UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS

Conceitos de Probabilidade e Estatística

Média Aritmética:

Soma dos valores do grupo de dados dividida pelo número de valores.

$\mu = \Sigma X / N$ (População) e $x = \Sigma X / n$ (Amostra)

Média Ponderada:

Média aritmética na qual cada valor é ponderado de acordo com sua importância no grupo.

$\mu_w = \Sigma (wX) / \Sigma w$

Mediana:

É o valor do item médio quando os itens do grupo foram ordenados.

Moda:

O valor que + freqüentemente ocorre em um conjunto de valores.

- Grupo de dados com 1 moda → *unimodal*
- Grupo de dados com 2 modas → *bimodal*
- Grupo de dados com mais de duas modas → *multimodal*

Exemplo:

Schaum

Tempo necessário para processar e preparar encomendas postais

Tempo (em minutos)	*Fronteiras de classe*	*Ponto médio (X)*	*Número de encomendas (f)*	*fX*	*Freqüência acumulada (F)*
5 e menor do que 8	5,0 8,0	6,5	10	65,0	10
8 e menor do que 11	8,0 11,0	9,5	17	161,5	27
11 e menor do que 14	11,0 14,0	12,5	12	150,0	39
14 e menor do que 17	14,0 17,0	15,5	6	93,0	45
17 e menor do que 20	17,0 20,0	18,5	2	37,0	47
			Total 47	$\Sigma(fX) = 506,5$	

n/2: tam. amostra/2; $\mathbf{F_B}$: freq. acum. da classe anterior a classe da mediana; $\mathbf{f_c}$: núm. de observ. da classe da mediana; **i**: amplitude da classe da mediana;

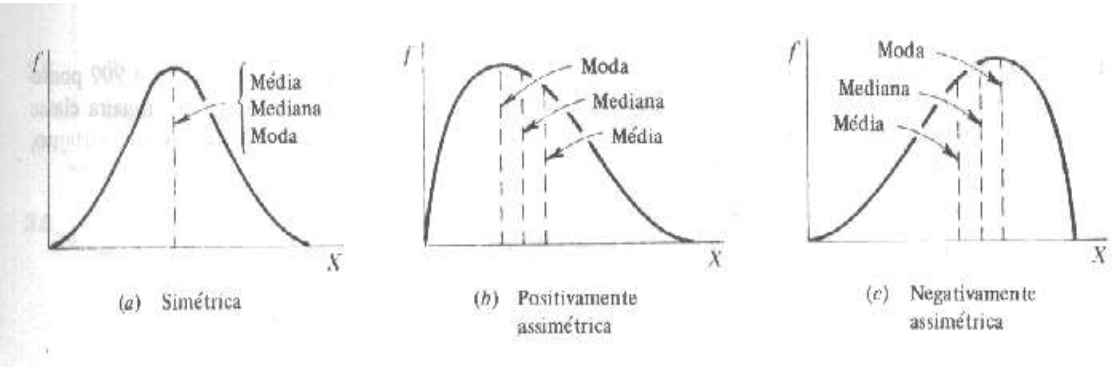
Resp. (a) $\bar{X} = \dfrac{\Sigma(fX)}{n} = \dfrac{506,5}{47} = 10,8 \text{ min}$

(b) $\text{Med} = B_L + \left(\dfrac{\frac{n}{2} - F_B}{f_c}\right) i = 8,0 + \left(\dfrac{23,5 - 10}{17}\right) 3,0 = 10,4 \text{ min}$

(*Nota:* 8,0 é a fronteira inferior da classe que contém a medida $\frac{n}{2}$, ou a 23,5ª medida.)

(c) $\text{Moda} = B_L + \left(\dfrac{d_1}{d_1 + d_2}\right) i = 8,0 + \left(\dfrac{7}{7 + 5}\right) 3,0 = 9,75 \cong 9,8 \text{ min}$

d1: freq. classe modal – freq. classe preced.; **d2**: freq. classe modal – freq. classe posterior. ; **i**: amplitude da classe da modal.



Medidas de Dispersão:

These tables list the number of sales they have made daily over a 12-day period.

Fancy Vanilla

15	12	18	10
12	13	13	15
14	16	20	22

Super Duper Chocolate Crunch

15	12	21	18
14	10	15	10
30	7	11	5

Find the mean deviation, variance and standard deviation of the above data. **Solution**:
Fancy Vanilla ice cream:

$$\bar{x} = \frac{15+12+18+10+12+13+13+15+14+16+20+22}{12} = 15$$

Super Duper Chocolate Crunch ice cream:

$$\bar{x} = \frac{15+12+21+18+14+10+15+10+30+7+11+5}{12} = 14$$

Fancy Vanilla

x_i	$\bar{x} - x_i$	$\left(\bar{x} - x_i\right)^2$
15	0	0
12	3	9
18	-3	9
10	5	25
12	3	9
13	2	4
13	2	4
15	0	0

