

Análise de Correlação e medidas de associação

1. Introdução

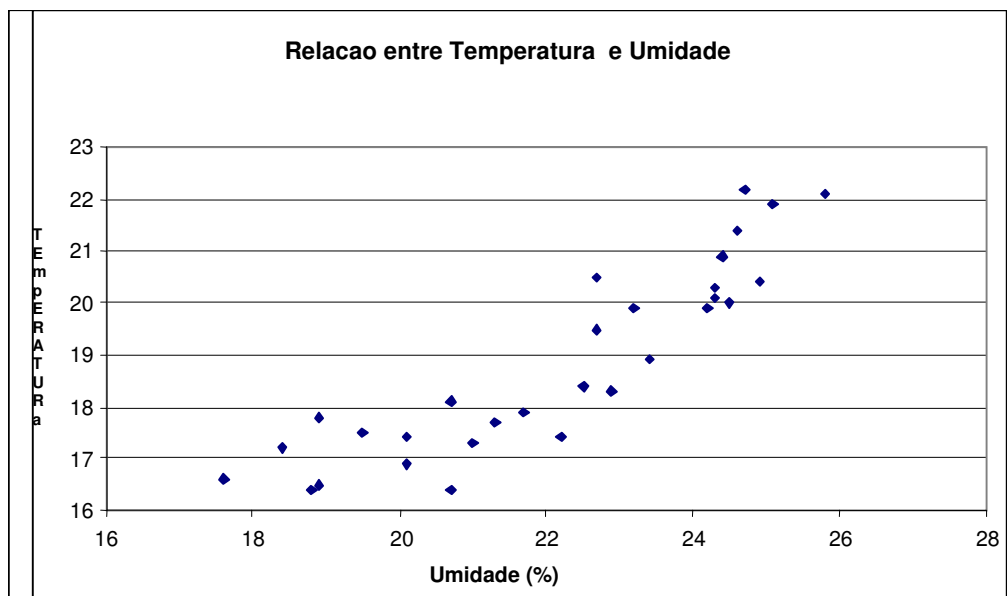
Muitas vezes precisamos avaliar o grau de relacionamento entre duas ou mais variáveis. É possível descobrir com precisão, o quanto uma variável interfere no resultado de outra. As técnicas associadas à Análise de Correlação representam uma ferramenta fundamental de aplicação nas Ciências Sociais e do comportamento, da Engenharia e das Ciências naturais. A importância de se conhecer os diferentes métodos e suas suposições de aplicação é exatamente pelo cuidado que se deve ter para não se utilizar uma técnica inadequada. Existem diversos critérios de avaliação desta relação, alguns próprios para variáveis que seguem uma distribuição normal e outros para variáveis que não seguem uma distribuição teórica conhecida. É comum a utilização do Coeficiente de correlação de Pearson. No entanto, existem situações em que o relacionamento entre duas variáveis não é linear, ou uma delas não é contínua ou as observações não são selecionadas aleatoriamente. Nestes casos, outras alternativas de coeficientes deverão ser aplicadas. Entre as diversas alternativas, veremos aqui algumas das mais importantes: Coeficiente de Spearman e coeficiente de Contingência.

Segundo o dicionário Aurélio, **correlação** significa *relação mútua entre dois termos*, qualidade de correlativo, correspondência. Correlacionar, significa estabelecer relação ou correlação entre; ter correlação. Enquanto que a palavra **regressão** significa: *ato ou efeito de regressar*, de voltar, retorno, regresso; dependência funcional entre duas ou mais variáveis aleatórias. A palavra *regredir* significa ir em marcha regressiva, retroceder.

Mas, onde e como surgiram os termos correlação e regressão? Foi Francis Galton (1822-1911), primo de Charles Darwin, quem usou pela primeira vez esses termos, cujo trabalho influenciou a Estatística e a Psicologia. Galton publicou o livro **Gênio Hereditário**, em 1869, onde aplicou conceitos estatísticos a problemas da hereditariedade. O primeiro relato onde Galton usou o termo “co-relações” foi em 1888.

