Ereignisanalyse

Sterbetafel-Methode
Kaplan-Meier-Verfahren
Cox-Regression

Kurt Holm

Almo Statistik-System
www.almo-statistik.de
holm@almo-statistik.de
kurt.holm@jku.at

2013

Im Text wird häufig auf das Dokument **P0** Bezug genommen. Dabei handelt es sich um das Almo-Dokument "Arbeiten mit Almo.PDF" (Dokument 0).

Weitere Almo-Dokumente

Die folgenden Dokumente können alle kostenlos von der Handbuchseite in

www.almo-statistik.de

heruntergeladen werden

```
0. Arbeiten_mit_Almo.PDF (1 MB)
1a. Eindimensionale Tabellierung.PDF  (1.8 MB)
1b. Zwei- und drei-dimensionale Tabellierung.PDF (1.1 MB)
2.  Beliebig-dimensionale Tabellierung.PDF (1.7 MB)
3.  Nicht-parametrische Verfahren.PDF (0.9 MB)
4.  Kanonische Analysen.PDF (1.8 MB)
    Diskriminanzanalyse.PDF (1.8 MB)
    enthält: Kanonische Korrelation, Diskriminanzanalyse, bivariate
    Korrespondenzanalyse, optimale Skalierung
5.  Korrelation.PDF (1.4 MB)
6.  Allgemeine multiple Korrespondenzanalyse.PDF (1.5 MB)
7.  Allgemeines ordinales Rasch-Modell.PDF (0.6 MB)
7a. Wie man mit Almo ein Rasch-Modell rechnet.PDF (0.2 MB)
8.  Tests auf Mittelwertsdifferenz,  t-Test.PDF (1,6 MB)
9.  Logitanalyse.pdf  (1,2MB)  enthält Logit- und Probitanalyse
10. Koeffizienten der Logitanalyse.PDF  (0,06 MB)
11. Daten-Fusion.PDF  (1,1 MB)
12. Daten-Imputation.PDF  (1,3 MB)
13. ALM Allgemeines Lineares Modell.PDF  (2.3 MB)
13a. ALM Allgemeines Lineares Modell II.PDF  (2.7 MB)
14. Ereignisanalyse: Sterbetafel-Methode, Kaplan-Meier-Schätzer,
    Cox-Regression.PDF (1,5 MB)
15. Faktorenanalyse.PDF (1,6 MB)
16. Konfirmatorische Faktorenanalyse.PDF (0,3 MB)
17. Clusteranalyse.PDF (3 MB)
18. Pisa 2012 Almo-Daten und Analyse-Programme.PDF (17 KB)
19. Guttman- und Mokken-Skalierung.PFD (0.8 MB)
20. Latent Structure Analysis.PDF (1 MB)
21. Statistische Algorithmen in C (80 KB)
22. Conjoint-Analyse (PDF 0,8 MB)
23. Ausreisser entdecken (PDF 170 KB)
24. Statistische Datenanalyse Teil I, Data Mining I
25. Statistische Datenanalyse Teil II, Data Mining II
26. Statistische Datenanalyse Teil III, Arbeiten mit Almo-
    Datenanalyse-System
27. Mehrfachantworten, Tabellierung von Fragen mit Mehrfach-
    antworten (0.8 MB)
28. Metrische multidimensionale Skalierung (MDS)  (0,4 MB)
29. Metrisches multidimensionales Unfolding (MDU)  (0,6 MB)
30. Nicht-metrische multidimensionale Skalierung (MDS)  (0,4 MB)
31. Pfadanalyse.PDF  (0,7 MB)
32. Datei-Operationen mit Almo  (1,1 MB)
33. Wählerstromanalyse und Wahlhochrechnung (1,6 MB)
34. Soziometrie. Auswerung soziometrischer Daten (0,5 MB)
```

INHALTSVERZEICHNIS

P39 Sterbetafel-Methode

Almo bietet folgende Verfahren zur Ereignisanalyse an:

1. Sterbetafel-Methode
2. Kaplan-Meier-Schätzer der Überlebenswahrscheinlichkeit
3. Cox-Regression

Der Begriff "Ereignisanalyse" ist nicht der einzige, unter dem diese Gruppe von Verfahren eingeordnet wird. In der englisch-sprachigen Literatur werden auch die Begriffe "survival analysis" und "lifetime analysis" verwendet.

Auch die obigen 3 Verfahren werden in der Literatur gelegentlich unter anderem Namen geführt: Das Kaplan-Meier-Verfahren wird auch "product-limit estimator" genannt, die Cox-Regression auch "proportional hazard model".

Wir wollen in diesem Kapitel P39 die Sterbetafel-Methode behandeln.

Wie lange überlebt ein Mensch eine Herztransplantation. Wie groß ist die Wahrscheinlichkeit nach einer Herztransplantation nach x Tage (z.B. nach 365 Tagen) noch zu leben?

Die Sterbetafel-Methode beantwortet diese Fragen. Sie ist aber auch auf Themen anwendbar, bei denen es nicht um Leben oder Tod geht, z.B. auf folgende sozialwissenschaftliche Fragestellungen: Wie lange verbleiben Menschen in ihrem 1. Beruf (siehe dazu Blossfeld, Hamerle, Mayer, 1986). Dem "Tod" entspricht hier der Berufswechsel, dem "Überleben" das Verbleiben im 1. Beruf.

Oder: Wann begehen Jugendliche aus einem Slum-Wohngebiet ihr erstes kriminelles Delikt. Dieses entspricht dem "Tod". Das "nicht kriminell werden" entspricht dem "Überleben".

Im technisch-naturwissenschaftlichen Bereich gibt es eine Fülle von Anwendungsmöglichkeiten der Ereignisanalyse; z.B: Wie lang ist die "Überlebensdauer" von elektrischen Glühbirnen?

Wir wollen im folgenden das Beispiel der Überlebensdauer von Herztransplantationen ausführlich betrachten. Die Daten sind folgende:

Untersuchungsperson	Überlebensdauer in Tagen
.	.
.	.
.	.
12	39
13	730
14	136
15	1775*
.	.
.	.

Siehe dazu die Datei ".\Almo\Testdat\survive1.fre" und das Almo-Maskenprogramm Prog39ma. Man findet es durch Klick auf den Knopf "Verfahren" am Oberrand des Almofensters, dann "Ereignisanalyse" oder durch Klick auf "alle Progs" ebenfalls am Oberrand des Almofensters.

Wir betrachten nur einen kleinen Ausschnitt aus der Datenmatrix. Person 12 ist nach 39 Tagen gestorben, Person 13 erst nach 730 Tagen etc. Person 15 ist nach 1775 Tagen aus der Beobachtung ausgeschieden. Wir kennzeichnen dies durch einen Stern. In der Sprache der Ereignisanalyse wird gesagt: Person 15 wurde nach 1775 Tagen "zensiert". Wir wissen nicht, ob Person 15 heute noch lebt. Wir wissen nur, daß Person 15 die Herztransplantation mindestens 1775 Tage überlebt hat - aber möglicherweise noch sehr viel mehr Tage überlebt hat.

Der Begriff "zensiert" meint also:

> das fragliche Ereignis, also in unserem Beispiel der Tod, ist nicht nachweisbar, weil die Untersuchungsperson aus der Beobachtung ausgeschieden ist.

kurz: "zensiert" heißt "ausgeschieden".

P39.1 Die Daten

Die Daten, die Almo benötigt um die Sterbetafel-Methode zu rechnen sind folgende: (siehe dazu die Datei ...testdat\survive1.fre. Die Daten sind Crowley/Hu entnommen).

```
    V1         V2
Zeitdauer  Ereignis
    15         1
    3          1
   624         0
    .          .
    .          .
    .          .
    39         0
   730         0
   136         0
   1775        1
    .          .
    .          .
    .          .
    13         1
    1          1
```

Beachte:

1. Selbstverständlich kann die Datenmatrix mehr als nur 2 Variable umfassen. Die Zeitvariable muß auch nicht unbedingt V1 und die Ereignisvariable muß nicht unbedingt V2 sein.

2. Die Ereignisvariable ist in folgender Weise kodiert:

0 = Ereignis eingetreten
1 = zensiert

Diese Codes sind im Prinzip beliebig wählbar. Jedoch muß der Code für "zensiert" höher sein als der für "Ereignis eingetreten". Almo bringt im Falle, daß 2 oder mehr Gruppen verglichen werden (wo diese Bedingung unbedingt erfüllt sein muß) eine Fehlermeldung.

In obigen Daten ist die 1. Person nach 15 Tagen ausgeschieden, die 2. Person nach 3 Tagen. Die 3. Person ist nach 624 Tagen gestorben. "Ausgeschieden" bedeutet bei der 1. und 2. Person wohl, daß diese an Komplikationen gestorben sind, die mit der eigentlichen Herztransplantation nichts zu tun haben. Sie wurden vermutlich deswegen als "zensiert" betrachtet und damit aus der Analyse herausgenommen.

3. Die Variablenwerte können als Ganzzahlwerte geschrieben sein aber auch als Dezimalwerte mit Dezimalpunkt und beliebig vielen Dezimalstellen.

4. Liegt der Zeitdauer-Wert genau auf der oberen Intervallgrenze, dann wird die Untersuchungseinheit in das nächst höhere Intervall eingeordnet.

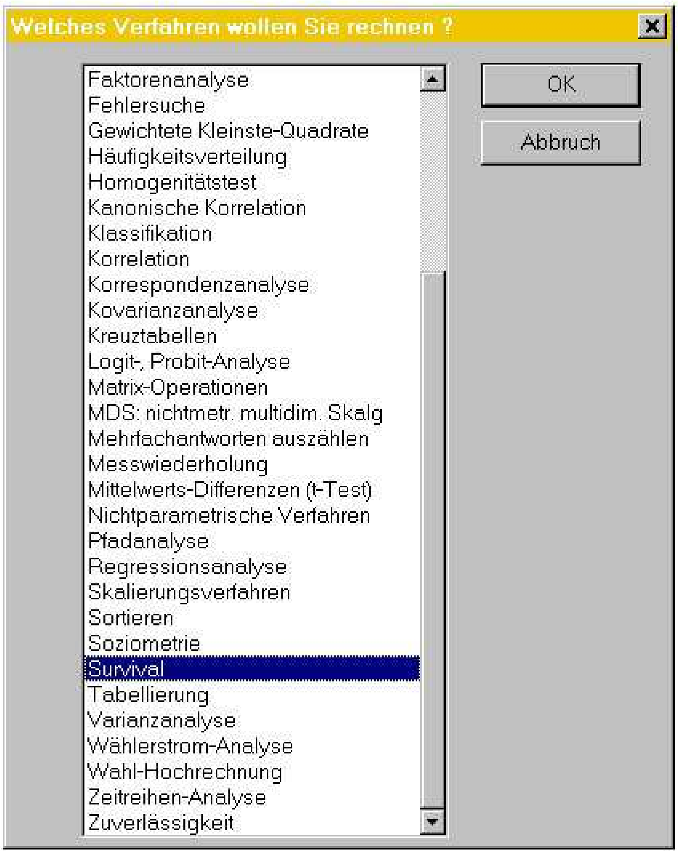
P39.2 Almo-Eingabe

P39.2.1 Die Programm-Maske **Prog 39ma**

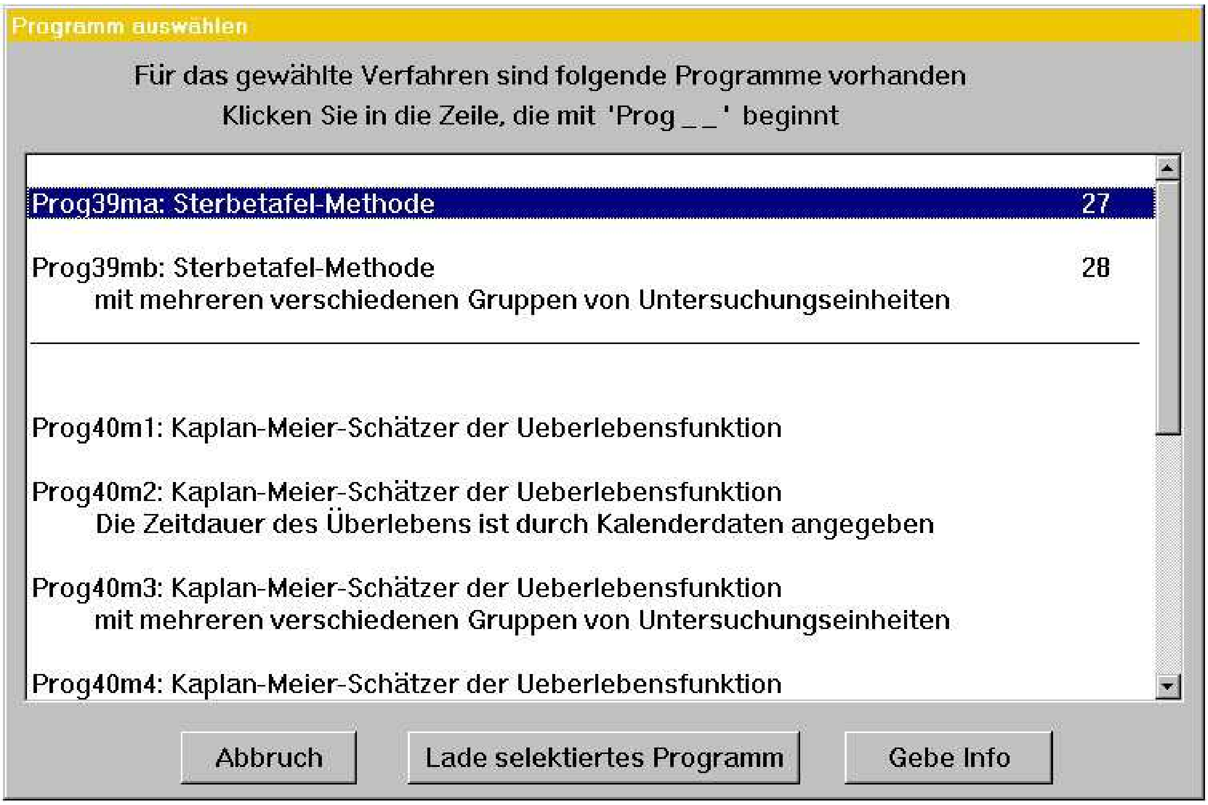
Kurzanleitung:

a. Starten Sie Almo. Es erscheint das Almo-Fenster uns eine Dialogbox. In ihr klicken Sie auf "Weiter".

b. Unterhalb der Menüleiste befindet sich eine Knopfleiste. In ih klicken Sie auf ddn Knopf "Verfahren". Almo präsentiert Ihnen dann eine Übersicht über die vorhandenen Verfahren.



Klicken Sie auf "Survival". Dann erscheint eine Übersicht über die Maskenprogramme zum Thema "Survival".



c. Klicken Sie nun auf Prog39ma: Sterbetafel-Methode' und dann am unteren Rand 'Lade Maskenprogramm' (oder gleich Doppelklick auf Prog39ma). Es erscheint eine Eingabemaske, die auf den folgenden Seiten abgebildet ist.

d. Nachdem Sie diese Eingabemaske entsprechend Ihren Analyseabsichten ausgefüllt haben, klicken Sie in der Knopfleiste auf den Knopf "Rechne".

e. Almo präsentiert Ihnen eine Dialogbox, in der Ihnen für das Abspeichern des Almo-Maskenprogramms ein Name vorgeschlagen wird. Klicken Sie OK.

f. Dann wird Ihnen eine weitere Dialogbox präsentiert, in der Ihnen für das Abspeichern des Ergebnisses ein Dateiname vorgeschlagen wird. Klicken Sie OK.

g. Almo bringt jetzt ein neues Fenster (mit dem Titel "Almo rechnet") auf den Bildschirm, in dem Sie die Rechenfortschritte Ihres Programms verfolgen können.

h. Wenn als Status 'fertig' gemeldet wird, dann klicken Sie in diesem Fenster auf den Knopf LADE ERGEBNIS.

i. Almo lädt das Ergebnis in ein Fenster des Editor, wo Sie es sich anschauen können.

j. Wenn Sie durch die "Ergebnis-Datei" scrollen oder blättern, dann werden Sie einen großen Knopf "Grafik" finden. Klicken Sie auf diesen.

k. Almo öffnet seinen Grafik-Editor und stellt die Tabelle aus der Ergebnis-Datei als 3D-Balkendiagramm dar.

l. Mit Klick auf den Knopf "Zurück" gelangen Sie wieder in die Ergebnis-Datei.

Die Programm-Maske Prog39ma zur Sterbetafel-Methode

Prog39ma.Msk Kurzprogramm
Sterbetafel-Methode

Sie können auch mit dem äquivalenten Programm Prog39mb rechnen
Oder doppelklicken Sie gleich auf: [".\Almo_Msk\Prog39mb.Msk"]
Sie müssen dort nur die Gruppierungsvariable herausnehmen

Beispiel:
Die Überlebensdauer von herztransplantierten Menschen wird berechnet.

Berechnet wird die Sterbetafel etwa für folgende Daten

Zeitdauer des Überlebens	Ereignis 0=tot 1=zensiert (=ausgeschieden)
15	0
3	0
624	1
46	0
.	.
.	.

Die 1. Person ist nach 15 Tagen gestorben, die 2. schon nach 3 Tagen. Die 3. Person ist nach 624 Tagen aus der Beobachtung ausgeschieden (man nennt das "zensiert")

Almo errechnet u.a. folgende Koeffizienten

1 Intervalldauer von bis Tag t1 Tag t2	2 Personen, die während dieses Intervalls starben	3 ausschied.	4 Wahrschlkt zu über-leben	5 Überlebensfunktion =kumulative Whrscheinlichkeit zu überleben
0 - 162	14	19	0.7477	0.7477
162 - 324	4	4	0.8667	0.6480
324 - 486	4	0	0.8333	0.5400
486 - 648	4	1	0.7949	0.4293
. .	.	.	.	.
. .	.	.	.	.

Median der Ueberlebensdauer = 544.5542 Tage
Standardfehler des Medians = 98.1493 Tage

Betrachten wir z.B. den 3. Wert in Spalte 4. Er sagt folgendes aus: Wenn ein Patient schon 2 Intervalle, also bis zum 324. Tag überlebt hat, dann ist seine Chance, das 3. Intervall, also bis zum 486. Tag zu überleben =0.8333 (=83%)

Der 3. Wert aus der Überlebensfunktion in Spalte 5 ist so zu interpretieren: Zum Zeitpunkt der Transplantation (Tag 0) ist die Wahrscheinlichkeit bis zum Ende des 3. Intervalls, also bis zum 486. Tag zu überleben =0.5400 (=54%)

Grafik: Liniendiagramm der Überlebenswahrscheinlichkeit
siehe Handbuch, Abschnitt P39

Was ist ein Kurzprogramm ? --> Hilfe
Bedienung --> Hilfe

Speicher fuer x Variable Hilfe

Vereinbare Variable= 20 ;

2 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Datei der Variablennamen | Hilfe

zeige = Namensdatei in Output zeigen
leer = nicht

4 Freie Namensfelder | Hilfe

Name 1=Ueberlebensdauer;
Name 2=Ereignis:tot,zensiert;

Gruppierungsvariable

erzeuge zusätzliche Namensfelder

5 Datei aus der gelesen wird | Hilfe bei Datei-Problemen

"C:\Almo7\Testdat\Survive1.fre"

frei — Format der Daten | Hilfe

V1:3 — der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

6 Wenn Dateiformat FIX oder Nicht-Standard-FREI | Hilfe

7 die Ereignis-Variable

Ereignis — die Ereignisvariable

0 — Code für "Ereignis hat stattgefunden"
z.B. Untersuchungseinheit ist gestorben

1 — Code für "zensiert"
z.B. Untersuchungseinheit hat ueberlebt oder ist ausgeschieden

die Zeit-Variable

Ueberlebensdauer — die Zeitvariable

---

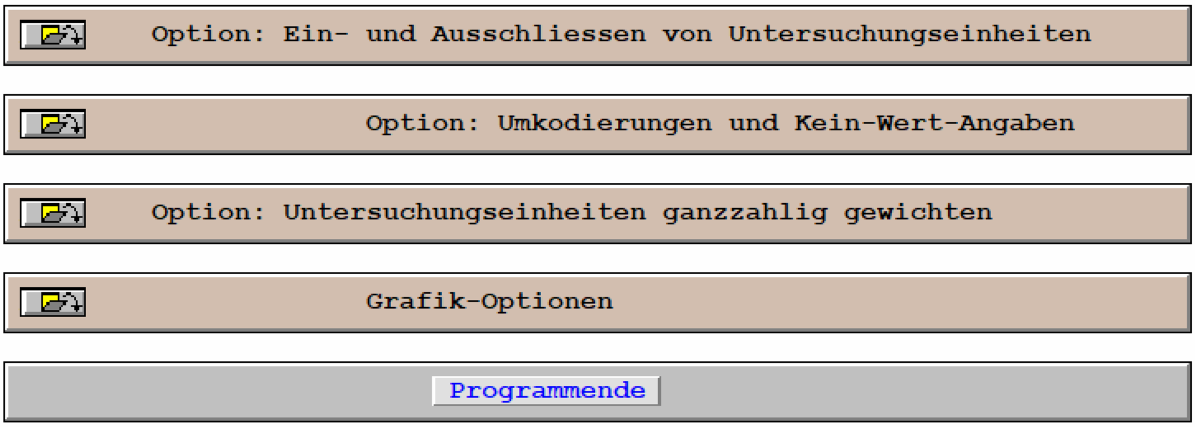
0 — Zeitanfang
1782 — Zeitende

Sie brauchen das Zeitende nicht anzugeben. Klicken Sie auf den Raus-Rein-Pfeil und leeren Sie damit das Eingabefeld. Almo ermittelt dann aus den Daten die Werte für das Zeitende

---

11 — Zahl der Zeitintervalle
Dauer des Zeitintervalle

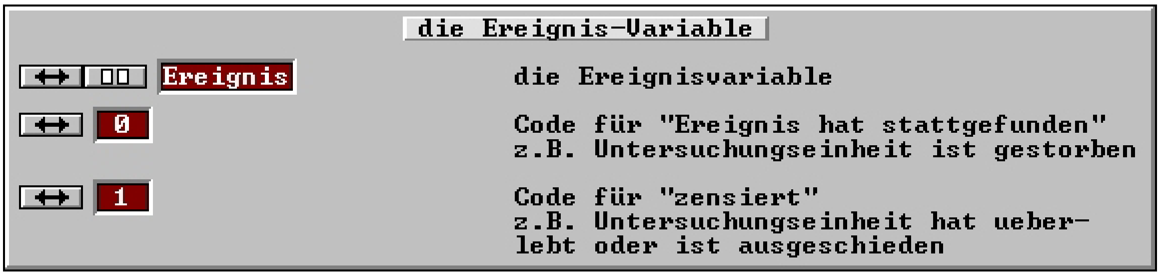
Sie können entweder die Zahl der Zeitintervalle festlegen, in die Almo die Zeit von Zeitanfang bis Zeitende untergliedern soll oder die Dauer des Zeitintervalls



Erläuterungen zu den Eingabeboxen

Eingabebox 1 bis 6. Siehe Almo-Dokument Nr. 0 "Arbeiten_mit_Almo", Abschnitt P0.1 bis P0.4

Eingabebox 7: Die Ereignis-Variable



Definieren Sie hier die Ereignisvariable. In unserem Beispiel ist dies die Variable 2 mit dem Namen "Ereignis".

Box 8: Die Zeit-Variable

Definieren Sie hier die Zeitvariable. Die Gesamtzeit wird bei der Sterbetafel-Methode in Zeitintervalle untergliedert. Sie haben dabei 2 Möglichkeiten: Entweder Sie geben an wieviele gleich große Intervalle Sie haben wollen oder Sie geben an wie lang ein Zeitintervall sein soll.

Box 9: Ein- und Ausschliessen von Untersuchungseinheiten. Siehe Almo-Dokument Nr. 0 "Arbeiten_mit_Almo", Abschnitt P0.7

Box 10: Umkodierungen und Kein-Wert-Angabe. Siehe Almo-Dokument Nr. 0 "Arbeiten_mit_Almo", Abschnitt P0.5. Siehe auch die Erläurerung dieser Eingabebox im nachfolgenden Abschnitt P39.4

Definieren Sie den Kein-Wert-Fall. Wenn Sie z.B. `-1` in der Ereignisvariablen als "Kein_Wert" definieren, dann werden Personen, die diesen Wert besitzen, nicht ausgewertet. Sie werden nicht als zensiert betrachtet sondern als nicht existent.

```
Ereignis (-1 = KeinWert)
```

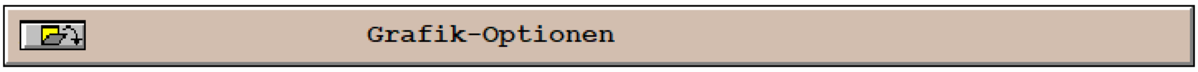
Nehmen Sie in dieser Box auch notwendige Umkodierungen vor. Es könnte z.B. sein, daß das Ereignis "tot" mit den Werten 10,11,12,13 kodiert ist - je nach konkreter Todesursache. In diesem Fall würden Sie umkodieren

```
Ereignis (10,11,12,13=0)
```

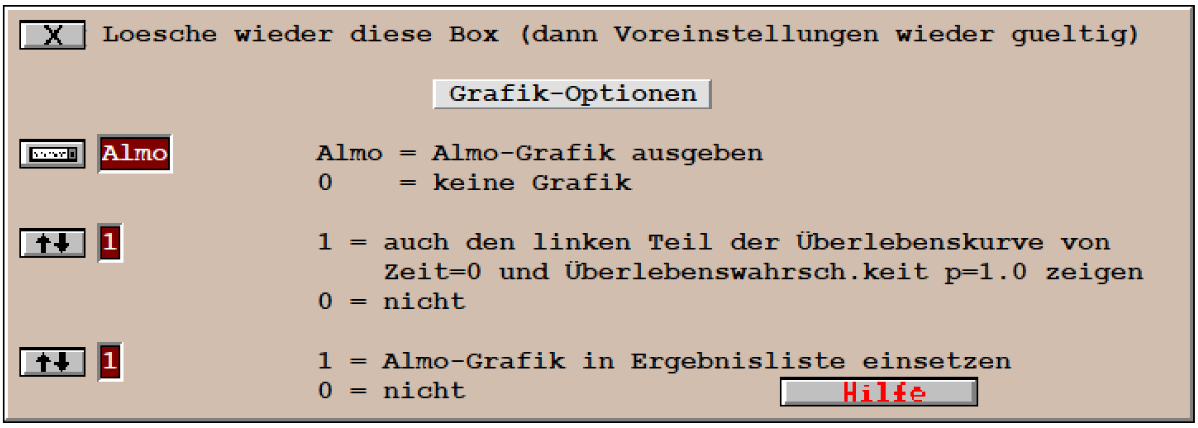
da Sie in Box 7 das Ereignis "tot" als mit 0 kodiert angegeben haben.

Box 11: Untersuchungseinheiten gewichten. Siehe Almo-Dokument Nr. Arbeiten_mit_Almo, Abschnitt P0.8

Box 12: Grafik-Optionen



Wird die Optionsbox geöffnet, dann sieht man folgendes:



Wird das 2. Eingabefeld auf "1" gesetzt, dann zeigt Almo auch den Linken Teil der Überlebenskurve.

