COS-R403. Special Research Methods. Forecasting I: Introduction

Lecturer: Christina Bohk-Ewald

Day 2 of intensive 5-day course

University of Helsinki, Finland
04.05.2020–08.05.2020

Second day's class:

Introduction to demographic forecasting

- Recap main concepts of last lecture
- Some findings of previous lab session
- Analyze forecasted population structure by age and sex
- Forecast population with the cohort-component method

Second day's class in the lab: Hands-on exercises in demographic forecasting with R

- Analyze global population structure by age and sex, 1950–2100,
  - Population (age) pyramid
  - Sex ratio
  - Mean age & dependency ratio
- Do this again for an assigned world region and put it into your report.
- Implement cohort-component method to project population with three age groups (slide 38).
  - Population size and structure after 1, 10, 100 years
  - Constant growth rate & constant age structure
- Present and discuss your findings in class.

Recap main concepts of last lecture

- What is the difference between projections and forecasts?
- What is the jump-off year?
- What are the base period and the forecast horizon?
- What is the difference between deterministic and probabilistic forecasts?
- What are typical sources of error in forecasting?
- ...

$\Rightarrow$ Questions?

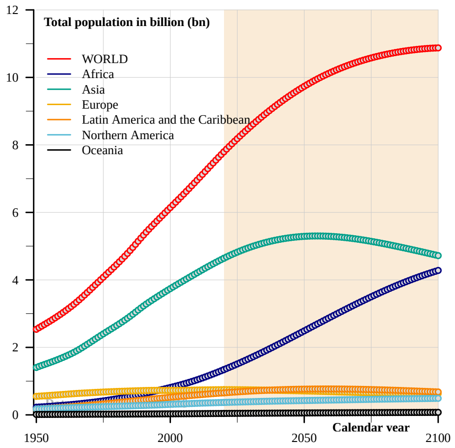
Previous lab session

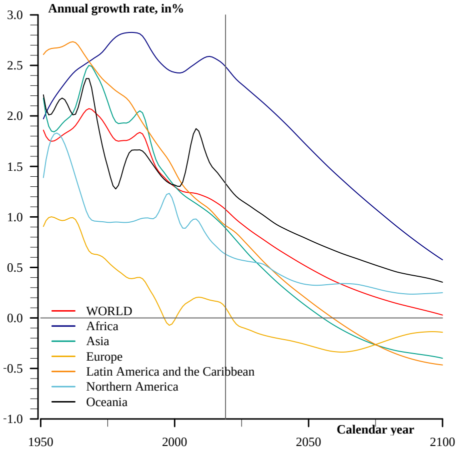
What can you tell us about your world region with respect to population size, life expectancy, and TFR? How might this relate to the global development?

- Africa
- Asia
- Europe
- Latin America and the Caribbean
- Oceania
- North America

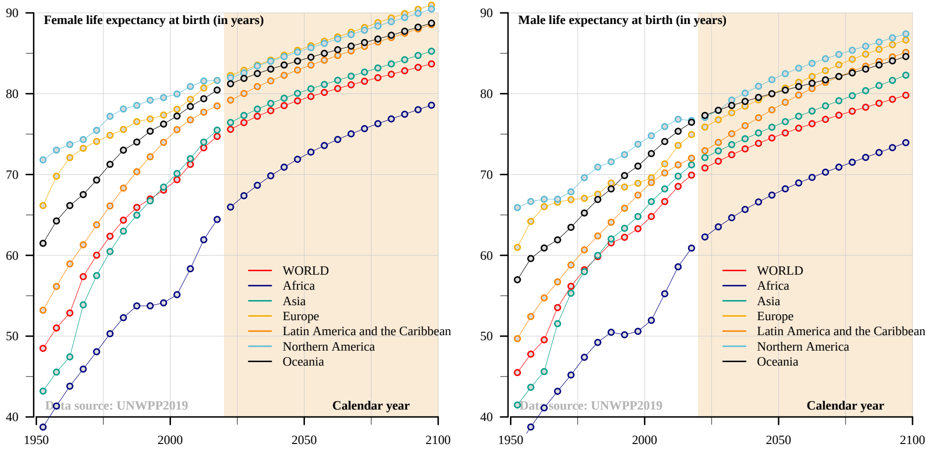
$\Rightarrow$ Questions?

