Business Forecasting with ForecastX 7th edition Keating Solutions Manual

SKU: 9781259903915-SOLUTIONS

Solutions to Chapter 2 Exercises

1. The mean volume of sales for a sample of 100 sales representatives is $25,350 per month. The sample standard deviation is $7,490. The vice president for sales would like to know whether this result is significantly different from $24,000 at a 95 percent confidence level. Set up the appropriate null and alternative hypotheses, and perform the appropriate statistical test.

$$H_0: \mu = 24{,}000 \quad \text{vs.} \quad H_1: \mu \neq 24{,}000$$

$$tc = (25{,}350 - 24{,}000) \div (7{,}490 \div \sqrt{100})$$

$$tc = (25{,}350 - 24{,}000) \div (7{,}490 \div 10)$$

$$tc = (1{,}350) \div (749) = 1.802$$

At n = 100 and for a 95% confidence level and a two tailed t-test the table value of t = 1.96. Since 1.802 is less than 1.96 we would fail to reject the null hypothesis (H_0:) and conclude that there is not sufficient statistical evidence to say that the two values are significantly different.

2. Larry Bomser has been asked to evaluate sizes of tire inventories for retail outlets of a major tire manufacturer. From a sample of 120 stores, he has found a mean of 310 tires. The industry average is 325. If the standard deviation for the sample was 72, would you say that the inventory level maintained by this manufacturer is significantly different from the industry norm? Explain why. (Use a 95 percent confidence level.)

$$H_0: \mu = 325 \quad \text{vs.} \quad H_1: \mu \neq 325$$

$$tc = (310 - 325) \div (72 \div \sqrt{120})$$

$$tc = (310 - 325) \div (72 \div 10.95)$$

$$tc = (-25) \div (6.58) = -3.80$$

At n = 120 and for a 95% confidence level and a two tailed t-test the table value of t = 1.96. Since -3.80 is less than -1.96 we would reject the null hypothesis (H_0:) and conclude that there is sufficient statistical evidence to say that the sample mean is significantly different from 325. This is because -3.80 is quite far into the left tail of the t-distribution.

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

3. Twenty graduate students in business were asked how many credit hours they were taking in the current quarter. Their responses are shown as follows (c2p3):

Student Number	Credit Hours	Student Number	Credit Hours
1	2	11	6
2	7	12	5
3	9	13	9
4	9	14	13
5	8	15	10
6	11	16	6
7	6	17	9
8	8	18	6
9	12	19	9
10	11	20	10

1. Determine the mean, median, and mode for this sample of data. Write a sentence explaining what each means.

Mean =	8.3
Median =	9
Mode =	9

The mean is the arithmetic average which is equal to the sum of the observations (166) divided by the number of observations (20). That is, (166/20) = 8.3.

The median is the middle value that splits the data into two equal parts when the data are arrayed from low to high (or high to low). In this case sorting from low to high would show us that observations 10 through 14 are all equal to 9. When there are an even number of observations there is no middle value. In such a case one uses the average of the two most middle values, which in this case are both 9 (observations 10 and 11). Thus, the median is 9.

The mode is the most frequently occurring value. In this sample there are five observations equal to 9. The next most would be a value of 6 which occurs four times. Thus the mode for this sample is 9.

2. It has been suggested that graduate students in business take fewer credits per quarter than the typical graduate student at this university. The mean for all graduate students is 9.1 credit hours per quarter, and the data are normally distributed. Set up the appropriate null and alternative hypotheses, and determine whether the null hypothesis can be rejected at a 95 percent confidence level.

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

To find a solution you first need to calculate the standard deviation of the 20 sample values. It is 2.64.

$$H_0: \mu \geq 9.1 \quad \text{vs.} \quad H_1: \mu < 9.1$$

$$tc = (9.0 - 9.1) \div (2.64 \div \sqrt{20})$$

$$tc = (9.0 - 9.1) \div (2.64 \div 4.47)$$

$$tc = (-0.1) \div (0.59) = -0.17$$

The number of degrees of freedom is n-1, which in this case is 20 – 1 = 19. Based on a t-table for a one tailed t-test at a 95% confidence level and 19 df, the table value of t is 1.729. Since the absolute value of t_c is less than 1.729 you would fail to reject the null hypothesis (H_0:). Thus, there is not statistical evidence at the desired 95% confidence level to conclude that graduate students in business take fewer credit hours than the typical graduate student.

4. Arbon Computer Corporation (ACC) produces a popular PC clone. The sales manager for ACC has recently read a report that indicated that sales per sales representative for other producers are normally distributed with a mean of $255,000. She is interested in knowing whether her sales staff is comparable. She picked a random sample of 16 salespeople and obtained the following results (c2p4):

Person	Sales	Person	Sales
1	177,406	9	110,027
2	339,753	10	182,577
3	310,170	11	177,707
4	175,520	12	154,096
5	293,332	13	236,083
6	323,175	14	301,051
7	144,031	15	158,792
8	279,670	16	140,891

At a 5 percent significance level, can you reject the null hypothesis that ACC's mean sales per salesperson was $255,000? Draw a diagram that illustrates your answer.

First calculate the mean and standard deviation for the 16 people in the sample. The results are: Mean = 219,018 and Standard Deviation = 76,621.28.

$$H_0: \mu = 255{,}000 \quad \text{vs.} \quad H_1: \mu \neq 255{,}000$$

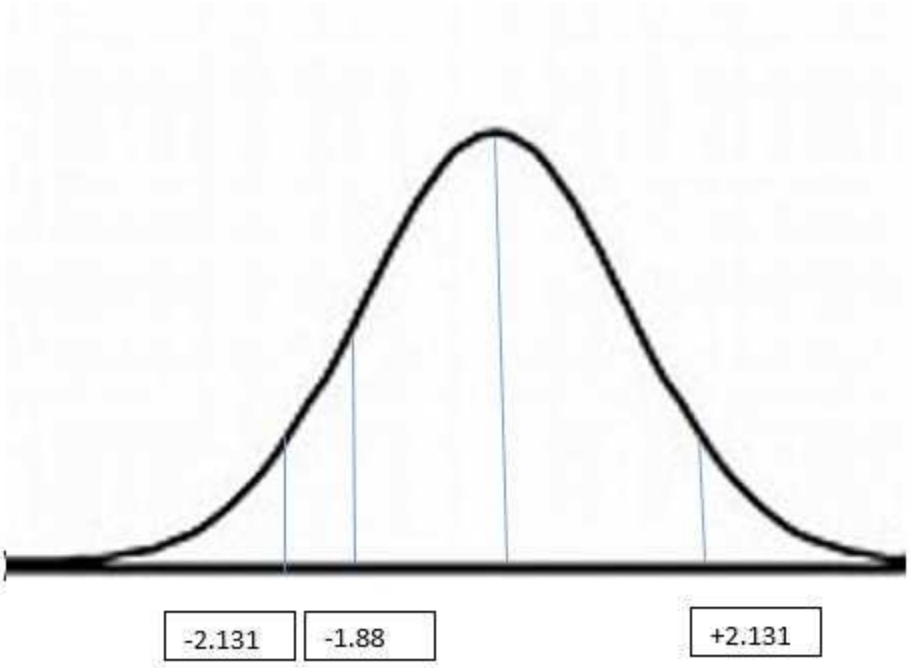
Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

$$tc = (219{,}018 - 255{,}000) \div (76{,}621.28 \div \sqrt{16})$$

$$tc = (219{,}018 - 255{,}000) \div (76{,}621.28 \div 4)$$

$$tc = (-35{,}982) \div (19{,}155.32) = -1.88$$

In this case with n=16, df=15. The corresponding value from the t-table at a 95% confidence level is ± 2.131. Since -1.88 lies between -2.131 and the middle of the distribution you would fail to reject H_0: and conclude that her sales people are similar to the others.



5. Assume that the weights of college football players are normally distributed with a mean of 205 pounds and a standard deviation of 30.

   1. What percentage of players would have weights greater than 205 pounds?

      If the population mean is 205 then 50% would be greater than 205.

   2. What percentage of players would weigh less than 250 pounds?

      If the population mean is 205 then 50% would be greater than 205.

   3. Ninety percentage of players would weigh more than what number of pounds?

      90% would include 40% to the left of the middle of the normal distribution and the 50% to the right of the middle. If you look in the body of a normal table the closest value to 0.40 (40%) is 0.3997. This value

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

corresponds to a z value of 1.28. Multiplying 1.28 times the standard deviation of 30 gives you 38.4. Subtracting 38.4 from the midpoint of 205 gives you 166.6. Thus, 90% of players would be expected to weigh more than 166.6 pounds.

4. What percentage of players would weigh between 180 and 230 pounds?

   With a mean of 205, 180 would be 25 below the mean and 230 would be 25 above the mean. Dividing 25 by the standard deviation of 30 gives you 0.83. So, each of the values of interest are 0.83 standard deviations from the mean. The value in the normal table that corresponds to 0.83 is 0.2967. This would be the area from the midpoint to =o.83 and also to -0.83. Thus, 2x0.2967 = 0.5934 or 59.34% as the percent of players expected to weigh between 180 and 230 pounds.

6. Mutual Savings Bank of Appleton has done a market research survey in which people were asked to rate their image of the bank on a scale of 1 to 10, with 10 being the most favorable. The mean response for the sample of 400 people was 7.25, with a standard deviation of 2.51. On this same question, a state association of mutual savings banks has found a mean of 7.01.

   1. Clara Weston, marketing director for the bank, would like to test to see whether the rating for her bank is significantly greater than the norm of 7.01. Perform the appropriate hypothesis test for a 95 percent confidence level.

      $$H_0: \mu \leq 9.1 \quad \text{vs.} \quad H_1: \mu > 9.1$$

      $$tc = (7.25 - 7.01) \div (2.51 \div \sqrt{400})$$

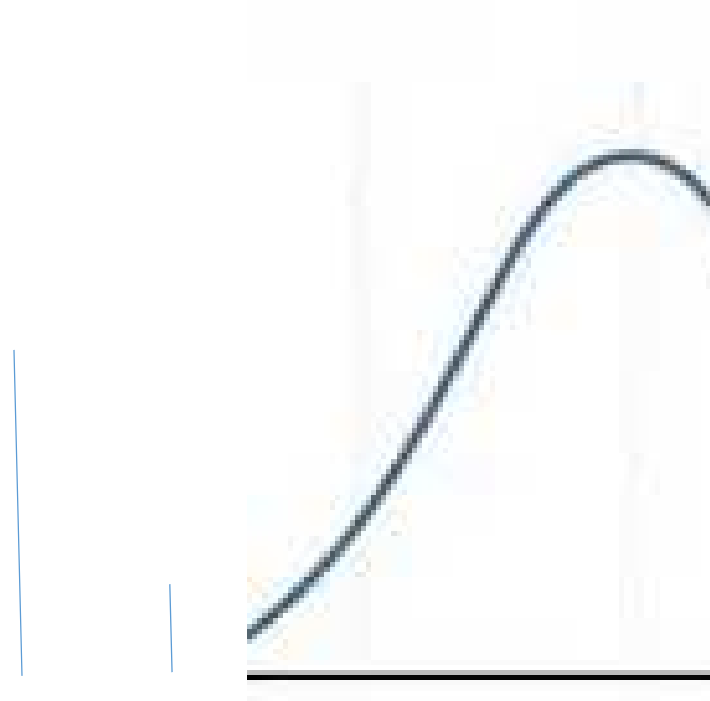
      $$tc = (7.25 - 7.01) \div (2.51 \div 20)$$

      $$tc = (0.24) \div (0.1255) = 1.91$$

      The t-value for a one tailed test at a 95% confidence level and df=399 (i.e. 400-1) is 1.645. Since 1.91 is greater than 1.645 you would reject the null hypothesis and conclude that the rating for Ms. Weston's bank is greater than the norm of 7.01.

