# Лабораторная работа 2. Проверка выполнения предпосылок метода наименьших квадратов (МНК)

**КРАТКАЯ СПРАВКА**

Проверка наличия или отсутствия систематической ошибки (выполнения предпосылок метода наименьших квадратов (МНК)) осуществляется на основе анализа ряда остатков.



## Решение типовых задач

**Пример**

Имеются следующие выборочные данные – выборка 20% случайная бесповоротная (табл. 2.1).

Таблица 2.1.

Среднедушевые денежные доходы населения и среднедушевой оборот розничной торговли по городам региона (месяц)

|  |  |  |
| --- | --- | --- |
| № города | Среднедушевой денежный доход, руб | Среднедушевой оборот розничной торговли, руб |
| 1 | 3357 | 2425 |
| 2 | 3135 | 2050 |
| 3 | 2842 | 1683 |
| 4 | 3991 | 2375 |
| 5 | 2293 | 1167 |
| 6 | 3563 | 2200 |
| 7 | 3219 | 1892 |
| 8 | 3308 | 2008 |
| 9 | 3724 | 2225 |
| 10 | 3416 | 1983 |
| 11 | 3340 | 1925 |
| 12 | 3089 | 1042 |
| 13 | 4372 | 2925 |
| 14 | 3022 | 2342 |
| 15 | 3383 | 2458 |
| 16 | 4267 | 2125 |

**ТРЕБУЕТСЯ**

1. Построить поле корреляции.
2. Построить уравнение регрессии, оценить его адекватность, сделать выводы.
3. Проверить выполнение предпосылок метода наименьших квадратов.
4. Оценить точность модели.

**Решение**

* 1. Постройте поле корреляции. Для этого воспользуйтесь мастером диаграмм (тип диаграммы – точечная) (рис. 2.1).



Рис. 2.1. Поле корреляции

* 1. Для получения решения воспользуйтесь инструментом **Анализа данных «Регрессия» (Данные / Анализ данных / Регрессия)**.

Результаты регрессионной статистики представлены на рисунке 2.2.1-2.2.4.

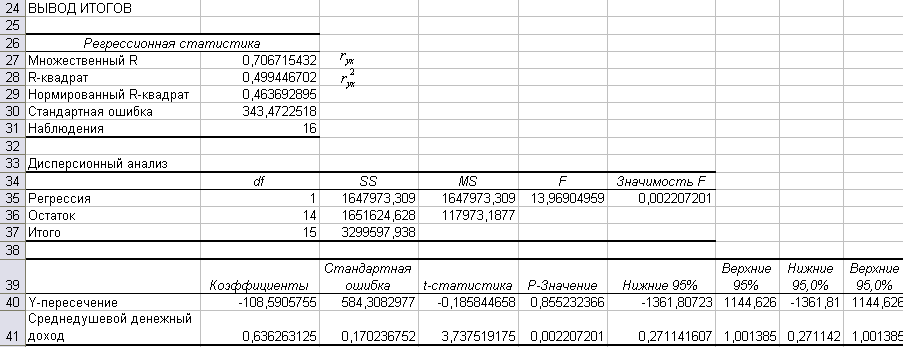


Рис. 2.2.1 Результаты регрессионной статистики

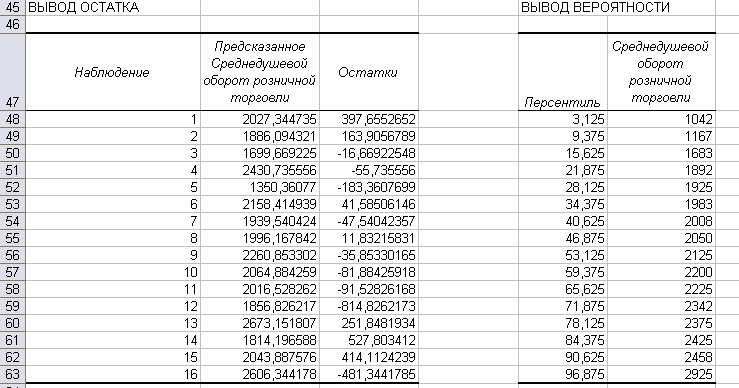


Рис. 2.2.2 Результаты регрессионной статистики (продолжение)



Рис. 2.2.3 График остатков



Рис. 2.2.4 График нормального распределения

**Выводы**

1. *Теснота связи между признаками определяется по величине парного коэффициента корреляции. Величина коэффициента корреляции () (рис. 2.2.1, множественный R) свидетельствует о тесной связи между среднедушевым денежным доходом и среднедушевым оборотом розничной торговли.*
2. *Парный коэффициент детерминации () (рис. 2.2.1, R-квадрат) показывает, что на 49,94% изменение оборота розничной торговли объясняется изменениями денежных доходов населения.*
   1. Значимость коэффициента корреляции проверяется с помощью t-критерия Стьюдента

(),

где *k* – число факторных признаков, включенных в модель (рис. 2.3).

