

Intervalos de Confiança

Departamento de Matemática
Escola Superior de Tecnologia de Viseu

Introdução

Introdução

Estimar o consumo médio de um automóvel, estimar o tempo médio que um funcionário leva a aprender uma nova tarefa ou estimar a percentagem (proporção) de pessoas que irão consumir um produto que vai ser lançado no mercado, são exemplos de estimação.

A estimação pode ser feita por dois processos:

- ▶ Estimação Pontual
- ▶ Estimação Intervalar

Estimação Pontual

Na **estimação pontual**, estima-se o parâmetro θ desconhecido (ou $\tau(\theta)$) usando o valor de um estimador $\hat{\theta}$, o qual é designado por **estimador pontual**.

Desvantagem

Não permite avaliar a precisão do estimador.

Exemplo

Parâmetro populacional	Exemplo de estimador pontual
Média (μ)	\bar{X}
Variância (σ^2)	S^2

Estimação Intervalar

Estimação Intervalar

A **estimação intervalar** consiste na determinação de um intervalo onde, com uma certa confiança (probabilidade), esteja o parâmetro θ desconhecido, tendo-se em conta um seu estimador.

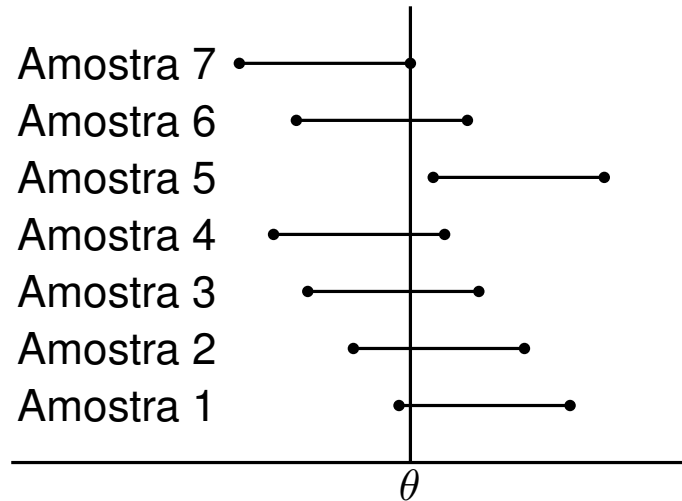
Assim, $P(L_1 < \theta < L_2) = \lambda$ significa que a probabilidade do intervalo aleatório (L_1, L_2) conter o valor exacto θ é λ .

O intervalo (L_1, L_2) é designado por **intervalo de confiança** para o parâmetro θ , com um **nível de confiança** λ .

Depois de recolhida uma amostra aleatória, usam-se os valores observados dessa amostra, para calcular os valores observados das variáveis aleatórias L_1 e L_2 , que se representam, respectivamente, por l_1 e l_2 .

(l_1, l_2) é o **intervalo de confiança concreto** para aquela amostra.

Estimação Intervalar



Estimação Intervalar

Vantagem

É possível determinar o erro máximo cometido na estimação, com uma certa confiança

Notas

- ▶ Tem em conta as variações das estatísticas amostrais de amostra para amostra.
- ▶ Nunca podemos ter intervalos com 100% de confiança.

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

Estamos perante uma situação em que temos conhecimento da distribuição de X e também da sua variância:

$$X \sim N(\mu, \sigma^2)$$

As variáveis aleatórias X_1, X_2, \dots, X_n que constituem a amostra, são independentes e têm distribuição $N(\mu, \sigma^2)$, donde, pelo [Teorema da aditividade da distribuição normal](#),

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

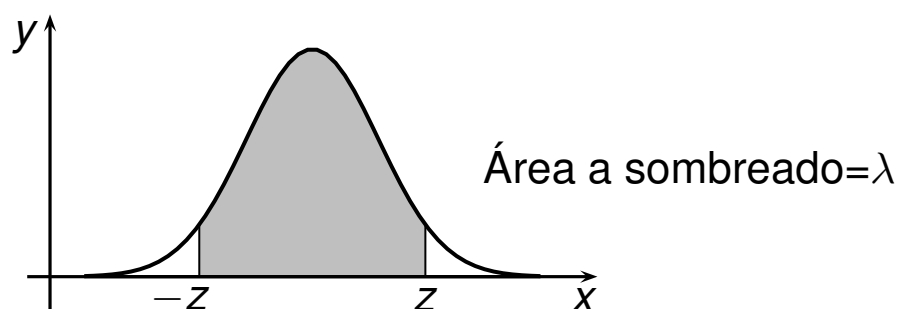
A v. a. Z é designada por [variável fulcral](#).

