



JORDAN GOLDMEIER
MISTRZ
ANALIZY DANYCH
OD DANYCH DO WIEDZY

WYDANIE II

Helion 

WILEY

Tytuł oryginału: Data Smart: Using Data Science to Transform Information into Insight, 2nd Edition

Tłumaczenie: Anna Mizerska z wykorzystaniem fragmentów poprzedniego wydania w przekładzie Konrada Matuka

ISBN: 978-83-289-1180-2

Copyright © 2024 by John Wiley & Sons, Inc., Hoboken, New Jersey.

All Rights Reserved. This translation published under license with the original publisher John Wiley & Sons, Inc.

Translation copyright © 2024 by Helion S.A.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, without either the prior written permission of the Publisher.

WILEY, and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

Autor oraz wydawca dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz wydawca nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres

<https://helion.pl/user/opinie/miand2>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Helion S.A.

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 230 98 63

e-mail: helion@helion.pl

WWW: <https://helion.pl> (księgarnia internetowa, katalog książek)

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

Spis treści

Wstęp	13
I Wszystko, co chciałeś wiedzieć o arkuszu kalkulacyjnym, ale bałeś się o to zapytać	21
Przykładowe proste dane	22
Szybkie statystyki opisowe	23
Tabele Excela	24
Filtrowanie i sortowanie	26
Formatowanie tabeli	27
Odwołania strukturalne	28
Dodawanie kolumn do tabeli	31
Formuły przeznaczone do wyszukiwania	32
WYSZUKAJ.POZIOMO	32
INDEKS/PODAJ.POZYCJĘ	34
X.WYSZUKAJ	36
Tabele przestawne	38
Korzystanie z formuł tablicowych	40
Rozwiązywanie problemów za pomocą narzędzia Solver	41
2 Ustaw i zapomnij! Wprowadzenie do Power Query	49
Czym jest Power Query?	49
Przykładowe dane	50
Rozpoczęcie pracy z Power Query	51
Filtrowanie wierszy	54
Usuwanie kolumn	55
Znajdź i zastąp	56
Zamknij i załaduj... do tabeli	58

3	Naiwny klasyfikator bayesowski i niezwykła lekkość bycia idiotą	61
	Najszybsze na świecie wprowadzenie do rachunku prawdopodobieństwa	61
	Obliczanie prawdopodobieństwa warunkowego	62
	Prawdopodobieństwo części wspólnej, reguła łańcuchowa i niezależność	62
	A co, jeżeli sytuacje są zależne od siebie?	63
	Twierdzenie Bayesa	64
	Oddzielanie sygnału od szumu	65
	Tworzenie modelu sztucznej inteligencji za pomocą twierdzenia Bayesa	66
	Zwykle zakłada się, że wysokopoziomowe prawdopodobieństwa klas są sobie równe	68
	Kilka innych drobnostek	68
	Czas rozpocząć zabawę z Excelem	70
	Porządkowanie danych za pomocą Power Query	70
	Dzielenie na znakach spacji: każde słowo musi dostać to, co mu się należy	73
	Zliczanie leksemów i obliczanie prawdopodobieństw	78
	Zbudowaliśmy model. Skorzystajmy z niego!	80
4	Analiza skupień. Część I — zastosowanie algorytmu centroidów do segmentowania bazy klientów	87
	Zabawy taneczne na obozie letnim	88
	Prawdziwy problem: implementacja algorytmu centroidów w e-mail marketingu	93
	Początkowy zbiór danych	93
	Określanie tego, co chcemy mierzyć	94
	Zacznij od czterech grup	97
	Odległość euklidesowa — pomiar odległości w linii prostej	98
	Określanie położenia środków klastrów	102
	Analiza uzyskanych wyników	104
	Ustalanie najlepszej oferty dla danego klastra	106
	Sylwetka podziału — dobry sposób na określenie optymalnej liczby klastrów	110
	A może potrzebujesz pięciu klastrów?	117
	Dzielenie klientów na pięć klastrów za pomocą narzędzia Solver	118
	Ustalanie najlepszych ofert dla wszystkich pięciu klastrów	119
	Określanie sylwetki podziału na pięć klastrów	122

Podział na grupy za pomocą algorytmu K-mediodów i asymetryczny pomiar odległości	123
Podział na grupy za pomocą metody K-mediodów	123
Stosowanie lepszego sposobu pomiaru odległości	123
Implementacja za pomocą Excela	126
Najlepsze oferty przy podziale na pięć klastrów za pomocą median	128
5 Analiza skupień. Część II — grafy i analiza sieci	133
Czym jest graf sieci?	134
Wizualizacja prostego grafu	134
Wyjście poza dodatek GiGraph i listy sąsiedztwa	138
Tworzenie grafu na podstawie danych sprzedaży wina	140
Tworzenie macierzy podobieństwa kosinusowego	142
Generowanie grafu r-sąsiedztwa	144
Wprowadzenie do Gephi	146
Tworzenie statycznej macierzy sąsiedztwa	146
Macierz r-sąsiedztwa w Gephi	147
Stopień rozgałęzienia	151
Edycja danych grafu	152
Jaka jest wartość krawędzi? Nagradzanie i karanie krawędzi — modularność grafu	154
Czym jest punkt, a czym kara?	155
Tworzenie arkusza punktacji	159
Czas dokonać podziału na grupy	161
Podział 1.	161
Podział 2. — kontratak	167
Podział 3. — zemsta	169
Grupy — kodowanie i analiza	170
Tam i z powrotem — czas na Gephi	174
6 Regresja jako przodek nadzorowanego uczenia maszynowego i sztucznej inteligencji	181
Przewidywanie ciąży klientów na podstawie regresji liniowej	183
Zbiór cech	184
Tworzenie treningowego zbioru danych	186

Tworzenie zmiennych fikcyjnych	188
Pobawmy się regresją liniową	191
Parametry regresji liniowej: współczynnik determinacji, test F i test t	198
Przewidywanie ciąży na nowym zbiorze danych i sprawdzanie jakości modelu	208
Przewidywanie ciąży klientów za pomocą regresji logistycznej	218
Najpierw musisz określić funkcję wiążącą	218
Tworzenie funkcji logistycznej i ponowna optymalizacja	219
Praca nad prawdziwą regresją logistyczną	222
7 Modele zespołowe — dużo nie najlepszej pizzy	229
Korzystanie z danych z rozdziału 6.	230
Agregacja — losuj, trenuj, powtórz	231
Pieniek decyzyjny to kolejne określenie słabego klasyfikatora	231
To wcale nie wydaje się takie słabe!	231
Więcej mocy!	234
Czas rozpocząć proces trenowania	235
Ocena działania modelu zespołowego	248
Wzmacnianie — jeżeli uzyskałeś niesatysfakcjonujące wyniki, to wzmocnij swój model i uruchom go jeszcze raz	252
Trenowanie modelu — każda cecha ma swoje pięć minut	253
Wydajność modelu wzmacnianych reguł decyzyjnych	261
8 Prognozowanie — oddychaj spokojnie, i tak nie wygrasz	265
Hossa na rynku sprzedaży mieczy	266
Szeregi czasowe	267
Zacznij od prostego wygładzania wykładniczego	268
Przygotowanie arkusza prognozy prostego wygładzania wykładniczego	271
Być może dane zawierają trend	280
Podwójne wygładzanie wykładnicze (metoda Holta)	282
Metoda Holta w arkuszu kalkulacyjnym	283
To wszystko? Analiza autokorelacji	290

Wielokrotne wygładzanie wykładnicze — model Holta-Wintersa	298
Określanie początkowych wartości poziomu, trendu i sezonowości	300
Tworzenie prognozy	307
Czas na optymalizację	313
Interwały prognozy	315
Tworzenie wykresu warstwowego wachlarza wartości	321
Arkusze prognozy w Excelu	323
9 Modelowanie optymalizacyjne — świeżo wyciśnięty sok	
nie zamieszka się sam	327
Ale czy to w ogóle jest analiza danych?	328
Zacznijmy od prostego kompromisu	329
Przedstawienie problemu w formie wielokomórki	330
Rozwiązywanie problemu poprzez przesuwanie poziomicy	331
Metoda simpleks — kręcenie się wokół rogów	333
Praca w Excelu	335
Szkłanka świeżego soku pomarańczowego prosto z drzewa...	
z przystankiem na modelowanie	340
Zacznijmy od specyfikacji soków	342
Stałość produktu wyjściowego	343
Wprowadzanie danych do Excela	344
Określanie problemu w dodatku Solver	347
Obniżanie standardów	349
Usuwanie cuchnącego problemu	
— minimalizacja maksymalnych odchyień	352
Warunki i ograniczenie „wielkiego M”	354
Mnożenie zmiennych — skorzystajmy ze 110% mocy Excela	359
Modelowanie ryzyka	366
Dane pochodzące z rozkładu normalnego	366
10 Wykrywanie obserwacji odstających	375
Element odstający to też człowiek	376
Fascynująca sprawa Hadlumów	376
Metoda Tukeya	378
Implementacja metody Tukeya w arkuszu kalkulacyjnym	379
Ograniczenia tej prostej techniki	381

Nie tragiczny, ale słaby we wszystkim	383
Przygotowywanie danych do utworzenia wykresu	384
Tworzenie grafu	386
Określanie k najbliższych sąsiadów	388
Pierwsza metoda wykrywania elementów odstających grafu — skorzystaj ze stopnia wchodzącego	389
Druga metoda wykrywania elementów odstających grafu — zgłębianie niuansów za pomocą k-odległości	393
Trzecia metoda wykrywania elementów odstających grafu — lokalny miernik stopnia oddalenia obserwacji	395
I I Przejście z arkusza kalkulacyjnego do języka R	401
Przygotowanie środowiska i początek pracy w języku R	402
Szybkie szkolenie z pisania skryptów w języku R	403
Działania matematyczne na wektorach i faktory	405
Najlepszy typ danych — dataframe	408
Pomoc dla języka R	409
Wyjście poza podstawowe możliwości R	410
Prawdziwa analiza danych	412
Wczytywanie danych do R	412
Sferyczny algorytm k-średnich wywołany za pomocą zaledwie kilku linii kodu	413
Budowanie modeli sztucznej inteligencji na podstawie danych zakupów (wykrywanie ciąży)	418
Prognozowanie w R	426
Wykrywanie elementów odstających	430
I 2 Wnioski	435
Gdzie ja jestem? Co się stało?	435
Zanim odłożysz tę książkę	435
Poznaj problem	436
Potrzebujemy więcej tłumaczy	436
Uważaj na trójgłowe monstrum: narzędzia, wydajność i perfekcjonizm	437
Nie jesteś najważniejszą osobą w firmie	440
Bądź kreatywny	440

8

Prognozowanie — oddychaj spokojnie, i tak nie wygrasz

W poprzednich rozdziałach dowiedziałeś się, że nadzorowane uczenie maszynowe jest związane z przewidywaniem wartości lub klasyfikacją obserwacji za pomocą modelu trenowanego na danych zebranych w przeszłości. Prognozy są tworzone w podobny sposób. Oczywiście jeżeli jesteś astrologiem, to możesz tworzyć prognozy bez danych, ale w przypadku prognoz ilościowych coś, co zaistnieje w przyszłości, jest prognozowane na podstawie danych zebranych w przeszłości. W przypadku prognozowania i nadzorowanego uczenia maszynowego stosuje się czasami te same techniki (np. regresję wielokrotną, o której pisałem w rozdziale 6. „Regresja jako prodek nadzorowanego uczenia maszynowego i sztucznej inteligencji”).

Prognozowanie i nadzorowane uczenie maszynowe charakteryzują się różnymi kanonicznymi przestrzeniami problemu. Zwykle kłopoty związane z przewidywaniem dotyczą zmian jakiejś wartości w czasie — wartością taką może być np.: popyt, podaż, PKB, emisja tlenków węgla lub zaludnienie. Z powodu trendów, cykli i zdarzeń losowych dane w przyszłości mogą wykraczać poza graniczne wartości z przeszłości.

Na tym polega właśnie problem z prognozowaniem: w poprzednich rozdziałach miałeś okazję zaobserwować nawyki zakupowe ciężarnych kobiet, które mniej więcej kupują to samo. Ale co w sytuacjach, w których przyszłość bardzo często zupełnie nie przypomina przeszłości? Szereg danych z przyszłości może i będzie wyglądać inaczej niż dane, które zaobserwowałeś wcześniej.

Nawet jeżeli opracujesz dobrą prognozę zapotrzebowania na mieszkania, może dojść do pęknięcia bańki mieszkaniowej, w wyniku czego Twoja prognoza nie będzie się do niczego nadawać.

Jeżeli uda Ci się opracować doskonałą prognozę popytu, to pracę Twoich dostawców może zakłócić pandemia, zmuszając Cię do podwyżki cen i pogodzenia się ze zmniejszeniem sprzedaży. Przyszłe wartości mogą wyglądać zupełnie inaczej od danych zebranych w przeszłości.

Jedynym pewnikiem związanym z prognozą jest to, że jest ona błędna.

Zapisz to sobie, wydrukuj i przyczep na biurku swojego szefa.

Nie oznacza to, że nie możesz podjąć próby dokonania prognoz. Są one niezbędne podczas planowania działalności gospodarczej. Bez względu na uzyskane wyniki

będziesz wiedzieć więcej niż nic. Nie chcesz się bawić w ganianego tylko dlatego, że Twój konkurenci przemyśleli problem, a Ty nie.

W tym rozdziale poznasz możliwości przewidywania przyszłości, a także określenia niepewności takich prognoz. Określanie niepewności prognoz za pomocą **interwałów** jest czymś bardzo praktycznym, o czym często się zapomina.

Pewien prognosta powiedział kiedyś: „Dobry prognosta nie jest osobą najmądrzejszą, po prostu jest osobą, której ignorancja jest lepiej zorganizowana”.

A więc, bez zbędnych ceregieli, spróbujmy zorganizować nieco lepiej naszą ignorancję.

Hossa na rynku sprzedaży mieczy

Wyobraź sobie, że jesteś fanem sagi *Władca Pierścieni*. Kilka lat temu, gdy pierwsza część tej trylogii wchodziła na ekrany kin, stałeś przez parę godzin w długiej kolejce, mając na nogach buty udające stopy hobbita, i czekałeś na rozpoczynającą się po północy premierę. Wkrótce potem uczestniczyłeś w zjazdach fanów fantastyki i kłóciłeś się na forach dyskusyjnych o to, czy Frodo mógł po prostu polecieć na Górę Przeznaczenia.

Pewnego dnia postanowiłeś dać całej społeczności fanów coś od siebie. Poszedłeś na kurs obróbki metalu i zacząłeś tworzyć miecze. Twoim ulubionym mieczem z *Władcy Pierścieni* był Anduril — Błask Zachodu. Stałeś się ekspertem potrafiącym wykuć takie opasłe pałasze w swojej przydomowej kuźni i zacząłeś je sprzedawać w różnych portalach aukcyjnych. Obecnie wykonane przez Ciebie repliki są mieczami, o których marzą wszyscy fani J.R.R. Tolkiena. Sprzedaż wzrasta błyskawicznie.

Zdarzały Ci się sytuacje, w których miałeś problem z zaspokojeniem popytu z powodu zbyt małej dostępności materiałów niezbędnych do wykucia miecza. W związku z tym postanowiłeś przewidzieć popyt. Umieściłeś dotychczasowe dane sprzedaży w arkuszu kalkulacyjnym, ale nie wiesz, jak na ich podstawie opracować prognozę.

W tym rozdziale skupię się na technikach prognozowania określanych mianem metod **wygładzania wykładniczego**. Metody te są jednymi z najprostszych i najczęściej stosowanych we współczesnym biznesie. Znam kilka spółek lokowanych w górnej części listy „Fortune” 500, które korzystają z tych metod, ponieważ w przypadku przetwarzanych przez nie danych metody te okazały się najbardziej dokładne.

Dokładność ta wynika częściowo z prostoty tych technik — są one odporne na nadmierne dopasowanie do danych z przeszłości, które często są niekompletne. (W ramach przypomnienia, przetrenowanie jest wynikiem potraktowania przez model danych treningowych jako całego zestawu, a co za tym idzie, model zbyt dobrze zna dane, na których został wytrenowany, i nie jest w stanie przewidzieć przyszłego zjawiska). Te techniki umożliwiają dość proste obliczenie interwałów prognoz utworzonych za pomocą wykładzania wykładniczego, a więc zaprezentuję również obliczenia tych interwałów.

Szeregi czasowe

Uwaga

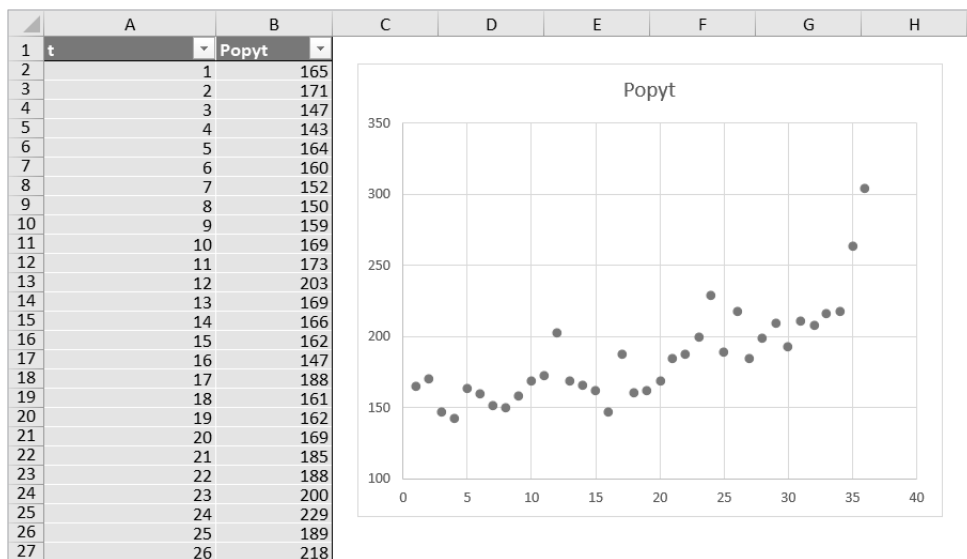
W tym rozdziale będę korzystał ze skróty programu Excel o nazwie *PrognozowaniePopytuNaMiecze.xlsm*. Możesz go pobrać ze strony: <https://ftp.helion.pl/przyklady/miand2.zip>.

Skoroszyt *PrognozowaniePopytuNaMiecze.xlsm* zawiera dane popytu na produkowany przez Ciebie miecz z ostatnich 36 miesięcy (zaczynając od stycznia 3 lata temu). Dane znajdujące się w arkuszu *Szereg czasowy* pokazano na rysunku 8.1. **Danymi szeregu czasowego** nazywamy dane obserwacji dokonywanych w regularnych odstępach czasu. Odstęp pomiędzy obserwacjami może być dowolny. Możesz analizować np. roczne zmiany populacji lub dobowe wahania cen paliw.

	A	B
1	t	Popyt
2	1	165
3	2	171
4	3	147
5	4	143
6	5	164
7	6	160
8	7	152
9	8	150
10	9	159
11	10	169
12	11	173
13	12	203
14	13	169
15	14	166
16	15	162
17	16	147
18	17	188
19	18	161
20	19	162
21	20	169
22	21	185
23	22	188
24	23	200
25	24	229
26	25	189
27	26	218
28	27	185
29	28	199
30	29	210
31	30	193
32	31	211
33	32	208
34	33	216
35	34	218
36	35	264
37	36	304

Rysunek 8.1. Dane szeregu czasowego

Dysponując miesięcznymi danymi popytu na miecze, powinieneś przedstawić je na wykresie (rysunek 8.2). W celu wygenerowania wykresu zaznacz kolumny A i B, a następnie wybierz opcję wstawiania wykresu znajdującą się na karcie *Wstawianie*. Zakres osi wykresu możesz zmienić, klikając je prawym przyciskiem myszy i wybierając opcję *Formatuj oś...*



Rysunek 8.2. Szereg czasowy przedstawiony za pomocą wykresu — wartości osi y zostały zmienione na zakres od 100 do 350

Co widzisz na wykresie pokazanym na rysunku 8.2? Trzy lata temu miesięczne zapotrzebowanie na miecz oscylowało w okolicy 150 sztuk, a miesiąc temu osiągnęło wartość 304 sztuk. W ciągu trzech lat zapotrzebowanie się podwoiło. Czy to oznacza, że masz do czynienia z tendencją wzrostową? Za chwilę odpowiemy na to pytanie.

Na wykresie widać pewne spadki i wzrosty zapotrzebowania, które mogą wskazywać na sezonowe zwiększenie popytu. Na przykład najwyższe zapotrzebowanie w każdym z trzech analizowanych lat panowało w grudniu (pod numerami 12, 24 i 36 kryje się grudzień). To przypadek czy trend? Sprawdź to.

Zacznij od prostego wygładzania wykładniczego

Techniki wygładzania wykładniczego umożliwiają prognozowanie przyszłych wartości na podstawie danych z przeszłości przy założeniu, że nowsze obserwacje są ważniejsze od starszych. Proces przypisywania wag do obserwacji odbywa się za pomocą **stałych wygładzających**. Na początek chciałbym przyjrzeć się metodzie

prostego wygładzania wykładniczego (ang. *Single Exponential Smoothing* — SES), w której stosuje się tylko jedną stałą wygładzającą.

Proste wygładzanie wykładnicze zakłada, że dane Twojego szeregu czasowego składają się z dwóch komponentów: **poziomu** (lub średniej) i jakiegoś błędu związanego z poziomem. Nie ma trendu, nie ma sezonowości, jest tylko poziom, wokół którego zmienia się popyt, charakteryzując się pewnymi odchyleniami występującymi tu i ówdzie. Proste wygładzanie wykładnicze bardziej dostosowuje się do nowszych obserwacji, co może wpływać na przesunięcia poziomu. Popyt możemy wyrazić za pomocą następującego wzoru:

popyt w chwili t = poziom + błąd losowy odchylenia od poziomu w chwili t

Najnowsza oszacowana wartość poziomu pełni funkcję prognozy na przyszłość. Jeżeli mamy 36. miesiąc, to jaka jest dobra prognoza dla 38. miesiąca? Najnowsza oszacowana wartość poziomu. A w 40. miesiącu? Poziom. To proste. Określenie „proste wygładzanie wykładnicze” nie jest gołosłowne.

Jak oszacować poziom?

Jeżeli założysz, że wszystkie wartości z przeszłości są tak samo ważne, to wystarczy policzyć ich średnią.

Średnia ta dyktowałaby poziom, a Ty mógłbyś prognozować przyszłe wartości, mówiąc: „Popyt w przyszłości jest równy wartości średniej popytu z przeszłości”. Istnieją firmy stosujące takie rozwiązanie. Spotkałem się z przedsiębiorstwami, których prognozy na kolejne miesiące były równe średniej wartości tych samych miesięcy w ciągu kilku lat, do której dodano dla zabawy „czynnik *ad hoc*”. Taaak... Prognozy w niektórych dużych spółkach publicznych są dokonywane w tak pokrętny sposób, że można tam wciąż spotkać określenia typu „czynnik *ad hoc*”. To straszne.

A co, jeżeli poziom zmienia się wraz z upływem czasu? Nie warto wtedy przypisywać każdemu punktowi danych z przeszłości takiej samej wagi, jak dzieje się to w przypadku średniej. Gdybyś chciał przewidzieć stan magazynowy w oparciu o dane z przeszłości, to czy dane od marca 2020 do grudnia 2021 roku miałyby taką samą wagę jak podczas prognozowania w 2024 roku? Prawdopodobnie nie, biorąc pod uwagę fakt, że w tym okresie na świecie panowała pandemia Covid-19. W związku z tym musiałbym określić poziom w sposób, który przywiązuje większą wagę do najnowszych danych popytu.

Jak w takim przypadku miałbyś obliczyć poziom? Zamiast wyliczać średnią ze wszystkich wartości, spróbuj obliczyć poziom dla wybranych danych z przeszłości. Na początku oblicz początkową szacunkową wartość poziomu jako średnią wartość np. danych z pierwszego roku działalności. Początkową szacunkową wartość poziomu nazwij *poziom₀*:

poziom₀ = średnia popytu 1. roku (miesiące 1. – 12.)

