Module 3

Segment 1

Populations, Processes, and
Descriptive Time Series Statistics
Populations

- A population is a collection of identifiable units (or a specified characteristic of these units).

- A frame is a listing of the units in the population.

- We take a sample from the frame and use the resulting data to make inferences about the population.

- Simple random sampling with replacement from a population implies independent and identically distributed (iid) observations.

- Standard statistical methods use the iid assumption for observations or residuals (in regression).
Processes

- A process is a system that transforms inputs into outputs, operating over time.

- A process generates a sequence of observations (data) over time.

- We can use a realization from the process to make inferences about the process.

- The iid model is rarely appropriate for processes (observations close together in time are typically correlated and the process often changes with time).
Process, Realization, Model, Inference, and Applications

- **Process**: Generates Data
- **Realization**: $Z_1, Z_2, \ldots, Z_n$
- **Statistical Model for the Process**: (Stochastic Process Model)
- **Estimates for Model Parameters**
- **Inferences**: Forecasts, Control
- **Description**
Stationary Stochastic Processes

$Z_t$ is a stochastic process. Some properties of $Z_t$ include mean $\mu_t$, variance $\sigma^2_t$, and autocorrelation $\rho_{Z_t,Z_{t+k}}$. In general, these can change over time.

- Strongly stationary (also strictly or completely stationary):

  $$F(z_{t_1}, \ldots, z_{t_n}) = F(z_{t_1+k}, \ldots, z_{t_n+k})$$

  Difficult to check.

- Weakly stationary (or 2nd order or covariance stationary) requires only that $\mu = \mu_t$ and $\sigma^2 = \sigma^2_t$ be constant and that $\rho_k = \rho_{Z_t,Z_{t+k}}$ depend only on $k$.

  Easy to check with sample statistics.

  Generally, “stationary” is understood to be weakly stationary.
Change in Business Inventories 1955-1969

Stationary?

Year

Billions of Dollars

## Estimation of Stationary Process Parameters

### Some Notation

<table>
<thead>
<tr>
<th>Process Parameter</th>
<th>Notation</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of $Z$</td>
<td>$\mu_Z = E(Z)$</td>
<td>$\hat{\mu}<em>Z = \bar{z} = \frac{\sum</em>{t=1}^n z_t}{n}$</td>
</tr>
<tr>
<td>Variance of $Z$</td>
<td>$\gamma_0 = \sigma_Z^2 = \text{Var}(Z)$</td>
<td>$\hat{\gamma}<em>0 = \frac{\sum</em>{t=1}^n (z_t - \bar{z})^2}{n}$</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$\sigma_Z$</td>
<td>$\hat{\sigma}_Z = \sqrt{\hat{\gamma}_0}$</td>
</tr>
<tr>
<td>Autocovariance</td>
<td>$\gamma_k$</td>
<td>$\hat{\gamma}<em>k = \frac{\sum</em>{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{n}$</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>$\rho_k = \frac{\gamma_k}{\gamma_0}$</td>
<td>$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\gamma_0}$</td>
</tr>
<tr>
<td>Variance of $\bar{Z}$</td>
<td>$\sigma_{\bar{Z}}^2 = \text{Var}(\bar{Z})$</td>
<td>$S^2_{\bar{Z}} = \frac{\hat{\gamma}_0}{n} [\cdots]$</td>
</tr>
</tbody>
</table>
Change in Business Inventories 1955-1969
with \( \bar{z} \) and \( \bar{z} \pm 3\hat{\sigma}_Z \)
Variance of $\bar{Z}$ with Correlated Data

The Variance of $\bar{Z}$

$$\sigma_{\bar{Z}}^2 = \text{Var}(\bar{Z}) = \frac{\gamma_0}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho_k$$

with correlated data is more complicated than with iid data where

$$\sigma_{\bar{Z}}^2 = \text{Var}(\bar{Z}) = \frac{\gamma_0}{n} = \frac{\sigma^2}{n}$$

Thus if one incorrectly uses the iid formula, the variance is under estimated, potentially leading for false conclusions.
Module 3

Segment 2

Correlation and Autocorrelation
Correlation [From “Statistics 101”]

Consider random data $y$ and $x$ (e.g., sales and advertising):

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$y_1$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$y_2$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$x_n$</td>
<td>$y_n$</td>
</tr>
</tbody>
</table>

$\rho_{x,y}$ denotes the “population” correlation between all values of $x$ and $y$ in the population.

