

Survival analysis

(Besprechung: Mo., 17.12.2012)

1. Betrachte den Datensatz `ovarian` aus dem R-Paket `survival`. Er enthält die Ergebnisse einer Studie über zwei unterschiedliche Behandlungen für Ovarialkarzinom:

<code>futime</code>	Überlebenszeit, in Tage, nach der Diagnose
<code>fustat</code>	0 = hat die studie überlebt, 1=ist gestorben
<code>age</code>	Alter in Jahre
<code>residual.dz</code>	Anwesenheit von 'residual disease' (1=nein, 2=ja)
<code>rx</code>	Behandlungsgruppe (1 für Behandlung A, 2 für B)
<code>...</code>	...

Man möchte sehen, ob die Behandlungen unterschiedliche Effekte auf das Überleben der Patienten haben, ohne Berücksichtigung der anderen Kovariablen. Wir möchten Folgendes testen:

$$H_0 : h^{(1)}(t) = h^{(2)}(t) \quad t \leq \tau,$$

wobei $h^{(i)}(t)$ die Hazardrate für die Behandlung i und τ der spätestmögliche Zeitpunkt ist, bei dem in beiden Fällen noch Patienten unter Risiko sind.

Verwende hierzu die Teststatistik:

$$T(\tau) = \frac{D_L(\tau)}{\sqrt{\widetilde{V}_0(\tau)}},$$

mit

$$\widetilde{V}_0(\tau) = \sum_{t_j \leq \tau} L(t_j)^2 \frac{D_j^{(\bullet)}}{Y_j^{(1)} Y_j^{(2)}} \cdot \frac{Y_j^{(\bullet)} - D_j^{(\bullet)}}{Y_j^{(\bullet)} - 1},$$
$$D_j^{(\bullet)} := D_j^{(1)} + D_j^{(2)}$$
$$Y_j^{(\bullet)} := Y_j^{(1)} + Y_j^{(2)}$$

Unter H_0 , $T(\tau)$ ist asymptotisch normalverteilt.

Zur Erinnerung: $(Y_j^{(\bullet)} - D_j^{(\bullet)}) / (Y_j^{(\bullet)} - 1)$ ist ein Korrekturfaktor für 'ties' (wenn $D(t_j) > 1$): Es ist gleich 1 wenn keine 2 Individuen einen gemeinsamen Ereigniszeit haben.

Je nachdem, welche Gewichtsfunktion L man verwendet, trägt der entsprechende Test einen anderen Namen, zum Beispiel:

- **Log-rank Test**, wenn

$$L(t) = \frac{Y_j^{(1)}(t) Y_j^{(2)}(t)}{Y_j^{(1)}(t) + Y_j^{(2)}(t)};$$

- und **Harrington- Fleming Tests** wenn

$$L(t) = [\hat{S}(t^-)]^\rho \cdot \frac{Y_j^{(1)}(t)Y_j^{(2)}(t)}{Y_j^{(1)}(t) + Y_j^{(2)}(t)}, \quad \rho \in [-1, 1],$$

wobei $\hat{S}(t)$ der Kaplan-Meier-Schätzer der Survivalfunktion ist, der auf der kombinierten Stichprobe berührt.

- Plotte die KM-Schätzungen der zwei Survivalkurven. Was könnte man raten, gibt es signifikative Unterschiede zwischen den zwei Kurven?
Hinweis: `survfit(Surv(...,...) ~ rx)` trennt die Daten nach der Variable rx.
- Fülle die Tabelle (*) auf.
- Bestimme τ und die Statistik $T(\tau)$ sowohl für den Log-Rank Test als auch für den Harrington-Fleming Test mit $\rho = 1$.
- Wird H_0 angenommen oder abgelehnt nach jedem Test?(Signifikanzniveau $\alpha = 5\%$).
- Die beide Tests sind in R als χ^2 -Tests implementiert. Das wichtige R Befehl ist `survdif()`. Man kann ρ bestimmen: $\rho = 0$ (default) ergibt den log-rank Test, $\rho = 1$ einen Harrington-Fleming Test (wie oben) mit $\rho = 1$. Führe beide Tests auch mit *R* durch und vergleiche die Ergebnisse.

(2+3+3+2+2 Punkte)

Tabelle (*)

t_j	$Y_j^{(1)}$	$D_j^{(1)}$	$Y_j^{(2)}$	$D_j^{(2)}$	$\frac{Y_j^{(1)}(t)Y_j^{(2)}(t)}{Y_j^{(1)}(t)+Y_j^{(2)}(t)}$	$\widehat{S}(t_{j-1})$	$\widehat{S}(t_{j-1}) \cdot \frac{Y_j^{(1)}(t)Y_j^{(2)}(t)}{Y_j^{(1)}(t)+Y_j^{(2)}(t)}$
59	13	1	13	0	6.500	1	6.500
115	12	1	13	0	6.240	0.962	6.000
156	11	1	13				
268							
329							
353							
365							
431							
464							
475							
563							
638							