(x_1-\bar{x}) & \dots & \omega_2^0(x_r-\bar{x}) \end{bmatrix} \cdot \begin{bmatrix} x_1-\bar{x} \\ \vdots \\ x_r-\bar{x} \\ x_1-\bar{x} \\ \vdots \\ x_r-\bar{x} \end{bmatrix} \\
&= \omega_1^0(x_1-\bar{x})^2 + \cdots + \omega_1^0(x_r-\bar{x})^2 + \omega_2^0(x_1-\bar{x})^2 + \cdots + \omega_2^0(x_r-\bar{x})^2 \\
&= \omega_1^0 \sum_{j=1}^{r} (x_j-\bar{x})^2 + \omega_2^0 \sum_{j=1}^{r} (x_j-\bar{x})^2 \\
&= (\omega_1^0 + \omega_2^0) \sum_{j=1}^{r} (x_j-\bar{x})^2
\end{aligned}
$$

Resultando nas seguintes expressões. A segunda mostra as estimativas necessárias para se estimar a variância assintótica de $\hat{\beta}$:

$$
\begin{aligned}
\mathrm{Var}_0(\hat{\beta}) &= \phi_0^{-1}(R_0^T W_0 R_0)^{-1} = \left(\phi_0(\omega_1^0 + \omega_2^0) \sum_{j=1}^{r} (x_j-\bar{x})^2 \right)^{-1} \\
\hat{\mathrm{Var}}_0(\hat{\beta}) &= \left(\hat{\phi}_0(\hat{\omega}_1^0 + \hat{\omega}_2^0) \sum_{j=1}^{r} (x_j-\bar{x})^2 \right)^{-1}
\end{aligned}
$$

2.2 Função score U_β

A função score é definida como $U_\beta(\theta) = \phi X_1^T W^{1/2} V^{-1/2}(\mathbf{y}-\mu)$. Porém podemos reescrevê-la em função do resíduo de Pearson $r_p = \phi^{1/2} V^{-1/2}(\mathbf{y}-\mu)$ e $U_\beta(\theta) = \phi^{1/2} X_1^T W^{1/2} r_p$

Iniciaremos o cálculo por $X_1^T W^{1/2}$:

$$
\begin{aligned}
X_1^T W^{1/2} &= \begin{bmatrix} x_1 & \dots & x_r & x_1 & \dots & x_r \end{bmatrix} \cdot \mathrm{diag}\left\{ \sqrt{\omega}_{11}, \cdots, \sqrt{\omega}_{1r}, \sqrt{\omega}_{21}, \cdots, \sqrt{\omega}_{2r} \right\} \\
&= \begin{bmatrix} x_1\sqrt{\omega}_{11} & \dots & x_r\sqrt{\omega}_{1r} & x_1\sqrt{\omega}_{21} & \dots & x_r\sqrt{\omega}_{2r} \end{bmatrix}
\end{aligned}
$$

Utilizando o resíduo de Pearson por obsevação, ficamos com:

$$
\begin{aligned}
X_1^T W^{1/2} r_p &= \begin{bmatrix} x_1\sqrt{\omega}_{11} & \dots & x_r\sqrt{\omega}_{1r} & x_1\sqrt{\omega}_{21} & \dots & x_r\sqrt{\omega}_{2r} \end{bmatrix} \cdot \begin{bmatrix} r_{11} \\ \vdots \\ r_{1r} \\ r_{21} \\ \vdots \\ r_{2r} \end{bmatrix} \\
&= \sum_{j=1}^{r} x_j \sqrt{\omega}_{1j} r_{1j} + \sum_{j=1}^{r} x_j \sqrt{\omega}_{2j} r_{2j}
\end{aligned}
$$

Onde $r_{ij} = \frac{\phi^{1/2}(y_{ij}-\mu_{ij})}{\sqrt{V_{ij}}}$. Assim o elemento da soma de cada um dos grupos, quando o resíduo é desmembrado:

$$x_j\sqrt{\omega}_{ij}r_{ij} = \frac{x_j\sqrt{\omega}_{ij}\phi^{1/2}(y_{ij}-\mu_{ij})}{\sqrt{V_{ij}}} \tag{1}$$

Nesse estágio do desenvolvimento da estatística de escore, precisamos assumir que a ligação utilizada no modelo de retas paralelas proposto é canônica. A razão disso é a seguinte: como os pesos são definidos como $\omega_{ij} = \left(\frac{d\mu_{ij}}{d\eta_{ij}}\right)^2 V_i^{-1}$. Se a ligação for canônica $\eta_{ij} = \theta_{ij}$, ou seja o componente linear sistemático coincide com o parâmetro canônico da família exponencial. Assim $\omega_{ij} = \left(\frac{d\mu_{ij}}{d\theta_{ij}}\right)^2 V_i^{-1}$ e, pela propriedade da família exponencial, $\frac{d\mu_{ij}}{d\theta_{ij}} = V_{ij}$. Por fim, os pesos sob a ligação canônica são expressos pela seguinte forma amigável $\omega_{ij} = V_{ij}^2 V_{ij}^{-1} = V_{ij}$.

