

Introdução aos Modelos Lineares em Ecologia

Prof. Adriano Sanches Melo - Dep. Ecologia – UFG
asm.adrimelo no gmail.com

Página do curso: www.ecologia.ufrgs.br/~adrimelo/lm/

Livro-texto: Crawley, M.J. 2005. Statistics: An Introduction using R.
John Wiley & Sons.

Página do livro na internet:

<http://www3.imperial.ac.uk/naturalsciences/research/statisticsusingr>

AULA 5

1. Planejamento e Desenho Experimental

“We emphasize that the proper design of a scientific study is far more important than the specific techniques used in the analysis. As we shall see, a well-designed study is usually simple to analyze. On the other hand, a poorly designed study or a botched experiment often cannot be salvaged, even with the most sophisticated analysis.”

Kutner et al. (2004)

[“The wide subject of experimental design was opened up, aimed at ... avoiding waste of effort in the accumulation of ill-planned, indecisive, or irrelevant observations.” **Fisher (1955)**

“The task now is to prevent it from being closed!” **Underwood (1997)]**

2. Estudos experimentais, observacionais e relação causa-e-efeito

2.1 Estudos Experimentais

Exemplo: Vitamina C previne gripe?

2.2 Estudos observacionais

Difere de estudos experimentais pois em geral não é possível aleatorizar o tratamento que determinada unidade experimental recebe.

Exemplo: Desempenho de professores que participaram e que não participaram de seminário sobre técnicas de ensino. Não era obrigatório; fez quem quis. É possível sugerir relação causa-e-efeito inequivocamente?

3. Conceitos básicos de estudos experimentais

3.1) Fator: a variável explanatória ou preditora ou independente. Em experimentação é comum ser categórica. O termo “fator” em geral é empregado para designar uma variável categórica, embora existam exceções. No R, o comando para transformar para um fator (categórico) é `factor()`.

3.2) Níveis de um fator: os estados que o fator pode assumir. No caso do exemplo no item 2.2 acima, tínhamos 2 níveis: professores que assistiram e os que não assistiram ao seminário.

3.3) Controles: Muitas vezes necessários, depende do estudo. Serve para assegurar que o *procedimento* experimental não interferiu com resultados (e.g. placebo) e também como base comparativa (e.g. perturbação alterou o que era antes [estado antes = controle]).

3.4) Unidades experimentais: “Uma unidade experimental é a menor unidade do experimento ao qual se pode designar um tratamento; unidades experimentais são portanto definidas pelo método de aleatorização”.

A idéia aqui é fundamental em estudos ecológicos!

Exemplo: Suponha um pesquisador queira estudar o efeito de fósforo em lagoas. Ele acredita que mais fósforo pode ocasionar eutrofização.

Experimento correto: Tratar como unidade experimental “lagoas”. Ele seleciona 10 lagoas; sorteia 5 e aplica fósforo e nas outras 5 deixa como está. Em cada lagoa coleta 1 *unidade amostral* (10 minutos com rede de plâncton)

Experimento errado: Ele seleciona 2 lagoas, sorteia 1 e aplica fósforo e na outra deixa como está. Em cada lagoa coleta 5 *unidades amostrais* (10´ rede plâncton).

Veja que os *tratamentos são aplicados em lagoas*. No caso do experimento errado, tem-se na verdade 1 réplica para cada tratamento com 5 unidades amostrais por réplica. O que importa na análise de um experimento é o número de réplicas. O número de unidades amostrais dependerá da representatividade desta(s). Em geral, quanto mais homogênea a unidade experimental, menor o esforço de amostragem (ou seja menor o número, área, volume das unidades amostrais).

Tente identificar em seu estudo quem é o objeto de estudo (= sua unidade experimental). Leia o texto (clássico!) de Hulbert (1984; disponível no sítio da disciplina) sobre pseudo-replicação.

3.5) Tamanho amostral e replicação: Maior o esforço amostral, maior a probabilidade de detectarmos um *efeito* significativo se ele existir. De forma simplificada, *efeito* = diferença na média dos tratamentos.

Usa-se o termo *poder do teste* para designar o quão adequado/poderoso um teste é para detectar um *efeito*, caso ele exista.

Maior a replicação, maior o poder do teste;

Maior o efeito, maior o poder do teste;

Menor a variância *dentro* dos tratamentos, maior o poder do teste;

Maior o nível de significância, maior o poder do teste.

3.6) Aleatorização: Enfatizada por R.A. Fisher. Uma das etapas mais importantes num experimento é a designação aleatória das unidades experimentais aos tratamentos.

Exemplo: Desempenho de professores que assistiram e os que não assistiram a seminário de didática.

O estudo pode ter sido *confundido*?

-- Um melhor desempenho de professores que assistiram ao seminário pode ser explicado pela maior motivação/interesse destes. Em geral, bons professores são mais motivados e portanto procuram se qualificar ainda mais. Neste caso, se o teste fosse aplicado antes do seminário talvez os resultados tivessem sido os mesmos. Aqui o efeito de confundimento era relativamente simples; o problema em situações reais é que sempre podem existir outros fatores de confundimento que não detectamos. A aleatorização é a solução para termos segurança em nossos resultados. Se houverem fatores de confundimento, estes serão alocados de maneira similar aos tratamentos e tenderão portanto a se contrabalancearem.

4. Relação entre Regressão e Análise de Variância

Análise de Variância difere de Regressão em dois aspectos:

1. A variável preditora pode ser qualitativa

2. Se a variável preditora é qualitativa, não existe a necessidade de se especificar o modelo mais adequado (e.g. linear ou quadrático?). Em situações em que se pode usar tanto Regressão como Anova, esta última será mais adequada se os dados não forem lineares ou não puderem ser “linearizados” com uma transformação (ver Cottingham et al. 2005 na página do curso).

