(Parte 1)

Capítulo 5- Introdução à Inferência estatística.

1.1) Introdução.(222)*[Vídeo 39]*

Na <u>inferência estatística</u>, analisamos e interpretamos amostras com o objetivo de tirar conclusões acerca da população de onde se extraiu a amostra.

(222) **Populaçã**o- é o conjunto de todos os elementos em estudo.

(222) **Amostr**a- é um subconjunto finito da população.

Nota: usamos amostras porque, muitas vezes não é viável utilizarmos toda a população.

Exemplo 1(223)

Exemplo 1

Impossibilidade de reutilização dos elementos da população analisados

Para testar a qualidade da produção diária (população) numa fábrica de bolachas seria impraticável prová-las todas, pois deixaria de haver artigo para comercializar.

Exemplo 2 (223)

Exemplo 2

Custos muito elevados para analisar todos os elementos da população

Consideremos uma unidade fabril que produz componentes eletrónicos e circuitos integrados. Seria praticamente impossível testar todos os componentes e circuitos, não só em termos de custos (muito dispendioso) como em termos de tempo.

Exemplo 3.

Exemplo 3

Impossibilidade de testar todos os elementos da população

Numa indústria metalúrgica, a produção de um determinado tipo de peças em aço é um processo contínuo, isto é, não se interrompe ao fim do dia para se retomar novamente no dia seguinte: produz-se dia e noite. A resistência das peças é testada antes de toda a produção estar disponível.

Nota: A amostra deverá ser representativa da população, caso contrário, não poderemos tirar conclusões fiáveis. Quando uma amostra não é representativa, dizemos que é **enviesada**.

Exemplo(extra) : Para sabermos a média das alturas dos alunos de uma escola, se apenas analisássemos alunos que jogam basquetebol.

Exercício 1(266).

1. Numa pequena composição, refira algumas das razões que levam à escolha de amostras para estudar uma população ao invés de utilizar a totalidade da mesma. Dê exemplos.

Resolução:

Usamos amostras porque, pode demorar demasiado tempo a recolher os dados da população, ter demasiados custos ou até ser impossível.

Exemplos: Obter a média dos pesos de todas as pedras de uma determinada praia de calhau. Pode não ser impossível pesar todas as pedras, mas seria muito difícil e pouco adequado realizar essa tarefa.

Exemplo: medir a temperatura do ar em todos os pontos da atmosfera terrestre. Seria impossível.

Exemplo: Provar todas as bolachas de uma determinada remessa de uma fábrica, para saber se têm bom sabor. Não seria impossível, mas não ficariam bolachas para vender.

Amostragem Probabilística / não probabilística.(224)

Existem vários métodos para selecionar uma amostra, que podemos dividir em dois grupos de amostragem:

- **Métodos de amostragem probabilística** (ou aleatória): qualquer elemento da população tem alguma probabilidade de fazer parte da amostra.
- **Métodos de amostragem não probabilística** (ou não aleatória): alguns elementos da população podem não ter possibilidade de ser selecionados para a amostra e a inclusão de elementos nesta tem, normalmente, a intervenção do fator humano.

1.1.1- Métodos de amostragem probabilística. (225).

1- amostragem aleatória simples de n elementos.

1. Amostragem aleatória simples de *n* elementos: é toda a amostra em que a probabilidade de qualquer outro conjunto com *n* elementos da população ser selecionado é a mesma (a recolha dos *n* elementos faz-se sequencialmente e sem reposição ou todos em simultâneo). Um processo muito comum na obtenção deste tipo de amostra é a utilização de números aleatórios (dados por tabelas ou gerados por computador).

Exemplo7 (225).

Exemplo 7

Para um determinado estudo, uma empresa precisa de uma amostra de dimensão 25, a extrair dos seus 2500 funcionários.

Para conseguir uma amostra aleatória simples basta atribuir a cada um dos funcionários da empresa um número entre 1 e 2500. Em seguida, com o auxílio de um computador, gera-se uma sucessão de 25 desses números e selecionam-se os funcionários a que correspondem os números extraídos.

2- amostragem aleatória de n elementos com reposição.

2. Amostragem aleatória de *n* elementos, com reposição: selecionam-se *n* elementos da população, utilizando um processo idêntico ao da seleção de uma amostra aleatória simples, mas, após a seleção de cada elemento, ele é reposto na população antes de se selecionar o próximo.

