# Лабораторная работа 4. Множественная линейная регрессия и корреляция: смысл и оценка параметров

**КРАТКАЯ СПРАВКА**

*Классическая (нормальная) модель линейной множественной регрессии (КЛММР)* имеет вид:

, (4.1)

где – коэффициенты регрессии;

*εi* – случайные остатки*.*

Коэффициент регрессии показывает, на какую величину в среднем изменится результативный признак , если переменную увеличить на единицу измерения при фиксированном (постоянном) значении других факторов, входящих в уравнение регрессии.

Для оценки параметров уравнения множественной регрессии применяют *метод наименьших квадратов (МНК).*

Переписывая выражение (4.1) в виде системы уравнений для различных значений *i=1,2,…,n*, их можно представить в матричном виде

(4.2)

где – вектор значений зависимой переменной размерности *(n×1)*;

– матрица значений независимых переменных , размерность матрицы *X* равна *(n×(k+1)).* Первыйстолбец является единичным, так как в уравнении регрессии умножается на единицу;

– подлежащий оцениванию вектор неизвестных параметров размерности *((k+1)×1);*

– вектор случайных отклонений размерности *(n×1).*

Формула для вычисления параметров регрессионного уравнения по методу наименьших квадратов (4.3):

(4.3)

где – транспонированная матрица ;

– обратная матрица.

## Решение типовых задач

**Пример 4.1.**

Изучается влияние стоимости основных и оборотных средств на величину валового дохода торговых предприятий. Для этого по 12 торговым предприятиям были получены данные, представленные в таблице 4.1.

Таблица 4.1.

|  |  |  |  |
| --- | --- | --- | --- |
| Номер предприятия | Валовой доход за год, млн. руб. | Среднегодовая стоимость,  млн. руб. | |
| основных фондов | оборотных средств |
| 1 | 203 | 118 | 105 |
| 2 | 63 | 28 | 56 |
| 3 | 45 | 17 | 54 |
| 4 | 113 | 50 | 63 |
| 5 | 121 | 56 | 28 |
| 6 | 88 | 102 | 50 |
| 7 | 110 | 116 | 54 |
| 8 | 56 | 124 | 42 |
| 9 | 80 | 114 | 36 |
| 10 | 237 | 154 | 106 |
| 11 | 160 | 115 | 88 |
| 12 | 75 | 98 | 46 |

**ТРЕБУЕТСЯ**

1. Построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.

2. На уровне значимости α = 0,05 оценить значимость коэффициентов регрессии и для значимых коэффициентов определить доверительные интервалы.

3. Рассчитать частные коэффициенты эластичности.

4. Определить стандартизованные коэффициенты регрессии. Сделать вывод о силе связи результата и факторов.

5. Определить парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделать выводы.

6. Дать оценку полученного уравнения на основе коэффициента детерминации и общего *F*-критерия Фишера.

7. Проверить полученные результаты с помощью программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.

8. Постройте уравнение регрессии со статистически значимыми факторами.

**Решение**

Решение задания будем выполнять с помощью средств MS Excel 2007.

Обозначим переменные: *X*1 – среднегодовая стоимость основных фондов; *X*2 – среднегодовая стоимость оборотных средств; *Y* – валовой доход за год.

1. Создайте рабочую книгу *ПРИМЕР4\_1.xlsx*. На листе *Задание\_1* рабочей книги *ПРИМЕР4\_1.xlsx* постройте таблицу 4.1, несколько изменив ее внешний вид (рисунок 4.1).

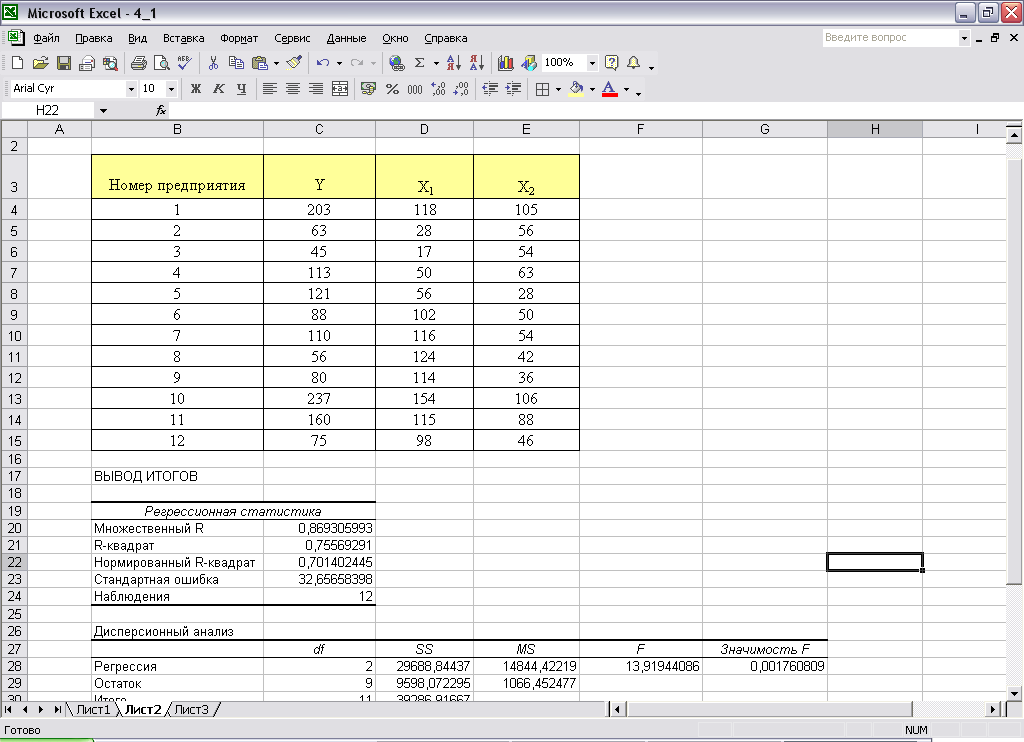


Рис. 4.1. Исходные данные

Для определения выборочных коэффициентов уравнения регрессии строим матрицу *X* объясняющих переменных размером (*12×3)*, в которой первый столбец с единичными элементами соответствует умножению коэффициента *b0* на единицу (рис. 4.2):

.

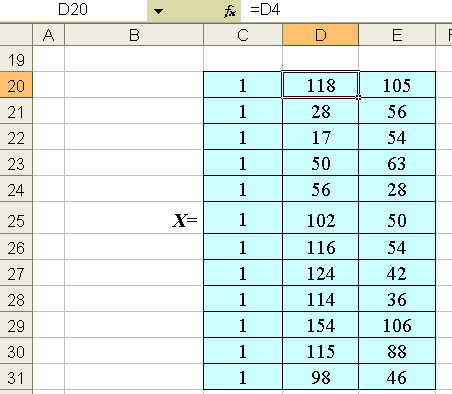


Рис. 4.2. Матрица *X*

Для построения транспонированной матрицы используем функцию **ТРАНСП()** из категории **Ссылки и массивы:**

* выделяем все ячейки, в которых будет содержаться матрица *XT* (*3×12*);
* вызываем **Мастер функций**, в категории **Ссылки и массивы** используем стандартную функцию **ТРАНСП(X),** в поле **Массив** вводим диапазон ячеек, который необходимо транспонировать (в нашем примере – это *C20:E31)***,** и контролируем результат в ее окне (первый элемент массива *XT* ) *x11=1*;
* для получения массива результатов сначала нажмите на клавишу **<F2>,** а затем – на комбинацию клавиш **<Ctrl+Shift+Enter>** (в выделенных ячейках появятся результаты вычислений) (рис. 4.3).

