

**IBI5086**  
**Introdução a Métodos Estatísticos**  
**para a Bioinformática**

*Profa. Júlia Maria Pavan Soler*  
*pavan@ime.usp.br*


IME/USP – 2º Semestre/2020

# Planejamento de Experimentos e Modelos ANOVA (Análise de Variância)

$$Y = f(X) + e$$

- Estrutura dos Fatores (Tratamentos – variável X):
  - Um único Fator e seus níveis (Ex. 1 Fator em 4 níveis)
  - Fatorial Cruzado (Ex. Fatorial 2x2)
- Estrutura das unidades amostrais (Aleatorização dos Trat):
  - Delineamento Completamente Aleatorizado (DCA)
  - Delineamento Aleatorizado em Blocos Completos (DABC):

*Já vimos!*

- 
- **Modelos (ANOVA e de Regressão) para avaliar Efeitos Genéticos:**
    - **Modelos Uniloco: de 1 Fator em 3 níveis (aa, Aa, AA)**
    - **Modelos Biloco: Fatorial 3x3**

# Modelos ANOVA

**Modelo de ANOVA: DCA com esquema de Um único fator em  $a$  níveis**

$$y_{ij} = \mu_0 + \tau_j + e_{ij}; \quad e_{ij} \stackrel{iid}{\sim} N(0, \sigma^2) \quad \Rightarrow \quad SQ_{Total} = SQ_{Trat} + SQ_{Res}$$

Modelo estrutural e distribucional

Esqueleto da ANOVA (FV)

**Modelo de ANOVA: DCA Fatorial Cruzado  $a \times b$**

$$y_{ijk} = \mu_{00} + \tau_j + \beta_k + \gamma_{jk} + e_{ijk}; \quad e_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\Rightarrow SQ_{Total} = SQ(A) + SQ(B) + SQ(A * B) + SQ_{Res}$$

Ef. Principal de A    Ef. Principal de B    Ef. Interação entre A e B    Ef. Residual

**Modelo de ANOVA: DABC com esquema de Um fator em  $J$  níveis**

$$y_{ij} = \mu_{00} + \tau_j + \beta_i + e_{ij}; \quad e_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\Rightarrow SQ_{Total} = SQ(A) + SQ(B) + SQ_{Res} \quad \text{Modelo aditivo (resíduo é a interação)}$$

# Motivação

**Considere um estudo de mapeamento de genes, cujo objetivo é avaliar o efeito de marcadores moleculares em fenótipos (variável Y) de interesse.**

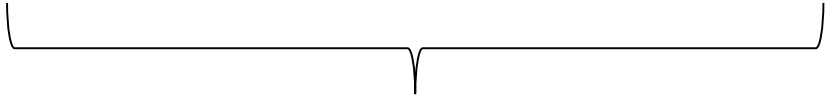
**Caso 1:** Avaliar o efeito do genótipo de um marcador molecular dialélico (SNP, Single Nucleotide Polimorphisms) em Y. Neste caso, considerando A o alelo de risco, as classes genotípicas do SNP são aa, Aa e AA, que correspondem, respectivamente, a 0, 1 e 2 alelos de risco.

**Caso 2:** Avaliar o efeito do genótipo conjunto de 2 marcadores moleculares dialélicos (SNP1 e SNP2) em Y. Neste caso, há 9 possíveis classes genotípicas:  $a_1a_1-a_2a_2$ ,  $a_1a_1-A_2a_2$ ,  $a_1a_1-A_2A_2$ ,  $a_1A_1-a_2a_2$ ,  $a_1A_1-A_2a_2$ ,  $a_1A_1-A_2A_2$ ,  $A_1A_1-a_2a_2$ ,  $A_1A_1-A_2a_2$ ,  $A_1A_1-A_2A_2$

**Que modelo de ANOVA seria apropriado no Caso 1? E no Caso 2?**

# Modelos Estatísticos:

**Modelos Lineares**  $\Rightarrow Y_{n \times 1} = X_{n \times a} \beta_{a \times 1} + e_{n \times 1}$



Flexibilidade na definição da matriz X:

- Modelos ANOVA (X com valores 0, 1 e -1)
- Modelos de Regressão (X com preditores gerais)

# Modelos ANOVA: Notação Matricial

$$y_{ij} = \mu_0 + \tau_j + e_{ij} \quad \Rightarrow \quad Y_{n \times 1} = X_{n \times a} \beta_{a \times 1} + e_{n \times 1}$$

Dados de “produção”: DCA, um Fator em 4 níveis e 6 réplicas

Vetor de observações

Matriz de Planejamento ou Preditores

Vetor de parâmetros

Vetor de erros

$$\begin{aligned} y_{11} &= \mu_1 + e_{11} \\ y_{21} &= \mu_1 + e_{21} \\ &\dots \\ y_{61} &= \mu_1 + e_{61} \\ y_{12} &= \mu_1 + \tau_2 + e_{12} \\ &\dots \\ y_{62} &= \mu_1 + \tau_2 + e_{62} \\ y_{13} &= \mu_1 + \tau_3 + e_{13} \\ &\dots \\ y_{63} &= \mu_1 + \tau_3 + e_{63} \\ y_{14} &= \mu_1 + \tau_4 + e_{14} \\ y_{24} &= \mu_1 + \tau_4 + e_{24} \\ &\dots \\ y_{64} &= \mu_1 + \tau_4 + e_{64} \end{aligned}$$

$$\begin{bmatrix} y_{11} \\ y_{21} \\ \dots \\ y_{61} \\ y_{12} \\ \dots \\ y_{62} \\ y_{13} \\ \dots \\ y_{63} \\ y_{14} \\ y_{24} \\ \dots \\ y_{64} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{21} \\ \dots \\ e_{61} \\ e_{12} \\ \dots \\ e_{62} \\ e_{13} \\ \dots \\ e_{63} \\ e_{14} \\ e_{24} \\ \dots \\ e_{64} \end{bmatrix}$$

# Exemplo: DCA com 1 Fator em 4 níveis

Passo1: Entender o estudo

Passo2: Adotar um modelo teórico de análise

Passo3: Ajuste do modelo e análise de diagnóstico

Dados de um DCA com 1 Fator em 4 níveis

T1	T2	T3	T4
6,2	12,7	7,0	8,3
4,8	11,3	4,4	7,1
3,0	9,3	3,8	11,7
5,6	9,5	5,0	10,0
7,1	11,7	5,5	8,5
4,8	15,3	3,2	12,4

$$H_0 : \mu_j = \mu$$

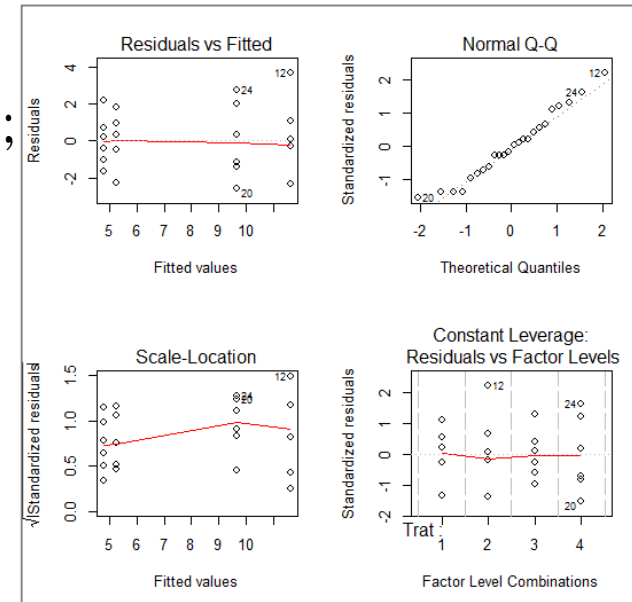
$H_1$  : Existe pelo menos uma diferença

$$y_{ij} = \mu_j + e_{ij} = \mu_1 + \tau_j + e_{ij};$$

$$j = 2, 3, 4; i = 1, \dots, 6$$

$$e_{ij} \stackrel{iid}{\sim} N(0; \sigma^2);$$

$$y_{ij} \sim N(\mu_j; \sigma^2)$$



Passo4: Análise dos resultados e Conclusão

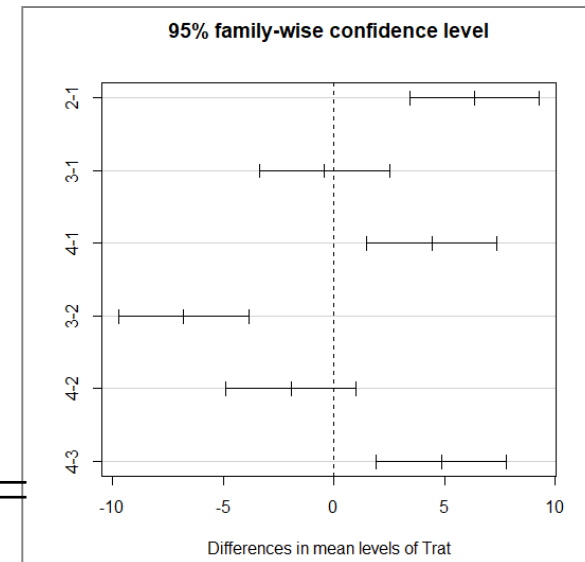
Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Trat	3	199.937	66.646	20.16	2.924e-06
Residuals	20	66.117	3.306		

