

СОВРЕМЕННЫЙ ИНСТРУМЕНТАРИЙ АНАЛИЗА И УПРАВЛЕНИЯ ЭКОНОМИЧЕСКИМИ ПРОЦЕССАМИ

УДК 311.332

А. П. Дарманян

ИСПОЛЬЗОВАНИЕ ПОКАЗАТЕЛЕЙ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ ДЛЯ ХАРАКТЕРИСТИКИ ЭМПИРИЧЕСКИХ ВЫБОРОК МАКРОЭКОНОМИЧЕСКИХ ИНДИКАТОРОВ

В статье выполнена статистическая проверка эмпирических распределений макроэкономических индикаторов РФ (среднемесячная номинальная начисленная зарплата, уровень безработицы) в 2011 г. и уровня безработицы стран Евросоюза в 2011 г. на соответствие их нормальному закону распределения. На основе расчета статистического критерия Колмогорова — Смирнова доказано, что указанные распределения не являются нормальными распределениями. Поэтому для их характеристики неправомерно использовать средние арифметические значения в качестве макроэкономических индикаторов состояния и развития. Для выбранных эмпирических распределений выполнены расчеты таких показателей описательной статистики, как медиана, нижний и верхний квартили, интерквартильный размах, выбросы и экстремальные значения, показаны преимущества графического представления данных в виде диаграммы «ящик с усами». На примере данных двух российских округов — Южного ФО и Приволжского ФО — проведено сравнение среднемесячных зарплат с помощью непараметрического критерия Манна — Уитни, показано различие в интерпретации данных по сравнению с использованием средних арифметических значений. Результаты выполненных расчетов могут быть использованы для анализа макроэкономических индикаторов среднемесячной зарплат и уровня безработицы в 83 субъектах РФ в 2011 г., уровня безработицы в странах Евросоюза в 2011 г., а также в качестве методологии статистического анализа различных эмпирических выборок.

Ключевые слова: макроэкономические индикаторы, среднемесячная зарплата, уровень безработицы, нормальное распределение, медиана, интерквартильный размах, квартиль, диаграмма «ящик с усами», критерий Колмогорова — Смирнова, критерий Манна — Уитни

В статистической отчетности, которую представляют органы государственной статистики (Росстат — в РФ [4]), (Евростат (Eurostat) — в странах Евросоюза [6]), имеются различные макроэкономические индикаторы, полученные методами описательной статистики в результате обработки эмпирических выборок. При этом в качестве центральной тенденции эмпирического распределения используются средние арифметические значения за определенный временной период (в Росстате — «в среднем за период» [4]). Анализ социально-экономического состояния развития РФ и отдельных ее субъектов (федеральных округов — ФО) при таком подходе сводится, как правило, к сравнению средних арифметических значений макроэкономических индикаторов для одного временного периода по различным качес-

твенным критериям (регионы, города, отрасли производства, виды деятельности и т. д.), а для разных временных периодов — как процентное изменение среднего значения выбранного макроэкономического показателя («в % к соответствующему периоду предыдущего года» [4]).

Однако использование среднеарифметических значений правомерно только в том случае, когда анализируемое эмпирическое распределение имеет нормальное распределение (распределение Гаусса) в виде колоколообразной кривой, а 68,3% значений распределения лежат в интервале «среднее значение плюс (минус) стандартное отклонение» [1, с. 188]. Только в этом случае среднее арифметическое значение эмпирического распределения может быть использовано в качестве макроэко-

номического индикатора, а его процентное изменение — как характеристика различия или темп роста (убыли).

В случае отсутствия нормального распределения более правильно и информативно использовать такие показатели описательной статистики [3, с. 71], как: медиана, минимальное и максимальное значение показателя, интерквартильный размах (*IQR* — разница значений нижнего 25-го и верхнего 75-го квартилей), а в качестве показателя различия двух несвязанных выборок — непараметрический критерий, например Манна — Уитни [5, с. 112] вместо часто используемого для этих целей *t*-критерия [1, с. 231], который справедлив только для сравнения средних значений нормальных распределений.