   2. Draw a diagram to illustrate your result.

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.



1.645	1.91

3. How would your result be affected if the sample size had been 100 rather than 400, with everything else being the same?

$$H_0: \mu \leq 9.1 \quad \text{vs.} \quad H_1: \mu > 9.1$$

$$tc = (7.25 - 7.01) \div (2.51 \div \sqrt{100})$$

$$tc = (7.25 - 7.01) \div (2.51 \div 10)$$

$$tc = (0.24) \div (0.251) = 0.956$$

In this case you would fail to reject the null hypothesis so would not have evidence at a 95% confidence level that Ms. Weston's bank had ratings above the norm of 7.01.

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

7. In a sample of 25 classes, the following numbers of students were observed (c2p7):

Class	Number of students
1	40
2	50
3	42
4	20
5	29
6	39
7	49
8	46
9	52
10	45
11	51
12	64
13	43

Class	Number of students
14	37
15	35
16	44
17	10
18	40
19	36
20	20
21	20
22	29
23	58
24	51
25	54

1. Calculate the mean, median, standard deviation, variance, and range for this sample.

   Mean = 40.16
   Median = 42
   Std, Dev. = 13.14
   Variance = 172.72

2. What is the standard error of the mean based on this information?

   The standard error of the mean = Std Dev / Square Root of n

   The standard error of the mean = 13.14 / 5 = 2.628

3. What would be the best point estimate for the population class size?

   The best point estimate would be the sample mean = 40.16.

4. What is the 95 percent confidence interval for class size? What is the 90 percent confidence interval? Does the difference between these two make sense?

   The 95% confidence interval would be the sample mean ± 1.96 (the standard error of the mean)

   40.16 ± 1.96(2.628) = 40.16 ± 5.15 = 35.01 to 45.31

   The 90% confidence interval would be the sample mean ± 1.645 (the standard error of the mean)

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

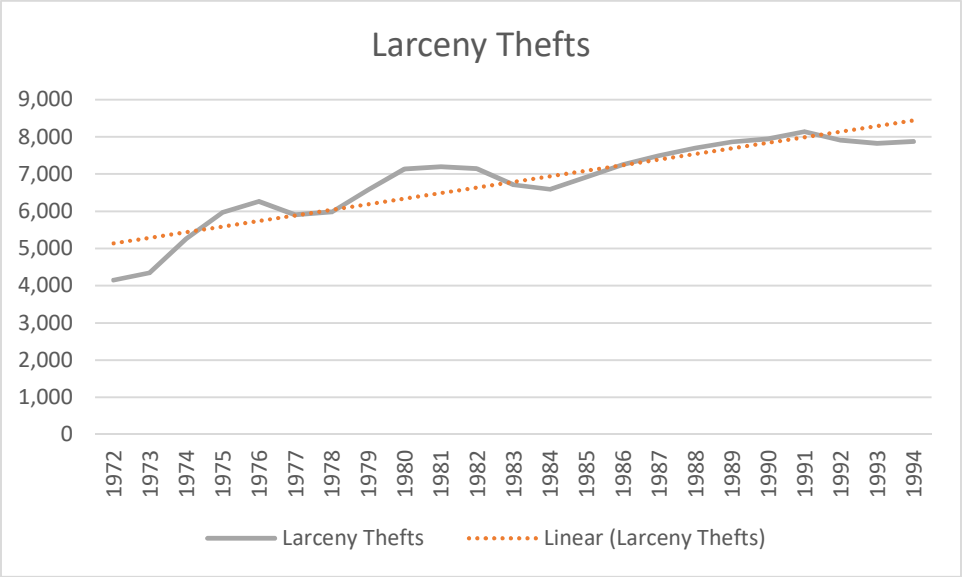
40.16 ± 1.645(2.628) = 40.16 ± 4.32 = 35.84 to 44.48

It makes sense that if you are less confident the confidence band would be more narrow.

8. CoastCo Insurance, Inc., is interested in forecasting annual larceny thefts in the United States using the following data (c2p8):

Year	Larceny Thefts		Year	Larceny Thefts
1972	4151		1984	6592
1973	4348		1985	6926
1974	5263		1986	7257
1975	5978		1987	7500
1976	6271		1988	7706
1977	5906		1989	7872
1978	5983		1990	7946
1979	6578		1991	8142
1980	7137		1992	7915
1981	7194		1993	7821
1982	7143		1994	7876
1983	6713			

1. Prepare a time-series plot of these data. On the basis of this graph, do you think there is a trend in the data? Explain.



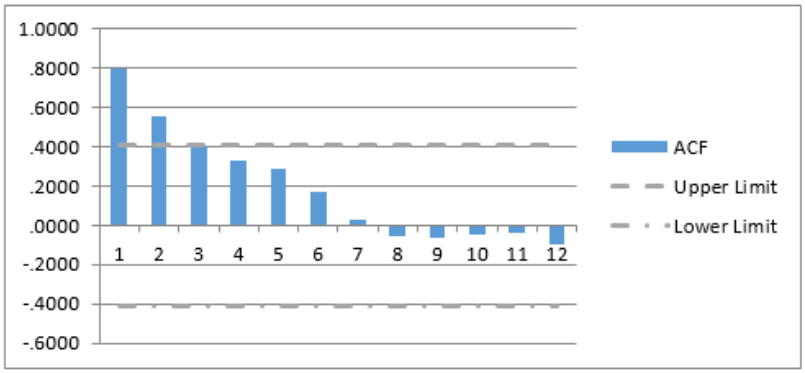
Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

As seen in the graph a dotted trend line has been added. The trend line shows a positive trend in the data.

2. Look at the autocorrelation structure of larceny thefts for lags of 1, 2, 3, 4, and 5. Do the autocorrelation coefficients fall quickly toward zero? Demonstrate that the critical value for r_k is 0.417. Explain what these results tell you about a trend in the data.

ACF of Larceny Thefts

Obs	ACF
1	.7986
2	.5594
3	.4045
4	.3298
5	.2844
6	.1705
7	.0312
8	-.0524
9	-.0626
10	-.0487
11	-.0360
12	-.0930



The ACF graph supports the idea that there is not a strong trend in the data. You see that after the first few bars the values fall quickly toward zero.

r_k = 2/ (square root of n) = 2/ square root 23 = 2/4.796 = 0.417.

3. On the basis of what is found in parts *a* and *b*, suggest a forecasting method from Table 2.1 that you think might be appropriate for this series.

Since the data are annual there can be no seasonality. A Holt's exponential smoothing model may be best although as one can see in the part (a) graph a linear trend is an alternative. Holt's MAPE = 3.90%, while the linear trend MAPE = 5.78%.

9. Use exploratory data analysis to determine whether there is a trend and/or seasonality in mobile home shipments (MHS). The data by quarter are shown in the following table (c2p9):

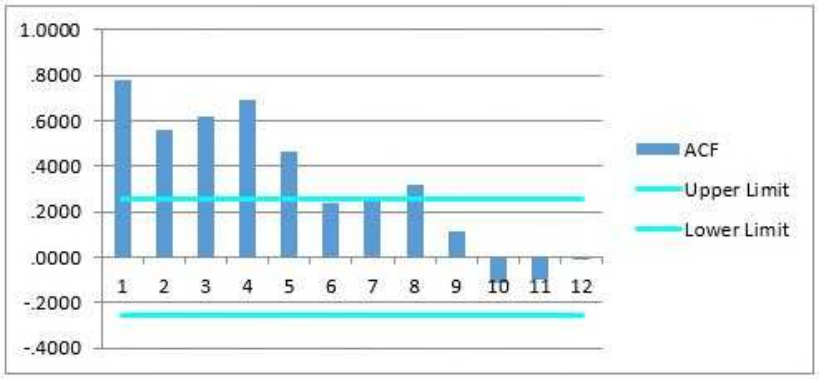
Period	MHS		Period	MHS		Period	MHS		Period	MHS
Mar-81	54.9		Dec-84	66.2		Sep-88	59.2		Jun-92	52.8
Jun-81	70.1		Mar-85	62.3		Dec-88	51.6		Sep-92	57
Sep-81	65.8		Jun-85	79.3		Mar-89	48.1		Dec-92	57.6
Dec-81	50.2		Sep-85	76.5		Jun-89	55.1		Mar-93	56.4
Mar-82	53.3		Dec-85	65.5		Sep-89	50.3		Jun-93	64.3
Jun-82	67.9		Mar-86	58.1		Dec-89	44.5		Sep-93	67.1

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

Sep-82	63.1		Jun-86	66.8		Mar-90	43.3		Dec-93	66.4
Dec-82	55.3		Sep-86	63.4		Jun-90	51.7		Mar-94	69.1
Mar-83	63.3		Dec-86	56.1		Sep-90	50.5		Jun-94	78.7
Jun-83	81.5		Mar-87	51.9		Dec-90	42.6		Sep-94	78.7
Sep-83	81.7		Jun-87	62.8		Mar-91	35.4		Dec-94	77.5
Dec-83	69.2		Sep-87	64.7		Jun-91	47.4		Mar-95	79.2
Mar-84	67.8		Dec-87	53.5		Sep-91	47.2		Jun-95	86.8
Jun-84	82.7		Mar-88	47		Dec-91	40.9		Sep-95	87.6
Sep-84	79		Jun-88	60.5		Mar-92	43		Dec-95	86.4

ACF of MHS

Obs	ACF
1	.7779
2	.5593
3	.6165
4	.6911
5	.4623
6	.2359
7	.2494
8	.3154
9	.1103
10	-.1118
11	-.0954
12	-.0136



On the basis of your analysis, do you think there is a significant trend in MHS? Is there seasonality?

The ACF values do not fall to zero quickly suggesting some trend. The relatively larger values at 1, 4, and 8 suggest seasonality.

What forecasting methods might be appropriate for MHS according to the guidelines in Table 2.1?

Based on Table 2.1 causal multiple regression, Winters' and time series decomposition would be good methods to consider.

