FIN307 MATLAB

TOOLBOX CHAPTER 1 – Financial Data Analysis PART02

Financial Data Analysis

Three properties:

✓Stationarity

✓Causality

✓Invertibility

Stationarity

A stationary time series is one whose properties do not depend on the time at which the series is observed.[1] So time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times. On the other hand, a white noise series is stationary — it does not matter when you observe it, it should look much the same at any period of time.

Some cases can be confusing — a time series with cyclic behaviour (but not trend or seasonality) is stationary. That is because the cycles are not of fixed length, so before we observe the series we cannot be sure where the peaks and troughs of the cycles will be.

In general, a stationary time series will have no predictable patterns in the long-term. Time plots will show the series to be roughly horizontal (although some cyclic behaviour is possible) with constant variance.

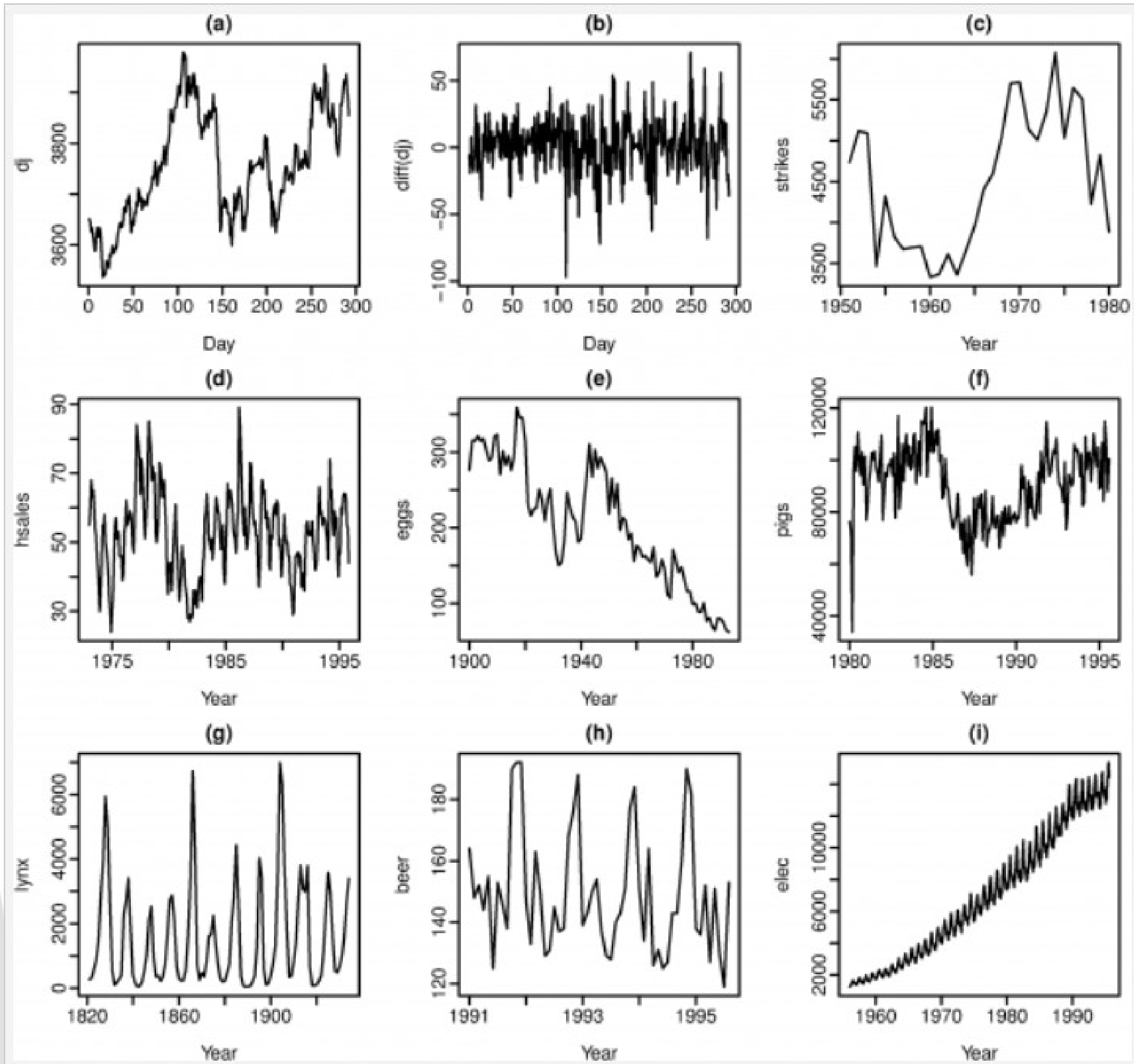


Figure 8.1: Which of these series are stationary? (a) Dow Jones index on 292 consecutive days; (b) Daily change in Dow Jones index on 292 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production.