Int. conf. para a média de uma população normal com variância conhecida

Usando as tabelas da distribuição normal padronizada determinamos o valor de z que satisfaz a seguinte condição:

$$P(-z < Z < z) = \lambda$$

onde λ é o nível de confiança desejado.



Int. conf. para a média de uma pop. normal com var. conhecida

$$\begin{aligned}
P(-z < Z < z) = \lambda &\Leftrightarrow P\left(-z < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z\right) = \lambda \\
&\Leftrightarrow P\left(-z\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z\frac{\sigma}{\sqrt{n}}\right) = \lambda \\
&\Leftrightarrow P\left(-\bar{X} - z\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + z\frac{\sigma}{\sqrt{n}}\right) = \lambda \\
&\Leftrightarrow P\left(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z\frac{\sigma}{\sqrt{n}}\right) = \lambda
\end{aligned}$$

Logo, o intervalo de confiança a $\lambda \times 100\%$ para μ é dado por:

$$\left(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}}\right)$$

Int. conf. para a média de uma pop. normal com var. conhecida

Sendo

$$\mu \in \left(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}}\right),$$

o erro que cometemos usando \bar{X} para estimar μ (Erro = $|\bar{X} - \mu|$) é, com probabilidade λ , inferior ou igual a $z\frac{\sigma}{\sqrt{n}}$ (metade da amplitude do intervalo).

Sendo assim, é possível escolher o tamanho de amostra, n , de modo a que o erro cometido seja menor ou igual a um valor especificado, e , com uma certa confiança $\lambda \times 100\%$.

Basta resolver a seguinte equação:

$$z\frac{\sigma}{\sqrt{n}} = e$$

Int. conf. para a média de uma pop. normal com var. conhecida

Exemplo (Exercício 7 da Ficha n.º 4: Intervalos de confiança)

Certo equipamento de empacotamento automático, encontra-se regulado para encher embalagens de um quilo de certo produto. O seu deficiente funcionamento origina prejuízo para a empresa: se a maioria das embalagens tem peso inferior ao estabelecido, haverá reclamações por parte dos clientes e perda de prestígio; peso excessivo será por outro lado anti-económico. Aceita-se da experiência passada que o peso das embalagens se comporta normalmente com desvio padrão de 12 gramas. Para verificar a afinação do equipamento, seleccionaram-se em determinada altura, nove embalagens cujos pesos exactos (em gramas) foram anotados:

983 992 1011 976 997 1000 1004 983 998

Exemplo (cont.)

1. Estime μ através de uma estimativa pontual.
2. Construa um intervalo de confiança para μ , com os seguintes graus de confiança: 90%, 95% e 99%. Como varia a precisão do intervalo (a sua amplitude) com o grau de confiança escolhido?
3. Qual deverá ser o tamanho da amostra a recolher, para que o erro que se comete ao considerar o valor da média amostral como estimativa para a média da população, não seja superior a 1. (utilize $\lambda = 0.95$).

Sol.:

1. $\bar{x} = 993.78$
2. $[I.C.0.9]_{\mu} = (987.2, 1000.36)$; $[I.C.0.95]_{\mu} = (985.94, 1001.62)$;
 $[I.C.0.99]_{\mu} = (983.476, 1004.084)$.

Quanto maior é a confiança, maior é a amplitude do intervalo, i.e., menor é a precisão do intervalo.

O que se ganha em “confiança” perde-se em “precisão”.

3. $n \geq 554$

Int. conf. para a média de uma pop. normal com var. desconhecida, usando amostras de pequena dimensão

Neste caso, não conhecemos o valor de σ^2 e como tal não podemos usar a variável fulcral do caso anterior.

Sabemos que, se $X \sim N(\mu, \sigma^2)$ então

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Agora, a **variável fulcral** é T .

Determina-se o valor de t tal que $P(-t < T < t) = \lambda$, recorrendo a uma tabela da distribuição t-Student (ou a um computador).

O **intervalo de confiança** a $\lambda \times 100\%$ para o valor esperado μ é dado por

$$\left(\bar{X} - t \frac{S}{\sqrt{n}}, \bar{X} + t \frac{S}{\sqrt{n}} \right)$$

Int. conf. para μ usando amostras de grande dimensão

Pelo Teorema Limite Central, quando a amostra é suficientemente grande ($n > 30$), a média amostral \bar{X} tem, aproximadamente, distribuição normal de média μ e variância σ^2/n , isto é,

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

A variável fulcral é então

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

e o **intervalo de confiança** a $\lambda \times 100\%$, é dado por

$$\left(\bar{X} - z \frac{\sigma}{\sqrt{n}}, \bar{X} + z \frac{\sigma}{\sqrt{n}} \right)$$

onde z é tal que $P(-z < Z < z) = \lambda$.