Total population size by world region

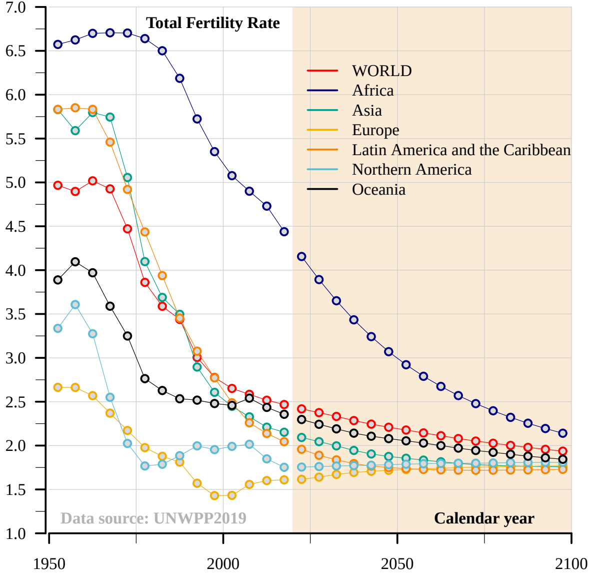




Life expectancy at birth by world region



Total fertility rate by world region



UNWPP2019 - key findings

World Population Prospects 2019: Highlights

June 2019 • https://population.un.org/wpp • #UNPopulation

Within little more than a decade there are likely to be around 8.5 billion people on earth, and almost 10 billion by 2050, compared to 7.7 billion today. A small number of countries will account for most of the increase. While some countries continue to grow rapidly, others are seeing their populations decline. At the same time, the world is growing older, as global life expectancy continues to rise and the fertility level continues to fall. Such changes in the size and distribution of the world's population have important consequences for achieving the Sustainable Development Goals (SDGs) and ensuring that no one is left behind.

---

https://population.un.org/wpp/Publications/Files/WPP2019_10KeyFindings.pdf

UNWPP2019 - key findings

2. Nine countries will make up more than half the projected population growth between now and 2050

The largest increases in population between 2019 and 2050 will take place in: India, Nigeria, Pakistan, the Democratic Republic of the Congo, Ethiopia, the United Republic of Tanzania, Indonesia, Egypt and the United States of America (in descending order of the expected increase). Around 2027, India is projected to overtake China as the world's most populous country.

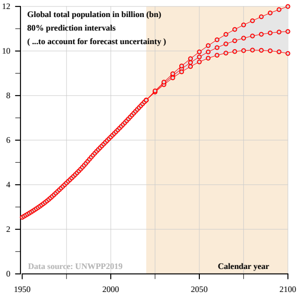
3. Rapid population growth presents challenges for sustainable development

Many of the fastest growing populations are in the poorest countries, where population growth brings additional challenges in the effort to eradicate poverty (SDG 1), achieve greater equality (SDGs 5 and 10), combat hunger and malnutrition (SDG 2), and strengthen the coverage and quality of health and education systems (SDGs 3 and 4).

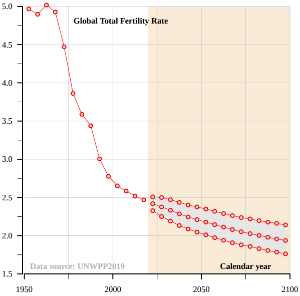
https://population.un.org/wpp/Publications/Files/WPP2019_10KeyFindings.pdf

So far we looked at aggregated measures

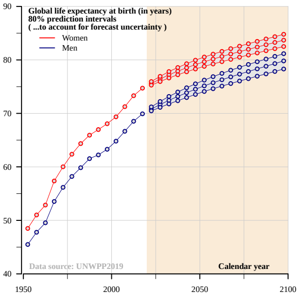
Total population



Total fertility rate



Life expectancy at birth

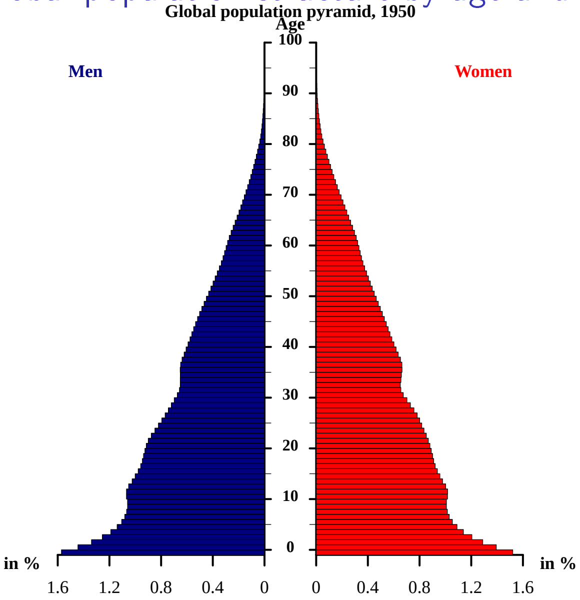


$\longrightarrow$ Disaggregate to analyze population structure by age and sex

Global population structure by age and sex, 1950

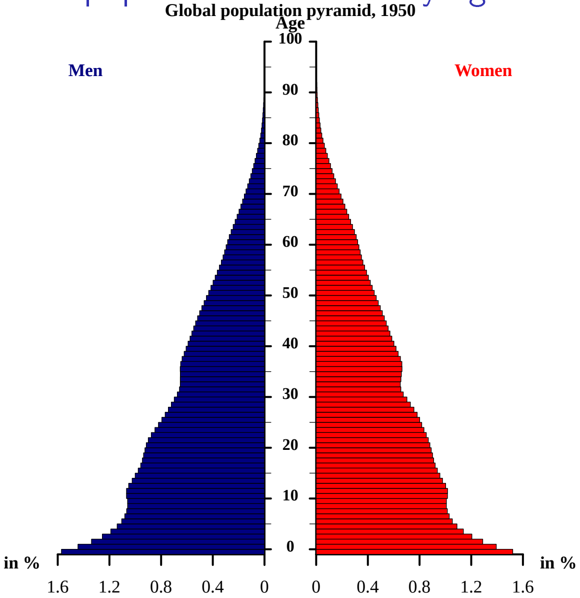
What can impact population structure by age and sex?