Stationary Models and the Autocorrelation Function

Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$. The **mean function** of $\{X_t\}$ is

$$\mu_X(t) = E(X_t).$$

The **covariance function** of $\{X_t\}$ is

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

for all integers r and s.

$\{X_t\}$ is **(weakly) stationary** if

(i) $\mu_X(t)$ is independent of t,

and

(ii) $\gamma_X(t + h, t)$ is independent of t for each h.

Mathematical Expectation

Definition

Let X be a random variable with probability distribution $f(x)$. The mean, or expected value, of X is

$$\mu = E(X) = \sum_x x f(x)$$

if X is discrete, and

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

if X is continuous.

Example A salesperson for a medical device company has two appointments on a given day. At the first appointment, he believes that he has a 70% chance to make the deal, from which he can earn \$1000 commission if successful. On the other hand, he thinks he only has a 40% chance to make the deal at the second appointment, from which, if successful, he can make \$1500. What is his expected commission based on his own probability belief? Assume that the appointment results are independent of each other.

Solution commission totals: \$0, \$1000, \$1500, and \$2500

$f(\$0) = (1-0.7)(1-0.4) = 0.18$ $\quad f(\$1000) = (0.7)(1-0.4) = 0.42,$

$f(\$2500) = (0.7)(0.4) = 0.28$ $\quad f(\$1500) = (1-0.7)(0.4) = 0.12$

$$E(X) = \sum_x x f(x)$$

the expected commission for the salesperson is

$$E(X) = (\$0)(0.18) + (\$1000)(0.42) + (\$1500)(0.12) + (\$2500)(0.28) = \$1300.$$

Mathematical Expectation

Definition

Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

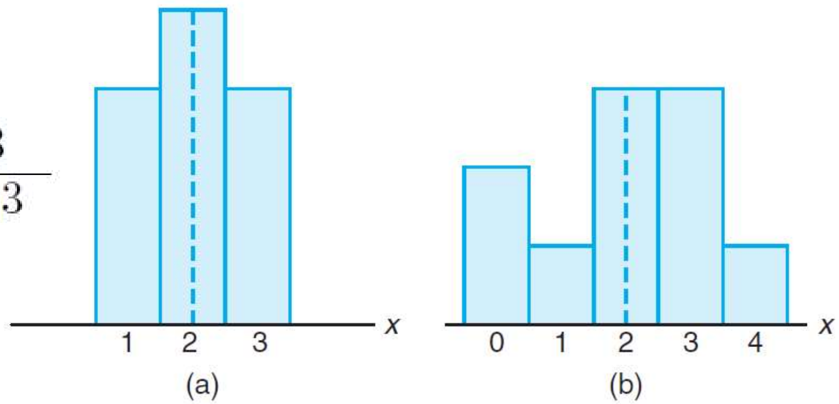
$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

Example Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A

x	1	2	3
$f(x)$	0.3	0.4	0.3



x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Figure : Distributions with equal means and unequal dispersions.

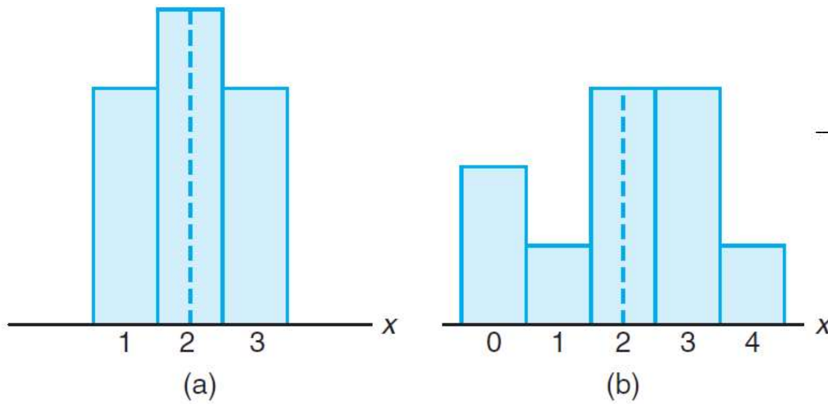
Solution

$$\mu_A = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

$$\sigma_A^2 = \sum_{x=1}^{3}(x-2)^2 = (1-2)^2(0.3) + (2-2)^2(0.4) + (3-2)^2(0.3) = 0.6.$$

Example Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A

x	1	2	3
$f(x)$	0.3	0.4	0.3



x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Figure : Distributions with equal means and unequal dispersions.

