Лабораторная работа: Регрессионный анализ в пакете SPSS

Выполнил:

Проверил: к.т.н. Панфилов Илья Александрович

Красноярск 20

Теоретические сведения

Регрессионный анализ — метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). [Регрессионная модель](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C) есть функция независимой переменной и параметров с добавленной [случайной переменной](http://www.machinelearning.ru/wiki/index.php?title=%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F_%D0%BF%D0%B5%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D0%B0%D1%8F&action=edit). Параметры модели настраиваются таким образом, что модель наилучшим образом приближает данные. Критерием качества приближения (целевой функцией) обычно является [среднеквадратичная ошибка](http://www.machinelearning.ru/wiki/index.php?title=%D0%A1%D1%80%D0%B5%D0%B4%D0%BD%D0%B5%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D0%B8%D1%87%D0%BD%D0%B0%D1%8F_%D0%BE%D1%88%D0%B8%D0%B1%D0%BA%D0%B0&action=edit): сумма квадратов разности значений модели и зависимой переменной для всех значений независимой переменной в качестве аргумента. Регрессионный анализ — раздел [математической статистики](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0) и [машинного обучения](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5). Предполагается, что зависимая переменная есть сумма значений некоторой модели и [случайной величины](http://www.machinelearning.ru/wiki/index.php?title=%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D0%B0). Относительно характера распределения этой величины делаются предположения, называемые гипотезой порождения данных. Для подтверждения или опровержения этой гипотезы выполняются [статистические тесты](http://www.machinelearning.ru/wiki/index.php?title=%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D1%82%D0%B5%D1%81%D1%82), называемые [анализом остатков](http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%BE%D1%81%D1%82%D0%B0%D1%82%D0%BA%D0%BE%D0%B2&action=edit). При этом предполагается, что независимая переменная не содержит ошибок. Регрессионный анализ используется для [прогноза](http://www.machinelearning.ru/wiki/index.php?title=%D0%9F%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7), [анализа временных рядов](http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%B2%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D1%85_%D1%80%D1%8F%D0%B4%D0%BE%D0%B2&action=edit), [тестирования гипотез](http://www.machinelearning.ru/wiki/index.php?title=%D0%A2%D0%B5%D1%81%D1%82%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%B3%D0%B8%D0%BF%D0%BE%D1%82%D0%B5%D0%B7&action=edit) и выявления скрытых взаимосвязей в данных.

Регрессия — зависимость [математического ожидания](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D0%BE%D0%B6%D0%B8%D0%B4%D0%B0%D0%BD%D0%B8%D0%B5) (например, среднего значения) случайной величины от одной или нескольких других случайных величин (свободных переменных), то есть E(y|\mathbf{x})=f(\mathbf{x}). Регрессионным анализом называется поиск такой функции f, которая описывает эту зависимость. Регрессия может быть представлена в виде суммы неслучайной и случайной составляющих.

 y=f(\mathbf{x})+\nu, 

где f — функция регрессионной зависимости, а \nu — аддитивная случайная величина с нулевым матожиданием. Предположение о характере распределения этой величины называется [гипотезой порождения данных](http://www.machinelearning.ru/wiki/index.php?title=%D0%93%D0%B8%D0%BF%D0%BE%D1%82%D0%B5%D0%B7%D0%B0_%D0%BF%D0%BE%D1%80%D0%BE%D0%B6%D0%B4%D0%B5%D0%BD%D0%B8%D1%8F_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85&action=edit). Обычно предполагается, что величина \nu имеет [гауссово распределение](http://www.machinelearning.ru/wiki/index.php?title=%D0%93%D0%B0%D1%83%D1%81%D1%81%D0%BE%D0%B2%D0%BE_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5&action=edit) с нулевым средним и дисперсией \sigma^2_\nu.

Задача нахождения регрессионной модели нескольких свободных переменных ставится следующим образом. Задана [выборка](http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0) — множество \{\mathbf{x}_1,...,\mathbf{x}_N|\mathbf{x}\in\mathbb{R}^M\}значений свободных переменных и множество \{y_1,...,y_N| y\in\mathbb{R}\}соответствующих им значений зависимой переменной. Эти множества обозначаются как D, множество исходных данных \{(\mathbf{x},y)_i\}. Задана [регрессионная модель](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C) — параметрическое семейство функций f(\mathbf{w},\mathbf{x}) зависящая от параметров \mathbf{w}\in\mathbb{R} и свободных переменных \mathbf{x}. Требуется найти наиболее вероятные параметры \bar{\mathbf{w}}:

\bar{\mathbf{w}}=\arg\max\limits_{\mathbf{w}\in\mathbb{R}^W}p(y|x,\mathbf{w},f)=p(D|\mathbf{w},f).

Функция вероятности pзависит от гипотезы порождения данных и задается [Байесовским выводом](http://www.machinelearning.ru/wiki/index.php?title=%D0%91%D0%B0%D0%B9%D0%B5%D1%81%D0%BE%D0%B2%D1%81%D0%BA%D0%B8%D0%B9_%D0%B2%D1%8B%D0%B2%D0%BE%D0%B4&action=edit) или [методом наибольшего правдоподобия](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%BD%D0%B0%D0%B8%D0%B1%D0%BE%D0%BB%D1%8C%D1%88%D0%B5%D0%B3%D0%BE_%D0%BF%D1%80%D0%B0%D0%B2%D0%B4%D0%BE%D0%BF%D0%BE%D0%B4%D0%BE%D0%B1%D0%B8%D1%8F).

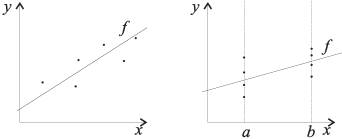


Рис. 1 Выборка может быть не функцией, а отношением. Например, данные для построения регрессии могут быть такими: \{(0,0),(0,1),(0,2),(1,1),(1,2),(1,3)\}. В такой выборке одному значению переменной x соответствует несколько значений переменной y.

Линейная регрессия

Линейная регрессия предполагает, что функция f зависит от параметров \mathbf{w} линейно. При этом линейная зависимость от свободной переменной \mathbf{x} необязательна,

 y=f(\mathbf{w},\mathbf{x})+\nu=\sum_{j=1}^N w_jg_j(\mathbf{x})+\nu. 

В случае, когда функция g\equiv\text{id} линейная регрессия имеет вид

 y=\sum_{j=1}^N w_jx_j+\nu=\langle\mathbf{w},\mathbf{x}\rangle +\nu, 

здесь x_j — компоненты вектора \mathbf{x}.

Значения параметров в случае линейной регрессии находят с помощью [метода наименьших квадратов](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%BD%D0%B0%D0%B8%D0%BC%D0%B5%D0%BD%D1%8C%D1%88%D0%B8%D1%85_%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D0%BE%D0%B2). Использование этого метода обосновано предположением о [гауссовском распределении](http://www.machinelearning.ru/wiki/index.php?title=%D0%93%D0%B0%D1%83%D1%81%D1%81%D0%BE%D0%B2%D1%81%D0%BA%D0%BE%D0%B5_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5&action=edit) случайной переменной.

Разности y_i-f(\mathbf{x}_i)между фактическими значениями зависимой переменной и восстановленными называются регрессионными остатками (residuals). В литературе используются также синонимы: *невязки* и *ошибки*. Одной из важных оценок критерия качества полученной зависимости является сумма квадратов остатков:

SSE=\|f(\mathbf{x}_i)-y_i\|_2=\sum_{i=1}^N(y_i-f(\mathbf{w},\mathbf{x}_i))^2.

Здесь SSE — Sum of Squared Errors.

Дисперсия остатков вычисляется по формуле

\bar{\sigma}^2_\nu=\frac{SSE}{N-2}=MSE.

Здесь MSE — Mean Square Error, среднеквадратичная ошибка.

