Métodos Estocásticos da Engenharia II

Capítulo 4 - Inferência: intervalos de confiança

Prof. Magno Silvério Campos

2024/2

Bibliografia

Essas notas de aulas foram baseadas nas seguintes obras:

1. BORNIA, A. C.; BARBETTA, P. A.; REIS, M. M. *Estatística para Cursos de Engenharia e Informática*. 2. ed. São Paulo: Atlas, 2009.
2. CANCHO, V.G. *Notas de Aulas sobre Noções de Estatística e Probabilidade*. São Paulo: USP, 2010.
3. HINES, W.W.; *et al. Probabilidade e Estatística na Engenharia*. 4. ed. Rio de Janeiro: LTC, 2006.
4. MONTGOMERY, D.C.; RUNGER, G.C. *Estatística Aplicada e Probabilidade para Engenheiros*. 6. ed. Rio de Janeiro: LTC, 2016.

Aconselha-se pesquisá-las para se obter um maior aprofundamento e um melhor aproveitamento nos estudos.

Conteúdo Programático

1. Seção 1 - Estimação de parâmetros
   - Estimação pontual;
   - Estimação por intervalos;
2. Seção 2 - Intervalos para uma única amostra
   - Intervalo para a média de uma população normal:
     - com variância conhecida;
     - com variância desconhecida;
     - para amostras grandes;
   - Intervalo para a variância de uma população normal;
   - Intervalo para a proporção populacional.
3. Seção 3 - Intervalos para duas amostras
   - Intervalo para a diferença de médias de populações normais:
   - Intervalo para a diferença de médias a partir de amostras emparelhadas;
   - Intervalo para a razão das variâncias de duas distribuições normais;
   - Intervalo para a diferença entre duas proporções.

Introdução

Estimação de parâmetros

Muitas vezes, sabe-se ou admite-se que uma variável aleatória X segue uma certa distribuição de probabilidade, mas não são conhecidos os valores dos parâmetros da distribuição.

Por exemplo, se X seguir a distribuição normal, pode-se querer saber o valor de seus parâmetros (a média e a variância).

Para estimar os parâmetros, considera-se uma amostra aleatória de tamanho n e, utiliza-se os dados amostrais para estimar os parâmetros desconhecidos. Isso é conhecido como o problema de estimação. E esse problema pode ser dividido em duas categorias: estimação pontual e estimação por intervalos.

Estimação Pontual

Sabemos de aulas passadas que, $\hat{\theta}$ é conhecido como um *estimador*, e um valor numérico particular assumido pelo estimador é conhecido como uma *estimativa*. Note que $\hat{\theta}$ pode ser tratado como uma variável aleatória, pois é uma função dos dados amostrais. O estimador $\hat{\theta}$ fornece uma regra, que diz como se pode estimar o θ verdadeiro. Por exemplo, ao se admitir que

$$\hat{\theta} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \bar{X}$$

temos que $\bar{X}$, a média amostral, é um estimador do valor médio verdadeiro (ou populacional), μ. Se em um caso específico, $\bar{X} = 50$, tem-se uma *estimativa* de μ.

O estimador $\hat{\theta}$ obtido acima nos fornece uma única estimativa pontual de θ.

Estimação por intervalos

Ao invés de se obter uma única estimativa de θ, suponha que obtém-se duas estimativas de θ por meio da construção de dois estimadores, $\hat{\theta}_1$ e $\hat{\theta}_2$, e considera-se com alguma confiança que o intervalo entre $\hat{\theta}_1$ e $\hat{\theta}_2$ inclui o verdadeiro θ.

Assim, em um estimativa por intervalo, em contraste com a estimativa pontual, fornecemos uma classe de possíveis valores dentro da qual se pode encontrar o verdadeiro θ.

Intervalos de confiança

Seja $X_1, \ldots, X_n$ uma amostra aleatória de população com a característica X, cuja distribuição de probabilidade é $f(x;\theta)$. Seja $L = G(X_1, \ldots, X_n)$ e $U = H(X_1, \ldots, X_n)$ duas estatísticas tais que $L < U$ e que

$$P(L \leq \theta \leq U) = \gamma = 1 - \alpha.$$

Então, o intervalo $(L \leq \theta \leq U)$ é chamado de intervalo de $100\gamma\%$ ou $(1-\alpha)100\%$ de confiança para θ.

