

# Introdução à Inferência Estatística

Eduardo Esteves

23 de Outubro de 2009

Departamento de Engenharia Alimentar, Instituto Superior de Engenharia da Universidade do Algarve, *Campus* da Penha, 8005-139 Faro, Portugal. E-mail: eesteves@ualg.pt URL: <http://w3.ualg.pt/~eesteves>.

## Resumo

Este "artigo" constitui o terceiro (grande) "capítulo" das matérias leccionadas no âmbito da disciplina de Estatística Aplicada do Curso de Engenharia Alimentar [*cf.* Esteves (2009a,b)<sup>1</sup>]. Foi preparado de forma a explicar (ainda que de forma sucinta e relativamente informal) e exercitar as metodologias que permitem, por um lado, "generalizar" os resultados da amostra (intervalos de confiança) e, por outro lado, testar hipóteses acerca dos resultados obtidos (testes de hipóteses).

## Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Amostragem</b>	<b>2</b>
2.1	Amostragem aleatória . . . . .	2
2.2	Distribuição da média na amostragem . . . . .	3
2.3	Teoria sobre distribuição de probabilidades da média na amostragem . . . . .	3
<b>3</b>	<b>Estimação da média da população</b>	<b>4</b>
3.1	Estimação pontual . . . . .	4
3.2	Estimação por intervalos de confiança . . . . .	5
<b>4</b>	<b>Teste (ou prova) de hipóteses</b>	<b>6</b>
4.1	Definição das hipóteses nula e alternativa . . . . .	6
4.2	Erros de inferência . . . . .	7
4.3	Definição de regras de decisão . . . . .	7
4.4	Testes de hipóteses bilaterais para a média a partir de uma amostra . . . . .	7
4.5	Testes de hipóteses unilaterais para a média a partir de uma amostra . . . . .	10
4.6	Testes de hipóteses para uma amostra através de "estatísticas de teste" . . . . .	11
4.7	Testes de hipóteses relativos às médias de duas amostras . . . . .	11
<b>5</b>	<b>Exercícios</b>	<b>11</b>

## 1 Introdução

Depois da colheita de informação (amostragem), da organização e resumo desses dados de forma a apresentá-los correctamente (estatística descritiva), é geralmente intenção do engenheiro(a)/investigador generalizar os resultados para toda a população. Um dos objectivos principais da análise estatística é derivar (ou inferir) conclusões válidas acerca de uma população através do exame de amostra(s) dessa população - INFERÊNCIA ESTATÍSTICA (Fig. 1). Com a inferência estatística pretende-se responder a dois tipos de questões:

<sup>1</sup>Esteves, E. (2009a) *Introdução à estatística aplicada e estatística descritiva*. Instituto Superior de Engenharia da Universidade do Algarve, Faro, 14 p.[disponível em <http://w3.ualg.pt/~eesteves>] e Esteves, E. (2009b) *Introdução à teoria das probabilidades*. Instituto Superior de Engenharia da Universidade do Algarve, Faro, 22 p. [disponível em <http://w3.ualg.pt/~eesteves>]

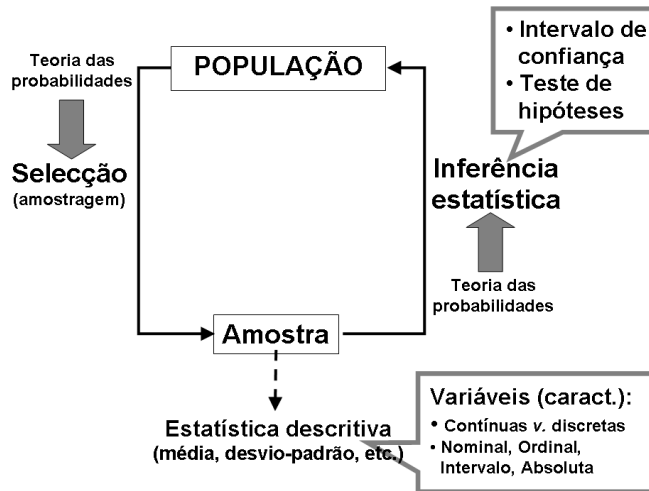


Figura 1: Esquema da relação entre os vários conceitos básicos em Estatística.

- 1) Qual é o valor de um certo parâmetro da população? (ESTIMAÇÃO DE PARÂMETROS, ver a seguir).
- 2) Pode considerar-se que um dado parâmetro da população tem determinado valor? (PROVA, OU TESTE, DE HIPÓTESES, ver mais adiante).

O primeiro caso designa-se, por vezes, teoria da estimação, enquanto o segundo caso é, muitas vezes, referido com dizendo respeito à teoria da decisão.

Exemplos de 1) Qual é a altura média dos alunos da EST? Qual é o peso líquido de iogurte nas embalagens produzidas em determinada fábrica?

Exemplos de 2) Será que a média das alturas dos alunos da EST é 183 cm? Será que as embalagens de iogurte duma dada marca têm um peso líquido de 125 mL?

Nesta versão do artigo, consideraremos apenas os casos que dizem respeito à média (geralmente o parâmetro de maior interesse para a maioria dos estudos) e que envolvem uma única amostra (os casos mais simples). Outros parâmetros (*e.g.* variância ou proporção) e casos com duas ou mais amostras serão incluídos em versões posteriores deste artigo e estudados no âmbito de outras disciplinas do curso de Engenharia Alimentar.

## 2 Amostragem

Define-se AMOSTRAGEM como o conjunto de todas as amostras, de um certo tamanho<sup>2</sup>  $n$ , possíveis de seleccionar com um determinado critério de uma população. Também se utiliza o termo amostragem como sinónimo do critério com que as amostras são seleccionadas ou, ainda, para indicar o procedimento prático de recolha dos dados. O seu significado fica, assim, dependente do contexto em que é utilizado.

### 2.1 Amostragem aleatória

O tipo de amostragem mais “básico” é a AMOSTRAGEM ALEATÓRIA SIMPLES (a.a.s.)<sup>3</sup> que dá igual probabilidade de ser seleccionado a qualquer um dos resultados (ou das amostras) possíveis. Poderemos amostrar aleatoriamente, isto é “ao acaso”, populações finitas ou infinitas (cujo significado é óbvio!).

**Populações finitas** Quando se pretendem estudar populações finitas, a amostragem aleatória simples pode ser feita:

- a) COM REPOSIÇÃO, se cada elemento da população pode ser escolhido mais de uma vez, isto é, após recolher informação (medir ou categorizar a variável em estudo) numa amostra, o elemento é devolvido à população.

<sup>2</sup>A decisão sobre a dimensão da amostra  $n$  também constitui um “problema” em si mesmo embora (por agora) não seja abordado aqui!

