

УДК 614.1:311.2

ТИПЫ ДАННЫХ, ПРОВЕРКА РАСПРЕДЕЛЕНИЯ И ОПИСАТЕЛЬНАЯ СТАТИСТИКА

© 2008 г. А. М. Гржибовский

Национальный Институт общественного здоровья, г. Осло, Норвегия

В статье рассматриваются некоторые аспекты первичной обработки данных и приемы описательной статистики, основной задачей которой является описание полученных в ходе исследования данных в максимально сжатом виде с минимальной потерей информации. Однако прежде чем приступить к описанию имеющихся материалов, следует определить их тип и распределение, так как различные типы данных описываются с помощью различных приемов описательной статистики. Кроме того, определение типа и распределения данных поможет в будущем выбрать наиболее подходящий статистический критерий для проверки гипотез. Изложенный материал дает общие сведения о первичной обработке данных и призван вызвать интерес читателей «Экологии человека» к прочтению специализированной литературы перед началом работы над будущими публикациями.

Ключевые слова: типы данных, распределение, описательная статистика.

Настоящей статьей журнал «Экология человека» открывает серию публикаций по основам биостатистики, предназначенную для начинающих исследователей. В этой серии будут рассмотрены наиболее часто встречающиеся в медицинской литературе статистические критерии, причем основное внимание будет уделяться не описанию формул или объяснению как тот или иной критерий «работает», а правилам выбора наиболее подходящего способа обработки данных в зависимости от имеющегося материала, алгоритмам применения выбранных статистических процедур с использованием пакета статистических программ SPSS и интерпретации полученных результатов. Основную информацию по использованию SPSS можно получить из недавно опубликованных пособий [4, 8]. Автор надеется, что, несмотря на то, что материал излагается простым языком и имеет прикладной характер, данная серия не будет использоваться как «пособие по нажиманию кнопок». Изложенный в серии материал призван вызвать интерес читателей «Экологии человека» к вопросам обработки данных и не заменяет специализированной литературы, тем более что недавно было выпущено несколько пособий, написанных специально для исследователей от медицины [1, 5]. Кроме того, перед началом проведения исследования, а не после, как это часто бывает, рекомендуется ознакомиться с литературой по эпидемиологии [2, 9].

Изучаемые признаки (переменные, *variables*) делятся на количественные и качественные. Количественными являются признаки, количественная мера которых четко определена, качественными — признаки, не поддающиеся числовому измерению.

Количественные признаки подразделяются на непрерывные (*continuous*) и дискретные (*discrete*). **Непрерывные признаки, например** рост, масса тела, артериальное давление, биохимические показатели крови и т. д., могут принимать любое значение на непрерывной шкале. Дискретные признаки, например количество детей в семье, число выкуриваемых сигарет в день, количество рецидивов заболевания и т. д., могут выражаться только целыми числами.

Качественные признаки, в свою очередь, делятся на номинальные (*nominal, categorical*) и порядковые или ранговые (*ordinal*). К номинальным признакам относятся характеристики, с которыми нельзя производить арифметические действия и которые нельзя расположить в порядке возрастания или убывания, например, идентификационный номер пациента, диагноз, название препарата, семейное положение и т. д. Порядковые (ранговые) признаки можно расположить (ранжировать) в логическом порядке, например, стадия болезни или оценка тяжести состояния пациента, однако невозможно количественно выразить, насколько или во сколько раз одно состояние лучше или хуже

другого. Несмотря на то, что при занесении порядковых признаков в компьютер их часто кодируют с помощью цифр, с ними, в отличие от количественных данных, нельзя производить арифметические действия. Например, оценка на экзамене в университете (по пятибалльной шкале) является типичным примером порядковой величины. Мы знаем, что оценка «отлично», традиционно выражаемая в виде «5», лучше, чем оценка «хорошо», выражаемая в виде «4», а оценка «удовлетворительно» или «3» лучше, чем оценка «неудовлетворительно» или «2», однако мы не можем сказать, что «5» лучше, чем «4», настолько же, насколько «4» лучше, чем «3», или «3» лучше, чем «2».