Coefficients

(Intercept)	Trat2	Trat3	Trat4
5.250	6.350	-0.433	4.417

$$(\mu_2 = \mu_4) > (\mu_1 = \mu_3) \leftarrow$$




# Modelos ANOVA

**Caso 1:** Avaliar o efeito do genótipo de um marcador molecular dialélico (SNP, Single Nucleotide Polimorphisms) em Y. Neste caso, considerando A o alelo de risco, as classes genotípicas do SNP são aa, Aa e AA, que correspondem, respectivamente, a 0, 1 e 2 alelos de risco.



**DCA, Um Fator em 3 níveis: 2 graus de liberdade para estudar o efeito do fator SNP**

**Caso 2:** Avaliar o efeito do genótipo conjunto de 2 marcadores moleculares dialélicos (SNP1 e SNP2) em Y. Neste caso, há 9 possíveis classes genotípicas:  $a_1a_1-a_2a_2$ ,  $a_1a_1-A_2a_2$ ,  $a_1a_1-A_2A_2$ ,  $a_1A_1-a_2a_2$ ,  $a_1A_1-A_2a_2$ ,  $a_1A_1-A_2A_2$ ,  $A_1A_1-a_2a_2$ ,  $A_1A_1-A_2a_2$ ,  $A_1A_1-A_2A_2$

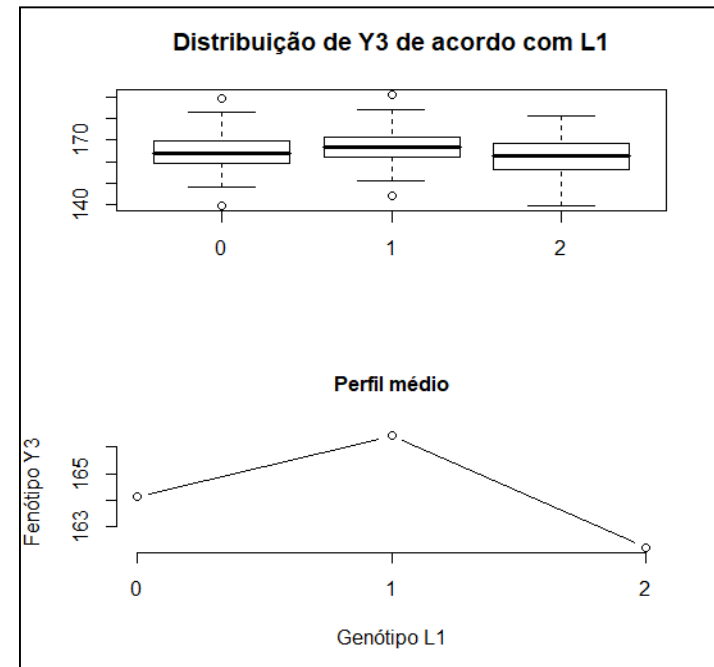
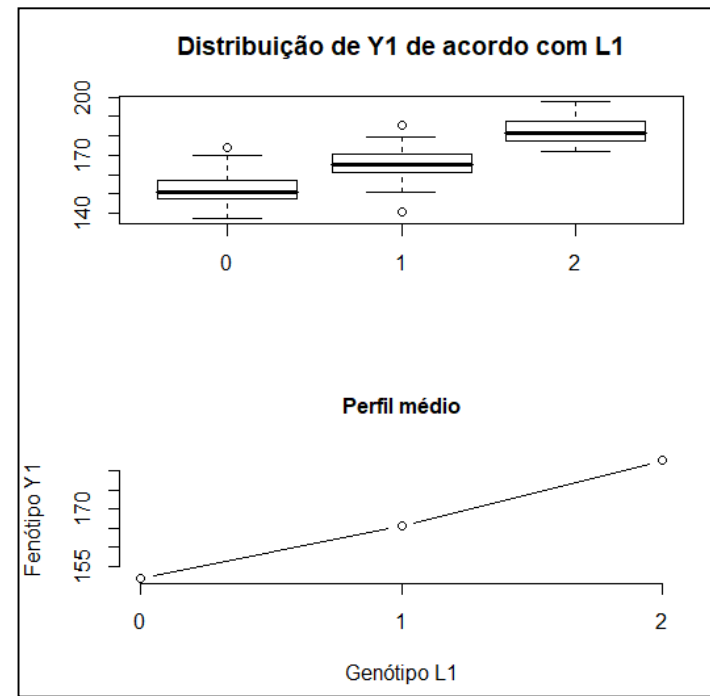
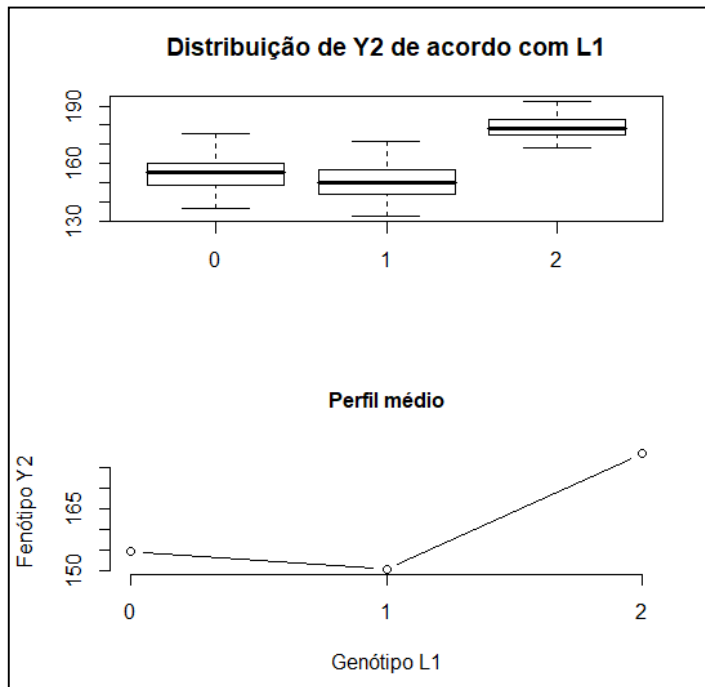


**DCA Fatorial 3x3: 8 graus de liberdade para estudar o efeito dos dois fatores (2 g.l. para o efeito principal do SNP1, 2 g.l. para o efeito principal do SNP2 e 4 g.l. para o efeito de interação)**



## Efeito do Fator gene (SNP) sobre Y:

Distribuição de diferentes fenótipos (Y1, Y2 e Y3) de acordo com o genótipo de um marcador molecular (SNP).



# Modelo DAC com um Fator em 3 Níveis

```
> phen2
      X      Y
[1,] 1 145.26
[2,] 0 154.42
[3,] 0 157.22
[4,] 0 139.25
[5,] 0 151.86
[6,] 0 156.80
[7,] 2 174.57
[8,] 0 167.70
[9,] 1 145.38
[10,] 0 146.48
...
[198,] 0 147.35
[199,] 1 150.46
[200,] 0 168.14
```

X: genótipo de um SNP (número de alelos de “risco”)  
X= 0 (aa), X=1 (Aa) e X=2 (AA)

Dados de pressão arterial (Y, em mmHg) e genótipo de um marcador SNP (variável X) para uma amostra aleatória de 200 pacientes.

Delineamento com indivíduos não relacionados (independentes) e variável de interesse quantitativa.