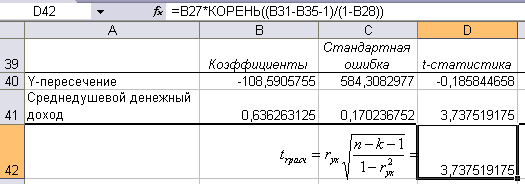


Рис.2.3. Фактическое значение t-критерия Стьюдента

* 1. Критическое (табличное) значение t-критерия Стьюдента при доверительной вероятности 0,95 (α=0,05) и числе степеней свободы составляет 2,1447 (для расчета используйте статистическую функцию **СТЬЮДРАСПОБР(α;n-k-1**)) (рис. 2.4).

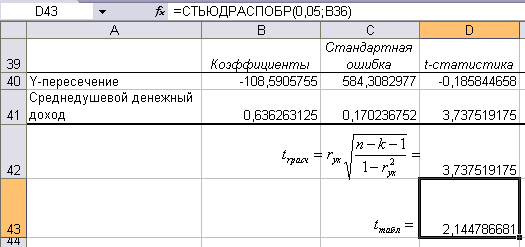


Рис.2.4. Табличное значение t-критерия Стьюдента

**Выводы**

*Так как , значение коэффициента корреляции признается значимым и делается вывод о том, что между среднедушевым денежным доходом и среднедушевым оборотом розничной торговли есть тесная статистическая взаимосвязь.*

1. Уравнение парной регрессии (рис.2.2.1 - 2.4) имеет вид

.

**Выводы**

*Коэффициент регрессии показывает, что с увеличением среднедушевого денежного дохода на 1 руб. среднедушевой оборот розничной торговли возрастает на 63,6 коп.*

1. Значимость параметров модели оценивается с помощью t-критерия Стьюдента (рис. 1.2.1, столбец t-статистика и рис. 2.4):
2. Для проверки значимости уравнения регрессии в целом используйте F-критерий Фишера (рис.2.2):

.

Критическое (табличное) значение F-критерия c и при доверительной вероятности 0,95 (α=0,05) равно 4,60 (для расчета используйте статистическую функцию **FРАСПОБР(α;k;n-k-1)**) (рис. 2.5).

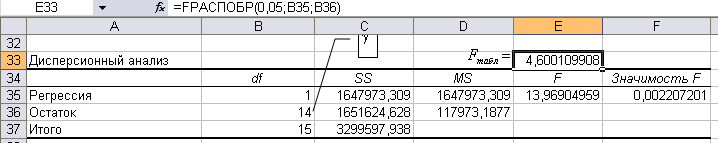


Рис. 2.5. Табличное значение F-критерия Фишера

**Выводы**

*Так как , уравнение парной регрессии с вероятностью 0,95 в целом статистически значимое.*

**Проверка выполнения предпосылок метода наименьших квадратов**

1. Проверка свойства случайности ряда остатков

**КРАТКАЯ СПРАВКА**

Для проверки свойства случайности ряда остатков можно использовать критерий поворотных точек (пиков).

Точка считается поворотной, если выполняются следующие условия:

.

Далее подсчитывается число поворотных точек.

Критерием случайности с 5%-ным уровнем значимости, т.е. с доверительной вероятностью 95%, является выполнение неравенства:

.

Квадратные скобки означают, что берется целая часть числа, заключенного в скобки.

Если неравенство выполняется, то модель считается адекватной.

Число поворотных точек *(p)* равно 8 (рис. 2.6).

