Introdução à Inferência estatística – estimação intervalar ou intervalos de confiança

1 – Introdução

Nas secções anteriores estudámos o comportamento da Média e da Proporção amostral, como estimadores, respectivamente, do valor médio de uma População e da proporção com que os elementos da População verificam determinada característica. Verificámos que, quando se consideram amostras diferentes, embora da mesma dimensão, a média ou a proporção variam de amostra para amostra, mas apresentam um **comportamento característico**, de uma distribuição *aproximadamente simétrica*, com *pequena variabilidade*, acentuando-se estas características à medida que a dimensão da amostra aumenta.

Por exemplo, no caso da estimação do valor médio, o facto de a média variar de amostra para amostra, não nos permite saber, recolhida uma amostra, se a média dessa amostra é uma boa estimativa do valor médio da população subjacente à amostra (como temos feito várias vezes, estamos a identificar população com a variável em estudo, cujo valor médio se pretende conhecer), isto é, não podemos atribuir nenhuma "confiança" a essa estimativa do valor médio.

2 – Intervalo de confiança para o valor médio

No estudo da distribuição de amostragem da média, concluímos ainda que quando se faz amostragem sem reposição e as Populações têm dimensão razoavelmente grande, ou no caso de a amostragem ser com reposição, Populações com qualquer dimensão, e as amostras também têm dimensão grande (maior ou igual a 30), <u>a distribuição de amostragem da Média pode ser aproximada pela distribuição Normal</u> (Teorema Limite Central).

Este comportamento da distribuição de amostragem da Média tem consequências muito importantes, no que diz respeito ao problema da estimação do parâmetro valor médio, já que vamos aproveitá-lo para encarar este problema de um outro ângulo. Em vez de procurarmos um valor **– estimativa pontual**, como aproximação do valor do parâmetro desconhecido, vamos procurar obter um intervalo – estimativa intervalar **ou intervalo de confiança**, que com uma determinada confiança contenha o valor do parâmetro.

Vamos então procurar um intervalo aleatório [A, B] que, com uma "grande probabilidade", por exemplo 0.95, contenha o parâmetro μ:

$$P([A, B] \text{ conter } \mu) = 0.95$$

Ora, é precisamente na construção destes intervalos de confiança, que vamos aproveitar o facto de a distribuição de amostragem da Média poder ser aproximada pelo modelo Normal, com valor médio igual ao valor médio μ da População (parâmetro que estamos a estimar) e desvio padrão igual a **$\sigma/\sqrt{n}$**, onde σ é o desvio padrão da população. Como o desvio padrão da População é quase sempre desconhecido, vamos também estimá-lo a partir do desvio padrão amostral, **s**, pelo que um valor aproximado para o desvio padrão da média, também conhecido como **erro padrão**, é $s/\sqrt{n}$.

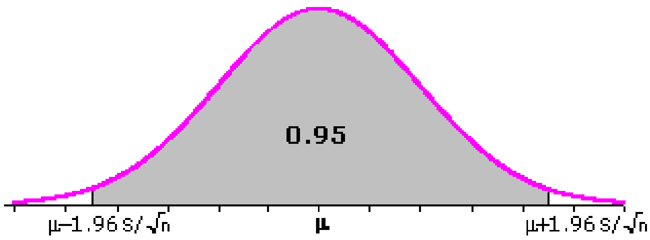
Então, tendo em consideração as propriedades da distribuição Normal, podemos escrever:

$$P(-1.96 \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq 1.96) \approx 0.95 \qquad (1)$$

O valor 1.96 pode ser obtido consultando uma tabela, a calculadora ou a folha de Excel.

De (1) vem

$P(\mu - 1.96\ S/\sqrt{n} \leq \overline{X} \leq \mu + 1.96\ S/\sqrt{n}) \approx 0.95$



ou

$$P(\overline{X} - 1.96\ S/\sqrt{n} \leq \mu \leq \overline{X} + 1.96\ S/\sqrt{n}) \approx 0.95$$

Então o intervalo aleatório que andávamos à procula é

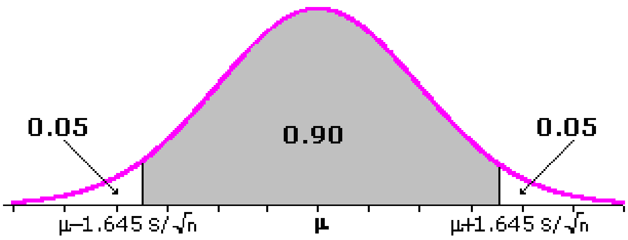
$$[\,\overline{x} - \mathbf{1.96} \times \mathbf{S}/\sqrt{n},\ \overline{x} + \mathbf{1.96} \times \mathbf{S}/\sqrt{n}\,]$$

Repare-se que o intervalo anterior é aleatório, já que o valor da média e do desvio padrão variam, dependendo da amostra que se recolher. Se recolhermos duas amostras diferentes, ambas da mesma dimensão, vamos obter valores diferentes para a média e para o desvio padrão. Dizemos que este intervalo é um **intervalo de confiança**, com uma **confiança ou um nível de confiança de 95%.**

Afinal, o que significa um intervalo de 95 % de confiança?

Significa que se recolhermos muitas amostras de dimensão n, calcularmos as médias e os desvios padrões dessas amostras e construirmos os intervalos de confiança respectivos, utilizando a expressão anterior, cerca de 95% desses intervalos conterão o valor médio μ, enquanto que os restantes 5% não conterão o parâmetro μ. Não temos a certeza que um dado intervalo, em particular, contenha o parâmetro desconhecido, mas estamos confiantes que assim aconteça, isto é estamos **95% confiantes** que o intervalo que calculámos a partir da amostra seleccionada (na prática só seleccionamos uma amostra), contenha o valor do parâmetro.

Se na expressão (1) da probabilidade, mudarmos a probabilidade de 0.95 para 0.90, por exemplo, então em vez de 1.96, devemos considerar 1.645:



Assim, um intervalo de confiança, com 90% de confiança terá o seguinte aspecto

$$[\,\overline{X} - 1.645 \times S/\sqrt{n},\ \overline{X} + 1.645 \times S/\sqrt{n}\,]$$

A forma geral do intervalo de confiança será,

$$[\,\overline{X} - \mathbf{z} \times S/\sqrt{n},\ \overline{X} + \mathbf{z} \times S/\sqrt{n}\,]$$

onde o valor de **z** dependerá da confiança com que se pretende construir o intervalo.

Alguns valores (obtidos a partir da distribuição da Normal(0,1)), incluindo os já considerados anteriormente, são:

Confiança	z
90%	1.645
95%	1.960
97.5%	2.326
99%	2.576
99.5%	3,090
99.9%	3.291
99.95%	3.891
99.995%	4.417

Como se verifica a partir da tabela anterior, quanto *maior for a confiança*, maior é o valor de **z**, pelo que *maior será a amplitude* do intervalo.

Como diminuir a amplitude de um intervalo de confiança?

De um modo geral pretende-se construir um intervalo com **pequena amplitude**, pois nos dá uma **maior precisão**. Como se depreende da forma desse intervalo, para diminuir a sua amplitude, que é dada por $2 \times z \times \frac{S}{\sqrt{n}}$ podemos fazê-lo de duas maneiras:

- ou diminuir a confiança (o que faz com que diminua o valor de z), o que não é aconselhável;
- ou aumentar a dimensão da amostra considerada para calcular o intervalo. Por exemplo, se aumentar 4 vezes a dimensão da amostra, a amplitude do intervalo reduz-se a metade.

