

Функции предсказания в Excel

Excel, как универсальный табличный редактор, давно и неплохо справляется с большинством задач прогнозирования (см. список литературы в конце заметки). Однако, не всегда вычисления в Excel являются простыми и понятными. И вот в версии 2016 года разработчики Microsoft добавили семейство функций ПРЕДСКАЗ (FORECAST), которые позволяют в несколько кликов решать большой круг задач прогнозирования на основе экспоненциального сглаживания.

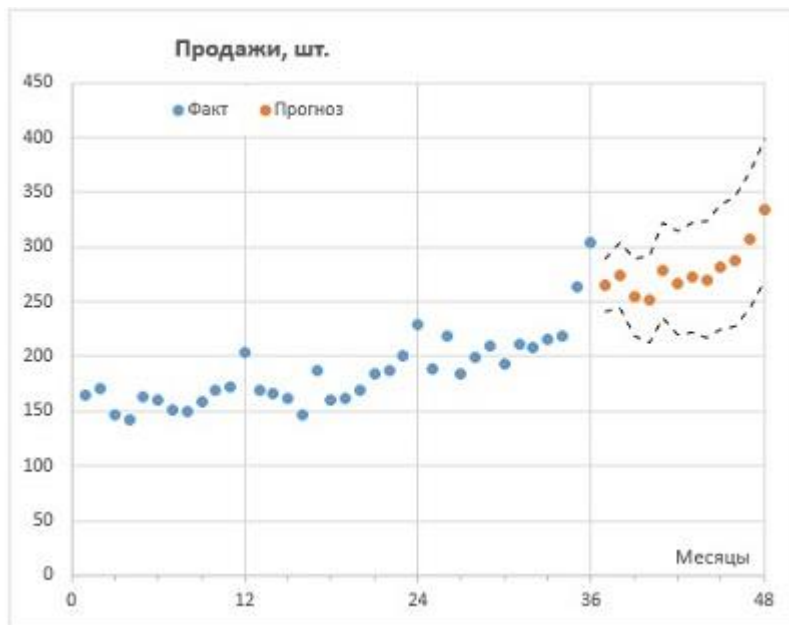


Рис. 1. Прогнозирование продаж в Excel с помощью семейства функций ПРЕДСКАЗ

Об экспоненциальном сглаживании

Экспоненциальное сглаживание также известно, как метод ETS: ошибки (Errors), тренд (Trend), сезонный фактор (Seasonal). Для составления прогноза используются все исторические данные, но коэффициенты, определяющие вклад, убывают в прошлое по экспоненте (отсюда и название). Это позволяет, с одной стороны, чутко реагировать на свежие данные, с другой стороны, сохранять информацию об историческом поведении всего временного ряда. Если данным присущ тренд, он вычисляется в каждой точке данных (а не на основе регрессии всего временного ряда). Наконец, с помощью автокорреляции в данных выявляется сезонность.

Преимущество модели в том, что она не использует никаких предположений относительно характера тренда (или его отсутствия) и периодичности сезонных колебаний (или их отсутствия). Все коэффициенты в модели подбираются на основе минимизации суммы квадратов ошибок, то есть, разности между прогнозом на исторических данных и самих данных. Если вас интересует, как это происходит, рекомендую работу Формана (см. список литературы).

Собственно, оптимизируются три коэффициента:

α – разброс относительно среднего

γ – тренд

δ – сезонность

Разработчики Microsoft не предоставили пользователям возможность влиять на выбор коэффициентов, за исключением периода сезонности (об этом ниже).

Обзор функций семейства ПРЕДСКАЗ

В Excel представлено 5 функций:

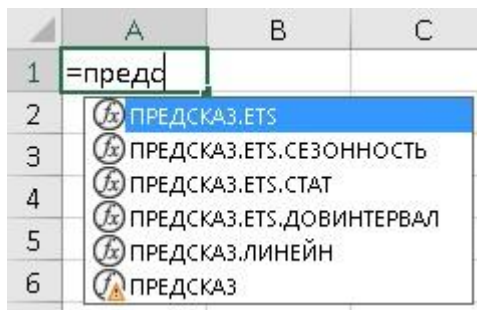


Рис. 2. Семейство функций ПРЕДСКАЗ в Excel

ПРЕДСКАЗ.ETS рассчитывает будущее значение на основе существующих (ретроспективных) данных методом экспоненциального сглаживания. Т.е., дает прогноз одним числом.

