Unidade 5. Análise de dados de classificação simples e dupla

5.1. Introdução e modelos estatísticos

5.2. Parâmetros do modelo de classificação simples e inferências sobre esses parâmetros

5.3. Parâmetros do modelo de classificação dupla e inferências sobre esses parâmetros

5.4. Discriminação da variação de tratamento: testes de comparações múltiplas: teste DMS de Fisher e teste de Tukey

5.5. Uso de programa estatístico para processamento das análises

Modelos de classificação dupla

Modelos de classificação dupla são aqueles que exprimem a relação entre uma variável resposta e dois fatores de tratamento.

Por que incluir mais um fator no experimento?

- aumentar a amplitude das conclusões
- estudar a influência de um fator sobre o outro (interação)

Os experimentos com dois ou mais fatores de tratamentos são denominados fatoriais e usualmente são assim representados:

Fatorial 2 X 2 $\rightarrow$ experimento com dois fatores, cada um com 2 níveis (mais simples)

Fatorial 2 X 3 $\rightarrow$ experimento com dois fatores, um com 2 e outro com 3 níveis

Fatorial 2 X 3 X 4 $\rightarrow$ experimento com três fatores, um com 2, outro com 3 e outro com 4 níveis

Experimento com dois fatores

Exemplo: Programas de computador modernos requerem capacidade de acesso rápido aos dados. Uma pesquisa foi conduzida para estudar o efeito do *tamanho do arquivo* de dados (Pequeno; Médio e Grande) e do *tamanho do buffer* (área de memória intermediária que serve para acelerar o acesso a dados que estão sendo transferidos entre dispositivos - memória de disco e memória RAM - que operam com velocidades diferentes), (20 kb e 40 kb) sobre a variável *tempo de acesso* aos arquivos, medido através do *tempo de leitura em milissegundos*.

Tamanho do arquivo	Tamanho do buffer	
	20	40
Pequeno	2,05; 2,04; 2,21; 2,12	2,32; 2,31; 2,48; 2,42
Médio	2,24; 2,21; 2,23; 2,09	2,52; 2,62; 2,57; 2,61
Grande	2,08; 2,34; 2,33; 2,24	2,71; 2,73; 2,90; 2,72

Experimento com dois fatores

Exemplo: Programas de computador modernos requerem capacidade de acesso rápido aos dados. Uma pesquisa foi conduzida para estudar o efeito do *tamanho do arquivo* de dados (Pequeno; Médio e Grande) e do *tamanho do buffer* (área de memória intermediária que serve para acelerar o acesso a dados que estão sendo transferidos entre dispositivos - memória de disco e memória RAM - que operam com velocidades diferentes), (20 kb e 40 kb) sobre a variável *tempo de acesso* aos arquivos, medido através do *tempo de leitura em milissegundos*.

Fator A: Tamanho do arquivo (pequeno, médio e grande)

Fator B: Tamanho do buffer (20 kb e 40 kb)

Tratamentos: combinações de níveis dos fatores A e B

Variável resposta (y): tempo de acesso aos arquivos (ms)

Objetivo: estudar os efeitos dos fatores tamanho do arquivo e tamanho do buffer sobre o tempo de acesso aos arquivos e a interação entre esses fatores

Estrutura dos dados

Os experimentos de classificação apresentam dois fatores de classificação (ou de agrupamento) das unidades de observação.

Tamanho do arquivo	Tamanho do buffer	
	20	40
P	y_{P21}; y_{P22}; y_{P23}; y_{P24}	y_{P41}; y_{P42}; y_{P43}; y_{P44}
M	y_{M21}; y_{M22}; y_{M23}; y_{M24}	y_{M41}; y_{M42}; y_{M43}; y_{M44}
G	y_{G21}; y_{G22}; y_{G23}; y_{G24}	y_{G41}; y_{G42}; y_{G43}; y_{G44}

Fator Tratamento, com $t = 6$ níveis : T = {P2, P4, M2, M4, G2, G4}

Fator A, com $n_a = 3$ níveis : A = {P, M, G}

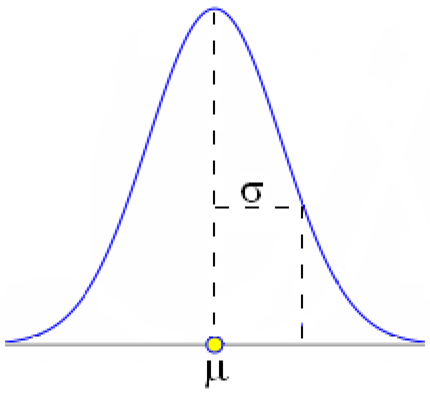
Fator B, com $n_b = 2$ níveis : B = {20, 40}

Número de repetições é constante para todos os tratamentos, $r = 4$

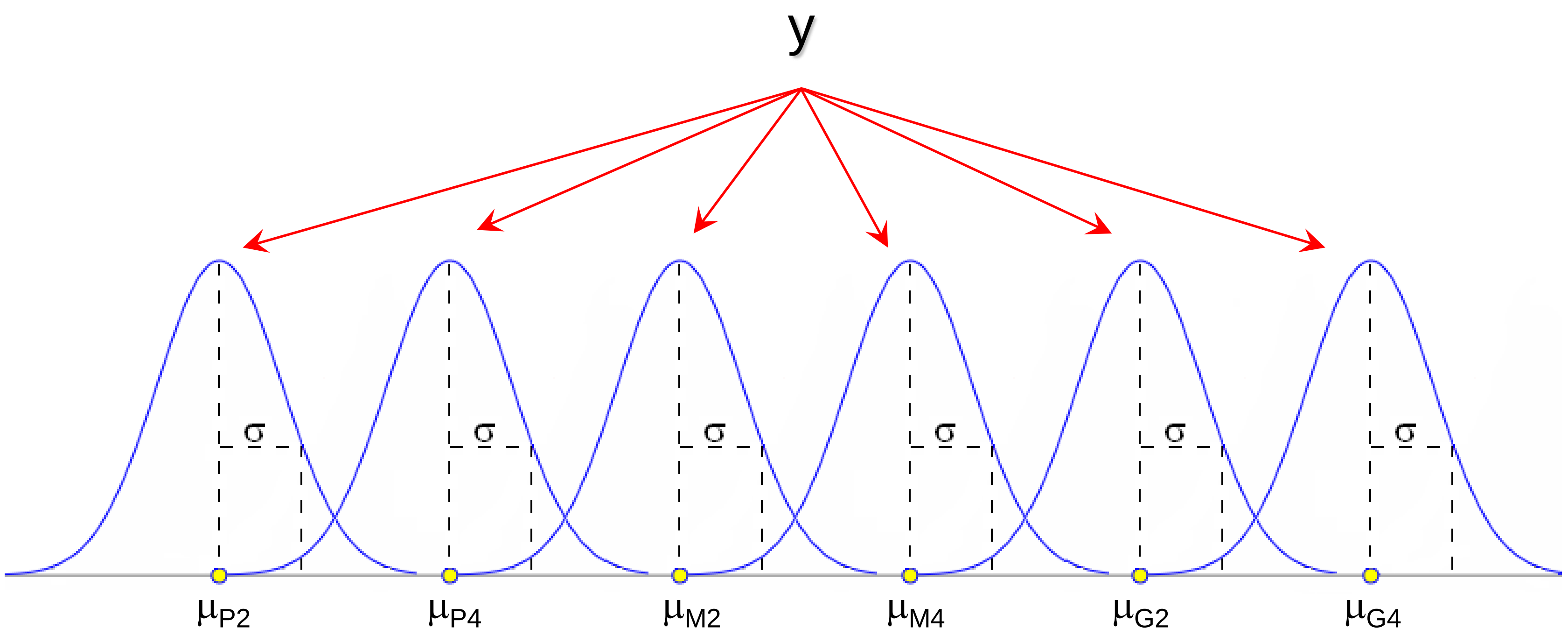
Populações estatísticas

y: variável resposta

y~N(μ, σ)



Tratamentos distintos geram populações de valores de y distintas?



Médias esperadas (populacionais)

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	μ_{P2}	μ_{P4}	μ_{P}
M	μ_{M2}	μ_{M4}	μ_{M}
G	μ_{G2}	μ_{G4}	μ_{G}
Médias marginais	μ_{2}	μ_{4}	μ

μ_{P2} $\rightarrow$ tempo médio esperado para arquivo pequeno e buffer de 20 kb

μ_{P4} $\rightarrow$ tempo médio esperado para arquivo pequeno e buffer de 40 kb

μ_{M2} $\rightarrow$ tempo médio esperado para arquivo médio e buffer de 20 kb

μ_{M4} $\rightarrow$ tempo médio esperado para arquivo médio e buffer de 40 kb

μ_{G2} $\rightarrow$ tempo médio esperado para arquivo grande e buffer de 20 kb

μ_{G4} $\rightarrow$ tempo médio esperado para arquivo grande e buffer de 40 kb

Médias observadas (amostrais)

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	$\bar{y}_{P2}$	$\bar{y}_{P4}$	$\bar{y}_{P}$
M	$\bar{y}_{M2}$	$\bar{y}_{M4}$	$\bar{y}_{M}$
G	$\bar{y}_{G2}$	$\bar{y}_{G4}$	$\bar{y}_{G}$
Médias marginais	$\bar{y}_{2}$	$\bar{y}_{4}$	$\bar{y}$

Estrutura dos dados (generalização)

Os experimentos de classificação apresentam dois fatores de classificação (ou de agrupamento) das unidades de observação.