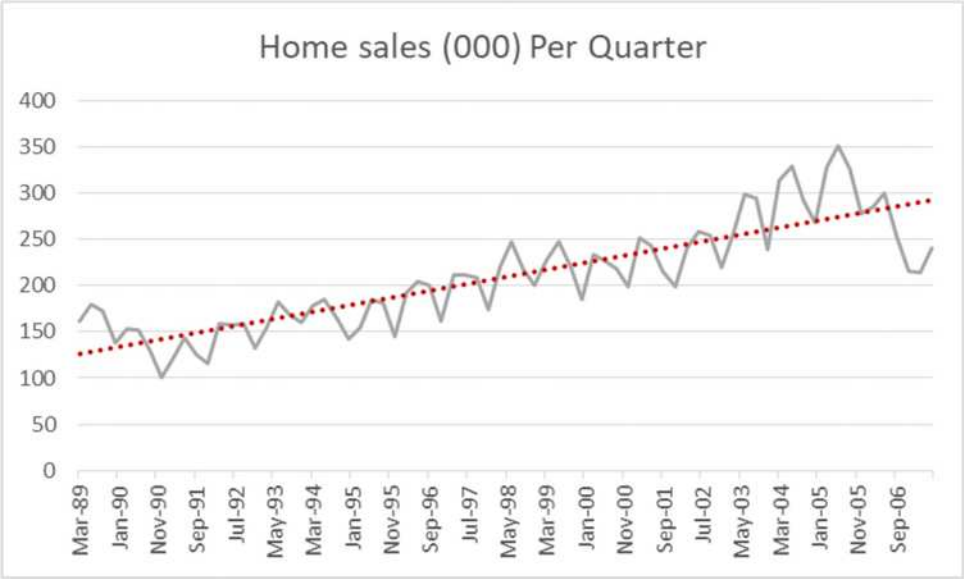
10. Home sales are often considered an important determinant of the future health of the economy. Thus, there is widespread interest in being able to forecast home sales (HS). Quarterly data for HS are shown in the following table in thousands of units (c2p10):

Date	Home sales (000) Per Quarter	Date	Home sales (000) Per Quarter	Date	Home sales (000) Per Quarter
Mar-89	161	Mar-95	154	Mar-01	251

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

Jun-89	179	Jun-95	185	Jun-01	243
Sep-89	172	Sep-95	181	Sep-01	216
Dec-89	138	Dec-95	145	Dec-01	199
Mar-90	153	Mar-96	192	Mar-02	240
Jun-90	152	Jun-96	204	Jun-02	258
Sep-90	130	Sep-96	201	Sep-02	254
Dec-90	100	Dec-96	161	Dec-02	220
Mar-91	121	Mar-97	211	Mar-03	256
Jun-91	144	Jun-97	212	Jun-03	299
Sep-91	126	Sep-97	208	Sep-03	294
Dec-91	116	Dec-97	174	Dec-03	239
Mar-92	159	Mar-98	220	Mar-04	314
Jun-92	158	Jun-98	247	Jun-04	329
Sep-92	159	Sep-98	218	Sep-04	292
Dec-92	132	Dec-98	200	Dec-04	268
Mar-93	154	Mar-99	227	Mar-05	328
Jun-93	183	Jun-99	248	Jun-05	351
Sep-93	169	Sep-99	221	Sep-05	326
Dec-93	160	Dec-99	185	Dec-05	278
Mar-94	178	Mar-00	233	Mar-06	285
Jun-94	185	Jun-00	226	Jun-06	300
Sep-94	165	Sep-00	219	Sep-06	251
Dec-94	142	Dec-00	199	Dec-06	216
				Mar-07	214
				Jun-07	240

1. Prepare a time-series plot of THS. Describe what you see in this plot in terms of trend and seasonality.



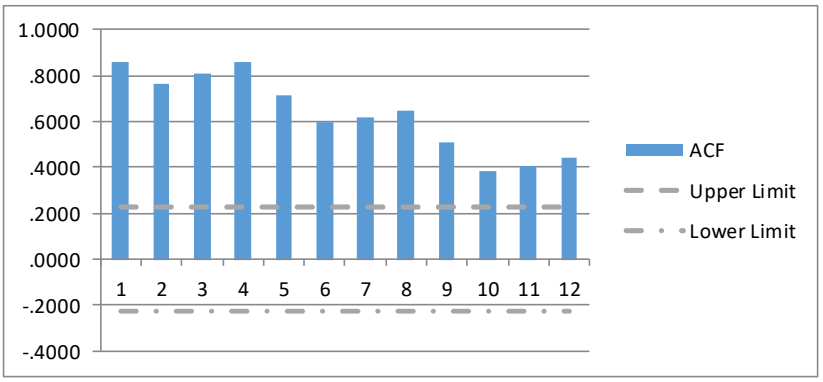
The graph shows a clear positive trend along with a regular seasonal pattern.

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

2. Calculate and plot the first 12 autocorrelation coefficients for HS. What does this autocorrelation structure suggest about the trend?

ACF of Home sales (000) Per Quarter

Obs	ACF
1	.8632
2	.7620
3	.8098
4	.8618
5	.7154
6	.5929
7	.6169
8	.6475
9	.5093
10	.3876
11	.4087
12	.4455



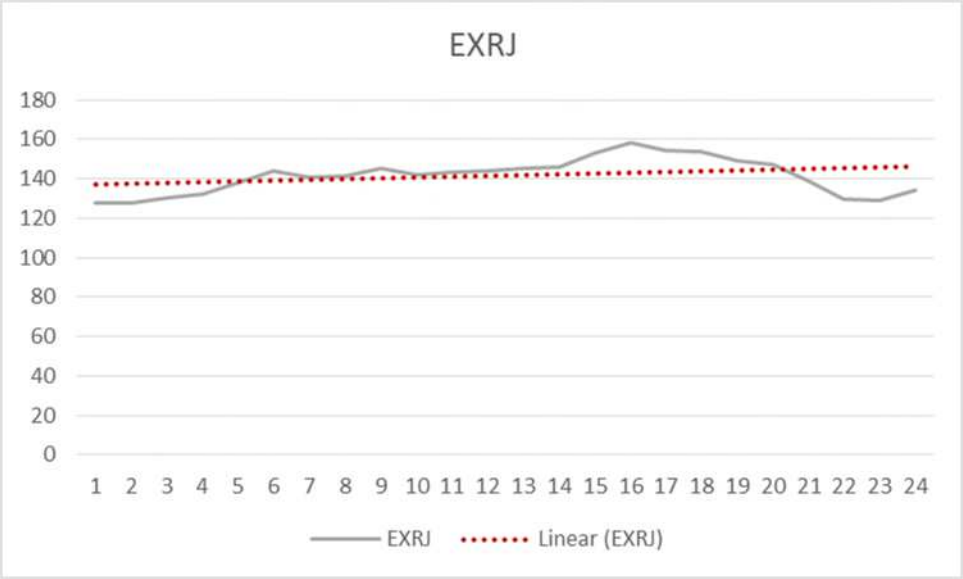
The ACF diagram shows the existing of a trend. The relatively high bars at 1, 4, 8 and 12 are suggestive of seasonality.

11. Exercise 12 of Chapter 1 includes data on the Japanese exchange rate (EXRJ) by month. On the basis of a time-series plot of these data and the autocorrelation structure of EXRJ, would you say the data are stationary? Explain your answer. (c2p11)

Period	EXRJ
1	127.36
2	127.74
3	130.55
4	132.04
5	137.86
6	143.98
7	140.42
8	141.49
9	145.07
10	142.21
11	143.53
12	143.69

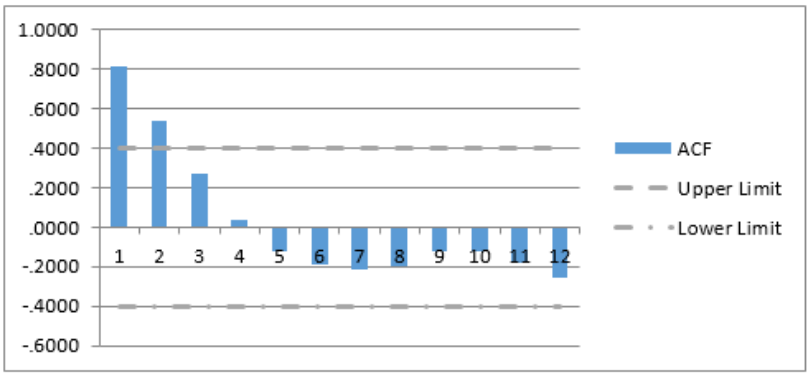
Period	EXRJ
13	144.98
14	145.69
15	153.31
16	158.46
17	154.04
18	153.7
19	149.04
20	147.46
21	138.44
22	129.59
23	129.22
24	133.89

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.



ACF of EXRJ

Obs	ACF
1	.8157
2	.5383
3	.2733
4	.0340
5	-.1214
6	-.1924
7	-.2157
8	-.1978
9	-.1215
10	-.1217
11	-.1823
12	-.2593

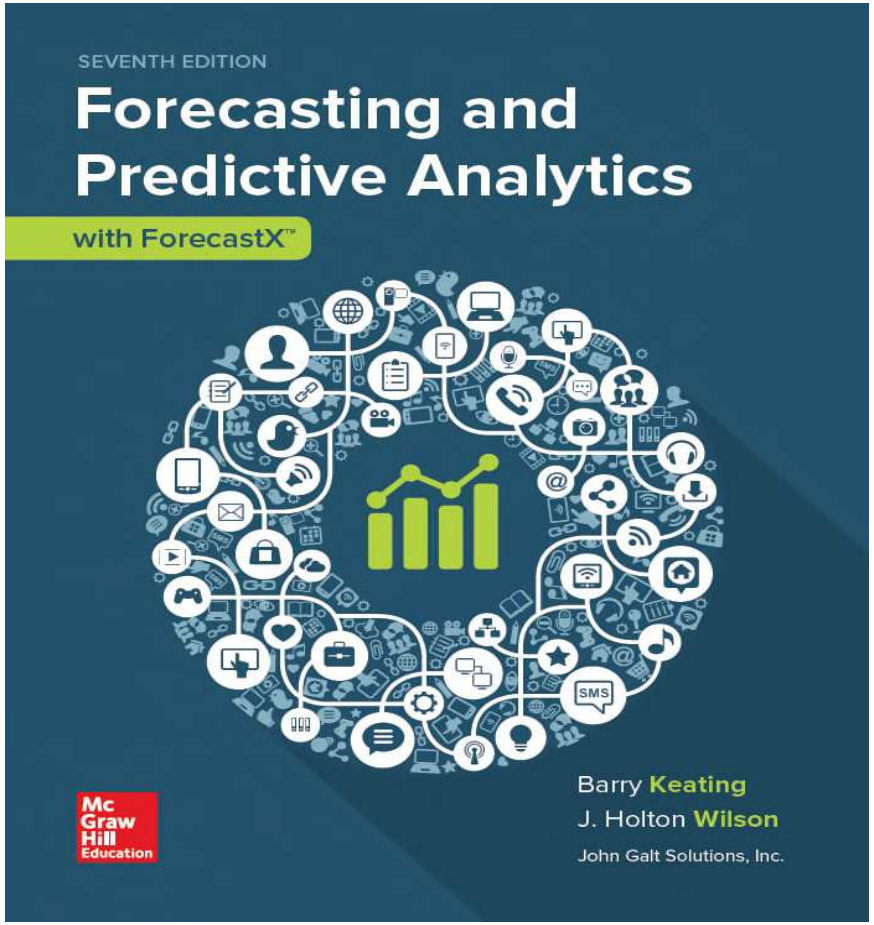


Both graphs suggest that there is not a significant trend in these exchange rate data.

Copyright © 2019 McGraw-Hill Education. All rights reserved. No reproduction or distribution without the prior written consent of McGraw-Hill Education.

Chapter 2

The Forecast Process, Data Considerations, and Model Selection



© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Business Forecasting with ForecastX 7th edition Keating Solutions Manual

The Forecast Process

1) Specify objectives.
2) Determine what to forecast.
3) Identify time dimensions.
4) Data considerations.
5) Model selection.
6) Model evaluation.
7) Forecast preparation.
8) Forecast presentation.
9) Tracking results.

