Matemática Computacional

Teórica 5

Departamento de Matemática
Instituto Superior de Engenharia do Porto

2º Semestre 20-21

Conteúdo

1 Inferência estatística

2 Estimação de parâmetros

3 Intervalos de confiança

4 Intervalo de confiança para a média

5 Intervalo de confiança para a diferença entre duas médias

6 Intervalo de confiança para a proporção

7 Intervalo de confiança para a diferença de proporções

Inferência estatística

A inferência estatística é um método científico que permite tirar conclusões sobre os parâmetros da população a partir da recolha, tratamento e análise dos dados de uma amostra, retirada dessa população.

Na inferência estatística ou se admite que a distribuição da população tem uma forma matemática conhecida, embora com alguns parâmetros desconhecidos - estatística paramétrica - ou se pretende conhecer a forma da distribuição - estatística não paramétrica.

Inferência estatística

Tipos de inferência estatística

- Estimação de parâmetros
  - Estimação pontual
  - Estimação intervalar
- Testes de hipóteses

Objetivo da estimação

Tirar conclusões sobre os parâmetros da distribuição de uma população, a partir de estatísticas realizadas sobre amostras dessa mesma população.

Objetivo dos testes de hipóteses

Decidir se o valor do parâmetro estimado pertence ou não a um domínio de valores especificado pelo investigador

Estimação de parâmetros

Estimação pontual

Determinação do verdadeiro valor desconhecido de um parâmetro θ da população estimando-o com base numa amostra de valores.

Estimação intervalar

Determinação de um intervalo que contém o valor exato do parâmetro que se pretende estimar, com um certo grau de confiança (probabilidade).
As estimativas intervalares designam-se intervalos de confiança (IC).

Estimação pontual

Seja X uma população com distribuição supostamente conhecida, dependente de um parâmetro θ desconhecido e $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de X.

Estimador de θ - $\hat{\Theta}$ - é uma função da amostra (ou estatística) $(X_1, X_2, \ldots, X_n)$ que permite estimar o valor do parâmetro θ.

Estimativa de θ - $\hat{\theta}$ - é o valor numérico do estimador $\hat{\Theta}$ para a amostra concreta $(x_1, \ldots, x_n)$.

Estimação pontual

Exemplo

Para uma dada população pretende-se estimar $\theta = \mu = E(X)$ a partir de uma amostra concreta $(x_1, \ldots, x_n)$.

Estimador de θ :

$$\hat{\Theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Estimativa de θ :

$$\hat{\theta} = \hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Estimação intervalar

A estimação pontual fornece uma estimativa $\hat{\theta}$ de um parâmetro θ mas desconhece-se a sua precisão. Na estimação intervalar, considera-se uma medida de erro, Δ, para indicar que o verdadeiro valor de θ pertence, muito provavelmente, a $\left]\hat{\theta}-\Delta, \hat{\theta}+\Delta\right[$.

Intervalo de confiança

Um intervalo de confiança para o parâmetro θ, com grau de confiança $1-\alpha$ é um intervalo

$$IC = \left]\hat{\theta}-\Delta, \hat{\theta}+\Delta\right[$$

tal que, $P\left(\hat{\theta}-\Delta < \theta < \hat{\theta}+\Delta\right) = 1-\alpha$.
α é o **nível de significância** e o **erro** Δ é igual à semi-amplitude de IC.

IC para a média de uma população normal (σ^2 conhecida)

Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de uma população X, com distribuição normal com média μ (desconhecida) e variância σ^2 conhecida.

Um estimador da média populacional μ é dado pela estatística $\overline{X}$.

