МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Государственное образовательное учреждение высшего профессионального образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР))

Кафедра автоматизированных систем управления (АСУ)

УТВЕРЖДАЮ

Заведующий кафедрой АСУ

___________________А.М.Кориков

«___» ____________________ 2015 г.

Методические указания

по практическим и самостоятельным работам

по дисциплине

Эконометрика

Направление подготовки: 230700-62 Прикладная информатика

Форма обучения: очная

Факультет систем управления (ФСУ)

Кафедра автоматизированных систем управления (АСУ)

Курс 4 Семестр 7

Разработчик

Доцент кафедры АСУ Е.Б. Грибанова

Практические работы 28 часов

Самостоятельная работа 54 часа

2015

Содержание

[Заголовок 1] **1. Модель парной регрессии** [.]

Задача 1. Используя данные таблицы 1.1, нужно определить наличие связи между баллом, полученным за экзамен и временем, потраченным на подготовку к экзамену.

Табл.1.1 – Исходные данные

Часы изучения	3	5	4	4	2	3
Балл за экзамен	86	95	92	83	78	82

Решение. Для решения задачи дополним таблицу вспомогательными столбцами (табл.1.2).

Табл.1.2 – Вычисление характеристик

	Часы изучения	Балл за экзамен			
	x	y	xy	x^2	y^2
	3	86	258	9	7396
	5	95	475	25	9025
	4	92	368	16	8464
	4	83	332	16	6889
	2	78	156	4	6084
	3	82	246	9	6724
Сумма	21	516	1835	79	44582

$$r = \frac{6 \cdot 1835 - 21 \cdot 516}{\sqrt{6 \cdot 79 - 21^2}\sqrt{6 \cdot 44582 - 516^2}} = 0{,}862.$$

Следовательно, связь существует.

Задача 2. По данным табл.1.3 оценить на уровне α=0,05 значимость уравнения регрессии Y по Х.

Табл.1.3 – Исходные данные

i	1	2	3	4	5	6	7	8	9	10
x_i	8	11	12	9	8	8	9	9	8	12
y_i	5	10	10	7	5	6	6	5	6	8

Решение. Найдем значения сумм:

$$\sum_{i=1}^{10} y_i = 5+10+10+7+5+6+6+5+6+8 = 68;$$

$$\sum_{i=1}^{10} y_i^2 = 5^2+10^2+10^2+7^2+5^2+6^2+6^2+5^2+6^2+8^2 = 496\,.$$

Для выполнения дисперсионного анализа вычислим промежуточную таблицу 1.4.

Табл.1.4 – Вычисление характеристик

x_i	8	11	12	9	8	8	9	9	8	12	$\sum$
$(x_i-\overline{x})^2$	1,96	2,56	6,76	0,16	1,96	1,96	0,16	0,16	1,96	6,76	24,40
$\widehat{y}_i = -2,75 + 1,016x_i$	5,38	8,43	9,44	6,39	5,38	5,38	6,39	6,39	5,38	9,44	
$\varepsilon_i^2 = (\widehat{y}_i - y_i)^2$	0,14	2,48	0,31	0,37	0,14	0,39	0,15	1,94	0,39	2,08	8,39

Вычислим необходимые суммы квадратов:

$$S_{\text{îòêë}} = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 = \sum_{i=1}^{10} \varepsilon_i^2 = 8,39$$

$$S_{\text{îáù}} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{10} y_i^2 - \frac{\left(\sum_{i=1}^{10} y_i\right)^2}{10} = 496 - \frac{68^2}{10} = 33,6;$$

$$S_{\text{îñò}} = S_{\text{îáù}} - S_{\text{îòêë}} = 33,6 - 8,39 = 25,21\,.$$

F-критерий:

$$F = \frac{25,21(10-2)}{8,39} = 24,04\,.$$

По таблице F-распределения (Приложение 1) $F_{0,05;1;8}$=4,20. Так как $F > F_{0,05;1;8}$, то уравнение регрессии значимо.

2-й способ. Учитывая, что $\widehat{\Theta}_1$=1,016, $\sum_{i=1}^{10}(x_i-\overline{x})^2 = 24,40$, $s^2 = \frac{8,39}{10-2} = 1,049$ определим значение t-статистики:

$$t = \frac{\Theta_1 - 0}{s}\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2} = \frac{1,016}{\sqrt{1,049}}\sqrt{24,40} = 4,90.$$

По таблицам t-распределения (Приложение 3) $t_{0,95;8} = 2,31$. Так как $t > t_{0,95;8}$? То коэффициент регрессии Θ_1, а значит, и уравнение парной линейной регрессии Y по X значимы.

Задача 3. Для анализа зависимости объема потребления Y (у. е.) домохозяйства в зависимости от располагаемого дохода X (у. е.) отобрана выборка объема n = 12 (помесячно в течение года), результаты которой приведены в табл. 1.5. Необходимо по методу наименьших квадратов оценить параметры уравнения регрессии Y на X; спрогнозировать потребление при доходе X = 160.

Табл.1.5 – Исходные данные

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150
y_i	102	105	108	110	115	117	119	125	132	130	141	144

Для наглядности вычислений по МНК построим следующую таблицу (табл. 1.6)

Табл.1.6 – Промежуточные вычисления

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\widehat{y}_i$	ε_i	ε_i^2
1	107	102	11449	10914	10404	103,63	-1,63	2,66

2	109	105	11881	11445	11025	105,49	-0,49	0,24
3	110	108	12100	11880	11664	106,43	1,57	2,46
4	113	110	12769	12430	12100	109,23	0,77	0,59
5	120	115	14400	13800	13225	115,77	-0,77	0,59
6	122	117	14884	14274	13689	117,63	-0,63	0,40
7	123	119	15129	14637	14161	118,57	0,43	0,18
8	128	125	16384	16000	15625	123,24	1,76	3,10
9	136	132	18496	17952	17424	130,71	1,29	1,66
10	140	130	19600	18200	16900	134,45	-4,45	19,8
11	145	141	21025	20445	19881	139,11	1,89	3,57
12	150	144	22500	21600	20736	143,78	0,22	0,05
$\sum$	1503	1448	190617	183577	176834	-		35,3
Среднее значение	125,5	120,67	15884675	15298,08	14736,17	-	-	-

По МНК имеем

$$\begin{cases} \widehat{\Theta}_1 = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \dfrac{15298,08 - 125,25 \cdot 120,67}{15884,75 - (125,25)^2} = \dfrac{184,1625}{197,1875} = 0,9339 \\ \widehat{\Theta}_0 = \overline{y} - \widehat{\Theta}_1 \overline{x} = 120,67 - 0,9339 \cdot 125,25 = 3,699 \end{cases}$$

Таким образом, уравнение парной линейной регрессии имеет вид: Y = 3.699 + 0.9339X.

Прогнозируемое потребление при располагаемом доходе x = 160 по данной модели составит $\widehat{y}(160) \approx 153{,}12$.

Коэффициент Θ_1 может трактоваться как предельная склонность к потреблению (МРС « 0.9339). Фактически он показывает, на какую величину изменится объем потребления, если располагаемый доход возрастает на одну единицу. На графике коэффициент Θ_1 определяет тангенс угла наклона прямой регрессии относительно положительного направления оси абсцисс (объясняющей переменной). Поэтому часто он называется *угловым коэффициентом.*

Свободный член Θ_0 уравнения регрессии определяет прогнозируемое значение *Y* при величине располагаемого дохода *X*, равной нулю (т. е. автономное потребление). Однако здесь необходима определенная осторожность. Очень важно, насколько далеко данные наблюдений за объясняющей переменной отстоят от оси ординат (зависимой переменной), так как даже при удачном подборе уравнения регрессии для интервала наблюдений нет гарантии, что оно останется таковым и вдали от выборки. В нашем случае значение Θ_0 = 3.699 говорит о том, что при нулевом располагаемом доходе расходы на потребление составят в среднем 3.699 у. е. Это можно объяснить в случае рассмотрения отдельного домохозяйства (оно может тратить накопленные или одолженные средства), но для совокупности домохозяйств это теряет смысл. В любом случае значение коэффициента Θ_0 определяет точку пересечения прямой регрессии с осью ординат и характеризует сдвиг линии регрессии вдоль оси *Y*.

Следует помнить, что эмпирические коэффициенты регрессии $\widehat{\Theta}_0$ и $\widehat{\Theta}_1$ являются лишь оценками теоретических коэффициентов Θ_0 и Θ_1, а само уравнение отражает лишь общую тенденцию в поведении рассматриваемых переменных. Индивидуальные значения переменных в силу различных причин могут отклоняться от модельных значений. В нашем примере эти отклонения выражены через значения $\widehat{\varepsilon}_i$. Эти отклонения являются оценками отклонений ε_i для генеральной совокупности.

Задача 4. В таблице 1.7 представлена информация, характеризующая зависимость объема выпуска продукции от объема капиталовложений.

Табл. 1.7 – Исходные данные

Объем выпуска, y_i, шт.	64	56	52	48	50
Объем капиталовложений, x_i, руб.	64	68	82	76	84

Найти оценки неизвестных параметров, построить уравнение регрессии.

Решение. Запишем исходные данные:

$$Y = \begin{pmatrix} 64 & 56 & 52 & 48 & 50 \end{pmatrix}^T,$$

$$X = \begin{pmatrix} 1 & \frac{1}{64} \\ 1 & \frac{1}{68} \\ 1 & \frac{1}{82} \\ 1 & \frac{1}{76} \\ 1 & \frac{1}{84} \end{pmatrix}.$$

Найдем оценки с помощью метода наименьших квадратов:

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^{n} \frac{1}{x_i} \\ \sum_{i=1}^{n} \frac{1}{x_i} & \sum_{i=1}^{n} (\frac{1}{x_i})^2 \end{pmatrix} = \begin{pmatrix} 5 & 0,068 \\ 0,068 & 9,24 \times 10^{-4} \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} \frac{1}{x_i} y_i \end{pmatrix} = \begin{pmatrix} 270 \\ 3,684 \end{pmatrix}$$

Для нахождения обратной матрицы воспользуемся формулой:

$$A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \frac{d}{ad - bc} & \frac{-b}{ad - bc} \\ \frac{-c}{ad - bc} & \frac{a}{ad - bc} \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 17,884 & -1,308 \times 10^3 \\ -1,308 \times 10^3 & 9,678 \times 10^4 \end{pmatrix}$$

Для нахождения оценок $\widehat{\Theta}$ перемножим полученные матрицы (см. Приложение 7)

$$(X^T X)^{-1} X^T Y = \begin{pmatrix} 17{,}884 & -1{,}308\times 10^3 \\ -1{,}308\times 10^3 & 9{,}678\times 10^4 \end{pmatrix} \begin{pmatrix} 270 \\ 3{,}684 \end{pmatrix} = \begin{pmatrix} 8{,}597 \\ 3{,}359\times 10^3 \end{pmatrix}$$

Умножение выполняется по правилу

$$\begin{pmatrix} a1 & b1 \\ a2 & b2 \end{pmatrix} \begin{pmatrix} c1 \\ c2 \end{pmatrix} = \begin{pmatrix} a1c1 + b1c2 \\ a2c1 + b2c2 \end{pmatrix}.$$

Следовательно, уравнение регрессии имеет вид:

$$y = 8{,}597 + 3359\frac{1}{x} + \varepsilon .$$

Задача 5. В таблице 1.8 приведена зависимость спроса от цены. Вычислить «невязки».