Na maior parte das situações estudadas em inferência estatística, a população é de grande dimensão ou de dimensão infinita, pelo que a recolha do mesmo elemento duas vezes tem probabilidade aproximadamente igual a zero. Assim, nestas condições, os processos de seleção sem reposição e com reposição são praticamente equivalentes. No entanto, em populações de pequena dimensão não é indiferente o processo de seleção utilizado, já que a subtração de alguns elementos poderá provocar alterações significativas na população.

3- Amostragem aleatória sistemática.

3. Amostragem aleatória sistemática: os elementos da população, ordenada por algum critério, são selecionados para a amostra através de um processo definido *a priori* (à exceção do primeiro, que é escolhido aleatoriamente).

Exemplo 8(226).

Exemplo 8

Consideremos a mesma empresa do Exemplo 4, na página 224, e os mesmos 2500 funcionários, dos quais continuamos a querer uma amostra de 25 elementos.

Se atribuirmos um número, de 1 a 2500, a cada um dos funcionários e escolhermos um elemento de 100 em 100, no final teremos uma amostra de 25 elementos selecionados de forma sistemática.

4- Amostragem aleatória estratificada. ***Exemplo 9***(226).

4. Amostragem aleatória estratificada: os elementos da população são pré-selecionados, segundo o estrato a que pertencem (social, cultural, etário, sexual, etc.). Em seguida, recolhem-se amostras aleatórias simples de cada estrato (proporcional à representatividade do estrato), que se juntam numa amostra única que é, sem dúvida, mais representativa da população.

Nota: Esta é a amostragem mais utilizada em estudos oficiais tais como eleições, onde temos em conta a o género (m/f), a região, a idade, o nível de escolaridade, o nível socioeconómico, etc… estas variantes costumam influenciar as escolhas dos partidos políticos… (sugestão: quando aceder a uma sondagem, consulte a ficha técnica.)

Exemplo 9

Suponhamos que se pretende realizar um estudo acerca de uma vasta região arbórea e que precisamos de uma amostra.

Podemos dividir a população de árvores segundo a espécie e, a partir daí, escolher uma amostra (proporcional) de cada espécie. Em seguida, juntamos as amostras de cada espécie numa única amostra representativa da população de árvores e, com esta, procedemos ao estudo visado.

5- Amostragem aleatória por conglomerados.(227)

5. Amostragem aleatória por conglomerados: consiste em selecionar, ao acaso, alguns conglomerados onde se encontram as características da população que queremos estudar. A amostra é constituída por todos os elementos dos conglomerados selecionados. A principal vantagem deste método de amostragem é a considerável redução de custos (por exemplo, em grandes pesquisas ou investigações científicas).

Exemplo: Quando estudamos a vida escolar dos estudantes, é costume estudar turmas inteiras. Isso permite compreender as interações … Ao estudar peixes, estudamos cardumes. Ao estudar lobos, estudamos alcateias, etc…

Exercício 5(266).

5. O agrupamento de escolas Pinheiro Manso tem 4365 alunos distribuídos da seguinte forma pelos diferentes ciclos:

- 1.º Ciclo: 937 alunos;
- 2.º Ciclo: 598 alunos;
- 3.º Ciclo: 1236 alunos;
- Secundário: 1594 alunos.

Pretende recolher-se uma amostra destes alunos com a finalidade de avaliar a qualidade das instalações escolares. Sabendo que a amostra deverá abranger 20% da população estudantil deste agrupamento, descreva como procederia, indicando o número de alunos, por ciclo, que deveriam fazer parte da amostra.
Apresente o resultado final arredondado às unidades. Nos cálculos intermédios, se proceder a arredondamentos, conserve pelo menos 4 c.d.

Resolução:

5. Como 20% dos alunos do agrupamento corresponde a $4365 \times 0{,}20 = 873$, este é o número total de alunos que deverão fazer parte da amostra.

Cálculo da proporção de alunos de cada ciclo:

1.º Ciclo: $\frac{937}{4365} \approx 0{,}2147$

2.º Ciclo: $\frac{598}{4365} \approx 0{,}1370$

3.º Ciclo: $\frac{1236}{4365} \approx 0{,}2832$

Ensino Secundário: $\frac{1594}{4365} \approx 0{,}3652$

Cálculo do número de alunos a selecionar de cada ciclo:

1.º Ciclo: $873 \times 0{,}2147 \approx 187$ alunos

2.º Ciclo: $873 \times 0{,}1370 \approx 120$ alunos

3.º Ciclo: $873 \times 0{,}2832 \approx 247$ alunos

Ensino Secundário: $873 \times 0{,}3652 \approx 319$ alunos

Assim, deverão fazer parte desta amostra 187 alunos do 1.º Ciclo, 120 do 2.º Ciclo, 247 do 3.º Ciclo e 319 do Ensino Secundário, num total de 873 alunos.

