

Лабораторная работа №5

«Электронные таблицы Excel и автоматизация вычислений на ПК»

РАЗДЕЛ №5. Статистические функции и графический анализ данных.

Вы менеджер фирмы по продажам подержанных автомобилей и постоянно ведете учет проданных машин. В распоряжении имеются две наблюдаемые величины: x – номер недели, y – число проданных за неделю автомобилей. Фирма совсем молодая, была создана шесть недель назад, и поэтому в Вашем распоряжении имеется статистика только за этот весьма ограниченный промежуток времени.

Наблюдаемые величины	Значения					
x	1	2	3	4	5	6
y	7	9	12	13	14	17

Вы хотите сначала смоделировать ту динамику продаж, которая имеет место, а на основе построенной модели затем попытаться заглянуть в будущее, то есть спрогнозировать ожидаемый объем продаж на ближайшие недели.

В качестве модели берем простейшую, то есть линейную. Таким образом, надо построить такую линейную модель $y=mx+b$, которая наилучшим образом описывает наблюдаемые значения. Обычно m и b подбираются так, чтобы минимизировать сумму квадратов разностей теоретических и наблюдаемых значений зависимой переменной y , то есть минимизировать $\sum_{i=1}^n (y_i - mx_i - b)^2$, где n – число наблюдений. Определить минимальное

значение из всех возможных значений выражения мы можем при помощи инструмента «Поиск решения». В этом случае неизвестными параметрами будут являться m и b , которые Excel подберет в ходе поиска решения. Однако параметры m и b линейной модели $y=mx+b$ можно определить при помощи функций НАКЛОН и ОТРЕЗОК.

НАКЛОН(массив1; массив2) – определяет коэффициент наклона линейного тренда;

ОТРЕЗОК(массив1; массив2) – определяет точку пересечения линии линейного тренда с осью ординат.

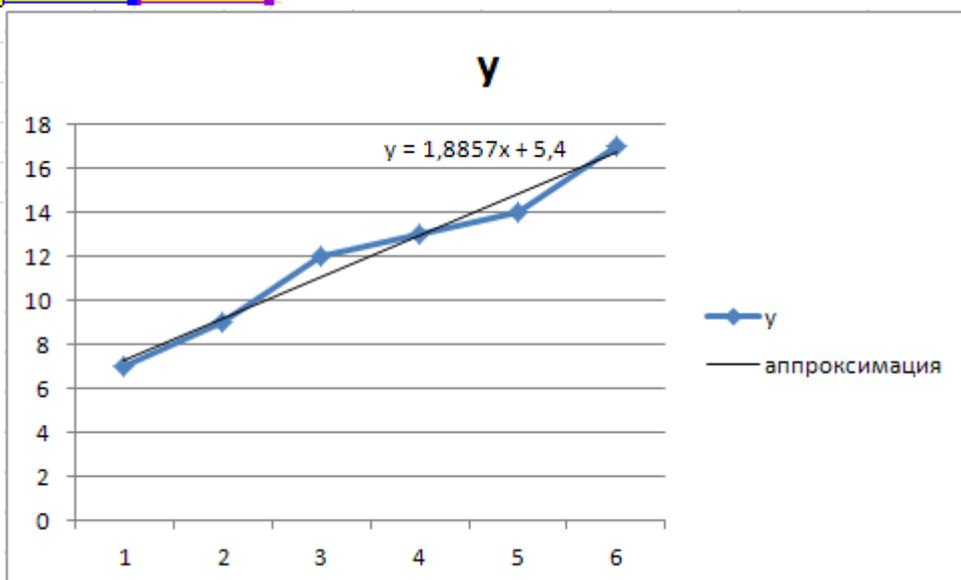
В обеих функциях *массив1* содержит ссылку на диапазон ячеек с известными значениями зависимой наблюдаемой величины y , а *массив2* – независимой наблюдаемой величины x .