ПРЕДСКАЗ.ETS.ДОВИНТЕРВАЛ возвращает доверительный интервал для прогнозной величины. Доверительный интервал следует отложить по обе стороны от среднего значения. Вместе с ПРЕДСКАЗ.ETS позволяет построить «коридор» прогноза.

ПРЕДСКАЗ.ETS.СЕЗОННОСТЬ возвращает длину повторяющегося фрагмента, обнаруженного в заданном временном ряду. Например, 12, если исторические данные представляют из себя продажи за месяц.

ПРЕДСКАЗ.ETS.СТАТ возвращает восемь статистических значений, являющихся результатом прогнозирования временного ряда. Вряд ли вы будете использовать эту функцию. Она нужна для более тонкого исследования параметров прогнозной модели.

ПРЕДСКАЗ.ЛИНЕЙН вычисляет будущее значение с помощью линейной регрессии исторических данных. До версии 2016 в Excel вместо семейства функций была единственная функция ПРЕДСКАЗ, которая работала также, как и ПРЕДСКАЗ.ЛИНЕЙН. Функция ПРЕДСКАЗ оставлена для обратной совместимости, но скоро перестанет поддерживаться. Далее в заметке ПРЕДСКАЗ.ЛИНЕЙН не рассматривается, так как не относится к функциям, использующим алгоритм экспоненциального сглаживания.

ПРЕДСКАЗ.ETS

Синтаксис:

ПРЕДСКАЗ.ETS

```
(
    целевая_дата;
    значения;
    временная_шкала;
    [сезонность];
    [заполнение_данных];
    [агрегирование]
)
```

В качестве примера рассмотрим месячный пассажиропоток в аэропорту (пример от [MS](#)). Исторические данные были собраны за период с января 2009 по декабрь 2012 г.

	А	В
1	Месяц	Число пассажиров
2	январь.09	2 644 539
3	февраль.09	2 359 800
4	март.09	2 925 918
5	апрель.09	3 024 973
6	май.09	3 177 100
7	июнь.09	3 419 595
8	июль.09	3 649 702
9	август.09	3 650 668
10	сентябрь.09	3 191 526
11	октябрь.09	3 249 428
12	ноябрь.09	2 971 484
13	декабрь.09	3 074 209
14	январь.10	2 785 466
44	июль.12	4 284 443
45	август.12	4 356 216
46	сентябрь.12	3 819 379
47	октябрь.12	3 844 987
48	ноябрь.12	3 478 890
49	декабрь.12	3 443 039

Рис. 3. Исторические данные

Продолжим временную шкалу еще на год, и создадим столбец для прогноза. Обычно прогноз располагают в отдельном столбце для того, чтобы при построении графика представить исторические и прогнозные значения разными линиями.

=ПРЕДСКАЗ.ETS(A50;\$B\$2:\$B\$49;\$A\$2:\$A\$49;1;1)

	А	В	С	Д
1	Месяц	Число пассажиров	Прогноз	
44	июль.12	4 284 443		
45	август.12	4 356 216		
46	сентябрь.12	3 819 379		
47	октябрь.12	3 844 987		
48	ноябрь.12	3 478 890		
49	декабрь.12	3 443 039		
50	январь.13		3 165 737	
51	февраль.13		2 890 877	
52	март.13		3 432 636	
53	апрель.13		3 505 637	
54	май.13		3 780 493	
55	июнь.13		4 026 030	
56	июль.13		4 208 821	
57	август.13		4 206 779	
58	сентябрь.13		3 782 632	
59	октябрь.13		3 858 961	
60	ноябрь.13		3 573 325	
61	декабрь.13		3 615 826	

Рис. 4. Прогнозные значения на основе функции ПРЕДСКАЗ.ETS

Подробнее о формуле в ячейке C50:

=ПРЕДСКАЗ.ETS(A50;\$B\$2:\$B\$49;\$A\$2:\$A\$49;1;1)

Первый аргумент – целевая_дата = A50 – янв.13, т.е., в ячейке C50 ищется прогноз пассажиропотока для января 2013 г. Ссылка относительная, что позволит при протягивании функции вниз по столбцу ссылаться на новое значение: в C51 – на A51, в C52 – на A52 и т.д.

