Array formulas

Working with arrays of data
Matrix operations
Using array functions

Arrays and matrices

- We encounter matrices in biology commonly
  - Multiple variables of interest in statistics
  - Age structure in populations
- Some types of calculations are more conveniently done with matrix approaches
- Some Excel functions that we might like to use are array formulas – that is, they work with ranges of cells instead of single cells
- We need to understand how matrices are used, and how Excel allows us to work with them

Array formulas

- In Excel, an array formula is any that repeats an operation on a range of cells
- Array formulas can return a single cell as a result, or may return a range of cells
- Best seen by example...

Single cell returned: calculating your grade in a class

- You have several different assignments worth varying numbers of points
- You want to know your average percentage
- You could:
  - Calculate your percentage for each assignment
  - Calculate the average of the percents
- Or, you could:
  - Use an array formula to calculate the average of the percentages in one formula

Example in Excel

C8 | {=AVERAGE(B2:B5/C2:C5)}

	A	B	C	D	E	F
1	Assignment	My score	Points possible		My percentages	
2	Homework 1	8	10		80%	
3	Homework 2	25	26		96%	
4	Homework 3	19	29		66%	
5	Homework 4	132	140		94%	
6						
7			Average by array formula		Average of percentages	
8			84%		84%	
9						
10						

Formula uses a range of cells in the numerator and denominator

When entered, use CTRL+SHIFT+ENTER

Curly braces show that this is an array formula

Unpacking the formula...

1. Calculate all the ratios, row by row:

B2/C2, B3/C3, B4/C4, B5/C5

C8 | {=AVERAGE(B2:B5/C2:C5)}

	A	B	C	D	
1	Assignment	My score	Points possible		M
2	Homework 1	8	10		
3	Homework 2	25	26		
4	Homework 3	19	29		
5	Homework 4	132	140		
6					
7			Average by array formula		A
8			84%		
9					
10					

2. Average all the ratios

If not entered as an array formula...

C8 =AVERAGE(B2:B5/C2:C5)

	A	B	C	D	E	F
1	Assignment	My score	Points possible		My percentages	
2	Homework 1	8	10		80%	
3	Homework 2	25	26		96%	
4	Homework 3	19	29		66%	
5	Homework 4	132	140		94%	
6						
7			Average by array formula		Average of percentages	
8			#VALUE!		84%	
9						
10						

Error message shown here

Sometimes you will get the result just for the first cell in each range – wrong, but not so obviously

Range of cells returned: calculating frequencies

- You know how to use PivotTables to count how often categories occur
- We could use it to calculate frequencies of discrete numbers as well
- We can't use it to get frequencies of binned continuous numbers (needed for histograms)
- We can use the frequency() array formula for that

Frequencies

=frequency(C2:C101,E2:E8)

C	D	E	F	G	H
Height		Bins	Frequency		
25.9		14	=frequency(C2:C101,E2:E8)		
21.6		18			
31.4		22			
28.1		26			
29.8		30			
27.1		34			
25.2		38			
30.2					
30.8					
29.2					
25.4					
26					
27.2					
25.3					
34					
22.2					
21.9					

Select the output range, arguments for frequency() identify the data range and bins range

{=FREQUENCY(C2:C101,E2:E8)}

C	D	E	F	G
Height		Bins	Frequency	
25.9		14	1	
21.6		18	3	
31.4		22	11	
28.1		26	35	
29.8		30	38	
27.1		34	11	
25.2		38	1	
30.2				
30.8				
29.2				
25.4				
26				
27.2				
25.3				
34				

CTRL+SHIFT+ENTER

Bins – upper limit
First is 14 and below (min in data set is 13.5)
Next is 14 to 18, and so on

Average of grouped data

- We have data on numbers of osprey seen at the San Elijo Lagoon during weekly surveys over two years
- We don't have the raw counts, but instead have a frequency table
- How do we calculate the mean and standard deviation of number of osprey?

