Парный регрессионный

1. Стохастическая зависимость, функция регрессии

2. Корреляция

3. Линейная парная регрессия

4. Коэффициент детерминации и средняя ошибка аппроксимации

5. Предпосылки регрессионного анализа. Условия Гаусса-Маркова

6. Свойства оценок коэффициентов классической линейной модели регрессии. Теорема Гаусса-Маркова

7. Значимость коэффициента регрессии и его доверительный интервал

8. Интервальная оценка функции регрессии. Прогноз на основе функции регрессии

9. Решение типовых задач с помощью MS Exsel

9.1 Построение линейной регрессии с помощью Мастер функций *f(x)*,

Статистические, ЛИНЕЙН (первый способ)

9.2 Построение линейной регрессии с помощью инструментов Сервис,

Анализ данных, Регрессия (второй способ)

9.3 Интерпретация результатов регрессионного анализа

9.4 Вычисление прогнозного значения *у*

9.5 Оценка точности прогноза

9. 6 Решение задач регрессии без помощи специальных средств

**1 Стохастическая зависимость, функция регрессии**

Стохастическая зависимость – зависимость между случайными величинами, при которой изменение закона распределения одной из них происходит под влиянием изменения другой.

В отличие от функциональной зависимости каждому из значений х независимой переменной X сопоставляется не одно конкретное значение другой переменной, а случайная величина Y(х). Например, при одном и том же количестве вносимых удобрений, урожайность является случайной величиной.

Поскольку случайная величина характеризуется своим законом распределения вероятностей, стохастическая зависимость понимается как соответствие х → Y(х), при котором величина Y(х) - распределение вероятностей случайной величины.

Одной из основных характеристик случайной величины служит ее среднее значение (или математическое ожидание), которое является числом. Таким образом, при заданной стохастической зависимости х → Y(х) возникает обычная числовая функция х → МY (х) (= Мх Y), значениями которой являются осредненные значения случайной величины Y.

В статистике часто вместо термина «случайная величина» используется термин «генеральная совокупность». При этом характеристики, относящиеся к случайным величинам, называются генеральными, чтобы отличать их от выборочных характеристик.

Регрессия – величина, выражающая зависимость среднего значения случайной величины y от значений случайной величины х.

Функцией регрессии, соответствующей стохастической зависимости х → Y(х), где х - число (значение независимой переменной); Y(х) - случайная величина, соответствующая числу х, называется функция, которая числу х ставит в соответствие условное математическое ожидание Мх(Y) = М Y (х), т.е. функция х → Мх (Y).

На практике, как правило, полная (генеральная) совокупность значений показателей отсутствует. Мы имеем только их выборки.

Рассмотрим пример стохастической зависимости и функции регрессии для случая, когда случайная величина Y является дискретной. Случайная величина называется дискретной, если ее множество значений счетно.

*Пример 1. Гипотетический пример стохастической зависимости*

Пусть переменная X обозначает недельные семейные доходы, а переменная Y - недельные семейные расходы на потребление.

В табл. 1 приведены данные о соответствии между значениями X и Y. Приведенное соответствие является стохастическим.

Таблица 1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Недельные доходы семьи X, $ | | | | | | | | | |
|  | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 260 |
| Недельные расходы семьи на потребление Y, $ | 55 | 65 | 79 | 80 | 102 | 110 | 120 | 135 | 137 | 150 |
|  | 60 | 70 | 84 | 93 | 107 | 115 | 136 | 137 | 145 | 152 |
|  | 65 | 74 | 90 | 95 | 110 | 120 | 140 | 140 | 155 | 175 |
|  | 70 | 80 | 94 | 103 | 116 | 130 | 144 | 152 | 165 | 178 |
|  | 75 | 85 | 98 | 108 | 118 | 135 | 145 | 157 | 175 | 180 |
|  | - | 88 | - | 113 | 125 | 140 | - | 160 | 189 | 185 |
|  | - | - | - | 115 | - | - | - | 162 | - | 191 |
| Итого | 325 | 462 | 445 | 707 | 678 | 750 | 685 | 1043 | 966 | 1211 |

В табл. 2 указаны условные вероятности р(Y / Хi) для данных табл. 1.

Таблица 2

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | X | | | | | | | | | |
|  | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 260 |
| Условные вероятности |  |  |  |  |  |  |  |  |  |  |
| р(Y / Хi) |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | - |  | - |  |  |  | - |  |  |  |
|  | - | - | - |  | - | - | - |  | - |  |
| Условное среднее МхY | 65 | 77 | 89 | 101 | 113 | 125 | 137 | 149 | 161 | 173 |

Полем корреляции называется графическое изображение стохастической зависимости, при этом точки плоскости соответствуют значениям (xi ,yi). Функция регрессии, соответствующая данной стохастической зависимости, задается формулой х → Мх Y.

На рис. 1 крестиками отмечены соответствующие значения Мх Y функции регрессии, а точками - значения Y(х). График этой функции называется линией регрессии, которую можно изобразить соединением соответствующих точек.

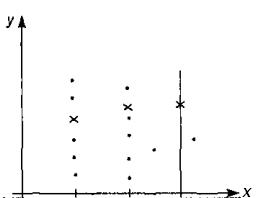


Рисунок 1

При исследовании стохастических зависимостей на практике оценку функции регрессии проводят по данным выборки.

Выборочной функцией регрессии называется оценка функции регрессии, построенная по выборке. Существуют различные методы построения такой оценки (будут рассматриваться в дальнейшем).

**2 Корреляция**

Корреляция – величина, отражающая наличие связи между явлениями, процессами и характеризующими их показателями.

Исследование зависимости между экономическими показателями начинается с корреляционного анализа.

Основная задача: выявление взаимосвязи между случайными переменными путем точечной и интервальной оценок парных (частных) коэффициентов корреляции, вычисление и проверка значимости множественных коэффициентов корреляции и коэффициента детерминации.

С помощью корреляционного анализа осуществляются:

1. отбор факторов, оказывающих наиболее существенное влияние на результативный признак на основании измерения тесноты связи между ними;

2. обнаружение ранее неизвестных причинных связей.

Но: корреляция непосредственно не выявляет причинных связей между параметрами, а устанавливает численное значение этих связей и достоверность суждений об их наличии.

При проведении корреляционного анализа вся совокупность данных рассматривается как множество переменных (факторов), каждая из которых содержит n наблюдений.

Для заданной пары случайных величин X и Y рассматривается выборка их значений (хiуi), i=1,...,n. На основании этой выборки мы будем делать заключение о взаимосвязи величин.

Ковариацией случайных величин X и Y называется число

cov (Х, Y) = М ((X - МХ)( Y - МХ)) = М{Х Y) - МХМ Y.

Ковариация - это ожидаемое значение произведения отклонений величин от их ожидаемых значений. Оценкой ковариации по выборке значений величин служит выборочная ковариация.

cov (,) =, (2.2)

где , - выборочные средние, которые находятся по следую­щим формулам:

= , = . (2.3)

При этом выборку значений случайной величины X мы пони­маем двояко; либо это набор случайных величин {Х1, Х2, …, Хn }, каждая из которых имеет такое же распределение вероятностей, как и сама X, либо это набор чисел {x1, x2,…,xn }, в котором хi (i=1,…,n) является значением случайной величины X i. Кроме того, предполагается, что пары (Xi, Yi) для i≠ j- независимые случайные величины.