Denota-se por $IC(\theta, 1-\alpha)$, o intervalo de $(1-\alpha)100\%$ de confiança para θ. Isto é,

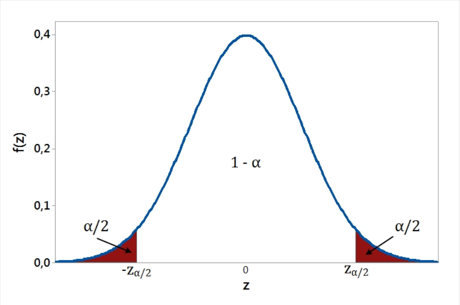
$$IC(\theta; 1-\alpha) = (\ L \leq \theta \leq U\ )$$

onde L e U são os limites inferior e superior de confiança, respectivamente, e $\gamma = 1 - \alpha$ é o coeficiente (ou nível) de confiança.

Intervalo de Confiança para a média de uma população normal com variância conhecida

Suponha que $X_1, \ldots, X_n$ seja uma amostra aleatória de tamanho n extraída de uma população, com a característica X, que tem distribuição normal com média μ e variância σ^2. Foi visto que a média amostral $\bar{X}$ tem distribuição normal com média μ e variância σ^2/n. Assim

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$



Agora, considere o cálculo da seguinte probabilidade, conforme indicada na figura acima:

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) =$$

Logo,

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$$

ou, o que é equivalente,

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha. \tag{1}$$

Que sendo manipulada matematicamente, permite-nos escrever:

$$-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \Rightarrow \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{2}$$

Intervalo de confiança para a média de uma população normal com variância conhecida

Se $\bar{x}$ for a média amostral de uma amostra aleatória, de tamanho n, proveniente de uma população com variância conhecida σ^2, um intervalo com $100(1-\alpha)\%$ de confiança pra μ é dados por

$$(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \tag{3}$$

Sendo $z_{\alpha/2}$ o ponto superior com $100\alpha/2\%$ da distribuição normal padrão.

Exemplo - [Montgomery e Runger(2016)]

Dez medidas de energia (J) de impacto em corpos-de-prova feitos de aço são:

64,1	67,7	64,5	64,6	64,5	64,3	64,6	64,8	64,2	64,3

Considere que a energia de impacto seja normalmente distribuída, com $\sigma = 1J$. Queremos encontrar um IC de 95% para μ, que é a energia média de impacto.

Interpretando um intervalo de confiança

De acordo com Montgomery e Runger (2016), um erro muito comum que se comete ao interpretar o intervalo de confiança é dizer que a probabilidade de μ estar no intervalo é $1 - \alpha$.

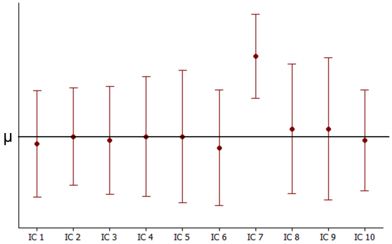
O erro resulta do fato de que μ não é uma variável aleatória e sim um parâmetro que caracteriza uma população. Ou seja, μ não varia e portanto, não tem uma distribuição de probabilidade.

Deve ficar claro que o intervalo é que é aleatório (**antes de que seja obtida a amostra e calculados os valores**).

Portanto, o correto seria dizer que a probabilidade do intervalo a **ser escolhido** conter o verdadeiro valor da média é igual a $1 - \alpha$.

Ou seja, obtendo-se várias amostras e, para cada uma delas, calcula-se o correspondente intervalo de confiança para μ, teremos que $100(1-\alpha)\%$ dessas amostras conterão o valor verdadeiro de μ e $100\alpha\%$ dessas amostras não conterão a média populacional.

Este fato está ilustrado na figura abaixo:



Repare que, por exemplo, se fosse calculado um intervalo de 90% de confiança, somente 10% dos intervalos falhariam em conter μ.

Agora na prática, obtemos somente uma amostra aleatória e calculamos um intervalo de confiança. No exemplo anterior, encontramos

$$64,14J \leq \mu \leq 65,38J$$

A observação apropriada é que o intervalo observado envolve o valor verdadeiro de μ, com confiança de $100(1-\alpha)\%$

Não sabemos se a afirmativa é verdadeira para essa amostra específica, mas o método usado para obtenção do intervalo fornece afirmativas corretas $100(1-\alpha)\%$ das vezes.

