

Módulo 3. Variabilidade - PRÁTICAS

Construção de modelos estocásticos em folha de cálculo

Bibliografia:

Donovan, TM, and JE Goetz. 1999. Exploring demographic and environmental stochasticity with spreadsheets. *Bull. Ecological Society of America* 80(4):235-239.

1. Estocasticidade Demográfica (ED): simulação duma população com 10 indivíduos.

Considere-se uma população em que se estima que a sobrevivência total é $S=0.85 \text{ ano}^{-1}$ e em que cada indivíduo não pode ter mais de 1 descendente por ano, sendo a probabilidade de ter 1 descendente dada por $b= 0.21 \text{ ano}^{-1}$. O produto bS representa a probabilidade de nascer um indivíduo e este sobreviver 1 ano. A taxa de incremento da população é $\lambda=S+bS$.

1) Abra uma folha de cálculo nova e introduza o valor de S em B1 e o de bS em D1. Em G2 introduza a fórmula $=10*(B1+D1)$ para obter uma estimativa determinística de N_{t+1} , assumindo que $N_t=10$.

Esta estimativa parece-lhe fiável com um valor de N_t tão pequeno ? Explique como é que a ED interfere no processo de projecção da população.

2) Nas células A5:A14 introduza os dígitos 1 a 10, representando os 10 indivíduos da população no instante t . Nas células B5:B14 introduza a fórmula $=RAND()$. Esta fórmula representa uma função pré-definida do Excel que gera npa's (números pseudo-aleatórios) com distribuição uniforme entre 0 e 1. Os npa's são recalculados de cada vez que a folha excel efectua cálculos.

O que significa dizer que a distribuição é uniforme entre 0 e 1 ?

Nota: se quisesse gerar npa's uniformes entre a e b , devia usar a fórmula $=rand()*(b-a)+a$

3) Na célula C5 escreva a fórmula $=IF(B5<B\$1;1;0)$. A fórmula compara o valor em B5 com o valor de S em B1. Se o valor em B5 for inferior, o indivíduo sobreviveu e, em C5, é escrito o valor 1; caso contrário morreu e é escrito 0. Copie e cole a mesma fórmula para todas as células C5:C14.

Na célula D5 escreva a fórmula $=IF(B5<D\$1;1;0)$. A fórmula compara o valor em B5 com o valor de bS em D1. Se o valor em B5 for inferior, o indivíduo teve um descendente que sobreviveu e, em D5, é escrito o valor 1; caso contrário, não teve descendente e aparece 0 em D5. Copie e cole a mesma fórmula para todas as células D5:D14.

Nas células C15 e D15 escreva, respectivamente, as fórmulas $=SUM(C5:C14)$ e $=SUM(D5:D14)$. Estas fórmulas somam o número de sobreviventes e de descendentes, respectivamente, no instante $t+1$, projectados a partir dos 10 indivíduos iniciais.

A soma dos sobreviventes com os descendentes são a população em $t+1$. Acaba portanto de projectar a população tendo em atenção o factor aleatório que se designa por ED. Pode confiar que esta única projecção lhe forneceu a população que vai haver em $t+1$?

4) Nas células A18:A67 introduza os dígitos 1 a 50, representando 50 réplicas da projecção da população de t para $t+1$. (Propõem-se 50 réplicas, mas poderiam ser feitas muitas mais). Copie o valor (usando Edit\Paste Special\Values) das células C15 e D15 para, respectivamente, B18 e C18. Esta forma de copiar ignora as fórmulas e copia apenas os valores numéricos. Repetir para as células B19 e C19, B20 e C20, B21 e C21, etc... 50 vezes. Notar que de cada vez que faz uma cópia de valores, os valores nas células C5:C15 e D5:D15 muda imediatamente porque as funções RAND() geram novos npa's.

Nota: Este trabalho moroso pode ser evitado fazendo uma “macro” no Excel. Ver abaixo como.

Na célula D18 escrever a fórmula =B18+C18, a qual dá o número de indivíduos da população em $t+1$, na réplica 1. Copiar e colar a fórmula para as células D18:D67, obtendo-se assim 50 réplicas da população em $t+1$.

Cada réplica fornece uma estimativa possível de N_{t+1} . Todos os valores de N_{t+1} parecem-lhe igualmente prováveis ?

5) Nas células E18:E36 introduza os dígitos 0, 1, 2 ... 18; 0 e 18 representam limites inferior e superior, respectivamente, para o número de indivíduos em $t+1$, determinados por inspecção visual das 50 réplicas em D18:D67.

Na célula F18 introduza a fórmula =COUNTIF(D\$18:D\$67;E18). Esta fórmula conta a quantidade de vezes que o valor da célula E18 aparece no conjunto das células D18:D67. Determina portanto a *frequência absoluta* de réplicas em que $N_{t+1}=0$. Copie e cole a fórmula para as células F18:F36. Obtém-se assim a distribuição de frequências absolutas de N_{t+1} .

