

Universidade Federal de Goiás

Bioestatística

Prof. Adriano Sanches Melo - Dep. Ecologia – ICB
asm.adrimelo@gmail.com

Página do curso:
<http://www.ecologia.ufrgs.br/~adrimelo/bioestat>

Teste t

Aulas anteriores (regressão linear):

O que é um modelo?

A equação da reta

Desvios (observado - esperado)

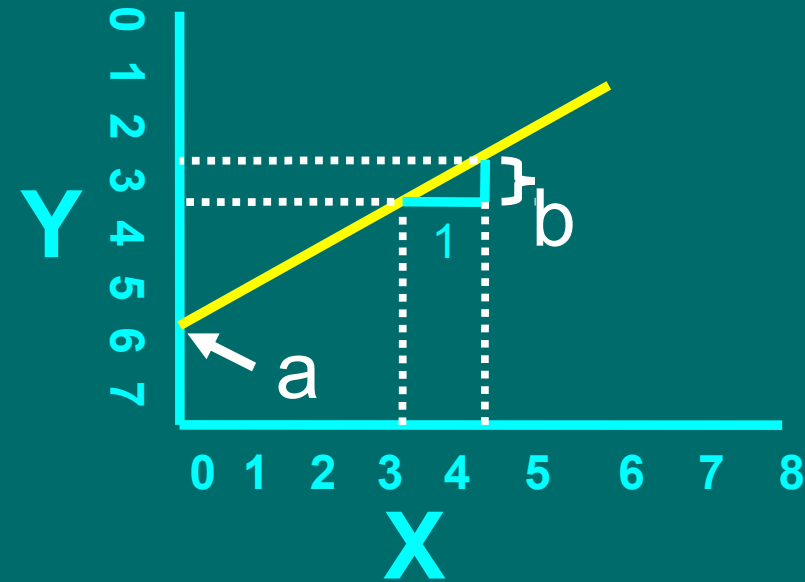
Como encontrar os coeficientes 'a' e 'b' da reta?

Iteração (tentativa-e-erro) ou fórmulas

Critério para encontrar melhor reta:

