

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ  
Государственное образовательное учреждение  
высшего профессионального образования  
«Оренбургский государственный университет»

Кафедра математических методов и моделей в экономике

А.Г. РЕННЕР, О.И. СТЕБУНОВА, Ю.А. ЖЕМЧУЖНИКОВА

# **МЕТОДЫ УСТРАНЕНИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ**

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ЛАБОРАТОРНОМУ ПРАКТИКУМУ И  
САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ

Рекомендовано к изданию Редакционно-издательским советом  
государственного образовательного учреждения высшего профессионального  
образования «Оренбургский государственный университет»

Оренбург 2005

УДК 33:517/519

ББК 65 в6

Р 39

Рецензент

кандидат экономических наук, доцент С.В. Дьяконова

Р 39      **Реннер А.Г., Стебунова О.И., Жемчужникова Ю.А.**  
**Методы устранения мультиколлинеарности [Текст]:**  
**методические указания к лабораторному практикуму и**  
**самостоятельной работе студентов/ А.Г. Реннер , О.И. Стебунова,**  
**Ю.А. Жемчужникова. – Оренбург: ГОУ ОГУ, 2005. – 25 с.**

Методические указания содержат описание работы по исследованию линейной регрессионной модели на мультиколлинеарность и варианты индивидуальных заданий для проведения лабораторной работы.

Методические указания предназначены студентам специальностей 061800, 061700, и других экономических специальностей, изучающих дисциплину «Эконометрика».

ББК 65 в6

© Реннер А.Г., 2005

© Стебунова О.И., 2005

© Жемчужникова Ю.А., 2005

© ГОУ ОГУ, 2005

## Содержание

Введение .....	4
1 Описание лабораторной работы №1 .....	5
2 Постановка задачи .....	5
3 Порядок выполнения работы. ....	5
4 Содержание письменного отчета. ....	19
5 Вопросы к защите. ....	19
Список использованных источников .....	19
Приложение А – Таблица А.1 – Выборочные данные. ....	20
Приложение А – Таблица А.2 – Варианты заданий. ....	23
Приложение А – Таблица А.3 – Наименование показателей. ....	25

## **Введение**

Поскольку несложно избежать нарушения второго из условий Гаусса-Маркова, ведущего к полной мультиколлинеарности, то ниже будем говорить о реальной (частичной) мультиколлинеарности, которая возникает в случаях существования тесных линейных статистических связей между объясняющими переменными.

Следствием реальной мультиколлинеарности является снижение точности оценок коэффициентов модели, рост дисперсий, численная неустойчивость оценок к незначительным изменениям исходных данных, а это говорит о низкой обоснованности эконометрической модели, о ее неадекватности описываемому процессу.

Цель предлагаемой работы заключается в выработке навыков выявления мультиколлинеарности и освоении приемов ее устранения.

## 1 Описание лабораторной работы №1

Лабораторная работа №1 включает следующие этапы:

- постановку задачи;
- ознакомление с порядком выполнения работы;
- выполнение расчетов индивидуальных задач на компьютере и анализ результатов;
- подготовку письменного отчета с выводами по работе;
- защиту лабораторной работы.

## 2 Постановка задачи

По данным Приложения А:

- 1) построить МНК-оценки коэффициентов линейной модели множественной регрессии и провести ее анализ;
- 2) провести анализ построенной модели на мультиколлинеарность;
- 3) устранить мультиколлинеарность одним из известных Вам методом.

## 3 Порядок выполнения работы

Рассмотрим пример построения линейной регрессионной модели на основе информации об ожидаемой продолжительности жизни мужчин, число лет ( $y$ ), рождаемости населения на 1000 человек ( $x_1$ ), смертности населения на 1000 человек ( $x_2$ ), числе браков на 1000 человек ( $x_3$ ), числе разводов на 1000 человек ( $x_4$ ), коэффициенте младенческой смертности ( $x_5$ ), соотношении денежного дохода и прожиточного минимума, % ( $x_6$ ), соотношении средней оплаты труда и прожиточного минимума трудоспособного населения, % ( $x_7$ ), численности населения с денежными доходами ниже прожиточного минимума в % от численности населения ( $x_8$ ), числа зарегистрированных преступлений на 100000 населения ( $x_9$ ).

**Запуск и подготовка данных.** Запустить ППП Statistica. После запуска на экране откроется основное окно системы Statistica, представленное на рисунке 1.

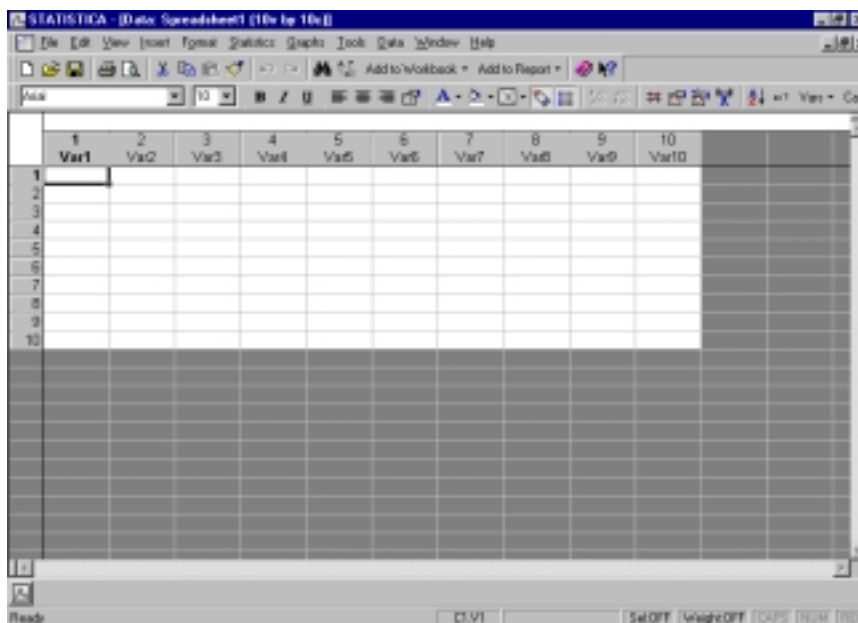


Рисунок 1 – Стартовое окно пакета Statistica

Стандартный вид исходной таблицы содержит 10 строк (10 cases) и 10 столбцов (10 variables). Так как исходная информация может быть представлена произвольного размера, то возникает необходимость в корректировке размерности таблицы. Если необходимо увеличить число столбцов, то в меню Insert, выбираем Add Variables, если необходимо изменить число строк, то – Add Cases. При этом откроется меню возможных операций со столбцами (строками).

