Survival Analysis

Applications

Wilfried Karmaus

LCS829

Survival analysis

• Time-to-event analysis
  Family of statistical models, in which both the event (occurrence of a new disease) and the time to the event are taken into consideration.

• Life-table analysis
  Time-to-event analysis with aggregative data.

Survival analysis: rationale

• Not only the event counts but also whether the event occurs early or late during the observation period.

• Cox regression (proportional hazard regression) provides a statistical model with only little assumptions.

• We have to collect information on the time of the event and the time to the event.

• Important e.g. for survival in clinical trials (treatment effect), reproductive epidemiology, life insurances etc.
Censoring

- Subject sometimes withdraws from a study
- The study is incomplete before the endpoint is reached

Thus the status of study participants beyond the observation time is unknown.

Censoring and exposure

- It is assumed that censoring is independent of exposure.
- For instance, subjects in a clinical trial do not withdraw because of the treatment.
- Complete-data analysis produces unbiased estimates only if the censored observations are missing completely at random.
- If not, survival analysis and e.g. Kaplan-Meier estimations can help.

Estimating survivor functions

- **Non-censored data**: \( S(t) = 1 - P(T \leq t) \)
  \( T \leq t \) = surviving until \( t \)

  This is a special case of the KM method

- **Censored data**:
  - Kaplan-Meyer (KM) Method
    (uses ungrouped ordered failure times)
  - Life-Table method (uses grouped failure times)
  - Step function (usual approach)
  - Continuous function
    (uses exact survival and censored times)
**Survival function - non-censored data**

- $S(t_{j-1}) = \frac{n - \# \text{failures at or before } t_{j}}{n} = \frac{\# \text{ surviving past } t_{j-1}}{n}$

**Step function**

- Constant $S(t)$ all through a time interval
- $P[\text{surviving past } t_{j} | \text{surviving until } t_{j}] = P[T > t_{j} | T \geq t_{j}] = \frac{n_{j} - m_{j}}{n_{j}}$

**Survival function - censored data**

Calculation of Survival according to the Kaplan-Meier method:

$$S_{t} = \Pi_{i=0}^{j} (n_{i} - m_{i})/n_{i}$$

- $l_{ij}$ = interval
- $n_{j} =$ number of women who survived the interval $t_{j}$ without conception.
- $m_{j} =$ number of women who conceived in interval $t_{j}$.
- $q_{j} =$ number of women who ended at risk in interval $t_{j}$. 
Calculate the:
- cumulative incidence
- incidence rate (incidence density)

for Cohort A & B

Calculate the:
- risk ratio
- incidence rate ratio (incidence density ratio)

comparing Cohort B to A and vice versa

Move all individuals to the starting line.
Diagrammatic presentation of a five-year follow-up, Cohort B

\[ \Delta = \text{first occurrence of the disease} \]

\[ \square = \text{death due to the disease} \]

Move all individuals to the starting line.

**Cox proportional hazard model**

Basic options:

- Compare survivor curves estimated within separate strata of an explanatory variable
- Use regression modeling to evaluate effects of explanatory variables (using hazard ratios)
- **Proportional Hazard Model** is most popular (requires time-independent covariates → stratification)
- Alternatives: parametric model, e.g. Exponential, Weibull, Log Normal etc.
- Extension for time-dependent covariates

**Proportional Hazard Model**

\[ X = (X_1, X_2, \ldots, X_p) = \text{set of explanatory variables} \]

\[ h(t,X) = \text{hazard function for persons with variable set } X \]

Then PH model:

\[ h(t,X) = h_0(t) \times e^{\sum \beta_i X_i} \]

Where \( h_0(t) = \text{an arbitrary unspecified base-line hazard function} \)

= hazard when all covariates are at average value

PH model can estimate \( \beta \)'s without specifying \( h_0(t) \).
Proportional Hazard Model

- PH model preferred over logistic model if the survival time information is available: more info than just whether event occurred.

\[
HR = \frac{h_0(t)^* e^{\sum \beta_i X_{2i}}}{h_0(t)^* e^{\sum \beta_i X_{1i}}} = e^{\hat{\beta}(X_2 - X_1)}
\]

\(\beta\) is not a function of time

To estimate \(\beta\)

- To estimate \(\beta\) Cox (1972, 1975) introduced the partial likelihood function, which eliminates the unknown baseline hazard and accounts for censored survival times.

Time-dependent covariates (1)

- An exploratory variable (exposure) is time-dependent if its value for a given individual can change over time.
- For instance, time-dependent variables can be used to model the effect of subjects changing treatment groups.
Time-dependent covariates (2)
- PHREG procedure in SAS performs a regression analysis of survival data based on the Cox proportional hazards model.
- If the hazard ratio changes with time then the proportional hazards model assumption is invalid.
  → Model the change over time to estimate time-dependent covariates
  → Perform a stratified analysis if time-dependent covariate is a confounder

Your homework for survival analysis:
- To describe the fecundability (fertility) in different European countries (ESIS).
- Copy the file qbase into one directory
- Open the program:
  C1_qbase_grf_class2.sas
- Declare the path/directory in which you copied the file qbase.
- Conduct TASK 1 to 6 that are in the titles of the SAS program.

<table>
<thead>
<tr>
<th>months</th>
<th>n</th>
<th>m</th>
<th>q</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>560</td>
<td>0</td>
<td>1</td>
<td>0.284</td>
</tr>
<tr>
<td>1</td>
<td>560</td>
<td>157</td>
<td>0</td>
<td>0.425</td>
</tr>
<tr>
<td>2</td>
<td>403</td>
<td>60</td>
<td>3</td>
<td>0.392</td>
</tr>
<tr>
<td>3</td>
<td>322</td>
<td>70</td>
<td>5</td>
<td>0.473</td>
</tr>
<tr>
<td>4</td>
<td>257</td>
<td>26</td>
<td>1</td>
<td>0.626</td>
</tr>
<tr>
<td>5</td>
<td>236</td>
<td>18</td>
<td>1</td>
<td>0.360</td>
</tr>
<tr>
<td>6</td>
<td>211</td>
<td>48</td>
<td>3</td>
<td>0.303</td>
</tr>
<tr>
<td>7</td>
<td>199</td>
<td>10</td>
<td>0</td>
<td>0.284</td>
</tr>
<tr>
<td>8</td>
<td>160</td>
<td>10</td>
<td>1</td>
<td>0.265</td>
</tr>
<tr>
<td>9</td>
<td>139</td>
<td>10</td>
<td>1</td>
<td>0.246</td>
</tr>
</tbody>
</table>

Survival analysis, m = # with conception, q = # who ended being at risk