Ковариация имеет положительный знак, если при увеличении (уменьшении) одной из переменных вторая переменная тоже увеличивается (уменьшается) (рис. 2, а). Отрицательный знак ковариации означает, что переменные изменяются в противоположные стороны (рис. 2, б).

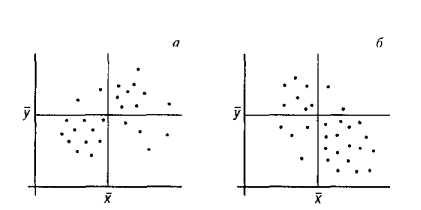


Рисунок 2

Если случайные величины Xи Yнезависимы, то cov (X,Y )= 0. Еcли cov (X,Y )≠ 0, то X и Y- зависимые случайные величины.

Для нормально распределенных случайных величин понятия независимости и некоррелированности (cov (X,Y)=0) совпадают.

Дисперсия случайной величины Х характеризует отклонение случайной величины от ее среднего значения. Она является частным случаем ковариации: =cov (X,X). Ее несмещенной оценкой служит исправленная выборочная дисперсия = (X,X)= .

Выборочная дисперсия =также является оценкой дисперсии , но смещенной.

Для измерения тесноты линейной статистической зависимости между двумя случайными величинами X и Y используется без­размерная статистическая характеристика, называемая коэффи­циентом корреляции:

Выборочным коэффициентом корреляции называется величина

где s2x, s2y- (исправленные) выборочные дисперсии величин X и Y (или оценки дисперсий ,этих величин).

Ковариация и корреляция представляют, по существу, одну и ту же информацию о взаимной вариации величин. Однако корреляция является более удобной безразмерной величиной.

Выборочный коэффициент корреляции обладает следующими свойствами:

1) -1 < r <1;

2) чем ближе | r | к единице, тем теснее линейная связь;

3) при | r | = 1 между переменными существует линейная зависимость.

Замечание: Выборочный коэффициент корреляции является оценкой генерального коэффициента корреляции.

Так как для построения оценки коэффициента корреляции мы пользуемся выборкой, возникает вопрос о значимости коэффициента корреляции.

Для этого тестируется нулевая гипотеза H0 об отсутствии линейной связи между величинами X и Y. Проверка гипотезы H0 проводится на основе t-статистики Стьюдента. Расчетное значение критерия определяется по формуле

С учетом заданного уровня значимости а и числа степеней свободы, равного (n - 2),по таблице Стьюдента находится критическое значение tкр.

Если |tр| >tкр , то коэффициент корреляции считается значимым. Нулевая гипотеза отвергается. Это означает, что выборочные данные на уровне значимости aне противоречат наличию линейной связи между X и Y.

Теснота линейной связи одной из случайных величин с совокупностью остальных величин проверяется с помощью выборочного множественного коэффициента корреляции, который вычисляется по формуле

где |R| - определитель матрицы R; Rjj- определитель матрицы, полученной из матрицы Rвычеркиванием j-го столбца и j-й строки. Крышка над индексом j в формуле означает пропуск этого индекса.

Для вычисления выборочного множественного коэффициента корреляции Rj:12,..^j..m в Ехсеl можно воспользоваться матрицей R-1, обратной к матрице R парных коэффициентов корреляции.

Величина является обратной величиной j-го элемента главной диагонали матрицы R-1 .

Выборочным коэффициентом детерминации называется квадрат выборочного множественного коэффициента корреляции:

Выборочный коэффициентом детерминации характеризует долю вариации зависимой переменной, обусловленной регрессией или изменчивостью объясняющих элементов.

Этот коэффициент принимает значения от 0 до 1.

Если коэффициент =0, то в выборке отсутствует связь.

Чем ближе коэффициент к 1, тем лучше регрессия описывает зависимость между объясняющими и зависимой переменными.

Он может увеличиваться при добавлении в набор новых переменных, и не увеличивается при исключении каких-либо переменных.

**3 Линейная парная регрессия**

Пусть между показателями X и Y существует стохастическая зависимость. Предполагая эту зависимость линейной, запишем ее в следующем виде: Y=β0 +β1X +ε

В этом уравнении X предполагается детерминированной (объ­ясняющей) переменной, β0, β1, - коэффициенты, а ε – случайная составляющая, вместе с которой переменная Y также является случайной величиной. На случайную величину εнакладывается условие обращения в нуль ее математического ожидания М ε = 0.

С учетом этого условия математическое ожидание Y принимает вид

Это уравнение и является теоретическим уравнением рег­рессии.

Таким образом, уравнение регрессии задает зависимость сред­него значения случайной переменной Y от значений детермини­рованной переменной X.

Основная задача регрессионного анали­за - построение оценок коэффициентов β0, β1 по данным выбор­ки и описание качества оценок.

Выборочным уравнением регрессии называется уравнение

где b0, b1- выборочные оценки коэффициентов ,,а ŷ - вы­борочная оценка MxY.

Одним из основных методов построения названных оценок является метод наименьших квадратов (МНК).

Обозначим через xi,yi(i = 1,...,n) выборку значений объема nпоказателей X и Y. Тогда, согласно предположению, между эти­ми значениями существует зависимость yi=β0 +β1xi +εi , i = 1,...,n.

Соответствующая выборочная оценка теоретического уравне­нии регрессии имеет вид ŷ = b0+b1xi, i = 1,...,n

В этих соотношениях нам пока известны только значения xi,величины b0 и b1, а вместе с ними и ŷ будут вычисляться по МНК.

Согласно МНК минимизируется сумма квадратов отклонений,

которая рассматривается как функция от b0,b1. Заметим, что это соотношение включает уже и переменные уi.

Записав необходи­мые условия минимума, приравняем нулю первые частные произ­водные

В силу того что S является квадратичной функцией от b0,b1, необходимые условия (которые находятся выше) являются также и достаточными условиями минимума.

С учетом минимизированной суммы квадратов отклонения предыдущее условие можно записать в виде системы нормальных уравнений:

где соответствующие выборочные средние определяются по формулам:

; ; ;

Решение линейной системы нормальных уравнений имеет вид:

Выборочным коэффициентом регрессии Y по X называется коэффициент b1 в уравнении ŷ = b0+b1xi.

Он показывает, на сколько единиц в среднем изменяется переменная Y при увеличение переменной Х на одну единицу.

Коэффициент регрессии b1 можно так же записать в виде:

b1 = ,

где – выборочная ковариация, а – исправленная выборочная дисперсия переменной Х.

Коэффициент b0 не имеет экономической интерпретации, хотя, если переменная Х может принимать нулевое значение, b0 является соответствующим значением переменной и может интерпретироваться как значение Y при х=0.

Уравнение регрессии (выборочное) записывается так же в виде:

*Пример 2.*

Для иллюстрации введенных понятий и расчета МНК-оценок рассмотрим выборку из n = 30 наблюдений (табл.3).

Таблица3

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| I | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| xi | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 260 |
| Y | 60  70  75 | 70  74  85 | 84  90  98 | 80  103  113 | 107  116  125 | 115  130  140 | 136  140  144 | 137  152  160 | 145  165  189 | 152  178  185 |

Для построения функции регрессии проведем необходимые вычисления, которые запишем в виде таблицы 4.

