

ОСНОВЫ СТАТИСТИКИ

Статистические функции анализа и
прогнозирования в MS Excel



В ходе изучения предстоящей лекции вы научитесь использовать функции, которые можно использовать для построения моделей и нахождения прогнозных значений. В частности, поговорим о функции ПРЕДСКАЗ, РОСТ, Тенденция, Отрезок, Наклон и т.д.

Функция ПРЕДСКАЗ в MS Excel позволяет с некоторой степенью точности предсказать будущие значения на основе существующих числовых значений, и возвращает соответствующие величины. Например, некоторый объект характеризуется свойством, значение которого изменяется с течением времени. Такие изменения могут быть зафиксированы опытным путем, в результате чего будет составлена таблица известных значений x и соответствующих им значений y , где x – единица измерения времени, а y – количественная характеристика свойства. С помощью функции ПРЕДСКАЗ можно предположить последующие значения y для новых значений x .

Функция ПРЕДСКАЗ используется как обычная формула. Если требуется предсказать сразу несколько значений, в качестве первого аргумента следует передать массив или ссылку на диапазон ячеек со значениями независимой переменной, а функцию ПРЕДСКАЗ использовать в качестве формулы массива. Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула.

Функция имеет синтаксис:

=ПРЕДСКАЗ(x ;известные_значения_у;известные_значения_х)

Описание аргументов:

- x – обязательный для заполнения аргумент, характеризующий одно или несколько новых значений независимой переменной, для которых требуется предсказать значения y (зависимой переменной). Может принимать числовое значение, массив чисел, ссылку на одну ячейку или диапазон;
- известные_значения_у – обязательный аргумент, характеризующий уже известные числовые значения зависимой переменной y . Может быть указан в виде массива чисел или ссылки на диапазон ячеек с числами;
- известные_значения_х – обязательный аргумент, который характеризует уже известные значения независимой переменной x , для которой определены значения зависимой переменной y .

Примеры использования функции ПРЕДСКАЗ в MS Excel.

Функция ПРЕДСКАЗ использует метод линейной регрессии, а ее уравнение имеет вид $y=ax+b$, где: Коэффициент a рассчитывается как $Y_{\text{ср.}}-bX_{\text{ср.}}$ ($Y_{\text{ср.}}$ и $X_{\text{ср.}}$ – среднее арифметическое чисел из выборок известных значений y и x соответственно).

Коэффициент b определяется по формуле:

Формула 22.1

$$b = \frac{\sum(x-x_{\text{ср}})(y-y_{\text{ср}})}{\sum(x-x_{\text{ср}})^2}$$

Пример 1. В таблице приведены данные экономически активного населения г. Алматы в динамике с 1991 по 2017 гг. Необходимо спрогнозировать динамику экономически активного населения до 2023 года.

Для этого в диапазоне A4:A36 набираем года с 1991 по 2023 гг., а в диапазоне B4:B30 вводим данные экономически активного населения г. Алматы относительно каждого года. Затем диапазону ячеек B\$4:B\$30 присваиваем имя «Выборка». Через ленту, перемещаемся во вкладку «Формулы». В группе инструментов на ленте «Библиотека функций» кликаем по кнопке «Другие функции». В открывшемся списке выбираем категорию «Статистические». Открытом перечне доступных элементов выбираем нужную нам функцию ПРЕДСКАЗ. В Аргументы функции: X – ссылка на ячейку A31, где A31:A36 диапазон ячеек с годами, для которых данные еще не определены; Известные_значения_у – «Выборка» диапазон ячеек, содержащих данные о экономически активном населении г. Алматы; Известные_значения_х – \$A\$4:\$A\$30 диапазон ячеек с годами, для которых уже известны данные. Кликаем на клавишу ОК.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула равная в нашем случае 846,41 тыс. человек в 2018 г. Используя маркер заполнения протянем формулу до B36, автоматически просчитав данные до 2023 года.