Табл.1.8 Исходные данные

Цена, X	50	100	150	200
Спрос, Y	100	60	40	20

Решение. Матрицы исходных данных:

$$Y = (100 \quad 60 \quad 40 \quad 20)^T$$

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 50 & 100 & 150 & 200 \end{pmatrix}^T$$

С помощью метода наименьших квадратов находим неизвестные оценки:

$$\widehat{\Theta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 120 \\ -0{,}52 \end{pmatrix}.$$

Определим модельные значения результирующей переменной:

$$\widehat{Y} = (94 \quad 68 \quad 42 \quad 16)^T$$

«Невязки»:

$$\widehat{\varepsilon} = Y - \widehat{Y} = (6 \quad -8 \quad -2 \quad 4)^T .$$

Задача 6. Используя данные таблицы 1.9 рассчитать среднюю ошибку аппроксимации и эластичность.

Табл.1.9 Исходные данные

Область	Средний размер назначенных ежемесячных пенсий, д.е., y	Прожиточный минимум в среднем на одного пенсионера в месяц, д.е., x
Орловская	232	166
Рязанская	215	199
Смоленская	220	180
Тверская	222	181
Тульская	231	186
Ярославская	229	250

Для определения неизвестных параметров парной регрессии используем стандартную систему нормальных уравнений, которая имеет вид:

$$\begin{cases} n\Theta_0 + \Theta_1 \sum x = \sum y \\ \Theta_0 \sum x + \Theta_1 \sum x^2 = \sum xy \end{cases}$$

Для её решения рассчитаем значения сумм (табл.1.10).

Табл.1.10 – Расчет промежуточных данных

Область	y	x	x^2	xy
Орловская	232	166	27556	38512
Рязанская	215	199	39601	42785
Смоленская	220	180	32400	39600
Тверская	222	181	32761	40182
Тульская	231	186	34596	42966
Ярославская	229	250	62500	57250
Сумма	1349	1162	229414	261295

Тогда система нормальных уравнений примет вид:

$$\begin{cases} 6\Theta_0 + 1162\Theta_1 = 1349 \\ 1162\Theta_0 + 229414\Theta_1 = 261295 \end{cases}.$$

Выражая из первого уравнения Θ_0 и подставляя полученное выражение во второе, получим:

$$\Theta_0 = \frac{1349 - 1162\Theta_1}{6},$$

$$1162\frac{1349 - 1162\Theta_1}{6} + 229414\Theta_1 = 261295$$

$$261256{,}3333 - 225041{,}054\Theta_1 + 229414\Theta_1 = 261295$$

$$4372{,}946\Theta_1 = 38{,}6667$$

$$\Theta_1 = 0{,}0088.$$

Тогда

$$\Theta_0 = \frac{1349 - 1162 \cdot 0{,}0088}{6} = 223.$$

Следовательно, уравнение регрессии будет иметь вид:

$$\hat{y} = 223 + 0{,}0088x.$$

Рассчитаем модельные значения результирующей переменной и величину $\left|\frac{(y-\hat{y})}{y}\right|$ (табл.1.11).

Табл.1.11 Расчет модельного значения результирующей переменной и ошибки аппроксимации

Область	Средний размер назначенных ежемесячных пенсий, д.е., у	Прожиточный минимум в среднем на одного пенсионера в месяц, д.е., х	$\hat{y}$	$\left\|\frac{(y-\hat{y})}{y}\right\|$
Орловская	232	166	224	0,032

Рязанская	215	199	225	0,045
Смоленская	220	180	225	0,021
Тверская	222	181	225	0,012
Тульская	231	186	225	0,028
Ярославская	229	250	225	0,017
Сумма	1349	1162		0,155

Тогда средняя ошибка аппроксимации равна

$$A = \frac{0,155}{6} \times 100\% = 2,6\% .$$

Эластичность вычисляем по формуле:

$$\acute{Y} = \Theta_1 \frac{\overline{x}}{\overline{y}} = 0,0088 \frac{1162/6}{1349/6} = 0,000758\% .$$

Следовательно, при изменении прожиточного минимума на 1% величина ежемесячной пенсии изменяется на 0,000758%.

Задача 7. По данным годовых отчетов десяти машиностроительных предприятий известна зависимость производительности труда y от объема производства x, которая представлена в таблице 1.12.

Табл.1.12 Исходные данные

№ предприятия	1	2	3	4	5	6	7	8	9	10
y_i (тыс.руб.)	2,1	2,8	3,2	4,5	4,8	4,9	5,5	6,5	12,1	15,1
x_i (млн.руб.)	3	4	5	5	5	5	6	7	15	20

Найти несмещенную оценку дисперсии.

Решение. Несмещенная оценка дисперсии определяется по формуле:

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\left(Y - X\hat{\Theta}\right)^T\left(Y - X\hat{\Theta}\right).$$

Найдем значения оценок.

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 5 & 5 & 5 & 6 & 7 & 15 & 20 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 5 \\ 1 & 5 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 15 \\ 1 & 20 \end{pmatrix} = \begin{pmatrix} 10 & 75 \\ 75 & 835 \end{pmatrix}.$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 5 & 5 & 5 & 6 & 7 & 15 & 20 \end{pmatrix} \begin{pmatrix} 2,1 \\ 2,8 \\ 3,2 \\ 4,5 \\ 4,8 \\ 4,9 \\ 5,5 \\ 6,5 \\ 12,1 \\ 15,1 \end{pmatrix} = \begin{pmatrix} 61,4 \\ 664,5 \end{pmatrix}.$$

Найдем обратную матрицу

$$\left(X^T X\right)^{-1} = \frac{1}{10 \cdot 835 - (75)^2} \begin{pmatrix} 835 & -75 \\ -75 & 10 \end{pmatrix} = \begin{pmatrix} 0,306422 & -0,0275229 \\ -0,0275229 & 0,0036697 \end{pmatrix}.$$

Тогда вектор оценок коэффициентов регрессии равен:

$$\hat{\Theta} = \begin{pmatrix} 0,306422 & -0,0275229 \\ -0,0275229 & 0,0036697 \end{pmatrix} \cdot \begin{pmatrix} 61,4 \\ 664,5 \end{pmatrix} = \begin{pmatrix} 0,5253430 \\ 0,7486096 \end{pmatrix}.$$

$$Y - X\hat{\Theta} = \begin{pmatrix} 2,1 \\ 2,8 \\ 3,2 \\ 4,5 \\ 4,8 \\ 4,9 \\ 5,5 \\ 6,5 \\ 12,1 \\ 15,1 \end{pmatrix} - \begin{pmatrix} 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 5 \\ 1 & 5 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 15 \\ 1 & 20 \end{pmatrix} \begin{pmatrix} 0,5253430 \\ 0,7486096 \end{pmatrix} = \begin{pmatrix} 2,1 \\ 2,8 \\ 3,2 \\ 4,5 \\ 4,8 \\ 4,9 \\ 5,5 \\ 6,5 \\ 12,1 \\ 15,1 \end{pmatrix} - \begin{pmatrix} 2,77 \\ 3,52 \\ 4,27 \\ 4,27 \\ 4,27 \\ 4,27 \\ 5,02 \\ 5,77 \\ 11,75 \\ 15,50 \end{pmatrix} = \begin{pmatrix} -0,67 \\ -0,72 \\ -1,07 \\ 0,23 \\ 0,53 \\ 0,63 \\ 0,48 \\ 0,73 \\ 0,35 \\ -0,40 \end{pmatrix}.$$

$$(Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) = \begin{pmatrix} -0,67 \\ -0,72 \\ -1,07 \\ 0,23 \\ 0,53 \\ 0,63 \\ 0,48 \\ 0,73 \\ 0,35 \\ -0,40 \end{pmatrix}^T \cdot \begin{pmatrix} -0,67 \\ -0,72 \\ -1,07 \\ 0,23 \\ 0,53 \\ 0,63 \\ 0,48 \\ 0,73 \\ 0,35 \\ -0,40 \end{pmatrix} = 0,4489 + 0,5184 + 1,449 + 0,0529 +$$

$$+0,2809 + 0,3969 + 0,2304 + 0,5329 + 0,1225 + 0,16 = 3,8887$$

Таким образом, несмещенная оценка дисперсии равна:

$$\hat{\sigma}^2 = \frac{1}{8} \cdot 3,8887 = 0,4861.$$

[Заголовок 1] 2. Модель множественной регрессии [.]

Задача 1. Имеются следующие данные (условные) о сменной добыче угля на одного рабочего Y(т), мощности пласта $X^{(1)}$(м) и уровне механизации работ $X^{(2)}$ (%), характеризующие процесс добычи угля в 10 шахтах (табл. 2.1).

Таблица 2.1 – Исходные данные

i	$x_i^{(1)}$	$x_i^{(2)}$	y_i	i	$x_i^{(1)}$	$x_i^{(2)}$	y_i
1	8	5	5	6	8	8	6
2	11	8	10	7	9	6	6
3	12	8	10	8	9	4	5
4	9	5	7	9	8	5	6
5	8	7	5	10	12	7	8

Необходимо оценить сменную добычу угля на одного рабочего для шахт с мощностью пласта 8м и уровнем механизации работ 6%; найти 95%-ые доверительные интервалы для индивидуального и среднего значений сменной добычи угля на 1 рабочего для таких же шахт. Проверить значимость коэффициентов регрессии построить для них 95%-ные доверительные интервалы.

Решение. С помощью МНК найдем неизвестные параметры, полученное уравнение регрессии будет иметь вид: $\hat{y} = -3,54 + 0,854x^{(1)} + 0,367x^{(2)}$.

По условию надо оценить $M_x(Y)$, где $\left(X^{(0)}\right)^T = \begin{pmatrix}1 & 8 & 6\end{pmatrix}$. Выборочной оценкой $M_x(Y)$ является групповая средняя, которую найдем по уравнению регрессии:

$$\hat{y} = -3,54 + 0,854 \cdot 8 + 0,367 \cdot 6 = 5,49 \text{ (т.)}$$

Для построения доверительного интервала для $M_x(Y)$ необходимо знать дисперсию его оценки— $\hat{\sigma}^2_{\hat{y}}$. Для ее вычисления обратимся к табл.2.2.

Табл.2.2 – Расчет регрессионные остатков

i	1	2	3	4	5	6	7	8	9	10	$\sum$
$\hat{y}_i$	5,13	8,79	9,64	5,98	5,86	6,23	6,35	5,61	5,13	9,28	
$\varepsilon_i^2 = (\hat{y}_i - y_i)^2$	0,016	1,464	1,127	1,038	0,741	0,052	0,121	0,377	0,762	1,631	6,329

$$\sigma^2 = \frac{6,329}{10-2-1} = 0,904 \text{ и } \sigma = \sqrt{0,904} = 0,951 \text{ (т.).}$$

Определяем стандартную ошибку групповой средней $\widehat{y}$. Вначале найдем

$$\left(X^{(0)}\right)^T (X^T X)^{-1} X^{(0)} = (1 \quad 8 \quad 6)\frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}\begin{pmatrix} 1 \\ 8 \\ 6 \end{pmatrix} =$$

$$= \frac{1}{3738}(2223 \quad -249 \quad 78)\begin{pmatrix} 1 \\ 8 \\ 6 \end{pmatrix} = \frac{1}{3738}(699) = 0,1870$$

Теперь $\widehat{\sigma}_{\widehat{y}} = 0,951\sqrt{0,1870} = 0,411$ (т.)

По табл.приложения 2 при числе степеней свободы k=10-2-1=7 находим $t_{0,95;7}$=2,36 . Доверительный интервал для $M_x(Y)$ равен

$$5,49 - 2,36 \cdot 0,411 \le M_x(Y) \le 5,49 + 2,36 \cdot 0,411$$

или

$$4,52 \le M_x(Y) \le 6,46 \text{ (т.).}$$

Итак, с надежностью 0,95 средняя сменная добыча угля на одного рабочего для шахт с мощностью пласта 8м и уровнем механизации работ 6% находится в пределах от 4,52 до 6,46т.

