INFERÊNCIA ESTATÍSTICA (EST0035)

AULA04: ESTIMAÇÃO INTERVALAR

Frederico Machado Almeida
frederico.almeida@unb.br

Departamento de Estatística
Instituto de Exatas
Universidade de Brasília (UnB)

Estimação Intervalar

- Quando se utiliza um estimador pontual como melhor representante do parâmetro na amostra, surge sempre a questão de precisão (capacidade da estimativa cair abaixo ou acima do verdadeiro valor).
- Na prática, costuma-se estimar o erro-padrão do estimador, $\sigma_{\hat{\theta}_n}$ para representar tal medida.
- No entanto, há uma outra forma de se ter uma ideia clara do grau de precisão de um estimador, por meio da construção de um intervalo de confiança (IC) para o parâmetro $\theta \in \Theta$.
- Ou seja, os IC's permitem assim medir a precisão de um estimador. Entretanto, para sua construção, é de suma importância que se conheça a distribuição do estimador, $\hat{\theta}_n$.

Estimação Intervalar

- Enquanto os métodos de estimação pontual pressupõem obter um valor único (estimador) da amostra que possa representar da melhor forma a quantidade de interesse. Neste tópico vamos apresentar o método de estimação intervalar, ou genericamente, *estimação por conjuntos*.
- Ou seja, a inferência por conjuntos é a declaração de que $\theta \in \mathcal{C}(\mathbf{X}) = [L_i(\mathbf{X}); L_s(\mathbf{X})]$, com $L_i(\mathbf{X})$ e $L_s(\mathbf{X})$ sendo quantidades que dependem apenas da amostra aleatória (*aa*).
- Note que, $\mathcal{C}(\mathbf{X}) \in \Theta$ denota um intervalo (ou conjunto) de possíveis valores de θ, no qual seus limites independem do parâmetro de interesse, θ (também conhecido com estimador intervalar).

Estimação Intervalar

- Por serem funções da *aa*, os limites dos intervalos/conjuntos de confiança são variáveis aleatórias, e portanto, se $\mathbf{x}$ for um o vetor observado de $\mathbf{X}$, segue que, $\mathcal{C}(\mathbf{x}) = [L_i(\mathbf{x})\,;L_s(\mathbf{x})]$ será igualmente observado (estimativa intervalar).

Definição 1

Um estimador intervalar para o parâmetro θ é qualquer par de funções $L_i(\mathbf{X})$ e $L_s(\mathbf{X})$ da aa, que satisfazem $L_i(\mathbf{X}) < L_s(\mathbf{X})$. Se $\mathbf{X}$ for observado, então, $L_i(\mathbf{x}) \leq \theta \leq L_s(\mathbf{x})$.

Observação 1

Embora a definição (1) esteja referenciando valores finitos para $L_i(\mathbf{X})$ e $L_s(\mathbf{X})$, há situações (dependendo do interesse do pesquisador) em que o interesse é trabalhar com intervalos infinitos, do tipo, $(-\infty, L_s(\mathbf{X})]$ ou $[L_i(\mathbf{X}), +\infty)$.

- *Equivalentemente, os intervalos semiabertos podem ser dados por:* $\theta \leq L_s(\mathbf{X})$ *ou* $\theta \geq L_i(\mathbf{X})$.
- *Uma outra forma que algumas vezes será mais natural utilizar um intervalo aberto, da seguinte forma:* $\mathcal{C}(\mathbf{X}) = (L_i(\mathbf{X}); L_s(\mathbf{X}))$.

Estimação Intervalar

Exemplo 1: Assuma que $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, com σ^2 sendo um valor conhecido. Suponha que duas amostras de tamanho n tenham sido obtidas, e que as mesmas tenham fornecido as seguintes quantidades:

- Amostra 1: $\hat{\mu}_n = \bar{X}_n = 70$ unidades,
- Amostra 2: $\hat{\mu}_n = \bar{X}_n = 20$ unidades,

- A pergunta de interesse é: qual das duas estimativas é a mais confiável (em algum sentido)?
- Observe que nada se sabe sobre o tamanho de amostra n, muito menos do valor (supostamente conhecido) de σ^2. Porém, já sabemos à priori que $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$.
- Observe também que o erro aleatório ξ é tal que, $\xi = \bar{X}_n - \mu \sim \mathcal{N}(0, \sigma^2/n)$. Logo, $\sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma \sim \mathcal{N}(0, 1)$.

Estimação Intervalar

- Ao considerar o método de estimação intervalar, esperamos que a quantidade,
  - $\mathbb{P}\left(|\bar{X}_n - \mu| > \xi\right)$ seja pequena, mais precisamente α, e
  - $\mathbb{P}\left(|\bar{X}_n - \mu| \leq \xi\right)$ seja grande o suficiente, isto é, $1-\alpha$.
- Assim, considerando o caso anterior, onde $X_i \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, com σ^2 (conhecido), segue imediatamente que,

$$\mathbb{P}\left(|\bar{X}_n - \mu| \leq \xi\right) = 1 - \alpha.$$

Desenvolvendo a expressão anterior obtemos o seguinte resultado,

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

Onde $C\left(\mathbf{X}\right) = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} ; \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]$.

Definição 2 (Probabilidade de Cobertura)

Para um estimador intervalar $\mathcal{C}(\mathbf{X}) = [L_i(\mathbf{X}); L_s(\mathbf{X})]$ do parâmetro θ, a probabilidade de cobertura do intervalo é a proporção de vezes que $\mathcal{C}(\mathbf{X})$ cobre o parâmetro de interesse. Simbolicamente isso significa que $\mathbb{P}(\theta \in \mathcal{C}(\mathbf{x}))$.

Definição 3

A partir de um estimador intervalar $\mathcal{C}(\mathbf{X})$ de um parâmetro θ, o coeficiente de confiança $(1-\alpha)$ é o ínfimo da probabilidade de cobertura, isto é, $(1-\alpha) = \underset{\theta}{Inf}\, \mathbb{P}(\theta \in \mathcal{C}(\mathbf{X}))$.