Fator A (i)	Fator B (j)			
	B_1	B_2	...	B_{n_b}
A_1	$y_{111}, y_{112}, ..., y_{11r}$	$y_{121}, y_{122}, ..., y_{12r}$	...	$y_{1n_b1}, y_{1n_b2}, ..., y_{1n_br}$
A_2	$y_{211}, y_{212}, ..., y_{21r}$	$y_{221}, y_{222}, ..., y_{22r}$	...	$y_{2n_b1}, y_{2n_b2}, ..., y_{2n_br}$
...	...	...	y_{ijk}	...
A_{n_a}	$y_{n_a11}, y_{n_a12}, ..., y_{n_a1r}$	$y_{n_a21}, y_{n_a22}, ..., y_{n_a2r}$	...	$y_{n_an_b1}, y_{n_an_b2}, ..., y_{n_an_br}$

Fator A, com n_a níveis: $\{A_1, A_2, ..., A_{na}\}$, i = nível do Fator A, sendo i=1,2, ..., n_a

Fator B, com n_b níveis: $\{B_1, B_2, ..., B_{nb}\}$, j = nível do Fator B, sendo j=1,2, ..., n_b

k = número de repetições da combinação de níveis, sendo k = 1, 2, ..., r

Notação geral: Médias esperadas (populacionais)

Fator A (i)	Fator B (j)				Médias marginais
	B_1	B_2	...	B_{n_b}	
A_1	μ_{11}	μ_{12}	...	μ_{1n_a}	$\mu_{1.}$
A_2	μ_{21}	μ_{22}	...	μ_{2n_a}	$\mu_{2.}$
...	...	...	μ_{ij}	...	...
A_{n_a}	$\mu_{n_a 1}$	$\mu_{n_a 2}$	...	$\mu_{n_a n_b}$	$\mu_{n_a .}$
Médias marginais	$\mu_{.1}$	$\mu_{.2}$	...	$\mu_{.n_b}$	μ

μ_{ij} = média populacional da combinação de níveis ij

$\mu_{i.}$ = média populacional do nível i do Fator A

$\mu_{.j}$ = média populacional do nível j do Fator B

μ = média populacional sem efeito

Notação geral: Médias observadas (amostrais)

Fator A (i)	Fator B (j)				Médias marginais
	B_1	B_2	...	B_{n_b}	
A_1	$\bar{y}_{11}$	$\bar{y}_{12}$	...	$\bar{y}_{1n_a}$	$\bar{y}_{1.}$
A_2	$\bar{y}_{21}$	$\bar{y}_{22}$	...	$\bar{y}_{2n_a}$	$\bar{y}_{2.}$
...	...	...	$\bar{y}_{ij}$	...	...
A_{n_a}	$\bar{y}_{n_a1}$	$\bar{y}_{n_a2}$	...	$\bar{y}_{n_an_b}$	$\bar{y}_{n_a.}$
Médias marginais	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.n_b}$	$\bar{y}$

$\bar{y}_{ij}$ = média da combinação do nível i do Fator A com o nível j do fator B

$\bar{y}_{i.}$ = média do nível i do Fator A

$\bar{y}_{.j}$ = média do nível j do Fator B

$\bar{y}$ = média geral

Análise de dados de classificação simples e dupla

Pesquisas em que as variáveis preditoras são, necessariamente, do tipo fator, ou seja, são variáveis categorizadas que possibilitam a classificação ou agrupamento das unidades de pesquisa.

Um fator de interesse $\rightarrow$ modelo de classificação simples

Dois fatores de interesse $\rightarrow$ modelo de classificação dupla

Modelos de classificação dupla são aqueles que exprimem a relação entre uma variável resposta e dois fatores de tratamento.

Por que incluir mais um fator no experimento?

- aumentar a amplitude das conclusões
- estudar a influência de um fator sobre o outro (interação)

Exemplo: Programas de computador modernos requerem capacidade de acesso rápido aos dados. Uma pesquisa foi conduzida para estudar o efeito do *tamanho do arquivo* de dados (Pequeno; Médio e Grande) e do *tamanho do buffer* (área de memória intermediária que serve para acelerar o acesso a dados que estão sendo transferidos entre dispositivos - memória de disco e memória RAM - que operam com velocidades diferentes), (20 kb e 40 kb) sobre a variável *tempo de acesso* aos arquivos, medido através do *tempo de leitura em milissegundos*.

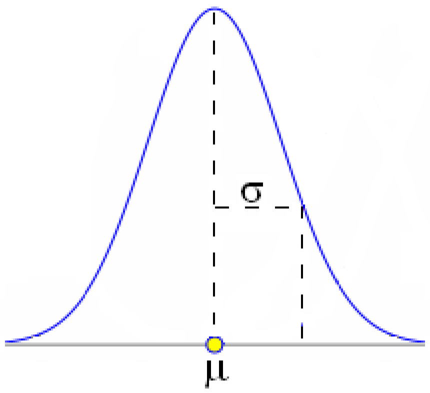
Tamanho do arquivo	Tamanho do buffer	
	20	40
Pequeno	2,05; 2,04; 2,21; 2,12	2,32; 2,31; 2,48; 2,42
Médio	2,24; 2,21; 2,23; 2,09	2,52; 2,62; 2,57; 2,61
Grande	2,08; 2,34; 2,33; 2,24	2,71; 2,73; 2,90; 2,72

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

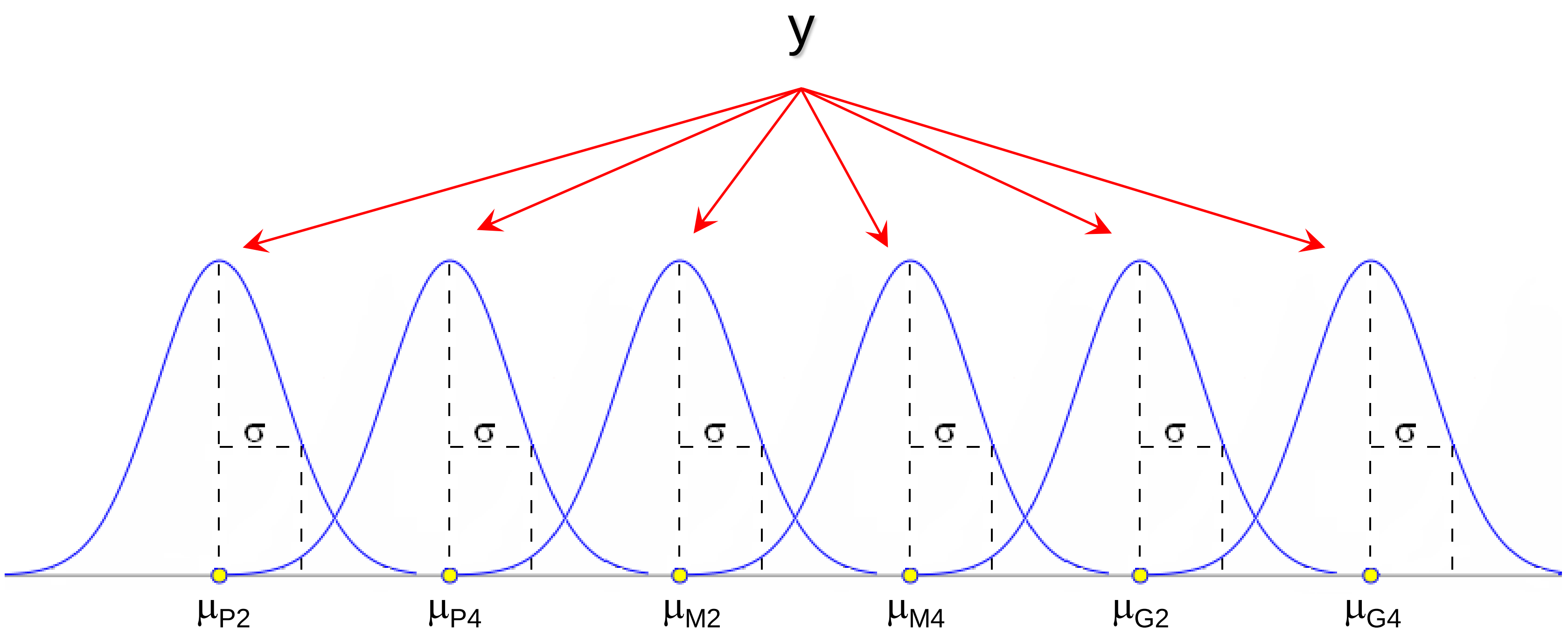
Populações estatísticas

y: variável resposta

y~N(μ, σ)



Tratamentos distintos geram populações de valores de y distintas?



Estrutura dos dados (generalização)

Os experimentos de classificação apresentam dois fatores de classificação (ou de agrupamento) das unidades de observação.