Медиана является центром распределения и лучше, чем среднее арифметическое, характеризует эмпирическое распределение при асимметричном распределении и малом числе значений выборки. Квартили делят выборку на четыре равные части, а интерквартильный размах, в котором располагается 50% всех данных, дает ясное представление о ширине распределения и таким образом, о социально-экономических различиях, например, регионов. Еще одно преимущество таких статистических показателей, как медиана и квартили, в том, что они не подвержены влиянию выбросов и экстремальных значений, которые всегда имеют место в эмпирическом распределении любого экономического показателя в такой огромной и разноплановой в социальном и экономическом смысле стране, как Россия.

Нелишне отметить, что такой подход уже много лет регламентирован для зарубежных медиков-исследователей, использующих медицинскую статистику для обработки распределений эмпирических данных, отличающихся (как и в экономике) большим разбросом значений и малыми выборками [2, с. 36].

Поэтому цель данной работы заключалась в проверке соответствия отдельных эмпирических распределений макроэкономических индикаторов РФ и стран ЕС нормальному закону распределения для доказательства правомерности (или неправомерности) использования средних арифметических значений распределений в качестве характеристик распределений.

Для исследования были выбраны следующие макроэкономические индикаторы: среднемесячная номинальная начисленная зарплата и уровень безработицы населения РФ в 2011 г., уровень безработицы в странах

Таблица 1

Проверка нормальности распределения социально-экономических показателей для уровня значимости $p = 0,05$

Показатель	Критерий Колмогорова — Смирнова (<i>K-S-test</i>)	Вывод
Среднемесячная номинальная начисленная зарплата работников в 83 субъектах РФ в 2011 г.	0,223	Не является
Уровень безработицы в 83 субъектах РФ в 2011 г.	0,300	Не является
Уровень безработицы в 27 странах ЕС в 2011 г.	0,194	Не является

Евросоюза (ЕС) в 2011 г. Проверка соответствия этих распределений нормальному закону распределения проводилась с помощью критерия Колмогорова — Смирнова (*K-S-test*), расчеты которого выполнялись с помощью программы Statistica v.6.0 [5].

По данным Росстата [4], на графике рис. 1 представлена гистограмма распределения среднемесячной зарплаты работников в 83 субъектах РФ, на которой также показан график нормального распределения. Как видно из этого графика, распределение среднемесячной зарплаты по субъектам РФ не является нормальным распределением, и поэтому утверждение о том, что «в среднем начисленная в РФ среднемесячная зарплата в 2011 г. составляла 21607 руб.» (табл. 2), некорректно. Причина этого в наличии регионов с большими значениями зарплат, превышающими 40000 руб. (рис. 1). С точки зрения описательной статистики такие значения являются «выбросами и экстремальными значениями», которые оказывают сильное влияние на вычисление средних значений эмпирических распределений.

Этот же вывод доказывают результаты расчетов критерия Колмогорова — Смирнова (табл. 1). Так как нормальное распределение для уровня значимости $p = 0,05$ (95% уровень достоверности) имеет место тогда, когда $K-S-test \leq 0,0895$ [5], то из данных табл. 1 видно, что анализируемое эмпирическое распределение не является нормальными, и к нему такие показатели описательной статистики, как среднее арифметическое значение, стандартное отклонение и доверительный интервал не применимы. Неправомерно также использовать процентные различия средних значений макроэкономических индикаторов для сравнения различных выборок (регионов РФ), а также использовать процентные изменения средних значений в качестве показателей темпов роста или убыли.

Таблица 2

Показатели описательной статистики среднемесячной номинальной начисленной зарплаты работников, руб.

