1. INTRODUÇÃO

Em engenharia, na maioria das vezes, pode-se descrever certo fenômeno por intermédio de um modelo matemático. Um Modelo pode ser entendido como a visão ou cenário de uma parte de um todo maior e completo. Para Leal (1999) dente as várias definições para Modelo Matemático a mais adequada é de que um modelo pode ser formulado em termos familiares, tais como, expressões numéricas ou fórmulas, diagramas, gráficos ou representações geométricas, equações algébricas tabelas, entre outros.

Assim, para se modelar os fenômenos de observação, utilizam-se os modelos matemáticos de dois tipos: determinístico e <u>não-determinístico</u>. O primeiro refere-se a experimentos que apresentem resultados adequados a um padrão matemático, onde pequenos desvios ou erros não são consideráveis a ponto de alterar o modelo. O segundo refere-se a modelos estocásticos, isto é, os resultados do modelo podem apresentar desvios ou erros que alteram a condição inicial do modelo, mesmo que se conheçam as suas possíveis respostas, denotando-se certa aleatoriedade do fenômeno.

É natural que se associe os Modelos Estocásticos aos Modelos Estatísticos. Por isso, cabe de destacar que os últimos são divididos em Descritivos e Inferenciais. Os descritivos envolvem coletas, apresentação, descrição e caracterização de dados; os inferenciais consideram estimativas e testes de hipótese objetivando à tomada de decisão sobre certo universo de dados.

Considerando-se então os experimentos em que os resultados não sejam previsíveis antecipadamente, tais como lançamento de uma moeda, jogar um dado, vida útil de um equipamento mecânico etc., pode-se considerar como espaço amostral os resultados possíveis destes. Sendo assim, toma-se o <u>Espaço Amostral</u> é o <u>conjunto universo</u> ou o conjunto de resultados possíveis de certo experimento aleatório.

Para exemplificar os experimentos aleatórios, considera-se uma moeda lançada. Deduz-se que o Espaço Amostral = {cara, coroa}; para um dado jogado, Espaço Amostral = {1, 2, 3, 4, 5, 6}; para vida útil de um equipamento mecânico, Espaço Amostral = $[0, \infty)$. Nos dois primeiros exemplos, tomou-se o espaço amostral como finito; no último, infinito.

Para tais experimentos, tem-se que o <u>subconjunto</u> de cada <u>espaço amostral</u> é denominado <u>evento</u>. Para o exemplo da moeda, evento 1 = {cara} e evento 2 = {coroa}; para o dado, podem-se considerar os resultados que são números pares, ou seja, evento 1 = {2, 4, 6}; para um equipamento que dure ao menos 1 anos, mas não complete o segundo, tem-se evento 1 = $[1,2)$.

Tomando-se, então, os resultados de um experimento que pode ser listado pelo <u>Espaço Amostral (S)</u> com os seus <u>Eventos (E)</u>, observa-se que a <u>probabilidade P</u> <u>de certo evento</u>

- P(E) – compreende-se entre 0 e 1, isto é: $0 \leq P(E) \leq 1$; Além disso, a probabilidade deste espaço amostral é igual a 1 - P(S) = 1.

Nestes termos, uma Variável Aleatória pode ser entendida como o resultado de uma medição de algum parâmetro que pode gerar um valor diferente a cada medida, ou seja, diz respeito à característica do experimento que se quer estudar. Matematicamente, ela é a função que associa cada elemento de um espaço amostral a um número real. Por exemplo, se ao lançar uma moeda três vezes, obtém-se o seguinte espaço amostral: S = {(ccc), (kcc), (ckc), (cck), (kkk), (kkc), (kck), (ckk)}, sendo *c* representando "cara" e *k*, "coroa". Para tal situação, imagina-se que se necessita avaliar a quantidade de caras possíveis. Assim, a variável aleatória *X*, que representa a quantidade de "caras", pode ser expressa da seguinte forma:

x = 0 (nenhuma cara) {(kkk)}
x = 1 (uma cara) {(kkc)(kck)(ckk)}
x = 2 (duas caras) {(kcc)(ckc)(cck)}
x = 3 (três caras) {(ccc)}

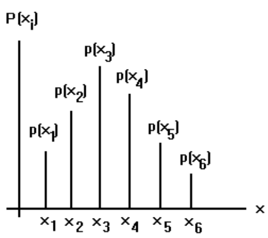
Uma Variável Aleatória Discreta assume cada um dos seus valores com certa probabilidade, conforme a seguir:

x	**0**	**1**	**2**	**3**
P(X = x)	1/8	3/8	3/8	1/8

Sendo assim, uma Distribuição de Probabilidade é um modelo matemático que relaciona certo valor da variável em estudo com a sua probabilidade de ocorrência.

As variáveis aleatórias podem ser classificadas em:

- Discretas (VAD) – a quantidade de valores possíveis, assumidos por X, for contável e finita.
- Contínua (VAC) – a quantidade de valores possíveis, assumidos por X, for formada por intervalos, ou seja, por valores não-contáveis. Podem ser determinadas por medição.



Para Variáveis Discretas (fig. ao lado), a probabilidade de que a variável X assuma um valor específico x é dada por: P(X = x) = P(x)

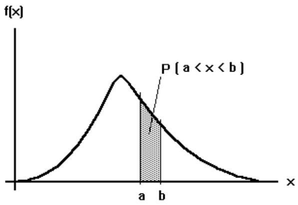
Essa função de probabilidade deve satisfazer as seguintes condições:

i) $P\{x_i\} \geq 0$, para todo x_i;

ii) $\sum_{i} P\{x_i\} = 1$.

Considerando-se <u>Variáveis Contínuas</u>, as probabilidades são especificadas em termos de intervalos, pois a probabilidade associada a um número específico é zero, ou seja,

$$P\{a \le x \le b\} = \int_a^b f(x)\,dx$$



Exemplos:

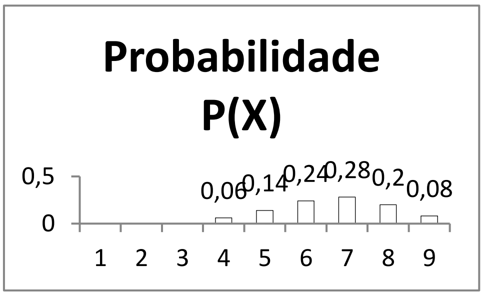
1) Para VAD:

a) Jogar um dado não viciado (não viesado):
S = {1, 2, 3, 4, 5, 6}
X = 1 se ponto for igual a 6
X = 0 caso contrário
X = {0, 1}

b) Jogar uma moeda até tirar uma cara:
X assume a quantidade de jogadas até tirar uma cara (incluindo-se a cara) - X = {1, 2, 3, ...}
X assume a quantidade de coroas até tirar uma cara - X = {0, 1, 2, ...}

c) Uso de certo veículo durante 50 dias:

Possibilidade de uso diário (X)	Dias de uso	Probabilidade P(X)
3	3	0,06
4	7	0,14
5	12	0,24
6	14	0,28
7	10	0,20
8	4	0,08



2) Para VAC:

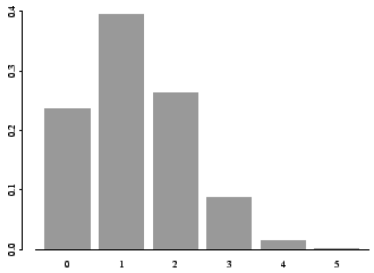
a) X distância entre dois pontos positivos - X = [0,+∞[
b) X distância entre dois pontos quaisquer - X =]-∞,+∞[

Quando nos depararmos com situações em que as variáveis aleatórias são dependentes umas das outras, ou suas distribuições de probabilidade mudam com o tempo, ou ambas as coisas acontecem; estudam-se tais situações baseando-se na <u>teoria de funções aleatórias</u>, ou seja, na <u>teoria de processos estocásticos</u>.

Os termos dos processos estocástico e aleatório são sinônimos e abrangem toda a teoria de probabilidades. Na prática, entretanto, o termo processo estocástico é reservado para quando o parâmetro temporal é introduzido.

2. DISTRIBUIÇÕES DE PROBABILIDADE DISCRETAS

2.1. DISTRIBUIÇÃO BINOMIAL



A distribuição binomial, extensão da distribuição de Bernoulli, é adequada para descrever situações em que os resultados de uma variável aleatória podem ser agrupados em apenas duas classes ou categorias. Essas categorias devem ser mutuamente excludentes, de forma que não haja dúvidas na classificação do resultado da variável nas categorias e coletivamente exaustivas, ou seja, se nenhum outro resultado for possível para o experimento em questão.

Por exemplo, certo produto, quando avaliado quanto à sua qualidade, pode-se classificá-lo por perfeito ou defeituoso; para certo questionamento a resposta pode ser verdadeira ou falsa. Assim, sabendo-se que, por exemplo, a probabilidade de sucesso em algum experimento é P(sucesso) = 0,4, a probabilidade de falha é P(falha) = 1-0,4 = 0,6.

Seguem, portanto, algumas premissas quanto aos experimentos:
- ✓ Considerar *n* repetições idênticas, onde *n* é uma constante;
- ✓ Há apenas dois resultados possíveis em cada repetição: sucesso e falha;
- ✓ As probabilidades de sucesso (p) e de falha (1-p) permanecem constantes em todas as repetições;
- ✓ As repetições devem ser independentes, ou seja, o resultado de uma repetição não é influenciado por outros resultados.

A sua função densidade de probabilidade, que representa a probabilidade *p* de certo evento ocorrer exatamente *x* vezes, em *n* repetições, ou seja, que ocorra *x* sucessos e *n – x* falhas (ou insucessos), é dada por $P(x) = \binom{n}{x} p^x (1-p)^{n-x}$ para x = 0, 1, 2,..., n, onde $\binom{n}{x}$ representa a quantidade de combinações de *n* repetições, *x* vezes, calculada por $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$.

