Session 6: Data sources, details, and derivations (oh my!)

Cat Anderson

Outline

Data sources:

- Tongan data:
  - Census data
  - Births data
  - Deaths data
  - Migration data

$$P_{t+1} = P_t + \underbrace{B - D}_{\text{Natural increase}} + \underbrace{I - O}_{\text{Net migration}}$$

- 3rd party sources
- Open up for discussion

Tongan data sources

Tongan data sources – census data

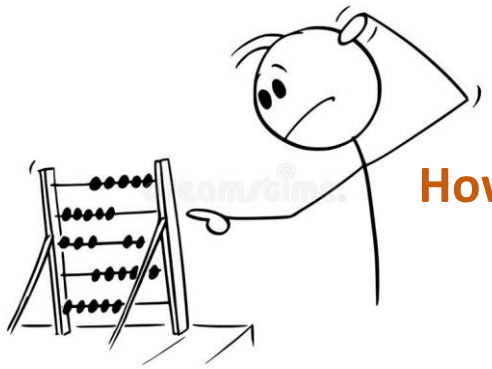
Census data is a count of the number of people where Tonga is their **place of usual residence.**

Census data is a good starting point for calculating the **base population.**

Census count is as at midnight of the census date, 30th November.

Base population counts can be:

- Annually
- Half-yearly
- Quarterly
- Monthly
- Lots of different dates!

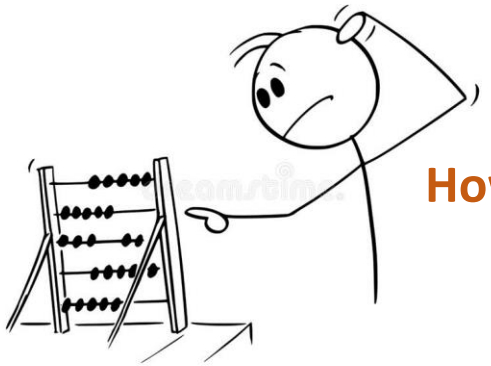


How do we derive the population at a different date?

Tongan data sources – births and deaths data

Births data can be grouped by **date of birth** OR **date of registration.**

Deaths data can be grouped by **date of death** OR **date of registration.**



Time between registration and occurrence varies

How do we account for a lag between occurrence and registration?

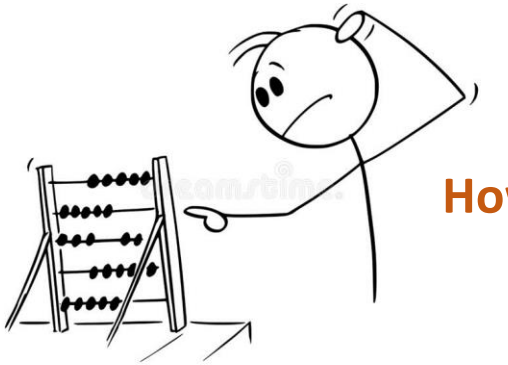
- Release data after complete
- Revise previous data
- Add x% to initial data

Tongan data sources – migration data

Net migration is comprised of two components: **arrivals** and **departures**.

Migration data for defined groups of travellers can follow different trends.

Migration data may need to be broken down into finer detail.



How do we derive data at a finer granularity?

- Applying trends/ratios
- Interpolation
- Modelling

Methods of Interpolation

Interpolation is often based on **mathematical formulae**, or the application of **coefficients**.

$$P^{2018}_{10\text{-}14} = P^{2013}_{5\text{-}9} + (M\text{-}D)^{2014}_{6\text{-}10} + (M\text{-}D)^{2015}_{7\text{-}11} + (M\text{-}D)^{2016}_{8\text{-}12} + (M\text{-}D)^{2017}_{9\text{-}13} + (M\text{-}D)^{2018}_{10\text{-}14}$$

where P = Population at 30 June
M = Net Migration for year ending 30 June
D = Deaths for year ending 30 June

Karup-King third-difference formula:

Karup-King splits the whole population into 3 “panels” (first, middle, last), and then applies a coefficient depending on the position of the single year in its group

- E.g. For the 5-9 year group, 5 is the first position, 6 is the second, 7 is the third etc.

	First panel		
Interpolated Subgroup	G1	G2	G3
1st	0.344	-0.208	0.064
2nd	0.248	-0.056	0.008
3rd	0.176	0.048	-0.024
4th	0.128	0.104	-0.032
5th	0.104	0.112	-0.016

	Middle panel		
Interpolated Subgroup	G1	G2	G3
1st	0.064	0.152	-0.016
2nd	0.008	0.224	-0.032
3rd	-0.024	0.248	-0.024
4th	-0.032	0.224	0.008
5th	-0.016	0.152	0.064

	Last panel		
Interpolated Subgroup	G1	G2	G3
1st	-0.016	0.112	0.104
2nd	-0.032	0.104	0.128
3rd	-0.024	0.048	0.176
4th	0.008	-0.056	0.248
5th	0.064	-0.208	0.344

Methods of Interpolation

Sprague fifth-difference formula:

Sprague splits the whole population into 5 “panels” (first, second from first, middle, second from last, last), and then applies a coefficient depending on the position of the single year in its group

- E.g. For the 5-9 year group, 5 is the first position, 6 is the second, 7 is the third etc.

