Departamento de Engenharia de produção
EESC Escola de Engenharia de São Carlos
USP
Grupo de Pesquisa em Gestão da Qualidade e Mudança
Research Group on Quality and Change Management

Métodos Estatísticos para a Melhoria da Qualidade

Disciplina: SEP-280
Controle da Qualidade de Processos de Fabricação

Research Group Leaders:
Luiz C. R. Carpinetti, Associate Professor
Mateus C. Gerolamo, Assistant Professor

- Estatística é um conjunto de técnicas úteis para a tomada de decisão sobre um processo ou população.

- Estatística é a língua na qual engenheiros, operários, compradores, administradores e outros integrantes da empresa se comunicam sobre qualidade.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 25

Inferência sobre a Qualidade do Processo

Inferência sobre a Qualidade do Processo

- Até o momento, supusemos que os parâmetros da distribuição de probabilidade e, por conseguinte, os parâmetros do processo, eram conhecidos.
- Esta é uma suposição, em geral, bastante irrealista!
- Em geral, os parâmetros de um processo são desconhecidos; além disso, eles podem mudar ao longo do tempo.
- Assim, é necessário desenvolver procedimentos para estimar os parâmetros das distribuições de probabilidade e para resolver outros problemas de inferência ou decisão relativos a eles.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência sobre a Qualidade do Processo

- As técnicas estatísticas tradicionais de estimação de parâmetros e de teste de hipóteses são úteis neste contexto.

- Tais técnicas são a base subjacente para a grande parte da metodologia do controle estatístico da qualidade.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência sobre a Qualidade do Processo

- As técnicas estatísticas tradicionais de estimação de parâmetros e de teste de hipóteses são úteis neste contexto.
- Tais técnicas são a base subjacente para a grande parte da metodologia do controle estatístico da qualidade.
- Nesta aula, apresentamos alguns resultados elementares da inferência estatística, indicando sua utilidade nos problemas referentes à melhoria da qualidade.
- Os principais tópicos incluem estimação pontual e intervalar das médias, variâncias e parâmetros da binomial, teste de hipóteses sobre médias, variâncias e parâmetros da binomial e o uso de gráficos da probabilidade normal.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência sobre a Qualidade do Processo

ESTATÍSTICAS E DISTRIBUIÇÕES AMOSTRAIS

Estatísticas e Distribuições Amostrais

- O objetivo da **inferência estatística** é a obtenção de resultados ou a tomada de decisão sobre uma população baseada em uma amostra selecionada desta população.
- Iremos supor, frequentemente, que *amostras aleatórias* são usadas na análise.
- Em geral, aplica-se a palavra “aleatória” a qualquer método de seleção de amostra que careça de direcionamento sistemático.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Estatísticas e Distribuições Amostrais

- Vamos definir uma amostra, digamos $x_1, x_2, \ldots, x_n$, como uma **amostra aleatória** de tamanho *n* se ela for selecionada de tal modo que as observações $\{x_i\}$ sejam independentes e identicamente distribuídas.
- Esta definição é apropriada para amostras aleatórias retiradas de populações infinitas ou de populações finitas quando a amostragem é feita *com reposição*.
- Na amostragem *sem reposição*, diremos que uma amostra de *n* itens, retirada de uma população finita de *N* itens, é uma amostra aleatória, se cada uma das $\binom{N}{n}$ amostras possíveis tiver igual probabilidade de ser escolhida.

Estatísticas e Distribuições Amostrais

- A Figura a seguir ilustra a relação entre população e amostra.

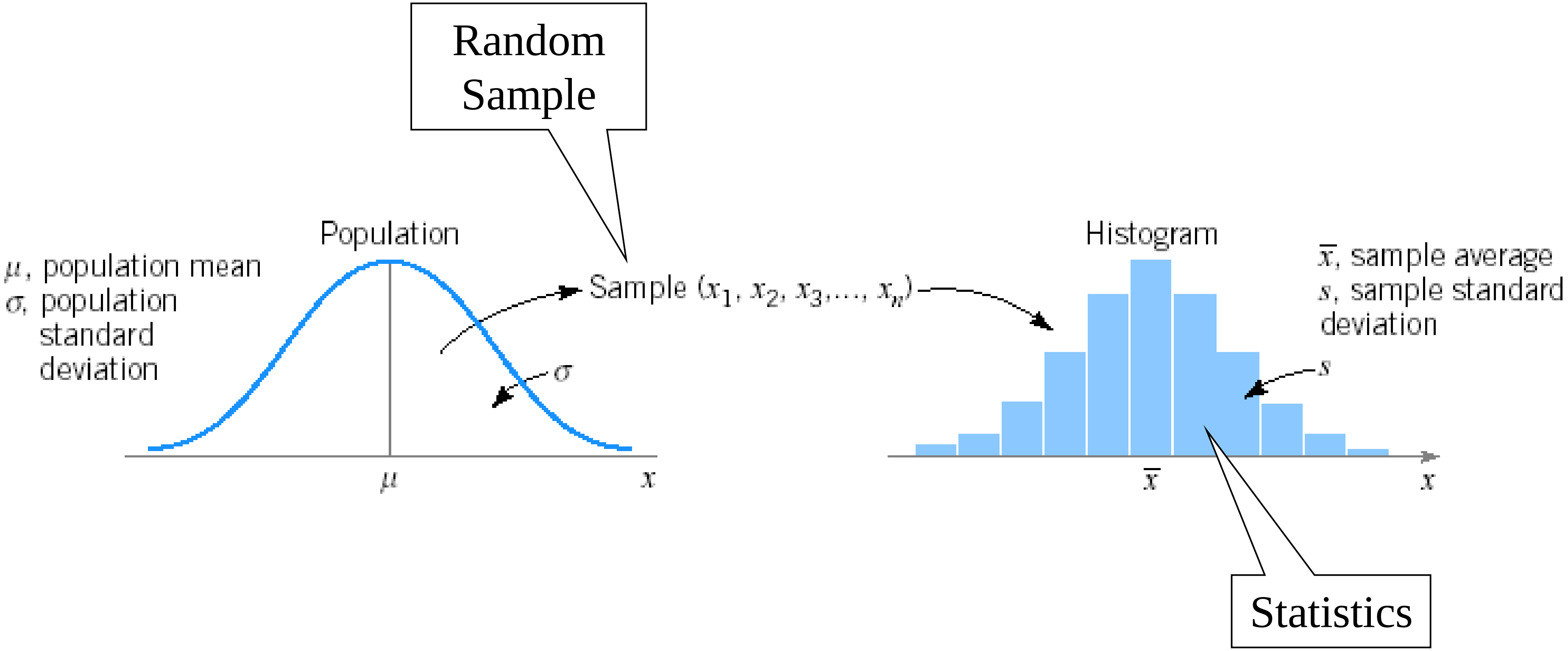


Figura: Relação entre uma População e uma Amostra

fonte: Montgomery, D. (2004). Introdução ao Controle Estatístico da Qualidade, 4a Edição, LTC - Livros Técnicos e Científicos Editora S.A. Rio de Janeiro-RJ. (pág. 54)

- A inferência estatística usa quantidades calculadas a partir das observações na amostra.

- Uma **estatística** é definida como sendo qualquer função dos dados amostrais que não contenha parâmetros desconhecidos.

Estatísticas e Distribuições Amostrais

- Por exemplo, se representamos por $x_1, x_2, \ldots, x_n$ as observações em uma amostra, então:

a média amostral	a variância amostral	e o desvio-padrão amostral
$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$	$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$	$S = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

são estatísticas.

- As estatísticas $\bar{x}$ e S (ou S^2) descrevem a tendência central e a variabilidade da amostra, respectivamente.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Estatísticas e Distribuições Amostrais

- Se a distribuição de probabilidade da população de onde foi retirada a amostra é conhecida, podemos determinar a distribuição de probabilidade de várias estatísticas calculadas a partir dos dados amostrais.

- A distribuição de probabilidade de uma estatística é chamada **distribuição amostral**.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Amostras da Distribuição Normal

- Supondo que x é uma variável aleatória normalmente distribuída com média μ e variância σ^2.

