Syntax SPSS: Deseasonalize serum 25(OH)D concentration - LASA wave C
May 2017 11th

*****LASA - deaseasonalize 25(OH)D levels*****

****
*deseasonalize 25(OH)D levels: an example of the first LASA-cohort at measurement wave C (1995/96).

*Some important remarks:.
**only deseasonalize 25OHD levels of persons with both a 25OHD level and a date of blood sampling. (We did not want to exclude persons with a 25OHD level but a missing for blood date (n=3). Therefore, we imputed the date for these persons by using the date of the medical interview, which was before blood sampling, and adding a couple of weeks).
**if you have 25OHD data at several measurement waves, deseasonalize for each separate wave.
**(same for different cohorts which differ in age/sex/generation/...) if you have 25OHD data from two cohorts at one wave, it is recommended to deseasonalize for each cohort separately.
**for all steps, select (by using a filter) only those persons that you will include in your analyses; i.e. not all persons with a 25OHD value and date of blood sampling (total sample), but only those who will be included in your study sample (after
exclusions for e.g. missings on the outcome or covariates).

****

*steps:
1. Sine term = sine[day of year/365.25 · 2 · pi]
   Cosine term = cosine[day of year/365.25 · 2 · pi]
2. Linear regression model: measured 25-OHD = beta0 + beta1*sine + beta2*cosine
3. Calculate the deseasonalized 25OHD value: residual + seasonal average = deseasonalized 25-OHD.

***STEP 1.

* Impute missing dates of blood sampling by data of medical interview + some extra weeks (in this case, we checked other respondents with the same medical interview date and checked when their blood sampling was).

```
SUMMARIZE
  /TABLES=respnr Cseizoen_dich   t3m_dat cmbldate
  /FORMAT=VALIDLIST NOCASENUM TOTAL LIMIT=5
  /TITLE='Case Summaries'
  /MISSING=VARIABLE
  /CELLS=COUNT.
*R15623: 18-12-95.
*R21226,    9-2-96.
*R31147   26-2-96.
if (respnr=15623) cmbldate = number("18.01.1996", date11).
if (respnr=21226) cmbldate = number("29.02.1996", date11).
if (respnr=31147) cmbldate = number("15.03.1996", date11).
```

*date of blood sampling (cmbldate = blood date at wave c).

```
FREQUENCIES VARIABLES=cmbldate
  /NTILES=4
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN MODE
  /ORDER=ANALYSIS.
```

*wave C: between jun 1995 and jan 1997, median may 1996. NB: also if the range of the data collection is short (just a part of the year which does not include January 1st), you can use January 1st as reference date (see next command).

* Model the influence of season with a sinus and cosinus function.
* The combination of sinus and cosinus makes the model to calculate 'by itself' the "amplitude of the wave" and to estimate the date of the minimum.
* Reference date baseline 1-1-1996 = day 0. NB: actually it does not matter which date you use: you do not 'calculate the 25OHD values to the first of January' but to 'a modelled yearly average' (based on a specific population sample). By using for example July 1st, you will retrieve the same B0 (intercept) (see step 2).

* 4*Arsin(1) = 2*pi.

```
*C-WAVE:.
COMPUTE datum1 = DATE.DMY(1,1,1996) .
execute.
format datum1 (edate11).
EXECUTE.
COMPUTE dagen1 = DATEDIFF(cmbldate, datum1, "days") .
EXECUTE.
COMPUTE hulp1 = 4*Arsin(1)*(DATEDIFF(cmbldate,datum1, "days")/365.25).
EXECUTE.
COMPUTE cos1 = cos(hulp1).
COMPUTE sin1 = sin(hulp1).
execute.
```

* Check sinus and cosinus function.

```
GRAPH
  /SCATTERPLOT(BIVAR)=dagen1 WITH sin1
  /MISSING=LISTWISE .
GRAPH
  /SCATTERPLOT(BIVAR)=dagen1 WITH cos1
  /MISSING=LISTWISE .
```

*a sinuscurve has at 0,0 an intersection point with the y-as and x-as and goes up from there
*a cosinuscurve has at 0,1 an intersection point with the y-as and goes down from there

*for this example: use all persons with 25OHD values
* in this example, ODIN-standardized 25OHD levels at wave C are used ('cmvitd25st'), but you can also use de original unstandardized values ('cmvitd25').DESCRIPTIVES VARIABLES= cmvitd25st cmbldate sin1 cos1
  /STATISTICS=MEAN STDDEV VARIANCE RANGE MIN MAX SEMEAN.
*n=1320 with vit D levels, 1328 with a blood date (after imputation for n=3) --> 1320 with 25OHD and blood date.

```
***STEP 2. LINEAR REGRESSION.
* original 25(OH)D = beta0  +  beta1*sine  +  beta2*cosine
**** 'original 25OHD' is here the ODIN-standardized 25-OHD level.

* select the n=1320:.
USE ALL.
COMPUTE filter_$=(Cmbldate>1 AND cmvitd25st>1).
VARIABLE LABELS filter_$ 'value(cmbldate) AND cmvitd25st>1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

*lin regr with n=1320 (save the residuals (and predicted values)).
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT cmvitd25st
  /METHOD=ENTER sin1 cos1
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /SAVE PRED RESID .
* mean cmvitD25st is 49.107,   B0=49.373.


***STEP 3.        Calculate the deseasonalized value:   residual  +  seasonal average (B0) =
deseasonalized 25-OHD.

*sum the residuals and the seasonal average.
COMPUTE cmvitd25stDes=RES_1 +  49.3729906051017.
VARIABLE LABELS cmvitd25stDes "deseasonalized standardized serum 25(OH)D at C (nmol/L)".
EXECUTE.

DESCRIPTIVES VARIABLES=cmvitd25st cmvitd25stDes
  /STATISTICS=MEAN STDDEV MIN MAX.
*mean of cmvitd25stDes is 49.3730 (B0).
```