C	D
Number of osprey	Frequency
0	40
1	20
2	30
3	6
4	4

Mean without an array formula

$$\frac{\sum x_i}{n}$$

=E8/E9

	C	D	E
	Number of osprey	Frequency	Products
	0	40	0
	1	20	20
	2	30	60
	3	6	18
	4	4	16
		Sum of products	114
		Total counts	100
		Mean	1.14

Osprey seen on the 40 counts with 0 osprey =c2*d2

Total osprey seen across all counts =sum(e2:e6)

Total counts =sum(d2:d6)

Mean with an array formula

$$\frac{\sum x_i}{n}$$

{=SUM(C2:C6*D2:D6)/SUM(D2:D6)}

	B	C	D	E	F	G
		Number of osprey	Frequency			
		0	40			
		1	20			
		2	30			
		3	6			
		4	4			
		Mean	1.14			

Standard deviation without an array formula

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

B	C	D	E	F	G
	Number of osprey	Frequency	Products	Squared deviations	Sq. dev. X freq.
	0	40	0	1.2996	51.984
	1	20	20	0.0196	0.392
	2	30	60	0.7396	22.188
	3	6	18	3.4596	20.7576
	4	4	16	8.1796	32.7184
		Sum of products	114		
		Total counts	100		
		Average	1.14		
		Sum of squared deviations	128.04		
		n-1	99		
		Standard deviation	1.137248		

1. Calculate the squared difference between the number of osprey and the mean number of osprey

=(c2-e$10)^2

2. Calculate the sum of squared deviations for each number of osprey

=d2*f2

3. Sum the squared deviations for each row

=sum(g2:g6)

4. sqrt(SS/df), =sqrt(e12/e13)

Standard deviation with an array formula

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

E16

	A	B	C	D	E	F	G
			Number of osprey	Frequency			
			0	40			
			1	20			
			2	30			
			3	6			
			4	4			
			Mean	1.14			
			Sum of Squared Deviations	128.04	{=SUM((C2:C6-D8)^2*D2:D6)}		
			Standard Deviation	1.1372481	=SQRT(D9/(SUM(D2:D6)-1))		

1. Squared deviations

2. How often they occurred

Matrix algebra

- Data that can be held in a matrix (rows and columns) can be manipulated using matrix algebra
- Certain types of calculations become much more efficient using a matrix approach
- But, there are rules for how to manipulate matrices

Matrices

- Made up of rows and columns
- A 2x2 matrix has two rows, two columns
- Element 2,1 is c
- Symbolized by bold, capital letters

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Rules of matrix algebra

- Addition
- Subtraction
- Multiplication
- Division
- Transposition

Addition

Add a scalar to a matrix

$$1 + \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1+a & 1+b \\ 1+c & 1+d \end{bmatrix}$$

Add two matrices (of equal dimensions)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

Is matrix addition commutative?

Subtraction

Subtract a scalar from a matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} - 1 = \begin{bmatrix} a-1 & b-1 \\ c-1 & d-1 \end{bmatrix}$$

Subtract two matrices (of equal dimensions)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a-e & b-f \\ c-g & d-h \end{bmatrix}$$

Is matrix subtraction commutative?

Multiplication

Multiply a scalar by a matrix

$$2\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 2a & 2b \\ 2c & 2d \end{bmatrix}$$

But, multiplication is not done with matching elements!

$$\cancel{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \end{bmatrix} \neq \begin{bmatrix} ae & bf \\ cg & dh \end{bmatrix}}$$

Multiplying matrices: across and down

For element 1,1 multiply row 1 by column 1, and add products

$$\begin{bmatrix} a & b \\ . & . \end{bmatrix} \times \begin{bmatrix} e & . \\ g & . \end{bmatrix} = \begin{bmatrix} ae+bg & . \\ . & . \end{bmatrix}$$

For element 1,2 multiply row 1 by column 2, and add products

$$\begin{bmatrix} a & b \\ . & . \end{bmatrix} \times \begin{bmatrix} . & f \\ . & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ . & . \end{bmatrix}$$

Multiplying matrices: across and down

For element 2,1 multiply row 2 by column 1, and add products

$$\begin{bmatrix} . & . \\ c & d \end{bmatrix} \times \begin{bmatrix} e & . \\ g & . \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & . \end{bmatrix}$$

For element 2,2 multiply row 2 by column 2, and add products

$$\begin{bmatrix} . & . \\ c & d \end{bmatrix} \times \begin{bmatrix} . & f \\ . & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

Is matrix multiplication commutative?