14	1	1
16	-1	1
20	-5	25
22	-7	49

Variance:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{12}[0+9+9+25+9+4+4+0+1+1+25+49]$$

$$= \frac{136}{12} = 11.3$$

Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{11.3} = 3.4$$

Super Duper Chocolate Crunch

x_i	$\bar{x} - x_i$	$(\bar{x} - x_i)^2$
15	-1	1
12	2	4
21	-7	49
18	-4	16
14	0	0
10	4	16
15	-1	1
10	4	16
30	-16	256
7	7	49
11	3	9
5	9	81

Variance:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{12}[1+4+49+16+0+16+1+16+256+49+9+81]$$

$$= \frac{498}{12} = 41.5$$

Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{41.5} = 6.4$$

Since the standard deviation for Fancy Vanilla is less than the standard deviation for Super Duper Chocolate (3.4 < 6.4), the sales from Fancy Vanilla deviate less from the mean. This indicates that ***the mean for Fancy Vanilla is a more reliable measure of its central tendency. The expected number of daily sales for Fancy Vanilla is more predictable.***

Modelagem (análise) de Dados de Entrada

Corresponde a **Fase de Preparação de Dados** no *Processo de Simulação*

No contexto discreto-estocástico:

Dados da realidade → Coleta de dados (amostragem) → análise dos dados → identificação da distribuição mais adequada a amostra → incorporar esta informação no modelo

Vantagens da abordagem estocástica:

- Manipulação dos "ruídos";
- Grande biblioteca de distribuições;
- Flexibilidade operacional;

Alternativas:

- Construír distribuição específica (sob medida): difícil, caro e demorado;
- Usar dados coletados (*trace-driven approach*): sem flexibilidade operacional.

Distribuições de Probabilidade:

Conjunto de valores que relacionam a freqüência relativa com a qual um determinado evento ocorre.

Distribuições Contínuas

- Simétricas : Normal, Uniforme, Beta
- Assimétricas (skewed): Exponencial, Gamma, Weibull, LogNormal, Erlang, Triangular

Distribuições Discretas: Poisson, Binomial, Uniforme

Escolha da melhor distribuição para um determinado conjunto de dados (amostra):

Goodness-of-Fit Tests (GOD) ou simplesmente "Fit Tests"

➢ **Testes Paramétricos**

Decisões baseadas em parâmetros derivados das distribuições; assumem que os dados podem ser descritos (aproximadamente) por uma distribuição Normal. Os parâmetros envolvidos são geralmente a média e o desvio-padrão.

Exemplos: *Qui-Quadrado* (χ^2), *t*, F.

Este tipo de teste se baseia no Teorema do Limite Central (TLC).

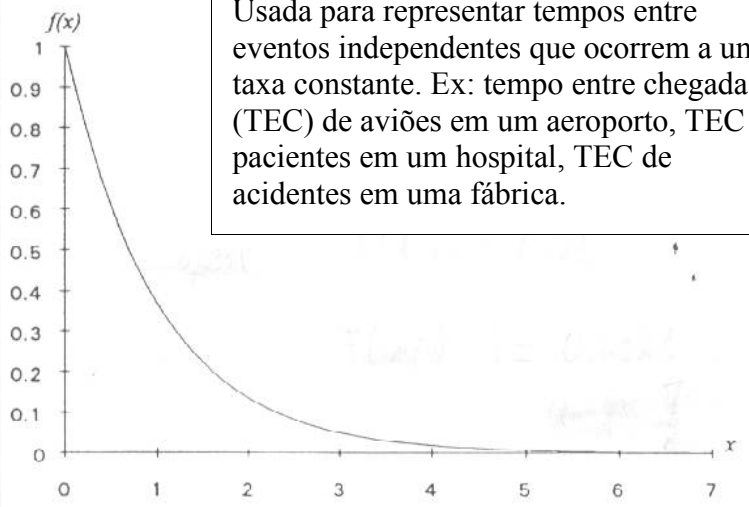
Teorema do Limite Central: à medida que o tamanho da amostra aumenta, a distribuição de amostragem da média se aproxima da forma da distribuição Normal, qualquer que seja a forma da distribuição da população. A distribuição pode ser considerada como aproximadamente Normal sempre que o tamanho da amostra for ≥ 30. ($n \geq 30$).

- **Testes não-paramétricos**

Não supõe nada em relação à forma da distribuição sendo testada. Não envolvem parâmetros da distribuição.
Podem ser usados quando o tamanho da amostra é pequeno ($n \leq 30$), o que exclui a aplicação do TLC.

Exemplos: *Kolmogorov-Smirnov* (KS) ; *Mann-Whitney*

Exponential Distribution β=1



Usada para representar tempos entre eventos independentes que ocorrem a uma taxa constante. Ex: tempo entre chegadas (TEC) de aviões em um aeroporto, TEC de pacientes em um hospital, TEC de acidentes em uma fábrica.

