

Correlação e Regressão

Comte Marcos Santos, Prof. Dr.

BRASIL

AGENDA



www.exercito.gov.br

Introdução;

Exemplos;

Conceito de Correlação e Regressão;

Diagrama de Dispersão;

Coeficiente de Correlação Linear;

Propriedades do Coeficiente de Correlação Linear;

Exercícios;

Regressão;

Interpolação e Extrapolação;

Exercícios.

Comte Marcos Santos, Prof. Dr.

INTRODUÇÃO



www.exercito.gov.br

Quando se trabalha com apenas uma variável, aprendemos a calcular as medidas de tendência central e variabilidade.

Quando porém, consideramos observações de duas ou mais variáveis, surge um novo problema: as relações que podem existir entre as variáveis estudadas.



Período de 20 anos entre Consumo de Cigarros e Câncer de Pulmão

Cigarros
Fumados
Por Pessoa
Por Ano

4000

3000

2000

1000

1900

1920

1940

1960

1980

Ano

Consumo de
Cigarros
(homens)

Câncer de
Pulmão
(homens)

Mortes por
Câncer de
Pulmão
(Por
100.000
Pessoas)

150

100

50

www.exercito.gov.br

Comte Marcos Santos, Prof. Dr.

OUTROS EXEMPLOS



www.exercito.gov.br

- Peso – altura;
- Tabagismo – incidência de câncer;
- Renda – analfabetismo;
- Câmbio – inflação;
- Obesidade – infarto;
- Entre outras.

Exemplo: Vendas e gasto com publicidade de uma loja de confecções



www.exercito.gov.br

X: Gasto com publicidade(em R\$ mil)

Y: Venda (em R\$ mil)

Pares de observação ($X_i; Y_i$)

Gasto	Venda
3	7
4	14
8	15
12	28
14	32



Exemplo: nota da prova e tempo de estudo



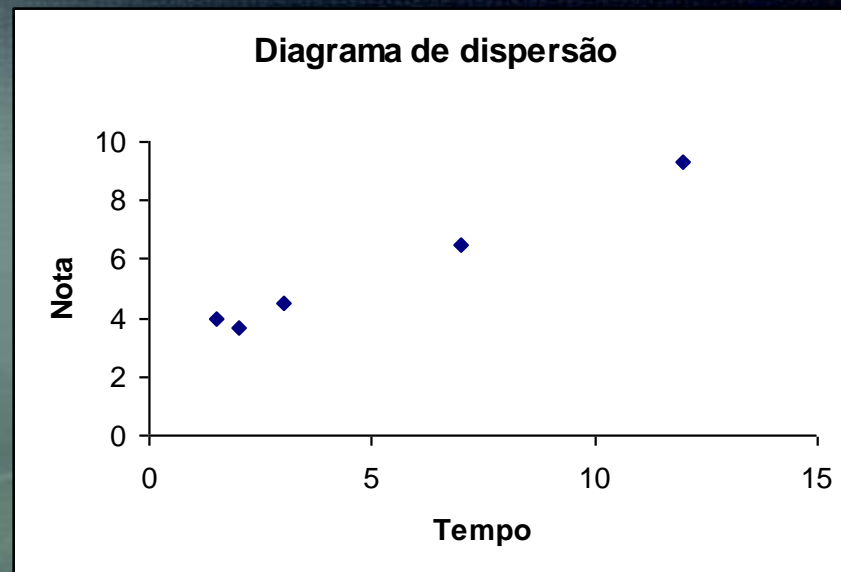
www.exercito.gov.br

X: tempo de estudo (em horas)

Y: Nota da prova

Pares de observação ($X_i; Y_i$)

Tempo	Nota
3,0	4,5
7,0	6,5
2,0	3,7
1,5	4,0
12,0	9,3



Cuidado! É preciso avaliar o contexto dessas correlações. Ex: nº de picolés vendidos e nº de afogamentos.

CORRELAÇÃO E REGRESSÃO



www.exercito.gov.br

Sendo a relação entre as variáveis de natureza quantitativa, a correlação é o instrumento adequado para descobrir e medir essa relação.

Uma vez caracterizada a relação, procuramos descrevê-la por meio de uma função matemática. A regressão é o instrumento adequado para a determinação dos parâmetros dessa função.

DIAGRAMA DE DISPERSÃO



www.exercito.gov.br

Consideremos uma amostra aleatória, formada por 10 dos 98 alunos de uma classe da faculdade A e pelas notas obtidas por eles em Matemática e Estatística.

N ^{os}	NOTAS	
	MATEMÁTICA (x_i)	ESTATÍSTICA (y_i)
01	5,0	6,0
08	8,0	9,0
24	7,0	8,0
38	10,0	10,0
44	6,0	5,0
58	7,0	7,0
59	9,0	8,0
72	3,0	4,0
80	8,0	6,0
92	2,0	2,0

DIAGRAMA DE DISPERSÃO



www.exercito.gov.br

Representando em um sistema de coordenadas ortogonais os pares ordenados (x, y) , obtemos uma nuvem de pontos que denominamos diagrama de dispersão. Esse diagrama nos fornece uma ideia grosseira, porém útil da correlação existente.

N ^{os}	NOTAS	
	MATEMÁTICA (x_i)	ESTATÍSTICA (y_i)
01	5,0	6,0
08	8,0	9,0
24	7,0	8,0
38	10,0	10,0
44	6,0	5,0
58	7,0	7,0
59	9,0	8,0
72	3,0	4,0
80	8,0	6,0
92	2,0	2,0

