Instituto Superior de Engenharia do Porto

# ESTAT

2019/2020

LEEC

# C4 - Estimação de parâmetros

Intervalos de confiança (IC) para médias e diferença de médias

IC para proporções e diferença de proporções grandes amostras.

# Estimação de parâmetros

## Objectivo:

Realização de Inferência sobre o valor dos parâmetros da distribuição de uma população, a partir de estatísticas realizadas sobre amostras dessa mesma população.

# Estimador e estimativa

Amostra aleatória: Conjunto de variáveis aleatórias $\mathrm{i.i.d.} \{X_1, X_2, \ldots.., X_n\}$ cuja distribuição é igual á da população.

Estimador pontual do parâmetro $\theta$: é uma estatística $\widehat{\Theta} = G(X_1, X_2, \ldots.., X_n)$ usada para estimar o valor do parâmetro $\theta$.

Estimativa pontual do parâmetro $\theta$ : é o valor observado da estatística $\hat{\theta} = G(x_1, x_2, \ldots.., x_n)$ correspondente a uma amostra em particular.

Nota: Qualquer "bom" estimador deverá originar uma estimativa "próxima" do valor que esta ser estimado. Existem "critérios"* que permitem avaliar a "qualidade" dos estimadores. Geralmente um "bom" estimador e aquele que e **centrado (não enviesado)** e tem uma boa **precisão.**

(*) Estes critérios assim como os métodos de desenvolvimento de estimadores tais como o método da máxima verosimilhança e/ou o método dos momentos estão fora do âmbito da UC.

# Estimação de parâmetros

Estimação pontual

- Determinação de um valor numérico correspondente ao parâmetro $\theta$ da população, que se pretende estimar com base numa amostra de valores.

Estimação intervalar

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

Os IC para o parâmetro $\theta$ com grau de confiança (1-α) que vamos estudar reduzem-se a intervalos na forma $[\hat{\theta}-\Delta, \hat{\theta}+\Delta]$ de modo que

$$P(\hat{\theta}-\Delta \le \theta \le \hat{\theta}+\Delta) = (1-\alpha)$$

onde, $1-\alpha$ *é o grau de confiança e* $\Delta$ *é a medida do erro*.

# 1. Intervalos de confiança para a média μ

Considere-se uma população com determinada característica em estudo que é uma v.a. X Normal com média μ (desconhecida) e variância $\sigma^2$ conhecida.

i) Sabe-se que a distribuição da médias de uma amostra de dimensão n é $\overline{X} \sim N\left(\mu\,; \frac{\sigma^2}{n}\right)$, de modo que $Z = \frac{\overline{X} - \mu}{\sigma \,/ \sqrt{n}} \sim \; N(0,1)$

ii) Pode afirmar-se, então que

$$\mu = \bar{X} - Z\,\frac{\sigma}{\sqrt{n}} \qquad (1)$$

iii) Pretende-se determinar um intervalo de confiança [A, B] para $\mu$ com um grau de confiança $(1-\alpha)\cdot 100\%$, ou seja

$$P(A \leq \mu \leq B) = 1 - \alpha \qquad (2)$$

# 1. Intervalos de confiança para a média μ

De (1) em (2) , *vem*

$$P\left(A \leq \bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq B\right) = 1 - \alpha. \qquad (3)$$

De (3) obtemos $$P\left(\frac{A-\bar{X}}{\frac{\sigma}{\sqrt{n}}} \leq -Z \leq \frac{B-\bar{X}}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

ou seja, $$P\left(\frac{\bar{X}-B}{\frac{\sigma}{\sqrt{n}}} \leq Z \leq \frac{\bar{X}-A}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha \qquad (4)$$

Procurando o intervalo centrado temos que $P(-z_c \leq Z \leq z_c) = 1 - \alpha \quad (5)$

De (4) e (5) resulta $$\begin{cases} A = \bar{X} - z_c\dfrac{\sigma}{\sqrt{n}} \\ B = \bar{X} + z_c\dfrac{\sigma}{\sqrt{n}} \end{cases}$$

