

Statistical Basics for Time Series Analysis

3

In Chapter 2 we discussed the topic of observing, smoothing, decomposing, seasonal adjusting, and forecasting time series data. We refer to the methodology used in that chapter as non-model based or data analytic methods. Most time series analysis techniques are based on statistical models. In this chapter we discuss statistical fundamentals that are basic to statistical analysis of time series, and this material will be used in future chapters as the basis for our time series modeling. In Section 3.1 we review some basic concepts from classical (not time series) statistical analysis.

3.1 STATISTICS BASICS

The following will be a quick review of basics for some readers and a brief introduction for others. We begin by discussing univariate statistical analysis. That is, we discuss the statistical concepts involved in the analysis of a single variable.

3.1.1 Univariate Data

Fundamental to statistical analysis is the *random variable*, which is defined in Definition 3.1.

Definition 3.1: *Random Variable:* A *random variable* is a variable whose values are the (numerical) outcomes of a random phenomenon associated with some probability distribution. There are two basic types of random variables:

- (a) discrete random variables: possible outcomes can be counted
- (b) continuous random variables: there is a *continuum* of possible outcomes

Key Point: The terms *random phenomenon* and *random process* will be used interchangeably from this point forward

3.1.1.1 Examples of Random Processes, Random Variables, and Probability Distributions

- (a) *Random Process*: Tossing a coin twice.

Random Variable: The number of heads. (discrete)

Probability Distribution: Prob(0)=.25, Prob(1)=.5, and Prob(2)=.25¹

- (b) *Random Process*: Randomly selecting male students at a large university

Random variable: Student height in inches. (continuous: actual height is a number along a continuum)

Probability Distribution: The true probability distribution would be the heights of all male students at the university. Assuming that the data on all male students are unavailable, we could randomly select several male students and plot a histogram of heights. It would probably tend to be bell-shaped with most heights between 64 inches and 76 inches, a few taller than 76 inches and a few shorter than 64 inches.

- (c) *Random Process*: Customers entering a store during a given day.

Random Variable: The number of customers who enter the store (discrete)

Probability Distribution: We would not know the true probability distribution, but it could be visually estimated using a histogram of daily number of customers over a period of several days.

Key Point: Histograms can be used to “visually estimate” a probability distribution.

The underlying probability distribution may be theoretical (such as in the case of the coin tossing example), or it may be the collection of all “things” about which we want to know information (such as the student height example). We often want to know characteristics about the underlying population or probability distribution. This information is often summarized using parameters that are estimated from observed data.

Definition 3.2: *Parameters* are numerical characteristics or descriptives of a population or probability distribution.

Definition 3.3: The *expected value* is a theoretical (population) average.² Parameters of interest related to the random variable X include the following, which we state in terms of expected values:

- (i) Mean, μ_x : Measures the “center” of the distribution and is denoted $\mu_x = E[X]$.
- (ii) Variance, σ_x^2 : A measure of how “spread out” the values of the underlying distribution are. The variance measures the average squared amount by which observations tend to differ from the mean. Variance is denoted $\sigma_x^2 = E[(X - \mu_x)^2]$.
- (iii) Standard Deviation, σ_x : We often use σ_x as a measure of spread because it can be interpreted as the typical amount by which values in the distribution differ from the mean (in the units of the problem). The standard deviation is simply the square root of the variance.

Random variables are typically denoted by capital letters, such as X, Y , or Z . A particular occurrence of the random variable produces a number. This leads to the following definition.

¹ The four possible, equally likely outcomes of two tosses are HH, HT, TH, TT.

² The calculation of expected values for continuous random variables involves the use of calculus.



Definition 3.4: *Observed Values* are specific outcomes of the underlying random phenomenon (process).

Observed values of random variables are numbers, and they are denoted by lower case letters, for example, x . Suppose our random variable, X , is the number of heads after two tosses of a coin. If after tossing the coin twice, the outcome “HT” is observed, then the particular outcome is $x = 1$. The classical study of random variables makes extensive use of collections of these observed values called a *sample*. A particularly useful (and often assumed) type of sample is known as a *random sample* and is defined next.

Definition 3.5: A *random sample*, say of size n , is the result of n *independent* instances of the underlying random phenomenon (process).

Random samples are often denoted, $x_j, j = 1, \dots, n$. They provide the statistical analyst with n independent (unrelated) observations of the random variable. Consequently, a random sample provides information about the underlying probability distribution of the phenomenon of interest.

Definition 3.6: A *statistic* is an estimate of a population parameter based on a sample.

Given the values, $x_i, i = 1, \dots, n$, in a random sample, the following are useful statistics:

- (i) Sample Mean: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. The symbol \bar{x} is read as “x-bar”, and it is simply the average of the observed data values.
- (ii) Sample Variance: $s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.³ The sample variance is approximately the average of the squared differences from the sample mean. This interpretation leaves most people “cold” regarding what is being measured, primarily because it is in squared units. For this reason we often use the standard deviation, the square root of the variance, for a more interpretable measure.
- (iii) Sample Standard Deviation: $s_x = \sqrt{s_x^2}$. The standard deviation is in the units of the observed data (not in squared units) and can be thought of as the typical amount by which data values differ from the sample mean.



QR 3.1
Unbiasedness

Key Points

1. Random variables are variables whose (numerical) values are determined by a random process.
2. Probability distributions are characterized by parameters such as the mean (μ_x), variance (σ_x^2), and standard deviation (σ_x).
3. We use random samples to obtain sample-based estimates of parameters
 - a single observation of a random variable gives limited information about the true mean and no information about the variance or standard deviation.

³ It is common to use the denominator $n-1$ which produces an unbiased estimator of σ_x^2 . Some practitioners use the denominator n , which produces the maximum likelihood estimator. These concepts are discussed in standard statistics textbooks, but will not be reviewed here.

Theorem 3.1: Central Limit Theorem Let $X_i, i = 1, \dots, n$ denote a random sample from a population with mean μ and variance σ^2 . Then for “large n ” the random variable

$$Z = \frac{\bar{X} - \mu_x}{\sigma_x / \sqrt{n}} \quad (3.1)$$

is approximately distributed as a Normal random variable with mean zero and variance one (denoted $N(0,1)$) even if the random variables, X_i , are not normally distributed.⁴

Definition 3.7: The standard error of \bar{X} (which is the standard deviation of the population of \bar{X} s) is given by

$$SE(\bar{X}) = \sigma_x / \sqrt{n}. \quad (3.2)$$

An important interpretation of Definition 3.7 is that σ_x / \sqrt{n} is the typical amount by which \bar{X} tends to differ from the true mean. Consequently, as n increases, \bar{X} tends to get closer and closer to the mean μ_x . We say that \bar{X} “converges to μ_x ” as n increases.

“Large” often simply means $n > 30$. Because we usually do not know the population variance, we often calculate the statistic

$$\tilde{Z} = \frac{\bar{X} - \mu_x}{s_x / \sqrt{n}}, \quad (3.3)$$



QR 3.2
R Simulation of CLT

which has an approximate normal distribution for large n (see Moore, McCabe, and Craig, 2017).⁵

Key Points

1. The standard error of \bar{X} (σ_x / \sqrt{n}) decreases as n increases, and \bar{X} converges to μ_x . That is, \bar{X} taken from a large sample should be a good estimate of μ and is probably a better estimate than \bar{X} calculated from a sample with smaller sample size.
2. According to the Central Limit Theorem, for “large n ” (typically greater than 30) the random variable

$$Z = \frac{\bar{X} - \mu_x}{\sigma_x / \sqrt{n}}$$

⁴ We are assuming that readers are familiar with the Normal (Gaussian) distribution that is characterized by the classic bell-shaped curve. If not, any introductory statistics text will provide a discussion of this distribution.

⁵ If the X_i s have a normal distribution, then the statistic in (3.3) has a Student’s t distribution. For large n the t distribution approaches the normal distribution, so we denote the statistic with a “ \tilde{Z} ” (suggesting that it is approximately $N(0,1)$). Statistical practitioners use this “normal approximation” even when the X_i s are not normally distributed.

has an approximate $N(0,1)$ distribution even if the random variables, X_i , are not normally distributed.

3. The random variable

$$\tilde{Z} = \frac{\bar{X} - \mu_x}{s_x / \sqrt{n}}$$

is approximately $N(0,1)$ if n is large.

Example 3.1 Consider the data in Table 3.1, which represent a sample of 30 students at a large university who were randomly assigned to an Introductory Statistics course. Table 3.1 shows the scores on a ten-point quiz given early in the semester.

The true mean and other parameters would be based on the population of students at the university who took the Introductory Statistics course. The histogram of the scores is shown in Figure 3.1 where it is seen that no students scored below five or greater than ten (a perfect score). The histogram is approximately bell-shaped, and thus we can conclude that the distribution is approximately normal. Also, from the histogram it is reasonable to guess that the average score is about 7.5. To calculate the sample average, we find $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{30}(222.6) = 7.42$. Letting \mathbf{x} denote the R vector containing the quiz scores, we can find the sample average using the R command **mean(x)** which yields 7.42.

TABLE 3.1 Quiz Scores

STUDENT	QUIZ	STUDENT	QUIZ	STUDENT	QUIZ
1	7.4	11	7.6	21	8.0
2	8.4	12	8.8	22	9.0
3	8.8	13	6.1	23	8.9
4	6.4	14	7.2	24	7.5
5	10.0	15	6.6	25	5.5
6	5.5	16	7.0	26	8.5
7	7.3	17	5.3	27	7.4
8	5.9	18	7.9	28	6.3
9	7.1	19	8.1	29	7.7
10	7.9	20	7.6	30	6.9
			Sum		222.6

The mean and standard deviation could be calculated by hand, but for datasets of this magnitude we would again use R, and the command **sd(x)** gives the result 1.147531. The interpretation of the standard deviation is that the values in the list tend to differ from the average ($\bar{x} = 7.42$) by about 1.15.

For a bell-shaped (normal) distribution, about 95% of the observations fall within two standard deviations of the mean and about 68% fall within one standard deviation of the mean. For more details, see any introductory text in statistics. Because the histogram for this example is approximately bell-shaped, we will assume that the distribution is approximately normal and thus that about 95% of the observations should be between $7.4 - 2(1.15) = 5.1$ and $7.4 + 2(1.15) = 9.7$. For the dataset in Table 3.1, 29 out of 30 or 96.7% fall within this range.



QR 3.3 Example 3.1

FIGURE 3.1 Quiz scores for a sample of 30 students selected randomly from the population of all students enrolled in Introductory Statistics.

3.1.2 Multivariate Data

When two or more random variables are calculated on members of a random sample, the resulting data are called *multivariate data*. The special case of two variables is referred to as *bivariate* data.

Table 3.2 is a generic illustration of a typical tabular presentation of a bivariate random sample. We see that Subject 1 has the value x_1 for the random variable X and the value y_1 for the random variable Y . The two observations for Subject 1 are considered as an ordered pair, (x_1, y_1) and the bivariate random sample is represented by $(x_i, y_i), i = 1 \dots, n$.

TABLE 3.2 Random Sample of Ordered Pairs

SUBJECT	X	Y
1	x_1	y_1
2	x_2	y_2
\vdots	\vdots	\vdots
n	x_n	y_n

Example 3.1 (revisited): Bivariate Data Consider the data in Table 3.1 which represent a sample of 30 students at a large university who were randomly assigned to an Introductory Statistics course. In Table 3.3 we again show the quiz scores from Table 3.1 along with the final exam score for each student.

TABLE 3.3 Quiz and Final Exam Scores

STUDENT	QUIZ	FINAL	STUDENT	QUIZ	FINAL	STUDENT	QUIZ	FINAL
1	7.4	79.8	11	7.6	80.7	21	8.0	84.2
2	8.4	82.0	12	8.8	94.5	22	9.0	87.8
3	8.8	76.1	13	6.1	50.1	23	8.9	94.1
4	6.4	62.7	14	7.2	68.3	24	7.5	78.2
5	10.0	98.2	15	6.6	64.4	25	5.5	62.4
6	5.5	43.0	16	7.0	67.2	26	8.5	85.1
7	7.3	76.5	17	5.3	53.9	27	7.4	77.8
8	5.9	61.4	18	7.9	78.8	28	6.3	67.6
9	7.1	78.5	19	8.1	85.7	29	7.7	70.2
10	7.9	88.7	20	7.6	81.7	30	6.9	73.6
						Sum	222.3	2253.2

The means and standard deviations are given below (again obtained using R functions **mean** and **sd**). The quiz scores are in R vector **x** and the final exam scores are in **y**. We have the following results:

```
mean(x) [1] 7.42
sd(x) [1] 1.15
mean(y) [1] 75.11
sd(y) [1] 13.15
```

The histograms for these two variables are shown in Figure 3.2. There it can be seen that the histograms for both variables are approximately bell shaped with the Final Exam scores being slightly skewed to the left. By examination of the histogram in Figure 3.2(b) we see that the mean of the final exam scores appears to be about 75 which is consistent with the results above. The majority of the scores are between 60 and 90 with a few above 90 and a few below 60.

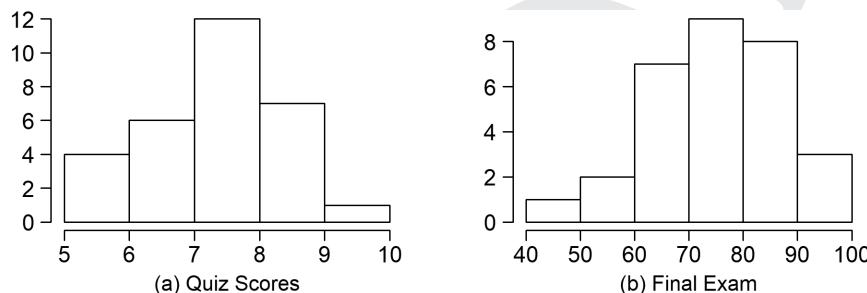


FIGURE 3.2 (a) Quiz and (b) Final Exam scores for a random sample of 30 students enrolled in Introductory Statistics.

Key Point: While the histograms and statistics derived from the sample discussed to this point are interesting and informative, they do not provide information about the relationship between the two random variables.

3.1.2.1 Measuring Relationships between Two Random Variables in a Bivariate Random Sample

The Key Point above emphasizes the fact that histograms, means, and standard deviations do not provide information about how two variables are related to each other. As an example, an instructor would probably like to know whether students who did well on the quiz also tended to do well on the final exam, and vice versa. She might also want to know if some students who did poorly on the quiz dramatically raised their grade on the final exam. The histograms and sample statistics shown above do not answer these questions.

What is needed are measures of the relationship between the two variables. The most common population parameters used to measure such relationships are the covariance, γ_{xy} and the correlation, ρ_{xy} .

Definition 3.8: The covariance, γ_{xy} , and correlation, ρ_{xy} , are parameters that measure the (linear⁶) relationship between two variables. These are defined as follows:

$$(a) \gamma_{xy} = E[(X - \mu_x)(Y - \mu_y)]$$

⁶ We will discuss the manner in which the correlation measures linear association as we discuss the sample correlation coefficient.

$$(b) \rho_{xy} = \frac{\gamma_{xy}}{\sigma_x \sigma_y}$$

Statistically, the covariance is the expected value (or theoretical average) of the cross-product $(X - \mu_x)(Y - \mu_y)$. It is a measure of how two variables “move together”. To facilitate the interpretation, we usually use the correlation, which is a “standardized” version of the covariance that has the property $-1 \leq \rho_{xy} \leq 1$ for any two random variables X and Y .

Notes

$$(a) \gamma_{xx} = E[(X - \mu_x)(X - \mu_x)] = E[(X - \mu_x)^2] = \sigma_x^2$$

$$(b) \rho_{xx} = \frac{\sigma_x^2}{\sigma_x \sigma_x} = 1$$

3.1.2.2 Assessing Association from a Bivariate Random Sample

Definition 3.8 and the notes that follow discuss (unknown) parameters. In a practical setting we estimate the association using the sample correlation coefficient.

Sample correlation coefficient: The sample statistic used to estimate ρ_{xy} based on a random sample of bivariate data, $(x_i, y_i), i = 1, \dots, n$, is the sample correlation coefficient (sometimes called the Pearson product-moment correlation coefficient). Given a bivariate dataset, the sample correlation coefficient, r_{xy} , is defined by

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}. \quad (3.4)$$

Values of r_{xy} share the property with the population correlation coefficient, ρ_{xy} , that $-1 \leq r_{xy} \leq 1$.

Scatterplots: Another tool for evaluating the association between two variables is the scatterplot. A scatterplot is a graph in which the ordered pairs, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ are plotted as n points on an X - Y plane. Figure 3.3 shows scatterplots for several pairs of variables. Figure 3.3(a) shows a strong positive linear relationship between X and Y . Specifically, as X increases, Y tends to increase and vice versa. Referring to the definition of r_{xy} , note that r_{xy} is a function of the products of the differences $x_i - \bar{x}$ and $y_i - \bar{y}$. In Figure 3.3(a) we see that nearly all of the points are in the upper right or lower left quadrants. Points in the upper right quadrant are such that $x_i > \bar{x}$ and $y_i > \bar{y}$, so $(x_i - \bar{x})$ and $(y_i - \bar{y})$ are both positive, and consequently, so is their product. Likewise, for points in the lower left quadrant, it follows that $x_i < \bar{x}$ and $y_i < \bar{y}$, so $(x_i - \bar{x})$ and $(y_i - \bar{y})$ are both negative, and again their product, $(x_i - \bar{x})(y_i - \bar{y})$, is positive. Therefore, since most of the products will be positive (and the ones that are negative will have relatively small magnitude), the overall sum $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ and thus r_{xy} will tend to be positive since the denominator is always positive. Based on the data in Figure 3.3(a), if you are told that an x - y pair has $x = 2$, then you would be fairly confident that the y value is about $y = 1.5$. In general, you could predict another value for y if you know the corresponding value for x . However, note that we only know that this relationship exists when x is between -3 and 3 . This leads to the following definition.



