**ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ ИМ. ПРОФ. М.А. БОНЧ-БРУЕВИЧА»**

**(СПбГУТ)**

**Факультет Информационных систем и технологий**

**Кафедра Автоматизации предприятий связи**

**ОТЧЕТ**

по лабораторной работе №1 на тему:

**«Оптимизация плана перевозок продукта от поставщиков к потребителям по критерию минимизации транспортных издержек»**

по дисциплине «**Системный анализ и принятие решений**»

**Исполнитель:**

студент 3 курса, группы ИСТ-831

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/Пономарев Е.И./

«\_\_\_»\_\_\_\_\_\_\_\_2020

**Проверил:**

д.т.н., доцент

\_\_\_\_\_\_\_\_\_\_\_\_/Бухарин В.В. /

«\_\_\_»\_\_\_\_\_\_\_\_2020

Санкт-Петербург

2020

### Цель выполнения работы

Цель выполнения лабораторной работы – освоение методологии краткосрочного прогнозирования объема продаж продукции с использованием моделей и методов регрессионного анализа и программы Statistica.

### Постановка задачи

Имеется совокупность результатов наблюдений за поведением переменной *Y* в зависимости от изменения одной или нескольких независимых переменных *X (X1, X2, …, Xn)*. Необходимо установить количественную взаимосвязь между показателем *Y* и факторами *X*, т.е. определить такую функциональную зависимость*Y\*=f(X1,X2, …, Xn)*, которая наилучшим образом описывает имеющиеся экспериментальные данные.На основании построенного уравнения регрессии требуется спрогнозировать значение зависимой переменной*Y*на шаг вперед (момент времени (*t+1*)) при условии, что значения влияющих факторов на этот периодизвестны.

### Описание метода решения задачи

Математическое уравнение, которое описывает линию простой (парной) линейной регрессии с учетом влияния одного фактора, имеет вид:

*Yt+1=b0+b1X*,

где*Yt+1* – прогнозное значение зависимой переменной на момент времени (*t+1*);

*b0,b1* – параметры, которые оцениваются на основе статистических данных (угловые коэффициенты или коэффициенты регрессии);

*X* – значение влияющего фактора (независимая переменная).

Однофакторная линейная регрессионная модель может быть расширена путем включения в нее более одной независимой переменной. При совместном влиянии на *Y* нескольких факторов (*X1, X2, …, Xn*), уравнениемножественной регрессии принимает вид:

*Yt+1=b0+b1X1+b2X2+…+bnXn*,

где*n* – число факторов.

Коэффициенты регрессии представляют собой независимые вклады каждой независимой переменной в предсказание зависимой переменной. Если коэффициент *b* положителен, то связь переменной с зависимой переменной положительна, если коэффициент отрицателен, то и связь носит отрицательный характер (чем меньше значение фактора, тем больше значение переменной *Y*). Если *b=0*, то связь между переменными отсутствует. Для проверки гипотезы о нулевых значениях коэффициентов регрессии (т.е. об отсутствии связи между *Y* и совокупностью факторов) анализируются значения *F*-статистики Фишера. *F*-критерий определяется отношением дисперсии оценки модели к дисперсии остатка и равен:

F= 𝑆𝑆𝑅/𝑞 ,

𝑆𝑆𝐸/(𝑛−(𝑞+1))

Где *SSR* – сумма квадратов, объясненная уравнением регрессии (Sumof Squaresabout Regression);

*SSE* – сумма квадратов остатков (Sum of Squares Errors);

*n* – число наблюдений;

*q* – число коэффициентов регрессии.

Гипотеза об отсутствии линейной зависимости между переменной *Y* и факторами *X* отклоняется при больших значениях *F*-критерия и значении *p-level* меньше *0,05* (вероятность ошибочной оценки относительно принятой гипотезы не превышает *5%* уровня). Наиболее простым методом определения коэффициентов регрессии является метод наименьших квадратов (МНК). С помощью этого метода параметры регрессионной модели вычисляются таким образом, чтобы сумма квадратов ошибок (расстояний от линии регрессии

до фактических значений данных) была бы минимальной.

