Subject: Reproductive calendar
Posted by kash on Wed, 18 Dec 2013 14:30:58 GMT
View Forum Message <> Reply to Message

I would appreciate some help regarding derivation of the duration of last contraceptive method used in Stata.

Subject: Re: Reproductive calendar
Posted by Liz-DHS on Thu, 20 Feb 2014 19:10:21 GMT
View Forum Message <> Reply to Message

Dear User,
I apologize for not catching your post. We are currently researching an appropriate response for you.
Thank you!

Subject: Re: Reproductive calendar
Posted by user-rhs on Thu, 20 Feb 2014 20:09:37 GMT
View Forum Message <> Reply to Message

Kash,
From a USAID 2009 report: Levels, Trends, and Reasons for Contraceptive Discontinuation (http://pdf.usaid.gov/pdf_docs/PNADQ639.pdf)

Quote:The DHS has created a system for generating events-based datasets from the calendar data,

event
in the calendar--an episode of contraceptive use, a pregnancy, a birth, a termination, or an episode of contraceptive non-use--is converted from the calendar string (the VCAL variables in individual recode or woman-based datasets) into a separate observation for analysis. The start and end date of each event is also recoded, allowing us to calculate directly the duration of the event, women's age, women's parity, and children ever born (using the birth history) at the start or end of each event.

As an alternative, you may want to read documentation on the string or substr function in Stata. Someone may have already answered your question on Statalist
http://www.stata.com/statalist/archive/

RHS

EDIT: I have attached a do-file that does what was mentioned in the report (takes the string and creates a new observation for each character). For this country, the length of the VCAL_1 variable was 65, so I generated 65 new vbls and reshaped from wide to long. From here, you can use -stset- or what have you to set the dataset as survival time data, panel, etc. I used Excel to help me repeat all of those command lines 65 times. I'm sure there's a more elegant way to do this, but this is one way you can get started

```
File Attachments
1) vcal.do, downloaded 1231 times
2) vcal.xlsx, downloaded 1226 times
```

Subject: Re: Reproductive calendar
Posted by Reduced-For(u)m on Fri, 21 Feb 2014 19:54:53 GMT
View Forum Message <> Reply to Message

Wow. So that is a cool thing. But I'm not exactly sure what it is doing (I only had a chance to glance at the documentation). Can I ask- I've been creating woman-year birth history panels using "expand", some re-numbering of within-person obs into years, and then some loops that go through the cross-sectional data and see if anything happened in that observation/year (did you have a baby this 1997, yes/no, in 1998, in 1999?). Does this method make that same panel (for whatever outcome), or is it making something else entirely. I can't really see the structure of the final data from this description.

Subject: Re: Reproductive calendar
Posted by user-rhs on Fri, 21 Feb 2014 23:44:33 GMT
View Forum Message <> Reply to Message

Hi Reduced, since the calendar is stored as a string variable, for it to be useful in survival analysis, it should be transformed so that each event is its own line (at least in my mind, that is how I think about it). Like I said earlier, I'm sure there's a more elegant way of separating the parts than what I did. I use Excel a lot when I'm managing data, especially for repetitious work, because I'm too lazy and impatient to do it manually.

Re: the panel question. In a sense it is a type of panel, in that births are nested within women. You can do survival analysis using the same data with the same long structure too. Of course creating this sort of "panel" is is only appropriate in some settings, i.e. when recall is reasonably accurate. I assume that recall of birth hx is generally accurate, especially if it happened in the past 5 years.

Subject: Re: Reproductive calendar
Posted by Reduced-For(u)m on Sat, 22 Feb 2014 20:39:51 GMT
View Forum Message <> Reply to Message

Thanks. I feel like sometimes when I read certain papers I don't know what the unit of observation is: woman, event, woman-year, potential-event-year... I think depending on your estimator and your strategy, any of those might be appropriate, but I was curious what the calendar variable creates. Got it.

"I use Excel a lot when I'm managing data, especially for repetitious work, because I'm too lazy and impatient to do it manually." - Me too! Auto-fill has saved me hours and hours of stupid code writing (followed by stupid typo de-bugging). Sometimes a loop is more elegant, but sometimes that trade-off isn't worth the time and energy.

Subject: Re: Reproductive calendar
Posted by user-rhs on Mon, 24 Feb 2014 19:42:24 GMT
View Forum Message <> Reply to Message

Sometimes one gets the sense that the authors of those papers don't know either what the unit of observation is :).