Nas considerações anteriores estamos a admitir que a dimensão da amostra inicial já é suficientemente grande, de modo que a estimativa **s** para o desvio padrão da população não se altera significativamente quando utilizamos mais informação (uma amostra de maior dimensão) para a calcular.

Como casos extremos de intervalos de confiança, temos:

- o intervalo de confiança, com uma **confiança 0%,** que se reduz a um ponto, que não é mais do que a estimativa pontual do valor médio, ou seja a média calculada a partir da amostra considerada;
- e temos ainda o intervalo com uma **confiança de 100%**, que é a recta real (porque vem o valor de z igual a infinito).

Obviamente que nenhum destes intervalos é de grande utilidade!

Margem de erro

A metade da amplitude de um intervalo de confiança, é costume chamar *margem de erro*.

Exemplo – Considerando a população dos deputados da X Legislatura, suponhamos que estávamos interessados em estimar o parâmetro idade média da população. Seleccionou-se uma amostra aleatória (com reposição) de dimensão 30 e registaram-se as idades dos elementos seleccionados. Os valores obtidos apresentam-se na seguinte tabela:

46	50	47
34	70	55
54	36	30
48	39	52
40	55	54
41	56	52
40	60	42
49	53	44
54	32	48
71	50	31

A média e o desvio padrão das idades anteriores são, respectivamente, 47.8 e 10.2 anos. Então, um intervalo de 95% de confiança para a idade média da população é

$$[47.8 - 1.96 \times 10.2/\sqrt{30},\ 47.8 + 1.96 \times 10.2/\sqrt{30}]$$

ou seja [44.2, 51.4], é um intervalo com uma confiança de 95%. Repare-se que o intervalo anterior contém o parâmetro em estudo (esta é uma situação de excepção, em que a população é tão pequena, que facilmente se obtém o valor do parâmetro valor médio da Idade).

Chamamos a atenção para que, se não conhecêssemos o valor do parâmetro em estudo, não poderíamos garantir que o intervalo que calculámos anteriormente o contivesse. Apenas estamos **confiantes** em que isso acontecesse, pois se calculássemos 100 amostras de dimensão 30, como a anterior, esperávamos que cerca de 95 dos intervalos que se poderiam construir com as médias e desvios padrões dessas amostras, contivessem o parâmetro em estudo.

Exemplo – Considere a população constituída pelos empregados da empresa X, em Anexo. Suponha que estamos interessados em estudar o parâmetro **altura média**.

a) Seleccione uma amostra de dimensão 30 e calcule um intervalo de 95% de confiança para o parâmetro em estudo.

b) Seleccione mais 99 amostras de dimensão 30, e a partir de cada uma delas construa um intervalo de 95% de confiança.

c) Quantos dos intervalos considerados anteriormente contêm o valor do parâmetro? Comente.

Resolução:

a) A seguir apresenta-se a amostra seleccionada pelo processo de amostragem com reposição

153	160	163	173	174	173	169	154	164	165	156	169	159	157	173
154	177	160	170	161	161	174	165	170	165	157	158	160	160	170

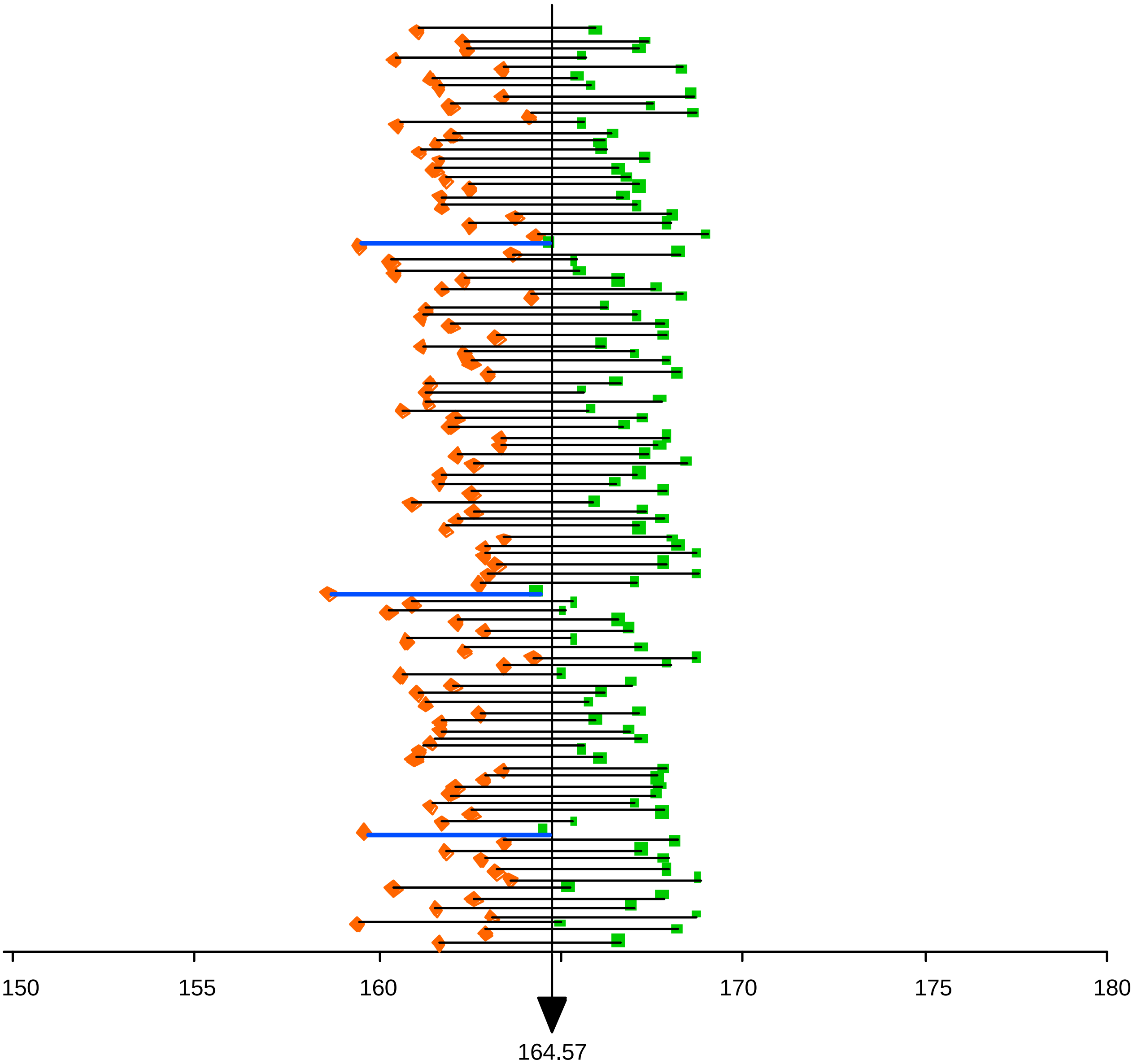
Média = 164.13

Desvio padrão = 6.95

Intervalo de 95% de confiança $[164.13 - 1.96 \times \frac{6.95}{\sqrt{30}}, 164.13 + 1.96 \times \frac{6.95}{\sqrt{30}}]$

[161.65, 166.62]

b) Seleccionámos mais 99 amostras aleatórias, para as quais calculámos a média e o desvio padrão e os respectivos intervalos de confiança, pelo mesmo processo que o calculado na alínea anterior. Apresentamos esses intervalos graficamente, na figura seguinte:



Na figura anterior, a seta indica a posição do valor do parâmetro a estimar, ou seja a altura média da população.

c) Verificamos que três dos intervalos construídos não contêm o valor do parâmetro (Esperarávamos encontrar um valor próximo de 5).

