

## **Aula 06 –(Aula Síncrona)** **(21/09)**

Davi R. de Moura Costa

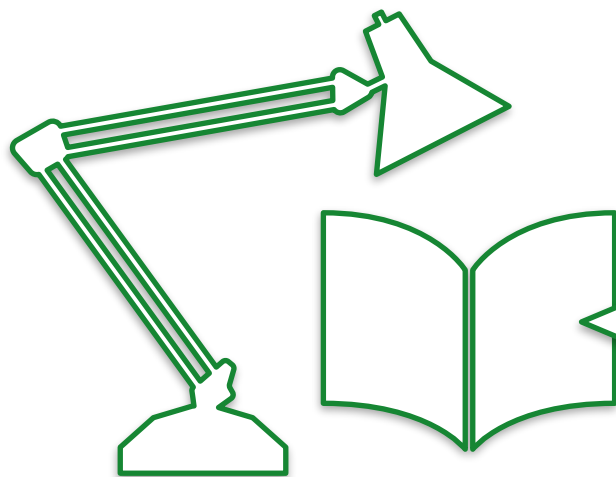


USP



fea-RP



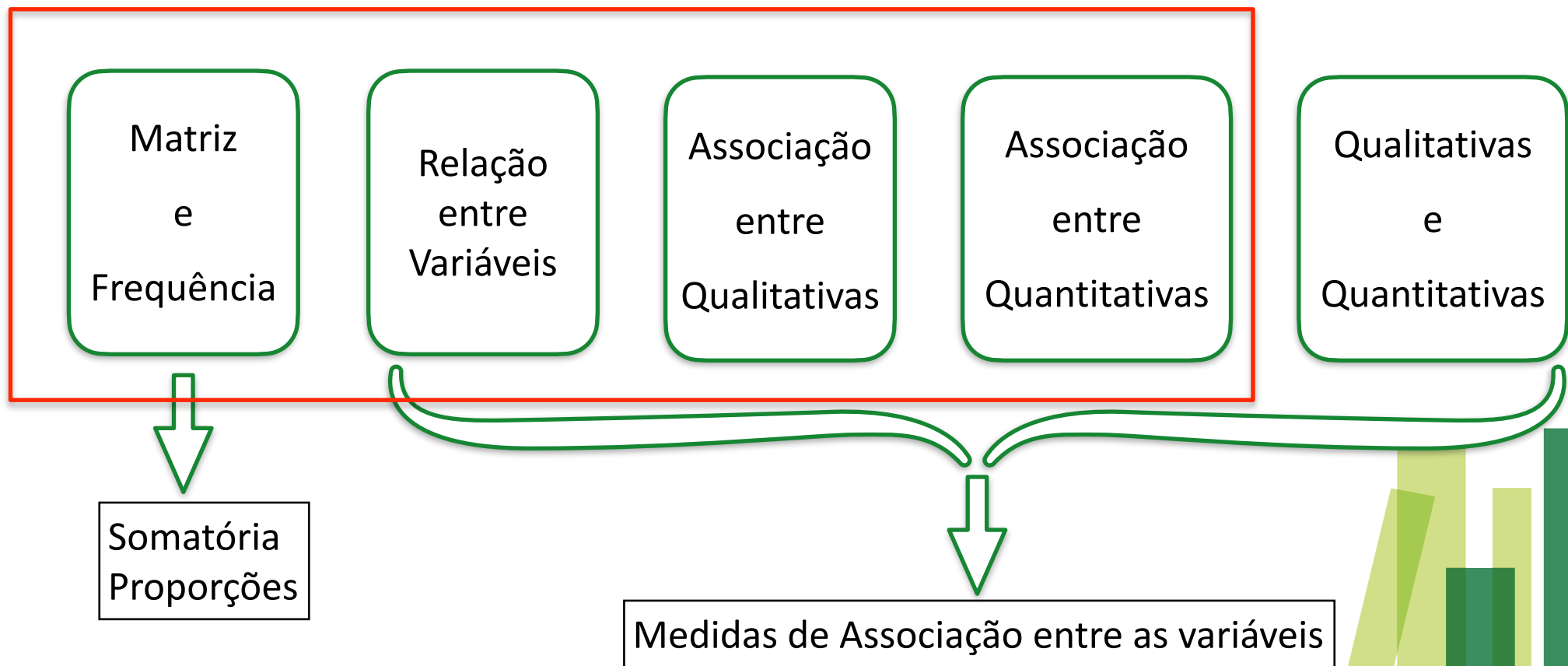


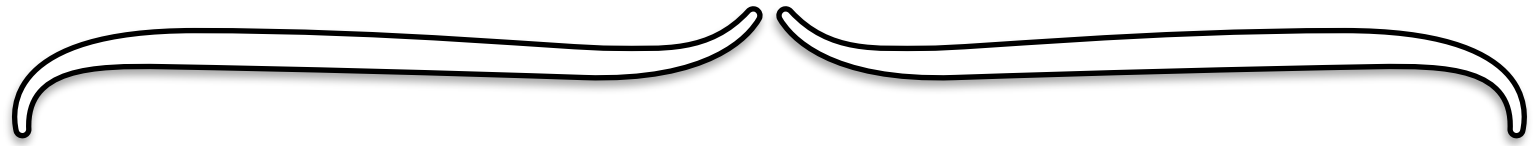
## Análise Bidimensional

- Revisão de Matriz
- Revisão de Dist de Frequência
- Relação entre as variáveis
- Associação entre Var. Qualitativas
- Associação entre Var. Quantitativas
- Associação entre Var. Qual/Quant

Leitura obrigatória

- Bussab & Morettin - Cap. 4 (cont)





Freqüência Absoluta

Freqüência Relativa (%)

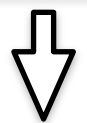
Freqüência Acumulada

$$n_i$$

Observações da variável

$$x$$

$$f_i = \frac{n_i}{N}$$



$$N = \begin{cases} n_{..} = \sum_{i=1}^L \sum_{j=1}^C n_{ij}, \\ n_{i.} = \sum_{j=1}^C n_{ij}, \\ n_{.j} = \sum_{i=1}^L n_{ij}, \end{cases}$$

