

LGN5830 - BIOMETRIA DE MARCADORES GENÉTICOS

AULA 2: MAPAS GENÉTICOS I

Antonio Augusto Franco Garcia
Roland Vencovsky

Departamento de Genética
ESALQ/USP
2007

CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

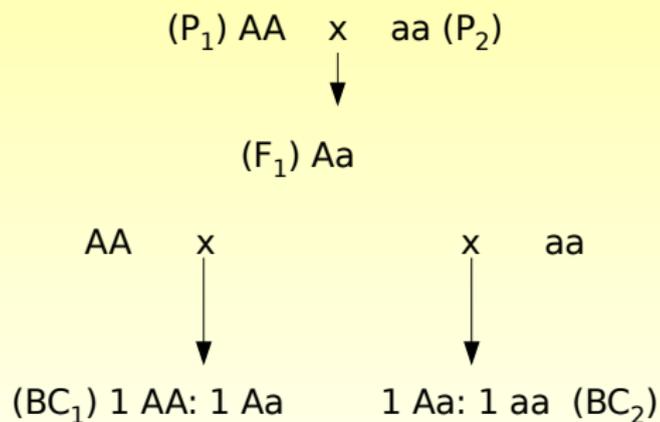
CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

CONTEÚDO

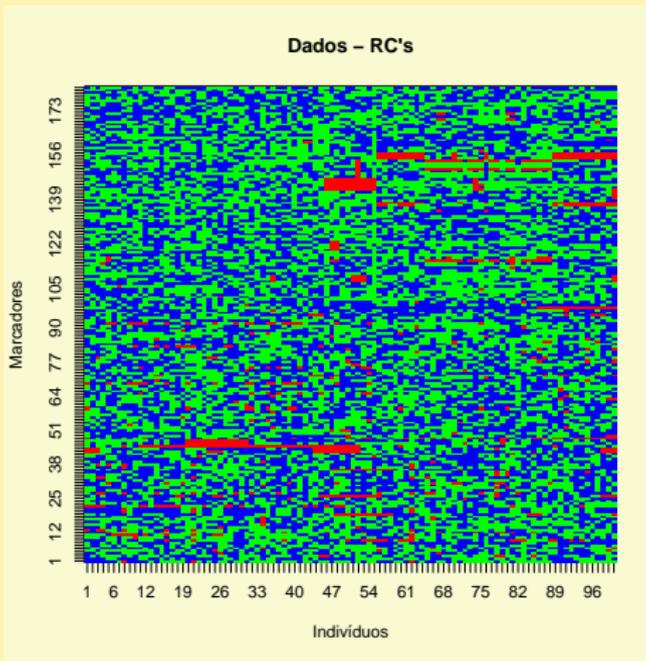
- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

FUNDAMENTOS



DADOS

CANA-DE-AÇÚCAR - GARCIA ET AL. 2006



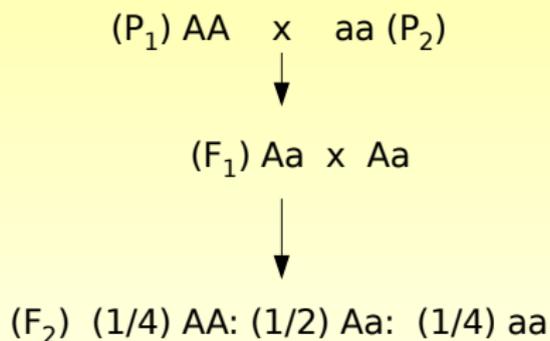
SEGREGAÇÃO

	<i>AA</i>	<i>Aa</i>
Freq. esperada	1/2	1/2
n. esp.	$n/2$	$n/2$
n. obs.	n_1	n_2

CONTEÚDO

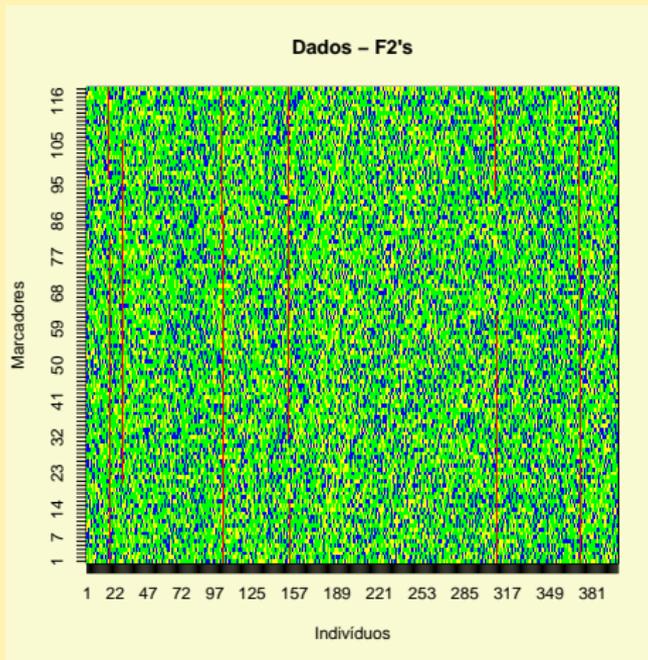
- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

FUNDAMENTOS



DADOS

MILHO - SIBOV ET AL. 2003



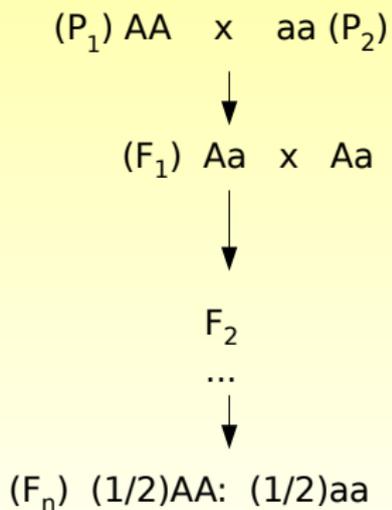
SEGREGAÇÃO

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Freq. esperada	1/4	1/2	1/4
n. esp.	$n/4$	$n/2$	$n/4$
n. obs.	n_1	n_2	n_3

CONTEÚDO

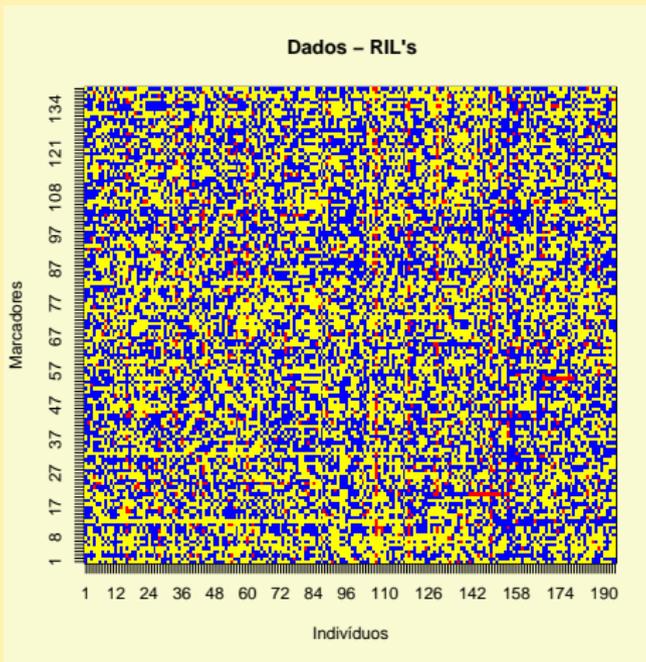
- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

