
Zeit-bis-zum-Ereignis Daten II

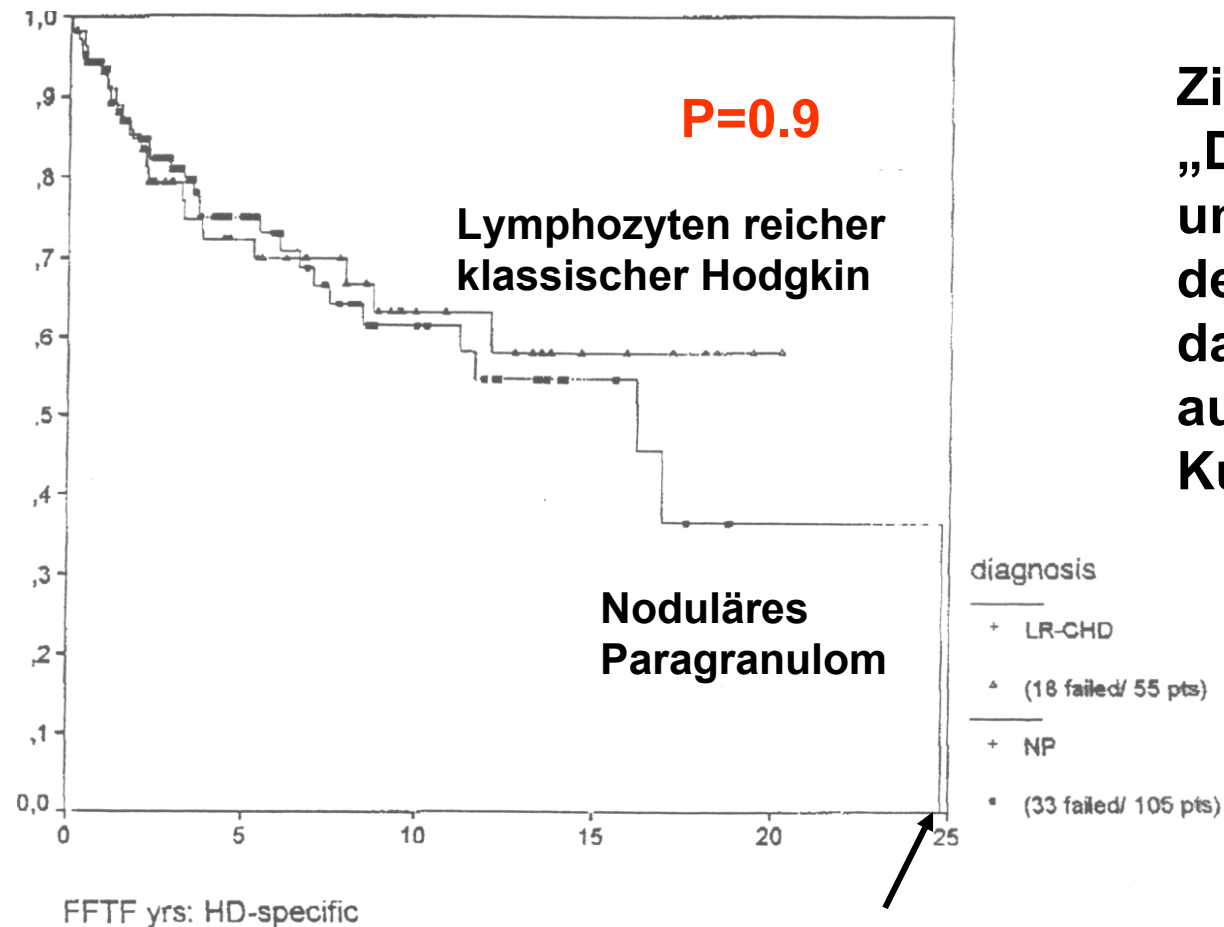
**(δ -Methode und Greenwood Formel für die
Variant des Kaplan Meier Schätzers)**

Dr. Dirk Hasenclever

IMISE, Universität Leipzig

Dirk.Hasenclever@IMISE.uni-Leipzig.de

Vorsicht: Kaplan Meier Kurven werden zum Ende hin instabil !