Já vimos que

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

e então

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

IC para a média de uma população normal (σ^2 conhecida)

Fixado o nível de significância α e designando por z_c o valor da v.a. Z tal que, $P\left(Z > z_c\right) = \alpha/2$, tem-se

$P\left(-z_c < Z < z_c\right) = 1 - \alpha \Leftrightarrow$

$P\left(-z_c < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_c\right) = 1 - \alpha$

$$\therefore\ P\left(\overline{X} - z_c\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_c\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

sendo $z_c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

IC para a média de uma população normal

σ^2 conhecida

Para uma amostra concreta $(x_1, x_2, \ldots, x_n)$, seja $\overline{x}$ o valor da estatística $\overline{X}$.
O intervalo de confiança a $(1-\alpha) \times 100\%$ para μ numa população normal com σ conhecido é

$$IC = \left]\overline{x} - z_c \frac{\sigma}{\sqrt{n}}, \overline{x} + z_c \frac{\sigma}{\sqrt{n}}\right[\tag{1}$$

sendo $z_c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

IC para a média de uma população normal (σ^2 desconhecida)

Sendo a variância da população desconhecida, se a **amostra é grande**, pode substituir-se σ por s (desvio-padrão amostral) em (1) obtendo-se um intervalo de confiança para μ.

No entanto, se **amostra é pequena**, a distribuição da v.a. $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$, que resulta de se substituir σ por um seu estimador, tem distribuição t-Student com $(n-1)$ graus de liberdade.

IC para a média de uma população normal (σ^2 desconhecida)

Sendo t_c o valor da v.a. T tal que $P\left(T > t_c\right) = \alpha/2$, tem-se

$P\left(-t_c < T < t_c\right) = 1 - \alpha \Leftrightarrow$

$P\left(-t_c < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_c\right) = 1 - \alpha$

$$\therefore \; P\left(\overline{X} - t_c\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_c\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

IC para a média de uma população normal

σ^2 desconhecida

Para uma amostra concreta $(x_1, x_2, \ldots, x_n)$, seja $\overline{x}$ o valor da estatística $\overline{X}$.

O intervalo de confiança a $(1-\alpha) \times 100\%$ para μ no caso de uma população normal com σ desconhecido é

$$IC = \left]\overline{x} - t_{c(n-1)}\frac{s}{\sqrt{n}}, \overline{x} + t_{c(n-1)}\frac{s}{\sqrt{n}}\right[\tag{2}$$

IC para a média de uma população não normal

Se n grande, tem-se pelo Teorema do Limite Central que, qualquer que seja a forma da distribuição da qual é retirada a amostra aleatória,

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

IC para a média de uma população não normal

σ^2 conhecida

O intervalo de confiança a $(1-\alpha)\times 100\%$ para μ numa população qualquer, com σ conhecido, desde que dimensão da amostra seja elevada, é

$$IC = \left]\overline{x} - z_c\frac{\sigma}{\sqrt{n}}, \overline{x} + z_c\frac{\sigma}{\sqrt{n}}\right[\tag{3}$$

IC para a média de uma população não normal

σ^2 desconhecida

Se σ é desconhecido, mas a dimensão da amostra é elevada, substitui-se σ por s e o intervalo de confiança a $(1-\alpha)\times 100\%$ para μ, numa população qualquer, é

$$IC = \left]\overline{x} - z_c\frac{s}{\sqrt{n}}, \overline{x} + z_c\frac{s}{\sqrt{n}}\right[\tag{4}$$

Exemplo

De acordo com a especificação, o diâmetro das peças produzidas por uma máquina é uma v.a. Normal com desvio padrão de 3.1 mm. O técnico de calibração retira uma amostra aleatória de 25 peças, calcula o diâmetro médio obtendo 21.4 mm Construa um intervalo de confiança a 95% para o diâmetro médio das peças da máquina.

