

Componentes de Variância

Profa. Cibebe Russo

(Referências: Montgomery (2012), Notas de aula de Roseli Leandro; Clarice Demétrio; Marinho Andrade)

Efeitos fixos x efeitos aleatórios

Efeitos fixos:

- Experimentador escolhe níveis específicos dos fatores
- Inferência se restringe aos níveis estudados
- Fatores são variáveis qualitativas

Efeitos aleatórios:

- Fatores são variáveis quantitativas
- Níveis são escolhidos de forma aleatória a partir de uma população de possíveis níveis
- Inferência pode ser extrapolada para outros níveis não avaliados dos fatores

Modelos com Efeitos Aleatórios

Os modelos lineares com efeitos mistos são uma extensão dos modelos lineares tradicionais que incorporam tanto efeitos fixos quanto efeitos aleatórios:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

$$i = 1, \dots, a,$$

$$j = 1, \dots, n$$

em que ambos τ_i e ϵ_{ij} são variáveis aleatórias independentes.

Os efeitos fixos representam fatores que são de interesse específico para o estudo, enquanto os efeitos aleatórios representam fatores cujas categorias são selecionadas aleatoriamente de uma população maior.

Modelos com Efeitos Aleatórios

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

$$i = 1, \dots, a,$$

$$j = 1, \dots, n$$

Suposições

$Var(\tau_i) = \sigma_\tau^2$, $Var(\epsilon_{ij}) = \sigma^2$, τ_i e ϵ_{ij} independentes.

Segue que

$$Var(y_{ij}) = \sigma_\tau^2 + \sigma^2$$

σ_τ^2 e σ^2 são chamados componentes de variância

Modelos com Efeitos Aleatórios

As somas de quadrados da ANOVA são dados por

$$SQT = SQTrat + SQRes$$

A variabilidade total nas observações em um componente que mede a variação entre tratamentos (SQTrat) e um componente que mede a variação dentro dos tratamentos (SQRes).

Testar hipóteses sobre efeitos individuais de tratamento não é muito significativo porque eles foram selecionados aleatoriamente; estamos mais interessados na população de tratamentos, então testamos hipóteses sobre o componente de variância σ_{τ}^2 .

Avaliação das componentes de variância

$$\begin{cases} H_0 : \sigma_\tau^2 = 0 \text{ contra} \\ H_1 : \sigma_\tau^2 > 0 \end{cases}$$

Se $\sigma_\tau^2 = 0$, todos os tratamentos são idênticos; mas se $\sigma_\tau^2 > 0$, existe variabilidade entre os tratamentos.

Como antes, $SQRes/\sigma^2$ é distribuído como qui-quadrado com $N - a$ graus de liberdade e, sob a hipótese nula, $SQTrat/\sigma^2$ é distribuído como qui-quadrado com $a - 1$ graus de liberdade.

Avaliação das componentes de variância

Ambas as variáveis aleatórias são independentes. Assim, sob a hipótese nula $\sigma_{\tau}^2 = 0$, a razão

$$F_0 = \frac{\text{SQTrat}/(a - 1)}{\text{SQRes}/(N - a)}$$

é distribuída como F com $a - 1$ e $N - a$ graus de liberdade.

Sobre os quadrados médios

$$E(QMTrat) = \frac{1}{a-1}E(SQTrat) = \dots = \sigma^2 + n\sigma_{\tau}^2$$

e

$$E(QMRes) = \sigma^2$$

Os estimadores das componentes de variância são:

$$\hat{\sigma}^2 = QMRes$$

e

$$\hat{\sigma}_{\tau}^2 = \frac{QMTrat - QMRes}{n}$$

Exemplo

Uma empresa têxtil produz tecido em um grande número de teares. Ela gostaria que os teares fossem homogêneos para que obtenha um tecido de resistência uniforme. O engenheiro de processo suspeita que, além da variação usual na resistência dentro das amostras de tecido do mesmo tear, também possa haver variações significativas na resistência entre os teares. Para investigar isso, ela seleciona quatro teares aleatoriamente e faz quatro determinações de resistência no tecido fabricado em cada tear.

