Métodos Estatísticos – L.EIC

Semana 8

Aula 6

9 de maio de 2022

Métodos Estatísticos – L.EIC

Aula 6

Inferência Estatística

Estimação Paramétrica e Testes de Hipóteses

Erro Padrão da Média Amostral

Intervalos de Confiança

Intervalo de Confiança para a Média

Caso de População Normal com Variância Conhecida

Caso de População Normal com Variância Desconhecida

Distribuição t de Student

Exemplos

5

INFERÊNCIA ESTATÍSTICA
– Intervalos de Confiança –

Introdução à Inferência Estatística

Inferência Estatística
Introdução

Os estudos e análises estatísticas desenvolvidos no domínio da **Estatística Descritiva** consistem, essencialmente, na **organização, apresentação e caracterização** de conjuntos de dados.

Na grande maioria das situações pretende-se, posteriormente, **inferir** **sobre características** de interesse **da população** da qual foram obtidos os dados. Esse é o domínio da designada **Inferência Estatística**.

Inferência Estatística
Introdução

Este é o ramo da Estatística que iremos agora abordar. O seu objetivo consiste em **tirar conclusões acerca de uma população a partir do estudo de uma amostra** dessa mesma população e que designamos então por Inferência Estatística.

Comecemos por constatar uma diferença fundamental entre a Inferência Estatística e as Probabilidades, que abordámos anteriormente.

Inferência Estatística
Introdução

Probabilidades – raciocínio dedutivo

Passamos do geral ao particular

Exemplo: Admitindo que uma moeda é não viciada (parâmetro $p = 0.5$), qual a probabilidade de se observarem 17 caras em 40 lançamentos?

Inferência Estatística – raciocínio indutivo

Passamos do particular ao geral

Exemplo: Dado que se observaram 17 caras em 40 lançamentos de uma moeda, que evidência é que este facto nos dá sobre se a moeda é viciada ou não?

Estimação de Parâmetros

Inferência Estatística
Estimação de Parâmetros

Estimação paramétrica

Na estimação paramétrica, pretende-se **calcular**, a partir da análise de uma amostra, **um valor** (ou intervalo de valores) que sirva como **aproximação do correspondente parâmetro** na população.

Testes de Hipóteses

Num problema de Teste de Hipóteses, pretende-se, através da **análise de uma amostra**, responder a uma **questão sobre a característica em estudo da população**. Essa questão é formulada como uma hipótese.

Estimativa de um Parâmetro de uma População

Na **Estimação Paramétrica**, a partir de uma amostra selecionada aleatoriamente da população em estudo, i.e., a partir de observações da população, pretende-se:

- **obter uma estimativa** para um parâmetro da população;
- **avaliar a qualidade** da estimativa.

Estimação Paramétrica – **Exemplo**

Foram registados os pesos à nascença de uma amostra aleatória de 28 cordeiros. Os resultados estão na tabela seguinte:

Peso à nascença (Kg)						
4.3	5.2	6.2	6.7	5.3	4.9	4.7
5.5	5.3	4.0	4.9	5.2	4.9	5.3
5.4	5.5	3.6	5.8	5.6	5.0	5.2
5.8	6.1	4.9	4.5	4.8	5.4	4.7

A partir destes dados que informação se pode extrair sobre μ e σ (respetivamente a média e o desvio padrão da população)?

Estimação Paramétrica – **Exemplo**

Naturalmente que as **estimativas mais naturais** para a média μ e para o desvio padrão σ da população são respetivamente a **média** $\overline{x}$**, e o desvio padrão** s**, da amostra**.
Para os dados da tabela anterior, temos então,

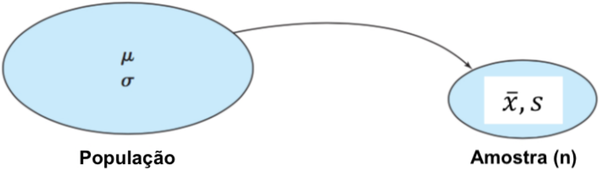
- $\overline{x} = 5,17\,\mathrm{Kg}$ é uma estimativa para μ
- $s = 0,65\,\mathrm{Kg}$ é uma estimativa para σ

Mas temos que ter em consideração que estas estimativas estão sujeitas ao **erro de amostragem**. Estamos a considerar apenas 28 cordeiros e não todos.

Como poderemos avaliar a qualidade daquelas estimativas?