3 – Intervalo de confiança para a proporção

No estudo da distribuição de amostragem da proporção, concluímos que quando se faz amostragem sem reposição e as Populações têm dimensão razoavelmente grande, ou no caso de a amostragem ser com reposição, Populações com qualquer dimensão, e as amostras também têm dimensão grande (maior ou igual a 30), <u>a distribuição de amostragem da Proporção amostral pode ser aproximada pela distribuição Normal</u> (Teorema Limite Central). Este comportamento da distribuição de amostragem da Proporção, tal como vimos anteriormente para a Média, tem consequências muito importantes, no que diz respeito ao problema da estimação do parâmetro proporção populacional, já que vamos aproveitá-lo para encarar este problema de um outro ângulo. Em vez de procurarmos um valor **– estimativa pontual**, como aproximação do valor do parâmetro desconhecido, vamos procurar obter um intervalo – estimativa intervalar **ou intervalo de confiança**, que com uma determinada confiança contenha o valor do parâmetro.

Representando por $\hat{p}$ a proporção amostral, estimador do parâmetro **p**, sabemos do módulo 2 – Introdução à Estimação que, se a recolha da amostra for feita com reposição de uma População de dimensão qualquer, ou sem reposição de uma população de grande dimensão, e se a dimensão, n, da amostra for grande, então

$$P\left(\frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z\right) \approx \Phi(z)$$

Por um processo perfeitamente idêntico ao considerado para obter o intervalo de confiança para o valor médio, em que no caso em que a variância σ^2 da população, é desconhecida, a substituímos pela variância amostral, também aqui, substituímos na variância da população p(1-p), o p por $\hat{p}$. Temos assim o intervalo de confiança para a proporção

$$[\hat{p} - \mathbf{z}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p} + \mathbf{z}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ]$$

Fazendo a analogia com o que se passa com o intervalo de confiança para o valor médio, no intervalo anterior, a proporção $\hat{p}$, substituiu a média, e considerou-se $\hat{p}(1-\hat{p})$ como estimador da variância populacional p(1-p).

Observação: Ao contrário do que é usual, em que se considera a variável aleatória com letra maiúscula e um seu valor observado com letra minúscula, no caso da proporção não é costume fazer essa distinção. Assim, representa-se indiferentemente por $\hat{p}$ tanto a variável aleatória como um seu valor observado, dependendo do contexto em que está a ser utilizado a sua interpretação como variável aleatória ou valor dessa variável aleatória.

O valor de **z** depende da confiança com que se quer construir o intervalo, como vimos para o caso do valor médio.

No caso particular de um intervalo de 95% de confiança, temos

$$\left[\hat{p} - \mathbf{1.96}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p} + \mathbf{1.96}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Em que condições é que se pode utilizar o intervalo anterior?

Dissemos anteriormente que era necessário que a dimensão da amostra fosse suficientemente grande. No entanto, também já vimos que quanto maior for a variabilidade presente na população de onde se recolhe a amostra, maior terá de ser a dimensão dessa amostra. Uma regra empírica aconselha-nos a considerar

$$n\hat{p} \geq 10 \text{ e } n(1 - \hat{p}) \geq 10$$

O intervalo anterior, obtido a partir de uma amostra de dimensão n, tem amplitude igual a

$$\mathbf{2 \times 1.96}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Como já referimos para o intervalo de confiança para o valor médio, a metade da amplitude do intervalo, ou seja à quantidade $1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, chamamos *margem de erro* da sondagem.

Exemplo – Suponha que para a população dos empregados da empresa X se pretende estimar a proporção de mulheres casadas. Seleccione uma amostra de dimensão 30 e obtenha uma estimativa pontual e uma estimativa intervalar ou intervalo de 95% de confiança para essa proporção.

Resolução:

Para facilitar o estudo, utilizámos uma folha de Excel e convertemos as categorias da característica populacional estado civil, da seguinte forma:

Casada	1	Solteira	0
Casado	0	Solteiro	0
Divorciada	0	Divorciado	0

Obtivemos uma população de 0's e 1's, em que um elemento da população assumia o valor 1 sempre que a característica em estudo se verificava.

Seleccionada uma amostra aleatória, com reposição, de dimensão 30, obtivemos os seguintes valores:

0	1	0	0	1	0	0	1	0	0	0	1
0	0	0	1	0	0	0	1	0	0	0	1
0	0	0	0	0	0						

Proporção de mulheres casadas na amostra= 7/30 = 0.233

Este valor é *uma estimativa pontual* da proporção p de mulheres casadas na população, cujo valor é 0.2268 (Mais uma vez estamos numa situação em que foi fácil calcular o valor do parâmetro, atendendo a que a população tinha uma dimensão muito pequena)..

Um intervalo de 95% de confiança para a proporção de mulheres casadas na população é

$$[0.233 - 1.96 \times \sqrt{\frac{0.233\,(1-0.233)}{30}}\ ,\ 0.233 + 1.96 \times \sqrt{\frac{0.233\,(1-0.233)}{30}}\ ],$$

ou seja [0.082, 0.384]. O intervalo anterior tem uma margem de erro de 0.151.

Exemplo – O Diário de Notícias na sua edição do dia 25 de Fevereiro de 1998, referia, relativamente a uma sondagem realizada em colaboração com a TSF/Universidade Moderna, que a confiança dos portugueses é "açambarcada" pelos docentes, logo seguidos pelos médicos, adiantando ainda que os políticos e sindicalistas partilham os últimos lugares. Apresentamos de seguida um excerto desse artigo.

Barómetro de profissões

Em relação às seguintes profissões com importância na vida nacional, diga se tem ou não confiança na sua acção

Profissão	Não tem confiança	Tem confiança	Não sabe/Não responde
Professores	9.1%	85.1%	5.8%
Médicos	9.7%	85.0%	5.3%
Juízes	26.9%	64.7%	8.4%
Militares	30.9%	61.6%	7.4%
Jornalistas	30.8%	58.2%	11.1%
Padres	31.8%	55.2%	13.0%
Empresários	33.5%	53.6%	13.0%
Polícias	38.6%	51.8%	9.6%
Sindicalistas	49.6%	39.4%	11.0%
Políticos	57.5%	30.9%	11.7%

Ficha técnica: Esta sondagem foi encomendada pelo DN e pela TSF ao Centro de Sondagens da Universidade Moderna. O trabalho de campo decorreu entre os dias 12 e 17 de Fevereiro de 1998. Os inquéritos foram realizados em 25 freguesias de Portugal continental. A amostra foi seleccionada aleatoriamente e, para cada uma das freguesias, foi feito um estudo demográfico com base nos dados do Censos 91 e do Stape. Foram validados 1303 inquéritos. O processo de informação utilizado foi a recolha directa (porta a porta), através de um inquérito estruturado onde estava anexado o boletim com o nome de várias profissões com importância na vida nacional. Foi feita uma primeira validação pelos monitores, no local, a 10 por cento dos inquéritos e, posteriormente, foram validados outros 20 por cento telefonicamente. O erro máximo, para um nível de confiança de 95%, é de 2.71. A análise dos resultados é da responsabilidade do DN.

a) O valor de 85.1% apresentado para os Professores é uma estatística ou um parâmetro?

b) Construa um intervalo de 95% de confiança para a percentagem da população que tem confiança nos Professores.

c) O intervalo anterior contém necessariamente a percentagem de indivíduos da população que têm confiança nos Professores?

d) Calcule a margem de erro dos intervalos de confiança, para a confiança associada às diferentes profissões consideradas.

e) Qual o valor máximo obtido para as margens de erro obtidas na alínea anterior? Isso estará de acordo com o que vem especificado na ficha técnica?

