

## **Analytics - Time Series I**

### **Preliminaries:**

### **Dates**

- Internal: # days since Jan 1, 1960
- Need format for reading, one for writing
- Often DATE is ID variable (extrapolates)
- Program has lots of examples:

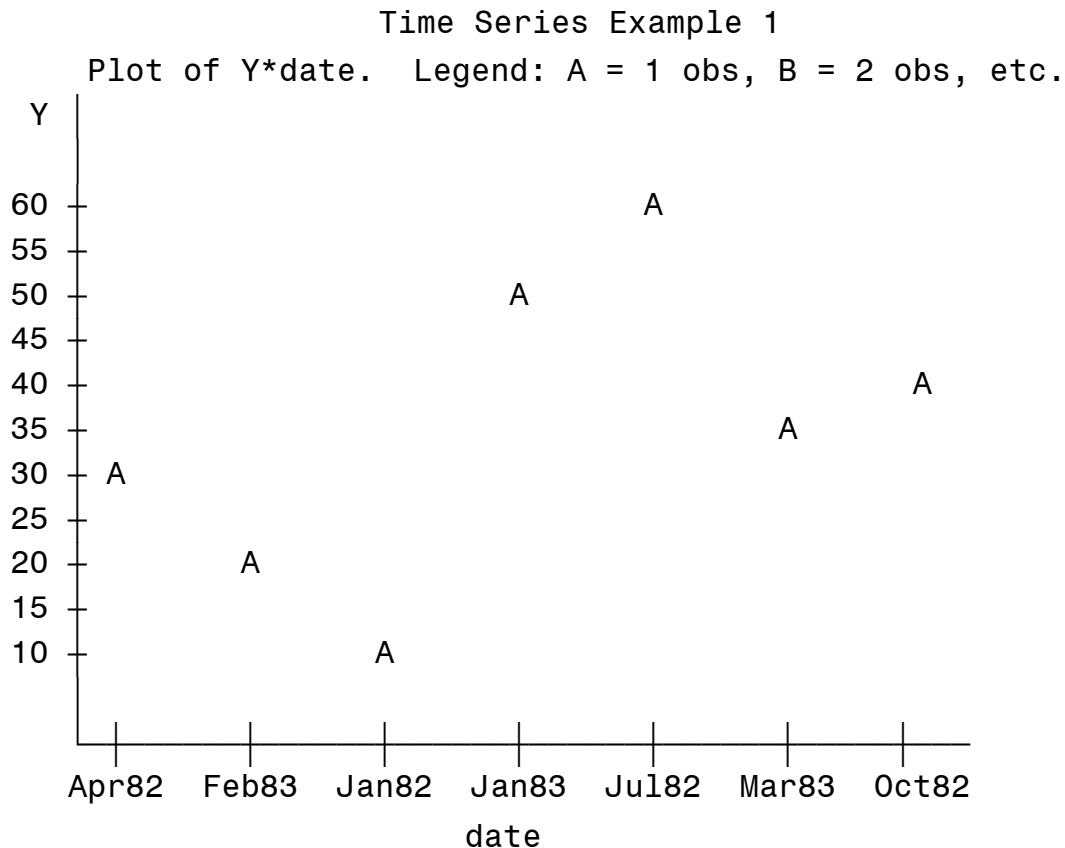
```
options ls=76 nodate; title "Time Series
Example 1";
data A; input date $ Y @@; cards;
Jan82 10      Apr82 30      Jul82 60      Oct82 40
Jan83 50      Feb83 20      Mar83 35
;
proc plot; plot Y*date/vpos=20 hpos=50;run;

data B;
Diff = '01jan80'D-'13nov1979'd;
dec1919='31dec1919'd;
dec19='31dec19'D; jan20='01jan20'D;
date1 = 18; date2=18; date3=18;
format date1 monyy.; format date2 date9.;
format date3 mmdyy.;
proc print; run;
```

```

data c; array date(3);
input x date7. ;
do i=1 to 3; twoi=2*i; date(i) =
intnx('month', '01jan1912'd, twoi); end;
cards;
01feb60
;
proc print;
proc print; format date1-date3 mmddyy.;
data next; set a;
  newdate= input(date, monyy.);
newdate2 = newdate; format newdate2 monyy.;
proc sort; by date; proc print;
proc sort; by newdate2; proc print; run;

```

**Program Output:**

Time Series Example 1

Obs	Diff	dec1919	dec19	jan20	date1	date2	date3
1	49	-14611	21914	-14610	JAN60	19JAN1960	01/19/60

Obs	date1	date2	date3	x	i	twoi
1	-17472	-17411	-17350	31	4	6

Obs	date1	date2	date3	x	i	twoi
1	03/01/12	05/01/12	07/01/12	31	4	6

Obs	date	Y	newdate	newdate2
1	Apr82	30	8126	APR82
2	Feb83	20	8432	FEB83
3	Jan82	10	8036	JAN82
4	Jan83	50	8401	JAN83
5	Ju182	60	8217	JUL82
6	Mar83	35	8460	MAR83
7	Oct82	40	8309	OCT82