Далее необходимо ввести данные для проведения регрессионного анализа. Если исходная информация уже имеется, то следует открыть нужный файл – для этого используется кнопка **Open Data – Открыть данные**. Окно с частью данных для анализа представлено на рисунке 2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	54,7	8,5	18,5	6,0	5,8	17,4	64,7	55	25,2	2044									
2	66,7	9,8	12,8	7,2	6,6	26,3	74,8	289	19,2	1809									
3	57,1	8,7	14,6	6,5	4,2	18,2	71,4	192	26,9	2486									
4	67,6	8,6	16,2	6,1	4,8	17,4	79,9	206	20,1	2032									
5	57,7	8,1	15,4	7,7	6,4	5,5	79,9	196	22,0	1419									
6	59,9	7,0	15,9	8,2	5,1	13,0	68,1	172	20,9	2184									
7	66,6	7,2	18,2	7,8	6,1	14,3	76,8	197	29,7	2489									
8	55,3	7,9	19,7	6,4	4,7	19,0	80,5	144	23,0	2426									
9	66,6	7,7	20,9	6,9	6,2	17,1	72,6	111	42,7	2084									
10	60,1	9,2	15,9	7,0	5,3	16,7	69,9	140	22,7	2084									
11	50,5	7,6	16,4	6,7	4,7	15,6	64,3	150	27,9	1760									
12	57,4	7,3	18,3	8,3	4,9	19,6	88,1	139	33,7	1982									
13	50,5	7,9	16,4	6,0	5,0	17,6	59,7	155	26,6	1621									
14	66,3	7,9	17,0	6,8	4,4	20,1	72,7	169	30,6	1631									
15	50,2	8,0	16,5	8,2	4,6	15,5	90,7	197	19,1	1386									
16	66,6	7,2	17,6	8,1	6,2	16,1	68,1	166	21,2	1182									
17	59,2	8,7	18,0	7,6	4,4	18,9	69,3	191	22,7	1386									
18	50,1	7,8	17,9	7,2	4,3	15,7	60,7	163	24,4	1476									
19	66,6	8,0	18,9	8,9	4,7	18,6	79,2	186	19,6	2081									
20	50,5	7,5	19,4	6,7	4,6	19,5	64,8	165	26,6	2189									
21	67,1	7,3	19,4	7,8	6,0	20,1	76,6	176	16,2	1767									
22	50,3	7,6	17,3	7,1	5,3	12,0	69,9	154	21,3	2111									
23	59,4	8,6	13,0	6,8	3,6	16,0	67,9	117	43,2	2112									
24	67,2	8,0	14,1	7,0	3,3	18,7	70,2	126	34,7	1794									
25	60,4	6,2	13,0	7,1	3,2	16,1	74,6	121	27,3	1680									
26	66,6	8,1	18,3	6,2	3,9	17,1	77,8	121	42,0	1774									
27	57,5	8,0	19,0	6,7	4,8	16,4	61,7	162	22,0	1773									
28	64,9	9,8	14,6	8,1	6,0	14,7	66,3	196	19,9	1126									
29	67,6	8,3	16,6	7,7	4,4	16,4	71,3	197	23,1	1286									
30	60,2	8,5	16,7	8,0	4,1	17,1	60,1	177	20,2	1883									
31	49,6	8,4	16,1	7,6	4,8	18,7	74,8	191	19,8	1386									
32	50,3	8,4	17,3	7,3	4,1	19,4	62,2	170	22,0	1549									
33	66,6	13,6	13,6	7,1	3,8	16,0	60,8	120	60,3	1817									

Рисунок 2 – Исходные данные

Для построения уравнения множественной регрессии в меню системы открыть **Statistics - Критерии** и выбрать в появившемся меню строку **Multiple Regression – Множественная регрессия**:

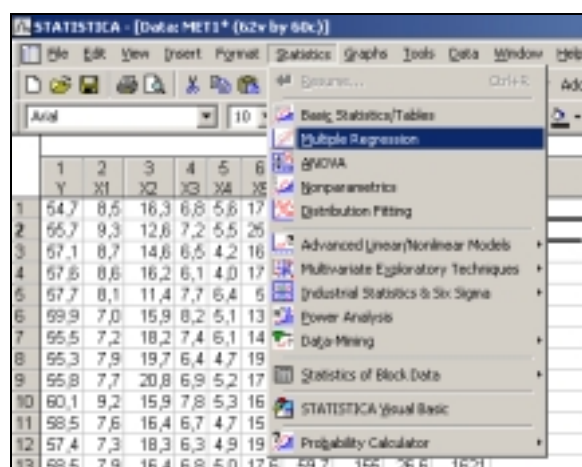


Рисунок 3 – Выбор пункта меню для проведения регрессионного анализа

На экране появится окно:

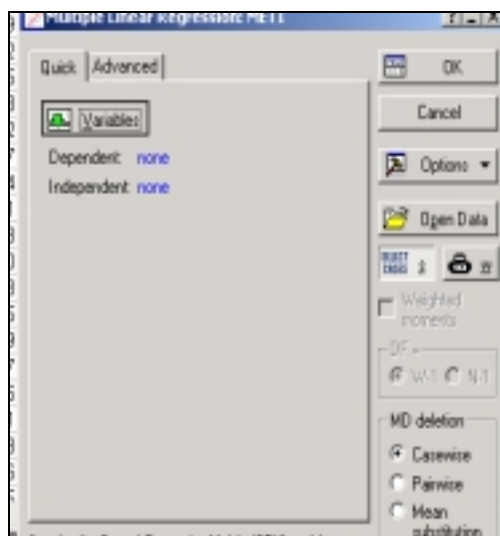


Рисунок 4 – Окно выбора переменных.

Далее необходимо выбрать зависимую (результатирующую, объясняемую) и независимые (объясняющие) переменные для анализа.

Для задания переменных воспользуемся кнопкой **Variables** – **Переменные** из панели **Multiple Regression – Множественная регрессия** (рисунок 4).

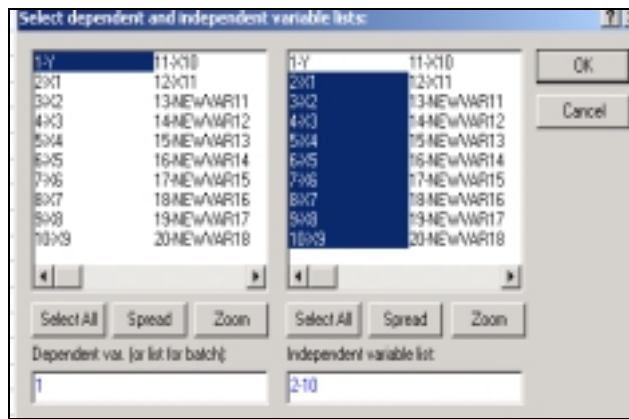


Рисунок 5 – Выбор зависимой и независимых переменных для проведения регрессионного анализа

В окне **Select dependent and independent variable list** – **Выбор зависимой переменной и списка независимых переменных**, выделяя имя переменной в левой части окна, производится выбор зависимой переменной **Dependent**. В правой части окна выбираем независимую переменную (**Independent**). Выбор нескольких несмежных переменных производят при нажатой клавише **CTRL**. После выбора переменных необходимо щелкнуть на кнопке **OK**, вновь окажемся в панели модуля Множественная регрессия. Нажатие на кнопку **Advanced** позволяет перейти к окну функциональных возможностей модуля Множественная регрессия.

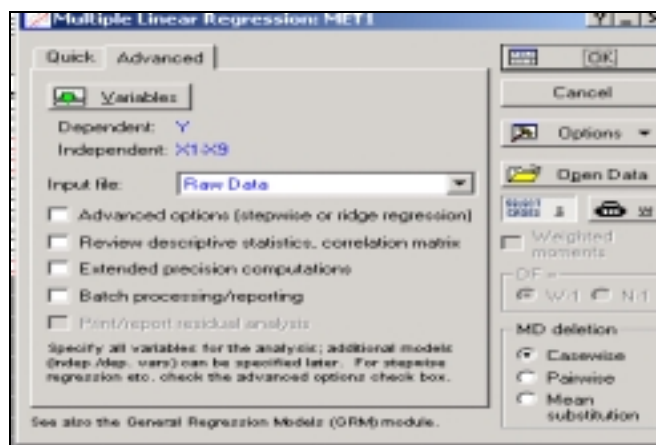


Рисунок 6 – Модуль множественная регрессия

Строка **Input file** определяет тип входной информации. Если входная информация представляет собой массив исходных данных, следует оставить **Raw Data** (необработанные данные). В поле окна **MD deletion** можно задать правило обработки пропущенных данных. Установка флажка в поле **Advanced options** позволит перейти к диалоговому окну **Model Defenition**, открывающему возможность выбора метода анализа, среди которых методы пошаговой регрессии и гребневой. Установка флажка в поле **Review descriptive statistics, correlations matrix** позволит провести предварительный анализ исходных переменных и построить корреляционную матрицу, анализ которой дает возможность сделать важные выводы о структуре связей между



выбранными переменными. Установка флажка в поле **Extended precision computations** позволит выбрать метод расчета с расширенной точностью. После определения всей необходимой информации для построения модели, щелкните по кнопке **ОК** в правом углу окна. Результаты расчетов приведены в виде отчета на рисунке 7.