número total de elementos  
Total de elementos por linha  
Total de elementos por coluna

$$N(x) \leq x$$

$$F_e(x) = F_n(x) = \frac{N(x)}{n}$$

$N(x)$  número de observações  
 $F_e(x)$  = Função de distribuição empírica



$X \backslash Y$	$B_1$	$B_2$	...	$B_j$	...	$B_s$	Total
$A_1$	$n_{11}$	$n_{12}$	...	$n_{1j}$	...	$n_{1s}$	$n_{1.}$
$A_2$	$n_{21}$	$n_{22}$	...	$n_{2j}$	...	$n_{2s}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_i$	$n_{i1}$	$n_{i2}$	...	$n_{ij}$	...	$n_{is}$	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_r$	$n_{r1}$	$n_{r2}$	...	$n_{rj}$	...	$n_{rs}$	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.s}$	$n_{..}$

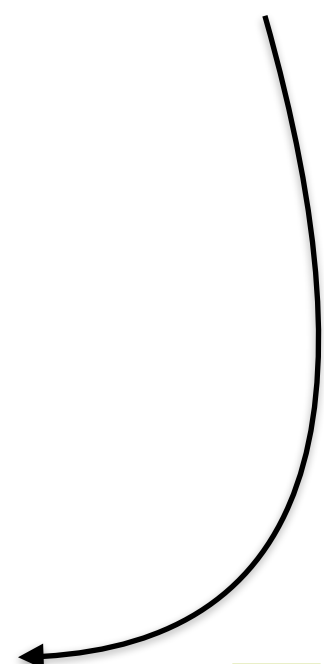
$$f_i = \frac{n_i}{N}$$

$n_{ij}$  = número de elementos pertencentes à  $i$ -ésima categoria de  $X$  e  $j$ -ésima categoria de  $Y$ ;