<sup>3</sup>Outro dos critérios possíveis é, por exemplo, amostragem aleatória estratificada.

Na amostragem a.s. com reposição, cada elemento de uma população com  $N$  elementos tem sempre uma probabilidade  $p = 1/N$  de ser seleccionado. Note-se que a amostragem com reposição duma população finita pode considerar-se teoricamente como infinita.

- b) SEM REPOSIÇÃO, se cada elemento da população não pode ser escolhido mais de uma vez. Nestes casos, o procedimento de amostragem é o seguinte: seleccionar o primeiro elemento da amostra dando a cada elemento da população igual probabilidade de ser escolhido, ou seja,  $p = 1/N$ ; seleccionar o segundo elemento da amostra, considerando que os  $N - 1$  elementos restantes possuem igual probabilidade de serem escolhidos, ou seja, a  $p = 1/(N - 1)$ ; repetir o processo até os  $N$  elementos da população serem seleccionados.

Em qualquer dos casos, a selecção dos elementos da amostra pode ser auxiliada pela utilização de uma tabela de números aleatórios (Tabela C, em anexo). Para cada posição na tabela, os dígitos 0, 1, ..., 9 têm igual probabilidade de ocorrerem e as várias posições na tabela são independentes.

**Populações infinitas** Nas populações infinitas, não é possível estabelecer um mecanismo semelhante ao descrito anteriormente para populações finitas. Assim, terá de se pressupor (ou verificar) que os  $n$  resultados possíveis  $x_1, x_2, \dots, x_n$  gerados por um processo (ou prova aleatória, ou experiência), constituem uma amostra aleatória uma vez que:

1. Os valores pertencem à mesma distribuição de probabilidades; e
2. Os valores são estatisticamente independentes.

Por exemplo, numa determinada linha de produção, a funcionar "continuamente", cada elemento seleccionado tem de ter a mesma probabilidade de ser defeituoso (1ª condição). O facto de um elemento seleccionado dessa linha de produção ser defeituoso não depende, nem implica, que os outros o sejam (2ª condição).

## 2.2 Distribuição da média na amostragem

As questões relacionadas com a amostragem têm muita importância, pois garantem a validade das conclusões que se pretendem elaborar em estatística. De facto, na maioria das "situações reais" não é possível estudar toda a população estatística, pelo que é necessário seleccionar amostra(s) e analisar características dessas amostra(s) de modo a "dizer coisas" acerca da população. De entre as "características" da amostra(s), a média da amostra é de utilização generalizada e constitui um dado importante na "prática estatística". Assim, é importante estudar a distribuição (de probabilidades) da média na amostragem.

Observe-se com atenção a Fig. 2. A partir de uma população estatística com  $N=8042$  elementos, seleccionaram-se aleatoriamente amostras de  $n=3$ ,  $n=10$  e  $n=100$  elementos e para cada amostra calculou-se a média amostral  $\bar{x}$ . Os polígonos de frequências absolutas dos resultados obtidos (a média para cada amostra seleccionada) indicam que:

- a) O valor médio da distribuição dos valores de  $\bar{x}$  das amostras aleatórias é igual à média da população  $\mu$ ;
- b) O desvio-padrão da distribuição dos valores de  $\bar{x}$  decresce com o aumento do tamanho amostra;
- c) A distribuição dos valores de  $\bar{x}$  vai ficando cada vez mais simétrica à medida que o tamanho da amostra aumenta; este último resultado é conhecido como TEOREMA DO LIMITE CENTRAL (de que falaremos mais adiante).

## 2.3 Teoria sobre distribuição de probabilidades da média na amostragem

Se a selecção de uma amostra for aleatória, então a média da amostra  $\bar{x}$  é resultado duma variável aleatória à qual está associada uma distribuição de probabilidades. Pode-se, portanto, recorrer ao valor médio e à variância para descrever resumidamente aquela distribuição de probabilidades.

Assim, qualquer que seja o tamanho da amostra  $n$ , o valor médio da distribuição de probabilidades de  $\bar{x}$  é dado por:

$$E\{\bar{X}\} = \mu$$

e a variância é igual a:

$$V\{\bar{X}\} = \frac{\sigma^2}{n}$$

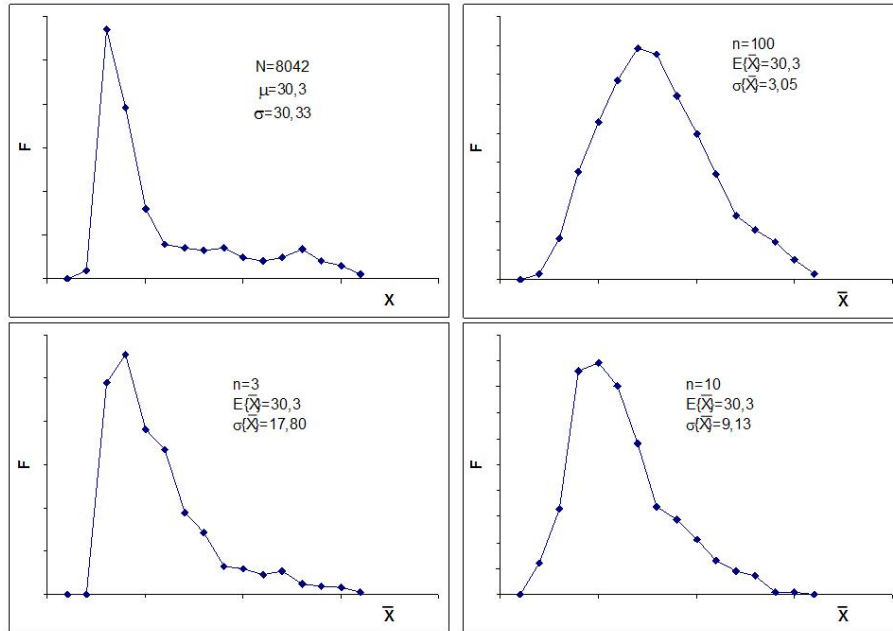


Figura 2: Distribuição na amostragem de  $\bar{x}$  nas amostras possíveis de tamanho  $n=3$ ,  $n=10$  e  $n=100$  que se podem obter duma população "original" com  $N=8042$  elementos.

O desvio-padrão das probabilidades de  $\bar{x}$ , que se designa erro-padrão da média<sup>4</sup> é igual a:

$$\sigma\{\bar{X}\} = \sqrt{V\{\bar{X}\}} = \frac{\sigma}{\sqrt{n}}$$

Por maioria de razão, para um certo tamanho de amostra  $n$ , quanto maior é a variabilidade da população maior é a variabilidade da distribuição das médias na amostragem.

