

Princípios Planejamento e Análise

PIAnEco

Alexandre Adalardo de Oliveira

PlanECO 2025

Modelos Lineares Mistos

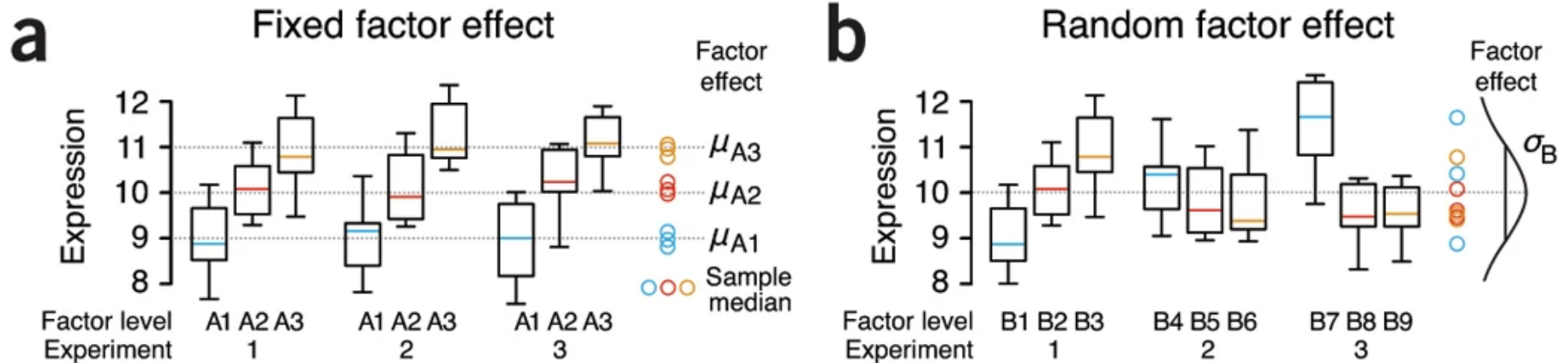
Princípios

dependência entre as observações

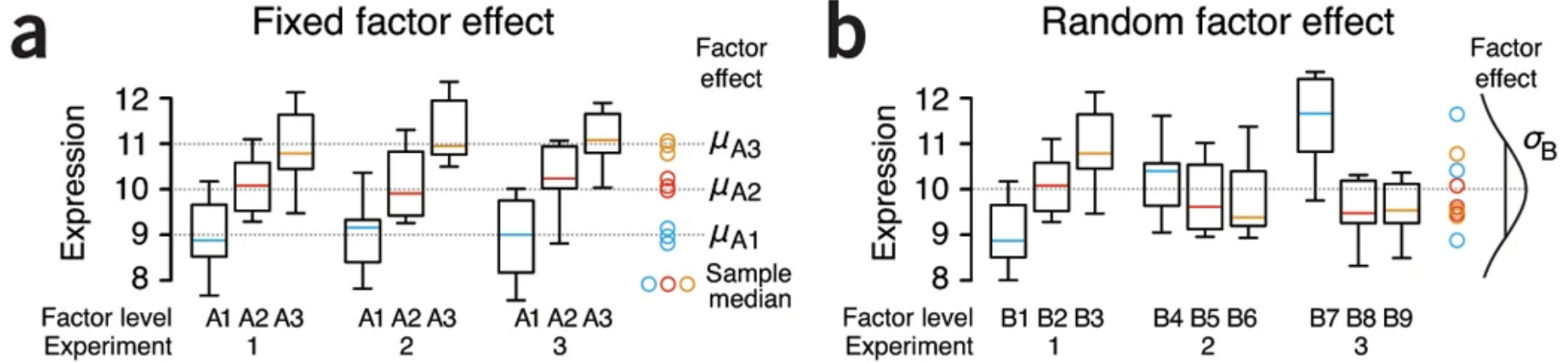
- espacial
- temporal
- biológica

preditora aleatória (categórica)

- amostras aleatória de níveis possíveis
- estima a variação entre os níveis



Preditora fixa x aleatória



LMM: exemplo

Exemplo: cresce árvore

Espécies decíduas crescem + ou - ?

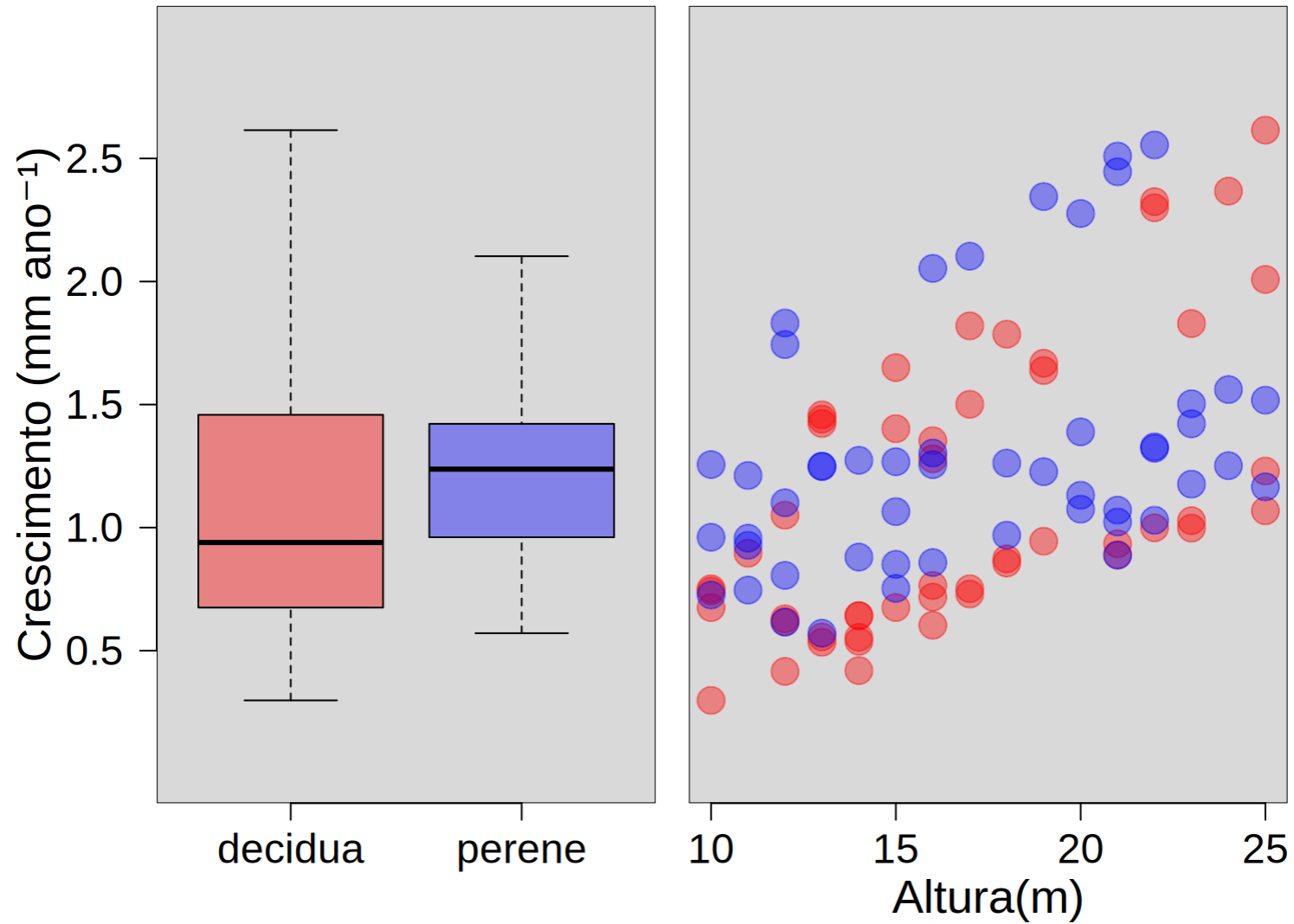
- 100 observações
- deciduidade: dois níveis (decídua, perene)
- 5 espécies em cada categoria
- 10 indivíduos em cada espécie
- crescimento controlado pelo tamanho

Exemplo: crescimento de árvores

cresc	sp	dec	alt
0.64	sp01	dec	14
0.75	sp01	dec	17
2.01	sp02	dec	25
1.44	sp03	dec	13
0.63	sp04	dec	12
0.55	sp05	dec	14
2.45	sp06	per	21
1.07	sp07	per	21
1.25	sp08	per	13
0.86	sp09	per	16

cresc	sp	dec	alt
1.52	sp10	per	25
1.25	sp10	per	13

Crescimento de árvores



lm: cresce árvores

- alt:dec

```
lmFull1 <- lm(cresc ~ alt + dec + alt:dec, data = dados)
lm01 <- lm(cresc ~ alt + dec, data = dados)
anova(lm01, lmFull1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cresc ~ alt + dec
```

```
## Model 2: cresc ~ alt + dec + alt:dec
```

```
##      Res.Df      RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1         97 21.501
```

```
## 2         96 20.807  1    0.69379 3.201 0.07675 .
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

lm: cresce árvores

- alt

