

Licenciatura em Gestão

Estatística 1

**Folhas de apoio às aulas de SPSS:
Medidas de Associação entre 2 Variáveis**

Elizabeth Reis

Ano lectivo de 2011 / 2012

ASSOCIAÇÃO LINEAR ENTRE 2 VARIÁVEIS QUANTITATIVAS – MENU GRAPHS

→ Diagrama de dispersão para as variáveis P4: Número de semanários lidos por mês e Pc: Número de diários lidos por semana com ajustamento da recta de regressão

Graphs

Legacy Dialogs

Interactive

Scatterplot:

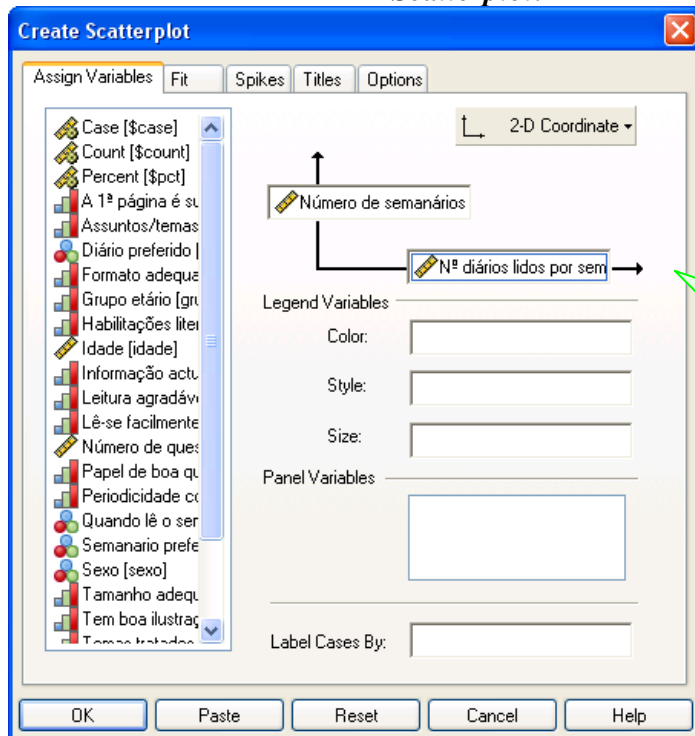
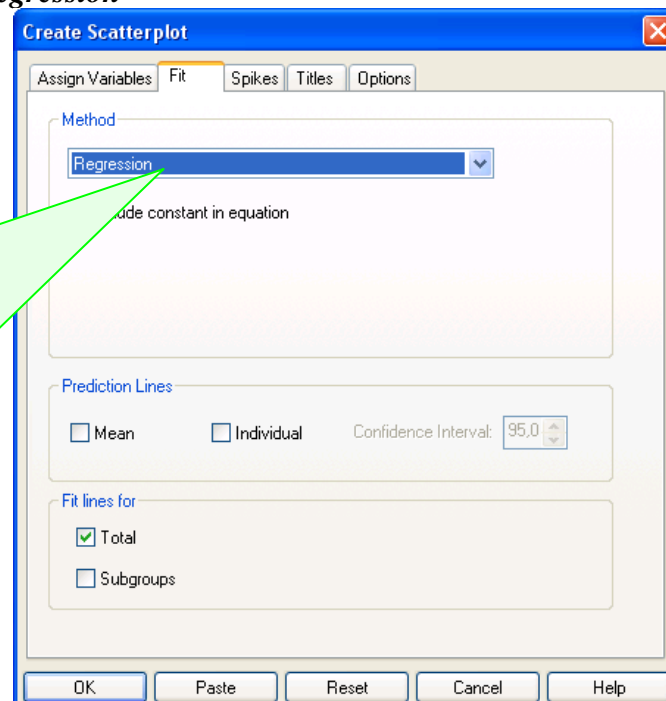
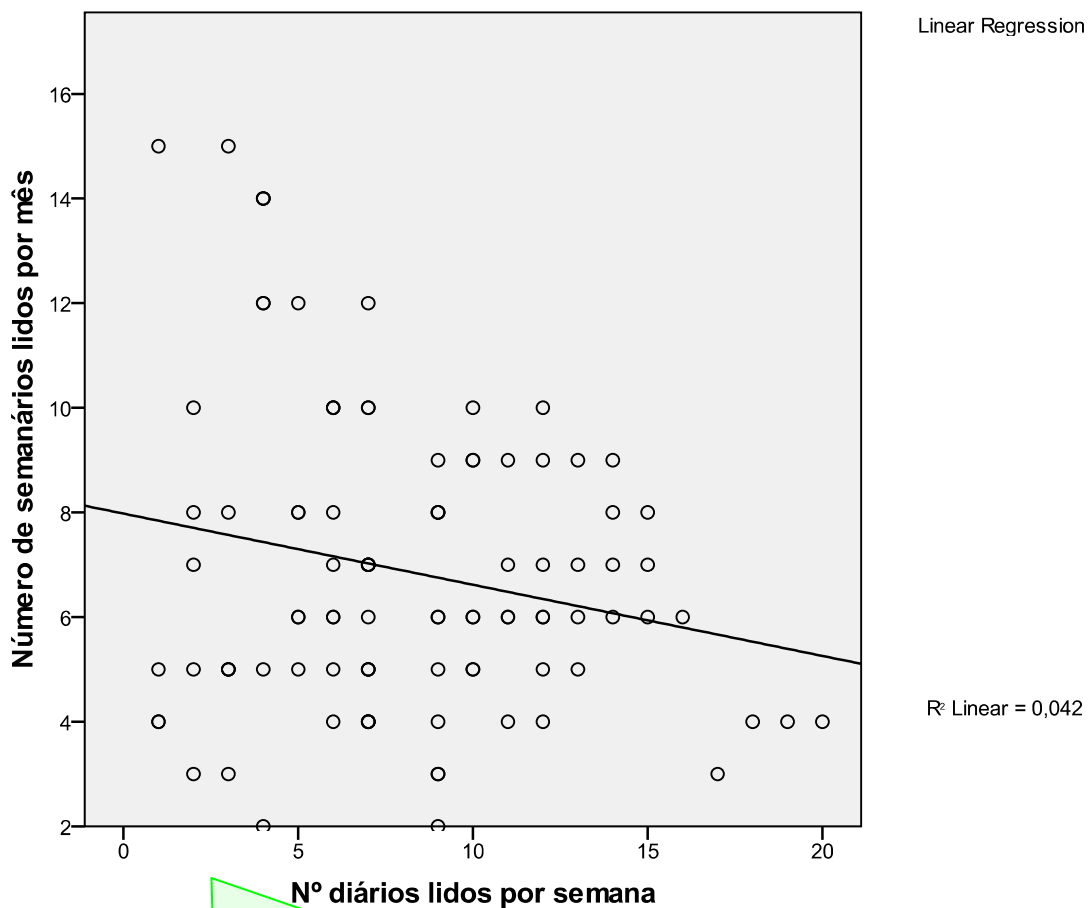