Global population structure by age and sex, 1950



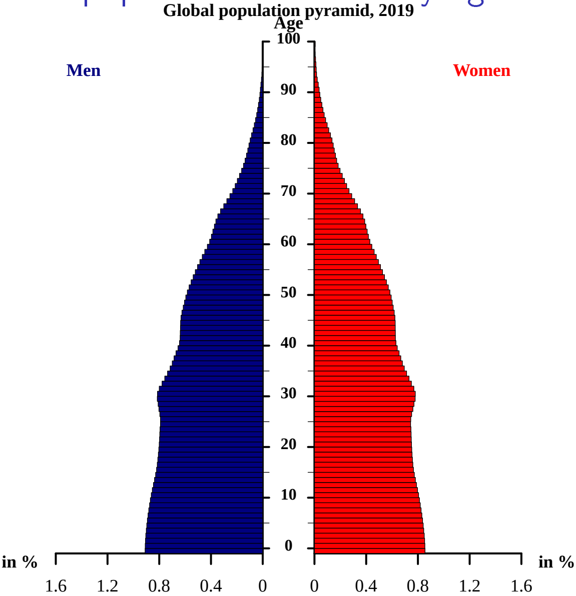
- In percent for sake of comparability.
-

Global population structure by age and sex, 1950



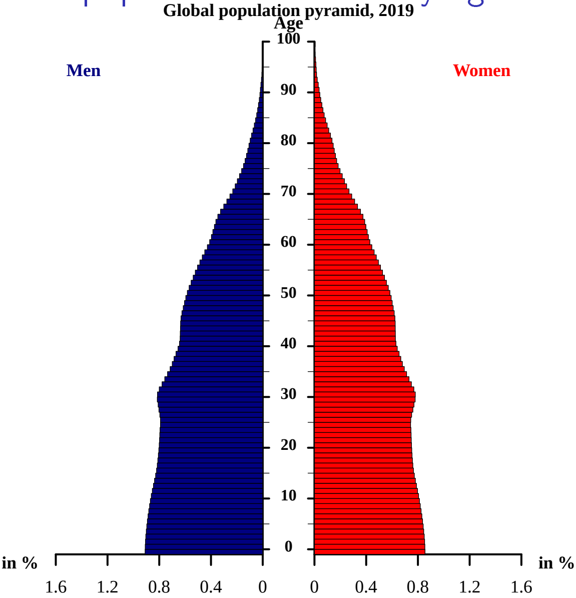
- In percent for sake of comparability.
- Young population with small shares of older people.
- Note highest attainable age: almost 90.
- Higher fertility: TFR is almost 5.
- Higher mortality: e_0 is 48 for women, 45 for men.

Global population structure by age and sex, 2019



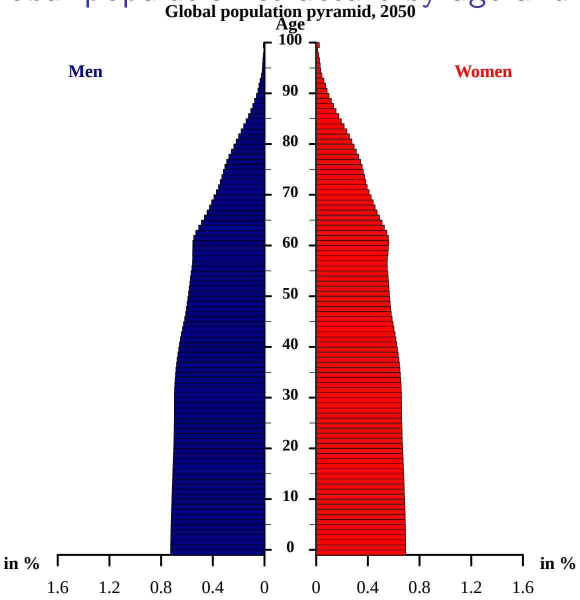
- In percent for sake of comparability.
-

Global population structure by age and sex, 2019



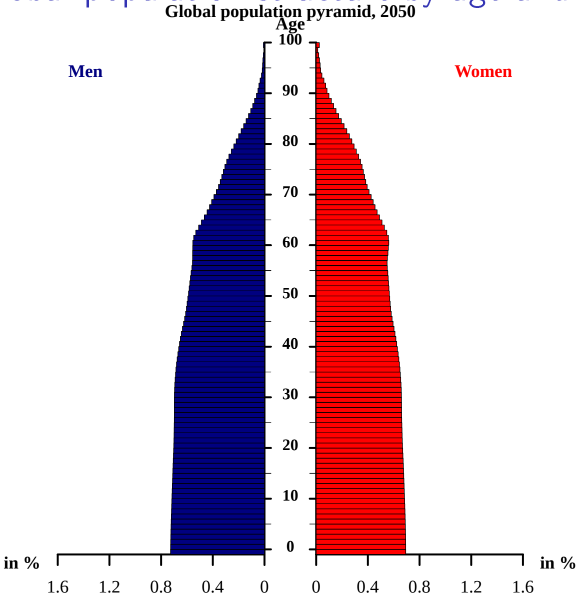
- In percent for sake of comparability.
- Older population with larger shares of older people.
- Note highest attainable age: almost 95.
- Lower fertility: TFR is almost 2.5.
- Lower mortality: e_0 is 75 for women, 70 for men.