**Há evidência de efeito significativo do SNP sobre o fenótipo Y?**



Oficina no R: Uso de modelos ANOVA (DCA com um fator em 3 níveis) ou, equivalentemente, modelos de regressão, para testar o efeito de genes!

# Dados fenotípicos de acordo com genótipos de um Marcador

Um fator em 3 níveis

DCA?

Garantido pela Lei da

Aleatorização

Mendeliana.

	X	Y
[1,]	1	145.26
[2,]	0	154.42
[3,]	0	157.22
[4,]	0	139.25
[5,]	0	151.86
[6,]	0	156.80
[7,]	2	174.57
[8,]	0	167.70
[9,]	1	145.38
[10,]	0	146.48
...	...	...
[198,]	0	147.35
[199,]	1	150.46
[200,]	0	168.14

```
> phen2[phen2[,1]=='0',2]
 [1] 154.4237 157.2263 139.2510 151.8692 156.8054 167.7073 146.4859
 [8] 151.3634 149.6427 160.0789 147.6343 161.4621 158.3081 143.3823
[15] 175.4925 148.5633 144.3721 170.2776 163.2950 150.0984 150.0065
[22] 142.2762 152.3247 156.3284 156.8510 140.6139 149.2869 162.6455
[29] 165.0624 155.8352 155.8515 158.0890 152.3135 164.0242 156.1974
[36] 168.1741 152.9114 158.0893 145.1791 158.2748 140.9773 172.7436
[43] 169.1738 157.6185 146.0102 159.9995 138.7326 151.8952 168.7321
[50] 153.7044 169.5174 155.4367 146.3344 153.6227 155.8537 156.3379
[57] 147.4218 141.8795 159.3708 147.0500 158.8498 166.3101 148.5914
[64] 142.8224 143.8362 163.9651 150.0744 149.9921 141.7130 136.1313
[71] 144.4088 158.4019 150.2478 147.8996 157.0903 144.1187 155.6270
[78] 165.1502 144.0674 156.5341 164.2521 162.4175 169.1474 154.9353
[85] 166.0011 155.1834 159.0337 163.4208 150.9837 155.6688 150.6998
[92] 151.9377 155.8304 158.1600 151.5347 149.2445 161.9220 151.7951
[99] 147.3532 168.1490
> phen2[phen2[,1]=='1',2]
 [1] 145.2637 145.3843 137.5895 167.7748 156.9217 149.6335 135.7536
 [8] 139.1928 164.0785 142.3166 157.0479 144.1613 144.0071 148.3424
[15] 149.2578 146.5655 147.0282 138.2769 150.5127 139.7317 141.2262
[22] 143.0031 153.8796 132.9858 151.2089 151.8173 159.0893 165.5663
[29] 134.3992 153.9076 160.9058 156.5199 159.2101 147.5772 145.3183
[36] 171.8219 132.0346 147.4122 163.7076 159.4918 136.2419 146.1509
[43] 147.8516 150.7990 150.4148 142.6881 164.1797 147.5689 142.4199
[50] 139.2421 153.1704 150.8772 144.8755 147.9154 143.9406 146.5751
[57] 164.8583 161.4599 142.6514 141.5781 154.3927 157.8455 161.3908
[64] 136.3838 142.9046 146.7168 155.7499 151.8689 161.5959 153.3873
[71] 150.1952 165.5565 160.6819 156.1825 155.6439 155.0257 159.4550
[78] 155.9889 139.8693 149.9790 153.6754 150.4695
> phen2[phen2[,1]=='2',2]
 [1] 174.5761 168.2239 183.1205 175.2907 178.5674 183.6826 170.7715
 [8] 169.6804 181.9329 192.5263 184.3214 178.2891 174.8969 175.3306
[15] 182.8309 180.1844 180.2545 177.6766
```

Dados simulados

A: alelo de risco

$P(A)=0.3$

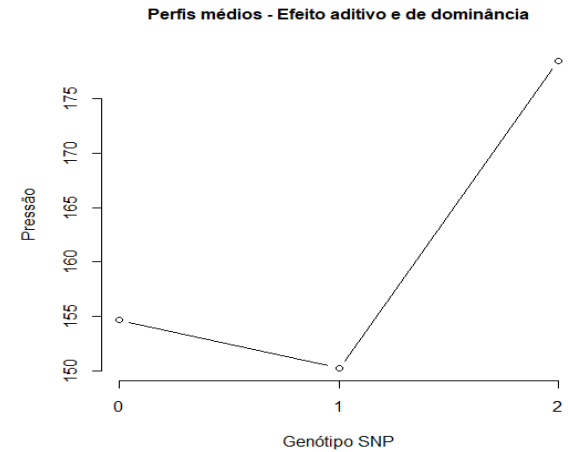
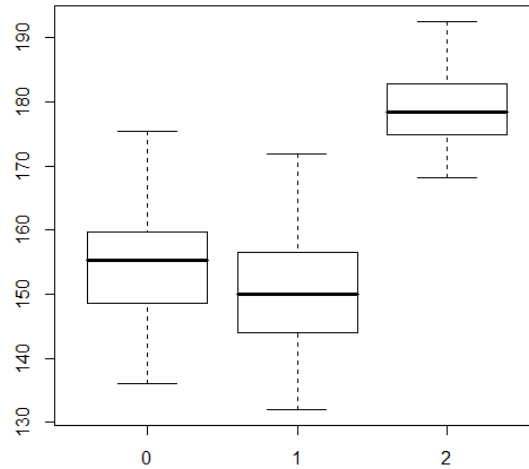
$P(aa)=0.3^2$ ,  $P(AA)=0.7^2$

$P(Aa)=2(0.3)(0.7)$ , sob

Equilíbrio de Hardy-Weinberg

# Efeitos Genéticos

	X	Y
[1,]	1	145.26
[2,]	0	154.42
[3,]	0	157.22
[4,]	0	139.25
[5,]	0	151.86
[6,]	0	156.80
[7,]	2	174.57
[8,]	0	167.70
[9,]	1	145.38
[10,]	0	146.48
...	...	...
[198,]	0	147.35
[199,]	1	150.46
[200,]	0	168.14



Modelo ANOVA: Um fator em 3 níveis (DCA??)

Tabela de ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SNP	2	11765	5882.7	81.418	< 2.2e-16
Resíduo	197	14234	72.3		

$$Y = X\beta + e$$

$$= \mu + \tau_1 X_1 + \tau_2 X_2 + e$$

Coefficients:

	Estimate	Std. Error	tvalue	Pr(> t )
(Intercept)	154.640	0.850	181.925	< 2e-16
(snp) 1	-4.392	1.266	-3.468	0.000644
(snp) 2	23.813	2.176	10.942	< 2e-16

No R: parametrização casela de referência

$$\hat{\mu}_{aa} = 154.64$$

$$\hat{\mu}_{Aa} = \hat{\mu}_{aa} + \hat{\tau}_1 = 150.25$$

$$\hat{\mu}_{AA} = \hat{\mu}_{aa} + \hat{\tau}_2 = 178.45$$

↑  $H_0 : \tau_j = 0$





# Distribuição de fenótipos (Y) de acordo com o genótipo de um marcador molecular (X).

Qual é o efeito do “gene” ?