EXAMPLE 3.11

A textile company weaves a fabric on a large number of looms. It would like the looms to be homogeneous so that it obtains a fabric of uniform strength. The process engineer suspects that, in addition to the usual variation in strength within samples of fabric from the same loom, there may also

be significant variations in strength between looms. To investigate this, she selects four looms at random and makes four strength determinations on the fabric manufactured on each loom. This experiment is run in random order, and the data obtained are shown in Table 3.17. The ANOVA is con-

■ **TABLE 3.17**
Strength Data for Example 3.11

Looms	Observations				y_i
	1	2	3	4	
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388

$$1527 = y_{..}$$

ducted and is shown in Table 3.18. From the ANOVA, we conclude that the looms in the plant differ significantly.

The variance components are estimated by $\hat{\sigma}^2 = 1.90$ and

$$\hat{\sigma}_\tau^2 = \frac{29.73 - 1.90}{4} = 6.96$$

Therefore, the variance of any observation on strength is estimated by

$$\hat{\sigma}_y^2 = \hat{\sigma}^2 + \hat{\sigma}_\tau^2 = 1.90 + 6.96 = 8.86.$$

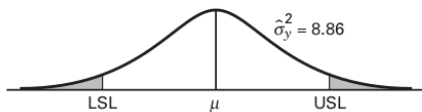
Most of this variability is attributable to differences *between* looms.

Fonte: Montgomery (2009)

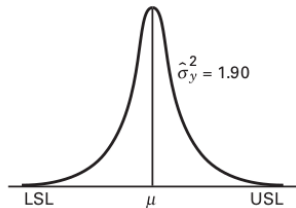
■ **TABLE 3.18**
Analysis of Variance for the Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P -Value
Looms	89.19	3	29.73	15.68	<0.001
Error	22.75	12	1.90		
Total	111.94	15			

Fonte: Montgomery (2009)



(a) Variability of process output.



(b) Variability of process output if $\sigma_\tau^2 = 0$.

Inferência sobre as componentes de variância

Podemos facilmente encontrar um intervalo de confiança para a componente de variância σ^2 . Se as observações forem independentes e provenientes de uma população normal, então $(N - a)Q\text{MRes}/\sigma^2$ é distribuído como χ^2_{N-a} . Assim, a probabilidade

$$P\left(\chi^2_{1-\alpha/2, N-a} < \frac{(N - a)Q\text{MRes}}{\sigma^2} < \chi^2_{\alpha/2, N-a}\right) = 1 - \alpha$$

e um intervalo de confiança de $100(1 - \alpha)\%$ para σ^2 é

$$\left(\frac{(N - a)\text{QMRes}}{\chi_{\alpha/2, N-a}^2}, \frac{(N - a)\text{QMRes}}{\chi_{1-\alpha/2, N-a}^2} \right)$$

Uma vez que $\text{QMRes} = 1.90$, $N = 16$, $a = 4$, $\alpha = 0.025$, $\chi_{0.025, 12}^2 = 23.3367$ e $\chi_{0.975, 12}^2 = 4.4038$, o intervalo de confiança de 95% para σ^2 é $[0.9770; 5.1775]$.

Inferência sobre as componentes de variância

Agora, considere o componente de variância σ_τ^2 . O estimador pontual de σ_τ^2 é

$$\hat{\sigma}_\tau^2 = \frac{\text{QMTrat} - \text{QMRes}}{n}$$

A variável aleatória $(a - 1)\text{QMTrat}/(\sigma^2 + n\sigma_\tau^2)$ é distribuída como χ_{a-1}^2 , e $(N - a)\text{QMRes}/\sigma^2$ é distribuída como χ_{N-a}^2 .

Assim, a distribuição de probabilidade de $\hat{\sigma}_\tau^2$ é uma combinação linear de duas variáveis aleatórias qui-quadrado, digamos $u_1\chi_{a-1}^2 + u_2\chi_{N-a}^2$, onde

$$u_1 = \frac{\sigma^2 + n\sigma_\tau^2}{n(a-1)}$$

e

$$u_2 = \frac{\sigma^2}{n(N-a)}$$

Infelizmente, uma expressão em forma fechada para a distribuição dessa combinação linear de variáveis aleatórias qui-quadrado não pode ser obtida.

