## 9.1. Conceito de Intervalo de Confiança

Através dos métodos de ***estimação pontual***, faz-se associar a cada parâmetro de uma população, ***um número isolado***, que constitui uma estimativa do parâmetro.

A grande limitação dos métodos de estimação pontual é que ***não fornecem qualquer informação relativa ao rigor ou à confiança das estimativas que através deles são obtidas***.

A ***limitação*** é ***ultrapassada recorrendo aos métodos de estimação por intervalo***. O conceito envolvido em tais métodos é o ***Intervalo de Confiança***.

DGI
2019

## Exemplo

Admita-se que a altura dos alunos do IPB segue uma distribuição normal com parâmetros $\mu = 1.69$ m e $\sigma^2 = (0.06\text{ m})^2$.

Para amostras de dimensão $n$, a média amostral $\bar{X}$ , segue uma distribuição normal:

$$\bar{X} \rightarrow \mathrm{N}\left(\mu,\ \sigma^2/n\right)$$

e

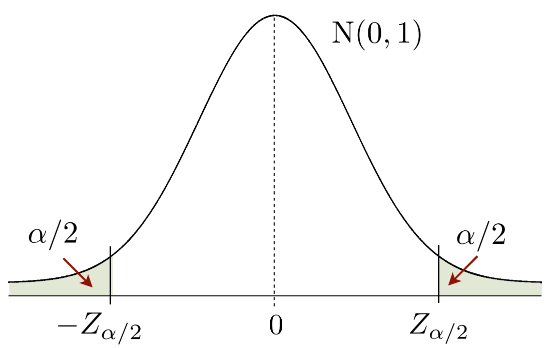
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \rightarrow \quad \mathrm{N}(0,\ 1)$$

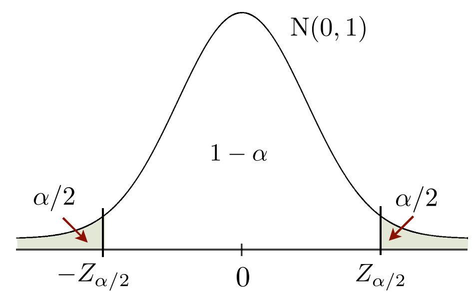
Definindo $Z_{\alpha/2}$ como o valor para o qual:

$$P\left(Z > Z_{\alpha/2}\right) = \alpha/2$$

e

$$P\left(Z < -Z_{\alpha/2}\right) = \alpha/2$$






vem: $P\left(-Z_{\alpha/2} < Z < Z_{\alpha/2}\right) = 1 - \alpha$

$$P\left(-Z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\mu - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Reescrevendo a desigualdade:

$$P\left(\bar{X} - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

De acordo com a expressão, *o intervalo*: $\bar{X} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

*Incluirá, com a probabilidade* $(1-\alpha)$ *o valor de* $\mu$ *(isto é, o parâmetro).* ***Tal intervalo designa-se por intervalo de confiança de $\mu$ a*** $(1-\alpha) \times 100\%$ ***.Os seus extremos designam-se por limites de confiança.***

Se para uma amostra de dimensão $n = 25$, se obtivesse $\bar{X} = 1.69$, definir o intervalo de confiança a 95%.

*Limite Inferior:* $\bar{X} - Z_{0.025} \times \dfrac{\sigma}{\sqrt{n}} = 1.69 - 1.96 \times \dfrac{0.06}{5} = 1.666$