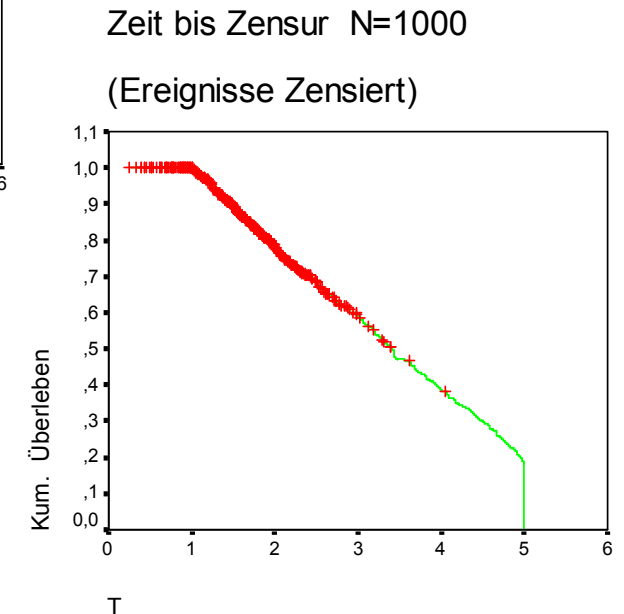
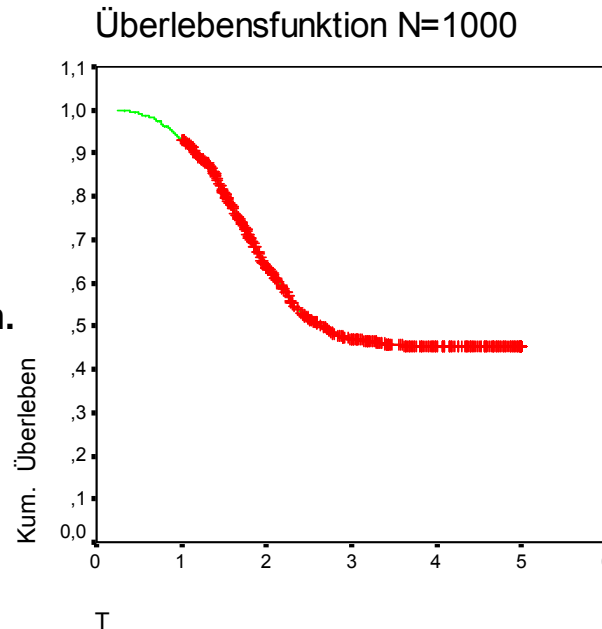


Zitat von int. Kongress:
„Die Subtypen unterscheiden sich deutlich in der Prognose, da LRCHD ein Plateau ausbildet und bei NP die Kurve auf Null fällt.“ (NN)

Patient mit längstem Follow-up hat Ereignis
-> Kurve fällt auf 0

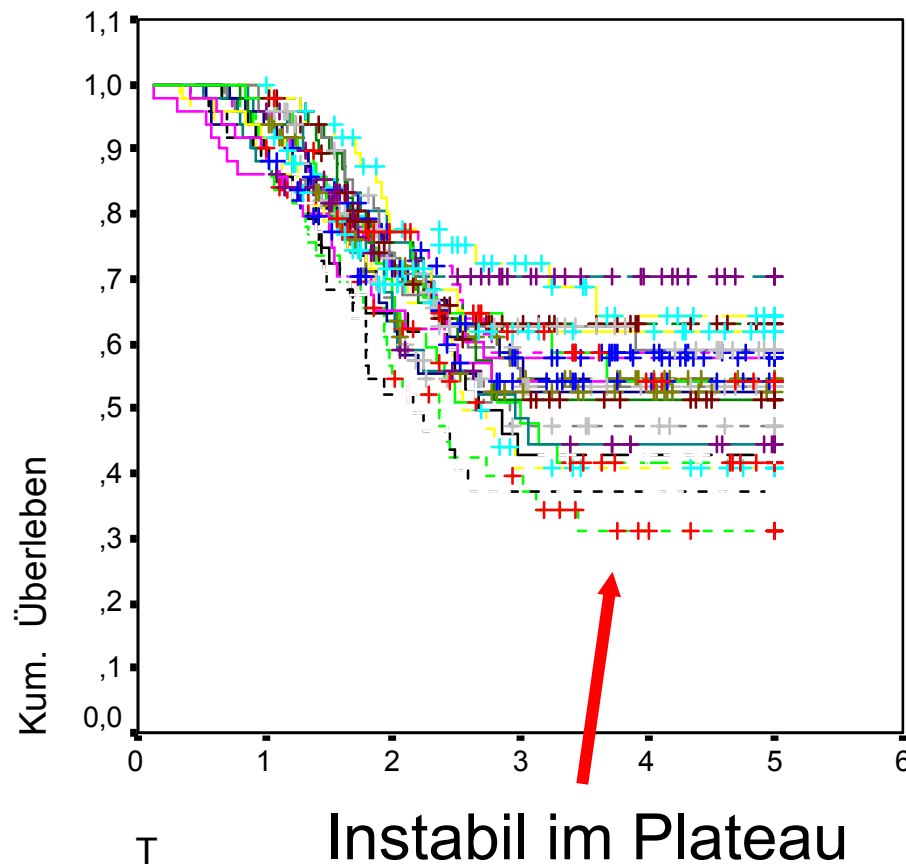
Simulation von Überlebenskurven mit Zensur

```
COMMENT Generiere Überlebenszeiten.  
COMPUTE ereignis = 1.  
COMPUTE SV=5.  
COMPUTE p=uniform(1).  
IF (p > 0.5) SV=rv.weibull(2,3).  
IF (p <= 0.5) ereignis=0.  
  
COMMENT Generiere Zensurzeiten.  
COMPUTE cens=1+uniform(5).  
IF (gruppe=0 ) cens=max(0.2,0.2*cens+0.5*SV).  
  
COMMENT Zensurierung.  
COMPUTE T=sv.  
IF(cens < SV) ereignis=0.  
IF(cens < SV) T=cens.
```