Resolução:

X_i- v.a. que representa o diâmetro da peça i produzida por uma máquina (mm)

$X_i \sim N\left(\mu, 3.1^2\right)$

Amostra: $n = 25$ e $\overline{x} = 21.4$

$\overline{X}$ - v.a. que representa o diâmetro médio das peças produzidas pela máquina quando considerada uma amostra aleatória de 25 peças.

$$\overline{X} = \frac{1}{25}\sum_{i=1}^{25} X_i \sim N\left(\mu, \frac{3.1^2}{25}\right) \Leftrightarrow Z = \frac{\overline{X}-\mu}{3.1/5} \sim N(0,1)$$

Grau de confiança:

$$1-\alpha = 0.95 \Leftrightarrow \alpha = 0.05 \Rightarrow z_c = z_{0.975} = \Phi^{-1}(1-0.025) = \Phi^{-1}(.975) = 1.96$$

Intervalo de confiança para μ:

$$IC_{95\%}(\mu) = \left]21.4 - 1.96 \times \frac{3.1}{5}, 21.4 + 1.96 \times \frac{3.1}{5}\right[= \left]20.185, 22.615\right[$$

IC para a diferença entre duas médias de populações normais (σ_1^2 e σ_2^2 conhecidas)

Consideremos duas populações normais, $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ e $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$.

Um estimador para $\mu_1 - \mu_2$ é $\overline{X}_1 - \overline{X}_2$.

Selecionando duas amostras aleatórias independentes, uma de cada população, de dimensões n_1 e n_2, respetivamente, tem-se

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

IC para a diferença entre duas médias de populações normais

σ_1^2 e σ_2^2 conhecidas

O intervalo de confiança a $(1-\alpha)\times 100\%$ para $\mu_1-\mu_2$ no caso de populações normais, com variâncias conhecidas, das quais foram extraídas amostras aleatórias independentes é

$$\left]\, (\overline{x}_1-\overline{x}_2) - z_c\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, (\overline{x}_1-\overline{x}_2) + z_c\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \,\right[\tag{5}$$

IC para a diferença entre duas médias de populações normais (σ_1^2 e σ_2^2 desconhecidas)

σ_1^2 e σ_2^2 desconhecidas - grandes amostras

O intervalo de confiança a $(1-\alpha)\times 100\%$ para $\mu_1-\mu_2$ no caso de populações normais, com variâncias desconhecidas, das quais foram extraídas amostras aleatórias independentes de dimensão elevada ($n_1 \geq 30$ e $n_2 \geq 30$) e $\sigma_1^2 \approx s_1^2$, $\sigma_2^2 \approx s_2^2$ é

$$\left]\left(\overline{x}_1-\overline{x}_2\right)-z_c\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}},\left(\overline{x}_1-\overline{x}_2\right)+z_c\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right[\tag{6}$$

IC para a diferença entre duas médias de populações normais ($\sigma_1^2 = \sigma_2^2$ desconhecidas) - pequenas amostras

Consideremos duas populações normais, com variâncias desconhecidas mas iguais, $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ e $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$.

Um estimador para $\mu_1 - \mu_2$ é $\overline{X}_1 - \overline{X}_2$.

Selecionando duas amostras aleatórias independentes, uma de cada população, de dimensões $n_1 < 30$ e $n_2 < 30$, respetivamente, tem-se

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n2}}} \sim t_{n_1+n_2-2}$$

com

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

IC para a diferença entre duas médias de populações normais ($\sigma_1^2 = \sigma_2^2$ desconhecidas)

σ_1^2 e σ_2^2 desconhecidas (iguais) - pequenas amostras

O intervalo de confiança a $(1-\alpha) \times 100\%$ para $\mu_1 - \mu_2$ no caso de populações normais, com variâncias desconhecidas mas supostas iguais, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, das quais foram extraídas amostras aleatórias independentes de pequena dimensão é

$$\left]\left(\overline{x}_1 - \overline{x}_2\right) - t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \left(\overline{x}_1 - \overline{x}_2\right) + t_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right[\tag{7}$$

IC para a diferença entre duas médias de populações normais ($\sigma_1^2 = \sigma_2^2$ desconhecidas)

