Demographic Methods

Monica Alexander

Week 1: Introduction

Overview

- Course structure and goals
- What is demography?
- Some important introductory demographic concepts
- Getting started with R

Palaver

Course structure and goals

- 6 week course
- Short project using a real dataset
- Introduce you to demographic methods
  - focus is on methods rather than substantive areas
- Get comfortable with R
- Learn about different sorts of datasets
- Illustrate real-world issues that you can help solve

Course materials:
https://github.com/MJAlexander/demographic-methods

Expectations

- Most readings are recommended, not compulsory
- Don't worry if no R experience, plenty of support materials
- Project should be something of interest to you (use your own data, etc)
  - improve data manipulation and graphing in R
  - research reproducibility (RMarkdown)
  - apply something demographic-y that we've covered in the course
  - short write up and presentation in week 6

Roadmap

1. Intro
2. Mortality
3. Fertility
4. Population projections
5. Migration, kinship, other
6. Project presentations

Hello!

Me:

- statistics $\cap$ chemistry —> social science $\cap$ statistics
- Not Canadian

What I work on: a mix of demography, applied stats, epi and computational social science

Hello!

What I work on:

1. Demographic methods (mostly Bayesian, but other stuff too)
   - estimates with no/bad data
   - age heaping
   - kinship size given fertility / mortality rates
2. Mortality
   - Maternal mortality for WHO (new group! w/ Dan and Patrick)
   - Opioid mortality
   - Seasonality
3. Non-traditional data in demography
   - Facebook for migration trends
   - Twitter and newspaper text data for migration sentiment

What is demography?

What is demography?

Demography is the scientific study of population dynamics. We are interested the size, composition and distribution of populations over time, and study these changes with respect to the three main population processes:

- Births (Fertility)
- Deaths (Mortality)
- Migration

What is demography?

Demography links the **individual** to the **aggregate**.

- Individual level behavior affected by prevailing social, cultural, environmental, economic conditions
- When aggregating individuals to populations (where a ‘population’ can be any collection of interest), we see clear patterns
- These patterns in turn tell us something about the likely demographic conditions faced by an individual in a population (e.g. life expectancy, fertility rates)

Different types of demography

- Economic demography
  - economic consequences of demographic change
  - Ron Lee, David Lam
- Family / Social demography
  - social / cultural effects on demographic change, how demography affects family and kinship structures
  - Peter Uhlenberg, Jenna Nobles
- Biodemography
  - biological mechanisms for ageing, genetics
  - Shripad Tuljapurkar, Jim Vaupel
- Mathematical demography
  - formal demographic relationships and models
  - Ken Wachter, Sam Preston
- Statistical demography
  - statistical models for demographic processes, agent-based studies
  - Adrian Raftery, Carlo Carmada

Why is demography important: developed countries

Social Security
Office of Policy

SEARCH

Coping with the Demographic Challenge: Fewer Children and Living Longer

by Gayle L. Reznik, Dave Shoffner, and David A. Weaver

Social Security Bulletin, Vol. 66, No. 4, 2005/2006

Why is demography important: developing countries

The Economist explains

Why nobody knows how many Nigerians there are

No census has yet arrived at an accurate figure

BILL GATES AND ALIKO DANGOTE SUPPORT POLIO ERADICATION EFFORTS IN NIGERIA

Gates and Dangote emphasized the need to eradicate polio, strengthen routine immunization, and improve primary health care.

Demographic methods

Useful to model demographic processes, for:

- **Understanding**
- **Generalization**
- **Projection**

Demographic methods

Demographic methods fall into one of two categories:

- **Mathematical models** based on simplifying assumptions, distributional assumptions
  - Can get closed-form estimates of demographic quantities, what is likely to happen in future
- Models based on **empirical regularities**, from observing patterns in real data
  - Can make assumptions and generalizations about demographic processes

We will look at both sorts of models in this course.