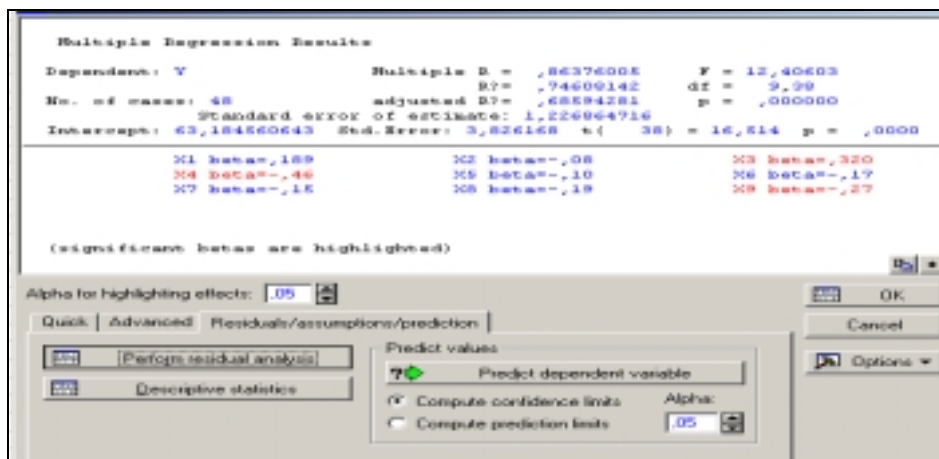


Рисунок 7 – Окно с результатами вычислений

В верхней информационной части окна результатов представлены основные характеристики построенной модели, а нижняя – содержит кнопки доступа к дополнительной информации, позволяющей провести исчерпывающий анализ модели, дать интерпретацию вычисленным параметрам и оценить адекватность модели исходным данным.

Рассмотрим содержание информационной части окна.

В левой части окна приводится имя зависимой переменной (**Dependent**) и число наблюдений, по которым построено уравнение регрессии (**No. Of Cases**).

В правой части окна приводится оценка коэффициента множественной корреляции (**Multiple R**) и значение квадрата этого коэффициента ( $R^2$ ) – коэффициента детерминации, несмещенная оценка  $R^2$  (**Adjusted  $R^2$** ), значение **F–критерий** /1/.

Также в верхней части окна результатов анализа приводится оценка свободного члена уравнения регрессии (**Intercept**), стандартная ошибка (среднеквадратическое отклонение) этой оценки (**Std. Error**), значение t-критерия и уровень значимости.

**Standard Error of estimate** является оценкой  $\sqrt{S_{\text{ост}}^2}$ , где  $S_{\text{ост}}^2$  – несмещенная оценка остаточной дисперсии.

Во второй части информационного окна подсвечены оценки значимых регрессионных коэффициентов (речь в данном случае идет о нормированных оценках: Beta-коэффициентах).

Более подробную информацию получим после нажатия на кнопку **Regression summary** (рисунок 8).

book6\* - Regression Summary for Dependent Variable: Y (MET1)

Insert Format Statistics Graphs Tools Data Workbook Window Help

Regression Summary for Dependent Variable: Y (MET1)  
 R= ,86376005 R²= ,74608142 Adjusted R²= ,68594281  
 F(9,38)=12,406 p<,00000 Std. Error of estimate: 1,2269

	Beta	Std. Err. of Beta	B	Std. Err. of B	t(38)	p-level
N=48						
Intercept			63,18456	3,826168	16,51380	0,000000
X1	0,188888	0,143370	0,16758	0,127199	1,31748	0,195565
X2	-0,081505	0,085997	-0,00769	0,008113	-0,94777	0,349237
X3	0,320317	0,110088	1,13001	0,368365	2,90966	0,006018
X4	-0,457356	0,154418	-1,12798	0,380840	-2,96181	0,005249
X5	-0,098744	0,085846	-0,07117	0,061873	-1,15024	0,257231
X6	-0,167099	0,095517	-0,03406	0,019467	-1,74941	0,088293
X7	-0,152619	0,134402	-0,00965	0,008496	-1,13554	0,263262
X8	-0,187146	0,133202	-0,04272	0,030406	-1,40498	0,168150
X9	-0,266202	0,117980	-0,00130	0,000577	-2,25633	0,029885

Рисунок 8 – Результаты оценивания параметров линейной модели множественной регрессии

В данном окне модуля представлены оценки параметров модели (**B**-обычные оценки и **Beta**- нормированные оценки), оценки их стандартных ошибок (**St. Error**) и уровни значимости (**p-level**) t-критерий Стьюдента /1-3/.

Далее можно приступить к исследованию остатков регрессионной модели. Остатки исследуются в специальном окне **Residuals analysis – Анализ остатков**. В нем приведен широкий набор статистических и визуальных методов исследования остатков модели. Для этого необходимо щелкнуть мышкой по кнопке **Residuals/assumptions/prediction – Остатки/распределение/предсказанные** в окне рисунка 7 (рисунок 9).

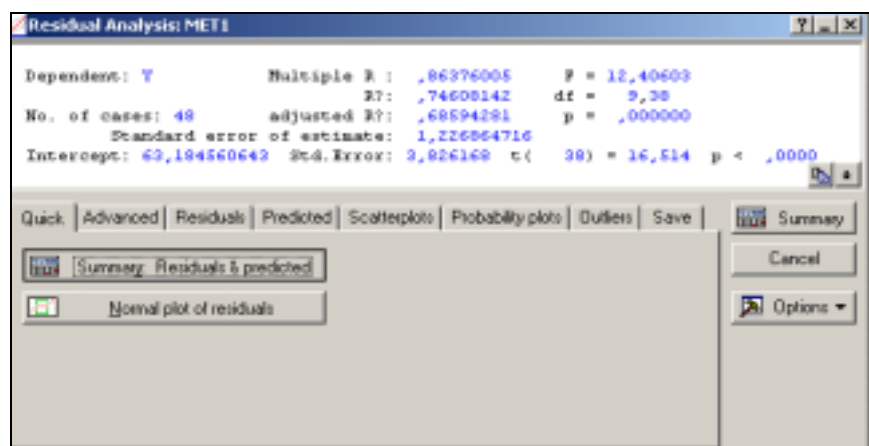


Рисунок 9 – Окно для анализа регрессионных остатков

Информация о значениях остатков может быть получена нажатием на кнопку **Summary: Residuals & predicted** (рисунок 10).

books\* - Predicted & Residual Values (MET1)

Dependent variable: Y

Case No.	Observed Value	Predicted Value	Residual	Standard Pred. y	Standard Residual	Std. Err. Pred. Val.	Mahalanobis Distance	Durbin Residual	Cook's Distance
1	54,70000	56,21296	-1,51296	-1,53777	-1,23318	0,562058	8,88513	-1,91483	0,0511
2	55,70000	56,75719	-1,05719	-1,24996	-0,86170	0,768218	17,44862	-1,73902	0,0788
3	57,10000	57,42084	-0,32084	-0,89900	-0,26151	0,591881	9,96973	-0,41817	0,0027
4	57,60000	57,45398	0,14602	-0,88148	0,11902	0,568884	9,12621	0,18602	0,0005
5	57,70000	58,09875	-0,39875	-0,54050	-0,32502	0,870725	22,68454	-0,80345	0,0216
6	59,90000	58,07185	1,82815	1,82816	-0,55473	1,49010	11,46983	2,48685	0,1088
7	55,50000	56,01296	-0,51296	-1,64353	-0,41811	0,522596	7,54863	-0,62666	0,0047
8	55,30000	56,34161	-1,04161	-1,46973	-0,84900	0,578350	9,46531	-1,33922	0,0265
9	55,80000	56,42005	-0,62005	-1,42825	-0,50539	0,551672	8,52394	-0,77719	0,0081
10	60,10000	58,78236	1,31764	-0,17900	1,07399	0,523481	7,57752	1,61092	0,0314
11	58,50000	58,36900	0,13100	-0,39759	0,10678	0,350604	2,85914	0,14265	0,0001
12	57,40000	56,84306	0,55694	-1,20455	0,45396	0,461144	5,66097	0,64857	0,0039
13	58,50000	58,39967	0,10033	-0,38137	0,08177	0,426808	4,70897	0,11414	0,0001
14	58,30000	57,66798	0,63202	-0,76831	0,51515	0,442326	5,13011	0,72645	0,0046
15	58,20000	59,90450	-1,70449	0,41442	-1,38931	0,567311	9,07037	-2,16807	0,0668
16	56,50000	58,38821	-1,88821	-0,38743	-1,53905	0,685727	13,70358	-2,74608	0,1955
17	59,20000	60,21419	-1,01419	0,57620	-0,82665	0,356279	2,98440	-1,10759	0,0069
18	58,10000	59,92093	-1,82093	0,42311	-1,48422	0,422034	4,58243	-2,06533	0,0335
19	58,80000	58,04236	0,75764	-0,57033	0,61754	0,400556	4,03077	0,84803	0,0051
20	56,50000	57,56338	-1,06338	-0,82362	-0,86675	0,360187	3,07182	-1,16368	0,0078
21	57,10000	58,31242	-1,21242	-0,42751	-0,98823	0,392317	3,80277	-1,35052	0,0124
22	58,30000	58,09826	0,20174	-0,54076	0,16443	0,499889	6,82364	0,24190	0,0006
23	59,40000	58,74685	0,65315	-0,19777	0,53238	0,509359	7,12210	0,78918	0,0071
24	61,20000	60,09682	1,10318	0,51613	0,89919	0,510295	7,15189	1,33396	0,0205
25	60,40000	60,99086	-0,59086	0,98892	-0,48160	0,489708	6,50906	-0,70284	0,0052
26	58,60000	58,31413	0,28587	-0,42661	0,23301	0,477503	6,14045	0,33691	0,0011
27	57,50000	57,28628	0,21372	-0,97016	0,17420	1,224763	45,85995	62,43660	258,1066
28	61,90000	60,00879	1,89121	0,48958	1,54150	0,433881	4,89906	2,16155	0,0388
29	61,00000	60,47494	0,52506	0,71609	0,42797	0,368717	3,26596	0,57720	0,0020
30	60,20000	60,44295	-0,24295	0,69917	-0,19803	0,583557	9,65422	-0,31399	0,0015

