統計學 (二)

Anderson's Statistics for Business & Economics (14/E)

Chapter 17: Time Series Analysis and Forecasting

上課時間地點: 二 D56, 資訊 140306
授課教師: 吳漢銘 (國立政治大學統計學系副教授)
教學網站: http://www.hmwu.idv.tw

系級: ________________ 學號: ________________ 姓名: ________________

17.1 Time Series Patterns

1. **time series**: A ____________ is a sequence of observations on a variable measured at successive points in time or over successive periods of time.

2. The measurements may be taken every hour, day, week, month, or year, or at any other ________________. (this textbook limits the discussion to time series in which the values of the series are recorded at equal intervals)

3. The __________ of the data is an important factor in understanding how the time series has behaved in the _______. If such behavior can be expected to continue in the _________, we can use the past pattern to guide us in selecting an appropriate ____________ method.

4. A ________________ is a graphical presentation of the relationship between time and the time series variable; _______ is on the horizontal axis and the time series _________ are shown on the vertical axis. A time series plot is useful to identify the underlying pattern in the data.

5. Some of the common types of data patterns that can be identified when examining a time series plot: horizontal pattern, trend pattern, seasonal pattern, trend and seasonal pattern, and cyclical pattern.

Horizontal Pattern

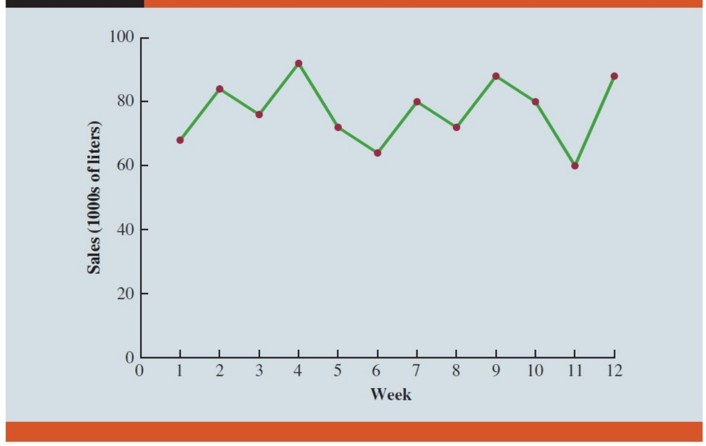
1. A horizontal pattern exists when the data ____________ around a _________________.

2. Example (Table 17.1) (Figure 17.1) These data show the number of gallons of gasoline sold by a gasoline distributor in Bennington, Vermont, over the past 12 weeks.

TABLE 17.1
Gasoline Sales Time Series

Week	Sales (1000s of gallons)
1	17
2	21
3	19
4	23
5	18
6	16
7	20
8	18
9	22
10	20
11	15
12	22

FIGURE 17.1 Gasoline Sales Time Series Plot



The average value or mean for this time series is 19.25 gallons (1000s) per week. Although ____________________ is present, we would say that these data follow a horizontal pattern.

3. The term _____________ time series is used to denote a time series whose statistical properties are _____________________.

4. In particular this means that
   (a) The process generating the data has a _________________.
   (b) The variability of the time series is ____________ over time.

5. A time series plot for a stationary time series will always exhibit a ____________________. But simply observing a horizontal pattern is not sufficient evidence to conclude that the time series is stationary.

6. More advanced texts on forecasting discuss procedures for determining if a time series is stationary and provide methods for transforming a time series that is not stationary into a stationary series.

7. Changes in business conditions can often result in a time series that has a horizontal pattern __________ to a new level.

   (a) $\boxed{\text{Example}}$ For instance, suppose the gasoline distributor signs a contract with the Vermont State Police to provide gasoline for state police cars located in southern Vermont. With this new contract, the distributor expects to see a major increase in weekly sales starting in week 13.

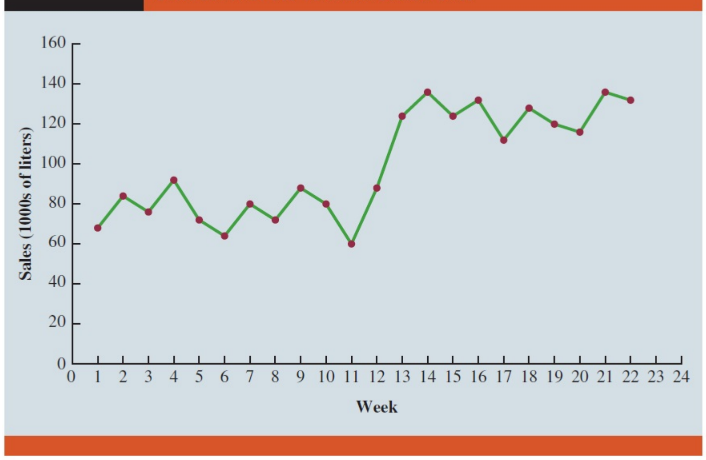
   (b) (Table 17.2) The number of gallons of gasoline sold for the original time series and for the 10 weeks after signing the new contract.

TABLE 17.2

Gasoline Sales Time Series After Obtaining the Contract with the Vermont State Police

Week	Sales (1000s of liters)
1	68
2	84
3	76
4	92
5	72
6	64
7	80
8	72
9	88
10	80
11	60
12	88
13	124
14	136
15	124
16	132
17	112
18	128
19	120
20	116
21	136
22	132

FIGURE 17.2 Gasoline Sales Time Series Plot After Obtaining the Contract with the Vermont State Police



   (c) (Figure 17.2) Note the increased level of the time series beginning in week 13. This change in the level of the time series makes it more __________ to choose an appropriate forecasting method.

8. Selecting a forecasting method that adapts well to ______________ of a time series is an important consideration in many practical applications.

Trend Pattern

1. Although time series data generally exhibit random fluctuations, a time series may also show gradual ______________________ to relatively higher or lower values over a __________ period of time.

2. If a time series plot exhibits this type of behavior, we say that a ________________ exists.

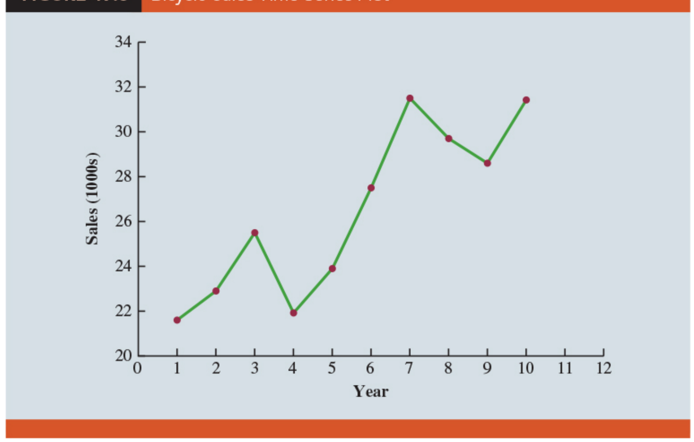
3. A trend is usually the result of ___________________ such as population increases or decreases, changing demographic characteristics of the population, technology, and/or consumer preferences.

4. $\boxed{\text{Example}}$ (Table 17.3) (Figure 17.3) Consider the time series of bicycle sales for a particular manufacturer over the past 10 years.

TABLE 17.3
Bicycle Sales Time Series

Year	Sales (1000s)
1	21.6
2	22.9
3	25.5
4	21.9
5	23.9
6	27.5
7	31.5
8	29.7
9	28.6
10	31.4

FIGURE 17.3 Bicycle Sales Time Series Plot



Visual inspection of the time series plot shows some up and down movement over the past 10 years, but the time series also seems to have a _________________________ or ________________. The trend for the bicycle sales time series appears to be _________ and increasing over time.

