MORE PREDICTIVE ANALYTICS:

Microsoft® Excel®

que®

Conrad Carlberg

FREE SAMPLE CHAPTER

SHARE WITH OTHERS

More Predictive Analytics: Microsoft Excel

Conrad Carlberg

Que®
800 E. 96th Street
Indianapolis, Indiana 46240

Contents at a Glance

Introduction 1
1 Smoothing and Its Alternatives 5
2 Diagnosing Trend and Seasonality 21
3 Working with Trended Time Series 59
4 Initializing Forecasts 91
5 Working with Seasonal Time Series 125
6 Names, Addresses, and Formulas 171
7 Multiplicative and Damped Trend Models 209
Index 241

More Predictive Analytics: Microsoft Excel

Copyright © 2016 by Que Publishing

All rights reserved. No part of this book shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher. No patent liability is assumed with respect to the use of the information contained herein. Although every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions. Nor is any liability assumed for damages resulting from the use of the information contained herein.

ISBN-13: 978-0-7897-5614-5
ISBN-10: 0-7897-5614-5

Library of Congress Control Number: 2015941441

Printed in the United States of America

First Printing: August 2016

Trademarks

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Que Publishing cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Warning and Disclaimer

Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied. The information provided is on an "as is" basis. The author and the publisher shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book.

Special Sales

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Editor-in-Chief
Greg Wiegand

Acquisitions Editor
Joan Murray

Development Editor
Charlotte Kughen

Managing Editor
Sandra Schroeder

Project Editor
Seth Kerney

Copy Editor
Cheri Clark

Indexer
Ken Johnson

Proofreader
Megan Wade

Technical Editor
Michael Turner

Editorial Assistant
Cindy Teeters

Cover Designer
Mark Shirar

Compositor
Jake McFarland

Contents

Introduction 1

1 Smoothing and Its Alternatives 5

Regression in Forecasting 6
- Curvilinear Relationships 7
- Piecewise Regression 8
- Polynomial Regression 9
- TREND() and LINEST() 13
- Using the Chart's Trendline 14

Regression or Smoothing? 16
- Comparing Smoothing with Regression with Forecasting Errors 17
- Additional Rationales for Smoothing 17
- Pushing the Forecast Horizon 19

2 Diagnosing Trend and Seasonality 21

Inside the Autocorrelation 22
- The Magnitude and the Direction of a Correlation 23
- The Definitional Formula for Pearson's Correlation 23
- Autocorrelation and Averages 26
- Looking at Individual Cases 26

Testing for Trend in the Observations 30
- Understanding the Meaning of ACFs 33
- Interpreting the ACFs 35
- Using the ACF Add-in 37
- Testing All the ACFs at Once 39

Testing for Seasonality in a Time Series 42
- Removing Trend 43
- Independent Residuals 49
- Analyzing the Regression with LINEST() 52
- Applying the Models Comparison Approach 56
- Time Series Diagnostics and the Analysis of Covariance 58

3 Working with Trended Time Series 59

Smoothing: The Basic Idea 59
- The Smoothing Equation's Error Correction Form 60
- Using Excel's Solver to Choose Alpha 62
- The Equation's Smoothing Form 64

About the Smoothing Approaches to Forecasting 65
- A Time Series' Pattern 65
- Additive and Multiplicative Models 66
- Initializing Values 67
- Initializing Versus Validation 68
- Terminology Problems 69
- More Smoothing Terminology 70

Dealing with Trend: Holt's Linear Exponential Smoothing 70
    Dealing with Trend by Differencing 73
    Forecasting the Differences 75
Using Holt's Linear Exponential Smoothing 76
    Forecasting a Level 77
    Forecasting a Trend 79
    Forecasting the Next Level and the Next Trend 80
    Extending the Forecast 81
    Optimizing the Constants 82
    Going Beyond the Next-Step-Ahead Forecast 84
    Validating the Results 86

4 Initializing Forecasts 91
Setting Initial Values 92
Getting an Initial Forecast with Backcasting 95
Getting an Initial Forecast with Optimization 99
    Initial Forecast in Simple Exponential Smoothing 99
    Initial Forecasts with Holt's Method 106
    Understanding the Relationship Between Holt's Method and Linear Regression 113
    About Multistart and Solver 116
Initializing with Regression 117
    Arranging for Curvilinear Regression 118
    Deriving the Regression Equation 119
    Confirming the Regression Equation 121
Deciding to Use Polynomial Regression 122

5 Working with Seasonal Time Series 125
Simple Seasonal Averages 125
    Identifying a Seasonal Pattern 126
    Calculating Seasonal Indexes 127
    Forecasting from Simple Seasonal Averages: No Trend 128
    Simple Seasonal Averages with Trend 129
    Evaluating Simple Averages 136
Moving Averages and Centered Moving Averages 137
    Using Moving Averages Instead of Simple Averages 138
    Understanding Specific Seasonals 140
    Aligning the Moving Averages 141
    Centered Moving Averages with Even Numbers of Seasons 141
    Detrending the Series with Moving Averages 143
Linear Regression with Coded Vectors 147
    About Effect Coding and Regression 148
    Effect Coding with Seasons 149
    Setting Up the Coded Vectors 153
    Dummy Coding 155

Simple Seasonal Exponential Smoothing....156
About the Level Smoothing Formulas....161
About the Season Smoothing Formulas....162
Dealing with the End of the Time Series....163
Holt-Winters Models....164
Initializing the Forecasts....166
Starting the Forecasts Earlier....169

6 Names, Addresses, and Formulas....171

We Interrupt This Program.......173
Implicit Intersections....174
The Relative Range Name....175
Establishing Names for Forecasting Formulas....178
Establishing the Fixed Range Names....180
Establishing the Mixed Range Names....182
Using the Names in Forecasting Formulas....184
Interpreting the Formula for Series Level....184
Interpreting the Formula for the Seasonal Effect....187
Interpreting the Forecast Formula....189
Deriving the Error Correction Formulas....190
The Level Equation....191
The Trend Equation....192
The Season Equation....194
Demonstrating the Formulas on the Worksheet....195
Deriving the Formulas on the Worksheet....199
Deriving Error Correction of Level and Season Effects....199
Deriving Error Correction of Level and Trend....201
Named Ranges for Holt-Winters Models....203
Estimating the Series Level in Holt-Winters Models....204
Estimating the Current Trend....205
Estimating the Current Seasonal Effect....205
Error Correction Formulas for Holt-Winters Models....206

7 Multiplicative and Damped Trend Models....209

About Multiplicative Models....209
Additive Models and Multiplicative Models....210
Why Use an Additive Model?....213
Multiplicative and Additive Models Compared....213
Comparing the Level Estimates....214
Comparing the Estimates of Seasonal Effects....217
Producing the Next Forecast....217
Using the Error Correction Formulas....220

Additive and Multiplicative Seasonal Models in Holt-Winters Analysis 222
    Holt-Winters Additive Models Revisited 224
    Multiplicative Holt-Winters Models 226
    Changing the Initial Seasonal Estimates 227
    Modifying the Level and Seasonal Smoothing Formulas 227
    Modifying the Trend Smoothing Formula 230
    The Multiplicative Model with Error Correction Formulas 232
Damped Trend Forecasts 234
    Formulas for the Damped Trend 236
    Implementing the Damped Trend on the Worksheet 238

Index 241

About the Author

Conrad Carlberg (www.conradcarlberg.com) is a nationally recognized expert on quantitative analysis and on data analysis and management applications such as Microsoft Excel, SAS, and Oracle. He holds a Ph.D. in statistics from the University of Colorado and is a many-time recipient of Microsoft's Excel MVP designation.

Carlberg is a Southern California native. After college he moved to Colorado, where he worked for a succession of startups and attended graduate school. He spent two years in the Middle East, teaching computer science and dodging surly camels. After finishing graduate school, Carlberg worked at US West (a Baby Bell) in product management and at Motorola.

In 1995 he started a small consulting business, which provides design and analysis services to companies that want to guide their business decisions by means of quantitative analysis—approaches that today we group under the term "analytics." He enjoys writing about those techniques and, in particular, how to carry them out using the world's most popular numeric analysis application, Microsoft Excel.

Dedication

In loving memory of Peter M. Messer, 1942-2015.

We Want to Hear from You!

As the reader of this book, *you* are our most important critic and commentator. We value your opinion and want to know what we're doing right, what we could do better, what areas you'd like to see us publish in, and any other words of wisdom you're willing to pass our way.

We welcome your comments. You can email or write to let us know what you did or didn't like about this book—as well as what we can do to make our books better.

Please note that we cannot help you with technical problems related to the topic of this book.

When you write, please be sure to include this book's title and author as well as your name and email address. We will carefully review your comments and share them with the author and editors who worked on the book.

Email: feedback@quepublishing.com

Mail: Que Publishing
ATTN: Reader Feedback
800 East 96th Street
Indianapolis, IN 46240 USA

Reader Services

Visit our website and register this book at quepublishing.com/register for convenient access to any updates, downloads, or errata that might be available for this book.

INTRODUCTION

In 2011, I wrote *Predictive Analytics: Microsoft Excel*. The book went into techniques that are heavily used in the field of predictive analytics but that we don't necessarily think of as "predictive."