Com isso, os pesos/variâncias dos componentes da soma na função escore de U_β descritos em (1) se cancelam quando a ligação do modelo é canônica:

$$\begin{aligned} x_j\sqrt{\omega}_{ij}r_{ij} &= \frac{x_j\sqrt{\omega}_{ij}\phi^{1/2}(y_{ij}-\mu_{ij})}{\sqrt{V_{ij}}} && (2)\\ &= \frac{x_j\sqrt{\omega}_{ij}\phi^{1/2}(y_{ij}-\mu_{ij})}{\sqrt{\omega_{ij}}} && (3)\\ &= \phi^{1/2}x_j(y_{ij}-\mu_{ij}) && (4) \end{aligned}$$

E a função escore possui então a seguinte forma:

$$\begin{aligned} U_\beta(\theta) &= \phi^{1/2}\left[\sum_{j=1}^{r}\phi^{1/2}x_j(y_{1j}-\mu_{1j}) + \sum_{j=1}^{r}\phi^{1/2}x_j(y_{2j}-\mu_{2j})\right] \\ &= \phi\left[\sum_{j=1}^{r}x_j(y_{1j}-\mu_{1j}) + \sum_{j=1}^{r}x_j(y_{2j}-\mu_{2j})\right] \end{aligned}$$

O próximo passo é entender como os componentes de $U_\beta(\theta)$ deveriam ser escritos sob H_0, ou seja, qual forma de $U_\beta(\theta^0)$. Como já expomos anteriormente, no cálculo da variância assintótica de $\hat{\beta}$:

$$\eta_{ij} = \alpha_i + \beta x_j \overset{H_0:\beta=0}{\rightarrow} \eta_i = \alpha_i^0$$

Adicionalmente, ainda sob H_0 temos

$$\mu_{ij} = g^{-1}(\eta_{ij}) \overset{H_0:\beta=0}{\rightarrow} \mu_i^0 = g^{-1}(\alpha_i^0) = \bar{y}_i$$

Portanto, o vetor de parâmetros θ sob a hipótese nula é:

$$\theta = \begin{bmatrix}\beta\\ \alpha_1\\ \alpha_2\\ \phi\end{bmatrix} \overset{H_0:\beta=0}{\rightarrow} \quad \theta^0 = \begin{bmatrix}0\\ \bar{y}_1\\ \bar{y}_2\\ \phi_0\end{bmatrix}$$

Culminando na seguinte função escore estimada:

$$U_\beta(\theta) = \phi\left[\sum_{j=1}^{r} x_j(y_{1j} - \mu_{1j}) + \sum_{j=1}^{r} x_j(y_{2j} - \mu_{2j})\right]$$
$$\downarrow$$
$$\hat{U}_\beta(\theta^0) = \hat{\phi}_0\left[\sum_{j=1}^{r} x_j(y_{1j} - \bar{y}_1) + \sum_{j=1}^{r} x_j(y_{2j} - \bar{y}_2)\right]$$

2.3 Estatística escore

Como estamos testando apenas um parâmetro, a estatística de escore resultante é a seguinte

$$\begin{aligned}\xi_{SR} &= \left[\hat{U}_\beta^0(\theta^0)\right]^T \hat{\text{Var}}_0(\hat{\beta})\ \hat{U}_\beta^0(\theta^0)\\ &= \left[\hat{U}_\beta^0(\theta^0)\right]^2 \hat{\text{Var}}_0(\hat{\beta})\end{aligned}$$

Juntando os componentes calculados nas seções anteriores, nomeadamente, a função escore e a variância assintótica, ambas para o parâmetro β e estimadas sob a hipótese nula. Vamos reproduzir ambas a seguir:

$$\hat{U}_\beta(\theta^0) = \hat{\phi}_0\left[\sum_{j=1}^{r} x_j(y_{1j} - \bar{y}_1) + \sum_{j=1}^{r} x_j(y_{2j} - \bar{y}_2)\right]$$
$$\hat{\text{Var}}_0(\hat{\beta}) = \left(\hat{\phi}_0(\hat{\omega}_1^0 + \hat{\omega}_2^0)\sum_{j=1}^{r}(x_j - \bar{x})^2\right)^{-1}$$

Finalmente, a expressão da algébrica estatística de escore é:

$$\begin{aligned}\xi_{SR} &= \frac{\left(\hat{\phi}_0\left[\sum_{j=1}^{r} x_j(y_{1j} - \bar{y}_1) + \sum_{j=1}^{r} x_j(y_{2j} - \bar{y}_2)\right]\right)^2}{\hat{\phi}_0(\hat{\omega}_1^0 + \hat{\omega}_2^0)\sum_{j=1}^{r}(x_j - \bar{x})^2}\\ &= \frac{\hat{\phi}_0\left[\sum_{j=1}^{r} x_j(y_{1j} - \bar{y}_1) + \sum_{j=1}^{r} x_j(y_{2j} - \bar{y}_2)\right]^2}{(\hat{\omega}_1^0 + \hat{\omega}_2^0)\sum_{j=1}^{r}(x_j - \bar{x})^2}\end{aligned}$$

Que de fato é a forma apresentada no enunciado. Note que, por usarmos a ligação canônica nos para se alcançar os resultados, a notação ω_i^0 pode ser substituída por V_i^0, uma vez que eles são iguais sob essas condições.

3 Exercício 3

A função densidade de uma variável $Y_i \overset{\text{iid}}{\sim} \text{NI}(\mu, \phi)$ tem a seguinte forma:

$$\frac{\phi^{1/2}}{\sqrt{2\pi y^3}}\exp\left[-\frac{\phi(y-\mu)^2}{2\mu^2 y}\right]$$

O enunciado pede que apresentemos a estatística da razão de verossimilhanças para testar a hipótese $H_0 : \phi = 1$ contra $H_1 : \phi \neq 1$. Assumo que o parâmetro μ é desconhecido, e por isso será necessário estimá-lo por máxima verossimilhança. O teste de razão de verossimilhança é dado pela seguinte expressão:

$$\xi_{RV} = 2(L(y_i;\hat{\mu},\hat{\phi}) - L(y_i;\hat{\mu},1))$$

Para tanto, utilizaremos a forma da função de densidade que além de provar que Y_i pertence à família exponencial, também facilita um pouco as computações:

$$\exp\left[\phi\left(-\frac{y}{2\mu^2}+\frac{1}{\mu}\right)-\frac{1}{2}\left[\log(2\pi y^3/\phi)+\frac{\phi}{y}\right]\right]$$