**Teorema do limite central** Para a maioria das populações estatísticas, a distribuição de  $\bar{x}$  na amostragem (isto é, o conjunto das médias das amostras seleccionadas aleatoriamente) é aproximadamente normal quando o tamanho da amostra  $n$  é suficientemente grande, ou seja, a distribuição de  $\bar{x}$  na amostragem é assintoticamente normal. Quando a distribuição da variável aleatória  $X$  na população é normal, a distribuição de  $\bar{x}$  na amostragem é sempre normal, qualquer que seja o tamanho da amostra  $n$ .

### 3 Estimação da média da população

Um problema importante da inferência estatística é a estimação de parâmetros, tais como a média ou variância populacionais, deduzidos da estatística amostral correspondente, isto é, da média ou da variância amostrais (Teoria da Estimação). Como se referiu anteriormente, vamos abordar aqui o caso da média.

Qual é a altura média dos alunos da EST? Qual é o peso líquido de iogurte nas embalagens produzidas em determinada fábrica?

#### 3.1 Estimação pontual

Quando uma característica da população é estimada por um simples valor, este é designado por ESTIMATIVA PONTUAL. Dito de outro modo: o ESTIMADOR é uma variável aleatória usada para estimar uma característica da população. O valor numérico do estimador designa-se por ESTIMATIVA.

<sup>4</sup>Usualmente o desvio-padrão do estimador de um parâmetro, *i.e.* duma "estatística", denomina-se erro-padrão.

Por exemplo, dizer que a média de alturas dos alunos do curso de Engenharia Alimentar é 170 cm é recorrer a um estimador (a média) e avançar com uma estimativa pontual (da média; no caso, 170 cm, que se obteve numa amostra aleatória "representativa").

A média da amostra é um estimador sem-vício (ou não-enviesado) da média da população  $\mu$ , porque o valor médio da distribuição de probabilidades de  $\bar{x}$ ,  $E\{\bar{X}\}$ , é igual a  $\mu$ . Prova-se, ainda, que a variância da amostra  $s^2$  (calculada com denominador igual a  $n-1$ ) também é um estimador sem-vício da variância da população  $\sigma^2$ , isto é,  $E\{s^2\} = \sigma^2$ , se a amostra for obtida de uma população infinita ou, então, de uma população finita mas com reposição. É usual designar os estimadores sem-vício por ESTATÍSTICAS.

### 3.2 Estimação por intervalos de confiança

Nos casos em que a distribuição de  $\bar{x}$  na amostragem é normal (*cf.* teorema do limite central), pode-se esperar que os intervalos  $\bar{x} \pm \sigma$ ,  $\bar{x} \pm 2\sigma$  e  $\bar{x} \pm 3\sigma$  incluam a média da população  $\mu$  em aproximadamente 68,3%, 95,5% e 99,7% das vezes, respectivamente (*cf.* Figura 12 de Esteves (2009b) *Introdução à Teoria das Probabilidades*). Por esse motivo, estes intervalos são denominados INTERVALOS DE CONFIANÇA de 68,3%, 95,5% e 99,7% (para avaliação) de  $\mu$ . Os números extremos desses intervalos são denominados LIMITES DE CONFIANÇA. De modo semelhante,  $\bar{x} \pm 1,96\sigma$  e  $\bar{x} \pm 2,58\sigma$  são limites de confiança de 95% e 99% de  $\mu$ . A percentagem de confiança  $c$  é frequentemente denominada NÍVEL DE CONFIANÇA e a probabilidade complementar  $\alpha$  (que se lê "alfa") é designada por NÍVEL DE SIGNIFICÂNCIA ou "risco"<sup>5</sup> sendo que  $c = 1 - \alpha$ . Os números 1,96 e 2,58 nos limites de confiança referidos acima são denominados COEFICIENTES DE CONFIANÇA ou VALORES CRÍTICOS. Quando a distribuição da média na amostragem é normal, é possível calcular a probabilidade da média populacional  $\mu$  ocorrer entre dois valores da distribuição, recorrendo por exemplo à DISTRIBUIÇÃO NORMAL REDUZIDA,  $Z$ .

Conhecendo-se a variância populacional  $\sigma^2$  e aplicando a transformação  $Z$  à média na amostragem, ou seja,

$$Z = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

teremos, então, que a probabilidade da média da população  $\mu$  assumir valores entre  $-z$  e  $+z$  será genericamente:

$$P\left\{-z < \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < +z\right\} = c$$

ou seja, o mesmo que

$$P\left\{\bar{x} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\frac{\sigma}{\sqrt{n}}\right\} = c$$

em que  $c = 1 - \alpha$  indica o nível de confiança e  $z$  é o percentil de ordem  $(1 - \frac{\alpha}{2})$  de  $Z$ ,  $z_{[1-\frac{\alpha}{2}]}$ .

Esta última expressão de probabilidade (ou afirmação probabilística) deve ser interpretada cuidadosamente! Não significa que a probabilidade do parâmetro  $\mu$  estar incluído no intervalo especificado é  $c$ ! A média populacional  $\mu$  é um parâmetro (e não uma v.a.), e portanto, está ou não dentro daquele intervalo. A expressão deve ser interpretada do seguinte modo:

**$c$  é a probabilidade do intervalo especificado conter  $\mu$ .**

O intervalo  $[\bar{x} - z(\sigma/\sqrt{n}); \bar{x} + z(\sigma/\sqrt{n})]$  é designado por INTERVALO DE CONFIANÇA de  $\mu$ . O valor de  $Z$  é arbitrário, isto é, pode ser escolhido (da Tabela A, em anexo) de modo a que a probabilidade (ou confiança) pretendida seja igual a  $1 - \alpha$ . Assim, para cada um dos limites de confiança,  $Z$  ficará definido para  $1 - \alpha/2$  (ou seja, metade de  $\alpha$  em cada extremo da distribuição) (*cf.* Figuras 18 e 19 de Esteves (2009b) *Introdução à Teoria das Probabilidades*). Duma forma geral, a probabilidade do intervalo de  $(1 - \alpha)100\%$  de confiança de  $\mu$  com nível de confiança de  $c = 1 - \alpha$ , é dada por:

$$P\left\{\bar{x} - z_{[1-\frac{\alpha}{2}]} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{[1-\frac{\alpha}{2}]} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

Suponha-se, por exemplo, que a v.a.  $X$  representa a duração de vida de cápsulas metálicas de garrafas de refrigerante. Admita-se que a duração de vida média das cápsulas é  $\mu = 5,00$  anos, com desvio-padrão  $\sigma = 0,40$  anos. O intervalo de confiança de 95% de  $\mu$  numa amostra 10 cápsulas será dado por  $P\{5,0 - 1,96\frac{0,40}{\sqrt{10}} < \mu < 5,0 + 1,96\frac{0,40}{\sqrt{10}}\} = 0,95$  em que  $z_{[0,975]} = 1,96$  se obtém da tabela de  $Z$  (Tabela A, anexa). Logo, o intervalo será [4,75 a; 5,25 a].