Int. conf. para μ usando amostras de grande dimensão

Na prática o valor de σ não é, em geral, conhecido. Uma vez que a amostra é suficientemente grande, a substituição de σ pelo seu estimador S na variável Z , não invalida que esta tenha aproximadamente distribuição normal. Então, a **variável fulcral** passa a ser,

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

e o **intervalo de confiança**

$$\left(\bar{X} - z \frac{S}{\sqrt{n}}, \bar{X} + z \frac{S}{\sqrt{n}} \right)$$

onde z é tal que $P(-z < Z < z) = \lambda$.

Estimação Intervalar da Diferença entre Valores Médios

Estimação intervalar da diferença entre valores médios $\mu_1 - \mu_2$

Consideram-se agora duas variáveis aleatórias, X_1 e X_2 , que representam uma certa característica em duas populações distintas, **População 1** e **População 2**, respectivamente.

Pretende-se construir um intervalo de confiança para a diferença $\mu_1 - \mu_2$, sendo μ_1 o valor médio de X_1 e μ_2 o valor médio de X_2 , ambos desconhecidos.

Mais notação: $\sigma_1, \sigma_2 \hookrightarrow$ desvios padrões de X_1 e X_2 ;
 $n_1, n_2 \hookrightarrow$ dimensão das amostras recolhidas.

Nota:

As amostras recolhidas devem ser independentes uma da outra.

Para estimar $\mu_1 - \mu_2$ pontualmente, usamos o valor do estimador pontual $\bar{X}_1 - \bar{X}_2$.

Int. conf. para a diferença entre valores médios de duas populações normais com variâncias conhecidas

Temos: $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$. Logo, pelo teorema da aditividade da distribuição normal,

$$\bar{X}_1 \sim N\left(\mu_1, \sigma_1^2/n_1\right) \quad \text{e} \quad \bar{X}_2 \sim N\left(\mu_2, \sigma_2^2/n_2\right)$$

Uma vez que as amostras são independentes uma da outra, \bar{X}_1 e \bar{X}_2 são independentes. Assim, mais uma vez pelo teorema da aditividade da distribuição normal,

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

e a **variável fulcral** é

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

Int. conf. para a diferença entre valores médios de duas populações normais com variâncias conhecidas

$$\begin{aligned} P(-z < Z < z) = \lambda &\Leftrightarrow P\left(-z < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z\right) = \lambda \\ &\Leftrightarrow P\left(-z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < (\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) < z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = \lambda \\ &\Leftrightarrow P\left((\bar{X}_1 - \bar{X}_2) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = \lambda \end{aligned}$$

Logo, o **intervalo de confiança a $\lambda \times 100\%$ para $\mu_1 - \mu_2$** é dado por:

$$\left((\bar{X}_1 - \bar{X}_2) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Int. conf. para a diferença entre as médias de 2 populações normais, com var. desconhecidas mas iguais, usando amostras de pequena dimensão

Se $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$ com $\sigma_1 = \sigma_2$, então

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Logo, o intervalo de confiança a $\lambda \times 100\%$ para $\mu_1 - \mu_2$ é dado por:

$$\left((\bar{X}_1 - \bar{X}_2) - t \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \right. \\ \left. (\bar{X}_1 - \bar{X}_2) + t \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

onde t é tal que $P(-t < T < t) = \lambda$.

Int. conf. para $\mu_1 - \mu_2$ usando amostras de grande dimensão

Pelo teorema Limite Central temos: $\bar{X}_1 \sim N(\mu_1, \sigma_1^2/n_1)$ e $\bar{X}_2 \sim N(\mu_2, \sigma_2^2/n_2)$. Uma vez que as amostras são independentes, \bar{X}_1 e \bar{X}_2 são independentes. Assim,

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

e a variável fulcral é

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

Logo, o intervalo de confiança a $\lambda \times 100\%$ para $\mu_1 - \mu_2$ é dado por:

$$\left((\bar{X}_1 - \bar{X}_2) - z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

onde z é tal que $P(-z < Z < z) = \lambda$.