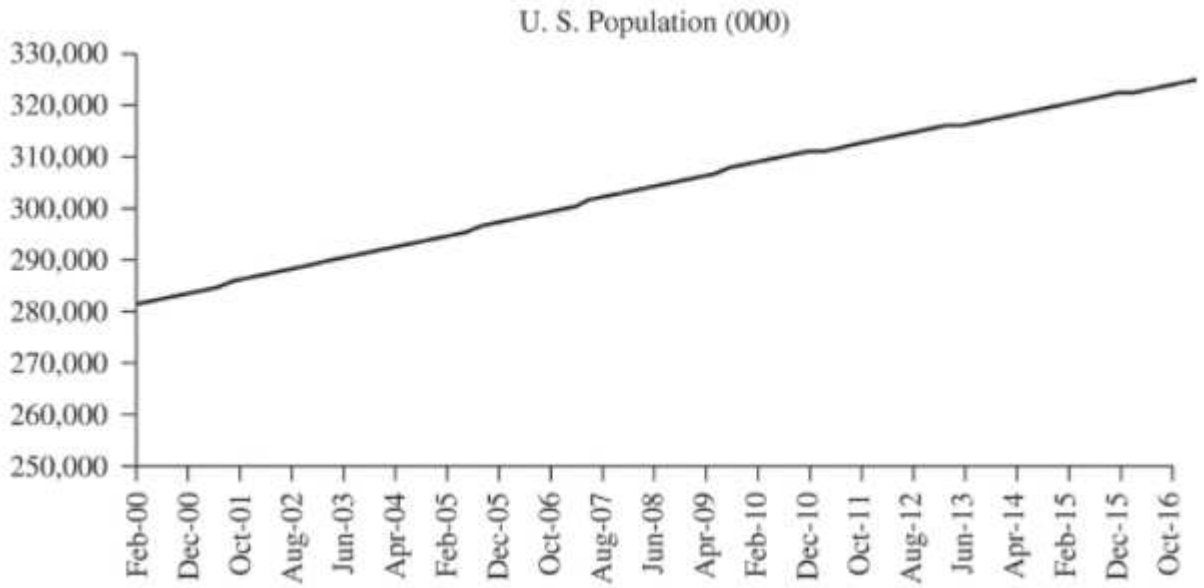
Forecasting Method	Data Pattern	Quantity of Historical Data (Number of Observations)	Forecast Horizon
Naive	Stationary	1 or 2	Very short
Moving averages	Stationary	Number equal to the periods in the moving average	Very short
Exponential smoothing			
Simple	Stationary	5 to 10	Short
Adaptive response	Stationary	10 to 15	Short
Holt's	Linear trend	10 to 15	Short to medium
Winters'	Trend and seasonality	At least 4 or 5 per season	Short to medium
Bass model	S-curve	Small, 3 to 10	Short to Medium
Regression-based			
Trend	Linear and nonlinear trend with or without seasonality	Minimum of 10 with 4 or 5 per season if seasonality is included	Short to medium
Causal	Can handle nearly all data patterns	Recommend a minimum of 10 per independent variable	Short, medium, and long
Time-series decomposition	Can handle trend, seasonal, and cyclical patterns	Enough to see two peaks and two troughs in the cycle	Short, medium, and long
ARIMA	Stationary or transformed to stationary	Minimum of 50	Short, medium, and long

Typical Time Series Patterns

- Trend
- Seasonal
- Cyclical
- Irregular (often called random)

© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Figure 2.1


U. S. Population (000)
330,000
320,000
310,000
300,000
290,000
280,000
270,000
260,000
250,000
Feb-00
Dec-00
Oct-01
Aug-02
Jun-03
Apr-04
Feb-05
Dec-05
Oct-06
Aug-07
Jun-08
Apr-09
Feb-10
Dec-10
Oct-11
Aug-12
Jun-13
Apr-14
Feb-15
Dec-15
Oct-16

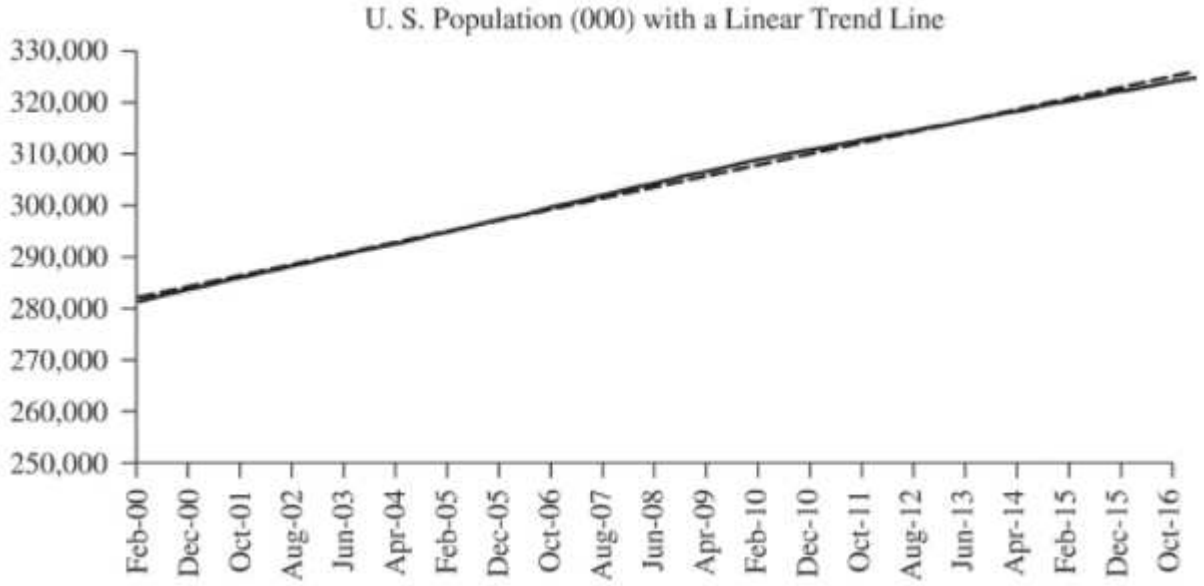

U. S. Population (000) with a Linear Trend Line
330,000
320,000
310,000
300,000
290,000
280,000
270,000
260,000
250,000
Feb-00
Dec-00
Oct-01
Aug-02
Jun-03
Apr-04
Feb-05
Dec-05
Oct-06
Aug-07
Jun-08
Apr-09
Feb-10
Dec-10
Oct-11
Aug-12
Jun-13
Apr-14
Feb-15
Dec-15
Oct-16

Figure 2.2

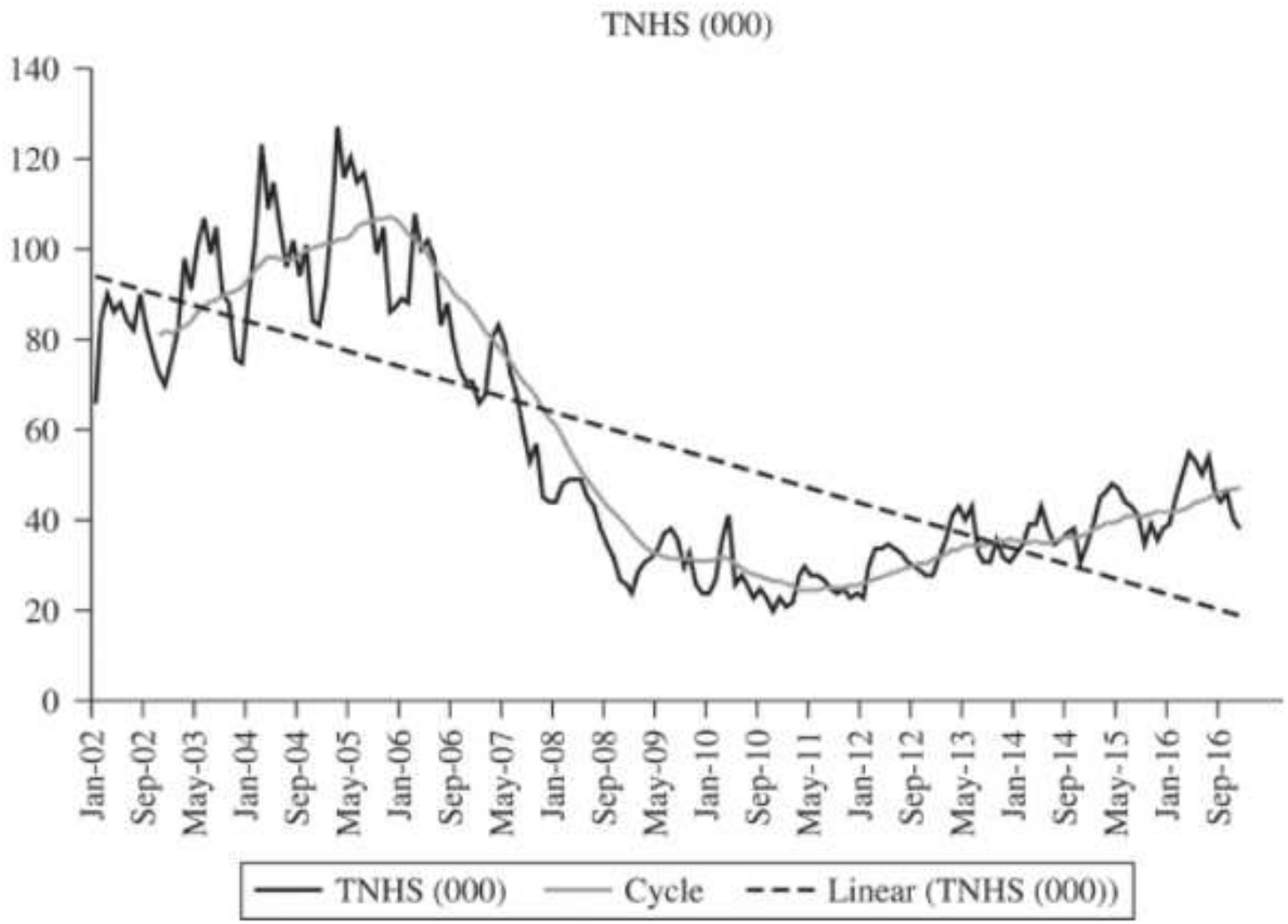
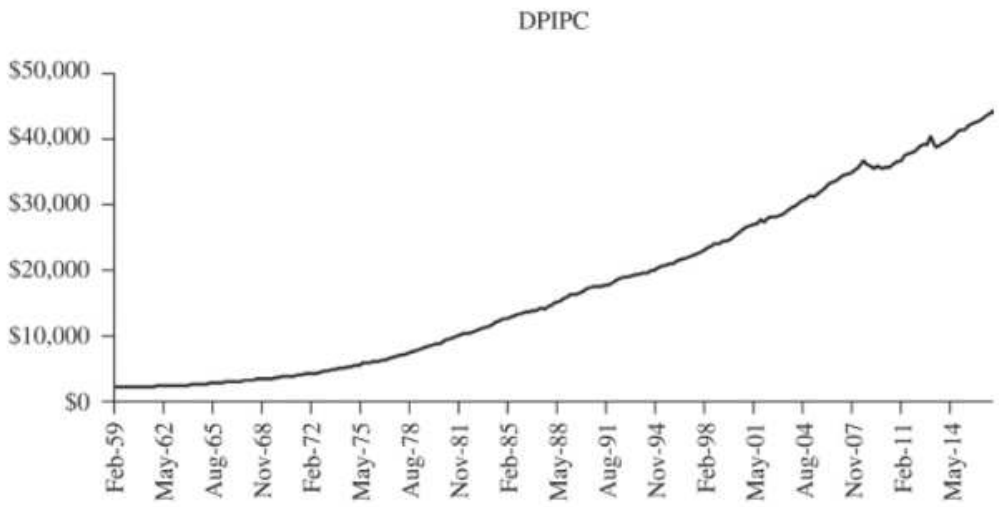
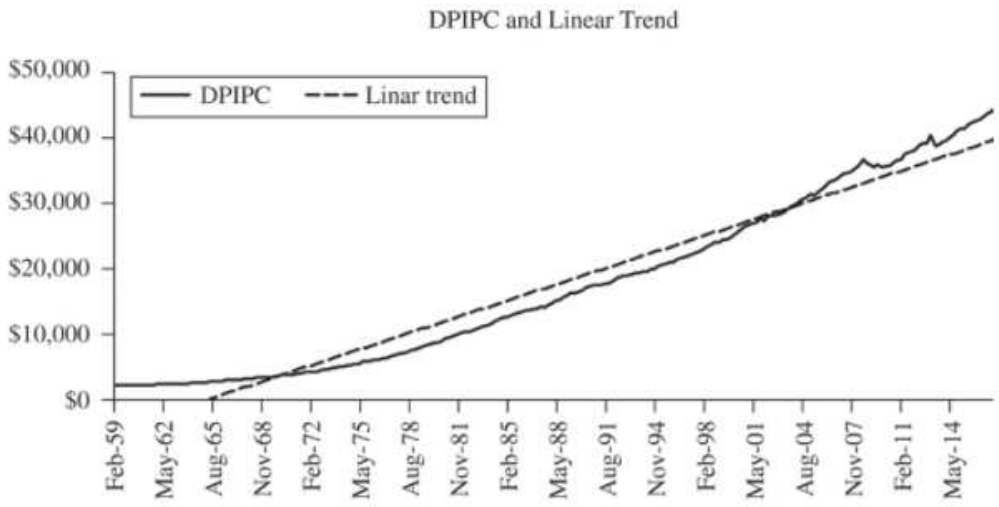
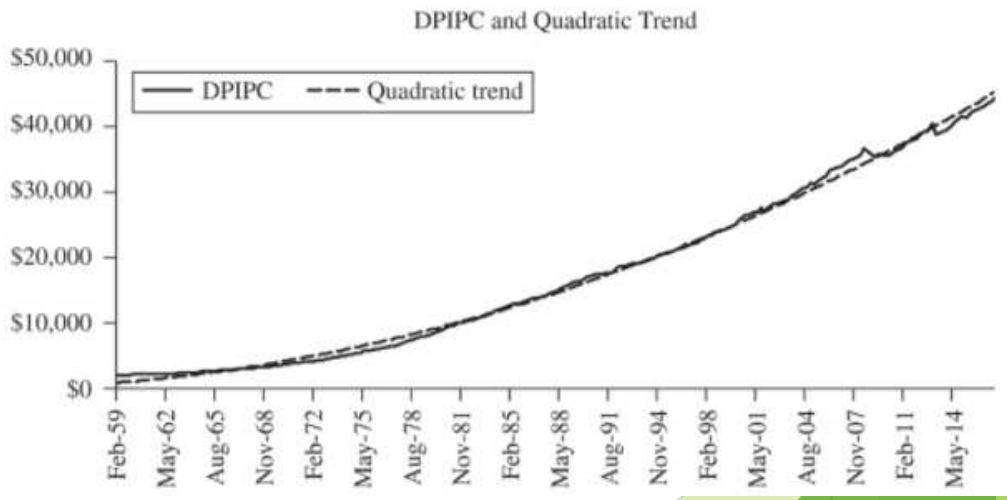