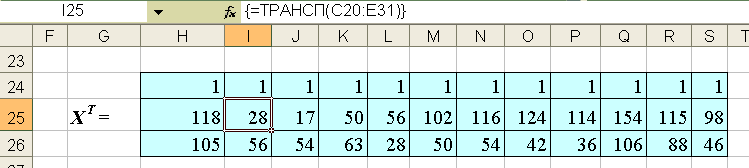


Рис. 4.3. Матрица *XT*

Определяем произведения матриц *XTX* с помощью стандартной функции **МУМНОЖ(*XT*; *X*)** из категории **Математические**:

* выделяем все ячейки, в которых будет содержаться матрица *XTX* (*3×3*);
* вызываем **Мастер функций**, в категории **Математические** используем стандартную функцию **МУМНОЖ(XT;X),** в поле **Массив1** вводим диапазон ячеек с матрицей ***XT***, в поле **Массив2** вводим диапазон ячеек с матрицей ***X*,** и контролируем результат в ее окне (первый элемент массива *XT X* равен 12) (рис.4.4);

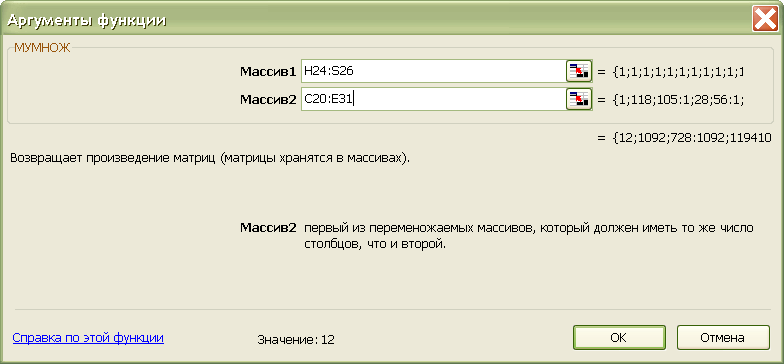


Рис. 4.4. Функция МУМНОЖ(XT;X)

* для получения массива результатов сначала нажмите на клавишу **<F2>,** а затем – на комбинацию клавиш **<Ctrl+Shift+Enter>** (в выделенных ячейках появятся результаты вычислений) (рис. 4.5).

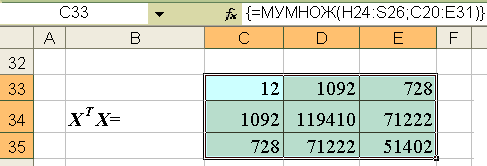


Рис. 4.5. Результат произведения матриц *XT* и *X*

Убеждаемся, что полученная матрица является симметричной.

Аналогичным образом находим произведение *XTY:*

* выделяем все ячейки, в которых будет содержаться матрица *XTY* (*3×1*);
* вызываем **Мастер функций**, в категории **Математические** используем стандартную функцию **МУМНОЖ(XT;Y),** в поле **Массив1** вводим диапазон ячеек с матрицей ***XT***, в поле **Массив2** вводим диапазон ячеек с вектром ***Y*,** и контролируем результат в ее окне (первый элемент массива *XT X* равен 1351) (рис.4.6);

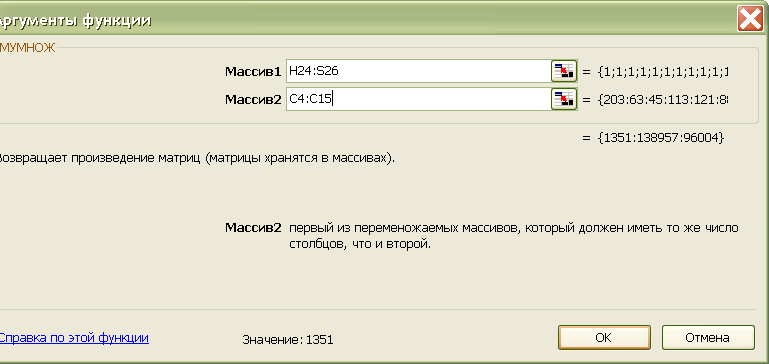


Рис. 4.6. Функция МУМНОЖ(XT;Y)

* для получения массива результатов сначала нажмите на клавишу **<F2>,** а затем – на комбинацию клавиш **<Ctrl+Shift+Enter>** (в выделенных ячейках появятся результаты вычислений) (рис. 4.7).

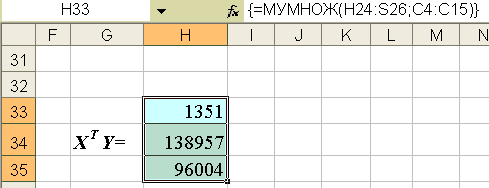


Рис. 4.7. Результат произведения матриц *XT* и *Y*

Рассчитываем обратную матрицу (*XTX)-1*, предварительно выделяя для вывода результата массив размером *3×3* и используя стандартную математическую функцию **МОБР(XTX)** (рис. 4.8):

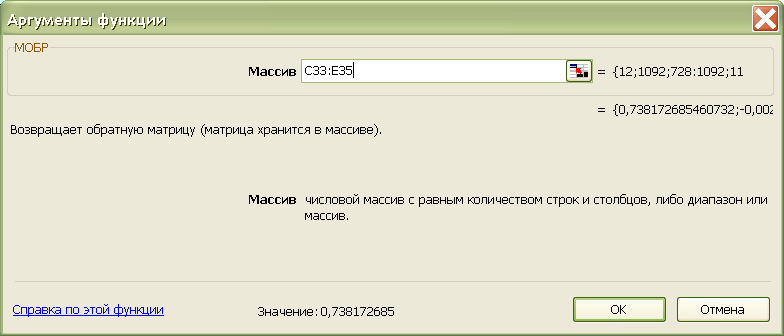


Рис. 4.8. Функция МОБР(XTX)

Для получения массива результатов сначала нажмите на клавишу **<F2>,** а затем – на комбинацию клавиш **<Ctrl+Shift+Enter>** (в выделенных ячейках появятся результаты вычислений) (рис. 4.9).

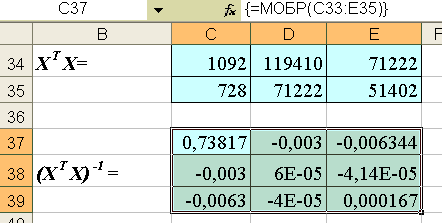


Рис. 4.9. Результат вычислений обратной матрицы (*XTX)-1*

С помощью математической функции **МУМНОЖ((XTX)-1; XTY)** по

формуле (4.3) определяем вектор выборочных оценок коэффициентов, предварительно выделив для вывода результата столбец из трех ячеек. Для получения массива результатов сначала нажмите на клавишу **<F2>,** а затем – на комбинацию клавиш **<Ctrl+Shift+Enter>** (в выделенных ячейках появятся результаты вычислений) (рис.4.10):

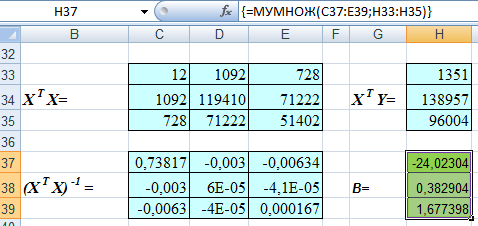


Рис. 4.10. Расчет коэффициентов регрессии

Таким образом, уравнение линейной множественной регрессии имеет вид (4.4): (4.4)

**ВЫВОДЫ**

*Из уравнения регрессии (4.4) видим, что коэффициенты при переменных X1 и X2 положительны, т.е. рост среднегодовой стоимости основных фондов, и оборотных средств увеличивает валовой доход. Причем, при прочих равных условиях:*

* *увеличение среднегодовой стоимости основных фондов на 1 ед. приводит к росту валового дохода на 0,383 ед.;*
* *увеличение среднегодовой стоимости оборотных средств на 1 ед. приводит к росту валового дохода на 1,677 ед.*

1. Используя выборочное уравнение регрессии (4.4), находим значения , *i=1, 2, …, n* и сводим их в 5-й столбец таблицы 4.1 (рис. 4.11).