Solution

$$\mu_B = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0,$$

$$\sigma_B^2 = \sum_{x=0}^{4}(x-2)^2 f(x)$$

$$= (0-2)^2(0.2) + (1-2)^2(0.1) + (2-2)^2(0.3) + (3-2)^2(0.3) + (4-2)^2(0.1) = 1.6.$$

Theorem

The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

***Proof*:**

$$\sigma^2 = \sum_x (x-\mu)^2 f(x)$$

$$= \sum_x (x^2 - 2\mu x + \mu^2) f(x)$$

$$\mu = \sum_x x f(x)$$

$$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$$

$$\sum_x f(x) = 1$$

$$= \sum_x x^2 f(x) - \mu^2 = E(X^2) - \mu^2.$$

Example Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Using Theorem , calculate σ^2.

Solution

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87$$

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

Theorem

The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

covariance

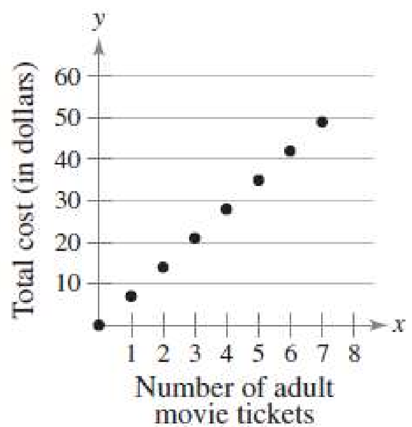
Let X and Y be random variables with joint probability distribution $f(x, y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_y) f(x, y)$$

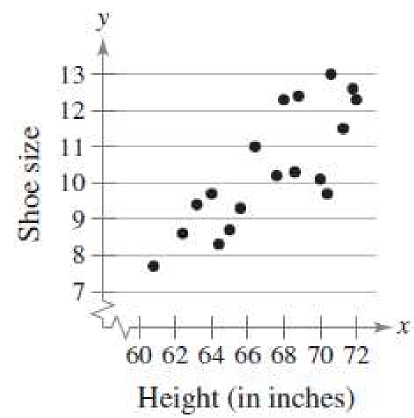
if X and Y are discrete, and

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_y) f(x, y) \, dx \, dy$$

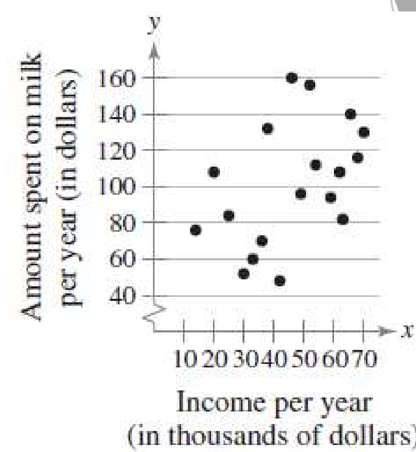
if X and Y are continuous.



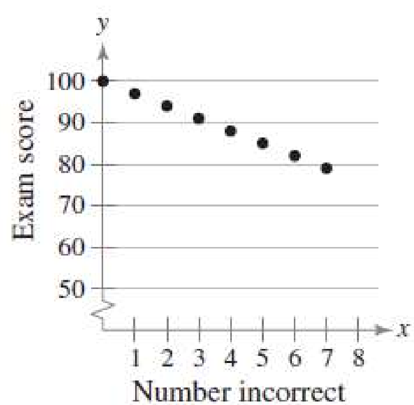
Perfect positive



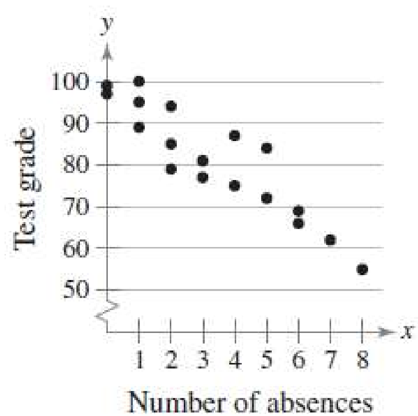
Strong positive



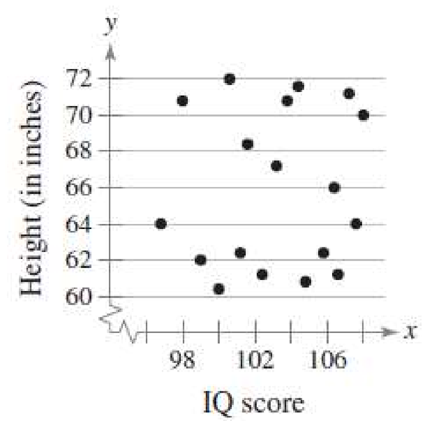
Weak positive



Perfect negative



Strong negative



No

sample covariance COV(x,y) $= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$