Найдем доверительный интервал для индивидуального значения y_0^* при $\left(X^{(0)}\right)^T = (1 \quad 8 \quad 6)$:

$$\widehat{\sigma}_{\widehat{y}_0} = 0,951\sqrt{1 + 0,1870} = 1,036 \text{ (т.)}$$

$$5,49 - 2,36 \cdot 1,036 \le y_0^* \le 5,49 + 2,36 \cdot 1,036, \text{ т.е.}$$

$$3,05 \le y_0^* \le 7,93 \text{ (т.)}$$

Итак, с надежностью 0,95 индивидуальное значение сменной добычи угля в шахтах с мощностью пласта 8м и уровнем механизации работ 6 % находится в пределах от 3,05 до 7,93 (т).

Проверим значимость коэффициентов регрессии $\widehat{\Theta}_1$ и $\widehat{\Theta}_2$. Были получены значения $\widehat{\Theta}_1$ =0,854 и $\widehat{\Theta}_2$ =0,367. Стандартная ошибка равна

$$\widehat{\sigma}_{\widehat{\Theta}_1} = 0,951\sqrt{\frac{1}{3738}\cdot 201} = 0,221.$$

Так как $t = \dfrac{0,854}{0,221} = 3,81 > t_{0,95;7} = 2,36$, то коэффициент $\widehat{\Theta}_1$ значим.

Аналогично вычисляем $\widehat{\sigma}_{\widehat{\Theta}_2} = 0,951\sqrt{\dfrac{1}{3738}\cdot 244} = 0,243$ и $t = \dfrac{0,367}{0,243} = 1,51 < t_{0,95;7} = 2,36$. Т.е. коэффициент $\widehat{\Theta}_2$ незначим на 5%-ном уровне.

Доверительный интервал имеет смысл построить только для значимого коэффициента регрессии $\widehat{\Theta}_1$:

$$0,854 - 2,36\cdot 0,221 \le \Theta_1 \le 0,854 + 2,36\cdot 0,221 \text{ или } 0,332 \le \Theta_1 \le 1,376.$$

Итак, с надежностью 0,95 за счет изменения на 1м мощности пласта $X^{(1)}$ (при неизменном $X^{(2)}$) сменная добыча угля на одного рабочего Y будет изменяться в пределах от 0,332 до 1,376 (т).

Формально переменные, имеющие незначимые коэффициенты регрессии, могут быть исключены из рассмотрения. В экономических исследованиях исключению переменных из регрессии должен предшествовать тщательный качественный анализ.

Поэтому может оказаться целесообразным все же оставить в регрессионной модели одну или несколько объясняющих переменных, не оказывающих существенного (значимого) влияния на зависимую переменную.

Задача 2. Для исследования зависимости между производительностью труда ($x^{(1)}$), возрастом ($x^{(2)}$) и производственным стажем ($x^{(3)}$) была произведена выборка из 100 рабочих одной и той же специальности.

Вычисленные парные коэффициенты корреляции оказались значимыми составили: r_{12}=0,20; r_{13}=0,41; r_{23}=0,82. Вычислить частные коэффициенты корреляции и оценить их значимость на уровне α=0,05.

Решение. По формуле (6.2) частные коэффициенты корреляции

$$r_{12.3}=\frac{0,20^2-0,41\cdot 0,82}{\sqrt{(1-0,41^2)(1-0,82^2)}}=-0,26$$

и аналогично $r_{13.2}=0,44$; $r_{23.1}=0,83$.

Оценим значимость $r_{12.3}$. Значение статистики t-критерия по при $n'=n-p+2$ =100-3+2=99 (по абсолютной величине)

$$|t|=\frac{|-0,26|\sqrt{99-2}}{\sqrt{1-(-0,26)^2}}=2,65$$

Больше табличного $t_{0,95;97}=1,99$, следовательно, частный коэффициент корреляции $r_{12.3}$ значим.

Аналогично устанавливается значимость других частных коэффициентов корреляции.

Сравнивая частные коэффициенты корреляции $r_{ij.k}$ с соответствующими парными коэффициентами, видим, что за счет «очищения связи» наибольшему изменению подвергся коэффициент корреляции между производительностью труда ($x^{(1)}$) и возрастом ($x^{(2)}$) рабочих (изменилось не только его значение, но и знак: r_{12}=0,20; $r_{12.3}$=-0,26, причем оба эти коэффициента значимы).

Итак, между производительностью труда ($x^{(1)}$) и возрастом ($x^{(2)}$) рабочих существует прямая корреляционная связь (r_{12}=0,20). Если же устранить (элиминировать) влияние переменной «производственный стаж» ($x^{(3)}$), то в чистом виде производительность труда ($x^{(1)}$) находится в обратной по направлению (и опять же слабой по тесноте) связи с возрастом рабочих ($x^{(2)}$) ($r_{12.3}$=-0,26). Это вполне объяснимо, если рассматривать возраст только как показатель работоспособности организма на определенном этапе его жизнедеятельности. Подобным образом могут быть интерпретированы и другие частные коэффициенты корреляции.

Задача 3. При оценке регрессии с двумя объясняющими переменными по 30 наблюдениям $R^2 = 0.65$. Оценить значимость уравнения регрессии.

Решение. Вычислим F- критерий*:*

$$F = \frac{0,65}{0,35} \cdot \frac{30-3}{3-1} \approx 25,07$$

По таблицам критических точек распределения Фишера найдем $F_{0,05;2;27} = 3,36$; $F_{0,01;2;27} = 5,49$. Поскольку $F_{\text{íàáë}}$ = 25.07 > $F_{\text{êðèò}}$ как при 5%, так и при 1% уровне значимости, то нулевая гипотеза в обоих случаях отклоняется.

Если в той же ситуации R^2=0,4, то $F = \frac{0,4}{0,6}\frac{27}{2} = 9$. Предположение о незначимости связи отвергается и здесь.

Задача 4. Рассматривается двухфакторная производственная функция Кобба-Дугласа

$$Y = AK^{\alpha}L^{\beta}.$$

Эта функция обычно применяется в макроэкономических исследованиях при анализе взаимосвязи между объемом полученного валового внутреннего продукта (Y) и используемыми ресурсами (K – основные фонды и L – затраты живого труда). Её “классический” вариант предполагает, что значения α и β удовлетворяют следующим ограничениям: $\alpha+\beta$=1; α, β≥0.

Пусть A=2,248, α=0,404, β=0,803. Необходимо определить эластичность производственной функции по фондам и по труду.

Решение. Представленная функция является степенной функцией вида:

$$y_i = \Theta_0 (x_i^{(1)})^{\Theta_1} (x_i^{(2)})^{\Theta_2} \varepsilon^e.$$

Расчет эластичности выполняется по формуле:

$$e = \frac{\partial(\ln y)}{\partial(\ln x)} = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y}$$

Получим:

$$e_{x^{(1)}} = \Theta_0 \Theta_1 x^{(1)(\Theta_1 - 1)} x^{(2)(\Theta_2)} \cdot \frac{x^{(1)}}{\Theta_0 x^{(1)\Theta_1} x^{(2)(\Theta_2)}} = \Theta_1,$$

$$e_{x^{(2)}} = \Theta_0 x^{(1)(\Theta_1)} \Theta_2 x^{(2)(\Theta_2 - 1)} \cdot \frac{x^{(2)}}{\Theta_0 x^{(1)\Theta_1} x^{(2)(\Theta_2)}} = \Theta_2.$$

Т.о. эластичность производственной функции по фондам равна 0,404, эластичность производственной функции по труду составляет 0,803.

В данном случае выполняется неравенство и β > α, таким образом, имеет место фондосберегающий (экстенсивный) рост экономики, когда увеличение трудовых ресурсов на 1% приводит к большему росту объема производства, нежели такое же увеличении фондов. Если же эластичность выпуска по фондам больше эластичности выпуска по труду, экономика имеет трудосберегающий (интенсивный) рост.

[Заголовок 1] 3. Нелинейная регрессия [.]

Задача 1. Используя данные табл.3.1 выполнить тест Зарембки.

Табл.3.1 Исходные данные

x_i	1	2	3	4
y_i	10	13	15	16

Решение.

Шаг 1. Вычисляем среднее геометрическое значений зависимой переменной и все её значения делятся на это среднее:

$$y^*_i = \frac{y_i}{\sqrt[n]{y_1 y_2 \ldots y_n}} = \begin{pmatrix} 0,752 \\ 0,978 \\ 1,129 \\ 1,204 \end{pmatrix}$$

Шаг 2. Рассчитаем линейную ($y^*_i = \Theta_0 + \Theta_1 x_i + \varepsilon$) и логарифмическую регрессии ($\ln y^*_i = \Theta_0 + \Theta_1 \ln x_i + \varepsilon$) и сравним значения их суммы квадратов остатков.

Исходные данные для линейной регрессии:

$$X1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix},$$

$$Y1 = \left(0,752, \quad 0,978, \quad 1,129, \quad 1,204\right)^T.$$

Оценки неизвестных параметров равны:

$$(X1^T X1)^{-1} X1^T Y1 = \begin{pmatrix} 0,64 \\ 0,15 \end{pmatrix}.$$

Линейная функция регрессии имеет вид:

$$\widehat{y} = 0,64 + 0,15x$$

Подставим значения x из таблицы и определим модельные значения результирующей величины:

$$\widehat{Y} = \begin{pmatrix} 0,79 \\ 0,94 \\ 1,09 \\ 1,24 \end{pmatrix}$$

Остаточная сумма квадратов отклонений:

$$S_{îñò} = \sum_{i=1}^{4} (y_i - \widehat{y}_i)^2 = 5,665 \times 10^{-3}$$

Аналогично выполним расчеты для логарифмической функции.

Исходные данные:

$$X2 = \begin{pmatrix} 1 & \ln(x_1) \\ 1 & \ln(x_2) \\ 1 & \ln(x_3) \\ 1 & \ln(x_4) \end{pmatrix} = \begin{pmatrix} 1 & \ln(1) \\ 1 & \ln(2) \\ 1 & \ln(3) \\ 1 & \ln(4) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0,693 \\ 1 & 1,099 \\ 1 & 1,386 \end{pmatrix},$$

$$Y2 = \left(\ln(0,752), \quad \ln(0,978), \quad \ln(1,129), \quad \ln(1,204)\right)^T =$$

$$\left(-0,284, \quad -0,022, \quad 0,121, \quad 0,186\right)^T$$

Оценки неизвестных параметров равны:

$$(X2^T X2)^{-1} X2^T Y2 = \begin{pmatrix} -0,275 \\ 0,346 \end{pmatrix}.$$

Логарифмическая функция регрессии имеет вид:

$$\widehat{y} = -0,275 + 0,346\ln(x)$$

Подставим значения x из таблицы и определим модельные значения результирующей величины:

$$\widehat{Y} = \begin{pmatrix} -0,275 \\ -0,035 \\ 0,105 \\ 0,205 \end{pmatrix}$$

Остаточная сумма квадратов отклонений:

$$S_{\text{îñò}} = \sum_{i=1}^{4} (Y2_i - \widehat{y}_i)^2 = 8,782 \times 10^{-4}.$$

Шаг 3. Вычислим хи квадрат статистику для оценки значимости различий

$$\chi^2 = \frac{n}{2} \left| \ln \frac{S1_{\text{îñò}}}{S2_{\text{îñò}}} \right| = 3,728.$$

Шаг 4. Сравним с критическим значением хи-квадрат распределения с одной степенью свободы для уровня значимости 0,95. Значение $\chi^2_{\text{êð}}$ равно 0,00393, что меньше полученного значения χ^2. Это означает, что значения различимы, т.е. есть различие между использованием линейной и логарифмической функции. Сравнивая значения остаточной суммы квадратов отклонений можно прийти к выводу, что предпочтительным является использование логарифмической функции регрессии.