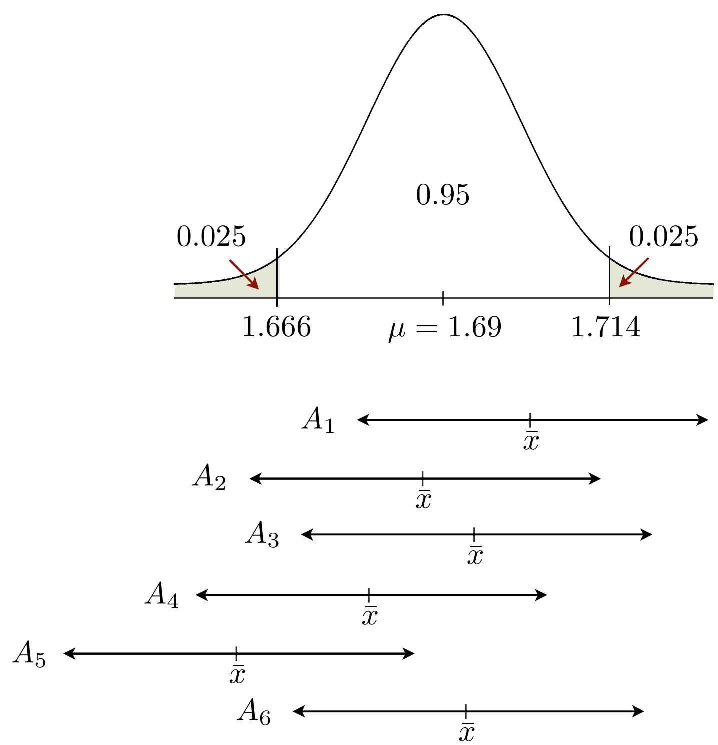
*Limite Superior:* $\bar{X} + Z_{0.025} \times \dfrac{\sigma}{\sqrt{n}} = 1.69 + 1.96 \times \dfrac{0.06}{5} = 1.714$

*Intervalo de Confiança:* $[1.666, 1.714]$ ou $1.666 < \mu < 1.714$

**Nota**

Neste caso admitiu-se que $\sigma^2$ (variância da população) era conhecida. Na prática é normalmente conhecida uma estimativa.

***O valor de μ é fixo*** (trata-se de um parâmetro da população a estimar). ***O que varia de amostra para amostra é o intervalo especificado*** *(por depender de* $\bar{X}$ *).*




- ***Numa amostra em cada vinte, o intervalo não conterá o valor do parâmetro estimado.***

## Simetria do Intervalo de Confiança relativamente ao valor da Estimativa Pontual

O intervalo de confiança de $\mu$ a $(1-\alpha)\cdot 100\%$:

$$\bar{X} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

é um intervalo simétrico em relação a $\bar{X}$

- Para quaisquer valores $\alpha_1$, $\alpha_2$ que satisfaçam a condição $\alpha = \alpha_1 + \alpha_2$ , os intervalos de confiança: $\left[\bar{X} - Z_{\alpha 1} \times \frac{\sigma}{\sqrt{n}},\ \bar{X} + Z_{\alpha 2} \times \frac{\sigma}{\sqrt{n}}\right]$

  são intervalos de confiança a $(1-\alpha)\cdot 100\%$.

- Quando a estatística a partir da qual se definem os intervalos tem uma distribuição unimodal simétrica(como sucede com $\bar{X}$ , quando a população é normal), então o *intervalo é simétrico* em relação à estatística (isto é, o que se obtém quando $\alpha_1 = \alpha_2 = \alpha/2$ ) *é o de menor amplitude e portanto o mais interessante*.

Considerando novamente a amostra de dimensão $n = 25$ com $\overline{X} = 1.69$, defina os intervalos de confiança a 95% para as três situações seguintes:

**(1)** $\alpha_1 = \alpha_2 = 0.025$

$[1.666,\ 1.714]$ $\qquad$ $\text{Amplitude} = 1.714 - 1.666 = 0.048$

**(2)** $\alpha_1 = 0.01,\ \alpha_2 = 0.04$

$[1.662,\ 1.711]$ $\qquad$ $\text{Amplitude} = 1.711 - 1.662 = 0.049\ (> 0.048)$

**(3)** $\alpha_1 = 0.005,\ \alpha_2 = 0.045$

$[1.659,\ 1.710]$ $\qquad$ $\text{Amplitude} = 1.710 - 1.659 = 0.051\ (> 0.048)$

## *Especificação de um Intervalo de Confiança*

***A especificação de um intervalo de confiança implica o conhecimento simultâneo de:***

**(1)** Um estimador do parâmetro em causa;

**(2)** A sua distribuição amostral;

**(3)** Um valor particular do estimador (*isto é, uma estimativa pontual*).

## 9.2. Intervalo de Confiança para o Valor Esperado

### *9.2.1. Amostra de Grande Dimensão, População Qualquer*

Neste caso, de acordo com o Teorema do Limite Central, a média amostral segue aproximadamente uma distribuição normal.