Średni popyt na miecze w tym okresie wynosił 163.

Wyglądanie wykładnicze polega na tym, że mimo iż znasz popyt w miesiącach 1. – 36., i tak prognozę na miesiąc znajdujący się po całej serii danych będziesz opierał głównie na ostatnich danych.

Wartość $poziom_0$ (163) możesz zastosować jako prognozę zapotrzebowania w miesiącu 1.

Przyjąłeś prognozowaną wartość dla okresu 1. Teraz przyjrzyj się liczbie mieczy sprzedanych od okresu 0 do okresu 1. W rzeczywistości popyt wyniósł 165 sztuk, a więc pomyliłeś się o 2 miecze. Powinieneś zaktualizować prognozowany poziom, biorąc pod uwagę tę pomyłkę. Korzystasz z prostego wykładzania wykładniczego, a więc możesz zastosować następujące równanie:

$$poziom_1 = poziom_0 + \text{jakiś procent} \cdot (popyt_1 - poziom_0)$$

Zauważ, że wartość $(popyt_1 - poziom_0)$ jest równa wartości błędu uzyskanego podczas prognozowania pierwszego okresu za pomocą początkowo szacowanej wartości poziomu. Idąc dalej, otrzymasz:

$$poziom_2 = poziom_1 + \text{jakiś procent} \cdot (popyt_2 - poziom_1)$$

A jeszcze dalej otrzymasz:

$$poziom_3 = poziom_2 + \text{jakiś procent} \cdot (popyt_3 - poziom_2)$$

Procent błędu, który chcesz uwzględnić w nowej wartości poziomu, określa się mianem **stałej wykładzania** lub **parametru alfa**. Może być to dowolna wartość od 0 do 100%.

Jeżeli przypiszesz parametrowi *alfa* wartość 1, to uwzględniasz błąd w całości, co oznacza, że poziom bieżącego okresu staje się zapotrzebowaniem tego okresu.

Jeżeli przypiszesz parametrowi *alfa* wartość 0, to nie przeprowadzasz żadnej korekcji błędu prognozy pierwszego poziomu.

Najprawdopodobniej chcesz zastosować jakąś wartość znajdującą się pomiędzy wspomnianymi wartościami skrajnymi, ale wybór najlepszej wartości parametru *alfa* opiszę nieco później.

Dla kolejnych punktów czasu wykonujesz obliczenia, które można opisać za pomocą ogólnego wzoru:

$$poziom_{\text{bieżącego okresu}} = poziom_{\text{poprzedniego okresu}} + \text{alfa} \cdot (popyt_{\text{bieżącego okresu}} - poziom_{\text{poprzedniego okresu}})$$

Na koniec uzyskasz prognozę ostatniego poziomu ($poziom_{36}$), na którą bardziej wpływają ostatnie obserwacje, ponieważ wartości te nie zostały pomnożone wielokrotnie przez współczynnik *alfa*:

$$poziom_{36} = poziom_{35} + \text{alfa} \cdot (popyt_{36} - poziom_{35})$$

Ta ostatnia prognozowana wartość poziomu zostanie użyta przez Ciebie do utworzenia prognozy dla kolejnych miesięcy. Jaka wartość prognozujesz dla 37. miesiąca? Jest ona po prostu równa wartości *poziom*₃₆. Jaka wartość prognozujesz dla 40. miesiąca? Tę samą. A jaką dla 45. miesiąca? Ponownie tę samą. Chyba już widzisz, na czym to polega. Ostatnia prognozowana wartość poziomu jest najlepszą wartością opisującą przyszłość, jaką dysponujesz, a więc możesz korzystać tylko z niej.

Spróbujmy ją uzyskać w arkuszu kalkulacyjnym.

Przygotowanie arkusza prognozy prostego wykładniczego

Zacznij od utworzenia w skoroszycie nowego arkusza o nazwie **SES**. Do kolumn A i B wklej dane szeregu czasowego. Zacznij od wiersza 4. — zostaw miejsce na wartości parametru *alfa*. Zamień dane na tabelę Excela i nazwij ją **SES.Miecze**.

CZY W NAZWIE TABELI MOŻE ZNAJDOWAĆ SIĘ KROPKA? A CO Z NAZWANYMI ZAKRESAMI?

Tak, może. I naprawdę cały czas stosuję taki zapis. Nazwy moich tabel i zakresów często opierają się na takim schemacie: [nazwa_arkusza].[nazwa_objektu]. W ten sposób w moich skoroszytach panuje większy porządek.

Ale musisz wiedzieć, że firma Microsoft poinformowała mnie, że kropki w nazwach mogą powodować problemy z takimi formułami jak LET czy LAMBDA. Osobiście nigdy nie miałem problemów, ale warto, żebyś o tym wiedział.

W komórce A2 umieść liczbę miesięcy, które są opisywane przez Twoje dane (36), a w komórce C2 podaj początkową wartość współczynnika *alfa*. W swoim arkuszu w komórce C2 wpisałem wartość 0,5, ponieważ znajduje się ona w połowie zakresu 0 – 1. W swoim arkuszu możesz również dodać etykiety takie jak na rysunku 8.3.

W kolumnie C umieść formuły obliczające wartości poziomów. Zanim to zrobisz, do danych szeregu czasowego wstaw nowy wiersz 5. (wstaw go nad miesiącem 1., a w nowej komórce A5 wpisz numer 0, tak jak na rysunku 8.3). Następnie utwórz trzy dodatkowe kolumny: **Szacowany poziom** (kolumna C), **Prognoza jednego kroku** (kolumna D) i **Błąd prognozy** (kolumna E), tak jak na rysunku 8.3. Teraz możesz przejść do obliczeń.

	A	B	C	D	E
			Parametr wygładzania poziomu (alfa)		
1	Liczba miesięcy				
2	36		0,50		
3					
4	t	Popyt	Szacowany poziom	Prognoza jednego	Błąd prognozy
5		0			
6		1	165		
7		2	171		
8		3	147		
9		4	143		
10		5	164		
11		6	160		
12		7	152		
13		8	150		
14		9	159		
15		10	169		
16		11	173		
17		12	203		
18		13	169		
19		14	166		
20		15	162		
21		16	147		
22		17	188		

Rysunek 8.3. Prognozowany początkowy poziom prostego wygładzania wykładniczego

Dodawanie kolumn prognozy jednego kroku i jej błędu

Aby utworzyć prognozę jednego kroku, a następnie zrozumieć błąd, musisz najpierw określić odpowiedni szacunkowy poziom. Ponieważ to obliczenie wiąże się z wieloma warunkami, będziesz je budować stopniowo. Zaczyniesz od części formuły, które następnie będziesz zmieniać. Na przykład powiem Ci, że masz wprowadzić zmiany w formule, którą wcześniej już wpisałeś w danej komórce. Dlatego wpisanie formuły do komórki nie oznacza, że praca nad nią została ukończona. Zobaczysz, jak to działa w praktyce.

Ponadto będziesz pracować głównie z formułami dynamicznymi. Co za tym idzie, będziesz mógł łatwo przetestować scenariusze spoza tej książki, wystarczy, że pozmieniasz wartości w komórkach i dostosujesz rozmiar tabeli Excela do swoich danych. Pozostałe elementy powinny pozostać niezmiennie i gotowe do użycia, ponieważ będziemy unikać wartości podanych na sztywno w formułach. Zatem zachęcam Cię do zabawy z przykładami z tego rozdziału, poza tym co tutaj jest opisane.

Jak dotąd masz kilka warunków, które pomogą Ci zbudować formułę: co się stanie, gdy $t = 0$, co się stanie gdy $t = 1$, co się stanie, gdy $t > 1$, i co się stanie gdy $t > 36$ (czyli całkowita liczba miesięcy).

Zacznij do warunku $t = 0$. Wszystkie warunki trafią do formuły WARUNKI. Na początek w komórce C5 umieść taką formułę:

```
=WARUNKI(
    [@t] = 0; ŚREDNIA(B6:B17)
)
```

Dzięki temu uzyskasz średnią z danych z pierwszego roku, która będzie stanowić początkowy poziom. Następnie w komórce D5 utwórz prognozę jednego kroku. W tym przypadku musisz rozpatrzyć dwa scenariusze, w czym pomoże Ci ta formuła:

```
=JEŻELI([@t] <> 0; PRZESUNIĘCIE([@Szacowany poziom];-1; 0); 0)
```

Ta formuła sprawdza, czy czas nie jest równy 0 (przypadek 1.). Jeśli nie (przypadek 2.), chcesz, by prognoza pierwszego kroku była równa poprzedniej wartości. Formuła PRZESUNIĘCIE idzie do komórki obok, a następnie idzie jeden wiersz do góry (za to odpowiada drugi parametr funkcji równy -1). Jeśli jesteś w punkcie 0, to nie ma poprzednich wartości, więc ze względów estetycznych zwracane jest zero. Teraz w komórce E5 umieść bardzo prostą formułę obliczającą błąd prognozy:

```
=[@Popyt]-[@Prognoza jednego kroku]]
```

ZASTOSOWANIE FUNKCJI LAMBDA

Przyjrzyj się nowej funkcji Excela — funkcji LAMBDA.

W tym rozdziale będziesz często zaciągał dane z poprzedniego okresu. To znaczy, że będziesz przesuwiał się o jedną lub więcej komórek w górę. Niestety pracując na tabeli Excela, łatwo pomylić odwołania. Jeśli jesteś w komórce C6 i musisz pobrać dane z komórki C5, to czy powinieneś użyć odwołania bezpośredniego (C5), czy raczej przesunąć się w górę o jeden wiersz za pomocą formuły =PRZESUNIĘCIE([@Szacowany poziom]; -1; 0). Druga formuła jest o wiele dłuższa, ale lepiej się sprawdzi w przypadku tabeli Excela. Przy tej okazji warto wspomnieć o funkcji LAMBDA.

Pokrótko opiszę tę funkcję, a Ty możesz ją spróbować zaimplementować samodzielnie. Funkcja LAMBDA umożliwia definiowanie własnych funkcji Excela.

Zobacz, jak to działa na tym przykładzie. Możesz zacząć od ogólnej formuły z użyciem funkcji LAMBDA:

```
=LAMBDA(bieżący_wiersz; PRZESUNIĘCIE(bieżący_wiersz; -1; 0))
```

Pierwsze parametry funkcji LAMBDA to wartości, które chcesz przekazywać do stworzonej przez siebie funkcji. Wystarczy, że podasz nazwę zmiennej. Ostatnim parametrem jest sama funkcja, którą chcesz utworzyć przy użyciu podanych wcześniej zmiennych. Jeśli wpiszesz podaną wyżej formułę do arkusza, prawdopodobnie będziesz miał błąd #OBL!. Wynika to z faktu, że to ogólna funkcja, do której nie przekazałeś żadnych wartości. Aby sprawdzić działanie tej funkcji, na końcu musisz podać przekazywane do niej wartości w nawiasie (traktując deklarację funkcji LAMBDA jak nazwę funkcji). Spójrz na wywołanie tej formuły pokazane na rysunku 8.4.

	A	B	C	D	E	F	G	H	I	J
1		3								
2		3								
3										

Rysunek 8.4. Test funkcji LAMBDA przez przekazanie parametrów

Jak widać, do funkcji LAMBDA została przekazana komórka B2 jako parametr *bieżący_wiersz*. Następnie podany argument jest przetwarzany przez funkcję PRZESUNIĘCIE. Ale tak naprawdę funkcja LAMBDA daje najwięcej korzyści podczas pracy z nazwanymi zakresami. I tak możesz wziąć wcześniej zdefiniowaną formułę (bez dodatkowych nawiasów z przekazywanymi wartościami) i przypisać ją do nazwanego zakresu. Po skopiowaniu formuły przejdź do zakładki *Formuły* i naciśnij przycisk *Definiuj nazwę*. W polu *Nazwa* wpisz np. **PobierzPoprzedniOkres** i w polu *Odwołuje się do* wklej funkcję LAMBDA (rysunek 8.5). Na koniec naciśnij *OK*.

Nowa nazwa — □ ×

Nazwa:

Zakres: ▾

Komentarz:

Odwołuje się do: ↕

Rysunek 8.5. Przypisywanie funkcji LAMBDA do nazwanego zakresu

Od tego momentu funkcja LAMBDA może być traktowana jak nowa funkcja Excelsa, którą możesz wywołać tak: `PobierzPoprzedniOkres(odwołanie_do_komórki)`. Ponieważ tematem niniejszej książki nie jest programowanie, nie będziemy używać funkcji LAMBDA, ale mam nadzieję, że dostrzeżasz jej potencjał w swoich analizach.

Twój arkusz powinien teraz wyglądać tak, jak pokazano na rysunku 8.6.

	A	B	C	D	E
1	Liczba miesięcy		Parametr wygładzania poziomu (alfa)		
2	36		0,50		
3					
4	t	Popyt	Szacowany poziom	Prognoza jednego	Błąd prognozy
5	0		163,000	0,000	0,000
6	1	165	#N/D	163,000	2,000
7	2	171	#N/D	#N/D	#N/D
8	3	147	#N/D	#N/D	#N/D
9	4	143	#N/D	#N/D	#N/D
10	5	164	#N/D	#N/D	#N/D
11	6	160	#N/D	#N/D	#N/D
12	7	152	#N/D	#N/D	#N/D
13	8	150	#N/D	#N/D	#N/D
14	9	159	#N/D	#N/D	#N/D
15	10	169	#N/D	#N/D	#N/D
16	11	173	#N/D	#N/D	#N/D
17	12	203	#N/D	#N/D	#N/D
18	13	169	#N/D	#N/D	#N/D
19	14	166	#N/D	#N/D	#N/D
20	15	162	#N/D	#N/D	#N/D
21	16	147	#N/D	#N/D	#N/D
22	17	188	#N/D	#N/D	#N/D
23	18	161	#N/D	#N/D	#N/D

Rysunek 8.6. Generowanie prognozy jednego kroku, błędu prognozy oraz szacunkowej wartości dla okresu I.

Czas zebrać całą prognozę razem. Wprowadź zmiany w formule w komórce C5, by móc przeciągnąć prognozę jednego kroku w dół. Gdy wartość t jest większa od 0, to następny szacowany poziom ma być równy poprzedniemu okresowi powiększonemu o parametr wygładzania pomnożony przez błąd prognozy. Teraz formuła w komórce C5 wygląda tak (dodana część kodu została pogrubiona):

```
=WARUNKI(
    [@t] = 0; ŚREDNIA($B$6:$B$17);
    [@t] <= $A$2; [@Prognoza jednego kroku] + $C$2 *[@Błąd prognozy]
)
```

Zastanów się, co ta formuła robi. Na tym etapie Twój model może zastosować szacowany poziom, gdy wartość t mieści się w zakresie [1, 36]. Widać to w powyżej podanej formule. W komórce A2 całkowita liczba miesięcy wynosi 36. Zatem jeśli t mieści się w zakresie od 1 do 36, szacowany poziom jest równy poprzedniej wartości zwiększonej o wykładniczy współczynnik wygładzania pomnożony przez błąd prognozy. To obliczenie widzimy w drugim warunku formuły WARUNKI.

Żeby było śmiesznie, wykonałeś już prawie wszystkie operacje, więc możesz nacisnąć *Enter*. Model prognozy wypełni Twoją tabelę wartościami. Jeśli chcesz przygotować prognozę dla okresu po 36 miesiącach, musisz użyć wartości z komórki C41, 271,648. Porównaj swój arkusz z moim, pokazanym na rysunku 8.7.

t	Popyt	Szacowany poziom	Prognoza jednego	Błąd prognozy	
22	17	188	172,449	156,897	31,103
23	18	161	166,724	172,449	-11,449
24	19	162	164,362	166,724	-4,724
25	20	169	166,681	164,362	4,638
26	21	185	175,841	166,681	18,319
27	22	188	181,920	175,841	12,159
28	23	200	190,960	181,920	18,080
29	24	229	209,980	190,960	38,040
30	25	189	199,490	209,980	-20,980
31	26	218	208,745	199,490	18,510
32	27	185	196,873	208,745	-23,745
33	28	199	197,936	196,873	2,127
34	29	210	203,968	197,936	12,064
35	30	193	198,484	203,968	-10,968
36	31	211	204,742	198,484	12,516
37	32	208	206,371	204,742	3,258
38	33	216	211,186	206,371	9,629
39	34	218	214,593	211,186	6,814
40	35	264	239,296	214,593	49,407
41	36	304	271,648	239,296	64,704
42					

Rysunek 8.7. Prognoza uzyskana w wyniku prostego wygładzania wykładniczego przy parametrze alfa równym 0,5

Czy to najlepsze rozwiązanie, jakie możesz uzyskać? Prognozę tę możesz zoptymalizować, modyfikując parametr *alfa* — im większa jego wartość, tym mniejszy wpływ starszych wartości popytu na prognozę.

Optymalizacja błędu pojedynczego kroku

Dopasowując regresję w rozdziale 6., minimalizowałeś sumę kwadratów błędów. W podobny sposób możesz znaleźć optymalną wartość parametru wygładzania — wystarczy, że zminimalizujesz sumę kwadratów błędów kolejnych prognoz.

W kolumnie F umieść formuły obliczające wartości błędów podniesione do kwadratu (kolumna E). Na koniec w komórce E2 oblicz sumę kwadratów błędów. Teraz Twój arkusz będzie wyglądał tak, jak pokazano na rysunku 8.8.

Uwaga

W gotowym pliku pobranym z materiałów dla tej książki, w kolumnie F, znajdziesz formułę =JEŻELI.BŁĄD([@[Błąd prognozy]]; 0)^2. Jeśli chcesz, możesz ją zmienić. Będziesz dodawać kolejne kolumny do tabeli i niektóre komórki z kolumny z błędem prognozy będą zawierać błędy #N/D. Mimo tego uzyskamy poprawne odpowiedzi, gdyż zoptymalizujemy wyniki, zanim wprowadzimy błąd #N/D. Jednak wciąż po zakończonej pracy w komórce E2 będzie widnieć błąd #N/D. Formuła JEŻELI.BŁĄD pomoże temu zapobiec.

WYSZUKAJ.PIONO... : X ✓ fx =SUMA(SES.Mieiec[Błąd podniesiony do kwadratu])						
	A	B	C	D	E	F
1	Liczba miesięcy		Parametr wygładzania poziomu (alfa)		Suma kwadratów błędów	
2	36		0,50		kwadratu])	
3						
4	t	Popyt	Szacowany poziom	Prognoza jedneg	Błąd prognoz	Błąd podniesiony do kwadratu
5	0		163,000	0,000	0,000	0,000
6	1	165	164,000	163,000	2,000	4,000
7	2	171	167,500	164,000	7,000	49,000
8	3	147	157,250	167,500	-20,500	420,250
9	4	143	150,125	157,250	-14,250	203,063
10	5	164	157,063	150,125	13,875	192,516

Rysunek 8.8. Suma kwadratów błędów prostego wygładzania wykładniczego

Ponadto w komórce F2 musisz umieścić formułę obliczającą **błąd standardowy**. Błąd ten jest równy pierwiastkowi kwadratowemu z sumy kwadratów błędów podzielonemu przez 35 (36 miesięcy minus liczba parametrów wygładzających modelu, która w przypadku prostego wygładzania wykładniczego jest równa 1).

Błąd standardowy jest szacowaną wartością standardowego odchylenia błędu kolejnego kroku. W skrócie jest to miara rozpiętości błędu. Innymi słowy, pokazuje, jak bardzo średnia wszystkich punktów danych będzie się (prawdopodobnie) różnić od rzeczywistej średniej całego zestawu.

Jeżeli uda Ci się uzyskać optymalnie dopasowany model, średnia wartość błędu będzie równa 0. Innymi słowy: taka prognoza będzie **nieobciążona**. Model taki będzie szacował zbyt wysoką wartość popytu tak samo często jak wartość zbyt niską. Błąd standardowy ujmuje ilościowo rozrzut wokół zera przy nieobciążonej prognozie.

Błąd standardowy możesz obliczyć, umieszczając następującą formułę w komórce F2 (rysunek 8.9):

$$=PIERWIASTEK(\$E\$2/(\$A\$2-1))$$

WYSZUKAJ.PIONO... : X ✓ fx =PIERWIASTEK(SES2/(\$A\$2-1))						
	A	B	C	D	E	F
1	Liczba miesięcy		Parametr wygładzania poziomu (alfa)		Suma kwadratów błędów	Błąd standardowy
2	36		0,50		15346,86	=PIERWIASTEK(SES2/(\$A\$2-1))
3						
4	t	Popyt	Szacowany poziom	Prognoza jedneg	Błąd prognoz	Błąd podniesiony do kwadratu
5	0		163,000	0,000	0,000	0,000
6	1	165	164,000	163,000	2,000	4,000
7	2	171	167,500	164,000	7,000	49,000
8	3	147	157,250	167,500	-20,500	420,250
9	4	143	150,125	157,250	-14,250	203,063
10	5	164	157,063	150,125	13,875	192,516
11	6	160	158,531	157,063	2,938	8,629

Rysunek 8.9. Obliczanie błędu standardowego

Błąd standardowy osiągnie wartość *20,94* przy parametrze *alfa* równym 0,5 (rysunek 8.7). Możliwe, że ze swoich zajęć ze statystyki pamiętasz regułę rozkładu

normalnego, która mówi, że trzy pierwsze standardowe odchylenia lub rozkład normalny pokrywają 68 procent, 95 procent i 99,7 procent wartości. Zgodnie z tą regułą 68% błędów prognoz kolejnego kroku powinno być mniejsze od 20,94 i większe od -20,94.

Zależy Ci na maksymalnym zmniejszeniu wartości błędu standardowego poprzez określenie optymalnej wartości parametru *alfa*. Mógłbyś sprawdzić wpływ różnych wartości parametru *alfa* na wartość błędu standardowego, ale optymalną wartość określisz najszybciej za pomocą narzędzia Solver.

Tym razem proces konfiguracji Solvera będzie bardzo prosty. Otwórz okno *Parametry dodatku Solver*. Jako cel optymalizacji określ błąd standardowy (komórka F2), a jako zmienianą komórkę zadeklaruj C2 zawierającą parametr *alfa*. Ponadto dodaj ograniczenie, przy którym wartość znajdująca się w komórce C2 będzie mniejsza od 1, i zaznacz opcję *Ustaw wartości nieujemne dla zmiennych bez ograniczeń*. Rekurencyjne obliczenia wykonywane przy określaniu błędu każdej kolejnej prognozy mają wysoce nieliniowy charakter, a więc w celu uzyskania optymalnej wartości parametru *alfa* musisz skorzystać z metody ewolucyjnej.

Okno *Parametry dodatku Solver* powinno wyglądać tak, jak pokazano na rysunku 8.10. Kliknij przycisk *Rozwiąż*. Uzyskasz parametr *alfa* równy 0,732, przy którym błąd standardowy będzie wynosił 20,393. To niewielka poprawa.

Parametry dodatku Solver

Ustaw cel:

Na: Maks Min Wartość:

Przez zmienianie komórek zmiennych:

Podlegających ograniczeniom:

Ustaw wartości nieujemne dla zmiennych bez ograniczeń

Wybierz metodę rozwiązywania:

Metoda rozwiązywania

W przypadku gładkich nieliniowych problemów dodatku Solver wybierz aparat nieliniowy GRG. Dla liniowych problemów dodatku Solver wybierz aparat LP simpleks, natomiast w przypadku problemów, które nie są gładkie, wybierz aparat ewolucyjny.

Pomoc Rozwiąż Zamknij

Rysunek 8.10. Konfiguracja dodatku Solver umożliwiająca optymalizację parametru *alfa*

Czas przedstawić to na wykresie

Najlepszym sposobem na sprawdzenie prognozy w praktyce jest umieszczenie jej na wykresie obok danych popytu z przeszłości. Zobaczysz wtedy, jak przewidywana wartość ma się do zgromadzonych danych popytu.