Второй аргумент – значения = \$B\$2:\$B\$49. Здесь расположены исторические данные пассажиропотока. Ссылка абсолютная, чтобы при протягивании формулы ячейки, на которые ссылаются не изменились.

Третий аргумент – временная_шкала = \$A\$2:\$A\$49. Здесь расположены даты временной шкалы или номера периодов. Важно чтобы они отстояли друг от друга на фиксированный интервал. Если интервал не будет фиксированным, Excel всё еще будет исходить из гипотезы, что интервал фиксированный, а некоторые данные пропущены. Как обрабатываются такие ситуации описано ниже. Сортировать массив по значениям временной шкалы не обязательно, так как ПРЕДСКАЗ.ETS сама отсортирует данные прежде, чем выполнить расчеты.

Четвертый аргумент – [сезонность] = 1. Это необязательный аргумент. Значение по умолчанию равно 1. Для него Excel автоматически определяет сезонность и использует положительные целые числа в качестве длины сезонного шаблона. Значение 0 предписывает не использовать фактор сезонности, в результате чего прогноз будет линейным. Если для этого параметра задано положительное целое число, алгоритм использует его в качестве длины шаблона сезонности. Например, вы знаете, что сезонность равна 4 (квартальная периодичность), но предполагаете, что она слабая, и автоматический алгоритм Excel может ее не выявить, и будет считать, что сезонности нет. Для начала я рекомендовал бы использовать значение по умолчанию.

Пятый аргумент – [заполнение_данных] = 1. Это необязательный аргумент. Хотя временная шкала требует постоянный шаг между точками данных, FORECAST.ETS поддерживает до 30% отсутствующих данных и автоматически настраивает их. 0 указывает, что алгоритм учитывает отсутствующие точки в качестве нулей. Если задано значение 1 (вариант по умолчанию), функция определяет отсутствующие значения как среднее между соседними точками.

Шестой аргумент – [агрегирование] – в нашем примере опущен. Это необязательный аргумент. Он нужен, если даты временной шкалы или номера периодов содержат дубли. Функция ПРЕДСКАЗ.ETS выполнит агрегирование точек с одинаковой меткой времени. Параметр агрегирования — это числовое значение, определяющее способ агрегирования нескольких значений с одинаковой меткой времени. Для значения по умолчанию 0 используется метод СРЗНАЧ; также доступны варианты СУММ, СЧЁТ, СЧЁТЗ, МИН, МАКС и МЕДИАНА.

ПРЕДСКАЗ.ETS.ДОВИНТЕРВАЛ

Синтаксис:

ПРЕДСКАЗ.ETS.ДОВИНТЕРВАЛ

(
целевая_дата;
значения;
временная_шкала;
[вероятность];
[сезонность];
[заполнение_данных];
[агрегирование]
)

Здесь один новый аргумент – [вероятность]. Он необязательный. И принимает значения от 0 до 1 (не включительно). Например, для значения 95% выполняется расчет доверительного интервала, в который с вероятностью 95% попадут прогнозные точки. Функция ПРЕДСКАЗ.ETS.ДОВИНТЕРВАЛ возвращает значение в тех же физических единицах, что и ПРЕДСКАЗ.ETS. Например, 50±8. Здесь 8 – доверительный интервал, возвращаемый функцией ПРЕДСКАЗ.ETS.ДОВИНТЕРВАЛ. Он откладывается от среднего прогнозного значения по обе стороны.