To estimate $\rho_{x,y}$, we use the sample correlation

$$\hat{\rho}_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \quad \text{(Stat 101)}$$

$$-1 \leq \hat{\rho}_{x,y} \leq 1$$
Sample Autocorrelation

Consider the random time series realization $z_1, z_2, \ldots, z_n$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$z_t$</th>
<th>$z_{t+1}$</th>
<th>$z_{t+2}$</th>
<th>$z_{t+3}$</th>
<th>$z_{t+4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$z_1$</td>
<td>$z_2$</td>
<td>$z_3$</td>
<td>$z_4$</td>
<td>$z_5$</td>
</tr>
<tr>
<td>2</td>
<td>$z_2$</td>
<td>$z_3$</td>
<td>$z_4$</td>
<td>$z_5$</td>
<td>$z_6$</td>
</tr>
<tr>
<td>3</td>
<td>$z_3$</td>
<td>$z_4$</td>
<td>$z_5$</td>
<td>$z_6$</td>
<td>$z_7$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$n-2$</td>
<td>$z_{n-2}$</td>
<td>$z_{n-1}$</td>
<td>$z_n$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$n-1$</td>
<td>$z_{n-1}$</td>
<td>$z_n$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$n$</td>
<td>$z_n$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

Assuming weak (covariance) stationarity, let $\rho_k$ denote the process correlation between observations separated by $k$ time periods. To compute the "order $k$" sample autocorrelation (i.e., correlation between $z_t$ and $z_{t+k}$)

$$
\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{n}(z_t - \bar{z})^2}, \quad k = 0, 1, 2, \ldots
$$

Note: $-1 \leq \hat{\rho}_k \leq 1$
Sample Autocorrelation (alternative formula)

Consider the random time series realization $z_1, z_2, \ldots, z_n$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$z_t$</th>
<th>$z_{t-1}$</th>
<th>$z_{t-2}$</th>
<th>$z_{t-3}$</th>
<th>$z_{t-4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$z_1$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>2</td>
<td>$z_2$</td>
<td>$z_1$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>3</td>
<td>$z_3$</td>
<td>$z_2$</td>
<td>$z_1$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>4</td>
<td>$z_4$</td>
<td>$z_3$</td>
<td>$z_2$</td>
<td>$z_1$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$n$</td>
<td>$z_n$</td>
<td>$z_{n-1}$</td>
<td>$z_{n-2}$</td>
<td>$z_{n-3}$</td>
<td>$z_{n-4}$</td>
</tr>
</tbody>
</table>

Assuming weak (covariance) stationarity, let $\rho_k$ denote the process correlation between observations separated by $k$ time periods.

To compute the “order $k$” sample autocorrelation (i.e., correlation between $z_t$ and $z_{t-k}$)

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=k+1}^{n}(z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^{n}(z_t - \bar{z})^2}, \quad k = 0, 1, 2, \ldots$$

Note: $-1 \leq \hat{\rho}_k \leq 1$
Wolfer Sunspot Numbers 1770-1869
Lagged Sunspot Data

<table>
<thead>
<tr>
<th>$t$</th>
<th>Spot</th>
<th>Spot1</th>
<th>Spot2</th>
<th>Spot3</th>
<th>Spot4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>101</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>82</td>
<td>101</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
<td>82</td>
<td>101</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>66</td>
<td>82</td>
<td>101</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>35</td>
<td>66</td>
<td>82</td>
<td>101</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>99</td>
<td>37</td>
<td>7</td>
<td>16</td>
<td>30</td>
<td>47</td>
</tr>
<tr>
<td>100</td>
<td>74</td>
<td>37</td>
<td>7</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>101</td>
<td>–</td>
<td>74</td>
<td>37</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>102</td>
<td>–</td>
<td>–</td>
<td>74</td>
<td>37</td>
<td>7</td>
</tr>
<tr>
<td>103</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>74</td>
<td>37</td>
</tr>
<tr>
<td>104</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>74</td>
</tr>
</tbody>
</table>
Wolfer Sunspot Numbers
Correlation Between Observations Separated by One Time Period

