Introdução

Embora a estimação pontual seja importante, sobretudo por ser expedita, a sua aplicação tal e qual foi apresentada no capítulo anterior carece da quantificação de erro. Isto é, quando obtemos uma estimativa sabemos que ela não coincidirá necessariamente com o valor verdadeiro do parâmetro estimado, mas estará muito longe desse valor? Até que ponto poderemos confiar no resultado obtido?

Há vários métodos para quantificar essa qualidade da estimativa. O método mais usado é sem dúvida a estimação intervalar. A sua principal vantagem é o facto de a incerteza associada às estimativas estar perfeitamente quantificada pelo nível de confiança, pelo que sabemos sempre exactamente o grau da confiança que poderemos depositar nos resultados obtidos. Ou seja, a estimação intervalar indica-nos não só o valor da estimativa, mas também o quanto podemos confiar em cada resultado.

1. Conceito de intervalo de confiança e outros conceitos relacionados

A estimação intervalar permite definir uma gama de valores – **intervalo de confiança** (**I.C.**) - na qual se encontrará o verdadeiro valor do parâmetro populacional a estimar (desconhecido), com uma confiança ou probabilidade pré-definida.

Os limites do I.C. designam-se por **limites de confiança** e serão aqui designados por l_{inf} e l_{sup}.

Um intervalo de confiança já construído apresentar-se-á com o seguinte aspeto:

$$[\ l_{inf}\ ;\ l_{sup}\ ]$$ [1]

A fiabilidade que podemos atribuir ao resultado obtido é o **grau** ou **nível de confiança**, n.c., cujo significado é: a probabilidade de o verdadeiro valor (desconhecido) do parâmetro que pretendemos estimar se encontrar dentro do I.C.

O valor mais usado para o nível de confiança é sem dúvida: 95%, 99% é também um valor muito usado quando se exige uma fiabilidade elevada e ainda 90 % quando não há grandes problemas em caso de erro. Em alguns casos pontuais, 98%, 99,9% , 80% e outros valores são também utilizados.

A **amplitude** ou **largura** do intervalo de confiança é uma medida da precisão (não da exactidão) do resultado e define-se como sendo a diferença entre o limite superior e o limite inferior:

$$A = l_{sup} - l_{inf}$$

Também muito importante é o erro máximo da estimativa ou incerteza (também vulgarmente referido como margem de erro), definido como sendo metade da amplitude:

$$\text{Incerteza} = A / 2 = (l_{sup} - l_{inf}) / 2$$

O interesse do erro máximo da estimativa, ou incerteza, é o seguinte: se necessitarmos de utilizar apenas um valor como estimativa do parâmetro em estudo, e não um intervalo de valores, o erro máximo que cometeremos, de acordo com o intervalo de confiança construído, é igual à Incerteza.

[1] Sendo l_{inf} e l_{sup} valores já calculados, claro !!

Quanto maior o nível de confiança, maior a fiabilidade dos resultados ; no entanto, também é maior a amplitude do intervalo de confiança, o que pode tornar os resultados inconclusivos. Por isso, ao seleccionar um valor para o n.c., há que ter sempre em atenção um compromisso entre confiança e precisão.

Quando pretendemos aumentar a precisão dos resultados, isto é, diminuir a amplitude, sempre que possível, devemos aumentar a dimensão da amostra em vez de baixar o nível de confiança.

Exemplo 1: *Com base num conjunto de resultados amostrais, construiu-se o seguinte intervalo de confiança para a média da população em estudo: [35,7 ; 43,1], a 95 % de confiança.*

Qual o significado deste resultado ?
Este resultado significa que, apesar de não conhecermos o verdadeiro valor da média populacional, a probabilidade de ele se encontrar entre 35,7 e 43,1 é de 95 %.
Além disso, a probabilidade desse valor estar abaixo do limite inferior, 35,7 , é de 2,5 % e a probabilidade de ele estar acima do limite superior, 43,1 , é também de 2,5 %.

A amplitude ou largura do intervalo é: $A = l_{sup} - l_{inf} = 43{,}1 - 35{,}7 = 7{,}4$.