I've only seen this vbl in action in one paper (Steele, Curtis, and Choe 1999; link: http://onlinelibrary.wiley.com/doi/10.1111/j.1728-4465.1999. 00028.x/abstract). I'm sure there are other papers out there, but I'm not an FP person, so I don't go out looking for these hairy things... :)

Subject: Re: Reproductive calendar
Posted by Liz-DHS on Wed, 05 Mar 2014 16:33:04 GMT
View Forum Message <> Reply to Message

Dear User,
Do you still need help with this post? If so, can you please provide a little more detail on what you are trying to accomplish and which data set you are working with?
Thank you!

Subject: Re: Reproductive calendar
Posted by user-rhs on Wed, 19 Mar 2014 00:20:58 GMT
View Forum Message <> Reply to Message

Hi Liz,
I have got one for you about VCAL (feeding my curiosity here). In my previous post, I trimmed the VCAL_1 variable down to 65 characters. Upon reading the documentation for VCAL more

closely, I see this is not the right thing to do. The recode manual says:
Quote:The first character in each variable represents the most recent point in time, while the 80th character
position represents data for January of the year in which the calendar started. The calendars
are fixed at the 80th character position, such that the first few entries in the calendar
represent points in time after the date of interview, and are consequently left blank.
I am wondering if you can help me understand what I'm seeing.

FYI, I'm looking at the Ethio 2011 data. When I separate the variable VCAL_1 into 80 separate variables (I call them 'en' where n=location of character in string, so that the first character, AKA the most recent time point, is 'e1' and the point where the calendar starts is 'e80'). I do this in preparation of converting the data from wide to long. Just looking at this step, I see that e1-e11 are blank for all observations. Variables e12-e80 all have non-blank values. Given that the recall period is typically 5 years (60 months), do those first 11 columns mean anything? I noticed that for column 12, 99.81% had " " as the value, .15% had "0", and .04% had "P", which makes me think this column was only populated because the women were interviewed at the tail-end of data collection or something like that. Is this a correct conclusion?

Is my "most recent timepoint" (month of interview) in fact column 12/13?

Thanks,
RHS

Subject: Re: Reproductive calendar
Posted by Liz-DHS on Wed, 19 Mar 2014 02:25:37 GMT

Dear RHS,
I will send your question over to one of our experts. The one thing I know about Ethiopia is that their calendar is different. That is they use a different system than the Gregorian calendar. Will get back to you as soon as I have an answer.
Thanks!

Subject: Re: Reproductive calendar
Posted by Liz-DHS on Wed, 19 Mar 2014 18:18:25 GMT

Dear RHS,
Here is a response from one of our experts, Guillermo Rojas:
Regarding the calendar in general (not just Ethiopia), the calendar should be used in conjunction with the date of interview.   Row 80 in the calendar correspond to the beginning of the calendar and that is fixed for all interviews. It the case of Ethiopia it correspond to January, 1998 (Variable V017 = 1177 (1998-1900)*12+1). It is important to know that Ethiopia uses a different calendar, 2011 in the western calendar corresponds to 2003 in Ethiopia.

The first row with information in the calendar correspond to the month of interview for that woman. Variable V018 tells where the calendar begins for a particular case. In the case of Ethiopia, row 12 correspond to woman interviews collected in September, 2003. (V006=9, V007=2003, V008=1245). Row 11 to interviews made in August, 2003 (V006=8, V007=2003, V008=1244) and so on. By looking at the frequencies only 32 women were interviewed in September , 2003 and thus the number of blanks in that row.

Thank you for your post. If this does not answer your question, please feel free to post again.

Subject: Re: Reproductive calendar
Posted by user-rhs on Wed, 19 Mar 2014 20:33:18 GMT
View Forum Message <> Reply to Message

Thank you, Liz and Guillermo.

Subject: Re: Reproductive calendar
Posted by Liz-DHS on Thu, 20 Mar 2014 13:38:34 GMT
View Forum Message <> Reply to Message

You are most welcome!

Subject: Re: Reproductive calendar
Posted by kash on Sat, 24 May 2014 20:34:12 GMT
View Forum Message <> Reply to Message

Dear Liz,
Thank you very much for your help. I'm trying to determine duration of contraceptive use from the 2010 Malawi DHS contraceptive calendar.
Best,

Subject: Re: Reproductive calendar
Posted by kash on Sat, 24 May 2014 21:06:09 GMT
View Forum Message <> Reply to Message

This was very helpful. I really appreciate your help. I am using the 2010 Malawi DHS. The command "di length(vcal_1)" suggests the length is 80, but then "tab v019" shows the length ranging from 66 to 70. Would you know what creates this discrepancy?