As medidas características dessa distribuição são:

- Média ou Esperança Matemática: E(x) $= np$
- Variância: Var(x) $= np(1 - p)$
- Desvio Padrão: $\sigma = \sqrt{\text{Var(x)}}$

Exemplos:

1) Certo componente industrial é utilizado em uma máquina embaladora sabendo-se que há 30% de chance de funcionar mais de 600h. Se uma amostra de 10 componentes desses for testada, qual será a probabilidade de que, entre eles, um funcione mais de

600h e que 3 funcionem mais de 600h. Determine a esperança matemática (média), a variância e o desvio padrão.
n=10; p=0,30; 1-p=0,70

a) 1 funcione mais de 600h: x=1

$$\binom{10}{1} = \frac{10!}{1!\,(10-1)!} = \frac{3.628.800}{362.880} = 10$$

$$P(1) = \binom{10}{1}\ p^1(1-p)^{10-1} = 10 \times 0{,}30^1(0{,}70)^9 = 0{,}121 = 12{,}1\%$$

b) 3 funcionem mais de 600h: x=3

$$\binom{10}{3} = \frac{10!}{3!\,(10-3)!} = \frac{3.628.800}{6 \times 5.040} = 120$$

$$P(1) = \binom{10}{3}\ p^3(1-p)^{10-3} = 120 \times 0{,}30^3(0{,}70)^7 = 0{,}267 = 26{,}7\%$$

c) Média: $\mu = np$ = 10 x 0,30 = 3

d) Variância: $\sigma^2 = np(1 - p)$ = 10 x 0,3 x 0,7 = 2,1

d) Desvio Padrão: $\sigma = \sqrt{\sigma^2}$ = 1,45

<u>No Excel:</u> Função DISTR.BINOM (núm_s,tentativas,probabilidade_s,cumulativo)
(para Excel v.2010)

Obs.: Caso a função seja DISTRBINOM (num_s;tentativas;probabilidade_s;FALSO) a expressão para cálculo da probabilidade é a apresentada anteriormente; para DISTRBINOM (num_s;tentativas;probabilidade_s;VERDADEIRO) a expressão para probabilidade será:

$$P(x) = \sum_{i=0}^{x} \binom{n}{i}\ p^i(1-p)^{n-i}$$

2) Considere que uma moeda é lançada 5 vezes seguidas e de forma independente. Calcule a probabilidade de serem obtidas 3 caras nessas 5 repetições. Determine a esperança matemática (média), a variância e o desvio padrão.

n=5 lançamentos; p=0,50; 1-p=0,50

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{120}{12} = 10$$

$$P(3) = \binom{5}{3}\ p^3(1-p)^{5-3} = 10 \times 0{,}50^3(0{,}50)^2 = 0{,}312 = 31{,}2\%$$

c) Média: $\mu = np$ = 5 x 0,50 = 2,5

d) Variância: $\sigma^2 = np(1 - p)$ = 5 x 0,5 x 0,5 = 1,25

d) Desvio Padrão: $\sigma = \sqrt{\sigma^2}$ = 1,12

3) Um engenheiro de produção detém 15 ações em uma carteira na bolsa de valores. Supondo-se que o pregão registrou queda de 75% das ações na bolsa e que o movimento das ações pode ser representado por uma distribuição binomial, determine:
a) Qual a probabilidade que as 15 ações da carteira tenham caído?
b) Qual a probabilidade que tenham caído de preço exatamente 10 ações?
c) Qual a probabilidade que treze ou mais ações tenham caído de preço?

n=15 quedas; p=0,75; 1-p=0,25

$$P(15) = \binom{15}{15} p^{15}(1-p)^{15-15} = 10 \times 0{,}75^{15}(0{,}25)^{0} = 0{,}0134 = 1{,}34\%$$

$$P(10) = 16{,}5\%$$

$$P(X \geq 13) = P(13) + P(14) + P(15) = 23{,}6\%$$

Exercícios:

1) Os sistemas militares de radar para detecção de mísseis são concebidos para um país precaver-se de ataques inimigos. Uma questão de confiabilidade é saber se um sistema de detecção será capaz de identificar um ataque inimigo e disparar um alarme. Considere que determinado sistema de detecção tenha 90% de probabilidade de detectar um ataque de mísseis. Use a distribuição binomial para responder as questões a seguir:

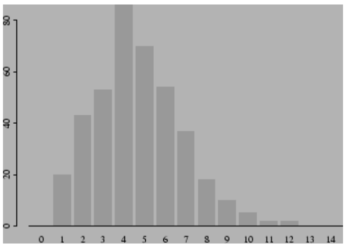
a) Qual a probabilidade de um único sistema de detecção detectar um ataque? R: 0,90
b) Se dois sistemas são instalados na área e operam de forma independente, qual é a probabilidade de pelo menos um deles detectar o ataque? R: 0,99
c) Se três sistemas ... de pelo menos um detectar? R: 0,999

2) Suponha que certa família tenha 25% de probabilidade de ter um filho (M ou F) louro, para 6 crianças, qual é a probabilidade que metade tenha cabelos louros?
n = 6, X = 3, p=25%, e 1-p=75% - R: 13%

3) A probabilidade de atingir um alvo em um único disparo de arma de fogo é de 0,3, qual é a probabilidade de que em 4 disparos o alvo seja atingido no mínimo 3 vezes?
n = 4, X ≥ 3, p=30%, e 1-p=70% - R: 8,37%

4) Um engenheiro de produção que esteja inspecionando uma área para fabricação de peças que serão utilizadas em indústrias metalúrgicas selecionou, aleatoriamente, 10 unidades como amostra em universo que imagina-se ter 20% de peças com problemas. Qual é a probabilidade de que não mais que 2 peças que sejam retiradas estejam com defeito?
n = 10, X ≤ 2, p=20%, e 1-p=80% - R: 67,78%

2.2. DISTRIBUIÇÃO DE POISSON



A distribuição de Poisson é uma aproximação da distribuição binomial, quando o número de repetições *n* é muito grande (tende ao infinito) e a probabilidade de sucesso *p* do evento, em uma tentativa, é muito pequeno (tende a zero), mantendo-se constante, finito e não nulo o produto entre *n* e *p* (média dos sucessos).

Essa distribuição pode ser utilizada para avaliar:

- ✓ Chamadas telefônicas por minuto;
- ✓ Carros que chegam ao estacionamento durante uma hora;
- ✓ Pessoas infectadas por unidade de área;
- ✓ Quantidade de peças defeituosas observadas em uma linha de produção em um determinado período de tempo;
- ✓ Quantidade de bactérias por unidade de área em certa lâmina;
- ✓ Modelagem de eventos ocorridos em um intervalo de tempo, quando os eventos ocorrem a uma taxa constante;
- ✓ Acidentes por dia.

A distribuição de Poisson é adequada para descrever eventos onde existe a probabilidade de ocorrência em um campo ou intervalo contínuo, geralmente tempo, área ou volume. Por exemplo, podem-se avaliar as quantidades de acidentes por mês, de peças defeituosas observadas em uma linha de produção em certo período de tempo e de carros que são atendidos ou que chegam a um posto de pedágio.

A sua função densidade de probabilidade para *x* ocorrências (sucessos) em um intervalo de tempo é

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

Onde x = 0,1,...; λ é a quantidade média de sucessos

As medidas características dessa distribuição são:

- Média ou Esperança Matemática: $E(x) = \lambda$
- Variância: $Var(x) = \lambda^2$

Em uma distribuição binomial quando o sucesso de certo evento é raro (*p* muito pequeno e *n* muito grande) há uma tendência para distribuição de Poisson. Na prática, considera-se essa aproximação quando $n \geq 50$ e $p \leq 0,10$. Então, neste caso, a esperança matemática - $E(X) = n \times p = \lambda$

<u>Exemplos:</u>

1) Em certo tipo de veículo ocorrem defeitos a uma taxa de 1 a cada 2000 metros. Qual a probabilidade, em 2000 metros, de que um veículo do mesmo tipo:
a) Não tenha defeitos?
b) Tenha no máximo dois defeitos?
c) Tenha pelo menos dois defeitos?

λ = 1 defeito/2000 metros
x = quant. de defeitos a cada 2000 metros

a) $P(x = 0) = \frac{e^{-1}1^{0}}{0!} = 36{,}8\%$

b) $P(x \le 2) = \frac{e^{-1}1^{0}}{0!} + \frac{e^{-1}1^{1}}{1!} + \frac{e^{-1}1^{2}}{2!} = 91{,}97\%$

c) $P(x \ge 2) = 1 - P(x \le 1) = 1 - (\frac{e^{-1}1^{0}}{0!} + \frac{e^{-1}1^{1}}{1!}) = 26{,}4\%$

2) Certa peça com formato de cubo para ser fabricada necessita de chapas de plástico de 10x10cm por face. Em média aparecem 50 defeitos por m^2 de chapa, segundo uma distribuição de Poisson. Verifique:
a) Qual a probabilidade da placa apresentar exatamente 2 defeitos?
b) Qual a probabilidade da peça apresentar no mínimo dois defeitos?

λ = 50 defeitos/m^2 = 50/10.000 defeitos/cm^2
Para chapa (face) de 10x10cm (100cm^2) >> λ = 50/10.000 defeitos/cm^2 x 100 =
0,5 defeitos/placa

x = quant. de defeitos em cada placa de 10x10cm (100cm^2)

a) $P(x = 2) = \frac{e^{-0,5}0{,}5^{2}}{2!} = 7{,}6\%$

b) λ = 0,5 defeitos/placa x 6 faces = 3 defeitos/peça

$P(x \ge 2) = 1 - P(x \le 1) = 1 - (\frac{e^{-3}3^{0}}{0!} + \frac{e^{-3}3^{1}}{1!}) = 80{,}1\%$

3) Em uma central telefônica a cada 1 hora recebem-se 2 chamadas, em média. Qual é probabilidade de, em uma hora, a central receber:
a) Nenhuma chamada.
b) Uma chamada.
c) Cinco chamadas.

λ = 2 chamadas/h

a) $P(x = 0) = \frac{e^{-2}2^{0}}{0!} = 13{,}3\%$

b) $P(x=1)=\dfrac{e^{-2}2^{1}}{1!}=27{,}1\%$

c) $P(x=5)=\dfrac{e^{-2}2^{5}}{5!}=3{,}6\%$

4) O corpo de bombeiros de um bairro recebe, em média, 3 chamados/dia. Qual a probabilidade de receber:
a) Nenhuma chamada.
b) 20 chamadas por semana.

λ = 3 chamadas/dia
x = quant. de chamados/dia

a) $P(x=0)=\dfrac{e^{-3}3^{0}}{0!}=5\%$

λ = 3 chamadas/7 dias = 21 chamadas/semana
y = quant. de chamados/semana

b) $P(y=20)=\dfrac{e^{-21}21^{20}}{20!}=8{,}7\%$

5) A probabilidade de um paciente sofrer reação alérgica por uso oral de certo medicamento é de 1%. Qual é a probabilidade de 200 pacientes, quando submetido a este remédio, não sofrer nenhuma reação alérgica?