A estimativa central é: $(35{,}7 + 43{,}1) / 2 = 39{,}4$

A incerteza a esta estimativa é: Incerteza = $A / 2 = 7{,}4 / 2 = 3{,}7$

O intervalo poderia ser apresentado também na seguinte forma: $39{,}4 \pm 3{,}7$ *, evidenciando-se assim a estimativa central e o erro associado.*

2. Construção de intervalos de confiança

2.1 Metodologia de construção de um intervalo de confiança

A metodologia de construção de um I.C. pode ser resumida nos seguintes passos:

1 – Seleção da fórmula adequada ao parâmetro a estimar e às circunstâncias populacionais e de recolha da amostra

2 – Consulta da tabela adequada de variáveis padrão (em anexo) para obtenção do valor crítico – ajuda elaborar previamente um gráfico e fazer a distribuição das áreas, de acordo com o valor do n.c – ver exemplos subsequentes.

3 – Cálculo dos limites de confiança e construção do I.C.

2.2. Intervalo de confiança para a média populacional

Dados necessários: - média amostral,

- dimensão amostral, n
- preferencialmente, desvio padrão populacional, σ ; no caso de este não ser conhecido (caso mais comum) desvio padrão amostral, s
- pré-definição do nível de confiança, n.c.

Caso população normal e σ seja conhecido, para qualquer n
ou
Caso população qualquer, σ seja conhecido e n > 30 (amostra grande):

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \qquad z \sim N(0,1)$$

Caso população normal, σ não seja conhecido (usa-se o valor de s)
e n ≤ 30 (amostra pequena):

$$\bar{x} \pm t \frac{s}{\sqrt{n}} \qquad t \sim T(GL) \quad \text{com} \quad GL = n - 1$$

Caso população qualquer, σ não seja conhecido (usa-se o valor de s)
e n > 30 (amostra grande):

$$\bar{x} \pm z \frac{s}{\sqrt{n}} \qquad z \sim N(0,1)$$

Quando seja exigido que a população seja normal para que se possa construir o I.C. e não haja certeza, poderemos verificar o pressuposto da normalidade pelo método simples apresentado no capítulo anterior (há outros métodos que não fazem parte do conteúdo deste módulo).

z ~ N (0,1) é um valor que segue a **distribuição normal reduzida** ou **padronizada**. Esta distribuição é semelhante às demais distribuições normais, sendo caracterizada por $\mu = 0$ e $\sigma = 1$. É a única distribuição normal cujos valores é comum encontrar tabelados na bibliografia (ver anexos).

A distribuição normal reduzida é equivalente à distribuição t de Student com um número de graus de liberdade igual a infinito. Em certas situações, como as de construção de intervalos de confiança, esta mostra-se de uma utilização mais simples.

t ~ T (GL) é um valor que segue a **distribuição t de Student**. Os valores desta distribuição encontram-se tabelados na bibliografia (ver anexos), em função de um parâmetro GL (graus de liberdade) que, em cada caso deverá ser correctamente calculado.

Exemplo 2: *Numa amostra de 25 valores, calculou-se: $\bar{x}$ = 187,2 e s = 31,9.*
Construa um intervalo de confiança para o valor médio da população da qual a amostra foi extraída, com uma confiança de 95 %.

1º passo: Seleção da fórmula a utilizar

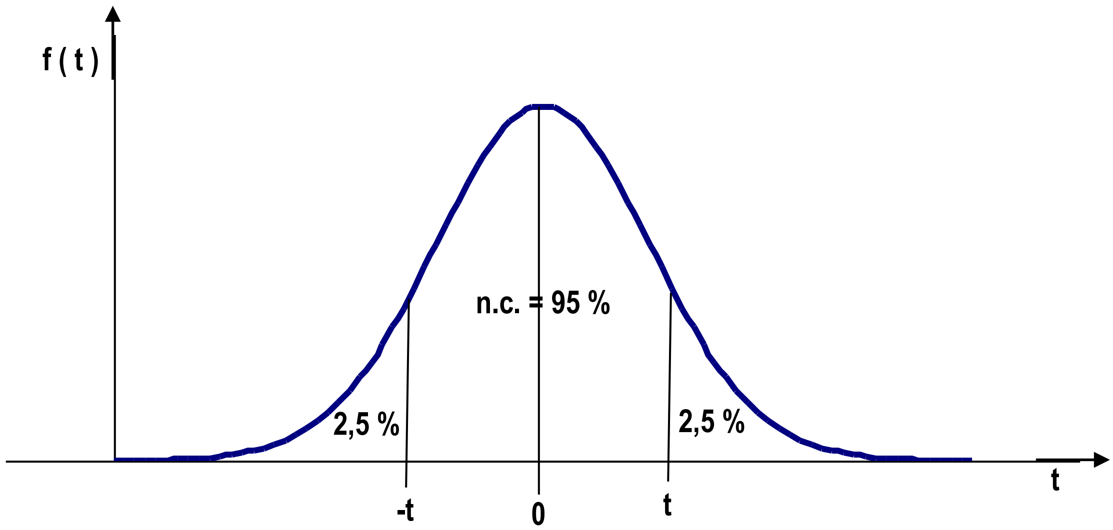
Como o desvio padrão populacional , σ , não é conhecido e a amostra disponível é pequena (n = 25 ≤ 30), devemos utilizar a 2ª fórmula. Porém, esta exige que a população em estudo seja uma população normal. Como não temos a certeza desse facto e não dispomos dos valores amostrais para realizar o teste simples, temos que o assumir. Assim:

