Planejamento de Experimentos
Prof.ª Dr.ª Sílvia Maria de Freitas
Aluno: Antônio Arthur Silva de Lima
Matrícula: 508492

1. Sabendo que $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$, onde $\epsilon_{ij} \sim N(\mu; \sigma^2)$ e considerando um DIC, encontre os estimadores de máxima verossimilhança de μ e τ_i.

Primeiro, temos que $\epsilon_{ij} \sim N(0; \sigma^2) \therefore M_{\epsilon_{ij}}(t) = e^{\frac{\sigma^2 t^2}{2}}$. Encontramos então a função geradora de momentos de Y_{ij}, demonstrando sua normalidade, i.e. $Y_{ij} \sim N(\mu + \tau_i; \sigma^2)$.

$$M_{Y_{ij}}(t) = e^{tY_{ij}} = E[e^{t(\mu+\tau_i+\epsilon_{ij})}] = E[e^{t(\mu+\tau_i)} \cdot e^{t\epsilon_{ij}}] = e^{t(\mu+\tau_i)}E[e^{t\epsilon_{ij}}] = e^{t(\mu+\tau_i)}M_{\epsilon_{ij}}(t) = e^{t(\mu+\tau_i)}e^{\frac{\sigma^2 t^2}{2}}$$

$$= e^{t(\mu+\tau_i)+\frac{\sigma^2 t^2}{2}}$$

Assim, temos que

$$f\left(y_{ij}; \mu; \tau_i; \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_{ij} - \mu - \tau_i)^2}{2\sigma^2}\right\}$$

Agora, conhecendo a distribuição da variável resposta, utilizamos o método da máxima verossimilhança para encontrar os estimadores:

$$L(\mu; \tau_i) = \prod_{i=1}^{a}\prod_{j=1}^{n} f\left(y_{ij}; \mu; \tau_i\right) = \prod_{i=1}^{a}\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_{ij} - \mu - \tau_i)^2}{2\sigma^2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{na} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \mu - \tau_i)^2\right\}$$

$$l(\mu; \tau_i) = na \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \mu - \tau_i)^2$$

- Estimação de μ

$$\frac{\partial l(\mu; \tau_i)}{\partial \mu} = -\frac{1}{2\sigma^2}(-2)\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \mu - \tau_i)$$

$$\frac{\partial l(\mu; \tau_i)}{\partial \mu} = 0 \rightarrow -\frac{1}{2\hat{\sigma}^2}(-2)\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0 \therefore$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0 \therefore$$

$$N\bar{y}_{..} - n\sum_{i=1}^{a}\hat{\tau}_i = N\hat{\mu} \therefore$$

$$\hat{\mu} = \bar{y}_{..}$$

- Estimação de τ_i, considerando um "i "(fator) qualquer

$$\frac{\partial l(\mu;\tau_i)}{\partial \tau_i} = -\frac{1}{2\sigma^2}(-2)\sum_{j=1}^{n}(y_{ij}-\mu-\tau_i)$$

$$\frac{\partial l(\mu;\tau_i)}{\partial \tau_i} = 0 \to -\frac{1}{2\hat{\sigma}^2}(-2)\sum_{j=1}^{n}(y_{ij}-\hat{\mu}-\hat{\tau}_i) = 0 \therefore$$

$$\sum_{j=1}^{n}(y_{ij}-\hat{\mu}-\hat{\tau}_i) = 0 \therefore$$

$$n\bar{y}_{i.} - n\hat{\mu} = n\hat{\tau}_i \therefore$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

Desta forma, temos que um estimador para a média populacional global é a média amostral global, enquanto que um estimador para a influência de um fator qualquer será a média amostral daquele fator subtraída da média amostral global. Os estimadores de máxima verossimilhança coincidem com os estimadores de mínimos quadrados.

2. Pesquise sobre testes de homogeneidade de variâncias.

- Teste de Levene (considerando um DIC)

Neste teste, considera-se uma nova variável, Z_{ij}, que é obtida pelo valor absoluto da diferença entre a observação ij e a média do fator, ou seja, $Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|$. A hipótese nula no teste é então dada por $H_0 : \mu_1 = \mu_2 = \cdots = \mu_i$, onde μ_i é a média populacional de Z para cada fator. A estatística do teste é dada por

$$W = \frac{\dfrac{\sum_{i=1}^{a} n\left(\bar{Z}_{i.} - \bar{Z}_{..}\right)^2}{a-1}}{\dfrac{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(Z_{ij} - \bar{Z}_{i.}\right)^2}{N-a}}$$

Sabe-se também que $W \sim F_{(a-1,N-a)}$, sendo equivalente à estatística F realizada na ANOVA.