Estimação Intervalar

- É importante lembrar que, apesar da notação que é frequentemente utilizada para denotar os IC's, digamos, $\mathbb{P}\left(\theta \in \mathcal{C}\left(\mathbf{X}\right)\right) \geq 1 - \alpha$, ou $\mathbb{P}\left(\theta \in \mathcal{C}\left(\mathbf{x}\right)\right) = 1 - \alpha$, tanto $\mathcal{C}\left(\mathbf{X}\right)$ quanto o parâmetro de interesse não estão associadas a nenhuma medida de probabilidade.
- Nisso, para não dar uma visão equivocada quanto a interpretação dos IC's, é conveniente denotá-lo da seguinte forma: $\mathbb{P}\left(L_i\left(\mathbf{X}\right) \leq \theta; L_s\left(\mathbf{X}\right) \geq \theta\right) = 1 - \alpha$.
- Associado ao fato de que, tanto $\mathcal{C}\left(\mathbf{X}\right)$ quanto θ não estão associados a nenhuma medida de probabilidade, a interpretação do IC baseia-se no princípio da repetitividade de amostras.

Estimação Intervalar

Interpretação (errada): É correto afirmar que existe uma probabilidade de $1-\alpha$ de que o verdadeiro valor (do parâmetro) esteja no intervalo $\mathcal{C}(\mathbf{x}) = [L_i(\mathbf{x}); L_s(\mathbf{x})]$. No entanto, observe que na prática esta probabilidade é de 1 se $\theta \in \mathcal{C}(\mathbf{x})$ ou 0 se $\theta \notin \mathcal{C}(\mathbf{x})$.

Interpretação (correta): Se obtivermos várias amostras de mesmo tamanho e, para cada uma delas, calcularmos os correspondentes intervalos de confiança com coeficiente $1-\alpha$, esperamos que a proporção de intervalos que contenham o verdadeiro valor seja igual $(1-\alpha) \times 100$.

Estimação Intervalar

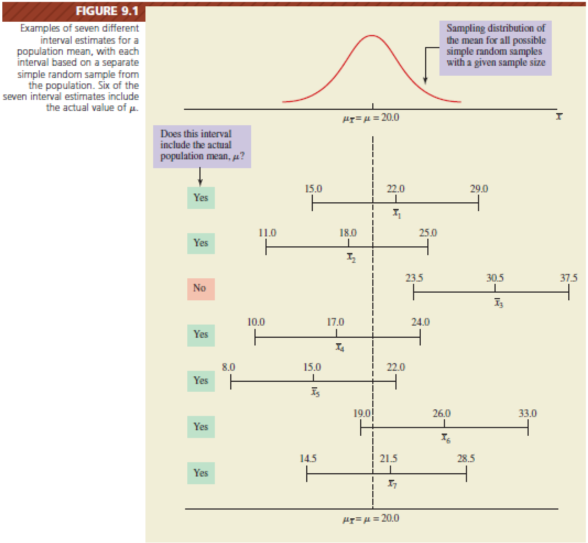


Figura: *Diferentes intervalos de confiança para μ.*

Quantidades Pivotais

- Existem vários métodos para construir IC's para algum parâmetro θ. Entre os quais se destacam:
  1. Quantidades pivotais.
  2. Métodos Bayesianos.
  3. Pivotando funções de distribuições acumuladas.
  4. Inversão dos testes de hipóteses.

Definição 4 (Quantidades Pivotais)

Seja $X_1, X_2, \cdots, X_n$ uma aa de $X \sim f(x|\theta)$. Uma variável $Q(\mathbf{X};\theta)$ é dita ser uma quantidade pivotal para o parâmetro θ, se sua distribuição não depender de θ, mesmo que sua forma funcional dependa.

Quantidades Pivotais

- A definição (4) indica que uma quantidade pivotal $Q\left(\mathbf{X};\theta\right)$ é uma função cuja sua forma é completamente especificada. Assim, para algum $a < b$, segue que, $\mathbb{P}\left(a \leq Q\left(\mathbf{X};\theta\right) \leq b\right)$.
- Exemplo de quantidades pivotais:
  - $Q\left(\mathbf{X};\theta\right) = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \sim \mathcal{N}\left(0,1\right)$.
  - $Q\left(\mathbf{X};\theta\right) = \left(\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}\right)^2 \sim \chi^2_{(1)}$, com μ conhecido.
  - $Q\left(\mathbf{X};\theta\right) = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$.
  - $Q\left(\mathbf{X};\theta\right) = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$, com μ conhecido.

Quantidades Pivotais

- Segue também que, se $X_i \sim Exp(\theta)$, com $\mathbb{E}(X_i) = \theta$. Então,
  - $Q(\mathbf{X};\theta) = \frac{1}{\theta}\sum_{i=1}^{n} X_i \sim Gama(n, 1)$, não depende de θ.
  - $Q(\mathbf{X};\theta) = \frac{2}{\theta}\sum_{i=1}^{n} X_i \sim \chi^2_{(2n)}$, não depende de θ.
- Observe que a distribuição da quantidade pivotal vai ser de suma importância, porque ela vai nos permitir obter os quantis de interesse.
- A obtenção da distribuição de $Q(\mathbf{X};\theta)$ pode ser exata (ou aproximada), dependendo do modelo onde os dados $\mathbf{X}$ foram extraídos, bem como do tamanho amostral.

Quantidades Pivotais

- Distribuição Normal Padrão: $Z \sim \mathcal{N}(0,1)$ então,
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

- Distribuição t-Student: $T \sim t_{(v)}$ então,
$$f_T(t|v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right), \text{ com } t \in \mathbb{R}, \ v \in \mathbb{N}.$$

- Distribuição Qui-Quadrado: $Y \sim \chi^2_{(v)}$ então,
$$f_Y(y|v) = \frac{1}{2^{v/2}\Gamma\left(\frac{v}{2}\right)} y^{v/2-1} \exp(-y/2), \text{ com } y > 0, \ v \in \mathbb{N}.$$

- Distribuição Gama: $W \sim Gama(n,1)$ então,
$$f_W(w|v) = \frac{1}{\Gamma(n)} w^{n-1} \exp(-w), \text{ com } w > 0.$$

Exemplo 2: Seja $X_1, \cdots, X_n$ uma *aa* de $X \sim f\left(x|\theta, \sigma^2 = 9\right)$, tal que $X \sim \mathcal{N}(\theta, 9)$. Então, $Q(\mathbf{X}, \theta) = \bar{X}_n - \theta$ é um pivô de μ. Tal que, $\mathbb{E}[Q(\mathbf{X}, \theta)] = 0$ e $Var[Q(\mathbf{X}, \theta)] = 9/n$. Logo, $Q(\mathbf{X}, \mu) \sim \mathcal{N}(0, 9/n)$ que não depende de θ.