$$\sum(x-\bar{x})(y-\bar{y}) = \sum(xy - x\bar{y} - y\bar{x} + \bar{x}\bar{y})$$

$$= \sum(xy) - \sum(x\bar{y}) - \sum(y\bar{x}) + \sum(\bar{x}\bar{y})$$

$$= \sum(xy) - \frac{\sum(y)}{n}\sum(x) - \frac{\sum(x)}{n}\sum(y) + n\frac{\sum(x)}{n}\frac{\sum(y)}{n}$$

$$= \sum(xy) - \frac{\sum(x)\sum(y)}{n}$$

$$COV(x,y) = \frac{1}{n-1}\left(\sum(xy) - \frac{\sum(x)\sum(y)}{n}\right)$$

Theorem

The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y.$$

***Proof*:**

$$\sigma_{XY} = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)f(x,y)$$

$$= \sum_x \sum_y xyf(x,y) \; - \mu_X \sum_x \sum_y yf(x,y) - \mu_Y \sum_x \sum_y xf(x,y) \; + \mu_X\mu_Y \sum_x \sum_y f(x,y)$$

$$= E(XY) \qquad - \mu_X\mu_Y \qquad - \mu_Y\mu_X \qquad + \mu_X\mu_Y$$

$$= E(XY) - \mu_X\mu_Y$$

$$\mu_X = \sum_x xf(x,y), \quad \mu_Y = \sum_y yf(x,y), \text{ and } \sum_x \sum_y f(x,y) = 1$$

Example Find the covariance of X and Y.

Table : Joint Probability Distribution

	$f(x,y)$	$x = 0$	$x = 1$	$x = 2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Solution

$$\mu_X = \sum_{x=0}^{2} x g(x)$$

$$= (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4}$$

$$\mu_Y = \sum_{y=0}^{2} y h(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}.$$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

Stationary Models and the Autocorrelation Function

Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$. The **mean function** of $\{X_t\}$ is

$$\mu_X(t) = E(X_t).$$

The **covariance function** of $\{X_t\}$ is

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

for all integers r and s.

$\{X_t\}$ is **(weakly) stationary** if

(i) $\mu_X(t)$ is independent of t,

and

(ii) $\gamma_X(t + h, t)$ is independent of t for each h.

Covariance - Properties

Given a constant a and random variables X, Y, and Z, the following properties hold:

- $\mathrm{Cov}(X, X) = \mathrm{Var}(X) \geq 0$
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X, a) = 0$
- $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$.

$$\mathrm{cov}(X, X) = \mathbb{E}\left[(X-\mu)^2\right] = \mathrm{var}(X)$$

$$\mathrm{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\mathrm{cov}(X+Y, Z) = \mathbb{E}\left[(X+Y)Z\right] - \mathbb{E}(X+Y)\mathbb{E}(Z)$$

$$= \mathbb{E}(XZ + YZ) - \left[\mathbb{E}(X) + \mathbb{E}(Y)\right]\mathbb{E}(Z)$$

$$= \left[\mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z)\right] + \left[\mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z)\right]$$

$$= \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$$

$$\mathrm{cov}(cX, Y) = \mathbb{E}(cXY) - \mathbb{E}(cX)\mathbb{E}(Y)$$

$$= c\mathbb{E}(XY) - c\mathbb{E}(X)\mathbb{E}(Y)$$

$$= c[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$= c\,\mathrm{cov}(X, Y)$$

$$\mathrm{Var}(X) = E[(X - E[X])^2]$$

$$\mathrm{Cov}(X,Y) = E\left[(X - E[X])(Y - E[Y])\right]$$

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[XY - E[X]Y - XE[Y] + E[X]E[Y]] \\ &= \boxed{E[XY] - E[X]E[Y].}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(X+Y) &= E\left[(X + Y - E[X] - E[Y])^2\right] \\ &= E[(X - E[X])\ ] + E[(Y - E[Y])\ ]^2 \\ &= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2E\left[(X - E[X])(Y - E[Y])\right] \\ &= \boxed{\mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y).}\end{aligned}$$

$\mathrm{Var}(X) = \sigma^2$ and $Y = aX$, where σ and a are constants

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,aX) = \mathrm{Cov}(aX,X) = a\mathrm{Cov}(X,X) = a\sigma^2$$

$$\mathrm{Cov}(X,Y) = E\left[(X - E[X])(Y - E[Y])\right]$$

$$\mathrm{Cov}(aX+b, cY)$$
$$= \mathrm{E}\left[\{aX + b - \mathrm{E}(aX+b)\}\{cY - \mathrm{E}(cY)\}\right]$$
$$= \mathrm{E}\left[\{aX + b - (a\mathrm{E}(X) + b)\}\{cY - c\mathrm{E}(Y)\}\right]$$
$$= \mathrm{E}\left[\{aX - a\mathrm{E}(X)\}c\{Y - \mathrm{E}(Y)\}\right]$$
$$= \mathrm{E}\left[a\{X - \mathrm{E}(X)\}c\{Y - \mathrm{E}(Y)\}\right]$$
$$= ac\,\mathrm{E}\left[\{X - \mathrm{E}(X)\}\{Y - \mathrm{E}(Y)\}\right]$$
$$= \quad ac\,\mathrm{Cov}(X,Y)$$

