V. Statistical Demography

- Demography: Statistical study of human populations
- By *statistical:* using the language of probability and statistical modeling for the population processes involved: mortality, fertility, population projections
- In this course, the topics:
  - Mortality
  - Life expectancy
  - Population projections

Mortality topics

- Crude Death Rate
  - Annual number of deaths per 1000
- Standardized crude death rate
- Age- and gender specific death rates
- Probability of death
- Mortality table, life table (dødelighetstabell)
- Modeling mortality rates (population projections)

Crude death rate, *CDR*

- D = number of deaths during one given year
- P_0 = population size at January 1
- P_1 = population size at December 31
- A measure of mortality: Annual number of deaths pr. 1000

$$CDR = \frac{D}{\frac{1}{2}(P_0 + P_1)} \cdot 1000$$

Mortality in the whole population regardless of age

CDR can be misleading with comparisons in time and space, because it does not take into account the age structure in the population

Standardize *CDR* for age and gender

- Example:
  - Norway 1950: *CDR* = 9,13
  - Norway 1987: *CDR* = 11,14
- *CDR* has increased by 2 per 1000, but has really the mortality increased?
- Age specific death rates in 1987 < 1950 for all ages
- Explanation: Population is older in 1987!

Age specific death or mortality **rate**:

$$\text{Agegroup}\, x, m(x) = \frac{D_x}{\frac{1}{2}(P_{0,x} + P_{1,x})}$$

D_x = number of deaths for agegroup x

$P_{0,x}$ and $P_{1,x}$:

Size of population for agegroup x at the start- and end-population for age x

Standardized (hypothetical) *CDR*

- *CDR* in a hypothetical population with 1950 age structure and 1987 age specific death rates

$$CDR_{50}^{(87)}$$

- Alternatively: *CDR* in a hypothetical population with 1987 age structure and 1950 age specific death rates

$$CDR_{87}^{(50)}$$

$N_{87,x} = (P_{0,x,87} + P_{1,x,87})/2$ and $N_{50,x} = (P_{0,x,50} + P_{1,x,50})/2$

$D_{87,x} =$ number of deaths in the age group x in 1987

$D_{50,x} =$ number of deaths in the age group x in 1950

Age specific mortality rates for 1987 and 1950 in age group :

$m(x,87) = D_{87,x} / N_{87,x}$, $m(x,50) = D_{50,x} / N_{50,x}$

Population 1987 with 1950 death rates:

The hypothetical number of deaths in the age group x:

$$D_{87,x}^{(50)} = N_{87,x} m(x,50)$$

Total number of deaths: $D_{87}^{(50)} = \sum_{\text{age groups } x} D_{87,x}^{(50)}$

Standardized CDR: $CDR_{87}^{(50)} = D_{87}^{(50)} / N_{87}$

Population 1950 with 1987 death rates:

The hypothetical number of deaths in the age group *x*:

$$D_{50,x}^{(87)} = N_{50,x} m(x,87)$$

Total number of deaths: $D_{50}^{(87)} = \sum_{\text{age groups } x} D_{50,x}^{(87)}$

Standardized *CDR*: $CDR_{50}^{(87)} = D_{50}^{(87)} / N_{50}$

We find :

$D_{87}^{(50)} = 57821$ (compared to observed $D_{87} = 46760$):

$CDR_{87}^{(50)} = 57821 / 4198 = 13.77$

$D_{50}^{(87)} = 23568$ (compared to observed $D_{50} = 29930$):

$CDR_{50}^{(87)} = 23568 / 3279 = 7.19$

Norway 1950 and 1987 standardized *CDR*

Alder	N_{50}	d_{50}	D_{50}	N_{87}	d_{87}	D_{87}	$N_{87}d_{50}$	$N_{50}d_{87}$
0–4	318	67,2	2137	259	21,1	546	1741	671
5–9	265	8,6	228	260	2,0	52	224	53
10–14	213	5,0	107	285	1,8	51	143	38
15–19	207	8,0	166	333	6,1	203	266	126
20–24	230	12,7	292	332	8,0	266	422	184
25–29	262	12,1	317	315	7,4	233	381	194
30–34	263	14,0	368	314	8,6	270	440	226
35–39	250	17,8	445	306	10,9	334	545	265
40–44	234	24,6	576	303	18,8	570	745	440
45–49	216	39,4	851	216	29,9	646	851	646
50–54	198	58,4	1156	186	50,8	945	1086	1006
55–59	167	89,9	1501	198	82,5	1634	1780	1378
60–64	136	141,3	1922	212	132,5	2809	2996	1802
65–69	110	228,0	2508	214	213,3	4565	4879	2346
70–74	90	393,7	3543	178	351,0	6248	7998	3159
75–79	62	715,5	4436	137	580,1	7947	9802	3597
80 —	57	1645,1	9377	149	1304,8	19441	24512	7437
Sum	3279		29930	4198		46760	57821	23568
s.d.			9,13			11,14	13,77	7,19

N: mean population size in 1000

D: number of deaths, $d = D/N$ per 10 000, s.d. = *CDR*

Comparison of observed and standardized mortality rates

- Compare the four crude deaths rates (two unstandardized and two standarized)
- How much of the difference between the *CDR* in 1950 and 1987 is due to:
  - Age structure
  - Mortality
- The change in *CDR* of 2 per 1000 is the sum of these two components

Standardized *CDR*

Alder	N_{50}	d_{50}	D_{50}	N_{87}	d_{87}	D_{87}	$N_{87}d_{50}$	$N_{50}d_{87}$
0–4	318	67,2	2137	259	21,1	546	1741	671
5–9	265	8,6	228	260	2,0	52	224	53
10–14	213	5,0	107	285	1,8	51	143	38
15–19	207	8,0	166	333	6,1	203	266	126
20–24	230	12,7	292	332	8,0	266	422	184
25–29	262	12,1	317	315	7,4	233	381	194
30–34	263	14,0	368	314	8,6	270	440	226
35–39	250	17,8	445	306	10,9	334	545	265
40–44	234	24,6	576	303	18,8	570	745	440
45–49	216	39,4	851	216	29,9	646	851	646
50–54	198	58,4	1156	186	50,8	945	1086	1006
55–59	167	89,9	1501	198	82,5	1634	1780	1378
60–64	136	141,3	1922	212	132,5	2809	2996	1802
65–69	110	228,0	2508	214	213,3	4565	4879	2346
70–74	90	393,7	3543	178	351,0	6248	7998	3159
75–79	62	715,5	4436	137	580,1	7947	9802	3597
80 —	57	1645,1	9377	149	1304,8	19441	24512	7437
Sum	3279		29930	4198		46760	57821	23568
s.d.			9,13			11,14	13,77	7,19

1: *CDR* (Pop87&mortality50) – *CDR*(50): due to effect of age structure = 4,64
2: *CDR* (87) – *CDR* (Pop87&mortality 50): due to change in age specific mortality rates = - 2,63

1. Change in age structure	13,77 – 9,13 = +4,64
2. Change in age specific mortality	11,14 – 13,77 = -2,63
3. Total = 1+2	4,64-2,63 = 2,01 (= 11,14-9,13)

Effect of changing factors

- *CDR* increased by 2 from 1950 to 1987, but with constant age specific mortality it would have increased by 4,64
  - Effect of an aging population
- The reduced age specific mortality has reduced the increase in *CDR* to only 2: 2,6 reduced increase
  - Effect of lower age specific mortality

Alternative computation

Alder	N_{50}	d_{50}	D_{50}	N_{87}	d_{87}	D_{87}	$N_{87}d_{50}$	$N_{50}d_{87}$
0–4	318	67,2	2137	259	21,1	546	1741	671
5–9	265	8,6	228	260	2,0	52	224	53
10–14	213	5,0	107	285	1,8	51	143	38
15–19	207	8,0	166	333	6,1	203	266	126
20–24	230	12,7	292	332	8,0	266	422	184
25–29	262	12,1	317	315	7,4	233	381	194
30–34	263	14,0	368	314	8,6	270	440	226
35–39	250	17,8	445	306	10,9	334	545	265
40–44	234	24,6	576	303	18,8	570	745	440
45–49	216	39,4	851	216	29,9	646	851	646
50–54	198	58,4	1156	186	50,8	945	1086	1006
55–59	167	89,9	1501	198	82,5	1634	1780	1378
60–64	136	141,3	1922	212	132,5	2809	2996	1802
65–69	110	228,0	2508	214	213,3	4565	4879	2346
70–74	90	393,7	3543	178	351,0	6248	7998	3159
75–79	62	715,5	4436	137	580,1	7947	9802	3597
80 —	57	1645,1	9377	149	1304,8	19441	24512	7437
Sum	3279		29930	4198		46760	57821	23568
s.d.			9,13			11,14	13,77	7,19

