# Лабораторная работа 1. Линейная регрессия и корреляция: смысл и оценка параметров

**КРАТКАЯ СПРАВКА**

Линейная регрессия имеет вид:

,

где ε – случайные остатки.

Построение уравнения регрессии сводится к оценке ее параметров.

Для оценки параметров регрессий, линейных по параметрам, используют метод наименьших квадратов (МНК).

МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических минимальна, т.е.

.

Для линейных уравнений решается следующая система относительно a и b:

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

, ,

где ,.

## Решение типовых задач

**Пример**

По группе предприятий, выпускающих один и тот же вид продукции, известны данные за ноябрь 2012 г. (табл. 1.1).

Таблица 1.1

|  |  |  |
| --- | --- | --- |
| *№ предприятия* | *Выпуск продукции, тыс. ед. (x)* | *Затраты на производство, млн. руб. (y)* |
| *1* | *1* | *30* |
| *2* | *2* | *70* |
| *3* | *4* | *150* |
| *4* | *3* | *100* |
| *5* | *5* | *170* |
| *6* | *3* | *100* |
| *7* | *4* | *150* |

**Требуется**

1. Построить поле корреляции.
2. Рассчитать параметры уравнения линейной парной регрессии.
3. Оценить тесноту связи с помощью показателей корреляции и детерминации.
4. Дать с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оценить качество уравнений с помощью средней ошибки аппроксимации.
6. Оценить статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера.
7. На уровне значимости α = 0,05 оцените значимость коэффициентов регрессии.
8. Рассчитать ожидаемое значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определить доверительный интервал прогноза для уровня значимости α = 0,05.
9. Проверить полученные результаты с помощью стандартных статистических функций ТЕНДЕНЦИЯ, ЛИНЕЙН и программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.
10. Оценить полученные результаты.

**Решение**

1. Постройте **поле корреляции**. Для этого воспользуйтесь программой MS Excel:

* Создайте рабочую книгу *ПРИМЕР1.xlsx*. На листе *Парная\_регрессия* постройте таблицу 1.1 (рис. 1.1).

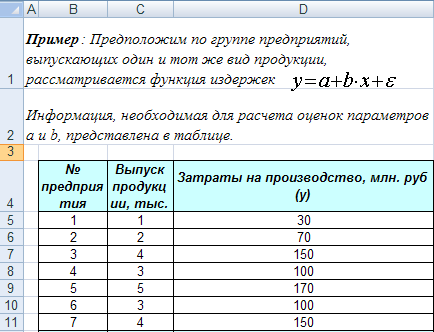


Рис. 1.1. Таблица с исходными данными

* Используя мастер диаграмм, постройте точечную диаграмму зависимости затрат на производство от выпуска продукции (**Вставка/ Диаграммы/ Точечная**) (рис. 1.2).



Рис. 1.2. Поле корреляции

**ВЫВОДЫ**

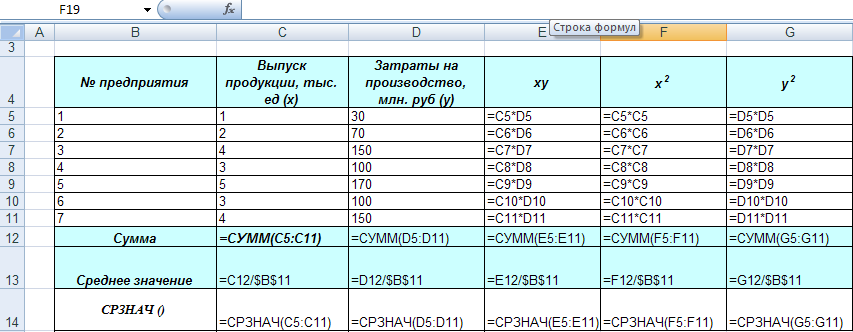
*По полю корреляции можно предположить, что корреляционная связь между признаками линейная.*

2. Рассчитайте параметры уравнения линейной парной регрессии.

Найдите выборочные средние (рис. 1.3):

.

Используя Мастер функций Microsoft Excel, проверьте полученные значения с помощью стандартной функции **СРЗНАЧ()** из категории **Статистические** (рис. 1.3).



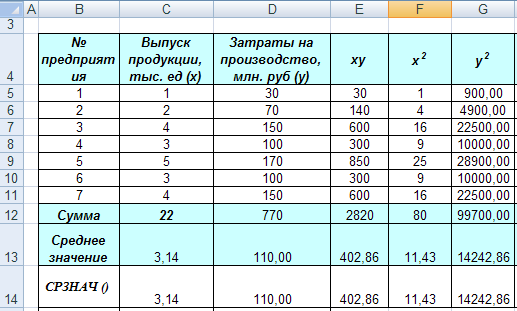


Рис. 1.3. Выборочные средние

3. Найдите значения выборочных дисперсий и средних квадратических отклонений (рис. 1.4):

;

.

Проверьте полученные значения с помощью стандартной функции **ДИСПР()** и **СТАНДОТКЛОНП()** из категории **Статистические** (рис. 1.4).

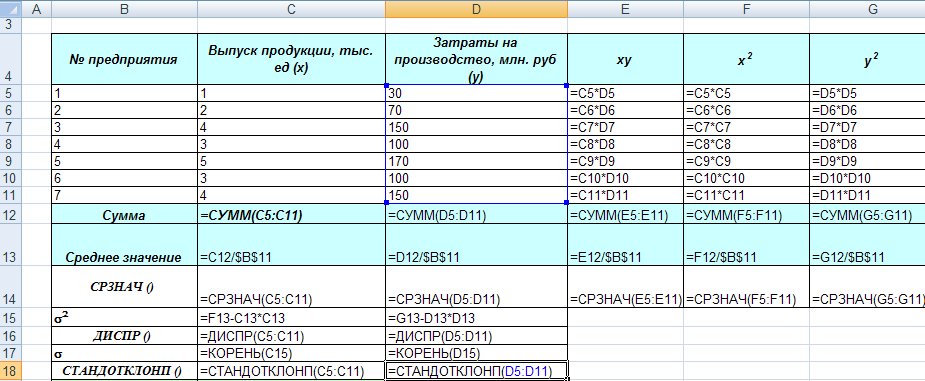




Рис. 1.4. Выборочные дисперсии и средние квадратические отклонения

Найдите выборочный коэффициент ковариации

и проверьте полученное значение с помощью стандартной статистической функции **КОВАР(x; y)** (рис. 1.5).

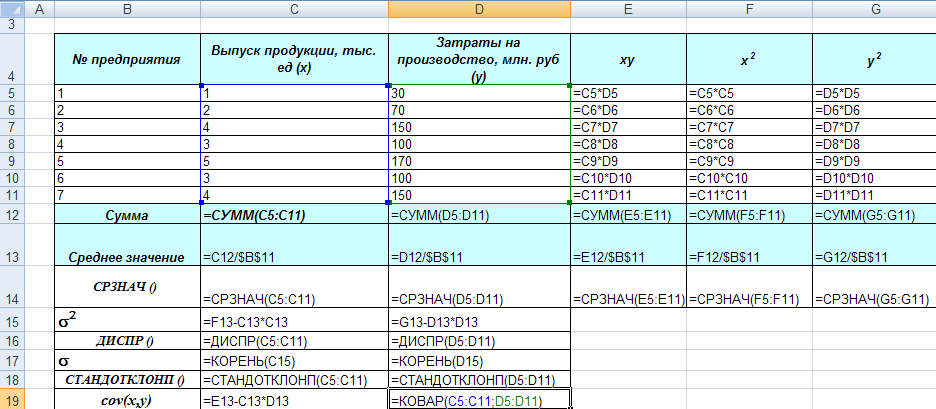




Рис. 1.5. Выборочный коэффициент ковариации

По исходным данным рассчитайте коэффициенты уравнения регрессии *a* и *b* (рис. 1.6):

,

,

где .



