

Adjustierung für die Sommerzeit

Guido Wendland
 Grüenthal GmbH
 52099 Aachen
 guido.wendland@grunenthal.com

Zusammenfassung

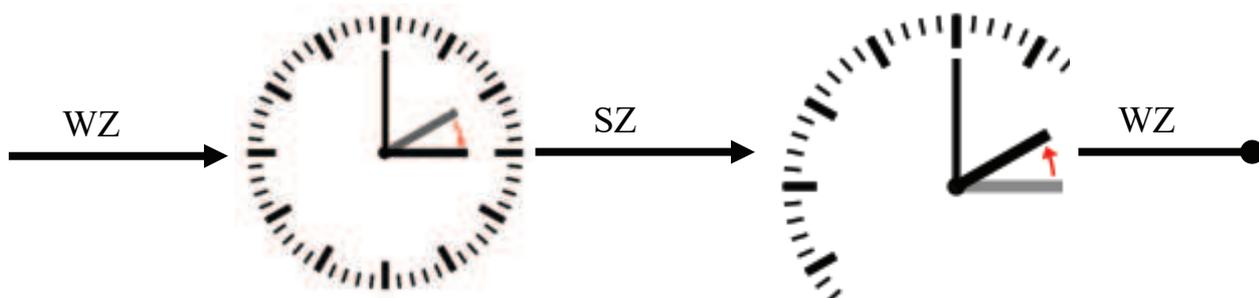
Bei der Analyse von Längsschnittdaten-wie pharmakokinetischen Daten-, bei denen es eher auf Minuten und Stunden als auf Tage ankommt, muss die Umstellung zwischen Winter- und Sommerzeit berücksichtigt werden.

Es wird unter Verwendung verschiedener SAS Codes gezeigt, wie Einflussgrößen wie Region und Jahr einbezogen werden können, wenn bestimmt wird, ob sich ein Ereignis während der Sommerzeit oder Winterzeit ereignet hat.

Schlüsselwörter: Winterzeit (WZ), Sommerzeit (SZ), Zeitumstellung

1 Einleitung

Die Sommerzeit (SZ) ist die Zeit des Jahres, in der die Uhrzeit eine Stunde vor der normalen Zeit, auch als Winterzeit (WZ) bezeichnet, verschoben wird.



In Regionen, die keine Zeitumstellung haben, gilt das ganze Jahr als WZ.

Bei der Analyse von Veränderungen im Laufe der Zeit, die mit einer Genauigkeit von Stunden oder sogar Minuten gemessen werden, kann die SZ berücksichtigt werden müssen, wie z.B. in pharmakokinetischen Studien. Die Unterschiede zwischen zwei Datumszeitangaben – eine während der SZ und eine während der WZ - müssen angepasst werden, indem man eine Stunde von der Datumszeit abzieht, die in die SZ fällt.

Ich veranschauliche, wie verschiedene Hindernisse bei der Bestimmung des Zeitraums (SZ oder WZ) durch einen SAS-Makro angegangen werden können.

2 Das Makro %_DST

Das hier vorgestellte SAS-Makro %_dst (DST steht für “daylight saving time“) kann z.B. auf klinische Studiendaten angewendet wird. Das Makro bestimmt, ob sich ein oder mehrere Ereignisse (mit Datumzeitangaben im Charakter- oder Zahlenformat) eines Patienten während der SZ oder WZ befinden und ergänzt diese Informationen in einem Output-Datensatz. Es gibt dabei zwei Faktoren, die berücksichtigt werden müssen: Region und Jahr.

2.1 Regionen

Länder in der Nähe des Äquators haben oft keine Zeitumstellung, während andere Länder ihre Uhrzeit in unterschiedlichen Wochen umstellen. Abbildung 1 illustriert dies für das Jahr 2018:

Country	Regions/States	DST Start Date	DST End Date
Belgium	All locations	Sunday, March 25	Sunday, October 28
Japan	All locations	No DST in 2018	
New Zealand	All locations	Sunday, September 30	Sunday, April 1
United States	Most locations	Sunday, March 11	Sunday, November 4
	Hawaii, most of Arizona	No DST in 2018	

Abbildung 1: Sommerzeiten im Jahr 2018 in ausgewählten Ländern

In Abbildung 1 wird Belgien als Musterland für Europa (mit einheitlicher Zeitumstellung, Stand 2019) gewählt. Abbildung 1 zeigt auch, dass Länder, wie z.B. die USA, manchmal unterschiedliche Zeitumstellungen haben. Das kann problematisch werden, wenn nur das Land, aber nicht die Region, bekannt ist.