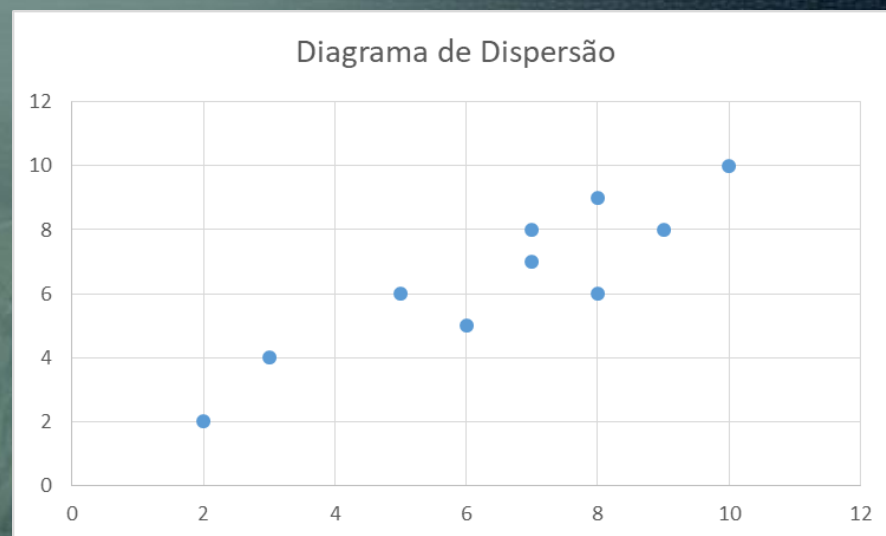
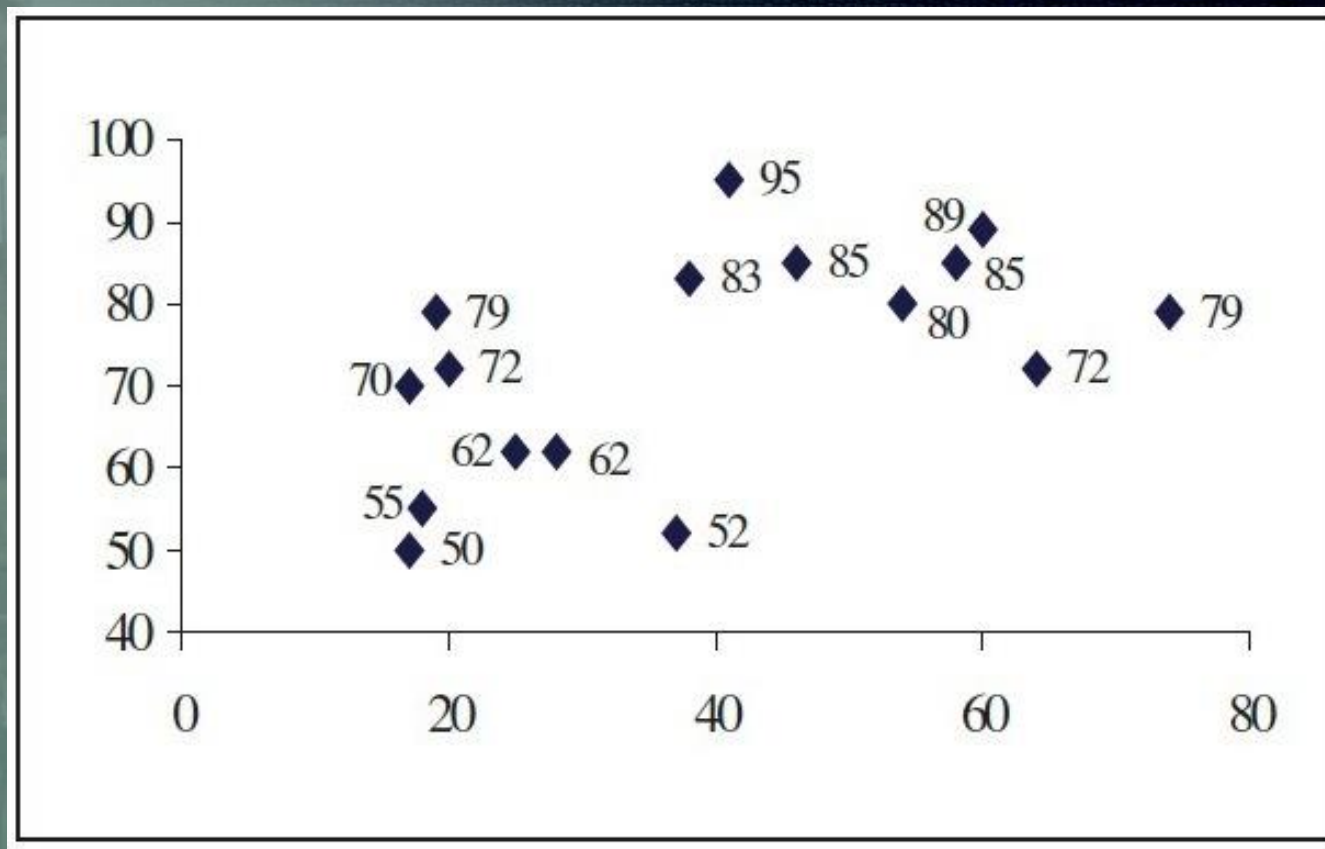


DIAGRAMA DE DISPERSÃO



www.exercito.gov.br



Comte Marcos Santos, Prof. Dr.

DIAGRAMA DE DISPERSÃO



www.exercito.gov.br

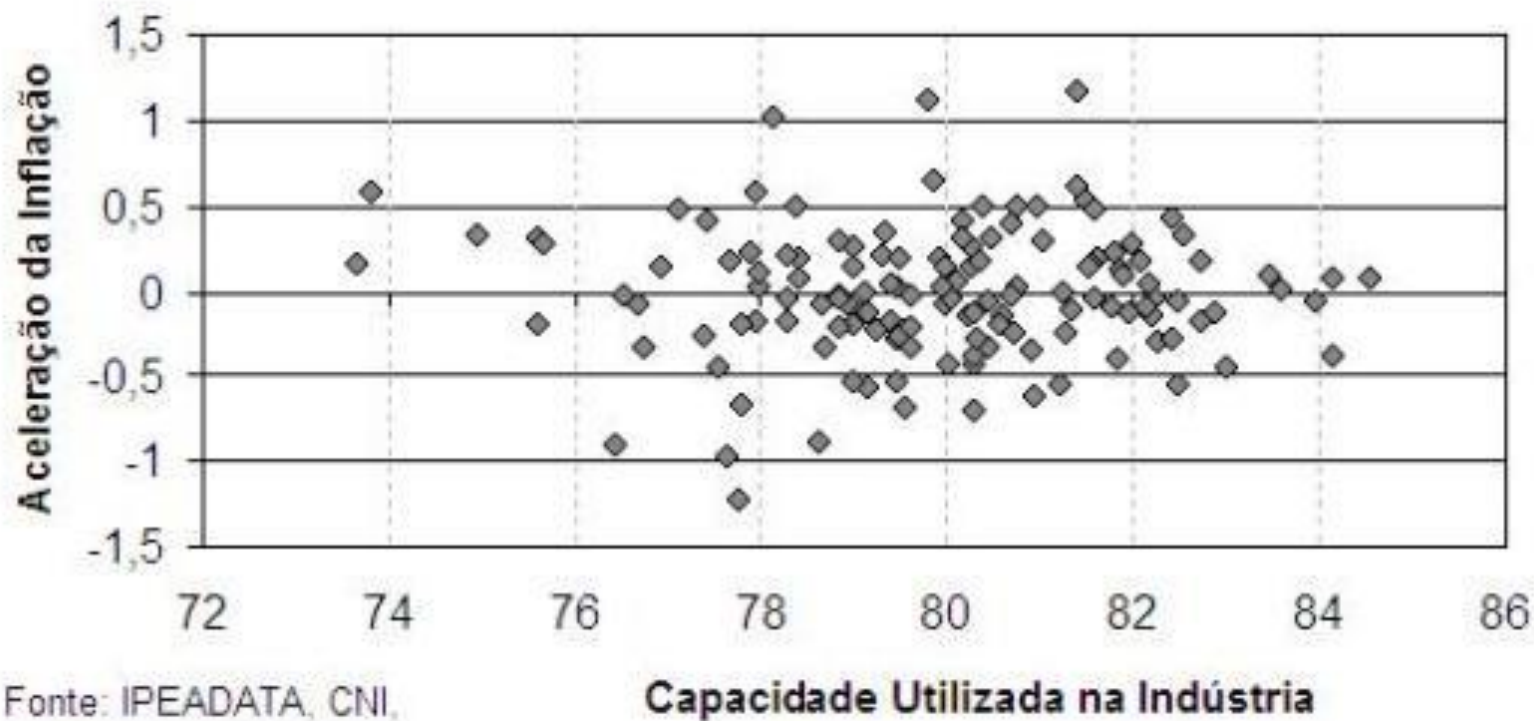
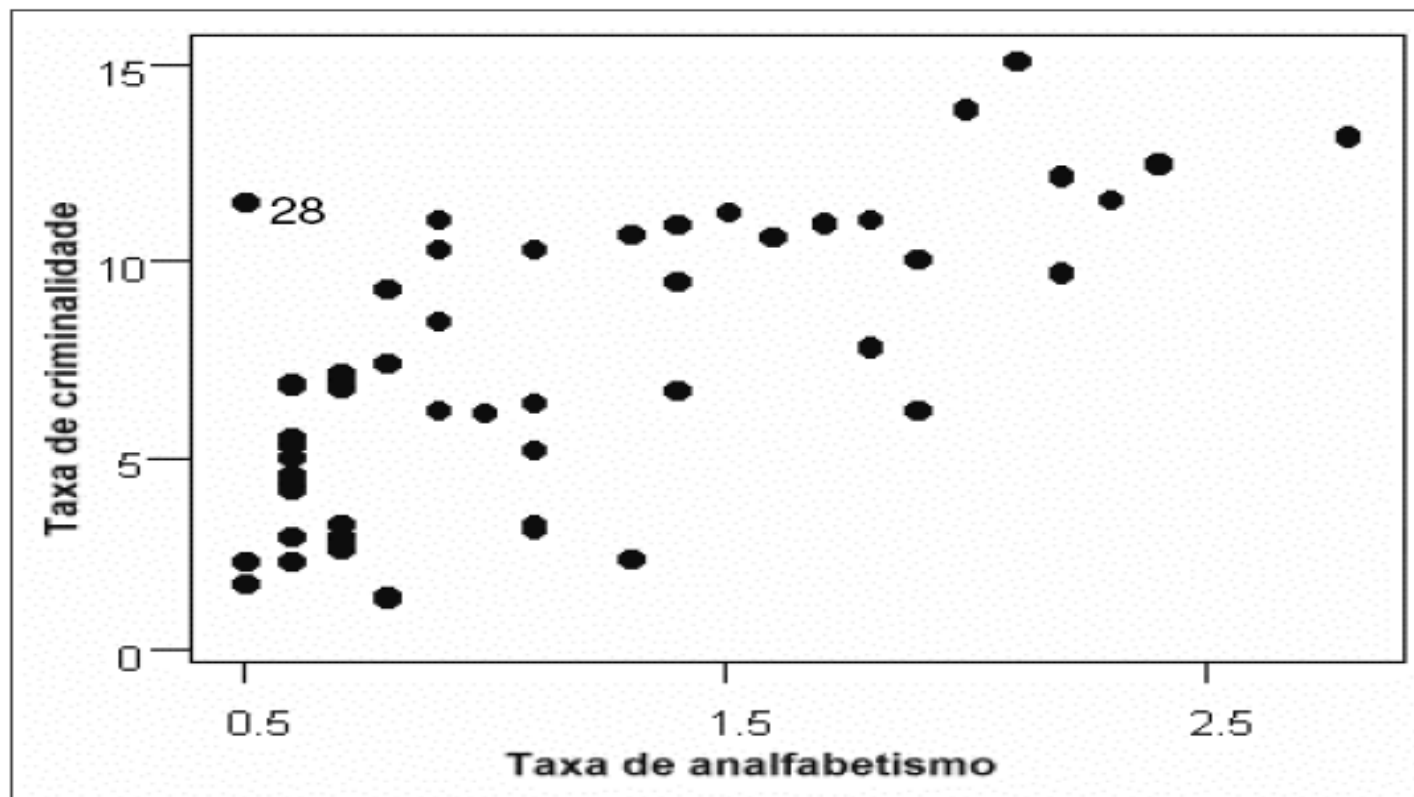