Correlation = 0.81708
Autocorrelation = 0.806244
Wolfer Sunspot Numbers
Correlation Between Observations Separated by $k$
Time Periods \texttt{[show.acf(spot.ts)]}
Module 3

Segment 3

Autoregression, Autoregressive Modeling, and Introduction to Partial Autocorrelation
Wolfer Sunspot Numbers 1770-1869
Autoregressive (AR) Models

• AR(0): \( z_t = \mu + a_t \) (White noise or “Trivial” model)
• AR(1): \( z_t = \theta_0 + \phi_1 z_{t-1} + a_t \)
• AR(2): \( z_t = \theta_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t \)
• AR(3): \( z_t = \theta_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + a_t \)
• AR(p): \( z_t = \theta_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t \)
  \( a_t \sim \text{nid}(0, \sigma_a^2) \)
Data Analysis Strategy

Tentative Identification

Estimation

Diagnostic Checking

Model ok?

Yes

Use the Model

No

Time Series Plot
Range-Mean Plot
ACF and PACF

Least Squares or
Maximum Likelihood

Residual Analysis and Forecasts

Forecasting
Explanation
Control
AR(1) Model for the Wolfer Sunspot Data

Data and 1-Step Ahead Predictions

Sunspot Residuals, AR(1) Model

Series : residuals(spot.ar1.fit)
AR(2) Model for the Wolfer Sunspot Data

Data and 1-Step Ahead Predictions

Sunspot Residuals, AR(2) Model

Series: residuals(spot.ar2.fit)
AR(3) Model for the Wolfer Sunspot Data

Data and 1-Step Ahead Predictions

Sunspot Residuals, AR(3) Model

Series: residuals(spot.ar3.fit)

ACF

Quantiles of Standard Normal

residuals(spot.ar2.fit)
AR(4) Model for the Wolfer Sunspot Data

Data and 1-Step Ahead Predictions

Sunspot Residuals, AR(4) Model

Series : residuals(spot.ar4.fit)
Sample Partial Autocorrelation Function

The sample PACF $\hat{\phi}_{kk}$ can be computed from the sample ACF $\hat{\rho}_1, \hat{\rho}_2, \ldots$ using the recursive formula from Durbin (1960):

$$\hat{\phi}_{1,1} = \hat{\rho}_1$$
$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j}, \quad k = 2, 3, \ldots$$

where

$$\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j} \quad (k = 3, 4, \ldots; \quad j = 1, 2, \ldots, k - 1)$$

Thus the PACF contains no new information—just a different way of looking at the information in the ACF.

This sample PACF formula is based on the solution of the Yule-Walker equations (covered module 5).
### Summary of Sunspot Autoregressions

<table>
<thead>
<tr>
<th>Model</th>
<th>Order</th>
<th>Regression Output</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R^2$</td>
<td>$S$</td>
</tr>
<tr>
<td>AR(1)</td>
<td>1</td>
<td>.6676</td>
<td>21.53</td>
</tr>
<tr>
<td>AR(2)</td>
<td>2</td>
<td>.8336</td>
<td>15.32</td>
</tr>
<tr>
<td>AR(3)</td>
<td>3</td>
<td>.8407</td>
<td>15.12</td>
</tr>
<tr>
<td>AR(4)</td>
<td>4</td>
<td>.8463</td>
<td>15.01</td>
</tr>
</tbody>
</table>

$\hat{\phi}_p$ is from ordinary least squares (OLS).

$\hat{\phi}_{pp}$ is from the Durbin formula.
Sample Partial Autocorrelation

The “true partial autocorrelation function,” denoted by $\phi_{kk}$, for $k = 1, 2, \ldots$ is the process correlation between observations separated by $k$ time periods (i.e., between $z_t$ and $z_{t+k}$) with the effect of the intermediate $z_{t+1}, \ldots, z_{t+k-1}$ removed.

