MAT02014 - Planejamento de Experimentos II
Efeitos aninhados

Rodrigo Citton P. dos Reis
rodrigocpdosreis@gmail.com

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2018

Onde estamos?

- Até aqui:
  - Efeitos fixos
  - Efeitos aleatórios
  - Ambos no contexto de delineamentos **fatoriais cruzados**
- Agora:
  - Estrutura de **fatores aninhados**
  - **Modelos mistos:** uma combinação de efeitos fixos e aleatórios.

Delineamentos fatoriais com efeitos aninhados

Relembrando: fatores cruzados

- Com os **fatores cruzados** A e B, **vemos** (por definição) **todas as combinações possíveis** de níveis de fatores, ou seja, podemos configurar uma tabela de dados da seguinte forma

■ **TABLE 5.2**

General Arrangement for a Two-Factor Factorial Design

		Factor *B*			
		1	2	. . .	*b*
Factor *A*	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$
	2	$y_{211}, y_{212}, \ldots, y_{21n}$	$y_{221}, y_{222}, \ldots, y_{22n}$		$y_{2b1}, y_{2b2}, \ldots, y_{2bn}$
	$\vdots$				
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots, y_{abn}$

Relembrando: fatores cruzados

- Isto significa: vemos **cada nível** do fator A em **todos os níveis** do fator B (e vice-versa).
- O nível 1 do fator A tem o mesmo significado em **todos os níveis** do fator B.

Exemplo: desempenho de alunos

- Queremos analisar o desempenho de alunos.
- Dados de diferentes **turmas** de diferentes **escolas** (a nível de **aluno**).
- Qual o (grau) de variabilidade
  - entre diferentes escolas?
  - entre turmas **dentro** da mesma escola?
  - entre alunos **dentro** da mesma turma?
- Isso parece um **novo delineamento**, pois as turmas claramente **não** são cruzadas com escolas, similarmente para alunos.
- Isso nos leva a uma nova definição …

Estrutura de fatores aninhados

- Nós chamamos fator B **aninhado** no fator A se tivermos níveis diferentes de B dentro de cada nível de A.

Factor A / Factor B	1	2	3	4	5	6	7	8	9	10	11	12
1	x	x										
2			x	x								
3					x	x						
4							x	x				
5									x	x		
6											x	x

- Por exemplo, pense em
  - A: escola
  - B: turma
- Também escrevemos $B(A)$.
- Os dados não são necessariamente apresentados neste formato.

Fatores aninhados: exemplo

- Dados apresentados

	Class 1	Class 2
School 1	x	x
School 2	x	x
School 3	x	x

- Estrutura dos dados subjacente

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
School 1	x	x				
School 2			x	x		
School 3					x	x

- Portanto, a turma é aninhada na escola, pois a classe 1 na escola 1 não tem nada a ver com classe 1 na escola 2.

Fatores aninhados

- Só porque as turmas são rotuladas 1 e 2 não significa que é um delineamento fatorial cruzado.
- Portanto, pergunte-se sempre se o nível 1 do fator realmente corresponde ao mesmo objeto em todos os níveis do outro fator.
- Geralmente, usamos parênteses no índice para indicar aninhamento, ou seja, o modelo é escrito como

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

- α_i: efeito de escola
- $\beta_{j(i)}$: efeito de turma dentro de escola
- $\epsilon_{k(ij)}$: erro aleatório; também escritos na "notação aninhada"

Por que aninhar?

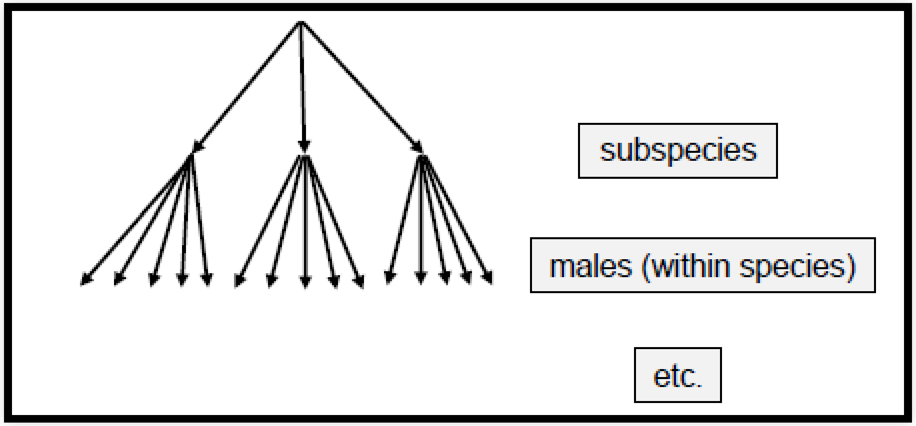
- Tipicamente usamos uma estrutura aninhada devido a restrições práticas/logísticas.
- Por exemplo:
  - Pacientes são aninhados em hospitais, pois não iremos enviar pacientes para todas as clínicas.
  - As amostras são aninhadas em lotes (controle de qualidade).