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \rightarrow \mathrm{N}(0,\ 1) \qquad I.C. \text{ a } (1-\alpha)\times 100\% \quad \rightarrow \quad \bar{x} \pm z(\alpha/2) \times \frac{\sigma}{\sqrt{n}}$$

Como a dimensão da amostra é grande, o erro de estimação é desprezável, vindo $\sigma \approx s$

com: $$s = \sqrt{\frac{1}{n-1}\sum_{n=1}^{N}(X_n - \bar{X})^2}$$

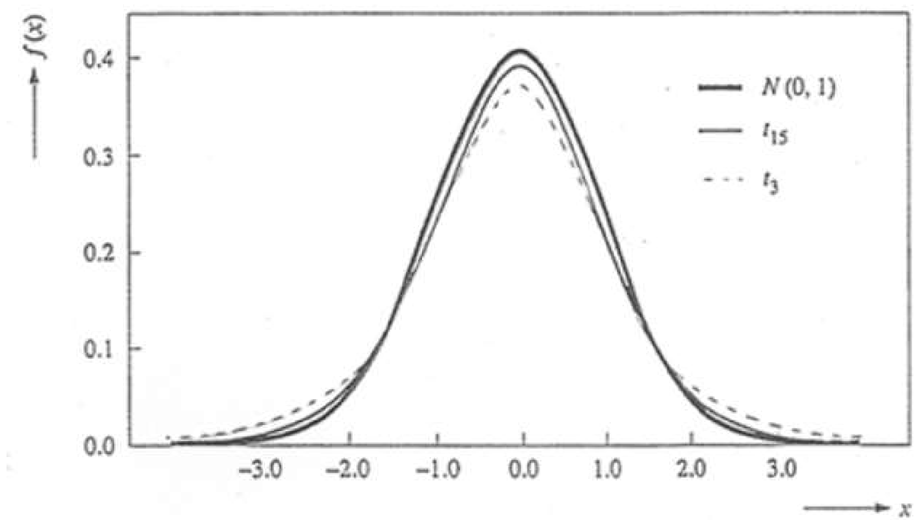
### 9.2.2. Amostra de Pequena Dimensão, População Normal

Como a dimensão da *amostra é pequena*, já *não é válida a aproximação* $\sigma \approx s$

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \to t_{n-1} \qquad I.C.\text{a}(1-\alpha)\times 100\% \quad \to \quad \bar{x} \pm t_{n-1}(\alpha/2) \times \frac{s}{\sqrt{n}}$$

#### 9.2.2.1. Distribuição Normal versus $t$-Student

A utilização da *distribuição* $t$-***Student* leva à definição de Intervalos de *Confiança de maior amplitude*** que os obtidos (inapropriadamente?) através da distribuição Normal.




- ***A distribuição $t$-Student apresenta maior variância em relação à distribuição Normal***.

  Isso significa que, se os valores de $z$ estão a ser usados quando se deveria usar $t$, o *Intervalo de Confiança* determinado apresentará uma *amplitude inferior* à que deveria ter.

Intervalo de confiança para o valor esperado a 90% . Dimensão da amostra , $n = 5$

$$t_4(\alpha/2 = 0.05): \quad \bar{x} \pm 2.1318 \times s \Big/ \sqrt{5}$$

$$Z(\alpha/2 = 0.05): \quad \bar{x} \pm 1.6449 \times s \Big/ \sqrt{5}$$

$$t_4(\alpha/2 = ?) = 1.6449 \quad \rightarrow \quad \alpha/2 = 0.088$$
$$\rightarrow \quad \alpha = 0.176$$
$$\rightarrow \quad (1-\alpha) \times 100\% = 82.4\%$$

Na realidade, usando inadvertidamente a distribuição Normal, definimos um intervalo de confiança a 82.4% em vez de 90%.

| | *Nível de Significância desejada do I.C. para populações Normais* | | |
|---|---|---|---|
| $\boldsymbol{n}$ | 90% | 95% | 99% |
| 5 | 82.4 | 87.8 | 93.8 |
| 10 | 86.5 | 91.8 | 97.0 |
| 15 | 87.8 | 92.9 | 97.8 |
| 20 | 88.3 | 93.5 | 98.1 |
| 25 | 88.7 | 93.8 | 98.3 |
| 30 | 88.9 | 94.0 | 98.4 |
| 50 | 89.3 | 94.4 | 98.6 |

***Os procedimentos propostos anteriormente cobrem todas as situações de obtenção de intervalos de confiança para o valor esperado, excepto quando a amostra é pequena e a população não é normal***. Neste caso, seria necessário recorrer à mediana amostral (fora do objectivo da disciplina).

