

УДК 330.47

Н. Х. Ворокова, А. Е. Сенникова

ФГБОУ ВО «Кубанский государственный аграрный университет
имени И.Т. Трубилина», Краснодар, e-mail: v.nodira@gmail.com

ИНФОРМАЦИОННАЯ ДИАГНОСТИКА СОЦИАЛЬНЫХ ОБЪЕКТОВ И ПРОЦЕССОВ С ПОМОЩЬЮ МЕТОДОВ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ

Ключевые слова: методы описательной статистики, проблематика описательной статистики, диагностика социальных объектов.

На основе большинства методов описательного статистического анализа проводится диагностика экономических процессов в социальной сфере. Такой подход обеспечивает подтверждение или опровержение, а также, установление причинно-следственных связей динамики экономических показателей социального развития общества в целом. Ввиду многообразия методов описательной статистики необходимо их теоретическое упорядочение для возможности применения на практике. Выбор применения соответствующего метода обеспечивает повышение информативности и качества получаемых выводов об экономике социальной сферы. Основываясь на группе статистических и экономических аргументов, согласно существующей общей практике, второстепенные значения исключаются из выборки, что обеспечивает ее структурность и однородность совокупных данных об экономическом субъекте. Не во всех случаях репрезентативность выборки обеспечивается путем исключения максимальных или минимальных неоднородных значений. Так, репрезентативность всегда ограничена в той мере, в какой представлена выборка. И именно репрезентативность выборки является основным критерием при определении границ генерализации выводов экономического исследования. Отсюда следует, что размер диапазона исследуемых значений, необходимый для достоверного исследования, очень зависит от качества выявленных неоднородных значений, которые в свою очередь можно проанализировать, выбрав соответствующий метод описательной статистики. Таким образом, целесообразно в статье представление исследования сильных и слабых сторон методов описательной статистики, представление которых возможно в максимально простой форме: графической, табличной, числовой.

N. H. Vorokova, A. E. Sennikova

FSBEI HE “Kuban State Agrarian University named after I.T. Trubilin”,
Krasnodar, e-mail: v.nodira@gmail.com

INFORMATION DIAGNOSTICS OF SOCIAL OBJECTS AND PROCESSES USING DESCRIPTIVE STATISTICS METHODS

Keywords: methods of descriptive statistics, problems of descriptive statistics, diagnostics of social objects.

Most of the methods of descriptive statistical analysis are used to diagnose economic processes in the social sphere. This approach provides confirmation or refutation, as well as the establishment of cause-and-effect relationships in the dynamics of economic indicators of social development of society as a whole. In view of the variety of descriptive statistics methods, their theoretical ordering is necessary for their application in practice. The choice of using the appropriate method provides an increase in the information content and quality of the conclusions obtained about the economy of the social sphere. Based on a group of statistical and economic arguments, according to existing general practice, secondary values are excluded from the sample, which ensures its structure and homogeneity of aggregate data about an economic entity. Not in all cases the representativeness of the sample is ensured by excluding the maximum or minimum non-uniform values. Thus, representativeness is always limited to the extent that the sample is represented. And it is the representativeness of the sample that is the main criterion in determining the boundaries of generalization of the conclusions of economic research. It follows that the size of the range of the studied values required for a reliable study depends very much on the quality of the identified heterogeneous values, which, in turn, can be analyzed by choosing the appropriate method of descriptive statistics. Thus, it is advisable in the article to present the study of the strengths and weaknesses of descriptive statistics methods, the presentation of which is possible in the simplest possible form: graphical, tabular, numerical.

Введение

Органы статистики и ведомственные организации регулярно осуществляют сбор больших массивов данных различных экономических и социальных объектов, процессов. А обработка этих данных возможна с помощью методов описательной статистики. Правильность выбранного метода обеспечивает своевременность выявления экономических и социальных проблем в обществе, что может повлиять на различные общественные явления.

Особое значение вопросу выбора методов описательной статистики, в том числе для информационной диагностики социальных объектов, процессов придается в эпоху развития цифровой экономики страны.

Цель исследования заключается в выявлении сильных и слабых сторон методов описательной статистики для диагностики социально-экономической инфраструктуры в организациях или обществе в целом.