Como $n \geq 50$ (200) e $p \leq 0{,}10$ (0,01) então E(X) = n x p = λ = 200 x 0,01 = 2, sendo assim:

$P(x=2)=\dfrac{e^{-2}2^{0}}{0!}=13{,}5\%$

<u>Exercícios</u>

1) O pessoal de inspeção de qualidade de uma fábrica que controla a quantidade de falhas na fabricação de fita adesiva plástica considera que, em média, há uma emenda a cada 50 metros fabricados. Admitindo-se que a distribuição de probabilidades da quantidade de emendas é dada por Poisson, calcule a probabilidade:
a) De nenhuma emenda acontecer em um rolo de 125 metros. R: 8,2%
b) De ocorrerem, no máximo, 2 emendas em um rolo de 125 metros. R: 20,5%
c) De ocorrer, pelo menos, uma emenda em um rolo de 100 metros. R: 86,5%

2) Um departamento de consertos de máquinas recebe-se, em média, duas chamadas para manutenção por hora. Determine as probabilidades de, em uma hora, este departamento receber nenhuma, uma, duas e nove chamadas.
a) R:13,5%   b) R: 27,1% c) R: 27,1% d) R: 0,02%

3) Uma fábrica de pneumáticos verificou que, ao testá-los nas pistas de prova, havia, em média, um estouro de pneu a cada 5.000 km. Por isso, responda:
a) Qual a probabilidade de que num teste de 3.000 km haja no máximo um pneu estourado? R: 87,8%
b) Qual a probabilidade de que um carro ande 8.000 km sem estourar nenhum pneu? R: 20,2%

4) No Japão acontece, em média, dois suicídios por ano numa população de 50.000. Em cada cidade japonesa de 100.000 habitantes, encontre a probabilidade de que em dado ano tenha havido:
a) zero suicídio. R: 1,8%
b) um suicídio. R: 7,3%
c) dois suicídios. R: 14,6%
d) dois ou mais suicídios. R: 90,8%

5) Suponha que 400 erros de impressão são distribuídos aleatoriamente em um livro de 500 páginas. Qual é probabilidade de que em uma página contenha:
a) nenhum erro. R: 44,9%
b) exatamente dois erros. R: 14,4%

6) O número médio de acidentes por mês em um determinado cruzamento no RJ é 3. Qual a probabilidade de que em um determinado mês ocorram 4 acidentes no mesmo cruzamento ? R: 16,8%

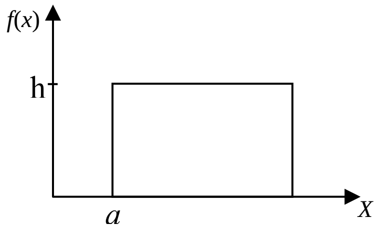
7) Se um banco espera receber, em média, 3 cheques sem fundo por dia, qual a probabilidade de num dia qualquer, receber:
a) 4 cheques sem fundo. R: 16,8%
b) no máximo 2 cheques sem fundo. R: 42,3%
c) 5 cheques sem fundo em dois dias consecutivos. R: 16,06%

8) Caminhões chegam a um depósito a razão de 2,8 caminhões/hora. Determine a probabilidade de chegarem três ou mais caminhões:
a) num período de 30 minutos. R: 16,6%
b) num período de 1 hora. R: 53,1%
c) num período de 2 horas. R: 91,8%

3. DISTRIBUIÇÕES DE PROBABILIDADE CONTÍNUAS

3.1. DISTRIBUIÇÃO UNIFORME

Usada comumente nas situações em que não há razão para atribuir probabilidades diferentes a um conjunto de valores da variável aleatória em um determinado intervalo de tempo.



Considera-se que uma variável aleatória contínua X, definida em um intervalo [a, b], tem distribuição uniforme (figura a seguir) se sua função densidade de probabilidade for especificada por f(x) = h = 1 / (b – a)

$Para\ X = [a, b] \ \Rightarrow\ P(a \le X \le b) = 1$

E(x) = (a + b) / 2
Var(x) = $(b - a)^2$ / 12
Desvio Padrão: $\sigma = \sqrt{}$ Var(x)

Exercício:

1) A dureza H de uma peça de aço pode ser pensada como uma variável aleatória com distribuição uniforme no intervalo [50,70] da escala Rockwel. Calcular a probabilidade de que uma peça tenha dureza entre 55 e 60. R: 25%

2) A distribuição da altura de plantas de Amaranthus hybridus, X, pode ser aproximada por uma distribuição normal de média 29,7 cm e desvio padrão 2,7 cm. A probabilidade de uma planta apresentar altura:
a) entre 29,7 e 32,0 cm? R:
b) acima de 32,0 cm? R:
c) abaixo de 30,0 cm? R:

3) Uma marcenaria corta toras de madeira com comprimento que variam uniformemente entre 30 cm e 90 cm. Determine:
a) A probabilidade de uma tora ter comprimento:
   a.1) maior que 80 cm; R:16,7%
   a.2) entre 65 cm e 70 cm; R: 8,3%
   a.3) exatamente 75 cm; R: 0 (não existe área – apenas um ponto)
b) Para 1200 toras cortadas, qual é a quantidade esperada com comprimento maior que 80 cm; R: 200 toras.
c) Sabendo que 90% das toras têm comprimento de k cm no máximo. Determine o valor de k. R: 84 cm

4) O tempo requerido para completar a montagem de um equipamento industrial, que segue a distribuição uniforme, pode ocorrer entre 30 a 40 minutos. Determine:

a) A probabilidade de uma montagem requerer:

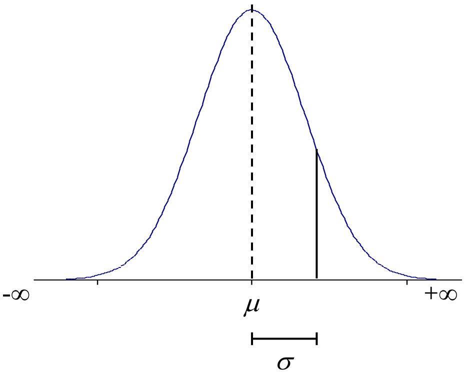
- Mais de 37 minutos para ser completado; R: 30%
- De 34 a 36 minutos; R:20%
- Exatamente 34 minutos. R: 0

b) Sabendo-se que 25% das vezes, o tempo de montagem é, no máximo, k segundos, determine o valor de k. R: 32,5 min.

c) Qual é a média e a variância do tempo de montagem. R: 35 min./8,3

3.2. DISTRIBUIÇÃO NORMAL

Também conhecida por distribuição Gaussiana, é considerada de uma família importante das distribuições contínuas de probabilidade, aplicável em muitas áreas. De forma geral, retrata bem fenômenos cujo efeito final corresponde à soma de múltiplas causas ou é afetado por diversas variáveis independentes (típico de variáveis físico químicas, socioeconômicas, psicossociais etc.).



Uma variável aleatória *X* possui uma distribuição Normal (ou Gaussiana) com média μ $(-\infty < \mu < \infty)$ e variância σ^2 $(\sigma > 0)$ se *X* possuir uma distribuição contínua com função densidade de probabilidade dada por:

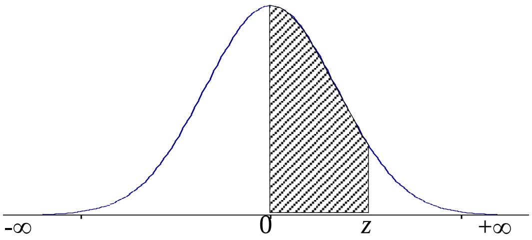
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

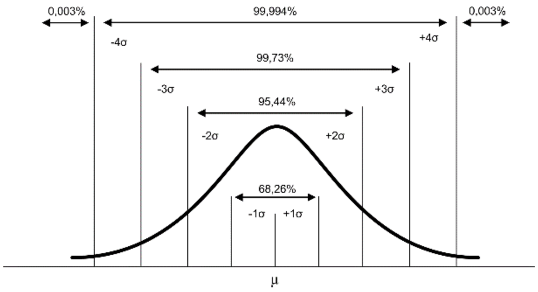
A distribuição normal com média zero ($\mu = 0$) e variância um ($\sigma^2 = 1$) é denominada distribuição normal padrão N(0,1). A função densidade de probabilidade de uma distribuição normal padrão fica da seguinte forma:

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

Onde $Z = \frac{x-\mu}{\sigma}$ (Variável Normal Reduzida)

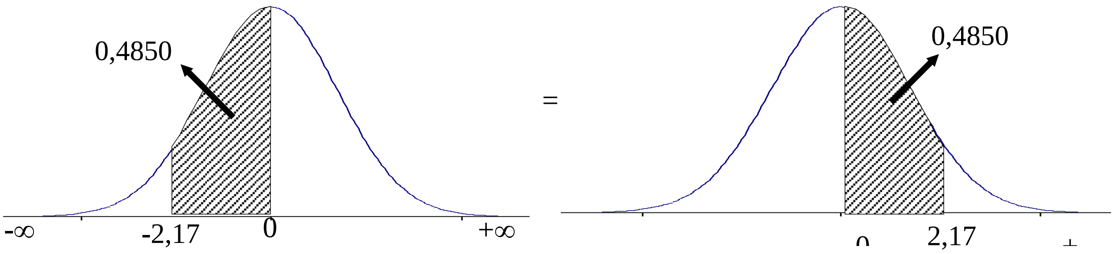
Para saber o valor da probabilidade utiliza-se a tabela da distribuição Normal (a seguir) que fornece a área acumulada até o valor de Z, ou seja, P(0 < Z < z).



