Medidas Estatísticas

Associação: Covariância, Correlação Linear Simples e Regressão linear simples

Discente: Murillo Almeida dos Santos Torres
Docente: José Cláudio Faria
Semestre: 2021/1

Universidade Estadual de Santa Cruz
CET 083 – Probabilidade e Estatística

1. Associação

Associação

Associação na estatística é o estudo de uma relação entre duas variáveis.

Este tipo de operação é útil para aplicações como:

- Possibilidade de estudar uma através da outra
- Tentar prever os valores de uma através da outra

As técnicas de associação são:

- Covariância
- correlação linear simples.

Conteúdo

1. Covariância
2. Correlação Linear Simples
3. Regressão Linear Simples

Conteúdo

Introdução

O termo Covariância no ramo da probabilidade e estatística está relacionado à medida da variabilidade conjunta entre duas variáveis aleatórias. Pelas regras da mesma se as variáveis apresentam um comportamento semelhante, ou seja, se a maioria dos maiores valores de uma variável correspondem com a maioria dos maiores valores da segunda, e o mesmo acontece para os menores valores, a covariância é positiva.
Caso aconteça o comportamento contrário, ou seja os maiores valores de uma variável corresponderem ao menores valores da segunda e vice-versa temos uma covariância negativa.

O sinal da covariância, portanto, vai mostrar a tendência na relação linear entre as variáveis.
A magnitude da covariância não é simples de se interpretar pois não é normalizada e, consequentemente, depende da magnitude das variáveis. Entretanto a versão normalizada da covariância, o coeficiente de correlação, mostra a força da relação linear pela sua magnitude.

Definição

A covariância entre duas variáveis reais aleatórias, X e Y, distribuídas em conjunto e com variância finita é definida como o produto esperado dos seus desvios a partir dos seus valores individuais esperados.

$$cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Onde E[X] é o valor esperado de X, também conhecido como média de X. Usando a propriedade de linearidade de expectativas podemos simplificar a fórmula para que a mesma seja o valor esperado do produto de X e Y menos o produto do valor esperado de X e de Y, ficando assim:

$$cov(X, Y) = E[XY] - E[X]\,E[Y]$$

Definição

Variáveis aleatórias as quais a covariância é zero são chamadas de variáveis não correlacionadas.

As unidades de medida de uma covariância cov(X, Y) são as de X vezes as de Y. Em contra-partida, coeficientes de correlação, os quais dependem da covariância são medidas adimensionais de associação linear.

Desdobramentos

Quando trabalhamos com **variáveis discretas**, ou seja variáveis que entre um valor e outro não existe valor intermediário (i.e., pontos), utilizamos a seguinte fórmula:

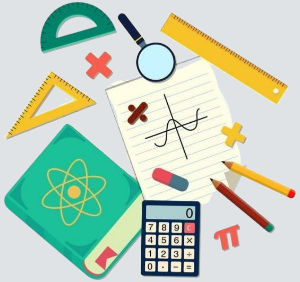
$$cov(X, Y) = \sum_X \sum_Y (X - E(X))(Y - E(Y))p(x,y)$$

A covariância padronizada, chama-se coeficiente de correlação entre X e Y, o qual denotaremos por p(x,y).

Quando porém as variáveis trabalhadas forem contínuas, ou seja, quando entre dois valores (X_1 e X_2) existirem infinitos valores intermediários (i.e., intervalos), não podemos mais utilizar a fórmula anterior pois agora precisamos de um método capaz de nos dar o resultado de toda uma região, de todo um intervalo, e para isso nós utilizamos integrais definidas como já visto em Cálculo 2:

$$cov(X,Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (X - E(X))(Y - E(Y)) f(x,y)\,dxdy$$

Exemplo



Y/X	0	1	2	P(y)
1	3/20	3/20	2/20	8/20
2	1/20	1/20	2/20	4/20
3	4/20	1/20	3/20	8/20
P(x)	8/20	5/20	7/20	1

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y)] = \sum (X - E(X))(Y - E(Y))p(X,Y)$$

$$E(X) = 0 \times \frac{8}{20} + 1 \times \frac{5}{20} + 2 \times \frac{7}{20} = \frac{19}{20}$$

$$E(Y) = 1 \times \frac{8}{20} + 2 \times \frac{4}{20} + 3 \times \frac{8}{20} = 2$$

$$Cov(X,Y) = (0 - \frac{19}{20}) \times (1 - 2) \times \frac{3}{20} + (1 - \frac{19}{20}) \times (1 - 2) \times \frac{3}{20} + (2 - \frac{19}{20}) \times (1 - 2) \times \frac{2}{20}$$

$$+ (0 - \frac{19}{20}) \times (2 - 2) \times \frac{1}{20} + (1 - \frac{19}{20}) \times (2 - 2) \times \frac{1}{20} + (2 - \frac{19}{20}) \times (2 - 2) \times \frac{2}{20}$$

$$+ (0 - \frac{19}{20}) \times (3 - 2) \times \frac{4}{20} + (1 - \frac{19}{20}) \times (3 - 2) \times \frac{1}{20} + (2 - \frac{19}{20}) \times (3 - 2) \times \frac{3}{20} = 0$$

Observação

- Se X e Y são variáveis aleatórias independentes a Cov(X,Y) será igual a 0.

- Porém, se a Cov(X,Y) for igual a 0, isso NÃO significa que as variáveis serão independentes.

Covariância

Covariância positiva

Obter uma covariância positiva significa, na prática, que as duas variáveis têm o mesmo comportamento, ou seja, quando uma delas aumenta, a segunda também aumentará e quando uma delas diminui a outra também se comportará da mesma forma.

Isto fará com que a maior parte das observações recaiam no 1º e 3º quadrante, demonstrando portanto um relacionamento positivo entre as variáveis.

Covariância negativa

Obter uma covariância negativa significa, na prática, que as duas variáveis têm o comportamento oposto, ou seja, quando uma delas aumenta, a segunda diminuirá e vice-versa. Isto fará com que a maior parte das observações recaiam no 2º e 4º quadrante, demonstrando portanto um relacionamento negativo entre as variáveis.