Estimação Paramétrica – **Exemplo**

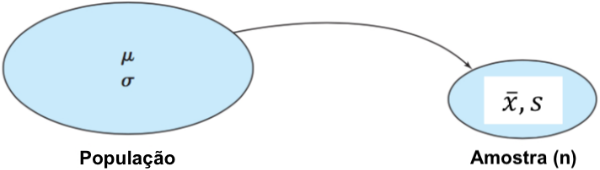
Em geral, com base numa amostra representativa da população, usa-se a média e variância amostrais para estimar os parâmetros correspondentes da população.



No exemplo anterior, as observações deveriam satisfazer certas condições. **Em particular:** carneiros nascidos no mesmo mês, pais com a mesma alimentação, carneiros do mesmo tipo escolhidos ao acaso, ...

Estimação Paramétrica – **Exemplo**

Na impossibilidade de se analisar a população completa, os valores fornecidos pelos estimadores serão sempre 'imperfeitos'.



Variabilidade: Já vimos, no estudo da distribuição por amostragem, que estas estimativas estão sujeitas a variabilidade, a qual é caracterizada pela distribuição por amostragem dos estimadores, como é o caso da estimação de μ, a média da população.

Erro Padrão da Média

Vimos anteriormente que sendo a média amostral $\overline{X}$, uma v.a. baseada em n v.a. $X_1, X_2, \ldots, X_n$, independentes e identicamente distribuídas (i.i.d.), com média μ e desvio padrão σ, temos:

$$E(\overline{X}) = \mu \quad \text{e} \quad V(\overline{X}) = \frac{\sigma^2}{n}$$

Podemos distinguir 2 casos.

- Caso 1 – σ conhecido

  O desvio padrão de $\overline{X}$ é neste caso dado por

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Erro Padrão da Média – **Exemplo**

- Caso 2 – σ desconhecido

  Como o desvio padrão da amostra que temos (s) é uma estimativa para σ, a estimativa natural para $\sigma_{\bar{X}}$ é:

$$s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

  onde s é o desvio padrão, e n a dimensão da amostra.

$s_{\bar{X}}$ designa-se por **desvio padrão da média amostral**, ou, **erro padrão da média**, e é normalmente representado por se, e o seu valor é dado então por:

$$se = \frac{s}{\sqrt{n}}$$

Erro Padrão da Média – **Exemplo**

Para o exemplo anterior (peso dos cordeiros), tínhamos

$$\bar{x} = 5,17\,\mathrm{Kg}; \quad s = 0,65\,\mathrm{Kg}; \quad n = 28$$

e o erro padrão da média é então:

$$s_{\bar{X}} \equiv se_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{0.65}{\sqrt{28}} = 0,12\,\mathrm{Kg}$$

$s \equiv s_X$ e $se_{\bar{X}}$ são duas medidas de dispersão:

s_X é relativo aos **pesos observados**;

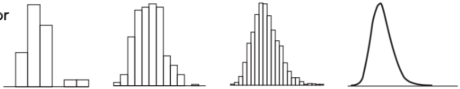
$s_{\bar{X}} \equiv se_{\bar{X}}$ é relativo ao **peso médio**.

Erro Padrão (*se*) e Desvio Padrão (*s*)

Sendo $\overline{X}$ um estimador de μ, podemos então observar que:

- *s* descreve a **variabilidade, na amostra, dos pesos à nascença**;
- *se* indica a **variabilidade associada com a média amostral**, considerada como estimativa da média da população dos pesos à nascença; i.e., fornece uma indicação sobre a **qualidade da estimação**.

Comportamento de *s* e de *se* com *n*:

	$n = 28$	$n = 280$	$n = 2{,}800$	$n \to \infty$
$\bar{x}$	5.17	5.19	5.14	$\bar{y} \to \mu$
s	0.65	0.67	0.65	$s \to \sigma$
se	0.12	0.040	0.012	$SE \to 0$
Distribuição por amostragem				

Erro Padrão (se) e Desvio Padrão (s)

Os termos **erro padrão** (se) e **desvio padrão** (s) são muitas vezes confundidos, mas estão associados a aspetos muito distintos:

- O desvio padrão s refere-se à **dispersão dos dados** que constituem a **amostra**.
- O erro padrão se, sendo o desvio padrão da distribuição (por amostragem) da média amostral $\overline{X}$, descreve a **fiabilidade da média da amostra** como estimativa da média da população.

Erro Padrão (se) e Desvio Padrão (s)

Assim,

- se pretendermos descrever a **variabilidade numa amostra**, devemos usar o desvio padrão s.
- Se o objetivo for indicar a **imprecisão associada à estimativa** $\overline{x}$, da média da população μ, utilizamos o erro padrão (**variabilidade associada com a média amostral**).

