Большие данные – Прогнозирование – часть 1

Контролируемое машинное обучение заключается в предсказании значения либо классификации наблюдений с помощью модели, обученной на данных из прошлого. Прогнозирование при обработке больших данных занимается ровно тем же. Конечно, если вы астролог, то можете «прозревать грядущее» и без данных, но в количественном прогнозировании для предсказания событий в будущем всегда используются данные из прошлого. Некоторые техники, например, множественная регрессия, применяются в обоих случаях.

А вот где они действительно различаются - так это в вопросе канонических задач. Типичные задачи прогнозирования построены на изменении неких данных во времени (продаж, спроса, поставок, ВВП, выбросов углерода или численности населения, к примеру) и проецировании этих изменений на будущее. А с учетом трендов, циклов и периодических вмешательств божьей воли данные в будущем могут существенно отличаться от произошедшего в обозримом прошлом.

В этом и состоит проблема прогнозирования.

Едва вы приходите к выводу, что у вас отличный прогноз спроса на жилье, пузырь недвижимости лопается и ваш прогноз отправляется в унитаз. Тем не менее ни одно бизнес-планирование не обходится без прогноза. Мы делаем попытку спрогнозировать рост наилучшим образом, чтобы иметь возможность спланировать инфраструктуру и кадровое обеспечение. Не собираетесь же вы все время играть в догонялки!

Как вы убедитесь на этом примере, можно не только попытаться предсказать будущее, но и выразить численно неопределенность всего, что связано с прогнозом. Численное выражение неопределенности с помощью создания *интервалов прогнозирования* поистине неоценимо, но часто игнорируется в прогностическом мире.

«Хороший специалист не умнее остальных, просто его невежество лучше организовано».

Так что давайте без лишних вступлений «организуем немного невежества»!

Торговля мечами

На минутку представьте себя ярым фанатом «Властелина Колец». Несколько лет назад, когда на основе трилогии вышел самый первый фильм, вы надели пару искусственных хоб- бичьих ножек и пошли стоять в очереди на первый полуночный показ. Затем вы стали посещать фестивали и бурно обсуждать на форумах, мог ли Фродо просто взять да полететь на орле к Роковой Горе.

В какой-то момент вы решаете внести свой вклад в «хоббитовую реальность». Вы проходите в вузе курс металлообработки и начинаете ковать мечи. Ваш самый любимый меч из романа - Андурил. Вы становитесь экспертом в выковывании этих здоровенных палашей в самодельной кузнице и организуете их продажу на Amazon, eBay и Etsy. Ваши высказывания приобретают популярность у искушенных ботанов, бизнес идет в гору семимильными шагами.

Раньше вам приходилось выискивать заказчиков с собственным материалом, поэтому вы решили спрогнозировать свой будущий спрос и вставить данные о предыдущих заказах в таблицу. Только откуда же взять данные о прошлых продажах, чтобы построить такой прогноз?

В данной примере рассматривается набор прогностических техник под названием *экспоненциальное сглаживание*. На сегодняшний день это одни из простейших техник, имеющие широчайшее применение в бизнесе.

Эта точность частично проистекает из простоты метода - переподгонке зачастую рассеянных данных из прошлого. Более того, с помощью этих методов значительно проще вычислить интервалы прогнозирования *вокруг* экспоненциально сглаженных прогнозов, что вам неизбежно придется делать.

Временная последовательность данных

Электронная таблица для примера содержит данные о торговле мечами за последние 36 месяцев, начиная с января. Данные записаны на листе Timeseries (рис. 1). Как уже отмечалось ранее, данные типа этих - регулярные наблюдения с фиксированными временными интервалами - называются *временными рядами данных*. Временной ряд может быть любым в зависимости от конкретной задачи - идет ли речь об изменении численности населения за год или ежедневной динамике цен на газ.

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

Рис. 1. Временной ряд данных

В нашем случае мы имеем ежемесячный спрос на мечи, и первое, что вам нужно сделать, - это отобразить его графически, как показано на рис. 2. Чтобы вставить график, выделите столбцы A и B в Excel и выберите «Точечную диаграмму» из меню диаграмм Excel (вкладка «Диаграммы» в MacOS и «Вставить диаграмму» в Windows). Длину осей можно настроить, кликнув на них правой клавишей мышки и выбрав «Форматирование». Что же вы видите на рис. 2? Данные варьируются от 140 три года назад до 304 в прошлом месяце. За три года спрос удвоился - может быть, это тренд роста? Мы вернемся к этой мысли чуть позже.



Рис. 2. Диаграмма временного ряда данных

На графике есть несколько пиков и спадов, что может быть признаком некоторой сезонности. В частности, пики спроса приходятся на месяцы с номерами 12, 24 и 36, которые оказываются декабрями. Но может быть это лишь случайность или особенность тренда? Давайте выясним.

Простое экспоненциальное сглаживание

Методы экспоненциального сглаживания основываются на прогнозировании будущего по данным из прошлого, где более новые наблюдения весят больше, чем старые. Такое взвешивание возможно благодаря константам сглаживания. Первый метод экспоненциального сглаживания, который мы опробуем, называется *простым* экспоненциальным сглаживанием (ПЭС, simple exponential smoothing, SES). Он использует лишь одну константу сглаживания.

При простом экспоненциальном сглаживании предполагается, что ваш временной ряд данных состоит из двух компонентов: уровня (или среднего) и некоей погрешности вокруг этого значения. Нет никакого тренда или сезонных колебаний - есть просто уровень, вокруг которого «висит» спрос, тут и там окруженный небольшими погрешностями. Отдавая предпочтение более новым наблюдениям, ПЭС может явиться причиной сдвигов этого уровня. Говоря языком формул,

Спрос в момент времени t = уровень + случайная погрешность около уровня в момент времени t

Самое последнее приближение уровня служит прогнозом на будущие периоды времени. Если вы обсчитываете месяц 36, каким будет хорошее приближение на период 38? Самое последнее приблизительное значение уровня. А на 40? Уровень. Все просто - ведь это простое экспоненциальное сглаживание.

Так как же найти приблизительное значение уровня?

Если принять все временные значения как имеющие одинаковую ценность, то следует просто вычислить их среднее значение.

Это среднее задаст вам уровень, и вы будете предсказывать будущее, говоря себе: «Будущий спрос — это средний спрос прошлого». Некоторые компании поступают именно так. Я видел ежемесячные прогнозы, в которых будущий спрос равнялся среднему за прошедшие месяцы последних нескольких лет. Плюс «поправочный коэффициент» смеха ради. Да-да, прогнозирование зачастую делается настолько криворуко, что даже крупные публичные компании до сих пор используют архаизмы вроде «поправочного коэффициента». Фу!

Но когда уровень изменяется во времени, то нет нужды одинаково оценивать каждую точку прошлого, как при использовании среднего. Должны ли все данные с 2008 по 2013 год весить одинаково, чтобы спрогнозировать 2014-й? Возможно, но для большей части компаний, вероятно, это не так. Так что нам нужно приблизительное значение уровня, которое дает больший вес недавним наблюдениям спроса.