- De forma similar, $Q(\mathbf{X}, \theta) = \frac{(\bar{X}_n - \theta)}{3/\sqrt{n}}$ é um pivô para θ.
- Note que, definindo $Q^*(\mathbf{X}, \theta) = \frac{\bar{X}_n}{\theta}$ temos que, $\mathbb{E}[Q^*(\mathbf{X}, \theta)] = 1$ e $Var[Q^*(\mathbf{X}, \theta)] = 9/n\theta^2$.
  Logo, $Q(\mathbf{X}, \theta) = \frac{\bar{X}_n}{\theta} \sim \mathcal{N}\left(1, \frac{9}{n\theta^2}\right)$ que depende de θ. Então, $Q^*(\mathbf{X}, \theta) = \frac{\bar{X}_n}{\theta}$ não é um pivô para θ.

Quantidades Pivotais

- Para construir um estimador intervalar para o parâmetro θ é de suma importância que se conheça à priori a distribuição do estimador pontual $\hat{\theta}_n$ de θ.
- Ademais, poder-se-á assim concluir igualmente que, para um intervalo de confiança, e necessário:

1. Encontrar um estimador pontual, $\hat{\theta}_n$.
2. Obter a respetiva quantidade pivotal, $Q\left(\mathbf{X};\theta\right)$.
3. Estabelecer um nível de confiança, $1-\alpha$.
4. Conhecer a dimensão da amostra, n.

Quantidades Pivotais

- Se $Q(\mathbf{X};\theta)$ é uma quantidade pivotal e tem *fdp*, então, para algum $1-\alpha$ fixo, denotando o nível de confiança, podemos escolher os quantis q_1 e q_2 que dependem de $1-\alpha$, tal que,

$$\mathbb{P}\left(q_1 \leq Q(\mathbf{X};\theta) \leq q_2\right) = 1-\alpha. \tag{1}$$

- Assim, para cada *aa* selecionada de $f(X|\theta)$, com $\theta \in \Theta$, temos que $q_1 \leq Q(\mathbf{X};\theta) \leq q_2$.
- Resolvendo a probabilidade apresentada na expressão (1), obtemos o seguinte intervalo, $L_i(\mathbf{X}) \leq g(\theta) \leq L_s(\mathbf{X})$. Com $L_i(\mathbf{X})$ e $L_s(\mathbf{X})$ dependendo de q_1 e q_2 mas não de θ.
- Uma estratégia para escolher os valores de q_1 e q_2 seria com o intuito de obter um par que garanta a obtenção de um intervalo ótimo.

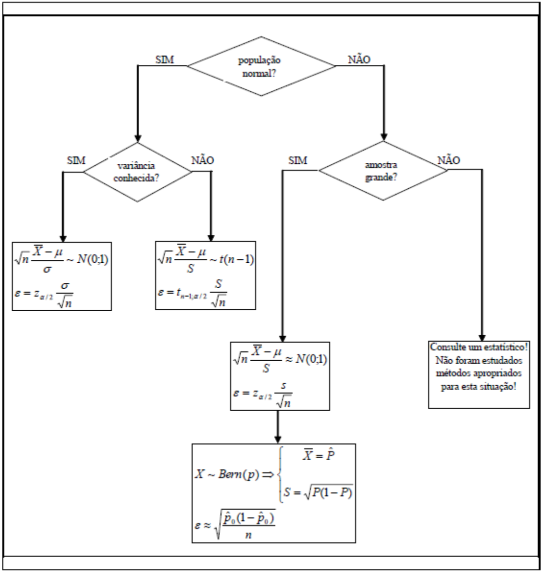
Alguns comentários importantes:

- O comprimento, ou amplitude de qualquer IC é dado por,

$$A = L_s(\mathbf{X}) - L_i(\mathbf{X}).$$

- Apesar de estamos cientes sobre a aleatoriedade dos limites de qualquer IC, o comprimento do mesmo, ou sua amplitude, nem sempre será aleatório.
- Caso a amplitude seja aleatória, precisamos escolher o par (q_1, q_2) que torna o comprimento médio menor possível.

Quantidades Pivotais



Amostragem de População Normal

Teorema 1

Seja $X_1, X_2, \cdots, X_n$ uma aa extraída de uma população normal, $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, com $\mu \in \mathbb{R}$ e $\sigma^2 > 0$. E seja $\bar{X}_n$ e S^2 a média e variância amostral, respectivamente. Então, segue que,

1. (i) *$\bar{X}_n$ e S^2 são estimadores não-tendenciosos para μ e σ^2.*
2. (ii) *A distribuição amostral de $\bar{X}_n$ é normal com média μ e variância σ^2/n.*
3. (iii) *$\bar{X}_n$ e S^2 são variáveis aleatórias independentes.*
4. (iv) *$(n-1)\, S^2/\sigma^2$ tem distribuição Qui-quadrado com $n-1$ graus de liberdade.*

Prova: a prova dos itens (*i*) e (*ii*) é trivial. A prova do item (*iii*) pode ser encontrada em *Bolfarine & Sandoval (2000, pag. 74)*. Por fim, a prova do item (*iv*) foi apresentada nas aulas anteriores.

Amostragem de População Normal

Exemplo 3: Seja $X_1, X_2 \cdots, X_n$ uma *aa* de $X \sim \mathcal{N}(\theta, 1)$. Considere a possibilidade de estimar a quantidade μ, com $Q(\mathbf{X}; \mu) = \frac{(\bar{X}_n - \mu)}{1/\sqrt{n}} = \sqrt{n}\left(\bar{X}_n - \mu\right) \sim \mathcal{N}(0, 1)$. Para algum par (q_1, q_2), e um nível de confiança $1 - \alpha$, derive o IC para μ, e apresenta o comprimento do mesmo.