1: *CDR* (Pop50&mortality87) – *CDR*(50): due to change in age specific mortality rates = -1,94

2: *CDR* (87) – *CDR* (Pop50&mortality 87): due to effect of age structure = 3,95

1. Change in age specific mortality	7,19 – 9,13 = -1,94
2. Change in age structure	11,14 – 7,19 = 3,95
3. Total = 1+2	-1,94+3,95 = 2,01

Some comments

- Standardized rates will vary with the standard use and so will the quantitative conclusion
- Can also control for gender
- When comparing two countries: A third country age structure can be used as standard
- Example:Mauritius vs England and Wales with Japan age structure as standard

Table 3.2 Calculation of standardized death rate, Mauritius and England and Wales

	Standard population Japan 1985 ('000)		Mauritius: Age-specific death rates 1986 (per 1,000)		Mauritius: Expected deaths in standard population		England and Wales: Age-specific death rates 1987 (per 1,000)		England and Wales: Expected deaths in standard population	
	M	F	M	F	M	F	M	F	M	F
Under 1	732	698	30.8	23.1	22,546	16,124	10.4	7.9	7,613	5,514
1–4	3,087	2,942	1.2	1.3	3,704	3,825	0.5	0.4	1,544	1,177
5–14	9,520	9,054	0.5	0.4	4,760	3,622	0.2	0.1	1,904	905
15–24	8,766	8,414	1.1	1.0	9,643	8,414	0.7	0.3	6,136	2,524
25–34	8,507	8,371	2.1	1.4	17,865	11,719	0.9	0.5	7,656	4,186
35–44	9,950	9,923	4.5	2.3	44,775	22,823	1.7	1.1	16,915	10,915
45–54	8,019	8,151	11.7	5.0	93,822	40,755	5.0	3.2	40,095	26,083
55–64	5,789	6,616	27.8	13.7	160,934	90,639	15.7	9.0	90,887	59,544
65–74	3,285	4,472	57.3	39.6	188,231	177,091	41.7	22.8	136,985	101,962
75–84	1,560	2,367	123.0	83.5	191,880	197,645	98.8	60.7	154,128	143,677
85+	256	529	249.5	192.5	63,872	101,833	212.7	168.6	54,451	89,189
	121,008				802,032	674,490			518,314	445,676

Standardized death rate:

Mauritius: $\frac{(802{,}032 + 674{,}490)}{121{,}008{,}000} \times 1{,}000$ = 12.20 per 1,000

England and Wales: $\frac{(518{,}314 + 445{,}676)}{121{,}008{,}000} \times 1{,}000$ = 7.97 per 1,000

Unstandardized (CDR) — Mauritius: 6.7 per 1,000 — England and Wales: 11.3 per 1,000

1. All age-specific death rates are higher in Mauritius for both men and women
2. Yet, unstandardized CDR is 6.7 compared to 11.3 for England and Wales
3. Reason: Population in Mauritius is much younger
4. The standardized *CDR*s are 12.20 for Mauritius and 7.97 for England and Wales, according to the age structure in Japan

Direct standardization

- <u>Same</u> age structures
- Different age specific mortality rates
- Here we ”control” for age, as in the previous example
- Another example:
  - *CDR* for Kuwait in 1996: 2.18 per 1000
  - *CDR* for United Kingdom (England, Wales, Scotland and Northern Ireland): 10.0 per 1000
  - If we use UK 1996 age structure as standard: Kuwaits CDR is 12.75 per 1000

Indirect standardization

- The most common approach in studies of mortality
- A standard of age specific death rates, combined with age structure (f.ex. from censuses)

1. Compute expected number of deaths based on actual age structure and standard age specific death rates
2. Compute standardized mortality ratio (SMR) =

   observed number of deaths/ expected number of deaths

SMR > 1: actual (but unknown) age specific death rates are higher than the standard

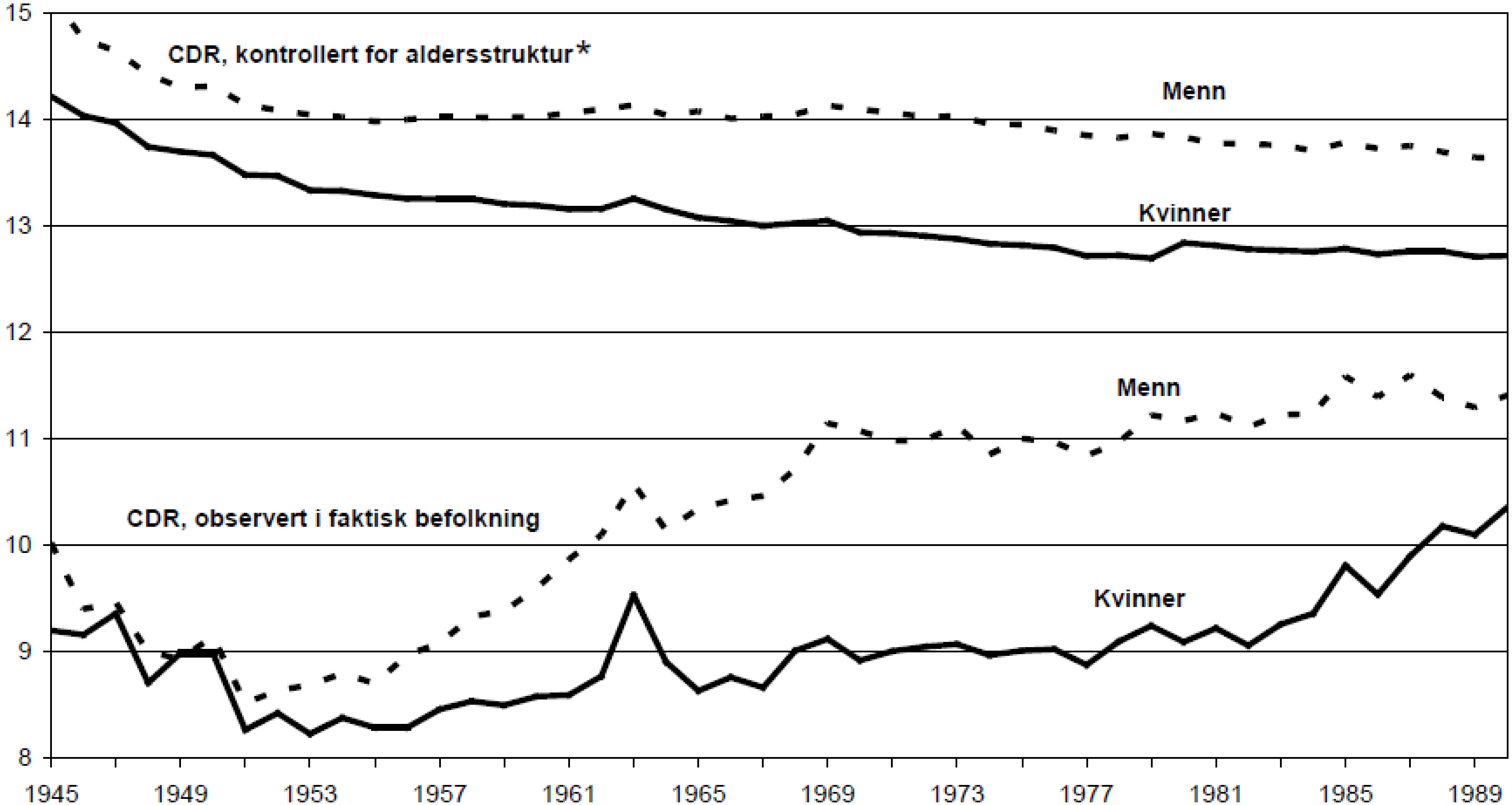
SMR < 1: actual (but unknown) age specific death rates are lower than the standard