We can estimate $\phi_{kk}$, $k = 1, 2, \ldots$ with $\hat{\phi}_{kk}$, $k = 1, 2, \ldots$

- $\hat{\phi}_{1,1} = \hat{\phi}_1$ from the AR(1) model
- $\hat{\phi}_{2,2} = \hat{\phi}_2$ from the AR(2) model
- $\hat{\phi}_{kk} = \hat{\phi}_k$ from the AR($k$) model

The Durbin formula gives somewhat different answers due to the basis of the estimator (ordinary least squares versus solution of the Yule-Walker equations).
Module 3

Segment 4

Introduction to ARMA Models and Time Series Modeling
Autoregressive-Moving Average (ARMA) Models (a.k.a. Box-Jenkins Models)

- **AR(0):** \( z_t = \mu + a_t \) (White noise or “Trivial” model)
- **AR(1):** \( z_t = \theta_0 + \phi_1 z_{t-1} + a_t \)
- **AR(2):** \( z_t = \theta_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t \)
- **AR(p):** \( z_t = \theta_0 + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + a_t \)
- **MA(1):** \( z_t = \theta_0 - \theta_1 a_{t-1} + a_t \)
- **MA(2):** \( z_t = \theta_0 - \theta_1 a_{t-1} - \theta_2 a_{t-2} + a_t \)
- **MA(q):** \( z_t = \theta_0 - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} + a_t \)
- **ARMA(1, 1):** \( z_t = \theta_0 + \phi_1 z_{t-1} - \theta_1 a_{t-1} + a_t \)
- **ARMA(p, q):** \( z_t = \theta_0 + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} \\
- \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} + a_t \)

\( a_t \sim \text{nid}(0, \sigma^2_a) \)
Graphical Output from RTSERIES Function
iden(spot.tsd)

Wolfer Sunspots
w= Number of Spots

Range-Mean Plot

ACF

Lag

PACF

Lag
Tabular Output from RTSERIES Function
iden(spot.tsd)

Identification Output for Wolfer Sunspots

w= Number of Spots

[1] "Standard deviation of the working series= 37.36504"

ACF

<table>
<thead>
<tr>
<th>Lag</th>
<th>Partial ACF</th>
<th>se</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.806243956</td>
<td>0.1000000</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.428105325</td>
<td>0.1516594</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.069611110</td>
<td>0.1632975</td>
</tr>
</tbody>
</table>

...
Standard Errors for Sample Autocorrelations and Sample Partial Autocorrelations

- **Sample ACF** standard error: \( S_{\hat{\rho}_k} = \sqrt{\left( \frac{1+2\hat{\rho}_1^2+\cdots+2\hat{\rho}_k^2}{n} \right)} \)

  Also can compute the “\( t \)-like” statistics \( t = \frac{\hat{\rho}_k}{S_{\hat{\rho}_k}} \).

- **Sample PACF** standard error: \( S_{\hat{\phi}_{kk}} = \frac{1}{\sqrt{n}} \)

  Use to compute “\( t \)-like” statistics \( t = \frac{\hat{\phi}_{kk}}{S_{\hat{\phi}_{kk}}} \) and set decision bounds.

- In long realizations from a stationary process, if the true parameter is really 0, then the distribution of a “\( t \)-like” statistic can be approximated by \( \text{N}(0,1) \).

- Values of \( \hat{\rho}_k \) and \( \hat{\phi}_{kk} \) may be judged to be different from zero if the “\( t \)-like” statistics are outside specified limits (±2 is often suggested; might use ±1.5 as a “warning”).
### Drawing Conclusions from Sample ACF’s and PACF’s

<table>
<thead>
<tr>
<th></th>
<th>PACF Dies Down</th>
<th>PACF Cutoff after $p &gt; 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACF Dies Down</td>
<td>ARMA</td>
<td>$AR(p)$</td>
</tr>
<tr>
<td>ACF Cutoff after $q &gt; 1$</td>
<td>$MA(q)$</td>
<td>Impossible</td>
</tr>
</tbody>
</table>

Technical details, background, and numerous examples will be presented in Modules 4 and 5.
Comments on Drawing Conclusions from Sample ACF’s and PACF’s

The “t-like” statistics should only be used as guidelines in model building because:

- An ARMA model is only an approximation to some true process
- Sampling distributions are complicated; the “t-like” statistics do not really follow a standard normal distribution (only approximately, in large samples, with a stationary process)
- Correlations among the $\hat{\rho}_k$ values for different $k$ (e.g., $\hat{\rho}_1$ and $\hat{\rho}_2$ may be correlated).
- Problems relating to simultaneous inference (looking at many different statistics simultaneously, confidence/significance levels have little meaning).