Probability Density Function

$$f(x) = \frac{1}{\beta} e^{-x/\beta}$$

for $x \geq 0$ else 0

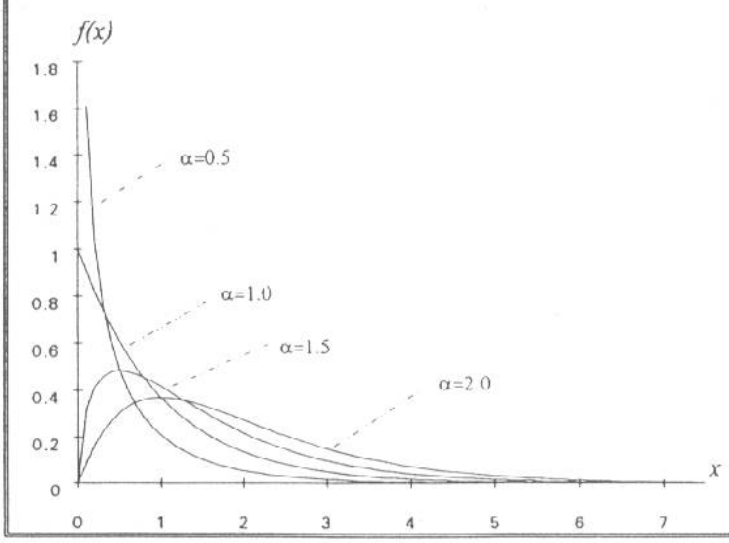
Cumulative Distribution Function

$$F(x) = 1 - e^{-x/\beta}$$

for $x \geq 0$ else 0

Mean=β

Gamma Distribution β=1, α=0.5, 1.0, and 2.0



Probability Density Function

$$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)}$$

for $x > 0$ else 0

Cumulative Distribution Function

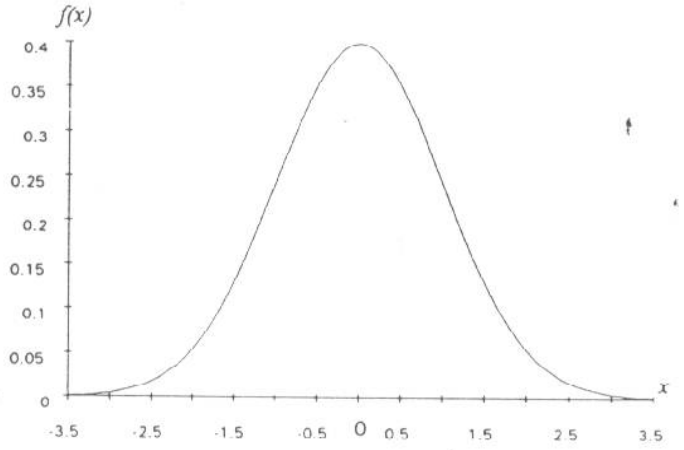
$$F(x) = 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} \frac{(x/\beta)^j}{j!}$$

for $x > 0$ else 0

Mean = $\alpha\beta$

Usar para tempos de tarefas que não conseguem ser muito mais rápidas que o tempo (duração) média, mas que eventualmente podem durar bem mais. Ex.: se durante uma tarefa houver um problema, esta instância vai durar bem mais que as outras ocorrências desta tarefa.

Normal Distribution $\mu=0$, $\sigma=1$



Probability Density Function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

for all real number x

Cumulative Distribution Function

No Closed Form

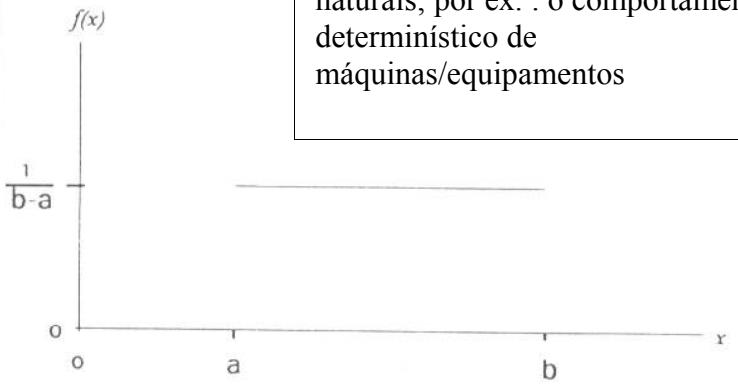
Mean=μ

$$\mu \approx \bar{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\sigma \approx \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \bar{X}(n)\right)^2}{n}}$$

Uniform Distribution

Para descrever processos não naturais; por ex. : o comportamento determinístico de máquinas/equipamentos