Global population structure by age and sex, 2050



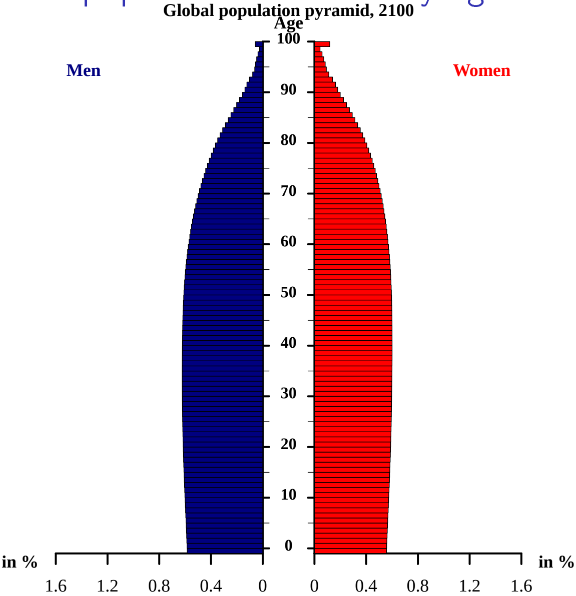
- In percent for sake of comparability.
-

Global population structure by age and sex, 2050



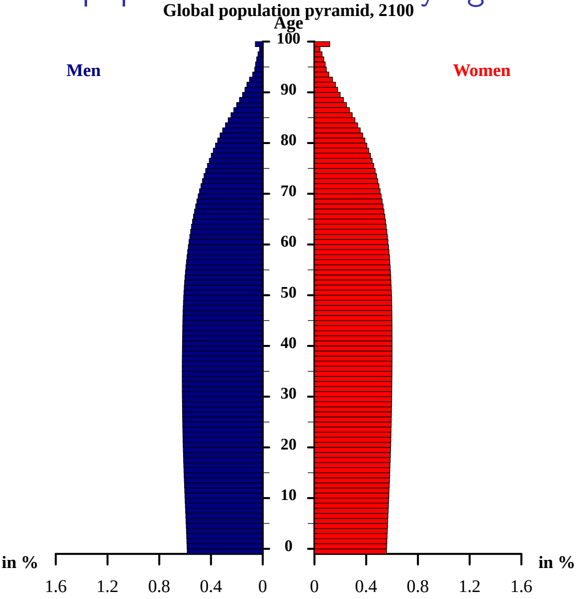
- In percent for sake of comparability.
- Older population with larger shares of older people.
- Note highest attainable age: almost 100.
- Lower fertility: TFR is almost 2.2.
- Lower mortality: e_0 is 79 for women, 75 for men.

Global population structure by age and sex, 2100



- In percent for sake of comparability.
-

Global population structure by age and sex, 2100



- In percent for sake of comparability.
- Old population with large shares of old people.
- Note highest attainable age: almost 100+.
- Lower fertility: TFR is almost 1.9.
- Lower mortality: e_0 is 84 for women, 80 for men.

Global population structure by sex

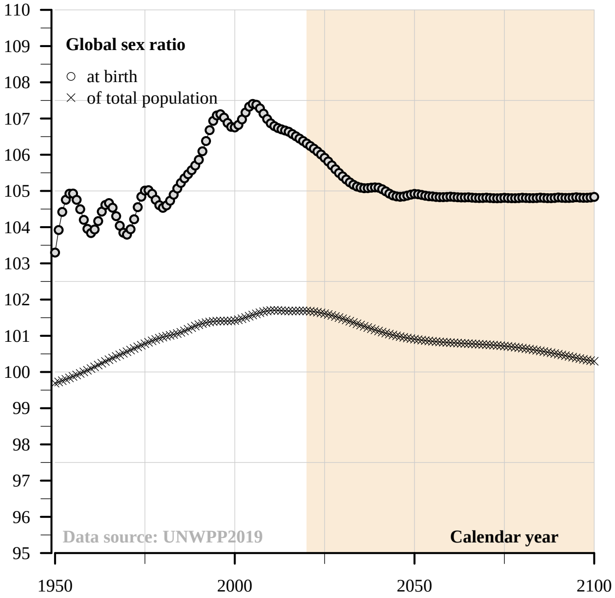
Measures to quantify population structure by sex:

- Sex ratio: $\frac{P^{males}}{P^{females}}$
- Sex ratio at birth: $\frac{P_0^{males}}{P_0^{females}}$
- Share of men in population: $\frac{P^{males}}{(P^{females}+P^{males})}$
- ...

