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:

```plaintext
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.

<table>
<thead>
<tr>
<th>Obs</th>
<th>Diff</th>
<th>dec1919</th>
<th>dec19</th>
<th>jan20</th>
<th>date1</th>
<th>date2</th>
<th>date3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>-14611</td>
<td>21914</td>
<td>-14610</td>
<td>JAN60</td>
<td>19JAN1960</td>
<td>01/19/60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>date1</th>
<th>date2</th>
<th>date3</th>
<th>x</th>
<th>i</th>
<th>twoi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-17472</td>
<td>-17411</td>
<td>-17350</td>
<td>31</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>date1</th>
<th>date2</th>
<th>date3</th>
<th>x</th>
<th>i</th>
<th>twoi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>03/01/12</td>
<td>05/01/12</td>
<td>07/01/12</td>
<td>31</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>date</th>
<th>Y</th>
<th>newdate</th>
<th>newdate2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Apr82</td>
<td>30</td>
<td>8126</td>
<td>APR82</td>
</tr>
<tr>
<td>2</td>
<td>Feb83</td>
<td>20</td>
<td>8432</td>
<td>FEB83</td>
</tr>
<tr>
<td>3</td>
<td>Jan82</td>
<td>10</td>
<td>8036</td>
<td>JAN82</td>
</tr>
<tr>
<td>4</td>
<td>Jan83</td>
<td>50</td>
<td>8401</td>
<td>JAN83</td>
</tr>
<tr>
<td>5</td>
<td>Jul82</td>
<td>60</td>
<td>8217</td>
<td>JUL82</td>
</tr>
<tr>
<td>6</td>
<td>Mar83</td>
<td>35</td>
<td>8460</td>
<td>MAR83</td>
</tr>
<tr>
<td>7</td>
<td>Oct82</td>
<td>40</td>
<td>8309</td>
<td>OCT82</td>
</tr>
</tbody>
</table>
Applied Time Series Notes

<table>
<thead>
<tr>
<th>Obs</th>
<th>date</th>
<th>Y</th>
<th>newdate</th>
<th>newdate2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan82</td>
<td>10</td>
<td>8036</td>
<td>JAN82</td>
</tr>
<tr>
<td>2</td>
<td>Apr82</td>
<td>30</td>
<td>8126</td>
<td>APR82</td>
</tr>
<tr>
<td>3</td>
<td>Jul82</td>
<td>60</td>
<td>8217</td>
<td>JUL82</td>
</tr>
<tr>
<td>4</td>
<td>Oct82</td>
<td>40</td>
<td>8309</td>
<td>OCT82</td>
</tr>
<tr>
<td>5</td>
<td>Jan83</td>
<td>50</td>
<td>8401</td>
<td>JAN83</td>
</tr>
<tr>
<td>6</td>
<td>Feb83</td>
<td>20</td>
<td>8432</td>
<td>FEB83</td>
</tr>
<tr>
<td>7</td>
<td>Mar83</td>
<td>35</td>
<td>8460</td>
<td>MAR83</td>
</tr>
</tbody>
</table>

PROC EXPAND (cubic spline)

```sas
data last; input Y @@; date=intnx('month','01dec83'd,_n_); format date monyy.; cards;
10 . . 12 18 40 . 13 18 . . . 10 . 10 10
```

```sas
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

<table>
<thead>
<tr>
<th>Obs</th>
<th>Y</th>
<th>date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>JAN84</td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>FEB84</td>
</tr>
<tr>
<td>3</td>
<td>.</td>
<td>MAR84</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>APR84</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>MAY84</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>JUN84</td>
</tr>
<tr>
<td>7</td>
<td>.</td>
<td>JUL84</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>AUG84</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>SEP84</td>
</tr>
<tr>
<td>10</td>
<td>.</td>
<td>OCT84</td>
</tr>
</tbody>
</table>
Plot of $Y$*date. Symbol used is '*'.
Plot of $y_{new}$*date. Symbol used is '-'.
Time Series Example 1