Resposta alínea d)

85.1%	0.0193
85.0%	0.0194
64.7%	0.0259
61.6%	0.0264
58.2%	0.0268
55.2%	0.0270
53.6%	0.0271
51.8%	0.0271
39.4%	0.0265
30.9%	0.0251

Reparou que: a margem de erro é máxima para um valor da proporção próximo de 0.5? Efectivamente se se fizer o estudo da função $f(\hat{p}) = 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, verifica-se que ela assume um valor máximo para $\hat{p}$ =0.5.

Qual a dimensão da amostra que se deve recolher para obter um intervalo com um nível de confiança de 95% e com uma determinada precisão?

Pretende-se que a margem de erro do intervalo de confiança seja menor ou igual que um valor d. Então, temos de resolver a seguinte desigualdade, em relação a n:

$$1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq d$$

de onde vem

$$n \geq \left(\frac{1.96}{d}\right)^2 \hat{p}(1-\hat{p}) \qquad (1)$$

Como, de um modo geral, não se conhece o valor da proporção amostral, antes de recolher a amostra, considera-se o valor máximo para a expressão $\hat{p}(1-\hat{p})$, que se obtém quando o valor da proporção é 0.5. Vem então

$$n \geq \left(\frac{1.96}{2d}\right)^2 \qquad (2)$$

No caso do exemplo anterior, a margem de erro relativamente aos portugueses que confiam nos professores, é de 1.93%. Se se pretendesse uma margem de erro não superior a 1.5%, teríamos de considerar uma amostra de dimensão **2165**, se entrarmos em consideração com o valor da estimativa 0.851, na fórmula (1).

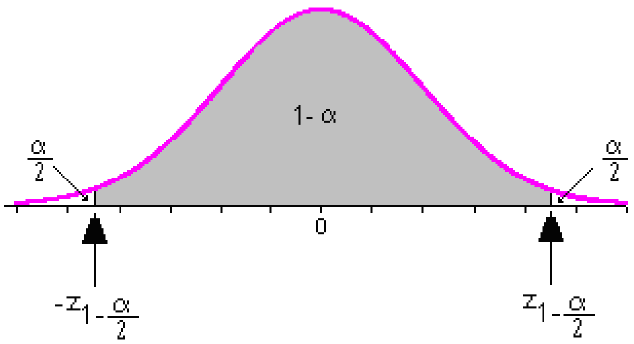
Se desconhecêssemos uma estimativa para a proporção, entraríamos com a fórmula (2) para calcular o valor da dimensão da amostra e obteríamos como valor necessário para a dimensão da amostra **4269**!

Qual a dimensão da amostra que se deve recolher para obter um intervalo com uma determinada precisão, com um nível de confiança de 100(1-α)%?

A confiança de um intervalo costuma exprimir-se na forma anterior, onde α é uma probabilidade relativamente pequena. Assim, se α=5%, temos um intervalo de 95% de confiança. Se representarmos a função densidade da Normal(0,1), o valor de z que aparece no intervalo de confiança genérico

$$[\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,,\ \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,]$$

não é mais do que o quantil de probabilidade 1-α/2, como se apresenta a seguir:



pelo que se se pretende um intervalo de 100(1-α)% de confiança, com uma precisão não inferior a **d**, a expressão (2) anterior, para a dimensão **n** da amostra necessária, toma a forma:

$$n \geq \left(\frac{z_{1-\alpha/2}}{2d}\right)^2$$

Ainda sobre intervalos de confiança

A interpretação do que é um intervalo de confiança nem sempre é feita correctamente. Vamos aproveitar o seguinte diálogo para ficarmos com as ideias um pouco mais claras sobre este assunto.

Suponhamos que um candidato à Câmara de Lisboa, o Dr. Gentil Alves, pretendia saber qual a percentagem **p**, de eleitores (lisboetas) que pensavam votar nele. Encomendou um estudo à empresa Sondagem, tendo esta questionado 785 lisboetas, escolhidos aleatoriamente, e verificado que a percentagem destes eleitores que pensavam votar no candidato era 56%. Se este valor dava um certo alento ao Dr. Gentil Alves para se candidatar, não o deixava, no entanto, descansado! Ele sabia que se fosse recolhida outra amostra, embora da mesma dimensão, quase de certeza obteria outro valor como estimativa de **p** e quem é que lhe garantia que não era um valor inferior a 50%, o que o deixaria infelicíssimo! Como interpretar este valor de 56%? O Prof. Amável, um amigo estatístico do Dr. Gentil Alves, ajudou-o nesta tarefa. Relatamos a seguir a conversa que se passou entre ambos.

Dr. Gentil Alves – Bom dia Amável, estás bem? Olha, ando um pouco preocupado com esta questão da candidatura à Câmara de Lisboa. Numa sondagem realizada ontem, deram-me uma percentagem de 56% de eleitores a votarem em mim. Mas com que confiança é que eu posso interpretar este resultado? Posso estar seguro que tenho a maioria?

Prof. Amável – Para te ser franco, a confiança que podes ter nesse resultado é nula! Tu próprio sabes que se tivessem sido outros eleitores escolhidos para a sondagem, quase certamente não obterias 56%. Mas não fiques muito preocupado, pois eu vou adiantar-te mais alguma coisa. O valor de 56% vai-me servir para obter um intervalo de 95% de confiança. Deixa-me fazer aqui umas contas que já te telefono.

Dr. Gentil Alves – Está bem. Muito obrigado.

Prof. Amável – Cá estou eu novamente. Com esse valor que me adiantaste construí o intervalo (52.5%; 59.5%), que é um intervalo de 95% de confiança para a percentagem de lisboetas que pensam votar em ti. Estás contente?

Dr. Gentil Alves – Significa isso que existe uma probabilidade de 95% desse intervalo conter essa percentagem (**p**) de eleitores que pensam votar em mim?

Prof. Amável – Nada disso!

Dr. Gentil Alves – Então 95% é a probabilidade de **p** estar contido no intervalo?