Apresentação de Resultados

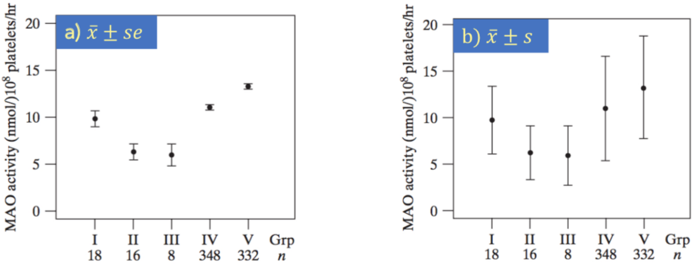
Em muitos relatórios científicos os dados são representados em tabelas.
Como apresentar as medidas s **(SD) e** se **(SE)?**
A tabela mostra as medições da atividade de uma enzima (MAO) nas plaquetas sanguíneas de 5 grupos de indivíduos (em nmol/108 plaquetas/hora):

- I, II e III: pacientes com diagnóstico de esquizofrenia;
- IV e V: grupos de controlo - pessoas saudáveis.

Grupo	n	Média	SE	SD
I	18	9.81	0.85	3.62
II	16	6.28	0.72	2.88
III	8	5.97	1.13	3.19
IV	348	11.04	0.30	5.59
V	332	13.29	0.30	5.50

Inferência Estatística
Estimação de Parâmetros

Apresentação de Resultados



a) Indicador da **variabilidade da média** em cada grupo (como estimativa da média da população correspondente)

b) Indicador da **variabilidade de MAO** dentro de cada grupo

Inferência Estatística
Estimação de Parâmetros

Uma **estimativa pontual de um parâmetro** desconhecido, mesmo que acompanhada do desvio padrão respetivo, fornece uma **informação** sobre o parâmetro a estimar que pode ser considerada bastante **incompleta**.

Por exemplo, uma **estimativa pontual** não é acompanhada por uma **medida de confiança**.

Os **intervalos de confiança** permitem tratar esse problema.

Na verdade, podemos estimar um parâmetro cujo valor se desconhece, utilizando **um intervalo (estimação intervalar)** e não **um valor (estimativa pontual)**

Intervalos de Confiança

Inferência Estatística
Intervalos de Confiança

Um intervalo de confiança (IC) **para um parâmetro** θ da população (cujo valor é desconhecido), é um **intervalo** construído a partir de uma amostra aleatória retirada da população e que **contém** θ **com uma certa garantia**.

A construção de um IC depende de diversos fatores, sendo que um deles é naturalmente o parâmetro da população que se pretende estimar.

Veremos a seguir o caso da construção de um **IC para a média** μ de uma **população** X **com distribuição normal**.

Intervalos de Confiança para μ

População Normal – Variância Conhecida

Inferência Estatística
Intervalos de Confiança

IC para μ – **População Normal – Variância Conhecida**

Suponha-se que queremos **estimar a média** μ de uma **população** normal **com variância conhecida** σ^2, através de um intervalo.

População: $X \sim N\left(\mu, \sigma^2\right)$
Amostra aleatória: $\left(X_1, X_2, \ldots, X_n\right) : X_i \sim N\left(\mu, \sigma^2\right)$

Como pretendemos **estimar a média populacional** μ, consideramos o estimador:

$$\overline{X} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

IC para μ – **População Normal – Variância Conhecida**

Neste caso, a construção de um intervalo de confiança é baseada no facto de **conhecermos a distribuição**, por amostragem, **da média amostral**:

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

Uma vez que para a população temos

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

já vimos antes que:

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

Inferência Estatística
Intervalos de Confiança

IC para μ – **População Normal – Variância Conhecida**

e como

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

sabemos que

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

Tentemos então construir um **IC centrado na média amostral**.

IC para μ – **População Normal – Variância Conhecida**

Um **IC centrado na média amostral**, é um intervalo da forma

$$]\overline{X} - a, \overline{X} + a[\,; \quad \text{com } a > 0$$

Vamos construir o intervalo impondo a condição:

$$P(\mu \in]\overline{X} - a, \overline{X} + a[) = 1 - \alpha$$

ou, escrito de outra forma,

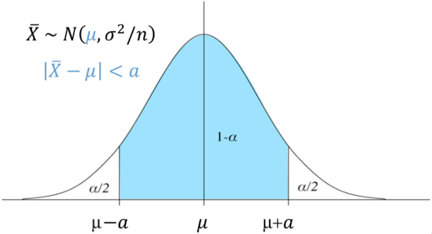
$$P(|\overline{X} - \mu| < a) = 1 - \alpha$$

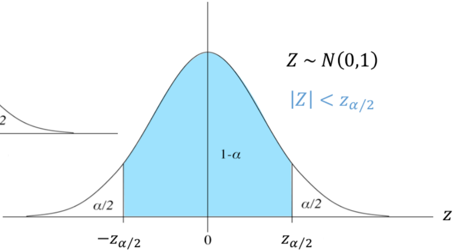
Ao valor $1 - \alpha$ chamamos **grau de confiança**.