→ Range of values and interpretation

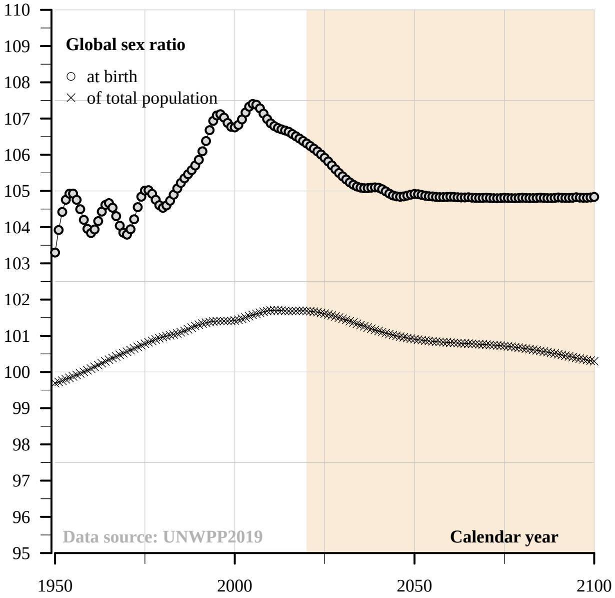
→ Related to societal relevance of demographic forecasts

Global population structure by sex



- Sex ratio above 100 indicates men outnumber women

Global population structure by sex



- Sex ratio above 100 indicates men outnumber women
- Sex ratio at birth substantially larger than for entire population
- Sex ratio at birth is forecasted to fluctuate around 105
- Sex ratio of total population is forecasted to slightly decline

Global population structure by sex

Sex ratio at birth:

- Sex ratio at birth is expected to be 105 [103,107], (it is more likely to have a boy than a girl)
- Sex ratio at conception appears to be 100
- Mortality during pregnancy appears to be higher for girls than for boys
- Probability of miscarriage varies by gender and course of pregnancy
- Selective abortion practices can bias sex ratio further

Ritchie 2019 (https://ourworldindata.org/sex-ratio-at-birth)

Orzack et al. 2015 (https://pdfs.semanticscholar.org/3986/f62b66ee8450859909e122fb2974c1c9fe6b.pdf)

Global population structure by age

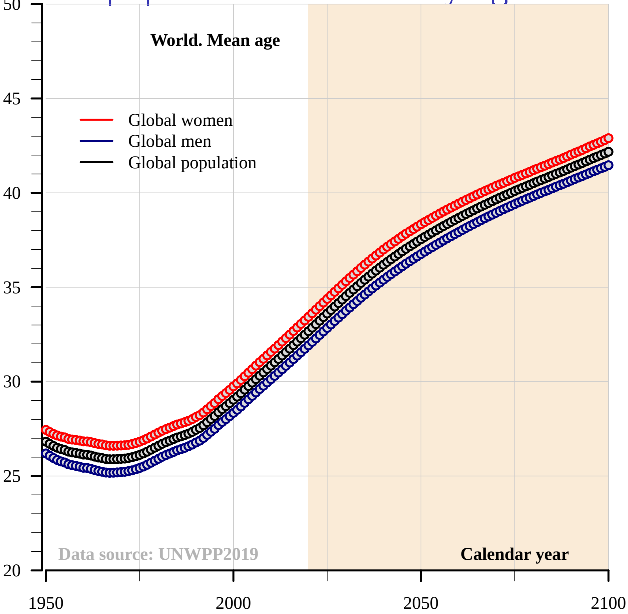
Retrospective measures:

- Mean age: $\frac{x \cdot P_x}{\sum_x P_x}$
- Median age: 50% of population are below / above median age.
- Share of old people: $\frac{P_{65+}}{P}$
- Dependency ratios
  - Old age dependency ratio: $\frac{P_{\mathbf{65}+}}{P_{\mathbf{15}-\mathbf{64}}}$
  - Child dependency ratio: $\frac{P_{\mathbf{0}-\mathbf{14}}}{P_{\mathbf{15}-\mathbf{64}}}$
  - Total dependency ratio: $\frac{P_{\mathbf{0}-\mathbf{14}}+P_{\mathbf{65}+}}{P_{\mathbf{15}-\mathbf{64}}}$
- Potential support ratio: $\frac{P_{15-64}}{P_{65+}}$
- ...

$\longrightarrow$ Range of values and interpretation

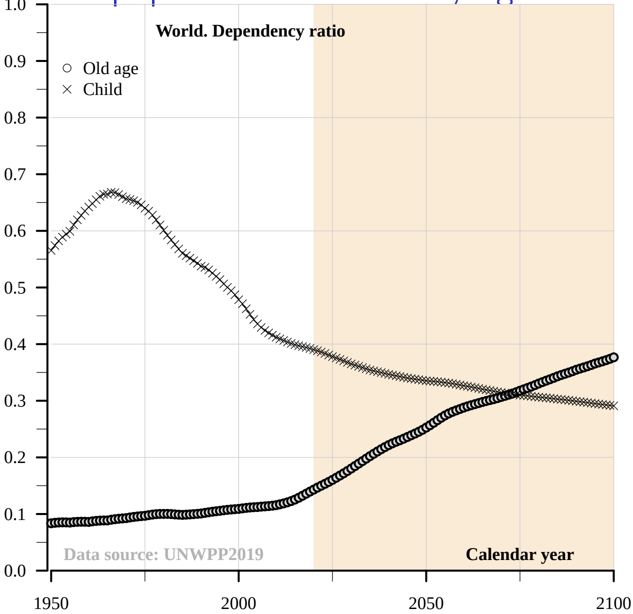
$\longrightarrow$ Related to societal & economic relevance of demographic forecasts