Таблица 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| I |  |  |  |  |
| 1 | 2 | 3 | 4 | 5 |
| 1 | 80 | 60 | 6400 | 4800 |
| 2 | 80 | 70 | 6400 | 5600 |
| 3 | 80 | 75 | 6400 | 6000 |
| 4 | 100 | 70 | 10000 | 7000 |
| 5 | 100 | 74 | 10000 | 7400 |
| 6 | 100 | 85 | 10000 | 8400 |
| 7 | 120 | 84 | 14400 | 10080 |
| 8 | 120 | 90 | 14400 | 10800 |
| 9 | 120 | 98 | 14400 | 11760 |
| 10 | 140 | 80 | 19600 | 11200 |
| 11 | 140 | 103 | 19600 | 14420 |
| 12 | 140 | 113 | 19600 | 15820 |
| 13 | 160 | 107 | 25600 | 17120 |
| 14 | 160 | 116 | 25600 | 18560 |
| 15 | 160 | 125 | 25600 | 20000 |
| 16 | 180 | 115 | 32400 | 20700 |
| 17 | 180 | 130 | 32400 | 23400 |
| 18 | 180 | 140 | 32400 | 25200 |
| 19 | 200 | 136 | 40000 | 27200 |
| 20 | 200 | 140 | 40000 | 28000 |
| 21 | 200 | 144 | 40000 | 28800 |
| 22 | 220 | 137 | 48400 | 30140 |
| 23 | 220 | 152 | 48400 | 33440 |
| 24 | 220 | 160 | 48400 | 35200 |
| 25 | 240 | 145 | 57600 | 34800 |
| 26 | 240 | 165 | 57600 | 39600 |
| 27 | 240 | 189 | 57600 | 45360 |
| 28 | 260 | 152 | 67600 | 39520 |
| 29 | 260 | 178 | 67600 | 48100 |
| 30 | 260 | 185 | 67600 | 48100 |
| Итого | 3618 | 5100 | 966000 | 674800 |
| Среднее | 120,6 | 170 | 32200 | 22493,33 |
|  |  |  |  |

Из формул и данных таблицы 4 получаем следующее уравнение регрессии:

Из этого уравнения следует, что при увеличении показателя Х на одну единицу показатель увеличивается на 0,603 единицы.

Графическое изображение уравнения регрессии изображено на рисунке 3.

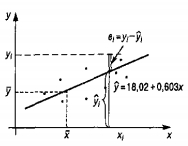


Рисунок 3

Вычисление коэффициентов регрессионных уравнений производятся по формулам ; ; ; ; .

Однако решение задач эконометрики, как правило, не возможно без использования информационных технологий и ПВЭМ. Необходимые вычисления осуществляем в среде Exсel.

Рассмотрим теперь другую выборку значений (таблица 5).

Таблица 5

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| I | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 260 |
| Y | 55  65  75 | 65  80  88 | 79  90  98 | 80  103  115 | 102  116  125 | 110  130  140 | 120  140  145 | 135  152  162 | 137  155  175 | 150  180  191 |

Для этой выборки уравнение регрессии имеет вид .

Данное уравнение значительно отличается от предыдущего, полученного по первой выборке. Для сравнения найдем уравнение регрессии и построим его график для всей (генеральной) совокупности из таблицы 1.

В этом случае уравнение регрессии имеет вид.

Эта прямая, согласно формуле , должна проходить через точку .

На рисунке 4 изображены три прямые регрессии, построенные по двум выборкам и самой генеральной совокупности.

Значительные расхождения между ними можно объяснить недостаточными репрезентативностями выборок. Этот пример показывает, какую роль играет выборка значений переменных.

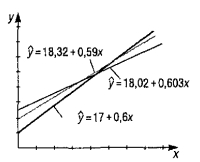


Рисунок 4

Для включения в уравнение регрессии коэффициента корреляции перепишем уравнение в следующем виде:,

где

Таким образом, уравнение регрессии имеет вид:

, где – выборочный коэффициент корреляции.

Проверим значимость коэффициента корреляции в приведенных примерах.

Для этих примеров коэффициент корреляции r (рассчитаем по формуле выборочного коэффициента корреляции) имеет следующее значение:

0,951 – для генеральной совокупности (табл.1);

0,956 – для первой выборки (табл. 3);

0,941 – для второй выборки (табл. 5).

Расчетные значения t- статистики равны, соответственно:

tp=23,54 – для генеральной совокупности,

tp=17,16 – для первой выборки,

tp=14,68 – для второй выборки.

На уровне значимости критические значения t-статистики соответственно равны: 2,002 – для генеральной совокупности, 2,048 – для первой и второй выборки.

Таким образом, для генеральной совокупности и для обеих выборок коэффициент корреляции значим.

Необходимо отметить, что хотя коэффициент корреляции Y на Х равен коэффициенту корреляции Х на Y (rxy=ryx), соответствующий коэффициент регрессии b1 Y на X не совпадает с коэффициентом регрессии X на Y.

4 **Коэффициент детерминации и средняя ошибка аппроксимации**

Пусть ŷ = b0+b1xвыборочное уравнение линейной регрессии, построенной по выборке .

Ошибки регрессии находятся по формуле:

Эти ошибки используются для проверки качества уравнения регрессии.

Разброс наблюдаемых значений зависимой переменной у характеризуется дисперсией

Из равенства + следует, что ,где , , – дисперсии переменных

Таким образом дисперсия  состоит из двух слагаемых:  - слагаемое, объясненное регрессией,  - необъясненная часть вариации y.

Коэффициент детерминации называется отношение

, которое характеризует долю вариации зависимой переменной у, объясненную уравнением регрессии.

Необъясненная часть вариации равна

Если , то , и все точки выборки лежат на прямой регрессии.

Если , то , и все точки выборки лежат на горизонтальной прямой у= в этом случае изменение переменной х не вызывает изменения переменной у, зависимость между ними отсутствует.

Чем ближе к единице, тем лучше регрессия объясняет у. В этом смысл коэффициента детерминации.Он является одной из эффективных оценок адекватности регрессионной модели.

Оценку качества модели регрессии можно также проводить с помощью средней ошибки аппроксимации.

Средней ошибкой аппроксимации (а) называется среднее про­центное относительное отклонение фактических значений зависи­мой переменной от ее расчетных значений:

А= .

Модель регрессии считается качественной, если ошибка аппроксимации не превышает 8-10%.

*Пример 3.*

Вычислим средние ошибки аппроксимации для генеральной совокупности и двух выборок из нее, приведенных в примерах 1-2.

В результате расчетов получена средняя ошибка аппроксима­ции, она равна соответственно: 8% - для генеральной совокупно­сти; 7,3% - для первой выборки; 9,4% - для второй выборки. Та­ким образом, качество всех трех моделей - удовлетворительное.

**5 Предпосылки регрессионного анализа. Условия Гаусса-Маркова**

Рассмотрим парную линейную регрессионную модельY=.

Для выборки наблюдений {() } модель регрессии записывается в виде , i= 1, ... , n.

Оценкой этой модели по МНК служит уравнение =, i= 1, ... , n, где находятся по формулам:

,

Как видно из последних формул, относительно переменных эти величины - линейные, что является важной особенностью оценок. Мы будем рассматривать и как (точечные) оценка параметров и модели регрессии.