Pode-se encontrar um intervalo de confiança para a razão $\frac{\sigma_\tau^2}{(\sigma_\tau^2 + \sigma^2)}$. Essa razão é chamada de coeficiente de correlação intraclass e reflete a proporção da variância de uma observação (lembre-se de que $V(y_{ij}) = \sigma_\tau^2 + \sigma^2$) que é resultado das diferenças entre tratamentos.

Para desenvolver este intervalo de confiança para o caso de um delineamento balanceado, observe que QMTrat e QMRes são variáveis aleatórias independentes e, além disso, pode-se mostrar que

$$\frac{\text{QMTrat}/(n\sigma_\tau^2 + \sigma^2)}{\text{QMRes}/\sigma^2} \sim F_{a-1, N-a}$$

Assim,

$$\frac{\text{QMTrat}}{\text{QMRes}} \frac{\sigma^2}{(n\sigma_\tau^2 + \sigma^2)} \frac{1}{\sigma_\tau^2} F_{1-\alpha/2, a-1, N-a} \leq \frac{\sigma_\tau^2}{(\sigma_\tau^2 + \sigma^2)} \leq$$

$$\frac{\text{QMTrat}}{\text{QMRes}} \frac{\sigma^2}{(n\sigma_\tau^2 + \sigma^2)} \frac{1}{\sigma_\tau^2} F_{\alpha/2, a-1, N-a}$$

Podemos então obter o seguinte:

$$P(L \leq \frac{\sigma_\tau^2}{\sigma^2} \leq U) = 1 - \alpha$$

onde

$$L = \frac{1}{n} \left(\frac{\text{QMTrat}}{\text{QMRes}} \frac{1}{F_{\alpha/2, a-1, N-a}} - 1 \right)$$

e

$$L = \frac{1}{n} \left(\frac{\text{QMTrat}}{\text{QMRes}} \frac{1}{F_{1-\alpha/2, a-1, N-a}} - 1 \right)$$

Observe que L e U são os limites de confiança inferiores e superiores, respectivamente, com $100(1 - \alpha)\%$ de confiança, para a razão $\frac{\sigma_\tau^2}{(\sigma_\tau^2 + \sigma^2)}$. Portanto, um intervalo de confiança de $100(1 - \alpha)\%$ para $\frac{\sigma_\tau^2}{(\sigma_\tau^2 + \sigma^2)}$ é

$$\left(\frac{L}{1 + L}, \frac{U}{1 + U} \right) \frac{\sigma^2}{\sigma_\tau^2}$$

Para ilustrar este procedimento, encontramos um intervalo de confiança de 95% para $\frac{\sigma_\tau^2}{(\sigma_\tau^2 + \sigma^2)}$.

Para os dados de resistência no Exemplo 3.11. Lembre-se de que $QM_{Trat} = 29.73$, $QM_{Res} = 1.90$, $a = 4$, $n = 4$, $F_{0.025,3,12} = 4.47$, e $F_{0.975,3,12} = 1/F_{0.025,12,3} = 1/14.34 = 0.070$. Portanto,

$$L = \frac{1}{4} \left(\frac{29.73}{1.90} \frac{1}{4.47} - 1 \right) = 0.625$$

$$U = \frac{1}{4} \left(\frac{29.73}{1.90} \frac{1}{0.070} - 1 \right) = 55.633$$

O intervalo de confiança de 95% para $\frac{\sigma_{\tau}^2}{(\sigma_{\tau}^2 + \sigma^2)}$ é dado por

$$\frac{0.625}{1.625} \leq \frac{\sigma_{\tau}^2}{(\sigma_{\tau}^2 + \sigma^2)} \leq \frac{55.633}{56.633}$$

ou

$$0.38 \leq \frac{\sigma_{\tau}^2}{(\sigma_{\tau}^2 + \sigma^2)} \leq 0.98$$

Concluimos que a variabilidade entre teares representa entre 38% e 98% da variabilidade na resistência observada do tecido produzido. Este intervalo de confiança é relativamente amplo devido ao pequeno número de teares utilizados no experimento. No entanto, claramente, a variabilidade entre teares (σ_{τ}^2) não é negligenciável.

