

ИСПОЛЬЗОВАНИЕ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА ДЛЯ ОБРАБОТКИ ЭКОНОМИЧЕСКИХ СТАТИСТИЧЕСКИХ ДАННЫХ

Обработка статистических данных уже давно применяется в самых разнообразных видах человеческой деятельности. Вообще говоря, трудно назвать ту сферу, в которой она бы не использовалась. Но, пожалуй, ни в одной области знаний и практической деятельности обработка статистических данных не играет такой исключительно большой роли, как в экономике, имеющей дело с обработкой и анализом огромных массивов информации о социально-экономических явлениях и процессах. Всесторонний и глубокий анализ этой информации, так называемых статистических данных, предполагает использование различных специальных методов, важное место среди которых занимает корреляционный и регрессионный анализы обработки статистических данных.

В экономических исследованиях часто решают задачу выявления факторов, определяющих уровень и динамику экономического процесса. Такая задача чаще всего решается методами корреляционного и регрессионного анализа. Для достоверного отображения объективно существующих в экономике процессов необходимо выявить существенные взаимосвязи и не только выявить, но и дать им количественную оценку. Этот подход требует вскрытия причинных зависимостей. Под причинной зависимостью понимается такая связь между процессами, когда изменение одного из них является следствием изменения другого.

Основными задачами корреляционного анализа являются оценка силы связи и проверка статистических гипотез о наличии и силе корреляционной связи. Не все факторы, влияющие на экономические процессы, являются случайными величинами, поэтому при анализе экономических явлений обычно рассматриваются связи между случайными и неслучайными величинами. Такие связи называются регрессионными, а метод математической статистики, их изучающий, называется регрессионным анализом.

Использование возможностей современной вычислительной техники, оснащенной пакетами программ машинной обработки статистической информации на ЭВМ, делает практически осуществимым

оперативное решение задач изучения взаимосвязи показателей биржевых ставок методами корреляционно-регрессионного анализа.

При машинной обработке исходной информации на ЭВМ, оснащенных пакетами стандартных программ ведения анализов, вычисление параметров применяемых математических функций является быстро выполняемой счетной операцией.

Данная работа посвящена изучению возможности обработки статистических данных биржевых ставок методами корреляционного и регрессионного анализа с использованием пакета прикладных программ Microsoft Excel.

Роль корреляционно-регрессионного анализа в обработке экономических данных. Корреляционный анализ и регрессионный анализ являются смежными разделами математической статистики, и предназначаются для изучения по выборочным данным статистической зависимости ряда величин; некоторые из которых являются случайными. При статистической зависимости величины не связаны функционально, но как случайные величины заданы совместным распределением вероятностей. Исследование взаимосвязи случайных величин биржевых ставок приводит к теории корреляции, как разделу теории вероятностей и корреляционному анализу, как разделу математической статистики. Исследование зависимости случайных величин приводит к моделям регрессии и регрессионному анализу на базе выборочных данных. Теория вероятностей и математическая статистика представляют лишь инструмент для изучения статистической зависимости, но не ставят своей целью установление причинной связи. Представления и гипотезы о причинной связи должны быть привнесены из некоторой другой теории, которая позволяет содержательно объяснить изучаемое явление.

Формально корреляционная модель взаимосвязи системы случайных величин $X' = (X_1, X_2, \dots, X_k)$ может быть представлена в следующем виде: $x = f(x, Z)$, где Z – набор случайных величин, оказывающих влияние на изучаемые случайные величины.

Экономические данные почти всегда представлены в виде таблиц. Числовые данные, содержащиеся в таблицах, обычно имеют между собой явные (известные) или неявные (скрытые) связи.

Явно связаны показатели, которые получены методами прямого счета, т. е. вычислены по заранее известным формулам. Например, проценты выполнения плана, уровни, удельные веса, отклонения в сумме, отклонения в процентах, темпы роста, темпы прироста, индексы и т. д.

Связи же второго типа (неявные) заранее неизвестны. Однако необходимо уметь объяснять и предсказывать (прогнозировать) сложные явления для того, чтобы управлять ими. Поэтому специалисты с помощью наблюдений стремятся выявить скрытые зависимости и выразить их в виде формул, т. е. математически смоделировать явления или процессы. Одну из таких возможностей предоставляет корреляционно-регрессионный анализ.

Математические модели строятся и используются для трех обобщенных целей:

- для объяснения;
- для предсказания;
- для управления.