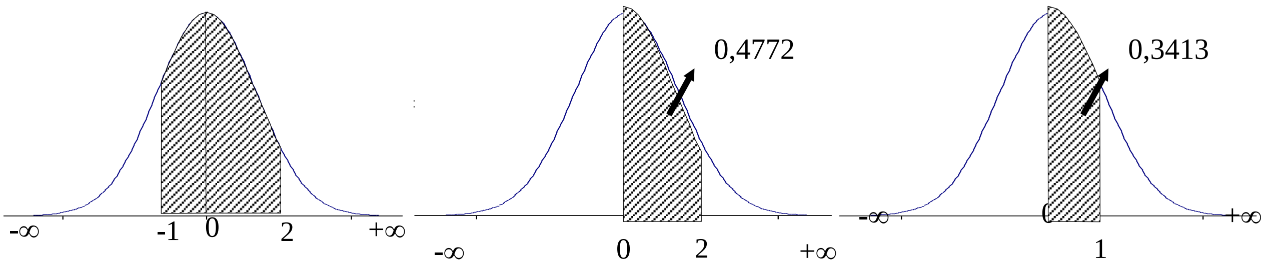


<u>Exemplos:</u>

1) $P(-2{,}17 < Z < 0) = ?$

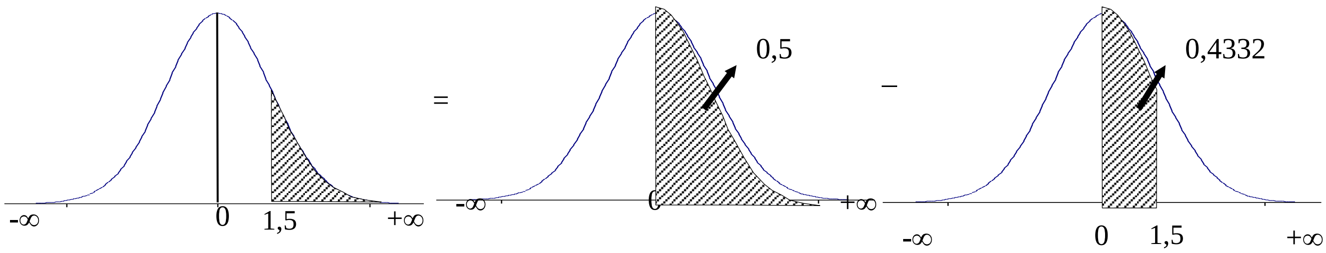


2) $P(-1 < Z < 2) = ?$



$P(-1 < Z < 2) = 0{,}4772 + 0{,}3413 = 0{,}8185$

3) $P(Z > 1{,}5) = ?$



$P(Z > 1{,}5) = 0{,}5 - 0{,}4332 = 0{,}0668$

Uma vez calculada a variável reduzida Z consulta-se a tabela da distribuição Normal padronizada para identificar a probabilidade acumulada à esquerda de Z, ou seja, a probabilidade de ocorrerem valores menores ou iguais a certo valor de Z consultado.

4) Testar para N(100,25):

a) P(100 ≤ x ≤ 106) = P(0 ≤ Z ≤ 1,2) = P(Z ≤ 1,2) − P(Z ≤ 0) = 0,8849 − 0,5000 = 0,3849

b) P(89 ≤ x ≤ 107) = P(-2,2 ≤ Z ≤ 1,4) = P(Z ≤ 1,4) − P(Z ≤ -2,2) = 0,9192 − 0,0139 = 0,9053

c) P(112 ≤ x ≤ 114) = P(2,4 ≤ Z ≤ 2,8) = P(Z ≤ 2,8) − P(Z ≤ 2,4) = 0,9974 − 0,9918 = 0,0056

d) P(x ≥ 108) = P(Z ≥ 1,6) = 1 - P(Z ≤ 1,6) = 1 – 0,9452 = 0,0548

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

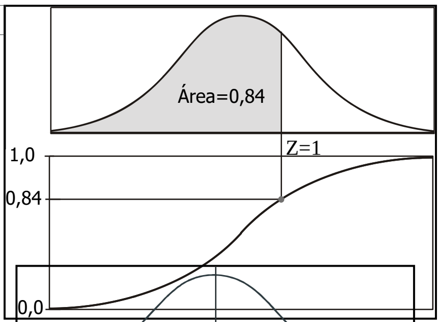
5) Considere que o peso de um rolo de arame seja normalmente distribuído com média 100 e desvio-padrão 10, ou seja, N(100,10). Então o peso (massa) está em torno de 100 variando entre ±10.

a) Calcular qual a probabilidade que um rolo, retirado ao acaso da produção, possuir peso menor ou igual a 110.

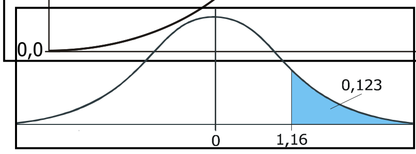
$$z = \frac{x - \mu}{\sigma} = \frac{110 - 100}{10} = 1$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

Probabilidade de ocorrência de valores abaixo de Z



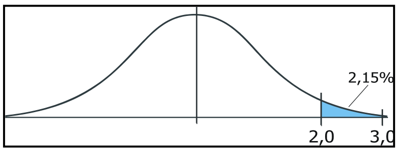
$$P(x \leq 110) = P(Z < 110) = 0{,}8413$$



b) Calcular a probabilidade do peso do rolo ser maior que 111,6?

$$z = \frac{x - \mu}{\sigma} = \frac{111{,}6 - 100}{10} = 1{,}16$$

	0.00	0.01	0.02	0.03	0 04	0.05	0.06	0 07	0.. 8	0 09
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	6 880
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0. 30	0 9309



$$P(x > 111{,}6) = 1 - P(Z \leq 111{,}6) = 1 - 0{,}8770 = 0{,}123$$

c) Calcular a probabilidade do peso do rolo estar entre 120 e 130?

$$P(120 < x < 130) = P(x < 130) - P(x > 120) = P(Z < 3) - P(Z > 2) = 0{,}9987 - 0{,}9772 = 0{,}0215$$

6) Considerando-se um grupo de indivíduos que tenha o seu peso distribuído normalmente com média 68 kg e desvio padrão 4 kg - N(68,4). Determinar a proporção de indivíduos:

a) abaixo de 66 kg; $P(X < 66) = P(Z < -0{,}5) = 0{,}3085$

b) acima de 72 kg; $P(X > 72) = P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0{,}8413 = 0{,}1587$

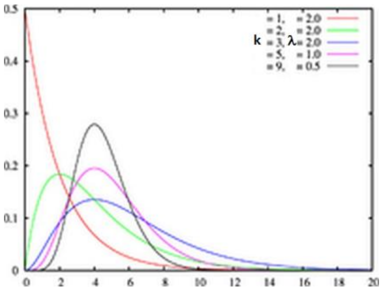
c) entre 66 e 72 kg. $P(66 < X < 72) = P(-0{,}5 < Z < 1) = P(Z < 1) - P(Z < -0{,}5) = 0{,}8413 - 0{,}3085 = 0{,}5328$

<u>Exercício:</u>

1) Uma fábrica de cimento produz sacos de 50 kg com variância de 0,25kg^2. Determine a probabilidade de que um saco selecionado aleatoriamente tenha:

a) entre 50 kg e 51 kg; R: 47,72%

b) entre 49,5 kg e 50 kg; R: 34,13%
c) entre 49 kg e 51 kg. R: 95,44%
d) acima de 51,5 kg. R: 0,13%
e) abaixo de 48,75 kg. R: 0,62%
f) entre 50,5 kg e 51,5 kg. R: 15,74%
g) entre 48,5 kg e 49,5 kg. R: 15,74%
h) abaixo de 48,5 kg ou acima de 51,5 kg. R: 0,26%
i) Em 1000 sacos saídos desta unidade de ensacamento, quantos serão esperados co m o peso entre 49,5 kg e 51,5 kg? R: 683 sacos



j) Calcule os limites, inferior e superior, do intervalo central onde existem 90% dos saco s saídos desta linha de ensacamento. R: entre 49,1775 e 50,8225.

3.3. DISTRIBUIÇÃO ERLANG

A distribuição Erlang foi desenvolvida para analisar a quantidade de chamadas telefônicas que poderiam ser feitas simultaneamente aos operadores das estações de comutação. Ela é utilizada como extensão da distribuição exponencial, especialmente quando o fenômeno aleatório é observado ao longo de diversas fases as quais podem ser descritas, de forma independente, com distribuições exponenciais. Desta forma, a soma destas *k* distribuições exponenciais de média 1/λ é uma distribuição Erlang com parâmetros 1/λ e k. A sua função densidade de probabilidade é dada por:

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

Onde x ≥ 0, λ > 0 e k > 0 inteiro.
A sua notação é Erl (k, λ), E(x) = 1/ λ e Var(x) = 1/(λ^2.k)

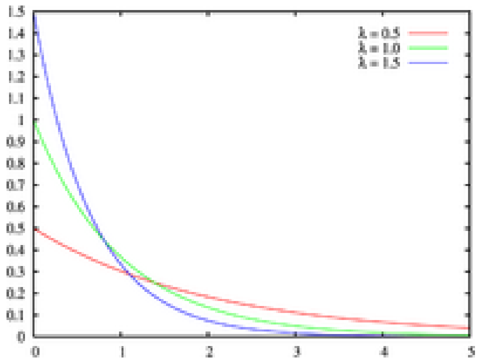
3.4. DISTRIBUIÇÃO GAMA

Esta distribuição é uma generalização da de Erlang, considerando-se que a quantidade de exponenciais somadas não precisa ser um número inteiro. Seja *X* uma variável aleatória contínua que considere somente valores não-negativos. Diz-se que *X* tem distribuição de probabilidade Gama (gráfico ao lado), se sua função de distribuição de probabilidade for dada por:

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

Onde x > 0, r ≥ 1 e α > 0.
Os parâmetros α e r são denominados, respectivamente, de escala e de forma. Cabe ainda ressaltar que a função Gama (Γ) é definida por Γ(n) = (n-1)! para n>0 (inteiro). A função Gama é uma generalização da função fatorial. Em particular Γ(0,5) = $\sqrt{\pi}$.

A distribuição de probabilidade Gama é bastante utilizada para análise de tempo de vida de equipamentos, de precipitação, de tempo de retorno de mercadorias com falhas, do tempo para falha de um sistema e em testes de confiabilidade. A esperança e a variância são E(x) = *r.α* e Var(x) = *r.α*^{2}. A notação para esta distribuição é Gama (r,α).



Existem, ainda, algumas propriedades especiais:

a) Se r = n / 2 e α = 1/2, em que *n* é inteiro e positivo, esta distribuição é denominada qui-quadrado com *n* graus de liberdade.
b) Se r = 1, $f(x) = \alpha e^{-\alpha x}$ que representa uma distribuição exponencial.

