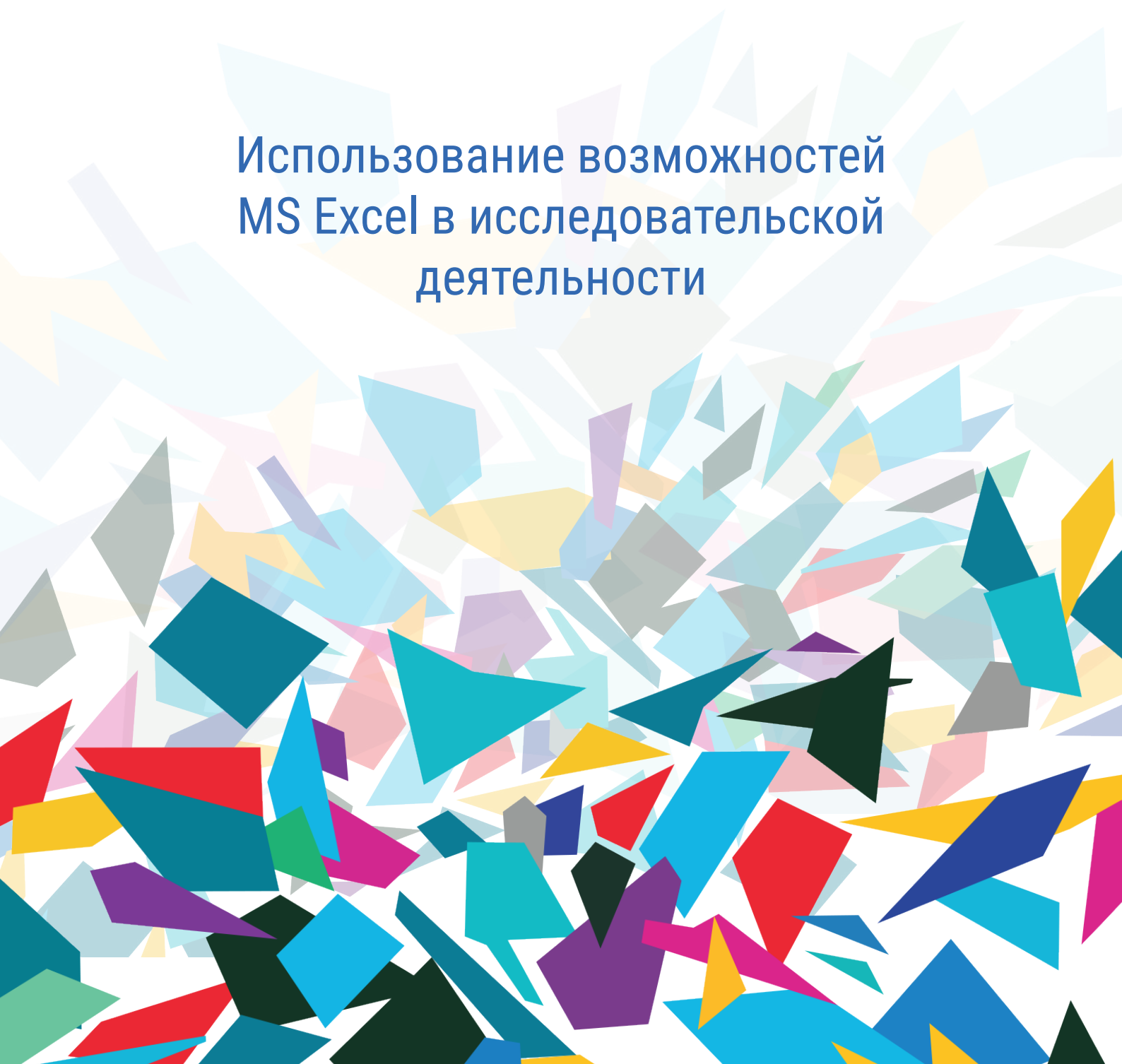


ОСНОВЫ СТАТИСТИКИ

Использование возможностей
MS Excel в исследовательской
деятельности





Используя полученные знания прочитанного курса мы, на примере решения задачи, можем извлечь информацию из данных, узнать их структуру и на основе полученной информации провести дальнейший анализ. Статистическая обработка нашей задачи позволит анализировать информацию с возможностью определения тенденции и прогноза по изучаемому явлению.

В нашем случае рассмотрим зависимость величины прожиточного минимума от уровня безработицы в РК в диапазоне 1997-2017 гг.

В этой лекции для решения нашей задачи будут использованы рассмотренные ранее несколько статистических функций приложения MS Excel:

НАИБОЛЬШИЙ;
НАИМЕНЬШИЙ;
РАНГ;
СРЗНАЧ;
СРЗНАЧЕСЛИ;
СРЗНАЧЕСЛИМН;
СЧЁТ;
СЧЁТЕСЛИ;
Стандартное отклонение;
Стандартная ошибка среднего;
Дисперсия выборки;
МОДА;
МЕДИАНА;
Корреляция;
Ковариации;
Доверительный интервал;
ПРЕДСКАЗ;
РОСТ;
ТЕНДЕНЦИЯ;
НАКЛОН;
ОТРЕЗОК;
Среднее значение выборки.

Если рассмотреть выборку показателей, нашей задачи, а именно величины прожиточного минимума и уровень безработицы в Республике Казахстан в диапазоне 1997-2017 гг., то на основании анализа этой выборки, мы можем вычислить среднее значение за 20 лет.

Для этого в диапазоне B2:B22 набираем данные величины прожиточного минимума по годам (в долларах США). Затем в диапазоне ячеек D2:D22 вводим уровень безработицы (в процентах). 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 СРЗНАЧ. В Аргументы функции: число1 вводим диапазон B2:B22 величины прожиточного минимума.

