Sejam bem vindos

Estatística e Probabilidade

Núcleo Temático VI

Tutor Prof. Luiz Angelo Cardinali Pansanato

1

Objetivos:

Entender como a ferramenta correção de regressão linear auxilia na análise dos relacionamentos entre variáveis, e também permite um panorama para previsão e o planejamento.

2

A **Correlação Linear Simples** é uma técnica estatística que nos ajuda a entender a relação entre duas variáveis.

3

CORRELAÇÃO LINEAR SIMPLES

O coeficiente de correlação, frequentemente denotado por “r”, é calculado usando a seguinte fórmula, conhecida como fórmula de Pearson:

$$r = \frac{\sum(X_i - \overline{X}) * (Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{\mathrm{X}})^2 * \sum(Y_i - \overline{\mathrm{Y}})^2}}$$

4

Onde:

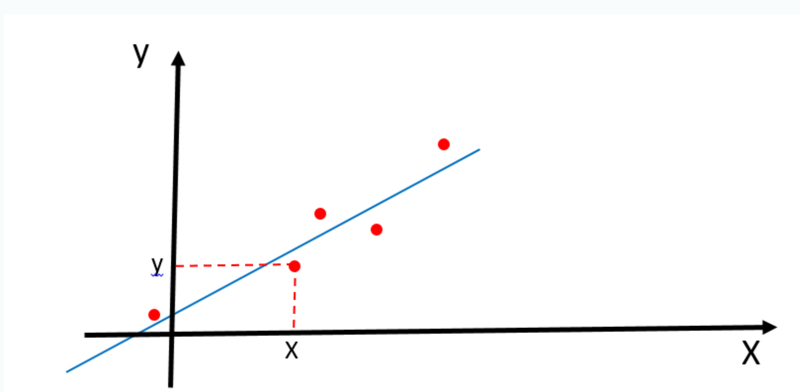
$$r = \frac{\sum(X_i - \overline{X}) * (Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 * \sum(Y_i - \overline{Y})^2}}$$

- X_i e Y_i são os valores das duas variáveis relacionadas
- $\overline{X}$ e $\overline{Y}$ são as médias das duas variáveis respectivamente
- $\sum$ representa a soma de todos os valores em um conjunto de dados

5

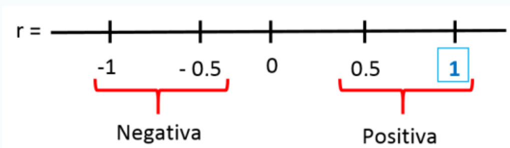
Coeficiente de Correção Linear

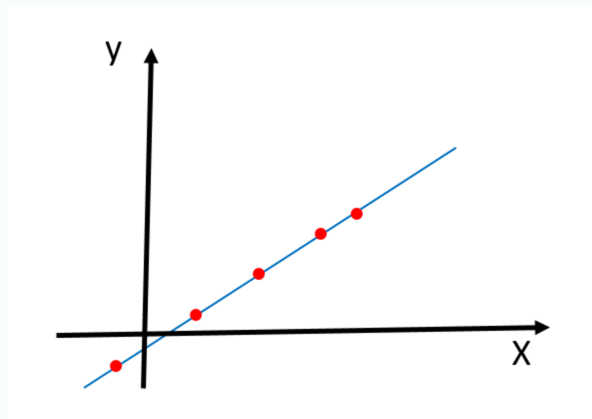
A cada x escolhido tem uma relação no Y



6

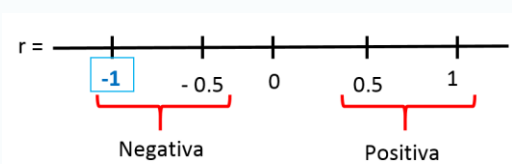
Quando o resultado da correlação linear for 1, ela será perfeita e positiva