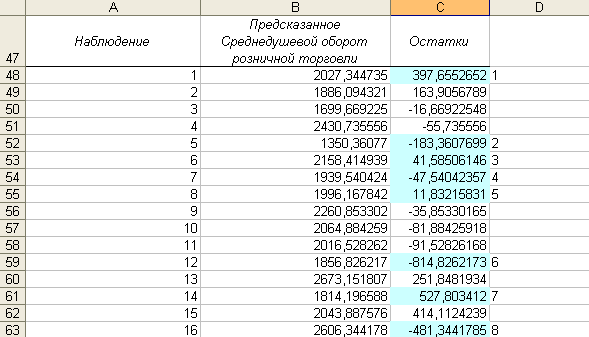


Рис. 2.6. Число поворотных точек

Критерием случайности с 5%-ным уровнем значимости, т.е. с доверительной вероятностью 95%, является выполнение неравенства:

.

Для расчета выражения (рис. 2.7) воспользуйтесь математической функцией

**ОТБР(число; число\_разрядов)**,

которая отбрасывает дробную часть числа, так что остается целое число,

где число – усекаемое число, число\_разрядов – число, определяющее точность усечения, по умолчанию используется значение 0.

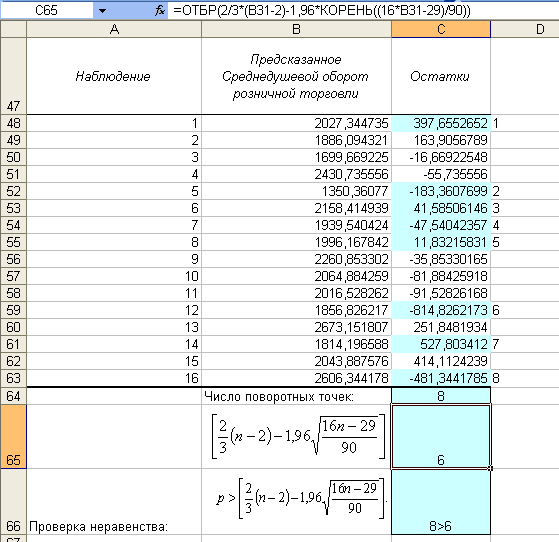


Рис. 2.7. Проверка свойства случайности ряда остатков

**ВЫВОДЫ**

*Неравенство выполняется (8 > 6), следовательно, модель может быть признана адекватной по критерию случайности.*

1. Проверка равенства математического ожидания остаточной последовательности нулю

**КРАТКАЯ СПРАВКА**

Для проверки равенства математического ожидания остаточной последовательности нулю вычисляется среднее значение ряда остатков:

.

Если , то считается, что модель не содержит постоянной систематической ошибки и адекватна по критерию нулевого среднего.

Если , то проверяется нулевая гипотеза о равенстве нулю математического ожидания. Для этого вычисляют t-критерий Стьюдента по формуле

,

где - стандартное отклонение остатков модели (стандартная ошибка).

Значение t-критерия сравнивают с табличным . Если выполняется неравенство , то модель неадекватна по данному критерию.

Вычислите среднее значение ряда остатков (рис. 2.8, столбец **Остатки**). Для вычисления среднего значения воспользуйтесь функцией **СРЗНАЧ()**:

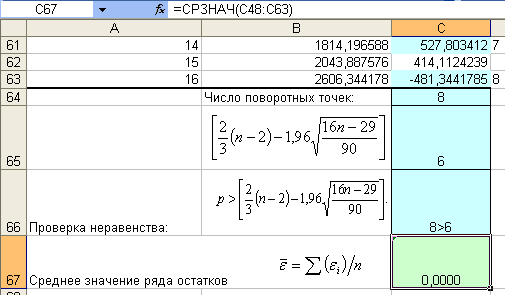


Рис. 2.8. Вычисление среднего значения ряда остатков

**ВЫВОДЫ**

*Так как , то модель не содержит постоянной систематической ошибки и адекватна по критерию нулевого среднего.*

1. Проверка свойства гомоскедастичности

**КРАТКАЯ СПРАВКА**

Дисперсия уровней ряда остатков должна быть одинаковой для всех значений *xi*  (свойствогомоскедастичности). Если это условие не соблюдается, то имеет место гетероскедастичность.

Для оценки гетероскедастичности при малом объеме выборки можно использовать метод Гольдфельда-Квандта, суть которого заключается в том, то необходимо:

1. расположить значения переменной *xi* в порядке возрастания;
2. разделить совокупность упорядоченных наблюдений на две группы c числом наблюдений *n1*каждая;
3. по каждой группе наблюдений построить уравнение регрессии;
4. определить остаточные суммы квадратов для первой и второй групп по формулам: ;
5. рассчитать критерии или (в числителе должна быть большая сумма квадратов);

При выполнении нулевой гипотезы о гомоскедастичности критерий будет удовлетворять F–критерию со степенями свободы и для каждой остаточной суммы квадратов (где *n1*– число наблюдений в группе; *m* – число оцениваемых параметров в уравнении регрессии).