- Se $x_1, x_2, x_3, \dots, x_n$ é uma amostra aleatória de tamanho n desse processo, então a distribuição da média amostral $\bar{x}$ é $N(\mu;\ {}^{\sigma^2}/_{n})$.

Amostras da Distribuição Normal

- Supondo que x é uma variável aleatória normalmente distribuída com média μ e variância σ^2.

- Se $x_1, x_2, x_3, \ldots, x_n$ é uma amostra aleatória de tamanho n desse processo, então a distribuição da média amostral $\bar{x}$ é $N(\mu;\ {\sigma^2}/{n})$.
  - Isso segue direto do resultado sobre a distribuição de combinações lineares de variáveis aleatórias normais.

Amostras da Distribuição Normal

- Esta propriedade da média amostral não está restrita exclusivamente ao caso de amostras de populações normais.

- Do *teorema central do limite*, sabemos que, independentemente da distribuição da população, a distribuição de $\sum_{i=1}^{n} \bar{x}_i$ é aproximadamente normal.

- Então, independente da distribuição da população, a distribuição amostral da média amostral é aproximadamente:

$$\bar{x} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$

Amostras da Distribuição Normal

- O desvio-padrão geralmente refere-se à distância média de uma observação em relação à média da população.
- Mas quando pegamos a média de *n* observações (média amostral), o desvio-padrão das médias das amostras será menor do que o desvio-padrão da população.
- A média amostral deveria variar menos de amostra para amostra.
- Na medida em que o tamanho (*n* observações) da amostra é maior, teremos aproximadamente a mesma média amostral todas as vezes.
- Assim, **a variância (o desvio-padrão) para uma distribuição das médias amostrais será dividido por *n* (raiz de *n*)**.

$$\bar{x} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$

Inferência sobre a Qualidade do Processo

ESTIMAÇÃO PONTUAL DE PARÂMETROS DE PROCESSOS

Estimação Pontual de Parâmetros de Processos

- Uma variável aleatória é caracterizada ou descrita pela sua distribuição de probabilidade.

- Esta distribuição é descrita pelos seus **parâmetros**.
  - Por exemplo, a média μ e a variância σ^2 da distribuição normal são seus parâmetros.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Estimação Pontual de Parâmetros de Processos

- No controle estatístico da qualidade, a distribuição de probabilidade é usada para descrever ou modelar alguma característica da qualidade, tal como a dimensão crítica de um produto ou a fração de defeitos em um processo de produção.

- Sendo assim, nosso interesse recai em fazer inferência sobre parâmetros de distribuições de probabilidade.

- Como os parâmetros em geral são desconhecidos, necessitamos de procedimentos para estimá-los a partir dos dados de uma amostra.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Estimação Pontual de Parâmetros de Processos

- Um estimador de um parâmetro desconhecido pode ser definido como uma estatística que corresponde a este parâmetro.

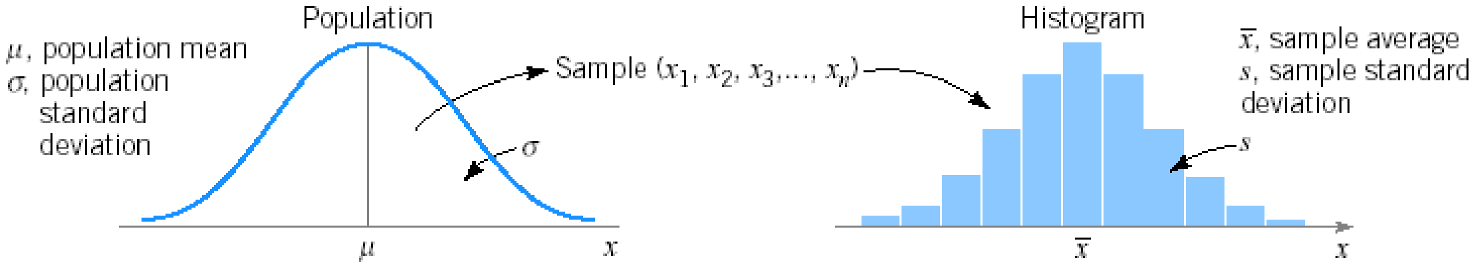
- Um valor numérico particular de tal estimador, calculado a partir dos dados de uma amostra, é chamado **estimativa**.

- Um **estimador pontual** é uma estatística que produz um único valor numérico como estimativa de um parâmetro desconhecido.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Estimação Pontual de Parâmetros de Processos

- Considere uma variável aleatória *x*, com distribuição de probabilidade *f*(*x*).



- Suponha que a média μ e a variância σ^2 desta distribuição sejam ambas desconhecidas.

- Se uma amostra aleatória de tamanho *n* é selecionada, então a **média amostral** $\bar{x}$ e a **variância amostral** S^2 são estimadores pontuais da **média populacional** μ e a da **variância populacional** σ^2, respectivamente.

Exemplo

- Suponha uma distribuição normal que represente um processo produzindo mancais e que *x* seja o diâmetro interno. Queremos obter estimativas pontuais da média e da variância do diâmetro interno dos mancais produzidos por este processo.

- Poderíamos, então, medir os diâmetros internos de uma amostra aleatória de *n* = 20 (por exemplo) mancais e calcular a média e a variância amostral.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Exemplo (continuação)

- Se resultasse $\bar{x} = 1{,}495$ e $S^2 = 0{,}001$, então:
  - a estimativa pontual de μ seria $\hat{\mu} = \bar{x} = 1{,}495$; e
  - a estimativa pontual de σ^2 seria $\hat{\sigma}^2 = S^2 = 0{,}001$.

- Lembre-se de que o símbolo “^” é usado para denotar uma estimativa de um parâmetro.

Estimação Pontual de Parâmetros de Processos

- Algumas propriedades importantes são requeridas de um bom estimador pontual.

- Duas das mais importantes são as propriedades seguintes:

1) O estimador pontual deve ser **não-viesado**. Isto é, o valor esperado do estimador pontual deve ser igual ao parâmetro sendo estimado, ou seja, na média o estimador acerta o valor do parâmetro.

2) O estimador pontual deve ter **variância mínima**. Qualquer estimador pontual é uma variável aleatória. Então, um estimador pontual de mínima variância deve ter uma variância que é menor que a variância de qualquer outro estimador pontual do parâmetro.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Estimação Pontual de Parâmetros de Processos

- A média e a variância amostral $\bar{x}$ e S^2 são estimadores não-viesados da média e da variância populacionais μ e σ^2, respectivamente.

  Isto é:

$$E(\bar{x}) = \mu \quad \text{e} \quad E(S^2) = \sigma^2$$

onde o operador *E* é simplesmente o operador do Valor Esperado, um modo abreviado de denotar o processo de calcular a média de uma variável aleatória.

Estimação Pontual de Parâmetros de Processos

Na Estatística, sabe-se que:

- a média amostral $\bar{x}$ é um estimador não-viesado da média populacional μ
- o desvio-padrão amostral *S* não é um estimador pontual não-viesado do desvio-padrão populacional σ.

- Na prática do controle de qualidade, pode-se obter uma estimativa não-viesada para o desvio-padrão através de:

$$\hat{\sigma} = \frac{S}{c_4}$$

- A Tabela do Apêndice VI (livro Montgomery, 2004, pág. 489) apresenta valores para o fator c_4, para tamanho de amostras $2 \leq n \leq 25$.