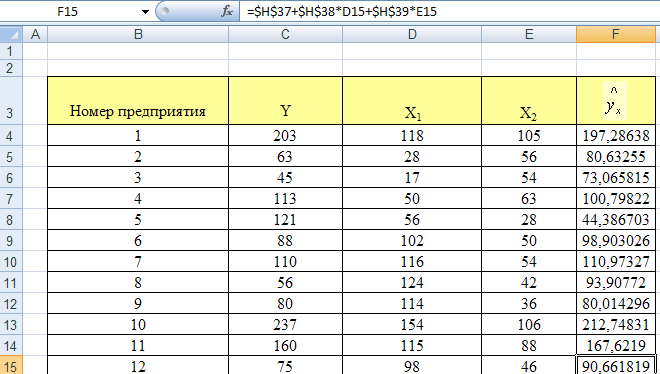
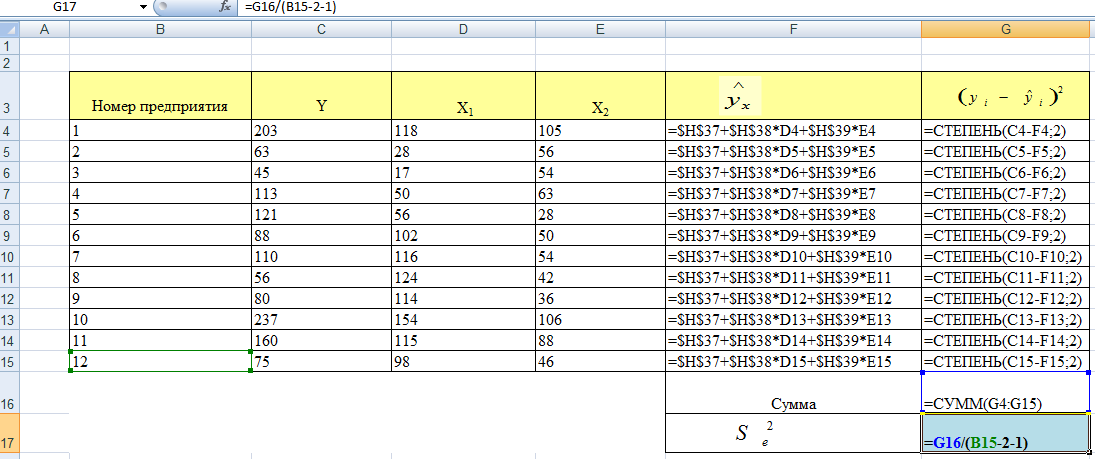


Рис. 4.10. Расчет

Вычисляем значения и рассчитываем несмещенную оценку дисперсии ошибок (несмещенную выборочную остаточную дисперсию) (рис. 4.11):

,

где *n* − число наблюдений, *k* − число регрессоров (в нашем примере *k=2)*.



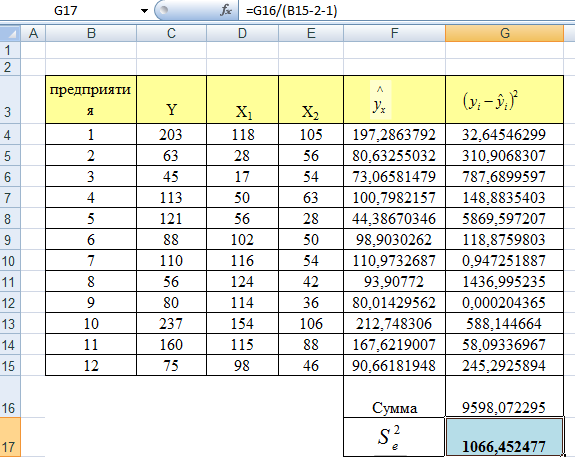
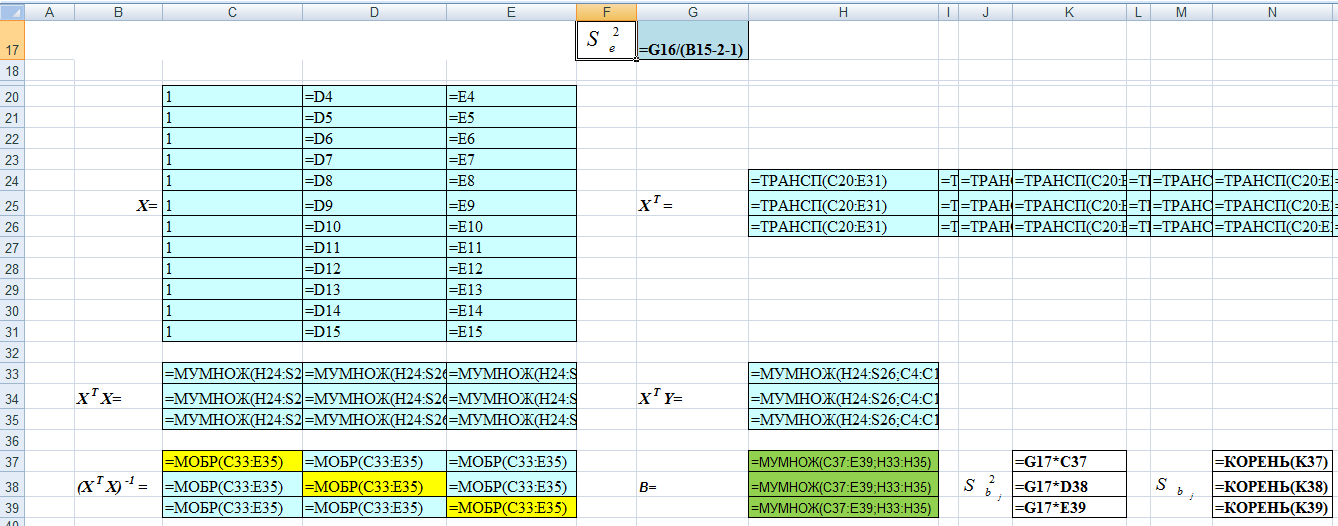


Рис. 4.11. Расчет

Находим векторы несмещенных оценок дисперсий и стандартных отклонений коэффициентов регрессии (диагональные элементы матрицы) (рис. 4.12):

,

.



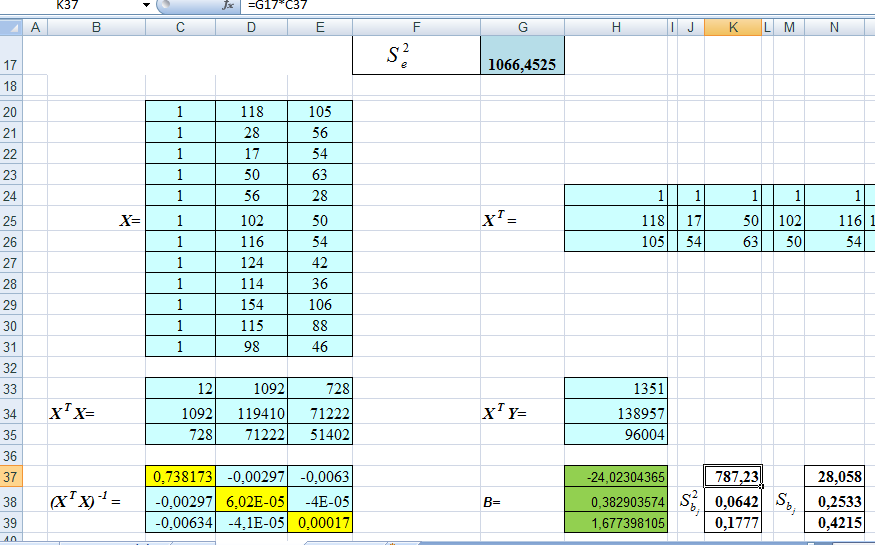
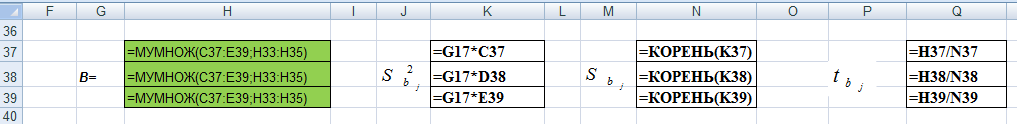


Рис. 4.12. Расчет и

Определяем вектор t−статистик критерия значимости коэффициентов регрессии (рис. 4.13):



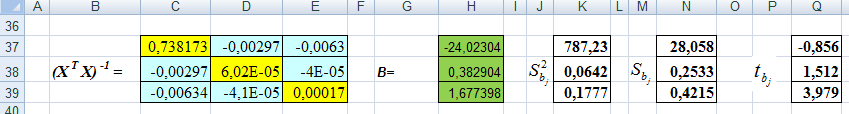


Рис. 4.13. Расчет

Для заданной доверительной вероятности *γ=0,95* (уровня значимости *α=0,05*) находим значение *tкр* с помощью стандартной статистической функции СТЬЮДРАСПОБР(α; *n-k-1*): *tкр=2,26* (рис. 4.14).