Que procedimento deve ser seguido?

| | ***População*** | |
|---|---|---|
| | *Normal* | *Não Normal* |
| $n \leq 30$ | $t$ de Student<br>(Normal, se σ conhecido) | Mediana |
| $n > 30$ | $t$ de Student<br>(Normal, se σ conhecido)<br>2ª escolha: Normal | Mediana<br>2ª escolha: Normal |

## 9.3. Intervalo de Confiança para a Proporção Binomial

### 9.3.1. *Amostra de Grande Dimensão*

Considere-se uma ***população dicotómica, constituída por elementos de dois tipos***.

O valor $p$, que corresponde à ***proporção de elementos de um dos dois tipos designa-se por Proporção Binomial***.

$Y$ – nº de ocorrências de elementos de um dado tipo contidos na amostra.

$$Y \to \mathrm{B}(n,\ p)$$

- De acordo com o ***Teorema do Limite Central, para valores de $n$ suficientemente elevados***:

$$Y \approx \mathrm{N}(n \times p,\ n \times p \times q)$$

como: $$p \approx \hat{p} = \frac{y}{n} \to \mathrm{N}\left(p, \frac{p \times q}{n}\right)$$

dessa forma:

$$\frac{\hat{p}-p}{\sqrt{\frac{p\times q}{n}}} \to \mathrm{N}(0,1) \qquad I.C.\mathrm{a}(1-\alpha)\times 100\% \quad \to \quad \hat{p} \pm z(\alpha/2)\sqrt{\frac{\hat{p}\times\hat{q}}{n}}$$

### 9.3.2. *Amostra de Pequena Dimensão (solução exacta para $n \leq 30$)*

Seja $Y$ a variável aleatória representando *o* ***número de sucessos*** numa amostra aleatória de tamanho $n$.

- O intervalo de confiança é determinado da seguinte forma:

*Limite Inferior:* $$P(Y \geq \hat{y}) \to \sum_{y=\hat{y}}^{n} \mathrm{C}_y^n \times p_i^y \times (1-p_i)^{n-y} = \alpha/2$$

*Limite Superior:* $$P(Y \leq \hat{y}) \to \sum_{y=0}^{\hat{y}} \mathrm{C}_y^n \times p_s^y \times (1-p_s)^{n-y} = \alpha/2$$

*Nota:* Em alternativa pode utilizar-se a *tabela A6* (R. L. Iman & W. J. Conover)

## Exemplo

$$n = 20$$
$$\hat{y} = 12$$

$$\text{Estimativa Pontual}: \quad \hat{p} = \hat{y}/n = 0.6 \quad \rightarrow \quad \text{Intervalo de Confiança: } [0.361, 0.809]$$

- Note-se que o ***intervalo de confiança não é simétrico*** em relação à estimativa pontual. Note-se também que o ***intervalo de confiança é bastante amplo,*** pelo que é pouco informativo. Como a ***amplitude do intervalo de confiança diminui com o aumento da dimensão da amostra, este exemplo ilustra a necessidade de amostras de dimensão elevada quando se pretende estimar a proporção binomial***.

DGI
2019

## 9.4. Intervalo de Confiança para a Variância de uma População Normal

Se, de uma população normal forem retiradas amostras de dimensão $n$, com variância amostral $s^2$:

$$\frac{(n-1)\times s^2}{\sigma^2} \to \chi^2_{n-1}$$

$$I.C.\text{a}(1-\alpha)\times 100\% \quad \to \quad \frac{(n-1)\times s^2}{\chi^2_{n-1}(\alpha/2)} \leq \sigma^2 \leq \frac{(n-1)\times s^2}{\chi^2_{n-1}(1-\alpha/2)}$$

DGI
2019

| Parâmetro | Condições | Distribuição | Intervalo de Confiança | Notas |
|---|---|---|---|---|
| Valor Esperado $\mu$ | Amostra de grande dimensão. População qualquer. | $\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$ | $\overline{x} \pm z(\alpha/2)\cdot\frac{\sigma}{\sqrt{n}}$ | Estimador do desvio padrão: $\sigma \approx s$ (1) |
| | Amostra de pequena dimensão. População Normal. | $\frac{\overline{x}-\mu}{s/\sqrt{n}} \to t_{n-1}$ | $\overline{x} \pm t_{n-1}(\alpha/2)\cdot\frac{s}{\sqrt{n}}$ | Não é válida a aproximação: $\sigma \approx s$ |
| Proporção Binomial $p$ | Amostra de grande dimensão. População de Bernoulli. (2) | $\frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} \to N(0,1)$ | $\hat{p} \pm z(\alpha/2)\cdot\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$ (3) | Estimador da proporção binomial: $p \approx \hat{p} = \frac{y}{n}$ |
| Variância $\sigma^2$ | População Normal. | $\frac{(n-1)\cdot s^2}{\sigma^2} \to \chi^2_{n-1}$ | $\frac{(n-1)\cdot s^2}{\chi^2_{n-1}(\alpha/2)} \le \sigma^2 \le \frac{(n-1)\cdot s^2}{\chi^2_{n-1}(1-\alpha/2)}$ | |

