



N.º 22 – AMOSTRA ALEATÓRIA

Por: Maria Eugénia Graça Martins
Departamento de Estatística e Investigação Operacional da FCUL
memartins@fc.ul.pt

Emília Oliveira
Escola Secundária de Tomaz Pelayo
ecmo.estp@gmail.com

Nesta ActivALEA apresentamos uma forma simples de utilizar o Excel para seleccionar uma amostra aleatória simples¹ e uma amostra aleatória com reposição², de uma População finita, de que se tenha uma listagem dos elementos.

A lista de escolas secundárias de Portugal Continental utilizada como exemplo nesta ActivALEA está disponível no ficheiro Excel que acompanha esta ficha³.

¹ **Amostra aleatória simples** - Dada uma população de dimensão N , uma amostra aleatória simples de dimensão n é um conjunto de n unidades da população, tal que qualquer outro conjunto dos $\binom{N}{n}$ conjuntos diferentes de n unidades, teria igual probabilidade de ser seleccionado.

² **Amostra aleatória com reposição** - Dada uma população de dimensão N , referir-nos-emos a uma amostra aleatória de dimensão n , com reposição, como um conjunto de n unidades da população, tal que qualquer outro conjunto dos N^n conjuntos diferentes de n unidades, teria igual probabilidade de ser seleccionado.

Os processos de selecção de uma amostra aleatória simples ou de uma amostra aleatória com reposição distinguem-se, fundamentalmente, pelo facto de no primeiro caso os elementos serem seleccionados sequencialmente ou todos de uma vez, enquanto que no segundo caso sempre que um elemento é seleccionado, é repostado antes de se seleccionar o próximo.

³ **Nota:** Caso pretenda informação mais detalhada e observar outros exemplos, sugerimos uma consulta ao capítulo 1 do Dossiê XIII – “Estatística Descritiva com Excel – Complementos”, disponível em http://www.alea.pt/html/statofic/html/dossier/html/meio_dossier13.html.



Considere a seguinte lista de Escolas Secundárias de Portugal continental. Utilizando o Excel, extraia uma amostra aleatória simples de 10 escolas.

Nome	Distrito	Concelho
Escola Secundária Alves Martins	Viseu	Viseu
Escola Secundária Amélia Rey Colaço	Lisboa	Oeiras
Escola Secundária com 3º ciclo do ensino básico José Afonso	Lisboa	Loures
Escola Secundária D. Afonso Henriques	Porto	Santo Tirso
Escola Secundária Vergílio Ferreira	Lisboa	Lisboa
Escola Secundária da Lourinhã	Lisboa	Lourinhã
Escola Secundária da Moita	Setúbal	Moita
Escola Secundária da Sertã	Castelo Branco	Sertã
Escola Secundária David Mourão Ferreira	Lisboa	Lisboa
Escola Secundária de Albufeira	Faro	Albufeira
Escola Secundária de Alves Redol	Lisboa	Vila Franca de Xira
Escola Secundária de Arganil	Coimbra	Arganil
Escola Secundária de Avelar Brotero	Coimbra	Coimbra
Escola Secundária de Benavente	Santarém	Benavente
Escola Secundária de Cantanhede	Coimbra	Cantanhede
Escola Secundária de Cascais	Lisboa	Cascais
Escola Secundária de Santa Maria da Feira	Aveiro	Santa Maria da Feira
Escola Secundária de D. Duarte	Coimbra	Coimbra
Escola Secundária de Alberto Sampaio	Braga	Braga
Escola Secundária de D. Pedro V	Lisboa	Lisboa
Escola Secundária de D. Sancho II	Portalegre	Elvas
Escola Secundária de D. Manuel I	Beja	Beja
Escola Secundária de Dom Manuel Martins	Setúbal	Setúbal
Escola Secundária de Domingos Sequeira	Leiria	Leiria
Escola Secundária de Francisco Rodrigues Lobo	Leiria	Leiria
Escola Secundária de Gabriel Pereira	Évora	Évora
Escola Secundária de Gago Coutinho	Lisboa	Vila Franca de Xira
Escola Secundária de Gil Eanes	Faro	Lagos
Escola Secundária de Homem Cristo	Aveiro	Aveiro
Escola Secundária de Jaime Cortesão	Coimbra	Coimbra
Escola Secundária de João de Deus	Faro	Faro
Escola Secundária de José Falcão	Coimbra	Coimbra
Escola Secundária de Júlio Dantas	Faro	Lagos
Escola Secundária de Loulé	Faro	Loulé
Escola Secundária de Manuel Teixeira Gomes	Faro	Portimão
Escola Secundária de Maria Amália Vaz de Carvalho	Lisboa	Lisboa
Escola Secundária de Montemor-o-Velho	Coimbra	Montemor-o-Velho
Escola Secundária de Odivelas	Lisboa	Odivelas
Escola Secundária de Oliveira do Bairro	Aveiro	Oliveira do Bairro
Escola Secundária de Pombal	Leiria	Pombal
Escola Secundária de S. João da Talha	Lisboa	Loures
Escola Secundária de S. João do Estoril	Lisboa	Cascais
Escola Secundária de Santa Maria - Sintra	Lisboa	Sintra
Escola Secundária de Seia	Guarda	Seia
Escola Secundária de Silves	Faro	Silves
Escola Secundária de Stº André	Setúbal	Barreiro
Escola Secundária de Tavira	Faro	Tavira
Escola Secundária de Tomás Cabreira	Faro	Faro
Escola Secundária de Vendas Novas	Évora	Vendas Novas
Escola Secundária de Vitorino Nemésio	Lisboa	Lisboa
Escola Secundária Diogo de Gouveia	Beja	Beja
Escola Secundária do Dr. Francisco Fernandes Lopes	Faro	Olhão
Escola Secundária do Eng. Acácio Calazans Duarte	Leiria	Marinha Grande
Escola Secundária do Forte da Casa	Lisboa	Vila Franca de Xira
Escola Secundária de Henriques Nogueira	Lisboa	Torres Vedras
Escola Secundária do Prof. Reynaldo dos Santos	Lisboa	Vila Franca de Xira





