

Tips and Tricks: Using Dates in STATA

Jen Cocohoba, Pharm.D., MAS
Health Sciences Associate Clinical Professor
UCSF School of Pharmacy

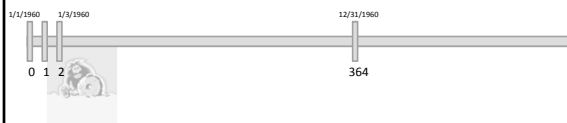
NOTE: tips tested for STATA v.12, though most info will be relevant for STATA 13

STATA dates

- Dates common data points in research
- STATA reads dates as string
- Do the “usual”
 - Open Excel spreadsheet, copy, paste into editor
 - (OR import the data)
 - Note color of variable

** DEMONSTRATION: DATES

How STATA thinks about dates



- “Counts” date as the # of days from a specific reference
 - January 1, 1960 = 0
 - January 2, 1960 = 1
 - January 3, 1960 = 2
 - December 31, 1960 = 364
- This makes it “easy” for STATA to manipulate mathematically
- We will come back to this when formatting dates

Cleaning STATA dates

- Need to convert to STATA-recognizable date to perform analysis
- 1) Generate a new date variable using `date` function
 - 2) Identify the “old” string variable which contains the date
 - 3) Tell STATA what format it was in (e.g. month, day, year)
 - 4) Compare old and new results

generate *dob* = `date`(*birthdate*, “MDY”, *topyear*)

Annotations:
 - New variable name: *dob*
 - Date function: `date`
 - Old variable name: *birthdate*
 - “top year” that should be interpreted: *topyear*
 - How the date is arranged: “MDY”

*NOTE: your original date variable can be “date-like” (e.g. 8/10/1970) or can be in a true string format (August 10, 1970) --- STATA can figure it out.

Number nonsense

dob
-2372
-4366
-3839
150
-4862
-3626
-2788
-3562
-1868
-5946
-5984
-1962
-4694
-6018
-4407
0

- Emerges as the date in STATA speak
- Can mask the numerical date so that it is easier for you to understand

Command:
`format dob %td`

* NOTE: Other formats aside from %td – in STATA help

Dates: series of commands

- Date conversions usually 2-commands plus checking


```
generate dob = date(birthdate, "MDY")
format dob %td
browse dob birthdate
drop birthdate
```
- NOTE: STATA issues with 2-digit years (8/10/76)
 - Will get "missing values" generated
 - Two ways to fix this
 - Format dates to 4 digit years in Excel, then copy to STATA
 - Add "topyear" cutoff to the STATA command.
 - Anything beyond topyear = previous century

```
generate dob = date(birthdate, "MDY", 2012)
```

```
9/10/11 = September 10, 2011
9/10/12 = September 10, 2012 ← top year
9/10/13 = September 10, 1913
9/10/14 = September 10, 1914
```

MDY command

- Date components housed in separate variables

birth_m	birth_d	birth_y
7	4	1953
1	18	1948
6	28	1949
5	30	1960

- STATA can concatenate these for you

Generate *datevar* = *mdy*(*birth_m*, *birth_d*, *birth_y*)

Diagram labels: "New variable name" points to *datevar*; "mdy date function" points to *mdy*; "Name of month, day, and year variables" points to the three arguments in parentheses.

Date is now formatted – what can you do with it?

- Extract components of the date into new variables (columns)
 - `gen nameofdayvariable = day(datevariable)`
 - `gen weekdayvariable = dow(datevariable)`
 - Lists as 0(Sunday) - 6(Saturday)
 - `gen monthvariable = month(datevariable)`
 - `gen yearvariable = year(datevariable)`

What else can you do with dates

- Find time elapsed between dates
- Suppose you wanted to find participants' age at the date of their study visit (or today)
 - Generate new variable called ageatvisit


```
gen ageatvisit = vdate - dob
```
 - Note this gives you their age in number of DAYS
 - Can do this more efficiently by


```
gen ageatvisit = (vdate - dob)/365.25
gen agevisityears = int(ageatvisit)
```

Comparing dates

- Suppose you wanted to categorize patients by a date
 - Patients starting ARV < 1996 = pre-HAART
- Using literal dates
 - Formatted as day month year (01jan1960)
 - Must be denoted by parenthesis
 - Must use pseudocommand `td`
 - Example: `td(01jan1960)`
- Example
 - `gen prehaart = 0`
 - `replace prehaart = 1 if artstart <= td(01jan2006)`
 - `replace prehaart = . if artstart == .`

Demonstration – dates

characteristics