P39.3 Ergebnisse

Almo liefert folgendes Ergebnis:

```
Fuer Analyse aus Datenvektor ausgewaehlte Variable

Zeitvariable:          V1   Zeitdauer
Ereignisvariable:      V2   Ereignis
```

Zahl der Intervalle: 11
Intervalldauer: 162.00

Zahl der eingelesenen Datensaetze: 65
Zahl der ausgewerteten Datensaetze: 65
Sterbetafel

1 Inter-vall Nr.	2 Intervalldauer von	bis	3 Pers. die in dieses Intervall eintreten	4 Pers. die waehrend dieses Intv ausscheiden (zensierte Beobachtgn)	5 Pers. die dem Risiko zu termi-nieren ausgesetzt sind	6 Pers. die waehrend dieses Intv. terminierten
1	0.00 -	162.00	65	14	58.0	19
2	162.00 -	324.00	32	4	30.0	4
3	324.00 -	486.00	24	4	22.0	0
4	486.00 -	648.00	20	4	18.0	1
5	648.00 -	810.00	15	1	14.5	1
6	810.00 -	972.00	13	3	11.5	1
7	972.00 -	1134.00	9	1	8.5	2
8	1134.00 -	1296.00	6	1	5.5	0
9	1296.00 -	1458.00	5	1	4.5	1
10	1458.00 -	1620.00	3	2	2.0	0
11	1620.00 -	1782.00	1	1	0.5	0

7 Inter-vall Nr.	8 Intervalldauer von	bis	9 Whrschlkt zu termi-nieren	10 Whrschlkt zu ueber-leben	11 Ueberleb.fkt. =kumulative Whrschlkt zu ueber-leben bei Intvallende	12 Dichte-funktion der Ueber-lebensdauer
1	0.00 -	162.00	0.3276	0.6724	0.6724	0.002022
2	162.00 -	324.00	0.1333	0.8667	0.5828	0.000553
3	324.00 -	486.00	0.0000	1.0000	0.5828	0.000000
4	486.00 -	648.00	0.0556	0.9444	0.5504	0.000200
5	648.00 -	810.00	0.0690	0.9310	0.5124	0.000234
6	810.00 -	972.00	0.0870	0.9130	0.4679	0.000275
7	972.00 -	1134.00	0.2353	0.7647	0.3578	0.000680
8	1134.00 -	1296.00	0.0000	1.0000	0.3578	0.000000
9	1296.00 -	1458.00	0.2222	0.7778	0.2783	0.000491
10	1458.00 -	1620.00	0.0000	1.0000	0.2783	0.000000
11	1620.00 -	1782.00	0.0000	1.0000	0.2783	0.000000

13 Inter-vall Nr.	14 Intervalldauer von	bis	15 Hazardrate Risikorate	16 Std.fehler der Hazardrate	17 Std.fehler der kumulativen Whrschlkt zu ueber-leben	18 Std.fehler der Dichte-funktion der Ueber-lebensdauer
1	0.00 -	162.00	0.002418	0.000544	0.061626	0.000380
2	162.00 -	324.00	0.000882	0.000440	0.067780	0.000263
3	324.00 -	486.00	0.000000	0.000000	0.067780	0.000000
4	486.00 -	648.00	0.000353	0.000353	0.071329	0.000196
5	648.00 -	810.00	0.000441	0.000441	0.075840	0.000228
6	810.00 -	972.00	0.000561	0.000561	0.081288	0.000266
7	972.00 -	1134.00	0.001646	0.001154	0.092183	0.000436
8	1134.00 -	1296.00	0.000000	0.000000	0.092183	0.000000
9	1296.00 -	1458.00	0.001543	0.001531	0.100286	0.000451
10	1458.00 -	1620.00	0.000000	0.000000	0.100286	0.000000
11	1620.00 -	1782.00	0.000000	0.000000	0.100286	0.000000

19 20 21 22

Inter- vall Nr.	Intervalldauer von	bis	Median der verbleibenden Lebensdauer bei Inter- vall-Beginn	Std.fehler des Median der verbleibenden Lebensdauer
1	0.00 -	162.00	855.1754	238.692060
2	162.00 -	324.00	1177.9577	125.071034
3	324.00 -	486.00	1107.2967	126.577968
4	486.00 -	648.00	945.2967	139.937209
5	648.00 -	810.00	-	-
6	810.00 -	972.00	-	-
7	972.00 -	1134.00	-	-
8	1134.00 -	1296.00	-	-
9	1296.00 -	1458.00	-	-
10	1458.00 -	1620.00	-	-
11	1620.00 -	1782.00	-	-

```
Median der Ueberlebensdauer =   855.1754
Standardfehler des Medians  =   238.6921
```

Der größte Teil dieser Ausgabe bedarf wohl keiner Interpretation. Wir wollen hier die wesentlichen Spalten erklären.

Spalte 5: Risikomenge. Das ist die Zahl der Personen, die dem Risiko zu "terminieren" ausgesetzt sind. Sie ergibt sich aus

(1) $n_i = l_i - c_i / 2$

n_i = Risikomenge für das Intervall i (Spalte 5)
l_i = Zahl der Personen, die in das Intervall i eintreten. Das ist Spalte 3
c_i = Zahl der Personen, die im Intervall i ausscheiden (= zensierte Beobachtungen). Das ist Spalte 4.

Bei der Sterbetafelmethode wird als "Stichtag" die Mitte des Zeitintervalls verwendet. Das ist der Grund dafür, daß man in obiger Gleichung 1 nur die Hälfte der Zensierten einsetzt.

Spalte 9 und 10

Die Wahrscheinlichkeit zu terminieren ergibt sich aus

(2) $q_i = \frac{d_i}{n_i}$

Die Wahrscheinlichkeit zu überleben ist dann residual

(3) $p_i^* = 1-q_i$

q_i = Wahrscheinlichkeit im Intervall i zu terminieren
d_i = Zahl der "Toten" im Intervall i. Sie stehen in Spalte 6
n_i = Risikomenge im Intervall i. Siehe (1). Sie steht in Spalte 5.
p_i^* = Wahrscheinlichkeit im Intervall i zu überleben.

Betrachten wir den 5. Wert von p_i^* = 0.9310 in Spalte 10. Er entsteht aus 1.0-1/14.5=0.9310. Dieser Wert sagt aus: Wenn eine Person das 4. Intervall überlebt hat, dann ist ihre Wahrscheinlichkeit auch das 5. Intervall zu überleben = 0.9310.

Spalte 11: Die Überlebensfunktion.

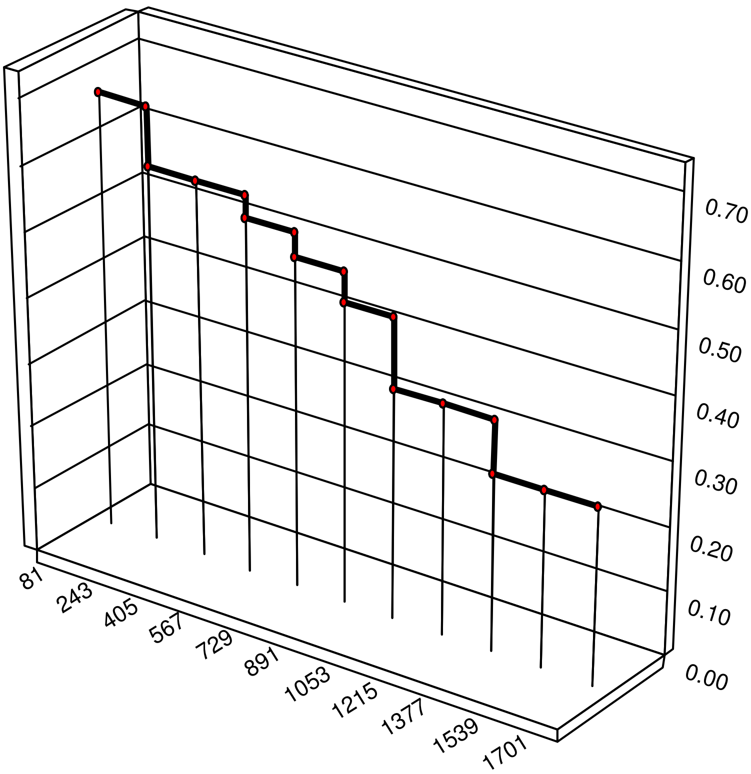
Dies ist die wichtigste Maßzahl aus der Sterbetafel. Sie ist definiert als kumulative Wahrscheinlichkeit zu überleben. In Almo wird dabei (ebenso wie in SPSS) das Intervallende als "Stichtag" verwendet. SAS verwendet den Intervallbeginn (bzw., was dasselbe ist, das Ende des vorhergehenden Intervalls).

(4) $p_i = \prod_{j=1}^{i} p_j^*$

Beispiel: Die kumulative Wahrscheinlichkeit des 3. Intervall (in unserem obigen Ein- und Ausgabe-Beispiel) zu überleben ist

$p_3 =$ 0.6724 * 0.8667 * 1.0 =0.5828

Der Wert von 0.5828 ist inhaltlich nicht sehr anschaulich erklärt, wenn wir sagen " dies ist die kumulative Wahrscheinlichkeit". Wir können ungefähr so interpretieren: Aus der Sicht des Zeitpunkts 0 ist dies die Wahrscheinlichkeit bis zum Ende des 3. Intervalls, also bis zum 486. Tag zu überleben - wobei die zensierten Fälle in einer spezifischen Art und Weise berücksichtigt werden. Gibt es keine zensierten Fälle, dann gilt diese Aussage nicht nur ungefähr sondern exakt.



Die Überlebensfunktion ist eine Treppenfunktion. In obiger Graphik fehlt der vordere und hintere Teil.

Wenn Sie das vordere Kurvenende doch sehen wollen, dann fügen Sie in den Programmparameter ein:

Option 2=1;

Spalte 12: Dichtefunktion der Überlebensdauer. Das ist die auf eine Zeiteinheit umgerechnete Überlebenswahrscheinlichkeit. Siehe Gross/Clark, 1975, S. 35

(5) $f_i = \frac{p_i - p_{i-1}}{v}$

f_i = Dichtefunktion für Intervall i

p_i, p_{i-1} = Kumulative Überlebenswahrscheinlichkeit für das Intervall i und das vorhergehende Intervall i-1. Für i=1 wird p_{i-1} also $p_0 = 1$ gesetzt.

v = Dauer des Intervalls. In Almo wird unterstellt, daß die Beobachtungsperiode in gleich große Intervalle unterteilt ist, v also für alle Intervalle gleich groß ist.

Spalte 15: Hazardrate oder Risikorate. Dies ist das Risiko einer Person (ausgedrückt als Wahrscheinlichkeit), die in das Intervall eintritt, im Verlauf der Zeit im Intervall zu terminieren. Dabei wird diese Wahrscheinlichkeit bezogen auf die Zeiteinheit. Siehe Gross/Clark, 1975, S. 35.

$$(6)\quad h = \frac{2q_i}{v\left(1 + p_i^*\right)}$$

h = Hazardrate, Risikorate

q_i = Wahrscheinlichkeit im Intervall zu terminieren. Siehe Gleichung 2

p_i^* = Wahrscheinlichkeit im Intervall i zu überleben. Siehe Gleichung 3

v = Intervalldauer.

Spalte 16: Standardfehler der Hazardrate

$$(7)\quad SE(h_i) = \sqrt{\frac{h_i^2}{n_i \cdot q_i}\left(1 - \left(\frac{h_i v}{2}\right)^2\right)}$$

Definition der Größen in den vorherigen Gleichungen.

Spalte 17: Standardfehler der Überlebensfunktion

$$(8)\quad SE(p_i) = p_i \sqrt{\sum_{j=1}^{i} \frac{d_j}{n_j\left(n_j - d_j\right)}}$$

siehe Gross/Clark, 1975, S. 41. Dort läuft fälschlicherweise j bis i-1.

Definition der Größen in den vorherigen Gleichungen.

Spalte 18: Standardfehler der Dichtefunktion

$$(9)\quad SE(fi) = fi \sqrt{\sum_{j=1}^{i-1} \frac{q_j}{n_j \cdot p_j} + \frac{p_i}{n_i \cdot q_i}}$$

Für das Intervall:

$$(9a)\quad SE(f_1) = f_i \cdot \sqrt{\frac{p_1}{n_1 \cdot q_1}}\ )$$

Definition der Größen in den vorherigen Gleichungen.

Median der Überlebenszeit: Das ist die mittlere Überlebenszeit (gerechnet als Median), d.h. also die Zahl der Tage, die die Personen im Durchschnitt erwarten dürfen.
Der Median liegt auf jenem Zeitpunkt, bei dem die Überlebensfunktion =0.5 ist. Dabei wird dieser Zeitpunkt linear interpoliert. In unserem Beispiel leben nach 544.5542 Tagen noch 50 % der Personen.

(10) $M = t_j + \frac{p_{j-1} - 0.5}{f_j}$

(11) Standardfehler des Median

$$SE(M) = \sqrt{\frac{1}{4 \cdot n_1 \cdot f_j^2}}$$

dabei ist j jenes Intervall für das gilt

(12) $p_j > .0$ und $p_j < 0.5$ und $p_{j-1} \geq 0.5$

$t_i =$ Zeitpunkt des Intervallbeginns j
j ist das Intervall das obige Bedingung 12 erfüllt.
p_j, p_{j-1} = Wert aus der Überlebensfunktion für das Intervall j bzw. j-1.

n_1 = Risikomenge im 1. Intervall
f_j = Dichtefunktion im Intervall j.

Spalte 21, 22: Median der *verbleibenden* Lebenszeit zu Beginn des Intervalls.

Beispiel: Wie lange ist noch die mittlere Überlebensdauer (gerechnet als Median) einer Person, die das 8. Intervall erreicht hat. In der Ausgabe zu unserem Beispiel erkennen wir, daß diese Person noch 379.5429 Tage leben wird. Genauer ist es zu fragen: Wieviele Tage noch leben 50 % der Personen, die das 8. Intervall erreicht haben. Und die Antwort ist dann, daß dies noch 379.5429 Tage sind. Siehe Gross/Clark, 1975, S. 42.

(13) $\text{Median}_i = t_{j-1} - t_{i-1} + \left(p_{j-1} - \frac{p_{i-1}}{2} \right) / f_j$

i = Intervall-Nr.
t_{i-1} = Zeitpunkt zu Intervallbeginn bzw. (was dasselbe ist) zu Endes des vorherigen Intervalls.
Beachte: Für das Intervall i=1 ist in der Regel $t_{i-1} = 0$

p_{i-1} = Wert aus der Überlebensfunktion aus dem Intervall i-1

j = das ist jenes Intervall, für das gilt:

(14) $p_{j-1} > \frac{p_{i-1}}{2} > p_j$ und $p_j > .0$

(14) $SE_{Median_i} = \sqrt{\frac{p^2_{i-1}}{n_i \cdot f_j^2}}$

n_i = Risikomenge im Intervall i
f_j = Wert der Dichtefunktion im Intervall j.

Beachte: Der Median und SE(Median) ergibt sich auch aus obigen Gleichungen (13) und (14) für i=1.

Der Median der verbleibenden Lebenszeit beim Eintritt in das 1. Intervall, also zu Beobachtungsbeginn ist selbstverständlich gleich dem allgemeinen Median aus Gleichung 10.

In unseren Beispieldaten ist interessanterweise festzustellen, daß wenn es einer Person gelingt, das 1. Intervall zu überleben und in das 2. Intervall einzutreten, ihre verbleibende mittlere Lebensdauer mit 1177 Tagen größer ist als zu Beginn des 1. Intervalls. Das ist natürlich eine Folge der hohen Sterblichkeit im 1. Intervall. Siehe Spalte 6.

P39.4 Vergleich von Gruppen

Ist die Überlebensdauer von herztransplantierten Männern signifikant anders als die von Frauen? Allgemein geht es also darum zu untersuchen, ob sich die Überlebensfunktionen mehrer Gruppen signifikant unterscheiden. Wir wollen unsere Testdaten daraufhin befragen, ob es einen Unterschied ausmacht, in welchem Hospital der Patient herztransplantiert wurde.

P39.4.1 Almo-Eingabe mit **Prog39mb**

Klicken Sie am Oberrand des Almo-Fenster auf den Knopf VERFAHREN. Es erscheint eine Übersicht über alle in Almo enthaltene Verfahren. Siehe dazu P39.2.1. Klicken Sie auf den Eintrag "Ereignisanalyse" oder "Survival". Almo präsentiert Ihnen dann eine Liste der Maskenprogramme zum Thema "Survival". Laden Sie Prog39mb. Es erscheint dann ein Maskenprogramm, das mit dem oben in Abschnitt P39.2.1 dargestellten Prog39ma identisch ist. Es ist nur die Box „Gruppierungs-Variable“ hinzugefügt.

Box „Gruppierungs-Variable

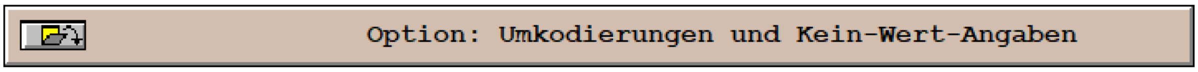
Beachte:

Die Zahl der Gruppierungsvariablen ist beliebig. Almo bildet dann alle Kombinationen, die sich aus den Ausprägungen der Gruppierungsvariablen zusammenstellen lassen. Sie können z.B. schreiben:

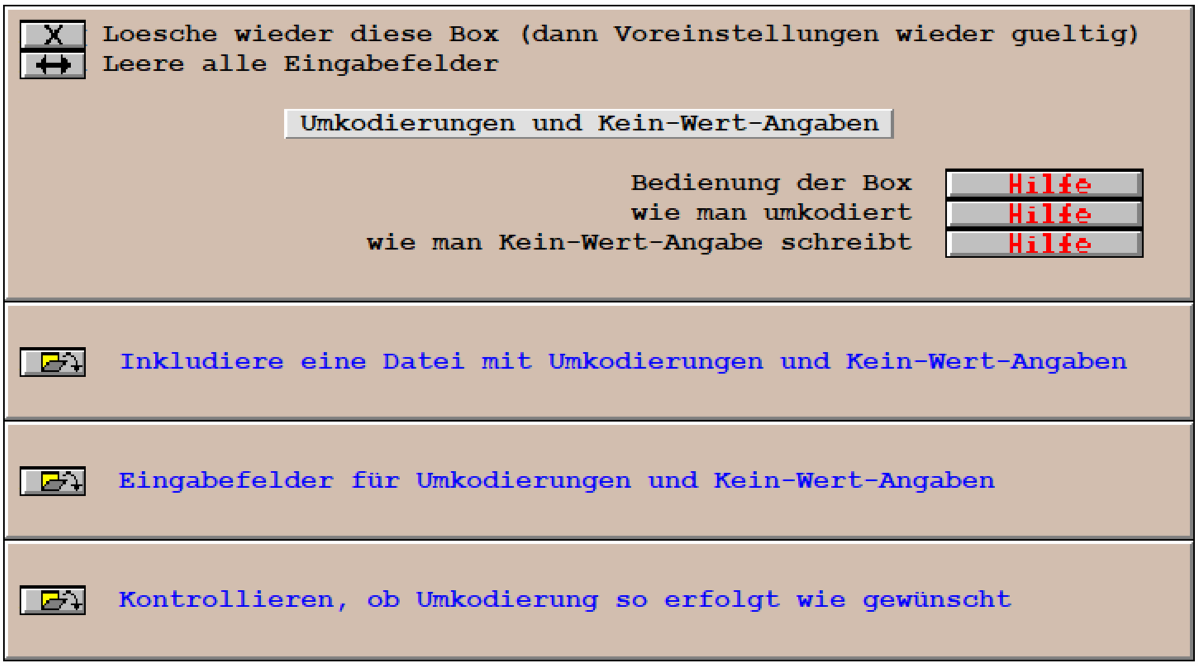
Hospital, Alter;

Da das Alter sehr viele Ausprägungen besitzt, würden außerordentlich viele Gruppen entstehen. In diesem Falle wäre es also sinnvoll das Alter zu dichotomisieren, z.B. durch eine Anweisung in der Umkodierungsbox

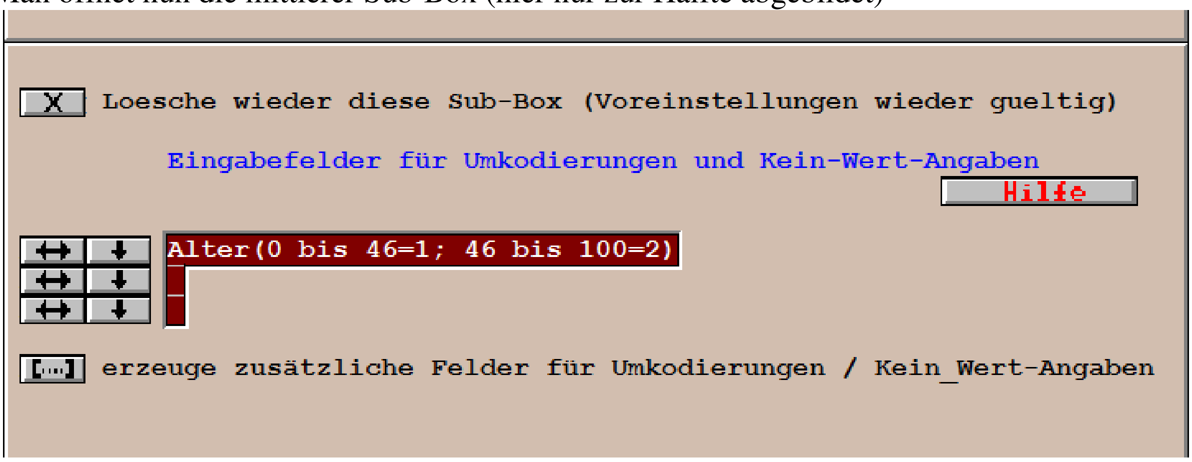
Öffnen Sie in diesem Fall die Optionsbox "Umkodierungen und Kein-Wert-Angaben"



Man sieht folgendes



Man öffnet nun die mittlerer Sub-Box (hier nur zur Hälfte abgebildet)



Zum Umkodieren siehe die ausführliche Erläuterung im Dokument P0 "Arbeiten_mit_Almo", Abschnitt P0.5.

P39.4.3 Almo-Ausgabe

```
Ergebnisse aus Almo
-------------------
Fuer Analyse aus Datenvektor ausgewaehlte Variable

Gruppierungsvariable: V3  Hospital  HospitalA HospitalB HospitalC
Zeitvariable:         V1  Zeitdauer
Ereignisvariable:     V2  Ereignis

Zahl der eingelesenen  Datensaetze:      65
Zahl der ausgewerteten Datensaetze:      65


Sterbetafel
Gruppe  1:  V3 Hospital: 1 HospitalA
-------------------------------------

  1          2             3          4            5           6
Inter-   Intervalldauer  Pers. die  Pers. die    Pers. die   Pers. die
vall      von     bis    in dieses  waehrend     dem Risiko  waehrend
 Nr.                     Intervall  dieses Intv  zu termi-   dieses Intv.
                         eintreten  ausscheiden  nieren      terminierten
                                    (zensierte   ausgesetzt
                                    Beobachtgn)  sind

  1      0.00 -  162.00     22          5          19.5          8
  2    162.00 -  324.00      9          2           8.0          1
  3    324.00 -  486.00      6          1           5.5          0
  4    486.00 -  648.00      5          2           4.0          1
  5    648.00 -  810.00      2          0           2.0          0
  6    810.00 -  972.00      2          0           2.0          0
  7    972.00 - 1134.00      2          1           1.5          0
  8   1134.00 - 1296.00      1          0           1.0          0
  9   1296.00 - 1458.00      1          0           1.0          0
 10   1458.00 - 1620.00      1          1           0.5          0


  7          8             9          10          11           12
Inter-   Intervalldauer  Whrschlkt  Whrschlkt   Ueberleb.fkt. Dichte-
vall      von     bis    zu termi-  zu ueber-   =kumulative   funktion
 Nr.                     nieren     leben       Whrschlkt     der Ueber-
                                                zu ueber-     lebensdauer
                                                leben bei
                                                Intvallende

  1      0.00 -  162.00   0.4103     0.5897       0.5897      0.002532
  2    162.00 -  324.00   0.1250     0.8750       0.5160      0.000455
  3    324.00 -  486.00   0.0000     1.0000       0.5160      0.000000
  4    486.00 -  648.00   0.2500     0.7500       0.3870      0.000796
  5    648.00 -  810.00   0.0000     1.0000       0.3870      0.000000
  6    810.00 -  972.00   0.0000     1.0000       0.3870      0.000000
  7    972.00 - 1134.00   0.0000     1.0000       0.3870      0.000000
  8   1134.00 - 1296.00   0.0000     1.0000       0.3870      0.000000
  9   1296.00 - 1458.00   0.0000     1.0000       0.3870      0.000000
 10   1458.00 - 1620.00   0.0000     1.0000       0.3870      0.000000

  13         14            15         16          17           18
Inter-   Intervalldauer  Hazardrate  Std.fehler  Std.fehler  Std.fehler
vall      von     bis    Risikorate  der         der         der
 Nr.                                 Hazardrate  kumulativen Dichte-
                                                 Whrschlkt   funktion
                                                 zu ueber-   der Ueber-
                                                 leben       lebensdauer

  1      0.00 -  162.00   0.003186    0.001088    0.111389    0.000688
  2    162.00 -  324.00   0.000823    0.000821    0.119392    0.000434
  3    324.00 -  486.00   0.000000    0.000000    0.119392    0.000000
  4    486.00 -  648.00   0.001764    0.001746    0.143179    0.000714
  5    648.00 -  810.00   0.000000    0.000000    0.143179    0.000000
  6    810.00 -  972.00   0.000000    0.000000    0.143179    0.000000
  7    972.00 - 1134.00   0.000000    0.000000    0.143179    0.000000
  8   1134.00 - 1296.00   0.000000    0.000000    0.143179    0.000000
```

9	1296.00 - 1458.00	0.000000	0.000000	0.143179	0.000000
10	1458.00 - 1620.00	0.000000	0.000000	0.143179	0.000000

19 Inter-vall Nr.	20 Intervalldauer von bis	21 Median der verbleibenden Lebensdauer bei Inter-vall-Beginn	22 Std.fehler des Median der verbleibenden Lebensdauer
1	0.00 - 162.00	506.1242	142.185864
2	162.00 - 324.00	-	-
3	324.00 - 486.00	-	-
4	486.00 - 648.00	-	-
5	648.00 - 810.00	-	-
6	810.00 - 972.00	-	-
7	972.00 - 1134.00	-	-
8	1134.00 - 1296.00	-	-
9	1296.00 - 1458.00	-	-
10	1458.00 - 1620.00	-	-

Median der Ueberlebensdauer = 506.1242
Standardfehler des Medians = 142.1859

Sterbetafel
Gruppe 2: V3 Hospital: 2 HospitalB
----

1 Inter-vall Nr.	2 Intervalldauer von bis	3 Pers. die in dieses Intervall eintreten	4 Pers. die waehrend dieses Intv ausscheiden (zensierte Beobachtgn)	5 Pers. die dem Risiko zu termi-nieren ausgesetzt sind	6 Pers. die waehrend dieses Intv. terminierten
1	0.00 - 162.00	22	6	19.0	8
2	162.00 - 324.00	8	1	7.5	1
3	324.00 - 486.00	6	2	5.0	0
4	486.00 - 648.00	4	0	4.0	0
5	648.00 - 810.00	4	1	3.5	1
6	810.00 - 972.00	2	1	1.5	0
7	972.00 - 1134.00	1	0	1.0	0
8	1134.00 - 1296.00	1	0	1.0	0
9	1296.00 - 1458.00	1	0	1.0	0
10	1458.00 - 1620.00	1	0	1.0	0
11	1620.00 - 1782.00	1	1	0.5	0

7 Inter-vall Nr.	8 Intervalldauer von bis	9 Whrschlkt zu termi-nieren	10 Whrschlkt zu ueber-leben	11 Ueberleb.fkt. =kumulative Whrschlkt zu ueber-leben bei Intvallende	12 Dichte-funktion der Ueber-lebensdauer
1	0.00 - 162.00	0.4211	0.5789	0.5789	0.002599
2	162.00 - 324.00	0.1333	0.8667	0.5018	0.000476
3	324.00 - 486.00	0.0000	1.0000	0.5018	0.000000
4	486.00 - 648.00	0.0000	1.0000	0.5018	0.000000
5	648.00 - 810.00	0.2857	0.7143	0.3584	0.000885
6	810.00 - 972.00	0.0000	1.0000	0.3584	0.000000
7	972.00 - 1134.00	0.0000	1.0000	0.3584	0.000000
8	1134.00 - 1296.00	0.0000	1.0000	0.3584	0.000000
9	1296.00 - 1458.00	0.0000	1.0000	0.3584	0.000000
10	1458.00 - 1620.00	0.0000	1.0000	0.3584	0.000000
11	1620.00 - 1782.00	0.0000	1.0000	0.3584	0.000000

13 Inter-vall Nr.	14 Intervalldauer von bis	15 Hazardrate Risikorate	16 Std.fehler der Hazardrate	17 Std.fehler der kumulativen Whrschlkt	18 Std.fehler der Dichte-funktion

				zu ueber- leben	der Ueber- lebensdauer
1	0.00 - 162.00	0.003292	0.001122	0.113269	0.000699
2	162.00 - 324.00	0.000882	0.000880	0.121659	0.000453
3	324.00 - 486.00	0.000000	0.000000	0.121659	0.000000
4	486.00 - 648.00	0.000000	0.000000	0.121659	0.000000
5	648.00 - 810.00	0.002058	0.002029	0.149101	0.000778
6	810.00 - 972.00	0.000000	0.000000	0.149101	0.000000
7	972.00 - 1134.00	0.000000	0.000000	0.149101	0.000000
8	1134.00 - 1296.00	0.000000	0.000000	0.149101	0.000000
9	1296.00 - 1458.00	0.000000	0.000000	0.149101	0.000000
10	1458.00 - 1620.00	0.000000	0.000000	0.149101	0.000000
11	1620.00 - 1782.00	0.000000	0.000000	0.149101	0.000000

19 Inter- vall Nr.	20 Intervalldauer von bis	21 Median der verbleibenden Lebensdauer bei Inter- vall-Beginn	22 Std.fehler des Median der verbleibenden Lebensdauer
1	0.00 - 162.00	649.9825	129.623900
2	162.00 - 324.00	-	-
3	324.00 - 486.00	-	-
4	486.00 - 648.00	-	-
5	648.00 - 810.00	-	-
6	810.00 - 972.00	-	-
7	972.00 - 1134.00	-	-
8	1134.00 - 1296.00	-	-
9	1296.00 - 1458.00	-	-
10	1458.00 - 1620.00	-	-
11	1620.00 - 1782.00	-	-