The screenshot shows an Excel spreadsheet with the following data:

Период	Величина прожиточного минимума, доллары США (Y)	Уровень безработицы, в процентах (X2)
1997	46,46	13,00
1998	47,46	13,10
1999	28,40	13,50
2000	28,19	12,80
2001	31,32	10,40
2002	31,06	9,50
2003	34,18	8,80
2004	39,59	8,40
2005	45,16	8,10
2006	66,70	7,80
2007	78,77	7,50
2008	102,78	6,60
2009	85,89	6,60
2010	91,53	5,80
2011	109,62	5,40
2012	112,77	5,30
2013	116,93	5,20
2014	16,41	5,00
2015	88,61	5,10
2016	63,16	5,00
2017	72,95	4,90

The 'Arguments of Function' dialog box for the AVERAGE function is open, showing:

- Function: СРЗНАЧ
- Argument 1: B2:B22 (selected)
- Argument 2: (empty)
- Result: = 63,73238095
- Value: 63,73



Рисунок 25.1 Исходные данные и диалоговое окно СРЗНАЧ.

Кликаем на клавишу ОК.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула среднего значение равная в нашем случае 62,73 долларов США.

Аналогично просчитаем среднее значение уровня безработицы (в процентах) в РК за 20 лет. В нашем случае он будет равен 8%.

Период	Величина прожиточного минимума, доллар США (X1)	Уровень безработицы, в процентах (X2)	Среднее значение
1997	46,46	13,00	63,73
1998	47,46	13,10	7,97
1999	28,40	13,50	
2000	28,19	12,80	
2001	31,32	10,40	
2002	31,06	9,30	
2003	34,28	8,80	
2004	39,89	8,40	
2005	45,26	8,10	
2006	66,70	7,80	
2007	78,77	7,30	
2008	102,78	6,60	
2009	85,83	6,60	
2010	91,53	5,80	
2011	109,62	5,40	
2012	112,77	5,30	
2013	116,93	5,20	
2014	16,41	5,00	
2015	88,61	5,10	
2016	63,16	5,00	
2017	72,95	4,90	

Рисунок 25.2 Результаты вычисления среднего значения.

СРЗНАЧЕСЛИ

Функция СРЗНАЧЕСЛИ имеет те же задачи, что и предыдущая, но в ней существует возможность задать дополнительное условие. В нашем случае, когда данные величины прожиточного минимума по годам (в долларах США) меньше 100 долларов США, среднее значение уровня безработицы меньше 10%.

Для этого в функцию СРЗНАЧЕСЛИ. В Аргументы функции: Диапазон - водим B2:B22 величины прожиточного минимума; Условие < 100.

Период	Величина прожиточного минимума, доллар США (X1)	Уровень безработицы, в процентах (X2)	Среднее значение	Среднее значение при условии < 100 прожиточный ур. и для <10 ур. безработицы	Количество условие < 100 и для <10 ур. безработицы
1997	46,46	13,00	63,73	СРЗНАЧЕСЛИ(B3:B22;"<100")	
1998	47,46	13,10	7,97	6,54	
1999	28,40	13,50			
2000	28,19	12,80			
2001	31,32	10,40			
2002	31,06	9,30			
2003	34,28	8,80			
2004	39,89	8,40			
2005	45,26	8,10			
2006	66,70	7,80			
2007	78,77	7,30			
2008	102,78	6,60			
2009	85,83	6,60			
2010	91,53	5,80			
2011	109,62	5,40			
2012	112,77	5,30			
2013	116,93	5,20			
2014	16,41	5,00			
2015	88,61	5,10			
2016	63,16	5,00			
2017	72,95	4,90			

Аргументы функции

СРЗНАЧЕСЛИ

Диапазон: B3:B22 = {47,46;28,4;28,19;31,32;31,06;3...}

Условие: "<100" = "<100"

Диапазон_усреднения: = ССЫЛКА = 53,11375

Вычисляет среднее (арифметическое) для ячеек, заданных указанным условием.

Диапазон: диапазон обрабатываемых ячеек.

Значение: 53,11

Справка по этой функции

ОК Отмена

Рисунок 25.3 Диалоговое окно функции СРЗНАЧЕСЛИ

Кликаем на клавишу ОК. Результат этого расчета – 53,11 долларов США.

Аналогично просчитаем среднее значение меньше 10% уровня безработицы (в процентах) в РК за 20 лет. В нашем случае он будет равен 6,64%.



Период	Величина прожиточного минимума, доллар США (Y)	Уровень безработицы, в процентах (X2)	Величина прожиточного минимума, доллар США	Среднее значение	Количество значений	Среднее значение при условии < 100 прожиточный ур. и для <10 ур. безработицы	Количество значений при условии < 100 прожиточный ур. и для <10 ур. безработицы
1997	46,46	13,00	63,73	21	53,11	17	
1998	47,46	13,10	7,97	21	6,54	16	
1999	38,40	13,50					
2000	38,19	12,80					
2001	31,32	10,40					
2002	31,06	9,30					
2003	34,28	8,80					
2004	39,89	8,40					
2005	45,26	8,10					
2006	66,70	7,80					
2007	78,77	7,50					
2008	102,78	6,60					
2009	85,83	6,60					
2010	91,33	5,80					
2011	109,62	5,40					
2012	112,77	5,30					
2013	116,93	5,20					
2014	16,41	5,00					
2015	88,61	5,10					
2016	63,16	5,00					
2017	72,99	4,90					

Рисунок 25.4 Расчетный лист задачи.

Рассчитаем значение МОДА для нашего примера. А именно, есть ли повторяющиеся значения уровня безработицы по годам.

Для этого в диапазоне ячеек D2:D22 вводим уровень безработицы (в процентах). 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 МОДА. В Аргументы функции: Число 1 – вводим D2:D22 – уровень безработицы в РК за период 1997–2017 гг.

Период	Величина прожиточного минимума, доллар США (Y)	Уровень безработицы, в процентах (X2)	МОДА
1998	47,46	13,10	=МОДА(D2:D22) 6,6
1999	38,40	13,50	
2000	38,19	12,80	
2001	31,32	10,40	
2002	31,06	9,30	
2003	34,28	8,80	
2004	39,89	8,40	
2005	45,26	8,10	
2006	66,70	7,80	
2007	78,77	7,50	
2011	109,62	5,40	
2012	112,77	5,30	
2013	116,93	5,20	
2014	16,41	5,00	
2015	88,61	5,10	
2016	63,16	5,00	
2017	72,99	4,90	

Рисунок 25.5. Расчет значения МОДА.