Давайте вместо этого подумаем о расчете уровня, пройдясь по точкам данных по порядку и обновляя его по мере продвижения. Для начала решим, что исходное значение уровня — это среднее от нескольких последних точек данных. В этом случае выберите значения первого года. Назовите их исходным уровнем, *уровень*0:

уровень0 = среднее значение спроса за первый год (месяцы 1-12)

Для спроса на мечи он равен 163.

Теперь о том, как работает экспоненциальное сглаживание. Даже при том, что вы знаете спрос за месяцы с 1 по 36, вам нужно взять компоненты вашего самого последнего прогноза и использовать их для предсказания на месяц вперед с помощью полного ряда данных.

Мы используем уровень0 (163) как прогноз спроса на месяц 1.

Предсказав период 1, вы шагаете вперед во времени от периода 0 к периоду 1. Текущий спрос равен 165, то есть он вырос на 2 меча. Стоит обновить приближение исходного уровня. Уравнение простого экспоненциального сглаживания выглядит так:

уровень1 = уровень O + несколько процентов *x* (спрос1 - уровень O)

Обратите внимание, что (спрос1 - уровень О) - ошибка, которую вы получаете, когда предсказываете период 1 с использованием первоначального уровня. Шаг вперед:

уровень2 = уровень1 + несколько процентов х (спрос2 - уровень1)

И еще один:

yровень3 = уровень2 + несколько процентов x (спрос3 - уровень2)

Теперь выясним, сколько процентов погрешности вы готовы заложить в уровень — это будет *константа сглаживания*, и для уровня ее название - *альфа* (так исторически сложилось). Это может быть любое число от О до 100 % (от О до 1).

Если вы сделаете *альфу* равной 1, то получится, что погрешностью является все, означающее, что уровень текущего периода равен спросу на текущий период.

Если сделать *альфу* равной О, то в первом прогнозе уровня не останется совсем никакой коррекции погрешности.

Наверное, нам нужно что-то между этими двумя крайностями, но выбирать лучшее значение альфы вы научитесь позже.

Пока же проведите это вычисление для разных моментов времени:

уровеньтекущий период = уровеньпредыдущий период + альфа * (спростекущий период - уровень предыдущий период)

В итоге вы имеете финальное приближение уровня - значение *уровень* 36, где последние наблюдения спроса весят больше, потому что их поправки на погрешность не умножались с каждым следующим значением *альфа*:

уровень36 = уровень35 + альфа * (спрос36 - уровень35)

Это итоговое приближение уровня - то, что вы будете использовать как прогноз на будущие месяцы. Спрос за месяц 37? Ну, это просто *уровень*36. А как насчет месяца 40? *Уровень*36. Месяц 45? *Уровень*36. У вас есть картинка. Итоговое приближение уровня - просто лучшее из тех, что у вас есть на будущее, так что его вы и используете.

Давайте взглянем на него в таблице.

Настраиваем прогноз простого экспоненциального сглаживания

Первое, что нужно сделать, — это создать новый лист в документе под названием SES. Вставьте временной ряд данных в столбцы А и В, начиная со строки 4, чтобы оставить немного места наверху для значений *альфа*. Вы можете ввести количество месяцев, которое у вас есть (36) в ячейку А2, а исходное значение *альфа* - в C2. Я начну с O,5, потому что это число между O и 1 и оно мне нравится.

Поместите расчет уровня в столбец С. Нужно вставить новую строку 5 во временной ряд данных вверху таблицы для исходного приближения уровня в момент времени О. В С5 используйте следующий расчет:

=AVERAGE (B6:B17)

=CP3HA4 (B6:B17)

Эти средние значения данных за первый год дают исходный уровень. Таблица с ними выглядит так, как показано на рис. 3.

		- Search in St	het	20
•	Home Layout	Tables C	Jharts	٥.
	C5 : 1	30 - fx	=AVERAGE(86:817)	
12	A	B	C	4
1	Total Months		Level smoothing parameter (alpha)	
2	36		0.50	
3				
4	t	Demand	Level Estimate	
5	0		163	
6	1	165		
7	2	171		9
8	3	147		
9	4	143		
10	5	164		
11	6	160		
12	7	152		а
13	8	150		
14	9	159		1
15	10	169		1
16	11	173		
17	12	203		
18	13	169		1
19	14	166		
20	15	162	1 (and)	

Рис. 3. Приблизительное значение исходного уровня для простого экспоненциального сглаживания

Добавление одношагового прогноза и погрешности

Теперь, когда вы добавили в последовательность значения первого уровня, можно идти вперед во времени с помощью следующей формулы - для этого добавьте еще два столбца: столбец одношагового прогноза (D) и погрешности прогноза (E). Одношаговый прогноз за период времени 1 — это уровень0 (ячейка C5), а расчет погрешности — это разница текущего спроса и прогноза:

=B6-D6

Приблизительное значение уровня за период 1 — это предыдущий уровень, с поправкой на *альфу*, умноженное на погрешность, то есть:

=C5+C\$2*E6

Обратите внимание, что у меня перед *альфой* стоит значок \$, поэтому при распространении формулы на всю таблицу абсолютные ссылки на строку позволяют *альфе* оставаться собой. Таким образом получается лист, изображенный на рис. 4.

00		VOA	SwordForecasting.xlsm		
•	Home Layout	Tables (Charts SmartArt I	Formulas Dat	ta 35 ~ 10
18	A	B	C	D	E
1	Total Months		Level smoothing parameter (alpha)		
2	36		0.50		/ · · · · · · · · · · · · · · · · · · ·
3					
4	t	Demand	Level Estimate	One-step Forecast	Forecast Error
5	0		163		
6	1	165	164	163	2
7	2	171			
8	3	147			
9	4	143			
10	5	164			2
11	6	160			
17	Normal View	Timeseries	s (+)		0

Рис. 4. Генерация одношагового прогноза, погрешности и расчета уровня на период 1

Растяните

Сделано уже практически все. Просто растяните С6:Е6 вниз на все 36 месяцев и - все! - у вас есть уровень.

Добавим месяцы 37-48 в столбец А. Прогноз на следующие 12 месяцев — это просто *уровень* 36. Так что в В42 можно добавить только

=C\$41

в качестве прогноза и оттянуть его на весь следующий год.

Таким образом получается прогноз, равный 272, как показано на рис. 5.

Но неужели нельзя сделать что-то получше? Сейчас наш способ оптимизации прогноза - установка значений *альфы*: чем *альфа* больше, тем меньше нам важны старые данные о спросе.

Оптимизация одношаговой погрешности

Точно так же, как при оптимизации суммы квадратов отклонений во время вычисления регрессии, можно найти лучшую константу сглаживания для прогноза путем минимизации суммы квадратов отклонений (погрешностей) для прогнозирования предстоящего периода.