5. Modelos de Anova

Tipo I ou de Efeitos Fixos. A situação mais comum. No caso dos níveis diferirem, pode-se fazer testes (*a priori* ou *a posteriori*) para distingui-los. É o tipo de Anova que estudaremos hoje e, salvo alguma exceção a ser notada, usaremos no restante do curso.

Exemplos:

Tipo II ou de Efeitos Aleatórios. Os níveis são amostras de possíveis níveis. Neste caso, outros níveis poderiam ter sido escolhidos. No caso de haver diferenças entre níveis, não se faz testes para descobrir quem é diferente de quem. Não existe razão para tais testes.

Exemplos:

-- Áreas diferem quanto a densidade da sp1 ?

6. Partição de Variância

Y_{ij} = Observação j do tratamento i

$Y_{..}$ = Média Geral

$Y_{i.}$ = Média do tratamento i

$$Y_{ij} - \bar{Y}_{..} = \bar{Y}_{i.} - \bar{Y}_{..} + Y_{ij} - \bar{Y}_{i.}$$

SSTO

SSR

SSE

$$SSTO = \sum (Y_{ij} - \bar{Y}_{..})^2$$

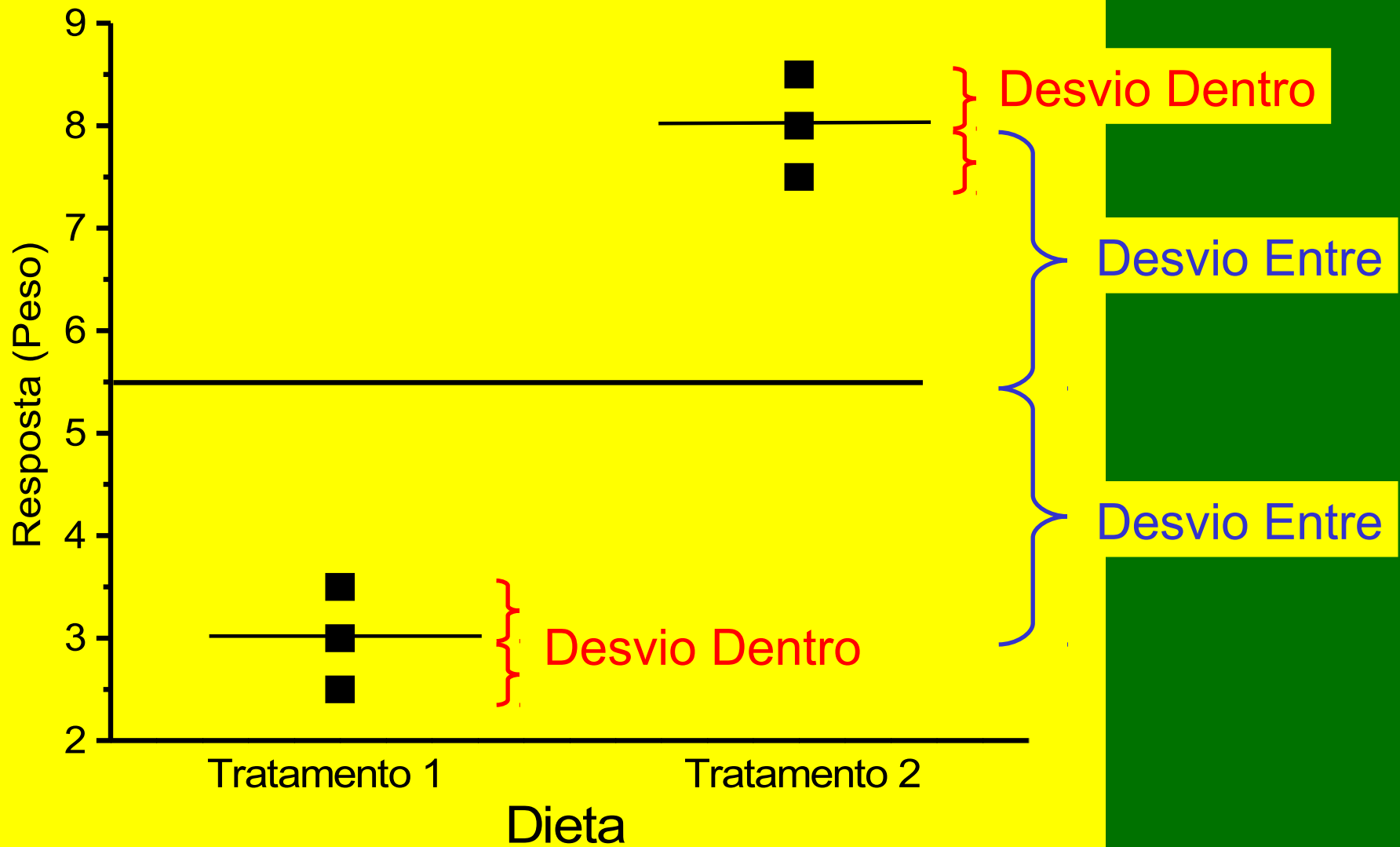
→ Variação Total

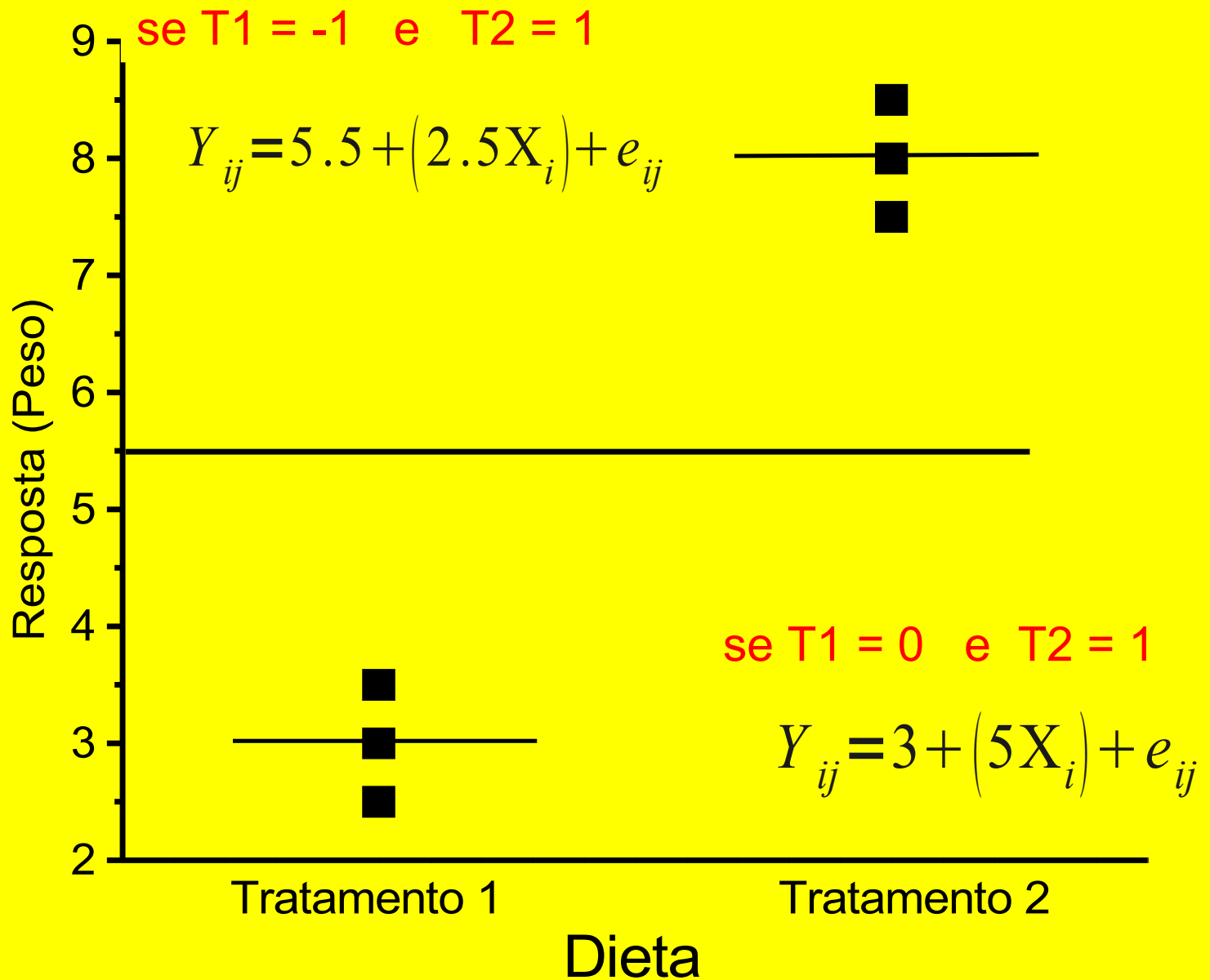
$$SSTR = \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

→ Desvio da média do tratamento em relação à média geral (Desvio Entre)

$$SSE = \sum (Y_{ij} - \bar{Y}_{i.})^2$$

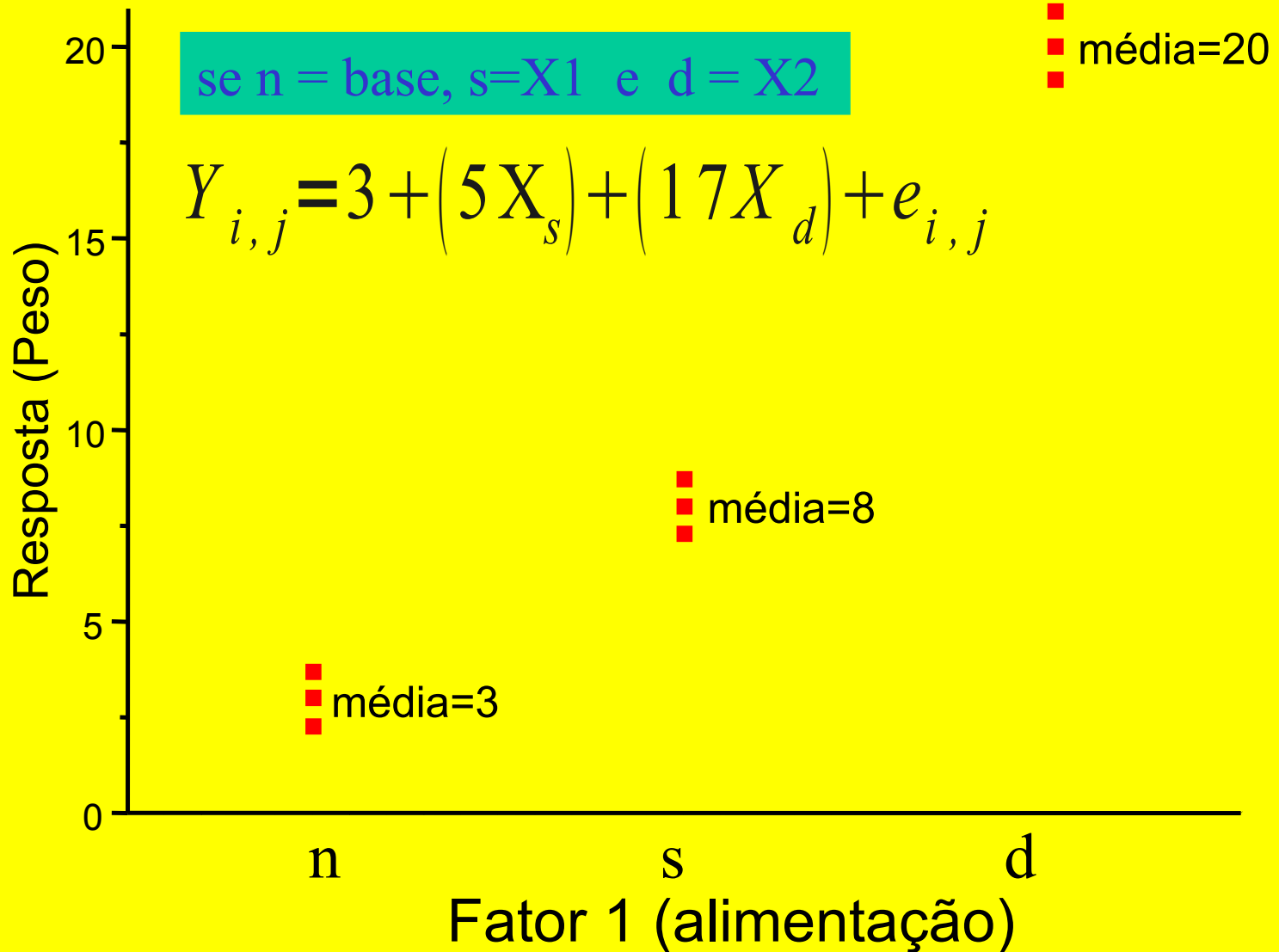
→ Desvio da observação em relação à média do tratamento (Desvio Dentro)