Estimação Pontual de Parâmetros de Processos

Factors for Constructing Variables Control Charts

Observations in Sample, n	Chart for Averages: Factors for Control Limits A	A_2	A_3	Factors for Center Line c_4	$1/c_4$	Chart for Standard Deviations: Factors for Control Limits B_3	B_4	B_5	B_6	Chart for Ranges: Factors for Center Line d_2	$1/d_2$	Factors for Control Limits d_3	D_1	D_2	D_3	D_4
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.686	0	3.267
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	4.358	0	2.574
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	4.698	0	2.282
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.114
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.004
7	1.134	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.204	5.204	0.076	1.924
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.388	5.306	0.136	1.864
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.547	5.393	0.184	1.816
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.687	5.469	0.223	1.777
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.922	5.594	0.283	1.717
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.203	5.741	0.347	1.653
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	1.424	5.856	0.391	1.608
19	0.688	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	1.487	5.891	0.403	1.597
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.605	5.951	0.425	1.575
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	1.659	5.979	0.434	1.566
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	1.710	6.006	0.443	1.557
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.759	6.031	0.451	1.548
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.806	6.056	0.459	1.541

For $n > 25$.

$$A = \frac{3}{\sqrt{n}} \qquad A_3 = \frac{3}{c_4\sqrt{n}} \qquad c_4 \cong \frac{4(n-1)}{4n-3}$$

$$B_3 = 1 - \frac{3}{c_4\sqrt{2(n-1)}} \qquad B_4 = 1 + \frac{3}{c_4\sqrt{2(n-1)}}$$

$$B_5 = c_4 - \frac{3}{\sqrt{2(n-1)}} \qquad B_6 = c_4 + \frac{3}{\sqrt{2(n-1)}}$$

Estimação Pontual de Parâmetros de Processos

- Outro estimador não-viesado do desvio-padrão σ de uma distribuição normal pode ser calculado a partir da Amplitude da amostra (R):

$$\hat{\sigma} = \frac{R}{d_2}$$

- A amplitude da amostra é:

$$R = máx(x_i) – mín(x_i) = x_{máx} - x_{mín}$$

- O método da amplitude é usado para estimar o desvio-padrão para certos tipos de gráficos de controle para variáveis;

- Valores para o fator d_2, para amostras de tamanho $2 \leq n \leq 25$ são dados na Tabela do Apêndice VI (livro Montgomery, 2004, pág. 489).

Estimação Pontual de Parâmetros de Processos

Factors for Constructing Variables Control Charts

Observations in Sample, n	Chart for Averages: Factors for Control Limits A	A_2	A_3	Factors for Center Line c_4	$1/c_4$	Chart for Standard Deviations: Factors for Control Limits B_3	B_4	B_5	B_6	Chart for Ranges: Factors for Center Line d_2	$1/d_2$	d_3	Factors for Control Limits D_1	D_2	D_3	D_4
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.686	0	3.267
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	4.358	0	2.574
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	4.698	0	2.282
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.114
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.004
7	1.134	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.204	5.204	0.076	1.924
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.388	5.306	0.136	1.864
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.547	5.393	0.184	1.816
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.687	5.469	0.223	1.777
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.922	5.594	0.283	1.717
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.203	5.741	0.347	1.653
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	1.424	5.856	0.391	1.608
19	0.688	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	1.487	5.891	0.403	1.597
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.605	5.951	0.425	1.575
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	1.659	5.979	0.434	1.566
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	1.710	6.006	0.443	1.557
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.759	6.031	0.451	1.548
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.806	6.056	0.459	1.541

For $n > 25$.

$$A = \frac{3}{\sqrt{n}} \qquad A_3 = \frac{3}{c_4\sqrt{n}} \qquad c_4 \cong \frac{4(n-1)}{4n-3}$$

$$B_3 = 1 - \frac{3}{c_4\sqrt{2(n-1)}} \qquad B_4 = 1 + \frac{3}{c_4\sqrt{2(n-1)}}$$

$$B_5 = c_4 - \frac{3}{\sqrt{2(n-1)}} \qquad B_6 = c_4 + \frac{3}{\sqrt{2(n-1)}}$$

Estimação Pontual de Parâmetros de Processos

- Em muitas aplicações da estatística nos problemas de engenharia da qualidade, é conveniente estimar o desvio-padrão pelo **método da amplitude**.

- Seja $x_1, x_2, \ldots, x_n$ uma amostra aleatória de n observações de uma distribuição normal com média μ e variância σ^2.

- A amplitude da amostra é:

$$R = máx(x_i) - mín(x_i) = x_{máx} - x_{mín}$$

Estimação Pontual de Parâmetros de Processos

- Isto é, a amplitude *R* é simplesmente a diferença entre o maior e o menor valor da amostra.
- A variável $W = R/\sigma$ é chamada **amplitude relativa**.
- Um estimador não-viesado do desvio-padrão σ de uma distribuição normal é:

$$\hat{\sigma} = \frac{R}{d_2}$$

- Valores para o fator d_2, para amostras de tamanho $2 \leq n \leq 25$ são dados na Tabela do Apêndice VI (livro Montgomery, 2004, pág. 489).

Estimação Pontual de Parâmetros de Processos

- A utilização da amplitude para estimar σ data dos primeiros dias do controle estatístico da qualidade e era popular por sua facilidade de cálculo.
- Com a advento dos computadores e calculadoras, esta não é mais uma consideração importante.
- Em geral, o "estimador quadrático" baseado em S é preferível.

Estimação Pontual de Parâmetros de Processos

- No entanto, se o tamanho *n* da amostra é relativamente pequeno, o método amplitude funciona bastante bem.

- A eficiência relativa do método da amplitude comparada com *S* é exibida na Tabela a seguir para diferentes tamanhos de amostra:

Tamanho da Amostra *n*	Eficiência Relativa
2	1,000
3	0,992
4	0,975
5	0,955
6	0,930
10	0,850

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Estimação Pontual de Parâmetros de Processos

- Para valores moderados de n, digamos $n \geq 10$, o método da amplitude perde eficiência rapidamente, uma vez que ele ignora toda a informação da amostra compreendida entre os dois valores extremos.

- No entanto, para valores pequenos do tamanho amostral, digamos $n \leq 6$, ele funciona bastante bem, sendo muito satisfatório.

- O método da amplitude será usado para estimar o desvio-padrão para certos tipos de gráficos de controle para variáveis.

Inferência sobre a Qualidade do Processo

INFERÊNCIA ESTATÍSTICA PARA UMA AMOSTRA

Inferência Estatística para Uma Amostra

- As técnicas de inferência estatística podem ser classificadas em duas amplas categorias:
  - **Estimação de parâmetros**; e
  - **Teste de hipóteses**.

- Anteriormente já introduzimos resumidamente as principais ideias da **estimação pontual** de parâmetros de processos.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Hipótese Estatística

- Uma **hipótese estatística** é uma afirmativa sobre os valores dos parâmetros de uma distribuição de probabilidade.

Hipótese Estatística

- Por exemplo, se pensamos que o diâmetro interno de um mancal é 1,500 in, podemos expressar essa afirmativa formalmente como:

$$H_0: \ \mu = 1{,}500$$
$$H_1: \ \mu \neq 1{,}500$$

Hipótese Estatística

- A afirmativa $H_0\colon\ \mu = 1{,}500$ é chamada **hipótese nula**;
- A afirmativa $H_1\colon\ \mu \neq 1{,}500$ é chamada **hipótese alternativa**;

- No exemplo, H_1 especifica os valores do diâmetro médio que são, ou maiores que 1,500 ou menores que 1,500 e, assim, ela é chamada **hipótese alternativa bilateral**;

- Dependendo do problema, diferentes hipóteses alternativas unilaterais podem ser apropriadas.

Inferência Estatística para Uma Amostra

- Os procedimentos de teste de hipótese são bastante úteis em muitos problemas de controle estatístico da qualidade.
- Eles também formam a base para a maioria das técnicas de controle estatístico de processos.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência Estatística para Uma Amostra

- Uma parte importante do problema de teste de hipótese é a determinação dos valores do parâmetro especificados nas hipóteses nula e alternativa.

- Em geral, isso é feito de uma entre as três maneiras a seguir:
  1. Os valores podem resultar de evidência ou conhecimento anterior; isso acontece frequentemente no controle estatístico da qualidade, onde usamos informação passada para especificar valores para o parâmetro correspondente ao estado sob controle, e periodicamente testamos a hipótese de que esse valor não mudou;
  2. Os valores escolhidos para o parâmetro podem resultar de especificações contratuais ou de projeto, situação que também ocorre frequentemente no controle estatístico da qualidade;
  3. Os valores podem resultar de alguma teoria ou modelo do processo.