Чем больше величина превышает табличное значение F–критерия, тем больше нарушена предпосылка о равенстве дисперсий остаточных величин.

Скопируйте исходную таблицу (табл. 2.1), например, в ячейки *B70:D86*.

Значения факторного признака *xi* расположите в порядке возрастания (**Данные /Сортировка)** (рис. 2.9).

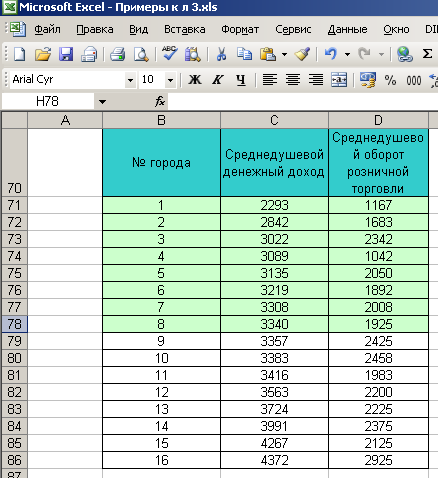


Рис. 2.9. Отсортированные в порядке возрастания по среднедушевому денежному доходу данные

Совокупность наблюдений разделите на две группы: первая группа – наблюдения с 1 по 8; вторая – с 9 по 16. Число наблюдений в каждой группе *n1*=8.

Для каждой группы с помощью команды **Данные/ Анализ данных/ Регрессия** определите параметры уравнений регрессий и остаточные суммы квадратов (рис. 2.10, табл. 2.2).

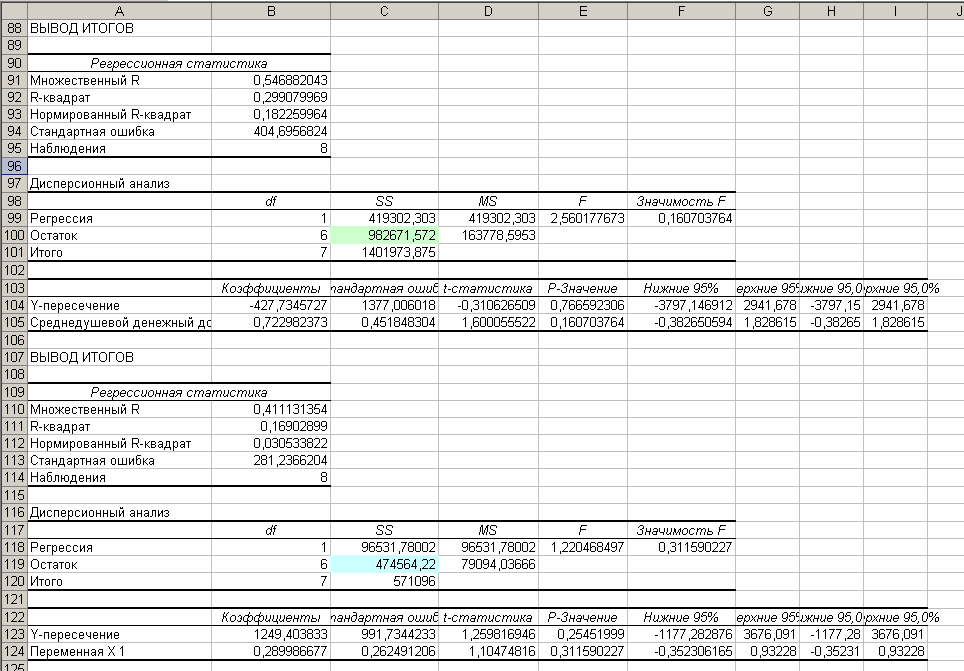


Рис. 2.10. Определение параметров уравнений регрессий и остаточных сумм квадратов

Таблица 2.2

Расчетные значения

|  |  |  |
| --- | --- | --- |
| Группа | Уравнение регрессии | Остаток |
| 1 |  |  |
| 2 |  |  |

Вычисляется расчетный критерий (рис.2.11):

.