Задание №1: Выполните построение линейной модели $y=mx+b$, предварительно

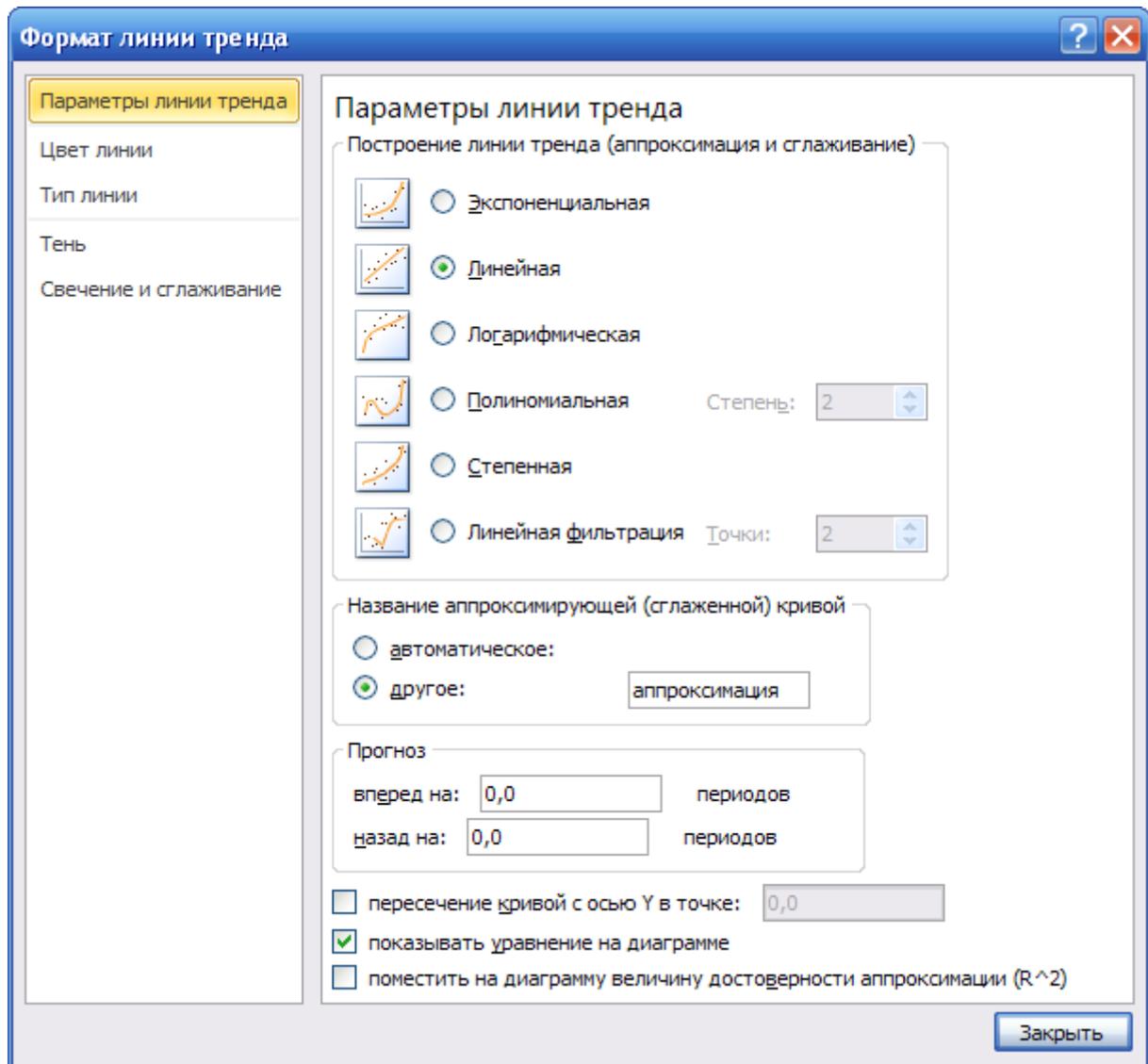
	A	B	C	D	E
1	x	y	линейная модель	наклон m	отрезок b
2	1	7	=D\$2*A2+E\$2		
3	2	9			
4	3	12			
5	4	13			
6	5	14			
7	6	17			

вычислив с помощью функций Excel значения величин наклона m и отрезка b . Пример таблицы показан на рисунке.