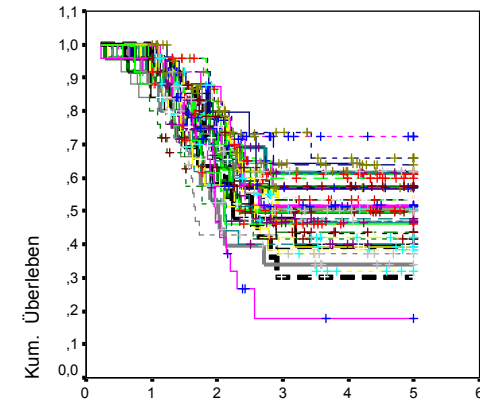


Simulierte Kurven

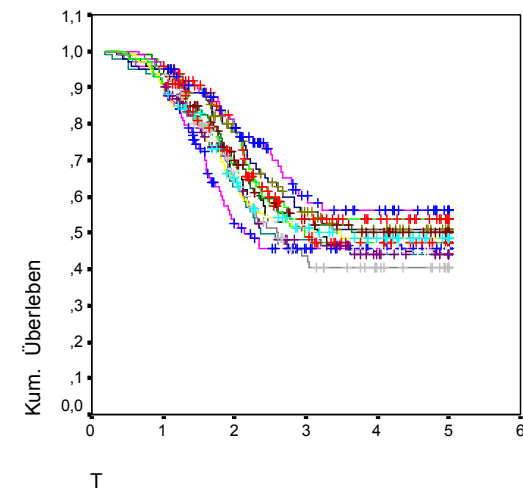
Überlebensfunktionen N=50



Überlebensfunktionen N=25



Überlebensfunktionen N=100



Greenwood-Formel für Standardfehler

Der Kaplan Meier Schätzer liefert asymptotisch konsistent Schätzwerte $\hat{S}(t)$ für die zugrundeliegende "wahre" Überlebenswahrscheinlichkeit $S(t)$.

$\hat{S}(t)$ ist asymptotisch normalverteilt um $S(t)$ mit

Standardfehler;

$$se(\hat{S}(t)) = \hat{S}(t) \cdot \sqrt{\sum_{t_j < t} \frac{d_j}{n_j(n_j - d_j)}}$$

95% Konfidenzintervall : $\hat{S}(t) \pm 1.96 \cdot se(\hat{S}(t))$

Wenn S ein Plateau entwickelt, so wird $\hat{S}(t)$ mit fallendem n_j zunehmend ungenauer.

δ -Methode I

Gegeben Schätzer $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k$, asymptotisch normal verteilt, mit $E(\hat{p}_i) = p_i$ und bekannter Kovarianzmatrix $\text{cov}(\hat{p}_i, \hat{p}_j)$.

Sei g eine differenzierbare Funktion von k Variablen.

