Презентация на тему: Парная регрессия и корреляция

Выполнила:
студент(ка) группы: 1к-пот.1 - МГЭ
Аликанова Татьяна

План:

1. Введение
2. Корреляционный анализ
3. Парная регрессия
4. Метод наименьших квадратов
5. Оценка качества уравнения регрессии

Основные понятия:

- Регрессионный анализ
- Корреляционный анализ
- Ковариация
- Стандартное отклонение
- Оценка значимости коэффициента корреляции
- МНК
- Коэффициент детерминации
- Проверка значимости моделей

Введение

Существуют три основных класса моделей, которые применяются для анализа и прогнозирования экономических систем:

- модели временных рядов,
- регрессионные модели с одним уравнением,
- системы одновременных уравнений.

Регрессионные модели с одним уравнением

В регрессионных моделях зависимая (объясняемая) переменная Y может быть представлена в виде функции f (X_1, X_2, X_3, … X_k), где - независимые (объясняющие) переменные, или факторы; k – количество факторов. В качестве зависимой переменной может выступать практически любой показатель, характеризующий, например, деятельность предприятия или курс ценной бумаги. В зависимости от вида функции

$$f(X_1, X_2, \dots, X_k)$$

модели делятся на линейные и нелинейные. В зависимости от количества включенных в модель факторов X модели делятся на однофакторные (парная модель регрессии) и многофакторные (модель множественной регрессии).

Регрессионный анализ

занимает ведущее место в математике статистических методах эконометрики.

До регрессионного анализа следует проводить **корреляционный анализ**, в процессе которого оценивается степень тесноты статистической связи между исследуемыми переменными. От степени тесноты связи зависит прогностическая сила регрессионной модели.

Регрессионный анализ предназначен для **исследования зависимости исследуемой переменной от различных факторов** и отображения их взаимосвязи в форме регрессионной модели.

Основная задача корреляционного анализа

заключается в выявлении взаимосвязи между случайными переменными путем точечной и интервальной оценки парных (частных) коэффициентов корреляции, вычисления и проверки значимости множественных коэффициентов корреляции и детерминации.

При проведении корреляционного анализа вся совокупность данных рассматривается как множество переменных (факторов), каждая из которых содержит *n* –наблюдений. При изучении взаимосвязи между двумя факторами их, как правило, обозначают

$$X = (x_1, x_2, \dots, x_n) \text{ и } Y = (y_1, y_2, \dots, y_n).$$

Ковариация - это статистическая мера взаимодействия двух переменных.

$$\boxed{\mathrm{Cov}(X,Y)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),}$$

где $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ — фактические значения переменных X и Y;

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, \quad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i.$$

Для двух переменных образом:
коэффициент парной корреляции определяется следующим

$$r_{y,x} = \frac{\mathrm{Cov}(X, Y)}{S_x S_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{S_x S_y} =$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

где S_x^2, S_y^2 — оценки дисперсий величин X и Y. Эти оценки характеризуют *степень разброса* значений $x_1, x_2, ..., x_n$ ($y_1, y_2, ..., y_n$) вокруг своего среднего $\bar{x}$ ($\bar{y}$ соответственно), или *вариабельность* (изменчивость) этих переменных на множестве наблюдений.

Дисперсия (оценка дисперсии)

характеризует вариабельность (изменчивость) этих переменных на множестве наблюдений.

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

В общем случае для получения несмещенной оценки дисперсии сумму квадратов следует делить на число степеней свободы оценки (n–p), где n - объем выборки, p - число наложенных на выборку связей. В данном случае p = 1, т.к. выборка уже использовалась один раз для определения среднего X, поэтому число наложенных связей равно единице, а число степеней свободы оценки (т.е. число независимых элементов выборки) равно (n –1).

Более естественно измерять степень разброса значений переменных в тех же единицах, в которых измеряется и сама переменная. Эту задачу решает показатель, называемый **среднеквадратическим отклонением или стандартным отклонением, или стандартной ошибкой переменной** X (переменной Y), определяемый соотношением:

$$\boxed{S_x = \sqrt{S_x^2}.}$$

Оценка значимости коэффициента корреляции

при малых объемах выборки выполняется с использованием ***t* - критерия Стьюдента**.

При этом фактическое (наблюдаемое) значение этого критерия определяется по формуле:

$$\boxed{t_{\text{набл}} = \sqrt{\frac{r_{y,x}^2}{1 - r_{y,x}^2}(n-2)}.}$$

Диаграмму рассеяния, на которой изображается совокупность значений двух признаков, называют еще ***корреляционным полем***. Каждая точка этой диаграммы имеет координаты X_i и Y_i. По мере того, как возрастает сила линейной связи, точки на графике будут лежать более близко к прямой линии, а величина r будет ближе к 1.