	A	B	C	D	E
D50	=C50-ПРЕДСКАЗ.ETS.ДОВИНТЕРВАЛ(A50;\$B\$2:\$B\$49;\$A\$2:\$A\$49;0,95;1;1)				
1	Месяц	Число пассажиров	Прогноз	Нижняя граница	Верхняя граница
44	июл.12	4 284 443			
45	авг.12	4 356 216			
46	сен.12	3 819 379			
47	окт.12	3 844 987			
48	ноя.12	3 478 890			
49	дек.12	3 443 039	3 443 039	3 443 039	3 443 039
50	январ.13		3 165 737	3 069 004	3 262 471
51	фев.13		2 890 877	2 760 671	3 021 084
52	мар.13		3 432 636	3 275 114	3 590 157
53	апр.13		3 505 637	3 325 493	3 685 781
54	май.13		3 780 493	3 580 266	3 980 721
55	июн.13		4 026 030	3 807 481	4 244 579
56	июл.13		4 208 821	3 973 370	4 444 271
57	авг.13		4 206 779	3 955 511	4 458 046
58	сен.13		3 782 632	3 516 438	4 048 825
59	окт.13		3 858 961	3 578 635	4 139 288
60	ноя.13		3 573 325	3 279 487	3 867 162
61	дек.13		3 615 826	3 308 977	3 922 675

Рис. 5. Расчет границ коридора прогноза

Теперь на основании расчета среднего прогноза, возвращаемого функцией ПРЕДСКАЗ.ETS и границ коридора, возвращаемого формулами...

=C50±ПРЕДСКАЗ.ETS.ДОВИНТЕРВАЛ(A50;\$B\$2:\$B\$49;\$A\$2:\$A\$49;0,95;1;1)

... можно построить график прогноза с доверительными границами:

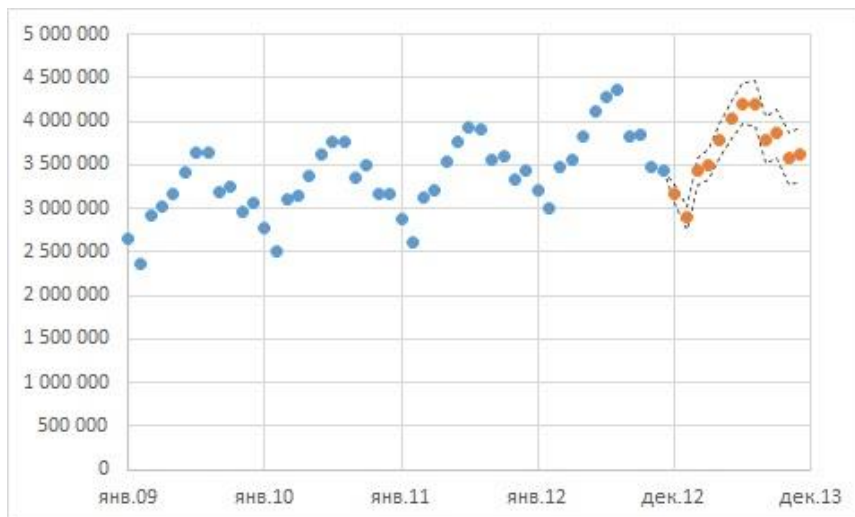


Рис. 6. Исторические и прогнозные данные пассажиропотока с 95%-ными доверительными границами

Благодаря тому, что исторические данные демонстрируют четкие закономерности, прогнозные границы получились довольно узкими.

ПРЕДСКАЗ.ETS.СЕЗОННОСТЬ

Синтаксис:

ПРЕДСКАЗ.ETS.СЕЗОННОСТЬ

(

значения;

временная_шкала;

[заполнение_данных];
 [агрегирование]
)

Здесь нет новых для нас аргументов. В простейшем случае функция ПРЕДСКАЗ.ETS.СЕЗОННОСТЬ принимает исторические данные и временную шкалу, и возвращает целое число, равное сезонности. В нашем примере – 12:

=ПРЕДСКАЗ.ETS.СЕЗОННОСТЬ(\$B\$2:\$B\$49;\$A\$2:\$A\$49)					
	A	B	C	D	E
1	Месяц	Число пассажиров			
2	янв.09	2 644 539			
3	фев.09	2 359 800		12	
4	мар.09	2 925 918			
5	апр.09	3 024 973			
6	май.09	3 177 100			
7	июн.09	3 419 595			
8	июл.09	3 649 702			
9	авг.09	3 650 668			
10	сен.09	3 191 526			
11	окт.09	3 249 428			

Рис. 7. Значение сезонности

ПРЕДСКАЗ.ETS.СТАТ

Синтаксис

ПРЕДСКАЗ.ETS.СТАТ

(
 значения;
 временная_шкала;
 тип_статистики;
 [сезонность];
 [заполнение_данных];
 [агрегирование]
)