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ea+fc & eb+fd \\ ga+hc & gc+hd \end{bmatrix}$$

Some matrix multiplication facts:

- To multiply matrices, the number of columns in the left matrix has to be equal to the number of rows on the right
- The matrix produced will have the number of rows of the left matrix, and the number of columns of the right matrix

A simple population model

- Imagine a species that:
  - Has a distinct breeding season
  - They mature to breed in a single year
  - All surviving females breed
  - After they breed all of them die
- If we know how many there are this year, can we predict how many there are next year?

How big will next year's population be?

- Ignore the males (total pop will be 2 x number of females)
- The number next year will be equal to the number of offspring produced this year that survive into next year
- The number of offspring produced this year will be a function of the number of females present multiplied by the reproductive rate, m
- Probability of survival to breeding is represented by s
- The number of females this year is n_t
- The number of females breeding next year is $s \times n_t$
- The population size next year, after the adults have bred and died, is:

  $n_{t+1} = m \times s \times n_t$

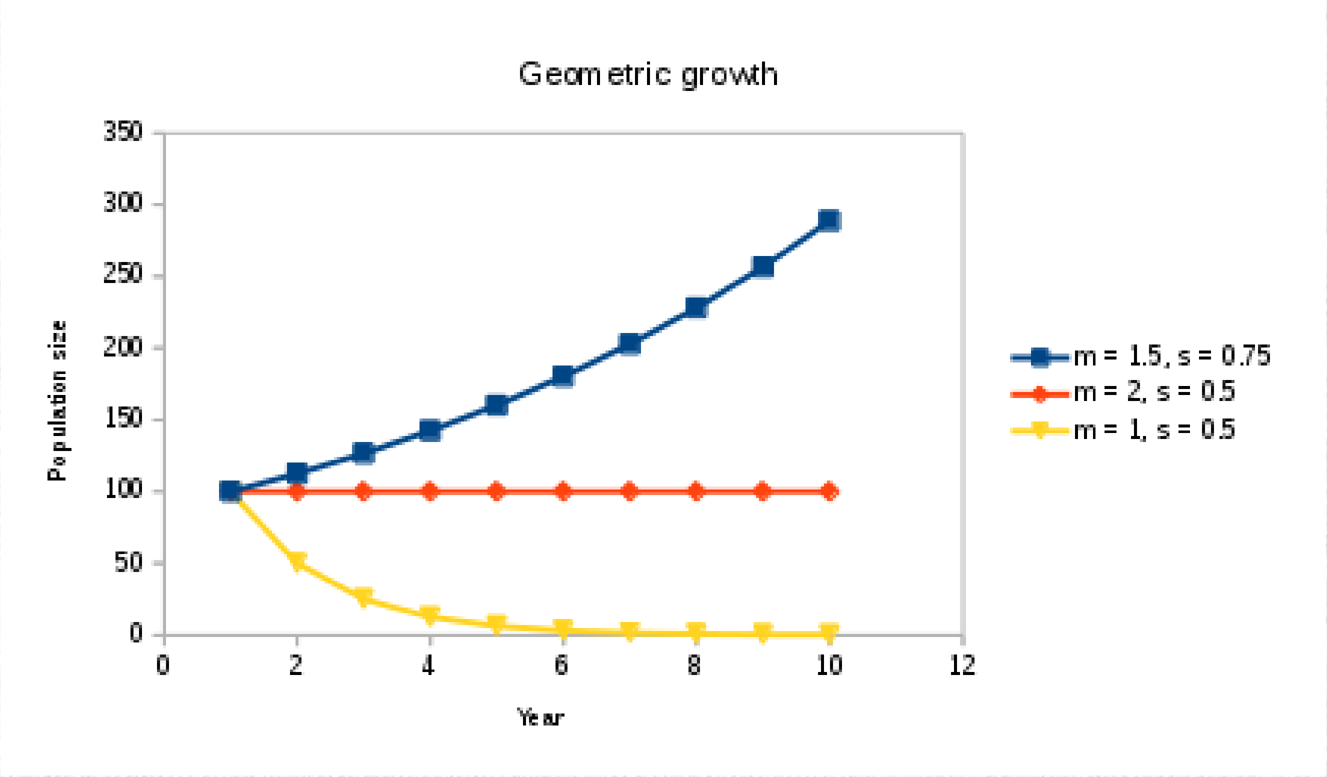
How big in two years?