Obs	date	Y	newdate	newdate2
1	Jan82	10	8036	JAN82
2	Apr82	30	8126	APR82
3	Ju182	60	8217	JUL82
4	Oct82	40	8309	OCT82
5	Jan83	50	8401	JAN83
6	Feb83	20	8432	FEB83
7	Mar83	35	8460	MAR83

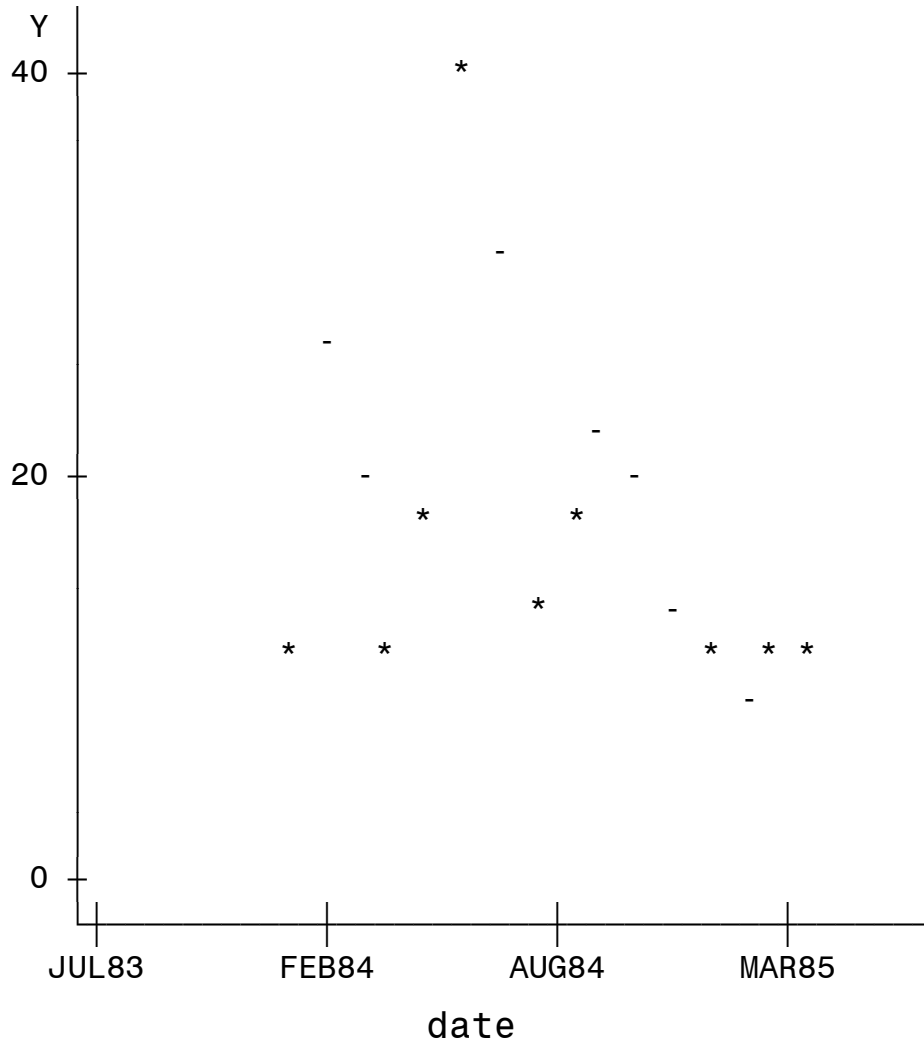
**PROC EXPAND (cubic spline)**

```
data last; input Y @@;
date=intnx('month','01dec83'd,_n_);
format date monyy.;
cards;
10 . . 12 18 40 . 13 18 . . . 10 . 10 10
;
proc print;
proc expand data=last from=month to=month
out=out1;
  convert Y = ynew; id date;
data out1; merge out1 last; by date; proc
plot data=out1;
  plot Y*date="*" Ynew*date = "-" /overlay
vpos=20 hpos=50;
proc expand data=last from=month to=week
out=out2 outest=spline;
  convert Y = Ywk; id date;
proc print data=out2(obs=5);
proc print data=spline(obs=5);
data out2; merge out2 last; by date; proc
plot data=out2;
  plot Y*date="*" Ywk*date = "-" /overlay
vpos=20 hpos=50;
run;
```

## Time Series Example 1

Obs	Y	date
1	10	JAN84
2	.	FEB84
3	.	MAR84
4	12	APR84
5	18	MAY84
6	40	JUN84
7	.	JUL84
8	13	AUG84
9	18	SEP84
10	.	OCT84
11	.	NOV84
12	.	DEC84
13	10	JAN85
14	.	FEB85
15	10	MAR85
16	10	APR85

Plot of Y\*date. Symbol used is '\*'.  
 Plot of ynew\*date. Symbol used is '-'.