Fundamental demographic concepts

The balancing equation of population change

- tracking population size (P) over time (t)
- enter a population: births (B), in-migration (I)
- exit a population: deaths (D), out-migration (O)

So we have the balancing equation, or demographic identity:

$$P(t+1) = P(t) + B[t, t+1] - D[t, t+1] + I[t, t+1] - O[t, t+1]$$

By definition, this must always hold. But issues:

- Data may not exist
- Data come from different sources
- Measurement, coverage error

Demographic rates

- Useful to standardize size of flows (births, deaths, migrants) based on the size of the population that is producing the flows
- Compare to ‘population at risk’
- Exposure has two features:
  - number of people in the population
  - length of time they were exposed to be counted
  - = ‘Person Years’

$$Rate = \frac{\text{Number of events}}{\text{Person-years of exposure to risk of event}}$$

Crude rates

The most simple demographic measures.

Crude birth rate:

$$CBR[t, t{+}1] = \frac{\text{Number of births in population from time } t \text{ to } t+1}{\text{Person-years lived in population from time } t \text{ to } t+1}$$

Crude death rate:

$$CDR[t, t{+}1] = \frac{\text{Number of deaths in population from time } t \text{ to } t+1}{\text{Person-years lived in population from time } t \text{ to } t+1}$$

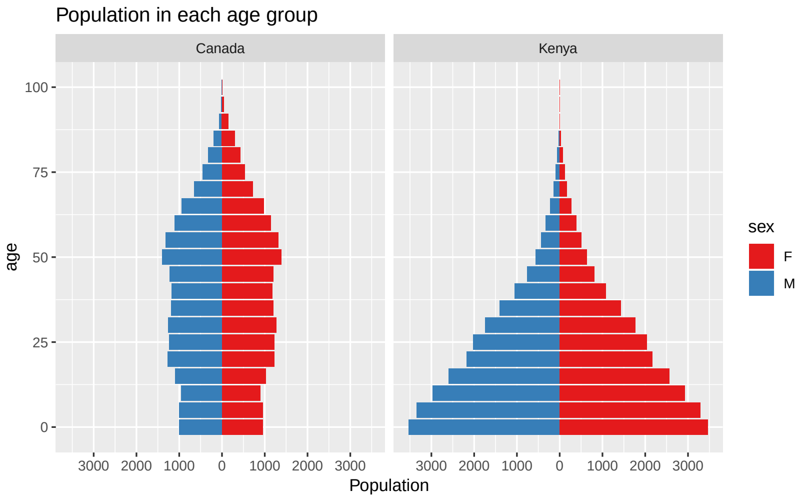
Easy to calculate: just need total counts! (don't need any info on age, sex etc)

Crude rates

- CDR in Canada: 3.8 per 1,000 people (trending up)
- CDR in Kenya: 3.3 per 1,000 people (trending down)

Which country has worse mortality conditions?

Population structures



Coffins and pyramids

Age-specific rates

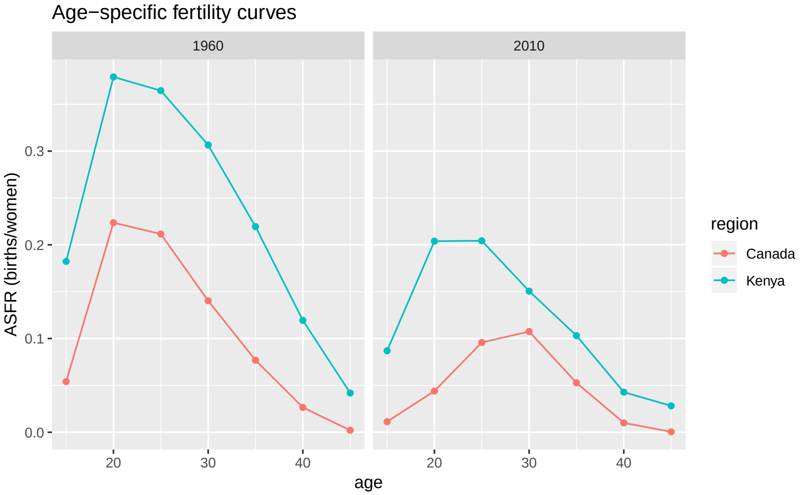
Same deal but by age:

Age-specific fertility rate for people aged x to $x+n$ (for a particular time period):

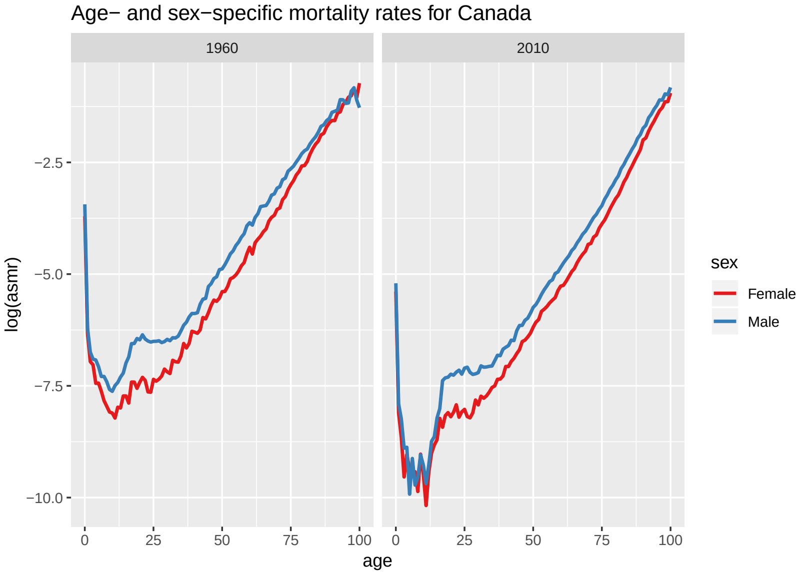
$$ASFR[x, x{+}n] = \frac{\text{Births to people aged } x \text{ to } x+n \text{ to people aged}}{\text{Person-years lived for population aged } x \text{ to } x+n}$$

etc

Characteristic shapes of age-specific rates



Characteristic shapes of age-specific rates



Standardization

- Given age-specific rates and the population structure, there are two things that can affect crude rates.
- We are (usually) interested in the effect of the outcome, not the effect of population age structure
- Pick a population to 'standardize' the population age structure. This has populations $P[x, x+n]$

Then

$$\text{Age-standardized rate} = \sum_{\text{all ages}} \text{rate}[x, x+n] \cdot \frac{P[x, x+n]}{\sum P[x, x+n]}$$

Standardization

For example,

$$\text{Age-standardized mortality rate} = \frac{\sum_{\text{all ages}} \text{ASMR}[x, x+n] \cdot P[x, x+n]}{\sum P[x, x+n]}$$

- CDR in Canada: 3.8 per 1,000 people - CDR in Kenya: 3.3 per 1,000 people

Using Canada's population:

- Age-standardized mortality rate in Canada: 3.8 per 1,000 people
- Age-standardized mortality rate in Kenya: 9.0 per 1,000 people

Age-standardized rates

$$\text{Age-standardized mortality rate} = \frac{\sum_{\text{all ages}} \text{ASMR}[x, x+n] \cdot P[x, x+n]}{\sum P[x, x+n]}$$