```
lm02 <- lm(cresc ~ dec, data = dados)
anova(lm02, lm01)
```

```
## Analysis of Variance Table
##
## Model 1: cresc ~ dec
## Model 2: cresc ~ alt + dec
##   Res.Df    RSS Df Sum of Sq   F    Pr(>F)
## 1      98 28.238
## 2      97 21.501   1    6.7371 30.394 2.913e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

lm: cresce árvores

- dec

```
lm03 <- lm(cresc ~ alt, data = dados)
anova(lm03, lm01)
```

```
## Analysis of Variance Table
##
## Model 1: cresc ~ alt
## Model 2: cresc ~ alt + dec
##   Res.Df    RSS Df Sum of Sq    F  Pr(>F)
## 1      98 22.173
## 2      97 21.501  1   0.67183 3.0309 0.08486 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

lm: cresce árvores

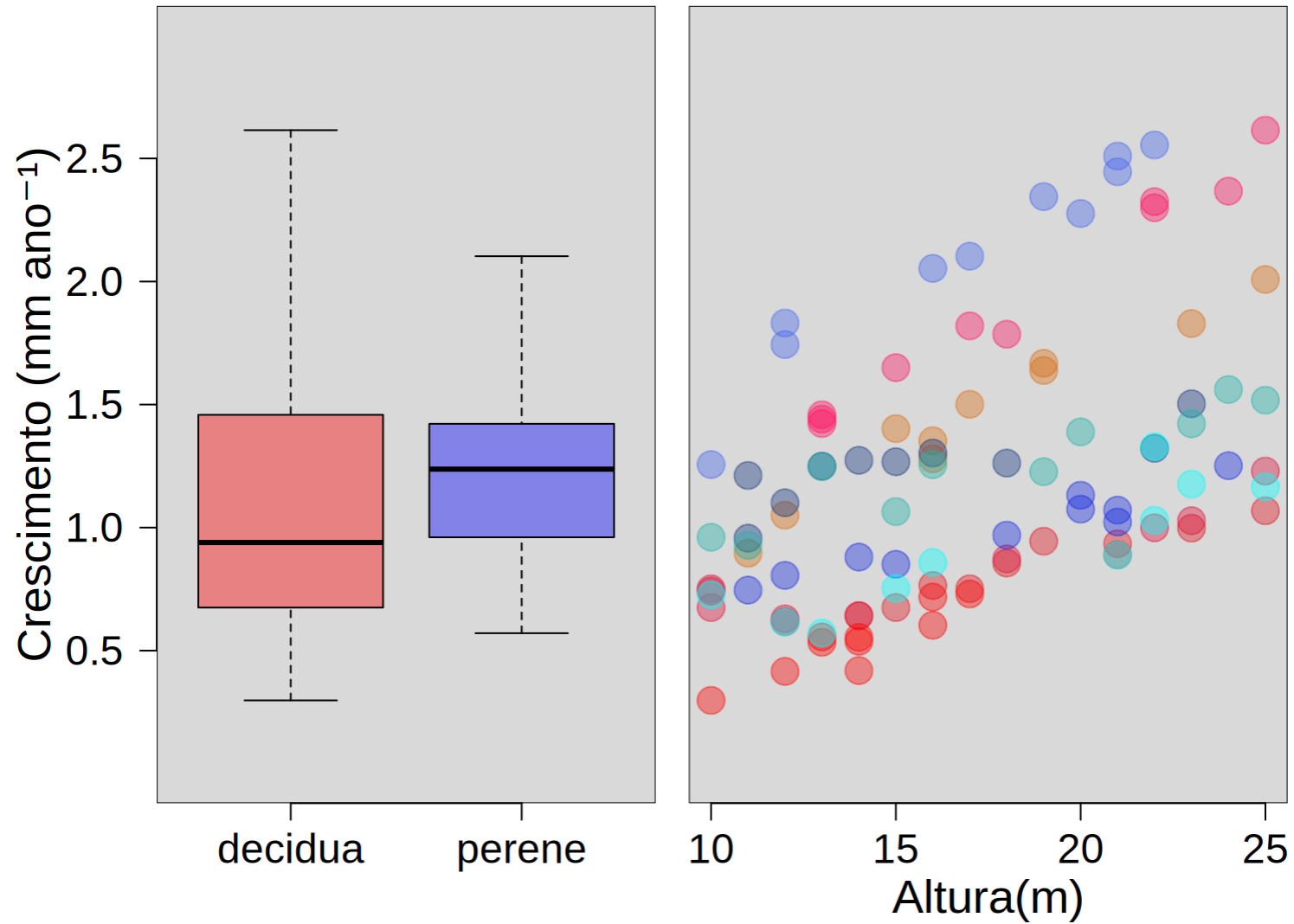
Mínimo adequado

```
summary(lm03)
```

```
##
## Call:
## lm(formula = cresc ~ alt, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6201 -0.4005 -0.1092  0.2928  1.0665
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.23052     0.18345   1.257    0.212
## alt         0.05775     0.01045   5.525 2.72e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4757 on 98 degrees of freedom
## Multiple R-squared:  0.2375, Adjusted R-squared:  0.2297
## F-statistic: 30.53 on 1 and 98 DF,  p-value: 2.716e-07
```

- **Apenas a altura importa**

Crescimento de árvores



Cresce Árvores

Problema:

- os indivíduos de uma mesma espécie não são independentes
- as espécies tem relações de parentescos (**ISSO FICA PARA DEPOIS!**)

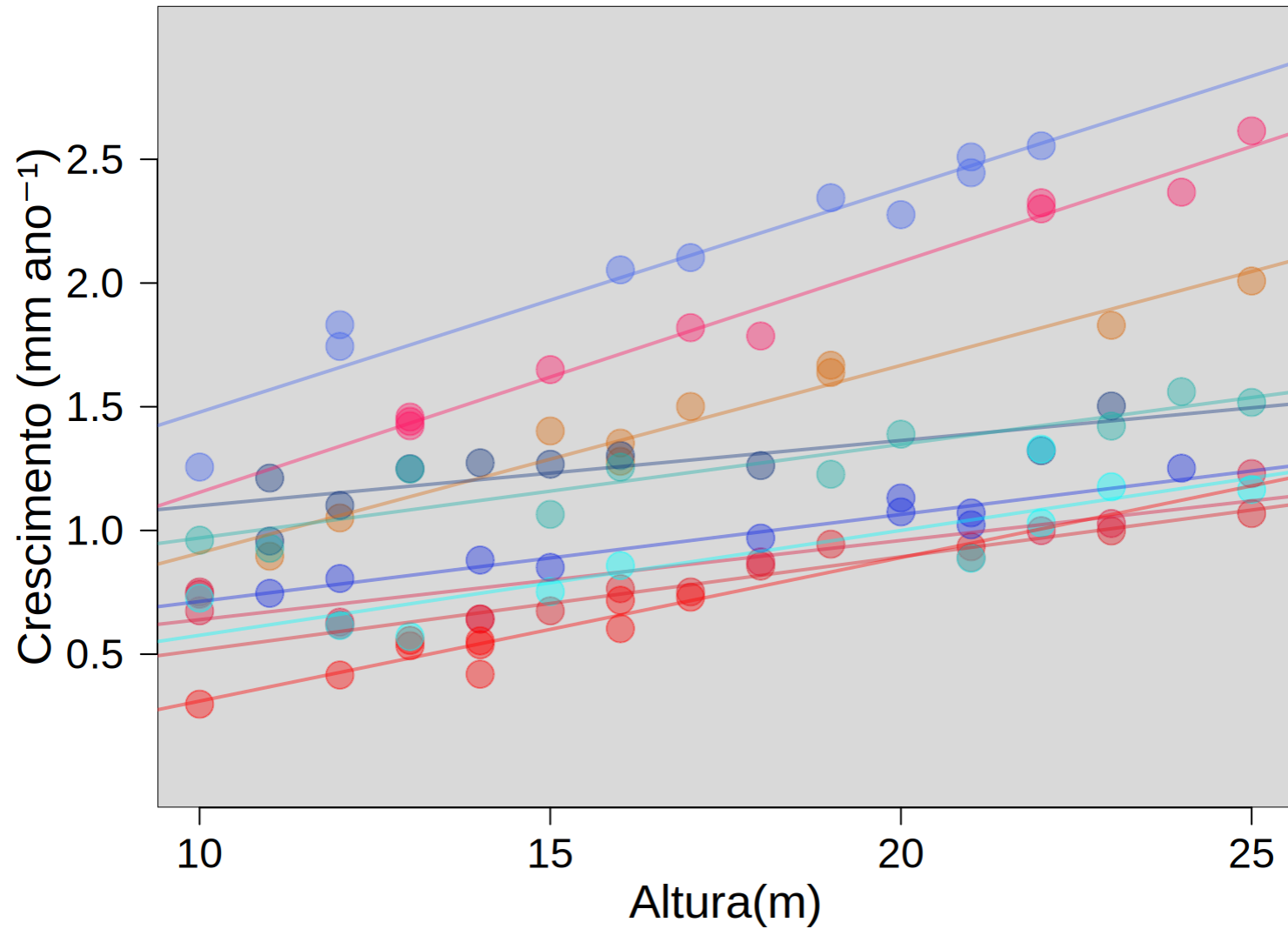
Soluções:

1. Procedimento em duas etapas

$$n : 100 \rightarrow 10$$

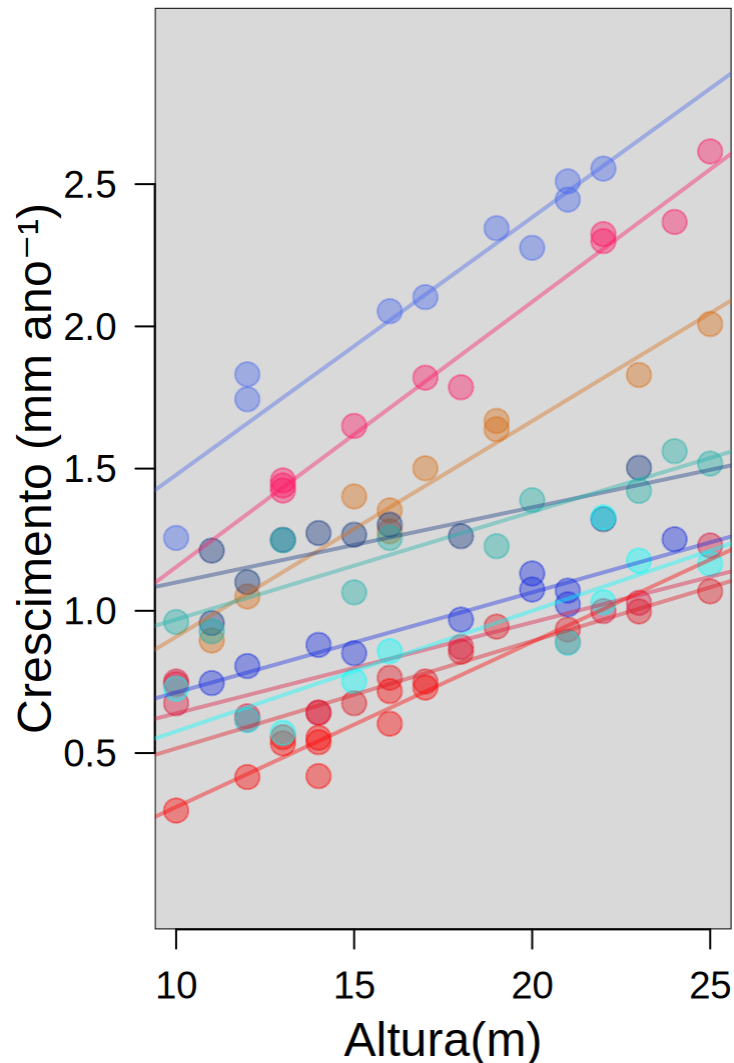
2. LMM com efeito aleatório de espécies

Diferentes Espécies



Procedimento: duas etapas

- Modelo Linear para cada espécie