Median der Ueberlebensdauer = 649.9825
Standardfehler des Medians = 129.6239

Sterbetafel
Gruppe 3: V3 Hospital: 3 HospitalC
----

1 Inter- vall Nr.	2 Intervalldauer von bis	3 Pers. die in dieses Intervall eintreten	4 Pers. die waehrend dieses Intv ausscheiden (zensierte Beobachtgn)	5 Pers. die dem Risiko zu termi- nieren ausgesetzt sind	6 Pers. die waehrend dieses Intv. terminierten
1	0.00 - 162.00	21	3	19.5	3
2	162.00 - 324.00	15	1	14.5	2
3	324.00 - 486.00	12	1	11.5	0
4	486.00 - 648.00	11	2	10.0	0
5	648.00 - 810.00	9	0	9.0	0
6	810.00 - 972.00	9	2	8.0	1
7	972.00 - 1134.00	6	0	6.0	2
8	1134.00 - 1296.00	4	1	3.5	0
9	1296.00 - 1458.00	3	1	2.5	1
10	1458.00 - 1620.00	1	1	0.5	0

7 Inter- vall Nr.	8 Intervalldauer von bis	9 Whrschlkt zu termi- nieren	10 Whrschlkt zu ueber- leben	11 Ueberleb.fkt. =kumulative Whrschlkt zu ueber- leben bei	12 Dichte- funktion der Ueber- lebensdauer

Intvallende

1	0.00 - 162.00	0.1538	0.8462	0.8462	0.000950
2	162.00 - 324.00	0.1379	0.8621	0.7294	0.000720
3	324.00 - 486.00	0.0000	1.0000	0.7294	0.000000
4	486.00 - 648.00	0.0000	1.0000	0.7294	0.000000
5	648.00 - 810.00	0.0000	1.0000	0.7294	0.000000
6	810.00 - 972.00	0.1250	0.8750	0.6383	0.000563
7	972.00 - 1134.00	0.3333	0.6667	0.4255	0.001313
8	1134.00 - 1296.00	0.0000	1.0000	0.4255	0.000000
9	1296.00 - 1458.00	0.4000	0.6000	0.2553	0.001051
10	1458.00 - 1620.00	0.0000	1.0000	0.2553	0.000000

13 Inter-vall Nr.	14 Intervalldauer von bis	15 Hazardrate Risikorate	16 Std.fehler der Hazardrate	17 Std.fehler der kumulativen Whrschlkt zu ueber-leben	18 Std.fehler der Dichte-funktion der Ueber-lebensdauer
1	0.00 - 162.00	0.001029	0.000592	0.081705	0.000504
2	162.00 - 324.00	0.000914	0.000645	0.104079	0.000478
3	324.00 - 486.00	0.000000	0.000000	0.104079	0.000000
4	486.00 - 648.00	0.000000	0.000000	0.104079	0.000000
5	648.00 - 810.00	0.000000	0.000000	0.104079	0.000000
6	810.00 - 972.00	0.000823	0.000821	0.124773	0.000533
7	972.00 - 1134.00	0.002469	0.001711	0.148349	0.000801
8	1134.00 - 1296.00	0.000000	0.000000	0.148349	0.000000
9	1296.00 - 1458.00	0.003086	0.002988	0.159073	0.000892
10	1458.00 - 1620.00	0.000000	0.000000	0.159073	0.000000

19 Inter-vall Nr.	20 Intervalldauer von bis	21 Median der verbleibenden Lebensdauer bei Inter-vall-Beginn	22 Std.fehler des Median der verbleibenden Lebensdauer
1	0.00 - 162.00	1077.2790	86.216338
2	162.00 - 324.00	1136.3143	105.750449
3	324.00 - 486.00	1029.8571	102.366844
4	486.00 - 648.00	867.8571	109.776210
5	648.00 - 810.00	705.8571	115.714286
6	810.00 - 972.00	543.8571	122.733534
7	972.00 - 1134.00	425.2500	124.005418
8	1134.00 - 1296.00	-	-
9	1296.00 - 1458.00	-	-
10	1458.00 - 1620.00	-	-

Median der Ueberlebensdauer = 1077.2790
Standardfehler des Medians = 86.2163

******************** MITTEILUNG
Fuer nachfolgenden logrank- und Wilcoxon-Test gilt: Alle Untersuchungs-einheiten, die im gleichen Zeitintervall "sterben" oder zensiert werden erhalten denselben "Todeszeitpunkt" bzw Zensierunszeitpunkt

Ereignis eingetreten in Gruppe 1 = 10 mal
Ereignis eingetreten in Gruppe 2 = 10 mal
Ereignis eingetreten in Gruppe 3 = 9 mal

Pauschaler logrank-Test auf Verschiedenheit aller Gruppen

Chi-Quadrat = 1.499863
df = 2
p = 0.476960
Signifikanz (1-p)100 der Verschiedenheit= 52.303995 %

Pauschaler Wilcoxon-(Breslow)-Test auf Verschiedenheit aller Gruppen
----
Chi-Quadrat = 3.142796
df = 2
p = 0.206401

```
Signifikanz (1-p)100 der Verschiedenheit= 79.359854 %


Chi-Quadrat-Werte der paarweisen Vergleiche der Gruppen
aus logrank-Test
  Oberer Wert ist Chi-Quadrat
  darunter p
  darunter die Signifikanz (1-p)*100 der Verschiedenheit der Gruppen
--------------------------------------------------------------------

Gruppe 1 = Hospital:HospitalA
Gruppe 2 = Hospital:HospitalB
Gruppe 3 = Hospital:HospitalC

            Gruppe 1    Gruppe 2
---------------------------------

Gruppe 2    0.000772
            0.977169
            2.283078%
---------------------------------

Gruppe 3    1.111535    1.530701
            0.291866    0.216168
           70.813360%  78.383222%
---------------------------------


Chi-Quadrat-Werte der paarweisen Vergleiche der Gruppen
aus Wilcoxon- (Breslow-) Test
  Oberer Wert ist Chi-Quadrat
  darunter p
  darunter die Signifikanz (1-p)*100 der Verschiedenheit der Gruppen
--------------------------------------------------------------------

Gruppe 1 = Hospital:HospitalA
Gruppe 2 = Hospital:HospitalB
Gruppe 3 = Hospital:HospitalC

            Gruppe 1    Gruppe 2
---------------------------------

Gruppe 2    0.000192
            0.990742
            0.925762%
---------------------------------


Gruppe 3    2.445397    2.584223
            0.118062    0.108145
           88.193782%  89.185504%
---------------------------------
```

P39.4.4 Berechnung des logrank-Test und des Wilcoxon (Breslow-) Test.

Almo verfährt hier in folgender Weise.
Alle Personen, die in einem Zeitintervall sterben oder ausscheiden (zensiert sind) erhalten denselben Todeszeitpunkt bzw. Zensierungszeitpunkt - obwohl ihre individuellen Todes- bzw. Zensierungszeitpunkte innerhalb des Intervalls aus den Daten bekannt sind.
Die Gründe für diese Vorgehensweise sind folgende:

1. Almo rechnet im Programm 40 (=Kaplan-Meier-Schätzer der Überlebensfunktion) den logrank- und Wilcoxon(Breslow)-Test aus den individuellen Daten. Will der Benutzer also diese beiden Tests mit individuellen Daten rechnen, dann braucht er nur Programm 40 zu rechnen.

2. Beim log-rank- und Wilcoxon(Breslow)-Test müssen die Daten nach aufsteigendem Todeszeitpunkt sortiert werden. Die Daten müssen zu diesem Zweck in einer Matrix im Speicher gehalten werden. Bei vielen Untersuchungspersonen - und für diesen Fall ist ja die Sterbetafel-Methode gedacht - kann es zu einem Speicherüberlauf kommen, mit der Folge, daß das Betriebssystem abstürzt oder auslagert ("swapping") und dann außerordentlich lange Rechenzeiten auftreten. Bei der Vorgehensweise, die in Almo gewählt wird kann dieser Fall nicht auftreten.
3. Die Untersuchungspersonen können bei der Sterbetafel-Methode in verschieden lange Zeitintervalle eingeteilt werden. Die Überlebensfunktion wird dadurch beeinflußt - nicht jedoch die Ergebnisse aus dem log-rank- und dem Wilcoxon(Breslow)-Test, sofern dieser die individuellen Daten verwendet; d.h. Überlebensfunktion und Tests werden aus verschiedenen Daten gerechnet. Das erscheint uns nicht sinnvoll. Bei der Vorgehensweise in Almo werden die Überlebensfunktion und die Tests aus denselben Daten berechnet.

Vergleich mit anderen Statistiksystemen:
SAS rechnet den log-rank-Test und den Wilcoxon(Breslow)-Test aus den individuellen Daten. SPSS und Statistica rechnen den Wilcoxon-Gehan-Test, der etwas andere Ergebnisse erbringt. Dies gilt für die genannten Statistiksysteme ab den Versionen von 1995 und später. Bei den jeweils aktuellen Versionen können Änderungen vorgenommen sein.
Die Formeln zur Berechnung des log-rank- und des Wilcoxon(Breslow)-Tests in Almo werden in Abschnitt P40.7.4 ausführlich dargestellt und kommentiert.

P39.5 Sterbetafel-Methode und Kaplan-Meier-Verfahren

Das Kaplan-Meier-Verfahren ist immer vorzuziehen, da es die individuellen "Todes- und Zensierungszeiten" berücksichtigt. Beim Kaplan-Meier-Verfahren müssen die individuellen Untersuchungseinheiten sortiert werden und in einer Matrix im Speicher des Rechners gehalten werden. Bei sehr vielen Daten kann der Rechner dann überfordert sein. Hier muß das Sterbetafel-Verfahren gerechnet werden. Dies gilt auch, wenn, bei vielen Daten, die graphische Darstellung der Kaplan-Meier-Überlebensfunktion, zu unübersichtlich wird.

Gelegentlich ist zu lesen, daß die Sterbetafel-Methode und das Kaplan-Meier-Verfahren dieselben Ergebnisse erbringt, wenn beim Sterbetafel-Verfahren die Intervalle nach der kleinsten Zeiteinheit gebildet werden. Das ist nur teilweise richtig. Der logrank- und Wilcoxon(Breslow)-Test aus beiden Verfahren liefert dieselben Ergebnisse (wenn eine Gruppierungsvariable eingeführt wurde).

Die Überlebensfunktionen die Almo ausgibt divergieren etwas, wenn in der Regel auch sehr wenig, weil beim Sterbetafel-Verfahren die Intervallmitte als Bezugszeit verwendet wird und dann nur die Hälfte der Zensierten in die Berechnung der Risikomenge eingeht. Siehe dazu das Almo-Beispielprogramm KM_SURV.Alm. Almo erkennt sozusagen nicht, daß die Sterbetafel-Methode mit der kleinsten Intervallbreite gerechnet wird und daß in diesem Fall keine Intervallmitte mehr existiert.

Es macht also wenig Sinn eine Sterbetafel-Methode zu rechnen, wenn die kleinst möglichen Intervalle verwendet wird. Hier ist das Kaplan-Meier-Verfahren zu rechnen.

Literatur:

H.-P. Blossfeld/A. Hamerle/K.u.Mayer: Ereignisanalyse, Campus, Frankfurt 1986
A. Diekmann/P. Mitter: Methoden zur Analyse von Zeitverläufen, Teubner, Stuttgart, 1984
A.J. Gross/V.A. Clark: Survival distributions. Reliability Applications in the Biomedical Sciences, Wiley, New York, 1975 (insbesondere Kap. 2.2, 2.3)
J.D. Kalbfleisch/R.L. Prentice: The statistical analysis of failure time data, Wiley, New York, 1980
J.F. Lawless: Statistical models and methods for lifetime data, Wiley, New York, 1982

P40 Der Kaplan-Meier-Schätzer der Überlebensdauer

Wie lange überlebt ein Mensch eine Herztransplation. Wir wollen etwas präziser fragen:

1. Was ist die durchschnittliche Überlebensdauer nach einer Herztransplantation?
2. Wie groß ist die Wahrscheinlichkeit nach einer Herztransplantation nach x Tagen (z.B. nach 365 Tagen) noch zu leben?

Das Kaplan-Meier-Verfahren liefert zur Beantwortung dieser Fragen Schätzwerte.

Es muß bei den Themen, auf die das Kaplan-Meier-Verfahren angewendet wird, nicht immer um Tod und Leben gehen. Das Kaplan-Meier-Verfahren ist etwa auch auf sozialwissenschaftliche Probleme anwendbar. So wenden z.B. Blossfeld, Hamerle, Mayer (1986) das Kaplan-Meier-Verfahren auf folgende Fragestellung an: Wie lange verbleiben Menschen in ihrem ersten Beruf. Dem "Tod" entspricht hier der Berufswechsel, dem "Überleben" der Verbleib im 1. Beruf.

Wir wollen im folgenden ausführlich ein Beispiel von Burns betrachten, das wir aus Altman (1991) entnehmen:

21 Personen wurden auf einer beweglichen Plattform 2 Stunden lang seekrank gemacht. Das Ereignis, auf das die Forscher dabei warteten, war natürlich nicht der Tod der Versuchspersonen. Man wollte wissen nach wievielen Minuten die Versuchspersonen erbrechen. Die "Überlebensdauer" war die Zeit bis zum Erbrechen, bzw. die vollen 120 Minuten bis zum Versuchsende.

Die Ergebnisse waren folgende:

Untersuchungsperson	Überlebensdauer in Minuten
1	30
2	50
3	50*
4	51
5	66*
6	82
7	92
8	120*
.	.
.	.
.	.
.	.
21	120*

Bei Person 1 ist nach 30 Minuten das Ereignis "Erbrechen" eingetreten. Person 2 mußte nach 50 Minuten erbrechen. Person 3 hat nach 50 Minuten "gestreikt" und ist von der Plattform gehüpft. In der Sprache des Kaplan-Meier-Verfahrens wird Person 3 als "zensierte" Beobachtung bezeichnet. Wir haben die zensierten Personen in obiger Tabelle durch einen Stern markiert.

Auch Person 5 ist eine derartige zensierte Beobachtung. Sie hatte nach 66 Minuten genug und entzog sich dem Experiment. Die Personen 8 bis 21 haben alle das Versuchsende erreicht ohne daß bei ihnen das Ereignis "Erbrechen" aufgetreten ist. Auch sie werden als "zensierte" Beobachtungen bezeichnet. Der etwas verwirrende Begriff "zensiert" meint also:

das fragliche Ereignis ist nicht aufgetreten - entweder weil die Untersuchungseinheit die Beobachtung verließ oder weil sie bis zum Beobachtungsende überlebte.

P40.1 Die Daten

Die Daten, die Almo benötigt um für unser Beispiel den Kaplan-Meier-Schätzer der Überlebensfunktion zu rechnen, sind folgende:

V1 Zeitdauer	V2 Ereignis
30	0
50	0
50	1
51	0
66	1
82	0
92	0
120	1
.	.
.	.
.	.
120	1

Beachte:

1. Selbstverständlich kann die Datenmatrix mehr als nur 2 Variable umfassen. Die Zeitvariable muß auch nicht unbedingt V1 und die Ereignisvariable muß nicht unbedingt V2 sein.

2. Die Ereignisvariable ist in folgender Weise kodiert:

   0 = Ereignis eingetreten
   1 = zensiert

Diese Codes sind im Prinzip beliebig wählbar. Jedoch muß der Code für "zensiert" höher sein als der für "Ereignis eingetreten".

3. Die Variablenwerte können als Ganzzahlwerte geschrieben sein aber auch als Dezimalwerte mit Dezimalpunkt und beliebig vielen Dezimalstellen.

P40.2 Almo-Eingabe

Die Eingabe für dieses Beispiel ist folgende:

P40.2.1 Die Programm-Maske **Prog40m1**

Klicken Sie am Oberrand des Almo-Fenster auf den Knopf VERFAHREN. Es erscheint eine Übersicht über alle in Almo enthaltene Verfahren. Scrollen Sie dann durch bis "Ereignisanalyse" oder "Survival".

Prog40m1.Msk
Kaplan-Meier-Schätzer der Überlebensfunktion

Beispiel:
Die Überlebensdauer von herztransplantierten Menschen wird berechnet.

Die Daten sind folgende:

Zeitdauer des Überlebens	Ereignis 0=tot 1=zensiert
30	0
50	0
50	1
61	0
.	.
.	.

Die 1. Person ist nach 30 Tagen gestorben, die 2. nach 50 Tagen. Die 3. Person ist nach 50 Tagen aus der Beobachtung ausgeschieden (man nennt das "zensiert") etc.

Almo errechnet u.a. folgende Koeffizienten

0	1	2		3	4	5
Pers	Zeit des Über-lebens	Ereignis	Zahl	nach dieser Zeit noch vor-handen	Anteil derer, die noch vor-handen	Kaplan-Meier-Überleb.fkt. =kumulierte Whrscheinlkt zu überleben
1	30	tot	1	20	0.9524	0.952381
2	50	tot	1	19	0.9500	0.904762
3	50	zens	1	18		
4	61	tot	1	17	0.9444	0.854497
	.	.	.	.	.	.
	.	.	.	.	.	.

Mittelwert der Ueberlebensdauer = 105.387897 Tage
Standardabweichung = 6.124213 Tage

Betrachten wir z.B. den 3. Wert in der Spalte 4. Er sagt folgendes aus: Wenn ein Patient im Intervall vom 50. Tag bis einen Tag vor dem 61. Tag, also bis zum 60. Tag, noch lebt, dann ist seine Chance, auch den 61. Tag zu überleben =0.9444

Der 3. Wert aus der Kaplan-Meier-Überlebensfunktion in Spalte 5 ist so zu interpretieren: Zum Zeitpunkt der Transplantation (Tag 0) ist die Wahrscheinlichkeit bis zum Ende des 61. Tags zu überleben =0.854497 (=85.4%)

Grafik: Liniendiagramm der Überlebenswahrscheinlichkeit
siehe Handbuch, Abschnitt P40

Was ist ein Kurzprogramm ? --> Hilfe
Bedienung --> Hilfe

1 Speicher fuer x Variable Hilfe
Vereinbare Variable= 20 ;

2 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Datei der Variablennamen Hilfe

zeige = Namensdatei in Output zeigen
leer = nicht

4 Freie Namensfelder Hilfe

Name1=Ueberlebensdauer; Zeitvariable

Name2=Ereignis:tot,zensiert; Ereignisvariable

erzeuge zusätzliche Namensfelder

5 Datei aus der gelesen wird Hilfe bei Datei-Problemen

"C:\Almo7\Testdat\Altman1.fre"

frei Format der Daten Hilfe

V1:2 der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

6 Wenn Dateiformat FIX oder Nicht-Standard-FREI Hilfe

7 die Analyse-Variablen

Ueberlebensdauer die Zeitvariable

Ereignis die Ereignisvariable

0 Code für "Ereignis hat stattgefunden"
z.B. Untersuchungseinheit ist gestorben

1 Code für "zensiert"
z.B. Untersuchungseinheit hat ueber-
lebt oder ist ausgeschieden

8 Option: Ein- und Ausschliessen von Untersuchungseinheiten

9 Option: Umkodierungen und Kein-Wert-Angaben

10 Option: Untersuchungseinheiten ganzzahlig gewichten

Ausgabe

1 0=volle Ausgabe.Jede einzelne Person wird ausgegeben
1=verkürzte Ausgabe.Personen mit gleicher Zeitdauer
und gleichem Ereignis werden zusammengefasst

Option: Zwischenergebnisse ausgeben

Grafik-Optionen

Programmende

Erläuterungen zu den Boxen

Box 1 bis 6: Siehe Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.1 bis P0.4

Box 7: Die Analyse-Variablen

```
                    die Analyse-Variablen

[↔] Ueberlebensdauer         die Zeitvariable

[↔] Ereignis                 die Ereignisvariable

[↔] 0                        Code für "Ereignis hat stattgefunden"
                             z.B. Untersuchungseinheit ist gestorben

[↔] 1                        Code für "zensiert"
                             z.B. Untersuchungseinheit hat ueber-
                             lebt oder ist ausgeschieden
```

Box 8: Ein- und Ausschliessen von Untersuchungseinheiten. Siehe P0.7
Box 9: Umkodierungen und Kein-Wert-Angabe. Siehe P0.5
Box 10: Option: Untersuchungseinheiten ganzzahlig gewichten. Siehe P0.8

Box 11: Ausgabe

```
                         Ausgabe

[↑↓] 1        0=volle Ausgabe.Jede einzelne Person wird ausgegeben
              1=verkürzte Ausgabe.Personen mit gleicher Zeitdauer
                und gleichem Ereignis werden zusammengefasst
```

Der Umfang der Ergebnis-Ausgabe kann gesteuert werden

Optionsbox 12: Zwischenergebnisse ausgeben
Wird die Optionsbox geöffnet, dann sieht man folgendes:

```
[X] Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

                 Zwischenergebnisse ausgeben

[↑↓] 0          1=verschiedene Zwischenergebnisse ausgeben
                  z.B. Zwischenergebnisse aus Berechnung der
                  mittleren Lebensdauer, des log rank Tests etc.
                  ACHTUNG: Es kann sehr viel Text entstehen !!
                0=keine Zwischenergebnisse
```

Zusaätzlich zur Standardausgabe können Zwischenergebnisse ausgegeben werden

Box 13: Grafik-Optionen
Wird die Optionsbox geöffnet, dann sieht man folgendes:

```
[X] Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

                      Grafik-Optionen

[    ] Almo         Almo = Almo-Grafik ausgeben
                    0    = keine Grafik

[↑↓] 1              1 = auch den linken Teil der Überlebenskurve von
                        Zeit=0 und Überlebenswahrsch.keit p=1.0 zeigen
                    0 = nicht

[↑↓] 1              1 = Almo-Grafik in Ergebnisliste einsetzen
                    0 = nicht                    [ Hilfe ]
```

Wird das 2. Eingabefeld auf "1" gesetzt, dann zeigt Almo auch den Linken Teil der Überlebenskurve.

P40.3 Ausgabe

Almo liefert folgende Ergebnisse:

```
Ergebnisse aus ALMO
-------------------
Fuer Analyse aus Datenvektor ausgewaehlte Variable

Zeitvariable:         V1
Ereignisvariable:     V2

Zahl der eingelesenen Untersuchungseinheiten:      21


Erlaeuterung zu Spalte 2 der nachfolgenden Tabelle:
  Code: 0.00 = Ereignis eingetreten (z.B. gestorben)
        1.00 = Untersuchungseinheit zensiert
                (z.B. ueberlebt oder ausgeschieden)
                zusaetzlich durch einen Stern markiert

Erlaeuterung zu Spalte 3:
  Dies ist die Zahl der Untersuchungseinheiten,die
  nach dem jeweiligen Zeitpunkt noch vorhanden sind

Erlaeuterung zu Spalte 4:
  Dies ist der Anteil der Untersuchungseinheiten,
  die nach dem jeweiligen Zeitpunkt noch verbleiben
  - gemessen an denjenigen, die zum Beginn dieses
  Zeitpunkts noch vorhanden waren

Erlaeuterung zu Spalte 5:
  Dies ist der Kaplan-Meier-Schaetzer der Ueber-
  lebensfunktion. Die Werte entstehen durch Multi-
  plikation der Werte aus Spalte 4

Erlaeuterung zu Spalte 6:
  Dies ist der Standardfehler der Kaplan-Meier-Ueberlebensfunktion

Erlaeuterung zu Spalte 7:
  Dies sind die kumulierten Untersuchungseinheiten, die
  das Ereignis erfahren haben (z.B. gestorben sind)

Erlaeuterung zu Spalte 8:
  Dies ist die einfache Wahrscheinlichkeit zum jeweiligen Zeitpunkt i
  noch zu leben. Entsteht aus Gesamtzahl der Untersuchungseinheiten
  minus den "Toten" (aus Spalte 7) dividiert durch die Gesamtzahl
  Beachte: Zensierte Untersuchungseinheiten werden als Ueberlebende betrachtet


     1       2        3       4        5              6          7      8
                    nach    Anteil  Kaplan-Meier-
                    dieser  derer,  Ueberleb.fkt.                     einf.
                    Zeit    die     = kumulierte  Stand.fehler        Whkeit
                    noch    noch    Whrscheinlkt  der           Tote  zu
           Ereignis vor-    vor-    zu ueber-     Kaplan-Meier  kumu- ueber-
   Zeit  Code       handen  handen  leben         Ueberleb.fkt  liert leben
   ----  ---------  ------  ------  ------------  ------------  ----- ------
  30.00     0.00        20  0.9524    0.952381      0.046471        1  0.9524
  50.00     0.00        19  0.9500    0.904762      0.064056        2  0.9048
  50.00     1.00*       18
  51.00     0.00        17  0.9444    0.854497      0.077757        3  0.8571
  66.00     1.00*       16
  82.00     0.00        15  0.9375    0.801091      0.089375        4  0.8095
  92.00     0.00        14  0.9333    0.747685      0.098084        5  0.7619
 120.00     1.00*       13
 120.00     1.00*       12
 120.00     1.00*       11
 120.00     1.00*       10
 120.00     1.00*        9
 120.00     1.00*        8
```

```
 120.00     1.00*       7
 120.00     1.00*       6
 120.00     1.00*       5
 120.00     1.00*       4
 120.00     1.00*       3
 120.00     1.00*       2
 120.00     1.00*       1
 120.00     1.00*       0

Mittelwert der Ueberlebensdauer        = 105.387897
Standardabweichung                     =   6.124213
95% Vertauensintervall fuer Mittelwert =  93.384439 - 117.391355

--Mittelwert wuerde unterschaetzt werden, da letzte Untersuchungseinheit
--zensiert ist. Deswegen wurde der Zeitraum bis zu dieser in den Kalkuel
--des Mittelwerts und der Standardabweichung miteinbezogen (Handbuch, P40)
--unterschaetzter Mittelwert = 84.452712   Standardabweichung =  4.259646

Quantile der Ueberlebensdauer     Stand.abwg.        95% Vertauensintervall
75%          =     92.00          nicht berechenbar  nicht berechenbar
50% (Median) = nicht berechenbar  nicht berechenbar  nicht berechenbar
25%          = nicht berechenbar  nicht berechenbar  nicht berechenbar
```

In der **4. Spalte** steht der Anteil der Personen, die nach dem jeweiligen Zeitpunkt noch verbleiben - gemessen an denjenigen, die zum Beginn dieses Zeitpunkts noch vorhanden waren.

Betrachten wir den 4. Wert von oben 0.9375. Er entsteht aus

$$\frac{15}{16} = 0.9375$$

Vor dem 82. Tag waren noch 16 Personen am Leben; nach dem 82. Tag waren es nur noch 15. Man könnte auch so formulieren: Wenn man im Intervall vom 51. Tag bis 1 Tag vor dem 82. Tag, also dem 81. Tag, noch lebt, dann ist die Wahrscheinlichkeit den 82. Tag zu überleben = 0.9357 (=93,75%).

In **Spalte 5** der Tabelle steht die Kaplan-Meier-Überlebensfunktion. Sie entsteht aus der Multiplikation der Werte der 4. Spalte. Betrachten wir als Beispiel den 4. Wert von oben 0.801091. Er entsteht aus

0.9524 * 0.9500 * 0.9444 * 0.9375
= 0.801091.

Wir bezeichnen deswegen die Werte der Kaplan-Meier Überlebensfunktion als "kumulierte Wahrscheinlichkeiten". Der Wert von 0.801091 ist inhaltlich nicht sehr anschaulich erklärt, wenn wir sagen " dies ist die kumulative Überlebenswahrscheinlichkeit". Wir können ungefähr in folgender Weise interpretieren:

Dies ist der Anteil der Personen, die den 82. Tag überleben - wobei die zensierten Fälle in einer spezifischen Weise berücksichtigt werden. Wir könnten auch formulieren: Gesehen vom Zeitpunkt 0 aus, ist die Wahrscheinlichkeit den 82. Tag zu überleben = 0.801091 - wobei die zensierten Fälle in einer spezifischen Weise berücksichtigt werden. Diese Formuliereungen gelten nicht nur ungefähr sondern exakt, wenn keine zensierten Fälle auftreten (siehe nachfolgende Erläuterung zu Spalte 8).

Die Formel für den Kaplan-Meier-Schätzer der Überlebenswahrscheinlichkeit lautet:

(1) $$p_i = \prod_{j=1}^{i} \left(1 - \frac{d_j}{n_j}\right)$$

i läuft von 1 bis k
i = der Index i bezeichnet den jeweiligen Zeitpunkt, zu dem jemand "gestorben" ist.

k = Zahl der Zeitpunkte zu dem eine oder auch gleich mehrere Personen gestorben sind. In unserem Beispiel ist k=5.
d_j = Zahl der "Toten" zum Zeitpunkt j.
n_j = Risikomenge zum Zeitpunkt j; das ist die Zahl der noch Lebenden zum Zeitpunkt j-1.

Beachte: Die Produktbildung in obiger Gleichung 1 läuft nur über die "Toten".

Beispiel: Zum Zeitpunkt i=1 ist n_j= 21 und d_j=1. Es entsteht p_1= (1-1/21) = 0.952381. Zum Zeitpunkt i=2 ist für j=2 n_j=20 und d_j=1. Es entsteht p_2= 0.952381*(1-1/20)=0.904762.

