Dates

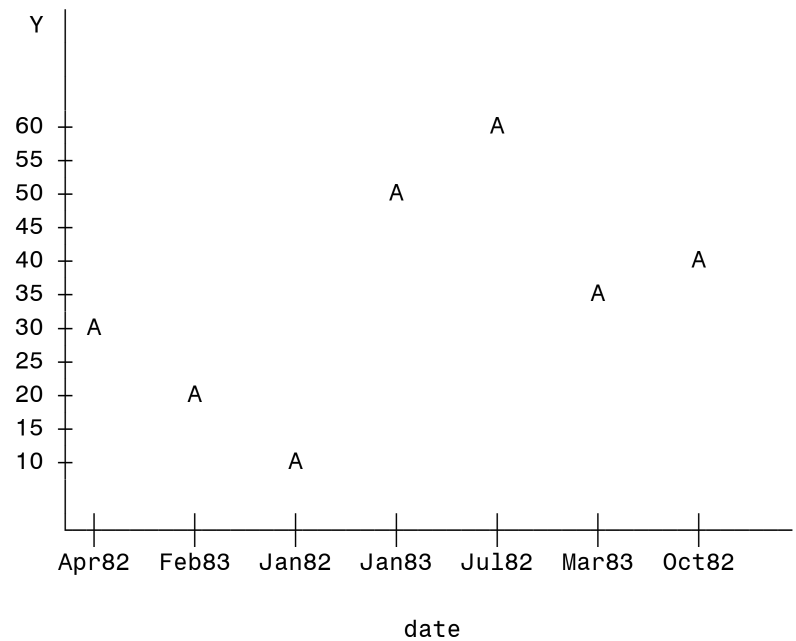
- Internal: # days since Jan 1, 1960
- Need format for reading, one for writing
- Often DATE is ID variable (extrapolates)
- Program has lots of examples:

```
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
mmddyy.;
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;
```

Program Output:

Time Series Example 1

Plot of Y*date. Legend: A = 1 obs, B = 2 obs, etc.



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	Jul82	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	Jul82	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)