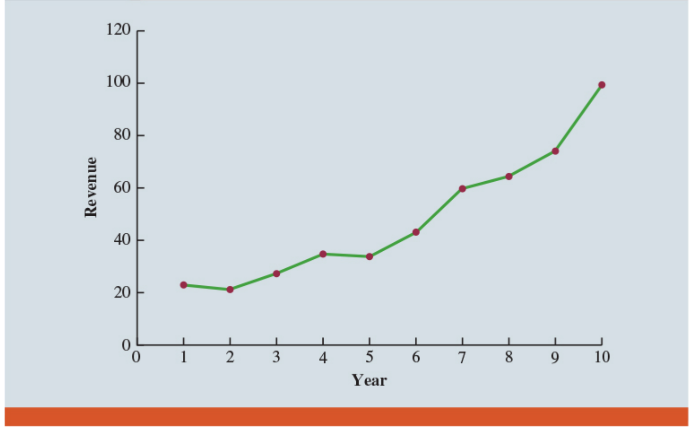
5. $\boxed{\text{Example}}$ (Table 17.4) (Figure 17.4) The data show the sales for a cholesterol drug since the company won FDA approval for it 10 years ago.

TABLE 17.4

Cholesterol Revenue Time Series ($Millions)

Year	Revenue
1	23.1
2	21.3
3	27.4
4	34.6
5	33.8
6	43.2
7	59.5
8	64.4
9	74.2
10	99.3

FIGURE 17.4 Cholesterol Revenue Times Series Plot ($Millions)



The time series increases in a nonlinear fashion; that is, the ________________ of revenue does not increase by a constant amount from one year to the next. In fact, the revenue appears to be growing in an ______________ fashion.

6. Exponential relationships such as this are appropriate when the percentage change from one period to the next is relatively ____________.

Seasonal Pattern

1. The trend of a time series can be identified by analyzing multiyear movements in ________________. Seasonal patterns are recognized by seeing the ________________________ over successive periods of time.

2. $\boxed{\text{Example}}$ For example, a manufacturer of swimming pools expects low sales activity in the fall and winter months, with peak sales in the spring and summer months. Manufacturers of snow removal equipment and heavy clothing, however, expect just the opposite yearly pattern.

3. The pattern for a time series plot that exhibits a repeating pattern over a one-year period due to seasonal influences is called a ____________ pattern.

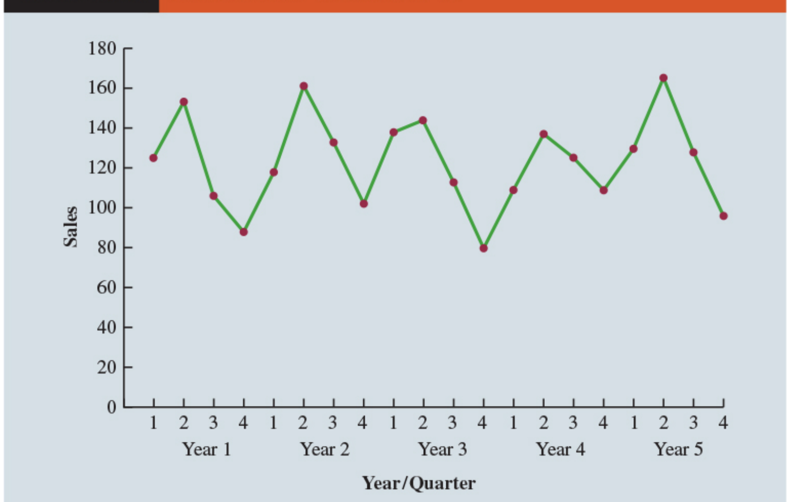
4. $\boxed{\text{Example}}$ Daily traffic volume shows within-the-day "seasonal" behavior, with peak levels occurring during rush hours, moderate flow during the rest of the day and early evening, and light flow from midnight to early morning.

5. Example (Table 17.5) (Figure 17.5) As an example of a seasonal pattern, consider the number of umbrellas sold at a clothing store over the past five years.

TABLE 17.5 Umbrella Sales Time Series

Year	Quarter	Sales
1	1	125
	2	153
	3	106
	4	88
2	1	118
	2	161
	3	133
	4	102
3	1	138
	2	144
	3	113
	4	80
4	1	109
	2	137
	3	125
	4	109
5	1	130
	2	165
	3	128
	4	96

FIGURE 17.5 Umbrella Sales Time Series Plot



The time series plot does not indicate any ______________ in sales. The data follow a ______________ pattern. But closer inspection of the time series plot reveals a ______________ in the data. That is, the first and third quarters have moderate sales, the second quarter has the highest sales, and the fourth quarter tends to have the lowest sales volume. Thus, we would conclude that a ______________ pattern is present.

Trend and Seasonal Pattern

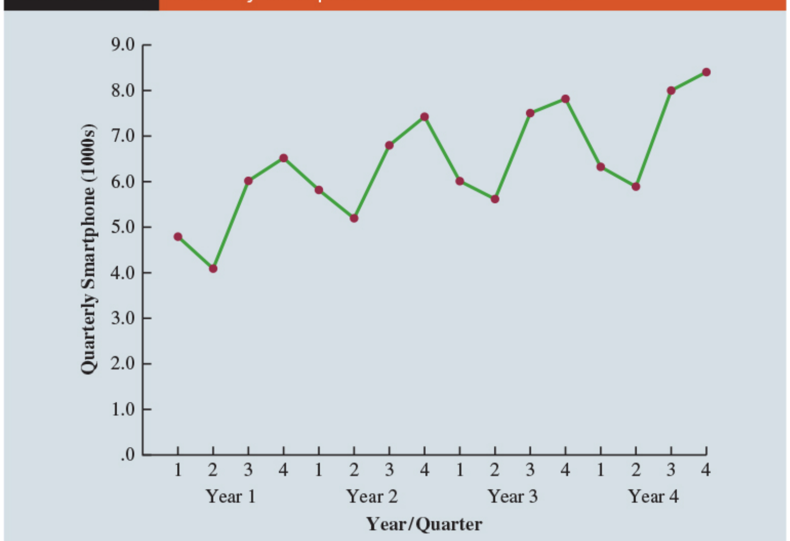
1. Some time series include a combination of a trend and seasonal pattern.

2. $\boxed{\text{Example}}$ (Table 17.6) (Figure 17.6) The smartphone sales for a particular manufacturer over the past four years.

TABLE 17.6 Quarterly Smartphone Sales Time Series

Year	Quarter	Sales (1000s)
1	1	4.8
	2	4.1
	3	6.0
	4	6.5
2	1	5.8
	2	5.2
	3	6.8
	4	7.4
3	1	6.0
	2	5.6
	3	7.5
	4	7.8
4	1	6.3
	2	5.9
	3	8.0
	4	8.4

FIGURE 17.6 Quarterly Smartphone Sales Time Series Plot



3. Clearly, an increasing trend is present.

4. But, Figure 17.6 also indicates that sales are lowest in the second quarter of each year and increase in quarters 3 and 4. Thus, we conclude that a seasonal pattern also exists for smartphone sales.

5. In such cases we need to use a forecasting method that has the capability to deal with both ______________________.

Cyclical Pattern

1. A __________ pattern exists if the time series plot shows an alternating sequence of points below and above the ____________ lasting more than one year.

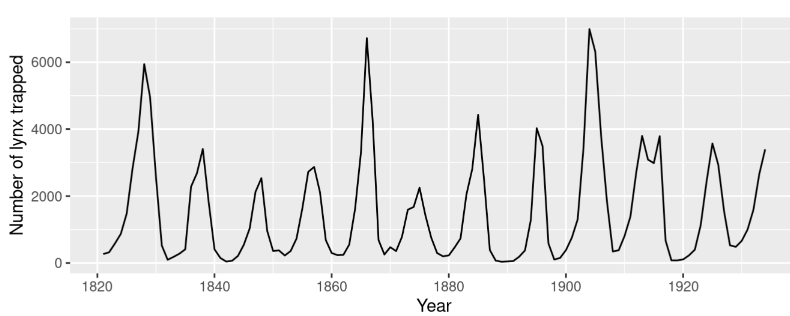
2. Often, the cyclical component of a time series is due to _________________________.

3. Example For example, periods of moderate inflation followed by periods of rapid inflation can lead to time series that alternate __________________ a generally increasing trend line (e.g., a time series for housing costs).