Subject: Re: Reproductive calendar
Posted by user-rhs on Mon, 26 May 2014 03:41:09 GMT
View Forum Message <> Reply to Message

Kash,
Since I originally posted the Stata code to extract each character into its own variable, I have revised my methods and would not recommend trimming the variables into a certain length x. Go ahead and extract each character in the variable, which means adding X number of repetitions of the lines that read gen ___ =substr(....) (I think the standard length was 80? I'm not at my desk right now and don't have the data in front of me). The different lengths in number of characters is normal and to be expected. It is a function of when the woman was interviewed. Remember that the start date for the calendar is the same for all women, so if a woman was interviewed toward the end of the data collection cycle, she will have more months of information than a woman who was interviewed at the beginning of the data collection cycle.

HTH,
RHS

Subject: Re: Reproductive calendar
Posted by user-rhs on Mon, 26 May 2014 03:48:18 GMT
View Forum Message <> Reply to Message

Sorry for the double post...

Subject: Re: Reproductive calendar
Posted by kash on Mon, 26 May 2014 15:41:12 GMT
View Forum Message <> Reply to Message

RHS,
Following our discussions, this is how I derive, in STATA, the duration to first pregnancy since the start of the reproductive calendar. I hope other users could find it useful or improve on it.
 -- ----------------------------

```
di length(vcal_1) /*length is 80*/
format %80s vcal_1
local i=1
forvalues n= 80(-1)1{
        gen str1 mcont`i' = substr(vcal_1,`n',1)
    local ++i
    }
reshape long mcont, i(caseid) j(event_time)
by caseid, sort: egen fecund = min(cond(mcont == "P", event_time,.))
```
 -- --------------------

Best regards,
Kash

Subject: Re: Reproductive calendar
Posted by Reduced-For(u)m on Mon, 26 May 2014 19:03:53 GMT
View Forum Message <> Reply to Message

That's lovely. I might even try using that reproductive calendar some time now. Thanks!

Subject: Re: Reproductive calendar
Posted by user-rhs on Tue, 27 May 2014 00:45:36 GMT
View Forum Message <> Reply to Message

Thanks, Kash. I second Reduced For(u)m's opinion above! One thing to add: I think the calendar starts at character #80 and ends at character #1 (you may have done this reverse sorting already...I admit I didn't read your code too carefully). Don't quote me on that, but I think that's how it works (the Recode Manual should clear this up).

RHS

Subject: Re: Reproductive calendar
Posted by kash on Tue, 27 May 2014 18:09:28 GMT
View Forum Message <> Reply to Message

Yes, it starts at #80 and that is why I made the loop count backwards from 80 to 1.

Subject: Re: Reproductive calendar
Posted by Kisaakye on Mon, 17 Aug 2015 10:32:44 GMT
View Forum Message <> Reply to Message

Hello..
I am just beginning to work with calendar data (for now am using the 2011 Uganda DHS). I have extracted the data from the standard recode file. In the meantime i am trying to study the data but as i browse through, i realize the first 11 columns have no record/event. I am struggling to give an explanation to this. The second question relates to the events column. What does "N" stand for? I seem to understand all other letters and figures.

Thanks in advance
Peter

Subject: Re: Reproductive calendar
Posted by Trevor-DHS on Mon, 17 Aug 2015 14:59:51 GMT
View Forum Message <> Reply to Message

The calendar (vcal_1) is set up as a string of 80 characters, with the 80th character representing the start point of the calendar, which, in the case of Uganda DHS 2011, is January 2006. V017 tells you the century month code for the start of the calendar (1273=January 2006). Position 1 in the calendar in this case would then be 79 months later, which would be August 2012, however, this is beyond the date of interview. Any months beyond the date of interview are left blank. If the interview, for example, took place in November 2011, then the information in the calendar would start in position 10, and positions 1-9 (corresponding to August 2012 back to December 2011) would all be blank. V018 tells you the position of the month of interview, and for an interview in November 2011 would be position 11.

You say that the first 11 positions have no record/event, but that is only the case for an interview taking place in September 2011. The Interviews took place between June 2011 and December 2011, so the number of blanks on the beginning of the record will differ, depending on the month of interview.

IMPORTANT NOTE: For some surveys, the calendar data are misaligned in the Stata data files. In those surveys affected, the calendar data are all trimmed so that they are left aligned. This only affects the Stata versions of the datasets, and does not affect all surveys (only a subset). I have written code to correct the misalignment, which I am attaching here. You will find a foreach statement in the code that lists the datasets affected (that I know of to date). There is a quick check to see if your dataset is affected:
gen v = substr(vcal_1,1,1)
tab v
If the calendar is correctly aligned, v will be mostly blank - in fact in the vast majority of surveys it will be all blank.
If v is not blank for most/all cases, then the calendar is misaligned. To correct the alignment, you can use the code attached.