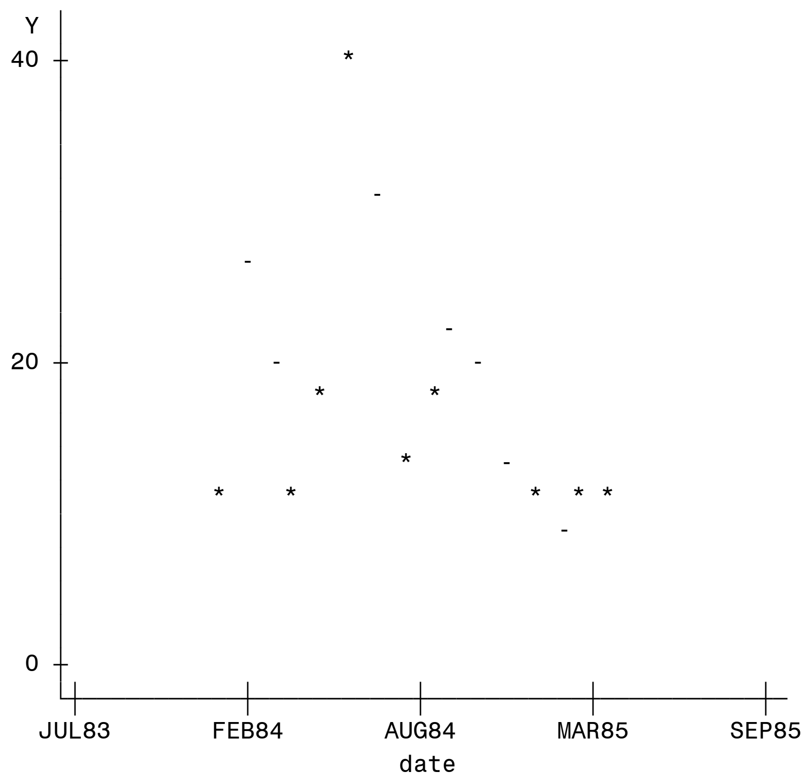
```
data last; input Y @@; date=intnx('month','01dec83'd,_n_);
format date monyy.;
cards;
10 . . 12 18 40 . 13 18 . . . 10 . 10 10
  ;
proc print;
proc expand data=last from=month to=month out=out1;
  convert Y = ynew;  id date;
data out1; merge out1 last; by date; proc plot data=out1;
  plot Y*date="*" 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 '-'.

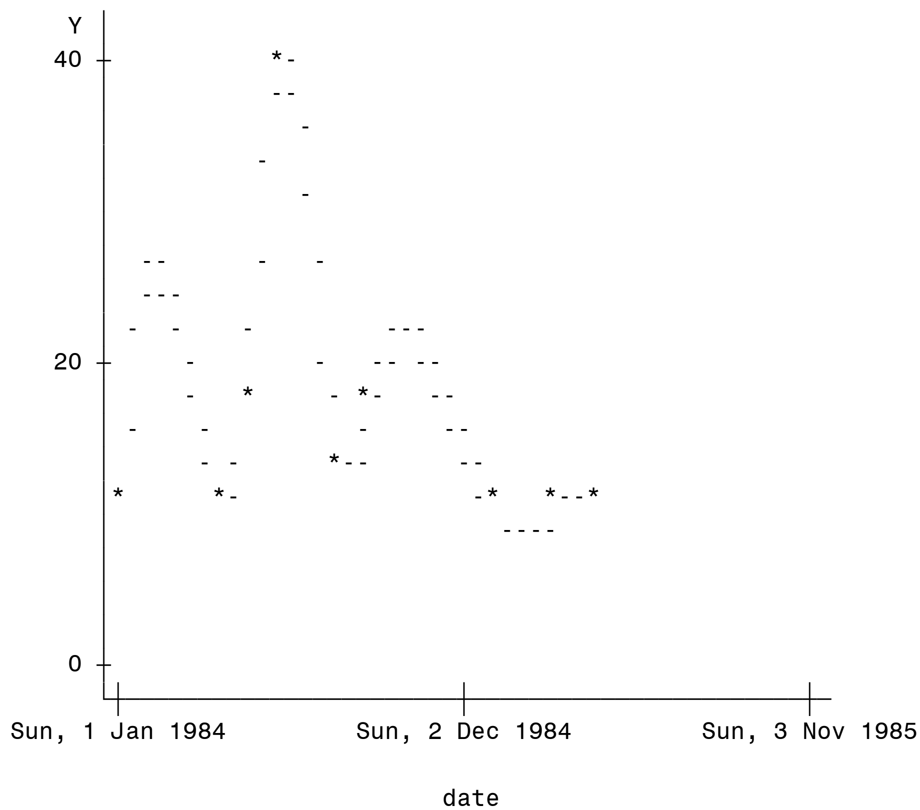


NOTE: 7 obs had missing values. 9 obs hidden.

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

Obs	VARNAME	METHOD	OBSERVED	date	CONSTANT	LINEAR	QUAD	CUBIC
1	Y	SPLINE	B	DEC1983	-23.4075	1.07766	0.000000	0
2	Y	SPLINE	B	JAN1984	10.0000	1.07766	-0.021432	0.000108034
3	Y	SPLINE	B	MAY1984	18.0000	0.63630	0.017784	-.000497339
4	Y	SPLINE	B	JUN1984	40.0000	0.30511	-0.028468	0.000265741
5	Y	SPLINE	B	AUG1984	13.0000	-0.20153	0.020162	-.000272852

```
                    Time Series Example 1

          Plot of Y*date.    Symbol used is '*'.
          Plot of Ywk*date.  Symbol used is '-'.
```




