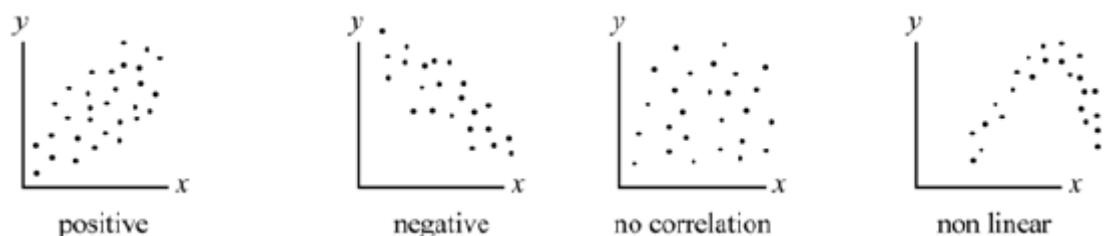


Correlação linear

Correlação linear é uma medida de associação linear entre duas variáveis contínuas, com

distribuição normal

Exemplos de correlações lineares, independência e não linear:



https://www.researchgate.net/profile/Mehdi_Farhadkhani/publication/259294558/figure/fig2/AS-297061200220165@1447836317956/A-representation-of-linear-vs-nonlinear-correlation-between-two-random-variables.png

Visualize o conceito de correlação: <http://rpsychologist.com/d3/correlation/>

A fórmula do coeficiente de correlação de Pearson é definida por:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

E a versão simplificada:

$$r = r_{xy} = \frac{\text{Cov}(x, y)}{S_x \times S_y}$$

Na versão simplificada pode-se observar que o cálculo da correlação envolve basicamente

três informações: a covariância (variância compartilhada das variáveis) e seus desvios-

padrão. Como a covariância é dividida pelos desvios-padrão (como ocorre no escore Z), ela

se torna uma medida padronizada, entre -1 e 1. O valor de -1 indica uma correlação negativa

perfeita, enquanto o valor 1 indica uma correlação positiva perfeita.

Como interpretar o coeficiente de correlação (r) em termos de sua magnitude:

Valor em módulo (intervalo)	Interpretação
0 a 0,32	Negligenciável
0,32 a 0,50	Baixa
0,50 a 0,70	Moderada
0,70 a 0,90	Elevada
0,90 +	Muito elevada

Prof. Dr. Wagner de Lara Machado

O coeficiente de determinação (r^2) indica a porção de variância comum entre duas variáveis.

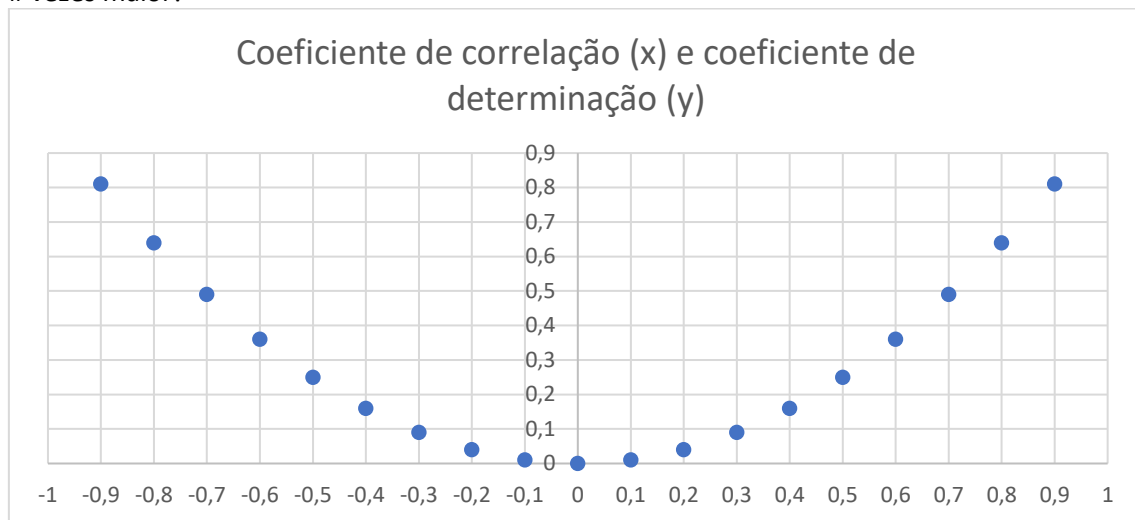
Observe os valores:

r	r^2 (%)
0,10	0,01
0,20	0,04
0,30	0,09
0,40	0,16
0,50	0,25
0,60	0,36
0,70	0,49
0,80	0,64
0,90	0,81

Note que a função da relação entre o coeficiente de correlação e o coeficiente de

determinação é quadrática. Logo uma correlação de 0,6 não é o dobro de 0,3, mas quatro

vezes maior!