2. Diagramas de Dispersão

Um dos métodos mais usados para a investigação de pares de dados é a utilização de diagramas de dispersão cartesianos (ou seja, os conhecidos diagramas x-y). Geometricamente, um diagrama de dispersão é simplesmente uma coleção de pontos num plano cujas duas coordenadas Cartesianas são os valores de cada membro do par de dados. E para quê fazemos um diagrama de dispersão? Este é o melhor método de examinar os dados no que se refere à ocorrência de tendências (lineares ou não), agrupamentos de uma ou mais variáveis, mudanças de espalhamento de uma variável em relação à outra e verificar a ocorrência dos valores discrepantes. Observe o exemplo a seguir:



Podemos notar pela análise da figura acima, a relação linear entre as duas variáveis. Os coeficientes apresentados a seguir nos auxiliam na quantificação do grau de relacionamento entre as variáveis de interesse.

3. A Covariância e o Coeficiente de correlação de Pearson

Quando estudamos a relação entre duas variáveis X e Y devemos primeiramente compreender o conceito de covariância. Se a variância é uma estatística através da qual chegamos ao desvio padrão que é uma medida de dispersão, da mesma maneira a covariância é uma estatística através da qual chegamos ao coeficiente de correlação que mede o grau de associação “linear” entre duas variáveis aleatórias X e Y.

Observe o exemplo abaixo. Sejam X e Y duas variáveis aleatórias quaisquer, que tomam os seguintes valores:

Tabela 1. Cálculo do coeficiente de correlação de Pearson

X	Y	DESVIOX ($X_i - \bar{X}$)	DESVIOY ($Y_i - \bar{Y}$)	DXDY ($X_i - \bar{X}$) * ($Y_i - \bar{Y}$)	DESVIOX2 ($X_i - \bar{X}$) ²	DESVIOY2 ($Y_i - \bar{Y}$) ²	PRE_1 Y=a+bX
1	0	-4,50	-6,00	27,00	20,25	36,00	,92727
2	2	-3,50	-4,00	14,00	12,25	16,00	2,05455
3	4	-2,50	-2,00	5,00	6,25	4,00	3,18182
4	5	-1,50	-1,00	1,50	2,25	1,00	4,30909
5	5	-,50	-1,00	,50	,25	1,00	5,43636
6	8	,50	2,00	1,00	,25	4,00	6,56364
7	7	1,50	1,00	1,50	2,25	1,00	7,69091
8	7	2,50	1,00	2,50	6,25	1,00	8,81818
9	11	3,50	5,00	17,50	12,25	25,00	9,94545
10	11	4,50	5,00	22,50	20,25	25,00	11,07273
55	60	0	0	93,00	82,50	114,00	60,0000

Na tabela acima está uma ilustração dos cálculos dos componentes da covariância e correlação.

A Figura abaixo mostra a relação entre as duas variáveis X e Y, bem como a linha ajustada a esses valores pelo método de mínimos quadrados. Observe que a média de X é 5,5 e a média de Y é 6,0, e que elas estão formadas pelas linhas paralelas ao eixo Y e ao eixo X respectivamente. Vejamos agora o que significa os desvios de cada ponto em relação a média. Observe que cada ponto está formado pelo par ordenado (X_i, Y_i) , onde X_i indica o valor da variável X e Y_i o valor da variável Y naquele ponto.