FUNDAMENTOS



DADOS

ARROZ - XIAO ET AL. 1995



SEGREGAÇÃO

	<i>AA</i>	<i>aa</i>
Freq. esperada	1/2	1/2
n. esp.	$n/2$	$n/2$
n. obs.	n_1	n_2

CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

FUNDAMENTOS

(P₁) A₁A₂ × A₃A₄ (P₂)



(F₁) (1/4)A₁A₃: (1/4)A₁A₄: (1/4)A₂A₃: (1/4)A₂A₄

VÁRIAS SEGREGAÇÕES

EXEMPLO

4 alelos	A_1A_3	A_1A_4	A_2A_3	A_2A_4
Freq. esperada	$1/4$	$1/4$	$1/4$	$1/4$
n. esp.	$n/4$	$n/4$	$n/4$	$n/4$
n. obs.	n_1	n_2	n_3	n_4

- Outros tipos de segregação são possíveis: 3:1, 1:2:1 e 1:1

CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 **TESTE DA SEGREGAÇÃO MENDELIANA**
 - **Teste de Aderência**
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

TESTES DE HIPÓTESES

- Abordagem frequentista: uso do *p-valor* e do α como evidências para se testar uma dada hipótese
- Para se testar hipóteses científicas geralmente dois modelos são comparados: $H_0 : \theta = 0$ e $H_1 : \theta \neq 0$
- Testes estatísticos são realizados para medir desvios da hipótese de nulidade
- Há muita confusão na literatura sobre p e α

VERDADEIRO OU FALSO?

- O *p-valor* é a probabilidade de estar errado caso H_0 seja verdadeiro

TESTES DE HIPÓTESES

- Abordagem frequentista: uso do *p-valor* e do α como evidências para se testar uma dada hipótese
- Para se testar hipóteses **científicas** geralmente dois modelos são comparados: $H_0 : \theta = 0$ e $H_1 : \theta \neq 0$
- Testes estatísticos são realizados para medir **desvios** da hipótese de nulidade
- Há muita confusão na literatura sobre p e α

VERDADEIRO OU FALSO?

- O *p-valor* é a probabilidade de estar errado caso H_0 seja verdadeiro

TESTES DE HIPÓTESES

- Abordagem frequentista: uso do *p-valor* e do α como evidências para se testar uma dada hipótese
- Para se testar hipóteses **científicas** geralmente dois modelos são comparados: $H_0 : \theta = 0$ e $H_1 : \theta \neq 0$
- Testes estatísticos são realizados para medir **desvios** da hipótese de nulidade
- Há muita confusão na literatura sobre p e α

VERDADEIRO OU FALSO?

- O *p-valor* é a probabilidade de estar errado caso H_0 seja verdadeiro

TESTES DE HIPÓTESES

- Abordagem frequentista: uso do *p-valor* e do α como evidências para se testar uma dada hipótese
- Para se testar hipóteses **científicas** geralmente dois modelos são comparados: $H_0 : \theta = 0$ e $H_1 : \theta \neq 0$
- Testes estatísticos são realizados para medir **desvios** da hipótese de nulidade
- Há muita confusão na literatura sobre p e α

VERDADEIRO OU FALSO?

- O *p-valor* é a probabilidade de estar errado caso H_0 seja verdadeiro

TESTES DE HIPÓTESES

- Abordagem frequentista: uso do *p-valor* e do α como evidências para se testar uma dada hipótese
- Para se testar hipóteses **científicas** geralmente dois modelos são comparados: $H_0 : \theta = 0$ e $H_1 : \theta \neq 0$
- Testes estatísticos são realizados para medir **desvios** da hipótese de nulidade
- Há muita confusão na literatura sobre p e α

VERDADEIRO OU FALSO?

- O *p-valor* é a probabilidade de estar errado caso H_0 seja verdadeiro

TESTES DE HIPÓTESES

- Abordagem frequentista: uso do *p-valor* e do α como evidências para se testar uma dada hipótese
- Para se testar hipóteses **científicas** geralmente dois modelos são comparados: $H_0 : \theta = 0$ e $H_1 : \theta \neq 0$
- Testes estatísticos são realizados para medir **desvios** da hipótese de nulidade
- Há muita confusão na literatura sobre p e α

VERDADEIRO OU FALSO?

- O *p-valor* é a probabilidade de estar errado caso H_0 seja verdadeiro

TESTES DE HIPÓTESES

- Abordagem frequentista: uso do *p-valor* e do α como evidências para se testar uma dada hipótese
- Para se testar hipóteses **científicas** geralmente dois modelos são comparados: $H_0 : \theta = 0$ e $H_1 : \theta \neq 0$
- Testes estatísticos são realizados para medir **desvios** da hipótese de nulidade
- Há muita confusão na literatura sobre p e α

VERDADEIRO OU FALSO?

- O *p-valor* é a probabilidade de estar errado caso H_0 seja verdadeiro

FALSO! Essa é a definição de α

TESTES DE HIPÓTESES

VERDADEIRO OU FALSO?

- O *p*-valor é a probabilidade observar falsos positivos
- Um erro muito comum é assumir que α é o *p*-valor observado

DEFINIÇÃO

O *p*-valor é definido como a probabilidade de observar valores mais extremos da estatística do teste sob H_0 do que o valor observado.

Sendo T a estatística do teste que assume valores positivos,

$$p = Pr(T \geq T_{obs} | H_0)$$

TESTES DE HIPÓTESES

VERDADEIRO OU FALSO?

- O *p*-valor é a probabilidade observar falsos positivos
- Um erro muito comum é assumir que α é o *p*-valor observado

DEFINIÇÃO

O *p*-valor é definido como a probabilidade de observar valores mais extremos da estatística do teste sob H_0 do que o valor observado.

Sendo T a estatística do teste que assume valores positivos,

$$p = Pr(T \geq T_{obs} | H_0)$$

TESTES DE HIPÓTESES

VERDADEIRO OU FALSO?

- O *p*-valor é a probabilidade observar falsos positivos

FALSO!

- Um erro muito comum é assumir que α é o *p*-valor observado

DEFINIÇÃO

O *p*-valor é definido como a probabilidade de observar valores mais extremos da estatística do teste sob H_0 do que o valor observado.