Global population structure by age



- Mean age:
- is forecasted to increase to over 40 years in 2100
- is higher for females than males

Global population structure by age



- Old age dependency ratio and child dependency ratio
- it is forecasted that fewer children will depend on working force, while more older people will depend on working force
- crossover around 2070

Global population structure by age

Prospective age measures:

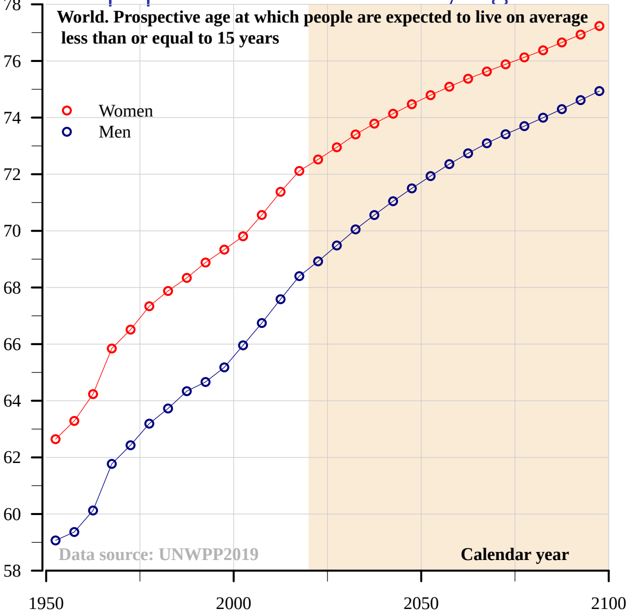
→ Chronological age (life years lived) vs thanatological age (life years left)

- Age at which people live on average 15 or less years: $x_{e_x \leq 15}$
- Population share of people who have 15 or less life years left: $\frac{P_{e_x \leq 15}}{P}$
- Prospective old age dependency ratio: $\frac{P_{e_x < 15}}{P_{e_x > 15}}$
- ...

→ Range of values and interpretation

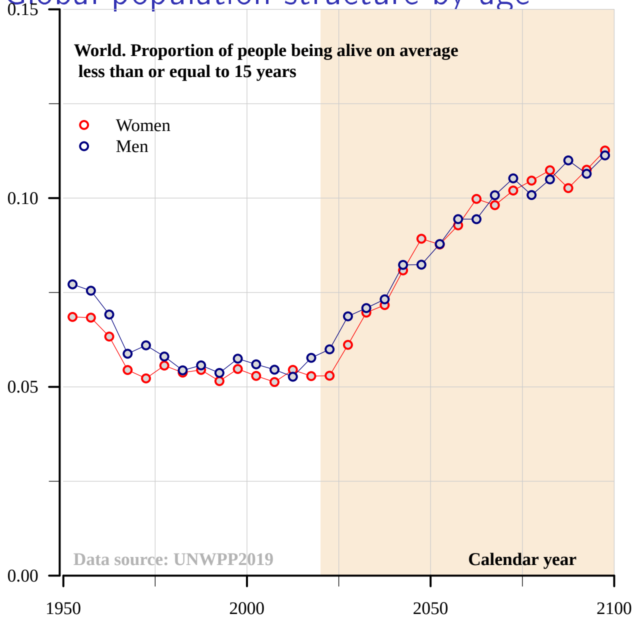
→ Related to societal & economic relevance of demographic forecasts

Global population structure by age



- Prospective perspective
- Age at which people live on average less than or equal to 15 years
- is forecasted to increase to 75 and 77 yrs old for women and men in 2100
- Based on life tables

Global population structure by age



- Prospective perspective
- Proportion of people living on average less than or equal to 15 years
- Based on life tables

→ How to best measure aging?

We now switch

from exploring existing forecasts

to generating forecasts in the first place :-)

Method to forecast population size and structure

Cohort-component method

- Cannan (1891), Whelpton (1928, 1936), Bowley (1924)
- Forecasts population (or subpopulations) by age
  based on age-specific assumptions about future development
  of fertility, mortality, migration
- Based on population balance equation:

$$P_{t+n} = P_t + B_{[t,t+n]} - D_{[t,t+n]} + I_{[t,t+n]} - E_{[t,t+n]}$$

- Implementation with matrix calculus via Leslie's projection matrix (1945)
- Also used by UNWPP 2019

Cohort-component method

$$\begin{pmatrix} P_{0,t+1} \\ P_{1,t+1} \\ \vdots \\ P_{14,t+1} \\ P_{15,t+1} \\ \vdots \\ P_{45,t+1} \\ P_{46,t+1} \\ \vdots \\ P_{100+,t+1} \end{pmatrix}$$