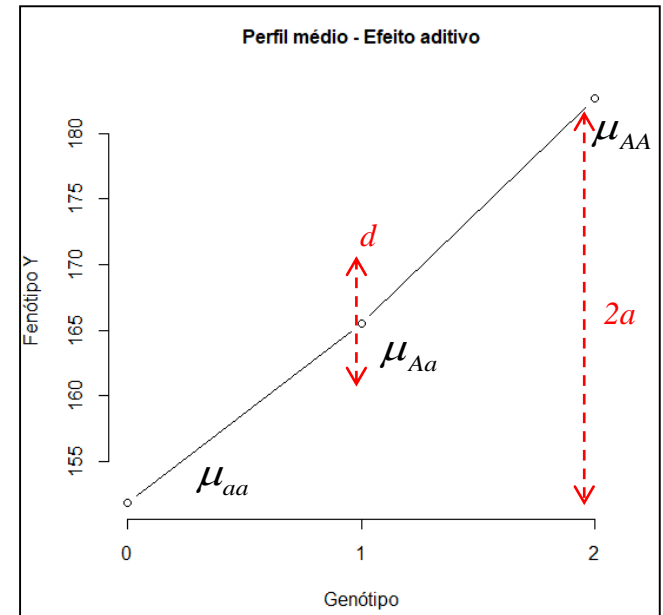
Genótipo	aa	Aa	AA
Média(Y)	$\mu_{aa}$	$\mu_{Aa}$	$\mu_{AA}$

$$a = \frac{\mu_{AA} - \mu_{aa}}{2}$$

Efeito aditivo

$$d = \mu_{Aa} - \frac{\mu_{AA} + \mu_{aa}}{2}$$

Efeito de dominância:  
desvio da aditividade



# Efeitos Genéticos

## ▪ Modelo estrutural e distribucional

$$y_i = \mu + aX_{ai} + dX_{di} + e_i; \quad e_i \sim N(0; \sigma^2)$$

$$X_{ai} = \begin{cases} -1 & \text{se } G=aa \\ 0 & \text{se } G=Aa \\ 1 & \text{se } G=AA \end{cases} \quad X_{di} = \begin{cases} 0 & \text{se } G=aa \\ 1 & \text{se } G=Aa \\ 0 & \text{se } G=AA \end{cases}$$

Parametrização conveniente que pode ser formulada via Modelos de Regressão para, diretamente, estimar e testar efeitos aditivo e de dominância do marcador.

$$E(y_i | G) = \begin{cases} \mu_{aa} = \mu - a & \text{se } G=aa \\ \mu_{Aa} = \mu + d & \text{se } G=Aa \\ \mu_{AA} = \mu + a & \text{se } G=AA \end{cases}$$

$$\Rightarrow a = \frac{\mu_{AA} - \mu_{aa}}{2}$$

**Efeito aditivo**

$$\Rightarrow d = \mu_{Aa} - \frac{\mu_{AA} + \mu_{aa}}{2}$$

**Efeito de dominância**

$$\Rightarrow \mu = \frac{\mu_{AA} + \mu_{aa}}{2}$$



# Mapeamento de ‘Genes’: Teste de Efeitos Genéticos

- **Modelo estrutural e distribucional: modelo aditivo (reduzido)**

$$y_i = \mu + aX_{ai} + e_i; \quad e_i \sim N(0; \sigma^2)$$

$$X_{ai} = \begin{cases} 0 & \text{se } G=aa \\ 1 & \text{se } G=Aa \\ 2 & \text{se } G=AA \end{cases}$$

$$E(y_i | G) = \begin{cases} \mu_{aa} = \mu & \text{se } G=aa \\ \mu_{Aa} = \mu + a & \text{se } G=Aa \\ \mu_{AA} = \mu + 2a & \text{se } G=AA \end{cases}$$

O Efeito linear ou aditivo do “gene” é avaliado **com 1 grau de liberdade** (o efeito de dominância está no erro)

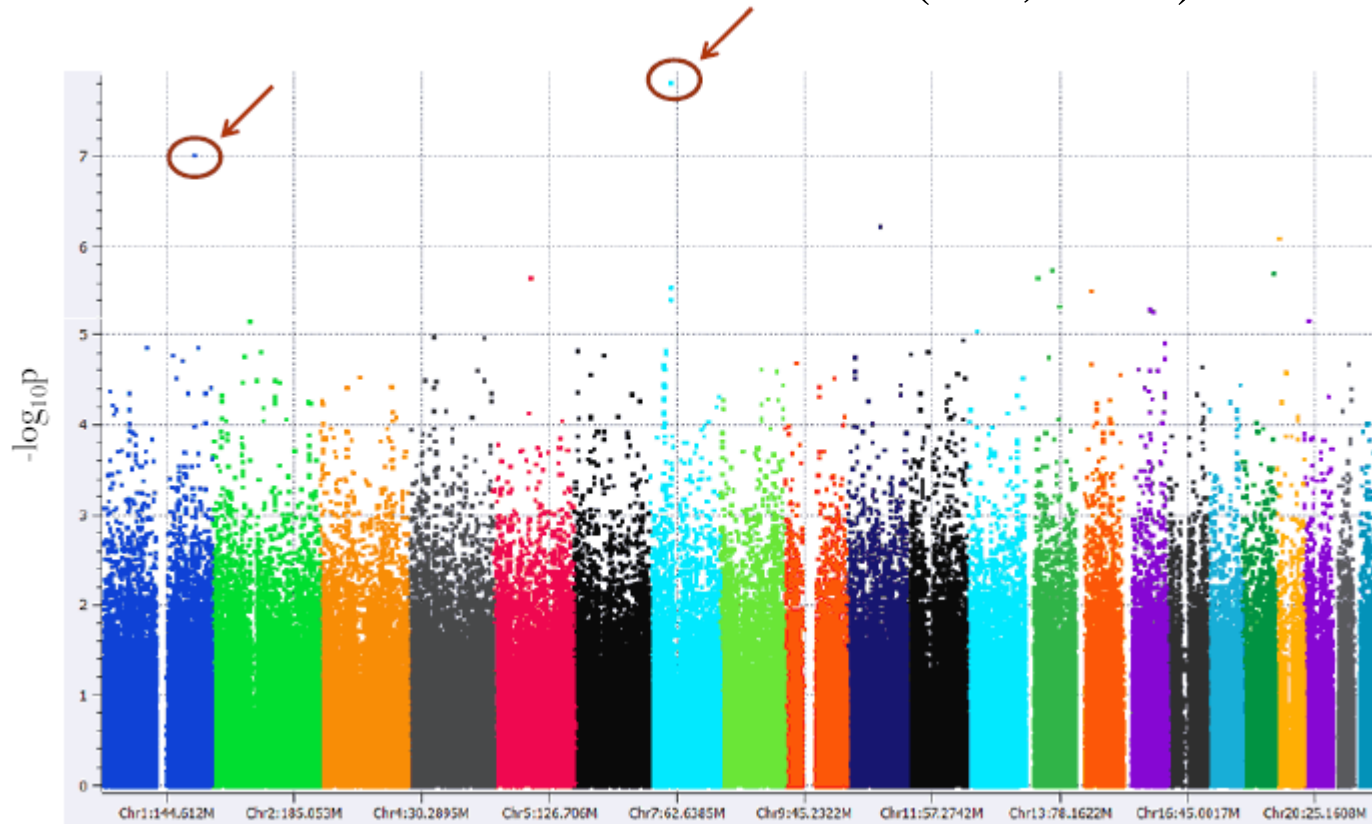
Pode ser usada a codificação de X como (-1, 0 e 1) ou (0, 1 e 2).

Interpretação: a cada aumento de uma cópia no alelo de risco, é esperada uma mudança de “a” unidades em Y.

**Teste do Efeito aditivo**  $\Rightarrow H_0 : a = 0 \quad \times \quad H_0 : a \neq 0$

# Mapeamento Genético

Gráfico Manhattan (Liu, 2011)



- Ajuste de *modelos aditivos unilocos* para estudar o efeito de milhares de marcadores
- Estatística de teste: F, t, Razão de Verossimilhança, Lod Score, dentre outras.
- $-\log(p\text{-valor})$  na ordenada e a posição de marcadores na abscissa
- Adotar correção para múltiplos testes (em geral, obtém-se poucos sinais significantes)

# Oficina: Simulação de Dados e Estimação de Efeitos Genéticos

## 1. Gerar dados de marcadores SNP em Equilíbrio de Hardy-Weinberg:

$$X_i \sim \text{Binomial}(n = 2; p = P(A))$$

$$X_i = \begin{cases} 0 \text{ se } aa & P(X_i = 0) = P(aa) = (1-p)^2 \\ 1 \text{ se } Aa & P(X_i = 1) = P(Aa) = 2p(1-p) \\ 2 \text{ se } AA & P(X_i = 2) = P(AA) = p^2 \end{cases}$$

## 2. Gerar variáveis quantitativas, Y, tal que:

$$y_i = \mu + aX_{ai} + dX_{di} + e_i; \quad e_i \stackrel{iid}{\sim} N(0; \sigma^2)$$

## 3. Realizar análises dos dados: ajustar modelos ANOVA e de Regressão

# Modelos Multilocos: Pares de Locos

	Loco2			
Loco1	<b>bb</b>	<b>Bb</b>	<b>BB</b>	
<b>aa</b>	$\mu_{aabb}$	$\mu_{aaBb}$	$\mu_{aaBB}$	$\mu_{aa.}$
<b>Aa</b>	$\mu_{Aabb}$	$\mu_{AaBb}$	$\mu_{AaBB}$	$\mu_{Aa.}$
<b>AA</b>	$\mu_{AAbb}$	$\mu_{AABb}$	$\mu_{AABB}$	$\mu_{AA.}$
	$\mu_{.bb}$	$\mu_{.Bb}$	$\mu_{.BB}$	$\mu_{..}$