```
lmsp1 <- lm(cres ~ alt,  
            data= arv, subset  
            = arv$sp=='sp01')  
lmsp2 <- lm(cres ~ alt,  
            data= arv, subset  
            = arv$sp=='sp02')]  
lmsp3 <- lm(cres ~ alt,  
            data= arv, subset  
            = arv$sp=='sp03')]  
...
```

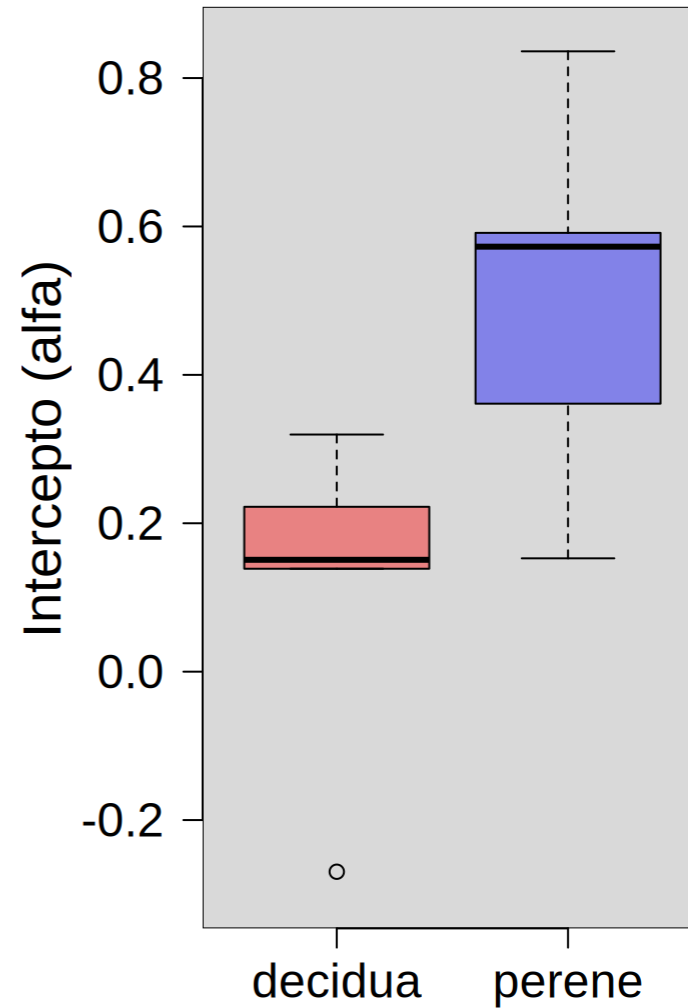
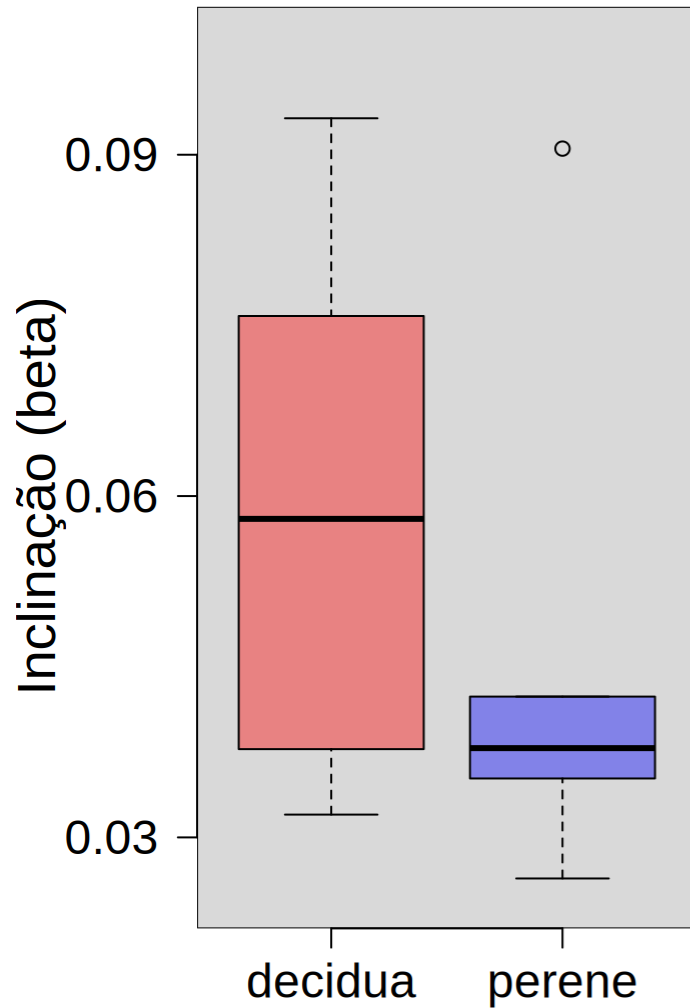
- **testa os coeficientes de inclinação** entre os níveis: decídua e perene
- se não forem significativos, **testa o intercepto**

Coeficientes do LM

sp	dec	alfa	beta
sp01	dec	0.14	0.04
sp02	dec	0.15	0.08
sp03	dec	0.22	0.09
sp04	dec	0.32	0.03
sp05	dec	-0.27	0.06
sp06	per	0.57	0.09
sp07	per	0.36	0.04
sp08	per	0.84	0.03
sp09	per	0.15	0.04

sp	dec	alfa	beta
sp10	per	0.59	0.04

LM dos coeficientes



Comparando as inclinações

```
lmspBeta <- lm(beta ~ dec, data=coefs)
lmspBetanull <- lm(beta ~ 1, data=coefs)
anova(lmspBetanull, lmspBeta)

## Analysis of Variance Table
##
## Model 1: beta ~ 1
## Model 2: beta ~ dec
##      Res.Df      RSS Df Sum of Sq      F Pr(>F)
## 1          9 0.0056125
## 2          8 0.0051970  1 0.00041547 0.6395  0.447
```

Não há diferença de inclinação entre decíduas e perenes

Comparando os interceptos

```
lmspAlfa <- lm(alfa~dec, data=coefs)
lmspAlfanull <- lm(alfa ~ 1, data=coefs)
anova(lmspAlfanull,lmspAlfa)

## Analysis of Variance Table
##
## Model 1: alfa ~ 1
## Model 2: alfa ~ dec
##      Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1          9 0.85061
## 2          8 0.46942  1    0.38119 6.4964 0.03424 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Resultado

