Корреляционный и регрессионный анализ данных. Множественный регрессионный анализ

СОДЕРЖАНИЕ

1. ЦЕЛЬ И ЗАДАЧИ МЕТОДИЧЕСКИХ РЕКОМЕНДАЦИЙ

Закрепить теоретические знания и приобрести практические навыки в построении регрессионных моделей объекта по экспериментальным данным с использованием программы Statistica.

По результатам наблюдений за функционированием объектов получены экспериментальные данные. Построить регрессионные модели объектов по заданным экспериментальным данным.

Решение общей задачи разбивается на несколько этапов:

а) предварительная обработка данных с целью стандартизации результатов наблюдения;

б) оценка параметров регрессионных моделей;

в) проверка значимости коэффициентов регрессии;

г) оценка точности регрессионных моделей;

д) выводы о возможности применения составленных регрессионных моделей.

Основные задачи корреляционно-регрессионного анализа

1. Измерение тесноты связи между результативным и факторным признаком (признаками). В зависимости от количества влияющих на результат факторов задача решается путем вычисления корреляционного отношения, коэффициентов парной, частной, множественной корреляции или детерминации.

2. Оценка параметров уравнения регрессии, выражающего зависимость средних значений результативного признака от значений факторного признака (признаков). Задача решается путем вычисления коэффициентов регрессии.

3. Определение важнейших факторов, влияющих на результативный признак. Задача решается путем оценки тесноты связи факторов с результатом.

4. Прогнозирование возможных значений результативного признака при задаваемых значениях факторных признаков. Задача решается путем подстановки ожидаемых значений факторов в регрессионное уравнение и вычисления прогнозируемых значений результата.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

2.1. Корреляционный анализ

Исторически первой и простейшей характеристикой тесноты связи является линейный коэффициент парной корреляции. Он используется при изучении парной корреляционной зависимости, т.е. когда оценивается связь между парой переменных.

Показатели корреляции основаны на оценке сопряженной вариации изучаемых переменных. Парный коэффициент корреляции *(r)* – это нормированный коэффициент ковариации. Ковариация, являясь мерой взаимосвязи двух переменных, рассчитывается как средняя величина произведения отклонений индивидуальных значений анализируемых данных от их средних значений:

$$Cov(y,x) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) \tag{1}$$

Недостаток этого показателя (числовое значение ковариации зависит от размерности переменных x и y) преодолевается в парном коэффициенте корреляции путем нормирования абсолютных отклонений.

$$\frac{(y_i - \bar{y})}{\sigma_y} \quad \text{и} \quad \frac{(x_i - \bar{x})}{\sigma_x},$$

где σ_y – среднее квадратическое отклонение переменной y;

σ_x – среднее квадратическое отклонение переменной x.

Парный коэффициент корреляции:

$$r = \frac{\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right)}{n}, \tag{2}$$

где n – число единиц в статистической совокупности.

Коэффициент корреляции изменяется в пределах

$$0 \le |r| \le 1 . \tag{3}$$

Если $r = 0$, линейная связь между изучаемыми переменными отсутствует. Если $|r| = 1$, связь функциональная, т.е. значение зависимой переменной полностью определяется независимой переменной. Положительное значение коэффициента свидетельствует о прямой зависимости между переменными, отрицательная – об обратной.

Парный коэффициент корреляции – это симметричная характеристика, т.е. $r_{yx} = r_{xy}$. Значение r отражает только степень тесноты корреляционной связи между изучаемыми переменными, но не свидетельствует о причинно-следственной зависимости между ними.

Квадрат коэффициента корреляции (r^2) называется коэффициентом детерминации. Его значение изменяется в пределах от 0 до 1, и означает долю вариации одной переменной, обусловленную вариацией другой.

2.2. Парный регрессионный анализ

Парной регрессией называется уравнение связи двух переменных x и y вида

$$y = f(x), \tag{4}$$

где y – зависимая переменная (результативный признак); x – независимая, объясняющая переменная (признак-фактор).

Парная регрессия применяется, если имеется доминирующий фактор, который и используется в качестве объясняющей переменной.

Различают линейные и нелинейные относительно фактора x регрессии.