Fator A (i)	Fator B (j)			
	B_1	B_2	...	B_{n_b}
A_1	$y_{111}, y_{112}, \ldots, y_{11r}$	$y_{121}, y_{122}, \ldots, y_{12r}$	...	$y_{1n_b1}, y_{1n_b2}, \ldots, y_{1n_br}$
A_2	$y_{211}, y_{212}, \ldots, y_{21r}$	$y_{221}, y_{222}, \ldots, y_{22r}$	...	$y_{2n_b1}, y_{2n_b2}, \ldots, y_{2n_br}$
...	...	...	y_{ijk}	...
A_{n_a}	$y_{n_a11}, y_{n_a12}, \ldots, y_{n_a1r}$	$y_{n_a21}, y_{n_a22}, \ldots, y_{n_a2r}$	...	$y_{n_an_b1}, y_{n_an_b2}, \ldots, y_{n_an_br}$

Fator A, com n_a níveis: $\{A_1, A_2, \ldots, A_{na}\}$, i = nível do Fator A, sendo i=1,2, ..., n_a

Fator B, com n_b níveis: $\{B_1, B_2, \ldots, B_{nb}\}$, j = nível do Fator B, sendo j=1,2, ..., n_b

k = número de repetições da combinação de níveis, sendo k = 1, 2, ..., r

Modelos estatísticos

Modelo de médias:

$$y_{ijk} = \mu_{ij} + e_{ijk},$$

onde:

μ_{ij} é a média esperada da combinação de níveis ij (parâmetro)
e_{ijk} é o erro aleatório da repetição k da combinação de níveis ij

Modelo de efeitos:

$$y_{ijk} = \mu + \tau_i + \tau_j + \tau_{ij} + e_{ijk},$$

onde:

μ é a média sem efeito ou efeito constante (parâmetro)
τ_i é o efeito do nível i do fator A (parâmetro)
τ_j é o efeito do nível j do fator B (parâmetro)
τ_{ij} é o efeito da interação dos níveis i e j (parâmetro)
e_{ijk} é o erro aleatório da repetição k da combinação de níveis ij

Parâmetros dos modelos

$$y_{ijk} = \mu_{ij} + e_{ijk}$$

$$y_{ijk} = \mu + \tau_i + \tau_j + \tau_{ij} + e_{ijk}$$

$$\mu_{ij} = \mu + \tau_i + \tau_j + \tau_{ij}$$

Fator A (i)	Fator B (j)				Médias marginais
	B_1	B_2	...	B_{n_b}	
A_1	μ_{11}	μ_{12}	...	μ_{1n_a}	$\mu_{1.}$
A_2	μ_{21}	μ_{22}	...	μ_{2n_a}	$\mu_{2.}$
...	...	...	μ_{ij}	...	...
A_{n_a}	μ_{n_a1}	μ_{n_a2}	...	$\mu_{n_an_b}$	$\mu_{n_a.}$
Médias marginais	$\mu_{.1}$	$\mu_{.2}$	...	$\mu_{.n_b}$	μ

Efeito do nível i do fator A $\rightarrow$ $\tau_i = \mu_i - \mu$

Efeito do nível j do fator B $\rightarrow$ $\tau_j = \mu_j - \mu$

Efeito da interação dos níveis ij $\rightarrow$ $\tau_{ij} = \mu_{ij} - (\mu + \tau_i + \tau_j)$

$$\tau_{ij} = \mu_{ij} - \mu - \tau_i - \tau_j$$

$$\tau_{ij} = \mu_{ij} - \mu - (\mu_i - \mu) - (\mu_j - \mu)$$

$$\tau_{ij} = \mu_{ij} - \mu - \mu_i + \mu - \mu_j + \mu$$

$$\tau_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$$

	Parâmetro		Estimador
Modelo de médias $y_{ijk} = \mu_{ij} + e_{ijk}$	μ_{ij}	$\longrightarrow$	$\hat{\mu}_{ij} = \bar{y}_{ij} = \dfrac{\sum_{ij} y_{ijk}}{r}$
Modelo de efeitos $y_{ijk} = \mu + \tau_i + \tau_j + \tau_{ij} + e_{ijk}$	μ	$\longrightarrow$	$\hat{\mu} = \bar{y} = \dfrac{\sum_{ijk} y_{ijk}}{n}$
Efeito do nível i do fator A $\rightarrow$	$\tau_i = \mu_i - \mu$	$\longrightarrow$	$\hat{\tau}_i = \bar{y}_i - \bar{y}$
Efeito do nível j do fator B $\rightarrow$	$\tau_j = \mu_j - \mu$	$\longrightarrow$	$\hat{\tau}_j = \bar{y}_j - \bar{y}$
Efeito da interação dos níveis ij $\rightarrow$	$\tau_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$	$\longrightarrow$	$\hat{\tau}_{ij} = \bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}$
Erro da observação ijk $\rightarrow$	$e_{ijk} = y_{ijk} - \mu_{ij}$	$\longrightarrow$	$\hat{e}_{ijk} = y_{ijk} - \bar{y}_{ij}$ (Resíduo)

Tamanho do arquivo	Tamanho do buffer		Efeitos marginais
	2	4	
P	τ_{P2}	τ_{P4}	τ_{P}
M	τ_{M2}	τ_{M4}	τ_{M}
G	τ_{G2}	τ_{G4}	τ_{G}
Efeitos marginais	τ_{2}	τ_{4}	-

Hipóteses de interesse

Efeito de tratamento

$$\begin{cases} H_0^T : \mu_{P2} = \mu_{P4} = \mu_{M2} = \mu_{M4} = \mu_{G2} = \mu_{G4} = \mu \\ H_1^T : \mu_{ij} \neq \mu \quad (\text{pelo menos uma média difere de } \mu) \end{cases}$$

Efeito da interação dos fatores A e B

$$\begin{cases} H_0^{AB} : \tau_{P2} = \tau_{P4} = \tau_{M2} = \tau_{M4} = \tau_{G2} = \tau_{G4} = 0 \\ H_1^{AB} : \tau_{ij} \neq 0 \ (\text{pelo menos uma combinação tem efeito difere de } 0) \end{cases}$$

Efeito principal do fator A

$$\begin{cases} H_0^A : \mu_P = \mu_M = \mu_G = \mu \\ H_1^A : \mu_i \neq \mu \end{cases} \qquad \begin{cases} H_0^A : \tau_P = \tau_M = \tau_G = 0 \\ H_1^A : \tau_i \neq 0 \end{cases}$$

Efeito principal do fator B

$$\begin{cases} H_0^B : \mu_2 = \mu_4 = \mu \\ H_1^B : \mu_j \neq \mu \end{cases} \qquad \begin{cases} H_0^B : \tau_2 = \tau_4 = 0 \\ H_1^B : \tau_j \neq 0 \end{cases}$$

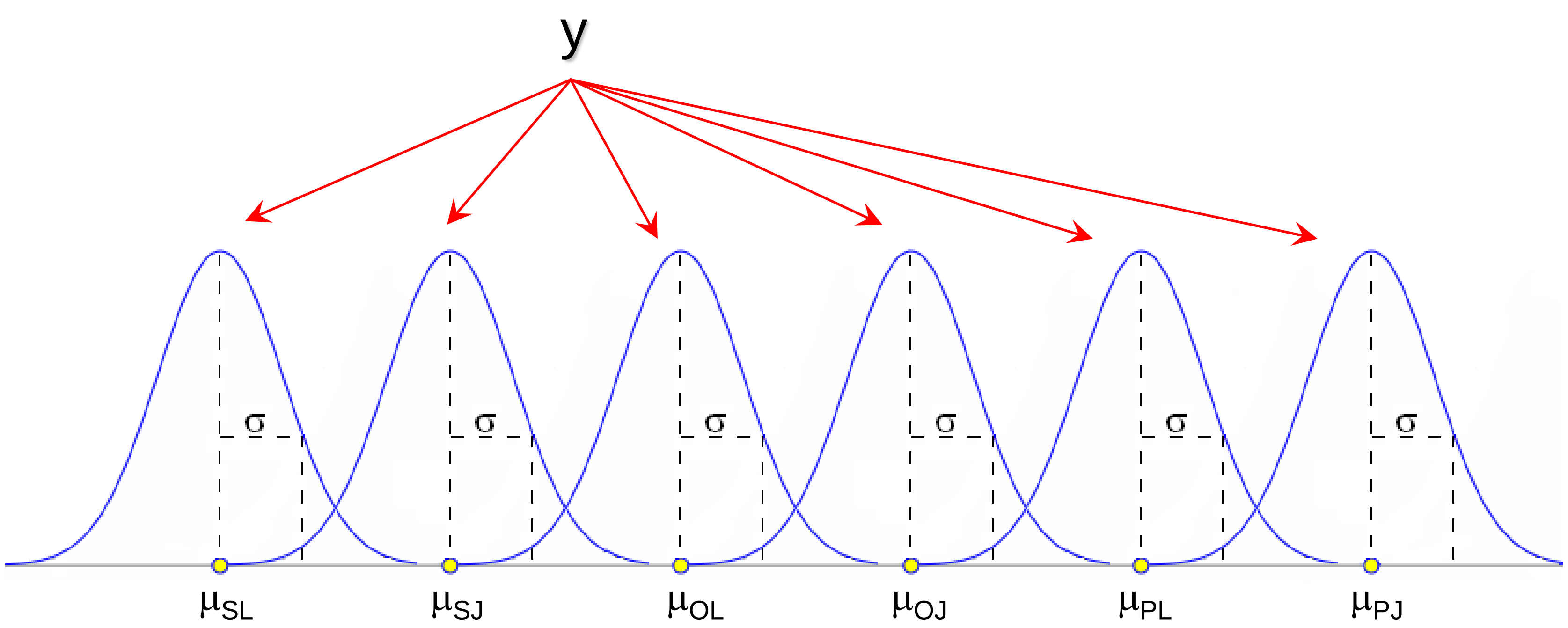
Populações estatísticas

y: variável resposta

Tratamentos distintos geram populações de valores de y distintas?