Регион	Среднее	Медиана	Минимум	Максимум	Нижний квартиль	Верхний квартиль	Размах (IQR)
РФ	21607	18416	11235	59095	15588	23302	7713
ЮФО	13269	12821	10244	16582	11663	14856	3193
ПрФО	14758	14489	11883	17438	13293	16378	3085

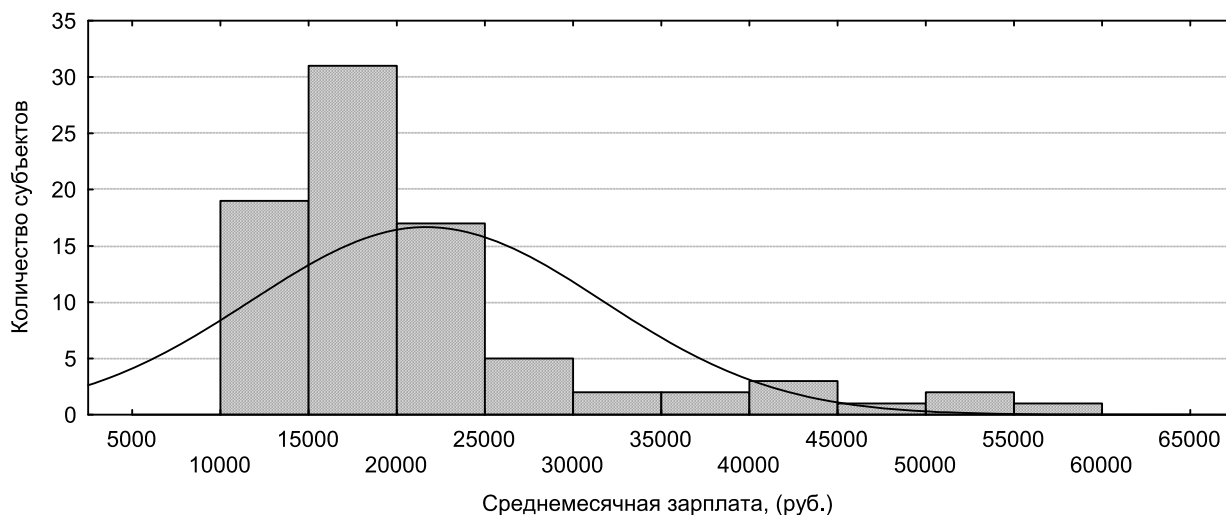


Рис. 1. Гистограмма распределения среднемесячной зарплаты работников в 83 субъектах РФ в 2011 г. (сплошная линия — нормальное распределение)

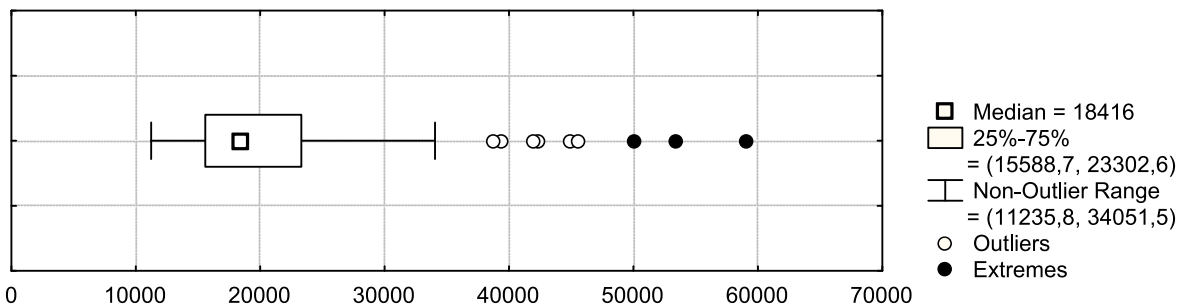


Рис. 2. Диаграмма «ящик с усами» распределения среднемесячной зарплаты работников в 83 субъектах РФ (руб.) в 2011 г.

При отсутствии нормального распределения необходимо использовать другие показатели описательной статистики. Поэтому в табл. 2 для РФ в целом (по данным Росстата для 83 субъектов РФ в 2011 г.) и для 2 регионов: Южного ФО (ЮФО) и Приволжского ФО (ПрФО) (по данным Росстата для 2010 г.) приведены результаты расчетов отдельных показателей описательной статистики.