Propriedades

Para quaisquer variáveis aleatórias X, Y, Z e uma constante *c*, temos:

Cov(X,X) = Var(X)

Cov(X,Y) = Cov(Y,X)

Cov(*c*X,Y) = *c*Cov(X,Y)

Cov(X,Y + Z) = Cov(X,Y) + Cov(X,Z)

Covariância Amostral

- Na **AUSÊNCIA** da distribuição de probabilidade, podemos trabalhar com uma **AMOSTRA** da população. Assim:

$$cov(X,Y) = \sum i \frac{(Xi - \bar{X})(Yi - \bar{Y})}{n - 1}$$

Exemplo

- Consideremos duas variáveis aleatórias:

  - M : rendimento acadêmico em matemática
  - L : rendimento acadêmico em línguas

Rendimento acadêmico:

$\sum M = 480$ $\sum L = 400$

$\bar{M} = 60$ $\bar{L} = 50$

Obs:	01	02	03	04	05	06	07	08
M:	36	80	50	58	72	60	56	68
L:	35	65	60	39	48	44	48	61

Exemplo

- Calculo do índice $\sum M * L$

Obs	M	L	m = (M - E(M))	l = (L - E(L))	m.l
01	36	35	-24	-15	360
02	80	65	20	15	300
03	50	60	-10	10	-100
04	58	39	-2	-11	22
05	72	48	12	-2	-24
06	60	44	0	-6	0
07	56	48	-4	-2	8
08	68	61	8	11	88

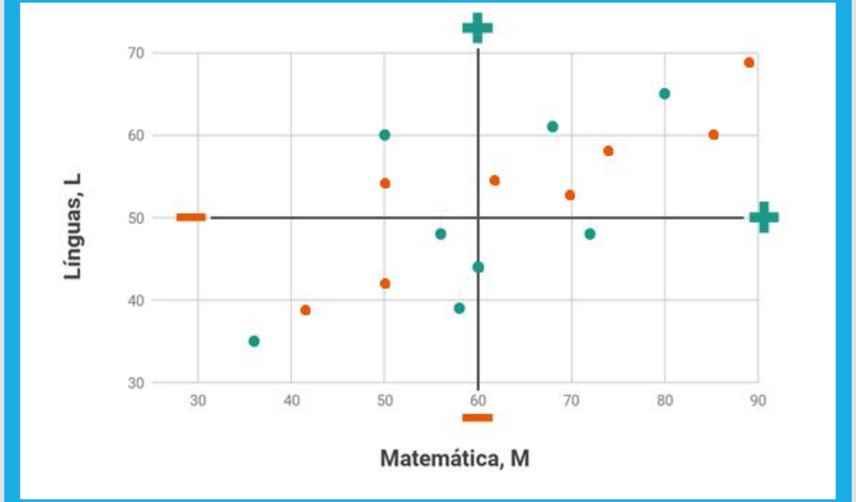
$$\sum M * L = 654$$

Covariância

Exemplo

$$Cov(M,L) = \frac{\sum_i (M_i - \bar{M})(L_i - \bar{L})}{n-1} = \frac{654}{7} = 93{,}43$$

Como possuem comportamentos semelhantes, ou seja, quando uma variável aumenta a outra também aumenta
e o mesmo acontece para quando uma diminui, a maior parte das observações recairão nos 1º e 3º quadrantes.
consequentemente, a maior parte dos produtos (m.I) serão positivos, bem como sua soma (Σml), demonstrando um relacionamento positivo entre M e L.



Conteúdo

Conteúdo

1. Covariância
2. Correlação Linear Simples
3. Regressão Linear Simples

Introdução

O termo correlação significa **relação** em dois sentidos (co + relação), e é usado em **estatística** para designar a força que mantém unidos dois conjuntos de valores. A verificação da existência e do grau de relação entre as variáveis é o objeto de estudo da correlação.

Frequentemente em pesquisas é procurado verificar se existe alguma relação entre duas ou mais variáveis, isto é, saber se as alterações sofridas por uma das variáveis são acompanhadas por alterações nas outras.

Por exemplo, peso **vs.** idade, consumo **vs.** renda, altura **vs.** peso, de um indivíduo.

Correlação Linear Simples

Fórmula

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}, \sigma_x \sigma_y > 0$$

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$\sigma_x \sigma_y = \sqrt{Var(x)\,Var(y)}$$

$$Var(y) = \sum (y - E(Y))^2 P(y) = [E(Y^2) - (E(Y))^2]$$

$$Var(x) = \sum (x - E(X))^2 P(x) = [E(X^2) - (E(X))^2]$$

Interpretando os resultados

O coeficiente de correlação r linear é um número puro que varia de **–1** a **+1** e sua interpretação dependerá do valor numérico e do sinal, como segue:

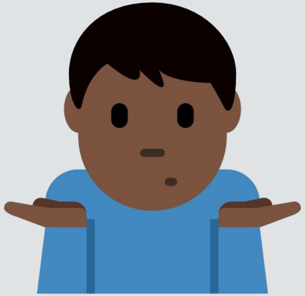
Coeficiente de Correlação	Correlação
$r = 1$	Perfeito positivo
$0,8 <= r < 1$	Forte positiva*
$0,5 <= r < 0,8$	Moderado positiva*
$0,1 <= r < 0,5$	Fraca positiva*
$0 <= r < 0,1$	Intima positiva*
$r = 0$	Nula
$-0,1 <= r < 0$	Intima negativa*
$-0,5 <= r < -0,1$	Fraca negativa*
$-0,8 <= r < -0,5$	Moderado negativa*
$-1 < r < -0,8$	Forte negativa*
$r = 1$	Perfeito negativo

Relação entre duas variáveis

Um coeficiente de correlação mede o grau pelo qual duas variáveis tendem a mudar juntas. O coeficiente descreve a força e a direção da relação. Uma medida do grau e do sinal da correlação é dada pela covariância entre as duas variáveis aleatórias **X** e **Y** que é uma medida numérica de associação existente entre elas. Para calcular o coeficiente de uma correlação entre duas variáveis, podemos recorrer a dois métodos já solidificados, sendo eles:

- **Pearson**
- **Spearman**

Pearson ou Spearman?



A correlação de **Pearson** avalia a relação linear entre duas variáveis contínuas. Uma relação é linear quando a mudança em uma variável é associada a uma mudança proporcional na outra variável. **Exemplo:** Você poderia usar uma correlação de **Pearson** para avaliar se aumentos na temperatura da instalação de produção estão associados a uma redução da espessura da cobertura de chocolate.

Já a correlação de **Spearman** avalia a relação monotônica entre duas variáveis contínuas ou ordinais. Em uma relação monotônica, as variáveis tendem a mudar juntas mas não necessariamente a uma taxa constante. A correlação de **Spearman** é muito usada para avaliar relações envolvendo variáveis ordinais. **Exemplo:** Você poderia usar a correlação de **Spearman** para avaliar se a ordem na qual os funcionários executam um teste está relacionada ao número de meses de emprego.