Definição 5

Seja $X_1, X_2, \cdots, X_n$ uma aa proveniente de uma distribuição normal com média $\theta \in \mathbb{R}$ e variância $\sigma^2 > 0$. O processo de estimação intervalar dos dois parâmetros, é feito de acordo com os seguinte casos:

- C1 *Obter o IC para μ, supondo σ^2 conhecido.*
- C2 *Obter o IC para μ, supondo σ^2 desconhecido.*
- C3 *Obter o IC para σ^2, supondo μ desconhecido.*
- C4 *Obter o IC para σ^2, supondo μ conhecido.*
- C5 *Obter o IC simultâneo de μ e σ^2.*

Intervalo de Confiança para μ

- Sob a suposição de normalidade dos dados, ou seja, se $X_i \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, com $i \in \{1, 2, \cdots, n\}$ temos que, $\bar{X}_n \sim \mathcal{N}\left(\mu, \sigma^2/n\right)$. Portanto, supondo $\theta = \left(\mu, \sigma^2\right)^\top$, segue que:

$$Q\left(\mathbf{X}, \theta\right) = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \sim F. \qquad (2)$$

- Observe que, a distribuição da quantidade pivotal $Q\left(\mathbf{X}, \theta\right)$ depende do conhecimento ou não do σ^2.

Intervalo de Confiança para μ

CASO I: supondo σ^2 conhecido, segue imediatamente que, $Q(\mathbf{X}, \theta) \sim \mathcal{N}(0, 1)$. Portanto, o IC para μ será dado por:

$$\mathbb{P}\left(q_1 \leq Q(\mathbf{X}, \theta) \leq q_2\right) = \mathbb{P}\left(-z_{\alpha/2} \leq \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \leq z_{\alpha/2}\right) = 1 - \alpha$$

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

- Ou seja, q_1 e q_2 devem ser simétricos. Tal que, $\mathbb{P}\left(Z \leq -z_{\alpha/2}\right) = \mathbb{P}\left(Z \geq z_{\alpha/2}\right) = \alpha/2$.

Intervalo de Confiança para μ

CASO II: supondo σ^2 desconhecido, vimos anteriormente que, $S^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2$ é o melhor estimador de σ^2. Desta forma,

$$Q(\mathbf{X}, \theta) = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S} \sim t_{n-1}.$$

Assim, o IC para μ será dado por:

$$\mathbb{P}\left(q_1 \leq Q(\mathbf{X}, \theta) \leq q_2\right) =$$
$$\mathbb{P}\left(\bar{X}_n - t_{(n-1,\alpha/2)} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{(n-1,\alpha/2)} \frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

- Como a distribuição t-Student é um caso particular da normal, aqui o q_1 e q_2 são tais que, $\mathbb{P}\left(T_{n-1} \leq -t_{(n-1,\alpha/2)}\right) = \mathbb{P}\left(T_{n-1} \geq t_{(n-1,\alpha/2)}\right) = \alpha/2.$

Intervalo de Confiança para σ^2

Exemplo 4: Uma máquina de bebidas está regulada de modo a servir uma quantidade de liquido que é uma variável aleatória com distribuição aproximadamente normal. Sabendo que numa amostra de 25 bebidas se obtiveram os seguintes resultados:

$$\sum_{i=1}^{n} x_i = 6250\, ml \text{ e } \sum_{i=1}^{n} (x_i - \bar{x})^2 = 384\, ml.$$

Pretende-se,

a. Construir um intervalo de confiança a 95% para a verdadeira quantidade média de liquido das bebidas servidas.
b. Determinar quantas bebidas deveriam ser incluídas na amostra, se se pretendesse reduzir a amplitude do intervalo para 2 ml (*dica: considere os quantis da distribuição normal, i.e,* $z_{\alpha/2} = 1,96$).

Intervalo de Confiança para σ^2

- Como no caso da média, a construção do IC para o parâmetro de escala da distribuição normal, digamos σ^2 requer avaliar duas situações (ou casos):

CASO III: supondo μ desconhecido, vimos que, a média amostral $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ é o melhor estimador para μ. Ademais, vimos igualmente que a variância amostral S^2 é o melhor estimador para σ^2, de tal forma que,

$$\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = (n-1)\,S^2 \Longleftrightarrow \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 = \frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Com $Q\left(\mathbf{X}, \sigma^2\right) = \frac{(n-1)S^2}{\sigma^2}$ sendo um pivô de σ^2.

Intervalo de Confiança para σ^2

- Assim, escolhendo os quantis q_1 e q_2 da distribuição Qui-Quadrado com $n-1$ graus de liberdade obtemos o seguinte IC para σ^2.

$$\mathbb{P}\left(q_1 \leq Q\left(\mathbf{X}, \sigma^2\right) \leq q_2\right) = \mathbb{P}\left(q_1 \leq \frac{(n-1)\,S^2}{\sigma^2} \leq q_2\right) = 1-\alpha.$$

$$\mathbb{P}\left(\frac{(n-1)\,S^2}{q_2} \leq \sigma^2 \leq \frac{(n-1)\,S^2}{q_1}\right) = 1-\alpha. \text{ Ou seja,}$$

$IC_{1-\alpha}\left(\sigma^2\right) = \left[\frac{(n-1)S^2}{q_2}; \frac{(n-1)S^2}{q_1}\right]$ denota o IC para σ^2.

- Como a distribuição Qui-Quadrado é assimétrica positiva, segue que, $\mathbb{P}\left(Q \leq q_1\right) = \mathbb{P}\left(Q \geq q_2\right) = \alpha/2$. Mais concretamente, $q_1 = \chi^2_{(n-1,\alpha/2)}$ e $q_2 = \chi^2_{(n-1,1-\alpha/2)}$.

Intervalo de Confiança para σ^2

- Como $S^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$ é uma estatística (ou estimador), o intervalo obtido anteriormente é aleatório. Nesse caso, um intervalo observado para σ^2 tem o seguinte formato:

$$IC_{1-\alpha}\left(\sigma^2\right) = \left[\frac{(n-1)\,s^2}{q_2}; \frac{(n-1)\,s^2}{q_1}\right].$$

  Com $s^2 = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$ denotando uma estimativa de S^2.