Plot of $Y$ vs $date$. Symbol used is '*'.
Plot of $Y_{wk}$ vs $date$. 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 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;
• 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
  \[ E = \text{expectation} = \text{average in population. } \mathbb{E}\{X\} = \mu, \mathbb{E}\{(X-\mu)^2\} = \sigma^2 \]

  • 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 ( \( \mathbb{E}(Y_{n+L} - \hat{Y}_{n+L}) = 0 \))
  • Best (pick \( b_j \) to minimize something, e.g. \( \mathbb{E}(Y_{n+L} - \hat{Y}_{n+L})^2 \))

\[
\begin{pmatrix}
Y_1 \\
Y_2 \\
Y_3
\end{pmatrix}
= \begin{pmatrix}
Y_1 \\
v_{12} \\
v_{13}
v_{21} \\
v_{22} \\
v_{23}
v_{31} \\
v_{32} \\
v_{33}
\end{pmatrix}
\text{ } \mathbb{E}\begin{pmatrix}
Y_1 \\
Y_2 \\
Y_3
\end{pmatrix}
= \begin{pmatrix}
\mu_1 \\
\mu_2 \\
\mu_3
\end{pmatrix}
\]

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

\[\text{Var}\{ Y_{3-\mu_3} - b_1(Y_1 - \mu_1) - b_2(Y_2 - \mu_2) \} = ( \text{from St 512 !!} ) \]

\[
\begin{pmatrix}
-b_1 \\
-b_2 \\
1
\end{pmatrix}
\begin{pmatrix}
v_{11} & v_{12} & v_{13} \\
v_{21} & v_{22} & v_{23} \\
v_{31} & v_{32} & v_{33}
\end{pmatrix}
\begin{pmatrix}
-b_1 \\
-b_2 \\
1
\end{pmatrix}
\]

• Minimize! (set \( \frac{\partial}{\partial b_j}(\ast) = 0 \) for \( j=1,2 \))

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

• Example: \( \mu_3 = \mu_2 = \mu_1 = 100, Y_1 = 120, Y_2 = 180 \)
\[ Y_1 = \begin{pmatrix} 8 & 4 & 2 \\ 4 & 8 & 4 \\ 2 & 4 & 8 \end{pmatrix} \]
\[ b_2 = \begin{pmatrix} b_{1} \\ b_{2} \\ b_{3} \end{pmatrix} = \begin{pmatrix} 1 \\ 64.16 \\ 2 \end{pmatrix} \begin{pmatrix} 8 & -4 & 0 \\ -4 & 8 & 24 \\ 0 & 24 & 1/2 \end{pmatrix} \]

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

(A) Toeplitz covariance matrix: \( \text{Cov}(Y_1, Y_3) = \gamma(|i-j|) \) is a function of \(|i-j|\) only.

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

(B) Means all the same

(A) \& (B) \overset{\text{defn.}}{\implies} "COVARIANCE STATIONARY" or just "STATIONARY"

True or false for Toeplitz covariance matrix:

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

- 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

\[
\begin{pmatrix}
10 & 8 & 0 \\
8 & 10 & 8 \\
0 & 8 & 10 \\
\end{pmatrix}
\]
looks like MA(1) but variance of \( W = (1, -1, 1) \)

\[
\begin{pmatrix}
Y_1 \\
Y_2 \\
Y_3 \\
\end{pmatrix}
\]

would then be

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

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 \( \gamma(h) \) are related to the \( \alpha \) and \( \theta \) parameters through the "Yule-Walker" equations.

4. The covariances \( \gamma(h) \) 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
   \[-1, 1, -1, 1, -1, 1, -1, 1, \cdots, 1\]
   and
   \[-1, -1, -1, -1, 1, 1, 1, 1, \cdots, 1\]
   both have mean 0 and variance 1 but the first fluctuates at a higher frequency than the second.