После заполнения ячеек диапазона C2:C7 формулой вычисления линейной модели, на основе значений ячеек диапазона B2:B7 постройте диаграмму «график с маркерами». Помимо отображаемых данных ряда чисел из



диапазона ячеек B2:B7, на диаграмме можно отобразить линию тренда (в контекстном меню на линии графика выбрать пункт «Добавить линию тренда»), которая аппроксимирует кривую графика согласно выбранному *типу уравнения регрессии* – позволяет на основе полученной зависимости вычислять прогнозируемые значения, находящиеся вне зоны наблюдения. В диалоге параметров тренда можно указать тип аппроксимации (сглаживания), а также уточнить подпись в легенде для линии тренда и отобразить уравнение аппроксимирующей зависимости.



Значения для линейной модели можно вычислить при помощи функции ПРЕДСКАЗ, не определяя предварительно коэффициенты линейной модели, в фиксированной точке, как в зоне наблюдений, так и вне ее.

ПРЕДСКАЗ(точка; массив1; массив2) – точка соответствует ячейке со значением для которого выполняется прогноз, массив1 содержит ссылку на диапазон ячеек с известными значениями зависимой наблюдаемой величины y , а массив2 – независимой наблюдаемой величины x .

Функция ТЕНДЕНЦИЯ вычисляет значения уравнения линейной регрессии для целого диапазона значений независимой переменной как для случая одномерного, так и многомерного уравнения регрессии. Многомерная модель регрессии соответствует параметрической зависимости (от нескольких неизвестных) и имеет вид: $y = m_1x_1 + \dots + m_nx_n + b$.

ТЕНДЕНЦИЯ(массив1; массив2; массив3; константа) – массив1 содержит ссылку на диапазон ячеек с известными значениями зависимой наблюдаемой величины y , массив2 – независимой наблюдаемой величины x , массив3 содержит ссылку на новые значения x , для которых прогнозируются соответствующие значения y , возвращаемые в качестве результата вычислений функции ТЕНДЕНЦИЯ. Логический аргумент константа принимает одно из двух значений ИСТИНА или ЛОЖЬ, и соответствует необходимости подбора значения отрезка b . При ложном значении аргумента константа значение отрезка b соответствует 0.

	A	B	C	D	E
1	x	y	линейная модель	наклон m	отрезок b
2	1	7	7,285714286	1,8857143	5,4
3	2	9	9,171428571		
4	3	12	11,05714286		
5	4	13	12,94285714		
6	5	14	14,82857143		
7	6	17	16,71428571		
8	7		=ПРЕДСКАЗ(A8;B\$2:B\$7;A\$2:A\$7)		
9	8				
10	9				
11			тенденция		

Если для расчетов необходимо знать значения коэффициентов m_1, \dots, m_n при независимых параметрах x_1, \dots, x_n уравнения линейной регрессии $y = m_1x_1 + \dots + m_nx_n + b$, а также значение отрезка b , используют функцию ЛИНЕЙН, которая возвращает массив значений m_n, \dots, m_1, b .

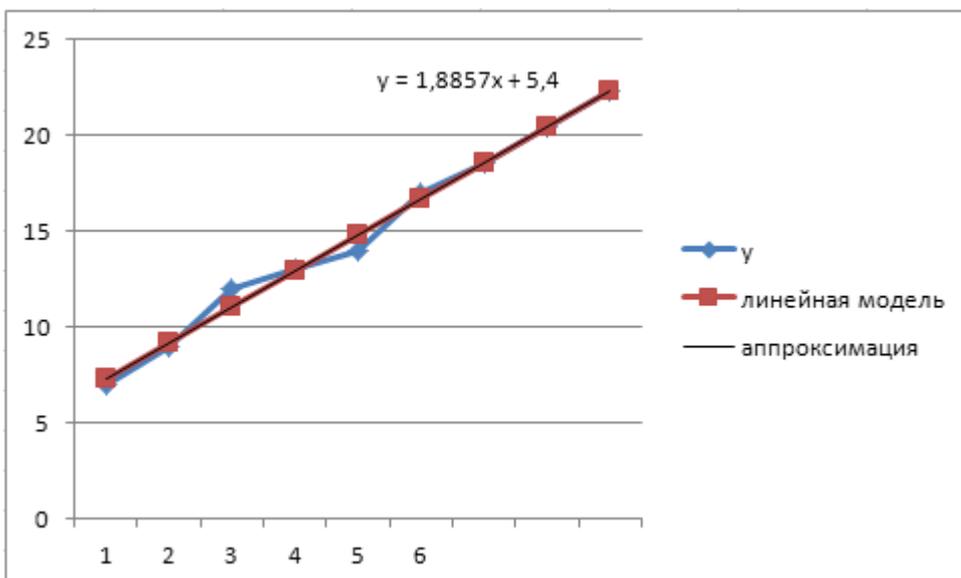
ЛИНЕЙН(массив1; массив2; константа) – массив1 содержит ссылку на диапазон ячеек с известными значениями зависимой наблюдаемой величины y , массив2 – независимой наблюдаемой величины x , константа имеет тоже самое значение что и в функции ТЕНДЕНЦИЯ.