I'm also attaching code I use for reshaping the calendar, which you can adapt for your own needs. Make sure you run the realign code first, though, before the reshape code, if the survey needs it [Uganda DHS 2011 is fine and does not need realigning - see the foreach for the datasets that I know do need realigning].

File Attachments
1) Realign vars.do, downloaded 949 times
2) Reshape vcal_1 to months.do, downloaded 1146 times

Subject: Re: Reproductive calendar
Posted by Kisaakye on Tue, 18 Aug 2015 09:57:35 GMT
View Forum Message <> Reply to Message

Thanks very much Trevor for this wonderful explanation. In the meantime, as i try to understand this, could you know what events "N" and "W" stand for?
Thanks.

Subject: Re: Reproductive calendar
Posted by user-rhs on Tue, 18 Aug 2015 14:11:28 GMT
View Forum Message <> Reply to Message

Peter, the country-specific documentation should have information on the coding of the calendar.

Subject: Re: Reproductive calendar
Posted by Kisaakye on Tue, 18 Aug 2015 14:17:53 GMT
View Forum Message <> Reply to Message

Yes it has but these two specific letters (N and W) are not defined and yet they appear somewhere?! I would not know what they stand for or the specific method they represent.

Subject: Re: Reproductive calendar
Posted by user-rhs on Tue, 18 Aug 2015 14:33:51 GMT
View Forum Message <> Reply to Message

By "country-specific documentation," I mean the Word document ("Individual Recode Documentation") that comes with the zipped dataset. I don't have access to the Uganda dataset, so I can't tell you what your codes mean, but in the Indonesia 2012 dataset, for VCAL1, "W" stands for "Other" and "N" stands for "Implants/Norplant," for VCAL2 (discontinuation), "W" stands for "Other" and "N" stands for "IUD expelled," for VCAL5 (source), "N" stands for "Friends/relatives," and for VCAL6, "W" stands for "Other" and "N" stands for "Implants/Norplant."

Subject: Re: Reproductive calendar
Posted by Kisaakye on Tue, 18 Aug 2015 15:11:46 GMT
View Forum Message <> Reply to Message

Thanks for this.

Subject: Re: Reproductive calendar
Posted by Liz-DHS on Tue, 18 Aug 2015 15:19:12 GMT
View Forum Message <> Reply to Message

Dear User,
In the Recode6 dictionary, the Col1 W is other N is implants/Norplant for Col2 discontinuation W is other and there is no N code. In this survey, there was only a 2 column calendar.
Thank you!

Subject: Re: Reproductive calendar

Posted by Kisaakye on Tue, 18 Aug 2015 15:24:51 GMT
View Forum Message <> Reply to Message

Thanks for the response but there is a big mix-up. According to the coding frame available to me (or the one used in the survey - derived from the questionnaire), 5 was for Implants in Col1 and X for other modern methods in Col1. However, for Col2 X was maintained for other. Still stuck on how to go about it

---

Subject: Re: Reproductive calendar
Posted by Liz-DHS on Tue, 18 Aug 2015 15:36:40 GMT
View Forum Message <> Reply to Message

We are forwarding your query to one of our experts. Will post as soon as possible. Thank you!

Subject: Re: Reproductive calendar
Posted by Trevor-DHS on Tue, 18 Aug 2015 17:35:13 GMT
View Forum Message <> Reply to Message

The codes found in the calendar in the recode file are not the same as the codes used at the time of data collection. The codes in the calendar are recoded to a common standard coding scheme. The DHS6 recode manual provides details of the coding used in the recode file (pages 98 and 99).

Subject: Re: Reproductive calendar
Posted by Kisaakye on Wed, 26 Aug 2015 12:17:30 GMT
View Forum Message <> Reply to Message

Trevor, the information(codes for specific methods) in the manual are very helpful in understanding what the letters stand for in the data-sets. However, "M" is not listed among or explained in the recode manual (page 98) and yet it appears in the data set (UDHS 2011). Could you help with this.
Thanks

Subject: Re: Reproductive calendar
Posted by Trevor-DHS on Thu, 27 Aug 2015 13:44:17 GMT
View Forum Message <> Reply to Message

Code "M" is a recent addition as a standard code and means "Other modern methods" (it is not specified which these are, but they are methods that do not already have other codes).

Subject: Re: Reproductive calendar
Posted by Kisaakye on Thu, 27 Aug 2015 13:48:48 GMT
View Forum Message <> Reply to Message

Thank you so much Trevor for this clarification.