DIAGRAMA DE DISPERSÃO



www.exercito.gov.br

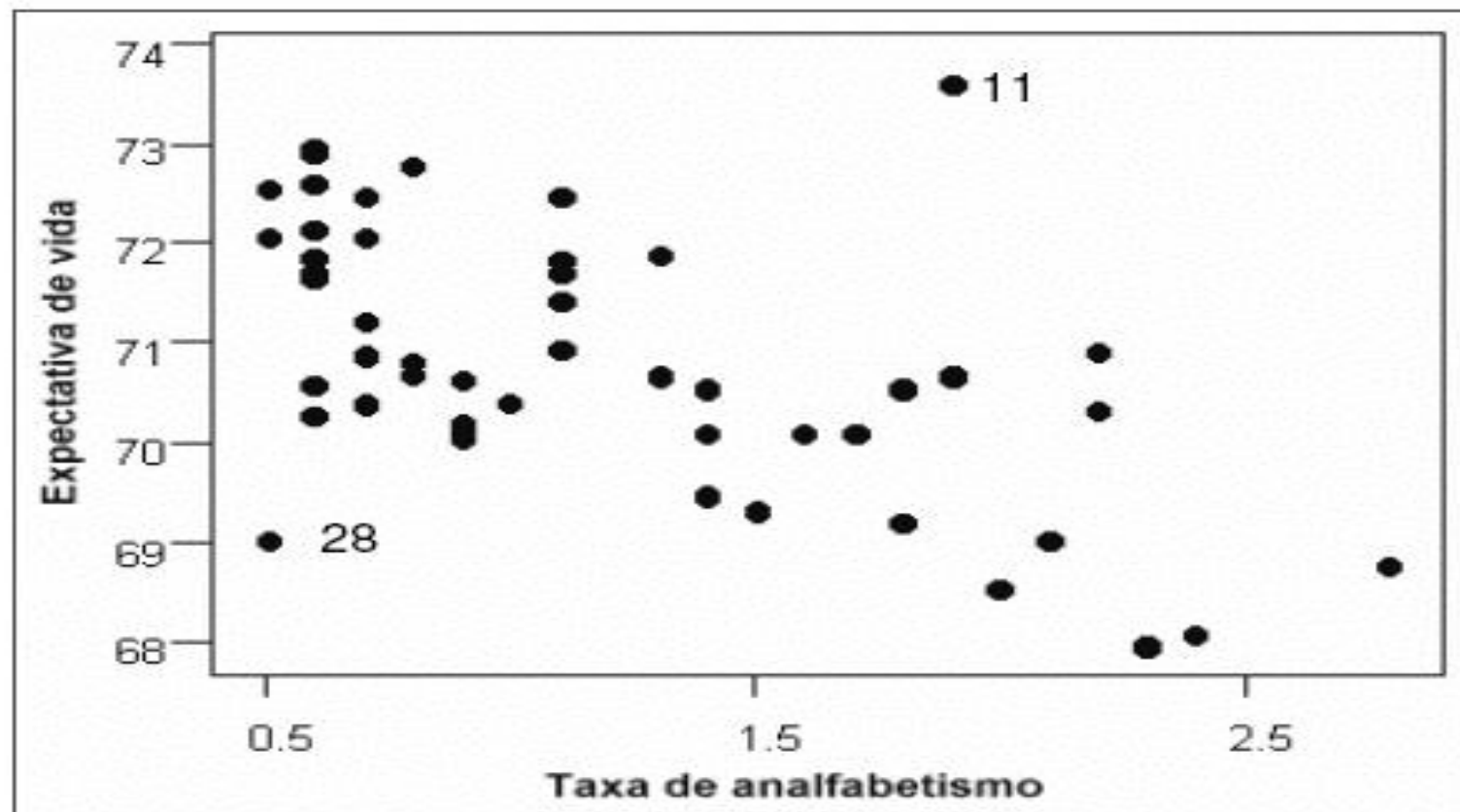


Comte Marcos Santos, Prof. Dr.

DIAGRAMA DE DISPERSÃO



www.exercito.gov.br



COEFICIENTE DE CORRELAÇÃO DE PEARSON



www.exercito.gov.br



É uma medida que avalia o quanto a “nuvem de pontos” no diagrama de dispersão aproxima-se de uma reta.

Comte Marcos Santos, Prof. Dr.

COEFICIENTE DE CORRELAÇÃO DE PEARSON



www.exercito.gov.br

O instrumento empregado para a medida da correlação linear é o coeficiente de correlação. Esse coeficiente deve indicar o grau de intensidade da correlação entre duas variáveis.

$$r = \frac{n \sum x_i y_i - (\sum x_i) (\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2] [n \sum y_i^2 - (\sum y_i)^2]}}$$

PROPRIEDADES DO COEFICIENTE DE CORRELAÇÃO DE PEARSON



www.exercito.gov.br

$$-1 \leq r \leq 1$$

Classificação da correlação:

- a) $r = 1$, correlação linear positiva e perfeita;
- b) $r = -1$, correlação linear negativa e perfeita; e
- c) $r = 0$, inexistência da correlação linear.



www.exercito.gov.br

PROPRIEDADES DO COEFICIENTE DE CORRELAÇÃO DE PEARSON

Para que uma relação possa ser descrita por meio do coeficiente de correlação de Pearson é imprescindível que ela se aproxime de uma função linear.