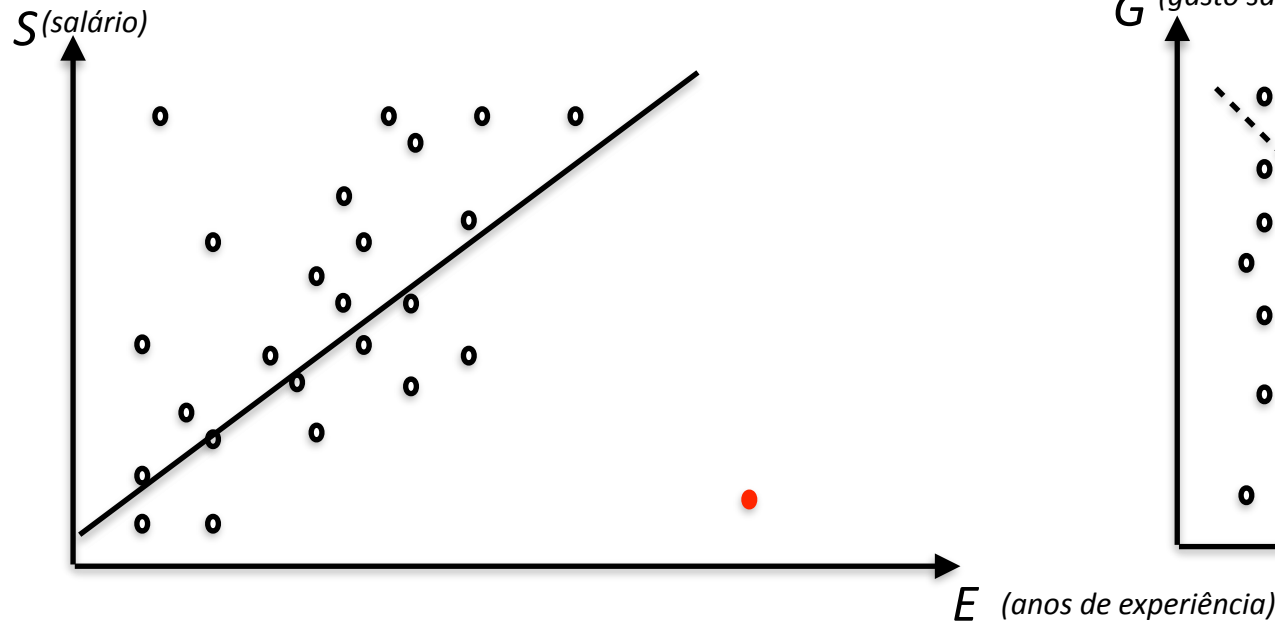
$n_{i.} = \sum_{j=1}^s n_{ij}$  = número de elementos da  $i$ -ésima categoria de  $X$ ;

$n_{.j} = \sum_{i=1}^r n_{ij}$  = número de elementos da  $j$ -ésima categoria de  $Y$ ;

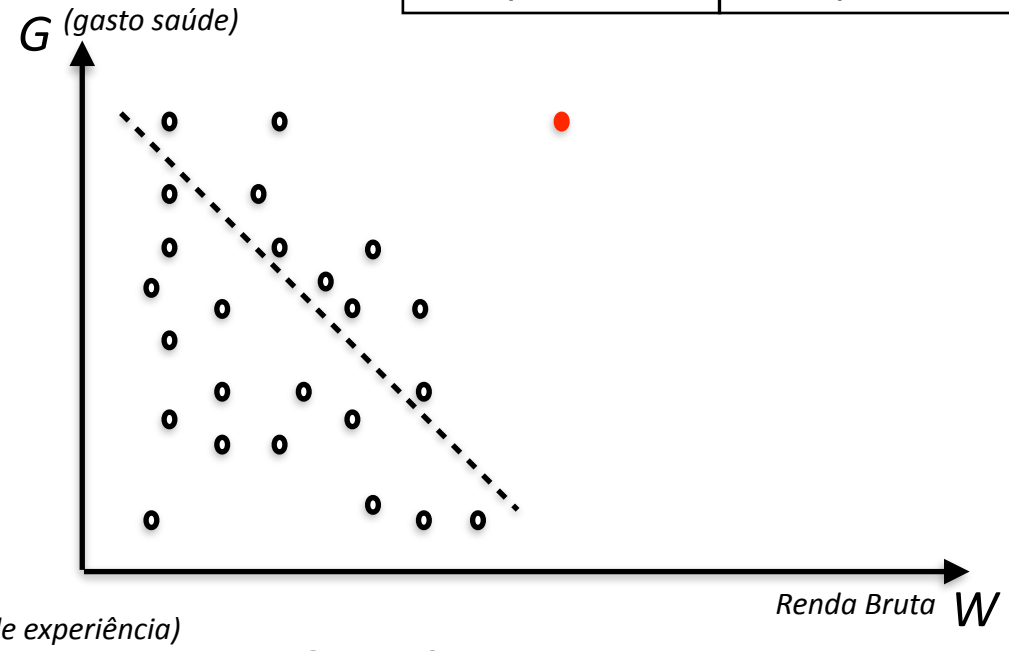
$n_{..} = n = \sum_{i=1}^r \sum_{j=1}^s n_{ij}$  = número total de elementos.