Calculando Pearson

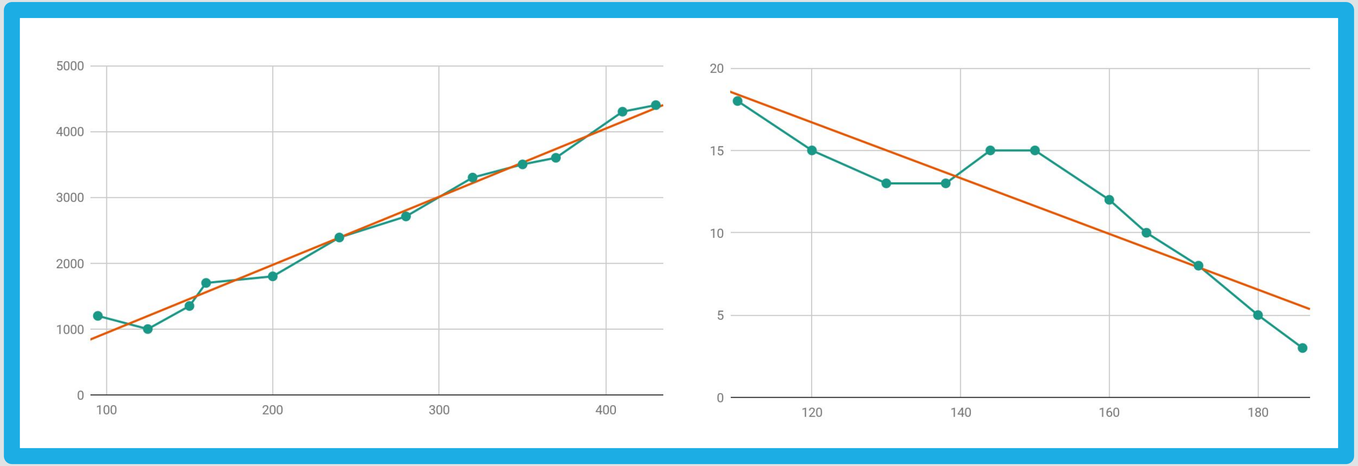
$$r = \frac{\sum (x - \bar{x}).(y - \bar{y})}{\sqrt{\left[\sum (x - \bar{x})^2\right] . \left[\sum (y - \bar{y})^2\right]}}$$

$$r = \frac{n . \sum x.y - (\sum x).(\sum y)}{\sqrt{n. \sum x^2 - (\sum x)^2} . \sqrt{n. \sum y^2 - (\sum y)^2}}$$

$$r = \frac{\sum x.y - \frac{(\sum x) . (\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] . \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

Diagrama de Dispersão

Os pares de valores das duas variáveis na correlação poderão ser colocados num diagrama cartesiano chamado "**diagrama de dispersão**". A vantagem de construir um diagrama de dispersão está em que, muitas vezes sua simples observação já nos dá uma idéia bastante boa de como as duas variáveis se relacionam.



Correlação **positiva** e **forte**
r = 0,984

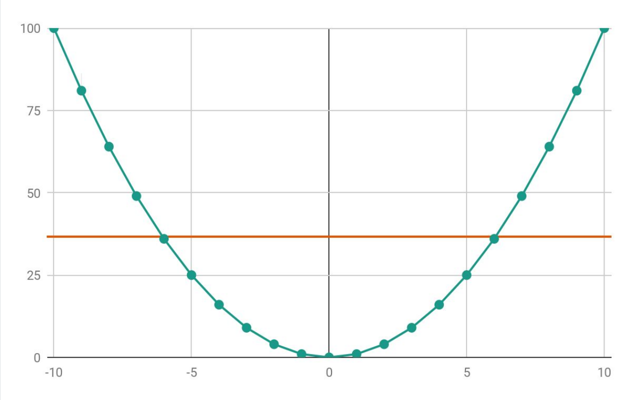
Correlação **negativa** e **forte**
r = -0,819

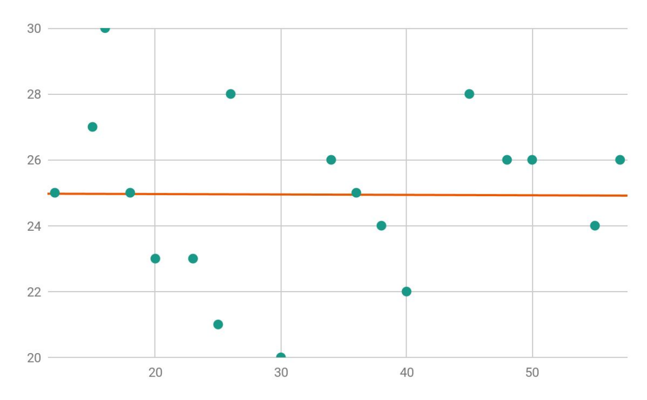
Correlação Linear Simples

Correlação fraca, quase nula r = -0,0068

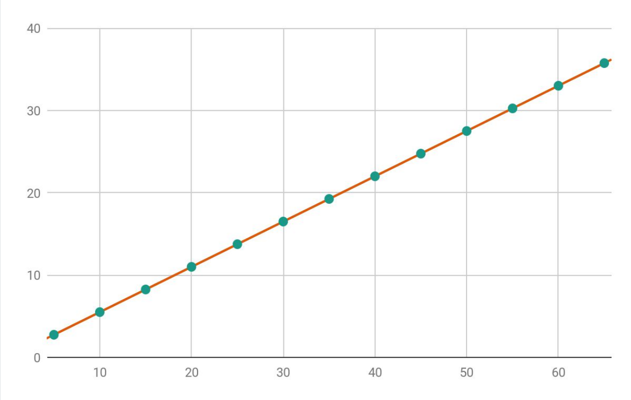
UESC

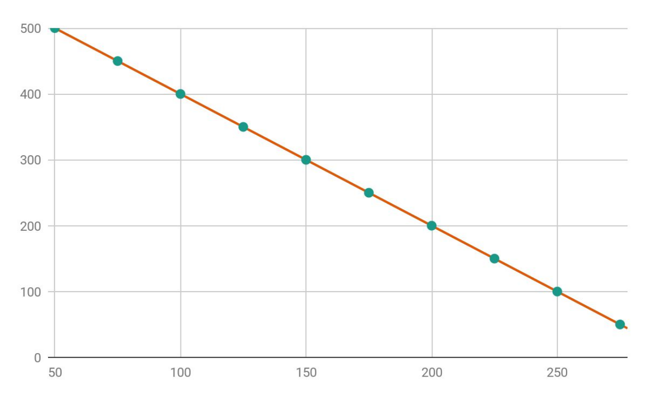
Correlação **nula** $r = 0$





Correlação **positiva** e perfeita $r = 1$





Correlação **negativa** e perfeita r = -1

CET 083 Probabilidade e Estatística

IN ALTUM

Vamos por em prática !