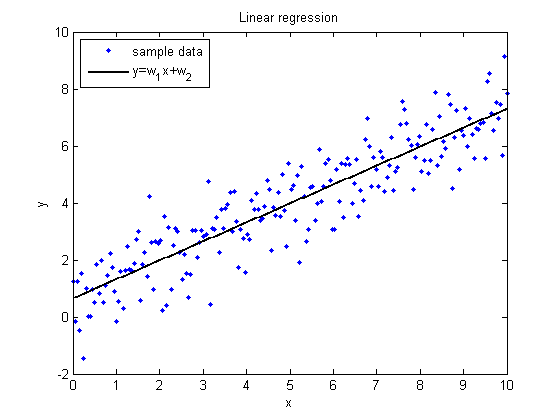


Рис. 2 Линия регрессии

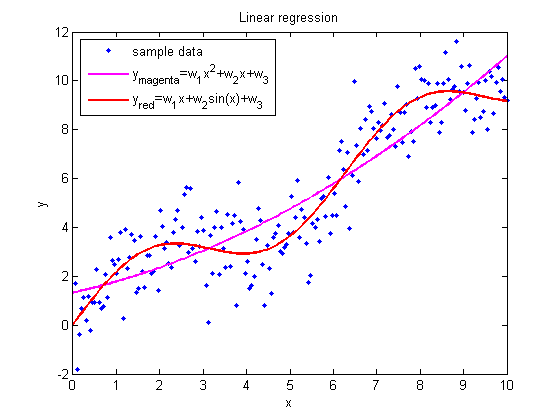


Рис. 3 Линия регрессии

На графиках представлены выборки, обозначенные синими точками, и регрессионные зависимости, обозначенные сплошными линиями. По оси абсцисс отложена свободная переменная, а по оси ординат — зависимая. Все три зависимости линейны относительно параметров.

## Нелинейная регрессия — частный случай [регрессионного анализа](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7), в котором рассматриваемая [регрессионная модель](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C) есть функция, зависящая от параметров и от одной или нескольких свободных переменных. Зависимость от параметров предполагается нелинейной.

Задана [выборка](http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0) из mпар (\mathbf{x}_i,y_i). Задана [регрессионная модель](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C) f(\mathbf{w},\mathbf{x}), которая зависит от параметров \mathbf{w}=(w_1,...,w_W) и свободной переменной x. Требуется найти такие значения параметров, которые доставляли бы минимум сумме квадратов [регрессионных остатков](http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D1%85_%D0%BE%D1%81%D1%82%D0%B0%D1%82%D0%BA%D0%BE%D0%B2)

S=\sum_{i=1}^mr_i^2,

где остатки r_i=y_i-f(\mathbf{w},\mathbf{x}_i) для i=1,\ldots,m.

Для нахождения минимума функции S, приравняем к нулю её первые частные производные параметрам \mathbf{w}:

\frac{\partial S}{\partial w_j}=2\sum_i r_i\frac{\partial r_i}{\partial w_j}=0 \ (j=1,\ldots,n). (*)

Так как функция S в общем случае не имеет единственного минимума, то предлагается назначить начальное значение вектора параметров w_0 и приближаться к оптимальному вектору по шагам:

w_j \approx w_j^{k+1} =w^k_j+\Delta w_j.

Здесь k - номер итерации, \Delta w_j - вектор шага.

На каждом шаге итерации линеаризуем модель с помощью приближения [рядом Тейлора](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D1%8F%D0%B4_%D0%A2%D0%B5%D0%B9%D0%BB%D0%BE%D1%80%D0%B0&action=edit) относительно параметров \mathbf{w}^k Здесь элемент матрицы Якоби J_{ij} - функция параметра w_j; значение свободной переменной \mathbf{x}_iфиксировано. В терминах линеаризованной модели

\frac{\partial r_i}{\partial w_j}=-J_{ij}

и регрессионные остатки определены как

r_i=\Delta y_i- \sum_{j=1}^{n} J_{ij}\Delta w_j; \ \Delta y_i=y_i- f(x_i,\mathbf{w}^k).

Подставляя последнее выражение в выражение (\*), получаем

-2\sum_{i=1}^{m}J_{ij} \left( \Delta y_i-\sum_{s=1}^{n} J_{is}\Delta w_s \right)=0.

Преобразуя, получаем систему из n линейных уравнений, которые называются [нормальным уравнением](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%BD%D0%B0%D0%B8%D0%BC%D0%B5%D0%BD%D1%8C%D1%88%D0%B8%D1%85_%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D0%BE%D0%B2)

\sum_{i=1}^{m}\sum_{s=1}^{n} J_{ij}J_{is}\Delta w_s=\sum_{i=1}^{m} J_{ij}\Delta y_i (j=1,n).

Запишем нормальное уравнение в матричном обозначении как

\mathbf{\left(J^TJ\right)\Delta \mathbf{w}=J^T\Delta y}.

В том случае, когда критерий оптимальности регрессионой модели задан как взвешенная сумма квадратов остатков

S=\sum_{i=1}^{m}W_{ii}r_i^2,

нормальное уравнение будет иметь вид

\mathbf{\left(J^TWJ\right)\Delta \mathbf{w}=J^TW\Delta y}.

Для нахождения оптимальных параметров нелинейных регрессионных моделей используются [метод сопряжённых градиентов](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D1%81%D0%BE%D0%BF%D1%80%D1%8F%D0%B6%D1%91%D0%BD%D0%BD%D1%8B%D1%85_%D0%B3%D1%80%D0%B0%D0%B4%D0%B8%D0%B5%D0%BD%D1%82%D0%BE%D0%B2), [метод Ньютона-Гаусса](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%9D%D1%8C%D1%8E%D1%82%D0%BE%D0%BD%D0%B0-%D0%93%D0%B0%D1%83%D1%81%D1%81%D0%B0) или [алгоритм Левенберга-Марквардта](http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC_%D0%9B%D0%B5%D0%B2%D0%B5%D0%BD%D0%B1%D0%B5%D1%80%D0%B3%D0%B0-%D0%9C%D0%B0%D1%80%D0%BA%D0%B2%D0%B0%D1%80%D0%B4%D1%82%D0%B0).

Постановка задачи

Существует организация, включающая в себя 3 филиала, которая занимается оптовой и розничной продажей стройматериалов. Подразделения разнесены географически. В этих городах разный социальный уровень, количество жителей и потребности тоже весьма отличаются. Для успешного ведения бизнеса требуется система, которая будет прогнозировать объем продаж, а так же давать рекомендации по реорганизации каждого подразделения и улучшения его деятельности и на основе имеющихся данных.

Прогнозом продаж фирмы можно считать предвидение будущих операционных доходов предприятия на протяжении определённого учётного периода и на основе предвидимой информации о действии факторов внутренней и внешней среды компании и исторических данных предыдущих периодов при условии, что устанавливается также допустимое несоответствие предвидимых будущих доходов и фактического их значения на основе вероятности.

Прежде чем начать регрессионный анализ, необходимо определить, какие факторы стоит учитывать на входе для решения поставленной задачи, чтобы понять, что мы ожидаем на выходе. Сначала определяется количество ресурсов, которые существенно оказывают влияние на выходной параметр. В качестве выходного параметра рассматривается выручка. В процессе анализа было выделено три основных параметра, влияющих на выручку с учетом специфики данного предприятия.

1) На каждый месяц расчитываеться сумма вырученных денежных средств сумма средств полученных от основных контрагентов. Основных контрагентов получаем при помощи реализации АВС анализа, в нашем случае берем 5 наиболее ярких представителей это те чья общая сумма полученная от реализации товара свыше 2 млн рублей за все существование организации. После выдвигаем гипотезу о влияние доли (суммы получаемый от основных контрагентов) на сумму получаемую в целом рис.4 (Выручку).

Так же представлен фрагмент отчета из которого видно, что лучшими в данной ситуации будут покупателе с общей выручкой за анализируемый период будут такие фирмы как ООО «ФАСАД», ООО ЭФП «Пилон», ООО «Мебель деталь», ООО «Строймастер», ООО «Сибирская тайга» с выручкой при отгрузке в более 2 млн. рублей (фрагмент отчете 1). Основной покупатель в данном случае является объединением суммы полученной от розничных покупателей. Анализировать используя эту цифру является не рентабельным ввиду того что фирма больше ориентирована на постоянных покупателей.