- We now know n_{t+1}
- That becomes the number of females next year
- So, n_{t+2} is just:

  $n_{t+2} = m \times s \times n_{t+1}$
  $n_{t+2} = m \times s \times (m \times s \times n_t)$

- We can project the population out as far as we want "iteratively" - that is, by calculating the next year from the year before
- This is a "discrete-time model", because time is treated as a discrete quantity – it passes in one year steps

Population growth over time



Stage structured populations

- Many species live more than one year
- They usually have survival and reproductive rates that change as they age
  - Juveniles do not breed
  - Juveniles have higher mortality
  - Older adults may stop breeding
  - Generations can overlap – adults and juveniles alive at the same time
- The stage is often more important than the age
  - Adults are more alike than they are like juveniles, regardless of the age of the adults
  - If years of life are important, we would use an “age structured” model
- Need a way to account for this stage structure

Simple, stage-structured population

- Imagine a species that spends a year as juveniles, then reaches adulthood in the second year
- Once they are adult, they have a higher survival rate than juveniles (they don't all just die)

Next year's juveniles

- Juveniles come from breeding adults
- The number of adult females at time t is A_t
- Adult fecundity (f_A) is reproductive output per female, multiplied by probability of survival to first year – substitutes for m x s
- J_{t+1} is the number of juveniles at time t+1, next year

Juveniles

$$J_{t+1} = f_A A_t$$

Next year's adults

- Two sources of adults for next year:
  - Juveniles that survive to adulthood
  - Adults that survive another year

Juveniles that become adults

$$A_{t+1} = s_J J_t + s_A A_t$$

Adults that survive

Next year's total population

Total

$$N_{t+1} = f_A A_t + s_J J_t + s_A A_t$$

Total population size at t+1 is the sum of the contributions of each age class to adults and juveniles

Lefkovitch matrices

- We can take advantage of some powerful matrix algebra if we express our age-structured population model as a matrix
- Method first developed by Patrick H. Leslie for age-structure
- Generalized by L.P. Lefkovitch to be used with stage-structured models (i.e. juvenile, adult – rather than 1 yr old, 2 yr old, etc.)
- We need to re-state our model in a matrix form to take advantage of these methods

Next year's population – matrix representation

$$J_{t+1} = f_J J_t + f_A A_t$$

$$A_{t+1} = s_J J_t + s_A A_t$$

No reproduction in juveniles, so f_J is 0

Next: separate coefficients from variables

$$\begin{bmatrix} f_J & f_A \\ s_J & s_A \end{bmatrix} \times \begin{bmatrix} J_t \\ A_t \end{bmatrix} = \begin{bmatrix} J_{t+1} \\ A_{t+1} \end{bmatrix}$$

$$\begin{array}{ccc} & \text{From juveniles} & \text{From adults} \\ \text{To juveniles} & \left[\begin{array}{c} f_J \\ s_J \end{array}\right. & \left.\begin{array}{c} f_A \\ s_A \end{array}\right] \\ \text{To adults} & & \end{array}$$

$$\begin{array}{rc} & \begin{array}{cc} \text{From juveniles} & \text{From adults} \end{array} \\ \begin{array}{r} \text{To juveniles} \\ \text{To adults} \end{array} & \begin{bmatrix} f_J & f_A \\ s_J & s_A \end{bmatrix} \end{array}$$

*The Lefkovitch matrix, **L***

Matrix multiplication – across first matrix, down the second

$$J_{t+1} \qquad \begin{bmatrix} f_J & f_A \\ s_J & s_A \end{bmatrix} \times \begin{bmatrix} J_t \\ A_t \end{bmatrix}$$

$$A_{t+1} \qquad \begin{bmatrix} f_J & f_A \\ s_J & s_A \end{bmatrix} \times \begin{bmatrix} J_t \\ A_t \end{bmatrix}$$

Now with parameters and starting population sizes

$$\begin{bmatrix} 0 & 2 \\ 0.4 & 0.8 \end{bmatrix} \times \begin{bmatrix} 66 \\ 33 \end{bmatrix} = \begin{bmatrix} J_{t+1} \\ A_{t+1} \end{bmatrix}$$

$$\begin{array}{rc} & \begin{array}{cc} \text{From juveniles} & \text{From adults} \end{array} \\ \begin{array}{r} \text{To juveniles} \\ \text{To adults} \end{array} & \begin{bmatrix} 0 & 2 \\ 0.4 & 0.8 \end{bmatrix} \end{array}$$

If they stayed juveniles for more than one year, what would we change?