In **Spalte 6** sind die jeweiligen Standardfehler der Kaplan-Meier-Überlebens-Funktion enthalten. Der 4. Wert von oben 0.089375 bedeutet, daß der Kaplan-Meier-Schätzer der Überlebenswahrscheinlichkeit für den 82. Tag von 0.801091 eine Standardabweichung von 0.089375 besitzt. Daraus läßt sich ein Vertrauensintervall mit einem Konfidenzniveau von 95.5 % (t=2) errechnen:

0.801091 ± 2 * 0.089375

dabei ist der Wertebereich auf 0 bis 1.0 zu beschränken.

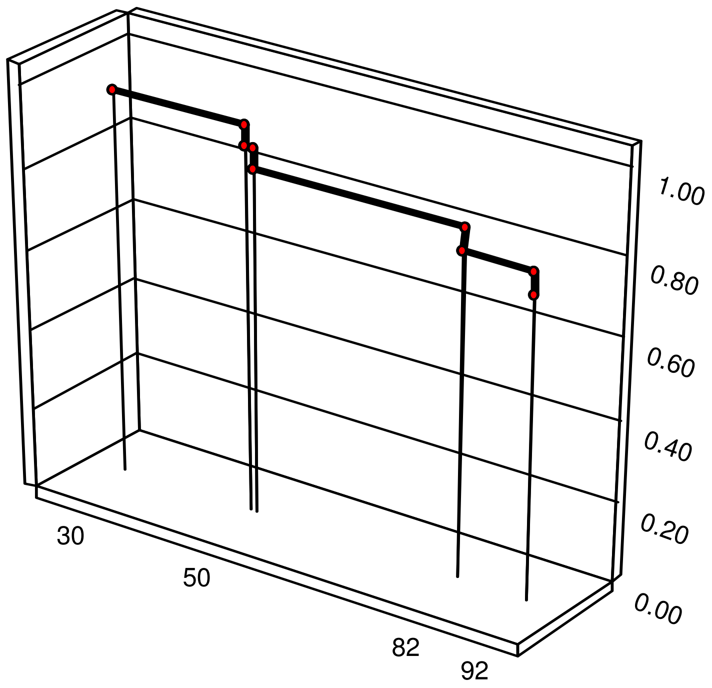
Wenn wir die Werte in Spalte 6 von oben nach unten vergleichen, dann erkennen wir, daß mit kleiner werdenden Zahlen von Personen, die noch verbleiben (Spalte 3) die Standardfehler (in Spalte 6) größer werden.

Der Standardfehler wird in Almo nach der Formel von Greenwood berechnet (siehe hierzu etwa Altmann, 1991, S. 379).

(2) $$SE(pi) = pi \; * \; \sqrt{\sum_{j=1}^{i} \left[\frac{d_j}{n_j(n_j - d_j)} \right]}$$

Wir wollen die Kaplan-Meier-Funktion jetzt noch graphisch darstellen:

Kaplan-Meier Ueberlebensfunktion



Die Überlebensfunktion ist eine Treppenfunktion. Die Wahrscheinlichkeit zu überleben bleibt solange konstant auf derselben Höhe - auch wenn ein zensierter Fall dazwischen liegt, bis wieder ein "Todesfall" eintritt. In obiger Graphik fehlt der vordere und hintere Teil. Vom 0. bis zum 30. Tag ist die Überlebenswahrscheinlichkeit 1.0. Vom 92. bis zum 120. Tag bleibt die Überlebenswahrscheinlichkeit auf dem Niveau des 92. Tags. In der Almo-Grafik haben wir diese beiden Kurvenenden weggelassen, weil die Dauer vom Beobachtungsbeginn (dem 0. Tag) bis zum Tag des 1. Ereignisses sehr lang sein kann und damit dieser eigentlich uninteressante Kurvenabschnitt sehr breit sein kann. Ebenso kann die Dauer vom letzten Ereignis bis zum Beobachtungsende (dem 120. Tag) sehr lang sein.

Wenn Sie das vordere Kurvenende doch sehen wollen, dann öffnen Sie die Optionsbox "Grafik-Optionen" und setzen im 2. Eingabefeld eine "1" ein:

In **Spalte 7** der Ergebnistabelle ist die kumulierte Zahl der "Toten" enthalten.

In **Spalte 8** ist die "einfache Wahrscheinlichkeit zu überleben" enthalten. Wir verwenden das Adjektiv "einfach", weil die zensierten Untersuchungseinheiten im Unterschied zur Kaplan-Meier-Überlebensfunktion nicht spezifisch behandelt werden. Sie werden als Überlebende betrachtet. Die Formel nach der die Werte in Spalte 8 berechnet werden, ist:

(3) $p_i^* = \dfrac{n - T_i}{n}$

p_i^* = einfache Überlebenswahrscheinlichkeit zum Zeitpunkt i
n = Gesamtzahl der Untersuchungseinheiten
T_i = bis zum Zeitpunkt i kumulierte Zahl der "Toten".

Die "einfache Überlebenswahrscheinlichkeit" ist also gleich dem Anteil der zum Zeitpunkt i noch Lebenden an der Gesamtzahl der Untersuchungseinheiten, wobei zensierte Untersuchungseinheiten als "noch lebend" betrachtet werden.

Nun gilt folgendes: Wenn keine zensierten Beobachtungen auftreten, dann sind die "einfache Überlebenswahrscheinlichkeit" und die Kaplan-Meier-Überlebenswahrscheinlichkeit gleich.

Wir erkennen, daß die Kaplan-Meier-Überlebensfunktion eine spezifische Form der "einfachen Überlebensfunktion" ist. Ihre Besonderheit besteht darin, daß sie in einer spezifischen Form die zensierten Untersuchungseinheiten berücksichtigt.

Validierung der Almo-Ergebnisse:
Die Kaplan-Meier-Überlebensfunktion und ihre Standardabweichung stimmen - für verschiedene Beispiele gerechnet - bei Almo, SPSS, SAS und Statistica voll überein.

P40.4 Mittelwert der Überlebensdauer

Almo ermittelt die mittlere Überlebensdauer. In unserem Beispiel beträgt sie 105.387897 min. Almo verwendet folgende Formel (siehe dazu Blossfeld u.a., 1986, S. 45)

$$(4)\ M = \sum_{i=0}^{k-1} p_i (t_{i+1} - t_i)$$

i = der Index i bezeichnet den jeweiligen Zeitpunkt zu dem jemand gestorben ist.

k = Zahl der Zeitpunkte zu denen ein oder gleich mehrere "Todesfälle" auftreten. Das sind in obiger Tabelle der Kaplan-Meier-Überlebensfunktion diejenigen Zeitpunkte, zu denen die Kaplan-Meier-Überlebenswahrscheinlichkeit ausgegeben wurde. In unserem Beispiel ist k=5.

p_i = Kaplan-Meier-Schätzer der Überlebenswahrscheinlichkeit der Gestorbenen zum Zeitpunkt i.

t_i = Überlebenszeit der Gestorbenen zum Zeitpunkt i.

Beachte: die Summierung bezieht nur die "Toten" ein. Die zensierten Fälle werden übersprungen.

Diese Formel unterschätzt die mittlere Überlebensdauer, wenn zum letzten Zeitpunkt (in unserem Beispiel ist dies 120 min) niemand gestorben ist, wenn also die letzte Person zensiert ist - wie dies in unserem Beispiel der Fall ist.

Tritt dieser Fall auf, dann wird der Zeitraum zwischen letztem Gestorbenen und letztem Zensierten in die Berechnung miteinbezogen. Die Formel lautet dann:

(5) $M = \sum_{i=0}^{k-1} p_i \left(t_{i+1} - t_i\right) + p_k \left(t_z - t_k\right)$

p_k = Überlebenswahrscheinlichkeit des letzten Gestorbenen (in unserem Beispiel = 0.747685).
t_k = Überlebenszeit des letzten Gestorbenen (in unserem Beispiel = 92).
t_z = Überlebenszeit des letzten Zensierten (in unserem Beispiel = 120).

Würden wir in unserem Beispiel die mittlere Überlebensdauer nach Formel 4 berechnen, entstünde = 84.452712. Die Überlebensdauer würde also drastisch unterschätzt.

Almo berechnet zur mittleren Überlebensdauer, berechnet nach Gleichung 4, noch die Standardabweichung nach der Formel

(6) $\operatorname{Var}\left(M_4\right) = \frac{T}{T-1} \sum_{i=1}^{k-1} \frac{A_i^2 \cdot d_i}{n_i \cdot s_i}$

Die Standardabweichung ist die Wurzel aus $\operatorname{Var}\left(M_4\right)$.
$\operatorname{Var}\left(M_4\right)$ = Varianz des Mittelwert s, berechnet nach Gleichung 4.
i = definiert wie oben bei Gleichung 4
k = definiert wie oben bei Gleichung 4
T = Gesamtzahl der "Toten". In unserem Beispiel ist T=5.
d_i = definiert wie bei Gleichung 1
n_i = definiert wie bei Gleichung 1
$s_i = n_i - d_i$

(6a) $A_i = \sum_{j=i}^{k-1} p_j \left(t_{j+1} - t_j\right)$

p_j = definiert wie bei Gleichung 1.

Ist die letzte Person zensiert und wird die mittlere Überlebensdauer nach Gleichung 5 berechnet, dann wird in die Berechnung der Standardabweichung der Zeitraum vom letzten "Toten" bis zum letzten "Zensierten" miteinbezogen. Die Formel lautet dann (siehe Lawless 1982, S. 98)

(7) $\operatorname{Var}(M_5) = \sum_{i=1}^{k} \frac{A_i^2 . d_i}{n_i . s_i}$

Beachte: i läuft hier bis k.

(7a) $A_i = \sum_{j=i}^{k-1} p_j \left(t_{j+1} - t_j\right) + p_k \left(t_z - t_k\right)$

p_k, t_z, t_k sind bei Gleichung 5 definiert.
Var (M_5) = Varianz des Mittelwerts, berechnet nach Gleichung 5.

P40.5 Ausgabe von Zwischenergebnissen

Wird in der Programm-Maske die Optionsbox "Zwischenergebnisse ausgeben" geöffnet

```
Option: Zwischenergebnisse ausgeben
```

dann sieht man folgendes:

```
X  Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

                  Zwischenergebnisse ausgeben

   1              1=verschiedene Zwischenergebnisse ausgeben
                    z.B. Zwischenergebnisse aus Berechnung der
                    mittleren Lebensdauer, des log rank Tests etc.
                    ACHTUNG: Es kann sehr viel Text entstehen !!
                  0=keine Zwischenergebnisse
```

Wird "1" eingesetzt, dann werden verschiedene Zwischenergebnisse aus der Berechnung des Mittelwerts und der Standardabweichung ausgegeben. Dabei wird die Notation verwendet, mit der wir hier die Formeln geschrieben haben. Für unser Beispiel aus Altman gibt Almo aus

```
--------- Zwischenergebnisse: Berechnung des Mittelwerts ---------
   p     t+1   t
  ---    ---  ---
1.0000*(  30-    0) =  30.0000
0.9524*(  50-   30) =  19.0476
0.9048*(  51-   50) =   0.9048
0.8545*(  82-   51) =  26.4894
0.8011*(  92-   82) =   8.0109
                      --------
                       84.4527  (=Mittelwert ohne letztes zensiertes Objekt)
  p(k)  t(z)  t(k)
  ----  ----  ----
0.7477*( 120-   92) =  20.9352
                      --------
                      105.3879
---------------Berrechnung des Mittelwerts: Ende ------------------

Mittelwert der Ueberlebensdauer        = 105.387897

------ Zwischenergebnisse: Berechnung der Standardabweichung ------
       Bilden der Summe A = p*( t(+1) - t )

0.9524*(50-30) + 0.9048*(51-50) + 0.8545*(82-51) + 0.8011*(92-82) +
0.7477*(120-92)   <----- p(k) *( t(z) - t(k) )
A=75.3879 d=1 n=21 s=20

0.9048*(51-50) + 0.8545*(82-51) + 0.8011*(92-82) +
0.7477*(120-92)   <----- p(k) *( t(z) - t(k) )
A=56.3403 d=1 n=20 s=19

0.8545*(82-51) + 0.8011*(92-82) +
0.7477*(120-92)   <----- p(k) *( t(z) - t(k) )
A=55.4355 d=1 n=18 s=17

0.8011*(92-82) +
0.7477*(120-92)   <----- p(k) *( t(z) - t(k) )
A=28.9461 d=1 n=16 s=15 +

0.7477*(120-92)   <----- p(k) *( t(z) - t(k) )
A=20.9352 d=1 n=15 s=14

---------- Berechnung der Standardabweichung: Ende -----------
```

```
Standardabweichung                    =   6.124213
95% Vertauensintervall fuer Mittelwert =  93.384439 - 117.391355

--Mittelwert wuerde unterschaetzt werden, da letzte Untersuchungseinheit
--zensiert ist. Deswegen wurde der Zeitraum bis zu dieser in den Kalkuel
--des Mittelwerts und der Standardabweichung miteinbezogen (Handbuch, P40)
--unterschaetzter Mittelwert = 84.452712   Standardabweichung =  4.259646
```

P40.6 Median und Quantile

Almo gibt zusätzlich zum Mittelwert noch die Quantile für die 75 %ige, 50 %ige und 25 %ige Überlebenswahrscheinlichkeit aus; sowie deren Standardabweichung und 95 %-Vertrauensintervall

```
Mittelwert der Ueberlebensdauer        = 105.387897
Standardabweichung                    =   6.124213
95% Vertauensintervall fuer Mittelwert =  93.384439 - 117.391355

--Mittelwert wuerde unterschaetzt werden, da letzte Untersuchungseinheit
--zensiert ist. Deswegen wurde der Zeitraum bis zu dieser in den Kalkuel
--des Mittelwerts und der Standardabweichung miteinbezogen (Handbuch, P40)
--unterschaetzter Mittelwert = 84.452712   Standardabweichung =  4.259646

Quantile der Ueberlebensdauer     Stand.abwg.        95% Vertauensintervall
75%          =     92.00          nicht berechenbar  nicht berechenbar
50% (Median) = nicht berechenbar  nicht berechenbar  nicht berechenbar
25%          = nicht berechenbar  nicht berechenbar  nicht berechenbar
```

Validierung der Almo-Ergebnisse:
SPSS und Almo verwenden für Mittelwert und Standardabweichung dieselben Formeln. Die Ergebnisse aus verschiedenen Beispielen stimmen überein (SPSS gibt nur 2 Kommastellen aus), SAS verwendet die Formeln 4 und 6. Im Falle, daß die letzte Person ein "Toter" ist, stimmen (für verschiedene Beispiele gerechnet) die Ergebnisse exakt überein. Statistica gibt keine Mittelwerte und Standardabweichungen aus.
Der Vergleich mit den genannten anderen Statistiksystemen gilt ab deren Version von 1995 und später. Bei jeweils aktuellen Versionen können Änderungen vorgenommen worden sein.

P40.7 Vergleich von Gruppen

Ist die Überlebensfunktion von Männern und Frauen, die an Leberkrebs erkrankt sind verschieden? Sind die Überlebensfunktionen von Krebskranken, die 3 verschiedenen Therapien unterzogen werden, verschieden?
Hier geht es also darum zu untersuchen, ob sich die Kaplan-Meier-Überlebensfunktion mehrerer Gruppen signifikant unterscheiden.

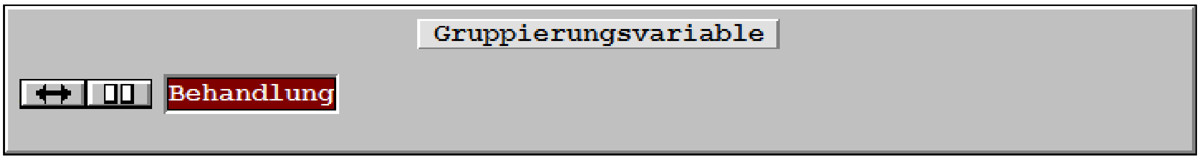
P40.7.1 Die Programm-Maske **Prog40m3**

Klicken Sie im Almo-Fenster auf den Knopf VERFAHREN. Es erscheint eine Übersicht über alle in Almo enthaltene Verfahren. Siehe dazu P0 und P39.2.1. Klicken Sie auf "Survival". Almo präsentiert Ihnen dann eine Liste der Maskenprogramme zum Thema "Survival". Laden Sie Prog40m3. Dieses Maskenprogramm ist identisch mit dem in Abschnitt P40.2.1 abgebildeten Prog40m1. Es kommt lediglich die Box für die Gruppierungs-Variable hinzu.

Box „Gruppierungsvariable“

Hier ist nicht nur 1 Variable als Gruppierungsvariable eingesetzt, sondern 2 Variable. Sie werden miteinander kombiniert. D.h. es werden alle Kombinationen aus den Ausprägungen der beiden Variablen gebildet. Jede Kombination ist eine eigene Gruppe. Ob das sinnvoll ist, muss der Benutzer entscheiden.

Im folgenden zeigen wir eine Analyse mit nur einer Gruppierungsvariablen. Das Programm liegt als Maskenprogramm unter dem Namen **KaplMei.Alm** vor. Es ist zu finden durch Klick auf den Knopf "alle Progs". Als Gruppierungsvariable wurde die Behandlungsart eingesetzt, mit der die Patienten therapiert wurden.



P40.7.3 Almo-Ausgabe zum Gruppenvergleich

Almo liefert folgende Ergebnisse

```
Fuer Analyse aus Datenvektor ausgewaehlte Variable

Gruppierungsvariable: V3  Behandlung
Zeitvariable:         V1  Ueberlebensdauer
Ereignisvariable:     V2  Ereignis

Zahl der eingelesenen Untersuchungseinheiten:      30


Erlaeuterung zu Spalte 2 der nachfolgenden Tabelle:
  Code: 0.00 = Ereignis eingetreten (z.B. gestorben)
        1.00 = Untersuchungseinheit zensiert
                (z.B. ueberlebt oder ausgeschieden)
               zusaetzlich durch einen Stern markiert

Erlaeuterung zu Spalte 3:
  Dies ist die Zahl der Untersuchungseinheiten,die
  nach dem jeweiligen Zeitpunkt noch vorhanden sind

Erlaeuterung zu Spalte 4:
  Dies ist der Anteil der Untersuchungseinheiten,
  die nach dem jeweiligen Zeitpunkt noch verbleiben
  - gemessen an denjenigen, die zum Beginn dieses
  Zeitpunkts noch vorhanden waren


Erlaeuterung zu Spalte 5:
  Dies ist der Kaplan-Meier-Schaetzer der Ueber-
  lebensfunktion. Die Werte entstehen durch Multi-
  plikation der Werte aus Spalte 4
```

Erlaeuterung zu Spalte 6:
Dies ist der Standardfehler der Kaplan-Meier-Ueberlebensfunktion

Erlaeuterung zu Spalte 7:
Dies sind die kumulierten Untersuchungseinheiten, die das Ereignis erfahren haben (z.B. gestorben sind)

Erlaeuterung zu Spalte 8:
Dies ist die einfache Wahrscheinlichkeit zum jeweiligen Zeitpunkt i noch zu leben. Entsteht aus Gesamtzahl der Untersuchungseinheiten minus den "Toten" (aus Spalte 7) dividiert durch die Gesamtzahl
Beachte: Zensierte Untersuchungseinheiten werden als Ueberlebende betrachtet

Gruppe 1: V3 Behandlung: 1 TherapieA

1	2	3	4	5	6	7	8
Zeit	Ereignis Code	nach dieser Zeit noch vor-handen	Anteil derer, die noch vor-handen	Kaplan-Meier-Ueberleb.fkt. = kumulierte Whrscheinlkt zu ueber-leben	Stand.fehler der Kaplan-Meier Ueberleb.fkt	Tote kumu-liert	einf. Whkeit zu ueber-leben
1.00	1.00*	8					
2.00	0.00	7	0.8750	0.875000	0.116927	1	0.8889
5.00	0.00	6	0.8571	0.750000	0.153093	2	0.7778
8.00	0.00	5	0.8333	0.625000	0.171163	3	0.6667
12.00	0.00	4	0.8000	0.500000	0.176777	4	0.5556
18.00	0.00	3	0.7500	0.375000	0.171163	5	0.4444
20.00	0.00	2	0.6667	0.250000	0.153093	6	0.3333
35.00	0.00	1	0.5000	0.125000	0.116927	7	0.2222
43.00	0.00	0	0.0000	0.000000	0.000000	8	0.1111

Mittelwert der Ueberlebensdauer = 17.875000
Standardabweichung = 5.138918
95% Vertauensintervall fuer Mittelwert = 7.802720 - 27.947280

Quantile der Ueberlebensdauer		Stand.abwg.	95% Vertauensintervall
75%	= 5.00	3.674235	0.000000 - 12.201500
50% (Median)	= 12.00	7.071068	0.000000 - 25.859293
25%	= 20.00	10.410331	0.000000 - 40.404250

Gruppe 2: V3 Behandlung: 2 TherapieB

1	2	3	4	5	6	7	8
Zeit	Ereignis Code	nach dieser Zeit noch vor-handen	Anteil derer, die noch vor-handen	Kaplan-Meier-Ueberleb.fkt. = kumulierte Whrscheinlkt zu ueber-leben	Stand.fehler der Kaplan-Meier Ueberleb.fkt	Tote kumu-liert	einf. Whkeit zu ueber-leben
2.00	0.00	8	0.8889	0.888889	0.104757	1	0.8889
10.00	0.00	7	0.8750	0.777778	0.138580	2	0.7778
12.00	0.00	6	0.8571	0.666667	0.157135	3	0.6667
12.00	1.00*	5					
12.00	1.00*	4					
35.00	0.00	3	0.7500	0.500000	0.186339	4	0.5556
43.00	0.00	2	0.6667	0.333333	0.184257	5	0.4444
72.00	0.00	1	0.5000	0.166667	0.149588	6	0.3333
72.00	1.00*	0					

Mittelwert der Ueberlebensdauer = 39.666667
Standardabweichung = 10.249300
95% Vertauensintervall fuer Mittelwert = 19.578039 - 59.755294

Quantile der Ueberlebensdauer		Stand.abwg.	95% Vertauensintervall
75%	= 12.00	7.071068	0.000000 - 25.859293
50% (Median)	= 35.00	17.329527	1.034127 - 68.965873

25%	=	72.00	26.028297	20.984539 - 123.015461

Gruppe 3: V3 Behandlung: 3 TherapieC

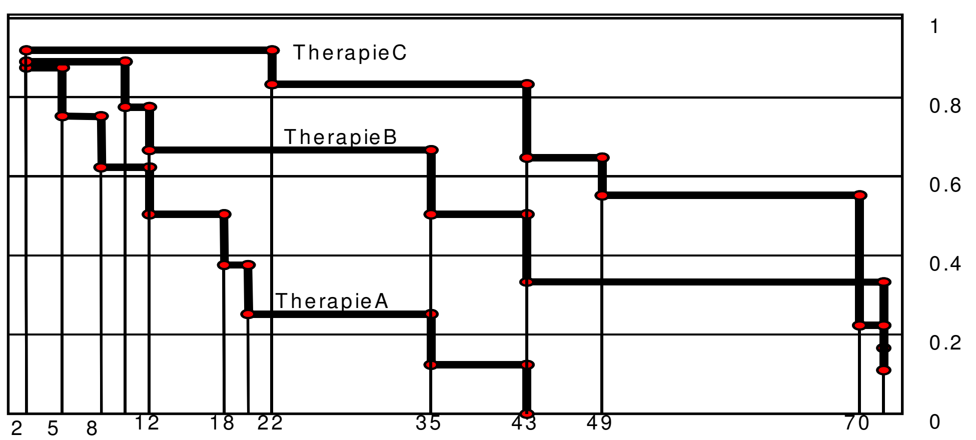
1	2	3	4	5	6	7	8
Zeit	Ereignis Code	nach dieser Zeit noch vor-handen	Anteil derer, die noch vor-handen	Kaplan-Meier-Ueberleb.fkt. = kumulierte Whrscheinlkt zu ueber-leben	Stand.fehler der Kaplan-Meier Ueberleb.fkt	Tote kumu-liert	einf. Whkeit zu ueber-leben
2.00	0.00	11	0.9167	0.916667	0.079786	1	0.9167
22.00	0.00	10	0.9091	0.833333	0.107583	2	0.8333
40.00	1.00*	9					
43.00	0.00	8					
43.00	0.00	7	0.8750	0.648148	0.142611	4	0.6667
49.00	0.00	6	0.8571	0.555556	0.149301	5	0.5833
69.00	1.00*	5					
70.00	0.00	4					
70.00	0.00	3					
70.00	0.00	2	0.6667	0.222222	0.135578	8	0.3333
72.00	0.00	1	0.5000	0.111111	0.103770	9	0.2500
85.00	1.00*	0					

Mittelwert der Ueberlebensdauer = 55.277778
Standardabweichung = 6.982423
95% Vertauensintervall fuer Mittelwert = 41.592228 - 68.963327

--Mittelwert wuerde unterschaetzt werden, da letzte Untersuchungseinheit
--zensiert ist. Deswegen wurde der Zeitraum bis zu dieser in den Kalkuel
--des Mittelwerts und der Standardabweichung miteinbezogen (Handbuch, P40)
--unterschaetzter Mittelwert = 53.833333 Standardabweichung = 6.893036

Quantile der Ueberlebensdauer			Stand.abwg.	95% Vertauensintervall
75%	=	43.00	16.172121	11.302643 - 74.697357
50% (Median)	=	70.00	8.541402	53.258851 - 86.741149
25%	=	70.00	7.016152	56.248342 - 83.751658

Kaplan-Meier Ueberlebensfunktion



```
Ereignis eingetreten in Gruppe 1 = 8 mal
Ereignis eingetreten in Gruppe 2 = 6 mal
Ereignis eingetreten in Gruppe 3 = 9 mal


Pauschaler logrank-Test auf Verschiedenheit aller Gruppen
---------------------------------------------------------
Chi-Quadrat =  11.305264
df          =   2
p           =   0.004015
Signifikanz (1-p)100 der Verschiedenheit= 99.598498 %


Pauschaler Wilcoxon-(Breslow)-Test auf Verschiedenheit aller Gruppen
--------------------------------------------------------------------
Chi-Quadrat =   9.451918
df          =   2
p           =   0.009183
Signifikanz (1-p)100 der Verschiedenheit= 99.081746 %


Chi-Quadrat-Werte der paarweisen Vergleiche der Gruppen
aus logrank-Test
  Oberer Wert ist Chi-Quadrat
  darunter p
  darunter die Signifikanz (1-p)*100 der Verschiedenheit der Gruppen
____________________________________________________________________

Gruppe 1 = Behandlung:TherapieA
Gruppe 2 = Behandlung:TherapieB
Gruppe 3 = Behandlung:TherapieC

            Gruppe 1    Gruppe 2
_________________________________


Gruppe 2    2.965274
            0.085210
           91.478985%
_________________________________

Gruppe 3   12.859612    0.332505
            0.000386    0.564047
           99.961401%  43.595262%
_________________________________


Chi-Quadrat-Werte der paarweisen Vergleiche der Gruppen
aus Wilcoxon- (Breslow-) Test
  Oberer Wert ist Chi-Quadrat
  darunter p
  darunter die Signifikanz (1-p)*100 der Verschiedenheit der Gruppen
____________________________________________________________________

Gruppe 1 = Behandlung:TherapieA
Gruppe 2 = Behandlung:TherapieB
Gruppe 3 = Behandlung:TherapieC

            Gruppe 1    Gruppe 2
_________________________________

Gruppe 2    1.744803
            0.186688
           81.331199%
_________________________________

Gruppe 3   10.532901    1.207561
            0.001357    0.271962
           99.864293%  72.803787%
_________________________________
```

P40.7.4 Berechnung des log rank- und Wilcoxon-Tests

In Almo werden folgende Formeln verwendet (siehe Kalbfleisch/Prentice, 1980, Kap1.4)

(8) Chi-Quadrat = $v' V^{-1} v$
mit Freiheitgrade df = g-1, dabei ist v ein Vektor, der aus g-1 Elementen besteht, wobei diese in folgender Weise gebildet werden:

$$(9)\ v_j = \sum_{i=1}^{k} w \cdot \left(d_{ij} - e_{ij}\right)$$

wobei

$$(10)\ e_{ij} = \frac{n_{ij.} d_i}{n_i}$$

w_i = ist beim log rank-Test = 1 und beim Wilcoxon-Test = n_i
k = definiert wie bei Gleichung 4.
g = Zahl der Gruppen
i = Zeitpunkt i, definiert wie bei (4)
j = der Index j läuft über die Gruppen von 1 bis g-1
d_{ij} = Zahl der Gestorbenen in Gruppe j zum Zeitpunkt i
d_i = Zahl aller Gestorbenen aus allen Gruppen zum Zeitpunkt i
n_{ij} = Risikomenge in Gruppe j zum Zeitpunkt i; das ist die Zahl der zum Zeitpunkt i-1 in Gruppe j noch Lebenden
n_i = gesamte Risikomenge aller Gruppen zum Zeitpunkt i. Siehe auch Gleichung 1.

e_{ij} ist der Erwartungswert der Toten (in Gruppe j zum Zeitpunkt i). Die Elemente des Vektors v sind also die Differenz der beobachteten Toten minus den erwarteten. Dies gilt für den log rank Test, bei dem w_i=1 ist.