QR 3.4 Visualizing Autocovariance

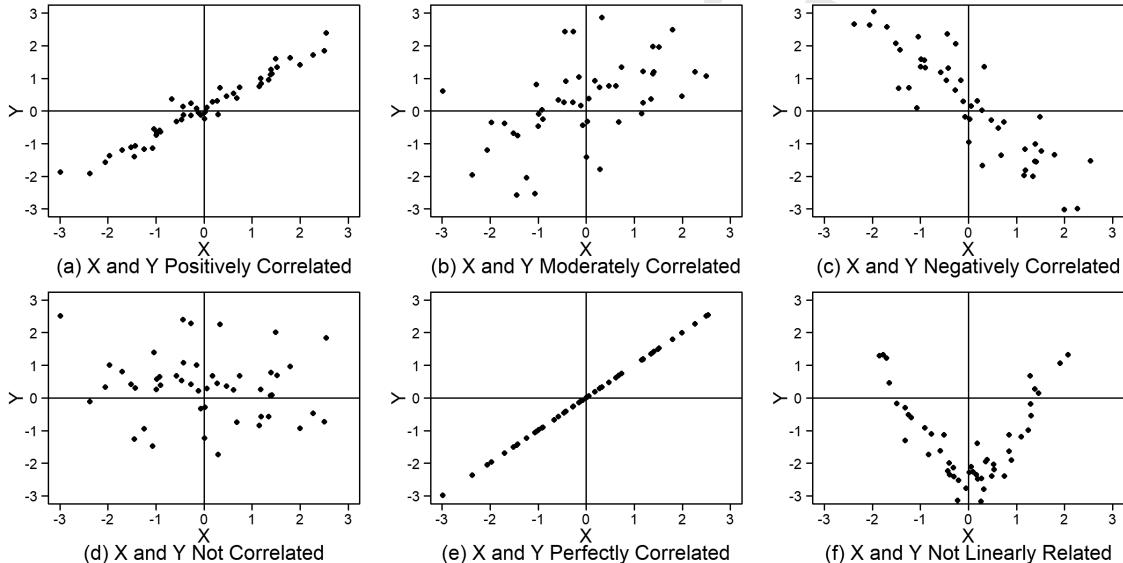


FIGURE 3.3 (a)–(f) are scatterplots with a variety of correlation structures.

Definition 3.9: Prediction of a y value based on an x value that is outside the range of observed x values is an *extrapolation*.

For example, to predict y if $x = 5$, which is outside the range of observed x values ($-3, 3$), is an example of extrapolation. Every introductory statistics book known to humankind discourages extrapolation because the fact that a linear relationship exists between X and Y in the range of the variables in the sample (in this case between -3 and 3) does not guarantee that the same relationship extends beyond the range of the original data.

Similar to Figure 3.3(a), the data in Figure 3.3(b) suggest a positive, although now moderate, correlation between X and Y . However, unlike Figure 3.3(a), for a given value of x , there is quite a range of plausible y values. Suppose you have sampled a new observation, $x = 2$, of the random variable X . Based on the plot, you would not be as confident that the corresponding y value is close to 1.8, although you would be justified in predicting that it would be greater than zero. That is, “knowing x ” helps in predicting y (but not as much as for the data in Figure 3.3(a)).

In contrast to Figures 3.3(a) and (b), Figure 3.3(c) shows an example of a fairly strong negative correlation, where large values of X are associated with small values of Y , and vice versa. We see that nearly all points are now in the upper left or lower right quadrants. Points in the upper left quadrant are such that $x_i < \bar{x}$ and $y_i > \bar{y}$, so $(x_i - \bar{x})(y_i - \bar{y})$ is negative, $(y_i - \bar{y})$ is positive, and thus the product $(x_i - \bar{x})(y_i - \bar{y})$ is negative. Likewise, for points in the lower right quadrant, it follows that that $x_i > \bar{x}$ and $y_i < \bar{y}$, so $(x_i - \bar{x})$ is positive, $(y_i - \bar{y})$ is negative, and again the product is negative. Therefore, because most of the products will be negative (and the ones that are positive will have relatively small magnitude), we expect the overall sum $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, and thus r_{xy} to be negative this time. This negative relationship suggests that if

you have a new randomly selected ordered pair with $x = 2$, then you can be relatively confident that the associated y value will be negative and likely less than -1 .

Figure 3.3(d) shows an example in which there seems to be no detectable correlation between X and Y . “Knowing x ” does not help in predicting y . Think about what this means in terms of $(x_i - \bar{x})$, $(y_i - \bar{y})$, $(x_i - \bar{x})(y_i - \bar{y})$, $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, and thus r_{xy} . See Problem 3.9.

Figure 3.3(e) shows an example of a “perfect” positive linear relationship. The points all lie on a line whose equation you could derive using elementary algebra. Given the equation, if you “know x ,” then you can calculate y (again, within the range of the observed x values).

Finally, Figure 3.3(f) shows another situation in which there is a relationship between X and Y . Specifically, y values tend to be high when x values are either low or high. On the other hand, for x values toward the middle of the range of x (that is, near zero), the y values tend to be low. Consequently, there is a relationship, but it is not linear. Think again about what this means for $(x_i - \bar{x})$, $(y_i - \bar{y})$, $(x_i - \bar{x})(y_i - \bar{y})$, $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, and thus r_{xy} . See Problem 3.10.

Calculating Correlation using R: Suppose the data for x are contained in the R vector \mathbf{x} and that y values are in \mathbf{y} . Then the correlation coefficient is calculated using the R command `cor(x, y)`. The sample correlation was calculated for the data in each plot in Figure 3.3, and the following values were obtained: (a) $r_{xy} = .976$, (b) $r_{xy} = .518$, (c) $r_{xy} = -.883$, (d) $r_{xy} = -.158$, (e) $r_{xy} = 1$, (f) $r_{xy} = -.211$.

Key Points

1. The data in Figure 3.3(d) were generated from a distribution for which the true correlation, $\rho_{xy} = 0$. The sample correlation coefficient r_{xy} is estimating this value and is doing a good job as it is small ($r_{xy} = -.158$).
2. The sample correlation for the data in Figure 3.3(f) is near zero. This emphasizes the fact that r_{xy} is measuring *linear* association. Again, for these data there is a relationship, but it is not linear and is not “detected” by the sample correlation coefficient.

Example 3.1 (revisited) Figure 3.4 shows a scatterplot of the quiz-final exam data in Example 3.1 along with vertical and horizontal lines specifying $\bar{x} = 7.42$ and $\bar{y} = 75.11$, respectively. Given the example scatterplots in Figure 3.3 it appears that the scatterplot is a “positively correlated” version of Figure 3.3(c), and, in fact, $r_{xy} \approx .9$. If the quiz scores are in R vector \mathbf{x} and the final exam scores are in \mathbf{y} , then the correlation coefficient can be obtained from issuing the command

```
cor(x, y)
[1] 0.9006451
```

As would be expected, there is convincing statistical and visual evidence of a strong positive linear relationship between quiz scores and final exam scores. Knowing the early quiz score helps predict performance on the final exam. From the scatterplot we see that there does not seem to be much evidence of students greatly improving on the final exam scores nor did many students underperform on the final.

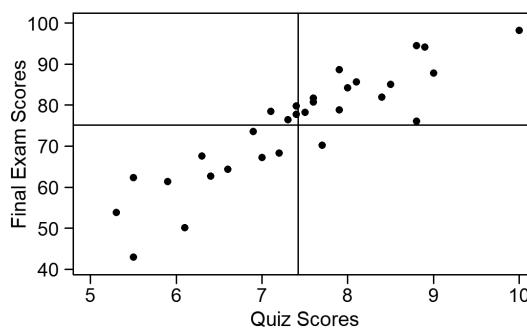


FIGURE 3.4 Scatterplot of quiz scores and final exam scores for the data in Table 3.3.

3.1.3 Independent vs Dependent Data

The main difference between a random sample and time series data is that a random sample is composed of independent observations while time series data have a dependence structure. To illustrate, we provide the following example.

Example 3.2

(a) **Independent Sum Data:** Consider the experiment of rolling two dice and finding the sum of the rolls. In this case, the random variable is:

X = the sum of the dots showing after rolling two dice.

Suppose the experiment of rolling two dice and finding the sum was performed and the result of the rolls was $\square\square$ and $\square\square$, so that the sum of the rolls was $x = 9$. If, after obtaining this single observation, you had to “guess” what the mean of the distribution is, your best guess would be 9. While a single observation of a random variable provides some information about the mean or center of the distribution, it makes sense that the sample mean of n unrelated observations is a better estimate of the mean. To illustrate further, the random process of rolling two dice and finding the sum of the rolls was repeated 30 times. The results are shown in Table 3.4.

To start with, we consider the first 10 outcomes (the first column). For the first data column, $\bar{x} = 8.2$. It makes sense that 8.2 is a reasonable estimate of the (to this point unknown) mean, and is likely more accurate than the estimate based simply on the first observation. Continuing the procedure, we expanded the sample until we obtained 30 repetitions; that is, we obtain a sample of size $n = 30$. Letting \mathbf{x} denote the R vector containing the 30 “sum of the rolls” dataset, we calculated \bar{x} using the base R command **mean(x)**. In this case $\bar{x} = 7.17$. The actual mean is $\mu = 7$ (see Appendix 3A), so the additional sample size improved the estimate. This will tend to be the case in general (although it is not a certainty).⁷

⁷ For example, if our original observation was $x = 7$, then there is no sample (no matter how large the sample size) for which the sample average is a better estimate of the true mean ($\mu_x = 7$).

TABLE 3.4 Sample Results for Dice-Rolling Process

REPETITION	SUM OF ROLLS	REPETITION	SUM OF ROLLS	REPETITION	SUM OF ROLLS
1	9	11	4	21	4
2	7	12	3	22	8
3	8	13	9	23	9
4	10	14	6	24	7
5	5	15	8	25	6
6	7	16	6	26	3
7	9	17	6	27	8
8	10	18	7	28	7
9	10	19	11	29	8
10	7	20	8	30	5

Again, letting \mathbf{x} denote the R vector containing the data in Table 3.4, the sample variance, s_x^2 , can be obtained using R command `var(x)`. For these data we obtained $s_x^2 = 4.35$. In our case, the sample standard deviation is $s_x = 2.09$ which has the interpretation that individual sums of the rolls of two dice tend to differ from the sample mean ($\bar{x} = 7.17$) by about 2. The sample standard deviation, s_x , can be obtained using R command `sd(x)`, or by using `var(x) = sqrt(var(x))`. In Appendix 3A we show that for the dice rolling distribution, $\mu_x = 7$, $\sigma_x^2 = 5.833$, and $\sigma_x = 2.415$. Note that, although a single observation gives us limited information about the true mean, μ_x , a single observation gives *no information* about the variability. This is because with a single observation, x , then $\bar{x} = x$ and $n = 1$ but the sample

$$\text{standard deviation } s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{0}{0}, \text{ which is undefined.}^8$$

Before discussing the statistical basics of time series analysis, we show Figure 3.5(a), which is a plot of the 30 outcomes in Table 3.4, indexed by time (that is, ordered by repetition number). The experiments were run sequentially in time, so that the first run of the experiment produced a sum of 9, the second time a sum of 7, and so forth. That is, $x_1 = 9$, $x_2 = 7$, $x_3 = 8$, etc., and we plot $x_i, i = 1, \dots, 30$. In this plot we see the randomness associated with a random sample. If we connect the points, as we have done in Figure 3.5(b), we get a better look at how the sums “progress through time”. We note that the randomness expresses itself in that the t th sum appears to have no predictive ability concerning the value of the $(t+1)^{\text{st}}$ sum.

As an illustration of this randomness, Figure 3.5(c) shows the mean line at 7. There are 14 instances in which the sum is greater than 7. In these cases, the next sum had the following characteristics:

- (i) in 5 cases, the next sum is also above 7
- (ii) in 4 cases, the next sum is equal to 7
- (iii) in 5 cases, the next sum is less than 7.

Consequently, if the sum at “time” t is greater than 7, there is no information in the dataset regarding whether the sum at $t+1$ is above, equal to, or less than 7. This is the essence of a random sample in a “time series” framework. Data such as that shown in Figure 3.5 are called *white noise*, for which a precise definition will be given in Definition 3.14.

⁸ It also makes intuitive sense that if only one observation is available, then you know nothing about how much different observations tend to differ from each other.

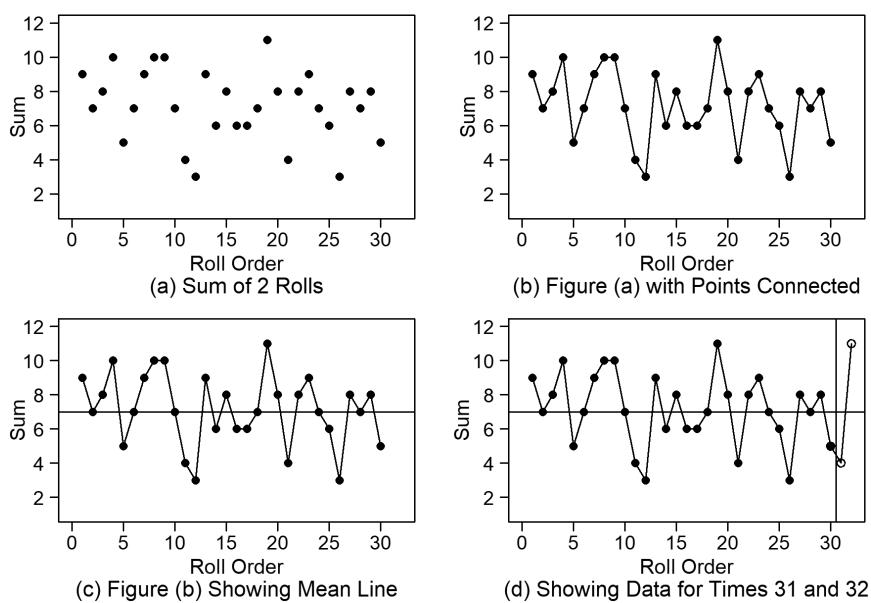


FIGURE 3.5 (a) Plot of the sum of two dice data in Table 3.4, (b) data in (a) with points connected, (c) plot (b) showing mean line at $\mu = 7$, and (d) plot showing data values for 31st and 32nd instances of the random process.

Figure 3.5(d) shows the original sum data for $t = 1, \dots, 30$ along with outcomes at $t = 31$ and $t = 32$. Note that the last observed outcome (at $t = 30$) is 5. However, this fact provides no information about the next two observations, which were 4 and 11, respectively. In fact, if you knew that the mean of the process is $\mu = 7$ (or if you observed that the average of the first 30 observations is $\bar{x} = 7.17$), then your best guess for the outcomes at times $t = 31$ and $t = 32$ would be 7 in both cases, regardless of the value at time $t = 30$.

We are now in position to make the following Key Points.

Key Points:

1. Random samples arranged in sequence produce a *white noise* time series (which will be formally defined in Definition 3.14)
2. The independent nature of a random sample sequenced by time leads to the fact that the value of a random variable at time t gives no information concerning the value at time $t + k$ for $k \geq 1$.

This leads to the following Key Points, which we have separated from the Key Points above for emphasis.

Key Points

1. In classical statistics, we typically work with random samples in which the observations are independent, and we avoid cases in which a sample is not truly random.
2. In time series analysis, we *focus* on data that are *dependent* (*that is, related*).
 - this is not a “bad thing” – in fact, it is quite the opposite!
 - we will use the dependence structure to our advantage for forecasting and other analyses.

(b) Dependent Sum Data: To emphasize Key Point Number 2 regarding the advantage of dependence, we simulated a “coin tossing experiment” using a random process producing the random variable, Y , which does not protect the independent nature of the data that occurs from simply rolling two dice and finding the sum.⁹ We denote this random process the “dependent sum” process, and from Footnote 9, we see that $\mu_y = 7$. We will refer to the random process that produced the data in Table 3.4 as the “original (independent) sum” process. The “dependent sum” data are shown in Figure 3.6(a) and again in Figure 3.6(b) with the points connected by straight lines to better show the pattern in time along with a horizontal line showing the mean line $\mu_y = 7$. There is something very different between the data in Figure 3.5 and Figure 3.6. We see in Figure 3.6 that “what happens at time $t + 1$ is related to what happened at time t ”. For example, between time $t = 10$ and time $t = 22$, all dependent sum outcomes are greater than the mean $\mu_y = 7$. Similarly, the four outcomes from $t = 24$ through $t = 27$ are all below the mean. The outcome at time $t = 30$ is 9. If you were asked to predict the 31st and 32nd outcomes of this process, because of the correlation structure of the data, you would probably guess the values to be somewhere in the neighborhood of 9 (and likely above the mean). We repeated the dependent dice tossing experiment two more times. Figure 3.6(c) shows that the outcomes at times $t = 31$ and $t = 32$ are both 10, which are consistent with our conjecture. That is, the dependence structure of the data provided information about the expected future outcomes.