Основными «хорошими» свойствами статистических оценок являются несмещенность, эффективность и состоятельность.

Для выполнения этих свойств оценок вводятся дополнительные условия на случай­ные остатки и переменные , в модели регрессии.

Приведем условия Гаусса-Маркова.

1. Объясняющая переменная , является детерминированной (неслучайной) величиной с ненулевой дисперсией, остаток явля­ется случайной величиной (вместе с переменной )(i = 1,... , п).

2. Математическое ожидание случайного остатка равно нулю, т.е. М = 0 (i = 1,..., n).

3. Дисперсия случайного остатка постоянна для всех наблюдений, т.е. D ( ) = M () = , (i = 1,..., n).

4. Случайные остатки некоррелированны между собой, т.е. ( ) = 0, при i ≠ j .

5. Случайный остаток , является нормально распределенной случайной величиной (с нулевым средним и постоянной дисперсией ). Вместе сзависимая переменная также является нормальной случайной величиной.

Классической нормальной линейной регрессионной моделью называется модель репрессии, для которой выполнены условия 1-5. Для такой модели некоррелированность случайных остатков (зависимых переменных) равносильна их независимости.

Гомоскедактичностью остатков называется условие 3 постоянства дисперсии случайных остатков (зависимых переменных).

**6 Свойства оценок коэффициентов классической линейной модели регрессии. Теорема Гаусса-Маркова**

Перейдем к описанию качества оценок модели регрессии.

Теорема Гаусса-Маркова. При выполнении условий 1- 4оценки , параметров , полученные по методу наимень­ших квадратов, обладают следующими свойствами:

1) несмещенность: М = β0 , М = β1 (отсутствие системати­ческого смещения оценки от самого параметра);

2) эффективность: оценки имеют наименьшую дисперсию в классе всехлинейных (относительно Y) несмещенных оценок;

3) состоятельность: при n → ∞ сходятся по вероят­ности к самим параметрам, т.е.

или подробнее

│≤ ɛ ) = 1, │≤ ε ) = 1.

Условие 5 о нормальности распределения случайной составляющей пока не использовано. Выполнение этого условия влечет нормальность распределений оценок и. Это свойство будет использовано в дальнейшем.

Доказательство теоремы Гаусса-Маркова будет приведено в общем случае для множественной регрессии.

Приведем доказательства состоятельности МНК-оценки коэффициента регрессии.

Теорема. Для модели парной репрессии = + и условия М() = 0 следует, что МНК-оценка параметра является состоятельной, т.е. предел по вероятности этой оценки равен самому параметру.

Доказательство. Установим связь между оценкой и параметром .

Используем полученную формулу для нахождения предела по вероятности:

В соответствии с законом больших чисел имеем:

Следовательно,

поскольку ковариация детерминированной переменной х со случайным остатком ε равна нулю. Теорема доказана.

Одной из важных характеристик оценок служит их дисперсия. Для оценок и дисперсии определяются по формулам

, где - общая дисперсия остатков; - выборочная дисперсия х.

Выборочная остаточная дисперсия служит несмещенной оценкой дисперсии остатков 2.

Стандартной ошибкой регрессии называется величина

Заменив в формулах дисперсии дисперсию остатков 2ее выборочной оценкой , получим выборочные оценки для дисперсий .

Стандартными ошибками оценок коэффициентов регрессии называются величины

полученные из предыдущей формулы.

При выполнении условий 5 о нормальности распределения случайных остатков: , оценки и коэффициентов регрессии β0 иβ1, соответственно, также имеют нормальные распределения

где дисперсии определяются по формулам .

**7 Значимость коэффициента регрессии и его доверительный интервал**

Качество регрессионной модели можно проверить не только на основе коэффициентов детерминации и ошибки аппроксима­цииА, но и на основе выборочного коэффициента регрессии .

Пусть для модели по выборке значений X и Yполучен коэффициент регрессии . Если выполняется условие 5 в нормальности распределения случайного остатка ε, то статистика

имеет t-распределение Стьюдента с степенями свободы. (При подсчете степеней свободы от объема выборки n вычитается количество связей, равное количеству уравнений в системе нор­мальных уравнений, т.е. 2.) По таблице критических точек Стью­дента по заданному уровню значимости (или доверительной вероятности ) и числу степеней свободы находится критическое значение . Критическое значение при заданном и определяется также в Ехсе1 с помощью функции = СТЬЮДРАСПОБР (, ). Для построения доверительного интервала коэффициента используется двусторонний критерий, основанный на статистике (27). Соответствующая доверительная область задается равенством

Доверительным интервалом на уровне значимости для коэффициента β1 служит интервал

который записывается также в виде

Для проверки значимости коэффициента регрессии тестируется гипотеза H0: = 0. Значимость коэффициента регрессии означает, что доверительный интервал не содержит числа нуль. Это равносильно условию, при котором гипотеза H0 отвергается. Таким образом, если , то коэффициент регрессии незначим ( = 0) на уровне . В доверительном интервале находятся статистически значимые значения коэффициента .

При вычислении оценок коэффициентов регрессии в Excel с помощью функции «Регрессия» находятся значения коэффициентов и, их стандартные ошибки и , t-статистики и . Кроме этого в столбце Р-значение указываются минимальные уровни значимости для коэффициентови. Например, еслиР-значение для равно 0,01, то коэффициент регрессии значимна всех уровнях, больших 0,01. В столбцах Нижние 95%, Верхние 95% указываются концы доверительных интервалов для на уровне значимости = 0,05 (0,05 = 1 – 0,95).

**8 Интервальная оценка функции регрессии. Прогноз на основе функции регрессии**

Прогнозирование на основе функции регрессии основано на формуле

Поскольку выборочная функция регрессии строится по выбор­ке и является оценкой функции регрессии, для нее можно постро­ить доверительный интервал. Для этого требуется определить дисперсию величины , которая и есть выборочная оценка для .С этой целью представим уравнение регрессии в виде:

Из данного уравнения дисперсия представляется в виде:

Дисперсия равна , а дисперсия по формуле равна от .

Поэтому

Заменив с ее выборочной оценкой , получим оценку дисперсии : .

Доверительный интервал для функции регрессии строится на основе статистики , которая имеет t-распределение Стьюдента с к = п- 2 степенями свободы. Таким образом, доверительный интервал на уровне для функции регрессии имеет вид , где - критическое значение t-распределения Стьюдента, а находится из формулы .

Индивидуальные значения зависимой переменной имеют относительно линии регрессии стандартную ошибку s.Поэтому доверительный интервал для индивидуальных значений вы­глядит следующий образом:

где - стандартная ошибка прогноза для индивидуального значения - прогнозное значение удля х= .