Prof. Amável – Que horror! Porventura o **p** é uma variável aleatória? Nem o p nem o intervalo que eu te dei. Assim não podemos falar na probabilidade do **p** estar contido no intervalo, nem do intervalo conter o **p**! Os 95% de confiança significam o seguinte: o processo que se utiliza para calcular os intervalos, como o que te apresentei, é um processo tal que se o utilizasse com todas as amostras possíveis (da mesma dimensão) que posso seleccionar da população, cerca de 95% das vezes produziria intervalos que contêm o p e cerca de 5% das vezes intervalos que não o contêm. No que diz respeito a um intervalo particular, como o que te dei, ficaremos sempre na dúvida se é um dos que contém p ou não! Temos “fé” que sim, pois já era preciso ter “azar” irmos obter um dos poucos intervalos que não contêm p.

Dr. Gentil Alves – Muito bem. Compreendi o que disseste, mas então porque é que não construo intervalos com, por exemplo, 99% de confiança? Assim, só 1% dos intervalos possíveis de construir é que não conteriam o p, não é verdade?

Prof. Amável – Muito bem observado! Mas nunca ouviste dizer que “sem ovos não se fazem omeletes” ou “que não há almoços grátis”? Pois é! A contrapartida para, com a mesma dimensão da amostra, termos intervalos de 99% de confiança, é que a margem de erro vem maior, isto é, vamos ter intervalos com maior amplitude, o que significa uma menor precisão. Em último caso construiríamos intervalos com uma confiança de 100%! Sabes ao que chegávamos? A R! Não tens nenhuma dúvida de que o intervalo está em R, pois não? Não nos adianta é nada! Já agora, com o valor de 56% obtido na amostra que a Sondagem recolheu, um intervalo de 99% de confiança seria (51.4%; 60.6%). Assim, enquanto que com o primeiro intervalo temos uma margem de erro de 3.5%, agora a margem de erro passou para 4.5%. Ficaste esclarecido?

Dr. Gentil Alves – Penso que sim. Só mais uma questão. Haveria algum processo de, com a confiança de 99%, obter um intervalo com a margem de erro que obtive para o intervalo de 95% de confiança?

Prof. Amável – Mais uma vez estás a colocar uma questão interessante. Efectivamente, podemos, mantendo a confiança, diminuir a margem de erro, agora à custa de recolhermos uma amostra de maior dimensão. Nada se faz sem custos, como estás a ver. Por exemplo, admitindo que a percentagem de lisboetas, que pensam votar em ti, não se alteraria muito se se recolhesse uma amostra de maior dimensão, então teria de ser recolhida uma amostra de 1335 lisboetas, em vez de 785 (estou a considerar que a proprção de votos a teu favor, obtida ao questionar os 1335 lisboetas, é aproximadamente igual a 56%).

Dr. Gentil Alves – Muito obrigada por estes esclarecimentos. Vou mesmo avançar com a minha candidatura.

Passados 8 dias realizaram-se as eleições. O Dr. Gentil Alves é o novo presidente da Câmara de Lisboa.

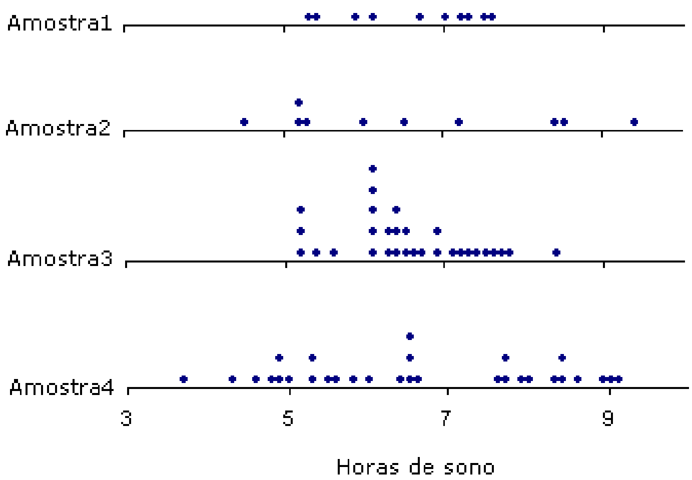
Exercícios

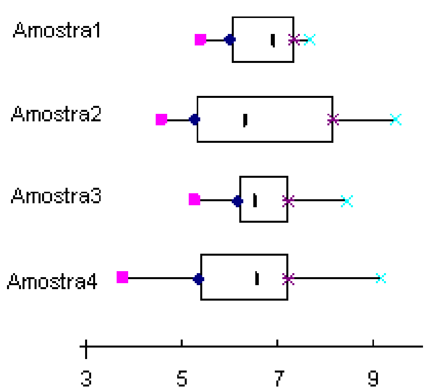
3.1 – Para cada uma das seguintes situações diga se o parâmetro de interesse é um valor médio ou uma proporção:

a) Numa sondagem perguntou-se a cada um de 100 estudantes quantas horas por semana, gastavam a ver televisão.
b) Numa sondagem perguntou-se a cada um de 100 estudantes se passavam mais de 8 horas por semana, a ver televisão.
c) Numa sondagem, perguntou-se a 100 agregados familiares qual a percentagem do orçamento familiar que era gasto com a renda da casa.
d) Num estudo sobre o consume de bebidas alcoólicas, investigou-se junto de 50 restaurantes, qual a percentagem de bebidas alcoólicas, de entre as bebidas consumidas por semana.
e) Junto dos mesmos restaurantes da alínea anterior, verificou-se que 35% dos restaurantes vendiam semanalmente mais bebidas alcoólicas, que não alcoólicas.

3.2 – Num Censo, em que a dimensão da amostra é igual à dimensão da população, o erro padrão da média (ou da proporção amostral) é igual a zero. Explique porquê.

3.3 – Suponha (Adaptado de Rossman, 2001) que pretende conhecer o tempo médio de sono que os alunos da sua escola dormiram, na última noite. Considere os seguintes diagramas que apresentam os tempos de sono de alunos da escola, referentes a 4 amostras:





a) As seguintes estatísticas descritivas foram calculadas com base nas amostras anteriores. Complete a tabela.

Amostra nº	Dimensão da amostra	Média	Desvio padrão amostral
	30	6.6	0.82
	10	6.6	0.82
	10	6.6	1.59
	30	6.6	1.59

b) O que é que todas as amostras têm em comum?
c) Qual a característica que sobressai quando comparamos as distribuições correspondentes às amostras 1 e 2?
d) Qual a característica que sobressai quando comparamos as distribuições correspondentes às amostras 1 e 3?

Na seguinte tabela apresentamos os intervalos de confiança para as amostras de dimensão 30 (consegue obter os intervalos de confiança correspondentes às amostras de dimensão 10?):

Amostra nº	Dimensão da amostra	Média	Desvio padrão amostral	Int. confiança
	30	6.6	0.82	(6.31; 6.89)
	10	6.6	0.82	-
	10	6.6	1.59	-
	30	6.6	1.59	(6.03; 7.17)

Qual das 2 amostras produz uma estimativa para o tempo médio de sono, mais precisa? Qual a influência da variabilidade apresentada pela amostra, para a amplitude do intervalo de confiança?

3.4 – Suponha que na sua escola, cada um dos 50 alunos de Matemática para as Ciências Sociais, foi encarregue de recolher informação junto de 10 adultos, se eram a favor do referendo da Constituição Europeia. O histograma construído com as 50 proporções obtidas terá um aspecto que faz lembrar o modelo normal? Justifique.