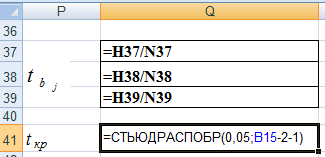
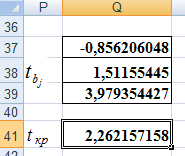
 

Рис. 4.14. Расчет *tкр*

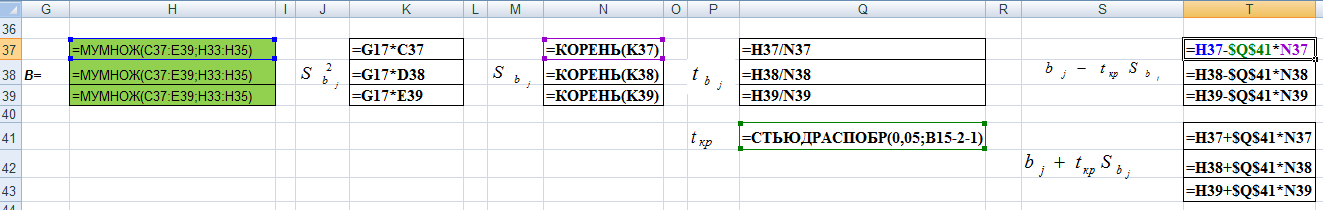
**ВЫВОДЫ**

*Так как и (0,8561<2,262 и 1,512<2,262), с доверительным уровнем 95% делаем вывод о том, что b0, b1 незначимы.*

*Так как (3,979>2,262), с доверительным уровнем 95% делаем вывод о том, что b2 значим.*

Определим векторы нижних и верхних границ доверительного интервала коэффициентов регрессии *bj* (рис. 4.15): .

Доверительные интервалы для значимого коэффициента: *0,724<b2<2,631.*



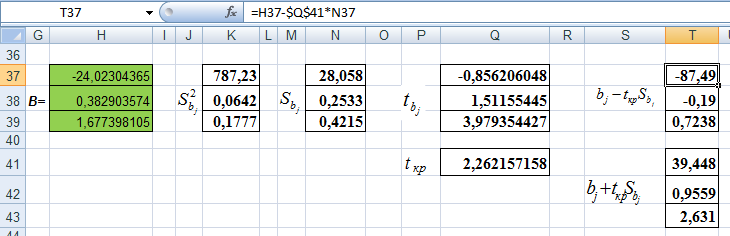
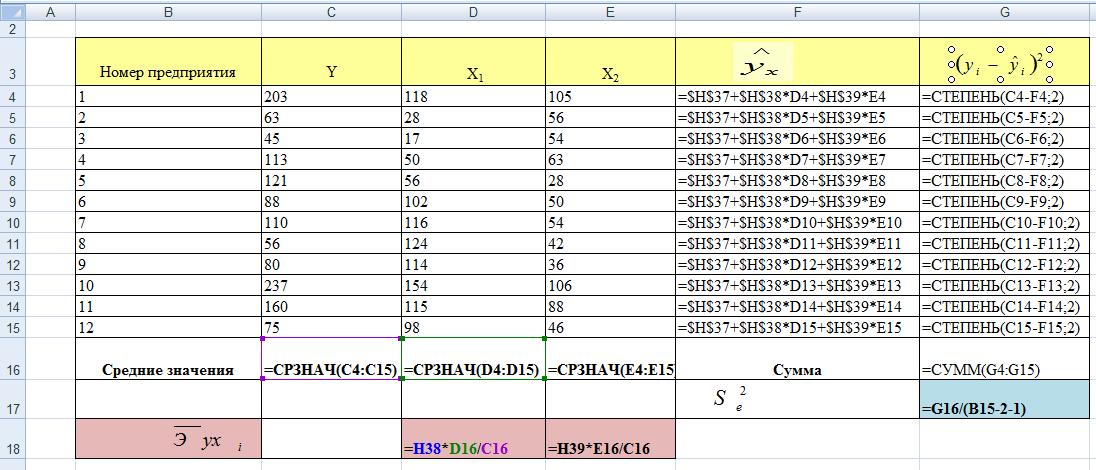


Рис. 4.15. Векторы нижних и верхних границ доверительного интервала коэффициентов регрессии *bj*

3. Определим частные коэффициенты эластичности. Для этого рассчитаем средние значения (рис. 4.16).



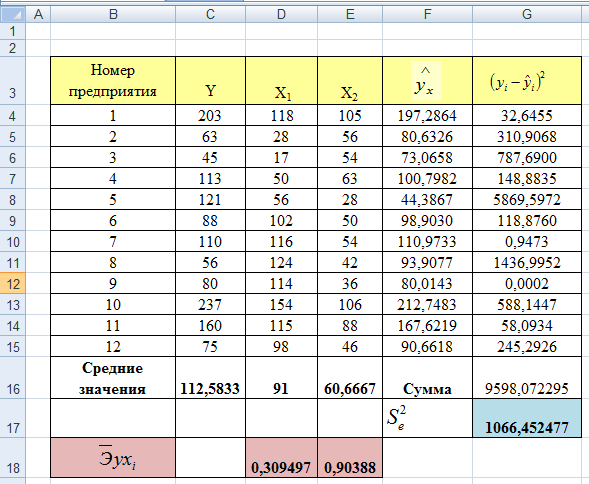


Рис. 4.16. Расчет частных коэффициентов эластичности

Частный коэффициент эластичности валового дохода относительно среднегодовой стоимости основных фондов равен:

,

где млн. руб. – среднее значение среднегодовой стоимости основных фондов; млн. руб. – среднее значение валового дохода.

Подставив рассчитанные значения в последнюю формулу, получим (рис. 4.16):

.

Частный коэффициент эластичности валового дохода относительно среднегодовой стоимости оборотных средств равен:

,

где млн. руб. – среднее значение среднегодовой стоимости основных фондов.

Подставив рассчитанные значения в последнюю формулу, получим (рис. 4.16):

.

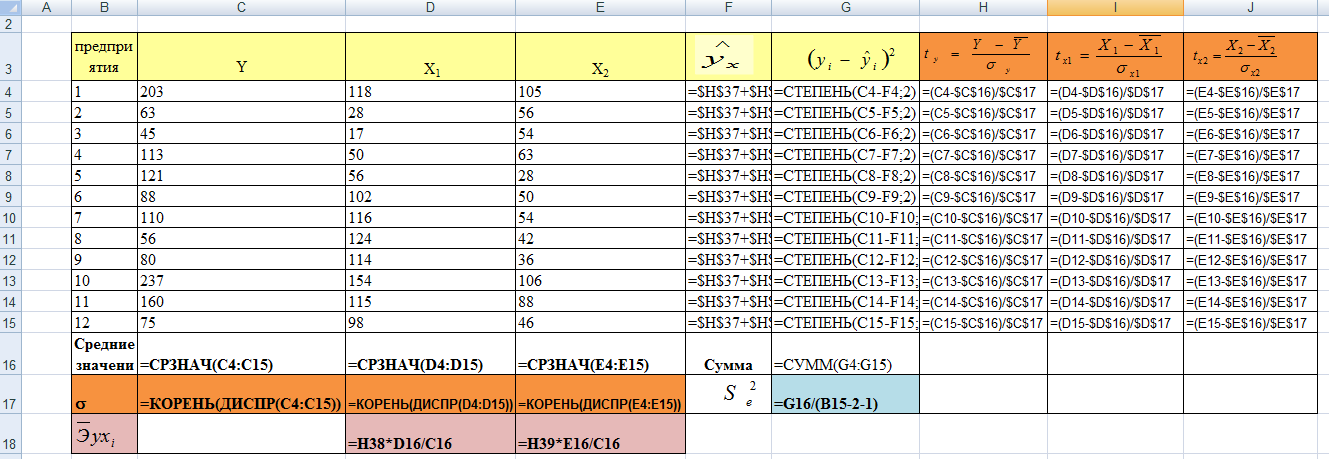
4. Определим стандартизованные коэффициенты регрессии.