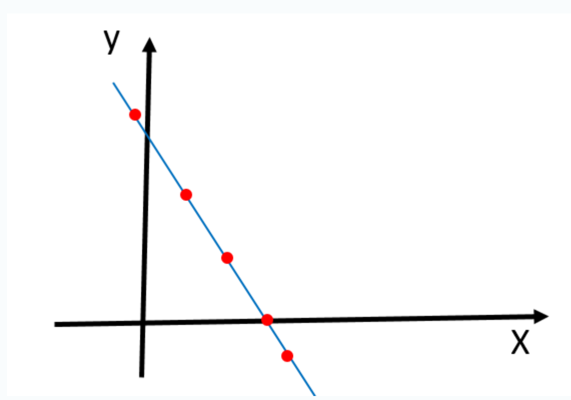




7

Quando o resultado da correlação linear for -1, ela será perfeita e negativa



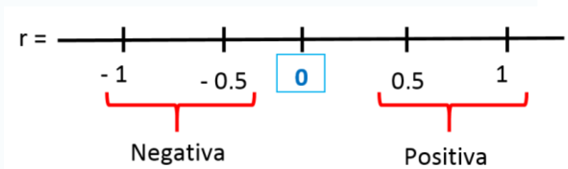


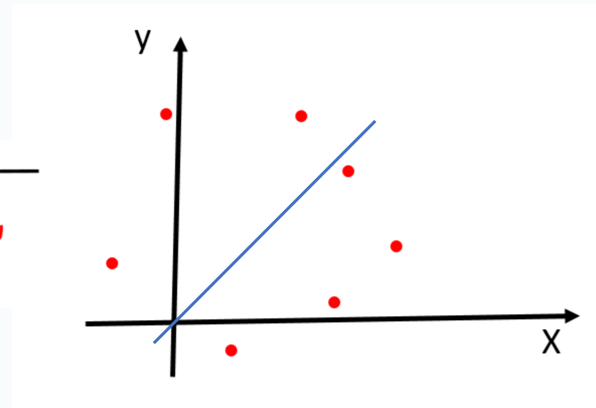
#vempraFaSouza

8

FaSouza

Quando o resultado da correlação linear for 0, neste caso não há uma relação entre X e Y, não há correlação (nula)



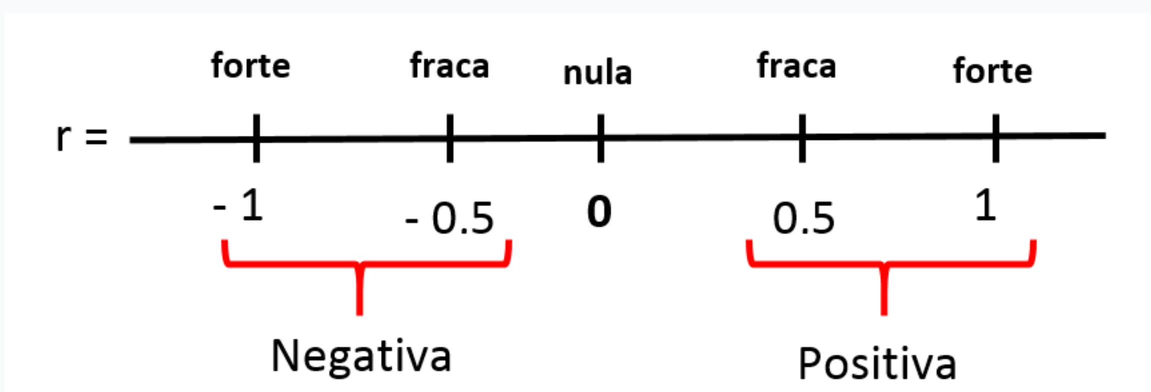


CREDENCIADA PELO MEC | EAD Ensino a distância | #vempraFaSouza

9

FaSouza

Então



CREDENCIADA PELO MEC | EAD Ensino a distância | #vempraFaSouza

10

FaSouza

Exemplo:

Uma determinada empresa quer analisar a correlação de duas variáveis

Treinamento de funcionários (X_i)	60	86	110	128	132	145
Produtividade (Y_i)	180	205	283	343	460	620