Регрессионный анализ предполагает теоретический анализ природы изучаемого явления с целью определения круга факторов, оказывающих влияние на поведение результативного признака. На базе корреляционного анализа выявляется наличие статистически значимых связей в конкретных условиях места и времени. Затем строится уравнение регрессии (аналитическая форма изучаемой зависимости), которое при определенных условиях может быть признано статистической моделью связи между признаками.

Уравнение регрессии – это математическая функция, описывающая зависимость условного среднего значения результативной (зависимой) переменной от заданных значений факторных (независимых) переменных. Таким образом, уравнение регрессии отражает основную тенденцию связи, характерную для изучаемой статистической совокупности в целом.

В регрессионном анализе можно выделить три составляющие:

- определение типа функции (структуры модели) для описания изучаемой зависимости;
- расчет неизвестных параметров уравнения регрессии;
- оценку качества модели.

До широкого распространения компьютерных технологий перечисленные элементы являлись последовательными этапами анализа. В современных условиях все процедуры выполняются комплексно. Представленное ниже раздельное их описание необходимо для понимания сути каждой процедуры.

Первый этап регрессионного анализа – поиск линии регрессии, которая бы лучшим образом аппроксимировала поле корреляции. При этом необходимо учитывать природу изучаемых показателей, специфику их взаимосвязи, свойства математических функций. Однако в настоящее время процедура выбора лучшего уравнения связи формализована. Современные ППП позволяют одновременно построить несколько видов уравнений, а затем, пользуясь специальными критериями, отобрать лучшую модель. В качестве критерия могут быть использованы: максимальное значение коэффициента детерминации, максимальное значение F-критерия Фишера, минимальное значение остаточной дисперсии, минимальное значение стандартной ошибки уравнения, минимальное значение средней ошибки аппроксимации.

Для аналитического описания связи между признаками могут быть использованы следующие виды уравнений:

1) $\bar{y} = a_0 + a_1 x$ – прямая, линейная функция;

2) $\bar{y} = a_0 + a_1 x^2 + a_2 x$ – парабола;

3) $\bar{y} = a_0 + a_1 \dfrac{1}{x}$ – гипербола;

4) $\bar{y} = a_0 x^{a_1}$ – степенная функция;

5) $\bar{y} = \exp(a_0 + a_1 x)$ – экспонента и др.

Некоторые задачи корреляционно-регрессионного анализа, а также возможности ППП, делают необходимым выполнение операции линеаризации уравнений, т.е. приведение их к линейному виду путем логарифмирования. Производится замена признака-фактора и признака-результата их натуральными логарифмами. При проведении анализа с использованием линеаризации необходимо помнить о том, что все показатели и графические изображения рассчитываются и строятся для логарифмов признаков. Рассмотрим экспоненциальную и степенную функции после линеаризации:

$$\ln \bar{y} = a_0 + a_1 \times x; \qquad \ln \bar{y} = a_0 + a_1 \times \ln x .$$

Простейшим видом уравнения регрессии является парная линейная регрессия

$$\bar{y} = a_0 + a_1 x + \varepsilon ,$$

где $\bar{y}$ – расчетное, теоретическое значение признака-результата; a_0, a_1 – параметры уравнения регрессии; ε – случайная величина.

Присутствие в уравнении ε связано с рядом причин, среди которых: наличие признаков-факторов, не включенных в данное уравнение; неправильное описание структуры модели; ошибки измерений и др.

Чтобы воспользоваться уравнением регрессии, необходимо рассчитать значения его параметров. Чаще всего расчет осуществляется по методу наименьших квадратов (МНК). Суть метода в том, что удается получить такие значения параметров, при которых минимизируется сумма квадратов отклонений эмпирических значений признака-результата от его теоретических значений, т.е.

$$S = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 = \sum[y_i - (a_0 + a_1 x)]^2 \to \min . \quad (5)$$

Третий этап регрессионного анализа предполагает оценку качества полученного уравнения связи. Поскольку уравнение регрессии строится, как правило, на основе выборочных данных, то следует оценить статистическую значимость параметров уравнения и уравнения в целом.

Оценка статистической значимости параметров модели означает проверку нулевых гипотез о равенстве параметров генеральной совокупности нулю, т.е. в условиях парной регрессии:

H_0: $A_0 = 0$, H_0: $A_1 = 0$.