Рисунок 10 – Наблюдаемые, модельные значения результирующего признака, значения регрессионных остатков

Для проведения теста на нормальный характер распределения регрессионных остатков, скопируем столбец **Residual** в окно с исходными данными. Затем в меню системы Statistica выберем пункт **Distribution Fitting**. На экране появится окно:

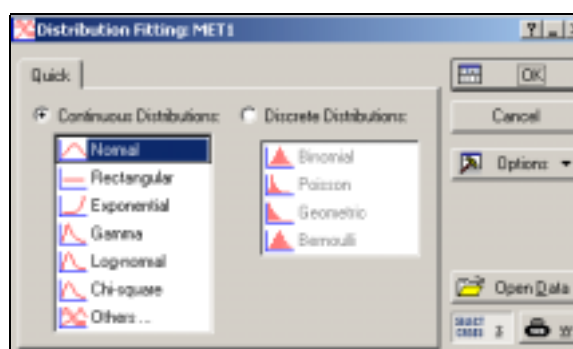


Рисунок 11 – Выбор вида распределения регрессионных остатков

В появившемся окне выберем распределение **Normal** – Нормальное и щелкнем по кнопке **OK**. После чего на экране появится окно:

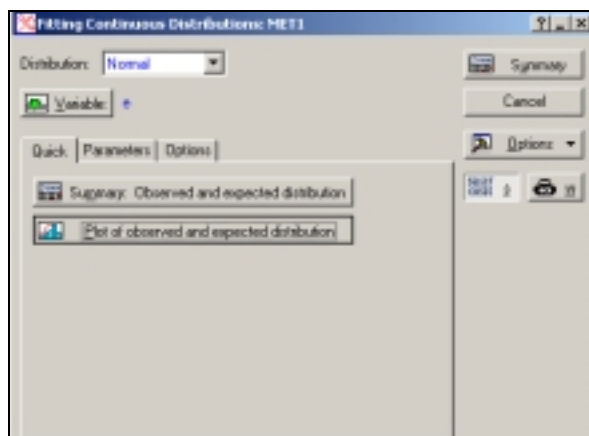


Рисунок 12 – Выбор пунктов для построения гистограммы регрессионных остатков

В данном окне сначала необходимо выбрать переменные, используя кнопку **Variable**. Кроме того, в данном модуле, используя кнопку **Parameters – Параметры**, можно изменить количество интервалов, верхнюю и нижнюю границы интервалов и т.д. Для получения графика нормального распределения, нажмем по кнопке **Plot of observed and expected distribution**.

На экране появится окно, содержащее гистограмму распределения, значение  $\chi^2$  – критерия, степени свободы, значимость нулевой гипотезы /1/.

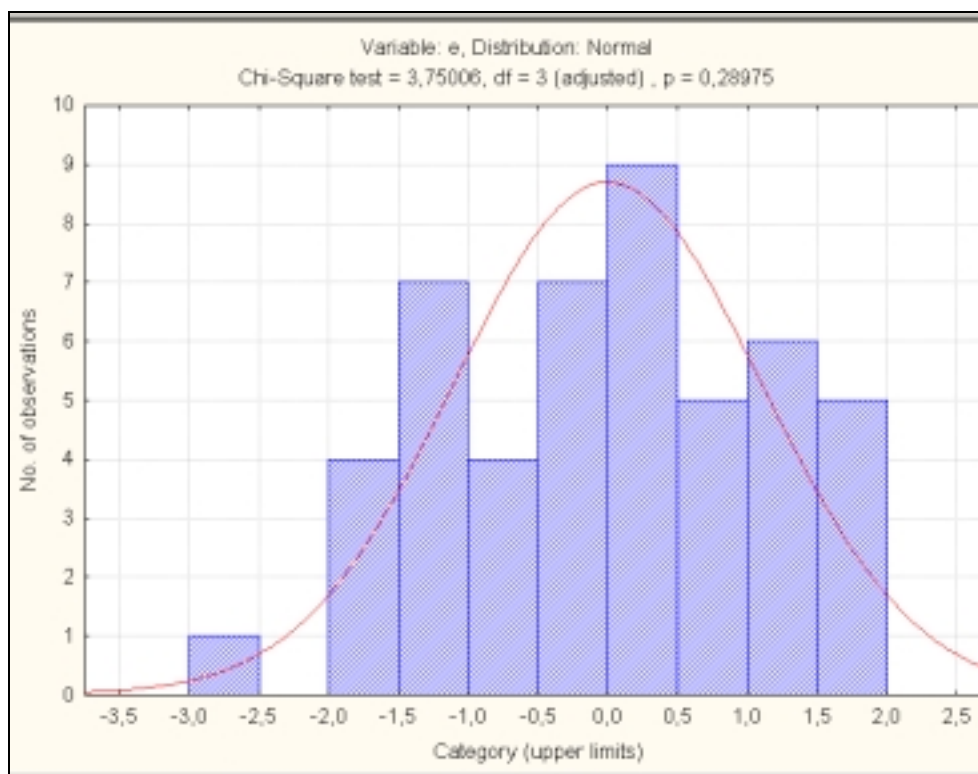
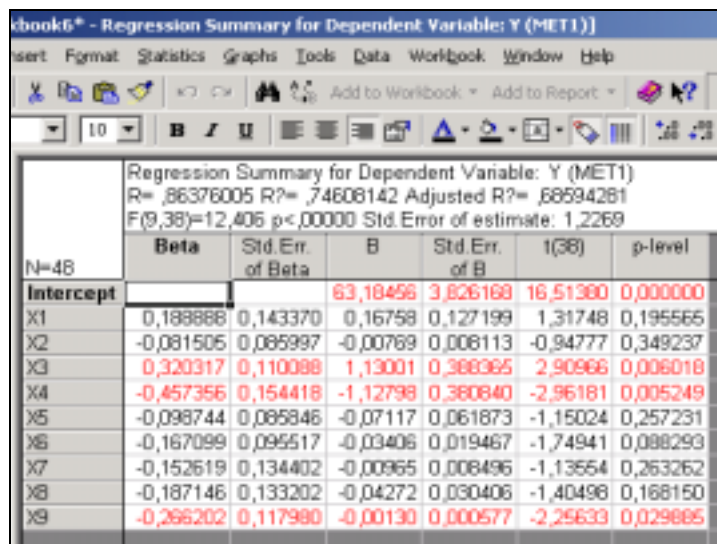


Рисунок 13 – График распределения регрессионных остатков

На уровне значимости 0,05 можно принять нулевую гипотезу о том, что распределение регрессионных остатков не отличаются от нормального, так как значимость нулевой гипотезы ( $p=0,29$ ).

Так как регрессионные остатки имеют нормальное распределение, то есть смысл проводить дальнейший анализ построенного уравнения множественной регрессии.