При отсутствии нормального распределения в качестве графической иллюстрации эмпирического распределения удобно использовать диаграмму типа «ящик с усами» («Box & Whisker Plot»), которая обеспечивает диагностическую, описательную и наглядную информацию об исследуемой совокупности эмпирических данных [1, с. 159; 3, с. 57]. Поэтому на рис. 2 приведена графическая иллюстрация

распределения среднемесячной зарплаты в РФ в 2011 г. в виде диаграммы «ящик с усами». Левая граница «ящика» соответствует нижнему квартилю, правая — верхнему квартилю, внутри «ящика» прямоугольная отметка — медиана (*median*), справа и слева от «ящика» помещены «усы» (*Non-Outlier Range*), длина которых равна $1,5 \times IQR$.

Как видно из данных табл. 2 и рис. 2, в 2011 г. в половине регионов РФ среднемесячная зарплата работников была меньше 18416 руб., а в половине — больше этой величины (значение медианы), причем в половине всех регионов РФ (50%) среднемесячная зарплата находилась в интервале значений от 15588 руб. до 23302 руб., т. е. интерквартильный размах (*IQR*) составлял 7713 руб. В то же время так называемая среднемесячная зарплата (сред-

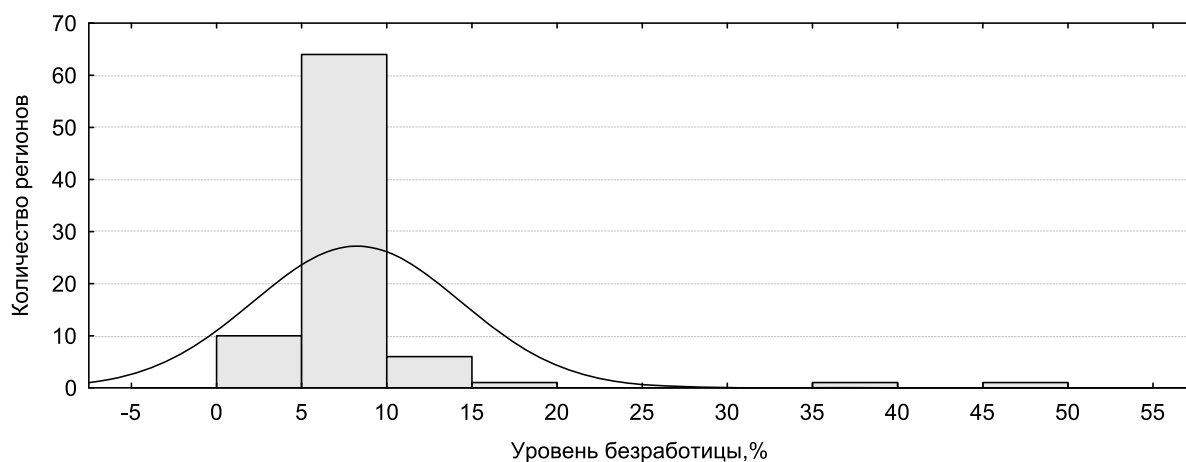


Рис. 3. Гистограмма распределения уровня безработицы в 83 субъектах РФ в 2011 г. (сплошная линия — нормальное распределение)

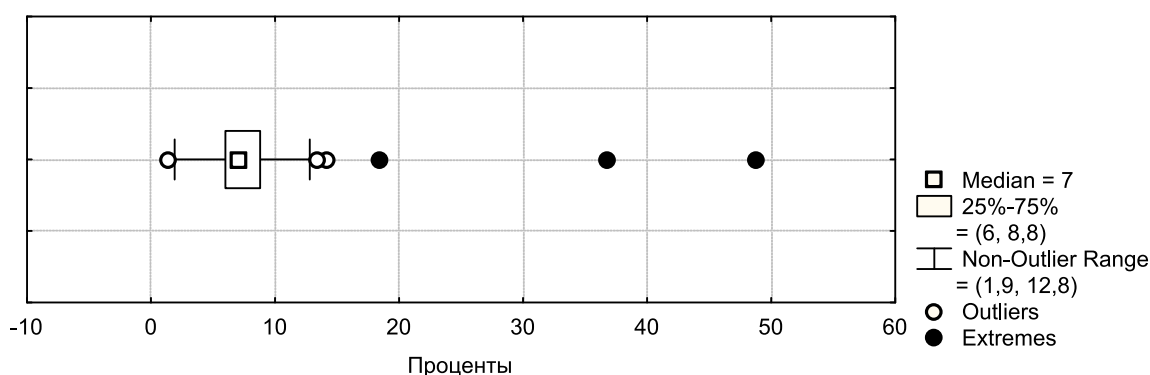


Рис. 4. Диаграмма «ящик с усами» распределения уровня безработицы в 83 субъектах РФ (%) в 2011 г.