Assim, o Intervalo de Confiança (IC) aleatório [A,B] com um grau de confiança de (1-$\alpha$)100% para a média populacional $\mu$ é dado por:

$$IC_{\mu} = [A, B] = \left[\bar{X} - z_c\frac{\sigma}{\sqrt{n}}, \bar{X} + z_c\frac{\sigma}{\sqrt{n}}\right]$$

# 1. Intervalos de confiança para a média μ

## 1.1 Cálculo de $z_c$

Uma vez escolhido o grau de confiança (1-$\alpha$) pode-se, de imediato, calcular $z_c$

$$Como\ P(-z_c \leq Z \leq z_c) = 1 - \alpha$$
$$\phi(z_c) - \phi(-z_c) = 1 - \alpha \Leftrightarrow \phi(z_c) - (1 - \phi(-z_c)) = 1 - \alpha$$
resulta que $\phi(z_c) = 1 - \alpha/2$

Assim, $z_c = z_{1-\alpha/2} = \phi^{-1}\left(1 - \alpha/_2\right)$

| $(1-\alpha)\cdot 100\%$ | 99% | 98% | 95% | 90% |
|---|---|---|---|---|
| $\alpha \cdot 100\%$ | 1% | 2% | 5% | 10% |
| $z_{1-\alpha/2}$ | 2,575 | 2,33 | 1,96 | 1,645 |

Para um grau de confiança de 95% O IC aleatório é

$$IC_\mu = \left[\bar{X} - 1.96\,\sigma/_{\sqrt{n}}, \bar{X} + 1.96\,\sigma/_{\sqrt{n}}\right]$$

# 1. Intervalos de confiança para a média μ: Erro, ou precisão, da estimativa

O intervalo de confiança para a média

$$IC_\mu = \left[ \bar{X} - z_{1-\alpha/2} \, \sigma\!\Big/\!\sqrt{n} \, , \bar{X} + z_{1-\alpha/2} \, \sigma\!\Big/\!\sqrt{n} \right]$$

pode ser colocado na forma

$$IC_\mu = \left[ \bar{X} - \Delta, \bar{X} + \Delta \right]$$

Onde $\Delta = z_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}}$ é a medida do erro, ou precisão da estimativa.

Observações:

1) A amplitude do IC é igual a $2\Delta$.

Considerando fixo o grau de confiança

2) Se a dimensão da amostra n aumenta (diminui) o erro diminui (aumenta) e consequentemente a precisão aumenta (diminui).

# Intervalos de confiança para a média μ: IC aleatório e IC determinístico

- Antes da amostragem, o IC tem uma probabilidade (1-$\alpha$) de conter o valor da média μ (desconhecida) e tem como limites de confiança as variáveis aleatórias

$$\bar{X} \pm z_{1-\alpha/2}\sigma/\sqrt{n}$$

- Depois de realizar a amostragem substitui-se, no IC, $\bar{X}$ por $\bar{x}_0$ (nº real – estimativa pontual) obtém-se, desta forma, um intervalo determinístico

$$IC_\mu = \left[\bar{x}_o - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x}_o + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

# Dimensionamento de amostras

A precisão de um IC para a média é dado pela sua semi-amplitude $\Delta$. Assim, é possível estimar-se com um grau de confiança $(1-\alpha)$x100% a dimensão da amostra (*n*), a recolher para garantir uma dada precisão $\varepsilon$, resolvendo a inequação

$$\Delta = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \varepsilon$$

Em ordem a n, obtendo

$$n \geq \left(\frac{z_{1-\alpha/2} \times \sigma}{\varepsilon}\right)^2, n \in \mathbb{N}$$

No caso da variância populacional ser desconhecida pode recolher-se uma amostra preliminar (amostra piloto) para obter a estimativa inicial, s, do desvio padrão $\sigma$, obtendo-se pela equação anterior uma estimativa para *n*.