Example UK vs. Kuwait

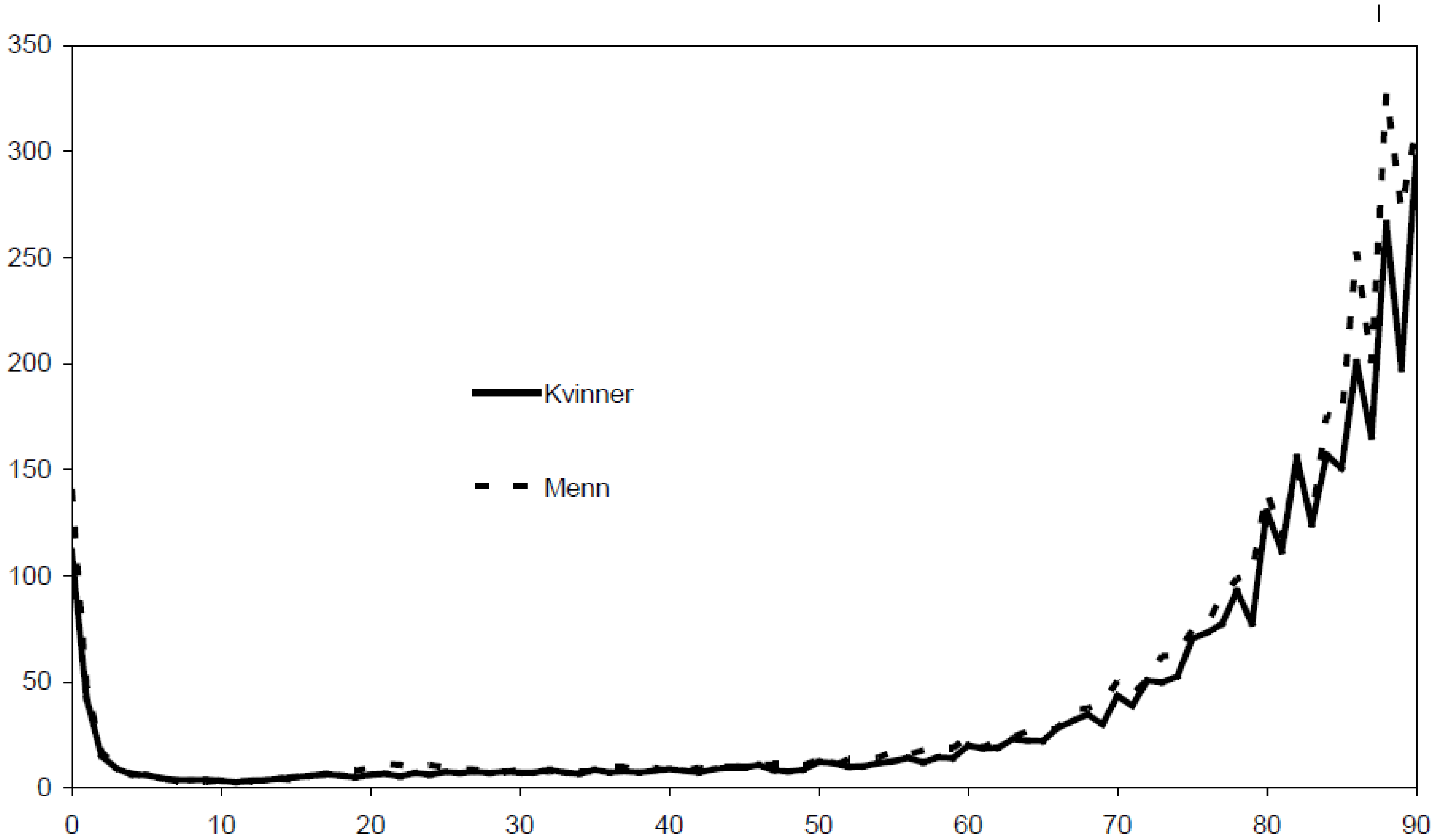
- UK 1996 *CDR* = 10.0, Kuwait 1996 *CDR* = 2.2
- Standard age specific death rates: UK 1996

1. Compute expected number of deaths for Kuwait, based on Kuwaits age structure and standard death rates
2. Result: 3459
3. Observed number of deaths: 3815
4. Standardized mortality ratio SMR = 3815/3459 = 1.10
5. SMR >1: Age specific mortality in Kuwait must be higher than the standard UK