Функция ошибки при этом равна:

*f = (b0 + b1x11 + b2x21 + … + bkxk1 – y1)2 + (b0 + b1x12 + b2x22 + … + bkxk2 – y2)2 + +*

*+ (b0 + b1x1n + b2x2n + … + bkxkn – yn)2*

Минимизируя функцию *f* положим:

*f*

*b*0

 *f*

*b*1

 ... 

*f*

*bn*

 0 .

Для определения коэффициентов модели множественной линейной регрессии, используя систему уравнений, получим систему нормальных линейных уравнений, которая в векторно–матричной форме имеет вид:



Отклонение отдельной точки от линии регрессии (предсказанного значения) называется остатком. Чем меньше разброс значений остатков около линии регрессии по отношению к общему разбросу значений, тем лучше прогноз. Оценка качества линейной регрессии проводится с помощью коэффициента детерминации *R2*, который показывает какая доля дисперсии отклика объясняется влиянием независимых переменных в построенной модели.

*R2=SSR / SST*, где

*SST* – полная сумма квадратов (Total Sumof Squares).

Если связь между переменными *X* и *Y* отсутствует, то отношение остаточной изменчивости переменной *Y* к исходной дисперсии равно *1*. Если *X* и *Y* коррелируют между собой, то остаточная изменчивость отсутствует и отношение дисперсий будет равно *0*. Например, если имеется *R2=0,4*, то изменчивость значений переменной *Y* около линии регрессии составляет *1-0,4* от исходной дисперсии, т.е. *40%* от исходной изменчивости могут быть объяснены, а *60%* остаточной изменчивости остаются необъясненными. Значение *R2* является индикатором степени подгонки модели к данным (значение *R2* близкое к *1* показывает, что модель объясняет почти всю изменчивость соответствующих переменных). При поиске лучшей регрессионной модели руководствуются требованием *R2≥0,8*.

### Функциональные возможности программы Statistica

Для решения задачи краткосрочного прогнозирования объема продаж предполагается использовать программу Statistica. Пакет прикладных программ, разработанный компанией StatSoft, позволяет проводить исчерпывающий, всесторонний анализ данных, представлять результаты анализа в виде таблиц и графиков, автоматически создавать отчеты о проделанной работе. Предоставляет мощные и удобные в использовании инструменты для статистического и графического анализа, реализует функции управления данными, добычи и визуализации данных, datamining и др.

Программа Statistica имеет модульную структуру, т.е. состоит из модулей, каждый из которых используется для решения конкретного класса задач, а именно: анализ временных рядов и прогнозирование, множественная регрессия, нелинейное оценивание, факторный анализ, кластерный анализ, канонический анализ, непараметрическая статистика, дисперсионный и дискриминантный анализ. Несколько модулей объединены в группу промышленная статистика: контроль качества, анализ процессов, планирование эксперимента.

Оценка коэффициентов однофакторной и многофакторной линейной регрессии осуществляется в отдельном окне системы Statistica, где представлены коэффициенты, оцененные методом наименьших квадратов, коэффициент детерминации, статистика Фишера оценки значимости регрессии, статистики Стьюдента, оценки значимости коэффициентов, коэффициент корреляции (матрица корреляций), статистика Дарбина-Уотсона.Можно анализировать большие модели, содержащие до 500 переменных.

### Пример (вариант **8**)

**Описание деловой ситуации**

Пусть предприятие работает на рынке определенного продукта. При формировании маркетингового решения возникла необходимость в прогнозировании уровня платежеспособного спроса (объема продаж) на выпускаемую продукцию.

Для решения задачи краткосрочного прогнозирования спроса предполагается использовать модели и методы регрессионного анализа.

Маркетологи фирмы располагают статистическими данными за период, равный 30 месяцам, о фактических значениях объемов продаж продукции по месяцам, расходах на рекламу по месяцам, ценах на продукцию фирмы и на продукцию конкурирующей фирмы (см. табл. 1). Такие факторы, как расходы на рекламу, цена на продукцию фирмы и ее основного конкурента были выбраны как наиболее значимые по степени влияния на выходной показатель – объем продаж продукции.

Требуется: 1. Рассчитать ожидаемый предприятием объем продаж продукции на 26-й месяц работы при условии, что предполагаемые на этот месяц значения факторов, влияющих на объем продаж, составят: расходы на рекламу – 87 тыс. руб., цена единицы продукции – 331 руб., цена единицы продукции конкурирующей фирмы – 337 руб.