	A	B	C	D	E
1	x	y	линейная модель	наклон m	отрезок b
2	1	7	7,285714286	1,8857143	5,4
3	2	9	9,171428571		
4	3	12	11,05714286		
5	4	13	12,94285714		
6	5	14	14,82857143		
7	6	17	16,71428571		
8	7		=ТЕНДЕНЦИЯ(B2:B7;A2:A7;A8:A10;ИСТИНА)		
9	8		20,48571429		
10	9		22,37142857		
11			тенденция		

Расширьте столбец наблюдаемых значений x (столбец A) в таблице новыми тремя величинами в диапазоне ячеек A8:A9, соответствующими 7-ой, 8-ой и 9-ой неделям. Используя функцию ПРЕДСКАЗ, продолжите заполнение линейной модели прогнозом на следующие три недели работы фирмы на рынке (рисунок выше).

В столбце наблюдаемых значений y (столбец B) в таблице добавьте прогнозируемые значения, возвращаемые функцией ТЕНДЕНЦИЯ, как показано на рисунке.

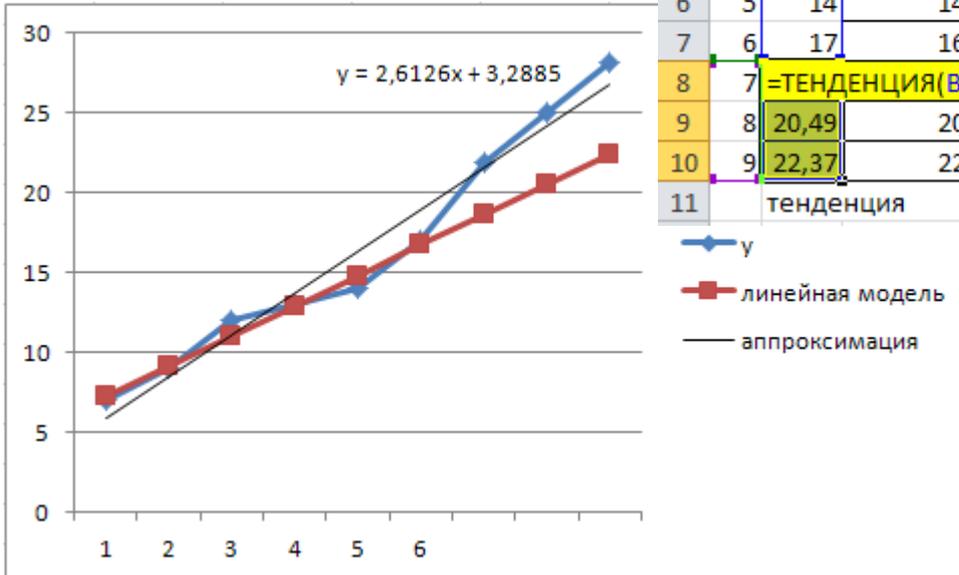
Имейте в виду, что, в отличие от функции ПРЕДСКАЗ, функция ТЕНДЕНЦИЯ возвращает массив прогнозируемых значений, согласно количеству ячеек в диапазоне (A8:A10), для которых составляется прогноз (необходимо выделить ячейки A8:A10, ввести формулу, по завершению ввода нажать комбинацию клавиш [Ctrl]+[Shift]+[Enter] вместо обычного нажатия [Enter]).



Измените диапазон построения диаграммы так, чтобы график охватывал дополнительные три недели продаж. Кроме этого добавьте еще один ряд данных, соответствующий линейной модели (диапазон ячеек C2:C10). Обратите внимание на совпадение графиков трех прогнозируемых недель продаж согласно расчетам по функциям ПРЕДСКАЗ, ТЕНДЕНЦИЯ и линии тренда.

