Transition Data Analysis (TDA)

Theoretische Grundlagen und Einführung in das Programm TDA
von G. Rohwer und U. Pötter

Institut für Biometrie der Medizinischen Hochschule Hannover

H. Hecker

Contents

Inhalt:

Transition Data Analysis (TDA)

1. Vorbemerkung

Die folgenden Beschreibungen beziehen sich auf das Programm TDA von G. Rohwer und U. Pötter, Bochum, siehe

http://www.stat.ruhr-uni-bochum.de/tda.html.

Es werden die Grundzüge der TDA und die Anwendung des Programms dargestellt. Es handelt sich dabei um eine Verallgemeinerung der Survivalanalyse:
In der Survivalanalyse wird die Dauer bis zum Eintreten eines ausgewählten Ereignisses untersucht. Man kann dies als die Verweildauer im Stadium 1 bis zum Übergang in ein Stadium 2 auffassen. In der TDA geht man nun davon aus, dass es sowohl mehrere Anfangszustände als auch mehrere Folgezustände gibt, in die hinein ein Übergang erfolgen kann. Darüber hinaus können je Beobachtungseinheit im Untersuchungszeitraum mehrere Übergänge nacheinander stattfinden, so dass hier auch mehrstufige Prozesse in die Analyse einbezogen werden.

Die theoretischen Grundlagen sind im Wesentlichen in Kapitel 3.3.1 und 6.17 des TDA-Manuals beschrieben. Eine allgemeine Einführung kann man auch in dem Buch von Kalbfleisch und Prentice nachlesen (*The Statistical Analysis of Failure Time Data, 1980*), insbesondere Kapitel 7: *Multivariate Failure Time Data and Competing Risks*. In beiden Quellen werden weitere Arbeiten zitiert, wobei die Analyse multipler Übergänge bereits seit 1961 (Chiang) in der hier beschriebenen Weise angesetzt wird.
Das Programm TDA von G. Rohwer und U. Pötter bietet eine große Auswahl verschiedener Optionen, insbesondere auch im Bereich parametrischer Modelle. Es ist unter den GNU-Bedingungen frei erhältlich.
Eine kurze Programmeinführung findet man auch unter

http://www.lrz-muenchen.de/~wlm/tdaframe.htm

2. Theoretische Grundlagen

2.1. Hazardrate und Survival

2.1.1. Einfacher Fall: Survivalanalyse

Im Fall der Survivalanalyse wird das zu untersuchende Geschehen durch nur *eine* Zufallsvariable und deren Verteilung bestimmt: die *Dauer T bis zum Eintreten des Ereignisses*; in der Terminologie von TDA: die *Verweildauer im Zustand 1.*
Daraus lassen sich dann Dichtefunktion und Hazardrate in Abhängigkeit von der Zeit t wie folgt definieren

$$\begin{aligned} T &= \text{Verweildauer im Zustand 1} \\ &= \text{Zeitpunkt des Übergangs von 1 auf 2} \end{aligned}$$

$$\begin{aligned} F(t) &= P(T \leq t) \\ &= \text{Verteilungsfunktion von } F \\ &= \text{Wahrscheinlichkeit, dass der Übergang bis Zeitpunkt } t \text{ passiert} \end{aligned}$$

$$\begin{aligned} S(t) &= 1 - F(t) = P(T > t) \\ &= \text{Survivalfunktion} \\ &= \text{Wahrscheinlichkeit, dass der Übergang bis Zeitpunkt } t \textit{ nicht} \text{ passiert} \end{aligned}$$

$$\begin{aligned} f(t) &= \lim_{\Delta \to 0} \frac{P(t \leq T < t + \Delta)}{\Delta} = F'(t) \\ &= \text{Dichtefunktion der Verteilungsfunktion} \\ &= \frac{1}{\Delta} \times \text{Wahrscheinlichkeit, dass der Übergang zwischen t und } t + \Delta \text{ passiert} \end{aligned}$$

$$\begin{aligned} r(t) &= \lim_{\Delta \to 0} \frac{P(t \leq T < t + \Delta | T \geqslant t)}{\Delta} \\ &= \text{Hazardfunktion} \\ &= \frac{1}{\Delta} \times \text{Wahrscheinlichkeit, dass der Übergang zwischen t und } t + \Delta \text{ passiert} \\ &\quad \text{unter der Bedingung, dass er bis } t \text{ noch nicht eingetreten ist} \end{aligned}$$