Представление экономических и других данных в электронных таблицах в наши дни стало простым и естественным. Оснащение же электронных таблиц средствами корреляционно-регрессионного анализа способствует тому, что из группы сложных, глубоко научных и потому редко используемых, почти экзотических методов, корреляционно-регрессионный анализ превращается для специалиста в повседневный, эффективный и оперативный аналитический инструмент. Однако, в силу его сложности, освоение его требует значительно больших знаний и усилий, чем освоение простых электронных таблиц.

Пользуясь методами корреляционно-регрессионного анализа, аналитики измеряют тесноту связей показателей с помощью коэффициента корреляции. При этом обнаруживаются связи, различные по силе (сильные, слабые, умеренные и др.) и различные по направлению (прямые, обратные). Если связи окажутся существенными, то целесообразно будет найти их математическое выражение в виде регрессионной модели и оценить статистическую значимость модели. В экономике значимое уравнение используется, как правило, для прогнозирования изучаемого явления или показателя.

Регрессионный анализ называют основным методом современной математической статистики для выявления неявных и завуалированных связей между данными наблюдений. Электронные таблицы делают такой анализ легко доступным. Таким образом, регрессионные вычисления и подбор хороших уравнений - это ценный, универсальный исследовательский инструмент в самых разнообразных отраслях деловой и научной деятельности (маркетинг, торговля, медицина и т. д.). Усвоив технологию использования этого инструмента, можно применять его по мере необходимости, получая знание о скрытых связях, улучшая аналитическую поддержку принятия решений и повышая их обоснованность.

Корреляционно-регрессионный анализ считается одним из главных методов в маркетинге, наряду с оптимизационными расчетами, а также математическим и графическим моделированием трендов (тенденций). Широко применяются как однофакторные, так и множественные регрессионные модели.

Корреляционно-регрессионный анализ и его возможности.

Корреляционный анализ является одним из методов статистического анализа взаимосвязи нескольких признаков.

Он определяется как метод, применяемый тогда, когда данные наблюдения можно считать случайными и выбранными из генеральной совокупности, распределенной по многомерному нормальному закону. Основная задача корреляционного анализа (являющаяся основной и в регрессионном анализе) состоит в оценке уравнения регрессии.

Корреляция – это статистическая зависимость между случайными величинами, не имеющими строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

1. Парная корреляция – связь между двумя признаками (результативным и факторным или двумя факторными).

2. Частная корреляция – зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков.

3. Множественная корреляция – зависимость результативного и двух или более факторных признаков, включенных в исследование.

Корреляционный анализ имеет своей задачей количественное определение тесноты связи между двумя признаками (при парной связи) и между результативным признаком и множеством факторных признаков (при многофакторной связи).

Теснота связи количественно выражается величиной коэффициентов корреляции. Коэффициенты корреляции, представляя количественную характеристику тесноты связи между признаками, дают возможность определить «полезность» факторных признаков при построении уравнений множественной регрессии. Величина коэффициентов корреляции служит также оценкой соответствия уравнению регрессии выявленным причинно-следственным связям.

Для применения корреляционного анализа необходимо, чтобы все рассматриваемые переменные были случайными и имели нормальный закон распределения. Причем выполнение этих условий необходимо только при вероятностной оценке выявленной тесноты связи.

Рассмотрим простейший случай выявления тесноты связи – двумерную модель корреляционного анализа.

Для характеристики тесноты связи между двумя переменными обычно пользуются парным коэффициентом корреляции ρ , если рассматривать генеральную совокупность, или его оценкой – выборочным парным коэффициентом r , если изучается выборочная совокупность. Парный коэффициент корреляции в случае линейной формы связи вычисляют по формуле

$$\rho = \frac{M[X - M(X)][Y - M(Y)]}{\sigma_X \sigma_Y},$$

а его выборочное значение – по формуле

$$r = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{S_X S_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n S_X S_Y}$$

При малом числе наблюдений выборочный коэффициент корреляции удобно вычислять по следующей формуле:

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \cdot \sum_{i=1}^n Y_i}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 \right] \left[n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i \right)^2 \right]}}$$

Величина коэффициента корреляции изменяется в интервале $-1 \leq r \leq 1$.

При $r = -1$ между двумя переменными существует функциональная связь, при $r = 1$ - прямая функциональная связь. Если $r = 0$, то значения X и Y в выборке некоррелированы; в случае, если система случайных величин (X, Y) имеет двумерное нормальное распределение, то величины X и Y будут и независимыми.

Если коэффициент корреляции находится в интервале $-1 \leq r \leq 0$, то между величинами X и Y существует обратная корреляционная связь. Это находит подтверждение и при визуальном анализе исходной информации. В этом случае отклонение величины Y от среднего значения взяты с обратным знаком.