Die einzelnen Elemente der Matrix V in Gleichung 8 werden nach folgender Formel gebildet:

$$(11a)\ V_{jj} = \sum_{i=1}^{k} \frac{w_i^2}{n_i^2} \frac{(n_i}{(n_i} \frac{n_{ij}}{} \frac{-n_{ij}^2}{-1)} \frac{)d_i}{} \frac{s_i}{}$$

die Außerdiagonalelemente jm

$$(11b)\ V_{jm} = \sum_{i=1}^{k} \frac{-w_i^2}{n_i^2} \frac{n_{ij}}{(n_i} \frac{n_{im}}{} \frac{d_i}{-1)} \frac{s_i}{}$$

m = der Index m läuft wie j über die Gruppen von 1 bis g-1.
$s_i = n_i - d_i$

Bei nur 2 Gruppen vereinfacht sich beim log rank Test Gleichung 8 auf

(12) Chi-Quadrat = $\frac{(D_1 - E_1)^2}{V_1}$

wobei

D_1 = Summe der Toten in Gruppe 1
E_1 = erwartete Tote in Gruppe 1
V_1 = Varianz

(12a) $D_1 = \sum_{i=1}^{k} d_{i1}$

(12b) $E_1 = \sum_{i=1}^{k} \frac{n_{i1} \cdot d_i}{n_i}$

(12c) $V_1 = \sum_{i=1}^{k} \frac{n_{i1}}{} \frac{n_{i2}}{n_i^2} \frac{d_i}{(n_i} \frac{s_i}{-1)}$

Siehe dazu Altman, 1991, S. 380.

Sind mehr als 2 Gruppen vorhanden, dann rechnet Almo noch zusätzlich den log-rank- und Wilcoxon-Test für jedes Gruppenpaar, wobei die jeweils anderen Gruppen negiert werden. Hier kann der Benutzer die von Almo ausgegebenen p-Werte mit g(g-1)/2 multiplizieren und damit die Bronferroni-Korrektur der Signifikanz von paarweisen Vergleichstests durchführen.

Validierung der Almo-Ergebnisse aus log rank- und Wilcoxon-Test.
Eine Vielzahl von Beispielen, die mit Almo, SAS und SPSS gerechnet wurden, erbrachten exakt dieselben Ergebnisse.

P40.7.4.1 Ausgabe von Zwischenergebnissen aus Berechnung des log rank-Test

Wurde „Zwischergeb=1;“ gesetzt, dann werden folgende Zwischenergebnisse zusätzlich ausgegeben.

```
------------- Zwischenergebnisse ---------------
Daten fuer log rank- und Wilcoxon-Test (sortiert)

   Zeit   noch da  Gestorben  Gruppe
           danach     (Zahl)

    1.00       8       0*         1
    2.00      11       1          3
    2.00       8       1          2
    2.00       7       1          1
    5.00       6       1          1
    8.00       5       1          1
   10.00       7       1          2
   12.00       6       1          2
   12.00       4       1          1
   12.00       5       0*         2
   12.00       4       0*         2
   18.00       3       1          1
   20.00       2       1          1
   22.00      10       1          3
   35.00       3       1          2
   35.00       1       1          1
   40.00       9       0*         3
   43.00       7       2          3
```

```
   43.00       2       1          2
   43.00       0       1          1
   49.00       6       1          3
   69.00       5       0*         3
   70.00       2       3          3
   72.00       1       1          2
   72.00       1       1          3
   72.00       0       0*         2
   85.00       0       0*         3
------------ Ende Daten (sortiert) ------------

Zeichenerklaerung fuer nachfolgende Tabelle
  n1, n2 = Risikomenge zu Begin der Zeit i in Gruppe 1, in Gruppe 2 etc.
  n      = Gesamtzahl der Risikomengen n1+n2+...
  d1, d2 = Zahl der "Toten" in Gruppe 1, in Gruppe 2 etc.
  d      = Gesamtzahl der "Toten"
  e1, e2 = Erwartungswert fuer Gruppe 1, fuer Gruppe 2 etc.

Tabelle zur Berrechnung des pauschalen logrank-/Wilcoxon-Tests
fuer alle Gruppen
---------------------------------------------------------------------
Zeit  n1   n2  n3    n  d1  d2  d3   d    e1      e2      e3
  1    8    9  12   29   0   0   0   0  0.0000  0.0000  0.0000
  2    8    9  12   29   1   1   1   3  0.8276  0.9310  1.2414
  5    7    8  11   26   1   0   0   1  0.2692  0.3077  0.4231
  8    6    8  11   25   1   0   0   1  0.2400  0.3200  0.4400
 10    5    8  11   24   0   1   0   1  0.2083  0.3333  0.4583
 12    5    7  11   23   1   1   0   2  0.4348  0.6087  0.9565
 12    4    4  11   19   0   0   0   0  0.0000  0.0000  0.0000
 18    4    4  11   19   1   0   0   1  0.2105  0.2105  0.5789
 20    3    4  11   18   1   0   0   1  0.1667  0.2222  0.6111
 22    2    4  11   17   0   0   1   1  0.1176  0.2353  0.6471
 35    2    4  10   16   1   1   0   2  0.2500  0.5000  1.2500
 40    1    3   9   13   0   0   0   0  0.0000  0.0000  0.0000
 43    1    3   9   13   1   1   2   4  0.3077  0.9231  2.7692
 49    0    2   7    9   0   0   1   1  0.0000  0.2222  0.7778
 69    0    2   5    7   0   0   0   0  0.0000  0.0000  0.0000
 70    0    2   5    7   0   0   3   3  0.0000  0.8571  2.1429
 72    0    2   2    4   0   1   1   2  0.0000  1.0000  1.0000
 72    0    0   1    1   0   0   0   0  0.0000  0.0000  0.0000
 85    0    0   0    0   0   0   0   0  0.0000  0.0000  0.0000

Varianz-Kovarianzmatrix V (ohne letzte Zeile/Spalte) fuer log rank Test
   2.2514  -0.8145
  -0.8145   3.9691

Vektor v von beobachteten minus erwarteten Toten (ohne letzte Zeile)
fuer log rank Test
  4.9675
 -0.6712
```

P40.8 Der Kein-Wert-Fall

Wenn ein Datensatz

a. in der Zeitvariablen
b. und/oder in der Ereignisvariablen
c. und/oder in einer der Gruppierungsvariablen

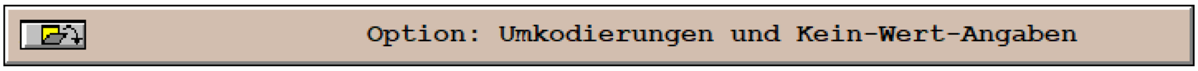
den Wert KEIN_WERT besitzt, dann wird der gesamte Datensatz aus der Analyse ausgeschlossen.

P40.9 Ausschluß von Untersuchungseinheiten

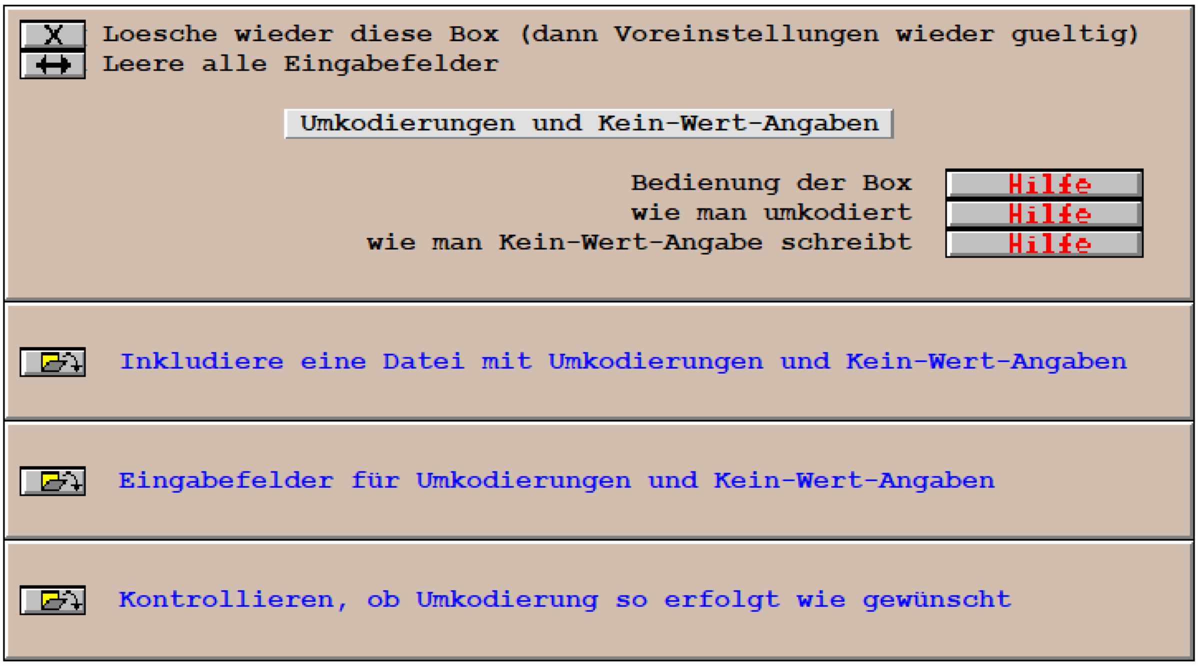
Wenn Sie bestimmte Datensätze ausschließen wollen, dann gehen Sie folgendermaßen vor:

1. Sie wollen z.B. bestimmte Zensierungen nicht berücksichtigen. Bei einer Studie über die Überlebensdauer nach Herztransplantationen kann es z.B. geschehen, daß einige Patienten durch einen Verkehrsunfall zu Tode kommen. Dies kann selbstverständlich nicht als Ereignis "tot" (mit Code=0) interpretiert werden, eher als das Ereignis "zensiert" (mit Code=1). Möglich ist es nun jedoch, dieses Ereignis mit 2 zu kodieren und diese 2 als Kein_Wert zu definieren - wodurch dieser Patient aus der Analyse ausgeschlossen wird.

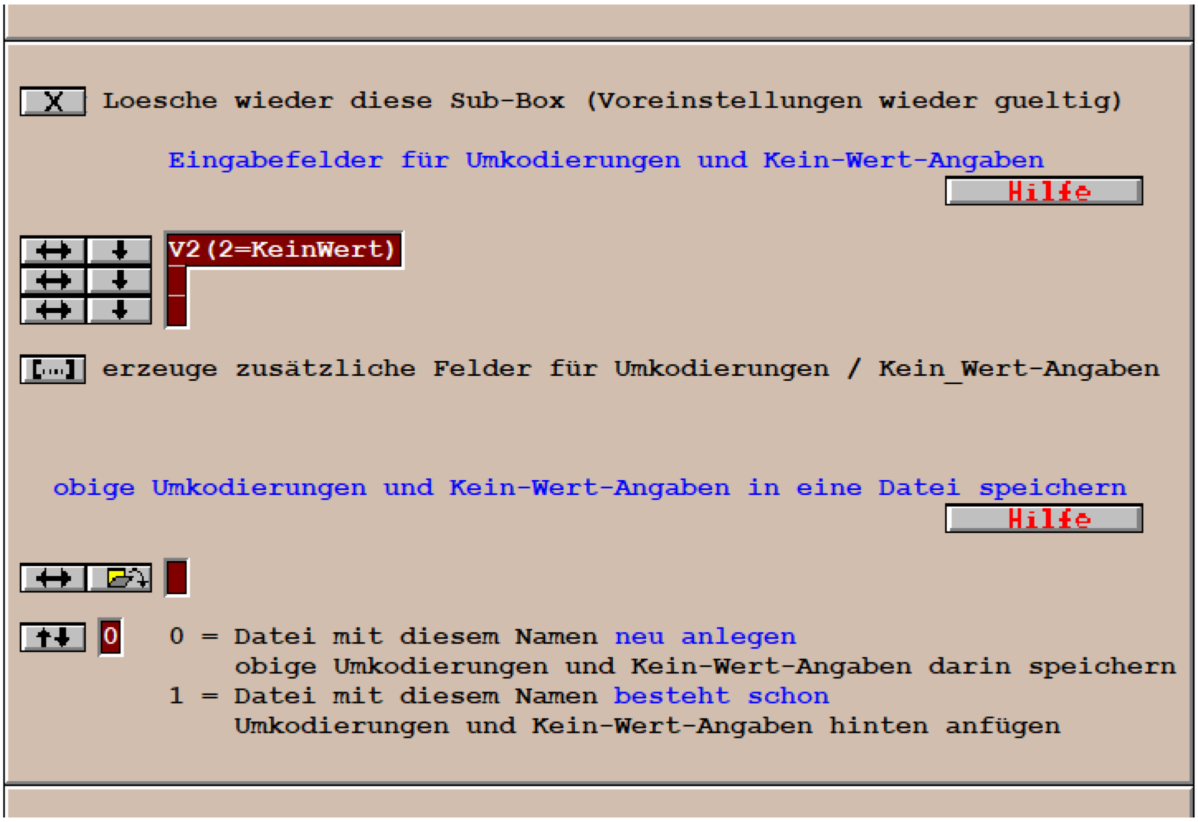
   Öffnen Sie in diesem Fall die Optionsbox "Umkodierungen und Kein-Wert-Angaben"



   Man sieht folgendes



Man öffnet nun die mittlerer Sub-Box



In das Eingabefeld wird die Kein-Wert-Anweisung geschrieben

```
V2 (2=Kein_Wert)
```

(V2 sei die Ereignisvariable). Almo schließt diesen Datensatz dann aus. Das Wort "Kein Wert" ist ein Almo-Schlüsselwort. Es muss als *ein* Wort geschrieben sein. Möglich sind z.B. `KeinWert` oder `Kein_Wert`. Der Unterstrich darf in Almo an beliebiger Stelle stehen. Er verbindet zwei Worte zu einem.

2. Wenn z.B. ein Patient kurz nach der Herztransplantation stirbt, so kann dies eine Folge des operativen Eingriffs und nicht spezifisch für die Transplantation sein. Sie können z.B. alle Fälle ausschließen wollen, bei denen es vor dem 4. Tag zum Tode oder einer Zensierung kommt. Schreiben Sie, wie oben beschrieben, in die Umkodierungs-Box

```
V1 (0 bis 4 = KeinWert).
```

(V1 sei die Zeitvariable). Almo schließt dann diesen Datensatz aus.

3. Die Patienten wurden nach dem Gruppierungsmerkmal "allgemeiner Gesundheitszustand" in folgende 3 Gruppen eingeteilt: gut, mittel, schlecht. Sie wollen nun die Patienten mit schlechtem Gesundheitszustand ausschließen. Schreiben Sie, wie oben beschrieben, in die Umkodierungs-Box

```
V3 (3=KeinWert).
```

(V3 sei die Gruppierungsvariable, 3 sei die Ausprägung "schlecht"). Almo schließt diesen Datensatz aus.

Den gleichen Effekt, den Ausschluß des Datensatzes, erreichen Sie durch folgende Anweisung in der Umkodierungsbox.

```
Wenn V1 (0 bis 4=Kein_Wert) gleich KeinWert
oder V2 gleich 2
oder V3 gleich 3
Dann gehezu Lese
Ende_Wenn
```

Literatur:

Altmann, Douglas G.: Practical statistics for medical research, Chapmann and Hall, London, 1991

Blossfeld, H.-P. / Hamerle A. / Mayer, K.U.: Ereignisanalyse, Campus, Frankfurt, 1986

Diekmann, A. / Mitter P.: Methoden zur Analyse von Zeitverläufen, Teubner, Stuttgart, 1984

Kalbfleisch J.D./Prentice R.L.: The stastical analysis of failure time data, Wiley N.Y:, 1980

Lawless, J.F.: Statistical models and methods for lifetime data, Wiley, New York, 1982.

P41 Cox-Regression

Die Cox-Regression wurde von Heinrich Potuschak programmiert. Die Anpassung an das Almo-System, die Grafik und die Programm-Masken, sowie die Übertragung in die Programmiersprache C wurde von Kurt Holm vorgenommen der auch diesen Text verfaßt hat. Joachim Gerich hat jene Programmteile an das Almo-System angepaßt, die es ermöglichen zeitabhängige Variable einzuführen.

Betrachten wir folgendes Beispiel:

Es soll die Frage untersucht werden: Warum ist die Überlebensdauer nach einer Herztransplantation unterschiedlich ?

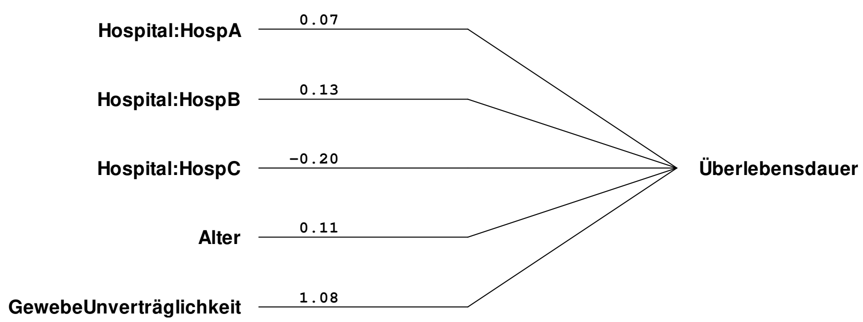
Als Ursachen der Überlebensdauer werden folgende Variable in das Modell eingeführt:

1. nominale unabhängige Variable: Hospital (in dem die Transplantation stattfand)

2. quantitative unabhängige Variable: Alter, GewebeUnvertraeglichkeit

Es wird also folgendes Modell gerechnet:

Regressionskoeffizienten



Wir haben auf die Linien dieses Flußdiagramms gleich die Regressionskoeffizienten geschrieben, die die Cox-Regression liefert.

P41.0 Überblick

Das Modell der Cox-Regression ist folgendes

(1) $S_t = (B_t)^p$

Häufig wird auch folgende Notation verwendet
(1a) $S(t) = (So(t))^P$

wobei

(2) $p = e^{(b_1 x_1 + b_2 x_2 + \ldots b_m x_m)}$

oder mathematisch äquivalent

(2a) $p = e^{b_1 x_1} * e^{b_2 x_2} * \ldots * e^{b_m x_m}$

Notation

S_t,S($_t$) = Survivalfunktion, Überlebensfunktion

Sie varriiert mit der Zeit t

Sie gibt die (kumulative) Überlebenswahrscheinlichkeit für die aufeinander folgenden Zeitpunkte nach dem Startzeitpunkt t_o (in unserem Beispiel: der Zeitpunkt der Transplantation) an. Siehe die ausführliche Erläuterung in Abschnitt ...

B_t, $S_o(t)$= Basis-Survivalfunktion, Basis-Überlebensfunktion. Sie variiert mit der Zeit t

Sie entspricht ungefähr der Konstanten in der linearen Regression. Wir können vorläufig definieren: Sie drückt die (kumulative) Überlebenswahrscheinlichkeit des „durchschnittlichen Probanden“ aus. Siehe die ausführliche Erläuterung in Abschnitt ...

x_1, x_2 ... x_m = unabhängige Variable. Sie sind um ihren Mittelwert zentriert.

b_1, b_2 ... b_m = Regressionskoeffizienten

e = die Zahl e

Für unser Beispiel würde die Survivalfunktion also folgendermaßen lauten:

(3) $S_t = (B_t)^p$

wobei

(4) $p = e^{(b_1 \cdot HA^* + b_2 \cdot HB^* + b_3 \cdot HC^* + b_4 \cdot A + b_5 \cdot G)}$

HA^*, HB^*, HC^* = Hospital A, B, C.
Dies sind 0-1 kodierte, mittelwertzentrierte Variable. Wenn ein Proband beispielsweise im Hospital C transplantiert wurde, dann ist für ihn
HC^*=1-0.2808, wobei 0.2808 der Mittelwert der 0-1 kodierten Variablen HC ist.
HB^{**}=0-0.3449, wobei 0.3449 der Mittelwert von HB ist,
HC^*=0-0.3743, wobei 0.3743 der Mittelwert von HC ist.

Das Cox-Modell kann äquivalent auch als Hazard-Funktion ausgedrückt werden.

(5) $h_t = H_t * e^{(b_1 x_1 + b_2 x_2 + \ldots b_m x_m)}$

oder mathematisch äquivalent

(6) $h_t = H_t * e^{b_1 x_1} * e^{b_2 x_2} * \ldots e^{b_m x_m}$

Häufig findet man auch folgende Nation
(6a) $h(t) = h_o(t) * e^{(b_1 x_1 + b_2 x_2 + \ldots + b_m x_m)}$

h_t, h(t)= Hazardfunktion
H_t, Ho(t)= Basis-Hazardfunktion

Die Formel für die Hazard-Funktion ist einfacher als die für die Survival-Funktion. Der Kalkül der Cox-Regression wird deswegen mit der Hazardfunktion entwickelt.

Auch unser Computerprogramm ist entsprechend ausgeführt. Wir werden später auf den Begriff der Hazardfunktion zurückkommen. Die Hazardfunktion kann kumuliert werden (durch Aufaddieren ihrer Werte über die aufeinanderfolgenden Zeitintervalle). Zwischen der kumulierten Hazardfunktion h^*_t und der Survivalfunktion S_t besteht folgender Zusammenhang:

$$h^*_t = -\ln(S_t)$$

$$S_t = e^{-h^*t}$$

h^*_t = kumulierte Hazardfunktion

Die kumulierte Hazardfunktion ist inhaltlich nicht sinnvoll definierbar. Zum Begriff der Hazardrate siehe auch P39.3 Spalte 15.

Der Zweck einer Cox-Regression ist es vor allem

1) den (unterschiedlich starken) Einfluß der unabhängigen Variablen auf die Überlebenswahrscheinlichkeit zu finden.
2) die Überlebenswahrscheinlichkeit von Probanden zu prognostizieren

Zu 1. Der (unterschiedlich starke) Einfluß der unabhängigen Variablen auf die Überlebenswahrscheinlichkeit wird durch die Regressionskoeffizienten b_1, b_2, ... b_m ausgedrückt bzw. - wie wir noch zeigen werden - durch die „Risiko-Koeffizienten" e^{b1}, e^{b2}, ... e^{bm}.
In unserem Beispiel ist der Regressionskoeffizient b_5 mit 1.08 am größten. Da bei der Cox-Regression ein positives Vorzeichen einen negativen Einfluß auf die Überlebenswahrscheinlichkeit ausdrückt, bedeutet dieser Regressionskoeffizient b_5, daß die Gewebe-Unverträglichkeit den ungünstigsten Einfluß auf die Überlebenswahrscheinlichkeit besitzt.

Zu 2. Prognose der Überlebenswahrscheinlichkeit. Bei der linearen Regression können wir für einen Probanden einen Wert in der abhängigen Variablen prognostizieren, wenn wir seine Werte in den unabhängigen Variablen kennen. Dies gilt auch für die Cox-Regression - jedoch mit dem Unterschied, daß wir nicht einen Wert prognostizieren, sondern für jeden Zeitpunkt nach t_o (in unserem Beispiel: der Zeitpunkt der Transplantation) die jeweilige Überlebenswahrscheinlichkeit. Betrachten wir die Prognose für einen 58-jährigen Probanden, d,h. für eine Person, die in der unabhängigen Variablen Alter den Wert 58 besitzt. Für die anderen unabhängigen Variablen setzen wir folgende Werte ein:

Hospital: Transplantation im Hospital A
Gewebeunverträglichkeit: 2.0 (das ist ein relativ hoher Wert)
Almo liefert folgendes Ergebnis:

Tage nach der Trans- plantation	Basis-Survival- Funktion Überlebenswahr- scheinlichkeit des "durchschnittlichen Probanden"	Überlebenswahr- scheinlichkeit des Probanden
10	0.9975	0.9734
25	0.9884	0.8834
29	0.9853	0.8539
39	0.9805	0.8112

```
 46          0.9679              0.7070
 47          0.9600              0.6476
 50          0.9325              0.4759
 51          0.9149              0.3886
 54          0.9058              0.3494
---------------------------------------
 60          0.8885              0.2845
---------------------------------------
 63          0.8667              0.2186
 64          0.8411              0.1589
 65          0.8300              0.1379
 66          0.8215              0.1236
 68          0.7984              0.0913
136          0.7765              0.0679
  .           .                   .
  .           .                   .
  .           .                   .
```

In der 3. Spalte stehen die Überlebenswahrscheinlichkeiten unseres Probanden. Seine Wahrscheinlichkeit, den 60. Tag nach der Transplantation lebend zu erreichen ist nur 0.2845. Die des „durchschnittlichen Probanden“ hingegen ist 0.8885.
Wir wollen den prognostizierten Wert errechnen. Wir ermitteln zunächst den Exponenten zu e in Gleichung 4.

```
y =  0.0701*(1-0.2808)      HA
    +0.1313*(0-0.3449)      HB
    -0.2014*(0-0.3743)      HC
    +0.1141*(58-45.7968)        Alter
    +1.0826*(2-1.1771)      G´unverträglichkeit

     = 2.3639
```

$p = e^{2.3639}$ ist = 10.6323429

Nun setzen wir in Gleichung 3 ein
$S_{60} = 0.8885^{10.6323} = 0.2845$

P41.1 Eingabe mit Programm-Maske Prog41m2

Klicken Sie im Almo-Fenster auf den Knopf VERFAHREN. Es erscheint eine Übersicht über alle in Almo enthaltene Verfahren. Klicken Sie auf den Eintrag „Ereignisanalyse". Siehe die ausführliche Darstellung in P39.2.1.

Prog41m2.Msk
Cox-Regression
mit vielen Optionen

Beispiel:
Es soll die Frage untersucht werden: Warum ist die Überlebens-dauer nach einer Herztransplantation unterschiedlich ?

Als Ursachen der Überlebensdauer werden folgende Variable in das Modell eingeführt:

1. nominale unabhängige Variable: Hospital (in dem die Trans-plantation stattfand)
2. quantitative unabhängige Variable: Alter, GewebeUnvertraeglichkeit

Es wird also folgendes Modell gerechnet:

```
                             b1=0.05
          Hospital_A  ───────────────────────┐
                             b2=0.14         │
          Hospital_B  ─────────────────────┐ │
                             b3=-0.19        │
          Hospital_C  ───────────────────────  Überlebenszeit
                             b4=0.11         │
               Alter  ─────────────────────┘ │
                             b5=0.06         │
GewebeUnvertraeglich  ───────────────────────┘
```

Folgende Besonderheiten werden noch in das Modell eingeführt:
a. Eine Gruppierungsvariable; z.B. der allgemeine Gesundheits-zustand. Für die Probanden mit gutem u. schlechtem Gesund-heitszustand werden verschiedene Überlebensfunktionen ermittelt.
b. Für Personen mit einem bestimmten 'Wertemuster' in den un-abhängigen Variablen (z.B. Patienten im Alter von 48 Jahren die im Hospital A operiert wurden) wird deren spezifische Überlebensfunktion ermittelt

Daten in Anlehnung an Crowley/Hu

Was ist ein Kurzprogramm ? --> Hilfe
Bedienung --> Hilfe

1 Speicher fuer x Variable Hilfe

2 Vereinbare Variable= 20 ;

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Datei der Variablennamen Hilfe

zeige = Namensdatei in Output zeigen
leer = nicht

Freie Namensfelder Hilfe

Name 1 =Haeufigkeit;
Name 2 =Ereignis:tot,zensiert; Ereignisvariable
Name 3 =Alter;
4 Name 4 =Antigen;
Name 5 =GewebeUnvertraeglichkeit;
Name 6 =Hospital:HospA,HospB,HospC;
Name 7 =Ueberlebensdauer; Zeitvariable
Name 8 =Zustand:gut,schlecht; Gruppierungsvariable

erzeuge zusätzliche Namensfelder

5

Datei aus der gelesen wird | Hilfe bei Datei-Problemen

"C:\Almo7\Testdat\Herzdat.fre"

frei | Format der Daten | Hilfe

V1:8 | der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

6 Wenn Dateiformat FIX oder Nicht-Standard-FREI | Hilfe

7

die Zeit- und Ereignis-Variable

Ueberlebensdauer | die Zeitvariable

Ereignis | die Ereignisvariable

0 | Code für "Ereignis hat stattgefunden"
z.B. Untersuchungseinheit ist gestorben

1 | Code für "zensiert"
z.B. Untersuchungseinheit hat ueber-
lebt oder ist ausgeschieden

8

die unabhängigen nominalen Variablen

Hospital

Hilfe

-1 | 0= die unabhängigen nominalen Variablen
werden in 0, 1 kodierte Dummies aufgelöst
-1= sie werden in 0, 1, -1 kodierte Dummies
aufgelöst

9

die unabhängigen quantitativen Variablen

Alter,GewebeUnvertraeglichkeit

10 Gruppierungsvariable

11

Häufigkeitsvariable | Hilfe

Haeufigkeit

diese Variable enthält die Häufigkeiten
je "Todes"- bzw. Zensierungs-Zeitpunkt

12 Option: Ein- und Ausschliessen von Untersuchungseinheiten

13 Option: Umkodierungen und Kein-Wert-Angaben

14 Option: Untersuchungseinheiten ganzzahlig gewichten

15 Option: Wertemuster

16 Optionen: Zusätzliche Ausgabe

0 — 1= Ausgeben der Analysevariable für alle Untersuchungseinheiten
0= nicht

1 — 1= Ausgeben der Basis-Survivalfunktion und der Basis-Hazardfunktion; wenn Wertemuster angegeben,dann auch Survivalfunktion u. Hazardrate je Wertemuster
0= nicht

17 Grafik-Optionen

Almo — Almo = obige Grafiken als Almo-Grafik ausgeben
0 = nicht

0 — 1= obige Grafiken zusätzlich als einfache Liniendiagramme mit Textzeichen ausgeben
0= nicht

1 — 1= in Almo-Grafik auch den linken Teil der Überlebenskurve von Zeit=0 und Überleb.whschkeit p=1.0 zeigen
0= nicht

18 spezielle Grafik-Optionen

19 Programmende

Erläuterungen zu den Eingabeboxen:

Wir werden zuerst ein Programm ohne Optionen rechnen und deswegen im folgenden nur jene Boxen erläutern, die in unserem Rechenbeispiel verwendet werden. Die Optionsboxen werden wir später in Abschnitt P41.4 erklären.