Новый аргумент один – тип_статистики. Является обязательным. Принимает целые значения от 1 до 8. Я использовал вложенную функцию ПОСЛЕД(8), которая задала массив целых чисел: 1, 2, ... , 8. И функция ПРЕДСКАЗ.ETS.СТАТ отработала как [функция динамического массива](#), создав столбец из 8 значений статистики:

=ПРЕДСКАЗ.ETS.СТАТ(B2:B49;A2:A49;ПОСЛЕД(8))				
	A	B	C	D
1	Месяц	Число пассажиров		
2	янв.09	2 644 539		
3	фев.09	2 359 800		0,9
4	мар.09	2 925 918		0,001
5	апр.09	3 024 973		0,099
6	май.09	3 177 100		0,227
7	июн.09	3 419 595		0,015
8	июл.09	3 649 702		53 973
9	авг.09	3 650 668		60 439
10	сен.09	3 191 526		31
11	окт.09	3 249 428		

Рис. 8. Статистики экспоненциального сглаживания

Первый параметр – альфа = 0,9. Чем он больше, тем больше вес более новых точек данных. Не следует путать коэффициенты, оптимизируемые в модели ETS (α , γ и δ), и значения, возвращаемые функцией ПРЕДСКАЗ.ETS.СТАТ. Это – о разном. Второй параметр – бета = 0,001. Возвращает значение параметра тренда: чем оно больше, тем больше вес более нового тренда. Третий параметр – гамма. Возвращает значение параметра сезонности: чем оно больше, тем больше вес более нового сезонного периода. С четвертого по седьмой параметр – это различные показатели точности прогноза. Восьмой параметр – величина шага, определенную во временной шкале ретроспективных значений. В нашем случае 31 день.

Лист прогноза

Вишенкой на торте является кнопка на ленте, доступная по меню Данные → Прогноз → Лист прогноза. Эта кнопка на основе исторических данных создает новый лист Excel с колонками прогноза, границ доверительного интервала, статистикой модели и довольно сноской диаграммой. Итак, встаньте курсором на любую клетку умной таблицы с исходными данными (как на рис. 3) и кликните Лист прогноза. Откроется окно Создание листа прогноза.