Проверка производится с использованием t-статистики, которая в этом случае представляет собой отношение значения параметра к его стандартной (среднеквадратической) ошибке S:

$$t_{a_0} = \frac{a_0 - A_0}{S_{a_0}} \quad \text{и} \quad t_{a_1} = \frac{a_1 - A_1}{S_{a_1}} , \quad (6)$$

поскольку $A_0 = 0$ и $A_1 = 0$, то

$$t_{a_0} = \frac{a_0}{S_{a_0}} ; \; t_{a_1} = \frac{a_1}{S_{a_1}} , \quad (7)$$

где S_{a_0} - стандартная ошибка параметра a_0: $S_{a_0} = \frac{\sigma_y \sqrt{1-r^2}}{\sqrt{n-2}}$;

S_{a_1} - стандартная ошибка параметра a_1: $S_{a_1} = \frac{\sigma_y \sqrt{1-r^2}}{\sigma_x \sqrt{n-2}}$.

Фактические значения t-критерия сравниваются с табличными (с учетом уровня значимости α и числа степеней свободы ($d.f.=n-k-1$)). Параметры признаются статистически значимыми, т.е. сформированными под воздействием неслучайных факторов, если $t_{\text{факт}} > t_{\text{табл}}$.

Значимость уравнения в целом оценивается на основе F-критерия Фишера.

F-критерий – это отношение объясненной вариации (факторной дисперсии) результативного признака, рассчитанной на одну степень свободы, $\left(\frac{\sum_{i-1}^{n}(\bar{y}_i - \bar{y})^2}{k} \right)$, к необъясненной вариации (остаточной дисперсии) признака-результата, рассчитанной на одну степень свободы, $\left(\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n-k-1} \right)$, таким образом

$$F = \frac{\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2}{k} : \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n-k-1} , \quad (8)$$

где k – число степеней свободы факторной дисперсии, равное числу независимых переменных (признаков-факторов) в уравнении регрессии;

n-*k*-1 - число степеней свободы остаточной дисперсии.

Если обе части соотношения разделить на общую дисперсию зависимой переменной (результата), то *F*-критерий может быть представлен следующим образом:

$$F = \frac{r_{yx}^2}{1-r_{yx}^2}\frac{n-k-1}{k}. \tag{9}$$

Расчетное значение критерия сопоставляется с табличным (с учетом числа степеней свободы: *d.f.* = *k* и *d.f.*=*n*-*k*-1) (приложение 3). Если $F_{\text{расч.}} \geq F_{\text{табл.}}$, то делается вывод о статистической значимости уравнения в целом. Поскольку *F*-критерий основан на соотношении факторной и остаточной дисперсий результативного признака, то вполне логично его использование для оценки качества модели. Если объясненная дисперсия существенно больше необъясненной, это означает, что в уравнение связи включены именно те факторы, которые играют определяющую роль в изменении значения признака-результата. Статистическая значимость уравнения одновременно означает статистическую значимость коэффициента детерминации.

Результаты оценки регрессионного уравнения могут быть разными. Возможен вариант, когда уравнение в целом статистически значимо, а некоторые параметры уравнения незначимы. Это означает, что описанная зависимость результата от аргументов может служить основой для принятия некоторых управленческих решений, но полученное уравнение регрессии нельзя использовать для прогнозирования. Уравнение связи признается моделью и может быть использовано в целях прогнозирования, если статистически значимы и параметры, и уравнение в целом.

2.3 Множественный корреляционный анализ

Изучение множественной корреляционной зависимости предполагает оценку влияния на результативный признак двух и более факторов. При этом рассчитываются множественные (совокупные) и частные коэффициенты корреляции, которые можно определить на основе парных коэффициентов корреляции r.

Так, при двухфакторной модели связи рекуррентная формула множественного коэффициента корреляции выглядит следующим образом:

$$R_{y_{x_1x_2}} = \sqrt{\frac{r_{y_{x_1}}^2 + r_{y_{x_2}}^2 - 2r_{y_{x_1}}r_{y_{x_2}}r_{x_1x_2}}{1-r_{x_1x_2}^2}}. \tag{10}$$

Значения *R* изменяются в пределах от 0 до 1. Величина совокупного коэффициента корреляции всегда больше любого из парных коэффициентов и включение в анализ новых факторов не может привести к уменьшению значения *R*.