Пример

Пример 3.2.1. *Вычисление коэффициентов парной, множественной и частной корреляции.*

В табл. 3.2.2 представлена информация об объемах продаж и затратах на рекламу одной фирмы, а также индекс потребительских расходов за ряд текущих лет.

1. Построить диаграмму рассеяния (корреляционное поле) для переменных «объем продаж» и «индекс потребительских расходов».
2. Определить степень влияния индекса потребительских расходов на объем продаж (вычислить коэффициент парной корреляции).
3. Оценить значимость вычисленного коэффициента парной корреляции.
4. Построить матрицу коэффициентов парной корреляции по трем переменным.
5. Найти оценку множественного коэффициента корреляции.
6. Найти оценки коэффициентов частной корреляции.

Рассмотрим теперь **решение примера 3.2.1 в Excel**.

Чтобы вычислить корреляцию средствами Excel, можно воспользоваться функцией **=КОРРЕЛ( )**, указав адреса двух столбцов чисел, как показано на рис. 3.2.2. Ответ помещен в D8 и равен 0,816.

(Примечание. Аргументы функции **КОРРЕЛ** должны быть числами или именами, массивами или ссылками, содержащими числа. Если аргумент, который является массивом или ссылкой, содержит текст, логические значения или пустые ячейки, то такие значения игнорируются; однако ячейки, которые содержат нулевые значения, учитываются.

Если **массив1** и **массив2** имеют различное количество точек данных, то функция **КОРРЕЛ** возвращает значение ошибки **#Н/Д**.

Если **массив1** либо **массив2** пуст или если σ (стандартное отклонение) их значений равно нулю, то функция **КОРРЕЛ** возвращает значение ошибки **#ДЕЛ/0!**.)

D8 fx =КОРРЕЛ(A2:A17;B2:B17)

	A	B	C	D
1	Объем продаж	Индекс потребительских расходов		
2	126	100		
3	137	98,4		
4	148	101,2		
5	191	103,5		
6	274	104,1		
7	370	107		
8	432	107,4		0,816018
9	445	108,5		
10	367	108,3		
11	367	109,2		
12	321	110,1		
13	307	110,7		
14	331	110,3		
15	345	111,8		
16	364	112,3		
17	384	112,9		
18				

Рис. 3.2.2. Вычисление коэффициента парной корреляции с помощью функции **КОРРЕЛ**