Inferência Estatística para Uma Amostra

- Para testar uma hipótese, toma-se uma amostra aleatória da população em estudo, calcula-se uma **estatística de teste** apropriada e, então, rejeita-se ou não a hipótese nula H_0.

- O conjunto de valores da estatística de teste que levam à rejeição de H_0 é chamado **região crítica** ou **região de rejeição** do teste.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência Estatística para Uma Amostra

- Dois tipos de erros podem ser cometidos quando testamos hipóteses.
- Se a hipótese nula é rejeitada quando ela é verdadeira, então dizemos que ocorreu um **Erro Tipo I**.
- Se a hipótese nula não é rejeitada (é aceita) quando ela é falsa, então temos um **Erro Tipo II**.

Inferência Estatística para Uma Amostra

- As probabilidades desses dois tipos de erro são denotadas como:

$$\alpha = P\{erro\ tipo\ I\} = P\{rejeitar\ H_0 | H_0\ \text{é}\ verdadeira\}$$

$$\beta = P\{erro\ tipo\ II\} = P\{não\ rejeitar\ H_0 | H_0\ \text{é}\ falsa\}$$

- Algumas vezes é mais conveniente trabalhar com o *poder* de um teste, onde:

$$Poder = 1 - \beta = P\{rejeitar\ H_0 | H_0\ \text{é}\ falsa\}$$

Inferência Estatística para Uma Amostra

- Então, o **poder** é a probabilidade de *corretamente* rejeitar H_0.

- No trabalho do controle da qualidade, α é às vezes chamado **risco do fabricante**, porque denota a probabilidade de um lote bom ser rejeitado ou a probabilidade de que um processo produzindo valores aceitáveis de uma particular característica da qualidade venha a ser rejeitado como produzindo insatisfatoriamente.

- De forma análoga, β é às vezes chamado de **risco do consumidor**, por denotar a probabilidade de aceitação de um lote de baixa qualidade, ou a probabilidade de permitir que um processo, operando em condições não satisfatórias com respeito a determinada característica da qualidade, continue em operação.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência Estatística para Uma Amostra

- O procedimento geral de teste de hipótese consiste em especificar um valor para a probabilidade α do erro tipo I e, então, planejar um procedimento de teste de tal forma que um valor pequeno da probabilidade β do erro tipo II seja obtido.

- Então, estamos falando de controlar ou escolher diretamente o risco α.

- O risco β é, geralmente, uma função do tamanho da amostra e é controlado indiretamente → quanto maior o tamanho da amostra utilizada no teste, menor o risco β.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência Estatística para Uma Amostra

- Nesta parte do curso, faremos uma revisão dos procedimentos de teste de hipótese quando **uma única amostra** de *n* observações é retirada do processo.

- Veremos também como a informação sobre os valores dos parâmetros do processo contida na amostra pode ser expressa em termos de uma estimativa intervalar chamada **intervalo de confiança**.

- Pode-se, ainda, considerar a inferência estatística para duas amostras de dois processos possivelmente diferentes.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

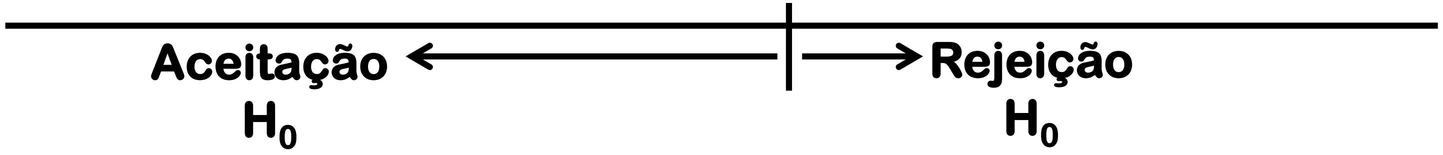
Quadro Resumo para Erros Tipos I e II e Poder de um Teste

α	Prob. Erro Tipo I	Probabilidade de Rejeitar H_0 quando H_0 é verdadeira	α é às vezes chamado risco do fabricante, porque denota a probabilidade de um lote bom ser rejeitado ou a probabilidade de que um processo produzindo valores aceitáveis de uma particular característica da qualidade venha a ser rejeitado como produzindo insatisfatoriamente
β	Prob. Erro Tipo II	Probabilidade de não Rejeitar H_0 quando H_0 é falsa	β é à vezes chamado de risco do consumidor, por denotar a probabilidade de aceitação de um lote de baixa qualidade, ou a probabilidade de permitir que um processo, operando em condições não satisfatórias com respeito a determinada característica da qualidade, continue em operação
Poder	$1 - \beta$	Probabilidade de Rejeitar H_0 quando H_0 é falsa	O poder é a probabilidade de *corretamente* rejeitar H_0

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

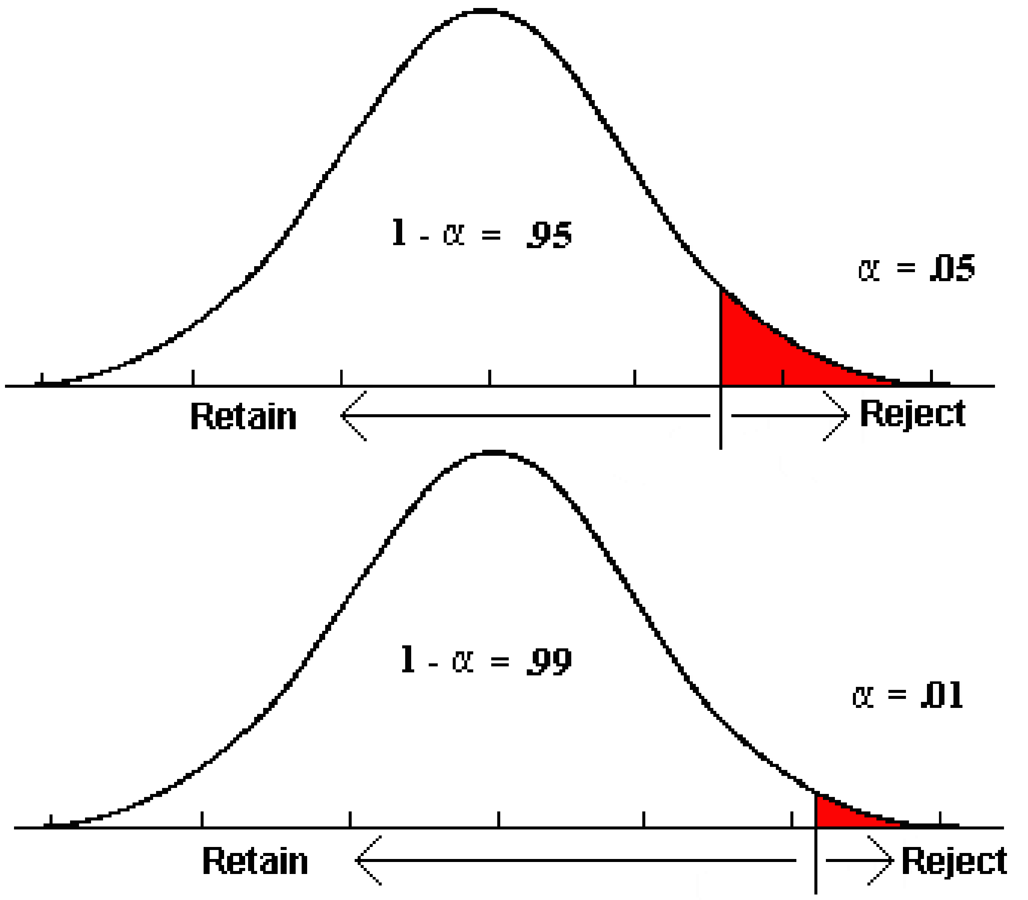
Visão Gráfica para Erros Tipos I e II e Poder de um Teste