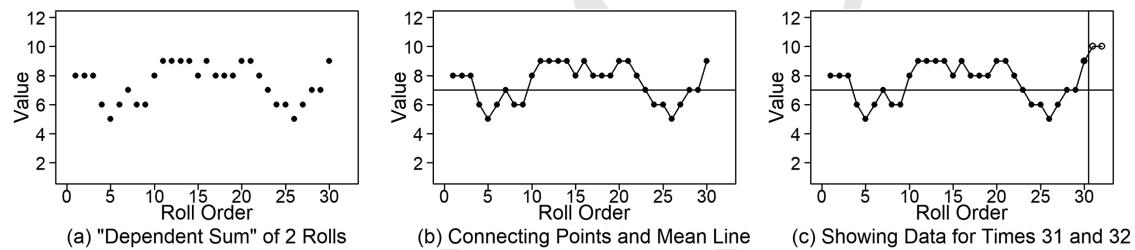


FIGURE 3.6 (a) Plot of the sum of “dependent sum” data, (b) data in (a) with points connected and showing mean line at $\mu = 7$, and (c) plot showing data values for 31st and 32nd instances of the random process.

(c) The “Oscillating” Process: As another example, consider the data in Figure 3.7(a). We denote the data with the notation, w_t . Figure 3.7(a) shows the data, w_t , $t = 1, \dots, 30$ while Figure 3.7(b) connects the points and plots the mean line, $\mu_w = 7$.¹⁰ The sample mean of the data in Figure 3.7(a) is $\bar{w} = 6.87$. Figure 3.7(b) shows the data connected by straight lines along with the mean line, which better illustrates the general oscillating behavior of the data, typically back and forth across the mean line. We note that the value of the random variable at $t = 30$ is 10. Consequently, the correlation structure of the data suggests that the 31st observation will probably be below 7 and the 32nd above 7. In Figure 3.7(c) we see that indeed these predictions are accurate. Again, the dependence structure aided us in predicting future outcomes.

9 To introduce dependence, the “dependent sum” process (denoted here by the random variable Y_t) was obtained as follows: at time $t = 1$, Y_1 sum of the rolls of the two dice as before. So $E(Y_1) = 7$. (See Appendix 3A). For $t = 2$, we roll two dice again and obtain a “new sum” random variable, Z_2 . Again $E(Z_2) = 7$. Then, we define $Y_2 = .6Y_1 + .4Z_2$, and using the properties of expectation discussed in Footnote 15, it follows that $E(Y_2) = .6(7) + .4(7) = 7$. Continuing this process we see that at time t for $t = 2, \dots, 20$, $Y_t = .4Y_{t-1} + .4Z_t$, and using the logic above, we see that $E(Y_{t-1}) = 7$ and $E(Z_t) = 7$. It follows that for $t = 2, \dots, 20$, we have $E(Y_t) = .6E(Y_{t-1}) + .4E(Z_t) = .6(7) + .4(7) = 7$. Consequently, for $t = 1, \dots, 20$, $E(Y_t) = \mu_y = 7$. Note that we round the value of y_t to the nearest integer to be consistent with the original sum-of-the-rolls variable, x_t .

10 The formula for W_t is $W_t = 7 - .97(W_{t-1} - 7) + a_t$, where the a_t s are random normal variables with mean 0 and variance 1. The population mean of the random variable W_t is 7. This will be discussed in Chapter 5. Please be patient! Again, we round the value of W_t to the nearest integer to be consistent with the original sum-of-the-rolls variable, x_t .

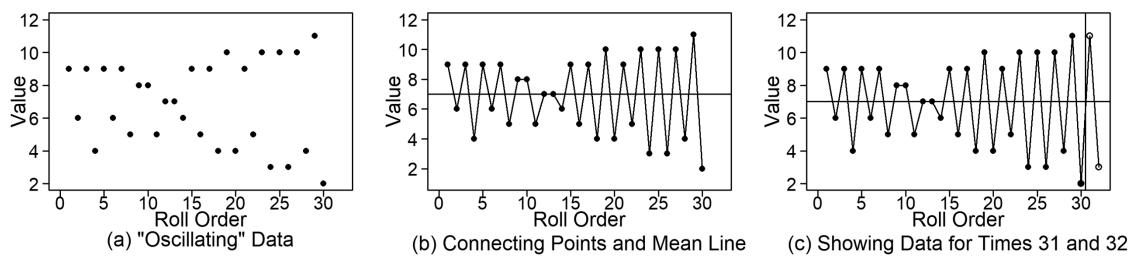


FIGURE 3.7 (a) Plot of the oscillating data, (b) data in (a) with points connected and a horizontal line at 7, and (c) plot showing data values for $t = 31$ and $t = 32$.

Key Points

1. For the “dependent sum” and “oscillating” data, an understanding of the dependence between consecutive runs of the “experiment” is helpful in predicting outcomes at $t = 31$ and $t = 32$.
2. Because of the independence (that is, lack of dependence structure) in the “original sum” data in Figure 3.5, knowing that the sum at time $t = 30$ is 5 provides no assistance in predicting the sums at $t = 31$ and $t = 32$.

3.2 TIME SERIES AND REALIZATIONS

Using the preceding review as a lead-in, we are now in a position to discuss time series as a collection of random variables. Informally, we can think of a time series as a collection of observations made sequentially in time. The datasets in Figures 1.1, 1.3, and so forth, are examples of time series data using this informal definition. However, to be more precise, we need to define time series in terms of random variables. We have the following definition.

Definition 3.10: A *time series* is a collection of random variables, $\{X_t\}$, indexed on time.¹¹

A collection of random variables is generically referred to as a *stochastic process*. When the index set is time, we refer to the stochastic process as a *time series*. We will focus on time series for which the index, t , takes on integer values. Definition 3.10 says that each time, t , is associated with “its own” random variable, X_t . Recall that it is common to denote a random variable using a capital letter.

Some Time Series Parameters: Because each time, t , has “its own” random variable, X_t , these random variables have their own distribution and thus possess their own parameters. Among these are the following:

- (a) Random variable X_t has a mean, μ_{X_t} (just as the random variable X has mean μ_X).
- (b) Random variable X_t has a variance, $\sigma_{X_t}^2$ (just as the random variable X has variance σ_X^2).
- (c) There is a covariance, $\gamma_{X_{t_1} X_{t_2}}$, and correlation, $\rho_{X_{t_1} X_{t_2}}$, between random variables X_{t_1} and X_{t_2} . The covariance and correlation between two random variables in the same time series are called the *autocovariance* and *autocorrelation*, respectively. Again, these correspond to the covariance and correlation between random variables X and Y .

¹¹ We usually omit the curly brace and simply refer to the time series X_t when there is no confusion.

Notes

- (a) If $t_1 \neq t_2$, it may follow (but not necessarily) that $\mu_{X_{t_1}} \neq \mu_{X_{t_2}}$ or $\sigma_{X_{t_1}}^2 \neq \sigma_{X_{t_2}}^2$, or both.
- (b) If $t_1 = t_2$, then $\rho_{X_{t_1} X_{t_2}} = \rho_{X_{t_1} X_{t_1}} = 1$.

In Section 3.1, we defined a random variable to be a variable whose values are the numerical outcomes of a random phenomenon, and we defined specific outcomes to be observed values. This leads to the following definition.

Definition 3.11: A *realization* of the time series $\{X_t\}$, is a set of specific outcomes of the random phenomenon (process).

We typically denote a realization of n outcomes using the notation, $x_t, t = 1, \dots, n$. Earlier we referred to the data sets plotted in Figures 1.1 among others as a time series. These are actually realizations of the underlying time series. For example, for the original sum of two dice example, if you decided to repeat the experiment (that is, roll the dice 30 more times and find the sums), this would produce another realization from the underlying time series, but it would be almost a certainty that the new realization would differ from the one plotted in Figure 3.5(b). Figure 3.8(a) shows the realization in Figure 3.5(b) along with two other realizations from the same random process. Although different from each other, the realizations exhibit similar behavior.

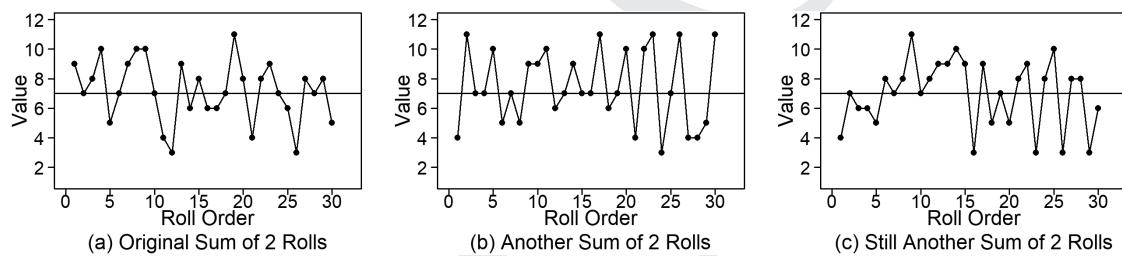


FIGURE 3.8 Three realizations from the random process of rolling the dice 30 times and finding the sums.

Definition 3.12: The collection of all possible realizations from a time series is called an *ensemble*.¹²

3.2.1 Multiple Realizations

Consider the case of a time series for which, as mentioned, each t “has its own” random variable, X_t . For example, X_3 has mean μ_{X_3} and standard deviation, σ_{X_3} . If we have a realization of length, say $n = 100$ (or $n = 1,000$), from the time series, we only have *one* observation, x_3 , on the random variable X_3 . Lacking other knowledge about the time series, our best estimate of is the single value of x_3 , and there is no information about σ_{X_3} . Multiple realizations would be needed to obtain reasonable estimates of the mean and standard deviation of X_3 .¹³

Suppose we have a time series, $\{X_t\}$, from which we have m (independent) realizations. Call these realizations $x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(m)}$, $t = 1, \dots, n$. Then, $x_1^{(i)}, i = 1, \dots, m$ are the observed values of a random sample from the random variable X_1 (the first observation of each realization). We can estimate the mean, μ_{X_1} ,

12 The ensemble is often a theoretical construct achievable only in the imagination.

13 Stationarity of a time series provides the ability to obtain information from a single realization. This will be discussed in Section 3.3.

variance, $\sigma_{X_{t_1}}^2$, and standard deviation, $\sigma_{X_{t_1}}$, along with the autocovariance, $\gamma_{X_{t_1}X_{t_2}}$, and autocorrelation, $\rho_{X_{t_1}X_{t_2}}$, using the following calculations:

- Sample Mean: $\bar{x}(t_1) = \frac{1}{m} \sum_{i=1}^m x_{t_1}^{(i)}$
- Sample Variance: $s_{X_{t_1}}^2 = \frac{1}{m-1} \sum_{i=1}^m (x_{t_1}^{(i)} - \bar{x}(t_1))^2$ and standard deviation: $s_{X_{t_1}} = \sqrt{s_{X_{t_1}}^2}$.
- Sample Autocovariance and Autocorrelation between Random Variables X_{t_1} and X_{t_2} :

$$\hat{\gamma}_{X_{t_1}X_{t_2}} = \frac{1}{m-1} \sum_{i=1}^m (x_{t_1}^{(i)} - \bar{x}(t_1))(x_{t_2}^{(i)} - \bar{x}(t_2))$$

$$r_{X_{t_1}X_{t_2}} = \frac{\sum_{i=1}^m (x_{t_1}^{(i)} - \bar{x}(t_1))(x_{t_2}^{(i)} - \bar{x}(t_2))}{(m-1)s_{X_{t_1}} s_{X_{t_2}}}$$
(3.5)



QR 3.5 Multiple Realizations

Note that in this case, we are assuming that we have multiple realizations $x_{t_1}^{(1)}, x_{t_1}^{(2)}, \dots, x_{t_1}^{(m)}$ and $x_{t_2}^{(1)}, x_{t_2}^{(2)}, \dots, x_{t_2}^{(m)}$ for each of the two random variables, X_{t_1} and X_{t_2} .

Example 3.3 Multiple Realizations

Figure 3.9(a) shows a realization, x_t , $t = 1, \dots, 20$ from a time series. For this realization, $x_1 = 17$, $x_2 = 19, \dots$, and $x_{20} = 60$. The realization has a somewhat wandering behavior, but the general tendency is to increase in time. Given the notation above, there are 20 random variables, X_t , $t = 1, \dots, 20$. The random variable at time t , that is, X_t , has a mean, μ_{X_t} . For example, X_2 is a random variable which has a mean, μ_{X_2} , and we have observed a single observation, 19, from its distribution. Similarly, the random variable X_{18} has a mean, $\mu_{X_{18}}$, and we have observed the single observation, $x_{18} = 47$.

Figure 3.9(b) shows a realization, y_t , $t = 1, \dots, 20$ from another time series, Y_t . This realization has characteristics that are similar to the one in Figure 3.9(a). The appearance again suggests that the time series involves a tendency for data values to increase in time. As in Figure 3.9(a), for a given value of t , the realization in Figure 3.9(b) gives us only one observation. For example, for the random variable X_5 we have the single observed value $x_5 = 23$ while X_{15} has the single observed value $x_{15} = 39$.

As noted in Section 3.1, a limited amount of information can be obtained from one observation from a random variable. In this example, we consider the information obtained by observing multiple realizations from each of these time series.

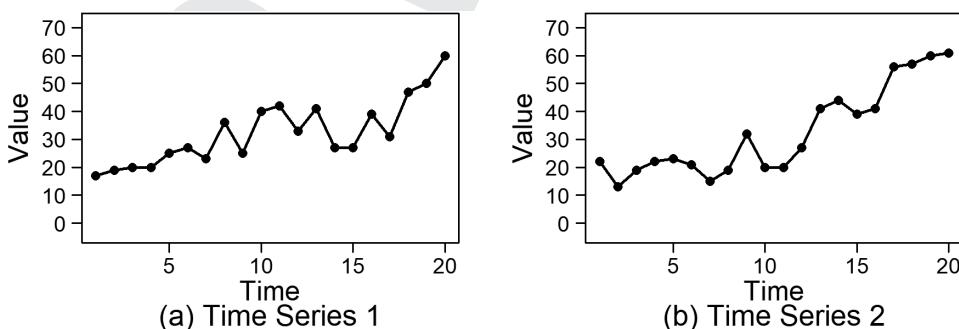


FIGURE 3.9 Realizations from two time series.

3.2.1.1 Time Series 1: X_t

The increase in time for the realization in Figure 3.9(a) suggests that the true means, μ_t , increase in time. To further examine this speculation about the increase, we randomly generated nine more realizations from the same random process (phenomenon) that generated the realization in Figure 3.9(a).¹⁴ In Figure 3.10(a) we show the realization in Figure 3.9(a), while Figure 3.10(b) shows this realization (in bold) along with nine new realizations from the same time series. In Figure 3.10(b) we see that indeed there is a tendency for realizations to increase in time (as was the case in the original realization). That is, the multiple realizations reinforce the conjecture that the true means increase over time. We estimate the population mean, μ_{X_t} and standard deviation, σ_{X_t} at time t , by finding the sample average ($\bar{x}(t)$) and sample standard deviation (s_{X_t}) of the 10 observations at time t . The sample data for times $t = 5, 10$, and 15 , highlighted in Figure 3.10(b), are shown in Table 3.5. This table and Figure 3.10(b) suggest that both the means and standard deviations of the random variables increase with time. To get a better understanding of the pattern of the means, we found the average \bar{x}_t for each t , where $t = 1, \dots, 20$. These sample averages are plotted in Figure 3.10(c) as open squares, and we see that they increase in a fairly linear manner. Also plotted are the true means, which follow a line with positive slope.

TABLE 3.5 Data from Multiple Realizations for $t = 5, 10$, and 15 Shown in Figure 3.10(b)

REALIZATION	X_5	X_{10}	X_{15}
1	25	40	27
2	21	27	48
3	25	35	49
4	21	34	33
5	21	31	51
6	23	27	39
7	20	26	48
8	25	34	42
9	23	37	46
10	22	47	31
\bar{x}_t	22.6	33.8	41.4
s_{X_t}	1.9	6.5	8.5

¹⁴ We generated the realizations from a time series model. A typical analysis would proceed knowing only the data, and we would attempt to understand the underlying “true model”. We will “disclose” the true model after we have discussed the analysis based on the data alone.