Int. conf. para $\mu_1 - \mu_2$ usando amostras de grande dimensão

Se não forem conhecidos os valores de σ_1 e σ_2 , estes são substituídos pelos seus estimadores S_1 e S_2 . Como as amostras são de grande dimensão, esta substituição não altera a distribuição assintótica de Z , vindo

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1)$$

Logo, o intervalo de confiança a $\lambda \times 100\%$ para $\mu_1 - \mu_2$ é dado por:

$$\left((\bar{X}_1 - \bar{X}_2) - z\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + z\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

onde z é tal que $P(-z < Z < z) = \lambda$.

Estimação Intervalar da Variância de Uma População Normal

Int. Conf. para a variância de uma população normal

Se $X \sim N(\mu, \sigma^2)$ então, para uma amostra de tamanho n ,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Sejam a e b tais que: $P(\chi_{n-1}^2 < a) = \frac{1-\lambda}{2}$ e $P(\chi_{n-1}^2 > b) = \frac{1-\lambda}{2}$

Assim, $P(a < \chi_{n-1}^2 < b) = \lambda$, donde

$$\begin{aligned} P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = \lambda &\Leftrightarrow P\left(\frac{1}{b} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{a}\right) = \lambda \\ &\Leftrightarrow P\left(\frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}\right) = \lambda \end{aligned}$$

i.e.,

$$\sigma^2 \in \left(\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a} \right) \text{ com } \lambda \times 100\% \text{ de confiança.}$$

Int. Conf. para a razão entre as variâncias de duas populações normais

Notação

- ▶ X_1 e X_2 : v.a.s que representam uma certa característica em duas populações distintas, **População 1** e **População 2**
- ▶ σ_1 : Desvio-padrão de X_1 ;
- ▶ σ_2 : Desvio-padrão de X_2 ;
- ▶ n_1 : Tamanho da amostra da **População 1** ;
- ▶ n_2 : Tamanho da amostra da **População 2** ;

Nota

As amostras recolhidas devem ser independentes uma da outra.

Para estimar σ_1^2/σ_2^2 pontualmente, usamos o valor do estimador pontual S_1^2/S_2^2 .

Int. Conf. para a razão entre as variâncias de duas populações normais

Temos: $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$. Então,

$$\frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1}^{n_2-1}$$

Sejam a e b tais que

$$P\left(F_{n_2-1}^{n_1-1} < a\right) = \frac{1-\lambda}{2} \quad \text{e} \quad P\left(F_{n_2-1}^{n_1-1} > b\right) = \frac{1-\lambda}{2}$$

Assim,

$$P\left(a < \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} < b\right) = \lambda \quad \Leftrightarrow \quad P\left(a \frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < b \frac{S_2^2}{S_1^2}\right) = \lambda$$

$$P\left(\frac{S_1^2}{b S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{a S_2^2}\right) = \lambda$$

Int. Conf. para a razão entre as variâncias de duas populações normais

Então,

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{S_1^2}{b S_2^2}, \frac{S_1^2}{a S_2^2} \right)$$

com $\lambda \times 100\%$ de confiança.

O intervalo de confiança a $\lambda \times 100\%$ para $\frac{\sigma_1^2}{\sigma_2^2}$ é então dado por

$$\left(\frac{S_1^2}{b S_2^2}, \frac{S_1^2}{a S_2^2} \right)$$

Estimação Intervalar de uma Proporção

Int. Conf. para uma proporção

Consideremos uma população cujos elementos podem ser classificados em dois tipos: **Sucesso** e **Insucesso**.

Pretende-se estimar a proporção p de sucessos na população.

Dada uma amostra de tamanho n , uma **estimativa pontual de p** é dada por $\hat{p} = x/n$, onde x é o nº de elementos do tipo sucesso contidos na amostra.

Esta estimativa é produzida pelo estimador $\hat{p} = X/n$, onde X é a v. a. que representa o **nº de sucessos contidos numa amostra aleatória de tamanho n** .

Tem-se,

$$X \sim B(n, p)$$

Int. Conf. para uma proporção

Se n for suficientemente grande, a distribuição binomial pode ser bem aproximada pela normal, vindo

$$X \sim N(np, npq) \Rightarrow \hat{p} = \frac{X}{n} \sim N\left(p, \frac{pq}{n}\right)$$

e conseqüentemente

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

Seja z tal que $P(-z < Z < z) = \lambda$. Então,

$$P(-z < \frac{\hat{p} - p}{\sqrt{pq/n}} < z) = \lambda \Leftrightarrow$$

$$\Leftrightarrow P(-z\sqrt{pq/n} < \hat{p} - p < z\sqrt{pq/n}) = \lambda \Leftrightarrow$$

$$\Leftrightarrow P(\hat{p} - z\sqrt{pq/n} < p < \hat{p} + z\sqrt{pq/n}) = \lambda$$

Int. Conf. para uma proporção

Da última igualdade poderíamos deduzir que o I.C., a $\lambda \times 100\%$, para p seria

$$\left(\hat{p} - z\sqrt{pq/n}, \hat{p} + z\sqrt{pq/n}\right)$$

No entanto, os limites deste intervalo contêm o parâmetro p que queremos estimar (e que é desconhecido).