α	Prob. Erro Tipo I
β	Prob. Erro Tipo II
Poder	$1 - \beta$



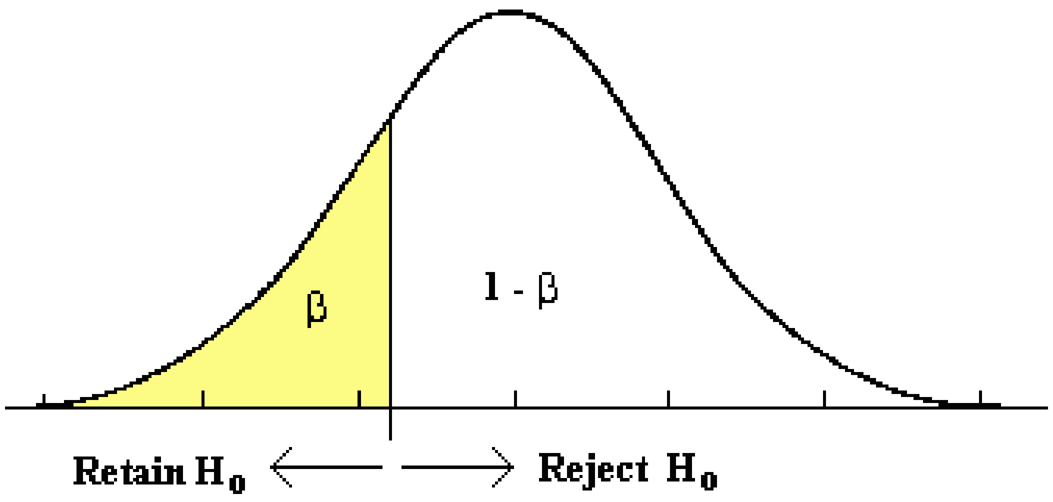
Visão Gráfica para Erros Tipos I e II e Poder de um Teste

α	Prob. Erro Tipo I
β	Prob. Erro Tipo II
Poder	1 – β



Visão Gráfica para Erros Tipos I e II e Poder de um Teste

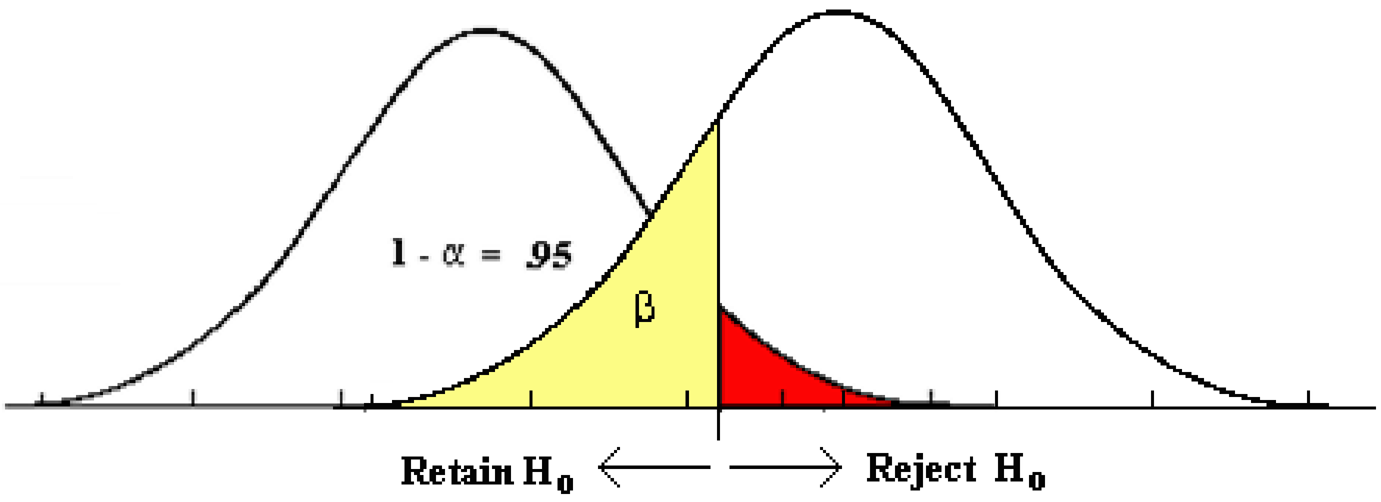
α	Prob. Erro Tipo I
β	Prob. Erro Tipo II
Poder	$1 - \beta$



Fonte: http://www.psychstat.missouristate.edu/IntroBook3/sbk20.htm

Visão Gráfica para Erros Tipos I e II e Poder de um Teste

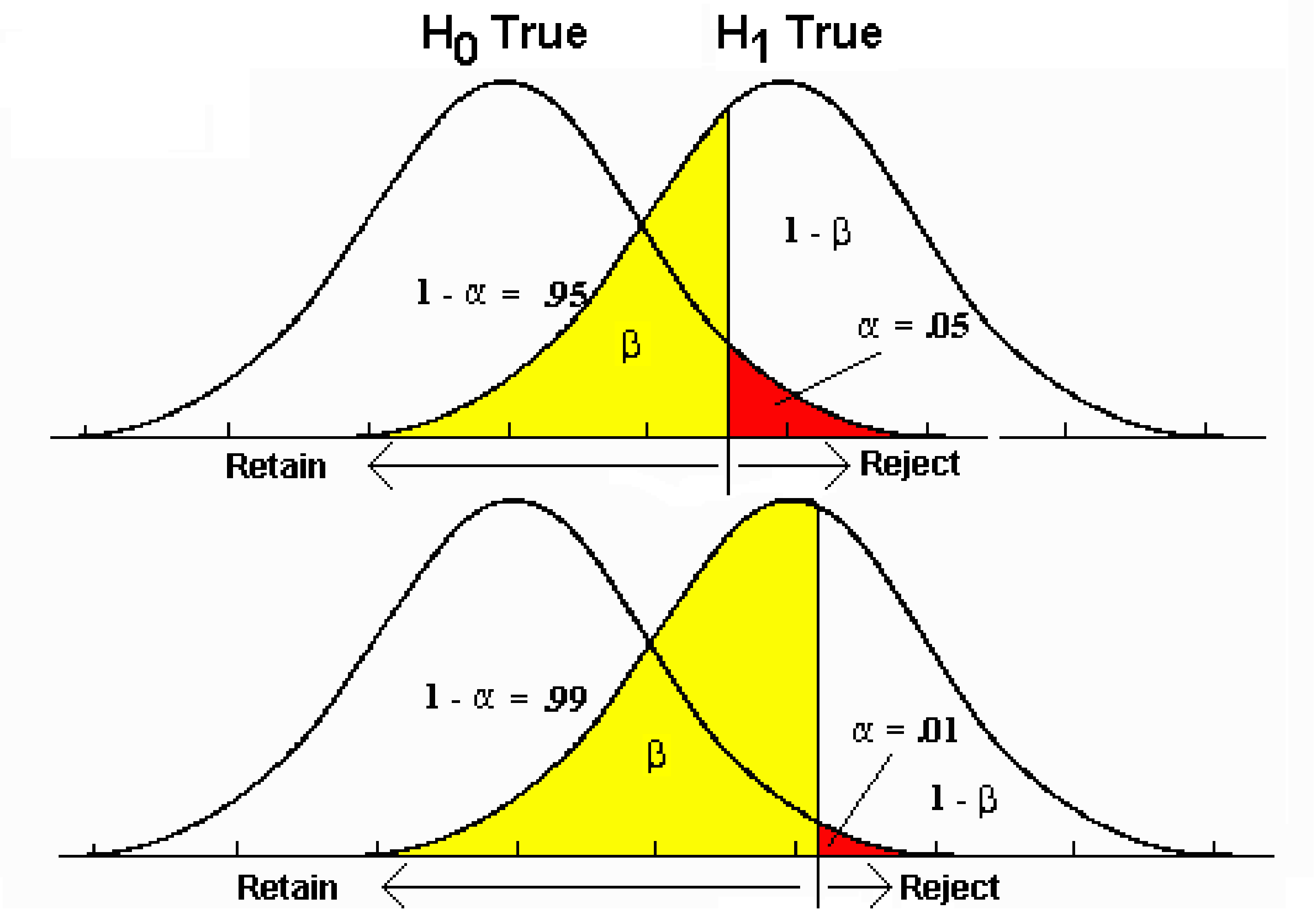
α	Prob. Erro Tipo I
β	Prob. Erro Tipo II
Poder	$1 - \beta$