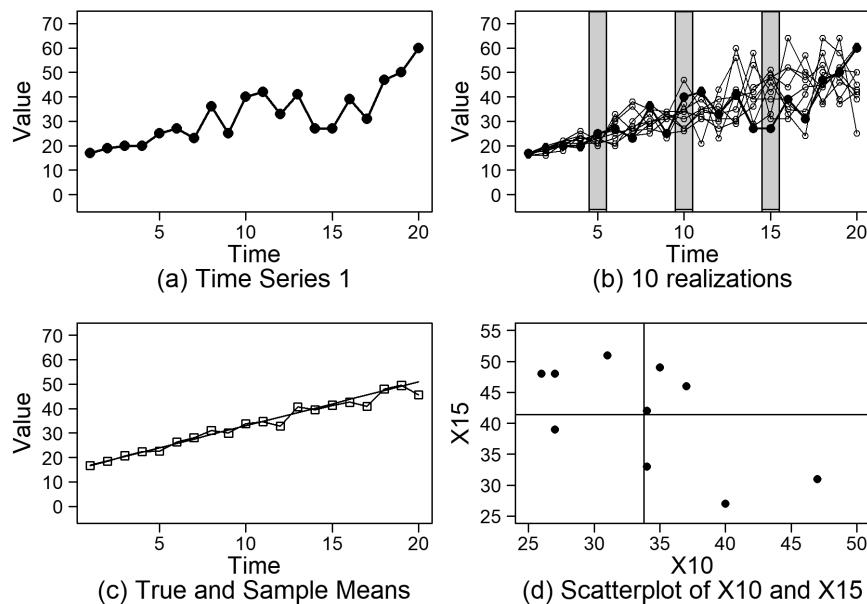


FIGURE 3.10 (a) A realization from Time Series 1, (b) multiple realizations from Time Series 1, highlighting $t = 5, 10, 15$, (c) plot of sample means (as open squares) and the true mean line, (d) scatterplot between X_{10} and X_{15} based on data in columns two and three in Table 3.5.

Estimating autocorrelations from multiple realizations. For any integers t_1 and t_2 between 1 and 20, we can also estimate the autocorrelation between random variables X_{t_1} and X_{t_2} because these are simply two random variables in the same sense that quiz and final exam scores are two random variables. For example, to estimate the autocorrelation between X_{10} and X_{15} , we use the associated columns in Table 3.5 and perform the following calculation:

$$r_{X_{10}X_{15}} = \frac{(40-33.8)(27-41.4)+(27-33.8)(48-41.4)+\dots+(47-33.8)(31-41.4)}{9(6.5)(8.5)} \\ = -0.62$$

These results can also be obtained using the following R code:

```
x10=c(40,27,35,34,31,27,26,34,37,47)
x15=c(27,48,49,33,51,39,48,42,46,31)
cor.x10x15=cor(x10,x15)
```

Figure 3.10(d) shows a scatterplot of the ten ordered pairs $(40,27), (27,48), \dots, (37,46), (47,31)$ along with a vertical line at the sample mean of X_{10} at 33.8 and a horizontal line at the sample mean of X_{15} at 41.4. The negative correlation can be seen because the “cloud of points” in Figure 3.10(d) slopes downward to the right. X_{10} values above its mean (33.8) seem to be paired with X_{15} values below its mean (41.4), and vice versa.

Inside Information:

The statistical model used to generate Time Series 1 has the form

$$X_t = 15 + 1.8t + .6ta_t, \quad (3.6)$$

where for each t , a_t is a randomly chosen observation from a normal (Gaussian) distribution with mean zero and standard deviation one. Consequently, μ_{X_t} follows the straight line, $\mu_{X_t} = 15 + 1.8t$.¹⁵ Also, at time t , $\sigma_{X_t} = .6t$.¹⁶ Thus, we have $\sigma_{X_5} = 3$, $\sigma_{X_{10}} = 6$, and $\sigma_{X_{15}} = 9$, indicating that the standard deviations also increase in time. These population standard deviations are well estimated by the sample standard deviations shown in Table 3.5.



QR 3.6
Explanation of
Footnotes 16

Keep this in mind:

In the analysis of real data, we will not in all likelihood know the “inside information”. We have the following Key Points based on the analysis of Time Series 1.

Key Points

1. The original realization appears to be upward trending, leading us to suspect that the true means, μ_{X_t} , are increasing in time.
2. If the original realization was our only realization, then while we would suspect that the time series has a general tendency to increase, our evidence would be weak.
3. The multiple realizations allow us to estimate the mean, μ_{X_t} , and standard deviation, σ_{X_t} , for each t using the sample mean, \bar{x}_t , and sample standard deviation, s_{X_t} , respectively.
4. The general tendency for each realization to trend upward in time and the fact that the sample means at each t tend to increase both suggest that the population means, μ_{X_t} , trend upward in time.
5. The fact that visually, the variability in the realizations seems to increase as t increases along with the fact that the sample standard deviations increase with time both suggest that the population standard deviations, σ_{X_t} , are increasing with time.
6. In this case, we can obtain several realizations from the time series only because we know its “formula” (Equation (3.6)). In other situations (for example, the shampoo sales data to be discussed in Example 3.5), multiple realizations can be obtained without a formula (different stores, different time periods).
7. By using the formula (inside information), we have shown that the true means actually trend linearly upward ($\mu_{X_t} = 15 + 1.8t$) and that the standard deviations $\sigma_{X_t} = .6t$ also increase with time.

3.2.1.2 Time Series 2: Y_t (Example 3.3 Continued)

As was the case with Figure 3.9(a), the increase in time for the realization in Figure 3.9(b) suggests that the true means, μ_{Y_t} , increase in time. To examine this supposition as we did for Time Series 1, we generated nine more realizations from the same random process that generated the realization in Figure 3.9(b). In Figure 3.11(a) we show the realization in Figure 3.9(b), and Figure 3.11(b) is a plot of the original realization (in bold) along with the nine new randomly selected realizations from the same time series. In Figure 3.11(b) we see that additional realizations from the underlying and “unknown” model go “all over the place”. There is no uniform tendency for the data to trend up or follow any other pattern.

¹⁵ Two facts about expectation are as follows: (1) if c is a constant, then $E[c] = c$ and (2) expectation is additive, that is, $E[aX + bY] = aE[X] + bE[Y]$. Thus, in our case, $X_t = 15 + 1.8t + .6ta_t$, so $\mu_t = E[X_t] = E[15 + 1.8t + .6ta_t] = E[15] + E[1.8t] + E[.6ta_t] = 15 + 1.8t + .6tE[a_t] = 15 + 1.8t$ because $15 + 1.8t$ is deterministic and $E[a_t] = 0$ (a_t has zero mean).

¹⁶ The variance at time t is $Var(t) = E[(X_t - \mu_{X_t})^2] = E[(15+1.8t+.6ta_t - (15+1.8t))^2] = E[(.6ta_t)^2] = E[(.36)t^2a_t^2] = .36t^2 E[a_t^2] = .36t^2$ because $E[a_t] = 0$ and $1 = Var[a_t] = E[(a_t - 0)^2] = E[a_t^2]$. $SD(X_t) = \sqrt{Var(X_t)} = .6t$.

Note: Although Figure 3.11(b) simply looks like an uninterpretable “mess”, there is much to learn from a closer examination of these realizations.

Figure 3.11(b) highlights the multiple values at times $t = 1$ and 2 and at times $t = 9$ and 10 . Table 3.6 shows the actual values along with the sample means ($\bar{x}_1 = 28.4$, $\bar{x}_2 = 28.9$, $\bar{x}_9 = 28.1$, $\bar{x}_{10} = 30.0$), sample standard deviations ($s_{x_1} = 18.3$, $s_{x_2} = 17.8$, $s_{x_9} = 14.9$, $s_{x_{10}} = 12.5$). The sample autocorrelations, $r_{x_1 x_2} = .90$ and $r_{x_9 x_{10}} = .79$ are also shown in the table.

While not following a recognizable pattern, the realizations from Time Series 2 show positive correlation between adjacent values. That is, y_t tends to be fairly close to y_{t-1} (or y_{t+1}). This is seen in Table 3.6 where the sample autocorrelations, $r_{x_1 x_2} = .90$ and $r_{x_9 x_{10}} = .79$ are fairly high. The scatterplots associated with both of these random variable pairs are shown in Figures 3.11(d) and (e). There it can be seen that (i) there is a strong positive autocorrelation in each case and (ii) the scatterplots look similar. We will return to Figure 3.11(f) in Section 3.3, but it is not too early to begin thinking about it.

TABLE 3.6 Data from Multiple Realizations for $t = 1$ and 2 and Again for $t = 9$ and 10 Shown in Figure 3.11

REALIZATION	X_1	X_2	X_9	X_{10}	
1	23	15	32	21	
2	13	24	31	25	
3	46	43	43	36	
4	35	25	8	26	
5	8	4	30	38	
6	35	33	9	10	
7	63	63	41	50	
8	3	7	6	16	
9	36	36	38	36	
10	22	39	43	42	
\bar{x}_1, \bar{x}_2	28.4	28.9	28.1	30.0	\bar{x}_9, \bar{x}_{10}
s_{x_1}, s_{x_2}	18.3	17.8	14.9	12.5	$s_{x_9}, s_{x_{10}}$
$r_{x_1 x_2}$	90		.79		$r_{x_9 x_{10}}$

Table 3.7 shows the sample means and standard deviations for the ten observations from each random variable X_t , $t = 1, \dots, 20$.¹⁷ Examination of Table 3.7 shows that the sample means are quite similar across time, and they tend to be between 28.1 and 34.3. The sample means are plotted in Figure 3.11(c) along with the true mean line at 30. Likewise, the sample standard deviations range from 7.4 to 18.3 with 7.4 being unusually low.

¹⁷ The actual observations are shown in Table 3.6 for X_1, X_2, X_9 , and X_{10} , whereas Table 3.7 gives only the sample statistics for all 20 of the time periods.

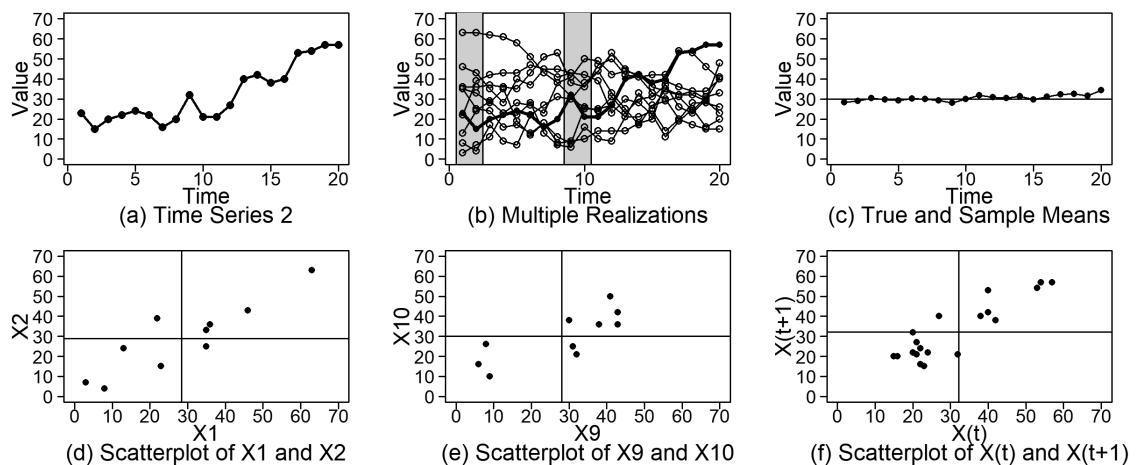


FIGURE 3.11 (a) A realization from Time Series 2, (b) multiple realizations from Time Series 2, (c) sample and true means, (d) and (e) scatterplots between X_1 and X_2 and between X_9 and X_{10} , respectively, and (f) scatterplot of X_t and X_{t+1} , $t = 1, \dots, 19$ for Time Series 2 realization in (a).

In addition to the sample statistics for each random variable, the estimated autocorrelations between X_t and X_{t+1} are given in the table for $t = 1, \dots, 19$. Note that the sample autocorrelations between X_1 and X_2 and between X_9 and X_{10} were already given in Table 3.6 (along with the data from which they were calculated). We noted that the estimated autocorrelations (.90 and .79), and scatterplots (Figures 3.11(d) and (e)) were quite similar for the random variable pairs X_1, X_2 and X_9, X_{10} respectively. This behavior is not unique to these two random variable pairs, and in fact the estimated autocorrelations between random variables X_t and X_{t+1} (that is, random variables that differ by one on the time axis) tended to range from about .75 to .95 (with one unusually small estimate of .56). Figure 3.11(f) is a scatterplot of the 19 pairs of data values plotted in Figure 3.11(a) that are separated by one time unit. That is, $(x_1, x_2), (x_2, x_3), \dots, (x_{19}, x_{20})$. These ordered pairs will be listed in Table 3.8, and are shown in Figure 3.11(f) as a scatterplot. This scatterplot is consistent with a correlation coefficient of about .85.

Key Point: In Section 3.3 we will discuss *stationary* processes, which are members of a broad class of time series that share the behavior exhibited by Time Series 2.

TABLE 3.7 Means and Standard Deviations for Table 3.6 Data $x_t^{(i)}, i = 1, \dots, 10$ for Time t

t	1	2	3	4	5	6	7	8	9	10
\bar{x}_t	28.4	28.9	30.5	29.6	29.3	30.5	29.9	29.2	28.1	30.0
s_{X_t}	18.3	17.8	14.6	14.9	14.9	13.8	14.5	16.6	14.9	12.5
$r_{X_t X_{t+1}}$.90	.95	.90	.89	.88	.85	.90	.91	.79	***
	11	12	13	14	15	16	17	18	19	20
\bar{x}_t	31.7	30.9	30.5	31.3	29.8	31.1	32.3	32.4	31.6	34.3
s_{X_t}	14.5	14.3	11.3	10.1	7.4	10.7	12.6	12.3	12.5	13.3
$r_{X_t X_{t+1}}$	*	.84	.80	.86	.82	.56	.75	.94	.91	.83

* $r_{X_{10} X_{11}} = .93$

Inside Information:

The realizations in Figure 3.11(b) are generated using the following:

$$X_t = 3 + .9X_{t-1} + a_t, \quad (3.7)$$

where for each t , a_t is a randomly selected normal random variable with zero mean and variance 40. This is our first example of an “autoregressive model”. Model (3.7) is known as an autoregressive process of order 1 (denoted AR(1)) and will be discussed in detail in Chapter 5. For the process in (3.7), it can be shown that the true means and standard deviations are constant across time ($\mu_{X_t} = 30$ and $\sigma_{X_t} = 14.5$ for each t , for $t = 1, \dots, 20$). Also, the autocorrelation $\rho_{X_t X_{t+1}} = .9$ for each t , for $t = 1, \dots, 19$.

Key Points

1. Both realizations in Figure 3.9 show a trending-type behavior.
2. The availability of multiple realizations revealed the fact that the trending in Figure 3.9(a) is “real” while the trending in Figure 3.9(b) is only temporary and is not repeatable in other realizations from the same model.
3. Points 1 and 2 above show that it is difficult to establish whether an observed trend should be predicted to continue based on a single realization, especially if n is small.
 - We will return to this topic in Chapter 8.

3.2.2 The Effect of Realization Length

The apparent trending in Time Series 1 and 2 in Example 3.3 was more fully understood because of the availability of multiple realizations. In this example, each realization is of length $n = 20$. The two realizations in Figure 3.9 seemed to be trending upward and were almost indistinguishable in terms of an interpretation. Recall however that the use of multiple realizations revealed the following:

- (a) *Time Series 1*: Realizations tended to consistently increase in time and to have increasing variability.
- (b) *Time Series 2*: Realizations wandered around aimlessly. There was a tendency for neighboring data values to be similar and the average of the ten multiple realizations at each time period was approximately 30.

Although we have spent some time discussing the information that can be obtained from multiple realizations, the fact is that we will usually have only one realization. (This will be a focus of future discussion.) Time series analysis usually involves longer realizations, and it is not uncommon for realizations of length $n = 20$ to contain insufficient information for a thorough analysis. In Figure 3.12, we show Time Series 1 and 2 extended to realization lengths $n = 100$. In Figure 3.12, the first 20 data values in each plot are the values of the realizations in Figure 3.9, and the following 80 data values are obtained by “following the formula” that produced the first 20. From single realizations of length $n = 100$, we see the following:

- (a) *Time Series 1*: The realization continues to increase in time and the increase in variability is very obvious.
- (b) *Time Series 2*: The realization continues to increase until about $n = 30$, then declines for about the next 10 time points, but in general, just wanders around. The tendency for neighboring data values to be similar continues and the realizations wander around, never getting very far from 30. $n = 100$.

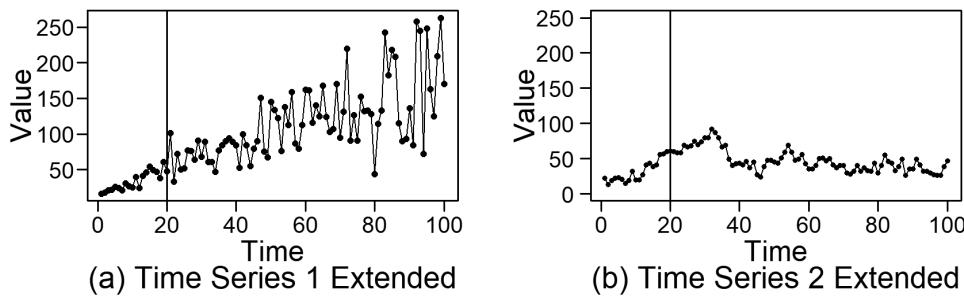


FIGURE 3.12 Figures 3.9 (a) and (b) extended to realizations of length $n = 100$.

Key Points

1. The results in this example are somewhat “sobering” because, in most time series applications these authors have encountered, only one realization is available.
2. Drawing conclusions from a single realization must be done with caution.
3. Individual time series of length as small as $n = 20$ rarely provide enough information for an informed analysis.

