

2 - CORRELAÇÃO PARCIAL E CORRELAÇÃO MÚLTIPLA

A correlação parcial estabelece o grau de associação entre duas variáveis, mantendo controlada ou desconsiderando a influência de uma terceira variável, ou seja, estudamos o grau de relação em particular de duas variáveis, tornando as demais constantes. É usada para capturar o efeito de variáveis sobre os pares correlacionados.

Quando o interesse não é apenas estudar a relação entre duas variáveis, mas entre duas ou mais variáveis, temos então um caso de correlação múltipla.

2.1 - CORRELAÇÃO PARCIAL

Suponha que queiramos estudar o grau de correlação linear existente entre X_1 , X_2 e X_3 . Para efeito de notação consideraremos X_1 como variável dependente e X_2 e X_3 como independentes, então:

$$Y = X_1 = f(X_2, X_3) \quad \text{ou} \quad X_1 = \beta_{1,23} + \beta_{12,3}X_2 + \beta_{13,2}X_3$$

Podemos notar que na correlação simples Y é a variável dependente e X a independente, mas isto é um pouco arbitrário, pois testamos várias combinações possíveis entre as correlações parciais.

Na expressão da reta anterior poderemos ter:

- ⇒ $\beta_{12,3}$ = coeficiente angular da reta, expressando a relação entre X_1 e X_2 , com X_3 constante ou controlada;
- ⇒ $\beta_{13,2}$ = coeficiente angular da reta, expressando a relação entre X_1 e X_3 , com X_2 constante ou controlada;
- ⇒ $\beta_{1,23}$ = intercepto da reta no eixo Y .

Os números colocados após a vírgula são chamados de variáveis sob controle ou as que se consideram constantes, também servindo para fornecer a ordem da correlação.

Assim, na análise de correlação parcial, os valores da variável dependente X_1 e de X_2 são “ajustados” tendo em conta os valores correspondentes da variável de controle X_3 (isto é, retirando o efeito de X_3 sobre as duas primeiras).

Operacionalização:

- a) Para retirar o efeito de X_3 sobre X_1 , é feita uma regressão entre estas duas variáveis e consideram-se os resíduos obtidos (que correspondem à parte de X_1 não explicada por X_3);
- b) Para retirar o efeito de X_3 sobre X_2 , faz-se uma regressão entre estas duas variáveis e consideram-se também os resíduos respectivos (que correspondem à parte de X_2 não explicada por X_3);
- c) O coeficiente de correlação parcial pretendido é simplesmente dado pela correlação entre os resíduos das regressões a) e b).

Quando trabalharmos com amostras a expressão será: $X_1 = b_{1,23} + b_{12,3}X_2 + b_{13,2}X_3$

Considerando os coeficientes $b_{12,3}$ e $b_{13,2}$ podemos dizer que:

- O coeficiente de correlação linear múltipla indica que um certo grau de variação total de uma das variáveis (X_1 , no caso) pode ser explicado por todas as outras variáveis (X_2 e X_3), consideradas independentemente, atuando conjuntamente;
- O termo correlação parcial designa a correlação entre duas variáveis quaisquer, quando os efeitos de outras variáveis forem controlados. Através de um procedimento adequado, obteremos o grau de relação entre uma variável X_1 , considerada dependente e qualquer das outras variáveis, X_2 ou X_3 , consideradas independentes, controlando uma delas. Sendo que o procedimento pode ser generalizado para mais de uma variável controlada;
- $r_{12,3}$ indica a correlação entre X_1 e X_2 com X_3 controlada ($r_{12,3} = r_{21,3}$);
 $r_{23,1}$ = correlação entre X_2 e X_3 com X_1 controlada;
 $r_{46,1235}$ = correlação entre X_4 e X_6 com X_1, X_2, X_3 e X_5 controladas;

Assim, o coeficiente de correlação parcial de 1ª ordem entre as variáveis X_1 e X_2 com controle de X_3 , é dado em função dos coeficientes de correlação de ordem zero (ou sem controle ou correlação total), pela seguinte fórmula:

$$r_{12,3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

A fórmula pode ser generalizada para qualquer correlação parcial, envolvendo 3 variáveis, fazendo $i = 1, j = 2, k = 3$, teremos:

$$r_{ij,k} = \frac{r_{ij} - r_{ik} \cdot r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}}$$

Exemplo 2.1: A tabela apresenta os pesos, as alturas e as idades de 10 meninos.