Уравнение регрессии в стандартизованном виде выглядит следующим образом:

,

где , , .

Для определения стандартизованных коэффициентов регрессии *βi* добавьте столбцы для расчета , , (рис. 4.17).



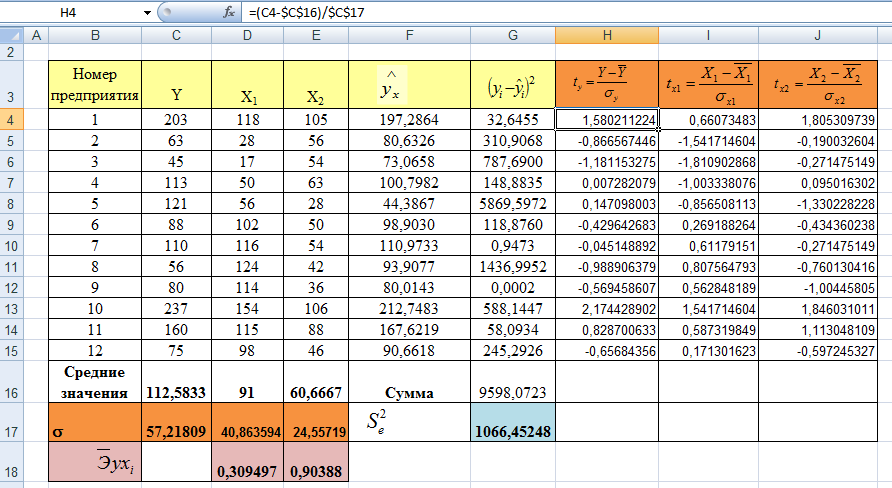


Рис. 4.17. Определение , ,

Затем выберите команду **Данные/ Анализ данных /Регрессия**. Заполните диалоговое окно **Регрессия** (рис. 4.18).

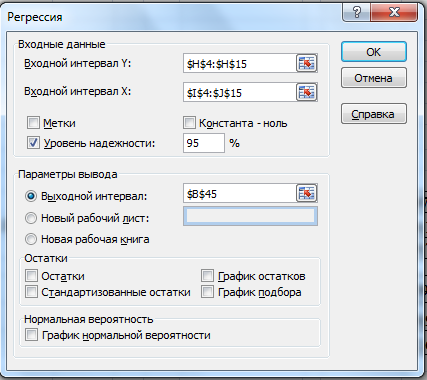


Рис. 4.18. Диалоговое окно **Регрессия**

Результаты расчета приведены на рисунке 4.19.

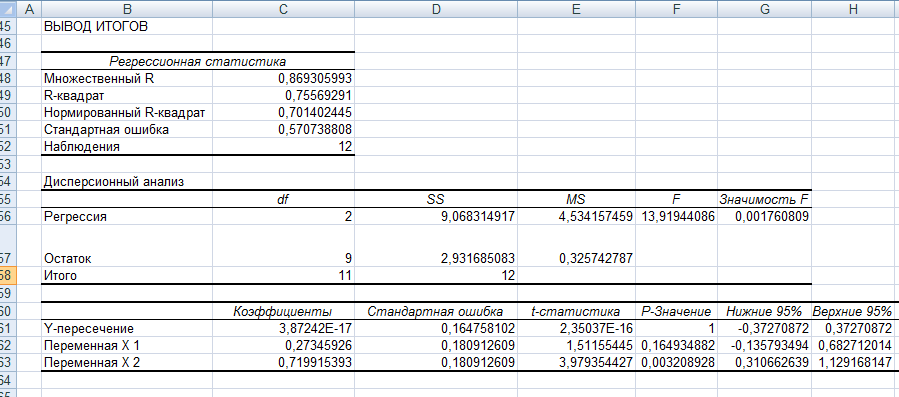


Рис. 4.19. Определение стандартизованных коэффициентов регрессии *βi*

Таким образом, стандартизованные коэффициенты регрессии равны:

; .

В «новых» переменных стандартизованное уравнение регрессии имеет вид:

.

**ВЫВОДЫ**

*Из уравнения видим, что сила связи результата с фактором X1 равна 0,273, а с фактором X2 равна 0,719.*

5. Определите парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделайте выводы.

Построим матрицу парных коэффициентов корреляции (**Данные/ Анализ данных/ Корреляция**). В диалоговое окно **Корреляция** введем необходимые данные (рис. 4.20).

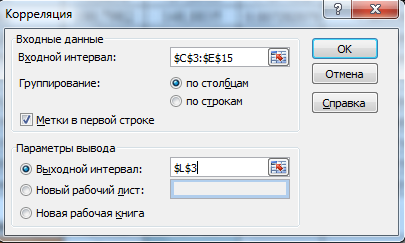


Рис. 4.20. Диалоговое окно Корреляция

Результаты представлены на рисунке 4.21.

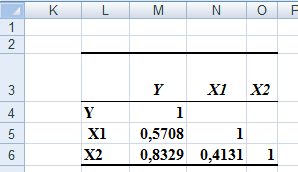
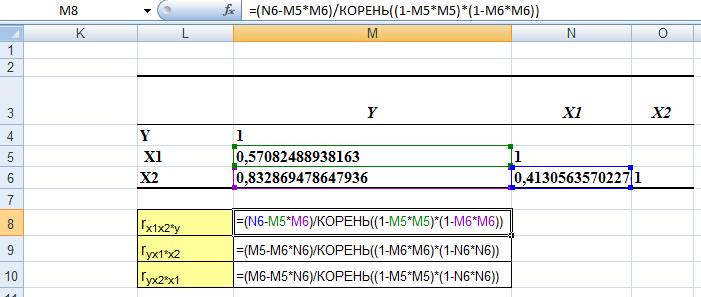


Рис. 4.21.Матрица парных коэффициентов корреляции

Частные коэффициенты корреляции определяются по формуле:

Подставив в последнюю формулу рассчитанные значения парной корреляции, получим (рис. 4.22): ; ; .



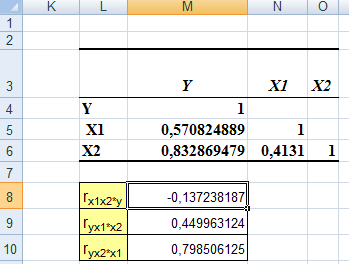


Рис. 4.22. Частные коэффициенты корреляции

Множественный коэффициент корреляции равен (рис. 4.23):

.

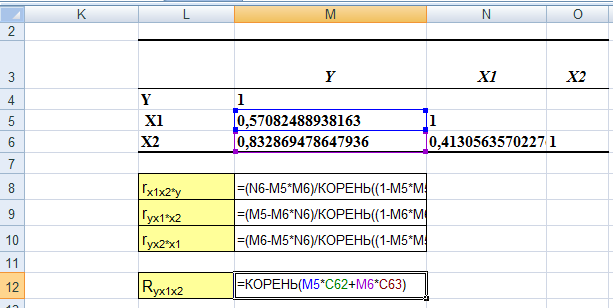
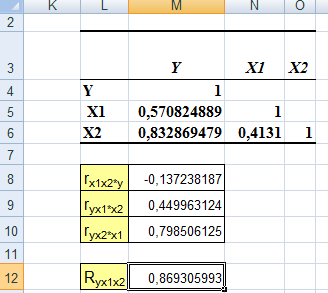
 

Рис. 4.23. Множественный коэффициент корреляции

**ВЫВОДЫ**

*После фиксирования одного из факторов частные коэффициенты корреляции , изменились незначительно, в то время как даже поменял знак.*

*Множественный коэффициент корреляции оказался достаточно высоким, что говорит об общей взаимосвязи между признаками.*

1. Дать оценку полученного уравнения на основе коэффициента детерминации и общего *F*-критерия Фишера.

**КРАТКАЯ СПРАВКА**

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации: .

Коэффициент детерминации показывает долю вариации результативного признака, находящегося под воздействием факторных признаков, т.е. определяет, какая доля вариации признака y учтена в модели и обусловлена влиянием на него факторов, включенных в модель. Чем ближе R2 к единице, тем выше качество модели.