1		26 4D 10	🕼 🖾 • (Q+ (Seat	ch in Sheet		h
•	Home Layout	Tables C	Charts SmartArt F	ormulas Dat	a >> 🗸 -	¢
_	853 \$ (3 © (= fx	=C\$41		-	1
-	A	В	L	D	E	H
1	Total Months		parameter (alpha)			
2	36		0.50			
3						
				One-step	Forecast	1
4	t	Demand	Level Estimate	Forecast	Error	
38	33	216	211.1855079	206.371016	9.62898412	
39	34	218	214.592754	211.185508	6.81449206	
40	35	264	239.296377	214.592754	49.407246	
41	36	304	271.6481885	239.296377	64.703623	U
42	37	271.64819				0
43	38	271.64819				
44	39	271.64819				
45	40	271.64819				
46	41	271.64819				
47	42	271.64819				
48	43	271.64819				
49	44	271.64819				ľ
50	45	271.64819				
51	46	271.64819				
52	47	271.64819				0
53	48	271.54819				

Рис. 5. Прогноз простого экспоненциального сглаживания с альфой, равной 0,5

Добавим расчет квадрата отклонения в столбец F — это просто значение из столбца E, возведенное в квадрат, - и растянем этот расчет на все 36 месяцев, а затем сложим их все в ячейке E 2, чтобы получить сумму квадратов отклонений (СКО). Таким образом у нас получается лист, изображенный на рис. 6.

Также в нашу таблицу нужно добавить само *стандартное отклонение*- в ячейку F2. Стандартное отклонение — это квадратный корень из СКО, разделенный на 35 (36 месяцев за вычетом количества сглаживающих параметров модели, которое для простого экспоненциального сглаживания равно 1).

Стандартное отклонение — это приблизительное значение стандартного распределения для одношаговой погрешности (отклонения). Это просто оценка распределения погрешности.

Если ваша прогностическая модель отлично подогнана, то среднее значение погрешности будет 0. Это, скажем так, *непредвзятый* прогноз. Он настолько же преувеличивает спрос, насколько и преуменьшает. Стандартное отклонение численно выражает распределение вокруг 0, если прогноз непредвзятый.

В ячейке F2 стандартное отклонение вычисляется как:

=SQRT (E2/(36-1)

=КОРЕНЬ (Е2/(36-1)

Для *альфы*, равной 0,5, оно получается равным 20,94 (рис. 7). 68 % погрешностей одношагового прогноза должны оказаться меньше 20,94 и больше -20,94.

Теперь ваша задача - сжать это распределение насколько возможно, найдя подходящее значение *альфы*. Попробуйте несколько разных значений, используя «Поиск решения» в который раз в этой книге.

Настройка «Поиска решения» для этой операции чрезвычайно проста. Откройте опцию, установите целевую функцию - стандартное отклонение в ячейке F2, и переменную решения - *альфа* в C2, добавьте ограничение, что C2 должна быть меньше 1, и поставьте галочку в поле, указывающем, что решение должно быть неотрицательным. Рекурсивные расчеты уровня, входящие в каждый расчет погрешности прогноза, очень нелинейны, поэтому для оптимизации *альфы* нам придется воспользоваться эволюционным алгоритмом.

•	Home Layout	Tables (Charts Sm	artArt F	ormulas Dat	ta Review	>> ~ <
	MIN ‡ €	3 0 (- fx	=SUM(F6:F41)				
1	I A	В	C		D	E	F
1	Total Months		Level smo parameter	othing r (alpha)		SSE	
2	36			0.50		=SUM(F6:F4:	1)
3	1						
4	t	Demand	Level Estin	nate	One-step Forecast	Forecast Error	Squared Error
5	0			163	1		
6	1	165		164	163	2	4
7	2	171		167.5	164	7	49
8	3	147		157.25	167.5	-20.5	420.25
9	4	143		150.125	157.25	-14.25	203.0625
10	5	164		157.0625	150.125	13.875	192.515625
11	6	160	1	58.53125	157.0625	2.9375	8.62890625
12	7	152	15	5.265625	158.53125	-6.53125	42.6572266
13	8	150	152	.6328125	155.265625	-5.265625	27.7268066
14	9	159	155	.8164063	152.632813	6.3671875	40.5410767
15	10	160	162	4082031	155 816406	13 1835038	173 807144

Рис. 6. Сумма квадратов отклонений для простого экспоненциального сглаживания

	0				SwordForecast	ing.xlsm			
21	To Da		200%	5 5	2 - ES - Q	· Q+ S	earch i	n Sheet	
•	Home La	yout	Tables (Charts	SmartArt	Formulas	Data	Review	>> ~ \$
	F2	: 6	3 © (= fx	=SQRT(E	2/(36-1))				
1	A		В		C	D		E	F
1	Total Mor	nths		Level : param	smoothing neter (alpha)		SS	E	Standard Error
2		36			0.5	D		15346.86	20.94
3						1.0			
4	t		Demand	Level	Estimate	One-step Forecast	Fo	ror	Squared Error
5		0			16	3			
6		1	165		16	4 10	63	2	4
7		2	171		167.	5 10	64	7	49
8		3	147		157.2	5 167	.5	-20.5	420.25
9		4	143		150.12	5 157.3	25	-14.25	203.0625
10		5	164		157.062	5 150.13	25	13.875	192.515625
11		6	160		158.5312	5 157.06	25	2.9375	8.62890625
12	114.4.4	7	152 Timeseries	s [+]	155 26562	158 531	25	-6 53125	42 6572266
	Normal V	New	Ready	-					

Рис. 7. Расчет стандартного отклонения

Настройка «Поиска решения» должна выглядеть так, как показано на рис. 8. Нажав «Выполнить», вы получаете значение *альфы*, равное 0,73, что дает нам новое стандартное отклонение 20,39. Не такое уж кардинальное улучшение!

00 So	Iver Parameters	
Set Objective: SFS2		ī.
To: 🔾 Max 💿 Min	O Value Of:)
By Changing Variable Cells	s:	
SCS2		
Subject to the Constraints	:	
\$C\$2 <= 1		Add
		Change
		Delete
		Reset All
		Load/Save
Make Unconstrained V	ariables Non-Negative	
Select a Solving Method:	Evolutionary 👻	Options
Solving Method		
Select the GRG Nonlinear el nonlinear. Select the LP Sim and select the Evolutionary smooth.	ngine for Solver Problems splex engine for linear Sol engine for Solver problem	that are smooth ver Problems, ns that are non-
	Card and and	and the second se

Рис. 8. Настройка «Поиска решения» для оптимизации альфы

А теперь графика!

Лучший способ проверить наш прогноз — это нарисовать его диаграмму рядом с графиком данных о спросе в прошлом и посмотреть, как спрогнозированный спрос отличается от реального. Мне нравится вид простой диаграммы Excel. Для начала выберите A6:B41, где расположены наши исторические данные, и простую линейную диаграмму из вариантов, предложенных Excel.