TNHS (000)
140
120
100
80
60
40
20
0
Jan-02
Sep-02
May-03
Jan-04
Sep-04
May-05
Jan-06
Sep-06
May-07
Jan-08
Sep-08
May-09
Jan-10
Sep-10
May-11
Jan-12
Sep-12
May-13
Jan-14
Sep-14
May-15
Jan-16
Sep-16
TNHS (000)
Cycle
Linear (TNHS (000))

Figure 2.3


DPIPC
$50,000
$40,000
$30,000
$20,000
$10,000
$0
Feb-59
May-62
Aug-65
Nov-68
Feb-72
May-75
Aug-78
Nov-81
Feb-85
May-88
Aug-91
Nov-94
Feb-98
May-01
Aug-04
Nov-07
Feb-11
May-14


DPIPC and Linear Trend
DPIPC
Linar trend
$50,000
$40,000
$30,000
$20,000
$10,000
$0
Feb-59
May-62
Aug-65
Nov-68
Feb-72
May-75
Aug-78
Nov-81
Feb-85
May-88
Aug-91
Nov-94
Feb-98
May-01
Aug-04
Nov-07
Feb-11
May-14


DPIPC and Quadratic Trend
DPIPC
Quadratic trend
$50,000
$40,000
$30,000
$20,000
$10,000
$0
Feb-59
May-62
Aug-65
Nov-68
Feb-72
May-75
Aug-78
Nov-81
Feb-85
May-88
Aug-91
Nov-94
Feb-98
May-01
Aug-04
Nov-07
Feb-11
May-14

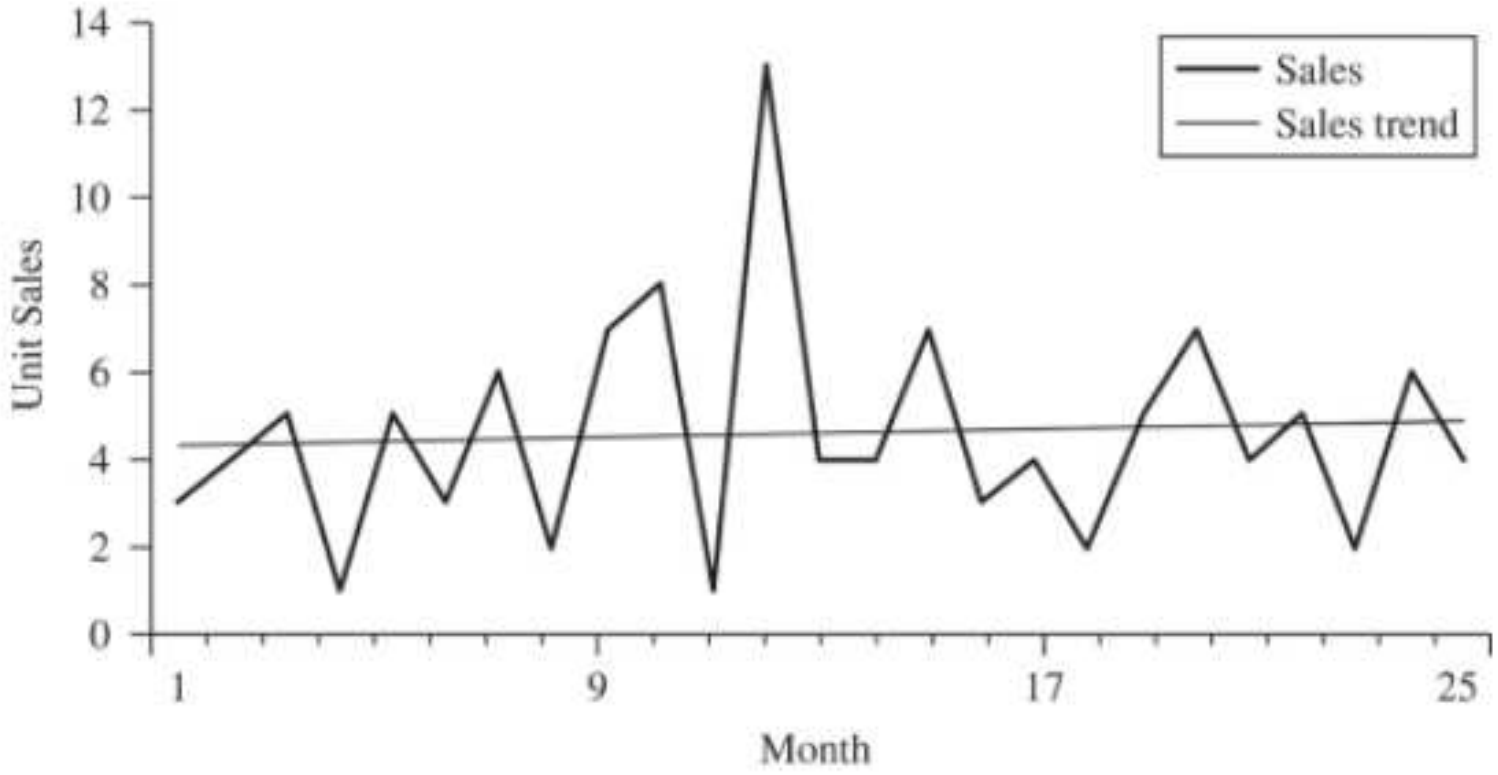
Statistical Review

Responses Arrayed from Low to High

$$\underbrace{1\ 1\ 2\ 2\ 2\ 3\ 3\ 3\ 4\ 4\ 4\ 4}_{\text{12 Values}} \quad \underset{\substack{\uparrow \\ \downarrow \\ \text{Median}}}{4} \quad \underbrace{4\ 5\ 5\ 5\ 5\ 6\ 6\ 7\ 7\ 7\ 8\ 13}_{\text{12 Values}}$$

© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Figure 2.4



© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

	For a Sample	For a Population
Standard deviation	$S = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}}$	$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$
Variance	$S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$	$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$

© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Distributions

Many everyday processes generate values of variables, and the likelihood of occurrence of specific values can be gauged with the help of one or another of certain common probability distributions, such as the **normal distribution** or the **student's t-distribution**.

© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Distributions…

The **distribution** of a variable **is a description** of the relative numbers of times each possible outcome will occur in a number of trials.

The function describing the probability that a given value will occur is called the probability function (or probability density function, abbreviated PDF).

© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Distributions…

The population of a continuous variable has an underlying distribution.

Throwing enough rocks at a spot on the earth will soon form a pile that would then represent the distribution of the process of the particular rock thrower throwing rocks at that spot from a fixed distance.