Diagrama de dispersão com a variável “Nº diários lidos por semana” no eixo do X e a variável “Nº de semanários” no eixo do Y

Fit: Regression

Pede-se que no diagrama de dispersão seja incluída a recta de regressão ajustada aos dados, sendo “Nº de semanários” a variável dependente (Y) e “Nº diários lidos por semana” a variável independente (X)





O diagrama de dispersão mostra uma relação linear muito reduzida entre as 2 variáveis.

A recta de regressão ajustada $\hat{Y} = 7,98 - 0,14X$,

com “Y = Nº de semanários lidos por mês”

e “X = Nº diários lidos por semana”

e o valor de $R^2 = 0,04$ mostram um muito reduzido poder explicativo do modelo de regressão: apenas 4% da variação no “Nº de semanários lidos por mês” pode ser explicada pela variação no “Nº de diários lidos por semana”; por cada jornal diário a mais lido por semana, o efeito sobre o número de semanários lidos por mês é apenas de menos 0,14 jornais.

ASSOCIAÇÃO LINEAR ENTRE 2 VARIÁVEIS QUANTITATIVAS – MENU ANALYZE ⇒ CORRELATE

→ Coeficiente de correlação linear de Pearson entre as variáveis P4: Número de semanários lidos por mês e Pc: Número de diários lidos por semana

Analyze

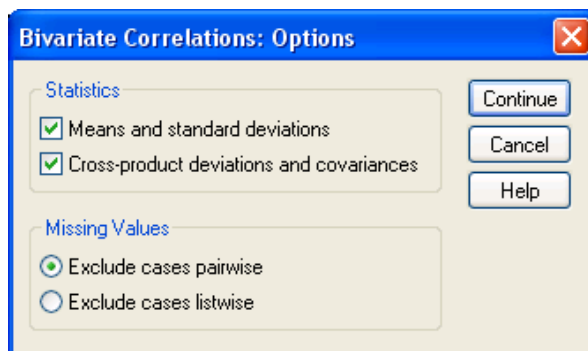
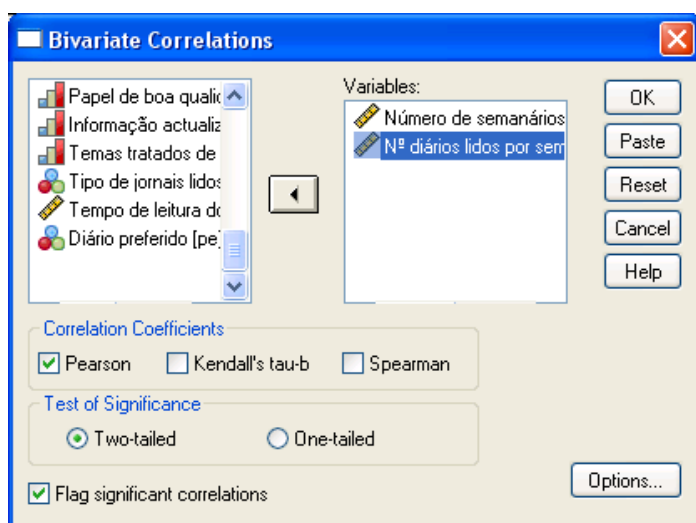
Correlate

Bivariate

Variables: P4 Pc

Correlation Coefficients: Pearson

OK



Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
Número de semanários lidos por mês	7,05	3,138	100
Nº diários lidos por semana	8,16	4,352	98

Para além do coeficiente de correlação linear de Pearson entre as variáveis “P₄: Nº de semanários lidos por mês” e “P_c: Nº de diários lidos por semana” foram também solicitadas medidas de estatística descritiva (média, desvio padrão e número de casos válidos) para cada variável e outras medidas de associação linear entre as 2 variáveis (covariâncias e somas de quadrados e produtos cruzados)

Correlations

		Número de semanários lidos por mês	Nº diários lidos por semana
Número de semanários lidos por mês	Pearson Correlation	1	-,205(*)
	Sig. (2-tailed)		,043
	Sum of Squares and Cross-products	974,750	-249,878
	Covariance	9,846	-2,576
	N	100	98
Nº diários lidos por semana	Pearson Correlation	-,205(*)	1
	Sig. (2-tailed)	,043	
	Sum of Squares and Cross-products	-249,878	1837,388
	Covariance	-2,576	18,942
	N	98	98

* Correlation is significant at the 0.05 level (2-tailed).