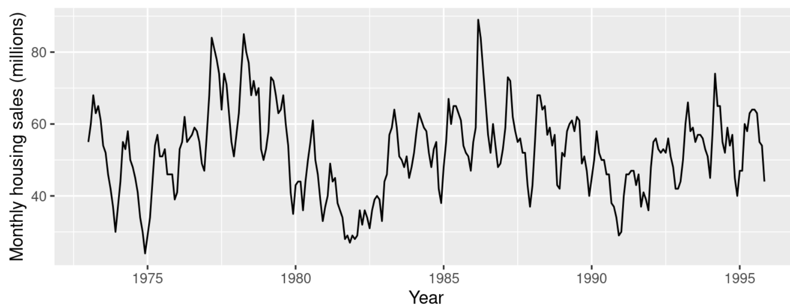
4. A cyclical pattern repeats with some ___________________________. Cyclical patterns differ from seasonal patterns in that cyclical patterns occur over multiple years, whereas seasonal patterns occur _________________.

5. More Example `https://robjhyndman.com/hyndsight/cyclicts/`

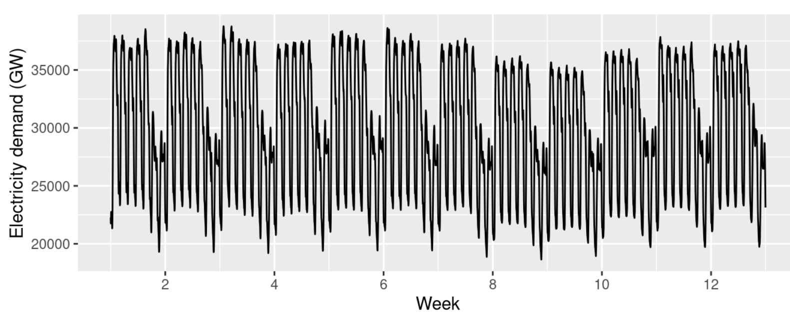
   (a) The plot shows the famous Canadian lynx (山貓) data –the number of lynx trapped each year in the McKenzie (麥肯錫) river district of northwest Canada (1821-1934). These show clear aperiodic (非週期性的) population cycles of approximately 10 years. The cycles are not of fixed length –some last 8 or 9 years and others last longer than 10 years.



(b) The plot shows the monthly sales of new one-family houses sold in the USA (1973-1995). There is strong seasonality within each year, as well as some strong cyclic behaviour with period about 6 − 10 years.



(c) The plot shows half-hourly electricity demand in England and Wales from Monday 5 June 2000 to Sunday 27 August 2000. Here there are two types of seasonality −a ________ pattern and a __________ pattern. If we collected data over a few years, we would also see there is an __________ pattern. If we collected data over a few decades, we may even see a longer cyclic pattern.



6. Business cycles are extremely difficult, if not impossible, to forecast. As a result, cyclical effects are often combined with long-term trend effects and referred to as ____________________.

Selecting a Forecasting Method

1. The underlying pattern in the time series is an important factor in selecting a forecasting method. Thus, a __________________ should be one of the first things developed when trying to determine which forecasting method to use.

2. The next two sections illustrate methods that can be used in situations where the underlying pattern is horizontal; in other words, no trend or seasonal effects are present. We then consider methods appropriate when trend and/or seasonality are present in the data.

17.2 Forecast Accuracy

1. The simplest of all the forecasting methods (a ____________________): an approach that uses the ___________ week's sales volume as the forecast for the next week.

2. (Table 17.7) The distributor sold 68 thousand gallons of gasoline in week 1; this value is used as the forecast for week 2. Next, we use 84, the actual value of sales in week 2, as the forecast for week 3, and so on.

TABLE 17.7 Computing Forecasts and Measures of Forecast Accuracy Using the Most Recent Value as the Forecast for the Next Period

Week	Time Series Value	Forecast	Forecast Error	Absolute Value of Forecast Error	Squared Forecast Error	Percentage Error	Absolute Value of Percentage Error
1	68						
2	84	68	16	16	256	19.05	19.05
3	76	84	−8	8	64	−10.53	10.53
4	92	76	16	16	256	17.39	17.39
5	72	92	−20	20	400	−27.78	27.78
6	64	72	−8	8	64	−12.50	12.50
7	80	64	16	16	256	20.00	20.00
8	72	80	−8	8	64	−11.11	11.11
9	88	72	16	16	256	18.18	18.18
10	80	88	−8	8	64	−10.00	10.00
11	60	80	−20	20	400	−33.33	33.33
12	88	60	28	28	784	31.82	31.82
		Totals	20	164	2864	1.19	211.69

3. The key concept associated with measuring forecast accuracy is ______________, defined as

____

(a) $\boxed{\text{Example}}$ Because the distributor actually sold 84 thousand gallons of gasoline in week 2 and the forecast, using the sales volume in week 1, was 68 thousand gallons, the forecast error in week 2 is

$$\text{Forecast Error in week 2} = \underline{\hspace{4cm}}$$

(b) The fact that the forecast error is positive indicates that in week 2 the forecasting method ________________ the actual value of sales. Next, we use 84, the actual value of sales in week 2, as the forecast for week 3. Since the actual value of sales in week 3 is 76, the forecast error for week 3 is $76-84 = -8$. In this case, the negative forecast error indicates that in week 3 the forecast ________________ the actual value.

(c) A simple measure of forecast accuracy is the ________________________________. Table 17.7 shows that the sum of the forecast errors for the gasoline sales time series is 20; thus, the mean or average forecast error is ________________.

(d) Because positive and negative forecast errors tend to _________ one another, the mean error is likely to be small; thus, the mean error is not a very useful measure of forecast accuracy.

4. The ______________________, denoted _________, is the average of the absolute values of the forecast errors.

   (a) MAE is a measure of forecast accuracy that avoids the problem of positive and negative forecast errors offsetting one another.

   (b) (Table 17.7) the sum of the absolute values of the forecast errors is 164:

$$\text{MAE} = \text{average of the absolute value of forecast errors} = \underline{\hspace{4cm}}$$

5. Another measure that avoids the problem of positive and negative forecast errors offsetting each other is obtained by computing the average of the ___________ forecast errors (______________________, is denoted _________.)

$$\text{MSE} = \text{average of the sum of squared forecast errors} = \underline{\hspace{4.5cm}}$$

6. The size of MAE and MSE depends upon the ___________________. As a result, it is difficult to make comparisons for different time intervals, such as comparing

a method of forecasting monthly gasoline sales to a method of forecasting weekly sales, or to make comparisons across different time series.

7. The ________________________________, denoted __________, is a percentage error corresponding to the ___________ of 84 in week 2 is computed by dividing the _______________ in week 2 by the ______________ in week 2 and multiplying the result by _______.

   (a) For week 2 the percentage error is computed as follows:

   $$\text{Percentage error for week 2} = \underline{\hspace{4cm}} = 16/84 \times (100) = 19.05\%$$

   Thus, the forecast error for week 2 is 19.05% of the observed value in week 2.

   (b) The sum of the absolute values of the percentage errors is 211.69:

   $$\begin{aligned}\text{MAPE} &= \text{average of the absolute value of percentage forecast errors}\\ &= \underline{\hspace{5cm}}\end{aligned}$$

8. Summarizing, using the naive (most recent observation) forecasting method, we obtained the following measures of forecast accuracy:

   $$\text{MAE} = 3.73, \quad \text{MSE} = 16.27, \quad \text{MAPE} = 19.24\%$$

9. These measures of forecast accuracy simply measure how well the forecasting method is able to _________________________ of the time series.

10. Suppose we want to forecast sales for a ____________________, such as week 13. In this case the forecast for week 13 is 88, the actual value of the time series in week 12. Is this an accurate estimate of sales for week 13? Unfortunately, there is no way to address the issue of ____________ associated with forecasts for ____________________. But, if we select a forecasting method that works well for the historical data, and we think that the historical pattern will continue into the future, we should obtain results that will ultimately be shown to be good.