неарифметическое значение или «в среднем за период») в РФ в 2011 г. составляла по тем же данным 21607 руб. [4], что, на наш взгляд, не очень информативно для такой большой страны как Россия.

В табл. 2 обращает на себя внимание сильное различие между минимальным и максимальным значением среднемесячной зарплаты в субъектах РФ, которые составляли 11235 руб. и 59095 руб., соответственно. Также на диаграмме рис. 2 обращают на себя внимание выбросы (*Outliers*), которые соответствуют среднемесячным зарплатам в регионах больше, чем верхняя граница «ящика» + $1,5 \times IQR$, а также — экстремальным значениям среднемесячных зарплат (*Extremes*), которые больше, чем верхняя граница «ящика + $2 \times IQR$ [3, с. 60]. Это говорит о том, что в отдельных регионах РФ «экстремальные среднемесячные зарплат» в 2011 г. были в 3 раза выше среднемесячной зарплаты, которую имели работники в половине субъектов РФ, и составляли более 40000 руб. в месяц. Это лишний раз подтверждает факт большого разброса данных среднемесячной зарплаты в различных регионах РФ, что и оказывает сильное влияние на вычисление средних арифметических значений.

Далее выполнен анализ уровня безработицы (доля безработных в экономически активном населении, %) в РФ в 2011 г. по данным Росстата и в странах Евросоюза по данным Евростата в 2011 г. [3, 6]. На диаграммах рис. 3 и рис. 4 показано распределение уровня безработицы в субъектах РФ, в табл. 1 приведены результаты проверки нормальности распределения, а в табл. 3 — результаты расчетов показателей описательной статистики.

Как видно из данных табл. 1 и рис. 3, распределение уровня безработицы в субъектах РФ не является нормальным распределением, и поэтому утверждение о том, что «в среднем уровень безработицы в РФ в 2010 г. составил 8,2%», некорректно. Произошло это из-за регионов с большими уровнями безработицы, превышающими 20 и 40% (см. рис. 3 и диаграмму на рис. 4), которые являются «экстремальными значениями».

Полученные данные говорят, скорее, о том, что в 2011 г. уровень безработицы в половине регионов РФ был ниже 7% (значение медианы в табл. 3), а в половине — больше 7%, причем для всех регионов имел место очень узкий интерквартильный размах величиной 2,8% (от 6,0% до 8,8%).

Таблица 3

Показатели уровня безработицы, %

Регион	Среднее	Медиана	Минимум	Максимум	Нижний квартиль	Верхний квартиль	Размах (IQR)
РФ	8,2	7,0	1,4	48,8	6,0	8,8	2,8
ЕС	9,9	8,2	4,2	21,7	7,2	12,9	5,7