<sup>5</sup>É vulgar usar proporções, em vez de percentagens, nas indicações e cálculos. Logo 95% de confiança indica-se por  $c=0,95$  e um nível de significância de 5% designa-se  $\alpha = 0,05$ .

E se a variância da população  $\sigma^2$  não é conhecida, como aliás acontece em muitos casos? Relembre-se que o valor médio de  $s^2$  é igual à variância populacional  $\sigma^2$ , quando  $s$  é calculado com  $n - 1$  graus de liberdade e  $X \cap N$ . Então pode-se utilizar  $s/\sqrt{n}$  como estimador sem-vício de  $\sigma/\sqrt{n}$ . Recorde-se, ainda, a utilidade da transformação  $t$  para  $\nu = n - 1$  graus de liberdade, quando não se conhece a variância  $\sigma^2$ , ou seja:

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

Assim, a probabilidade de determinado intervalo de confiança da média da população  $\mu$  é dada por:

$$P\left\{\bar{x} - t_{[\nu, 1-\frac{\alpha}{2}]} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{[\nu, 1-\alpha]} \frac{s}{\sqrt{n}}\right\} = 1 - \alpha$$

De modo similar a  $z$ , o valor de  $t$  é obtido da tabela apropriada para  $\nu = n - 1$  graus de liberdade e probabilidade igual a  $(1 - \alpha/2)$  (ver Tabela B, anexa).

São realizadas dez medições da resistência de determinada embalagem para produtos alimentares (variável aleatória  $X$  com distribuição normal). Suponha-se que  $\bar{x} = 10,48$  psi e que  $s = 1,36$  psi. O intervalo de confiança da média  $\mu$  com nível de confiança de 90% é dado por  $P\{10,48 - 1,83 \frac{1,36}{\sqrt{10}} < \mu < 10,48 + 1,83 \frac{1,36}{\sqrt{10}}\} = 0,90$  em que  $t_{[9;0,95]} = 1,83$  se obtém da Tabela de  $t$  (Tabela B, anexa). O intervalo de confiança será [9,693 psi; 11,267 psi].

Os intervalos de confiança permitem fazer afirmações probabilísticas acerca de parâmetros da população com base em estatísticas amostrais. Com uma dada confiança, o intervalo de confiança calculado deverá conter "o verdadeiro valor" do parâmetro populacional em questão, neste caso a média  $\mu$ .

## 4 Teste (ou prova) de hipóteses

A inferência estatística também pretende responder a questões do tipo: a) *Pode considerar-se que um dado parâmetro da população tem determinado valor?* b) *Será que duas amostras têm igual média*<sup>6</sup>. A resposta questões deste tipo diz respeito à Teoria da Decisão.

Por exemplo, para a) Será que a média das alturas dos alunos da EST é 183 cm? Será que, em média, as embalagens de iogurte da marca  $W$  têm um peso líquido de 125 mL? No caso de b) Será que os alunos de Eng<sup>a</sup> Alimentar do ISE são, em média, mais altos do que os alunos de Eng<sup>a</sup> Mecânica? Verifica-se, ou não, que o nível de um metabolito é igual após tratamento dum produto com duas enzimas diferentes?

O método científico (muito simplificado) pressupõe a definição prévia de hipóteses, a sua "experimentação" (com base numa amostra ou em várias) e a elaboração de conclusões. Em estatística o processo não é diferente.

"Classicamente", os testes de hipóteses podem ser resumidos em fases: 1) Definição das hipóteses nula e alternativa; 2) Definição do nível de significância,  $\alpha$ ; 3) Seleção e obtenção de informação, de estatísticas, a partir da(s) amostra(s) (nestes apontamentos, consideraremos unicamente a média da amostra, mas poderia analisar-se a variância, a proporção, etc.); 4) Determinação de valores-limite ou duma estatística de teste; e 5) Utilização de "regras" de decisão e elaboração de conclusões.

### 4.1 Definição das hipóteses nula e alternativa

O primeiro passo de um teste de hipóteses é definir duas HIPÓTESES (OU CONCLUSÕES) ALTERNATIVAS, mutuamente exclusivas, que se designam (geralmente) por  $H_0$  e  $H_1$ . Se a HIPÓTESE NULA  $H_0$  é verdadeira, então a HIPÓTESE ALTERNATIVA  $H_1$  será falsa e vice-versa (alguns autores utilizam a notação  $H_A$  para se referirem à hipótese alternativa). Saliente-se que, classicamente, ambas as hipóteses são definidas *a priori*. Esta será "a regra" seguida neste texto.

No caso de hipóteses bilaterais, a hipótese  $H_0$  contém o "valor padrão"  $\mu_0$  em relação ao qual se faz o teste (ver a seguir) ou refere-se à situação em que as médias de duas populações são iguais (ver mais diante). A designação de hipótese nula está relacionada com o conceito de "não diferente de" um valor-padrão  $\mu_0$  (*i.e.*  $H_0 : \mu = \mu_0$  ou  $H_1 : \mu \neq \mu_0$ ) ou então de "não existir" diferença entre (a média) de duas amostras,  $\mu_A$  e  $\mu_B$  (ou seja,  $H_0 : \mu_A = \mu_B$  ou  $H_1 : \mu_A \neq \mu_B$ ).

<sup>6</sup>Para já não se discutem estes casos!

Tabela 1: Resumo dos erros tipo I e II possíveis de acontecer em inferência estatística.

Teste de hipóteses	Alternativa verdadeira	
	$H_0$	$H_1$
$H_0$	Conclusão correcta	Erro de tipo II ( $\beta$ )
$H_1$	Erro de tipo I ( $\alpha$ )	Conclusão correcta

Quanto às hipóteses unilaterais, é "prática comum" definir na hipótese alternativa  $H_1$  o resultado que se "deseja obter". Por exemplo, para testar se, em média, a pressão em determinado equipamento é superior àquela especificada, as hipóteses nula e alternativa seriam  $H_0 : \mu \leq \mu_0$  ou  $H_1 : \mu > \mu_0$ , respectivamente. Por outro lado, quando se pretende testar se a classificação média do "aspecto"<sup>7</sup> de um produto alimentar é menor do que aquela atribuída a um produto concorrente, definem-se as seguintes hipóteses:  $H_0 : \mu \geq \mu_0$  ou  $H_1 : \mu < \mu_0$ .