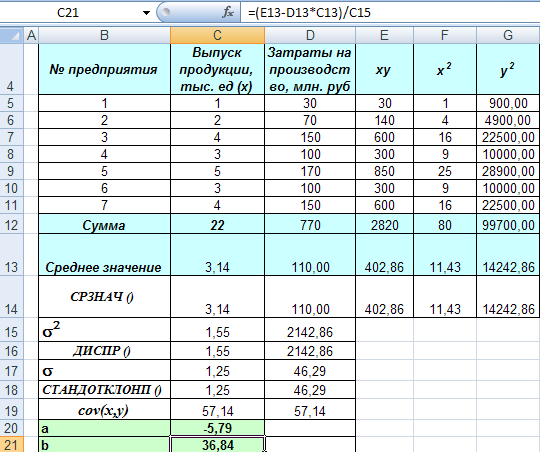
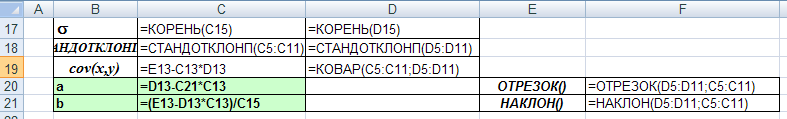


Рис. 1.6. Параметры уравнения регрессии

Проверьте полученные значения с помощью стандартных статистических функций **НАКЛОН(y; x)** и **ОТРЕЗОК(y; x)** соответственно (рис. 1.7).



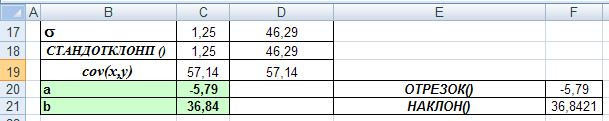


Рис. 1.7. Использование функций НАКЛОН(y; x) и ОТРЕЗОК(y; x)

Отсюда эмпирическое линейное уравнение регрессии имеет вид:

.

В данном случае величина параметра *a* не имеет экономического смысла. То, что *a<0*, соответствует опережению изменения результата над изменением фактора.

**ВЫВОДЫ**

*С увеличением выпуска продукции на 1 тыс. ед. затраты на производство возрастают в среднем на 36,84 млн. руб., т.е. дополнительный выпуск продукции на 1 тыс. ед. потребует увеличение затрат в среднем на 36,84 млн. руб.*

4. Оцените тесноту связи с помощью показателей корреляции и детерминации.

**КРАТКАЯ СПРАВКА**

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции *rxy*  для линейной регрессии :

.

В общем случае, когда величины x и y связаны произвольной вероятностной зависимостью, линейный коэффициент корреляции принимает значения в пределах , тогда качественная оценка тесноты связи величин может быть выявлена на основе шкалы Чеддока:

|  |  |  |
| --- | --- | --- |
| Теснота связи | Значение коэффициента корреляции при наличии | |
| Прямой связи | Обратной связи |
| Слабая | 0,1 – 0,3 | (-0,1) – (-0,3) |
| Умеренная | 0,3 – 0,5 | (-0,3) – (-0,5) |
| Заметная | 0,5 – 0,7 | (-0,5) – (-0,7) |
| Высокая | 0,7 – 0,9 | (-0,7) – (-0,9) |
| Весьма высокая | 0,9 – 0,99 | (-0,9) – (-0,99) |

Оценку качества построенной модели даст коэффициент (индекс) детерминации.

Задача дисперсионного анализа состоит в анализе дисперсии зависимой переменной:

,

где – общая сумма квадратов отклонений;

– сумма квадратов отклонений, обусловленная регрессией («объясненная» или «факторная»);

– остаточная сумма квадратов отклонений.

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака *y* характеризует *коэффициент (индекс) детерминации* :

.

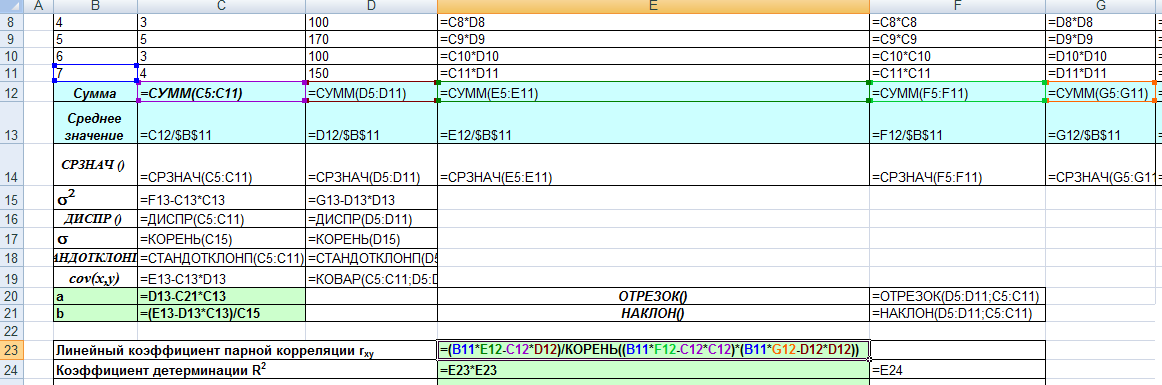
*Коэффициент детерминации* – квадрат коэффициента или индекса корреляции.

Коэффициент линейной корреляции между двумя переменными *x* и *y* при небольших значениях *n* (*n<30*) вычисляется по формуле:

.

Для парной линейной регрессии коэффициент детерминации равен квадрату коэффициента корреляции: .

На рисунке 1.8 показаны рассчитанные коэффициент линейной корреляции и коэффициент детерминации.



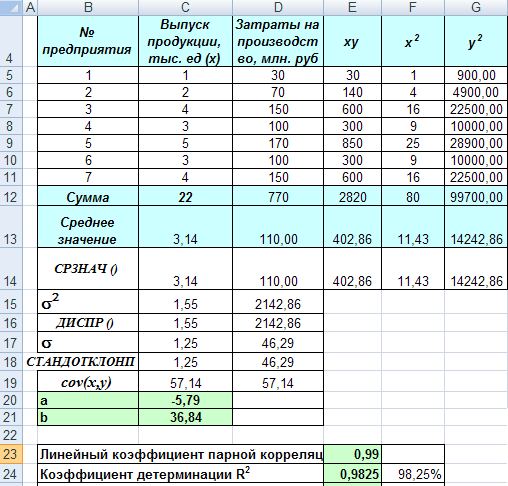
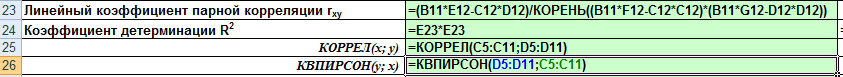


Рис. 1.8. Коэффициенты линейной корреляции и детерминации

Проверьте полученные значения с помощью стандартных статистических функций **КОРРЕЛ(x; y)** и **КВПИРСОН(y; x)** (рис. 1.9).