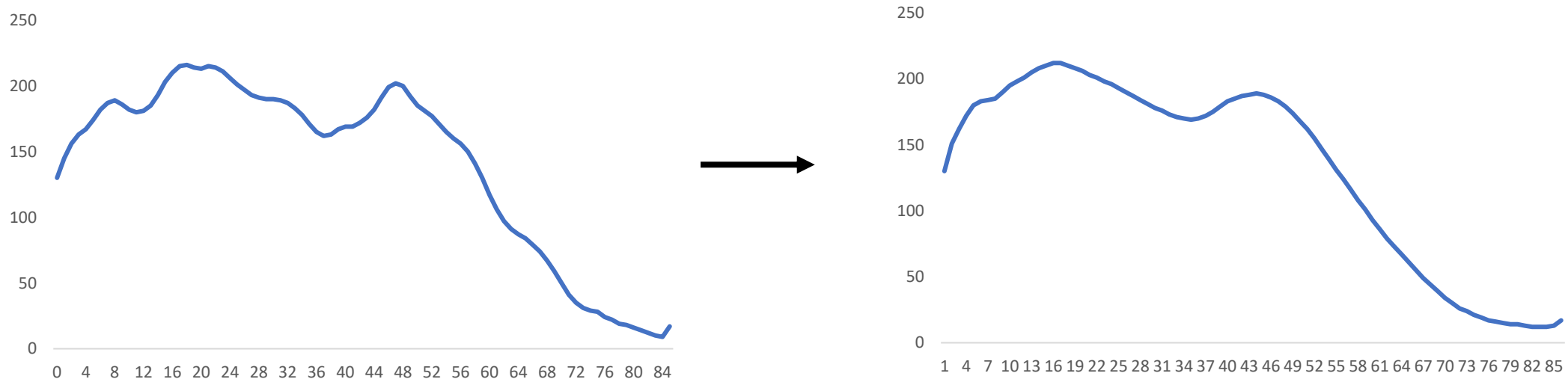
Other formulae:

- Beers Ordinary Formula
- Beers Modified Formula
- Grabill modification of Sprague formula
- And many more!

	First panel				
Interpolated Subgroup	G1	G2	G3	G4	G5
1st	0.3616	-0.2768	0.1488	-0.0336	
2nd	0.2640	-0.0960	0.0400	-0.0080	
3rd	0.1840	0.0400	-0.0320	0.0080	
4th	0.1200	0.1360	-0.0720	0.0160	
5th	0.0704	0.1968	-0.0848	0.0176	
	Second from first panel				
Interpolated Subgroup	G1	G2	G3	G4	G5
1st	0.0336	0.2272	-0.0752	0.0144	
2nd	0.008	0.232	-0.048	0.008	
3rd	-0.008	0.216	-0.008	0	
4th	-0.016	0.184	0.04	-0.008	
5th	-0.0176	0.1408	0.0912	-0.0144	
	Middle panel				
Interpolated Subgroup	G1	G2	G3	G4	G5
1st	-0.0128	0.0848	0.1504	-0.024	0.0016
2nd	-0.0016	0.0144	0.2224	-0.0416	0.0064
3rd	0.0064	-0.0336	0.2544	-0.0336	0.0064
4th	0.0064	-0.0416	0.2224	0.0144	-0.0016
5th	0.0016	-0.024	0.1504	0.0848	-0.0128
	Second from last panel				
Interpolated Subgroup	G1	G2	G3	G4	G5
1st		-0.0144	0.0912	0.1408	-0.0176
2nd		-0.008	0.04	0.184	-0.016
3rd		0	-0.008	0.216	-0.008
4th		0.008	-0.048	0.232	0.008
5th		0.0144	-0.0752	0.2272	0.0336
	Last panel				
Interpolated Subgroup	G1	G2	G3	G4	G5
1st		0.0176	-0.0848	0.1968	0.0704
2nd		0.016	-0.072	0.136	0.12
3rd		0.008	-0.032	0.04	0.184
4th		-0.008	0.04	-0.096	0.264
5th		-0.0336	0.1488	-0.2768	0.3616

Smoothing

Smoothing techniques can be used to make the data flow more smoothly from one age group to the next, and reducing random variation.