Итак, вернемся к окну **Multiple Regression Results** - Результаты множественной регрессии:



	Beta	Std. Err. of Beta	B	Std. Err. of B	t(38)	p-level
N=48						
Intercept			63,18456	3,826168	16,51380	0,000000
X1	0,188888	0,143370	0,16758	0,127199	1,31748	0,195565
X2	-0,081505	0,085997	-0,00769	0,008113	-0,94777	0,349237
X3	0,320317	0,110088	1,13001	0,368365	2,90966	0,006018
X4	-0,457356	0,154418	-1,12798	0,360840	-2,96181	0,005249
X5	-0,098744	0,085846	-0,07117	0,061873	-1,15024	0,257231
X6	-0,167099	0,095517	-0,03406	0,019467	-1,74941	0,088293
X7	-0,152619	0,134402	-0,00965	0,008496	-1,13554	0,263262
X8	-0,187146	0,133202	-0,04272	0,030406	-1,40498	0,168150
X9	-0,266202	0,117980	-0,00130	0,000577	-2,25633	0,029885

Рисунок 14 - Результаты оценивания параметров линейной модели множественной регрессии

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y} = 63,18 + 0,17X_1 - 0,008X_2 + 1,13X_3 - 1,13X_4 - 0,07X_5 - 0,034X_6 - 0,0097X_7 - 0,04X_8 - 0,001X_9$$

(3,83)   (0,13)   (0,008)   (0,39)   (0,38)   (0,06)   (0,02)   (0,008)   (0,03)   (0,0006)

Как видно из отчета, уравнение регрессии значимо, т.е. модель адекватна экспериментальным данным, значимыми оказались только коэффициенты при переменных  $X_3$ ,  $X_4$ ,  $X_9$ , среднеквадратические ошибки  $S_{b_j}$  оказались того же порядка, что и коэффициенты регрессии при переменных  $X_1$ ,  $X_2$ ,  $X_5$ ,  $X_6$ ,  $X_7$ ,  $X_8$ . Это свидетельствует о том, что коэффициенты при этих переменных могут иметь доверительный интервал, включающий в себя точку нуль. Согласно полученной модели при увеличении соотношения денежного дохода и прожиточного минимума на 1% ожидаемая продолжительность жизни мужчин уменьшится в среднем на 0,034 (коэффициент при переменной  $X_6$  имеет отрицательный знак), что противоречит экономическому смыслу. Все эти признаки позволяют нам заподозрить наличие мультиколлинеарности между объясняющими переменными. Итак, перейдем к рассмотрению критериев по выявлению мультиколлинеарности.

1. В первую очередь анализируют оценку матрицы парных коэффициентов корреляции между объясняющими переменными. Считается, что наличие значимых коэффициентов корреляции, по абсолютной величине превосходящих 0,7-0,8, свидетельствуют о присутствии мультиколлинеарности /1/.



Для вычисления оценки матрицы парных коэффициентов корреляции в окне множественная регрессия (рисунок 6) установим флажок в поле **Review descriptive statistics, correlations matrix**. После нажатия на кнопку **OK** на экране откроется окно.

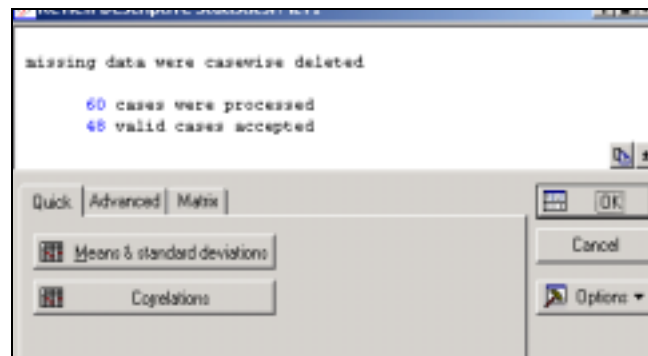


Рисунок 17 – Окно для вычисления оценки матрицы парных коэффициентов корреляции

В открывшемся окне нажимаем кнопку **Correlations** для вычисления оценки матрицы парных коэффициентов корреляции.

Variable	X1	X2	X3	X4	X5	X6	X7	X8	X9	Y
X1	1.000000	-0.163471	-0.033347	-0.734697	0.065213	-0.303998	-0.485993	0.566683	-0.554129	0.687535
X2	-0.163471	1.000000	-0.119662	-0.005496	-0.028424	0.130073	0.136329	-0.155531	0.107924	-0.187556
X3	-0.033347	-0.119662	1.000000	0.324062	-0.155090	0.320078	0.214232	-0.113756	-0.310372	0.208605
X4	-0.734697	-0.005496	0.324062	1.000000	-0.155051	0.338975	0.513321	-0.463331	0.481303	-0.652968
X5	0.065213	-0.028424	-0.155090	-0.155051	1.000000	-0.163029	0.025993	0.054247	0.031511	-0.058139
X6	-0.303998	0.130073	0.320078	0.338975	-0.163029	1.000000	0.388647	-0.368066	0.101189	-0.288899
X7	-0.485993	0.136329	0.214232	0.513321	0.025993	0.388647	1.000000	-0.745229	0.148294	-0.389195
X8	0.566683	-0.155531	-0.113756	-0.463331	0.054247	-0.368066	-0.745229	1.000000	-0.235037	0.340489
X9	-0.554129	0.107924	-0.310372	0.481303	0.031511	0.101189	0.148294	-0.235037	1.000000	-0.697877
Y	0.687535	-0.187556	0.208605	-0.652968	-0.058139	-0.288899	-0.389195	0.340489	-0.697877	1.000000

Рисунок 18 – Оценка матрицы парных коэффициентов корреляции

На основе вычисленной матрицы есть основания подозревать тесную связь между  $X_1$  и  $X_4$  ( $r(x^{(1)}, x^{(4)}) = 0,73$ ) и  $X_7$  и  $X_8$  ( $r(x^{(7)}, x^{(8)}) = -0,75$ ).

2. Более внимательное изучение этого вопроса достигается с помощью расчета значений коэффициентов детерминации  $\hat{R}_{x^{(j)}.X(j)}^2$  каждой из объясняющих переменных  $x^{(j)}$  по всем остальным переменным  $X(j) = (x^{(1)}, \dots, x^{(j-1)}, x^{(j+1)}, \dots, x^{(p)})$ .

Для определения коэффициентов детерминации следует воспользоваться модулем множественная регрессия, где в качестве зависимой переменной выбрать  $x^{(j)}$ , все остальные объясняющие переменные в качестве независимых (рисунок 19).

Multiple Regression Results			
Dependent: X1	Multiple R =	0,82153385	F = 10,12115
	R² =	0,67491655	df = 8,39
No. of cases: 48	adjusted R² =	0,60823277	p = 0,000000
Standard error of estimate: 1,544474752			
Intercept: 12,830009051	Std. Error: 4,356569	t( 39) = 2,9450	p = 0,0054

Рисунок 19 – Оценка коэффициента детерминации переменной  $x_1$

Все расчеты остальных коэффициенты детерминации производятся аналогичным образом. В результате получили:

$$R^2_{x_1 / x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9} = 0.67498$$

$$R^2_{x_2 / x_1 x_3 x_4 x_5 x_6 x_7 x_8 x_9} = 0.0965$$

$$R^2_{x_3 / x_1 x_2 x_4 x_5 x_6 x_7 x_8 x_9} = 0.4486$$

$$R^2_{x_4 / x_1 x_2 x_3 x_5 x_6 x_7 x_8 x_9} = 0.7198$$

$$R^2_{x_5 / x_1 x_2 x_3 x_4 x_6 x_7 x_8 x_9} = 0.0933$$

$$R^2_{x_6 / x_1 x_2 x_3 x_4 x_5 x_7 x_8 x_9} = 0.2676$$

$$R^2_{x_7 / x_1 x_2 x_3 x_4 x_5 x_6 x_8 x_9} = 0.6301$$

$$R^2_{x_8 / x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_9} = 0.6234$$

Рисунок 20 – Результаты вычислений оценок коэффициента детерминации

Анализ оценок коэффициентов детерминации показал наличие тесной линейной связи между объясняющей переменной  $X_4$  и всеми остальными признаками, то же самое можно сказать о переменных  $X_7$ ,  $X_8$ ,  $X_1$ .