The second year

$$\begin{bmatrix} 0 & 2 \\ 0.4 & 0.8 \end{bmatrix} \times \begin{bmatrix} 66 \\ 33 \end{bmatrix} = \begin{bmatrix} J_{t+1} \\ A_{t+1} \end{bmatrix}$$

$$J_{t+1} = 0 \times 66 + 2 \times 33 = 66$$

$$A_{t+1} = 0.4 \times 66 + 0.8 \times 33 = 52.8$$

Matrix multiplication in Excel

C9 | {=MMULT($B$4:$C$5,B9:B10)}

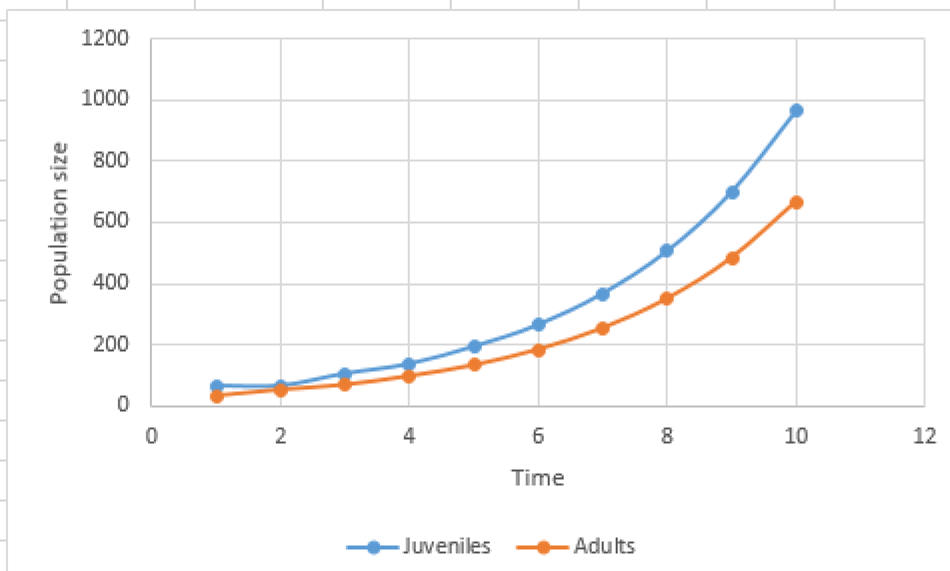
	A	B	C	D	E	F	G	H	I	J	K
1		Matrix of coefficients (Lefkovich matrix)									
2											
3		Juveniles	Adults								
4	Fecundity	0	2								
5	Survival	0.4	0.8								
6											
7		Population sizes of juveniles and adults									
8	t =	1	2	3	4	5	6	7	8	9	10
9	Juveniles	66	66								
10	Adults	33	52.8								
11											
12	Total N	99									
13											
14											

mmult(array1, array2) is the Excel formula for matrix multiplication

It is an array formula – select cells for output matrix, type in formula, hit CTRL+SHIFT+ENTER

Projected out 10 years

	A	B	C	D	E	F	G	H	I	J	K	L
1		Matrix of coefficients (Lefkovich matrix)										
2												
3		Juveniles	Adults									
4	Fecundity	0	2									
5	Survival	0.4	0.8									
6												
7		Population sizes of juveniles and adults										
8	t =	1	2	3	4	5	6	7	8	9	10	
9	Juveniles	66	66	105.6	137.3	194.3	265.3	367.7	506.3	699.2	964.4	
10	Adults	33	52.8	68.6	97.2	132.6	183.8	253.2	349.6	482.2	665.5	
11												
12	Total N	99	118.8	174.24	234.432	326.9376	449.0957	620.8266	855.9378	1181.412	1629.88	