Peso (kg)	X_1	30	32	24	30	26	35	25	23	35	31
Altura (cm)	X_2	145	150	125	157	127	140	132	107	155	145
Idade (anos)	X_3	7	10	7	11	8	10	10	6	12	9

Considerando esses dados calcule: a) r_{12}, r_{13} e r_{23} b) $r_{12,3}$ c) $r_{23,1}$ d) $r_{13,2}$

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	$X_1 \cdot X_2$	$X_1 \cdot X_3$	$X_2 \cdot X_3$
30	145	7	900	21025	49	4350	210	1015
32	150	10	1024	22500	100	4800	320	1500
24	125	7	576	15625	49	3000	168	875
30	157	11	900	24649	121	4710	330	1727
26	127	8	676	16192	64	3302	208	1016
35	140	10	1225	19600	100	4900	350	1400
25	132	10	625	17424	100	3300	250	1320
23	107	6	529	11499	36	2461	138	642
35	155	12	1225	24025	144	5425	420	1860
31	145	9	961	21025	81	4495	279	1305
Σ 291	1383	90	8641	193451	844	40743	2673	12660

$$(\sum x_1)^2 = 291^2 = 84.681 \quad (\sum x_2)^2 = 1.383^2 = 1.912.689 \quad (\sum x_3)^2 = 90^2 = 8.100$$

$$\text{Como: } r_{12,3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$\text{a) } r_{12} = \frac{S_{12}}{\sqrt{S_{11} \cdot S_{22}}} = \frac{40.734 - \frac{(291)(1383)}{10}}{\sqrt{\left(8641 - \frac{84681}{10}\right)\left(193451 - \frac{1912689}{10}\right)}} = 0,81$$

$$r_{13} = \frac{S_{13}}{\sqrt{S_{11} \cdot S_{33}}} = \frac{2673 - \frac{(291)(90)}{10}}{\sqrt{\left(8641 - \frac{84681}{10}\right)\left(844 - \frac{8100}{10}\right)}} = 0,704$$

$$r_{23} = \frac{S_{23}}{\sqrt{S_{22} \cdot S_{33}}} = \frac{12660 - \frac{(1383)(90)}{10}}{\sqrt{\left(193451 - \frac{1912689}{10}\right)\left(844 - \frac{8100}{10}\right)}} = 0,783$$

$$\text{b) } r_{12,3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{0,81 - (0,704)(0,783)}{\sqrt{(1 - 0,496)(1 - 0,612)}} = \frac{0,258}{0,442} = 0,586$$

$$\text{c) } r_{23,1} = \frac{r_{23} - r_{21} \cdot r_{31}}{\sqrt{(1 - r_{21}^2)(1 - r_{31}^2)}} = \frac{0,783 - (0,81)(0,704)}{\sqrt{(1 - 0,656)(1 - 0,496)}} = \frac{0,212}{0,416} = 0,509$$

$$\text{d) } r_{13,2} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{(1 - r_{12}^2)(1 - r_{32}^2)}} = \frac{0,704 - (0,81)(0,783)}{\sqrt{(1 - 0,656)(1 - 0,612)}} = \frac{0,069}{0,365} = 0,189$$

2.2 - CORRELAÇÃO PARCIAL ENVOLVENDO MAIS DE TRÊS VARIÁVEIS

A fórmula refere-se ao coeficiente de correlação parcial de 1ª ordem, ou seja, com controle de uma única variável, identificada, genericamente, pelo índice k..

Desejamos, por exemplo, obter o coeficiente de correlação parcial entre X₁ e X₂, conservando constante X₃ e X₄, a fórmula seria dada por:

$$r_{12,34} = \frac{r_{12,4} - r_{13,4} \cdot r_{23,4}}{\sqrt{(1 - r_{13,4}^2)(1 - r_{23,4}^2)}} = \frac{r_{12,3} - r_{14,3} \cdot r_{24,3}}{\sqrt{(1 - r_{14,3}^2)(1 - r_{24,3}^2)}}$$

A 1ª expressão pode ser simbolizado por $r_{12,34}$ e a 2ª por $r_{12,43}$. No 1º caso, temos a correlação entre as variáveis X_1 e X_2 quando X_3 e X_4 são mantidas constantes, no 2º caso, a correlação entre X_1 e X_2 quando X_4 e X_3 são mantidas constantes, de forma que a igualdade permanece.