Inferência Estatística
Intervalos de Confiança

IC para μ – **População Normal – Variância Conhecida**

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



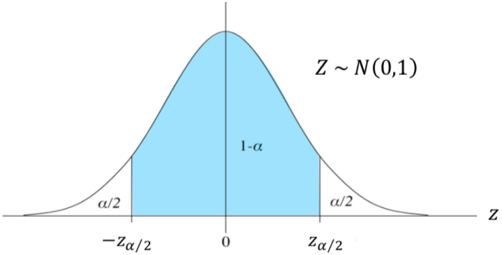


IC para μ – **População Normal – Variância Conhecida**

$$P(|\overline{X} - \mu| < a) = 1 - \alpha \Leftrightarrow P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

onde $z_{\alpha/2} = a \cdot \sqrt{n}/\sigma$

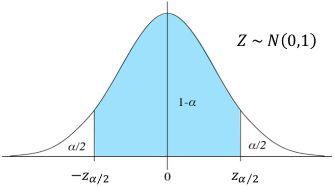


IC para μ – **População Normal – Variância Conhecida**

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

Alguns casos particulares:

$1-\alpha$	$z_{\alpha/2}$
90%	1.65
95%	1.96
99%	2.58



$$P\left(\overline{X} - 1.65\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.65\frac{\sigma}{\sqrt{n}}\right) = 0.90$$

$$P\left(\overline{X} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2.58\frac{\sigma}{\sqrt{n}}\right) = 0.99$$

Inferência Estatística
Intervalos de Confiança

IC para μ – **População Normal – Variância Conhecida**

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

Para $1 - \alpha = 0.95$, temos $Z_{\alpha/2} = 1.96$, e escrevemos:

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

E portanto, podemos dizer que 95% dos IC's

$$\overline{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

construídos a partir das amostras de dimensão n, contêm a média da população μ.

Nota: $\overline{X} \pm 1.96 \cdot \sigma/\sqrt{n}$, é um **intervalo aleatório**, não é um IC.

IC para μ – **População Normal – Variância Conhecida**

Ao substituirmos $\overline{X}$ por um valor observado da média de uma amostra específica $(\overline{x})$ passamos a ter um intervalo concreto, designado por **intervalo de confiança**

$$(\overline{x} \pm 1,96 \cdot \sigma/\sqrt{n})$$

Deste modo, um intervalo para a média de uma população normal $N\left(\mu, \sigma^2\right)$, com σ conhecido, com grau de confiança 95%, é:

$$(\overline{x} - 1,96 \cdot \sigma/\sqrt{n}, \overline{x} + 1,96 \cdot \sigma/\sqrt{n})$$

Também designado por **intervalo de confiança para μ a 95%.**

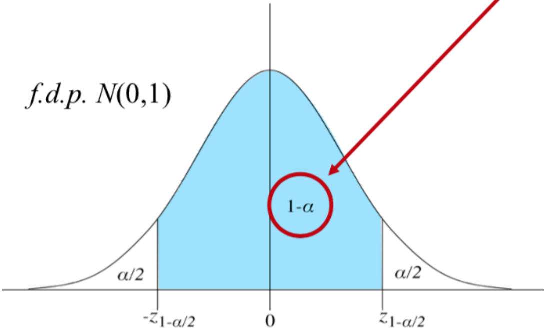
Inferência Estatística

Intervalos de Confiança

IC para μ – **População Normal – Variância Conhecida**

$$1-\alpha = 0.95; \quad z_{\alpha/2} = 1.96$$

$$Z \sim N(0,1) \Rightarrow P(-1.96 \leq Z \leq 1.96) = 0.95$$



IC para μ – **População Normal – Variância Conhecida**

Um **intervalo de confiança para** μ **a** 99%, **é construído do mesmo modo**, mas usando agora o **quantil adequado** (de ordem 0,995), da distribuição normal.