O coeficiente de correlação linear de Pearson entre as variáveis “P₄: Nº de semanários lidos por mês” e “P_c: Nº diários lidos por semana” é igual a -0,205 ⇒ identifica uma associação linear reduzida e negativa entre as duas variáveis.

Para o cálculo da correlação foram utilizados os valores da covariância (-2,576) e das variâncias de P₄ (9,846) e de P_c (18,942)

$$r_{P_4P_c} = \frac{Cov(P_4, P_c)}{\sqrt{Var(P_4)Var(P_c)}} = \frac{-249,878/97}{\sqrt{\frac{974,75}{99} \times \frac{1837,388}{97}}} = \frac{-2,576}{\sqrt{9,846 \times 18,952}} = -0,205$$

$$Cov(P_4, P_c) = \frac{\sum_{i=1}^n (P_{4i} - \bar{P}_4)(P_{ci} - \bar{P}_c)}{(n-1)} = \frac{-249,878}{97} = -2,576$$

$$Var(P_4) = \frac{\sum_{i=1}^n (P_{4i} - \bar{P}_4)^2}{(n-1)} = \frac{974,75}{99} = 9,846$$

$$Var(P_c) = \frac{\sum_{i=1}^n (P_{ci} - \bar{P}_c)^2}{(n-1)} = \frac{1837,388}{97} = 18,952$$

Sendo $\sum_{i=1}^n (P_{4i} - \bar{P}_4)^2$ a soma de quadrados para a variável P₄, $\sum_{i=1}^n (P_{ci} - \bar{P}_c)^2$ a soma de quadrados para P_c e $\sum_{i=1}^n (P_{4i} - \bar{P}_4)(P_{ci} - \bar{P}_c)$ a soma dos produtos cruzados entre P₄ e P_c.

ASSOCIAÇÃO LINEAR ENTRE 2 VARIÁVEIS ORDINAIS – MENU ANALYZE ⇒ CORRELATE

➔ Coeficiente de correlação linear de Spearman entre as variáveis P5a: Lê-se facilmente e P5b: Leitura agradável

Analyze

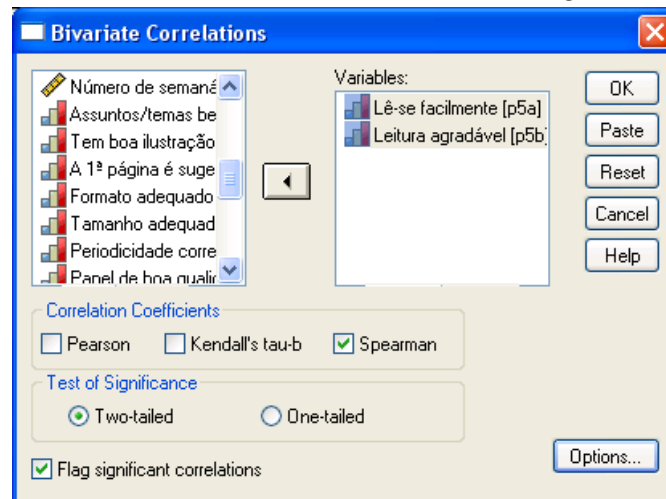
Correlate

Bivariate

Variables: P5a P5b

Correlation Coefficients: Spearman

OK



Correlations

			Lê-se facilmente	Leitura agradável
Spearman's rho	Lê-se facilmente	Correlation Coefficient	1,000	,582(**)
		Sig. (2-tailed)	.	,000
		N	100	100
	Leitura agradável	Correlation Coefficient	,582(**)	1,000
		Sig. (2-tailed)	,000	
		N	100	

O coeficiente de correlação de Spearman entre as variáveis “P_{5a}: Lê-se facilmente” e “P_{5b}: Leitura agradável” é igual a 0,582 ⇒ identifica uma associação linear média e positiva entre as duas variáveis.