Aus den Definitionen ergeben sich die folgenden Beziehungen:

$$
\begin{aligned}
r(t) &= \lim_{\Delta \to 0} \frac{P(t \leq T < t + \Delta | T \geqslant t)}{\Delta} = \lim_{\Delta \to 0} \frac{1}{\Delta} \frac{P(t \leq T < t + \Delta)}{P(T \geqslant t)} \\
&= \frac{f(t)}{S(t)}
\end{aligned}
$$

Andererseits ist

$$
\begin{aligned}
\frac{d}{dt} [\log S(t)] &= \frac{S'(t)}{S(t)} = \frac{-F'(t)}{S(t)} = \frac{-f(t)}{S(t)}, \quad \text{also} \\
\frac{d}{dt} [\log S(t)] &= -r(t)
\end{aligned}
$$

Es folgt durch Integration:

$$
\begin{aligned}
\log S(t) &= \int_0^t -r(u) du, \\
S(t) &= \exp \left[- \int_0^t r(u) du \right]
\end{aligned}
$$

Folgerung:

> **Die Survivalfunktion ist durch die Hazardfunktion vollständig bestimmt. Daher ist es im Prinzip möglich, die Untersuchung des Prozesses über die Untersuchung der Hazardfunktion (Modellierung, Schätzung) zu führen.**

2.1.2. Allgemeiner Fall: Multiple Übergänge

Je Beobachtungseinheit werden eine oder mehrere aufeinanderfolgende *Episoden* betrachtet. Deren Grenzen (Beginn, Ende) sind durch die Zeitpunkte der einzelnen Übergänge gekennzeichnet. Es sei

$\mathcal{O}$ die Menge der möglichen Ausgangszustände einer Episode ($\mathcal{O}$ steht für *Origin)*

$\mathcal{D}$ die Menge der möglichen Zustände nach Ende einer Episode ($\mathcal{D}$ steht für *Destination).* Darin ist auch der Fall eingeschlossen, dass wegen "Zensierung" am Ende der Episode noch keine Änderung eingetreten ist.

Für eine Episode , die im Zustand $\mathcal{O} = j$ beginnt, sei

$$\begin{aligned} T_j &= \text{der Zeitpunkt des nächsten Übergangs} \\ D_j &= \text{der Zustand nach dem nächsten Übergang} \\ t_s &= \text{Zeitpunkt des Beginns der Episode} \\ t &= \text{die "Prozesszeit"} \end{aligned}$$

Dann wird die *Rate* oder *Übergangsintensität* für den Wechsel von Zustand j auf Zustand k $(k \neq j)$ zum Zeitpunkt t $(t \geq t_s)$ definiert durch:

$$r_{jk}(t) = \lim_{\Delta \to 0} \frac{P(t \leq T_j < t + \Delta, \quad D_j = k \,|T_j \geqslant t)}{\Delta}$$

Man geht hier also von der Wahrscheinlichkeit aus, dass zwischen t und $t+\Delta$ ein Übergang passiert und *zwar in den Zustand j,* unter der Bedingung, dass bis zur Zeit t noch kein Wechsel stattgefunden hat.
Die Hazardrate für *irgendeinen Wechsel* setzt sich dann aus diesen spezifischen Übergangsintensitäten einfach als Summe zusammen:

$$\begin{aligned} r_j(t) &= \lim_{\Delta \to 0} \frac{P(t \leq T_j < t + \Delta \,|T_j \geqslant t)}{\Delta} \\ &= \sum_{k \neq j} \lim_{\Delta \to 0} \frac{P(t \leq T_j < t + \Delta, \quad D_j = k \,|T_j \geqslant t)}{\Delta} \\ &= \sum_{k \neq j} r_{jk}(t) \end{aligned}$$

Es gibt auch eine analoge Zerlegung der Survivalfunktion (definiert für den Fall, dass der Zustand j zur Zeit $t = 0$ beginnt):

$$
\begin{aligned}
S_j(t) &= P(T_j > t) = \exp\left[-\int_0^t r(u)du\right] \\
&= \exp\left[-\sum_{k\neq j}\int_0^t r_{jk}(u)du\right] = \prod_{k\neq j}\exp\left[-\int_0^t r_{jk}(u)du\right] \\
&= \prod_{k\neq j} S_{jk}(t)
\end{aligned}
$$

wobei die $S_{jk}(t) = \exp\left[-\int_0^t r_{jk}(u)du\right]$ als "*Pseudo-Survival-Funktionen*" bezeichnet werden.