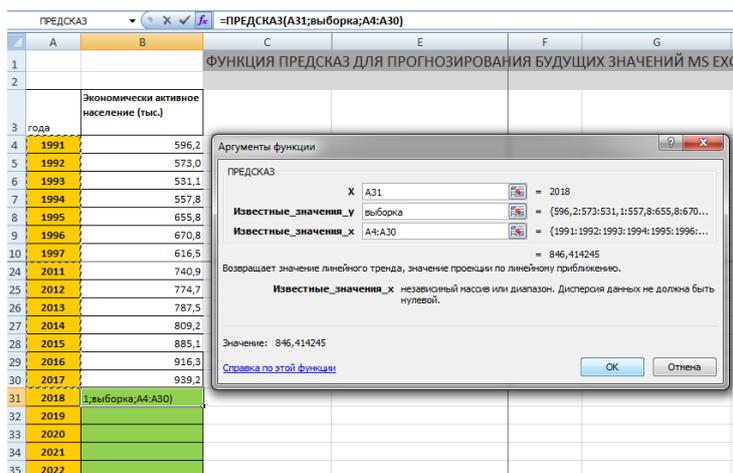


Рисунок 22.1. Диалоговое окно функции ПРЕДСКАЗ

Рассчитаем среднее значение на основании имеющихся и расчетных данных с помощью функции: =СРЗНАЧ(В4:В36).

Результат: Для этого посчитаем «Среднее значение на 2017 год» =СРЗНАЧ(В4:В30), которое будет равно 679,2 тыс. человек. Затем аналогично, рассчитаем «Предполагаемое среднее значение» =СРЗНАЧ(В4:В36) равное 715,0 тыс. человек. Учтем также «Возможные изменения» = E7 - E5, отняв от «Предполагаемого среднего значения» - «Среднее значение на 2017 год» результат составит 35,8 тыс. человек. Просчитаем «Возможные изменения в % отношении» = E10/ E5, а именно «Возможные изменения» / «Среднее значение на 2017 год» получим 5 %.

	A	B	C	E	F
1				ФУНКЦИЯ ПРЕДСКАЗ ДЛЯ ПРОГНОЗИРОВАНИ	
2					
3	года	Экономически активное население (тыс.)			
4	1991	596,2		Среднее значение на 2017 г.	
5	1992	573,0		679,2 =СРЗНАЧ(В4:В30)	
6	1993	531,1		Предполагаемое среднее значение	
7	1994	557,8		715,0 =СРЗНАЧ(В4:В36)	
8	1995	655,8			
9	1996	670,8		Возможные изменения	
10	1997	616,5		35,8 = E7 - E5	
26	2013	787,5		Возможные изменения в %	
27	2014	809,2		5% = E10/ E5	
28	2015	885,1			
29	2016	916,3			
30	2017	939,2			
31	2018	846,4			
32	2019	858,4			
33	2020	870,3			
34	2021	882,2			
35	2022	894,2			
36	2023	906,1			

Рисунок 22.2. Расчет среднего значения.

Можно сделать вывод о том, что динамика экономически активного населения до 2023 года сохранится. Анализ прогноза изменения динамики в MS Excel по функции ПРЕДСКАЗ.

Пример 2. Используем данные с первого примера, а именно, данные экономически активного населения г. Алматы с 1991 по 2017 гг. Предположим, каким будет динамика изменения на протяжении 6 последующих лет.

Как видно из исходной таблицы данных, в первые годы динамика изменения экономически активного населения г. Алматы была небольшой, затем она росла достаточно большими темпами, а на протяжении последних двух лет изменялась незначительно. Это свидетельствует о том, что основным фактором роста динамики на данный момент является не увеличение экономически активного населения, а развитие уже



существующих экономически активных граждан. В таких случаях рекомендуют использовать не линейную регрессию, а логарифмический тренд, чтобы результаты прогнозов были более точными.

Рассчитаем значения логарифмического тренда с помощью функции ПРЕДСКАЗ следующим способом: =ПРЕДСКАЗ(LN(A31:A36); «Выборка»; \$A\$4:\$A\$30),

где A31:A36 – диапазон ячеек с годами, для которых данные еще не определены;

«Выборка» диапазон ячеек, содержащих данные о экономически активные населения г. Алматы;

Известные значения x – \$A\$4:\$A\$30 диапазон ячеек с годами (в порядковом виде), для которых уже известны данные. Кликаем на клавишу ОК.

B31		fx =ПРЕДСКАЗ(LN(A31);Выборка;A4:A30)	
A	B	C	D
1			
2			
3	года	Экономически активное население (тыс.)	
4	1991	596,2	
5	1992	573,0	
6	1993	531,1	
7	1994	557,8	
8	1995	655,8	
9	1996	670,8	
10	1997	616,5	
28	2015	885,1	
29	2016	916,3	
30	2017	939,2	
31	2018	551,8	
32	2019	552,2	
33	2020	552,6	
34	2021	553,0	
35	2022	553,4	
36	2023	553,7	

Рисунок 22.3. Пример использования функции ПРЕДСКАЗ.