- Sem perda de generalidade, um intervalo de confiança para o desvio-padrão, σ é dado por:

$$IC_{1-\alpha}(\sigma) = \left[S\sqrt{\frac{(n-1)}{q_2}}; S\sqrt{\frac{(n-1)}{q_1}}\right].$$

Intervalo de Confiança para σ^2

- Adicionalmente, o intervalo de confiança observado para o desvio-padrão tem o seguinte formato:

$$IC_{1-\alpha}(\sigma) = \left[s\sqrt{\frac{(n-1)}{q_2}}; s\sqrt{\frac{(n-1)}{q_1}} \right].$$

CASO IV: o caso em que μ é conhecido será omitido aqui, mas segue a mesma ideia do CASO III.

Intervalo de Confiança para σ^2

Exemplo 5: A administração do transporte Metropolitano defronta uma situação de irregularidade na hora de passagem dos comboios pelas diversas estações. Essa irregularidade (em segundos) pode ser descrita por uma variável aleatória normal cuja media μ variância σ^2 ambos são desconhecidos. Com $n = 22$ e $s^2 = 9$, pretende-se saber entre que valores se situa a variância, com nível de confiança de $0,99$.

Intervalos de Confiança Assintóticos

Definição 6

Seja $X_1, X_2, \cdots, X_n$ uma aa extraída de uma população caracterizada por uma fdp (ou fp) $f(x|\theta)$, com $\theta \in \Theta$. Assuma que $\hat{\theta}_n = \hat{\theta}(X_1, \cdots, X_n)$ é um estimador pontual de θ, e que $I_n(\theta) = nI_1(\theta) < \infty$ denota a informação de Fisher.

- *Segue do TCL que, $\hat{\theta}_n \overset{a}{\sim} \mathcal{N}\left(\theta, Var\left(\hat{\theta}_n\right)\right)$ com $Var\left(\hat{\theta}_n\right) = I_n^{-1}(\theta)$.*
- *Como consequência do resultado anterior, obtemos:*

$$Q(\mathbf{X}, \theta) = \frac{\left(\hat{\theta}_n - \theta\right)}{\sqrt{I_n^{-1}(\theta)}} \overset{a}{\sim} \mathcal{N}(0, 1).$$

Intervalos de Confiança Assintóticos

- Segue da normalidade assintótica que,

$$\mathbb{P}\left(-z_{\alpha/2} \leq Q\left(\mathbf{X}, \theta\right) \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\left(\hat{\theta}_n - \theta\right)}{\sqrt{I_n^{-1}\left(\theta\right)}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$\mathbb{P}\left(\hat{\theta}_n - z_{\alpha/2}\sqrt{I_n^{-1}\left(\theta\right)} \leq \theta \leq \hat{\theta}_n + z_{\alpha/2}\sqrt{I_n^{-1}\left(\theta\right)}\right) = 1 - \alpha.$$

Fornecendo o seguinte intervalo de confiança assintótico:

$$IC_{1-\alpha}\left(\theta\right) = \left[\hat{\theta}_n - z_{\alpha/2}\sqrt{I_n^{-1}\left(\theta\right)}; \hat{\theta}_n + z_{\alpha/2}\sqrt{I_n^{-1}\left(\theta\right)}\right]$$

Intervalos de Confiança Assintóticos

- Como θ é uma quantidade desconhecida, e porque os limites dos IC's não podem depender de θ, obtemos o seguinte *intervalo de confiança assintótico* para θ.

$$Q(\mathbf{X}, \theta) = \frac{\left(\hat{\theta}_n - \theta\right)}{\sqrt{I_n^{-1}\left(\hat{\theta}_n\right)}} \stackrel{a}{\sim} \mathcal{N}(0, 1).$$

Fornecendo o seguinte intervalo de confiança assintótico para θ.

$$IC_{1-\alpha}(\theta) = \left[\hat{\theta}_n - z_{\alpha/2}\sqrt{I_n^{-1}\left(\hat{\theta}_n\right)}; \hat{\theta}_n + z_{\alpha/2}\sqrt{I_n^{-1}\left(\hat{\theta}_n\right)}\right].$$

Intervalos de Confiança Assintóticos

Exemplo 6: Seja $X \sim Ber\left(\theta\right)$, tal que, $\theta = \mathbb{E}\left(X\right) = \mathbb{P}\left(X = 1\right)$. Supondo que $X_1, X_2, \cdots, X_n$ são cópias iid's de X, então, $\hat{\theta}_n = \sum_{i=1}^{n} X_i/n \equiv \bar{X}_n$ é um estimador não-viesado para θ. Portanto, segue do TCL que,
$\hat{\theta}_n \overset{a}{\sim} \mathcal{N}\left(\theta, \theta\left(1-\theta\right)/n\right)$.

Desta forma, $Q\left(\mathbf{X}, \theta\right) = \frac{\hat{\theta}_n - \theta}{\sqrt{\frac{\hat{\theta}_n\left(1-\hat{\theta}_n\right)}{n}}} \overset{a}{\sim} \mathcal{N}\left(0, 1\right)$. Resultando no seguinte IC para θ:

$$IC_{1-\alpha}\left(\theta\right) = \left[\hat{\theta}_n - z_{\alpha/2}\sqrt{\hat{\theta}_n\left(1-\hat{\theta}_n\right)/n}; \hat{\theta}_n + z_{\alpha/2}\sqrt{\hat{\theta}_n\left(1-\hat{\theta}_n\right)/n}\right].$$

- Quando $\hat{\theta}_n \neq 1/2$ o intervalo de confiança será otimista.
- Quando $\hat{\theta}_n = 1/2$ a variância será máxima e o intervalo de confiança será conservador.

Intervalos de Confiança Assintóticos

Exemplo 7: Numa pesquisa com 50 eleitores, o candidato José João obteve 0,34 da preferência dos eleitores. Construa, para a confiança 94%, os intervalos otimista e conservador de confiança para a proporção de votos a serem recebidos pelo candidato.

Inferência para Duas Populações

- Em determinadas situações é de interesse do pesquisador comparar (ou fazer inferências) sobre as características de duas ou mais populações (distribuições).
- Em particular, vamos apresentar formas de como podemos construir intervalos de confiança para a diferença de médias e proporções populacionais, bem como a razão de variâncias.
- Num primeiro momento, nosso interesse reside em comparar as médias de duas distribuições, digamos $\mathcal{D}_1$ e $\mathcal{D}_2$.