Если каждая пара значений величин X и Y чаще всего одновременно оказывается выше (ниже) соответствующих средних значений, то между величинами существует прямая корреляционная связь и коэффициент корреляции находится в интервале $0 \leq r \leq 1$.

Если же отклонение величины X от среднего значения одинаково часто вызывают отклонения величины Y вниз от среднего значения и при этом отклонения оказываются все время различными, то можно предполагать, что значение коэффициента корреляции стремится к нулю.

Следует отметить, что значение коэффициента корреляции не зависит от единиц измерения и выбора начала отсчета. Это означает, что если переменные X и Y уменьшить (увеличить) в K раз либо на одно и то же число C , то коэффициент корреляции не изменится.

Пакет анализа Microsoft Excel. В состав Microsoft Excel входит набор средств анализа данных (так называемый пакет анализа), предназначенный для решения сложных статистических и инженерных задач. Для проведения анализа данных с помощью этих инструментов следует указать входные данные и выбрать параметры; анализ будет проведен с помощью подходящей статистической или инженерной макрофункции, а результат будет помещен в выходной диапазон. Другие средства позволяют представить результаты анализа в графическом виде.

Графические изображения используются прежде всего для наглядного представления статистических данных, благодаря им существенно облегчается их восприятие и понимание. Существенна их роль и тогда, когда речь идет о контроле полноты и достоверности исходного статистического материала, используемого для обработки и анализа.

Статистические данные приводятся в виде длинных и сложных статистических таблиц (см., например, табл.1), поэтому бывает весьма трудно обнаружить в них имеющиеся неточности и ошибки.

Графическое же представление статистических данных помогает легко и быстро выявить ничем не оправданные пики и впадины, явно не соответствующие изображаемому статистическим данным, аномалии и отклонения. На графике, построенном по данным таблицы 1 (рис.1), наглядно показано распределение курса биржевых ставок в зависимости от времени совершения сделки и цены сделки в рублях.

Однако, чтобы эффективнее использовать графические изображения статистических данных, необходимо овладеть методикой и техникой их построения. К этому следует добавить, что построенное графическое изображение статистических данных биржевых ставок в наибольшей степени соответствует характеру и содержанию изображаемых данных и поставленной задаче их анализа.

Таблица 1. Выборка биржевых ставок относительно времени совершения сделки и цены сделки в рублях за один день работы биржи

	Время	Цена сделки в рублях
1.	11:16:45	99,45
2.	11:21:53	99,4
3.	11:23:09	99,31
4.	11:23:37	99,31
5.	11:24:49	99
6.	11:24:57	99
7.	11:48:40	98,61
8.	11:49:45	98,99
9.	11:53:51	98,66
10.	11:55:05	98,65
11.	11:55:24	98,7
12.	11:58:18	98,8
13.	11:58:18	98,8
14.	11:58:24	98,65
15.	11:58:35	98,8

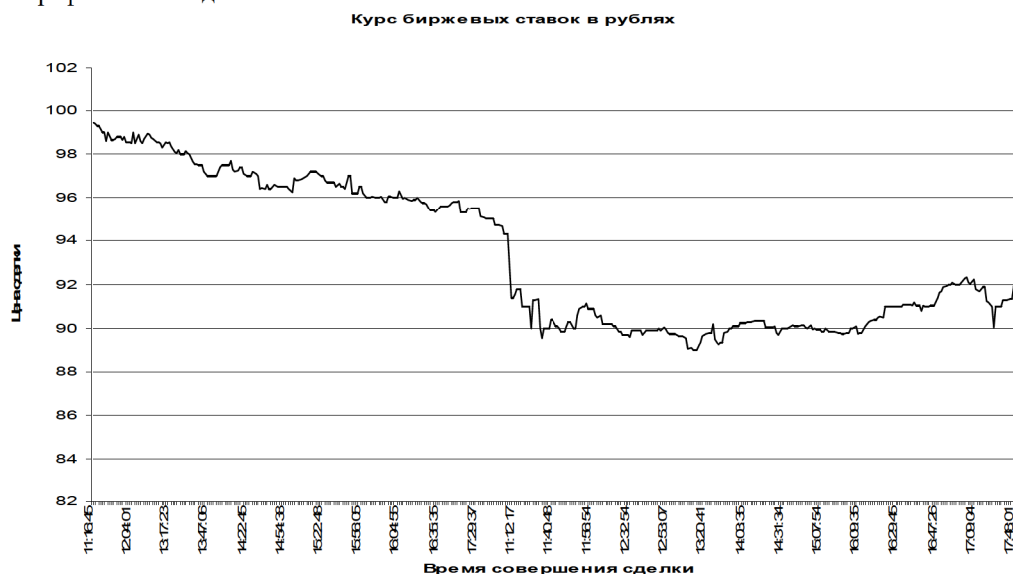


Рис.1 Распределение курса биржевых ставок в зависимости от времени совершения сделки и цены сделки в рублях.