- Teste de Brown-Forsythe (considerando um DIC)

O teste de Brown-Forsythe é praticamente o mesmo que o teste de Levene, diferindo apenas na construção da variável Z, que ao invés de utilizar a média de cada fator, considera agora a mediana dos mesmos, o que acaba sendo mais robusto nos casos em que há forte assimetria e/ou caudas pesadas nas distribuições amostrais, ou seja, não normalidade.

- Teste de Bartlett (considerando um DIC)

O teste de Bartlett é o teste mais usual e estatisticamente mais poderoso entre os testes quando há simetria nos dados. A hipótese nula é a de que as variâncias de todos os fatores sejam iguais, enquanto a alternativa é a de que pelo menos duas delas sejam distintas. A estatística é dada por

$$\chi^2 = \frac{(N-a)\log\left(S_p^2\right) - \sum_{i=1}^{a}(n-1)\log\left(S_i^2\right)}{1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a}\frac{1}{n-1} - \frac{1}{N-a}\right)}$$

onde $S_p^2 = \frac{1}{N-a}\sum_{i=1}^{a}(n-1)S_i^2$, e S_i^2 são as variâncias dos fatores. A estatística possui distribuição chi-quadrado com a-1 graus de liberdade.

3. Pesquise sobre testes de normalidade.

- Teste de Shapiro-Wilk

Testa diretamente se há normalidade em um conjunto de dados através da estatística

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

com $x_{(i)}$ sendo a estatística de ordem, e $a_i = \dfrac{m^T V^{-1}}{C}$, de onde $C = \|V^{-1}m\|$, m é o vetor transposto dos valores esperados das estatísticas de ordem retiradas de uma simulação de números aleatórios gerados de uma normal padrão, e V é a matriz de covariância das estatísticas de ordem simuladas.

A distribuição da estatística não é conhecida, e os valores tabelados para avaliar as hipóteses são encontrados a partir de simulações.

- Teste de Anderson-Darling

Diferente do teste de Shapiro-Wilk, que busca saber diretamente se determinado conjunto provém de uma distribuição normal, o teste de Anderson-Darling busca saber se os dados seguem uma distribuição qualquer. Ele faz parte de um conjunto de testes que utilizam a função de distribuição empírica para encontrarem suas respectivas estatísticas. Assim, sua estatística é dada por

$$A^2 = -n - \sum_{i=1}^{n}\frac{2i-1}{n}\log\left(F(Y_i)\right) + \log\left(1 - F(Y_{n+1-i})\right)$$

onde $F(.)$ é a função de distribuição acumulada da distribuição hipotetizada.

- Teste de Kolmogorov–Smirnov

Assim como no teste de Anderson-Darling, o teste de Kolmogorov–Smirnov também testa a hipótese nula de que certo conjunto de dados veio de uma distribuição hipotética, e também pode testar se dois conjuntos têm a mesma distribuição (desconhecida). O teste também utiliza a comparação entre função de distribuição empírica e função de distribuição acumulada da distribuição de interesse. A estatística de teste é dada então por

$$D_n = \sup_x |F_n(x) - F(x)|$$

onde $F_n(.)$ é a função de distribuição empírica, e $F(.)$ é a função de distribuição acumulada da distribuição hipotética. Pelo teorema de Glivenko-Cantelli, se a amostra advém de uma distribuição F, então D_n converge para 0 quando n tende ao infinito.

4. Demonstre que $F = t^2$.

Seja t uma estatística que segue distribuição t de Student com v graus de liberdade. Então, $t = \dfrac{Z}{\sqrt{\frac{V}{v}}}$, onde $Z \sim N(0,1)$ e $V \sim \chi^2_v$. Assim:

$$t^2 = \left(\frac{Z}{\sqrt{\frac{V}{v}}}\right)^2 = \frac{Z^2}{\dfrac{V}{v}} = \frac{\dfrac{Z^2}{1}}{\dfrac{V}{v}} = \frac{\dfrac{\chi^2_1}{1}}{\dfrac{\chi^2_v}{v}} \sim F_{(1,v)}$$