	X_i	Y_i
	60	180
	86	205
	110	283
	128	343
	132	460
	145	620
Σ	**661**	**2091**

Média de X_i = 661÷ 6 = 110,17

Média de Y_i = 2091÷ 6 = 348,50

CREDENCIADA PELO MEC EAD Ensino a distância #vempraFaSouza

11

FaSouza

$$r = \frac{\boxed{\sum(X_i - \overline{X})} * \boxed{(Y_i - \overline{Y})}}{\sqrt{\sum(X_i - \overline{X})^2 * \sum(Y_i - \overline{Y})^2}}$$

	X_i	Y_i	$X_i - \overline{X}$	$Y_i - \overline{Y}$	$(X_i - \overline{X}).(Y_i - \overline{Y})$
	60	180	12	72	864
	86	205	38	97	3686
	110	283	62	175	10850
	128	343	80	235	18800
	132	460	84	352	29568
	145	620	97	512	49664
Σ	**661**	**2091**	**373**	**1443**	**113432**

$$r = \frac{113432}{\sqrt{\sum(X_i - \mathrm{X})^2 * \sum(Y_i - \mathrm{Y})^2}}$$

CREDENCIADA PELO MEC EAD Ensino a distância #vempraFaSouza

12

FaSouza

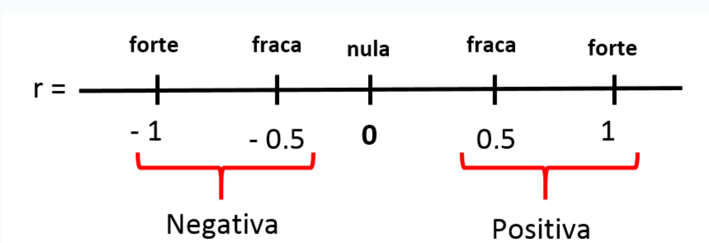
$$r = \frac{\sum(X_i - \overline{X}) * (Y_i - \overline{Y})}{\sqrt{\boxed{\sum(X_i - \overline{X})^2} * \boxed{\sum(Y_i - \overline{Y})^2}}}$$

	X_i	Y_i	$X_i - \overline{X}$	$Y_i - \overline{Y}$	$(X_i-\overline{X}).(Y_i-\overline{Y})$	$(X_i-\overline{X})^2$	$(Y_i-\overline{Y})^2$
	60	180	12	72	864	144	5184
	86	205	38	97	3686	1444	9409
	110	283	62	175	10850	3844	30625
	128	343	80	235	18800	6400	55225
	132	460	84	352	29568	7056	123904
	145	620	97	512	49664	9409	262144
Σ	**661**	**2091**	**373**	**1443**	**113432**	**28297**	**486491**

13

FaSouza

$$r = \frac{113432}{\sqrt{28297 * 486491}} = 0{,}97$$



R: Correlação positiva forte, ou seja, quanto mais tempo de treinamento maior será a produtividade

14

A **Regressão Linear Simples** leva essa análise um passo adiante, permitindo-nos prever uma variável (variável independente) com base em outra (variável dependente).

15

Exemplo:

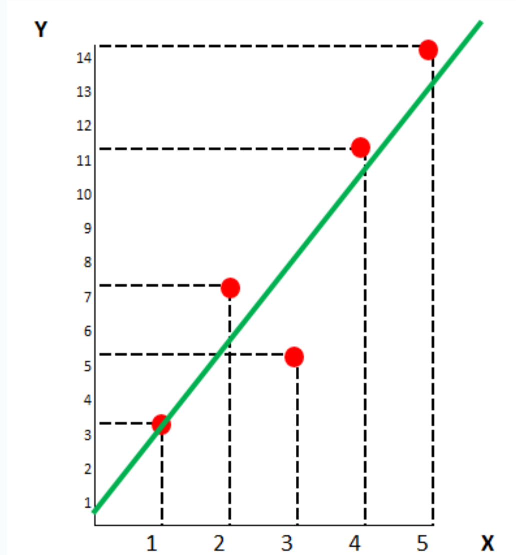
Suponha que você tem um conjunto de dados que relaciona o número de horas de treinamento (variável independente) com a produtividade de produção (variável dependente). Você deseja realizar uma regressão linear simples para entender melhor essa relação e prever a produtividade dos trabalhadores com base no tempo de treinamento.