Assumindo população normalmente distribuída:

$$\bar{x} \pm t \frac{s}{\sqrt{n}} \qquad t \sim T\,(\,GL\,) \quad \text{com} \quad GL = n - 1$$

2º passo: Obtenção do valor crítico, t

Sendo o nível de confiança, n.c., igual a 95 %, devemos fazer a seguinte esquematização da distribuição t de Student [2]:



Por comparação com o esquema presente no topo da tabela t de Student, vemos que α = 2,5 % = 0,025.
GL = n – 1 = 25 – 1 = 24
Por consulta da tabela, vem: ***t = 2,064.***

3º passo: Cálculo dos limites do I.C.

Por aplicação direta da fórmula selecionada, vem:

$$\bar{x} \pm t \frac{s}{\sqrt{n}} = 187{,}2 \pm 2{,}064 \times \frac{31{,}9}{\sqrt{25}} = 187{,}2 \pm 13{,}2 = 174{,}0 \;;\; 200{,}4$$

Logo, o I.C. para a média populacional a 95 % de confiança é:

[174,0 200,4] *ou* ***187,2 ± 13,2***.

O significado deste resultado é o seguinte: Apesar de não conhecermos a média populacional, sabemos agora que a probabilidade de ela se encontrar entre 174,0 e 200,4 é de 95 % (o valor do nível de confiança utilizado na construção do intervalo).

[2] NOTE: No esquema, indicam-se percentagens da área total sob o gráfico.

Exemplo 3: *Tome novamente a amostra anterior, mas considere agora que o desvio padrão populacional é conhecido e é: $\sigma = 31,9$.*
Construa novamente um intervalo de confiança para o valor médio da população da qual a amostra foi extraída, com uma confiança de 95 %.

1º passo: Seleção da fórmula a utilizar

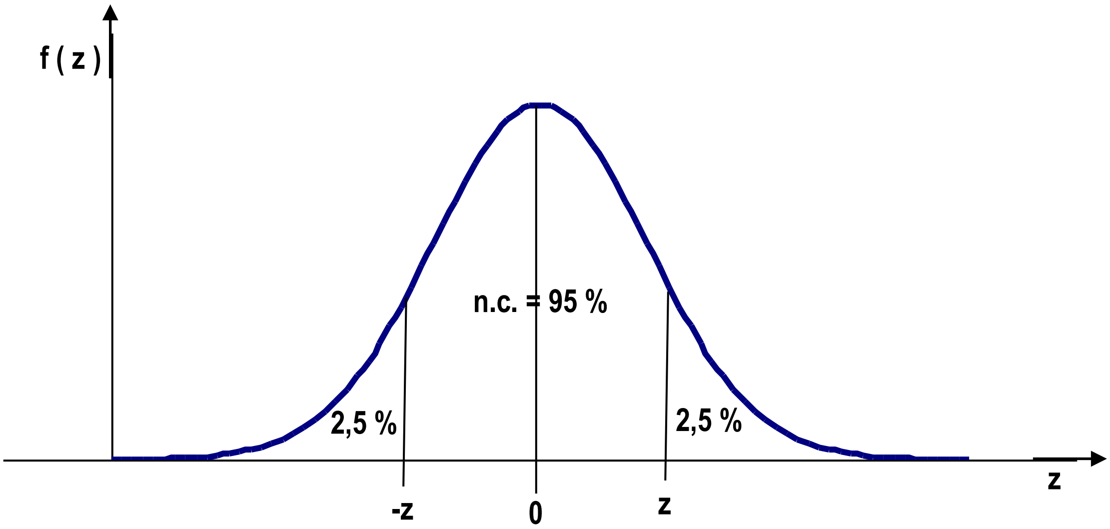
Como agora o desvio padrão populacional , σ , é conhecido devemos utilizar a 1ª fórmula. Novamente, como a amostra é de pequena dimensão exige-se que a população em estudo seja uma população normal. Como não temos a certeza desse facto, temos que o assumir. Assim:

Assumindo população normalmente distribuída:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \qquad z \sim N(0,1)$$

2º passo: Obtenção do valor crítico, z

Sendo o nível de confiança, n.c., igual a 95 %, devemos fazer a seguinte esquematização da distribuição normal reduzida, aliás semelhante à realizada no exemplo anterior para a distribuição t de Student:



Como já foi referido, a distribuição normal reduzida é equivalente à distribuição t de Student com GL = ∞. Esta consulta é mais cómoda nestes casos.