CDR Norway 1945- 1990, number of deaths per 1000



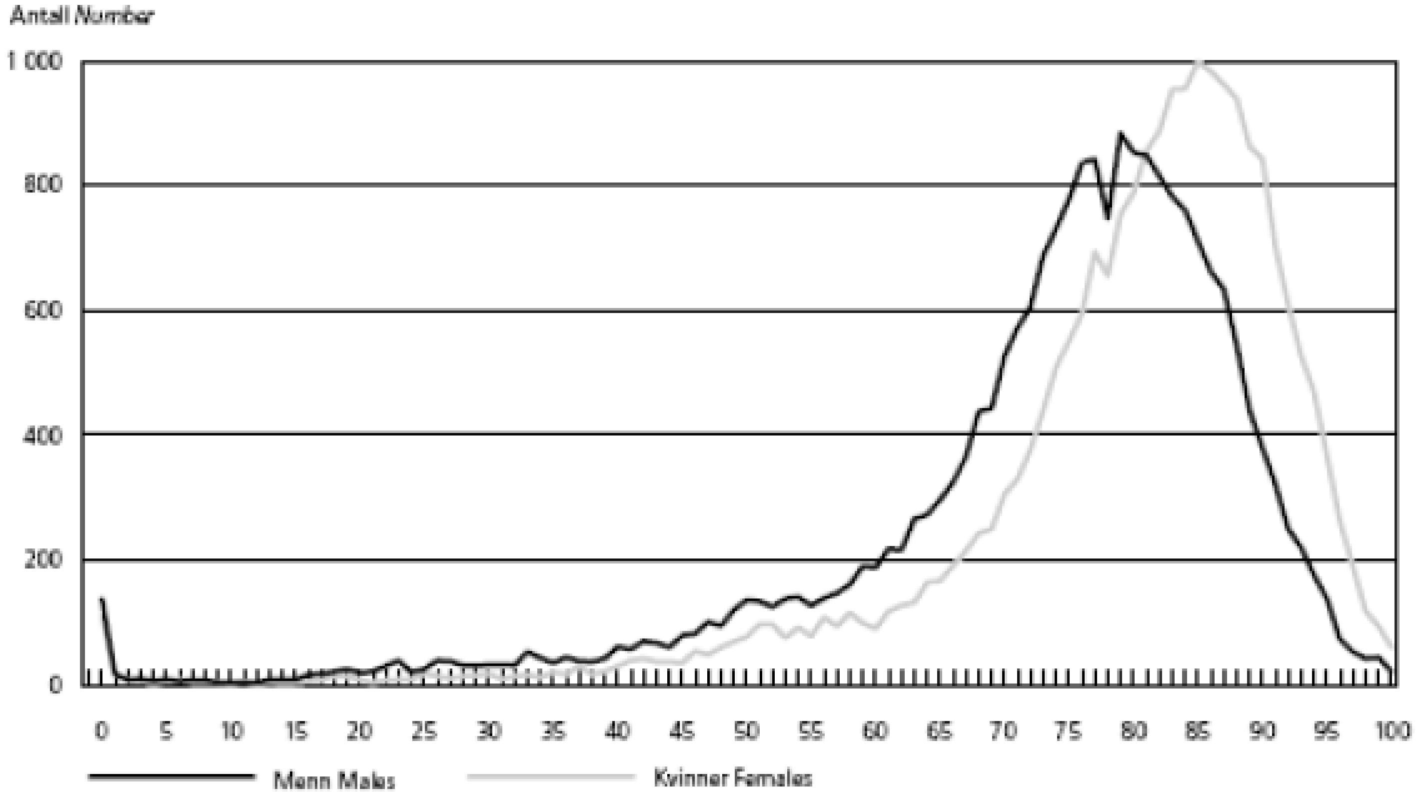
Age specific death rates, Norway 1900



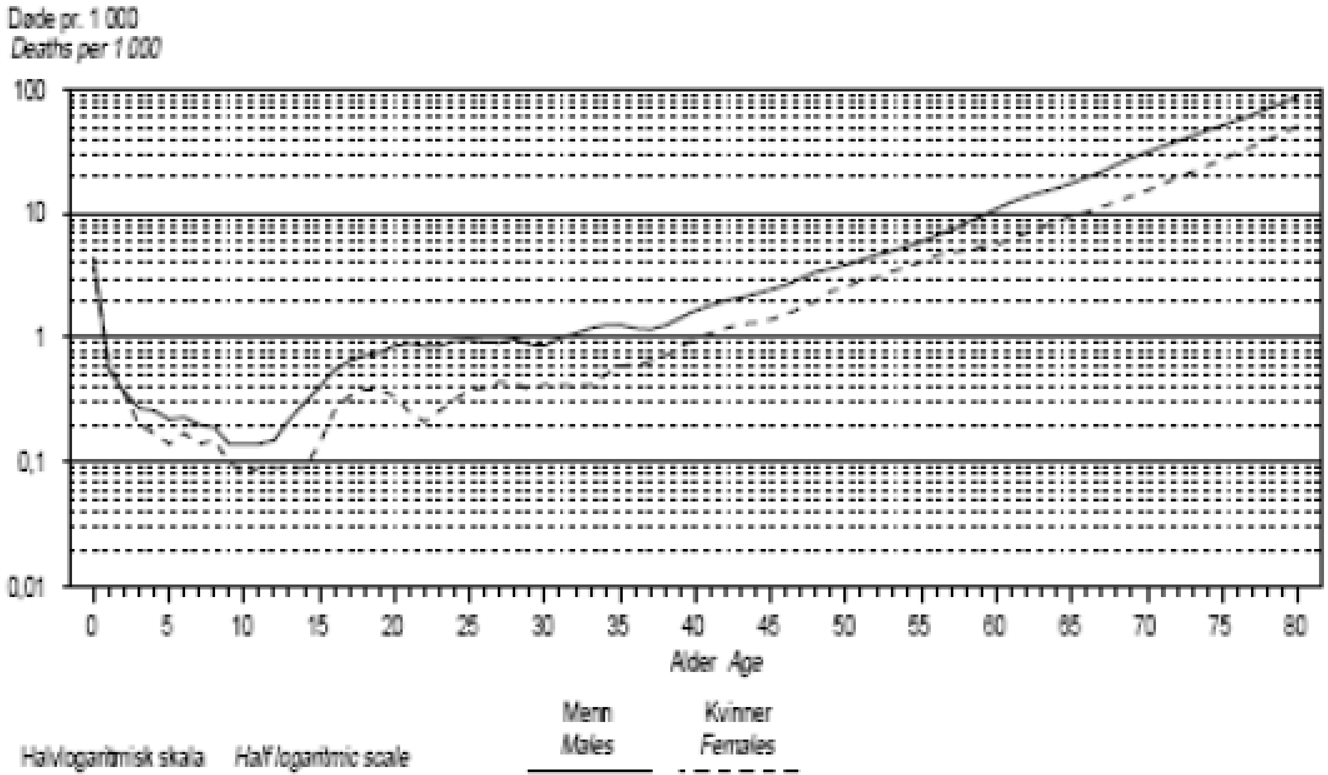
age

Døde etter kjønn og alder. 1997

Deaths by sex and age. 1997



Aldersavhengige dødsfallsrater. 1997
Age-specific death rates. 1997



Probability of death

$P_{0,x}$: number of persons alive at exact age x

D_x = number of deaths for this agegroup x

and $P_{1,x}$: population size for agegroup x at the end for age x

$q(x)$ = estimated probability of dying before the age of $x+1$ given that the person is alive at age x, the one year death **probability**

If we observe D_x :

$$q(x) = D_x / P_{0,x}$$

Let $D_{x,total}$ be the total number of deaths for this agegroup

If only aggregated numbers are available: $q(x)$ is the ratio of $D_{x,total}$ to the "middle" population half-way:

Middle population:

$$P_{0,x} + \frac{1}{2}(P_{1,x} - P_{0,x} + D_{x,total}) = \frac{1}{2}(P_{0,x} + P_{1,x} + D_{x,total})$$

and $q(x) = \dfrac{D_{x,total}}{\frac{1}{2}(P_{0,x} + P_{1,x} + D_{x,total})}$

Age specific mortality rate

$$\text{Agegroup}\, x, m(x) = \frac{D_{x,total}}{\frac{1}{2}(P_{0,x} + P_{1,x})}$$

In either case: $$q(x)=\frac{m(x)}{1+\frac{1}{2}m(x)}$$

Proof : a) no immigratio n in this age group, $D_x = D_{x,total}$

$$P_{1,x}=P_{0,x}-D_x$$

$$\Rightarrow m(x)=\frac{D_x}{P_{0,x}-\frac{1}{2}D_x}$$

and $$\frac{m(x)}{1+\frac{1}{2}m(x)}=\frac{D_x}{P_{0,x}-\frac{1}{2}D_x+\frac{1}{2}D_x}=q(x)$$

b) immigratio n and we do not know D_x, only the aggregated $P_{1,x}$ and $D_{x,total}$:

Then $$q(x)=\frac{D_{x,total}}{(P_{0,x}+P_{1,x}+D_{x,total})/2}=\frac{m(x)}{1+m(x)/2}$$

Example

$\text{At age } 60: P_{0,60} = 1000 \text{ and } D_{60} = 20$

$\text{Direct}: q(60) = 20/1000 = 0.02$

$\text{Note}: m(60) = 20/\tfrac{1}{2}(1000 + 980) = 20/990$

$\text{and } m(60)/(1 + \tfrac{1}{2}m(60)) = 20/(10+990) = 20/1000 = 0.02$

$\text{With immigration of 200 and only aggregated numbers}:$

$P_{1,60} = 1000 + 200 - 20 = 1180:$

$m(60) = 20/\tfrac{1}{2}(1000 + 1180) = 0.01835 = 18.35 \text{ per } 1000$

$q(60) = 0.01835/(1 + 0.01835/2) = 0.01818 = 18.18 \text{ per } 1000$

Individual data: $q(60)$ can be calculated directly

Aggregated data, knows only P_0, P_1 and the total number of deaths during one year of this age group: first mortality rate $m(x)$ and then $q(x)$

If $m(x) = 0$ then $q(x) = 0$,

While if all die, $m(x) = 2$ and $q(x) = 1$

When all die: $D_x = P_{0,x}$ and $P_{1,x} = 0$

and $m(x) = P_{0,x} /(P_{0,x}/2) = 2$.

Then: $q(x) = 2/(1+2/2) =1$.

Otherwise: $q(x)< m(x)$

Mortality table (life table)

- First time: John Graunt in 1662
- A method for summarizing age dependent death rates/probabilities for a given year
- A hypothetical cohort (for example 100 000 persons) experience deaths in accordance with the mortality rates:

  simulate the lifecareer to a table population (life table population)
- Can answer several questions using a standard mortality table:
  - How many are alive after 1, 2, 3, …years?
  - What is the life expectancy (forventet levealder)?
  - What are the chance of dying between two given ages?

Life expectancy

- The number of years a person born today can be expected to live under the *current* age specific mortality rates.
- Specifically: Given age specific death rates (or death probabilities) for ages 0, 1, 2, 3, …
- **Remaining life expectancy** for a certain age x, under the current age specific death rates (death probabilities) for ages x, $x+1$, $x+2$, …
- Notation: e_x for $x = 0, 1, 2, \ldots$
- Hence: e_0 is life expectancy at birth

NB!

- Life expectancy and remaining life expectancy is a hypothetical (also called synthetical : kunstig) measure of mortality.
- Example: In 2012 in Norway, for male of age 64:
  - e_{64} = 19.03 years. So 64 year old Norwegian men can expect to be 83.03 years old
  - But mortality will decrease in the coming years so the true expected age will be higher than 83.03
- There is a difference between remaining life years based on synthetic mortality and actual given age cohort

Let X be the length of life for a person.

Special case of variable *waiting time* until a specific event. Here the event is death. Let the density function of X be $f(x)$. Then Life expectancy is

$$E(X) = \int_0^\infty x f(x) dx$$

Alternative expression:

Indicator process: $I(t) = 1$ if $X > t$, and 0 otherwise.