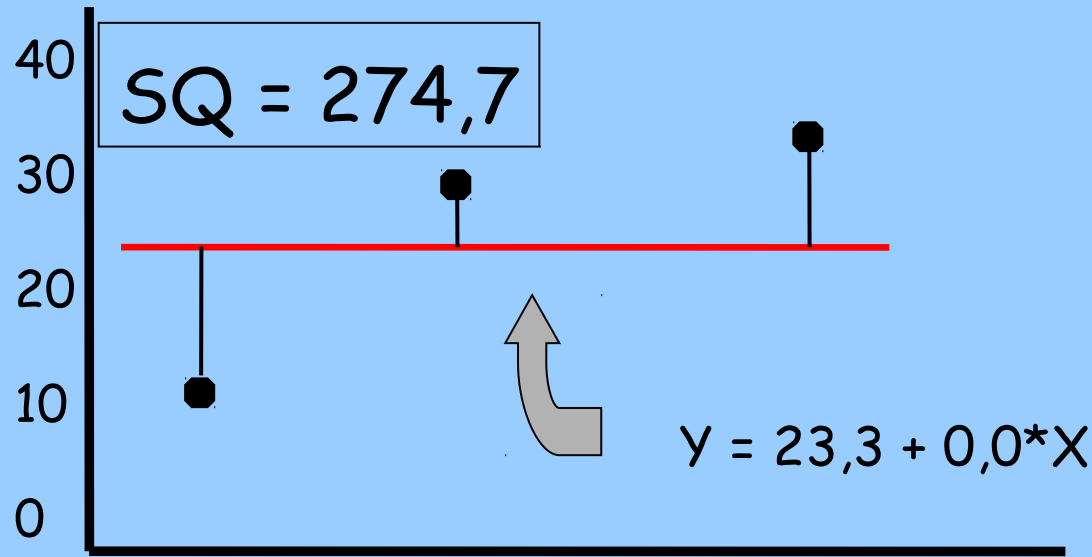
A menor soma de quadrados

A tabela de análise de variância

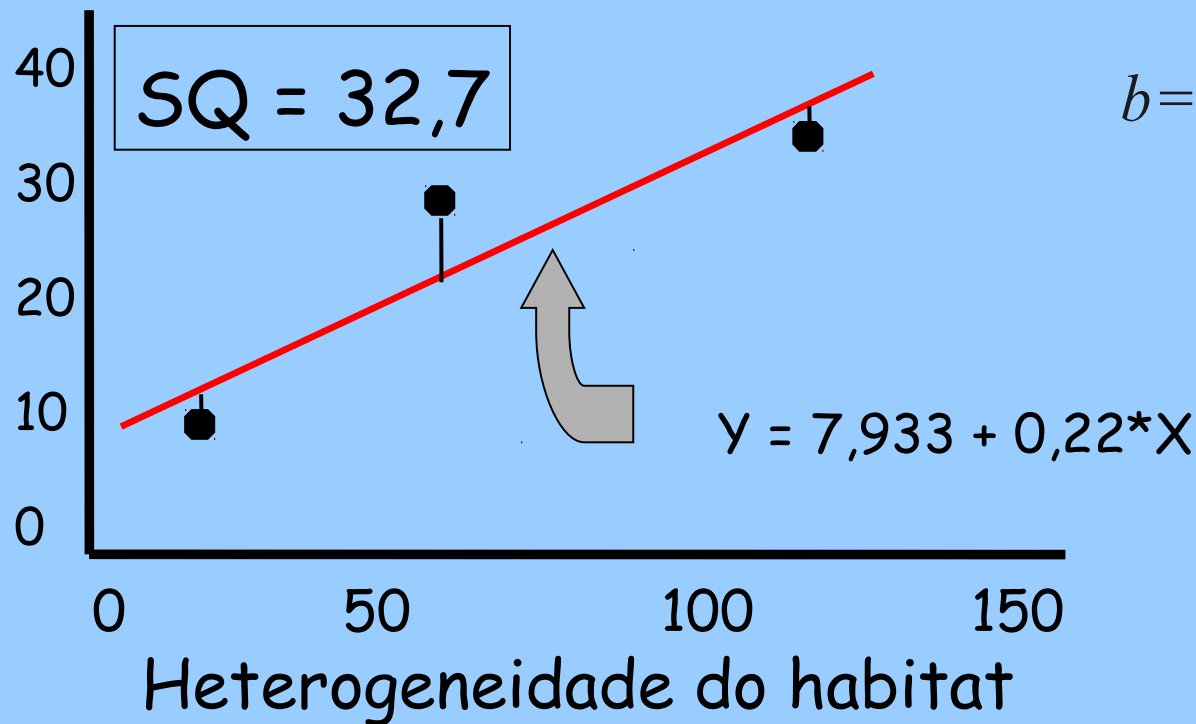


Hoje: Mais um modelo --> teste t

Riqueza de espécies



$Y_1 = 10$	$X_1 = 20$
$Y_2 = 28$	$X_2 = 70$
$Y_3 = 32$	$X_3 = 120$



$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$a = \bar{Y} - b\bar{X}$$

Teste t --> Comparação de duas médias

As alturas de homens e mulheres na UFG são diferentes?

Alunos submetidos às técnicas de ensino A e B diferem quanto a aprendizagem?

A taxa de herbivoria na planta X depende desta estar no sol ou na sombra?

Indivíduos de uma espécie de planta produzem quantidades distintas de sementes quando estão ou não em associação com formigas?

As notas dadas por consumidores à duas massas de pizza são diferentes?

Qual o modelo?