Autocorrelation function

Let $\{X_t\}$ be a stationary time series. The **autocovariance function** (ACVF) of $\{X_t\}$ at lag h is

$$\gamma_X(h) = \operatorname{Cov}(X_{t+h}, X_t).$$

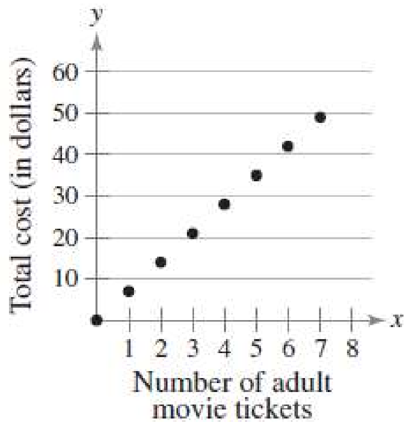
The **autocorrelation function** (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = \operatorname{Cor}(X_{t+h}, X_t).$$

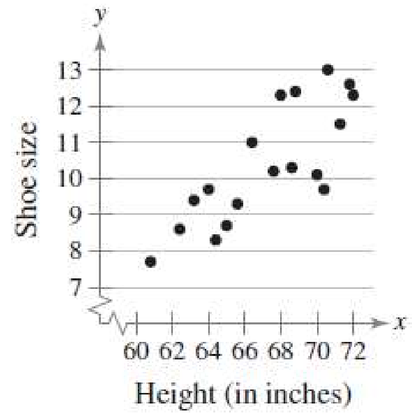
The value of r_k can be written as

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2},$$

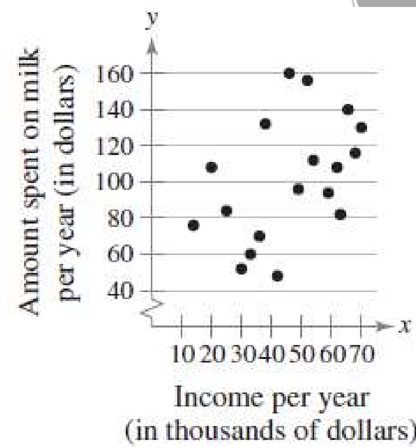
where T is the length of the time series.



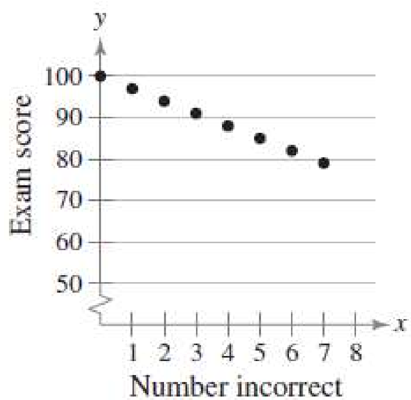
Perfect positive correlation
$r = 1$



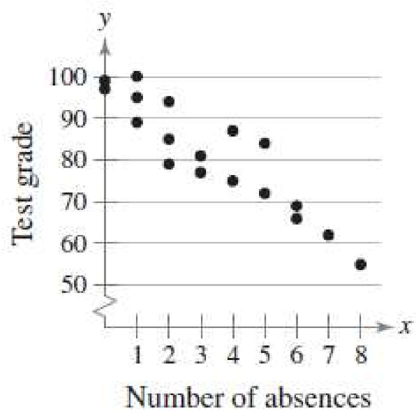
Strong positive correlation
$r = 0.81$



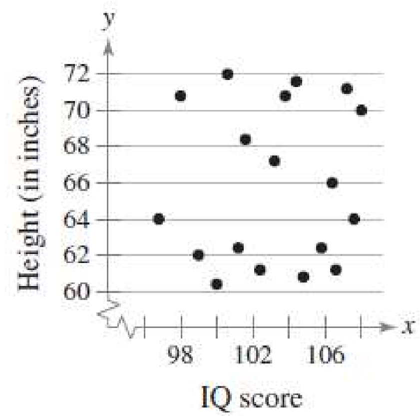
Weak positive correlation
$r = 0.45$



Perfect negative correlation
$r = -1$



Strong negative correlation
$r = -0.92$



No correlation
$r = 0.04$

The value of r_k can be written as

$$r_k = \frac{\sum\limits_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2},$$