Задача 2. Используя данные табл.3.2, найти оценки параметров, считая, что рассматривается статистическая зависимость вида

$$y = \frac{1}{\Theta_0 + \Theta_1 x + \varepsilon}, \qquad (-\frac{\Theta_0}{\Theta_1} < x < \infty).$$

Табл. 3.2 – Исходные данные

Объем выпуска, y_i, шт.	64	56	52	48	50
Объем капиталовложений, x_i, руб.	64	68	82	76	84

Решение. Исходные данные:

$$Y = \begin{pmatrix} \frac{1}{64} & \frac{1}{56} & \frac{1}{52} & \frac{1}{48} & \frac{1}{50} \end{pmatrix}^T,$$

$$X = \begin{pmatrix} 1 & 64 \\ 1 & 68 \\ 1 & 82 \\ 1 & 76 \\ 1 & 84 \end{pmatrix}.$$

Найдем оценки с помощью метода наименьших квадратов:

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} (x_i)^2 \end{pmatrix} = \begin{pmatrix} 5 & 374 \\ 374 & 2{,}828 \times 10^4 \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} \sum_{i=1}^{n} \frac{1}{y_i} \\ \sum_{i=1}^{n} x_i \frac{1}{y_i} \end{pmatrix} = \begin{pmatrix} 0,094 \\ 7,055 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 18,801 & -0,249 \\ -0,249 & 3,324 \times 10^{-3} \end{pmatrix}$$

Для нахождения оценок $\widehat{\Theta}$ перемножим полученные матрицы

$$(X^T X)^{-1} X^T Y = \begin{pmatrix} 18,801 & -0,249 \\ -0,249 & 3,324 \times 10^{-3} \end{pmatrix} \begin{pmatrix} 0,094 \\ 7,055 \end{pmatrix} = \begin{pmatrix} 4,465 \times 10^{-3} \\ 1,904 \times 10^{-4} \end{pmatrix}$$

Следовательно, уравнение регрессии имеет вид:

$$\frac{1}{y} = 4,465 \times 10^{-3} + 1,904 \times 10^{-4} x + \varepsilon .$$

[Заголовок 1] 4. Характеристики случайных величин [.]

Задача 1. В таблице 4.1 представлены значения цены акции. Определить среднее значение, дисперсию по выборке и дисперсию по генеральной совокупности, среднее квадратическое отклонение.

Табл.4.1 Исходные данные

Месяц	1	2	3	4	5	6	7	8	9	10
Цена	100	102	103	99	97	101	107	115	125	127

Решение. Среднее значение:

$$M(X) = (100 + 102 + 103 + 99 + 97 + 101 + 107 + 115 + 125 + 127)/10 = 107,6$$

Дисперсия по генеральной совокупности:

$$D(X) = ((100 - 107,6)^2 + (102 - 107,6)^2 + (103 - 107,6)^2 + (99 - 107,6)^2 + \\ +(97 - 107,6)^2 + (101 - 107,6)^2 + (107 - 107,6)^2 + (115 - 107,6)^2 + (125 - 107,6)^2 + \\ +(127 - 107,6)^2)/10 = 107,44$$

Среднее квадратическое отклонение:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{107,44} = 10,37 .$$

Дисперсия по выборке:

$D(X) = ((100-107,6)^2 + (102-107,6)^2 + (103-107,6)^2 + (99-107,6)^2 +$
$+(97-107,6)^2 + (101-107,6)^2 + (107-107,6)^2 + (115-107,6)^2 + (125-107,6)^2 +$
$+(127-107,6)^2)/9 = 119,38$

$\sigma(X) = \sqrt{D(X)} = \sqrt{119,38} = 10,93$

Задача 2. На станции технического обслуживания анализируются затраты времени на ремонт автомобилей. На основании данных, полученных по 100 автомобилям, выяснилось, что для 25 из них требуется 1 ч для проведения профилактических работ. Мелкий ремонт требуется для 40 автомобилей, что занимает 2 ч. Для 20 автомобилей требуется ремонт с заменой отдельных узлов, что занимает в среднем 5 ч. 10 автомобилей могут быть отремонтированы за 10 ч. Для 5 автомобилей необходимое время ремонта составляет 20 ч. Построить закон распределения СВ X - времени обслуживания случайно выбранного автомобиля. Вычислить мат.ожидание, дисперсию, среднее квадратическое отклонение.

Решение. Закон распределения представлен в таблице.

x_i	1	2	5	10	20
p_i	0,25	0,40	0,20	0,10	0,05

Математическое ожидание:

$M(X) = 1\cdot 0,25 + 2\cdot 0,4 + 5\cdot 0,2 + 10\cdot 0,1 + 20\cdot 0,05 = 4,05$

Дисперсия:

$D(X) = (1-4,05)^2\cdot 0,25 + (2-4,05)^2\cdot 0,4 + (5-4,05)^2\cdot 0,2 +$
$+(10-4,05)^2\cdot 0,1 + (20-4,05)^2\cdot 0,05 = 20,4475$

Среднее квадратическое отклонение:

$\sigma(X) = \sqrt{20,4475} = 4,5219$.

[Заголовок 1] **5. Гетероскедастичность** [.]

Задача 1. В таблице 5.1 представлена зависимость среднедушевых сбережений от среднедушевого дохода. С помощью графического анализа остатков определить наличие гетероскедастичности.

Табл.5.1 Исходные данные

Среднедушевые сбережения, д.е., Y	15,2	10,7	18,5	14,9	24,1	10,3	14,2	31	20,4	20
Среднедушевой доход, д.е., X	80	80	80	80	80	80	80	80	80	80
Среднедушевые сбережения, д.е., Y	60,1	35	43	29	17	48,2	18,9	53	39,4	66,2
Среднедушевой доход, д.е., X	120	120	120	120	120	120	120	120	120	120
Среднедушевые сбережения, д.е., Y	49,6	69,4	77,8	43	31,8	62,6	100,2	68,8	78	29,6
Среднедушевой доход, д.е., X	160	160	160	160	160	160	160	160	160	160
Среднедушевые сбережения, д.е., Y	125,5	88,3	52	58,8	84	79	95,5	120,8	98,1	19,7
Среднедушевой доход, д.е., X	200	200	200	200	200	200	200	200	200	200

Решение. С помощью метода наименьших квадратов определяем оценки:

$$\widehat{\Theta} = \begin{pmatrix} -23,95 \\ 0,53 \end{pmatrix}.$$

В таблице приведены значения оценок результирующей переменной $\widehat{Y} = -23,95 + 0,53X$ и остатки $\widehat{\varepsilon} = Y - \widehat{Y}$ (табл. 5.2):

Табл.5.2 Значения оценок результирующей переменной и остатков

X	80	80	80	80	80	80	80	80	80	80
$\widehat{Y}$	18,6	18,6	18,6	18,6	18,6	18,6	18,6	18,6	18,6	18,6
$\widehat{\varepsilon}$	-3,4	-7,9	-0,1	-3,7	5,5	-8,3	-4,4	12,4	1,8	1,4
X	120	120	120	120	120	120	120	120	120	120
$\widehat{Y}$	39,9	39,9	39,9	39,9	39,9	39,9	39,9	39,9	39,9	39,9
$\widehat{\varepsilon}$	20,2	-4,9	3,1	-10,9	-22,9	8,3	-21,0	13,1	-0,5	26,3
X	160	160	160	160	160	160	160	160	160	160
$\widehat{Y}$	61,2	61,2	61,2	61,2	61,2	61,2	61,2	61,2	61,2	61,2

$\hat{\varepsilon}$	-11,6	8,2	16,6	-18,2	-29,4	1,4	39,0	7,6	16,8	-31,6
X	200	200	200	200	200	200	200	200	200	200
$\hat{Y}$	82,5	82,5	82,5	82,5	82,5	82,5	82,5	82,5	82,5	82,5
$\hat{\varepsilon}$	43,0	5,8	-30,5	-23,7	1,5	-3,5	13,0	38,3	15,6	-62,8

На основе полученных значений остатков построим график (рис.5.1).

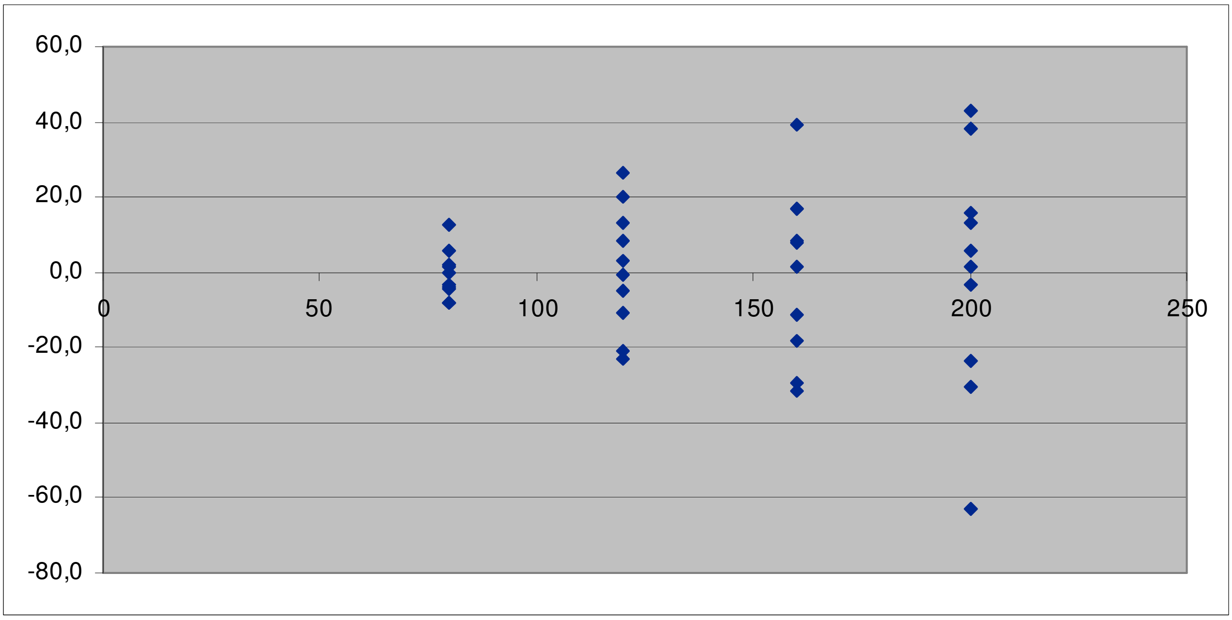

Рис. 5.1 – График остатков

Согласно построенному графику разброс значений случайных остатков увеличивается с увеличением значения объясняющей переменной. Можно сделать вывод о наличии гетероскедастичности.

Задача 2. Используя исходные данные таблицы 5.3, найти оценки неизвестных параметров функции регрессии а) используя обобщенный МНК, б) путем нормирования переменных.

Табл.5.3- Исходные данные

X	8	11	12	9
Y	5,11	8,81	9,71	5,82
σ^2	0,22	0,66	0,79	0,3

Решение.

