Problem Set 2

Due March 15, 2006

[1] U.S. Mortality

File `us1992a.dat` in the dataset section of the course website has data on deaths and exposure for U.S. white and black males in 1992 (Monthly Vital Statistics Report, vol 42, Number 6S, 22 March 1996, NCHS). The file has variable names and columns are separated by commas, so you can read it into Stata using `insheet`.

(a) Calculate age-specific period mortality rates for 1992 for blacks and whites.

(b) Plot the rates against age on a log scale. Exclude the last group, which is actually 85+.

(c) Fit a Gompertz model over an appropriate age range.

(d) Interpret the resulting parameters in terms of the underlying hazard rate and the rate of aging.

(e) Could one include the 85+ group? What assumption(s) might be needed?

[2] Model Life Tables

Consider the life table of South African females you estimated in problem set 1. (We return to it again in question 5.)

(a) Compute Brass's logit transformation and plot both the data and a linear fit against the general standard (available in `brassrlm5`).

(b) Can you find a table in the Coale-Demeny family that does a better job? (Your search doesn't have to be exhaustive, just pick one or two indicators to match in each family.)

(c) Why do you think the model life tables have a bit of trouble fitting this life table?

[3] Child Survival

An analysis of infant and child mortality in Guatemala using proportional hazards models treats the hazard as constant within narrow age intervals, and considers two explanatory variables: age of mother (<20, 20-39 and 40p) and birth order (1-5 and 6+). The table below shows the baseline hazard for age in months and the estimated effects of the two predictors, with mothers 20-39 and births of order 1-5 as the reference group.

Age of Child	Hazard
0	.0457006
1-5	.0033105
6-11	.0040912
12-23	.0024781
24+	.0004648

Age of Mother	Hazard Ratio
<20	1.200587
20-39	1
40+	1.532668

Birth Order	Hazard Ratio
1-5	1
6+	1.309283

(a) What's the probability of infant death (by 12.0 months) for the reference group?

(b) Describe the effect of age of mother on child survival

(c) What's the probability of infant death for a 6th child of a 44 year old mother?

(These are real data from Guatemala, analyzed by Pebley and Stupp and later Guo and Rodriguez. The model considered in those papers is considerably more complex, but the two variables considered here are important predictors. The length of the preceeding birth interval is also very important.)

[4] Unobserved Heterogeneity

Consider two groups of individuals who face a constant hazard from age 20 onwards but are heterogenous, with gamma distributed frailty that averages 0.02 in group one and 0.03 in group two at age 20, and has a variance of one in both groups. Note that we only work from age 20 onwards.

(a) Group two has 50% higher risk at age 20. What's the risk ratio at age 40? At age 60? At age 80?

(b) Suppose the high risk group represented half the total population at age 20. What fraction of the total would it be at age 40? 60? 80?

Challenge: Can you answer these questions under the more realistic assumption that the baseline hazard is Gompertz from age 20 onwards, with a log hazard of -7 at age 20 and an aging rate of 0.08 for group one?

[5] Competing Risks

I have updated file `safed.dat` used in Problem Set 1 to include counts of deaths due to AIDS for South African females in 1998.

(a) Construct a multiple decrement life table distinguishing deaths due to AIDS and to other causes.

(b) What's the probability that a woman who dies before age 50 dies of AIDS?

(c) Construct the associated single decrement table in the absence of AIDS.

(d) How many years of life expectancy are lost to AIDS?

(e) What assumption underlies the calculations in c and d? How do you think the true impact of AIDS on life expectancy differs from your estimate in d?