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы: , где n – число наблюдений, k – число факторов.

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью F-критерия Фишера:

Из найденной матрицы (рис. 4.23) формируем матрицу межфакторной парной корреляции (рис. 4.24) (или используя инструмент КОРРЕЛЯЦИЯ (рис. 4.25)).

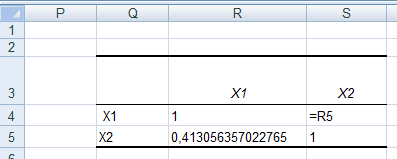
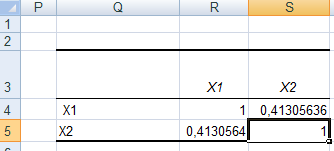
 

Рис. 4.24. Матрица межфакторной парной корреляции

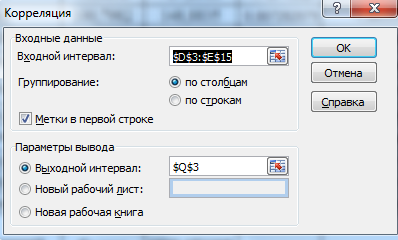


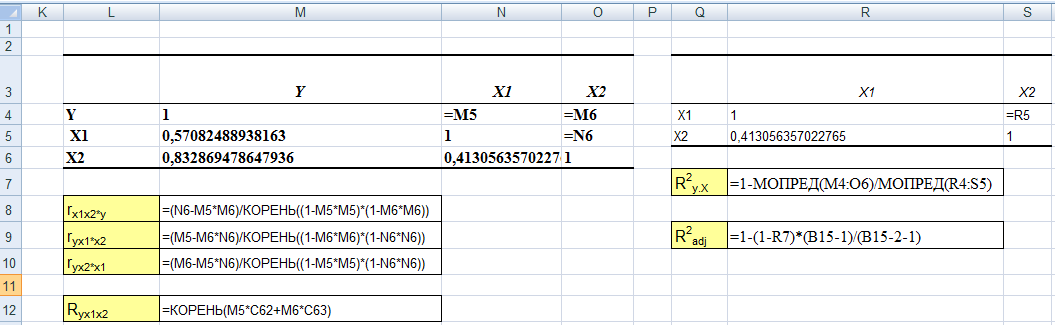
Рис. 4.25. Диалоговое окно Корреляция для определения

Используя математическую функцию МОПРЕД() с соответствующими массивами и в качестве аргументов, получаем (рис. 4.26):

,

**ВЫВОДЫ**

*Из полученного результата делаем вывод: 75,6% вариации валового дохода описывается стоимостью основных и оборотных средств.*



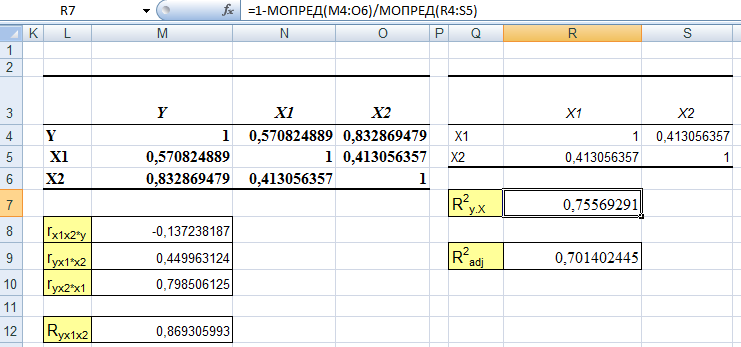


Рис. 4.26. Определение коэффициента детерминации и скорректированного коэффициент детерминации

Скорректированный коэффициент детерминации определяем по формуле (рис. 4.26): .

Рассчитываем значение *F*-статистики (рис. 4.27):

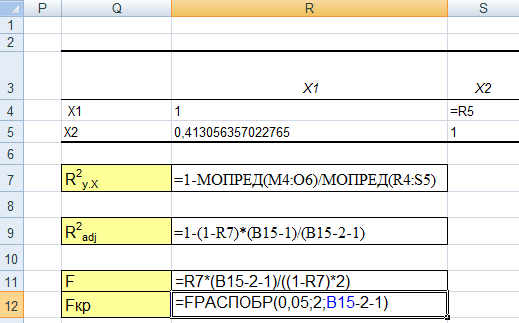
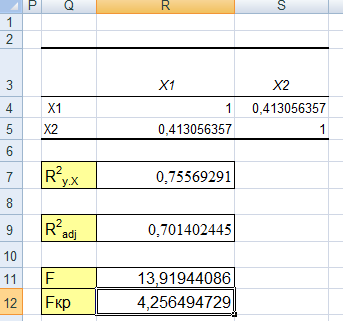
 

Рис. 4.27. Определение *F*-статистики

**ВЫВОДЫ**

*По F-критерию Фишера модель оказалась значимой на уровне значимости α = 0,05. Расчетное значение F-статистики равно: F = 13,919, табличное (критическое) Fкр=FРАСПОБР(0,05;k; n-k-1)=4,26 (F > Fкр с доверительным уровнем 0,95)* (рис. 4.27).

7. Для вычисления коэффициентов множественной регрессии по данным, представленным в таблице на рисунке 4.1, с помощью подпрограммы РЕГРЕССИЯ необходимо:

* Выбрать команду **Данные / Анализ данных /Регрессия**.
* Выбрать в раскрывающемся списке Анализ данных пункт Регрессия, ОК.
* Находясь в диалоговом окне Регрессия, заполнить поля (рис. 4.28):

в поле *Входной интервал Y:* ввести диапазон ячеек *C3:C15;*

в поле *Выходной интервал X:* ввести диапазон ячеек *D3:E15*;

установить флажок *Метки*;

установить флажок *Уровень надежности*;

установить переключатель *Параметры вывода* в положение *Входной интервал*: *B45*. ОК.

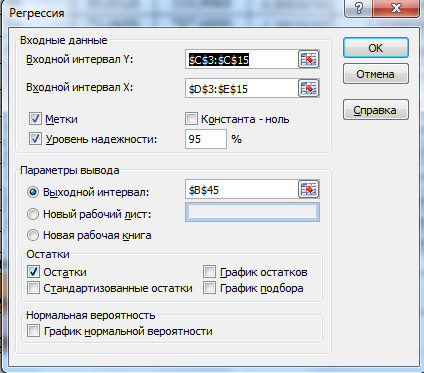


Рис. 4.28. Диалоговое окно подпрограммы РЕГРЕССИЯ

Результаты расчета приведены на рисунке 4.29.

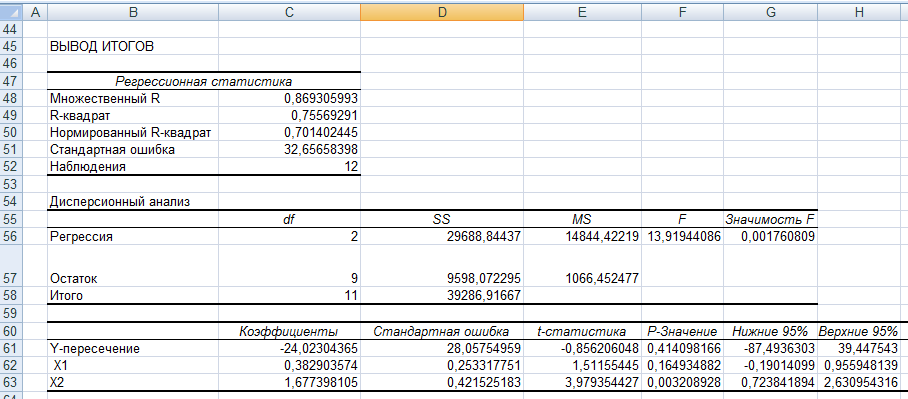


Рис. 4.29. Результаты расчета с использованием подпрограммы РЕГРЕССИЯ

8. Так как статистически значимым является коэффициент *b2*, тоуравнение регрессии со статистически значимыми факторами будет иметь вид, показанный на рисунке 4.31 (с использованием подпрограммы РЕГРЕССИЯ (рис. 4.30)).