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

Assim

$$P(|\overline{X} - \mu| \leq 2.58 \cdot \sigma/\sqrt{n}) \approx 0,99$$

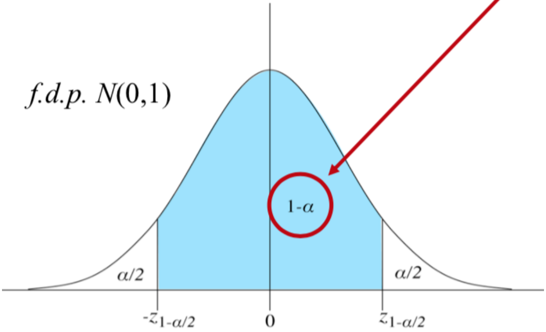
e obtemos o **intervalo aleatório**:

$$(\overline{X} \pm 2.58 \cdot \sigma/\sqrt{n})$$

IC para μ – **População Normal – Variância Conhecida**

$$1-\alpha = 0.99; \quad z_{\alpha/2} = 2.58$$

$$Z \sim N(0,1) \Rightarrow P(-2.58 \leq Z \leq 2.58) = 0.99$$



IC para μ – **População Normal – Variância Conhecida**

Para uma amostra particular, para a qual se observou uma média $\overline{x}$, obtemos o intervalo:

$$(\overline{x} \pm 2.58 \cdot \sigma/\sqrt{n})$$

que é, um **intervalo de confiança a** 99%, para a média μ de uma população normal, com desvio padrão conhecido.

Notar que este intervalo é **do mesmo tipo do intervalo anterior**, mas o fator **1.96** foi substituído por **2.58**, obtendo-se um **intervalo de maior amplitude**.

IC para μ – **População Normal – Variância Conhecida**

Notas Finais

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

Conduz ao **intervalo aleatório**:

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Esta é a **expressão geral do intervalo aleatório**, a partir do qual se obtém o IC que contém a média da população μ, para $(1-\alpha) \times 100\%$ das amostras de dimensão n.

Inferência Estatística
Intervalos de Confiança

IC para μ – **População Normal – Variância Conhecida**

Notas Finais

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

Conduz ao **intervalo aleatório**:

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Uma **amostra específica** fornece um **valor concreto** $\overline{x}$ para a **v.a.** $\overline{X}$, obtendo-se um **intervalo específico** designado por **intervalo de confiança** para μ **com grau de confiança** $1 - \alpha$.

Ao valor de $z_{\alpha/2}\,\sigma/\sqrt{n}$ chamamos a **margem de erro**, cuja **aproximação** deve ser feita **sempre por excesso**.

IC para μ – **População Normal – Variância Conhecida**

Notas Finais

Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de uma população, com uma distribuição que depende do parâmetro θ, cujo valor é desconhecido.

Suponhamos que Y_1 e Y_2 são **duas estatísticas**, com

$$Y_1 < Y_2$$

tais que

$$P\left(Y_1 < \theta < Y_2\right) \approx 1 - \alpha$$

IC para μ – **População Normal – Variância Conhecida**

Notas Finais

Então, para uma **amostra concreta** $(x_1, \ldots, x_n)$, o correspondente intervalo

$$(y_1, y_2)$$

diz-se um **intervalo de confiança** para θ a $100(1-\alpha)\%$

Nota: O intervalo (y_1, y_2), contém ou não θ. **Não existe nada de aleatório neste intervalo**.

Portanto, **não faz sentido falar na probabilidade de um intervalo particular** conter θ.

IC para μ – **População Normal – Variância Conhecida**

Notas Finais

Por exemplo, considere-se

$$P\left(Y_1 < \theta < Y_2\right) = 0.95$$

onde Y_1 e Y_2 são duas estatísticas construídas a partir de uma amostra aleatória $(X_1, \ldots, X_n)$.

A interpretação é a seguinte:

Considerando todas as amostras aleatórias de tamanho n que é possível retirar da população, e construindo os respetivos I.C., 95% desses intervalos irão conter o parâmetro θ.

Nota: $(1-\alpha)$ é o grau de confiança e α o nível de significância.

IC para μ – **População Normal – Variância Conhecida**

Exemplo 1

Numa certa fábrica, as máquinas de embalar cereais estão reguladas de modo a que a distribuição do peso das embalagens seja aproximadamente normal, com desvio padrão 5 g.

Escolhem-se ao acaso 36 embalagens, para as quais se registou um peso médio de 749.5 g.

A partir destes dados, pretende-se **estimar o peso médio** μ, de uma embalagem, através de um intervalo de confiança a 95%.

