Chapter 13 Multivariate Survival Analysis

PubH 7450

©Wei Pan

Email: weip@biostat.umn.edu

Http: www.biostat.umn.edu/∼weip

§13.1 Introduction

- So far, assume all T_j's (or X_j's) are independent
- Not always true: e.g. litter effects
  Table 13.1
  Data: $(T_{ij}, \delta_{ij}, Z_{ij})$, mouse j from litter i;
  because shared genetic and environmental effects, $T_{i1}, ..., T_{in_i}$
  are usually *correlated*!—-well-known litter effects!
  valid analysis needs to account for within-litter correlations.

ctical Note

1. An SAS macro to compute this test is available on our web site.

TABLE 13.1
Data On 50 Litters of Rats

Group	*Treated Rat*	*Control Rats*	*Group*	*Treated Rat*	*Control Rats*
1	101^+	104^+, 49	26	89^+	104^+, 104^+
2	104^+	104^+, 102^+	27	78^+	104^+, 104^+
3	104^+	104^+, 104^+	28	104^+	81, 64
4	77^+	97^+, 79^+	29	86	94^+, 55
5	89^+	104^+, 104^+	30	34	104^+, 54
6	88	104^+, 96	31	76^+	87^+, 74^+
7	104	94^+, 77	32	103	84, 73
8	96	104^+, 104^+	33	102	104^+, 80^+
9	82^+	104^+, 77^+	34	80	104^+, 73^+
10	70	104^+, 77^+	35	45	104^+, 79^+
11	89	91^+, 90^+	36	94	104^+, 104^+
12	91^+	92^+, 70^+	37	104^+	104^+, 104^+
13	39	50, 45^+	38	104^+	101, 94^+
14	103	91^+, 69^+	39	76^+	84, 78
15	93^+	104^+, 103^+	40	80	80, 76^+
16	85^+	104^+, 72^+	41	72	104^+, 95^+
17	104^+	104^+, 63^+	42	73	104^+, 66
18	104^+	104^+, 74^+	43	92	104^+, 102
19	81^+	104^+, 69^+	44	104^+	98^+, 78^+
20	67	104^+, 68	45	55^+	104^+, 104^+
21	104^+	104^+, 104^+	46	49^+	83^+, 77^+
22	104^+	104^+, 104^+	47	89	104^+, 104^+
23	104^+	83^+, 40	48	88^+	99^+, 79^+
24	87^+	104^+, 104^+	49	103	104^+, 91^+
25	104^+	104^+, 104^+	50	104^+	104^+, 79

$^+$ Censored observation

2-1

- Correlated data
  - Clustered data: multiple possibly correlated observations from each cluster, and many independent clusters.
    e.g., familial data with multiple family members from each family; observations on the two eyes/kidneys from the same subject;......
  - Longitudinal data/repeated measures: multiple measurements from the same subject over time.
    e.g. recurrent events: onset of disease, smoking, etc.
- Two (or three?) general approaches:
  - Random-effect model: use random-effects to explicitly account for within-cluster correlation; usually need a parametric assumption on the distribution of the random effects.
    called frailty models in survival analysis.

- Marginal model: do not model within-cluster correlation explicitly in a regression model; but some adjustment is made for inference; less assumptions.
  analogous to GEE.

- Here we consider semi-parametric PHM.
  parametric PHM/AFT is straightforward: ML

§13.3 Frailty models

- Data: $(T_{ij}, \delta_{ij}, Z_{ij})$, $j = 1, ..., n_i$, $i = 1, ..., n$.
- Model:
  $h_{ij}(t) = h(t|Z_{ij}, i) = h_0(t)u_i \exp(\beta' Z_{ij}) = h_0(t) \exp(v_i + \beta' Z_{ij})$,
  where $u_i > 0$, $u_i \overset{iid}{\sim} Gamma(\theta)$ with mean=1 and var=θ, and
  $v_i = \log(u_i)$.
- Feature of the model: explicitly to model cluster-specific effects u_i, which account for within-cluster correlations.
- Interpretation of u_i: ......
- A larger θ: larger heterogeneity and stronger within-cluster association.
  $S(x_{i1}, ..., x_{in_i}) = Pr(X_{i1} > x_{i1}, ..., X_{in_i} > x_{in_i}) =$
  $[1 + \theta \sum_{j=1}^{n_i} H_0(x_{ij}) \exp(\beta' Z_{ij})]^{-1/\theta} \neq \prod_{j=1}^{n_i} Pr(X_{ij} > x_{ij})$.
- Interpetation of β:

Effect of a covariate (i.e. log HR) after adjusting for other covariates **and** ......
–subject-specific effect!

- Technical difficulty: as in any random-effects model, need to integrate out u_i's (which is hard) to get a marginal likelihood.
- Model fitting is complicated; see text for an EM-type algorithm (Klein 1992).
S+/R uses a penalized likelihood as an approximation.

§13.4 Marginal models

- Data: $(T_{ij}, \delta_{ij}, Z_{ij})$, $j = 1, ..., n_i$, $i = 1, ..., n$.
- Model:
  $h_{ij}(t) = h(t|Z_{ij}) = h_0(t) \exp(\beta' Z_{ij})$.
- Feature of the model: marginal; why called marginal?
  no explicit cluster-specific effects.
- No explicit account of within-cluster correlations in the model; but to draw inference, one has to account for it.
- Can use the *working independence assumption* (i.e. **incorrectly** assuming that all X_{ij}'s or T_{ij}'s are independent) to get $\hat{\beta}_I$
- Based on estimating function theory, $\hat{\beta}_I$ is consistent and asymptotically Normal.
- However, the usually information matrix under the *working*

independence model **cannot** be used to estimate $Cov(\hat{\beta}_I)$; use so-called empirical/robust/sandwich estimator, see p.437.

- A simple analog: $X_i \sim N(\mu, \sigma)$ for $i = 1, ..., n$; X_i's may be correlated.
  $\hat{\mu} = \bar{X}$ is unbiased for μ;
  but S^2/n is biased for $Var(\hat{\mu})$, where S^2 is the sample variance.
- Interpretation of β:
  log HR as before;
  Is it the log HR after adjusting for the other covariates?
  ......
  population averaged effects, as compared to *subject-specific* effects in a frailty model.
- The above method is called Wei-Lin-Weisfeld method, analogous to the GEE for GLMs with correlated data.
- In SAS:

```
Proc Phreg covs(aggregate);
Model ...;
ID subj;
```

- Example 13.1: R

§Fixed-effects models

- Data: $(T_{ij}, \delta_{ij}, Z_{ij})$, $j = 1, ..., n_i$, $i = 1, ..., n$.
- Model:
  $h_{ij}(t) = h(t|Z_{ij}, i) = h_0(t) \exp(v_i + \beta' Z_{ij}) =$
  $h_0(t) u_i \exp(\beta' Z_{ij}) = h_{0,i}(t) \exp(\beta' Z_{ij})$,
  where $h_{0,i}(t)$ is the baseline hazard function for cluster i.
- Contrast to the corresponding RE model:
  In RE-model: i) $h_{0,i}(t) = h_0(t) u_i$;
  ii) u_i is random with ...
  Then, why use RE- models?
- Contrast to the corresponding marginal model:
  Difference:
- How to fit?
- Example 13.1b: SAS

- Comments: the above three approaches, in logistic regression, correspond to logistc RE model, marginal logistic model (i.e. GEE) and conditional logistic regression!