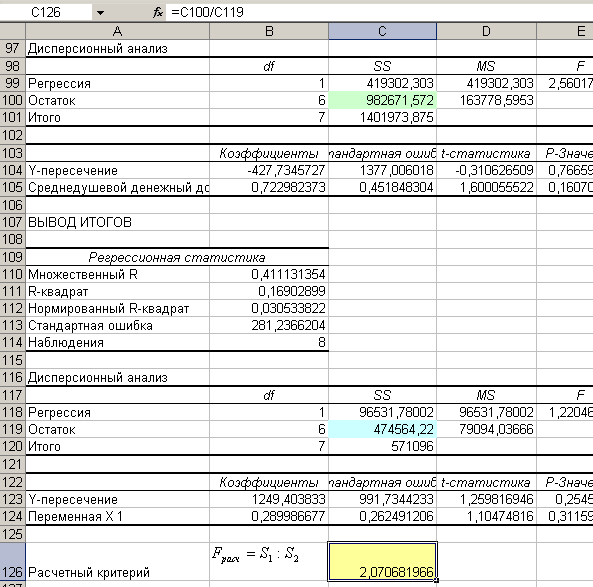


Рис. 2.11. Определение

Табличное значение F-критерия c и при доверительной вероятности 0,95 (α=0,05) равно 4,2838 (для расчета используйте статистическую функцию **FРАСПОБР(α;n1-m;n-n1-m)***)* (рис.1.12).

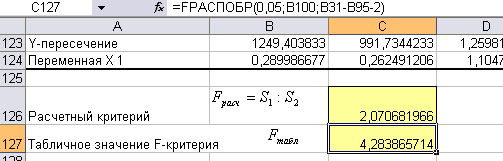


Рис. 2.12. Определение

**ВЫВОДЫ**

*Величина табличного значения F-критерия превышает , следовательно, свойство гомоскедастичности выполняется.*

1. Проверка независимости последовательности остатков (отсутствие автокорреляции) с помощью d-критерия Дарбина-Уотсона

**КРАТКАЯ СПРАВКА**

Проверка независимости последовательности остатков (отсутствие автокорреляции) осуществляют с помощью d-критерия Дарбина-Уотсона. Он определяется по формуле:

.

Расчетное значение критерия сравнивается с нижним *dL* и верхним *dU* критическими значениями статистики Дарбина-Уотсона.



Возможны следующие случаи:

1. если *d < dL*,то гипотеза о независимости остатков отвергается и модель признается неадекватной по критерию независимости остатков;
2. если *dL < d < dU* (включая сами эти значения), то считается, что нет достаточных оснований сделать тот или иной вывод и нужно использовать дополнительный критерий;
3. если *dU < d < 2*, то гипотеза о независимости остатков принимается и модель признается адекватной по данному критерию;
4. если *d > 2*, то это свидетельствует об отрицательной автокорреляции остатков.

В этом случае расчетное значение критерия необходимо преобразовать по формуле *d′=4-d* и сравнить с критическим значением *d′* , а не *d.*