3.5. DISTRIBUIÇÃO EXPONENCIAL

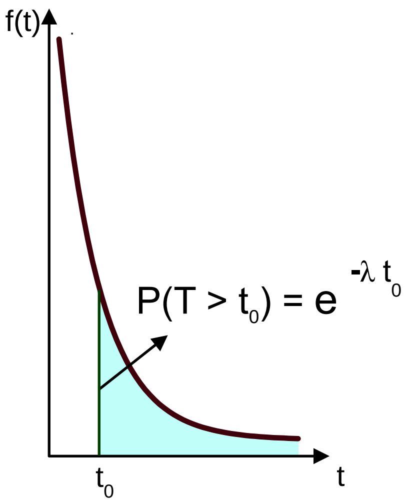
Todo fenômeno aleatório descrito por esta distribuição se caracteriza pela total imprevisibilidade e assimetria, mesmo que se conheça seu passado, por isso ela tem grande aplicabilidade em sistemas de filas. Neste caso a distribuição exponencial é muito utilizada na modelagem de tempos decorridos entre dois eventos, particularmente se estes forem causados por grande quantidade de fatores independentes.

A distribuição Exponencial também é importante na teoria da confiabilidade, pois serve para descrever as características da vida útil de certo componente, principalmente, os eletrônicos. Destaca-se ainda, para esta aplicação, que esta distribuição não tem "memória", isto é, pode ser usada para modelos de duração de vida que não desgastam com o tempo. Sendo assim, um componente novo não é mais confiável do que outro que já esteja em funcionamento.

A expressão que denota a densidade de probabilidade Exponencial, com parâmetro de distribuição λ, é dada por:

$$f(x) = \lambda e^{-\lambda x}$$

Onde x > 0 e λ > 1.



As medidas características dessa distribuição são:

- Média ou Esperança Matemática: $E(x) = 1/\lambda$
- Variância: $Var(x) = 1/\lambda^2$
- Desvio Padrão: $\sigma = \sqrt{Var(x)}$

Obs.: há forte relação entre a distribuição Exponencial e a distribuição de Poisson. Se uma variável aleatória *x* de Poisson tem média de λ ocorrências em um intervalo de tempo, então o intervalo de tempo T entre ocorrências segue uma distribuição exponencial e tem média de 1/λ.

Exemplos:

1) Uma máquina opera, em média, durante 2 horas sem necessitar de paradas para reajustes de configuração. Considerando-se que esse fenômeno possa ser representado por uma distribuição Exponencial, qual é a probabilidade dela funcionar durante 1h sem paradas?

$P(x \geq 1h) = ?$
$E(x) = 1/\lambda = 2 >> \lambda = 0,5$
$P(x \geq 1h) = e^{-0,5.1} = 0,6065 = 60,65\%$

2) O tempo de espera entre a solicitação de uma requisição ao almoxarifado de uma indústria e o atendimento é, em média, de 10 minutos. Considerando-se que esse fenômeno possa ser representado por uma distribuição Exponencial, qual é a probabilidade de um pedido exceder 10 minutos?

$P(x \geq 10 \text{ minutos}) = ?$
$E(x) = 1/\lambda = 10 >> \lambda = 0,1$
$P(x \geq 10 \text{ minutos}) = e^{-0,1.10} = 36,79\%$

3) Uma máquina falha, em média, uma vez a cada dois anos. Considerando-se que esse fenômeno possa ser representado por uma distribuição Exponencial, calcule a probabilidade da máquina falhar durante o próximo ano.

$P(x \leq 1 \text{ ano}) = ?$

E(x) = 1/ λ = 2 >> λ = 0,5
P(x ≤ 1 ano) = 1 - $e^{-0,5.1}$ = 39,35%

4) Um setor de manutenção de uma fábrica registrou que certo equipamento, em um ano, em média, teve 0,75 falha. Considerando-se que o tempo entre falhas siga distribuição Exponencial, determine a probabilidade desse equipamento não falhar no próximo ano. P(x ≥ 1 ano) = $e^{-0,75.1}$ = 47,24%

5) Um componente eletrônico importante para o funcionamento do equipamento do item 4 tem, em média, 10.000h de vida útil. Considerando-se que esse fenômeno possa ser representado por uma distribuição Exponencial, calcule a quantidade estimada de componentes que apresentarão falhas em menos de 10.000h.

P(x < 10.000 horas) = ?
E(x) = 1/ λ = 10.000 >> λ = 0,0001
P(x < 10.000 horas) = 1 - $e^{-0,0001.1}$ = 63,21%

6) Após quantas horas se espera que 25% dos componentes do item 5 tenham falhado?

E(x) = 1/ λ = 10.000 >> λ = 0,0001
P(x > t) = $e^{-0,0001.t}$ = 1 – 0,25 = 0,75
-0,0001.t = ln(0,75) >> t = 2.876,82 horas

7) A bancada de testes de qualidade dos componentes eletrônicos citados no item 5 utiliza certo tipo de bateria descarrega, em média, a cada 7 dias para demanda normal de serviço. Tome que o tempo de vida útil das baterias são distribuídas exponencialmente, determine:
a) A probabilidade de uma bateria durar pelo menos 2 semanas;
b) A probabilidade de uma bateria falhar dentro de 3 dias;
c) A probabilidade de uma bateria durar de 3 a 4 semanas.

a) 7 dias = 1 semana
E(x) = 1/ λ = 1 semana >> λ = 1
P(x ≥ 2 semanas) = $e^{-1.2}$ = 13,53%

b) E(x) = 1/ λ = 7 >> λ = 0,1428
P(x ≤ 3 dias) = 1 - $e^{-0,1428.3}$ = 1 - 65,15% = 34,85%

c) 7 dias = 1 semana
E(x) = 1/ λ = 1 semana >> λ = 1
P(3 semanas ≤ x ≤ 4 semanas) = P(x ≤ 4 semanas) - P(x ≤ 3 semanas) =

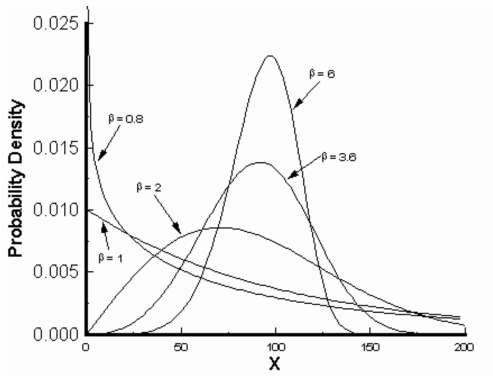
$= 1 - e^{-1.4} – (1 - e^{-1.3}) = 0,9817 – 0,9502 = 0,0315 = 3,15\%$

8) Certo tipo de fusível tem vida média de 100h e segue uma distribuição exponencial. Cada um deles tem um custo de R$10,00 e, se durar menos de 200 horas, existe um custo adicional de R$8,00. Sendo assim, determine:
a) Qual a probabilidade de um fusível durar mais de 150 horas?
b) Foi proposta a substituição do estoque por outra marca que tem o dobro de vida média, mas custa R$15,00, com o mesmo custo adicional. Verifique se é viável a substituição da marca anterior?

a) $E(x) = 1/ \lambda = 100$ horas >> $\lambda = 0,01$
$P(x > 150$ horas$) = e^{-0,01.150} = 22,31\%$

b) Custo total (CT) = Custo fusível (CF) -> $x \geq 200$ horas
Custo total (CT) = Custo fusível (CF) + Custo adicional (CA) -> $x < 200$ horas

Análise da 1ª marca:
$E(x) = 1/ \lambda = 100$ horas >> $\lambda = 0,01$
$E(CT) = CF1.P(x \geq 200) + (CF1+CA).P(x < 200)$
$E(CT) = 10,00.\ e^{-0,01.200} + (10,00 + 8,00).\ (1 - e^{-0,01.200}) = 10.0,1353 + 18.0,8647 =$
R$ 16,92



Análise da 2ª marca:
$E(x) = 1/ \lambda = 200$ horas >> $\lambda = 0,005$
$E(CT) = CF2.P(x \geq 200) + (CF2+CA).P(x < 200)$
$E(CT) = 15,00.\ e^{-0,005.200} + (15,00 + 8,00).\ (1 - e^{-0,005.200}) = 15.0,3679 + 23.0,6321 =$
R$ 20,06

Resultado: a 1ª marca é a mais econômica.

3.6. DISTRIBUIÇÃO WEIBULL

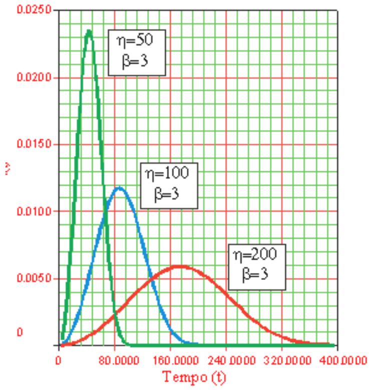
A distribuição de Weibull é uma das mais flexíveis, pois pode assumir várias outras formas. É bastante utilizada para modelagem de tempos de processo ou tempos até a falha (TPF) de componentes elétricos, componentes mecânicos, elementos estruturais e sistemas complexos. A função de densidade de probabilidade de Weibull é dada por:

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

Onde β (parâmetro de forma) > 0 e α (parâmetro de escala) > 0

No tratamento para avaliação de falhas, tomando-se uma curva de ciclo de vida (curva da banheira) genérica, ou seja, constante de três etapas, o parâmetro β pode assumir o seguinte:

- β < 1: Mortalidade Infantil
- β = 1: Falhas aleatórias
- β > 0: Falhas por desgaste (final da vida útil)



Cabe ainda observar que o valor de β pode indicar outras coisas, dependendo da análise de falhas efetuada:

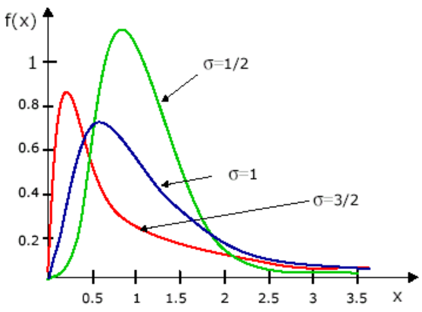
- β = 1: Indicação de modos de falhas múltiplos; suspeita de que os dados de TPF estão inadequados (geralmente para componentes com diferentes tempos de vida), indicação de a falha pode ser iniciada por agente externo ao sistema.
- β > 1: Pode ocorrer quando a amostra com os componentes testados contém alguns itens imperfeitos, acarretando a ocorrência de falhas antes de um tempo determinado em projeto.

