

СМЕШАННЫЙ МЕТОД ОПИСАТЕЛЬНОЙ СТАТИСТИКИ НА ПРИМЕРЕ ДАННЫХ О СРЕДНЕЙ ПЕНСИИ В КЫРГЫЗСТАНЕ В 2014 г.

СУЛЕЙМАНОВА Г.А.
Кыргызско-Турецкий Университет Манас

УДК 311.1+519.2

Аннотация: Целью настоящей работы является описание смешанного метода исследования, позволяющего на основе единой группировки исходных данных решать задачи описательной статистики и построения гистограммы.

Такого рода задачи являются основой практически каждого количественного анализа данных. Например, классический подход к фундаментальной задаче нахождения средних и построения гистограммы, назовем ее смешанной задачей, предусматривает разделение этой задачи на две отдельные типовые подзадачи: (1) нахождение средних величин; (2) построение гистограммы. Эти две отдельные задачи затем решаются по двум разным методам, в каждом из которых исходные данные группируются. Таким образом, группировка данных при решении смешанной задачи, описанной выше, производится дважды, что трудоемко. Возникает вопрос: можно ли смешанную задачу решить, группируя данные только один раз? Преимущества такого решения очевидны – это позволит снизить объем вычислений вдвое. По мнению автора, такое решение возможно, поскольку оба метода решения типовых подзадач имеют в своем решении одинаковые части – группировку данных. Автором, на основе широко известных и популярных методов описательной статистики: *stemplot* и классического метода построения гистограммы, описанными John Tukey и Richard Thomas соответственно; предлагается смешанный метод описательной статистики и построения гистограммы. На примере данных из открытых источников о средней пенсии в Кыргызстане, в работе приводится решение смешанной задачи классическими методами и смешанным методом, анализируются ограничения и преимущества методов.

Abstract: *The purpose of this paper is to describe mixed method of data analysis that let you use sole algorithm to solve the problem of descriptive statistics and to build histogram. First of all we would present a sole formulation: (1) find measures of central tendency and (2) build histogram. Then we would solve above solution to solving the same problem using our method. Data used in the problem about average pension benefit in Kyrgyzstan was acquired from the listed public sources and measures of center. Secondly, we would compare it using both the method of descriptive statistics – steam and leaf (stemplot) and the classical method of building histograms, which were proposed by professors John Tukey of Princeton University and Richard Thomas, respectively. According to the authors, the proposed mixed method of data analysis lets you simplify the difficulty and decrease the number of calculations, build the histogram that is easier to analyze.*

In this paper we used the following methods: retrospective, descriptive, comparative methods of analysis, descriptive statistics and histogram.

Ключевые слова: *средние величины, средняя пенсия в Кыргызстане, гистограмма, stemplot, смешанный метод анализа данных.*

Key words: *Measures of central tendency, average pension in Kyrgyzstan, histogram, stem-and-leaf diagram, mixed method of data analysis*

1. Теоретические основы исследования

В последнее время появилось много методов анализа данных, получивших название «смешанных», которые сочетают в себе несколько методов исследования[1-2]. Р. Берк Джонсон и Энтони Дж Онwuegbuzie в своем исследовании «Mixed Methods Research: A Research Paradigm Whose Time Has Come» дают следующее определение такого рода методам: «Смешанные методы исследования можно определить как методы исследования

в котором исследователь смешивает или комбинирует количественные и качественные техники, методы, подходы, концепции в едином исследовании...предлагая логическую и практическую альтернативу»[1]. По мнению авторов, смешанные методы, также, есть попытка узаконить использование нескольких подходов и концепций в едином исследовании для получения логической и практической альтернативы.

Авторы [2], говоря о фундаментальных свойствах смешанных методов, отмечают их следующую отличительную черту: «методы исследования должны следовать исследовательским вопросам таким образом, чтобы предлагать лучший шанс, чтобы получить полезные ответы.» Следовательно, при разработке смешанных методов исследования авторы следуют эклектическому подходу, т.е. не следуют какой-то одной системе, но выбирают и используют из всех систем то, что считается лучшим.