Those techniques included logistic regression, a technique that's often used in place of least squares regression when the outcome variable is measured with categories; principal components analysis, which groups not records but variables into mutually independent components of related variables; and factor rotation, a technique that helps make the interpretation of principal components more straightforward.

These topics are important ones in predictive analytics, and I personally find them intrinsically interesting and fun to write about. But they soft-pedal the predictive part. You can use logistic regression to predict—on the basis of such variables as state of residence, amount loaned, and annual income—who is likely to repay a loan and who isn't. But that's not the main purpose of logistic regression.

You can use principal components and factor analysis to reduce a database that contains hundreds of related variables down to a manageable few, without significant loss of information. Doing so often makes a forecast feasible, when—using hundreds of variables—it would not have been sensible even to attempt one.

So these techniques get you ready to make predictions, but none of them actually looks at the history of a diagnosis-related component in a given hospital and tells you how many patients to expect next quarter. None looks at a principal component that quantifies customers' interest in particular types of products and predicts what they'll spend on given products next year. Any of those techniques might help you develop useful information about which sorts of patients are likely to develop symptoms, without giving you a clue about when that will happen.

In that earlier book on predictive analytics, I did discuss the use of regression, moving averages, and smoothing to predict upcoming results. I also discussed the use of Box-Jenkins techniques to identify different sorts of time series patterns: autoregressive and moving average. But there just wasn't enough room in the book to cover the handling of time series that occur frequently and that are among the most interesting: series that are trending up or down, and series that have distinct seasonal components.

So I decided to focus this book on the predictive end of things. That still leaves some choices because even from 30,000 feet you can see three distinct approaches to quantitative forecasting: least squares regression, smoothing, and autoregressive integrated moving averages (or *ARIMA*).

It made sense to me that this book's topic ought to be the second of those three, smoothing. I didn't want to spend my time and yours on simple regression, whether multiple or not. Projecting an outcome variable forward in time based on its relationship to a predictor variable has limited applicability. This is particularly true when you begin to consider all the reasons that it can be a bad idea.

Spurious correlation is one such reason. That effect takes place when two variables that are not causally related are nevertheless correlated because each rises (or falls) as time passes. Suppose the per capita incidence of a disease has risen each year during the past decade and the per capita incidence of vaccination has risen similarly. Standard least squares regression might convince you that the correlation is too strong to be due to anything other than causation. Your analysis might argue that the vaccinations are causing the disease that they are meant to prevent.

Regression itself isn't to blame for this kind of misunderstanding, but it is susceptible to the confusion of spurious correlation. It seems to me that the opportunities for legitimate and informative use of the regression technique, at least for predictive purposes, simply don't fill a book.

That's not to say that multiple regression doesn't have a place in predictive analytics. It does, particularly when it comes to time series that are both trended and seasonal. I discuss the use of multiple regression in support of seasonal smoothing in Chapter 5, "Working with Seasonal Time Series."

ARIMA is another matter. I've been using it successfully since the mid-1980s in various business situations, from product sales to inventory control, and I intend to continue doing so. But it's not at all transparent, at least when it comes to the more complex models. ARIMA analysis can become heavily loop-dependent, and that makes it difficult to visualize specifically what's going on in the black box. The loop dependency also makes it difficult to discuss in the context of Excel worksheets.

Certainly Excel can accommodate iterative analysis by means of Visual Basic for Applications (VBA), add-ins coded in C, and similar tools. But it's difficult to learn anything other than coding from looking at code. Excel is a nearly ideal environment for learning how numerics that are in place at time 15 affect those in place at time 16. The Excel worksheet, properly designed, can illuminate those relationships better than anything except, perhaps, for an Excel chart.

Furthermore, there's some evidence—buttressed by sentiment—that for certain common business applications ARIMA is overkill, that it's unnecessarily complex and burdensome as compared to the somewhat simpler smoothing approaches. As I noted earlier, I'm not about to give up ARIMA just because some authors think it's too much of a good thing. I particularly like having a largely objective tool in the form of ARIMA's correlograms to help me select a model. But I'd just as soon wait for the jury to weigh in some more before I take the time to write another book about ARIMA analysis, or ask you to read it.

I do ask you to read this one. Smoothing approaches have come under some criticism since the mid-1980s. They've been called ad hoc, not grounded in careful statistical theory, and there's some truth to that. The mathematical and statistical groundwork that underlies, for example, multivariate statistical analysis outweighs the theory behind smoothing many times over.

Another drawback is the variety of methods touted by writers and practitioners of prediction via smoothing. For example, smoothing requires that forecasts be *initialized*. You can take many different routes to arrive at those initial values, and your choice can have a significant impact on your forecasts—particularly early in the time series. There's little in the way of theory to guide you here, and the result is that many sources use different methods, often without explanation, to initialize their forecasts. A stronger theoretical superstructure would help steer a path through the rubble.

And yet forecasts from smoothing have turned out to be accurate far more than you might expect of a technique that's so transparent and intuitively appealing. Excel even has a tool in its Data Analysis add-in that performs a (very rudimentary) version of simple smoothing. So this book pays only brief attention to the basics of smoothing. I assume that you know something about what it is, and I review the very basics primarily to make sure that we're using the same vocabulary. (If you don't yet have simple exponential smoothing in your bag of tricks, you might want to check out that *Predictive Analytics: Microsoft Excel* book I mentioned earlier.)

The topics of diagnosing time series (for example, to decide whether a time series is in fact trended), dealing with trend, dealing with seasonality, using additive or multiplicative models, and extending damped trend forecasts constitute most of this book.

You won't find any VBA here, for good or ill. (One exception: There's an Excel workbook named `ACF.xls`, which contains code to calculate and chart autocorrelation functions for you.) Everything I have to say about the smoothing part of predictive analytics can take place on the worksheet and can be illustrated in Excel charts. You can download all the Excel workbooks, and there's one for each chapter, from the publisher's website. That's so you don't have to enter data to explore the examples in this book.

After you've downloaded them, I suggest that you open the workbook for Chapter 1, "Smoothing and Its Alternatives," and then have a look at Chapter 1 itself, which has more to say about how regression and smoothing differ and how they help one another out.

Moving Averages and Centered Moving Averages

A couple of points about seasonality in a time series bear repeating, even if they seem obvious. One is that the term “season” does not *necessarily* refer to the four seasons of the year that result from the tilting of the Earth’s axis. In predictive analytics, “season” often means precisely that, because many of the phenomena that we study do vary along with the progression of spring through winter: sales of winter or summer gear, incidence of certain widespread diseases, weather events caused by the location of the jet stream and changes in the temperature of the water in the eastern Pacific ocean, and so on.

Equally, events that occur regularly can act like meteorological seasons, even though they have only a tenuous connection to the solstices and equinoxes. Eight-hour shifts in hospitals and factories often get expressed in the incidence of intakes and expenditures of energy; there, a season is eight hours long and the seasons cycle every day, not every year. Due dates for taxes signal the beginning of a flood of dollars into municipal, state, and federal treasuries; there, the season might be one year long (personal income taxes), six months (property taxes in many states), quarterly (many corporate taxes), and so on.

NOTE It’s a little strange that we have the word “season” to refer generally to the regularly recurring period of time, but no general term for the period of time during which one full turn of the seasons occurs. “Cycle” is possible, but in analytics and forecasting that term is usually taken to mean a period of indeterminate length, such as a business cycle. In the absence of a better term, I’ve used “encompassing period” in this and subsequent chapters.

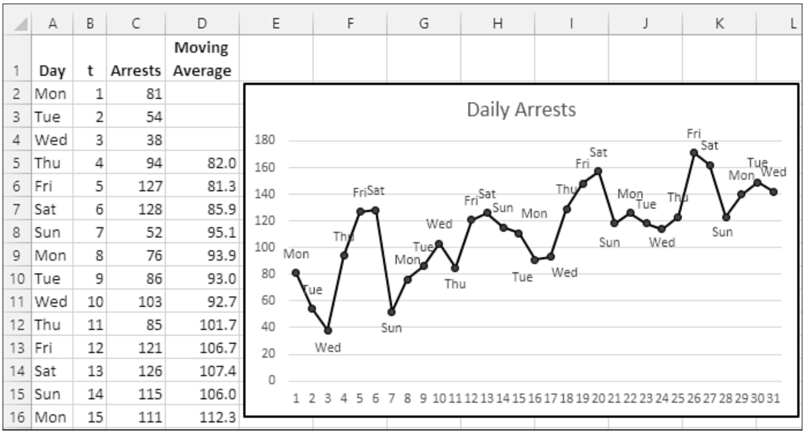
This isn’t just terminological musing. The ways we identify seasons and the period of time during which the seasons turn have real, if often minor, implications for how we measure their effects. The following sections discuss how some analysts vary the way they calculate moving averages according to whether the number of seasons is odd or even.

