HT 2007: Statistical Lifetime Models, Sheet 6

1. Coronary Heart Disease (CHD) remains the leading cause of death in many countries. The evidence is substantial that males are at higher risk than females, but the role of genetic factors versus the gender factor is still under investigation. A study was performed to assess the gender risk of death from CHD, controlling for genetic factors. A dataset consisting of non-identical twins was assembled. The age at which each person died of CHD was recorded. Individuals who either had not died or had died from other causes had censored survival times (age). A randomly selected subsample from the data is as follows. (* indicates a censored observation.)

Age male twin	Age female twin
50	63*
49*	52
56*	70*
68	75
74*	72
69*	69*
70*	70*
67	70
74*	74*
81*	81*
61	58
75*	73*

   (a) Assuming the censoring mechanism is independent of death times due to CHD, perform the log-rank test for equivalence of hazard amongst these two groups. Contrast the test statistic and associated p-value with the results from the Fleming Harrington test using a weight $W(t_i) = \widehat{S}(t_{i-1})$.

   (b) Do you think the assumption of an independent censoring mechanism is appropriate? Give reasons.

2. A life office wishes to compare mortality of its male assurance policyholders with males in a standard table. A chi-squared test has been used to test the hypothesis that the true underlying mortality rates of the policyholders are the same as those for the standard table. Based on the results of this test the null hypothesis has been accepted. Explain how the Chi-squared test statistic is constructed.

   (a) Give three possible defects that the Chi-squared test may fail to detect, giving reasons.

   (b) Describe a suitable test for dealing with one of these defects, again giving reasons.

3. A large investigation has been carried out into mortality among people of working age. They are to be compared with a well-known standard table.

Age	Exposed to risk E_x	Observed deaths d_x	Expected deaths $E_x q_x^s$	Standardised deviations z_x
$20-24$	35000	35	34	0.17150
$25-29$	33000	30	29	0.18569
$30-34$	30000	31	35	-0.67612
$35-39$	30000	45	52	-0.97072
$40-44$	31000	84	80	0.44721
$45-49$	28000	138	130	0.70165
$50-54$	25000	229	213	1.09630
$55-59$	23000	360	348	0.64327
$60-64$	20000	522	505	0.75649

The standardised deviations are calculated using the approximation $E_x q_x^s \simeq E_x q_x^s (1 - q_x^s)$ to estimate the variance within each age range.

Perform the following three tests finding the p-values and the test statistic (where appropriate):

a) Chi-squared b) sign test c) cumulative deviations test, commenting on the outcomes.

What might you do if you wished to use this dataset in order to estimate the number of deaths (possibly for establishing premium levels)?

4. The following is an investigation carried out by a (medium-sized) UK pension scheme into the mortality of its pensioners between 2000-2002.

   (a) Explain why the crude rates are usually graduated.

   (b) The data used to produce the crude rates and the proposed gradauted rates are as follows.

Age	Central ExpRisk	Deaths	crude hazard	graduated hazard	Stand Deviation
x	E_x^c	d_x	$\mu_{x+0.5}$	$\overset{o}{\mu}_{x+0.5}$	z_x
$60-64$	1388.9	10	0.0072	0.0061	0.5249
$65-69$	1188.8	17	0.0143	0.0131	0.3615
$70-74$	880.5	28	0.0318	0.0262	1.0266
$75-79$	841.6	34	0.0404	0.0487	-1.0912
$80-84$	402.8	41	0.1018	0.0839	1.2394
$85-89$	123.9	19	0.1533	0.1338	0.5949
$90-94$	27.9	7	0.2509	0.1975	0.6346
$95-99$	10.0	3	0.3000	0.2706	0.1787
$100+$	7.5	2	0.2666	0.3455	-0.3673

   Assume that the Gompertz-Makeham model has been used for graduation. Is this a sensible choice of model?

   Test the proposed graduation for:

   i. overall goodness of fit; and

   ii. bias.

   (c) Comment on the use of the graduated hazard rates in calculating benefits paid from the scheme.