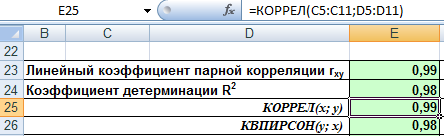


Рис. 1.9. Использование функций КОРРЕЛ(x; y) и КВПИРСОН(y; x)

**ВЫВОДЫ**

*В результате вычислений нами получены результаты:*

* . *Связь весьма высокая, прямая.*
* . *Вариация результата на 98,25% объясняется вариацией фактора x.*

4. Дайте с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

**КРАТКАЯ СПРАВКА**

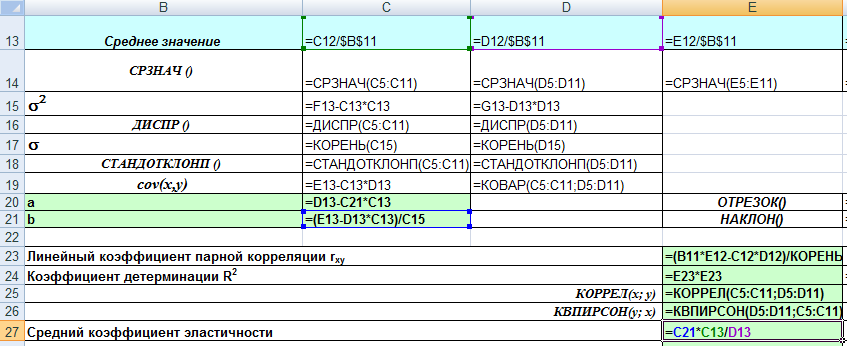
*Средний коэффициент эластичности* показывает, на сколько процентов в среднем по совокупности изменится результат *y* от своей средней величины при изменении фактора *x* на *1%* от своего среднего значения:

.

Средний коэффициент эластичности для линейной модели равен:

.

Используя исходные данные, рассчитайте коэффициент эластичности (рис. 1.10).



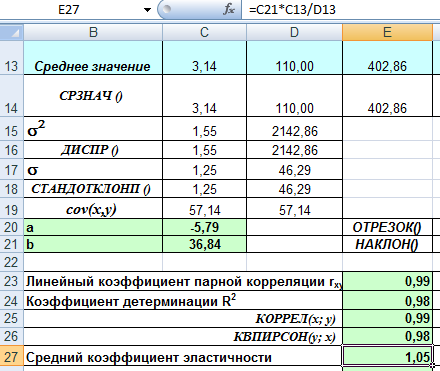


Рис. 1.10. Средний коэффициент эластичности

**ВЫВОДЫ**

. *Результат y в среднем по совокупности вырастет на 1,05% от своей средней величины при изменении фактора x на 1% от своего среднего значения.*

5. Качество модели определяет средняя ошибка аппроксимации.

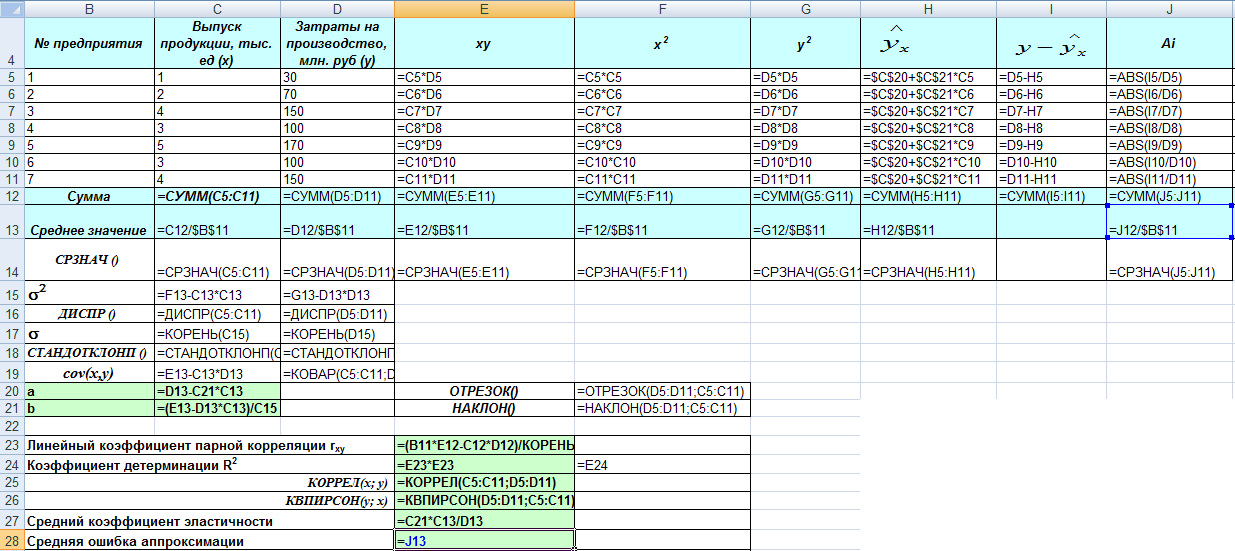
**КРАТКАЯ СПРАВКА**

*Средняя ошибка аппроксимации* – среднее отклонение расчетных значений от фактических:

.

Допустимый предел значений – не более *8-10%.*

Для ее расчета добавим три столбца в таблицу 1.1, в которых соответственно найдем теоретические значения *y* (), рассчитанные по уравнению регрессии, разницу между фактическим и теоретическим значением *y* и среднюю ошибку аппроксимации . Результаты приведены на рисунке 1.11.



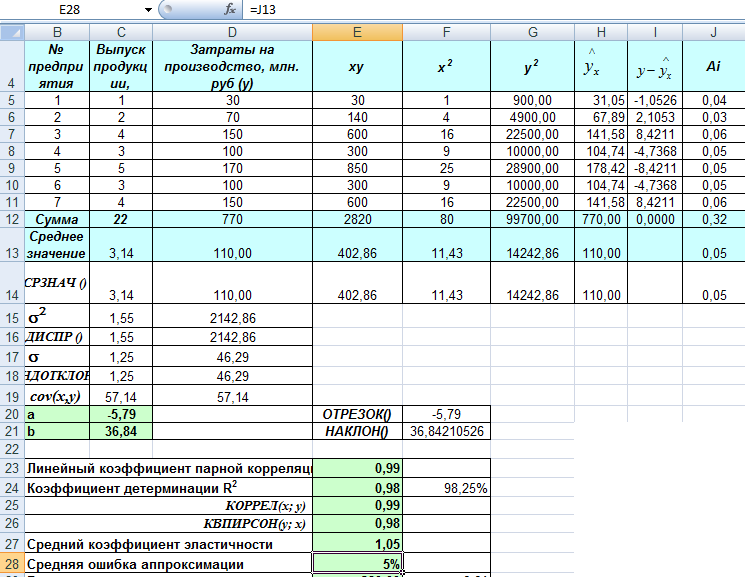


Рис. 1.11. Средняя ошибка аппроксимации

**ВЫВОДЫ**

*Для парной линейной регрессии: . В среднем расчетные значения отклоняются от фактических на 5%. Качество построенной модели оценивается как хорошее, т.к.*  *не превышает 8 – 10%.*

6. Оцените статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера.

**КРАТКАЯ СПРАВКА**

F-тест – оценивание качества уравнения регрессии – состоит в проверке гипотезы о статистической незначимости уравнения регрессии и показателя тесноты связи.

Для этого выполняется сравнение фактического и критического (табличного) значений F-критерия Фишера.

определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

,

где *n* – число единиц совокупности;

*m* – число параметров при переменных x.