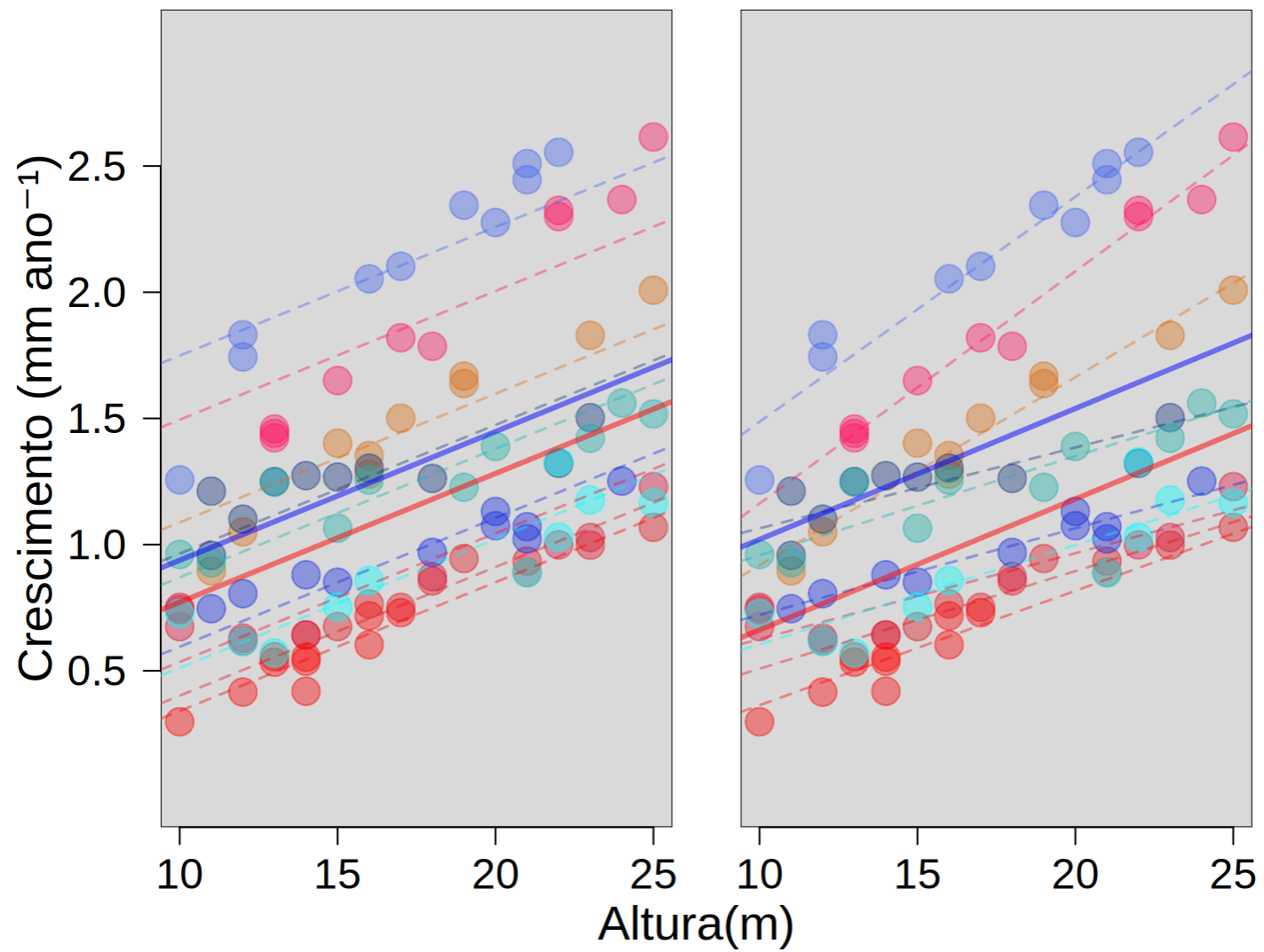
```
summary(lmspAlfa)
```

```
##
## Call:
## lm(formula = alfa ~ dec, data = coefs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3819 -0.0996  0.0542  0.1045  0.3332
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1123     0.1083   1.037   0.3302
## decper         0.3905     0.1532   2.549   0.0342 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2422 on 8 degrees of freedom
## Multiple R-squared:  0.4481, Adjusted R-squared:  0.3792
## F-statistic: 6.496 on 1 and 8 DF,  p-value: 0.03424
```

Perenes crescem em média + 0.4 mm ano⁻¹

LMM: cresce árvores

Intercepto ou Inclinação



LMM: efeito aleatório

inclinação (alt|sp)

```
lmmincl <- lmer(cresc ~ alt + dec + alt:dec + (alt|sp),  
               data = dados, REML=TRUE)
```

intercepto (1|sp)

```
lmmint <- lmer(cresc ~ alt + dec + alt:dec + (1|sp),  
              data = dados, REML=TRUE)  
anova(lmmincl, lmmint, refit=FALSE)
```

```
## Data: dados  
## Models:  
## lmmint: cresc ~ alt + dec + alt:dec + (1 | sp)  
## lmmincl: cresc ~ alt + dec + alt:dec + (alt | sp)  
##          npar      AIC      BIC logLik -2*log(L)  Chisq Df  
Pr(>Chisq)  
## lmmint      6  -40.971  -25.340  26.486   -52.971  
## lmmincl     8 -108.018  -87.176  62.009  -124.018  71.046  2  
3.736e-16 ***
```

```
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '  
' 1
```

retemos o modelo mais complexo (alt|sp)

LMM: efeito fixo

Retirando a interação alt:dec

```
lmmssp02 <- lmer(cresc ~ alt + dec + (alt|sp), data=dados)
anova(lmmssp02, lmmincl)
```

```
## Data: dados
## Models:
## lmmssp02: cresc ~ alt + dec + (alt | sp)
## lmmincl: cresc ~ alt + dec + alt:dec + (alt | sp)
##          npar      AIC      BIC logLik -2*log(L)  Chisq Df
Pr(>Chisq)
## lmmssp02      7 -129.42 -111.18  71.710    -143.42
## lmmincl      8 -127.90 -107.06  71.952    -143.90 0.4841  1
0.4866
```

O modelo com interação não fornece informação adicional relevante

LMM: efeito fixo

Retirando deciduidade

```
lmmssp01 <- lmer(cresc ~ alt + (alt|sp), data=dados)
```

```
anova(lmmssp01, lmmssp02)
```

```
## Data: dados
## Models:
## lmmssp01: cresc ~ alt + (alt | sp)
## lmmssp02: cresc ~ alt + dec + (alt | sp)
##          npar      AIC      BIC logLik -2*log(L)  Chisq Df
Pr(>Chisq)
## lmmssp01      6 -125.98 -110.35  68.993   -137.99
## lmmssp02      7 -129.42 -111.18  71.710   -143.42  5.4349  1
0.01974 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Deciduidade é informativa e deve ser retida no modelo

LMM: simplificando o efeito fixo

Retirando altura

```
lmmssp03 <- lmer(cresc ~ dec + (alt|sp), data=dados)
```

```
anova(lmmssp03, lmmssp02)
```

```
## Data: dados
## Models:
## lmmssp03: cresc ~ dec + (alt | sp)
## lmmssp02: cresc ~ alt + dec + (alt | sp)
##          npar      AIC      BIC logLik -2*log(L)  Chisq Df
Pr(>Chisq)
## lmmssp03      6 -114.15   -98.516  63.074    -126.15
## lmmssp02      7 -129.42 -111.184  71.710    -143.42 17.273   1
3.238e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Altura é informativa e deve ser retida no modelo

LMM: resultado

```
summary(lmmsp02)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cresc ~ alt + dec + (alt | sp)
## Data: dados
##
## REML criterion at convergence: -130
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.72998 -0.53247 -0.05237  0.58674  2.97258
##
## Random effects:
##   Groups      Name      Variance  Std.Dev.  Corr
##   sp          (Intercept) 0.0416225 0.20402
##           alt          0.0005977 0.02445   0.07
##   Residual          0.0071125 0.08434
## Number of obs: 100, groups:  sp, 10
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 0.143127   0.107096  1.336
## alt         0.051886   0.008027  6.464
## decper      0.358505   0.148234  2.419
##
## Correlation of Fixed Effects:
##      (Intr) alt
## alt    -0.065
```

```
## decper -0.721 0.026
```