NOTE: 7 obs had missing values. 9 obs hidden.

Obs	date	Ywk
1	Sun, 1 Jan 1984	10.0000
2	Sun, 8 Jan 1984	16.5305
3	Sun, 15 Jan 1984	21.1830
4	Sun, 22 Jan 1984	24.1798
5	Sun, 29 Jan 1984	25.7433



## Data Sets

- Expect column for id (date)
- Dependent (response, target) variable and explanatory (input) variable columns
- May need to transpose, combine.

```

data a; input y1-y5; cards;
12 11 16 19 999
data b; retain date;
if _n_=1 then do;
input month day year @@;
date=mdy(month,day,year); end;
input Y Z @@; if _n_>1 then date=date+1; format
date mmdyy.;
cards;
10 28 1987    16 1    19 2    15 3    18 4    21 4
25 3    28 2    26 1
;
proc print data=a; proc print data=b;
proc transpose data=a out=aa; var y1-y5;
data aa; set aa (rename=(col1=Y)); *drop _name_;
date=date+1; retain date '29oct1987'd;
proc print data=aa;
data both; merge b aa; by date; proc print; run;

```

Obs	y1	y2	y3	y4	y5
1	12	11	16	19	999

Obs	date	month	day	year	Y	Z
1	10/28/87	10	28	1987	16	1
2	10/29/87	.	.	.	19	2
3	10/30/87	.	.	.	15	3
4	10/31/87	.	.	.	18	4
5	11/01/87	.	.	.	21	4
6	11/02/87	.	.	.	25	3
7	11/03/87	.	.	.	28	2
8	11/04/87	.	.	.	26	1

Obs	_NAME_	Y	date
1	y1	12	10164
2	y2	11	10165
3	y3	16	10166
4	y4	19	10167
5	y5	999	10168

Obs	date	month	day	year	Y	Z	_NAME_
1	10/28/87	10	28	1987	16	1	
2	10/29/87	.	.	.	19	2	
3	10/30/87	.	.	.	12	3	y1
4	10/31/87	.	.	.	11	4	y2
5	11/01/87	.	.	.	16	4	y3
6	11/02/87	.	.	.	19	3	y4
7	11/03/87	.	.	.	999	2	y5
8	11/04/87	.	.	.	26	1	

- Merge: Format is first one encountered (aa has none)
- Merge: Value is last one encountered (like overrecording a tape)

## Transformations

- Most common: no transformation or log  
Try log if variation increases as mean increases.
- Log difference often used in economics (Log is natural log, ln)

$$\text{Log}(Y_t) - \text{Log}(Y_{t-1}) = \text{Log}(Y_t/Y_{t-1})$$

Taylor's series:  $\text{Log}(1+\epsilon) = \log(1) + 1\epsilon - \epsilon^2/2 + \dots \approx \epsilon$   
for  $\epsilon$  small.

Thus  $100 \text{Log}(Y_t/Y_{t-1})$  is approximate percentage change if small.

```
data a; array X(11);
do i=1 to 11; X(i) = .88+i/50; end; output;
do i=1 to 11; x(i) =log(x(i)); end; output;
proc print noobs; var X1-X11; format X1-X11 5.2;
run;
  X1    X2    X3    X4    X5    X6    X7    X8    X9    X10   X11
  0.90  0.92  0.94  0.96  0.98  1.00  1.02  1.04  1.06  1.08  1.10
-0.11 -0.08 -0.06 -0.04 -0.02  0.00  0.02  0.04  0.06  0.08  0.10
```

•  $\log(Y)=X \sim N(\mu, \sigma^2) \Rightarrow E\{Y\} = E\{e^X\} =$

$$\frac{1}{\sqrt{2\pi}\sigma} \int e^X e^{-\frac{(X-\mu)^2}{2\sigma^2}} dx =$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int e^{-\frac{X^2-2(\mu+\sigma^2)X+\mu^2}{2\sigma^2}} dx =$$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{\mu+\sigma^2/2} \int e^{-\frac{X^2-2(\mu+\sigma^2)X+(\mu+\sigma^2)^2}{2\sigma^2}} dx =$$

$$e^{\mu+\sigma^2/2} \text{ (not just } e^\mu\text{)}$$

• Although exponentiating mean of  $\log(Y)$  does not give mean of  $Y$ , it is true that  $\Pr\{\log(Y) < C\} = \Pr\{Y < e^C\}$  from which we see that exponentiating median of  $\log(Y)$  gives the median of  $\log(Y)$ , that is,  $e^\mu$  is median of  $Y$  (not mean).