1.2- Parâmetro e estimador (ou estatística). Estimativa.(228).

Exemplo 1(228).

Exemplo 1

Nas duas escolas secundárias de À-das-Frações, realizou-se um estudo que visava determinar o valor médio μ das idades dos alunos do 11.º ano.

Para isso, recolheu-se uma amostra aleatória de 70 alunos (dos 350 que frequentam o 11.º ano nas duas escolas), tendo-se obtido um valor para a média amostral de $\overline{x} = 16{,}4$.

Símbolos a utilizar e seu significado:

	População	Amostra
Medida	**Parâmetro**	**Estatística**
Dimensão	N	n
Média	μ	$\hat{\mu} = \bar{x}$
Proporção	p	$\hat{p}$
Desvio padrão	σ	$\hat{\sigma} = s$

NOTA

É usual utilizar-se o acento circunflexo, ^, sobre a letra que representa um parâmetro para designar uma correspondente estimativa.

Parâmetro, θ, é um valor que caracteriza uma população.

Estatística, $\hat{\theta}$, é um valor que caracteriza a amostra.

NOTA

É usual representar o estimador por uma letra maiúscula (por exemplo, $\overline{X}$) e a correspondente estimativa pela respetiva letra minúscula (por exemplo, $\overline{x}$).

Nota: No caso da média, temos:

μ **Parâmetro** (populacional)

$\bar{X}$ **<u>Estimador</u>**(fórmula ou processo para estimar). O estimador é uma variável aleatória, pois os seus valores variam de amostra para amostra.

$\bar{x}$ **Estimativa**- resultado concreto de uma amostra particular.

Exemplo 2(230)-Estimativa pontual.

Exemplo 2

Suponhamos que pretendemos saber a proporção, p, de alunos de uma universidade que preferem uma determinada marca de chocolates. Para isso, selecionamos uma amostra aleatória de 250 alunos, dos quais 75 responderam afirmativamente.
Indique o parâmetro e a estatística.

Resolução

O parâmetro que se pretende estimar é a proporção de alunos de uma universidade que preferem determinada marca de chocolates.
A estatística é $\hat{p} = \frac{75}{250} = 30\%$, isto é, 30% dos alunos inquiridos preferem a tal marca de chocolates. Podemos acrescentar que, neste caso, 30% é uma estimativa pontual da proporção de alunos da universidade que preferem aquela marca de chocolates.

Uma estimativa pontual de um parâmetro θ é o valor numérico assumido pela estatística $\hat{\theta}$ para a amostra selecionada.

1.3- Distribuição de amostragem de uma estatística (ou estimador). (231)

Nota: No exemplo anterior, se recolhêssemos outras amostras com a mesma dimensão, muito provavelmente iríamos obter diferentes valores para a proporção amostral. No entanto, e por desconhecermos o valor exato do parâmetro, é impossível dizer qual dessas estimativas é a melhor.

Estimador é uma variável aleatória.

Distribuição de amostragem de um estimador $\hat{\theta}$ para estimar um determinado parâmetro θ é a distribuição de todos os valores que o estimador pode tomar na estimação desse parâmetro para todas as amostras possíveis de igual dimensão.

No tópico seguinte, vamos concretizar esta variabilidade de estimativas para o caso de um valor médio.

1.4- Estimação de um valor médio. (232).

Exemplo(Extra)

Para melhor compreender a distribuição de amostragem referida no tópico anterior, consideremos uma **população** só com 3 elementos: **{1; 2; 3}.**

Para esta população, vamos agora obter todas as **amostras** com 2 elementos**(n=2)**, com reposição: {1, 1}; {1, 2; {1, 3}; {2, 1}; {2, 2}; {2, 3}; {3, 1}; {3, 2}; {3, 3}. Ao todo são 9 amostras possíveis. Repare que foi 3×3=9, ou 3^2=9.

Questão: como relacionar o valor médio populacional com as médias amostrais?