Для достижения данной цели необходимо решить следующие задачи:

- отразить теоретические алгоритмы наиболее распространенных описательных методов статистики обработки данных о социально-экономических процессах

- представить сильные и слабые стороны применения различных методов описательной статистики экономических, социальных объектов и процессов, дать рекомендацию их своевременного применения.

Материал и методы исследования

Для достижения поставленных задач на первом этапе статистического анализа необходимо представить общую характеристику собранных сведений о социальном объекте исследования. Это возможно осуществить посредством применения множества методов и методик описательной статистики, которые различаются по типу шкалы.

Для номинальных и порядковых шкал описательная статистика представлена круговыми и столбиковыми диаграммами, а также, частотными таблицами.

В свою очередь, частотные таблицы используются для наглядной группировки имеющихся для исследования данных. Далее, для упрощения работы с таблицей применяют графическое представление полученной информации путем построения столбиковых и круговых диаграммах. Следует отметить, что в столбиковых диаграм-

мах представлены абсолютные значения, а в круговых – процентные.

Результаты методов описательной статистики социальных объектов и процессов представляют собой значения таких данных, меры которых выражены как в усредненная оценка типичных явлений, так и на сколько точно подходят полученные значения для соответствующей выборки. Так, выделяют меры среднего уровня и меры рассеяния. Примерами мер среднего уровня являются: минимум, максимум, интервал, мода и так далее. Мерами рассеяния признаются: размах вариации, дисперсия случайной величины, среднее абсолютное отклонение и так далее.

Для количественных шкал описательных статистик гораздо больше. Рассмотрим некоторые из них.

Выборочная средняя арифметическая – самая распространенная описательная статистика, которая позволяет быстро охарактеризовать выборку, содержащую количественные значения. Это ее основной плюс. Тем не менее, среднее арифметическое значение не всегда адекватно отражает значение выборки (например, при изучении уровня заработных плат в разных организациях). Для оптимизации искаженных количественных данных выполняются следующие способы решения:

- разделение величин заработной платы исследуемой организации на группы в зависимости от занимаемой должности сотрудника.

- расчет медианы, то есть выполнение упорядочения выборки, расположив ее значения от большего к меньшему.

Если в выборке нечетное количество элементов, то необходимо найти число, которое делит медиану пополам. Если в выборке количество значений представлено четным числом, то необходимо найти два срединных элемента и найти их среднее арифметическое значение.

Если отразить такую отсортированную выборку в виде отрезков с минимумом в начале и максимум в конце, то медиана будет делить его пополам. Точка, отделяющая порядка 25 процентов выборки, начиная с минимального значения, называется нижняя четверть, а точка, отделяющая 75 процентов выборки, начиная с нижнего значения – верхняя четверть.

Квартильным размахом представляются 50 процентов имеющейся выборки данных.

Основываясь на данных оставшихся числовых значений выборки строится диаграмма размаха или ящик с усами (box plot). Она необходима для диагностики возникновения ошибочных значений, медианы и распределения данных. На диаграмме размаха есть: тело ящика, которое характеризует квартильный размах, медиана, представленная линией или точкой внутри тела ящика. Нижняя и верхняя стенки ящика – это нижняя и верхняя квартили. Расстояние между кончиками усов – это размах выборки без учета выбросов грубых ошибок.

Выбросы и грубые ошибки – это значения, существенно отличающиеся от большинства значений выборки, которые находятся на расстоянии 1,5 квартильных размаха ниже и выше верхней и нижней квартилей. Наиболее грубые ошибки находятся на расстоянии трех квартильных размаха, именуемая также зоной минимальных и максимальных значений. По итогам анализа диаграммы экстремальные значения должны быть исключены из анализа и рассматриваться, при необходимости, отдельно. По поводу выбросов решение об их исключении может быть принято с учетом знания об изучаемом явлении или для реализации определенной цели анализа.

Помимо квартильного размаха рассчитывают обычный размах выборки. Это разница между максимальным и минимальным значением выборки.

Размах выборки используют, чтобы понять, насколько сильно варьируется изменяемый признак: чем больше значение размаха, тем сильнее он варьируется. Чем больше ящик по размерам, тем больше количество наблюдений в выборке, по которой строился этот ящик, что касается доверительного интервала, то он представляет собой выемки на каждом из ящиков, в случае, если полученные выемки разных ящиков не пересекаются, их медианы статистически значимо различаются.