Smoothing techniques include:

- Rectangular smoothing
- Exponential smoothing
- Moving averages

Summary

- We are estimating the population at a point in time
- Our components may be incomplete
    - Base – some people may be missing from the census counts
    - Births – some may take a while to be registered or remain unregistered
    - Deaths – some may take a while to be registered or remain unregistered
    - Migration – some travellers may be missing from the data; migrants may be swamped in number by short-term travellers
- Our components may have errors
    - Respondent errors
    - Processing errors
    - Missing and incomplete responses
- All components of population estimation potentially need adjustment

Stats NZ
Tatauranga Aotearoa

www.stats.govt.nz

United Nations

population.un.org

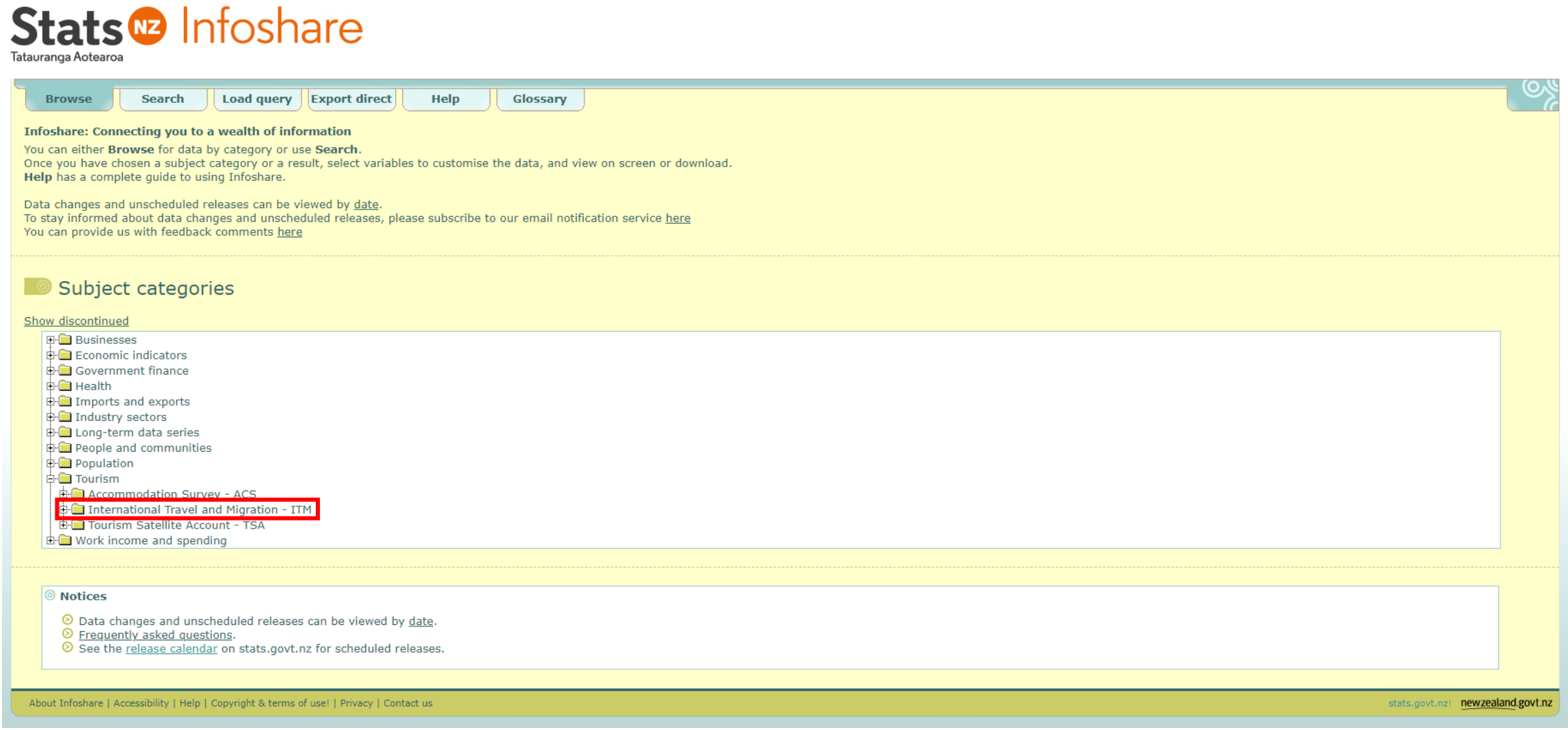
www.unescap.org/stat/data

Pacific Community
Communauté du Pacifique

www.spc.int

3rd party data sources

Extracting information from Infoshare (Stats NZ)



Extracting information from Infoshare (Stats NZ)