Estrutura Fatorial 3x3:  
há 9 níveis dos fatores

8 graus de Liberdade para  
estudar os possíveis efeitos  
dos Marcadores

Moleculares: 2 g.l. para os  
efeitos principais e 4 g.l.  
para interações

## Efeitos Marginais (Aditivo e de Dominância):

$$a_1 = \frac{\mu_{AA} - \mu_{aa}}{2}; \quad d_1 = \mu_{Aa} - \frac{\mu_{AA} + \mu_{aa}}{2} \quad a_2 = \frac{\mu_{BB} - \mu_{bb}}{2}; \quad d_2 = \mu_{Bb} - \frac{\mu_{BB} + \mu_{bb}}{2}$$

## Teste do Efeito de Interação (Aditivo/Aditivo):

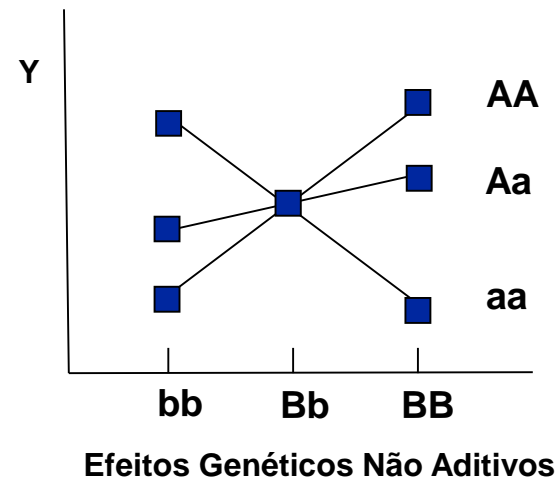
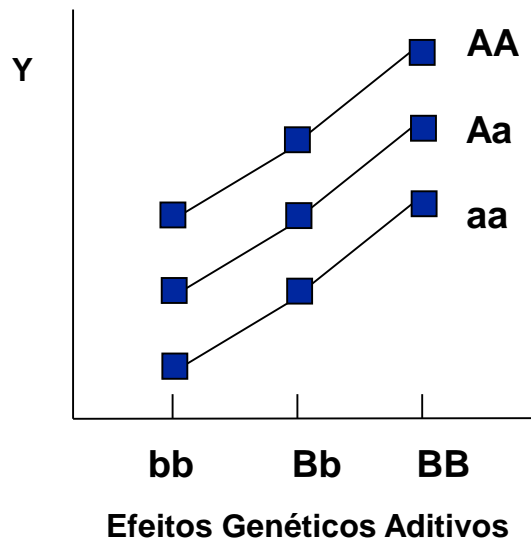
$$H_0 : \mu_{AAbb} - \mu_{aabb} = \mu_{AABB} - \mu_{aaBB}$$

$$\Rightarrow (\mu_{AAbb} - \mu_{aabb}) - (\mu_{AABB} - \mu_{aaBB}) = 0$$

Há outros efeitos de interação  
que podem ser definidos  
(4 graus de Liberdade para os  
possíveis efeitos de interação)

# Modelos Bilocos (Pares de Locos): Fatorial 3x3

**Efeito de “Interação” entre os 2 Fatores (cada um em 3 níveis):  
visualização por meio de gráficos de perfis de médias**



# Modelos para Pares de Locos

$$y = \mu + \underbrace{a_1 X_{a_1} + d_1 X_{d_1}}_{\text{Loco1}} + \underbrace{a_2 X_{a_2} + d_2 X_{d_2}}_{\text{Loco2}} + \underbrace{i_{aa} X_{a_1} * X_{a_2}}_{\text{Interação nos ef. aditivos}} + \underbrace{i_{dd} X_{d_1} * X_{d_2}}_{\text{Interação nos ef. de dominância}} + \underbrace{i_{ad} X_{a_1} * X_{d_2} + i_{da} X_{d_1} * X_{a_2}}_{\text{Interação nos ef. aditivo e de dominância}} + \varepsilon$$

⇒ **Modelo com 2 Locos: Fatorial 3x3 (8 graus de liberdade)**

⇒ **Suposição de Marcadores não correlacionados: em Genética denomina-se, Loco1 e Loco2 em “Equilíbrio de Ligação (LD)”**

⇒ **Interação entre fatores: em Genética este é Epistasia entre Locos**

## Oficina no R:

Gerar  $n=200$  observações  
de dados genotípicos e fenotípicos

#parâmetros fixados

#Probabilidade do alelo de risco

Loco1:  $P(A) = 0.3$

Loco2:  $P(A) = 0.4$

#Distribuição do fenótipo

$\mu_i=165$

$aL1=15$ ,  $dL1=0$

$aL2=10$ ,  $dL2=0$

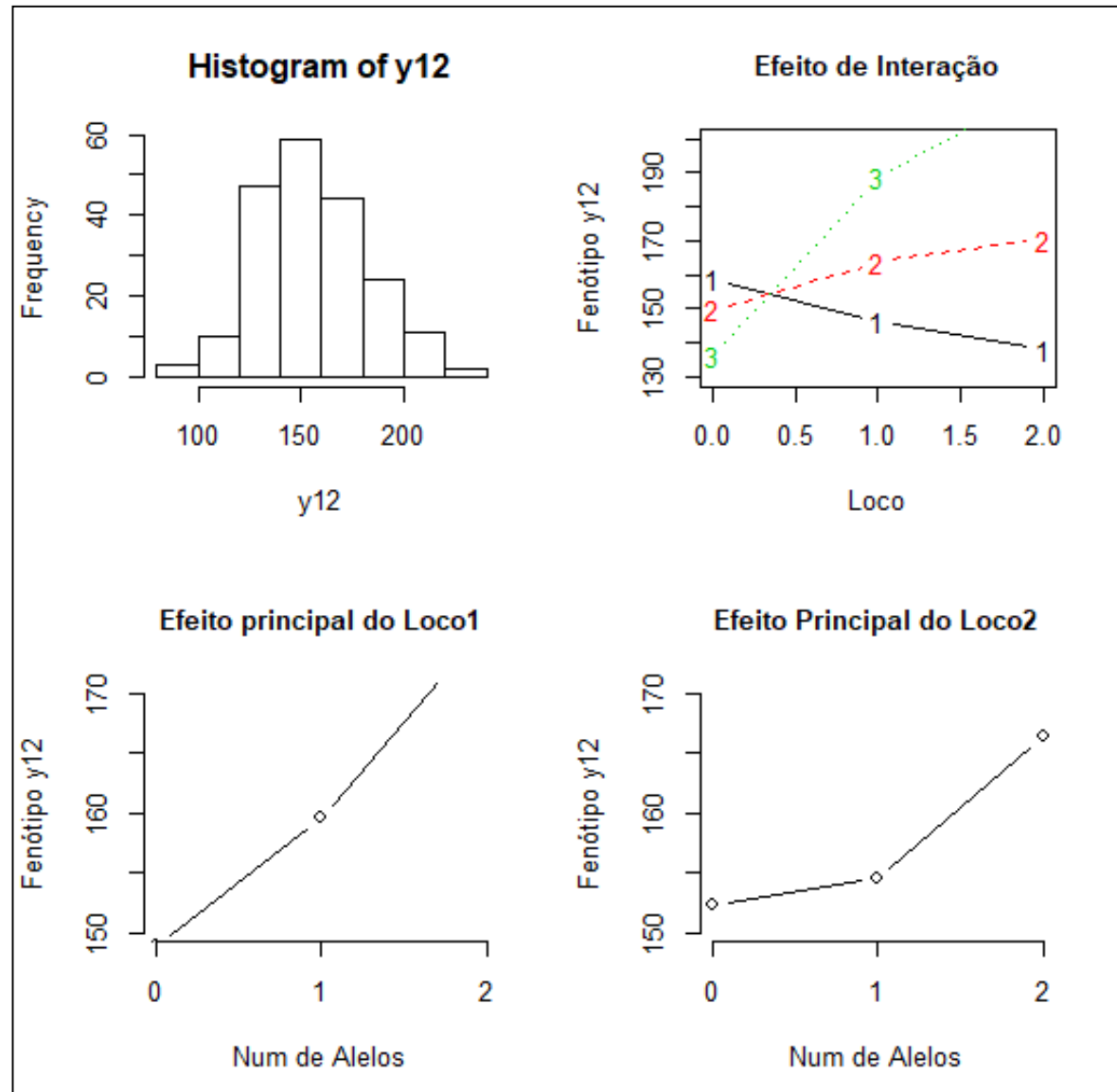
$aL1aL2=20$ ,  $aL1dL2=0$

$dL1aL2=0$ ,  $dL1dL2=0$

$s^2=25$

```
> perf12
```

```
      [,1]      [,2]      [,3]  
[1,] 158.8418 149.7181 136.3804  
[2,] 146.1166 163.7302 188.5351  
[3,] 138.2871 170.4384 215.1540
```