Sim

$$\begin{cases} H_0^T: \mu_{P2} = \mu_{P4} = \mu_{M2} = \mu_{M4} = \mu_{G2} = \mu_{G4} = \mu \\ H_1^T: \mu_{ij} \neq \mu \;\; (\text{pelo menos uma média difere de } \mu) \end{cases}$$



Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	μ_{P2}	μ_{P4}	μ_P
M	μ_{M2}	μ_{M4}	μ_M
G	μ_{G2}	μ_{G4}	μ_G
Médias marginais	μ_2	μ_4	μ

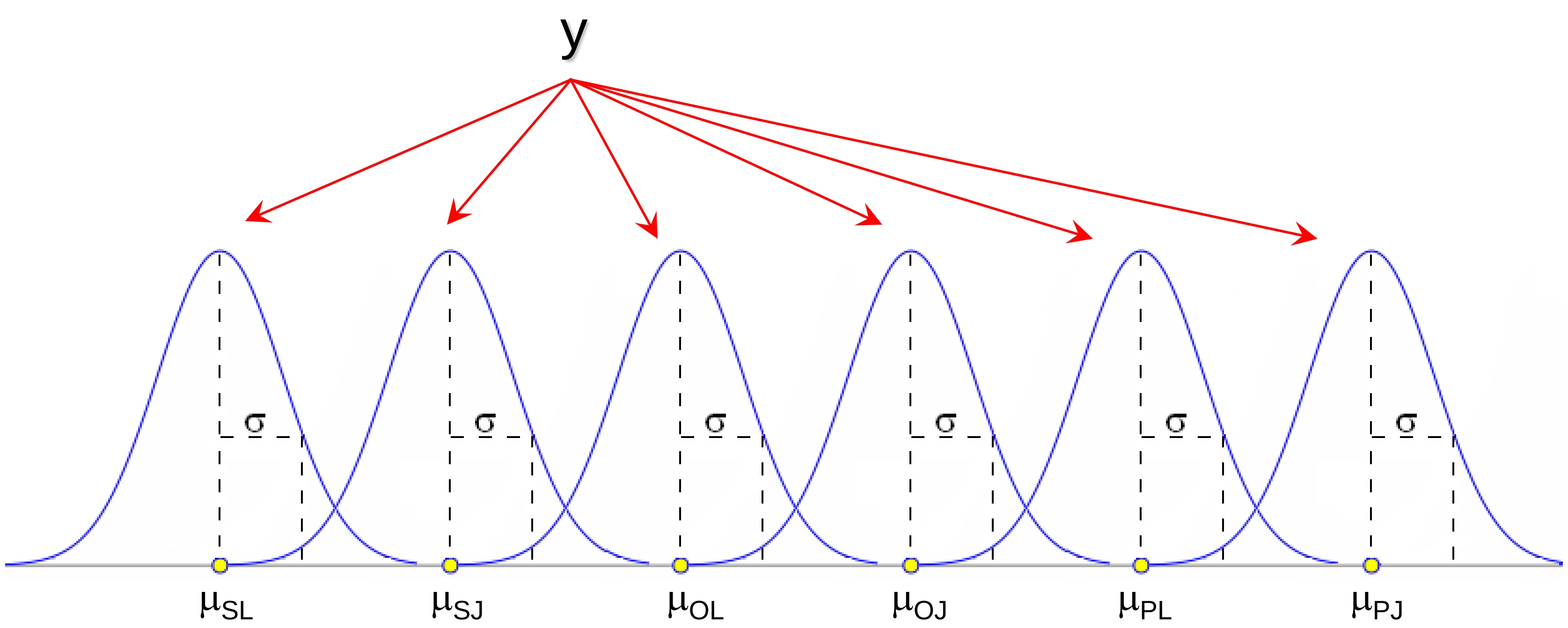
Populações estatísticas

y: variável resposta

Tratamentos distintos geram populações de valores de y distintas?

Sim ⇨ A.B: interação significativa?

$$\begin{cases} H_0^{AB}: \tau_{P2} = \tau_{P4} = \tau_{M2} = \tau_{M4} = \tau_{G2} = \tau_{G4} = 0 \\ H_1^{AB}: \tau_{ij} \neq 0 \end{cases}$$



Populações estatísticas

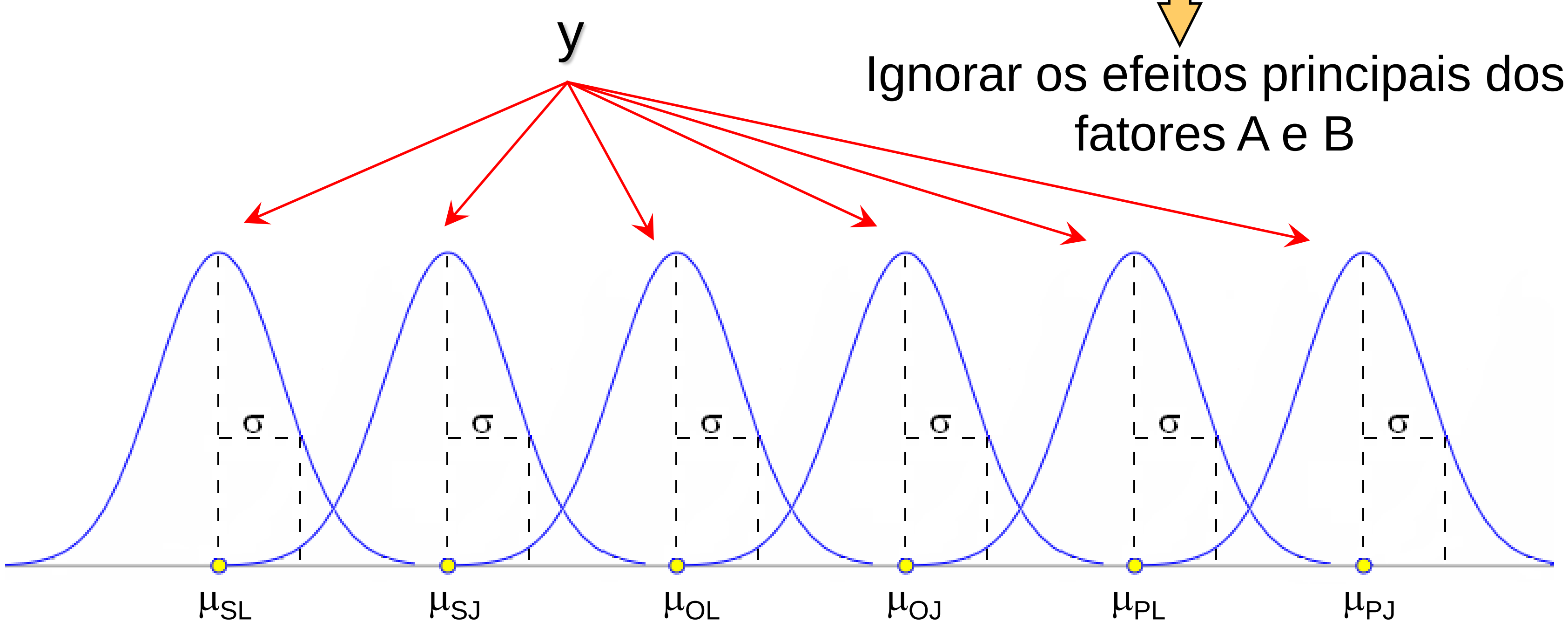
y: variável resposta

Tratamentos distintos geram populações de valores de y distintas?

Sim → A.B: interação significativa?

Sim

Ignorar os efeitos principais dos fatores A e B



Populações estatísticas

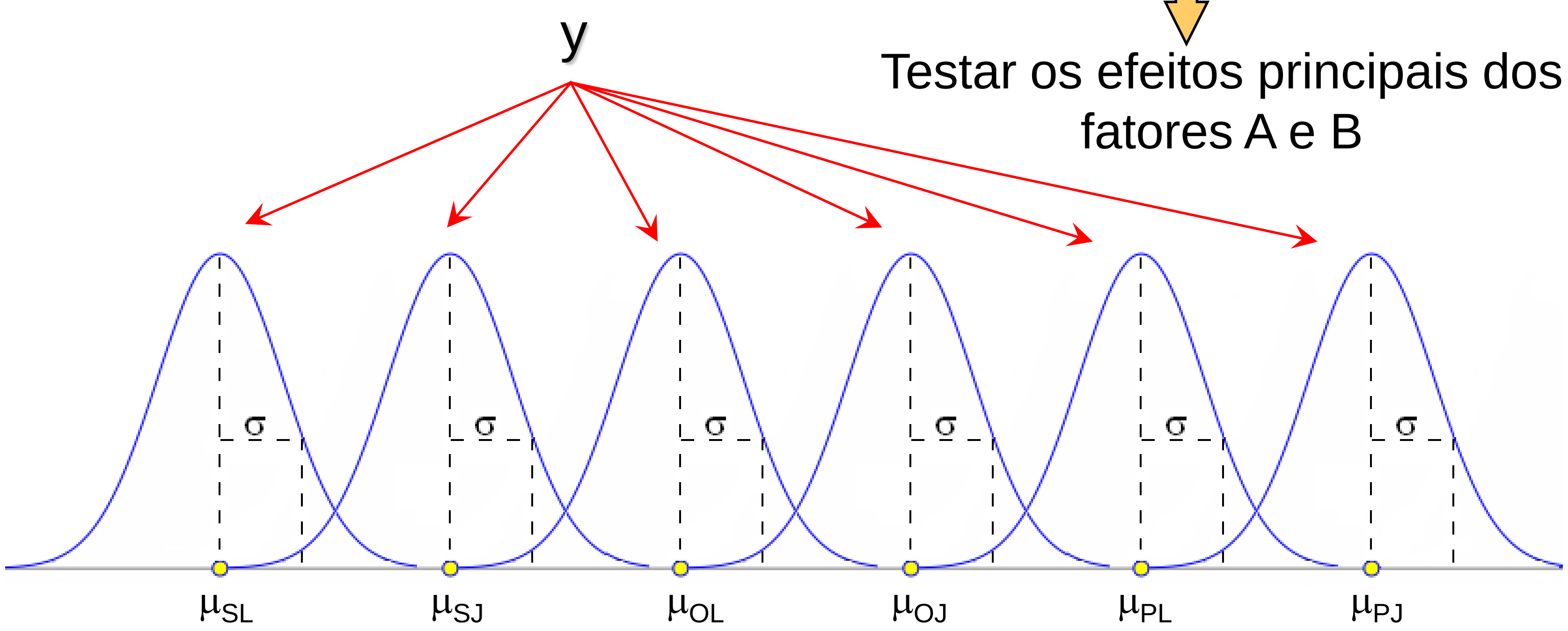
y: variável resposta

Tratamentos distintos geram populações de valores de y distintas?