Observação 1

No caso de a amostragem ser sem reposição de uma população finita de N elementos, o intervalo de $(1-\alpha)100\%$ de confiança para μ é:

$$IC(\mu; 1-\alpha) = \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}; \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right). \quad (4)$$

Observação 2

Se $\gamma \to 1$, o comprimento do intervalo de confiança é maior.

Observação 3

O tamanho da amostra aparece no denominador de $z_{\frac{\alpha}{2}}\sigma$. Para amostras grandes os intervalos de confiança têm comprimentos mais curtos, portanto, informação mais precisa.

Exemplo - [Cancho(2010)]

De um lote de 2200 produtos, com um desvio padrão de duração igual a 900 horas, foram sorteados, sem reposição, 81 produtos. O tempo médio de duração dos produtos sorteados foi 3200 horas . Construa um intervalo de 95% de confiança para o tempo médio dos produtos do lote.

Determinação do tamanho da amostra para estimar a média μ

Do intervalo de confiança para a média populacional

$$IC(\mu; 1-\alpha) = \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

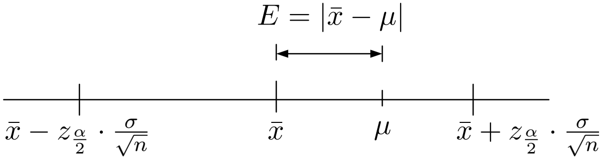
deseja-se que o comprimento do intervalo seja o mais curto possível. Para isso, tem-se duas opções:

(i) Diminuir o coeficiente de confiança: $\gamma = 1 - \alpha$

(ii) Aumentar o tamanho da amostra.

Dessas duas opções, a primeira não é recomendável porque aumenta-se α,que é o risco de que μ não esteja no intervalo.

Para resolver esse impasse, considere a seguinte figura que esquematiza a idéia de Erro de Estimação.



O comprimento ou amplitude do intervalo é:

$$L = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \tag{5}$$

O Erro Máximo da Estimação, denotado por E, é:

$$E = \frac{L}{2} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \tag{6}$$

Da equação 6 é possível obter n se o erro máximo de estimação E, o risco α e a variância populacional são conhecidos. Ou seja,

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} \tag{7}$$

Se a amostragem é sem reposição, é introduzido o fator de correção de população finita: $\sqrt{\frac{N-n}{N-1}}$, de onde:

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \tag{8}$$

que ao resolver em n, tem-se

$$n = \frac{N z_{\alpha/2}^2 \sigma^2}{E^2 (N-1) + z_{\alpha/2}^2 \sigma^2} \tag{9}$$

Observação: se o tamanho da população finita N é muito maior em comparação com n (isto é, $\frac{n}{N} < 0,10$) o fator de correção de população finita pode ser ignorado.

Exemplo - [Montgomery e Runger(2016)]

Para ilustrar o uso desse procedimento, considere o primeiro exemplo e suponha que quiséssemos determinar qual o tamanho amostral para assegurar que o IC de 95% para μ do aço, tivesse um erro de estimação de 0,5 J. Como proceder nesse caso?

Intervalo de Confiança para a média de uma população normal com variância desconhecida

Se $X_1, \ldots, X_n$ é uma amostra aleatória de tamanho n, de uma população normal com média μ e variância desconhecida σ^2 a variável aleatória,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

tem distribuição t-Student com $n-1$ graus de liberdade. Seguindo o procedimento anterior, para o nível de confiança fixado,$100(1-\alpha)\%$ tal que $0 < \alpha < 1$, pode-se encontrar um valor de $t_{\frac{\alpha}{2}, n-1}$, tal que

$$P\left(-t_{\frac{\alpha}{2}, n-1} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2}, n-1}\right) = 1 - \alpha, \tag{10}$$

onde $t_{\frac{\alpha}{2}, n-1}$, é obtido da tabela de distribuição t-Student com $n-1$ graus de liberdade.

