

© Галина Викторовна Фетисова
кандидат экономических наук, доцент,
Институт экономики, управления и
права, кафедра экономики, Новгородский
государственный университет имени
Ярослава Мудрого
Galina.Fetisova@novsu.ru

Fetisova Galina Viktorovna
Candidate of Economic Sciences,
Associate Professor,
Institute of Economics, Management and
Law, Department of Economics,
Yaroslav the Wise Novgorod State
University

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ПЕРВИЧНЫХ ДАННЫХ КАК ОСНОВНОЕ ТРЕБОВАНИЕ К ЭМПИРИЧЕСКОЙ БАЗЕ ПРОГНОЗИРОВАНИЯ¹

Statistical processing of primary data as a basic requirement for an empirical forecasting base

Аннотация: в настоящем исследовании автором представлена характеристика понятия «эмпирическая база прогнозирования», принципы и требования к экономической информации, используемой для целей прогнозирования. В статье рассмотрены основные методы первичной и вторичной обработки экономической информации, позволяющие сформировать однородную качественную эмпирическую базу, необходимую для получения достоверных прогнозов.

Summary: In this study, the author presents a description of the concept of “empirical forecasting base”, principles and requirements for economic information used for forecasting purposes. The article discusses the basic methods of primary and secondary processing of economic information, allowing to form a homogeneous, high-quality empirical base necessary to obtain reliable forecasts.

Ключевые слова: первичная информация, эмпирическая база, экономическая информация, дескриптивная статистика, аномальные наблюдения, однородность выборки, нормальное распределение, прогнозирование.

Keywords: primary information, empirical base, economic information, descriptive statistics, abnormal observations, sample uniformity, normal distribution, forecasting.

Развитие прогнозтики в настоящее время характеризуется беспрецедентными масштабами прогнозирования. Прогнозы создаются как на уровне глобальных территориальных социально-экономических систем, так и на уровне локальных территорий. Роль прогнозирования в современной России значительно возрастает в связи с ускорением научно-технического прогресса, усложнением управленческих задач, отсутствием стабильности и большой степенью неопределенности. При принятии управленческих решений прогнозирование выступает первоосновой, поскольку оно направлено на предсказание результатов и последствий внедрения принятых стратегий.

Успешность прогнозирования зависит от множества условий: объема и качества информации о прогнозируемом процессе, объекте управления; правильности формулирования задачи прогнозирования и обоснованности выбора способа ее

¹ Исследование выполнено при финансовой поддержке РФФИ и правительства Новгородской области в рамках научного проекта «Обоснование дифференциации территорий по уровню социально-экономического развития как детерминанты демографической ситуации (на примере муниципальных образований Новгородской области)» № 18-410-530003.

решения; наличия необходимых вычислительных средств и вычислительного аппарата в соответствии с выбранным методом.

Важнейшей предпосылкой получения достоверного и качественного прогноза является наличие соответствующей эмпирической (информационной) базы, под которой «понимается совокупность сведений, данных, фактов, знаний относительно исследуемого объекта» [5]. Эмпирическая база прогнозирования отражает состояние объекта и определяет направления его изменения и развития. Важнейшее требование к эмпирической базе – ее соответствие целям и задачам прогнозирования и планирования.

Основным видом используемой в планировании и прогнозировании информации выступает экономическая информация. Она представляет собой совокупность различных сведений экономического характера, на основе которых осуществляется анализ, разрабатываются плановые и прогнозные документы, отражающие развитие народного хозяйства в целом и его отдельных звеньев [3]. Экономическая информация обладает следующим признаками: массовость, повторение циклов получения и обработки в установленные периоды, большой удельный вес сравнительно стабильных элементов, объединяемых для дальнейшей обработки или длительного хранения. В экономическую информацию входит и статистическая информация, отражающая процессы и явления в хозяйственной деятельности как отдельных предприятий и организаций, так и в отраслях национальной экономики. Получаемая из всех звеньев и необходимая для экономического анализа и сопоставления данных, экономическая информация должна отвечать принципам, вытекающим из методологии планирования и прогнозирования: сравнение величин, агрегирование и дезагрегирование единиц информации, установление информационных отношений между ними, она должна быть достоверной, полной, но без избыточной информации и своевременной [1, 3].