```
NOTE: 83 obs had missing values.  17 obs hidden.
```

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 mmddyy.;
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)

Prediction

- BLUP Best Linear Unbiased Predictor
  E = expectation = average in population. $E\{X\}=\mu$, $E\{ (X-\mu)^2 \}= \sigma^2$
  - Predictor (of Y_{n+L})
  - Linear ($\hat{Y}_{n+L} = b_n Y_n + b_{n-1} Y_{n-1} + \cdots + b_1 Y_1$, assuming means 0)
  - Unbiased ($E(Y_{n+L}-\hat{Y}_{n+L}) = 0$)
  - Best (pick b_j to minimize something, e.g. $E(Y_{n+L}-\hat{Y}_{n+L})^2$)

$$\begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \end{array} \begin{array}{c} \begin{array}{ccc} Y_1 & Y_2 & Y_3 \end{array} \\ \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} \end{array} \qquad E\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

$\hat{Y}_3=\mu_3 + b_1(Y_1-\mu_1) + b_2(Y_2-\mu_2)$ unbiased

$Var\{ Y_3-\mu_3 - b_1(Y_1-\mu_1) - b_2(Y_2-\mu_2)) \}$ = (from St 512 !!)

$$(-b_1 \quad -b_2 \quad 1) \begin{pmatrix} v_{11} & v_{12} & v_{12} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} \begin{pmatrix} -b_1 \\ -b_2 \\ 1 \end{pmatrix}$$

- Minimize! (set $\frac{\partial}{\partial b_j}(*) = 0$ for j=1,2)

Solution: $\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}^{-1} \begin{pmatrix} v_{13} \\ v_{23} \end{pmatrix}$

- Example: $\mu_3 = \mu_2 = \mu_1 = 100$, $Y_1=120$, $Y_2=180$

$$\begin{array}{c|ccc} & Y_1 & Y_2 & Y_3 \\ \hline Y_1 & 8 & 4 & 2 \\ Y_2 & 4 & 8 & 4 \\ Y_3 & 2 & 4 & 8 \end{array} \qquad \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \frac{1}{64-16}\begin{pmatrix} 8 & -4 \\ -4 & 8 \end{pmatrix}\begin{pmatrix} 2 \\ 4 \end{pmatrix} = 48^{-1}\begin{pmatrix} 0 \\ 24 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

MARKOV $\hat{Y}_3 = 100 + (1/2)(180-100) = 140$

(A) Toeplitz covariance matrix: $\text{Cov}(Y_i,Y_j) = \gamma(\,|i-j|\,)$ = fn of $|i-j|$ only.

$$\begin{pmatrix} A & B & C \\ B & A & B \\ C & B & A \end{pmatrix} \qquad \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{pmatrix}$$

(B) Means all the same

(A)&(B) $\overset{\text{defn.}}{<==>}$ *"COVARIANCE STATIONARY"* or just *"STATIONARY"*