Analog gibt es auch eine Definition für die "*Pseudo-Dichte-Funktionen*":

Wie wahrscheinlich ist es im Zustand j, dass im Intervall $[t, < t+\Delta)$ ein Übergang in den Zustand $k \neq j$ stattfindet (wenn der Zustand j zur Zeit $t = 0$ beginnt)? Diese Wahrscheinlichkeit ist

$$
\begin{aligned}
& P(t \leq T_j < t+\Delta, \quad D_j = k\ ) \\
=\ & P(t \leq T_j < t+\Delta, \quad D_j = k \mid t \leq T_j)P(\ t \leq T_j) \\
=\ & P(t \leq T_j < t+\Delta, \quad D_j = k \mid t \leq T_j)\ S_j(t)
\end{aligned}
$$

Daraus folgt:

$$
\begin{aligned}
f_{jk}(t) &= \lim_{\Delta\to 0}\frac{P(t \leq T_j < t+\Delta, \quad D_j = k\ )}{\Delta} \\
&= r_{jk}(t)\ S_j(t)
\end{aligned}
$$

Insgesamt wird deutlich, dass sich das gesamte Geschehen auch hier aus Hazardraten, nun aber den spezifischen Raten für die einzelnen Übergänge, zusammensetzt. Modellannahmen und Parameterschätzungen werden sich daher auf diese "Übergangsintensitäten" beziehen.

Unterdessen ist eine Modellannahme in die Formulierung eingegangen:

Die Übergangsintensitäten können von der Zeit abhängen, sind aber **unabhängig von der Vorgeschichte des Prozesses**.

Diese Modellannahme kann aber in der Anwendung des Programmes wieder abgeschwächt werden.

2.2. Modellschätzung

2.2.1. Likelihood und parametrische Modelle

Die "Likelihood" einer Stichprobe ist die Wahrscheinlichkeit (oder Wahrscheinlichkeitsdichte) für die Werte der Stichprobe unter den vorliegenden Modellannahmen. Die *Schätzung der Modellspezifizierungen* erfolgt so, dass die Spezifizierungen so gewählt werden, dass die davon abhängige Likelihood *maximiert* wird.
In TDA bilden die einzelnen *Episoden* die Stichprobenelemente. (Mehrfache Episoden je Patient können dadurch berücksichtigt werden, dass die Patientenidentifikation und die interne laufende Nummer der Episoden erfasst werden. Eine alternative Vorgehensweise wird in den Anwendungsbeipielen dargestellt.)
Zur Formulierung der Likelihhod benötigen wir noch die folgenden Bezeichnungen:

$\mathcal{E}_{jk}$: Die Menge der Übergänge von j nach k $(j \neq k)$

$\mathcal{Z}_j$: Zensierung im Zustand j

$\mathcal{N}_j := \mathcal{E}_{jk} \cup \mathcal{Z}_j$

i : die Nummer der Episode $(i =, \ldots, n)$

t_i : die Zeitpunkte der Übergänge bzw. der Zensierung nach Beginn der Episode

Dann ist die Likelihood der Stichprobe:

$$
\begin{aligned}
L &= \prod_{j,k} \prod_{i \in \mathcal{E}_{jk}} f_{jk}(t_i) \prod_{i \in \mathcal{Z}_j} S_j(t_i) \\
&= \prod_{j,k} \prod_{i \in \mathcal{E}_{jk}} r_{jk}(t_i) S_j(t_i) \prod_{i \in \mathcal{Z}_j} S_j(t_i) \\
&= \prod_{j,k} \prod_{i \in \mathcal{E}_{jk}} r_{jk}(t_i) \prod_{i \in \mathcal{N}_j} S_j(t_i) \quad \text{mit} \\
S_j(t_i) &= \exp \left[- \int_0^{t_i} r_{jk}(u) du \right]
\end{aligned}
$$

Hier erscheinen die Übergangsintensitäten r_{jk} an den Zeitpunkten t_i, aber auch im Integranden. In parametrischen Modellen, in denen $r_{jk}(t)$ als parametrische Funktion dargestellt werden, resultiert aus den Integralen wiederum eine parametrische Funktion, so dass L insgesamt als Funktion der Modellparameter erscheint, nach denen L maximiert werden kann.