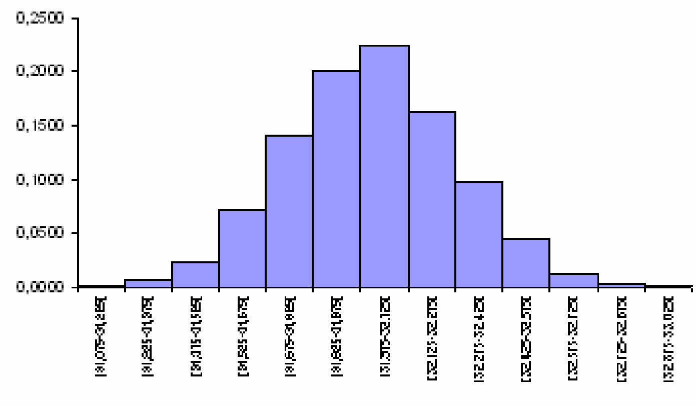
3.5 – Suponha que na sua escola, cada um dos 45 alunos de Matemática A. Foi encarregue de recolher informação, junto de 30 alunos de outras escolas, se eram a favor do "Novo Estatuto para o Aluno". O histograma construído com as 45 proporções obtidas terá um aspecto que faz lembrar o modelo normal? Justifique.

3.6 – Na correcção de certo tipo de exames, feitos a nível nacional, em que cada exame é constituído por uma parte fechada e uma parte aberta, utiliza-se um leitor óptico para corrigir a parte fechada. Cada exame tem 50 questões, e a probabilidade de a máquina ler erradamente uma destas questões é p, a qual é constante de questão para questão e de exame para exame. Desconhece-se este valor de p.

a) Admitindo que em 10 destes exames, a máquina leu erradamente 15 questões, obtenha uma estimativa pontual para p.
b) Utilizando o resultado da alínea anterior:
   i) Obtenha um intervalo, com uma confiança de 95%, para p;
   ii) Qual a margem de erro do intervalo que obteve?
c) A empresa que vende as máquinas de leitura óptica diz que a percentagem de erros que a máquina comete, anda à volta de 1%. Tendo em conta o intervalo de confiança obtido na alínea anterior, pensa que a empresa tem razão no que afirma? Justifique a sua resposta. (Se na alínea anterior não conseguiu determinar o intervalo de confiança pretendido, admita o seguinte intervalo (1.5%; 4.5%)).

3.7 - Uma fábrica de calçado para adultos, pretende começar a produzir sapatos para criança. Encarregou uma empresa de sondagens, de lhe fazer um estudo sobre qual seria o tamanho médio (em cm) do pé de crianças de determinada classe etária. Mesmo antes da empresa apresentar as conclusões, o dono da fábrica (que há muitos anos tinha tido uma disciplina de Estatística) teve acesso à seguinte tabela de frequências e correspondente histograma, dos valores calculados para as médias de 500 amostras, de dimensão 30, recolhidas pela empresa:

Classes	**Freq.rel.**
[31,075-31,225[	0,0020
[31,225-31,375[	0,0075
[31,375-31,525[	0,0250
[31,525-31,675[	0,0735
[31,675-31,825[	0,1410
[31,825-31,975[	0,2005
[31,975-32,125[	0,2250
[32,125-32,275[	0,1635
[32,275-32,425[	0,0990
[32,425-32,575[	0,0445
[32,575-32,725[	0,0130
[32,725-32,875[	0,0040
[32,875-33,025[	0,001 5



Então, na posse destes elementos, pediu ao filho, que tinha frequentado a disciplina de MACS do 11º ano, que lhe respondesse às seguintes questões:

a) Este histograma pretende representar a distribuição de amostragem, aproximada, de uma certa variável. Que variável?

b) Utilizando a tabela anterior, obtenha um valor aproximado para o valor médio da distribuição de amostragem da Média, para amostras de dimensão 30 (considere o valor aproximado às unidades).

c) Tendo em consideração que a estatística Média $\overline{X}$, é um estimador centrado do valor médio da população X, de onde se retiram as amostras, sugira um valor para o valor médio μ, da população X, constituída pelo tamanho do pé, das crianças da classe etária considerada.

d) Sabendo que o desvio padrão de $\overline{X}$, é igual a $\frac{\sigma}{\sqrt{30}}$, onde σ é o desvio padrão da população X, utilize a tabela dada para sugerir um valor para este desvio padrão σ.

e) Como o histograma anterior sugere, e o Teorema Limite Central justifica, a distribuição de amostragem da Média pode ser aproximada por uma distribuição Normal (para amostras de dimensão n, suficientemente grande, ou seja, $n \geq 30$). Admitindo que um dos valores obtidos para a média de uma das 500 amostras de dimensão 30 consideradas, foi 32.125, obtenha um intervalo de 95% de confiança para o valor médio do comprimento do pé. (Se na alínea d) não conseguiu determinar o valor de σ, admita que é igual a 1.5).

f) Admitindo que a população X tem distribuição normal, com o valor médio e desvio padrão obtidos, respectivamente, nas alíneas c) e e), calcule a probabilidade de uma criança, escolhida ao acaso, da classe etária em estudo, ter um comprimento do pé superior a 32.5 cm. (Se não resolveu as alíneas c) e e) considere os valores 32 cm e 1.5 cm, respectivamente para valor médio e desvio padrão de X.

3.8 – Nas últimas eleições legislativas, passada uma hora do fecho das mesas de voto, apareceram os resultados para o concelho de Sintra, dando uma percentagem de votos para JS e FS, respectivamente de 39% e 42%, com uma margem de erro de 3.5% e uma confiança de 95%.

a) O locutor afirmou, ao apresentar aqueles resultados, que os candidatos estavam empatados tecnicamente. Explique, por palavras suas, o que quereria o locutor dizer.
b) Passadas duas horas a margem de erro, diminuiu para 2.5%. Admitindo que a confiança era a mesma, dê uma explicação para a diminuição da margem de erro.
c) Numa sondagem realizada antes das eleições, JS tinha encomendado uma sondagem, que lhe dava a vitória, quando afinal veio a perder as eleições. Teremos que deixar de acreditar nas sondagens?

3.9 – Uma sondagem da TSF/DN publicada na edição do DN de 2 de Julho de 2004, dizia:

Portugueses querem referendo

Maioria mostra-se favorável à eleição de um presidente e de um governo da União Europeia. E também quer exército comum

Os portugueses manifestam tendência para o federalismo europeu: a maioria defende um presidente e um governo europeus, eleitos pelos cidadãos. São igualmente favoráveis à criação de um exército da União Europeia (UE). E, na análise que fazem sobre o futuro comunitário, dizem ainda que querem referendar a próxima reforma institucional da UE. A maioria já ouviu falar do Tratado de Nice, mas está longe de saber o que ele contempla. Talvez por isso, a larga maioria não sabe se o documento deve ou não ser aprovado pelos deputados.

O Barómetro de Junho do DN/TSF/Marktest não incluiu qualquer pergunta directa sobre o federalismo europeu, mas os portugueses acabaram por pronunciar-se nesse sentido. Senão vejamos: 62 por cento dos inquiridos mostrou-se favorável à eleição de um presidente da UE e 53 por cento disse também estar a favor de um governo europeu. É uma tese defendida equitativamente por mulheres e homens no que diz respeito à eleição de um presidente europeu.