а) Запишем исходные данные:

$$X = \begin{pmatrix} 1 & 8 \\ 1 & 11 \\ 1 & 12 \\ 1 & 9 \end{pmatrix}, \quad Y = (5{,}11 \quad 8{,}81 \quad 9{,}71 \quad 5{,}82)^T, \quad V = \begin{pmatrix} 0{,}22 & 0 & 0 & 0 \\ 0 & 0{,}66 & 0 & 0 \\ 0 & 0 & 0{,}79 & 0 \\ 0 & 0 & 0 & 0{,}3 \end{pmatrix}.$$

Используя МНК, находим оценки неизвестных параметров:

$$\widehat{\Theta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y = \begin{pmatrix} -4{,}616 \\ 1{,}196 \end{pmatrix}.$$

Выполним также расчет с помощью формул (4.1):

$\lambda_i = \left(0{,}22 \quad 0{,}66 \quad 0{,}79 \quad 0{,}3\right)$,

$x_i = \left(8 \quad 11 \quad 12 \quad 9\right)$,

$y_i = (5{,}11 \quad 8{,}81 \quad 9{,}71 \quad 5{,}82)$.

$$\widehat{\Theta}_1 = \frac{\left(\sum_{i=1}^{4} \lambda_i\right)\left(\sum_{i=1}^{4} \lambda_i x_i y_i\right) - \left(\sum_{i=1}^{4} \lambda_i x_i\right)\left(\sum_{i=1}^{4} \lambda_i y_i\right)}{\left(\sum_{i=1}^{4} \lambda_i\right)\left(\sum_{i=1}^{4} \lambda_i x_i^2\right) - \left(\sum_{i=1}^{4} \lambda_i x_i\right)^2} = 1{,}196,$$

$$\widehat{\Theta}_0 = \frac{\sum_{i=1}^{4} \lambda_i y_i}{\sum_{i=1}^{4} \lambda_i} - \widehat{\Theta}_1 \frac{\sum_{i=1}^{4} \lambda_i x_i}{\sum_{i=1}^{4} \lambda_i} = 4{,}616,$$

Что совпадает с решением, полученным с использованием матричных преобразований.

б) Выполним нормирование переменных:

$$X = \begin{pmatrix} \frac{1}{\sqrt{0,22}} & \frac{8}{\sqrt{0,22}} \\ \frac{1}{\sqrt{0,66}} & \frac{11}{\sqrt{0,66}} \\ \frac{1}{\sqrt{0,79}} & \frac{12}{\sqrt{0,79}} \\ \frac{1}{\sqrt{0,3}} & \frac{9}{\sqrt{0,3}} \end{pmatrix}, \quad Y = \left(\frac{5,11}{\sqrt{0,22}} \quad \frac{8,81}{\sqrt{0,66}} \quad \frac{9,71}{\sqrt{0,79}} \quad \frac{5,82}{\sqrt{0,3}}\right)^T$$

Используя МНК, находим оценки неизвестных параметров:

$$\widehat{\Theta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} -4,616 \\ 1,196 \end{pmatrix}.$$

Задача 3. Используя данные табл.5.4 выполнить тест Парка.

Таблица 5.4 –Исходные данные

Цена, X	15,91	15,54	16,76	15,21	15,28	15,92	15,95	16,69
Спрос, Y	117,088	119,864	110,023	123,809	121,175	116,17	118,344	110,106
Цена, X	15,09	15,62	16,31	16,33	16,60	15,49	15,70	
Спрос, Y	125,178	118,068	116,201	111,457	115,103	116,914	123,589	

Решение.

1. Рассчитаем оценки параметров уравнения регрессии $y_i = \Theta_0 + \Theta_1 x + \varepsilon_i$. Вычисление можно сделать либо с помощью метода наименьших квадратов, либо с помощью стандартных функций (в Excel – ЛИНЕЙН, ОТРЕЗОК, НАКЛОН).

Получим, что значения оценок параметров: Θ_0=239,96, Θ_1=-7,7.

2. Вычислим значения $\varepsilon_i = y_i - \widehat{y}_i$, $\ln \varepsilon_i^2$, $\ln x_i$ (табл. 5.5).

Табл.5.5- Данные для теста Парка

Цена X(р.)	Спрос Y (тыс. шт.)	Остатки ε	$\ln \varepsilon^2$	$\ln X$
15,91р.	117,088	-0,322884172	-2,260923238	2,766947842
15,54р.	119,864	-0,396975543	-1,847761208	2,743417345
16,76р.	110,023	-0,840376969	-0,347809429	2,818995095
15,21р.	123,809	1,006051072	0,012065676	2,721953106
15,28р.	121,175	-1,088742452	0,170046634	2,726544784
15,92р.	116,17	-1,163854676	0,303474985	2,76757618
15,95р.	118,344	1,241233814	0,432211792	2,769458829
16,69р.	110,106	-1,296583445	0,519465371	2,814809738
15,09р.	125,178	1,450697114	0,744088418	2,714032273
15,62р.	118,068	-1,576739571	0,910718305	2,748552144
16,31р.	116,201	1,871295688	1,253262144	2,791778417
16,33р.	111,457	-2,718645319	2,000267424	2,793003907
16,60р.	115,103	3,007151087	2,201986295	2,809402695
15,49р.	116,914	-3,732123026	2,633954495	2,740194654
15,70р.	123,589	4,560496401	3,034862955	2,753660712

3. Оценим вспомогательное уравнение регрессии

$$\ln \varepsilon_i^2 = a + b \cdot \ln x_i + \nu_i,$$

где x_i - i - е значение фактора;

v_i - случайный остаток.

Аналогично п.1. находим оценки a и b. , используя в качестве исходных данных 4 и 5 столбцы таблицы 4.4. Получим: a=-11,126, b=4,259.

4. Проверим значимость коэффициента b.

Поставляя значения последнего столбца в полученное уравнение, определим оценки результирующей переменной и разность реальных и модельных значений (табл. 5.6).

Табл.5.6 Определение разности реальных и модельных данных

$z \quad (z = \ln \varepsilon^2)$	$\ln X$	$\hat{z}$	$(z - \hat{z})^2$
-2,260923238	2,766947842	0,657443518	8,516864525
-1,847761208	2,743417345	0,55723204	5,783992525
-0,347809429	2,818995095	0,879101905	1,505311423
0,012065676	2,721953106	0,465820327	0,205893284
0,170046634	2,726544784	0,485375323	0,099432183
0,303474985	2,76757618	0,660119479	0,127195295

0,432211792	2,769458829	0,668137287	0,055660839
0,519465371	2,814809738	0,861277342	0,116835423
0,744088418	2,714032273	0,43208715	0,097344791
0,910718305	2,748552144	0,57910008	0,109970647
1,253262144	2,791778417	0,763191752	0,240168989
2,000267424	2,793003907	0,768410859	1,517470595
2,201986295	2,809402695	0,838249877	1,859777019
2,633954495	2,740194654	0,543507274	4,369969585
3,034862955	2,753660712	0,600856404	5,924387888
		Сумма	30,53

Определим t -критерий Стьюдента:

$$S^2_{\hat{\imath}\,\tilde{n}\grave{o}} = \frac{\sum_{i=1}^{n}(z-\hat{z})^2}{n-m-1} = \frac{30,53}{13} = 1,5325,$$

$$\sigma_x^2 = \overline{x^2} - \overline{x}^2 = 0,00116\,,$$

$$m_b = \frac{\sqrt{S^2_{\hat{\imath}\,\tilde{n}\grave{o}}}}{\sigma_x \cdot \sqrt{n}} = \frac{\sqrt{1,5325}}{\sqrt{0,00116}\sqrt{15}} = 11,59,$$

$$t_b = \frac{b}{m_b} = \frac{4,259}{11,59} = 0,37$$

5. Условие принятия гипотезы*:* $t_b > t_{\alpha,n-2}$.

Рассчитаем t - статистику:

Т.к. условие $t_b > t_{0,95;13}$ ($t_{0,95;13} = 2,16$) не выполняется, то нулевая гипотеза об отсутствии гетероскедастичности принимается при уровне значимости $\alpha = 0,95$.

Задача 4. Используя данные задачи 3 выполнить тест Голфелда-Квандта.

Решение. Упорядочим исходные данные по мере возрастания значений x и разделим исходную модель на три равных части (табл.5.7).

Табл.5.7 – Разбиение таблицы на 3 части

Цена X (р.)	Спрос Y (тыс. шт.)	
15,09р.	125,178	1
15,21р.	123,809	1
15,28р.	121,175	1
15,49р.	116,914	1
15,54р.	119,864	1
15,62р.	118,068	2
15,70р.	123,589	2
15,91р.	117,088	2
15,92р.	116,17	2
15,95р.	118,344	2
16,31р.	116,201	3
16,33р.	111,457	3
16,60р.	115,103	3
16,69р.	110,106	3
16,76р.	110,023	3

Построим регрессионную модель для первых пяти и последних пяти наблюдений. Используя метод наименьших квадратов, определим неизвестные оценки параметров, так что уравнения регрессии будут иметь вид:

$\hat{y}1 = 245,6 - 8,043x$,

$\hat{y}3 = 2,727 + 22,303x$.

Определим сумму квадратов отклонений в каждой регрессии (промежуточные данные представлены в табл.5.8).

Табл.5.8 – Данные для выполнения теста

Цена X (р.)	Спрос Y (тыс. шт.)	$\hat{Y}$	$(Y-\hat{Y})^2$
15,09	125,178	124,976	0,040692434
15,21	123,809	123,120	0,474347171
15,28	121,175	122,038	0,744081782
15,49	116,914	118,790	3,51784908
15,54	119,864	118,016	3,414152066
		Сумма	8,19
16,31	116,201	114,412	3,200906604
16,33	111,457	114,251	7,806572866

16,6	115,103	112,079	9,142701174
16,69	110,106	111,355	1,5610133
16,76	110,023	110,792	0,591927257
		Сумма	22,303

Вычислим статистику:

$$F = \frac{S^{(3)}}{S^{(1)}} = \frac{22,303}{8,19} = 2,72$$

Т.к. условие $F > F_{0,95}(3,3)$ не выполняется, то гипотеза об отсутствии гетероскедастичности принимается.

[Заголовок 1] 6. Автокорреляция [.]

Задача 1. Анализируется объем S сбережений некого домохозяйства за 10 лет. Предполагается, что его размер St в текущем году t зависит от величины y_{t-1} располагаемого дохода *Y* в предыдущем году и от величины r_t реальной процентной ставки *R* в рассматриваемом году. Статистические данные представлены в табл. 6.1.

Табл.6.1 Исходные данные

Год	80	81	82	83	84	85	86	87	88	89	90
Y(тыс.у.е.)	100	110	140	150	160	160	180	200	230	250	260
Z (%)	2	2	3	2	3	4	4	3	4	5	5
S (тыс.у.е.)	20	25	30	30	35	38	40	38	44	50	55

Необходимо:

а) по МНК оценить коэффициенты линейной регрессии;

б) вычислить коэффициент детерминации R;

в) определить, какой процент разброса зависимой переменной объясняется данной регрессией;

г) сравнить коэффициент детерминации R со скорректированным коэффициентом детерминации R;

д) вычислить статистику DW Дарбина-Уотсона и оценить наличие автокорреляции;

е) сделать выводы по качеству построенной модели;

ж) спрогнозируйте средний объем сбережений в 1991 г., если предполагаемый доход составит 270 тыс. у. е., а процентная ставка будет равна 5.5.

Решение. Оценки неизвестных параметров определяем с помощью метода наименьших квадратов:

$$\widehat{\Theta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 2,96 \\ 0,12 \\ 3,55 \end{pmatrix}.$$

Таким образом, уравнение регрессии будет иметь вид:

$s_t = 2,96 + 0,12 y_t + 3,55 z_t$.

Найденное уравнение позволяет рассчитать модельные значения s_t зависимой переменной S и вычислить отклонения ε_i реальных значений от модельных (табл.6.2):

Табл.6.2 Промежуточные вычисления

Год	S	$\widehat{S}$	ε_i	ε_i^2	$\varepsilon_i - \varepsilon_{i-1}$	$(\varepsilon_i - \varepsilon_{i-1})^2$
80	20	22,49	-2,49	6,19		
81	25	23,73	1,27	1,61	3,76	14,12
82	30	31,01	-1,01	1,02	-2,28	5,20
83	30	28,70	1,30	1,70	2,31	5,35
84	35	33,49	1,51	2,27	0,20	0,04
85	38	37,05	0,95	0,91	-0,55	0,31
86	40	39,53	0,47	0,22	-0,48	0,23
87	38	38,46	-0,46	0,21	-0,93	0,86
88	44	45,74	-1,74	3,03	-1,28	1,64
89	50	51,78	-1,78	3,16	-0,04	0,00
90	55	53,02	1,98	3,92	3,76	14,12
Сумма	405	405	0	24,24		41,87

Коэффициент детерминации рассчитаем по формуле:

$R^2 = 1 - \frac{24,2408}{1087,636} = 0,9777$, где число 1087,636 было получено по формуле $\sum (s - \overline{s})^2$ (величина $\overline{s} = \frac{\sum s}{11}$).