$$Y_{i,j} = \mu + T_i + e_{i,j}$$

X	Y
n	2.5
n	3
n	3.5
s	7.5
s	8
s	8.5
d	19.5
d	20
d	20.5



Note que numa Análise de Variância não estamos interessados nos coeficientes de cada variável *dummy* ou indicadora. Nossa pergunta é: existe (qualquer) diferença entre as médias?

Para Análise de Variância usamos portanto apenas a partição de variâncias e o teste F, obtidos com o comando `aov(y~x)`

10. Partição dos Graus de Liberdade

$$SSTO = n_T - 1$$

$$SSR = r - 1$$

$$SSE = n_T - r$$

onde,

SSR = Soma de Quadrados dos Tratamentos (ou do modelo).

n_T = número total de unidades experimentais.

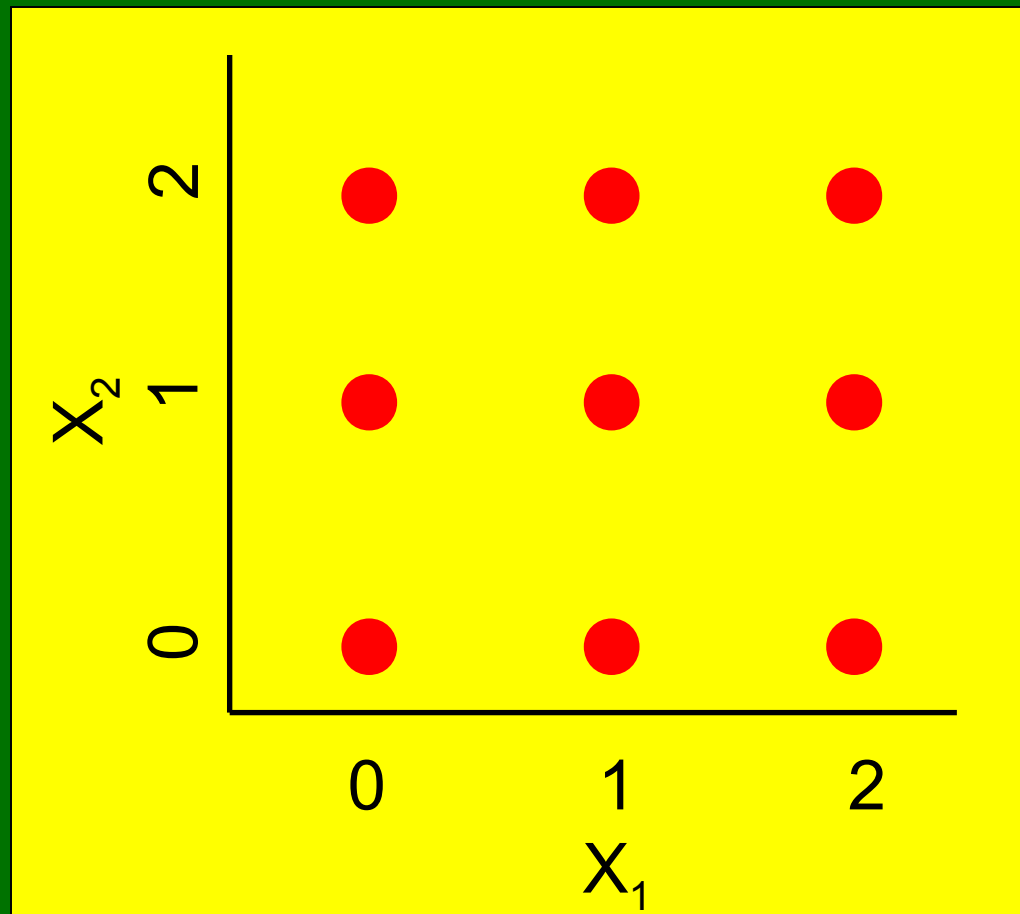
r = número de níveis do fator em estudo.