Em regressão linear, vimos que nosso modelo era o de uma reta:

Simplificando (retirando a constante) --> $Y \sim X$

onde, Y = variável resposta contínua

X = variável explanatória contínua

No teste t, nosso modelo é: $Y \sim X$

onde, Y = variável resposta contínua

X = variável explanatória categórica com 2 níveis

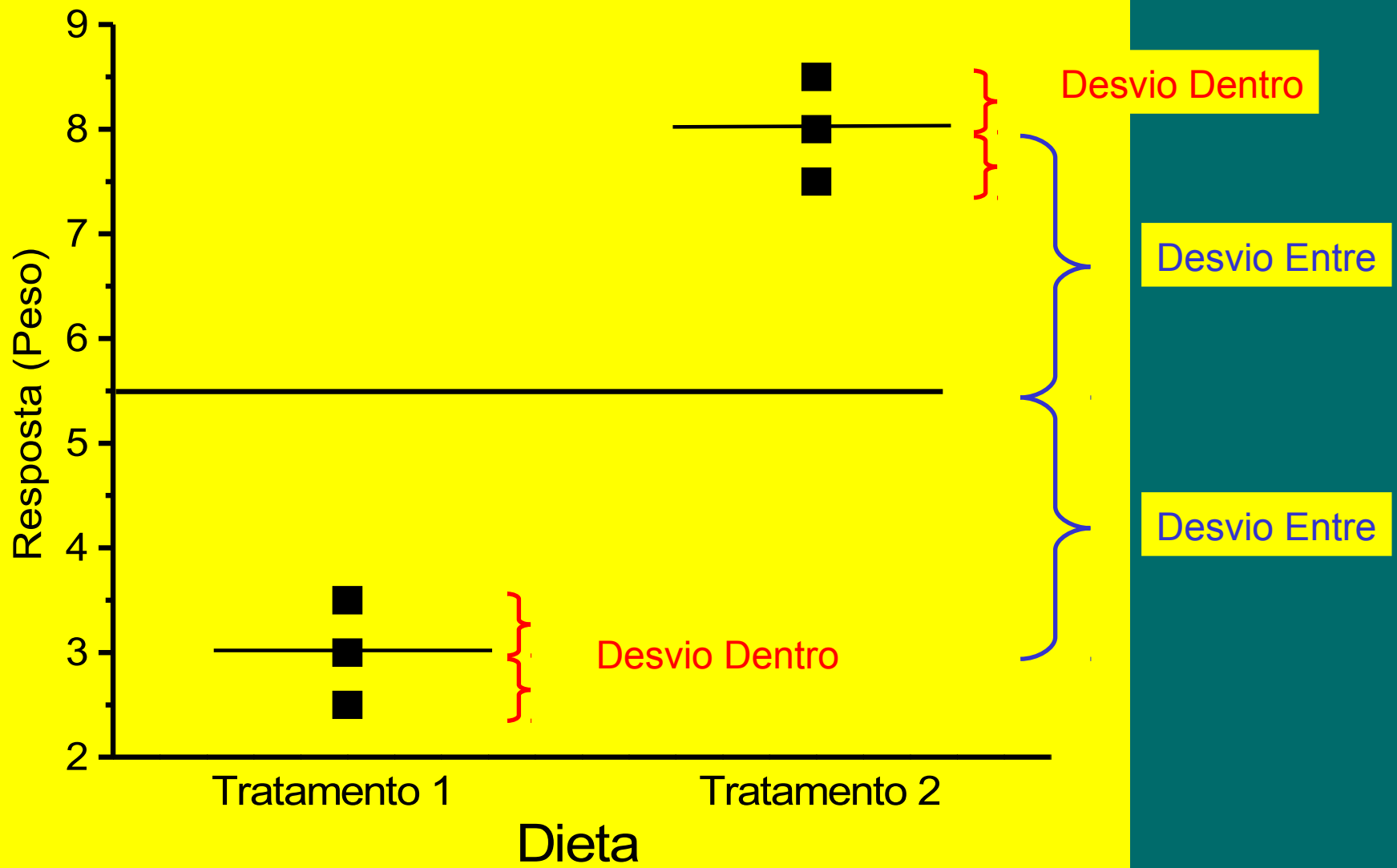
Valores ajustados:

Regressão: Valores sobre a reta

Teste t: Médias de cada nível da variável explanatória

Então é mais simples que regressão linear ???