Добавив график, кликните на нем правой кнопкой мыши, выберите «Выбрать данные» и вставьте туда новую серию предсказанных значений A42:B53. Если хотите, добавьте также названия осей, после чего у вас получится что-то похожее на рис. 9.



Рис. 9. Диаграмма итогового прогноза простого экспоненциального сглаживания

Возможно, у вас есть тренд

Достаточно беглого взгляда на этот график, чтобы понять несколько вещей. Вопервых, простое экспоненциальное сглаживание — это прямая линия, уровень. Но если посмотреть на данные о спросе за прошедшие 36 месяцев, то окажется, что он поднялся. Похоже на восходящий тренд, особенно в конце.

Наши глаза, определенно, не врут, но все же как это доказать?

Для доказательства достаточно подогнать линейную регрессию под данные спроса и выполнить тест на соответствие критерию Стьюдента на подъеме этой линии тренда.

Если уклон линии ненулевой и статистически значимый (в проверке по критерию Стьюдента величина p менее 0,05), то можете быть уверены: у данных есть тренд. Если в последнем предложении вы не видите никакого смысла, загляните в раздел главы 6 о статистических тестах.

Переключитесь во вкладку Timeseries вашей электронной таблицы для выполнения проверки тренда.

В этой главе для подгонки линейной регрессии, а также для проверки уклона, расчета стандартного отклонения коэффициента уклона и количества степеней свободы (если не понимаете эти термины - загляните в главу 6) мы будем использовать функцию LINEST/ ЛИНЕЙН, встроенную в Excel. Затем можно будет рассчитать статистику Стьюдента и пропустить ее через функцию TDIST/CTьюдРАСП.

Если вы никогда раньше не пользовались функцией LINEST/ЛИНЕЙН, ознакомьтесь со справочным материалом Excel по этой функции. Вы задаете LINEST/ЛИНЕЙН данные зависимых переменных (спроса в столбце В) и независимых (у вас только одна независимая переменная, и это - время в столбце А).

Отметив «Истина», вы даете функции понять, что свободный член надо подгонять как часть линии регрессии, и еще раз - чтобы получить детализированный отчет по стандартному отклонению и R-квадрату. После этого линейную регрессию на листе Timeseries можно запустить следующим образом:

=LINEST (B2:B37, A2:A37, TRUE, TRUE)

=ЛИНЕЙН (B2:B37, A2:A37, ИСТИНА, ИСТИНА)

Эта формула вернет уклон линии регрессии, потому что LINEST/ЛИНЕЙН - формула массива. LINEST/ЛИНЕЙН возвращает всю статистику регрессии в массиве, чтобы вытряхнуть все содержимое в выбранную область таблицы. Вы также можете воспользоваться формулой INDEX/ИНДЕКС для LINEST/ЛИНЕЙН, чтобы вытащить только интересующие вас значения одно за другим.

Например, первые компоненты линии регрессии, которые выдает LINEST/ЛИНЕЙН, это коэффициенты регрессии, так что можно рассчитать уклон регрессии в ячейке B39 таблицы Timeseries, пропуская LINEST/ЛИНЕЙН через INDEX/ИНДЕКС:

=INDEX(LINEST(B2:B37,A2:A37,TRUE,TRUE),1,1)

=ИНДЕКС (ЛИНЕЙН (В2:В37, А2:А37, ИСТИНА, ИСТИНА), 1, 1)

Получается уклон, равный 2,54, что означает тренд линии регрессии к повышению спроса на 2,54 дополнительных меча в месяц. То есть уклон действительно есть. Но значим ли он статистически?

Чтобы проверить уклон критерием Стьюдента, нужно узнать стандартное отклонение уклона и количество степеней свободы регрессии. LINEST/ЛИНЕЙН помещает значения стандартного отклонения в строку 2, столбец 1 массива результатов. К примеру, в В40 вы можете это написать как:

=INDEX(LINEST(B2:B37,A2:A37,TRUE,TRUE),2,1)

=ИНДЕКС (ЛИНЕЙН (В2:В37, А2:А37, ИСТИНА, ИСТИНА), 2, 1)

Единственное отличие от вычисления уклона заключается в том, что вместо ряда 1 столбца 1 мы задаем формуле INDEX/ИНДЕКС строку 2 столбец 1 и находим стандартное отклонение.

Стандартное отклонение уклона получается равным 0,34. Таблица на этом этапе показана на рис. 10.

		ET (MER) SCA	HER IN SHEET		10
	Home Layout Tables	Charts Sm	artArt For	mulas 33 🛩	0
	139 1 3 0	fx =INDEX(LINES)	T(82:837,A2:A	37, TRUE, TRUE), 1.1	
2	A	8	С	D	
29	28	199			
30	29	210			
51	30	193			
52	31	211			
55	32	208			
54	33	216			_
52	34	218			
50	35	264			-1
\$1	36	304			
38					
39	Slope	2.54			
40	Slope Standard Error	0.34			
	and the state of the second				-1

Рис. 10. Уклон и стандартное отклонение линии регрессии, подогнанное под исторический спрос

Точно так же функция LINEST/ЛИНЕЙН знает, что степени свободы регрессии помещены в строку 4 столбца 2 итогового массива. Поэтому в В41 их можно вычислить следующим образом:

=INDEX(LINEST(B2:B37,A2:A37,TRUE,TRUE),4,2)

=ИНДЕКС (ЛИНЕЙН (В2:В37, А2:А37, ИСТИНА, ИСТИНА), 4, 2)

У вас должно получиться 34 степени свободы (как отмечено в главе 6, они рассчитываются как 36 точек данных минус 2 коэффициента линейной регрессии).

Теперь известны все три значения для проверки критерием Стьюдента статистической значимости вашего подогнанного тренда. Вычислим статистику этого критерия как абсолютную величину уклона, разделенную на стандартное отклонение уклона. Рассчитайте величину p для этой статистики из распределения Стьюдента с 34 степенями свободы, воспользовавшись функцией TDIST/CTьЮДРАСП в B42:

=TDIST(ABS(B39/B40), B41, 2)

=СТЬЮДРАСП (ABS (B39/B40), B41, 2)

Функция выдает значение *p*, близкое к 0, намекая на то, что если бы тренд был не существующим в реальности (уклон 0), то ни при каких условиях мы бы не получили настолько отличающийся от регрессии уклон. Это показано на рис. 11.

	00	002	SwordFor	ecasting.xlsr	m		
21	•		Q+ 50	arch in Sheet			
*	Home	Layout	Tables	Charts	SmartArt	33 V	٥,
	B42	11	00	fx =TDIST(A	85(839/840),841,2)	
-	1	A		6	(2	-
32	-		31	21	1		
33			32	20	8		
34			33	21	.6		
35			34	21	.8		
36			35	26	54		
37			36	30	04		
38							
39	Slope			2.5	i4		18
40	Slope	Standard	Error	0.3	84		
41	Degre	es of Free	edom	3	34		- 10
42	t Test	p value		1.17E-0	8		
12		A	Timesories	(ES) A	1	-	
		mail Minut	Paadu	AL IT			111

Рис. 11. Ваш тренд существует!