Probability Density Function

$$f(x) = \frac{1}{b-a}$$

for $a \le x \le b$ else 0

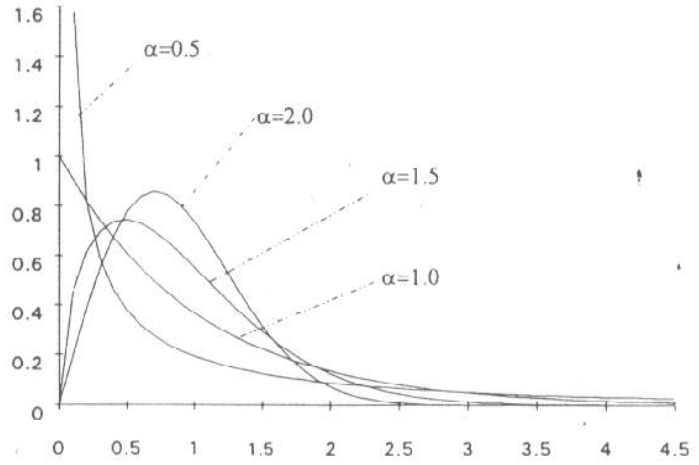
Cumulative Distribution Function

$$F(x) = \frac{x-a}{b-a} \quad \text{for } a \le x \le b$$

0 for $x < a$ 1 for $b < x$

Mean $= \frac{(a+b)}{2}$

Weibull Distribution $\beta=1$, $\alpha=0.5, 1.0, 1.5, 2.0$



Probability Density Function

$$f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}}$$

for $x > 0$ else 0

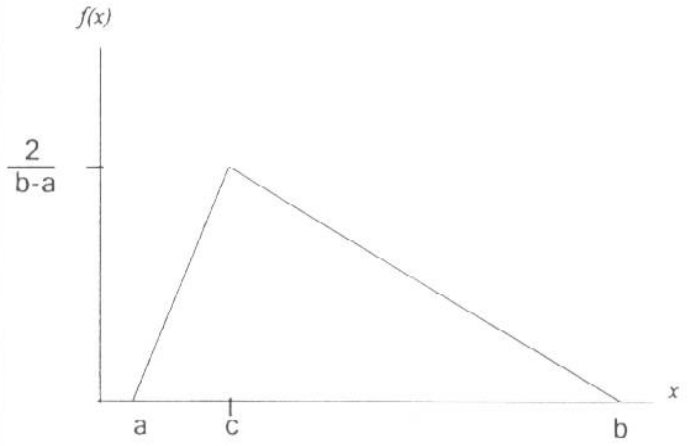
Cumulative Distribution Function

$$F(x) = 1 - e^{-(x/\beta)^{\alpha}}$$

for $x > 0$ else 0

$$\text{Mean} = \frac{\beta}{\alpha}\Gamma(1/\alpha)$$

Triangular Distribution



Probability Density Function

$$f(x) = \frac{2(b-x)}{(b-a)(b-c)} \quad \text{for } c \le x \le b$$

$$f(x) = \frac{2(x-a)}{(b-a)(c-a)} \quad \text{for } a \le x \le c$$

0 otherwise

Cumulative Distribution Function

$$F(x) = \frac{(x-a)^2}{(b-a)(c-a)} \quad \text{for } a \le x \le c$$

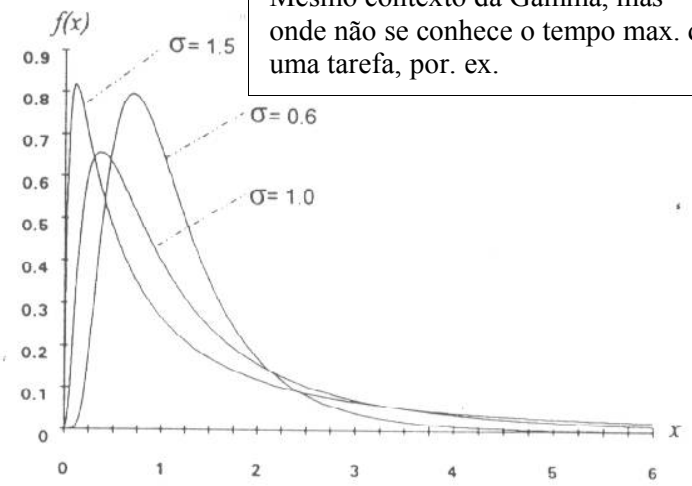
$$F(x) = 1 - \frac{(b-x)^2}{(b-a)(b-c)} \quad \text{for } c < x \le b$$

0 for $x < a$ and 1 for $b < x$

$$\text{Mean} = \frac{a+b+c}{3}$$

Lognormal Distribution μ=0, σ=0.6, 1.0, 1.5

Descreve valores cujos logaritmos naturais são dist. Normalmente. Mesmo contexto da Gamma, mas onde não se conhece o tempo max. de uma tarefa, por. ex.