We can represent X as:

$$X = \int_0^\infty I(t) dt$$

the integral representation of X

$$E(X) = E[\int_0^\infty I(t)dt] = \int_0^\infty [\int_0^\infty I(t)dt] f(x)dx$$

$$= \int_0^\infty [\int_0^\infty I(t) f(x)dx]dt = \int_0^\infty [\int_t^\infty f(x)dx]dt = \int_0^\infty [P(X > t)]dt$$

Let

$$p(t) = P(X > t)$$

such that

$$e_0 = E(X) = \int_0^\infty p(t)dt$$

Remaining life expectancy at age x:

$$e_x = E(X - x \mid X > x)$$

Conditional probability of surviving age $x+t$ given survival to age x:

$$P(X > x+t \mid X > x) = \frac{P(X > x+t \cap X > x)}{P(X > x)} = \frac{P(X > x+t)}{P(X > x)}$$

Hence,

$$e_x = \int_0^{\infty} \frac{p(x+t)}{p(x)} dt$$

Assume $p(x)$ is specified for the ages $x = 0, 1, 2, ..$

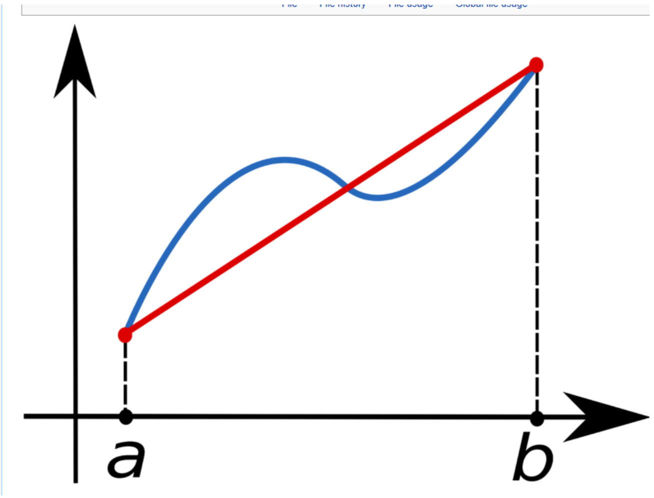
Approximation, assuming linearity of $p(t)$ in each interval $[x,x+1)$:

$$e_x \approx 0.5 + \sum_{t=1}^{\infty} p(x+t)/p(x) = 0.5 + \frac{1}{p(x)} \sum_{t=1}^{\infty} p(x+t)$$

This is the trapezoidal method of numerical integration

The trapezoidal method of numerical integration

$$\int_a^b f(x)dx \approx (b-a)\left[\frac{f(a)+f(b)}{2}\right]$$



The function (in blue) is approximated by a linear function (in red)

The area under the curve $f(x)$ is approximated by a trapezoid (only two parallell lines), norsk: trapes

Applied to e_x :

$$e_x = \frac{1}{p(x)}\int_0^\infty p(x+t)dt = \frac{1}{p(x)}\sum_{k=0}^\infty [\int_k^{k+1} p(x+t)dt]$$

$$\approx \frac{1}{p(x)}\sum_{k=0}^\infty [1\cdot \frac{p(x+k)+p(x+k+1)}{2}]$$

$$= \frac{1}{p(x)}[\frac{p(x)+p(x+1)}{2} + \frac{p(x+1)+p(x+2)}{2} + ...]$$

$$= 0.5 + \frac{1}{p(x)}[p(x+1)+p(x+2)+...]$$

Estimation of $p(x+t) = P(X > x+t)$, the probability of survival at age $x+t$.

$q(x)$ is the estimated probability of death at age x, i.e., the probability of death between the ages of x and $x+1$ so the probablity of death *before* age $x+t$ is the sum of $q(k)$ for $k =$ 0,…, $x+t-1$. Hence an estimate of the probability of survival at age $x+t$ is given by:

$$\hat{p}(x+t) = 1 - \sum_{k=0}^{x+t-1} q(k).$$

Based on these estimates, we can compute the number of persons alive at age $x+t$ based on a hypothetical (synthetic) population of 100 000.

For example, assume $q(0) = 2.73$ and $q(1) = 0.25$ per 1000.
Then an estimate of $P(X>2) = 1- (q(0) + q(1))= 0.99702$
Then the number of persons remaining alive at age 2 will be
$100\,000 \cdot 0.99702 = 99702$.

Let I_{x+t} be the number of survivals at age $x + t$ in the synthetic population of 100 000. Then

$$I_{x+t} = \hat{p}(x+t) \cdot 100000$$

It follows that

$$\hat{p}(x+t) / \hat{p}(x) = I_{x+t} / I_x$$

Estimated life expectancy

$$\hat{e}_x = 0.5 + \frac{1}{I_x}\sum_{t=1}^{K-x} I_{x+t}$$

where K is the highest age recorded in the mortality table.

This is obtained from a life table or mortality table

Mortality table 2008 Norway

dix

lder	Levende ved alder x			Døde i alder x til x+1			Forventet gjenstående levetid ved alder x			Dødssannsynlighet for alder x, Promille		
	l_x			d_x			e_x			q_x		
	Begge kjønn[2]	Menn	Kvinner	Begge kjønn[2]	Menn	Kvinner	Begge kjønn[2]	Menn	Kvinner	Begge kjønn[2]	Menn	Kvinner
0	100 000	100 000	100 000	273	330	214	80,67	78,31	82,95	2,73	3,30	2,14
1	99 727	99 670	99 786	25	29	21	79,89	77,57	82,13	0,25	0,30	0,21
2	99 702	99 641	99 766	27	33	21	78,91	76,59	81,14	0,27	0,33	0,21
3	99 675	99 608	99 745	8	10	7	77,93	75,62	80,16	0,09	0,10	0,07
4	99 666	99 598	99 738	7	3	10	76,94	74,62	79,17	0,07	0,03	0,10
5	99 659	99 595	99 728	12	13	10	75,94	73,63	78,17	0,12	0,13	0,11
6	99 647	99 581	99 717	9	10	7	74,95	72,64	77,18	0,09	0,10	0,07
7	99 639	99 571	99 710	12	16	7	73,96	71,64	76,19	0,12	0,16	0,07
8	99 627	99 555	99 703	8	9	7	72,97	70,66	75,19	0,08	0,10	0,07
9	99 619	99 546	99 697	8	3	13	71,97	69,66	74,20	0,08	0,03	0,13
10	99 611	99 543	99 683	6	6	7	70,98	68,66	73,21	0,06	0,06	0,07
11	99 605	99 536	99 677	6	9	3	69,98	67,67	72,21	0,06	0,09	0,03
12	99 598	99 527	99 673	11	9	13	68,99	66,67	71,21	0,11	0,09	0,13
13	99 587	99 518	99 661	8	6	10	68,00	65,68	70,22	0,08	0,06	0,10
14	99 579	99 512	99 651	17	22	13	67,00	64,69	69,23	0,17	0,22	0,13
15	99 562	99 490	99 638	20	25	16	66,01	63,70	68,24	0,21	0,25	0,16
16	99 542	99 466	99 622	16	21	10	65,03	62,71	67,25	0,16	0,21	0,10
17	99 526	99 445	99 612	39	51	25	64,04	61,73	66,26	0,39	0,51	0,26
18	99 487	99 394	99 587	53	73	32	63,06	60,76	65,27	0,53	0,73	0,32
19	99 434	99 320	99 554	53	75	30	62,09	59,80	64,29	0,53	0,75	0,30
20	99 381	99 246	99 525	68	117	17	61,13	58,85	63,31	0,68	1,18	0,17
21	99 314	99 129	99 508	38	60	14	60,17	57,92	62,32	0,38	0,61	0,14
22	99 276	99 069	99 494	66	88	43	59,19	56,95	61,33	0,66	0,88	0,43
23	99 210	98 981	99 451	57	78	36	58,23	56,00	60,36	0,58	0,79	0,36