Logo, o intervalo de confiança para μ, com coeficiente de confiança $100(1-\alpha)\%$ é dado por:

$$IC(\mu; 1-\alpha) = \left(\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}; \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right). \quad (11)$$

Exemplo - [Montgomery e Runger(2016)]

Um artigo no periódico *Materials Engineering* (1989, Vol. II, N^o. 4, pp. 275-281) descreve os resultados de testes trativos de adesão em 22 corpos-de-prova. A média amostral para a carga no ponto de falha do corpo-de-prova é $\bar{x} = 13,71$ megapascal, e o desvio-padrão da amostra é $s = 3,55$ megapascal. Sabendo-se que a população segue uma distribuição normal, encontre um IC de 95% para μ.

Intervalo de Confiança para a média com amostras grandes

A aplicação do Teorema Central do Limite permite a obtenção de intervalos de confiança para μ, quando a distribuição das variáveis aleatórias que constituem a amostra não é dada pelo modelo normal.

Exemplo - [Cancho(2010)]

Um provedor de acesso à internet está monitorando a duração do tempo das conexões de seus clientes com o objetivo de dimensionar seu equipamento. Suponha que são desconhecidos a média e a distribuição de probabilidade desse tempo, mas a variância, por analogia com outra serviços é considerada como sendo igual a 50 $(minutos)^2$. Uma amostra de 500 conexões resultou num valor observado médio de 25 minutos. O que dizer da verdadeira média com confiança de 95%?

Intervalo de Confiança para a variância de uma população normal

Algumas vezes, são necessários intervalos de confiança para a variância da população. O seguinte resultado fornece a base para construir esses intervalos de confiança:

Se $X_1, \ldots, X_n$ é uma amostra aleatória de tamanho n, de uma população normal com média μ e variância σ^2, ambas desconhecidas, e seja S^2 a variância da amostra. Então a variável aleatória

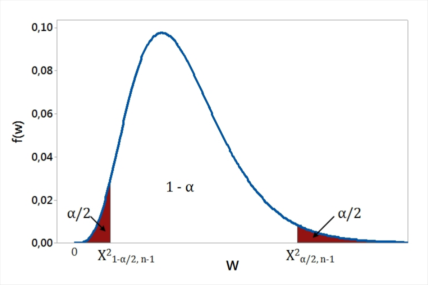
$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Ou seja, a variável aleatória W tem distribuição Qui-quadrado (χ^2) com $n-1$ graus de liberdade.

A construção do intervalo de confiança de $100 \times (1-\alpha)\%$ para σ^2 é direta. Uma vez que

$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

e, observando as probabilidades da figura a seguir, podemos escrever:



$$P\left(\chi^2_{1-\frac{\alpha}{2},n-1} \leq W \leq \chi^2_{\frac{\alpha}{2},n-1}\right) = P\left(\chi^2_{1-\frac{\alpha}{2},n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},n-1}\right) =$$

$$= 1 - \alpha$$

Essa última equação pode ser rearranjada como

$$P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right) = 1-\alpha.$$

Logo, o intervalo de $(1-\alpha)\times 100\%$ de confiança para σ^2 é dado por

$$IC(\sigma^2; 1-\alpha) = \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right).$$

Exemplo - [Montgomery e Runger(2016)]

Uma máquina automática de enchimento é usada para encher garrafas *pet*. Uma amostra aleatória de 20 garrafas resulta em uma variância amostral de volume de enchimento de $s^2 = 0,0153$ (onça fluida)^2. Se a variância do volume de enchimento for muito grande, existirá uma proporção inaceitável de garrafas cujo enchimento não foi completo e cujo enchimento foi em demasia. Consideraremos que o volume de enchimento seja distribuído de forma aproximadamente normal, construa um intervalo de 95% de confiança para a variância populacional.

Intervalo de Confiança para a proporção populacional com amostra grande

Considere uma população dicotômica, constituída apenas por elementos de dois tipos (por exemplo, peças defeituosas ou não defeituosas). O valor de p, que corresponde à proporção de elementos de um dos dois tipos na população (por exemplo, peças defeituosas) é definido como **proporção populacional**.

Se dessa população for retirada uma amostra aleatória de tamanho n, então $\hat{p} = Y/n$ será uma proporção amostral, sendo Y o número de elementos de um tipo na amostra (por exemplo, número de peças defeituosas), o que pode ser interpretado como número de sucessos em n ensaios de Bernoulli.