Как видно, в качестве первого аргумента представлен массив натуральных логарифмов последующих лет. Таким образом, получаем функцию логарифмического тренда, которая записывается как $y = \ln(x) + b$.

Для визуального сравнительного анализа построим простой график с использованием данных функции линейного тренда и функции логарифмического тренда.

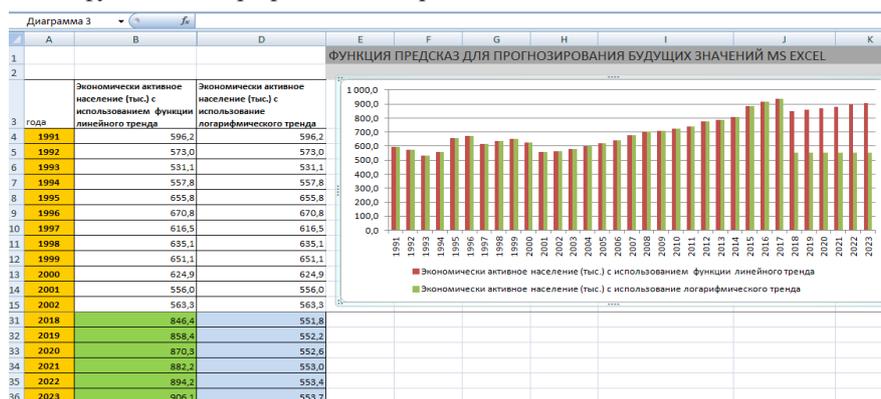


Рисунок 22.4. Графический анализ данных

Как видно, функцию линейной регрессии следует использовать в тех случаях, когда наблюдается постоянный рост какой-либо величины. В данном случае, на наш взгляд функция логарифмического тренда позволяет получить более правдоподобные данные.

Примечания: Второй и третий аргументы рассматриваемой функции должны принимать ссылки на непустые диапазоны ячеек или такие диапазоны, в которых число ячеек совпадает. Иначе функция ПРЕДСКАЗ вернет код ошибки #Н/Д.



Если одна или несколько ячеек из диапазона, ссылка на который передана в качестве аргумента x , содержит нечисловые данные или текстовую строку, которая не может быть преобразована в число, результатом выполнения функции ПРЕДСКАЗ для данных значений x будет код ошибки #ЗНАЧ!.

Статистическая дисперсия величин (можно рассчитать с помощью формул ДИСП.Г, ДИСП.В и др.), передаваемых в качестве аргумента известные_значения_x, не должна равняться 0 (нулю), иначе функция ПРЕДСКАЗ вернет код ошибки #ДЕЛ/0!.

Рассматриваемая функция игнорирует ячейки с нечисловыми данными, содержащиеся в диапазонах, которые переданы в качестве второго и третьего аргументов.

Для предсказания только одного будущего значения на основании известного значения независимой переменной функция ПРЕДСКАЗ используется как обычная формула. Если требуется предсказать сразу несколько значений, в качестве первого аргумента следует передать массив или ссылку на диапазон ячеек со значениями независимой переменной, а функцию ПРЕДСКАЗ использовать в качестве формулы массива.

Функция РОСТ и прогноз экспоненциального роста в MS Excel.

Функция РОСТ используется для расчета прогнозируемого экспоненциального роста на основе принимаемых на вход известных массивов данных X и Y , и возвращает массив значений для зависимой переменной Y на основе полученных новых данных для массива независимой переменной X .

Функция РОСТ имеет следующую синтаксическую следующий синтаксис:

`=РОСТ(известные_значения_y;[известные_значения_x];[новые_значения_x];[конст]).`

Описание аргументов:

- известные_значения_y – массив данных, элементы которого характеризуют значения зависимой переменной y в уравнении $y=bkx$. Аргумент обязателен для заполнения.
- [известные_значения_x] – массив данных, элементы которого соответствуют известным значениям независимой переменной x в записи $y=bkx$. Аргумент является необязательным.
- [новые_значения_x] – массив с новыми значениями независимой переменной x , на основе которых функция выполняет расчет новых значений зависимой переменной y . Аргумент необязателен для заполнения.
- [конст] – данные логического типа (ИСТИНА или ЛОЖЬ), определяющие значение константы b в уравнении $y=bkx$. По умолчанию (если аргумент явно не указан), а также при явном указании логического ИСТИНА, коэффициент b вычисляется обычным способом. Если данный параметр принимает значение ЛОЖЬ, дальнейшие расчеты проводятся для уравнения $y=kx$, поскольку значение b принимается равным единице.