Sendo T a estatística do teste que assume valores positivos,

$$p = Pr(T \geq T_{obs} | H_0)$$

TESTES DE HIPÓTESES

VERDADEIRO OU FALSO?

- O *p*-valor é a probabilidade observar falsos positivos

FALSO!

- Um erro muito comum é assumir que α é o *p*-valor observado

DEFINIÇÃO

O *p*-valor é definido como a probabilidade de observar valores mais extremos da estatística do teste sob H_0 do que o valor observado.

Sendo T a estatística do teste que assume valores positivos,

$$p = Pr(T \geq T_{obs} | H_0)$$

TESTES DE HIPÓTESES

VERDADEIRO OU FALSO?

- O *p*-valor é a probabilidade observar falsos positivos

FALSO!

- Um erro muito comum é assumir que α é o *p*-valor observado

DEFINIÇÃO

O *p*-valor é definido como a probabilidade de observar valores mais extremos da estatística do teste sob H_0 do que o valor observado.

Sendo T a estatística do teste que assume valores positivos,

$$p = Pr(T \geq T_{obs} | H_0)$$

RETROCRUZAMENTOS

	AA	Aa
Freq. esperada	1/2	1/2
n. esp.	$n/2$	$n/2$
n. obs.	n_1	n_2

$$\begin{aligned}\chi^2 &= \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/2)^2}{n/2} + \frac{(n_2 - n/2)^2}{n/2} = \\ &= \frac{(n_1 - n_2)^2}{n} \sim \chi_1^2\end{aligned}$$

QUANTOS GLs?: 1 GL para $\theta = 1/2$

RETROCRUZAMENTOS

	AA	Aa
Freq. esperada	1/2	1/2
n. esp.	$n/2$	$n/2$
n. obs.	n_1	n_2

$$\begin{aligned}\chi^2 &= \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/2)^2}{n/2} + \frac{(n_2 - n/2)^2}{n/2} = \\ &= \frac{(n_1 - n_2)^2}{n} \sim \chi_1^2\end{aligned}$$

QUANTOS GLs?: 1 GL para $\theta = 1/2$

RETROCRUZAMENTOS

	AA	Aa
Freq. esperada	1/2	1/2
n. esp.	$n/2$	$n/2$
n. obs.	n_1	n_2

$$\begin{aligned}\chi^2 &= \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/2)^2}{n/2} + \frac{(n_2 - n/2)^2}{n/2} = \\ &= \frac{(n_1 - n_2)^2}{n} \sim \chi_1^2\end{aligned}$$

QUANTOS GLs?: 1 GL para $\theta = 1/2$

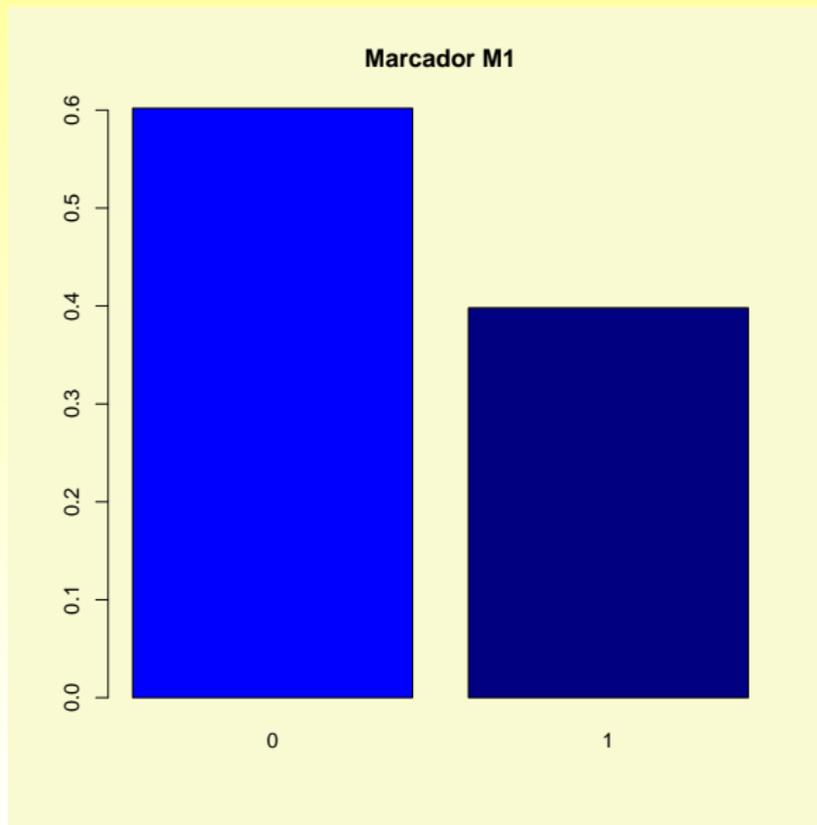
RETROCRUZAMENTOS

	AA	Aa
Freq. esperada	1/2	1/2
n. esp.	$n/2$	$n/2$
n. obs.	n_1	n_2

$$\begin{aligned}\chi^2 &= \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/2)^2}{n/2} + \frac{(n_2 - n/2)^2}{n/2} = \\ &= \frac{(n_1 - n_2)^2}{n} \sim \chi_1^2\end{aligned}$$

QUANTOS GLs?: 1 GL para $\theta = 1/2$

DISTRIBUIÇÃO DE FREQUÊNCIAS



RESULTADOS

	chi.square	p.valor
[1,]	4.281553398	0.03852812
[2,]	3.504854369	0.06118923
[3,]	2.805825243	0.09392250
[4,]	3.504854369	0.06118923
[5,]	2.805825243	0.09392250
[6,]	2.184466019	0.13940941
[7,]	1.640776699	0.20021894
[8,]	0.242718447	0.62224957
[9,]	0.087378641	0.76753650
[10,]	0.087378641	0.76753650
[11,]	0.087378641	0.76753650
[12,]	0.009708738	0.92150913
[13,]	0.786407767	0.37518855
[14,]	0.242718447	0.62224957

F_2

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Freq. esperada	1/4	1/2	1/4
n. esp.	$n/4$	$n/2$	$n/4$
n. obs.	n_1	n_2	n_3

$$\chi^2 = \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/4)^2}{n/4} + \frac{(n_2 - n/2)^2}{n/2} + \frac{(n_3 - n/4)^2}{n/4} \sim \chi_2^2$$

QUANTOS GL? Dois: θ_1 e θ_2 (multinomial)

F_2

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Freq. esperada	1/4	1/2	1/4
n. esp.	$n/4$	$n/2$	$n/4$
n. obs.	n_1	n_2	n_3

$$\chi^2 = \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/4)^2}{n/4} + \frac{(n_2 - n/2)^2}{n/2} + \frac{(n_3 - n/4)^2}{n/4} \sim \chi_2^2$$