# Estimação intervalar da média. Desvio padrão populacional desconhecido.

Caso a população tenha distribuição normal e se n≥30 (grande amostra) pode-se utilizar, dentro de certas condições gerais, a aproximação $\sigma^2 \approx s^2$.

O intervalo determinístico é

$$IC_\mu = [\bar{x}_0 - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{x}_0 - z_{1-\alpha/2}\frac{s}{\sqrt{n}}]$$

Caso a população não seja normal se n≥30 (grande amostra) usa-se também a aproximação $\sigma^2 \approx s^2$. Contudo, o IC é aproximado pois foi usado o teorema do limite central.

# Exemplo 1

De acordo com a especificação, o diâmetro das peças produzidas por uma máquina é uma v.a. Normal com média 20 mm desvio padrão igual a 3,1 mm.

O técnico de calibração retira uma amostra aleatória de 25 peças, calcula o diâmetro médio obtendo 21,4 mm

a) Construa um intervalo de confiança a 95% para o diâmetro médio das peças da máquina.

b) Baseando a sua resposta no IC obtido anteriormente poder-se-á decidir que a máquina está desalinhada.

# Exemplo 1. Resolução

a) $Amostra: \begin{cases} \bar{x}_o = 21{,}4 \\ n = 25 \end{cases}$

$\bar{X}$ – "Diâmetro médio de 25 peças (mm)"

$$\bar{X} = \frac{\sum_{i=1}^{x} X_i}{n} \sim N\left(\mu, \frac{3.1^2}{25}\right)$$ aditividade Normal

$X_i$ – Diâmetro da i–esima peça (mm)

i=1,2,…,25; $X_i \sim N(\mu, 3.1^2)$

$$Z = \frac{\bar{X} - \mu}{3.1/\sqrt{25}} \sim N(0;1)$$

$Confiança: (1-\alpha) = 0.95 \Leftrightarrow \alpha = 0.05$

$Erro: \Delta = z_{1-\alpha/2}\, {}^{\sigma}/_{\sqrt{n}}$

$IC_\mu: [\bar{x}_0 - \Delta; \bar{x}_0 + \Delta]$

Cálculo de $z_{1-\alpha/2}$: $z_{0.975} = \Phi^{-1}(0.975) = 1.96$

$$\Delta = 1.96 \cdot \frac{3.1}{\sqrt{25}} \approx 1.2mm$$

$R: Após\ amostragem: IC_\mu = [21.4 - 1.2; 21.4 + 1.2] = [20.2; 22.6]$

b) R: Como o valor de referência 20 mm não está contido no IC então podemos admitir que há diferença significativa entre a verdadeira média e o valor de referência. Existem evidência estatísticas que permitem afirmar que a máquina está desalinhada, com um grau de confiança de 95%.

# Intervalos de confiança para a diferença de médias

Intervalo de Confiança para a diferença das médias ($\sigma_1^2$ e $\sigma_2^2$ conhecidos)

Se $\overline{X}_1$ e $\overline{X}_2$ são as médias amostrais de duas amostras aleatórias mutuamente independentes de tamanhos $n_1$ e $n_2$, de duas populações independentes com médias $\mu_1$ e $\mu_2$ desconhecidas e variâncias $\sigma_1^2$ e $\sigma_2^2$ conhecidas, sabemos, dentro de certas condições, que

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

Pretende-se determinar um intervalo de confiança aleatório [A,B] para $\mu_1 - \mu_2$ com um nível de confiança $(1-\alpha)100\%$.