Introduzir em F39 a fórmula =SUM(F18:F36) para obter o total de réplicas (50).

6) Em G18 introduzir a fórmula =F18/F\$39, a qual calcula a *frequência relativa* de réplicas em que $N_{t+1}=0$. Copie e cole para as células G18:G36, a fim de obter a distribuição de frequências relativas dos valores de N_{t+1} . Estas frequências podem ser interpretadas como as *probabilidades* de obter cada valor de N_{t+1} .

Na célula H18 introduza =G18 e na célula H19 introduza =G19+H18. Copie e cole a célula H19 para a gama de células H19:H36. Obtem-se assim as *frequências relativas acumuladas* dos valores de N_{t+1} . Por exemplo, o valor da célula H30, é interpretado como a probabilidade de que a população em $t+1$ tenha 12 *ou menos* indivíduos (12 é o valor em E30).

Acaba de construir a chamada função de distribuição cumulativa. Assegure-se de que compreendeu o seu significado probabilístico. Se compreendeu, não terá dificuldade em compreender o que é a curva de risco.

7) Construa um gráfico de barras, usando as gamas de células E18:E36 para abcissas e F18:F36 para ordenadas. O gráfico representa as frequências absolutas de N_{t+1} . Construa um gráfico do tipo “scatter” (dispersão de pontos), usando as gamas de células E18:E36 (abcissas) e H18:H36 (ordenadas). O gráfico tem forma logística (curva em S) e representa as probabilidades acumuladas de N_{t+1} . Em abcissas estão números de indivíduos em $t+1$ e, em ordenadas, a probabilidade de a população em $t+1$ ter um efectivo *igual ou menor* que o correspondente valor em abcissa. Acaba de fazer a curva de risco desta população.

O que significa “risco” ? dê um exemplo, usando a sua curva, no qual faça uso da palavra “probabilidade” e da palavra “tempo”.

8) Na célula B70 introduza a função =**AVERAGE(B18:B67)**, a qual calcula o número médio de sobreviventes em N_{t+1} , a partir das 50 réplicas.

Na célula B71 introduza a função =**STDEV(B18:B67)**, a qual calcula a dispersão dos sobreviventes em torno da sua média, através do desvio-padrão.

Na célula B72 introduza a função =**B71/B70**, a qual calcula o coeficiente de variação (CV), i.e., expressa o desvio-padrão como uma proporção da média.

Faça cópia das células B70:B72 e cole em C70:C72, obtendo a média, d.-p. e CV do número de descendentes em $t+1$.

Faça cópia das células B70:B72 e cole em D70:D72, obtendo a média, d.-p. e CV do número de indivíduos em $t+1$.

Qual era o valor que *esperava* obter em B70 ? compare com o que obteve.

Qual era o valor que *esperava* obter em D70 ? compare com o que obteve.

Suponha que em vez duma população inicial de 10 indivíduos tinha simulado ED numa população inicial de 100 indivíduos. Se tudo o resto (*S* e *b*) se mantivesse inalterado, a população de 100 devia ter valores mais altos ou mais baixos em:

- B70 e D70 ? (médias)
- B71 e D71 ? (desvios-padrões)
- B72 e D72 ? (CV's)

MACRO para o ponto 4)

Macro é um termo usado na gíria do Excel para designar uma sequência de operações que a folha de Excel memoriza e repete, sempre que o utilizador lhe dá ordem para repetir. Um macro é portanto um programa de computador, com a particularidade de ser escrito numa linguagem que o Excel entende (Visual Basic ou VB) e de ser criado e compilado no ambiente do próprio Excel. Ao contrário da generalidade das linguagens de computador, neste caso o utilizador não tem necessariamente de estudar a linguagem antes de escrever o programa. É possível ordenar ao Excel que inicie a memorização duma sequência de operações, à medida que estas se vão *fazendo* (por oposição a serem escritas). No fim, dá-se ordem para parar a memorização e informa-se o Excel de quais as teclas que futuramente serão usadas para ordenar que o programa seja executado. Exemplifica-se em seguida como proceder para construir uma Macro que faça o ponto 4) acima.

Suponha que terminou o ponto 3) acima,

1. No menu geral seleccione Tools\Macro\Record New Macro. Na caixa de diálogo que aparece, dê um nome ao Macro que vai ser feito (e.g., réplicas). Seleccione também a tecla que no futuro iniciará o macro (e.g. Ctrl t). O macro está pronto para ser memorizado. Se se fizer um erro ao gravá-lo, as correcções do erro também serão memorizadas.

2. No menu geral seleccione Edit\Go To. Na caixa de diálogo que aparece, escrever a identificação das células onde quer o cursor posicionado, B18:C18. Clicar OK e estas células são realçadas.