Probability Density Function

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\ln x-\mu)^2/2\sigma^2}$$

for x > 0 else 0

Cumulative Distribution Function

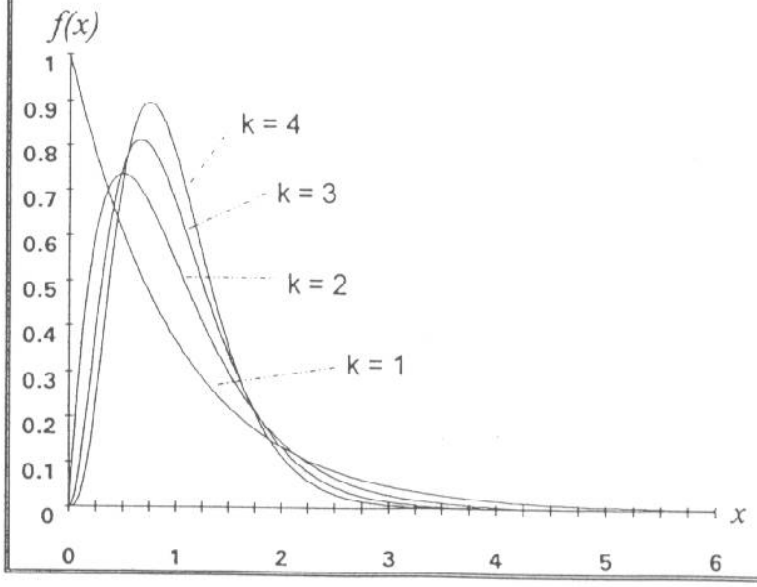
No Closed Form

$$\text{Mean} = e^{\mu+\sigma^2/2}$$

$$\hat{\mu} \approx \frac{\sum_{i=1}^{n} \ln X_i}{n}$$

$$\hat{\sigma} \approx \sqrt{\frac{\sum_{i=1}^{n} (\ln X_i - \hat{\mu})^2}{n}}$$

Erlang Distribution μ=1, k=1, 2, 3, and 4 (k must be a positve integer)



Probability Density Function

$$f(x) = \frac{(\mu k)^k}{(k-1)!} x^{k-1} e^{-k\mu x}$$

for x > 0 else 0

Cumulative Distribution Function

$$F(x) = 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} \frac{(x/\beta)^j}{j!}$$

for x > 0 else 0

$$\text{Mean} = \frac{1}{\mu} = \frac{1}{k\beta}$$

Discrete Probability Distributions

Poisson Distribution

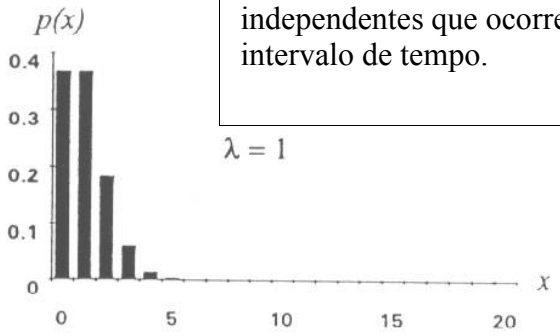
Modela o número de eventos independentes que ocorrem em um intervalo de tempo.

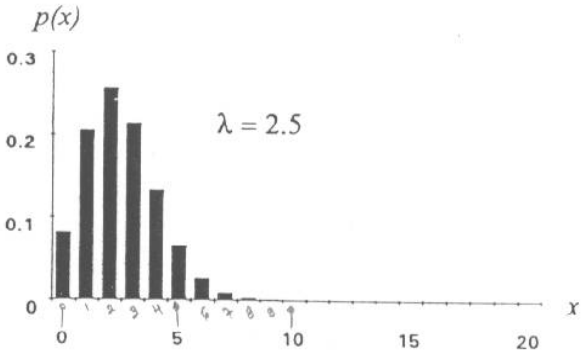
Probability Density Function

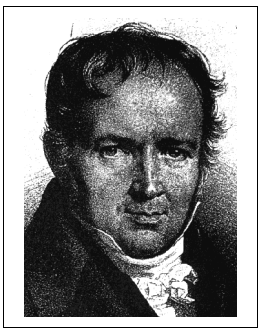
$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

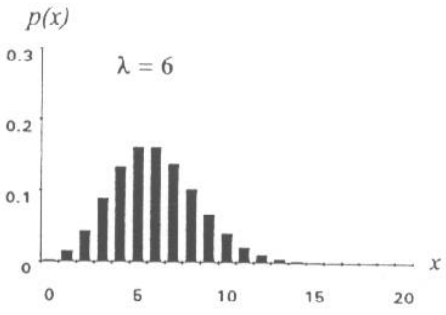
If x is a positive integer ≥ 0 else $p(x) = 0$