Example 3.4 December Sales Data with Multiple Realizations

Based on the analysis of the two datasets in Figure 3.9, it is clear that the behavior of the time series data can be misleading if not properly understood. For example, suppose the realizations in Figure 3.9 represent daily sales (in hundreds of dollars) of a certain product in two different stores for the days December 1–20, 2009. Let’s assume that Figures 3.9(a) and (b) refer to sales data for Store 1 and Store 2, respectively. In each case the store owner is led to believe that sales of the product will increase during the first 20 days of December. We let the nine additional realizations represent the sales data for December 1–20, for the years 2010–2018. Although not identified as such in the plots, we now assume that each of the realizations in Figure 3.10(b) and Figure 3.11(b) corresponds to one of the years between 2009 and 2018. For Store 1, the multiple realizations shown in Figure 3.10(b) give the store owner some confidence that the first 20 days of December 2019 will behave in a similar manner. However, given the data in Figure 3.11(b) for Store 2, the owner should see that the increase in sales in 2009 was more of a “random occurrence”, and an increase in sales during the first 20 days of December should not be expected to occur on an annual basis.

Forecasting Using Multiple Realizations

In many applications, data (such as the hypothetical sales data in Figure 3.9) are collected for the purpose of making “forecasts”. We consider two situations:

- (a) In November 2019, the store owner might want to forecast sales for the first 20 days of December using the information from the previous years.
 - (i) *Store 1:* The multiple realizations should provide the owner of Store 1 with some confidence that sales will increase in a similar manner *given that conditions have not changed*.
 - For example, in November 2020 (the COVID pandemic year) it would be advisable for the store owner to proceed cautiously and not assume similar behavior.
 - (ii) *Store 2:* For the owner of Store 2, about the only information suggested by the multiple realizations is that sales seem to average about \$3,000 per day with no discernible pattern.
 - Again, this prediction assumes that conditions do not change (which would likely not have been true in the pandemic year).

- (b) Data are often collected for purposes of predicting into the immediate future. That is, in the sales data example, store owners might want to predict sales in 2018 for December 21–24 supposing that historical data are only available for December 1–20.
- Store 1:* The sales for December 1–20 seem to follow a linearly increasing pattern. Fitting a line to the data (which seems justified for days 1–20 of each year) and extending it for the next four days is a type of *extrapolation*. As we have discussed, extrapolation is discouraged (or should at least be done very cautiously). In this case, extrapolation is a bad idea unless there is additional information suggesting that conditions will not change. However, one might speculate that sales would increase more rapidly in these days due to last-minute shopping.
 - Store 2:* It seems that the best guess for the future is to predict the average of observed data (that is, \$3,000 per day). However, last minute shopping might alter this expectation.

Key Points

- Forecasting strategies will be a main focus of this book.
- The forecasting techniques we will discuss are based on forecasting the immediate future given a *single realization*.
- Forecasts based on time series models will be accompanied by prediction limits that acknowledge the inherent uncertainty.

Example 3.5 Shampoo Sales Data

Figure 3.13 is a plot of shampoo sales at a hypothetical store with sales at days 10 and 33 highlighted. In time series terminology, the data in Figure 3.13 represent a realization, $x_t, t = 1, \dots, 50$. We note that $x_5 = 9$ and $x_{33} = 20$. For this particular store, sales seemed to be higher on day 33 than on day 10.

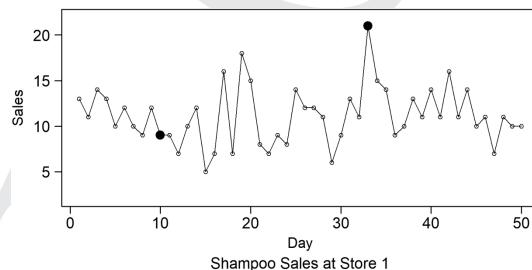


FIGURE 3.13 Shampoo sales recorded at a particular store.

We may question whether sales at day 33 are typically higher on average than at day 10, that is, whether $\mu_{x_{33}}$ is greater than $\mu_{x_{10}}$. In order to better understand the situation, we obtained corresponding sales data for four randomly selected, similar stores. Figure 3.14 shows shampoo sales data for these four stores. The sales data for times $t = 10$ and $t = 33$ are highlighted. In this case, these realizations represent a sample from the ensemble of corresponding sales data for all the stores of interest. Averaging the data over the five realizations at times $t = 10$ and $t = 33$, we obtain $\bar{x}_{10} = (9 + 10 + 7 + 16 + 9) / 5 = 10.2$ and $\bar{x}_{33} = (21 + 23 + 21 + 17 + 20) / 5 = 20.4$. The evidence suggests that $\mu_{x_{33}} > \mu_{x_{10}}$. We cannot be sure of this fact because we only had a sample of five observations and not the entire ensemble itself. Given that we only have a sample of five stores at each value of t , we do not know either $\mu_{x_{10}}$ or $\mu_{x_{33}}$ for certain. However, we do have evidence (the data).

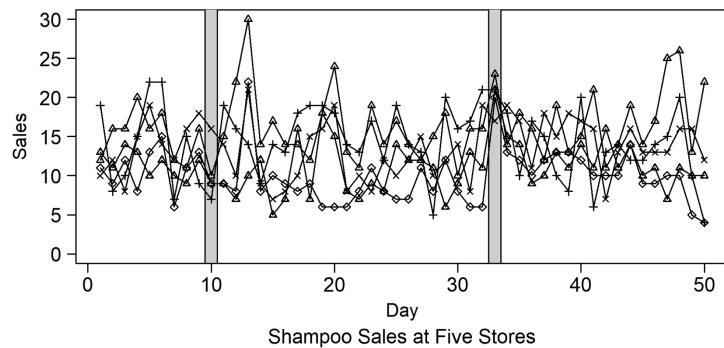


FIGURE 3.14 Shampoo sales recorded over the same time period at the store in Figure 3.13 and four similar stores.

3.3 STATIONARY TIME SERIES

In Section 3.2 we discussed the fundamentals of a statistical approach to time series analysis. We noted that in most time series scenarios these authors have encountered, there is only one realization available for analysis. For example, the data for monthly DOW closing averages, the price of West Texas Intermediate crude oil, Texas unemployment data, Dallas-Ft. Worth temperature data, and sunspots are only available as single realizations. We cannot somehow turn back the clock and obtain other realizations. In Example 3.3 we discovered the fact that the availability of only single, admittedly short ($n = 20$), realizations caused problems with interpretation. So, what do we do? We discussed the fact that longer realizations provide increased information. Also, *stationary time series* are a type of time series for which we can make meaningful analysis given only a single realization. This statement is sufficiently important to state separately as a Key Point.

Key Point: Meaningful analysis of a *stationary time series* can often be made given only a single realization.

Time Series 2 in Example 3.3 is an example of a realization from a stationary time series process. In Figure 3.11(b) and Table 3.7, we can see that the ten realizations from Time Series Model 2, although seemingly dissimilar, actually share the following properties:

- The sample means for each t were all in the neighborhood of 30 and showed no tendency to increase or decrease with time.
- The sample standard deviations were similar for each $t, t = 1, \dots, 20$.
- The autocorrelations $r_{X_t X_{t+1}}, t = 1, \dots, 19$ were similar and showed no tendency to change with time.

Another way to express the behavior of Time Series 2 is that it seems to be in a sort of *equilibrium*. That is, the basic behavior of the time series does not change with time. Recall that we have said that a time series is a sequence of random variables X_t with population parameters, mean, μ_{X_t} , and standard deviation, σ_{X_t} . Also, the population correlation between random variables X_{t_1} and X_{t_2} is denoted $\rho_{X_{t_1} X_{t_2}}$,

and is called the population autocorrelation. With this notation in mind, a stationary process is defined in Definition 3.13.

Definition 3.13 (Stationarity):¹⁸ The time series $\{X(t); t \in T\}$ is said to be *covariance stationary* if

1. $\mu_{X_t} = \mu$ (constant mean for all t)
2. $\sigma_{X_t}^2 = \sigma^2 < \infty$ (i.e., a finite, constant variance for all t)
3. $\gamma_{X_1 X_2}$ and $\rho_{X_1 X_2}$ depend only on $t_2 - t_1$.



QR 3.7 Conditions
of Stationarity

Condition 1: $\mu_{X_t} = \mu$ (constant mean for all t)

Figure 3.10(b) strongly suggests that for Time Series 1 in Example 3.3, the means for each t are not constant. In fact, we showed (using our “inside information”) that μ_{X_t} is linearly increasing in time which is supported by the plot of sample means in Figure 3.10(c). Consequently, Time Series 1 is *not* a stationary process. However, the realizations from Times Series 2 shown in Figure 3.11(b) wander around 30 but show no uniform tendency to trend up or down. This behavior is consistent with Condition 1 of stationarity. In Example 3.3 we stated that the realizations in Figure 3.11(b) were from an AR(1) model with $\mu = 30$. We will study AR(1) processes in Chapter 5 where we will discuss the fact that these processes are stationary.

The Air Passenger data (Figure 2.7(a)) and annual DOW data (Figure 1.9) increase in time which makes the constant mean questionable. The global temperature data show an increase, and the recent debate about global warming is in essence a debate about whether there is an increasing “population mean”.

Condition 2: $\sigma_{X_t}^2 = \sigma^2 < \infty$ (constant and finite variance)

Processes that satisfy this condition seem to have a variability that does not change with time. An example of changing variability is Time Series 1 in Figure 3.10 and even more dramatically in Figure 3.12(a). While the variability change is not very apparent in the original realization, by examining the ten realizations in Figure 3.10(b) we see that the variability is increasing with time.

As mentioned in Chapter 2, the Air Passenger data (Figure 2.7(a)) show more variability in later years, and one method for adjusting for this increase in variability is to analyze the “log airline data” (Figure 2.7(b)) which has a more stable variability across time. Also, the bat echolocation data in Figure 3.15(a)¹⁹ and the seismic data in Figure 3.15(b)²⁰ show a distinct decrease in variability across time toward the end of the realizations.²¹

¹⁸ Time series satisfying the conditions in Definition 3.13 are sometimes referred to as covariance stationary, weakly stationary, or second-order stationary. There is another more restrictive type of stationarity called strict stationarity. In this book, when we refer to a stationary process, we will be referring to one that satisfies the conditions of Definition 3.13.

¹⁹ Figure 3.15(a) shows big brown bat echolocation data furnished by Al Feng of the Beckman Center at the University of Illinois. The data consist of 381 points sampled at 7-microsecond intervals.

²⁰ Figure 3.15(b) is a seismic Lg wave from an earthquake known as Massachusetts Mountain earthquake (August 5, 1971) which was recorded at the Mina, Nevada station.

²¹ We stated that a random sample of size $n = 1$ gave no information about variability. Because a realization represents “a single time series observation” it may be confusing that we are using the “single observation” to discuss variability change. However, in the seismic data, for example, by looking “down the realization” we can obtain the following information: (a) The variability changes with time. (b) In this particular case, the variability change down the timeline is gradual, so that some information about the variability at time t_k may be gleaned from the time-to-time variability of observations close to t_k .

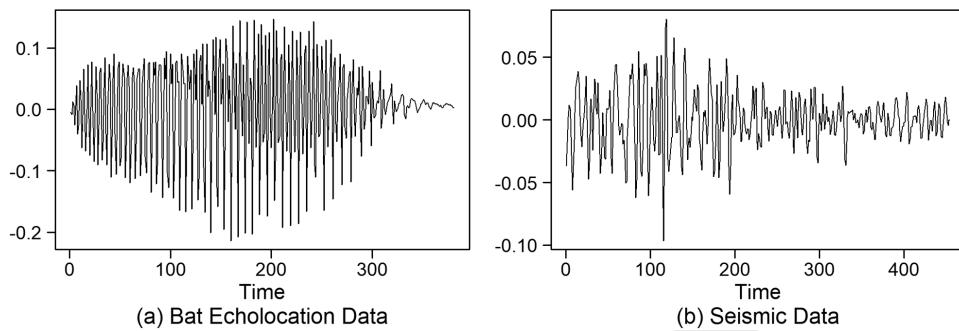


FIGURE 3.15 (a) Bat Echolocation data and (b) Seismic Lg wave from an earthquake.

Condition 3: $\gamma_{X_1 X_2}$ and $\rho_{X_1 X_2}$ depend only on $t_2 - t_1$.

This condition can be restated as follows:

“The autocorrelation between X_{t_1} and X_{t_2} depends on how far apart t_1 and t_2 are, not where they are in time.”

If Condition 3 is satisfied, there is no confusion in replacing the notation, $\gamma_{X_1 X_2}$ with $\gamma_{t_2 - t_1}$, and similarly denoting $\rho_{X_1 X_2}$ with $\rho_{t_2 - t_1}$. Letting $t_2 - t_1 = k$, we refer to γ_k and ρ_k as the *autocovariance and autocorrelation at lag k*, respectively. Also, it follows that $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$.

See Key Point 1 below.

Key Points

- $\gamma_{-k} = E[(X_t - \mu)(X_{t-k} - \mu)] = E[(X_{t-k} - \mu)(X_t - \mu)] = \gamma_k$
 - $\rho_{-k} = \rho_k$
 - $\gamma_0 = E[(X_t - \mu)(X_t - \mu)] = E[(X_t - \mu)^2] = \sigma^2$
 - In the notations γ_k and ρ_k , the integer k is referred to as the *lag*.

A realization for which Condition 3 seems reasonable is Time Series 2, multiple realizations of which are shown in Figure 3.11(b). Table 3.7 shows the correlation estimates between X_t and X_{t+1} for $t = 1, 2, \dots, 19$. These estimated correlations were obtained using the ten realizations that were available. The autocorrelations tended to be in the range .7 to .9. That is, the autocorrelations between adjacent variables did not seem to change across the timeline.

As mentioned, the typical situation is that only one realization is available (and for our example, this is Time Series 2 in Figure 3.11(a)). Table 3.8 lists the available pairs of observations in Time Series 2 that are one time unit apart. These are plotted in Figure 3.11(f) and in the discussion following Figure 3.11 we noted that the associated correlation coefficient is in the neighborhood of .85. Section 3.3.2.3 will provide a formula for estimating the “lag 1” autocorrelation based on the scatterplot data in Figure 3.11(f).

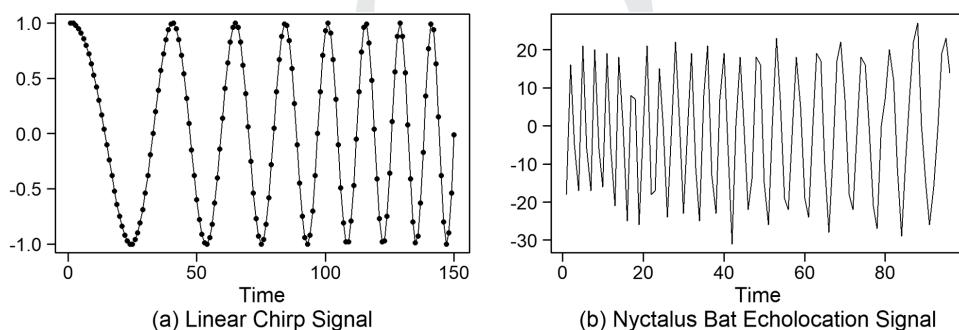
TABLE 3.8 19 Pairs of Observations in Time Series 2 that Differ by One Time Unit

t	x_t	x_{t+1}	t	x_t	x_{t+1}
1	22	13	11	20	20
2	13	19	12	27	41
3	19	22	13	41	44
4	22	23	14	44	56
5	23	21	15	39	57
6	21	15	16	41	60
7	15	19	17	56	57
8	19	32	18	57	60
9	32	20	19	60	61
10	20	27			



QR 3.8 Positive and Negative Autocorrelations

While it may be difficult to “tell by looking” that a correlation structure stays constant across time, there are obvious examples in which this does not happen. An example of a time series realization that clearly violates this condition is the linear chirp data in Figure 3.16(a). Examination of the data shows a strong positive autocorrelation between X_t and X_{t+1} early in the data set. This positive autocorrelation appears to become smaller in time to the point that toward the end of the realization the autocorrelation between X_t and X_{t+1} is near zero or even negative. The Nyctalus noctula hunting bat echolocation signal sampled at 44×10^{-3} seconds shown in Figure 3.16(b) has a varying correlation structure that is not quite so obvious. Time series associated with DFW temperatures, lynx trappings, and Texas Unemployment have realizations for which it is not apparent that there is a violation of Condition 3.

**FIGURE 3.16** (a) Linear chirp signal and (b) echolocation signal from a Nyctalus noctula bat.

Definition 3.14: White Noise Process

A white noise process satisfies the following conditions.

1. The X_t s are independent and identically distributed with mean zero and finite variance.
2. $\gamma_{X_{t_1} X_{t_2}} = \rho_{X_{t_1} X_{t_2}} = 0$ whenever $t_1 \neq t_2$. ($\gamma_k = \rho_k = 0$ if $k \neq 0$)
That is, the autocovariances and autocorrelations are equal to zero at lags $k \neq 0$.
3. $\gamma_{X_{t_1} X_{t_2}} = \sigma^2$ when $t_1 = t_2$ where $0 < \sigma^2 < \infty$. ($\gamma_0 = \sigma^2$ and $\rho_0 = 1$)

In words, whenever lag $k = 0$, the autocovariance is equal to the process variance and the autocorrelation is equal to one.