Notando que $P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha \;\; e$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

# Intervalos de confiança para a diferença de médias

vem,

$$P\left( -z_{1-\alpha/2} \leq \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{1-\alpha/2} \right) = 1 - \alpha$$

Procedendo de forma análoga ao que fizemos para dedução do IC para a média, obtém-se o intervalo de confiança aleatório:

$$IC_{\mu_1-\mu_2} = \left[ (\overline{X}_1 - \overline{X}_2) - z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} ; (\overline{X}_1 - \overline{X}_2) + z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

*Uma vez realizada a amostragem, obtém-se um IC determinístico:*

$$IC_{\mu_1-\mu_2} = \left[ (\bar{x}_1 - \bar{x}_2) - z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} , (\bar{x}_1 - \bar{x}_2) + z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

# Intervalos de confiança para a diferença de médias

**Exemplo 2**

Pretende-se comparar dois tipos de baterias A e B.

Realizaram-se 30 percursos aleatórios com veículos equipados com baterias do tipo A e 30 percursos aleatórios com veículos equipados com baterias do tipo B.(com carga máxima)

Obtiveram-se os resultados:

| Tipo bateria | Distância média (km) | Desvio padrão (km) | nº de percursos |
|---|---|---|---|
| A | 425,3 | 51,3 | 30 |
| B | 441,2 | 55,1 | 30 |

a) Determine um IC a 95% para a verdadeira diferença entre as distâncias médias percorridas por ambos os tipos de bateria.

b) Baseando a sua resposta no resultado obtido anteriormente, pode admitir-se que haja diferença significativa entre a verdadeira distância média percorrida usando baterias do tipo A e do tipo B?

# Estimação intervalar para proporções

Considere-se uma população caracterizada por uma distribuição binomial relativamente a determinada característica.

Pretende-se determinar um IC para a proporção p da distribuição de indivíduos que têm determinada característica em toda a população:

Estimar: $IC_p$=[A, B]: $P(A \leq p \leq B) = 1 - \alpha$

O número de indivíduos que tem a característica em estudo (sucesso) numa amostra de tamanho n é caracterizado pela v.a

$$X \sim B_i(n, p) \;\; aprox \; N(np, npq) \;\; \text{se n é “ grande”}$$

onde q=1-p.

A percentagem de indivíduos que tem a característica em estudo (sucesso) numa amostra de tamanho n (n>>30) é a v.a

$$\hat{P} = \frac{X}{n} aprox \quad N(p, \frac{pq}{n})$$

# Estimação intervalar para proporções

Considerando grandes amostras n>>30, pretende-se construir o Intervalo de confiança aleatório IC=[A,B] contendo uma probabilidade 1-α de conter o valor do parâmetro desconhecido p.

$$P(A \leq p \leq B) = 1 - \alpha$$

A distribuição da proporção amostral é dada por:

$$\hat{P} = \frac{X}{n}\ aprox.N\left(p, \frac{pq}{n}\right)$$

Por conseguinte

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

Substituindo Z na expressão seguinte

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

Obtém-se

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

# Estimação intervalar para proporções

rearranjando as desigualdades, vem

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

Neste caso é impossível manipular as desigualdades para remover *p*, que desconhecemos, então para *n* grande substitui-se *p* pelo estimador $\hat{P}$ e q por 1-$\hat{P}$, resultando

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

obtendo-se assim o IC aleatório para *p* correspondente ao grau de confiança de (1-*α*)100%:

$$IC_p = \left[\hat{P} - z_{1-\alpha/2}\sqrt{\hat{P}(1-\hat{P})/n}; \hat{P} + z_{1-\alpha/2}\sqrt{\hat{P}(1-\hat{P})/n}\right]$$

onde, ,

$$z_{1-\alpha/2} = \phi^{-1}(1-\alpha/2)$$

# Intervalo de Confiança determinístico para *p*

Depois de realizar a amostragem substitui-se $\hat{p}_o$ por $\hat{P}$ (nº real - estimativa pontual baseada na amostra) e obtém-se o intervalo de confiança determinístico:

$$IC_p = \left[\hat{p}_0 - \Delta; \hat{p}_0 + \Delta\right]$$

onde

$$\Delta = z_{1\text{-}\alpha/2}\sqrt{\frac{\hat{p}_0\hat{q}_0}{n}}$$

$$\hat{q}_0 = 1 - \hat{p}_0$$

# Exemplo 3

Uma cadeia de supermercados tenciona abrir mais uma loja num bairro. Antes de tomar a decisão entrevistou 400 pessoas e 310 indicaram que frequentariam o novo supermercado. Encontre um IC a 95% para a proporção de pessoas que poderá ser cliente da loja.