Using Moving Averages Instead of Simple Averages

Suppose that a large city is considering the reallocation of its traffic police to better address the incidence of driving while impaired, which the city believes has been increasing. Four weeks ago, new legislation went into effect, legalizing the possession and recreational use of marijuana. Since then, the daily number of traffic arrests for DWI seems to be trending up. Complicating matters is the fact that the number of arrests appears to spike on Fridays and Saturdays. To help plan for manpower requirements into the future, you'd like to forecast any underlying trend that's being established. You'd also like to time the deployment of your resources to take account of any weekend-related seasonality that's taking place.

Figure 5.9 has the relevant data you have to work with.

Figure 5.9
With this data set, each day of the week constitutes a season.



	A	B	C	D
1	Day	t	Arrests	Moving Average
2	Mon	1	81	
3	Tue	2	54	
4	Wed	3	38	
5	Thu	4	94	82.0
6	Fri	5	127	81.3
7	Sat	6	128	85.9
8	Sun	7	52	95.1
9	Mon	8	76	93.9
10	Tue	9	86	93.0
11	Wed	10	103	92.7
12	Thu	11	85	101.7
13	Fri	12	121	106.7
14	Sat	13	126	107.4
15	Sun	14	115	106.0
16	Mon	15	111	112.3

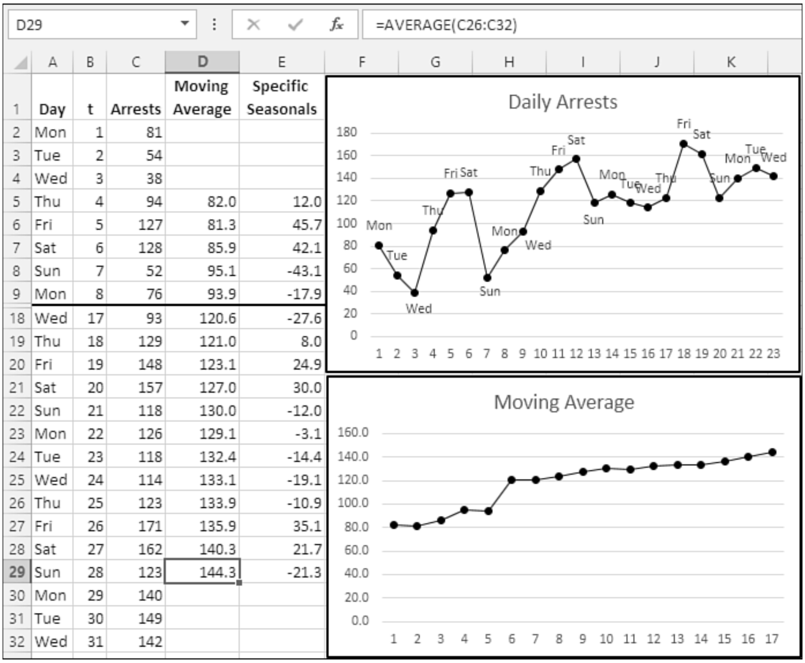
Even by just eyeballing the chart in Figure 5.9, you can tell that the trend of the number of daily arrests is up. You'll have to plan to expand the number of traffic officers, and hope that the trend levels off soon. Further, the data bears out the notion that more arrests occur routinely on Fridays and Saturdays, so your resource allocation needs to address those spikes.

But you need to quantify the underlying trend, to determine how many additional police you'll have to bring on. You also need to quantify the expected size of the weekend spikes, to determine how many additional police you need watching for erratic drivers on those days.

The problem is that as of yet you don't know how much of the daily increase is due to trend and how much is due to that weekend effect. You can start off by detrending the time series. Earlier in this chapter, in "Simple Seasonal Averages," you saw an example of how to detrend a time series in order to isolate the seasonal effects using the method of simple averages. In this section you'll see how to do so using moving averages—very likely, the moving-averages approach is used more often in predictive analytics than is the simple-averages approach.

There are various reasons for the greater popularity of moving averages, among them, that the moving-averages approach does not ask you to collapse your data in the process of quantifying a trend. Recall that the earlier example made it necessary to collapse quarterly averages to annual averages, calculate an annual trend, and then distribute one-fourth of the annual trend across each quarter in the year. That step was needed in order to remove trend from the seasonal effects. In contrast, the moving-averages approach enables you to detrend the time series without resorting to that sort of machination.

Figure 5.10 shows how the moving-averages approach works in the present example.


D29
=AVERAGE(C26:C32)
A
B
C
D
E
F
G
H
I
J
K
Daily Arrests
180
160
140
120
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Mon
Wed
Thu
Fri
Sat
Sun
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Mon
Tue
Wed
Moving Average
160.0
140.0
120.0
100.0
80.0
60.0
40.0
20.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

	Day	t	Arrests	Moving Average	Specific Seasonals
2	Mon	1	81		
3	Tue	2	54		
4	Wed	3	38		
5	Thu	4	94	82.0	12.0
6	Fri	5	127	81.3	45.7
7	Sat	6	128	85.9	42.1
8	Sun	7	52	95.1	-43.1
9	Mon	8	76	93.9	-17.9
18	Wed	17	93	120.6	-27.6
19	Thu	18	129	121.0	8.0
20	Fri	19	148	123.1	24.9
21	Sat	20	157	127.0	30.0
22	Sun	21	118	130.0	-12.0
23	Mon	22	126	129.1	-3.1
24	Tue	23	118	132.4	-14.4
25	Wed	24	114	133.1	-19.1
26	Thu	25	123	133.9	-10.9
27	Fri	26	171	135.9	35.1
28	Sat	27	162	140.3	21.7
29	Sun	28	123	144.3	-21.3
30	Mon	29	140		
31	Tue	30	149		
32	Wed	31	142		

Figure 5.10
The moving average in the second chart clarifies the underlying trend.

Figure 5.10 adds a moving average column, and a column for *specific seasonals*, to the data set in Figure 5.9. Both additions require some discussion.

The spikes in arrests that take place on weekends gives you reason to believe that you're working with seasons that repeat once each week. Therefore, start by getting the average for the encompassing period—that is, the first seven seasons, Monday through Sunday. The formula for the average in cell D5, the first available moving average, is as follows:

```
=AVERAGE (C2:C8)
```

That formula is copied and pasted down through cell D29, so you have 25 moving averages based on 25 runs of seven consecutive days. Notice that in order to show both the first and the last few observations in the time series, I have hidden rows 10 through 17. You can unhide them, if you want, in this chapter's workbook, available from the publisher's website.

Make a multiple selection of visible rows 9 and 18, right-click one of their row headers, and choose Unhide from the shortcut menu.

NOTE When you hide a worksheet's rows, as I've done in Figure 5.10, any charted data in the hidden rows is also hidden on the chart. The x-axis labels identify only the data points that appear on the chart.

Because each moving average in Figure 5.10 encompasses seven days, no moving average is paired with the first three or final three actual observations. Copying and pasting the formula in cell D5 up one day to cell D4 runs you out of observations—there is no observation recorded in cell C1. Similarly, there is no moving average recorded below cell D29. Copying and pasting the formula in D29 into D30 would require an observation in cell C33, and no observation is available for the day that cell would represent.

It would be possible, of course, to shorten the length of the moving average to, say, five instead of seven. So doing would mean that the moving-average formulas in Figure 5.10 could start in cell D4 instead of D5. However, in this sort of analysis, you want the length of the moving average to equal the number of seasons: seven days in a week for events that recur weekly implies a moving average of length seven, and four quarters in a year for events that recur annually implies a moving average of length four.

Along similar lines, we generally quantify seasonal effects in such a way that they total to zero within the encompassing time period. As you saw in this chapter's first section, on simple averages, this is done by calculating the average of (say) the four quarters in a year, and then subtracting the average for the year from each quarterly figure. So doing ensures that the total of the seasonal effects is zero. In turn, that's useful because it puts the seasonal effects on a common footing—a summer effect of +11 is as far from the mean as a winter effect of –11. If you want to average five seasons instead of seven to get your moving average, you're better off finding a phenomenon that repeats every five seasons instead of every seven.

NOTE However, when you take the average of the seasonal effects later in the process, those averages are unlikely to sum to zero. It's necessary at that point to recalibrate, or *normalize*, the averages so that their sum is zero. When that's done, the averaged seasonal averages express the effect on a time period of belonging to a particular season. Once normalized, the seasonal averages are termed the *seasonal indexes* that this chapter has already mentioned several times. You'll see how it works later in this chapter, in "Detrending the Series with Moving Averages."

Understanding Specific Seasonals

Figure 5.10 also shows what are called *specific seasonals* in column E. They are what's left after subtracting the moving average from the actual observation. To get a sense of what the specific seasonals represent, consider the moving average in cell D5. It is the average of the observations in C2:C8. The deviations of each observation from the moving average (for example, C2 – D5) are guaranteed to sum to zero—that's a characteristic of an average. Therefore, each deviation expresses the effect of being associated with that particular day in that particular week. It's a specific seasonal, then—specific because the deviation applies to that particular Monday or Tuesday and so on, and seasonal because in this example we're treating each day as though it were a season in the encompassing period of a week.

Because each specific seasonal measures the effect of being in that season vis-à-vis the moving average for that group of (here) seven seasons, you can subsequently average the specific seasonals for a particular season (for example, all the Fridays in your time series) to estimate that season's general, rather than specific, effect. That average is not confounded by an underlying trend in the time series, because each specific seasonal expresses a deviation from its own particular moving average.