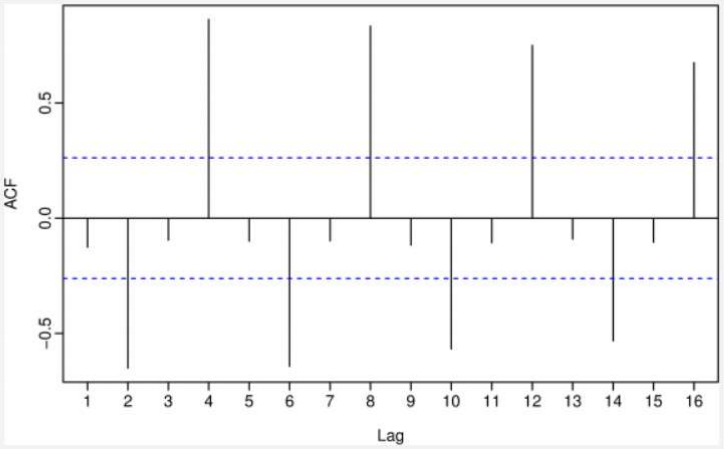


Figure 2.10: Autocorrelation function of quarterly beer production

where T is the length of the time series.

The first nine autocorrelation coefficients for the beer production data are given in the following table.

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
-0.126	-0.650	-0.094	0.863	-0.099	-0.642	-0.098	0.834	-0.116

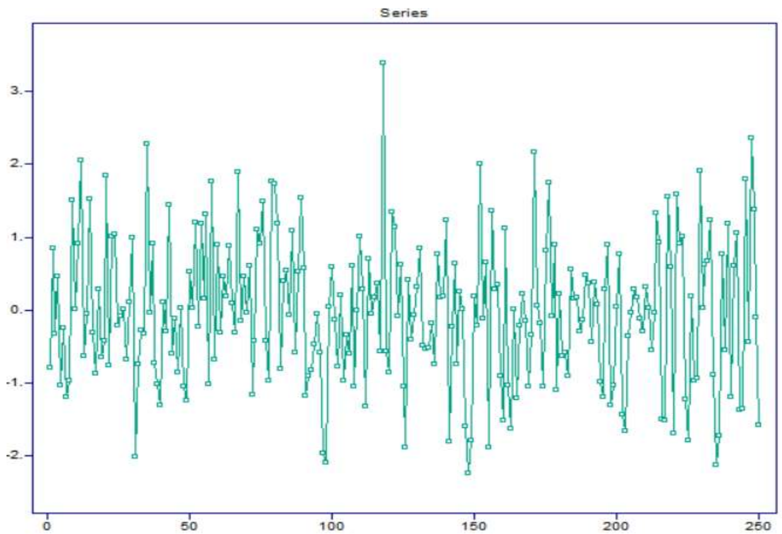
These correspond to the nine scatterplots in the graph above. The autocorrelation coefficients are normally plotted to form the *autocorrelation function* or ACF. The plot is also known as a *correlogram*.

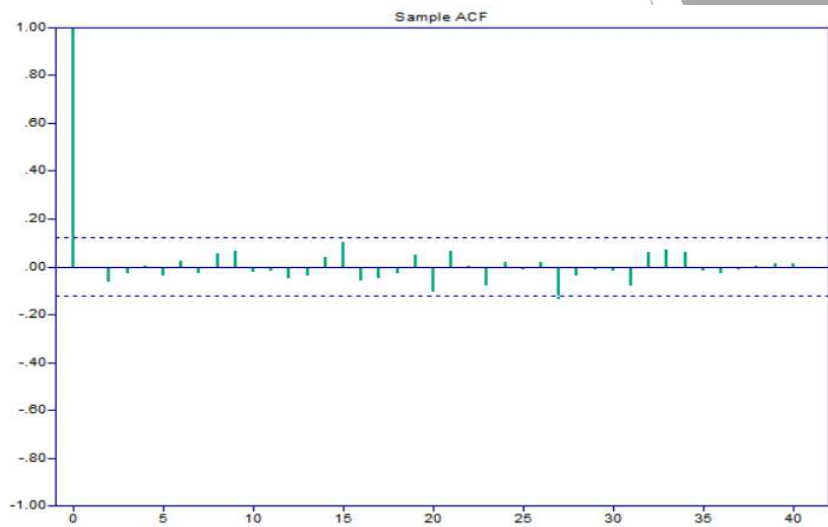
In this graph:

- r_4 is higher than for the other lags. This is due to the seasonal pattern in the data: the peaks tend to be four quarters apart and the troughs tend to be two quarters apart.
- r_2 is more negative than for the other lags because troughs tend to be two quarters behind peaks.