Nessas condições, a variável aleatória Y segue uma distribuição Binomial com parâmetros n e p.

De acordo com Teorema Central do Limite, para n suficientemente grande, a distribuição de Y (número de elementos de um tipo contidos na amostra) aproxima-se de uma distribuição normal com média np e variância $np(1-p)$.

Daí é imediato verificar , que a proporção amostral $\hat{p}$ também aproxima-se da distribuição normal com média p e variância $p(1-p)/n$, ou seja,

$$\hat{p} \sim N(p, \frac{p(1-p)}{n}), \tag{12}$$

Para construir o intervalo de confiança para p, note que

$$P\left(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) \approx 1-\alpha.$$

Fixando o nível de confiança $(1-\alpha)\times 100\%$ tal que $0 < \alpha < 1$, o intervalo de confiança para p, para amostras suficientemente grandes, é dado por:

$$IC(p; 1-\alpha) = \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right). \qquad (13)$$

Note que, nesse caso, os limites do intervalo dependem de p que é desconhecido. Assim sendo, o intervalo não pode ser calculado diretamente. Uma possível solução é substituir $p(1-p)$ por $\hat{p}(1-\hat{p})$ em (13). Assim o intervalo se reduz a

$$IC(p; 1-\alpha) = \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \qquad (14)$$

Uma outra abordagem é baseada no fato de que a expressão $p(1-p)$ assume o valor máximo igual $1/4$ quando $0 \leq p \leq 1$. Logo, o intervalo (13) se reduz a

$$IC(p; 1-\alpha) = \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{1}{4n}}\right) \tag{15}$$

Exemplo - [Montgomery e Runger(2016)]

Em uma amostra aleatória de 85 mancais de eixos de manivelas de motores de automóveis, 10 têm um acabamento de superfície que é mais rugoso do que as especificações permitidas. Determine um intervalo de 95% de confiança para a proporção de mancais na população que excede a especificação de rugosidade.

Determinação do tamanho da amostra para estimação de uma proporção populacional

A determinação do tamanho da amostra quando se quer estimar a proporção populacional é essencialmente a mesma descrita nos *slides* 20 e 21 para a determinação do tamanho da amostra na estimação de uma média populacional. Para isto, considere o erro máximo de estimação

$$E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}.$$

Supondo que p e risco α são conhecidos, tem-se:

$$n = \frac{z_{\alpha/2}^2 p(1-p)}{E^2}.$$

Se a população é finita e a amostragem é sem reposição

$$E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}},$$

de onde tem-se:

$$n = \frac{Nz^2_{\alpha/2}p(1-p)}{E^2(N-1)+z^2_{\alpha/2}p(1-p)}$$

Quando não se tem nenhuma informação de p, considera-se $p = 0,5$ Nesse caso o tamanho da amostra é:

$$n = \frac{0,25z^2_{\alpha/2}}{E^2}.$$

e

$$n = \frac{0,25Nz^2_{\alpha/2}}{E^2(N-1)+0,25z^2_{\alpha/2}}$$

Se $\frac{n}{N} < 0,1$ o fator de correção de população finita pode ser ignorado.

Exemplo 1 - [Montgomery e Runger(2016)]

Considere a situação do exemplo anterior. Quão grande deverá ser a amostra, se quisermos estar 95% confiantes de que o erro em usar $\hat{p}$ para estimar p é no máximo 0,05?

Exemplo 2 - [Montgomery e Runger(2016)]

E se quiséssemos estar no mínimo 95% confiantes de que nossa estimativa $\hat{p}$ da proporção verdadeira p estivesse dentro de 0,05 de erro, independentemente do valor de p?

Intervalo de Confiança para a diferença de médias de populações normais com variâncias conhecidas

Se $\bar{x_1} - \bar{x_2}$ forem as médias de duas amostras aleatórias independentes de tamanhos n_1 e n_2, provenientes de populações normais com variâncias conhecidas σ_1^2 e σ_2^2, respectivamente, então pode-se provar que um intervalo de confiança de $100 \times (1-\alpha)\%$ para $\mu_1 - \mu_2$ é

$$IC(\mu_1 - \mu_2; 1-\alpha) =$$

$$\left(\bar{x_1} - \bar{x_2} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; \bar{x_1} - \bar{x_2} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \tag{16}$$

Exemplo - [Montgomery e Runger(2016)]

Testes de resistência à tensão foram feitos em dois tipos diferentes de estruturas de alumínio. Essas estruturas foram usadas na fabricação das asas de um avião comercial.