- International Travel and Migration - ITM
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Apr)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Aug)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Dec)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Feb)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Jan)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Jul)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Jun)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Mar)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-May)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Nov)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Oct)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Annual-Sep)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Monthly)
  - Estimated migrant arrivals by citizenship, visa type and CLPR, 12/16-month rule (Qrtly-Mar/Jun/Sep/Dec)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Apr)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Aug)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Dec)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Feb)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Jan)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Jul)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Jun)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Mar)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-May)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Nov)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Oct)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Annual-Sep)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Monthly)
  - Estimated migrant arrivals by visa type, 12/16-month rule (Qrtly-Mar/Jun/Sep/Dec)
  - Estimated migrant arrivals to Australia by CLPR, citizenship and birthplace (Qrtly-Mar/Jun/Sep/Dec)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Apr)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Aug)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Dec)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Feb)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Jan)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Jul)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Jun)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Mar)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-May)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Nov)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Oct)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Sep)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Monthly)
  - Estimated migration by direction and country of citizenship, 12/16-month rule (Qrtly-Mar/Jun/Sep/Dec)
  - Estimated migration by direction, 12/16-month rule (Annual-Apr)

Extracting information from Infoshare (Stats NZ)

Stats NZ Tatauranga Aotearoa

Stats NZ Tatauranga Aotearoa Infoshare

Browse | Search | Load query | Export direct | Help | Glossary

Select variables

Group: International Travel and Migration - ITM

Table: Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Dec)

Mark your selections and choose between table on screen and file format. Help
For variables marked you need to select at least one value.

Travel Direction
Select all | Clear selection |
Total: 3. Selected: 2
Arrivals
Departures
Net
Search Find Text start

Citizenship
Select all | Clear selection |
Total: 46. Selected: 1
New Zealand
Australia
Fiji
Tonga
Samoa
China, People's Republic of
Hong Kong (Special Administrative Region
Search Find Text start

Estimate, 12/16-month rule
Select all | Clear selection |
Total: 4. Selected: 1
Estimate
Seasonally adjusted
Trend
Standard error of estimate
Search Find Text start

Time
Select all | Clear selection |
Total: 21. Selected: 1
2021
2020
2019
2018
2017
2016
2015
Search Find Text start

Number of selected data rows 1
Number of selected data columns 2
Number of selected cells 2
Table on screen is limited to 50,000 cells.
Downloaded files are limited to 100,000 cells.

Options Help Table on screen Help Go

ts.govt.nz

Extracting information from Infoshare (Stats NZ)

Stats NZ Tatauranga Aotearoa

Stats NZ Tatauranga Aotearoa Infoshare

Browse | Search | Load query | Export direct | Help | Glossary

View table

Group: International Travel and Migration - ITM

Table: Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Dec)

Edit table: --Please select-- Save table: Table on screen

Footnotes | Information | Print table | Help | Status flags

Estimated migration by direction and country of citizenship, 12/16-month rule (Annual-Dec)		
	Estimate	
	2021	
	Arrivals	Departures
Tonga	276	687

Table information:
Units:
Number, Magnitude = Units

Footnotes:
The standard error indicates the uncertainty of the provisional migration estimate. For example, the 95 percent confidence interval is +/-1.96 standard errors.
These intervals reflect the model uncertainty, not the extent of future revisions to provisional data. Standard errors are non-additive.
Owing to rounding, individual figures may not sum to stated totals.

Symbols:
.. figure not available
C: Confidential
E: Early Estimate
P: Provisional
R: Revised
S: Suppressed

Status flags are not displayed

Table reference:
ITM454AA

Last updated:
20 January 2023 10:45am

Source: Statistics New Zealand
Contact: Information Centre
Telephone: 0508 525 525
Email:info@stats.govt.nz

Questions?

Malo ‘aupito!

Thanks for listening!

Session 6 (cont.): Practical exercises for deriving data

Cat Anderson & Kim Dunstan

Exercise 1: Deriving Births and Deaths Data

Time between registration and occurrence varies

→ Need to account for lag

Lag Data: "Workshop Materials – Births and Deaths lag data.xlsx"

- 2020H2 has incomplete data.
  1. Using the lag data, calculate a plausible births estimate for 2020H2, for males, females, and total sex.
  2. Do the same for deaths x sex x SYOA in 2021H1.

- No data is available for 2021.
  1. Using previous year and half-year trends, calculate plausible births estimates for 2021H1 and 2021H2.
  2. Do the same for deaths x sex x SYOA in 2021H2.

Exercise 2: Deriving SYOA from 5-year age groups

We have data for sex and SYOA for base population, births, and deaths.

We only have data for sex for migration.

- Derive SYOA data for arrivals and departures for each sex (male, female, total) using:
    1. Previous year trends for males, females, and total sex (“2019 SYOA data.xlsx”).
    2. The Karup-King method of interpolation (“Component Data.xlsx”, sheet = “Karup-King”).
    3. The Sprague method of interpolation (“Component Data.xlsx”, sheet = “Sprague”).

- Calculate net migration x sex x SYOA for all periods.