11. (Table 17.8) Suppose we use the ________________________________ available as the forecast for the next period. We begin by developing a forecast for week 2. Since there is only one historical value available prior to week 2, the forecast for week 2

is just the time series value in week 1; thus, the forecast for week 2 is 84 thousand gallons of gasoline. To compute the forecast for week 3, we take the average of the sales values in weeks 1 and 2. Thus,

$$\text{Forecast for week } 3 = \underline{\hspace{3cm}}$$

TABLE 17.8 Computing Forecasts and Measures of Forecast Accuracy Using the Average of All the Historical Data as the Forecast for the Next Period

Week	Time Series Value	Forecast	Forecast Error	Absolute Value of Forecast Error	Squared Forecast Error	Percentage Error	Absolute Value of Percentage Error
1	68						
2	84	68.00	16.00	16.00	256.00	19.05	19.05
3	76	76.00	.00	.00	.00	.00	.00
4	92	76.00	16.00	16.00	256.00	17.39	17.39
5	72	80.00	−8.00	8.00	64.00	−11.11	11.11
6	64	78.40	−14.40	14.40	207.36	−22.50	22.50
7	80	76.00	4.00	4.00	16.00	5.00	5.00
8	72	76.57	−4.57	4.57	20.90	−6.35	6.35
9	88	76.00	12.00	12.00	144.00	13.64	13.64
10	80	77.33	2.67	2.67	7.11	3.33	3.33
11	60	77.60	−17.60	17.60	309.76	−29.33	29.33
12	88	76.00	12.00	12.00	144.00	13.64	13.64
		Totals	18.10	107.24	1425.13	2.76	141.34

12. Comparing the values of MAE, MSE, and MAPE for each method:

	Naive Method	Average of Past Values
MAE	14.91	9.75
MSE	260.36	129.56
MAPE	19.24%	12.85%

13. For every measure, the average of past values provides ________________ forecasts than using the most recent observation as the forecast for the next period.

14. In general, if the underlying time series is _____________, the average of all the historical data will always provide the best results.

    (a) (Recall Table 17.2) But suppose that the underlying time series is not stationary. Note the ________________ in week 13 for the resulting time series. When a shift to a new level like this occurs, it takes a long time for the forecasting method that uses the average of all the historical data to adjust to the new level of the time series.

(b) In this case, the simple naive method adjusts very rapidly to the change in level because it uses the most recent observation available as the forecast.

(c) Measures of forecast accuracy are important factors in comparing different forecasting methods, but we have to be careful not to rely upon them too heavily.

(d) Good judgment and knowledge about business conditions that might affect the forecast also have to be carefully considered when selecting a method. And ______________________ is not the only consideration, especially if the time series is likely to change in the future.

☺ **EXERCISES 17.2**: 1, 4

17.3 Moving Averages and Exponential Smoothing

1. Three forecasting methods that are appropriate for a time series with a horizontal pattern: __________ averages, __________ moving averages, and __________ smoothing.

2. The objective of each of these methods is to smooth out the ______________ in the time series, they are referred to as __________ methods.

3. These methods are easy to use and generally provide a high level of __________ for short-range __________, such as a forecast for the next time period.

Moving Averages

1. (**Moving Average Forecast of Order** k) The moving averages method uses the average of the most recent k data values in the time series as the forecast for the next period:

$$F_{t+1} = \frac{\sum(\text{most recent } k \text{ data values})}{k} = \underline{\hspace{6cm}} \tag{17.1}$$

where F_{t+1} is the forecast of the times series for period $t+1$ and Y_t is the actual value of the time series in period t.

2. The average will change, or move, as new observations become available.

   (a) To use moving averages to forecast a time series, we must first select the ________, or number of time series values, to be included in the moving average.

   (b) If only the ______________ values of the time series are considered relevant, a small value of k is preferred.

   (c) If _____________ values are considered relevant, then a larger value of k is better.

   (d) A time series with a horizontal pattern can shift to a new level over time. A moving average will adapt to the new level of the series and resume providing good forecasts in k periods.

   (e) Thus, a smaller value of k will ______________ in a time series more quickly. But larger values of k will be more effective in ________________ the random fluctuations over time.

3. $\boxed{\text{Example}}$ (Recall Table 17.1 and Figure 17.1) the gasoline sales data

   (a) The time series plot in Figure 17.1 indicates that the gasoline sales time series has a ____________________. Thus, the smoothing methods of this section are applicable.

   (b) Use a three-week moving average ($k=3$), the forecast of sales in week 4 using the average of the time series values in weeks 1−3:

   $$F_4 = \text{average of weeks 1–3} = \underline{\hspace{4cm}}$$

   Thus, the moving average forecast of sales in week 4 is 76 or 76,000 liters of gasoline.

   (c) The actual value observed in week 4 is 92, the ________________ in week 4 is $92-76=16$.

(d) (Table 17.9) The forecast of sales in week 5 by averaging the time series values in weeks 2–4.

$$F_5 = \text{average of weeks 2–4} = \frac{84 + 76 + 92}{3} = 84$$

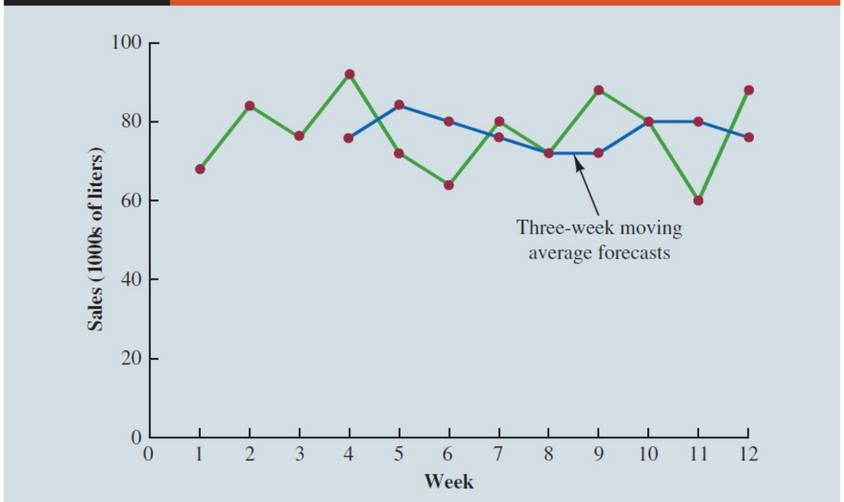
Hence, the forecast of sales in week 5 is 84 and the error associated with this forecast is $72 - 84 = -12$.

TABLE 17.9 Summary of Three-Week Moving Average Calculations

Week	Time Series Value	Forecast	Forecast Error	Absolute Value of Forecast Error	Squared Forecast Error	Percentage Error	Absolute Value of Percentage Error
1	68						
2	84						
3	76						
4	92	76	16	16	256	17.39	17.39
5	72	84	−12	12	144	−16.67	16.67
6	64	80	−16	16	256	−25.00	25.00
7	80	76	4	4	16	5.00	5.00
8	72	72	0	0	0	.00	.00
9	88	72	16	16	256	18.18	18.18
10	80	80	0	0	0	.00	.00
11	60	80	−20	20	400	−33.33	33.33
12	88	76	12	12	144	13.64	13.64
		Totals	0	96	1472	−20.79	129.21

(e) (Figure 17.7) Note how the graph of the moving average forecasts has tended to ________________ the random fluctuations in the time series.

FIGURE 17.7 Gasoline Sales Time Series Plot and Three-Week Moving Average Forecasts



(f) To forecast sales in week 13, the next time period in the future, we simply compute the average of the time series values in weeks 10, 11, and 12.

$$F_{13} = \text{average of weeks 10–12} = \frac{80+60+88}{3} = 76$$

(g) **Forecast Accuracy** Using the three-week moving average calculations in Table 17.9, the values for these three measures of forecast accuracy (MAE, MSE, and MAPE) are

$$\begin{aligned}
\text{MAE} &= \frac{96}{9} = 10.67 \quad \text{(mean absolute error)} \\
\text{MSE} &= \frac{1472}{9} = 163.56 \quad \text{(mean squared error)} \\
\text{MAPE} &= \frac{129.21}{9} = 14.36\% \quad \text{(mean absolute percentage error)}
\end{aligned}$$

(h) (Recall Section 17.2) Using the most recent observation as the forecast for the next week (a moving average of order $k = 1$) resulted in values of MAE = 14.91, MSE = 260.36, and MAPE = 19.24%. Thus, in each case the three-week moving average approach provided ________________ forecasts than simply using the most recent observation as the forecast.

4. To determine if a moving average with a different order k can provide more accurate forecasts, we recommend using ________________ to determine the value of k that minimizes MSE.