## 4.2 Erros de inferência

É aconselhável "fazer aqui um parêntesis" no texto para se abordarem os "erros estatísticos". Não se está isento de errar quando se infere estatisticamente, pois existe sempre risco de elaborar uma conclusão incorrecta quando recorremos a medida(s) da amostra para decidir estatisticamente em favor duma das hipóteses alternativas (relativas a um parâmetro da população).

De facto, teremos de considerar que uma hipótese nula verdadeira será ocasionalmente rejeitada. Este tipo de erro (conhecido como erro do tipo I) será cometido com probabilidade/frequência  $\alpha$  (em alguns contextos  $\alpha$  é designado por risco), isto é, se  $H_0$  é na realidade uma afirmação verdadeira acerca da população estatística, no caso de  $\alpha = 0,05$ , concluir-se-á incorrectamente que é falsa em 5% das afirmações. Se, pelo contrário,  $H_0$  é, de facto, falsa, um teste de hipóteses não detectará este resultado algumas vezes e dele se derivará uma conclusão errada, não-rejeitando  $H_0$ , o que se designa por erro de tipo II. A probabilidade de cometer este tipo de erro é  $\beta$  (que se lê "beta" e por vezes se designa risco  $\beta$ )<sup>8</sup>. A Tab. 1 resume estes conceitos.

Não podemos evitar completamente estes erros, mas podemos tentar minimizar a probabilidade de os cometer. Os erros de tipo I poderão ser reduzidos, directamente, se diminuirmos  $\alpha$  (o nível de significância) e por conseguinte aumentando a confiança. No entanto, sabe-se que, para determinado tamanho de amostra  $n$ , o valor de  $\alpha$  está inversamente relacionado com o valor de  $\beta$ , ou seja, à menor probabilidade de cometer erros de tipo I está associada a maior probabilidade de cometer erros de tipo II. O único modo de reduzir ambos os erros é aumentar o tamanho da amostra  $n$ .

## 4.3 Definição de regras de decisão

Depois de definidas as hipóteses nula e alternativa, é possível calcular o(s) limite(s) de confiança em torno de  $\mu_0$  (caso de uma amostra) ou da diferença das médias (no caso de duas amostras). Esse(s) valor(es) limitam "áreas" ou "regiões", que se designam "regiões de aceitação" e "de rejeição", e que permitem decidir qual das hipóteses concluir em face da estatística em estudo, neste caso a média da amostra  $\bar{x}$  (Fig. 3). Na prática, se estiver  $\bar{x}$  entre o(s) limite(s) calculado(s) decide-se em favor da hipótese nula (ver adiante). Caso contrário, rejeita-se a hipótese nula. Alguns autores referem "região de não-rejeição" em vez de "região de aceitação", pois em estatística "não existem certezas" e o resultado pode variar consoante o nível de confiança. Essa será também a terminologia utilizada neste texto.

É, no entanto, muito mais vulgar decidir-se um teste de hipóteses com base em "estatísticas de teste" (e.t.) (vd. Secções 4.6 e 4.7).

## 4.4 Testes de hipóteses bilaterais para a média a partir de uma amostra

Com este "tipo" de hipóteses pretende-se responder a perguntas do género: "Será que o volume médio de sumo nas embalagens de determinada marca é igual a 20 cl, ou não?". Observe-se a Fig. 22(a), onde se ilustra este tipo de testes de hipóteses. O testes que se apresentam a seguir só se aplicam se a variável se distribui normalmente na população ou quando o tamanho da amostra é grande, por regra  $n > 30$  elementos. O procedimento que a seguir se apresenta utiliza conceitos relacionados com intervalos de confiança (mais à frente utilizam-se estatísticas de teste).

<sup>7</sup> Em Análise Sensorial, utiliza-se um painel de provadores para avaliar/classificar características de produtos (alimentares) recorrendo a escalas; por exemplo, pode-se usar uma escala de 5 pontos, de "Muito mau" ou 1 até "Muito Bom" ou 5.

<sup>8</sup> A quantidade  $1 - \beta$  é designada potência, ou "poder estatístico", do teste e refere-se à probabilidade de se rejeitar uma  $H_0$  falsa (i.e. de não se cometer um erro de Tipo II).

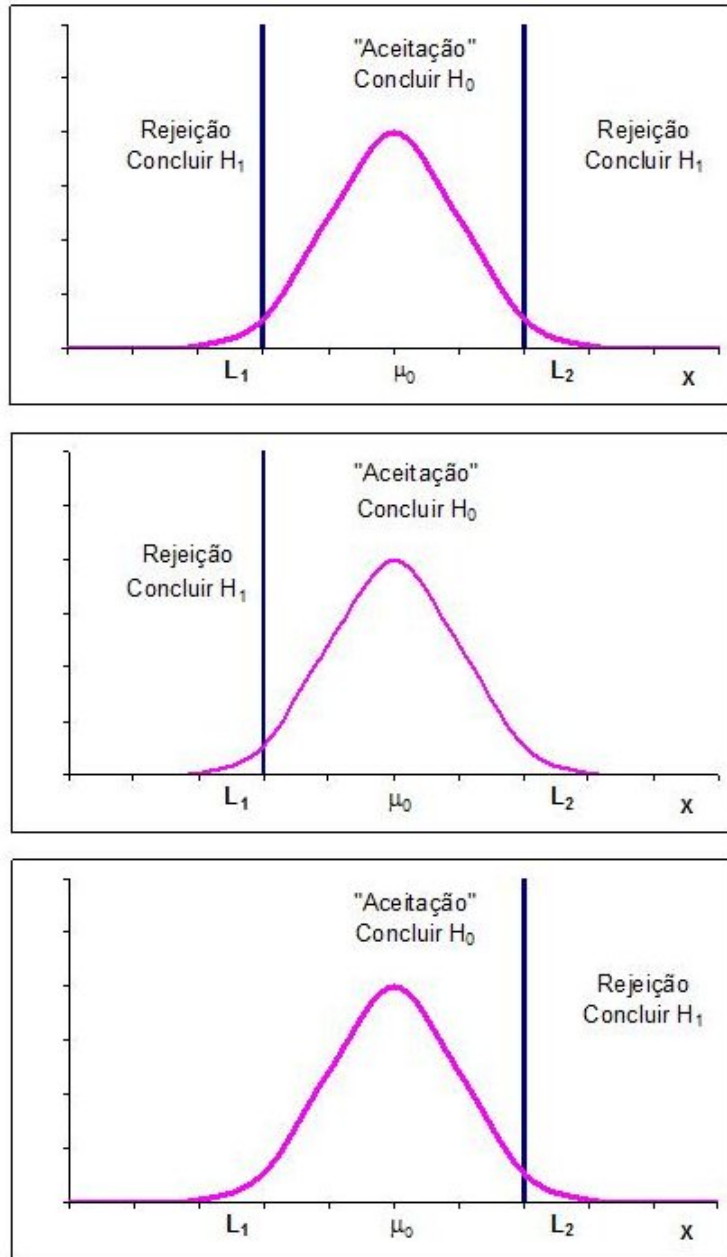


Figura 3: Representação esquemática do processo de decisão em três tipos de testes de hipóteses diferentes para um caso que diz respeito a uma amostra (ver texto para mais detalhes): (a) hipóteses bilaterais e (b) e (c) hipóteses unilaterais.