As computações a seguir a função de verosimilhança que usa função de densidade apresentada. Já partindo do seu logarítimo, temos

$$\begin{aligned}
L(y_i;\mu,\phi) &= \sum_{i=1}^{n}\left[\phi\left(-\frac{y_i}{2\mu^2}+\frac{1}{\mu}\right)-\frac{1}{2}\left[\log(2\pi y_i^3/\phi)+\frac{\phi}{y_i}\right]\right]\\
&= \sum_{i=1}^{n}\left[\phi\left(\frac{2\mu-y_i}{2\mu^2}\right)-\frac{1}{2}\left[\log(2\pi y_i^3/\phi)+\frac{\phi}{y_i}\right]\right]\\
&= \frac{\phi}{2\mu^2}\left(2n\mu-\sum_{i=1}^{n}y_i\right)-\frac{1}{2}\sum_{i=1}^{n}\log(2\pi y_i^3)+\frac{n\log(\phi)}{2}-\frac{\phi}{2}\sum_{i=1}^{n}\frac{1}{y_i}
\end{aligned}$$

Derivando com relação ao parâmetro μ, obtemos a função score do parâmetro

$$\begin{aligned}
\frac{\partial L}{\partial\mu} &= \frac{\phi}{2}\frac{-2}{\mu^3}\left(2n\mu-\sum_{i=1}^{n}y_i\right)+\frac{2n\phi}{2\mu^2}\\
&= -\frac{\phi}{\mu^3}\left(2n\mu-\sum_{i=1}^{n}y_i\right)+\frac{n\phi}{\mu^2}
\end{aligned}$$

Igualando a zero a expressão anterior, chegamos na estimativa de máxima verossimilhança para μ:

$$\begin{aligned}
&\frac{\partial L}{\partial\mu} = 0 && \Leftrightarrow\\
&-\frac{\phi}{\mu^3}\left(2n\mu-\sum_{i=1}^{n}y_i\right)+\frac{n\phi}{\mu^2} = 0 && \Leftrightarrow\\
&-\frac{(2n\mu-\sum_{i=1}^{n}y_i)}{\mu^3}+\frac{n\mu}{\mu^2\mu} = 0 && \Leftrightarrow\\
&-2n\mu+\sum_{i=1}^{n}y_i+n\mu = 0 && \Leftrightarrow\\
&\sum_{i=1}^{n}y_i-n\mu = 0 && \Leftrightarrow\\
&\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}y_i = \bar{y}
\end{aligned}$$

O próximo passo é calcular uma expressão para a função de verossimilhança quando as estimativas estão sob o modelo proposto (em H_0 e H_1). Calcularemos primeiro $L(y_i;\hat{\mu},\hat{\phi})$:

$$L(y_i;\hat{\mu},\hat{\phi}) = \sum_{i=1}^{n}\left[\hat{\phi}\left(-\frac{y_i}{2\hat{\mu}^2}+\frac{1}{\hat{\mu}}\right)-\frac{1}{2}\left[\log(2\pi y_i^3/\hat{\phi})+\frac{\hat{\phi}}{y_i}\right]\right] \tag{1}$$

Antes de reduzirmos a expressão de $L(y_i;\hat{\mu},\hat{\phi})$, iremos reescrever o fator que multiplica o primeiro ϕ de maneira conveniente. Note que omitimos o subscrito i e a notação que indica estimativas para mantermos clareza nas manipulações:

$$\begin{aligned}
-\frac{y}{2\mu^2}+\frac{1}{\mu} &= \frac{-y+2\mu}{2\mu^2}\frac{(-y)}{(-y)} \\
&= -\frac{y^2-2\mu y+\mu^2-\mu^2}{2\mu^2 y} \\
&= (-1)\left[\frac{(y-\mu)^2}{2\mu^2 y}-\frac{\mu^2}{2\mu^2 y}\right] \\
&= -\frac{(y-\mu)^2}{2\mu^2 y}+\frac{1}{2y}
\end{aligned}$$

Desta maneira podemos escrever (1) da seguinte forma:

$$L(y_i;\hat{\mu},\hat{\phi}) = \sum_{i=1}^{n}\left[\hat{\phi}\left(-\frac{(y_i-\hat{\mu})^2}{2\hat{\mu}^2 y_i}+\frac{1}{2y_i}\right)-\frac{1}{2}\left[\log(2\pi y_i^3/\hat{\phi})+\frac{\hat{\phi}}{y_i}\right]\right]$$

Desenvolvendo os somatórios temos:

$$\begin{aligned}
L(y_i;\hat{\mu},\hat{\phi}) &= \sum_{i=1}^{n}\left[\hat{\phi}\left(-\frac{(y_i-\hat{\mu})^2}{2\hat{\mu}^2 y_i}+\frac{1}{2y_i}\right)-\frac{1}{2}\left[\log(2\pi y_i^3/\hat{\phi})+\frac{\hat{\phi}}{y_i}\right]\right] \\
&= -\frac{\hat{\phi}}{2}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu})^2}{\hat{\mu}^2 y_i}+\frac{\hat{\phi}}{2}\sum\frac{1}{y_i}-\frac{1}{2}\sum_{i=1}^{n}\log(2\pi y_i^3)+\frac{n\log(\hat{\phi})}{2}-\frac{\hat{\phi}}{2}\sum_{i=1}^{n}\frac{1}{y_i} \qquad (2) \\
&= \frac{n\log(\hat{\phi})}{2}-\frac{\hat{\phi}}{2}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu})^2}{\hat{\mu}^2 y_i}-\frac{1}{2}\sum_{i=1}^{n}\log(2\pi y_i^3)
\end{aligned}$$

Note que, para o caso de uma variável aleatória distribuída por uma Normal Inversa, a função desvio é dada por

$$D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n}\frac{(y_i-\hat{\mu})^2}{\hat{\mu}^2 y_i}$$