Por comparação com o esquema presente no topo da tabela t de Student, vemos que α = 2,5 % = 0,025.
Tomando GL = ∞ , por consulta da tabela, vem: ***z = 1,960.***

3º passo: Cálculo dos limites do I.C.

Por aplicação direta da fórmula selecionada, vem:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} = 187,2 \pm 1,960 \times \frac{31,9}{\sqrt{25}} = 187,2 \pm 12,5 = 174,7 \; ; \; 199,7$$

Logo, o I.C. para a média populacional a 95 % de confiança é:

[174,7 199,7] *ou* ***187,2 ± 12,5****.*

Cujo significado é: a probabilidade de o verdadeiro valor da média populacional se encontrar entre 174,7 e 199,7 é de 95 %.

Note que, apesar de os valores empregues nos exemplos 2 e 3 serem semelhantes, o segundo I.C. é mais estreito do que o primeiro. Isto deve-se ao facto de que, no último, conhecemos o desvio padrão populacional, valor que nos oferece maior confiança do que o do desvio padrão amostral.

Esta observação levanta-nos uma nova questão: Como podemos conhecer o verdadeiro valor do desvio padrão populacional, σ ?

Em algumas situações já muito estudadas, pode-se admitir sem grande erro que, apesar de o valor médio variar, o desvio padrão mantém-se constante ou aproximadamente constante. Assim, juntando valores recolhidos ao longo de muito tempo, consegue-se uma estimativa do desvio padrão populacional já tão segura que podemos afirmar sem grande erro que encontrámos o verdadeiro valor na população.

Há ainda outros casos nos quais, conhecendo-se bem populações semelhantes, podemos assumir sem grande erro que o desvio padrão da população em estudo será igual.

As estimativas da variância e do desvio padrão populacionais são o tema da próxima secção.

2.3 Intervalo de confiança para a variância populacional

Dados necessários: - dimensão amostral, n
- desvio padrão amostral, s
- pré-definição do nível de confiança, n.c.

<u>Caso população normal, para qualquer n:</u>

$$\frac{(n-1)\,s^2}{b} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{a} \qquad a\,,\,b\ \sim \mathrm{X}^2\,(\,GL\,) \quad \text{com} \quad GL = n-1 \ \text{ e } \ a < b$$

Quando não haja certeza de que a população seja normal, poderemos verificar o pressuposto da normalidade pelo método simples apresentado no capítulo anterior (há outros métodos que não fazem parte do conteúdo deste módulo).

$a\,,\,b\ \sim \mathrm{X}^2$ (GL) são dois valores (diferentes, não simétricos) que seguem a **distribuição do qui-quadrado**. Esta distribuição difere bastante das anteriores nos valores da variável, que são sempre positivos, e também por não ser uma distribuição simétrica. Por esse motivo, há sempre que determinar separadamente os dois valores a e b.

Os valores desta distribuição encontram-se tabelados na bibliografia (ver anexos), em função de um parâmetro GL (graus de liberdade) que, em cada caso deverá ser correctamente calculado.

Exemplo 4: *Considere novamente a amostra do exemplo 2.*
Supondo novamente que o desvio padrão populacional não é conhecido, construa um intervalo de confiança para o desvio padrão da população da qual a amostra foi extraída, com uma confiança de 95 %.

Como não existe uma fórmula para a construção de um I.C. para o desvio padrão, temos que estimar primeiro a variância e posteriormente relacioná-la com o desvio padrão.