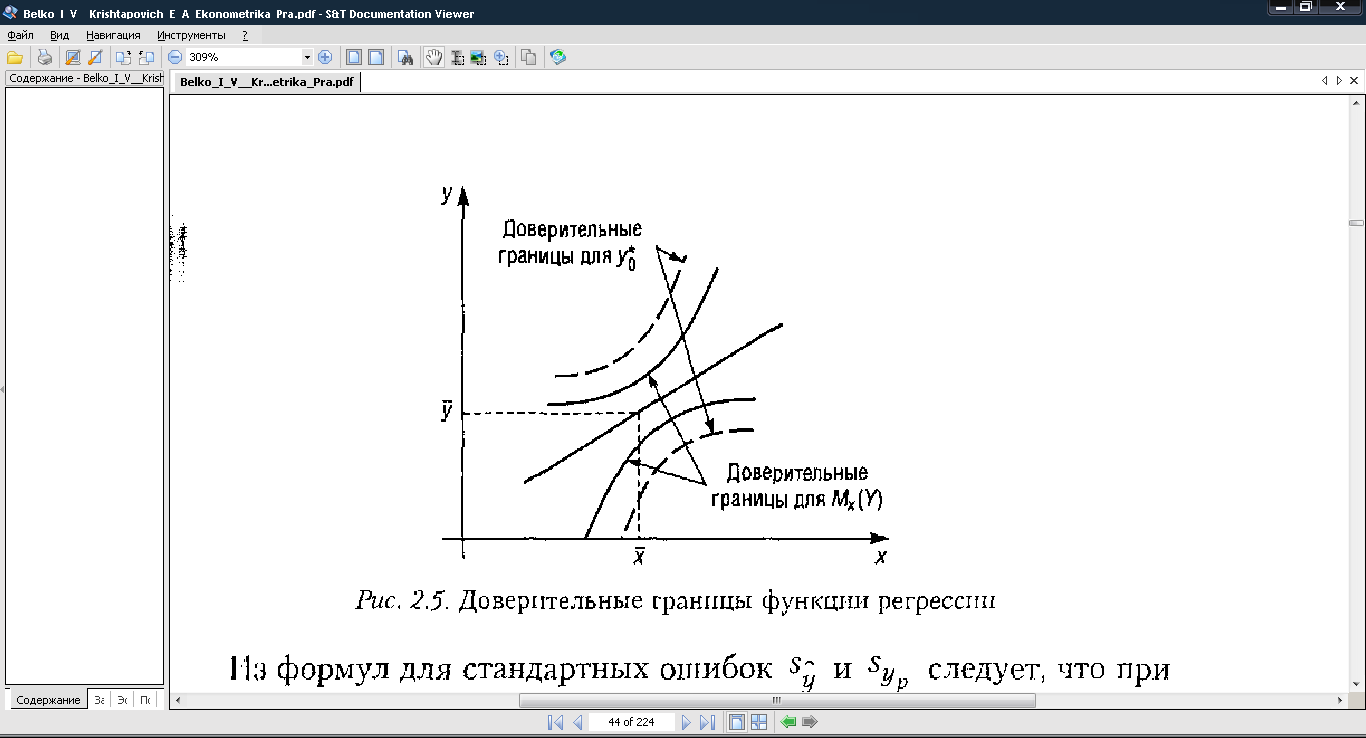


Рисунок 5

Из формул для стандартных ошибок и следует, что при они минимальны, а при увеличении отклонения от. Возрастают (рис. 5).

**9 Решение типовых задач с помощью МSЕхсеl**

В этом пункте будут рассмотрены способы построения уравнения парной линейной регрессии, вычисления прогноза и оценки его точности на основе следующего примера.

*Пример 4.*

Имеются данные по 50 предприятиям о стоимости ОПФ (X, млн р.) и среднесуточной производительности (Y, т).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| X | 35,03 | 36,79 | 37,40 | 36,68 | 37,62 | 38,48 | 39,09 | 39,23 | 39,26 | 39,31 | 39,72 | 40,19 |
| Y | 12,98 | 15,35 | 15,81 | 17,13 | 18,64 | 18,60 | 24,29 | 23,41 | 22,21 | 19,80 | 21,84 | 17,53 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| X | 40,10 | 40,19 | 40,41 | 40.76 | 41,54 | 41,68 | 45,12 | 48,42 | 48,29 | 48,65 | 51,93 | 53,23 |
| Y | 16,29 | 22,13 | 19,81 | 21,30 | 20,18 | 24,17 | 24,97 | 26,44 | 24,19 | 22,05 | 22,16 | 27,72 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
| X | 54,35 | 54,69 | 56,61 | 56,76 | 56,80 | 57,30 | 59,26 | 59,84 | 61,45 | 61,72 | 62,91 | 65,21 |
| Y | 25,86 | 21,15 | 25,37 | 24.83 | 30,80 | 28,00 | 30,55 | 31,66 | 29,77 | 31,12 | 35,13 | 33,73 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | ' 45 | 46 | 47 | 48 |
| X | 67,92 | 67,63 | 68,45 | 68,24 | 68,46 | 68,58 | 68,53 | 69,17 | 69,45 | 70,98 | 71,40 | 74,23 |
| Y | 29,08 | 32,66 | 34,12 | 33,88 | 26,33 | 34,54 | 36,34 | 29,19 | 28,50 | 35,60 | 32,19 | 30,76 |

|  |  |  |
| --- | --- | --- |
| N | 49 | 50 |
| X | 74, 97 | 75,78 |
| Y | 31,87 | 38,00 |

Парный регрессионный анализ начинается с построения уравнения регрессии 

**9.1 Построение линейной регрессии с помощью Мастер функций f(x), Статистические, ЛИНЕЙН (первый способ)**

Порядок действий следующий:

1. Введите исходные данные или откройте существующий файл, содержащий анализируемые данные (данные примера).

2. Выделите область пустых ячеек 5x2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1 х2 -для получения только оценок коэффициентов регрессии.

3. Активизируйте Мастер функций любым из способов;

а) в главном меню выберите Вставка/Функция;

б) на панели инструментов Стандартная щелкните но кнопке  
Вставка функции.

4. В окне «Категория» (рис. 6) выберите Статистические, в  
окне «Функция» - ЛИНЕЙН. Щелкните по кнопке ОК.

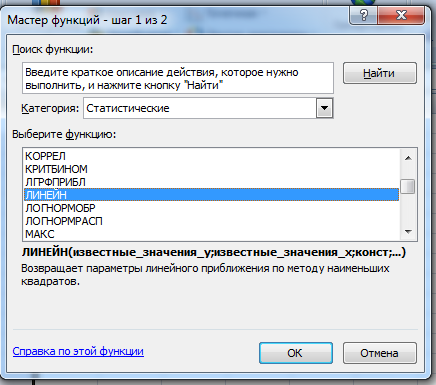


Рисунок 6. Диалоговое окно «Мастер функции»

5. Заполните аргументы функции (рис. 7):

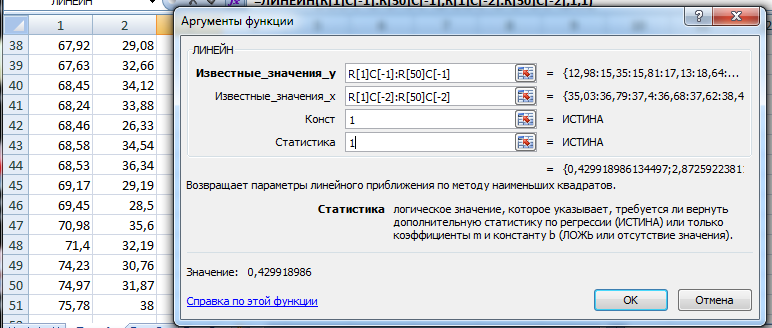


Рисунок 7. Диалоговое окно ввода аргументов функции ЛИНЕЙН

Известные значения y — диапазон, содержащий данные результативного признака (среднесуточной производительности).