2.2 Jahre

Die Zeitumstellung findet derzeit an einem bestimmten Sonntag statt, wie nachfolgend für Europa und die USA zu sehen ist:



Um die beiden Einflussfaktoren, Region und Jahr, zu berücksichtigen, verwendet das Makro einen Referenzdatensatz. Dieser enthält das Datum und die Uhrzeit für den Start und das Ende der SZ für alle benötigten Regionen und Jahre.

Der Standard-Referenzdatensatz, der die SZ-Start- und Endzeiten Europas und der USA enthält, wird innerhalb des Makros erstellt. Es kann aber auch ein abweichender Referenzdatensatz separat erstellt und im Makro verwendet werden, wenn z.B. andere Regionen berücksichtigt werden müssen.

_dst	USA 			EU 		
	Start		End	Start		End
_year	start1		end1	start2		end2
2016	2016-03-13T02:00:00		2016-11-06T02:00:00	2016-03-27T02:00:00		2016-10-30T02:00:00
2017	2017-03-12T02:00:00		2017-11-05T02:00:00	2017-03-26T02:00:00		2017-10-29T02:00:00
2018	2018-03-11T02:00:00		2018-11-04T02:00:00	2018-03-25T02:00:00		2018-10-28T02:00:00
2019	2019-03-10T02:00:00		2019-11-03T02:00:00	2019-03-31T02:00:00		2019-10-27T02:00:00

Abbildung 2: SZ Startzeiten und Endzeiten für ausgewählte Jahre in den USA und Europa (Ausschnitt der Jahre 2016-2020)

3 Parameter des Makros %_DST

Das Makro %_dst verwendet die folgenden fünf Makro-Parameter:

```
%_dst(ds      =, /* Input SAS Datensatz */
      dsout   =, /* Output SAS Datensatz */
      var     =, /* Variable(n), die überprüft werden */
      refds   =, /* Referenz Datensatz */
      region =); /* (E oder U oder fehlend.) */
```

Die Parameter DS und VAR müssen angegeben werden. Der Parameter VAR kann eine Kombination aus Datum/Uhrzeit Variablen im character Format (ISO8601-Format) und numerischen Format sein.

Der Parameter DSOUT ist der Output-Datensatz, bei dem im Vergleich Input-Datensatz pro Datumszeitvariable zwei Indikatorvariablen (je eine numerisch und character) ergänzt werden.

Der Parameter REFDS ist der Referenzdatensatz, der nur dann zur Verfügung gestellt werden muss, wenn andere Regionen als Europa oder die USA in den Input-Datensatz aufgenommen werden. Sonst wird der Standard-Referenzdataset _dst innerhalb des Makros erstellt und verwendet.

Der Parameter REGION muss nur dann angegeben werden, wenn die Region-Informationen nicht aus den Datensatzvariablen "usubjid" oder "region0" abgeleitet werden. Falls der Parameter angegeben ist, müssen alle Patienten aus der gleichen Zeitumstellungszone kommen. Ansonsten extrahiert das Makro die Ländercode-Informationen, die in der Variable "usubjid" enthalten sind, oder verwendet eine Standard-Variable ("region0"), welche die Region enthält. Diese muss dem Input-Datensatz gegebenenfalls hinzugefügt werden.

4 Einige Code Highlights im Makro %_DST

4.1 Erstellung des Referenz SAS-Datensatz _dst

Wie bereits gezeigt, startet und endet die SZ in der EU und in den USA an festgelegten Sonntagen. Im Folgenden wird beispielsweise die Stunde des zweiten Sonntags um 2:00 (Variable z1) programmatisch bestimmt (Start der SZ in den USA)

Erster tag im März (x1):

```
x1=mdy(3,1,year);
```

z.B. 03/01/2018

Inkrement für den 2ten Sonntag (y1):

```
y1=(weekday(x1) ne 1) +1;
```

Sonntag ist Wochentag1!

Datum->Datum-Zeit Formatierung:

```
z1=put(86400*intnx("week",x1,y1)+7200,E8601DT.);
```

Ein Auszug aus diesem Referenzdatensatz ist in Abbildung 2 zu sehen.