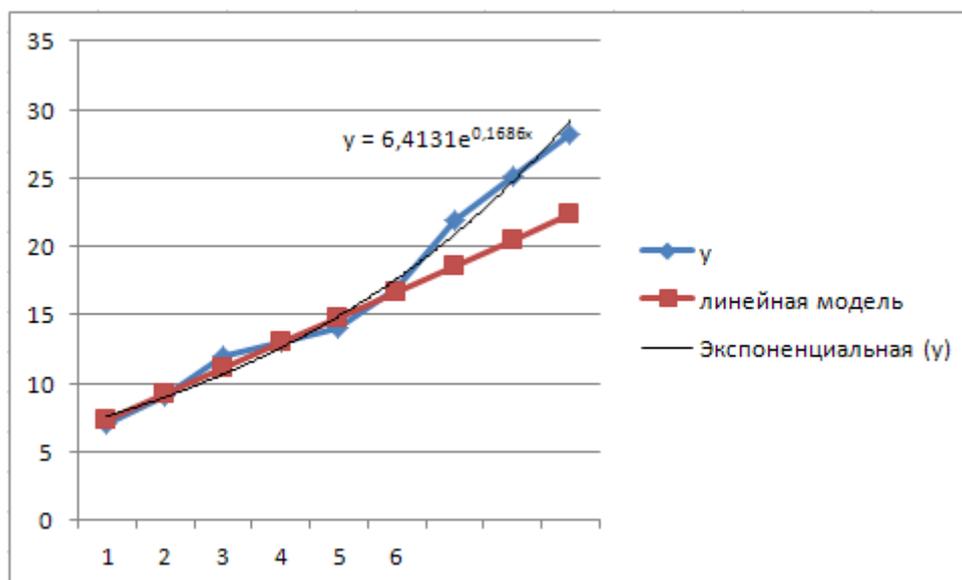
Выделите диапазон ячеек A8:A10 и внесите изменения в вычисления по функции ТЕНДЕНЦИЯ, указав значение последнего аргумента ЛОЖЬ. Не забудьте завершить изменение формулы нажатием комбинации клавиш [Ctrl]+[Shift]+[Enter].

	A	B	C	D	E
1	x	y	линейная модель	наклон m	отрезок b
2	1	7	7,285714286	1,8857143	5,4
3	2	9	9,171428571		
4	3	12	11,05714286		
5	4	13	12,94285714		
6	5	14	14,82857143		
7	6	17	16,71428571		
8	7	=ТЕНДЕНЦИЯ(B2:B7;A2:A7;A8:A10;ЛОЖЬ)			
9	8	20,49	20,48571429		
10	9	22,37	22,37142857		
11	тенденция				



Обратите внимание на изменения, которые произошли в линиях на диаграмме. Угол для графика линии тренда отклонился от угла направления линейной модели. В столбце прогноза по функции ТЕНДЕНЦИЯ новые значения устремились вверх.

Если действительно положительный прогноз исполнится и продажи пойдут вверх, то наиболее подходящим типом уравнения регрессии будет не столько линейная аппроксимация, сколько экспоненциальная. Выполните смену в диаграмме типа линии тренда на экспоненциальный. Диаграмма должна измениться, как показано на рисунке.