Описательная статистика используется для описания основных характеристик данных в исследовании. Они обеспечивают простые сводки о выборке и мерах. Вместе с простой графикой анализа, они составляют основу практически каждого количественного анализа данных.[3]

Описание смешанной задачи и алгоритм ее решения единственной группировкой были предложены кыргызским математиком профессором Американского Университета Центральной Азии Кыдыралиевым С. К. на курсе «Статистический анализ и прогнозирование», который проводился Академией TeachEx в Бишкеке для математиков преподавателей статистики в апреле 2006 года.

Автором настоящей работы, впервые делается попытка описания этого метода решения для того, чтобы единый/универсальный метод исследования для решения исследовательских задач, описанных выше, и учебных исследований стал доступен более широкой аудитории.

К примеру, метод отображения данных stemplot, ставший классическим для решения задач описательной статистики и широко применяемый в настоящее время на практике, в том числе и учебных исследованиях, с которым и будем сравнивать смешанный метод, или stem-and-leaf diagram, был изобретен профессором John Tukey из Princeton University в 1960-х годах[5]. В качестве метода построения гистограммы, который был принят за основу предлагаемого смешанного метода, рассматривается классический метод, описанный профессором Ричардом Томасом [6].

2. Средние величины. Проблема выбора средней величины.

Проблема выбора средней, которая изначально выделялась статистиками: средняя арифметическая – число дающее представление о многих числах, различных, хотя и однородных[7,8], продолжает быть актуальной и в настоящее время. Поскольку именно эта особенность средней арифметической, что она может быть характеристикой большого набора данных, только в случае, если набор данных однороден, сознательно или по «незнанию» игнорируется на практике.

Об этой проблеме выбора средней указывается и в [4], [6], [8]. В [4] на примере показано, что в случае неоднородных данных, с большим разбросом, в качестве средней величины такого набора наиболее репрезентативной, по сравнению со средней арифметической, наилучшим образом характеризующей данный набор как средняя величина будет медиана, а не средняя арифметическая. В [6] приводится сравнение трех методов получения средних – отмечаются достоинства и недостатки каждого метода. «Метод, основанный на вычислении средней арифметической, обычно считается наиболее приемлемым. Однако, несмотря на всю простоту его вычисления, зачастую этот метод наименее адекватен. Значение моды соответствует максимальному значению частот представленных в распределении и поэтому также не является полностью репрезентативным. Значение медианы, как центральное значение, выступает в роли компромиссного решения и часто считается наилучшим показателем»[6].

2.1. Средняя пенсия в Кыргызстане (2014 г.)

На сегодняшний день пенсии в Кыргызстане состоят из трех частей:

- базовой – это гарантированная часть пенсии, которую граждане получают по возрасту и с учетом страхового стажа. Ее размер на сегодня составляет 1,5 тысячи сомов;

- первой страховой – эту часть получают пенсионеры, имеющие страховой стаж до 1996 года. (Рассчитывается она по следующему принципу – среднемесячная зарплата за любые 60 месяцев трудового стажа подряд умножается на количество проработанных лет, по 1 % за год с учетом поправочных коэффициентов. При этом, по желанию оформителя пенсии, эти месяцы может выбрать как он сам, так и Соцфонд).

- второй страховой – это часть пенсии, которая начисляется с учетом суммы накопленных взносов с 1996 года.

Отметим, размер социальных отчислений на сегодня составляет 27,25 %. Из них 17,25 % выплачивает работодатель, туда входят 15 % отчислений в пенсионный фонд, 2 % - в Фонд обязательного медицинского страхования, 0,25 % - в Фонд оздоровления трудящихся. Из 10 %, выплачиваемых работником, 8 % идет в пенсионный фонд, а остальные 2 % - в накопительный фонд, созданный в 2010 году. [9]

Как действует накопительный фонд? Накопительный фонд был создан с той целью, чтобы в будущем средства, которые на него поступают, приносили трудящемуся инвестиционный доход. Что самое важное, эту сумму, в отличие от базовой и страховых частей пенсии, после смерти получателя смогут наследовать его родственники.