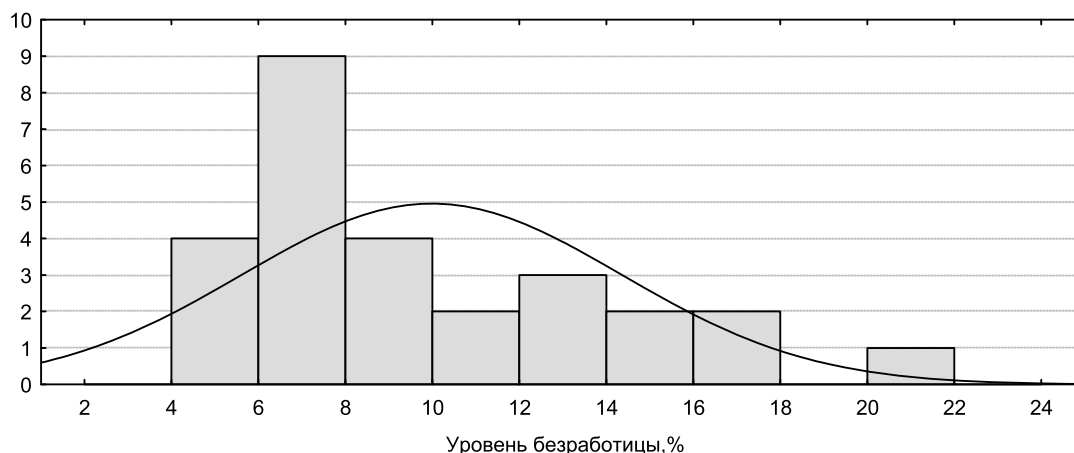


Рис. 5. Гистограмма распределения уровня безработицы в 27 странах ЕС в 2011 г. (сплошная линия — нормальное распределение)

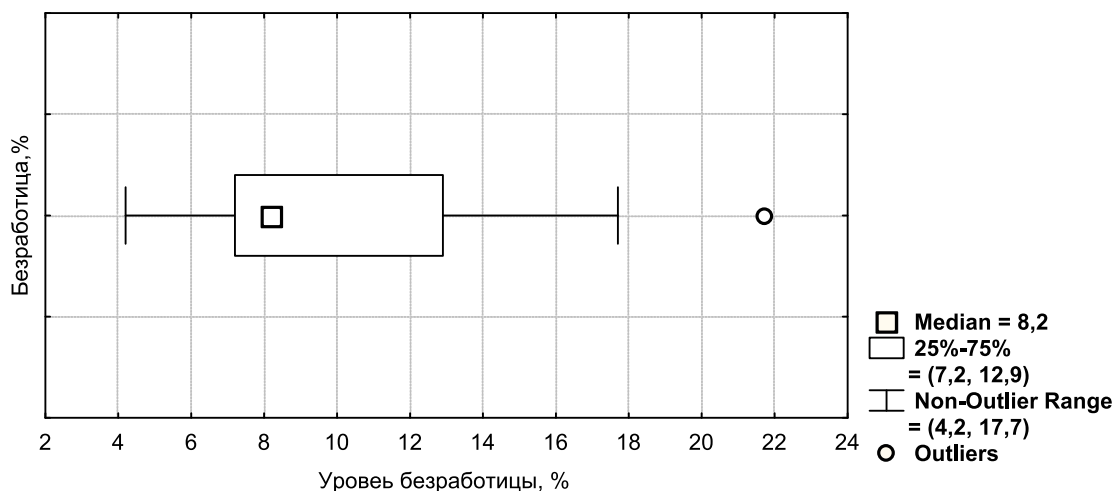


Рис. 6. Диаграмма «ящик с усами» распределения уровня безработицы в 27 странах ЕС в 2011 г.

Для анализа уровня безработицы в странах Евросоюза были использованы статистические данные 27 стран ЕС, которые даны в табл. 4. На диаграммах рис. 5 и рис. 6 показано распределение уровня безработицы в странах ЕС, в табл. 1 приведены результаты проверки нормальности распределения, а в табл. 3 представлены результаты расчетов показателей описательной статистики.

Как видно из данных табл. 1 и рис. 5, распределение уровня безработицы в странах ЕС не является нормальным распределением и поэтому утверждение о том, что «в среднем уровень безработицы в ЕС в 2011 г. составил 8,0%» некорректно и не отражает реальной безработицы в странах Евросоюза.

Скорее, можно утверждать, что в половине стран Евросоюза уровень безработицы в 2011 г. был ниже 8,2% (значение медианы в табл. 3), а в половине стран — больше 8,2%, причем половина (50%) стран Евросоюза имели уровень безработицы в диапазоне от 7,2% до 12,9% (интерквартильный размах — 5,7%). Кроме того, наличие больших «усов» на распределении (диаграмма на рис. 6) свидетельствует о большом разбросе уровня безработицы в странах ЕС (в отличие от регионов РФ!), причем на распределении имеется только один явный «выброс» (*Outliers*) величиной 22% (см. Испания в табл. 4).