Sim → A.B: interação significativa?

Não

Testar os efeitos principais dos fatores A e B



Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	μ_{P2}	μ_{P4}	μ_{P}
M	μ_{M2}	μ_{M4}	μ_{M}
G	μ_{G2}	μ_{G4}	μ_{G}
Médias marginais	μ_{2}	μ_{4}	μ

Populações estatísticas

y: variável resposta

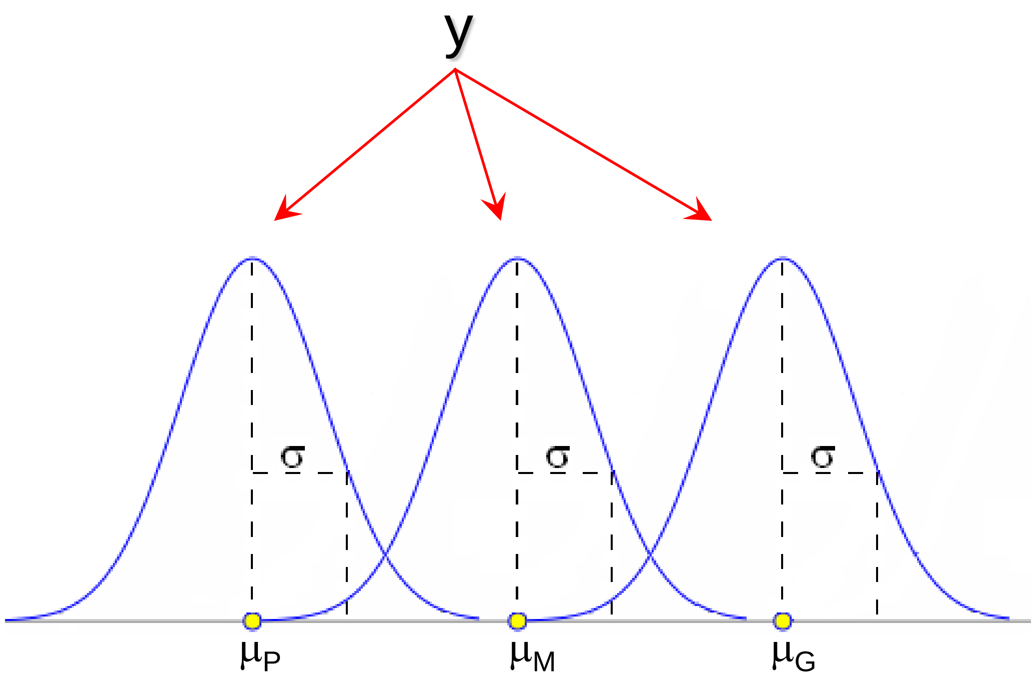
Tratamentos distintos geram populações de valores de y distintas?

$$\begin{cases} H_0^A: \mu_P = \mu_M = \mu_G = \mu \\ H_1^A: \mu_i \neq \mu \end{cases}$$

Sim

A.B : interação não significativa
A : efeito principal significativo

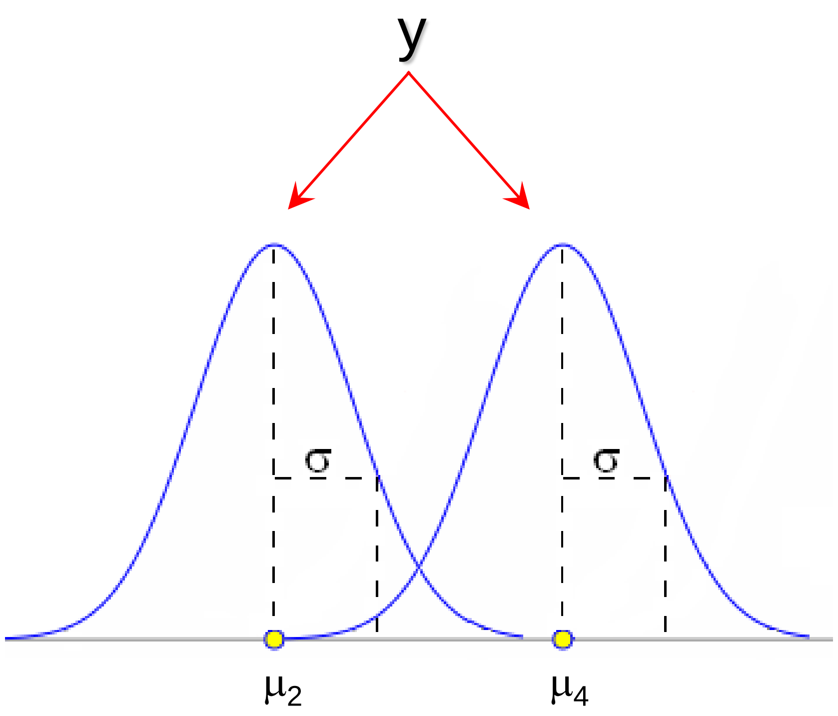
Considerar o fator A



Estrutura cruzada: populações estatísticas

y: variável resposta

$$\begin{cases} H_0^B : \mu_2 = \mu_4 = \mu \\ H_1^B : \mu_2 \neq \mu_4 \end{cases}$$



Tratamentos distintos geram populações de valores de y distintas?

Sim

A.B : interação não significativa
B : efeito principal significativo

Considerar o fator B

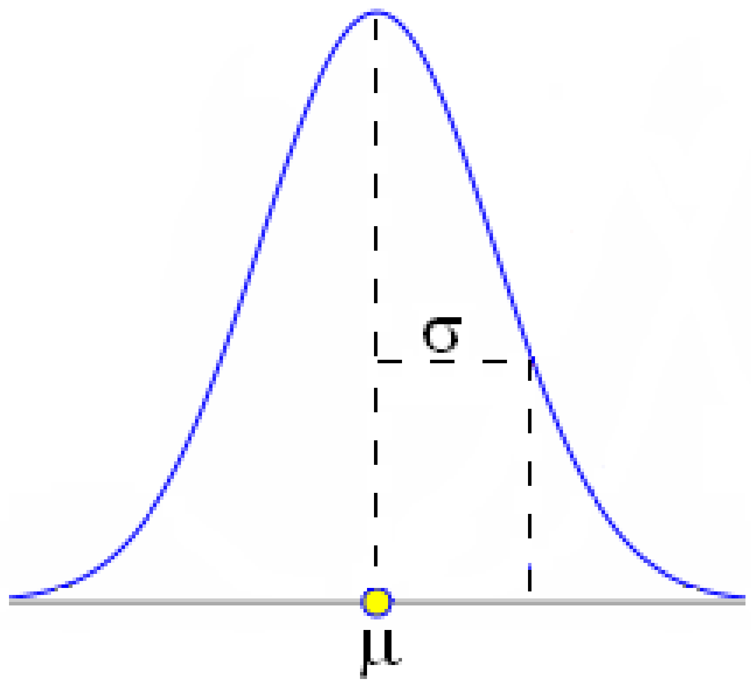
Populações estatísticas

y: variável resposta

Tratamentos distintos geram populações de valores de y distintas?

Não

$Y \sim N(\mu, \sigma)$



F.A : interação não significativa
F : efeito principal não significativo
A : efeito principal não significativo

Análise da variância

A análise da variância decompõe a variação total das observações, representada pelos desvios $(y_{ijk} - \bar{y})$, em duas partes:

→ a variação provocada pelo efeito do fator de tratamento, representada pelos desvios $(\bar{y}_{ij} - \bar{y})$;

→ a variação aleatória, representada pelos desvios$(y_{ijk} - \bar{y}_{ij})$.