LMM: resultado

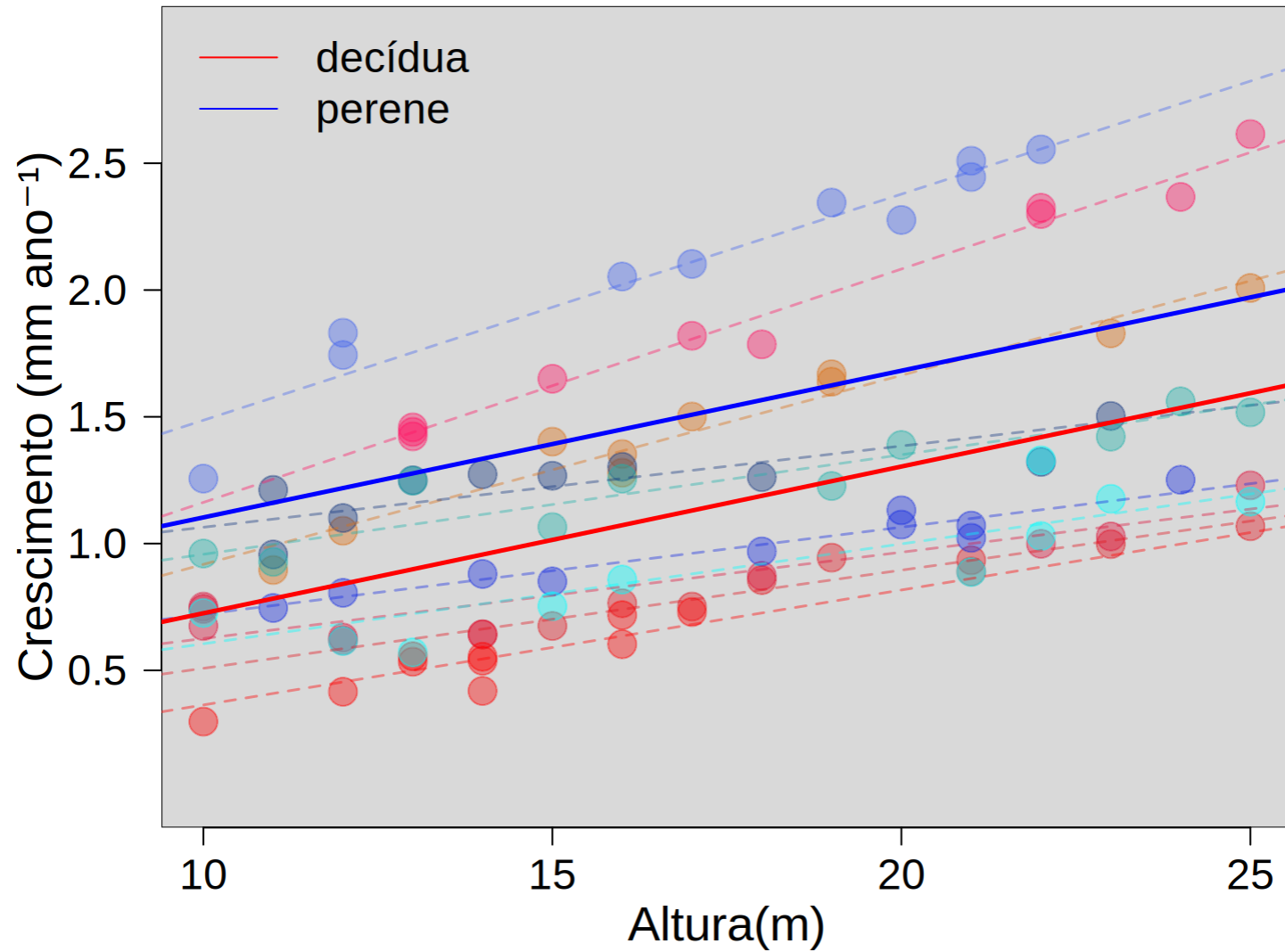
```
lmmNull <- lmer(cresc ~ 1 + (alt|sp), data = dados)
anova(lmmNull, lmmsp02)
```

```
## Data: dados
## Models:
## lmmNull: cresc ~ 1 + (alt | sp)
## lmmsp02: cresc ~ alt + dec + (alt | sp)
##          npar      AIC      BIC logLik -2*log(L)  Chisq Df
Pr(>Chisq)
## lmmNull      5 -110.69   -97.662  60.344   -120.69
## lmmsp02      7 -129.42  -111.184  71.710   -143.42 22.732  2
1.158e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

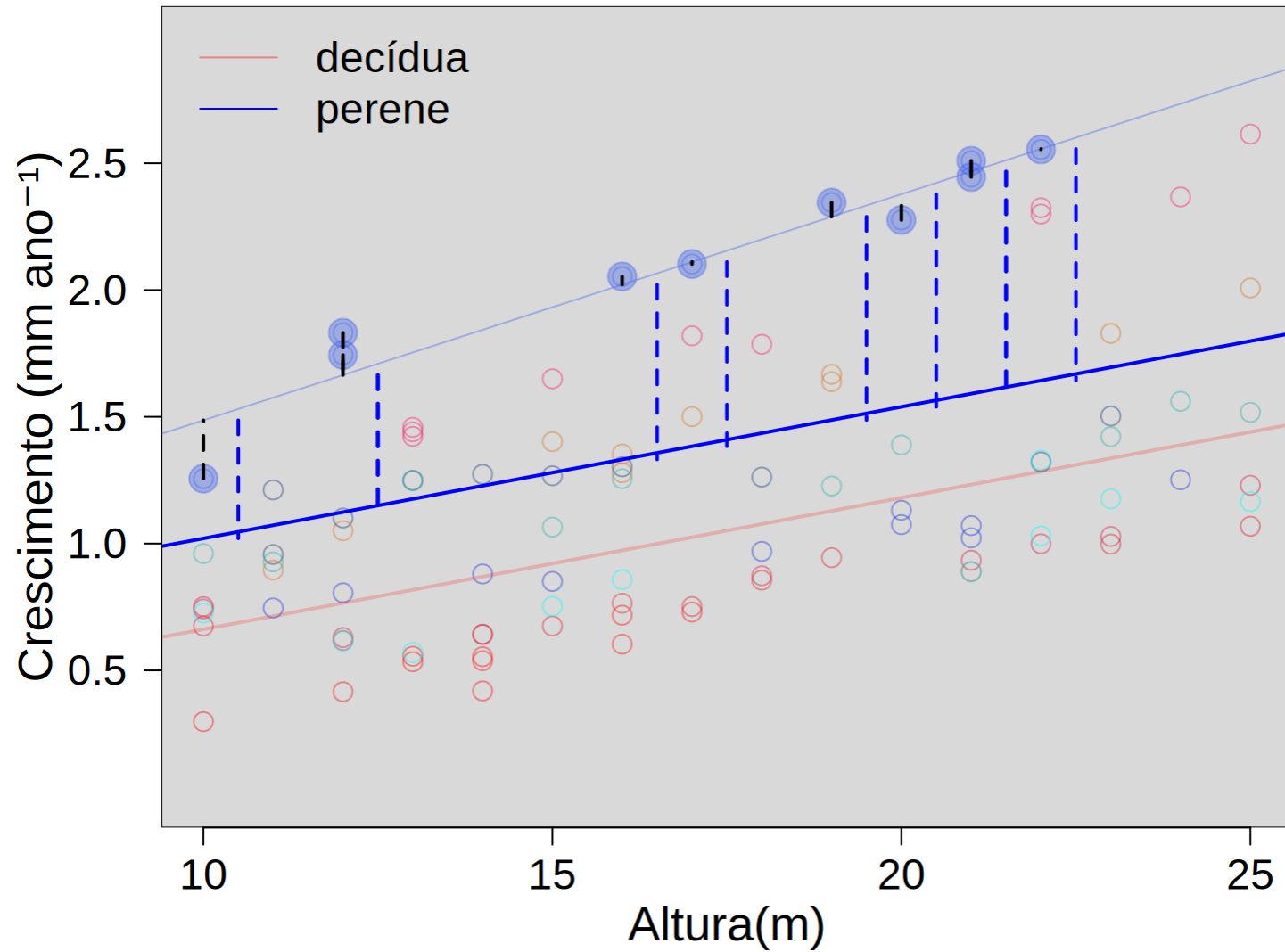
LMM: condicional

	(Intercept)	alt
sp01	-0.0221555	-0.0131752
sp02	0.0294034	0.0226638
sp03	0.0990827	0.0401401
sp04	0.1419220	-0.0178436
sp05	-0.2328465	-0.0065543
sp06	0.0930178	0.0372900
sp07	-0.1236096	-0.0175543
sp08	0.2424243	-0.0198471
sp09	-0.2908253	-0.0124891
sp10	0.0635867	-0.0126303

LMM: árvores



Resíduos do modelo: sp06



Interpretação Biológica