Кроме анализа среднего значения какого-либо макроэкономического показателя для

Уровень безработицы в странах Евросоюза в 2011 г. [6]

Страна	Уровень безработицы, %	Страна	Уровень безработицы, %
Австрия	4,2	Германия	5,9
Бельгия	7,2	Греция	17,7
Болгария	11,3	Венгрия	10,9
Кипр	7,9	Ирландия	14,4
Чехия	6,7	Италия	8,4
Дания	7,6	Латвия	16,2
Эстония	12,5	Люксембург	4,8
Финляндия	7,8	Мальта	6,5
Франция	9,6	Нидерланды	4,4
Польша	9,7	Румыния	7,4
Португалия	12,9	Словакия	13,6
Словения	8,2	Швеция	7,5
Испания	21,7	Великобритания	8,0
Литва	15,4		

всей страны в целом, часто бывает необходимо провести сравнительный анализ макроэкономических показателей по разным регионам. Для этих целей нужно использовать, как отмечалось выше, непараметрические критерии ввиду отсутствия нормального распределения данных.

В качестве примера в настоящей работе выполнено сравнение среднемесячных зарплат работников в двух округах (ЮФО и ПрФО) с целью ответа на вопрос: а имелись ли различия в среднемесячных зарплатах в этих округах в 2010 г.?

Согласно выполненным расчетам (см. табл. 2), в ЮФО средняя зарплата составляла 13269 руб., а в ПрФО — 14758 руб., т. е. в 2010 г. имело место отличие среднемесячных зарплат в этих регионах. Можно ли это отличие считать существенным? Для ответа на этот вопрос был вычислен непараметрический критерий Манна — Уитни, который используется для сравнения несвязанных выборок и является менее чувствительным к экстремальным значениям [5, с. 112]. Расчет показал, что данное отличие статистически незначимо на уровне достоверности $p = 0,0133$, и поэтому можно доказательно заключить, что в сравниваемых округах ЮФО и ПрФО среднемесячные зарплаты работников в 2010 г. были практически одинаковы. Этот пример лишний раз подтверждает некорректность тех выводов, которые делаются на основании сравнения средних арифме-

тических значений эмпирических распределений макроэкономических индикаторов в случае отсутствия нормальности распределения.

На основе выполненного исследования можно сделать следующие выводы:

1. Прежде чем вычислять показатели описательной статистики по эмпирическим выборкам макроэкономических индикаторов, необходимо проверить исследуемую выборку на соответствие нормальному закону распределения.

2. В случае отсутствия нормального распределения для характеристики эмпирических выборок неправомерно использовать средние арифметические значения и доверительные интервалы, а также процентные изменения средних арифметических значений в качестве темпа роста или убыли.

3. Более информативно и корректно для описания эмпирических распределений макроэкономических индикаторов использовать такие показатели описательной статистики, как медиана, нижняя и верхняя квартиль, интерквартильный размах, выбросы и экстремальные значения, а в качестве графических иллюстраций распределений — диаграмму типа «ящик с усами».

4. Для выявления различий между эмпирическими выборками различных регионов (или других несвязанных выборок) необходимо использовать непараметрические критерии, например, критерий Манна — Уитни.

Список источников

1. Берк К., Кейри П. Анализ данных с помощью Microsoft Excel : пер с англ. — М.: Издательский дом «Вильямс», 2005. — 560 с.