Eingabebox 1 und 2: Speicher für x Variable, Weitere Vereinbarungen. Siehe P0.1 und P0.2

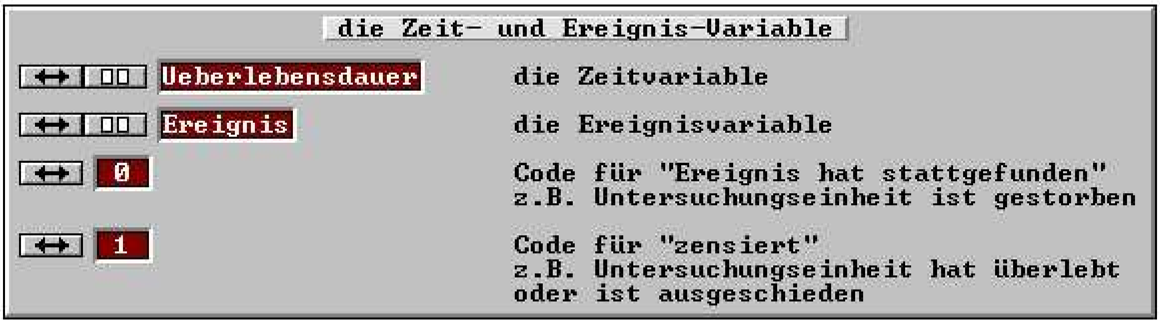
Box 3: Datei der Variablennamen. Siehe P0.3.

Box 4: Freie Namensfelder

Hier geben Sie den Variablen, die in die Analyse eingehen, einen Namen. Natürlich wäre es auch möglich diese Namen in eine Datei einzuschreiben und den Namen dieser Datei in Box 3 anzugeben.

Box 5 und 6: Datei aus der gelesen wird. Siehe P0.4

Box 7: Die Zeit- und Ereignisvariable.



Die Zeitvariable ist die abhängige Variable, also die Zeitdauer bis zum Eintritt des Ereignisses. Die Ereignisvariable enthält die Codes für "Ereignis hat stattgefunden" bzw. "Ereignis hat nicht stattgefunden".

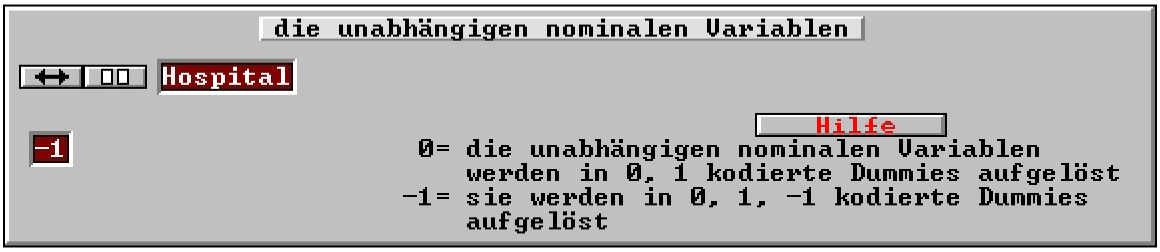
"Ereignis hat stattgefunden" ist in unserem Beispiel der Tod des Patienten.

"Ereignis hat nicht stattgefunden" bedeutet in unserem Beispiel, daß der Patient bis zum Beobachtungsende überlebt hat oder aus der Beobachtung ausgeschieden ist.

Hat bei einer Untersuchungsperson das Ereignis nicht stattgefunden, dann sagt man auch "die Person ist zensiert". Siehe dazu die ausführliche Darstellung in P39 oder P40.

Box 8: Die unabhängigen nominalen Variablen

Box 9: Die unabhängigen quantitativen Variablen



In diesen beiden Boxen werden die unabhängigen Variablen (von denen angenommen wird, daß sie die Überlebensdauer bestimmen) mitgeteilt. Die nominalen Variablen werden von Almo in Dummy-Variable aufgelöst, wobei entweder die 0-1-Kodierung oder die eher vorzuziehende 0,1, -1 -Kodierung verwendet wird. Siehe dazu im Almo-Handbuch „Das Allgemeine Lineare Modell“, Abschnitt P20.3 und im Handbuch, Teil 4, Abschnitt P22.2.3.1.

Bei der 0,1 -Kodierung wird die letzte Dummy eliminiert. Wir könnten auch sagen: der Regressionskoeffizient der eliminierten Dummy wird auf .0 gesetzt. Bei der 0,1,-1 -Kodierung summieren sich die Regressionskoeffizienten zu .0.

Wir empfehlen die 0,1,-1 -Kodierung, da hier auch für die eliminierte Dummy Koeffizienten (Beta, Standardfehler etc.) berechnet werden können. Sind die nominalen Variablen dichotom, dann hat allerdings die 0,1-Kodierung einige Vorteile.

Die 0,1 -Kodierung wird in der Literatur auch "Indikator-Kodierung" (bei SPSS: indicator-contrast) und die 0,1,-1 -Kodierung "Effekt-Kodierung" (bei SPSS: deviation-contrast) genannt.

Die Ergebnisse müssen bei den beiden Kodierungsweisen unterschiedlich interpretiert werden. Bei der 0,1-Kodierung ist die letzte Dummy die Referenzkategorie, bei 0,1,-1-Kodierung ist dies der Durchschnitt. Betrachten wir von den verschiedenen Ergebnissen, die Almo ausgibt das "relative Risiko".
Bei der 0,1-Kodierung der unabhängigen nominalen Variablen Hospital erhalten wir folgendes Ergebnis

	relatives Risiko
HospA	31,19%
HospB	39,46%

Für die Referenzkategorie HospC wird nichts ausgegeben. Die Ergebnisse sind so zu interpretieren: Das Risiko zu sterben ist in HospA um 31,19% und in HospB um 39,46% größer als in HospC.

Bei der 0,1,-1-Kodierung entsteht folgendes Ergebnis:

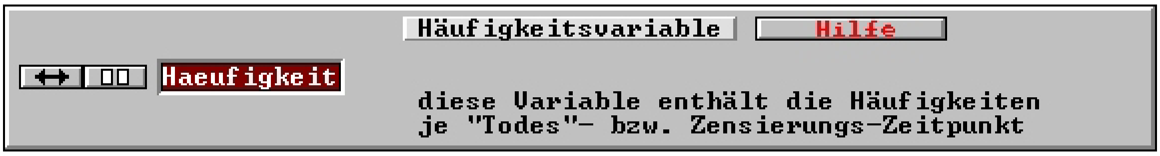
	relatives Risiko
HospA	7,26%

```
HospB            14,03%
HospC           -18,24%
```

Die Interpretation ist nun folgende: Im Vergleich zur durchschnittlichen Wirkung der 3 Hospitäler ist das Risiko zu sterben für jene die in HospA transplantiert wurden um 7,26% größer; bei HospB um 14,03% größer und bei HospC um 18,24% kleiner.

Box 10: Gruppierungsvariable. Siehe später in Abschnitt P41.4

Box 11: Häufigkeitsvariable



Betrachten wir einen Ausschnitt aus unseren Beispieldaten

V1 Häufigkeit	V2 Ereignis	V3 Alter	V7 Überlebensdauer
1	1	54	15
15	1	40	3
.	.	.	.
.	.	.	.

1 Person im Alter von 54 Jahren ist nach 15 Tagen gestorben.

15 Personen im Alter von 40 Jahren sind nach 3 Tagen gestorben.

V1 enthält die Häufigkeit der Personen, die dasselbe Ereignis nach derselben Überlebensdauer erfahren haben und in den unabhängigen Variablen dieselben Werte besitzen.

Normalerweise würden die 15 Personen, die nach 3 Tagen gestorben sind, als 15 Datensätze in der Datei enthalten sein. Almo erlaubt es jedoch auch diese als nur einen Datensatz zu verarbeiten. Dann muß aber eine Variable mitgeführt werden, die die Häufigkeit enthält und diese muß als "Häufigkeitsvariable" angegeben werden. Zu große Häufigkeiten sollten nicht auftreten, da dies nicht modellgemäß ist.

Stellt jeder Datensatz eine einzelne Person dar, besteht also keine "Häufigkeitsvariable", dann löschen Sie den Eintrag in dieser Box.

Box 12: Option: Ein- und Ausschliessen von Untersuchungseinheiten. Siehe P0.7
Box 13: Option: Umkodierungen und Kein-Wert-Angaben. Siehe P0.5
Box 14: Option: Untersuchungseinheiten ganzzahlig gewichten. Siehe P0.8
Box 15: Option: Wertemuster. Siehe die Darstellung im folgenden Abschnitt P41.4
Box 16:Optionen: Zusätzliche Ausgaben. Siehe im folgenden Abschnitt P41.4
Box 17 und 18: Grafik-Optionen. Spezielle Grafikoptionen. Siehe Abschnitt P41.4

P41.3 Ausgabe

Almo liefert aus Prog41m2.Msk folgende Ergebnisse:

```
Fuer Analyse aus Datenvektor ausgewaehlte Variable

Unabh. nominale Variable: V6    Hospital:  HospA, HospB, HospC
Unabh. quantitat. Variab: V3    Alter
                          V5    GewebeUnvertraeglichkeit
Zeitvariable:             V7    Ueberlebensdauer
Ereignisvariable:         V2    Ereignis

Beachte: Fuer die unabh. nominalen Variablen
wird die 0,1,-1 Dummy-Kodierung verwendet.

Zahl der eingelesenen Datensaetze = 64
(ohne Beruecksichtigung der Gewichtung)


                                         ******************** MITTEILUNG
         von den 64 Datensaetzen wurden 3 aus der Analyse ausgeschlossen
                   da diese zensiert sind bevor das 1. Ereignis eintritt

Keine Gruppierungsvariable

Probanden (gewichtet)                  374
Zahl der Todeszeitpunkte                27
Haeufigkeit "Tote" (gewichtet)         179
Haeufigkeit Zensierte (gewichtet)      195
```

******* Erläuterung:**
Es wurden 64 Datensätze eingelesen. Jeder Datensatz entsprach in unserem Beispiel mehreren Probanden. Wird jeder Datensatz mit der Häufigkeitsvariable gewichtet, so sind insgesamt 374 Probanden für die Analyse vorhanden, wobei 179 gestorben sind und 195 zensiert wurden (d.h. überlebt oder ausgeschieden sind).

```
Mittelwerte der unabhaengigen Variablen
---------------------------------------
Nominale Variable
  die Mittelwerte der Dummies der nominalen Variable
  sind gleich den Anteilswerten

  Hospital:HospA         0.2807
  Hospital:HospB         0.3449
  Hospital:HospC         0.3743

Quantitative Variable
           Alter        45.7968
GewebeUnvertraeg         1.1771
```

******* Erläuterung:**
Die Mittelwerte der quantitativen Variablen Alter und GewebeUnvertraeglichkeit werden mitgeteilt. Die nominale Variable "Hospital" wird von Almo in 3 Dummies aufgelöst. Dabei wird die 0,1,-1 Dummy-Kodierung verwendet. Möglich wäre auch die 0,1 Dummy-Kodierung. Diese so kodierten Variablen besitzen auch einen Mittelwert. Die Mittelwerte der Dummies sind gleich den Anteilswerten. So besitzt HospA einen Mittelwert von 0.2807. Das bedeutet: 28.07 % der Probanden wurde in Hospital A transplantiert.

```
Konvergenz nach 5 Schritten, maximaler Fehler = 0.000214
Determinante = 1.466520e+009

log-likelihood(0)    = -935.1313
```

```
log-likelihood(4)     = -861.5488
logratio              =  147.1649  p= 0.0000  (1-p)*100= 100.0000
Wald-Statistik        =  110.0006
Score-Statistik       =  269.1684

                                Stand.                           Signifik
Variable              Beta      fehler       z        p     (1-p)100
-------------------------------------------------------------------
Hospital:HospA       0.070100   0.117376   0.5972  0.5504   44.9628
Hospital:HospB       0.131254   0.107986   1.2155  0.2242   77.5787
Hospital:HospC      -0.201354   0.115733  -1.7398  0.0819   91.8079
Alter                0.114110   0.014266   7.9989  0.0000  100.0000
GewebeUnvertraegl    1.082602   0.167250   6.4730  0.0000  100.0000
-------------------------------------------------------------------

                     part.    "Risiko"   relatives
Variable               r      exp(Beta)  Risiko in %
----------------------------------------------------
Hospital:HospA       0.000000   1.0726     7.26
Hospital:HospB       0.000000   1.1403    14.03
Hospital:HospC      -0.023433   0.8176   -18.24
Alter                0.182047   1.1209    12.09
GewebeUnvertraegl    0.146060   2.9524   195.24
----------------------------------------------------
```

******* Erläuterung:**

Spalte 1: Beta (Regressionskoefizienten)
In der Spalte "Beta" befinden sich die Regressionskoeffizienten. Der Regressionskoeffizient des Alters beträgt 0.11411. Der der GewebeUnvertraeglichkeit ist mit 1.0826 der höchste. Die Regressionskoeffizienten der Dummies der nominalen Variablen "Hospital" summieren sich zu 0. Dies ist nur bei der 0,1,-1 Kodierung der Fall.

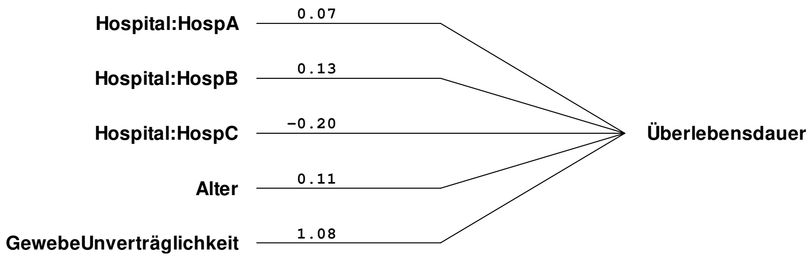
Sind die Variablen in unterschiedlichen Einheiten gemessen, dann sind ihre Regressionskoeffizienten nicht vergleichbar. Dies gilt in unserem Beispiel für Alter (in Jahren) und Gewebe-Unverträglichkeit (ein Index). Die Regressionskoeffizienten der Dummies einer nominalen Variablen sind allerdings vergleichbar. Zum Vergleich zwischen allen Variablen eignet sich der z-Wert und noch besser das Risiko (siehe unten).

Die Survival-Funktion für unser Beispiel lautet also:

$S_t = (B_t)^p$
wobei
$p = e^y$
und

```
y =   0.07*(HA-0.28)        Hospital A
    + 0.13*(HB-0.34)        Hospital B
    - 0.20*(HC-0.38)        Hospital C
    + 0.11*(Alter-45.79)    Alter
    + 1.08*(GU-1.17)        Gewebe-Unverträglichkeit (GU)
```

Grafisch als Flussdiagramm dargestellt:



Ein negativer Koeffizient bedeutet, daß sich die Überlebenswahrscheinlichkeit erhöht, wenn sich die unabhängige Variable um 1 Einheit erhöht; und umgekehrt bedeutet ein positiver Koeffizient, daß sich die Überlebenswahrscheinlichkeit verringert, wenn sich die unabhängige Variable um 1 Einheit erhöht. Der Regressionskoeffizient sowohl des Alters als auch der GewebeUnvertraeglichkeit ist positiv. Wenn sich diese Variable erhöhen, dann verringert sich die Überlebenswahrscheinlichkeit.

Diese Interpretation der Regressionskoeffizienten gilt auch für die Dummies der nominalen Variablen des "Hospitals". Wird die Dummy um 1 Einheit von 0 auf 1 erhöht, dann bedeutet dies, dass der Proband in dem betreffenden Hospital transplantiert wurde.
Wir sehen, daß HospC mit –0.201 einen Regressionskoeffizienten besitzt, der zeigt, daß dort die Transplantationen am erfolgreichsten durchgeführt werden.

Spalte 7 und Spalte 8: Risiko
Die relative Stärke, mit der die Variablen die Überlebenswahrscheinlichkeit bestimmen, kann besser in den beiden letzten Spalten abgelesen werden. Sie sind überschrieben mit
"Risiko exp(Beta)" und mit "relatives Risiko".

Das "Risiko" exp(Beta) ist in folgender Weise zu interpretieren. Ist exp(Beta) = 1 dann bedeutet dies, dass durch die unabhängige Variable die Hazardrate nicht verändert wird, d.h. dass sie keinen Einfluß auf die Überlebenswahrscheinlichkeit besitzt.

Betrachten wir die GewebeUnvertraeglichkeit. Sie besitzt ein Risiko exp(Beta) von 2.9524. Das bedeutet, dass bei einer Zunahme der GewebeUnvertraeglichkeit um 1 Einheit, die Hazardrate um den Faktor 2.9524 zunimmt. Das wiederum bedeutet, dass das Risiko zu sterben größer wird oder umgekehrt: die Überlebenswahrscheinlichkeit abnimmt.

Die Hazardrate ist in folgender Weise definiert:

Betrachten wir dazu ein Beispiel: Unsere Beispieldaten weisen folgende Zeitintervalle auf:

Nr	Zeit	Risiko menge	Tote
1	10	374	2
2	25	348	6
3	29	333	2
4	39	328	3
5	46	324	8
6	47	316	5
7	50	309	16

```
  8   51  293  10
  9   54  283   5
 10   60  278   9
  .    .    .   .
  .    .    .   .
  .    .    .   .
```

Das Intervall 4 beispielsweise reicht vom Tag 29 bis zum Tag 39 (nach der Transplantation).

> Die Hazardrate ist das Risiko eines Probanden, der in das Intervall eintritt, im Verlauf einer Zeiteinheit in diesem Intervall, also innerhalb eines Tages, zu terminieren.

Siehe dazu unsere Darstellung in P39.3 Spalte 15

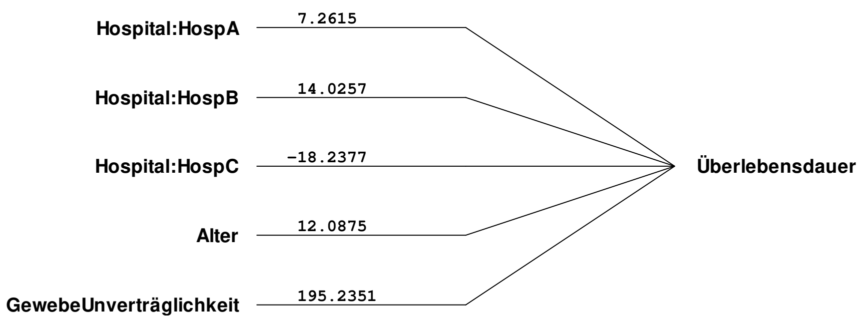
Das "relative Risiko" ergibt sich dann sehr einfach als

relative Risiko = 100 * (exp(Beta) - 1)

Betrachten wir wieder die GewebeUnvertraeglichkeit. Sie besitzt ein relatives Risiko von 195.24 %. D.h. daß bei einer Zunahme der GewebeUnvertraeglichkeit um 1 Einheit, das Risiko zu sterben zunimmt, genauer: die Hazardrate um 195 % zunimmt.

Flussdiagramm des relativen Risikos:

Relatives Risiko (in %)



Betrachten wir die Dummy-Variable HospC. Sie besitzt ein relatives Risiko von 100*(0.8176-1) = -18.24. Da das Vorzeichen negativ ist, bedeutet dies, dass das Risiko zu sterben verringert wird, wenn der Proband in Hospital C transplatiert wird, genauer: dass die Hazardrate um 18.24 % kleiner wird. Das numerische Ergebnis ändert sich wenn anstelle der 0,1-Kodierung, die 0,1,-1-Kodierung verwendet wird. Siehe dazu unsere ausführliche Darstellung in P41.4, Erläuterung zu Box 5.

Siehe zum Begriff des "Risikos" auch unsere Darstellung zur Logit-Analyse in Abschnitt P22.2.3.2. Dort wird der Begriff in derselben Weise verwendet.

Spalte 2: Standardfehler
In der 2. Spalte der obigen Tabelle wird der Standardfehler der Regressionskoeffizienten ausgegeben.

Spalte 3 und 4 und 5: z-Wert, p-Wert, Signifikanz
Der z-Wert ergibt sich gemäß

z = Beta / Standardfehler (zweiseitig zu testen)

Die "Wald"-Statistik, die z.B. in SPSS ausgegeben wird, ist z*z

Betrachten wir wieder die GewebeUnvertraeglichkeit.
Ihr z-Wert ist = 6.4730
der p-Wert = 0.0000
die Signifikanz 100*(1-p) = 100

Selbstverständlich ist der p-Wert nicht exakt 0 und die Signifikanz nicht exakt 100. Diese Werte entstehen durch runden.

Spalte 6: Partielle Korrelation
Aus der z-, bzw. Wald-Statistik läßt sich ein partielles r errechnen. Da die Cox-Regression eine nichtlineare Regression ist, kann dieses r nicht als ein PRE-Koeffizient interpretiert werden, was seine Sinnhaftigkeit schmälert.

Korrelationen zwischen den Kovariablen
(in der Hauptdiagonale stehen die Standardabweichungen)

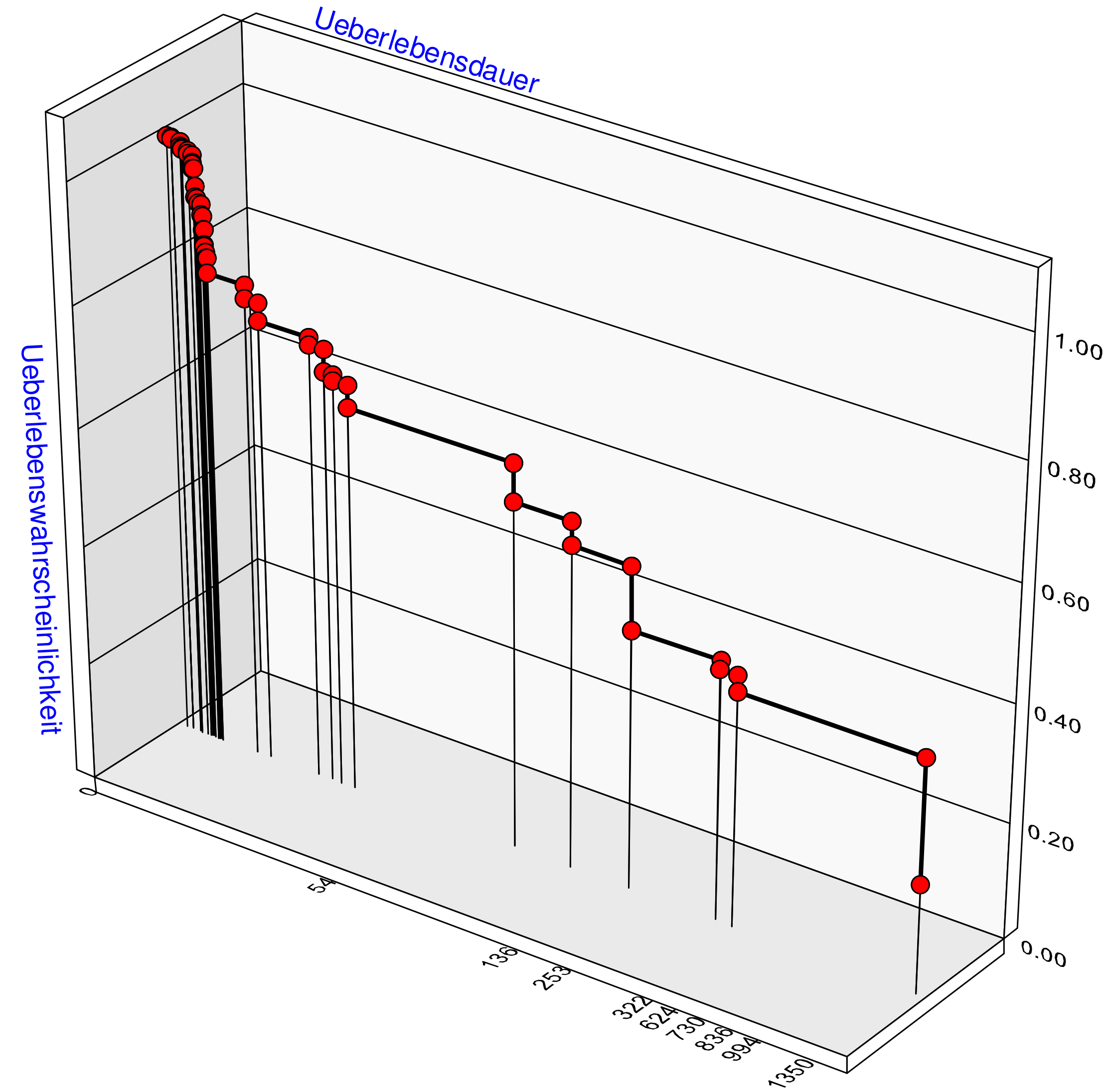
		Hospita HospA V6-1	Hospita HospB V6-2	Alter V3	GewebeU V5
Hospital HospA	V6-1	0.1174	-0.4751	-0.1596	0.0585
Hospital HospB	V6-2	-0.4751	0.1080	0.0908	-0.1007
Alter	V3	-0.1596	0.0908	0.0143	0.0247
GewebeUn	V5	0.0585	-0.1007	0.0247	0.1672

Survival-Funktionen

Nr	Zeit	Risiko menge	Tote	Basis Surviv Funkt.	Stand. Fehler.	Basis Hazard Funkt.
1	10	374	2	0.9975	0.0018	0.0025
2	25	348	6	0.9884	0.0042	0.0091
3	29	333	2	0.9853	0.0048	0.0032
4	39	328	3	0.9805	0.0057	0.0048
5	46	324	8	0.9679	0.0077	0.0129
6	47	316	5	0.9600	0.0088	0.0083
7	50	309	16	0.9325	0.0122	0.0290
8	51	293	10	0.9149	0.0141	0.0191
9	54	283	5	0.9058	0.0151	0.0100
10	60	278	9	0.8885	0.0169	0.0193
11	63	269	10	0.8667	0.0188	0.0248
12	64	259	10	0.8411	0.0208	0.0300
13	65	249	4	0.8300	0.0217	0.0134
14	66	245	3	0.8215	0.0223	0.0102
15	68	242	8	0.7984	0.0239	0.0285
16	136	213	7	0.7765	0.0253	0.0279
17	161	206	9	0.7473	0.0270	0.0382
18	253	174	4	0.7342	0.0278	0.0178
19	280	170	11	0.6978	0.0296	0.0508
20	297	159	3	0.6879	0.0301	0.0143
21	322	151	10	0.6525	0.0316	0.0529
22	624	91	8	0.5903	0.0353	0.1002
23	730	77	4	0.5483	0.0379	0.0737
24	836	72	9	0.4426	0.0425	0.2142
25	994	40	1	0.4285	0.0438	0.0323
26	1024	39	2	0.4009	0.0458	0.0666
27	1350	24	10	0.1872	0.0386	0.7618

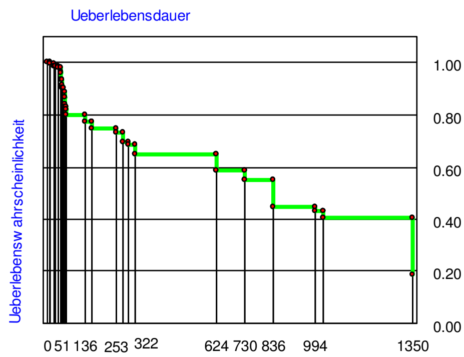
```
Almo liefert folgendes Liniendiagramm zur Basis-Überlebensfunktion
```

Basis-Ueberlebensfunktion



Wenn Sie im Grafik-Editor auf der linken Seite auf den Knopf "Diverse Positionen" klicken und in der dann von Almo angebotenen Auswahl verschiedener Positionen der Grafik die dritte wählen, dann erhalten Sie die Basis-Überlebensfunktion in folgender "Draufsicht" (hier verkleinert):

Basis-Ueber-
lebensfunktion



Erläuterung:

Spalte 5: Basis-Survivalfunktion
Wir sagten, dass der wesentliche Zweck der Cox-Regression darin besteht, (1) die relative Einflußstärke der unabhängigen Variablen abzuschätzen und (2) eine Prognose der Überlebenschance von Probanden abzugeben. Für die Prognose benötigen wir die Basis-Survivalfunktion. Betrachten wir noch einmal das Modell der Cox-Regression:

$$S_t = (B_t)^p$$

wobei

$$p = \mathrm{e}^y$$

und

$$y = b_1 x_1 + b_2 x_2 + ... + b_m x_m$$

B(t) = das ist die Basis-Survivalfunktion.