## Oficina no R:

```
> #modelo completo
> fit12.c<- lm(y12 ~
xaL1+xdL1+xaL2+xdL2+xaL1*xaL2+xaL1*xdL2+xdL1*xaL2+xdL1*xdL2)

> anova(fit12.c)
Analysis of Variance Table


```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
xaL1	1	12760	12760.3	24.7958	1.420e-06	***
xdL1	1	244	244.5	0.4751	0.49151	
xaL2	1	3421	3420.7	6.6470	0.01069	*
xdL2	1	89	89.2	0.1734	0.67759	
xaL1:xaL2	1	25155	25155.2	48.8815	4.435e-11	***
xaL1:xdL2	1	787	787.2	1.5296	0.21769	
xdL1:xaL2	1	378	378.2	0.7349	0.39238	
xdL1:xdL2	1	18	17.8	0.0345	0.85274	
Residuals	191	98292	514.6			

```

> summary(fit12.c)
Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	162.166	3.881	41.787	< 2e-16	***
xaL1	13.601	3.881	3.505	0.000569	***
xdL1	-2.088	5.162	-0.404	0.686378	
xaL2	14.555	3.881	3.750	0.000234	***
xdL2	5.160	6.217	0.830	0.407559	
xaL1:xaL2	24.832	3.881	6.399	1.18e-09	***
xaL1:xdL2	7.608	6.217	1.224	0.222542	
xdL1:xaL2	-4.195	5.162	-0.813	0.417494	
xdL1:xdL2	-1.508	8.114	-0.186	0.852743	

*Partição dos 8 graus  
de Liberdade em  
efeitos específicos!*



## Oficina no R:

```
> #modelo reduzido
> fit12.a12 <- lm(y12 ~ xaL1 + xaL2 + xaL1*xaL2)

> anova(fit12.a12)
Analysis of Variance Table

Response: y12
          Df Sum Sq Mean Sq F value    Pr(>F)
xaL1       1  12760  12760.3  25.0966 1.214e-06 ***
xaL2       1   3461   3460.5   6.8061 0.009785 **
xaL1:xaL2  1  25268  25268.1  49.6964 2.993e-11 ***

> summary(fit12.a12)
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept)  162.506     1.927   84.312 < 2e-16 ***
xaL1         15.959     2.513    6.351 1.46e-09 ***
xaL2         12.488     2.489    5.016 1.18e-06 ***
xaL1:xaL2    24.521     3.478    7.050 2.99e-11 ***
```

## Oficina no R - Ainda, considerando os dados de Y12, interprete os seguintes ajustes:

```
> lm(formula = y12 ~ as.factor(SNP1):as.factor(SNP2))
```

### Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(SNP1):as.factor(SNP2)	8	51395	6424.4	452.71	< 2.2e-16 ***
Residuals	191	2710	14.2		

```
> Coefficients: (1 not defined because of singularities)
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	209.234	1.424	146.95	<2e-16
as.factor(SNP1)0:as.factor(SNP2)0	-47.502	2.014	-23.59	<2e-16
as.factor(SNP1)1:as.factor(SNP2)0	-53.721	2.014	-26.68	<2e-16
as.factor(SNP1)2:as.factor(SNP2)0	-61.139	2.600	-23.52	<2e-16
as.factor(SNP1)0:as.factor(SNP2)1	-59.255	1.541	-38.46	<2e-16
as.factor(SNP1)1:as.factor(SNP2)1	-44.349	1.526	-29.06	<2e-16
as.factor(SNP1)2:as.factor(SNP2)1	-32.355	2.206	-14.67	<2e-16
as.factor(SNP1)0:as.factor(SNP2)2	-69.698	1.522	-45.79	<2e-16
as.factor(SNP1)1:as.factor(SNP2)2	-34.724	1.564	-22.21	<2e-16
as.factor(SNP1)2:as.factor(SNP2)2	NA	NA	NA	NA

Parametrização casela de referência!

Interprete os coeficientes!

## Oficina no R - Ainda, considerando os dados de Y12, interprete os seguintes ajustes:

```
> lm(formula = y12 ~ as.factor(SNP1) + as.factor(SNP2) + as.factor(SNP1) *  
as.factor(SNP2))
```

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
as.factor(SNP1)	2	35494	17747.2	1250.599	< 2.2e-16	***
as.factor(SNP2)	2	553	276.3	19.469	2.017e-08	***
as.factor(SNP1):as.factor(SNP2)	4	15348	3837.0	270.385	< 2.2e-16	***
Residuals	191	2710	14.2			

---

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

> **Coefficients:**

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	161.731	1.424	113.589	< 2e-16
as.factor(SNP1)1	-6.219	2.014	-3.088	0.00231
as.factor(SNP1)2	-13.636	2.600	-5.246	4.11e-07
as.factor(SNP2)1	-11.752	1.541	-7.628	1.09e-12
as.factor(SNP2)2	-22.196	1.522	-14.582	< 2e-16
as.factor(SNP1)1:as.factor(SNP2)1	21.124	2.169	9.741	< 2e-16
as.factor(SNP1)2:as.factor(SNP2)1	40.536	3.153	12.856	< 2e-16
as.factor(SNP1)1:as.factor(SNP2)2	41.193	2.182	18.878	< 2e-16
as.factor(SNP1)2:as.factor(SNP2)2	83.335	3.012	27.664	< 2e-16

# Modelos para Pares de Locos

Médias fenotípicas de acordo com o genótipo conjunto

	P1		
	bb	Bb	BB
aa	20	18	16
Aa	17	15	13
AA	14	12	10

	P2		
	bb	Bb	BB
aa	20	20	16
Aa	20	20	16
AA	14	14	10

	P3		
	bb	Bb	BB
aa	20	20	10
Aa	20	20	10
AA	10	10	10

	P4		
	bb	Bb	BB
aa	38.7	6.6	3.2
Aa	4.0	2.0	1.3
AA	2.7	3.4	1.6

	P5		
	aa	Aa	AA
aa	20	19	16
Aa	19	18	15
AA	14	13	10

	P6		
	aa	Aa	AA
aa	20	21	16
Aa	22	23	18
AA	14	15	10

	P7		
	aa	Aa	AA
aa	20	20	20
Aa	20	20	20
AA	20	20	10

- P1: Ação gênica intermediária
- P2: Dominância completa
- P3: Ação gênica complementar
- P4: Epistasia complexa
- P5: Dominância parcial
- P6: Superdominância
- P7: Ação gênica duplicada

**Exercício da Lista!**

# Inferência sobre Efeitos Genéticos

## **Modelos Unilocos: Um fator em 3 níveis**

- Até 2 graus de liberdade para estudar o efeito do Loco

## **Modelos Bilocos: Fatorial $3 \times 3 = 3^2$**

- Até 8 graus de liberdade para estudar o efeito do Loco



## **Inferências:**

- Sob premissas clássicas: modelos ANOVA
- Soluções Não-Paramétricas: Kruskal-Wallis
- Aleatorização Mendeliana (mimetiza um DCA). Logo pode ser usado Testes de Aleatorização

# Modelos mais Gerais

$$Y_{1 \times 1} = \mu + X_{G1} \beta_{G1} + e$$

$$Y_{1 \times 1} = \mu + X_{G1} \beta_{G1} + X_E \beta_E + e$$

$$Y_{1 \times 1} = \mu + X_{G1} \beta_{G1} + X_E \beta_E + (X_E * X_{G1}) \beta_{EG} + e$$

$$Y_{1 \times 1} = \mu + X_{G1} \beta_1 + X_{G2} \beta_2 + (X_{G1} * X_{G2}) \beta_{12} + X_E \beta_E + e$$

Modelos univariados ( $Y_1$ )


$$(Y_1 \quad Y_2)_{n \times 2} = (\mu_1 \quad \mu_2) + X_{G1} (\beta_{11} \quad \beta_{12}) + (e_1 \quad e_2)$$

Modelos multivariados ( $Y_1, Y_2, \dots, Y_p$ )

Modelos considerando o “gene” como um fator de efeito fixo sobre Y

⇒ Modelos mais gerais podem ser definidos com o “gene” como aleatório!