Следовательно, уравнение регрессии объясняет 97,77% разброса зависимой переменной S.

Скорректированный коэффициент детерминации рассчитывается по формуле:

$$\widehat{R}^2 = 1 - (1 - 0,9777)\frac{11 - 1}{8} = 0,9721$$

Как и следовало ожидать, он меньше обычного коэффициента детерминации.

Вычислим статистику Дарбина-Уотсона:

$$d = \frac{\sum_{i=2}^{n} (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n} \varepsilon_i^2} = \frac{41,87375}{24,24058} = 1,72742$$

Для проверки статистической значимости d воспользуемся таблицей критических точек (приложение 3), при уровне значимости α=0,05 и числе наблюдений n=11 имеем: $d_{í} = 0,658$, $d_{â} = 1,604$.

Т.к. $1,604 < d < 2,396$ $(d_{â} < d < 4 - d_{â})$, то гипотеза об отсутствии автокорреляции не отклоняется, т.е. есть основания считать, что автокорреляция остатков отсутствует. Это является одним из подтверждений высокого качества модели.

Средний объем сбережений в 1991 г., если предполагаемый доход составит 270 тыс.у.е., а процентная ставка равна 5,5:

$$2,96 + 0,12 \cdot 270 + 3,55 \cdot 5,5 = 56,039.$$

Задача 2. В таблице 6.3 представлены значения отклонений функции регрессии. С помощью метода рядов выполнить тест на автокорреляцию:

Табл.6.3 Исходные данные

Номер наблюдения	1	2	3	4	5	6	7	8	9	10
Отклонение ε	-0,01	-0,03	0,14	0,08	0,07	-0,09	-0,12	-0,11	0,05	0,07
Номер наблюдения	11	12	13	14	15	16	17	18	19	20
Отклонение ε	-0,23	0,08	0,02	0,07	-0,26	-0,11	0,06	-0,21	-0,32	-0,19

Решение. Согласно таблице n_1 = 9 (число знаков «+»), $n_2 = 11$ (число знаков «-»), количество рядов k =9. Тогда при α = 0.05 нижняя граница k_1 = 6, верхняя граница k_2 = 16 (Приложение 2). Т.к. $6 < k < 16$, то гипотеза об отсутствии автокорреляции не отклоняется.

[Заголовок 1] 7. Авторегрессионная модель [.]

Задача 1. В табл.7.1 представлены сведения о курсе акции. С помощью авторегресионной модели выполнить прогноз курса акции на 5 месяцев.

Табл.7.1 – Данные о курсе акции

Месяц, i	1	2	3	4	5	6	7
Курс акции, y_i	133,77	132,88	132,2	134,66	134,29	131,05	127,57
Месяц, i	8	9	10	11	12	13	14
Курс акции, y_i	129,36	129,83	133,94	132,66	131,75	130,44	126,5
Месяц, i	15	16	17	18	19	20	

Курс акции, y_i	125,91	128,2	128,42	128,19	127,15	125,69	

Решение. С помощью метода наименьших квадратов находим оценки неизвестных параметров Θ_0, Θ_1 уравнения $y_i = \Theta_0 + \Theta_1 y_{i-1} + \varepsilon_i$. Для расчета будем использовать данные, начиная со второго месяца, т.к. для первого месяца не дано предыдущее значение. В качестве значений x_i используются y_{i-1}. Получим: Θ_0=30,246, Θ_1=0,765.

С помощью уравнения регрессии вычисляем прогноз:

$y_{21} = 30,246 + 0,765 \cdot y_{20} = 30,246 + 0,765 \cdot 125,69 = 126,39$,

$y_{22} = 30,246 + 0,765 \cdot y_{21} = 30,246 + 0,765 \cdot 126,39 = 126,92$,

$y_{23} = 30,246 + 0,765 \cdot y_{22} = 30,246 + 0,765 \cdot 126,92 = 127,33$, и т.д.

Данные сведены в табл.7.2.

Табл.7.2 Построение прогноза

Месяц, i	Курс акции, y_i	Курс акции, $x_i = y_{i-1}$	Прогноз $\widehat{y}_i$
1	133,77		
2	132,88	133,77	
3	132,2	132,88	
4	134,66	132,2	
5	134,29	134,66	
6	131,05	134,29	
7	127,57	131,05	
8	129,36	127,57	
9	129,83	129,36	
10	133,94	129,83	
11	132,66	133,94	
12	131,75	132,66	

13	130,44	131,75	
14	126,5	130,44	
15	125,91	126,5	
16	128,2	125,91	
17	128,42	128,2	
18	128,19	128,42	
19	127,15	128,19	
20	125,69	127,15	
21		125,69	126,3865
22		126,3865	126,9193
23		126,9193	127,3268
24		127,3268	127,6385
25		127,6385	127,8769

[Заголовок 1] 8. Фиктивные переменные [.]

Задача 1. Необходимо исследовать зависимость между результатами письменных вступительных и курсовых (на I курсе) экзаменов по математике. Получены следующие данные о числе решенных задач на вступительных экзаменах X (задание -10 задач) и курсовых экзаменах Y (задание—7 задач) 12студентов, а также распределение этих студентов по фактору «пол» (табл.8.1).

Табл.8.1 – Исходные данные

№ студента	Число решенных задач		Пол студента	№ студента	Число решенных задач		Пол студента
i	x_i	y_i		i	x_i	y_i	
1	10	6	муж.	7	6	3	жен.
2	6	4	жен.	8	7	4	муж.

3	8	4	муж.	9	9	7	муж.
4	8	5	жен.	10	6	3	жен.
5	6	4	жен.	11	5	2	муж.
6	7	7	муж.	12	7	3	жен.

Построить линейную регрессионную модель Y по X с использованием фиктивной переменной по фактору «пол». Можно ли считать, эта модель одна и та же для юношей и девушек?

Решение. Вначале рассчитаем уравнение парной регрессии Y по X:

$$\widehat{y} = -1,437 + 0,815x .$$

Коэффициент детерминации R^2=0,530, т.е. 53,0% вариации зависимой переменной Y обусловлено регрессией.

Однако полученное уравнение не учитывает влияние качественного признака—фактора «пол».

Для ее учета введем в регрессионную модель фиктивную (бинарную) переменную $z^{(1)}$,

где $z_i^{(1)} = \begin{cases} 1, & \text{åñëè } i-\text{é ñòóäåíò ìóæñêîãî ïîëà}, \\ 0, & \text{åñëè } i-\text{é ñòóäåíò æåíñêîãî ïîëà}. \end{cases}$

Матрица плана будет иметь вид:

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 10 & 6 & 8 & 8 & 6 & 7 & 6 & 7 & 9 & 6 & 5 & 7 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}^T .$$

Найдем вектор оценок параметров регрессии:

$$\widehat{\Theta} = (-1,165 \quad 0,743 \quad 0,466)^T .$$

Выборочное уравнение множественной регрессии примет вид:

$$\widehat{y} = \underset{(2,410)}{-1,165} + \underset{(0,053)}{0,743}x + \underset{(0,405)}{0,466}z^{(1)} .$$

Коэффициент детерминации: $R^2 = 0,449$.

Из полученного уравнения следует, что при том же числе решенных задач на вступительных экзаменах *X*, на курсовых экзаменах юноши решают в среднем на 0,46 задачи больше.

Однако коэффициент регрессии α_1 при фиктивной переменной $z^{(1)}$ незначим по t-критерию (т.к. $t = \dfrac{0,466}{0,405} = 1,15$ и $t < t_{0,95;9} = 2,26$), возможно, из-за недостаточного объема выборки либо в силу того, что гипотеза H_0: α_1=0 верна.

В этом также легко убедиться, если учесть, что скорректированный коэффициент детерминации уменьшился от значения R^2=0,483 для парной модели до значения R^2=0,449 для множественной модели. Следовательно, по имеющимся данным влияние фактора «пол» оказалось несущественным, и у нас есть основания считать, что регрессионная модель результатов курсовых экзаменов по математике в зависимости от вступительных одна и та же для юношей и девушек.

Замечание. Если бы в регрессионной модели мы хотели учесть другие факторы с большим, чем две, числом k_i градаций, то, как отмечено выше, следовало бы ввести в модель (k_i-1) бинарных переменных. Например, если было бы необходимо изучить влияние на результаты курсового экзамена фактора «тип учебного заведения», оконченного студентом (школа, техникум, ПТУ), то в регрессионную модель следовало ввести k_i-1=3-1=2 бинарные переменные $z^{(21)}$ и $z^{(22)}$:

$$y_i = \Theta_0 + \Theta_1 x_i^{(1)} + \alpha_1 z_i^{(1)} + \alpha_{21} z_i^{(21)} + \alpha_{22} z_i^{(22)} + \varepsilon_i,$$

где $z_i^{(21)} = \begin{cases} 1, & \text{åñëè ñòóäåíò îêîí÷èë øêîëó;} \\ 0, & \text{â îñòàëüíûõ ñëó÷àÿõ;} \end{cases}$

$$z_i^{(22)} = \begin{cases} 1, & \text{åñëè ñòóäåíò îêîí÷èë òåõíèêóì;} \\ 0, & \text{â îñòàëüíûõ ñëó÷àÿõ.} \end{cases}$$

Но при этом, конечно, следовало увеличить объем выборки n, так как надежность статистических выводов существенно зависит от отношения объема выборки n к общему числу всех параметров регрессионной модели: чем больше величина отношения $n/(p+1)$,тем точнее соответствующие оценки, тем надежнее статистические выводы.

[Заголовок 1] 9. Тест Чоу [.]

Задача 1. По данным табл.9.1, используя критерий Г.Чоу, выяснить, можно ли считать одной и той же линейную регрессию Y по X для юношей и девушек.

Табл.9.1 – Исходные данные

№ студента	Число решенных задач		Пол студента	№ студента	Число решенных задач		Пол студента
i	x_i	y_i		i	x_i	y_i	
1	10	6	муж.	7	6	3	жен.
2	6	4	жен.	8	7	4	муж.
3	8	4	муж.	9	9	7	муж.
4	8	5	жен.	10	6	3	жен.
5	6	4	жен.	11	5	2	муж.
6	7	7	муж.	12	7	3	жен.

Решение. По n_1=6 парам наблюдений (x_i, y_i) для юношей— (10;6), (8;4), (7;7), (7;4), (9;7), (5;2) (1-ая выборка) и по n_2= 6 парам наблюдений для девушек—(6;4), (8;5), (6;4), (6;3), (6;3), (7;3) (2-ая выборка) рассчитаем уравнения регрессии:

$$\hat{y} = -1{,}000 + 0{,}783x \quad (\textit{äëÿ} \;\; 1-\textit{é} \;\; \textit{âûáîðêè});$$

$$\hat{y} = -0{,}048 + 0{,}571x \quad (\textit{äëÿ} \;\; 2-\textit{é} \;\; \textit{âûáîðêè}).$$

По всем $n = n_1 + n_2 = 12$ парам наблюдений рассчитаем уравнение регрессии для объединенной выборки:

$\hat{y} = -1,437 + 0,815x$.

Так как вычисленное по (7.4) значение

$F = 0,21 < F_{0,05;2;11} = 4,46$,

то влияние фактора «пол» несущественно, и в качестве оценки регрессионной модели Y по X можно рассматривать уравнение регрессии, полученное по объединенной выборке.