(1) O desvio padrão $\sigma$, sendo desconhecido, é estimado através do erro padrão $s = \sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\overline{x})^2}$

(2) No caso de amostras de pequena dimensão é necessário recorrer à solução exacta através da distribuição binomial.

(3) Se a população é finita e a amostragem é efectuada sem reposição, a aproximação da distribuição de $Y$ pela normal continua válida se $n \ge 20,\, n\cdot p \ge 7$ e desde que a dimensão da população $M$ seja grande e a dimensão da amostra, $n$, muito menor que $M$. Em rigor, nesta situação, a estimativa do erro padrão de $Y/n$ deverá ter em conta o factor de redução $(M-n)/(M-1)$

(4) A igualdade das variâncias deverá ser testada recorrendo ao teste à razão das variâncias (Teste $\mathrm{F}$).

| Parâmetro | Condições | | Distribuição | Intervalo de Confiança | Notas |
|---|---|---|---|---|---|
| Diferença de valores esperados<br>$\mu_A - \mu_B$<br>(4) | Amostras independentes de grande dimensão.<br>População qualquer. | Variâncias diferentes | $\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}} \to N(0,1)$ | $(\overline{x}_A - \overline{x}_B) \pm z(\alpha/2) \cdot \sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}$ | Estimadores dos desvios padrão:<br>$\sigma_A^2 \approx s_A^2,\ \sigma_B^2 \approx s_B^2$ |
| | | Variâncias iguais | $\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{\sigma \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} \to N(0,1)$ | $(\overline{x}_A - \overline{x}_B) \pm z(\alpha/2) \cdot \sigma \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}$ | $\sigma_A^2 = \sigma_B^2 = \sigma^2 \approx s^2$<br>$s^2 = \dfrac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2}$ |
| | Amostras independentes de pequena dimensão.<br>Populações Normais. | Variâncias diferentes | $\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} \to t_{GL}$ | $(\overline{x}_A - \overline{x}_B) \pm t_{GL}(\alpha/2) \cdot \sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}$ | $GL = \dfrac{\left(s_A^2/n_A + s_B^2/n_B\right)^2}{\dfrac{\left(s_A^2/n_A\right)^2}{n_A - 1} + \dfrac{\left(s_B^2/n_B\right)^2}{n_B - 1}}$ |
| | | Variâncias iguais | $\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{s \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} \to t_{GL}$ | $(\overline{x}_A - \overline{x}_B) \pm t_{GL}(\alpha/2) \cdot s \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}$ | $GL = n_A + n_B - 2$<br>$s^2 = \dfrac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2}$ |
| Diferença de proporções binomiais<br>$p_A - p_B$ | Amostras independentes de grande dimensão.<br>Populações de Bernoulli. | | $\dfrac{(\hat{p}_A - \hat{p}_B) - (p_A - p_B)}{\sqrt{\dfrac{p_A \cdot q_A}{n_A} + \dfrac{p_B \cdot q_B}{n_B}}} \to N(0,1)$ | $(\hat{p}_A - \hat{p}_B) \pm z(\alpha/2) \cdot \sqrt{\dfrac{\hat{p}_A \cdot \hat{q}_A}{n_A} + \dfrac{\hat{p}_B \cdot \hat{q}_B}{n_B}}$ | Estimadores das proporções binomiais:<br>$\hat{p}_A = \dfrac{y_A}{n_A},\ \hat{p}_B = \dfrac{y_B}{n_B}$ |
| Razão de variâncias<br>$\sigma_A^2/\sigma_B^2$ | Populações Normais. | | $\dfrac{s_A^2/\sigma_A^2}{s_B^2/\sigma_B^2} \to F_{GL_1,GL_2}$ | $\dfrac{s_A^2/s_B^2}{F_{GL_1,GL_2}(\alpha/2)} \le \dfrac{\sigma_A^2}{\sigma_B^2} \le \dfrac{s_A^2/s_B^2}{F_{GL_1,GL_2}(1-\alpha/2)}$ | $GL_1 = n_A - 1,\ GL_2 = n_B - 1$<br>$F_{GL_1,GL_2}(1-\alpha/2) = \dfrac{1}{F_{GL_2,GL_1}(\alpha/2)}$ |