Квадрат множественного коэффициента корреляции R^2 является множественным коэффициентом детерминации и характеризует долю дисперсии результативного признака, объясненную вариацией всех факторов, включенных в анализ, в общей дисперсии результата.

При изучении множественной корреляционной зависимости наряду с оценкой совокупного влияния всего набора интересующих исследователя факторов возникает необходимость получить количественную характеристику влияния каждой объясняющей переменной, «очищенную» от опосредованного воздействия других факторов. Эта задача решается с помощью так называемых частных (парциальных) коэффициентов корреляции. При их построении применяется прием элиминирования влияния всех факторов кроме фактора, оцениваемого в данный момент. Элиминирование осуществляется путем закрепления значений признаков-факторов на неизменном (среднем) уровне. Таким образом, частные коэффициенты корреляции позволяют измерить «очищенное» влияние конкретного фактора.

Частные коэффициенты также могут быть рассчитаны по рекуррентной формуле. Если элиминируется влияние одного фактора, то частный коэффициент корреляции называется коэффициентом первого порядка и в условиях двухфакторной модели рассчитывается следующим образом:

$$R_{y_{x_1 . x_2}} = \frac{r_{y_{x_1}} - r_{y_{x_2}} r_{x_1 x_2}}{\sqrt{(1 - r^2_{y_{x_2}})(1 - r^2_{x_1 x_2})}}, \tag{11}$$

где $R_{y_{x_1 . x_2}}$ – частный коэффициент корреляции первого порядка; точка между x_1 и x_2 означает, что элиминируется влияние x_2.

Если элиминируется влияние фактора x_1, то частный коэффициент корреляции рассчитывается следующим образом:

$$R_{y_{x_2 . x_1}} = \frac{r_{y_{x_2}} - r_{y_{x_1}} r_{x_1 x_2}}{\sqrt{(1 - r^2_{y_{x_1}})(1 - r^2_{x_1 x_2})}}. \tag{12}$$

2.4. Множественный регрессионный анализ

Модель множественной регрессии в общем виде записывается следующим образом:

$$\bar{y} = a + a_1 x_1 + a_2 x_2 + a_3 x_3 + ... + a_k x_k + \varepsilon \tag{13}$$

Специфической проблемой, решаемой при построении множественной регрессии, является отбор факторов, включаемых в уравнение регрессии. Напомним, что для получения надежных оценок параметров необходимо, чтобы число факторов, включаемых в модель, было по меньшей мере в 5 – 6 раз меньше объема изучаемой совокупности. Из-за ограниченности объема совокупности не стоит «засорять» модель факторами, связь которых с зависимой переменной слабая (r < 0,3).

Для получения оценки «очищенного» влияния каждого фактора в уравнение не следует одновременно включать факторы, между которыми существует тесная линейная зависимость (коллинеарность). Отбор факторов может быть осуществлен на основе матрицы парных коэффициентов корреляции (таблица 1). Поскольку парный коэффициент корреляции – мера симметричная, матрица симметрична относительно единичной диагонали, поэтому достаточно заполнить либо верхний, либо нижний сегмент корреляционной матрицы.

Анализ первой строки матрицы, содержащей показатели тесноты связи между признаком-результатом и каждым из признаков-факторов, позволяет исключить из анализа факторы, практически не влияющие на поведение зависимой переменной (r < 0,3).

В остальных клетках выделенного сегмента матрицы содержатся коэффициенты, оценивающие зависимость между факторами. Анализ этих характеристик позволяет выявить наличие мультиколлинеарности. Из двух коллинеарных факторов (r ≥ 0,7) один следует исключить из анализа. Предпочтение отдается признаку, связь которого с результатом более тесная.

Таблица 1

Матрица парных коэффициентов корреляции

Признак	y	x_1	x_2	x_3	…	x_K
y	1	r_{yx_1}	r_{yx_2}	r_{yx_3}	…	r_{yx_k}
x_1	⋮	1	$r_{x_1x_2}$	$r_{x_1x_3}$	…	$r_{x_1x_k}$
x_2	⋮	⋮	1	$r_{x_2x_3}$	…	$r_{x_2x_k}$

x_3	$\vdots$	$\vdots$	$\vdots$	1	…	$r_{x_3x_k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	1	$\vdots$
$x_к$	…	…	…	…	…	1

Расчет параметров уравнения множественной регрессии осуществляется на основе МНК.