Die Regressionskoeffizienten b1, b2, ..., bm sind uns bekannt.
Die Formel für unser Beispiel lautet also

$$S_t = (B_t)^p$$

wobei

$$p = \mathrm{e}^y$$

```
y =   0.07*(HA-0.28)        Hospital A
    + 0.13*(HB-0.34)        Hospital B
    - 0.20*(HC-0.38)        Hospital C
    + 0.11*(Alter-45.79)    Alter
    + 1.08*(GU-1.17)        Gewebe-Unverträglichkeit (GU)
```

Wenn wir nun die Werte eines Probanden in den unabhängigen Werten kennen, dann können wir seine Überlebenschance prognostizieren - und dies für die aufeinander folgenden Zeitpunkte an denen mindestens ein Proband terminierte.

```
Tage nach der Transplantation
-----------------------------
              10
              25
              29
              39
              46
               .
               .
               .
```

Man beachte den Unterschied: Bei der gängigen linearen Regression wird 1 Wert prognostoziert. Bei der Cox-Regression wird eine Folge von Werten prognostiziert.

Um diese Prognose leisten zu können, benötigen wir die Basis-Survivalfunktion.

Betrachten wir einen Probanden, der

1. im Hospital A transplantiert wurde

```
                      2. im Alter von 58 Jahren
                      3. mit einer Gewebe-Unverträglichkeit von 2

Wir setzen also in obiger Gleichung HC    =  1   HA = HB = 0
                                    Alter = 58
                                    GU    =  2
```

Die Basis-Survivalfunktion ist in obiger Almo-Ausgabe ablesbar. Wir können damit die gewünschten Prognosewerte errechnen:

Tage nach der Transplantation	Basis-Survivalfunktion	prognostizierte Überlebens-wahrscheinlichkeit
10	0.9975	0.9734
25	0.9884	0.8834
29	0.9853	0.8539
39	0.9805	0.8112
46	0.9679	0.7070
.	.	.
.	.	.
.	.	.

Die Basis-Survivalfunktion ist, wie in P41.... schon ausgeführt als die "Survival-Funktion des durchschnittlichen Probanden" interpretierbar. Das ist der Proband, der in allen unabhängigen Werten den Mittelwert besitzt. In unserem Beispiel shen wir nun, daß die Prognosewerte für unseren Probanden deutlich schlechter sind als die für den "durchschnittlichen Probanden". Nach 46 Tagen beispielsweise hat der "durch-schnittlichen Probanden" noch eine Chance zu leben von 0.9679, unser Proband jedoch nur von 0.7070.

Es ist mühsam die prognostizierte Überlebenswahrscheinlichkeit zu berechnen. Almo übernimmt diese Aufgabe. Wir werden das im folgenden Maskenprogramm Prog41m2.Msk zeigen. Dort besteht die Möglichkeit, die Überlebenswahrscheinlichkeit von "Wertemustern" zu errechnen.

P41.4 Optionen für Programm-Maske Prog41m2

Wir verwenden wieder das Beispiel mit der Herztransplantation. Dabei fügen wir jetzt noch folgende Besonderheiten dazu:

1. Eine Gruppierungsvariable (bzw. "Schichtungsvariable"); z.B. der allgemeine Gesund-heitszustand. Für die Probanden mit gutem u. schlechtem Gesundheitszustand werden verschiedene Überlebensfunktionen ermittelt.

   Wir wollen hier gleich folgende wichtige Aussage vorweg nehmen: Eine Variable sollte als normale unabhängige Variable in die Analyse eingeführt werden und nicht als Gruppierungsvariable – es sei denn sie ist "zeitabhängig" und verletzt damit die Forderung des "proportionalen Hazard". Wir werden auf dieses Thema in P41.7 zurückkommen.

2. Für Personen mit einem bestimmten 'Wertemuster' in den unabhängigen Variablen (z.B. Patienten im Alter von 48 Jahren die im Hospital A operiert wurden) wird deren spezifische Überlebensfunktion ermittelt

3. Optionen für zusätzliche Ausgabeteile
   a) Ausgeben der Analysevariable für alle Untersuchungseinheiten

4. Grafik-Optionen
   a) Liniendiagramm der Basis-Survivalfunktionen (wenn Gruppen vorhanden, dann je Gruppe)
   b) Liniendiagramm der log-minus-log-Funktionen (wenn Gruppen vorhanden, dann je Gruppe)
   c) Liniendiagramm der Survivalfunktionen der Wertemuster
      Sind Gruppen vorhanden, dann wird je Gruppe ein Diagramm gezeichnet,in dem die Wertemuster verglichen werden
   d) Liniendiagramm der Survivalfunktionen der Wertemuster
      Sind Gruppen vorhanden, dann wird je Wertemuster ein Diagramm gezeichnet, in dem die Gruppen verglichen werden.
   e) Weibull-Test je Gruppe

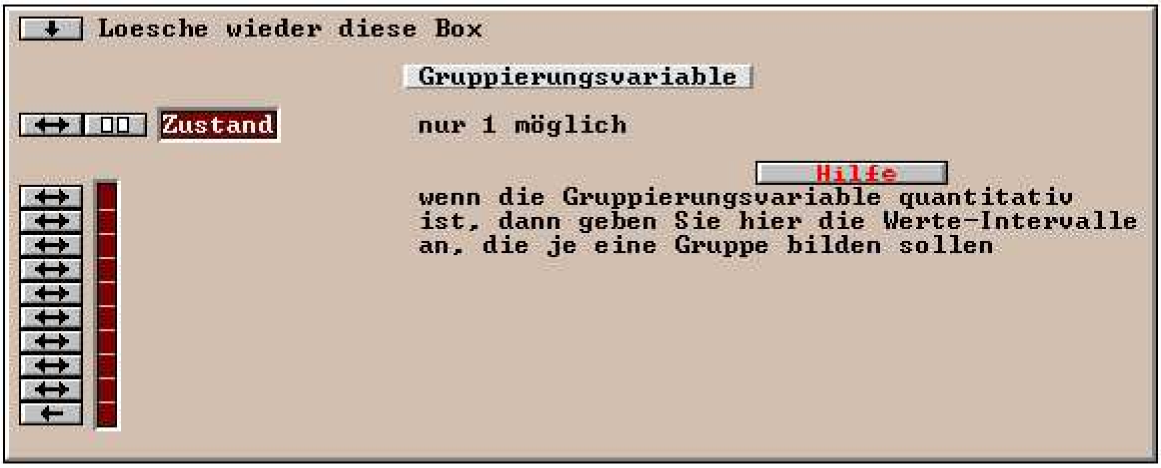
Die Optionsboxen sind bereits für das nachfolgend gerechnete Beispiel entsprechend ausgefüllt. Der Benutzer braucht die Boxen nur zu öffnen.

Erläuterung zu den Options-Boxen von Prog41m2

Wir erläutern im folgenden die Boxen, die im vorausgegangenen Abschnitt P41.1 ausgespart worden sind..

Box 10: Gruppierungsvariable

Optionsbox geöffnet:



Eine Variable wird sinnvollerweise nur dann als Gruppierungsvariable (Schichtungsvariable) eingesetzt, wenn sie die Bedingung der Zeitunabhängigkeit bzw. der proportionalen Hazardfunktion verletzt. Wir werden im nächsten Abschnitt ausführlich darauf zurückkommen.

Almo ermittelt, wenn Sie die entsprechenden Optionen setzen (siehe unten) die Basis-Surviaval-Funktion und die Survival-Funktion für Wertemuster.

Wenn Sie nun eine Gruppierungsvariable angeben, z.B. den Gesundheitszustand (gut, schlecht) dann werden diese Funktionen separat für die Gruppen ermittelt, in unserem Beispiel also für Personen mit gutem und mit schlechtem Gesundheitszustand.

Die Gruppierungsvariable wird normalerweise nominal sein. Sie darf aber auch quantitativ sein. Betrachten wir ein Beispiel: Die Gruppierungsvariable variiere zwischen 0.5 und 3.0. Sie wollen drei Gruppen bilden

```
1. Gruppe von 0.5 bis 1.5
2. Gruppe von 1.5 bis 2.5
3. Gruppe von 2.5 bis 3.0
```

Schreiben Sie dann in die Eingabefelder folgendes:

```
0.5 ,  1.5
1.5 ,  2.5
2.5 ,  3.0
```

Beachten Sie: Zwischen die Werte einer Gruppe muss ein Beistrich als Trennzeichen gesetzt werden

Erlaubt ist nur **eine** Gruppierungsvariable. Sie können über die MIT-Operationen jedoch mehrere Variable kombinieren. Betrachten wir ein Beispiel. Sie wollen Gesundheitszustand mit Geschlecht kombinieren. In der Box "Umkodierungen und Kein-Wert-Angabe" schreiben Sie in ein Eingabefeld

V99 = Zustand mit Geschlecht;

und geben V99 dann als Gruppierungsvariable an. Almo bildet dann folgende 4 Gruppen

Zustand			
gut		schlecht	
Geschlecht		Geschlecht	
männlich	weiblich	männlich	weiblich
1	2	3	4

Die "Kombinations"-Variable besitzt die Werte 1, 2, 3, 4.
Beispiel: Gruppe 3 sind Männer mit schlechterem Gesundheitszustand.
Siehe dazu die Darstellung der MIT-Operation in Almo-Handbuch, Teil 2, Abschnitt 23.

Box 15: Optionen: Wertemuster

Optionsbox geöffnet:

Es soll die Überlebensfunktion und die Hazardfunktion für eine Person ermittelt werden, die

```
im Hospital A transplantiert wurde
und 48 Jahre alt ist
```

sowie für eine Person, die

```
im Hospital B transplantiert wurde
und 58 Jahre alt ist
```

Wir haben also 2 Wertemuster

Wertemuster 1: `Hospital = 1      Alter = 48`
Wertemuster 2: `Hospital = 2      Alter = 58`

Die Variable „Gewerbe-Unverträglichkeit“ verwenden wir nicht. Almo setzt für diese Variable deren Mittelwert ein. Allgemein gilt: Von den unabhängigen nominalen und quantitativen Variablen kann der Benutzer beliebig viele verwenden, um ein

Wertemuster zu definieren. Die nicht explizit vom Benutzer angegebenen unabhängigen Variablen werden von Almo mit deren Mittelwert in das Wertemuster eingesetzt.

Betrachten wir ein weiteres Beispiel:

Die unabhängigen nominalen Variablen seien:

```
Gechlecht: männlich (=1),  weiblich (=2)
Hospital: Hosp1, Hosp2, Hosp3
```

Die unabhängigen quantitativen Variablen seien:

```
Alter
Körpergewicht
```

Sie wollen nun die Überlebensfunktion dargestellt erhalten für

1. Männer im Alter von 48
2. Frauen im Alter von 58

Geben Sie als Zahl der Wertemuster = 2 an und schreiben Sie in die Editfelder der Matrix der Wertemuster

```
Geschlecht ,       1,  2
Alter ,           48, 58
```

Zuerst wird also der Variablenname (oder -nummer) geschrieben, dann der Wert des 1. Wertemusters, dann der des 2. Es können beliebig viele Wertemuster angefordert werden.

Beachte: Als Trennzeichen innerhalb eines Eingabefeldes muss ein Beistrich geschrieben werden. Am Zeilenende wird kein Beistrich geschrieben.

Almo setzt automatisch für die anderen unabhängigen Variablen, die der Benutzer nicht für die Wertemuster verwendet, deren Mittelwerte ein.

Das gilt auch für die nicht verwendeten nominalen Variablen. In unserem Beispiel wird die nominale Variable "Hospital" nicht verwendet. Almo löst intern diese Variable in 2 Dummies auf (die 3. Dummy ist redundant) und setzt für diese Dummies deren Mittelwert ein. Der Mittelwert einer Dummy-Variablen ist gleich dem Anteilswert der Probanden, die sich in der betreffenden Ausprägung befinden.

Hinweis: Wenn Sie 3 oder mehr Wertemuster angeben, dann erzeugt Almo eine Ausgabe, die so breit ist, daß sie nicht mehr auf ein A4-Blatt ausgedruckt werden kann. Es ist also sinnvoll, die Wertemuster auf 2 oder mehr Programmläufe zu verteilen.

Box 16: Optionen: Zusätzliche Ausgabe

Optionen: Zusätzliche Ausgabe

[0] 1= Ausgeben der Analysevariable für alle Untersuchungseinheiten
0= nicht

[1] 1= Ausgeben der Basis-Survivalfunktion und der Basis-Hazardfunktion; wenn Wertemuster angegeben,dann auch Survivalfunktion u. Hazardrate je Wertemuster
0= nicht

Eingabefeld 1 = 1. Alle Untersuchungseinheiten werden mit ihren Werten in den Analysevariablen in einer ”schönen Form” ausgegeben. Es entsteht folgender Output.

Daten aller Untersuchungseinheiten:

Nr	Zeit	Ereignis 1=tot 0=zens	Haeufig keit	Hospital	Hospital	Alter	GewebeUn
1	29	1	2	0.0000	1.0000	54.0000	1.0800
2	30	0	3	0.0000	1.0000	45.0000	0.1600
3	39	1	3	1.0000	0.0000	42.0000	1.3800
4	44	0	1	0.0000	1.0000	36.0000	0.0000
5	46	1	8	0.0000	1.0000	42.0000	0.6100
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Eingabefeld 2 = 1. Die Basis-Survivalfunktion und die log-minus-log-Funktion werden für alle ”Todeszeitpunkte” ausgegeben. Wurden Wertemuster in Box 15 angegeben, dann wird für diese die Survivalfunktion ausgegeben.

Almo liefert folgende Ausgabe:

Survival-Funktionen

Survival-Funktion fuer Untersuchungsobjekte mit folgendem Wertemuster

	Wertemuster 1	Wertemuster 2
Hospital:HospA -->	ja	nein
Hospital:HospB -->	nein	ja
Hospital:HospC -->	nein	nein
Alter -->	48.0000	58.0000
GewebeUnvertraegli -->	Mittelwert	Mittelwert

Nr	Zeit	Risiko menge	Tote	Basis Surviv Funkt.	Stand. Fehler.	Basis Hazard Funkt.	Surviv Funkt.	Stand. fehler	Hazard Funkt.	Surviv Funkt.	Stand. fehler	Hazard Funkt.
Gruppe 1 = V8 Zustand gut												
1	29	238	2	0.9916	0.0059	0.0084	0.9890	0.0078	0.0111	0.9635	0.0254	0.0372
2	39	233	3	0.9789	0.0095	0.0129	0.9723	0.0124	0.0170	0.9102	0.0390	0.0568
3	46	229	8	0.9454	0.0156	0.0348	0.9288	0.0202	0.0458	0.7808	0.0568	0.1533
4	47	221	5	0.9240	0.0185	0.0229	0.9013	0.0237	0.0301	0.7059	0.0622	0.1009
5	50	216	11	0.8759	0.0241	0.0535	0.8400	0.0304	0.0704	0.5576	0.0677	0.2359
6	51	205	2	0.8672	0.0250	0.0100	0.8291	0.0315	0.0131	0.5336	0.0679	0.0439
7	60	203	9	0.8250	0.0286	0.0499	0.7764	0.0354	0.0656	0.4283	0.0654	0.2200
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.

Box 17 und 18. Grafik-Optionen
Wir werden diese Grafikoptionen bei der Ausgabe erläutern.

P41.6 Ausgabe aus Programm Prog41m2 mit Optionen

Almo liefert aus dem Maskenprogramm Prog41m2 folgende Ergebnisse:

```
Fuer Analyse aus Datenvektor ausgewaehlte Variable

Unabh. nominale Variable: V6    Hospital:  HospA, HospB, HospC
Unabh. quantitat. Variab: V3    Alter
                          V5    GewebeUnvertraeglichkeit
Gruppierungsvariable:     V8    Zustand:  gut, schlecht
Zeitvariable:             V7    Ueberlebensdauer
Ereignisvariable:         V2    Ereignis

Beachte: Fuer die unabh. nominalen Variablen
wird die 0,1,-1 Dummy-Kodierung verwendet.

Zahl der eingelesenen Datensaetze = 64
(ohne Beruecksichtigung der Gewichtung)


                                        ******************** MITTEILUNG
         von den 64 Datensaetzen wurden 7 aus der Analyse ausgeschlossen
                   da diese zensiert sind bevor das 1. Ereignis eintritt

Auspraegungen der Gruppierungsvariablen:  0.0000   1.0000
                                          -------  -------
Haeufigkeit (ungewichtet)                      40       17
Probanden (gewichtet)                         238      114
Zahl der Todeszeitpunkte                       18       11
Haeufigkeit "Tote" (gewichtet)                103       76
Haeufigkeit Zensierte (gewichtet)             135       38



Mittelwerte der unabhaengigen Variablen
---------------------------------------
Nominale Variable
  die Mittelwerte der Dummies der nominalen Variable
  sind gleich den Anteilswerten

  Hospital:HospA         0.2614
  Hospital:HospB         0.3409
  Hospital:HospC         0.3977

Quantitative Variable
           Alter       46.1960
GewebeUnvertraeg        1.2043

Konvergenz nach 5 Schritten, maximaler Fehler = 0.000429
Determinante = 3.473063e+008

log-likelihood(0)    = -801.5114
log-likelihood(4)    = -748.5521
logratio             =  105.9186  p= 0.0000  (1-p)*100= 100.0000
Wald-Statistik       =   76.1407
Score-Statistik      =  215.6671


                                Stand.                       Signifik
Variable             Beta       fehler       z        p      (1-p)100
-------------------------------------------------------------------
Hospital:HospA       0.054645   0.121416   0.4501  0.6527    34.7326
Hospital:HospB       0.131344   0.111337   1.1797  0.2381    76.1853
Hospital:HospC      -0.185989   0.121382  -1.5323  0.1255    87.4510
Alter                0.113267   0.014577   7.7705  0.0000   100.0000
GewebeUnvertraegl    1.309094   0.317127   4.1280  0.0000    99.9963
-------------------------------------------------------------------

                     part.     "Risiko"   relatives
Variable               r      exp(Beta)  Risiko in %
----------------------------------------------------
Hospital:HospA       0.000000   1.0562      5.62
Hospital:HospB       0.000000   1.1404     14.04
Hospital:HospC      -0.014730   0.8303    -16.97
```

```
Alter                  0.190838   1.1199    11.99
GewebeUnvertraegl      0.096863   3.7028   270.28
----------------------------------------------------

Korrelationen zwischen den Kovariablen
(in der Hauptdiagonale stehen die Standardabweichungen)
-------------------------------------------------------
```

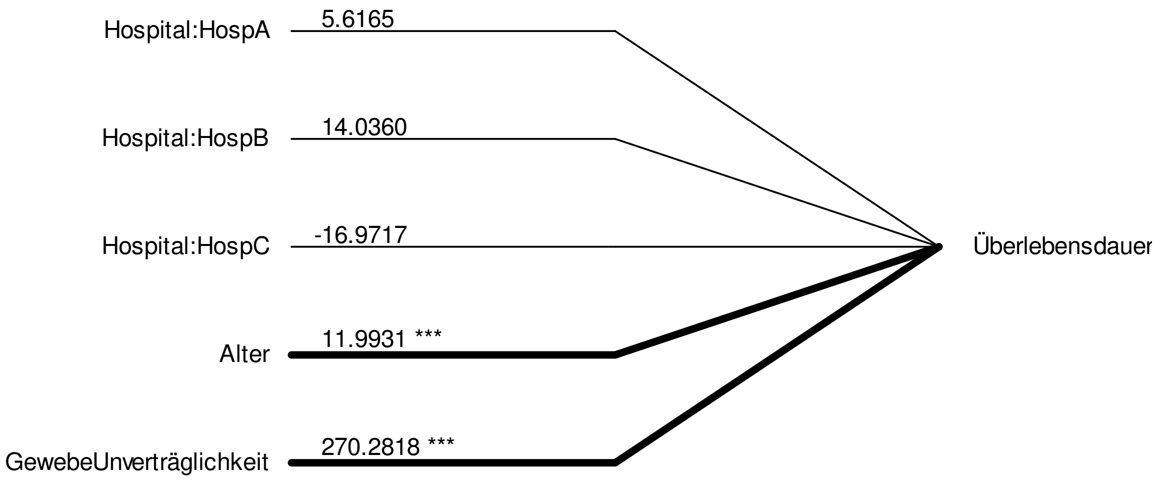
		Hospital HospA V6-1	Hospital HospB V6-2	Alter V3	GewebeUn V5
Hospital HospA	V6-1	0.1214	-0.4587	-0.2080	-0.0272
Hospital HospB	V6-2	-0.4587	0.1113	0.0926	0.0082
Alter	V3	-0.2080	0.0926	0.0145	0.1540
GewebeUn	V5	-0.0272	0.0082	0.1540	0.3171

Beachte: Dir redundante Dummy HospC wird nicht ausgegeben.

Almo liefert als Grafiken folgende zwei Flussdiagramme:

Regressionskoeffizienten

Relatives Risiko (in %)



Survival-Funktionen

Survival-Funktion fuer Untersuchungsobjekte mit folgendem Wertemuster

	Wertemuster 1	Wertemuster 2
Hospital:HospA -->	ja	nein
Hospital:HospB -->	nein	ja
Alter -->	48.0000	58.0000
GewebeUnvertraegli -->	Mittelwert	Mittelwert

Nr	Zeit	Risiko menge	Tote	Basis Surviv Funkt.	Stand. Fehler.	Basis Hazard Funkt.	Wertemuster 1 Surviv Funkt.	Wertemuster 1 Stand. fehler	Wertemuster 1 Hazard Funkt.	Wertemuster 2 Surviv Funkt.	Wertemuster 2 Stand. fehler	Wertemuster 2 Hazard Funkt.
Gruppe 1 = V8 Zustand gut												
1	29	238	2	0.9916	0.0059	0.0084	0.9890	0.0078	0.0111	0.9635	0.0254	0.0372
2	39	233	3	0.9789	0.0095	0.0129	0.9723	0.0124	0.0170	0.9102	0.0390	0.0568
3	46	229	8	0.9454	0.0156	0.0348	0.9288	0.0202	0.0458	0.7808	0.0568	0.1533
4	47	221	5	0.9240	0.0185	0.0229	0.9013	0.0237	0.0301	0.7059	0.0622	0.1009
5	**50**	**216**	**11**	**0.8759**	**0.0241**	**0.0535**	**0.8400**	**0.0304**	**0.0704**	**0.5576**	**0.0677**	**0.2359**
6	51	205	2	0.8672	0.0250	0.0100	0.8291	0.0315	0.0131	0.5336	0.0679	0.0439
7	60	203	9	0.8250	0.0286	0.0499	0.7764	0.0354	0.0656	0.4283	0.0654	0.2200
8	66	194	3	0.8096	0.0297	0.0188	0.7575	0.0366	0.0247	0.3943	0.0638	0.0826
9	68	191	8	0.7672	0.0324	0.0538	0.7058	0.0392	0.0708	0.3110	0.0578	0.2372
10	161	174	9	0.7171	0.0353	0.0676	0.6458	0.0418	0.0888	0.2309	0.0501	0.2978
11	253	142	4	0.6944	0.0364	0.0321	0.6191	0.0427	0.0422	0.2005	0.0463	0.1415
12	280	138	11	0.6319	0.0387	0.0943	0.5468	0.0441	0.1241	0.1323	0.0357	0.4158
13	297	127	3	0.6150	0.0393	0.0271	0.5277	0.0444	0.0357	0.1174	0.0331	0.1195
14	624	81	8	0.5409	0.0428	0.1284	0.4457	0.0463	0.1688	0.0667	0.0232	0.5658
15	730	67	4	0.4880	0.0462	0.1029	0.3893	0.0485	0.1353	0.0424	0.0177	0.4533
16	994	40	1	0.4705	0.0485	0.0366	0.3710	0.0503	0.0482	0.0360	0.0164	0.1615
17	1024	39	2	0.4362	0.0520	0.0756	0.3359	0.0526	0.0995	0.0258	0.0136	0.3333
18	1350	24	10	0.1811	0.0433	0.8790	0.1057	0.0332	1.1560	0.0005	0.0006	3.8740
Gruppe 2 = V8 Zustand schlecht												
19	10	114	2	0.9964	0.0027	0.0036	0.9953	0.0035	0.0047	0.9843	0.0116	0.0159
20	25	101	6	0.9820	0.0079	0.0146	0.9763	0.0103	0.0192	0.9229	0.0326	0.0644
21	50	93	5	0.9679	0.0119	0.0145	0.9580	0.0155	0.0190	0.8659	0.0468	0.0637
22	51	88	8	0.9441	0.0178	0.0248	0.9272	0.0230	0.0327	0.7761	0.0645	0.1095
23	54	80	5	0.9284	0.0217	0.0168	0.9069	0.0278	0.0221	0.7206	0.0741	0.0742
24	63	75	10	0.8902	0.0292	0.0420	0.8581	0.0370	0.0553	0.5988	0.0866	0.1852
25	64	65	10	0.8360	0.0378	0.0628	0.7901	0.0470	0.0826	0.4540	0.0905	0.2768
26	65	55	4	0.8093	0.0425	0.0324	0.7571	0.0523	0.0426	0.3936	0.0911	0.1427
27	136	39	7	0.7567	0.0498	0.0672	0.6931	0.0600	0.0884	0.2927	0.0849	0.2961
28	322	32	10	0.6614	0.0599	0.1347	0.5806	0.0692	0.1771	0.1617	0.0645	0.5935
29	836	9	9	0.0000	0.0000	KW	0.0000	0.0000	KW	0.0000	0.0000	KW

******* Erläuterung zur obigen Tabelle**

Risikomenge. Das ist die Zahl der Probanden, die beim Eintritt in das betreffende Zeitintervall dem Risiko ausgeseztzt sind zu terminieren (zu sterben).

"**Tote**". Das ist die Zahl der Probanden, die im Verlauf des Intervalls terminiert haben (gestorben sind).

Beispiel: Beim Eintritt in das Zeitintervall 39 ist die Risikomenge=233. Von ihnen sterben 3 Probanden bis zum Ende des Intervalls, also bis zum Zeitpunkt 46.

Die Basis-Hazardfunktion

Die Basis-Hazardfunktion H_t ist die Hazardfunktion des „durchschnittlichen Probanden“. Sie ist gleich der Wahrscheinlichkeit im Zeitintervall t zu terminieren (zu sterben) – wenn keine unabhängigen Variablen sich im Modell befinden. Natürlich ist es der Sinn der Cox-Regression, unabhängige Variable im Modell zu haben und ihre Wirkung zu untersuchen. Trotzdem wollen wir kurz den Grenzfall der Cox-Regression ohne unabhängige Variable betrachten – um das Konzept der Basis-Hazardfunktion verständlich zu machen.

(7) $H'_t = q_t = \dfrac{d_t}{n_t}$

H'_t = Basis-Hazardfunktion, wenn keine Kovariaten vorhanden
q_t = Wahrscheinlichkeit im Zeitintervall t zu terminieren (zu sterben)
d_t = Zahl der „Toten“ in Intervall t
n_t = Risikomenge im Intervall t, d.h. Menge der Probanden, die dem Risiko ausgesetzt sind, im Intervall t zu terminieren.

Siehe dazu P39.3, Gleichung 2

Wir berechnen die Basis-Hazardfunktion für den Zeitpunkt 29. Die Zahl der Toten in diesem Intervall ist 2. Die Risikomenge 238. Also gilt

$$H_{29} = q_{29} = \frac{2}{238} = 0.0084$$

Gleichung 7 ist der Nelson-Aalen-Schätzer.
Almo verwendet den (weniger einsichtigen) Produkt-Limit-Schätzer

(7a) $H_t = -\ln(1 - H'_t)$

der nur einen minimal anderen Wert erbringt.

Befinden sich Kovariate in der Analyse, dann ist die Berechnung der Basis-Hazardfunktion aufwendiger.

Die Hazardfunktion kann kumuliert werden. Wir sprechen dann von der kumulierten Basis-Hazardfunktion H^*_t. Zwischen dieser und der Basis-Survivalfunktion B_t besteht folgender Zusammenhang.

(8) $B_t = e^{-H^*_t}$

(9) $H^*_t = -\ln(B_t)$

B_t = Basis-Survivalfunktion
H^*_t = komulierte Basis-Hardfunktion

Gleichung 8 und 9 gelten für den Fall, daß Kovariate vorhanden sind.