Alder x	Levende ved alder x l_x			Døde i alder x til x+1 d_x			Forventet gjenstående levetid ved alder x e_x			Dødssannsynlighet for alder x, Promille q_x		
	Begge kjønn	Menn	Kvinner	Begge kjønn	Menn	Kvinner	Begge kjønn	Menn	Kvinner	Begge kjønn	Menn	Kvinner
82	55 923	47 685	63 897	3 630	4 096	3 255	7,53	6,46	8,19	64,91	85,90	50,94
83	52 293	43 588	60 642	3 785	3 872	3 758	7,02	6,02	7,60	72,39	88,84	61,98
84	48 507	39 716	56 884	4 066	4 121	4 091	6,53	5,55	7,07	83,82	103,77	71,93
85	44 441	35 595	52 793	3 937	4 032	3 954	6,08	5,14	6,58	88,59	113,27	74,89
86	40 505	31 563	48 839	4 330	4 282	4 503	5,62	4,73	6,07	106,90	135,68	92,19
87	36 175	27 281	44 336	4 147	4 122	4 307	5,23	4,40	5,64	114,63	151,08	97,14
88	32 028	23 159	40 029	4 219	3 763	4 703	4,84	4,09	5,19	131,71	162,48	117,48
89	27 809	19 396	35 327	3 654	3 281	4 052	4,50	3,79	4,81	131,38	169,15	114,70
90	24 156	16 115	31 274	3 807	2 982	4 585	4,11	3,46	4,37	157,58	185,03	146,60
91	20 349	13 133	26 690	3 796	3 141	4 452	3,78	3,13	4,04	186,53	239,18	166,80
92	16 554	9 992	22 238	3 279	2 471	4 048	3,54	2,95	3,75	198,08	247,29	182,02
93	13 275	7 521	18 190	2 772	1 914	3 534	3,29	2,76	3,47	208,80	254,48	194,27
94	10 503	5 607	14 656	2 493	1 703	3 168	3,02	2,53	3,18	237,35	303,74	216,16
95	8 010	3 904	11 488	2 038	1 212	2 740	2,81	2,42	2,92	254,45	310,38	238,51
96	5 972	2 692	8 748	1 713	895	2 398	2,59	2,28	2,68	286,86	332,55	274,09
97	4 259	1 797	6 350	1 205	544	1 766	2,44	2,16	2,51	282,89	302,93	278,10
98	3 054	1 253	4 584	1 075	503	1 557	2,20	1,89	2,28	351,92	401,45	339,65
99	1 979	750	3 027	645	300	938	2,12	1,81	2,19	325,71	399,89	309,85
100	1 335	450	2 089	521	167	824	1,91	1,69	1,95	390,21	371,29	394,19
101	814	283	1 266	353	176	495	1,81	1,39	1,90	434,19	623,85	391,04
102	460	106	771	153	49	242	1,81	1,88	1,79	333,15	463,25	313,72
103	307	57	529	111	0	219	1,47	2,06	1,38	361,37	0,00	414,19
104	196	57	310	95	25	153	1,02	1,06	1,01	483,60	435,28	493,81
105	101	32	157	48	10	79	0,50	0,50	0,50	475,42	304,86	506,33

60,6% chance of reaching 83

40% chance of reaching 88

For a 90 year old, the chance of reaching the age of 100 = 2089/31274 = 6.7%.

Computation of remaining life expectancy –some examples:

$$\text{For men and women}: \hat{e}_{99} = 0.5 + \frac{1}{1979}(1335 + 814 + ... + 101)$$

$$= 0.5 + \frac{3213}{1979} = 0.5 + 1.62 = 2.12$$

For women :

$$\hat{e}_{99} = 0.5 + \frac{1}{3027}(2089 + 1266 + ... + 157)$$

$$= 0.5 + \frac{5122}{3027} = 0.5 + 1.69 = 2.19$$

Construction of life table

- Compute the age specific mortality rates $m(x)$
- Compute $q(x) = m(x)/[1+m(x)/2]$
- Derive the estimated $p(x + t)$
- Derive I_x in the synthetic population of 100 000
  - Start with $I_0 = 100\ 000$
- Compute the number of deaths d_x at the same time as I_x
- Finally compute estimated e_x

Historic development of life expectancy at birth, estimated e_0

Women

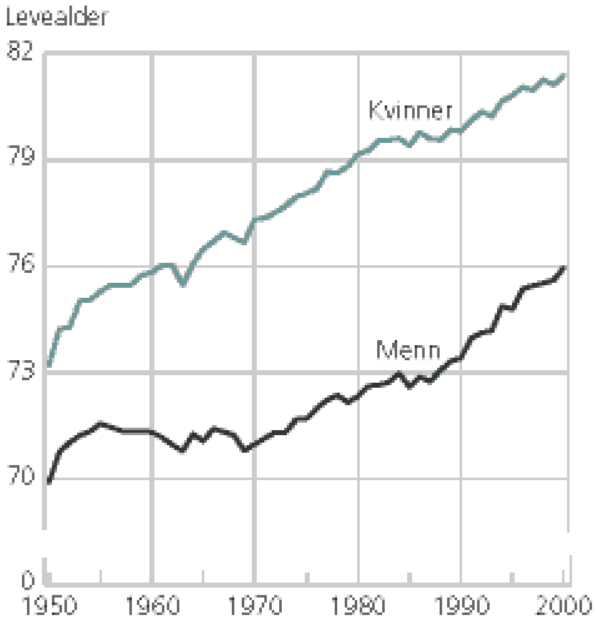
	1960	1970	1980	1990	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Noreg	76,0	77,5	79,2	79,8	81,5	81,6	81,6	82,1	82,5	82,7	82,9	82,9	83,2	83,2	83,3	83,6
Danmark	74,4	75,9	77,3	77,7	79,2	79,3	79,4	79,8	80,2	80,5	80,7	80,6	81,0	81,1	81,4	81,9
Finland	72,5	75,0	77,6	78,9	81,2	81,7	81,6	81,9	82,5	82,5	83,1	83,1	83,3	83,5	83,5	83,8
Island	76,4	77,3	80,1	80,5	81,6	83,2	82,5	82,5	83,2	83,5	82,9	83,4	83,3	83,8	84,1	84,1
Sverige	74,9	77,1	78,8	80,4	82,0	82,2	82,1	82,5	82,8	82,9	83,1	83,1	83,3	83,5	83,6	83,8

Men

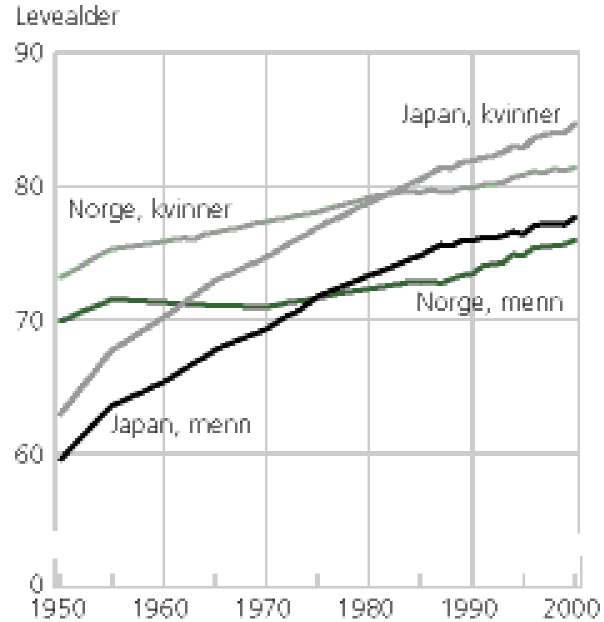
	1960	1970	1980	1990	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Noreg	71,6	71,2	72,3	73,4	76,0	76,2	76,4	77,1	77,6	77,8	78,2	78,3	78,4	78,7	79,0	79,1
Danmark	70,4	70,7	71,2	72,0	74,5	74,7	74,8	75,0	75,4	76,0	76,1	76,2	76,5	76,9	77,2	77,8
Finland	65,5	66,5	69,2	70,9	74,2	74,6	74,9	75,1	75,4	75,6	75,9	76,0	76,5	76,6	76,9	77,3
Island	71,3	71,2	73,4	75,4	77,8	78,3	78,6	79,5	78,9	79,6	79,5	79,6	80,0	79,8	79,8	80,7
Sverige	71,2	72,2	72,8	74,8	77,4	77,6	77,7	78,0	78,4	78,5	78,8	79,0	79,2	79,4	79,6	79,9

Increase in life expectancy every decade for Norway

	1960-1970	1970-1980	1980-1990	1990-2000	2000-2010
Women	1.5	1.7	0.6	1.7	2.0
Men	-0.4	1.1	1.1	2.6	3.1



Kilde: Mamelund og Borgan (1996) oppdatert med tall fra Statistisk sentralbyrå.