Fonte: http://www.psychstat.missouristate.edu/IntroBook3/sbk20.htm

Visão Gráfica para Erros Tipos I e II e Poder de um Teste

α	Prob. Erro Tipo I
β	Prob. Erro Tipo II
Poder	$1 - \beta$



Fonte: http://www.psychstat.missouristate.edu/IntroBook3/sbk20.htm

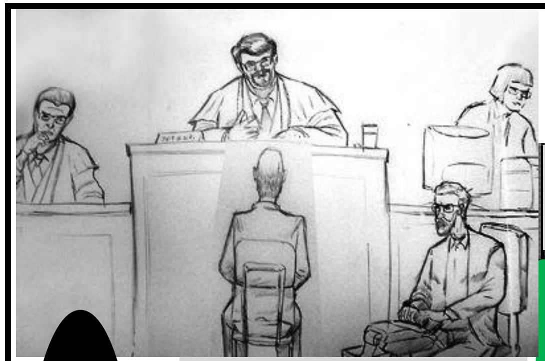
Quadro Resumo para Erros Tipos I e II e Poder de um Teste

HYPOTHESIS TESTING OUTCOMES		Reality: The Null Hypothesis Is True	Reality: The Alternative Hypothesis is True
Research	The Null Hypothesis Is True	Accurate $1 - \alpha$ = nível de confiança	Type II Error β
Research	The Alternative Hypothesis is True	Type I Error α = nível de significância	Accurate $1 - \beta$ = poder

Fonte: http://allpsych.com/researchmethods/errors.html

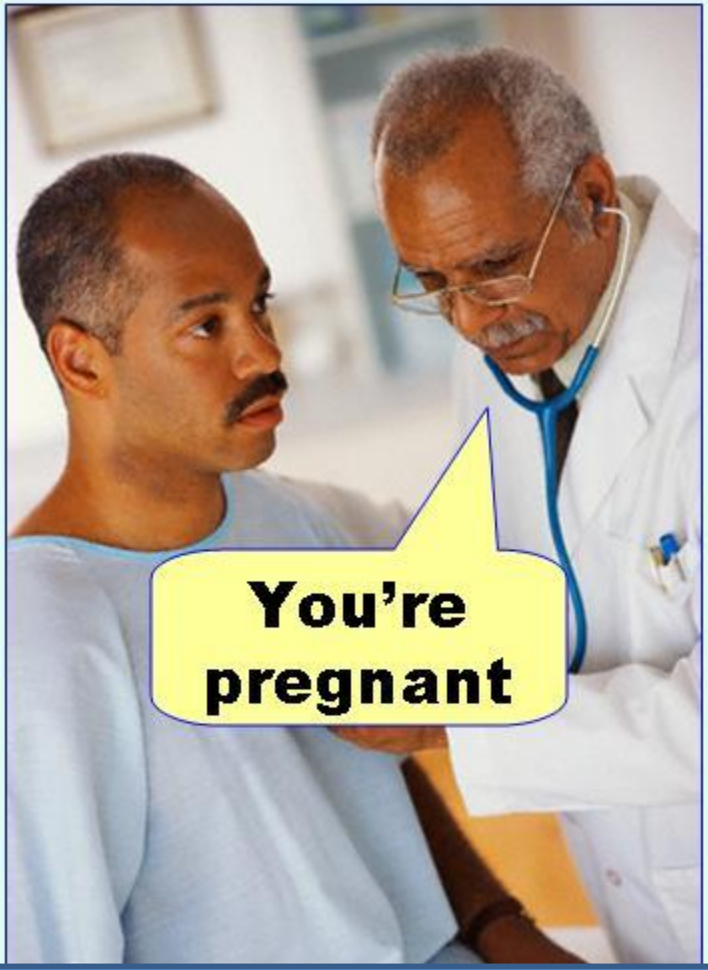
Inferência Estatística para Uma Amostra

Quadro Resumo para Erros Tipos I e II e Poder de um Teste

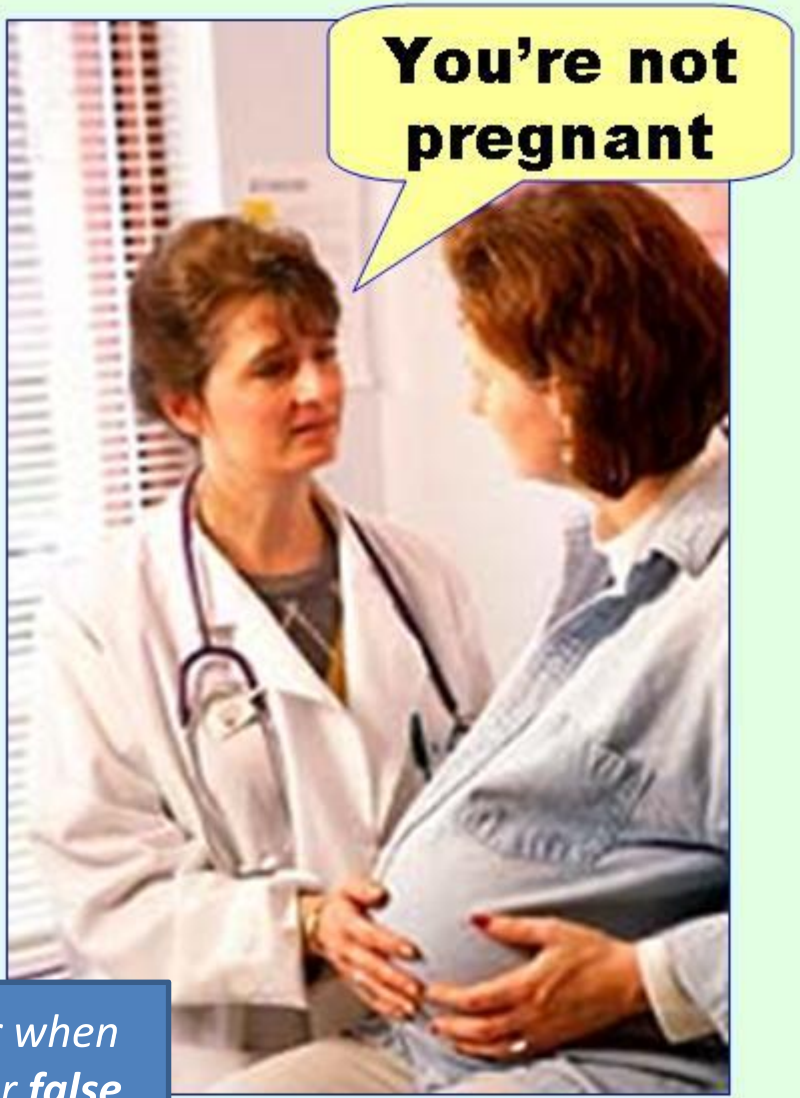


		Realidade	
		Inocente	Culpado
Julgamento	Inocente	Accurate $1 - \alpha$ = nível de confiança	Type II Error β **Risco para a sociedade!** **Falso Negativo**
	Culpado **Risco para o réu!** **Falso Positivo**	Type I Error α = nível de significância	Accurate $1 - \beta$ = poder

Type I error
(false positive)



Type II error
(false negative)



***Type I errors**, also known as **false positives**, occur when you see things that are not there. **Type II errors**, or **false negatives**, occur when you don't see things that are there*

Inferência sobre a Qualidade do Processo

INFERÊNCIA ESTATÍSTICA PARA UMA AMOSTRA

(CONTINUAÇÃO)

Inferência para a Média de uma População com Variância Conhecida

Inferência para a Média de uma População com Variância Conhecida

Teste de Hipóteses

- Suponha que x seja uma variável aleatória com média desconhecida μ e variância conhecida σ^2.
- Queremos testar a hipótese de que a média é igual a um valor nominal, digamos μ_0.
- As hipóteses podem ser formuladas como:

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

Inferência para a Média de uma População com Variância Conhecida

<u>Teste de Hipóteses</u>

- O procedimento para testar essa hipótese é tomar uma amostra aleatória de *n* observações da variável aleatória *x*, calcular a estatística de teste:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

e **rejeitar H_0 se $|Z_0| > Z_{\alpha/2}$, onde $Z_{\alpha/2}$ é o ponto da distribuição normal padrão correspondendo à porcentagem superior $\alpha/2$**.