Cohort-component method

$$\begin{pmatrix} P_{0,t+1} \\ P_{1,t+1} \\ \vdots \\ P_{14,t+1} \\ P_{15,t+1} \\ \vdots \\ P_{45,t+1} \\ P_{46,t+1} \\ \vdots \\ P_{100+,t+1} \end{pmatrix} \qquad \begin{pmatrix} P_{0,t} \\ P_{1,t} \\ \vdots \\ P_{14,t} \\ P_{15,t} \\ \vdots \\ P_{45,t} \\ P_{46,t} \\ \vdots \\ P_{100+,t} \end{pmatrix}$$

Cohort-component method

$$
\begin{pmatrix} P_{0,t+1} \\ P_{1,t+1} \\ \vdots \\ P_{14,t+1} \\ P_{15,t+1} \\ \vdots \\ P_{45,t+1} \\ P_{46,t+1} \\ \vdots \\ P_{100+,t+1} \end{pmatrix}
=
\begin{pmatrix}
0 & 0 & \dots & 0 & f_{15} & \dots & f_{45} & 0 & \dots & 0 \\
{}_1p_0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \ddots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\
0 & 0 & \ddots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & {}_1p_{14} & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \dots & \vdots & \ddots & \dots & \vdots & \vdots & \dots & \vdots \\
0 & 0 & \dots & 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & \dots & {}_1p_{45} & 0 & \dots & 0 \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \ddots & \dots & \vdots \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \ddots & p_{100+}
\end{pmatrix}
\cdot
\begin{pmatrix} P_{0,t} \\ P_{1,t} \\ \vdots \\ P_{14,t} \\ P_{15,t} \\ \vdots \\ P_{45,t} \\ P_{46,t} \\ \vdots \\ P_{100+,t} \end{pmatrix}
$$

Cohort-component method

$$
\begin{pmatrix} P_{0,t+1} \\ P_{1,t+1} \\ \vdots \\ P_{14,t+1} \\ P_{15,t+1} \\ \vdots \\ P_{45,t+1} \\ P_{46,t+1} \\ \vdots \\ P_{100+,t+1} \end{pmatrix}
=
\begin{pmatrix}
0 & 0 & \dots & 0 & f_{15} & \dots & f_{45} & 0 & \dots & 0 \\
{}_1p_0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \ddots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\
0 & 0 & \ddots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & {}_1p_{14} & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \dots & \vdots & \ddots & \dots & \vdots & \vdots & \dots & \vdots \\
0 & 0 & \dots & 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & \dots & {}_1p_{45} & 0 & \dots & 0 \\
\vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \ddots & \dots & \vdots \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \ddots & p_{100+}
\end{pmatrix}
\cdot
\begin{pmatrix} P_{0,t} \\ P_{1,t} \\ \vdots \\ P_{14,t} \\ P_{15,t} \\ \vdots \\ P_{45,t} \\ P_{46,t} \\ \vdots \\ P_{100+,t} \end{pmatrix}
$$

$$
P_{t+1} \quad = \quad L \quad \cdot \quad P_t
$$

Cohort-component method

Mini example for population with 3 age groups:

$$\begin{pmatrix} ? \\ ? \\ ? \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Leslie matrix

- Fertility in first row (non-zero for age groups 2 & 3)
- Survival probabilities on subdiagonal

Population vector

- In jump-off year t_0: 1 person per age group
- How many people will be there in t_1?

Cohort-component method

Mini example for population with 3 age groups:

$$\begin{pmatrix} 3 \\ 0.8 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Leslie matrix

- Fertility in first row (non-zero for age groups 2 & 3)
- Survival probabilities on subdiagonal

Population vector

- In jump-off year t_0: 1 person per age group
- How many people will be there in t_1?

Cohort-component method

- Time-invariant assumptions on fertility and mortality:

$$P_{t+n} = L^n \cdot P_t$$

- Time-variant assumptions on fertility and mortality:

$$P_{t+4} = L_4 \cdot L_3 \cdot L_2 \cdot \underbrace{L_1 \cdot P_t}_{P_{t+1}}$$

Cohort-component method

Considering...

- Net migration at the end of each projection interval:

$$P_{t+1} = L \cdot P_t + NM_t$$

- Net migration at the beginning of each projection interval:

$$P_{t+1} = L \cdot (P_t + NM_t)$$

- Gross migration with specific assumptions for immigrants and emigrants:

$$P_{t+1} = N \cdot L^N + I \cdot L^I - E \cdot L^E$$

$\Rightarrow$ Bohk et al. (2012, 2009),

https://bitbucket.org/Christina_Bohk/p3j/wiki/Home

Cohort-component method

Properties of the Leslie (projection) matrix

- Time-invariant assumptions on fertility and mortality:

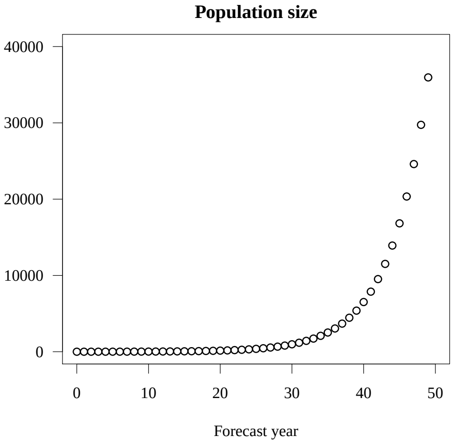
$$P_{t+n} = L^n \cdot P_t$$

  - Constant growth rate λ
  - Constant age structure
  - ...