Корреляция - один из инструментов пакета анализа Microsoft Excel. Используется для количественной оценки взаимосвязи двух наборов данных, представленных в безразмерном виде. Коэффициент корреляции выборки представляет собой ковариацию двух наборов данных, деленную на произведение их стандартных отклонений.

Корреляционный анализ дает возможность установить ассоциированы ли наборы данных по величине, то есть: большие значения из одного набора данных связаны с большими значениями другого набора (положительная корреляция); или, наоборот, малые значения одного набора связаны с большими значениями другого (отрицательная корреляция); или данные двух диапазонов никак не связаны (корреляция близка к нулю).

Регрессия также является инструментом пакета анализа данных Microsoft Excel. Линейный регрессионный анализ заключается в подборе графика для набора наблюдений с помощью метода наименьших квадратов. Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или более независимых переменных. Например, на курс биржевых ставок влияют несколько факторов, включая такие, как время совершения сделки и ее цена. Регрессия пропорционально распределяет меру качества по этим двум факторам на основе данных функционирования курса биржевых ставок. Результаты регрессии могут быть использованы для предсказания качеств новых, не совершенных еще биржевых сделок. Например, используя результаты таблицы 1, можно с помощью регрессии предсказать цены следующих сделок.

Табл.2. Предсказанная цена сделки в рублях

Наблюдение	Предсказанная цена сделки в рублях	Остатки
1	72,22015	27,22985
2	72,76796	26,63204
3	72,90313	26,40687
4	72,95293	26,35707
5	73,08099	25,91901
6	73,09522	25,90478
7	75,62617	22,98383
8	75,74178	23,24822
9	76,17932	22,48068
10	76,31094	22,33906
11	76,34473	22,35527
12	76,65421	22,14579
13	76,65421	22,14579
14	76,66488	21,98512
15	76,68444	22,11556

Наиболее сложным этапом, завершающим регрессионный анализ, является интерпретация полученных результатов, т.е. перевод их с языка статистики и математики на язык экономики.

Интерпретация моделей регрессии осуществляется методами той отрасли знаний, к которой относятся исследуемые явления. Всякая интерпретация начинается со статистической оценки уравнения регрессии в целом и оценки значимости входящих в модель факторных признаков, т.е. с изучения, как они влияют на величину результирующего признака. Чем больше величина коэффициента регрессии, тем значительнее влияние данного признака на моделируемую обработку биржевых ставок. Особое значение при этом имеет знак перед коэффициентом регрессии. Знаки коэффициентов регрессии говорят о характере влияния на результирующий признак статистической обработки биржевых ставок. Если факторный признак имеет плюс, то с увеличением данного фактора результирующий признак возрастает; если факторный признак со знаком минус, то с его увеличением результирующий признак уменьшается. Интерпретация этих знаков полностью определяется социально-экономическим содержанием моделируемого признака. Если его величина изменяется в сторону увеличения, то плюсовые знаки факторных признаков имеют положительное влияние. При изменении результирующего признака в сторону снижения положительные значения имеют минусовые знаки факторных признаков. Если экономическая теория подсказывает, что факторный признак должен иметь положительное значение, а он со знаком минус, то необходимо проверить расчеты параметров уравнения регрессии.

Корреляционный и регрессионный анализ позволяет определить зависимость между факторами, а так же проследить влияние задействованных факторов. Эти показатели имеют широкое применение в обработке статистических данных для достижения наилучших показателей биржевых ставок.

Литература

1. Колемаев В.А., Староверов О.В., Турундаевский В.Б. Теория вероятностей и математическая статистика / - М., 1991.
2. Теория Статистики / Под ред. Р.А. Шмойловой / «ФиС», 1998.
3. Многомерный статистический анализ на ЭВМ с использованием пакета Microsoft Excel / - М., 1997.
4. Френкель А.А., Адамова Е.В. Корреляционно регрессионный анализ в экономических приложениях / - М., 1987.
5. Одинцов И.Д. Теория статистики / - М., 1998.
6. Кленин А.Н., Шевченко К.К. Математическая статистика для экономистов-статистиков / - М., 1990.