QUANTOS GL? Dois: θ_1 e θ_2 (multinomial)

F_2

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Freq. esperada	1/4	1/2	1/4
n. esp.	$n/4$	$n/2$	$n/4$
n. obs.	n_1	n_2	n_3

$$\chi^2 = \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/4)^2}{n/4} + \frac{(n_2 - n/2)^2}{n/2} + \frac{(n_3 - n/4)^2}{n/4} \sim \chi_2^2$$

QUANTOS GL? Dois: θ_1 e θ_2 (multinomial)

F_2

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Freq. esperada	1/4	1/2	1/4
n. esp.	$n/4$	$n/2$	$n/4$
n. obs.	n_1	n_2	n_3

$$\chi^2 = \sum \frac{(n.obs - n.esp)^2}{n.esp} = \frac{(n_1 - n/4)^2}{n/4} + \frac{(n_2 - n/2)^2}{n/2} + \frac{(n_3 - n/4)^2}{n/4} \sim \chi_2^2$$

QUANTOS GL? **Dois:** θ_1 e θ_2 (multinomial)

CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 **TESTE DA SEGREGAÇÃO MENDELIANA**
 - Teste de Aderência
 - **Correção para Múltiplos Testes**
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

BONFERRONI

- Mapeamento: normalmente, os testes são realizados repetidas vezes
- $1 - \alpha$: probab. de não cometer erro tipo I em um teste
- $(1 - \alpha)^m$: prob. de não cometer erro tipo I nesses m testes
- Note que estamos assumindo que os m testes são independentes
- α^* : erro conjunto tipo I
- Logo, $1 - \alpha^* = (1 - \alpha)^m$ e $\alpha^* = 1 - (1 - \alpha)^m$

BONFERRONI

- Mapeamento: normalmente, os testes são realizados repetidas vezes
- $1 - \alpha$: probab. de não cometer erro tipo I em um teste
- $(1 - \alpha)^m$: prob. de não cometer erro tipo I nesses m testes
- Note que estamos assumindo que os m testes são independentes
- α^* : erro conjunto tipo I
- Logo, $1 - \alpha^* = (1 - \alpha)^m$ e $\alpha^* = 1 - (1 - \alpha)^m$

BONFERRONI

- Mapeamento: normalmente, os testes são realizados repetidas vezes
- $1 - \alpha$: probab. de não cometer erro tipo I em um teste
- $(1 - \alpha)^m$: prob. de não cometer erro tipo I nesses m testes
- Note que estamos assumindo que os m testes são independentes
- α^* : erro conjunto tipo I
- Logo, $1 - \alpha^* = (1 - \alpha)^m$ e $\alpha^* = 1 - (1 - \alpha)^m$

BONFERRONI

- Mapeamento: normalmente, os testes são realizados repetidas vezes
- $1 - \alpha$: probab. de não cometer erro tipo I em um teste
- $(1 - \alpha)^m$: prob. de não cometer erro tipo I nesses m testes
- Note que estamos assumindo que os m testes são **independentes**
- α^* : erro **conjunto** tipo I
- Logo, $1 - \alpha^* = (1 - \alpha)^m$ e $\alpha^* = 1 - (1 - \alpha)^m$

BONFERRONI

- Mapeamento: normalmente, os testes são realizados repetidas vezes
- $1 - \alpha$: probab. de não cometer erro tipo I em um teste
- $(1 - \alpha)^m$: prob. de não cometer erro tipo I nesses m testes
- Note que estamos assumindo que os m testes são **independentes**
- α^* : erro **conjunto** tipo I
- Logo, $1 - \alpha^* = (1 - \alpha)^m$ e $\alpha^* = 1 - (1 - \alpha)^m$

BONFERRONI

- Mapeamento: normalmente, os testes são realizados repetidas vezes
- $1 - \alpha$: probab. de não cometer erro tipo I em um teste
- $(1 - \alpha)^m$: prob. de não cometer erro tipo I nesses m testes
- Note que estamos assumindo que os m testes são **independentes**
- α^* : erro **conjunto** tipo I
- Logo, $1 - \alpha^* = (1 - \alpha)^m$ e $\alpha^* = 1 - (1 - \alpha)^m$

BONFERRONI

EXEMPLO

- $m = 14$
- $\alpha = 0,05$
- Qual a probabilidade de ocorrer pelo menos um falso positivo nos 14 testes?
- Resp: $\alpha^* = 0,51$

EXEMPLO

- Simulação usando o R

BONFERRONI

EXEMPLO

- $m = 14$
- $\alpha = 0,05$
- Qual a probabilidade de ocorrer pelo menos um falso positivo nos 14 testes?
- Resp: $\alpha^* = 0,51$

EXEMPLO

- Simulação usando o R

BONFERRONI

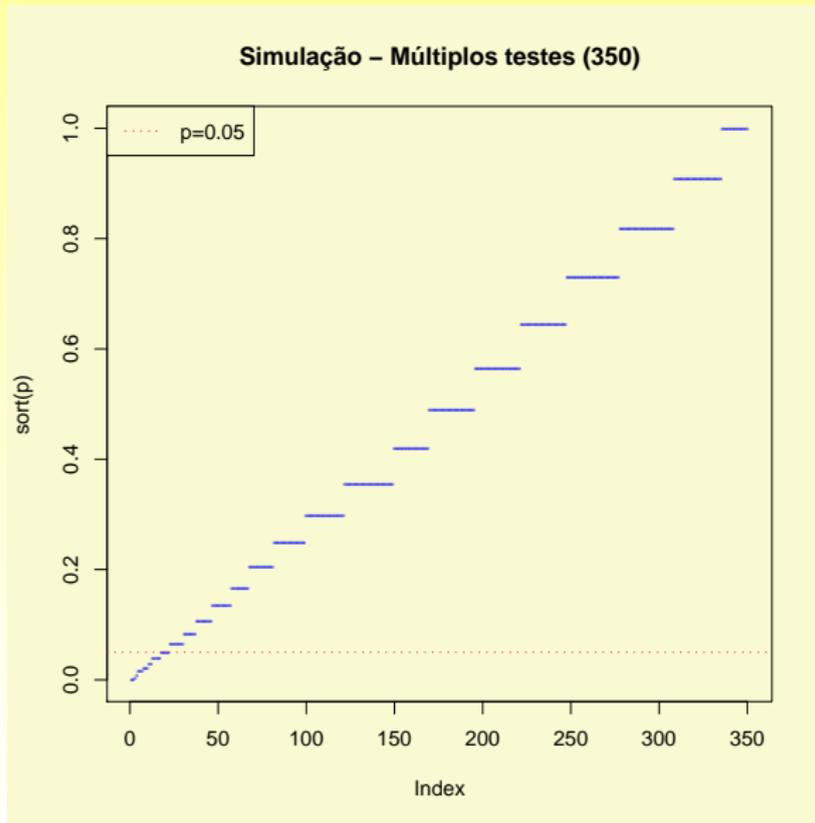
EXEMPLO

- $m = 14$
- $\alpha = 0,05$
- Qual a probabilidade de ocorrer pelo menos um falso positivo nos 14 testes?
- Resp: $\alpha^* = 0,51$