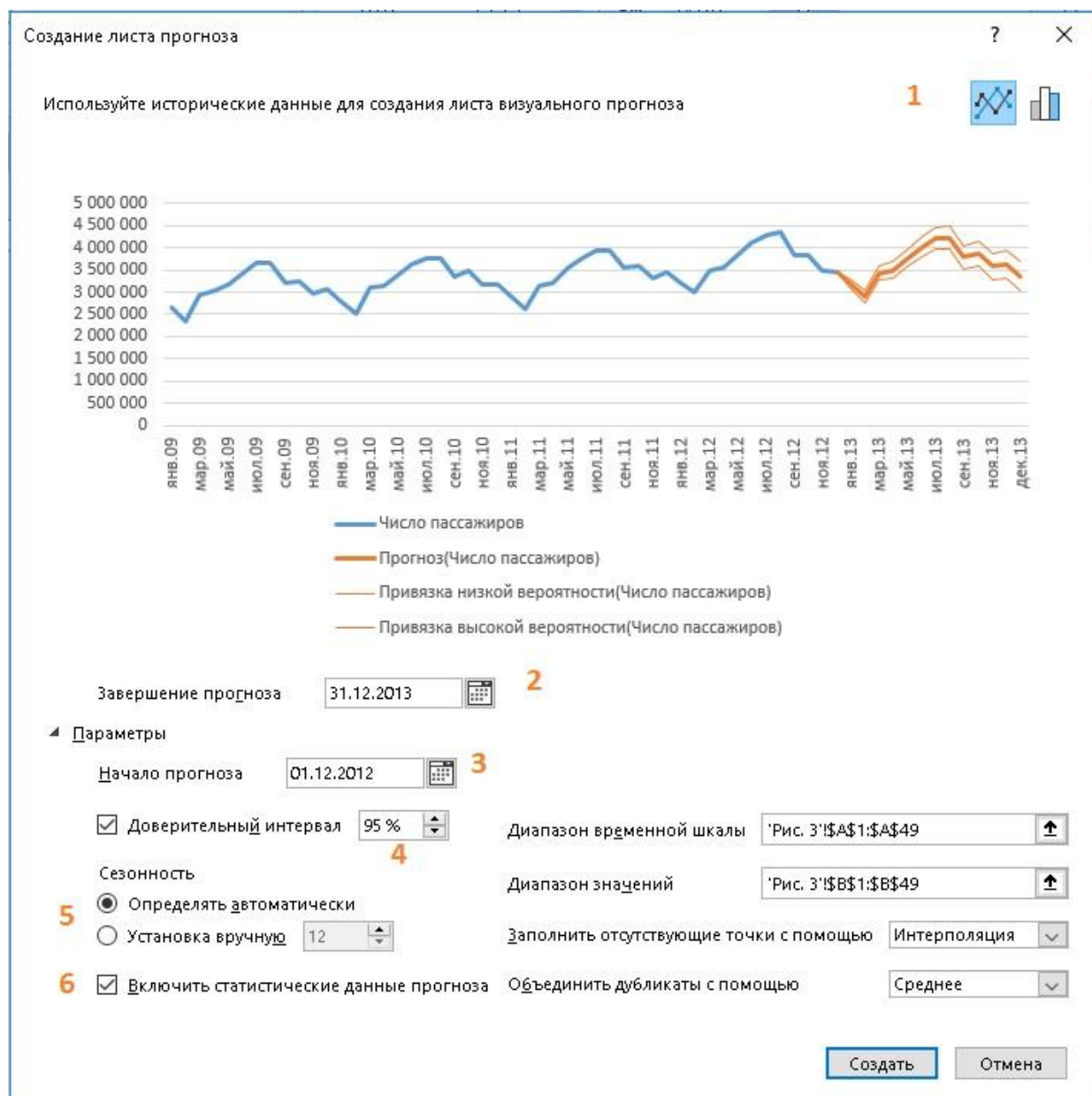


Рис. 9. Настройка параметров листа прогноза

Здесь (1) можно выбрать тип диаграммы: график или гистограмма, (2) задать период прогнозирования, (3) выбрать на временной шкале начало прогноза; (4) выбрать величину доверительного интервала, (5) использовать параметр сезонности (число), определяемый функцией ПРЕДСКАЗ.ETS автоматически, или установить сезонность принудительно, (6) указать, выводить ли статистику прогнозной модели.

Если выбрать дату начала прогноза внутри временной шкалы исторических данных, то прогноз для каждой точки будет использовать только данные, предшествующие ей (это называется «ретроспективным прогнозированием»).

Если начать прогнозирование перед последней точкой, можно получить оценку точности прогноза, сравнив прогнозируемый ряд с фактическими данными. Но если начать прогнозирование со слишком ранней даты, построенный прогноз может отличаться от созданного на основе всех статистических данных. При использовании всех статистических данных прогноз будет более точным.

Если в ваших данных прослеживаются сезонные тенденции, то рекомендуется начинать прогнозирование с даты, предшествующей последней точке статистических данных.

Если вы хотите задать сезонность вручную, не используйте слишком короткие ряды исторических значений. Нужно как минимум два полных цикла статистических данных. На коротких исторических данных лучше довериться автоматическому определению сезонности. Если алгоритм не выявит сезонность, прогноз примет вид линейного тренда.

Настроив параметры в окне *Создание листа прогноза*, нажмите *Создать*. В нашем примере Excel создаст новый лист, расположенный слева (перед) листом с историческими данными:

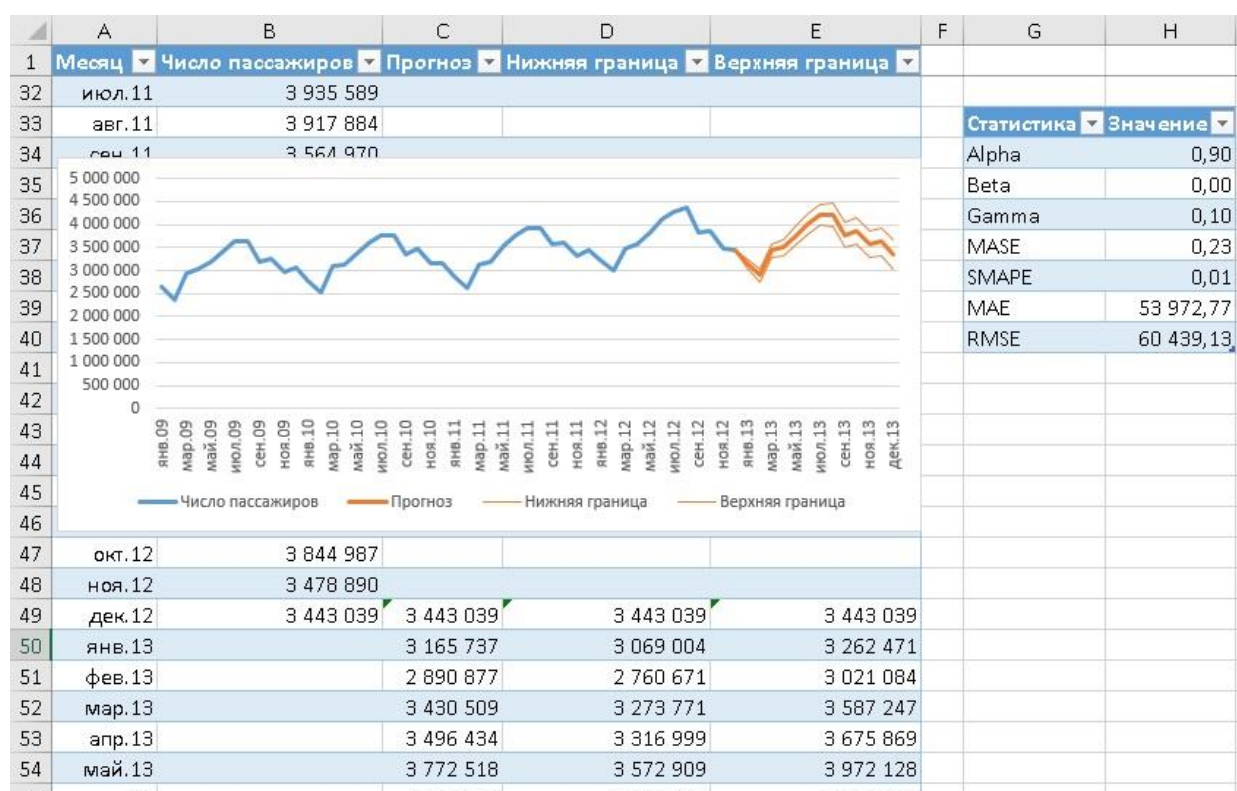


Рис. 10. Лист прогноза

Литература

Джон Форман. Много цифр: Анализ больших данных при помощи Excel. Глава 8. [Прогнозирование на основе экспоненциального сглаживания](#). В заметке на сквозном примере в Excel рассмотрено простое экспоненциальное сглаживание, ПЭС (simple exponential smoothing, SES); экспоненциальное сглаживание Холта с корректировкой тренда; анализ автокорреляций в данных; мультипликативное экспоненциальное сглаживание Холта-Винтерса; построение доверительного интервала прогноза методом Монте-Карло.

Левин и др. Статистика для менеджеров. Глава 15. [Анализ временных рядов](#). В заметке: что такое временные ряды; компоненты классической мультипликативной модели временных рядов; сглаживание временных рядов; скользящее среднее; экспоненциальное сглаживание; вычисление трендов с помощью метода наименьших квадратов; модели линейной, квадратичной и экспоненциальной регрессии; выбор модели на основе разностей первого и второго порядка; вычисление тренда с помощью авторегрессии; прогнозирование временных рядов на основе сезонных данных; ловушки, связанные с анализом временных рядов; индексы.

[Прогнозирование спроса и продаж. Записки дилетанта](#). В заметке: общие рассуждения о прогнозировании; постановка задачи прогнозирования; наивные методы прогнозирования (простое среднее, скользящее среднее, взвешенное скользящее среднее); простое экспоненциальное сглаживание; составляющие прогнозной модели (тренд, сезонность); классическая сезонная декомпозиция; экспоненциальное сглаживание с учетом тренда и сезонности (метод Холта-Винтерса); прогнозирование редких продаж (метод Кростона и метод Виллемейна); метод Кростона и страховой запас; функция распределения спроса.