1 Sex-Age-CalendarTime Patterns in population mortality rates in Denmark

Exercise 1: Use the same informal approach as earlier (OR – only if interested– a median polish), to fit a multiplicative model to the slightly larger dataset consisting of the 24 rates for all 3 periods i.e., to the data involving the 3 periods 1980-84, 2000-2004 and 2005-2007.

Yrs	Age	Female (F)			Male (M)			
	70-	R_F			R_F		$\times M_M$	
'80-	75-	R_F	$\times M_{75}$		R_F	$\times M_{75}$	$\times M_M$	
'84	80-	R_F	$\times M_{80}$		R_F	$\times M_{80}$	$\times M_M$	
	85-	R_F	$\times M_{85}$		R_F	$\times M_{85}$	$\times M_M$	
	70-	R_F		$\times M_{20y}$	R_F		$\times M_M$	$\times M_{20y}$
'00-	75-	R_F	$\times M_{75}$	$\times M_{20y}$	R_F	$\times M_{75}$	$\times M_M$	$\times M_{20y}$
'04	80-	R_F	$\times M_{80}$	$\times M_{20y}$	R_F	$\times M_{80}$	$\times M_M$	$\times M_{20y}$
	85-	R_F	$\times M_{85}$	$\times M_{20y}$	R_F	$\times M_{85}$	$\times M_M$	$\times M_{20y}$
	70-	R_F		$\times M_{25y}$	R_F		$\times M_M$	$\times M_{25y}$
'05-	75-	R_F	$\times M_{75}$	$\times M_{25y}$	R_F	$\times M_{75}$	$\times M_M$	$\times M_{25y}$
'07	80-	R_F	$\times M_{80}$	$\times M_{25y}$	R_F	$\times M_{80}$	$\times M_M$	$\times M_{25y}$
	85-	R_F	$\times M_{85}$	$\times M_{25y}$	R_F	$\times M_{85}$	$\times M_M$	$\times M_{25y}$

R = rate. M = multiplier. The array called 'r' in the R code (which fits additive models to the rates and logs of the rates) can be used to calculate ratios.

...Year.......Age...Female...Male.....Total... Observed rates

1980-1984	70-74	0.02725	0.05213	0.03814
1980-1984	75-79	0.04592	0.08235	0.06042
1980-1984	80-84	0.08098	0.12163	0.09561
1980-1984	85-89	0.13680	0.18202	0.15193
2000-2004	70-74	0.02666	0.03972	0.03261
2000-2004	75-79	0.04179	0.06586	0.05189
2000-2004	80-84	0.06923	0.10584	0.08279
2000-2004	85-89	0.11970	0.16773	0.13480
2005-2007	70-74	0.02359	0.03468	0.02874
2005-2007	75-79	0.03934	0.05815	0.04750
2005-2007	80-84	0.06559	0.09622	0.07730
2005-2007	85-89	0.11462	0.15808	0.12860

Age multipliers:

The rate in the (females 70-74, 1980-84) cell is 0.02725, while that in the cell one below it (75-79) is 0.04592, yielding an empirical rate ratio of 1.69 for the pure 75-79 vs 70-74 contrast. We can repeat the same 75-79 vs 70-74 contrast for each of the other 5 sex-calendar year combinations, to obtain in all six 75-79 vs 70-74 ratios:

Years	Age	Female (F)	Male (M)
1980-1984	70-74	1	1
	75-79	**1.69**	**1.57**
2000-2004	70-74	1	1
	75-79	**1.58**	**1.66**
2005-2007	70-74	1	1
	75-79	**1.67**	**1.68**

One way, without even using a calculator, to arrive at a best estimate of the M_{75} multiplier is to make the median, 1.66, of these 6 estimates.

Moving on to the the pure 80-84 versus 70-74 contrast, we obtain 6 rate ratio estimates: 2.97, 2.60, 2.33, 2.66, 2.78 and 2.77; their median is 2.72.

For the 85-89 versus 70-74 contrast, the median of the 6 estimates is 4.52.

These three multipliers can be used to derive multiplicative rate (i.e., insurance premium) increases for the higher age categories, using the rates in the 70-74 group as the reference or 'starter' or 'corner' category ('corner' is Clayton and Hills terminology in their chapter 22).

It seems that rates double about every 7 years or so. Note also that the estimated 10 year increase of 2.72 is virtually the same as 1.66^2, so in fact we could use two 66% 5-year increases, 1 each per 5 years of age, and avoid having (to memorize/estimate) a separate multiplier for the 10 years of age increase. Note also that $1.66^3 = 4.57$ which is quite close to the fitted 4.52. So, in fact we could save having to memorize not just 1 but 2 multipliers, and simply say the rates in those ages 75-79, 80-84 and 85-89 are 1.66, 1.66^2, and 1.66^3 times the rates in those aged 70-74.

Another way to say this is that the *logs* of the mortality rates are *linear* in *age*. This finding is not new: The actuary Benjamin Gompertz described this pattern as a Law of Mortality (that now bears his name) in a paper in 1825. And William Farr and Thomas R Edmonds, and Gompertz, used this smooth functions relationship to save a lot of steps in the otherwise tedious lifetable calculations used in actuarial and population-lifetable analyses. When we come to formally fitting multiplicative rate (ie log linear) models for rates, the fact that the log rates seem to be close to linear over this age range means that we do not have to model age as a 'categorical' variable with 3 indicator variables (3 separate coefficients) but instead can be parsimonious (economical, even frugal) and use just 1 linear age term and its 1 associated regression coefficient.