**Значения статистик Дарбина-Уотсона**  **при 5%-ном уровне значимости**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | |  | |  | |  | |  | |
|  |  |  |  |  |  |  |  |  |  |
| 6 | 0,61 | 1,40 |  |  |  |  |  |  |  |  |
| 7 | 0,70 | 1,36 | 0,47 | 1,90 |  |  |  |  |  |  |
| 8 | 0,76 | 1,33 | 0,56 | 1,78 | 0,37 | 2,29 |  |  |  |  |
| 9 | 0,82 | 1,32 | 0,63 | 1,70 | 0,46 | 2,13 |  |  |  |  |
| 10 | 0,88 | 1,32 | 0,70 | 1,64 | 0,53 | 2,02 |  |  |  |  |
| 11 | 0,93 | 1,32 | 0,66 | 1,60 | 0,60 | 1,93 |  |  |  |  |
| 12 | 0,97 | 1,33 | 0,81 | 1,58 | 0,66 | 1,86 |  |  |  |  |
| 13 | 1,01 | 1,34 | 0,86 | 1,56 | 0,72 | 1,82 |  |  |  |  |
| 14 | 1,05 | 1,35 | 0,91 | 1,55 | 0,77 | 1,78 |  |  |  |  |
| 15 | 1,08 | 1,36 | 0,95 | 1,54 | 0,82 | 1,75 | 0,69 | 1,97 | 0,56 | 2,21 |
| 16 | 1,10 | 1,37 | 0,98 | 1,54 | 0,86 | 1,73 | 0,74 | 1,93 | 0,62 | 2,15 |
| 17 | 1,13 | 1,38 | 1,02 | 1,54 | 0,90 | 1,71 | 0,78 | 1,90 | 0,67 | 2,10 |
| 18 | 1,16 | 1,39 | 1,05 | 1,53 | 0,93 | 1,69 | 0,82 | 1,87 | 0,71 | 2,06 |
| 19 | 1,18 | 1,40 | 1,08 | 1,53 | 0,97 | 1,68 | 0,85 | 1,85 | 0,75 | 2,02 |
| 20 | 1,20 | 1,41 | 1,10 | 1,54 | 1,00 | 1,68 | 0,90 | 1,83 | 0,79 | 1,99 |
| 21 | 1,22 | 1,42 | 1,13 | 1,54 | 1,03 | 1,67 | 0,93 | 1,81 | 0,83 | 1,96 |
| 22 | 1,24 | 1,43 | 1,15 | 1,54 | 1,05 | 1,66 | 0,96 | 1,80 | 0,86 | 1,94 |
| 23 | 1,26 | 1,44 | 1,17 | 1,54 | 1,08 | 1,66 | 0,99 | 1,79 | 0,90 | 1,92 |
| 24 | 1,27 | 1,45 | 1,19 | 1,55 | 1,10 | 1,66 | 1,01 | 1,78 | 0,93 | 1,99 |
| 25 | 1,29 | 1,45 | 1,21 | 1,55 | 1,12 | 1,66 | 1,04 | 1,77 | 0,95 | 1,89 |
| 26 | 1,30 | 1,46 | 1,22 | 1,55 | 1,14 | 1,65 | 1,06 | 1,76 | 0,98 | 1,88 |
| 27 | 1,32 | 1,47 | 1,24 | 1,56 | 1,16 | 1,65 | 1,08 | 1,76 | 1,01 | 1,86 |
| 28 | 1,33 | 1,48 | 1,26 | 1,56 | 1,18 | 1,65 | 1,10 | 1,75 | 1,03 | 1,85 |
| 29 | 1,34 | 1,48 | 1,27 | 1,56 | 1,20 | 1,65 | 1,12 | 1,74 | 1,05 | 1,84 |
| 30 | 1,35 | 1,49 | 1,28 | 1,57 | 1,21 | 1,65 | 1,14 | 1,74 | 1,07 | 1,83 |

Скопируйте таблицу ВЫВОД ОСТАТКА (рис. 2.2.2), например, в ячейки *A129:C145.*

Добавьте столбцы, вычисляющие , , (рис.1.13). Посчитайте и (рис. 2.13). Рассчитайте значение d-критерия (рис. 2.13).

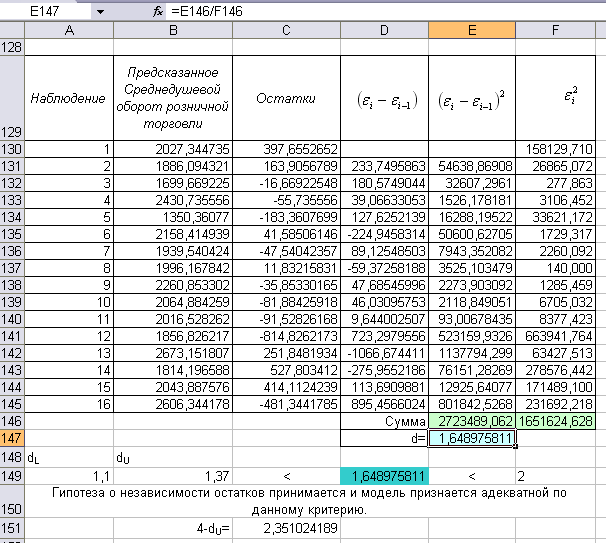


Рис. 2.13. Расчет d-критерия Дарбина-Уотсона

Расчетное значение критерия () сравнивается с нижним *dL* и верхним *dU* критическими значениями статистики Дарбина-Уотсона. При *n = 16* и уровне значимости 5% dL=1,10, dU=1,37.