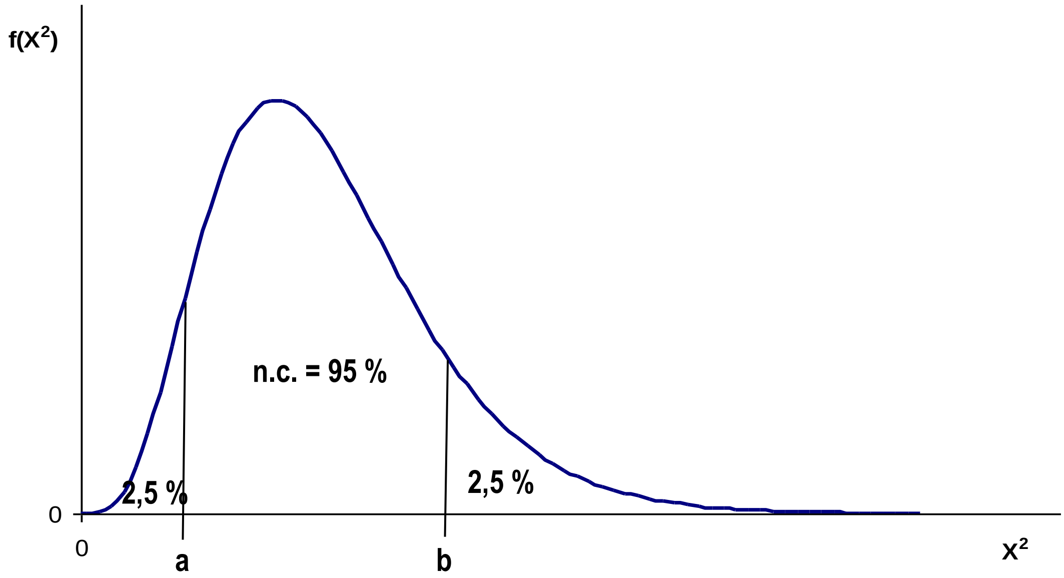
<u>1º passo: Seleção da fórmula a utilizar</u>

Assumindo população normalmente distribuída:

$$\frac{(n-1)\,s^2}{b} \le \sigma^2 \le \frac{(n-1)\,s^2}{a} \qquad a\,,\,b \sim X^2\,(\,GL\,) \quad \text{com} \quad GL = n-1 \;\; e \;\; a < b$$

<u>2º passo: Obtenção dos valores críticos, a e b</u>

Sendo o nível de confiança, n.c., igual a 95 %, devemos fazer a seguinte esquematização da distribuição do qui-quadrado:



Por comparação com o esquema presente no topo da tabela X^2, vemos que α = 2,5 % = 0,025, para b e α = 95 % + 2,5 % = 97,5 % = 0,975, para a.

Quando usamos esta distribuição, há a necessidade de determinar ambos os valores, (a e b) visto ela não ser simétrica (o que não acontecia com as distribuições normal reduzida e t de Student, simétricas).

Tomando GL = n – 1 = 25 – 1 = 24 , por consulta da tabela, vem:
a = 12,40 *e* ***b = 39,36.***

<u>3º passo: Cálculo dos limites do I.C.</u>

Por aplicação direta da fórmula selecionada, vem:

$$\frac{(n-1)\,s^2}{b} \le \sigma^2 \le \frac{(n-1)\,s^2}{a} \Leftrightarrow \frac{(\,25-1)\times 31{,}9^{\,2}}{39{,}36} \le \sigma^2 \le \frac{(\,25-1)\times 31{,}9^{\,2}}{12{,}40}$$

$$\Leftrightarrow 620{,}91 \le \sigma^2 \le 1969{,}6$$

Logo, o I.C. para a variância populacional a 95 % de confiança é:

***[620,91 ; 1969,6]**.*

Cujo significado é: a probabilidade de o verdadeiro valor da variância populacional se encontrar entre 620,91 e 1969,6 é de 95 %.

Como, por definição, $\sigma = \sqrt{\sigma^2}$, temos então que o intervalo de confiança para o desvio padrão populacional é:

$$[\sqrt{620{,}91}\;;\;\sqrt{1969{,}6}\,] = \boldsymbol{[\,24{,}9\,;\,44{,}4]}$$

2.4 Intervalo de confiança para a proporção populacional, π

Dados necessários: - dimensão amostral, n

- proporção amostral, f ou número de ocorrências na amostra, x , sendo f = x / n
- pré-definição do nível de confiança, n.c.

<u>Para n > 30:</u>

$$f \pm z \sqrt{\frac{\pi.(1-\pi)}{n}} \;,\; f = \frac{x}{n} \qquad z \sim N(0,1)$$

Como π não é conhecido, estima-se o seu valor por f : $\hat{\pi} = f$, pelo que na prática aquela fórmula transforma-se em:

$$f \pm z \sqrt{\frac{f.(1-f)}{n}}$$

Se desejado, pode multiplicar-se ambos os limites por 100, para obter a estimativa em percentagem de ocorrências.

z ~ N (0,1) é um valor que segue a **distribuição normal reduzida** ou **padronizada** e tem o significado já atrás apresentado.