Пример. Если рассмотреть выборку, например, экономически активного населения г. Алматы в динамике с 1991 по 2023 годы. При этом имея данные только по 2017 год, мы можем использовать метод прогнозируемого экспоненциального роста вычислить значения 6 последующих лет.

Для это выполним следующие действия: Выделяем диапазон ячеек В31:В36 и через ленту, перемещаемся во вкладку «Формулы». В группе инструментов на ленте «Библиотека функций» кликаем по кнопке «Другие функции». В открывшемся списке выбираем категорию «Статистические». Открытом перечне доступных элементов выбираем нужную нам функцию РОСТ.

Итак, `= РОСТ (Выборка;А4:А30;А31)`, где аргументы функции:

- известные_значения_y В4:В30 – диапазон известных значений, содержащих данные об экономически активном населения г. Алматы;
- известные_значения_x – А4:А30 – диапазон лет, для которых известны значения; новые_значения_x - А31 – значение, соответствующее номеру года, для которого необходимо выполнить расчет.

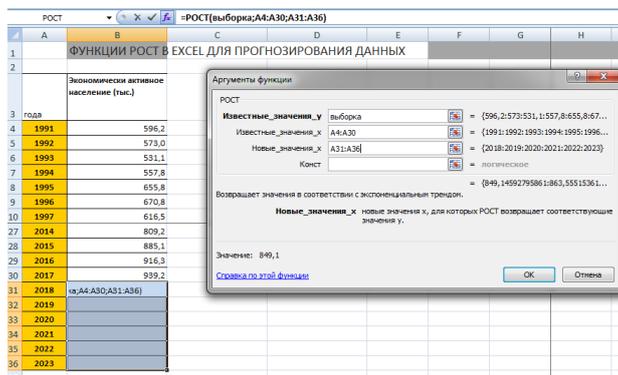


Рисунок 22.5. Диалоговое окно функции РОСТ.

Для ввода формулы используем комбинацию клавиш CTRL+SHIFT+Enter, так как формула должна выполняться в массиве. В результате получим:

	A	B	C	D	E
1		ФУНКЦИИ РОСТ В	EXCEL ДЛЯ ПРОГНОЗИРОВАНИЯ ДАННЫХ		
2		Экономически активное население (тыс.)			
3	год				
4	1991	596,2			
5	1992	573,0			
6	1993	531,1			
7	1994	557,8			
8	1995	655,8			
9	1996	670,8			
10	1997	616,5			
27	2014	809,2			
28	2015	885,1			
29	2016	916,3			
30	2017	939,2			
31	2018	849,1			
32	2019	863,6			
33	2020	878,2			
34	2021	893,1			
35	2022	908,3			
36	2023	923,7			

Рисунок 22.6 Результат использования функции РОСТ.

Мы видим явное несоответствие найденных величин диапазону уже известных значений. В MS Excel существует еще одна функция для прогнозирования на основе известных значений – Тенденция.

Примечание: Элементы массива известные_значения_у должны быть взяты из диапазона положительных чисел. При наличии отрицательных значений или значений, равных 0 (нулю), результатом выполнения функции РОСТ будет код ошибки #ЧИСЛО!

В качестве аргумента [известные_значения_x] может быть передано одно либо несколько множеств значений. Размерности множеств, передаваемых в качестве первого и второго аргументов должны совпадать, если используется единственная переменная. При вводе нескольких переменных в качестве аргумента известные_значения_у должен быть передан вектор. В MS Excel вектором считается интервал значений, высота которого составляет одну строку, либо ширина которого равна только одному столбцу).

Функция РОСТ интерпретирует каждый столбец или каждую строку массива [известные_значения_x] в качестве отдельной переменной, если массив известные_значения_у содержит только один столбец или только одну строку соответственно.