```
fixef(lmmsp02)
```

```
## (Intercept)          alt          decper  
## 0.14312740 0.05188638 0.35850519
```

- **crescimento perenes + 0.36 mm ano⁻¹**
- **a cada metro de altura há um incremento de 0.05 mm ano⁻¹**

Variação dos dados

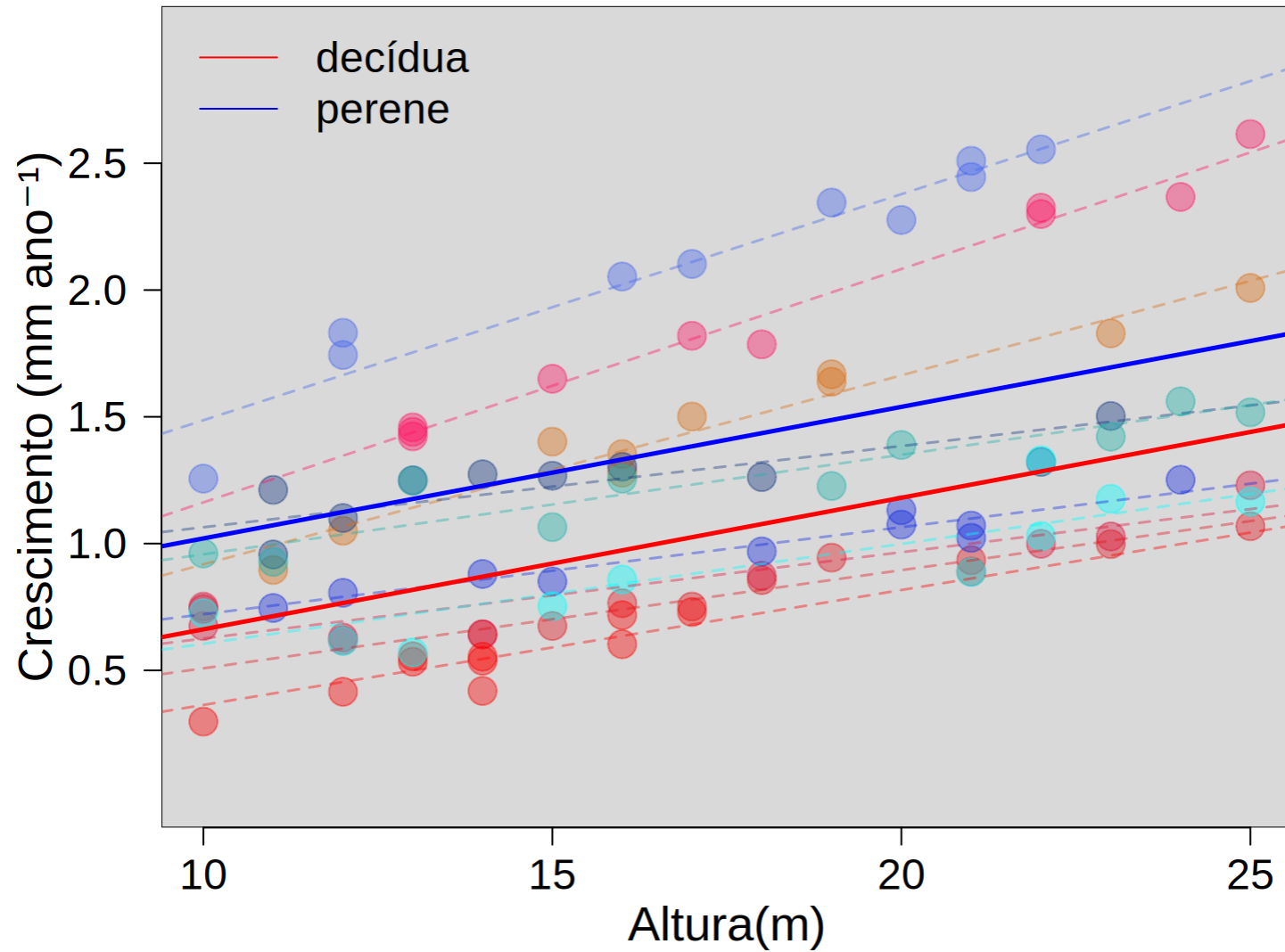
$$sd_{a_{sp}} = 0.20$$

$$sd_{b_{sp}} = 0.02$$

$$sd_{res} = 0.08$$

- **intercepto com altura varia entre espécies**
- **inclinação com altura varia pouco entre espécies**
- **variação não explicada (altura, sp, dec) é pequena**
- **variação entre espécies é maior que a intra**

LMM: árvores



LMM: cresce árvores fim

Modelos Lineares Mistos

Princípios

Testes Clássicos

Resposta	Preditoras	Teste	Hipótese
Contínua	Categórica(2)	Teste-t	$\mu_1 = \mu_2$
Contínua	Categórica (>2)	Anova	$\mu_1 = \mu_2 = \mu_3$
Contínua	1 Contínua	Regressão	$\beta_1 = 0$
Contínua	>1 Contínua	Reg. múltipla	$\beta_1 = 0; \beta_n = 0$
Contínua	Cont + Cat	Ancova	$\beta_1 = \beta_2; \alpha_1 = \alpha_2$
Proporção	Contínua	Reg. Logística	$\text{logit}^{-1}(\beta_1) = 0.5$

- **independência entre observações**

Modelos Lineares (LM)

Características:

- resposta: variável contínua
- preditora: múltiplas (contínuas, discretas, fator)
 - fator: variável indicadora (nível basal no intercepto)
- ajuste do modelo:
 - desvios quadráticos
 - F (razão da variância)

Premissas:

- relação linear: $y = \alpha + \beta x$
- estrutura dos resíduos: $N(0, \sigma)$
- **independência entre observações**

Modelos Lineares Mistos:

Variáveis preditoras: fixas e aleatórias

- incorporam dependência entre observações
 - espaço
 - tempo
 - biologia
- desenho em bloco/aninhado

Efeito (Fator) Aleatório:

- dependência entre observações (**correlação**)
- **agrupamento de observações**
- não há interesse em interpretá-la (**pode ter!**)
- preditora aleatória: $N(\mu, \sigma_{entre}^2)$
- **depende da sua pergunta**

Efeito Aleatório

Box 13.1 WHEN TO TREAT A PREDICTOR VARIABLE AS A RANDOM EFFECT

You may want to treat a predictor variable as a random effect if you:

- don't want to test hypotheses about differences between responses at particular levels of the grouping variable;
- do want to quantify the variability among levels of the grouping variable;
- do want to make predictions about unobserved levels of the grouping variable;
- do want to combine information across levels of the grouping variable;
- have variation in information per level (number of samples or noisiness);
- have levels that are randomly sampled from/representative of a larger population;
- have a categorical predictor that is a nuisance variable (i.e., it is not of direct interest, but should be controlled for).

Cf. Crawley (2002); Gelman (2005)

If you have sampled fewer than five levels of the grouping variable, you should strongly consider treating it as a fixed effect even if one or more of the criteria above apply.

- Ben Bolker, 2015

Fox et. al. Ecological Statistics Princípios

Definições: fator aleatório

Frequentista

Modelos Mistos ou multinível

- variáveis categóricas em que os níveis são amostras aleatórias da população

Bayesiana

Modelos hierárquicos

- parâmetros são realizações de distribuição aleatória

Modelo Misto ou Multinível

Efeito aleatório no intercepto

$$y_{ij} = (\bar{\alpha} + \epsilon_j) + \beta x_{ij} + \epsilon_{ij}$$

Efeito aleatório

$$\epsilon_j = N(0, \sigma_{entre})$$

$$\epsilon_{ij} = N(0, \sigma_{intra})$$

Variância total

$$\sigma_{total}^2 = \sigma_{entre}^2 + \sigma_{intra}^2$$

Modelo Misto ou Multinível

Efeito aleatório no intercepto

Modelo Médio

$$y = \bar{\alpha} + \bar{\beta}x + \epsilon$$

Modelo Condicional

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \epsilon_{\text{res}_{ij}}$$

-

$$\alpha_j = N(\bar{\alpha}, \sigma_{\alpha})$$

-

$$\beta_j = N(\bar{\beta}, \sigma_{\beta})$$

$$y_{ij} = \bar{\alpha} + \epsilon_{\alpha_j} + (\bar{\beta} + \epsilon_{\beta_j})x_{ij} + \epsilon_{\text{res}_{ij}}$$

-

$$\epsilon_{\alpha} = N(0, \sigma_{\alpha})$$

-

$$\epsilon_{\beta} = N(0, \sigma_{\beta})$$

Variância do LMM

Dependência entre observações;

- parte da variação total incorporada entre níveis

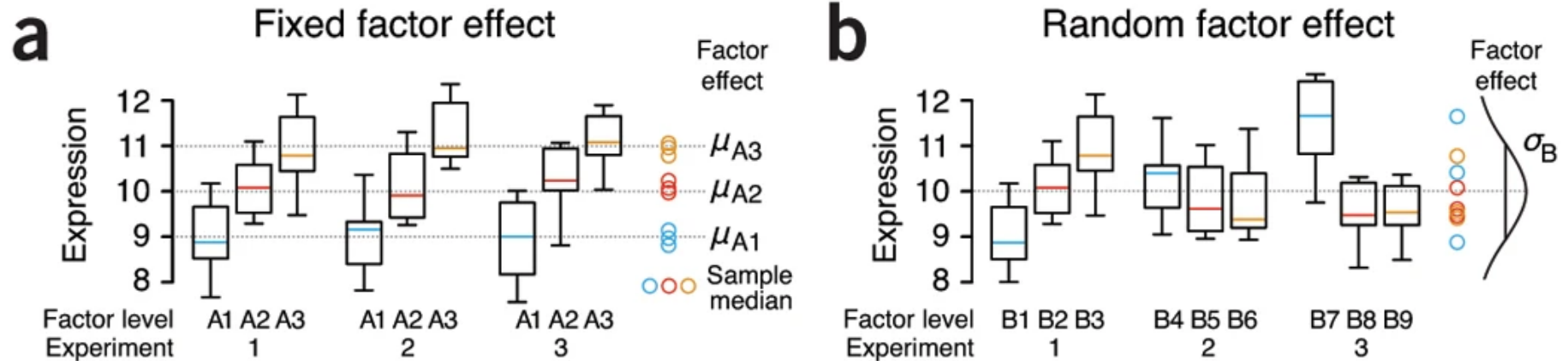
$$\sigma_{total}^2 = \sigma_{entre}^2 + \sigma_{intra}^2$$

- correlação entre 2 observação de um mesmo nível

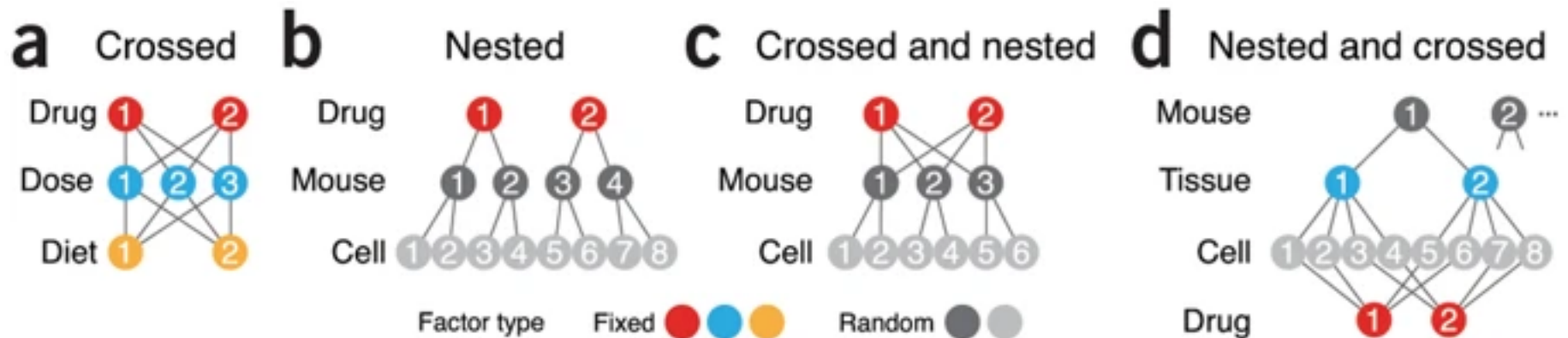
$$\rho = \frac{\sigma_{entre}}{\sigma_{total}}$$

- observações entre níveis $\rho = 0$ (**independentes!**)

Multiplos Efeitos Aleatórios



Cruzado ou Aninhado



Krzywinski et al. 2014. Nature Methods 11(10)

Estrutura aleatória lme4

Intercepto