Assim, podemos reescrever (2)

$$\begin{aligned}
L(y_i;\hat{\mu},\hat{\phi}) &= \frac{n\log(\hat{\phi})}{2}-\frac{\hat{\phi}}{2}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu})^2}{\hat{\mu}^2 y_i}-\frac{1}{2}\sum_{i=1}^{n}\log(2\pi y_i^3) \\
&= \frac{n\log(\hat{\phi})}{2}-\frac{\hat{\phi}}{2}D(\boldsymbol{y},\hat{\boldsymbol{\mu}})-\frac{1}{2}\sum_{i=1}^{n}\log(2\pi y_i^3)
\end{aligned}$$

Assim, podemos partir para o cálculo de ξ_{RV} a partir de sua definição e substituindo os casos particulares de H_0 e H_1:

$$\begin{aligned}
\xi_{RV} &= 2(L(y_i;\hat{\mu},\hat{\phi}) - L(y_i;\hat{\mu},1)) \\
&= 2\left[\frac{n\log(\hat{\phi})}{2} - \frac{\hat{\phi}}{2}D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) - \frac{1}{2}\sum_{i=1}^{n}\log(2\pi y_i^3) - \left(\frac{n\log(1)}{2} - \frac{1}{2}D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) - \frac{1}{2}\sum_{i=1}^{n}\log(2\pi y_i^3)\right)\right] \\
&= n\log(\hat{\phi}) - \hat{\phi}D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) + D(\boldsymbol{y},\hat{\boldsymbol{\mu}})
\end{aligned} \tag{3}$$

Porém, nos é fornecido no enunciado que $\hat{\phi} = \frac{n}{D(\boldsymbol{y},\hat{\boldsymbol{\mu}})}$, levando ao fato que $D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) = n\hat{\phi}^{-1}$. Munidos disso, podemos desenvolver (3) ainda mais:

$$\begin{aligned}
\xi_{RV} &= n\log(\hat{\phi}) - \hat{\phi}D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) + D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) \\
&= n\log(\hat{\phi}) - \hat{\phi}\hat{\phi}^{-1}n + n\hat{\phi}^{-1} \\
&= n\log(\hat{\phi}) + n(\hat{\phi}^{-1} - 1)
\end{aligned}$$

Que é o resultado requisitado pelo enunciado.

4 Exercício 4

O modelo proposto no enunciado consiste em $Y_{ij} \overset{ind}{\sim} Poisson(\mu_{ij})$ para $i, j = 1, 2$. O componente sistmático do modelo é

$$\log(\mu_{ij}) = \alpha + \beta_i + \gamma_j + \delta_{ij}$$

Onde, para permitir a identificabilidade do modelo, temos $\beta_1 = \gamma_1 = 0$ e $\delta_{11} = \delta_{12} = \delta_{21} = 0$ e $\delta_{22} = \delta$. Onde β_1 e β_2 correspondem aos efeitos do fator A e γ_1 e γ_2 correspondem aos efeitos do fator B.

4.1 Modelo Multinomial

Com relação aos fatores A e B, e seus respectivos níveis, A_1 e A_2 (correspondentes à $i = 1, 2$) e B_1 e B_2 (correspondentes à $j = 1, 2$), podemos pensar no modelo proposto para Y_{ij} como a tabela de contingência 2x2 a seguir.

		A	
		A_1	A_2
B	B_1	Y_{11}	Y_{21}
	B_2	Y_{12}	Y_{22}

E as partes sistemáticas associadas à cada uma dessas combinações de fatores, quando substituímos as restrições do impostas pela identificabilidade do modelo, ficam dadas por

$$\log(\mu_{11}) = \alpha$$

$$\log(\mu_{12}) = \alpha + \gamma_2$$

$$\log(\mu_{21}) = \alpha + \beta_2$$

$$\log(\mu_{22}) = \alpha + \beta_2 + \gamma_2 + \delta$$

Com essa especificação do modelo Poisson em termos dos fatores A e B, se condicionarmos o problema em $n = \sum_{i=1}^{2} \sum_{j=1}^{2} Y_{ij}$, temos que

$$\boldsymbol{Y} \sim \text{Multinomial}(n, \boldsymbol{\pi})$$

Com $\boldsymbol{Y} = (Y_{11}, Y_{12}, Y_{21}, Y_{22})^T$ e $\boldsymbol{\pi} = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$. Onde π_{ij} é, quando em função das médias de Y_{ij}, dado por

$$\pi_{ij} = \frac{\mu_{ij}}{\sum_{i=1}^{2} \sum_{j=1}^{2} \mu_{ij}}$$

E em seguida, mostramos a expressão de π_{ij} em termos das partes sistemática que englobam os efeitos dos fatores A e B:

$$\pi_{ij} = \frac{e^{\alpha + \beta_i + \gamma_j + \delta_{ij}}}{\sum_{i=1}^{2} \sum_{j=1}^{2} e^{\alpha + \beta_i + \gamma_j + \delta_{ij}}} \tag{1}$$

4.2 Independência probabilística entre os fatores

Partindo da tabela de contingência apresentada na seção anterior e condicionados à $n = \sum_{i=1}^{2}\sum_{j=1}^{2} Y_{ij}$ podemos construir a seguinte tabela de contingência relativa entre os fatores A e B

		A		Marginal
		A_1	A_2	B
B	B_1	π_{11}	π_{21}	π_{+1}
	B_2	π_{12}	π_{22}	π_{+2}
Marginal	A	π_{1+}	π_{2+}	1

Adicionamos margens à tabela, que correspondem às distribuições marginais dos fatores. Além disso $\pi_{i+} = \sum_{j=1}^{2} \pi_{ij}$ e $\pi_{+j} = \sum_{1=1}^{2} \pi_{ij}$.