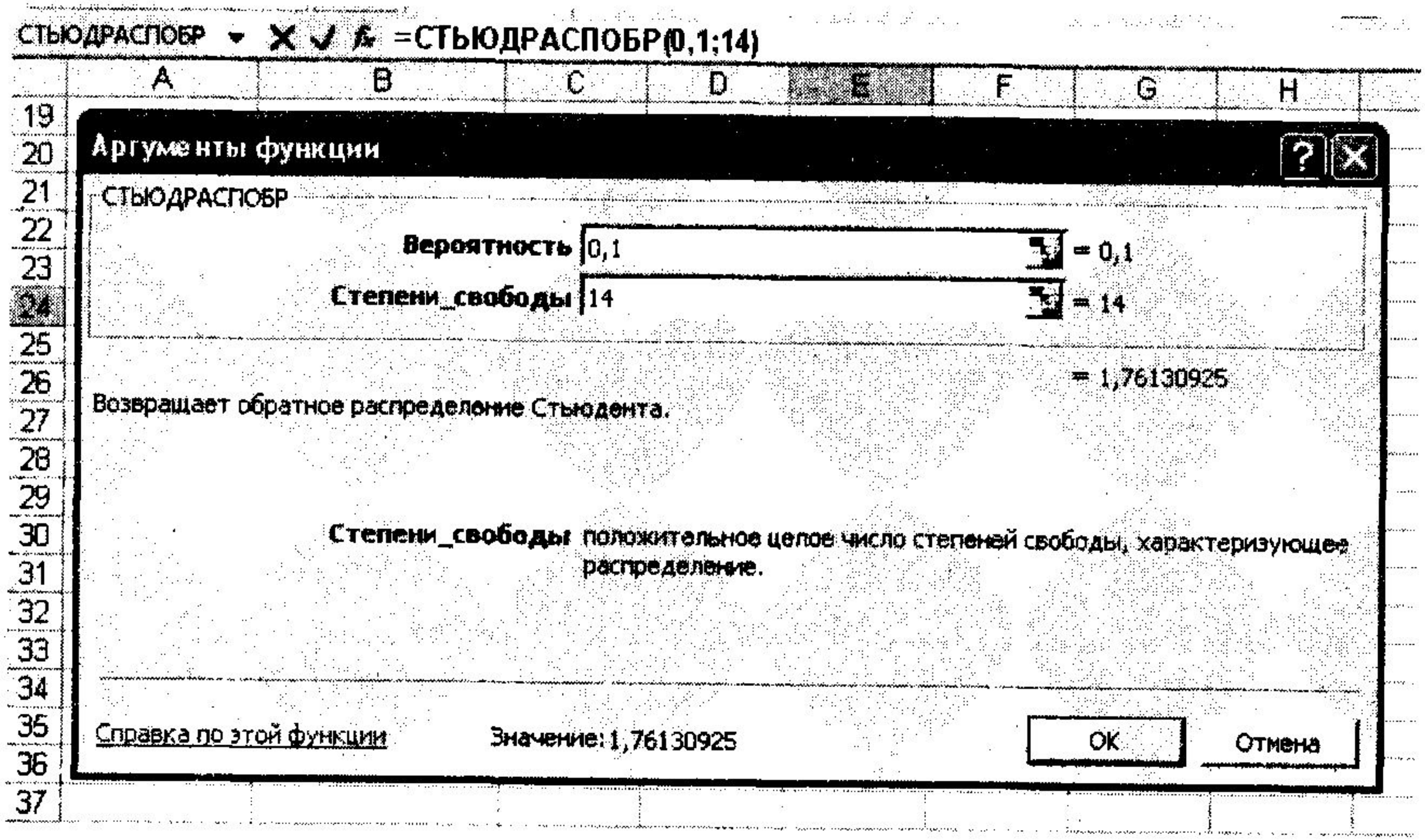

Рис. 3.2.3. Критическое значение *t*-статистики равно 1,7613

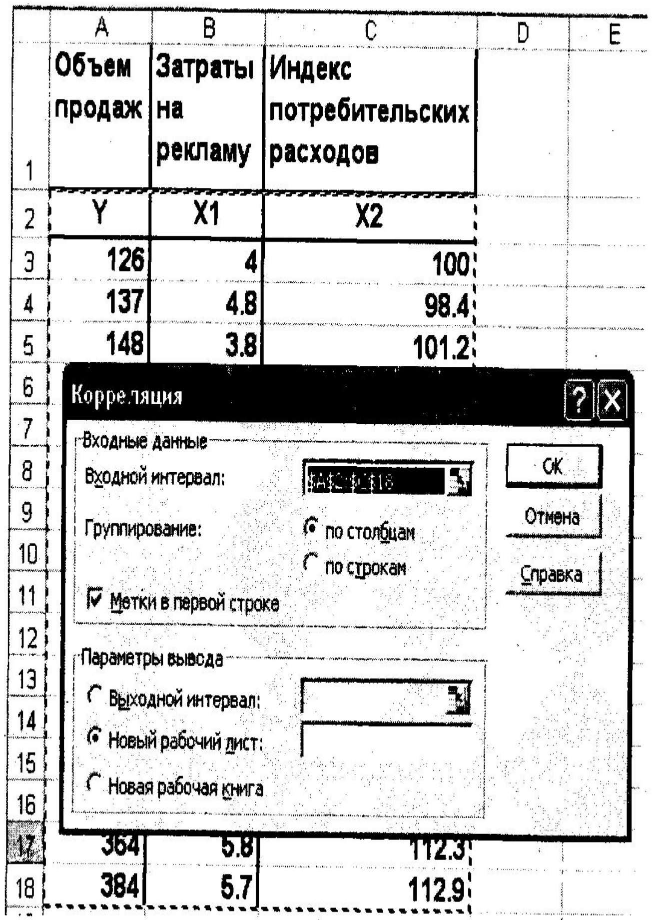

Рис. 3.2.4. Вычисление матрицы коэффициентов парной корреляции с помощью инструмента **Корреляция**

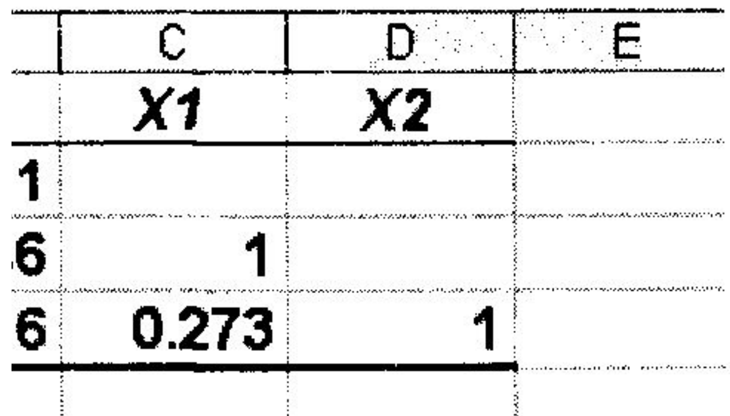

Рис. 3.2.5. Матрица коэффициентов парной корреляции

Основная задача регрессионного анализа заключается в исследовании зависимости изучаемой переменной от различных факторов и отображении их взаимосвязи в форме регрессионной модели.