Assim, a variação de cada observação pode ser representada pela seguinte expressão:

$$(y_{ijk} - \bar{y}) = (\bar{y}_{ij} - \bar{y}) + (y_{ijk} - \bar{y}_{ij})$$

E a variação total das observações pode ser representada por:

$$\sum_{ijk}(y_{ijk} - \bar{y})^2 = \sum_{ij} r(\bar{y}_{ij} - \bar{y})^2 + \sum_{ijk}(y_{ijk} - \bar{y}_{ij})^2$$

$$SQ_{Total} = SQ_{Tratamento} + SQ_{Resíduo}$$

$$y_{ijk} = \mu_{ij} + e_{ijk}$$

Fonte de variação	ν	SQ	S^2	F
Tratamento	ν_T= t - 1	$\sum_{ij} r(\bar{y}_{ij} - \bar{y})^2$	$s_T^2 = \frac{SQ_T}{\nu_T}$	$f_T = \frac{s_T^2}{s^2}$
Resíduo	ν=(r-1) t	$\sum_{ijk} (y_{ijk} - \bar{y}_{ij})^2$	$s^2 = \frac{SQ_{Res}}{\nu}$	-
Total	n-1	$\sum_{ijk} (y_{ijk} - \bar{y})^2$		

Hipóteses estatísticas: $\begin{cases} H_0^T : \mu_{11} = \mu_{12} = \ldots = \mu_{n_a n_b} = \mu \\ H_1^T : \mu_{ij} \neq \mu \end{cases}$

Estatística do teste: $f_T = \frac{s_T^2}{s^2} \sim F(\nu_T, \nu)$

Critério de decisão $\begin{cases} \text{Se } f_T > f_\alpha, \text{ rejeitamos } H_0 \\ \text{Se } f_T < f_\alpha, \text{ não temos motivos para rejeitar } H_0 \end{cases}$

Fonte de variação	ν	SQ	S^2	F
Tratamento	$\nu_T = t - 1$	$\sum_{ij} r(\bar{y}_{ij} - \bar{y})^2$	$s_T^2 = \frac{SQ_T}{\nu_T}$	$f_T = \frac{s_T^2}{s^2}$
Resíduo	$\nu=(r-1)\ t$	$\sum_{ijk} (y_{ijk} - \bar{y}_{ij})^2$	$s^2 = \frac{SQ_{Res}}{\nu}$	-
Total	n-1	$\sum_{ijk} (y_{ijk} - \bar{y})^2$		

$$SQ_{Total} = \sum_{ijk} (y_{ijk} - \bar{y})^2$$

$$SQ_{Trat} = \sum_{ij} r(\bar{y}_{ij} - \bar{y})^2$$

$$SQ_{Res} = \sum_{ijk} (y_{ijk} - \bar{y}_{ij})^2 = \sum_{ijk} \hat{e}_{ijk}^2 \quad \text{(por diferença)}$$

Exemplo: Programas de computador modernos requerem capacidade de acesso rápido aos dados. Uma pesquisa foi conduzida para estudar o efeito do *tamanho do arquivo* de dados (Pequeno; Médio e Grande) e do *tamanho do buffer* (área de memória intermediária que serve para acelerar o acesso a dados que estão sendo transferidos entre dispositivos - memória de disco e memória RAM - que operam com velocidades diferentes), (20 kb e 40 kb) sobre a variável *tempo de acesso* aos arquivos, medido através do *tempo de leitura em milissegundos*.

Tamanho do arquivo	Tamanho do buffer	
	20	40
Pequeno	2,05; 2,04; 2,21; 2,12	2,32; 2,31; 2,48; 2,42
Médio	2,24; 2,21; 2,23; 2,09	2,52; 2,62; 2,57; 2,61
Grande	2,08; 2,34; 2,33; 2,24	2,71; 2,73; 2,90; 2,72

$$SQ_{Total} = \sum_{ijk} \left(y_{ijk} - \bar{y}\right)^2$$

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

$$SQ_{Trat} = \sum_{ij} r\left(\bar{y}_{ij} - \bar{y}\right)^2$$

Fonte de variação	GL	SQ	S^2	F	$f\alpha$
Tratamento	5	1,2662	0,2532	35,83	2,77
Resíduo	18	0,1272	0,00707	-	-
Total	23	1,3935	-	-	-

$$\begin{cases} H_0^T : \mu_{P2} = \mu_{P4} = \mu_{M2} = \mu_{M4} = \mu_{G2} = \mu_{G4} = \mu \\ H_1^T : \mu_{ij} \neq \mu \end{cases}$$ $\leftarrow$ Rejeita-se H_0

$$SQ_{Total} = \sum_{ijk} \left(y_{ijk} - \bar{y}\right)^2 = (2{,}05 - 2{,}379)^2 + (2{,}04 - 2{,}379)^2 + \ldots + (2{,}72 - 2{,}379)^2 = 1{,}3935$$

$$SQ_{Trat} = \sum_{ij} r\left(\bar{y}_{ij} - \bar{y}\right)^2 = 4 \times (2{,}105 - 2{,}379)^2 + 4 \times (2{,}383 - 2{,}379)^2 + \ldots + 4 \times (2{,}765 - 2{,}379)^2 = 1{,}2662$$

$$SQ_{Res} = SQ_{Total} - SQ_{Trat} = 1{,}3935 - 1{,}2662 = 0{,}1272$$

Conclusão: Concluímos ao nível de 5% de significância, que existe efeito das combinações de níveis dos fatores tamanho do buffer e tamanho do arquivo sobre o tempo acesso ao arquivo.

Hipóteses de interesse

Efeito de tratamento

$$\begin{cases} H_0^T: \mu_{P2} = \mu_{P4} = \mu_{M2} = \mu_{M4} = \mu_{G2} = \mu_{G4} = \mu \\ H_1^T: \mu_{ij} \neq \mu \ \ \text{(pelo menos uma média difere de } \mu\text{)} \end{cases}$$

Efeito da interação dos fatores A e B

$$\begin{cases} H_0^{AB}: \tau_{P2} = \tau_{P4} = \tau_{M2} = \tau_{M4} = \tau_{G2} = \tau_{G4} = 0 \\ H_1^{AB}: \tau_{ij} \neq 0 \ \ \text{(pelo menos uma combinação tem efeito difere de 0)} \end{cases}$$

Efeito principal do fator A

$$\begin{cases} H_0^A: \mu_P = \mu_M = \mu_G = \mu \\ H_1^A: \mu_i \neq \mu \end{cases}$$

$$\begin{cases} H_0^A: \tau_P = \tau_M = \tau_G = 0 \\ H_1^A: \tau_i \neq 0 \end{cases}$$

Efeito principal do fator B

$$\begin{cases} H_0^B: \mu_2 = \mu_4 = \mu \\ H_1^B: \mu_j \neq \mu \end{cases}$$

$$\begin{cases} H_0^B: \tau_2 = \tau_4 = 0 \\ H_1^B: \tau_j \neq 0 \end{cases}$$

$$y_{ijk} = \mu + \tau_i + \tau_j + \tau_{ij} + e_{ijk}$$

Fonte de variação	ν	SQ	S^2	F
Tratamento	$\nu_T = t - 1$	$\sum_{ij} r(\bar{y}_{ij} - \bar{y})^2$	$s_T^2 = \frac{SQ_T}{\nu_T}$	$f_T = \frac{s_T^2}{s^2}$
Fator A	$\nu_A = n_a - 1$	$\sum_{i} n_b r(\bar{y}_i - \bar{y})^2$	$s_A^2 = \frac{SQ_A}{\nu_A}$	$f_A = \frac{s_A^2}{s^2}$
Fator B	$\nu_B = n_b - 1$	$\sum_{j} n_a r(\bar{y}_j - \bar{y})^2$	$s_B^2 = \frac{SQ_B}{\nu_B}$	$f_B = \frac{s_B^2}{s^2}$
A.B	$\nu_{AB} = \nu_a \times \nu_b$	$\sum_{ij} r(\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$	$s_{AB}^2 = \frac{SQ_{AB}}{\nu_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s^2}$
Resíduo	$\nu = (r-1)\,t$	$\sum_{ijk} (y_{ijk} - \bar{y}_{ij})^2$	$s^2 = \frac{SQ_{Res}}{\nu}$	-
Total	n-1	$\sum_{ijk} (y_{ijk} - \bar{y})^2$	-	-

Somas de quadrados

$$SQ_A = \sum_i n_b r \left(y_i - \bar{y}\right)^2$$

$$SQ_B = \sum_j n_a r \left(y_j - \bar{y}\right)^2$$

$$SQ_{AB} = SQ_T - SQ_A - SQ_B \quad \text{(por diferença)}$$

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

Hipóteses estatísticas

Estatística do teste

Interação entre os fatores A e B

$$\begin{cases} H_0^{AB}: \tau_{11} = \tau_{12} = \ldots = \tau_{n_a n_b} = 0 \\ H_1^{AB}: \tau_{ij} \neq 0 \end{cases}$$

$$f_{AB} = \frac{s_{AB}^2}{s^2}$$

Se a interação não é significativa:

- Testar o efeito principal do fator A

$$\begin{cases} H_0^{A}: \mu_1 = \mu_2 = \ldots = \mu_{n_a} = \mu \\ H_1^{A}: \mu_i \neq \mu \end{cases}$$

$$f_{A} = \frac{s_{A}^2}{s^2}$$

- Testar o efeito principal do fator B

$$\begin{cases} H_0^{B}: \mu_1 = \mu_2 = \ldots = \mu_{n_b} = \mu \\ H_1^{B}: \mu_j \neq \mu \end{cases}$$

$$f_{B} = \frac{s_{B}^2}{s^2}$$

- Testar o efeito principal do fator A

$$\begin{cases} H_0^A : \mu_1 = \mu_2 = \ldots = \mu_{n_a} = \mu \\ H_1^A : \mu_i \neq \mu \end{cases}$$

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

- Testar o efeito principal do fator B

$$\begin{cases} H_0^B : \mu_1 = \mu_2 = \ldots = \mu_{n_b} = \mu \\ H_1^B : \mu_j \neq \mu \end{cases}$$

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	2	4	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

Hipóteses estatísticas

Interação entre os fatores A e B

$$\begin{cases} H_0^{AB}: \tau_{11} = \tau_{12} = \ldots = \tau_{n_a n_b} = 0 \\ H_1^{AB}: \tau_{ij} \neq 0 \end{cases}$$

Estatística do teste

$$f_{AB} = \frac{s_{AB}^2}{s^2}$$

Se a interação é significativa:

- Testar o efeito simples de A dentro de cada nível de B

- Testar o efeito simples de B dentro de cada nível de A

Hipóteses estatísticas

Interação entre os fatores A e B

$$\begin{cases} H_0^{AB} : \tau_{11} = \tau_{12} = \ldots = \tau_{n_a n_b} = 0 \\ H_1^{AB} : \tau_{ij} \neq 0 \end{cases}$$

Estatística do teste

$$f_{AB} = \frac{s_{AB}^2}{s^2}$$

Se a interação é significativa:

- Testar o efeito simples de A dentro de cada nível de B

$$\begin{cases} H_0^{A|B_j} : \mu_{1|B_j} = \mu_{2|B_j} = \ldots = \mu_{n_a|B_j} = 0 \\ H_1^{A|B_j} : \mu_{i|B_j} \neq 0 \end{cases}$$

- Testar o efeito simples de B dentro de cada nível de A

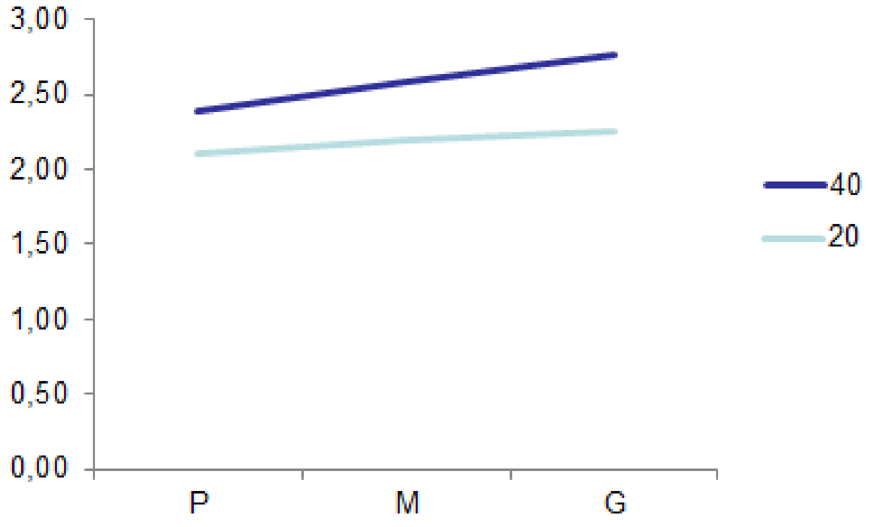
$$\begin{cases} H_0^{B|A_i} : \mu_{B|A_i} = \mu_{2|A_i} = \ldots = \mu_{n_b|A_i} = 0 \\ H_1^{B|A_i} : \mu_{j|A_i} \neq 0 \end{cases}$$

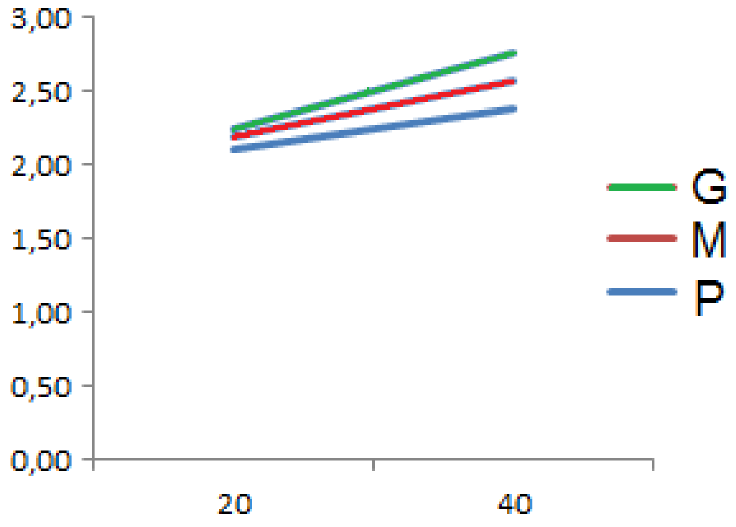
Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

No exemplo: Tabela de média observadas

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379





Gráficos de médias: indicam a presença ou não de interação

Fonte de variação	GL	SQ	S^2	F	$f\alpha$
Tratamento	5	1,2662	0,2532	35,83	
Arquivo (A)	2	0,2763	0,13815	19,55	
Buffer (B)	1	0,9322	0,9322	131,89	
Arquivo x Buffer (AB)	2	0,0577	0,02887	4,08	3,55
Resíduo	18	0,1272	0,00707	-	-
Total	23	1,3935	-	-	-

$$SQ_A = \sum_i n_b r\left(\bar{y}_i - \bar{y}\right)^2 = 2\times 4\times(2{,}244-2{,}379)^2 + 2\times 4\times(2{,}386-2{,}379)^2 + 2\times 4\times(2{,}506-2{,}379)^2 = 0{,}2763$$

$$SQ_B = \sum_j n_a r\left(y_j - \bar{y}\right)^2 = 3\times 4\times(2{,}182-2{,}379)^2 + 3\times 4\times(2{,}576-2{,}379)^2 = 0{,}9322$$

$$SQ_{AB} = SQ_T - SQ_A - SQ_B = 1{,}2662 - 0{,}9322 - 0{,}2763 = 0{,}0577$$

Fonte de variação	GL	SQ	S^2	F	fα
Tratamento	5	1,2662	0,2532	35,83	
Arquivo (A)	2	0,2763	0,13815	19,55	
Buffer (B)	1	0,9322	0,9322	131,89	
Arquivo x Buffer (AB)	2	0,0577	0,02887	4,08	3,55
Resíduo	18	0,1272	0,00707	-	-
Total	23	1,3935	-	-	-

Efeito significativo da interação dos fatores A e B

$$\begin{cases} H_0^{AB} : \tau_{P2} = \tau_{P4} = \tau_{M2} = \tau_{M4} = \tau_{G2} = \tau_{G4} = 0 \\ H_1^{AB} : \tau_{ij} \neq 0 \end{cases}$$

$\leftarrow$ Rejeita-se H_0

Conclusão: Concluímos ao nível de 5% de significância, que existe interação entre tamanho do arquivo e tamanho do buffer. Isso significa que a mudança nos níveis de tamanho do arquivo irá modificar a forma como os níveis de tamanho do buffer afetam a variável resposta (e vice-versa).

Hipóteses estatísticas

Interação entre os fatores A e B

$$\begin{cases} H_0^{AB}: \tau_{11} = \tau_{12} = \ldots = \tau_{n_a n_b} = 0 \\ H_1^{AB}: \tau_{ij} \neq 0 \end{cases}$$

Estatística do teste

$$f_{AB} = \frac{s_{AB}^2}{s^2}$$

Se a interação é significativa:

- Testar o efeito simples de A dentro de cada nível de B

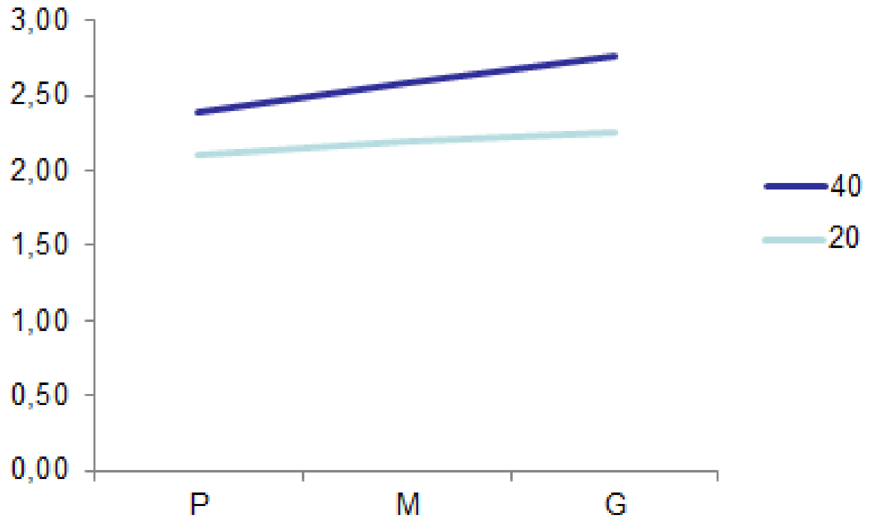
- Testar o efeito simples de B dentro de cada nível de A

Se a interação é significativa:

- Testar o efeito simples de A (tamanho do Arquivo) dentro de cada nível de B (Tamanho do Buffer)

$$\begin{cases} H_0^{A|2} : \mu_{P|2} = \mu_{M|2} = \mu_{G|2} = \mu_2 \\ H_1^{A|2} : \mu_{i|2} \neq \mu_2 \end{cases}$$

$$\begin{cases} H_0^{A|4} : \mu_{P|4} = \mu_{M|4} = \mu_{G|4} = \mu_4 \\ H_1^{A|4} : \mu_{i|4} \neq \mu_4 \end{cases}$$



Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	**2,383**	2,244
M	2,193	**2,580**	2,386
G	2,248	**2,765**	2,506
Médias marginais	2,182	**2,576**	2,379