# Modelos mais Gerais: Efeitos Genéticos Ajustados


$$Y_i = \mu + aX_{Gi} + e_i$$

Modelo reduzido, sem efeito de “covariáveis”

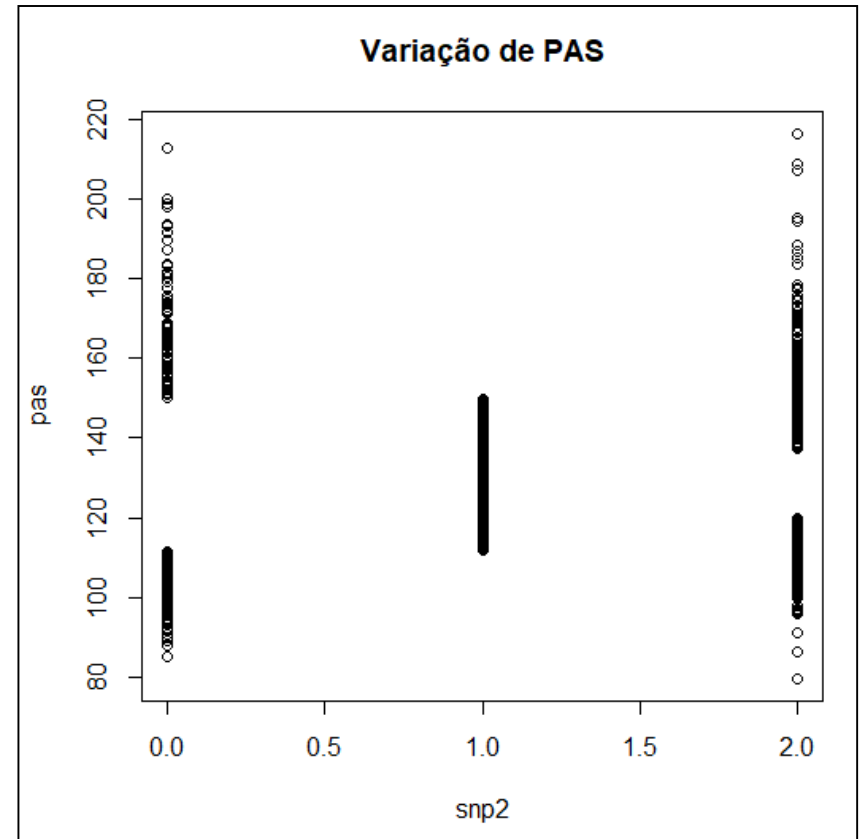
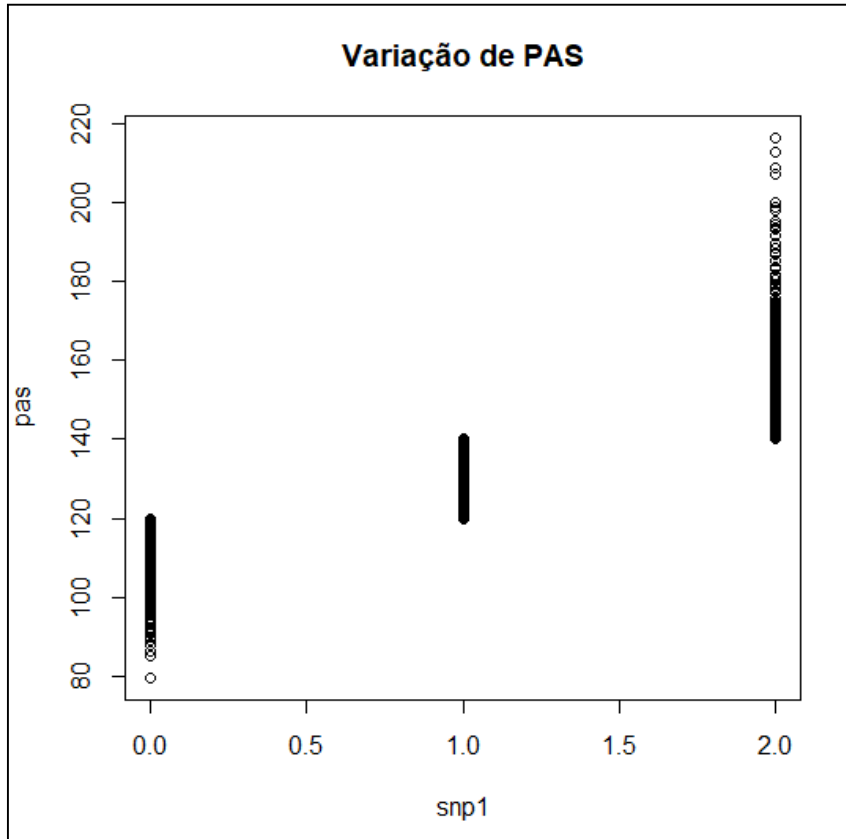
$$Y_i = \mu + aX_{Gi} + \beta_{Sexo} X_{Sexoi} + e_i$$

Modelo (complete) com efeito de covariáveis

**Avaliar o efeito da variável de interesse ( $X_G$ ) sobre um fenótipo ( $Y$ ), controlando (ajustando) pelo efeito de variáveis como o Sexo ( $X_{Sexo}$ ) ou outras covariáveis (idade, peso, etc.)**

# Padrão de Variação Fenotípica

Dados simulados:  $Y_{n \times 1} = \mu + X_G \beta_G + e$



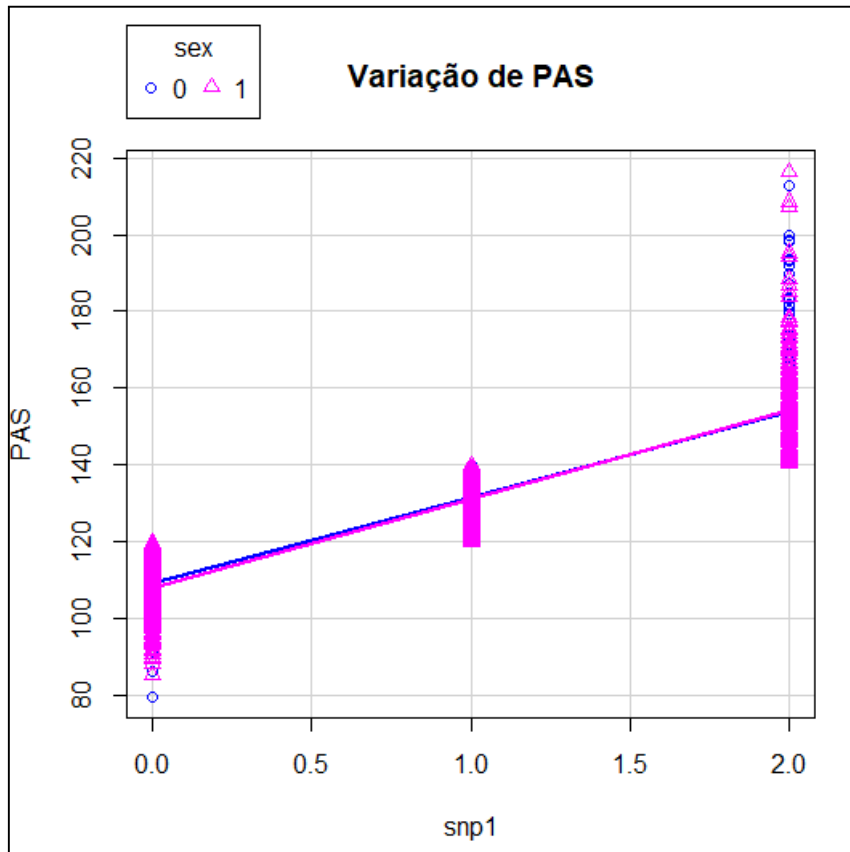
```
> fit1 <- lm(pas ~ snp1, dat=bd)
      Estimate Std.Error tvalue Pr(>|t|)
Mi    108.4466    0.3138   345.60 <2e-16
snp1   22.8416    0.2833   80.62  <2e-16
```

```
> fit2 <- lm(pas ~ snp2, dat=bd)
      Estimate Std.Error tvalue Pr(>|t|)
Mi    120.4360    0.8179  147.256 <2e-16
snp2    6.3436    0.6629   9.569  <2e-16
```