IC para μ – **População Normal – Variância Conhecida**

X – "v.a. que representa o peso de uma embalagem de cereais"

$$X \sim N\left(\mu, 5^2\right)$$

Como vimos, o intervalo de confiança a 95% para μ, será obtido a partir de:

$$\left(\overline{x} - 1.96\,\frac{5}{\sqrt{n}}, \overline{x} + 1.96\,\frac{5}{\sqrt{n}}\right)$$

ou seja:

$$\left(\overline{x} \pm 1.96\,\frac{5}{\sqrt{n}}\right)$$

IC para μ – **População Normal – Variância Conhecida**

IC a 95% para μ:

$$\left(\overline{x} \pm 1.96 \cdot 5/\sqrt{n}\right)$$

Nesta amostra, $n = 36$ e $\overline{x} = 749.5\,$g.

Então o IC pretendido é

$$(749.5 - 1.63; 749.5 + 1.63) \equiv (747.86; 751.13)$$

Assim, esta amostra permite-nos **ter 95% de confiança** de que o peso médio das embalagens empacotadas nas máquinas consideradas, se encontre entre 747.86 e 751.13 g.

IC para μ – **População Normal – Variância Conhecida**

Exemplo 2

Suponha-se agora que, **com o mesmo grau de confiança**, se pretendia estimar μ com uma determinada precisão.

Considere-se, por exemplo, que se pretendia o IC a 95%:

$$(\overline{x} \pm 0.9)$$

Vimos que o IC a 95% para o peso médio μ da população, é dado por:

$$\left(\overline{x} \pm 1.96 \, \frac{5}{\sqrt{n}}\right)$$

IC para μ – **População Normal – Variância Conhecida**

O IC a 95% é tal que $(\overline{x} \pm 0.9)$, e é dado por:

$$\left(\overline{x} \pm 1.96\,\frac{5}{\sqrt{n}}\right)$$

Então:

$$1.96\,\frac{5}{\sqrt{n}} \leq 0.9$$

Resolvendo em ordem a n virá

$$n \geq 118.6$$

O novo IC deveria então ser baseado numa amostra de pelo menos 119 embalagens.

IC para μ – **População Normal – Variância Conhecida**

Exercício Proposto

Considerar o IC do exemplo anterior, obtido a partir da amostra de 36 embalagens:

$$(747.86; 751.13)$$

Pretende-se estimar μ através de um IC a 95%, cuja **amplitude seja metade** da do IC anterior.

Qual deverá ser o valor de n, para se obter um tal intervalo?

População Normal – Variância Desconhecida

IC para μ – **Pop. Normal – Variância Desconhecida**

Esta é a situação mais comum. Como construir neste caso um intervalo de confiança para a média μ ?

população: $X \sim N\left(\mu, \sigma^2\right)$

amostra aleatória $(X_1, X_2, \ldots, X_n)$**:** $X_i \sim N\left(\mu, \sigma^2\right)$

Como a **variância populacional é desconhecida**, consideramos a variável aleatória:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \quad ; \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

IC para μ – **Pop. Normal – Variância Desconhecida**

Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de uma população com distribuição $N\left(\mu, \sigma^2\right)$.

Nesta situação (σ^2 desconhecido), para construir o IC **é necessário recorrer** a uma outra distribuição, a distribuição t **de Student**, em vez da distribuição normal.

À semelhança da distribuição normal, a distribuição t de Student também é uma **distribuição simétrica**, mas depende de **um só parâmetro**, o número de “**graus de liberdade**”.

IC para μ – **Pop. Normal – Variância Desconhecida**

Não trataremos aqui da justificação teórica, mas **pode mostrar-se** que a variável aleatória,

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tem uma **distribuição** t **de Student** com $n - 1$ graus de liberdade. Isto é:

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

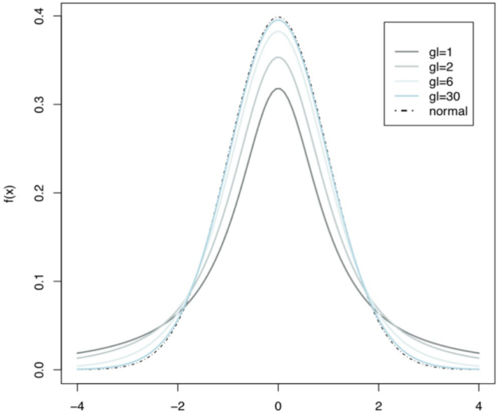
A forma da distribuição depende do número de graus de liberdade