Quanto ao parâmetro α, a sua variação, mantendo-se o β constante, influencia no seguinte:

- α crescente: a curva estica para direita, diminuindo a altura.
- α decrescente: a curva encolhe para esquerda, aumentando a altura.

As medidas características dessa distribuição são:

- Média ou Esperança Matemática: $E(x) = \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)$
- Variância: $Var(x) = \frac{\beta^2}{\alpha}\left\{2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha}\left[\Gamma\left(\frac{2}{\alpha}\right)\right]^2\right\}$
- Desvio Padrão: $\sigma = \sqrt{Var(x)}$



3.7. DISTRIBUIÇÃO LOGNORMAL

É utilizada quando o logaritmo da variável aleatória segue uma distribuição Normal. A distribuição Lognormal é bastante conhecida nas áreas atuantes do mercado financeiro e em avaliações de tempo para completar tarefas (ex. TPF) e processos com grande quantidade de valores representativos. A função de densidade de probabilidade da distribuição Lognormal é dada por:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\ln x - \mu)^2/(2\sigma^2)}$$ Para x ≥ 0.

As medidas características dessa distribuição são:

- Média ou Esperança Matemática: $E(x) = e^{\mu+\sigma^2/2}$
- Variância: $Var(x) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$
- Desvio Padrão: $\sigma = \sqrt{Var(x)}$

4. AMOSTRAGEM E DISTRIBUIÇÕES AMOSTRAIS

A teoria da amostragem é um estudo das relações existentes entre uma população e as amostras dela extraídas. É por meio das amostras que se pode, por exemplo, avaliar grandezas desconhecidas da população (parâmetros), tais como a sua média, variância etc., por intermédio das correspondentes grandezas amostrais, denominadas de estatísticas amostrais.

Geralmente, os parâmetros são expressos por:

Parâmetros	População	Amostra	Expressões Para Amostra
Tamanho	N	n	-
Média Aritmética	μ	$\overline{x}$	$\frac{\sum x}{n}$
Variância Absoluta	σ^2	s^2	$\frac{\sum\left(x-\overline{X}\right)^2}{n-1}$
Desvio Padrão	σ	s	$\sqrt{S^2}$
Proporção	π	p	-

Uma amostra é um subconjunto finito de elementos extraído de uma população. Para que seja representativa a amostra tem que apresentar a capacidade de reproduzir as mesmas características importantes da população de origem.

Uma amostra é formada por processo de seleção dos elementos da população e pode ser obtida pelos métodos probabilísticos e não-probabilísticos.

Os métodos não-probabilísticos são aqueles onde há escolha deliberada dos elementos que irão compor a amostra. Com estas amostras não é possível generalizar os resultados da pesquisa, de vez que elas não garantem a representatividade da população.

Os métodos probabilísticos são aqueles onde a amostragem é caracterizada por cada elemento da população possuir a mesma probabilidade de ser escolhido para compor a amostra. Desta forma, se conhece a distribuição de probabilidade de todas as combinações amostrais, viabilizando a determinação da variabilidade amostral, o que por sua vez permite estimar o erro amostral, garantindo assim a cientificidade do método. Este método é caracterizado por amostragens aleatórias simples, sistemática, estratificada e por conglomerado.

Uma amostragem aleatória simples é um dos métodos probabilísticos mais básicos. Os elementos da amostra são rotulados individualmente, sendo objeto de sorteio com base nesses rótulos.

Na amostragem sistemática, indicada quando a população esteja ordenada segundo algum critério, tal como as chegadas de caminhões ordenadas por horário de chegada, determina-se o intervalo de periodicidade da amostragem. Posteriormente, seleciona-se a ordem do primeiro elemento da amostra por sorteio dentro do intervalo de periodicidade. A seguir, repete-se a ordem de seleção dentro de cada intervalo de periodicidade.

Para população heterogênea formada por grupos homogêneos, aos quais denominamos de estratos, é recomendado que se utilize a amostragem estratificada. Atributos de estratificação comuns são sexo, idade, classe social, profissão etc.. Uma vez determinados os estratos, retira-se uma amostra aleatória de cada estrato, de tamanho proporcional à participação de cada estrato na população.

A amostragem por conglomerado é utilizada quando é difícil ou até impossível identificar todos os elementos da população, mas é possível identificar facilmente os grupos que apresentem as mesmas características da população. Neste caso, extrai-se uma amostra aleatória destes grupos, denominados conglomerados, e amostra-se os elementos do conglomerado. Utilizam-se estas amostragens em pesquisa de população de uma cidade, quando se pode sortear quarteirões e contar todos os moradores de cada quarteirão.

Geralmente, as pesquisas são conduzidas pela análise dos elementos de uma amostra extraída de uma população que se deseja estudar. Essa amostra depende do seu tamanho (quantidade de elementos) e dos aspectos metodológicos que devem nortear a extração dos elementos da população.

Quanto à quantidade de elementos que a amostra deve ter, deve-se considerar o nível de confiança e a margem de erro que se pretenda para os resultados. Isso constitui a precisão de uma estimativa. O cálculo do tamanho da amostra para cada caso, é feito sempre em função dos Intervalos de Confiança correspondentes.

O intervalo de confiança para a média, para uma população infinita e a variância da população conhecida, é dado por:

$$P\left(\bar{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

O erro padrão da estimativa, ou seja, em quanto a média populacional pode diferir da média amostral, é dado por:

$$e = z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

Sendo assim, a quantidade de elementos de uma amostra com população infinita é dada por:

$$n_0 = \left(\frac{z_{\alpha/2} \cdot \sigma(x)}{e}\right)^2$$

Para a variância da população desconhecida deve-se substituir o desvio padrão populacional (σ_x) utilizado anteriormente, por uma estimativa (desvio padrão amostral s_x), calculado numa amostra piloto de n_1 elementos. Além disso, substitui-se a distribuição normal, com coeficiente de confiança Z pela distribuição t (STUDENT) com n_1 -1 graus de liberdade.

A quantidade de elementos de uma amostra com população infinita é dada por:

$$n = \left(\frac{t_{\alpha/2} \times s(x)}{e} \right)^2$$

Observa-se, entretanto, que se o valor de *n* calculado for menor que n1, então a amostra piloto com n1 elementos já satisfaz a precisão desejada. Se o valor de n calculado for maior que n1, então se deve complementar a amostra piloto com mais (n – n1) elementos.

4.1. DISTRIBUIÇÃO T-STUDENT

O uso da distribuição de Student, ou simplesmente distribuição t, está associado a estudos com pequenas amostras (n<30). É um modelo de distribuição contínua que se assemelha à distribuição normal padrão, ou seja, N(0;1), mas reflete a maior variabilidade (com curvas mais alargadas) que é de se esperar em amostras pequenas. Uma variável aleatória contínua X tem distribuição t de Student com *gl* graus de liberdade, se sua função densidade de probabilidade for dada por:

$$f(x) \frac{\Gamma\left(\frac{gl+1}{2}\right)}{\sqrt{gl.\pi}.\Gamma\left(\frac{gl}{2}\right)} \left(1 + \frac{x^2}{gl}\right)^{-\left(\frac{gl+1}{2}\right)}$$

Onde gl > 0 e x > 0.

As medidas características dessa distribuição são:

- Média ou Esperança Matemática: $E(x) = 0$

- Variância: $Var(x) = \frac{gl}{gl-2}$

- Desvio Padrão: $\sigma = \sqrt{}$ Var(x)

4.2. DISTRIBUIÇÃO QUI-QUADRADO

Considere *X* variáveis aleatórias independentes, não-negativas, dependente do grau de liberdade (gl), tem-se a sua função de distribuição de probabilidade dada por:

$$f(x) = \frac{e^{-\frac{x}{2}} x^{(gl/2)-1}}{2^{gl/2} \Gamma\left(\frac{n}{2}\right)}$$

Onde gl > 0 e x > 0.

As medidas características dessa distribuição são:

$$E(x) = gl$$

- Média ou Esperança Matemática:

- Variância: $Var(x) = 2gl$

- Desvio Padrão: $\sigma = \sqrt{Var(x)}$

4.3. DISTRIBUIÇÃO F (DE SNEDECOR) - FISHER-SNEDECOR

Ela depende de dois parâmetros denominados também de graus de liberdade. O primeiro (m) é o grau de liberdade do numerador e o segundo (n) do denominador. Na estatística ela é caracterizada como o quociente de duas variâncias e, portanto de duas distribuições qui-quadrado. Cada parâmetro, da mesma forma que nos modelos anteriores, é associado ao tamanho amostral menos um. A sua função de distribuição de probabilidade é dada por:

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(mx+n)^{\frac{m+n}{2}}}$$

Onde m,n > 0 e x ≥ 0.

As medidas características dessa distribuição são:

- Média ou Esperança Matemática: $E(x) = \frac{n}{n-2}$ para n>2

- Variância: $Var(x) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ para n>4

- Desvio Padrão: $\sigma = \sqrt{Var(x)}$

5. TEOREMA DO LIMITE CENTRAL

A possibilidade da utilização de amostras para fazer inferências sobre parâmetros populacionais (estimação) depende do conhecimento do tipo de distribuição amostral. Para se obter a distribuição amostral faz-se necessário repetir *n* vezes um experimento e, após, calcular a média das amostras, permitindo-se obter a distribuição amostral.

Em uma estimação considera-se como estimativa pontual a estimativa de um único valor para um parâmetro populacional. Essa estimativa não permite avaliar a precisão do parâmetro. A estimativa pontual menos enviesada (menos não-representativa da população) da média populacional μ é a média amostral $\overline{X}$.

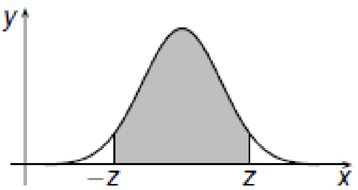
Uma estimativa intervalar é um intervalo de valores usado para estimar um parâmetro populacional (μ, σ etc.) com certo nível de confiança (1-α). O nível de confiança é a probabilidade de que o intervalo estimado contenha o parâmetro populacional (μ). Com isso é possível determinar o erro máximo cometido na estimação, com certa confiança.

Assim, P (L1 < μ < L2) = K significa que a probabilidade do intervalo aleatório (L1, L2) conter o valor exato μ é K. O intervalo (L1, L2) é denominado <u>intervalo de confiança</u> para o parâmetro populacional μ, com um nível de confiança K.