Nota-se, contudo, alguma diferença quando a questão é a eleição de um governo europeu. Aqui, já são os homens que se mostram mais favoráveis. Sobre um e outro assunto é, claramente, a classe média a maior defensora de um executivo europeu.

Quando questionados sobre a criação de um exército na UE, uma questão que até aqui tem levantado alguma polémica, 45 por cento dos inquiridos afirmam ser defensores desta ideia. Embora o número daqueles que se opõem não seja muito inferior - 36 por cento. Significativa é também a percentagem dos que não sabem o que responder - 19 por cento. Esta hipótese acolhe mais adeptos entre os entrevistados do sexo masculino (53 por cento) e na faixa etária que poderá ser contemplada pelas incorporações (igualmente 53 por cento). E se a maioria dos portugueses refere já ter ouvido falar do Tratado de Nice, também são peremptórios a afirmar que não fazem a mais pequena ideia das suas linhas gerais: 65 por cento sublinha que não sabe o que está consagrado no documento. Uma resposta que justifica a elevada percentagem (62 por cento) daqueles que não sabe se os deputados devem ou não aprovar o Tratado.

A larga maioria dos inquiridos (60 por cento) defende, por outro lado, que as mudanças na organização da União Europeia devem ser referendadas no nosso País. O que não deixa de ser curioso, já que as duas experiências anteriores (aborto e regiões) revelaram uma grande falta de participação dos cidadãos. Só 18 por cento tem opinião contrária e 22 por cento optou por não responder a esta questão. O alargamento da União Europeia aos países do Centro e de Leste do continente merece o acordo da maioria (64 por cento), que se mostram convencidos de que essa reestruturação

interna vai tirar poderes a Portugal no seio da UE (46 por cento). Mais de dois terços (67 por cento) considera também que o processo de alargamento poderá reduzir a atribuição de fundos comunitários para Portugal.
Embora não seja referido no artigo anterior, segundo a notícia da TSF, a sondagem envolveu **813** indivíduos adultos, dos quais 421 eram mulheres e foi realizada via telefone. É referido no artigo que **62**% dos **inquiridos** se mostra favorável à eleição de um presidente da UE.

a) Este valor de 62% é uma *estatística* ou um *parâmetro*?
b) Seria possível ter obtido este valor, se a percentagem de portugueses adultos que se mostra favorável à eleição de um presidente da UE fosse 65%?
c) Tendo em conta o resultado obtido pela sondagem da TSF/DN, acha plausível que a proporção de portugueses que se mostra favorável à eleição de um presidente da UE seja 68%? Porquê?

3.10 – No dia x do mês y do ano z realizar-se-ão as Eleições Autárquicas. Relativamente à cidade de Lisboa, há dois candidatos sobre os quais se criaram mais expectativas, nomeadamente TT e MM. Suponha que, no dia das eleições, passado uma hora sobre o fecho das urnas, altura em que começam a contar os votos para cada candidato, surgiram os primeiros resultados nos canais televisivos. Relativamente a um daqueles candidatos, o candidato TT, apresentaram o seguinte resultado: - *O candidato TT tem, neste momento, uma percentagem de 48.4%, com um erro máximo de 3.45% e uma confiança de 95%.*
1. Explique, por palavras suas, o que significa o resultado anterior.
2. Qual a amplitude do intervalo de confiança, que pode construir com os resultados apresentados no enunciado do problema, para a percentagem de lisboetas que votaram no candidato TT?
3. Acha razoável admitir que o candidato TT, ao ouvir aquele resultado, pense que tem alguma "Chance" de ganhar a Câmara de Lisboa, admitindo que para ganhar essa Câmara eram necessários, pelo menos, 50% de votos favoráveis?
4. Passadas três horas do fecho das urnas, o resultado anunciado para o candidato TT era: - *O candidato TT tem, neste momento, uma percentagem de 49.8%, com um erro máximo de 1.23% e uma confiança de 95%.*
a) Compare a amplitude do intervalo de confiança considerado na alínea 2, com a amplitude do intervalo de confiança, que pode construir com os resultados agora anunciados.
b) Como é que interpreta o resultado a que chegou na alínea anterior?
5. Quando todos os votos tiverem sido escrutinados, obtém o resultado para a percentagem de eleitores que votaram no candidato TT, na forma de um intervalo de confiança, ou na forma de um valor? Explique porquê.

3.11 - Numa altura em que se discutia o problema dos touros de morte, em Portugal, nomeadamente por causa das festas de Barrancos, uma conhecida estação de televisão propôs a seguinte questão aos telespectadores, no final do telejornal de uma 6ª feira:

- Se é a favor dos touros de morte, em Portugal, envie uma mensagem para 7771
- Se é contra os touros de morte, em Portugal, envie uma mensagem para 7772

No telejornal do dia seguinte, sábado, apresentaram a seguinte notícia, como sendo o resultado da sondagem efectuada: *72% dos portugueses são a **favor** dos touros de morte, em Portugal, enquanto que 28% são **contra***!
Acontece que o jornal Expresso, desse sábado, publicou o seguinte resultado de uma sondagem, encomendada a uma conceituada empresa de sondagens: *81% dos portugueses são **contra** os touros de morte, em Portugal*!

1. Alguma das amostras consideradas para obter os resultados anteriores, pode ser considerada enviesada? Isso poderá explicar a discrepância obtida, nas duas sondagens, relativamente às percentagens obtidas para os portugueses, que são contra os touros de morte?
2. Qual dos resultados anteriores, 28% ou 81%, estará mais perto da percentagem de portugueses que são contra os touros de morte em Portugal? Explique porquê
3. Admitindo que o resultado obtido pela empresa de sondagens, foi baseado numa amostra aleatória de dimensão 150, obtenha um intervalo de 95% de confiança para a percentagem de portugueses que são contra os touros de morte, em Portugal.
4. Calcule a margem de erro do intervalo obtido anteriormente. O que é que aconselharia a alguém, que lhe perguntasse como poderia obter um intervalo de confiança, com uma margem de erro inferior?

3.12 – O Sr. Silva, fabricante de camisas para homem, recebeu uma encomenda proveniente de Macau. Ficou um pouco preocupado, pois quando visitou este território, na sua viagem de lua-de-mel, apercebeu-se que os homens tinham, de um modo geral, os braços mais curtos. Sendo assim, não poderia utilizar os moldes habituais. Pediu, então, a uma empresa de sondagens que lhe fornecessem uma estimativa do comprimento médio dos braços dos naturais de Macau. A empresa apresentou um estudo, que se pode resumir da seguinte forma:
Sr. Silva
Apresentando os nossos cumprimentos, vimos apresentar os resultados do nosso estudo: recolhemos uma amostra de dimensão 70, de outros tantos indivíduos adultos, do sexo masculino, a quem medimos o tamanho do braço, tendo obtido como média dos 70 valores observados, o valor 52 cm.
Reiterando os nossos cumprimentos, aproveitamos para dizer que segue, em anexo, a factura do trabalho prestado.
Atenciosamente o gerente (assinatura irreconhecível)
O Sr. Silva ficou um pouco menos preocupado, mas continuava sem saber o que fazer:

1. Efectivamente, qual a confiança que poderia atribuir à estimativa obtida? Se tivesse sido outra a amostra obtida, seria de esperar obter o mesmo valor para a média? Explique porquê.
2. O Sr. Silva resolveu questionar a empresa e esta forneceu-lhe os seguintes intervalos de confiança para o tamanho médio do braço dos naturais de Macau, com uma confiança de 50% e 75%, respectivamente, e obtidos a partir da mesma amostra: [51.4, 52.6] e [51.0, 53.0].
    a. Qual a margem de erro dos intervalos anteriores?
    b. Se fosse o Sr. Silva, qual o intervalo que escolhia? O de menor amplitude ou o de maior amplitude? Explique porquê.