5. For the gasoline sales time series, it can be shown that the minimum value of MSE corresponds to a moving average of order ________________________. If we are willing to assume that the order of the moving average that is best for the historical data will also be best for future values of the time series, the most accurate moving average forecasts of gasoline sales can be obtained using a moving average of order $k = 6$.

Weighted Moving Averages

1. In the moving averages method, each observation in the moving average calculation receives the _______________.

2. One variation, known as weighted moving averages, involves selecting a __________________ for each data value and then computing a weighted average of the most recent k values as the forecast.

3. In most cases, the ______________ observation receives the ______________, and the weight decreases for older data values.

4. A moving average forecast of order $k = 3$ is just a special case of the weighted moving averages method in which each weight is equal to 1/3. Note that for the weighted moving average method the sum of the weights is equal to _____.

5. $\boxed{\text{Example}}$ We assign a weight of _______ to the most recent observation, a weight of _______ to the second most recent observation, and a weight of _______ to the third most recent observation. Using this weighted average, our forecast for week 4 is:

   Forecast for week 4 = ______________________________

6. To use the weighted moving averages method, we must first select the number of data values to be included in the weighted moving average and then choose weights for each of the data values. In general, if we believe that the _____________ is a better predictor of the future than the distant past, ________________ should be given to the more recent observations. However, when the time series is highly variable, selecting approximately _______________ for the data values may be best.

7. **Forecast Accuracy** To determine whether one particular combination of number of data values and weights provides a more accurate forecast than another combination, we recommend using ________ as the measure of forecast accuracy. That is, if we assume that the combination that is best for the ________ will also be best for the _________, we would use the combination of number of data values and weights that minimizes MSE for the historical time series to forecast the next value in the time series.

Exponential Smoothing

1. Exponential smoothing also uses a weighted average of past time series values as a forecast; it is a special case of the weighted moving averages method in which we select _________________—the weight for the ______________ observation.

2. The weights for the other data values are computed automatically and become smaller as the observations move farther into the past.

3. **Exponential Smoothing Forecast**

$$F_{t+1} = \underline{\hspace{4cm}} \qquad (17.2)$$

where

F_{t+1}: forecast of the time series for period $(t+1)$

Y_t: actual value of the time series in period t

F_t: forecast of the time series for period t

α: ____________________ $(0 \leq \alpha \leq 1)$

4. Equation (17.2) shows that the forecast for period $t+1$ is a weighted average of the actual value in period t and the forecast for period t.

5. The weight given to the actual value in period t is the smoothing constant ______ and the weight given to the forecast in period t is _________.

6. Let us illustrate by working with a time series involving only three periods of data: Y_1, Y_2, and Y_3.

   (a) To initiate the calculations, we let F_1 equal the actual value of the time series in period 1; that is, $F_1 = Y_1$. Hence, the forecast for period 2 is

   $$F_2 = \underline{\hspace{4cm}} = \underline{\hspace{5cm}}$$

   We see that the exponential smoothing forecast for period 2 is equal to the actual value of the time series in period ______.

   (b) The forecast for period 3 is

   $$F_3 = \underline{\hspace{7cm}}$$

   (c) Finally, substituting this expression for F3 in the expression for F4, we obtain

   $$\begin{aligned} F_4 &= \alpha Y_3 + (1-\alpha)F_3 \\ &= \alpha Y_3 + (1-\alpha)[\alpha Y_2 + (1-\alpha)Y_1] \\ &= \underline{\hspace{6cm}} \end{aligned}$$

(d) We now see that F_4 is a weighted average of the first three time series values. The sum of the coefficients, or weights, for Y_1, Y_2, and Y_3 equals 1.

(e) A similar argument can be made to show that, in general, any forecast F_{t+1} is a weighted average of all the previous time series values.

✏ Question .... (p876)

Use exponential smoothing approach with a smoothing parameter $\alpha = 0.2$ to obtain F_2, F_3, F_4 and F_{13} for the gasoline sales time series in Table 17.1 and Figure 17.1. Start the calculations, set the exponential smoothing forecast for period 2 equal to the actual value of the time series in period 1.

sol:

TABLE 17.10 Summary of the Exponential Smoothing Forecasts and Forecast Errors for the Gasoline Sales Time Series with Smoothing Constant $\alpha = .2$

Week	Time Series Value	Forecast	Forecast Error	Squared Forecast Error
1	68			
2	84	68.00	16.00	256.00
3	76	71.20	4.80	23.04
4	92	72.16	19.84	393.63
5	72	76.13	−4.13	17.06
6	64	75.30	−11.30	127.69
7	80	73.04	6.96	48.44
8	72	74.43	−2.43	5.90
9	88	73.95	14.05	197.40
10	80	76.76	3.24	10.50
11	60	77.41	−17.41	303.11
12	88	73.92	14.08	198.25
		Totals	43.70	1581.02

FIGURE 17.8 Actual and Forecast Gasoline Sales Time Series with Smoothing Constant $\alpha = .2$

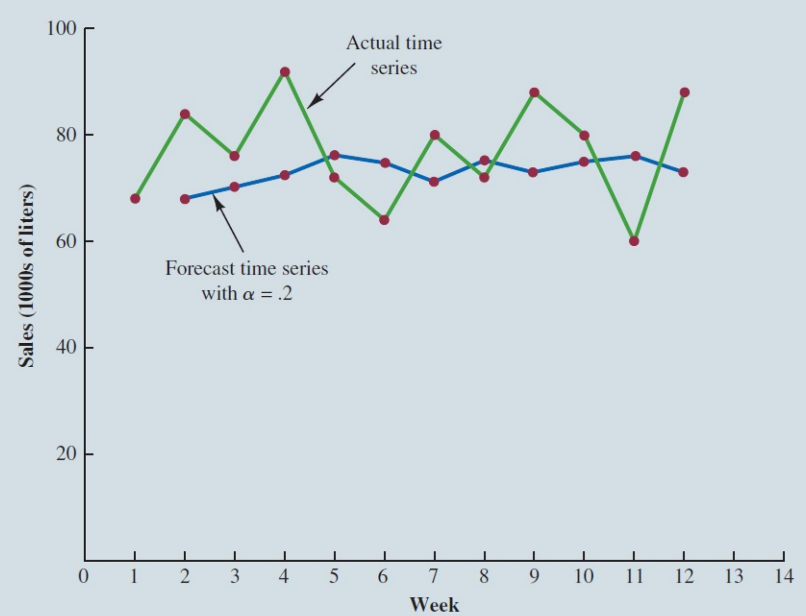


TABLE 17.11 Summary of the Exponential Smoothing Forecasts and Forecast Errors for the Gasoline Sales Time Series with Smoothing Constant $\alpha = .3$

Week	Time Series Value	Forecast	Forecast Error	Squared Forecast Error
1	68			
2	84	68.00	16.00	256.00
3	76	72.80	3.20	10.24
4	92	73.76	18.24	332.70
5	72	79.23	−7.23	52.27
6	64	77.06	−13.06	170.56
7	80	73.14	6.86	47.06
8	72	75.20	−3.20	10.24
9	88	74.24	13.76	189.34
10	80	78.37	1.63	2.66
11	60	78.86	−18.86	355.70
12	88	73.20	14.80	219.04
		Totals	32.14	1645.81

1. **Forecast Accuracy** (Table 17.10)(Figure 17.8)(Table 17.11) The criterion we will use to determine a desirable value for the smoothing constant a is the same as the criterion we proposed for determining the order or number of periods of data to include in the moving averages calculation. That is, we choose the value of a that ____________________.

2. The exponential smoothing results with $\alpha = 0.2$: the value of the sum of squared forecast errors is 98.80; hence ________________________. The exponential smoothing results with $\alpha = 0.3$: the value of the sum of squared forecast errors is 102.83; hence ________________________.

3. Thus, we would be inclined to prefer the original smoothing constant of $\alpha = 0.2$. Using a ________________ calculation with other values of α, we can find a "good value for the smoothing constant.

☺ **EXERCISES 17.3**: 5, 9, 11, 14

17.4 Trend Projection

1. We present two forecasting methods in this section that are appropriate for time series exhibiting a ____________.
   (a) First, we show how __________________ can be used to forecast a time series with a linear trend.
   (b) Next we show how the ___________ capability of regression analysis can also be used to forecast time series with a __________ or _________ trend.