– это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости .

Уровень значимости – вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно принимается равной *0,005* или *0,01*.

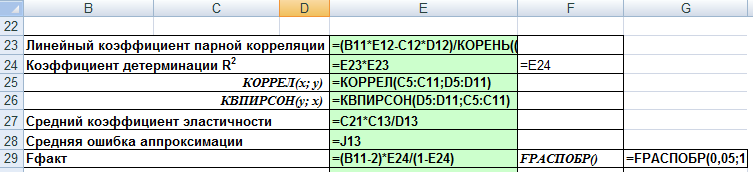
***Если , то -гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность.***

***Если , то гипотеза не отклоняется и признается статистическая незначимость, ненадежность уравнений регрессии.***

Определите расчетное значение *F*-статистики через найденные коэффициенты детерминации, используя соотношение:

.

Результаты расчетов приведены на рисунке 1.12.



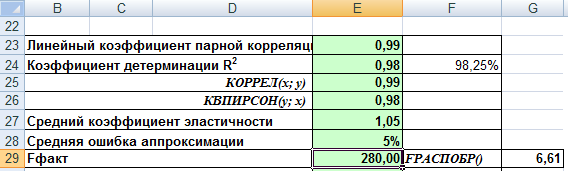


Рис. 1.12. Фактическое и критическое значения *F*-статистики

Отсюда расчетное значение *F*-статистики для линейной модели равно: .

Для расчета критического (табличного) значения использовали статистическую функцию **FРАСПОБР(0,05;1;5)** *(α=0,05; m=1; n-m-1=5)* (рис. 1.12).

Критическое значение *F*-статистики равно: .

**ВЫВОДЫ**

*Так как , линейная модель по F-критерию Фишера статически значима и надежна при доверительной вероятности 95%.*

7. Оцените статистическую значимость коэффициентов регрессии на уровне значимости α = 0,05.

**КРАТКАЯ СПРАВКА**

Для оценки с*татистической значимости коэффициентов регрессии и корреляции* рассчитываются *t-критерий Стьюдента* и *доверительные интервалы* каждого из показателей.

Выдвигается гипотеза о случайной природе показателей, т.е. о незначительном их отклонении от нуля. Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

; ; .

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

;

;

.

Сравнивая фактическое и критическое (табличное) значения *t*-статистики – и – принимаем или отвергаем гипотезу .

Связь между F-критерием Фишера и *t-*статистикой Стьюдента выражается равенством

.

***Если , то отклоняется, т.е. a, b и rxy не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x.***

***Если , то не отклоняется и признается случайная природа формирования a, b или rxy .***

Для расчета доверительного интервала определяем *предельную ошибку*  для каждого показателя:

, .

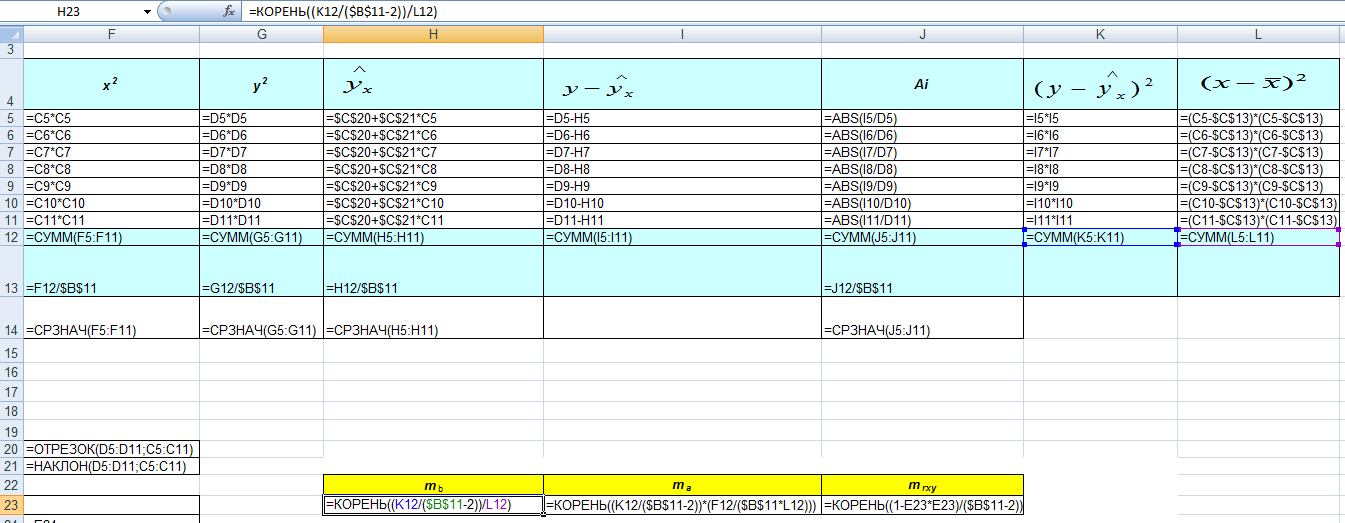
Формулы для расчета доверительных интервалов имеют следующий вид:

; ;

; ; .

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

Для расчета случайных ошибок параметров линейной регрессии и коэффициента корреляции дополнительно введите два столбца, вычисляющих  и  (рис. 1.13). Подсчитайте их сумму.



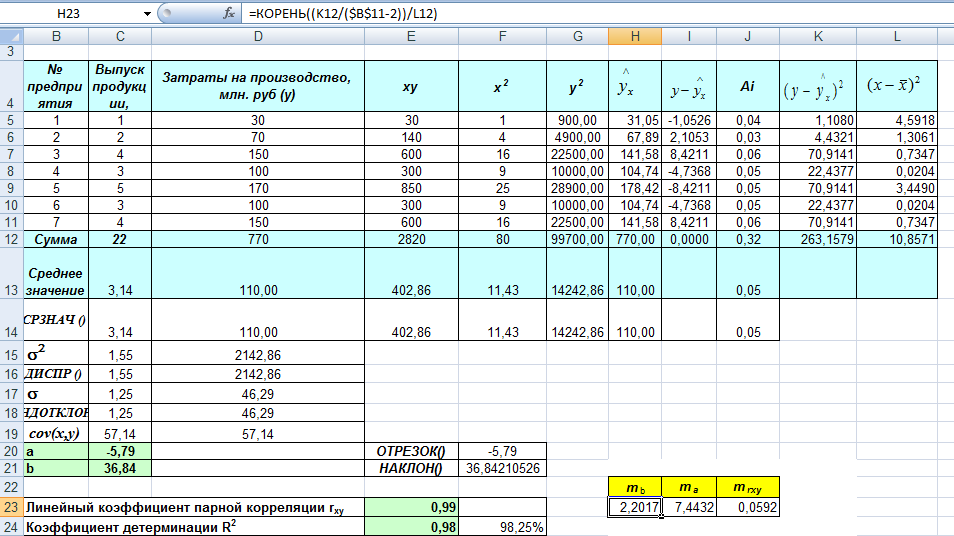
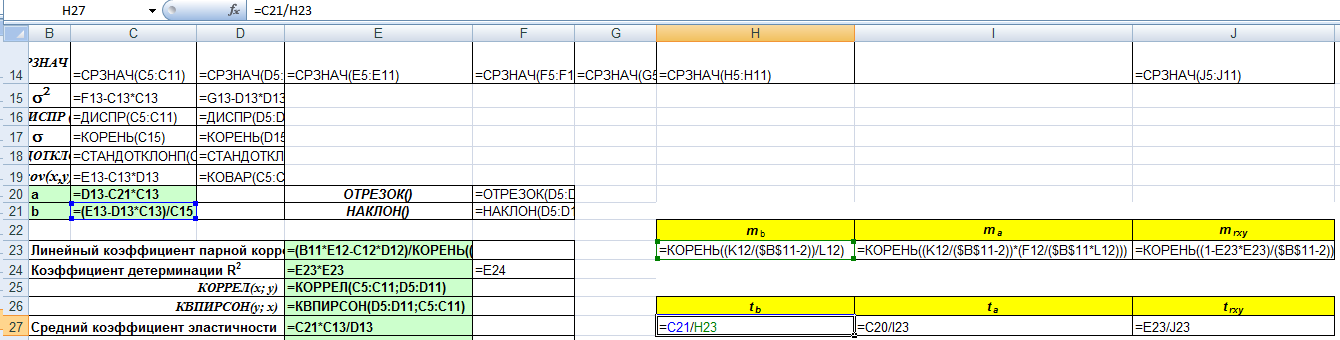