Reparemos que a média populacional é dada por $\mu = \frac{1+2+3}{3} = \frac{6}{3} = 2$

Valor médio populacional: $\boldsymbol{\mu = 2}$

Calculemos as médias de cada uma das amostras, isto é, as médias amostrais:

	{1, 1}	{1, 2}	{1, 3}	{2, 1}	{2, 2}	{2, 3}	{3, 1}	{3, 2}	{3, 3}
$\bar{x}$	1	1.5	2	1.5	2	2.5	2	2.5	3

A média de todas as médias amostrais é dada por

$$E(\bar{X}) = \frac{1 + 1.5 + 2 + 1.5 + 2 + 2.5 + 2 + 2.5 + 3}{9} = \frac{18}{9} = 2$$

Repare que $\boldsymbol{E(\bar{X}) = \mu}$ a média das médias amostrais é igual à média populacional.

Questão: como relacionar o desvio padrão populacional, com o desvio padrão das médias amostrais?

O desvio padrão populacional é dado por

$$\sigma = \sqrt{\frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3}} = \sqrt{\frac{1+0+1}{3}} = \sqrt{\frac{2}{3}} \approx 0.816$$

Nota: pode obter este mesmo valor pelo desvio padrão populacional da calculadora.

Se tentar fazer o desvio padrão populacional de todas as médias amostrais, obtemos:

(pode usar a calculadora- indique uma coluna com os valores das médias amostrais obtidas acima: 1; 1.5; 2; …)

Obtém um desvio padrão populacional de $\sigma_{\bar{X}}$ ≈0.577, que não é igual a 0.816.

No entanto, se dividir 0.816 pela raiz quadrada de 2, obtém $\frac{0.816}{\sqrt{2}} \approx 0.577$.

Neste caso, escolhemos dividir por $\sqrt{2}$ porque 2 é a dimensão das amostras.

De um modo geral,

$$\boldsymbol{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

onde n é a dimensão da amostra.

Exemplo 1(232).

Exemplo 1

Numa empresa de transportes, cada um dos seis camionistas regista, semanalmente, o número de horas que passa ao volante do camião. O registo de uma determinada semana encontra-se na tabela seguinte:

Camionista	Número de horas
C_1	28
C_2	26
C_3	24
C_4	26
C_5	30
C_6	22

1.1 Quantas amostras diferentes de dimensão 2 (com dois camionistas) é possível definir?

1.2 Calcule a média do número de horas de cada uma das amostras obtidas na alínea anterior.

1.3 Elabore uma tabela com a distribuição de amostragem do estimador média.

1.4 Calcule a média das médias amostrais e o valor médio da população. Que conclusões pode tirar?

Resolução

1.1 Vamos considerar todas as amostras de dois camionistas, escolhidos aleatoriamente de entre os seis da população (com reposição).

Amostras					
C_1, C_1	C_1, C_2	C_1, C_3	C_1, C_4	C_1, C_5	C_1, C_6
C_2, C_1	C_2, C_2	C_2, C_3	C_2, C_4	C_2, C_5	C_2, C_6
C_3, C_1	C_3, C_2	C_3, C_3	C_3, C_4	C_3, C_5	C_3, C_6
C_4, C_1	C_4, C_2	C_4, C_3	C_4, C_4	C_4, C_5	C_4, C_6
C_5, C_1	C_5, C_2	C_5, C_3	C_5, C_4	C_5, C_5	C_5, C_6
C_6, C_1	C_6, C_2	C_6, C_3	C_6, C_4	C_6, C_5	C_6, C_6

São 36 amostras de dimensão 2.

Podia fazer $6 \times 6 = 36$

1.2 Na tabela seguinte encontram-se todas as amostras de dimensão 2 e as respetivas médias:

Amostras	$\bar{x}$	Amostras	$\bar{x}$	Amostras	$\bar{x}$
C_1, C_1	28	C_3, C_1	26	C_5, C_1	29
C_1, C_2	27	C_3, C_2	25	C_5, C_2	28
C_1, C_3	26	C_3, C_3	24	C_5, C_3	27
C_1, C_4	27	C_3, C_4	25	C_5, C_4	28
C_1, C_5	29	C_3, C_5	27	C_5, C_5	30
C_1, C_6	25	C_3, C_6	23	C_5, C_6	26
C_2, C_1	27	C_4, C_1	27	C_6, C_1	25
C_2, C_2	26	C_4, C_2	26	C_6, C_2	24
C_2, C_3	25	C_4, C_3	25	C_6, C_3	23
C_2, C_4	26	C_4, C_4	26	C_6, C_4	24
C_2, C_5	28	C_4, C_5	28	C_6, C_5	26
C_2, C_6	24	C_4, C_6	24	C_6, C_6	22