Aligning the Moving Averages

There's also the question of aligning the moving averages with the original data set. In Figure 5.10, I have aligned each moving average with the midpoint of the range of observations that it includes. So, for example, the formula in cell D5 averages the observations in C2:C8, and I have aligned it with the fourth observation, the midpoint of the averaged range, by placing it in row 5. This arrangement is termed a *centered moving average*, and many analysts prefer to align each moving average with the midpoint of the observations that it averages.

NOTE

Bear in mind that in this context, "midpoint" refers to the middle of a time span: Thursday is the midpoint of Monday through Sunday. It does *not* refer to the median of the observed values, although of course it might work out that way in practice.

Another approach is the *trailing moving average*. In that case, each moving average is aligned with the final observation that it averages—and therefore it trails behind its arguments. This is often the preferred arrangement if you want to use a moving average as a forecast, as is done with exponential smoothing, because your final moving average occurs coincident with the final available observation.

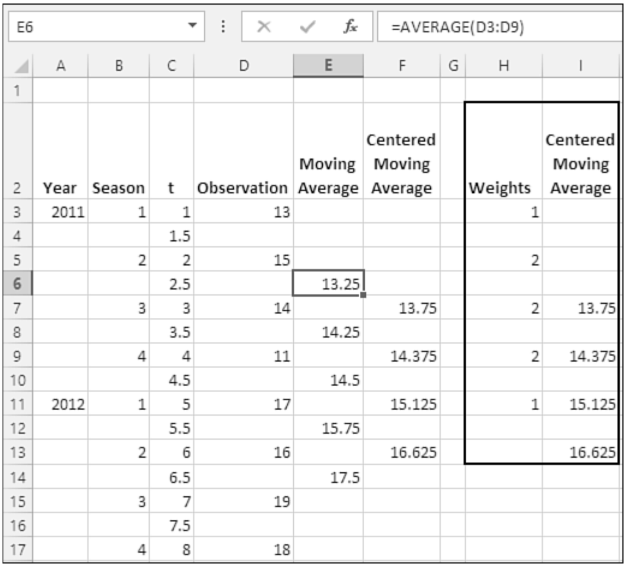
Centered Moving Averages with Even Numbers of Seasons

We usually adopt a special procedure when the number of seasons is even rather than odd. That's the typical state of affairs: There tend to be even numbers of seasons in the encompassing period for typical seasons such as months, quarters, and quadrennial periods (for elections).

The difficulty with an even number of seasons is that there is no midpoint. Two is not the midpoint of a range starting at 1 and ending at 4, and neither is 3; if it can be said to have one, its midpoint is 2.5. Six is not the midpoint of 1 to 12, and neither is 7; its purely theoretical midpoint is 6.5.

To act as though a midpoint exists, you need to add a layer of averaging atop the moving averages. See Figure 5.11.

Figure 5.11
Excel offers several ways to calculate a centered moving average.



	A	B	C	D	E	F	G	H	I
1									
2	Year	Season	t	Observation	Moving Average	Centered Moving Average		Weights	Centered Moving Average
3	2011	1	1	13				1	
4			1.5						
5		2	2	15				2	
6			2.5		13.25				
7		3	3	14		13.75		2	13.75
8			3.5		14.25				
9		4	4	11		14.375		2	14.375
10			4.5		14.5				
11	2012	1	5	17		15.125		1	15.125
12			5.5		15.75				
13		2	6	16		16.625			16.625
14			6.5		17.5				
15		3	7	19					
16			7.5						
17		4	8	18					

The idea behind this approach to getting a moving average that's centered on an existing midpoint, when there's an even number of seasons, is to pull that midpoint forward by half a season. You calculate a moving average that would be centered at, say, the third point in time if five seasons instead of four constituted one full turn of the calendar.

That's done by taking two consecutive moving averages and averaging *them*. So in Figure 5.11, there's a moving average in cell E6 that averages the values in D3:D9. Because there are four seasonal values in D3:D9, the moving average in E6 is thought of as centered at the imaginary season 2.5, half a point short of the first available candidate season, 3. (Seasons 1 and 2 are unavailable as midpoints for lack of data to average prior to Season 1.)

Note, though, that the moving average in cell E8 averages the values in D5:D11, the second through the fifth in the time series. That average is centered at (imaginary) point 3.5, a full period ahead of the average centered at 2.5. By averaging the two moving averages, so the thinking goes, you can pull the center point of the first moving average forward by half a point, from 2.5 to 3.

That's what the averages in column F of Figure 5.11 do. Cell F7 provides the average of the moving averages in E6 and E8. And the average in F7 is aligned with the third data point in the original time series, in cell D7, to emphasize that the average is centered on that season.

If you expand the formula in cell F7 as well as the moving averages in cells E6 and E8, you'll see that it turns out to be a weighted average of the first five values in the time series, with the first and fifth value given a weight of 1, and the second through fourth values given a weight of 2. That leads us to a quicker and simpler way to calculate a centered moving average with an even number of seasons.

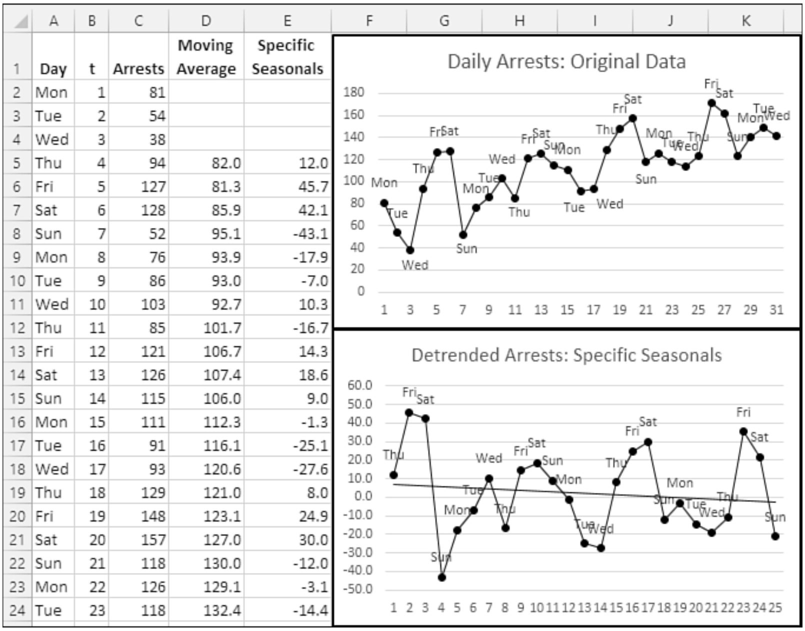
Still in Figure 5.11, the weights are stored in the range H3:H11. This formula returns the first centered moving average, in cell I7:

```
=SUMPRODUCT(D3:D11,$H$3:$H$11)/8
```

That formula returns `13.75`, which is identical to the value calculated by the double-average formula in cell F7. Making the reference to the weights absolute, by means of the dollar signs in `$H$3:$H$11`, you can copy the formula and paste it down as far as necessary to get the rest of the centered moving averages.

Detrending the Series with Moving Averages

When you have subtracted the moving averages from the original observations to get the specific seasonals, you have removed the underlying trend from the series. What's left in the specific seasonals is normally a stationary, horizontal series with two effects that cause the specific seasonals to depart from an absolutely straight line: the seasonal effects and random error in the original observations. Figure 5.12 shows the results for this example.



	A	B	C	D	E
1	Day	t	Arrests	Moving Average	Specific Seasonals
2	Mon	1	81		
3	Tue	2	54		
4	Wed	3	38		
5	Thu	4	94	82.0	12.0
6	Fri	5	127	81.3	45.7
7	Sat	6	128	85.9	42.1
8	Sun	7	52	95.1	-43.1
9	Mon	8	76	93.9	-17.9
10	Tue	9	86	93.0	-7.0
11	Wed	10	103	92.7	10.3
12	Thu	11	85	101.7	-16.7
13	Fri	12	121	106.7	14.3
14	Sat	13	126	107.4	18.6
15	Sun	14	115	106.0	9.0
16	Mon	15	111	112.3	-1.3
17	Tue	16	91	116.1	-25.1
18	Wed	17	93	120.6	-27.6
19	Thu	18	129	121.0	8.0
20	Fri	19	148	123.1	24.9
21	Sat	20	157	127.0	30.0
22	Sun	21	118	130.0	-12.0
23	Mon	22	126	129.1	-3.1
24	Tue	23	118	132.4	-14.4

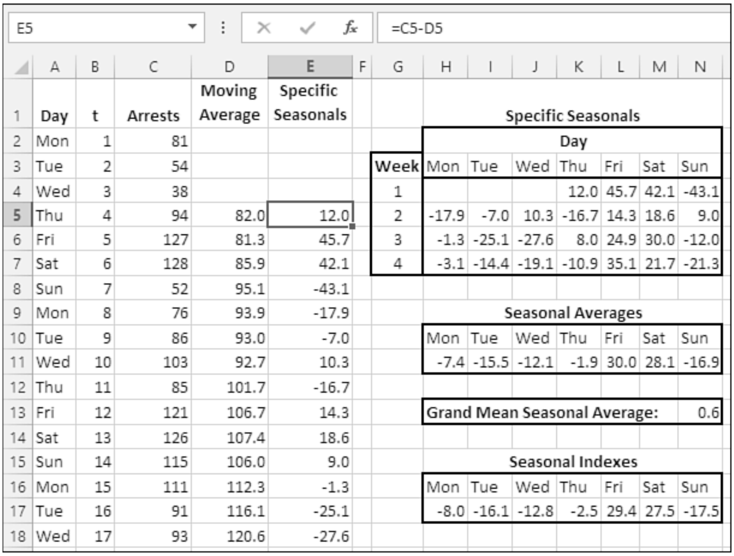
F G H I J K
Daily Arrests: Original Data
180 160 140 120 100 80 60 40 20 0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31
Mon Tue Wed Thu Fri Sat Sun
Detrended Arrests: Specific Seasonals
60.0 50.0 40.0 30.0 20.0 10.0 0.0 -10.0 -20.0 -30.0 -40.0 -50.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 5.12
The specific seasonal effects for Friday and Saturday remain clear in the detrended series.