Pede se para mostrar que a indepêndencia probabilistica entre os fatores A e B ocorre quando $\delta_{22} = \delta = 0$. Mostrar a independência probabilística nesse contexto implica em mostrar que

$$\pi_{ij} = \pi_{i+} \cdot \pi_{+j} \quad \forall\ i, j = 1, 2$$

Ou, em outras palavras, equivale à mostrar com o produtos das distribuições maginais coincide com a probabilidade conjunta para todos as combinações de níveis entre A e B. Mostraremos cada um dos pares $i, j = 1, 2$, porém antes precisamos levar as expressões de π_{ij}, definidas em (1), para versões mais amigáveis. Primeiro vamos nos livrar do intercepto dos modelos poisson, uma vez que a equivalência entre os modelos multinomial e poisson só se da às variáveis dentro do modelo, e as equivalências entre interceptos não é relevante para o que queremos mostrar:

$$\begin{aligned}
\pi_{ij} &= \frac{e^{\alpha+\beta_i+\gamma_j+\delta_{ij}}}{\sum_{i=1}^{2}\sum_{j=1}^{2} e^{\alpha+\beta_i+\gamma_j+\delta_{ij}}} \\
&= \frac{e^{\alpha} e^{\beta_i+\gamma_j+\delta_{ij}}}{e^{\alpha}\sum_{i=1}^{2}\sum_{j=1}^{2} e^{\beta_i+\gamma_j+\delta_{ij}}} \\
&= \frac{e^{\beta_i+\gamma_j+\delta_{ij}}}{\sum_{i=1}^{2}\sum_{j=1}^{2} e^{\beta_i+\gamma_j+\delta_{ij}}}
\end{aligned} \tag{2}$$

Em seguida, manipulamos o denominador de (2), uma vez que ele é comum à todos os π_{ij}

$$\sum_{i=1}^{2}\sum_{j=1}^{2} e^{\beta_i+\gamma_j+\delta_{ij}} = 1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta}$$

E então podemos chegar às espressões de cada π_{ij}:

$$\pi_{11} = \frac{1}{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta}}$$

$$\pi_{12} = \frac{e^{\beta_2}}{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta}}$$

$$\pi_{21} = \frac{e^{\gamma_2}}{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta}}$$

$$\pi_{22} = \frac{e^{\beta_2+\gamma_2+\delta}}{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta}}$$

4.2.1 $\pi_{11} = \pi_{1+} \cdot \pi_{+1}$

$$\begin{aligned}
\pi_{1+} \cdot \pi_{+1} &= (\pi_{11} + \pi_{12}) \cdot (\pi_{11} + \pi_{21}) \\
&= \frac{(1 + e^{\gamma_2})(1 + e^{\beta_2})}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta})^2} \\
&= \frac{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2}}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta})^2}
\end{aligned}$$

Se $\delta = 0$:

$$\begin{aligned}
\pi_{1+} \cdot \pi_{+1} &= \frac{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2}}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2})^2} \\
&= \frac{1}{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2}} \\
&= \pi_{11}
\end{aligned}$$

4.2.2 $\pi_{12} = \pi_{1+} \cdot \pi_{+2}$

$$\begin{aligned}
\pi_{1+} \cdot \pi_{+2} &= (\pi_{11} + \pi_{12}) \cdot (\pi_{12} + \pi_{22}) \\
&= \frac{(1 + e^{\gamma_2})(e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta})}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta})^2} \\
&= \frac{e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta} + e^{\gamma_2}e^{\gamma_2} + e^{\gamma_2}e^{\beta_2+\gamma_2+\delta}}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta})^2} \\
&= \frac{e^{\gamma_2}\left(1 + e^{\gamma_2} + e^{\beta_2+\delta} + e^{\beta_2+\gamma_2+\delta}\right)}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta})^2}
\end{aligned}$$

Tomando $\delta = 0$:

$$\begin{aligned}
\pi_{1+} \cdot \pi_{+2} &= \frac{e^{\gamma_2}\left(1 + e^{\gamma_2} + e^{\beta_2} + e^{\beta_2+\gamma_2}\right)}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2})^2} \\
&= \frac{e^{\gamma_2}}{1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2}} \\
&= \pi_{12}
\end{aligned}$$

4.2.3 $\pi_{21} = \pi_{2+} \cdot \pi_{+1}$

$$\begin{aligned}
\pi_{2+} \cdot \pi_{+1} &= (\pi_{21} + \pi_{22}) \cdot (\pi_{11} + \pi_{21}) \\
&= \frac{(e^{\beta_2} + e^{\beta_2+\gamma_2+\delta})(1 + e^{\beta_2})}{(1 + e^{\beta_2} + e^{\gamma_2} + e^{\beta_2+\gamma_2+\delta})^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{e^{\beta_2} + e^{\beta_2}e^{\beta_2} + e^{\beta_2+\gamma_2+\delta} + e^{\beta_2}e^{\beta_2+\gamma_2+\delta}}{(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2+\delta})^2} \\
&= \frac{e^{\beta_2}\left(1+e^{\gamma_2}+e^{\gamma_2+\delta}+e^{\beta_2+\gamma_2+\delta}\right)}{(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2+\delta})^2}
\end{aligned}
$$

Com $\delta = 0$:

$$
\begin{aligned}
\pi_{2+} \cdot \pi_{+1} &= \frac{e^{\beta_2}\left(1+e^{\gamma_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2}\right)}{(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2})^2} \\
&= \frac{e^{\beta_2}}{1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2}} \\
&= \pi_{21}
\end{aligned}
$$

4.2.4 $\pi_{22} = \pi_{2+} \cdot \pi_{+2}$

$$
\begin{aligned}
\pi_{2+} \cdot \pi_{+2} &= (\pi_{21}+\pi_{22}) \cdot (\pi_{12}+\pi_{22}) \\
&= \frac{(e^{\beta_2}+e^{\beta_2+\gamma_2+\delta})(e^{\gamma_2}+e^{\beta_2+\gamma_2+\delta})}{(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2+\delta})^2} \\
&= \frac{e^{\beta_2}e^{\gamma_2} + e^{\beta_2}e^{\beta_2+\gamma_2+\delta} + e^{\gamma_2}e^{\beta_2+\gamma_2+\delta} + e^{\beta_2}e^{\beta_2+\gamma_2+\delta} + e^{2(\beta_2+\gamma_2+\delta)}}{(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2+\delta})^2} \\
&= \frac{e^{\beta_2+\gamma_2+\delta}\left(e^{-\delta}+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2+\delta}\right)}{(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2+\delta})^2}
\end{aligned}
$$