В регрессионных моделях *зависимая* (объясняемая) переменная Y может быть представлена в виде функции $f(X_1, X_2, ..., X_k)$, где $X_1, X_2, ..., X_k$ — *независимые* (объясняющие) переменные, или факторы.

Связь между переменной Y и k независимыми факторами X можно охарактеризовать ***функцией регрессии*** $Y = f(X_1, X_2, ..., X_k)$, которая показывает, каково будет в среднем значение переменной Y, если переменные X_i примут конкретные значения. Данное обстоятельство позволяет использовать модель регрессии не только для анализа, но и для прогнозирования экономических явлений.

Сформулируем **регрессионную задачу для случая одного факторного признака.**

Пусть имеется набор значений двух переменных: $Y = (y_1, y_2, \dots, y_n)$ — объясняемая переменная и $X = (x_1, x_2, \dots, x_n)$ — объясняющая переменная, каждая из которых содержит n наблюдений. Пусть между переменными X и Y теоретически существует некоторая линейная зависимость

$$Y = f(X) = f(x_1, x_2, \dots, x_n) = \alpha + \beta x.$$

Это уравнение будем называть *«истинным» уравнением регрессии*.

Отклонения от предполагаемой формы связи, естественно, могут возникнуть и в силу неправильного выбора вида самого уравнения, описывающего эту зависимость. Учитывая возможные отклонения, линейное уравнение связи двух переменных (парную регрессию) представим в виде:

$$\boxed{y_i = \alpha + \beta x_i + \varepsilon_i,}$$

где α — постоянная величина (или свободный член уравнения); β — коэффициент регрессии, определяющий наклон линии, вдоль которой рассеяны данные наблюдений; ε_i — случайная переменная (случайная составляющая, остаток, или возмущение).

Коэффициент регрессии β характеризует изменение переменной y_i при изменении значения x_i на единицу. Если $\beta > 0$, переменные x_i и y_i положительно коррелированны, если $\beta < 0$ — отрицательно коррелированны.

Случайная составляющая ε_i отражает тот факт, что изменение y_i будет неточно описываться изменением x_i, так как присутствуют другие факторы, неучтенные в данной модели.

Предпосылки МНК

Свойства коэффициентов регрессии существенным образом зависят от свойств случайной составляющей. Для того чтобы регрессионный анализ, основанный на обычном методе наименьших квадратов, давал наилучшие из всех возможных результаты, должны выполняться следующие условия, известные как

условия Гаусса – Маркова.

1. *Математическое ожидание случайной составляющей в любом наблюдении должно быть равно нулю.*

$$\boxed{M(\varepsilon_i) = 0.}$$

2. *В модели (3.3.1) возмущение ε_i (или зависимая переменная y_i) есть величина случайная, а объясняющая переменная x_i — величина неслучайная.* Если это условие выполнено, то теоретическая ковариация между независимой переменной и случайным членом равна нулю.

3. *В любых двух наблюдениях отсутствует систематическая связь между значениями случайной составляющей.* Например, если случайная составляющая велика и положительна в одном наблюдении, это не должно обусловливать систематическую тенденцию к тому, что она будет большой и положительной в следующем наблюдении. Случайные составляющие должны быть независимы друг от друга.

В силу того что $M(\varepsilon_i) = M(\varepsilon_j) = 0$, данное условие можно записать следующим образом:

$$\boxed{M(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j).} \qquad (3.3.3)$$

Возмущения ε_i и ε_j некоррелированны (**условие независимости случайных составляющих** в различных наблюдениях). Это условие означает, что отклонения регрессии (а значит, и сама зависимая переменная) не коррелируют. В случае временного ряда y_t условие (3.3.3) означает отсутствие автокорреляции ряда ε_t.

4. *Дисперсия случайной составляющей должна быть постоянна для всех наблюдений.* Иногда случайная составляющая будет больше, иногда меньше, однако не должно быть априорной причины для того, чтобы она порождала бо́льшую ошибку в одних наблюдениях, чем в других.

Постоянная дисперсия обычно обозначается $\sigma^2(\varepsilon)$ или σ^2_ε, а условие записывается следующим образом:

$$\boxed{D(\varepsilon_i) = \sigma^2_\varepsilon.} \qquad (3.3.4)$$

Это **условие гомоскедастичности**, или **равноизменчивости**, случайной составляющей (возмущения).

