

Survival analysis

(Besprechung: Mo., 04.02.2013)

1. Gegeben sei der Datensatz $D = (T_i^*, \delta_i, Z_i)_{i=1}^n$. Die Kovariable Z ist eindimensional (und nimmt nur die Werte 0 und 1 an).

Vorsicht: wenn man die (partielle) Log-Likelihoodfunktion wie in der Vorlesung definiert ($\ell_p(\vec{\beta}) = -\log(L_p(\vec{\beta}))$), werden die Größen aus Aufgabe 1, Blatt 13:

$$\ell_p(\beta) = -d_1\beta + \sum_{j=1}^R \log(R_j^{(0)} + R_j^{(1)}e^\beta) \quad (1)$$

$$U(\beta) = -\left(d_1 - \sum_{j=1}^R \frac{R_j^{(1)}e^\beta}{R_j^{(0)} + R_j^{(1)}e^\beta}\right); \quad (2)$$

$$I(\beta) = \sum_{j=1}^R \frac{R_j^{(1)}e^\beta}{R_j^{(0)} + R_j^{(1)}e^\beta} - \left(\frac{R_j^{(1)}e^\beta}{R_j^{(0)} + R_j^{(1)}e^\beta}\right)^2 = \sum_{j=1}^R \frac{R_j^{(0)}R_j^{(1)}e^\beta}{(R_j^{(0)} + R_j^{(1)}e^\beta)^2}. \quad (3)$$

- (a) Wie sieht die Nullhypothese aus, wenn man prüfen möchte, ob es überhaupt Unterschiede zwischen den zwei Gruppen (Gruppe 1: $Z = 0$ und Gruppe 2: $Z = 1$) gibt?
- (b) Bestimme die Score-Teststatistik χ_{SC}^2 für diese Nullhypothese. Zeige, dass sie folgende Gleichung erfüllt:

$$\chi_{SC}^2 = \chi_{LRT}^2 = \frac{Z_1^2(\tau)}{V_{11}(\tau)},$$

wobei χ_{LRT}^2 die Teststatistik eines Log-Rank-Tests ist; τ ist so dass $\tau \geq \max\{t_1, \dots, t_R\}$.

Hinweis: $R_i^{(0)}$ entspricht $Y^{(1)}(t_i)$ und $R_i^{(1)}$ entspricht $Y^{(2)}(t_i)$ (für $i = 1, \dots, R$) beim Log-Rank-Test.

- (c) Betrachte den Datensatz **dataset** aus Blatt 13: Bestimme die Werte von $d^{(1)}, \vec{R}^{(0)} = (R_1^{(0)}, \dots, R_R^{(0)})^T$ und $\vec{R}^{(1)}$.
- (d) Wir wollen nun die Nullhypothese:

$$H_0 : \{\beta = 0.5\}$$

für **dataset** zum Niveau $\alpha = 5\%$ untersuchen. Berechne (mir R) die drei Teststatistiken χ_{LR}^2, χ_{SC}^2 und χ_W^2 aus der Vorlesung. Benutze hierfür die gegebenen Formeln (1), (2) und (3). Entscheide jeweils, ob die Nullhypothese angenommen oder abgelehnt wird.

- (e) Wiederhole das Ganze für die Nullhypothese

$$H_0 : \{\beta = 1.5\}.$$

- (f) Führe nun die drei Tests mithilfe von geeigneten Funktionen in R durch. Beachte hierzu insbesondere die **init**-Option bei **coxph**.

(8 Punkte)

2. Betrachte wieder den Datensatz `larynx` (aus `KMsurv`) und die drei Kovariablen $Z^{(1)}, Z^{(2)}, Z^{(3)}$. Benutze den Likelihood-Ratio-Test (mit verallgemeinerten Hypothesen) um zu bestimmen, ob die einzelnen Kovariablen signifikant zum Niveau $\alpha = 5\%$ sind. Definiere hierfür eine Funktion die, gegeben das vollständige und das 'partielle' Modell, Teststatistik und p -Wert berechnet. Beachte hierzu insbesondere die `loglik` und `coef` Argumente der zu vergleichenden Regressionsmodelle.

(4 Punkte)