3. Достаточным условием плохой обусловленности матрицы (наличия мультиколлинеарности) является большое значение числа обусловленности:

$$M = \frac{|\lambda_{\max}|}{|\lambda_{\min}|},$$

$\lambda_{\max}$  - максимальное собственное число матрицы  $X^T X$  /1/. Для вычисления собственных чисел матрицы  $X^T X$  воспользуемся функцией `eigenvals` из Mathcad (рисунок 21).

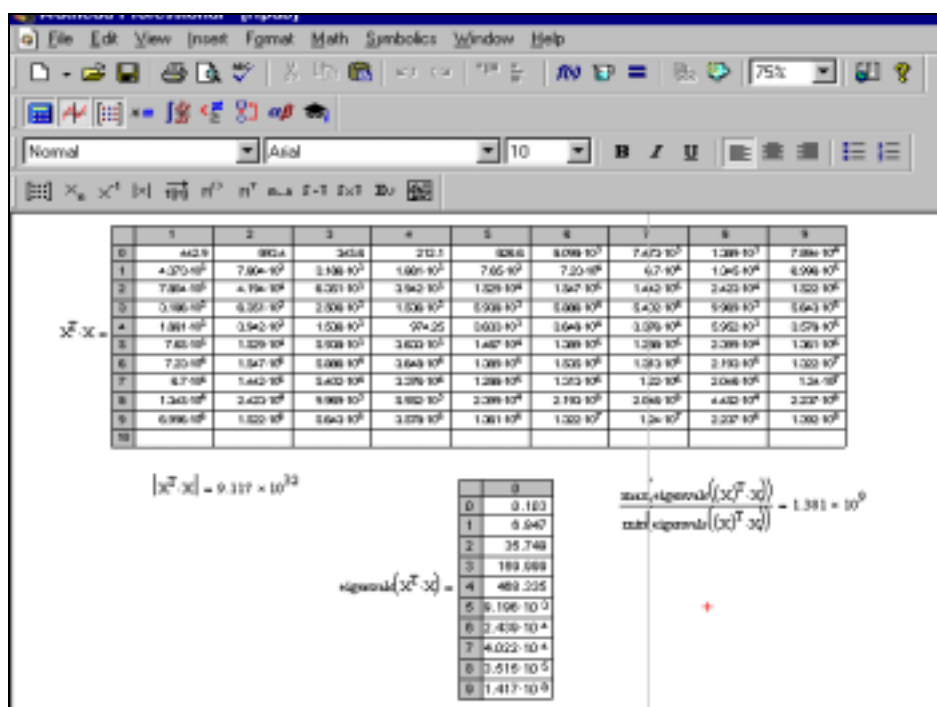


Рисунок 21 – Результаты вычислений в программе Mathcad

Таким образом, можно говорить о наличии мультиколлинеарности между объясняющими переменными  $X_1, \dots, X_{17}$ .

В случае, если между объясняющими переменными существует частичная мультиколлинеарность, то оценки коэффициентов линейной модели, полученные по МНК, становятся неустойчивыми, незначительное изменение состава выборки или состава объясняющих переменных может вызвать кардинальное изменение модели, что делает модель непригодной для практических целей. Наиболее распространенные в таких случаях приемы оценивания параметров регрессионной модели: методы пошаговой регрессии, использование гребневой регрессии (ридж-регрессии), переход от первоначальных переменных к их главным компонентам [1-3]. Все вышеприведенные методы реализуются в ППП Statistica. Рассмотрим некоторые методы устранения регрессии, используя модуль «множественная регрессия».

Установка флажка в поле **Advanced options** модуля множественная регрессия (рисунок 6) позволит перейти к диалоговому окну **Model Defenition**, открывающему возможность выбора метода анализа, среди которых методы пошаговой регрессии и гребневой (метод ридж-регрессии). В прокручиваемом списке методов можно выбрать один из методов пошаговой регрессии. В модуле реализованы две процедуры отбора переменных, каждая из которых может давать различный конечный набор переменных: последовательное включение (**Forward stepwise**) и последовательное исключение (**Backward stepwise**).

В данном случае выбран пошаговый метод включения:



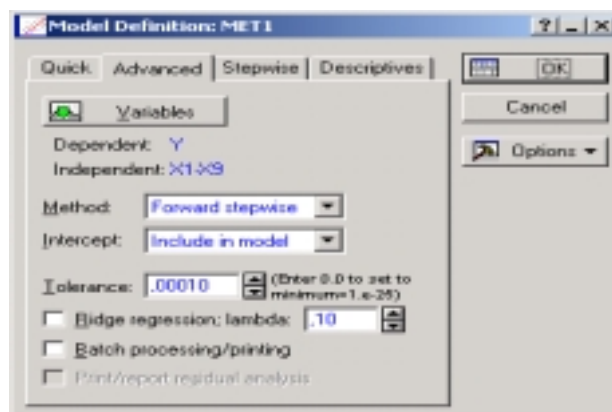


Рисунок 22 – Выбор метода оценивания параметров регрессионной модели

Результаты расчетов приведены в виде отчета на рисунке 23.

book8\* - Regression Summary for Dependent Variable: Y (MET1)

Insert Format Statistics Graphs Tools Data Workbook Window Help

Regression Summary for Dependent Variable: Y (MET1)  
 $R = .82174263$   $R^2 = .67526096$  Adjusted  $R^2 = .65311966$   
 $F(3,44) = 30.498$   $p < .00000$  Std. Error of estimate: 1.2894

	Beta	Std. Err. of Beta	B	Std. Err. of B	t(44)	p-level
N=48						
Intercept			60.41614	2.771840	21.79640	0.000000
X9	-0.313403	0.118526	-0.00153	0.000579	-2.64416	0.011308
X4	-0.601357	0.119100	-1.48313	0.293737	-5.04917	0.000008
X3	0.306210	0.109621	1.08024	0.387426	2.78626	0.007798

Рисунок 23 – Результаты оценивания параметров линейной модели множественной регрессии методом пошаговой регрессии

Были исследованы также регрессионные остатки, анализ которых показал нормальность их распределения (рисунок 24).

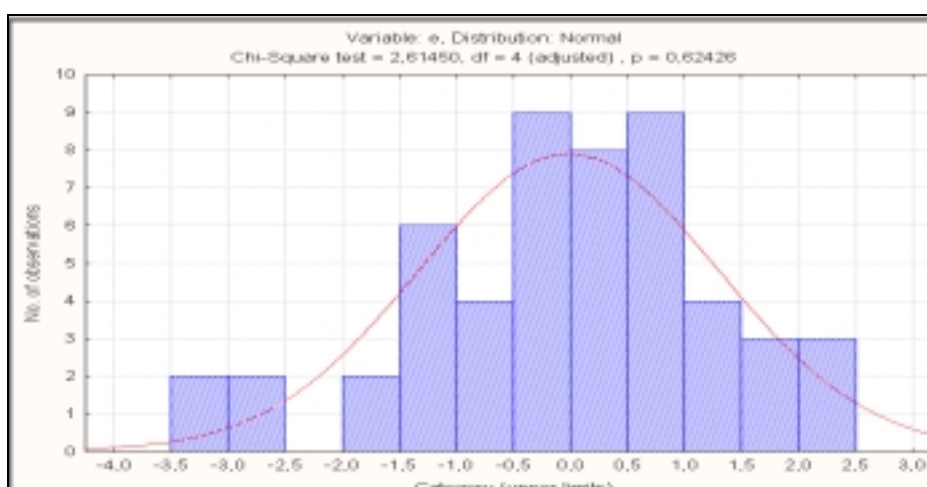


Рисунок 24 – Гистограмма распределения регрессионных остатков

В результате проведения пошаговой регрессии получили следующую оценку уравнения регрессии:

$$\hat{y} = 60,42 + 1,08x_3 - 1,48x_4 - 0,0015x_9 \quad (2)$$

(2,77) (0,39) (0,29) (0,0006)

Оценка уравнения регрессии значима т.к. нулевая гипотеза отклонена; коэффициенты при переменных также значимы. Коэффициент детерминации составил 0,675, т.е. 67,5% доли вариации результирующей переменной объясняется переменными  $X_3$ ,  $X_4$  и  $X_9$ , а 32,5% доли вариации, вероятно, объясняется неучтенными в модели факторами.