Resolução:

Dados amostrais:

$$\begin{cases} \hat{\mathrm{p}}_0 = \dfrac{310}{400} = 0{,}775 \\ n = 400 \end{cases}$$

$\hat{\mathrm{P}}$ – "Proporção de pessoas que afirmam frequentar o supermercado, em 400"

$$\hat{P} \sim N\left(p; \frac{pq}{400}\right), \text{ aproximando } \frac{\mathrm{pq}}{400} \approx \frac{\hat{\mathrm{p}}_0 \hat{q}_0}{400} = \frac{0{,}775 \cdot 0{,}225}{400}$$

$= 0{,}0004359.$

$\hat{P} \sim N(p; 0{,}0004359)$

Cálculo do erro $\Delta$

$1 - \alpha = 0{,}95 \Leftrightarrow \alpha = 0{,}05$

$z_{1-\alpha/2} = z_{0,975} = \Phi^{-1}(0{,}975) = 1{,}96$

$$\Delta = z_{1-\alpha/2} \sqrt{\frac{\hat{\mathrm{p}}_0 \hat{q}_0}{n}} = z_{0,975} \sqrt{\frac{0{,}775 \cdot 0{,}225}{400}} = 1{,}96 \cdot \sqrt{0{,}0004359} = 0{,}041$$

$R$: *Após amostragem* : $IC_p = [\hat{\mathrm{p}}_0 - \Delta; \hat{\mathrm{p}}_0 + \Delta] = [0{,}734; 0{,}816]$

# Exemplo 4

Pretende-se conhecer a % de componentes eletrónicos defeituosos num contentor com cerca de 1000000 de unidades. Para o efeito recolheu-se uma amostra de 300, com reposição, tendo-se verificado que 18 eram defeituosos.

a) Determine um IC para a % de componentes defeituosos em todo o contentor, para um grau de confiança de 98%.

b) Estime qual o tamanho da amostra a recolher de modo a ser possível determinar um IC com um grau de confiança de 98% com um erro máximo de 1%.

$\hat{\mathrm{P}}$ – "Proporção de unidades defeituosas, numa amostra de 300"

$$\text{Dados amostrais:} \begin{cases} \hat{\mathrm{p}}_0 = \dfrac{18}{300} = 0{,}06 \\ n = 300 \end{cases}$$

$$\hat{P} \sim N\left(p; \frac{pq}{300}\right), \text{ aproximando } \frac{\mathrm{pq}}{300} \approx \frac{\hat{\mathrm{p}}_0 \hat{q}_0}{300} = \frac{0{,}06 \cdot 0{,}94}{300}$$

$= 0{,}0137^2$ temos que $\hat{P} \sim N\left(p; 0{,}0137^2\right)$

Cálculo do erro $\Delta$

$1 - \alpha = 0{,}98 \Leftrightarrow \alpha = 0{,}02$

$z_{1-\alpha/2} = z_{0{,}99} = \varphi^{-1}(0{,}99) = 2{,}33,$ logo $\Delta = 2{,}33 \cdot 0{,}0137 \approx 0{,}032$

$a) R: Após\ amostragem: IC_p = [\hat{\mathrm{p}}_0 - \Delta; \hat{\mathrm{p}}_0 + \Delta] = [0{,}028; 0{,}092]$

$$b) \Delta \leq 0{,}01 \Leftrightarrow 2{,}33 \sqrt{\frac{0{,}06 \cdot 0{,}94}{n}} \leq 0{,}01 \Leftrightarrow n \geq 3062$$