Anova II

1. Estudos Fatoriais

Vantagens Estudos Fatoriais

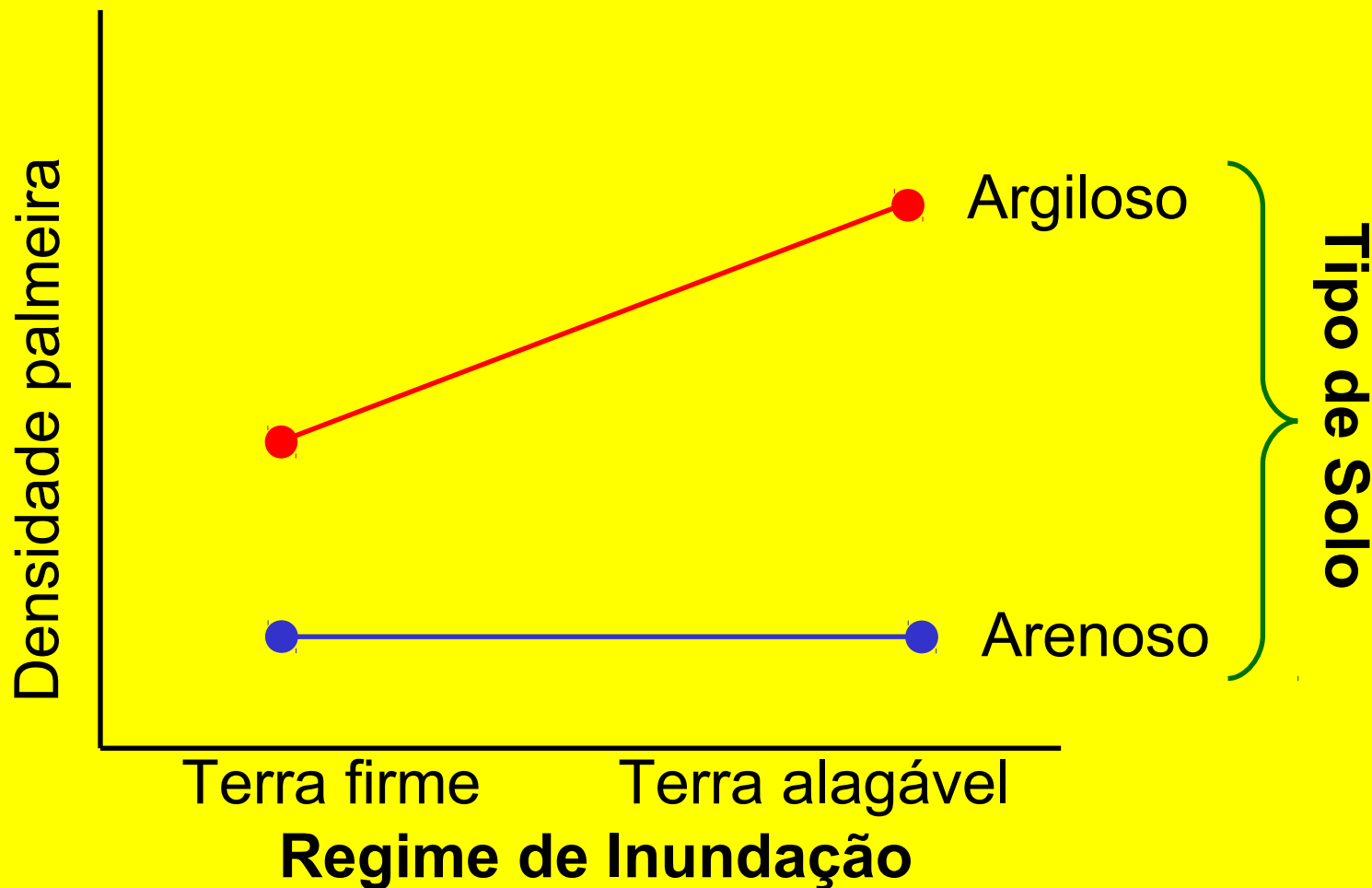
1.1 Replicações “escondidas”