Regra de Decisão!

Inferência para a Média de uma População com Variância Conhecida

Teste de Hipóteses

- O procedimento para testar essa hipótese é tomar uma amostra aleatória de *n* observações da variável aleatória *x*, calcular a estatística de teste:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Por que usamos $\sqrt{n}$?

Porque agora estamos falando de uma amostra e não mais de valores individuais. Portanto, a distribuição dessa refere-se à distribuição das médias amostrais e não dos valores individuais de *x*.

e **rejeitar H_0 se $|Z_0| > Z_{\alpha/2}$, onde $Z_{\alpha/2}$ é o ponto da distribuição normal padrão correspondendo à porcentagem superior $\alpha/2$**.

Regra de Decisão!

Inferência para a Média de uma População com Variância Conhecida

Teste de Hipóteses

- O procedimento para testar essa hipótese é tomar uma amostra aleatória de *n* observações da variável aleatória *x*, calcular a estatística de teste:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Note que:
- Quanto menor for o *n* (tamanho da amostra) ➔ menor será o Z_0,
- Assim, p/ um $|Z_{\alpha/2}|$ dado, diminui-se a chance de $|Z_0|$ ser maior do que $|Z_{\alpha/2}|$,
- Logo, diminui-se a chance de rejeitar H_0.

e **rejeitar H_0 se $|Z_0| > Z_{\alpha/2}$, onde $Z_{\alpha/2}$ é o ponto da distribuição normal padrão correspondendo à porcentagem superior $\alpha/2$**.

Regra de Decisão!

Inferência para a Média de uma População com Variância Conhecida

Teste de Hipóteses

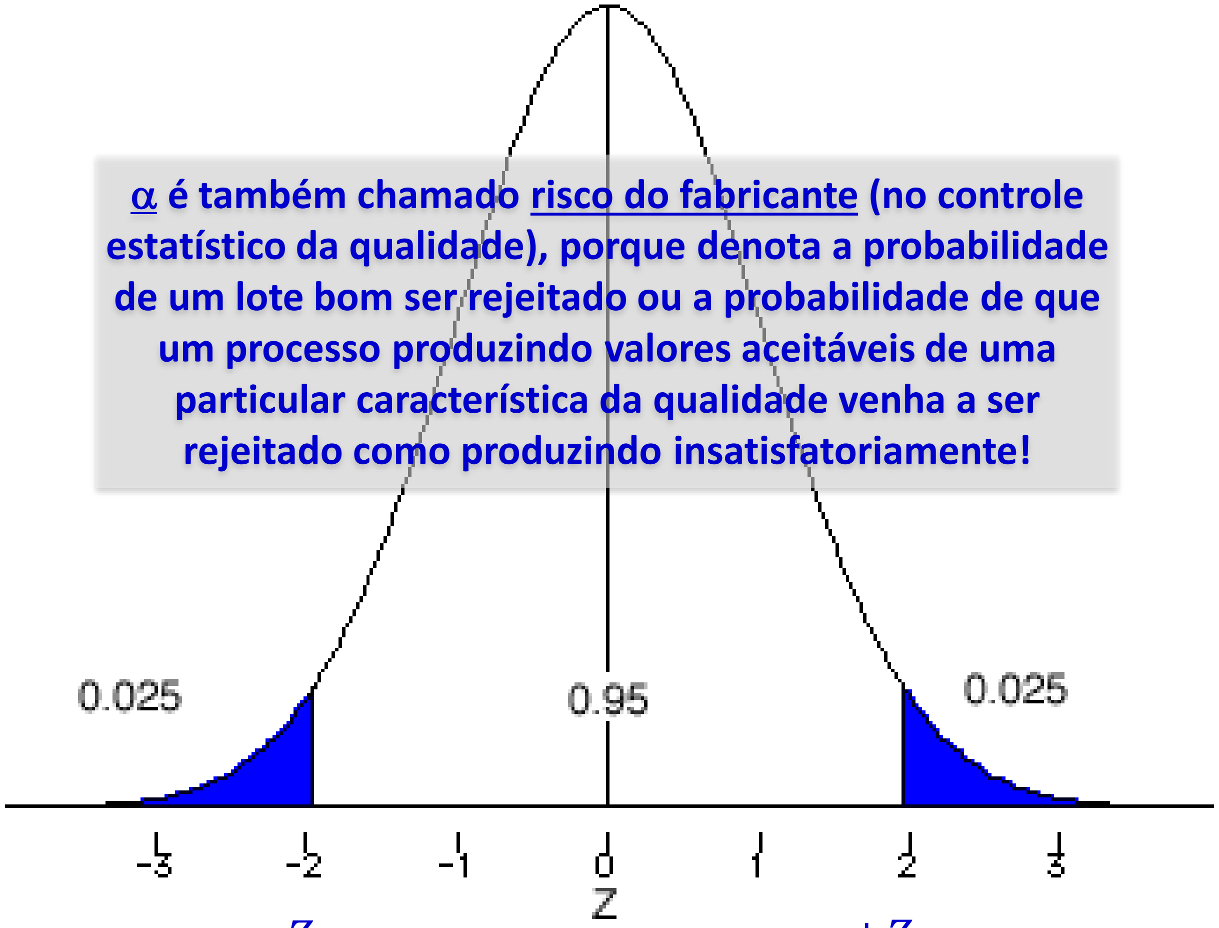
- Podemos dar uma justificativa intuitiva para esse procedimento de teste.
- Pelo teorema central do limite, sabemos que a média amostral $\bar{x}$ é aproximadamente distribuída como uma $N(\mu; {}^{\sigma^2}/_{n})$.
- Agora, se $H_0: \mu = \mu_0$ é verdadeira (realidade), a estatística de teste Z_0 é distribuída aproximadamente como uma $N(0; 1)$.
- Consequentemente, devemos esperar que $100 \times (1 - \alpha)\%$ dos valores de Z_0 caiam entre $-Z_{\alpha/2}$ e $+Z_{\alpha/2}$.
- ...

Inferência para a Média de uma População com Variância Conhecida

Teste de Hipóteses

- Devemos esperar que $100 \times (1-\alpha)\%$ dos valores de Z_0 caiam entre $-Z_{\alpha/2}$ e $+Z_{\alpha/2}$.
- Assim, uma amostra que produz um valor de Z_0 fora desses limites deveria ser considerada pouco comum se a hipótese nula fosse verdadeira e seria, então, considerada evidência para a rejeição de $H_0: \mu = \mu_0$.
- Note que α é a probabilidade de erro tipo I para o teste, e os intervalos $(+Z_{\alpha/2}; +\infty)$ e $(-\infty; -Z_{\alpha/2})$ formam a região crítica para o teste.

Inferência Estatística para Uma Amostra



<u>α</u> é também chamado <u>risco do fabricante</u> (no controle estatístico da qualidade), porque denota a probabilidade de um lote bom ser rejeitado ou a probabilidade de que um processo produzindo valores aceitáveis de uma particular característica da qualidade venha a ser rejeitado como produzindo insatisfatoriamente!

$-\infty$ $\quad -Z_{\alpha/2}$ $\quad +Z_{\alpha/2}$ $\quad +\infty$

$-1{,}96$ $\quad +1{,}96$

α = 5% ou 0,05

$\alpha/2$ = 2,5% ou 0,025

α é a probabilidade de erro tipo I para o teste, e os intervalos $(Z_{\alpha/2}; \infty)$ e $(-\infty; -Z_{\alpha/2})$ formam a região crítica para o teste

Da Tabela de Distribuição Normal Acumulada, temos que para um $\alpha/2$ = 0,025, precisamos encontrar o *Z* de tal forma que $\Phi(Z_{\alpha/2})$ = 0,975.