Najpierw zwiększ liczbę miesięcy do 48. W rezultacie tabela automatycznie się powiększy. Następnie dodaj dwie kolumny: **Wartość rzeczywista** i **Prognoza** (rysunek 8.11). Na podstawie tej tabeli utworzysz wykres.

t	Popyt	Szacowany poziom	Prognoza jednego	Błąd prognozy	Błąd podniesiony do kwadratu	Wartość rzeczywista	Prognoza
34	29	210	206,575	197,221	-12,779	163,302	
35	30	193	196,638	206,576	-13,576	184,318	
36	31	211	207,151	196,637	14,363	206,288	
37	32	208	207,773	207,152	0,848	0,719	
38	33	216	213,795	207,773	8,227	67,686	
39	34	218	216,873	213,796	4,204	17,675	
40	35	264	251,370	216,874	47,126	2220,891	
41	36	304	289,896	251,374	52,626	2769,457	
42	37		77,693	289,901	-289,901	84042,617	
43	38		#N/D	#N/D	#N/D	#N/D	
44	39		#N/D	#N/D	#N/D	#N/D	
45	40		#N/D	#N/D	#N/D	#N/D	
46	41		#N/D	#N/D	#N/D	#N/D	
47	42		#N/D	#N/D	#N/D	#N/D	
48	43		#N/D	#N/D	#N/D	#N/D	
49	44		#N/D	#N/D	#N/D	#N/D	
50	45		#N/D	#N/D	#N/D	#N/D	
51	46		#N/D	#N/D	#N/D	#N/D	
52	47		#N/D	#N/D	#N/D	#N/D	
53	48		#N/D	#N/D	#N/D	#N/D	

Rysunek 8.11. Kolumny Wartość rzeczywista i Prognoza dodane do tabeli

Teraz w kolumnie Wartość rzeczywista wpisz taką formułę:

```
=JEŻELI(ORAZ([@t] >0;
    [@t] <= $A$2); [@Popyt];
    BRAK())
```

A w kolumnie Prognoza umieść następującą formułę:

```
=JEŻELI([@t] > $A$2;
    INDEKS([Szacowany poziom]; $A$2 + 1);
    BRAK())
```

Formuła dla wartości rzeczywistych po prostu zaciąga znane wartości popytu. Natomiast formuła dla prognozy zaciąga ostatni szacowany poziom. Zwróć uwagę, że liczba miesięcy (36) została zwiększona o 1, by uwzględnić to, że dane zaczynają się od t = 0, a nie t = 1. Teraz możesz wykreślić obie serie danych na jednym wykresie.

Podoba mi się wygląd wykresu punktowego z prostymi liniami. Najpierw dodaj czas (t) z kolumny A. Następnie przytrzymując przycisk *Ctrl*, zaznacz kolumnę Wartość rzeczywista (kolumna G). Mając zaznaczone te dwie kolumny, dodaj wykres punktowy z prostymi liniami, który znajdziesz w zakładce *Wstawianie*.

Teraz zaznacz kolumnę Prognoza. Naciśnij *Ctrl+C* (na komputerze Mac: *Cmd+C*), by skopiować dane. Potem kliknij wykres i naciśnij *Ctrl+V* (na komputerze Mac: *Cmd+V*), by wkleić dane. Możesz dodać etykiety i popracować nad formatowaniem, a uzyskasz wykres podobny do tego, który przedstawiłem na rysunku 8.12.



Rysunek 8.12. Wykres przedstawiający efekty ostatniej prognozy uzyskanej za pomocą prostego wygładzania wykładniczego

Być może dane zawierają trend

Przyglądając się temu wykresowi, można zauważyć kilka rzeczy. Przede wszystkim po przeprowadzeniu prostego wygładzania wykładniczego uzyskałeś linię prostą — prognozę pewnego poziomu — ale popyt z ostatnich 36 miesięcy wydaje się rosnać. Trend ten szczególnie uwidacznia się w ciągu kilku ostatnich miesięcy.

Ale jak udowodnić istnienie trendu?

Dokonasz tego, dopasowując do danych popytu regresję liniową i przeprowadzając test t na linii trendu. Jeżeli nachylenie linii trendu nie jest zerowe ani istotne z punktu widzenia statystyki (charakteryzuje się wartością p mniejszą od 0,05 w teście t), możesz uznać, że dane charakteryzują się pewnym trendem.

Wróć do arkusza *Szereg czasowy*. Czas przeprowadzić test trendu. Tym razem skorzystasz z wbudowanej w Excela funkcji zwracającej parametry trendu liniowego REGLINP, określisz współczynnik regresji, jego błąd standardowy, a także stopnie swobody. Następnie obliczysz wartość współczynnika t zastosujesz ją jako argument funkcji ROZKŁAD.T.

Jeżeli nigdy wcześniej nie korzystałeś z funkcji REGLINP, to warto, abyś zajrzał do jej opisu w pomocy programu Excel. Do funkcji tej należy skierować dane zmiennych zależnych (dane popytu umieszczone w kolumnie B) i dane zmiennej niezależnej (w analizowanym przykładzie masz tylko jedną zmienną niezależną — czas zdefiniowany w kolumnie A).

Ponadto do wspomnianej funkcji musisz podać również argument PRAWDA w celu normalnego obliczenia ostatniego parametru równania opisującego linię regresji, a także drugi argument PRAWDA, dzięki któremu uzyskasz szczegółowe statystyki, takie jak błąd standardowy i współczynnik determinacji. W przypadku danych znajdujących się w arkuszu *Szereg czasowy* trend regresji liniowej można wyznaczyć za pomocą formuły:

=INDEKS(REGLINP(MieczeDane[Popyt];MieczeDane[t];PRAWDA;PRAWDA);1;1)

Wpisz tę formułę w komórce B39 (a w komórce A39 umieść etykietę **Współczynnik regresji**).

Wywołanie tej funkcji zwróci tylko współczynnik regresji. Stanie się tak, ponieważ formuła REGLINP jest tablicowa — dane statystyczne regresji są przez nią zwracane w postaci tablicy. W związku z tym możesz wywołać ją jako formułę tablicową i zapisać zwrócone dane w określonym zakresie komórek arkusza lub umieścić ją wewnątrz formuły INDEKS w celu uzyskania tylko interesujących Cię pojedynczych wartości. Wolę tę drugą metodę, gdyż łatwiej się z nią pracuje.

W celu odczytania wartości błędu standardowego w komórce B40 umieść formułę:

=INDEKS(REGLINP(MieczeDane[Popyt];MieczeDane[t];PRAWDA;PRAWDA);2;1)

W celu odczytania wartości stopni swobody regresji w komórce B41 umieść formułę:

=INDEKS(REGLINP(MieczeDane[Popyt];MieczeDane[t];PRAWDA;PRAWDA);4;2)

Formuła ta powinna zwrócić wartość równą 34. Jest ona obliczana jako różnica liczby danych (36) i liczby współczynników regresji liniowej (2).

Dysponujesz już trzema wartościami niezbędnymi do przeprowadzenia testu *t* określającego statystyczną istotność trendu. Wartość współczynnika *p* testu *t* można obliczyć za pomocą funkcji ROZKŁ.T.DS, do której należy jako argumenty podać moduł ilorazu współczynnika regresji i błędu standardowego i liczbę stopni swobody. W związku z tym w komórce B42 umieść formułę:

=ROZKŁ.T.DS(MODUŁ.LICZBY(B39/B40);B41)

Formuła ta zwróci wartość bliską zeru, z czego wynika, że jeżeli trend by w rzeczywistości nie istniał (współczynnik regresji w takim przypadku jest równy 0), to nie ma możliwości uzyskania bardzo odbiegającej wartości współczynnika regresji (rysunek 8.13).

	A	B
39	Współczynnik regresji	2,54
	Błąd standardowy współczynnika	
40	regresji	0,34
41	Stopnie swobody	34
42	Współczynnik p testu	1,17E-08

Rysunek 8.13. Zauważony przez Ciebie trend istnieje w rzeczywistości

No dobrze, udowodniłeś istnienie trendu. Teraz musisz wziąć go pod uwagę w swoich prognozach.

Podwójne wygładzanie wykładnicze (metoda Holta)

Podwójne wygładzanie wykładnicze (metoda Holta) to rozszerzona wersja metody wygładzania wykładniczego umożliwiająca tworzenie prognoz na podstawie danych, które charakteryzują się trendem liniowym. Metoda ta jest również określana mianem **szeregu liniowego Holta**. W przypadku prostego wygładzania wykładniczego korzystałeś tylko z jednego parametru wygładzającego (*alfa*) i jednego komponentu redukującego błąd prognozy. Tym razem będziesz stosował po dwa takie parametry.

Jeżeli szereg czasowy ma trend liniowy, to możesz go zapisać za pomocą następującego równania:

zapotrzebowanie w chwili $t = \text{poziom} + t \cdot \text{trend} + \text{losowy błąd wokół poziomu w chwili } t$

Najnowsze oszacowane wartości poziomu i trendu umożliwiają prognozowanie popytu w przyszłości. Jeżeli masz 36. miesiąc, to w jaki sposób możesz utworzyć dobrą prognozę na miesiąc 38.? Do najnowszej szacowanej wartości poziomu należy dodać dwa miesiące trendu. A co z 40. miesiącem? Do poziomu należy *dodać* cztery miesiące trendu. Może nie jest to tak łatwe jak proste wygładzanie wykładnicze, ale wciąż nie jest to nic skomplikowanego.

Podobnie jak w przypadku prostego wygładzania wykładniczego musisz oszacować pewne wartości początkowe. Tym razem potrzebujesz początkowych wartości poziomu i trendu — powinieneś określić *poziom₀* i *trend₀*. Jednym z popularnych sposobów na ustalenie tych wartości jest przedstawienie pierwszej połowy danych popytu na wykresie i wygenerowanie linii trendu. Krzywizna wykresu wyznacza *trend₀*, a punkt przecięcia wykresu z osią *y* wyznacza *poziom₀*.

Korzystając z metody Holta, musisz obliczyć wartości wynikające z dwóch równań: jedno umożliwia określenie zmiany wartości poziomu w czasie, a drugie umożliwia określenie zmiany trendu w czasie. W równaniu określającym poziom wciąż znajduje się parametr wygładzający *alfa*, a w równaniu określającym trend znajduje się parametr *gamma*. Oba parametry działają tak samo — mogą przyjmować wartości z zakresu od 0 do 1 i określają wpływ błędu szacunku jednego kroku na całość prognozy.

Oto równanie umożliwiające obliczenie wartości kolejnego poziomu:

$$\text{poziom}_1 = \text{poziom}_0 + \text{trend}_0 + \text{alfa} \cdot (\text{popyt}_1 - (\text{poziom}_0 + \text{trend}_0))$$

Zauważ, że $(\text{poziom}_0 + \text{trend}_0)$ jest prognozą kolejnego kroku — wartości pierwszego miesiąca znajdujących się bezpośrednio po wartościach początkowych, a więc

$(popyt_1 - (poziom_0 + trend_0))$ jest wartością błędu kolejnego kroku. Równanie to wygląda bardzo podobnie do równania poziomu prostego wygładzania wykładniczego, ale tym razem przy każdym kolejnym kroku brany jest pod uwagę trend. W związku z tym ogólne równanie pozwalające na oszacowanie poziomu wygląda następująco:

$$poziom_{bieżący\ okres} = poziom_{poprzedni\ okres} + trend_{poprzedni\ okres} + alfa \cdot (popyt_{bieżący\ okres} - (poziom_{poprzedni\ okres} + trend_{poprzedni\ okres}))$$

Nowa technika wygładzania wymaga również korzystania z równania aktualizującego wartość trendu. Dla pierwszego miesiąca będzie ono miało postać:

$$trend_1 = trend_0 + gamma \cdot alfa \cdot (popyt_1 - (poziom_0 + trend_0))$$

Jak widzisz, równanie trendu wygląda podobnie do równania poziomu. Poprzednia szacunkowa wartość trendu jest sumowana z iloczynem współczynnika *gamma* oraz błędu związanego z aktualizacją poziomu, co na pierwszy rzut oka wydaje się sensownym zabiegiem, ponieważ tylko część wartości błędu używanej do dopasowania poziomu wynika z niedoszacowania trendu.

W związku z tym równanie służące do szacowania trendu można zapisać za pomocą następującej ogólnej postaci:

$$trend_{bieżący\ okres} = trend_{poprzedni\ okres} + gamma \cdot alfa \cdot (popyt_{bieżący\ okres} - (poziom_{poprzedni\ okres} + trend_{poprzedni\ okres}))$$

Metoda Holta w arkuszu kalkulacyjnym

Zacznij od utworzenia nowego arkusza o nazwie **Metoda Holta**. W tym arkuszu, podobnie jak w arkuszu *SES*, umieść, zaczynając od 4. wiersza, dane szeregu czasowego, a w wierszu 5. utwórz pusty wiersz na szacunkowe wartości początkowe. Tabelę nazwij **Holt.MieczyDane**.

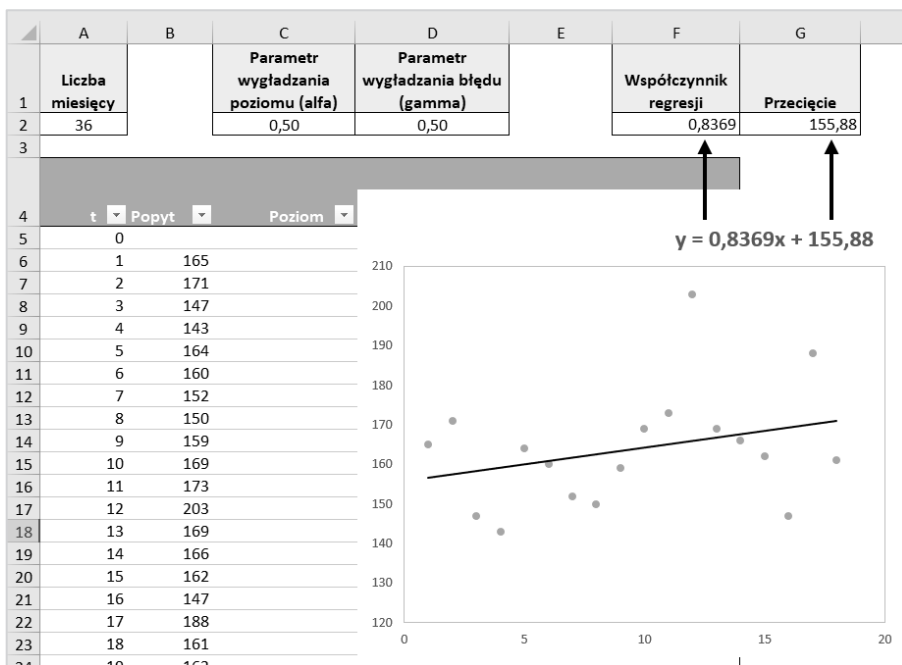
Kolumna C będzie ponownie zawierała szacunkowe wartości poziomu, a w kolumnie D umieścisz szacunkowe wartości trendu. Dlatego nad tymi kolumnami umieść wartości parametrów *alfa* i *gamma*. Za chwilę uzyskasz optymalne wartości tych parametrów za pomocą dodatku Solver, ale na razie podaj wstępne wartości równe 0,5. Teraz Twój wykres powinien wyglądać tak, jak pokazano na rysunku 8.14.

Czas ustalić początkowe wartości poziomu i trendu, które należy umieścić w komórkach C5 i D5. W tym celu utwórz wykres danych z pierwszych 18 miesięcy zawierający linię trendu i równanie tej linii (kliknij prawym przyciskiem myszy dane i wybierz *Dodaj linię trendu...*, następnie w panelu *Formatowanie linii trendu* zaznacz opcję *Wyświetl równanie na wykresie*). Okaże się, że początkowa wartość trendu jest równa 0,8369, a początkowy poziom jest równy 155,88 (początkowy poziom określa miejsce przecięcia linii trendu z osią *y*). Zapisz te wartości w komórkach F5 i G2. W ten sposób uzyskasz arkusz przedstawiony na rysunku 8.15.

284 Mistrz analizy danych

	A	B	C	D
1	Liczba miesięcy		Parametr wygładzania poziomu (alfa)	Parametr wygładzania błędu (gamma)
2	36		0,50	0,50
3				
4	t	Popyt		
5	0			
6	1	165		
7	2	171		
8	3	147		
9	4	143		
10	5	164		
11	6	160		
12	7	152		
13	8	150		
14	9	159		
15	10	169		
16	11	173		
17	12	203		
18	13	169		
19	14	166		
20	15	162		
21	16	147		
22	17	188		
23	18	161		
24	19	162		

Rysunek 8.14. Zacznij od przypisania parametrom wygładzającym wartości równych 0,5



Rysunek 8.15. Początkowe wartości poziomu i trendu

Tak jak wcześniej dodaj cztery nowe kolumny: **Poziom**, **Trend**, **Prognoza jednego kroku** i **Błąd prognozy**, odpowiednio, w kolumnach C, D, E i F.

Tak jak w poprzednim podrozdziale, dodamy formuły tablicowe w celu dalszej analizy tych danych.

Zacznij od wypełnienia kolumn dla wiersza $t = 0$. Pamiętaj, że $t = 0$ to swoisty przypadek, gdyż jest pierwszym analizowanym okresem. W komórkach C5 i D5 umieść następujące formuły:

C5: =JEŻELI([@t]=0; \$G\$2;)

D5: =JEŻELI([@t]=0; \$F\$2;)

Dla $t = 0$ wiadomo, że Poziom i Trend są równe przecięciu i współczynnikowi regresji. Natomiast prognoza jednego kroku jest po prostu równa wartości poprzedniego poziomu zsumowanej z poprzednią wartością trendu. Zatem w pustej komórce kolumny dla prognozy jednego kroku umieść następującą formułę:

=JEŻELI([@t]<>0;
 PRZESUNIĘCIE([@Poziom];-1;0) + PRZESUNIĘCIE([@Trend];-1;0);
 0
)

Błąd prognozy to proste wygładzanie wykładnicze, czyli wartość komórki F6 pomniejszona o prognozę jednego kroku:

=JEŻELI([@t]<>0;
 [@Popyt] - [@[Prognoza jednego kroku]];
 0
)

Teraz możesz zaktualizować wartość poziomu — nowa wartość będzie równa wartości poprzedniego poziomu zsumowanej z poprzednią wartością trendu, a także iloczynem współczynnika *alfa* i błędu prognozy:

=JEŻELI([@t]=0;
 \$G\$2;
 SUMA(PRZESUNIĘCIE([@[Poziom]:[Trend]];-1;0) + C\$2*[@[Błąd prognozy]])
)

Wartość trendu (komórka D6) należy zaktualizować, czyli zdefiniować ją jako sumę poprzedniego trendu i iloczynu współczynnika *gamma*, współczynnika *alfa* i błędu prognozy:

=JEŻELI([@t]=0;
 \$F\$2;
 PRZESUNIĘCIE([@Trend]; -1; 0) +D\$2*C\$2*[@[Błąd prognozy]]
)

Formuły przeciągną się automatycznie w dół, a Twój arkusz powinien wyglądać tak, jak pokazano na rysunku 8.16.

	A	B	C	D	E	F	G
1	Liczba miesięcy		Parametr wygładzania poziomu (alfa)	Parametr wygładzania błędu (gamma)		Współczynnik regresji	Przecięcie
2	36		0,50	0,50		0,8369	155,88
3							
4	t	Popyt	Poziom	Trend	Prognoza jednego kroku	Błąd prognozy	
5	0		155,88	0,837	0,000	0,000	
6	1	165	160,86	2,908	156,717	8,283	
7	2	171	167,38	4,716	163,766	7,234	
8	3	147	159,55	-1,559	172,099	-25,099	
9	4	143	150,50	-5,306	157,991	-14,991	
10	5	164	154,59	-0,604	145,189	18,811	
11	6	160	157,00	0,899	153,991	6,009	
12	7	152	154,95	-0,575	157,894	-5,894	
13	8	150	152,19	-1,668	154,372	-4,372	
14	9	159	154,76	0,453	150,518	8,482	
15	10	169	162,11	3,900	155,212	13,788	
16	11	173	169,50	5,648	166,005	6,995	
17	12	203	189,08	12,611	175,151	27,849	
18	13	169	185,34	4,439	201,686	-32,686	
19	14	166	177,89	-1,506	189,782	-23,782	
20	15	162	169,19	-5,103	176,385	-14,385	

Rysunek 8.16. Obliczanie poziomu, trendu, prognozy i błędu

Prognozowanie popytu w przyszłości

Tak jak wcześniej, dodaj dwie nowe kolumny do tabeli: **Wartość rzeczywista** i **Prognoza**. Następnie przeciągnij wartości w kolumnie A aż do 48. miesiąca. Umieść następujące formuły w kolumnach, które przed chwilą dodałeś.

W kolumnie **Wartość rzeczywista** umieść taką formułę:

```
=JEŻELI(ORAZ([@t] >0; [@t] <= $A$2);
    [@Popyt];
    BRAK()
)
```

A w kolumnie **Prognoza** wpisz tę formułę:

```
=JEŻELI([@t] > $A$2;
    INDEKS([Poziom]; $A$2+1) + ([@t] - $A$2) * INDEKS([Trend]; $A$2+1);
    BRAK()
)
```

To są prognozy dla 37. miesiąca i później. Dodajesz ostateczny poziom do liczby prognozowanych miesięcy *pomnożonej* przez ostateczny szacowany trend.

Na tym etapie Twój arkusz powinien wyglądać jak ten na rysunku 8.17.

WYSZUKAJ.PIONO... X ✓ fx

=JEŻELI([@t]>\$A\$2;
INDEKS([Poziom];\$A\$2+1) + ([@t]-\$A\$2) * INDEKS([Trend];\$A\$2+1);
BRAK()
)

t	Popyt	Poziom	JEŻELI(test_logiczny; [wartość_jeżeli_prawda]; [wartość_jeżeli_fałsz])	ść rzecz	Prognoza			
33	28	199	198,18	1,478	196,602	2,398	199,000	#N/D
34	29	210	206,48	1,840	199,660	10,340	210,000	#N/D
35	30	193	198,22	1,304	208,315	-15,315	193,000	#N/D
36	31	211	207,09	1,706	199,525	11,475	211,000	#N/D
37	32	208	208,27	1,678	208,794	-0,794	208,000	#N/D
38	33	216	213,94	1,890	209,948	6,052	216,000	#N/D
39	34	218	217,26	1,966	215,827	2,173	218,000	#N/D
40	35	264	248,74	3,533	219,225	44,775	264,000	#N/D
41	36	304	286,37	5,344	252,270	51,730	304,000	#N/D
42	37		99,44	-4,868	291,709	-291,709	#N/D	291,709
43	38		32,24	-8,179	94,575	-94,575	#N/D	297,054
44	39		8,20	-9,022	24,062	-24,062	#N/D	302,398
45	40		-0,28	-8,993	-0,819	0,819	#N/D	307,743
46	41		-3,16	-8,668	-9,272	9,272	#N/D	313,087
47	42		-4,03	-8,254	-11,829	11,829	#N/D	318,431
48	43		-4,19	-7,824	-12,287	12,287	#N/D	323,776
49	44		-4,10	-7,403	-12,013	12,013	#N/D	329,120
50	45		-3,92	-7,001	-11,499	11,499	#N/D	334,465
51	46		-3,72	-6,619	-10,921	10,921	#N/D	339,809
52	47		-3,53	-6,257	-10,342	10,342	#N/D	345,153
53	48		-3,33	-5,914	-9,782	9,782	#N/D	BRAK()

Rysunek 8.17. Prognozy na kolejne miesiące uzyskane w wyniku podwójnego wygładzania wykładniczego

Podobnie jak w przypadku prostego wygładzania wykładniczego możesz przedstawić dane popytu z przeszłości oraz dane prognozowanego popytu na wspólnym wykresie punktowym z liniami prostymi (rysunek 8.18).



Rysunek 8.18. Wykres prognozy przy domyślnych wartościach parametrów alfa i gamma

Przy parametrach *alfa* i *gamma* równych 0,5 prognoza nie wygląda na najbardziej przekonującą. Rozpoczyna się od wartości, na której skończył się ostatni miesiąc, a kolejne prognozowane wartości wzrastają dość gwałtownie. Być może trzeba zoptymalizować parametry wygładzania.