1. Построить последовательно однофакторную линейную регрессионную модель (с учетом только одного фактора – расходов на рекламу), двухфакторную линейную регрессионную модель (с учетом таких факторов, как расходы на рекламу и цена единицы продукции предприятия) и трехфакторную линейную регрессионную модель (с учетом трех факторов - расходы на рекламу, цена единицы продукции предприятия и цена на единицу продукции фирмы-конкурента).
2. Рассчитать оценки прогнозов объема продаж в 26 месяце, используя построенные регрессионные модели.
3. Провести сравнительный анализ прогнозных оценок объемов продаж продукции, полученных с помощью однофакторной, двухфакторной и трехфакторной регрессионных моделей.

При поиске лучшей регрессионной модели следует руководствоваться следующими наиболее общими требованиями:

1. 47Регрессионная модель должна объяснять не менее 80% вариации зависимой переменной, т.е. R2 0.8.
2. Стандартная ошибка оценки зависимой переменной по уравнению должна составлять не более 5% среднего значения зависимой переменной;
3. Коэффициенты уравнения регрессии и его свободный член должны быть значимы на 5%-ом уровне.
4. Остатки от регрессии должны быть нормально распределены и без систематической составляющей.

Таблица 1. Исходные данные для прогнозирования спроса

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер месяца (t) | Фактический объем продаж за месяц (тыс.руб.)  (Y) | Расходы на рекламу за месяц  (тыс. руб.)  (X1) | Цена продукта (руб./ед. продукта) (X2) | Цена продукта фирмы-конкурента (руб./ед. продукта)  (X3) |
| 1 | 605 | 58 | 287 | 290 |
| 2 | 610 | 52 | 289 | 291 |
| 3 | 615 | 50 | 290 | 292 |
| 4 | 620 | 51 | 279 | 295 |
| 5 | 625 | 53 | 278 | 293 |
| 6 | 630 | 49 | 283 | 296 |
| 7 | 635 | 53 | 291 | 295 |
| 8 | 640 | 59 | 293 | 295 |
| 9 | 645 | 60 | 293 | 297 |
| 10 | 650 | 61 | 291 | 293 |
| 11 | 655 | 57 | 289 | 292 |
| 12 | 660 | 55 | 294 | 295 |
| 13 | 665 | 62 | 299 | 302 |
| 14 | 670 | 68 | 301 | 304 |
| 15 | 675 | 67 | 301 | 304 |
| 16 | 680 | 75 | 299 | 305 |
| 17 | 685 | 63 | 315 | 318 |
| 18 | 690 | 64 | 318 | 322 |
| 19 | 695 | 69 | 313 | 320 |
| 20 | 700 | 70 | 302 | 308 |
| 21 | 705 | 75 | 321 | 327 |
| 22 | 710 | 74 | 334 | 339 |
| 23 | 715 | 79 | 328 | 330 |
| 24 | 720 | 80 | 320 | 325 |
| 25 | 725 | 83 | 329 | 334 |

### Анализ результатов решения задачи

В программе Statistica задаются исходные данные по 25 месяцам из таблицы 1.

* 1. Проводится построение однофакторной линейной регрессионной модели (с учетом одного фактора – расходов на рекламу) вида *Y=b0+b1X1*. Коэффициенты регрессии *b0, b1*, рассчитанные для однофакторной модели представлены на рисунке 1.

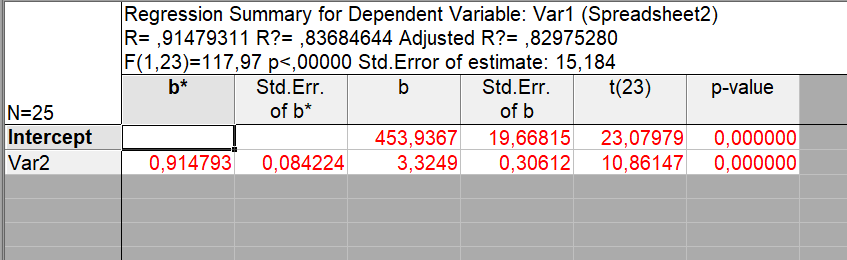


Рисунок 1. Итоги регрессионного анализа однофакторной модели

Проведен расчет значений коэффициентов регрессии для модели с учетом влияния фактора «Расходы на рекламу». Выражение однофакторной линейной регрессии можно представить в виде: *Y26=453,9+3,32\*X1*.