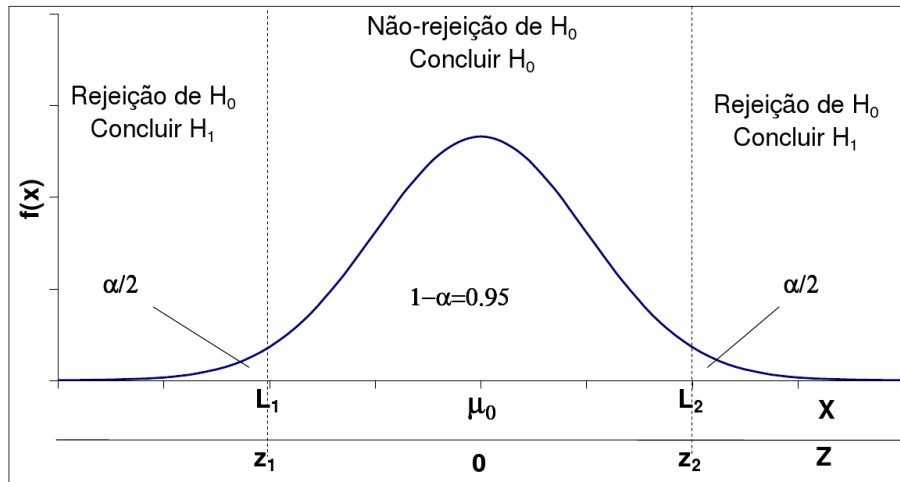


Figura 4: Ilustração da regra de decisão para  $H_0 : \mu = 500$  e  $H_1 : \mu \neq 500$  i.e.  $\mu_0 = 500$  (com  $\alpha = 0,05$ ). Apresentam-se duas escalas para as abcissas: os valores-limite na distribuição de  $Z$  ( $z_1 = z_{[\frac{\alpha}{2}]} = -1,96$  e  $z_2 = z_{[1-\frac{\alpha}{2}]} = +1,96$ ) e os valores da variável "original"  $X$  correspondentes àqueles limites ( $L_1 = 500 - 1,96(\frac{\sigma}{\sqrt{n}})$  e  $L_2 = 500 + 1,96(\frac{\sigma}{\sqrt{n}})$ ).

**Casos em que a variância da população  $\sigma^2$  é conhecida** Considerem-se as hipóteses nula e alternativa:

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0$$

A partir da amostra, calcular a média (das observações) e confrontar essa estimativa (amostral) com os limites de confiança apropriados,  $L_1$  e  $L_2$ . Se  $L_1 \leq \bar{x} \leq L_2$  concluir em favor de  $H_0$ ; caso contrário, isto é, se  $\bar{x} < L_1$  ou  $\bar{x} > L_2$ , concluir  $H_1$  (é prática corrente usar a expressão "rejeitar  $H_0$ " neste caso). Com referência à distribuição de  $\bar{x}$  na amostragem, e recorrendo à distribuição normal reduzida  $Z$ , os limites inferior e superior de confiança, respectivamente  $L_1$  e  $L_2$ , podem obter-se através de

$$L_1 = \mu_0 - z_{[1-\frac{\alpha}{2}]} \frac{\sigma}{\sqrt{n}}$$

e

$$L_2 = \mu_0 + z_{[1-\frac{\alpha}{2}]} \frac{\sigma}{\sqrt{n}}$$

em que  $z$  se obtém da Tabela A (anexa) para uma probabilidade  $1 - \alpha/2$  (daí a notação utilizada acima). Na Fig. 4, ilustra-se a regra de decisão num caso simples.

**Casos em que a variância da população  $\sigma^2$  não é conhecida** Quando não se conhece a variância populacional, é necessário recorrer à distribuição  $t$  de Student, ou seja, utilizar a transformação  $t$  com  $\nu = n - 1$  graus de liberdade. Assim, no teste de hipóteses sobre a média da população  $\mu$ , as hipóteses nula e alternativa mantêm-se, assim como a regra de decisão enunciado no caso anterior. Contudo, neste caso os limites inferior e superior de confiança,  $L_1$  e  $L_2$ , calculam-se da seguinte forma:

$$L_1 = \mu_0 - t_{[n-1, 1-\frac{\alpha}{2}]} \frac{s}{\sqrt{n}}$$

e

$$L_2 = \mu_0 + t_{[n-1, 1-\frac{\alpha}{2}]} \frac{s}{\sqrt{n}}$$

O valor de  $t$  obtém-se da Tabela B (anexa) para  $n - 1$  g.l. e uma probabilidade  $1 - \alpha/2$  (daí a notação usada nas equações anteriores).

Depois de um certo período experimental, em que 12 ratos-cobaia foram sujeitos a uma dieta composta por um complexo químico, será que a alteração de peso observada é significativa, com 95% de confiança? Sabe-se que após oito dias a diferença média de peso nos ratos-cobaia foi  $\bar{x} = -0,65$  g e que  $s^2 = 1,5682$  g<sup>2</sup>.

As hipóteses alternativas serão:  $H_0 : \mu = \mu_0$  ou  $H_1 : \mu \neq \mu_0$  (considere-se que  $\mu_0 = 0$  representa uma situação em que a diferença de peso foi nula *i.e.* a dieta não teve efeito sobre o peso dos ratos-cobaia). Neste caso, da Tabela B (anexa) obtém-se  $t=2,201$ . Os limites são, por conseguinte,  $L_1 = 0 - 2,201(\sqrt{1,5682/12}) = -0,796$  e  $L_2 = 0 + 2,201(\sqrt{1,5682/12}) = +0,796$ . Um vez que  $(L_1 = -0,796) < (\bar{x} = -0,65) < (L_2 = +0,796)$  conclui-se em favor de  $H_0$ , ou seja, com 95% de confiança a alteração de peso não foi significativamente diferente de zero.