Optymalizacja błędu prognozy pojedynczego kroku

W kolumnie I umieść wartości błędu podniesione do kwadratu (to samo rozwiązanie stosowałeś w przypadku prostego wygładzania wykładniczego). Następnie za pomocą formuły pokazanej poniżej przy porównaniu prognozy z wartością rzeczywistą podnieś błąd do kwadratu:

```
=JEŻELI([@t] <= $A$2;
        [Błąd prognozy]^2;
        0
)
```

Zwróć uwagę, że ta formuła sprawdza, czy wartości są mniejsze niż całkowita liczba analizowanych przez Ciebie punktów danych, a nie niż prognozowane wartości ($t \geq 36$). W komórkach H2 i I2 oblicz tak jak wcześniej sumę kwadratów błędów oraz wartość błędu standardowego, aczkolwiek tym razem pracujesz na modelu z dwoma parametrami wygładzającymi, a więc obliczając błąd standardowy, przed pierwiastkowaniem będziesz musiał podzielić sumę kwadratów przez $36 - 2$:

```
=PIERWIASTEK(H2/(A2-2))
```

Po wykonaniu tych operacji powinieneś uzyskać arkusz przedstawiony na rysunku 8.19.

	F	G	H	I
1	Współczynnik regresji	Przecięcie	Suma kwadratów błędów	Błąd standardowy
2	0,8369	155,88	15315,315	21,224
3	Błąd			
4	=SUMA(Holt.MieczeDane[Błąd podniesiony do kwadratu])			
5	0,000	#N/D	#N/D	0,000

Rysunek 8.19. Obliczanie sumy kwadratów błędów oraz błędu standardowego

Dodatek Solver skonfiguruj tak samo jak w przypadku prostego wygładzania wykładniczego, ale tym razem wybierz jednoczesną optymalizację parametrów *alfa* i *gamma* (rysunek 8.20).

Parametry dodatku Solver

Ustaw cel:

Na: Maks Mjn Wartość:

Przez zmienianie komórek zmiennych:

Podlegających ograniczeniom:

Ustaw wartości nieujemne dla zmiennych bez ograniczeń

Wybierz metodę rozwiązywania:

Metoda rozwiązywania
 W przypadku gładkich nieliniowych problemów dodatku Solver wybierz aparat nieliniowy GRG. Dla liniowych problemów dodatku Solver wybierz aparat LP simpleks, natomiast w przypadku problemów, które nie są gładkie, wybierz aparat ewolucyjny.

Pomoc Rozwiąż Zamknij

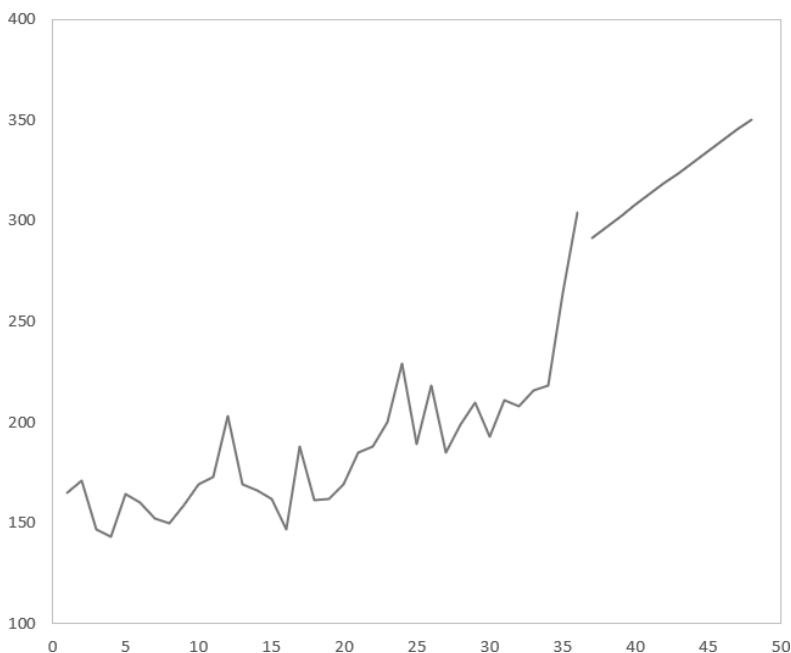
Rysunek 8.20. Optymalizacja parametrów podwójnego wygładzania wykładniczego

Gdy dodatek Solver skończy pracę, uzyskasz optymalną wartość parametru *alfa* (0,66) i parametru *gamma* (0,05). Graficzną interpretację prognozy opartej na tych parametrach przedstawiono za pomocą wykresu punktowego z liniami prostymi na rysunku 8.21.

W obliczeniach zastosowałeś prognozowaną wartość trendu dodającą co miesiąc pięć mieczy. Trend został podwojony w stosunku do trendu z poprzedniego arkusza, ponieważ podwójne wygładzanie wykładnicze opiera prognozę w większym stopniu na ostatnich wartościach popytu, a te miały tendencję wzrostową.

Obie prognozy mają podobne wartości początkowe (290 i 292 dla 37. miesiąca), ale prognoza podwójnego wygładzania wykładniczego zgodnie z oczekiwaniem wzrasta razem z trendem.

Prognoza podwójnego wykładniczego



Rysunek 8.21. Wykres prognozy zoptymalizowanego podwójnego wykładniczego

To wszystko? Analiza autokorelacji

Czy to najlepsza prognoza, jaką możesz uzyskać? Czy wzięłeś pod uwagę wszystko?

Poprawność uzyskanego modelu tworzącego prognozy można sprawdzić, analizując błędy kolejnych kroków prognozy. Jeżeli błędy te mają charakter losowy, to znaczy, że osiągnąłeś najlepszą możliwą prognozę, ale jeżeli wartości błędów są powtarzalne (istnieje jakiś regularny wzór), może to oznaczać, że w danych popytu kryje się jakaś sezonowość, której jeszcze nie wzięłeś pod uwagę.

Pisząc „powtarzalne wartości błędów”, mam na myśli zsynchronizowane zmiany wartości błędów powtarzające się np.: co miesiąc, co dwa miesiące lub co dwanaście miesięcy. Koncepcję korelacji błędów wartości w różnych okresach określamy mianem **autokorelacji**.

Zacznij od utworzenia nowego arkusza o nazwie **Autokorelacja metody Holta**. W tym arkuszu wklej numery miesięcy (1 – 36) albo wypełnij nimi za pomocą funkcji SEKWENCJA. Pamiętaj, że zakres [1, 36] **nie** uwzględnia przypadku $t = 0$, który tylko pomaga w uzyskaniu prognozy. Następnie za pomocą funkcji INDEKS do kolumny B zaciągnij błędy prognozy obliczone w arkuszu *Metoda Holta*.

Pod wartościami błędów, w komórce B38, oblicz ich średnią wartość. Powinienesz teraz dysponować arkuszem wyglądającym tak, jak pokazano na rysunku 8.22.

	A	B
1	t Błąd prognozy jednego kroku	
26	25	-32,23762045
27	26	16,13259637
28	27	-29,94280246
29	28	2,398387897
30	29	10,33952568
31	30	-15,31532409
32	31	11,47512572
33	32	-0,793754823
34	33	6,051575658
35	34	2,173284481
36	35	44,7750903
37	36	51,73048169
38		3,576404834

Rysunek 8.22. Miesiące i przypisane im błędy pojedynczych kroków prognozy

W kolumnie C oblicz odchylenia wszystkich wartości błędów podanych w kolumnie B od średniej w komórce B38. Odchylenia te pozwolą na odkrycie powtarzalności błędów prognozy. Być może co roku popyt wzrasta w grudniu (jest wyraźnie większy od wartości średniej). Tego typu sezonowe powtarzalności zostaną uwidocznione za pomocą tych liczb.

W komórce C2 umieść formułę określającą odchylenie błędu zdefiniowanego w komórce B2 od wartości średniej:

=B2-B\$38

Formułę tę przeciągnij w dół, w wyniku czego uzyskasz odchylenia od średniej wszystkich błędów. W komórce C38 oblicz sumę odchyleń podniesionych do kwadratu za pomocą formuły:

=SUMA.ILOCZYNÓW(\$C2:\$C37;C2:C37)

Po wykonaniu tych czynności otrzymasz arkusz pokazany na rysunku 8.23.

Teraz w kolumnie D „opóźnij” odchylenia błędów o jeden miesiąc. Oznacz kolumnę D etykietą 1. Nie musisz niczego wpisywać w komórce D2, ale w komórce D3 umieść formułę:

=C2

Przeciągnij tę formułę w dół. Twój arkusz powinien teraz wyglądać tak, jak pokazano na rysunku 8.24.

	A	B	C
1	t Błąd prognozy jednego kroku		Odchylenia od średniej
19	18	-16,34746246	-19,924
20	19	-5,157759634	-8,734
21	20	4,837378087	1,261
22	21	17,07536392	13,499
23	22	7,649490212	4,073
24	23	13,16840714	9,592
25	24	31,58878508	28,012
26	25	-32,23762045	-35,814
27	26	16,13259637	12,556
28	27	-29,94280246	-33,519
29	28	2,398387897	-1,178
30	29	10,33952568	6,763
31	30	-15,31532409	-18,892
32	31	11,47512572	7,899
33	32	-0,793754823	-4,370
34	33	6,051575658	2,475
35	34	2,173284481	-1,403
36	35	44,7750903	41,199
37	36	51,73048169	48,154
38		3,576404834	13636,81665

Rysunek 8.23. Suma podniesionych do kwadratu odchyżeń błędów prognozy metody Holta od średniej

	A	B	C	D
1	t Błąd prognozy jednego kroku		Odchylenia od średniej	1
28	27	-29,94280246	-33,519	12,556
29	28	2,398387897	-1,178	-33,519
30	29	10,33952568	6,763	-1,178
31	30	-15,31532409	-18,892	6,763
32	31	11,47512572	7,899	-18,892
33	32	-0,793754823	-4,370	7,899
34	33	6,051575658	2,475	-4,370
35	34	2,173284481	-1,403	2,475
36	35	44,7750903	41,199	-1,403
37	36	51,73048169	48,154	=C36
38		3,576404834	13636,81665	-474,1

Rysunek 8.24. Odchylenia błędów opóźnione o jeden miesiąc

Aby opóźnić błędy o dwa miesiące, zaznacz komórki D1:D37, a następnie przeciągnij je do kolumny E. Aby uzyskać opóźnienie o 12 miesięcy, przeciągnij wspomniane komórki do kolumny O. To proste! Po obliczeniu wszystkich opóźnień uzyskasz macierz pokazaną na rysunku 8.25.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	t	Błąd prognozy jednego kroku	Odchylenia od średniej	1	2	3	4	5	6	7	8	9	10	11	12
2	1	8,2831	4,707												
3	2	7,69681953	-4,120	4,707	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
4	3	-22,77250526	-26,349	4,120	4,707	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
5	4	-12,36223655	-15,939	-26,349	4,120	4,707	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
6	5	16,6194175	13,043	-15,939	-26,349	4,120	4,707	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
7	6	0,917418805	-2,659	13,043	-15,939	-26,349	4,120	4,707	0,000	0,000	0,000	0,000	0,000	0,000	0,000
8	7	-8,467507241	-12,044	-2,659	13,043	-15,939	-26,349	4,120	4,707	0,000	0,000	0,000	0,000	0,000	0,000
9	8	-5,370383003	-8,947	-12,044	-2,659	13,043	-15,939	-26,349	4,120	4,707	0,000	0,000	0,000	0,000	0,000
10	9	6,873440424	3,297	-8,947	-12,044	-2,659	13,043	-15,939	-26,349	4,120	4,707	0,000	0,000	0,000	0,000
11	10	11,80672113	8,230	3,297	-8,947	-12,044	-2,659	13,043	-15,939	-26,349	4,120	4,707	0,000	0,000	0,000
12	11	7,075127991	3,499	8,230	3,297	-8,947	-12,044	-2,659	13,043	-15,939	-26,349	4,120	4,707	0,000	0,000
13	12	31,21443194	27,638	3,499	8,230	3,297	-8,947	-12,044	-2,659	13,043	-15,939	-26,349	4,120	4,707	0,000
14	13	-25,64928706	-29,226	27,638	3,499	8,230	3,297	-8,947	-12,044	-2,659	13,043	-15,939	-26,349	4,120	4,707
15	14	-13,1361464	-16,713	-29,226	27,638	3,499	8,230	3,297	-8,947	-12,044	-2,659	13,043	-15,939	-26,349	4,120
16	15	-9,410359582	-12,987	-16,713	-29,226	27,638	3,499	8,230	3,297	-8,947	-12,044	-2,659	13,043	-15,939	-26,349
17	16	-18,81100961	-22,387	-12,987	-16,713	-29,226	27,638	3,499	8,230	3,297	-8,947	-12,044	-2,659	13,043	-15,939
18	17	34,64293327	31,067	-22,387	-12,987	-16,713	-29,226	27,638	3,499	8,230	3,297	-8,947	-12,044	-2,659	13,043
19	18	-16,34746246	-19,924	31,067	-22,387	-12,987	-16,713	-29,226	27,638	3,499	8,230	3,297	-8,947	-12,044	-2,659
20	19	-5,157759634	-8,734	-19,924	31,067	-22,387	-12,987	-16,713	-29,226	27,638	3,499	8,230	3,297	-8,947	-12,044
21	20	4,837378087	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713	-29,226	27,638	3,499	8,230	3,297	-8,947
22	21	17,07536392	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713	-29,226	27,638	3,499	8,230	3,297
23	22	7,649490212	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713	-29,226	27,638	3,499	8,230
24	23	13,16840714	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713	-29,226	27,638	3,499
25	24	31,58878508	28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713	-29,226	27,638
26	25	-32,23762045	-35,814	28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713	-29,226
27	26	16,13259637	12,556	-35,814	-28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713
28	27	-29,94280246	-33,519	12,556	-35,814	-28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987
29	28	2,398387897	-1,178	-33,519	12,556	-35,814	-28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387
30	29	10,33952568	6,763	-1,178	-33,519	12,556	-35,814	-28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067
31	30	-15,31532409	-18,892	6,763	-1,178	-33,519	12,556	-35,814	-28,012	9,592	4,073	13,499	1,261	-8,734	-19,924
32	31	11,47512572	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	-28,012	9,592	4,073	13,499	1,261	-8,734
33	32	-0,793754823	-2,470	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	-28,012	9,592	4,073	13,499	1,261
34	33	6,051575658	4,375	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	-28,012	9,592	4,073	13,499
35	34	2,173284481	-1,403	2,475	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	-28,012	9,592	4,073
36	35	44,7309903	41,199	-1,403	2,475	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	-28,012	9,592
37	36	51,73048169	48,154	41,199	-1,403	2,475	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	-28,012
38		3,576404834	13636,81665	-474,1	753,4	-229,3	-2223,6	1544,3	-2442,5	1790,5	-4382,6	-167,7	-1460,4	733,1	[02:037]

Rysunek 8.26. Sumy iloczynów kolumn zawierających wartości przesunięte w czasie oraz kolumny zawierającej wartości rzeczywiste

Autokorelację danego przesunięcia w czasie można obliczyć, dzieląc znajdującą się w danej kolumnie sumę iloczynów przez sumę odchylen podniesionych do kwadratu znajdującą się w kolumnie C38.

W komórce D40 możesz obliczyć autokorelację przesunięcia o jeden miesiąc za pomocą formuły:

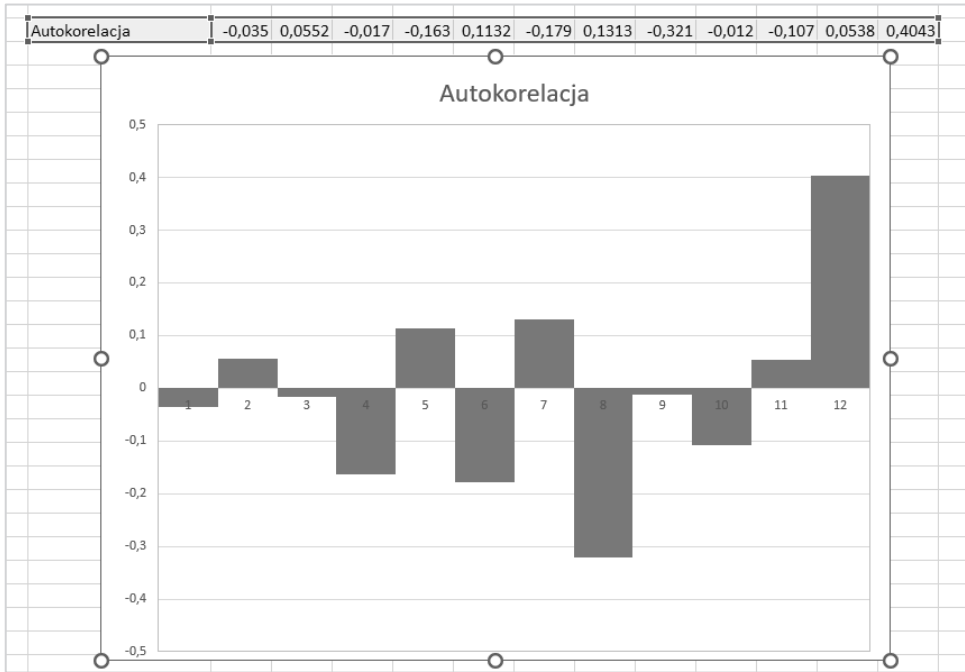
$$=D38/£C38$$

Przeciągając tę formułę w bok, określisz autokorelację przesunięcia o dowolną liczbę miesięcy.

Zaznaczając komórki C40:O40, możesz wstawić do arkusza wykres słupkowy pokazany na rysunku 8.27. Taki wykres słupkowy określa się mianem **korelogramu**. Pokazano na nim autokorelację przesunięć w czasie co jeden miesiąc aż do roku. (Uważam, że słowo *korelogram* brzmi naprawdę dobrze).

Które autokorelacje są dla Ciebie istotne? Czy możemy zmienić wykres z rysunku 8.27 tak, by pomógł odpowiedzieć na to pytanie?

Odpowiadając na pierwsze pytanie — w statystyce przyjęto twierdzenie, że warto interesować się autokorelacjami, których wartości są większe od wartości zwracanej przez operację arytmetyczną $2/\text{PIERWIASTEK}(\text{liczba punktów danych})$, czyli w Twoim przypadku $2/\text{PIERWIASTEK}(36) = 0,333$. Powinieneś zwrócić również uwagę na przypadki autokorelacji o wartościach mniejszych od $-0,333$.



Rysunek 8.27. Uzyskany przeze mnie korelogram

Możesz przyjrzeć się wykresowi i wyłączyć gołym okiem wartości autokorelacji, które są większe lub mniejsze od tych **wartości krytycznych**, ale zwykle podczas tworzenia prognoz wartości te nanosi się na wykres za pomocą linii przerywanych (dzięki temu obraz sytuacji staje się wyraźniejszy). Chciałbym pokazać Ci sposób na dodanie wspomnianych linii do wykresu.

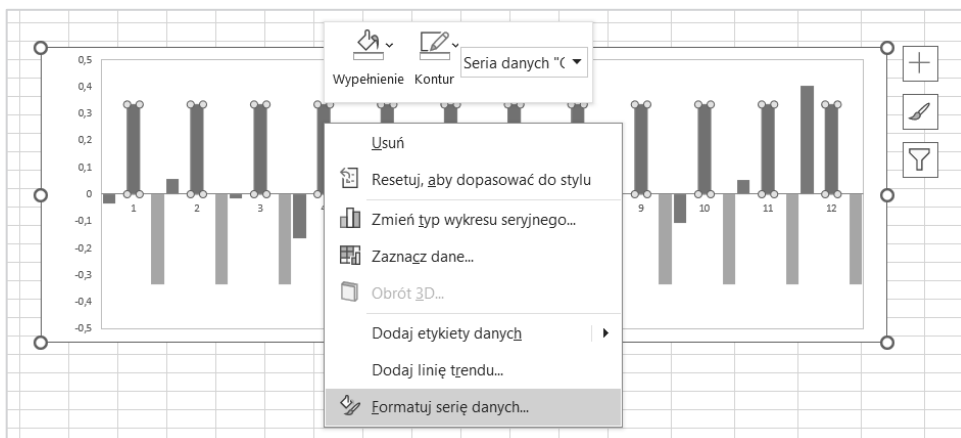
W komórce D41 wprowadź formułę =2/PIERWIASTEK(36), a następnie przeciągnij ją w bok aż do kolumny O. W komórce D42 wprowadź formułę, która będzie generowała identyczną, ale ujemną wartość. W tym celu możesz po prostu użyć takiego odwołania = -D41. Przeciągnij go w bok aż do kolumny O. W ten sposób utworzyłeś krytyczne punkty autokorelacji (rysunek 8.28).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Błąd prognozy jednego kroku	Odchylenia od średniej	1	2	3	4	5	6	7	8	9	10	11	12
27	26	16,13259637	12,556	-35,814	28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987	-16,713
28	27	-29,94280246	-33,519	12,556	-35,814	28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387	-12,987
29	28	2,398387897	-1,178	-33,519	12,556	-35,814	28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067	-22,387
30	29	10,33952568	6,763	-1,178	-33,519	12,556	-35,814	28,012	9,592	4,073	13,499	1,261	-8,734	-19,924	31,067
31	30	-15,3152409	-18,892	6,763	-1,178	-33,519	12,556	-35,814	28,012	9,592	4,073	13,499	1,261	-8,734	-19,924
32	31	11,47512572	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	28,012	9,592	4,073	13,499	1,261	-8,734
33	32	-0,793754823	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	28,012	9,592	4,073	13,499	1,261
34	33	6,051575658	2,475	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	28,012	9,592	4,073	13,499
35	34	2,173284481	-1,403	2,475	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	28,012	9,592	4,073
36	35	44,7750903	41,199	-1,403	2,475	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	28,012	9,592
37	36	51,73048169	48,154	41,199	-1,403	2,475	-4,370	7,899	-18,892	6,763	-1,178	-33,519	12,556	-35,814	28,012
38	37	3,576404834	13636,81665	-474,1	753,4	-229,3	-2223,6	1544,3	-2442,5	1790,5	-4382,6	-167,7	-1460,4	733,1	5512,8
39															
40			Autokorelacja	-0,03476	0,055245	-0,01682	-0,16306	0,113244	-0,17911	0,131301	-0,32138	-0,0123	-0,10709	0,053756	0,404259
41			Górna wartość progowa	0,333333	0,333333	0,333333	0,333333	0,333333	0,333333	0,333333	0,333333	0,333333	0,333333	0,333333	0,333333
42			Dolna wartość progowa	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333	-0,333333

Rysunek 8.28. Krytyczne punkty autokorelacji

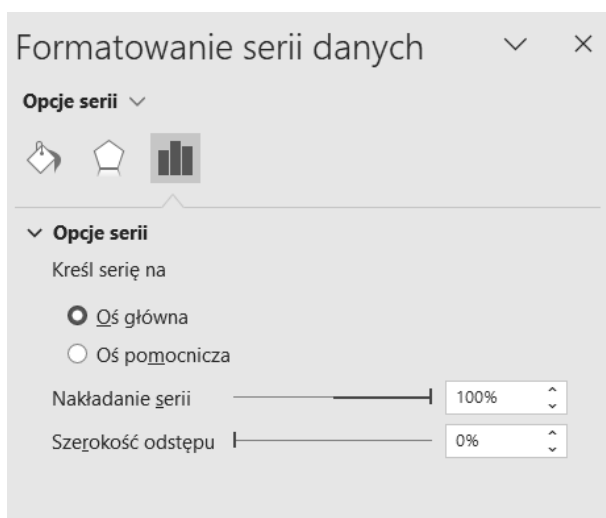
Zaznacz komórki C41:042. Skopiuj zaznaczony zakres za pomocą skrótu klawiszowego *Ctrl+C* (na komputerze Mac: *Cmd+C*). Następnie kliknij wykres i naciśnij *Ctrl+V* (na komputerze Mac: *Cmd+V*), by wkleić skopiowane przed chwilą dane. Po wykonaniu tej operacji na wykresie pojawią się dwa kolejne zestawy słupków. To nie jest jeszcze gotowy wykres, ale stanowi dobrą podstawę do dalszej pracy.