Exemplo 5: *Numa amostra de 100 vacas, 7 mostraram-se infectadas pela "doença das vacas loucas". Estime a taxa (percentagem) de vacas infectadas com aquela doença. Utilize um nível de confiança de 99 %.*

<u>1º passo: Seleção da fórmula a utilizar</u>

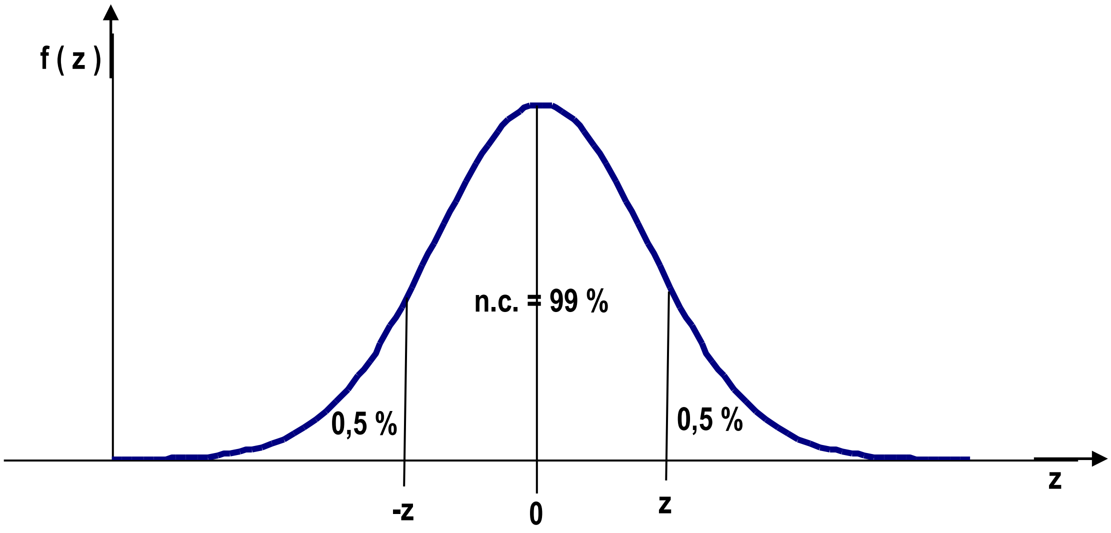
Sendo n = 100 > 30, podemos utilizar a fórmula atrás apresentada:

$$f \pm z \sqrt{\frac{f.(1-f)}{n}} \;,\; f = \frac{x}{n} \qquad z \sim N(0,1) \text{ , sendo } \hat{\pi} = f$$

.

<u>2º passo: Obtenção do valor crítico, z</u>

Sendo o nível de confiança, n.c., igual a 99 %, devemos fazer a seguinte esquematização da distribuição normal reduzida:



Usaremos novamente a tabela t de Student com GL = ∞.

Por comparação com o esquema presente no topo da tabela t de Student, vemos que α = 0,5 % = 0,005.
GL = ∞
Por consulta da tabela, vem: ***z = 2,576.***

<u>3º passo: Cálculo dos limites do I.C.</u>

Sendo n = 100 e x = 7, por aplicação direta da fórmula selecionada, vem:

$$f = \frac{x}{n} = \frac{7}{100} = 0{,}070$$

$$f \pm z\sqrt{\frac{f.(1-f)}{n}} = 0{,}070 \pm 2{,}576 \times \sqrt{\frac{0{,}07 \times (1-0{,}07)}{100}} = 0{,}070 \pm 0{,}066 = 0{,}004 ; 0{,}136$$

Logo, o I.C. para a proporção populacional a 99 % de confiança é:

[0,004 ; 0,136] *ou* ***0,070 ± 0,066****.*

Multiplicando ambos os limites por 100, obtemos a estimativa em percentagem:

[0,4 % ; 13,6 %] *ou* ***7,0 % ± 6,6****.*

O significado deste resultado é o seguinte: Apesar de não conhecermos a verdadeira percentagem de animais infectados na população, sabemos agora que há 99 % de probabilidade de se encontrarem infectados entre 0,4 % a 13,6 % da totalidade dos animais.

3. Dimensionalização de amostras

Embora existam vários métodos de dimensionalização de amostras, vamos aqui concentrar-nos apenas em dois – os mais comuns - um relacionado com a estimação de uma média populacional e outro adequado para a estimação de uma proporção populacional.