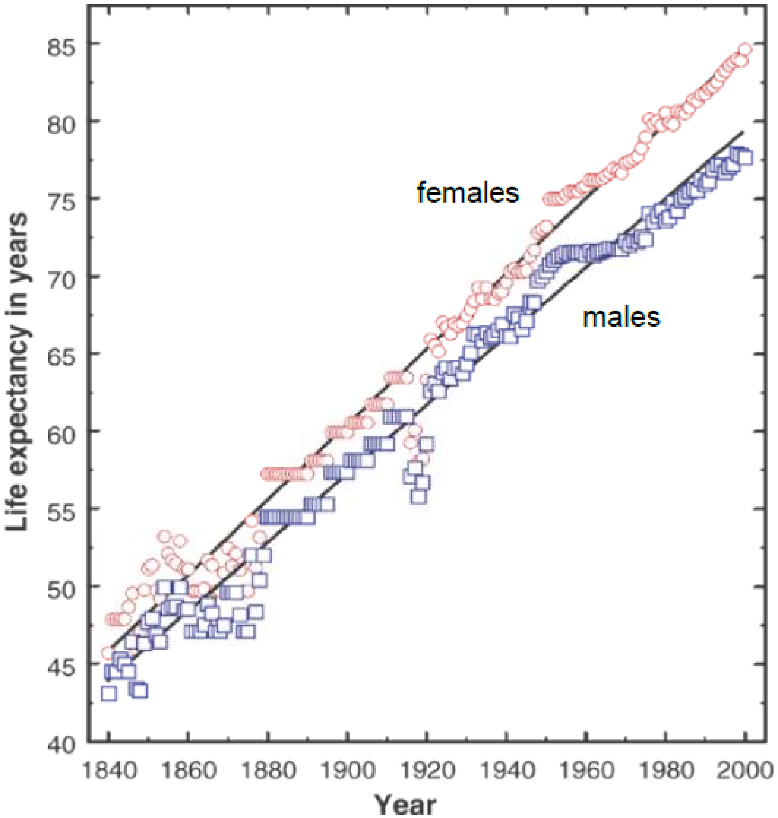


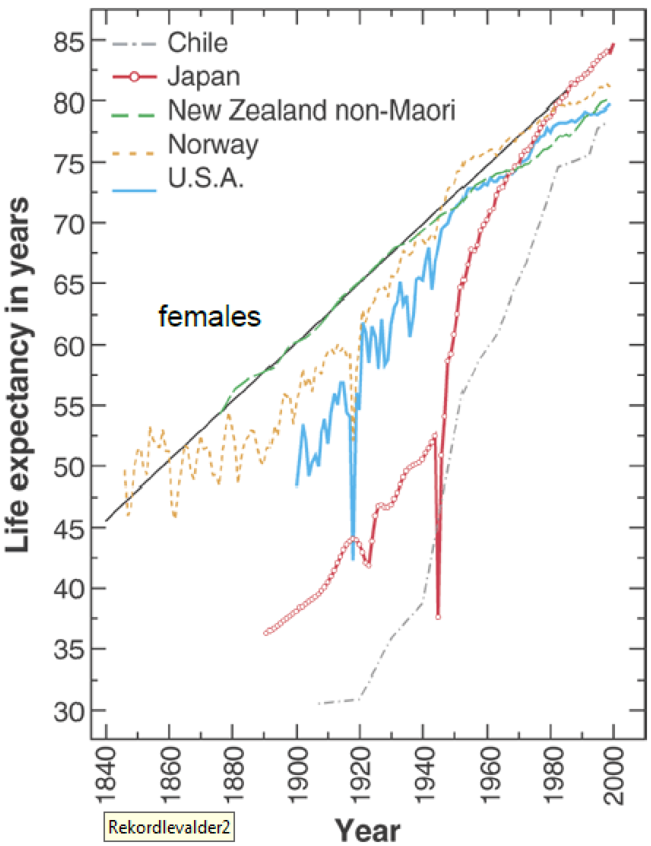
Kilde: Nettsidene til de statistiske sentralbyråer i Norge og Japan.

Norway. Difference between men and women in e_0 1850-2000

Kilde: Mamelund og Borgan (1996) oppdatert med tall fra Statistisk sentralbyrå.

Record life expectancy: has increased approximately linearly the last 160 years (2,5 year each decade)





Kilde: Oeppen og Vaupel (2002) *Science*

38

Population projections: forecast life expectancy

- Need to predict mortality rates in the future. Shall describe a method suggested by Lee-Carter (1992), the most used approach

- To model and forecast mortality:

    Standard methods for forecasting time series, together with a simple model for the age-time surface of the log of mortality. A forecast is produced for the probability distribution of each future age specific death rate

- We have data of mortality rates for the years $t = T_0, \ldots, T_1$

Lee-Carter model

Let $m(x,t)$ be the mortality rate for age x in the year t

and $a(x)$ be the average over time (T_0, T_1) of $\log m(x,t)$

Standard Lee-Carter model: $\log m(x,t) = a(x) + b(x)k(t) + \varepsilon(x,t)$

where $E[\varepsilon(x,t)] = 0$ and $Var[\varepsilon(x,t)] = \sigma_x^2$

Hence, $e^{a(x)}$ is the geometric mean of $m(x,t)$ over t

$a(x)$: age-specific constants describing the general pattern of mortality for the whole base period

$b(x)k(t)$ is an age(row) by time (column) matrix and the columns are *proportional*

Hence, the model will fit the data well, if the columns of $\{\log m(x,t)-a(x)\}$ are close to proportional *k*(*t*): index of the level of mortality capturing the main trend in death rates

b(*x*): age-specific constants describing the relative speed of change in mortality at each age

We see that $\sum_t b(x)k(t) = \sum_t \{\log m(x,t) - a(x)\} = 0$

$$\Leftrightarrow \sum_{t=T_0}^{T_1} k(t) = 0$$

The model is undetermined, e.g. if *b*(.) and *k*(.) are one solution, then so are *b*(.)*c* and *k*(.)/*c* for any constant *c*.

Normalize $b(x)$: $\sum_x b(x) = 1$

Estimation of the parameters in the Lee-Carter model

Unique least squares (LS) estimates :

LS estimates minimizes

$$\sum_x \sum_t [log\, m(x,t) - a(x) - b(x)k(t)]^2$$

Under conditions:

$\sum_x b(x) = 1$ and $\sum_t k(t) = 0$

Forecasting

Having fitted the demographic model we need a model for the mortality index $k(t)$

Typical model, in most applications: random walk with drift fits very well:

$$k(t) = k(t-1) + c + e(t)\sigma$$

where $e(t) \sim N(0,1)$ and uncorrelated

The drift term c represents an assumed linear trend in the change of $k(t)$ while $e(t)\sigma$ represents the deviations from this linear trend as random fluctuations

Negative c corresponds to a constant rate of decline for $m(x,t)$, reflecting a stable reduction of mortality

Seen as follows:

$$log \frac{m(x,t)}{m(x,t-1)} = log\, m(x,t) - log\, m(x,t-1)$$

$$= b(x)k(t) - b(x)k(t-1) = b(x)[k(t) - k(t-1)] = b(x) \cdot c$$

and $\frac{m(x,t)}{m(x,t-1)} = e^{c \cdot b(x)}$, independent of t

Estimation of c: the average of all observed $k(t)$-$k(t$-1)

Since [$k(t)$-$k(t$-1)] are i.i.d with mean c and standard deviation σ

$$\hat{c} = \frac{1}{T_1 - T_0} \sum_{t=T_0+1}^{T_1} [k(t) - k(t-1)] = \frac{k(T_1) - k(T_0)}{T_1 - T_0}$$

and

$$\hat{\sigma}^2 = \frac{1}{T_1 - T_0} \sum_{t=T_0+1}^{T_1} [k(t) - k(t-1) - \hat{c}]^2$$

and

$$SE(\hat{c}) = \hat{\sigma} / \sqrt{(T_1 - T_0)}$$

Point forecasts of mortality rates

$$\text{For } t > T_1 : k(t) = k(t-1) + \hat{c}$$

$$= k(t-2) + \hat{c} + \hat{c} = k(t-2) + 2\hat{c}$$

$$= .... = k(T_1) + (t - T_1)\hat{c}$$

$$\Rightarrow$$

$$\log m(x,t) - \log m(x,T_1) = b(x)[k(t) - k(T_1)]$$

and hence:

$$\log m(x,t) = \log m(x,T_1) + b(x)[k(t) - k(T_1)]$$

$$= \log m(x,T_1) + b(x)[t - T_1]\hat{c}$$

Stochastic forecasts of mortality

$$\hat{c} \sim N(c, \hat{\sigma} / \sqrt{T_1 - T_0}), \quad SE(\hat{c}) = \hat{\sigma} / \sqrt{T_1 - T_0}$$

$$\Rightarrow \hat{c} = c + SE(\hat{c})Z \text{ where } Z \sim N(0,1)$$

$$\Rightarrow k(t) = k(T_1) + c(t - T_1) + \hat{\sigma} \sum_{s=T_1+1}^{t} e(s)$$

$$= k(T_1) + [\hat{c} - SE(\hat{c})Z](t - T_1) + \hat{\sigma} \sum_{s=T_1+1}^{t} e(s)$$

and

$$\log m(x,t) = \log m(x,T_1) + b(x)[k(t) - k(T_1)]$$

Usually as basis for prediction intervals, 1000 simulations

Forecasting life expectancy

For each $t > T_1$: Derive life table with l_x = the number of survivals at age x and use the formula for estimating e_x