Kliknij prawym przyciskiem myszy każdą z tych nowych serii słupków i wybierz opcję *Formatuj serię danych...* (rysunek 8.29).



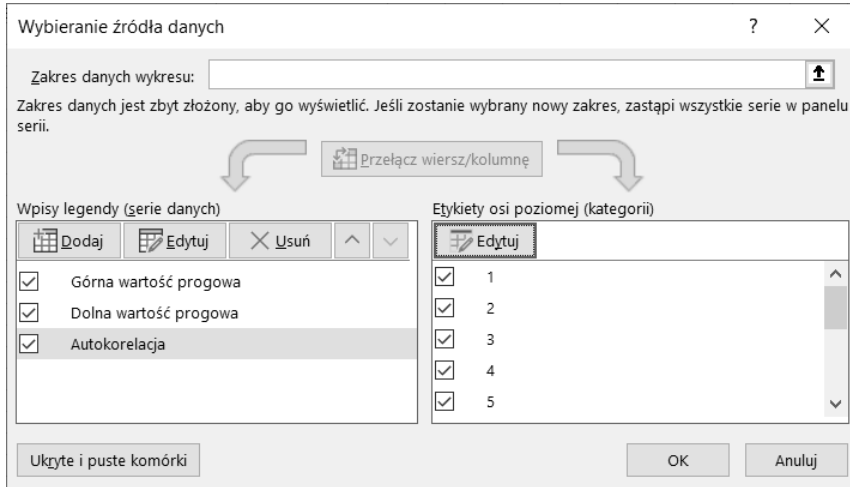
Rysunek 8.29. Kliknij wklejone serie danych i popraw ich wygląd za pomocą formatowania serii danych

W panelu *Formatowanie serii danych* (rysunek 8.30) *Nakładanie serii* ustaw na 100%, a *Szerokość odstępu* na 0%.



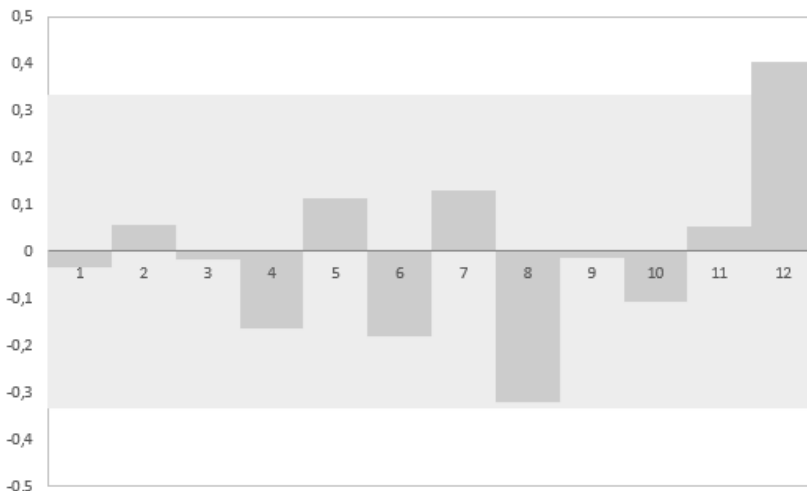
Rysunek 8.30. Panel Formatowanie serii danych

Następnie kliknij wykres prawym przyciskiem myszy i kliknij *Zaznacz dane...* To spowoduje otwarcie nowego okna, w którym za pomocą strzałek przesuń serię Autokorelacja na sam dół, tak by wyświetlała się przed pozostałymi seriami. Po wykonaniu tych operacji naciśnij *OK* (rysunek 8.31).



Rysunek 8.31. Okno Wybieranie źródła danych. Seria danych Autokorelacja została przesunięta na dół listy, dzięki czemu będzie wyświetlana na pierwszej warstwie wykresu

Teraz to już tylko kwestia formatowania. Lubię podkreślać główne analizowane przeze mnie wartości i wyróżniać je na tle ogólnego kontekstu (czyli wartości, z którymi porównywałeś swoje główne wartości). Spójrz na rysunek 8.32, gdzie pokazano wykres sformatowany tak, by było łatwiej zrozumieć przedstawione na nim dane.



Rysunek 8.32. Wyróżnienie wartości krytycznych

Co widzisz na tym wykresie?

Poza wartości krytyczne wykracza tylko jedna autokorelacja. Jest to autokorelacja dla przesunięcia o 12 miesięcy.

Błąd przesunięty o rok jest skorelowany z samym sobą, co wskazuje na istnienie dwunastomiesięcznego cyklu. Chyba nie jest to nic dziwnego. Jeżeli przyjrzesz się wykresowi popytu znajdującemu się w arkuszu *Szereg czasowy*, to zauważysz wyraźny wzrost popytu w okolicy świąt Bożego Narodzenia oraz wyraźny spadek popytu w okolicy kwietnia i maja.

W związku z tym potrzebujesz techniki prognozowania biorącej pod uwagę sezonowość popytu. Istnieje specjalna wersja techniki wygładzania wykładniczego, która potrafi uwzględnić sezonowe wahania popytu.

Wielokrotne wygładzanie wykładnicze — model Holta-Wintersa

Model Holta-Wintersa jest logicznym rozszerzeniem opisanego wcześniej modelu Holta. Uwzględnia on poziom, trend i potrzebę dostosowania popytu do sezonowych wahań. Fluktuacje te mogą dotyczyć dowolnego okresu i nie muszą zachodzić co 12 miesięcy, jak w opisywanym przykładzie. Na przykład w firmie wysyłającej automatyczne wiadomości e-mail do klientów wzrost popytu może występować w każdy czwartek (ludziom chyba wydaje się, że czwartek to dobry dzień na wysyłanie wiadomości marketingowych). Model Holta-Wintersa umożliwia uwzględnienie tego siedmiodniowego cyklu.

W większości sytuacji nie można po prostu dodać stałej wartości sezonowego popytu do ustalonej wcześniej prognozy (lub odjąć). Jeżeli Twoja firma rozwijała się, zaczynając od sprzedaży 200 mieczy, a obecnie sprzedaje co miesiąc 2000 mieczy, to popytu w święta Bożego Narodzenia nie można określić, dodając do obu tych wartości 20 mieczy. Wahania sezonowe zwykle ustala się za pomocą mnożników. Zamiast dodawać 20 mieczy, być może trzeba *pomnożyć* prognozę przez 120%. W związku z tym model opisany w tym podrozdziale można określić również mianem **modelu mnożnika Holta-Wintersa**. Oto ogólny wzór pozwalający na prognozowanie popytu:

*zapotrzebowanie w czasie $t = (\text{poziom} + t \cdot \text{trend}) \cdot \text{wyrównanie sezonowe}$
właściwe dla czasu t · wszystkie pozostałe nieregularne wyrównania, których nie można przewidzieć*

W tym wzorze ponownie występują struktury takie jak poziom i trend, ale tym razem popyt uwzględnia wyrównanie sezonowe. Nie możesz przewidzieć wszystkich czynników wpływających na popyt. Na popyt wpływają również zmienne losowe.

Model Holta-Wintersa określany jest też mianem **potrójnego wygładzania wykładniczego**. Jak się zapewne domyślasz, wynika to z obecności trzech parametrów wygładzających. Model wciąż korzysta z parametrów *alfa* i *gamma*, ale zawiera również dodatkowy parametr uwzględniający sezonowość — *delta*.

Trzy równania korygujące błędy są bardziej rozbudowane od równań zaprezentowanych wcześniej, ale wciąż da się je zrozumieć.

Zanim zacznę opis modelu Holta-Wintersa, chciałbym zaznaczyć, że dotąd prognozy na kolejne miesiące opierałeś na poziomach i trendach z poprzedniego okresu, a teraz nie będziesz się opierać na danych poprzedniego okresu — będziesz korzystać z poprzedniej prognozy czynnika korygującego dla danego punktu cyklu. Innymi słowy: tym razem nie będziesz opierać prognozy na jednym okresie, a na danych z 12 okresów.

W związku z tym, jeżeli jesteś w 36. miesiącu i chcesz określić prognozę na miesiąc 39. (3 miesiące do przodu), prognoza ta będzie wyglądała następująco:

$$\text{prognoza na 39. miesiąc} = (\text{poziom}_{36} + 3 \cdot \text{trend}_{36}) \cdot \text{sezonowość}_{27}$$

Tak, to nie błąd. Chodzi o *sezonowość*₂₇. To ostatnia wartość korekty sezonowej dla marca. Nie można tu skorzystać z korekty *sezonowość*₃₆, ponieważ dotyczy ona grudnia.

Wiesz już, jak działa mechanizm przewidywania przyszłości, a więc czas zająć się równaniami aktualizującymi parametry używane podczas tworzenia prognoz. Zacznijmy od poziomu. Potrzebujesz wartości wstępnych *poziom*₀ i *trend*₀, a także 12 początkowych współczynników sezonowości — *sezonowość*₋₁₁ – *sezonowość*₀.

Na przykład równanie aktualizujące *poziom*₁ korzysta z początkowego współczynnika sezonowości dla stycznia:

$$\text{poziom}_1 = \text{poziom}_0 + \text{trend}_0 + \text{alfa} \cdot (\text{popyt}_1 - (\text{poziom}_0 + \text{trend}_0) \cdot \text{sezonowość}_{-11}) / \text{sezonowość}_{-11}$$

Wzór ten zawiera wiele znanych Ci elementów. Bieżący poziom jest sumą poprzedniego poziomu, poprzedniego trendu (podobnie jak w przypadku podwójnego wygładzania wykładniczego) oraz iloczynu współczynnika *alfa* i błędu prognozy kolejnego kroku (*popyt*₁ – (*poziom*₀ + *trend*₀) · *sezonowość*₋₁₁). Błąd jest korygowany poprzez podzielenie go przez współczynnik sezonowości *sezonowość*₋₁₁.

Równanie określające kolejny poziom ma postać:

$$\text{poziom}_2 = \text{poziom}_1 + \text{trend}_1 + \text{alfa} \cdot (\text{popyt}_2 - (\text{poziom}_1 + \text{trend}_1) \cdot \text{sezonowość}_{-10}) / \text{sezonowość}_{-10}$$

Jak się zapewne domyślasz, ogólny wzór na poziom ma postać:

$$\text{poziom}_{\text{bieżący okres}} = \text{poziom}_{\text{poprzedni okres}} + \text{trend}_{\text{poprzedni okres}} + \text{alfa} \cdot (\text{popyt}_{\text{bieżący okres}} - (\text{poziom}_{\text{poprzedni okres}} + \text{trend}_{\text{poprzedni okres}}) \cdot \text{sezonowość}_{\text{ostatni analogiczny okres}}) / \text{sezonowość}_{\text{ostatni analogiczny okres}}$$

Trend jest aktualizowany w zależności od poziomu (dokładnie tak samo jak w przypadku podwójnego wygładzania wykładniczego):

$$\text{trend}_{\text{bieżący okres}} = \text{trend}_{\text{poprzedni okres}} + \text{gamma} \cdot \text{alfa} \cdot (\text{popyt}_{\text{bieżący okres}} - (\text{poziom}_{\text{poprzedni okres}} + \text{trend}_{\text{poprzedni okres}}) \cdot \text{sezonowość}_{\text{ostatni analogiczny okres}}) / \text{sezonowość}_{\text{ostatni analogiczny okres}}$$

Podobnie jak w przypadku podwójnego wygładzania wykładniczego bieżący trend jest sumą poprzedniego trendu i iloczynu współczynnika *gamma* oraz błędu wprowadzonego do równania aktualizującego poziom.

Równanie aktualizujące czynnik sezonowości przypomina nieco równanie aktualizujące trend, ale czynnik sezonowości ostatniego analogicznego okresu jest mnożony przez współczynnik *delta* oraz błąd, który został pominięty w równaniach aktualizujących poziom oraz trend:

$$\text{sezonowość}_{\text{bieżący okres}} = \text{sezonowość}_{\text{ostatni analogiczny okres}} + \text{delta} \cdot (1 - \text{alfa}) \cdot (\text{popyt}_{\text{bieżący okres}} - (\text{poziom}_{\text{poprzedni okres}} + \text{trend}_{\text{poprzedni okres}}) \cdot \text{sezonowość}_{\text{ostatni analogiczny okres}}) / (\text{poziom}_{\text{poprzedni okres}} + \text{trend}_{\text{poprzedni okres}})$$

W tym przypadku aktualizujesz czynnik sezonowości, korzystając z analogicznego czynnika sprzed 12 miesięcy, ale uwzględniasz również iloczyn współczynnika *delta* oraz błędu pozostałego po aktualizacji poziomu. Zauważ, że zamiast uwzględniać czynnik sezonowości, dzielisz przez poprzedni poziom oraz poprzedni trend. Dzięki „regulacji poziomu i trendu” błędu kolejnego kroku błąd ten jest mnożony przez tę samą wartość co czynniki sezonowości.

Określanie początkowych wartości poziomu, trendu i sezonowości

Określanie początkowych wartości prostego wygładzania wykładniczego oraz podwójnego wygładzania wykładniczego było bardzo proste. Teraz musisz określić trend i sezonowość na podstawie szeregu czasowego, co oznacza, że określanie początkowych wartości dla tej prognozy (1 poziom, 1 trend i 12 czynników wyrównania sezonowego) jest dość trudne. Istnieją proste sposoby na zdefiniowanie tych wartości, ale są one nieprawidłowe. Pokażę Ci dobry sposób inicjalizacji modelu Holta-Wintersa, który zakłada, że dysponujesz danymi przynajmniej dwóch sezonowych cykli. W tym przypadku dysponujesz danymi trzech cykli.

Oto lista czynności, które w związku z tym wykonasz:

1. Wygładzisz dane opisujące popyt w przyszłości za pomocą średniej ruchomej 2×12 .
2. Porównasz wygładzoną wersję szeregu czasowego z jego oryginalną wersją w celu określenia sezonowości.

3. Skorzystasz z oszacowanych początkowych wartości sezonowości dla danych opisujących popyt w przeszłości.
4. Oszacujesz poziom i trend za pomocą linii trendu danych, z których usunięto sezonowość.

Zacznij od utworzenia nowego arkusza o nazwie **Holt-Winters inicjalizacja** i wklejenia do jego dwóch pierwszych kolumn danych szeregu czasowego. Po wklejeniu danych zmień nazwę tabeli na **HoltWintersInicjalizacja.Dane**. Następnie wstaw cztery puste wiersze nad tą tabelą, w których umieścisz dwie zmienne: liczbę miesięcy w cyklu oraz całkowitą liczbę miesięcy w danych. Dzięki temu będziesz mógł tworzyć dynamiczne formuły (rysunek 8.33).

	A	B	C	D	E	F	G
1	Długość cyklu	12					
2	Liczba miesięcy	36	=ILE.NIEPUSTYCH(HoltWintersInicjalizacja.Dane[t])				
3							
4							
5	t	Popyt					
6	1	165					
7	2	171					
8	3	147					
9	4	143					
10	5	164					
11	6	160					
12	7	152					
13	8	150					
14	9	159					
15	10	169					
16	11	173					
17	12	203					
18	13	169					

Rysunek 8.33. Przygotowanie danych dla algorytmu Holta-Wintera

Teraz musisz wygładzić część danych szeregu czasowego za pomocą średniej ruchomej. Sezonowość ma formę dwunastomiesięcznych cykli, a więc warto przetworzyć dane za pomocą dwunastomiesięcznej ruchomej średniej. Śmiało więc możesz wpisać liczbę 12 do komórki B1.

Co mam na myśli, pisząc o dwunastomiesięcznej ruchomej średniej?

Obliczając średnią ruchomą, należy wziąć pod uwagę zapotrzebowanie w danym miesiącu, a także zapotrzebowanie sąsiednich miesięcy w obu kierunkach, a następnie policzyć średnią tych wartości. Spowoduje to wygładzenie wszelkich gwałtownych zmian serii danych.

Obliczając dwunastomiesięczną ruchomą średnią, napotykamy pewien problem. Dwanaście to liczba parzysta. Czy wygładzając popyt 7. miesiąca, powinieneś obliczyć średnią miesięcy od 1. do 12., czy od 2. do 13.? Tak czy inaczej, 7. miesiąc nie leży w środku. Nie mamy elementu środkowego!

W celu rozwiązania tego problemu będziesz wygładzać popyt metodą „ruchomej średniej 2×12” — będziesz obliczać średnią wartość obu możliwości: miesięcy 1. – 12. i 2. – 13. Tego typu rozwiązanie należy stosować w przypadku dowolnej innej parzystej liczby okresów w cyklu. W przypadku pracy z cyklem składającym się z nieparzystej liczby okresów element „2×” ruchomej średniej jest zbędny (możliwe jest korzystanie z prostej ruchomej średniej).

Zauważ, że to rozwiązanie nie może być zastosowane w przypadku 6 pierwszych i 6 ostatnich miesięcy Twojego zbioru danych. Nie ma 6 miesięcy danych sąsiadujących po obu stronach tych zakresów. Możesz wygładzić tylko środkowe miesiące zbioru danych (miesiące 7. – 30.). Teraz już wiesz, dlaczego do wygładzenia danych potrzebujesz zbioru z przynajmniej dwóch lat.

Do tabeli dodaj nową kolumnę zatytułowaną **Wygładzanie**.

Następnie wpisz tę formułę w komórce C6, by wypełnić nowo dodaną kolumnę:

```
=LET(
    PołowaCyklu; $B$1/2;
    MożliwośćWygładzania; ORAZ([@t] > PołowaCyklu; [@t] < $B$2-PołowaCyklu + 1);
    JEŻELI(MożliwośćWygładzania;
        (ŚREDNIA(PRZESUNIĘCIE([@Popyt]; -PołowaCyklu; 0; $B$1)) +
         ŚREDNIA(PRZESUNIĘCIE([@Popyt]; -PołowaCyklu + 1; 0; $B$1)) *
         CZY.PARZYSTE($B$1))/2;
        BRAK()
    )
)
```

Ta formuła może wydawać się przydługawa, ale uwierz mi, nie jest tak źle. Najpierw za pomocą funkcji LET deklarujesz zmienne. Pierwsza z nich, PołowaCyklu, dzieli długość cyklu (B1) na dwa. Natomiast MożliwośćWygładzania jest zmienną logiczną, czyli przechowuje wartość PRAWDA albo FAŁSZ i informuje o tym, czy w danym momencie można zastosować wygładzanie, do czego potrzebne są dane z 6 miesięcy (przed danym miesiącem i po nim). Nie będziesz wygładzać danych w punkcie $t = 3$ lub $t = 35$, gdyż przed i po nie ma wystarczającej liczby miesięcy, by przygotować prognozę.

Następnie formuła sprawdza, czy dany miesiąc może być wygładzony, a jeśli tak, wykonywana jest ta formuła:

```
ŚREDNIA(PRZESUNIĘCIE([@Popyt]; -PołowaCyklu; 0; $B$1)) +
ŚREDNIA(PRZESUNIĘCIE([@Popyt]; -PołowaCyklu + 1; 0; $B$1)) *
CZY.PARZYSTE($B$1))/2
```

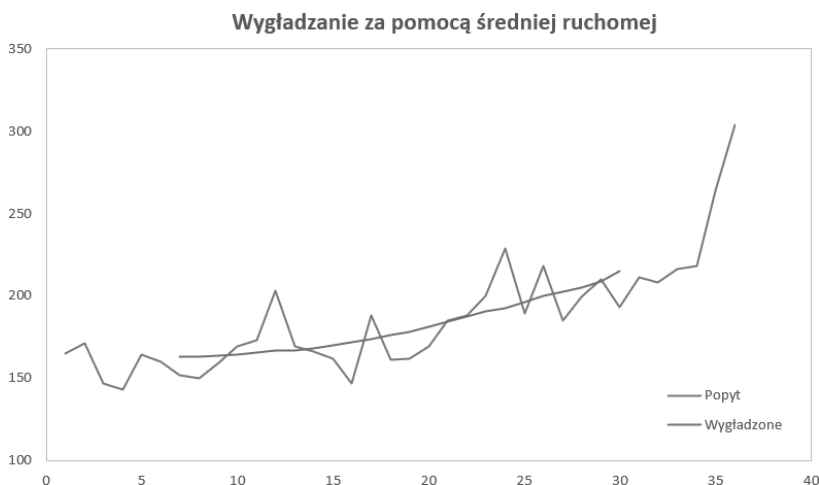
To tak naprawdę nasza funkcja wygładzająca za pomocą 2 · 12 średnich ruchomych. Funkcja PRZESUNIĘCIE używa bieżącego okresu jako punktu zaczepienia. Potem sprawdza dane z połowy cyklu wstecz i zwiększa ten zakres o całą długość cyklu

znajdującą się w komórce B1 (w tym przypadku 12). Pamiętaj, że jeśli średnia ruchoma jest parzysta, to brakuje punktu środkowego. Dlatego dla $t = 7$ uśrednisz popyt dla miesięcy od 1 – 12 i 2 – 13. Ta formuła jest dostosowana do takiego rozwiązania dzięki mnożeniu wartości zwracanej przez drugą funkcję ŚREDNIA przez wartość zwracaną przez funkcję CZY.PARZYSTE. Jeśli cykl składa się z nieparzystej liczby miesięcy, ten iloczyn będzie równy zero. Wyniki działania tej formuły pokazano na rysunku 8.34.

	A	B	C
5	t	Popyt	Wyglądanie
6	1	165	#N/D
7	2	171	#N/D
8	3	147	#N/D
9	4	143	#N/D
10	5	164	#N/D
11	6	160	#N/D
12	7	152	163,167
13	8	150	163,125
14	9	159	163,542
15	10	169	164,333
16	11	173	165,500
17	12	203	166,542
18	13	169	167,000
19	14	166	168,208

Rysunek 8.34. Funkcja wygładzająca biorąca pod uwagę to, że średnia ruchoma może być liczbą parzystą

Teraz możesz wykreślić dane popytu, by zobaczyć, jak wyglądają po wygładzeniu (rysunek 8.35).



Rysunek 8.35. Wygładzone dane popytu

Teraz w kolumnie D możesz podzielić oryginalną wartość popytu przez jego wygładzoną wartość i uzyskać w ten sposób czynnik wyrównania sezonowego. Do tabeli dodaj nową kolumnę o nazwie **Czynnik wyrównania sezonowego**. W dowolnej komórce nowej kolumny umieść taką formułę:

$$=[@Popyt]/[@Wygładzanie]$$

Formuła automatycznie wypełni całą kolumnę. Zauważ, że w miesiącach 12. i 24. (grudzień) popyt jest większy o ok. 20%, a na wiosnę jest nieco niższy niż w pozostałych miesiącach.

Technika wygładzania pozwoliła otrzymać dwupunktowe szacunki dla każdego czynnika sezonowości. Oblicz wartość średnią tych dwóch punktów. Uzyskana w ten sposób wartość może być używana jako początkowy czynnik sezonowości w modelu Holta-Wintersa.

Na przykład w komórce D masz trzy punkty związane ze styczniem (komórki D6, D18 i D30). Jednak tylko D18 i D30 zawierają rzeczywiste dane, gdyż wartość komórki D6 to #N/D. Dlatego należy obliczyć średnią komórek punktów dla stycznia z rzeczywistymi danymi i pominąć #N/D. W tym celu najłatwiej utworzyć nową tabelę Excela, taką jak na rysunku 8.36. W swoim arkuszu nazwałem tę tabelę **HoltWintersInicjacja.CzynnikiSezonowości**.