True or false for Toeplitz covariance matrix:

- Covariance matrix of $Y_1, Y_2, \cdots, Y_n$ same as that of $Y_n, Y_{n-1}, \cdots, Y_1$
- BLUP of Y_1 based on $Y_2, \cdots, Y_n$ uses same weights (b_j) as
  BLUP of Y_n based on $Y_{n-1}, \cdots, Y_1$

- Stationary: Toeplitz matrix still has n entries, with mean that's n+1 parameters to estimate - still too many.

- Idea: Express $\gamma(h)$ as function of just a few unknowns.

Example 1: $\gamma(0)$, $\gamma(1)$, $\gamma(2)$, and $\gamma(h)=0$ if h>2. "MA(2)"
Example 2: $\gamma(0)$, and $\gamma(h)=\rho^h\gamma(0)$ for h>0. "AR(1)"
Example 3: $\gamma(0)$, $\gamma(1)$, and $\gamma(h)=\rho^h\gamma(1)$ for h>1. "ARMA(1,1)"
Example 4: $\gamma(0)$, and $\gamma(h)=0$ if h>0. "White Noise"

Check: $\gamma(0) = 100$, $\gamma(1)=80$, $\gamma(2)=72$, $\gamma(2)=64.8$, $\gamma(3)=58.32$ etc.
What type is this???

- Matrices must be "positive semi-definite" (conditions that prevent negative variances)

Example

$$\begin{pmatrix} 10 & 8 & 0 \\ 8 & 10 & 8 \\ 0 & 8 & 10 \end{pmatrix} \text{ looks like MA(1) but variance of } W = (1, -1, 1)\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

would then be

$$\begin{pmatrix} 1 & -1 & 1 \end{pmatrix}\begin{pmatrix} 10 & 8 & 0 \\ 8 & 10 & 8 \\ 0 & 8 & 10 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = 30 - 32 = -2 < 0$$

Can't have negative variance!!

- We will see in the course that

(1) All AR, MA, and ARMA models can be expressed in terms of white noise as
$Y_t-\mu = \alpha_1(Y_{t-1}-\mu) + \alpha_2(Y_{t-2}-\mu) +\cdots +\alpha_p(Y_{t-p}-\mu) + e_t-\theta_1 e_{t-1}-\cdots-\theta_q e_{t-q}$

(2) Stationarity for these models is ensured if roots of a certain "characteristic polynomial" are in the right region (no unit roots!)

(3) The associated covariances γ(h) are related to the α and θ parameters through the "Yule-Walker" equations.

(4) The covariances γ(h) can be estimated without assuming a model and thus can serve as identifying functions to show what kind of model is appropriate.

(5) Each such series has a "spectral density" that decomposes the variation in a series into components at different frequencies. For example the series