De experiências passadas com o processo de fabricação dessas estruturas e com o procedimento de testes, os desvios-padrão das resistências à tensão são considerados conhecidos.

Os dados obtidos são os seguintes:

$n_1 = 10$	$\bar{x}_1 = 87,6$	$\sigma_1 = 1$
$n_2 = 12$	$\bar{x}_2 = 74,5$	$\sigma_2 = 1,5$

Se μ_1 e μ_2 denotarem as resistências médias verdadeiras à tensão para os dois tipos de estrutura, então escreva um intervalo de 90% de confiança para a diferença na resistência média $\mu_1 - \mu_2$.

Intervalo para a diferença de médias de populações normais com variâncias diferentes e desconhecidas

Se $\bar{x_1}$, $\bar{x_2}$, s_1^2 e s_2^2 forem as médias e as variâncias amostrais de duas amostras aleatórias de tamanho n_1 e n_2, respectivamente, provenientes de duas populações normais independentes, com variâncias desconhecidas e desiguais, então um intervalo de confiança de $100 \times (1-\alpha)\%$ para a diferença de médias $\mu_1 - \mu_2$ é

$$IC(\mu_1 - \mu_2; 1-\alpha) =$$

$$\left(\bar{x_1} - \bar{x_2} - t_{(\alpha/2,\nu)}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}; \bar{x_1} - \bar{x_2} + t_{(\alpha/2,\nu)}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) \tag{17}$$

Em que $t_{(\alpha/2,\nu)}$ é o ponto percentual superior $\alpha/2$ da distribuição t-Student, com ν graus de liberdade e,

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}} - 2.$$

Se ν não for um número inteiro, arredonde para o menor inteiro mais próximo.

Exemplo - Adaptado de [Cancho(2010)]

Um certo estudo descreve a quantidade de protopectina em tomates durante o armazenamento. Foram considerados dois períodos de armazenamento e analisaram-se as amostras de 10 lotes de tomates em cada período. Os dados resumidos apresentam-se na continuação:

Tempo de armazenamento	Média	Desvio Padrão
7 Dias	500	100
21 Dias	200	40

Considerando que o conteúdo de protopectina para os tempos de armazenamento tenha distribuição normal e que as variâncias verdadeiras são diferentes, construa um intervalo de 95% de confiança para a diferença de médias entre o tempo de armazenamento de 7 dias e 21 dias.

Intervalo de Confiança para a diferença de médias a partir de amostras emparelhadas

Um caso especial de testes t para duas amostras ocorre quando as observações nas duas populações de interesse são coletadas em pares.

Observação	X_{1j}	X_{2j}	d_j
1	x_{11}	x_{21}	d_1
2	x_{12}	x_{22}	d_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_{1n}	x_{2n}	d_n

Cada par de observação, como (X_{1j}, X_{2j}), é tomado sob condições homôgeneas, mas essas condições podem mudar de um par para outro. O procedimento de teste consistiria então em analisar as diferenças (d_j) entre as leituras de cada par de observação. Esse procedimento de teste é chamado de teste t emparelhado.

Se $\bar{d}$ e s_d forem a média e o desvio-padrão amostrais da diferença de n pares aleatórios de medidas distribuídas normalmente, então um intervalo de confiança de $100 \times (1-\alpha)\%$ para a diferença de médias $\mu_d = \mu_1 - \mu_2$ será

$$IC(\mu_d; 1-\alpha) = \left(\bar{d} - t_{\alpha/2,n-1} \cdot \frac{S_d}{\sqrt{n}}; \bar{d} + t_{\alpha/2,n-1} \cdot \frac{S_d}{\sqrt{n}}\right) \qquad (18)$$

sendo $t_{\alpha/2,n-1}$ o ponto percentual superior a $\alpha/2\%$ da distribuição t, com $n-1$ graus de liberdade.

Esse intervalo de confiança é válido também para o caso em que $\sigma_1^2 \neq \sigma_2^2$. Também é válido para amostras grandes ($n \geq 30$) (TLC).