Предположение о нормальности

Несмещенность оценки означает, что математическое ожидание остатков равно нулю. Если оценки обладают свойством несмещенности, то их можно сравнивать по разным исследованиям.

Для практических целей важна не только несмещенность, но и эффективность оценок.

Оценки считаются эффективными, если они характеризуются наименьшей дисперсией. Поэтому несмещенность оценки должна дополняться минимальной дисперсией.

Степень достоверности доверительных интервалов парамет ров регрессии обеспечивается, если оценки будут не только несмещенными и эффективными, но и *состоятельными*. ***Состоятельность оценок*** характеризует увеличение их точности с увеличением объема выборки.

Метод наименьших квадратов

дает оценки, имеющие наименьшую дисперсию в классе всех линейных оценок, если выполняются предпосылки нормальной линейной регрессионной модели.

Оценки $\hat{\alpha}$, β находят путем минимизации суммы квадратов

$$Q(\alpha, \beta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$$

по всем возможным значениям α и β при заданных (наблюдаемых) значениях $x_1, \ldots, x_n, y_1, \ldots, y_n$. Задача сводится к математической задаче поиска *точки минимума* функции двух переменных. Точка минимума находится путем приравнивания к нулю частных производных функции $z = Q(\alpha, \beta)$ по переменным α и β. Это приводит к системе нормальных уравнений

$$\begin{cases} \dfrac{\partial Q(\alpha, \beta)}{\partial \alpha} = 0, \\ \dfrac{\partial Q(\alpha, \beta)}{\partial \beta} = 0, \end{cases}$$

решением которой и является пара $\hat{\alpha}$, $\hat{\beta}$.

В результате применения МНК получаем формулы для вычисления параметров модели парной регрессии:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Такое решение может существовать только при выполнении условия

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \neq 0, \qquad (3.3.6)$$

что равносильно отличию от нуля определителя системы нормальных уравнений. Действительно, этот определитель равен

$$n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2 = n\left[\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right] = n\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Условие (3.3.6) называется **условием идентифицируемости** модели наблюдений $y_i = (\alpha + \beta x_i) + \varepsilon_i$, $i = \overline{1, n}$, и означает, что не все значения $x_1, \ldots, x_n$ совпадают между собой. При нарушении этого условия в с е точки (x_i, y_i), $i = \overline{1, n}$, лежат на одной вертикальной прямой $x = \bar{x}$.

Коэффициент детерминации

показывает долю вариации результативного признака, находящегося под воздействием изучаемых факторов, т. е. определяет, какая доля вариации признака Y учтена в модели и обусловлена влиянием на него факторов.

Коэффициент детерминации определяют следующим образом:

$$\boxed{R^2 = \frac{\text{Объясненная сумма квадратов}}{\text{Общая сумма квадратов}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.} \qquad (3.3.12)$$

Для **проверки значимости модели** регрессии используется **F-критерий Фишера**, вычисляемый как отношение дисперсии исходного ряда и несме щенной дисперсии остаточной компоненты.

Для модели парной регрессии

$$\boxed{F = \frac{R^2}{1-R^2}(n-2) = \frac{r_{y,x}^2}{1-r_{y,x}^2}(n-2).} \qquad (3.3.15)$$

В качестве *меры точности* применяют ***несмещенную оценку дисперсии остаточной компоненты*** S^2, которая представляет собой отношение суммы квадратов уровней остаточной компоненты к величине $(n-k-1)$, где k — количество факторов, включенных в модель. Квадратный корень из этой величины называется ***стандартной ошибкой***:

$$\boxed{S_e = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2}.} \qquad (3.3.16)$$

Пример

Пример 3.3.1. В табл. 3.3.1 приведена информация о среднедушевых месячных доходах и расходах по Центральному федеральному округу в 2002 г. [17]. Требуется:

1) построить однофакторную модель регрессии зависимости расходов от доходов;

2) проверить значимость параметров модели регрессии ($\alpha = 0{,}1$);

3) построить доверительный интервал для полученной модели регрессии ($\alpha = 0{,}05$). Отобразить на графике исходные данные, результаты моделирования и прогнозирования;

4) оценить расходы, если доход составит 3600 руб.

Таблица 3.3.1

Область	*№ п/п*	*Доходы, руб.*	*Расходы, руб.*
Белгородская	1	2784	2478
Брянская	2	2255	2034
Владимирская	3	2062	2019
Воронежская	4	2553	2501
Ивановская	5	1595	1668
Калужская	6	2254	2188
Костромская	7	2371	2217
Курская	8	2518	2202
Липецкая	9	2742	2392
Московская (без г. Москва)	10	3416	3354
Орловская	11	2540	2347
Рязанская	12	2510	2309
Смоленская	13	2843	2671
Тамбовская	14	2648	2201
Тверская	15	2204	1932
Тульская	16	2561	2160
Ярославская	17	3311	2921

Решение.