Согласно полученной модели, можно сделать вывод о том, что увеличение количества браков приводит к росту ожидаемой продолжительности жизни мужчин в среднем на 1,08 лет, при росте количества разводов ожидаемая продолжительность жизни мужчин в среднем сокращается на 1,48 лет, при увеличении числа зарегистрированных преступлений ожидаемая продолжительность жизни мужчин в среднем также сокращается на 0,0015 лет.

Для реализации метода гребневой регрессии (ридж-регрессии), необходимо в окне **Model Defenition** (рисунок 22) установить флажок в поле **Ridge regression** и указать величину «гребня», «хребта» в диапазоне значений от 0,1 до 0,4 /1/.

## **4 Содержание письменного отчета**

Отчет должен быть оформлен на листах формата А4 с титульным листом, оформленным соответствующим образом и содержать следующее:

- 1) постановку задачи с вариантом выборок;
- 2) краткое изложение теории по исследованию ЛММР на мультиколлинеарность;
- 3) результаты компьютерной обработки данных;
- 4) анализ полученных результатов;
- 5) выводы по полученным результатам.

## **5 Вопросы к защите**

- 1 Раскройте понятие полной и частичной мультиколлинеарности.
- 2 Укажите причины и признаки мультиколлинеарности.
- 3 Укажите формальные признаки мультиколлинеарности.
- 4 К чему ведет отбрасывание незначимых коэффициентов в модели регрессии.
- 5 Поясните суть пошаговой регрессии.
- 6 Поясните суть «ридж-регрессии».
- 7 В чем суть метода главных компонент, как средства устранения мультиколлинеарности.

## **Список использованных источников**

- 1 **Айвазян С.А., Мхитарян В.С.** Прикладная статистика и основы эконометрики [Текст]: учебник для вузов/ С.А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022 с.
- 2 **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика. Начальный курс [Текст]: учебник/ Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 57 с.
- 3 **Тихомиров Н.П., Дорохина Е.Ю.** Эконометрика [Текст]: учебник/ Н.П. Тихомиров, Е.Ю. Дорохина. – М.: Издательство «Экзамен», 2003. – 512 с.

## Приложение А (обязательное)

### Исходные данные для анализа

Таблица А.1 – Выборочные данные

	Y	x1	x2	x3	x4	x5	x6	x7	x8	x9
1	2	3	4	5	6	7	8	9	10	11
<b>Российская Федерация</b>	<b>58,3</b>	<b>9,3</b>	<b>15</b>	<b>7,3</b>	<b>4,5</b>	<b>18,1</b>	<b>202</b>	<b>179</b>	<b>24,7</b>	<b>1860</b>
<b>Северный район</b>	<b>56,8</b>	<b>8,7</b>	<b>14,2</b>	<b>6,8</b>	<b>5</b>	<b>18,5</b>	<b>75</b>			
Республика Карелия	54,7	8,5	16,3	6,8	5,6	17,4	163	151	23,6	2344
Республика Коми	5,7	9,3	12,6	7,2	5,5	25,3	194	239	9,2	1809
Архангельская область	7,1	8,7	4,6	6,5	4,2	16,2	152	192	26,9	2406
Вологодская область	57,6	8,6	6,2	6,1	4	17,4	190	205	20,1	2023
Мурманская область	57,7	8,1	11,4	7,7	6,4	5,9	183	198	22	1419
<b>Северо-Западный район</b>	<b>58</b>	<b>7,2</b>	<b>17,3</b>	<b>7,7</b>	<b>5,3</b>	<b>14,9</b>				
г. Санкт-Петербург	59,9	7	15,9	8,2	5,1	13,8	229	172	20	2104
Ленинградская область	55,5	7,2	18,2	7,4	6,1	14,3	146	167	29,1	2489
Новгородская область	55,3	7,9	19,7	6,4	4,7	19,8	174	144	22,8	2428
Псковская область	55,8	7,7	20,8	6,9	5,2	17,1	128	111	42,7	2494
<b>Центральный район</b>	<b>57,8</b>	<b>7,7</b>	<b>17,3</b>	<b>7,6</b>	<b>4,8</b>	<b>16,6</b>				
Брянская облсть	60,1	9,2	15,9	7,8	5,3	16,7	169	148	22,7	2094
Владимирская область	58,5	7,6	16,4	6,7	4,7	15,5	144	150	27,9	1768
Ивановская область	57,4	7,3	18,3	6,3	4,9	19,6	138	133	33,7	1982
Калужская область	58,5	7,9	16,4	6,8	5	17,6	197	155	26,6	1621
Костромская область	58,3	7,9	17	6,3	4,4	20,1	182	159	30,5	1631
г.Москва	58,2	8	16,9	8,2	4,6	15,5	520	197	19,1	1066
Московская область	56,5	7,2	17,6	8,1	5,2	16,1	143	165	31,2	1183
Орловская область	59,2	8,7	16	7,6	4,4	18,9	214	161	22,7	1308
Рязанская область	58,1	7,8	17,9	7,2	4,3	15,7	158	163	24,4	1475
Смоленская область	58,8	8	16,9	6,9	4,7	16,8	185	146	19,8	2081
Тверская область	56,5	7,5	19,4	6,7	4,6	19,3	153	165	28,6	2109
Тульская область	57,1	7,3	19,4	7,4	5	20,1	200	175	16,2	1757
Ярославская область	58,3	7,6	17,3	7,1	5,3	12	180	154	21,3	2111
<b>Волго-Вятский район</b>	<b>58,7</b>	<b>8,6</b>	<b>15,8</b>	<b>6,6</b>	<b>3,7</b>	<b>16,4</b>				
Республика Марий Эл	59,4	9,6	13	6,4	3,5	16,8	120	117	43,2	2112
Республика Мордовия	61,2	9	14,1	7	3,3	15,2	132	126	34,7	1794
Чувашская Республика	60,4	10,2	13	7,1	3,2	16,1	145	121	27,3	1688
Кировская область	58,6	8,1	16,3	6,2	3,9	17,1	137	121	32	1774
Нижегородская область	57,5	8	17,5	6,7	4	16,4	181	182	22	1773
<b>Центрально-Чернозмый район</b>	<b>60,5</b>	<b>8,5</b>	<b>16,3</b>	<b>7,8</b>	<b>4,4</b>	<b>16,4</b>				
Белгородская область	61,9	9,4	14,8	8,1	5	14,7	200	195	19,9	1128
Воронежская область	61	8,3	16,6	7,7	4,4	15,4	182	157	23,1	1295
Курская область	60,2	8,5	16,7	8	4,1	17,1	179	177	20,2	1803
Липецкая область	59,5	8,4	16,1	7,6	4,6	16,7	181	191	18,6	1358
Тамбовская область	59,3	8,4	17,3	7,3	4,1	19,4	183	170	22	1549
<b>Поволжсий район</b>	<b>60,2</b>	<b>9,3</b>	<b>14,1</b>	<b>7,1</b>	<b>4,4</b>	<b>18,5</b>				
Республика Калмыкия	59,8	13,5	10,5	7,1	3,4	15,8	100	120	60,3	1417
Республика Татарстан	60,2	10,4	12,9	7	3,9	18,5	194	225	22,1	1581