Переменные, которые могут быть отнесены к противоположным категориям, то есть могут принимать только одно из двух значений (здоров/болен, умер/выжил, курит/не курит и т. д.), называются дихотомическими (*dichotomous*).

Количественные данные при необходимости могут быть представлены в виде ранговых или номинальных. Например, индекс массы тела измеряется на непрерывной шкале, однако можно провести разделение выборки на лиц с недостаточной, нормальной и избыточной массой тела, создав, таким образом, порядковый признак. В дальнейшем признак можно превратить в номинальный или дихотомический, объединив первую и вторую группы.

Ранговые переменные можно представить в виде номинальных, но не наоборот. В некоторых случаях, например при применении визуально-аналоговых шкал, ранговые переменные представляют и анализируют как количественные, однако в таких случаях следует с большой осторожностью относиться к интерпретации результатов, так как различия между значениями на одном конце шкалы (например, между 1 и 2) могут быть более выражены, чем на другом (например, между 9 и 10) несмотря на то, что числовое значение различий в обоих случаях равно единице.

Перед тем как описывать количественные данные, всегда следует проводить проверку распределения. Под видом распределения понимают функцию, связывающую значения переменной случайной величины с вероятностью их появления в совокупности [6]. В биомедицинских исследованиях чаще всего проводится «проверка распределения на нормальность». Под нормальным распределением понимают симметричное распределение колоколообразной формы, при котором около 68 % данных отличается от среднего арифметического не более чем на одно, а примерно 95 % — не более чем на два стандартных отклонения в каждую сторону. Несмотря на то, что нормальное (Гауссово) распределение встречается очень часто и играет важную роль в статистике, существуют и другие распределения данных (биномиальное, Пуассона, Максвелла, Шарлье, и др.), о которых можно прочитать в специальной литературе.

Проверка распределения производится тремя способами: с помощью описательной статистики, графики и с использованием статистических критериев.

Все способы проверки рассматриваются на меренно измененном материале, полученном в ходе Северодвинского когортного исследования [11]. Файл с данными (*Human_Ecology_1_2008.sav*) доступен на сайте журнала: http://www.nsmu.ru/nauka_sgmu/rio/eco_human/. Проверим распределение семейного дохода (переменная *dohod*) в семьях первородящих женщин г. Северодвинска Архангельской области и массы тела их новорожденных детей (переменная *ves*). В файле содержатся данные только по детям, рожденным в срок 37–42 недели.

Для получения описательной статистики в SPSS следует в меню «**Analyze**» выбрать «**Descriptive statistics**», а затем «**Explore**». В открывшемся диалоговом окне слева будет список переменных, из них следует выбрать те, для которых планируется провести проверку распределения (*vozzrast*, *ves*). Выделите их с помощью левой кнопки мыши и переместите в окно **Dependent list** путем нажатия на кнопку «**►**» в диалоговом окне. Нажатие кнопки «**Statistics**» позволяет открыть новое диалоговое окно, в котором помимо уже отмеченной по умолчанию описательной статистики можно выбрать процентиля (**Percentiles**) и выскакивающие величины (**Outliers**). В этом же окне можно изменить установленный по умолчанию 95 % доверительный интервал для средней арифметической. Закрытие окна «**Statistics**» осуществляется нажатием кнопки «**Continue**» в том же окне. Для запроса графической информации и формальных тестов для проверки распределения следует после возвращения в диалоговое окно «**Explore**» открыть окно «**Plots**». Для проверки распределения графически выберите только гистограмму (**Histogram**) и графики и тесты для проверки распределения (**Normal probability plots with tests**). Возвращение в диалоговое окно «**Explore**» осуществляется нажатием на кнопку «**Continue**». Убедитесь, что в нижнем левом углу окна под заголовком «**Display**» отмечено «**Both**», что позволит вывести результаты всех запрашиваемых процедур одновременно. Запуск анализа производится нажатием на «**ОК**».