В качестве примера рассмотрим определение величины средней пенсии в Кыргызстане. Данные о величине пенсии и количестве пенсионеров в Кыргызстане, приведенные в работе, взяты автором на основе информации из открытых источников [9]-[13].

2.2. Пример.

В Кыргызстане 300 тыс 336 пенсионеров получают пенсию ниже прожиточного минимума, а самая высокая пенсия составляет 100 тыс сомов. Об этом (05.12.2014) на пресс-конференции в агентстве «Кабар» сообщил председатель Социального фонда Мухамметкалый Абулгазиев. По его словам, эти пенсионеры получают пенсию в размере 2 тыс.сом. «В пределах 30 тыс человек получают пенсию в размере 50-60 тыс сомов.» Также он отметил, что за 11 месяцев на выплаты средств пенсионных накоплений профинансировано 85 млн сомов 19 тыс 543 пенсионерам.[9], [10]. За последние 5 лет средний размер пенсии в Кыргызстане увеличился на 164,3% и на 2015 год составляет 4,6 тыс. сомов, - Соцфонд [11]. В Кыргызстане насчитывается 585 тыс. пенсионеров, 57 % из которых получают пенсии меньше прожиточного минимума, установленного Министерством экономики на уровне 4 тыс. 200 сомов.[12]. Большие пенсии в основном начисляются не по отчислениям, а по заслугам. Получающие пенсии более 25 тысяч сомов в месяц - в основном бывшие чиновники»[13]. Согласно открытым источникам приблизительно 62% пенсионеров получают пенсию ниже средней пенсии. У 52 % пенсионеров пенсия ниже прожиточного уровня 4100 сомов (\$71). Минимальная пенсия составляет 1500 сомов (\$26) , максимальная пенсия достигает 100000 сомов (\$1700). Средняя пенсия на 2014 год 4600 сомов(\$79). Пенсию более 25000 сомов (\$431) получают бывшие госчиновники. Пенсию 40000 сомов(\$650) и выше получают бывшие судьи и другие представители власти. На основе вышеприведенных данных, в качестве примера, рассмотрим следующий набор данных:

53	60	79	82	26	27	60	41
60	60	79	81	27	26	78	51
26	60	79	80	26	26	80	78
26	27	60	82	26	40	431	81
83	26	79	80	26	50	650	52

3. Постановка задачи.

Задача. Приводимая первичная информация содержит данные о месячной пенсии выборки из 40 пенсионеров, проживающих в городе. (Суммы указаны в долларах США):

53	60	79	82	26	27	60	41
----	----	----	----	----	----	----	----

60 60 79 81 27 26 78 51
 26 60 79 80 26 26 80 78
 26 27 60 82 26 40 431 81
 83 26 79 80 26 50 650 52

Найдите средние значения (1) и постройте гистограмму(2).

3.1. Решение задачи классическими методами: stemplot и построение гистограммы

(1) Для определения средних величин воспользуемся методом stemplotJohnTukey[5]:

Stems	Leaves
2	6 6 7 6 6 7 6 6 6 7 6 6
4	0 1
5	3 1 0 2
6	0 0 0 0 0 0
7	8 8 9 9 9 9
8	0 1 3 2 1 0 2 0
43	1
65	0

Определяем средние величины: Среднее арифметическое значение $\mu = \frac{(x_1+x_2+\dots+x_n)}{n}$

$\mu \cong 79$. Среднее арифметическое значение \$79. Медиана равна \$60. Мода равна \$26.