Если второй аргумент функции явно не указан, то по умолчанию используется массив данных {1;2;3;...;n}, размерность которого соответствует размерности массива известные_значения_у.

Массив [новые_значения_x] должен быть аналогичен по своей структуре массиву [известные_значения_x], то есть содержать строку либо столбец для каждого элемента массива известные_значения_у.

Если третий аргумент рассматриваемой функции явно не указан, считается, что он тождественен значению второго аргумента данной функции. Если второй и третий аргументы опущены, они оба являются массивами типа {1;2;3;...;n} с требованиями, указанными в пункте 4.

Если массив значений передается в качестве константы массива, по правилам записи массивов в Excel необходимо использовать знак «;» для разделения значений, содержащихся в одной строке, и знак «:» для разделения строк.



Примечание 2: Функция РОСТ часто используется для аппроксимации (упрощения) значений независимой (x) и зависимой (y) переменных экспоненциальной кривой.

Данная функция принадлежит к классу формул массивов, поэтому при ее использовании необходимо выделить соответствующее количество ячеек, а после ввода всех требуемых аргументов следует нажать сочетание клавиш Ctrl+Shift+Enter, для корректного отображения результатов.

В качестве функции экспоненциального роста используется уравнение типа $y=bkx$.

Функция ТЕНДЕНЦИЯ в Excel для прогнозирования данных.

Функция ТЕНДЕНЦИЯ в MS Excel используется при расчетах последующих значений для рассматриваемого события и возвращает данные в соответствии с линейным трендом. Функция выполняет аппроксимацию (упрощение) прямой линией диапазона известных значений независимой и зависимой переменных с использованием метода наименьших квадратов и прогнозирует будущие значения зависимой переменной Y для указанных последующих значений независимой переменной X . Рассматриваемая функция не используется для получения статистической характеристики модели тренда и математического описания.

Линейным трендом называется распределение величин в изучаемой последовательности, которое может быть описано функцией типа $y=ax+b$. Поскольку функция ТЕНДЕНЦИЯ выполняет аппроксимацию прямой линией, точность результатов ее работы зависит от степени разброса значений в рассматриваемом диапазоне.

Функция ТЕНДЕНЦИЯ используется наряду с прочими функциями прогноза в Excel (ПРЕДСКАЗ, РОСТ) и имеет следующий синтаксис:

`=ТЕНДЕНЦИЯ(известные_значения_y; [известные_значения_x]; [новые_значения_x]; [конст])`

Описание аргументов:

- известные_значения_y – обязательный аргумент, характеризующий диапазон исследуемых известных значений зависимой переменной y из уравнения $y=ax+b$.
- [известные_значения_x] – необязательный для заполнения аргумент, характеризующий диапазон известных значений независимой переменной x из уравнения $y=ax+b$.
- [новые_значения_x] – необязательный аргумент, характеризующий одно значение или диапазон данных, для которых необходимо определить соответствующие значения зависимой переменной y .
- [конст] – необязательный аргумент, принимающий на вход логические значения:

ИСТИНА (значение по умолчанию, если явно не указано обратное) – функция ТЕНДЕНЦИЯ выполняет расчет коэффициента b из уравнения $y=ax+b$ обычным методом

ЛОЖЬ – функция ТЕНДЕНЦИЯ использует упрощенный вариант уравнения – $y=ax$ (коэффициент $b = 0$).

Пример. Используем наш рассмотренный ранее пример. В таблице Excel содержатся данные экономически активного населения г. Алматы с 1991 по 2017 гг. Необходимо спрогнозировать на 2018-2023 годы.

Для прогноза данных экономически активного населения г. Алматы на 6 последующих лет выполним следующие действия: Через ленту, перемещаемся во вкладку «Формулы». В группе инструментов на ленте «Библиотека функций» кликаем по кнопке «Другие функции». В открывшемся списке выбираем категорию «Статистические». Открытом перечне доступных элементов выбираем нужную нам функцию ТЕНДЕНЦИЯ.

Итак, `=ТЕНДЕНЦИЯ(Выборка;A4:A30;A31)`, где аргументы функции:

- известные_значения_y B4:B30 – диапазон известных значений, содержащих данные о экономически активном населении г. Алматы;
- известные_значения_x – A4:A30 – диапазон лет, для которых известны значения; новые_значения_x – A31 – значение, соответствующее номеру года, для которого необходимо выполнить расчет. Кликаем на клавишу ОК.