3.13 – Lançou-se uma moeda 50 vezes e saiu cara 20 vezes. Tem a certeza que a moeda não é equilibrada? Justifique. Sugestão: Construa um intervalo de 95% de confiança para a probabilidade de sair cara.

3.14 – Pretende-se determinar um intervalo de confiança para a proporção p, de peças defeituosas produzidas por determinada máquina. Pretende-se que a precisão seja grande, pelo que não queremos que a margem de erro seja superior a 3%. Tem-se a informação de que a máquina em questão costuma produzir cerca de 2% de peças defeituosas, mas não se tem a garantia que este valor não esteja um pouco alterado. Qual a dimensão da amostra que deve recolher para construir o intervalo pretendido?

3.15 (continuação de 3.14)– Considere de novo o exercício 3.14, mas admita que a percentagem de peças defeituosas que a máquina costuma produzir, anda à volta de 20%. Qual a dimensão da amostra que se deve recolher?

3.16 (continuação de 3.14 e 3.15) – Admita agora que não tinha qualquer informação sobre a percentagem de peças defeituosas produzidas pela máquina. Qual a dimensão da amostra que teria de recolher? Compare o valor obtido para a dimensão da amostra com os valores obtidos em 3.14 e 3.15. Comente os resultados obtidos.

3.18 – Suponha que num curso com 350 raparigas e 150 rapazes quer seleccionar uma amostra de 50 alunos.
a) Pode-se considerar que a população de onde está a seleccionar a amostra, tem dimensão suficientemente grande para poder ser considerada uma população infinita?
b) Considere a amostragem com reposição. Calcule o valor médio e o desvio padrão do estimador da proporção amostral de raparigas. Pode continuar a utilizar o mesmo valor médio e/ou o mesmo desvio padrão para o estimador da proporção amostral de raparigas, se a amostragem for feita sem reposição?
c) Qual a distribuição de amostragem, aproximada, da proporção de raparigas numa amostra de dimensão 50?
d) Qual a probabilidade, aproximada, do número de raparigas na amostra, estar entre 25 e 35?
e) Se quisesse garantir na amostra, com uma probabilidade de 0.5, 25 raparigas, quantos alunos devia seleccionar?

Lista de algumas funções usadas no Excel:

Inglês	**Português**	
And()	*E()*	Devolve verdadeiro se todos os argumentos forem verdadeiros e devolve falso se algum dos argumentos for falso
Average()	*Media()*	Calcula a média dos valores existentes num conjunto de células
Count()	*Contar()*	Conta as células com valores numéricos, incluindo datas e fórmulas cujos resultados são números
Counta()	*Contar.val()*	Conta todas as células não vazias
Countblank()	*Contar.vazio()*	Conta as células vazias
Countif()	*Contar.se()*	Conta as ocorrências verificadas num conjunto de célula, que obedecem a um critério
If()	*Se()*	Executa uma de duas acções possíveis, em função do resultado da condição
Int()	*Int()*	Devolve a parte inteira de um número
Max()	*Maximo()*	Devolve o maior valor de um conjunto de células
Min()	*Minimo()*	Devolve o menor valor de um conjunto de células
Mod()	*Resto()*	Devolve o resto de uma divisão
Or()	*Ou()*	Devolve verdadeiro se um dos argumentos for verdadeiro e devolve falso se todos os argumentos forem falsos
Product()	*Produto()*	Multiplica os valores de um conjunto de células, ignorando as células vazias e/ou com texto
Rand()	*Aleatório()*	Devolve um número pseudo-aleatório (no intervalo (0,1))
Randbetween()	*Aleatórioentre()*	Devolve um número pseudo-aleatório no intervalo especificado
Round()	*Arred()*	Devolve um número arredondado, na posição indicada
Rounddown()	*Arred.para.baixo()*	Devolve um número arredondado, por defeito, na posição indicada
Roundup()	*Arred.para.cima()*	Devolve um número arredondado, por excesso, na posição indicada
Sum()	*Soma()*	Soma os valores de um conjunto de células
Sumif()	*Soma.se()*	Soma as ocorrências verificadas num conjunto de células que obedecem a um critério
Sumproduct()	*Somarproduto()*	Multiplica dois conjuntos de células e devolve a soma total dos produtos
Vlookup()	*Procv()*	Procura um valor na coluna mais à esquerda de uma tabela e devolve um valor na mesma linha na coluna indicada

Bibliografia

Graça Martins, M. E. (2005) – *Introdução à Probabilidade e à Estatística*. Sociedade Portuguesa de Estatística.

Graça Martins, M.E. et al (2001) – *Estatística* – 10º ano de escolaridade, Edição do Ministério da Educação – Departamento do Ensino Secundário.

Graça Martins. M. E. and al (1999) – *Introdução às Probabilidades e à Estatística*. Universidade Aberta.

Graça Martins, M.E. e Loura, M. E. (2001) – *Matemática para as Ciências Sociais* – Anexo para apoio à interpretação do programa.

Moore, D. (1977) – *Statistics. Concepts and Controversies*. W.H. Freeman and Company, New York.

Moore, D. and al (1992) – Perspectives in Contemporary Statistics. The Mathematical Association of America.

Moore, D. and al (1993) – *Introduction to the Practice of Statistics*. W.H. Freeman and Company, New York.

Murteira, B. (1993) – *Análise Exploratória de Dados. Estatística Descritiva*. McGraw-Hill de Portugal.

Rossman, A. and al (2001) – *Workshop Statistics. Discovery with data*. Key College Publishing/Spinger-Verlag. New York, Inc.

Tannenbaum, P. and al (1998) – *Excursions in Modern Mathematics*. Prentice Hall.

Velleman, P. and al (2004) – *Intro Stats*. Pearson Education, Inc.

Artigos da revista TEACHING STATISTICS

AGEEL, M.I. – Spreadsheets as a Simulation Tool for Solving Probability Problems, Vol 24, 2, 51-54.
Hodgson, T., and Borkowski, J. - Why Stratify? Vol 20, 1, 68-71.
NEVILLE, H. – Handling Continuous Data in Excel, Vol 25, 2, 42-45.
NEVILLE, H. – Charts in Excel, Vol 26, 2, 49-53.

Páginas na Internet

INSTITUTO NACIONAL DE ESTATÍSTICA E ESCOLA SECUNDÁRIA TOMAZ PELAYO
PROJECTO ALEA – http://www.alea.pt

(Desta página recomenda-se, em especial, o dossier didáctico "ESTATÍSTICA COM EXCEL" da autoria de Luís Miguel Cunha).

INSTITUTO NACIONAL DE ESTATÍSTICA – www.ine.pt/

Tem informação sobre Portugal, ao nível da freguesia.

EUROSTAT – europa.eu.int/comm/eurostat/

Tem informação relativa aos diversos países da Europa.