-1, 1, -1 ,1, -1, 1, -1, 1, -1 , 1 , $\cdots$, 1

and

-1, -1, -1 , -1, -1, 1, 1, 1, 1 , 1 ,(<-repeat) $\cdots$, 1

both have mean 0 and variance 1 but the first fluctuates at a higher frequency than the second.

Regression

- Regression *may* be appropriate for time series
- Time t and seasonal dummies often used

- $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. Example, n=40, trend & quarterly effects

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ \vdots \\ Y_{40} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 1 \\ 1 & 4 & 0 & 0 & 0 \\ 1 & 5 & 1 & 0 & 0 \\ 1 & 6 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 40 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ \vdots \\ e_{40} \end{pmatrix}$$

Quarter 1: $Y_t = \alpha + \beta t + \delta_1 + e_t$
Quarter 2: $Y_t = \alpha + \beta t + \delta_2 + e_t$ Four parallel lines
Quarter 3: $Y_t = \alpha + \beta t + \delta_3 + e_t$
Quarter 4: $Y_t = \alpha + \beta t + 0 + e_t$

- $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ B.L.U.E. ***if errors are iid***

- $(\mathbf{X}'\mathbf{X})^{-1}$(MSE) is proper variance-covariance matrix ***if errors are iid***

- ***if errors are not iid then:***

  - $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ unbiased but not best
  - $(\mathbf{X}'\mathbf{X})^{-1}$(MSE) not appropriate
  - t tests, P-values, F tests *all wrong* (they use $(\mathbf{X}'\mathbf{X})^{-1}$(MSE))

- How to tell?

  Durbin-Watson test:
  Run regression as usual (Ordinary Least Squares, OLS)
  Get residuals r_t. Compute $D=\Sigma_{t=2}^{n}(r_t-r_{t-1})^2 / \Sigma_{t=1}^{n} r_t^2$
  For i.i.d. r_t you'd have $E\{ \Sigma_{t=2}^{n}(r_t-r_{t-1})^2 \} = 2(n-1)\sigma^2$ and $E\{\Sigma_{t=1}^{n} r_t^2\} = n\sigma^2$
  For i.i.d. D should be near 2.
  If r_t and r_{t-1} alike (positively correlated) $\Sigma_{t=2}^{n}(r_t-r_{t-1})^2$ smaller and D<2.
  Durbin-Watson give bounds on critical value and computationally intensive method to get exact p-values.

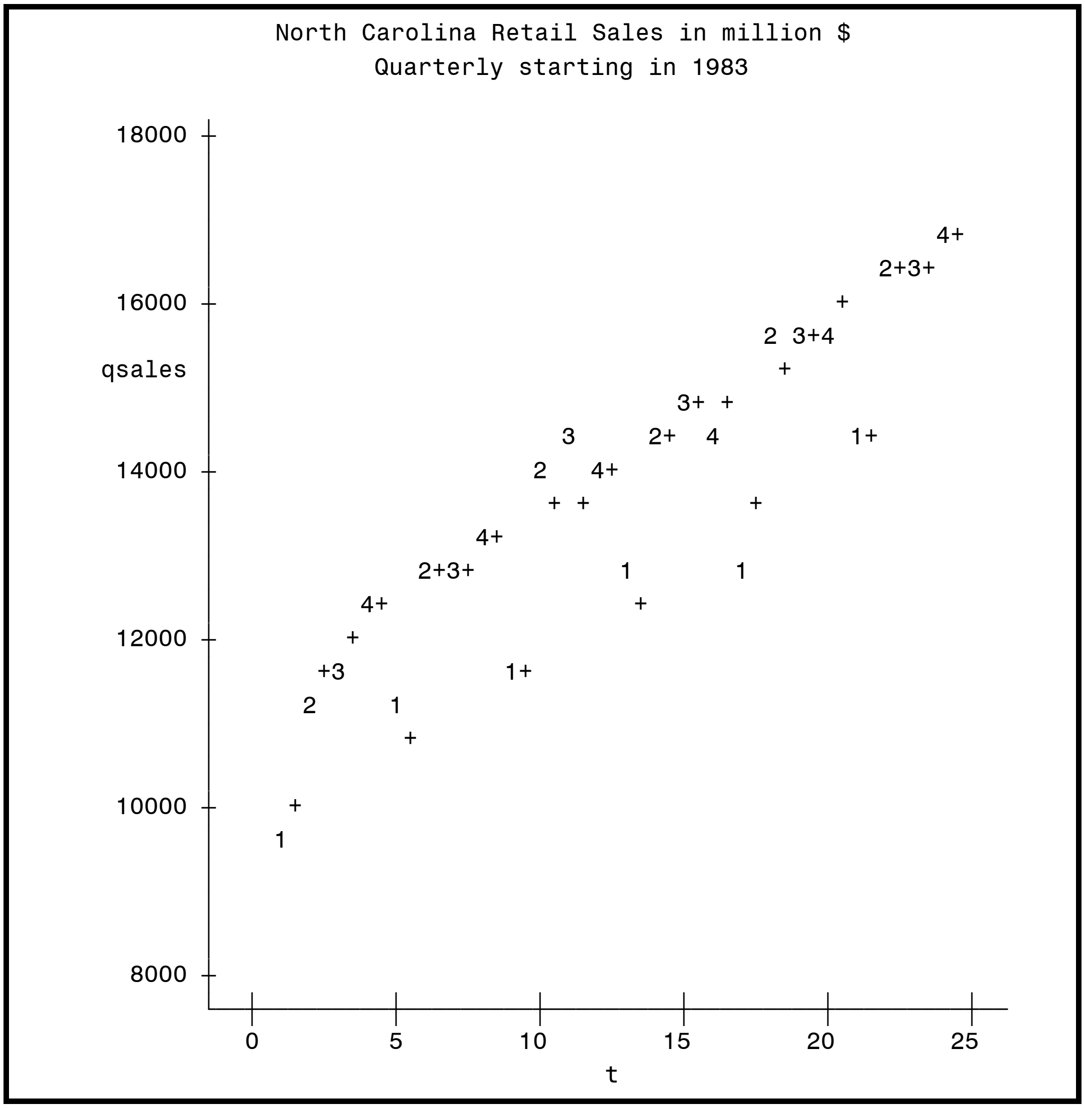
- Example: Quarterly NC retail sales.