In a white noise process, each observation is uncorrelated with all other observations. It is analogous to a random sample indexed in time.²² An important fact is that white noise process is stationary. (See Problem 3.8.)

Key Points

1. Stationary time series models, such as autoregressive or autoregressive moving average models, are associated with model-based parameters such as the:
 - mean (μ)
 - variance $\sigma^2(\gamma_0)$
 - autocorrelations and autocovariances (ρ_k and γ_k)
2. It is common practice to plot the autocorrelations, $\rho_k, k = 0, 1, \dots, K$ for some integer K .
 - There is no need to plot autocorrelations for negative lags since $\rho_{-k} = \rho_k$ for all lags.
3. In Section 3.3.1 we will discuss the estimation of these parameters for a given realization.

3.3.1 Plotting the Autocorrelations of a Stationary Process

It is important to understand the extent to which a plot of the autocorrelations describes the behavior of a time series realization. Figure 3.17 shows realizations from four stationary, autoregressive time series models, while Figure 3.18 shows the corresponding model-based autocorrelations for lags 0–30.

Realization 1 in Figure 3.17(a) has a random wandering behavior. Note that it is typical for x_t and x_{t+1} to be relatively close to each other with a few exceptions, e.g. around $t = 50$. That is, the value of the random variable X_{t+1} is usually not very far from the value of X_t , and, as a consequence, there is a noticeably strong positive autocorrelation between the random variables X_t and say X_{t+1} (ρ_1 is at least .9). Note also that the autocorrelation between X_t and X_{t+k} decreases as the lag k increases. This decrease leads to the fact that by lag 30 there is very slight correlation between X_t and X_{t+30} . By examining Figure 3.18(a) it appears that $\rho_{30} \approx .10$.

²² From a mathematical perspective, we know that “independence” and “uncorrelated” are equivalent concepts if the data are normally distributed. In general, independence implies uncorrelated, but not vice versa.

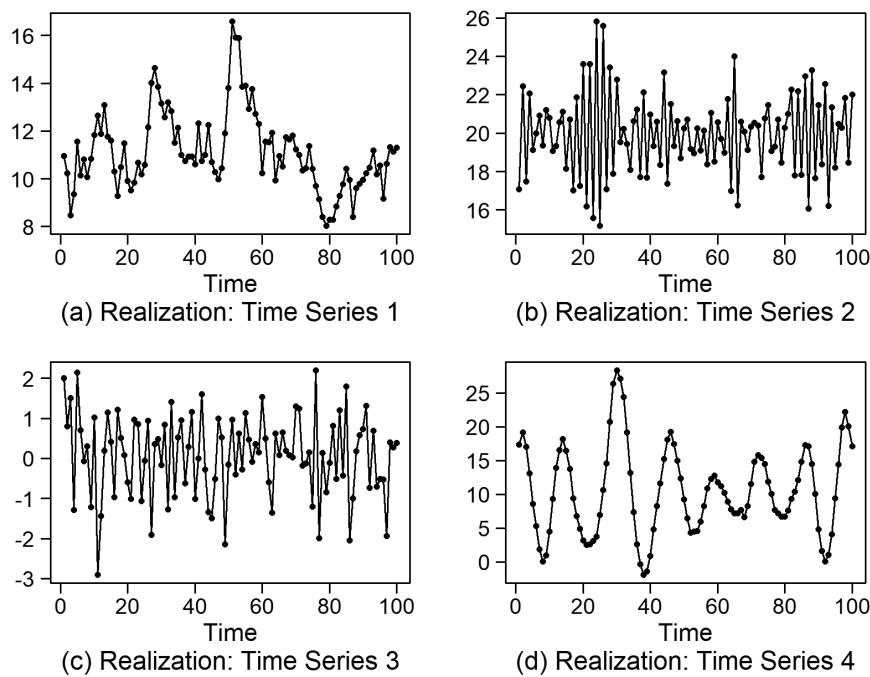


FIGURE 3.17 Four realizations from stationary processes. (a) Realization 1 (b) Realization 2 (c) Realization 3 (d) Realization 4.

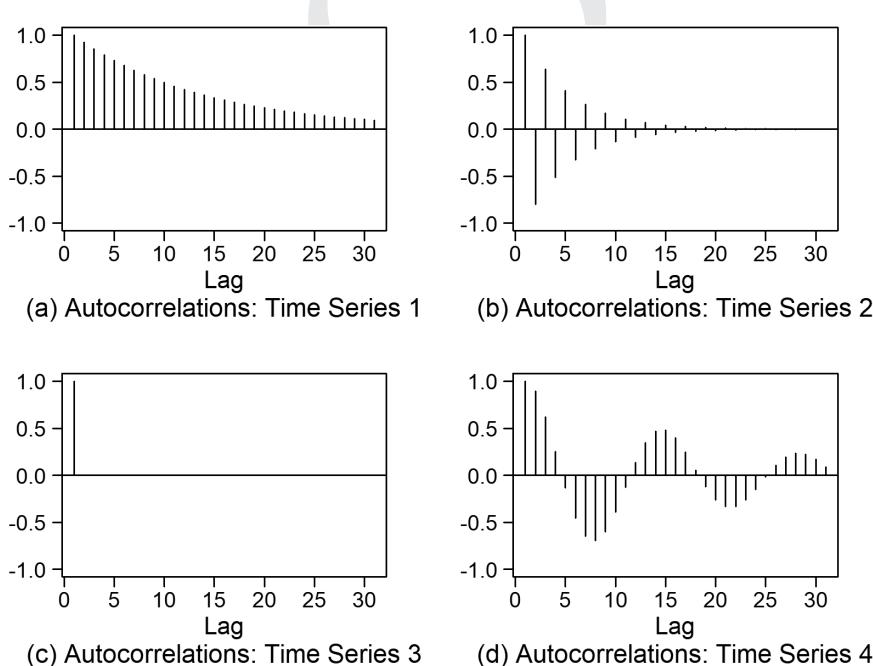


FIGURE 3.18 True autocorrelations from models associated with realizations in Figure 3.17 (a) Realization 1 (b) Realization 2 (c) Realization 3 (d) Realization 4.

Realization 2 in Figure 3.17(b) shows a highly oscillatory behavior. In fact, if x_t is above average, then x_{t+1} tends to be below average, x_{t+2} tends to be above average, and so forth. The autocorrelations in Figure 3.18(b) describe this behavior where we see that $\rho_1 \approx -0.8$ while $\rho_2 \approx 0.6$, and $\rho_3 \approx -0.5$, and so on. Note also that the up-and-down pattern is sufficiently imperfect that the autocorrelations damp to near zero by about lag 15.

Realization 3 in Figure 3.17(c) shows an absence of pattern. That is, there appears to be no relationship between X_t and X_{t+1} and as a matter of fact, there is seemingly no correlation between X_t and X_{t+k} for any $k \neq 0$. In actual fact, this is a realization from a white noise model, and for this model, $\rho_k = 0$ whenever $k \neq 0$, as can be seen in Figure 3.18(c). Notice that in all autocorrelation plots, $\rho_0 = 1$.

Realization 4 in Figure 3.17(d) is characterized by cyclic behavior with an average cycle-length of about 14 time points. The corresponding autocorrelations in Figure 3.18(d) show a damped sinusoidal behavior. Note that, not surprisingly, because of the cycle length of about 15, there is a positive correlation between X_t and X_{t+15} . Also, there is a substantial negative correlation at lags 7 and 8, because within the sinusoidal cycle of length 15, if x_t is above the average, then x_{t+7} would be expected to be below average and vice versa. Note also that this autocorrelation structure holds up fairly “solidly” as far as $k = 30$.

3.3.2 Estimating the Parameters of a Stationary Process

In this section we discuss the estimation of the parameters of a stationary process given a realization of length n .

3.3.2.1 Estimating μ

Note that if the means do not depend on time (that is, Condition 1 is satisfied), then the given realization provides n observations from random variables that share the same population mean (μ). Thus, because each observation is estimating the same mean, it makes sense that we can estimate μ with the sample mean of X_1, \dots, X_n , that is, by “averaging down the realization time line”. The formula looks familiar: $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$. This looks like a sample average from a random sample.

Key Point: For a stationary time series, we can use the data across time to estimate the mean because the mean is assumed to be the same for each time, t .

The difference, however, is that the observations are not independent, and instead are related. The sample mean from a random sample is based on n independent observations of the phenomenon. However, if observations in a time series are highly dependent, then a realization of length 30 does not provide 30 independent pieces of information.

Recall that when data are from a random sample of size n with variance σ^2 , then we have the classical result that $Var(\bar{X}) = \sigma^2 / n$ (and $SE(\bar{X}) = \sigma / \sqrt{n}$). However, if X_t is a stationary time series, then the variance of \bar{X} based on a realization of length n is given by

$$Var(\bar{X}) = \frac{\sigma^2}{n} + 2 \frac{\sigma^2}{n} \sum_{k=1}^{n-1} \left(1 - \frac{|k|}{n} \right) \rho_k \quad (3.8)$$

Equation 3.8 is written in the form $\text{Var}(\bar{X}) = \frac{\sigma^2}{n} + Q$ where clearly the standard error of \bar{X} increases over that of a random sample of size n (that is, $Q > 0$) when the autocorrelations are positive. The following example illustrates the effect of dependence on the variability of \bar{X} .

Example 3.6 Consider the data plotted in Figure 3.19 where, in each case, the true mean is 30 and the population variance is $\sigma_x^2 = 50$. Figure 3.19(a) is a plot of a random sample, X_t , (with data values connected). Figure 3.19(b) is a “sample” (actually a realization) from the AR(1) model that generates data values using the formula

$$X_t = .3 + .99X_{t-1} + a_t, \quad (3.9)$$

where the a_t s are a white noise process with mean zero and $\sigma_a^2 = 1$ and $\sigma_x^2 = 50$. In Figure 3.19(a), it can be seen that the independent observations provide unrelated observations “hovering” around the mean $\mu = 30$. On the other hand, note that in Figure 3.19(b), the first observation is about 40 and subsequent observations are correlated with (sort of “tied to”) this observation. In fact, the correlation structure is such that all of the 50 data values are above 30.

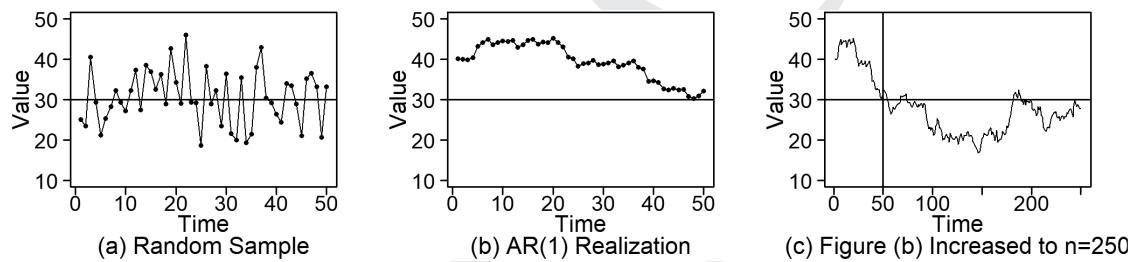


FIGURE 3.19 (a) Random sample (white noise) of length $n = 50$ (b) AR(1) realization of length $n = 50$ from the model $X_t = 9 + .99X_{t-1} + a_t$, with $\sigma_a^2 = 1$. In both cases, the theoretical mean is $\mu = 30$ and $\sigma_x^2 = 50$ (c) the realization in (b) extended to $n = 250$ observations.

Analogy: The following trivial analogy may be helpful. (Take it or leave it.) Consider the situation in which two sets of 50 testers are attempting to estimate the weight of a metal object. They are allowed to pick up and carefully examine the object but are not allowed to place it on a scale.

Figure 3.19(a) Data: The situation in Figure 3.19(a) is analogous to each of the 50 testers holding the object and independently and carefully using their common sense and expertise to come up with their “best guesses”.

Figure 3.19(b) Data: The data in Figure 3.19(b) is analogous to the situation in which we assume that Tester 1 uses the same carefully thought-out techniques employed by *each* of the 50 testers represented in Figure 3.19(a). However, Tester 2 holds the object, finds out the first tester’s estimate, and instead of coming up with an independent well thought-out estimate, estimates the weight to be fairly close to the estimate given by Tester 1. Similarly, after holding the object but before giving an estimate, Tester 3 learns the second tester’s estimate and gives a similar guess. The 49 testers following Tester 1 use the same strategy on down the line.

Question: Which estimate, (a) or (b), would you put the most faith in?²³ With regard to estimate (b), the only tester who gave an independent and seriously thought-out estimate was the first one. The others “just went along”.

The sample mean of the data in Figure 3.19(a) is 30.6 (quite close to 30), while the mean for the data in Figure 3.19(b) is 34.7, leaving the impression that the true mean of this process is really above $\mu = 30$. Figure 3.19(c) shows the realization in Figure 3.19(b) extended to 250 data values. There we continue to see a solid connection between neighboring data values, but the realization does eventually “wander” back toward $\mu = 30$. Because the testers did hold the object, the estimates never strayed extremely far from the true value.

Calculating $Var(\bar{X})$ (and $SE(\bar{X})$): We have obtained \bar{X} from three realizations:

- (a) From the random sample of size $n = 50$
- (b) From the highly dependent realization of length $n = 50$
- (c) From the highly dependent realization of length $n = 250$

We calculate $Var(\bar{X})$ (and $(SE(\bar{X}))$ in each case.

$$\begin{aligned} \text{(a)} \quad Var(\bar{X}) &= \frac{\sigma^2}{n} = \frac{50}{50} = 1 \quad (SE(\bar{X}) = 1) \\ \text{(b)} \quad Var(\bar{X}) &= \frac{50}{50} + 2 \sum_{k=1}^{49} \left(1 - \frac{k}{50}\right) \cdot 99^k = 42.6 \quad (SE(\bar{X}) = 6.5) \\ \text{(c)} \quad Var(\bar{X}) &= \frac{50}{250} + 2 \frac{50}{250} \sum_{k=1}^{249} \left(1 - \frac{k}{250}\right) \cdot 99^k = 25.2 \quad (SE(\bar{X}) = 5.0) \end{aligned}$$

Two other bits of information are of interest:

- (a) In order for the “dependent group” to achieve an $SE(\bar{X}) = 1$, it would require a realization length of $n = 9750$.²⁴
- (b) Suppose there was only one observation in the “random sample” group. In this case, $\bar{X} = X_1$, in which case $Var(\bar{X}) = 50 / 1 = 50$ with $SE(\bar{X}) = 7.1$. Surprisingly, this is not much greater than the corresponding quantities for a realization of length $n = 50$ in the “dependent group”.

In the previous example, we generated realizations from a model that introduced a strong autocorrelation structure, $\rho_k = .99^{|k|}$, that dies out very slowly. For comparison, we consider the following autocorrelation structures (that are associated with AR(1) models). Each of these models is designed to have mean $\mu = 30$ and variance $\sigma^2 = 50$.²⁵

- (a) $X_t = .3 + .99X_t + a_t$, $\sigma_a^2 = 1$
- (b) $X_t = 1.5 + .95X_t + a_t$, $\sigma_a^2 = 4.9$
- (c) $X_t = 3 + .9X_t + a_t$, $\sigma_a^2 = 9.5$
- (d) $X_t = 6 + .8X_t + a_t$, $\sigma_a^2 = 18$

The autocorrelation plots for these four models are given in Figure 3.20.

²³ Testers in group (a) would probably have the better overall estimate of the mean.

²⁴ The correct reaction at this point is “WOW”.

²⁵ Autoregressive processes and their properties will be discussed in Chapter 5.

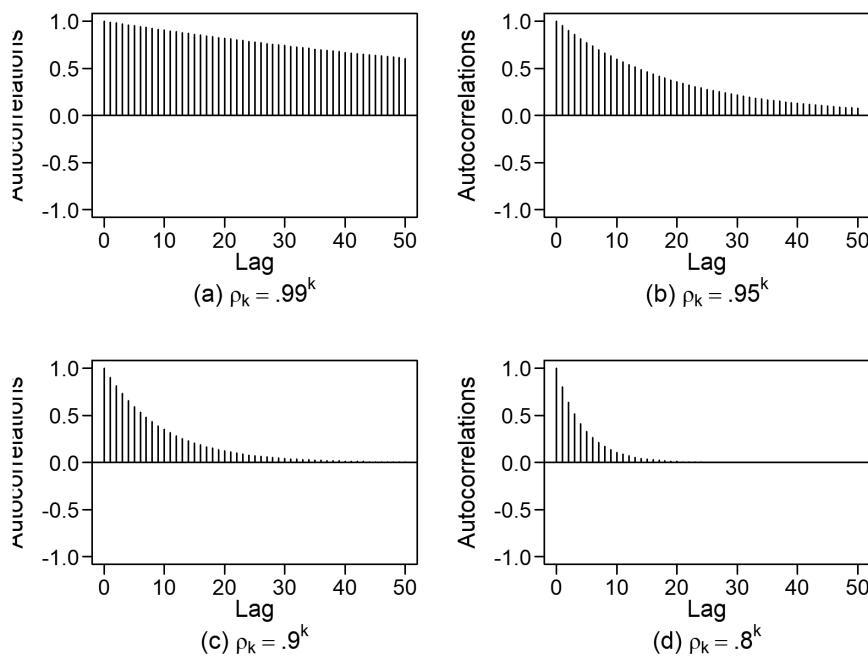


FIGURE 3.20 Model-based autocorrelations for AR(1) models $X_t = \beta + \phi_1 X_{t-1} + a_t$, with (a) $\phi_1 = .99$, (b) $\phi_1 = .95$, $\phi_1 = .9$, and $\phi_1 = .8$.