EXEMPLO

- Simulação usando o R

DISTRIBUIÇÃO DE FREQUÊNCIAS



BONFERRONI

CORREÇÃO

- Princípio: fixar α^* e a partir desse valor determinar o α individual

$$\log(1 - \alpha^*) = \log(1 - \alpha)^m = m \cdot \log(1 - \alpha)$$

$$\alpha = -e^{\frac{\log(1-\alpha^*)}{m}} + 1$$

APROXIMAÇÃO: $\alpha = \frac{\alpha^*}{m}$

BONFERRONI

CORREÇÃO

- Princípio: fixar α^* e a partir desse valor determinar o α individual

$$\log(1 - \alpha^*) = \log(1 - \alpha)^m = m \cdot \log(1 - \alpha)$$

$$\alpha = -e^{\frac{\log(1 - \alpha^*)}{m}} + 1$$

APROXIMAÇÃO: $\alpha = \frac{\alpha^*}{m}$

BONFERRONI

CORREÇÃO

- Princípio: fixar α^* e a partir desse valor determinar o α individual

$$\log(1 - \alpha^*) = \log(1 - \alpha)^m = m \cdot \log(1 - \alpha)$$

$$\alpha = -e^{\frac{\log(1-\alpha^*)}{m}} + 1$$

APROXIMAÇÃO: $\alpha = \frac{\alpha^*}{m}$

BONFERRONI

CORREÇÃO

- Princípio: fixar α^* e a partir desse valor determinar o α individual

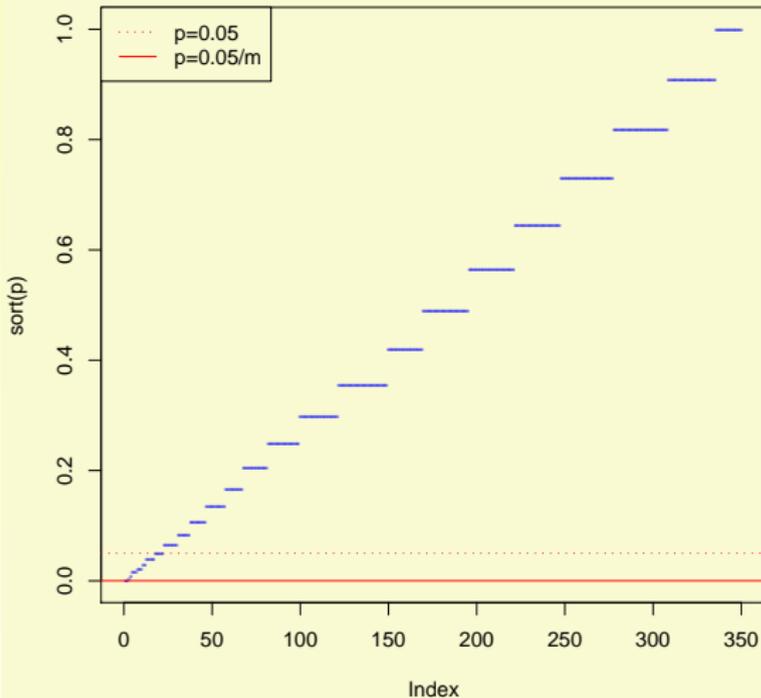
$$\log(1 - \alpha^*) = \log(1 - \alpha)^m = m \cdot \log(1 - \alpha)$$

$$\alpha = -e^{\frac{\log(1-\alpha^*)}{m}} + 1$$

APROXIMAÇÃO: $\alpha = \frac{\alpha^*}{m}$

BONFERRONI

Simulação - Múltiplos testes(350)



BONFERRONI

EXEMPLO

- $m = 14$
- $\alpha^* = 0,05$
- Qual valor de α deve ser usado em cada teste para garantir esse valor global de 5%?
- Resp: $\alpha = 0,00357$ (bem menor que 0,05)

PONTOS PARA REFLEXÃO

- Os m testes são independentes no caso dos mapas genéticos?
- São graves as conseqüências de não descartar marcas que não segregam mendelianamente?
- A correção de Bonferroni é conhecida como **conservativa**.

BONFERRONI

EXEMPLO

- $m = 14$
- $\alpha^* = 0,05$
- Qual valor de α deve ser usado em cada teste para garantir esse valor global de 5%?
- Resp: $\alpha = 0,00357$ (bem menor que 0,05)

PONTOS PARA REFLEXÃO

- Os m testes são independentes no caso dos mapas genéticos?
- São graves as conseqüências de não descartar marcas que não segregam mendelianamente?
- A correção de Bonferroni é conhecida como **conservativa**.

BONFERRONI

EXEMPLO

- $m = 14$
- $\alpha^* = 0,05$
- Qual valor de α deve ser usado em cada teste para garantir esse valor global de 5%?
- Resp: $\alpha = 0,00357$ (bem menor que 0,05)

PONTOS PARA REFLEXÃO

- Os m testes são independentes no caso dos mapas genéticos?
- São graves as conseqüências de não descartar marcas que não segregam mendelianamente?
- A correção de Bonferroni é conhecida como **conservativa**.

BONFERRONI

EXEMPLO

- $m = 14$
- $\alpha^* = 0,05$
- Qual valor de α deve ser usado em cada teste para garantir esse valor global de 5%?
- Resp: $\alpha = 0,00357$ (bem menor que 0,05)

PONTOS PARA REFLEXÃO

- Os m testes são independentes no caso dos mapas genéticos?
- São graves as conseqüências de não descartar marcas que não segregam mendelianamente?
- A correção de Bonferroni é conhecida como **conservativa**.

BONFERRONI

EXEMPLO

- $m = 14$
- $\alpha^* = 0,05$
- Qual valor de α deve ser usado em cada teste para garantir esse valor global de 5%?
- Resp: $\alpha = 0,00357$ (bem menor que 0,05)

PONTOS PARA REFLEXÃO

- Os m testes são independentes no caso dos mapas genéticos?
- São graves as conseqüências de não descartar marcas que não segregam mendelianamente?
- A correção de Bonferroni é conhecida como **conservativa**.