2.2.2. Semiparametrische Modelle und Partielle Likelihood

Im "*Proportional Hazard*" Modell nach Cox wird folgende Modellannahme getroffen:

$$r_{jk}(t) = h_{jk}(t)\ \exp\left[\sum_l \beta^l_{jk} X_{il}\right]$$

Dabei ist

$r_{jk}(t)$ eine für den Übergang von j nach k spezifische "*Baseline-*" Hazardfunktion, die *nicht parametrisiert* wird, also beliebige Formen annehmen kann

X_{il} der Wert der Kovariablen X_l in der Epsode i und

β^l_{jk} der Regressionskoeffizient zur Kovariablen X_l für den Übergang $j \to k$ (Modellparameter)

In Modellen mit zeitabhängigen Kovariablen kann X_l auch noch vom Zeitpunkt t der Beobachtung abhängen.
Es wird also zunächst davon ausgegangen, **dass für jeden Übergang $j \to k$ sowohl eine eigene "Baseline-Hazardfunktion" als auch übergangsspezifische Modellparameter β^l_{jk} bereitgestellt werden.**
Für die Likelihood ergibt sich dann die folgende Darstellung:

$$\begin{aligned}
L &= \prod_{j,k} \prod_{i \in \mathcal{E}_{jk}} r_{jk}(t_i) \prod_{i \in \mathcal{N}_j} S_j(t_i) \\
&= \prod_{j,k} \prod_{i \in \mathcal{E}_{jk}} h_{jk}(t_i)\ \exp\left[\sum_l \beta^l_{jk}\ X_{il}\right] \quad \prod_{i \in \mathcal{N}_j} S_j(t_i) \\
&= \prod_{j,k} \prod_{i \in \mathcal{E}_{jk}} h_{jk}(t_i)\ \exp\left[\sum_l \beta^l_{jk}\ X_{il}\right] \quad \prod_{i \in \mathcal{N}_j} \exp\left[-\int_0^{t_i} r_{jk}(u)du\right] \\
&= \prod_{j,k} \prod_{i \in \mathcal{E}_{jk}} h_{jk}(t_i)\ \exp\left[\sum_l \beta^l_{jk}\ X_{il}\right] \quad \prod_{i \in \mathcal{N}_j} \exp\left\{-\int_0^{t_i} h_{jk}(u)\ \exp\left[\sum_l \beta^l_{jk}\ X_{il}\right] du\right\}
\end{aligned}$$

In dieser Form erscheint die Maximierung der Likelihoodfunktion zunächst aussichtslos, da die Parameter β^l_{jk} mit der Funktion $h_{jk}(u)$ verknüpft sind. Die Lösung erfolgt üblicherweise dadurch, dass man eine *getrennte* Maximierung der Baseline-Hazardfunktionen $h_{jk}(t)$ und der Modellparameter β^l_{jk} durchführt: Man betrachtet die "*Partielle Likelihood*", eine Funktion, in der nur noch die Parameter β^l_{jk} erscheinen. Die Herleitung wird hier nur für den Fall des einfachen Cox-Modells (mit nur *einem* Übergang) und ohne zensierte Beobachtungen angedeutet (Näheres siehe Kalbfleisch und Prentice, S. 71 ff., S. 167 ff.):

1. *Invarianz*: Seien $t_{(1)} < t_{(2)} < \ldots < t_{(i)} < \ldots < t_{(n)}$ die geordneten Zeitpunkte der beobachteten Übergänge, und
   $X_l^{(i)}$ die Kovariablenwerte aus den zugehörigen Beobachtungseinheiten.
   Man betrachte dann irgendeine monotone Transformation auf der Zeitachse, in der also die *Reihenfolge* der Übergangszeiten $t_{(i)}$ erhalten bleiben
   Diese Transformation würde die Schätzung der Modellparameter β_l *nicht verändern,* da sie innnerhalb der Hazardfunktion

$$r(t_{(i)}) = h_0(t_{(i)})\ \exp\left[\sum_l \beta_l X_l^{(i)}\right]$$

   durch die *Veränderung der Baselinerate* $h_0(t)$ aufgefangen wird.