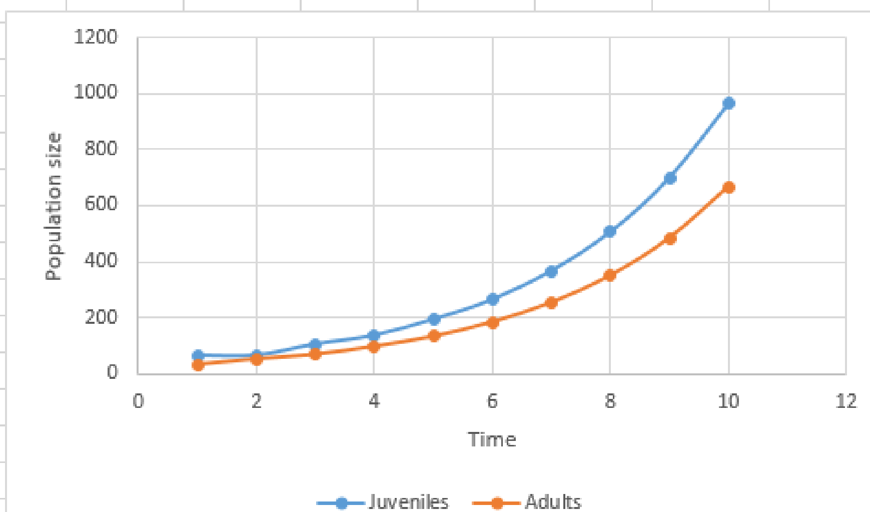


Stable stage distribution

- Can start with any number of adults and juveniles
- Over multiple years of projection, the relative number of adults and juveniles will reach a stable state that doesn't change over time = stable stage distribution
- At the stable stage distribution the population grows smoothly, by a constant multiple each year
  - Growth rate = λ
  - Represents population size at t+1 over population size at t
- If not at stable stage distribution, the population can vary in size even though the demographic rates are the same each year

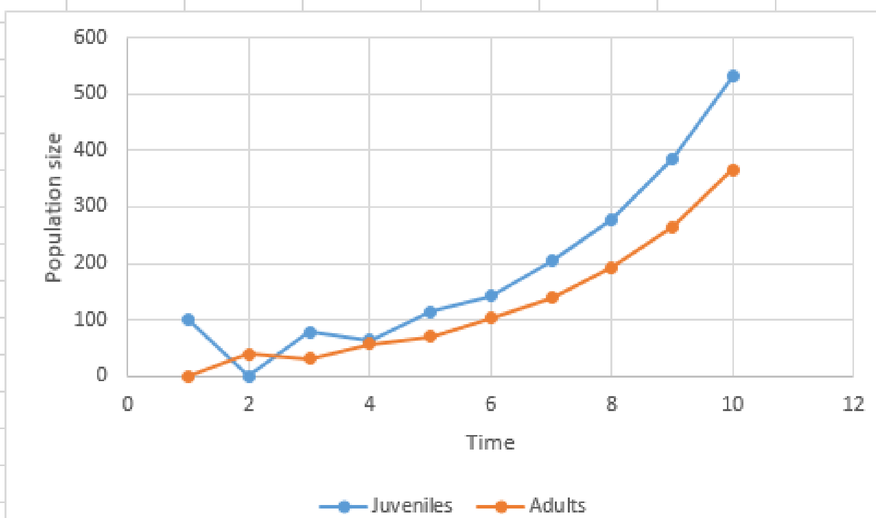
Stable stage distribution

	A	B	C	D	E	F	G	H	I	J	K
1		Matrix of coefficients (Lefkovich matrix)									
2											
3		Juveniles	Adults								
4	Fecundity	0	2								
5	Survival	0.4	0.8								
6											
7		Population sizes of juveniles and adults									
8	t =	1	2	3	4	5	6	7	8	9	10
9	Juveniles	66	66	105.6	137.3	194.3	265.3	367.7	506.3	699.2	964.4
10	Adults	33	52.8	68.6	97.2	132.6	183.8	253.2	349.6	482.2	665.5
11											
12	Total N	99	118.8	174.24	234.432	326.9376	449.0957	620.8266	855.9378	1181.412	1629.88
13											
14		Proportion of juveniles and adults									
15											
16	Juvenile	0.67	0.56	0.61	0.59	0.59	0.59	0.59	0.59	0.59	0.59
17	Adult	0.33	0.44	0.39	0.41	0.41	0.41	0.41	0.41	0.41	0.41