```
lmer(y ~ x + (1|z))
```

```
lmer(y ~ x + k + x:k + (1|z))
```

Intercepto e inclinação

```
lmer(y ~ x + (x|z))
```

```
lmer(y ~ x + k + x:k + (x|z))
```

Multiplas aleatória lme4

Efeito Cruzado

Serve para efeito aninhado se os código que representam os níveis são únicos

```
lmer(y ~ x + (1|z) + (1|w))  
lmer(y ~ x + (x|z) + (x|w))  
lmer(y ~ x + k + (x|z) + (1|w))
```

Efeito Aninhado

```
lmer(y ~ x + (1|z/w))  
lmer(y ~ x + (x|z/w))  
lmer(y ~ x + k + (x|z/w) + (k|z/w))
```

Modelo Mistos: uma abordagem

1. Parta com efeitos fixos e aleatórios **cheios**
2. Simplifique a estrutura aleatória:
 - comparação com anova usando REML: refit= FALSE
 - retenha a estrutura aleatória minima adequada
 - mantenha termos associada ao desenho experimental (correlação entre amostras)
3. Definido a estrutura do efeito aleatório:
 - simplifique a estrutura fixa do modelo
 - retenha as variáveis e interações minimas adequadas
 - use comparação por anova com ML (padrão)
4. Diagnóstico do Modelo
 - resíduos: homogeneidade, normalidade

LMM: fim dos princípios

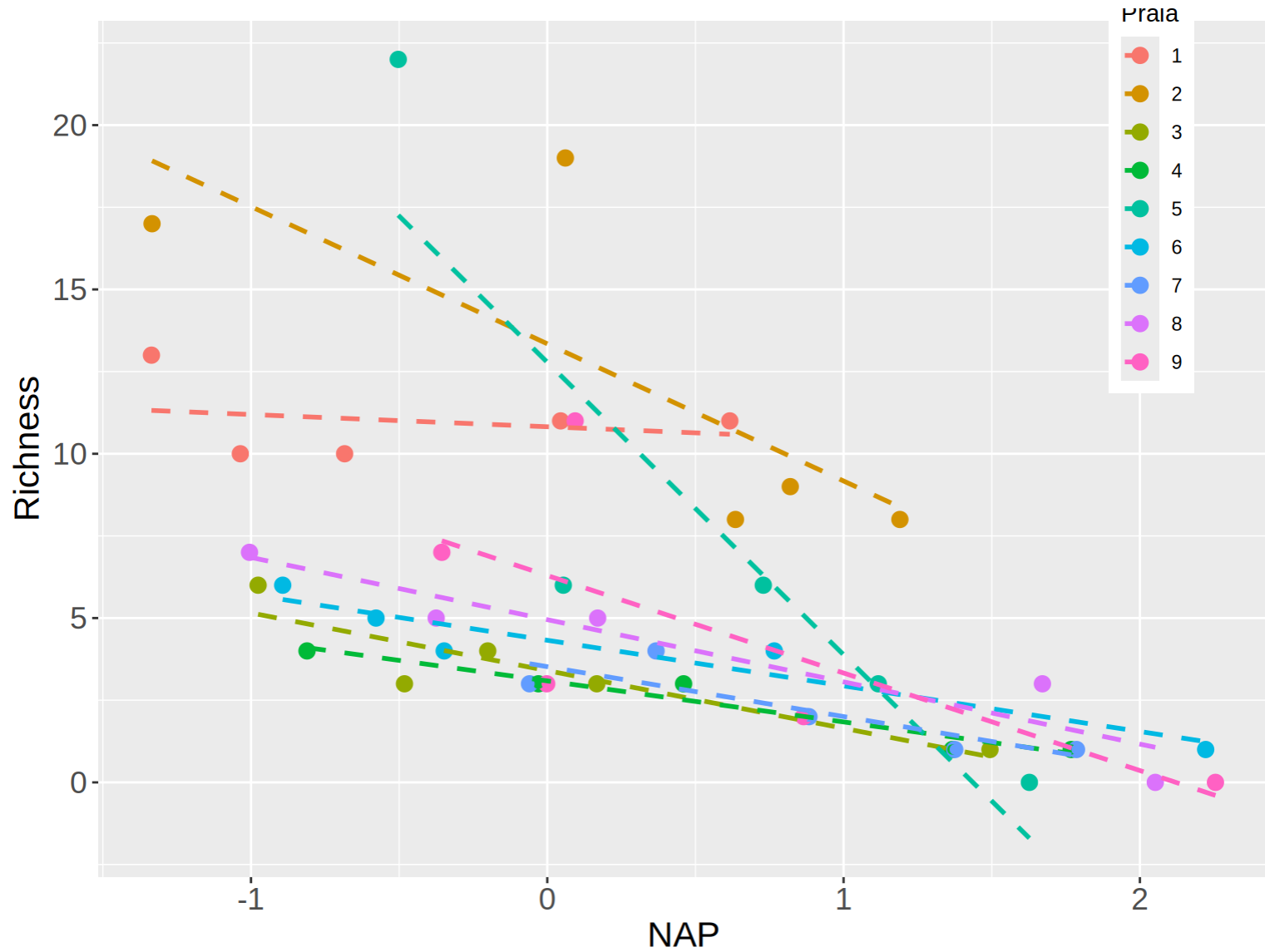
LMM: riqueza na praia

A riqueza da macrofauna varia em função da altura e exposição da praia

- 45 observações
- 9 praias
 - **NAP**: contínuo, altura em relação ao NMM
- **fExp**: 2 níveis de exposição da praia (10 , 11)

Sample	Beach	Richness	NAP	Praia	fExp
1	1	11	0.045	1	10
6	2	8	1.190	2	10
11	3	6	-0.976	3	11
16	4	1	1.768	4	11
21	5	3	1.117	5	10
26	6	5	-0.578	6	11
31	7	2	0.883	7	11
36	8	3	1.671	8	10
41	9	7	-0.356	9	10

LMM: riqueza na praia



LMM: riqueza da praia

Comparando efeitos aleatórios

```
lmm00s <- lmer(Richness ~ NAP * fExp + (NAP|Beach),
              data=praia)
lmm00i <- lmer(Richness ~ NAP * fExp + (1|Beach),
              data=praia)
anova(lmm00i, lmm00s, refit= FALSE)

## Data: praia
## Models:
## lmm00i: Richness ~ NAP * fExp + (1 | Beach)
## lmm00s: Richness ~ NAP * fExp + (NAP | Beach)
##          npar      AIC      BIC   logLik -2*log(L)  Chisq Df
Pr(>Chisq)
## lmm00i      6 236.49 247.33 -112.25      224.49
## lmm00s      8 237.13 251.59 -110.57      221.13 3.3594  2
0.1864
```

LMM: modelo mínimo adequado

Comparando LMM: **estrutura fixa**

- os modelos devem ser comparados por ML
- devem ser apresentados com REML

```
lmm01p <- lmer(Richness ~ NAP + fExp + NAP:fExp + (1|Beach), data=praia)
lmm02p <- lmer(Richness ~ NAP + fExp + (1|Beach), data=praia)
anova(lmm01p, lmm02p)
```

```
## Data: praia
## Models:
## lmm02p: Richness ~ NAP + fExp + (1 | Beach)
## lmm01p: Richness ~ NAP + fExp + NAP:fExp + (1 | Beach)
##      npar      AIC      BIC    logLik -2*log(L)  Chisq Df Pr(>Chisq)
## lmm02p     5 244.76 253.79 -117.38      234.76
## lmm01p     6 242.11 252.95 -115.06      230.11 4.6454  1    0.03114 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A interação NAP:fExp fornece informação relevante