Нередко при использовании тех или иных математико-статистических методов в экономических исследованиях модели трудно интерпретировать, т.к. выводы идут вразрез с элементарными представлениями об исследуемом процессе. Это связано, в первую очередь, с нарушениями сбора информации и отсутствием ее первичной обработки. После принятия решения о том, какая информация будет служить для целей исследования, возникает вопрос о качестве данных со статистической (репрезентативность) и источниковедческой (надежность) точек зрения.

В настоящем исследовании обобщены теоретические положения о первичной статистической обработке информации, используемой в качестве эмпирической базы при прогнозировании социально-экономических явлений. Основная цель проверки – выделение для дальнейших исследований только той информации, которая несет необходимые сведения о явлении или процессе. Все методы первичной обработки информации можно разделить на первичные и вторичные.

Первичная статистическая обработка нацелена на упорядочение информации об объекте и предмете изучения. На этой стадии разрозненные сведения группируются по тем или иным признакам, обобщаются в виде сводных таблиц. Это позволяет исследователю в первом приближении иметь представление о характере всей совокупности данных: об их однородности (или неоднородности), компактности (или разбросанности), четкости и т.д. Эта информация хорошо считывается с наглядных форм представления данных и дает сведения об их распределении. К основным методам первичной статистической обработки можно отнести: вычисление показателей центра распределения (мер центральной тенденции) и показателей размера и интенсивности вариации (мер разброса данных). Показатели центра распределения: выборочная

средняя, мода и медиана – это величины, вокруг которых группируются остальные данные. По этим обобщающим показателям можно судить обо всей выборке, а также сравнивать разные выборки между собой. Сопоставление всех трех показателей между собой позволяет сделать вывод о симметричности (асимметричности) распределения. При нормальном распределении значения показателей выборочного среднего, медианы и моды равны или очень близки друг к другу [4].

Показатели размера и интенсивности вариации характеризуют различия между отдельными единицами выборки. Они позволяют сделать вывод о степени однородности выборочной совокупности, ее компактности и как следствие надежности полученных данных. В группу данных показателей можно отнести следующие, как наиболее используемые, меры вариации: размах вариации, среднее линейное отклонение, дисперсия, стандартное (среднее квадратическое) отклонение. Все показатели первичной обработки данных реализуются в ППП в разделе «Дескриптивная (описательная) статистика» [6].

Вторичная статистическая обработка позволяет выявить скрытые статистические закономерности в данных и состоит из трех последовательных этапов: на первом этапе проверяют, имеется ли между исходными данными автокорреляция; на втором этапе исходной информации исключаются резко выделяющиеся (аномальные) наблюдения; третьим этапом первичной статистической обработки исходных данных является их проверка на однородность информации, содержащейся в отдельных группах наблюдения.

Проверку на автокорреляцию необходимо проводить для всех исходных данных, представленных в виде временных рядов. Это необходимо для того, чтобы выяснить независимость некоторых членов статистического ряда в смысле теории вероятности, т.к. только при этом условии можно использовать классические методы математической статистики. Независимость уровней ряда означает, что они между собой не коррелируют, т.е. между соседними членами ряда отсутствует корреляционная связь или она незначительна. В случае, если она имеется и значима, ее необходимо устранить.

Чтобы определить наличие автокорреляции в данном временном ряду, вычисляется коэффициент автокорреляции, который рассматривает тесноту связи рассматриваемых переменных. Для совершенно случайных рядов последовательные значения должны быть независимы друг от друга и теоретическое значение коэффициента автокорреляции должно быть равно нулю. Чем сильнее связь между последовательными членами ряда, тем величина коэффициента автокорреляции выше.