4.2 Bestimmung der Variablen und Variable Typen im Input SAS-Datensatz

Nehmen wir zur Veranschaulichung an, dass wir den folgenden Input SAS-Datensatz haben mit jeweils 2 Patienten aus Deutschland (EU) und 2 Patienten aus den USA.

Input SAS-Datensatz

	usubjid	dt1dtc	dt2n	
KF101	DE1001001	2017-03-25T03:25	1806024300	=2017-03-25T01:25
KF101	US2001001	2017-03-25T01:25	1806031500	=2017-03-25T03:25
KF101	DE1002001	2018-03-25T03:25	1837560300	=2018-03-25T01:25
KF101	US2001002	2018-03-25T01:25	1837567500	=2018-03-25T03:25

Darüber hinaus gibt es zwei Datumsvariablen – dt1dtc (character) und dt2n (numerisch) – die auf SZ überprüft werden sollten. In der numerischen Darstellung der Variable dt2n werden die Zeitpunkte der character Variable dt1dtc zu Illustrationszwecken ausgetauscht, wie von den Pfeilen angegeben.

In den folgenden zwei Schritten werden zunächst zwei lokale Makrovariablen &_VARS und &_VARTYPES erstellt, die später verwendet werden.

```
data _vcolumn;
  set sashelp.vcolumn (where=(upcase(libname)="%upcase(&_libname.)"
    and upcase(memname)="%upcase(&_memname.)"
    and upcase(name) in
    (%f_list(list=%upcase(&VAR.),quotes=Y)));
  keep libname memname name type;
run;
```

Der obige Code verwendet die lokalen Makro-Variablen `&_libname` und `&_memname`, die den Libref und Namen des Datensatzes darstellen. Für temporäre Datensätze wird der Libref auf "WORK" gesetzt. Das verwendete Makro `%f_list` setzt alle angegebenen Variablen (`&VAR`) in doppelte Anführungszeichen, wie für den IN-Operator benötigt. Der resultierende Datensatz `_vcolumn` sieht wie folgt aus:

Library Name	Member Name	Column Name	Column Type
WORK	DS2	dt1dte	char
WORK	DS2	dt2n	num

Temporary dataset `_vcolumn`

Dieser SAS-Datensatz wird verwendet, um die lokalen Makro-Variablen `"&_VARS=dt1dte dt2n"` und `"&_VARTYPES = char num"` mittels folgender SQL Anweisung zu generieren:

```
proc sql noprint;
  select name, type
  into :_vars separated by " ",
       :_vartypes separated by " "
  from _vcolumn;
quit;
```

Diese beiden Makrovariablen werden nachfolgend in `%do`-Schleifen verwendet.

4.3 Bestimmung des Referenzjahres und der Region für jede Variable und jede Beobachtung

Um den Referenzdataset `_dst` mit dem Input-Datensatz zu verknüpfen, müssen wir die Regionen für jede Beobachtung und die Jahre für die beiden angegebenen Zeitvariablen `dt1dte` und `dt2n` bestimmen. Die Ableitung des Jahres hängt von dem Variablentyp ab:

```
_yearref1=input(substr(dt1dte,1,4),best.);   Character Datumsvariable
_yearref2=year(datepart(dt2n));              Numerische Datumsvariable
```

Die variable `USUBJID`, die Patienten-ID, enthält einen Ländercode, der wiederum in der folgenden Liste erscheint (Tabelle 1). Dadurch ist der Beginn und das Ende der SZ festgelegt.

Tabelle 1: Sommerzeit (SZ) nach Landesvorwahl und Region

Region	Ländercodes (gemäß ISO 3166)	Start und Ende der SZ
E(U)	AX, AL, AD, BE, BA, BG, DK, DE, EE, FO, FI, SF, FR, GI, GR, GL, GG, IM, IE, IT, JE, XK, HR, LV, LB, LI, LT, LU, MT, MA, MK, MD, MC, ME, NL, NO, AT, PL, PT, RO, SM, SE, CH, RS, SK, SI, ES, CZ, UA, HU, VA, UK, EH, CY	Letzter Sonntag im März bis zum letzten Sonntag im Oktober
U(S)	BS, BM, HT, CA, CU, TC, US	2ter Sonntag im März bis zum ersten Sonntag im November

Als Ergebnis erhalten wir einen intermediären SAS-Datensatz, der die Jahre pro Datums-Zeit-Variable und die Regionen für jeden Beobachtung im Input-Datensatz enthält.