LMM: resultado do modelo

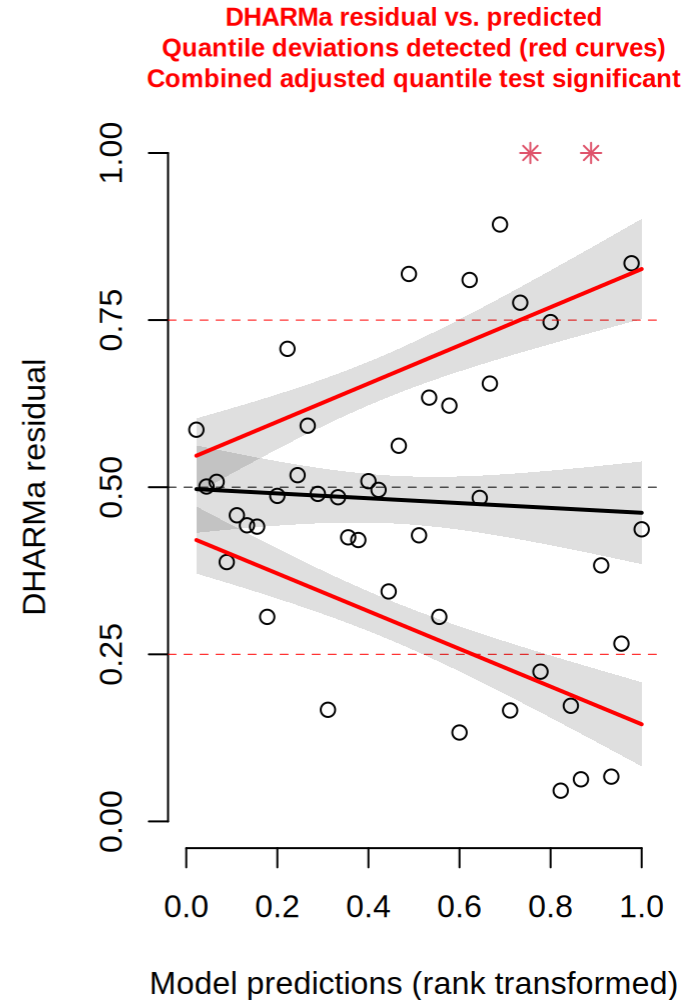
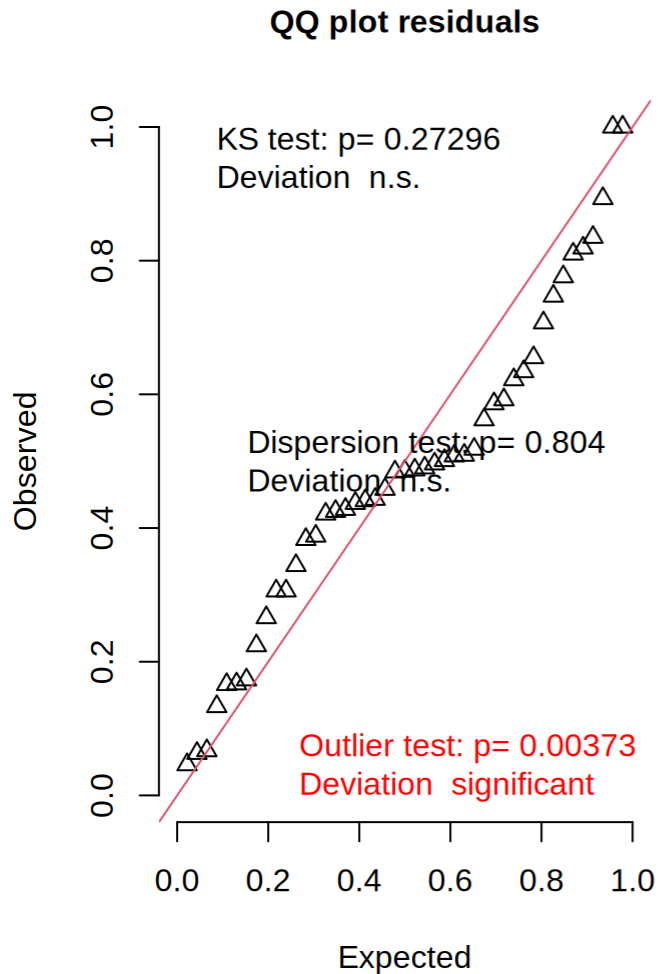
```
summary(lmm01p)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Richness ~ NAP + fExp + NAP:fExp + (1 | Beach)
## Data: praia
##
## REML criterion at convergence: 224.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.4849 -0.4161 -0.0770  0.1521  3.7313
##
## Random effects:
##   Groups      Name              Variance Std.Dev.
##   Beach      (Intercept)    3.307      1.819
##   Residual                        8.660      2.943
## Number of obs: 45, groups: Beach, 9
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   8.8611     1.0208   8.680
## NAP          -3.4637     0.6279  -5.517
## fExp11        -5.2556     1.5452  -3.401
## NAP:fExp11     2.0005     0.9461   2.114
##
```

```
## Correlation of Fixed Effects:
##          (Intr)  NAP      fExp11
```

LMM: diagnóstico do modelo

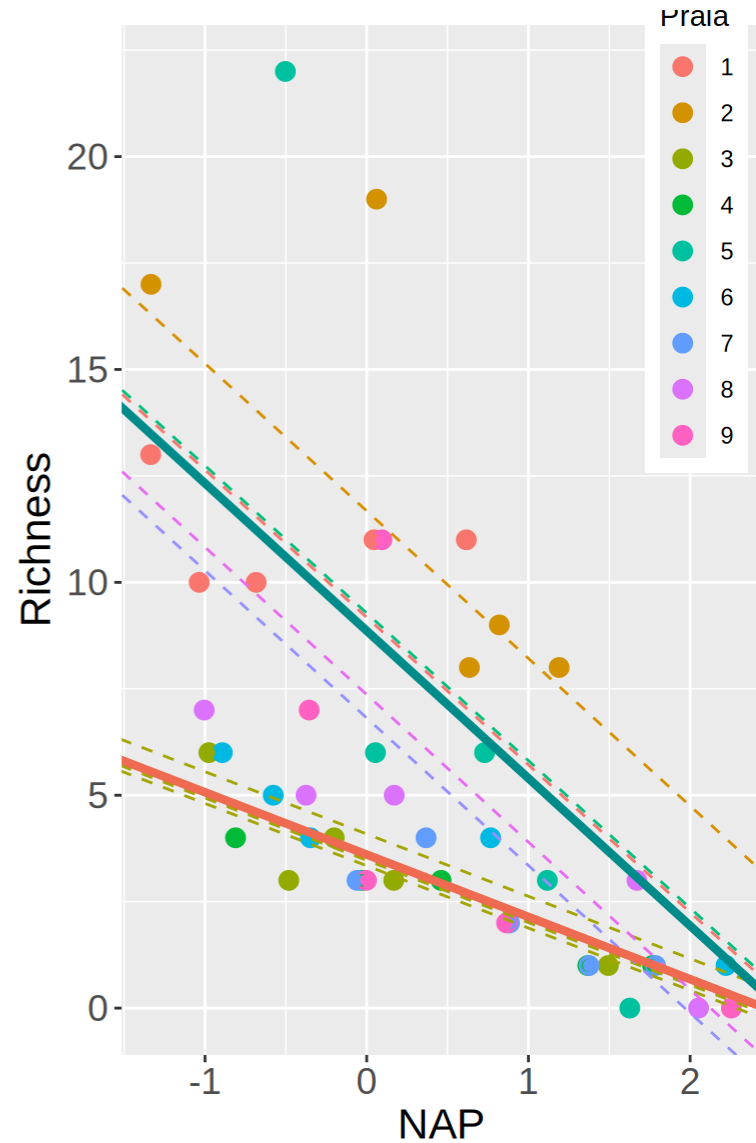
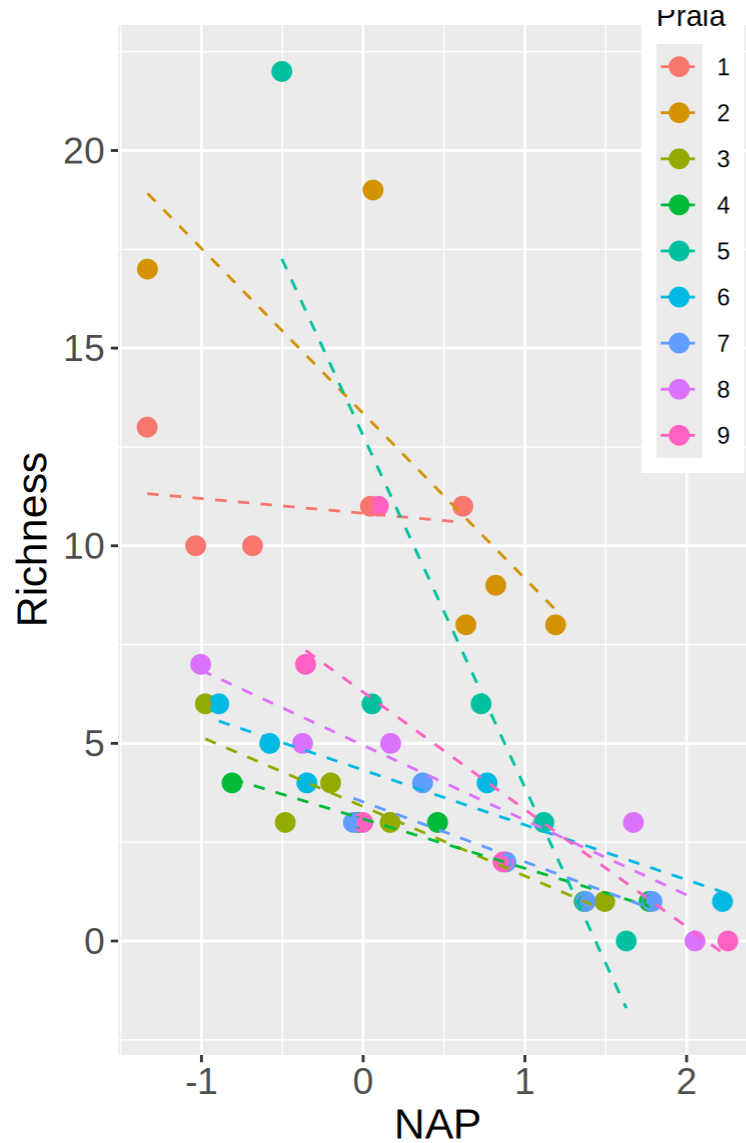
DHARMA residual



LMM: resultado do modelo

```
lmmNull <- lmer(Richness ~ (1|Beach), data=praia)  
anova(lmmNull, lmm01p)
```

LMM: resultado do modelo



Linear Mixed Models: FIM