Escola Secundária do Professor Herculano de Carvalho	Lisboa	Lisboa
Escola Secundária Dr. Bernardino Machado	Coimbra	Figueira da Foz
Escola Secundária Dr. Manuel Candeias Gonçalves -Odemira	Beja	Odemira
Escola Secundária Emídio Navarro	Viseu	Viseu
Escola Secundária Infanta D. Maria	Coimbra	Coimbra
Escola Secundária Jacôme Ratton	Santarém	Tomar
Escola Secundária José Belchior Viegas - São Brás de Alportel	Faro	São Brás de Alportel
Escola Secundária José Saramago	Lisboa	Mafra
Escola Secundária Marques de Castilho	Aveiro	Águeda
Escola Secundária Martinho Árias	Coimbra	Soure
Escola Secundária Monserrate	Viana do Castelo	Viana do Castelo
Escola Secundária Poeta António Aleixo	Faro	Portimão
Escola Secundária Santa Maria Maior	Viana do Castelo	Viana do Castelo
Escola Secundária Sebastião e Silva	Lisboa	Oeiras

Começámos por criar um ficheiro em Excel com os dados das escolas, considerando ainda uma coluna onde inserimos o número da escola, segundo a ordem pela qual as escolas são apresentadas (esta ordem – que, no caso presente, é a ordem crescente - não é importante para o que se segue), a que chamámos *EscolasSec* e do qual apresentamos um pequeno excerto:

	A	B	C	D
		Nome	Distrito	Concelho
1	1	Escola Secundária Alves Martins	Viseu	Viseu
2	2	Escola Secundária Amélia Rey Colaço	Lisboa	Oeiras
3	3	Escola Secundária c/ 3.º ciclo do ensino básico José Afonso	Lisboa	Loures
4	4	Escola Secundária D. Afonso Henriques	Porto	Santo Tirso
5	5	Escola Secundária Vergílio Ferreira	Lisboa	Lisboa
6	6	Escola Secundária da Lourinhã	Lisboa	Lourinhã
7	7	Escola Secundária da Moita	Setúbal	Moita
8	8	Escola Secundária da Sertã	Castelo Branco	Sertã
9	9	Escola Secundária David Mourão Ferreira	Lisboa	Lisboa