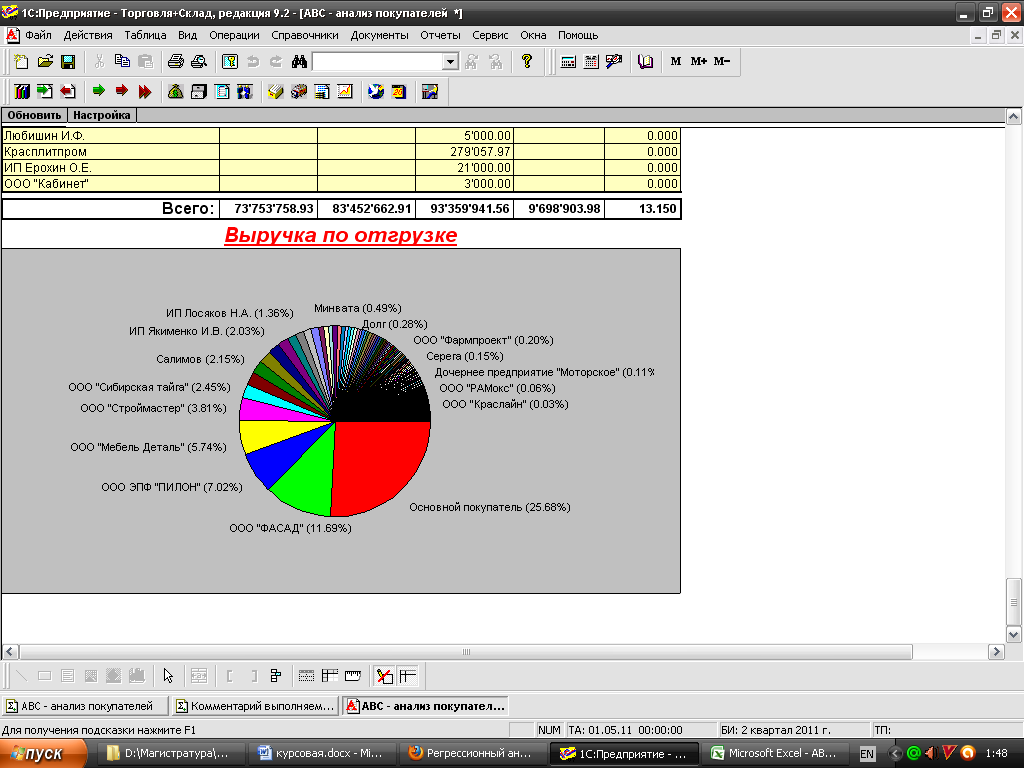


Рис.5 Клиенты фавориты

Фрагмент отчета 1 -ABC- анализ клиентов.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Покупатель | Себестоимость продаж | Выручка по отгрузке | Получена оплата | Прибыль по отгрузке | Прибыль- ность (%) |
| Основной покупатель | 19 215 193,99 | 21427 458,00 | 21427 599,00 | 2212 264,01 | 11,513 |
| ООО "ФАСАД" | 8 213 488,21 | 9 757 443,76 | 9 424 019,40 | 1543 955,55 | 18,798 |
| ООО ЭПФ "ПИЛОН" | 5 478 259,77 | 5 855 409,00 | 5 441 934,00 | 377 149,23 | 6,884 |
| ООО "Мебель Деталь" | 4 253 564,82 | 4 791 560,00 | 4 791 560,00 | 537 995,18 | 12,648 |
| ООО "Строймастер" | 2 677 768,29 | 3 181 618,80 | 3 393 633,50 | 503 850,51 | 18,816 |
| ООО "Сибирская тайга" | 1 790 751,15 | 2 046 154,00 | 2 019 924,00 | 255 402,85 | 14,262 |
| ООО ТД "Сибирь" | 1 780 722,03 | 1 860 977,00 | 524 382,00 | 80 254,97 | 4,507 |
| Салимов | 1 579 857,85 | 1 791 213,00 | 668 777,50 | 211 355,15 | 13,378 |
| ООО"Вира" | 1 389 061,76 | 1 760 919,90 | 1 760 919,90 | 371 858,14 | 26,770 |
| ИП Якименко И.В. | 1 487 943,83 | 1 691 770,00 | 1 569 000,00 | 203 826,17 | 13,699 |
| Антипин П. | 1 456 494,75 | 1 575 299,00 | 1 581 599,00 | 118 804,25 | 8,157 |
| ИП Резинкин А.И. | 1 073 710,67 | 1 194 350,00 | 1 194 350,00 | 120 639,33 | 11,236 |

2) Точно так же выявляем наиболее значимые единицы товаров, и строим гипотезу о влиянии суммы продаваемых «основных» товаров на полученные средства в целом рисунок 6.

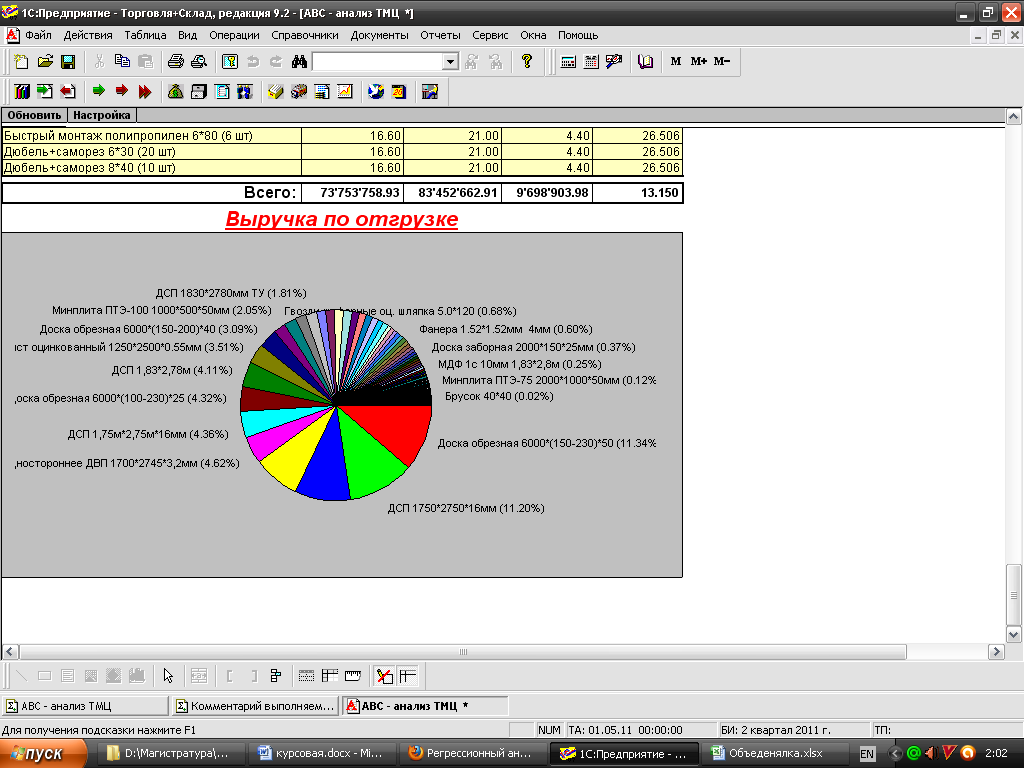


Рисунок 6

Проведя анализ товаров, были выявлены такие товары выручка при отгрузке которых составила более 6 млн. рублей (фрагмент отчета 2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ТМЦ | Себестоимость продаж | Выручка по отгрузке | Прибыль по отгрузке | Прибыль- ность (%) |
| Доска обрезная 6000\*(150-230)\*50 | 8 296 827,88 | 9467 239,70 | 1170411,82 | 14,107 |
| ДСП 1750\*2750\*16мм | 8 305 334,25 | 9 349 005,00 | 1 043 670,75 | 12,566 |
| ДВП 1.22\*2.75\*2.5мм | 7 455 316,00 | 7 923 296,00 | 467 980,00 | 6,277 |
| 1220\*2710\*6мм Двустороннее ДВП | 6 220 765,00 | 6 574 684,00 | 353 919,00 | 5,689 |
| Одностороннее ДВП 1700\*2745\*3,2мм | 3 638 651,00 | 3 858 596,16 | 219 945,16 | 6,045 |
| ДСП 1,75м\*2,75м\*16мм | 3 221 635,00 | 3 642 285,00 | 420 650,00 | 13,057 |
| Доска обрезная 6000\*(100-230)\*25 | 3 138 686,82 | 3 608 004,51 | 469 317,69 | 14,953 |
| ДСП 1,83\*2,78м | 2 852 501,37 | 3 432 850,00 | 580 348,63 | 20,345 |
| Лист оцинкованный 1250\*2500\*0.55мм | 2 530 030,00 | 2 925 965,00 | 395 935,00 | 15,649 |
| Доска обрезная 6000\*(150-200)\*40 | 2 276 049,90 | 2 577 772,60 | 301 722,70 | 13,256 |
| Минплита ПТЭ-100 1000\*500\*50мм | 1 407 327,00 | 1 714 427,50 | 307 100,50 | 21,822 |
| Брусок 50\*50 | 1 312 708,00 | 1 690 068,84 | 377 360,84 | 28,747 |

3)Затраты производимые организацией ежемесячно так же влияют на общую выручку.