1.3 A distribuição de amostragem do estimador média consiste em atribuir a cada valor da média amostral a respetiva probabilidade. Assim, neste caso, a distribuição de amostragem para o estimador média é:

X	22	23	24	25	26	27	28	29	30
Probabilidade	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

1.4 Em geral, o valor do parâmetro a estimar, que neste caso é o número médio de horas que os camionistas passaram naquela semana ao volante (do camião), é um valor desconhecido, em virtude, sobretudo, da grande dimensão das populações. Neste exemplo, o tamanho da população permite-nos, sem qualquer dificuldade, determinar o seu valor médio, em horas:

$$\mu = \frac{28+26+24+26+30+22}{6} = 26$$

Para determinar a média da estatística $\bar{X}$, basta recorrer à sua distribuição de amostragem, ou seja:

$$E(\bar{X}) = 22 \times \frac{1}{36} + 23 \times \frac{2}{36} + 24 \times \frac{5}{36} + 25 \times \frac{6}{36} + 26 \times \frac{8}{36} +$$

$$+ 27 \times \frac{6}{36} + 28 \times \frac{5}{36} + 29 \times \frac{2}{36} + 30 \times \frac{1}{36} = 26$$

Mostra-se que a média da distribuição amostral do estimador média coincide com o parâmetro a estimar, isto é, com o valor médio μ . Escrevendo simbolicamente: $E(\overline{X}) = \mu_{\overline{X}}$

Questão importante: Como relacionar o desvio padrão populacional com o desvio padrão das médias amostrais?

No caso do desvio padrão populacional, obtemos o valor $\sigma \approx 2{,}582$ horas.

O desvio padrão da distribuição da amostragem é $\sigma_{\overline{X}} \approx 1{,}826$.

Como a dimensão da amostra é n=2,

$$\frac{\sigma}{\sqrt{n}} \approx \frac{2{,}582}{\sqrt{2}} \approx 1{,}826 \approx \sigma_{\overline{X}}$$

Conclusões:

$$\boldsymbol{E(\overline{X}) = \mu} \qquad \textbf{(235).}$$

Dizemos que $\bar{X}$ é um **estimador centrado** ou cêntrico ou não enviesado, pois o seu valor médio é igual ao parâmetro que pretendemos estimar. $E(\bar{X}) = \mu$.

O desvio padrão da distribuição de amostragem da média: (235)

$$\boldsymbol{\frac{\sigma}{\sqrt{n}} = \sigma_{\overline{X}},}$$ também se designa **<u>erro padrão</u>.**

Nota: O <u>aumento da dimensão da amostra</u>, a partir da qual é calculado o estimador, faz diminuir a variabilidade das estimativas. (236).

1.5- Importância do teorema do Limite Central.

Teorema do Limite Central.(241)

Seja X uma população com valor médio μ e desvio padrão σ, da qual se recolhem amostras de dimensão n.

Então, se n ≥ 30 , a distribuição de amostragem da média $\bar{X}$ pode ser aproximada a uma distribuição normal com valor médio μ e desvio padrão $\frac{\sigma}{\sqrt{n}}$,

isto é, para **n ≥ 30** $$\overline{\boldsymbol{X}} \sim \boldsymbol{N}\left(\boldsymbol{\mu}, \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right)$$

Nota: este teorema diz-nos que as médias amostrais, apesar de variarem de amostra para amostra, tendem a concentrar-se em torno do valor médio da população à medida que a dimensão das amostras aumenta, uma vez que o desvio-padrão $\frac{\sigma}{\sqrt{n}}$ diminui.

Exemplo 1(241).

Exemplo 1

Consideremos uma população de uma determinada espécie animal cuja esperança média de vida é 17,25 anos, com um desvio padrão de 3,05 anos.

Se tivermos uma amostra dessa população de dimensão 40, como será de esperar que seja a distribuição de amostragem da média?

Resolução

Uma vez que a dimensão da amostra é maior do que 30, o teorema do limite central garante que a distribuição de amostragem para a média pode ser representada por um modelo normal, com valor médio 17,25 anos e desvio padrão:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{3,05}{\sqrt{40}} \approx 0,482$$

☞***Exercício(267):13***, ***14***.

☞*Exercício(267):13*.

13. Da produção de pacotes de um tipo de bolacha, de uma determinada fábrica, sabe-se que o peso médio é 151,35 gramas, com um desvio padrão de 9,25 gramas. Recolhe-se, aleatoriamente, uma amostra de 40 pacotes de bolachas.