Известные значения x — диапазон, содержащий данные факто­ров независимого признака (стоимости ОПФ).

Константа — логическое значение, которое указывает на нали­чие или отсутствие свободного члена в уравнении. Если Константа = 1, то свободный член рассчитывается обычным образом, еслиКонстанта = 0,то свободный член равен 0.

Статистика — логическое значение, которое указывает, выво­дить дополнительную информацию по регрессионному анализу или нет. Если Статистика = 1,то дополнительная информация выводится, если Статистика = 0,то выводятся только оценки па­ра метров уравнения.

Щелкните по кнопке ОК.

6. В левой верхней ячейке выделенной области появится пер­вый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, на­жмите на клавишу F2,а затем на комбинацию клавиш CTRL+SHIFT+ЕNTER. Дополнительная регрессионная статисти­ка будет выводиться в порядке, указанном в следующей схеме:

|  |  |
| --- | --- |
| Значение коэффициента | Значение коэффициента |
| Среднеквадратическое отклонение коэффициента | Среднеквадратическое отклонение коэффициента |
| Коэффициент детерминации | Стандартная ошибка регрессии s |
| F-статистика | Число степеней свободыn-2 |
| Объясненная регрессией сумма квадратов отклонений RSS | Остаточная сумма квадратов ESS |

Для данных из рассматриваемого примера результат вычисле­ния функции ЛИНЕЙН представлен на рис. 8.

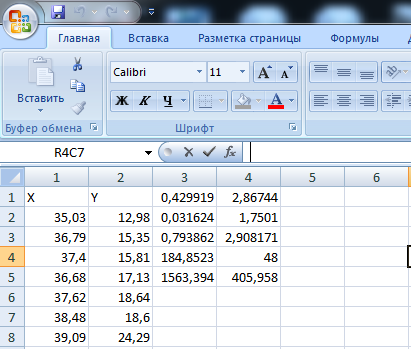


Рисунок 8. Результат вычисления функции ЛИНЕЙН

Функция ЛИНЕЙН рассчитывает статистику для ряда с при­менением метода наименьших квадратов, чтобы вычислить пря­мую линию, которая наилучшим образом аппроксимирует имею­щиеся данные. Функция возвращает массив, описывающий полу­ченную прямую, уравнение которой можно записать в виде= 2,86744 + 0,429963x.

Свободный член экономического значения не имеет, в отличие от коэффициента регрессии, который показывает, что при увели­чении стоимости ОПФ на 1млн. р. среднесуточная производи­тельность увеличится на 0,429963 т.

**9.2 Построение линейной регрессии с помощью инструментов Сервис, Анализ данных, Регрессия (второй способ)**

С помощью инструмента параметров Регрессия можно получить результаты регрессионной статистики, дисперсионного ана­лиза и доверительных интервалов, а также остатки и графики подбора линии регрессии.

Порядок действий следующий:

1. Активизируйте доступ к Пакету анализа (если он не был активизирован

до этого). В главном меню последовательно выбе­рите Сервис, Надстройки. Установите флажок Пакет анализа (рис. 9).



Рис. 9. Подключение надстройки Пакет анализа

1. В главном меню выберите Сервис, Анализ данных, Регрес­сия. Щелкните по кнопке ОК.
2. Заполните диалоговое окно ввода данных и параметров вывода (рис. 2.10):

Входной интервалУ,X —аналогичны известным значениям Y и X, рассмотренным в пункте 9.1.

Метки— флажок, который указывает, содержит ли первая стро­ка названия столбцов или нет.

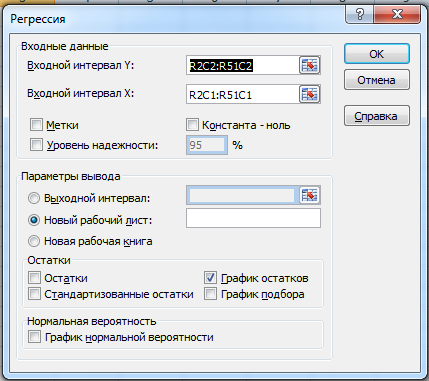


Рис. 10.Диалоговое окно ввода параметров инструмента Регрессия

Константа-ноль — флажок, указывающий на отсутствие сво­бодного члена в уравнении.

Выходной интервал — достаточно указать левую верхнюю ячейку будущего диапазона.

Новый рабочий лист — можно задать произвольное имя ново­го листа.

Если необходимо получить информацию об остатках и их графиках, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке ОК.

Результаты регрессионного анализа для данных из примера представлены на рис. 11.

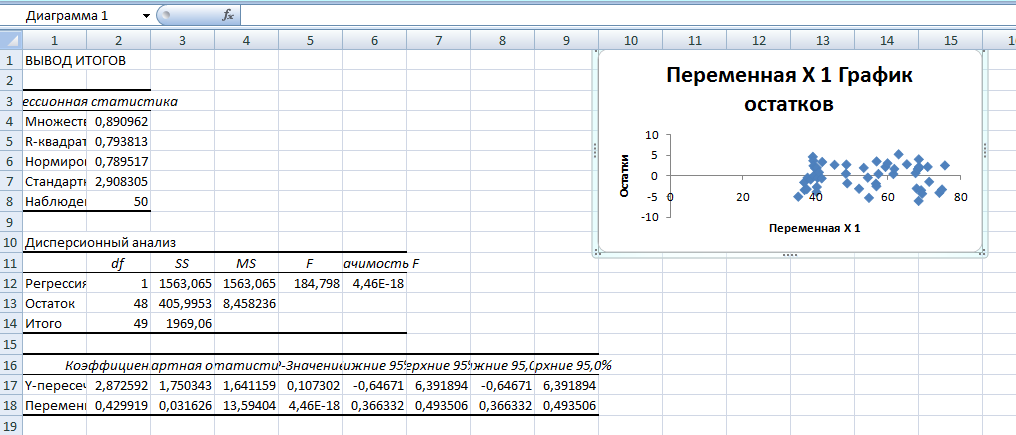


Рис. 11.Результаты регрессионного анализа

**9.3 Интерпретация результатов регрессионного анализа**

1. Регрессионная статистика

|  |  |
| --- | --- |
| Множественный R | 0,890989405 |
| R-квадрат | 0,793862119 |
| Нормированный R-квадрат | 0,78956758 |
| Стандартная ошибка | 2,90817145 |
| Наблюдения | 50 |

Множественный R— коэффициент корреляции, который явля­ется показателем тесноты линейной связи между переменными. В нашем случае r=0,89. Значение близко к единице, что свидетельствует о достаточно тесной связи между у и х.

R-квадрат — коэффициент детерминации — показатель качества подбора линейной функции.

Нормированный R-квадрат— скорректированный коэффициент детерминации (в отличие от предыдущего его значение не зависит от числа факторов, включенных в модель). Этот показатель также достаточно близок к единице, что говорит о правильности использования линейной функции в данном случае.

Стандартная ошибка— оценка для среднего квадратического отклонения случайных остатков , рассчитывается по формуле

Наблюдения— количество наблюдений в данном конкретном случае.