Exemplo - [Montgomery e Runger(2016)]

O gerente de uma frota de carros está testando duas marcas de pneus radiais. Ele coloca, ao acaso, um pneu de cada marca nas duas rodas traseiras de oito carros e anda com carros até que os pneus se desgatem. Os dados (em quilômetros) são mostrados a seguir.

Carro	Marca 1	Marca 2	d_j
1	36.925	34.318	
2	45.300	42.280	
3	36.240	35.500	
4	32.100	31.950	
5	37.210	38.015	
6	48.360	47.800	
7	38.200	37.810	
8	33.500	33.215	

Encontre um intervalo de 99% de confiança para a diferença na vida-média. Baseado nos seus cálculos, qual marca você você prefere?

Intervalo de Confiança para a razão de variâncias de duas populações normais

Se s_1^2 e s_2^2 forem as variâncias amostrais de duas amostras aleatórias de tamanho n_1 e n_2, respectivamente, provenientes de duas populações normais independentes, com variâncias desconhecidas σ_1^2 e σ_2^2, então um intervalo de confiança de $100 \times (1-\alpha)\%$ para a razão $\frac{\sigma_1^2}{\sigma_2^2}$ é

$$IC(\frac{\sigma_1^2}{\sigma_2^2}; 1-\alpha) = \left(\frac{s_1^2}{s_2^2} \cdot f_{(1-\alpha/2, n_2-1, n_1-1)}; \frac{s_1^2}{s_2^2} \cdot f_{(\alpha/2, n_2-1, n_1-1)} \right). \quad (19)$$

Em que $f_{(\alpha/2, n_2-1, n_1-1)}$ e $f_{(1-\alpha/2, n_2-1, n_1-1)}$ são os pontos percentuais $\alpha/2$ superior e inferior da distribuição F, com n_2-1 graus de liberdade no numerador e n_1-1 graus de liberdade no denominador, respectivamente.

Exemplo - [Montgomery e Runger(2016)]

Uma companhia fabrica propulsores para uso em motores de turbinas a jato. Uma das operações envolve dar um acabamento, esmerilhando uma determinada superfície de um componente feito com liga de titânio.

Dois processos diferentes para esmerilhar podem ser usados, podendo produzir peças com iguais rugosidades médias na superfície.

Uma amostra aleatória de $n_1 = 11$ peças, proveniente do primeiro processo, resulta em um desvio-padrão de $s_1 = 5,1$ micropolegadas.

Uma amostra aleatória de $n_2 = 16$ peças, proveniente do segundo processo, resulta em um desvio-padrão de $s_2 = 4,7$ micropolegadas.

Encontre um intervalo de 90% de confiança para a razão dos dois desvios-padrão $\frac{\sigma_1}{\sigma_2}$.

Intervalo de Confiança para a diferença de proporções populacionais

Se $\hat{p_1}$ e $\hat{p_2}$ forem as proporções amostrais de observações em duas amostras aleatórias e independentes, de tamanho n_1 e n_2 que pertençam a uma classe de interesse, então um intervalo de $100 \times (1-\alpha)\%$ de confiança para a diferença das proporções verdadeiras $p_1 - p_2$ será

$$IC(p_1 - p_2; 1-\alpha) =$$

$$\left(\hat{p_1} - \hat{p_2} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}\right). \tag{20}$$

Sendo $z_{\alpha/2}$ o ponto percentual superior $\alpha/2$ da distribuição normal padrão.

Exemplo - [Montgomery e Runger(2016)]

Considere novamente o exemplo do slide 36 que trata da fabricação de mancais para eixos de manivelas.

Suponha que uma modificação seja feita no processo de acabamento da superfície e que, subsequentemente, obtenha-se uma segunda amostra aleatória de 85 eixos. O número de eixos defeituosos nessa segunda amostra é 8.

Obtenha um intervalo de 95% de confiança para a diferença da proporção de mancais defeituosos produzidos pelos dois processos $(p_1 - p_2)$.

Cancho, V., 2010. Notas de aulas sobre noções de estatística e probabilidade - São Paulo: USP.

Montgomery, D., Runger, G., 2016. Estatística Aplicada e Probabilidade para Engenheiros. Rio de Janeiro: LTC.