



## N.º 30 – ESTIMAR A PERCENTAGEM DE...

Por: Maria Eugénia Graça Martins  
Faculdade de Ciências da Universidade de Lisboa  
[memartins@fc.ul.pt](mailto:memartins@fc.ul.pt)  
Emília Oliveira  
Agrupamento de Escolas Tomaz Pelayo  
[ecmo.estp@gmail.com](mailto:ecmo.estp@gmail.com)

### ALUNOS DO 1.º CICLO QUE TÊM TELEMÓVEL

O Pedro é um aluno do 1.º ciclo e pediu à mãe um telemóvel, pois a maioria dos seus colegas do 1.º ciclo já tem telemóvel!  
Será que o Pedro tem razão? Como estimar a percentagem de alunos do 1.º ciclo que têm telemóvel?

### ALUNOS DO AGRUPAMENTO DE ESCOLAS *ESTUDA BEM*, QUE PENSA VOTAR NA LISTA *PROMETO E CUMPRO*



No agrupamento de escolas *Estuda Bem*, a lista *Prometo e Cumpro* pensa candidatar-se à Associação de Estudantes, pois tem recebido muitas manifestações de apoio!

A um mês das eleições, os estudantes que integram a lista *Prometo e Cumpro* gostariam de ter uma ideia da percentagem de alunos que pensam votar na sua lista. Como proceder, uma vez que o agrupamento tem mais de 4000 alunos?

O que têm as situações anteriores em comum? Em qualquer uma delas, é pressuposto fazer uma sondagem a uma população, em que se procura estimar uma quantidade numérica (desconhecida) dessa população. Em ambos os casos, procura-se estimar uma percentagem.

Vamos começar por estudar o segundo caso e de seguida abordaremos o primeiro, que serve de exemplo para situações em que é necessário fazer uma sondagem em populações cujos elementos estão listados, situação frequente, como por exemplo quando se consideram as populações constituídas pelos elementos de ordens profissionais, os elementos da lista telefónica, etc...



obtendo-se o intervalo  $[50,2\%;79,8\%]$ , que, com uma confiança de 95%, contém o valor da proporção **p** de alunos do agrupamento de escolas que pensam votar na lista *Prometo e Cumpro*.

#### 4. Conclusão

Efetivamente, os resultados indiciam que a lista *Prometo e Cumpro* terá a maioria, já que a percentagem de 50% fica aquém do intervalo de confiança.

**Nota** - Como obter um intervalo de confiança com uma maior precisão, ou seja, com menor amplitude? Se pretendermos manter a confiança de 95%, teremos de aumentar a dimensão da amostra a recolher para que a amplitude do intervalo seja mais pequena.

(Para explicação mais detalhada sobre a dimensão da amostra, consultar o Curso de Inferência Estatística do ALEA [www.alea.pt/html/statofic/html/dossier/doc/Modulo3-Int\\_InferenciaEstatisticaFinal.pdf](http://www.alea.pt/html/statofic/html/dossier/doc/Modulo3-Int_InferenciaEstatisticaFinal.pdf), página 75).

### Qual a percentagem de alunos do 1.º ciclo que tem telemóvel?

Pretende-se fazer uma sondagem com o objetivo de averiguar qual a percentagem de alunos do 1.º ciclo que tem telemóvel.

#### 1. Definição da população, variável e parâmetro a estimar

Vamos considerar que a **população** é constituída pelos 23054 alunos do 1.º ciclo que responderam ao inquérito do Minicensos – ficheiro *RespostasMiniCensosNovo*<sup>3</sup>. Temos uma lista, em Excel, com algumas das características destes alunos, os quais estão identificados unicamente por um número, na coluna ID. No caso presente, a **variável** em estudo é “Ter telemóvel”. É uma variável de tipo qualitativo, que assume as categorias *Sim* e *Não*, que vamos representar, respetivamente por 1 e 0. Pretende-se, a partir de uma amostra selecionada da população, estimar o parâmetro “percentagem de alunos” que tem telemóvel.

#### 2. Como selecionar uma amostra a partir de uma lista<sup>4</sup>

Vamos utilizar o processo descrito na página 18 do Dossiê XIIIa – Estatística descritiva com Excel. Assim, considera-se uma folha de Excel em que na Coluna A colocamos as identificações dos elementos da nossa população – coluna *Id* do ficheiro *RespostasMiniCensosNovo*. Chamamos a atenção para o facto de a situação que estamos a considerar ser especial, na medida em que temos a resposta da variável em estudo para todos os elementos da população, o que numa situação real não se verifica! Neste caso, não necessitaríamos de recolher uma amostra para estimar a percentagem de alunos com telemóvel, pelo que podemos obter o verdadeiro valor dessa percentagem. Contudo, este conhecimento vai permitir verificar em que medida é que o valor estimado está

<sup>3</sup> Consultar este ficheiro na área dos dossiês e recursos do ALEA, Dossiê XVI- Mini-Censos 2011.