Variáveis Quantitativas	
Dependente	Independente



Correlação



Covariância

$$\text{Corr}(X, Y) = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{dp(X) dp(Y)} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}}$$

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$-1 \leq \text{corr}(X, Y) \leq 1$$

$$\text{Corr}(X, Y) = \frac{\text{cov}(X, Y)}{dp(X) dp(Y)} =$$



Tipos de Associação entre Variáveis

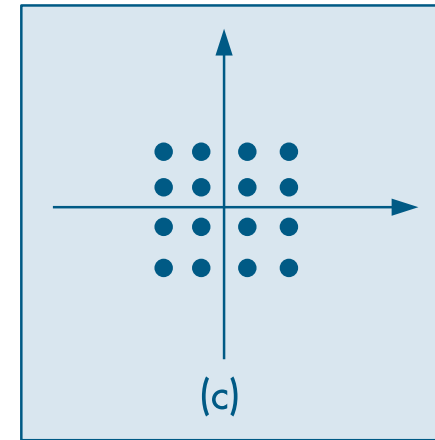
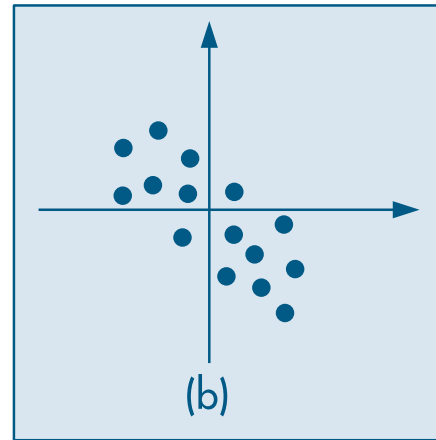
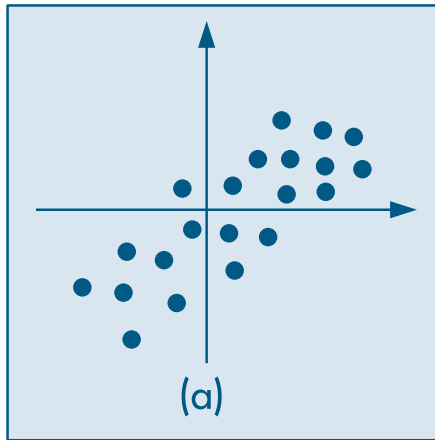


Tabela de Contingência

Câncer		Fumante	
Sim	Não	Sim	Não

Independentes  
x  
Pareado

Variáveis Qualitativas	
Dependente	Independente
Nominal	Nominal

	Câncer		
Fumante	Sim	Não	Total
Sim	S,S	S,N	$n_{1j}$
Não	N,S	N,N	$n_{2j}$
Total	$n_{i1}$	$n_{i2}$	$n_{..}$

- $n_{1j}$  → núm. indivíduos que fumam
- $n_{2j}$  → núm. indivíduos que não fumam
- $n_{i1}$  → núm. indivíduos com câncer
- $n_{i2}$  → núm. indivíduos sem câncer
- $n_{..}$  → núm. total de indivíduos