		A	B	C	D	E	F	G	H
1	Długość cyklu		12						
2	Liczba miesięcy		36						
3									
4									
5	t	Popyt	Wygładzanie	Czynnik wyrównania sezonowego	Początkowy czynnik sezonowy		ID miesiąca	Początkowe czynniki sezonowe	
6	1	165	#N/D	#N/D	0,988		1	0,988	
7	2	171	#N/D	#N/D	1,039		2	1,039	
8	3	147	#N/D	#N/D	0,933		3	0,933	
9	4	143	#N/D	#N/D	0,913		4	0,913	
10	5	164	#N/D	#N/D	1,043		5	1,043	
11	6	160	#N/D	#N/D	0,906		6	0,906	
12	7	152	163,167		0,932		7	0,921	
13	8	150	163,125		0,920		8	0,927	
14	9	159	163,542		0,972		9	0,988	
15	10	169	164,333		1,028	1,016	10	1,016	
16	11	173	165,500		1,045	1,048	11	1,048	
17	12	203	166,542		1,219	1,204	12	1,204	

Rysunek 8.36. Tabela z początkowymi czynnikami sezonowymi

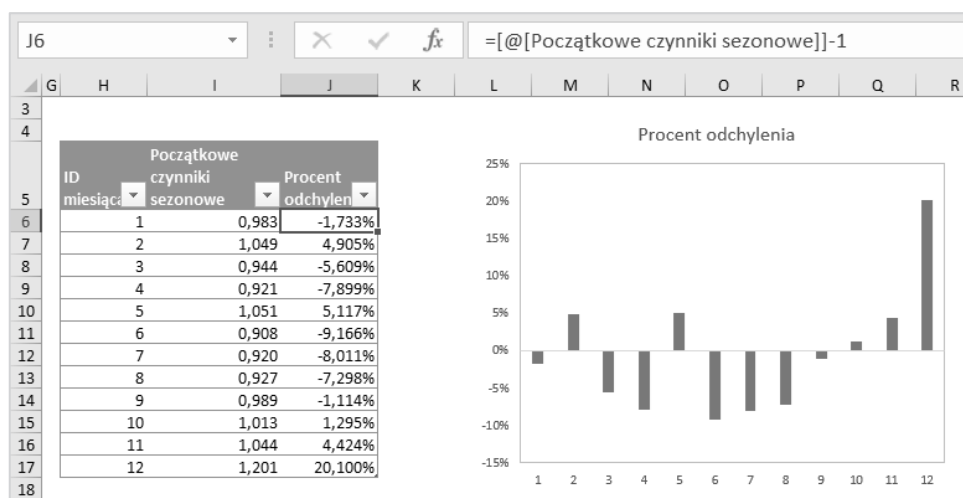
Najpierw jest kolumna z 12 czynnikami sezonowymi, w której wartości od 1 do 12 są wpisane na sztywno. W kolumnie G znajdują się początkowe czynniki sezonowe obliczane za pomocą tej formuły:

```
=ŚREDNIA(FILTRUJ(
    HoltWintersInicjalizacja.Dane[Czynnik wyrównania sezonowego];
    CZY.LICZBA(HoltWintersInicjalizacja.Dane[Czynnik wyrównania
    sezonowego]) * [@[ID miesiąca]] = MOD(HoltWintersInicjalizacja.Dane[t]
    - 1;12) + 1
))
```

Na początku przyjrzyj się funkcji FILTRUJ. Funkcja FILTRUJ umożliwia wyfiltrowanie określonych elementów zwracanych w postaci dynamicznej tabeli (doskonale!). Pierwszy parametr funkcji FILTRUJ to zakres, dla którego chcesz zastosować filtr, i w tym przypadku jest to kolumna Czynnik wyrównania sezonowego.

Następnie podane są warunki filtra. Tutaj mamy dwa warunki, z czego pierwszy z nich sprawdza, czy dany element musi być liczbą. Zatem dopóki nie ma wartości #N/D, jest to liczba. Drugi warunek sprawdza miesiąc. Na przykład t = 13 z pierwszej tabeli odnosi się do stycznia. Możemy odjąć 1 od wartości t (by uwzględnić fakt, że lista nie zaczyna się od 0), wykonać działanie modulo, by sprawdzić bieżący miesiąc, a na koniec dodać 1 (ponownie z uwagi na fakt, że lista nie zaczyna się od 0).

Po obliczeniu 12 czynników sezonowych możesz napisać formułę odejmującą jeden od każdej z tych wartości w celu obliczenia odchylenia. Sformatuj komórki kolumny odchylenia, aby znajdujące się w nich dane były wyświetlane jako wartości procentowe. W ten sposób uzyskasz wartości pokazujące, jak bardzo czynniki wpływają na popyt w każdym miesiącu. Możesz utworzyć wykres słupkowy prezentujący te odchylenia (rysunek 8.37).



Rysunek 8.37. Wykres oszacowanych odchyłeń sezonowych

Dysponujesz już początkowymi wartościami wyrównania sezonowego, a więc możesz ich użyć w celu *pozbycia się sezonowości* z danych szeregu czasowego. Po usunięciu sezonowości z całego szeregu danych będziesz mógł wytyczyć jego linię trendu, a następnie wyznaczyć jej nachylenie i punkt przecięcia z osią y (początkowy poziom i trend).

Zacznij od dodania nowej kolumny z prawej strony kolumny D. Nazwij ją **Początkowy czynnik sezonowy x**. Sprowadza się to do ponownego wklejenia pierwotnych czynników sezonowych do tej nowej kolumny dla każdego miesiąca. Zatem w nowej kolumnie (powinna być to kolumna E tak jak na rysunku 8.38) użyj takiej formuły:

```
=INDEKS(
    HoltWintersInicjalizacja.CzynnikiSezonowości[Początkowe czynniki sezonowe];
    MOD([@t] - 1;12) + 1
)
```

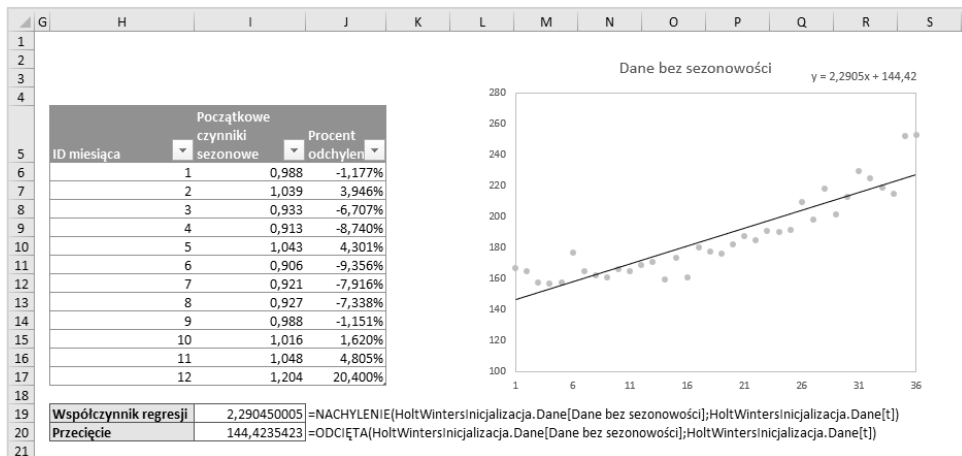
t	Popyt	Wyglądani	Czynnik wyrównania sezonowego	Początkowy czynnik sezonowy x	ID miesiąca	Początkowe czynniki sezonowe	Procent odchylen
1	165	#N/D	#N/D	0,988	1	0,988	-1,177%
2	171	#N/D	#N/D	1,039	2	1,039	3,946%
3	147	#N/D	#N/D	0,933	3	0,933	-6,707%
4	143	#N/D	#N/D	0,913	4	0,913	-8,740%
5	164	#N/D	#N/D	1,043	5	1,043	4,301%
6	160	#N/D	#N/D	0,906	6	0,906	-9,356%
7	152	163,167	0,932	0,921	7	0,921	-7,916%
8	150	163,125	0,920	0,927	8	0,927	-7,338%
9	159	163,542	0,972	0,988	9	0,988	-1,151%
10	169	164,333	1,028	1,016	10	1,016	1,620%
11	173	165,500	1,045	1,048	11	1,048	4,805%
12	203	166,542	1,219	1,204	12	1,204	20,400%
13	169	167,000	1,012	0,988			
14	166	168,208	0,987	1,039			
15	162	170,083	0,952	0,933			
16	147	171,958	0,855	0,913			
17	188	173,875	1,081	1,043			
18	161	176,083	0,914	0,906			

Rysunek 8.38. Powtórzenie początkowych czynników sezonowości w celu usunięcia sezonowości z każdego miesiąca w zestawie danych

Formuła $\text{MOD}([\text{@t}] - 1; 12) + 1$ umożliwi zamianę wartości t na dowolny numer miesiąca z zakresu od 1 do 12. Ta informacja posłuży do wyszukania początkowych czynników sezonowych powiązanych z danym miesiącem. Porównaj swój arkusz z rysunkiem 8.38.

Wstaw kolejną kolumnę po prawej stronie kolumny E. Nazwij ją **Dane bez sezonowości**. Teraz podziel pierwotną serię danych z kolumny B przez czynniki sezonowe z kolumny E, by z zestawu danych usunąć szacowaną sezonowość.

Teraz, podobnie jak w przypadku poprzednich arkuszy, narysuj wykres punktowy danych umieszczonych w kolumnie F, a następnie nanieś na niego linię trendu. Ponadto wyświetl na wykresie równanie linii trendu. W ten sposób uzyskasz początkową wartość trendu (miesięczny wzrost sprzedaży o 2,29 miecza) i początkowy poziom (144,42). Wykres ten pokazano na rysunku 8.39. W swoim arkuszu, w komórkach H19: I20, umieściłem również obliczenia dla współczynnika regresji i przecięcia. Zamiast funkcji REGLINP użyłem funkcji NACHYLENIE i ODCIĘTA, ponieważ nie chcę się mierzyć ze wszystkimi parametrami funkcji REGLINP.



Rysunek 8.39. Wartości początkowego poziomu i trendu oszacowane na podstawie linii trendu danych szeregu czasowego po usunięciu czynników sezonowych

Tworzenie prognozy

Dysponujesz już wartościami początkowymi wszystkich parametrów. W związku z tym utwórz nowy arkusz i nadaj mu nazwę **Holt-Winters prognoza**. Do 4. wiersza arkusza wklej dane szeregu czasowego, tak jak robiłeś to w przypadku dwóch poprzednich technik prognozowania. Tabelę w tym arkuszu nazwij **HoltWinters Prognoza .Dane**.

Obok danych szeregu czasowego — w kolumnach C, D i E — umieścisz kolejno wartości poziomu, trendu i wyrównania sezonowego. Aby to zrobić, będziesz musiał wstawić puste wiersze w miejscu wiersza 5. W poprzednich arkuszach wstawiałeś jeden pusty wiersz, a tym razem musisz wstawić 12 pustych wierszy (wiersze o numerach 5 – 16 muszą być puste). Po wykonaniu tych operacji umieść etykiety nowych wierszy (od -11 do 0) w kolumnie A. W komórce A1 umieść etykietę **Liczba miesięcy**, a w komórce A2 liczbę miesięcy, za które masz dane. Teraz Twój arkusz powinien wyglądać tak jak na rysunku 8.40.

	A	B	C	D	E
1	Liczba miesięcy				
2	36				
3					
4	t	Rzeczywisty popyt	Pozio	Trend	Wyrównanie sezonowe
5	-11				
6	-10				
7	-9				
8	-8				
9	-7				
10	-6				
11	-5				
12	-4				
13	-3				
14	-2				
15	-1				
16	0				
17	1	165			
18	2	171			

Rysunek 8.40. Arkusz zawierający wszystkie wartości początkowe modelu Holta-Wintersa

Najlepiej tworzyć tę tabelę krok po kroku, tak jak robiłeś to wcześniej. Dlatego nastaw się na kilkukrotne wprowadzanie zmian w formułach.

Zacznij od wyrównania sezonowego. W tym przypadku, gdy $t = [-11, 0]$, wyrównanie sezonowe ma być równe początkowym wartościom z poprzedniego arkusza. Zatem w tej kolumnie umieść taką formułę:

```
=WARUNKI(
  [@t] < 1; INDEKS(
    HoltWintersInicjalizacja.CzynnikiSezonowości[Początkowe czynniki
    sezonowe];
    [@t] + 'Holt-Winters inicjalizacja'!$B$1
  )
)
```

Ta formuła działa w dynamiczny sposób. Jeśli $t < 1$, wtedy wiadomo, że mieści się w zakresie od -11 do 0 . W rezultacie używana jest bieżąca wartość t , powiedzmy $t = -11$, a następnie zwiększamy ją o całkowitą liczbę miesięcy z arkusza *Holt-Winters inicjalizacja* (czyli w tym przypadku 12). Wynikiem dodawania $12 + -11$ jest 1. Uzyskany wynik posłuży jako indeks dla naszej początkowej tabeli z czynnikami sezonowymi. Jak się przekonasz, to działa i pierwsze 12 elementów zostanie wyszukane.

Następnie ustaw poziom i trend, zdefiniowane przez przecięcie i współczynnik regresji z poprzedniego arkusza. Te wartości umieść w wierszu dla $t = 0$, gdyż będą stanowić wartości początkowe. W kolumnach Poziom i Trend umieść następujące formuły:

```
C5 (Poziom): =WARUNKI(
    [0t] = 0; 'Holt-Winters inicjalizacja'!$I$20
)
D5 (Trend): =WARUNKI(
    [0t] = 0; 'Holt-Winters inicjalizacja'!$I$19
)
```

Po wykonaniu tych wszystkich kroków Twój arkusz będzie wyglądał tak jak na rysunku 8.41. Zwróć szczególną uwagę na wiersz 16., gdyż tam znajdują się wartości początkowe.

	A	B	C	D	E
1	Liczba miesięcy				
2	36				
3					
4	t	Rzeczywisty popyt	Pozio	Trend	Wyrównanie sezonowe
5	-11		#N/D	#N/D	0,988
6	-10		#N/D	#N/D	1,039
7	-9		#N/D	#N/D	0,933
8	-8		#N/D	#N/D	0,913
9	-7		#N/D	#N/D	1,043
10	-6		#N/D	#N/D	0,906
11	-5		#N/D	#N/D	0,921
12	-4		#N/D	#N/D	0,927
13	-3		#N/D	#N/D	0,988
14	-2		#N/D	#N/D	1,016
15	-1		#N/D	#N/D	1,048
16	0		144,4235	2,29045	1,204
17	1	165	152,9518	3,73103	0,988
18	2	171	159,0934	4,287731	1,039

Rysunek 8.41. Przypisanie wartości początkowych trendu i poziomu do prognozy sezonowej

W kolumnach F i G dodaj takie nagłówki: **Prognoza jednego kroku** i **Błąd prognozy**.

Jeśli chodzi o prognozę jednego kroku dla miesiąca 1., to jest ona równa poprzedniemu poziomowi z komórki C16 powiększonemu o poprzedni trend z komórki D16. Jednak obie wartości są wyrównane za pomocą odpowiedniej szacunkowej sezonowości dla stycznia (komórka E5 znajdująca się 12 wierszy wyżej). Oczywiście to wszystko musisz zawrzeć w formule. Dlatego w kolumnie F wpisz taką formułę:

310 Mistrz analizy danych

```
=JEŻELI ([@t]>0;
SUMA(
    PRZESUNIĘCIE(HoltWintersPrognoza.Dane[@Poziom]:[Trend]]; -1; 0)) *
    PRZESUNIĘCIE([@Wyrównanie sezonowe]]; - 'Holt-Winters
    inicjalizacja'!$B$1; 0);
BRAK()
)
```

Tak jak poprzednio, formuła sprawdza, czy t jest większe od 0, gdyż prognoza jednego miesiąca ma się zaczynać od punktu $t = 1$. Potem wartości poziomu i trendu są sumowane i mnożone przez wartość wyrównania sezonowego (poprzedzające 12 miesięcy podane w komórce 'Holt-Winters inicjalizacja'!\$B\$1).

Teraz możesz obliczyć błąd prognozy za pomocą tej formuły:

```
=[@Rzeczywisty popyt]-[@Prognoza jednego kroku]
```

Teraz możesz rozpocząć obliczanie poziomu trendu i sezonowości dla kolejnych miesięcy. W komórkach E2:G2 umieść wartości parametrów *alfa*, *gamma* i *delta* (jak zwykle zacznę od 0,5). Na rysunku 8.42 pokazano arkusz przygotowany do dalszej pracy.

	A	B	C	D	E	F	G
1	Liczba miesięcy				Parametr wygładzania poziomu (alfa)	Parametr wygładzania a błędu (gamma)	Parametr wygładzania sezonowego (delta)
2	36				0,5	0,5	0,5
3							
4	t	Rzeczywisty popyt	Pozio	Trenc	Wyrównanie sezonow	Prognoza jednego kroku	Błąd prognozy
5	-11	#N/D	#N/D	#N/D	0,988	#N/D	#N/D
6	-10	#N/D	#N/D	#N/D	1,039	#N/D	#N/D
7	-9	#N/D	#N/D	#N/D	0,933	#N/D	#N/D
8	-8	#N/D	#N/D	#N/D	0,913	#N/D	#N/D
9	-7	#N/D	#N/D	#N/D	1,043	#N/D	#N/D
10	-6	#N/D	#N/D	#N/D	0,906	#N/D	#N/D
11	-5	#N/D	#N/D	#N/D	0,921	#N/D	#N/D
12	-4	#N/D	#N/D	#N/D	0,927	#N/D	#N/D
13	-3	#N/D	#N/D	#N/D	0,988	#N/D	#N/D
14	-2	#N/D	#N/D	#N/D	1,016	#N/D	#N/D
15	-1	#N/D	#N/D	#N/D	1,048	#N/D	#N/D
16	0		144,4235	2,29045	1,204	#N/D	#N/D
17	1	165	156,8393	7,353103	1,022	144,987667	20,01233271

Rysunek 8.42. Arkusz zawierający parametry wygładzania oraz prognozę wartości dla pierwszego kroku i jej błąd

Pierwszą wartością, którą musisz obliczyć, jest nowa szacunkowa wartość poziomu dla okresu 1. W związku z tym w kolumnie C umieść formułę:

```
=WARUNKI(
    [@t] = 0; 'Holt-Winters inicjalizacja'!$I$20;
```

$[@t] > 0$; SUMA(PRZESUNIĘCIE([@Poziom]:[@Trend]; -1;0)) + \$E\$2*[@Błąd prognozy] / PRZESUNIĘCIE([@Wyrównanie sezonowe]); -'Holt-Winters inicjalizacja'!\$B\$1; 0))

Dla $t > 0$ sumowane są dwie poprzednie komórki z kolumny Poziom i Trend. Następnie dodawany jest parametr wygładzania podzielony przez wyrównanie sezonowe sprzed 12 miesięcy. W poprzedniej sekcji dowiedziałeś się, że wartość nowego poziomu jest równa sumie wartości poprzedniego poziomu, poprzedniego trendu i parametru *alfa* pomnożonego przez błąd prognozy uwzględniającej sezonowość.

Wprowadź zmiany w formule w kolumnie D. Formuła jest podobna do tej z kolumny C.

```
=WARUNKI(
    [@t] = 0; 'Holt-Winters inicjalizacja'!$I$19;
    [@t] > 0;
    PRZESUNIĘCIE([@Trend]; -1; 0) + $F$2 * $E$2 * [@Błąd prognozy] /
    PRZESUNIĘCIE([@Wyrównanie sezonowe]); -'Holt-Winters
    inicjalizacja'!$B$1; 0)
)
```

Nowy trend jest równy sumie poprzedniego trendu i iloczynu parametru *gamma* oraz błędu prognozy uwzględniającej sezonowość.

Aktualną wartość wyrównania sezonowego dla stycznia oblicz za pomocą formuły:

```
=WARUNKI(
    [@t] < 1; INDEKS(
        HoltWintersInicjalizacja.CzynnikiSezonowości[Początkowe czynniki
        sezonowe];
        [@t] + 'Holt-Winters inicjalizacja'!$B$1
    );
    ORAZ([@t] >= 1; [@t]<=$A$2);
    PRZESUNIĘCIE([@Wyrównanie sezonowe]); -'Holt-Winters
    inicjalizacja'!$B$1; 0) +
    $G$2 * (1 - $E$2) * [@Błąd prognozy] /
    SUMA(PRZESUNIĘCIE(HoltWintersPrognoza.Dane[@Poziom]:
    [Trend]); -1;0))
)
```

Wartość ta jest równa czynnikowi wyrównania sezonowego z poprzedniego stycznia skorygowanemu o iloczyn parametru *delta* i błędu pominiętego podczas korekcji poziomu, który był skalowany tak samo jak czynniki sezonowe — poprzez dzielenie przez wartości przypisane poprzedniemu poziomowi i poprzedniemu trendowi. Jeżeli wszystko wykonałeś poprawnie, Twój arkusz powinien wyglądać tak jak na rysunku 8.43.

312 Mistrz analizy danych

	A	B	C	D	E	F	G
1	Liczba miesięcy				Parametr wygładzania poziomu (alfa)	Parametr wygładzania błędów (gamma)	Parametr wygładzania sezonowego (delta)
2	36				0,5	0,5	0,5
3							
4	t	Rzeczywisty popyt	Pozycja	Trend	Wyrównanie sezonowe	Prognoza jednego kroku	Błąd prognozy
5	-11		#N/D	#N/D	0,988	#N/D	#N/D
6	-10		#N/D	#N/D	1,039	#N/D	#N/D
7	-9		#N/D	#N/D	0,933	#N/D	#N/D
8	-8		#N/D	#N/D	0,913	#N/D	#N/D
9	-7		#N/D	#N/D	1,043	#N/D	#N/D
10	-6		#N/D	#N/D	0,906	#N/D	#N/D
11	-5		#N/D	#N/D	0,921	#N/D	#N/D
12	-4		#N/D	#N/D	0,927	#N/D	#N/D
13	-3		#N/D	#N/D	0,988	#N/D	#N/D
14	-2		#N/D	#N/D	1,016	#N/D	#N/D
15	-1		#N/D	#N/D	1,048	#N/D	#N/D
16	0		144,424	2,290	1,204	#N/D	#N/D
17	1	165	156,839	7,353	1,022	144,988	20,012
18	2	171	164,350	7,432	1,040	170,671	0,329
19	3	147	164,675	3,878	0,914	160,262	-13,262
20	4	143	162,624	0,914	0,897	153,822	-10,822
21	5	164	160,388	-0,661	1,033	170,572	-6,572
22	6	160	168,120	3,536	0,930	144,783	15,217
23	7	152	168,361	1,888	0,912	158,067	-6,067
24	8	150	166,064	-0,204	0,915	157,757	-7,757
25	9	159	163,355	-1,457	0,981	163,951	-4,951
26	10	169	164,102	-0,355	1,023	164,522	4,478
27	11	173	164,408	-0,025	1,050	171,616	1,384
28	12	203	166,493	1,031	1,212	197,918	5,082
29	13	169	166,416	0,477	1,019	171,266	-2,266

Rysunek 8.43. Aktualizowanie wartości dla 36 miesięcy

Zauważ, że w trzech ostatnich formułach odwołania do wartości parametrów *alfa*, *gamma* i *delta* są bezwzględne, co umożliwia przeciąganie tych formuł w dół arkusza. Jednak naprawdę powinieneś rozważyć stosowanie nazwanych zakresów. (Tutaj ich unikam z uwagi na to, że wielokrotnie kopiuję arkusze, a co za tym idzie, wprowadzenie nazwanych arkuszy mogłoby zaciemnić to, co chcę przekazać, i wydłużyć instrukcje, ale w rzeczywistości ten przykład doskonale nadaje się do zastosowania nazwanych zakresów).