Estes métodos permitem-nos encontrar um valor mínimo para a dimensão amostral, $\mathbf{n_{mín}}$, mediante condições pré-definidas.

Cabe depois ao investigador definir uma dimensão amostral igual ou superior a esta, consoante os custos e o esforço envolvido na obtenção de cada valor, mas nunca deverá aplicar uma dimensão inferior, sob pena de não obter a precisão pretendida.

3.1 Dimensionalização com base no erro da estimativa da média populacional

A largura ou amplitude de um intervalo de confiança para a média pode ser calculada por (caso população normal e/ou amostra grande):

$$A = l_{sup} - l_{inf} = 2.z.\frac{\sigma}{\sqrt{n}}$$

Na mesma situação, a incerteza da estimativa (definida como sendo metade da amplitude intervalar) vem dado por:

$$\text{Incerteza} = ½ (l_{sup} - l_{inf}) = z.\frac{\sigma}{\sqrt{n}}$$

Para dimensionar uma amostra impõe-se que a amplitude ou então aincerteza da estimativa não sejam superiores a um valor pré-definido que estamos dispostos a aceitar. Assim:

<u>Se definimos $A < A_{máx}$</u> $\rightarrow$ $\mathbf{n >} \left(\frac{2.z.\sigma}{A_{máx}} \right)^2$

<u>Se definimos $\text{Incerteza} < \text{Incerteza}_{máx}$</u> $\rightarrow$ $n \geq \left(\frac{z.\sigma}{\text{Incerteza}_{máx}} \right)^2$

<u>Em qualquer caso, o resultado do cálculo deve ser arredondado para o inteiro imediatamente superior, dado que n tem que ser um número inteiro.</u>

Obtém-se assim a **dimensão mínima amostral**, $\mathbf{n_{mín}}$.

Este valor será válido desde que maior 30 e/ou desde que a população seja normal.

Quando σ for conhecido, o seu valor deverá ser usado na dimensionalização da amostra. Quando isso não acontecer, recolhe-se uma amostra pequena, designada por **amostra piloto**, e calcula-se o seu desvio padrão. Estima-se o valor de σ a partir deste resultado: $\hat{\sigma} = s_{piloto}$ e realizam-se os cálculos de forma semelhante.

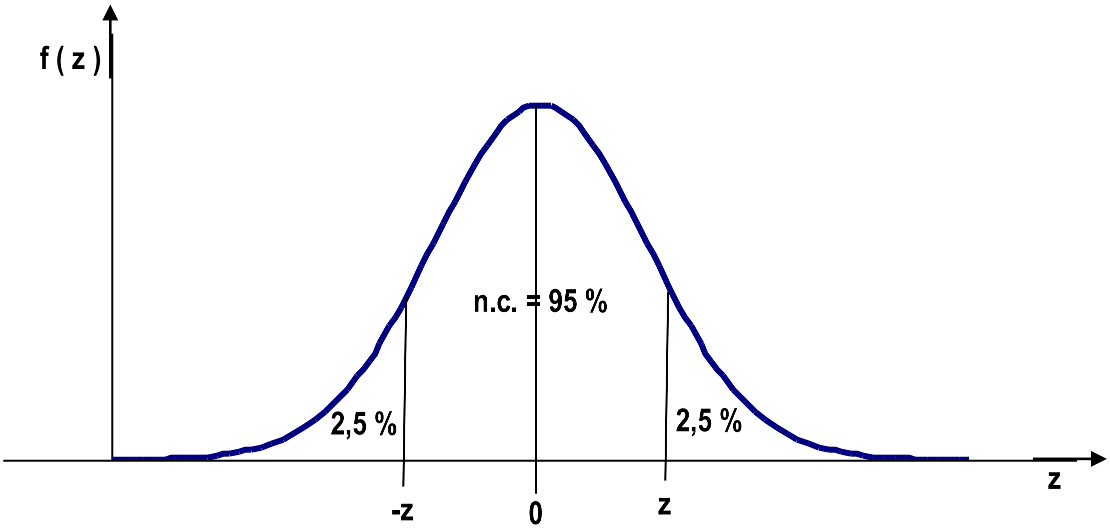
Exemplo 10: *Considere novamente os dados do exemplo 2.*
Pretende-se obter uma nova estimativa para a média populacional, mas desta vez mais precisa, com uma incerteza não superior a 5 unidades, também com um nível de confiança de 95 %. Qual deverá ser a dimensão da nova amostra a recolher ?