Exemplo: delineamento completamente aninhado

- Chamamos um delineamento completamente aninhado se cada fator estiver aninhado em seu predecessor.

Exemplo de genômica:

- Considere três subespécies.
- Escolha aleatoriamente cinco machos de cada subespécie (= 15 machos).
- Cada macho é acasalado com quatro fêmeas diferentes da mesma subespécie (= 60 fêmeas).
- Observe 3 filhotes por acasalamento (= 180 filhotes).
- Faça duas medições por filhote (= 360 medições)


subspecies
males (within species)
etc.

Exemplo: delineamento completamente aninhado

- Usamos o modelo

$$Y_{ijklm} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \delta_{l(ijk)} + \epsilon_{m(ijkl)}$$

- **Exercício:** liste os efeitos do modelo para o exemplo discutido.
- Para calcular a correspondente soma de quadrados, usamos a decomposição

$$\begin{aligned}(y_{ijklm} - \bar{y}_{.....}) = (\bar{y}_{i....} - \bar{y}_{.....}) + (\bar{y}_{ij...} - \bar{y}_{i....}) \\ +(\bar{y}_{ijk..} - \bar{y}_{ij...}) + (\bar{y}_{ijkl.} - \bar{y}_{ijk..}) \\ +(\bar{y}_{ijklm} - \bar{y}_{ijkl.}).\end{aligned}$$

elevamos ao quadrado e somamos sobre todos os índices.

Tabela de ANOVA para delineamentos completamente aninhados

- Temos a seguinte decomposição

$$SS_T = SS_A + SS_{B(A)} + SS_{C(AB)} + + SS_{D(ABC)} + SS_E$$

- Assumindo que temos apenas **efeitos aleatórios** e um delineamento **balanceado**, temos a seguinte tabela de ANOVA

Source	df	E[MS]
A	$a-1$	$\sigma^2 + n\sigma_\delta^2 + nd\sigma_\gamma^2 + ncd\sigma_\beta^2 + nbcd\sigma_\alpha^2$
$B(A)$	$a(b-1)$	$\sigma^2 + n\sigma_\delta^2 + nd\sigma_\gamma^2 + ncd\sigma_\beta^2$
$C(AB)$	$ab(c-1)$	$\sigma^2 + n\sigma_\delta^2 + nd\sigma_\gamma^2$
$D(ABC)$	$abc(d-1)$	$\sigma^2 + n\sigma_\delta^2$
Error	$abcd(n-1)$	σ^2

- Com esta informação podemos construir **testes** e **estimadores** para diferentes componentes de variância.

Tabela de ANOVA para delineamentos completamente aninhados

- Testes F são construídos considerando a razão dos quadrados médios dos erros "vizinhos", já que eles diferem apenas pelo componente de variação de interesse.
- Isso significa que sempre usamos o quadrado médio do do sucessor na hierárquia como denominador.
- Por exemplo, $F = \frac{MS_A}{MS_{B(A)}}$ para testar $H_0 : \sigma_\alpha^2 = 0$ *vs.* $H_1 : \sigma_\alpha^2 > 0$.

Exemplo: força da pasta

- Conjunto de dados do pacote `lme4`.
- Produto de pasta química contido em barris.
- 10 **entregas** (**lotes**) foram selecionadas **aleatoriamente**.
- De cada entrega, 3 **barris** foram selecionados **aleatoriamente**.
- Por barril: faça duas **medições**.