Полученные результаты описательной статистики для переменных «*dohod*» и «*ves*» представлены в том виде, как их выдает SPSS, в табл. 1 и 2.

В таблицах представлены значения средней арифметической (**Mean**) и ее стандартной ошибки (**Std. Error**), верхняя (**Upper Bound**) и нижняя (**Lower Bound**) границы 95 % доверительного интервала для средней, медиана (**Median**), дисперсия (**Variance**), стандартное отклонение (**Std. Deviation**), минимальное (**Minimum**) и максимальное (**Maximum**) значения переменной, размах (**Range**), межквартильный интервал (**Interquartile Range**), а также коэффициенты асимметрии (**Skewness**) и эксцесса (**Kurtosis**).

Таблица 1

Описательная статистика для переменной «dohod»
Descriptives

			Statistic	Std. Error
dohod	Mean		11966,63	242,932
	95% Confidence Interval for Mean	Lower Bound Upper Bound	11489,83 12443,43	
	5% Trimmed Mean		11219,81	
	Median		10200,00	
	Variance		5E+007	
	Std. Deviation		7161,336	
	Minimum		1900	
	Maximum		59000	
	Range		57100	
	Interquartile Range		6800	
	Skewness		2,196	,083
	Kurtosis		7,480	,166

Таблица 2

Описательная статистика для переменной «ves»
Descriptives

			Statistic	Std. Error
ves	Mean		3388,20	14,784
	95% Confidence Interval for Mean	Lower Bound Upper Bound	3359,18 3417,21	
	5% Trimmed Mean		3385,75	
	Median		3370,00	
	Variance		189927,0	
	Std. Deviation		435,806	
	Minimum		1900	
	Maximum		4720	
	Range		2820	
	Interquartile Range		560	
	Skewness		,101	,083
	Kurtosis		,145	,166

При нормальном распределении, которое симметрично, значения медианы и среднего арифметического будут одинаковы, а значения асимметрии и эксцесса равны нулю. Если средняя арифметическая больше медианы, а коэффициент асимметрии > 0 , то распределение имеет правостороннюю асимметрию (скошено вправо). При левосторонней асимметрии средняя арифметическая меньше медианы, а коэффициент асимметрии < 0 . По величине коэффициента эксцесса говорят об островершинном ($Kurtosis > 0$) или плосковершинном ($Kurtosis < 0$) распределении. Однако ситуаций, когда средняя арифметическая равна медиане, а коэффициенты асимметрии и эксцесса равны нулю, практически не встречается, поэтому необходимо решить, какие отклонения от идеального сценария допустимы для того, чтобы считать распределение полученных данных нормальным или близким к нормальному.

Для такой проверки существуют графические методы и статистические критерии. SPSS проводит проверку распределения с помощью критериев Shapiro-Wilk (для небольших выборок) и Kolmogorov-Smirnov с поправкой Lilliefors (для больших выборок). В различных версиях SPSS существуют различные алгоритмы по поводу того, какую выборку считать большой. Версия 14 по умолчанию предоставляет результаты критерия Shapiro-Wilk для выборок объ-

емом до 5 000 наблюдений, хотя обычно этот критерий применяется для проверки распределения в значительно меньших выборках (до 50 наблюдений). При применении вышеупомянутых критериев за нулевую гипотезу принимается гипотеза о том, что изучаемое распределение не отличается от нормального, значит, если достигнутый уровень значимости при проверке гипотезы будет меньше, чем критический уровень значимости (p , обычно 0,05), обозначаемый SPSS как Sig. (от англ. *significance*), то нулевая гипотеза о сходстве распределений отвергается, значит, распределение отличается от нормального. Соответственно если $p > 0,05$, то распределение не отличается от нормального. Результаты проверки гипотез о соответствии распределения переменных «dohod» и «ves» нормальному представлены в табл. 3 и 4.