Região	Taxa de Mortalidade Infantil (X)	Taxa de Analfabetismo (Y)	X Y	X^2	Y^2
Norte	18,1	9,1	164,71	327,61	82,81
Nordeste	17,5	16,2	283,5	306,25	262,44
Centro-Oeste	14,8	5,7	84,36	219,04	32,49
Sudeste	10,7	4,3	46,01	114,49	18,49
Sul	9,7	4,1	39,77	94,09	16,81
Somatório	**70,8**	**39,4**	**618,35**	**1061,48**	**413,04**
(Somatório)²				**1126740**	**170602**

$$r = \frac{\sum x.y - \frac{(\sum x)\,.\,(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\,.\,\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

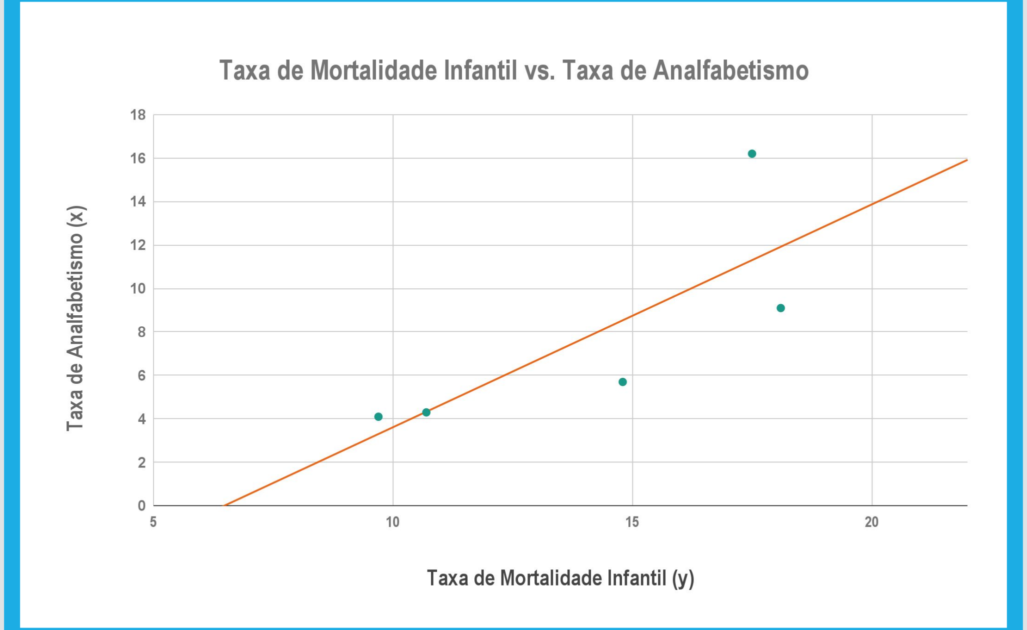
$$r = \frac{618{,}35 - \frac{70{,}8\ .\ 39{,}4}{5}}{\sqrt{\left[1061{,}48 - \frac{1126740}{5}\right]\,.\,\left[413{,}04 - \frac{170602}{5}\right]}}$$

$$r = \frac{60.446}{\sqrt{[-224.286{,}52]\ .\ [-33.707{,}36]}}$$

$$r = 0{,}69$$

Correlação Linear Simples

Gráfico de Dispersão



Coeficiente de Correlação	Correlação
0,5 <= r < 0,8	Moderado positiva

$r = 0{,}69$

Conteúdo

1. Covariância
2. Correlação Linear Simples
3. Regressão Linear Simples

Conteúdo

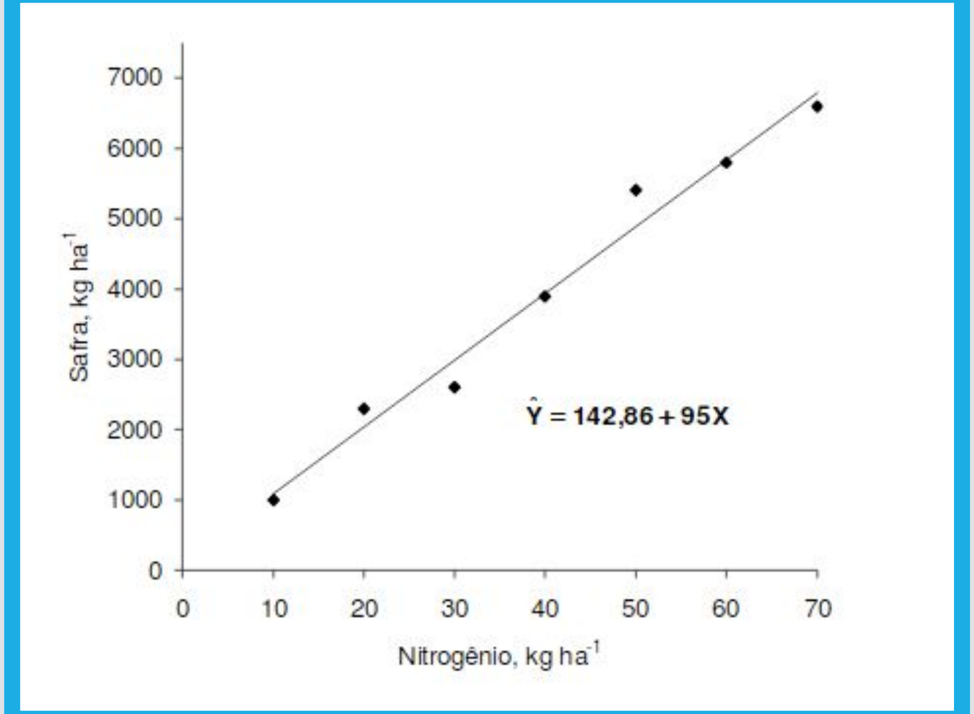
Introdução

- A Regressão Linear Simples, ou RLM, é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis quantitativa de forma que uma variável pode ser predita a partir de outra ou outras.

- O modelo de regressão é um dos métodos estatísticos mais usados para investigar a relação entre variáveis.

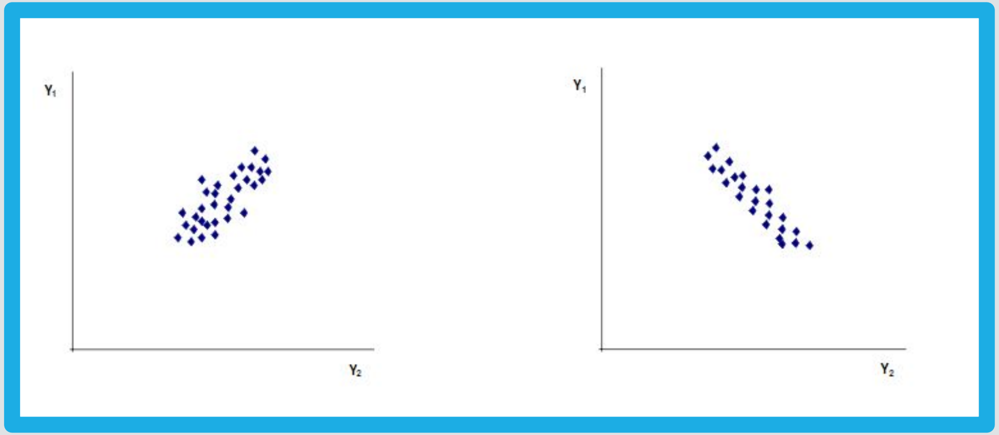
Aplicação

Figura 14.2 – Exemplo ilustrativo de regressão linear simples. A safra do milho em função de doses crescentes de adubo nitrogenado aplicado em cobertura.



Aplicação

A análise de correlação é indicada para estudar o grau de associação linear entre variáveis aleatórias. Ou seja, essa técnica é empregada, especificamente, para se avaliar o grau de covariação entre duas variáveis aleatórias: se uma variável aleatória **Y1** aumenta, o que acontece com uma outra variável aleatória **Y2**: aumenta, diminui ou não altera?