A idéia de correlação parcial pode ser estendida ao caso de mais de três variáveis:

$$r_{12/34\dots n} = \frac{r_{12/34\dots(n-1)} - r_{1n/34\dots(n-1)} \cdot r_{2n/34\dots(n-1)}}{\sqrt{(1 - r_{1n/34\dots(n-1)}^2)(1 - r_{2n/34\dots(n-1)}^2)}}$$

Exemplo 2.2:

Cem alunos submeteram-se aos exames finais de: Complementos de Matemática, Estatística, Pesquisa Operacional e Processamento de Dados. Façamos inicialmente:

X_1 : Notas de Matemática

X_2 : Notas de Estatística

X_3 : Notas de Pesquisa Operacional

X_4 : Notas de Processamento de Dados

Obtendo-se os seguintes resultados:

$$\begin{array}{llllll} \bar{X}_1 = 80 & \bar{X}_2 = 60 & \bar{X}_3 = 30 & \bar{X}_4 = 50 & & \\ S_{x1} = 10 & S_{x2} = 5 & S_{x3} = 15 & S_{x4} = 5 & & \\ r_{12} = 0,50 & r_{13} = 0,75 & r_{14} = 0,80 & r_{23} = 0,60 & r_{24} = 0,60 & r_{34} = 0,90 \end{array}$$

a) Determinar: (i) $r_{12,34}$ (ii) $r_{13,24}$ (iii) $r_{14,23}$

Primeiro calculamos as correlações de 1ª ordem:

$$r_{12,4} = \frac{r_{12} - r_{14} \cdot r_{24}}{\sqrt{(1 - r_{14}^2)(1 - r_{24}^2)}} = \frac{0,5 - (0,8)(0,60)}{\sqrt{(1 - 0,8^2)(1 - 0,6^2)}} = 0,042$$

$$r_{13,4} = \frac{r_{13} - r_{14} \cdot r_{34}}{\sqrt{(1 - r_{14}^2)(1 - r_{34}^2)}} = \frac{0,75 - 0,8 \cdot 0,9}{\sqrt{(1 - 0,8^2)(1 - 0,9^2)}} = 0,115$$

$$r_{23,4} = \frac{r_{23} - r_{24} \cdot r_{34}}{\sqrt{(1 - r_{24}^2)(1 - r_{34}^2)}} = \frac{0,6 - 0,6 \cdot 0,9}{\sqrt{(1 - 0,6^2)(1 - 0,9^2)}} = 0,172$$

Logo:

$$r_{12,3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{0,5 - 0,75 \cdot 0,6}{\sqrt{(1 - 0,75^2)(1 - 0,6^2)}} = \frac{0,05}{0,529} = 0,094$$

$$r_{24,3} = \frac{r_{24} - r_{23} \cdot r_{43}}{\sqrt{(1 - r_{23}^2)(1 - r_{43}^2)}} = \frac{0,6 - 0,6 \cdot 0,9}{\sqrt{(1 - 0,6^2)(1 - 0,9^2)}} = \frac{0,06}{0,349} = 0,172$$

$$r_{14,3} = \frac{r_{14} - r_{13} \cdot r_{43}}{\sqrt{(1 - r_{13}^2)(1 - r_{43}^2)}} = \frac{0,8 - 0,75 \cdot 0,9}{\sqrt{(1 - 0,75^2)(1 - 0,9^2)}} = \frac{0,125}{0,288} = 0,434$$

$$(i) \quad r_{12,34} = \frac{r_{12,4} - r_{13,4} \cdot r_{23,4}}{\sqrt{(1 - r_{13,4}^2)(1 - r_{23,4}^2)}} \rightarrow r_{12,34} = \frac{0,042 - 0,115 \cdot 0,172}{\sqrt{(1 - 0,115^2)(1 - 0,172^2)}} = \frac{0,022}{0,978} = 0,022$$

$$(ii) \quad r_{13,24} = \frac{r_{13,4} - r_{12,4} \cdot r_{23,4}}{\sqrt{(1 - r_{12,4}^2)(1 - r_{23,4}^2)}} \rightarrow r_{13,24} = \frac{0,115 - 0,042 \cdot 0,172}{\sqrt{(1 - 0,042^2)(1 - 0,172^2)}} = \frac{0,108}{0,984} = 0,109$$

$$(iii) \quad r_{14,23} = \frac{r_{14,3} - r_{12,3} \cdot r_{24,3}}{\sqrt{(1 - r_{12,3}^2)(1 - r_{24,3}^2)}} \rightarrow r_{14,23} = \frac{0,434 - 0,094 \cdot 0,172}{\sqrt{(1 - 0,094^2)(1 - 0,172^2)}} = \frac{0,417}{0,981} = 0,426$$