The upper chart in Figure 5.12 shows the original daily observations. Both the general upward trend and the weekend seasonal spikes are clear.

The lower chart shows the specific seasonals: the result of detrending the original series with a moving-average filter, as described earlier in "Understanding Specific Seasonals." You can see that the detrended series is now virtually horizontal (a linear trendline for the specific seasonals has a slight downward drift), but the seasonal Friday and Saturday spikes are still in place.

The next step is to move beyond the specific seasonals to the seasonal indexes. See Figure 5.13.



E5 =C5-D5

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Day	t	Arrests	Moving Average	Specific Seasonals			Specific Seasonals						
2	Mon	1	81					Day						
3	Tue	2	54				Week	Mon	Tue	Wed	Thu	Fri	Sat	Sun
4	Wed	3	38				1				12.0	45.7	42.1	-43.1
5	Thu	4	94	82.0	12.0		2	-17.9	-7.0	10.3	-16.7	14.3	18.6	9.0
6	Fri	5	127	81.3	45.7		3	-1.3	-25.1	-27.6	8.0	24.9	30.0	-12.0
7	Sat	6	128	85.9	42.1		4	-3.1	-14.4	-19.1	-10.9	35.1	21.7	-21.3
8	Sun	7	52	95.1	-43.1									
9	Mon	8	76	93.9	-17.9			Seasonal Averages						
10	Tue	9	86	93.0	-7.0			Mon	Tue	Wed	Thu	Fri	Sat	Sun
11	Wed	10	103	92.7	10.3			-7.4	-15.5	-12.1	-1.9	30.0	28.1	-16.9
12	Thu	11	85	101.7	-16.7									
13	Fri	12	121	106.7	14.3			Grand Mean Seasonal Average:						0.6
14	Sat	13	126	107.4	18.6									
15	Sun	14	115	106.0	9.0			Seasonal Indexes						
16	Mon	15	111	112.3	-1.3			Mon	Tue	Wed	Thu	Fri	Sat	Sun
17	Tue	16	91	116.1	-25.1			-8.0	-16.1	-12.8	-2.5	29.4	27.5	-17.5
18	Wed	17	93	120.6	-27.6									

Figure 5.13
The specific seasonals effects are first averaged and then normalized to reach the seasonal indexes.

In Figure 5.13, the specific seasonals in column E are rearranged in the tabular form shown in the range H4:N7. The purpose is simply to make it easier to calculate the seasonal averages. Those averages are shown in H11:N11.

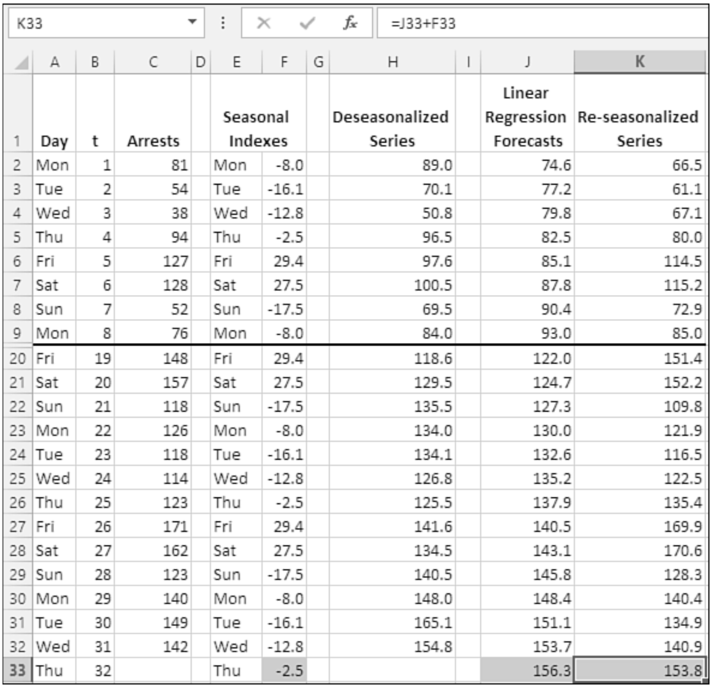
However, the figures in H11:N11 are averages, not deviations from an average, and therefore we can't expect them to sum to zero. We still need to adjust them so that they express deviations from a grand mean. That grand mean appears in cell N13, and is the average of the seasonal averages.

We can arrive at the seasonal indexes by subtracting the grand mean in N13 from each of the seasonal averages. The result is in the range H17:N17. These seasonal indexes are no longer specific to a particular moving average, as is the case with the specific seasonals in column E. Because they're based on an average of each instance of a given season, they express the average effect of a given season across the four weeks in the time series.

Furthermore, they are measures of a season's—here, a day's—effect on traffic arrests vis-à-vis the average for a seven-day period.

We can now use those seasonal indexes to deseasonalize the series. We'll use the deseasonalized series to get forecasts by way of linear regression or Holt's method of smoothing trended series (discussed in Chapter 4). Then we simply add the seasonal indexes back into the forecasts to reseasonalize them. All this appears in Figure 5.14.

Figure 5.14
After you have the seasonal indexes, the finishing touches as applied here are the same as in the method of simple averages.



1	Day	t	Arrests	Seasonal Indexes		Deseasonalized Series	Linear Regression Forecasts	Re-seasonalized Series
2	Mon	1	81	Mon	-8.0	89.0	74.6	66.5
3	Tue	2	54	Tue	-16.1	70.1	77.2	61.1
4	Wed	3	38	Wed	-12.8	50.8	79.8	67.1
5	Thu	4	94	Thu	-2.5	96.5	82.5	80.0
6	Fri	5	127	Fri	29.4	97.6	85.1	114.5
7	Sat	6	128	Sat	27.5	100.5	87.8	115.2
8	Sun	7	52	Sun	-17.5	69.5	90.4	72.9
9	Mon	8	76	Mon	-8.0	84.0	93.0	85.0
20	Fri	19	148	Fri	29.4	118.6	122.0	151.4
21	Sat	20	157	Sat	27.5	129.5	124.7	152.2
22	Sun	21	118	Sun	-17.5	135.5	127.3	109.8
23	Mon	22	126	Mon	-8.0	134.0	130.0	121.9
24	Tue	23	118	Tue	-16.1	134.1	132.6	116.5
25	Wed	24	114	Wed	-12.8	126.8	135.2	122.5
26	Thu	25	123	Thu	-2.5	125.5	137.9	135.4
27	Fri	26	171	Fri	29.4	141.6	140.5	169.9
28	Sat	27	162	Sat	27.5	134.5	143.1	170.6
29	Sun	28	123	Sun	-17.5	140.5	145.8	128.3
30	Mon	29	140	Mon	-8.0	148.0	148.4	140.4
31	Tue	30	149	Tue	-16.1	165.1	151.1	134.9
32	Wed	31	142	Wed	-12.8	154.8	153.7	140.9
33	Thu	32		Thu	-2.5		156.3	153.8

The steps illustrated in Figure 5.14 are largely the same as those in Figures 5.6 and 5.7, discussed in the following sections.

Deseasonalizing the Observations

Subtract the seasonal indexes from the original observations to deseasonalize the data. You can do this as shown in Figure 5.14, in which the original observations and the seasonal indexes are arranged as two lists beginning in the same row, columns C and F. This arrangement makes it a little easier to structure the calculations.

You can also do the subtraction as shown in Figure 5.6, in which the original quarterly observations (C12:F16), the quarterly indexes (C8:F8), and the deseasonalized results (C20:F24) are shown in a tabular format. That arrangement makes it a little easier to focus on the seasonal indexes and the deseasoned quarterlies.

Forecast from the Deseasonalized Observations

In Figure 5.14, the deseasonalized observations are in column H, and in Figure 5.7 they're in column C. Regardless of whether you want to use a regression approach or a smoothing approach to the forecast, it's best to arrange the deseasonalized observations in a single-column list.