Задача 2. В таблице 9.2 приведены данные о зависимости потребления шоколада в неделю (y) от дохода (x). Нужно определить, можно ли объединить выборки для мужчин и женщин.

Табл.9.2 Исходные данные

Пол	ж	ж	ж	ж	ж	ж	м	м	м	м	м	м
y	6	4	7	4	7	2	4	5	4	3	3	3
x	10	8	7	7	9	5	6	8	6	6	6	7

Решение. С помощью метода наименьших квадратов находим оценки параметров и функции регрессии для трех выборок

1) женщин:

$\hat{y}_{\text{æ åí}} = -1 + 0,783x$

2) мужчин:

$\hat{y}_{\text{æ åí}} = 0,571 - 0,048x$

3) объединенной:

$\hat{y}_{\text{æ åí}} = 0,815 - 1,436x$.

Найдем значения квадратов случайных остатков $\hat{\varepsilon} = y - \hat{y}_{\text{æ åí}}$.

1) функции регрессии выборки женщин (табл.9.3)

Табл.9.3 Расчет значений квадратов случайных остатков

x	10	8	7	7	9	5	
y	6	4	7	4	7	2	
$\hat{y}_{\text{æ åí}}$	6,826	5,261	4,478	4,478	6,043	2,913	
$\hat{\varepsilon}$	-0,826	-1,261	2,522	-0,478	0,956	-0,913	
$\hat{\varepsilon}^2$	0,682	1,590	6,360	0,229	0,915	0,834	$\sum =$ 10,609

2) функции регрессии выборки мужчин (табл.9.4)

Табл.9.4 Расчет значений квадратов случайных остатков

x	6	8	6	6	6	7	
y	4	5	4	3	3	3	
$\hat{y}_{\text{ì óæ}}$	3,381	4,524	3,381	3,381	3,381	3,952	
$\hat{\varepsilon}$	0,619	0,476	0,619	-0,381	-0,381	-0,952	
$\hat{\varepsilon}^2$	0,383	0,227	0,383	0,145	0,145	0,907	$\sum =$ 2,190

3) функции регрессии общей выборки (табл. 9.5)

Табл.9.5 Расчет значений квадратов случайных остатков

x	10	8	7	7	9	5	6	8	6	6	6	7	
y	6	4	7	4	7	2	4	5	4	3	3	3	
$\hat{y}_{\text{î áù}}$	6,709	5,080	4,265	4,265	5,895	2,636	3,451	5,080	3,451	3,451	3,451	4,265	
$\hat{\varepsilon}$	-0,709	-1,080	2,735	-0,265	1,105	-0,636	0,549	-0,080	0,549	-0,451	-0,451	-1,265	
$\hat{\varepsilon}^2$	0,503	1,166	7,478	0,070	1,222	0,405	0,302	0,006	0,302	0,203	0,203	1,601	$\sum =$ 13,462

Вычислим F - статистику:

$$F=\frac{\left(\sum_{i=1}^{n}\varepsilon_i^2-\sum_{i=1}^{n_1}\varepsilon_i^2-\sum_{i=n_1}^{n}\varepsilon_i^2\right)(n-2p-2)}{\left(\sum_{i=1}^{n_1}\varepsilon_i^2+\sum_{i=n_1}^{n}\varepsilon_i^2\right)(p+1)}=$$

$$\frac{(13{,}462-10{,}609-2{,}190)(12-2\cdot 1-2)}{(10{,}609+2{,}190)(1+1)}=0{,}207$$

Т.к. условие $F>F_{\alpha;p+1;n-2;p-2}(=4{,}46)$, не выполняется, то две регрессионные модели можно объединить в одну, следовательно, выборки мужчин и женщин можно рассматривать как одну выборку.

[Заголовок 1] 10. Мультиколлинеарность [.]

Задача 1. В таблице 10.1 представлены данные о квартирах на рынке недвижимости. Найти оценки неизвестных параметров. Определить индекс детерминации. Выполнить тест на мультиколлинеарность.

Табл.10.1 Исходные данные

Адрес	Цена, тыс.руб.	Кол-во комнат	Общая площадь	Отделка (0-черновая, 1- есть)	Год постройки
Интернационалистов ул, 29	3950	4	80	1	1976
Урожайный пер, 30	2050	1	35	1	2009
Ленина пр-кт, 126	2250	1	38	1	2004
Мира пр-кт, 33	5500	5	100	1	1988
Профсоюзная ул, 7	1270	1	26	0	2014
Интернационалистов ул, 6	1680	1	30	1	1982

Ленская ул, 47	2690	2	57	1	2008
Розы Люксембург ул, 101	2200	1	38	1	2008
Профсоюзная ул, 7	1290	1	31	0	2014
Дербышевский пер, 26а	2800	2	50	1	1993
Профсоюзная ул, 7	1520	1	35	0	2014
Карла Ильмера ул, 21	2500	2	54	1	1983
79-й Гвардейской Дивизии, 20	2650	2	54	1	1978
Смирнова, 23	1900	2	44	1	1962

Решение. Запишем исходные данные в матричном виде:

$$Y := (3950\ \ 2050\ \ 2250\ \ 5500\ \ 1270\ \ 1680\ \ 2690\ \ 2200\ \ 1290\ \ 2800\ \ 1520\ \ 2500\ \ 2650\ \ 1900)^T$$

$$X := \begin{pmatrix} 1 & 4 & 80 & 1 & 1976 \\ 1 & 1 & 35 & 1 & 2009 \\ 1 & 1 & 38 & 1 & 2004 \\ 1 & 5 & 100 & 1 & 1988 \\ 1 & 1 & 26 & 0 & 2014 \\ 1 & 1 & 30 & 1 & 1982 \\ 1 & 2 & 57 & 1 & 2008 \\ 1 & 1 & 38 & 1 & 2008 \\ 1 & 1 & 31 & 0 & 2014 \\ 1 & 2 & 50 & 1 & 1993 \\ 1 & 1 & 35 & 0 & 2014 \\ 1 & 2 & 54 & 1 & 1983 \\ 1 & 2 & 54 & 1 & 1978 \\ 1 & 2 & 44 & 1 & 1962 \end{pmatrix}$$

С помощью метода наименьших квадратов находим оценки:

$$Q := \left(X^T \cdot X\right)^{-1} X^T \cdot Y$$

$$Q = \begin{pmatrix} -3.45 \times 10^4 \\ 597.58 \\ 17.658 \\ 753.483 \\ 17.241 \end{pmatrix}$$

Вычислим оценки *y*:

$$j := 0, 1 .. 13$$

$$ym_j := Q_0 + Q_1 \cdot X_{j,1} + Q_2 \cdot X_{j,2} + Q_3 \cdot X_{j,3} + Q_4 \cdot X_{j,4}$$

ym =

	0
0	$4.122 \cdot 10^3$
1	$2.104 \cdot 10^3$
2	$2.071 \cdot 10^3$
3	$5.28 \cdot 10^3$
4	$1.278 \cdot 10^3$
5	$1.55 \cdot 10^3$
6	$3.073 \cdot 10^3$
7	$2.14 \cdot 10^3$
8	$1.366 \cdot 10^3$
9	$2.69 \cdot 10^3$
10	$1.437 \cdot 10^3$
11	$2.589 \cdot 10^3$
12	$2.502 \cdot 10^3$
13	$2.05 \cdot 10^3$

По формуле вычислим индекс детерминации

$$R^2 = 1 - \frac{S_{\text{î ñò}}}{S_{\text{î áù}}} = \frac{S_{\text{î ò êë}}}{S_{\text{î áù}}} = \frac{\sum(\hat{y} - \overline{y})^2}{\sum(y - \overline{y})^2}$$

$$ys := \frac{\left(\sum_{i=0}^{13} Y_i\right)}{14}$$

$$ys = 2.446 \times 10^{3}$$

$$Sot := \sum_{i=0}^{13} \left(ym_i - ys\right)^2$$

$$Sob := \sum_{i=0}^{13} \left(Y_i - ys\right)^2$$

$$R := \frac{Sot}{Sob}$$

$$R = 0.978$$

Для выявления мультиколлинеарности вычислим *vif*.

Построим уравнение регрессии

$$x_i^{(1)} = \alpha_0 + \alpha_2 x_i^{(2)} + \alpha_3 x_i^{(3)} + \alpha_4 x_i^{(4)}$$

Вычислим оценки неизвестных параметров:

$$Y1 := (4\ \ 1\ \ 1\ \ 5\ \ 1\ \ 1\ \ 2\ \ 1\ \ 1\ \ 2\ \ 1\ \ 2\ \ 2\ \ 2)^{T}$$

$$X1 := \begin{pmatrix} 1 & 80 & 1 & 1976 \\ 1 & 35 & 1 & 2009 \\ 1 & 38 & 1 & 2004 \\ 1 & 100 & 1 & 1988 \\ 1 & 26 & 0 & 2014 \\ 1 & 30 & 1 & 1982 \\ 1 & 57 & 1 & 2008 \\ 1 & 38 & 1 & 2008 \\ 1 & 31 & 0 & 2014 \\ 1 & 50 & 1 & 1993 \\ 1 & 35 & 0 & 2014 \\ 1 & 54 & 1 & 1983 \\ 1 & 54 & 1 & 1978 \\ 1 & 44 & 1 & 1962 \end{pmatrix}$$

$$Q1 := \left(X1^{T} \cdot X1\right)^{-1} X1^{T} \cdot Y1$$

$$Q1 = \begin{pmatrix} 23.44 \\ 0.058 \\ -0.486 \\ -0.012 \end{pmatrix}$$

Вычислим вектор оценок результирующей переменной:

$j := 0, 1 .. 13$

$ym1_j := Q1_0 + Q1_1 \cdot X1_{j,1} + Q1_2 \cdot X1_{j,2} + Q1_3 \cdot X1_{j,3}$

ym1 =

	0
0	3.854
1	0.827
2	1.063
3	4.879
4	0.727
5	0.86
6	2.125
7	1.015
8	1.019
9	1.897
10	1.253
11	2.251
12	2.311
13	1.919

Наконец, определим *vif*:

$$ys1 := \frac{\left(\sum_{i=0}^{13} Y1_i\right)}{14}$$

$ys1 = 1.857$

$$Sot1 := \sum_{i=0}^{13} \left(ym1_i - ys1\right)^2$$

$$Sob1 := \sum_{i=0}^{13} \left(Y1_i - ys1\right)^2$$

$$R1 := \frac{Sot1}{Sob1}$$

$R1 = 0.979$

$$vif1 := \frac{1}{(1 - R1)}$$

$vif1 = 46.861$

Т.к. полученное значение больше 10, то в модели наблюдается мультиколлинеарность.

[Заголовок 1] 11. Процедура отбора существенных факторов [.]

Задача 1. По данным таблицы 11.1 выполнить отбор наиболее существенных факторов с помощью процедуры пошагового отбора.

Табл.11.1 Исходные данные

№ заемщика	Просрочки кредита (шт.), (Y)	Стаж, лет (X1)	Срок кредита, мес. (X2)	Сумма кредита, руб. (X3)
1	0	7,5	12	170 000
2	0	4,5	12	120 000
3	0	6,5	12	85 000
4	1	2,5	12	160 000
5	1	3,5	24	105 000
6	0	6,5	12	90 000
7	3	2	24	80 000
8	2	3,5	24	395 000
9	2	6	36	150 000
10	4	2	60	70 000

Решение.

Шаг 1. Вычислим квадрат коэффициента корреляции зависимой переменной с каждой объясняющей переменной:

$$r=\Theta_1\frac{\sigma_x}{\sigma_y}=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{\left(n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2\right)}},$$

$r_{yx_1}^2=0,5204$,

$r_{yx_2}^2=0,7582$,

$r_{yx_3}^2=0,0003$.