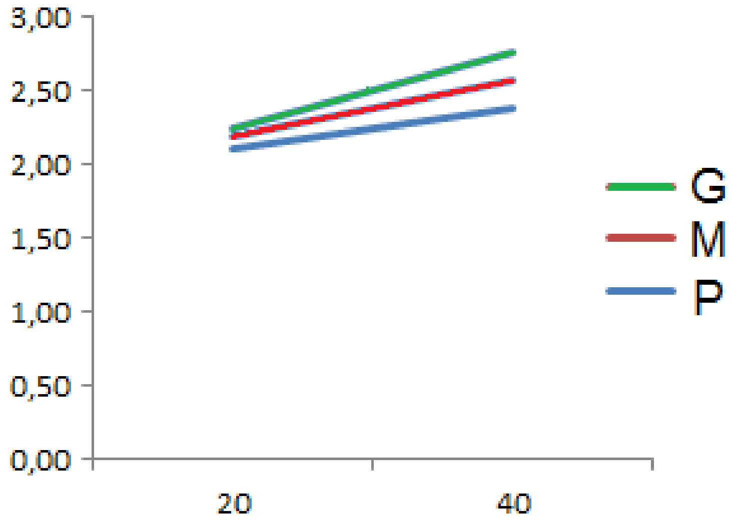
Se a interação é significativa:

- Testar o efeito simples de B (tamanho do Buffer) dentro de cada nível de A (Tamanho do Arquivo)

$$\begin{cases} H_0^{B|P}: \mu_{2|P} = \mu_{4|P} = \mu_P \\ H_1^{B|P}: \mu_{j|P} \neq \mu_P \end{cases}$$

$$\begin{cases} H_0^{B|M}: \mu_{2|M} = \mu_{4|M} = \mu_M \\ H_1^{B|M}: \mu_{j|M} \neq \mu_M \end{cases}$$

$$\begin{cases} H_0^{B|G}: \mu_{2|G} = \mu_{4|G} = \mu_G \\ H_1^{B|G}: \mu_{j|G} \neq \mu_G \end{cases}$$



Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	**2,248**	**2,765**	**2,506**
Médias marginais	2,182	2,576	2,379

Fonte de variação	GL	SQ	S^2	F	$f\alpha$
Tratamento	5	1,2662	0,2532	35,83	
Arquivo (A)	2	0,2763	0,13815	19,55	
Buffer (B)	1	0,9322	0,9322	131,89	
Arquivo x Buffer (AB)	2	0,0577	0,02887	4,08	3,55
Resíduo	18	0,1272	0,00707	-	-
Total	23	1,3935	-	-	-

Efeito significativo da interação dos fatores A e B

$$\begin{cases} H_0^{AB}: \tau_{P2} = \tau_{P4} = \tau_{M2} = \tau_{M4} = \tau_{G2} = \tau_{G4} = 0 \\ H_1^{AB}: \tau_{ij} \neq 0 \end{cases}$$

$\leftarrow$ Rejeita-se H_0

Conclusão: Concluímos ao nível de 5% de significância, que existe interação entre tamanho do arquivo e tamanho do buffer. Isso significa que a mudança nos níveis de tamanho do arquivo irá modificar a forma como os níveis de tamanho do buffer afetam a variável resposta (e vice-versa).

No exemplo: Tabela de médias

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	2	4	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

$$\hat{\tau}_i = \bar{y}_i - \bar{y}$$

$$\hat{\tau}_j = \bar{y}_j - \bar{y}$$

$$\hat{\tau}_{ij} = \bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}$$

Tabela de efeitos

Tamanho do arquivo	Tamanho do buffer		Efeitos marginais
	20	40	
P	0,058	-0,058	-0,135
M	0,004	-0,004	0,007
G	-0,061	0,061	0,127
Efeitos marginais	-0,197	0,197	0

Tabela de médias

Tamanho do arquivo	Tamanho do buffer		Médias marginais
	20	40	
P	2,105	2,383	2,244
M	2,193	2,580	2,386
G	2,248	2,765	2,506
Médias marginais	2,182	2,576	2,379

$$\begin{cases} H_0^A : \mu_P = \mu_M = \mu_G = \mu \\ H_1^A : \mu_i \neq \mu \end{cases}$$

$$\begin{cases} H_0^B : \mu_2 = \mu_4 = \mu \\ H_1^B : \mu_j \neq \mu \end{cases}$$

Tabela de efeitos

Tamanho do arquivo	Tamanho do buffer		Efeitos marginais
	20	40	
P	0,058	-0,058	-0,135
M	0,004	-0,004	0,007
G	-0,061	0,061	0,127
Efeitos marginais	-0,197	0,197	0

$$\begin{cases} H_0^A : \tau_P = \tau_M = \tau_G = 0 \\ H_1^A : \tau_i \neq 0 \end{cases}$$

$$\begin{cases} H_0^B : \tau_2 = \tau_4 = 0 \\ H_1^B : \tau_j \neq 0 \end{cases}$$

$$\begin{cases} H_0^{AB} : \tau_{P2} = \tau_{P4} = \tau_{M2} = \tau_{M4} = \tau_{G2} = \tau_{G4} = 0 \\ H_1^{AB} : \tau_{ij} \neq 0 \text{ (pelo menos uma combinação tem efeito difere de 0)} \end{cases}$$

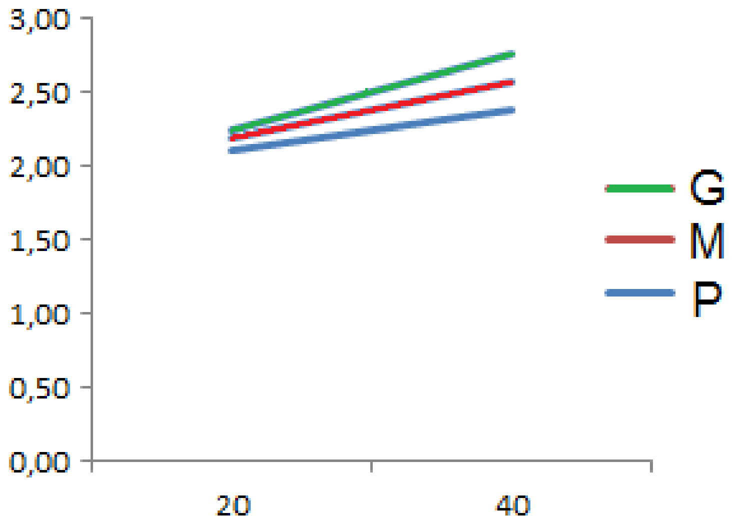
No exemplo: Tabela de efeitos

Tamanho do arquivo	Tamanho do buffer		Efeitos marginais
	20	40	
P	0,058	-0,058	-0,135
M	0,004	-0,004	0,007
G	-0,061	0,061	0,127
Efeitos marginais	-0,197	0,197	0

$$\hat{\tau}_i = \bar{y}_i - \bar{y}$$

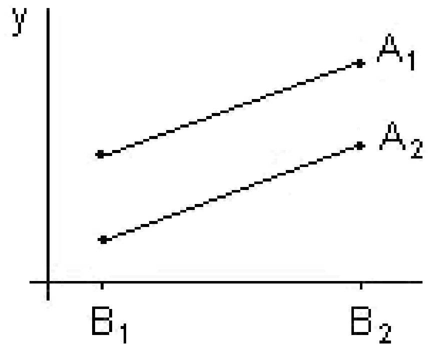
$$\hat{\tau}_j = \bar{y}_j - \bar{y}$$

$$\hat{\tau}_{ij} = \bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}$$

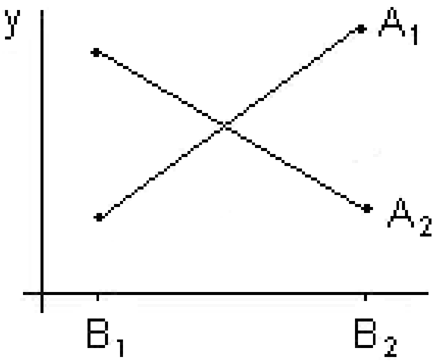


Quando o efeito simples de um fator não varia entre os níveis do outro fator, a interação A x B é nula ou inexistente.

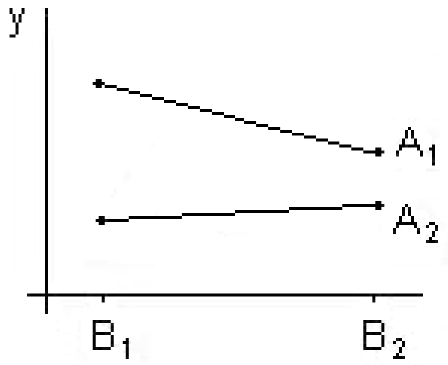
Abaixo são apresentadas quatro situações distintas referentes à magnitude da interação A x B:

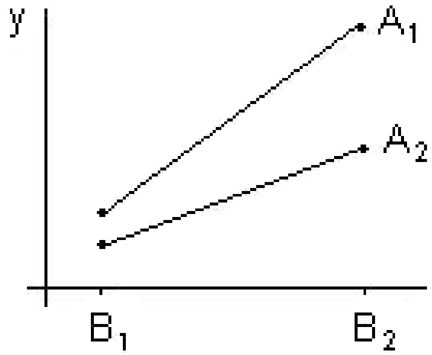


Ausência de interação
($B|A_1 = B|A_2$)



Interação elevada
($B|A_1 = B|A_2$, com sinais contrários)





Interação média ($B|A_1 \neq B|A_2$)

Bibliografia consultada

SILVA, J.G.C. da **Estatística experimental: análise estatística de experimentos.** Pelotas, RS: Instituto de Física e Matemática, Universidade Federal de Pelotas, 2000. 318p.

Sistema Galileu de Educação Estatística. Disponível em: http://www.galileu.esalq.usp.br

VIEIRA, S. **Estatística Experimental**. 2 ed. São Paulo: Atlas, 1999. 185p.