Linear Trend Regression

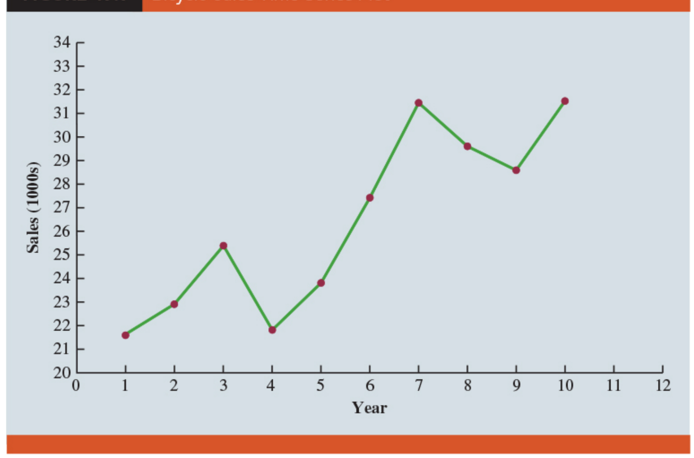
1. (Table 17.12) (Figure 17.9) the bicycle sales time series: the linear trend line provides a reasonable approximation of the long-run movement in the series.

TABLE 17.12

Bicycle Sales Time Series

Year	Sales (1000s)
1	21.6
2	22.9
3	25.5
4	21.9
5	23.9
6	27.5
7	31.5
8	29.7
9	28.6
10	31.4

FIGURE 17.9 Bicycle Sales Time Series Plot



2. The estimated regression equation describing a ___________ relationship between an independent variable x and a dependent variable y is written as

$$\underline{\hspace{3cm}}$$

where $\hat{y}$ is the estimated or predicted value of y.

3. To emphasize the fact that in forecasting the independent variable is time, we will replace ____ with ____ and ____ with ____ to emphasize that we are estimating the trend for a time series.

4. **Linear Trend Equation**

$$\underline{\hspace{4cm}} \qquad (17.4)$$

where

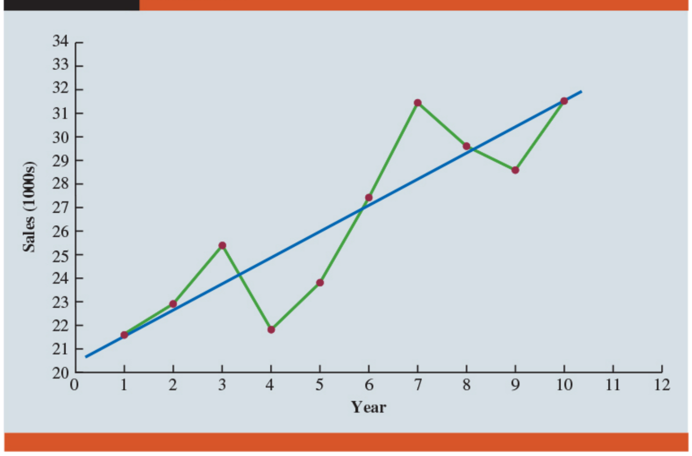
T_t = linear trend forecast in period t

b_0 = intercept of the linear trend line

b_1 = slope of the linear trend line

t = time period, $t = 1 (t = n)$ corresponding to the first time (most recent) series observation

FIGURE 17.10 Trend Represented by a Linear Function for the Bicycle Sales Time Series



5. **Computing the Slope and Intercept for a Linear Trend**

$$\begin{aligned} b_1 &= \frac{\hspace{5cm}}{\underline{\hspace{5cm}}} \\ &= \frac{\hspace{4cm}}{\underline{\hspace{4cm}}} \qquad (17.5) \\ b_0 &= \underline{\hspace{3cm}} \qquad (17.6) \end{aligned}$$

where

Y_t = value of the time series in period t

n = number of time periods (number of observations)

$\bar{Y}$ = average value of the time series

$\bar{t}$ = average value of t

6. $\boxed{\text{Example}}$ the bicycle sales time series

TABLE 17.13 Summary of Linear Trend Calculations for the Bicycle Sales Time Series

	t	Y_t	$t-\bar{t}$	$Y_t-\bar{Y}$	$(t-\bar{t})(Y_t-\bar{Y})$	$(t-\bar{t})^2$
	1	21.6	−4.5	−4.85	21.825	20.25
	2	22.9	−3.5	−3.55	12.425	12.25
	3	25.5	−2.5	−.95	2.375	6.25
	4	21.9	−1.5	−4.55	6.825	2.25
	5	23.9	−.5	−2.55	1.275	.25
	6	27.5	.5	1.05	.525	.25
	7	31.5	1.5	5.05	7.575	2.25
	8	29.7	2.5	3.25	8.125	6.25
	9	28.6	3.5	2.15	7.525	12.25
	10	31.4	4.5	4.95	22.275	20.25
Totals	55	264.5			90.750	82.50

(a) The linear trend equation is

$$T_t = \underline{\hspace{6em}}$$

(b) The slope of 1.1 indicates that over the past 10 years the firm experienced an ____________________ of about ____________ per year.

(c) If we assume that the past 10-year trend in sales is a good indicator of the future, this trend equation can be used to develop forecasts for future time periods. For example, the next year's trend projection or forecast, T_{11}.

$$T_{11} = \underline{\hspace{10em}}$$

Thus, using trend projection, we would forecast sales of 32,500 bicycles next year.

7. (Table 17.14) Use MSE to compute the accuracy associated with the trend projection forecasting method.

TABLE 17.14 Summary of the Linear Trend Forecasts and Forecast Errors for the Bicycle Sales Time Series

Year	Sales (1000s) Y_t	Forecast T_t	Forecast Error	Squared Forecast Error
1	21.6	21.5	.1	.01
2	22.9	22.6	.3	.09
3	25.5	23.7	1.8	3.24
4	21.9	24.8	−2.9	8.41
5	23.9	25.9	−2.0	4.00
6	27.5	27.0	.5	.25
7	31.5	28.1	3.4	11.56
8	29.7	29.2	.5	.25
9	28.6	30.3	−1.7	2.89
10	31.4	31.4	.0	.00
			Total	30.70

$$MSE = \underline{\hspace{8cm}}$$

8. Because ____________________ in forecasting is the same as the standard regression analysis procedure applied to time-series data, we can use statistical software to perform the calculations.

9. (Figure 17.11) the value of MSE in the ANOVA table is

$$MSE = \frac{\text{Sum of Squares Due to Error}}{\text{Degrees of Freedom}} = \underline{\hspace{4cm}}$$

FIGURE 17.11 Regression Output for the Bicycle Sales Time Series

Regression Analysis: Sales Versus Year

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	99.82	99.825	26.01	.001
Error	8	30.70	3.837		
Total	9	130.52			

Model Summary

S	R-sq
1.95895	76.48%

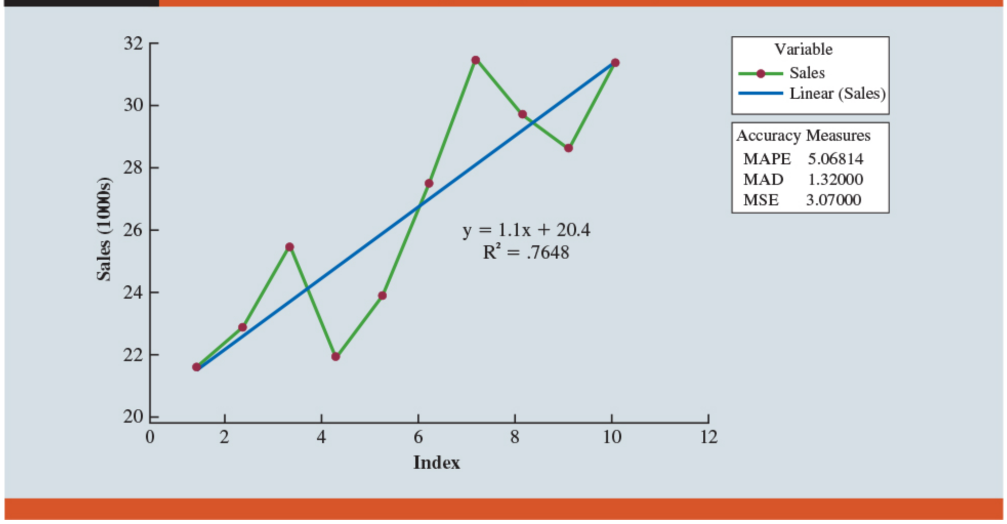
Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	20.40	1.34	15.24	.000
Year	1.10	.216	5.10	.001

Regression Equation

Sales = 20.40 + 1.100 Year

FIGURE 17.12 Time Series Linear Trend Analysis Output for the Bicycle Sales Time Series



10. This value of MSE __________ from the value of MSE that we computed previously because the sum of squared errors is divided by ______ instead of _______; thus, MSE in the regression output is not the ______________________________________.