**

O coeficiente de correlação de Spearman entre X_1 e X_2 calcula-se do seguinte modo, caso não existam empates nas ordenações das duas variáveis:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad \text{onde } d_i^2 = (r_{1i} - r_{2i})^2, \text{ sendo } r_{1i} \text{ e } r_{2i} \text{ os valores ordenados de } X_1 \text{ e } X_2,$$

respectivamente. Caso existam empates, r_s é calculado do seguinte modo:

$$r_s = \frac{(n^3 - n) - 6 \sum_{i=1}^n d_i^2 - (T_{X_1} + T_{X_2})/2}{\sqrt{(n^3 - n)^2 - (T_{X_1} + T_{X_2})(n^3 - n) + (T_{X_1} T_{X_2})}} \quad \text{onde } T_{X_i} = \sum_{j=1}^{g_i} (t_j^3 - t_j) \text{ é o factor de correcção}$$

para os empates da variável X_i e t_j é o número de observações empatadas em cada grupo de empates da variável X_i .

ASSOCIAÇÃO LINEAR ENTRE 2 VARIÁVEIS NOMINAIS E TABELA DE CONTIGÊNCIA – MENU ANALYZE ⇒ CROSSTABS

➔ Cruzamento de duas variáveis (Sexo e P1) e medidas de associação entre as 2 variáveis qualitativas