Данные показатели в совокупности наилучшим образом отражаются на итоговой деятельности предприятия.

Для выявления групп наиболее значимых товаров и клиентов использовался элементарный ABC анализ. ABC-анализ — метод, позволяющий классифицировать ресурсы фирмы по степени их важности. В его основе лежит принцип Парето — 20 % всех товаров дают 80 % оборота. По отношению к ABC-анализу правило Парето может прозвучать так: надёжный контроль 20 % позиций позволяет на 80 % контролировать систему, будь то запасы сырья и комплектующих, либо продуктовый ряд предприятия и т. п. Часто ABC-анализ путают с ABC-методом, расшифровывая ABC как Activity Based Costing, что в корне не верно.

ABC-анализ — анализ товарных запасов путём деления на три категории:

А — наиболее ценные, 20 % — тов.запасов; 80 % — продаж

В — промежуточные, 30 % — тов.запасов; 15 % — продаж

С — наименее ценные, 50 % — тов.запасов; 5 % — продаж

Реализация регрессионного анализа в SPSS

SPSS Statistics (аббревиатура англ. «Statistical Package for the Social Sciences» — «статистический пакет для социальных наук») — компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в социальных науках. Проведем регрессионный анализ в программе SPSS. Пакет SPSS оснащен в полнее интуитивно понятным интерфейсом, в нем предложено несколько разновидностей регрессионного анализа для начала воспользуемся простой множественной линейной регрессией.

Алгоритм проведения анализа:

1. Сначала выгрузим данные, с помощью которых будем проводить анализ (рис.7).

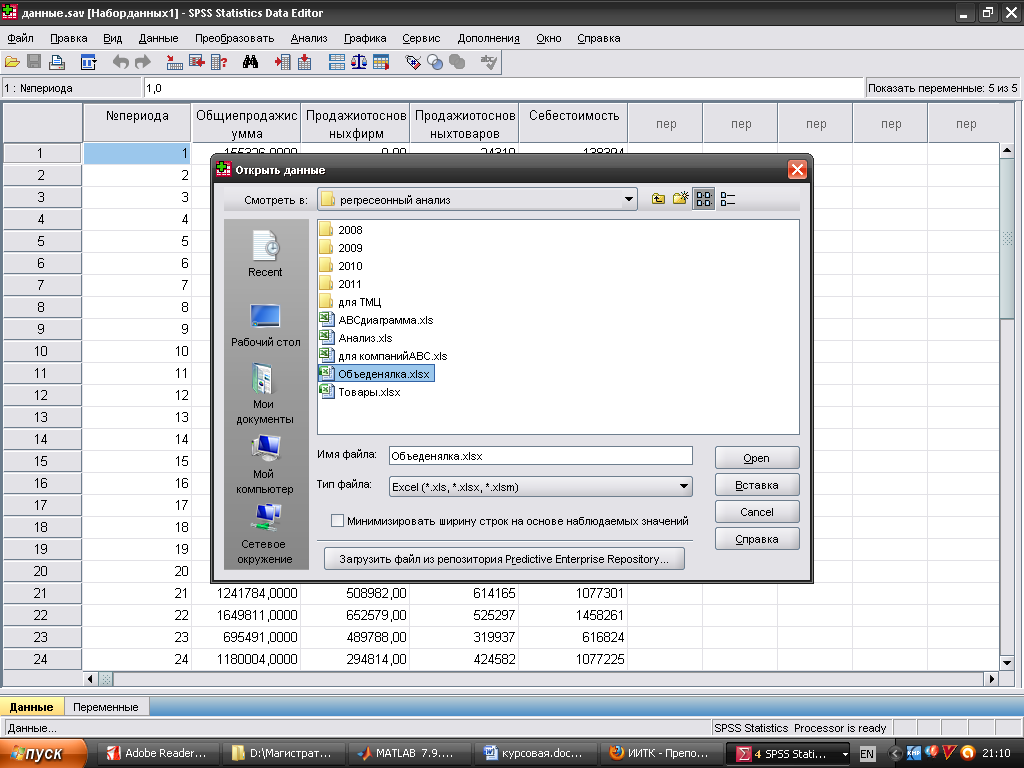


Рисунок 7. Открываем файл с данными.

1. После того как данные выгружены заходим в верхнее меню во вкладку *Анализ\_Регрессия\_Линейная,* как показано на рисунке 8.

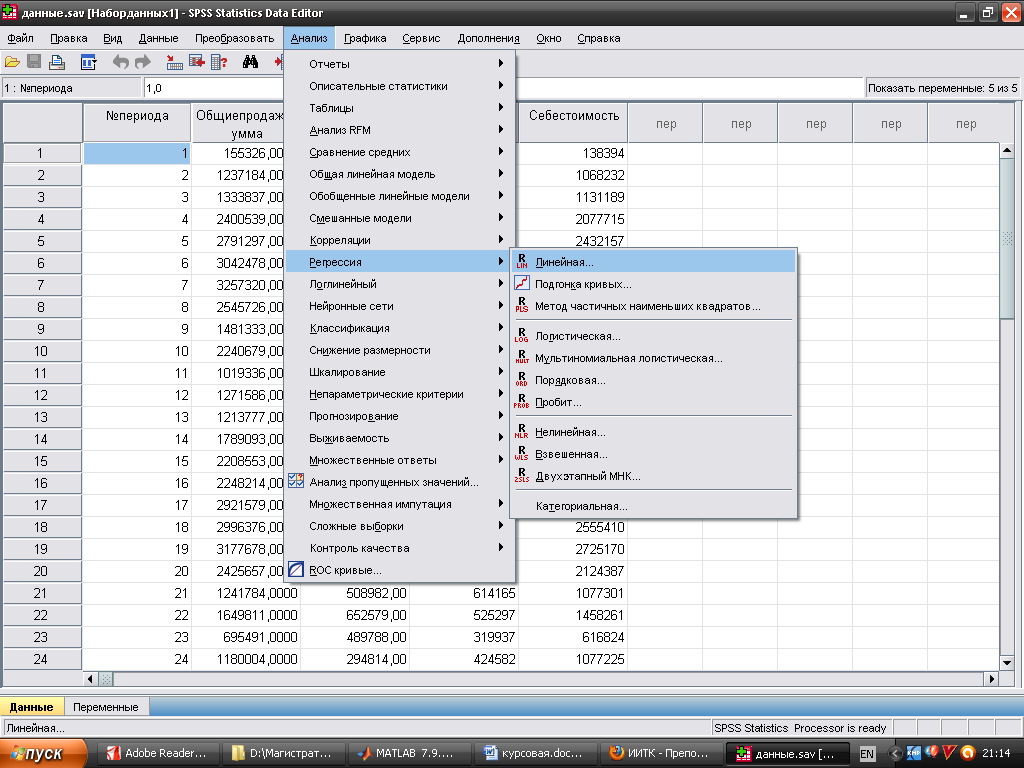


Рисунок 8.Выбор анализа данных.