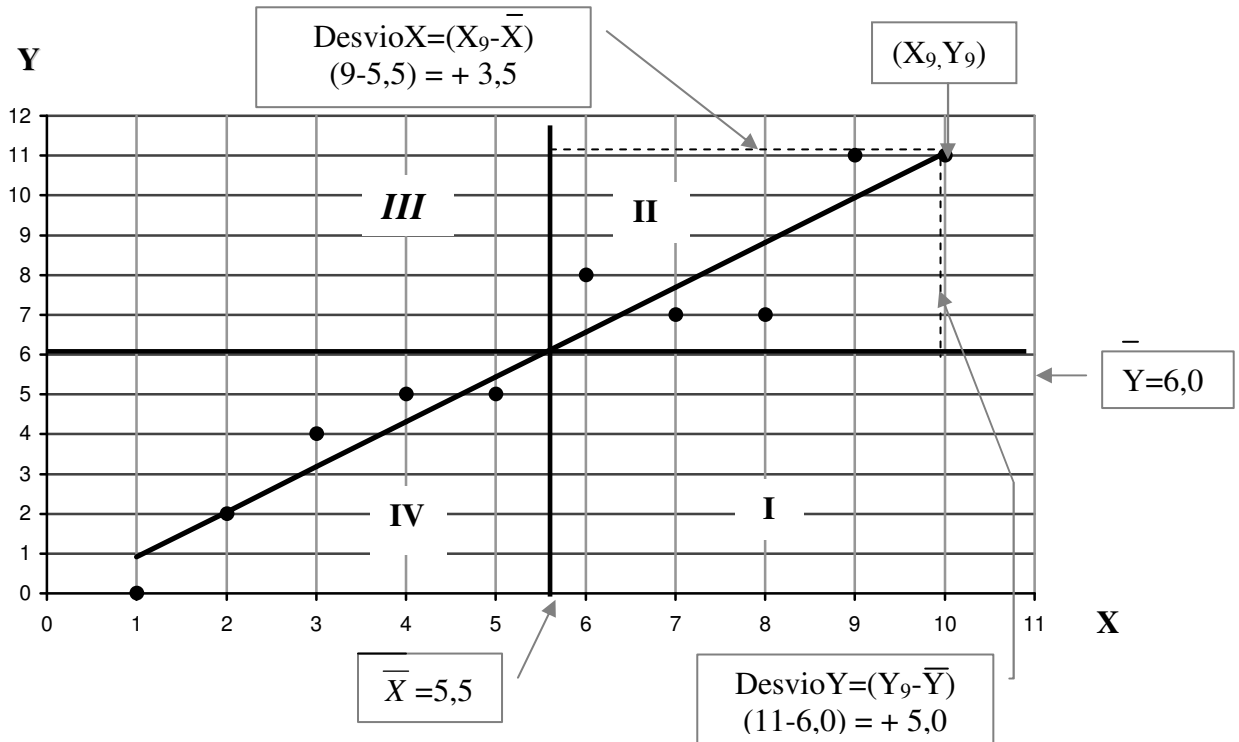


Figura 41. Relação entre X e Y

Tome agora, por exemplo,

$$\text{Desvio X} = (X_9 - \bar{X}) = (9 - 5,5) = + 3,5 \quad \text{e} \quad \text{Desvio Y} = (Y_9 - \bar{Y}) = (11 - 6,0) = + 5,0$$

O produto dos desvios:

$$\text{Desvíos X Desvio Y} = (X_9 - \bar{X}) * (Y_9 - \bar{Y}) = (9 - 5,5) * (11 - 6,0) = (+ 3,5) * (+ 5,0) = 17,5$$

Se calcularmos esses produtos para todos os valores de X e Y e somarmos temos o numerador da covariância de X e Y:

$$C(X, Y) = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{n} = \frac{93}{10} = 9,3$$

(1)

Logo, covariância significa co-variação, como as duas variáveis variam de forma conjunta. Agora, vejamos o que acontece se os pontos estivessem no quadrante I. Neste caso, os desvios de X seriam todos positivos, enquanto que os desvios de Y seriam todos negativos, logo os produtos tomarão valores negativos. O mesmo vai acontecer com os pontos do quadrante III, nele, os desvios de X tomarão valores negativos e os desvios de Y, valores positivos, logo os produtos tomarão valores negativos. Assim, se a maioria dos pontos caem nos quadrantes I e III a covariância tomará valores negativos, indicando que essas duas variáveis se relacionam de forma negativa ou inversa, ou seja, que quando uma cresce a outra diminui e vice-versa.

Quando os pontos se distribuem nos quatro quadrantes, haverá valores positivos e negativos, logo a soma tenderá para zero, e neste caso, afirmaremos que não existe relação linear entre essas variáveis. Observamos que esta estatística tenderá para zero, mesmo havendo uma relação que não for linear, por exemplo se os dados tivessem o formato de uma parábola, ou relação quadrática.