Кликаем на клавишу ОК.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула и он равен 6,6.

Аналогично просчитаем МОДА величины прожиточного минимума по годам (в долларах США) в РК за 20 лет. В нашем случае, так как в массиве нет повторяющихся значений, то функция вернет значение ошибки #Н/Д.



N2		fx =МОДА(B2:B22)			
A	B	D	E	F	N
Период	Величина прожиточного минимума,долларов США (Y)	Уровень безработицы, в процентах (X2)			МОДА
1997	46,46	13,00		Величина прожиточного минимума,долларов США	#Н/Д
1998	47,46	13,10		Уровень безработицы, в процентах	6,60
1999	28,40	13,50			
2000	28,19	12,80			
2001	31,32	10,40			
2002	31,06	9,30			
2003	34,28	8,80			
2004	39,89	8,40			
2005	45,26	8,10			
2006	66,70	7,80			
2007	78,77	7,30			
2011	109,62	5,40			
2012	112,77	5,30			
2013	116,93	5,20			
2014	16,41	5,00			
2015	88,61	5,10			
2016	63,16	5,00			
2017	72,95	4,90			

Рисунок 25.6 Второй вариант значения МОДА.

Рассчитаем значение МЕДИАНА для нашего примера. А именно серединное значения множества чисел по годам величины прожиточного минимума (в долларах США) в РК за 20 лет.

Для этого в функцию МЕДИАНА водим следующие Аргументы функции: Число 1 – диапазон ячеек B2:B22 величины прожиточного минимума по годам.

O3		fx =МЕДИАНА(D2:D22)			
A	B	D	E	F	O
Период	Величина прожиточного минимума,долларов США (Y)	Уровень безработицы, в процентах (X2)			МЕДИАНА
1997	46,46	13,00		Величина прожиточного минимума,долларов США	63,16
1998	47,46	13,10		Уровень безработицы, в процентах	7,30
1999	28,40	13,50			
2000	28,19	12,80			
2001	31,32	10,40			
2002	31,06	9,30			
2003	34,28	8,80			
2004	39,89	8,40			
2005	45,26	8,10			
2006	66,70	7,80			
2007	78,77	7,30			
2011	109,62	5,40			
2012	112,77	5,30			
2013	116,93	5,20			
2014	16,41	5,00			
2015	88,61	5,10			
2016	63,16	5,00			
2017	72,95	4,90			

Рисунок 25.7 Расчет значения медианы выборки.

Кликаем на клавишу ОК.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула, который равен 63,16 долларов США.

Аналогично просчитаем МЕДИАНА для уровня безработицы (в процентах) в РК в динамике 1997–2017 гг. В нашем случае он будет равен 7,30%.

O3		fx =МЕДИАНА(D2:D22)			
A	B	D	E	F	O
Период	Величина прожиточного минимума,долларов США (Y)	Уровень безработицы, в процентах (X2)			МЕДИАНА
1997	46,46	13,00		Величина прожиточного минимума,долларов США	63,16
1998	47,46	13,10		Уровень безработицы, в процентах	7,30
1999	28,40	13,50			
2000	28,19	12,80			
2001	31,32	10,40			
2002	31,06	9,30			
2003	34,28	8,80			
2004	39,89	8,40			
2005	45,26	8,10			
2006	66,70	7,80			
2007	78,77	7,30			
2011	109,62	5,40			
2012	112,77	5,30			
2013	116,93	5,20			
2014	16,41	5,00			
2015	88,61	5,10			
2016	63,16	5,00			
2017	72,95	4,90			



Рисунок 25.8 Расчет медианы второй выборки данных.

Рассчитаем значения Стандартного отклонения и Стандартную ошибку среднего для нашего примера. В категории функций «Статистические», выбираем нужную нам функцию СТАНДОТКЛОН. В Аргументы функции: Число 1 – водим диапазон ячеек B2:B22 величины прожиточного минимума по годам.

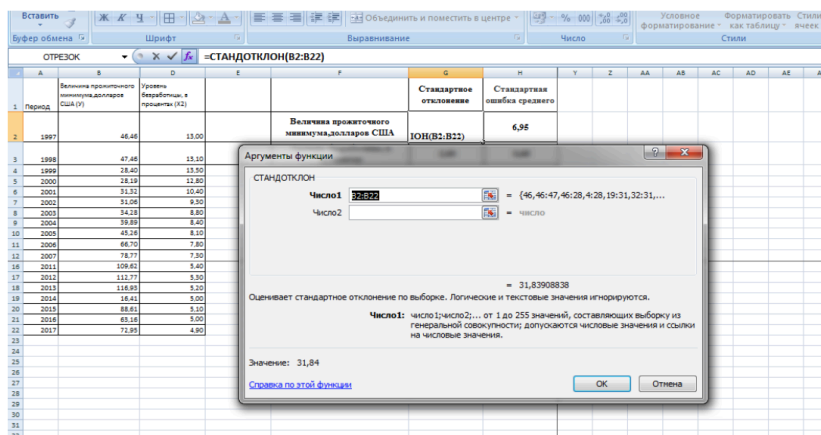


Рисунок 25.9 Вычисление стандартного отклонения.

Кликаем на клавишу ОК.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула. Таким образом, стандартное отклонение равно 31,84 долларов США.

Для того чтобы рассчитать «Стандартную ошибку среднего» значение «Стандартного отклонения» делим на корень функции Счет, в аргументы которой водим B2:B22 величины прожиточного минимума по годам.

Аналогично просчитаем Стандартное отклонение и Стандартную ошибку среднего для уровня безработицы (в процентах) в РК в динамике 1997–2017 гг. В нашем случае он будет равен 3% и 0,65% соответственно.

Рассчитаем НАИБОЛЬШИЙ и НАИМЕНЬШИЙ, а также Ранг для нашего примера.

В списке функций выбираем категорию «Статистические». Открытом перечне доступных элементов выбираем нужную нам функцию =НАИБОЛЬШИЙ В Аргументы функции Массив – B2:B22 величины прожиточного минимума по годам, k – это порядковый номер величины, а именно 2008 года (12). Затем находим функцию =НАИМЕНЬШИЙ, в аргументы которой Массив водим B2:B22 величины прожиточного минимума по годам, k – это порядковый номер величины -12, а именно 2008 года. Аналогично вводим =РАНГ – в аргументы Массив водим имя «Выборка». Число – B13 величину прожиточного минимума в 2008 году.