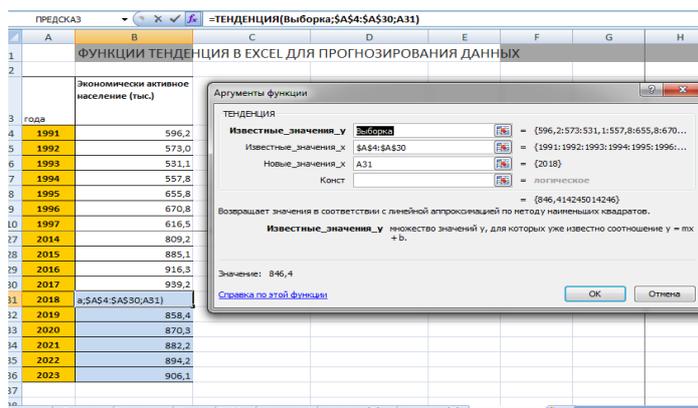


Рисунок 22.7 Диалоговое окно функции ТЕНДЕНЦИЯ.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула равная в нашем случае 846,41 тыс. человек в 2018 г. Аналогично рассчитаем все остальные величины.

Сравним полученные результаты нашего примера с использованием данных функции РОСТ и ТЕНДЕНЦИЯ. Как видно, синтаксические записи функций РОСТ и ТЕНДЕНЦИЯ идентичны, однако они используют различные алгоритмы для вычислений.

год	Экономически активное население (тыс.)	ФУНКЦИИ РОСТ
1991	596,2	
1992	573,0	
1993	531,1	
1994	557,8	
1995	655,8	
1996	670,8	
1997	616,5	
2014	809,2	
2015	885,1	
2016	916,3	
2017	939,2	
2018	846,4	849
2019	858,4	864
2020	870,3	878
2021	882,2	893
2022	894,2	908
2023	906,1	924

Рисунок 22.8 Сравнение результатов.

Как видно из данных изображенных на рисунке, значения функции ТЕНДЕНЦИЯ меньше функции РОСТ и более реальные относительно динамики изменения по годам. То есть, в данном примере функция ТЕНДЕНЦИЯ дает более точный прогноз и целесообразно использовать именно ее.

Примечания:

Рассматриваемая функция интерпретирует каждый столбец или каждую строку из диапазона известных значений x в качестве отдельной переменной, если аргументом известное $_u$ является диапазон ячеек из только одного столбца или только одной строки соответственно.

Аргументы [известное $_x$] и [новое $_x$] должны содержать одинаковое количество строк либо столбцов соответственно. Если новые значения независимой переменной явно не указаны, функция ТЕНДЕНЦИЯ выполняет расчет с условием, что аргументы [известное $_x$] и [новое $_x$] принимают одинаковые значения. Если оба эти аргумента явно не указаны, рассматриваемая функция использует массивы $\{1;2;3;\dots;n\}$ с размерностью, соответствующей размерности известное $_u$.

Не забывайте, что ТЕНДЕНЦИЯ является массивной функцией, поэтому после ее ввода не забудьте выполнить ее в массиве. Для этого ждем не просто Enter, а комбинацию клавиш Ctrl+Shift+Enter. Если в строке формул по краям функции появились фигурные скобки $\{ \}$, значит функция выполняется в массиве и все сделано правильно.



В качестве аргумента [известное_ x] могут быть переданы:

Только одна переменная, при этом два первых аргумента функции ТЕНДЕНЦИЯ могут являться диапазонами любой формы, но обязательным условием является одинаковая размерность (количество элементов).

Несколько переменных, при этом в качестве аргумента известное_у должен быть передан вектор значений (диапазон из только одной строки или только одного столбца).

Функция НАКЛОН для определения наклона линейной регрессии в Excel.

Функция НАКЛОН в MS Excel предназначена для определения угла наклона прямой, используемой для аппроксимации данных методом линейной регрессии, и возвращает значение коэффициента a из уравнения $y=ax+b$. Для определения наклона используются две любые точки на прямой. При этом вычисляется частное от деления длины отрезка, полученного при проецировании этих двух точек на ось Ординат (OY), на длину отрезка, образованного проекциями этих же двух точек на ось Абсцисс (OX).