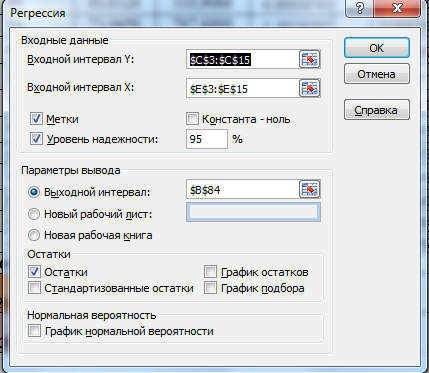


Рис. 4.30. Диалоговое окно подпрограммы РЕГРЕССИЯ

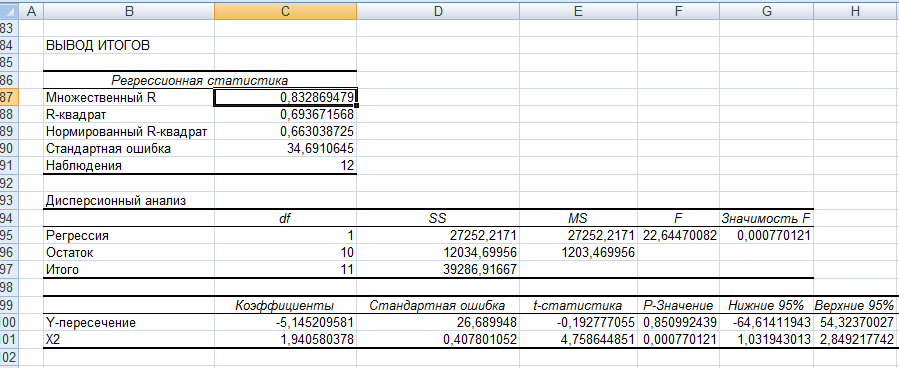


Рис. 4.31. Результаты расчета со статистически значимыми параметрами

## Задания для самостоятельной работы

**Задача 4.1.** Имеются данные о деятельности крупнейших компаний США в 1996 г.

|  |  |  |  |
| --- | --- | --- | --- |
| № | Чистый доход, млрд. долл. США, *y* | Оборот капитала, млрд. долл. США, *x*1 | Использованный капитал, млрд. долл. США, *х*2 |
| 1 | 0,9 | 31,3 | 18,9 |
| 2 | 1,7 | 13,4 | 13,7 |
| 3 | 0,7 | 4,5 | 18,5 |
| 4 | 1,7 | 10,0 | 4,8 |
| 5 | 2,6 | 20,0 | 21,8 |
| 6 | 1,3 | 15,0 | 5,8 |
| 7 | 4,1 | 137,1 | 99,0 |
| 8 | 1,6 | 17,9 | 20,1 |
| 9 | 6,9 | 165,4 | 60,6 |
| 10 | 0,4 | 2,0 | 1,4 |
| 11 | 1,3 | 6,8 | 8,0 |
| 12 | 1,9 | 27,1 | 18,9 |
| 13 | 1,9 | 13,4 | 13,2 |
| 14 | 1,4 | 9,8 | 12,6 |
| 15 | 0,4 | 19,5 | 12,2 |
| 16 | 0,8 | 6,8 | 3,2 |
| 17 | 1,8 | 27,0 | 13,0 |
| 18 | 0,9 | 12,4 | 6,9 |
| 19 | 1,1 | 17,7 | 15,0 |
| 20 | 1,9 | 12,7 | 11,9 |
| 21 | -0,9 | 21,4 | 1,6 |
| 22 | 1,3 | 13,5 | 8,6 |
| 23 | 2,0 | 13,4 | 11,5 |
| 24 | 0,6 | 4,2 | 1,9 |
| 25 | 0,7 | 15,5 | 5,8 |

**Задание**

1. Постройте линейное уравнение множественной регрессии и поясните экономический смысл его параметров.

2. Рассчитайте частные коэффициенты эластичности.

3. Определите стандартизованные коэффициенты регрессии.

4. Сделайте вывод о силе связи результата и факторов.

5. Определите парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделайте выводы.

6. Дайте оценку полученного уравнения на основе коэффициента детерминации и общего *F*-критерия Фишера.

7. Проверить полученные результаты с помощью программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.

8. Постройте уравнение регрессии со статистически значимыми факторами.

**Задача 4.2.** Имеются данные о деятельности крупнейших компаний США в 1996 г.

|  |  |  |  |
| --- | --- | --- | --- |
| № | Чистый доход, млрд. долл. США, *y* | Оборот капитала, млрд. долл. США, *x*1 | Численность служащих, тыс. чел., *х*2 |
| 1 | 6,6 | 6,9 | 222,0 |
| 2 | 3,0 | 18,0 | 32,0 |
| 3 | 6,5 | 107,9 | 82,0 |
| 4 | 3,3 | 16,7 | 45,2 |
| 5 | 0,1 | 79,6 | 299,3 |
| 6 | 3,6 | 16,2 | 41,6 |
| 7 | 1,5 | 5,9 | 17,8 |
| 8 | 5,5 | 53,1 | 151 |
| 9 | 2,4 | 18,8 | 82,3 |
| 10 | 3,0 | 35,3 | 103,0 |
| 11 | 4,2 | 71,9 | 225,4 |
| 12 | 2,7 | 93,6 | 675 |
| 13 | 1,6 | 10 | 43,8 |
| 14 | 2,4 | 31,5 | 102,3 |
| 15 | 3,3 | 36,7 | 105,0 |
| 16 | 1,8 | 13,8 | 49,1 |
| 17 | 2,4 | 64,8 | 50,4 |
| 18 | 1,6 | 30,4 | 480,0 |
| 19 | 1,4 | 12,1 | 71,0 |
| 20 | 0,9 | 31,3 | 43,0 |

**Задание**

1. Постройте линейное уравнение множественной регрессии и поясните экономический смысл его параметров.

2. Рассчитайте частные коэффициенты эластичности.

3. Определите стандартизованные коэффициенты регрессии.

4. Сделайте вывод о силе связи результата и факторов.

5. Определите парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделайте выводы.

6. Дайте оценку полученного уравнения на основе коэффициента детерминации и общего *F*-критерия Фишера.

7. Проверить полученные результаты с помощью программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.

8. Постройте уравнение регрессии со статистически значимыми факторами.

**Задача 4.3.** Имеются данные о деятельности крупнейших компаний США в 1996 г.

|  |  |  |  |
| --- | --- | --- | --- |
| № | Чистый доход, млрд. долл. США, *y* | Численность служащих, тыс. чел., *х*3 | Рыночная капитализация компании, млрд. долл. США, *х*4 |
| 1 | 0,9 | 43,0 | 40,9 |
| 2 | 1,7 | 64,7 | 40,5 |
| 3 | 0,7 | 24,0 | 38,9 |
| 4 | 1,7 | 50,2 | 38,5 |
| 5 | 2,6 | 106,0 | 37,3 |
| 6 | 1,3 | 96,6 | 26,5 |
| 7 | 4,1 | 347,0 | 37,0 |
| 8 | 1,6 | 85,6 | 36,8 |
| 9 | 6,9 | 745,0 | 36,3 |
| 10 | 0,4 | 4,1 | 35,3 |
| 11 | 1,3 | 26,8 | 35,3 |
| 12 | 1,9 | 42,7 | 35,0 |
| 13 | 1,9 | 61,8 | 26,2 |
| 14 | 1,4 | 212,0 | 33,1 |
| 15 | 0,4 | 105,0 | 32,7 |
| 16 | 0,8 | 33,5 | 32,1 |
| 17 | 1,8 | 142,0 | 30,5 |
| 18 | 0,9 | 96,0 | 29,8 |
| 19 | 1,1 | 140,0 | 25,4 |
| 20 | 1,9 | 59,3 | 29,3 |
| 21 | -0,9 | 131,0 | 29,2 |
| 22 | 1,3 | 70,7 | 29,2 |
| 23 | 2,0 | 65,4 | 29,1 |
| 24 | 0,6 | 23,1 | 27,9 |
| 25 | 0,7 | 80,8 | 27,2 |

**Задание**

1. Постройте линейное уравнение множественной регрессии и поясните экономический смысл его параметров.

2. Рассчитайте частные коэффициенты эластичности.