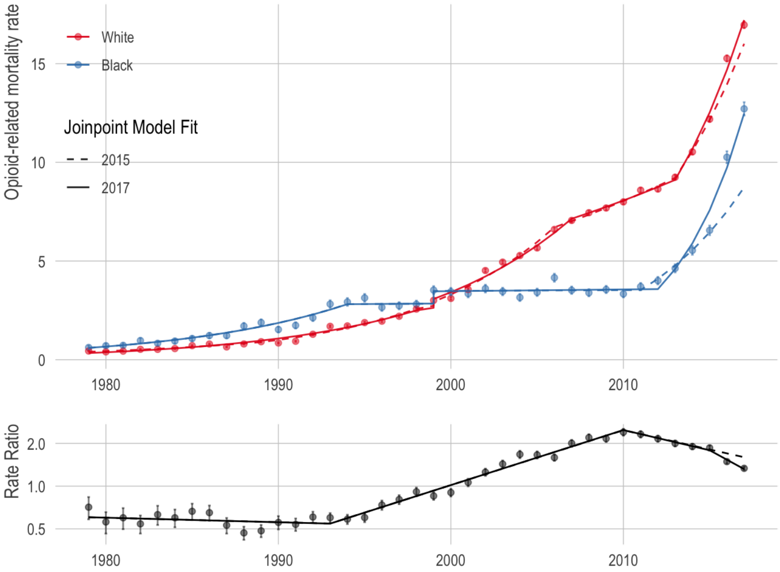
Side note I: if we moved into continuous form, call mortality rates at age x $m(x)$ and population weights $w(x)$, then

$$\text{Age-standardized mortality rate} = \frac{\int_0^\omega m(x)w(x)dx}{\int_0^\omega w(x)dx}$$

where ω is the maximum age. We will see this weighted average form a lot.

Side note II: we will switch between discrete and continuous a fair bit

Standardized mortality rates

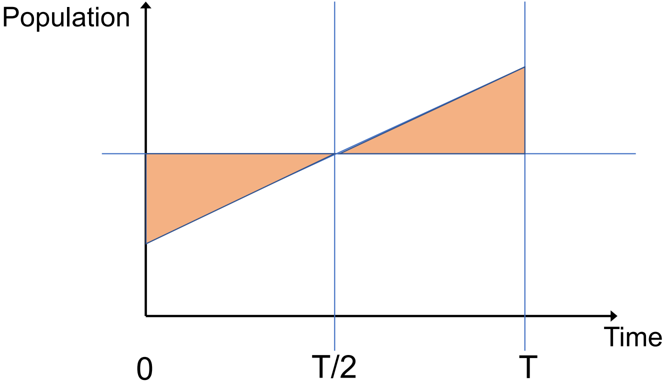


Standardization

- Standardizing age structure is called **direct standardization**
- **Indirect standardization** also exists, more common in epidemiology
  - don't know age-specific rates for a population
  - calculate 'expected deaths' based on another population's mortality
  - look at standardized mortality ratio (observed / expected deaths)

Approximation to person-years

- So far we been using person-years (PY) on the denominator
- But actually almost never have this information
- Usually just have population estimates at one point in time
- Common to use mid-year population times the period length as an **estimate** of PY



Population growth

Crude growth rate

Divide the balancing equation by population years:

$$
\begin{aligned}
\frac{P(t+1) - P(t)}{PY(t)} &= \frac{B(t) - D(t) + I(t) - O(t)}{PY(t)} \\
CGR(t) &= CBR(t) - CDR(t) + CRI(t) - CRO(t) \\
CGR(t) &= CRNI(t) + CRNM(t)
\end{aligned}
$$

Growth rate is natural increase + net migration.

Geometric growth

Let's ignore migration for now.

$$\begin{aligned} P(t+1) &= P(t) + B(t) - D(t) \\ &= P(t)\left(1 + \frac{B(t)}{P(t)} - \frac{D(t)}{P(t)}\right) \end{aligned}$$

So

$$\begin{aligned} P(1) &= P(0)\left(1 + \frac{B(0)}{P(0)} - \frac{D(0)}{P(0)}\right) \\ P(2) &= P(1)\left(1 + \frac{B(1)}{P(1)} - \frac{D(1)}{P(1)}\right) \\ &= P(0)\left(1 + \frac{B(0)}{P(0)} - \frac{D(0)}{P(0)}\right)\left(1 + \frac{B(1)}{P(1)} - \frac{D(1)}{P(1)}\right) \end{aligned}$$

etc

This is called **geometric growth**

Constant growth rate over time

If the birth and death rates (i.e. $b = \frac{B}{P}$ and $d = \frac{D}{P}$) are not changing over time

$$\begin{aligned} P(1) &= P(0)\,(1+b-d) \\ P(2) &= P(1)\,(1+b-d) \\ &= P(0)\,(1+b-d)\,(1+b-d) \\ &= P(0)\,(1+b-d)^2 \end{aligned}$$

In general, $P(t) = A^t P(0)$ where $A = (1+b-d)$.