SIM!!!



Novamente, a tabela de análise de variância!

Média F: 21,5

Média M: 24,25

Média Geral: 22,875

Idade	Sexo
23	F
21	F
22	F
20	F
24	M
22	M
25	M
26	M

SQT	$Y - \bar{Y}$	e	e^2
	$23 - 22,875 =$	0,125	0,015
	$21 - 22,875 =$	-1,875	3,515
	$22 - 22,875 =$	-0,875	0,765
	$20 - 22,875 =$	-2,875	8,265
	$24 - 22,875 =$	1,125	1,265
	$22 - 22,875 =$	-0,875	0,765
	$25 - 22,875 =$	2,125	4,515
	$26 - 22,875 =$	3,125	<u>9,765</u>
			28,875

SQE	$Y - \hat{Y}$	e	e^2
	$23 - 21,5 =$	1,5	2,25
	$21 - 21,5 =$	-0,5	0,25
	$22 - 21,5 =$	0,5	0,25
	$20 - 21,5 =$	-1,5	2,25
	$24 - 24,25 =$	-0,25	0,06
	$22 - 24,25 =$	-2,25	5,06
	$25 - 24,25 =$	0,75	0,56
	$26 - 24,25 =$	1,75	<u>3,06</u>
			13,75

A Tabela de Análise de Variância ou Partição de Variância

Fonte	SQ	g.l.	QM	F	p
Regressão (SQR)	15,125	1	15,125	6,6	0,04
Erro (SQE)	13,75	6	2,29		
Total (SQT)	28,875	n-1			

A Tabela de Análise de Variância ou Partição de Variância

Fonte	SQ	g.l.	QM	F	p
Regressão (SQR)	15,125	1	15,125	6,6	0,04
Erro (SQE)	13,75	6	2,29		
Total (SQT)	28,875	n-1			

Ou podemos usar a fórmula tradicional:

$$t^2 = F$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_p^2}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Usando a fórmula tradicional:

$$t^2 = F$$

Média F: 21,5

Média M: 24,25

Média Geral: 22,875

Var F = 1,66

Var M = 2,92

Variância ponderada = $\frac{(4-1)*1,66 + (4-1)*2,92}{4 + 4 - 2} = \frac{4,98 + 8,76}{6} = \frac{13,74}{6} = 2,29$

$t = 24,25 - 21,5 / \text{sqrt}((1/4) + (1/4) * 2,29) = 2,75 / 1,07 = 2,57$

$2,57^2 = 6,6$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_p^2}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Exercício



1. Introdução

2. Objetivos

3. Metodologia

4. Resultados

5. Conclusão

6. Referências

7. Anexos

8. Bibliografia

9. Conclusão

10. Referências

11. Anexos

12. Bibliografia

13. Conclusão

14. Referências

15. Anexos

16. Bibliografia

17. Conclusão

18. Referências

19. Anexos

Estudos experimentais, observacionais e relação causa-e-efeito

Estudos Experimentais

Exemplo: Vitamina C previne gripe?

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?

Conceitos básicos de estudos experimentais

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.

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.

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]).

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!

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).

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.