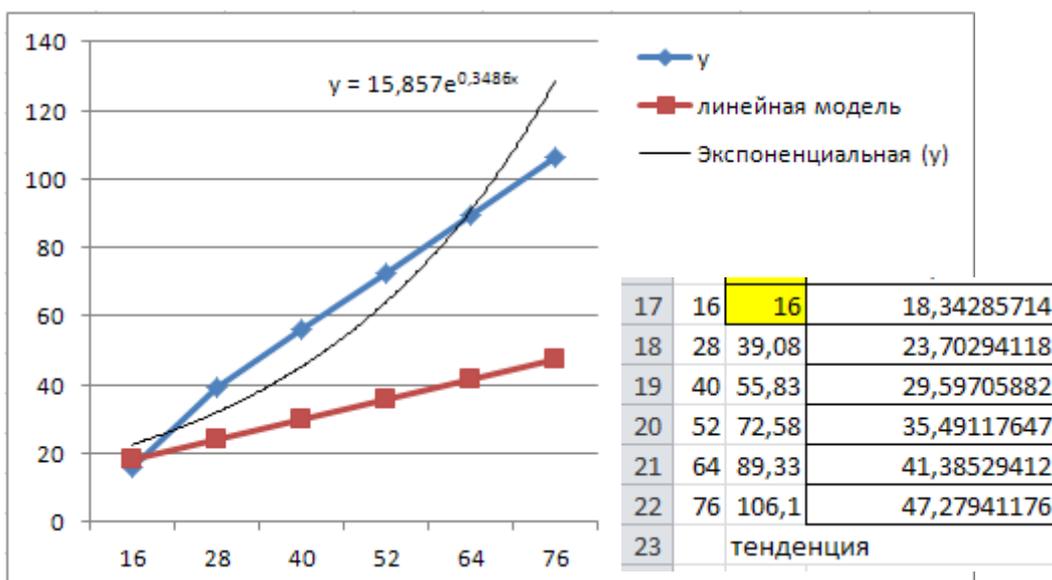
Вопрос: Как Вы считаете, имеет ли значение выбор типа регрессионной модели для точности прогнозирования? Если да, что выступает в качестве основного аргумента при выборе варианта аппроксимации? Если нет, какие объемы показателей (долгосрочные или

краткосрочные) достаточны для выработки наиболее правильного прогноза финансовой ситуации?

Задание №2: Определив ближайшие перспективы развития, топ-менеджмент фирмы решил накапливать финансовые показатели в дальнейшем, чтобы делать не только краткосрочные прогнозы, но и долгосрочные, обеспечивая свою конкурентоспособность на рынке. Продажи действительно увеличивались, но ненадолго. Внешним негативным фактором к благополучной экономической модели фирмы явилась программа правительства об утилизации автомобилей старше 10 лет. Рынок сбыта подержанных автомобилей резко стал уменьшаться. Прямо пропорционально на порядок уменьшился и спрос, перейдя в сектор рынка новых автомобилей. Фирма расширила сферу деятельности и через месяц охватила новый вид деятельности, став официальным дилером одного из автоконцернов. На протяжении всего этого времени накапливалась финансовая отчетность, на порядок отличающаяся от финансовых показателей, которые прогнозировались вначале. Большую работу проделали бизнес-аналитики фирмы, которые изменили формулу в линейной модели так, что прогноз стал краткосрочным и стал охватывать свежие финансовые показатели прошедшей недели, включив их в состав *плавающего диапазона*. Обновите формулу в ячейках C8:C17 как показано на рисунке. Найдите отличия в выражении формулы.

	A	B	C	D	E
1	x	y	линейная модель	наклон m	отрезок b
2	1	7	7,285714286	1,8857143	5,4
3	2	9	9,171428571		
4	3	12	11,05714286		
5	4	13	12,94285714		
6	5	14	14,82857143		
7	6	17	16,71428571		
8	7	17	18,6		
9	8	15	20,48571429		
10	9	12	22,37142857		
11	10	11			
12	11	13			
13	12	16			
14	13	18			
15	14	19			
16	15	17			
17	16	16			

	A	B	C	D	E
1	x	y	линейная модель	наклон m	отрезок b
2	1	7	7,285714286	1,8857143	5,4
3	2	9	9,171428571		
4	3	12	11,05714286		
5	4	13	12,94285714		
6	5	14	14,82857143		
7	6	17	16,71428571		
8	7	17	=ПРЕДСКАЗ(A8;B\$2:B7;A\$2:A7)		
9	8	15	20,48571429		



К пятому месяцу топ-менеджмент фирмы принял решение делать долгосрочные прогнозы на год вперед с интервалом в квартал. Для этого отчетными точками стали календарные кварталы, продолжительностью в 12 недель, из которых формируется долгосрочный прогноз на год. Начиная с ячейки В18, используется функция прогнозирования ТЕНДЕНЦИЯ, учитывающая весь предыдущий период. Ячейки диапазона А17:С22 должны отображать динамику долгосрочного прогноза, как показано на рисунке выше. Измените в диаграмме диапазон ячеек, по которым строятся графики на строки с 17 по 22 (квартальные точки 16, 28, 40, 52, 64 и 76 недели).