2.3 – CORRELAÇÃO MÚLTIPLA

O coeficiente de correlação múltipla é o indicador de quanto da variação total da variável dependente (Y) é explicado pelo conjunto das variáveis independentes (explicativas X_1, X_2 e X_3, \dots, X_n).

De outra forma, o coeficiente de correlação múltipla mede a relação entre certo número de variáveis tomadas conjuntamente em vez da relação entre certo número de variáveis tomadas separadamente, ou seja, mede o relacionamento de Y com todos os regressos conjuntamente e pode ser expressa da seguinte maneira, considerando apenas três variáveis, X_1, X_2 e X_3 :

$$R_{1,23} = \frac{\sqrt{r_{12}^2 + r_{13}^2 - 2 \cdot r_{12} \cdot r_{13} \cdot r_{23}}}{\sqrt{1 - r_{23}^2}},$$

Genericamente:

$$R_{i,jk} = \frac{\sqrt{r_{ij}^2 + r_{ik}^2 - 2 \cdot r_{ij} \cdot r_{ik} \cdot r_{jk}}}{\sqrt{1 - r_{jk}^2}}$$

É fácil ver que, se as variáveis controladas tiverem correlação nula, $\text{Cov}(X_j, X_k) = 0$, o que equivale a dizer $r_{jk} = 0$.

$$R_{i,jk} = \sqrt{r_{ij}^2 + r_{ik}^2}$$

A vantagem de se definir o coeficiente de correlação dessa forma reside no fato dos elementos que compõe sua fórmula serem os coeficientes de correlação de ordem zero, os quais são facilmente calculáveis.

Uma outra fórmula utilizada para o cálculo do coeficiente de correlação linear múltipla é a quantidade sempre não-negativa ($0 \leq R \leq 1$)

$$R = \sqrt{\frac{\sum_{i=1}^k b_i S_{iy}}{S_{yy}}}$$

Essa expressão parte da existência de uma variável dependente Y e diversas variáveis independentes X_i .

Pode-se notar que R é uma generalização da expressão $r = \sqrt{\frac{b \cdot S_{xy}}{S_{yy}}}$ para duas variáveis $\left(r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \right)$ no caso de duas variáveis independentes X_1, X_2 .

O coeficiente de determinação, R^2 indica a parcela da variação total de Y, expressa por S_{yy} , que é explicada pelo hiperplano da regressão e tem o mesmo campo de variação do caso simples: $0 \leq R^2 \leq 1$.

Por analogia ao caso da reta a variância residual em torno da regressão é:

$$S^2_M = \frac{S_{yy} - \sum_{i=1}^k b_i \cdot S_{iy}}{n - k - 1}$$

Sendo que $(k + 1)$ parâmetros devem ser estimados.

Quanto mais próximo de 1 estiver $R_{i,jk}$, mais bem definida será a relação linear entre as variáveis; quanto mais próximo de zero, menos acentuada será essa relação. Esse coeficiente varia de 0 a 1 e no caso em que $R_{i,jk} = 1$ a correlação é denominada perfeita.

Exemplo 2.3: Calcular o coeficiente de correlação $R_{1,23}$ do exemplo do peso, altura e idade

$$r_{12} = 0,810 \Rightarrow r_{12}^2 = 0,656 \quad r_{13} = 0,704 \Rightarrow r_{13}^2 = 0,496 \quad r_{23} = 0,783 \Rightarrow r_{23}^2 = 0,613$$

$$R_{1,23} = \frac{\sqrt{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}}{\sqrt{1 - r_{23}^2}}$$

$$R_{1,23} = \frac{\sqrt{0,656 + 0,496 - 2 \cdot 0,810 \cdot 0,704 \cdot 0,783}}{\sqrt{1 - 0,613}} = \sqrt{\frac{0,260}{0,388}} = \sqrt{0,670} = 0,818$$

(correlação forte)