Несмотря на свою простоту и удобство, первоначальная форма ящика с усами обладает и некоторыми недостатками. Один из таких существенных недостатков заключается в полном отсутствии на графике информации о количестве наблюдений по исследуемой выборке. Ящик с усами позволяет сравнить медианы, квартили, минимумы и максимумы по различным выборкам, но если необходимо получение вывода об общей медиане по всей совокупности

выборок, то появляются некоторые затруднения, связанные с невозможность в полной мере осуществить расчеты, не прибегая к исходным данным [1].

Дисперсия – мера рассеивания переменной, которая рассчитывается как средняя арифметическая квадратов отклонения переменной от ее среднего значения. Таким образом, дисперсия отражает варьирование значений переменной вокруг среднего значения. То есть, разброс вокруг среднего значения [6]. С ее помощью можно сравнивать выборки, которые различаются по объему.

Применение дисперсии, как метода описательной статистики целесообразно использовать для анализа инвестиций. Свойство волатильности данного метода заключается в том, что появляется возможность определения инвесторами возможных рисков при покупке или продаже того или иного актива, а также, предположить величину прибыли на краткосрочный период времени. Полученная информация об инвестициях может свидетельствовать о совокупной эффективности совершения финансовой сделки в целом.

Отметим что дисперсия не удобна для интерпретации, так как имеет другую размерность, нежели исходные данные. Например, если исходные данные даны в метрах, то при расчете дисперсии подойдут метры квадратные, то есть линейная единица длины станет единицей площади. В связи с этим рассчитываются отдельно выборочные стандартные отклонения.

Производная от дисперсии, но имеющая такую же размерность, как и исходные данные – стандартное отклонение. Расчет данных согласно принципу стандартного отклонения представляет совокупность как близкую или нет от среднего значения. Плюсом применения расчета стандартного отклонения является возможность анализа данных во всей выборке или только ее части.

Производная от дисперсии, но имеющая такую же размерность, как и исходные данные – стандартное отклонение. Расчет данных согласно принципу стандартного отклонения представляет совокупность как близкую или нет от среднего значения. То есть, применение данного метода описательной статистики исследование результаты исследования можно предопределить или принять варьируемые значения относительно средней величины.

Чем меньше значение стандартного отклонения, тем ниже вероятность возникновения широкого колебания значений исследуемого процесса социальных объектов. То есть, сокращаются риски существенного изменения прогнозируемого числа на краткосрочный период времени. Плюсом применения расчета стандартного отклонения является возможность анализа данных, как во всей выборке, так и только ее части.

Следует отметить, что использование данного метода дает возможность определения факторов риска принятия каких-либо экономических или социальных мер. Тем не менее, данный метод не достаточно емок, чтобы говорить об однозначности полученных результатов или сделанных выводов.

Заключение

Согласно проведенным исследованиям, представленные в статье методы описательной статистики оценки социальных и экономических процессов наиболее приемлемы, поскольку результаты, возможно преобразить в доступную для расшифровки форму посредством визуализации данных.

Тем не менее, представленные методы не являются универсальными и содержат определенный набор приоритетных задач расчета. Таким образом, для полноты, качества исследования следует применять соответствующую методiku, наиболее подходящую для того или иного анализа экономического исследования.

Библиографический список

1. Benjamini Y. Opening the Box of a Boxplot (англ.) // The American Statistician. 1988. Vol. 42. № 4. P. 257-262.
2. Алибеков И.Ю. Теория вероятностей и математическая статистика в среде MATLAB: учебное пособие. М.: Лань, 2019. 184 с.
3. Боровков А.А. Теория вероятностей. Глава 4: Числовые характеристики случайных величин. Глава 5: Дисперсия. 5-е изд. М.: Либроком, 2009. 656 с.
4. Горленко О.А., Борбаць Н.М. Статистические методы в управлении качеством: учебник и практикум. М.: Юрайт, 2020. 306 с.
5. Горобец Б.С. Теория вероятностей, математическая статистика и элементы случайных процессов. Упрощенный курс. М.: Едиториал УРСС, 2020. 232 с.
6. Медведева С.И. Влияние уровня образованности на качество жизни населения России // Известия Оренбургского государственного аграрного университета. 2007. № 4 (16). С. 183-188.
7. Шушерина О.А. Математическая статистика: учеб. пособие / НОУ ВПО Университет Российской академии образования, Красноярский филиал. Красноярск, 2011. 60 с.