The graphs in Figure 3.20 show various levels of exponential damping associated with AR(1) models given by $X_t = \beta + \phi_1 X_{t-1} + a_t$, where ϕ_1 is positive and less than 1. As previously mentioned, Figure 3.20(a) shows that the autocorrelations for .99 die out very slowly. The correlation between adjacent values is $\rho_1 = .99$, while the autocorrelation between variables 25 time units apart is $\rho_{25} = .99^{25} = .78$ (still a substantial correlation). At lag 49, we have $\rho_{49} = .99^{49} = .61$. Consequently, in the previous example with a realization of length $n = 50$, the correlation between the observation made by Tester 1 and by Tester 50 would be .61. That is, all observations in the realization of length 50 were markedly related. The autocorrelation structure in Figure 3.20(b) shows that the first and last observations are only very slightly correlated. Finally, in Figures 3.20(c) and (d), the correlation dies out by lags 30 and 15, respectively. That is, if the testers had autocorrelations similar to those in Figure 3.20(d), then Tester 20's estimate would essentially be independent of that of Tester 1. Table 3.9 provides analogous results to those shown earlier for the $\phi_1 = .99$ case.

TABLE 3.9 $SE(\bar{X})$ for Realizations with $\rho_k = \phi_1^k$ and Realization Lengths Necessary to Obtain $SE(\bar{X}) = 1$, which is Associated with a White Noise Realization of Length $n = 50$

MODEL	$n = 50$	$n = 250$	$n = 500$	$n = 1000$	n FOR $SE(\bar{X}) = 1$
(a) $\phi_1 = .99$	6.53	5.02	4.00	2.99	9750
(b) $\phi_1 = .95$	5.00	2.68	1.94	1.38	1925
(c) $\phi_1 = .90$	3.93	1.91	1.37	0.97	940
(d) $\phi_1 = .80$	2.86	1.33	0.94	0.67	445

Even with autocorrelations such as those in Figure 3.20(d) that “die out” by lag 20, a realization of length $n = 445$ is “equivalent” in terms of standard error of \bar{X} for a white noise realization (random sample) of length $n = 50$. The take-away from this example is summarized in the following Key Point.

Key Points

- When analyzing a random sample, we are often able to consider $n = 30$ as a “large sample” for purposes of normal approximation, and so forth.
- The presence of autocorrelations, especially ones such as shown most dramatically in Figures 3.20(a) and (b), reduces the amount of information about the mean that is available in a realization.
- Time series data with autocorrelation structures such as those illustrated in Figure 3.20 may require n to be 100, 1000, 5000, or more to provide the same amount of information about the population mean that is available in a random sample of length $n = 50$.

3.3.2.2 Estimating the Variance

The estimate of the variance of a stationary process is $\hat{\gamma}_0$ where $\hat{\gamma}_k$ is defined in Section 3.3.2.3.

3.3.2.3 Estimating the Autocovariance and Autocorrelation

As mentioned, because of stationarity, $\gamma_k = E[(X_t - \mu_X)(X_{t+k} - \mu_X)]$ does not depend on t . Consequently, it seems reasonable to estimate γ_k from a single realization by “moving down the time axis and finding the average of all products of observations in the realization separated by k time units”. There are $n - k$ such pairs, and these are shown in Table 3.10.

TABLE 3.10 Lag k Data Pairs

LAG k PAIRS	
X_t	X_{t+k}
X_1	X_{1+k}
X_2	X_{2+k}
\vdots	\vdots
X_{n-k-1}	X_{n-1}
X_{n-k}	X_n



QR 3.9 Sample Autocovariance and Autocorrelation

The estimator or γ_k is denoted by $\hat{\gamma}_k$, and is defined by

$$\begin{aligned}\hat{\gamma}_k &= \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}), \quad 0 \leq k \leq n \\ &= 0, \quad k \geq n \\ &= \hat{\gamma}_{-k}, \quad k < 0\end{aligned}\tag{3.10}$$

Notes:

- Viewing Table 3.10 makes it easy to see that if $k = n - 1$, there is only one pair of data values that differ by $n - 1$ time units, namely X_1 and X_n . In this case, the “sum” in estimator $\hat{\gamma}_{n-1}$ has only a single term: $\hat{\gamma}_{n-1} = \frac{1}{n}(X_1 - \bar{X})(X_n - \bar{X})$. Consequently, the estimator of γ_{n-1} would be expected to

be of poorer quality than, for example, the estimator of γ_1 which is estimated using an “average” of $n - 1$ cross-products of data pairs separated by one time unit.

- (2) From Figure 3.20, we see that model-based autocorrelations damp toward zero (at differing rates). Because the autocorrelations are simply standardized versions of the autocovariances, it follows that both $\rho_k \rightarrow 0$ and $\gamma_k \rightarrow 0$ as n gets large.²⁶
- (3) Using (3.10) it follows that

$$\hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})^2 \quad (= \hat{\sigma}^2). \quad (3.11)$$

- (4) The estimator of the autocorrelation, ρ_k is given by

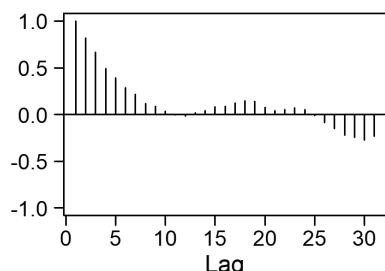
$$\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0. \quad (3.12)$$

This estimator is called the *sample autocorrelation*. From examination of (3.10), it is clear that $\hat{\gamma}_k$ (and $\hat{\rho}_k$) values will tend to be “small” when k is large relative to n . For example, in Note (1) above we showed that $\hat{\gamma}_{n-1}$ is one term divided by n . If n is large, then $\hat{\gamma}_{n-1}$ (and $\hat{\rho}_{n-1}$) are likely to be small, which mimics the behavior of the true values.

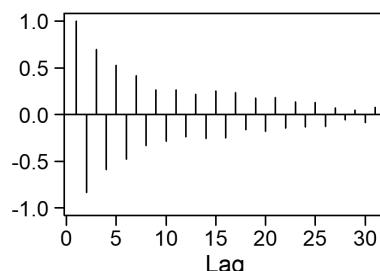
3.3.2.4 Plotting Sample Autocorrelations

In Figures 3.17 and 3.18 we showed plots of four realizations from autoregressive time series models and their associated model-based (theoretical) autocorrelations. In the discussion regarding Figures 3.17 and 3.18, we described how patterns in the autocorrelations provide information about the basic nature of realizations from time series models. Of course, given a set of data, we will not know the underlying model, so we will need to estimate the autocorrelations. Figure 3.21 shows plots of the sample autocorrelations, $\hat{\rho}_k, k = 0, \dots, 40$ calculated from the four realizations in Figure 3.17. These sample autocorrelations are estimates of the true autocorrelations, shown in Figure 3.18, and they approximate the behavior of the true autocorrelations. In particular, Figures 3.18(d) and 3.21(d) are very similar and show a damped sinusoidal behavior with a period of about 15. Figure 3.21(b) has a damped oscillating behavior similar to Figure 3.18(b) but the damping is more extreme in the true autocorrelations. In Figure 3.21(a) we note that there is correlation among the $\hat{\rho}_k$ values. For example, notice that when a value of $\hat{\rho}_k$ tends to underestimate ρ_k , values of $\hat{\rho}_j$ for j “near” k also tend to underestimate the true value. This can produce a cyclic behavior in sample autocorrelation plots when no such behavior is present in the plot of ρ_k or the data. Also, it should be noted that sample autocorrelations for a white noise realization will of course not be exactly equal to zero. In Sections 6.1.2 and 9.1.1 we will discuss tests for white noise that will help conclude whether sample autocorrelations such as those shown in Figure 3.21(c) are larger in magnitude than would be expected for a realization from a white noise process.

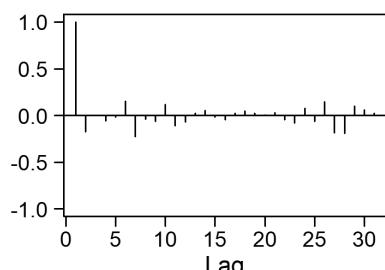
²⁶ Remember that in time series analysis, “large” n may be several thousand time units.



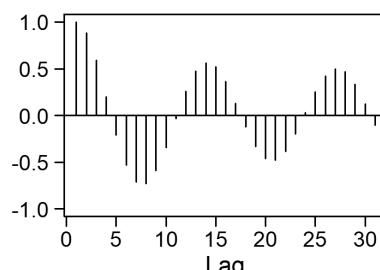
(a) Autocorrelations: Time Series 1



(b) Autocorrelations: Time Series 2

QR 3.10 Finding the
ACF in R

(c) Autocorrelations: Time Series 3



(d) Autocorrelations: Time Series 4

FIGURE 3.21 Sample autocorrelations, $\hat{\rho}_k$, calculated from realizations in Figure 3.17 (a), (b), (c), and (d), respectively.

Key Points

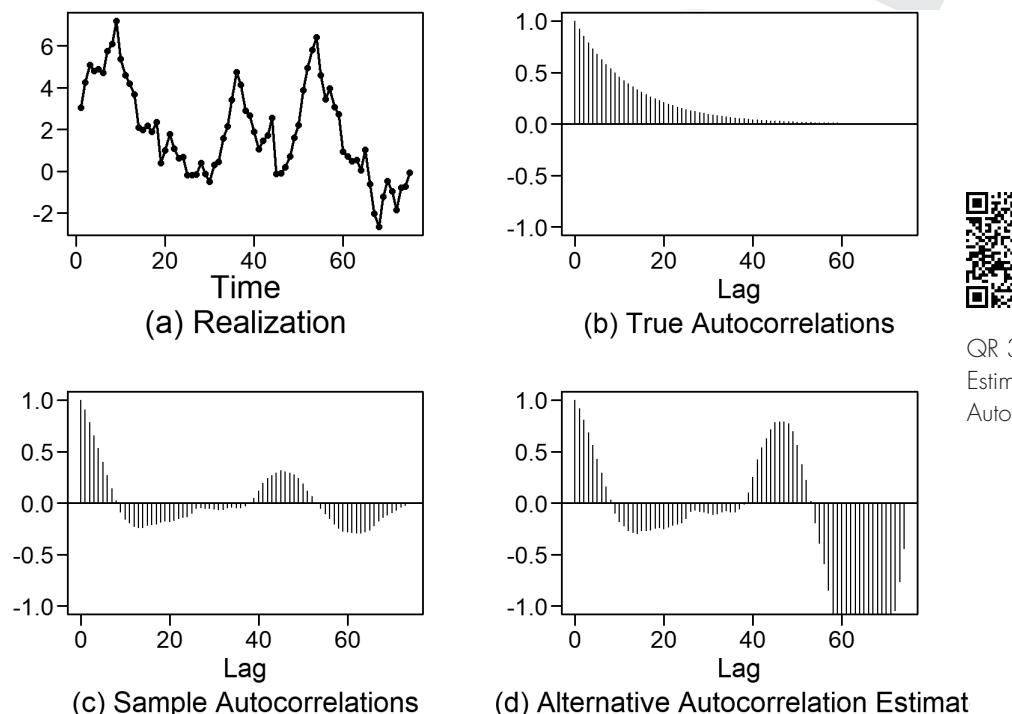
- Given a realization of length n , the estimators of γ_k and ρ_k are 0 for $k \geq n$.
- $-1 \leq \hat{\rho}_k \leq 1$
- The $\hat{\rho}_k$ s are referred to as the *sample autocorrelations*.
- Model-based or “true” autocorrelations such as those in Figure 3.18 and 3.20 are only available when we know the “inside information” concerning the model used to generate the realization.
- In practice, given a set of time series data such as sales data or crude oil price data, we will not know the “true” autocorrelations, but will only be able to calculate the sample autocorrelations such as those in Figure 3.21.
- An analysis of a set of time series data almost always begins by examining plots of the data and the sample autocorrelations.

Example 3.7 Why we use $\hat{\rho}_k$ to estimate ρ_k

You may have noticed that the estimator $\hat{\gamma}_k$ in (3.10) always has a divisor of n regardless of the number of terms in the sum. This seems odd; admittedly, a more natural estimate of γ_k would have been to replace μ_x by \bar{X} and find the *average* of the $n-k$ (instead of only n) cross products

$$\begin{aligned}\tilde{\gamma}_k &= \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}), \quad 0 \leq k < n \\ &= 0, \quad k \geq n \\ &= \tilde{\gamma}_{-k}, \quad k < 0.\end{aligned}\tag{3.13}$$

Then, the corresponding autocorrelation estimator would be $\tilde{\rho}_k = \tilde{\gamma}_k / \tilde{\gamma}_0$. Figures 3.22(a), (b), and (c) show a realization of length $n = 75$, the true autocorrelations associated with the model that generated the data in Figure 3.22(a), and the sample autocorrelations (defined by (3.10)) up to lag $k = 74$, respectively. We see that the true autocorrelations damp fairly rapidly and that the sample autocorrelations damp somewhat more rapidly and then “hover” around zero. In Figure 3.22(d), we plot the alternatively proposed autocorrelation estimates, $\tilde{\rho}_k$, based on (3.13). In the plot, we see that $\hat{\rho}_k$ and $\tilde{\rho}_k$ are similar for small k , but as k gets closer to n , the sample autocorrelations, $\hat{\rho}_k$, begin to damp toward zero while the estimates, $\tilde{\rho}_k$, become very erratic. Even more disturbing is the fact that nearly all $\tilde{\rho}_k$ for lag $k = 57$ and above are less than -1 . This is particularly concerning because correlations and autocorrelations must fall between -1 and 1 . For these reasons, the alternate formula $\tilde{\rho}_k$, which is based on an autocovariance $\tilde{\gamma}$ that divides by $n - k$, is not used to estimate $\tilde{\rho}_k$.



QR 3.11 Two Estimates of Autocovariance

FIGURE 3.22 (a) Realizations from a time series model, (b) and (c) theoretical (ρ_k) and sample ($\hat{\rho}_k$) autocorrelations, respectively, and (d) sample autocorrelations using proposed estimates, $\tilde{\rho}_k$.

3.4 CONCLUDING REMARKS

In this chapter, several pivotal concepts which are influential in the analysis of time series data have been introduced. In particular, central to the understanding of time series methodology is the dependence across time that is inherent in these data. Comparisons and contrasts were made between dependent and independent data structures, and interesting relationships were explored. Also, of relevance to future chapters are the autocovariance and autocorrelation, which are the time series versions of the familiar covariance and correlation, respectively. These terms were defined and considered both in mathematical form and by graph. A significant amount of time series methodology has been developed assuming that data are

stationary, which will become apparent very soon! The term “stationary” was defined and illustrated in this chapter, and stationary datasets were compared to those for which data are nonstationary. Several examples of both simulated and actual data were presented to illustrate this wide array of important topics.

APPENDIX 3A

In this appendix we obtain the mean, variance, and standard deviation associated with the random process of rolling two dice and finding the sum of the rolls. Table 3A.1 shows the 36 equally likely possible outcomes when we roll a white die and a black die and find the sum.

TABLE 3A.1 Probability Distribution for the Sum of the Rolls of Two Dice

	•	••	•••	••••	•••••	••••••
•	2	3	4	5	6	7
••	3	4	5	6	7	8
•••	4	5	6	7	8	9
••••	5	6	7	8	9	10
•••••	6	7	8	9	10	11
••••••	7	8	9	10	11	12

Consequently, we see that the probability of a sum of 12 is $1/36$ (which can only happen if the white die and black die both land on 6) while the probability of 6 is $5/36$, and so forth. Table 3A.2 shows the possible values and associated probabilities.

TABLE 3A.2 Possible Outcomes and Probabilities for the Random Process of Rolling Two Dice and Finding the Sum

SUM	PROBABILITY
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$
9	$4/36$
10	$3/36$
11	$2/36$
12	$1/36$

This is an example of a discrete probability distribution in which there are 11 possible outcomes. Suppose we have a discrete probability distribution with possible values y_1, y_2, \dots, y_k and the probability

of y_i is p_i , $i = 1, \dots, k$. Then from introductory statistics, we know that the mean is given by $\mu = \sum_{i=1}^k p_i y_i$

and the variance is given by $\sigma^2 = \sum_{i=1}^k p_i (y_i - \mu)^2$. So, for the “sum” distribution we have

$$\begin{aligned}\mu &= \frac{1}{36}(2) + \frac{2}{36}(3) + \frac{3}{36}(4) + \frac{4}{36}(5) + \frac{5}{36}(6) + \frac{6}{36}(7) + \frac{5}{36}(8) + \frac{4}{36}(9) + \frac{3}{36}(10) + \frac{2}{36}(11) + \frac{1}{36}(12) \\ &= 7\end{aligned}$$

This seems reasonable because 7 is the most likely sum and the distribution is symmetric about 7. The variance is

$$\begin{aligned}\mu &= \frac{1}{36}(2-7)^2 + \frac{2}{36}(3-7)^2 + \frac{3}{36}(4-7)^2 + \frac{4}{36}(5-7)^2 + \frac{5}{36}(6-7)^2 + \frac{6}{36}(7-7)^2 + \frac{5}{36}(8-7)^2 \\ &\quad + \frac{4}{36}(9-7)^2 + \frac{3}{36}(10-7)^2 + \frac{2}{36}(11-7)^2 + \frac{1}{36}(12-7)^2 \\ &= 5.833,\end{aligned}$$

and the standard deviation is $\sigma = 2.415$.