Para podermos tirar algumas conclusões significativas sobre o comportamento simultâneo das variáveis analisadas, é necessário que:

$$0,6 < |r| < 1$$

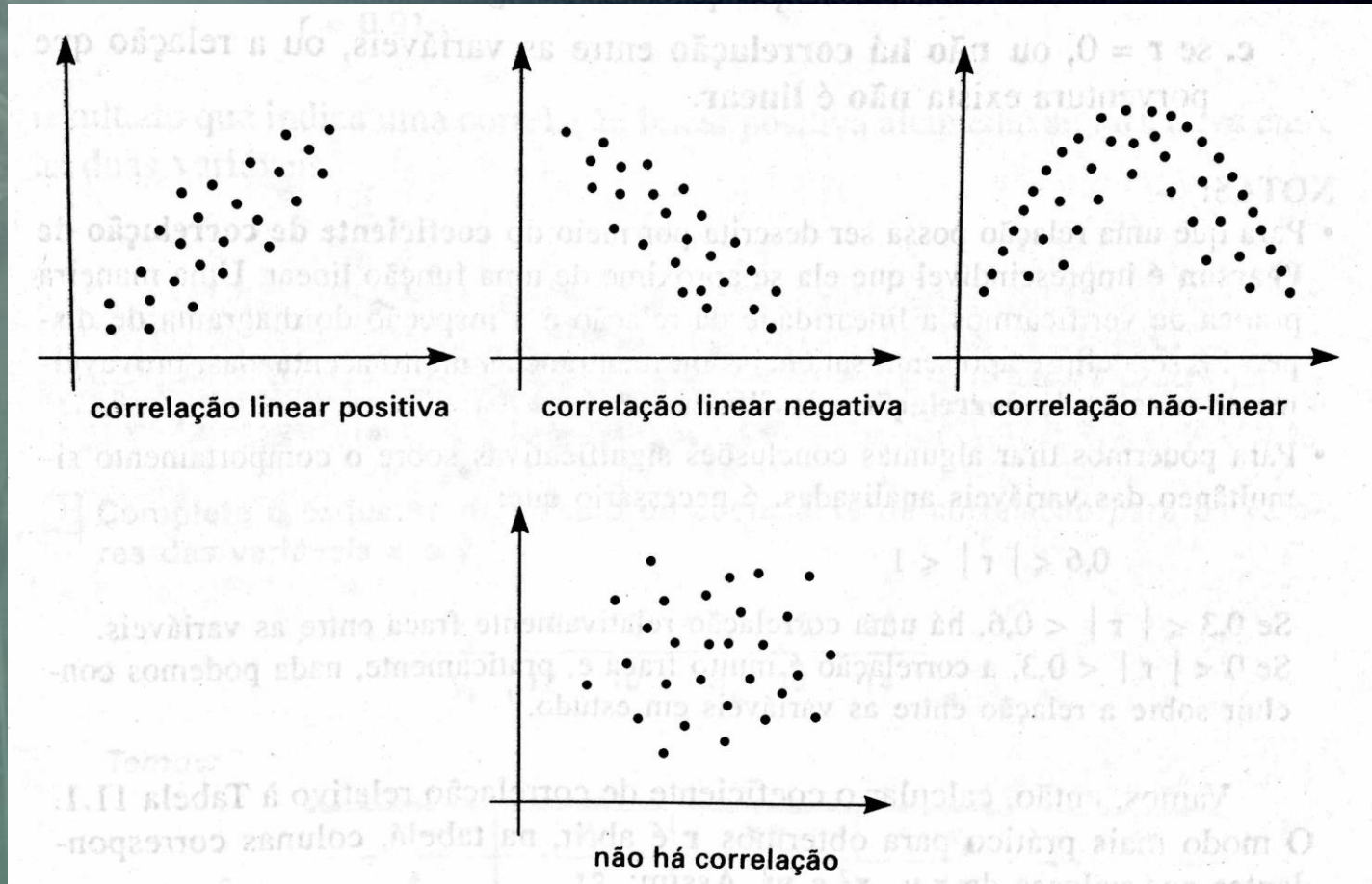
Se $0,3 < |r| < 0,6 \rightarrow$ há uma correlação relativamente fraca entre as variáveis.

Se $0 < |r| < 0,3 \rightarrow$ a correlação é muito fraca e, praticamente nada podemos concluir sobre a relação entre as variáveis em estudo.

PROPRIEDADES DO COEFICIENTE DE CORRELAÇÃO DE PEARSON



www.exercito.gov.br

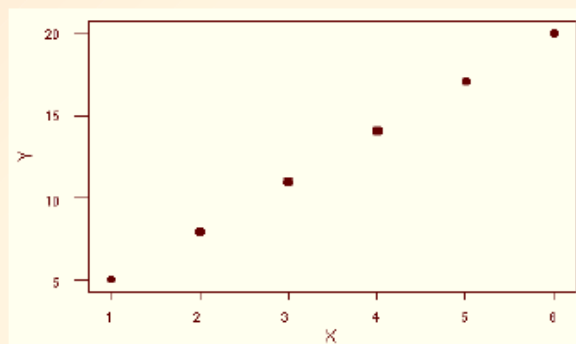


PROPRIEDADES DO COEFICIENTE DE CORRELAÇÃO DE PEARSON

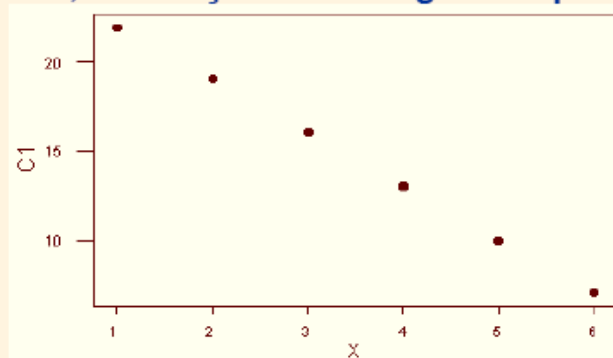


www.exercito.gov.br

$r = 1$, correlação linear positiva e perfeita



$r = -1$, correlação linear negativa e perfeita



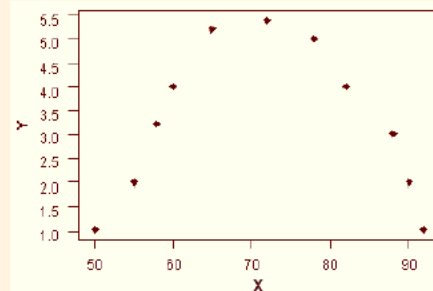
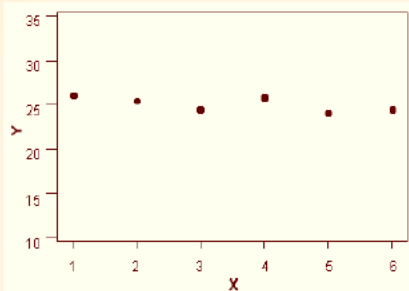
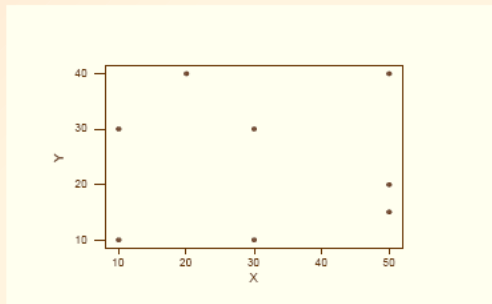
Comte Marcos Santos, Prof. Dr.