Рис.1.13. Случайные ошибки параметров линейной регрессии и коэффициента корреляции

Статистики критерия значимости коэффициентов регрессии и корреляции, таким образом, равны (рис. 1.14):

; ; .



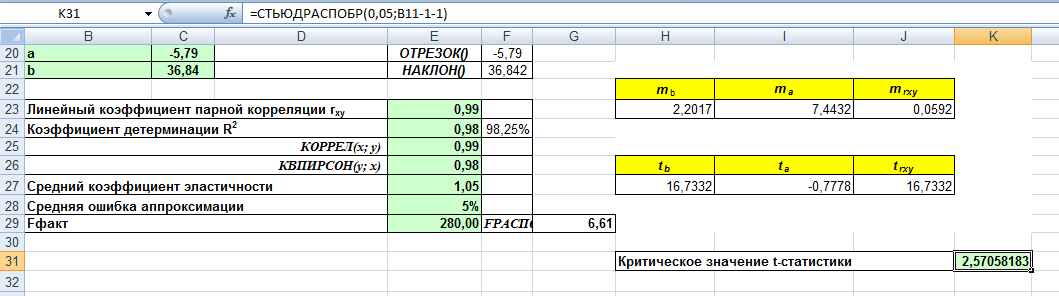


Рис.1.14. Статистики критерия значимости коэффициентов регрессии и корреляции

Табличное (критическое) значение t-критерия *tкр* для числа степеней свободы *df=n-2=5 и α=0,05* составит *tкр = t0,05;5 = 2,5706*. Для его расчета использовали статистическую функцию =**СТЬЮДРАСПОБР(0,05;5)** *(α=0,05; m=1; n-m-1=5)* (рис. 1.14).

**ВЫВОДЫ**

*Так как* () *, то не отклоняется и признается случайная природа формирования a.*

*Так как и* ()*, то отклоняется, т.е. b и rxy не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x.*

8. Рассчитайте ожидаемое значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости α = 0,05.)

**КРАТКАЯ СПРАВКА**

*Прогнозное значение yp*определяется путем подстановки в уравнение регрессии соответствующего (прогнозного) значения *xp.* Вычисляется средняя стандартная ошибка прогноза :

,

где ;

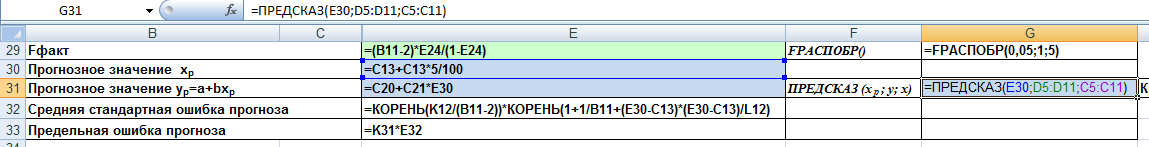
и строится *доверительный интервал прогноза*:

; ; ,

где .

Прогноз по линейной модели для получаем, подставив значение *xр* в уравнение линейной регрессии (рис. 1.15):  млн. руб.

Проверим полученное решение, используя стандартную статистическую функцию **ПРЕДСКАЗ (xp; y; x)** (рис. 1.15)**.**



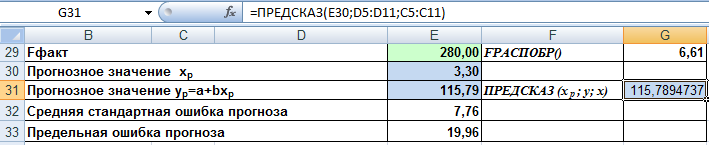


Рис. 1.15. Прогноз по линейной модели

Средняя стандартная ошибка прогноза вычисляется по формуле и равна: ,

где (рис.1.15).

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит: .

Доверительный интервал прогноза для уровня значимости *α=0,05* определяется в виде:



9. Проверьте полученные результаты с помощью стандартных статистических функций ТЕНДЕНЦИЯ, ЛИНЕЙН и программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.

Для определения значений результативного признака по линейному уравнению регрессии с помощью стандартной статистической функции **ТЕНДЕНЦИЯ**() выполните следующие операции:

* в расчетной таблице добавьте столбец *Тенденция* и выделите 7 значащих позиций этого столбца (i=1,2,…,n);
* с помощью Мастера функций выберите статистическую функцию **ТЕНДЕНЦИЯ**();
* в поля **Известные\_значения\_y** и **Известные\_значения\_x** введите значения векторов*y*и *x* соответственно;
* поле **Новые\_значения\_x** оставьте пустым (при этом предполагается, что **Новые\_значения\_x** совпадают с **Известные\_значения\_x**);
* поле **Конст** оставьте пустым (если **Конст** имеет значение ИСТИНА, 1 или опущена, то коэффициент *a* вычисляется обычным образом, если **Конст** имеет значение ЛОЖЬ или 0, то коэффициент *a* полагается равным нулю);
* контролируем результат решения в окне функции (первый элемент массива) (рис. 1.16). **ОК**;

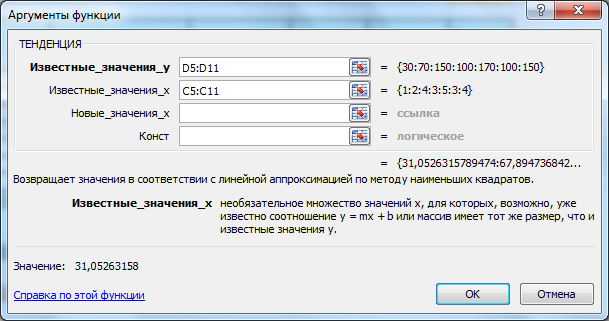


Рис. 1.16. Статистическая функция ТЕНДЕНЦИЯ

* для получения массива результатов сначала нажмите на клавишу **<F2>,** а затем – на комбинацию клавиш **<Ctrl+Shift+Enter>** (в выделенном столбце появятся результаты вычислений) (рис. 1.17).