#### 4.5 Testes de hipóteses unilaterais para a média a partir de uma amostra

Em muitas circunstâncias, interessa saber se  $\mu$  é significativamente maior (ou menor) do que  $\mu_0$ . Nestes casos utilizam-se testes de hipótese unilaterais. Na Fig. 3(b,c) ilustram-se os dois casos possíveis, respectivamente  $\mu > \mu_0$  e  $\mu < \mu_0$ . A definição das hipóteses iniciais e alguns dos cálculos envolvidos são diferentes do caso dos testes de hipóteses bilaterais.

Recorrendo a um exemplo relacionado com dietas, mas desta vez a questão é: Será que ocorreu redução significativa de peso em 12 voluntários para testar o efeito de um novo produto de naturopatia (com nível de confiança de 95%)? As hipóteses alternativas serão:  $H_0 : \mu \geq \mu_0$  vs.  $H_1 : \mu < \mu_0$  (em que  $\mu_0 = 0$ ).

Repare-se que, curiosamente, para se estudar a ocorrência de redução significativa de peso neste exemplo se definiram as hipóteses como  $H_0 : \mu \geq \mu_0$  ou  $H_1 : \mu < \mu_0$ . A hipótese alternativa  $H_1$  contém o resultado que "se desejava obter" (pelo menos, do ponto de vista do fabricante do produto em questão!). Esta "estratégia" resulta na área de rejeição ficar visualmente à esquerda de  $\mu_0$ , o que facilita a compreensão do teste (na medida em que se lê da esquerda para a direita!!). De modo similar, se se pretendesse estudar se o tempo de processamento de determinado produto agro-alimentar é maior do que o especificado no manual de qualidade da empresa (15 min.) as hipóteses poderiam ser:  $H_0 : \mu \leq 15$  min ou  $H_1 : \mu > 15$  min.

Os cálculos envolvidos são diferentes dos apresentados anteriormente em virtude das diferenças na definição das hipóteses iniciais. De facto, recorre-se apenas a um limite em vez de um intervalo e, portanto, consideram-se os valores (críticos) teóricos  $z_{[1-\alpha]}$  ou  $t_{[\nu,1-\alpha]}$  em que  $\nu = n - 1$ . Assim, para os casos de hipóteses unilaterais as regras de decisão são:

1. Para as hipóteses  $H_0 : \mu \geq \mu_0$  ou  $H_1 : \mu < \mu_0$ : Concluir  $H_0$  se  $\bar{x} \geq L$ ; Caso contrário (ou seja, se  $\bar{x} < L$ ), concluir  $H_1$ . O limite (inferior) de confiança calcula-se de uma das seguintes formas:

$$L = \mu_0 - z_{[1-\alpha]} \frac{\sigma}{\sqrt{n}}$$

quando se conhece a variância  $\sigma^2$ , ou então, através de

$$L = \mu_0 - t_{[n-1,1-\alpha]} \frac{s}{\sqrt{n}}$$

2. Para as hipóteses  $H_0 : \mu \leq \mu_0$  vs.  $H_1 : \mu > \mu_0$ : Se  $\bar{x} \leq L$ , concluir  $H_0$ ; caso contrário, *i.e.* se  $\bar{x} > L$ , concluir  $H_1$ . O limite (superior) de confiança, quando se conhece a variância  $\sigma^2$ , calcula-se da seguinte forma:

$$L = \mu_0 + z_{[1-\alpha]} \frac{\sigma}{\sqrt{n}}$$

ou então recorrendo à distribuição *t*-Student e à variância amostral  $\sigma^2$ , através de:

$$L = \mu_0 + t_{[n-1,1-\alpha]} \frac{s}{\sqrt{n}}$$

(continuação do exemplo anterior) Agora, sabe-se que a diferença média dos pesos ao fim de quinze dias de dieta foi  $\bar{x} = -0,61$  kg e que  $s^2 = 0,4800$  kg<sup>2</sup>. Da Tabela B (anexa) obtém-se, neste caso, que  $t = 1,796$  para 11 g.l. e  $\alpha = 0,05$ , logo o limite (inferior) será:  $L = 0 - 1,796(\sqrt{0,4800/12}) = -0,3282$ . Assim, como  $\bar{x} < L$  conclui-se  $H_1$ , isto é, com 95% de confiança ocorreu redução significativa do peso médio.

Note-se, mais uma vez, que no caso de testes de hipóteses unilaterais os valores de *z* (ou *t*) são obtidos para probabilidades de  $(1 - \alpha)$  em vez de  $(1 - \alpha/2)$ , em virtude de interessar apenas um dos limites do intervalo de confiança.

## 4.6 Testes de hipóteses para uma amostra através de “estatísticas de teste”

Nos tópicos anteriores acerca de testes de hipóteses, apresentámos um modo de realizar aqueles testes acerca da média populacional que se baseia no conceito de intervalo de confiança. A estatística amostral, neste caso  $\bar{x}$ , é confrontada com determinados limites para se decidir rejeitar, ou não, a  $H_0$ . No entanto, é possível abordar a questão recorrendo a "estatísticas de teste" (o procedimento habitual em estatística aplicada!). Depois de se estabelecerem as hipóteses nula e alternativa, calcula-se a "estatística de teste" (e.t.) a partir da(s) amostra(s). Esta é comparada com valores teóricos ou críticos apropriados (tabelados ou obtidos através de *software*) para se concluir em favor de  $H_0$  ou, caso contrário, em favor de  $H_1$ .

Para testar hipóteses relativas à média  $\mu$  de uma amostra, podem usar-se duas e.t. consoante se conhece, ou não, a “verdadeira” variabilidade dos dados (i.e. a variância  $\sigma^2$ ). Aquelas e.t. derivam das transformações  $Z$  ou  $t$ , e usam as respectivas distribuições teóricas de probabilidades. Assim, se se conhecer  $\sigma^2$  pode utilizar-se a e.t.  $z_0$ , ou seja:

$$z_0 = \frac{\bar{x} - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

No caso de  $H_1 : \mu \neq \mu_0$ , se  $|z_0| > z_{[1-\frac{\alpha}{2}]}$ , então pode-se concluir em favor de  $H_1$  (i.e. rejeita-se  $H_0$  com  $(1-\alpha) \times 100\%$  de confiança). Para hipóteses unilaterais, por exemplo se  $H_1 : \mu < \mu_0$ , rejeita-se  $H_0$  no caso de  $z_0 < -z_{[1-\alpha]}$  (pelo contrário, para  $H_1 : \mu > \mu_0$  rejeita-se  $H_0$  se  $z_0 > z_{[1-\alpha]}$ ). Os valores críticos obtêm-se da tabela apropriada (Tabela A anexa) para uma probabilidade  $(1-\alpha/2)$  no caso de hipóteses bilaterais ou  $(1-\alpha)$  no caso de hipóteses unilaterais.