A distribuição normal pode ser utilizada sempre que se tiver uma das seguintes situações:

1ª - se n ≥ 30, conforme o Teorema do Limite Central.
2ª - se n < 30, sendo a população estudada normalmente distribuída e o desvio padrão populacional σ conhecido.

Nessa situação, o <u>nível de confiança</u> é a área sob a curva normal padrão entre os valores críticos -Z e +Z, sendo Z definido como <u>coeficiente de confiança</u> (fig. ao lado).

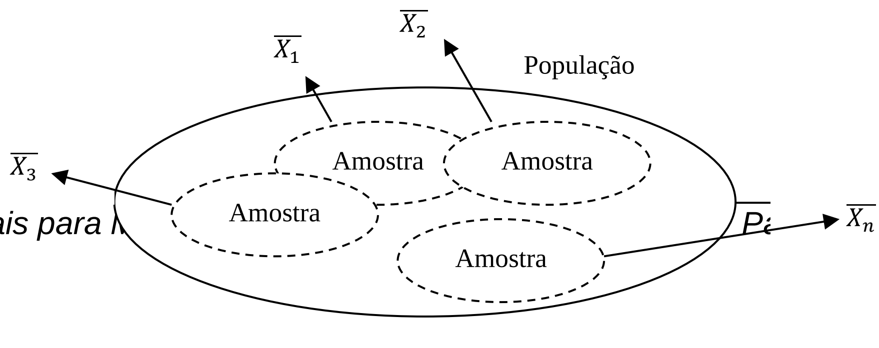


Os valores de Z mais utilizados são:

Nível de confiança (1 - α)	**Z**
0,80	1,28
0,90	1,64
0,95	1,96
0,99	2,58

O Teorema do Limite Central considera que na medida em que o tamanho da amostra aumenta, a distribuição amostral das médias amostrais tende para uma distribuição normal. Mesmo no caso de uma distribuição não-normal, a distribuição das médias amostrais será aproximadamente normal, desde que a amostra seja grande. Isto quer dizer que não é necessário conhecer a distribuição de uma população para que seja possível fazer inferências sobre ela a partir de dados amostrais.

Sendo assim, supondo-se que certa variável aleatória *x* tenha, ou não, o comportamento da distribuição normal, e que a média dos valores *x* seja μ e o desvio-padrão seja σ e que se coletem dados para compor amostras de tamanho *n*, pode-se calcular as médias amostrais. O Teorema do Limite Central caracteriza que na medida em que o tamanho *n* de amostras aumenta (na prática, a distribuição de amostragem da média pode se considerada como normal sempre que *n* ≥ 30), a distribuição amostral das médias amostrais tente para uma distribuição normal com média μ e desvio-padrão $\sigma / \sqrt{n}$.



Observar a situação de um dado não viciado em certo experimento. Sendo *x* o lançamento de um dado, tem-se f(x) = 1/6 (0,167) $\forall$ x. Agora considere *x* a média de lançamento de dois dados, obteve-se N (3,5; $1,58^2$). Tomando-se ainda os mesmos dois dados, mas testando-os 36 vezes obtêm a seguinte distribuição das médias:

Média dos valores dos dados	Frequência
1,0	1
1,5	2
2,0	3
2,5	4
3,0	5
3,5	6
4,0	5
4,5	4
5,0	3
5,5	2
6,0	1

Observar que a frequência se distribui da forma normal após vários testes.

<u>**Exemplo 1:**</u> Certo equipamento automático encontra-se regulado para encher embalagens de um quilo de certo produto. Quando ele está desregulado pode provocar os seguintes problemas: caso as embalagens tenham massa inferior ao estabelecido, haverá reclamações dos clientes; no sentido contrário se estará permitindo disponibilizar mais do que está especificado, promovendo aumento de custo no processo. A experiência da empresa mostra que a massa das embalagens se comporta normalmente com desvio padrão de 12 gramas. Para verificar a precisão do equipamento, selecionaram-se em determinada altura, nove embalagens com as seguintes medições de massa: 983/992/1011/976/997/1000/1004/983/998. Determine a estimativa pontual (µ) e o intervalo de confiança para 90%, 95% e 99%.

a) µ = 993,78
b)
   - Para 90%: 1 – 0,9 = 0,1 / 2 = 0,05 >> ± 1,64.12/$\sqrt{9}$
     = ±1,64.4 = ±6,56
     P(993,78 - 6,56 < μ < 993,78 + 6,56) = 0,95
     P(987,22 < μ < 1000,34) = 0,95
   - Para 95%: (985.94, 1001.62);

- Para 99%: (983.476, 1004.084).

<u>Exemplo 2:</u> Medições do comprimento de agulhas realizadas numa amostra de 100 unidades permitiram calcular o seguinte: μ=3,1cm; σ=0,7cm. O investigador pretende estabelecer a probabilidade dessa média ser adequada para representar a população com um intervalo de confiança de 95%.

Nestas condições da amostra é possível utilizar o Teorema do Limite Central. Sendo assim, para o intervalo de confiança de 95% tem-se 5%÷2, determinando os limites L_1 e L_2, ou seja 2,5%. Observando-se a tabela a eguir, chega-se a:

P($-Z_{0,025}$ < Z amostra < $Z_{0,025}$) = 0,95
P(-1,96 < Z amostra < 1,96) = 0,95

$$Z = \frac{\bar{x} - \mu}{\sigma}$$

Considerando-se que o Teorema do Limite Central segue que a distribuição amostral das médias amostrais tente para uma distribuição normal com média μ e desvio-padrão chega-se a $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

P(-1,96. $\sigma / \sqrt{n}$ < $\bar{x} - \mu$ < 1,96. $\sigma / \sqrt{n}$) = 0,95
P($\bar{x}$ -1,96. $\sigma / \sqrt{n}$ < μ < $\bar{x}$ +1,96. $\sigma / \sqrt{n}$) = 0,95
P(3,1-1,96. $0{,}7 / \sqrt{100}$ < μ < 3,1+1,96. $0{,}7 / \sqrt{100}$) = 0,95
P(2,96 < μ < 3,24) = 0,95

$P(Z<z)$

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Desta forma, ao se selecionar amostras com tamanho *n*=100, tem-se 95% de confiança de que a média estará entre 2,96 e 3,24.

Para intervalos de confiança de 90% e 99% chega-se a:

$P(-Z_{0,050} < Z \text{ amostra} < Z_{0,050}) = 0,90$, onde $Z_{0,050} = 1,645$;

$P(-Z_{0,005} < Z \text{ amostra} < Z_{0,005}) = 0,99$, onde $Z_{0,005} = 2,575$.

<u>Exemplo 3:</u> Numa amostra de 64 pessoas foi perguntado o peso de cada uma delas. A média amostral obtida foi 50 Kg com desvio padrão do peso de 16Kg. Pede-se estimar o valor da média da população para um intervalo de confiança de 95%.

O coeficiente de confiança (Z) para 95% de confiança é 1,96. Sendo assim:

$P(50-1,96.\ 16/\sqrt{64} \ < \ \mu < 50+1,96.\ 16/\sqrt{64}\ ) = 0,95$

$P(46{,}08 < \mu < 53{,}92) = 0{,}95$

Exemplo 4: 100 domicílios, selecionados ao acaso, foram entrevistados para saber a quantidade de pessoas residiam. Obteve-se média amostral de 5 pessoas. O desvio padrão da amostra foi de 4 pessoas. Estime a média de pessoas por domicílio na cidade como um todo, para um intervalo de confiança de 80%, 90%, 95%, 99%.

Limites da μ: $\bar{x} \pm Z.\ \sigma / \sqrt{n}$
Para 80%: Z=1,28; 5 - (1,28 x 4 / 10) = 4,488; 5 + (1,28 x 4 / 10) = 5,512
Para 90%: 4,34; 5,65.
Para 95%: 4,21; 5,78.
Para 99%: 3,96; 6,03.

Quando o desvio padrão da população (σ) não for conhecido e o tamanho da amostra for inferior a 30 (n < 30), sendo a população estudada normalmente distribuída, a distribuição normal não é apropriada para determinação de intervalos de confiança para a média. Nesse caso usamos a distribuição de Student (*t*) para determinação desse intervalo.

A distribuição de Student tem a forma semelhante a da distribuição Normal, e a principal diferença é que a distribuição *t* tem maior área nas caudas. Então, para um dado nível de confiança, o valor de *t* será um pouco maior que o correspondente valor de z. A determinação do valor de *t* na tabela depende no nível de confiança e do nº do grau de liberdade.

O grau de liberdade (*gl*) é definido como a quantidade de observações independentes da amostra subtraindo-se o número dos parâmetros populacionais que devem ser estimados por meio das observações amostrais. Quanto maior o grau de liberdade, mas a distribuição se aproxima da normal. É uma medida de credibilidade ou segurança em cada estimativa do parâmetro de cada fonte de incerteza considerada no cálculo.

$$\boxed{gl = n - 1}$$

Por exemplo, em uma distribuição retangular, que é fechada em um intervalo [a;b] definido, pode-se dizer que há alta segurança na estimativa do valor verdadeiro (pois há 100% de probabilidade de que o valor esteja dentro desse intervalo). Por isso, atribui-se um alto grau de liberdade (no caso da retangular, infinito). Esse é apenas um exemplo para se ter a noção do grau de liberdade, já que a distribuição retangular não é caracterizada pelo grau de liberdade.

Exemplo 5: Em um laboratório testou-se um produto novo em 20 cobaias, com idade compreendida entre 56 e 84 dias de idade, e se obteve a massa média de 200 g com um desvio-padrão de 26 g. O investigador pretende avaliar a média para um índice de confiança de 99%.