Inferência Estatística

Distribuição t de Student

IC para μ – **Pop. Normal – Variância Desconhecida**

A Distribuição t de Student

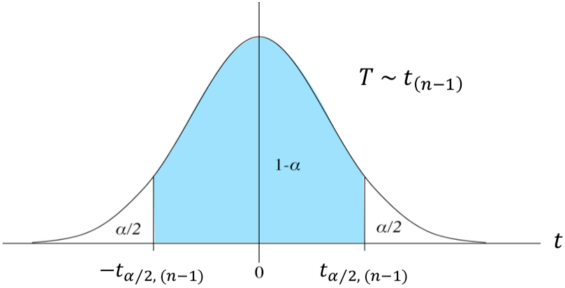


Inferência Estatística
Distribuição t de Student

IC para μ – **Pop. Normal – Variância Desconhecida**

Então

$$X \sim N\left(\mu, \sigma^2\right) \quad \Rightarrow \quad T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$



Inferência Estatística
Intervalos de Confiança

IC para μ – **Pop. Normal – Variância Desconhecida**

Procedendo como anteriormente

$$T \sim t_{(n-1)} \Rightarrow P\left(-t_{\alpha/2,(n-1)} \leq T \leq t_{\alpha/2,(n-1)}\right) = 1-\alpha$$

$$P\left(-t_{\alpha/2,(n-1)} \cdot S/\sqrt{n} \leq \overline{X} - \mu \leq t_{\alpha/2,(n-1)} \cdot S/\sqrt{n}\right) = 1-\alpha$$

$$P\left(\overline{X} - t_{\alpha/2,(n-1)} \cdot S/\sqrt{n} \leq \mu \leq \overline{X} + t_{\alpha/2,(n-1)} \cdot S/\sqrt{n}\right) = 1-\alpha$$

e obtemos então o IC aleatório

$$\left(\overline{X} \pm t_{\alpha/2,(n-1)} \cdot \frac{S}{\sqrt{n}}\right)$$

Se n é a dimensão da amostra, o número de graus de liberdade é $n-1$.

IC para μ – **Pop. Normal – Variância Desconhecida**

$$X \sim N\left(\mu, \sigma^2\right) \Rightarrow T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Para o caso $1 - \alpha = 0.95$, (**95% de confiança**), temos:

$$P\left(-t_{0.025,(n-1)} \cdot S/\sqrt{n} \leq \overline{X} - \mu \leq t_{0.025,(n-1)} \cdot S/\sqrt{n}\right) = 0.95$$

$$P\left(\overline{X} - t_{0.025,(n-1)} \cdot S/\sqrt{n} \leq \mu \leq \overline{X} + t_{0.025,(n-1)} \cdot S/\sqrt{n}\right) = 0.95$$

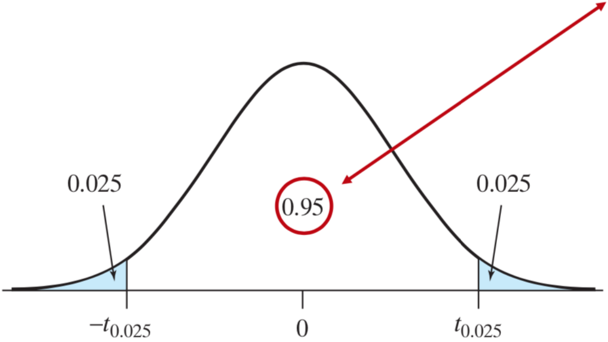
e o **IC aleatório**

$$\left(\overline{X} \pm t_{0.025,(n-1)} \cdot \frac{S}{\sqrt{n}}\right)$$

IC para μ – **Pop. Normal – Variância Desconhecida**

Caso $1 - \alpha = 0.95$, **(95% de confiança)**

$$T \sim t_{(n-1)} \Rightarrow P\left(-t_{0.025,(n-1)} \leq T \leq t_{0.025,(n-1)}\right) = 0.95$$



IC para μ – **Pop. Normal – Variância Desconhecida**

Para uma amostra concreta, de dimensão n, com média $\overline{x}$ e desvio padrão s, obtemos o intervalo

$$\left(\overline{x} - t_{\alpha/2,(n-1)} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2,(n-1)} \cdot \frac{s}{\sqrt{n}}\right)$$

que é um **intervalo de confiança**, com grau de confiança $1 - \alpha$ **para a média** μ de uma população normal com **desvio padrão desconhecido**, baseado na amostra referida.

IC para μ – Pop. Normal – Variância Desconhecida

Exemplo: Considerar novamente a amostra dos pesos registados à nascença, de uma amostra de 28 cordeiros e supor que se pretende **estimar o peso médio** à nascença da população através de um IC a 95%.