PROPRIEDADES DO COEFICIENTE DE CORRELAÇÃO DE PEARSON



www.exercito.gov.br

$$r = 0$$

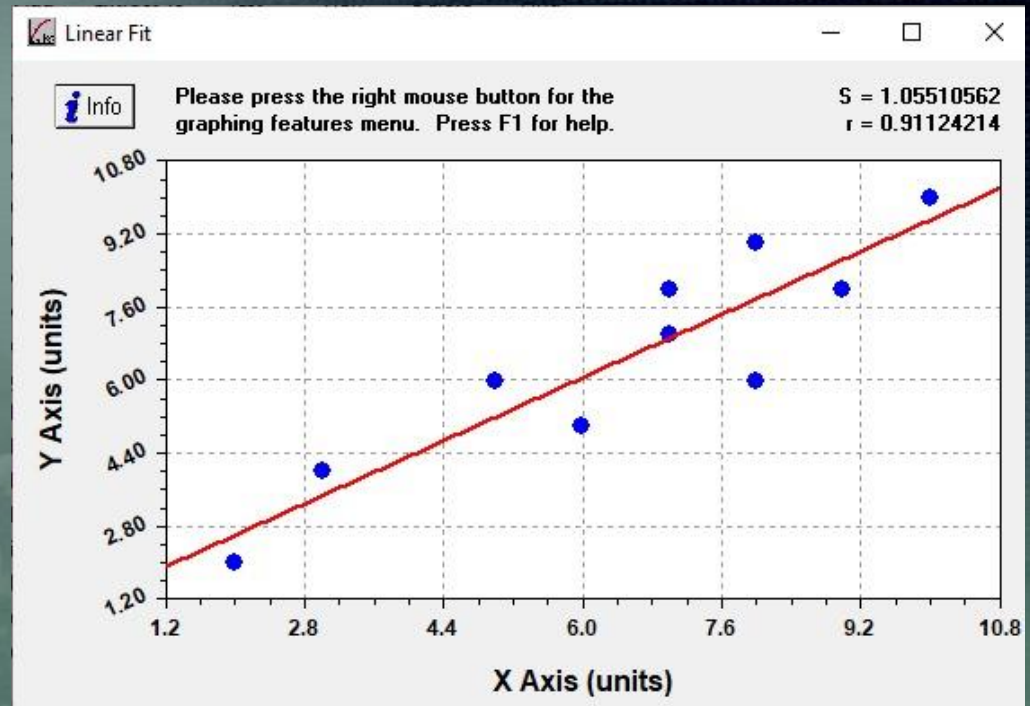


Comte Marcos Santos, Prof. Dr.

EXERCÍCIO

Vamos então calcular o coeficiente de correlação relativo à tabela abaixo:

MATEMÁTICA (x_i)	ESTATÍSTICA (y_i)
5	6
8	9
7	8
10	10
6	5
7	7
9	8
3	4
8	6
2	2
$\Sigma = 65$	$\Sigma = 65$

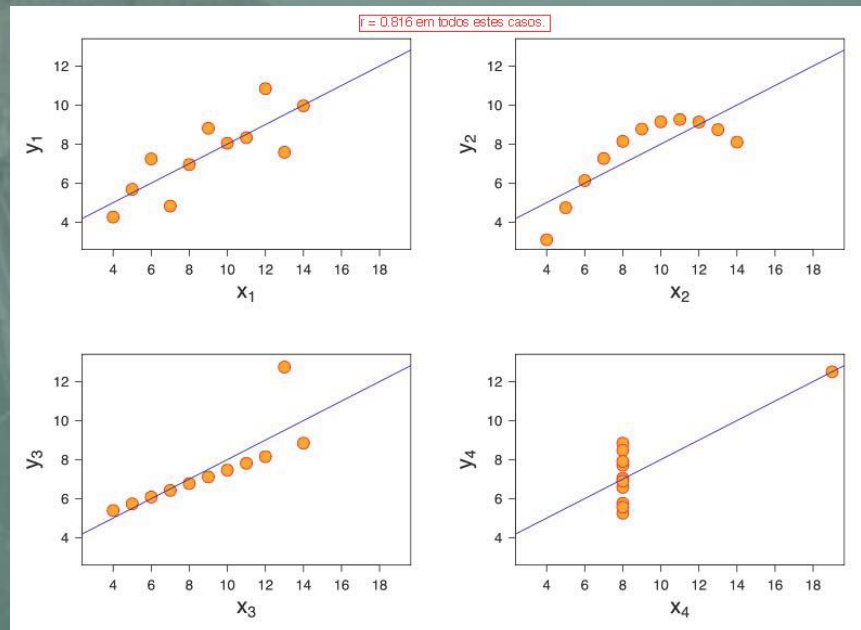


CUIDADO!



www.exercito.gov.br

Só o coeficiente de correlação não é suficiente para afirmarmos que há uma tendência linear. É preciso interpretá-lo em conjunto com o Gráfico de Dispersão. Veja os exemplos abaixo:



Comte Marcos Santos, Prof. Dr.

EXERCÍCIO



www.exercito.gov.br

(Cesgranrio/Engenheiro de Produção Jr/Petrobras/2010) Analisando os diagramas de correlação das variáveis y_1 , y_2 e y_3 , afirma-se que o coeficiente de correlação, em relação a x , da variável:

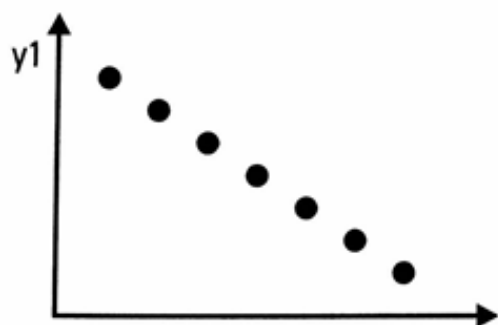


Gráfico I

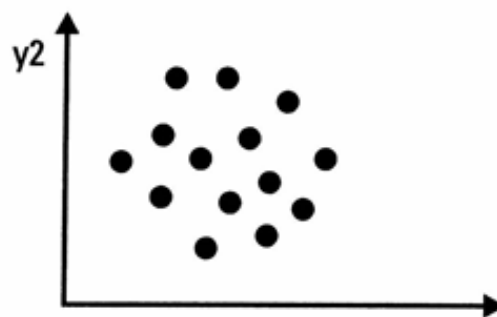


Gráfico II



Gráfico III

- a) y_1 é igual a 1;
- b) y_1 é igual a 0;
- c) y_2 é igual a -1;
- d) y_3 está entre 0 e -1;
- e) y_3 está entre 0 e 1.

REGRESSÃO LINEAR



www.exercito.gov.br

Sempre que desejamos estudar determinada variável em função de outra, fazemos uma **análise de regressão**.

Podemos dizer que a análise de regressão tem por objetivo descrever, através de um modelo matemático, a relação entre duas variáveis, partindo de n observações das mesmas.

A variável sobre a qual desejamos fazer uma estimativa recebe o nome de **variável dependente** e a outra recebe o nome de **variável independente**.

Assim, supondo X a variável independente e Y a variável dependente, vamos procurar determinar o ajustamento de uma reta à relação entre essas variáveis, ou seja, vamos obter uma função definida por:

$$\hat{Y} = aX + b$$

REGRESSÃO LINEAR

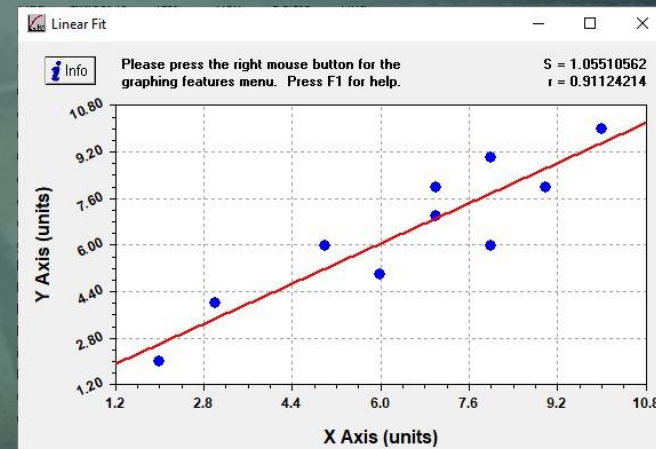


www.exercito.gov.br

Sejam duas variáveis X e Y , entre as quais exista uma correlação acentuada, embora não perfeita, como, por exemplo, na tabela abaixo. Daí temos:

Nº	NOTAS	
	MATEMÁTICA (x_i)	ESTATÍSTICA (y_i)
01	5,0	6,0
08	8,0	9,0
24	7,0	8,0
38	10,0	10,0
44	6,0	5,0
58	7,0	7,0
59	9,0	8,0
72	3,0	4,0
80	8,0	6,0
92	2,0	2,0

Cujo diagrama de dispersão é dado por:



Comte Marcos Santos, Prof. Dr.