σ_1^2 e σ_2^2 desconhecidas (iguais) - grandes amostras

O intervalo de confiança a $(1-\alpha)\times 100\%$ para $\mu_1-\mu_2$ no caso de populações normais, com variâncias desconhecidas mas supostas iguais, $\sigma_1^2=\sigma_2^2=\sigma^2$, das quais foram extraídas amostras aleatórias independentes de grande dimensão é

$$\left](\overline{x}_1-\overline{x}_2)-Z_c s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},(\overline{x}_1-\overline{x}_2)+Z_c s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right[\tag{8}$$

IC para a diferença entre duas médias de populações normais (σ_1^2 e σ_2^2 desconhecidas)

σ_1^2 e σ_2^2 desconhecidas (diferentes) - pequenas amostras

O intervalo de confiança a $(1-\alpha) \times 100\%$ para $\mu_1 - \mu_2$ no caso de populações normais, com variâncias diferentes desconhecidas, das quais foram extraídas amostras aleatórias independentes de pequena dimensão, é

$$\left]\,(\overline{x}_1 - \overline{x}_2) - t_c\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\overline{x}_1 - \overline{x}_2) + t_c\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\,\right[\tag{9}$$

IC para a diferença entre duas médias de populações não normais (σ_1^2 e σ_2^2 conhecidas)

σ_1^2 e σ_2^2 conhecidas - grandes amostras

Quando as populações não são normais de variâncias conhecidas e as dimensões das amostras elevadas ($n_1 \geq 30$ e $n_2 \geq 30$), o intervalo de confiança a $(1-\alpha) \times 100\%$ para $\mu_1 - \mu_2$ é

$$\left] (\overline{x}_1 - \overline{x}_2) - z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\overline{x}_1 - \overline{x}_2) + z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right[\quad (10)$$

IC para a diferença entre duas médias de populações não normais (σ_1^2 e σ_2^2 desconhecidas)

σ_1^2 e σ_2^2 desconhecidas (iguais) - grandes amostras

O intervalo de confiança a $(1-\alpha)\times 100\%$ para $\mu_1 - \mu_2$ no caso de populações normais, com variâncias desconhecidas mas supostas iguais, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, das quais foram extraídas amostras aleatórias independentes de grande dimensão é

$$\left]\left(\overline{x}_1 - \overline{x}_2\right) - Z_c s_p^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \left(\overline{x}_1 - \overline{x}_2\right) + Z_c s_p^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right[\tag{11}$$

* vêr slide 22

IC para a diferença entre duas médias de populações não normais (σ_1^2 e σ_2^2 desconhecidas)

σ_1^2 e σ_2^2 desconhecidas (diferentes) - grandes amostras

Se σ_1^2 e σ_2^2 são desconhecidas e $n_1 \geq 30$ e $n_2 \geq 30$, podemos substituir σ_1^2 por s_1^2 e σ_2^2 por s_2^2 e o intervalo de confiança a $(1-\alpha) \times 100\%$ para $\mu_1 - \mu_2$ é

$$\left] (\overline{x}_1 - \overline{x}_2) - z_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\overline{x}_1 - \overline{x}_2) + z_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right[\tag{12}$$

Exemplo

Pretende-se testar a eficácia de um medicamento de ajuda ao combate da obesidade. Criaram-se dois grupos I e II de 40 indivíduos interessados em perder peso. Ao grupo I foi fornecido um placebo e ao grupo II o medicamento em teste. Após o período de controlo obtiveram-se os dados conforme a tabela seguinte

	média($\overline{x}$)	desvio padrão(s)
Grupo I	$2.3kg$	$1.9kg$
Grupo II	$3.5kg$	$3.1kg$

Construa um intervalo de confiança a 95% para a diferença da perda média de peso entre os grupos I e II.

Resolução:

X_{I_i} - v.a. que representa a perda de peso do elemento i do grupo I

X_{II_i} - v.a. que representa a perda de peso do elemento i do grupo II

Amostras: $n_1 = n_2 = 40$, $\overline{x}_1 = 2.3$, $s_1 = 1.9$, $\overline{x}_2 = 3.5$ e $s_2 = 3.1$

$$\overline{X}_I - \overline{X}_{II} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Como as amostras são de dimensão elevada, não é necessário exigir a normalidade das populações, assim como o conhecimento das suas variâncias.