RAZÃO DE FALSAS DESCOBERTAS (*FDR*)

- *False Discovery Rate*: proposto como uma alternativa para controlar o erro tipo I
- Benjamini e Hochberg, 1995, 2000; Benjamini e Yekutieli, 2001
- Seu uso é cada vez mais freqüente em experimentos com microarranjos
- Motivação: usar $\alpha = 0,05$ (ou $\alpha = 0,01$) fornece muitos falso positivos; usar α^* elimina muitos positivos verdadeiros
- Idéia: dados os resultados significativos, determina-se quantos deles (proporção) são verdadeiramente significativos ($1-FDR$)

RAZÃO DE FALSAS DESCOBERTAS (*FDR*)

- *False Discovery Rate*: proposto como uma alternativa para controlar o erro tipo I
- Benjamini e Hochberg, 1995, 2000; Benjamini e Yekutieli, 2001
- Seu uso é cada vez mais freqüente em experimentos com microarranjos
- Motivação: usar $\alpha = 0,05$ (ou $\alpha = 0,01$) fornece muitos falso positivos; usar α^* elimina muitos positivos verdadeiros
- Idéia: dados os resultados significativos, determina-se quantos deles (proporção) são verdadeiramente significativos ($1-FDR$)

RAZÃO DE FALSAS DESCOBERTAS (*FDR*)

- *False Discovery Rate*: proposto como uma alternativa para controlar o erro tipo I
- Benjamini e Hochberg, 1995, 2000; Benjamini e Yekutieli, 2001
- Seu uso é cada vez mais freqüente em experimentos com microarranjos
- Motivação: usar $\alpha = 0,05$ (ou $\alpha = 0,01$) fornece muitos falso positivos; usar α^* elimina muitos positivos verdadeiros
- Idéia: dados os resultados significativos, determina-se quantos deles (proporção) são verdadeiramente significativos ($1-FDR$)

RAZÃO DE FALSAS DESCOBERTAS (*FDR*)

- *False Discovery Rate*: proposto como uma alternativa para controlar o erro tipo I
- Benjamini e Hochberg, 1995, 2000; Benjamini e Yekutieli, 2001
- Seu uso é cada vez mais freqüente em experimentos com microarranjos
- Motivação: usar $\alpha = 0,05$ (ou $\alpha = 0,01$) fornece muitos falso positivos; usar α^* elimina muitos positivos verdadeiros
- Idéia: dados os resultados significativos, determina-se quantos deles (proporção) são verdadeiramente significativos ($1-FDR$)

RAZÃO DE FALSAS DESCOBERTAS (*FDR*)

- *False Discovery Rate*: proposto como uma alternativa para controlar o erro tipo I
- Benjamini e Hochberg, 1995, 2000; Benjamini e Yekutieli, 2001
- Seu uso é cada vez mais freqüente em experimentos com microarranjos
- Motivação: usar $\alpha = 0,05$ (ou $\alpha = 0,01$) fornece muitos falso positivos; usar α^* elimina muitos positivos verdadeiros
- Idéia: dados os resultados significativos, determina-se quantos deles (proporção) são verdadeiramente significativos ($1-FDR$)

FDR

DEFINIÇÃO

FDR é a proporção esperada de falsas descobertas dentre as hipóteses H_0 rejeitadas

- *FDR* positivo (Storey, 2003):

$$pFDR = E(V/R | R > 0)$$

- V : número de falsos positivos quando H_0 é rejeitada
- R : número de hipóteses H_0 rejeitadas

FDR

DEFINIÇÃO

FDR é a proporção esperada de falsas descobertas dentre as hipóteses H_0 rejeitadas

- *FDR* positivo (Storey, 2003):

$$pFDR = E(V/R | R > 0)$$

- V : número de falsos positivos quando H_0 é rejeitada
- R : número de hipóteses H_0 rejeitadas

CÁLCULO

- 1 Ordene todos os *p*-valores ($p_{(i)}$ denota os valores ordenados)
- 2 Encontre o último *p*-valor $p_{(k)}$ que fica abaixo da linha

$$p_{(i)} = \frac{\alpha}{\pi_0 m} \times i$$

- 3 π_0 é a proporção de efeitos verdadeiros (????)
 - Na prática, para microarranjos, $\pi_0 \approx 1$
- 4 H_0 é rejeitado para todos $p_{(i)} \leq p_{(k)}$

FDR é conservativo

CÁLCULO

- 1 Ordene todos os p -valores ($p_{(i)}$ denota os valores ordenados)
- 2 Encontre o último p -valor $p_{(k)}$ que fica abaixo da linha

$$p_{(i)} = \frac{\alpha}{\pi_0 m} \times i$$

- 3 π_0 é a proporção de efeitos verdadeiros (????)
 - Na prática, para microarranjos, $\pi_0 \approx 1$
- 4 H_0 é rejeitado para todos $p_{(i)} \leq p_{(k)}$

FDR é conservativo

CÁLCULO

- 1 Ordene todos os p -valores ($p_{(i)}$ denota os valores ordenados)
- 2 Encontre o último p -valor $p_{(k)}$ que fica abaixo da linha

$$p_{(i)} = \frac{\alpha}{\pi_0 m} \times i$$

- 3 π_0 é a proporção de efeitos verdadeiros (????)
 - Na prática, para microarranjos, $\pi_0 \approx 1$
- 4 H_0 é rejeitado para todos $p_{(i)} \leq p_{(k)}$

FDR é conservativo

CÁLCULO

- 1 Ordene todos os p -valores ($p_{(i)}$ denota os valores ordenados)
- 2 Encontre o último p -valor $p_{(k)}$ que fica abaixo da linha

$$p_{(i)} = \frac{\alpha}{\pi_0 m} \times i$$

- 3 π_0 é a proporção de efeitos verdadeiros (????)
 - Na prática, para microarranjos, $\pi_0 \approx 1$
- 4 H_0 é rejeitado para todos $p_{(i)} \leq p_{(k)}$

FDR é conservativo

CÁLCULO

- 1 Ordene todos os p -valores ($p_{(i)}$ denota os valores ordenados)
- 2 Encontre o último p -valor $p_{(k)}$ que fica abaixo da linha

$$p_{(i)} = \frac{\alpha}{\pi_0 m} \times i$$

- 3 π_0 é a proporção de efeitos verdadeiros (????)
 - Na prática, para microarranjos, $\pi_0 \approx 1$
- 4 H_0 é rejeitado para todos $p_{(i)} \leq p_{(k)}$

FDR é conservativo

CÁLCULO

- 1 Ordene todos os p -valores ($p_{(i)}$ denota os valores ordenados)
- 2 Encontre o último p -valor $p_{(k)}$ que fica abaixo da linha

$$p_{(i)} = \frac{\alpha}{\pi_0 m} \times i$$

- 3 π_0 é a proporção de efeitos verdadeiros (????)
 - Na prática, para microarranjos, $\pi_0 \approx 1$
- 4 H_0 é rejeitado para todos $p_{(i)} \leq p_{(k)}$