Формула расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1) \times \sum_{t=2}^n (y_{t-1} - \bar{y}_2)}} \quad (1)$$

где y_t – уровни исходного временного ряда;

y_{t-1} – уровни ряда, сдвинутого на один лаг;

$\bar{y}_1$ – среднее значение исходного временного ряда;

$\bar{y}_2$ – среднее значение ряда, сдвинутого на один лаг.

Существует несколько способов выявления автокорреляции уровней временного ряда. Например, на основе формулы Г. Тинтнера и дальнейшей оценки коэффициента автокорреляции на основе таблиц Андерсона. Но чаще используется проверка автокорреляции на основе статистики Дарбина-Уотсона, поскольку она реализована во многих пакетах прикладных программ [5].

Выявление аномальных наблюдений из статистической информации – второй этап статистической обработки первичных данных. Для его реализации существуют

специально разработанные методы математической статистики, различающиеся сложностью применяемого математического аппарата.

Наиболее приемлемый способ – использование критерия Граббса для выявления аномальных наблюдений (грубых погрешностей) [2, 5]. Как правило, в поле зрения попадают значительно отличающиеся минимальное и максимальное значения совокупности. При его применении используются показатели выборочной средней и стандартного отклонения.

Формула критерия Граббса имеет вид:

$$t_{расч} = \frac{|x_{min} - \bar{x}|}{\sigma}, t_{расч} = \frac{|x_{max} - \bar{x}|}{\sigma} \quad (2)$$

где $\bar{x}$ – выборочная средняя;

σ – стандартное отклонение;

x_{min}, x_{max} – минимальное (максимальное) значение выборки.

Расчётное значение сравнивают с табличным t_{α} , при $t_{расч} > t_{\alpha}$ результат x считают грубой ошибкой и отбрасывают.

Таким образом, сравнение минимального и максимального значений экономического показателя со значениями дескриптивной статистики позволяет предположить наличие (отсутствие) грубых погрешностей в выборке.

Последним этапом вторичной статистической обработки исходных данных является проверка исследуемого ряда на однородность. При проверке ряда на однородность пользуются сравнением дисперсий двух или нескольких выборок исследуемого ряда [2, 5]. Однородность выборок проверяется по F-критерию Фишера при сопоставлении выборочных дисперсий между собой:

$$F = \frac{S_1^2}{S_2^2}, \quad (3)$$

где S_1^2, S_2^2 – выборочные дисперсии, определяемые с учетом числа степеней свободы (n_1-1) и (n_2-1) . При этом $S_1^2 > S_2^2$.

При заданном уровне значимости α определяют критические значения критерия и оценивают однородность совокупности.

Помимо рассмотренных методов, важно учитывать источники и возможности появления искажений экономической информации, а также влияние факторов на вариативность показателей.

Предложенные методы позволяют улучшить качество эмпирической базы исследования, что позволит получить более точные и надежные прогнозы экономических показателей.

Список литературы

1. Вишнев С. М. Основы комплексного прогнозирования. - М. Наука, 1977. – 279 с.
2. Дубров А. М. Последовательный анализ в статистической обработке информации. - М. : Статистика, 1976. – 160 с.
3. Лугачев М. И., Ляпунов Ю. П. Методы социально-экономического прогнозирования. - М. : Статистика, 1999. – 179 с.
4. Васильева Э. К. Статистика: учебник для студентов вузов, обучающихся по специальностям экономики и управления. - М. : ЮНИТИ-ДАНА, 2012. - 399 с.
5. Статистический факторный анализ в экономике: учебное пособие / сост. Л. И. Бернасовская, Г. В. Лебедева, О. Д. Притула. - Великий Новгород, 2002. - 144 с.
6. Статистическое моделирование и прогнозирование / под ред. А. Г. Гранберга. - М. : Наука, 1990. – 341 с.