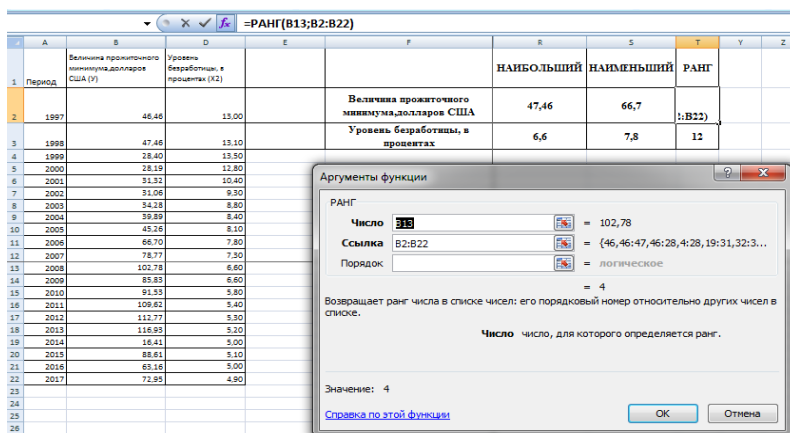


Рисунок 25.10 Вычисление значения ранг.

Кликаем на клавишу ОК.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула.

Наибольшее – 47,46, наименьшее значение – 66,7 а также Ранг = 4.



Вычисление доверительного интервала в Microsoft Excel.

Теперь давайте рассчитаем доверительный интервал на нашем примере. Была проведена выборка за 20 лет величины прожиточного минимума и уровня безработицы в Республике Казахстан в диапазоне 1997–2017 гг., вследствие которых были получены различные результаты, занесенные в таблицу. Это и есть наша совокупность. Стандартное отклонение равно 31,84 долларов США и 3% данное значение было посчитано нами ранее). Нам нужно рассчитать доверительный интервал при уровне доверия 97%.

Для этого в диапазоне выделяем ячейку, куда будет выводиться результат обработки данных. Кликаем по кнопке «Вставить функцию». Появляется Мастер функций. Переходим в категорию «Статистические» и выделяем наименование «ДОВЕРИТ.НОРМ». После этого кликаем по кнопке «ОК».

Открывается окошко аргументов. Его поля закономерно соответствуют наименованиям аргументов. Устанавливаем курсор в первое поле – «Альфа». Тут нам следует указать уровень значимости. Как мы помним, уровень доверия у нас равен 97%. В то же время мы говорили, что он рассчитывается таким путем:

$$(1 - \text{«Альфа»}) * 100$$

Значит, чтобы посчитать уровень значимости, то есть, определить значение «Альфа» следует применить формулу такого вида: $(1 - \text{уровень доверия}) / 100$.

То есть, подставив значение, получаем: $(1 - 97) / 100$.

Путем нехитрых расчетов узнаем, что аргумент «Альфа» равен 0,03. Вводим данное значение в поле.

Как известно, по условию стандартное отклонение равно 31,84 долларов США. Поэтому в поле «Стандартное отклонение» просто записываем это число.

В поле «Размер» нужно ввести количество наблюдаемых лет. Как мы помним, их 20. Но чтобы автоматизировать формулу и не редактировать её каждый раз при проведении нового испытания, давайте зададим данное значение не обычным числом, а при помощи функции СЧЁТ. Итак, устанавливаем курсор в поле «Размер», а затем кликаем по треугольнику, который размещен слева от строки формул.

Появляется список недавно применяемых функций. Если функция СЧЁТ применялась вами недавно, то она должна быть в этом списке. В таком случае, нужно просто кликнуть по её наименованию. В обратном же случае, если вы её не обнаружите, то переходите к пункту «Другие функции...».

Появляется уже знакомый нам Мастер функций. Опять перемещаемся в группу «Статистические». Выделяем там наименование «СЧЁТ». Кликаем по кнопке «ОК».

Появляется окно аргументов вышеуказанной функции.

Группа аргументов «Значения» представляет собой ссылку на диапазон, в котором нужно рассчитать количество заполненных числовыми данными ячеек. Всего может насчитываться до 255 подобных аргументов, но в нашем случае понадобится лишь один.

Устанавливаем курсор в поле «Значение1» и, нажав левую кнопку мыши, выделяем на листе диапазон, который содержит нашу совокупность. Затем его адрес будет отображен в поле. Кликаем по кнопке «ОК».

The screenshot shows an Excel spreadsheet with the following data:

Период	Величина прожиточного минимума, долларов США (P)	Уровень безработицы, в процентах (K2)	Стандартное отклонение	Стандартная ошибка среднего	Среднее значение	Доверительный интервал
1997	46,46	13,00	31,84	6,95	63,73	СЧЁТ(B2:B22)
1998	47,46	12,19	3,00	0,65	7,97	1,420
1999	28,40	12,80				
2000	28,19	12,80				
2001	31,22	10,40				
2002	31,06	9,20				
2003	34,28	8,80				
2004	39,89	8,40				
2005	42,14	8,10				
2006	66,70	7,80				
2007	78,77	7,50				
2008	102,78	6,60				
2009	82,83	6,60				
2010	91,83	5,80				
2011	109,62	5,40				
2012	112,77	5,20				
2013	116,93	5,20				
2014	16,41	5,00				
2015	88,61	5,10				
2016	83,16	5,00				
2017	72,85	4,90				

The 'Arguments of the Function' dialog box for CONFIDENCE.NORM is open, showing the following values:

- Альфа: 0,03
- Станд_откл: СТАНДОТКЛОН(B2:B22) = 31,83908838
- Размер: СЧЁТ(B2:B22) = 21
- Result: = 15,0774813

Возвращает доверительный интервал для среднего генеральной совокупности.
Альфа - уровень значимости, используемый для вычисления доверительного интервала - число, большее 0 и меньше 1.

Значение: 15,077

Справка по этой функции

Buttons: OK, Отмена

Рисунок 25.11 Расчет доверительного интервала.