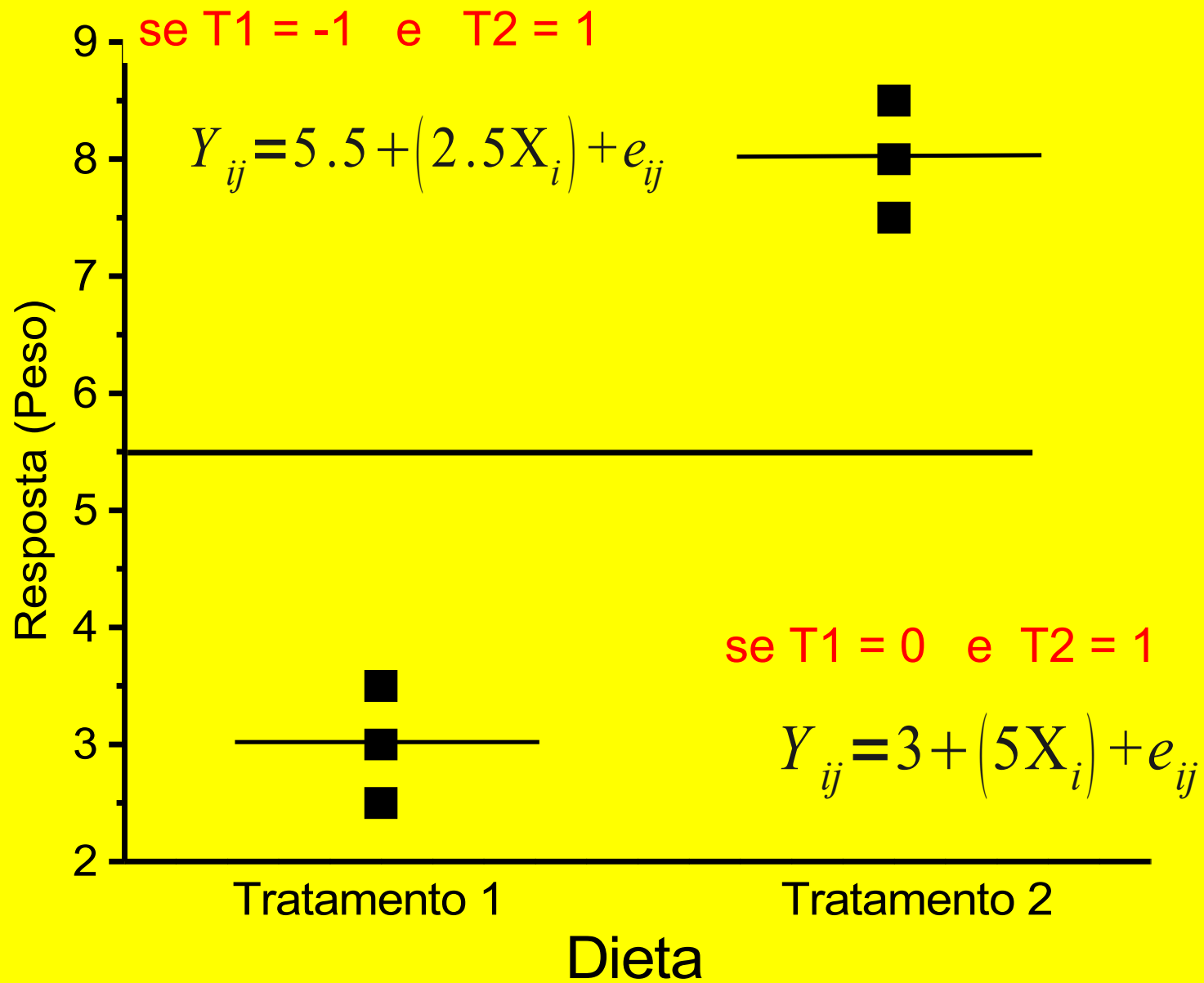
Usa-se o termo *poder do teste* para designar o quão adequado/poderoso um teste é para detectar um *efeito*, caso ele exista.