1. Дисперсионный анализ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | SS | MS | F | Значимость F |
| Регрессия | 1 | 1563,394295 | 1563,394295 | 184,8538541 | 4,43348Е-18 |
| Остаток | 48 | 405,9581366 | 8,45746118 |  |  |
| Итого | 49 | 1969,352432 |  |  |  |

В этой части результатов в столбце dfвыписаны значения степеней свободы для суммы квадратов остатков RSS, ЕSS и TSS. В столбце SS выписаны суммы квадратов остатков RSS, ЕSSи TSS. В столбце МS - частные отделения суммs квадратов остатков на число их степеней свободы. В двух послед них столбцах выписаны F-статистика Фишера для проверки значимости коэффициента детерминации и ее Р-значение. Для проверки значимости коэффициента детерминации (и уравнения в целом) значение F-статистики сравнивается с критическим.

Обычно при проверке значимости уравнения с помощью F-статистики сначала выбирают уровень значимости (%), затем определяют число степеней свободы = 1 (количество факторов),= п-2 (разность между числом наблюдений, уменьшенным на единицу, и числом факторов). С учетом этих заданных параметров находят критическое значение Fкр. Если F>Fкр, то уравнение в целом является значимым при выбранном уровне . В противном случае - нет.

Для вычисления критического значения F-статистики используют функцию FРАСПОБР с параметрами = 1,= п-2 (рис. 12).

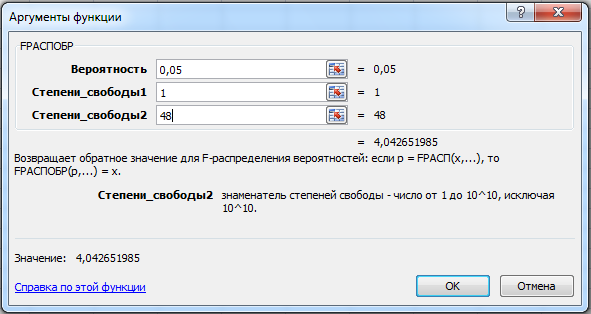


Рис. 12. Диалоговое окно функции FРАСПОБР

По данным нашего примера Fкр = 4,0427, следовательно, уравнение значимо. Гипотеза о значимости подтверждается также значением F, равным 4.43348Е-18.

1. Коэффициенты уравнения, параметры оценки, доверитель­ные

интервалы

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Коэффициенты | Стандартная ошибка | t-статистика | P-значение | Нижние 95% | Верхние 95% |
| y-пересечение | 2,867439647 | 1,7501009 | 1,638443232 | 0,1078687 | -0,6513723 | 6,38625 |
| Переменная | 0,429963383 | 0,03162403 | 13,59609702 | 4,4335E-18 | 0,36637901 | 0,49354776 |

Коэффициенты— коэффициенты уравнения регрессии y==2,86744 + 0,429963х.

В столбце Стандартная ошибка выписаны стандартные ошибки коэффициентов.

В столбце t-статистика приведены частные от деления коэффициентов на их стандартные ошибки.

Столбец Р-значение указывает минимальные уровни зна­чимости коэффициентов.

Столбцы Нижние 95% и Верхние 95% показывают границы доверительных интервалов для коэффициен­тов уравнения регрессии с доверительной вероятностью 0,95.

Для проверки значимости коэффициентов уравнения регрессии вычисляем критическое значение tкр. Для этого можно вос­пользоваться статистической функцией СТЬЮДРАСПОБР (рис. 13), в диалоговом окне которой указывается уровень доверия (например, 0,05) и число степеней свободы, равное n-2.В нашем случае tкр = 2,0106 . Если |t| >tкр, то коэффициент является зна­чимым при выбранном уровне . В противном случае выбранный фнктор х не влияет на результативный признак у.

Для коэффици­ента t-статистика меньше критического значения, поэтому незначим, его доверительный интервал (-0,6513723; 6,38625165) содержит значение нуль.

Поэтому с вероятностью 0,95 коэффици­ент может быть равен нулю.

Коэффициент является значи­мым, поскольку его t-статистика больше tкр, Р-значение меньше0,05, а доверительный интервал (0,36637901; 0,49354776) не содер­жит нуля.

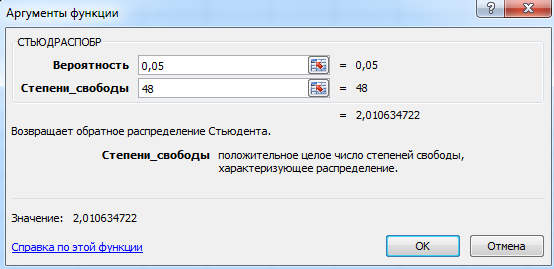


Рис. 13. Диалоговое окно ввода параметров функции СТЬЮДРАСПОБР

1. Некоторые взаимосвязи

Связь между значимостью коэффициента регрессии и уравне­ния в целом: в парной регрессии F-статистика является квадратом t-статистики, то же верно и для критических уровней этих показа­телей (т.е. значимость коэффициента регрессии и значимость уравнения в целом эквивалентны) .

Взаимосвязь между коэффициентом корреляции и коэффици­ентом детерминации: в парном регрессионном анализе коэффици­ент корреляции по абсолютной величине совпадает с квадратным корнем коэффициента детерминации , t-статистики для коэффициентов корреляции и детерминации совпадают (про­верка значимости коэффициента регрессии эквивалентна провер­ке наличия линейной связи).

**9.4 Вычисление прогнозного значения у**

Пусть стоит задача рассчитать прогнозную производительность труда на вновь создаваемом предприятии, где сумма, выделяемая наприобретение ОПФ, планируется на уровне 90% от среднего уровня стоимости ОПФ по данной отрасли предприятий. Решить поставленную задачу можно двумя способами.

Первый способ:

1. Откройте файл, содержащий анализируемые данные (данные примера).
2. В свободную ячейку введите формулу для вычисления стои­мости вновь вводимых ОПФ, используя статистическую функцию СРЗНАЧ.
3. Выделите свободную ячейку, где впоследствии будет располагаться прогнозное значение у, в главном меню выберите Вставка/Функция/Статистические/ТЕНДЕНЦИЯ и заполните пара­метры функции, как показано на рис. 14 (новые значения x – новое значение признака, для которого мы делаем прогноз, длянашего примера =48,41) .

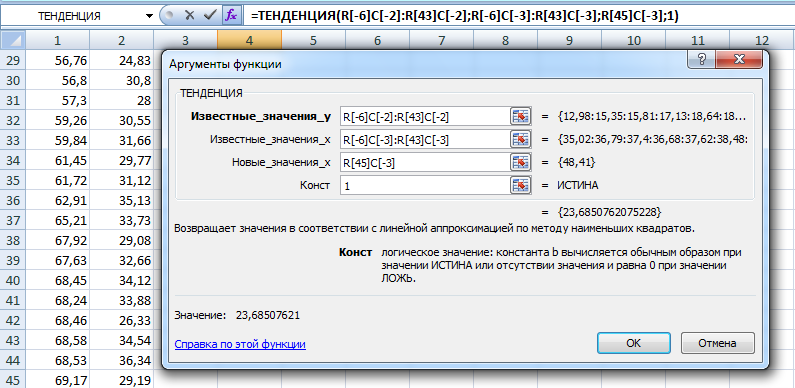


Рисунок 14. Диалоговое окно функции ТЕНДЕНЦИЯ

Второй способ:

Введите в свободную ячейку формулу уравнения регрессии =2,86744 + 0,429963x, полученного в пункте 9.3. Подставив вместо x значение 48,41, получим = 23,68278.