<sup>4</sup> O processo que se vai descrever pode ser utilizado para obter uma amostra aleatória simples a partir de uma listagem de números de telefone. As pessoas cujos telefones forem selecionados serão objeto de contacto da empresa encarregue do estudo.

próximo do verdadeiro valor. Assim, à coluna A acrescentamos uma nova coluna, coluna B, com os dados da variável em estudo (coluna *telemóvel* do ficheiro *RespostasMiniCensosNovo*). Seguindo o processo descrito no dossiê, vamos inserir a função `rand()` na coluna C. Obteremos um ficheiro idêntico ao seguinte:

	A	B	C
1	Id	telemovel	Função rand()
2	1	0	0,44099848
3	2	1	0,10707905
4	3	1	0,27202996
5	4	0	0,86554584
6	5	0	0,41837630
7	6	1	0,79894893
8	7	1	0,81978242
9	8	0	0,13586401
10	9	1	0,57151122
11	10	0	0,33732302
12	11	1	0,28480708
13	12	1	0,76482621
14	...	...	...

Procedemos de seguida como indicado no dossiê, isto é, colamos os valores devolvidos pela função `rand()` na coluna D. Ordenamos o ficheiro obtido, utilizando como critério esta coluna.

	A	B	C	D
1	Id	Telem.	Função rand()	
2	2419	0	0,158157080	4,84E-05
3	16435	0	0,931208700	7E-05
4	2912	0	0,301699186	0,000171
5	6693	0	0,668546604	0,00018
6	15004	0	0,883300991	0,00018
7	13689	0	0,392135861	0,000187
8	9447	1	0,372289288	0,000201
9	7448	0	0,823884238	0,000209
10	18996	1	0,450931133	0,000224
11	9692	1	0,403735576	0,000252
12	5219	1	0,573187333	0,000279
13	20248	1	0,163342416	0,000299
14	...	...	...	...

Selecionamos tantos elementos da coluna B, quantos os desejados para a dimensão da amostra. Admitindo que se pretende uma amostra de dimensão 100, selecionamos os 100 primeiros elementos da coluna B, que constituirão as 100 respostas desejadas. Embora não fosse necessário, apresentamos a seguir os números dos elementos da população que foram selecionados e as respetivas respostas:

	K	L	M	N	O	P	Q	R
1	Id	Telem.	Id	Telem.	Id	Telem.	Id	Telem.
2	2419	0	14287	0	638	1	1610	1
3	16435	0	13816	1	15602	0	10599	0
4	2912	0	22453	0	20246	0	2598	1
5	6693	0	2207	1	18537	1	19136	0
6	15004	0	9936	0	22815	1	18665	1
7	13689	0	9914	0	22632	0	23231	0
8	9447	1	19908	1	17974	1	19744	0
9	7448	0	11442	0	8082	1	9492	1
10	18996	1	2440	1	21570	0	7154	1
11	9692	1	3111	1	1236	1	7852	0
12	5219	1	22774	1	2143	0	5959	0
13	20248	1	13923	0	17010	0	3465	1
14	22831	0	5401	0	17476	0	19394	0
15	12577	1	17966	0	9446	0	7940	0
16	18438	0	981	0	20613	1	4579	0
17	2268	1	3779	0	11375	1	15622	0
18	4169	0	15897	0	13051	1	6346	0
19	9239	0	19786	0	23181	0	20147	1
20	7515	0	1768	1	22633	1	22405	0
21	2714	0	14551	0	8261	0	13898	0
22	10093	0	15961	0	8645	1	17563	1
23	15597	0	1948	1	5099	1	2938	0
24	19808	1	238	0	14198	0	6638	0
25	11321	0	20433	0	10911	0	14393	0
26	19734	0	252	1	3594	0	18118	1

A percentagem de alunos da amostra que tem telemóvel é igual a 38%.

### 3. Como obter um intervalo de confiança

Para obter o intervalo de 95% de confiança para a percentagem de alunos do 1.º ciclo com telemóvel, o procedimento é idêntico ao considerado no caso anterior, de modo que se obtém o seguinte intervalo [28,5%;47,5%].

### 4. Conclusão

O Pedro não tem razão! Com uma confiança de 95%, pode argumentar-se que o intervalo anterior contém a verdadeira proporção de alunos do 1.º ciclo com telemóvel, que nem chega a conter a proporção de 50%!

**Nota** – Uma vez que, como referimos atrás, temos as respostas de todos os elementos da população, isto é, temos os dados relativos a todos os alunos sobre se têm ou não telemóvel, temos a possibilidade de obter o verdadeiro valor do parâmetro que procurámos estimar e que é igual a 37,9%.