Есть тренд. Осталось включить его в прогноз.

Экспоненциальное сглаживание Холта с корректировкой тренда

Экспоненциальное сглаживание Холта с корректировкой тренда распространяет простое экспоненциальное сглаживание на создание прогноза из данных, имеющих линейный тренд. Часто оно называется двойным экспоненциальным сглаживанием, потому что в отличие от простого ЭС, имеющего один параметр сглаживания - *альфа*, и один компонент, не являющийся отклонением, двойное экспоненциальное сглаживание имеет их два.

Если у временной последовательности линейный тренд, то его можно записать так:

спрос за время t = уровень+ t * тренд + случайное отклонение уровня в момент времени t

Самое последнее приближение уровня и тренда (умноженные на количество периодов) служит прогнозом на грядущие периоды времени. Если сейчас месяц 36, то каким будет хорошее приближение спроса на период 38? Приближение последнего периода времени *плюс* два месяца тренда. Не настолько просто, как ПЭС, но близко.

Теперь, как и в простом экспоненциальном сглаживании, вам нужно получить несколько исходных приближений значений уровня и тренда, которые мы обозначим *уровень*0 и *тренд*0. Один из простейших способов их найти - это просто изобразить на графике первую половину ваших данных о спросе и определить линию тренда (как мы делали в главе 6 в примере с аллергией на кошек). Уклон этой линии - *тренд*0, а *у*-начальный отрезок (свободный член) - *уровень*0.

Экспоненциальное сглаживание Холта с корректировкой тренда имеет два новых уравнения, одно - для уровня по мере его продвижения во времени, а другое - собственно тренд. Уравнение уровня также содержит сглаживающий параметр под названием *альфа*, а в уравнении тренда используется параметр, часто называемый *гамма*. Они абсолютно одинаковые - просто значения от 0 до 1, регулирующие вмешательство погрешности одношагового прогноза в дальнейшие расчеты приближений.

Вот как выглядит новое уравнение уровня:

уровень 1 = уровень о + тренд о + альфа * (спрос 1 - (уровень о + тренд о))

Обратите внимание, что *уровень*0 + *тренд*0 — это просто одношаговый прогноз от исходных значений к месяцу 1, поэтому (*спрос*₁ - (*уровень* • + *тренд* •)) - это одношаговое отклонение. Это уравнение очень похоже на уравнение уровня из ПЭС, за исключением того, что считается значение тренда за один период, когда бы вы ни рассчитывали следующий шаг. Таким образом, основное уравнение приближения уровня будет следующим:

уровень текущий период= уровень предыдущий период + тренд предыдущий период + альфа * (спрос текущий период - (уровень предыдущий период + тренд предыдущий период))

Для этой новой техники сглаживания нам понадобится новое уравнение обновления тренда. Для первого шага оно будет таким:

тренд1= тренд о + гамма * альфа * (спрос1 - (уровеньо + трендо))

То есть уравнение тренда очень похоже на уравнение обновления уровня. Берется предыдущее приближение тренда и изменяется на *гамму*, умноженную на размер отклонения, заложенного в соответствующее обновление уровня (которое имеет интуитивный смысл, потому что только некоторые из отклонений, используемых для изменения уровня,

можно отнести к небольшим или скачкообразным приближениям тренда).

Таким образом основное уравнение приближения тренда такое:

Тренд текущий период = тренд предыдущий период + гамма * альфа * (спрос текущий период - (уровень предыдущий период + тренд предыдущий период))

Настройка холтовского сглаживания с коррекцией тренда в электронной таблице

Для начала создайте новый лист под названием Holt's Trend-Corrected. Затем, как и в случае с таблицей простого экспоненциального сглаживания, скопируйте в строку 4 временной ряд данных и вставьте пустую строку 5 для исходных приближений.

Столбец С снова будет содержать приближения уровня, а столбец D - приближения тренда. Наверху в этих столбцах находятся значения *альфы* и *гаммы*. Вы оптимизируете их «Поиском решения» за секунду, но сейчас пусть они будут около 0,5. Так получается таблица, изображенная на рис. 12.

*	Home Layout	Tables 6	harts Sesart	Art 33 Y
		R R	0.5	D
1	Total months		Level smoothing parameter (alpha)	Trend smoothing parameter (gamma)
2	36		0.5	0.5
3		S		
4	t	Actual Demand	Level	Trend
5	0			
6	1	165		
7	2	171		
8	3	147		
9	4	143		
10	5	164		
11	6	160		
12	7	152		

Рис. 12. Начинаем с параметрами сглаживания 0,5

А для исходных значений уровня и тренда в C5 и D5 мы построим график за первые 18 месяцев и добавим в него линию тренда с уравнением. Так мы получаем исходное значение

тренда 0,8369 и исходный уровень (свободный член линии тренда), равный 155,88. Добавив данные в D5 и C5 соответственно, получаем лист, изображенный на рис. 13.



Рис. 13. Исходные значения уровня и тренда

Теперь добавьте в столбцы Е и F столбцы прогноза и погрешности прогноза на один шаг вперед. Если посмотреть на строку 6, одношаговый прогноз получается примерно равным предыдущему уровню, к которому прибавлен тренд за один месяц с помощью предыдущего приближения — это C5+D5. Погрешность прогноза вычисляется так же, как и в случае простого экспоненциального сглаживания: F6 - это просто текущий спрос, из которого вычтен одношаговый прогноз - B6-E6.

Теперь вы можете обновить значение уровня в ячейке C6 - оно будет равно предыдущему значению, к которому прибавлен предыдущий тренд и *альфа*, умноженная на погрешность:

=C5+D5+C\$2*F6

Тренд в D6 обновляется точно так же, только нужно еще прибавить *гамму*, умноженную на *альфу*, умноженную на погрешность:

=D5+D\$2*C\$2*F6

Обратите внимание на следующую деталь: необходимо использовать абсолютные ссылки на *альфу* и *гамму*, чтобы была возможность растянуть формулу на целые столбцы. Сделайте это прямо сейчас - растяните C6:F6 вниз до месяца 36. Операция показана на рис. 14.

•	Home L	ayout T	ables Charts	SmartArt	Formulas D	ata >> ~ 3
	C41	: 00	fx =C40	+D40+C\$2*F41	c	E I
1	Total	D	Level smoothing parameter (alpha)	Trend smoothing parameter (gamma)	L	
2	36		0.5	0.5		
3		Actual			One step	Foreset
4	t	Demand	Level	Trend	Forecast	Error
35	30	193	197.585842	-2.65641058	202.171684	-9.1716838
36	31	211	202.964716	1.361231594	194.929431	16.0705687
37	32	208	206.162974	2.279744781	204.325947	3.67405275
38	33	216	212.221359	4.169065179	208.442718	7.55728159
39	34	218	217.195212	4.571459083	216.390424	1.60957562
40	35	264	242.883336	15.12979126	221.766671	42.2333287
41	36	304	281.006563	26.62650954	258.013127	45,9868731

Рис. 14. Растягивание расчетов уровня, тренда, прогноза и погрешности

Прогнозируем будущее

Чтобы прогнозировать дальше, чем на 36 месяцев, добавьте итоговый уровень (который для *альфы* и *гаммы*, равных 0,5, будет равняться 281) к количеству месяцев после 36го, на которые вы хотите рассчитать прогноз, *умноженному* на итоговое приближение тренда. Вы можете подсчитать количество месяцев от 36 до нужного вам, вычтя один месяц в столбце А из другого.