A pesar da covariância ser uma estatística adequada para medir relação linear entre duas variáveis, ela não é adequada para comparar graus de relação entre variáveis, dado que ela está influenciada pelas unidades de medida de cada variável, que pode ser metros, quilometro, quilogramas, centímetros, etc. Para evitar a influência da ordem de grandeza e unidades de cada variável, dividimos a covariância pelo desvio padrão de X e de Y, dando origem ao coeficiente de correlação de Pearson:

Notação:

Coeficiente de correlação amostral: **r**

Coeficiente de correlação populacional: **ρ**

$$r = \frac{C(X, Y)}{S_Y * S_X} \qquad r = \frac{9,3}{2,8723 * 3,3764} = 0,95896$$

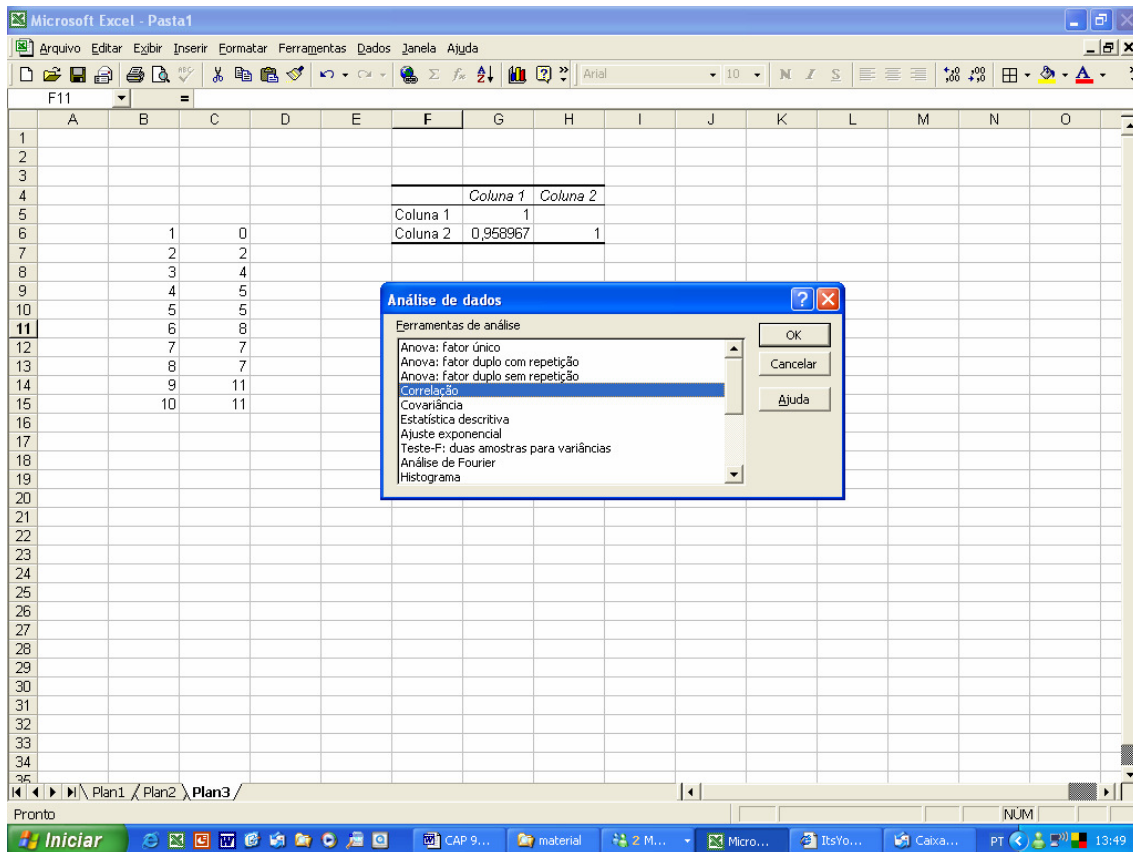
(2)

Onde: $S_x^2 = 82,5 / 10 = 8,25 \rightarrow S_x = 2,8723$
 $S_y^2 = 114,0 / 10 = 11,4 \rightarrow S_y = 3,3764$

Como o coeficiente de correlação está isento de unidades e da ordem de grandeza das variáveis, este toma valores entre -1 e 1.

Quando a relação é positiva r tomará o valor 1 quando a relação é perfeita.
 Quando a relação é negativa r tomará o valor -1 quando a relação é perfeita.
 Quando a relação é difusa ou não linear r será igual a 0.