Этот метод соответствует одновременной обработке всех независимых переменных, выбранных для анализа, и поэтому он может рекомендоваться для использования только в случае простого анализа с одной независимой переменной. Для множественного анализа следует выбрать один из пошаговых методов. При прямом методе независимые переменные, которые имеют наибольшие коэффициенты частичной корреляции с зависимой переменной пошагово увязываются в регрессионное уравнение.

1. Далее появляется диалоговое окно, в котором указываем в строке зависимые переменные выручка, в строке не зависимые переменные все остальные кроме номера периода как на рисунке 9.

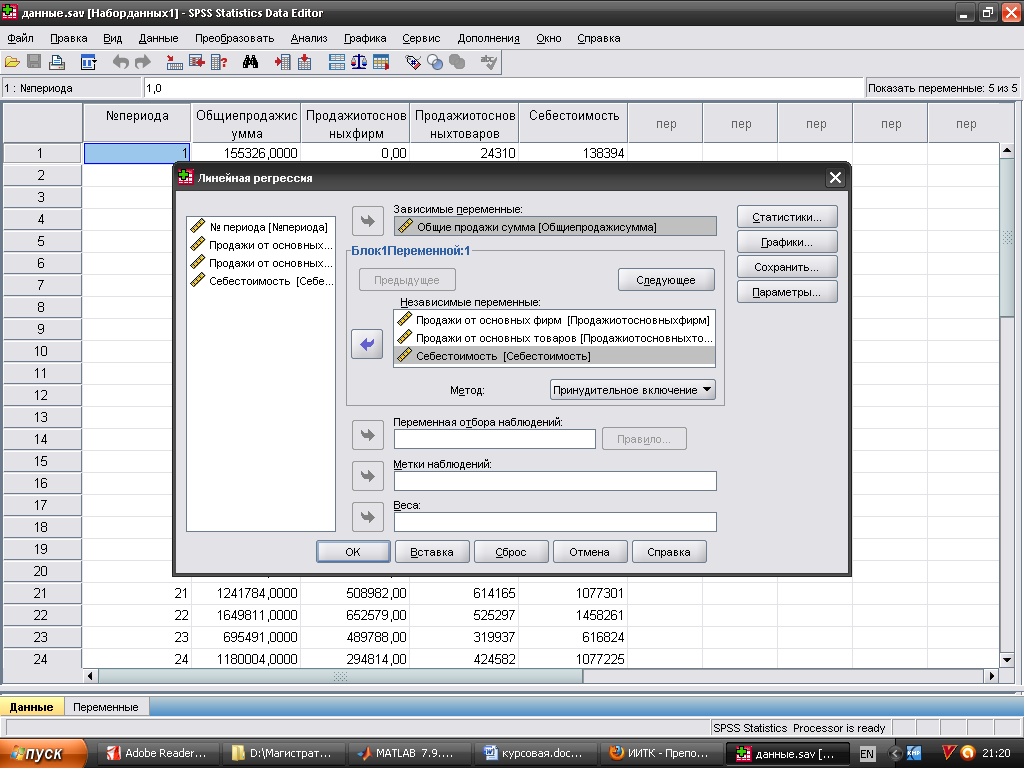


Рисунок 9 параметры для линейной регрессии.

Представим результаты первой таблицы 6:

1. Заходим в пункт статистика выбираем все параметры как на рисунке 10.

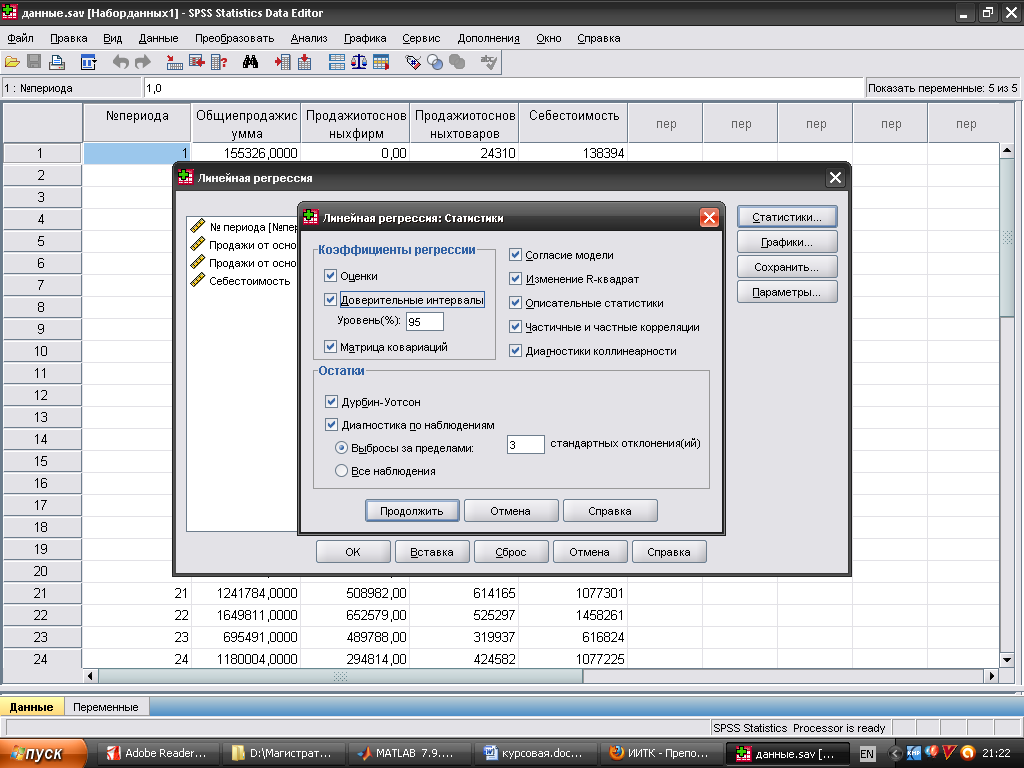


Рисунок 10 Выбор параметров

1. Во вкладке графики выбираем все как показано на рисунке 11.

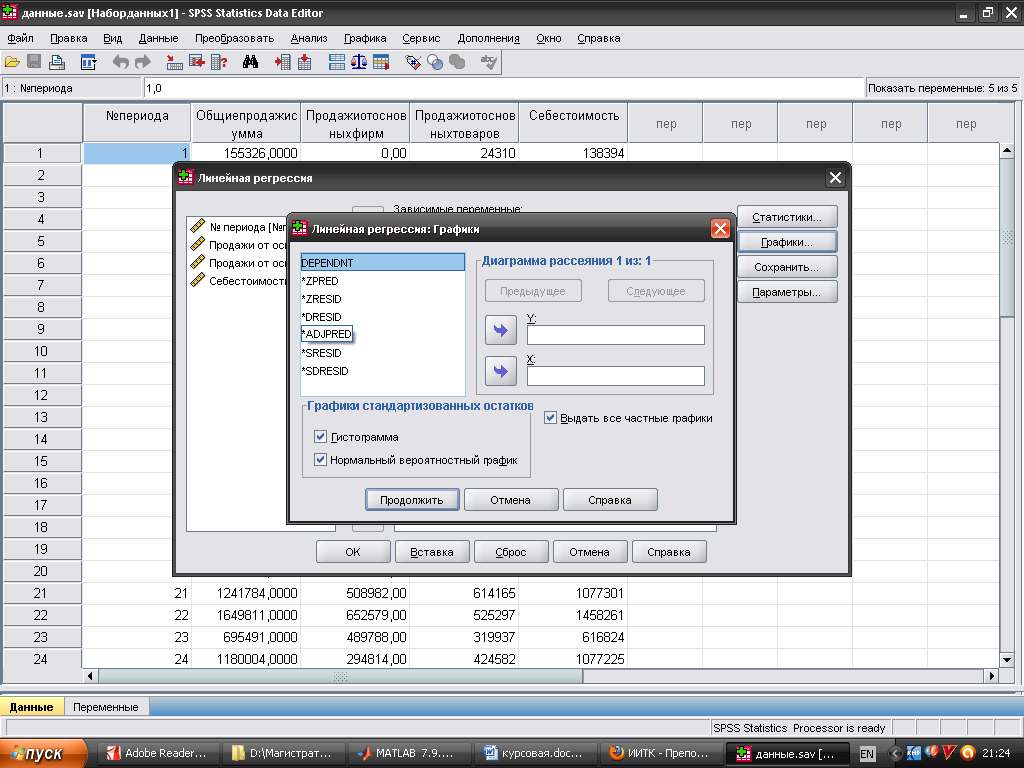


Рисунок 11 Графики

После чего нажимаем кнопку ОК.

