Class Notes for Chapter 8

- **Survival data (i.e., 'time until' data) is typically skewed and censored. Censored data means right (most common), left, and/or interval. We'll focus mostly on right-censored data analysis, although the nonlinear model in Section 8.6 addresses interval-censored data.**
- **Notation – need to get at an understanding of the survival and hazard functions. Survival function gives the probability of "survival" beyond some time-point. Hazard function gives the hazard/risk of 'dying' in the next instance given survival up to time t.**
- **The Weibull distribution is considered and illustrated here since it is useful in applications (especially reliability and engineering) and since it contains the Exponential distribution as a special case.**
- **Equation (8.7) is a key result – it gives the all-important LL!**
- **Example 8.1 (Carcinoma) – fit the Weibull distribution to these data; two values are right censored. MLE's are given in Output 8.1a. Non-parametric (i.e., without any distributional assumption) estimate of the survival function: this estimate is called the Kaplan-Meier estimate. We can use SAS' LIFETEST procedure to get it. Using both parametric and non-parametric methods, we obtain point and interval estimates of the median survival time (like LD_{50}) – see pp.5-6 for KM approach.**
- **Example 8.2 (IUD) – first graph is of N2LL for Exponential distribution, and second graph is obtained by profiling out the nuisance parameter (γ) and gives the N2PLL for the median survival time; intervals obtained by 'cutting'.**

- **Cox's Proportional Hazards (CPH) model is used to relate hazard functions which vary with x, h(t,x), to a baseline hazard (h_0) – the covariate vector (x) is assumed to enter as in Equation 8.8, or equivalently, as in Equation 8.9; covariates may include age, gender, etc. When there are two treatments (e.g., drugs A & B), x is just a dummy variable, and the slope parameter is given in Equation 8.10.**
- **Another model for survival data is the Accelerated Failure Time [AFT] model in Equation 8.12; it too brings covariate(s) into the model. Equations 8.12 and 8.13 are just for a dummy x, but can easily be extended to include other covariates.**
- **Example 8.3 (Breast Cancer) – the estimated survival curves on p.10 (obtained by KM estimation) appear to differ for the positively and negatively stained tumors: is this difference statistically significant or just an artifact?**

  (1) SAS LIFETEST procedure provides nonparametric estimates of quartiles and medians for the two groups, and, in Output 8.2c, three distinct tests of this question (null: no difference between the two curves). Many researchers use the log-rank test, but this may merely be by habit.

  (2) SAS PHREG procedure fits the CPH model, gives an estimate of β in Output 8.2d, $\widehat{\beta}$ = 0.9080 – note that the likelihood test gives marginal significance here (p = 0.0491). The fitted result is given in Equation 8.11: the 'positive' survival curve is shifted down from the 'negative' survival curve. Important interpretation: the hazard associated with the positive stain group is 2.479 times the hazard associated with the negative stain group – for all values of t.

(3) SAS LIFEREG (with Weibull distribution) fits the *parametric* Cox PH model – here the likelihood test that $\beta = 0$ gives p = 0.0418. *Note the doubling used in finding the TS here!*

(4) SAS' LIFEREG (with Log-logistic distribution) fits the (parametric) AFT model – fitted results appear in Equation 8.14.

- Parametric, non-Normal SLR with censored data – use the LL given above to handle the censored measurements. For the Weibull distribution, median is more natural, so we use median in Equation 8.16 instead of mean in Equation 8.15.
- Example 8.4 – Y = time until death (no censored data here) and X is dose: wish to connect by a line. Intercept (β_0), slope (β_1) and variance-type parameter (γ) are estimated in Output 8.3a; the test of no significant relationship (H_0: $\beta_1 = 0$) is simple. Also simple is the test of LOF (lack of fit) of the assumed line since we could fit the full model: this is either the one-way ANOVA with 4 levels or a cubic polynomial. The ANOVA results are in Output 8.3b, and the line is fine here ($\chi_2^2 = 266.0 - 265.7 = 0.3$, NS).
- Interval-censored data are addressed in Example 8.5, so the binary logistic modelling ($\chi_1^2 = 31.1 - 26.4 = 4.7$, p = 0.0302) in Appendix 8.8.2 is wrong. Using the nonlinear model of Farrington (1996) and Collett (2003:286) in conjunction with the log-logistic distribution, we obtain Appendix 8.8.3 (p.22). Here the results are NS ($\chi_1^2 = 128.4 - 126.7 = 1.7$, p = 0.1932). We conclude there is no difference between drugs A and B.
- Next issues:
  - nonlinear modelling with censored and skewed data
  - detecting synergy with censored and skewed data.