Problem : Bestimme Verteilung, Erwartungswert und Varianz von $g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$

SATZ :

a) $g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$ ist asymptotisch normal verteilt

b) $E(g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)) \approx g(p_1, p_2, \dots, p_k)$ [n groß]

c) $\text{var}(g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)) \approx \sum_{i=1}^k \sum_{j=1}^k \frac{\partial g}{\partial p_i} \frac{\partial g}{\partial p_j} \text{cov}(\hat{p}_i, \hat{p}_j)$

δ -Methode II

Beweis : Taylor - Entwicklung!

$$g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k) \doteq g(p_1, p_2, \dots, p_k) + \sum_{i=1}^k \frac{\partial g}{\partial p_i} (\hat{p}_i - p_i) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \frac{\partial^2 g}{\partial p_i \partial p_j} (\hat{p}_i - p_i)(\hat{p}_j - p_j) + \dots$$

$$E(g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)) \doteq g(p_1, p_2, \dots, p_k) + \sum_{i=1}^k \frac{\partial g}{\partial p_i} E((\hat{p}_i - p_i)) + \sum_{i=1}^k \sum_{j=1}^k \frac{\partial^2 g}{\partial p_i \partial p_j} E((\hat{p}_i - p_i)(\hat{p}_j - p_j))$$

$= 0$ *geht mit $n \rightarrow \infty$ gegen 0*

$$\text{var}(g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)) = E((g(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k) - g(p_1, p_2, \dots, p_k))^2) \approx E\left(\left(\sum_{i=1}^k \frac{\partial g}{\partial p_i} (\hat{p}_i - p_i)\right)^2\right) =$$

$$\sum_{i=1}^k \sum_{j=1}^k \frac{\partial g}{\partial p_i} \frac{\partial g}{\partial p_j} E((\hat{p}_i - p_i)(\hat{p}_j - p_j)) = \sum_{i=1}^k \sum_{j=1}^k \frac{\partial g}{\partial p_i} \frac{\partial g}{\partial p_j} \text{cov}(\hat{p}_i, \hat{p}_j)$$

Einfache Anwendungen

Funktion g	Approximierte Varianz
$X \cdot Y$	$E(Y)^2 \text{var}(X) + E(X)^2 \text{var}(Y) + 2E(X)E(Y)\text{cov}(X, Y)$
$\frac{X}{Y}$	$\left(\frac{E(X)}{E(Y)} \right)^2 \left(\frac{\text{var}(X)}{E(X)^2} + \frac{\text{var}(Y)}{E(Y)^2} + \frac{\text{cov}(X, Y)}{E(X)E(Y)} \right)$
$\sqrt{X}$	$\frac{\text{var}(X)}{4E(X)}$
X^2	$4E(X)^2 \text{var}(X)$
$\ln(X)$	$\frac{\text{var}(X)}{E(X)^2}$

Varianz des Kaplan-Meier Schätzers (Greenwood-Formel)

Wende δ - Methode auf Kaplan - Meier Schätzer an :

$$\hat{S}(t) = \hat{p}_1 \cdot \hat{p}_2 \dots \cdot \hat{p}_k$$

Die Schätzer der bedingten Wahrscheinlichkeiten $\hat{p}_j = \frac{n_j - d_j}{n_j}$

sind unabhängig mit $\text{var}(\hat{p}_j) = \frac{\hat{p}_j(1 - \hat{p}_j)}{n_j}$.

$$\frac{\partial \hat{S}}{\partial \hat{p}_j} = \frac{\hat{p}_1 \cdot \hat{p}_2 \dots \cdot \hat{p}_k}{\hat{p}_j} = \frac{\hat{S}(t)}{\hat{p}_j}$$

$$\begin{aligned} \text{var}(\hat{S}(t)) &\doteq \sum_{i=1}^k \sum_{j=1}^k \frac{\hat{S}(t)^2}{\hat{p}_i \hat{p}_j} \cdot \text{cov}(\hat{p}_i, \hat{p}_j) = \hat{S}(t)^2 \cdot \sum_{i=1}^k \frac{\text{var}(\hat{p}_i)}{\hat{p}_i^2} = \hat{S}(t)^2 \cdot \sum_{i=1}^k \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i \hat{p}_i^2} \\ &= \hat{S}(t)^2 \cdot \sum_{i=1}^k \frac{d_i}{n_i(n_i - d_i)} \end{aligned}$$

δ -Methode



Lernziel

- δ -Methode erlaubt approximativ die Varianz einer Funktion von approximativ normal-verteilten Zufallsgrößen zu bestimmen.
- Die Greenwood-Formel erklärt, warum das Plateau einer Kaplan Meier Schätzkurve zunehmend instabil wird.