Параметры при факторах в уравнении множественной регрессии называются условно-чистыми коэффициентами регрессии. Их можно было бы назвать «чистыми», если бы удалось включить в модель все факторы, определяющие значение признака-результата, что на практике не может быть реализовано.

Условно-чистые коэффициенты регрессии оценивают силу влияния каждой независимой переменной при условии элиминирования других факторов, включенных в модель. Интерпретация значений коэффициентов аналогична интерпретации коэффициентов в уравнении парной регрессии.

Независимые переменные могут иметь разные единицы измерения, поэтому получаемые коэффициенты регрессии не сопоставимы и не позволяют ранжировать аргументы по силе их влияния на зависимую переменную.

3. РЕШЕНИЕ ЗАДАЧ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА С ИСПОЛЬЗОВАНИЕМ ППП STATISTICA

3.1. Общие сведения об интегрированном статистическом пакете общего назначения STATISTICA

В настоящем разделе дано краткое описание системы STATISTICA, более подробные сведения о пакете приведены в [3, 4], а также в поставляемой вместе с системой документацией фирмы-разработчика StatSoft и кратком руководстве. Следует отметить, что в процессе работы в среде STATISTICA студент может воспользоваться экранным справочником, содержащим практически всю информацию печатной документации.

STATISTICA полностью удовлетворяет основным стандартам среды Windows:

- стандартам пользовательского интерфейса;

- технологии DDE — динамического обмена данными из других приложений. Благодаря поддержке DDE нетрудно выполнить командные сценарии изнутри других приложений. Например, можно в Excel написать минипрограмму (макрос), которая запускает пакет STATISTICA. После добавления в макрос специальных SQL-команд можно импортировать в пакет данные;

- технологии OLE — связывания и внедрения объектов, поддержка основных операций с буфером обмена и др. Использование OLE технологии обмена между Windows-приложениями позволяет легко интегрировать результаты, например, между WinWord и STATISTICA.

Статистический анализ данных в системе STATISTICA можно представить в виде следующих основных этапов:

- ввод данных в электронную таблицу с исходными данными и их предварительное преобразование перед анализом (структурирование, построение необходимых выборок, ранжирование и т. д.);

- визуализация данных при помощи того или иного типа графиков;

- применение конкретной процедуры статистической обработки;

- вывод результатов анализа в виде графиков и электронных таблиц с численной и текстовой информацией;

- подготовка и печать отчета;
- автоматизация процессов обработки при помощи макрокоманд, языка SCL или STATISTICA BASIC.

Интегрированный статистический пакет общего назначения

STATISTICA состоит из следующих основных компонент:

- многофункциональной системы для работы с данными, которая включает в себя электронные таблицы для ввода и задания исходных данных, а также специальные таблицы (*Scroolsheet* ™) для вывода численных результатов анализа. Для сложной обработки данных в STATISTICA имеется модуль *Управления данными*;
- графической системы для визуализации данных и результатов статистического анализа;
- набора статистических модулей, в которых собраны группы логически связанных между собой статистических процедур:
- основные статистики и таблицы;
- непараметрическая статистика;
- дисперсионный анализ;
- множественная регрессия;
- нелинейное оценивание;
- анализ временных рядов и прогнозирование;
- кластерный анализ;
- управление данными;
- факторный анализ и др.

После запуска системы STATISTICA на экране появляется **Переключатель модулей**. Модули взаимодействуют друг с другом, имея одинаковый формат системных файлов. Если пользователю нужен, например, раздел линейной регрессии, то следует выбрать модуль **Multiple Regression - Множественной регрессии**. В любом конкретном модуле можно выполнить определенный способ статистической обработки, не обращаясь к процедурам из других модулей. Все основные операции при работе с данными и графические возможности доступны в любом статистическом модуле и на любом шаге анализа;

- специального инструментария для подготовки отчетов. При помощи текстового редактора, встроенного в систему, можно готовить полноценные отчеты. В пакете STATISTICA также имеется возможность автоматического создания отчетов;
- встроенных языков SCL и STATISTICA BASIC, которые позволяют автоматизировать рутинные процессы обработки данных в системе.