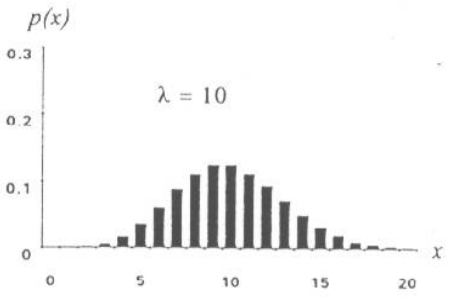
Mean-λ

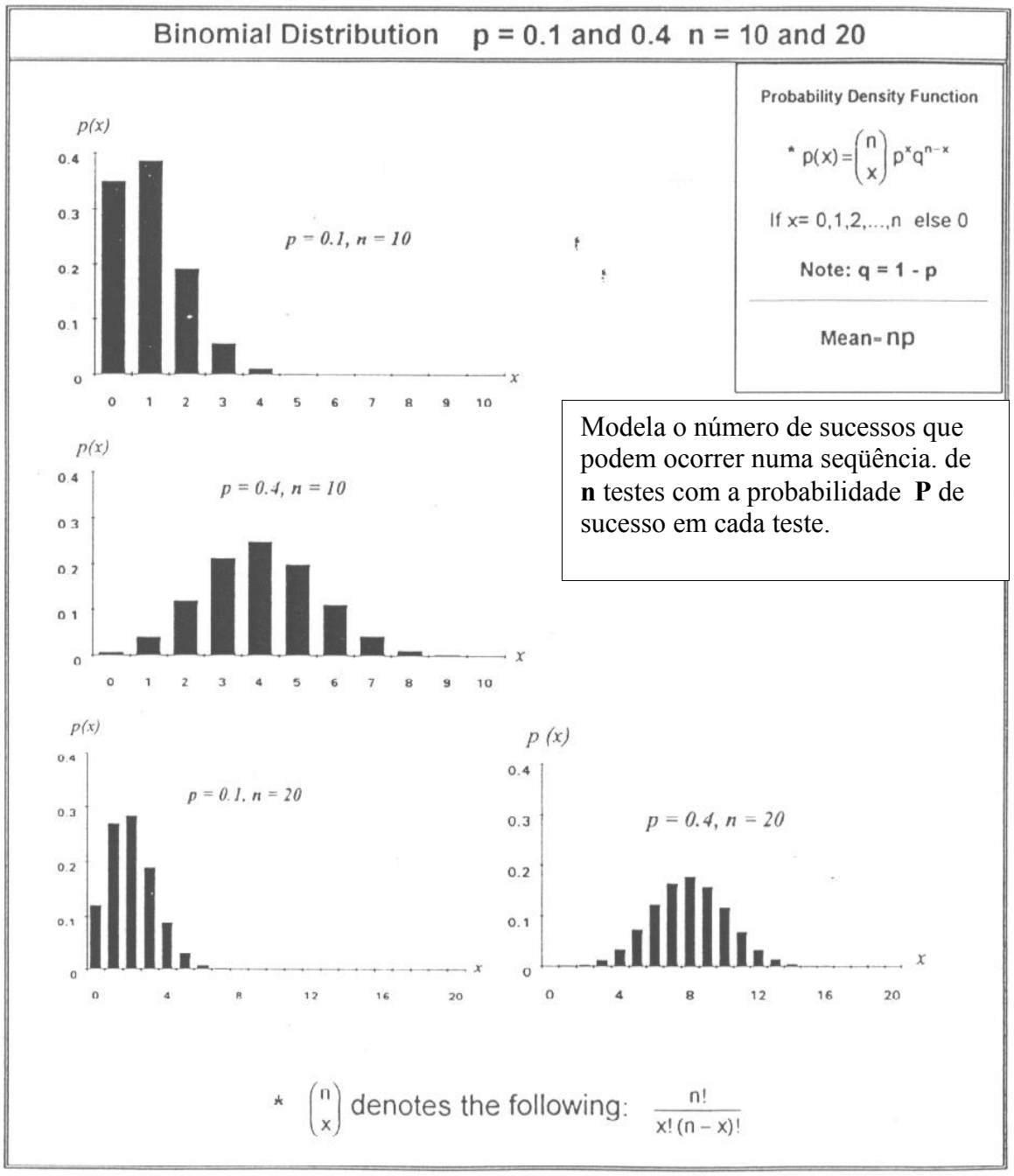






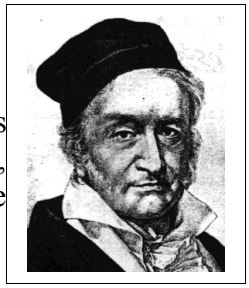






Modela o número de sucessos que podem ocorrer numa seqüência. de **n** testes com a probabilidade **P** de sucesso em cada teste.

➢ **Mais sobre a Distribuição Normal:**



Although the distribution was discovered earlier, the normal curve is often called the **Gaussian curve** because the German mathematician, Carl Friedrich Gauss (1777 – 1855) found so many applications for the normal curve.

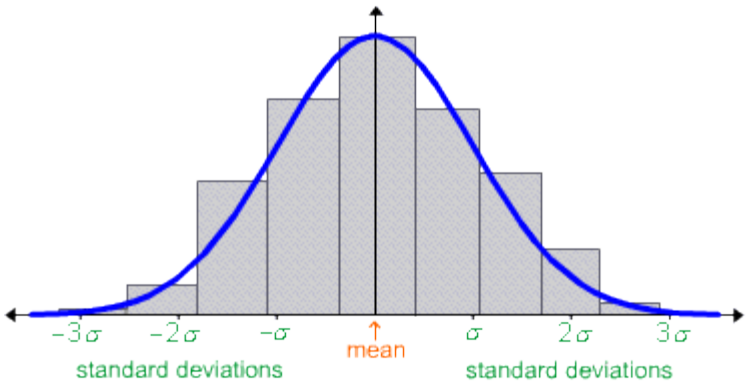
Gauss

The *normal curve* was developed mathematically in **1733** by DeMoivre as an approximation to the binomial distribution. His paper was not discovered until 1924 by Karl Pearson. Laplace used the normal curve in 1783 to describe the *distribution of errors*. Subsequently, Gauss used the normal curve to analyze astronomical data in 1809.

Mean and Standard Deviation

- $\bar{x}$ - the population mean of the distribution. Remember: this occurs at the peak of the distribution

- σ - the population standard deviation of the distribution



Properties of Normal Distributions

- The curve is symmetric about the mean, $\bar{x}$.