APPENDIX 3B

R COMMANDS FOR CHAPTER 3

The R package *tswge* is available on GitHub. Some of the commands listed below are Base R commands that do not require the installation of a special library. These will be noted as Base R commands.

1. Base R Computation Routines used in Chapter 3

- (a) **mean(x)**: calculates the mean of a time series realization where **x** is a vector containing the time series realization.

Example: The basic R command

```
data(AirPassengers); mean.air=mean(AirPassengers)
```

calculates the sample mean of the **AirPassengers** data in *tswge* dataset **AirPassengers** and places it in variable **mean.air**. In this case **mean.air** is 280.2986.

- (b) **var(x)**: calculates the variance of a time series realization where **x** is a vector containing the time series realization.

Example: The basic R command

```
data(AirPassengers); var.air=var(AirPassengers)
```

calculates the sample variance (using denominator $n-1$) of the **AirPassengers** data and places it in variable **var.air**. In this case **var.air** is 14391.92.

- (c) **sd(x)** : calculates the standard deviation of a time series realization where **x** is a vector containing the time series realization.

Example: The basic R command

```
data(AirPassengers); sd.air=sd(AirPassengers)
```

calculates the sample standard deviation (using denominator $n - 1$) of the **AirPassengers** data in **tswge** data set **AirPassengers** and places it in variable **sd.air**. In this case **sd.air** is 119.9663, which is the square root of **var.air**.

- (d) **acf(x, lag.max, type, plot)** (Base R function): calculates (and optionally plots) the sample autocorrelations (or sample autocovariances and sample variance) of a time series realization where

x is a vector containing the time series realization

lag.max is the maximum lag at which to calculate the sample autocorrelations (or autocovariances)

type='correlation' (default) outputs the sample autocorrelations ($\hat{\rho}_k$ in our notation)

='covariance' specifies the autocovariances

='partial' designates the partial autocorrelations (discussed in Section 6.1)

plot is a logical variable. **plot='TRUE'** (default) produces a plot of the sample autocorrelations (or autocovariances)

Example:

```
data(lynx); aut=acf(x=lynx, lag.max=25)
```

calculates and plots the first 25 sample autocorrelations for the dfw annual temperature data in **ts** object **dfw.yr** shown in Figure 1.19(b). The sample autocorrelations are placed in the vector **aut\$acf** (which has 26 elements counting **aut\$acf[1]=1**, i.e. the sample autocorrelation at lag 0). If **type='covariance'** is selected, then the vector **aut\$acf** contains the sample autocovariances, and **aut\$acf[1]** contains the sample variance (using the divisor n).

2. *tswge functions*

plotts.sample.wge(x,lag.max,M,arlimits,periodogram) : plots a realization, sample autocorrelations, and Parzen spectral density (both plotted in dB). Note: See Chapter 4 for discussion of the Parzen spectral density. (The periodogram is not discussed in this book.) The periodogram is plotted as an option. See Woodward et al. (2017).

x is a vector containing the time series realization **lag.max** is the maximum lag at which to plot the sample autocorrelations (default is 25) **M** (integer ≥ 0) specifies the truncation point for the Parzen window **M=0** (default) (or if no value for **M** specified) indicates default truncation point $M = 2\sqrt{n}$ **M>0** is a user specified truncation point, i.e. calculations will be based on the user supplied value of **M** **arlimits** (default=**FALSE**) is a logical variable specifying whether 95% limit lines will be included on sample autocorrelation plots **periodogram**(default=**FALSE**) specifies whether the periodogram will be plotted

Example:

```
data(AirPassengers); airlog=log(AirPassengers);
sp=plotts.sample.wge(x=airlog,lag.max=50)
```

plots the realization(**x=airlog**), sample autocorrelations (lags 0 to **lag.max=50**), and the Parzen spectral estimator (in dB) using truncation point $M = 2\sqrt{n}$. The vector **\$autplt** contains the sample autocorrelation values while **\$freq** and **\$pgram** contain the frequencies and Parzen-based spectral estimates in dB, respectively.

PROBLEMS

- 3.1 The following data are annual sales of a hypothetical company in millions of dollars:

Period	Sales
1	76
2	70
3	66
4	60
5	70
6	72
7	76
8	80

Compute by hand (i.e., calculator) the estimates $\hat{\gamma}_0$, $\hat{\gamma}_1$, $\hat{\rho}_0$, and $\hat{\rho}_1$.

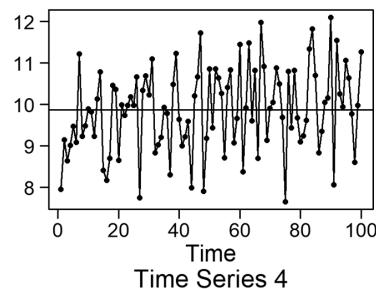
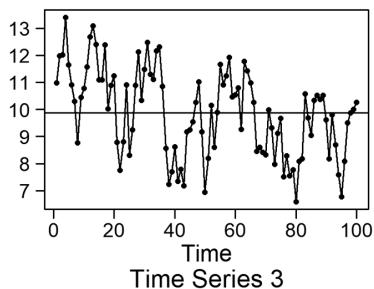
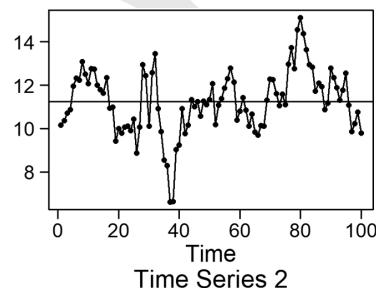
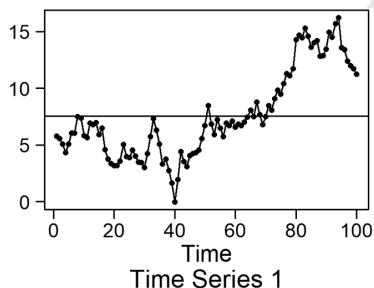
- 3.2 Figure 1.16(b) is a plot of the **tswge** dataset **WTICrude** which contains the monthly West Texas intermediate crude oil prices from January 1990 through December 2020. Figure 1.3 is a plot of the monthly average temperatures in the Dallas-Ft. Worth area from January 1900 through December 2020. These data are stored in **tswge** file **dfw.mon**. For each of these datasets, plot the realization and the sample autocorrelations. Explain how these plots describe (or fail to describe) the behavior in the data.
- 3.3 Dice-Rolling
- (A) The following R code can be used to repeat the random process of rolling two dice, finding the sum, and plotting the resulting data. This generates *another* set of 30 “sum-of-two-rolls” data and plots the new data in the format of Figure 3.5(c).

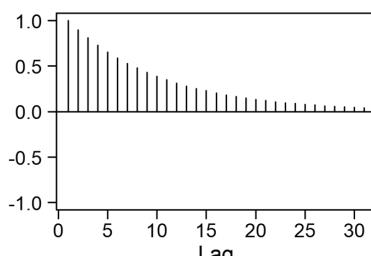
```
set.seed(8327)
x=rep(0,30)
for(i in 1:30) {
  roll=sample(1:6,2)
  x[i]=roll[1]+roll[2]}
plotts.wge(x)
abline(h=7)
```

Note: The command **roll=sample(1:6,2)** instructs R to randomly choose two numbers (**roll[1]** and **roll[2]**) between 1 and 6, inclusive. The following line adds the two rolls.

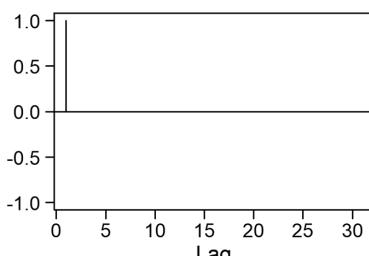
- (a) Using the above commands, plot the resulting 30 new “sum of rolls”.
 (b) Do these realizations have the same random appearance as Figure 3.5? Explain.

- (c) In the plot in (a) (using seed 8327), there are 11 instances in which the sum of the rolls is greater than the true mean of 7. In these situations, the next sum was
- above 7 _____ times
 - below 7 _____ times
 - equal to 7 _____ times
- (B) Change the seed from 8327 to another positive integer and repeat the process. Answer (a) and (b) for the data using the seed you selected. How many instances were there in which the sum of the rolls is greater than 7? In each of these situations answer (i)–(iii) above.
- (C) What do your results in (A) and (B) tell you about the ability to predict the sum at time $t+1$ if you know the sum at time t ?
- (D) Use **plots.sample.wge** to plot the datasets in (A) and (B) along with their sample autocorrelations. What do the autocorrelations say about the relationship among outcomes in the “sum of the roll” data? Note: The Parzen spectral plot will be discussed in Chapter 4.
- 3.4 Following are displayed two sets of figures, each containing four plots. The first set shows four realizations of length $n = 100$ each generated from a time series model. The sample mean is also plotted for each realization. The second set contains four autocorrelation functions based on the models used to generate the four realizations (in random order). Match each realization with its corresponding autocorrelation plot. Explain your answers.

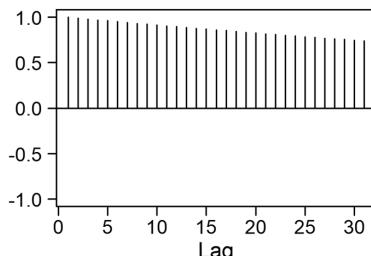




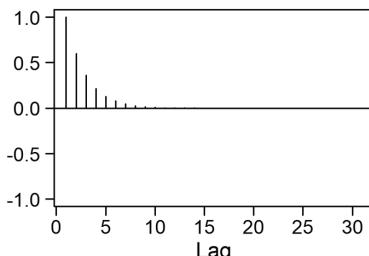
(a) Autocorrelations A



(b) Autocorrelations B

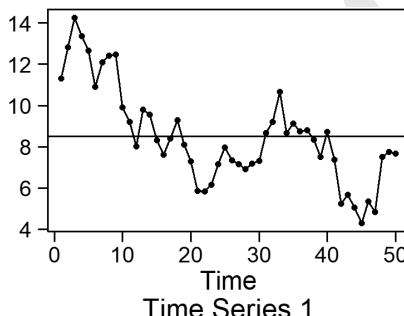


(c) Autocorrelations C

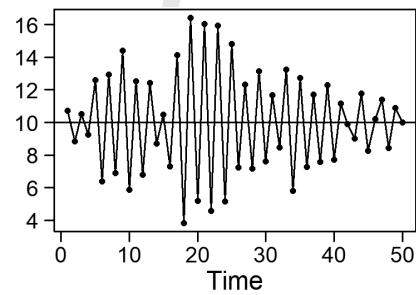


(d) Autocorrelations D

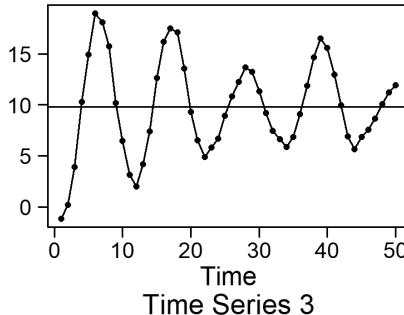
- 3.5 Following are displayed two sets of figures, each containing four plots. The first set shows four realizations of length $n = 50$ each generated from a time series model. The second set contains four autocorrelation functions based on the models used to generate the four realizations (in random order). Match each realization with its corresponding autocorrelation plot. Explain your answers.



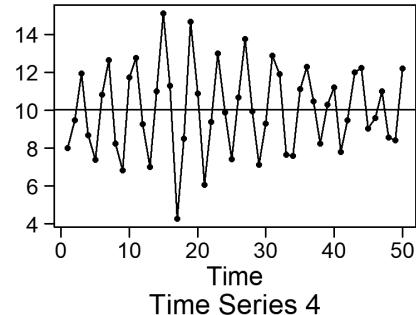
Time Series 1



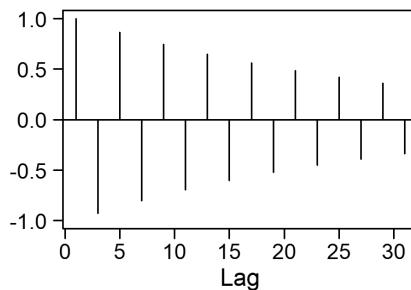
Time Series 2



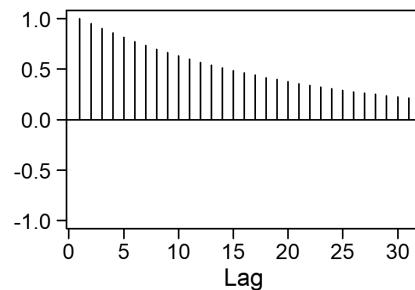
Time Series 3



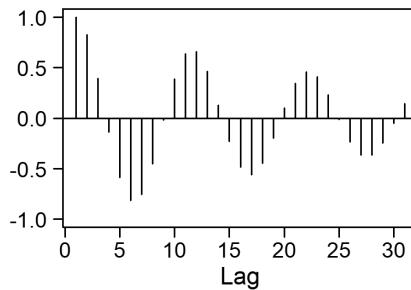
Time Series 4



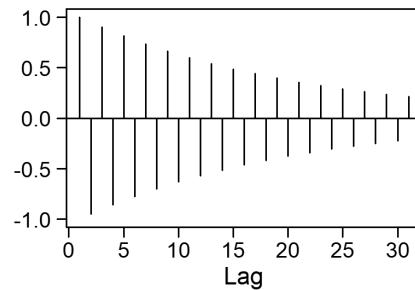
(a) Autocorrelations A



(b) Autocorrelations B

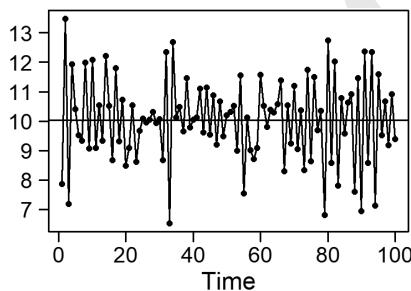


(c) Autocorrelations C

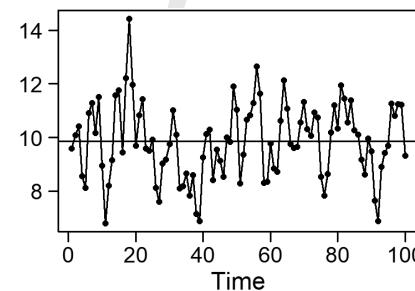


(d) Autocorrelations D

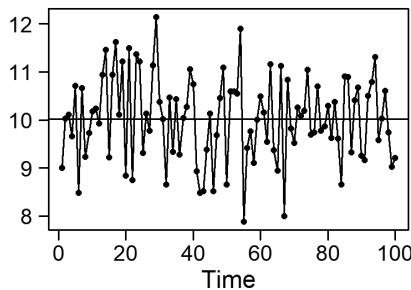
- 3.6 The following plots are realizations of length 50. One of these realizations is white noise and the other three are not. Which is the white noise realization?



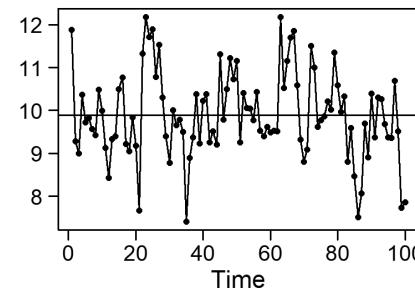
(a) Realization: Time Series 1



(b) Realization: Time Series 2



(c) Realization: Time Series 3



(d) Realization: Time Series 4

- 3.7 Find a time series dataset (on the internet, etc.) not discussed in the book or in class. For this time series:
- Plot the time series realization and sample autocorrelations.
 - Describe the behavior (wandering, periodic, etc.) in the realization and how that behavior manifests itself in the sample autocorrelations.

Hint: Use the *tswge* function `plotts.sample.wge`.

- 3.8 Show that the three conditions of stationarity are satisfied by a white noise process.
- 3.9 Recall that Figure 3.3(d) is a plot in which there is no discernible linear relationship between X and Y , and “knowing x ” does not help in predicting y . Discuss what this means in terms of $(x_i - \bar{x})$, $(y_i - \bar{y})$, $(x_i - \bar{x})(y_i - \bar{y})$, $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, and thus r_{XY} .
- 3.10 Figure 3.3(f) is a plot in which there is a relationship between X and Y , but it is not a linear relationship. However, in this case, “knowing x ” does help in predicting y , although the correlation coefficient is very close to $r_{XY} = 0$. Explain the reason for the near zero correlation by discussing the terms $(x_i - \bar{x})$, $(y_i - \bar{y})$, $(x_i - \bar{x})(y_i - \bar{y})$, and $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$.

PROOF