Расчет прогнозного значения объема продаж на 26-й месяц в ручном режиме:

*Y26=453,9+3,32\*87=742,74*

Прогнозное значение объема продаж на 26-й месяц с учетом влияния расходов на рекламу, полученное в автоматическом режиме *Y26=743,2*тыс.руб. (рис. 2).

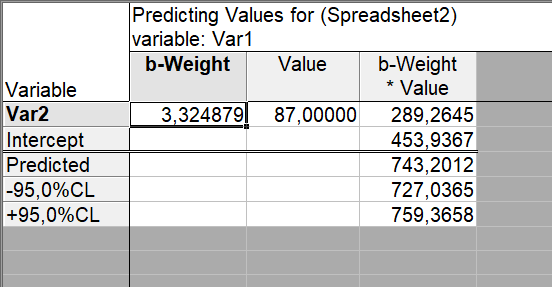


Рисунок 2. Расчет прогнозного значения объема продаж с использованием однофакторной линейной регрессионной модели

* 1. Строится двухфакторная линейная регрессионная модель (с учетом двух факторов – расходов на рекламу и цены единицы продукции) вида *Y=b0+b1X1+b2X2*. Значения коэффициентов регрессии *b0, b1, b2* представлены на рисунке 3.

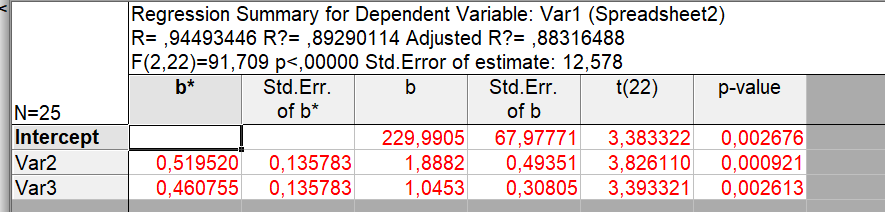


Рисунок 3. Итоги регрессионного анализа двухфакторной модели

Проведен расчет значений коэффициентов регрессии для модели с учетом влияния факторов «Расходы на рекламу» и «Цена единицы продукции». Выражение двухфакторной линейной регрессии можно представить в виде: *Y26=229,99+1,88\*X1+1,04\*X2.*

Расчет прогнозного значения объема продаж на 26-й месяц в ручном режиме:

*Y26=229,99+1,88\*87+1,04\*331=737,79*

Прогнозное значение объема продаж на 26-й месяц с учетом влияния расходов на рекламу и цены единицы продукции предприятия, полученное в автоматическом режиме составляет *Y26=740,2692* тыс.руб.

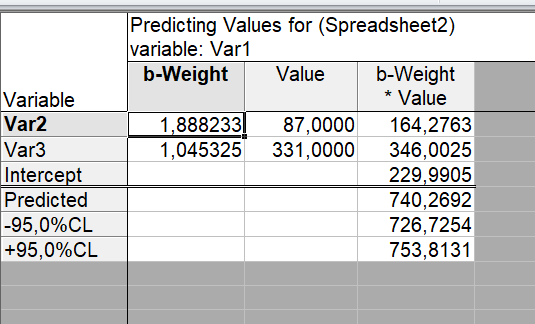


Рисунок 4. Расчет прогнозного значения объема продаж при использовании двухфакторной регрессионной модели

* 1. Значения коэффициентов регрессии *b0, b1, b2, b3,* рассчитанные для трехфакторной линейной регрессионной модели (с учетом трех факторов – расходов на рекламу, цены единицы продукции предприятия и цены единицы продукции фирмы- конкурента) вида *Y=b0+b1X1+b2X2+b3X3*, представлены на рисунке 5.