13.1 O valor médio da distribuição de amostragem da média é:

A. 150 gramas. **C.** 160,60 gramas.

B. 151,35 gramas. **D.** Não se sabe.

13.2 O desvio padrão da distribuição de amostragem da média é, aproximadamente:

A. 9,25 gramas. **C.** 1,463 gramas.

B. 0,231 gramas. **D.** 6,143 gramas.

Resolução:

13.1 B

13.2 C

$$\sigma_{\overline{X}} = \frac{9,25}{\sqrt{40}} \approx 1,463 \text{ g}$$

☞*Exercício(267):14*.

14. Considere uma população de cães de raça *Boxer* cujo peso médio é 17,5 kg e o desvio padrão é 2,75 kg. Para uma amostra de 35 destes animais, determine:

14.1 O valor médio da distribuição de amostragem da média.

14.2 O erro padrão da distribuição. Apresente o resultado final arredondado às milésimas.

Resolução:

14.1 $E\left(\overline{X}\right) = 17,5 \text{ kg}$

14.2 $\sigma_{\overline{X}} = \dfrac{2,75}{\sqrt{35}} \approx 0,465 \text{ kg}$

Exames: Inferência/distribuição de amostragem.

https://www.pedronoia.net/ExPrAss8Infer.htm

1.6- Intervalos de confiança para o valor médio.(245)

***[Vídeo 40]*.**

Intervalo de confiança de um parâmetro é um intervalo no qual temos alguma confiança que contenha o verdadeiro valor do parâmetro. A essa confiança chamamos nível ou grau de confiança.

Qualquer intervalo de confiança tem a forma:

]estatística – margem de erro; estatística + margem de erro[

Nota: a partir do teorema do Limite central e das propriedades da distribuição Normal, vamos obter intervalos de confiança.

Recordemos:

Para amostras com dimensão suficientemente grande (com pelo menos 30 elementos), a distribuição de amostragem da média pode ser aproximada pela distribuição normal.

Como

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right), \text{ então}$$

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

(podemos consultar a tabela da pág. 193)

Vamos obter um intervalo de confiança da forma:

$$\left]\bar{x} - z \times \frac{\sigma}{\sqrt{n}};\ \bar{x} + z \times \frac{\sigma}{\sqrt{n}}\right[$$

Comecemos por deduzir um intervalo com 90% de confiança. (pág. 246)

Procuremos o valor de z para o qual:

$$P\left(-z \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le z\right) = 0{,}9$$

$$\Leftrightarrow \phi(z) - \phi(-z) = 0{,}9 \Leftrightarrow$$
$$\Leftrightarrow \phi(z) - [1 - \phi(z)] = 0{,}9 \Leftrightarrow$$
$$\Leftrightarrow 2\phi(z) - 1 = 0{,}9 \Leftrightarrow$$
$$\Leftrightarrow \phi(z) = 0{,}95 \Leftrightarrow z = 1{,}645$$

(Consultamos a tabela da página 193)

O valor é z=1.645.

$$P\left(-z \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le z\right) = 0{,}9$$

Nota: A partir de, podemos isolar o μ, seguindo vários passos:

$$p\left(-z < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z\right) = p\left(-z \times \frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z \times \frac{\sigma}{\sqrt{n}}\right) =$$

$$= p\left(\bar{X} - z \times \frac{\sigma}{\sqrt{n}} < -\mu < \bar{X} + z \times \frac{\sigma}{\sqrt{n}}\right) =$$

$$= p\left(-\bar{X} - z \times \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + z \times \frac{\sigma}{\sqrt{n}}\right) =$$

$$= p\left(+\bar{X} + z \times \frac{\sigma}{\sqrt{n}} > \mu > +\bar{X} - z \times \frac{\sigma}{\sqrt{n}}\right) =$$

$$= p\left(+\bar{X} + z \times \frac{\sigma}{\sqrt{n}} > \mu > +\bar{X} - z \times \frac{\sigma}{\sqrt{n}}\right) =$$

Mudando os sinais e e reordenando:

$$\boldsymbol{= p\left(\bar{X} - z \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z \times \frac{\sigma}{\sqrt{n}}\right)}$$

Dizer que esta probabilidade é 0.9, com z= 1.645, é equivalente a afirmar que a probabilidade de μ estar no intervalo:

$$\left]\boldsymbol{\bar{X} - 1.645} \times \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}};\ \boldsymbol{\bar{X} + Z} \times \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right[\text{ é 90\%.}$$

Nota: do mesmo modo podemos obter intervalos com outras probabilidades.