Exemplo (R)

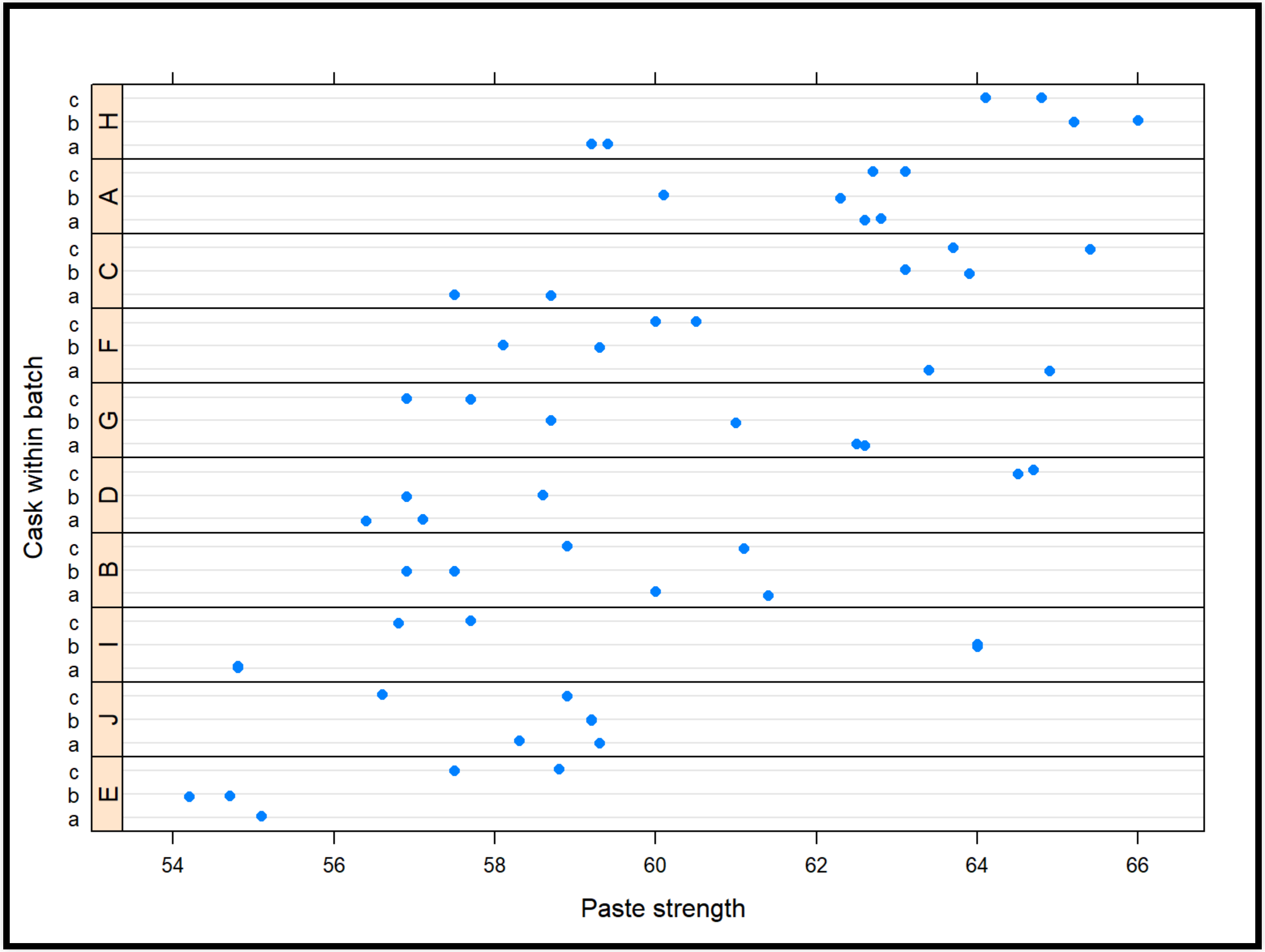
```r
library(lme4)
data(Pastes)
head(Pastes)
```

```
##   strength batch cask sample
## 1     62.8     A    a    A:a
## 2     62.6     A    a    A:a
## 3     60.1     A    b    A:b
## 4     62.3     A    b    A:b
## 5     62.7     A    c    A:c
## 6     63.1     A    c    A:c
```

```r
tail(Pastes)
```

```
##    strength batch cask sample
## 55     58.3     J    a    J:a
## 56     59.3     J    a    J:a
## 57     59.2     J    b    J:b
## 58     59.2     J    b    J:b
## 59     58.9     J    c    J:c
## 60     56.6     J    c    J:c
```