REGRESSÃO LINEAR



www.exercito.gov.br

Podemos concluir, pela forma do diagrama, que se trata de uma correlação retilínea, definida pela função:

$$\hat{Y} = aX + b$$

Vamos então calcular os valores dos parâmetros a e b:

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \bar{y} - a\bar{x}$$

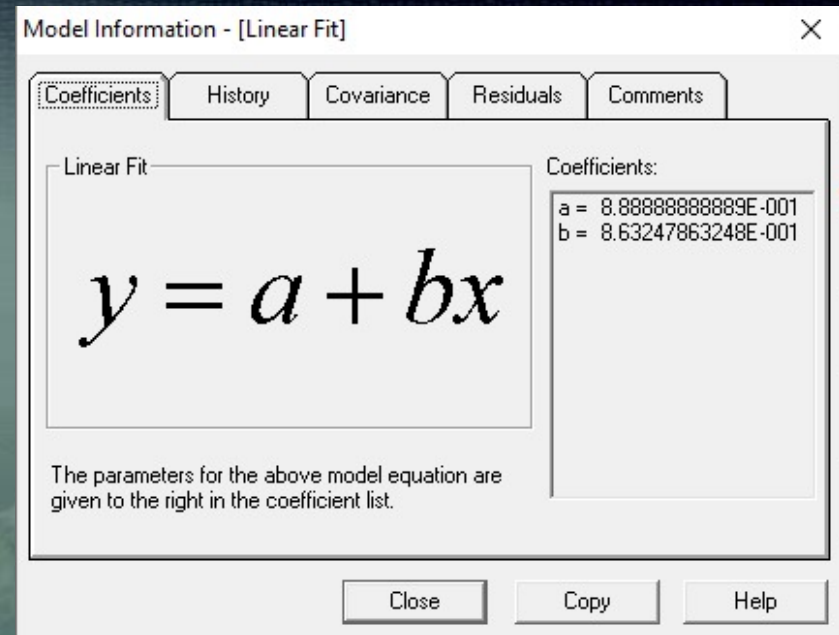
REGRESSÃO LINEAR



www.exercito.gov.br

Vamos então à equação de regressão:

N ^{os}	NOTAS	
	MATEMÁTICA (x _i)	ESTATÍSTICA (y _i)
01	5,0	6,0
08	8,0	9,0
24	7,0	8,0
38	10,0	10,0
44	6,0	5,0
58	7,0	7,0
59	9,0	8,0
72	3,0	4,0
80	8,0	6,0
92	2,0	2,0



INTERPOLAÇÃO E EXTRAPOLAÇÃO



www.exercito.gov.br

Voltando à tabela do exemplo anterior, vemos que 4,0 não figura entre as notas de Matemática. Entretanto, podemos estimar a nota correspondente, fazendo $X = 4,0$ na equação da reta:

INTERPOLAÇÃO E EXTRAPOLAÇÃO



www.exercito.gov.br

Repetindo o mesmo procedimento para $X = 1$, teríamos o seguinte valor de Y :

Como $4 \in [2, 10]$ dizemos que foi feita uma INTERPOLAÇÃO e como 1 não pertence ao intervalo $[2, 10]$ dizemos que foi feita uma EXTRAPOLAÇÃO.

CUIDADO! Uma norma fundamental no uso de equações de regressão é a de nunca extrapolar, exceto quando considerações teóricas ou experimentais demonstrem a possibilidade de extrapolação.

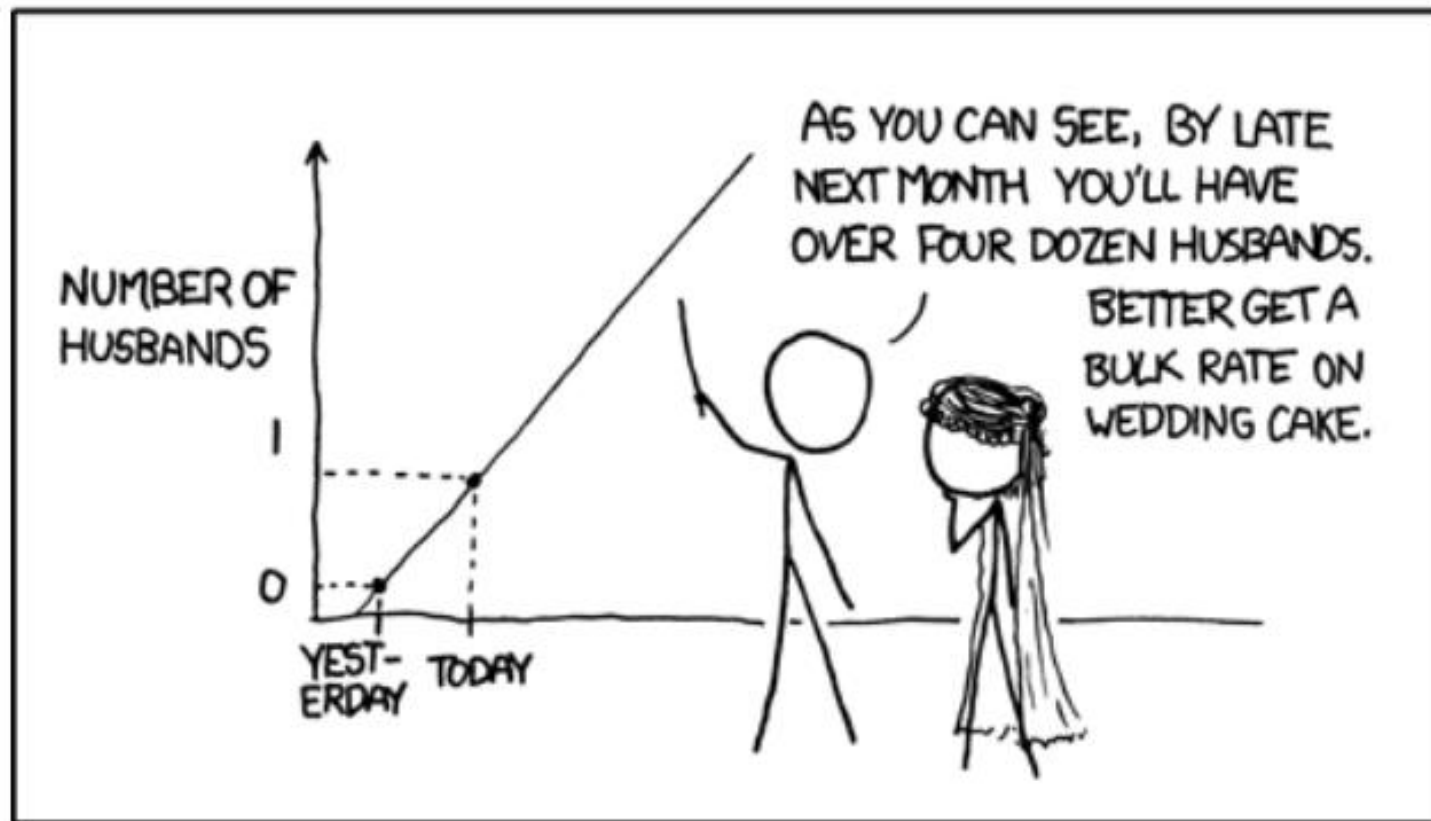
Comte Marcos Santos, Prof. Dr.

INTERPOLAÇÃO E EXTRAPOLAÇÃO



www.exercito.gov.br

MY HOBBY: EXTRAPOLATING



EXERCÍCIO



www.exercito.gov.br

A tabela a seguir, apresenta o tempo de funcionamento (x - em horas) de uma máquina seladora de sacos plásticos e a quantidade de sacos plásticos que saem danificados da referida máquina (y - em unidades).

Verifique se há uma correlação linear entre as referidas variáveis. Caso exista, estabeleça a fórmula da regressão.

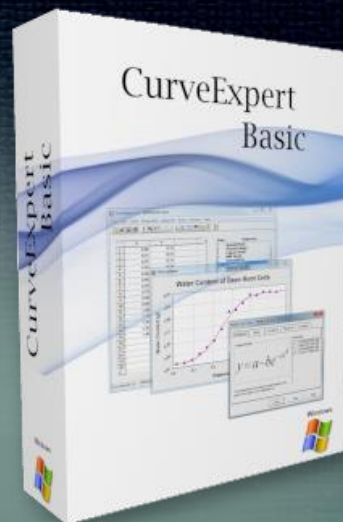
- a. Manualmente;
- b. Utilizando o Excel;
- c. Utilizando o software Curve Expert.