Male multiplier:

The rate in the (females 70-74, 1980-84) cell is 0.02725, while that in the cell to the right of it (Males) is 0.05213, yielding an empirical rate ratio of 1.91 for the pure M vs F contrast. We can repeat the same M vs F contrast for each of the other 11 age-calendar year combinations, to obtain in all twelve M vs F ratios:

Yrs	Age	Female (F)	Male (M)
	70-74	1	**1.91**
'80-	75-79	1	**1.79**
'84	80-85	1	**1.50**
	85-90	1	**1.33**
	70-74	1	**1.49**
'00-	75-79	1	**1.58**
'04	80-84	1	**1.53**
	85-	1	**1.40**
	70-74	1	**1.47**
'05-	75-79	1	**1.48**
'07	80-84	1	**1.47**
	85-	1	**1.38**

The median of these 12 estimates is 1.48; one interpretation is that males should pay 48% higher life insurance premiums than females!

20-year multiplier: unchanged from in smaller dataset

The rate in the (females 70-74, 1980-84) cell is 0.02725, while that in the cell 4 cells below it (also females-70-74, but 20 years later) is 0.02666, yielding an empirical rate ratio of 0.98 for the pure '20 calendar years' contrast. We can repeat the same contrast for each of the other 7 age-sex combinations, to obtain in all eight 2000-2004 vs 1980-1984 ratios:

Age	Female (F)	Male (M)
70-74	**0.98**	**0.76**
75-79	**0.91**	**0.80**
80-84	**0.85**	**0.87**
85-89	**0.88**	**0.92**

The median of these 8 estimates is 0.88 representing a reduction of 12% in mortality in the 20 years between 198-1984 and 2000-2004.

25 (24?)-year multiplier:

The rate in the (females 70-74, 1980-84) cell is 0.02725, while that in the cell 8 cells below it (also females-70-74, but 24 years later) is 0.02359, yielding an empirical rate ratio of 0.87 for the pure '24 calendar years' contrast. We can repeat the same contrast for each of the other 7 age-sex combinations, to obtain in all eight 2005-2007 vs 1980-1984 ratios:

Age	Female (F)	Male (M)
70-74	**0.98**	**0.66**
75-79	**0.86**	**0.71**
80-84	**0.81**	**0.79**
85-89	**0.84**	**0.87**

The median of these 8 estimates is 0.82 representing a reduction of 18% in mortality in the 24 years between 1980-1984 and 2005-2007.

corner term (a.k.a. the 'intercept':

Whereas all of the other estimates used a synthesis of several estimates, it is not immediately obvious whether we are forced to use the one observed value in the 'corner' cell as the best fitted value for that cell. But for now, lets use it as the corner estimate, so that we can write a master equation for all 24 rates

The equation is for the rate in any given age-group in a given gender in a given calendar period:

Rate =	0.02725	×1.66	×2.72	×4.52	×1.48	×0.88	×0.82
		if	if	if	if	if	if
		75-79	80-84	85-89	male	2000-04	2005-07
$\log[Rate] =$	-3.603	+0.509	+1.000	+1.509	+0.395	−0.136	−0.194
		if	if	if	if	if	if
		75-79	80-84	85-89	male	2000-04	2005-07
$\log[Rate] =$	β_0	$+\beta_{\cdot 75'}$	$+\beta_{\cdot 80'}$	$+\beta_{\cdot 84'}$	$+\beta_M$	$+\beta_{\cdot 20y'}$	$+\beta_{\cdot 25y'}$
		×	×	×	×	×	×
		I_{75-79}	I_{80-84}	I_{85-89}	I_{male}	$I_{2000-04}$	$I_{2005-07}$

where each 'I'' is a (0/1) indicator of the category in question.

By using both the 0 and 1 values of each I, this 7-parameter equation produces a fitted value for each of the $4 \times 2 \times 3 = 24$ cells.

You can also think of I_{75-79}, I_{80-84}, and I_{85-89} as 'radio buttons': at most 1 of them can be 'on' at the same time, since there are 4 age levels in all.

1.1 More formal fitting of 6 parameter values

It shouldn't have to be, in the model fitting above, that the intercept was forced to go through an observed value, when we know that that value (like each of the 15 others) is subject to sampling variation. A fitted regression line or curve that goes *between* the dots [as opposed to one that actually *joins* the (error-containing!) dots] recognizes the fact that none of the observed data-points is 'perfect.' Also the purpose of the line is as a 'line of means' or 'line of centres.'

One option to avoid the arbitrariness in fitting an intercept is to apply a **median polish** to the log-rates. You can look up this procedure on the web, and the c634 course website provides some code for carrying it out (It seems that the `medpolish` function in `R` just handles 2 dimensional arrays, whereas the homemade R function is designed for ≥ 2 dimensions.

The fitted values from the median polish of the $4 \times 2 \times 3$ array of log rates are given in the next column

Converting them back to rates, and scaling them all so that the corner is 1, we get the following fitted rate ratio model:-

RateRatio =	1	×1.68	×2.71	×4.49	×1.49	×0.88	×0.79
		if	if	if	if	if	if
		75-79	80-84	85-89	male	2000-04	2005-07

-3.505	-3.106
-3.005	-2.606
-2.510	-2.111
-2.002	-1.603
-3.633	-3.234
-3.133	-2.734
-2.637	-2.239
-2.130	-1.731
-3.739	-3.340
-3.240	-2.841
-2.744	-2.345
-2.236	-1.838