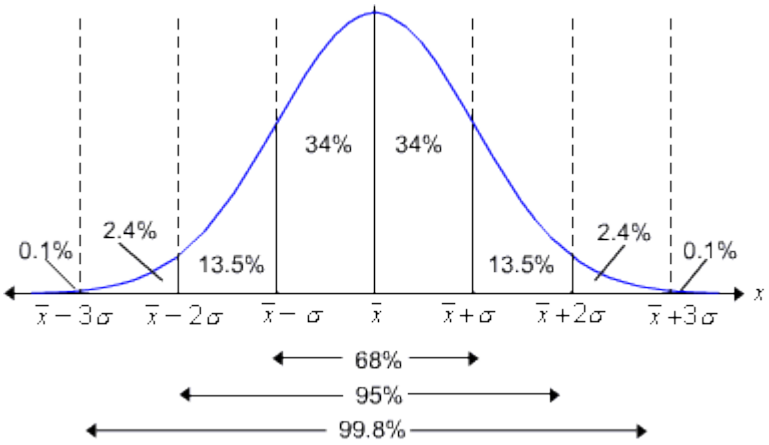
- The total area under the curve is 1.

- The mean, median and mode are equal.

- About 68% of the area is within 1 standard deviation (σ) of the mean.

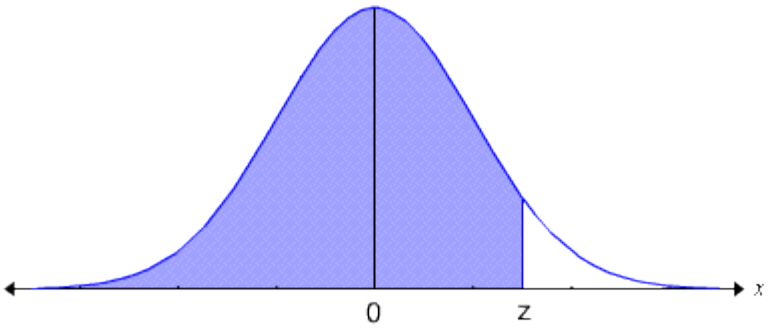
- About 95% of the area is within 2 standard deviations (2σ) of the mean.

- About 99.8% of the area is within 3 standard deviations (3σ) of the mean.



The **Standard Normal Distribution** has a mean of 0 and a standard deviation of 1. The letter Z is often used to refer to a standard normal random variable.

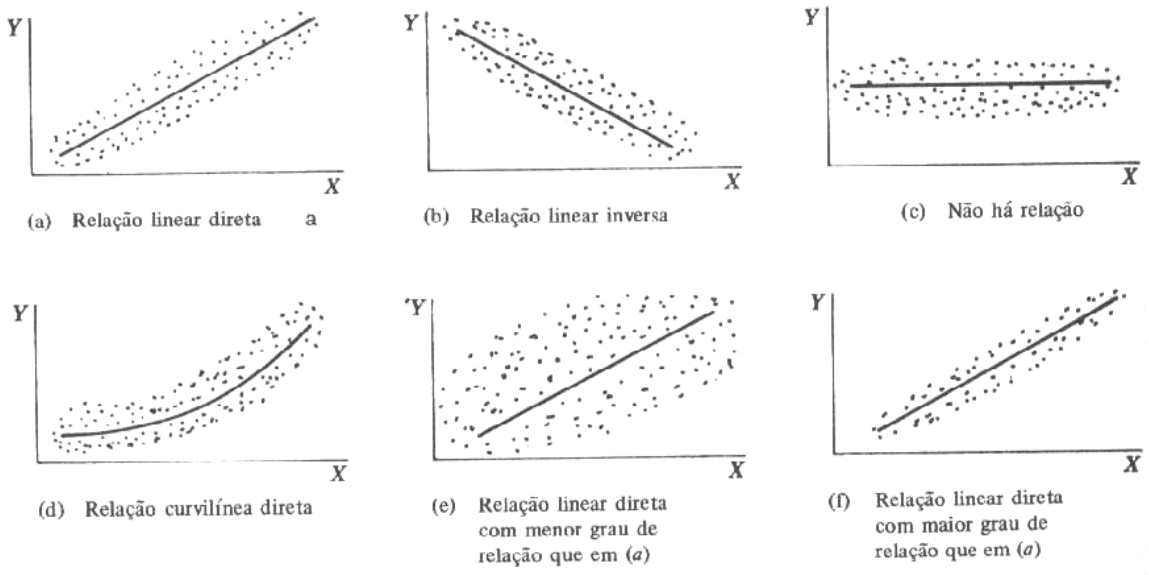
Note that, although many applications in the real world have a normal distribution, rarely does anything in the real world follow a *standard* normal distribution. This is a convenient distribution that can be used (after some transformations) for ANY normal distribution.