Darüber hinaus werden für den späteren Vergleich beide Zeitvariablen nun im Character-Format dargestellt.

```
dt1dtcc=dt1dtc;
dt2nc =substr(put(dt2n,e8601dt.),1,16);
```

	usubjid	dt1dtc	dt2n	_yearref1	dt1dtcc	_yearref2	dt2nc	region
KF101	DE1001001	2017-03-25T03:25	1806024300	2017	2017-03-25T03:25	2017	2017-03-25T01:25	E
KF101	US2001001	2017-03-25T01:25	1806031500	2017	2017-03-25T01:25	2017	2017-03-25T03:25	U
KF101	DE1002001	2018-03-26T03:25	1837560300	2018	2018-03-26T03:25	2018	2018-03-25T01:25	E
KF101	US2001002	2018-03-26T01:25	1837567500	2018	2018-03-26T01:25	2018	2018-03-25T03:25	U

4.4 Bestimmung für jede Beobachtung und für jede Datumsvariable, ob sie sich in der Sommerzeit befindet.

Um die SZ in den verschiedenen Jahren anzuwenden, wird der Referenzdatensatz (voreingestellt ist `_dst`, siehe Abbildung 2) verwendet, um die entsprechenden Start- und Endzeiten der SZ mit der Kombination aus Jahr und Region im intermediären Datensatz zu verknüpfen.

Das Referenzjahr wird für jede verwendete Datumszeit-Variable separat bestimmt (Variablen: `_yearref(i)`), die Region für jede Beobachtung (Variable: `Region`). Es wird jeweils ein Start- und eine Endzeitpunkt für die beiden Datumzeit-Variablen beibehalten.

Im folgenden Code-Ausschnitt enthält die lokale Makro-Variable `_CNT` die Anzahl der beteiligten Zeitvariablen und die lokale Variable `_NOBS` die Anzahl der Regionen. Der intermediäre Datensatz `refdate0&_cnt` wird iterativ überschrieben bis alle spezifizierten Datumsvariablen abgearbeitet sind. Für jede Region wird jeweils ein Variablen-Paar für den Start and das Ende zum neuen Datensatz hinzugefügt.

```

%let _cnt=1;
%do %while(%qscan(&_vars.,&_cnt,%str( )) ne %str( ));
  %let _cnt1=%eval(&_cnt.+1);
  proc sql;
    create table _refdate0&_cnt1. as
      select a.*,%do _i=1 %to &nobs1;
        start&_i. as start&_i.&_cnt.,
        end&_i. as end&_i.&_cnt.,
        %end;
        b._year as _year_&_cnt.
      from _refdate0&_cnt. a left join &REFDS. b
      on a._yearref&_cnt.=b._year;
  quit;
  %let _cnt=%eval(&_cnt+1);
%end;

```

Der daraus resultierende Datensatz – `_refdate03` in unserem Beispiel mit 2 Datumsvariablen – enthält 5 zusätzliche Hilfs-Variablen für jede Input-Variable:

Zusätzlich generierte Variablen zur Datumsvariable `d1 dtc`:

start1_1	end1_1	start2_1	end2_1	_year_1
2017-03-12T02:00:00	2017-11-05T02:00:00	2017-03-26T02:00:00	2017-10-29T02:00:00	2017
2017-03-12T02:00:00	2017-11-05T02:00:00	2017-03-26T02:00:00	2017-10-29T02:00:00	2017
2018-03-11T02:00:00	2018-11-04T02:00:00	2018-03-25T02:00:00	2018-10-28T02:00:00	2018
2018-03-11T02:00:00	2018-11-04T02:00:00	2018-03-25T02:00:00	2018-10-28T02:00:00	2018

Zusätzlich generierte Variablen zur Datumsvariable `dt2n`:

start1_2	end1_2	start2_2	end2_2	_year_2
2017-03-12T02:00:00	2017-11-05T02:00:00	2017-03-26T02:00:00	2017-10-29T02:00:00	2017
2017-03-12T02:00:00	2017-11-05T02:00:00	2017-03-26T02:00:00	2017-10-29T02:00:00	2017
2018-03-11T02:00:00	2018-11-04T02:00:00	2018-03-25T02:00:00	2018-10-28T02:00:00	2018
2018-03-11T02:00:00	2018-11-04T02:00:00	2018-03-25T02:00:00	2018-10-28T02:00:00	2018

In diesem Beispiel sind die Werte der zusätzlichen Variablen für beide Datumzeitvariablen identisch, da das Jahr der beiden Zeitvariablen identisch ist (`_year_1=_year_2`).