Regression

- Regression \textit{may} be appropriate for time series

- Time \( t \) and seasonal dummies often used
• \( Y = X\beta + e \). Example, \( n=40 \), trend & quarterly effects

\[
\begin{pmatrix}
Y_1 \\
Y_2 \\
Y_3 \\
Y_4 \\
Y_5 \\
Y_6 \\
\vdots \\
Y_{40}
\end{pmatrix} =
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 \\
1 & 2 & 0 & 1 & 0 \\
1 & 3 & 0 & 0 & 1 \\
1 & 4 & 0 & 0 & 0 \\
1 & 5 & 1 & 0 & 0 \\
1 & 6 & 0 & 1 & 0 \\
\vdots \\
1 & 40 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\delta_1 \\
\delta_2 \\
\delta_3 \\
\delta_4 \\
\vdots \\
e_{40}
\end{pmatrix} +
\begin{pmatrix}
e_1 \\
e_2 \\
e_3 \\
e_4 \\
e_5 \\
e_6 \\
\vdots \\
e_{40}
\end{pmatrix}
\]

Quarter 1: \( Y_1 = \alpha + \beta t + \delta_1 + e_t \)
Quarter 2: \( Y_1 = \alpha + \beta t + \delta_2 + e_t \) \hspace{1cm} Four parallel lines
Quarter 3: \( Y_1 = \alpha + \beta t + \delta_3 + e_t \)
Quarter 4: \( Y_1 = \alpha + \beta t + 0 + e_t \)

\( \hat{\beta} = (X'X)^{-1}(X'Y) \) \hspace{1cm} B.L.U.E. \hspace{0.5cm} if errors are iid

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

if errors are not iid then:

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

• How to tell?

Durbin-Watson test:
Run regression as usual (Ordinary Least Squares, OLS)
Get residuals \( r_t \). Compute \( D = \sum_{t=2}^{n}(r_t-r_{t-1})^2 / \sum_{t=1}^{n} r_t^2 \)
For i.i.d. \( r_t \) you'd have \( E\{ \sum_{t=2}^{n}(r_t-r_{t-1})^2 \} = 2(n-1)\sigma^2 \) and \( E\{ \sum_{t=1}^{n} r_t^2 \} = n\sigma^2 \)
For i.i.d. \( D \) should be near 2.
If \( r_t \) and \( r_{t-1} \) alike (positively correlated) \( \sum_{t=2}^{n}(r_t-r_{t-1})^2 \) 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.

```plaintext
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.;
qutr=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

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>78085176</td>
<td>19521294</td>
<td>189.90</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>19</td>
<td>1953171</td>
<td>102798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>23</td>
<td>80038346</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard 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
To predict, concatenate future values of all X's to data (set Y = missing (.)

Advantages:
- Easy to understand and implement
- Picks up very regular 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)

  \[ Y^\lambda \], e.g. \( Y^{\frac{1}{2}} = \sqrt{Y} \)

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

  Plot MSE or likelihood versus \( \lambda \) and pick optimal \( \lambda \) 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)

  \[ \log(Y_t) - \log(Y_{t-1}) = \log(Y_t/Y_{t-1}) \]
  Taylor’s series: \( \log(1+\epsilon) = \log(1) + \epsilon - \epsilon^2/2 + \cdots \approx \epsilon \) for \( \epsilon \) small.
  Thus 100 \( \log(Y_t/Y_{t-1}) \) 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;
```

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
<th>X8</th>
<th>X9</th>
<th>X10</th>
<th>X11</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>0.92</td>
<td>0.94</td>
<td>0.96</td>
<td>0.98</td>
<td>1.00</td>
<td>1.02</td>
<td>1.04</td>
<td>1.06</td>
<td>1.08</td>
<td>1.10</td>
</tr>
<tr>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.08</td>
<td>0.10</td>
</tr>
</tbody>
</table>

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

  \[
  \frac{1}{\sqrt{2\pi}\sigma} \int e^{-\frac{x^2-2\mu x+\mu^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} e^{\mu+\sigma^2/2} \int e^{-\frac{x^2-2\mu x+\mu^2}{2\sigma^2}} dx = e^{\mu+\sigma^2/2} \text{ (not just } e^\mu)\]

- 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, \( e^\mu \) is median of \( Y \) (not mean).