1. Что бы построить график рассевания с линией тренда заходим на страницу в SPSS с нашими данными выбираем вкладку графика\_устаревшие диалоговые окна\_ рассевание точки, как показано на рисунке 12.

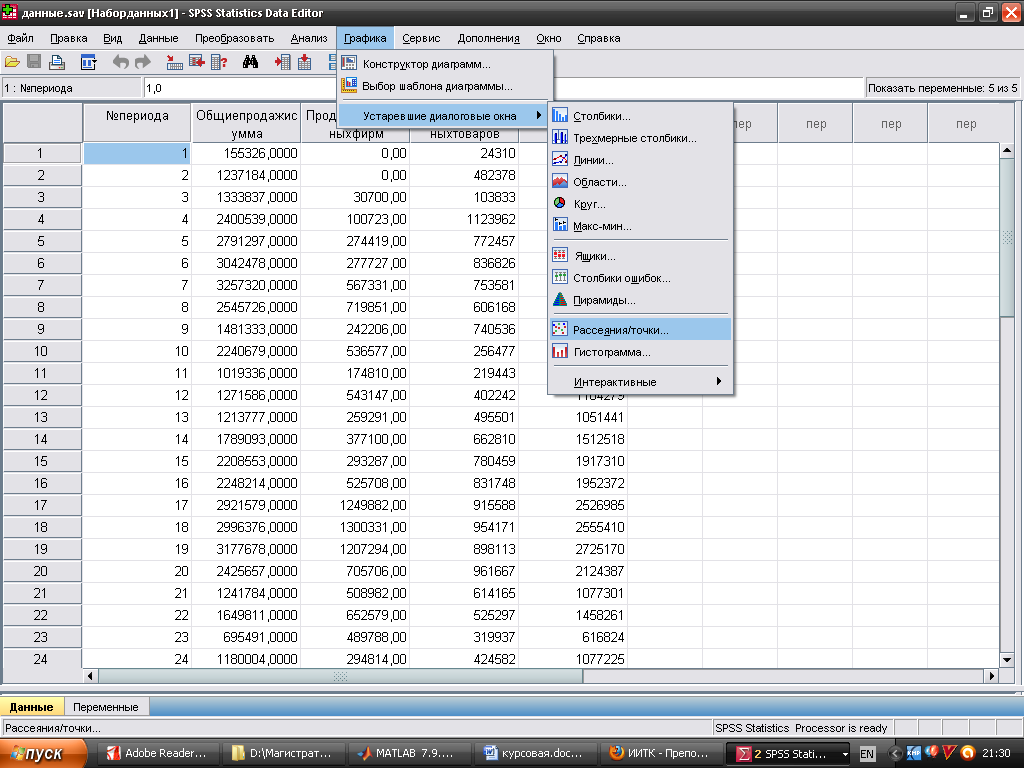


Таблица 7 Выбор графика

Далее выбираем:

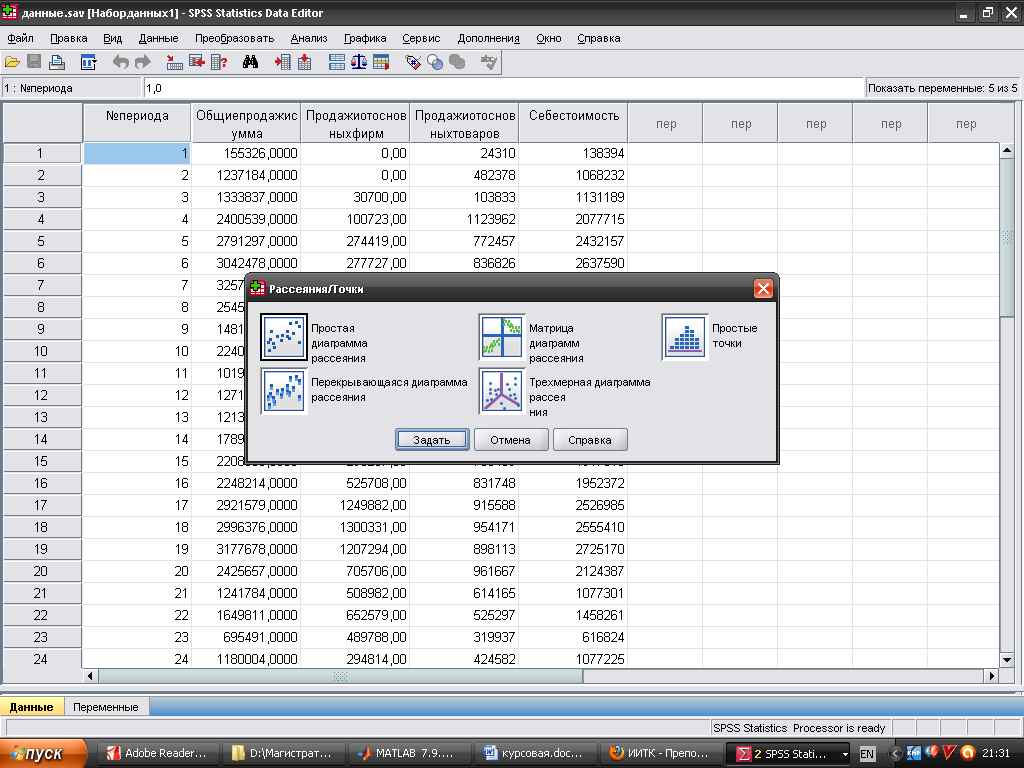


Рисунок 8

Задаем значения:

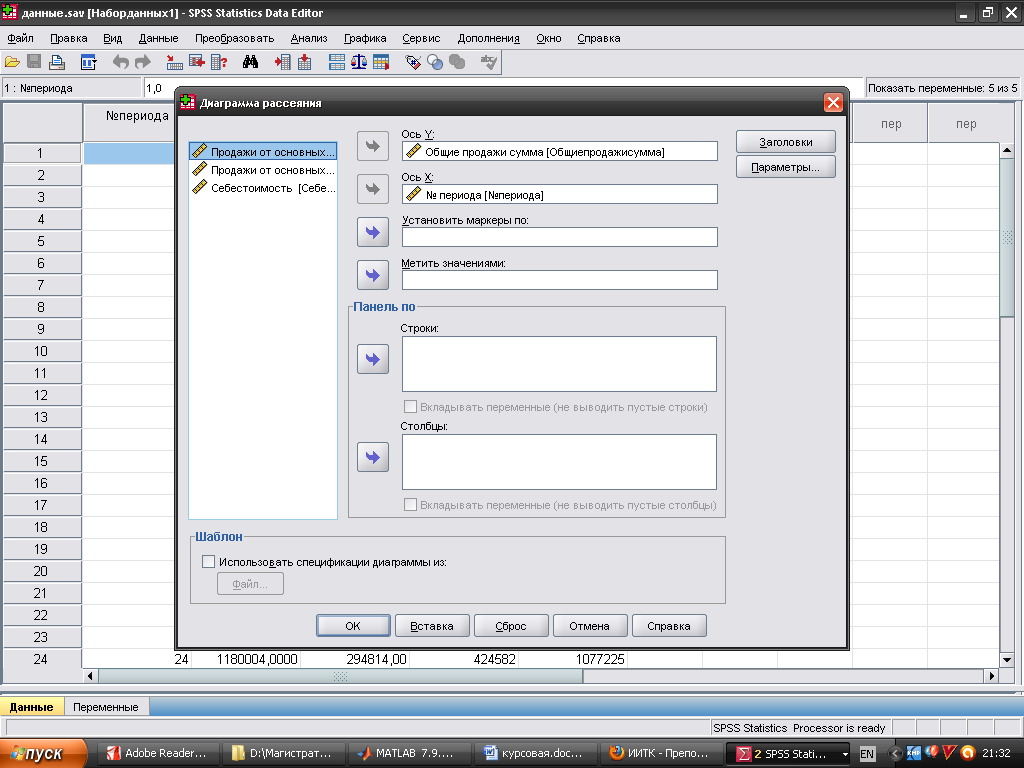


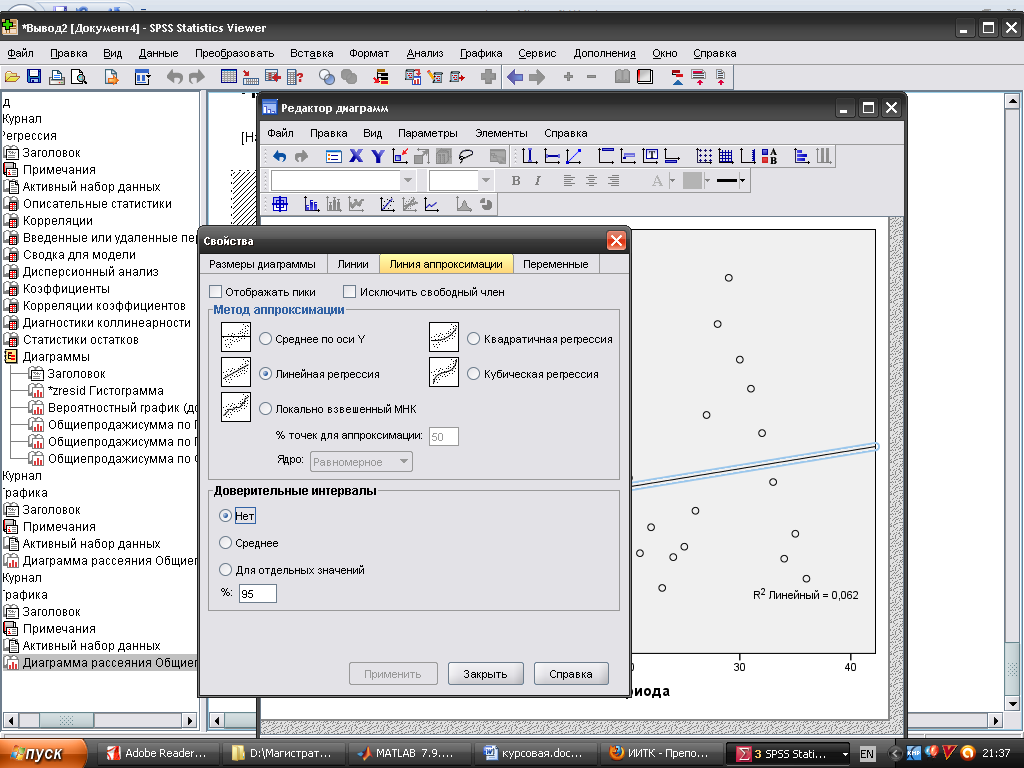
Рисунок 9

Далее на странице где приведены результаты регрессионного анализа будет выведен график 1:



График 1

Далее, нажимаем левой кнопкой мыши по графику выбираем: Изменить содержимое\_ В отдельном окне. Далее нажимам на значок свойства и выбираем такую линию, которая будет наилучшим способом описывать процесс. Пусть в этом случае у нас буде Линейная регрессия. (рис10)



| Рисунок 10 Выбор линии тренда  Далее разберемся с полученными результатами | | | |
| --- | --- | --- | --- |
|  | Для среднего | Стд. Отклонение | M |
| Общие продажи cумма | 2,243514E6 | 1,2033653E6 | 36 |
| Продажи от основных фирм | 666030,2500 | 5,79745E5 | 36 |
| Продажи от основных товаров | 958708,78 | 761318,827 | 36 |
| Себестоимость | 2089139,06 | 1164624,980 | 36 |

Далее так же следует таблица корреляции:

Таблица 2 Корреляция

| Корреляции | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  | Общие продажи cумма | Продажи от основных фирм | Продажи от основных товаров | Себестоимость |
| Корреляция Пирсона | Общие продажи cумма | 1,000 | ,810 | ,846 | ,940 |
| Продажи от основных фирм | ,810 | 1,000 | ,826 | ,772 |
| Продажи от основных товаров | ,846 | ,826 | 1,000 | ,894 |
| Себестоимость | ,940 | ,772 | ,894 | 1,000 |
| Знч. (1-сторонняя) | Общие продажи cумма | . | ,000 | ,000 | ,000 |
| Продажи от основных фирм | ,000 | . | ,000 | ,000 |
| Продажи от основных товаров | ,000 | ,000 | . | ,000 |
| Себестоимость | ,000 | ,000 | ,000 | . |
| M | Общие продажи cумма | 36 | 36 | 36 | 36 |
| Продажи от основных фирм | 36 | 36 | 36 | 36 |
| Продажи от основных товаров | 36 | 36 | 36 | 36 |
| Себестоимость | 36 | 36 | 36 | 36 |

Как мы видим по верхней характеристики коэффициента Пирсона выходящее значение выручка в значительной степени зависит от всех выявленных нами влияющих на нее параметров.

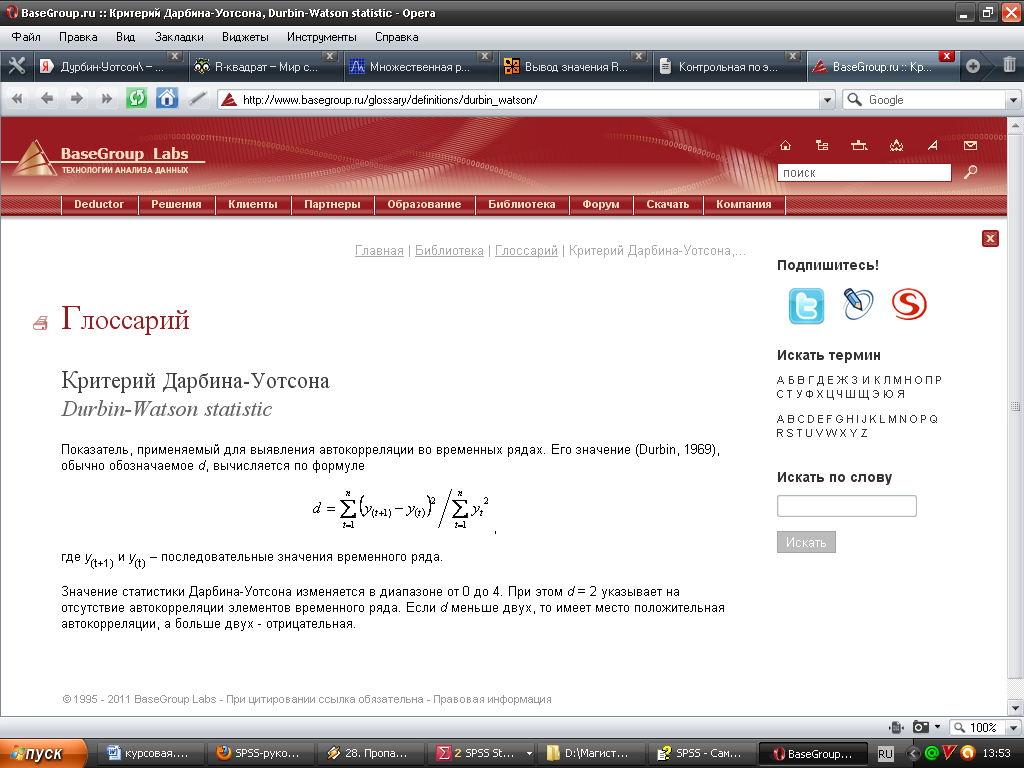
Таблица 8

| Сводка для моделиb | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель | Н | R-квадрат | Скорректированный R-квадрат | Стд. ошибка оценки | Изменения статистик | | | | | Дурбин-Уотсон |
| Изменение R квадрат | изменения F | ст.св.1 | ст.св.2 | Знч. изменения F |
| 1 | ,951a | ,904 | ,895 | 3,9007350E5 | ,904 | 100,366 | 3 | 32 | ,000 | 1,739 |
| a. Предикторы: (конст) Себестоимость , Продажи от основных фирм , Продажи от основных товаров | | | | | | | | | | |
| b. Зависимая переменная: Общие продажи cумма | | | | | | | | | | |

Поясним некоторые значения таблицы3.