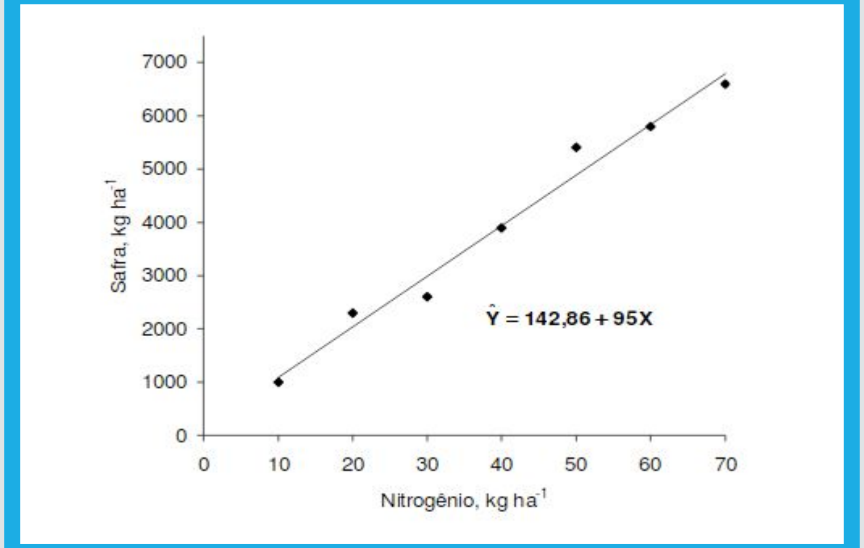
Informações

- Na análise de regressão uma resposta unilateral é esperada: alterações em X (fator quantitativo) podem implicar em alterações em Y, mas alterações em Y não resultam em alterações em X.
- Enquanto a análise de regressão linear nos mostra como as variáveis se relacionam linearmente, a análise de correlação vai nos mostrar apenas o grau desse mesmo relacionamento.
- Na análise de regressão estimamos toda uma função Y = f(X)
- Quando se deseja verificar a existência de alguma relação estatística entre uma ou mais variáveis fixas, independentes, sobre uma variável aleatória, denominada dependente, utiliza-se a análise de regressão.

Exemplo

- Para exemplificar, vamos considerar que conduzimos um experimento submetendo plantas de milho a doses crescentes de nitrogênio.

- Naturalmente, a produção será dependente da quantidade aplicada desse fertilizante, X.



- Assim, o fertilizante nitrogenado aplicado é a variável independente, e cada uma das quantidades aplicadas são seus níveis, xi(10 ... 70 kg ha-^{1}).

- Cada variável aleatória mensurada na cultura do milho, sujeita a influência dos níveis xi da variável independente, ou seja, das doses de nitrogênio, é chamada "variável dependente" ou "fator resposta".

Exemplo

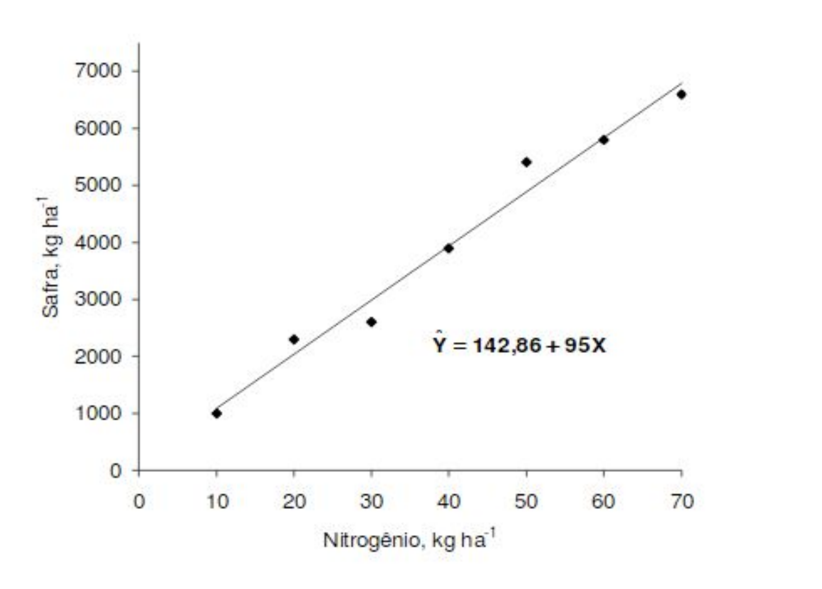
- Poderia-se medir, por exemplo, o número de espigas por planta (Y1), a altura média das plantas (Y2), o peso de 1.000 grãos (Y3), o teor de proteínas dos grãos (Y4), o teor de gordura dos grãos (Y5), etc.

- Como a aplicação do fertilizante não depende da safra, sendo, ao contrário, determinada independentemente pelo pesquisador, designamo-la “variável independente” ou “regressor”.

- Podemos estudar via análise de regressão o efeito da variável, neste caso, fixa, independente, X (dose de nitrogênio), sobre as variáveis aleatórias, ou dependentes, Yi (produção de matéria seca, teor de proteínas dos grãos, teor de gordura dos grãos, etc.). Diz-se regressão de Y sobre X.

Exemplo

Se grafarmos a safra, Y, decorrente das diversas aplicações, X, de nitrogênio, poderemos observar uma dispersão análoga a Figura 14.3



A aplicação de nitrogênio afeta a safra.

Podemos, por meio de uma equação, relacionando X e Y, descrever como afeta.

Estimar uma equação é geometricamente equivalente a ajustar uma curva àqueles dados dispersos, isto é, a "regressão de Y sobre X".

Esta equação será útil como descrição breve e precisa de predizer a safra Y para qualquer quantidade X de nitrogênio.

Regressão Linear Simples

Exemplo

Vamos considerar um estudo sobre a influência do N (nitrogênio) aplicado em cobertura sobre a safra do milho.

Suponhamos que só dispomos de recursos para fazer sete observações experimentais.