No Excel, usando a opção Correlação em “Análise de dados”, obtemos:



O coeficiente de Determinação

Outro coeficiente amplamente utilizado para mensurar o grau de correlação entre duas variáveis é o coeficiente de determinação. É definido elevando o valor do coeficiente de Pearson ao quadrado e denotado por r^2 . Pode ser interpretado como sendo a proporção da variação de Y que é explicada pela variável X (e vice versa).

Muito embora o coeficiente de determinação seja relativamente fácil de interpretar, ele não pode ser testado estatisticamente. Contudo, a raiz quadrada do coeficiente de determinação, que é o coeficiente de correlação (r), pode ser testada estatisticamente, pois está associada a uma estatística de teste que é distribuída segundo uma distribuição t de Student, quando a correlação populacional $\rho = 0$.

O coeficiente de correlação para dados populacionais é:

$$\text{População: } \rho = \sqrt{\rho^2}$$

O coeficiente de correlação para dados amostrais é:

$$\text{Amostra: } r = \sqrt{r^2}$$

Significância do coeficiente de correlação

Para comprovarmos se o coeficiente de correlação é significativo, devemos realizar o seguinte teste de hipóteses:

Hipóteses:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

A estatística de teste é

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

com n-2 graus de liberdade na tabela t de Student. Caso o valor de t_c seja superior ao valor crítico de t, devermos rejeitar a hipótese nula. Se a hipótese nula, ao nível de significância α , for rejeitada podemos concluir que efetivamente existe uma relação significativa entre as variáveis.

Exemplo:

1. Para estudar a poluição de um rio, um cientista mediu a concentração de um determinado composto orgânico (Y) e a precipitação pluviométrica na semana anterior (X):

X	Y
0,91	0,10
1,33	1,10
4,19	3,40
2,68	2,10
1,86	2,60
1,17	1,00

Existe alguma relação entre o nível de poluição e a precipitação pluviométrica?
Teste sua significância, ao nível de 5%.

Calculando a média de X e de Y temos $\bar{X} = 2,023$ e $\bar{Y} = 1,717$.

Calculando a covariância entre X e Y pela expressão (1),

$$C(X, Y) = \frac{(0,91 - 2,023)(0,10 - 1,717) + (1,33 - 2,023)(1,10 - 1,717) + \dots + (1,17 - 2,023)(1,00 - 1,717)}{6}$$

$$= 1,0989$$

Calculando os desvios padrões de X e Y temos: $S_X = 1,125$ e $S_Y = 1,10$

E assim, pela expressão (2),

$$r = \frac{C(X, Y)}{S_Y * S_X} = \frac{1,0989}{1,125 * 1,1} = 0,888$$

Testando a significância do coeficiente,

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,888\sqrt{6-2}}{\sqrt{1-(0,888)^2}} = 3,86$$

O valor crítico de t para $n-2 = 4$ graus de liberdade e 5% de nível de significância é 2,78. Note que o teste de significância do coeficiente será sempre bilateral.

Como o valor calculado de t é superior ao valor crítico, podemos concluir que existem evidências suficientes para afirmar que o composto orgânico (Y) e a precipitação pluviométrica (X) estejam correlacionados.

2. Procurando quantificar os efeitos da escassez de sono sobre a capacidade de resolução de problemas simples, um agente tomou ao acaso 10 sujeitos e os submeteu a experimentação. Deixou-os sem dormir por diferentes números de horas, após o que solicitou que os mesmos resolvessem os itens "contas de adicionar" de um teste. Obteve, assim, os seguintes dados:

No de erros - Y	Horas sem dormir - X
8	8
6	8
6	12
10	12
8	16
14	16
14	20
12	20
16	24
12	24

Calcule o coeficiente de correlação linear de Pearson e teste a sua significância ao nível de 1%.

Calculando a média de X e de Y temos $\bar{X} = 16$ e $\bar{Y} = 10,6$.

Calculando a covariância entre X e Y pela expressão (1),

$$C(X, Y) = \frac{(8-16)(8-10,6) + (8-16)(6-10,6) + \dots + (24-16)(12-10,6)}{10} = 15,2$$

Calculando os desvios padrões de X e Y temos: $S_X = 5,656854$ e $S_Y = 3,52611$

E assim, pela expressão (2),

$$r = \frac{C(X,Y)}{S_y * S_x} = \frac{15,2}{5,656854 * 3,52611} = 0,801467$$

Obs: procure sempre usar o maior número de casas decimais possível.