1. Estudos Fatoriais

Vantagens Estudos Fatoriais

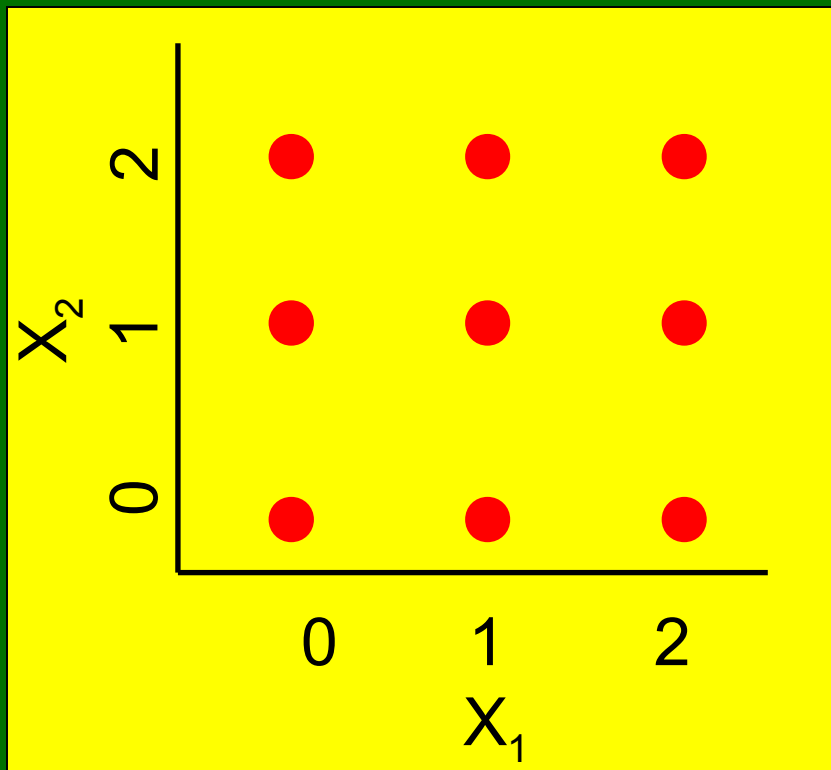
1.2 Avaliação de Interações



2. Estudos Fatoriais e Não Fatoriais

Análises fatoriais

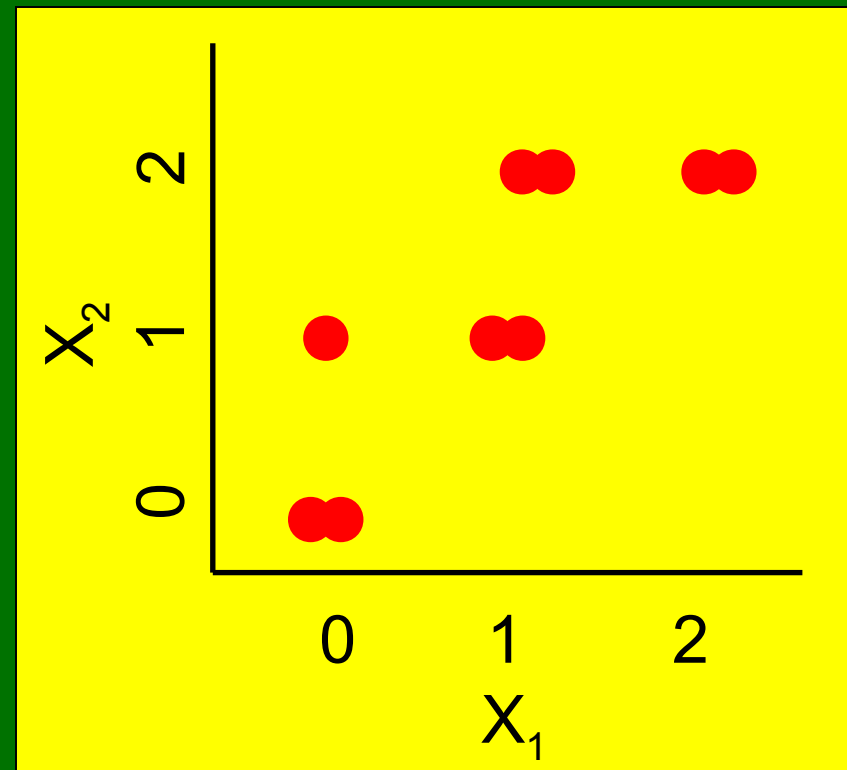
Correlação entre variáveis preditoras é '0'



Análises não fatoriais

Correlação entre variáveis preditoras é diferente de '0'

Problemas com multicolinearidade ...



3. Comandos no R:

```
aov (y ~ x1 + x2 + x1:x2)
```

ou

```
aov (y ~ x1*x2)
```

4. Blocos ao acaso

Experimento feito por Flecker (1996; exercício da aula 6). Avaliação dos efeitos do peixe *Prochilodus mariae* sobre invertebrados, diatomáceas e algas filamentosas.

Suponha que o rio é bastante heterogêneo e isto cause muita variação indesejada dentro dos grupos. Ou seja, poderíamos desconfiar de que a fauna entre corredeiras seja distinta.

No planejamento, poderíamos incluir blocos para controlar tal variação.

Neste caso, temos 3 níveis do fator de estudo (Co = controle; Ce2 = Controle de procedimento; Ce4 gaiola de exclusão com 4 lados).

O trabalho é feito por corredeiras, cada corredeira sendo 1 bloco. Note que em cada corredeira existe uma réplica de cada tratamento.

Poderíamos ter mais de uma réplica por corredeira. Note também que a posição do tratamento dentro da corredeira é aleatória.