O pesquisador fixa então sete valores de X (sete níveis do regressor), fazendo apenas uma observação Y (fator resposta), em cada caso, tal como se vê na Figura 14.4

X Nitrogênio kg ha^{-1}	Y Safra kg ha^{-1}
10	1.000
20	2.300
30	2.600
40	3.900
50	5.400
60	5.800
70	6.600

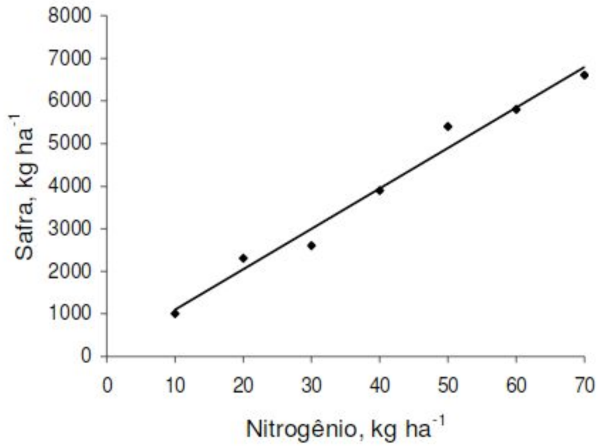
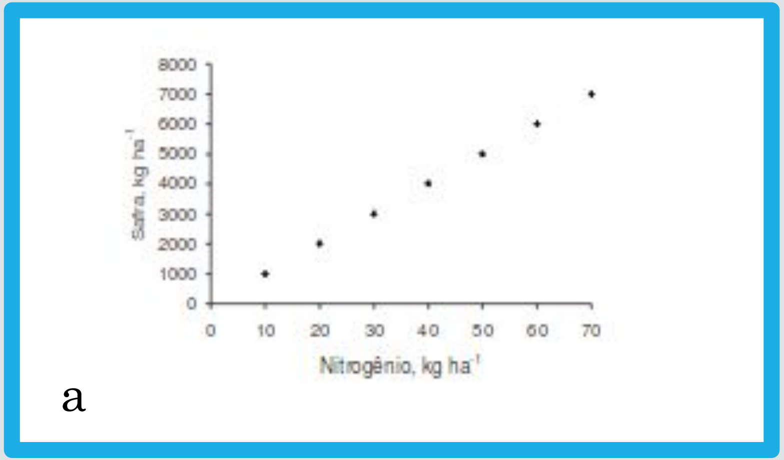
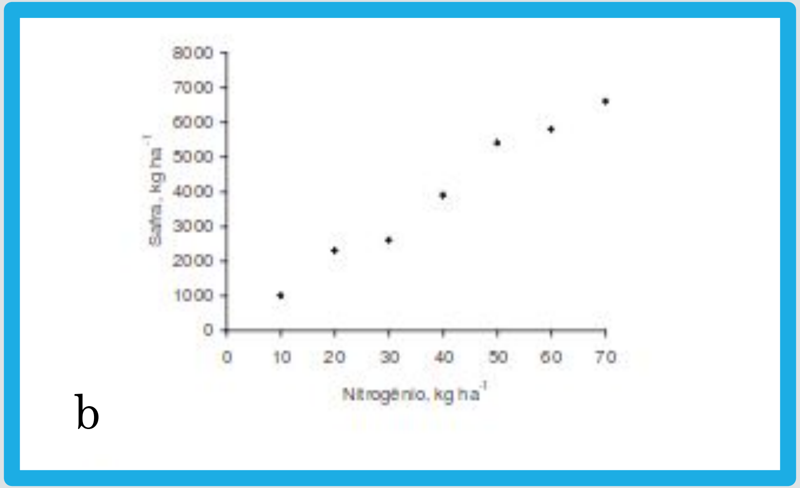


Figura 14.4 - Dados e reta ajustada a olho aos dados apresentados.

Regressão Linear Simples





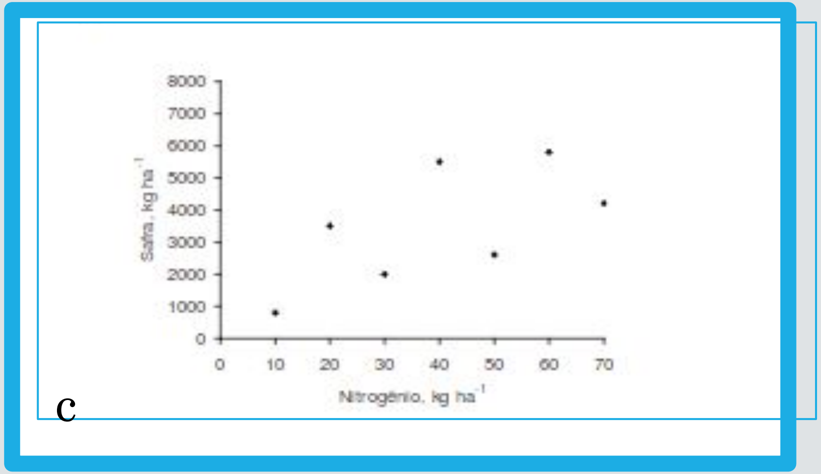


Figura 14.5 - Ilustração de diversos graus de dispersão.

Erro ou a falta de ajustamento é definido como a distância vertical entre o valor observado Yi e o valor ajustado $\hat{Y}_i$ na reta, isto é, $(Y_i - \hat{Y}_i)$:

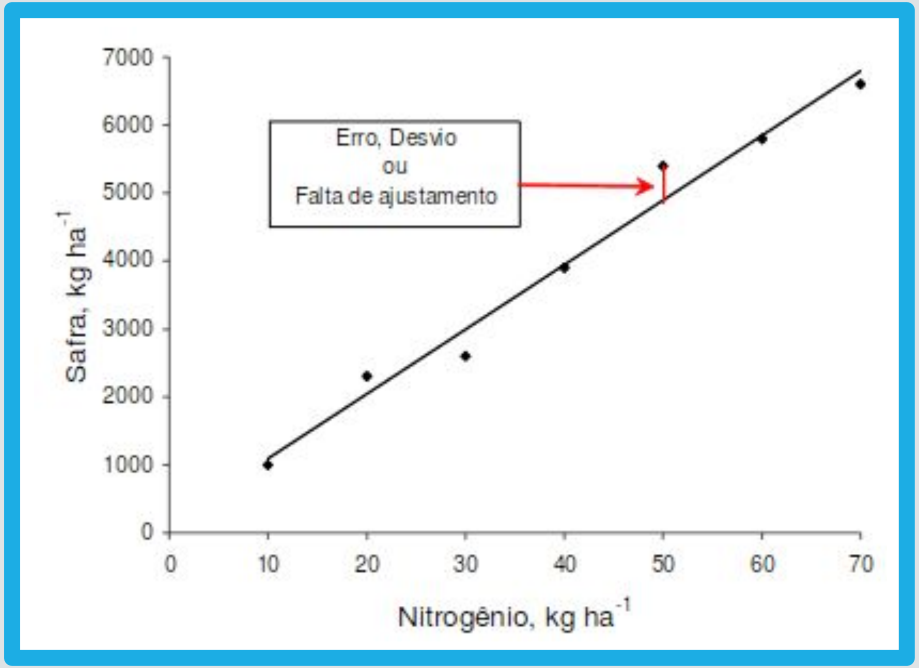


Figura 14.6 - Erro típico no ajustamento de uma reta.