К примеру, для прогноза месяца 37 в ячейке В42 вы будете пользоваться такой формулой:

=C\$41+(A42-A\$41)*D\$41

С помощью абсолютных ссылок на месяц 36, итоговый тренд и итоговый уровень, прогноз можно растянуть вниз до месяца 48, что показано на рис. 15.

*	Home L	ayout Ta	ables Charts	SmartArt	Formulas D	iata 33 🗸	¢
	MIN	:00	C 1x =C34	1+(A53-A541)*05	41 F	F	r
1	Total months		Level smoothing parameter (alpha)	Trend smoothing parameter (gamma)			
2	36		0.5	0.5			
4	t	Actual Demand	Level	Trend	One-step Forecast	Forecast Error	
38	33	216	212.221359	4.169065179	208.442718	7.55728159	T
39	34	218	217.195212	4.571459083	216.390424	1.60957562	
40	35	264	242.883336	15.12979126	221.766671	42.2333287	
41	36	304	281.006563	26.62650954	258.013127	45.9868731	
42	37	307.633					
43	38	334.26					
44	39	360.885					
45	40	387.513					
46	41	414.139					
47	42	440.766					
48	43	467.392					
49	44	494.019					
50	45	520.645					
51	46	547.272					
52	47	573.898					
53	48	D\$41	6				

Рис. 15. Прогнозы на будущие месяцы, сделанные методом экспоненциального сглаживания Холта с коррекцией тренда

Как и в таблице простого экспоненциального сглаживания, здесь можно построить график исторического спроса и прогноза как двух серий данных на одной простой координатной плоскости, как показано на рис. 16.

При альфе и гамме, равных 0,5, прогноз, конечно, выглядит немного странно. Начиная с конца последнего месяца он внезапно возрастает, гораздо быстрее, чем ранее. Похоже, стоит оптимизировать параметры сглаживания.



Рис. 16. График прогноза со значениями альфа и гамма по умолчанию

Оптимизация одношаговой погрешности

Точно так же, как и в случае с простым экспоненциальным сглаживанием, здесь нужно добавить квадрат погрешности прогноза в столбец G. B F2 и G2 можно рассчитать сумму квадратов погрешностей (отклонений) и стандартное отклонение для одношагового прогноза, в точности как и раньше. За исключением того, что в этот раз у модели два параметра сглаживания, так что придется разделить эту сумму на 36-2 перед извлечением из нее корня:

=SQRT(F2/(36-2)) =КОРЕНЬ(F2/(36-2))

Так получается лист, изображенный на рис. 17.

Настройки оптимизации идентичны таковым в простом экспоненциальном сглаживании. Единственное отличие - в этот раз вы оптимизируете оба параметра, *альфу* и *гамму*, вместе, как показано на рис. 18.

После запуска «Поиска решения» вы получите оптимальное значение *альфы*, равное 0,66, и *саммы*, равное 0,05. Оптимальный прогноз показан на графике на рис. 19.

•	Home L	ayout T	ables Charts	SmartArt	Formulas D	ata Review	>> ~ \$
	A	B	C	D	E	F	G
1	Total months		Level smoothing parameter (alpha)	Trend smoothing parameter (gamma)		SSE	Standard Error
2	36		0.5	0.5		15315.3154	21.2238181
4	t	Actual Demand	Level	Trend	One-step Forecast	Forecast Error	Squared Error
5	0		155.88	0.8369			
6	1	165	160.85845	2.907675	156.7169	8.2831	68.6097456
7	2	171	167.383063	4.71614375	163.766125	7.233875	52.3289475
8	3	147	159.549603	-1.55865781	172.099206	-25.099206	629.970154
9	4	143	150.495473	-5.30639414	157.990945	-14.990945	224.728441
10	5	164	154.594539	-0.60366377	145.189079	18.8109215	353.850767
11	6	160	156.995438	0.898617358	153.990875	6.00912451	36.1095774
12	7	152	154.947028	-0.57489642	157.894055	-5.8940551	34.7398856
13	8	150	152.186066	-1.6679292	154.372131	-4.3721311	19.1155307
1.4	0	150	154 750069	0 453536700	150 519126	9 49196363	71 0420107

Рис. 17. Расчет суммы квадрата погрешности и стандартного отклонения

0.0	So	lver Parameter	15	_
Set Objective:	\$G\$2			
To: Max	Min Min Cells	Value Of:	0	
\$C\$2:\$D\$2				T
Subject to the Co	nstraints			
\$C\$2:\$D\$2 <=	1			Add
				Change
			I	Delete
			1	Reset All
				Load/Save
Make Unconst	trained V	ariables Non-Ne	gative	
Select a Solving M	lethod:	Evolutionary	-	Options
Solving Method Select the GRG No nonlinear. Select the and select the Evo smooth.	nlinear ei he LP Sim lutionary	ngine for Solver Pr plex engine for li engine for Solver	oblems tha near Solver problems t	t are smooth Problems, hat are non-
		_		

Рис. 18. Настройки оптимизации для экспоненциального сглаживания Холта с корректировкой тренда



Рис. 19. График оптимального холтовского прогноза

Тренд, который вытекает из вашего прогноза, — это пять дополнительных мечей, проданных за месяц. Этот тренд дублирует уже найденный вами ранее в предыдущей таблице, потому что сглаживание с корректировкой тренда ценит более поздние данные выше, и в этом случае спрос за недавнее время оказывается более «трендовым».

Обратите внимание, что этот прогноз начинается очень близко к прогнозу ПЭС для месяца 37-290 против 292. Но довольно быстро прогноз с корректировкой тренда начинает расти вместе с трендом, как вы и предполагали.

Так мы закончили? Взгляд на автокорреляции

Все ли возможное мы сделали? Все ли учли?

Есть один способ испытать вашу прогностическую модель на прочность - проверить погрешности одного шага вперед. Если эти отклонения случайны, то вы хорошо поработали. Но если найдется спрятанная в них закономерность - что-то вроде повторяющегося поведения на регулярных интервалах - возможно, в данных о спросе есть сезонный фактор, который мы не учли.

Под закономерностью я понимаю следующее: если взять погрешность и поместить ее рядом с ней же, но взятой за следующий месяц, или через два месяца, или даже 12 - будет ли она меняться синхронно?