**ВЫВОДЫ**

*Поскольку dU < d < 2, то гипотеза о независимости остатков принимается и модель признается адекватной по данному критерию.*

1. Проверка соответствия распределения остаточной последовательности нормальному закону распределения с помощью R/S-критерия

**КРАТКАЯ СПРАВКА**

Проверку соответствия распределения остаточной последовательности нормальному закону распределения можно осуществить с помощью *R/S*-критерия, который определяется по формуле:

,

где – стандартное отклонение остатков модели (стандартная ошибка).

Расчетное значение *R/S*-критерия сравнивается с табличными значениями (нижней и верхней границами данного отношения), и если значение не попадает в интервал между критическими границами, то с заданным уровнем значимости гипотеза о нормальности распределения отвергается; в противном случае гипотеза принимается.

**Критические границы отношения R/S**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n | Нижние границы | | Верхние границы | |
| α=0,05 | α=0,01 | α=0,05 | α =0,01 |
| 8 | 2,50 | 2,59 | 3,308 | 3,399 |
| 10 | 2,67 | 2,76 | 3,57 | 3,685 |
| 12 | 2,80 | 2,90 | 3,78 | 3,91 |
| 14 | 2,92 | 3,02 | 3,95 | 4,09 |
| 16 | 3,01 | 3,12 | 4,09 | 4,24 |
| 18 | 3,10 | 3,21 | 4,21 | 4,37 |
| 20 | 3,18 | 3,29 | 4,32 | 4,49 |
| 25 | 3,34 | 3,45 | 4,53 | 4,71 |
| 30 | 3,47 | 3,59 | 4,70 | 4,89 |
| 35 | 3,58 | 3,70 | 4,84 | 5,04 |
| 40 | 3,67 | 3,79 | 4,96 | 5,16 |
| 45 | 3,75 | 3,88 | 5,06 | 5,26 |
| 50 | 3,83 | 3,95 | 5,14 | 5,35 |

Рассчитайте . Для этого воспользуйтесь статистическими функциями **МИН()** и **МАКС()** (в нашем случае =МИН(C130:C145) и =МАКС(C130:C145)) (рис. 2.14): .

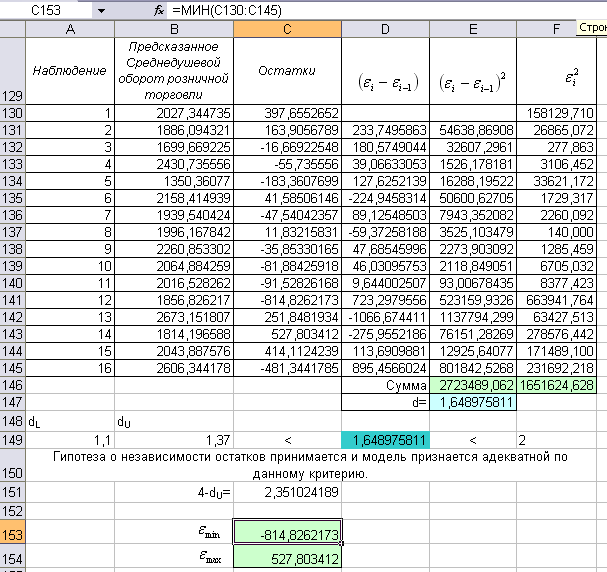


Рис. 2.14. Расчет значений

Стандартное отклонение остатков модели (стандартная ошибка) определена с помощью инструмента анализа **Регрессия** *Sε=343,4722 (*рис. 2.2.1).

Проверка соответствия распределения остаточной последовательности нормальному закону распределения осуществляется по формуле

(рис. 2.15).

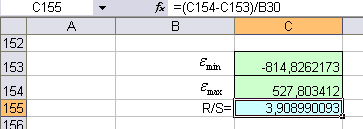


Рис. 2.15. R/S-критерий

Нижняя и верхняя границы отношения при уровне значимости α=0,05 равны соответственно 3,01 и 4,09.

**ВЫВОДЫ**

*Расчетное значение отношения попадает в интервал между критическими границами (3,01<3,9089<4,09), следовательно, с заданным уровнем значимости гипотеза о нормальности распределения принимается.*

1. Оценка точности модели

**КРАТКАЯ СПРАВКА**

Средняя относительная ошибка аппроксимации определяется по формуле:

.

Допустимый предел значений составляет не более 8-10%.

Добавьте к таблице, показанной на рисунке 2.13, столбец вычисляющий . Для определения абсолютной величины воспользуйтесь функцией **ABS(),** которая возвращает модуль (абсолютную величину) числа (рис. 2.16).