O método mais comumente utilizado para se ajustar uma reta aos pontos dispersos é o que minimiza a soma de quadrados dos erros:

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Conhecido como critério dos "mínimos quadrados" ou "mínimos quadrados dos erros".
Sua justificativa inclui as seguintes observações:

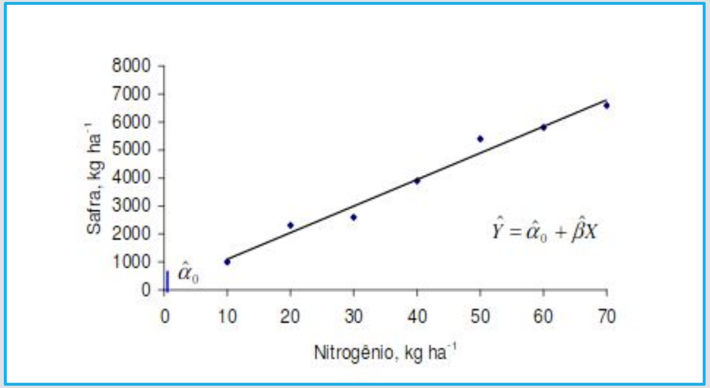
- O quadrado elimina o problema do sinal, pois torna positivos todos os erros.
- A álgebra dos mínimos quadrados é de manejo relativamente fácil.

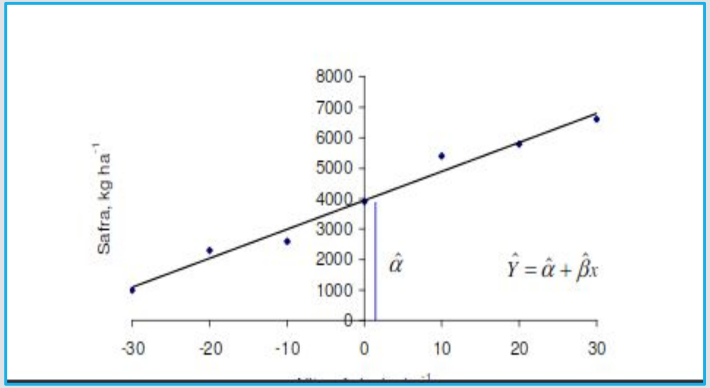
Regressão Linear Simples

Ajustando uma reta:

- Estágio 1: Exprimir X em termos de desvios a contar de sua média, isto é, definir uma nova variável x (minúsculo), tal que:

$$x = X - \bar{X}$$





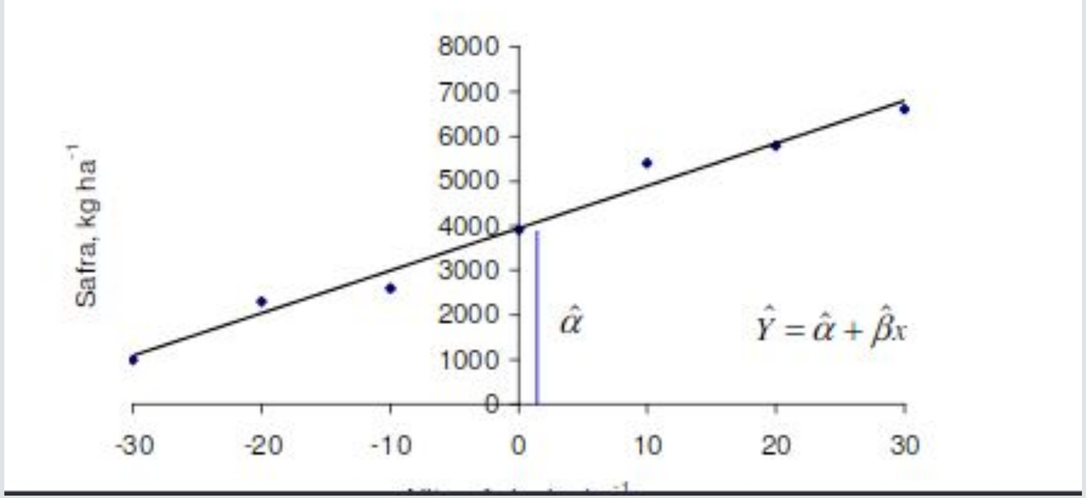
Observa-se que o eixo Y foi deslocado para a direita, de 0 a X.
O novo valor x torna-se positivo, ou negativo, conforme X esteja a direita ou a esquerda de X. Não há modificação nos valores de Y. O intercepto α difere do intercepto original, α0, mas o coeficiente angular, permanece o mesmo.

Regressão Linear Simples

- Medir X como desvio a contar de X simplifica os cálculos porque a soma dos novos valores x é igual a zero, isto é:

$$\sum x_i = 0 \qquad \therefore \qquad \sum x_i = \sum (X_i - \bar{X}) = \sum X_i - n\bar{X} = n\bar{X} - n\bar{X} = 0$$

Estágio 2: Ajustar a reta:



Regressão Linear Simples

Devemos ajustar a reta aos dados, escolhendo valores para α e β, que satisfaçam o critério dos mínimos quadrados. Ou seja, escolher valores de α e β que minimizem

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \qquad \text{Equação 01}$$

Cada valor ajustado $\hat{Y}_i$ estará sobre a reta estimada:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i \qquad \text{Equação 02}$$

Assim, estamos diante da seguinte situação: devemos encontrar os valores α e β de modo a minimizar a soma de quadrados dos erros.
Considerando as Equações 01 e 02, isto pode ser expresso algebricamente como:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i \qquad \text{Equação 02}$$

$$S(\hat{\alpha},\hat{\beta}) = \sum\left(Y_i - (\hat{\alpha} + \hat{\beta}x_i)\right)^2 = \sum\left(Y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2$$

A técnica de minimização mais simples é fornecida pelo cálculo. A minimização de S(α, β) exige a anulação simultânea de suas derivadas parciais:

Igualando a zero a derivada parcial em relação a α:

$$\frac{\partial}{\partial\hat{\alpha}}\sum\left(Y_i-\hat{\alpha}-\hat{\beta}x_i\right)^2=\sum 2(-1)\left(Y_i-\hat{\alpha}-\hat{\beta}x_i\right)^1=0$$

Dividindo ambos os termos por (-2) e reagrupando:

$$\sum Y_i-n\hat{\alpha}-\hat{\beta}\sum x_i=0 \qquad \therefore \qquad \sum x_i=0$$

$$\sum Y_i-n\hat{\alpha}-0=0$$

$$\sum Y_i-n\hat{\alpha}=0$$

$$n\hat{\alpha}=\sum Y_i$$

$$\hat{\alpha}=\frac{\sum Y_i}{n}=\bar{Y}$$

Assim, a estimativa de mínimos quadrados para â é simplesmente o valor médio de Y.