Applied to US data 1900-1989, from Lee-Carter (1992)
Was an influenza epidemic in 1918, used an intervention model for $k(t)$:

$$k(t) = k(t-1) + c + d \cdot I(t = 1918) + \sigma e_t$$

Estimates (standard error):

$$\hat{c} = -0.365\ (0.069),\ \ \hat{d} = 5.24\ (0.461),\ \ \hat{\sigma} = 0.655$$

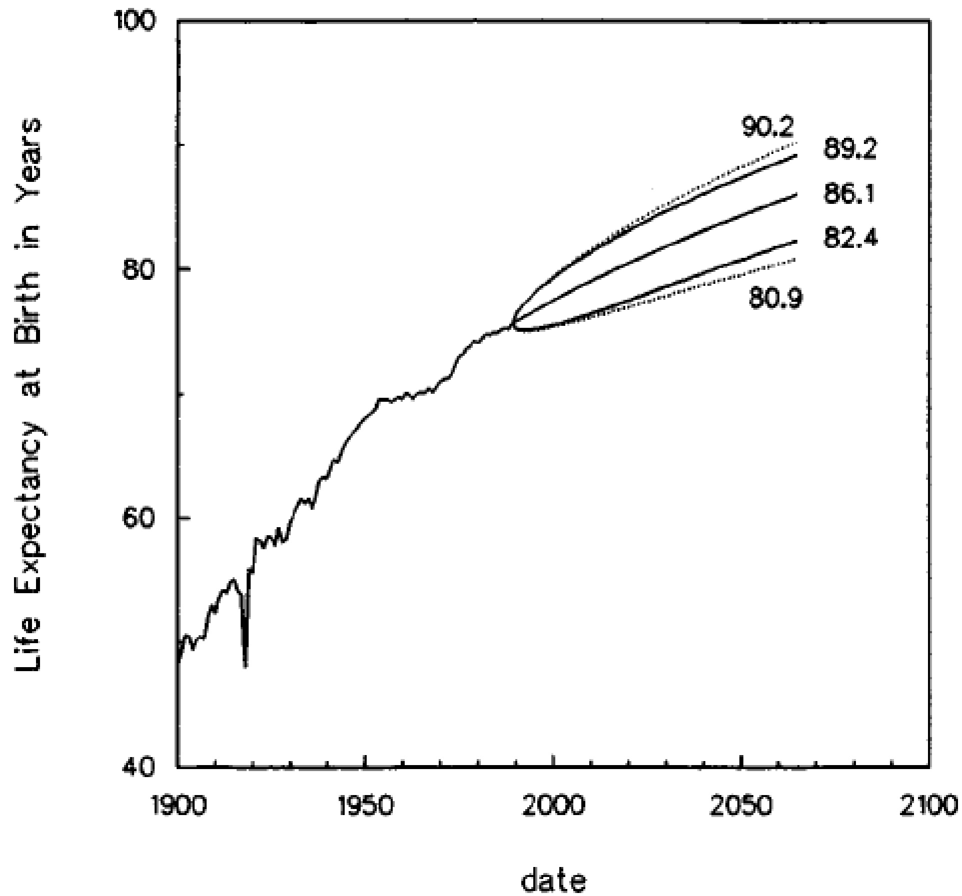


Figure 1. Actual U.S. Life Expectancy and Forecasts (95% Confidence Intervals With and Without Uncertainty From Trend Term). The forecasts use a (0, 1, 0) model with a flu dummy estimated on mortality data from 1900 to 1989. The 95% confidence intervals are shown with and without uncertainty from drift.

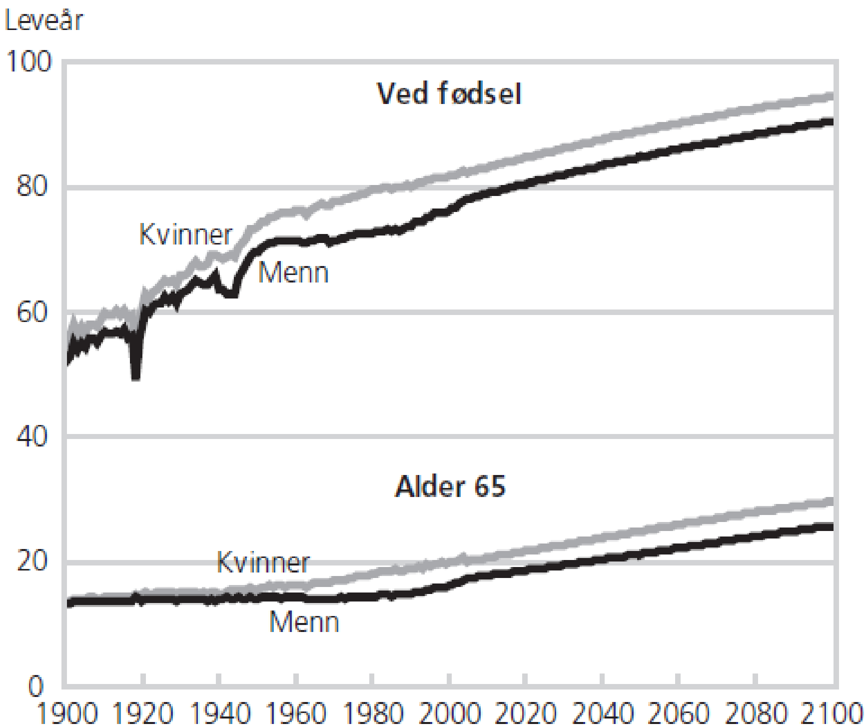
Applied to Norway, data from 1900-2004.
The simple Lee-Carter model did not fit the data satisfactorily. Needed to add one more *b.k* term in the model

Then the model fits the data well

$$\log m(x,t) = a(x) + b_1(x)k_1(t) + b_2(x)k_2(t) + \varepsilon(x,t)$$

Keilman og Pham (2005), Økonomiske analyser 6/2005

Figur 7. Forventet levealder ved fødsel og forventet gjenstående levetid på alder 65, 1900-2100



Population size forecasts

- Need also to forecast fertility rates, can use a Lee-Carter model
- Need to forecast migration (emigration and immigration)

R-packages for survey sampling

- Survey analysis in R:

http://r-survey.r-forge.r-project.org/survey/index.html

File Edit View Favorites Tools Help

Survey analysis in R

This is the homepage for the "survey" package, which provides facilities in R for analyzing data from complex surveys. The current version is 3.29. A much earlier version (2.2) was published in Journal of Statistical Software

An experimental package for very large surveys such as the American Community Survey can be found here

A port of a much older version of the survey package (version 3.6-8) to S-PLUS 8.0 is available from CSAN (thanks to Patrick Aboyoun at Insightful).

Features:

- Means, totals, ratios, quantiles, contingency tables, regression models, loglinear models, survival curves,rank tests, for the whole sample and for domains.
- Variances by Taylor linearization or by replicate weights (BRR, jackknife, bootstrap, multistage bootstrap, or user-supplied)
- Multistage sampling with or without replacement.
- PPS sampling with or without replacement: Horvitz-Thompson and Yates-Grundy estimators and a range of approximations.
- Post-stratification, generalized raking/calibration, GREG estimation, trimming of weights.
- Two-phase designs. Estimated weights for augmented IPW estimators.
- Graphics
- Support for using multiply imputed data
- Database-backed design objects for large data sets (now with replicate weights, too)
- Some support for parallel processing on multicore computers.
- Multivariate analysis: principal components, factor analysis (experimental).
- Likelihood ratio (Rao-Scott) tests for glms, Cox models, loglinear models.

The NEWS file gives a history of features and bug fixes.

Comparison shopping:
Alan Zaslavsky keeps a comprehensive list of survey analysis software for the ASA Section on Survey Research Methods.

User-generated ratings and reviews of this package (and others) at crantastic.

Using the survey package:

- Specifying a survey design
- Creating replicate weights
- Simple summary statistics
- Using supplied replicate weights
- Domain (subpopulation) estimation
- Tables of summary statistics
- Post-stratification and calibration
- Lonely PSUs
- Regression models
- Tests of association
- Stratification within PSUs
- Graphics
- Multiple imputation and ordinal logistic regression
- Database-backed survey objects
- Programming with survey objects