После этого приложение произведет вычисление и выведет результат в ту ячейку, где она находится сама. В нашем конкретном случае формула получилась такого вида:

=ДОВЕРИТ.НОРМ(0,03;111,98;СЧЁТ(B2:B22)), где B2:B22 величины прожиточного минимума по годам. Общий результат вычислений составил 15,07 долларов США.



Но это ещё не все. Как мы помним, граница доверительного интервала вычисляется путем сложения и вычитания от среднего выборочного значения результата вычисления ДОВЕРИТ.НОРМ. Таким способом рассчитывается соответственно правая и левая граница доверительного интервала. Само среднее выборочное значение было рассчитано нами ранее при помощи функции СРЗНАЧ. Оно равно 63,73 долларов США.

Производим расчет правой границы доверительного интервала. Для этого выделяем отдельную ячейку, ставим знак «=» и складываем содержимое элементов листа, в которых расположены результаты вычислений функций СРЗНАЧ и ДОВЕРИТ.НОРМ. Для того, чтобы выполнить расчет, кликаем на клавишу Enter. В нашем случае получилась следующая формула: =J2+U2. Результат вычисления: 78,81 долларов США.

Таким же образом производим вычисление левой границы доверительного интервала, только на этот раз от результата вычисления СРЗНАЧ отнимаем результат вычисления функции ДОВЕРИТ.НОРМ. Получается формула для нашего примера следующего типа: =J2-U2. Результат вычисления: 48,6 долларов США.

Аналогичное вычисление проведем и для уровня безработицы в Республике Казахстан в диапазоне 1997-2017гг.

Корреляция и ковариация в MS EXCEL.

Теперь давайте попробуем посчитать коэффициент корреляции на нашем примере. Имеем таблицу, в которой расписана в отдельных колонках рассматриваемые года с 1997 по 2017 г. и величины прожиточного минимума по годам. Нам предстоит выяснить степень зависимости величины прожиточного минимума в РК от периода лет.

Выделяем ячейку, в которой должен выводиться результат расчета. Кликаем по кнопке «Вставить функцию», которая размещается слева от строки формул.

В списке, который представлен в окне Мастера функций, ищем и выделяем функцию КОРРЕЛ. Кликаем на кнопку «ОК».

Открывается окно аргументов функции. В поле «Массив1» вводим координаты диапазона ячеек одного из значений, зависимость которого следует определить. В нашем случае это будут значения в колонке «года». Для того, чтобы внести адрес массива в поле, просто выделяем все ячейки с данными в вышеуказанном столбце.

В поле «Массив2» нужно внести координаты второго столбца. У нас величины прожиточного минимума в РК. Точно так же, как и в предыдущем случае, заносим данные в поле. Кликаем на кнопку «ОК».

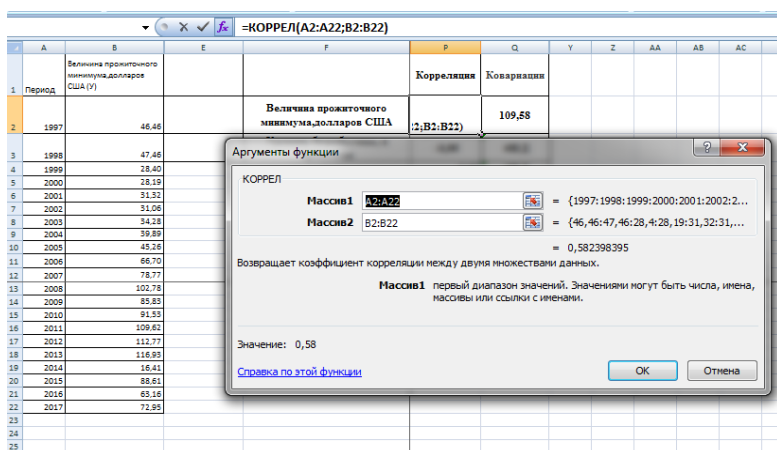


Рисунок 25.12 Расчет тесноты связи между показателями.

Как видим, коэффициент корреляции в виде числа появляется в заранее выбранной нами ячейке. В данном случае он равен 0,58, что является недостаточно высоким признаком зависимости одной величины от другой.

Использование MS EXCEL для расчета ковариации.

Теперь давайте попробуем посчитать коэффициент ковариации на нашем примере.

Кликаем по кнопке «Вставить функцию», которая размещается слева от строки формул.

В списке, который представлен в окне Мастера функций, ищем и выделяем функцию КОВАР(). Жмем



на кнопку «ОК».

Открывается окно аргументов функции. В поле «Массив1» вводим координаты диапазона ячеек одного из значений, зависимость которого следует определить. В нашем случае это будут значения в колонке «года». Для того, чтобы внести адрес массива в поле, просто выделяем все ячейки с данными в вышеуказанном столбце.

В поле «Массив2» нужно внести координаты второго столбца. У нас величины прожиточного минимума в РК от периода лет. Точно так же, как и в предыдущем случае, заносим данные в поле.

Кликаем на кнопку «ОК».

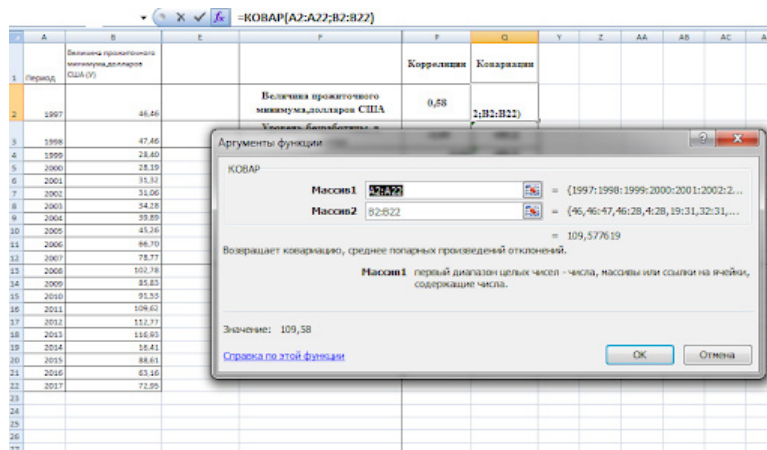


Рисунок 25.13 Расчет ковариации.