1. Для вычисления параметров модели воспользуемся формулами (3.3.5). Промежуточные расчеты приведены в табл. 3.3.2. Получаем

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{2\,544\,843{,}76}{2\,993\,601{,}06} = \mathbf{0{,}85},$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 2329{,}06 + 0{,}85 \cdot 2539{,}24 = \mathbf{170{,}47}.$$

Представим расчет параметров модели в матричной форме:

$$A = (X'X)^{-1}X'Y = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix} =$$

$$= \begin{pmatrix} 2{,}2127 & -0{,}0008 \\ -0{,}0008 & 0{,}0000 \end{pmatrix} \cdot \begin{pmatrix} 39\,594{,}0 \\ 103\,083\,326{,}0 \end{pmatrix} = \begin{pmatrix} 170{,}47 \\ 0{,}85 \end{pmatrix}.$$

Таблица 3.3.2

№ п/п	Доходы X	Расходы Y	$y_i - \bar{y}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})(x_i - \bar{x})$	$\hat{Y}$	Остатки e_i	e_i^2
1	2784	2478	148,94	244,76	59 909,76	36 455,54	2537,13	–59,13	3496,59
2	2255	2034	–295,06	–284,24	80 789,70	83 866,13	2087,43	–53,43	2854,98
3	2062	2019	–310,06	–477,24	227 753,53	147 971,01	1923,36	95,64	9146,30
4	2553	2501	171,94	13,76	189,47	2366,72	2340,76	160,24	25 676,82
5	1595	1668	–661,06	–944,24	891 580,29	624 195,07	1526,37	141,63	20 059,17
6	2254	2188	–141,06	–285,24	81 359,17	40 234,96	2086,58	101,42	10 285,64
7	2371	2217	–112,06	–168,24	28 303,11	18 852,25	2186,04	30,96	958,34
8	2518	2202	–127,06	–21,24	450,94	2698,13	2311,01	–109,01	11 882,49
9	2742	2392	62,94	202,76	41 113,53	12 762,25	2501,43	–109,43	11 974,48
10	3416	3354	1024,94	876,76	768 716,35	898 632,25	3074,39	279,61	78 180,82
11	2540	2347	17,94	0,76	0,58	13,72	2329,71	17,29	298,98
12	2510	2309	–20,06	–29,24	854,70	586,43	2304,21	4,79	22,98
13	2843	2671	341,94	303,76	92 273,00	103 869,66	2587,29	83,71	7007,78
14	2648	2201	–128,06	108,76	11 829,76	–13 928,28	2421,52	–220,52	48 628,67
15	2204	1932	–397,06	–335,24	112 382,70	133 108,13	2044,08	–112,08	12 561,29
16	2561	2160	–169,06	21,76	473,70	–3679,52	2347,56	–187,56	35 179,08
17	3311	2921	591,94	771,76	595 620,76	456 839,31	2985,13	–64,13	4112,88
Сумма	43 167,00	39 594,00	0	0	**2 993 601,06**	**2 544 843,76**	39 594,00	0	**282 327,28**
Среднее	**2539,24**	**2329,06**	0	0	—	149 696,69	—	0	—

Построена модель зависимости расходов от дохода:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i = \mathbf{170{,}47 + 0{,}85}x_i.$$

При увеличении дохода на 1 руб. расходы увеличиваются в среднем на 0,85 руб.

2. Расчетное значение t-критерия ($t_{\beta\,\text{расч}}$) вычислим по формулам (3.3.17) и (3.3.18), используя данные табл. 3.3.2:

$$S_\beta = \sqrt{\frac{S_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{137{,}19^2}{2\,993\,601{,}06}} = 0{,}079,$$

$$t_{\beta\,\text{расч}} = \frac{|\hat{\beta}|}{S_\beta} = \frac{0{,}85}{0{,}079} = \mathbf{10{,}72},$$

где $S_e = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2} = \sqrt{\frac{282\,327{,}28}{15}} = 137{,}19.$

Табличное значение t-критерия при 10%-ном уровне значимости и 15 степенях свободы (17 − 1 − 1 = 15) составляет 1,75 (см. Приложение 2). Так как $|t_{\text{расч}}| > t_{\text{табл}}$, то коэффициент β значим.