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

gl = 20 – 1 = 19

Degrees of Freedom	Probability of a Larger Value, Sign Ignored								
	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657		
2	0.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	1.330	1.734	2.101	2,445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850
21	.686	.859	1.323	1.721	2.080	2.414	2.831	3.135	3.819
22	.686	.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	.685	.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	.685	.857	1.318	1.711	2.064	2.391	2.797	3.090	3.745
25	.684	.856	1.316	1.708	2.060	2.385	2.787	3.078	3.725
26	.684	.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	.684	.855	1.314	1.703	2.052	2.373	2.771	3.056	3.690
28	.683	.855	1.313	1.701	2.048	2.368	2.763	3.047	3.674
29	.683	.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659
30	.683	.854	1.310	1.697	2.042	2.360	2.750	3.030	3.646
35	.682	.852	1.306	1.690	2.030	2.342	2.724	2.996	3.591
40	.681	.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	.680	.850	1.301	1.680	2.014	2.319	2.690	2.952	3.520
50	.680	.849	1.299	1.676	2.008	2.310	2.678	2.937	3.496
55	.679	.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	.679	.848	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	.678	.847	1.294	1.667	1.994	2.290	2.648	2.899	3.435
80	.678	.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	.678	.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	.677	.846	1.290	1.661	1.982	2.276	2.625	2.871	3.390
120	.677	.845	1.289	1.658	1.980	2.270	2.617	2.860	3.373
∞	.6745	.8416	1.2816	1.6448	1.9600	2.2414	2.5758	2.8070	3.2905

P($-t_{0,005}$ < t amostra < $t_{0,005}$) = 0,99, onde $t_{0,005}$ = 2,861.
P(-2,861 < t amostra < 2,861) = 0,99.
P(200-2,861. $26/\sqrt{20}$ < μ < 200+2,861. $26/\sqrt{20}$) = 0,99
P(183,37 < μ < 216,63) = 0,99

Ao se selecionar da população das cobaias sucessivas amostras aleatórias de tamanho *n*=20 e se calcular a média para um intervalo de confiança de 99% a partir de cada uma delas, é de esperar que 99% destes intervalos contenham a média populacional μ .

Exemplo 6: Coletou-se uma amostra com 16 pacientes para se verificar a existência de certa enzima no sangue. Obteve-se média de 13mg para cada 100ml de sangue e desvio padrão 4,6mg/100ml. Pede-se para se estimar, por intermédio do intervalo de confiança, com grau de confiança de 95%, a média de enzimas no sangue.

Gl=16-1=15

IC[μ, 95%] = $13 \pm t_{(15;0,025)} (4,6/\sqrt{16}) = 13 \pm 2,131 \times 1,15 = [10,55;15,45]$

6. REGRESSÃO LINEAR SIMPLES

Análise de regressão é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis de tal forma que uma variável pode ser predita a partir da outra ou outras. Por isso, pode-se dizer que a analise de regressão estuda o relacionamento entre uma variável chamada dependente e outras variáveis denominadas independentes. Neste caso caracteriza-se esta relação como Regressão Linear Múltipla.

O caso mais simples de regressão é quando temos duas variáveis e a relação entre elas pode ser representada por uma linha reta. Esta metodologia é assim denominada Regressão Linear Simples (RLS). Nesta situação define-se a relação entre uma variável dependente e outra independente.

Para RLS pode-se efetuar a Análise de Correlação que permite inferir, estatisticamente, as medidas de associação entre as duas variáveis.

Os dados para a análise de regressão e correlação, análises que estão intimamente ligadas são da seguinte forma: (x1, y1), (x2, y2), . . . ,(xi, yi), . . . , (xn, yn). A partir deles constrói-se um Diagrama de Dispersão que permite decidir, empiricamente, se um há relacionamento linear entre as variáveis e, ainda, se há um relacionamento “forte” ou “fraco” entre elas.

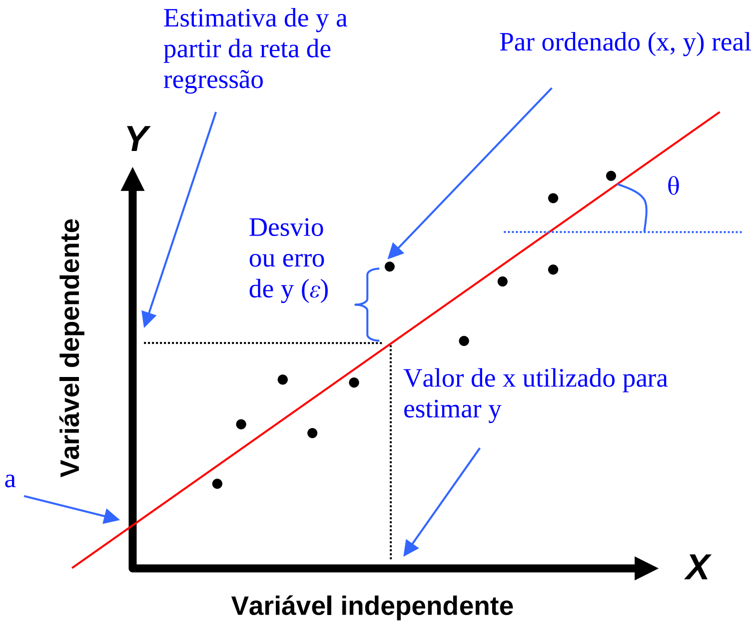
Para se determinar a equação da reta que relacione as variáveis depende e independente pode-se usar o método dos mínimos quadrados. Este método se baseia em encontrar os coeficientes angular e linear da reta de regressão que minimizem a soma dos quadrados dos desvios. Cabe ainda considerar o seguinte:

- ✓ A soma dos desvios verticais dos pontos em relação à reta é zero;
- ✓ A soma dos quadrados desses desvios é mínima (isto é, nenhuma outra reta daria menor soma de quadrados de tais desvios).

A equação da reta de regressão é do tipo $Y = b + aX + \varepsilon$, onde:

- ✓ Y é a variável dependente;
- ✓ X é a variável independente
- ✓ ε são os desvios de Y em relação ao valor esperado;
- ✓ b é o coeficiente linear, ou seja, é o ponto onde a reta de regressão intercepta a ordenada (o valor de Y quando X = 0) e;
- ✓ a é o coeficiente angular (tg θ).

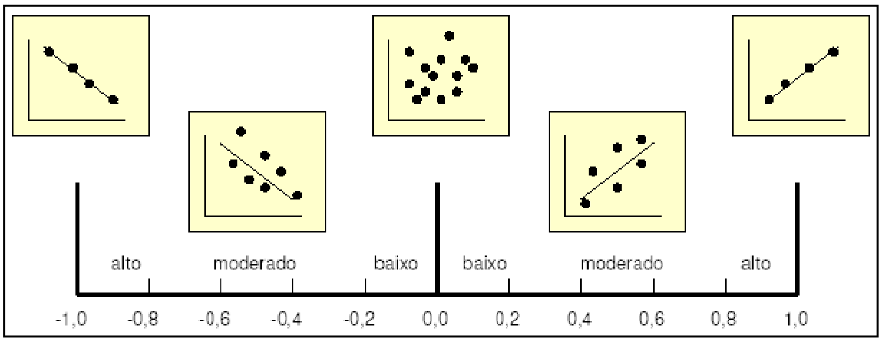
Sendo assim, deseja-se ajustar a reta estimando-se os coeficientes a e b. A figura a seguir apresenta este ajustamento, bem como as características das variáveis dependente e independente.



Para se calcular os coeficientes linear (b) e angular (a) pelo Método dos Mínimos Quadrados utilizam-se as seguintes expressões:

$$a = \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sum X^2 - n\overline{X}^2} \qquad b = \overline{Y} - a\overline{X}$$

Para saber o grau de relacionamento entre as variáveis dependente e independente utiliza-se o Coeficiente de Correlação de Pearson (r). Ele indica o grau em que uma equação linear descreve a relação entre essas duas variáveis. Varia entre -1 a 1, e assume valor negativo quando as variáveis são inversamente proporcionais e, positivo quando diretamente proporcionais. Assume valor zero quando não há relação entre as duas variáveis. A figura adiante facilita a visualização desses valores.



O Coeficiente de Correlação é calculado pela seguinte expressão:

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

Exemplo

Período	Y	X
1	264	2,5
2	116	1,3
3	165	1,4
4	101	1,0
5	209	2,0

Período	Y	X	XY	X^2	Y^2
1	264	2,5	660,00	6,25	69.696
2	116	1,3	150,80	1,69	13.456
3	165	1,4	231,00	1,96	27.225
4	101	1,0	101,00	1,00	10.201
5	209	2,0	418,00	4,00	43.681
Total	855	8,2	1560,80	14,90	164.259
Média	171	1,64			

$$a = \frac{1560{,}80 - 5 \times (1{,}64) \times (171)}{14{,}90 - 5 \times (1{,}64)^2} = 109{,}23 \quad b = 171 - 109{,}23 \times (1{,}64) = -8{,}37$$

Sendo assim, $Y = -8{,}37 + 109{,}23X$ com r = 0,98.

BIBLIOGRAFIA

Agnaldo, José **Estatística Básica** – Un. 05 - Distribuição de Probabilidade, utilização de exercícios, capturado de http://pt.scribd.com/joanathibau/d/53696690/14-Distribuicao-Uniforme, disponível em 23/05/2012.

Bertolo, Luiz Antonio **Distribuição Contínua** utilização de exercícios, capturado de http://www.bertolo.pro.br/FinEst/Estatistica/DistribuicaoContinua.pdf, disponível em 23/05/2012.

Farias, Ana Maria Lima, **Inferência Estatística** Universidade Federal Fluminense - Instituto de Matemática – Departamento de Estatística, Apontamentos da Disciplina, Niterói, 2008.

Fernandes, Edite Manuela da G.P. **Estatística Aplicada,** Universidade do Minho, Braga, Portugal, 1999.

Leal, Simone **Modelação Matemática uma Proposta Metodológica para o Curso de Economia** Dissertação de Mestrado em Engenharia de Produção, UFSC, Capturado de http://www.eps.ufsc.br/disserta99/leal/, Disponível em 08/05/2012, 1999, Florianópolis.

Reboita, Michelle S. **Introdução à Estatística Aplicada à Climatologia**, Parte II – Distribuições de Probabilidades, Projeto PAE, São Paulo, 2005.

Rennó, Camilo Daleles **Estatística: Aplicação ao Sensoriamento Remoto** Instituto Nacional de Pesquisas Espaciais, utilização de exercícios, capturado de http://www.dpi.inpe.br/~camilo/estatistica/, disponível em 23/05/2012, 2010.

Sérgio, Carlos **Estatística Básica** – Un. 04 - Distribuições Contínuas, Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia Agroalimentar - Campus Pombal, utilização de exercícios, capturado de http://www.ccta.ufcg.edu.br/admin.files.action.php?action=download&id=886, disponível em 23/05/2012.

Souto, Marcílio **Estatística Aplicada à Indústria** utilização de exercícios, capturado de www.cin.ufpe.br/~mcps/AM/distribuicao-prob.ppt, disponível em 23/05/2012.