Unidades experimentais
1 2 3

Bloco 1



Co Ce2 Ce4

Bloco 2



Ce2 Ce4 Co

Bloco 3



Ce4 Ce2 Co

Bloco 4

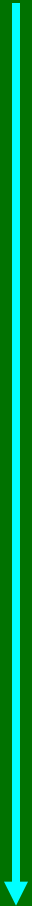


Ce2 Co Ce4

Bloco 5

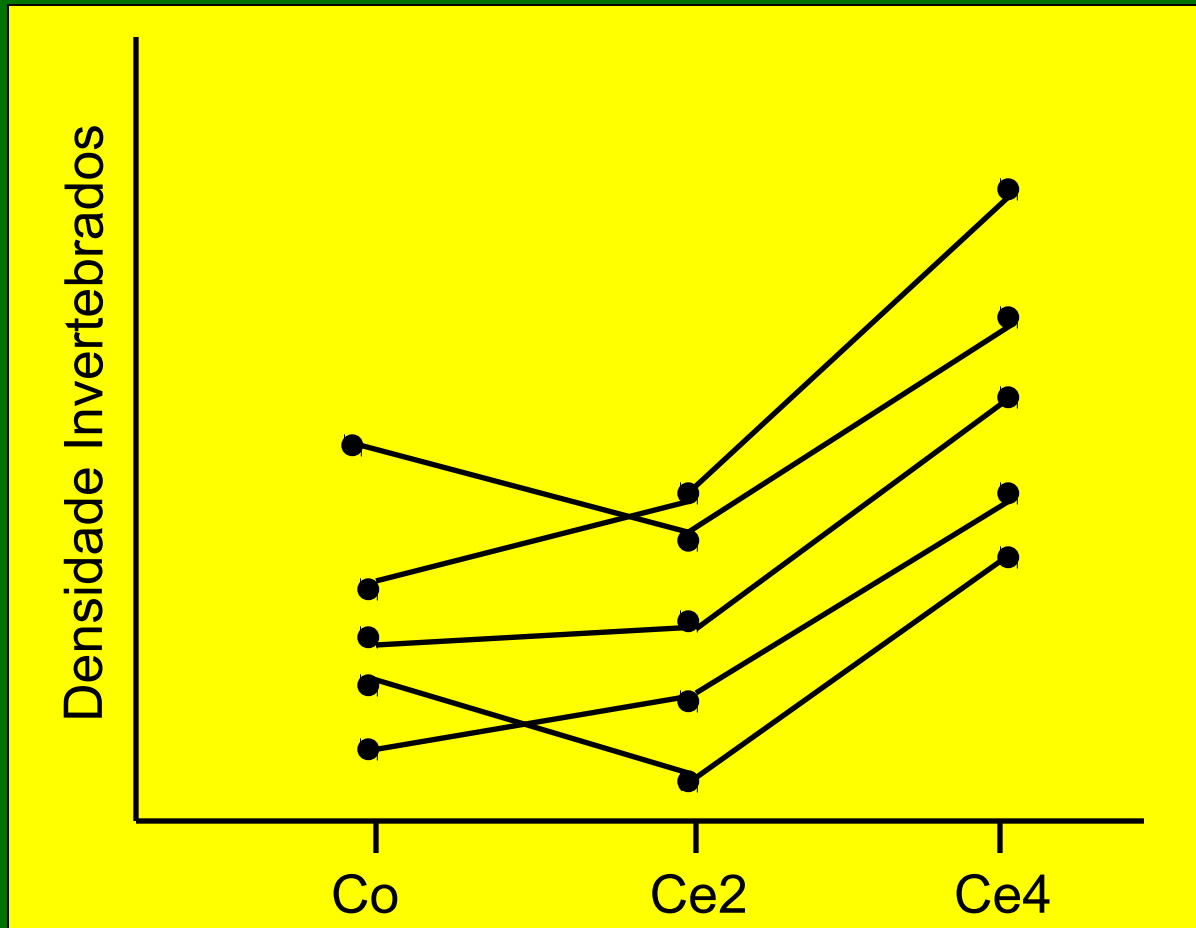


Co Ce4 Ce2



Fluxo

Possível resultado



Neste caso o uso de blocos é bastante útil. Se analisássemos como 3 amostras independentes (Anova com 1 fator e 3 níveis) talvez não detectaríamos uma diferença significativa, pois existe muita variabilidade dentro de cada tratamento. Note sobreposição do tratamento Ce4 com os outros dois.

5. Comentários sobre blocos

Note que o modelo para análise de variância de 1 fator + bloco é parecido com o modelo de análise de variância com dois fatores:

$$\text{aov } (y \sim x1 + x2)$$

Mas com duas diferenças:

5.1 O modelo linear para desenhos em blocos não inclui interação. Este é um pressuposto que deve ser feito. Nem todos os autores concordam com isto...

5.2 Não avaliamos a significância dos blocos. A rigor, sabíamos a priori que existiam variações e daí a decisão de usar blocos. Não faz sentido portanto avaliar a significância de algo que já sabíamos antes do experimento e que a rigor não faz parte da hipótese do trabalho.

Discutimos aqui o uso de uma variável bloco. Podemos ter mais. Pode-se imaginar que o experimento com *Prochilodus mariae* poderia ser feito em dois riachos. Riacho portanto seria a segunda variável bloco.

6. Exemplo: 2-Anova

```
ajuste<-aov(dens~solo*alag)
```

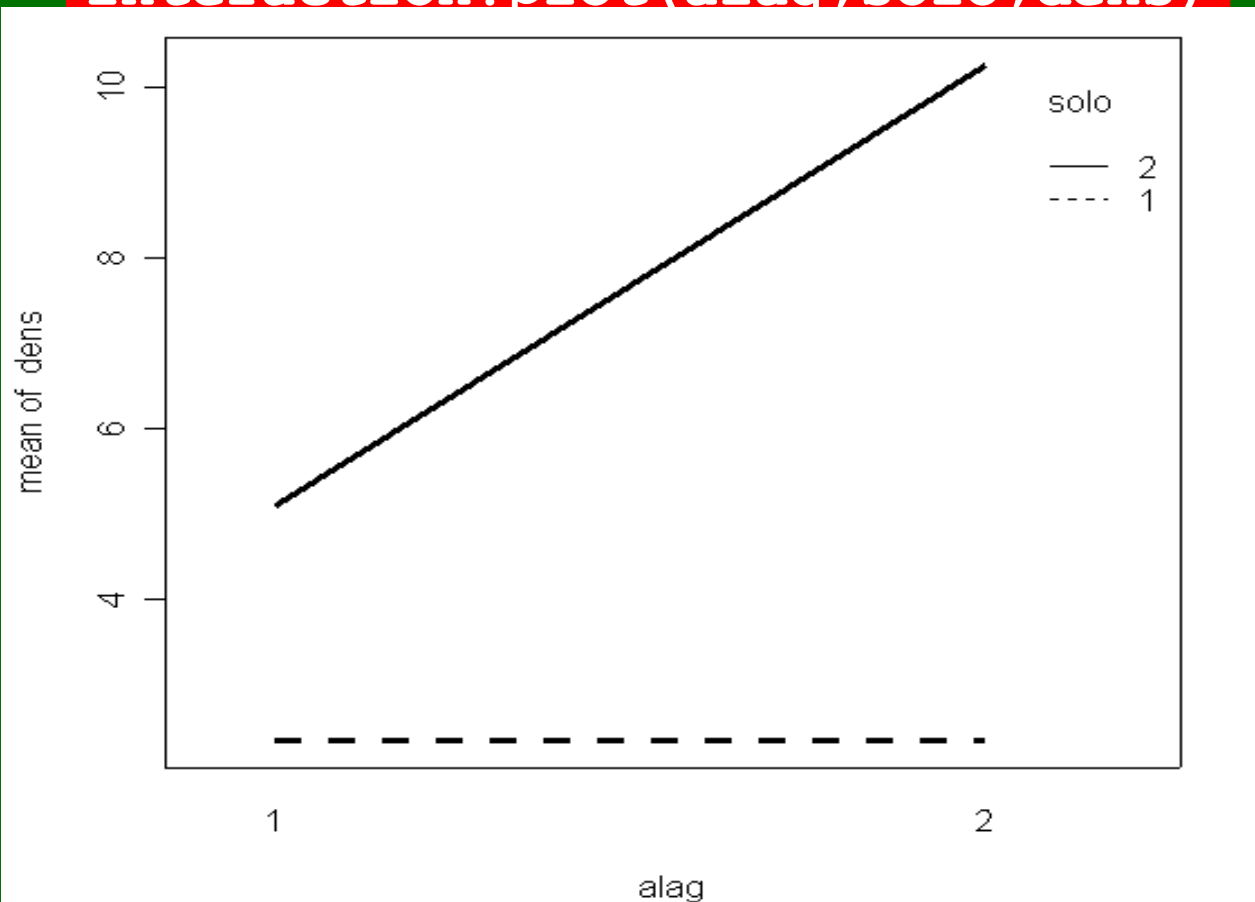
```
summary(ajuste)
```

```
plot(ajuste)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
solo	1	54.601	54.601	693.35	1.236e-05
alag	1	12.251	12.251	155.57	0.0002376
solo:alag	1	13.261	13.261	168.40	0.0002035
Residuals	4	0.315	0.079		