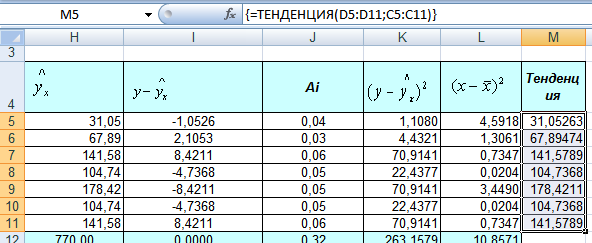


Рис. 1.17. Результат использования функции ТЕНДЕНЦИЯ

Для определения параметров линейного уравнения регрессии с помощью стандартной статистической функции **ЛИНЕЙН()** выполните следующие операции:

* выделите область пустых ячеек  (5 строк, 2 столбца) для вывода результатов регрессионной статистики;
* с помощью Мастера функций выберите статистическую функцию **ЛИНЕЙН ()**. **ОК.**
* заполните аргументы функции (рис. 1.18):

|  |  |
| --- | --- |
| **Известные\_значения\_y** | – диапазон, содержащий данные результативного признака (в нашем примере: *D5:D11*). |
| **Известные\_значения\_x** | – диапазон, содержащий данные факторов независимого признака (в нашем примере: *C5:C11*). |
| **Конст** | – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении; если *Конст=1,* то свободный член рассчитывается обычным образом, если *Конст=0*, то свободный член равен 0 (в нашем случае *Конст=1)*. |
| **Статистика** | – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. Если *Статистика=1,*то дополнительная информация выводится, если *Статистика=0,* то выводятся только оценки параметров уравнения (в нашем случае *Статистика =1).* |

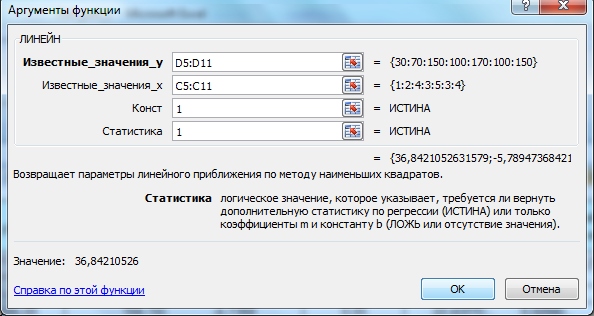


Рис. 1.18.Статистическая функция ЛИНЕЙН()

* В левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу **<F2>,** а затем – на комбинацию клавиш **<Ctrl+Shift+Enter>.**

Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

|  |  |
| --- | --- |
| Значение коэффициента b | Значение коэффициента a |
| Стандартное значение ошибки для коэффициента b | Стандартное значение ошибки для постоянной a |
| Коэффициент детерминации R2 | Стандартная ошибка для оценки y |
| F - статистика | Число степеней свободы |
| Регрессионная сумма квадратов | Остаточная сумма квадратов |

На рисунке 1.19 представлен результат вычисления функции ЛИНЕЙН.

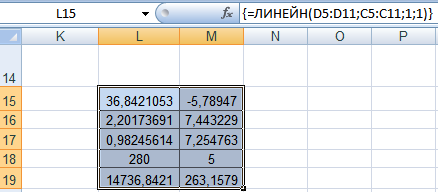


Рис. 1.19. Результат вычисления функции ЛИНЕЙН()

Режим работы **Регрессия** служит для расчета параметров уравнения линейной регрессии и проверки его адекватности исследуемому процессу. С его помощью, помимо результатов регрессионной статистики, дисперсионного анализа и доверительных интервалов, можно получить остатки и графики подбора линии регрессии, остатков и нормальной вероятности. Порядок действий следующий:

* проверьте доступ к пакету анализа (**Данные / Анализ данных**) (рис. 1.20).

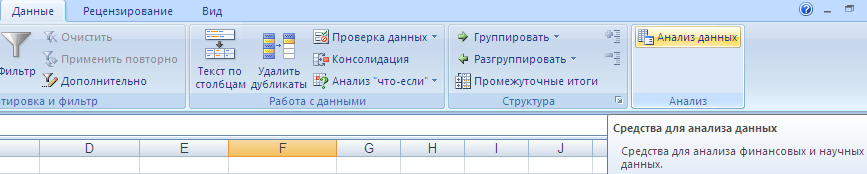


Рис. 1.20. Доступ к пакету анализа

**КРАТКАЯ СПРАВКА**

Если надстройка Анализ данных не установлена, то необходимо проделать следующую последовательность действий:

1. нажать на кнопку **Настройка панели быстрого доступа** (рис. 1.21) и выбрать **Другие команды…**;

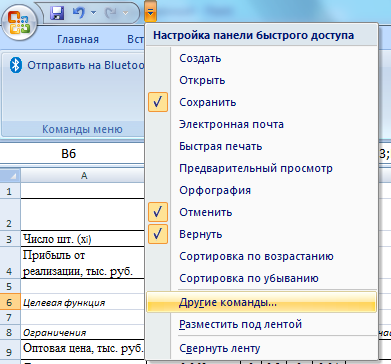


Рис. 1.21 Настройка панели быстрого доступа

1. в диалоговом окне **Параметры Excel** выбрать **Надстройки** инажать на кнопку **Перейти** (рис. 1.22);

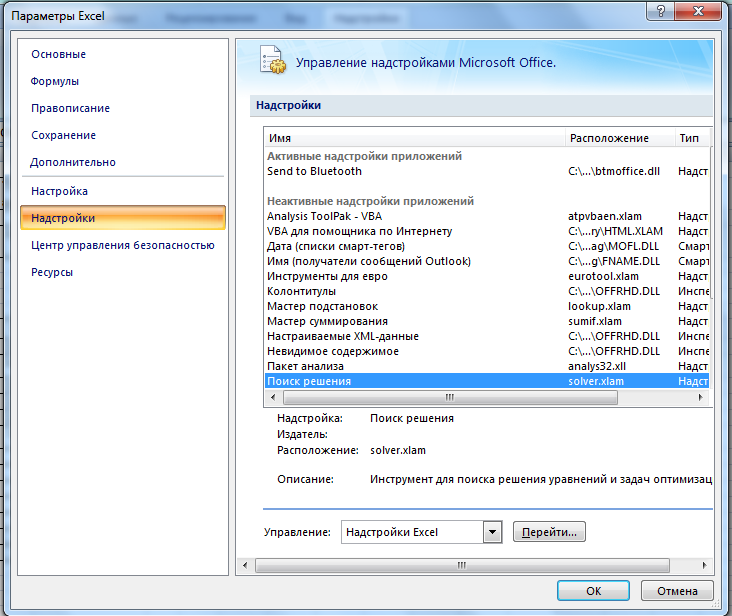


Рис. 1.22. Окно Параметры Excel

1. в диалоговом окне **Надстройки** установить галочку **Пакет анализа** (рис. 1.23). **ОК**.