Intervalo de confiança:

$$\boldsymbol{\left]\bar{X} - Z\frac{\sigma}{\sqrt{n}};\ \bar{X} + Z\frac{\sigma}{\sqrt{n}}\right[}$$

$\boldsymbol{\bar{X}}$ é a <u>média amostral</u> $\boldsymbol{\sigma}$ desvio padrão

n é o número de elementos da amostra

Z é o valor relacionado com a confiança. Os valores usuais para Z são os que constam da tabela seguinte:

Nível de confiança	90%	95%	99%
Valor de Z	1,645	1,960	2,576

A diferença entre os dois extremos do intervalo: $\left(\bar{X} + Z\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - Z\frac{\sigma}{\sqrt{n}}\right)$ é a

amplitude do intervalo(A).

$\varepsilon = Z\frac{\sigma}{\sqrt{n}}$ representa a **margem de erro** do intervalo.

Reparemos que:

A **amplitude** é igual ao dobro da margem de erro. **A = 2×ε**

A **margem de erro** é igual a metade da amplitude **ε = A/2**

Nota: Quando afirmamos que o intervalo tem 95% confiança, significa que, para 100 amostras diferentes, da mesma dimensão, selecionadas da mesma população, esperamos que 95 dos correspondentes intervalos obtidos, contenham o valor médio da população a estudar.

Exemplo 1 (246).

Exemplo 1

O responsável pelo ginásio «Gordinhos» está interessado em estudar o parâmetro «peso médio dos seus utentes». Para isso, selecionou uma amostra aleatória de 80 utentes e concluiu que a média era 72 kg e o desvio padrão 11 kg.

Determine um intervalo de 90% de confiança para o peso médio dos utentes do «Gordinhos».

Resolução

Tal como se viu anteriormente, um intervalo de 90% de confiança para o peso médio dos utentes é da forma $\left]\bar{x} - 1{,}645 \times \frac{\sigma}{\sqrt{n}};\ \bar{x} + 1{,}645 \times \frac{\sigma}{\sqrt{n}}\right[$. Neste caso, será $\left]72 - 1{,}645 \times \frac{11}{\sqrt{80}};\ 72 + 1{,}645 \times \frac{11}{\sqrt{80}}\right[=]69{,}97;\ 74{,}02[$.

Nota muito importante:

Usamos o símbolo **σ** para o desvio padrão populacional, mas quando se trata de uma amostra devemos usar o símbolo "**s**".

No exemplo anterior, é mais correto usar: s=11 e escrever o intervalo

$$\left]\bar{X} - Z\frac{s}{\sqrt{n}};\ \bar{X} + Z\frac{s}{\sqrt{n}}\right[.$$

Atividade 1(247).

ATIVIDADE 1

Numa fábrica foi recolhida uma amostra aleatória de 50 peças cilíndricas para avaliação do seu diâmetro. Concluiu-se que tinham diâmetro médio de 120 mm e desvio padrão 10 mm. Determine o intervalo de 90% de confiança para μ .

Resolução:

Atividade 1 (pág. 247)

Como se trata de uma amostra, consideramos s=10

$\mu = 120$ mm

Um intervalo de confiança de 90% para o parâmetro μ é de forma:

$$\left]\overline{x} - 1,645 \times \frac{s}{\sqrt{n}};\ \overline{x} + 1,645 \times \frac{s}{\sqrt{n}}\right[= \left]120 - 1,645 \times \frac{10}{\sqrt{50}};\ 120 + 1,645 \times \frac{10}{\sqrt{50}}\right[= \left]117,67;\ 122,33\right[$$

Atividade 2(247)

ATIVIDADE 2

O que significa um intervalo de 90% de confiança para o parâmetro μ ?

Resolução:

Significa que se recolhermos muitas amostras de dimensão *n*, calcularmos as médias e os desvios-padrão dessas amostras e construirmos os intervalos de confiança respetivos, cerca de 90% desses intervalos conterão o valor médio μ, enquanto os restantes 10% não conterão o parâmetro μ. Não temos a certeza de que um determinado intervalo contenha o parâmetro desconhecido, mas temos 90% de confiança que o intervalo que calculámos contenha o valor do parâmetro.

Exemplo 2(248).

Exemplo 2

O Marcelino pretende estimar o parâmetro média das idades na população dos professores da escola «Estatística». Para isso, selecionou uma amostra aleatória de dimensão 50 e registou as idades dos professores selecionados. Verificou que a média e o desvio padrão eram, respetivamente, 41 e 10.