Фактически, функция НАКЛОН вычисляет значение, которое характеризует скорость изменения данных вдоль линии регрессии. Зная наклон (коэффициент a) и значение коэффициента b можно рассчитать приближенные будущие значения какого-либо свойства y , которое меняется при изменении характеристики x .

Для расчета наклона линии регрессии используется рассмотренное ранее уравнение 1.

Функция НАКЛОН не может быть использована для анализа коллинеарных данных и будет возвращать код ошибки #ДЕЛ/0! в отличие от функции ЛИНЕЙН, которая использует иной алгоритм расчета и возвращает как минимум одно полученное значение.

Функция имеет следующий синтаксис:

=НАКЛОН(известные_значения_у;известные_значения_х)

Описание аргументов (все являются обязательными для заполнения):

- известные_значения_у – аргумент, принимающий массив числовых значений или ссылку на диапазон ячеек, которые содержат числа, характеризующие значения зависимой переменной y , которые определены для известных значений x ;
- известные_значения_х – аргумент, который может быть указан в виде массива чисел или ссылки на диапазон ячеек, содержащих числовые значения, которые характеризуют известные значения независимой переменной x .

Пример. Определить наклон аппроксимирующей прямой для показателей экономически активного населения г. Алматы с на протяжении нескольких лет (1991-2017гг.).

Для нахождения наклона используем следующую формулу из категории «Статистические» - НАКЛОН. Итак, = НАКЛОН (Выборка;A4:A30), где аргументы функции:

- известные_значения_у В4:В30 – диапазон известных значений, содержащих данные о экономически активном населении г. Алматы; характеризующие зависимую переменную y ;
- известные_значения_х – А4:А30 – диапазон ячеек с данными об годах, характеризующие независимую переменную x .

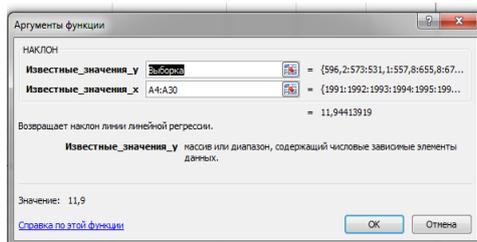


Рисунок 22.9 Диалоговое окно функции НАКЛОН.

В результате вычислений получим значение, которое, свидетельствует о том, что на протяжении обозначенного периода показатель экономически активного населения г. Алматы средним увеличивался примерно на 11,9 тыс. человек.

**Примечание:**

1. В качестве аргументов должны быть переданы массивы чисел либо ссылки на диапазоны ячеек с числовыми значениями или текстовыми строками, которые могут быть преобразованы к числам. Строки, не являющиеся текстовыми представлениями числовых данных, а также логические ИСТИНА и ЛОЖЬ в расчете не учитываются.
2. Если в качестве аргументов были переданы массивы, содержащие разное количество элементов, или ссылки на диапазоны с разным количеством ячеек, функция НАКЛОН вернет код ошибки #Н/Д. Аналогичный код ошибки будет возвращен в случае, если оба аргумента принимают пустые массивы или ссылки на диапазоны пустых ячеек.
3. Если оба аргумента ссылаются на нечисловые данные, функция НАКЛОН вернет код ошибки #ДЕЛ/0!.
4. Если в диапазоне, переданном в качестве любого из аргументов, содержатся пустые ячейки, они игнорируются в расчете. Однако ячейки, содержащие значение 0 (нуль) будут учтены.

Функции ОТРЕЗОК в MS Excel на оси ординат.

Функция ОТРЕЗОК в MS Excel используется для прогнозирования некоторого события. Она находит координаты точки пересечения графика с осью ординат (ОУ), построенного по координатам точек X (независимая переменная) и Y (зависимая переменная).

Данная функция имеет седующий синтаксис:

`=ОТРЕЗОК(известные_значения_y; известные_значения_x)`

Описание аргументов:

- известные_значения_y – массив данных или ссылка на диапазон ячеек, хранящих некоторые данные, которые характеризуют изменчивость одной величины относительно другой. Аргумент обязателен для заполнения.
- известные_значения_x – массив либо ссылка на ячейки, в которых хранятся данные, на основании которых выполняются вычисления значений, содержащихся в первом аргументе данной функции.

Прогноз экономически активного населения с учетом численности населения города Алматы в MS Excel.

