I. Causal inference for an internal variable is difficult

3. Likelihood:
   a. Model:
      i. Ex., remission status
      ii. Ex., transplant status
   b. Estimating base line hazard rate \( h_0 \) is very difficult.

3. Advantages and Disadvantages of Stratification
   a. do calculations separately in each strata
   b. Expect some variable to influence hazard in all strata in the same way
   c. Testing is straightforward
   d. Derivatives and variances sum
      i. Testing is straightforward
      ii. As alternative to putting interaction with time in for stratification variable
   e. Artificial variables used to assess proportional hazards assumption
      i. Random variables who change with no causal relation to response
      ii. Examples
         a. market rates for replacement mortgages when modeling mortgage prepayments
         b. effect of heat or pollution on mortality
   f. As alternative to putting interaction with time in for stratification variable

J. Stratification:
   1. Suppose
      a. expect hazards from different strata not to be proportional
      b. Expect some variable to influence hazard in all strata in the same way
      c. Strata act independently
   2. Then
      a. do calculations separately in each strata
      b. Partial log likelihood is sum of individual contributions
      c. Estimator maximizes this
      d. Derivatives and variances sum
      i. Testing is straightforward
      e. As alternative to putting interaction with time in for stratification variable
   3. Advantages and Disadvantages of Stratification
      a. Bonuses for stratification:
         i. Doesn’t depend on correct effect being linear
         ii. Computationally easier
      b. Drawbacks
         i. Won’t work for continuous covariate
         ii. Can’t simultaneously estimate effect of stratification variable

K. Diagnostics

1. Diagnostic we have seen before:
   a. Are cumulative hazard functions parallel?
   b. Testing time dependent covariate R Code SAS Code

2. What we will look for
   a. Shape of relationship with covariate
      i. log scale or original scale
      ii. change points
      iii. etc.
   b. Adequacy of proportional hazards
   c. Presence of outliers
   d. Leverage

KM: 11.2a

3. Cox and Snell residuals for general regression models:
   a. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
   b. Residuals are $R_j = h_j(T_j, \beta)$
   c. Examples
      i. for normal theory regression $h_j(t, \beta) = t - z_j \beta$
      ii. For logistic regression, $h_j(t, \beta) = (t - n_j \pi_j) / \sqrt{n_j \pi(1 - \pi)}$
   d. General definition precedes PH regression.
   e. Distribution of Hazard at Event times exponential.
      i. may be determined through transformaton
      ii. $P[T > t] = S(t)$
      iii. $P[S(T) < S(t)] = S(t)$
      iv. $P[- \log(S(T)) > - \log(S(t))] = S(t)$

4. Cox and Snell residuals for proportional hazards:
   a. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
      i. Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$
   b. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$
   c. Let $R_j = H_0(T_j) \exp(z_j \beta)$, possibly censored.
   d. Analogy with regression:
      i. For small samples you would want to adjust for negative correlation
      ii. There is no corresponding proportional hazards adjustment, so do not use for small samples.
   e. Can also be used for time dependent covariates, if you had $\hat{H}_0$
   f. Can be done within a stratum for a stratified model
   g. $\sum_j R_j \approx \sum_j \delta_j$
      i. Exact for $\hat{H}(t_i) = \sum_{k=1}^{i} \sum_{j \in R(t_i)} \frac{D_k}{\exp(z_j \beta)}$
         KM: 11.2b
   h. See if CDF of $\hat{H}(T_i)$ like a line through 0 with slope 1
      i. Use Nelson–Aalen estimator to account for censoring.

This page intentionally left blank.