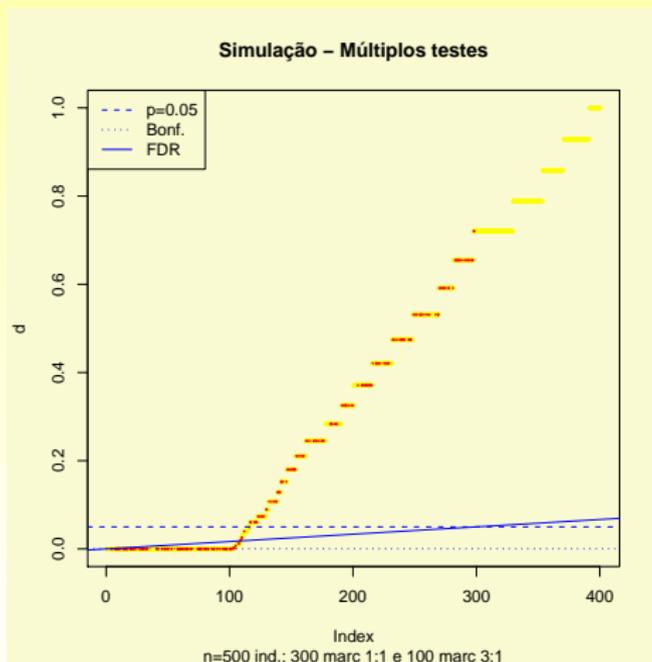
FDR é conservativo

SIMULAÇÕES

- Várias situações podem ser estudadas via simulação no R

FDR

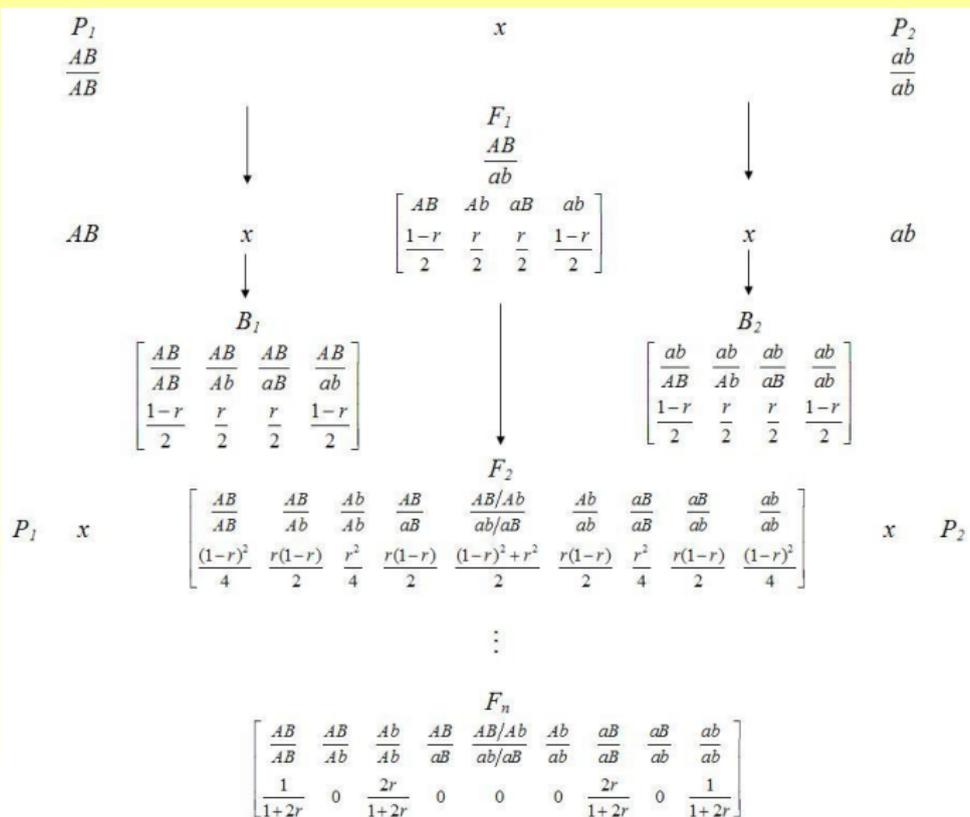
- Simulando 500 indivíduos, genotipados com 400 marc. (300 1:1 e 100 3:1)



CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

DELINEAMENTOS GENÉTICOS



CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

RC

- Qual a função de verossimilhança para r no RC1?

$$L(r) = \left(\frac{1-r}{2}\right)^{n_1} \cdot \left(\frac{r}{2}\right)^{n_2} \cdot \left(\frac{r}{2}\right)^{n_3} \cdot \left(\frac{1-r}{2}\right)^{n_4}$$

- Qual o estimador de máxima verossimilhança para r ?

$$\hat{r} = \frac{n_2 + n_3}{n_1 + n_2 + n_3 + n_4} = \frac{n_R}{n_R + n_{NR}}$$

RC

- Qual a função de verossimilhança para r no RC1?

$$L(r) = \left(\frac{1-r}{2}\right)^{n_1} \cdot \left(\frac{r}{2}\right)^{n_2} \cdot \left(\frac{r}{2}\right)^{n_3} \cdot \left(\frac{1-r}{2}\right)^{n_4}$$

- Qual o estimador de máxima verossimilhança para r ?

$$\hat{r} = \frac{n_2 + n_3}{n_1 + n_2 + n_3 + n_4} = \frac{n_R}{n_R + n_{NR}}$$

RC

- Qual a função de verossimilhança para r no RC1?

$$L(r) = \left(\frac{1-r}{2}\right)^{n_1} \cdot \left(\frac{r}{2}\right)^{n_2} \cdot \left(\frac{r}{2}\right)^{n_3} \cdot \left(\frac{1-r}{2}\right)^{n_4}$$

- Qual o estimador de máxima verossimilhança para r ?

$$\hat{r} = \frac{n_2 + n_3}{n_1 + n_2 + n_3 + n_4} = \frac{n_R}{n_R + n_{NR}}$$

RC

- Qual a função de verossimilhança para r no RC1?

$$L(r) = \left(\frac{1-r}{2}\right)^{n_1} \cdot \left(\frac{r}{2}\right)^{n_2} \cdot \left(\frac{r}{2}\right)^{n_3} \cdot \left(\frac{1-r}{2}\right)^{n_4}$$

- Qual o estimador de máxima verossimilhança para r ?

$$\hat{r} = \frac{n_2 + n_3}{n_1 + n_2 + n_3 + n_4} = \frac{n_R}{n_R + n_{NR}}$$

MOUSE

- Qual a estimativa de r entre os marcadores 1 e 2?