De modo semelhante, se utiliza a e.t.  $t_0$  nos casos em que não se conhece  $\sigma^2$  (embora se assuma que a distribuição da variável é normal) e se usa, por isso, a variância amostral  $s^2$  (com denominador  $(n-1)$ ):

$$t_0 = \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$$

Para  $H_1 : \mu \neq \mu_0$ , rejeita-se  $H_0$  com  $(1-\alpha) \times 100\%$  de confiança se  $|t_0| > t_{[n-1; 1-\frac{\alpha}{2}]}$ . Por outro lado, conclui-se em favor de  $H_1 : \mu > \mu_0$  se  $t_0 > t_{[n-1; 1-\alpha]}$ . Considerem-se os exemplos anteriores, agora tratados recorrendo a estatísticas de teste.

Depois de um período determinado em que 12 ratos-cobaia foram sujeitos a uma dieta composta por certo complexo químico, será que a alteração de peso observada é significativa, com 95% de confiança? Sabe-se que  $\bar{x} = -0,65$  g e que  $s^2 = 1,5682$  g<sup>2</sup>. As hipóteses alternativas são  $H_0 : \mu = \mu_0$  ou  $H_1 : \mu \neq \mu_0$ . Calcula-se a estatística de teste  $t_0 = -1,798$ . Da Tabela B (anexa) obtém-se  $t=2,201$  para 11 g.l. e  $\alpha=0,05$ . Como  $|t_0| < t$ , não se rejeita a  $H_0$ . A conclusão é idêntica à obtida anteriormente recorrendo à noção de intervalo de confiança.

Recorrendo a um exemplo relacionado com dietas, mas desta vez a questão é: será que ocorreu redução significativa de peso em 12 voluntários para testar o efeito de um novo produto de naturopatia (com nível de confiança de 95%)? E agora, sabe-se que  $\bar{x} = -0,61$  kg e que  $\sigma^2 = 0,4008$  kg<sup>2</sup>. As hipóteses alternativas (unilaterais) são:  $H_0 : \mu \geq \mu_0$  e  $H_1 : \mu < \mu_0$ . A estatística de teste  $t_0 = -3.338$  e da Tabela B obtém-se que  $t = 1,796$  para  $\alpha = 0,05$ . Como  $t_0 < -t$ , rejeita-se a  $H_0$ , isto é, a conclusão é igual àquela que se obteve usando limites.

Este é o procedimento mais vulgar para realizar testes de hipóteses não só acerca da média como para outros parâmetros populacionais importantes. Nos manuais de estatística aplicada, os autores referem-se a estes testes de hipóteses como teste de  $z$  e teste de  $t$  para uma amostra (em inglês, "one-sample z-test" ou "one-sample t-test").

## 4.7 Testes de hipóteses relativos às médias de duas amostras

**Amostras independentes** (em preparação)

**Amostras emparelhadas** (em preparação)

## 5 Exercícios

1. Considere a seguinte população estatística, constituída pelas alturas (em cm) de 39 alunos de Estatística do ano lectivo 1992/1993 (vd. Tab. 2). a) Calcule a média, a variância e o desvio-padrão da população; b)

Tabela 2: Alturas (em cm) dos 39 alunos que frequentaram a disciplina de Estatística no ano lectivo de 1992/1993.

N <sup>o</sup>	Altura	N <sup>o</sup>	Altura	N <sup>o</sup>	Altura
1	162	14	162	27	157
2	170	15	184	28	156
3	165	16	176	29	167
4	160	17	158	30	156
5	157	18	153	31	180
6	168	19	165	32	153
7	160	20	166	33	163
8	158	21	174	34	161
9	156	22	160	35	172
10	162	23	170	36	182
11	163	24	165	37	172
12	174	25	185	38	160
13	182	26	172	39	170

Selecione 5 amostras com critério aleatório simples de tamanho  $n=12$ , com reposição (utilize a tabela de números aleatórios; Tabela C anexa, para obter as amostras) e calcule para cada amostra a média  $\bar{x}$  e o desvio-padrão  $s$ ; c) Para cada amostra, determine uma estimativa pontual e uma estimativa por intervalo de confiança da média da população  $\mu$  (considere  $\sigma$  conhecido). Compare estas estimativas com o valor de  $\mu$  calculado anteriormente; d) Repita a alínea anterior considerando  $\sigma$  desconhecido.

- Determine o tamanho da amostra  $n$  de modo a estimar a média de uma população  $\mu$ , sabendo que a sua variância  $\sigma^2 = 40$  e que o intervalo de confiança de 95% tem amplitude  $h=4$ .
- O conteúdo (em litros) de garrafas de óleo alimentar segue distribuição normal. Admita-se que os respectivos parâmetros são  $\mu = 0,99$  L e  $\sigma = 0,02$  L. Nestas condições: a) Qual é a probabilidade de o conteúdo médio numa amostra de 16 garrafas seleccionadas para inspecção ser superior a 1 L? b) Qual é a probabilidade de, numa amostra de 100 garrafas, o conteúdo médio ser inferior a 9,85 dL? c) Tendo recolhido uma amostra de 100 garrafas e determinado um conteúdo médio inferior a 9,85 dL, que pensaria da hipótese avançada de início (isto é,  $\mu = 0,99$  L)? d) Encontre um intervalo de confiança tal que a probabilidade de  $\bar{x}_{100}$  nele estar contida seja de 0,95, isto é encontrar  $a$  e  $b$  tais que  $P\{a < \bar{x}_{100} < b\} = 0,95$ .
- Supõe-se que a média de uma população é  $\mu = 50$  cm. Seleccionou-se uma amostra de tamanho  $n=10$  com um critério a.s. com reposição. Obtiveram-se os seguintes valores:  $\bar{x} = 53$  cm e  $s=3$  cm. Para um nível de confiança de 99%, verifique se a suposição inicial feita sobre a média populacional é verdadeira.
- A dose diária recomendada (DDR) de cálcio para adultos da classe etária 25 – 50 anos é 800 mg. Duma população Algarvia, seleccionaram-se com critério a.s., treze indivíduos desse intervalo de idades, para os quais se determinou a quantidade daquele mineral que ingerem por dia. Os valores (em mg) para essa amostra da população foram: 987, 888, 741, 698, 654, 666, 678, 543, 567, 555, 639, 693 e 654. a) Será que as pessoas ingerem menor quantidade de cálcio do que a DDR? b) Noutra população, a quantidade média de cálcio ingerida (obtida a partir duma amostra a.s. de vinte adultos) foi de 876 mg (com  $s = 54$  mg). Será que nesta população se ingere mais cálcio do que a dose diária recomendada?