ID	x_i	y_i
1	56	147
2	42	125
3	72	160
4	36	118
5	63	149
6	47	128
7	55	150
8	49	145
9	38	115
10	42	140
11	68	152
12	60	155
Totais	628	1684



ATIVIDADE NO CURVE EXPERT

ID	x_i	y_i
1	56	147
2	42	125
3	72	160
4	36	118
5	63	149
6	47	128
7	55	150
8	49	145
9	38	115
10	42	140
11	68	152
12	60	155
Totais	628	1684



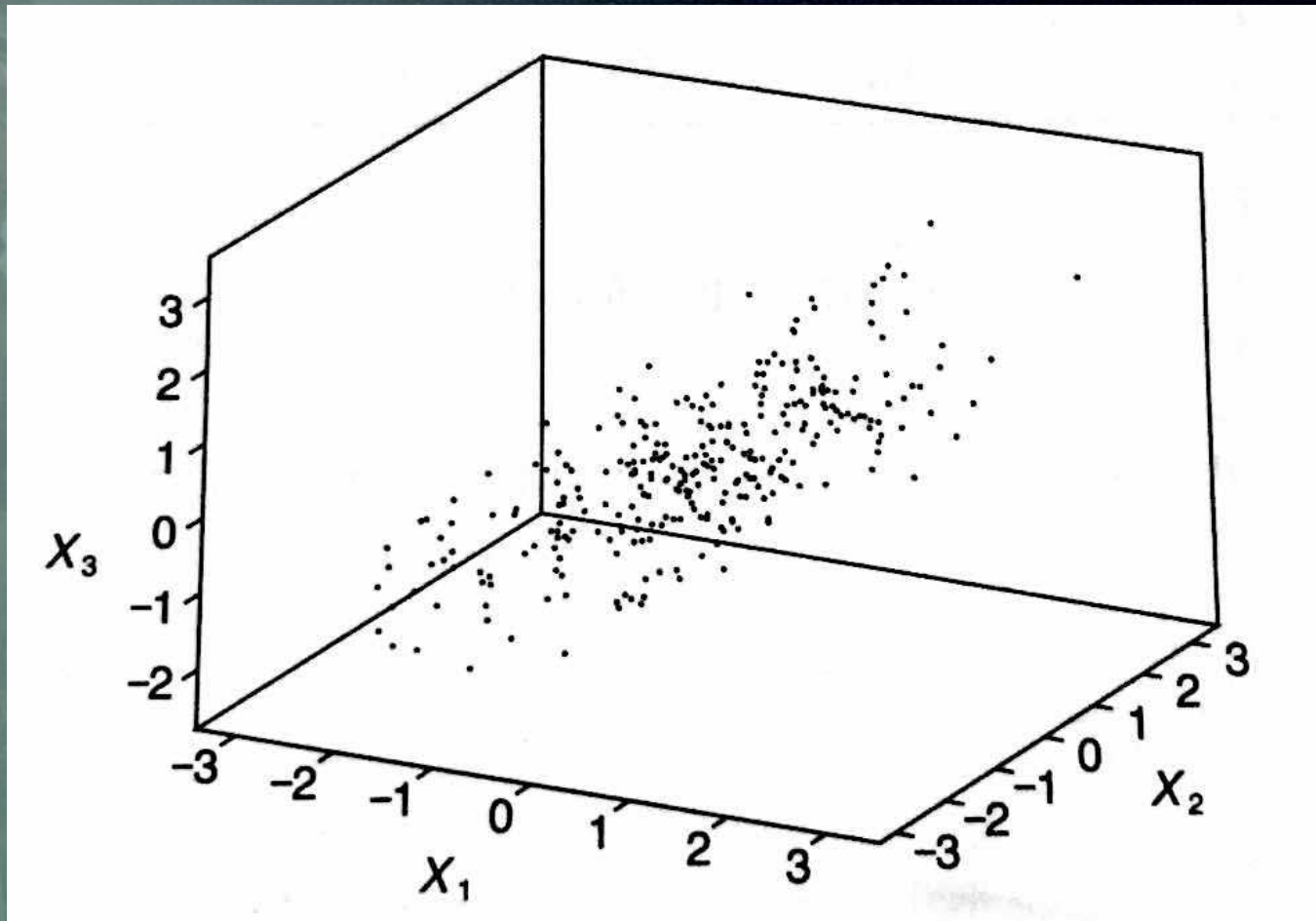
www.exercito.gov.br

Comte Marcos Santos, Prof. Dr.

ESTATÍSTICA MULTIVARIADA



www.exercito.gov.br



Comte Marcos Santos, Prof. Dr.

ARTIGO CIENTÍFICO INTERNACIONAL



www.exercito.gov.br




[New Global Perspectives on Industrial Engineering and Management](#) pp 57-64 | [Cite as](#)

Mapping the Perception of Users as the Usability of Smartphones: Benchmarking Features Through the Borda Count Method

Authors

[Authors and affiliations](#)

M. Santos  , T. Silva, C. Gomes, J. Vieira, R. Walker

Conference paper

First Online: 30 August 2018

508

Downloads

Part of the [Lecture Notes in Management and Industrial Engineering](#) book series (LNMIE)

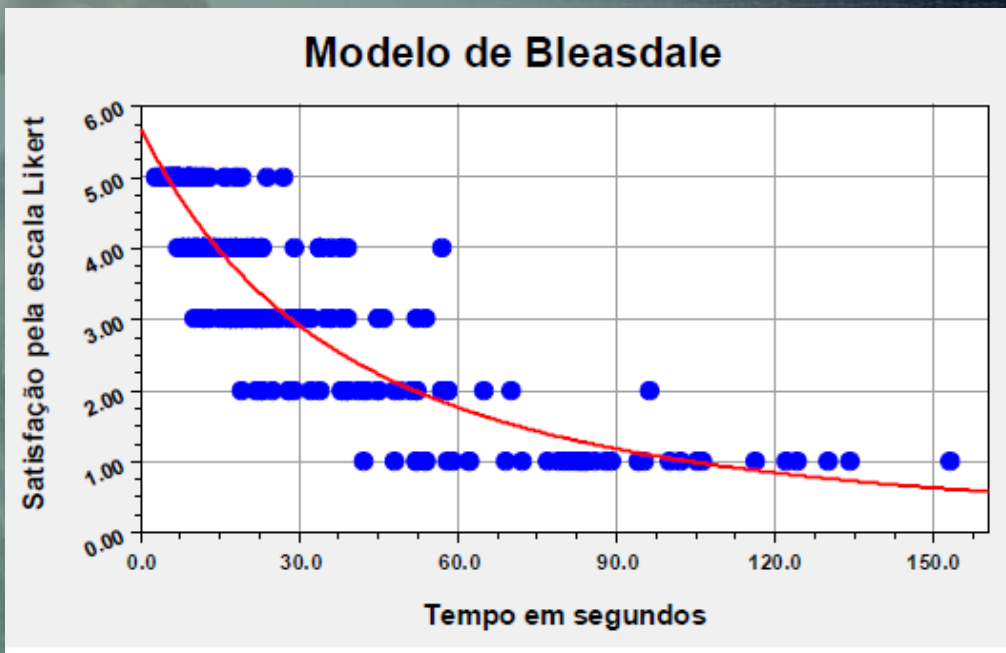


Comte Marcos Santos, Prof. Dr.

ARTIGO DO ENG. THIAGO PORTO



www.exercito.gov.br



$$y = (a + bx)^{\frac{-1}{c}}$$

$$a = 4.25462725229e^{-1}$$

$$b = 5.54199247344e^{-3}$$

$$c = 4.92789764985e^{-1}$$

Comte Marcos Santos, Prof. Dr.

NESSA AULA VIMOS:



www.exercito.gov.br

- ✓ Conceito de Correlação e Regressão;
- ✓ Diagrama de Dispersão;
- ✓ Coeficiente de Correlação Linear;
- ✓ Propriedades do Coeficiente de Correlação Linear;
- ✓ Exercícios;
- ✓ Regressão;
- ✓ Interpolação e Extrapolação;
- ✓ Exercícios.