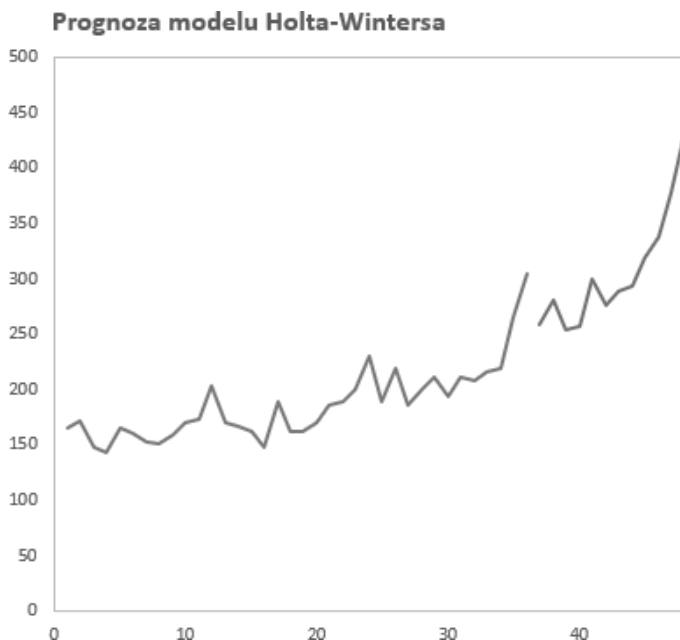
Dysponując ostatnim poziomem, trendem i wszystkimi wyrównaniami sezonowymi, możesz dokonać prognozy popytu na kolejny rok. Najpierw zwiększ tabelę do 48 rekordów. Tak samo jak poprzednio, dodaj kolumny **Wartość rzeczywista** i **Prognoza**. W kolumnie dla wartości rzeczywistych umieść taką formułę:

$$=JEŻELI(ORAZ([@t] >0; [@t] <= \$A\$2); [@Rzeczywisty popyt]); BRAK())$$

Natomiast w kolumnie prognozy umieść formułę pokazaną poniżej. Ta formuła zacznie zwracać wartości po $t > 37$.


```
=JEŻELI([@t] > $A$2;
    (X.WYSZUKAJ($A$2; [t]; [Poziom]) + ([@t]-$A$2) * X.WYSZUKAJ($A$2; [t];
    [Trend])) * PRZESUNIĘCIE([@Wyrównanie sezonowe]); - 'Holt-Winters
    inicjalizacja'!$B$1;0);
    BRAK()
)
```

Podobnie jak w przypadku podwójnego wygładzania wykładniczego do wartości szacunkowej ostatniego poziomu dodajesz trend *pomnożony* przez liczbę miesięcy, które upłynęły od ostatniego oszacowania trendu. Jediną różnicą jest to, że skalujesz całą prognozę przez ostatni mnożnik sezonowości właściwy dla stycznia. Po naciśnięciu klawisza *Enter* uzyskasz prognozę. Tak jak w przypadku dwóch poprzednich technik prognozę i wartość rzeczywistą możesz przedstawić na wykresie (rysunek 8.44).



Rysunek 8.44. Wykres prognozy modelu Holta-Wintersa

Czas na optymalizację

Myślisz, że to koniec? Jeszcze nie! Czas określić optymalne wartości parametrów wygładzających. Do tabeli dodaj nową kolumnę (jako kolumnę J) i nazwij ją **Błąd podniesiony do kwadratu**. Umieść w niej taką formułę:

```
=JEŻELI.BŁĄD([@Błąd prognozy])^2 * ORAZ([@t]>0;[@t]<=$A$2);0)
```

Funkcja *ORAZ* to proste sprawdzenie, czy dane mieszczą się w odpowiednim zakresie. Jeśli nie, formuła zwraca 0. Funkcja *JEŻELI.BŁĄD* obsługuje błąd #N/D. Podobnie

jak w dwóch poprzednich technikach w komórce I2 umieścić formułę obliczającą sumę kwadratów błędów, a w komórce J2 umieścić formułę określającą błąd standardowy.

Zauważ, że tym razem masz trzy parametry wygładzające, a więc błąd standardowy należy obliczać za pomocą formuły:

$$=PIERWIASTEK(I2/(A2-3))$$

Uzyskana przez Ciebie wartość błędu standardowego powinna być równa 12,568.

Konfigurację dodatku Solver przedstawiono na rysunku 8.45. Tym razem optymalizujesz zawartość komórki J2, zmieniając wartości przypisane trzem parametrom wygładzającym. W wyniku optymalizacji możesz uzyskać błąd standardowy równy połowie błędu powstającego podczas korzystania z dwóch opisanych wcześniej technik. Na rysunku 8.46 przedstawiono wykres zoptymalizowanej prognozy. Wygląda dobrze, nieprawdaż? Tym razem Twoja prognoza bierze pod uwagę trend oraz sezonowe wahania popytu. Świetnie!

Parametry dodatku Solver

Ustaw cel:

Na: Maks Min Wartość:

Przez zmienianie komórek zmiennych:

Podlegających ograniczeniom:

Ustaw wartości nieujemne dla zmiennych bez ograniczeń

Wybierz metodę rozwiązywania:

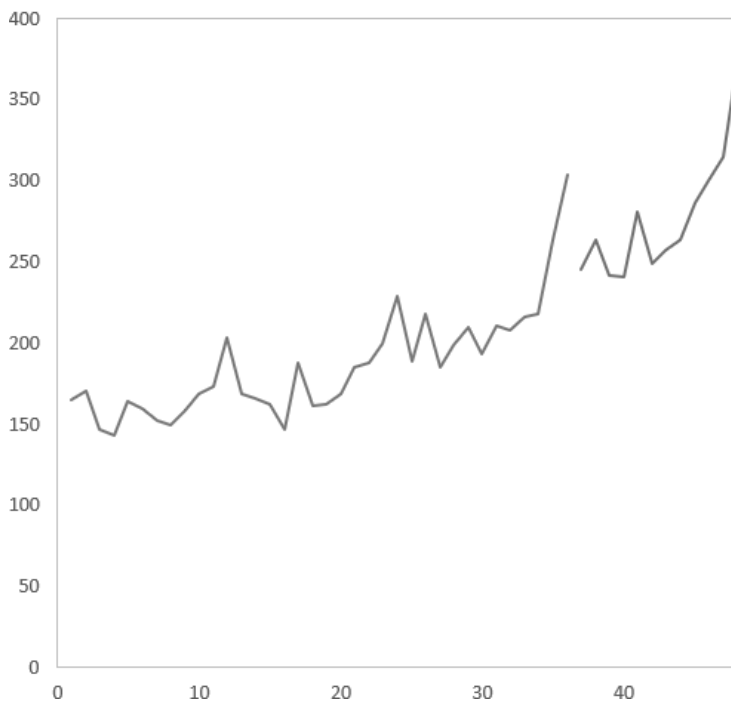
Metoda rozwiązywania

W przypadku gładkich nieliniowych problemów dodatku Solver wybierz aparat nieliniowy GRG. Dla liniowych problemów dodatku Solver wybierz aparat LP simpleks, natomiast w przypadku problemów, które nie są gładkie, wybierz aparat ewolucyjny.

Pomoc Rozwiąż Zamknij

Rysunek 8.45. Konfiguracja dodatku Solver dla modelu Holta-Wintersa

Prognoza modelu Holta-Wintersa



Rysunek 8.46. Zoptymalizowana prognoza modelu Holta-Wintersa

Teraz musisz sprawdzić autokorelację tej prognozy. Arkusz autokorelacji utworzyłeś już wcześniej, a więc tym razem wystarczy, że go skopiujesz i wkleisz do niego nowe wartości błędów.

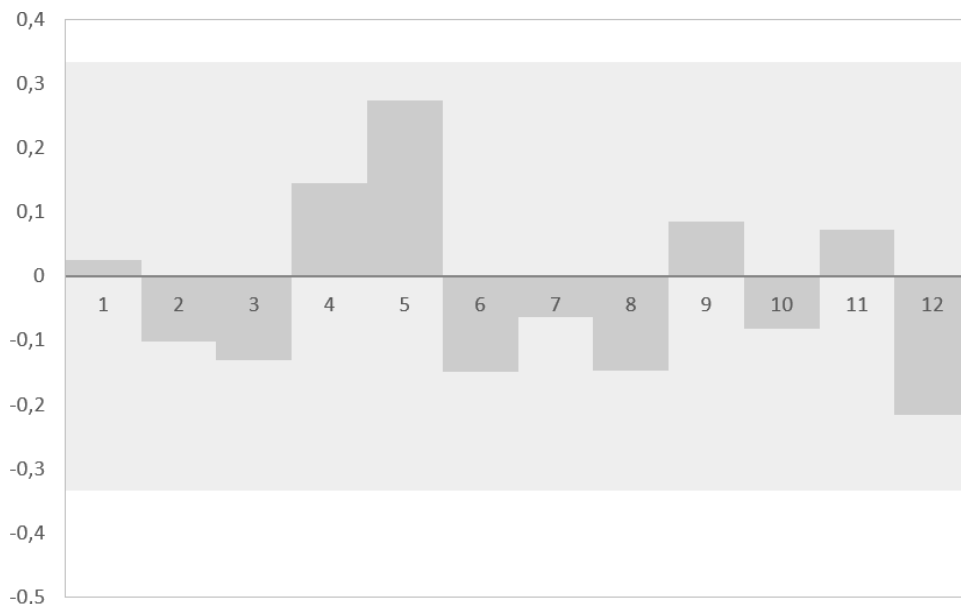
Utwórz kopię arkusza *Autokorelacja metody Holta* i nazwij go **Holt-Winters autokorelacja**. Skopiuj wartości błędów z kolumny *Błąd prognozy* arkusza *Holt-Winters prognoza* (komórki G17:G52) i wklej je do kolumny B arkusza *Holt-Winters autokorelacja* (skorzystaj z opcji wklejania specjalnego). W ten sposób uzyskasz korelogram pokazany na rysunku 8.47.

Bum! Nie ma autokorelacji przekraczających wartość krytyczną równą 0,33, a więc możesz uznać, że model poprawnie odzwierciedla strukturę wartości popytu.

Interwały prognozy

Uzyskałeś dobrze dopasowany model prognozujący. Jak dodać do niego górne i dolne ograniczenia, które pozwolą na definiowanie rzeczywistych oczekiwań?

Można to zrobić za pomocą symulacji Monte Carlo. Ogólnie rzecz biorąc, metoda ta sprowadzi się w tym przypadku do wygenerowania scenariuszy popytu w przyszłości i określenia granic, w których mieści się 95% tych scenariuszy. Jak rozpocząć generowanie symulacji popytu w przyszłości? To nic trudnego.



Rysunek 8.47. Autokorelacja modelu Holta-Wintersa

Zacznij od utworzenia kopii arkusza *Holt-Winters prognoza* i nazwania jej **Interwały prognozy**. Usuń wszystkie wykresy z nowego arkusza (nie będą Ci już potrzebne). Ponadto zmień nazwę tabeli na **InterwałyPrognozy.Dane**.

Na początku tego rozdziału napisałem, że prognoza jest zawsze błędna. Podczas prognozowania nie da się uniknąć błędów, ale w tym przypadku znasz dystrybucję błędów. Opracowałeś dobrze dopasowaną prognozę i możesz założyć, że charakteryzuje się ona średnią błędu jednego kroku równą 0 (prognoza jest nieobciążona) oraz standardowym odchyleniem równym 10,38 (wartość tę obliczyłeś w poprzednim arkuszu).

Funkcja `ROZKŁAD.NORMALNY.ODW` umożliwia wygenerowanie symulowanych błędów w przyszłych. Jako argumenty funkcji `ROZKŁAD.NORMALNY.ODW` możesz podać średnią (0), odchylenie standardowe (wartość 10,38 znajdującą się w komórce J2), a także liczbę losową z zakresu od 0 do 1. Spowoduje to wygenerowanie wartości błędu należącej do krzywej dzwonowej. Więcej informacji na temat krzywej dzwonowej znajdziesz w rozdziale 9. „Modelowanie optymalizacyjne — »świeżo wyciśnięty« sok nie zamiesza się sam”.

Z uwagi na obecny kształt modelu mógłbyś od razu zmienić formułę w kolumnie **Błąd prognozy**. Ale może to spowodować błąd odwołania cyklicznego. Aby tego uniknąć, zmień formułę obliczającą sumę kwadratów błędów w komórce I2. To zapobiegnie błędom cyklicznych odwołań. Przyjrzyj się tej formule i spróbuj zrozumieć, jak to jest zaimplementowane.

```
=SUMA.WARUNKÓW(InterwałyPrognozy.Dane[Błąd podniesiony do kwadratu];
InterwałyPrognozy.Dane[t]; ">=1";
InterwałyPrognozy.Dane[t]; "<=" & $A$2
)
```

Teraz zmień formułę dla błędu prognozy z tej:

```
=[@Rzeczywisty popyt] - [@Prognoza jednego kroku]
```

na tę:

```
= JEŻELI([@t] <= $A$2;
[@Rzeczywisty popyt]-[@Prognoza jednego kroku]];
ROZKŁ.NORMALNY.ODWR(LOS(); 0; $J$2)
)
```

Teraz masz symulowane wartości błędów dla prognozy na 12 miesięcy. Po wykonaniu tych operacji uzyskasz arkusz pokazany na rysunku 8.48 (Twój arkusz będzie z pewnością zawierał inne wartości wygenerowane w wyniku symulacji).

t	Rzeczywisty	Poziom	Trend	Wyrównanie st	Prognoza j	Błąd prognozy	Błąd podniesiony	Wartość rz	Prognoza	
49	33	216	225,46362	3,5303198	0,988	225,926339	-9,926338637	98,53219873	216	#N/D
50	34	218	224,53686	2,5009911	1,016	232,703978	-14,70397751	216,2069545	218	#N/D
51	35	264	234,69488	4,2693214	1,048	237,947618	26,05238207	678,7266118	264	#N/D
52	36	304	243,13079	5,2315613	1,204	287,71407	16,28593012	265,2315197	304	#N/D
53	37		247,34231	4,9959924	0,988	245,439968	-3,272490831	10,70919624	#N/D	242,1674772
54	38		255,40118	5,7033378	1,039	262,295453	10,33570391	106,8267753	#N/D	272,631157
55	39		258,51775	5,1059448	0,933	243,593094	-7,834507599	61,37950931	#N/D	235,7585865
56	40		259,99195	4,2672233	0,913	240,58239	-10,75964952	115,7700579	#N/D	229,8227406
57	41		264,66106	4,3600363	1,043	275,625122	1,360812839	1,851811583	#N/D	276,9859344
58	42		265,14955	3,465932	0,906	243,852145	-11,39276995	129,7952072	#N/D	232,4593752
59	43		264,11954	2,4276312	0,921	247,35123	-13,44024511	180,6401887	#N/D	233,9109845
60	44		264,08375	1,8587236	0,927	246,988192	-7,410453246	54,91481731	#N/D	239,5777386
61	45		264,41101	1,5050446	0,988	262,881676	-4,914540187	24,15270525	#N/D	257,9671357
62	46		264,74094	1,2336608	1,016	270,224282	-3,876722784	15,02897954	#N/D	266,3475591
63	47		265,09187	1,029802	1,048	278,755385	-3,003401468	9,020420377	#N/D	275,7519836
64	48		275,00787	3,0819969	1,204	320,411802	34,73344149	1206,411958	#N/D	355,1452432

Rysunek 8.48. Wartości błędów poszczególnych kroków uzyskane w wyniku symulacji

Musisz wprowadzić zmiany w kolumnie Wyrównanie sezonowe. Teraz formuła w tej kolumnie wygląda tak:

```
=WARUNKI(
  [@t] < 1; INDEKS(
    HoltWintersInicjalizacja.CzynnikiSezonowości[Początkowe czynniki
    sezonowe];
    [@t] + 'Holt-Winters inicjalizacja'!$B$1
  );
  ORAZ([@t] >= 1; [@t]<=$A$2);
  PRZESUNIĘCIE([@Wyrównanie sezonowe]); -'Holt-Winters
  inicjalizacja'!$B$1; 0) +
  $H$2 * (1 - $F$2) * [@Błąd prognozy] /
  SUMA(PRZESUNIĘCIE(HoltWintersPrognoza.Dane[@Poziom];
  [Trend]]; -1;0)
)
```

Należy wprowadzić drobną zmianę w drugim warunku. Nie potrzebujesz już funkcji ORAZ. Jeśli t jest większe od 1, ma zostać obliczona wartość wyrównania sezonowego:

```
=WARUNKI(
  [@t] < 1; INDEKS(
    HoltWintersInicjalizacja.CzynnikiSezonowości[Początkowe czynniki sezonowe];
    [@t] + 'Holt-Winters inicjalizacja'!$B$1
  );
  [@t] >= 1;
  PRZESUNIĘCIE([@Wyrównanie sezonowe]); -'Holt-Winters inicjalizacja'!$B$1; 0)
  + $G$2 * (1 - $E$2) * [@Błąd prognozy] /
  SUMA(PRZESUNIĘCIE(InterwałyPrognozy.Dane[@Poziom]:[Trend]); -1;0)
)
```

Dotarłeś do punktu, w którym rzeczy zaczynają się komplikować z analitycznego punktu widzenia. Dysponujesz symulowanym błędem prognozy oraz prognozą jednego kroku, a więc jeżeli do prognozy dodasz błąd, możesz odtworzyć wartości popytu dla danego okresu.

W tym celu w kolumnie Prognoza zmień formułę na taką:

```
=JEŻELI([@t] > $A$2;
  [@Prognoza jednego kroku] +[@Błąd prognozy];
  BRAK())
)
```

Formuła automatycznie wypełni całą kolumnę (rysunek 8.49).

t	Rzeczywii	Poziom	Trend	Wyrównanie sez	Prognoza	Błąd prognozy	Wartość rz	Prognoza	Błąd podniesion	
50	34	218	224,53686	2,5009911	1,016	232,703978	-14,70397751	218	#N/D	216,2069545
51	35	264	234,69488	4,2693214	1,048	237,947618	26,05238207	264	#N/D	678,7266118
52	36	304	243,13079	5,2315613	1,204	287,71407	16,28593012	304	#N/D	265,2315197
53	37		249,73523	5,5486182	0,988	245,439968	4,404509775	#N/D	249,8444778	19,39970636
54	38		256,00627	5,7154549	1,039	265,357229	2,437812623	#N/D	267,7950414	5,942930384
55	39		266,25363	6,7620623	0,933	244,168911	13,72572618	#N/D	257,894637	188,3955593
56	40		272,97222	6,7520221	0,913	249,153509	-0,128801997	#N/D	249,0247074	0,016589954
57	41		281,55742	7,1753793	1,043	291,75535	6,207213101	#N/D	297,9625627	38,52949448
58	42		292,23001	7,9830325	0,906	261,719663	10,29120198	#N/D	272,0108653	105,9088382
59	43		300,46401	8,0409916	0,921	276,447453	0,750248215	#N/D	277,197701	0,562872384
60	44		309,33628	8,2329687	0,927	285,867194	2,500648731	#N/D	288,3678432	6,253244078
61	45		318,72613	8,5001422	0,988	313,914265	3,712503455	#N/D	317,6267683	13,78268191
62	46		328,24746	8,7359759	1,016	332,527817	3,368888372	#N/D	335,8967051	11,34940886
63	47		335,89692	8,4850538	1,048	353,176382	-3,696774335	#N/D	349,4796076	13,66614048
64	48		346,37561	8,9454695	1,204	414,637583	7,792545449	#N/D	422,4301289	60,72376457

Rysunek 8.49. Popyt w przyszłości (wartości uzyskane w wyniku symulacji)

Wygenerowałeś dane jednego scenariusza. W celu wygenerowania danych innych scenariuszy wystarczy odświeżyć arkusz (spowoduje to zmianę wartości). W związku z tym możesz wygenerować wiele scenariuszy popytu w przyszłości, kopiując w inne miejsce dane scenariuszy i odświeżając arkusz.

Zaczynaj od umieszczenia etykiety **Symulacja popytu** w komórce A69, a w komórkach A70:L70 umieść numery miesięcy od 37 do 48. Możesz to zrobić, kopiując zawartość komórek A53:A64 i wklejając ją do komórek A70:L70 za pomocą opcji wklejania

specjalnego z transpozycją wartości. Za pomocą tej samej techniki (wklejania specjalnego z transpozycją wartości) skopiuj dane popytu pierwszego scenariusza do komórek A71:L71. W celu wstawienia danych drugiego scenariusza kliknij wiersz 71. i wybierz opcję *Wstaw*, za pomocą której dodasz do arkusza nowy pusty wiersz. Teraz możesz umieścić w tym wierszu kolejne wartości popytu uzyskane w wyniku kolejnej symulacji (wartości powinny zostać zmienione po wykonaniu ostatniej operacji wklejania danych).

Możesz wykonywać te operacje w kółko w celu wygenerowania tak dużej liczby scenariuszy popytu, jak tylko zechcesz. W praktyce będzie to czasochłonne, a więc zoptymalizuj ten proces, rejestrując makro.

Tak jak robiłeś poprzednim razem, rozpocznij proces rejestracji makra, a następnie wykonaj poniższe operacje:

1. Wstaw pusty wiersz w miejscu wiersza 71.
2. Skopiuj komórki I53:I64.
3. Wklej specjalnie transponowane wartości do wiersza 71.
4. Kliknij przycisk zatrzymujący rejestrację makra.

Teraz możesz przypisać do makra dowolny skrót klawiszowy i używać go aż do wygenerowania tony scenariuszy. Możesz również wcisnąć i przytrzymać kombinację klawiszy pełniącą funkcję skrótu uruchamiającego makra — wygenerowanie 1000 scenariuszy w ten sposób nie powinno przysporzyć żadnego problemu. Jeżeli pomysł przytrzymywania klawiszy wydaje Ci się okropny, to możesz zapętlić makro za pomocą języka Visual Basic for Applications. Odsyłam Cię do mojej innej książki (w języku angielskim) *Dashboards for Excel* (Apress Berkeley, CA, 2015).

Po wykonaniu tych czynności Twój arkusz powinien wyglądać tak, jak pokazano na rysunku 8.50.

Dysponujesz już scenariuszami dla każdego miesiąca, a więc możesz skorzystać z funkcji PERCENTYL w celu ustalenia górnej i dolnej wartości granicznej — wartości, które wyznaczają 95% środkowych scenariuszy, co umożliwi utworzenie interwału prognozy.

Na przykład nad miesiącem 37., w komórce A66, możesz umieścić następującą formułę:

```
=PERCENTYL(A71:A1070;0.975)
```

Formuła ta zwróci percentyl 97,5% popytu w miesiącu 37. W moim arkuszu wartość ta wynosi ok. 264. W komórce A67 możesz ustalić percentyl 2,5% analizowanego miesiąca za pomocą formuły:

```
=PERCENTYL(A71:A1070;0.025)
```

	A	B	C	D	E	F	G	H	I	J	K	
69	Symulacja popytu											
70	37	38	39	40	41	42	43	45	46	47	48	
71	253,895	257,511	236,488	235,120	287,359	242,202	274,590	297,017	299,914	341,439	373,500	
72	231,758	255,289	225,791	235,913	250,314	236,774	237,426	245,404	244,942	261,897	294,143	
73	240,580	264,497	247,951	225,703	285,539	237,321	240,553	262,647	281,693	291,788	344,456	
74	233,041	248,897	255,755	246,005	280,663	259,272	248,713	285,545	293,171	321,717	362,808	
75	251,231	259,940	243,998	227,327	261,124	240,857	253,848	277,300	272,353	294,868	344,160	
76	243,992	253,569	229,566	230,657	285,372	248,917	269,643	298,597	299,779	332,537	388,326	
77	253,822	264,231	242,581	245,745	287,768	246,623	278,859	310,957	320,524	346,655	398,342	
78	244,476	260,429	242,490	246,857	285,153	240,988	246,383	262,764	299,037	315,513	352,067	
79	248,293	259,015	224,712	230,235	270,781	231,518	246,563	264,667	270,446	277,273	324,619	
80	247,360	276,465	232,034	250,241	290,641	238,455	275,009	292,218	292,567	308,872	360,384	
81	239,725	241,731	234,028	245,837	260,824	243,518	254,293	291,041	312,064	320,296	385,007	
82	234,466	270,167	254,208	252,937	300,907	272,314	268,044	320,363	349,755	355,494	434,294	
83	226,041	243,004	221,190	227,270	262,297	236,761	246,115	265,499	282,100	293,436	329,548	
84	233,617	250,219	235,092	234,366	274,737	250,769	243,602	280,448	264,899	284,426	335,669	
85	245,389	267,375	253,714	243,003	285,408	248,194	246,892	286,884	287,445	285,555	342,356	
86	257,981	285,241	253,476	259,011	303,546	271,008	287,233	304,301	325,119	320,955	394,402	
87	223,674	251,071	240,125	230,299	265,992	224,344	249,717	266,287	259,803	271,325	331,494	
88	250,134	267,969	242,633	228,918	278,568	217,628	231,286	262,525	255,339	267,834	313,889	
89	233,698	263,667	231,835	221,360	278,181	247,600	241,371	260,084	278,769	282,528	328,107	
90	244,542	268,148	250,695	248,337	290,564	252,890	271,379	298,017	306,527	329,829	392,881	
91	253,914	266,332	243,997	241,723	296,675	260,145	271,436	291,062	314,480	331,023	375,289	
92	259,729	267,947	256,404	241,333	288,138	232,299	253,475	260,297	288,834	295,441	352,550	
93	257,479	257,151	248,954	244,475	283,647	266,331	267,060	305,544	303,701	332,612	398,070	
94	240,014	277,102	244,872	238,515	293,375	238,212	246,828	281,220	283,292	320,390	345,049	
95	245,990	246,702	261,146	239,871	295,607	239,748	254,952	282,264	302,956	308,133	347,038	
96	251,564	284,875	249,332	239,219	301,270	256,162	274,130	310,519	337,141	352,430	411,914	
97	231,628	272,354	236,726	243,399	292,557	234,452	258,983	277,563	278,802	327,611	359,695	
98	245,159	269,642	263,441	260,822	299,035	257,125	262,837	282,602	304,133	322,434	369,303	
99	245,693	246,672	214,159	225,291	258,867	212,156	218,112	236,838	228,856	236,076	266,164	

Rysunek 8.50. Wygenerowałem 1000 scenariuszy popytu

W tej formule podałem zakres komórek A71:A1070, ponieważ dysponuję danymi 1000 symulowanych scenariuszy. W zależności od zrzętności palca wskazującego mogłeś wygenerować więcej lub mniej scenariuszy. W moim arkuszu to dolne ograniczenie wyniosło ok. 224.