Exemplo (R)



Cask within batch
Paste strength
H
A
C
F
G
D
B
I
J
E
c
b
a
54
56
58
60
62
64
66

Exemplo (R)

```
fit <- aov(strength ~ batch + batch:cask, data = Pastes)
#fit <- aov(strength ~ batch/cask, data = Pastes) ## equivalent formulat:
#fit <- aov(strength ~ batch + cask %in% batch, data = Pastes)
summary(fit)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## batch           9  247.4  27.489   40.55 2.28e-14 ***
## batch:cask     20  350.9  17.545   25.88 9.79e-14 ***
## Residuals      30   20.3   0.678
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exemplo (R)

	E[MS]
batch	$\sigma^2 + 2 \cdot \sigma_\beta^2 + 2 \cdot 3 \cdot \sigma_\alpha^2$
cask(batch)	$\sigma^2 + 2 \cdot \sigma_\beta^2$
Error	σ^2

- Da saída podemos manualmente calcular os diferentes componentes de variância resolvendo as equações que correspondem às esperanças dos quadrados médios.
- **Exercício:** (1) qual é este método de estimação? (2) Estime os componentes de variância.

```
##               Df Sum Sq Mean Sq F value   Pr(>F)    
## batch          9  247.4  27.489   40.55 2.28e-14 ***
## batch:cask    20  350.9  17.545   25.88 9.79e-14 ***
## Residuals     30   20.3   0.678                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exemplo (R)

Método dos momentos

```r
sigma2 <- .678
sigma2bc <- (17.545 - sigma2) / 2
sigma2b <- (27.489 - sigma2 - 2 * sigma2bc ) / (2*3)
cat("Method of Moments Variance Component Estimates","\n",
    "Var(error)=",sigma2,"\n","Var(C(B))=",sigma2bc,"\n",
    "Var(B)=",sigma2b,"\n")
```

```
## Method of Moments Variance Component Estimates
##  Var(error)= 0.678
##  Var(C(B))= 8.4335
##  Var(B)= 1.657333
```

Exemplo (R)

- Similarmente, testes têm que ser computados manualmente.
- Por exemplo, para o componente de variância de lote:

```
pf(27.489 / 17.545, 9, 20, lower.tail = FALSE)
```

```
## [1] 0.1925487
```

- Assim, não podemos rejeitar $H_0 : \sigma_A^2 = 0$.

Exemplo (R)

```
## lmer approaches (all equivalent) ####
fm1 <- lmer(strength ~ (1 | batch/cask), data = Pastes)
summary(fm1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: strength ~ (1 | batch/cask)
##    Data: Pastes
##
## REML criterion at convergence: 247
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4798 -0.5156  0.0095  0.4720  1.3897
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  cask:batch (Intercept) 8.434    2.9041
##  batch      (Intercept) 1.657    1.2874
##  Residual               0.678    0.8234
## Number of obs: 60, groups:  cask:batch, 30; batch, 10
##
## Fixed effects:
```

Exemplo (R)

```
confint(fm1, oldNames = FALSE)
```

```
##                                   2.5 %    97.5 %
## sd_(Intercept)|cask:batch     2.1579337  4.053589
## sd_(Intercept)|batch          0.0000000  2.946591
## sigma                         0.6520234  1.085448
## (Intercept)                  58.6636504 61.443016
```

Exemplo (R)

```r
fm2 <- lmer(strength ~ (1 | batch) + (1 | cask:batch), data = Pastes)
library(lmerTest)
rand(fm2)
```