Outros tipos de correlação

Há variações do coeficiente de correlação quando os dados não possuem distribuição normal

ou são de outros níveis de mensuração (e.g. categorias ordenadas ou variáveis dicotômicas):

Características das variáveis	Tipo de correlação
Duas variáveis contínuas sem distribuição normal ou discretas ordinais	Spearman
Duas variáveis dicotômicas	Tetracórica
Duas variáveis categóricas ordinais	Policórica
Uma variável dicotômica e outra contínua	Bisserial ou Ponto bisserial
Uma variável categórica ordinal e uma contínua	Polisserial

Mais sobre correlações:

<https://bookdown.org/burak2358/SARP-EN/correlation.html>

<https://www.andrews.edu/~calkins/math/edrm611/edrm13.htm>

<http://personality-project.org/r/psych/help/tetrachor.html>

Prof. Dr. Wagner de Lara Machado

Funções para calcular outros tipos de correlação no R <https://www.personality-project.org/r/psych/help/tetrachor.html>

Como reportar uma correlação: $r(n) = .55, p$

Em que r é o coeficiente de correlação, n é o tamanho amostral e p é a probabilidade

associada.

Exercícios:

Com o banco anexo ("banco_revisão.csv"), faça:

(Legendas: SV = satisfação de vida, AU = autonomia, AA = autoaceitação, Ans = ansiedade e Dep = depressão)

a) Salve o banco em um objeto do R

`banco_r<-read.csv(file.choose(),sep=";")`

b) Usando a função "plot", crie um *scatterplot* e observe a relação entre as variáveis:

SV e AU, SV e Dep, AA e Ans:

c) Usando o comando "abline(lm(x~y), col="red")", faça uma reta que melhor represente a

associação entre SV e AU:

d) Usando a função "pairs", observe a associação entre todas as variáveis do banco:

e) Utilizando a função "cor.test", calcule a correlação de Pearson entre todos os pares de

variáveis:

f) Calcule a correlação entre SV e AU usando as funções "cov" e "sd":

g) Utilizando a função "shapiro.test" verifique se as variáveis cumprem o pressuposto de

normalidade para o cálculo da correlação de Pearson:

h) Utilizando a função "cor", calcule a correlação de Spearman (não-paramétrica) entre todos

os pares de variáveis:

i) Com base na seguinte tabela de correlações, responda:

Tabela 5
Correlações de Spearman entre as Seis Subescalas da EBEP

Indicador	Subescala da EBEP							Média (DP)
	RP	AU	DA	CP	PV	AA	α	
Satisfação de Vida	0,22	0,23	0,42	0,22	0,47	0,51	0,79	25,64(5,26)
Afeto positivo	0,38	0,21	0,40	0,33	0,41	0,43	0,86	68,90(11,05)
Afeto negativo	-0,28	-0,33	-0,40	-0,05*	-0,16**	-0,37	0,89	43,64(12,40)
Balanço entre afetos	0,41	0,33	0,49	0,22	0,36	0,49	-	25,26(19,27)
Depressão (BDI)	-0,35	-0,36	-0,23	-0,24	-0,41	-0,49	0,77	5,38(4,74)

Nota: EBEP = Escala de Bem-estar Psicológico, RP = Relações positivas com outros, AU = Autonomia, DA = Domínio sobre o ambiente, CP = Crescimento pessoal, PV = Propósito na vida, AA = Autoaceitação, BDI = Inventário Beck de depressão, α = Alpha de Cronbach, DP = Desvio padrão. * = Não significativo ao nível de $p \leq 0,05$, ** = $p \leq 0,05$. Para todos os demais coeficientes de correlação, $p \leq 0,001$.

(<http://pepsic.bvsalud.org/pdf/avp/v12n2/v12n2a17.pdf>)

1) Qual a associação positiva de maior magnitude?

2) Qual a associação negativa de maior magnitude?

3) Qual o maior tamanho de efeito observado entre as variáveis?

4) Existem correlações que não possuem relevância teórico-prática, mesmo significativas?

Quais?