Instantaneous growth rate

Consider the growth rate $r(t)$ between two times points that are very close together, Δt, and then look at the limit.

$$r(t) = \lim_{\Delta t \to 0} \frac{\Delta P(t)}{P(t)\Delta t} = \frac{\frac{dP(t)}{dt}}{P(t)} = \frac{d \ln(P(t))}{dt}$$

Taking integrals and exponents and doing a bit of rearranging we have the population at time T

$$P(T) = P(0)e^{\int_0^T r(t)dt}$$

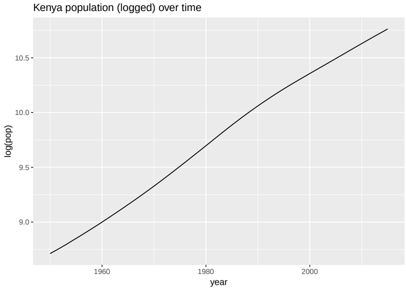
Constant growth rates

If r is constant over time then

$$P(T) = P(0)e^{Tr}$$

And so assuming constant growth and given two time points, we can calculate the implied growth rate as

$$r = \frac{log(P(t_2)) - log(P(t_1))}{t_2 - t_1}$$



Age, periods cohorts

Three dimensions of demographic time

We can express an individual's relative position in time based on three different dimensions:

- their age
- the period (year) we are in
- their cohort (e.g. birth cohort)

Demographers, sociologists, epidemiologists often interested in one, some or all of these effects on outcomes

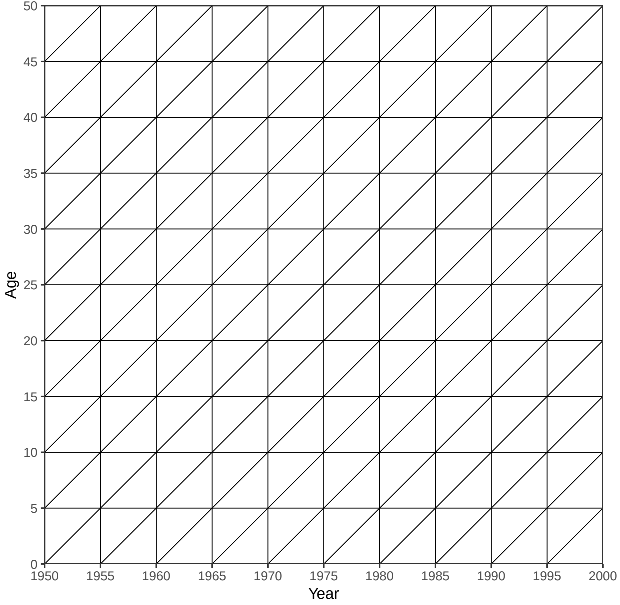
Age effect


ALEXANDER, MONICA 16.11.1986
C1-5-RS/OB
MI 0.9
TNI BAY ST. RM 3
1743521 GA=20w2d
13.3cm / 1.6 / 35Hz
TIb 0.3
21.11.2018 11:25:43 AM
Routine
Har-mid
93
Gn 0
C6 / M4
P4 / E3
SRI II 3
Voluson
S8