Вопрос: Как Вы считаете, какой вид прогноза наиболее точно описывается линейной регрессионной моделью, краткосрочный или долгосрочный?

Другой часто встречающейся на практике регрессионной моделью является экспоненциальная, описываемая уравнением $y=bt^x$. Значения экспоненциального тренда можно предсказывать при помощи функции РОСТ, которая вернет массив значений в качестве результата. При этом сама модель рассматривает вариант составления долгосрочного прогноза на достаточном количестве краткосрочных наблюдаемых данных.

РОСТ(массив1; массив2; массив3; константа) – массив1 содержит ссылку на диапазон ячеек с известными значениями зависимой наблюдаемой величины y , массив2 – независимой наблюдаемой величины x , массив3 содержит ссылку на новые значения x , для которых прогнозируются соответствующие значения y , возвращаемые в качестве результата вычислений функции РОСТ. Логический аргумент константа принимает одно из двух значений ИСТИНА или ЛОЖЬ, и соответствует необходимости подбора значения отрезка b . При ложном значении аргумента константа значение коэффициента b соответствует 1.

Задание №3: По истечению некоторого времени фирма выросла по объемам финансового оборота до крупной корпорации, охватив при этом всю сферу подержанных и

	A	B	C	D
1		новые	подержанные	оборот
2	год			
3	1	9,000	=10%*B3	
4	2	11,000		
5	3	13,000		
6	4	15,000		
7	5	17,000		
8	6	20,000		
9	7	23,000		
10	8	22,000		
11	9	25,000		
12	10	21,000		
13	11	24,000		
14	12	22,000		
15	13	20,000		
16	14	19,000		
17	15	21,000		

	A	B	C	D
1		новые	подержанные	оборот
2	год			
3	1	9,000	0,900	=D2+B3+C2
4	2	11,000	1,100	

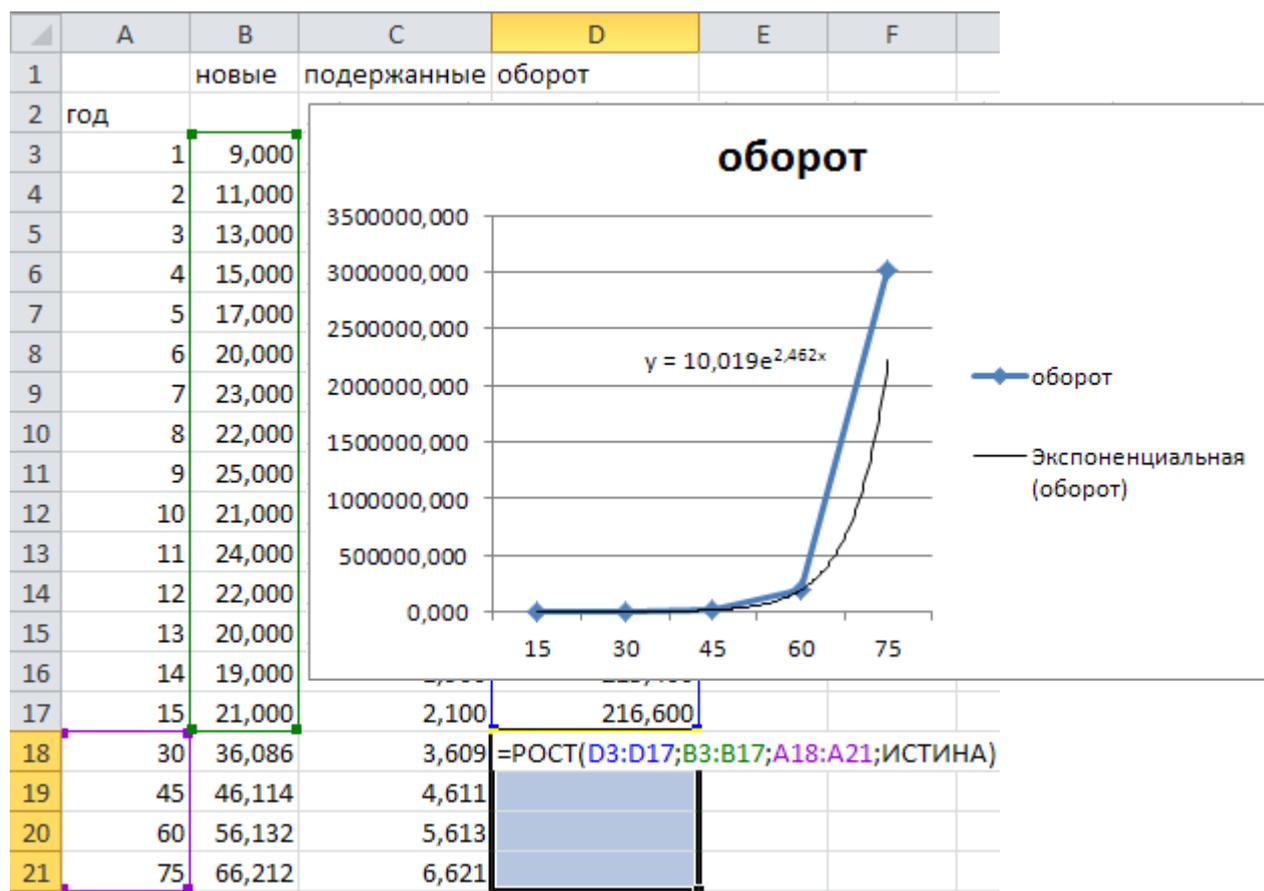
	A	B	C	D	E
1		новые	подержанные	оборот	
2	год				
3	1	9,000	0,900		9,000
4	2	11,000	1,100		20,900
5	3	13,000	1,300		35,000
6	4	15,000	1,500		51,300
7	5	17,000	1,700		69,800
8	6	20,000	2,000		91,500
9	7	23,000	2,300		116,500
10	8	22,000	2,200		140,800
11	9	25,000	2,500		168,000
12	10	21,000	2,100		=D11+B12+C11-B3-C2
13	11	24,000	2,400		