3. Определите стандартизованные коэффициенты регрессии.

4. Сделайте вывод о силе связи результата и факторов.

5. Определите парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделайте выводы.

6. Дайте оценку полученного уравнения на основе коэффициента детерминации и общего *F*-критерия Фишера.

7. Проверить полученные результаты с помощью программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.

8. Постройте уравнение регрессии со статистически значимыми факторами.

## Проведение исследования рынка недвижимости с помощью фиктивных (бинарных) переменных

**КРАТКАЯ СПРАВКА**

Как правило, в регрессионных моделях переменные имеют непрерывные области изменения (совокупный доход, размер заработной платы и др.). В том случае, если необходимо включение в модель факторов, не принимающих количественные значения (профессия, пол, образование, климатические условия, принадлежность к региону), им должны быть присвоены те или иные цифровые метки, то есть качественные переменные должны быть преобразованы в количественные. Такого вида сконструированные переменные принято называть ***фиктивными переменными***.

Переменные, которые могут принимать только два значения: 1 (если данный качественный признак присутствует) и 0 (при его отсутствии), называются ***дихотомические или бинарные***.

***Условие задачи***

Необходимо провести аналитическое исследование рынка недвижимости в г.Челябинска.

Перечень полученных зависимостей:

1. Средняя цена квадратного метра 1,2,3,4 – комнатных квартир в зависимости от района города, вида (старые квартиры/новостройки).
2. Эластичность стоимости квартиры (рассмотреть отдельно старые квартиры и новостройки, каждый район и количество комнат) по:
3. Планировке
4. Материалу стен
5. Местоположению (центр района или нет)
6. Этажу (первый, последний или средний)
7. Общей площади
8. Площади кухни
9. Жилой площади
10. Проводя каждое исследование, вычислить коэффициент корреляции и его статистическую достоверность.

Исходные данные можно найти на любом сайте, на котором опубликованы объявления о продаже недвижимости.

***Описание метода использования бинарных переменных***

Используем методы введения бинарных переменных для исследования рынка жилья в Челябинске.

Как известно стоимость квартиры зависит от ее площади: жилой, нежилой, кухни. Предположим, что собрана стати­стика по стоимости жилья, в которой:

* C – стоимость квартиры;
* S1 – общая площадь;
* S2 – жилая площадь;
* S3– площадь кухни.

В процессе многочисленных исследований установлено, что наибо­лее точно исходные данные описываются логарифмической моделью:

***lnC = a + b1 lnS1+b2 ln S2* + *b3 ln S3.***

Кроме перечисленных факторов, на стоимость квартиры оказывают влияние материал стен, район, планировка и этаж, на котором она расположена (с одной стороны, большинство предпочитает не покупать квартиры на первых и последних этажах домов, но, с другой стороны, квартиры на первых этажах ценятся выше из-за возможности устройства в них офисных помещений).

Однако, если стоимость квартиры и ее площадь, как жилая, так и нежилая, являются количественными характеристиками, которые можно измерить, то материал стен, планировка и этаж – характеристики качественные.

Введем бинарные переменные:

планировка: Р = 1, если квартира 97-й или 121-й серии,

Р =0, если квартира хрущ/бреж. планировки;

район: L = 1, если квартира расположена в центре района,

L = 0 в остальных случаях;

этаж: M=1, если этаж не крайний,

M=0, если этаж крайний.

С учетом фиктивных переменных Р, М, и L модель примет вид:

***lnC=a+ b1 P + b2 L + b3 M + b4 lnS1+b5 ln S2 + b6 ln S3 .***

***Алгоритм проведения исследования***

1. С любого сайта объявлений о продаже недвижимости скопируйте текущую информацию и представьте ее в следующем виде (табл. 4.2):

Таблица 4.2

Информация о продаже квартир



**КРАТКАЯ СПРАВКА**

Результаты многих исследований подтверждают, что число наблюдений должно в 6-7 раз превышать число рассчитываемых параметров. Поэтому, если мы хотим найти влияние шести факторов (общей, жилой площади, площади кухни, материалу стен, району и этажу), то количество анализируемых записей должно быть не менее 36.

Строки с неполными данными (например, строка 10 не содержит информации о жилой площади и площади кухни) из таблицы удаляем.

1. Для проведения регрессионного анализа необходимо преобразовать исходные данные и представить их в виде таблицы (табл. 4.3).

Таблица 4.3

Таблица с данными, подготовленными для регрессионного анализа



1. Далее необходимо прологарифмировать количественные характеристики (стоимость квартиры, все виды площадей) и ввести бинарные переменные для учета качественных признаков (планировка, этаж, район). В следующей таблице представ­лены преобразованные данные, которые могут быть использованы для проведения регрессионного анализа (табл. 4.3).

Таблица 4.3

Таблица с прологарифмированными количественными характеристиками



1. Проводим регрессионный анализ. В качестве входного интервала Y выбираются значения столбца «Стоимость», в качестве входного интервала X выбираются значения столбцов «Планировка», «Район», «Этаж», «Общая площадь», «Жилая площадь», «Площадь кухни».

Процедура выполнения регрессионного анализа является стандартной (рис. 4.32).

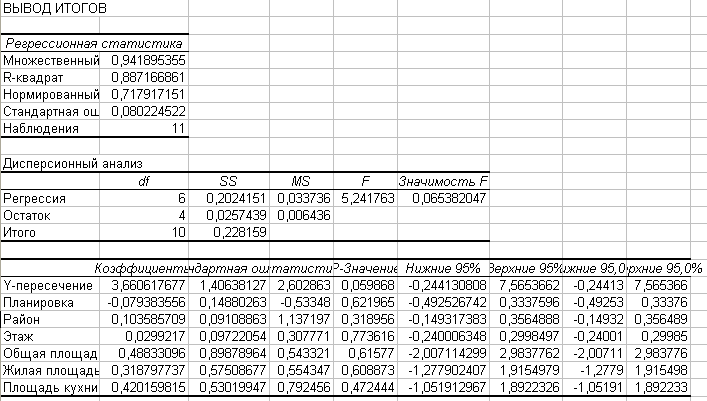


Рисунок 4.32 Результат выполнения регрессионного анализа

Используя метод наименьших квадратов, получаем оценки соответствующих параметров регрессионной модели, на основе которых записывается уравнение регрес­сии:

***ln С = 3,66 - 0,08 ln S1 + 0,10 ln S2 + 0,02 ln S3 + 0,49 P +0,31 L+0,42 M.***

***Анализ полученных данных***

В процессе исследования качественных характеристик полученной регрессионной модели установлено, что коэффициент корреляции достаточно высокий и статистически достоверен (объем статистической информации доста­точен).

Отрицательный коэффициент **b1,** равный -0,08, означает, что улучшение планировки уменьшает ее стоимость. Как это можно объяснить? Дело в том, что если бы квартиры в панельных домах имели бы такие же площади, что квартиры хрущ/бреж. планировки, то они стоили бы дешевле! Таким образом, мы выявили один из парадоксов челябинского рынка недвижимости.

Коэффициент **b2** (район) при бинарной переменной L, равный 0,1, означает, что квартиры, находящиеся в центре Центрального района стоят на 10% дороже аналогичных, но находящихся не в центре района.

Коэффициент **b3** (этаж) при бинарной переменной M, равный 0,02, означает, что стоимость квартир, расположенных на средних этажах, всего на 2% превышает стоимость аналогичных квартир, находящихся на крайних этажах. Это объясняется тем, как было указано выше, квартиры на 1-м этаже в основном покупаются для создания офиса. Поэтому, для достоверного анализа необходимо в модель ввести дополнительные бинарные переменные: дом находится во дворе или выходит окнами на улицу, квартира на первом или последнем этаже.

Далее перейдем к анализу количественных показателей.

Коэффициенты b4, b5, b6 показывают эластичность стоимости квартиры по общей, жилой площади и площади кухни.

Мы видим, что в наибольшей степени на стоимость квартиры влияет общая площадь (что естественно) – увеличение на 1м общей площади ведет к увеличению стоимости на 0,49%.

Увеличение жилой площади квартиры на 1% увеличивает ее стоимость на 0,31%. Иначе говоря, эластичность стоимости квартиры по жилой площади равна 0,31.