In Figure 5.14, the forecasts are in column J. The following array formula is entered in the range J2:J32.

```
=TREND(H2:H32)
```

Earlier in this chapter, I pointed out that if you omit the x-values argument from the `TREND()` function's arguments, Excel supplies the default values `1`, `2`, ..., `n`, where `n` is the number of y-values. In the formula just given, H2:H32 contains 31 y-values. Because the argument normally containing the x-values is missing, Excel supplies the default values `1`, `2`, ..., `31`. Those are the values we would want to use anyway, in column B, so the formula as given is equivalent to `=TREND(H2:H32,B2:B32)`. And that's the structure used in D5:D24 of Figure 5.7:

```
=TREND(C5:C24,B5:B24)
```

Making the One-Step-Ahead Forecast

So far you have arranged for forecasts of the deseasonalized time series from `t = 1` through `t = 31` in Figure 5.14, and from `t = 1` through `t = 20` in Figure 5.7. These forecasts constitute useful information for various purposes, including assessing the accuracy of the forecasts by means of an RMSE analysis.

But your main purpose is forecasting at least the next, as yet unobserved time period. To get that, you could first forecast from the `TREND()` or `LINEST()` function if you're using regression, or from the exponential smoothing formula if you're using Holt's method. Then you can add the associated seasonal index to the regression or smoothing forecast, to get a forecast that includes both the trend and the seasonal effect.

In Figure 5.14, you get the regression forecast in cell J33 with this formula:

```
=TREND(H2:H32,,B33)
```

In this formula, the y-values in H2:H32 are the same as in the other `TREND()` formulas in column J. So are the (default) x-values of `1` through `32`. Now, though, you supply a new x-value as the function's third argument, which you tell `TREND()` to look for in cell B33. It's `32`, the next value of `t`. And Excel returns the value `156.3` in cell J33. The `TREND()` function in cell J33 is telling Excel, in effect, "Calculate the regression equation for the values in H2:H32 regressed on the `t` values `1` through `31`. Apply that regression equation to the new x-value of `32` and return the result."

You'll find the same approach taken in cell D25 of Figure 5.7, where the formula to get the one-step-ahead forecast is this:

```
=TREND(C5:C24,B5:B24,B25)
```

Adding the Seasonal Indexes Back In

The final step is to reseasonalize the forecasts by adding the seasonal indexes to the trend forecasts, reversing what you did four steps back when you subtracted the indexes from the original observations. This is done in column F in Figure 5.7 and column K in Figure 5.14.

Don't forget to add the appropriate seasonal index for the one-step-ahead forecast, with the results shown in cell F25 in Figure 5.7 and in cell K33 in Figure 5.14. (I've shaded the one-step-ahead cells in both Figure 5.7 and Figure 5.14 to highlight the forecasts.)

You can find charts of three representations of the traffic arrest data in Figure 5.15: the deseasonalized series, the linear forecast from the deseasonalized data, and the reseasonalized forecasts. Note that the forecasts incorporate both the general trend of the original data and its Friday/Saturday spikes.

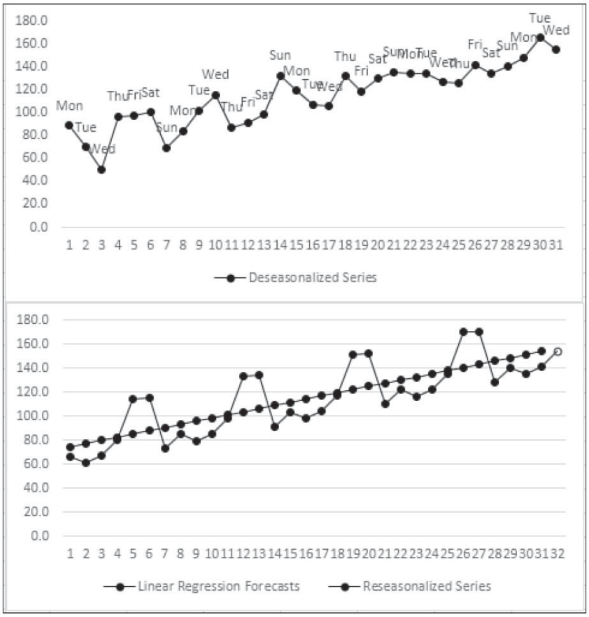


Figure 5.15
Charting the forecasts.

Index

A

ACF (autocorrelation function), 22-23
- ACF add-in, 37-38
- averages and, 26
- calculating, 33-35
- drift, 26-29
- PACF (Partial Autocorrelation Function), 37-38
- Pearson's correlation versus, 29-30
- testing all at once, 39-42
- testing seasonal time series, 49-51
- trend testing
  - ACF add-in, 37-38
  - interpreting ACF, 35-37
  - removing trends from time series, 43-48
  - testing all ACF at once, 39-42
  - understanding the ACF meaning, 33-35

additive models
- Holt-Winters models, 222-224
  - forecast equations, 226
  - level equations, 224-225
  - seasonal effect equations, 225-226
  - trend equations, 225
- multiplicative models and, 210-214
  - comparing level estimates, 214-217
  - comparing seasonal effect estimates, 217
  - error correction formulas, 220-222

        producing forecasts, 219-220
    smoothing and forecasting, 66-67
    uses of, 213
aligning moving averages, 141
Alpha (fixed range names), 181, 187, 204, 206
ANCOVA (Analysis of Covariance), 58
ANOVA (Analysis of Variance), 148-149
ARIMA analysis, 37
autofilling names in forecasting formulas, 184
averages and ACF, 26

B

backcasting, 95-98
    constraints, 97
    RMSE, 97-98
    smoothing constants, 96
baselines, 91-92
Box-Ljung tests, 39-42

C

centered moving averages, 141-143
chart trendlines and polynomial regression, 14-16
constants. ***See*** **intercepts**
converting adjusted seasonal means to seasonal effects, 132-133
copying/pasting names into forecasting formulas, 184
correlations
    ACF, 22-23
        ACF add-in, 37-38
        averages and, 26
        calculating, 33-35
        drift, 26-29
        PACF, 37-38
        Pearson's correlation versus, 29-30
        removing trends from time series, 43-48
        testing all ACF at once, 39-42
        testing seasonal time series, 49-51
        trend testing, 33-42
PACF (Partial Autocorrelation Function), 37-38
    Pearson's correlation
        ACF versus, 29-30
        definitional formula of, 23-26
correlograms and ACF drift, 29
covariance. ***See*** **ANCOVA**
CurrentLevel (fixed range names), 182, 188-190, 205
CurrentObs (fixed range names), 181-182, 187-188, 204-206
curvilinear regression, 7-8, 118-119

D

damped trend forecasts, 234-236
    formulas, 236-238
    implementing on worksheets, 238-240
damping factor in smoothing and forecasting, 70
Data Analysis add-in, damping factor, 70
Delta (fixed range names), 181, 188, 205- 206
DEVSQ () function, linear regression with coded vectors, 152
differencing trends, 73-76
drift and ACF, 26-29
dummy coding. ***See also*** **effect coding, 155-156**
Durbin-Watson statistic, interpreting ACF, 37

E

effect coding. *See also* **dummy coding**
  regression and, 148-149
  with seasons, 149-153
error handling
  Error (mixed range names), 207
  error correction formulas, 190-191
    additive models, 220-222
    Holt-Winters models, 206-208, 232-234
    level equations, 191-192, 199-203
    multiplicative models, 220-222, 232-234
    range names, 206-208
    season equations, 194-195, 199-201
    trend equations, 192-194, 201-203
    worksheet demonstration, 195-199
  regression, 17
  smoothing, 17, 60-64
exponential smoothing (linear). *See* **Holt's linear exponential smoothing**
exponential smoothing (simple), 59
  equation smoothing form, 64-65
  error handling, 60-62
    RMSE, 62-64
    Solver, 62-64
  golf example, 59-60

F

fixed range names, 180-182
forecasting
  additive models, 219-220, 226
  ANCOVA (Analysis of Covariance), 58
  backcasting, 95-98
    constraints, 97
    RMSE, 97-98
    smoothing constants, 96
  damped trend forecasts, 234-240
  damping factor, 70
  formulas, 178-180
    autofilling names, 184
    copying/pasting names, 184
    fixed range names, 180-182
    interpreting forecast formulas, 189-190
    interpreting formula for seasonal effect, 187-189
    interpreting formula for series level, 184-187
    mixed range names, 182-184
  Holt-Winters models, 226
  initializing forecasts, 92
    backcasting, 95-98
    baselines, 91-92
    confirming regression equations, 121-122
    curvilinear regression, 118-119
    deriving regression equations, 119-121
    Holt-Winters models, 166-168
    optimizing forecasts, 99-117
    polynomial regression, 118-119, 122-124
    seasonal time series, 166-168
    setting initial values, 92-95
  levels, Holt's linear exponential smoothing, 77-81
  multiplicative models, 217-219
  next-step-ahead forecasting, Holt's linear exponential smoothing, 84-86
  optimizing initial forecasts, 99
    Holt's linear exponential smoothing, 106-113
    linear regression, 112-116