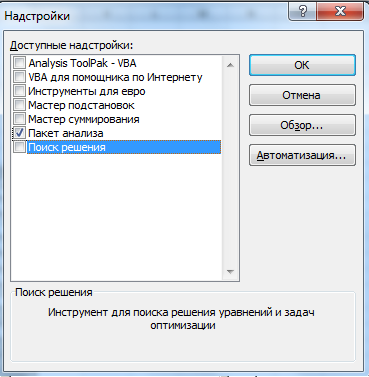


Рис. 1.24. Надстройки. Пакет анализа

* Выберите **Данные/Анализ данных/Регрессия**. ОК.
* Заполните диалоговое окно ввода данных и параметров вывода (рис. 1.25):

|  |  |
| --- | --- |
| **Входной интервал Y:** | *–* вводится ссылка на ячейки, содержащие данные по результативному признаку. Диапазон должен состоять из одного столбца. В нашем примере – это диапазон *D4:D11*. |
| **Входной интервал X:** | *–* вводится ссылка на ячейки, содержащие факторные признаки. Максимальное число входных диапазонов (столбцов) равно 16. В нашем примере – это диапазон *C4:C11.* |
| Флажок**Метки** | – устанавливается в активное состояние, если первая строка (столбец) во входном диапазоне содержит заголовки. Если заголовки отсутствуют, флажок следует деактивизировать. В этом случае будут автоматически созданы стандартные названия для данного выходного диапазона. В нашем примере – флажок установлен. |
| **Уровень надежности** | – установите данный флажок в активное состояние, если в поле, расположенное напротив флажка, необходимо ввести уровень надежности, отличный от уровня 95%, применяемого по умолчанию. Установленный уровень надежности используется для проверки значимости коэффициента детерминации *R2* и коэффициентов регрессии. |
| **Константа ноль** | – установите данный флажок в активное состояние, если требуется, чтобы линия регрессии прошла через начало координат (т.е. *a=0*). В нашем примере – флажок снят. |
| **Выходной интервал/ Новый рабочий лист/ Новая рабочая книга** | – указывают параметры вывода. В нашем примере – выберите **Выходной интервал**, введите ссылку на левую верхнюю ячейку выходного диапазона (например, *В37*). |
| **Остатки** | – установите данный флажок в активное состояние, если требуется включить в выходной диапазон столбец остатков. В нашем примере – флажок установлен. |
| **Стандартизованные остатки** | *–* установите данный флажок в активное состояние, если требуется включить в выходной диапазон столбец стандартизованных остатков. В нашем примере – флажок установлен. |
| **График остатков** | *–* установите данный флажок в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости остатков от факторных признаков *xj*. В нашем примере – флажок установлен. |
| **График подбора** | *–* установите данный флажок в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости теоретических результативных значений  от факторных признаков *xj*. В нашем примере – флажок установлен. |
| **График нормальной вероятности** | *–* установите данный флажок в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости наблюдаемых значений *y* от автоматически формируемых интервалов персентилей. График строится на основе генерируемой таблицы «Вывод вероятности». В нашем примере – флажок установлен. |

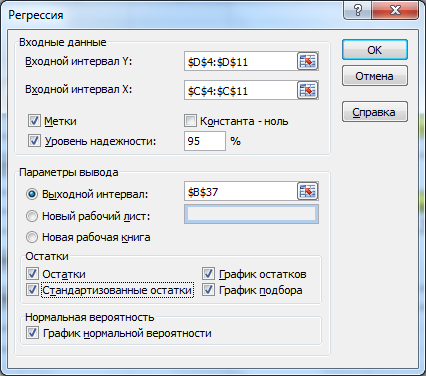


Рис. 1.25. Регрессия

Результаты регрессионной статистики представлены на рисунке 1.26.

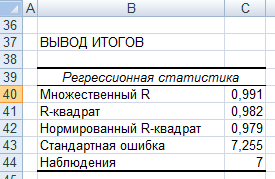


Рис. 1.26. Таблица вывод итогов

Эти результаты соответствуют следующим статистическим показателям:

1. Множественный R – коэффициент корреляции *rxy* ;
2. R-квадрат – коэффициенту детерминации *R2*;
3. Нормированный R-квадрат – для парной регрессии определяется выражением (вычисление этого коэффициента целесообразно только для множественной регрессии);
4. Стандартная ошибка – остаточному стандартному отклонению ;
5. Наблюдения – числу наблюдений *n.*

Результаты дисперсионного анализа, которые используются для проверки значимости коэффициента детерминации *R2* , представлены на рисунке 1.27.

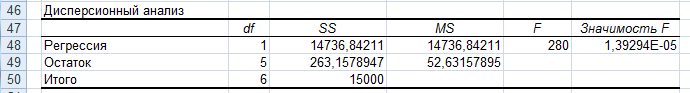


Рис. 1.27. Дисперсионный анализ

*df –* число степеней свободы.

Для строки *Регрессия* число степеней свободы определяется количеством факторных признаков *m* в уравнении регрессии *kф=m*.

Для строки *Остаток* число степеней свободы определяется числом наблюдений *n* и количеством переменных в уравнении регрессии *m+1: k0=n-(m+1).*

Для строки *Итого* число степеней свободы определяется суммой *kY= kф+ k0.*

*SS –* сумма квадратов отклонений.

Для строки *Регрессия -* это сумма квадратов отклонений теоретических данных от среднего: .

Для строки *Остаток* *-* это сумма квадратов отклонений эмпирических данных от теоретических : .

Для строки *Итого* *-* это сумма квадратов отклонений эмпирических данных от среднего .

*MS -* дисперсии, рассчитанные по формуле .

Для строки *Регрессия –* это факторная дисперсия .

Для строки *Остаток* *–* это остаточная дисперсия .

*F –* расчетное значение F-критерия Фишера , вычисляемое по формуле .

*Значимость F –* значение уровня значимости, соответствующее вычисленному значению . Определяется с помощью функции *=FРАСП(Fp; df(регрессия); df(остаток)).*

На рисунке 1.28 представлены значения коэффициентов регрессии и их статистические оценки.

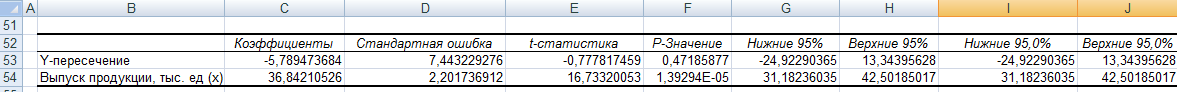


Рис. 1.28. Значения коэффициентов регрессии и их статистические оценки

*Коэффициенты –* значения коэффициентов регрессии.

*Стандартная ошибка –* стандартные ошибки коэффициентов регрессии.

*t-статистика* – расчетные значения t-критерия, вычисляемые по формуле .

*P*-*значение* – значения уровней значимости, соответствующие вычисленным значениям *tp.* Определяется с помощью функции *=СТЬЮДРАСП(tp;n-m-1).*

*Нижние 95% и Верхние 95%* соответственно нижние и верхние границы доверительных интервалов для коэффициентов регрессии.

На рисунке 1.29 представлены теоретические значения  результативного признака *Y* и значения остатков. Последние вычисляются как разность между эмпирическими *y* и теоретическими  значениями результативного признака Y.

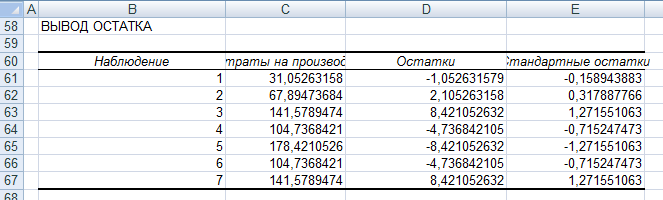


Рис. 1.29. Теоретические значения результативного признака и значения остатков

На рисунке 1.30 представлены интервалы персентилей и соответствующие им эмпирические значения y.

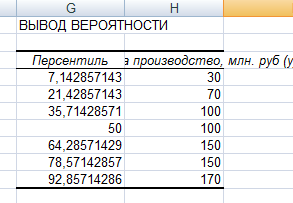


Рис. 1.30. Эмпирические значения *y*

## Анализ полученных данных

Рассчитанные в таблице на рисунке 1.28 коэффициенты регрессии позволяют построить уравнение, выражающее зависимость затрат на производство от выпуска продукции на i-том предприятии: .