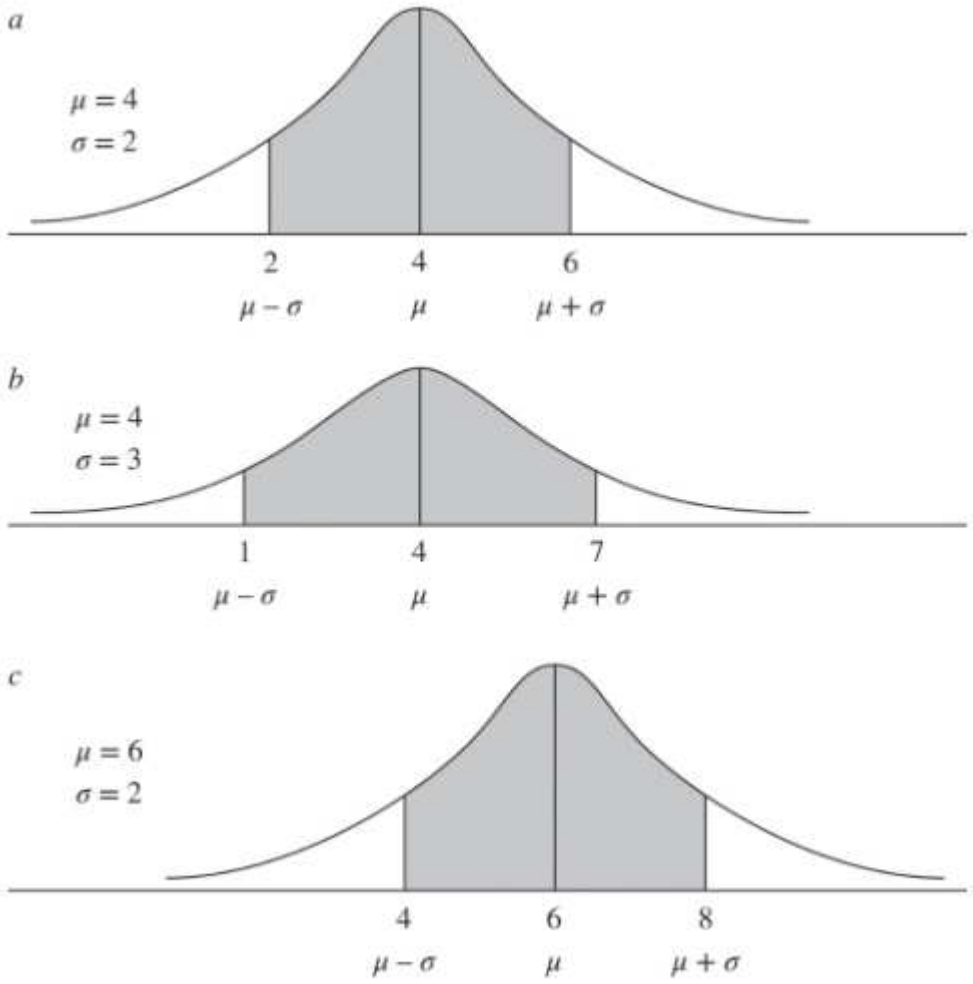
If the rocks are magnetic the pile would be more peaked (higher kurtosis).

If the distance to the point is increased, the pile will be less peaked and likely have a more skewed, less bell shape (normal/gaussian) distribution.

If half the rocks are thrown by one person and the other half by another, the pile is likely to have two peaks (bi-modal).

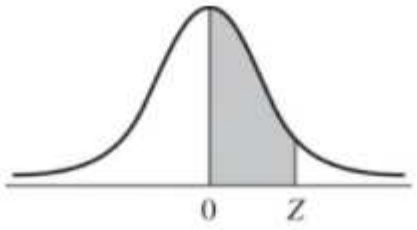
© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Figure 2.5 Three Normal Distributions



© 2019 McGraw-Hill Education. All rights reserved
reproduction or further distribution permitted wit

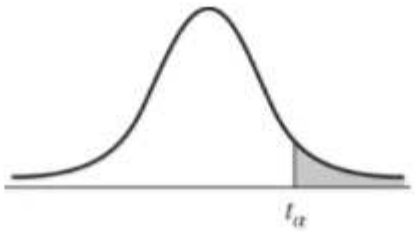
Table 2.4 Standard Normal Distribution



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2109	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.2051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621

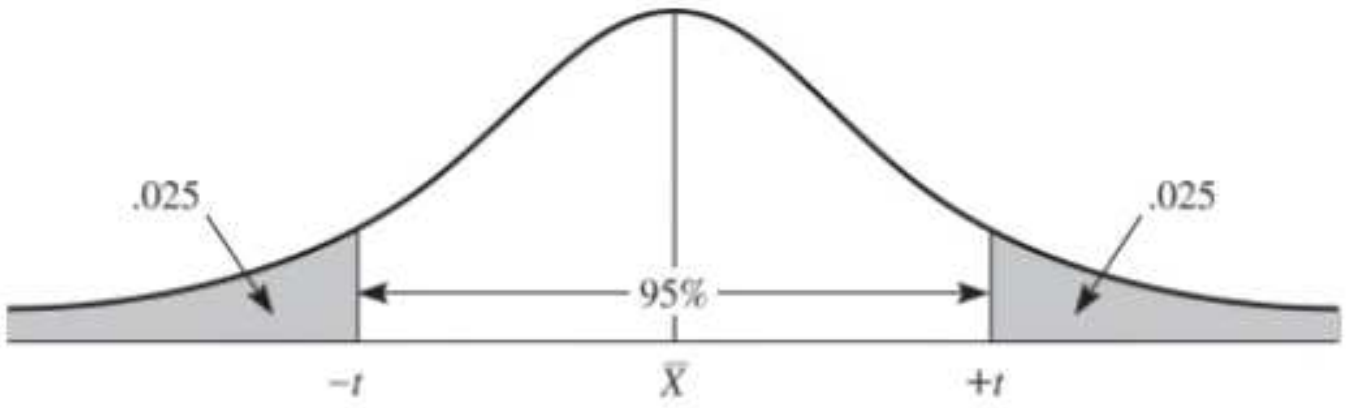
A common bell-shaped curve used to calculate probabilities of events that tend to occur around a mean value (zero for the standard normal) and trail off with decreasing likelihood. Also called Gaussian Distribution. Values are grouped about a central value with symmetrical variance (bell curve).

Figure 2.5 The Student's t-Distribution



df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

When the population standard deviation is not known, or when the sample size is small, the Student's *t*-distribution should be used rather than the normal distribution. The Student's *t*-distribution resembles the normal distribution but is somewhat more spread out for small sample sizes.



Number of Degrees of Freedom	*t*-Value for 95% Confidence Interval
5	2.571
10	2.228
20	2.086
50	1.960
100	1.960

Hypothesis Testing

The process begins by setting up two hypotheses, the null hypothesis (designated H_0:) and the alternative hypothesis (designated H_1:). These two hypotheses should be structured so that they are mutually exclusive and exhaustive.

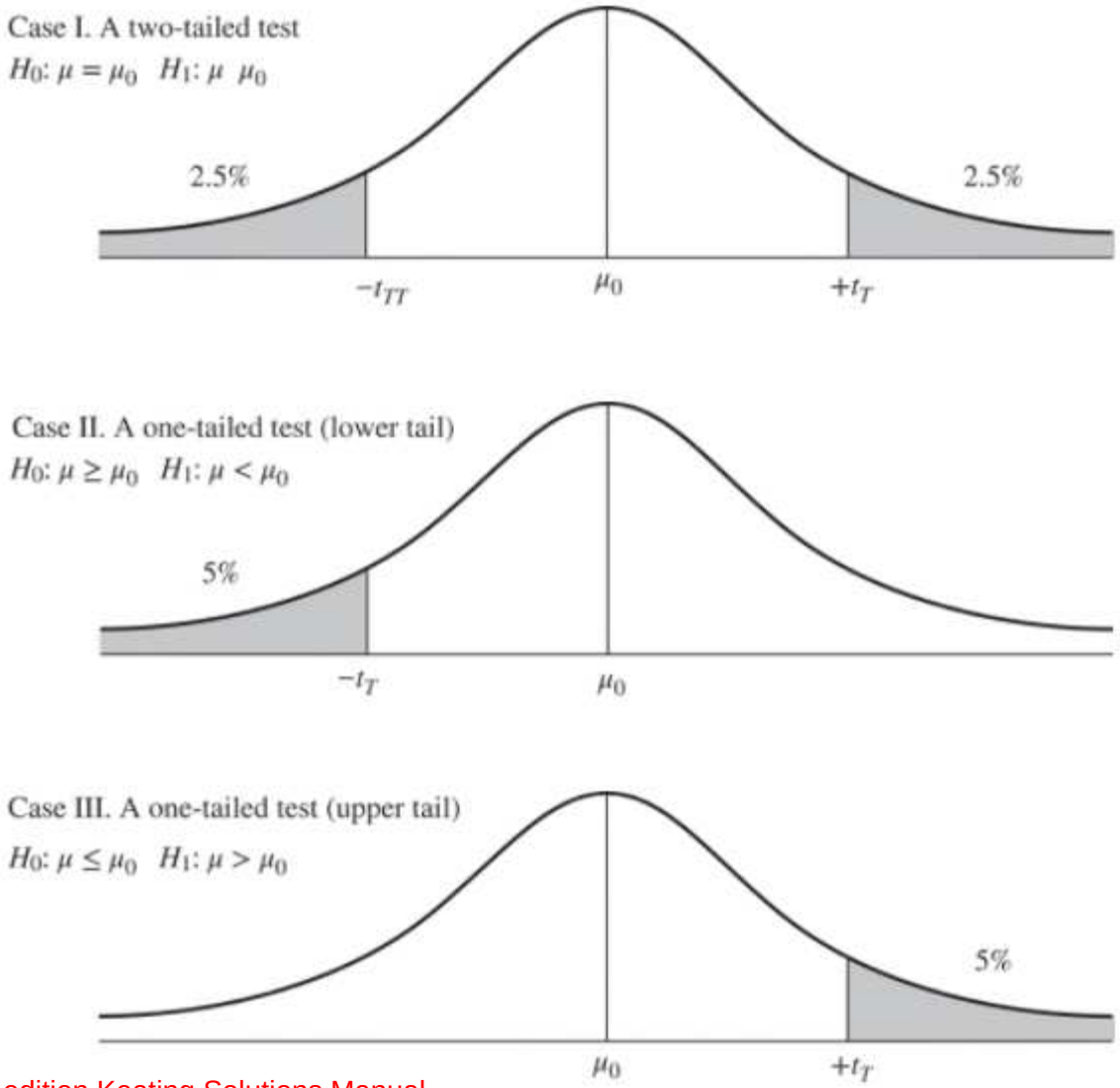
$$\text{Case I}\begin{cases} \mathrm{H}_0: & \mu = \mu_0 \\ \text{i.e., } \mathrm{H}_0: & \text{The city mean equals the national mean.} \\ \mathrm{H}_1: & \mu \neq \mu_0 \\ \text{i.e., } \mathrm{H}_1: & \text{The city mean is not equal to the national mean.} \end{cases}$$

$$\text{Case II}\begin{cases} \mathrm{H}_0: & \mu \geq \mu_0 \\ \text{i.e., } H_0: & \text{The mean for retired people is greater than or equal to the national average.} \\ H_1: & \mu < \mu_0 \\ \text{i.e., } H_1: & \text{The mean for retired people is less than the national average.} \end{cases}$$

$$\text{Case III}\begin{cases} \mathrm{H}_0: & \mu \leq \mu_0 \\ \text{i.e., } H_0: & \text{The mean for professional women is less than or equal to the standard.} \\ H_1: & \mu > \mu_0 \\ \text{i.e., } H_1: & \text{The mean for professional women is greater than the standard.} \end{cases}$$

Table 2.6 Type I and Type II Errors

	The Truth	
Statistical Decision	**H_0: Is True**	**H_0: Is Not True**
Reject H_0:	Type I error	No error
Fail to Reject H_0:	No error	Type II error



The Correlation Coefficient and its Significance

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{[\sum (X - \overline{X})^2][\sum (Y - \overline{Y})^2]}}$$

$$t = \frac{r - 0}{\sqrt{(1 - r^2)/(n - 2)}}$$

© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Representative scatterplots with the corresponding correlation coefficients. These scatterplots show correlation coefficients that range from a perfect positive correlation (*A*) and a perfect negative correlation (*B*) to zero correlations (*E* and *F*).