3. No menu geral seleccione Insert\Cells. Na caixa de diálogo que aparece, seleccione “move cells down”.

4. No menu geral seleccione Edit\Go To. Na caixa de diálogo que aparece, escrever as células C15:D15, onde estão o número de sobreviventes e descendentes recém-calculados.

5. No menu geral seleccione Edit\Copy. No menu geral seleccione Edit\Go To e na caixa de diálogo escreva B18:C18.
6. No menu geral seleccione Edit\Paste Special. Na caixa de diálogo que aparece, seleccione “Paste values” (ou “Values” nalgumas versões do Excel) para colar os valores copiados de C15:D15.
7. Para terminar a memorização do macro, no menu geral seleccione Tools\Macro\Stop Recording.

Para correr o macro, teclar **Ctrl t**. Se o macro estiver bem feito, a réplica (sobreviventes e descendentes) é inserida nas colunas B e C da linha 18 e, de cada vez que se pressiona Ctrl t, uma nova réplica é inserida, empurrando a anterior para baixo.

Para os estudantes curiosos em ver a linguagem (também designada por “código”) de VB deste macro, no menu geral seleccione Tools\Macro.

2. Variabilidade ambiental

A tabela seguinte apresenta valores de λ_t ($=N_{t+1}/N_t$) obtidos a partir de uma população recenseada periodicamente ao longo de $n=20$ anos:

ANOS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Taxa incr	1.02	1.07	1.04	0.92	0.95	1.03	1.12	1.00	1.09	0.92	0.92	0.97	0.92	1.05	0.91	1.06	1.12	1.13	0.95	1.01

Sabendo que a população em causa tem, no ano corrente, aproximadamente 500 indivíduos, pretende-se construir uma curva de risco para esta população daqui a 5 anos. Assume-se que os factores ambientais e erros de medição que têm provocado oscilações em λ , mantêm as suas características.

- 1) Abra uma folha de cálculo nova e insira os n valores de λ_t nas células **B2:B21**.
- 2) Construa intervalos de classe para a variável λ_t seguindo as seguintes regras:
 - a) Calcular o número de intervalos pela regra de Sturges $C = 1 + \text{Log}_{10} n / \text{Log}_{10} 2$
A amplitude do intervalo de classe é dada por $h \approx (\lambda_{\max} - \lambda_{\min}) / C$
 - b) Calcular a média dos n valores de λ_t com 3 casas decimais (mais uma que os valores de λ_t tabelados). O intervalo central fica centrado na média: $[\lambda_{\text{médio}} - h/2, \lambda_{\text{médio}} + h/2]$. Os outros intervalos são construídos para a esquerda e direita do intervalo central.

Escrever os intervalos obtidos por ordem ascendente em **D2:D6**.

Nas células **E2:E6**, escrever o ponto médio de cada intervalo (este calcula-se adicionando $h/2$ ao limite inferior do intervalo). Se tudo correu bem, estes valores devem ser: 0.92, 0.97, 1.01, 1.05, 1.10.

- 3) Em **F2:F6**, escreva as frequências absolutas de λ_t em cada intervalo. Em **G2:G6** as frequências relativas e em **H2:H6** as frequências acumuladas. A tabela seguinte sumariza os resultados esperados:

Intervalos da taxa	Ponto medio	Freq absolu	Freq relativas	Freq Cumul
0.9-0.944	0.92	5	0.25	0.25
0.944-0.988	0.97	3	0.15	0.4
0.988-1.032	1.01	4	0.2	0.6
1.032-1.076	1.05	4	0.2	0.8
1.076-1.13	1.10	4	0.2	1
		20		

- 4) Para simular os valores que λ_t virá a tomar no futuro, recorrendo à geração de npa's, estas frequências acumuladas implicam que, se um npa gerado estiver...

- ... entre 0 e 0.25, então λ_t toma o valor 0.92
- entre 0.25 e 0.40, então λ_t toma o valor 0.97
- entre 0.40 e 0.6, então λ_t toma o valor 1.01
- entre 0.6 e 0.8, então λ_t toma o valor 1.05
- entre 0.8 e 1.00, então λ_t toma o valor 1.10

Na linha 9, em **D9:J9**, escreva da esquerda p/ a direita: “Replica”, “t”, “t+1”, “t+2”, “t+3”, “t+4”, t+5

Na coluna E, em **E10:E12**, escreva de cima p/ baixo: NPA’s, TAXA, N

Em **F10** introduza a fórmula =RAND(). Copie e cole em **G10:J10**.

Em **F11** introduza a fórmula:

=IF(F\$10<0.25; 0.92; IF(F\$10<0.4; 0.97; IF(F\$10<0.6; 1.01; IF(F\$10<0.8; 1.05; 1.1))))

Esta fórmula compara o NPA em F10 com os limites definidos pelas frequências acumuladas, atribuindo à taxa λ_t , o valor correspondente.