```
interaction.plot(alag,solo,dens)
```

dens	solo	alag
2.3	1	1
2.6	1	1
2.2	1	2
2.5	1	2
4.8	2	1
5.4	2	1
10.3	2	2
10.0	2	2

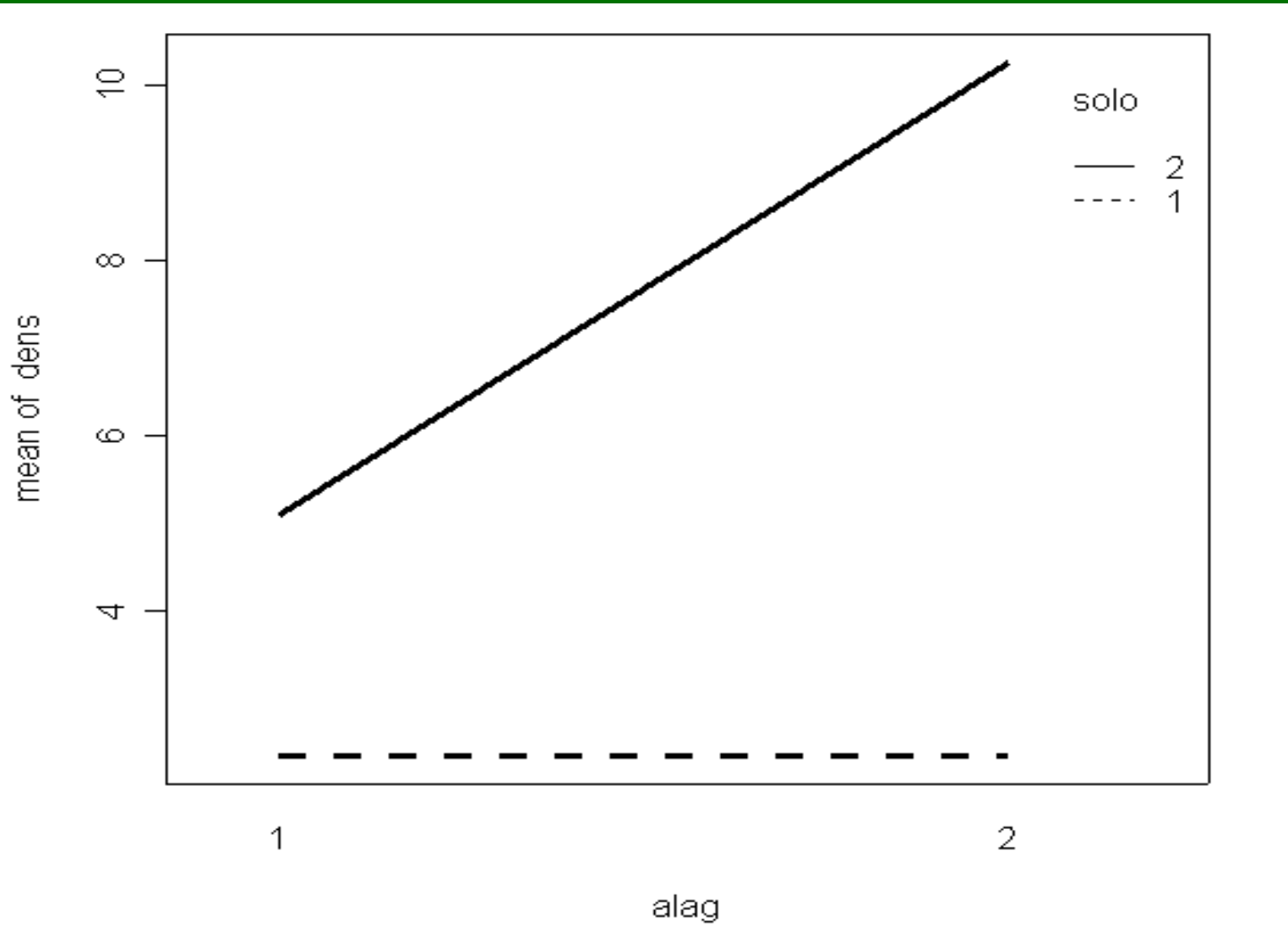


6. Exemplo:

Interpretação

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
solo	1	54.601	54.601	693.35	1.236e-05
alag	1	12.251	12.251	155.57	0.0002376
solo:alag	1	13.261	13.261	168.40	0.0002035
Residuals	4	0.315	0.079		

dens	solo	alag
2.3	1	1
2.6	1	1
2.2	1	2
2.5	1	2
4.8	2	1
5.4	2	1
10.3	2	2
10.0	2	2



Exercícios e estudo individual:

- Lista em sala de aula
- Crawley: Cap. 9 (pp. 155-175)
- Gotelli & Ellison: Cap. 10 (pp. 289-348)