Identificação de relações entre variáveis

Diagrama de dispersão:

Eixo **y**: variável dependente
Eixo **x**: variável independente



(a) Relação linear direta a

(b) Relação linear inversa

(c) Não há relação

(d) Relação curvilínea direta

(e) Relação linear direta com menor grau de relação que em (*a*)

(f) Relação linear direta com maior grau de relação que em (*a*)

***Ajuste de uma linha de regressão*:** método dos **Mínimos Quadrados**

$$a = \bar{Y} - b\bar{X} \qquad b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2}$$

Coeficiente de Correlação:

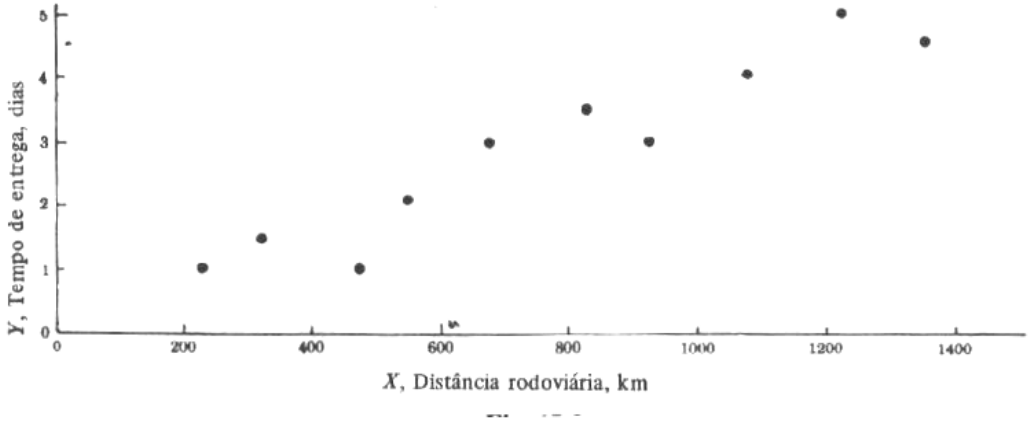
$$r = \frac{n \Sigma XY - \Sigma X \Sigma Y}{\sqrt{n \Sigma X^2 - (\Sigma X)^2}\sqrt{n \Sigma Y^2 - (\Sigma Y)^2}}$$

Exemplo:

Tabela 17.1 Amostra de observações de distâncias rodoviárias e tempo de de entrega para 10 carregamentos aleatoriamente selecionados

Carregamento amostrado	1	2	3	4	5	6	7	8	9	10
Distância X, em km	825	215	1070	550	480	920	1350	325	670	1215
Tempo de entrega, Y, em dias	3,5	1,0	4,0	2,0	1,0	3,0	4,5	1,5	3,0	5,0

Resp. O diagrama de dispersão para tais dados encontra-se na Fig. 17-3. O primeiro par de valores apresentado na tabela é representado pelo ponto acima de 825 no eixo dos X e alinhado com 3,5 com respeito ao eixo dos Y. Os outros nove pontos do diagrama de dispersão foram plotados de maneira similar. Pelo diagrama, parece que os pontos seguem, de modo geral, uma relação linear. Então, parece apropriada ao caso a análise de regressão linear.



Determinar a equação de regressão de mínimos quadrados para os dados no Problema 17.1, e traçar a linha de regressão no diagrama de dispersão para os dados.

Resp. Com referência à Tabela 17.2,

$$b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} = \frac{(26.370) - (10)(762)(2,85)}{7.104.300 - (10)(762)^2} = \frac{4653}{1.297.860} = 0,0035851 \cong 0,0036$$

$$a = \bar{Y} - b\bar{X} = 2,85 - (0,0036)(762) = 0,1068 \cong 0,11$$

Portanto,

$$\bar{Y}_x = a + bX = 0,11 + 0,0036X$$

Tabela 17.2 Cálculos para a determinação da equação de regressão linear para estimar o tempo de entrega com base na distância rodoviária

Carregamento amostrado	*Distância X, em km*	*Tempo de entrega, Y, em dias*	XY	X^2	Y^2
1	825	3,5	2887,5	680.625	12,25
2	215	1,0	215,0	46.225	1,00
3	1070	4,0	4280,0	1.144.900	16,00
4	550	2,0	1100,0	302.500	4,00
5	480	1,0	480,0	230.400	1,00
6	920	3,0	2760,0	846.400	9,00
7	1350	4,5	6075,0	1.822.500	20,25
8	325	1,5	487,5	105.625	2,25
9	670	3,0	2010,0	448,900	9,00
10	1215	5,0	6075,0	1.476.225	25,00
Totais	7620	28,5	26.370,0	7.104.300	99,75
Media	$\bar{X} = \frac{\Sigma X}{n} = \frac{7620}{10} = 762$	$\bar{Y} = \frac{\Sigma Y}{n} = \frac{28,5}{10} = 2,85$			

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

$$= \frac{(10)(26.370) - (7.620)(28,5)}{\sqrt{(10)(7.104.300) - (7.620)^2}\sqrt{(10)(99,75) - (28,5)^2}} =$$

$$= \frac{46.530}{(3.602,5824)(13,6107)} = \frac{46.530}{49.033,668} = +0,9489 \cong +0,95$$