Inferência para Duas Populações

Definição 7 (População Normal)

Assuma agora que X e Y são variáveis aleatórias proveniente de uma distribuição normal, tal que, $X \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ e $Y \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$, com $\mu_1, \mu_2 \in \mathbb{R}$ e $\sigma_1^2, \sigma_2^2 < \infty$. Supondo que $X_1, X_2, \cdots, X_{n_1}$ e $Y_1, Y_2, \cdots, Y_{n_2}$ denotam cópias iid's de X e Y, segue imediatamente que, $\bar{X}_n \sim \mathcal{N}\left(\mu_1, \sigma_1^2/n_1\right)$ e $\bar{Y}_n \sim \mathcal{N}\left(\mu_2, \sigma_2^2/n_2\right)$.

- Em decorrência ao resultado da definição 7, segue imediatamente que:

$$Q\left(\mathbf{X}, \mathbf{Y}, \Delta\right) = \frac{\hat{\Delta}_n - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\left(\bar{X}_n - \bar{Y}_n\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (3)$$

- Como X e Y são variáveis aleatórias com distribuição normal, a inferência sobre Δ será feita considerando três casos.

Inferência para Duas Populações

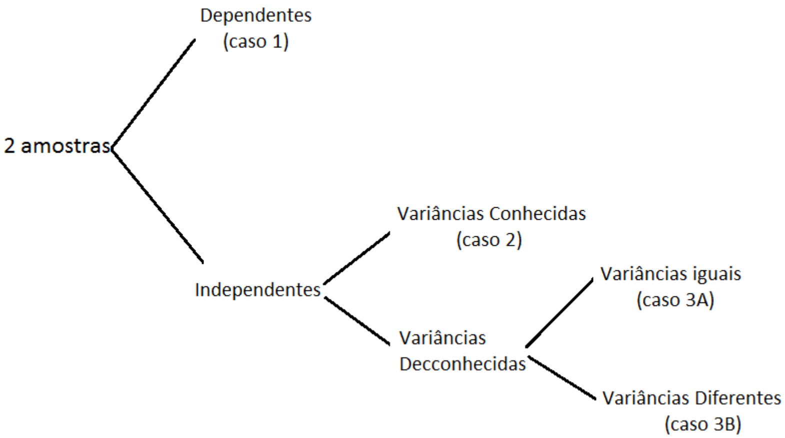


Figura: Inferência para a diferença de médias de suas populações.

CASO I: supondo $\sigma_1^2 \neq \sigma_2^2$ sendo ambos conhecidos, a quantidade pivotal apresentada na equação 3 terá uma distribuição normal padrão. A construção dos intervalos de confiança é trivial. Ou seja,

$$Q(\mathbf{X}, \mathbf{Y}, \Delta) = \frac{\hat{\Delta}_n - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\left(\bar{X}_n - \bar{Y}_n\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

Assim, o intervalo de confiança para Δ será dado por:

$$\mathbb{P}(q_1 \leq Q(\mathbf{X}, \mathbf{Y}, \Delta) \leq q_2) = \mathbb{P}\left(q_1 \leq \frac{\hat{\Delta}_n - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq q_2\right) = 1 - \alpha$$

$$\mathbb{P}\left(\hat{\Delta}_n - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \Delta \leq \hat{\Delta}_n + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha.$$

Inferência para Duas Populações

- Desta forma, o intervalo de confiança assintótico para $\Delta = \mu_1 - \mu_2$ é dado por,

$$IC_{1-\alpha}(\Delta) \;=\; \left[\hat{\Delta}_n - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}};\, \hat{\Delta}_n + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right],$$

  onde $\hat{\Delta}_n = \bar{X}_n - \bar{Y}_n$ denota um estimador pontual para o parâmetro Δ.

- Supondo que $\hat{\delta}_n = \bar{x}_n - \bar{y}_n$ é uma estimativa para Δ (depois de observar os dados), uma estimativa intervalar será dado por:

$$IC_{1-\alpha}(\Delta) \;=\; \left[\hat{\delta}_n - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}};\, \hat{\delta}_n + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right],$$

Exemplo 8: Suponha que estamos testando a resistência à tração de 10 barras de aço produzidas pelo fabricante 1 e 15 barras produzidas pelo fabricante 2. A partir de experiências anteriores sabemos que o fabricante 1 produz barras de aço cuja resistência à tração tem variância de 900 kgf^2/cm^2 enquanto para o fabricante 2 este valor é 625 kgf^2/cm^2. As amostras nos forneceram resistências médias de 5000 kgf/cm^2 e 4800 kgf^2/cm^2 respectivamente para os fabricantes 1 e 2. Construa um intervalo de confiança de 90% para a diferença entre as médias $\Delta = \mu_1 - \mu_2$.

CASO II: supondo $\sigma_1^2 \neq \sigma_2^2$ ambos desconhecidos, os estimadores pontuais S_1^2 e S_2^2 podem ser considerados para construir o intervalo de confiança para Δ.

- Portanto, a quantidade pivotal é dada por:

$$Q(\mathbf{X}, \mathbf{Y}, \Delta) = \frac{\hat{\Delta}_n - \Delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{\left(\bar{X}_n - \bar{Y}_n\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(v)},$$

  com $v = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$ denotando os graus de liberdade da distribuição t-Student.