Not at stable stage to begin

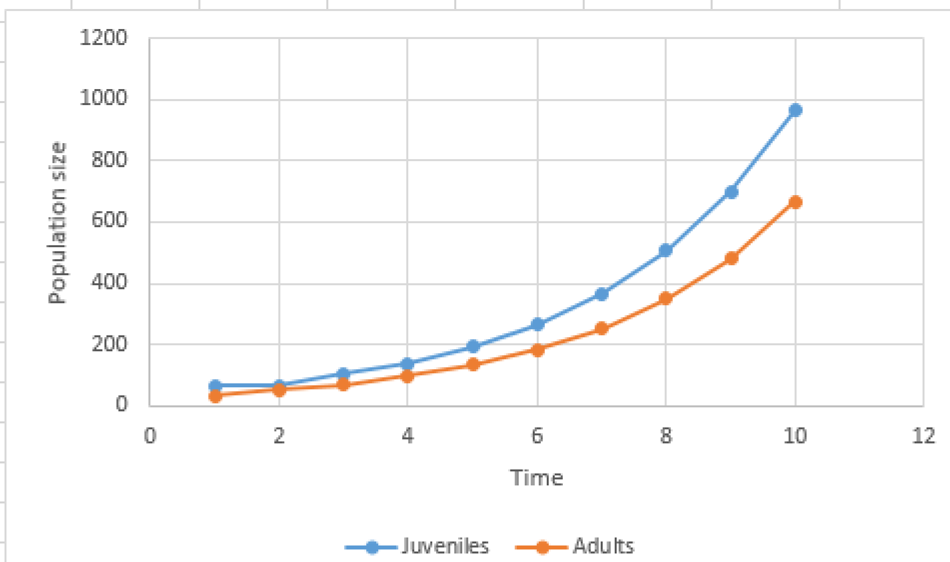
	A	B	C	D	E	F	G	H	I	J	K
1		Matrix of coefficients (Lefkovich matrix)									
2											
3		Juveniles	Adults								
4	Fecundity	0	2								
5	Survival	0.4	0.8								
6											
7		Population sizes of juveniles and adults									
8	t =	1	2	3	4	5	6	7	8	9	10
9	Juveniles	99	0	79.2	63.4	114.0	141.9	204.8	277.4	385.7	530.5
10	Adults	0	39.6	31.7	57.0	71.0	102.4	138.7	192.9	265.2	366.5
11											
12	Total N	99	39.6	110.88	120.384	185.0112	244.3162	343.4619	470.2224	650.9475	896.9359
13											
14		Proportion of juveniles and adults									
15											
16	Juvenile	1.00	0.00	0.71	0.53	0.62	0.58	0.60	0.59	0.59	0.59
17	Adult	0.00	1.00	0.29	0.47	0.38	0.42	0.40	0.41	0.41	0.41



Growth rate

- The growth rate, λ, can be estimated in two ways
  - Calculate directly from the Lefkovich matrix (eigenanalysis)
  - Project the population until it reaches stable age distribution, and divide the population size at the final time by the population size the prior year
- The first is mathematically elegant, shows that growth rate depends **only** on the demographic rates (Lefkovitch matrix), not on the population size
- The second is easy to do in a spreadsheet, mathematically simpler – that's what we'll do today

	A	B	C	D	E	F	G	H	I	J	K
1		Matrix of coefficients (Lefkovich matrix)									
2											
3		Juveniles	Adults								
4	Fecundity	0	2								
5	Survival	0.4	0.8								
6											
7		Population sizes of juveniles and adults									
8	t =	1	2	3	4	5	6	7	8	9	10
9	Juveniles	66	66	105.6	137.3	194.3	265.3	367.7	506.3	699.2	964.4
10	Adults	33	52.8	68.6	97.2	132.6	183.8	253.2	349.6	482.2	665.5
11											
12	Total N	99	118.8	174.24	234.432	326.9376	449.0957	620.8266	855.9378	1181.412	1629.88
13											
14		Proportion of juveniles and adults								Lambda =	1.379603
15											
16	Juvenile	0.67	0.56	0.61	0.59	0.59	0.59	0.59	0.59	0.59	0.59
17	Adult	0.33	0.44	0.39	0.41	0.41	0.41	0.41	0.41	0.41	0.41



Reasons to use array formulas

- Some functions can only be used as array formulas (e.g. frequency())
- Using array formulas can simplify the spreadsheet
- Using array formulas can save you work

Reasons to be cautious with array formulas

- They are easy to make mistakes with
  - Recommend you do an example calculation without array formula, then do it with array formula to check accuracy
- They need to be handled differently in the spreadsheet
  - If an array formula returns multiple cells, you can't edit part of the array