Impondo $\delta = 0$:

$$
\begin{aligned}
\pi_{2+} \cdot \pi_{+2} &= \frac{e^{\beta_2+\gamma_2}\left(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2}\right)}{(1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2})^2} \\
&= \frac{e^{\beta_2+\gamma_2}}{1+e^{\beta_2}+e^{\gamma_2}+e^{\beta_2+\gamma_2}} \\
&= \pi_{22}
\end{aligned}
$$

Provando assim que se $\delta = 0$, $\pi_{ij} = \pi_{i+} \cdot \pi_{+j} \quad \forall\ i,j = 1,2$ e os fatores A e B são probablisticamente independentes.

5 Exercício 5

O modelo descrito pelo enunciado consiste em $Y_{ij}|z_i \stackrel{ind}{\sim} \text{FE}(\mu, \phi_i)$ com $i = 1, 2$ e $j = 1, ..., m$. A parte sistemática é:

$$\log(\phi_1) = \lambda_1 = \alpha - \beta$$

$$\log(\phi_2) = \lambda_2 = \alpha + \beta$$

5.1 Função Escore U_β

Desejamos encontrar uma forma algébrica fechada para a estatística se escore que testa as hipóteses $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$. A estatística de score para o nosso conjunto de hipóteses tem a seguinte forma:

$$\xi_{SR} = \left(\hat{U}_\beta(\hat{\beta}^0)\right)^2 \hat{\text{Var}}_0(\hat{\beta})$$

Iremos calcular cada um desses componentes de maneira genérica e em seguida avaliá-los sob as restrições das hipóteses testadas. Como se trata de hipóteses com relação ao parâmetro de precisão de uma variável pertencente à família exponencial, vamos trabalha com a verossimilhança restrita da variável $T_{ij} = t_{ij}$, onde $t_{ij} = y_{ij}\theta - b(\theta) + u(y_{ij})$, onde as estimativas do parâmetro de localização θ e suas funções partirão de um modelo apenas com intercepto.

Assim, a derivação da função escore U_β será com base na função de verossimilhança da variável transformada T_{ij}.

$$L(t_{ij}; \phi_i) = \sum_{i=1}^{2}\sum_{j=1}^{m} \phi_i t_{ij} + \text{d}(\phi_i) + \text{u}(y_{ij})$$

$$\begin{aligned}
U_\beta &= \frac{\partial L}{\partial \beta} \\
&= \frac{\partial}{\partial \beta}\sum_{i=1}^{2}\sum_{j=1}^{m} \phi_i t_{ij} + \text{d}(\phi_i) + \text{u}(y_{ij}) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{m} \frac{\partial}{\partial \beta}\left[\phi_i t_{ij} + d(\phi_i) + u(y_{ij})\right] \\
&= \sum_{i=1}^{2}\sum_{j=1}^{m} \left[\frac{d\phi_i}{d\lambda_i}\frac{\partial \lambda_i}{\partial \beta} t_{ij} + \text{d}'(\phi_i)\frac{d\phi_i}{d\lambda_i}\frac{\partial \lambda_i}{\partial \beta}\right] \qquad (1)
\end{aligned}$$

Onde das derivadas em (1) são:

$$\frac{d\phi_i}{d\lambda_i} = \left(\frac{d\lambda_i}{d\phi_i}\right)^{-1} = \left(\frac{d}{d\phi_i}\log(\phi_i)\right)^{-1} = \left(\frac{1}{\phi_i}\right)^{-1} = \phi_i$$

$$\lambda_1 = \alpha - \beta \quad \rightarrow \quad \frac{\partial \lambda_1}{\partial \beta} = -1 \lambda_2 = \alpha + \beta \quad \rightarrow \quad \frac{\partial \lambda_2}{\partial \beta} = \quad 1$$

Onde denotaremos que, simbolicamente, $\frac{\partial \lambda_i}{\partial \beta} = (-1)^i$, por conveniência. Então, a função escore do parâmetro β é

$$
\begin{aligned}
U_\beta &= \sum_{i=1}^{2}\sum_{j=1}^{m}\left[\frac{d\phi_i}{d\lambda_i}\frac{\partial\lambda_i}{\partial\beta}t_{ij} + \mathrm{d}'(\phi_i)\frac{d\phi_i}{d\lambda_i}\frac{\partial\lambda_i}{\partial\beta}\right] \\
&= \sum_{i=1}^{2}\sum_{j=1}^{m}\left[\phi_i(-1)^i t_{ij} + \phi_i(-1)^i\mathrm{d}'(\phi_i)\right] \\
&= \sum_{i=1}^{2}\sum_{j=1}^{m}\phi_i(-1)^i\left[t_{ij} + \mathrm{d}'(\phi_i)\right]
\end{aligned}
$$

5.2 Variância assintótica de $\hat{\beta}$

Em seguida, calculamos o componente da hessiana de L referente à β

$$
\begin{aligned}
\frac{\partial^2 L}{\partial\beta^2} &= \sum_{i=1}^{2}\sum_{j=1}^{m}\frac{\partial}{\partial\beta}\phi_i(-1)^i\left[t_{ij} + \mathrm{d}'(\phi_i)\right] \\
&= \sum_{i=1}^{2}\sum_{j=1}^{m}(-1)^i\left[(-1)^i\phi_i\left[t_{ij} + \mathrm{d}'(\phi_i)\right]\right] + \phi_i\mathrm{d}''(\phi_i)\phi_i(-1)^i \\
&= \sum_{i=1}^{2}\sum_{j=1}^{m}(-1)^{2i}\left[\phi_i^2\mathrm{d}''(\phi_i) + \phi_i\left[t_{ij} + \mathrm{d}'(\phi_i)\right]\right] \\
&= \sum_{i=1}^{2}\sum_{j=1}^{m}\left[\phi_i^2\mathrm{d}''(\phi_i) + \phi_i\left[t_{ij} + \mathrm{d}'(\phi_i)\right]\right] \qquad (2)
\end{aligned}
$$