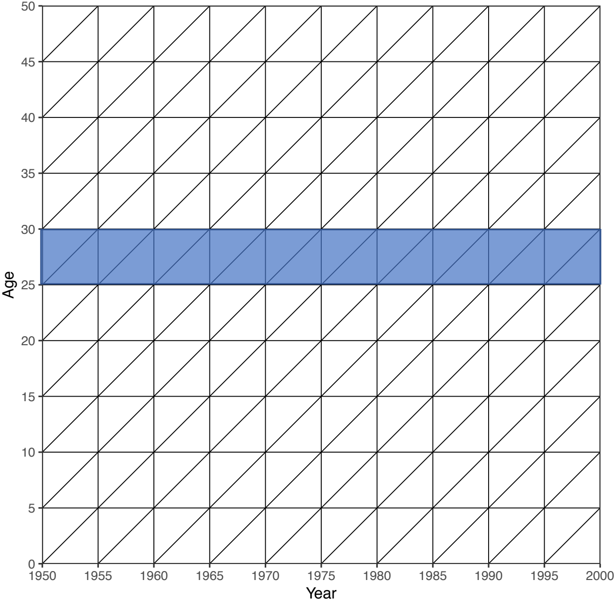
Period effect

Cohort effect

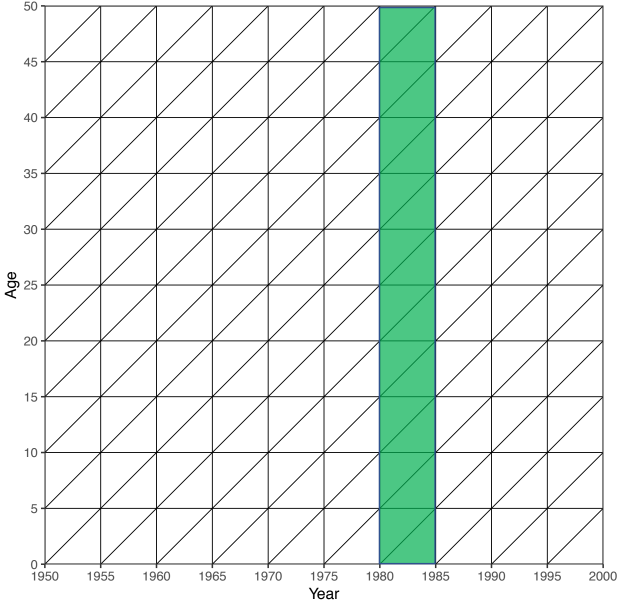
Lexis diagrams



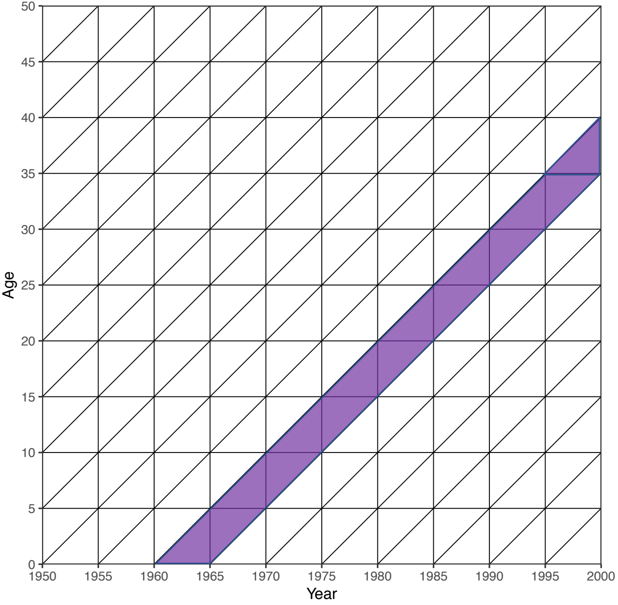
Lexis diagrams



Lexis diagrams



Lexis diagrams



Measurement issues

May want to estimate age, period and cohort effects on outcome of interest (e.g. smoking)

$$Y_i = \alpha_i + \beta_i + \gamma_i$$

But

$$Age = Period - Cohort$$

etc

so regression matrix is singular (not identifiable)

- APC models is a big research area
- e.g. Ethan Fosse here at UofT
- Bayesian approaches seem under-utilized (future work??)

Intro to R