    - Multistart (Solver), 116-117
    - simple exponential smoothing, 99-105
  - regression, 6-7
    - confirming regression equations, 121-122
    - curvilinear regression, 7-8, 118-119
    - deriving regression equations, 119-121
    - error handling, 17
    - forecast horizons, 19-20
    - initializing forecasts, 117-122
    - intercepts, 45-46
    - linear regression and forecast horizons, 19-20
    - LINEST () function, 10-13, 52-58
    - model comparisons, 56-58
    - piecewise regression, 8-9
    - polynomial regression, 9-16, 118-119, 122-124
    - smoothing versus regression, 17-18
    - testing seasonal time series, 49-58
    - TREND () function, 7, 10-14, 19-20
  - seasonal time series
    - Holt-Winters models, 164-170
    - initializing forecasts, 166-168
    - starting forecasts earlier, 169-170
  - slope/trend, 70
  - smoothing, 16-17
    - additive model, 66-67
    - damping factor, 70
    - error handling, 17
    - forecasting differences, 75-76
    - Holt's linear exponential smoothing, 70-89
    - initializing values, 67-69
    - multiplicative model, 66-67
    - regression versus smoothing, 17-18
    - slope/trend, 70
    - smoothing constant, 70
    - terminology, 69-70
    - time series, 65-66
    - validating values, 68-69
  - trends, Holt's linear exponential smoothing, 79-81
  - validating forecasts, Holt's linear exponential smoothing, 86-89

ForecastSeason (mixed range names), 190

formulas, 173
  - damped trend forecasts, 236-238
  - error correction formulas, 190-191
    - additive models, 220-222
    - Holt-Winters models, 206-208, 232-234
    - level equations, 191-192, 199-203
    - multiplicative models, 220-222, 232-234
    - range names, 206-208
    - season equations, 194-195, 199-201
    - trend equations, 192-194, 201-203
    - worksheet demonstration, 195-199
  - forecasting formulas, 178-180
    - autofilling names, 184
    - copying/pasting names, 184
    - fixed range names, 180-182
    - interpreting forecast formulas, 189-190
    - interpreting formula for seasonal effect, 187-189
    - interpreting formula for series level, 184-187

        mixed range names, 182-184
    implicit intersections, 174-175
    smoothing formulas, worksheet demonstration, 195-199

G

Gamma (fixed range names), 205-206

GLM (General Linear Models), 148-149

H

Holt's linear exponential smoothing, 70-73, 76-77, 106-113
    differencing, 73-76
    forecasting
        extending forecasts, 81-82
        levels, 77-81
        next-step-ahead forecasting, 84-86
        optimizing constants, 82-84
        trends, 79-81
        validating forecasts, 86-89

Holt-Winters models
    additive models and, 222-224
        forecast equations, 226
        level equations, 224-225
        seasonal effect equations, 225-226
        trend equations, 225
    error correction formulas, 206-208
    multiplicative models and, 222-224, 226-227
        changing initial season estimates, 227
        error correction formulas, 232-234
        modifying level smoothing formulas, 227-230
        modifying seasonal smoothing formulas, 227-230
        modifying trend smoothing
        formulas, 230-232
    named ranges, 203
        estimating current trend, 205
        estimating seasonal effect, 205
        estimating series level, 204
    seasonal time series, 164-166
        initializing forecasts, 166-168
        starting forecasts earlier, 169-170

I

implicit intersections, 174-175

independent residuals, testing seasonal time series, 49-51

indexes (seasonal), calculating, 127-128, 135

inferential statistics, 21-22
    ACF, 22-23
        ACF add-in, 37-38
        averages and, 26
        calculating, 33-35
        drift, 26-29
        PACF, 37-38
        Pearson's correlation versus, 29-30
        removing trends from time series, 43-48
        testing all ACF at once, 39-42
        testing seasonal time series, 49-51
        trend testing, 33-42
    Pearson's correlation
        ACF versus, 29-30
        definitional formula of, 23-26
    testing seasonal time series, 49-51

initializing forecasts, 92
    backcasting, 95-98
        constraints, 97
        RMSE, 97-98
        smoothing constants, 96
    baselines, 91-92

  - Holt-Winters models, seasonal time series, 166-168
  - optimizing forecasts, 99
    - Holt's linear exponential smoothing, 106-113
    - linear regression, 112-116
    - Multistart (Solver), 116-117
    - simple exponential smoothing, 99-105
  - regression
    - confirming regression equations, 121-122
    - curvilinear regression, 118-119
    - deriving regression equations, 119-121
    - linear regression, 112-116
    - polynomial regression, 118-119, 122-124
  - seasonal time series, Holt-Winters models, 166-168
  - setting initial values, 92-95
  - smoothing values, 67-69

intercepts (constants), 45-46

J - K - L

Level (mixed range names), 206

level equations (error correction formulas), 191-192
  - level/seasonal effects, 199-201
  - level/trend effects, 201-203

linear regression
  - coded vectors and, 147-148
    - dummy coding, 155-156
    - effect coding with seasons, 149-153
    - setting up coded vectors, 153-155
  - forecast horizons, 19-20
  - optimizing initial forecasts, 112-116
  - seasonal time series, 147-148
    - dummy coding, 155-156
    - effect coding with seasons, 149-153
    - setting up coded vectors, 153-155

LINEST () function, 45
  - curvilinear regression, 119
  - initializing forecasts
    - curvilinear regression, 119
    - deriving regression equations, 120-121
    - polynomial regression, 123
  - linear regression, 112-113
  - linear regression with coded vectors
    - effect coding with seasons, 150-153
    - setting up coded vectors, 155
  - optimizing initial forecasts, 112-113
  - polynomial regression, 10-13
  - regression in seasonal time series, 52-58

M

means, simple seasonal averages
  - calculating
    - annual means, 130-131
    - trend based on annual means, 131
  - converting adjusted seasonal means to seasonal effects, 132-133

mixed range names, 182-184

model comparisons and regression, 56-58

moving averages, 138-140
  - aligning, 141
  - centered moving averages, 141-143
  - detrending series, 143-147
  - specific seasonals, 140-141

multiplicative models, 209-210
  - additive models and, 210-214

        comparing level estimates, 214-217
        comparing seasonal effect estimates, 217
        error correction formulas, 220-222
        producing forecasts, 217-219
    Holt-Winters models, 222-224, 226-227
        changing initial season estimates, 227
        error correction formulas, 232-234
        modifying level smoothing formulas, 227-230
        modifying seasonal smoothing formulas, 227-230
        modifying trend smoothing formulas, 230-232
    smoothing and forecasting, 66-67

Multistart (Solver), 116-117

N

names, 172-173
    fixed range names, 180-182
    forecasting formulas
        copying/pasting names, 184
        fixed range names, 180-182
        interpreting forecast formulas, 189-190
        interpreting formula for seasonal effect, 187-189
        interpreting formula for series level, 184-187
        mixed range names, 182-184
    Holt-Winters models, 203
        error correction formulas, 206-208
        estimating current trend, 205
        estimating seasonal effect, 205
        estimating series level, 204
        smoothing formulas, 206
    mixed range names, 182-184
    relative range names, 175-178
    smoothing formulas, Holt-Winters models, 206

next-step-ahead forecasting, Holt's linear exponential smoothing, 84-86

O

optimizing initial forecasts, 99
    Holt's linear exponential smoothing, 106-113
    linear regression, 112-116
    Multistart (Solver), 116-117
    simple exponential smoothing, 99-105

P

PACF (Partial Autocorrelation Function), 37-38

patterns (seasonal), identifying, 126

Pearson's correlation
    ACF versus, 29-30
    definitional formula of, 23-26

piecewise regression, 8-9

polynomial regression, 9-10
    chart trendlines, 14-16
    initializing forecasts, 118-119, 122-124
    LINEST () function, 10-13
    TREND () function, 10-14

PriorLevel (mixed range names), 183, 187, 204-207

PriorTrend (mixed range names), 204-207

prorating trends across seasons, 131-132

R

range names, 172-173
  fixed range names, 180-182
  forecasting formulas
    copying/pasting names, 184
    fixed range names, 180-182
    interpreting forecast formulas, 189-190
    interpreting formula for seasonal effect, 187-189
    interpreting formula for series level, 184-187
    mixed range names, 182-184
  Holt-Winters models, 203
    error correction formulas, 206-208
    estimating current trend, 205
    estimating seasonal effect, 205
    estimating series level, 204
    smoothing formulas, 206
  mixed range names, 182-184
  relative range names, 175-178
  smoothing formulas, Holt-Winters models, 206

regression, 6-7
  ANCOVA (Analysis of Covariance), 58
  curvilinear regression, 7-8, 118-119
  effect coding and, 148-149
  error handling, 17
  forecast horizons, 19-20
  initializing forecasts, 117
    confirming regression equations, 121-122
    curvilinear regression, 118-119
    deriving regression equations, 119-121
    polynomial regression, 118-119
  intercepts, 45-46
  linear regression
    coded vectors and, 147-156
    forecast horizons, 19-20
    optimizing initial forecasts, 112-116
    seasonal time series, 147-156
  LINEST () function, 10-13, 52-58
  model comparisons, 56-58
  piecewise regression, 8-9
  polynomial regression, 9-10
    chart trendlines, 14-16
    initializing forecasts, 118-119, 122-124
    LINEST () function, 10-13
    TREND () function, 10-14
  seasonal time series, 147-156
  smoothing versus regression, 17-18
  testing seasonal time series, 52-58
  TREND () function, 7, 10-14, 19-20

relative range names, 175-178

removing trends from time series, 43-48

residuals (independent), testing seasonal time series, 49-51

RMSE (root mean square error)
  backcasting, 97-98
  Holt's linear exponential smoothing, 76
    optimizing constants, 82-84
    validating forecasts, 86-89
  linear regression, 113-116
  optimizing initial forecasts, 112-116
  polynomial regression, 122
  simple exponential smoothing
    error handling, 62-64
    optimizing initial forecasts, 99-105