Usando a planilha Excel poderíamos também obter uma matriz de covariância, que irá nos fornecer a covariância entre X e Y além da variância de X e de Y.

The screenshot shows an Excel spreadsheet with data in columns C and D (rows 6-15). A dialog box titled 'Covariância' is open, showing the input range as '\$C\$6:\$D\$15' and the output range as '\$F\$22'. Below the dialog, a small table shows the resulting covariance matrix:

	Coluna 1	Coluna 2
Coluna 1	11,24000	
Coluna 2	15,2000	32,000

Agora testando a significância do coeficiente,

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,801467 \sqrt{10-2}}{\sqrt{1-(0,801467)^2}} = 3,79$$

O valor crítico de t para $n-2 = 8$ graus de liberdade e 1% de nível de significância é 3,355 (bilateral).

Como o valor calculado de t é superior ao valor crítico, podemos concluir que existem evidências suficientes para afirmar que o número de horas sem dormir (X) influencia significativamente o número de erros (Y).

4. Medidas de Associação

Freqüentemente estamos interessados em verificar a existência de associação entre dois conjuntos de escores e também o grau desta associação. No caso paramétrico, a medida usual é o coeficiente de correlação r de Pearson que exige mensuração dos escores no mínimo ao nível intervalar. Ainda, se estivermos interessados em comprovar a significância de um valor observado de r de Pearson deveremos supor que os escores provenham de uma distribuição normal. Quando estas suposições não são atendidas podemos utilizar um dos coeficientes de correlação não-paramétricos e suas respectivas provas de significância.

4.1 Coeficiente de Contingência C

Este coeficiente mede a associação entre dois conjuntos de atributos quando um ou ambos os conjuntos são medidos em escala nominal.

Considere uma tabela de contingência $k \times r$, que representa as freqüências cruzadas dos escores A (divididos em k categorias) e escores B (divididos em r categorias). O grau de associação entre dois conjuntos de atributos é calculado por:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \text{ onde } \chi^2 \text{ é a estatística Qui-quadrado.}$$

O p -valor associado ao valor da estatística Qui-quadrado com $(r-1) \times (k-1)$ graus de liberdade é a prova de significância do coeficiente de contingência C .

O coeficiente C se caracteriza por assumir valor zero quando há inexistência de associação porém nunca será igual à 1. O limite superior do coeficiente é dado por

$\sqrt{\frac{k-1}{k}}$ (quando $k = r$). Note que para calcular o coeficiente C , a tabela de contingência deve satisfazer as restrições do teste Qui-quadrado.

Exemplo:

Estudantes de escolas particulares e de escolas públicas selecionados aleatoriamente foram submetidos a testes padronizados de conhecimento, e produziram os resultados abaixo. Verifique o grau de associação entre as variáveis mensuradas e teste a significância ao nível de 5%.

Escola	Escore			
	0 - 275	276-350	351-425	426-500
Particular	6	14	17	9
Pública	30	32	17	3

Queremos aqui verificar o grau de associação entre as variáveis “Escola” e “Escore de conhecimento”. A variável Escola é mensurada em nível nominal, o que inviabiliza a utilização do coeficiente r de Pearson.

Obtendo então o coeficiente de Contingência, necessitamos inicialmente calcular o valor da estatística χ^2 :

Freq.	6	14	17	9
Obs.	30	32	17	3
Freq.	12,94	16,53	12,22	4,31
Esp.	23,06	29,47	21,78	7,69

$$\chi^2 = \frac{(6-12,94)^2}{12,94} + \frac{(14-16,53)^2}{16,53} + \dots + \frac{(3-7,69)^2}{7,69} = 17,28$$

O coeficiente de contingência será:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{17,28}{128 + 17,28}} = 0,345$$

Para testar a significância do coeficiente precisamos verificar o valor crítico de χ^2 considerando $\alpha=0,05$ e $(r-1) \times (k-1) = 3$ graus de liberdade. Este valor é igual a 7,81. Comparando com o valor calculado de 17,28, podemos admitir a existência de associação significativa entre a Escola e o escore de conhecimento. Analisando atentamente, poderíamos acrescentar que o fato de um estudante pertencer a uma escola particular faz com que ele obtenha um escore de conhecimento mais alto.

