

Population projections vary in complexity. Simplest one is extending by hand a line of a graph showing how the total population changes over time. The most complex one may be a computer program for projecting population by age, sex, marital status, and other characteristics using varying assumptions for different periods about the future courses of mortality, fertility, migration, marriage, etc.

The purposes of projections

- Answer to if ...
 - What would happen if a cure for some important cancers were to be discovered?
 - What would happen if the age of marriage were sharply risen?
 - What would happen if the family planning program were implemented?
 - Frejka (1973) estimated the size of ultimate population if fertility were to fall to an NRR of 1.0 in various speeds.
- Planning
 - National level: Most central government departments need population projections as an input to their estimates of future income and expenditure, the level of demand for their services (like pensions or health services)
 - Regional level: Growth of towns and cities caused by rural-to-urban migration or differential natural increase must be projected to prepare housings, transport and other services.
- The actions of planners are influenced by the existing projections, which in turn resulted in actual population change. To use best available data for projection and to present the projections with uncertainty and interactions involved are important.
- Projections made using realistic assumptions are called "forecasts". It's important to remember that forecasts are usually no more than educated guesses dressed up in sophisticated ways. If the guesses ("assumptions") are wrong, then the forecasts will be wrong.

Techniques of projection (discussions below is only targeted to population projection, but applicable to births, marriages, deaths and hospital admissions, similarly)

- The simplest one based on $P_{t+1} = f(P_t)$ is exponential growth, $P_{t+1} = P_t \cdot e^r$, $P_t = P_0 \cdot e^{rt}$ (r is growth rate): $dP/dt=rP \leftrightarrow dP/P=r dt \leftrightarrow \int dP/P = \int r dt \leftrightarrow \log(P)=rt+C \leftrightarrow P=\exp(rt+C) \leftrightarrow P=\exp(C)\exp(rt)$
- In R, if $P_0=1000$ and $r=0.02$, the growth curve for years from 0 to 300 will be drawn by typing **curve(1000*exp(0.02*x), 0, 300)**
- Slightly more realistic curve is logistic growth, setting upper limit of population growth (K) $dP/dt=rP(1-P/K) \leftrightarrow \{1/(K-P)+1/P\}dP=r dt \leftrightarrow \log|P/(K-P)|=rt+C \leftrightarrow P=K/(1+\exp(-rt-C))$
R code: **curve(5000/(1+exp(-0.02*x)/0.05), 0, 500)**
- Flaws of such deterministic models are: (1) assuming population growth to follow simple mathematical rules, (2) using very little historical information
- Better model is fitting a simple descriptive model to the existing data and extrapolate into the future.

```
t <- 0:10*10
P <- c(1000,1020,1030,1050,1080,1130,1150,1200,1250,1400,1520)
summary(lm(P~t))
summary(lm(P~t+I(t*t)))
(res <- summary(lm(P~t+I(t*t)+I(t*t*t))))
rc <- coef(res)
plot(P~t, xlim=c(0,200), ylim=c(0,5000))
curve(rc[1]+rc[2]*x+rc[3]*x*x+rc[4]*x*x*x, 0, 200, add=T)
```
- This kind of fitting gives equal weight to each data, but more recent figures will contribute more to future patterns than earlier in fact. Thus taking weighted moving average is better idea.
- (Not given in the text) Considering auto-correlation is another idea. In R, using forecast package, it's easy.

```
library(forecast)
PP <- ts(P, start=0, end=100, freq=0.1)
plot(forecast(auto.arima(PP), h=10))
```
- All those projections are only useful for theoretical purpose, not for real forecasts. Instead, component method is usually applied.

Component method of projection (a.k.a. cohort component method)

- Basic idea is very simple: Population change = Natural increase + net migration
 $P_{t+1} = P_t + \text{Births} - \text{Deaths} + \text{Immigrants} - \text{Emigrants}$
- This "basic demographic" equation is not only applicable to whole population, but also to component parts.
- Splitting the population by age, sex, marital status, and whatever else is necessary. Then the basic demographic equation should be applied to each part, then integrate them after projection.
- Requiring 3 things: (1) Base population from which the projection starts, (2) a set of assumptions about the course of events during the period covered by the projection, (3) a method by which the assumptions are applied to the base population.
- (See, <http://www.mhlw.go.jp/topics/bukyoku/seisaku/syousika/030819/2b.html> [in Japanese])
- (e.g.) 1981 female population of India.
 - Initially project for 5 years, because the base population (Table 15.1) is divided into 5 year age groups.
 - Assuming net migration being zero in every age-group.
 - The base population from 1981 census is the first column of Table 15.1.
 - The assumptions about fertility and mortality may vary widely from quite simple to very detailed
 - Fertility: Total number of births, CBR, ASFR of period/cohort, fertility model parameters; breakdown by parity, legitimacy are possible. However, for component method, ASFR is required (For 1981 India, given in Table 15.2)
 - Mortality: Actual life table in 1981 India can be used, but usually model life table is applied because the changes in mortality can be specified in terms of changes in model parameters or levels. Usually survivorship ratio P (the probability of surviving from one age group to another, the ratio of two adjacent nL_x values, such as $P = \frac{{}_5L_{x+5}}{{}_5L_x$ (given in Table 15.1)
 - Based on 1981 female population and P (of Coale and Demeny model life table West level 15), 1986 female population for ages above 5 will be calculated as shown in Table 15.1.
 - For 0-4 female population, taking average of 1981 and 1986 female populations aged 15-49, multiplying FASFRs times 5 for every age groups, then sum up them.

There are many other new projection or forecasting methods. The document given by Dr. Rob Hyndman, "Prospective Life Tables" is very concise and informative, available from <https://robjhyndman.com/papers/prospect.pdf>. It gives the instruction how to use the function `lca()` of demography package.

At first, it's necessary to call demography package (If you have not installed it, please install from cran).

Then, read mortality rates and population data from database. "Japanese Mortality Database", National Institute of Population and Social Security Research, Available at <http://www.ipss.go.jp/p-toukei/JMD/index-en.asp>.

Number of births 1x1 in the nation: <http://www.ipss.go.jp/p-toukei/JMD/00/STATS/Births.txt>

Mortality rates 1x1 in the nation: http://www.ipss.go.jp/p-toukei/JMD/00/STATS/Mx_1x1.txt

Population at risk 1x1 in the nation: http://www.ipss.go.jp/p-toukei/JMD/00/STATS/Exposures_1x1.txt

```
library(demography)
```

```
Japan <- read.demogdata("Mx_1x1.txt", "Exposures_1x1.txt", type="mortality", label="Japan")
```

```
# type is either of "mortality", "fertility", or "migration"
```

```
lc.f <- lca(Japan, series="female")
```

```
forecast.lc.f <- forecast(lc.f, h=20)
```

```
plot(Japan, series="female", lty=2, col="lightgray")
```

```
lines(forecast.lc.f, lty=1, col="red")
```

See the R code, will be given as <http://minato.sip21c.org/demography-special/Chap15.R>