Para se chegar ao componente da informação de Fisher de β, $\mathrm{K}_{\beta\beta}$, calculamos o valor esperado do negativo de (2):

$$
\begin{aligned}
\mathrm{K}_{\beta\beta} &= \mathrm{E}\left(-\frac{\partial^2 L}{\partial\beta^2}\right) \\
&= -\sum_{i=1}^{2}\sum_{j=1}^{m}\phi_i^2\mathrm{d}''(\phi_i)
\end{aligned}
$$

Em seguida, vamos desenvolver os somatórios de $\mathrm{K}_{\beta\beta}$:

$$
\begin{aligned}
\mathrm{K}_{\beta\beta} &= -\sum_{i=1}^{2}\sum_{j=1}^{m}\phi_i^2\mathrm{d}''(\phi_i) \\
&= -\sum_{i=1}^{2}m\phi_i^2\mathrm{d}''(\phi_i) \\
&= -\left[m\phi_1^2\mathrm{d}''(\phi_1) + m\phi_2^2\mathrm{d}''(\phi_2)\right] \\
&= -m\left[\phi_1^2\mathrm{d}''(\phi_1) + \phi_2^2\mathrm{d}''(\phi_2)\right] \qquad (3)
\end{aligned}
$$

Para chegar ao fato de que $\mathrm{Var}(\hat{\beta}) = \mathrm{K}_{\beta\beta}^{-1}$ sem precisar calcular a matriz de informação de Fisher por completo, basta notar que a matriz modelo Z que modela $\log(\phi_i)$ é:

$$
Z = \begin{bmatrix} 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}
$$

E além disso $Z^T Z = 2m \cdot I$, ou seja uma matriz diagonal. Nesses casos a matriz de informação de Fisher também é diagonal. Assim para se obter variância assintótica do parâmetro de interesse basta obter o inverso do seu componente correspondente na matriz de Fisher, sem precisar calcular a matriz inteira e inverte-la em seguida. Nesse sentido, temos o seguinte resultado, partindo de (3)

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}) &= \left[-\sum_{i=1}^{2}\sum_{j=1}^{m} \phi_i^2 \mathrm{d}''(\phi_i) \right]^{-1} \\
&= \left[-m \left(\phi_1^2 \mathrm{d}''(\phi_1) + \phi_2^2 \mathrm{d}''(\phi_2) \right) \right]^{-1}
\end{aligned}
$$

5.3 Estatística de escore

Primeiro vamos desenvolver os somatórios da função escore da mesma maneira que fizemos em (3):

$$
\begin{aligned}
U_\beta &= \sum_{i=1}^{2}\sum_{j=1}^{m} \phi_i (-1)^i \left[t_{ij} + \mathrm{d}'(\phi_i) \right] \\
&= \sum_{j=1}^{m} \phi_1(-1) t_{1j} - \phi_1 \mathrm{d}'(\phi_1) + \sum_{j=1}^{m} \phi_2 t_{2j} + \phi_2 \mathrm{d}'(\phi_2) \\
&= \sum_{j=1}^{m} \phi_2 t_{2j} + \phi_2 \mathrm{d}'(\phi_2) - \sum_{j=1}^{m} \phi_1 t_{1j} + \phi_1 \mathrm{d}'(\phi_1) \\
&= \phi_2 \sum_{j=1}^{m} [t_{2j}] + m\phi_2 \mathrm{d}'(\phi_2) - \phi_1 \sum_{j=1}^{m} [t_{1j}] - m\phi_1 \mathrm{d}'(\phi_1) \\
&= \phi_2 m \bar{t_2} + m\phi_2 \mathrm{d}'(\phi_2) - \phi_1 m \bar{t_1} - m\phi_1 \mathrm{d}'(\phi_1)
\end{aligned}
$$

Sob $H_0 : \beta = 0$, temos que as estimativas $\phi_1 = \phi_2 = \phi_0$. Portanto, a função escore U_β é escrita da seguinte forma

$$
\begin{aligned}
U_\beta &= \phi_2 m \bar{t_2} + m\phi_2 \mathrm{d}'(\phi_2) - \phi_1 m \bar{t_1} - m\phi_1 \mathrm{d}'(\phi_1) \\
&\overset{H_0}{=} \phi_0 m \bar{t_2} + m\phi_0 \mathrm{d}'(\phi_0) - \phi_0 m \bar{t_1} - m\phi_1 \mathrm{d}'(\phi_0) \\
&\overset{H_0}{=} \phi_0 m \bar{t_2} - \phi_0 m \bar{t_1} \\
&\overset{H_0}{=} \phi_0 m \left(\bar{t_2} - \bar{t_1} \right)
\end{aligned}
$$

E a variância assintótica, também sob a hipótese nula

$$
\mathrm{Var}(\hat{\beta}) = \left[-m \left(\phi_1^2 \mathrm{d}''(\phi_1) + \phi_2^2 \mathrm{d}''(\phi_2) \right) \right]^{-1}
$$

$$
\begin{aligned}
&\overset{H_0}{=} \left[-m\left(\phi_0^2\mathrm{d}''(\phi_0)+\phi_0^2\mathrm{d}''(\phi_0)\right)\right]^{-1}\\
&\overset{H_0}{=} \left[-m\phi_0^2\left(2\mathrm{d}''(\phi_0)\right)\right]^{-1}
\end{aligned}
$$