3. Доверительный интервал для прогнозов индивидуальных значений y_i определяется из соотношения:

$$y_i \in [\hat{y}_i \pm U_i] = \left[\hat{y}_i \pm S_e t_\alpha \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right].$$

Коэффициент Стьюдента t_α для $\nu = 15$ степеней свободы ($\nu = n - 2$) и уровня значимости 0,05 равен 2,13, тогда

$$U_{(x_i; n=17; \alpha=0{,}05)} = 137{,}19 \cdot 2{,}13 \cdot \sqrt{1 + \frac{1}{17} + \frac{(x_i - 2539{,}24)^2}{2\,993\,601{,}06}}.$$

Результаты вычислений приведены в табл. 3.3.3 и на рис. 3.3.1.

4. Для того чтобы определить расходы при доходе 3600 руб., необходимо получить точечный и интервальный прогноз аналогично тому, как это сделано в пункте 3. Для этого следует подставить значение $x_{\text{прогн}}$, равное 3600, в полученную модель:

$$\hat{y}_{\text{прогн}} = 170{,}47 + 0{,}85 \cdot 3600 \approx \mathbf{3230{,}81}.$$

Таблица 3.3.3

i	U_i	Верхняя граница	Нижняя граница	i	U_i	Верхняя граница	Нижняя граница
1	303,73	2840,86	2233,40	10	335,41	3409,80	2738,99
2	304,71	2392,14	1782,72	11	300,90	2630,61	2028,81
3	311,52	2234,88	1611,84	12	300,94	2605,14	2003,27
4	300,91	2641,67	2039,85	13	305,25	2892,53	2282,04
5	340,60	1866,97	1185,77	14	301,46	2722,98	2120,06
6	304,73	2391,32	1781,85	15	306,18	2350,26	1737,89
7	302,24	2488,28	1883,81	16	300,92	2648,48	2046,64
8	300,92	2611,93	2010,09	17	327,95	3313,08	2657,18
9	302,84	2804,27	2198,59	.			

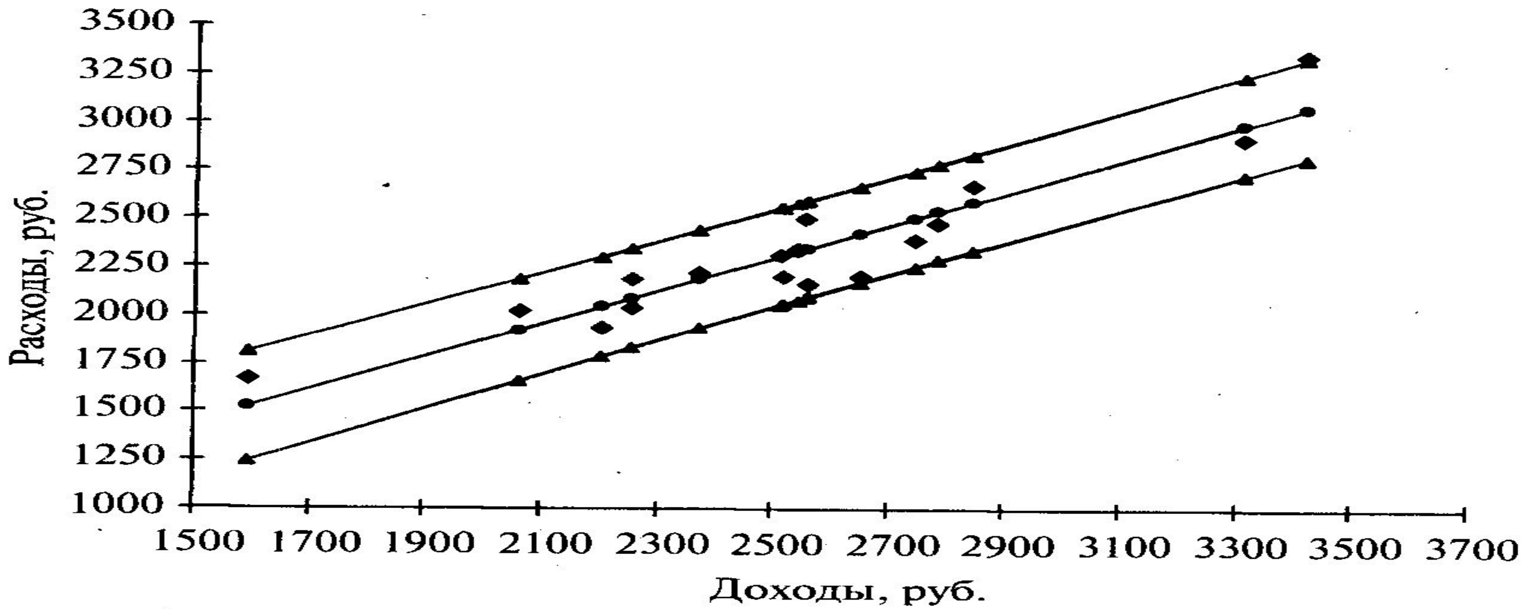

Рис. 3.3.1. График исходных данных (—◆—), результаты моделирования (—●—) и доверительные интервалы (—▲—)

Таким образом, *прогнозное значение* $\hat{y}_{\text{прогн}} = 3230{,}81$ с вероятностью 90% будет находиться между *верхней границей*, равной $3230{,}81 + 265{,}49 =$ **3496,30**, и *нижней границей*, равной $3230{,}81 - 265{,}49 =$ **2965,32** (рис. 3.3.2).