Sabemos que

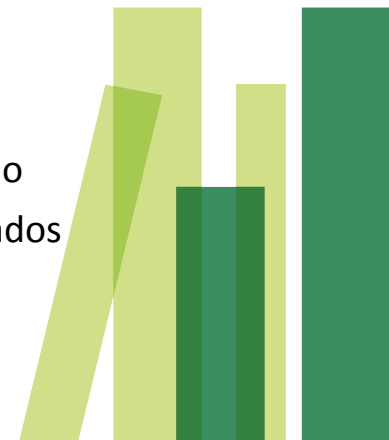
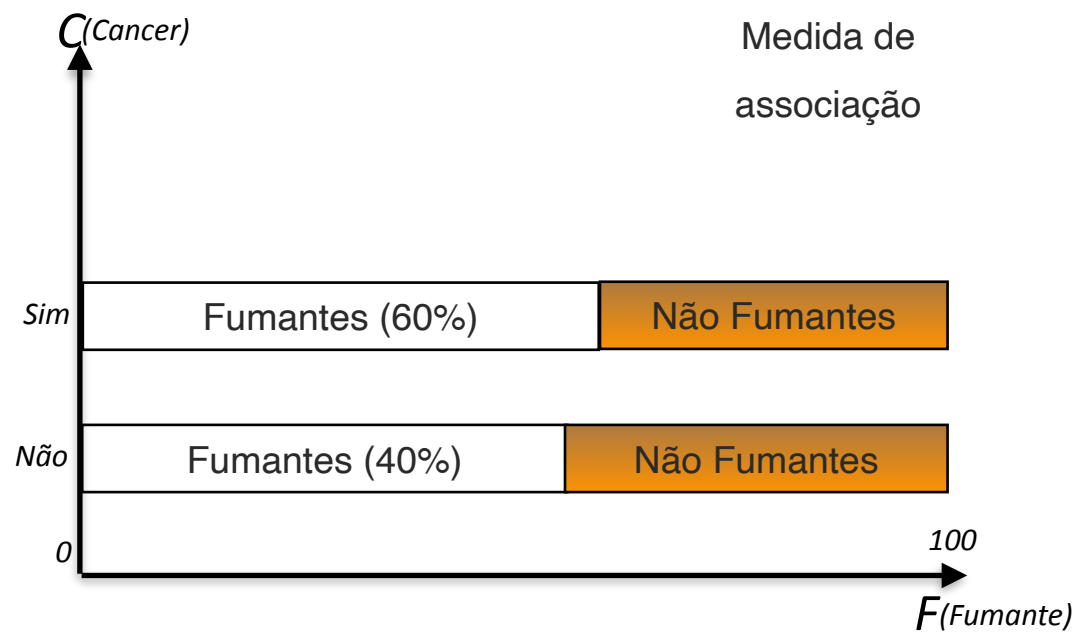
$$f_i = \frac{n_i}{N}$$

Então:

$$f_{ij} = f_i \cdot f_j$$

$$\chi^2 = n \sum_{i=1}^r \sum_{j=1}^s \frac{((f_{ij} - f_{ij}^*))^2}{f_{ij}^*} \Rightarrow$$

se,  $\chi^2 \rightarrow$  alto  $\Rightarrow$  associado  
se,  $\chi^2 \rightarrow 0 \Rightarrow$  não associados