16

FaSouza

Exemplo

	Horas / treinamento	Produtividade
	1	3
	2	7
	3	5
	4	11
	5	14
$\sum$	**15**	**40**
Média	**3**	**8**



17

FaSouza

Para encontrarmos a equação dessa reta na regressão linear utilizamos as seguintes fórmulas:

Estimativa → $$\hat{y} = b_0 + b_1 . X_i$$

Para calcularmos $$b_0 = \bar{Y}_i - b_1 . \bar{X}_i$$

Para calcularmos $$b_1 = \frac{\sum(X_i - \bar{X}) * (Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

18

	Horas / treinamento	Produtividade	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$X_i - \bar{X} * Yi - \bar{Y}$	$(X_i - \bar{X})2$
	1	3	-2	-5	10	4
	2	7	-1	-1	1	1
	3	5	0	-3	0	0
	4	11	1	3	3	1
	5	14	2	6	12	4
Σ	**15**	**40**			**26**	**10**
Média	**3**	**8**				

CREDENCIADA PELO MEC EAD Ensino a distância #vempraFaSouza

19

Exemplo

$$b_1 = \frac{\sum(X_i - \bar{X}) * (Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

$$b_1 = \frac{26}{10} = 2,6$$

$$b_0 = \bar{Y}_i - b_1 . \bar{X}_i$$

$$b_0 = 8 - 2,6.3 = 0,2$$

CREDENCIADA PELO MEC EAD Ensino a distância #vempraFaSouza

20

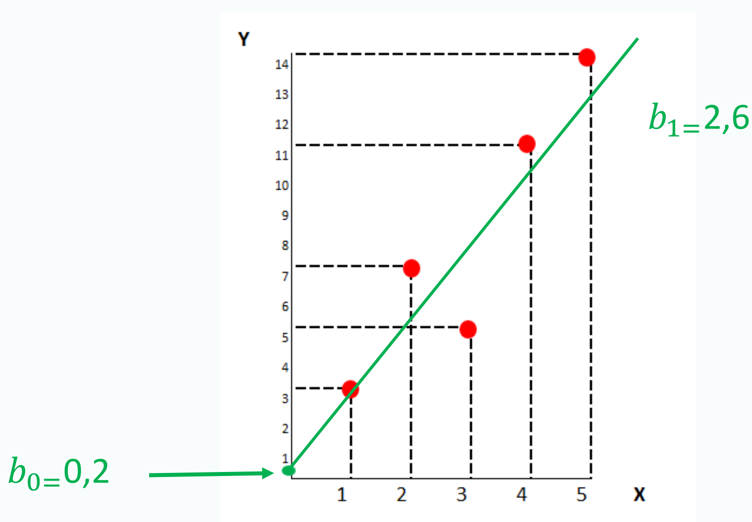
Então:

$$\hat{y} = b_0 + b_1 . X_i$$

Equação para estimativa

$$y\hat{} = 0{,}2 + 2{,}6 . X_i$$

21

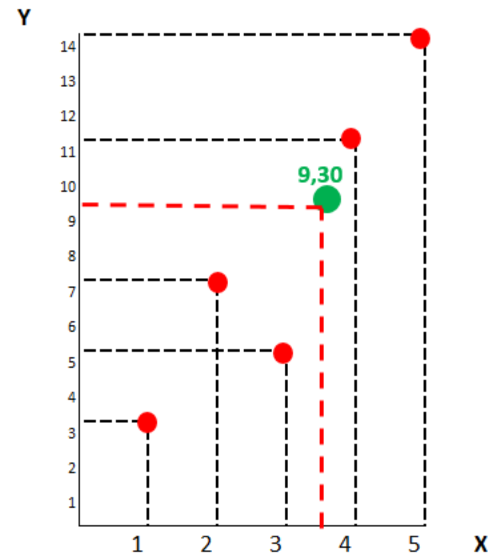


22

Utilizando a equação para estimativa:

$$\hat{y} = 0{,}2 + 2{,}6.3{,}5 = 9{,}30$$

Conclui-se que, para 3,5 horas de treinamento temos uma produtividade de 9,3.



#vempraFaSouza

23

- ✓ **Fórum tira dúvidas**

#vempraFaSouza

24