Так что начнем с создания нового листа под названием Holt's Autocorrelation. Вставьте туда месяцы с 1 по 36 вместе с их одношаговыми погрешностями из холтовского прогноза - в столбцы А и В.

Под погрешностями в В38 рассчитайте среднюю погрешность. Получившийся лист показан на рис. 20.

•	Home Layo	Tab	les	Charts	33 ~	¢.,
	838 3	00	= fx	=AVE	RACE(82:837)	
2	A	-	E		C	
1	t	One-st	ep er	or		
26	2	-32.2	37607	21		
27	2	16.1	32593	18		
28	2	-29.9	42797	72		1
29	2	2.39	83816	525		
30	2	10.3	39524	42		
31	3	-15	5.3153	321		1
32	3	11.4	75122	271		
33	3	-0.79	37522	33		
34	3	6.05	15768	331		
35	3	2.17	32872	29		
36	3	44.7	75092	259		
37	3	51.7	30497	18		
38		3.57	64060	42		

Рис. 20. Месяцы и ассоциированные с ними одношаговые погрешности

В столбце С рассчитайте отклонения каждой погрешности в столбце В от среднего в В38. Эти отклонения одношаговых погрешностей от средней - то «пространство», где закономерности могут поднимать свои страшненькие головы. К примеру, возможно, что каждый декабрь погрешность прогноза значительно больше средней - и такая сезонная закономерность обнаружится в этих цифрах.

Итак, в C2 рассчитаем отклонение погрешности в B2 от средней: =B2-B\$38

Затем можно растянуть эту формулу вниз, чтобы получить все отклонения. В ячейке С38 рассчитаем сумму квадратов отклонений:

=SUMPRODUCT (\$C2:\$C37, C2:C37)

=СУММПРОИЗВ(\$C2:\$C37,C2:C37)

Так получается таблица, изображенная на рис. 21.

-			an can water a	
•	Home Layou	t Tables Chu	irts SmartArt 30 🗸	0
-	C38 ;	00 - Jx -5	UMPRODUCT(\$C2:\$C37,C2:C37)	
-	A	B	De latin de	
1	t	Une-step error	Deviations from mean	
27	26	16.13259318	12.55618714	-
28	27	-29.94279772	-33.51920376	
29	28	2.398381625	-1.178024417	
30	29	10.33952442	6.763118375	
31	30	-15,315321	-18.89172704	
32	31	11.47512271	7,898716673	
33	32	-0.793752233	-4.370158274	
34	33	6.051576831	2.47517079	
35	34	2.173287229	-1.403118813	
36	35	44.77509259	41.19868655	
37	36	51.73049718	48.15409114	
38		3.576406042	13636.81634	

Рис. 21. Сумма квадратов отклонений от средней погрешности холтовского прогноза

Теперь поместим в столбец D отклонения погрешностей со сдвигом на месяц. Назовите столбец D «1». Ячейку D2 можете оставить пустой, а в D3 поместить

=C2

А теперь просто растяните эту формулу до D37, оказавшегося равным C36. Таким образом получается таблица, изображенная на рис. 22.

*	Home Layou	Tables Cha	arts SmartArt Formul	as D V š
	MIN ‡	③ ⊙ (= fx =0	36	
2	A	В	С	D
1	t	One-step error	Deviations from mean	1
26	25	-32.23760721	-35.81401325	28.0123854
27	26	16.13259318	12.55618714	-35.814013
28	27	-29.94279772	-33.51920376	12.5561871
29	28	2.398381625	-1.178024417	-33.519204
30	29	10.33952442	6.763118375	-1.1780244
31	30	-15.315321	-18.89172704	6.76311838
32	31	11.47512271	7.898716673	-18.891727
33	32	-0.793752233	-4.370158274	7.89871667
34	33	6.051576831	2.47517079	-4.3701583
35	34	2.173287229	-1.403118813	2.47517079
36	35	44.77509259	41.19868655	-1.4031188
37	36	51.73049718	48.15409114	=C36
38		3.576406042	13636.81634	



Чтобы сдвинуть отклонения на 2 месяца, просто выделите D1:D37 и перетяните в столбец Е. Точно так же можно сдвинуть погрешности на 12 месяцев - просто протащить выделенную область до столбца О. Элементарно! У вас получилась каскадная матрица сдвинутых отклонений погрешности, как показано на рис. 23.

Теперь, имея эти сдвиги, подумайте о том, что может значить для одного из этих столбцов «синхронное движение» со столбцом С. Возьмем, к примеру, сдвиг на 1 месяц в столбце D. Если эти два столбца синхронны, то число, отрицательное в одном из них, должно быть отрицательно и в другом. Собственно, и положительное должно быть таковым в обоих столбцах. Это означает, что произведение двух столбцов в результате будет иметь много положительных значений (отрицательное значение, умноженное на отрицательное, дает положительное, как и положительное, умноженное на положительное).

16		H # 10	5	0 🛃	50.	121.	2.1	10 *	Q.+	Search	n Sheet			
•	Home	Layout T	ables	Chart	ts S	martAn	Fo	rmulas	Dat	a B	leview	Deve	eloper	4
	037	: 0.0	2 - 1	x =N3	6	_		_	_					_
		C	D	E	F	G	н	1	1	K	L	M	N	0
1 [Deviatio	ns from mean	1	2	3	4	5	6	7	8	9	10	11	12
2		4.71												
5		4.12	4.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4		-20.35	4.12	4./1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		-15.94	-20.35	4.12	4./1	4.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9		13.04	-13.94	-20.33	9.12	4.71	4.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0		-2.00	15.04	12.04	-20.33	76.26	4.71	4 71	0.00	0.00	0.00	0.00	0.00	0.00
9		-12.04	-12.00	-2.66	13.04	-15 04	-26.35	4.12	4.71	0.00	0.00	0.00	0.00	0.00
0		3 30	-12.04	-12.00	.2.66	13.04	-15 04	-26 35	4.12	4.71	0.00	0.00	0.00	0.00
1		8 23	3 30	.9 05	-12.00	.2 66	13.04	-15.94	-76 35	A 12	4 71	0.00	0.00	0.00
2		3.50	8.23	3.30	-8.95	-12.00	.2.66	13.04	-15.94	-26 35	4 12	4 71	0.00	0.00
3		27.64	3.50	8.23	3.30	.8.95	-12.04	-2.66	13.04	-15.94	-76.35	4.12	4.71	0.00
4		-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04	-2.66	13.04	-15.94	-26 35	4.12	4.71
ŝ		-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04	-2.66	13.04	-15.94	-26.35	4.12
6		-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04	-2.66	13.04	-15.94	-26.35
7		-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04	-2.66	13.04	-15.94
8		31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04	-2.66	13.04
9		-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04	-2.66
0		-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04
1		1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95
Z		13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30
3		4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23
4		9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50
5		28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64
б		-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23
7		12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71
8		-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99
9		-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39
0		6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07
1		-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92
2		7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73
3		-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26
4		2.48	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50
2		-1.40	2.48	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07
0		41.20	-1.40	2.48	-4.37	7.90	-18.89	6.75	-1.18	-33.52	12.56	-35.81	28.01	9.59
7.6		48.15	41.20	-1.40	Z.48	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01
8		13636.81634	1				_	_	_	1000				1

Рис. 23. Прекрасно подогнанная каскадная матрица сдвинутых отклонений погрешности

Сложите эти произведения. Чем ближе окажется сумма (SUMPRODUCT / СУММПРОИЗВ) столбца со сдвигом исходных отклонений к сумме квадратов отклонений в C38, тем более синхронны, более коррелированы сдвинутые погрешности будут к исходным.