Wynika z tego, że pomimo przewidywanego popytu wynoszącego 245 (w miesiącu 37.) 95% interwału prognozy mieści się pomiędzy popytem wynoszącym 224 a popytem wynoszącym 264.

Formuły obliczające percentyle możesz przeciągnąć w bok aż do miesiąca 48. (kolumny L). W ten sposób ustalisz całość interwału (rysunek 8.51). Teraz możesz przedstawić swojemu przełożonemu prognozę wraz z zakresem wartości, pomiędzy którymi będzie się ona wahać. W formułach możesz zmienić wartości 0,025 i 0,975 na np. 0,1 i 0,9 (interwał 90%) lub 0,2 i 0,8 (interwał 80%).

	A	B	C	D	E	F	G	H	I	J	K	
65												
66	263,592	284,285	264,700	266,197	309,379	279,546	291,232	326,654	348,215	370,744	436,817	
67	223,960	240,793	218,676	216,820	251,661	220,213	224,330	243,527	254,129	263,293	302,768	
68												
69	Symulacja popytu											
70	37	38	39	40	41	42	43	45	46	47	48	
71	253,895	257,511	236,488	235,120	287,359	242,202	274,590	297,017	299,914	341,439	373,500	
72	231,758	255,289	225,791	235,913	250,314	236,774	237,426	245,404	244,942	261,897	294,143	
73	240,580	264,497	247,951	225,703	285,539	237,321	240,553	262,647	281,693	291,788	344,456	
74	233,041	248,897	255,755	246,005	280,663	259,272	248,713	285,545	293,171	321,717	362,808	

Rysunek 8.51. Interwał prognozy uzyskanej za pomocą modelu Holta-Wintersa

Tworzenie wykresu warstwowego wachlarza wartości

Wykonywanie tej ostatniej czynności nie jest koniecznością, ale prognozy z interwałami są często przedstawiane za pomocą **wykresu wachlarza**. Wykres ten możesz utworzyć w Excelu.

Zacznij od utworzenia nowego arkusza i nadania mu nazwy **Wykres wachlarza**. Następnie skopiuj do jego 1. wiersza etykiety miesięcy (numery od 37 do 48). Do 2. wiersza wklej dane górnego ograniczenia interwału prognoz (skopiuj je z 66. wiersza arkusza *Interwały prognozy*). Do 3. wiersza wklej za pomocą opcji wklejania specjalnego z transpozycją wartości prognoz z arkusza *Holt-Winters prognoza*. Do 4. wiersza wklej wartości dolnego ograniczenia interwału prognoz (skopiuj je z 67. wiersza arkusza *Interwały prognozy*).

Po wykonaniu tych czynności w kolejnych wierszach Twojego arkusza będą się znajdowały numery miesięcy, wartości górnego ograniczenia interwału, prognozowane wartości i wartości dolnego ograniczenia interwału (rysunek 8.52).

	A	B	C	D	E	F	G	H	I	J	K	L
1	37	38	39	40	41	42	43	44	45	46	47	48
2	263,5925	284,2848	264,7001	266,1974	309,379	279,5458	291,2321	300,1703	326,6543	348,2147	370,7438	436,8169
3	245,44	263,6006	241,4669	240,9779	280,8709	248,8367	257,6059	264,0715	286,8747	300,2331	315,1263	368,3166
4	223,9602	240,7928	218,6765	216,8198	251,6615	220,213	224,3303	230,7295	243,5266	254,1289	263,2932	302,7681

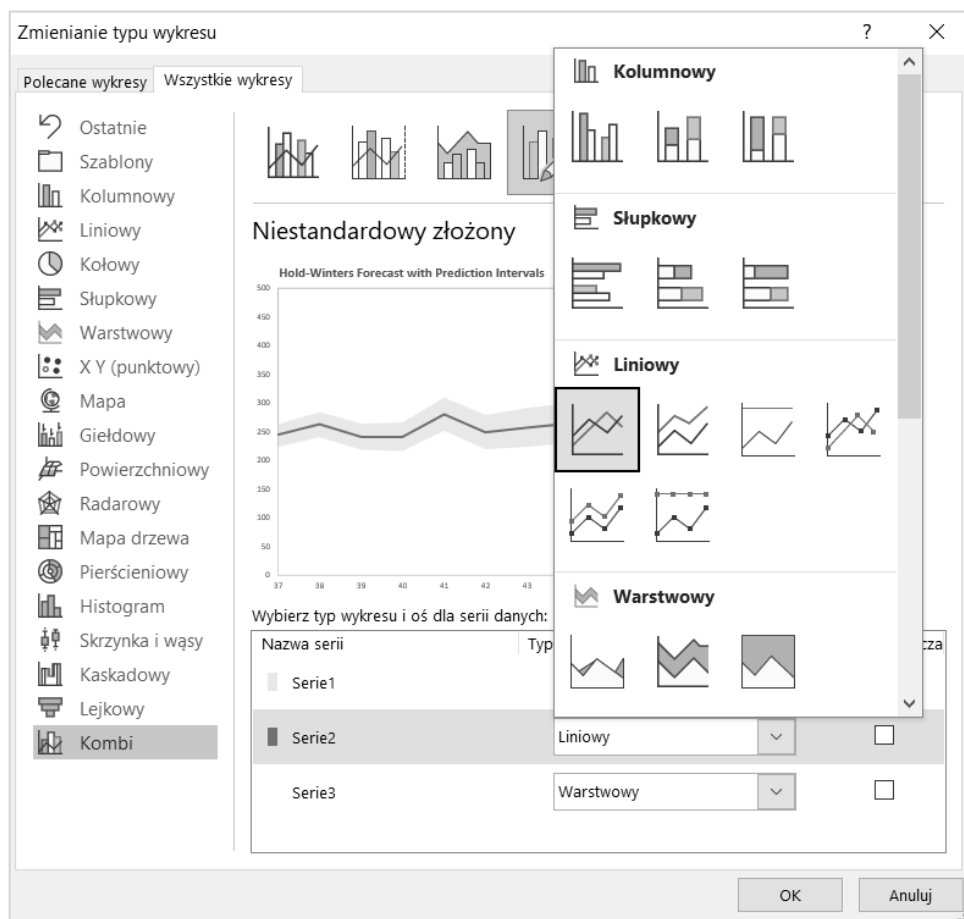
Rysunek 8.52. Dane prognozy otoczone wartościami granicznymi interwału

Zaznacz komórki A2:L4 i wybierz opcję tworzenia wykresu warstwowego. Spowoduje to wygenerowanie wykresu składającego się z trzech ułożonych na sobie ciągłych warstw. Kliknij jedną z serii prawym przyciskiem myszy i wybierz opcję *Zaznacz dane...* Zmień etykiety osi poziomej (kategorii) — w tej części okna umieść serię danych znajdującą się w komórkach A1:L1. Spowoduje to wyświetlenie na wykresie poprawnych etykiet z numerami miesięcy. (Nie zostawiaj tego jako 1 – 12, gdyż nie wygląda to dobrze).

Teraz upiększ nieco ten wykres. Kliknij prawym przyciskiem myszy serię danych dolnych wartości granicznych i sformatuj ją tak, aby charakteryzowała się białym wypełnieniem. Aby wykres ten wyglądał podobnie do poprzednich, usuń z niego linie siatki. Następnie kliknij prawym przyciskiem myszy dowolną serię danych i wybierz *Zmień typ wykresu seryjnego*. Zmień drugą serię (prawdopodobnie nazywaną *Serie2*) na wykres liniowy, tak jak pokazano na rysunku 8.53.

Na koniec naciśnij OK.

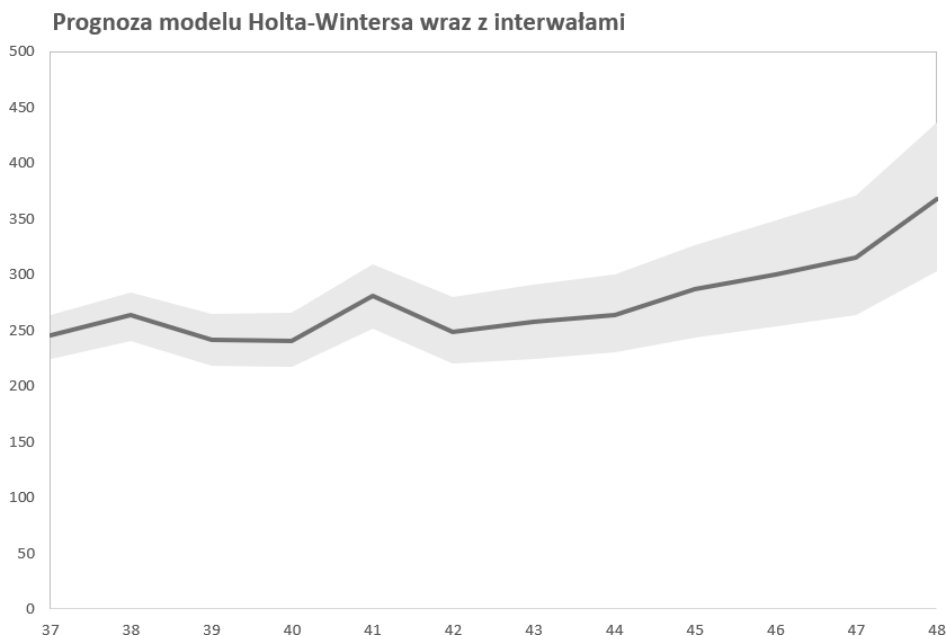
Za każdym razem, gdy zmieniasz typ wykresu na liniowy, Excel dodaje pustą przestrzeń po lewej i prawej stronie wykresu. Aby się tego pozbyć, kliknij prawym przyciskiem myszy oś y i wybierz *Formatuj oś...*



Rysunek 8.53. Okno Zmianianie typu wykresu

By usunąć białe puste pole, w panelu *Formatowanie osi*, w zakładce *Opcje osi*, poszukaj pola *Położenie etykiety* i wybierz *Obok osi*. Usunąłem również znaczniki osi, ponieważ nie są potrzebne. Tę opcję znajdziesz w tym samym menu. Rozwiń opcję *Znaczniki osi* i zarówno dla *Typ jednostki głównej*, jak i dla *Typ jednostki pomocniczej* wybierz *Brak*.

Na koniec możesz dodać etykiety osi i tytuł wykresu. Spójrz na mój wykres wachlarzowy widoczny na rysunku 8.54. Dzięki formatowaniu możesz sprawić, by ten wykres był bardzo czytelny. Zgodnie z dobrymi praktykami wizualizacji danych zachęcam Cię do wybrania jaśniejszych kolorów, z wyjątkiem głównej linii prognozy, która powinna się wyróżniać.



Rysunek 8.54. Wykres wachlarza wygląda pięknie

Największą zaletą tego wykresu jest to, że prezentuje on w prosty sposób jednocześnie prognozę oraz interwały. Przyglądając się wykresowi, można zauważyć dwie ciekawe rzeczy:

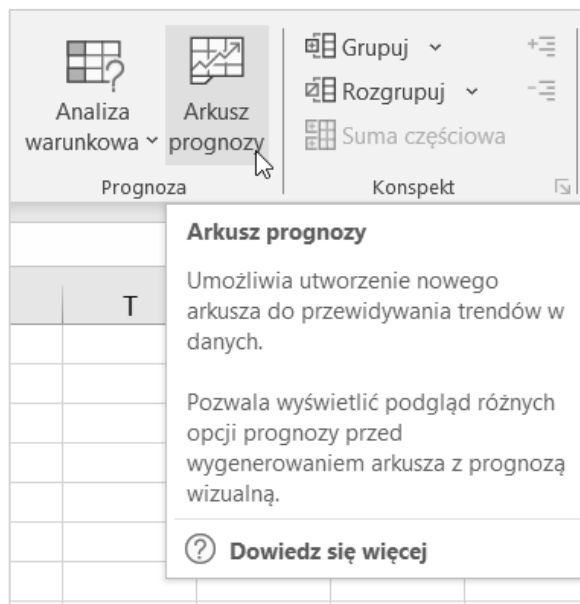
- Błąd poszerza się wraz z upływem czasu. Jest to dość logiczne. Niepewność prognoz na kolejne miesiące nakłada się na siebie.
- Granice błędu są szersze również podczas okresów sezonowego wzrostu popytu. Przy niższym popycie granice błędu są węższe.

Arkusze prognozy w Excelu

Pewnie myślałeś, że to już koniec, ale jest jeszcze jedna, ostania rzecz, którą chcę Ci pokazać w tym rozdziale. W 2010 roku firma Microsoft dodała nowe narzędzie o nazwie *Arkusze prognozy*. Tak naprawdę to narzędzie nie jest rozbudowane, a firma Microsoft ostatnio bardziej skupia się na rozwijaniu funkcjonalności prognozowania za pomocą języków R i Python oraz programu Power BI. W rezultacie to narzędzie jest w połowie zrealizowanym pomysłem. Ale i tak chcę Ci je pokazać.

Excel oferuje bardzo prosty sposób tworzenia prognozy. Jednak jest jeden minus. Podobnie jak w przypadku innych narzędzi, które robią wszystko za Ciebie, nie wiesz, co dzieje się pod spodem, przez co możesz przygotować niedorzeczne wyniki. Bez dokładnego zrozumienia, coś co wygląda w danym momencie dobrze, później może przestać mieć sens.

W celu utworzenia arkusza prognozy przejdź do arkusza *Szereg czasowy* i naciśnij dowolną komórkę tabeli. Następnie z zakładki *Dane* wybierz *Arkusz prognozy* (który znajdziesz w grupie *Prognoza*). Podpowiedź pokazano na rysunku 8.55.



Rysunek 8.55. Arkusz prognozy znajdujący się w zakładce *Dane*, w grupie *Prognoza*

Gdy klikniesz opcję *Arkusz prognozy*, Excel utworzy podgląd wykresu wachlarzowego z kilkoma opcjami (rysunek 8.56).

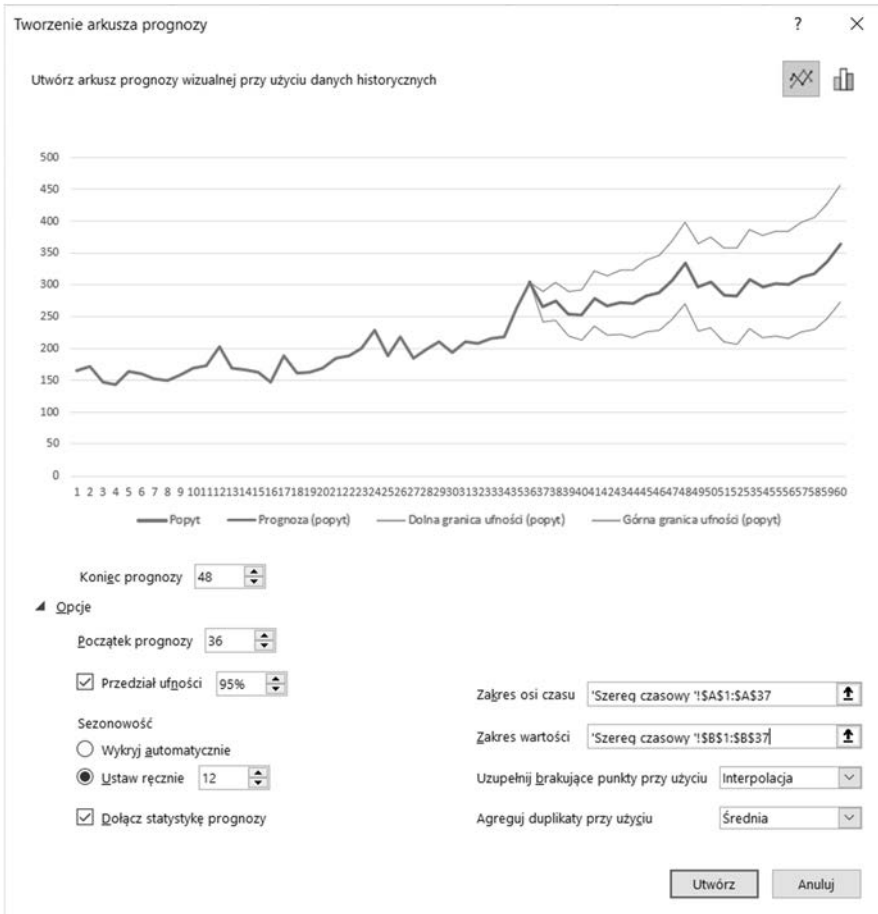
Aby odtworzyć analizę z tego rozdziału, *Koniec prognozy* ustawiłem na 48. Rozwinąłem także listę *Opcje*, by ustawić kilka dodatkowych parametrów. *Sezonowość* ustaw ręcznie na 12 miesięcy i zaznacz opcję *Dołącz statystykę prognozy*, by uzyskać kilka statystyk, które obliczyliśmy ręcznie w tym rozdziale. (Dlaczego ta opcja nie jest domyślnie zaznaczona?). Na koniec naciśnij przycisk *Utwórz*.

Po utworzeniu arkusza (rysunek 8.57) możesz zmienić jego nieelegancką nazwę *Arkusz4* (czy inną domyślną nazwę nadaną przez Excela) na **Arkusz prognozy**.

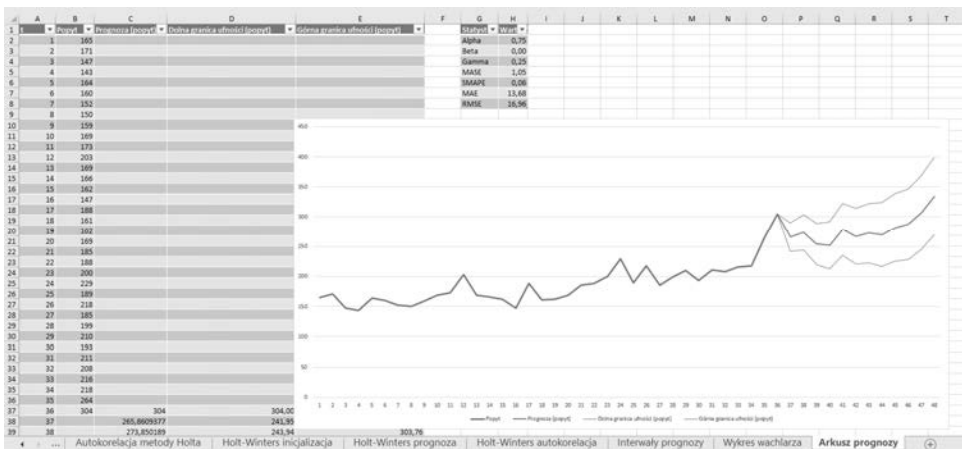
Dzięki temu, czego dowiedziałeś się w tym rozdziale, będziesz w stanie wprowadzić zmiany w tym arkuszu. Tak samo jak ja możesz się zastanawiać, jaki algorytm jest używany do prognozowania. Po przeczytaniu dokumentacji na temat arkusza prognozy już wiem, że to metoda o nazwie **potrójne wygładzanie wykładnicze**. Jak sobie może przypominasz, to po prostu inna nazwa algorytmu Holta-Wintersa.

Jeśli przeczytałeś ten rozdział w całości, to gratuluję. Poważnie. Przedstawiłem Ci naprawdę dużo technik prognozowania.

Mam nadzieję, że czujesz się na siłach, aby iść dalej i „organizować swoją ignorancję”!



Rysunek 8.56. Okno Tworzenie arkusza prognozy



Rysunek 8.57. Gotowy arkusz prognozy

PROGRAM PARTNERSKI

— GRUPY HELION —

- 
1. ZAREJESTRUJ SIĘ
 2. PREZENTUJ KSIĄŻKI
 3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

GRUPA
Helion 

Mimo zainteresowania danymi i ich analizą zbyt wiele projektów z obszaru data science kończy się fiaskiem. Firmy zatrudniają analityków danych, kupują drogie narzędzia, traktują analizę danych jak magię. Niestety, bez dogłębnego zrozumienia sposobu i celu przetwarzania danych nie masz co liczyć na sukces. Jeśli chcesz podjąć się tego wyzwania, Twoim najlepszym sojusznikiem okaże się... Excel, który świetnie się nadaje do wyjaśniania najważniejszych zagadnień nauki o danych.

Analitycy, którzy w pełni wykorzystują potencjał Excela, będą odnosić sukcesy.

— **George Mount**, Excel MVP, autor książki *Zaawansowana analiza danych*

Dzięki tej książce zrozumiesz techniki analizy danych, poznasz szczegóły ich implementacji, a także nauczysz się z nich korzystać w biznesie niezależnie od jego skali. To przewodnik po budowaniu modeli statystycznych, uczeniu maszynowym i użyciu sztucznej inteligencji bezpośrednio w Twoim arkuszu kalkulacyjnym. Jeśli chcesz stosować naukę o danych do odczytywania ukrytej w nich praktycznej wiedzy, to *Mistrz analizy danych* jest lekturą, po którą koniecznie musisz sięgnąć!

Najciekawsze zagadnienia:

- nowe funkcje Excela i Power Query, takie jak X.WYSZUKAJ, LET i LAMBDA
- ogólny model liniowy, metody zespołowe i naiwny klasyfikator Bayesa
- optymalizacja matematyczna, programowanie nieliniowe i algorytmy genetyczne
- szeregi czasowe i generowanie prognoz
- algorytmy k-średnich, modularność grafu i symulacja Monte Carlo
- analiza danych za pomocą języka R

JORDAN GOLDMEIER jest światowej klasy ekspertem w dziedzinie analityki i wizualizacji danych, a także autorem cenionych książek. Od sześciu lat regularnie otrzymuje Excel MVP Award. Był wykładowcą analityki na Wake Forest University, z jego wiedzy korzystały największe firmy świata i NATO.

Jeśli chcesz dołączyć do nowego pokolenia analityków, ta książka będzie Twoim przewodnikiem!
— **dr Alex J. Gutman**, współautor książki *Analityk danych. Przewodnik po data science, statystyce i uczeniu maszynowym*

	KOD KORZYŚCI Sięgnij po więcej! ▶ 
 helion.pl	ISBN 978-83-289-1180-2
 HELION S.A. ul. Kościuszki 1c 44-100 Gliwice tel.: 32 230 98 63 helion@helion.pl	 9 788328 911802
Cena: 99,00 zł	

WILEY