Colar e copiar a fórmula de F11 para **G11:J11**.

5) Na célula **F12** introduza a fórmula =INT(500*F\$11) e, na célula **G12**, a fórmula =INT(F12*G\$11. Copie e cole esta última fórmula para **H12:J12**.

A fórmula calcula o valor de N_t ao longo dos próximos 5 anos, usando em cada transição (t para t+1) um novo valor de λ_t gerado a partir dos npa’s. A função INT() faz com que o resultado seja arredondado para o inteiro mais próximo.

6) Nas células **D14:D63**, escreva a sucessão de números 1, 2, ..., 50 que numeram 50 réplicas da população.

Em **E14:E63** escreva 500, o número inicial de indivíduos (presentes no ano t)

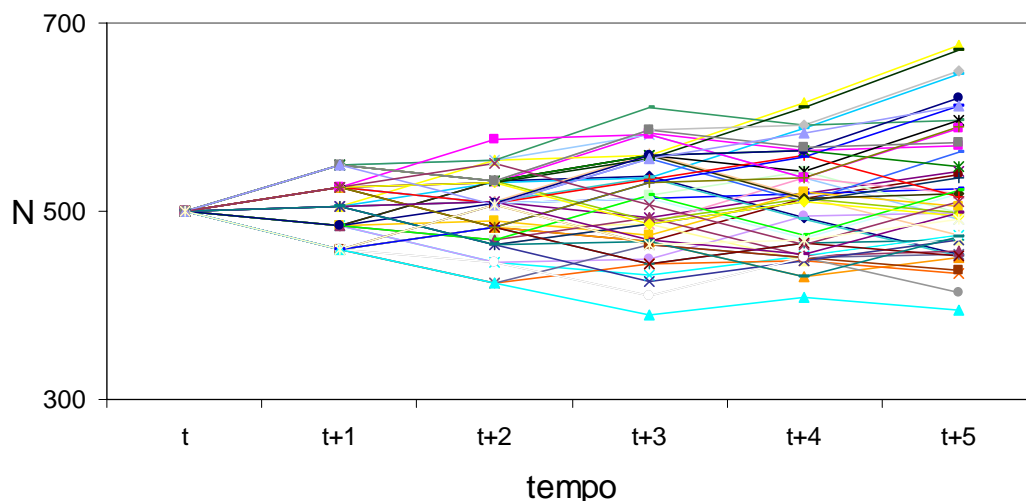
Copie e cole os valores das células F12:J12 para **F14:J14** usando Edit/Paste Special/Values, a fim de copiar apenas os valores numéricos e não as fórmulas.

Repita mais 49 vezes, copiando as células F12:J12 para as colunas F a J, linhas 15 a 63.

6) Na célula **F65** introduza a fórmula =STDEV(F14:F63) para calcular o desvio-padrão de N_{t+1} nas 50 réplicas. Copie e cole em **G65:J65**.

Note a forma como o desvio-padrão e, portanto, a incerteza sobre o futuro da população, aumenta com o tempo.

7) Faça um gráfico com a evolução das 50 réplicas da população entre t e t+5. O aspecto deve ser mais ou menos este:



8) Pretende-se agora construir a curva de risco para a população daqui a 5 anos. É necessário, em primeiro lugar, saber quais os valores possíveis que ela pode tomar daqui a 5 anos. Uma forma simples de o fazer consiste em copiar toda a coluna J14:J63 para o lado, por exemplo para L14:L63, e ordenar esta última. Para isso, seleccionar L14:L63 e, no menu principal seleccionar Data/Sort A↓Z.

Em seguida, escrever na coluna N, em N14:N_ , a sucessão de valores *diferentes* (omitir as repetições) que N_{t+5} pode tomar.

9) Em O14 escrever a fórmula =COUNTIF(L\$14:L\$63;N14) que conta o número de vezes que o número de indivíduos em N14 apareceu no ano t+5. Copiar e colar pela coluna O abaixo: O15:O_ . Ao fundo da coluna O somar as frequências, para verificar se somam 50.

10) Em P14:P_ colocar as frequências relativas da coluna N14:N_ . Em Q14:Q_ calcular e escrever as frequências cumuladas.

11) Fazer o gráfico da coluna N14:N_ contra a coluna Q14:Q_ . Escolher o gráfico do tipo “scatter”e colocar “gridlines” no gráfico, a fim de tornar a leitura mais fácil.

Esta é a curva de risco da população no ano t+5. Qual é a probabilidade de, em t+5, haver menos de 500 indivíduos na população ?

O aspecto do gráfico deve ser mais ou menos este:

Curva de risco a 5 anos