Comte Marcos Santos, Prof. Dr.

Referências



www.exercito.gov.br

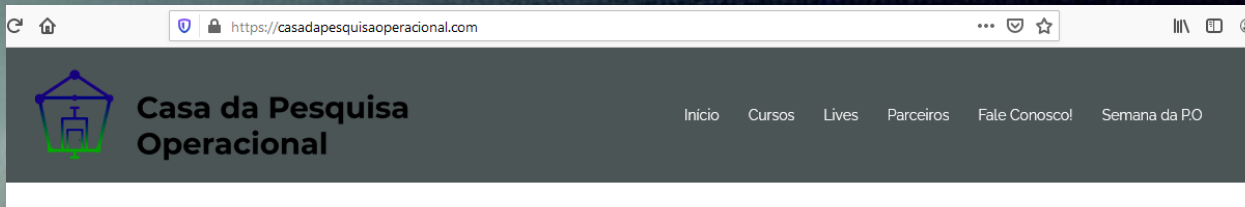
SANTOS, Marcos dos; SILVA, T. L. P.; GOMES, C. F. S.; VIEIRA, J. A. M.; WALKER, R. A. Mapping the Perception of Users as the Usability of Smartphones: Benchmarking Features Through the Borda Count Method. In: MULA, J.; BARBASTEFANO, R.; DIAZ-MANDROÑERO, M.; POLER, R. **New Global Perspectives on Industrial Engineering and Management.** Springer International Publishing, 2018. p.57 – p.64.

Comte Marcos Santos, Prof. Dr.

Casa da Pesquisa Operacional



www.exercito.gov.br



A Pesquisa Operacional é uma ferramenta necessária para a aplicação da pesquisa em contextos reais, permitindo aprender a desenvolver soluções para problemas complexos.

A Casa da Pesquisa Operacional atua no âmbito acadêmico, quanto no âmbito profissional, contextualizada resolvendo problemas reais.

Na CPO, o foco é que você aprenda a aplicar a pesquisa em contextos reais.

Você sabia que a Casa da Pesquisa Operacional tem um canal no Youtube?

Saiba mais em

<https://www.youtube.com/casada pesquisaoperacional>

Casa da Pesquisa Operacional

Acesse nossas redes sociais

<https://youtube.com/c/CasadaPesquisaOperacional>

Comte Marcos Santos, Prof. Dr.

Casa da Pesquisa Operacional



www.exercito.gov.br

● Live 26/05 18:00h



Casa da Pesquisa Operacional

Mediador:

Prof. Dr. Marcos dos Santos



APLICAÇÃO DA LÓGICA FUZZY NA AVALIAÇÃO DA CONFIABILIDADE HUMANA NAS INDÚSTRIAS



Palestrante:

Prof. Dr. Jesus Domech More
Professor Titular IV no IBMEC

Possui graduação em Thermoenergetic Engineering - Moscow Power Engineering Institute (1987), Mestrado em Ciências - Moscow Power Engineering Institute (1987) e Doutorado em Engenharia Metalúrgica e de Materiais pela Universidade Federal de Rio de Janeiro (2004). Atualmente sou Professor Titular IV no Centro Universitário de IBMEC (Instituto Brasileiro de Mercado de Capitais), Pesquisador Visitante na Pontifícia Universidade Católica do Rio de Janeiro e Pesquisador do Laboratório de Lógica Fuzzy na UFRJ. Prêmio Cientista de Nosso Estado (2011); Professor homenageado (3 vezes); Professor paraninfo (6 vezes) e Prêmio de Excelência de Ensino no IBMEC (11 vezes). Tenho experiência na área de Engenharia de Produção, com ênfase em Aplicação de Técnicas Inteligentes, atuando principalmente nos seguintes temas: lógica fuzzy, melhoria de processos, confiabilidade humana, inferência estatística e inteligência computacional aplicada.

CURSO Método de Estabilização de Processos Produtivos: UMA ESTRATÉGIA PARA REDUÇÃO DE CUSTOS

27/05 e 03/06 20:00h às 22:00h



Casa da Pesquisa Operacional

Conteúdo Programático:

- Abordagem à gestão de processos;
- Indicadores de processos Critical to Quality (CTQ);
- Séries Temporais: enxergando oportunidades no processo;
- Elementos de Estabilização do Processo: Gráfico de Controle, Capabilidade e Análise da Causa Raiz;
- Aplicações dos conceitos abordados num estudo de caso.

Curso Exclusivo para Membros da Casa da Pesquisa Operacional

Palestrante:
Prof. David Costa

Bônus: atividades
práticas com o
software Minitab



Os links das aulas serão enviados
pelo grupo Casa da Pesquisa Operacional do Telegram

<https://t.me/casadapesquisaoperacional>

Comte Marcos Santos, Prof. Dr.

Casa da Pesquisa Operacional



www.exercito.gov.br

1^{os} passos com a Pesquisa Operacional e a Ciência de Dados: APRENDENDO A DECIDIR DE MANEIRA ESTRUTURADA E ASSERTIVA
Evento Gratuito e com Certificado

Inscrições pelo Sympla
VAGAS LIMITADAS!

Casa da Pesquisa Operacional

Eng. Prod. Leandro Soares (Instituto de Pesquisas da Marinha)
31/05 20:00H
Otimização da localização de instalações com o método das p-medianas

Prof. Andrei Gomes (Instituto INFNET)
01/06 20:00H
Quadro para tomada de decisões estratégicas: compreensão e assertividade no processo decisório

Tenente-Coronel (FAB) Marcelo Zawadzki, Prof. Dr. (U.S. Air Force Institute of Technology)
02/06 20:00H
Estruturação de Problemas a partir da abordagem Value-Focused Thinking (VFT)

Eng^a Prod. Lana Priscila (University of Arkansas)
03/06 18:00H
Carreira internacional na Engenharia: intercâmbio acadêmico in a nutshell

Prof. Dr. Marcos dos Santos (IME/ITA)
04/06 20:00H
Como resolver problemas reais da sociedade utilizando a Pesquisa Operacional e a Ciência de Dados

Prof. Thiago Marques (ENCE/USP)

•Live



02/06 20:00h



Casa da Pesquisa Operacional

Mediador:

Capitão de Fragata Marcos dos Santos, Prof. Dr.
(Professor do Instituto Militar de Engenharia)



Estruturação de Problemas a partir da abordagem Value-Focused Thinking (VFT)

Palestrante:

Tenente-Coronel da FAB Marcelo Zawadzki, Prof. Dr.
(Professor do U.S. Air Force Institute of Technology)



Possui Bacharelado em Ciências Aeronáuticas pela Academia da Força Aérea (1999). Realizou o curso de Tática Aérea (2000). É Mestre em Ciências pelo Instituto Tecnológico da Aeronáutica - ITA (2009) com ênfase na área de Apoio à decisão. Possui MBA em Gestão Pública pela Universidade Federal Fluminense - UFF (2011). É Doutor pelo Instituto Tecnológico de Aeronáutica - ITA em colaboração com a London School of Economics and Political Science - LSE com ênfase nas Áreas de Alocação de Recursos e Análise de Risco Inteligente. Foi Chefe da Subdivisão de Sistemas de Apoio a Decisão e Chefe da Divisão de Geointeligência do Instituto de Estudos Avançados (IEAV), no Comando da Aeronáutica. Atualmente, é Professor no Air Force Institute of Technology (AFIT) e Diretor do Cost Capability Analysis Certificate Program (CCACP) do AFIT. Como avia-dor militar formou-se Ala operacional da Aviação de Ataque, piloto Instrutor de Transporte Aerotático e Aéreo Logístico. Líder de Seção da Aviação de Transporte de Tropa. Líder de elemento de instrução aérea (AFA).

https://www.sympla.com.br/primeiros-passos-com-a-pesquisa-operacional-e-a-ciencia-de-dados__1226299

Comte Marcos Santos, Prof. Dr.



www.exercito.gov.br

TIRE suas DUVIDAS

Comte Marcos Santos, Prof. Dr.