(2) **Построение гистограммы** используем алгоритм описанный RichardThomas [6]:

а) Находим наибольшее и наименьшее значения. В нашем примере самое большое число -\$650, самое маленькое \$26. Таким образом, определяем общий диапазон таблицы частот. б) Как правило весь диапазон разбивают на 5-10 групп из соображений целесообразности. Обычно группы имеют одинаковую протяженность, что иногда доставляет неудобство. Итак, разобьем на интервалы следующим образом. На диапазоне от \$20 до \$100. интервалы группировки данных примем по \$10. Далее по \$100. Таким образом, можем подсчитать количество пенсионеров, получающих пенсию от \$20 до \$30, от \$30 до \$40 и т.д. Сведем данные по интервалам в таблицу частот, каждому значению данных соответствует «х» в соответствующем интервале

Размер пенсии	Точки	Количество
от \$20 до \$29	x xxxxxxxxxxxx	12
от \$30 до \$39		
от \$40 до \$49	x x	2
от \$50 до \$59	x xxx	4
от \$60 до \$69	x xxxxx	6
от \$70 до \$79	x xxxxx	6
от \$80 до \$89	x xxxxxxx	8
от \$400 до \$499	x	1

от \$500 до \$599		
от \$600 до \$699	x	1

Размах: $650 - 26 = 624$. Межквартильный размах: $79 - 27 = 52$. и т.д., т.е. все, что можем рассчитать, используя stemplotJohnTukey[5].

Итак, за счет чего достигается лучший результат? Смешанный метод упрощает классический метод построения гистограммы [6] - данные сразу относятся в соответствующий интервал. Этим достигаются сразу две цели, во-первых, таблица частот не строится, поскольку данные сразу относятся в соответствующий интервал, во-вторых, значения данных не теряются и могут быть использованы для анализа описательной статистики.

Метод позволяет упростить и известный метод stem-and-leafdiagram (stemplot) [5]. В предлагаемом смешанном методе интервалы разбиения могут быть произвольные, тогда как в stemplot, у которого возможность разбиения на интервалы ограничена самим алгоритмом этого метода, т.е. привязкой к «стволам», которые кратные «10». И данные представлены в их естественном виде, что также упрощает анализ.

Литература:

[1] Johnson B.R. and Onwuegbuzie, A.J.(2004), Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educational Researcher, Vol. 33, No. 7 pp. 14-26, Published by: American Educational Research Association

[2] Onwuegbuzie, A.J. and Leech, N.L. (2006) Linking Research Questions to Mixed Methods Data Analysis, The Qualitative Report, 11 (3)ss. 474-498, <http://www.nova.edu/ssss/QR/QR11-3/onwuegbuzie.pdf>, adresinden 22.12.2014 tarihinde alındı

[3] Trochim, William M. K. (2006). "Descriptive statistics". Research Methods Knowledge Base. Retrieved 14 March 2011.

[4] Кыдыралиев С.К., Урмамбетов Б.М. (2006) Сборник заданий по современной статистике. Бишкек, АУЦА, 181.

[5] Rubenstein, R.N. vb (1992) Functions, Statistics, and Trigonometry The University of Chicago School Mathematics Project (Teachers Edition) by Foresman Scott, Illinois, 844.

[6] Thomas R.(1997) Quantitative Methods for Business Studies. Published by Financial Times/ Prentice Hall. ISBN 10: 0132311194

[7] Descriptivestatics, http://en.wikipedia.org/wiki/Descriptive_statistics (10.04.2015)

[8] Среднее значение, https://ru.wikipedia.org/wiki/Среднее_значение (05.04.2015)

[9] Возможности Соцфонда не позволяют уменьшить ставки отчислений в пенсионный фонд. (22.12.2014) <http://www.knews.kg/society/51829>

[10] Абдулгазиев М.(05.12.2014) *В Кыргызстане более 300 тысяч пенсионеров получают пенсии ниже прожиточного уровня, а самая высокая пенсия 100 тысяч сомов* <http://www.kabar.kg/economics/full/87338>, (22.12.2014)

[11] Средний размер пенсии в Кыргызстане 4,6 тыс. сомов.-Соцфонд <http://kg.akipress.org/news:612681>

[12] ОНС Соцфонда КР (2014) *Бешенные пенсии* <http://www.adc.kg/> (20.12.2014)

[13] <http://aif.kg/novosti/obschestvo/700-v-kyrgyzstane-pensii-bolee-25-tysyach-somov-v-osnovnom-poluchayut-byvshie-chinovniki.html>