Таблица 3

Результат проверки распределения переменной «dohod» с помощью статистических критериев
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
dohod	,140	869	,000	,825	869	,000

a. Lilliefors Significance Correction

Таблица 4

Результат проверки распределения переменной «ves» с помощью статистических критериев
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ves	,029	869	,076	,997	869	,203

a. Lilliefors Significance Correction

Достигнутый уровень значимости (Sig.) для переменной «dohod» представляет собой малую величину ($p < 0,001$) и позволяет отвергнуть нулевую гипотезу о подчинении данных закону нормального распределения. Для переменной «ves» нулевую гипотезу при критическом уровне значимости 0,05 отвергнуть нельзя, значит, можно сделать вывод о том, что масса тела новорожденных в исследуемой выборке подчиняется закону нормального распределения.

К аналогичному заключению можно прийти на основании результатов анализа графиков. Гистограммы обеих переменных представлены на рис. 1 и 2.

Гистограмма показывает, что распределение переменной «dohod» смещено вправо, что соответствует результатам описательной статистики. Непрерывная линия на рисунке показывает нормальное распределение при значениях средней арифметической и стандартного отклонения, полученных для имеющихся данных. Таким образом, гистограмма наглядно показывает, что распределение доходов в семьях не подчиняется закону нормального распределения, а значение средней арифметической больше медианы из-за более высоких доходов небольшого количества семей. Нелишне упомянуть, что в масштабах страны распределение доходов еще сильнее смещено вправо, чем на рис. 1, то есть средние значения доходов не являются реальным отражением доходов большинства населения.

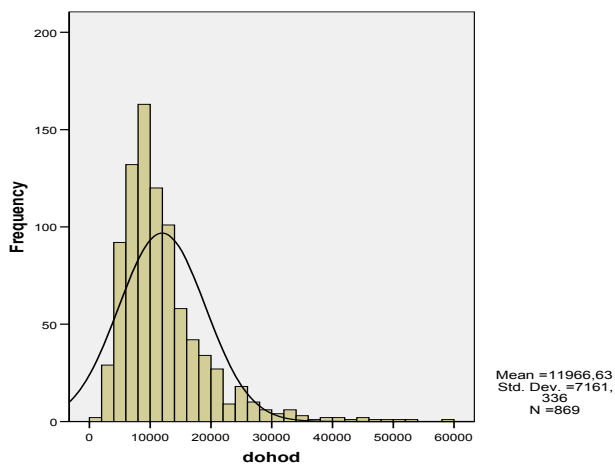


Рис. 1. Гистограмма переменной «dohod»

Гистограмма для переменной «ves» имеет симметричный вид вокруг средней величины, и большинство частот находится под кривой нормального распределения, что было ранее показано с помощью критерия Shapiro-Wilk.

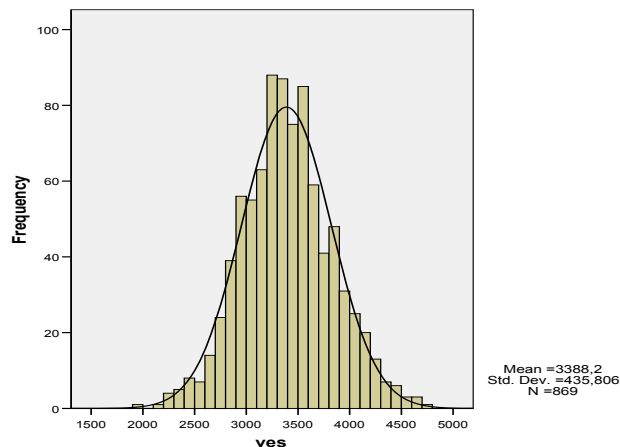


Рис. 2. Гистограмма переменной «ves»

Несмотря на то, что гистограмма является хорошим способом проверки нормальности распределения, автоматическое создание программой шкалы может привести к неверным выводам. Более четкую картину распределения данных и соответствия распределения данных закону нормального распределения дают квантильные диаграммы (Q-Q plots).