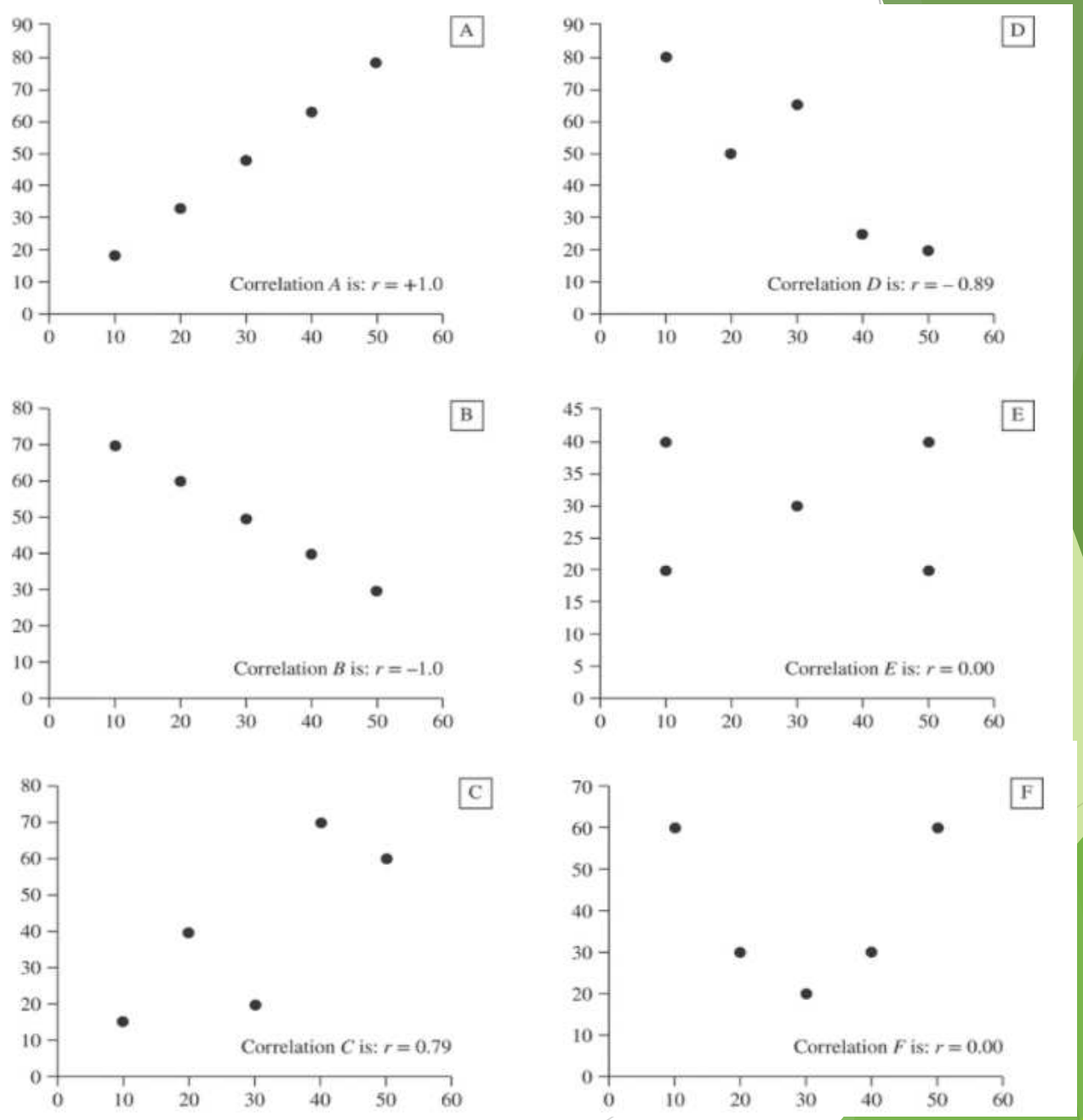


Figure 2.7

In evaluating a time series of data, it is useful to look at the correlation between successive observations over time. This measure of correlation is called an *autocorrelation*.

$$r_k = \frac{\sum_{t=1}^{n-k}(Y_{t-k} - \overline{Y})(Y_t - \overline{Y})}{\sum_{t-1}^{n}(Y_t - \overline{Y})^2}$$

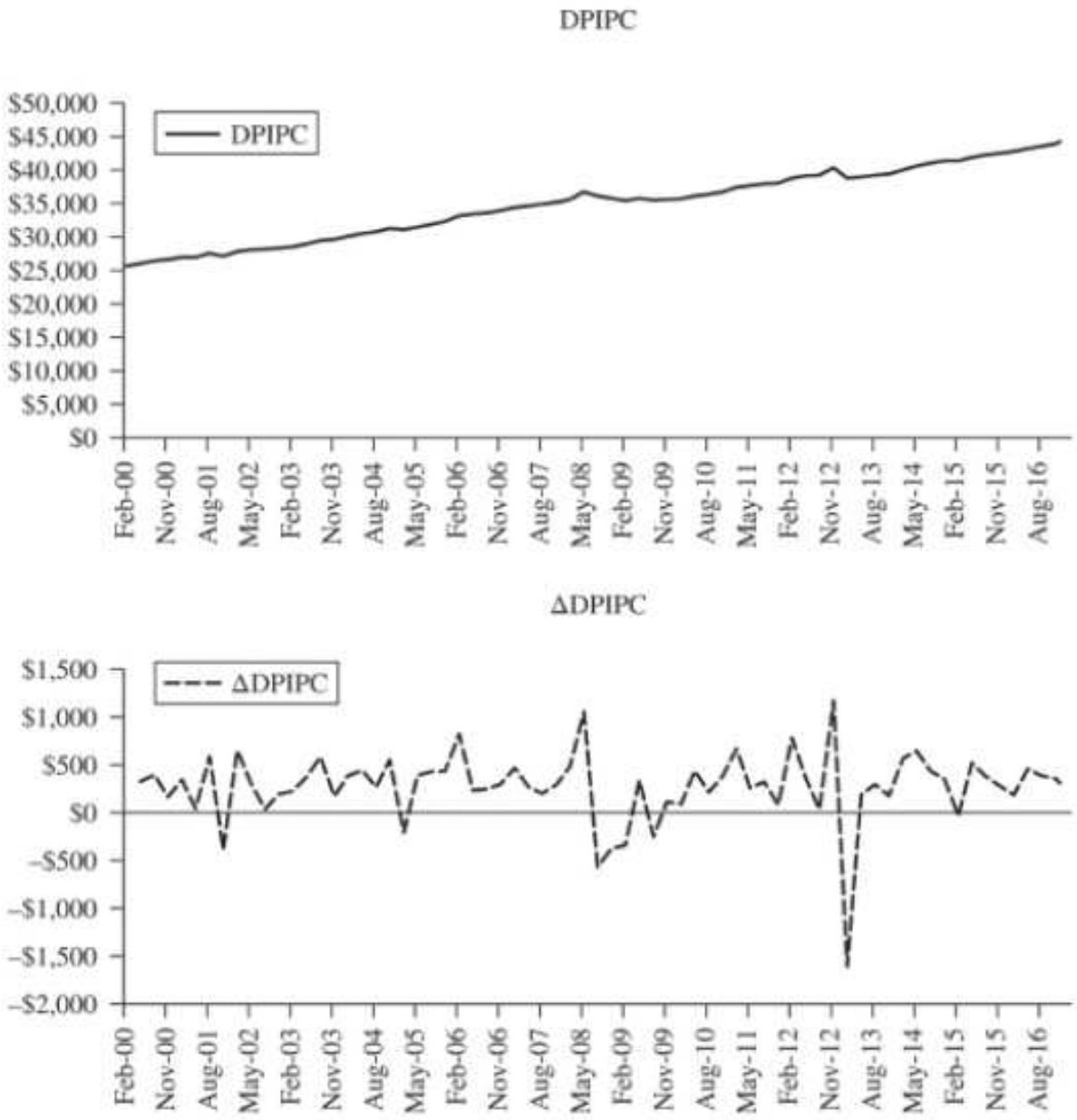
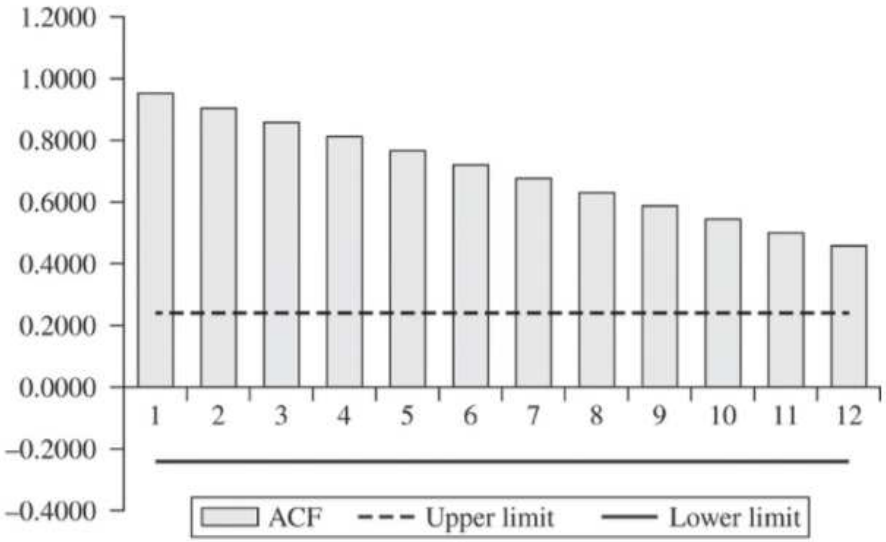


Figure 2.8 ACF Graphs for DPIPC and ΔDPIPC

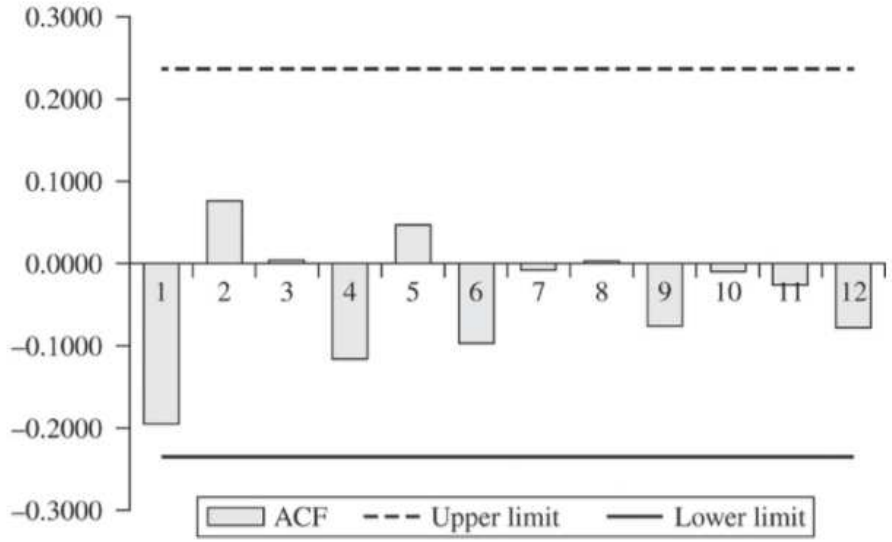
ACF Values for DPIPC

Obs	ACF
1	0.951
2	0.903
3	0.857
4	0.810
5	0.765
6	0.719
7	0.675
8	0.629
9	0.586
10	0.543
11	0.499
12	0.457



ACF Values for ΔDPIPC

Obs	ACF
1	−0.195
2	0.076
3	0.004
4	−0.116
5	0.047
6	−0.097
7	−0.008
8	0.003
9	−0.076
10	−0.010
11	−0.026
12	−0.078



ACF is the correlation between points separated by time periods.