S

- **SameSeasonLastYear (mixed range names), 182-183, 187-188, 204-207**
- **season equations (error correction formulas), 194-195**
  - level/seasonal effects, 199-201
  - level/trend effects, 201-203
- **seasonal time series, 48, 137**
  - centered moving averages, 141-143
  - Holt-Winters models, 164-166
    - initializing forecasts, 166-168
    - starting forecasts earlier, 169-170
  - independent residuals, 49-51
  - linear regression with coded vectors, 147-148
    - dummy coding, 155-156
    - effect coding with seasons, 149-153
    - setting up coded vectors, 153-155
  - moving averages, 138-140
    - aligning, 141
    - centered moving averages, 141-143
    - detrending series, 143-147
    - specific seasonals, 140-141
  - removing trends from time series, 43-48
  - simple exponential smoothing, 156-161
    - handling ends of time series, 163-164
    - level smoothing formulas, 161-162
    - season smoothing formulas, 162-163
  - simple seasonal averages, 125-126
    - applying seasonal effects to observed quarterlies, 133-134
    - calculating annual means, 130-131
    - calculating seasonal indexes, 127-128, 135
    - calculating trend based on annual means, 131
    - converting adjusted seasonal means to seasonal effects, 132-133
    - evaluating, 136-137
    - forecasting from no trend, 128-129
    - forecasting with trends, 129-136
    - identifying seasonal patterns, 126
    - prorating trends across seasons, 131-132
    - regressing deseasonalized quarterlies onto time periods, 134-135
  - specific seasonals, 140-141
  - testing for, 42-43
- **simple exponential smoothing, 59**
  - equation smoothing form, 64-65
  - error handling, 60-62
    - RMSE, 62-64
    - Solver, 62-64
  - golf example, 59-60
  - optimizing initial forecasts, 99-105
  - seasonal time series, 156-161
    - handling ends of time series, 163-164
    - level smoothing formulas, 161-162
    - season smoothing formulas, 162-163
- **simple seasonal averages, 125-126**
  - applying seasonal effects to observed quarterlies, 133-134
  - calculating seasonal indexes, 127-128, 135
  - evaluating, 136-137
  - forecasting from
    - no trend, 128-129
    - with trends, 129-136

- identifying seasonal patterns, 126
- means
  - calculating annual means, 130-131
  - calculating trend based on annual means, 131
  - converting adjusted seasonal means to seasonal effects, 132-133
- regressing deseasonalized quarterlies onto time periods, 134-135
- trends
  - calculating trend based on annual means, 131
  - prorating trends across seasons, 131-132

slope/trend in smoothing and forecasting, 70

smoothing, 16-17

- backcasting, 96
- damping factor, 70
- error handling, 17
- forecasting techniques
  - additive model, 66-67
  - damping factor, 70
  - forecasting differences, 75-76
  - initializing values, 67-69
  - multiplicative model, 66-67
  - slope/trend, 70
  - smoothing constant, 70
  - terminology, 69-70
  - time series, 65-66
  - validating values, 68-69
- formulas, worksheet demonstration, 195-199
- Holt's linear exponential smoothing, 70-73, 76-77, 106-113
  - differencing, 73-76
  - extending forecasts, 81-82
  - forecasting levels, 77-81
  - forecasting trends, 79-81
  - next-step-ahead forecasting, 84-86
  - optimizing constants, 82-84
  - validating forecasts, 86-89
- level smoothing formulas
  - Holt-Winters models, 227-230
  - multiplicative models, 227-230
- regression versus smoothing, 17-18
- seasonal smoothing formulas
  - Holt-Winters models, 227-230
  - multiplicative models, 227-230
- simple exponential smoothing, 59
  - equation smoothing form, 64-65
  - error handling, 60-64
  - golf example, 59-60
  - optimizing initial forecasts, 99-105
  - seasonal time series, 156-164
- slope/trend, 70
- smoothing constant, 70
- trend smoothing formulas
  - Holt-Winters models, 230-232
  - multiplicative models, 230-232

Solver

- Holt's linear exponential smoothing
  - optimizing constants, 82-84
  - validating forecasts, 86-89
- linear regression, 112-116
- Multistart, 116-117
- optimizing initial forecasts
  - Holt's linear exponential smoothing, 106-113
  - linear regression, 112-116
  - simple exponential smoothing, 99-105
- simple exponential smoothing error handling, 62-64

statistical inference, 21-22
- ACF, 22-23
  - ACF add-in, 37-38
  - averages and, 26
  - calculating, 33-35
  - drift, 26-29
  - PACF, 37-38
  - Pearson's correlation versus, 29-30
  - removing trends from time series, 43-48
  - testing all ACF at once, 39-42
  - testing seasonal time series, 49-51
  - trend testing, 33-42
- Pearson's correlation
  - ACF versus, 29-30
  - definitional formula of, 23-26
- testing seasonal time series, 49-51

T

testing for
- Box-Ljung tests, 39-42
- seasonal time series, 42-48
- trends, 30-33
  - ACF add-in, 37-38
  - interpreting ACF, 35-37
  - testing all ACF at once, 39-42
  - understanding the ACF meaning, 33-35

time series
- ANCOVA, 58
- removing trends from, 43-48
- seasonal time series, 48, 137
  - centered moving averages, 141-143
  - detrending series with moving averages, 143-147
  - Holt-Winters models, 164-170
  - independent residuals, 49-51
  - linear regression with coded vectors, 147-156
  - moving averages, 138-141, 143-147
  - removing trends from time series, 43-48
  - simple exponential smoothing, 156-164
  - simple seasonal averages, 125-137
  - specific seasonals, 140-141
  - testing for, 42-43
- smoothing and forecasting, 65-66
- testing for trends, 30-33
  - ACF add-in, 37-38
  - interpreting ACF, 35-37
  - testing all ACF at once, 39-42
  - understanding the ACF meaning, 33-35
- trends, 48
  - differencing, 73-76
  - Holt's linear exponential smoothing, 70-76

TREND () function, 45
- forecast horizons, 19-20
- initializing forecasts, 120-121
- linear regression with coded vectors, 151-152
- polynomial regression, 10-14
- regression, 7, 19-20

trend equations (error correction formulas), 192-194, 201-203

trendlines, 113
- chart trendlines, polynomial regression, 14-16
- initializing forecasts, curvilinear regression, 118-119

trends
- additive models, 225

- damped trend forecasts, 234-236
  - formulas, 236-238
  - implementing on worksheets, 238-240
- differencing, 73-76
- estimating current trend in Holt-Winters models, 205
- forecasting, Holt's linear exponential smoothing, 79-81
- Holt-Winters models, 225
- removing from time series, 43-48
- simple seasonal averages
  - calculating trend based on annual means, 131
  - forecasting from no trend, 128-129
  - forecasting with trends, 129-136
  - prorating trends across seasons, 131-132
- slope/trend, smoothing and forecasting, 70
- smoothing formulas
  - Holt-Winters models, 230-232
  - multiplicative models, 230-232
- testing for, 30-33
  - ACF add-in, 37-38
  - interpreting ACF, 35-37
  - testing all ACF at once, 39-42
  - understanding the ACF meaning, 33-35
- time series, 48, 70-76

U - V

validating

- forecasts, Holt's linear exponential smoothing, 86-89
- values, smoothing and forecasting, 68-69

variance. *See* **ANOVA**

VLOOKUP () function, linear regression with coded vectors, 153-155

W - X - Y

Year_1_Mean (fixed range names), 181

Z

z-scores, ACF drift, 27-28

REGISTER THIS PRODUCT

SAVE 35%*

ON YOUR NEXT PURCHASE!

How to Register Your Product

- Go to quepublishing.com/register
- Sign in or create an account
- Enter ISBN: 9780134070926

Benefits of Registering

- Ability to download product updates
- Access to bonus chapters and workshop files
- A 35% coupon to be used on your next purchase – valid for 30 days
  To obtain your coupon, click on "Manage Codes" in the right column of your Account page
- Receive special offers on new editions and related Que products

Please note that the benefits for registering may vary by product. Benefits will be listed on your Account page under Registered Products.

We value and respect your privacy. Your email address will not be sold to any third party company.

** 35% discount code presented after product registration is valid on most print books, eBooks, and full-course videos sold on QuePublishing.com. Discount may not be combined with any other offer and is not redeemable for cash. Discount code expires after 30 days from the time of product registration. Offer subject to change.*

quepublishing.com

QUE®