Estimação da Média Geral μ

Em muitos experimentos de efeitos aleatórios, o experimentador está interessado em estimar a média geral μ . A partir das suposições básicas do modelo, é fácil ver que o valor esperado de qualquer observação é apenas a média geral. Consequentemente, um estimador não-viesado da média geral é:

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

Então, para o Exemplo 3.11, a estimativa da média geral de resistência é:

$$\hat{\mu} = \frac{\bar{y}..}{N} = \frac{1527}{16} = 95.44$$

Também é possível encontrar um intervalo de confiança de $100(1 - \alpha)\%$ para a média geral. A variância de y é:

$$Var(\bar{y}_{..}) = Var\left(\frac{\sum_{i=1}^a \sum_{j=1}^n y_{ij}}{an}\right) = \frac{n\sigma_{\tau}^2 + \sigma^2}{an}$$

O numerador desta razão é estimado pelo quadrado médio do tratamento, então um estimador não-viesado de $Var(\bar{y}_{..})$ é:

$$\hat{V}(\bar{y}_{..}) = \frac{QMTrat}{an}$$

Portanto, o intervalo de confiança de $100(1 - \alpha)\%$ para a média geral é:

$$\bar{y}_{..} - t_{\alpha/2, a(n-1)} \sqrt{\frac{\text{QMTrat}}{an}} \leq \mu \leq \bar{y}_{..} + t_{\alpha/2, a(n-1)} \sqrt{\frac{\text{QMTrat}}{an}}$$

Portanto, o intervalo de confiança de $100(1 - \alpha)\%$ para a média geral é:

$$\bar{y}_{..} - t_{\alpha/2, a(n-1)} \sqrt{\frac{QMTrat}{an}} \leq \mu \leq \bar{y}_{..} + t_{\alpha/2, a(n-1)} \sqrt{\frac{QMTrat}{an}}$$

Para encontrar um intervalo de confiança de 95% para a média geral no experimento de resistência do tecido do Exemplo 3.11, precisamos de $QMTrat = 29.73$ e $t_{0.025, 12} = 2.18$. O intervalo de confiança é calculado da seguinte forma:

$$95.44 - 2.18 \sqrt{\frac{29.73}{20}} \leq \mu \leq 95.44 + 2.18 \sqrt{\frac{29.73}{20}}$$
$$92.78 \leq \mu \leq 98.10$$

Portanto, com 95% de confiança, a resistência média do tecido produzido pelos teares nesta instalação está entre 92.78 e 98.10.

Este é um intervalo de confiança relativamente amplo porque um pequeno número de teares foi amostrado e há uma grande diferença entre os teares, como refletido pela grande parte da variabilidade total que é explicada pelas diferenças entre os teares.

Pode-se mostrar que os Estimadores de Máxima Verossimilhança (MLEs) são:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$$

e

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

Note que o MLE $\hat{\sigma}^2$ não é o familiar desvio padrão amostral. Ele não leva em conta a estimativa do parâmetro de localização μ . O estimador REML seria:

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

O estimador REML é não-viesado.

Estimação das Componentes de Variância

Para estimar as componentes de variância em um modelo linear com efeitos mistos, podem ser utilizados métodos de máxima verossimilhança restrita (REML) ou métodos de mínimos quadrados generalizados (GLS).

Para ilustrar o método REML (máxima verossimilhança restrita), a Tabela 3.19 apresenta a saída do JMP para o experimento de teares no exemplo. As estimativas REML dos parâmetros do modelo μ , σ^2 , e σ_τ^2 estão mostradas na saída.

Observe que as estimativas REML dos componentes de variância são idênticas às encontradas anteriormente pelo método ANOVA. Esses dois métodos concordarão para delineamentos balanceados. No entanto, a saída REML também contém a matriz de covariância dos componentes de variância.

As raízes quadradas dos elementos diagonais principais desta matriz são os erros padrão dos componentes de variância.