2. Daraus folgt, dass für die Schätzung von β_l die *Ränge* der Beobachtungen bereits suffizient sind. Man berechne also die Wahrscheinlichkeit für die gegebene Rangfolge:

$$P(\text{Beobachtung Nr. } j \text{ mit Kovariablenwert } \mathrm{X}_{lj}$$
$$\text{hat bezüglich der Übergangszeit den Rang } r_j)$$

3. Die Formel hierfür lässt sich sequentiell herleiten. In den einzelnen Schritten geht dabei folgende Beziehung ein:
   Gegeben bis zur Zeit $t_{(j)}$ sind genau j Ereignisse eingetreten. Dann ist die Wahrscheinlichkeit dafür, dass Beobachtung Nr. i den nächsten Übergang liefert, proportional zu $\exp\left[\sum_l \beta_l X_{li}\right]$, also gleich

$$\frac{\exp(\sum_l \beta_{li}\ X_{li})}{\left[\sum_{k\in R(t_{(i)})} \exp \sum_l (\beta_{lk}\ X_{lk})\right]}$$

   wobei $R(t_{(j)})$ die "Risikomenge ab $t_{(j)}$" bezeichnet, also alle Beobachtungseinheiten die unmittelbar nach $t_{(j)}$ noch unter Beobachtung stehen.

4. Insgesamt ergibt sich daraus die Wahrscheinlichkeit für die beobachtete Rangfolge als

$$L = \frac{\exp(\sum_i \sum_l \beta_{li}\ X_{li})}{\prod_{i=1}^{n} \left[\ \sum_{k\in R(t_{(i)})} \exp \sum_l \beta_{lk}\ X_{lk})\right]}$$

Dies ist die Basisformel zur Berechnung der partiellen Likelihood. In dieser Form kann sie auch auf den allgemeinen Fall multipler Übergänge übertragen werden.

Ebenso ist es in dieser Form möglich, zeitabhängige Kovariablen zu berücksichtigen. Diese sind dann mit den Regressionskoeffizienten in der Form $\sum_l \beta_{li}\ X_{li}(t_{(i)})$ verbunden.
Das bedeutet, dass die Werte der zeitabhänigen Kovariablen nur für die Zeitpunkte der Übergänge zur Verfügung gestellt werden müssen.
Näheres zur Maximierung der partiellen Likelihood findet man im TDA-Manual unter 6.17.7.1.

3. TDA Programmbeschreibung

3.1. Übersicht

TDA ist ein Programm, das die Analyse multipler Übergänge im oben beschriebenen Sinne ermöglicht. Daneben stellt es aber noch eine Reihe weiterer statistischer Prozeduren zur Verfügung.
Für die Analyse von Übergangsdaten benötigt man:

- Eine Inputdatei mit den zu analysierenden Episoden und Übergängen (ASCII oder SPSS)
- Eine Syntaxdatei, in der die gewünschten Prozeduren mit Optionen und Spezifizierungen stehen (im ASCII-Format)
- ein DOS-Fenster, in dem das Programm unter Angabe der Syntaxdatei `(infname)` und einer Outputdatei `(outfname)` aufgerufen wird

Der Aufruf erfolgt mit dem Kommando

```
tda cf=(infname) > (outfname)
```

Darin steht `cf` für "command file". In dieser Version wird ein bereits vorhandener Output file überschrieben. In der Version

```
tda cf=(infname) >> (outfname)
```

wird ein bereits vorhandener Outputfile um den neuen Output ergänzt.
Die Ergebnisdatei enthält neben einer Beschreibung der Übergängeg $j \rightarrow k$ und deren Häufigkeiten u.a.:

- Die Likelihoodschätzungen der Regressionskoeffizienten zusammen mit deren asymptotischen Standardfehlern
- Daraus abgeleitet einen asymptotischen Test gegen den Wert 0 für jeden Koeffizienten. (Achtung: In der Spalte "`Signif''` steht nich der "P-Wert" sondern 1-P-Wert; für $\alpha = 0.05$ ist der Test also ab 0.95 signifikant.)
- Die (partielle) Likelihood vor und nach Modellanpassung. Daraus sind die Vergleiche zweier Modelle ableitbar, in denen das eine durch zusätzliche Spezifizierung aus dem anderen hervorgeht.

Darüber hinaus kann eine Datei mit den Pseudo-Survivalfunktionen $S_{jk}(t)$ und den kumulativen Hazards $\int_0^t r_{jk}(u)du$ erzeugt werden.