Die Hazardfunktion eines Wertemusters

Sie wird entsprechend definiert als die Wahrscheinlichkeit eines Probanden mit dem betreffenden Wertemuster im Zeitintervall t zu terminieren.

Erläuterung zu Wertemuster

Wir haben in der Erläuterung zur Ergebnis-Ausgabe von Prog41m2.Msk in Abschnitt P41.2 ausgeführt, daß die Basis-Survivalfunktion und die Basis-Hazardfunktion die Überlebenswahrscheinlichkeit und die Hazardfunktion des „durchschnittlichen Probanden“ sind.
Zum Zeitpunkt 50 ist die Überlebenswahrscheinlichkeit eines „durchschnittlichen Probanden“ aus der Gruppe 1 (guter Allgemeinzustand) 0.8759.

Ein Proband, der mit 48 Jahren geringfügig älter ist und in Hospital 1 transplantiert wurde ist sie mit 0.8400 nur etwas geringer. Der Hazardwert des „durchschnittlichen Probanden“ ist 0.0535. Für das Wertemuster 1 finden wir einen etwas ungünstigeren Hazardwert von 0.0704.
Der mit 58 Jahren ältere Proband, der allerdings im günstigeren Hospital 3 transplantiert wurde, hat wesentlich schlechtere Erfolgsaussichten. Seine Wahrscheinlichkeit ist nur 0.5576. Sein Hazardwert ist 0.2359.
Wir wollen für den Probanden des Wertemusters 1 den Hazardwert gemäß Gleichung 5 berechnen.

$$h_t = H_t * e^{b_1 x_1} * e^{b_2 x_2} * \ldots * e^{b_m x_m}$$

Beachte: Die x sind um ihren Mittelwert zentriert. D.h. von jedem x-Wert muß der jeweilige Mittelwert subtrahiert werden.

$h_{50} =$	0.0535	Basis-Hazardrate
	$*e^{0.0546(1-0.2614)}$	Hospital A
	$*e^{0.1313*0}$	Hospital B
	$*e^{-0.1860*0}$	Hospital C
	$*e^{0.1133(48-46.1960)}$	Alter
	$*e^{1.3091*0}$	Gewerbe-Unverträglichkeit
	$= 0.07$	

Der Proband wurde in Hospital A transplantiert. Der Mittelwert dieser Dummy ist 0.2614, d.h. ein Anteil von 0.2614 der Probanden wurde in Hospital A transplantiert. Der Regressionskoeffizient für diese Dummy ist 0.0546. Der entsprechende Ausdruck in obiger Gleichung lautet deswegen

$e^{0.0546(1-0.2614)}$

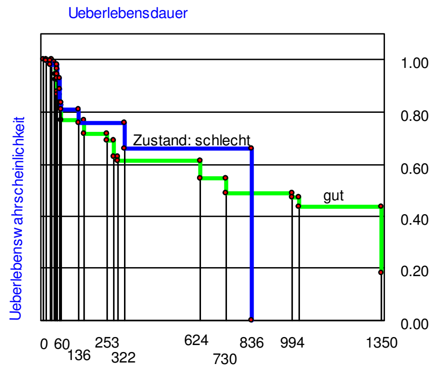
Für die Dummy für Hospital B ist der Mittelwert 0.3407 und der Regressionskoeffizient 0.1313. Da der Proband nicht in Hospital B transplantiert wurde setzt Almo den Mittelwert ein.

$$\mathbf{e^{0.1313(0.3407-0.3407)} = e^{0.1313*0} = e^{o} = 1}$$

Ebenso wird für Hospital C verfahren.

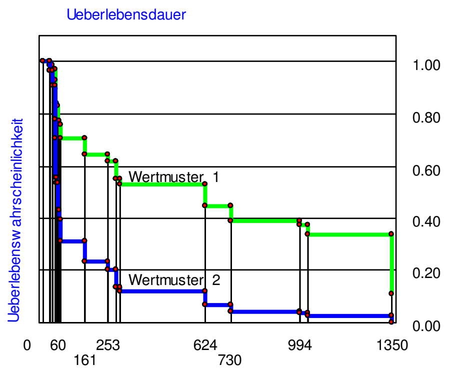
Wenn bei den Grafik-Optionen das Eingebfeld 1 auf =1 gesetzt wurde, dann erzeugt Almo ein Liniendiagramm der Basis-Ueberlebensfunktion für die 2 Gruppen der Gruppierungsvariablen

Basis-Ueber-
lebensfunktion

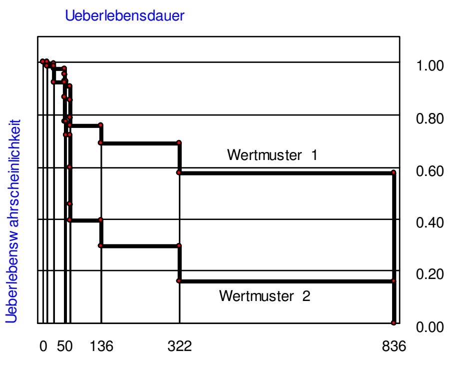


Wenn bei den Grafik-Optionen das Eingebfeld 3 auf =3 gesetzt wurde, dann erzeugt Almo folgende Liniendiagramme der Survivalfunktion je Wertemuster und je Gruppe.

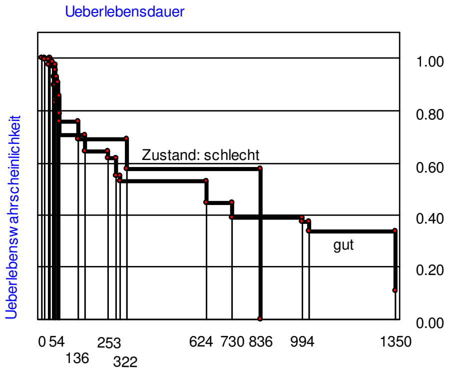
Ueberlebensfunktion
fuer Wertemuster
fuer 1-te Gruppe



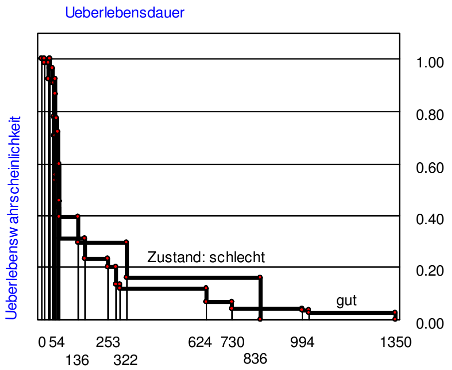
Ueberlebensfunktion
fuer Wertemuster
fuer 2-te Gruppe



Ueberlebensfunktion
fuer Gruppen
fuer 1-tes Wertemuster



Ueberlebensfunktion
fuer Gruppen
fuer 2-tes Wertemuster



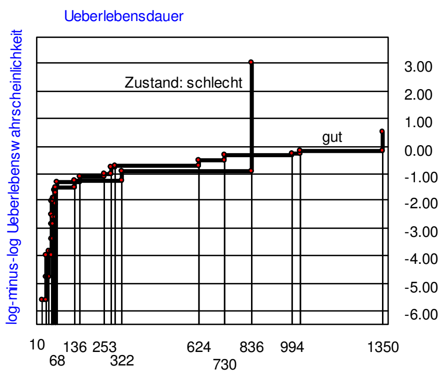
Wenn bei den Grafik-Optionen das Eingabefeld 2 auf =1 gesetzt wurde, dann gibt Almo die log-minus-log-Funktion der Survivalfunktion aus und erzeugt ein Liniendiagramm für die 2 Gruppen der Gruppierungsvariablen

```
 Zeit       log-minus-log
 ----       -------------

Gruppe 1 = V8 Zustand  gut

   29         -4.7749
   39         -3.8473
   46         -2.8801
   47         -2.5381
   50         -2.0209
   51         -1.9484
   60         -1.6481
   66         -1.5551
   68         -1.3281
  161         -1.1010
  253         -1.0088
  280         -0.7787
  297         -0.7213
  624         -0.4870
  730         -0.3322
  994         -0.2824
 1024         -0.1868
 1350          0.5357

Gruppe 2 = V8 Zustand  schlecht

   10         -5.6263
   25         -4.0060
   50         -3.4214
   51         -2.8559
   54         -2.5991
   63         -2.1511
```

64	-1.7195
65	-1.5533
136	-1.2775
322	-0.8834
836	2.9998

Für die log-minus-log-Funktion wird dann noch folgendes Liniendiagramm erzeugt:

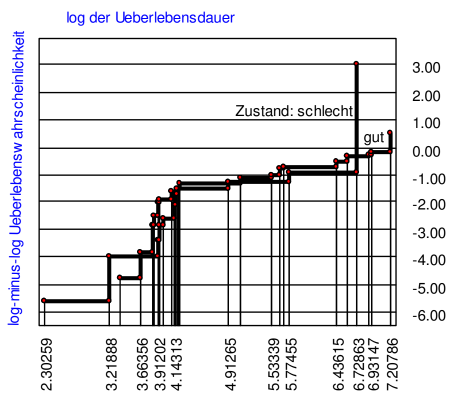
log-minus-log-Diagramm



Der Zweck dieser Grafik ist es, die Annahme der "proportionalen Hazardfunktionen" zu überprüfen. Die beiden Kurven verlaufen weitgehend parallel, sodaß die Annahme als erfüllt gelten darf. Die Abweichungen am rechten Ende sind vernachlässigbar, da sie auf nur wenigen Daten beruhen. Die Variable des Zustands (gut, schlecht) braucht deshalb nicht als Gruppierungsvariable (Schichtungsvariable) in die Analyse eingeführt zu werden. Sie kann als normale unabhängige Variable eingesetzt werden.

Wenn bei den Grafik-Optionen das Eingebfeld 5 auf =1 gesetzt wurde, dann erzeugt Almo ein Liniendiagramm für den Weibull-Test.

Weibull-Test
(log-minus-log-Survival
gegen logarithmierte Zeit)



Zweck dieser Darstellung ist die graphisch Überprüfung, ob nicht doch ein parametrisches Weibull-Modell statt dem verteilungsfreien Cox-Modell verwendet werden kann. Wurde keine Verletzung der PH-Eigenschaft und kein abweichender Verlauf der Residuen festgestellt, zeigt diese Graphik den Verlauf des lln(S_0), d.h den log-minus-log der Basissurvivorfunktion gegen die logarithmierte Zeit aufgetragen. Ergibt sich eine Gerade (im Falle einer Gruppierung : ebensoviele parallele Gerade), ist auf Weibull-verteilte Lebensdauern zu schließen. Die Steigung dieser Geraden entspricht dem Formparameter α, weswegen bei $\alpha \cong 1$ eine Exponentialverteilung anzusetzen wäre. In unserem Fall spricht der Verlauf deutlich dagegen.

Wurde in Box 14 (allgemeine Grafik-Optionen) das 2. Eingabefeld auf 1 gesetzt, dann erzeugt Almo zusätzlich zu allen obigen Grafiken einfache Diagramme mit Textzeichen. Als Beispiel geben wir hier die Überlebenswahrscheinlichkeit für die 2. Gruppe für die 2 Wertemuster
aus:

```
In nachfolgender Grafik bedeuten die Achsen:

waagrecht : Ueberlebenswahrscheinlichkeit fuer 2-te Gruppe
            fuer 2 Wertemuster
senkrecht : Ueberlebensdauer

     0  ..... 0.2 ..... 0.4 ..... 0.6 ..... 0.8 .....  1
  10 .                                                21
  25 .                                              2  1.
  50 .                                           2    1 .
  51 .                                       2      1   .
  54 .                                    2       1     .
  63 .                              2           1       .
  64 .                         2               1        .
  65 .                      2                1          .
 136 .                 2                  1             .
 322 .          2                   1                   .
 836 2                                                  .
     0  ..... 0.2 ..... 0.4 ..... 0.6 ..... 0.8 .....  1
```

P41.7 Zeitabhängige Kovariate

Betrachten wir folgendes (konstruiertes) Beispiel. Ab dem Zeitpunkt der Herztransplantation erhält ein Teil der Patienten täglich ein Medikament x, der Rest der Patienten keines. Die Variable "Medikament x: nein, ja" wird als unabhängige nominale Variable in die Analyse aufgenommen. Wir erhalten folgendes Ergebnis

	Regress Koeff ß	Risiko exp (ß)	Signifikanz p (1-p) 100
Medikament nein	0.3	1.35	0.03 97%

Das Medikament hat eine hochsignifikante Wirkung. Bei Nichteinnahme des Medikaments erhöht sich die Wahrscheinlichkeit zu sterben um den Multiplikationsfaktor von 1.35.

Bei genauerer Betrachtung der Daten stellen wir jedoch fest, daß das Medikament in der Anfangszeit nach der Transplantation zwar die Überlebenswahrscheinlichkeit sehr stark erhöht, in der späteren Zeit jedoch gegenteilig wirkt und die Überlebenswahrscheinlichkeit (im Vergleich zur Nichteinnahme) sogar verringert.

Damit ist eine der Grundannahme der Cox-Regression verletzt. Diese Grundannahme können wir in folgender Weise formulieren.

> Eine unabhängige Variable darf nicht "zeitabhängig" sein.
>
> Eine Variable ist zeitabhängig, wenn sie im Verlauf der Überlebenszeit ihre Wirkung auf die Überlebenswahrscheinlichkeit ändert.
>
> Das geht auch unmittelbar aus der Grundgleichung der Cox-Regression hervor, wie wir sie in Gleichung 6 in P41.0 formuliert haben:
>
> $$h_t = H_t * e^{b_1 x_1} * e^{b_2 x_2} * .... * e^{b_m x_m}$$
>
> Die Hazardfunktion h_t ist (multiplikativ) bestimmt durch die Basis-Hazardfunktion H_t und die "Risiko-Koeffizienten $e^{b_i x_i}$. Nur die Basis-Hazardfunktion H_t ist zeitabhängig. In den Risiko-Koeffizienten ist die Zeit t nicht enthalten. Der Regressionskoeffizient b_i variiert nicht mit der Überlebenszeit.

Nicht gemeint ist mit dem Begriff der Zeitabhängigkeit, daß die Variable selbst sich in ihrem Wert im Verlauf der Zeit ändert, wie dies etwa beim Lebensalter der Fall ist. Diese Art von Variablen könnte man "zeitveränderliche" nennen.

In unserem Beispiel ändert die Variable "Medikament x: nein, ja" ihren Wert im Zeitverlauf nicht. Ein Teil der Patienten erhält täglich das Medikament, der Rest nicht. Und das bis zum Tod bzw. zum Ausscheiden aus der Beobachtung. Die Variable ändert jedoch in der beschriebenen Weise ihre Wirkung auf die Überlebenswahrscheinlichkeit.

Das Lebensalter ist eine "zeitveränderliche" Variable. Die Patienten werden täglich älter. Das bedeutet aber nicht notwendig, daß das Lebensalter "zeitabhängig" ist.

Es wäre zeitabhängig, wenn folgender Fall gegeben wäre:

Das Lebensalter wird zum Zeitpunkt der Transplantation erhoben. Die Überlebenswahrscheinlichkeit ist anfänglich umso besser je jünger die Patienten sind, wohl weil ihr allgemeiner Gesundheitszustand besser ist. Im Verlauf der Zeit kehrt sich dieser Zusammenhang jedoch um: Die Überlebenswahrscheinlichkeit wird umso schlechter je jünger die Patienten sind, wohl weil deren Immunabwehr gegen das Transplantat stärker ist. (Wir wollen nochmals darauf hinweisen, daß unser Beispiel konstruiert ist). Die Variable des Lebensalters ist somit eine "zeitabhängige", da ihre Wirkung auf die Überlebenswahrscheinlichkeit sich im Zeitverlauf ändert.

Der Sachverhalt der "Zeit-Unabhängigkeit" läßt sich formalisiert und strenger ausdrücken: Für 2 oder mehrere beliebige Wertemuster gilt, daß das Verhältnis ihrer Hazardfunktion an jedem Punkt der Zeitachse dasselbe sein muß. Die Cox-Regression wird aus diesem Grund auch "proportionales Hazardmodell" genannt. Wir wollen diesen Begriff der Proportionalität an einem Beispiel erläutern.

Wir rechnen mit Maskenprogramm Prog41m1.Msk eine Analyse, wobei wir das Lebensalter als unabhängige Variable einsetzten und 2 Wertemuster berechnen, eines für 48-jährige und eines für 58-jährige Patienten. Wir erhalten folgendes Ergebnis (stark gekürzt)

Überlebens-zeit	Wertemuster 1 48 Jahre alt Hazardfunktion h1	Wertemuster 2 58 Jahre alt Hazardfunktion h2	Verhältnis h2/h1
10	0.0138	0.0374	2.71
25	0.0154	0.0417	2.71
29	0.0163	0.0444	2.72
39	0.0170	0.0462	.
46	0.0173	0.0470	.
47	0.0181	0.0492	.
50	0.0392	0.1065	.
.	.	.	.
.	.	.	.
.	.	.	.
297	0.0399	0.1084	2.72

Das Verhältnis der beiden Hazardfunktionen ist konstant 2.71 bis 2.72.

Nun stellen sich 2 Fragen:

1. Wie kann die Annahme der "Zeit-Unabhängigkeit" bzw. der "Proportionalität der Hazardfunktionen" überprüft werden?

2. Wenn diese Annahme nicht bestätigt werden kann, welche Möglichkeit gibt es dann trotzdem eine Cox-Regression zu rechnen?

Wir wollen diese beiden Fragen zunächst nur kurz beantworten.

Zu 2.: Erweist sich eine Variable als zeitabhängig, d.h. verletzt sie die Annahme der proportionalen Hazardfunktionen, dann kann sie als Gruppierungsvariable verwendet werden.

Zu 1.: Die Kurven der Überlebensfunktion verschiedener Wertemuster oder Gruppen dürfen sich nicht kreuzen.

Oder: Die Kurven der log-minus-log-Funktion der Basis-Survivalfunktion verschiedener Gruppen sollen annähernd parallel verlaufen.

Dies alles sind grafische Verfahren, die der Interpretation durch den Auswerter viel Raum lassen.

Eine eindeutige Lösung erhalten wir, wenn wir eine Variante der Cox-Regression mit zeitabhängigen Variablen rechnen.

Das Modell dieser Variante lautet für unser Beispiel

$$h_t = H_t * e^{b_1 \cdot \text{Alter}} * e^{b_2 \cdot (\text{Alter*Zeit})}$$

h_t = Hazardfunktion

Siehe dazu Gleichung 6 in P41.0

Die Besonderheit ist, daß eine Art Interaktion von Alter mit der Zeit als zusätzliche Variable eingeführt wird. b_2 ist deren Regressionskoeffizient. Ist b_2 signifikant, dann ist das Alter eine zeitabhängige Variable und darf nicht als unabhängige Variable in die Cox-Regression eingeführt werden. Alter muß dann in 2 oder mehrere Altersgruppen unterteilt werden und als Gruppierungsvariable, neben anderen unabhängigen Variablen, eingeführt werden.

Almo liefert folgendes Ergebnis (gekürzt)

Variable	Beta	Stand. fehler	z	p	Signifik (1-p)100
Alter	0.098153	0.037886	2.5908	0.0096	99.0410
Alter *	0.000034	0.000094	0.3576	0.7207	27.9316

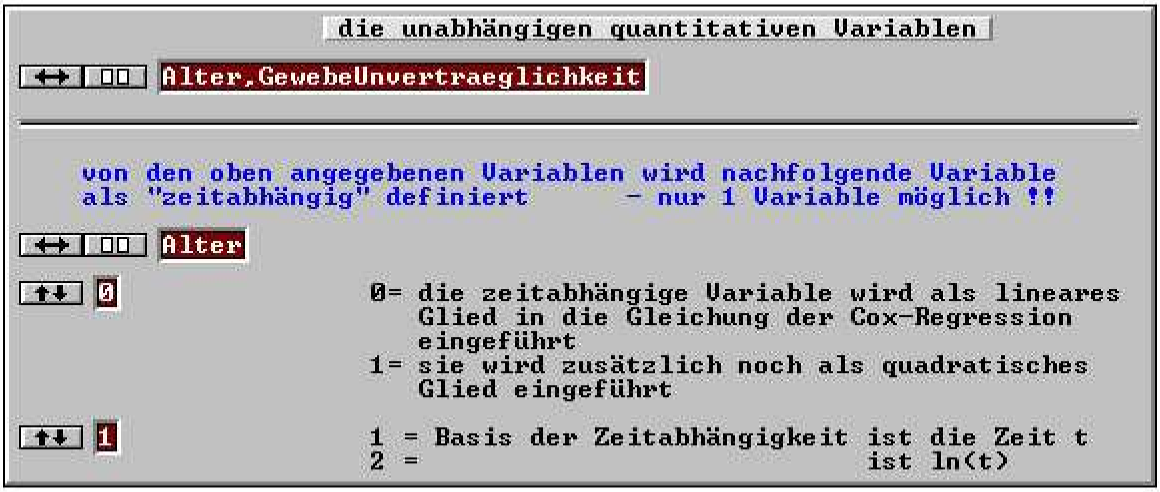
* =definiert linear zeitabhaengig

Alter * ist nicht signifikant, d.h. Alter ist nicht zeitabhängig.

P41.8 Programm-Maske mit definiert zeitabhängiger Variablen Prog41m3

Prog41m3 ist identisch mit Prog41m2, das in Abschnitt P41.1 abgebildet ist. Nur die Box „Unabhängige quantitative Variable“ ist anders.

Box: Die unabhängigen quantitativen Variablen



Eingabefeld 1: Die Variablen werden angegeben.

Eingabefeld 2: Aus den in Eingabefeld 1 angegebenen Variablen wird eine Variable als zeitabhängig definiert. In unserem Beispiel ist dies "Alter".

Es ist nur 1 Variable möglich. Unabhängige **nominale** Variable können nicht als zeitabhängig definiert werden. Sehr wohl ist es aber möglich eine zweiwertige nominale Variable (z.B. Gesundheitszustand: gut, schlecht) als unabhängige **quantitative** Variable anzugeben (das ist bei zweiwertigen Variablen immer zulässig) und dann als zeitabhängig zu definieren.

Eingabefeld 3

0 = die zeitabhängige Variable wird als lineares Glied in die Gleichung der Cox-Regression eingeführt
1 = sie wird zusätzlich noch als quadratisches Glied eingeführt

Eingabefeld 4

1 = als Basis der Zeitabhängigkeit wird die Zeit t verwendet
2 = wird ln (t) verwendet

P41.10 Ausgabe für Analyse mit definiert zeitabhängiger Variablen

Almo liefert für das Maskenprogramm Prog41m3 folgendes Ergebnis:

```
Ergebnisse aus Almo: Cox-Regression
----------------------------------
Fuer Analyse aus Datenvektor ausgewaehlte Variable

Unabh. nominale Variable: V6    Hospital:  HospA, HospB, HospC
Unabh. quantitat. Variab: V3    Alter
                          V5    GewebeUnvertraeglichkeit
Zeitvariable:             V7    Ueberlebensdauer
Ereignisvariable:         V2    Ereignis

Beachte: Fuer die unabh. nominalen Variablen
wird die 0,1,-1 Dummy-Kodierung verwendet.

Definiert zeitabhaengige Variable
---------------------------------
Die Variable V3  Alter
wurde linear definiert-zeitabhaengig zur Basis t gesetzt.

Zahl der eingelesenen Datensaetze = 64


                                            ******************** MITTEILUNG
         von den 64 Datensaetzen wurden 3 aus der Analyse ausgeschlossen
                   da diese zensiert sind bevor das 1. Ereignis eintritt

Keine Gruppierungsvariable

Probanden                             61
Zahl der Todeszeitpunkte              27
Haeufigkeit "Tote"                    29
Haeufigkeit Zensierte                 32



Mittelwerte der unabhaengigen Variablen
---------------------------------------
Nominale Variable
  die Mittelwerte der Dummies der nominalen Variable
  sind gleich den Anteilswerten

  Hospital:HospA         0.3443
  Hospital:HospB         0.3279
  Hospital:HospC         0.3279

Quantitative Variable
           Alter       45.8852
GewebeUnvertraeg        1.1808

Konvergenz nach 5 Schritten, maximaler Fehler = 0.000309
Determinante = -1.514794e+014

log-likelihood(0)    =  -99.0270
log-likelihood(5)    =  -86.3786
logratio             =   25.2970  p= 0.0001  (1-p)*100=  99.9876
Wald-Statistik       =   20.7077
Score-Statistik      =   37.5090


                                Stand.                     Signifik
Variable             Beta      fehler       z       p     (1-p)100
-------------------------------------------------------------------
Hospital:HospA       0.379615   0.280212   1.3547  0.1755   82.4470
Hospital:HospB       0.095252   0.277064   0.3438  0.7310   26.8989
Hospital:HospC      -0.474867   0.315455  -1.5053  0.1323   86.7731
```

Alter	0.083057	0.042696	1.9453	0.0518	94.8233
GewebeUnvertraegl	1.128453	0.381993	2.9541	0.0031	99.6857
Alter *	0.000075	0.000093	0.7995	0.4240	57.5984

* =definiert linear zeitabhaengig

Variable	part. r	"Risiko" exp(Beta)	relatives Risiko in %
Hospital:HospA	0.000000	1.4617	46.17
Hospital:HospB	0.000000	1.0999	9.99
Hospital:HospC	-0.036651	0.6220	-37.80
Alter	0.094914	1.0866	8.66
GewebeUnvertraegl	0.184295	3.0909	209.09
Alter *	0.000000	1.0001	0.01

* =definiert linear zeitabhaengig

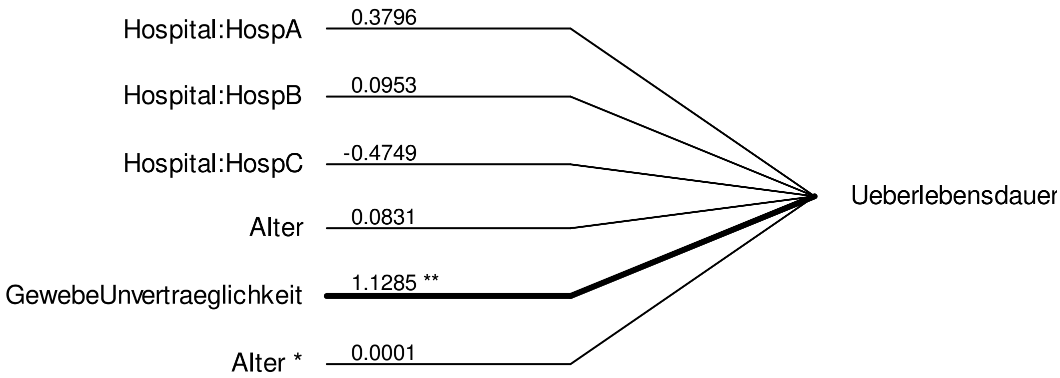
Korrelationen zwischen den Kovariablen
(in der Hauptdiagonale stehen die Standardabweichungen)

Hospital:HospA	0.280				
Hospital:HospB	-0.359	0.277			
Alter	-0.099	-0.125	0.043		
GewebeUnvertraegl	0.167	-0.096	-0.013	0.382	
Alter *	0.182	0.057	-0.634	0.067	0.000

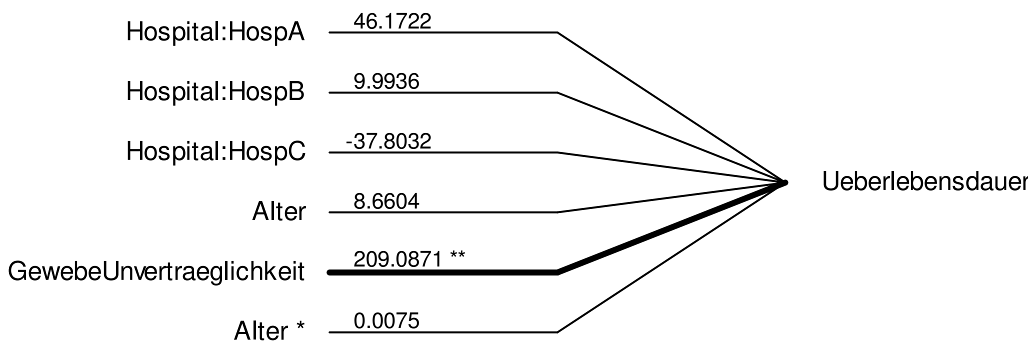
* =definiert linear zeitabhaengig

Almo liefert folgende Flussdiagramme
der Regressionskoeffizienten und des relativen Risikos

Regressionskoeffizienten



Relatives Risiko (in %)



Literatur:

H.-P. Blossfeld/A. Hamerle/K.u.Mayer: Ereignisanalyse, Campus, Frankfurt 1986
H.-P. Blossfeld/G. Rohwer: Techniques of Event History Modeling, Lawrence Erlbaum, Mahwah, New Jersey, 1995
J.D. Kalbfleisch/R.L. Prentice: The statistical analysis of faiture time data, Wiley, New York, 1980
J.P. Klein/M.L. Moeschberger: Survival Analysis, Springer, New York, 1997

SCHLAGWORTVERZEICHNIS