R-квадрат показывает процентную долю колебаний стоимости, объясняемых колебаниями индекса. Значение может находиться в диапазоне от 0 до 100.

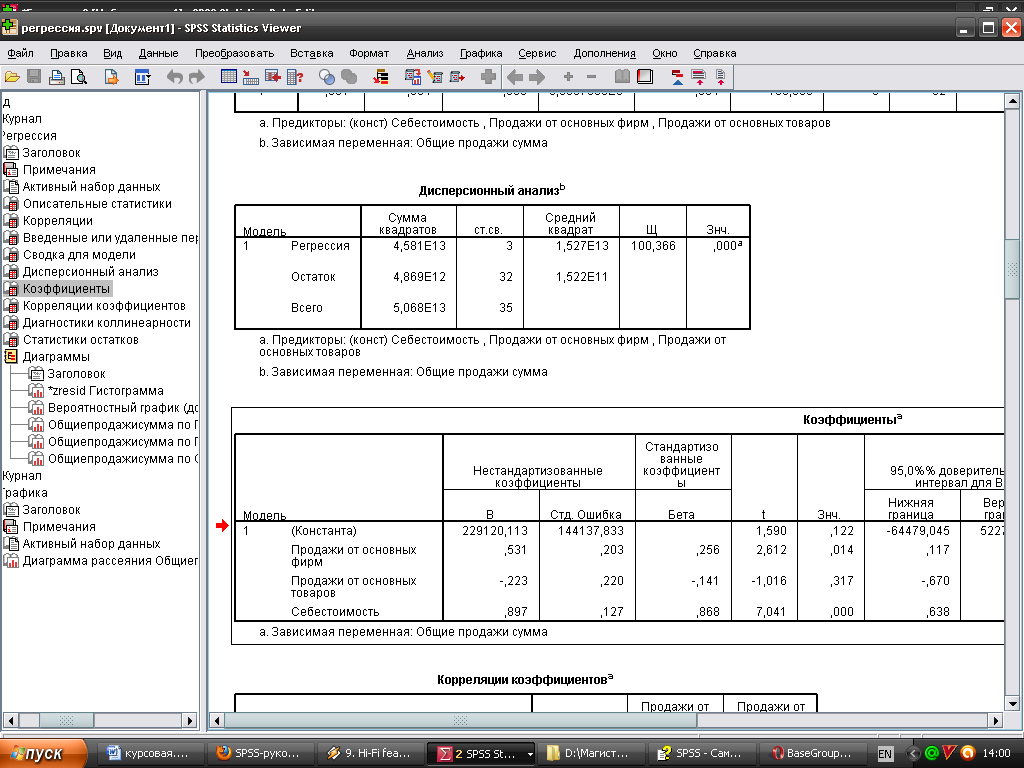
Дурбин-Уотсон показатель, применяемый для выявления автокорреляции во временных рядах. Его значение (Durbin, 1969), обычно обозначаемое d, вычисляется по формуле



где y(t+1) и y(t) – последовательные значения временного ряда.

Значение статистики Дарбина-Уотсона изменяется в диапазоне от 0 до 4. При этом d = 2 указывает на отсутствие автокорреляции элементов временного ряда. Если d меньше двух, то имеет место положительная автокорреляции, а больше двух - отрицательная.

Таблица 9 Коэффициенты



Получаем уравнение регрессии из таблицы 9:

Y= x1\*0,531+ x2\*-0,223+ x3\*0,897+229120

**В данном случае бета-коэффициент** — показатель, характеризующий общее влияние на динамику.

Проведем еще один вид регрессионного анализа называемого Взвешенным методом наименьших квадратов.

Получаем следующие значения таблица 10 сводка по модели:

Таблица 10 сводка по модели

| Сводка по модели | |
| --- | --- |
| Множественный R | ,986 |
| R-квадрат | ,972 |
| Скорректированный R-квадрат | ,969 |
| Стд.ошибка оценки | 48430,155 |
| Значение функции максимального правдоподобия | -509,235 |

По результатам видно, что значения R квадрата и скорректированного R квадрата увеличились в методе последнем методе.

Таблица 11 коэффициенты имеет следующий вид:

| Коэффициенты | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Нестандартизованные коэффициенты | | Стандартизованные коэффициенты | | t | Знач. |
|  | B | Стд. ошибка | Бета | Стд. ошибка |
| (Константа) | 29848,903 | 48599,869 |  |  | ,614 | ,543 |
| Продажиотосновныхфирм | ,110 | ,150 | ,033 | ,044 | ,735 | ,468 |
| Продажиотосновныхтоваров | -,229 | ,168 | -,096 | ,070 | -1,367 | ,181 |
| Себестоимость | 1,178 | ,078 | 1,048 | ,070 | 15,017 | ,000 |

Уравнение регрессии в этом случае выглядит так:

Y= x1\*0,110+ x2\*-0,229+ x3\*1,178+29848

Ниже представлен прогноз на несколько периодов вперед:

Таблица 12 прогноз

|  |  |
| --- | --- |
| № периода | прогноз |
| 1 | 187310,0177 |
| 2 | 1177761,62 |
| 3 | 1341987,865 |
| 4 | 2231089,219 |
| 5 | 2748223,089 |
| 6 | 2975846,005 |
| 7 | 3243710,017 |
| 8 | 2470526,991 |
| 9 | 2181999,874 |
| 10 | 1454068,976 |
| 11 | 1032076,669 |
| 12 | 1298322,384 |
| 13 | 1183498,177 |
| 14 | 1701292,656 |
| 15 | 2141976,003 |
| 16 | 2197100,806 |
| 17 | 2934454,173 |
| 18 | 2964653,255 |
| 19 | 3167233,3 |

Важным моментом является анализ остатков, то есть отклонений наблюдаемых значений от теоретически ожидаемых. Остатки должны появляться случайно (то есть не систематически) и подчиняться нормальному распределению. Это можно проверить, если с помощью кнопки Charts... (Диаграммы) построить гистограмму остатков. В приведенном примере наблюдается довольно хорошее согласование гистограммы остатков с нормальным распределением.



График 4 Гистограмма остатков

Множественная линейная регрессия

Y= x1\*0,531+ x2\*-0,223+ x3\*0,897+229120

Взвешенный метод наименьших квадратов

Y= x1\*0,110+ x2\*-0,229+ x3\*1,178+29848

В итоге был построен прогноз:

Таблица 13 Сравнение полученных прогнорзов

|  |  |
| --- | --- |
| Прогноз линейная регрессия | Прогноз взвешенный метод |
| 347861,5676 | 187310,0177 |
| 1079950,618 | 1177761,62 |
| 1237130,653 | 1341987,865 |
| 1896093,726 | 2231089,219 |
| 2384722,537 | 2748223,089 |
| 2556434,3 | 2975846,005 |
| 2894073,363 | 3243710,017 |
| 2380631,44 | 2470526,991 |
| 1940627,008 | 2181999,874 |
| 1541462,44 | 1454068,976 |
| 1059984,682 | 1032076,669 |
| 1418714,148 | 1298322,384 |
| 1199715,309 | 1183498,177 |
| 1638662,023 | 1701292,656 |
| 1931066,541 | 2141976,003 |
| 2074568,077 | 2197100,806 |
| 2956126,261 | 2934454,173 |
| 2999829,259 | 2964653,255 |
| 3115198,302 | 3167233,3 |
| 2295556,644 | 2389782,262 |
| 1329115,638 | 1214253,19 |
|  |  |

На графике 7 можно увидеть полученные данные в графической интерпретации.

График 7

Прогноз 1 Множественная линейная регрессия

Прогноз 2 Взвешенный метод наименьших квадратов