- O intervalo de confiança para $\Delta = \mu_1 - \mu_2$ nas circunstâncias supracitadas é dado por:

$$\mathbb{P}\left(q_1 \leq Q\left(\mathbf{X}, \mathbf{Y}, \Delta\right) \leq q_2\right) = \mathbb{P}\left(q_1 \leq \frac{\hat{\Delta}_n - \Delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \leq q_2\right) = 1 - \alpha,$$

onde os quantis q_1 e q_2 da distribuição t-Student são tais que, $q_2 = t_{(v,\alpha/2)}$ e $q_1 = -q_2 = -t_{(v,\alpha/2)}$. Desta forma,

$$IC_{1-\alpha}\left(\Delta\right) = \left[\hat{\Delta}_n - t_{(v,\alpha/2)}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}; \hat{\Delta}_n + t_{(v,\alpha/2)}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right].$$

CASO III: supondo $\sigma_1^2 = \sigma_2^2 = \sigma^2$, isso implica que $\bar{X}_n \sim \mathcal{N}\left(\mu_1, \sigma^2/n\right)$ e $\bar{Y}_n \sim \mathcal{N}\left(\mu_2, \sigma^2/n\right)$. Portanto, segue dos resultados anteriores que,

$$Var\left(\hat{\Delta}_n\right) = Var\left(\bar{X}_n\right) + Var\left(\bar{Y}_n\right) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Desta forma, segue que $\hat{\Delta}_n \sim \mathcal{N}\left(\Delta, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$. Assim, (i) supondo σ^2 conhecido, obtemos:

$$Q\left(\mathbf{X}, \mathbf{Y}, \Delta\right) = \frac{\hat{\Delta}_n - \Delta}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}\left(0, 1\right),$$

o intervalo de confiança para Δ será dado por:

$$IC_{1-\alpha}\left(\Delta\right) = \left[\hat{\Delta}_n - z_{\alpha/2}\sigma\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right); \hat{\Delta}_n + z_{\alpha/2}\sigma\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)\right],$$

(ii) supondo σ^2 desconhecido (o que acontece com frequência), a quantidade pivotal $Q(\mathbf{X}, \mathbf{Y}, \Delta)$ será obtida considerando o estimador não-viesado $\hat{\sigma}^2$ de σ^2. Isto é,

$$Q(\mathbf{X}, \mathbf{Y}, \Delta) = \frac{\hat{\Delta}_n - \Delta}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\hat{\Delta}_n - \Delta}{S_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2},$$

com $\hat{\sigma}^2 = S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ denotando a variância ponderada (pooled variance).

- Como S_1^2 e S_2^2 são estimadores consistentes para σ^2, segue imediatamente que S_p^2 é igualmente consistente para σ^2. Adicionalmente,

$$\frac{(n_1 + n_2 - 2)\, S_p^2}{\sigma^2} = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{\sigma^2} \sim \chi^2_{n-2}.$$

Se q_1 e q_2 denotam os quantis da distribuição t-Student com $n_1 + n_2 - 2$ graus de liberdade, segue que,

$$\mathbb{P}\left(q_1 \leq Q\left(\mathbf{X}, \mathbf{Y}, \Delta\right) \leq q_2\right) = \mathbb{P}\left(q_1 \leq \frac{\hat{\Delta}_n - \Delta}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \leq q_2\right) = 1 - \alpha.$$

Resultando no seguinte intervalo de confiança para Δ,

$$IC_{1-\alpha}\left(\Delta\right) = \left[\hat{\Delta}_n - t_{(n-2,\alpha/2)} S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}; \hat{\Delta}_n + t_{(n-2,\alpha/2)} S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right]$$

onde $n = n_1 + n_2$.

Exemplo 9: suponha que $n_1 = 10$; $n_2 = 7$; $\bar{x} = 4,2$; $\bar{y} = 3,4$; $s_1^2 = 49$ e $s_2^2 = 32$. Supondo que as amostras aleatórias foram extraídas de duas distribuições normais, tais que, $X \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ e $X \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$, pede-se para:

a. Construir um intervalo de confiança para Δ supondo $\sigma_1^2 \neq \sigma_2^2$ e desconhecidos.
b. Construir um intervalo de confiança para Δ supondo $\sigma_1^2 = \sigma_2^2 = \sigma^2$, sendo σ^2 desconhecido.

Observação 2

Obs3.1: Observe que, supondo que X e Y são variáveis aleatórias que não seguem uma distribuição normal, é suficiente usar a consistência de S_p^2 para construir um intervalo de confiança assintótico para Δ.

Obs3.2: Quando os intervalos de confiança são aproximados (assintóticos), as quantidades pivotais que forneceram tais intervalos são igualmente assintóticos.

Definição 8 (IC para a Diferença de Proporções)

Seja X e Y duas variáveis aleatórias independentes, e ambas com distribuição Bernoulli. Isto é, $X \sim Ber\left(\theta_1\right)$ e $Y \sim Ber\left(\theta_2\right)$, tal que, $\mathbb{P}\left(X = 1\right) = \theta_1 \in \Theta$ e $\mathbb{P}\left(Y = 1\right) = \theta_2 \in \Theta$. Portanto, nosso interesse é fazer inferências sobre a quantidade $\Delta = \theta_1 - \theta_2 \in \Theta$.

Assuma que $X_1, X_2, \cdots, X_{n_1}$ e $Y_1, Y_2, \cdots, Y_{n_2}$ denotam cópias iid's de X e Y, segue que, $\hat{\theta}_{1n} = \sum_{i=1}^{n_1} X_i/n_1 \equiv \bar{X}_n$ e $\hat{\theta}_{2n} = \sum_{i=1}^{n_2} Y_i/n_2 \equiv \bar{Y}_n$ são estimadores pontuais para θ_1 e θ_2, respectivamente.

Inferência para Duas Populações

- Como a variância de $\bar{X}_n$ e $\bar{Y}_n$ é tal que, $Var\left(\bar{X}_n\right) = \theta_1\left(1-\theta_1\right)/n_1$ e $Var\left(\bar{Y}_n\right) = \theta_2\left(1-\theta_2\right)/n_2$, segue do TCL que, $\bar{X}_n \overset{a}{\sim} \mathcal{N}\left(\theta_1, \theta_1\left(1-\theta_1\right)/n_1\right)$ e $\bar{Y}_n \overset{a}{\sim} \mathcal{N}\left(\theta_2, \theta_2\left(1-\theta_2\right)/n_2\right)$.