Determine um intervalo de confiança de 95% para a idade média da população dos professores da escola.

Resolução

Nota: como agora pretendemos a confiança de 95%, o valor de z é 1.96

Um intervalo de confiança de 95% de confiança para a idade média da população é, como vimos anteriormente,

$$\left]41 - 1{,}96 \times \frac{10}{\sqrt{50}}; 41 + 1{,}96 \times \frac{10}{\sqrt{50}}\right[=]38{,}22; 43{,}77[$$

Nota: para o intervalo do exemplo 2, qual é o valor amplitude? E da margem de erro?

A= 43.77-38.22= 5.55   **ε**=A/2 = 5.55/2 = 2.775

Exemplo (Extra).

Construa um intervalo semelhante ao que vimos no exemplo 2, mas utilize n=500 em vez de n=50. O que acontece à margem de erro? Aumenta ou diminui?

Resolução: *(…)*

]40.12; 41.88[A= 1.76 ε = 0.88   Resposta: a margem de erro diminui

Atividade 3(248)

ATIVIDADE 3

Mediu-se o pulso de 1000 pessoas saudáveis, em repouso, e verificou-se que a média das batidas por minuto era de 75, com um desvio padrão de 10 batidas por minuto.

Determine um intervalo de confiança de 95% para a pulsação média, em repouso, de pessoas saudáveis, com base nestes dados.

Resolução:

$n = 1000 \qquad \bar{x} = 75 \qquad s = 10$

Um intervalo de confiança para μ com 95% de confiança é dado por:

$$\left]75 - 1{,}96 \times \frac{10}{\sqrt{1000}};\ 75 + 1{,}96 \times \frac{10}{\sqrt{1000}}\right[= \left]74{,}38;\ 75{,}62\right[$$

Exemplo 3(249).

(usar os dados do exemplo 2, exceto a confiança de 99% em vez de 95%)

Exemplo 3

Para a mesma situação do exemplo anterior, determine um intervalo de confiança de 99% para μ .

Resolução

Substituindo os valores conhecidos no intervalo anterior, vem:

$$\left]41 - 2{,}576 \times \frac{10}{\sqrt{50}};\ 41 + 2{,}576 \times \frac{10}{\sqrt{50}}\right[= \left]37{,}36;\ 44{,}64\right[$$

Nota: para o intervalo do exemplo 3, qual é o valor amplitude? E da margem de erro?

A= 44.64-37.36= 7.28 ε=A/2 = 7.28/2 = 3.64.

Se comparar com o exemplo 2, pode constatar que a confiança passou de 95% para 99% e a margem de erro aumentou. Passou de 2.775 para 3.64.

Atividade 4(249)

ATIVIDADE 4

Mediu-se os QI de 200 crianças da cidade X, numa determinada faixa etária, e verificou-se que o QI médio era 110 e o desvio padrão 15.

4.1 Determine um intervalo de confiança de 99% para o QI médio populacional das crianças da faixa etária considerada.

4.2 Foi necessário, neste caso, assumir que os QI seguem uma distribuição aproximadamente normal? Justifique.

Resolução:

4.1 Um intervalo de confiança para μ com 99% de confiança é dado por:

$$\left]110-2,576\times\frac{15}{\sqrt{200}};\ 110+2,576\times\frac{15}{\sqrt{200}}\right[=\left]107,27;\ 112,73\right[$$

Nota(para 4.2): *Esta questão é teórica e apenas alerta para as condições de utilização dos intervalos de confiança dados. Os dados têm que estar de acordo com o Teorema do Limite Central (TLC), isto é, n≥ 30 ou então, se a dimensão for menor que 30, os dados precisam ter uma distribuição, pelo menos aproximadamente Normal.*

Resolução:

4.2 Não, porque a amostra tem dimensão superior a 30.

, Nota: Quando aumentamos o grau de confiança, a amplitude do intervalo aumenta, e consequentemente a informação fica menos precisa.

Nota: Quando aumentamos o número de elementos da amostra, a margem de erro $Z\frac{\sigma}{\sqrt{n}}$ diminui e a amplitude do intervalo também diminuir. Assim, melhoramos a precisão da nossa estimação.

Sugestão: Ir resolvendo os Exercícios (266): 16 a 26.

Exames: Inferência/ Intervalo para o valor médio.

https://www.pedronoia.net/ExPrAss8Infer.htm

(Parte 2→)

1.7- Estimativa pontual de proporção.(250)(…)