Как видим, ковариации в виде числа появляется в заранее выбранной нами ячейке. В данном случае он равен 109,58.

Функция ПРЕДСКАЗ для прогнозирования будущих значений.

В таблице приведены данные уровня безработицы в Республике Казахстан в диапазоне 1997–2017 гг. Необходимо спрогнозировать динамику уровня безработицы до 2023 года.

Для этого используем функцию ПРЕДСКАЗ. В Аргументы функции вводим: X – ссылка на ячейку A25, где A25:A30 диапазон ячеек с годами, для которых данные еще не определены; Известные_значения_y – «Выборка» диапазон ячеек, содержащих данные о уровня безработицы; Известные_значения_x – \$A\$4:\$A\$24 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 2,9 % в 2018 г. Используя маркер заполнения протянем формулу до B30, автоматически просчитав данные до 2023 года.

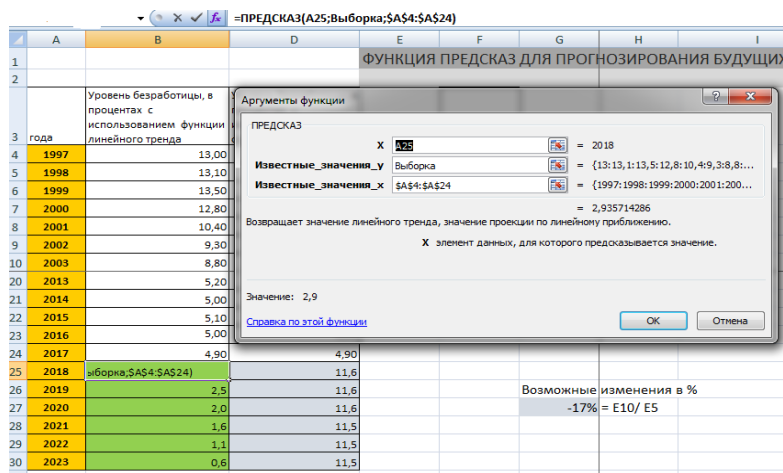


Рисунок 25.14 Демонстрация функции ПРЕДСКАЗ.



Рассчитаем среднее значение на основании имеющихся и расчетных данных с помощью функции: =СРЗНАЧ(В4:В30).

Для этого посчитаем «Среднее значение на 2017 год» =СРЗНАЧ(В4:В24), которое будет равно 8%. Затем аналогично, рассчитаем «Предполагаемое среднее значение» =СРЗНАЧ(В4:В30), равное 6,6%. Учтем также «Возможные изменения» = G7 -GE5, отняв от «Предполагаемого среднего значения» – «Среднее значение на 2017 год» результат составит -1,4%. Просчитаем «Возможные изменения в % отношении» = G10/ G5, а именно «Возможные изменения» / «Среднее значение на 2017 год» получим -17%.

G7		fx		=СРЗНАЧ(В4:В30)			
A	B	E	F	G	H	I	
1		ФУНКЦИЯ ПРЕДСКАЗ ДЛЯ ПРОГНОЗИРОВАНИЯ БУДУЩИХ ЗНА					
2							
3	года	Уровень безработицы, в процентах с использованием функции линейного тренда					
4	1997	13,00			Среднее значение на 2017 г.		
5	1998	13,10			8,0 =СРЗНАЧ(В4:В24)		
6	1999	13,50			Предполагаемое среднее значение		
7	2000	12,80			6,6 =СРЗНАЧ(В4:В30)		
8	2001	10,40					
9	2002	9,30			Возможные изменения		
10	2003	8,80			-1,4 = G7 - G5		
26	2019	2,5			Возможные изменения в %		
27	2020	2,0			-17% = G10/ G5		
28	2021	1,6					
29	2022	1,1					
30	2023	0,6					

Рисунок 25.15. Аналитическая часть исследования.

Можно сделать вывод о том, что динамика снижения уровня безработицы в РК до 2023 года сохранится.

Если рассмотреть выборку из нашего примера, величины прожиточного минимума в РК в динамике с 1997 по 2023 годы. При этом имея данные только по 2017 год, мы можем использовать метод прогнозируемого экспоненциального роста вычислить значения б последующих лет.

Для это выполним следующие действия: Выделяем диапазон ячеек C25:C30 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;A4:A24;A25:A30), где аргументы функции:

известные_значения_у C4:C24 – диапазон известных значений, содержащих данные о величине прожиточного минимума;

известные_значения_x – A4:A24 – диапазон лет, для которых известны значения; новые_значения_x – A25:A30 – значение, соответствующее номеру года, для которого необходимо выполнить расчет.

x		fx		=РОСТ(Выборка;A4:A24;A25:A30)		
A	C	F	G	H	I	J
1						
2						
3	года	Величина прожиточного минимума,долларов США, РОСТ				
4	1997	46,46				
14	2007	78,77				
15	2008	102,78				
16	2009	85,83				
17	2010	91,53				
18	2011	109,62				
19	2012	112,77				
20	2013	116,93				
21	2014	16,41				
22	2015	88,61				
23	2016	63,16				
24	2017	72,95				
25	2018	97				
26	2019	101				
27	2020	106				
28	2021	111				
29	2022	116				
30	2023	116				

Аргументы функции

РОСТ

Известные_значения_у: Выборка = {46;46;47;46;28;4;28;19;31;32;3...}

Известные_значения_x: A4:A24 = {1997;1998;1999;2000;2001;200...}

Новые_значения_x: A25:A30 = {2018;2019;2020;2021;2022;2023}

Конст: = логическое

Возвращает значения в соответствии с экспоненциальным трендом.

Известные_значения_y: множество значений y, которые уже известны для соотношения y=b^m*x, массив или диапазон положительных чисел.

Значение: 92

Справка по этой функции

OK Отмена

Рисунок 25.16 Демонстрация использования функции РОСТ.