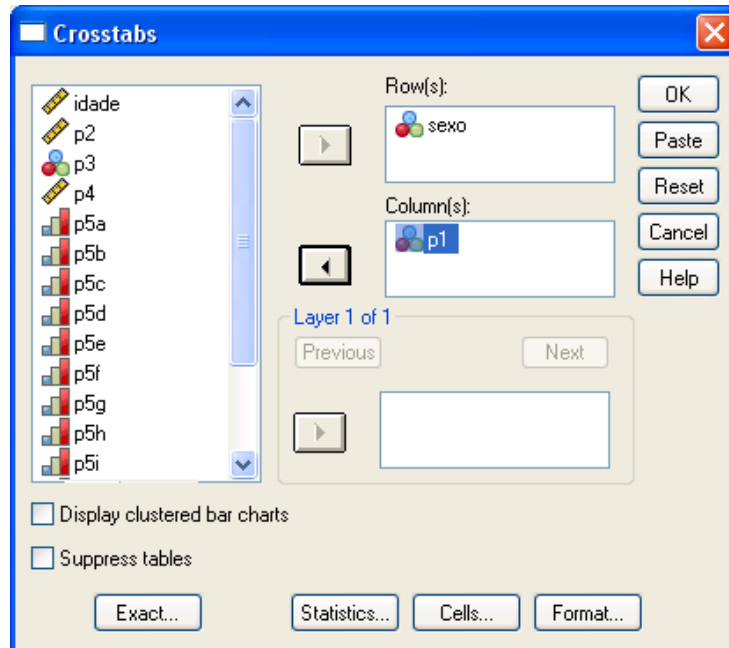
Analyze

Descriptive statistics

Crosstabs

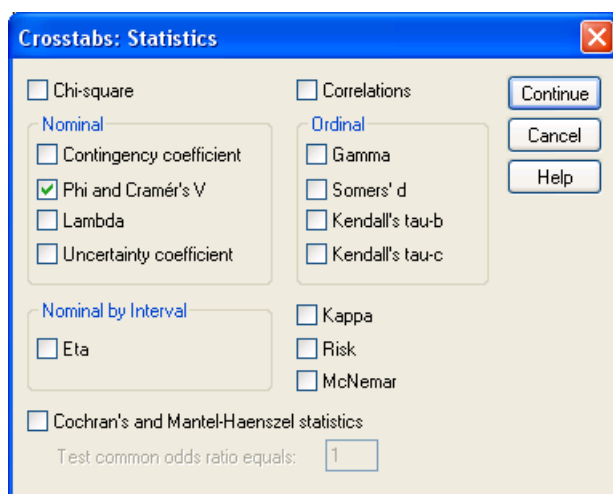
Row: Sexo

Column: P1



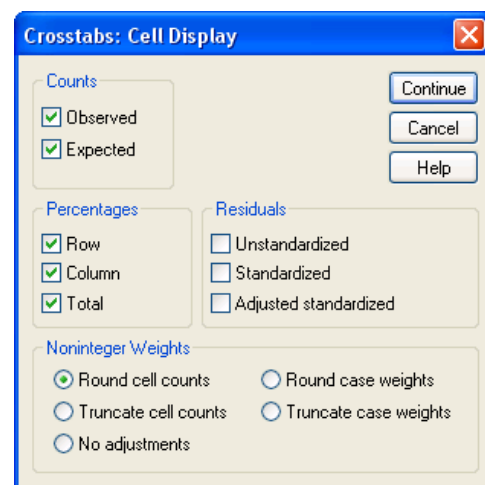
Statistics:

Nominal: Phi e Cramér's V



Cells

Counts: Observed Expected
Percents: Row Column Total



Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sexo * Semanario preferido	100	100,0%	0	,0%	100	100,0%

Sexo * Semanario preferido Crosstabulation

		Semanario preferido				
			expresso	semanario	sol	Total
Sexo	feminino	Count	14	5	15	34
		Expected Count	17,0	8,5	8,5	34,0
		% within sexo Sexo	41,2%	14,7%	44,1%	100,0%
		% within p1	28,0%	20,0%	60,0%	34,0%
		Semanario preferido				
		% of Total	14,0%	5,0%	15,0%	34,0%
	masculino	Count	36	20	10	66
		Expected Count	33,0	16,5	16,5	66,0
		% within sexo Sexo	54,5%	30,3%	15,2%	100,0%
		% within p1	72,0%	80,0%	40,0%	66,0%
		Semanario preferido				
		% of Total	36,0%	20,0%	10,0%	66,0%
Total	Count	50	25	25	100	
	Expected Count	50,0	25,0	25,0	100,0	
	% within sexo Sexo	50,0%	25,0%	25,0%	100,0%	
	% within p1	100,0%	100,0%	100,0%	100,0%	
	Semanario preferido					
	% of Total	50,0%	25,0%	25,0%	100,0%	

Tabela de contingência ou de dupla entrada para as categorias das variáveis Sexo e P₁ com a seguinte informação:

- frequências observadas, O_{ij} (*count*) correspondem ao número de observações que pertencem simultaneamente à categoria i da Variável Sexo e à categoria j da variável P₁
- frequências esperadas no caso de independência entre as 2 variáveis, E_{ij} (*expected count*), calculado do seguinte modo:

$$E_{ij} = \frac{O_i \times O_j}{n} \text{ sendo } O_i \text{ o total observado para a linha } i, O_j \text{ o total marginal para a coluna } j \text{ e } n \text{ a dimensão total da amostra.}$$

- frequência relativa para cada categoria da variável Sexo (*% within Sexo*)
- frequência relativa para cada categoria da variável Semanário Preferido (*% within P1*)
- frequência relativa em relação à amostra total (*% of total*)
- frequências marginais em linha e em coluna (*total*)

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,324	,005
	Cramer's V	,324	,005
N of Valid Cases		100	

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

O coeficiente V de Cramer é uma medida de associação não simétrica que toma valores entre 0 e 1 ($0 \leq V \leq 1$): valores próximos de 0 significam associação muito reduzida entre as variáveis; valores próximos de 1 apontam para associação elevada

$$V = \sqrt{\frac{X^2}{N[\min(c_1, c_2) - 1]}}$$

onde

$$X^2 = \sum_{i=1}^{c_1} \sum_{j=1}^{c_2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

com

c_1 = nº categorias da variável 1

c_2 = nº categorias da variável 2

Coeficiente Phi (Φ) para variáveis dicotómicas (tabelas 2 x 2):

$$\Phi = \sqrt{\frac{X^2}{N}}$$