Пример. Два массива данных характеризуют показатели одной величины (Y) относительно другой (X). Предположить, каким будет значение величины Y, если X примет значение 0 (нуль).

Если рассмотреть выборку демографических показателей, например, Численность населения (X) и Экономически активное население (Y) г. Алматы в динамике с 1991 по 2017 г.

Для нахождения координаты пересечения с осью Ординат введем следующую формулу из категории «Статистические» - ОТРЕЗОК.

Итак, `=ОТРЕЗОК(Выборка;A4:A30)`, где аргументы функции:

- известные_значения_y В4:В30 – диапазон известных значений, содержащих данные о экономически активном населении г. Алматы, характеризующие зависимую переменную y;
- известные_значения_x – А4:А30 – диапазон ячеек с численность населения г. Алматы, характеризующие независимую переменную x.

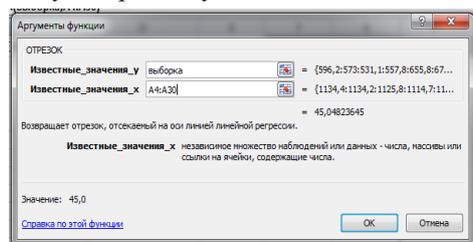


Рисунок 22.10 Диалоговое окно функции ОТРЕЗОК

В результате получим:

То есть, координата Y точки пересечения графика с осью Ординат равна примерно 45,0.



Построим график на основе известных значений:

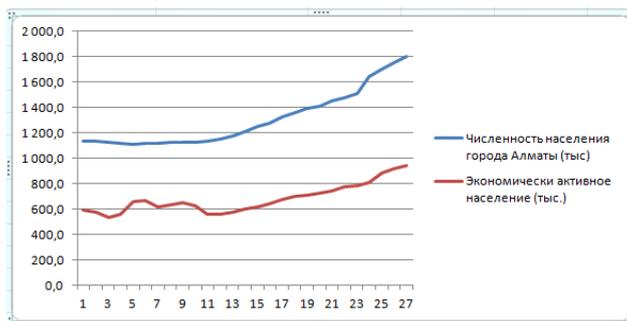


Рисунок 22.11 Оценка зависимости показателей

Как видно из графика, между двумя значениями установилась почти линейная зависимость.

Примечания:

В алгебре при построении графиков функций типа $y(x)$ составляют таблицу величин Y (зависимой) и X (независимой), при этом расчет значений Y выполняют на основании известных значений X путем их подстановки в уравнение функции. Аналогичный смысл имеют аргументы функции известные_значения_y и известные_значения_x, то есть их можно рассматривать как готовые таблицы известных значений, используемых для построения графика функции.

Функция ОТРЕЗОК принимает в качестве аргументов числовые данные, имена, содержащие числа, ссылки на диапазоны значений и массивы данных.

Данная функция пропускает объекты данных (ячейки), содержащие логические ИСТИНА или ЛОЖЬ, текстовые строки, пустые значения. Нулевые значения функцией ОТРЕЗОК учитываются.

Если в массивах данных, принимаемых в качестве аргументов рассматриваемой функции, содержится неодинаковое количество элементов, либо эти массивы являются пустыми, результатом работы функции ОТРЕЗОК будет код ошибки #Н/Д.

Функция ОТРЕЗОК завершит свою работу с кодом ошибки #ДЕЛ/0!, если в качестве аргумента известные_значения_y будет передан массив данных, содержащий только нули, а в качестве известные_значения_x – массив данных, все элементы которого являются единицами.

При вычислении точки пересечения определяется оптимальная линия регрессии, при этом координаты X и Y вычисляются с использованием функции СРЗНАЧ.

Функция ОТРЕЗОК позволяет определить значение переменной Y в тот момент, когда переменная X принимает значение 0 (нуль). Например, функция может быть использована для предсказания значения относительной влажности воздуха при температуре 0°C , если имеются известные значения этой величины при температурах 10°C , 15°C , 20°C и т. д.

Подведем итоги, сегодня мы рассмотрели основные функции. Которые можно использовать при проведении статистического анализа и нахождения прогнозного значения. В зависимости от исследования и поставленных перед Вами задачами Вы можете использовать ту или иную функцию, причем можете дополнить расчеты мерами центральной тенденции, тогда Ваше исследование будет иметь законченную работу. Для закрепления материалы, выполните домашнее задание.