Começamos por assumir que n > 30, pois não temos a certeza de se tratar de uma população normal. Este pressuposto poderá ser verificado no final.

Como não conhecemos o desvio padrão populacional, para efeitos da determinação da dimensão amostral, estimamos o seu valor usando a amostra disponível que assume assim aqui o papel de amostra piloto:

$$\hat{\sigma} = s_{piloto} = 31,9$$

Precisamos também do valor de z. Para um nível de confiança de 95 %, vem:



Por consulta da tabela t de Student, com GL = ∞ e α = 0,025, tem-se que z = 1,960.
Portanto:

$$n > \left(\frac{z.\sigma}{\text{Incerteza}_{\text{máx}}}\right)^2 \Leftrightarrow n > \left(\frac{1,960 \times 31,9}{5}\right)^2 \Leftrightarrow n > 156,4 \rightarrow \boldsymbol{n_{mín} = 157}$$

Para atingir a precisão pretendida, sem baixar o nível de confiança, deverá ser recolhida uma nova amostra com dimensão n = 157 ou superior.
Note ainda que se confirma o pressuposto de que n > 30, pelo que este resultado é válido, ainda que a população em causa não seja normal.

3.2 Dimensionalização com base no erro da estimativa da proporção populacional

Sendo n > 30, a largura ou amplitude de um intervalo de confiança para a proporção vem dada por:

$$A = l_{sup} - l_{inf} = 2 . z \sqrt{\frac{\pi.(1-\pi)}{n}}$$

e a incertza da estimativa vem dado por:

$$\text{Incerteza} = ½ \,(l_{sup} - l_{inf}) = z \sqrt{\frac{\pi.(1-\pi)}{n}}$$

Tal como no caso anterior, para dimensionar uma amostra impõe-se que a amplitude ou então a incerteza da estimativa não sejam superiores a um valor pré-definido que estamos dispostos a aceitar. Assim:

Se definimos $A < A_{máx}$ → **n >** $\frac{4.z^2.\pi.(1-\pi)}{A_{máx}{}^2}$

Se definimos Incerteza < $\text{Incerteza}_{máx}$ → $n \geq \left(\frac{z}{\text{Incerteza}_{máx}} \right)^2 . \pi(1-\pi)$

Em qualquer caso, o resultado do cálculo deve ser arredondado para o inteiro imediatamente superior, dado que n tem que ser um número inteiro.

Obtém-se assim a **dimensão mínima amostral** , $\mathbf{n_{mín}}$.

Como π não é conhecido, mas é necessário no cálculo, recorre-se a uma das seguintes estratégias para obter um valor:

- Recolhe-se de uma pequena amostra, designada por **amostra piloto**, e calcula-se o respectivo f. Estima-se o valor de π a partir deste resultado: $\hat{\pi} = f_{piloto}$.

- Caso se desconfie fortemente de um determinado valor para π que apenas se pretende confirmar ou precisar, usa-se esse valor para efeitos de dimensionalização da amostra.

- Caso nada seja conhecido e não haja tempo ou seja dispendioso ou incómodo recolher uma amostra piloto, usa-se simplesmente $\hat{\pi} = 0{,}5$.

Exemplo 11: *Considere novamente o exemplo 5. Pretende-se obter uma nova estimativa da taxa de incidência da doença, mas desta vez com um erro não superior a 2,5 %, mantendo-se o mesmo nível de confiança. Quantas vacas deverão ser submetidas ao teste ?*

Vamos assumir que n > 30.
Como não conhecemos o valor de π, mas existe uma amostra anterior, vamos utilizar o resultado aí obtido. A amostra anterior assume assim o papel de uma amostra piloto.

$\hat{\pi} = f_{piloto} = 0{,}07$

Dado que nem o nível de confiança, nem a distribuição (normal reduzida) se alteraram em relação ao exemplo 5, temos que . z = 2,576.
Portanto:

$$n > \frac{z^2 . \pi . (1 - \pi)}{\text{Incerteza}_{\text{máx}}^{\;2}} \Leftrightarrow n > \frac{2{,}576^2 . 0{,}07 . (1 - 0{,}07)}{0{,}025^2} \Leftrightarrow n > 691{,}2 \rightarrow \boldsymbol{n_{mín} = 692}$$

Para obter uma estimativa da taxa de incidência da doença, a um nível de confiança de 99 %, com um erro máximo de 2,5 %, deverão ser submetidas ao teste pelo menos 692 vacas.
Note ainda que se confirma o pressuposto assumido inicialmente de que n > 30.

- FIM -