**9.5 Оценка точности прогноза**

Оценить полученный прогноз можно, рассчитав его среднюю ошибку и построив доверительный интервал для заданного уров­ня надежности (0,05).

Формула для расчета средней ошибки индивидуального значе­ния уимеет вид:

где s- стандартная ошибка; xпр - новое значение признака, для которого был сделан прогноз; - среднее значение gо столбцу значений независимого признака Х; - столбец значений незави­симого признака X; п- количество наблюдений.

Используя стандартные математические операции, а также функции СРЗНАЧ и КОРЕНЬ, производим расчет ошибки по указанной формуле.

Следующим шагом будет расчет доверительного интервала для вычисленного прогноза =23,68278, который строится с помощью найденной ранее t-статистики Стьюдента и имеет вид:

.

По этой формуле легко найти, что с вероятностью 95% при сто­имости ОПФ, равной 48,41 млн. р. (x), среднесуточная производительность (у) вновь создаваемого предприятия будет находиться в пределах от 17,76743394 до29,59813 т.

**9.6 Решение задач регрессии без помощи специальных средств**

*Пример 5.*

По условным данным, приведенным в следующей таблице, необходимо:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| x | 2 | 2,1 | 2,3 | 2,4 | 2,9 | 3.3 | 3,8 | 4,6 |
| y | 14,3 | 18,6 | 18,7 | 20,9 | 22,3 | 24,2 | 25,7 | 27 |

1) построить уравнение регрессии, описывающее зависимость между среднесуточной производительностью труда Y (т) и стои­мостью основных производственных фондов Х (млн р.). Пояснить экономический смысл коэффициента регрессии;

2) оценить степень тесноты связи между переменными с по­мощью коэффициента корреляции;

3) проверить значимость коэффициентов уравнения регрессии и построить их интервальные оценки на уровне значимости 0,05;

4) найти стандартную ошибку и доверительный интервал для уравнения регрессии в целом, а также для индивидуального про-гнозного значения упр для хпр = 3.

Решение.

1. Вычислим все необходимые суммы:

Затем найдем выборочные характеристики и параметры уравнения регрессии:

= = = 2,925 (млн. р)

= = = 21,4625 (т)

= – x2 = – x2 = - 2,9252 = 0,7394

cov (X,Y) = xy- ͞x ·͞y = - x͞ ·͞y = - 2,925· 21,4625=3,1247

^

b = = = = 4,226

b0=͞y – b1͞x = 21,4625 - 4,226 · 2,925 = 9,101

где s2x - выборочная дисперсия переменной X; cov (X,Y) - вы­борочная ковариация; b1 - коэффициент регрессии упо X.

Коэффициент регрессии Y по X показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной xнаодну единицу.

Уравнение регрессии Y по X: ŷ = 9,101 + 4,226 х.

Из полученного уравнения регрессии следует, что при увеличе­ниистоимости ОПФ X на 1 млн р. среднесуточная производительность труда Y увеличивается в среднем на 4,226 т. Свободный членв данном уравнении регрессии не имеет реального смысла.

2. Коэффициент корреляции r является показателем тесноты связи и рассчитывается по формулам (2.3):

r= b1 = =

·

Для практических расчетов наиболее удобна последняя формула, так как по ней r находится непосредственно из данных наблюдений и на значении r не скажутся округления данных, связанные с расчетом средних и отклонений от них.

Коэффициент корреляции принимает значения на отрезке [-1; 1], т.е., -1 <r< 1.

Чем ближе |r| к единице, тем теснее линейная связь.

Выше были вычислены:

Вычислим сумму

14, 32 + 192 +18,72 + 20,92 + 22,32 + 24,22 + 25, 72 + 272 = 3809,37

r = = 0,922, т.е. связь между переменными достаточно тесная.

3. Вычислим стандартную ошибку регрессии s. Для этого будем использовать формулу

S2=, в которой ошибки ei находятся с использованием уравнения регрессии. Составим вспомогательную таблицу.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| xi | 2 | 2,1 | 2,3 | 2,4 | 2,9 | 3,3 | 3,8 | 4,6 | y |
| (xi-͞x2) | 0,86 | 0,68 | 0,39 | 0,28 | 0,001 | 0,14 | 0.77 | 2,81 | 5,915 |
| ŷ=9,101099+ 4,1703·xj | 17,55 | 17,98 | 18,82 | 19,24 | 21,36 | 23,05 | 25,16 | 28,54 | - |
| = | 3,25 | -0,62 | 0,12 | -1,65 | -0,94 | -1,15 | -0,54 | 1,54 |  |
|  | 10,58 | 0,39 | 0,01 | 2,74 | 0,89 | 1,33 | 0,29 | 2,37 | 18,616 |

Теперь имеем

S2===3,103; s=1,76

Найдем расчетные значения t-статистик для коэффициентов уравнения регрессии, учитывая, что b1= 4,226, , s2=3,103

s=1,76, = =

== ==5,83

Пo таблицам t-распределения t0,95:6= 2,45. Так как t>t0,95:6 коэффициент регрессии b1 значим, а следовательно, и уравнение парной линейной регрессии Y пo Xзначимо.

Аналогично, b0=9,101, sb₀= s= 1,76=1,76·1,25=2,21, == = 4,12

Сравнивая это значение с tкр, делаем вывод о значимости коэф­фициента b₀. Интервальные оценки для коэффициентов уравне­ния регрессии строятся по формулам:

Найдем 95%-й доверительный интервал для параметра рег­рессионной модели. Для этого подставим полученные выше значе­ния величин, входящих в эти формулы

4,226 - 2,45·0,72 ≤  1 ≤4,226 + 2,45 ·0,72

2,45 ≤ ≤5,99;

9,101-2,45·2,21 ≤ 0 ≤ 9,101+ 2,45·2,21

3,69 ≤ ≤ 14,5.

Значит, с надежностью 0,95 при изменении стоимости ОПФ X на 1 млн р. среднесуточная производительность Y будет изменяться на величину, заключенную в интервале от 2,45 до 5,99 (т).

4. Стандартная ошибка для ŷх = 9,101 + 4,226x находится по формуле

= s

С учетом данных нашего примера стандартная ошибка имеет вид

=1,76

Доверительный интервал для уравнения регрессии имеет вид

Этот интервал зависит от значений х.

Построим прогноз для хпр = 3, используя полученное ранее уравнение регрессии ŷпр= 9,101 + 4,226xпр = 21,78.

Стандартная ошибка для индивидуального значения ух имеет вид

sŷₓ = s

В нашем случае она равна =1,76=1,87

Доверительный интервал для индивидуального прогнозного значения равен

или

(21,78-2,45·1,87; 21,78 + 2,45·1,87) = (17,2; 26,4).

Таким образом, среднесуточная производительность труда на предприятии со стоимостью ОПФ, равной 3 млн. р., с надежностью 0,95 находится в пределах от 17,2 до 26,4 т.