3.2. Порядок выполнения индивидуального задания

1. Ввод исходных данных.

Статистическую обработку данных следует предварить открытием уже существующего файла с данными через команду **Open Data** или ввести данные в компьютер через команду **File/ New Data** .

2. Получить дискриптивные статистики по каждому признаку. Оценить показатели вариации каждого признака и сделать вывод о возможностях применения метода наименьших квадратов для их изучения.

Для получения дискриптивных статистик необходимо в **Переключателе модулей**, появившемся после запуска пакета STATISTICA, выбрать команду **Basic Statistics/Tables**, при этом на экране появится стартовая панель модуля **Основные статистики и таблицы**, в которой следует выбрать команду **Descriptive statistics.**

3. Составить уравнение множественной регрессии, оценить его параметры, пояснить их смысл.

Для построения уравнения множественной регрессии необходимо в **Переключателе модулей** выбрать команду **Multiple Regression**.

4. Проанализировать линейные коэффициенты парной и частной корреляции.

Для построения матрицы парных коэффициентов корреляции и матрицы линейных коэффициентов частной корреляции в модуле **Multiple Regression** следует последовательно выбрать команды **Correlations** and **desc.stats** (для построения матрицы парных коэффициентов корреляции), **Partial correlations** (для построения матрицы линейных коэффициентов частной корреляции).

5. Оценить значения скорректированного и нескорректированного линейных коэффициентов множественной корреляции.

6. С помощью F-критерия Фишера оценить статистическую надежность уравнения регрессии в целом.

7. Оформление отчета. В отчете следует отразить основные этапы выполненного задания, полученные результаты и сделать выводы по каждому этапу. Для этой цели можно использовать распечатки отчета, полученного средствами пакета STATISTICA (файл с расширением *.rtf), включая его широкие графические возможности.

4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1 Что изучает корреляционно-регрессионный анализ?
2 Определение парного, частного, множественного коэффициентов корреляции.
3 Как рассчитывается частный коэффициент корреляции?
4 Как рассчитывается множественный коэффициент корреляции?
5 Как проверяется значимость оценки коэффициента парной корреляции?
6 Как проверяется значимость частного коэффициента корреляции?
7 Как проверяется значимость множественного коэффициента корреляции?
8 Для чего используется коэффициент детерминации?
9 Как строится корреляционная матрица?
10 Как строятся доверительные интервалы для коэффициентов корреляции?
11Как оцениваются коэффициенты регрессии?
12 Как проверяется значимость уравнения регрессии?
13 Как проверяется значимость коэффициентов регрессии?
14 Как строятся интервальные оценки коэффициентов уравнения регрессии?

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Айвазян С. А. Прикладная статистика. Основы эконометрики: Учебник для вузов/ С. А. Айвазян, В. С. Мхитарян. М. : ЮНИТИ-ДАНА, 2001. –656 с.

2. Куприенко Н. В. Статистические методы изучения связей. Корреляционно-регрессионный анализ/ Н. В. Куприенко, О. А. Пономарева, Д. В. Тихонов. СПб. : Изд-во политехн. ун-та, 2008. – 118 с.

3. Боровиков В.П., Боровиков И.П. STATISTICA. Статистический анализ и обработка данных в среде Windows. М.: Филин, 1997.

4. Боровиков В.П. Популярное введение в программу STATISTICA. М., 1998

5. www.statsoft.ru (сайт компании StatSoft Russia – документация по STATISTICA).

6. www.exponenta.ru (примеры решения практических задач в ППП STATISTICA).

7. Статистика: Корреляционно-регрессионный анализ статистических связей на персональном компьютере: Методические указания к практическим занятиям для студентов всех форм обучения специальности «Менеджмент» / Калинингр. ун-т; Сост. Н.Ю. Лукьянова. - Калининград, 1999. - 35 с.

8. Шанченко, Н. И. Эконометрика: лабораторный практикум : учебное пособие / Н. И. Шанченко. – Ульяновск : УлГТУ, 2011. – 117 с.

9. Реннер А.Г. Корреляционно-регрессионый анализ / А.Г. Реннер, Г.Г. Аралбаева, О. А. Зиновьева // Изд-во Оренбургского государственного университета, 2002. – 26 с.