Assim, munido dos fatos $U_\beta(\beta^0)=\phi_0 m\left(\bar{t_2}-\bar{t_1}\right)$ e $\mathrm{Var}_0(\hat{\beta})=\left[-m\phi_0^2\left(2\mathrm{d}''(\phi_0)\right)\right]^{-1}$, podemos montar a estatística de escore ξ_{SR}:

$$
\begin{aligned}
\xi_{SR} &= \left[\hat{U_\beta}(\beta^0)\right]^2\cdot\left[\mathrm{Var}_0(\hat{\beta})\right]\\
&= \frac{\left[\hat{\phi}_0 m\left(\hat{\bar{t_2}}-\hat{\bar{t_1}}\right)\right]^2}{-m\hat{\phi}_0^2\left(2\mathrm{d}''(\hat{\phi}_0)\right)}\\
&= -\frac{\hat{\phi}_0^2 m^2\left(\hat{\bar{t_2}}-\hat{\bar{t_1}}\right)^2}{m\hat{\phi}_0^2\left(2\mathrm{d}''(\hat{\phi}_0)\right)}\\
&= -\frac{m\left(\hat{\bar{t_2}}-\hat{\bar{t_1}}\right)^2}{2\mathrm{d}''(\hat{\phi}_0)}
\end{aligned}
$$

Onde $\hat{\bar{t_i}}=\frac{1}{m}\sum_{j=1}^{m}\hat{t_{ij}}$ e $\hat{t}_{ij}=y_{ij}\hat{\theta}-b(\hat{\theta})+u(y_{ij})$ e $\hat{\theta}$ foi estimado anteriormente e tratado como fixo nos cálculos que culminam na variável $\hat{t}_{ij}$.

Como T_{ij} é membro da família exponencial, os mesmos resultados assintóticos se aplicam. Nesses moldes $\xi_{SR}\to\chi^2_{(1)}$, pois apenas um parâmetro é testado.

5.4 Caso normal

Particularizando para o caso normal, os componentes presentes da estatística de escore são:

$$
\hat{t}_{ij}=y_{ij}\hat{\mu}-\frac{1}{2}\left(\hat{\mu}^2+y_{ij}^2\right)
$$

$$
\mathrm{d}''(\hat{\phi}_0)=\left(-2\hat{\phi}_0^2\right)^{-1}
$$

Partindo de (5.4), podemos ver que os componentes de média $\hat{\bar{t_i}}$ estão diretamente relacionados com a função desvio

$$
\begin{aligned}
\hat{\bar{t_i}} &= \frac{1}{m}\sum_{j=1}^{m}y_{ij}\hat{\mu}-\frac{1}{2}\left(\hat{\mu}^2+y_{ij}^2\right)\\
&= \frac{1}{m}\sum_{j=1}^{m}\frac{2}{2}y_{ij}\hat{\mu}-\frac{\hat{\mu}^2}{2}-\frac{y_{ij}^2}{2}\\
&= \frac{1}{m}\sum_{j=1}^{m}\frac{2y_{ij}\hat{\mu}-\mu^2-y_{ij}^2}{2}\\
&= -\frac{1}{m}\sum_{j=1}^{m}\frac{\mu^2+y_{ij}^2-2y_{ij}\hat{\mu}}{2}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2m} \sum_{j=1}^{m} \mu^2 + y_{ij}^2 - 2y_{ij}\hat{\mu} \\
&= -\frac{1}{2m} \sum_{j=1}^{m} (y_{ij} - \hat{\mu})^2 \\
&= -\frac{1}{2m} \mathrm{D}(\boldsymbol{y_i}, \boldsymbol{\hat{\mu}})
\end{aligned}
$$

Assim, ξ_{SR} pode ser escrita como uma diferença de desvios entre os grupo $i = 1, 2$:

$$
\begin{aligned}
\xi_{SR} &= -\frac{m\left(\hat{\hat{t}}_2 - \hat{\hat{t}}_1\right)^2}{2\mathrm{d}''(\hat{\phi}_0)} \\
&= -\frac{m\left(\frac{1}{2m}\mathrm{D}(\boldsymbol{y_1}, \boldsymbol{\hat{\mu}}) - \frac{1}{2m}\mathrm{D}(\boldsymbol{y_2}, \boldsymbol{\hat{\mu}})\right)}{\left(-4\hat{\phi}_0^2\right)^{-1}} \\
&= \frac{1}{2}\left(\mathrm{D}(\boldsymbol{y_1}, \boldsymbol{\hat{\mu}}) - \mathrm{D}(\boldsymbol{y_2}, \boldsymbol{\hat{\mu}})\right) 4\hat{\phi}_0^2 \\
&= 2\hat{\phi}_0^2\left(\mathrm{D}(\boldsymbol{y_1}, \boldsymbol{\hat{\mu}}) - \mathrm{D}(\boldsymbol{y_2}, \boldsymbol{\hat{\mu}})\right)
\end{aligned}
$$

Uma outra forma para possível ξ_{SR} é notando que $\frac{1}{m}\mathrm{D}(\boldsymbol{y_i}, \boldsymbol{\hat{\mu}}) = \left(\hat{\phi}_i\right)^{-1}$ e portanto:

$$
\begin{aligned}
\xi_{SR} &= -\frac{m\left(\hat{\hat{t}}_2 - \hat{\hat{t}}_1\right)^2}{2\mathrm{d}''(\hat{\phi}_0)} \\
&= -\frac{m}{2}\left(\frac{1}{2\hat{\phi}_1} - \frac{1}{2\hat{\phi}_2}\right) \cdot \left(-2\hat{\phi}_0^2\right) \\
&= \frac{\hat{\phi}_0^2 m}{2}\left(\frac{1}{\hat{\phi}_1} - \frac{1}{\hat{\phi}_2}\right)
\end{aligned}
$$