- Como $\bar{X}_n$ e $\bar{Y}_n$ são estimadores consistentes para θ_1 e θ_2, segue que,

$$\hat{\Delta}_n \overset{a}{\sim} \mathcal{N}\left(\Delta, Var\left(\hat{\Delta}_n\right)\right),$$

  com $Var\left(\hat{\Delta}_n\right) = Var\left(\bar{X}_n\right) + Var\left(\bar{Y}_n\right) = \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}$. Como θ_1 e θ_2 são quantidades desconhecidas, segue que,

$$Var\left(\hat{\Delta}_n\right) = \frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n_1} + \frac{\bar{Y}_n\left(1-\bar{Y}_n\right)}{n_2}.$$

Inferência para Duas Populações

- Com base na informação anterior, a quantidade pivotal é dada por:

$$Q(\mathbf{X},\mathbf{Y},\Delta) = \frac{\hat{\Delta}_n - \Delta}{\sqrt{Var\left(\hat{\Delta}_n\right)}} = \frac{\left(\bar{X}_n - \bar{Y}_n\right) - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n_1} + \frac{\bar{Y}_n\left(1-\bar{Y}_n\right)}{n_2}}} \overset{a}{\sim} \mathcal{N}(0,1),$$

e portanto, o intervalo de confiança para Δ é dado por:

$$\mathbb{P}\left(q_1 \leq \frac{\left(\bar{X}_n - \bar{Y}_n\right) - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n_1} + \frac{\bar{Y}_n\left(1-\bar{Y}_n\right)}{n_2}}} \leq q_2\right) = 1 - \alpha.$$

Resultando no seguinte intervalo de confiança assintótico para $\Delta = \theta_1 - \theta_2$,

$$IC_{1-\alpha}(\Delta) = \left[\hat{\Delta}_n - z_{\alpha/2}\sqrt{\frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n_1} + \frac{\bar{Y}_n\left(1-\bar{Y}_n\right)}{n_2}}; \hat{\Delta}_n + z_{\alpha/2}\sqrt{\frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n_1} + \frac{\bar{Y}_n\left(1-\bar{Y}_n\right)}{n_2}}\right].$$

Inferência para Duas Populações

Exemplo 9: Suponha que X e Y são variáveis aleatórias independentes com distribuição Bernoulli, i.e., $X \sim Ber(\theta_1)$ e $Y \sim Ber(\theta_2)$. Assuma que, $n_1 = 100$; $n_2 = 400$; $S_x = \sum_{i=1}^{n_1} X_i = 30$ e $S_y = \sum_{i=1}^{n_2} Y_i = 80$. Então, pede-se para construir um intervalo de confiança de 93% para a diferença de proporções.

Intervalo de Confiança para a Razão de Variâncias (σ_1^2/σ_2^2)

Para encontrar o intervalo de confiança para σ_1^2/σ_2^2, lembremos que a primeira suposição é que $X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_n$ são amostras aleatórias proveniente de distribuições normais normal com média μ_1 e μ_2, e variâncias σ_1^2 e σ_2^2. Supondo μ_1 e μ_2 ambas desconhecidas, segue que,

$$F = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \sim F_{(v_2, v_1)},$$

é uma distribuição F com $v_2 = n_2 - 1$ e $v_1 = n_1 - 1$ os graus de liberdades para o numerador e denominador, respectivamente.

Denotemos por $f_{(1-\alpha/2, v_2, v_1)}$ e $f_{(\alpha/2, v_2, v_1)}$ denotam os quantis (ou valores críticos da distribuição F). Sendo,

$$f_{(1-\alpha/2, v_2, v_1)} = 1/f_{(\alpha/2, v_1, v_2)}$$

Intervalo de Confiança para a Razão de Variâncias (σ_1^2/σ_2^2)

Segue então que, se S_1^2 e S_2^2 são as variâncias amostrais. Assim,

$$\begin{aligned}
1-\alpha &= P\left(f_{(\alpha/2,v_2,v_1)} \leq F \leq f_{(1-\alpha/2,n_2-1,n_1-1)}\right)\\
&= P\left(f_{(\alpha/2,v_2,v_1)} \leq \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \leq f_{(1-\alpha/2,n_2-1,n_1-1)}\right)\\
&= P\left(f_{(\alpha/2,v_2,v_1)} \leq \frac{S_2^2}{\sigma_2^2}\times\frac{\sigma_1^2}{S_1^2} \leq f_{(1-\alpha/2,v_2,v_1)}\right)\\
&= P\left(\frac{S_1^2}{S_2^2}f_{(\alpha/2,v_2,v_1)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2}f_{(1-\alpha/2,v_2,v_1)}\right).
\end{aligned}$$

Ou seja, o intervalo de $(1-\alpha)\,100\%$ de confiança para $\frac{\sigma_1^2}{\sigma_2^2}$ é dado por:

$$IC_{1-\alpha}\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left[\frac{S_1^2}{S_2^2}f_{(\alpha/2,v_2,v_1)};\,\frac{S_1^2}{S_2^2}f_{(1-\alpha/2,v_2,v_1)}\right]$$

Intervalo de Confiança para a Razão de Variâncias (σ_1^2/σ_2^2)

O intervalo de confiança para a razão de variâncias σ_1^2/σ_2^2 é dado por:

$$IC_{1-\alpha}(\sigma_1^2/\sigma_2^2) = \left[\frac{S_1^2}{S_2^2} f_{(\alpha/2, v_2, v_1)}; \frac{S_1^2}{S_2^2} f_{(1-\alpha/2, v_2, v_1)}\right]. \qquad (4)$$

Os valores críticos são tais que $f_{(\alpha/2, v_2, v_1)}$ é o valor crítico inferior e $f_{(1-\alpha/2, v_2, v_1)}$ o superior, respectivamente.

Entretanto, o IC para a razão dos desvios-padrão σ_1/σ_2 é obtido extraindo a raiz quadrada da equação 4. Isto é,

$$IC_{1-\alpha}(\sigma_1/\sigma_2) = \left[\sqrt{\frac{S_1^2}{S_2^2} f_{(\alpha/2, v_2, v_1)}}; \sqrt{\frac{S_1^2}{S_2^2} f_{(1-\alpha/2, v_2, v_1)}}\right].$$

Intervalo de Confiança para a Razão de Variâncias (σ_1^2/σ_2^2)

Exemplo 10: Suponha que para uma determinada turma do curso de Estatística estejamos interessados em estudar o tempo que os alunos de sexo feminino levam para executar uma determinada tarefa comparando com os do sexo masculino. Uma amostra aleatória de $n_H = 11$ (homens) e $n_M = 16$ (mulheres) foi extraída dos dois grupos, tendo resultado em $s_H = 5,1$ mins e $s_M = 4,7$ mins, respectivamente. Obtenha o IC a 95% para σ_1^2/σ_2^2 e σ_1/σ_2. Interprete-os