Series and its ACF (correlogram)

White noise

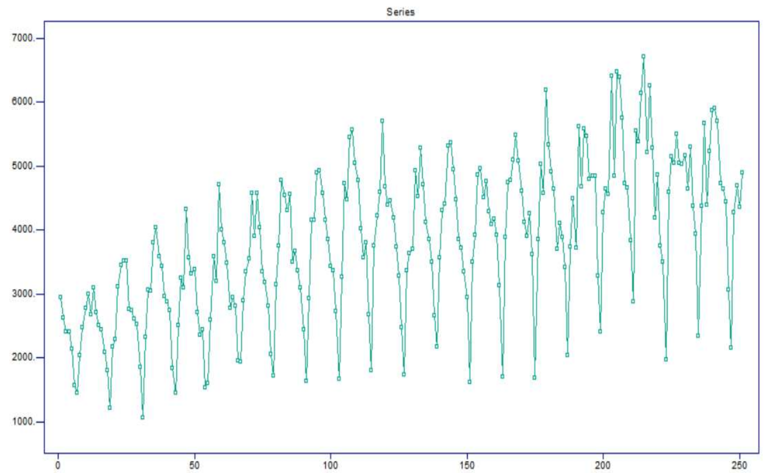


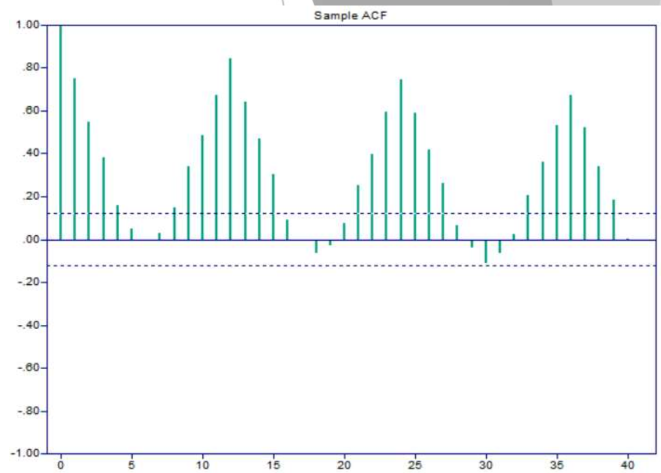



```
===============
ITSM::(ACF/PACF)
===============
# of Lags =        10

Sample Autocorrelations:

Sample Variance = .98908800

       1.0000          .0013          -.0645          -.0260          .0077
       -.0376          .0278          -.0273           .0599          .0717
```