Amostra aleatória simples

1.º Passo - Utilizando a função **RAND()** [ALEATÓRIO()], atribuir um número aleatório⁴, entre 0 e 1, a cada escola. Para isso, basta inserir a função na célula E2 e replicá-la tantas vezes, quantas as escolas (ou seja, 70 vezes):

	A	B	E
		Nome	
2	1	Escola Secundária Alves Martins	=RAND()
3	2	Escola Secundária Amélia Rey Colaço	=RAND()
4	3	Escola Secundária c/ 3.º ciclo do ensino básico José Afonso	=RAND()
5	4	Escola Secundária D. Afonso Henriques	=RAND()
6	5	Escola Secundária Vergílio Ferreira	=RAND()
7	6	Escola Secundária da Lourinhã	=RAND()
8	7	Escola Secundária da Moita	=RAND()
9	8	Escola Secundária da Sertã	=RAND()
10	9	Escola Secundária David Mourão Ferreira	=RAND()
11	10	Escola Secundária de Albufeira	=RAND()
12	11	Escola Secundária de Alves Redol	=RAND()
13	12	Escola Secundária de Arganil	=RAND()

⁴ Nota: Embora os números gerados sejam referidos como aleatórios, convém ter presente que os números que se obtêm são *pseudo-aleatórios*, já que é um mecanismo determinista que lhes dá origem. No entanto, comportam-se como números aleatórios (passam uma bateria de testes destinados a confirmar a sua aleatoriedade) daí a sua utilização como tal.





Para visualizar as fórmulas na folha de Excel, bastou seleccionar:

Tools [Ferramentas] → **Options** [Opções] → **View** [Ver] → **Formulas** [Fórmulas] → Ok

Uma vez que a função **RAND()** [ALEATÓRIO()] é uma função volátil, isto é, muda quando se recalcula a folha, no caso de pretendermos ficar com os valores gerados convém ir ao **Edit** [Editar], copiar e fazer um **Paste Special – Values** [Colar especial - Valores], como se indica a seguir:

	A	B	E	F
1		Nome		
2	1	Escola Secundária Alves Marti	0,618064	
3	2	Escola Secundária Amélia Rey	0,052667	
4	3	Escola Secundária c/ 3.º ciclo	0,119493	
5	4	Escola Secundária D. Afonso H	0,41474	
6	5	Escola Secundária Vergílio Fer	0,33154	
7	6	Escola Secundária da Lourinhã	0,065462	
8	7	Escola Secundária da Moita	0,946472	
9	8	Escola Secundária da Sertã	0,319074	
10	9	Escola Secundária David Mour	0,333699	
11	10	Escola Secundária de Albufeira	0,739753	
12	11	Escola Secundária de Alves Red	0,959635	

	A	B	E	F
1		Nome		
2	1	Escola Secundária Alves Marti	0,676871	0,618064
3	2	Escola Secundária Amélia Rey	0,834374	0,052667
4	3	Escola Secundária c/ 3.º ciclo	0,282121	0,119493
5	4	Escola Secundária D. Afonso H	0,862547	0,41474
6	5	Escola Secundária Vergílio Fer	0,191305	0,33154
7	6	Escola Secundária da Lourinhã	0,799682	0,065462
8	7	Escola Secundária da Moita	0,307812	0,946472
9	8	Escola Secundária da Sertã	0,03148	0,319074
10	9	Escola Secundária David Mour	0,883317	0,333699
11	10	Escola Secundária de Albufeira	0,944583	0,739753
12	11	Escola Secundária de Alves Red	0,871099	0,959635

Colámos os valores na coluna F e fizemos o **Save** [Guardar]. Repare-se que os valores que estavam inicialmente na coluna E foram alterados, dando origem a novos valores (devido ao facto de a função **RAND()** [ALEATÓRIO()] ser volátil, como referimos anteriormente);

2.º Passo – Ordenar o ficheiro, utilizando como critério a coluna F;

3.º Passo – Como pretendemos uma amostra de dimensão 10, seleccionar as primeiras 10 escolas do ficheiro ordenado:

	A	B	E	F
1		Nome		
2	24	Escola Secundária de Domingos	0,01211949	
3	34	Escola Secundária de Loulé	0,0368761	
4	37	Escola Secundária de Montemor-	0,0461299	
5	2	Escola Secundária Amélia Rey	0,0526671	
6	6	Escola Secundária da Lourinhã	0,0654624	
7	28	Escola Secundária de Gil Eanes	0,0696173	
8	33	Escola Secundária de Júlio Dante	0,0740176	
9	62	Escola Secundária Jacôme Ratto	0,085772	
10	30	Escola Secundária de Jaime Cort	0,0995526	
11	13	Escola Secundária de Avelar Brot	0,1023393	

As escolas seleccionadas foram as que estão identificadas pelos números 24, 34, 37, 2, 6, 28, 33, 62, 30 e 13.

Amostra aleatória com reposição

Vamos apresentar dois processos para seleccionar uma amostra aleatória com reposição, utilizando a função **RANDBETWEEN** [ALEATÓRIOENTRE] ou a função **Sampling** [Amostragem]:

1. Função **RANDBETWEEN** [ALEATÓRIOENTRE]

Para seleccionar aleatoriamente uma escola da lista anterior, utilizamos a função **RANDBETWEEN(m;n)**, que devolve um número pseudo-aleatório entre os números **m** e **n** especificados ($m < n$). Como o nosso ficheiro é constituído por 70 escolas e pretendemos obter uma amostra de dimensão 10, escolhemos as células E2:E11 para replicar a função **RANDBETWEEN(1;70)** [ALEATÓRIOENTRE(1;70)]:



	E
2	=RANDBETWEEN(1,70)
3	=RANDBETWEEN(1,70)
4	=RANDBETWEEN(1,70)
5	=RANDBETWEEN(1,70)
6	=RANDBETWEEN(1,70)
7	=RANDBETWEEN(1,70)
8	=RANDBETWEEN(1,70)
9	=RANDBETWEEN(1,70)
10	=RANDBETWEEN(1,70)
11	=RANDBETWEEN(1,70)

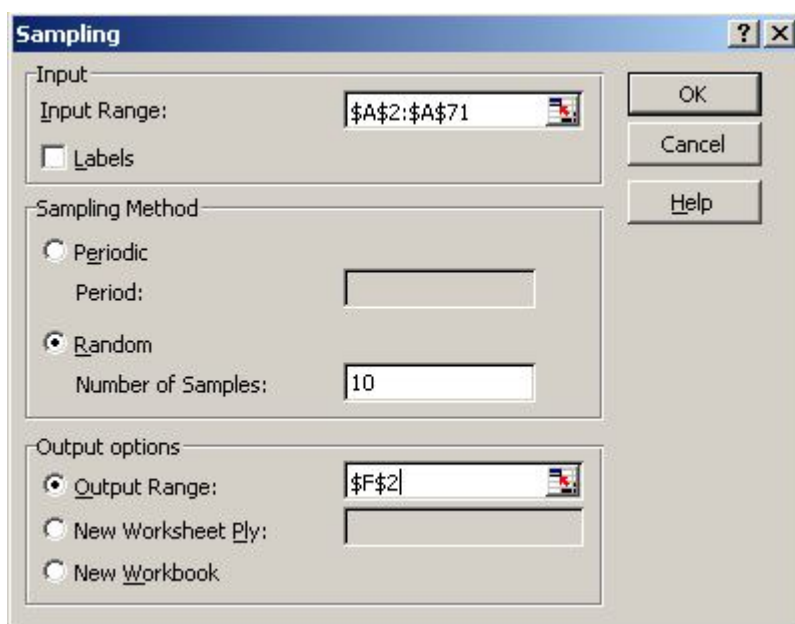
	E
2	45
3	10
4	62
5	37
6	47
7	23
8	32
9	17
10	50
11	11

	E	F
2	15	45
3	56	10
4	63	62
5	10	37
6	23	47
7	62	23
8	28	32
9	43	17
10	54	50
11	38	11

A amostra obtida é constituída pelas escolas com os números obtidos nas células E2:E11. Uma vez que a função **RANDBETWEEN** é uma função volátil, isto é, muda quando se recalcula a folha, como pretendíamos ficar com os valores gerados fizemos um *Paste Special* [Colar especial] dos valores obtidos para as células F2:F11 e fizemos o *Save* [Guardar]. Repare-se que os valores que estavam inicialmente nas células E2:E11 foram alterados, dando origem a uma nova amostra. Esta amostra alterar-se-á sempre que procedermos a alguma operação na folha de cálculo.