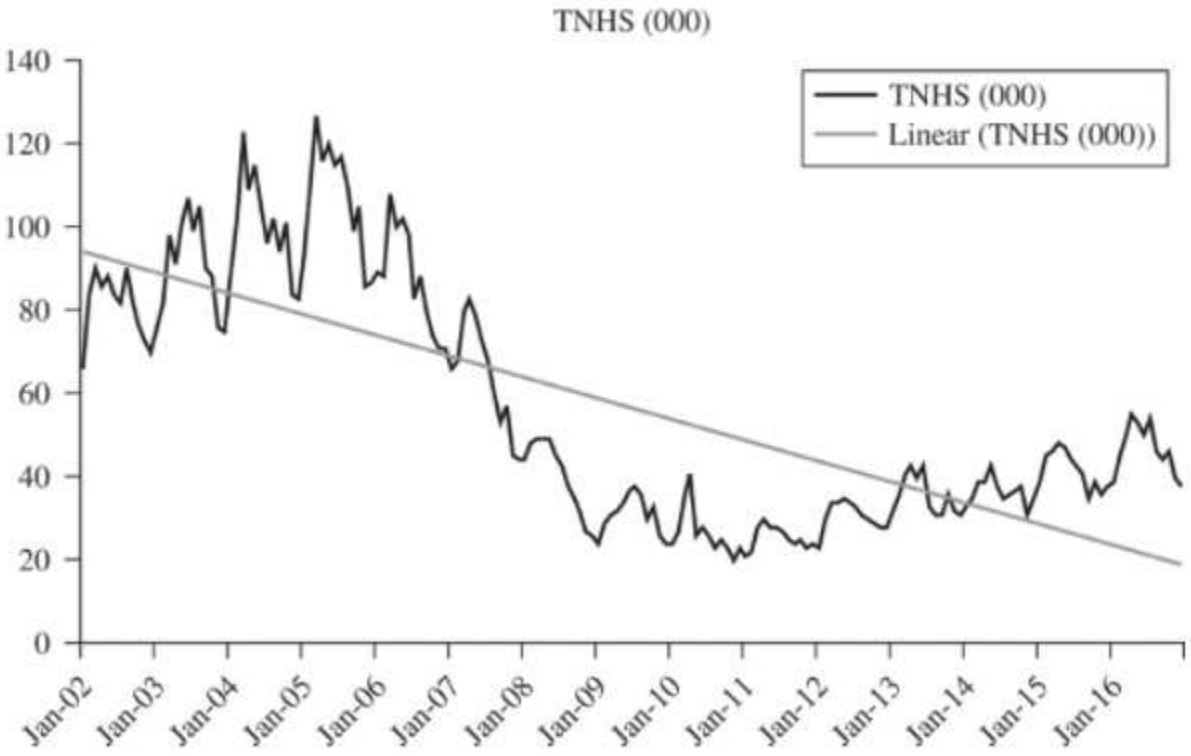
$$r_k = \frac{\sum_{t=1}^{n-k}(Y_{t-k} - \overline{Y})(Y_t - \overline{Y})}{\sum_{t-1}^{n}(Y_t - \overline{Y})^2}$$

PACF is the ACF for the *change* in the variable lagged.

Figure 2.9 Total New Houses Sold

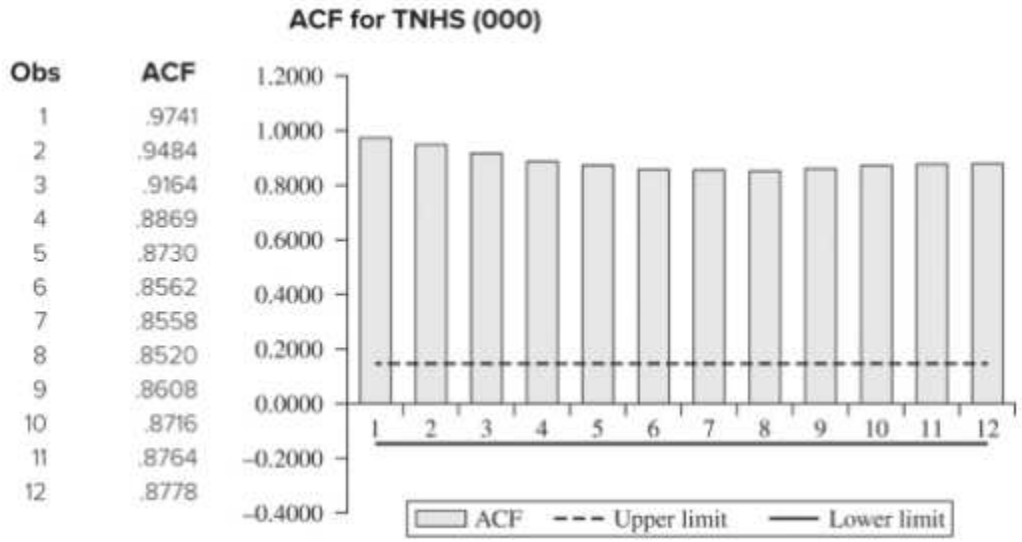
In evaluating a time series of data, it is useful to look at the correlation between successive observations over time. This measure of correlation is called an *autocorrelation*.

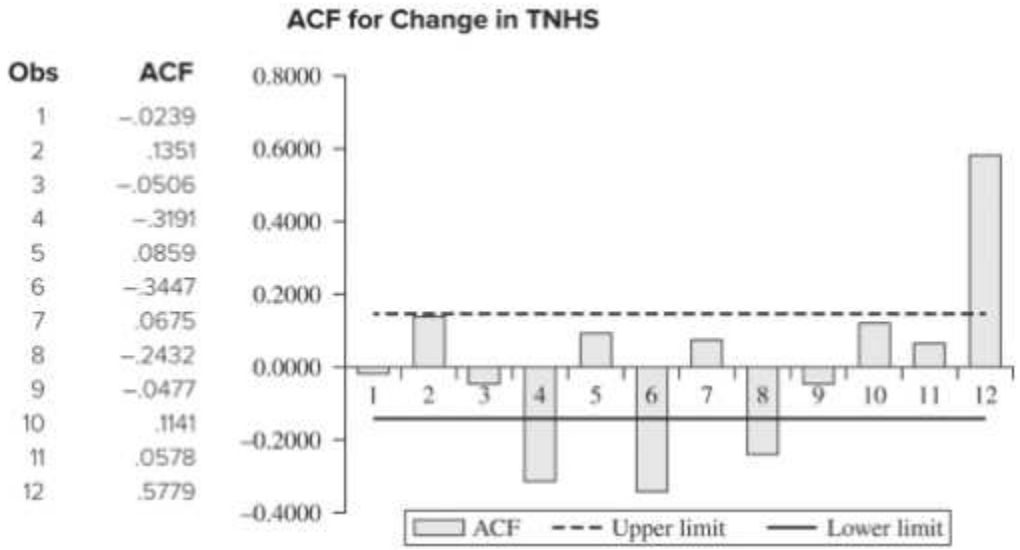
$$r_k = \frac{\sum_{t=1}^{n-k}(Y_{t-k} - \overline{Y})(Y_t - \overline{Y})}{\sum_{t-1}^{n}(Y_t - \overline{Y})^2}$$



© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

Figure 2.10 ACF Graphs for TNHS and ΔTNHS





ACF is the correlation between points separated by time periods.

$$r_k = \frac{\sum_{t=1}^{n-k}(Y_{t-k} - \overline{Y})(Y_t - \overline{Y})}{\sum_{t-1}^{n}(Y_t - \overline{Y})^2}$$

PACF is the ACF for the *change* in the variable lagged.

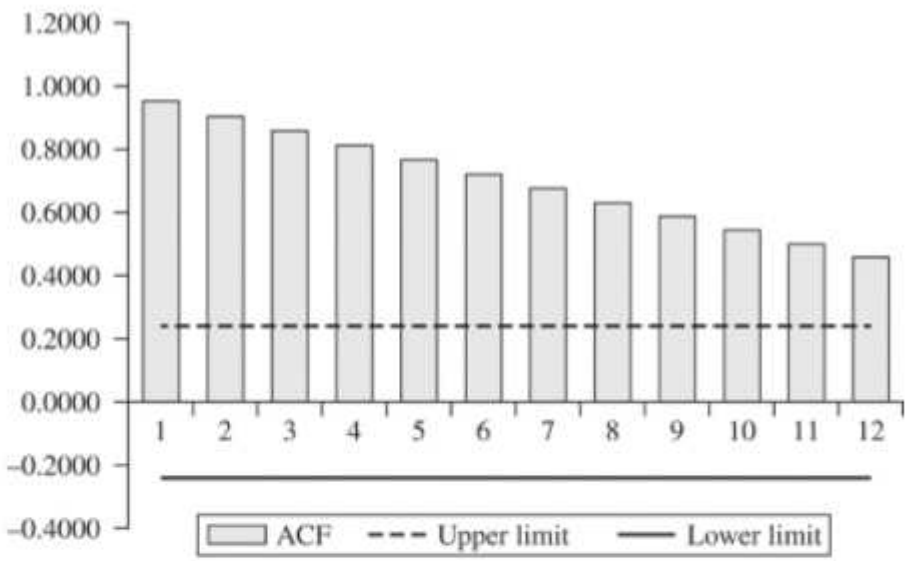
Figure 2.8

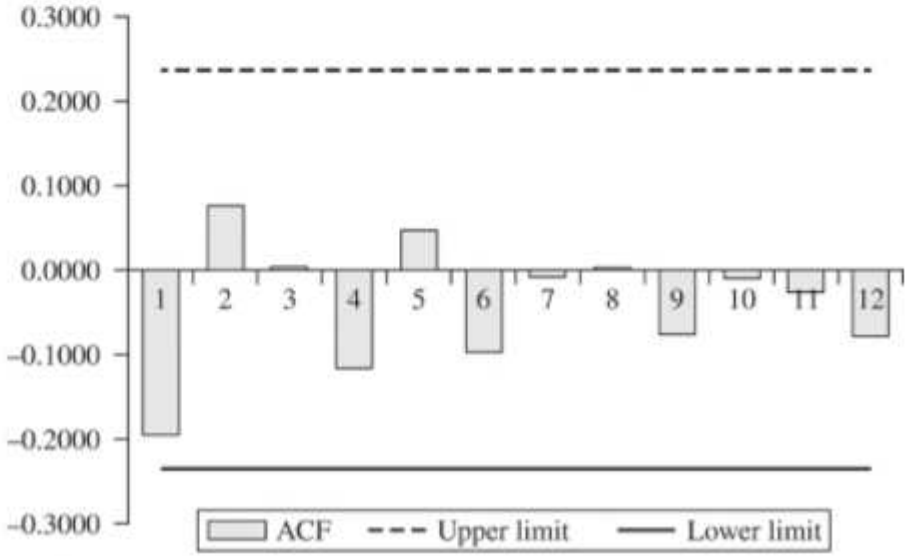
ACF is the correlation between points separated by time periods.

PACF is the ACF for the *change* in the variable lagged.

ACF Values for DPIPC

Obs	ACF
1	0.951
2	0.903
3	0.857
4	0.810
5	0.765
6	0.719
7	0.675
8	0.629
9	0.586
10	0.543
11	0.499
12	0.457



ACF Values for ΔDPIPC

Obs	ACF
1	−0.195
2	0.076
3	0.004
4	−0.116
5	0.047
6	−0.097
7	−0.008
8	0.003
9	−0.076
10	−0.010
11	−0.026
12	−0.078