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



C29		fx {=РОСТ(Выборка;A4:A24;A25:A30)}			
	A	C	F	G	H
1					
2					
3	года	Величина прожиточного минимума, долларов США, РОСТ			
4	1997	46,46			
14	2007	78,77			
15	2008	102,78			
16	2009	85,83			
17	2010	91,53			
18	2011	109,62			
19	2012	112,77			
20	2013	116,93			
21	2014	16,41			
22	2015	88,61			
23	2016	63,16			
24	2017	72,95			
25	2018	92			
26	2019	97			
27	2020	101			
28	2021	106			
29	2022	111			
30	2023	116			

Рисунок 25.17 результат использования функции РОСТ.

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

Используем наш рассмотренный ранее пример. В таблице MS Excel содержатся данные величины прожиточного минимума в РК с 1997 по 2017 г. Необходимо спрогнозировать на 2018–2023 годы.

Для прогноза данных величины прожиточного минимума в РК на 6 последующих лет выполним следующие действия: в категории функций «Статистические» выбираем нужную нам функцию ТЕНДЕНЦИЯ.

Итак, =ТЕНДЕНЦИЯ(Выборка;A4:A24;A25), где аргументы функции:

- известные_значения_у В4:В24 – диапазон известных значений, содержащих данные о величины прожиточного минимума;
- известные_значения_х – А4:А24 – диапазон лет, для которых известны значения; новые_значения_х – А25 – значение, соответствующее номеру года, для которого необходимо выполнить расчет.

Кликаем на клавишу ОК.

Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула равная в нашем случае 96,6 долларов США 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.

G25		fx		
	A	B	C	F
1		ФУНКЦИИ ТЕНДЕНЦИЯ и РОСТ В EXCEL ДЛЯ ПРОГНОЗ		
2				
3	года	Величина прожиточного минимума, долларов США, ТЕНДЕНЦИЯ	Величина прожиточного минимума, долларов США, РОСТ	
4	1997	46,46	46,46	
20	2013	116,93	116,93	
21	2014	16,41	16,41	
22	2015	88,61	88,61	
23	2016	63,16	63,16	
24	2017	72,95	72,95	
25	2018	96,6	92	
26	2019	99,6	97	
27	2020	102,6	101	
28	2021	105,6	106	
29	2022	108,6	111	
30	2023	111,5	116	

Рисунок 25.18 Анализ результатов использования функций Тенденция и РОСТ.



Как видно из данных значения функции ТЕНДЕНЦИЯ отличаются от значений 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.

Определить наклон аппроксимирующей прямой для показателей величины прожиточного минимума на протяжении нескольких лет (1997–2017гг.).

Для нахождения наклона используем следующую формулу из категории «Статистические» – НАКЛОН.

Итак, = НАКЛОН (B2:B22;A2:A22), где аргументы функции:

известные_значения_y B2:B22 – диапазон известных значений, содержащих данные о величине прожиточного минимума; характеризующие зависимую переменную y;

известные_значения_x – A4:A30 – диапазон ячеек с данными об годах, характеризующие независимую переменную x.

В результате вычислений получим:

B23		fx =НАКЛОН(B2:B22;A2:A22)			
A	B	E	F	G	
Период	Величина прожиточного минимума,долларов США				
3	2003	34,28			
9	2004	39,89			
0	2005	45,26			
1	2006	66,70			
2	2007	78,77			
3	2008	102,78			
4	2009	85,83			
5	2010	91,53			
6	2011	109,62			
7	2012	112,77			
8	2013	116,93			
9	2014	16,41			
0	2015	88,61			
1	2016	63,16			
2	2017	72,95			
3	угол	3,0	наклон		

Рисунок 25.19 Результат использования функции НАКЛОН.

Полученное значение свидетельствует о том, что на протяжении обозначенного периода показатель величины прожиточного минимума в Республики Казахстан в среднем увеличивался примерно на 3 доллара США.

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

Рассмотрим выборку зависимости величины прожиточного минимума от уровня безработицы в РК в диапазоне 1997–2017 гг.

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

Итак, = ОТРЕЗОК (B2:B22;A2:A22), где аргументы функции:

известные_значения_y B2:B22 – диапазон известных значений, содержащих данные о величине прожиточного минимума; характеризующие зависимую переменную y;

известные_значения_x – A4:A30 – диапазон ячеек с данными об годах, характеризующие независимую переменную x.

В результате получим – 117,2 долларов США.

Рассмотрим теперь корреляционно- регрессионный анализ.

Рассмотрим проблему, зависит ли величина прожиточного минимума, в долларах от среднегодового обменного курса и уровня безработицы. Для анализа и построения модели нам необходимо прежде всего рассмотреть коэффициент корреляции, т.е. оценить существует ли связь между независимым фактором и независимыми факторами, кроме того необходимо из модели исключить интрекоррелированные факторы, если таковые имеются. Для начала используем уже известный нам инструмент Данные-Анализ данных. Если есть необходимость настройте его. ПО итогам работы данного инструмента получили следующие



результаты.

Период	Величина прожиточного минимума, доллар США (Y)	Среднегодовой обменный курс доллара США (X1)	Уровень безработицы, в процентах (X2)	Величина прожиточного минимума, доллар США (Y)	Среднегодовой обменный курс доллара США (X1)	Уровень безработицы, в процентах (X2)
1997	46,46	75,44	13,00	1		
1998	47,46	78,30	13,10	0,107804001	1	
1999	28,40	119,52	13,50	-0,631262974	-0,578413678	1
2000	28,19	142,13	11,80			
2001	31,32	146,74	10,40			
2002	31,06	153,28	9,30			
2003	34,28	149,58	8,80			
2004	39,89	136,04	8,40			
2005	45,26	132,88	8,10			
2006	66,70	126,09	7,80			
2007	78,77	122,55	7,30			
2008	102,78	120,30	6,60			
2009	85,83	147,50	6,60			
2010	91,93	147,35	5,80			
2011	109,62	146,62	5,40			
2012	112,77	149,11	5,30			
2013	116,93	152,13	5,20			
2014	16,41	179,19	5,00			
2015	88,61	221,73	5,10			
2016	63,16	342,16	5,00			
2017	72,95	326,00	4,90			

Рисунок 25.20 Исходные данные и корреляционная матрица.