Se $\hat{\sigma}$ é o MLE do erro, então o intervalo de confiança aproximado de $100(1 - \alpha)\%$ em σ^2 é:

$$\hat{\sigma} \pm Z_{\alpha/2} \hat{\sigma} \sqrt{\frac{1}{2n}}$$

O JMP utiliza este método para encontrar os ICs aproximados para σ_{τ}^2 e σ^2 mostrados na saída. O IC de 95% do REML para σ^2 é muito semelhante ao intervalo baseado em qui-quadrado calculado anteriormente.

■ TABLE 3.19

JMP Output for the Loom Experiment in Example 3.11

Response Y

Summary of Fit

RSquare	0.793521
RSquare Adj	0.793521
Root Mean Square Error	1.376893
Mean of Response	95.4375
Observations (or Sum Wgts)	16

Parameter Estimates

Term	Estimate	Std Error	DFDen	t Ratio	Prob> t
Intercept	95.4375	1.363111	3	70.01	<.0001*

REML Variance Component Estimates

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
X1	3.6703297	6.9583333	6.0715247	-4.941636	18.858303	78.588
Residual		1.8958333	0.7739707	0.9748608	5.1660065	21.412
Total		8.8541667				100.000

Covariance Matrix of Variance Component Estimates

Random Effect	X1	Residual
X1	36.863412	-0.149758
Residual	-0.149758	0.5990307

Máxima Verossimilhança Restrita (REML)

- REML é uma abordagem estatística amplamente utilizada para a estimação de parâmetros em modelos de efeitos mistos.
- Ele ajusta o modelo para estimar os parâmetros fixos e, em seguida, “restringe” a verossimilhança para estimar os efeitos aleatórios.
- REML leva em consideração o fato de que os parâmetros fixos são estimados e não conhecidos com certeza.
- Os estimadores REML geralmente têm menos viés do que os estimadores de máxima verossimilhança (ML) quando aplicados a dados de amostra pequena.

Mínimos Quadrados Generalizados (GLS)

- GLS é uma técnica alternativa que também pode ser usada para estimar parâmetros em modelos de efeitos mistos.
- Envolve a minimização de uma função de perda ponderada, em que os pesos são derivados da estrutura de correlação dos erros.
- Os estimadores GLS são eficientes sob as suposições do modelo e da estrutura de correlação dos erros.

A escolha entre REML e GLS pode depender das características dos dados, do tamanho da amostra e das suposições do modelo.

Exemplo em R

```
# Carregando pacotes necessários
library(lme4)

## Loading required package: Matrix

# Definindo o tamanho da amostra
n_operadores <- 5 # Número de operadores
n_pecas <- 10 # Número de peças por operador
n_total <- n_operadores * n_pecas

# Definindo os operadores
operadores <- rep(1:n_operadores, each = n_pecas)

# Definindo as peças
pecas <- rep(1:n_pecas, times = n_operadores)
```

```
# Simulando as medições  
set.seed(123) # Fixa semente para reprodutibilidade  
medicoes <- rnorm(n_total, mean = 100, sd = 5)  
  
# Criando o dataframe  
dados <- data.frame(Operador = factor(operadores),  
                    Peca = factor(pecas),  
                    Medicao = medicoes)
```

```
# Ajustando o modelo com efeitos aleatórios  
modelo <- lmer(Medicacao ~ (1 | Operador) + (1 | Peca),  
              data = dados)  
  
## boundary (singular) fit: see help('isSingular')
```



```
# Resumo do modelo
```

```
summary(modelo)
```

```
## Linear mixed model fit by REML ['lmerMod']  
## Formula: Medicao ~ (1 | Operador) + (1 | Peca)  
## Data: dados  
##  
## REML criterion at convergence: 293.1  
##  
## Scaled residuals:  
##      Min      1Q  Median      3Q      Max  
## -2.1612 -0.6413 -0.1156  0.7169  2.3055  
##  
## Random effects:  
## Groups   Name                Variance  Std.Dev.  
## Peca     (Intercept)  0.000e+00 0.000000  
## Operador (Intercept) 1.087e-06 0.001042  
## Residual                2.143e+01 4.629350
```