имеет постоянную планку выпуска, варьирующую в определенных границах. Известен спрос на новые автомобили за последние 15 лет. Каждый год около 10% новых приобретенных ранее автомобилей выходят на рынок подержанных машин. Действующая правительственная программа утилизации изымает из

оборота все ранее выпущенные автомобиле старше 10 лет. Необходимо на основе наблюдаемых данных построить вычислительную модель общего ежегодного оборота автомобильного рынка. На основе этих данных построить прогноз с периодом в 15 лет до 75 года.

После внесения наблюдаемых данных по выпуску новых машин в ячейках В3:В17, внесем и растиражируем в ячейках С3:С17 формулу расчета количества подержанных автомобилей. Общее количество находящихся в обороте автомобилей рассчитывается как накопительная сумма из количества новых машин текущего года и подержанных прошлого периода. Начиная с исчисления 10 первых лет и далее от ячейки D12 оборот рынка автомобилей рассчитывается также, но от суммы отнимается количество выпущенных машин 10 лет назад и подержанных, которые утилизируются, и на рынке появиться более не могут.

В ячейках В18:В21 рассчитывается прогноз выпускаемых каждые 15 последующих лет новых автомобилей (30, 45, 60 и 75 годы) с использованием функции ПРЕДСКАЗ. Ячейки D18:D21 рассчитываются при помощи экспоненциальной регрессионной модели, функцией РОСТ. В качестве экспериментальных значений независимых параметров берется выпуск новых автомобилей, который более всего коррелирует с общим оборотом машин на автомобильном рынке. Если модель многомерная по количеству параметров, то в качестве независимой наблюдаемой величины x берется наиболее коррелирующая (имеющая на наблюдаемом отрезке параметров самую адекватную степень изменений, как в сторону увеличения, так и в сторону уменьшения). Постройте график прогнозируемых объемов автомобильного рынка. Добавьте к ряду отображаемых данных на графике линию тренда.



Вопрос: Если экспоненциальная регрессионная модель $y=bt^x$, на основе которой будет прогнозироваться линейно коррелируемый (пропорциональный) набор значений, имеет множество возможных для выбора параметров, которые можно подставлять в качестве x , то лучше всего подойдет ступенчатая увеличивающаяся/уменьшающаяся величина или периодическая величина с математическим ожиданием в определенной точке?