Para contornar esta dificuldade podemos substituir p pelo seu estimador, o que conduz ao seguinte intervalo de confiança

$$\left(\hat{p} - z\sqrt{\hat{p}\hat{q}/n}, \hat{p} + z\sqrt{\hat{p}\hat{q}/n}\right)$$

onde $\hat{q} = 1 - \hat{p}$.

Int. Conf. para uma proporção

Questão

Qual deverá ser o tamanho da amostra, de modo a que, com uma certa confiança, o erro que se comete ao estimar p usando \hat{p} , seja inferior a e ?

Como vimos anteriormente,

$$P(\hat{p} - z\sqrt{pq/n} < p < \hat{p} + z\sqrt{pq/n}) = \lambda$$

isto é, com $\lambda \times 100\%$ de confiança a distância máxima entre p e \hat{p} é:

$$z\sqrt{pq/n} = z\sqrt{p(1-p)/n}$$

Pretende-se calcular o valor de n tal que

$$e = z\sqrt{p(1-p)/n}$$

Resolvendo em ordem a n , vem $n = z^2 p(1-p)/e^2$.

Int. Conf. para uma proporção

Como o valor de p não é conhecido, podemos substituí-lo por uma sua estimativa \hat{p} , conhecida *a priori* (caso exista), obtendo

$$n = z^2 \frac{\hat{p}(1-\hat{p})}{e^2}$$

Quando p não pode ser estimada *a priori*, um procedimento alternativo consiste em substituir $p(1-p)$ pelo máximo valor que pode tomar que é 0.25 , o que conduz a uma dimensão máxima da amostra, uma vez que se tomou para $p(1-p)$ o valor mais desfavorável,

$$n = z^2 \frac{0.25}{e^2}$$

Int. Conf. para a diferença entre proporções

Notação

- ▶ X_1, X_2 : v.a.s que representam o número de sucessos contidos nas amostras retiradas, respectivamente, da **População 1** e da **População 2**;
- ▶ n_1 : Tamanho da amostra da **População 1** ;
- ▶ n_2 : Tamanho da amostra da **População 2** ;

Nota

As amostras recolhidas devem ser independentes uma da outra.

$$\hat{p}_1 = X_1/n_1 \quad \text{e} \quad \hat{p}_2 = X_2/n_2$$

Para estimar $p_1 - p_2$ pontualmente, usamos o valor do estimador pontual $\hat{p}_1 - \hat{p}_2$.

Int. Conf. para a diferença entre proporções

Tem-se: $X_1 \sim B(n_1, p_1)$ e $X_2 \sim B(n_2, p_2)$

Se $n_1 > 30$ e $n_2 > 30$,

$$X_1 \sim N(n_1 p_1, n_1 p_1 q_1) \quad \text{e} \quad X_2 \sim N(n_2 p_2, n_2 p_2 q_2)$$

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1 q_1}{n_1}\right) \quad \text{e} \quad \hat{p}_2 \sim N\left(p_2, \frac{p_2 q_2}{n_2}\right)$$

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

Logo a **variável fulcral** é:

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

Int. Conf. para a diferença entre proporções

Para calcular o intervalo de confiança, a $\lambda \times 100\%$, para $p_1 - p_2$, determinamos z tal que $P(-z < Z < z) = \lambda$, donde,

$$P\left((\hat{p}_1 - \hat{p}_2) - z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right) = \lambda$$

Os limites do intervalo sugerido pela igualdade anterior contêm os parâmetros desconhecidos p_1 e p_2 . Contornamos esta dificuldade substituindo p_1 e p_2 por \hat{p}_1 e \hat{p}_2 , respectivamente.

Assim, obtemos o seguinte intervalo de confiança para $p_1 - p_2$ a $\lambda \times 100\%$:

$$\left((\hat{p}_1 - \hat{p}_2) - z\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}, (\hat{p}_1 - \hat{p}_2) + z\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right)$$

onde $\hat{q}_1 = 1 - \hat{p}_1$ e $\hat{q}_2 = 1 - \hat{p}_2$.