11. **NOTE**: Most forecasting packages, however, compute MSE by taking the average of the squared errors. Thus, when using time series packages to develop a trend equation, the value of MSE that is reported may differ slightly from the value you would obtain using a general regression approach.

Nonlinear Trend Regression

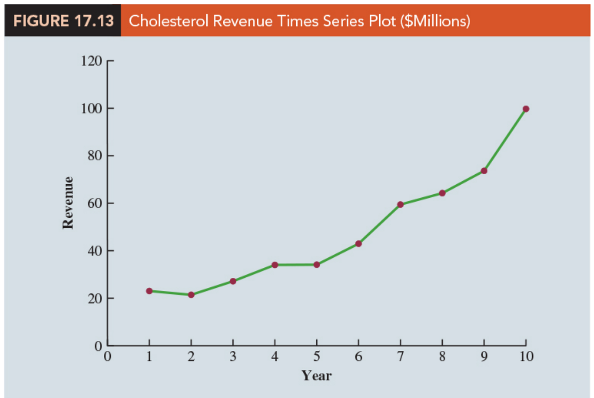
1. Example (Table 17.15) (Figure 17.13) Consider the annual revenue in millions of dollars for a cholesterol drug for the first 10 years of sales.

TABLE 17.15

Cholesterol Revenue Time Series ($ Millions)

Year (t)	Revenue ($ millions)
1	23.1
2	21.3
3	27.4
4	34.6
5	33.8
6	43.2
7	59.5
8	64.4
9	74.2
10	99.3

FIGURE 17.13 Cholesterol Revenue Times Series Plot ($Millions)



2. The time series plot indicates an ______________ or __________ trend. A curvilinear function appears to be needed to model the long-term trend.

3. **Quadratic Trend Equation** A variety of nonlinear functions can be used to develop an estimate of the trend for the cholesterol time series. For instance, consider the following quadratic trend equation:

$$\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \quad (17.7)$$

4. (Figure 17.14) a portion of the multiple regression output for the quadratic trend model;

FIGURE 17.14 Quadratic Trend Regression Output for the Cholesterol Revenue Time Series

Regression Analysis: Revenue Versus Year, YearSq

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	5770.13	2885.06	182.52	.000
Error	7	110.65	15.81		
Total	9	5880.78			

Model Summary

S	R-sq
3.97578	98.12%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	24.18	4.68	5.17	.001
Year	−2.11	1.95	−1.08	.317
YearSq	.922	.173	5.33	.001

Regression Equation

Revenue = 24.18 − 2.11 Year + .922 YearSq

The estimated regression equation is

$$\text{Revenue (\$millions)} = 24.18 - 2.11\text{Year} + 0.922\text{YearSq}$$

5. **Exponential Trend Equation**

$$\_\_\_\_\_\_\_\_\_\_\_\_ \quad (17.8)$$

6. Suppose $b_0 = 16.71$, and $b_1 = 0.1697$, T_t is not increasing by a constant amount as in the case of the linear trend model but by a ________________.

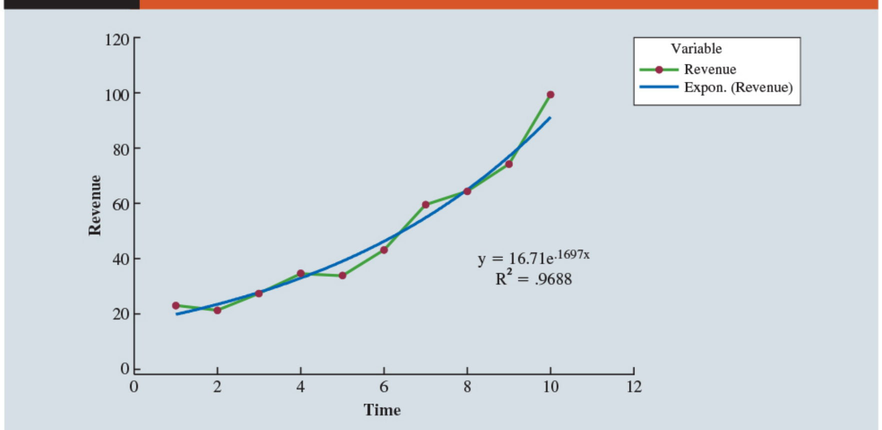
7. In this exponential trend model, multiplicative factor is ________________, so the constant percentage increase from time period to time period is _________.

8. Many statistical software packages have the capability to compute an exponential trend equation directly. Some software packages only provide linear trend, but by applying a natural *log* transformation to both sides of the equality in equation (17.8) we can apply the equivalent linear form:

____

(Figure 17.15)

FIGURE 17.15 Time Series Exponential Growth Trend Analysis Output for the Cholesterol Sales Time Series



☺ **EXERCISES 17.4**: 17, 20, 22, 26

17.5 Seasonality and Trend

Seasonality Without Trend

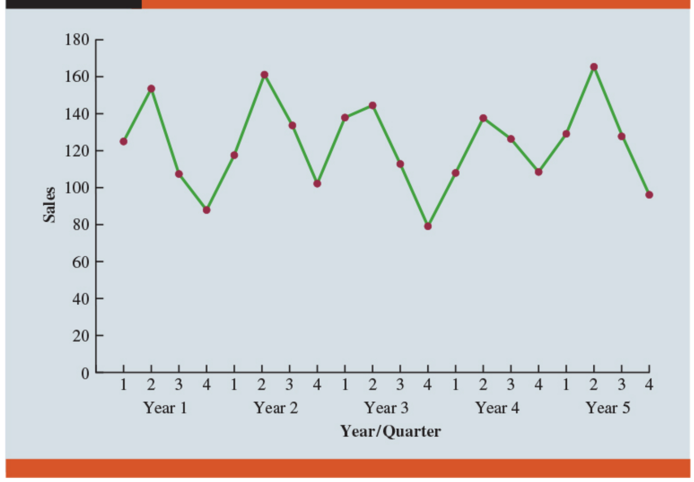
1. $\boxed{\text{Example}}$ (Table 17.16)(Figure 17.16) Consider the number of umbrellas sold at a clothing store over the past five years.

TABLE 17.16

Umbrella Sales Time Series

Year	Quarter	Sales
1	1	125
	2	153
	3	106
	4	88
2	1	118
	2	161
	3	133
	4	102
3	1	138
	2	144
	3	113
	4	80
4	1	109
	2	137
	3	125
	4	109
5	1	130
	2	165
	3	128
	4	96

FIGURE 17.16 Umbrella Sales Time Series Plot



2. The time series plot does not indicate any ____________ trend in sales. The first and third quarters have moderate sales, the second quarter has the highest sales, and the fourth quarter tends to be the lowest quarter in terms of sales volume. Thus, we would conclude that a ___________________ pattern is present.

3. Just like using __________________ to deal with an independent variable in a standard regression analysis, we can use the same approach to model a time series with a seasonal pattern by treating the season as a ____________________.

4. Recall that when a categorical variable has k levels, ________ dummy variables are required. Thus, to model the ________________ in the umbrella time series we need 4−1 = 3 dummy variables:

$$Qtr1 = \begin{cases} 1 & \text{if Quarter 1} \\ 0 & \text{otherwise} \end{cases}, Qtr2 = \begin{cases} 1 & \text{if Quarter 2} \\ 0 & \text{otherwise} \end{cases}, Qtr3 = \begin{cases} 1 & \text{if Quarter 3} \\ 0 & \text{otherwise} \end{cases}$$

5. Using $\hat{Y}$ to denote the estimated or forecasted value of sales, the general form of the estimated regression equation relating the number of umbrellas sold to the quarter

the sales take place:

---

6. (Table 17.17) the umbrella sales time series with the coded values of the dummy variables.