Cohort-component method

Mini example for population with 3 age groups:

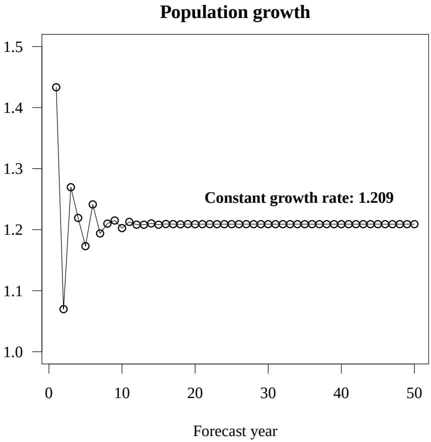
$$\begin{pmatrix} 3 \\ 0.8 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$



Cohort-component method

Mini example for population with 3 age groups:

$$\begin{pmatrix} 3 \\ 0.8 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$



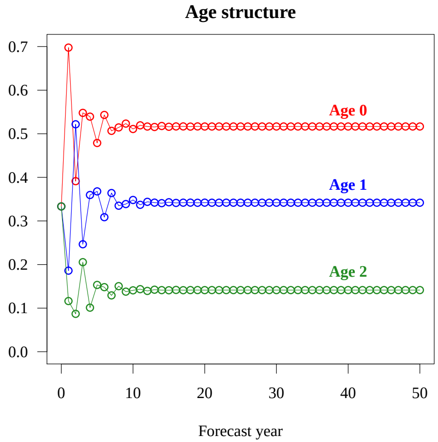
Cohort-component method

Mini example for population with 3 age groups:

$$\begin{pmatrix} 3 \\ 0.8 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Constant age structure:

Age 0: 0.5167014+0i
Age 1: 0.3419012+0i
Age 2: 0.1413974+0i



What you have learned today about demographic forecasting

- Describe structure of population age pyramid
- Explain metrics to quantify population structure by age and sex
- Describe basic procedure of cohort-component method
- Project population size and structure with cohort-component method (using matrix calculus)
- Within Leslie matrix framework: derive constant growth rate and constant age structure based on time-invariant vital rates

Second day's class in the lab: Hands-on exercises in demographic forecasting with R

- Analyze global population structure by age and sex, 1950–2100,
  - Population age pyramid
  - Sex ratio
  - Mean age & dependency ratio
- Do this again for your assigned world region and put it into your report.
- Implement cohort-component method to project population with three age groups (slide 38).
  - Population size and structure after 1, 10, 100 years
  - Constant growth rate & constant age structure
- Present and discuss your findings in class.

Course learning materials

Course learning materials on GitHub:

https://github.com/christina-bohk-ewald/2020-course-COS-R403-forecasting-1-introduction

R programming

Some functions we will use:

- matrix(), nr=, nc=, rbind(), cbind()

$\longrightarrow$ Get information about what they are and how to use them

Recommended learning material for today's class

- **Christina Bohk (2012)**
  Ein probabilistisches Bevölkerungsprognosemodell. Entwicklung und Anwendung für Deutschland
  Springer VS
  `https://link.springer.com/book/10.1007%2F978-3-531-19267-3`
- **Rau, R., Bohk-Ewald, C., Muszyńska, M. M. and Vaupel, J. W. (2017)**
  Visualizing mortality dynamics in the Lexis diagram
  The Springer Series on Demographic Methods and Population Analysis, 44. Cham, Springer International Publishing AG.
- **Alho, J. and Spencer, B. (2006)**
  Statistical demography and forecasting
  Springer Science & Business Media.
- **Preston, S., Heuveline, P., and Guillot, M. (2000)**
  Demography. Measuring and modeling population processes
  Blackwell Publishers Ltd.

Recommended learning material for today's class

- **UNWPP2019**: https://population.un.org/wpp/
  Publications, Graphs, & Data files.
- **Bohk, C., Ewald, R., and Uhrmacher, A. M. (2009)**
  Probabilistic population projection with JAMES II.
  Proceedings of the 2009 Winter Simulation Conference (WSC), Austin, TX, USA, pp. 2008–2019.
  https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5429715
- **Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., and Heilig, G. K. (2012)**
  Bayesian probabilistic population projections for all countries.
  Proceedings of the National Academy of Sciences, 109(35), 13915–13921.
- **UNWPP2019**: World Population Prospects 2019: Highlights.
  https://population.un.org/wpp/Publications/Files/WPP2019_10KeyFindings.pdf

Thank you for your attention!

christina.bohk-ewald@helsinki.fi

Second day's class in the lab: Hands-on exercises in demographic forecasting with R

- Analyze global population structure by age and sex, 1950–2100,
  - Population age pyramid
  - Sex ratio
  - Mean age & dependency ratio
- Do this again for your assigned world region and put it into your report.
- Implement cohort-component method to project population with three age groups (slide 38).
  - Population size and structure after 1, 10, 100 years
  - Constant growth rate & constant age structure
- Present and discuss your findings in class.