( Usa–se a % de defeituosas obtida anteriormente como estimativa da % futura – amostra piloto)

# Intervalos de confiança para a diferença de proporções

- O problema é: $\text{Determinar}[A, B]: P(A \leq p_1 - p_2 \leq B) = 1 - \alpha$

- Neste caso usamos a diferença $\hat{P}_1 - \hat{P}_2 = \dfrac{X_1}{n_1} - \dfrac{X_2}{n_2}$

  para estimar $p_1 - p_2$

- sendo $n_1$ ≥30 e $n_2$ ≥30 os respetivos tamanhos das amostras aleatórias i.i.d. e mutuamente independentes, pelo T.L.C, tem-se

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

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}} \sim \quad N(0,1)$$

# Estimação intervalar para diferença proporções

Partindo de

$$P\left(-z_{1-\alpha/2} \le \frac{(\hat{P}_1-\hat{P}_2)-(p_1-p_2)}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}} \le z_{1-\alpha/2}\right)=1-\alpha$$

obtemos, pelo procedimento habitual, o IC com (1-$\alpha$)100% de confiança para $p_1 - p_2$

$$IC_{p_1-p_2}=\left[(\hat{P}_1-\hat{P}_2)-z_{1-\alpha/2}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1}+\frac{\hat{P}_2(1-\hat{P}_2)}{n_2}},(\hat{P}_1-\hat{P}_2)+z_{1-\alpha/2}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1}+\frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}\right]$$

e , após amostragem, obtém-se o IC determinístico

$$IC_{p_1-p_2}=\left[(\hat{p}_1-\hat{p}_2)-z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\ \hat{q}_1}{n_1}+\frac{\hat{p}_2\ \hat{q}_2}{n_2}},(\hat{p}_1-\hat{p}_2)+z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\ \hat{q}_1}{n_1}+\frac{\hat{p}_2\ \hat{q}_2}{n_2}}\right]$$

$\hat{q}_1=1-\hat{p}_1; \hat{q}_2=1-\hat{p}_2$

# Exemplo de aplicação

Realizou-se um estudo sobre a % de pessoas que concorda com uma determinada medida governamental.

Região N: Número de pessoas inquiridas:50
Concordam:27 Sem opinião:4 Discordam:19
Região S: Número de pessoas inquiridas:120
Concordam:48 Sem opinião:16 Discordam:56

a) Poder-se-á admitir que a percentagem de opiniões concordantes é idêntica nas duas regiões ? Considere um grau de confiança de 95%.

b) Idem a) para uma confiança de 90%.

c) Pode afirmar-se, com uma confiança de 95%, que há diferença significativa entre a % de pesssoas sem opinião formada, em ambas as regiões?

# Resumo das fórmulas

IC para grandes amostras (n≥30)

| Tipo IC | Estatística teste | Erro | IC determinístico |
|---|---|---|---|
| Média | $\overline{X} \sim N(\mu; \frac{\sigma^2}{n})$ | $\Delta = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ | $[\overline{x}_0 - \Delta; \overline{x}_0 + \Delta]$ |
| Dif de médias | $\overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$ | $\Delta = z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ | $[(\overline{x}_1 - \overline{x}_2) - \Delta; (\overline{x}_1 - \overline{x}_2) + \Delta]$ |
| proporções | $\hat{P} \sim N(p; \frac{pq}{n})$ | $\Delta = z_{1-\alpha/2} \sqrt{\frac{\hat{p}_0 \hat{q}_0}{n}}$ | $[\hat{p}_0 - \Delta; \hat{p}_0 + \Delta]$ |
| Dif. proporções | $\hat{P}_1 - \hat{P}_2 \sim N(p_1 - p_2; \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2})$ | $\Delta = z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$ | $[(\hat{p}_1 - \hat{p}_2) - \Delta; (\hat{p}_1 - \hat{p}_2) + \Delta]$ |

$q = 1 - p$

$\hat{q}_0 = 1 - \hat{p}_0$