Maior a replicação, maior o poder do teste;

Maior o efeito, maior o poder do teste;

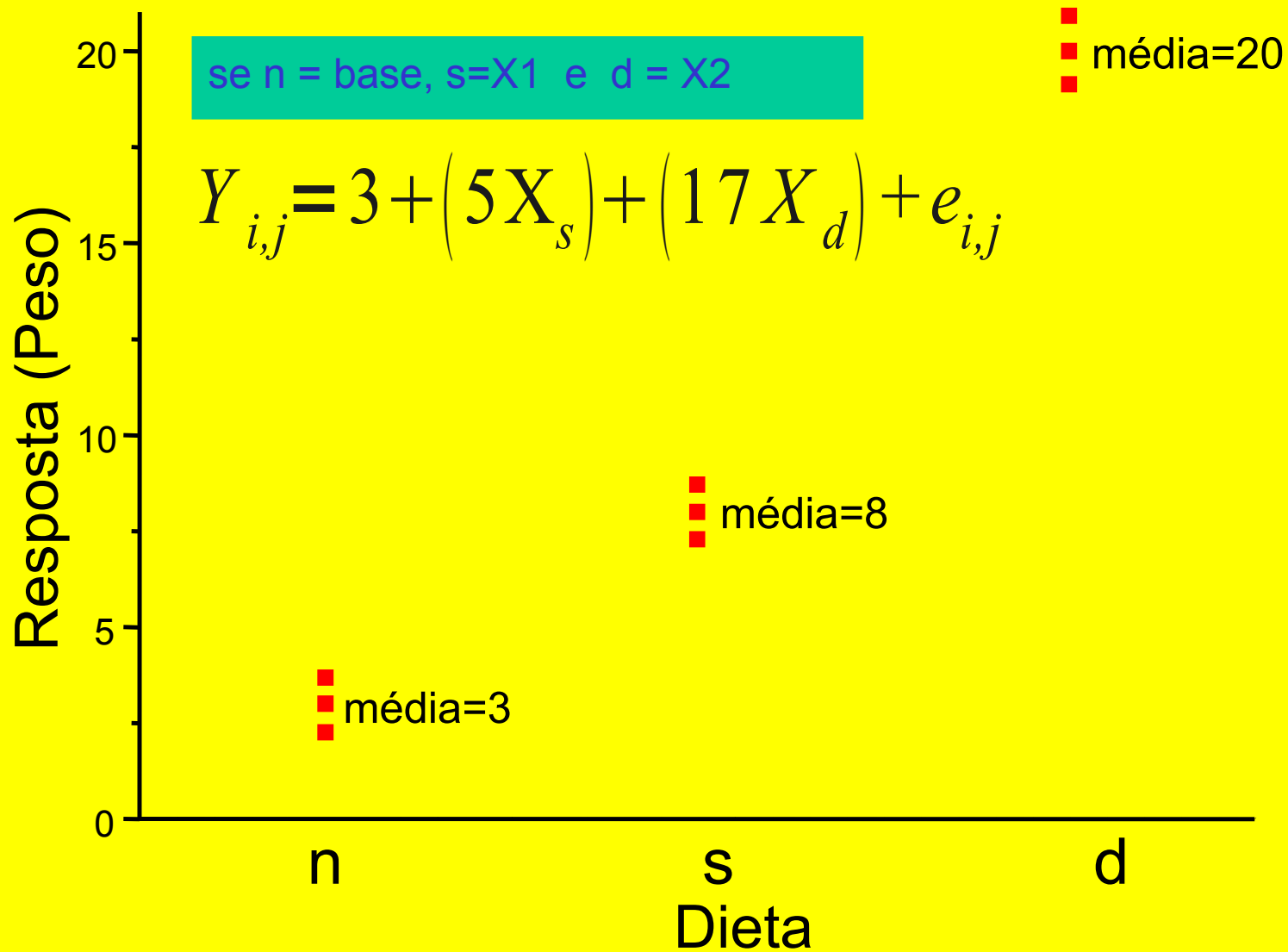
Menor a variância *dentro* dos tratamentos, maior o poder do teste;

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



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

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?

Partição dos Graus de Liberdade

$$SQTO = n_T - 1$$

$$SQR = r - 1$$

$$SQE = n_T - r$$

onde,

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

n_T = número total de unidades experimentais.

r = número de níveis do fator em estudo.

Sugestão de estudo:

Vieira, S. 2008. Introdução à Bioestatística. 4a ed. Elsevier. (Cap. 13)