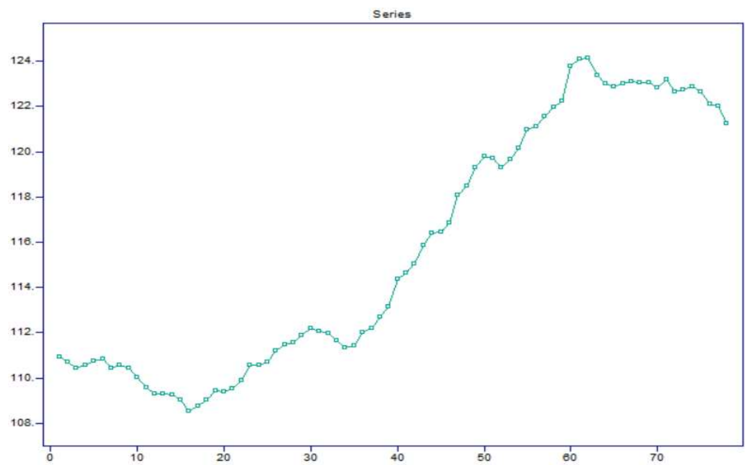
Series and its ACF

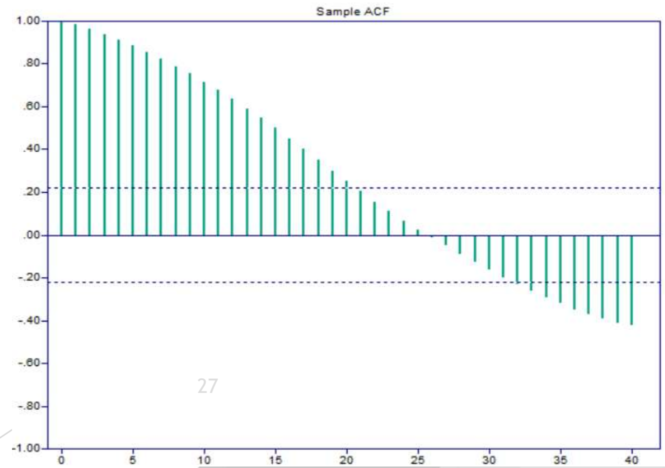
Seasonal series





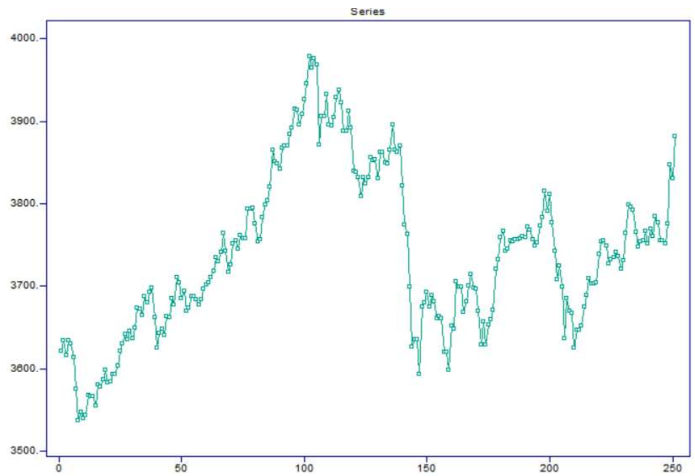
Trend series

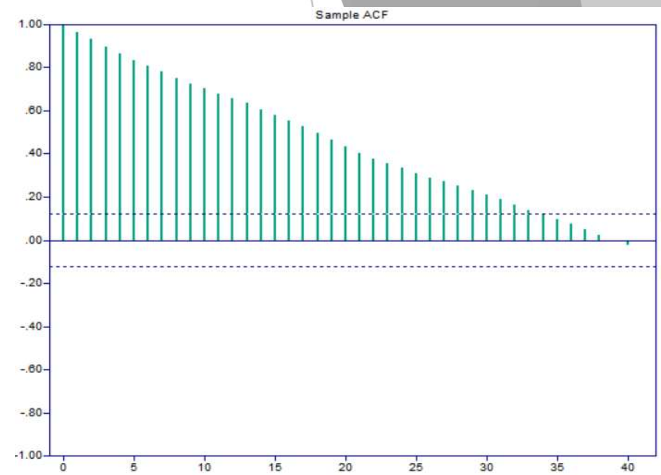




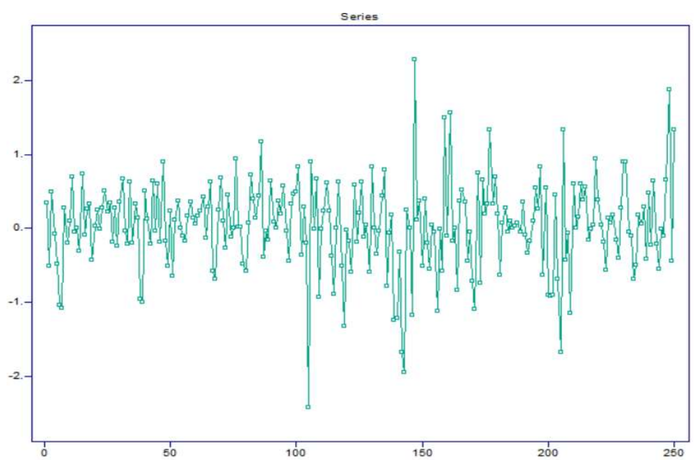
Series and its ACF

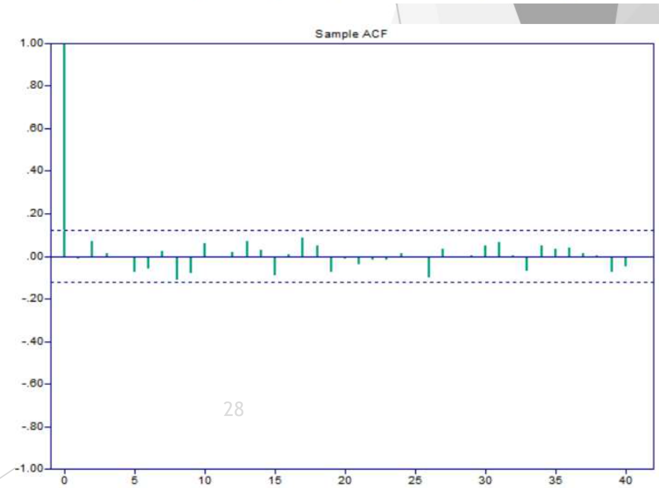
Dow Jones Price Index





Dow Jones Return





MATLAB: ACF and PACF

Plot ACF and PACF for

- MA(1), MA(2), AR(1), AR(2), ARMA(1,1)

Build in functions: **autocorr, parcorr**


THE
END