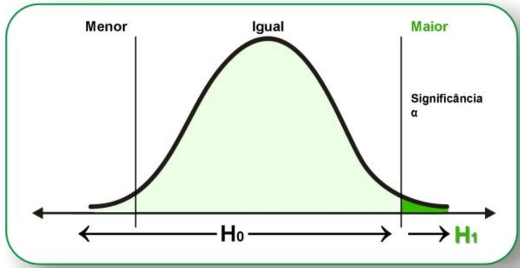
Logo $Z_{\alpha/2}$ = 1,96.

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência para a Média de uma População com Variância Conhecida

Teste de Hipóteses

- Em algumas situações podemos querer rejeitar a hipótese nula H_0 apenas se a verdadeira média for maior que μ_0.



- Então, a hipótese alternativa **unilateral** é $H_1: \mu > \mu_0$ e rejeitaríamos $H_0: \mu = \mu_0$ apenas se $Z_0 > Z_\alpha$.
- Se a rejeição é desejada apenas quando $\mu < \mu_0$ então a hipótese alternativa é $H_1: \mu < \mu_0$ e rejeitamos H_0 quando $Z_0 < -Z_\alpha$.

Exemplo

- A força de pressão interna de garrafas de vidro usadas para embalar refrigerantes é uma importante característica da qualidade. O engarrafador deseja saber se a pressão média excede 175 psi. Da experiência passada, ele sabe que o desvio padrão da força de pressão é 10 psi. O fabricante do vidro submete lotes dessas garrafas ao engarrafador, que está interessado em testar a hipótese:

$$H_0\colon \mu = 175$$
$$H_1\colon \mu > 175$$

- É importante notar que um lote é aceito se a hipótese nula $H_0\colon \mu = 175$ for rejeitada. Uma **amostra** aleatória de 25 garrafas é selecionada e as garrafas são colocadas em uma máquina que aumenta a pressão na garrafa até que ela quebre. A pressão média para o estouro das garrafas na amostra foi de $\bar{x} = 182\ psi$. Determine se as garrafas deste lote são apropriadas para o uso no processo de envasamento de refrigerantes.

Considerar um nível de confiança de 95%.

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança

- Uma estimativa intervalar de um parâmetro é um intervalo entre duas estatísticas que inclui o verdadeiro valor do parâmetro com uma dada probabilidade.

- Por exemplo, para construir um estimador intervalar para a média μ, temos que encontrar duas estatísticas *I* e *S* tais que:

$$P\{I \leq \mu \leq S\} = 1 - \alpha$$

Fonte: Montgomery (2004). Introdução ao Controle Estatístico da Qualidade, p. 53-91

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança

- O intervalo: $I \leq \mu \leq S$

é chamado **intervalo de confiança de nível 100*(1 – α)%** para a média desconhecida μ. *I* e *S* são os limites inferior (*I*) e superior (*S*) de confiança e $1 - \alpha$ é o nível ou coeficiente de confiança.

- Às vezes, a largura da metade do intervalo $S - \mu$ ou $\mu - I$ é chamada **precisão** do intervalo de confiança

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança

- A interpretação de um **intervalo de confiança** é que, se um grande número de tais intervalos é construído, cada um resultante de uma amostra aleatória, então 100*(1 – α)% deles irão conter o verdadeiro valor de μ.

- Assim, intervalos de confiança têm uma interpretação frequencial.

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança

- O intervalo de confiança determinado pela equação $P\{I \leq \mu \leq S\} = 1 - \alpha$ é mais corretamente chamado de intervalo de confiança **bilateral**, uma vez que especifica limites inferior (I) e superior (S) para μ.

- Algumas vezes, em aplicações de controle da qualidade, um intervalo de confiança **unilateral** pode ser mais apropriado.

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança (inferior)

- Um intervalo de confiança unilateral inferior de nível 100*(1 – α)% para μ é dado pelo intervalo:

$$I \leq \mu$$

onde *I*, o limite inferior de confiança, é escolhido de modo que:

$$P\{I \leq \mu\} = 1 - \alpha$$

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança (superior)

- Um intervalo de confiança unilateral superior de nível 100*(1 – α)% para μ é dado pelo intervalo:

$$\mu \leq S$$

onde *S*, o limite superior de confiança, é escolhido de modo que:

$$P\{\mu \leq S\} = 1 - \alpha$$

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança para a Média com Variância Conhecida

- Considere a variável aleatória *x* com média desconhecida μ e variância conhecida σ^2.
- Suponha que uma amostra aleatória de *n* observações é sorteada, digamos $x_1, x_2, \ldots, x_n$, e $\bar{x}$ calculada.
- Então, o intervalo de confiança bilateral de nível 100*(1 – α)% para μ é:

$$\bar{x} - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

onde $Z_{\alpha/2}$ é a abscissa da distribuição $N(0;1)$ tal que $P\{z \geq Z_{\alpha/2}\} = \alpha/2$.

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança para a Média com Variância Conhecida

- Note que $\bar{x}$ é distribuída aproximadamente como uma $N(\mu; {}^{\sigma^2}/_{n})$ independentemente da distribuição de x, pelo teorema central do limite.

- Consequentemente, a equação anteriormente apresentada:

$$\bar{x} - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

é um intervalo de confiança de nível aproximado 100*(1 – α)% para μ, independentemente da distribuição de x.

- Se x é $N(\mu; \sigma^2)$, então a equação anterior é um intervalo de confiança de nível exato 100*(1 – α)%.

Inferência para a Média de uma População com Variância Conhecida

Intervalos de Confiança para a Média com Variância Conhecida

- O intervalo de confiança superior de nível 100*(1 – α)% para μ é:

$$\mu \leq \bar{x} + Z_\alpha \times \frac{\sigma}{\sqrt{n}}$$

- O intervalo de confiança inferior de nível 100*(1 – α)% para μ é:

$$\bar{x} - Z_\alpha \times \frac{\sigma}{\sqrt{n}} \leq \mu$$

Inferência Estatística para Uma Amostra


MENU G1 ELEIÇÕES 2018 BUSCAR
ELEIÇÃO EM NÚMEROS

Pesquisa Ibope de 23 de outubro para presidente por sexo, idade, escolaridade, renda, região, religião e cor

Confira os números da pesquisa por segmento. Levantamento foi feito entre os dias 21 e 23 de outubro e ouviu 3.010 eleitores.

Por G1
24/10/2018 05h00 · Atualizado há 11 meses

Pesquisa Ibope de 23 de outubro para presidente por sexo, idade, escolaridade, renda, região, religião e cor

Fonte: https://g1.globo.com/politica/eleicoes/2018/eleicao-em-numeros/noticia/2018/10/24/pesquisa-ibope-de-23-de-outubro-para-presidente-por-sexo-idade-escolaridade-renda-regiao-religiao-e-cor.ghtml

Exemplo

- Reconsidere a situação do teste das garrafas do Exemplo anterior. Como $\bar{x} = 182\ psi$, sabemos que uma estimativa pontual razoável para a pressão média de ruptura é $\hat{\mu} = \bar{x} = 182\ psi$. Podemos também achar um intervalo de confiança de nível 100*(1 – α)% para μ .
  a) Suponha que um intervalo de confiança **bilateral** de nível 95% seja especificado. Calcule os intervalos de confiança.

Exemplo (cont.)

- Reconsidere a situação do teste das garrafas do Exemplo anterior. Como $\bar{x} = 182\ psi$, sabemos que uma estimativa pontual razoável para a pressão média de ruptura é $\hat{\mu} = \bar{x} = 182\ psi$. Podemos também achar um intervalo de confiança de nível $100*(1 - \alpha)\%$ para μ .
  b) Suponha que um intervalo de confiança **unilateral** de nível 95% seja especificado. Calcule o intervalo de confiança.

Inferência Estatística para Uma Amostra

- Descobrimos que a pressão média de ruptura da garrafa excede 178,71 psi com 95% de confiança!

Questão:

- Qual é a real probabilidade da pressão média da garrafa ser maior ou igual a 175?
  - Note que essa probabilidade seria um "α" calculado, e não definido;
  - Para isso usaremos o conceito dos **valores *P***.