IC para $\mu_1 - \mu_2$ a 95%:

$$\left]\left(\overline{x}_1 - \overline{x}_2\right) - z_{0.975}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \left(\overline{x}_1 - \overline{x}_2\right) + z_{0.975}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right[=$$

$$\left]2.3 - 3.5 - 1.96\sqrt{\frac{1.9^2}{40} + \frac{3.1^2}{40}}, 2.3 - 3.5 + 1.96\sqrt{\frac{1.9^2}{40} + \frac{3.1^2}{40}}\right[=$$

$$]-2.327, -0.073[$$

Como IC apresenta apenas valores negativos para $\mu_1 - \mu_2$, podemos afirmar, com uma confiança de 95%, que a perda média do grupo II é superior à do grupo I, isto é, o medicamento está a revelar-se eficaz.

IC para a proporção populacional

Seja X uma variável aleatória com uma distribuição Binomial de parâmetros (n, p) com $n \geq 30$. Um estimador de p é

$$\hat{P} = \frac{X}{n}$$

sendo X o número de sucessos em n provas de Bernoulli.
Segundo o teorema de De Moivre, sabemos que

$$X \sim N\left(np, npq\right)$$

IC para a proporção populacional

Portanto

$$\hat{P} \sim N\left(p, \frac{pq}{n}\right) \Leftrightarrow Z = \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

Fixado o nível de significância α e designando por z_c o valor da v.a. Z tal que, $P\left(Z > z_c\right) = \alpha/2$, tem-se

$$P\left(-z_c < Z < z_c\right) = 1 - \alpha \Leftrightarrow P\left(-z_c < \frac{\hat{P}-p}{\sqrt{\frac{pq}{n}}} < z_c\right) = 1 - \alpha$$

$$\therefore \; P\left(\hat{P} - z_c\sqrt{\frac{pq}{n}} < p < \hat{P} + z_c\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

sendo $z_c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

IC para a proporção populacional

IC para a proporção

Para uma amostra concreta $(x_1, x_2, \ldots, x_n)$, seja $\hat{p} = \frac{x}{n}$ o valor da estatística $\hat{P}$.
O intervalo de confiança a $(1-\alpha) \times 100\%$ para p, para amostras de dimensão elevada é

$$IC = \left] \hat{p} - z_c\sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + z_c\sqrt{\frac{\hat{p}\hat{q}}{n}} \right[\qquad (13)$$

sendo $z_c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

Exemplo

Numa certa cidade A foi recolhida uma amostra aleatória de 150 mulheres, tendo 57 afirmado que viam telenovela todos os dias. Poder-se-à afirmar, com 90% de confiança, que a proporção de mulheres nesta cidade, que vêem telenovela todos os dias é 42%?

Exemplo: resolução

X - v.a. que representa o nº de mulheres da cidade A que vêem telenovela todos os dias.

Amostra: $n = 150$ e $\hat{p} = \frac{57}{150} = 0.38$

$\hat{P}$ - proporção de mulheres da cidade A que vêem telenovela todos os dias quando considerada uma amostra aleatória de 150.

$\hat{P} = \frac{X}{150} \sim N\left(p, \frac{pq}{150}\right)$, sendo p a proporção de mulheres da cidade A que que vêem telenovela todos os dias

IC a 90% para p: $IC_{90\%} = \left]\hat{p} - z_{0.95}\sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + z_{0.95}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right[=$

$\left]0.38 - 1.645\sqrt{\frac{0.38\times0.62}{150}}, 0.38 - 1.645\sqrt{\frac{0.38\times0.62}{150}}\right[=]0.315, 0.445[$

Podemos afirmar, com 90% de confiança, que a percentagem de mulheres que vê telenovela todos os dias situa-se entre 31.5% e 44.5%. Assim, não é de rejeitar a hipótese de que a proporção de mulheres que vêem telenovela todos os dias é de 42%.