Продолжение таблица А.1

1	2	3	4	5	6	7	8	9	10	11
Астраханская область	60	10,1	13,5	7,1	4,6	18,6	143	137	32,1	1938
Волгоградская область	60,7	9,1	14,6	7,5	5,1	19,1	141	160	33,2	1443
Пензенская область	60,9	8,2	15	7,1	4	14,7	148	121	30,2	1121
Самарская область	59,1	8,6	14,8	7,3	4,8	14	188	207	21,2	1511
Саратовская область	60,3	8,9	14,5	6,9	4,2	23,6	138	117	35,3	1601
Ульяновская область	60,7	8,9	13,4	6,7	3,9	21,8	198	206	16,3	1193
<b>Северо-Кавказский район</b>	<b>60,5</b>	<b>12</b>	<b>13,6</b>	<b>7,9</b>	<b>4,1</b>	<b>19</b>				
Республика Адыгея	60,6	10,7	14,4	8	4	18,7	129	130	46,3	1344
Республика Дагестан	65,9	21,8	7,5	6,9	1,3	17,6	86	79	41,5	673
Кабардино-Балкарская Республика	62,6	13,7	10,4	7,1	3,4	14,5	128	102	42,5	859
Карачаево-Черкесская	63,9	12,9	10,3	7	3,3	16,3	123	107	45,7	925
Северная осетия	59,5	13,3	13	6,6	2,6	17,8	128	101	42,8	968
Красноярский край	59,4	10	15,3	8,8	5	19,2	175	160	32,4	1565
Ставропольский край	61	10,7	13,5	8,1	4,5	21,7	151	154	39,6	1325
Ростовская область	59,4	9,2	15,8	8	4,8	18,7	146	140	33,4	1497
<b>Уральский район</b>	<b>58,2</b>	<b>9,5</b>	<b>14,5</b>	<b>6,9</b>	<b>4,2</b>	<b>18,3</b>				
Республика Башкортостан	60,2	11,2	12,7	7,3	3,7	18,3	158	191	32,4	1059
Удмуртская республика	57,5	9,4	13,7	6,7	3,4	18,4	158	144	26,1	1915
Курганская область	68,9	9	14,6	7,3	4,3	22,6	113	132	50,4	2660
Оренбургская область	59	10,3	13,5	7,5	4	19,7	115	145	49,3	1534
Пермская область	56,9	9,2	15,8	5,9	3,8	18,9	184	175	25,7	2654
Свердловская область	57,7	8,5	15,6	6,7	4,7	17,5	163	169	29,5	2508
Челябинская область	58	9	14,8	7	4,7	16,6	171	182	27,9	1987
<b>Западно-Сибирский</b>	<b>58</b>	<b>9,4</b>	<b>13,5</b>	<b>7,3</b>	<b>4,9</b>	<b>19,3</b>				
Республика Алтай	55,1	14,2	13,1	7,1	3,8	27,9	188	148	26,2	2176
Алтайский край	58,3	8,7	14,7	7,3	4,4	20,8	158	146	33,7	1871
Кемеровская область	55,4	8,9	16,6	7	4,9	19,6	254	260	16,1	1563
Новосибирская область	59,1	8,5	14,1	7	4,5	15,9	136	156	39,8	2665
Омская область	60,6	10,2	12,3	7,3	4,6	16,3	157	170	29,7	2273
Томская область	58,2	9,1	13	7	5,3	21,2	173	190	30,6	2635
Тюменская область	57,8	10,6	9,8	7,9	5,7	21,3	290	293	19,2	2478
<b>Восточно-Сибирский</b>	<b>55,5</b>	<b>11</b>	<b>13,7</b>	<b>6,8</b>	<b>4</b>	<b>19,6</b>				
Республика Бурятия	57,2	11,7	12	6,5	3,5	15,2	122	155	55,2	2580
Республика Тыва	49,7	20	13	5,9	1,9	28	84	101	73,2	2713
Республика Хакасия	56	9,9	14	7,1	4,4	24,6	161	201	25,3	2222
Красноярский край	56	9,8	14	7,2	4,8	19,8	246	296	24,2	2417
Иркутская область	54,7	10,6	14,6	6,3	3,3	18,1	170	215	32,3	2317
Читинская область	56,2	12,2	12,8	6,9	4	20,8	99	112	66,5	2784
<b>Дальневосточный</b>	<b>56,7</b>	<b>10,2</b>	<b>12,6</b>	<b>7,1</b>	<b>5,3</b>	<b>20,5</b>				
Республика Саха	57	15,3	9,8	8	4,7	19,5	170	201	29,2	1483
Еврейская автономная область	55	10,9	13,6	7,3	5,2	26,4	130	125	28,5	3276
Чукотский автономный округ	57,8	9,8	8,6	7,3	8,9	34	85	73	26,4	1148
Приморский край	57,8	9,4	13,1	6,6	4,7	21,5	144	170	31,8	3095

Продолжение таблица А.1

1	2	3	4	5	6	7	8	9	10	11
Хабаровский край	57,2	9,3	13,1	6,6	5,7	17,8	153	171	29,4	2881
Амурская область	58	10,1	12	7,2	4,9	23,6	175	187	37,9	2017
Камчатская область	56,3	9,1	11,2	7,9	6,7	15,4	211	228	22,7	2064
Магаданская область	55,7	8,3	10,9	7,2	7,1	14,2	202	187	24,6	3068
Сахалинская область	50,7	8,9	17	7,2	5,7	22,7	145	169	24,6	3588
Калининградская область	58,9	8,6	13,6	7,8	6	15,4	155	145	26,6	2471

Таблица А.2 – Варианты заданий

<b>№ варианта</b>	<b>Название районов РФ</b>
1	2
1	Северный, Северо-Западный, Центральный, Волго-Вятский, Центрально-Черноземный, Северо-Кавказский, Уральский районы
2	Северный, Центральный, Волго-Вятский, Центрально-Черноземный, Поволжский, Уральский, Западно-Сибирский районы
3	Северный, Северо-Западный, Волго-Вятский, Центрально-Черноземный, Поволжский, Северо-Кавказский, Уральский, Западно-Сибирский районы
4	Северный, Северо-Западный, Центральный, Центрально-Черноземный, Поволжский, Северо-Кавказский, Уральский районы
5	Северный, Северо-Западный, Центральный, Центрально-Черноземный, Поволжский, Северо-Кавказский, Западно-Сибирский районы
6	Северный, Северо-Западный, Центральный, Центрально-Черноземный, Северо-Кавказский, Уральский, Западно-Сибирский районы
7	Северный, Северо-Западный, Центральный, Центрально-Черноземный, Уральский, Западно-Сибирский, Восточно-Сибирский районы
8	Северный, Центральный, Волго-Вятский, Северо-Кавказский, Восточно-Сибирский, Дальневосточный районы
9	Северный, Центральный, Центрально-Черноземный, Поволжский, Восточно-Сибирский, Дальневосточный районы
10	Северный, Центральный, Волго-Вятский, Поволжский, Восточно-Сибирский, Дальневосточный районы
11	Северо-Западный, Центральный, Волго-Вятский, Северо-Кавказский, Уральский, Западно-Сибирский, Восточно-Сибирский районы
12	Северо-Западный, Центральный, Центрально-Черноземный, Поволжский, Восточно-Сибирский, Дальневосточный районы
13	Северо-Западный, Центральный, Поволжский, Северо-Кавказский, Уральский, Восточно-Сибирский, Дальневосточный районы,

Продолжение таблицы А.2

1	2
14	Центральный, Волго-Вятский, Центрально-Черноземный, Поволжский, Северо-Кавказский, Уральский, Западно-Сибирский районы
15	Центральный, Центрально-Черноземный, Северо-Кавказский, Западно-Сибирский, Восточно-Сибирский, Дальневосточный районы
16	Центральный, Поволжский, Северо-Кавказский, Уральский, Восточно-Сибирский, Дальневосточный районы
17	Центральный, Северо-Кавказский, Уральский, Западно-Сибирский, Восточно-Сибирский, Дальневосточный районы
18	Волго-Вятский, Центрально-Черноземный, Поволжский, Северо-Кавказский, Уральский, Западно-Сибирский, Восточно-Сибирский районы
19	Волго-Вятский, Поволжский, Северо-Кавказский, Уральский, Западно-Сибирский, Восточно-Сибирский, Дальневосточный районы
20	Центрально-Черноземный, Поволжский, Северо-Кавказский, Уральский, Западно-Сибирский, Восточно-Сибирский, Дальневосточный район



Таблица А.3 – Наименование показателей

Наименование показателей	Обозначение
1 Ожидаемая продолжительность жизни мужчин при рождении (число лет)	$y$
2 Рождаемость населения (на 1000 человек)	$x_1$
3 Смертность населения (на 1000 человек);	$x_2$
4 Браки на 1000 населения	$x_3$
5 Разводы на 1000 населения	$x_4$
6 Коэффициент младенческой смертности ( число детей, умерших в возрасте до 1 года, на 1000 родившихся)	$x_5$
7 Соотношение денежного дохода и прожиточного минимума (%)	$x_6$
8 Соотношение средней оплаты труда с учетом выплат социального характера и прожиточного минимума трудоспособного населения (%)	$x_7$
9 Численность населения с денежными доходами ниже прожиточного минимума в % от численности населения региона	$x_8$
10 Число зарегистрированных преступлений по регионам РФ (на 100000 населения)	$x_9$