3.2. Die Inputdatei

Jede Zeile der Inputdatei beschreibt eine *Episode*. Als Variable *müssen* darin enthalten sein:

- Der Zeitpunkt des *Beginns* der Episode
- Der Zeitpunkt des *Endes* der Episode (auch bei zensierten Daten)
- Der Status bei *Beginn* der Episode
- Der Status bei *Ende* der Episode

Je nach Fragestellung werden weiterhin benötigt:

- Zeitunabhängige Kovariablen
- Kovariablen, aus denen im Programm eine zeitabhängige Funktion erzeugt werden kann (dies entspricht dem Vorgehen in Standardprogrammen zur Cox-Regression und wird später an Beispielen erläutert).
- Patientenidentifizierung und laufende Nummer der Episode

3.3. Die Kommandos

3.3.1. Daten Einlesen

Bei SPSS-Dateien wird mit dem Kommando

```
rspss1 (optionen) = (SPSS-Dateiname)
```

die SPSS-Datei eingelesen.

Achtung: TDA ersetzt fehlende Angaben durch den Wert -5 und bezieht sie mit diesem numerischen Wert in die Analyse ein. Es besteht derzeit keine Möglichkeit, auf TDA-Ebene die entsprechenden Episoden aus der Analyse auszuschließen (dieses ist nur beim Einlesen von ASCII-Datien möglich). Der Benutzer muss also vorher auf SPSS-Ebene alle Episoden ausschließen, die in einer der später benutzten Variablen Missings haben.

3.3.2. Definition der Episoden

Mit dem Kommando

```
edef ( Spezifikationen)
```

wird definiert, welche Variablen den Zeitpunkt und den Status von Beginn und Ende einer Episode beschreiben.

Beispiel:

```
edef (ts = BEGINNE,
 tf = ENDEE,
 org = ORIGIN,
 des = DESTIN,
 );
```

Innerhalb dieses Kommandos werden auch die zeitabhängigen Kovariablen definiert. Dazu wird die "Prozesszeit" als Variable `time` zur Verfügung gestellt. Soll beispielsweise die Variable X bis zum Zeitpunkt 5 den Wert X1 und danach den Wert X2 annnehmen, so definiert man:

```
X = X1*le(time,5)+ X2* gt(time,5)
```

Darin ist z.B. `le(time,5)` eine Indikatorfunktion für die Beziehung `time`$\leq$ 5.

3.3.3. Modellwahl und Spezifikationen

Mit dem Kommando

```
rate (Spezifikationen) = (n)
```

werden u.a. die Kovariablen definiert und wird mit der Nummer (n) wird das Modell ausgewählt. Dabei kennzeichnet z.B. die 1 das Cox-Modell.

Beispiel:

```
rate (xa(0,1) = GESCHL,V1,
  xa(0,2) = GESCHL,V1,
  xa(0,3) = GESCHL,V1,
  xa(1,2) = GESCHL,V1,
  xa(1,3) = GESCHL,V1,
  xa(1,4) = GESCHL,V1,
  xa(2,3) = GESCHL,V1,
  xa(2,4) = GESCHL,V1,
  xa(3,4) = GESCHL,V1,
 ) = 1; # 1 bedeutet Cox-Regression
```

Darin bedeutet z.B. `xa(0,1) = GESCHL,V1,` dass für den Übergang von Zustand 1 nach Zustand 2 die Variablen `GESCHL` und `V1` als Kovariablen eingesetzt werden.

3.4. Anwendungen

Ergänzend zu diesem Skript wird eine vollständige Kommandodatei mit dem Namen `test5in.cf` und mit (sparsamer) Kommentierung zur Verfügung gestellt. Den damit erzeugten Output findet man unter dem Namen `test5out.txt`. Die dabei benutzte SPSS-Datei `tda5.sav` wird ebenfalls zur Verfügung gestellt. Wenn Sie das Programm `tda.exe` in Ihrem Vezeichnis zur Verfügung haben, können Sie mit dem Befehl

```
tda cf=test5in.cf > test.out
```

das Programm starten und unter dem neuen Namen `test.out` den Output noch einmal erzeugen. (*Achtung*: kein Blank vor oder nach dem Gleicheitszeichen!!) In der Datei `test5in.cf` muss ggf. zuvor noch der Pfadname für die SPSS-Datei angepasst werden.

4. Schlussfolgerung

Aus den beigefügten Anwendungsdateien wird deutlich: TDA ist ein flexibles Programm zur Analyse von Übergängen, in dem auch zeitabhängige Kovariablen berücksichtigt werden können. Es kann Hinweise darüber liefern, welche Variablen bei der Veränderung von Zuständen in den beobachteten Prozessen eine Rolle spielen und als prognostische Faktoren für Veränderungen gewertet werden können.