Im letzten Schritt werden die enthaltenen Region-Informationen verwendet, um das zur jeweiligen Region gehörige Paar aus Start- und Endezeitpunkt auf Beobachtungsebene auszuwählen.

Im Standard-Referenzdatensatz wird das erste Paar (`start1_x` und `end1_x`, $x \in \{1,2\}$) für die Region EU verwendet und das zweite Paar (`start2_x` und `end2_x`, $x \in \{1,2\}$) für die Region USA.

Die Datum/Zeit Variablen werden mit dem jeweiligen Variablenpaar für den Start und das Ende der SZ verglichen. Das Ergebnis dieses Vergleichs wird mit einer character und einer numerischen Indikatorvariable im Output-Datensatz gespeichert, indem zu der Variable der Suffix `_dst` (character) und `_dstn` (numerisch) hinzugefügt wird. Die Hilfs-Variablen werden nun nicht mehr benötigt und deswegen gelöscht.

Der Output-Datensatz (Parameter DSOUT) des Makros %_dst sieht beispielsweise wie folgt aus:

USUBJID	DT1DTC	DT2N	DT1DTC_DST	DT2N_DST	REGION	DT1DTC_DSTN	DT2N_DSTN
KF101	DE001001	2017-03-25T03:25	1806024300	SST	E	0	0
KF101	US2001001	2017-03-25T01:25	1806031500	DST	U	-1	-1
KF101	DE002001	2018-03-26T03:25	1837560300	DST	E	-1	0
KF101	US2001002	2018-03-26T01:25	1837567500	DST	U	-1	-1

5 Fehlerbehandlung

Alle Makro-Inputparameter werden überprüft, wenn das Makro ausgeführt wird.

Folgende Prüfungen werden für das Makro %_dst durchgeführt:

- Die erforderlichen Makroparameter DS und VAR sind spezifiziert
- Der Datensatz, der für den Parameter DS angegeben ist, existiert und enthält mindestens eine der Variablen, die für Parameter VAR angegeben sind
- Wenn ein Referenzdatensatz angegeben ist
 - muss er existieren und die Variablen start1 und end1 enthalten
 - müssen die Start- und Ende Variablen paarweise auftreten
- Wenn der Parameter REGION nicht angegeben ist, können die Region-Informationen aus den Variablen usubjid oder region0 abgeleitet werden.

6 Fazit

Die Anpassung an die Sommerzeit (ST) kann komplex werden, wenn mehrere Regionen beteiligt sind. Daher ist es sinnvoll, das Problem mit einem Makro anzugehen. Mögliche Datenprobleme löst dies jedoch nicht:

- In einem Land gibt es Regionen mit unterschiedlichen Zeitumstellungszeiten, aber die Region ist nicht bekannt, z.B. verwendet Arizona (US), die SZ nur teilweise.
- Darüber hinaus können Länder die SZ zu einem bestimmten Zeitpunkt abschaffen, verändern oder etablieren.

Die aktuelle Diskussion in der EU deutet darauf hin, dass die SZ in Zukunft abgeschafft werden könnten. Klinische Studien können sich daher teilweise über Jahren mit und teilweise über Jahre ohne SZ erstrecken. Wenn das Datum solche Verschiebungen im Laufe der Zeit enthält, könnte die beste Option die Verwendung einer SZ-Referenz-Datei sein, in der die spezifischen Änderungen im Laufe der Zeit umgesetzt werden.

Ein weiteres potenzielles Problem kann in seltenen Situationen auftreten, wenn Messungen in der Stunde während der Umstellung von SZ auf WZ durchgeführt werden. Während dieser Stunde können keine verlässlichen Zeiten aufgezeichnet werden, da die Uhr für eine Stunde angehalten wird.

Literatur

- [1] Sascha Ahrweiler (2011) Impact of daylight saving time shifts on data quality of a pharmacokinetic study, *Pharmaceutical Programming*, 4:1-2, 31-34, DOI: 10.1179/175709311X13147964247017
- [2] John van Bemmelen. Timezone correction of dates. Basel: PhUSE; 2009, <https://www.lexjansen.com/phuse/2009/cc/CC04.pdf>