É preciso também anular a derivada parcial em relação a β:

$$\frac{\partial}{\partial\hat{\beta}}\sum\left(Y_i-\hat{\alpha}-\hat{\beta}x_i\right)^2=\sum 2(-x_i)\left(Y_i-\hat{\alpha}-\hat{\beta}x_i\right)^1=0$$

Dividindo ambos os termos por (-2):

$$\sum x_i\left(Y_i-\hat{\alpha}-\hat{\beta}x_i\right)=0$$

Reagrupando:

$$\sum x_iY_i-\hat{\alpha}\sum x_i-\hat{\beta}\sum x_i^2=0 \qquad \therefore \qquad \sum x_i=0$$

$$\sum x_iY_i-0-\hat{\beta}\sum x_i^2=0$$

$$\sum x_iY_i-\hat{\beta}\sum x_i^2=0$$

$$\hat{\beta}\sum x_i^2=\sum x_iY_i$$

$$\hat{\beta}=\frac{\sum x_iY_i}{\sum x_i^2}$$

Regressão Linear Simples

Para os dados da Figura 14.4 (Dados e reta ajustada a olho aos dados apresentados da Safra em função do Nitrogênio), α e β acham-se calculados no Quadro 14.1.

Quadro 14.1 - Cálculos dos valores necessários

X	$x = X - \bar{X}$ $x = X - 40$	Y	xY	x^2
10	- 30	1.000	- 30.000	900
20	- 20	2.300	- 46.000	400
30	- 10	2.600	- 26.000	100
40	0	3.900	0	0
50	10	5.400	54.000	100
60	20	5.800	116.000	400
70	30	6.600	198.000	900
$\sum X = 280$ $\bar{X} = \frac{1}{N}\sum X$ $\bar{X} = \frac{280}{7} = 40$	$\sum x = 0$	$\sum Y = 27.600$ $\bar{Y} = \frac{1}{N}\sum \bar{Y}$ $\bar{Y} = \frac{27.600}{7}$ $\bar{Y} = 3.942,86$	$\sum xY = 266.000$	$\sum x^2 = 2.800$

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \bar{Y} \quad \therefore \quad \hat{\alpha} = \frac{27.600}{7} = 3.942,86$$

$$\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2} \quad \therefore \quad \hat{\beta} = \frac{266.000}{2.800} = 95,00$$

$$\hat{Y} = 3.942,86 + 95x \qquad \text{Equação 03}$$

Regressão Linear Simples

Estágio 3: A regressão pode agora ser transformada para o sistema original de referência:

$$\hat{Y} = 3.942{,}86 + 95x \quad \therefore \quad x = \left(X - \bar{X}\right)$$

$$\hat{Y} = 3.942{,}86 + 95(X - \bar{X})$$

$$\hat{Y} = 3.942{,}86 + 95(X - 40)$$

$$\hat{Y} = 3.942{,}86 + 95X - 3.800$$

$$\hat{Y} = 142{,}86 + 95X \qquad \text{Equação 04}$$

$$\hat{Y} = 3.942{,}86 + 95x \qquad \text{Equação 03}$$

Comparando as Equações 03 e 04, observa-se que:

- O coeficiente angular da reta de regressão ajustada (β = 95X) permanece inalterado.
- A única diferença é o intercepto, α, onde a reta tangencia o eixo Y.
- O intercepto original foi facilmente obtido.

Regressão Linear Simples

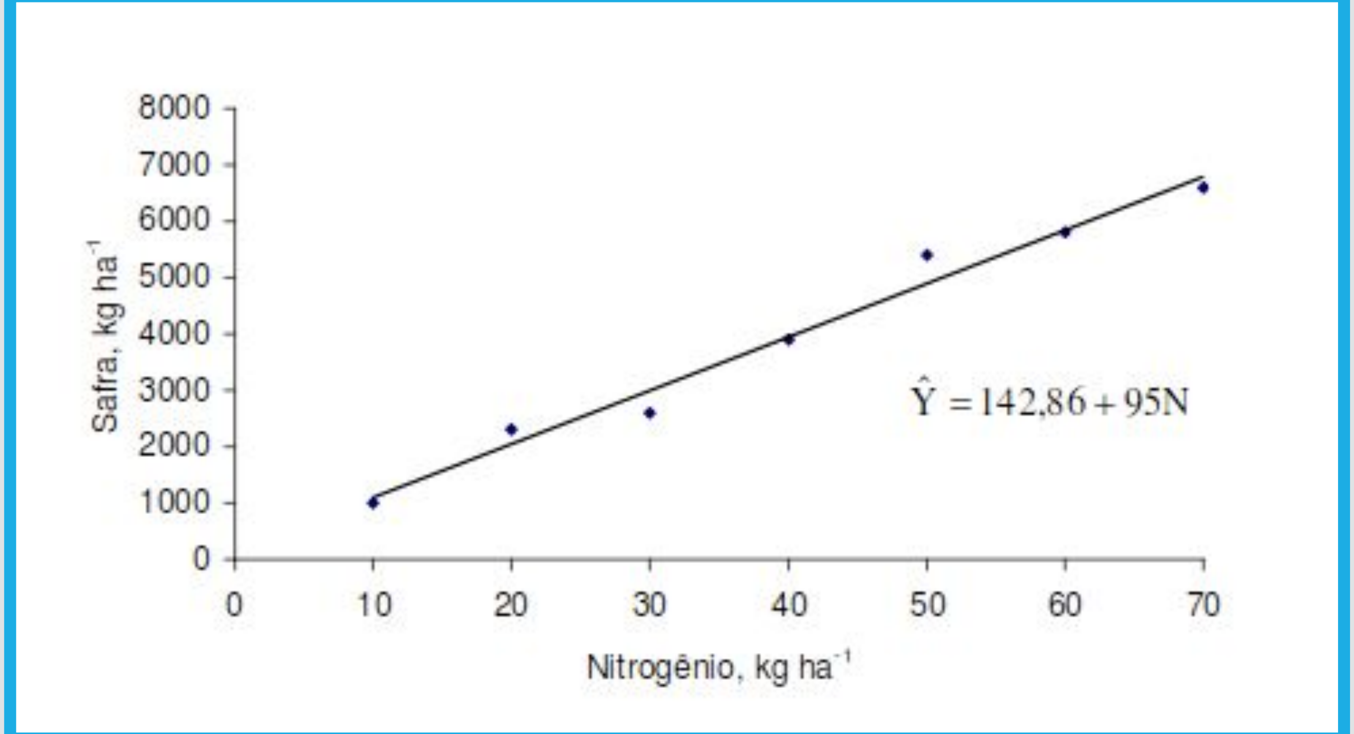


Figura 14.8 - Gráfico dos pontos dispersos com a reta ajustada.

Esta equação é útil como descrição breve e precisa de predizer a safra, em kg ha^{-1} para qualquer quantidade de nitrogênio, também em kg ha^{-1}, aplicada.

Observar que:

- Se nenhum nitrogênio for aplicado à cultura, a safra estimada será de 142,86 kg.
- Esta safra se deve a absorção pela cultura do N disponível no solo, possivelmente associado ao ciclo orgânico.
- No intervalo das doses aplicadas (10 a 70 kg), considerando-se um hectare,para cada kg de nitrogênio aplicado, a cultura responde com 95 kg de grãos.

Referências

FARIA, José Cláudio. Notas de aulas expandidas – Ilhéus, UESC/DCET, 10 ed. 2009.

LARSON, Ron; FARBER, Betsy. Estatística Aplicada 4ª Edição – São Paulo.