Из объясняющих переменных $x^{(1)}-x^{(3)}$ выделяется переменная $x^{(2)}$, имеющая с зависимой переменной Y наибольший коэффициент детерминации R^2 (равный для парной модели квадрату коэффициента корреляции r). Очевидно, это переменная $x^{(2)}$, так как коэффициент детерминации $R_{yx_2}^2=r_{yx_2}^2=0,7582$-максимальный. С учетом поправки на несмещенность скорректированный коэффициент детерминации $\hat{R}_{yx_2}^2=1-\frac{n-1}{n-m}\cdot(1-R^2)=1-\frac{10-1}{10-2}(1-0,7582)=0,728$.

Шаг 2. Среди возможных пар объясняющих переменных $x^{(2)}$, $x^{(j)}$, $j=1,3$, выбирается пара, имеющая с зависимой переменной Y наиболее высокий коэффициент детерминации.

Рассмотрим пару ($x^{(2)},x^{(1)}$)

Уравнение регрессии: $\hat{y}=1,1124-0,2792x^{(1)}+0,0627x^{(2)}$. Вычисление данных для расчета индекса детерминации приведено в табл. 11.2.

Табл.11.2 Вычисление данных для расчета индекса детерминации

Просрочки кредита (шт.), (Y)	Стаж, лет (X1)	Срок кредита, мес. (X2)	$\hat{Y}$	$(Y-\hat{Y})^2$	$(Y-\overline{Y})^2$
0	7,5	12	-0,2290	0,0524	1,6900
0	4,5	12	0,6086	0,3704	1,6900

0	6,5	12	0,0502	0,0025	1,6900
1	2,5	12	1,1670	0,0279	0,0900
1	3,5	24	1,6405	0,4103	0,0900
0	6,5	12	0,0502	0,0025	1,6900
3	2	24	2,0593	0,8849	2,8900
2	3,5	24	1,6405	0,1292	0,4900
2	6	36	1,6952	0,0929	0,4900
4	2	60	4,3173	0,1007	7,2900
$\sum$				2,07381	18,1

$$R^2_{yx_1x_2} = 1 - \frac{S_{\text{î ñò}}}{S_{\text{î áù}}} = 1 - \frac{2,07381}{18,1} = 0,8854,$$

$$\widehat{R}^2_{yx_1x_2} = 1 - \frac{10-1}{10-3}(1 - 0,8854) = 0,8527.$$

Рассмотрим пару ($x^{(2)}, x^{(3)}$)

Уравнение регрессии: $\widehat{y} = -0,7968 + 0,08116x^{(2)} + 0x^{(3)}$. Вычисление данных для расчета индекса детерминации приведено в табл. 11.3.

Табл.11.3 Вычисление данных для расчета индекса детерминации

Просрочки кредита (шт.), (Y)	Срок кредита, мес. (X2)	Сумма кредита, руб. (X3)	$\widehat{Y}$	$(Y-\widehat{Y})^2$	$(Y-\overline{Y})^2$
0	12	170 000	0,4710	0,2218	1,6900
0	12	120 000	0,3845	0,1479	1,6900
0	12	85 000	0,3240	0,1050	1,6900
1	12	160 000	0,4537	0,2984	0,0900
1	24	105 000	1,3326	0,1106	0,0900
0	12	90 000	0,3327	0,1107	1,6900
3	24	80 000	1,2893	2,9264	2,8900
2	24	395 000	1,8340	0,0276	0,4900
2	36	150 000	2,3843	0,1477	0,4900
4	60	70 000	4,1939	0,0376	7,2900
$\sum$				4,1336	18,1000

$$R^2_{yx_2x_3} = 1 - \frac{S_{\text{î ñò}}}{S_{\text{î áù}}} = 1 - \frac{4,1336}{18,1} = 0,7716,$$

$$\widehat{R}^2_{yx_2x_3} = 1 - \frac{10-1}{10-3}(1-0,7716) = 0,7064.$$

Следовательно, выбираем пару ($x^{(2)}, x^{(1)}$).

Шаг 3. Рассмотрим тройку объясняющих переменных ($x^{(1)}, x^{(2)}, x^{(3)}$).

Уравнение регрессии: $\widehat{y} = 0,875 - 0,2735x^{(1)} + 0,064x^{(2)} + 0x^{(3)}$. Вычисление данных для расчета индекса детерминации приведено в табл. 11.4.

Табл.11.4 Вычисление данных для расчета индекса детерминации

Просрочки кредита (шт.), (Y)	Стаж, лет (X1)	Срок кредита, мес. (X2)	Сумма кредита, руб. (X3)	$\widehat{Y}$	$(Y-\widehat{Y})^2$	$(Y-\overline{Y})^2$
0	7,5	12	170 000	-0,1895	0,0359	1,6900
0	4,5	12	120 000	0,5665	0,3209	1,6900
0	6,5	12	85 000	-0,0258	0,0007	1,6900
1	2,5	12	160 000	1,1653	0,0273	0,0900
1	3,5	24	105 000	1,5882	0,3460	0,0900
0	6,5	12	90 000	-0,0193	0,0004	1,6900
3	2	24	80 000	1,9662	1,0688	2,8900
2	3,5	24	395 000	1,9627	0,0014	0,4900
2	6	36	150 000	1,7299	0,0729	0,4900
4	2	60	70 000	4,2558	0,0654	7,2900
$\sum$					1,9396	18,1000

$$R^2_{yx_1x_2x_3} = 1 - \frac{S_{\text{îñò}}}{S_{\text{îáù}}} = 1 - \frac{1,9396}{18,1} = 0,8928,$$

$$\widehat{R}^2_{yx_1x_2x_3} = 1 - \frac{10-1}{10-4}(1-0,8928) = 0,8393.$$

Так как скорректированный коэффициент детерминации на 3-м шаге не увеличился, то в регрессионной модели достаточно ограничиться лишь двумя отобранными ранее объясняющими переменными $x^{(1)}$ и $x^{(2)}$.

[Заголовок 1] ПРИЛОЖЕНИЕ 1. Таблица критерия Фишера для $\alpha = 0{,}05$ [.]

ν_1 / ν_2	1	2	3	4	5	6	7	8	9	10
1	162	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,26	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,69	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,38	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93
200	3,92	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

ν_1 – число степеней свободы числителя, ν_2 – число степеней свободы знаменателя.

[Заголовок 1] ПРИЛОЖЕНИЕ 2. Таблица критических значений количества рядов [.]

Нижняя граница k_1

n_1	n_2																	
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	2	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	6	7	7	8	8	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

Верхняя граница k_2

n_1	n_2																	
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4				9	9														
5			9	10	10	11	11												
6			9	10	11	12	12	13	13	13	13								
7				11	12	13	13	14	14	14	14	15	15	15					
8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
16							17	18	19	20	21	21	22	23	23	24	25	25	25
17							17	18	19	20	21	22	23	23	24	25	25	26	26
18							17	18	19	20	21	22	23	24	25	25	26	26	27
19							17	18	20	21	22	23	23	24	25	26	26	27	27
20							17	18	20	21	22	23	24	25	25	26	27	27	28

[Заголовок 1] ПРИЛОЖЕНИЕ 3. Таблица критерия Дарбина-Уотсона для α= 0,05 [.]

	m=1		m=2		m=3		m=4		m=5	
n	$d_{í}$	$d_{â}$	$d_{í}$	$d_{â}$	$d_{í}$	$d_{â}$	$d_{í}$	$d_{â}$	$d_{í}$	$d_{â}$
6	0,610	1,400	–	–	–	–	–	–	–	–
7	0,700	1,356	0,467	1,896	–	–	–	–	–	–
8	0,763	1,332	0,559	1,777	0,368	2,287	–	–	–	–
9	0,824	1,320	0,624	1,699	0,455	2,128	0,296	2,588	–	–
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,814	0,243	2,822
11	0,927	1,324	0,758	1,604	0,595	1,928	0,444	2,283	0,316	2,645
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,379	2,506
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,445	2,390
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,752	2,023
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,824	1,964
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,895	1,920
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786
50	1,503	1,585	1,452	1,628	1,421	1,674	1,378	1,721	1,335	1,771
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772
90	1,635	1,679	1,612	1,703	1,589	1,726	1,568	1,751	1,542	1,776
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780
150	1,720	1,746	1,706	1,760	1,693	1,774	1,679	1,789	1,665	1,802
200	1,758	1,778	1,748	1,789	1,736	1,799	1,728	1,810	1,718	1,820

n - число наблюдений, m -число объясняющих переменных.

n	$m=6$		$m=7$		$m=8$		$m=9$		$m=10$	
	$d_{\acute{\imath}}$	$d_{\hat{a}}$	$d_{\acute{\imath}}$	$d_{\hat{a}}$	$d_{\acute{\imath}}$	$d_{\hat{a}}$	$d_{\acute{\imath}}$	$d_{\hat{a}}$	$d_{\acute{\imath}}$	$d_{\hat{a}}$
6	–	–	–	–	–	–	–	–	–	–
7	–	–	–	–	–	–	–	–	–	–
8	–	–	–	–	–	–	–	–	–	–
9	–	–	–	–	–	–	–	–	–	–
10	–	–	–	–	–	–	–	–	–	–
11	0,203	3,005	–	–	–	–	–	–	–	–
12	0,268	2,832	0,171	3,149	–	–	–	–	–	–
13	0,328	2,692	0,230	2,985	0,147	3,266	–	–	–	–
14	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,360	–	–
15	0,447	2,472	0,343	2,727	0,251	2,979	0,175	3,216	0,111	3,438
16	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,090	0,155	3,304
17	0,554	2,318	0,451	2,537	0,356	2,757	0,272	2,975	0,198	3,184
18	0,603	2,257	0,502	2,461	0,407	2,667	0,321	2,873	0,244	3,073
19	0,649	2,206	0,549	2,396	0,456	2,589	0,369	2,783	0,290	2,974
20	0,692	2,162	0,595	2,339	0,502	2,521	0,416	2,704	0,336	2,885
21	0,732	2,124	0,637	2,290	0,547	2,460	0,461	2,633	0,380	2,806
22	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571	0,424	2,734
23	0,804	2,061	0,715	2,208	0,628	2,360	0,545	2,514	0,465	2,670
24	0,837	2,035	0,751	2,174	0,666	2,318	0,584	2,464	0,506	2,613
25	0,868	2,012	0,784	2,144	0,702	2,280	0,621	2,419	0,544	2,560
30	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251	0,712	2,363
40	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,945	2,149
50	1,219	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110	2,044
60	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939	1,222	1,984
70	1,433	1,802	1,401	1,837	1,369	1,873	1,337	1,910	1,305	1,948
80	1,480	1,801	1,428	1,831	1,425	1,861	1,397	1,893	1,369	1,925
90	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881	1,420	1,909
100	1,550	1,803	1,528	1,826	1,506	1,850	1,484	1,874	1,462	1,898
150	1,651	1,817	1,637	1,832	1,622	1,847	1,608	1,862	1,594	1,877
200	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863	1,665	1,874

Список литературы

1. Тихомиров Н.П. Эконометрика. Учебник для вузов.-М.:ЭКЗАМЕН, 2007.-510с. (11 экз.).

2. Сидоренко М.Г. Эконометрика. Учебное пособие. – Томск: ТМЦДО, 2004. – 119с. (8 экз.).

3. Кристофер Доугерти. Введение в эконометрику. Учебник для вузов. – М.: Инфра-М, 1999 – 402с. (2 экз.).

4. Грибанова Е.Б. Эконометрика. Учебное пособие. – Томск: ФДО, 2014.- 152 с.

5. Лузина Л.И. Эконометрика. Методические указания к практическим работам по дисциплине «Эконометрика».- Томск: ТУСУР, 2007.-32 с. (95 экз.)