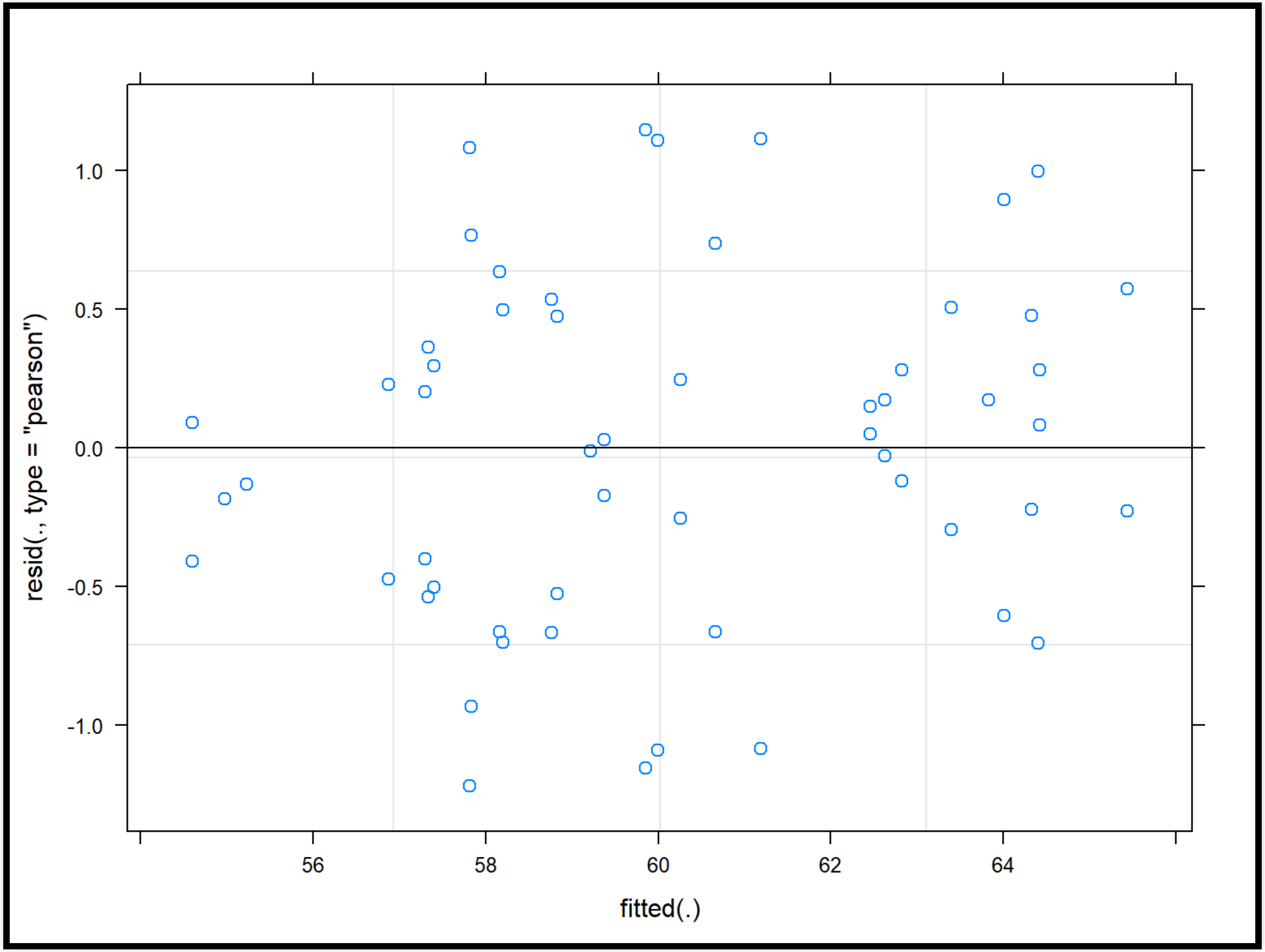
```
## Analysis of Random effects Table:
##            Chi.sq Chi.DF p.value
## batch       0.658      1     0.4
## cask:batch 54.605      1   1e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exemplo

- **Pergunta:** o que podemos concluir?

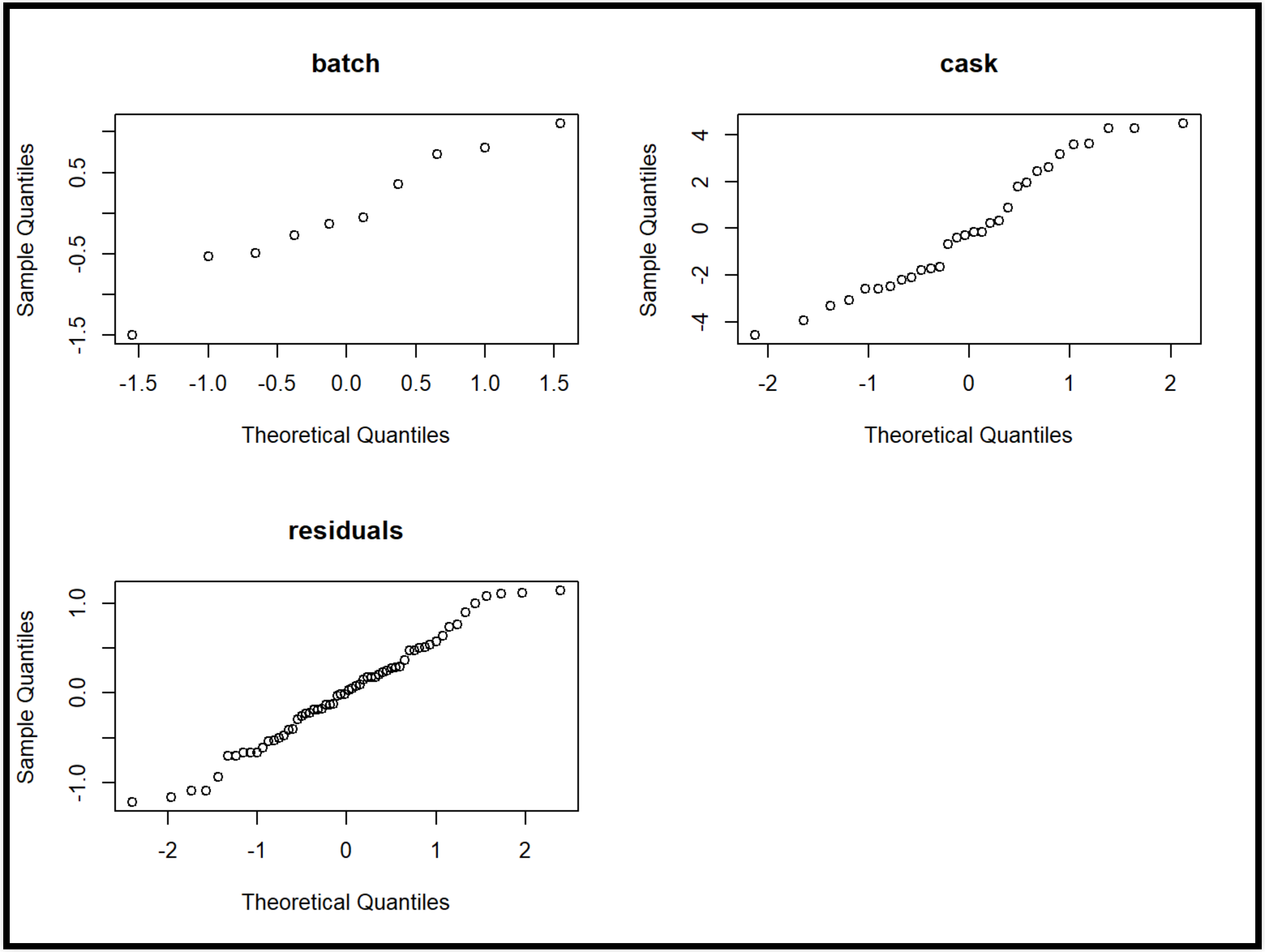
Exemplo (R)

```
## residual analysis ####
plot(fm1) ## Tukey-Anscombe plot
```


1.0
0.5
0.0
-0.5
-1.0
resid(., type = "pearson")
56
58
60
62
64
fitted(.)

Exemplo (R)

```r
par(mfrow = c(2,2))
qqnorm(ranef(fm1)$batch[,1], main = "batch")
qqnorm(ranef(fm1)$'cask:batch'[,1], main = "cask")
qqnorm(resid(fm1), main = "residuals")
```


batch
Sample Quantiles
Theoretical Quantiles
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
cask
Sample Quantiles
Theoretical Quantiles
-4
-2
0
2
4
-2
-1
0
1
2
residuals
Sample Quantiles
Theoretical Quantiles
-1.0
0.0
1.0
-2
-1
0
1
2

Próxima aula

- Lista de exercícios 2 **(não haverá aula presencial)**