4.2 – Coeficiente de correlação de Spearman

É uma medida de associação que exige que ambas as variáveis se apresentem em escala de mensuração pelo menos ordinal. Basicamente equivale ao coeficiente de correlação de Pearson aplicado a dados ordenados. Assim,

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = r_s$$

ou seja, o coeficiente de correlação de Spearman se utiliza da expressão do coeficiente de Pearson porém calculado com postos. Esta expressão equivale à

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad \text{onde } d_i = x_i - y_i \text{ a diferença de postos dos escores X e Y.}$$

Para verificar a significância do valor observado de r_s podemos usar a expressão de t de Student

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}} \quad \text{onde } t \text{ tem } n-2 \text{ graus de liberdade.}$$

Exemplo:

As notas obtidas por 10 estudantes de Administração e o seu QI (quociente de inteligência) são apresentadas no quadro abaixo

Notas	8	9,5	10	9,1	6,5	9	9,5	5,2	9,1	9,3
QI	127	149	150	135	122	129	142	100	136	139

Utilize o coeficiente de Spearman para verificar se as variáveis estão associadas e qual o seu grau de associação.

Inicialmente ordenamos os valores originais, transformando-os em postos. Aqui então substituímos os valores originais pelos seus respectivos postos, ou seja, o menor valor da variável em questão será substituído pelo valor 1 e assim por diante. Em seguida, calculamos as diferenças de postos:

Notas	3	8,5	10	5,5	2	4	8,5	1	5,5	7
QI	3	9	10	5	2	4	8	1	6	7
d_i	0	-0,5	0	0,5	0	0	0,5	0	-0,5	0
$(d_i)^2$	0	0,25	0	0,25	0	0	0,25	0	0,25	0

Calculando o coeficiente:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} = 1 - \frac{6(0^2 + 0,25^2 + \dots + 0^2)}{10^3 - 10} = \frac{6 \cdot 0,25}{990} = 0,998$$

Verificando a significância estatística do coeficiente:

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}} = 0,998 \sqrt{\frac{8}{1-(0,998)^2}} = 0,998 \sqrt{\frac{8}{0,004}} = 44,63$$

O valor crítico da estatística t de Student é obtido definindo-se $n-2 = 8$ graus de liberdade e o nível de significância, que admitiremos igual a 1%. Este valor é igual a 3,36. Mais uma vez temos aqui um teste bilateral pois estamos verificando se o coeficiente é diferente de zero.

Assim, podemos comprovar que o coeficiente de associação é altamente significativo, ou seja, existem fortes indícios que apontam para notas altas obtidas por aqueles que possuem maiores quocientes de inteligência.

5. Ampliando seus conhecimentos

Teste de Kappa

O Teste de Kappa é uma medida de concordância interobservador e mede o grau de concordância além do que seria esperado tão somente pelo acaso.

Para descrevermos se há ou não concordância entre dois ou mais avaliadores, ou entre dois métodos de classificação, utilizamos a medida Kappa que é baseada no número de respostas concordantes, ou seja, no número de casos cujo resultado é o mesmo entre os avaliadores. Esta medida de concordância assume valor máximo igual a 1, que representa total concordância ou ainda pode assumir valores próximos e até abaixo de 0, os quais indicam nenhuma concordância.

O coeficiente Kappa é calculado a partir da seguinte fórmula:

$$\text{Kappa} = \frac{P_O - P_E}{1 - P_E}$$

Onde $P_O = \frac{\text{número de concordâncias}}{\text{número de concordâncias} + \text{número de discordâncias}}$

e $P_E = \sum_{i=1}^n (p_{i1} \times p_{i2})$ sendo que:

- n é o número de categorias;
- i é o índice da categoria (que vale de 1 a n);
- p_{i1} é a proporção de ocorrência da categoria i para o avaliador 1;
- p_{i2} é a proporção de ocorrência da categoria i para o avaliador 2;

Para avaliar se a concordância é razoável, Landis, JR e Koch, GG (1977) sugerem a seguinte interpretação:

Valores obtidos de Kappa	Interpretação
<0	Nenhuma concordância
0-0,19	Concordância pobre
0,20-0,39	Concordância leve
0,40-0,59	Concordância moderada
0,60-0,79	Concordância substancial
0,80-1,00	Concordância quase perfeita

Exemplo:

Em certo órgão de financiamento, em cada edital aberto se apresentam diversos pesquisadores que enviam projetos solicitando recursos para desenvolvê-los. Estes projetos recebem uma avaliação, muitas vezes subjetiva, baseada na opinião de um consultor.