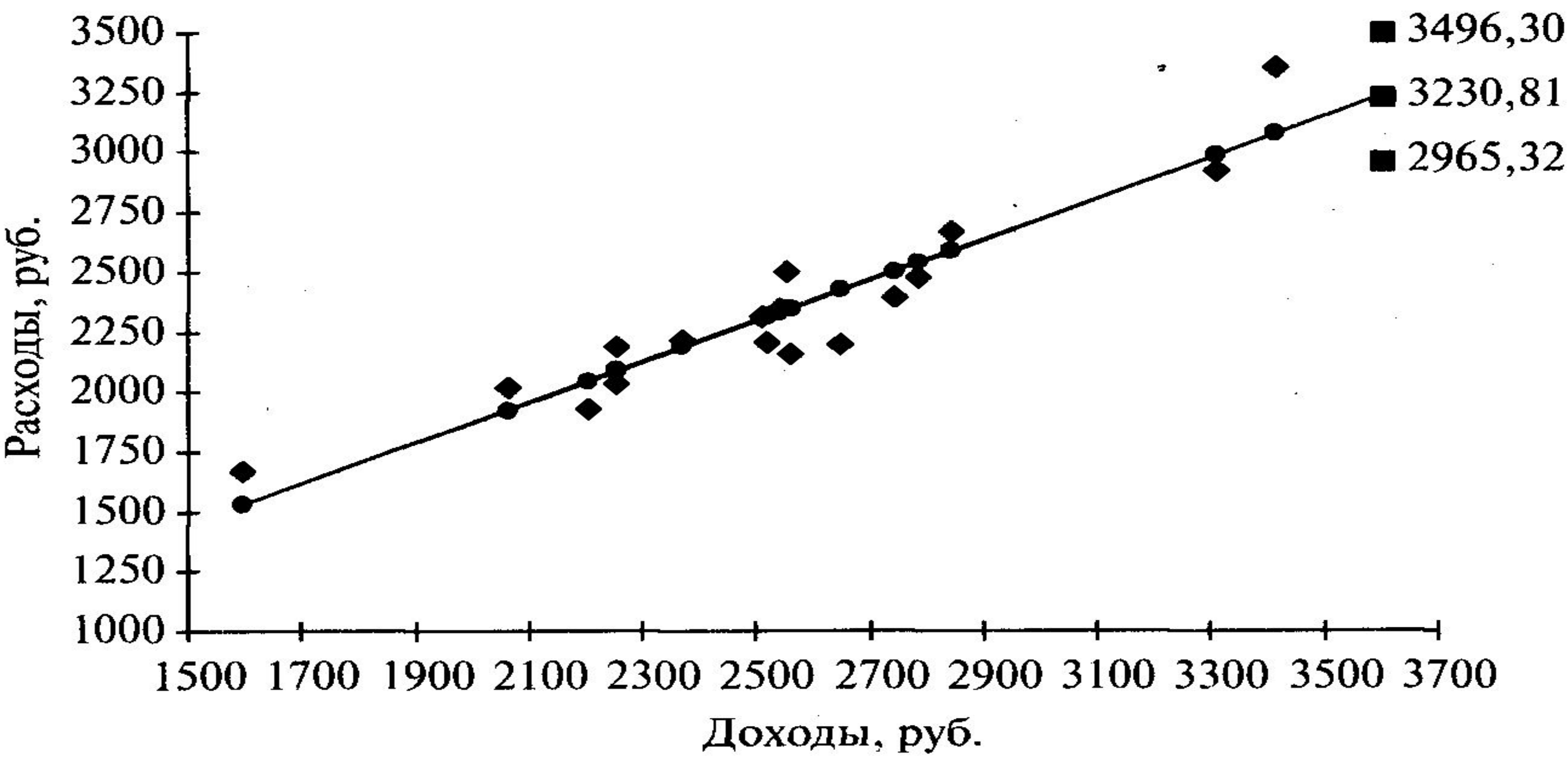

Рис. 3.3.2. График модели парной регрессии зависимости расходов от дохода, точечный и интервальный прогноз при $x = 3600$ руб.

СПАСИБО ЗА ВНИМАНИЕ!