Значение коэффициента детерминации *R2 = 0,9825* показывает, что вариация результата на 98,25% объясняется вариацией фактора *x*. Значит, выбранный фактор существенно влияет на затраты на производство, что подтверждает правильность их включения в построенную модель.

Рассчитанный уровень значимости *1,39294E-05 < 0,05 (*показатель ***Значимость F*** в таблице на рисунке 1.27 подтверждает значимость *R2 .*

Другой подход к проверке значимости *R2* основан на проверке попадания *Fфакт (*показатель *F* в таблице на рисунке 1.27) в критическую область (F0,05;1;5, +∞). Для рассматриваемого примера , который рассчитывался по формуле *=FРАСПОБР(0,05;1;5).* Т.к. *Fфакт = 280* попадает в критический интервал (6,61; +∞), то гипотеза H0: *R2 = 0* о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность*.*

Показатель средней ошибки аппроксимации  также подтверждает достаточно высокую адекватность построенного уравнения, т.к.  не превышает 8 – 10%.

Следующим этапом является проверка значимости коэффициентов регрессии. Сравнивая попарно элементы массивов *C53:C54* и *D53:D54* (рис. 1.28), видим, что абсолютное значение свободного члена *a* меньше, чем его стандартная ошибка. Таким образом, свободный член следует исключить из уравнения регрессии. Стандартная ошибка коэффициента регрессии b, меньше его абсолютного значения. К тому же этот коэффициент является значимым, о чем можно судить по значению показателя *P-значение* (рис. 1.28), которое меньше заданного уровня значимости α=0,05.

Другой распространенный способ проверки значимости коэффициентов регрессии основан на проверке попадания *tpасч* (показатель *t-статистика* в таблице на рис. 1.28) в критическую область . В нашем примере табличное значение *tтабл* рассчитано с помощью функции *=СТЬЮДРАСПОБР(0,05;5).* Табличное значение t-критерия  *tтабл* для числа степеней свободы *df=n - 2=5 и α=0,05* составит .

Так как попадает в критический интервал , то коэффициент регрессии *b* является значимым.

Подводя итог предварительному анализу уравнения регрессии, можно сделать вывод, что целесообразно пересчитать без свободного члена *a* , который не является статически значимым.

Полученные результаты, в целом, удовлетворительные.

## Задания для самостоятельной работы

**Задача 1.1.**

По территориям Центрального района известны данные за 2012 г.

|  |  |  |
| --- | --- | --- |
| Область | Доля денежных доходов, направленных на прирост сбережений во вкладах, займах, сертификатах и на покупку валюты, в общей сумме среднедушевого денежного дохода, %, *y* | Среднемесячная начисленная заработная плата, тыс.руб., *х* |
| Брянская обл. | 6,9 | 289 |
| Владимирская обл. | 8,7 | 334 |
| Ивановская обл. | 6,4 | 300 |
| Калужская обл. | 8,4 | 343 |
| Костромская обл. | 6,1 | 356 |
| Орловская обл. | 9,4 | 289 |
| Рязанская обл. | 11 | 341 |
| Смоленская обл. | 6,4 | 327 |
| Тверская обл. | 9,3 | 357 |
| Тульская обл. | 8,2 | 352 |
| Ярославская обл. | 8,6 | 381 |

**ТРЕБУЕТСЯ:**

1. Построить поле корреляции.
2. Рассчитать параметры уравнения линейной парной регрессии.
3. Оценить тесноту связи с помощью показателей корреляции и детерминации.
4. Дать с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оценить качество уравнений с помощью средней ошибки аппроксимации.
6. Оценить статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера.
7. На уровне значимости α = 0,05 оцените значимость коэффициентов регрессии.
8. Рассчитать ожидаемое значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определить доверительный интервал прогноза для уровня значимости α = 0,05.
9. Проверить полученные результаты с помощью стандартных статистических функций ТЕНДЕНЦИЯ, ЛИНЕЙН и программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.
10. Оценить полученные результаты.

**Задача 1.2.** По территориям Северного, Северо-Западного и Центрального районов известны данные за ноябрь 2011 г.

|  |  |  |
| --- | --- | --- |
| Область | Потребительские расходы на душу населения, тыс. руб., *y* | Денежные доходы на душу населения, тыс. руб., *x* |
| Республика Карелия | 596 | 913 |
| Республика Коми | 417 | 1095 |
| Архангельская обл. | 354 | 606 |
| Вологодская обл. | 526 | 876 |
| Мурманская обл. | 934 | 1314 |
| Ленинградская обл. | 412 | 593 |
| Новгородская обл. | 525 | 754 |
| Псковская обл. | 367 | 528 |
| Брянская обл. | 364 | 520 |
| Владимирская обл. | 336 | 539 |
| Ивановская обл. | 409 | 540 |
| Калужская обл. | 452 | 682 |
| Костромская обл. | 367 | 537 |
| Московская обл. | 328 | 589 |
| Орловская обл. | 460 | 626 |
| Рязанская обл. | 380 | 521 |
| Смоленская обл. | 439 | 626 |
| Тверская обл. | 344 | 521 |
| Тульская обл. | 401 | 658 |
| Ярославская обл. | 514 | 746 |

**ТРЕБУЕТСЯ**

1. Построить поле корреляции.
2. Рассчитать параметры уравнения линейной парной регрессии.
3. Оценить тесноту связи с помощью показателей корреляции и детерминации.
4. Дать с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оценить качество уравнений с помощью средней ошибки аппроксимации.
6. Оценить статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера.
7. На уровне значимости α = 0,05 оцените значимость коэффициентов регрессии.
8. Рассчитать ожидаемое значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определить доверительный интервал прогноза для уровня значимости α = 0,05.
9. Проверить полученные результаты с помощью стандартных статистических функций ТЕНДЕНЦИЯ, ЛИНЕЙН и программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.
10. Оценить полученные результаты.

**Задача 2.4.** По территориям Уральского и Западно-Сибирского районов известны данные за ноябрь 2011 г.

|  |  |  |
| --- | --- | --- |
| Область | Потребительские расходы на душу населения, тыс. руб., *y* | Денежные доходы на душу населения, тыс. руб., *x* |
| Республика Башкортостан | 461 | 632 |
| Удмуртская Республика | 524 | 738 |
| Курганская обл. | 298 | 515 |
| Оренбургская обл. | 351 | 640 |
| Пермская обл. | 624 | 942 |
| Свердловская обл. | 584 | 888 |
| Челябинская обл. | 425 | 704 |
| Республика Алтай | 277 | 603 |
| Алтайский край | 321 | 439 |
| Кемеровская обл. | 573 | 985 |
| Новосибирская обл. | 576 | 735 |
| Омская обл. | 588 | 760 |
| Томская обл. | 497 | 830 |
| Тюменская обл. | 863 | 2093 |

**ТРЕБУЕТСЯ**

1. Построить поле корреляции.
2. Рассчитать параметры уравнения линейной парной регрессии.
3. Оценить тесноту связи с помощью показателей корреляции и детерминации.
4. Дать с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оценить качество уравнений с помощью средней ошибки аппроксимации.
6. Оценить статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера.
7. На уровне значимости α = 0,05 оцените значимость коэффициентов регрессии.
8. Рассчитать ожидаемое значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определить доверительный интервал прогноза для уровня значимости α = 0,05.
9. Проверить полученные результаты с помощью стандартных статистических функций ТЕНДЕНЦИЯ, ЛИНЕЙН и программы РЕГРЕССИЯ из пакета анализа Microsoft Excel 2007.
10. Оценить полученные результаты.