Также вы можете получить негативную автокорелляцию, в которой сдвинутые отклонения становятся отрицательными, независимо от того, положительны исходные значения или отрицательны. SUMPRODUCT/CYMMПPOИЗВ в этом случае будет довольно большим отрицательным числом.

Начнем с растягивания формулы SUMPRODUCT (\$C2:\$C37, C2:C37) / СУММ- ПРОИЗВ (\$C2:\$C37, C2:C37) в ячейке C38 вправо до столбца О. Обратите внимание на то, как абсолютная ссылка на столбец С сохраняет его на месте. У вас появляется SUMPRODUCT/СУММПРОИЗВ каждого столбца со сдвигом и исходного столбца, как показано на рис. 24.

A Home	Lavout To	bles d	Charts	SmartArt	Formulas	Data	Review	Develo	ner l				
MIN	: 0.0	1- 6x	-SUMPRO	DUCT/SC2	102 02 03	71				_		_	10000
1	C	D	E	F.	G	H	1	1	K	L	м	N	
Deviat	ions from mean	1	2	3	4	5	6	7	8	9	10	11	1
3	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.04	-2.6
)	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.95	-12.0
Ē	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.30	-8.9
5	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.23	3.3
	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.50	8.3
	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.64	3.5
	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.23	27.
1	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.99	-16.71	-29.2
	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	+12.99	-16.
	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.07	-22.39	-12.
	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.25	-8.73	-19.92	31.07	-22.
	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.92	31.0
	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.73	-19.
	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.26	-8.
	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.50	1.
1	2.48	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.07	13.
	-1.40	2.48	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.59	4.(
	41.20	-1.40	2.48	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.01	9.1
	48.15	41.20	-1.40	2,48	-4.37	7.90	-18.89	6.76	-1.18	-33.52	12.56	-35.81	28.0
	13636.81634	-474.069	753.363	-229.306	-2223.642	1544.284	-2442.531	1790.525	-4382.609	-167.700	-1460.419	733.062	02:037

Рис. 24. SUMPRODUCT/СУММПРОИЗВ сдвинутых отклонений и исходных значений

Теперь рассчитайте автокорреляцию данных со сдвигом на месяц: SUMPRODUCT/ СУММПРОИЗВ отклонений сдвига, разделенная на сумму квадратов отклонений в C38.

Для примера можно рассчитать автокорреляцию сдвига на один месяц в ячейке C40: =D38/\$C38

Растянув ее по горизонтали, можно получить автокорреляцию для каждого сдвига.

Теперь выделите D40:040 и вставьте столбчатую диаграмму, как показано на рис. 25 (кликните правой клавишей мышки на серии данных и отформатируйте их, чтобы столбики диаграммы стали прозрачнее и номера месяцев лучше читались под осью). Эта столбчатая диаграмма называется коррелограммой. Она показывает автокорреляции для каждого сдвига на месяц до самого конца года.



Рис. 25. Это коррелограмма.

Какие автокорреляции нам важны? Как правило, все автокорреляции, большие, чем 2/ корень из количества точек данных, что в нашем случае будет $2/\sqrt{36} = 0,333$. Также нас интересуют отрицательные автокорреляции меньше -0,333.

Посмотрите на эту диаграмму автокорреляций, находящихся выше и ниже *критических значений*. В прогнозировании на коррелограмме принято изображать критические значения пунктирной линией. Из любви к красивым картинкам я, так и быть, покажу, как это сделать здесь.

В ячейку D42 добавьте =2/SQRT (36) /=2/КОРЕНЬ (36) и растяните вправо до столбца О. В D43 сделайте то же самое, только с отрицательным значением: =-2/SQRT (36) /=-2/КОРЕНЬ (36) и тоже растяните вправо до О. Так получаются критические

точки автокорреляций, показанные на рис. 26.

	fome Layout Ti	ables (Charts	SmartArt	Formula	s Data	Review	Develo	per.				× (
	D42 1 0 0	- fx	=-2/SQRT	(36)											
11	C	Ð	E	F	G	н	11	3	×	L	M	N	0		
1 D	eviations from mean	1	2	3	4	5	6	7	8	9	10	11	12		
7	48.15	41.20	-1.40	2.48	-4.37	7.90	-18.89	6.75	-1.18	-33.52	12.56	-35.81	28.01		
8	13636.81634	-474.069	753.363	-229.306	-2223.642	1544.284	-2442.531	1790.525	-4382.609	-167.700	-1460.419	733.062	5512.801		
9.															
0 A	utocorrelation	-0.035	0.055	-0.017	-0.163	0.113	-0.179	0.131	-0.321	-0.012	-0.107	0.054	0.404		
1		0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333		
51		.0 333	.0 333	.0 333	.0 333	.0 333	.0 333	.0 333	.0 333	.0 333	.0 333	.0 333	.0 333		

Рис. 26. Критические точки автокорреляций

Кликните правой клавишей мышки на столбчатой диаграмме и выберите «Выбрать данные». В выплывшем окошке нажмите кнопку «Добавить», чтобы создать новые серии данных.

Для первой из них выберите промежуток D42:O42 для оси *у*. А для второй - D43:O43. Это добавит новые столбцы на диаграмму.

Правый щелчок на каждой из этих новых серий столбцов даст возможность изменить тип отображения данных на линейную диаграмму. Затем кликните правой кнопкой мыши на эти линии и выберите «Формат данных». Переключитесь на линию (тип линии в некоторых версиях Excel) в выплывшем окне. Здесь вы можете сделать свою линию пунктирной, как показано на рис. 27.

Таким образом получается коррелограмма с пунктирными критическими значениями. Она изображена на рис. 28.

И что же мы видим?

Совершенно точно есть одна автокорреляция выше критического значения, и это 12 месяцев.

Погрешность, сдвинутая на год, коррелирует сама с собой. Это означает 12-месячный сезонный цикл. И это неудивительно. Если вы посмотрите на график спроса во вкладке Timeseries, то окажется, что есть пики спроса на каждое декабрь-январь и провалы в апреле-мае.

Вам нужна такая техника прогнозирования, которая способна учитывать сезонность. Кроме того, существует аналогичная техника экспоненциального сглаживания.



Рис. 27. Изменение отображения критических значений со столбцов на пунктир



Рис. 28. Коррелограмма с критическими значениями