- $\hat{r} = \frac{7}{7+96} = 0,0680$

MOUSE

- Qual a estimativa de r entre os marcadores 1 e 2?
- $\hat{r} = \frac{7}{7+96} = 0,0680$

CONTEÚDO

- 1 POPULAÇÕES DE MAPEAMENTO
 - Retrocruzamentos
 - Populações F_2
 - RIL's
 - F_1 's segregantes
- 2 TESTE DA SEGREGAÇÃO MENDELIANA
 - Teste de Aderência
 - Correção para Múltiplos Testes
- 3 ANÁLISE DE LIGAÇÃO
 - Fundamentos
 - Verossimilhança e Estimação (RC)
 - Teste de Dois Pontos

TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Na maioria das situações, não é suficiente apenas obter estimativas ML
- Geralmente, temos interesses em **testar** hipóteses de interesse relacionadas com os parâmetros
- Uma estatística usada para tanto é a **razão de verossimilhanças** (LR)
- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros

TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Na maioria das situações, não é suficiente apenas obter estimativas ML
- Geralmente, temos interesses em **testar** hipóteses de interesse relacionadas com os parâmetros
- Uma estatística usada para tanto é a **razão de verossimilhanças** (LR)
- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros

TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Na maioria das situações, não é suficiente apenas obter estimativas ML
- Geralmente, temos interesses em **testar** hipóteses de interesse relacionadas com os parâmetros
- Uma estatística usada para tanto é a **razão de verossimilhanças** (LR)
- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros

TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Na maioria das situações, não é suficiente apenas obter estimativas ML
- Geralmente, temos interesses em **testar** hipóteses de interesse relacionadas com os parâmetros
- Uma estatística usada para tanto é a **razão de verossimilhanças** (LR)
- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros

LRT

- Seja θ o valor do parâmetro sob H_0 e $\hat{\theta}$ o valor sob H_1
- A estatística razão de verossimilhanças é

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})}$$

- Note que $LRT = -2[l(\theta) - l(\hat{\theta})]$
- Importante: $LRT \sim \chi^2$, com número de GL dado pelo número de parâmetros sendo testados
- Note que podemos escrever também

$$LRT = 2 \log \frac{L(\hat{\theta})}{L(\theta)}$$

LRT

- Seja θ o valor do parâmetro sob H_0 e $\hat{\theta}$ o valor sob H_1
- A estatística razão de verossimilhanças é

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})}$$

- Note que $LRT = -2[l(\theta) - l(\hat{\theta})]$
- Importante: $LRT \sim \chi^2$, com número de GL dado pelo número de parâmetros sendo testados
- Note que podemos escrever também

$$LRT = 2 \log \frac{L(\hat{\theta})}{L(\theta)}$$

LRT

- Seja θ o valor do parâmetro sob H_0 e $\hat{\theta}$ o valor sob H_1
- A estatística razão de verossimilhanças é

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})}$$

- Note que $LRT = -2[l(\theta) - l(\hat{\theta})]$
- Importante: $LRT \sim \chi^2$, com número de GL dado pelo número de parâmetros sendo testados
- Note que podemos escrever também

$$LRT = 2 \log \frac{L(\hat{\theta})}{L(\theta)}$$

LRT

- Seja θ o valor do parâmetro sob H_0 e $\hat{\theta}$ o valor sob H_1
- A estatística razão de verossimilhanças é

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})}$$

- Note que $LRT = -2[l(\theta) - l(\hat{\theta})]$
- Importante: $LRT \sim \chi^2$, com número de GL dado pelo número de parâmetros sendo testados
- Note que podemos escrever também

$$LRT = 2 \log \frac{L(\hat{\theta})}{L(\theta)}$$

LRT

- Seja θ o valor do parâmetro sob H_0 e $\hat{\theta}$ o valor sob H_1
- A estatística razão de verossimilhanças é

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})}$$

- Note que $LRT = -2[l(\theta) - l(\hat{\theta})]$
- Importante: $LRT \sim \chi^2$, com número de GL dado pelo número de parâmetros sendo testados
- Note que podemos escrever também

$$LRT = 2 \log \frac{L(\hat{\theta})}{L(\theta)}$$

LOD SCORE

- A razão

$$\frac{L(\theta)}{L(\hat{\theta})}$$

é conhecida como *odds ratio*, ou *razão de chances*

- Usando o logaritmo na base 10, podemos calcular o *log of the odds* (LOD) no lugar do LRT:

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

- Supostamente, a interpretação é mais intuitiva (mas note que não há *p-valor* envolvido, necessitando cautela para as interpretações)

LOD SCORE

- A razão

$$\frac{L(\theta)}{L(\hat{\theta})}$$

é conhecida como *odds ratio*, ou *razão de chances*

- Usando o logaritmo na base 10, podemos calcular o *log of the odds* (LOD) no lugar do LRT:

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

- Supostamente, a interpretação é mais intuitiva (mas note que não há *p-valor* envolvido, necessitando cautela para as interpretações)

LOD SCORE

- A razão

$$\frac{L(\theta)}{L(\hat{\theta})}$$

é conhecida como *odds ratio*, ou *razão de chances*

- Usando o logaritmo na base 10, podemos calcular o *log of the odds* (LOD) no lugar do LRT:

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

- Supostamente, a interpretação é mais intuitiva (mas note que não há *p-valor* envolvido, necessitando cautela para as interpretações)

LOD

- Usar $LOD = 3$ significa que

$$\log_{10} \frac{L(\hat{\theta})}{L(\theta)} = 3$$

ou seja,

$$L(\hat{\theta}) = 10^3 L(\theta)$$

- $LOD = 0, 2172$ LRT (verifique!)
- Note que o LOD aumenta numa escala logarítmica, ou seja, cada aumento de 1 unidade no LOD significa um aumento de 10 vezes na razão

LOD

- Usar $LOD = 3$ significa que

$$\log_{10} \frac{L(\hat{\theta})}{L(\theta)} = 3$$

ou seja,

$$L(\hat{\theta}) = 10^3 L(\theta)$$

- $LOD = 0, 2172$ LRT (verifique!)
- Note que o LOD aumenta numa escala logarítmica, ou seja, cada aumento de 1 unidade no LOD significa um aumento de 10 vezes na razão

LOD

- Usar $LOD = 3$ significa que

$$\log_{10} \frac{L(\hat{\theta})}{L(\theta)} = 3$$

ou seja,

$$L(\hat{\theta}) = 10^3 L(\theta)$$

- $LOD = 0, 2172$ LRT (verifique!)
- Note que o LOD aumenta numa escala logarítmica, ou seja, cada aumento de 1 unidade no LOD significa um aumento de 10 vezes na razão

CÁLCULO

EXERCÍCIO

- 1 Teste se os marcadores 1 e 2 (mouse) estão ligados
- 2 Interprete o resultado do ponto de vista genético

CÁLCULO

EXERCÍCIO

- 1 Teste se os marcadores 1 e 2 (mouse) estão ligados
- 2 Interprete o resultado do ponto de vista genético

$$LRT = 91,63$$

$$p = 1,04 \times 10^{-21}$$

$$LOD = 19,90$$