Considere a tabela abaixo que resume as avaliações feitas por dois avaliadores a 30 projetos que concorrem ao financiamento. O interesse deste estudo é saber qual é a concordância entre estes dois profissionais e se há alguma classificação com concordância maior do que as demais.

		AVALIADOR 2			Total
		A	B	C	
AVALIADOR 1	A	14 (0,47)	1 (0,03)	1 (0,03)	16 (0,53)
	B	3 (0,10)	3 (0,10)	2 (0,07)	8 (0,27)
	C	0 (0,00)	1 (0,03)	5 (0,17)	6 (0,20)
Total		17 (0,57)	5 (0,16)	8 (0,27)	30 (1,00)

* entre parênteses as proporções

Calculando o coeficiente Kappa:

$$P_o = \frac{14 + 3 + 5}{30} = \frac{22}{30} = 0,7333$$

$$P_E = \sum_{i=1}^n (p_{i1} \times p_{i2}) = (0,57 \cdot 0,53) + (0,16 \cdot 0,27) + (0,27 \cdot 0,20) = 0,3021 + 0,0432 + 0,054 =$$

$$= 0,3993$$

$$\text{Kappa} = \frac{0,733 - 0,3993}{1 - 0,3993} = 0,556$$

Note que a concordância geral pode ser considerada apenas moderada. Avaliando cada uma das três classificações, notamos que a concordância é alta quando os avaliadores atribuem o conceito A e o conceito C. No entanto, para atribuir o conceito B, um conceito intermediário, a concordância já não é tão satisfatória.

Fonte: Landis JR, Koch GG. *The measurement of observer agreement for categorical data*. Biometrics

6. Atividades de aplicação

1. Foi tomada uma amostra aleatória de 10 carregamentos recentes por caminhão feitos por uma companhia, anotada a distância em quilômetros e o tempo de entrega. Os dados seguem abaixo:

Carregamento	1	2	3	4	5	6	7	8	9	10
Distância em Km (X)	825	215	1070	550	480	920	1350	325	670	1215
Tempo de entrega em dias (Y)	3,5	1,0	4,0	2,0	1,0	3,0	4,5	1,5	3,0	5,0

- a) Construa o diagrama de dispersão;
- b) Calcule o coeficiente de correlação de Pearson para os dados desta amostra;
- c) Calcule o coeficiente de determinação;
- d) Verifique se o coeficiente de correlação é significativo ($\alpha=0,05$).

2. Para uma amostra de $n = 10$ tomadores de empréstimos em uma companhia financeira, o coeficiente de correlação entre a renda familiar média e débitos a descoberto de curto prazo foi calculado $r = 0,50$.

Teste a hipótese de que não existe correlação entre as duas variáveis, usando um nível de significância de 5%.

3. Para avaliar a relação entre habilidade verbal e habilidade matemática, escores de 8 estudantes foram obtidos, gerando a tabela abaixo:

	estudantes							
Escore	1	2	3	4	5	6	7	8
Matemática	80	50	36	58	72	60	56	68
Verbal	65	60	35	39	48	44	48	61

Calcule o coeficiente de correlação e teste sua significância.

4. Em um estudo conduzido com 10 pacientes, estes foram colocados sob uma dieta de baixas gorduras e altos carboidratos. Antes de iniciar a dieta, as medidas de colesterol e de triglicerídeos foram registradas para cada indivíduo .

- a) Construa um gráfico de dispersão para esses dados.
- b) Há alguma evidência de relação linear entre os níveis de colesterol e de triglicerídeos?
- c) Calcule o coeficiente de correlação de Spearman e teste sua significância.

paciente	Colesterol (mmol/l)	Triglicerídeos (mmol/l)
1	5,12	2,30
2	6,18	2,54
3	6,77	2,95
4	6,65	3,77
5	6,36	4,18
6	5,90	5,31
7	5,48	5,53
8	6,02	8,83
9	10,34	9,48
10	8,51	14,20