IC para a diferença entre duas proporções

Consideremos as variáveis aleatórias X_1 e X_2 com distribuição Binomial e duas amostras aleatórias independentes de dimensões $n_1 \geq 30$ e $n_2 \geq 30$.

$X_1 \sim B_i(n_1, p_1)$ e $X_2 \sim B_i(n_2, p_2)$

Um estimador para $p_1 - p_2$ é $\hat{P}_1 - \hat{P}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$.

Sabemos que $\hat{P}_1 \sim N\left(p_1, \frac{p_1 q_1}{n_1}\right)$ e $\hat{P}_2 \sim N\left(p_2, \frac{p_2 q_2}{n_2}\right)$

logo,

$$\hat{P}_1 - \hat{P}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

IC para a diferença de proporções

IC para a diferença de proporções

O intervalo de confiança a $(1-\alpha)\times 100\%$ para a diferença $p_1 - p_2$ de duas populações Binomiais das quais se retiram duas amostras de dimensão elevada é

$$\left](\hat{p}_1-\hat{p}_2) - z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}, (\hat{p}_1-\hat{p}_2) + z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}\right[\tag{14}$$

Exemplo

Numa certa cidade A foi recolhida uma amostra aleatória de 150 mulheres, tendo 57 afirmado que viam telenovela todos os dias. Numa outra cidade do mesmo país, cidade B, foi recolhida uma amostra aleatória de 200 mulheres, das quais 80 afirmaram que vêem telenovela todos os dias.

Com 90% de confiança, poder-se-á considerar que a proporção de mulheres que vêem telenovela todos os dias é igual nas duas cidades?

Exemplo: resolução

X_A - v.a. que representa o número de mulheres da cidade A que afirmam ver telenovela todos os dias.

X_B - v.a. que representa o número de mulheres da cidade B que afirmam ver telenovela todos os dias.

Amostras: $n_A = 150$, $\hat{p}_A = \frac{57}{150} = 0.38$, $n_B = 200$, $\hat{p}_B = \frac{80}{200} = 0.40$

$\hat{P}_A$ - proporção de mulheres da cidade A que vêem telenovela todos os dias quando considerada uma amostra aleatória de 150.

$\hat{P}_B$ - proporção de mulheres da cidade B que vêem telenovela todos os dias quando considerada uma amostra aleatória de 200.

$\hat{P}_A - \hat{P}_B \sim N\left(p_A - p_B, \frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}\right)$, sendo p_A e p_B a proporção de mulheres de cada umas das cidades que vêem telenovela todos os dias

Exemplo: resolução

$$IC_{90\%}\left(p_A - p_B\right) =$$

$$\left]\left(\hat{p}_A - \hat{p}_B\right) - z_{0.95}\sqrt{\frac{\hat{p}_A\hat{q}_A}{n_A} + \frac{\hat{p}_B\hat{q}_B}{n_B}}, \left(\hat{p}_A - \hat{p}_B\right) + z_{0.95}\sqrt{\frac{\hat{p}_A\hat{q}_A}{n_A} + \frac{\hat{p}_B\hat{q}_B}{n_B}}\right[=$$

$$\left]0.38 - 0.40 - 1.645\sqrt{\tfrac{0.38\times0.62}{150} + \tfrac{0.4\times0.6}{200}}, 0.38 - 0.40 + 1.645\sqrt{\tfrac{0.38\times0.62}{150} + \tfrac{0.4\times0.6}{200}}\right[$$

$$IC_{90\%}\left(p_A - p_B\right) = \left]-0.107, 0.067\right[$$

Como o IC contém valores positivos e negativos, podemos afirmar, com 90% de confiança, que não há diferença significativa entre as proporções de mulheres das duas cidades que vêem telenovela todos os dias.