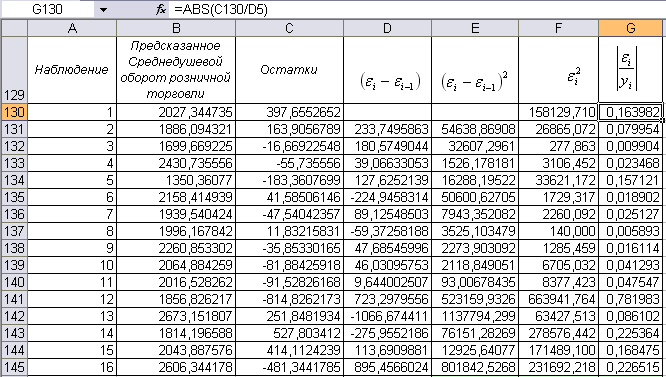


Рис. 2.16. Расчет

В качестве показателя точности модели используется средняя относительная ошибка аппроксимации (функция **СРЗНАЧ())** (рис. 2.17):

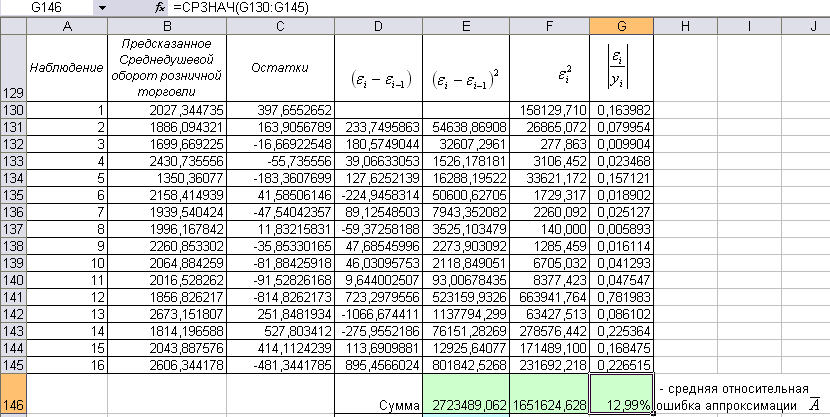


Рис. 2.17. Оценка точности модели

**ВЫВОДЫ**

*Уровень точности модели можно признать приемлемым.*

## Задания для самостоятельной работы

**Задача 2.1.**

для прогноза возможного объема экспорта на основе ВНП предложено использовать линейную регрессионную модель. При этом используются данные за 1989-1998 годы.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Годы | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
| ВНП | 1000 | 1090 | 1150 | 1230 | 1300 | 1360 | 1400 | 1470 | 1500 | 1580 |
| Экспорт | 190 | 220 | 240 | 240 | 260 | 250 | 280 | 290 | 310 | 350 |

**ТРЕБУЕТСя**

1. Построить поле корреляции.
2. Построить уравнение регрессии, оценить его адекватность, сделать выводы.
3. Проверить выполнение предпосылок метода наименьших квадратов.
4. Оценить точность модели.

**Задача 2.2.**

Имеется информация за 7 лет относительно среднего дохода и среднего потребления (млн. руб.):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Годы | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 |
| Доход | 14,56 | 15,70 | 16,30 | 18,50 | 20,34 | 21,70 | 23,50 |
| Потребление | 12,00 | 12,70 | 13,00 | 15,50 | 16,70 | 17,30 | 20,00 |

**ТРЕБУЕТСЯ**

1. Построить поле корреляции.
2. Построить уравнение регрессии, оценить его адекватность, сделать выводы.
3. Проверить выполнение предпосылок метода наименьших квадратов.
4. Оценить точность модели.

**Задача 2.3.**

Проводится анализ взаимосвязи количества населения (РОР) и количества практикующих врачей (MED).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Годы | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 |
| РОР, млн.чел. | 10,0 | 10,3 | 10,4 | 10,55 | 10,6 | 10,7 | 10,75 | 10,9 | 10,9 | 11,0 |
| MED, тыс. чел. | 12,1 | 12,6 | 13,0 | 13,8 | 14,9 | 16,0 | 18,0 | 20,0 | 21,0 | 22,0 |

**ТРЕБУЕТСЯ**

1. Построить поле корреляции.
2. Построить уравнение регрессии, оценить его адекватность, сделать выводы.
3. Проверить выполнение предпосылок метода наименьших квадратов.
4. Оценить точность модели.