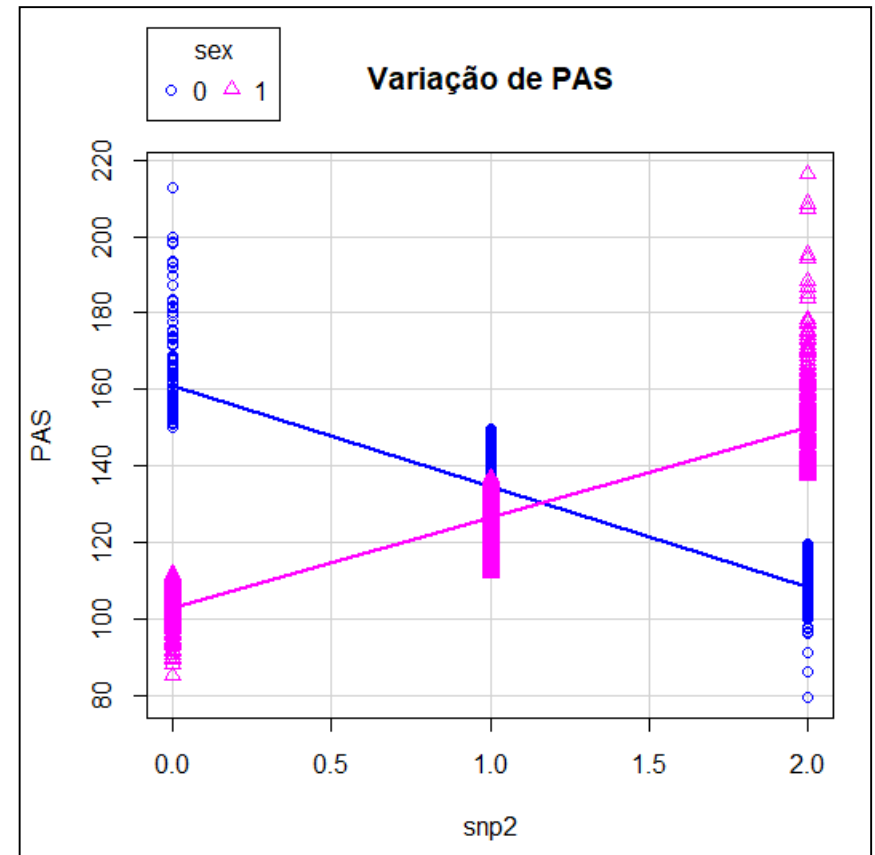


# Efeito de Interação Gene\*Sexo

$$Y_{n \times 1} = \mu + X_{sex} \beta_{sex} + X_G \beta_G + (X_{sex} * X_G) \beta_{G1*sex} + e$$

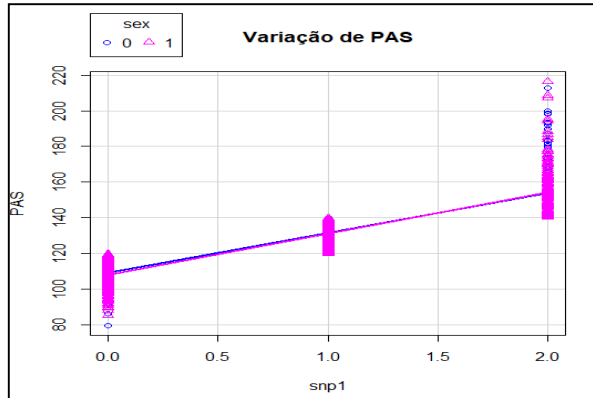


	Estimate	Std.Error	tvalue	Pr(> t )
Mi	109.0887	0.5359	203.573	<2e-16
Sex	-0.9778	0.6611	-1.479	0.139
<b>snp1</b>	<b>22.4298</b>	<b>0.4503</b>	<b>49.805</b>	<b>&lt;2e-16</b>
Sex:snp1	0.6127	0.5843	1.048	0.295



	Estimate	Std.Error	tvalue	Pr(> t )
Mi	160.8702	0.7186	223.86	<2e-16
Sex	-58.0629	0.8607	-67.46	<2e-16
snp2	-26.2363	0.5439	-48.24	<2e-16
<b>Sex:snp2</b>	<b>49.9609</b>	<b>0.6798</b>	<b>73.49</b>	<b>&lt;2e-16</b>

# Efeito de Interação Gene\*Sexo



	Estimate	Std.Error	tvalue	Pr(> t )
Mi	109.0887	0.5359	203.573	<2e-16
Sex	-0.9778	0.6611	-1.479	0.139
<b>snp1</b>	<b>22.4298</b>	<b>0.4503</b>	<b>49.805</b>	<b>&lt;2e-16</b>
Sex:snp1	0.6127	0.5843	1.048	0.295

Considerando indivíduos do mesmo sexo, para cada aumento de um alelo de risco, há um aumento esperado na pas de 22,43mmHg

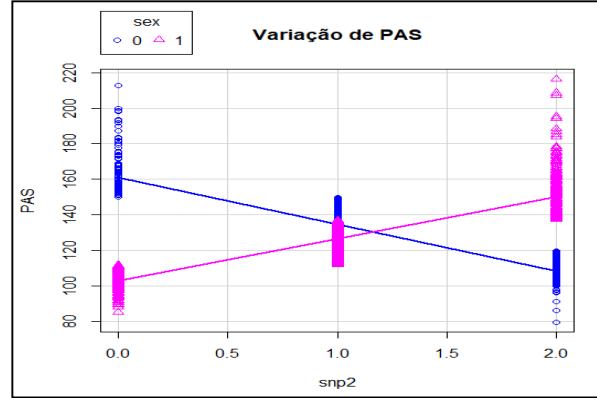
Modelo reduzido:

	Estimate	Std.Error	tvalue	Pr(> t )
Mi	108.4466	0.3138	345.60	<2e-16
<b>snp1</b>	<b>22.8416</b>	<b>0.2833</b>	<b>80.62</b>	<b>&lt;2e-16</b>

$$\hat{y}|_{G1} = 108.45 + 22.84 X_{G1}$$

$$\hat{y}|_{G1=0} = 108.45 \quad \hat{y}|_{G1=1} = 131.29 \quad \hat{y}|_{G1=2} = 154.13$$

valor esperado de pas para indivíduos que carregam 0, 1 ou 2 cópias do alelo de risco, respectivamente!



	Estimate	Std.Error	tvalue	Pr(> t )
Mi	160.8702	0.7186	223.86	<2e-16
Sex	-58.0629	0.8607	-67.46	<2e-16
snp2	-26.2363	0.5439	-48.24	<2e-16
<b>Sex:snp2</b>	<b>49.9609</b>	<b>0.6798</b>	<b>73.49</b>	<b>&lt;2e-16</b>

$$\hat{y}|_{Sex,G2} = 160.87 - 58.06 X_{sex} - 26.24 X_{G2} + 49.96 (X_{sex} * X_{G2})$$

$$\hat{y}|_{M,G2} = 160.87 - 26.24 X_{G2}$$

$$\hat{y}|_{F,G2} = (160.87 - 58.06) + (-26.24 + 49.96) X_{G2}$$

$$= 102.81 + 23.71 X_{G2}$$

$$\hat{y}|_{M,0} = 160.87 \quad \hat{y}|_{M,1} = 134.63 \quad \hat{y}|_{M,2} = 108.39$$

$$\hat{y}|_{F,0} = 102.81 \quad \hat{y}|_{F,1} = 126.53 \quad \hat{y}|_{F,2} = 150.25$$

Considerando indivíduos Masculinos, para cada aumento de um alelo de risco, há uma redução esperada na pas de 26,24mmHg. Já para femininos, há um aumento esperado na pas de 23,71mmHg.