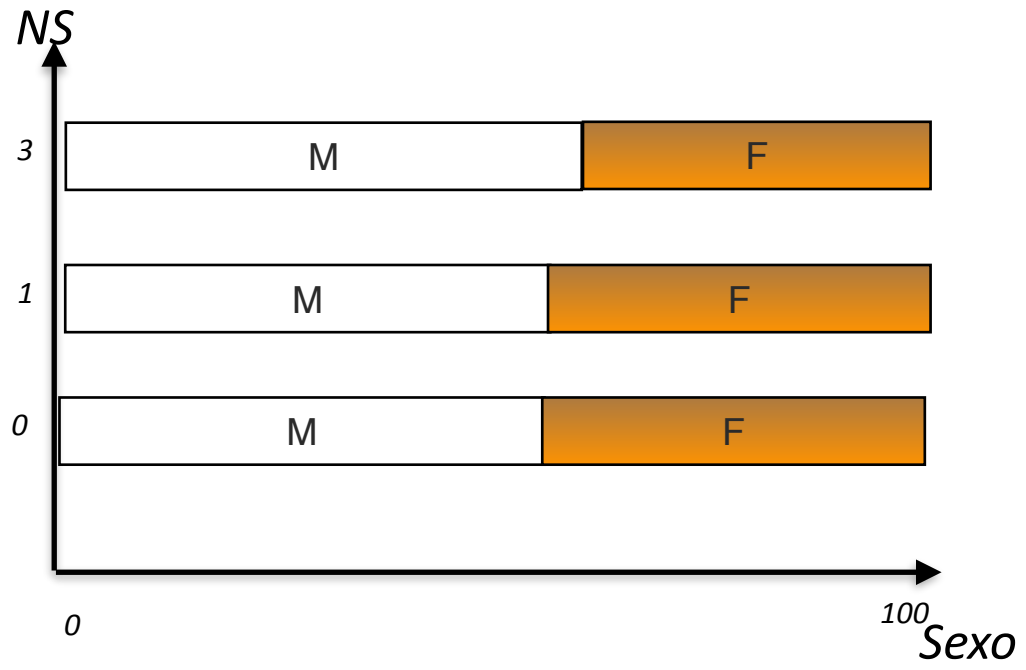


Nível de Satisfação			Sexo	
0	1	3	M	F

Variáveis Qualitativas	
Dependente	Independente
Ordinal	Nominal

Comparação entre os grupos

Testes Não Paramétricos  
Mann-Withney  
Wilcoxon



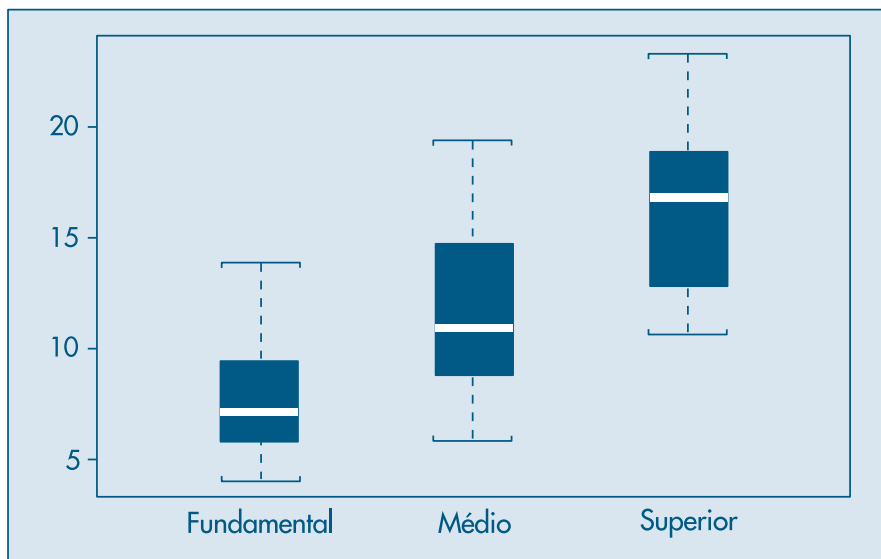
Existe diferença no nível de satisfação com o nosso produto, dado o sexo do cliente?



Grau de instrução	$n$	$\bar{s}$	$dp(S)$	$var(S)$	$s_{(1)}$	$q_1$	$q_2$	$q_3$	$s_{(n)}$
Fundamental	12	7,84	2,79	7,77	4,00	6,01	7,13	9,16	13,65
Médio	18	11,54	3,62	13,10	5,73	8,84	10,91	14,48	19,40
Superior	6	16,48	4,11	16,89	10,53	13,65	16,74	18,38	23,30
Todos	36	11,12	4,52	20,46	4,00	7,55	10,17	14,06	23,30

Dependente	Independente
Ordinal	Qualitativa
	Nominal

Testes Paramétricos  
T-student  
Anova



Educação afeta o salário?