Admitindo a normalidade da população, o IC é da forma:

$$\left(\overline{x} \pm t_{\alpha/2,(n-1)} \cdot s/\sqrt{n}\right)$$

$n = 28; \quad \overline{x} \approx 5.1679; \quad s \approx 0.6544; \quad \alpha = 0.05; \quad t_{0.025,27} \approx 2.0518$

e o **IC será** então:

$$5.17 \pm 0.25 \quad \text{(por excesso)}$$

Inferência Estatística
Intervalos de Confiança

IC para μ – **Pop. Normal – Variância Desconhecida**

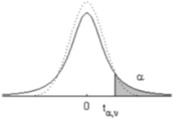
$$\alpha/2 = 0.025; \quad n-1 = 27 \text{ g.l.}; \quad t_{\alpha/2,(n-1)} = t_{0.025,27}$$

Distribuição t-Student: $t_{\nu,\alpha}$ **= quantil (1-α) de** t_ν

$X \sim t_\nu, \nu \in \mathbb{N}$

$P(X > t_{\alpha,\nu}) = 1 - F(t_{\alpha,\nu}) = \alpha.$

Isto é, $t_{\alpha,\nu}$ é o quantil $(1-\alpha)$ de t_ν.



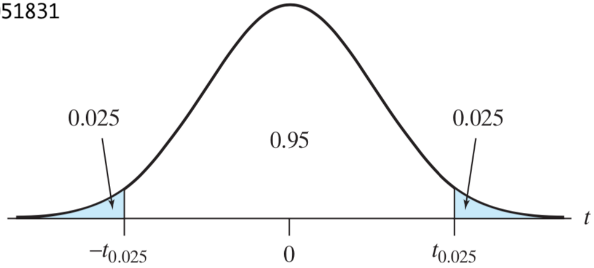
ν \ α	0.4	0.3	0.25	0.1	0.05	0.025	0.02	0.01	0.005	0.001
1	0.3249	0.7265	1.0000	3.0777	6.3138	12.7062	15.8945	31.8205	63.6567	318.3088
2	0.2887	0.6172	0.8165	1.8856	2.9200	4.3027	4.8487	6.9646	9.9248	22.3271
3	0.2767	0.5844	0.7649	1.6377	2.3534	3.1824	3.4819	4.5407	5.8409	10.2145
4	0.2707	0.5686	0.7407	1.5332	2.1318	2.7764	2.9985	3.7469	4.6041	7.1732
23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
24	0.2562	0.5314	0.6848	1.3178	1.7109	2.0639	2.1715	2.4922	2.7969	3.4668
25	0.2561	0.5312	0.6844	1.3163	1.7081	2.0595	2.1666	2.4851	2.7874	3.4502
26	0.2560	0.5309	0.6840	1.3150	1.7056	2.0555	2.1620	2.4786	2.7787	3.4350
27	0.2559	0.5306	0.6837	1.3137	1.7033	2.0518	2.1578	2.4727	2.7707	3.4210
28	0.2558	0.5304	0.6834	1.3125	1.7011	2.0484	2.1539	2.4671	2.7633	3.4082
29	0.2557	0.5302	0.6830	1.3114	1.6991	2.0452	2.1503	2.4620	2.7564	3.3962

IC para μ – **Pop. Normal – Variância Desconhecida**

$$\alpha/2 = 0.025; \quad n-1 = 27 \text{ g.l.}; \quad t_{\alpha/2,(n-1)} = t_{0.025,27}$$

No **R**, $t_{0.025,27}$ é dado por:

```
t = qt(0.975, df=27)
[1] 2.051831
```



IC para μ – **Pop. Normal – Variância Desconhecida**

Que **tipo de confiança** 'devemos' ter num **intervalo de confiança**?

No exemplo anterior o que significa dizer que temos uma confiança de 95% de que $\mu \in (4.92, 5.42)$?

O que podemos dizer é que 95% de **todos os intervalos possíveis** de construir a partir de todas as amostras aleatórias de dimensão 28, **contêm** μ.

Mas não podemos dizer que a probabilidade de μ pertencer ao intervalo $(4.92,\ 5.42)$ é 0.95.

Na verdade, μ **pertence, ou não, a esse intervalo**, uma vez que é um parâmetro (valor numérico) embora desconhecido.

IC para μ – **Pop. Normal – Variância Desconhecida**

Em resumo, sendo $\overline{x}$ a média, s o desvio padrão, e n a dimensão da amostra, temos 2 casos distintos:

- variância $\left(\sigma^2\right)$ conhecida

$$\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \longrightarrow \quad N(0,1)$$

- variância $\left(\sigma^2\right)$ desconhecida

$$\overline{x} \pm t_{\alpha/2, n-1} se \quad \longrightarrow \quad t_{(n-1)}; \qquad \left(se = \frac{s}{\sqrt{n}}\right)$$