DGI
2019

## 9.5. Amostras Dependentes

*Intervalo de Confiança a (1-α)·100%* para a diferença dos valores esperados, $\mu_1 - \mu_2$, quando as *amostras não podem ser consideradas independentes*: ***observações emparelhadas***.

- Define-se uma nova variável aleatória:

$$D = X_1 - X_2$$

Os valores da variável $D$, são considerados como constituindo uma amostra e o Intervalo de Confiança determina-se como na situação em que existe apenas uma amostra (Intervalo de Confiança para o valor esperado).

## Exemplo

Numa experiência industrial, uma certa tarefa foi realizada por 10 operários, primeiro de acordo com o método A e em seguida de acordo com o método B. Os dados que se seguem fornecem os resultados obtidos (tempos em minutos)

| ***operário*** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ***método A*** | 5.4 | 5.6 | 5.8 | 6.2 | 4.9 | 4.9 | 5.9 | 6.1 | 5.3 | 5.9 |
| ***método B*** | 5.2 | 5.7 | 5.3 | 5.9 | 4.9 | 4.8 | 5.7 | 6.0 | 5.3 | 5.7 |
| ***Diferença*** | -0.2 | 0.1 | -0.5 | -0.3 | 0 | -0.1 | -0.2 | -0.1 | 0 | -0.2 |

## 9.6. Dimensionamento de Amostras

*Qual o interesse em dimensionar convenientemente uma amostra?*

- Recolher e tratar uma ***amostra grande demais*** para os resultados que se pretendem obter constitui um ***desperdício de recursos***.
- É destituída de interesse uma amostra cuja ***dimensão não seja suficiente*** para se poderem tirar *conclusões* ***relevantes***.

O ***dimensionamento correcto*** de uma amostra com base na qual se pretende estimar um ou mais parâmetros de uma população passa pela ***identificação de qual a fiabilidade desejada para a(s) estimativa(s)***.

INSTITUTO POLITÉCNICO DE BRAGANÇA Escola Superior de Tecnologia e Gestão



DGI
2019

### 9.6.1. Dimensionamento de Amostras para o Valor Esperado

População Normal com $\sigma^2$ conhecida.

$$\bar{x} \pm z(\alpha/2) \times \frac{\sigma}{\sqrt{n}}$$

***Objectivo:*** Qual deverá ser a dimensão da amostra a recolher para que a amplitude do intervalo de confiança para o valor esperado não exceda $\omega$?

$$\omega = 2 \times z(\alpha/2) \times \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad n = 4 \cdot \sigma^2 \times \left[\frac{z(\alpha/2)}{\omega}\right]^2$$

### 9.6.1. Dimensionamento de Amostras para o a Proporção Binomial

Considerando que a dimensão da amostra a retirar é grande ($n$ *grande*):

$$\hat{p} \pm z(\alpha/2) \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}}$$

$$\omega = 2 \times z(\alpha/2) \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}} \quad \rightarrow \quad n = 4 \times \hat{p} \times \hat{q} \times \left[\frac{z(\alpha/2)}{\omega}\right]^2$$

A expressão anterior depende de $\hat{p}$ e $\hat{q} = 1 - \hat{p}$ que são geralmente desconhecidos antes da obtenção da amostra. Nestes casos, utiliza-se o valor mais conservativo: $\hat{p} = 1/2$

$$n = \left[\frac{z(\alpha/2)}{\omega}\right]^2$$

***Nota:*** $\phi = \hat{p} \times \hat{q} = \hat{p} \times (1 - \hat{p})$

$$\frac{d\phi}{d\hat{p}} = (1 - \hat{p}) - \hat{p} = 1 - 2 \times \hat{p} \qquad \frac{d\phi}{d\hat{p}} = 0 \rightarrow \hat{p} = 1/2 \qquad \frac{d^2\phi}{d\hat{p}^2} = -2 < 0 \Rightarrow \text{máximo}$$