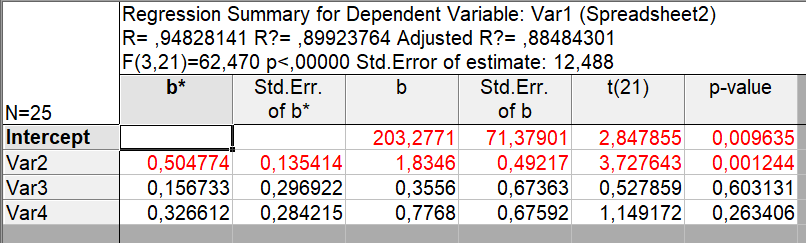


Рисунок 5. Итоги регрессионного анализа трехфакторной модели

Проведен расчет значений коэффициентов регрессии для модели с учетом влияния факторов «Расходы на рекламу», «Цена единицы продукции предприятия» и «Цена единицы продукции фирмы-конкурента». Выражение трехфакторной линейной регрессии можно представить в виде:*Y26=203,28+1,83\*X1+0,35\*X2+0,77\*Х3*

Расчет прогнозного значения объема продаж на 26-й месяц в ручном режиме:

*Y26=203,28+1,83\*87+0,35\*331+0,77\*336=737,06*

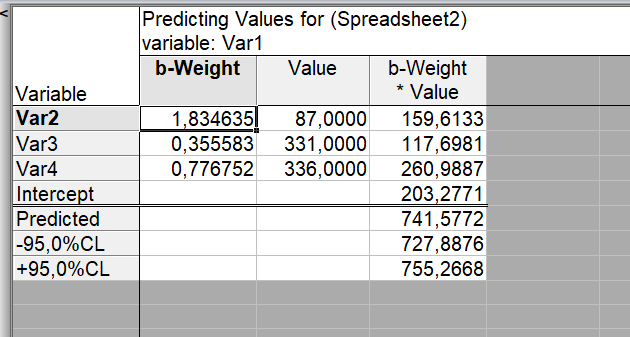


Рисунок 6. Расчет прогнозного значения объема продаж с использованием трехфакторной регрессионной модели

Прогнозное значение объема продаж на 26-й месяц с учетом влияния расходов на рекламу, цены единицы продукции предприятия и цены единицы продукции фирмы- конкурента, полученное в автоматическом режиме *Y26=741,5772* тыс.руб.

### Заключение

В ходе лабораторной работы была решена задача прогнозирования объема продаж с применением метода регрессионного анализа. Осуществлено последовательное построение моделей регрессионного анализа с учетом влияния одного, двух и трех рассматриваемых факторов на значение объема продаж для 26 месяца работы предприятия. Анализ построенных моделей показал, что все три модели корректны, имеют значения *R2>0,85*, т.е. объясняют больше *85%* разброса значений переменной *Y* относительно среднего:

* для однофакторной модели *R2=0,836*;
* для двухфакторной модели *R2=0,892*;
* для трехфакторной модели*R2=0,899*.

Во всех построенных регрессионных моделях стандартная ошибка оценки зависимой переменной составляет порядка 2,5%, что является допустимой нормой. Построенные регрессии значимы, а гипотеза об отсутствии связи между переменными может быть отклонена, т.к. большим значениям *F*-критериев соответствуют уровни значимости (*p-level*) меньше *5%*. Результаты получены на последнем шаге регрессии и проведен анализ остатков для каждой регрессионной модели (табл. 2). По графикам можно сделать вывод, что остатки нормально распределены (в пределах ±18 ед.), заметных выбросов нет.

Фактор *X2* (цена единицы продукции предприятия) в двухфакторной и трехфакторной моделях имеет низкий уровень значимости (*p-level* больше *0,05*), т.е. этот фактор в меньшей степени влияет на изменение уровня объема продаж, чем остальные факторы.

Таблица 2. Распределение остатков, полученных по результатам

построенных регрессионных моделей

|  |  |  |
| --- | --- | --- |
| Распределение остатков однофакторной регрессионной  модели | Распределение остатков двухфакторной регрессионной  модели | Распределение остатков трехфакторной регрессионной  модели |
|  |  |  |

Таким образом, каждая из построенных регрессионных моделей может быть использована для решения задачи прогнозирования объема продаж продукции предприятия. Окончательным решением задачи прогнозирования будем считать прогнозное значение *Y26=733,26* тыс.руб., полученное при использовании трехфакторной регрессионной модели, т.к. этой модели соответствует наиболее высокий уровень коэффициента детерминации *R2* и наименьший разброс остатков (в пределах ±15 ед.).