```
options ls=76;
title 'North Carolina Retail Sales in million $';
title2 "Quarterly starting in 1983";
```

```
Data NCSALES;
input qsales t t2 s1 s2 s3 s4 date :yyq6.;
qtr=qtr(date);  x=t+.3; *(for graphs);
cards;
           9485.68     1       1     1     0     0     0    1983Q1
          11164.09     2       4     0     1     0     0    1983Q2
                    (more data)
          16829.22    24     576     0     0     0     1    1988Q4
  ;
data next; set ncsales;
proc reg;
model qsales = t S1 S2 S3/dw;
output out=out1 p=pred;
* Autoreg shows Durbin-Watson 1.2190    Pr < DW 0.0289;
proc plot; plot  qsales*t=qtr pred*x="+"/overlay vpos=26;
run;
```

```
                         The REG Procedure
                           Model: MODEL1
                    Dependent Variable: qsales

                       Analysis of Variance

                                 Sum of          Mean
Source               DF         Squares        Square    F Value    Pr > F
Model                 4        78085176      19521294     189.90    <.0001
Error                19         1953171        102798
Corrected Total      23        80038346

                        Parameter Estimates

                          Parameter     Standard
   Variable     DF         Estimate        Error    t Value    Pr > |t|
   Intercept     1            11342    187.41080      60.52      <.0001
   t             1        224.91923      9.58041      23.48      <.0001
   s1            1      -1737.00230    187.32915      -9.27      <.0001
   s2            1        -49.68154    186.10022      -0.27      0.7924
   s3            1        -82.12577    185.35894      -0.44      0.6627

                 Durbin-Watson D                 1.219
                 Number of Observations             24
                 1st Order Autocorrelation       0.358
```

```
* Autoreg shows Durbin-Watson 1.2190    Pr < DW 0.0289;
```



- To predict, concatenate future values of all X's to data (set Y = missing (.))
- Advantages:
  - Easy to understand and implement
  - Picks up very regualr trends and seasonal patterns (dummy variables)
  - Has check for autocorrelation (Durbin_Watson)
- Disadvantages
  - Not flexible (changing trend, seasonal difficult to model)
  - Autocorrelation destroys inference (but see PROC AUTOREG later)
  - Need future values of input variables.

Transformations

- Most common: no transformation or log
  Try log if variation increases as mean increases.

- Box-Cox family (see Steel et al, St 512 text, page 246)

  Power transformation Y^{λ}, e.g. $Y^{\frac{1}{2}} = \sqrt{Y}$

  Fit model to X where X= $(Y^{\lambda}\text{-}1)/(\lambda \dot{Y}^{\lambda})$ for grid of λ values
  Let X= $\dot{Y}$ ln(Y) for λ=0.

  Plot MSE or likelihood versus λ and pick optimal λ from plot.

- Plot on next page is

  Dow Jones (upper left) Log(Dow) upper right
  Log difference (lower panel)
  Transformation can have big effect.

- 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: $\text{Log}(1+\epsilon) = \log(1) + 1\epsilon - \epsilon^2/2 + \cdots \approx \epsilon$ for ϵ 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
```

- $\log(Y)=X \sim N(\mu,\sigma^2)$ => $E\{Y\} = E\{e^X\} = \frac{1}{\sqrt{2\pi}\,\sigma}\int e^X e^{-\frac{(X-\mu)^2}{2\sigma^2}}dx =$

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

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

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