TABLE 17.17 Umbrella Sales Time Series with Dummy Variables

Year	Quarter	Qtr1	Qtr2	Qtr3	Sales
1	1	1	0	0	125
	2	0	1	0	153
	3	0	0	1	106
	4	0	0	0	88
2	1	1	0	0	118
	2	0	1	0	161
	3	0	0	1	133
	4	0	0	0	102
3	1	1	0	0	138
	2	0	1	0	144
	3	0	0	1	113
	4	0	0	0	80
4	1	1	0	0	109
	2	0	1	0	137
	3	0	0	1	125
	4	0	0	0	109
5	1	1	0	0	130
	2	0	1	0	165
	3	0	0	1	128
	4	0	0	0	96

7. (Figure 17.17) the computer output: the estimated multiple regression equation obtained is

$$Sales = 95.00 + 29.00\ Qtr1 + 57.00\ Qtr2 + 26.00\ Qtr3$$

We can use this equation to forecast quarterly sales for next year.

Quarter 1: $Sales = 95.0 + 29.0(1) + 57.0(0) + 26.0(0) = 124.$

Quarter 2: $Sales = 95.0 + 29.0(0) + 57.0(1) + 26.0(0) = 152.$

Quarter 3: $Sales = 95.0 + 29.0(0) + 57.0(0) + 26.0(1) = 121.$

Quarter 4: $Sales = 95.0 + 29.0(0) + 57.0(1) + 26.0(0) = 95.$

FIGURE 17.17 Regression Output for the Umbrella Sales Time Series

Term	Coef	SE Coef	T-Value	P-Value
Constant	95.00	5.06	18.76	.000
Qtr1	29.00	7.16	4.05	.001
Qtr2	57.00	7.16	7.96	.000
Qtr3	26.00	7.16	3.63	.002

Regression Equation

Sales = 95.00 + 29.00 Qtr1 + 57.00 Qtr2 + 26.00 Qtr3

8. The regression output shown in Figure 17.17 provides additional information that can be used to assess the ____________ of the forecast and determine the ______________ of the results.

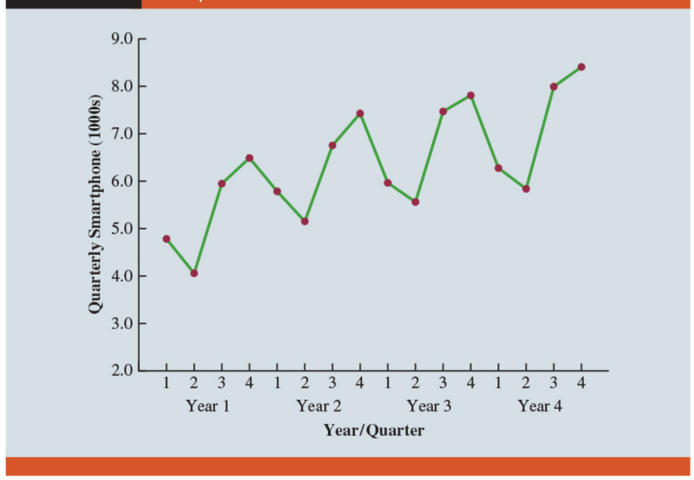
Seasonality and Trend

1. $\boxed{\text{Example}}$ (Table 17.18) (Figure 17.18) The quarterly smartphone sales.

TABLE 17.18
Smartphone Sales Time Series

Year	Quarter	Sales (1000s)
1	1	4.8
	2	4.1
	3	6.0
	4	6.5
2	1	5.8
	2	5.2
	3	6.8
	4	7.4
3	1	6.0
	2	5.6
	3	7.5
	4	7.8
4	1	6.3
	2	5.9
	3	8.0
	4	8.4

FIGURE 17.18 Smartphone Sales Time Series Plot



2. The sales are lowest in the second quarter of each year and increase in quarters 3 and 4. Thus, we conclude that a __________________ exists for smartphone sales.

3. But the time series also has an ____________________ that will need to be accounted for in order to develop accurate forecasts of quarterly sales.

4. This is easily handled by combining the __________________________ for seasonality with the time series ____________________ for hand-ling linear trend.

5. The general form of the estimated multiple regression equation for modeling both the quarterly seasonal effects and the linear trend in the smartphone time series:

____

where

$\hat{Y}_t$ = estimate or forecast of sales in time period t

$Qtr1 = 1$ $(Qtr2 = 1)(Qtr3 = 1)$ if time period t corresponds to the first (second) (third) quarter of the year; 0 otherwise.

6. (Table 17.19) revised smartphone sales time series that includes the coded values of the dummy variables and the time period t.

TABLE 17.19 Smartphone Sales Time Series with Dummy Variables and Time Period

Year	Quarter	Qtr1	Qtr2	Qtr3	Period	Sales (1000s)
1	1	1	0	0	1	4.8
	2	0	1	0	2	4.1
	3	0	0	1	3	6.0
	4	0	0	0	4	6.5
2	1	1	0	0	5	5.8
	2	0	1	0	6	5.2
	3	0	0	1	7	6.8
	4	0	0	0	8	7.4
3	1	1	0	0	9	6.0
	2	0	1	0	10	5.6
	3	0	0	1	11	7.5
	4	0	0	0	12	7.8
4	1	1	0	0	13	6.3
	2	0	1	0	14	5.9
	3	0	0	1	15	8.0
	4	0	0	0	16	8.4

7. (Figure 17.19) The estimated multiple regression equation is

$$Sales(1000s) = 6.069 - 1.363\ Qtr1 - 2.034\ Qtr2 - 0.304Qtr3 + 0.1456\ t \qquad (17.9)$$

8. Forecast for Time Period 17 (Quarter 1 in Year 5):

$$Sales(1000s) = 6.069 + 1.363(1) + 2.034(0) + 0.304(0) + 0.1456(17) = 7.18$$

Thus, accounting for the seasonal effects and the linear trend in smartphone sales, the estimates of quarterly sales in year 5 are 7180, 6660, 8530, and 8980.

9. The dummy variables in the estimated multiple regression equation actually provide ______________________ regression equations, one for each quarter. If time period t corresponds to quarter 1, the estimate of quarterly sales is

$$\begin{aligned}
Quarter1: Sales &= 6.069 - 1.363(1) - 2.034(0) - 0.304(0) + 0.1456(t) \\
&= 4.71 + 0.1456t \\
Quarter2: Sales &= 4.04 + 0.1456t \\
Quarter3: Sales &= 5.77 + 0.1456t \\
Quarter4: Sales &= 6.07 + 0.1456t
\end{aligned}$$

10. The ________ of the trend line for each quarterly forecast equation is __________, indicating a __________ in sales of about 146 sets per quarter.

11. The intercept for the Quarter 1 equation is 4.71 and the intercept for Quarter 4 equation is 6.07. Thus, sales in Quarter 1 are ____________________ or ____________ ____________ in Quarter 4.

12. The estimated regression coefficient for $Qtr1$ in equation (17.9) provides an estimate of the difference in sales between Quarter _____ and Quarter _____.

13. Similar interpretations can be provided for _________, the estimated regression coefficient for dummy variable $Qtr2$, and __________, the estimated regression coefficient for dummy variable $Qtr3$.

Models Based on Monthly Data

1. For monthly data, season is a categorical variable with 12 levels and thus _____________ dummy variables are required.

$$Month1 = \begin{cases} 1 & \text{if January} \\ 0 & \text{otherwise} \end{cases}, \ldots, \; Month11 = \begin{cases} 1 & \text{if November} \\ 0 & \text{otherwise} \end{cases}$$

2. Other than this change, the multiple regression approach for handling seasonality remains ____________.

☺ **EXERCISES 17.5**: 28, 30, 33

17.6 Time Series Decomposition*

☺ **SUPPLEMENTARY EXERCISES**: 41, 44, 47