2. Ланг Т. А., Сесик М. Как описывать статистику в медицине. Аннотированное руководство для авторов, редакторов и рецензентов : пер. с англ. / Под ред. В. П. Леонова. — М.: Практическая медицина, 2011. — 480 с.
3. Методы анализа распределений. Выборочное наблюдение : учеб. пособие. / Н. В. Куприенко и др.; 3-е изд. — СПб.: Изд-во Политехн. ун-та, 2009. — 138 с.
4. Российский статистический ежегодник. 2011 : стат. сб. / Росстат. — М., 2011. — 795 с.
5. Халафян А. А. STATISTICA 6. Статистический анализ данных : учебник; 3-е изд. — М.: ООО «Бином-Пресс», 2007. — 512 с.
6. EU employment statistics. [Электронный ресурс]. URL: http://epp.eurostat.ec.europa.eu/statistics_explained//index/php (дата обращения 09.01.2013)

Информация об авторе

Дарманян Анатолий Петрович (Волгоград, Россия) — доктор технических наук, профессор кафедры математики и информатики, Волгоградский филиал Финансового университета при Правительстве РФ (400078, г. Волгоград, ул. Кубинская, 26, e-mail: adarma@inbox.ru).

A. P. Darmanyan

The use of descriptive statistics for the characteristics of the empirical samples macroeconomic indicators

In the article, the statistical tests of the empirical distributions of economic indicators of the Russian Federation in 2011 (average monthly nominal accrued salary, the unemployment rate), the unemployment rate of the countries of the European Union in 2011, on conformity to their normal distribution law (distribution fitting) are made. On the basis of the calculation of the statistical criterion of the Kolmogorov-Smirnov, (K-S test) the distributions being not normal distributions is proved. Therefore, for their characteristics it is inappropriate to use the arithmetic mean as indicators of economic status and development. For the selected empirical distributions of the calculations of such indicators descriptive statistics, as the median, the lower and upper quartiles, interquartile range are conducted, the advantages of graphical representation of the data in the form of a chart «box and whisker» are shown. On the example of the two Russian districts of the Southern Federal district and the Volga Federal district, the comparison of the average monthly salary in 2010 with the help of nonparametric test Mann-Whitney U test is carried out, the difference in the interpretation of data compared with the use of simple averages is shown. The results of the performed calculations can be used for the analysis of economic indicators average monthly salary and the unemployment rate in 83 subjects of the Russian Federation in 2011, the unemployment rate in the EU countries in 2011, as well as the methodology of statistical analysis of the different empirical distribution economic indicators.

Keywords: economic indicators, average salary, unemployment rate, descriptive statistics, the arithmetic mean, the normal distribution, median, interquartile range, outlier, quartile, chart «Box-and-Whisker Plot», nonparametric test Mann-Whitney U test.

References

1. Berk K., Keyri P. (2005). Analiz dannykh s pomoshchyu Microsoft Excel: per.s angl. [Data Analysis with the Help of Microsoft Excel: translated from Engl], Moscow, Vilyams Publ, 560.
2. Lang T. A., Sesik M.; Leonov V. P. (Ed.) (2011). Kak opisyyvat statistiku v meditsine. Annotiravannoye rukovodstvo dlya avtorov, redaktorov i retsenzentsov: per. s angl. [How to Describe Statistics in Medicine. The Annotated Management for Authors, Editors and Reviewers:]. Prakticheskaya meditsina [Applied medicine], 480.
3. Kuprienko N. V. et al. (2009). Metody analiza raspredeleniy. Vyborochnoye nablyudeniye: ucheb posobiye [Methods of the Distributions Analysis. Selective Supervision: studies. book]. 3rd edition. St.-Petersburg. Polytechnical University Publ., 138.
4. Rossiyskiy statisticheskiy ezhegodnik. 2011: stat. sb. [Statistical Yearbook of Russia. 2011; statistic book]. (2011). Rosstat, Moscow, 795.
5. Khalafyan A. A. (2007). STATISTICA 6. Statisticheskiy analiz dannykh: uchebnik; 3-e izd. [Statistical analysis of data: text-book; 3rd ed.]. Moscow, Ltd. «Binom-Press», 512 с.
6. EU employment statistics. Available at: http://epp.eurostat.ec.europa.eu/statistics_explained//index/php (date of access: 09.01.2013)

Information about the author

Darmanyan Anatoly Petrovich (Volgograd, Russia) — Doctor of Technical Sciences, Professor, Department of Mathematics and Computer Science, Volgograd Branch of Financial University under the Government of the Russian Federation (400078, Volgograd, Kubinskaya St., 26, e-mail: adarma@inbox.ru).