2. Função *Sampling* [Amostragem]

No Excel existe uma função que permite seleccionar aleatoriamente um subconjunto de números de um conjunto mais vasto de números. Acede-se a esta função seleccionando *Tools* [Ferramentas] → *Data Analysis* [Análise de dados] → *Sampling* [Amostragem] (se o comando *Data Analysis* [Análise de dados] não constar do menu, seleccione *Tools* [Ferramentas] e, na opção *Add-Ins* [Suplementos], seleccione *Analysis ToolPack*) e procedendo como se indica a seguir:



Na caixa de diálogo, proceder do seguinte modo:



- Em *Input Range* [Intervalos de entrada]: colocar o endereço da população de onde pretendemos seleccionar a amostra;
- Em *Sampling Method* [Método de amostragem]: seleccionar *Random* [Aleatório], em *Number of Samples* [Número de amostras], seleccionar a dimensão da amostra (nós escolhemos a dimensão 10, como no caso anterior);
- Em *Output options* [Opções de saída]: seleccionar a localização para onde pretendemos colocar a amostra (nós optámos por seleccionar a célula \$F\$2, para que a amostra ficasse colocada nas células \$F\$2:\$F\$11, como no caso da selecção através da função *Randbetween* [Aleatórioentre]).

Função **VLOOKUP** [PROCV]

Para seleccionar o nome das escolas correspondentes aos elementos da amostra obtida – células F2:F11 –, vamos utilizar a função *VLOOKUP(a;b;c)* [PROCV (a;b;c)], que vai à tabela das escolas, constituída pelas 2 colunas com o número e nome das escolas – células A2:B71 (argumento b) –, seleccionar os nomes das escolas – 2ª coluna da tabela seleccionada (argumento c) –, que correspondem aos números das células F2:F11 (argumento a):

	F	G
2	45	= VLOOKUP(F2;\$A\$2:\$B\$71;2)
3	10	= VLOOKUP(F3;\$A\$2:\$B\$71;2)
4	62	= VLOOKUP(F4;\$A\$2:\$B\$71;2)
5	37	= VLOOKUP(F5;\$A\$2:\$B\$71;2)
6	47	= VLOOKUP(F6;\$A\$2:\$B\$71;2)
7	23	= VLOOKUP(F7;\$A\$2:\$B\$71;2)
8	32	= VLOOKUP(F8;\$A\$2:\$B\$71;2)
9	17	= VLOOKUP(F9;\$A\$2:\$B\$71;2)
10	50	= VLOOKUP(F10;\$A\$2:\$B\$71;2)
11	11	= VLOOKUP(F11;\$A\$2:\$B\$71;2)

	F	G
2	45	Escola Secundária de Silves
3	10	Escola Secundária de Albufeira
4	62	Escola Secundária Jacôme Ratton
5	37	Escola Secundária de Montemor-o-Velho
6	47	Escola Secundária de Tavira
7	23	Escola Secundária de Dom Manuel Martins
8	32	Escola Secundária de José Falcão
9	17	Escola Secundária de Santa Maria da Feira
10	50	Escola Secundária de Vitorino Nemésio
11	11	Escola Secundária de Alves Redol

Se pretendêssemos seleccionar, além do nome da escola, o seu distrito, então teríamos de considerar como “argumento b” da função *VLOOKUP* [PROCV] a tabela A2:C71 e como “argumento c”, o valor 3, uma vez que nos estamos a referir à 3ª coluna da tabela considerada.

Tarefa

Selecciona na tua turma 3 alunos para pertencerem à Associação de Estudantes. A selecção deve ser tal que:

- Não permita que o mesmo aluno seja seleccionado 2 vezes;
- Dê a cada aluno igual probabilidade de ser seleccionado.

Utiliza um dos processos considerados anteriormente, que julgues conveniente.