Связь между величиной прожиточного минимума (Y) и среднегодового курса валюты связь отсутствует так как коэффициент корреляции равен 0,11, что свидетельствует о слабой зависимости между факторами. Коэффициент корреляции между факторами величина прожиточного минимума и уровень безработицы равен -0,63, что свидетельствует об обратной связи, причем данная связь является так сказать заметной, т.е. при увеличении уровня безработицы снижается величина прожиточного минимума, согласно шкале о который мы уже говорили. Кроме этого рассмотрим существует ли связь между независимыми факторами, т.е. среднегодовым обменным курсом валюты и уровнем безработицы. Да она имеется, причем имеем обратная и заметная, и равна -0,59. Отсюда следует, что необходимо исключить один из показателей, исключаем то, который дает слабую межфакторную связь с независимым фактором, т.е. среднегодовой обменный курс валюты. В итоге мы с вами будем рассматривать уравнение парной регрессии. Для этого применим инструмент Анализ Данных – Регрессия. В результате получим оценочную таблицу показателей уравнения регрессии.

1	Вывод итогов						
2							
3	Регрессионная статистика						
4	Множественный R	0,631262974					
5	R-квадрат	0,398492943					
6	Нормированный R-квадрат	0,366834677					
7	Стандартная ошибка	25,33489947					
8	Наблюдения	21					
9							
10	Дисперсионный анализ						
11		df	SS	MS	F	Значимость F	
12	Регрессия	1	8079,265485	8079,265485	12,58732682	0,00214865	
13	Остаток	19	12195,2855	641,8571314			
14	Итого	20	20274,55098				
15							
16		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
17	У-пересечение	117,1760515	16,04610398	7,302461185	6,32104E-07	83,59116989	150,7609331
18	Уровень безработицы, в процентах (X2)	-6,704403116	1,889702184	-3,547862289	0,00214865	-10,65959524	-2,749210989
19							

Рисунок 25.21 Вывод итоговых результатов по модели.

Из таблицы видно что $a=117,18$, $b=-6,7$, т.е. модель будет выглядеть следующим образом $Y=117,18-6,7 \cdot X+E$, т.е. при увеличении уровня безработицы на 1 % снижается величина прожиточного минимума на 6,7 доллара. Множественный R, R-квадрат и нормированный R, характеризуют модель не с лучшей стороны, но нельзя сказать, что совсем плохо, так как множественный R больше чем 0,5, но меньше 0,7. Поэтому оценим модель по F-критерию Фишера, согласно данного критерия модель статистически значима так как величина значимость F меньше чем 0,05. Параметры модели так же статистически значимы по критерию Стьюдента. Для этого, мы сравниваем значения в столбце P-значение также с 0,05 и видим, что величины для каждого параметра модели меньше чем 0,05. Аналогично можно провести анализ вышеуказанных показателей сравнив значения в столбце F и t статистика с табличными. получим аналогичный результат. Оценивая доверительный интервал, мы видим, что наши параметры попадают в данные интервалы, и они являются корректными. Что можно сказать о модели в целом, в принципе показатели не совсем худшие, но не выдерживают требований к оценке модели, кроме того ошибка аппроксимации достаточно большая и составляет 18%. что свидетельствует о плохом качестве модели.



Таким образом, фактор уровень безработицы не является ключевым при формировании величины прожиточного минимума. Т.е., все операции, проделанные выше становятся бессмысленными в плане проведения прогнозирования (функции ПРЕДСКАЗ, РОСТ и т.д.). Что необходимо для решения проблемы – нужно пересмотреть перечень независимых факторов, и перед проведением каких-либо операций необходимо рассчитать коэффициент корреляции. Возникнет вопрос, а почему в качестве примера взят материал, который не принес на должного результат? Этим самым хотелось показать, что в профессиональной и исследовательской деятельности не всегда исследование дает положительный результат. Наличие отрицательного результата, дает свой ответ на поставленные вопросы, т.е. чего не нужно делать и рекомендации обществу какие показатели не стоит рассматривать при моделировании того или иного процесса.

В доказательство сделанных нами выводов построим линию тренда, причем отметим на графике уравнение регрессии и значение R квадрат и уравнение регрессии, выбираем линейную форму тренда, из рисунка видно, что фактические данные и линия регрессии не отражают сущности изучаемого явления, и к тому же мы видим не очень высокий коэффициент детерминации, если рассмотреть исходный график зависимости показателей, то имеются незначительные выбросы, которые ухудшают качество модели, что подтверждает выше сказанное: показатель прожиточного минимума не зависит от уровня инфляции.

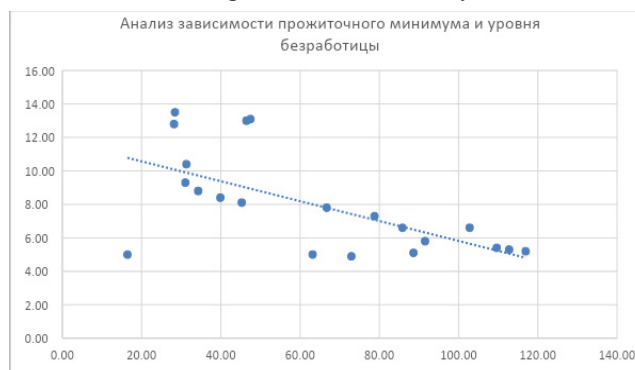


Рисунок 25.22 Линия тренда.

На этом мы завершаем изучение курса «Основы статистики», основой которого являлась книга Джозефа Хили «Основы статистики». Для более глубокого понимания были использованы дополнительные источники – это лекции профессора Шалабанова А.К. и Роганова Д.А, ну и конечно же не обошлось и без официальных интернет источников. Книга сама по себе очень интересная, и может стать настольной, если у вас есть интерес к статистическим исследованиям, в ней много красочных примеров, правда они далеки от нашей реальности, нас волнуют немного другие проблемы. Для более глубокого понимания необходимы базовые знания математики, освоив основы статистики вы сможете уже сегодня попытаться дать анализ вашим доходам, или рассчитать среднюю стоимость поездки за границу, а возможно и предсказать на какой уровень зарплаты вы выйдете, если освоите дополнительные компетенции, например, основы статистики. Статистика – она везде, она вокруг нас, надо ее просто понять и заговорить с ней на языке цифр.