В случае нормального распределения квантильная диаграмма имеет вид прямой линии. Любое отклонение от прямой свидетельствует об отклонении данных от нормальности. Распределение переменной «dohod» значительно отличается от нормального, что подтверждается квантильной диаграммой, на которой прямой линией обозначено, как выглядело бы нормальное распределение, а фактическое распределение видимо отклоняется от этой прямой (рис. 3).

Для переменной «ves» большинство значений переменной находится на прямой линии, что говорит о близости фактического распределения нормальному (рис. 4), что было ранее показано с помощью критерия Shapiro-Wilk.

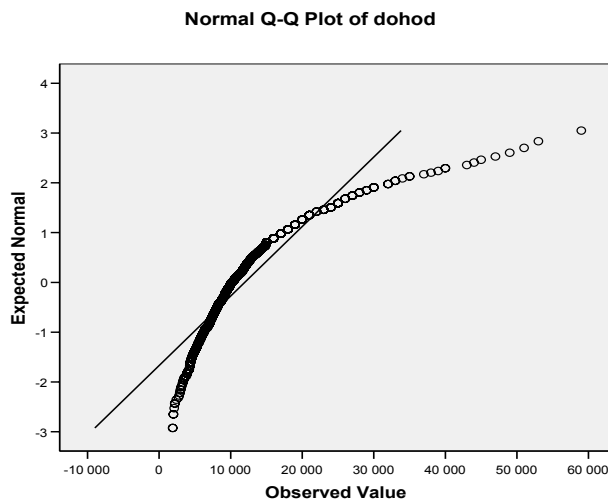


Рис. 3. Квантильная диаграмма переменной «dohod»

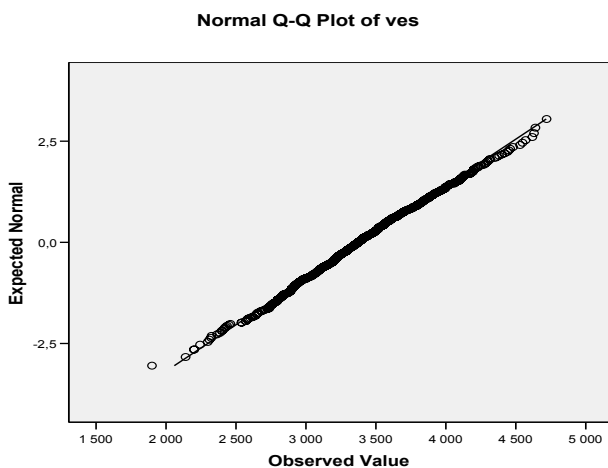


Рис. 4. Квантильная диаграмма переменной «ves»

Кроме того, исходя из определения нормального распределения, 95 % значений должны находиться в пределах от 2,5 до 97,5 перцентиля. SPSS автоматически не рассчитывает эти процентиля, но их можно задать, зайдя в меню Analyze, выбрать Descriptive Statistics, а затем Frequencies. В появившемся окне следует выбрать интересующие переменные, нажать на «Statistics» и отметить «Percentiles», после чего вручную ввести интересующие процентиля. При желании также можно выбрать квантили (Quartiles) и/или определить значения для разбития выборки на любое количество групп (Cut-off points). Кроме того, в этом окне можно выбрать все упомянутые ранее средние величины и меры рассеяния. В диалоговом окне Frequencies также можно открыть меню «Charts» и запросить построение гистограммы «Histogram» с кривой нормального распределения «With normal curve».

Итак, на основании проверки распределения можно сделать вывод, что значения переменной «dohod» не подчиняются закону нормального распределения. Такие данные не рекомендуется описывать с помощью средней арифметической и стандартного отклонения, которые сильно подвержены влиянию крайних вари-

ант. Медиана значительно меньше подвержена такому воздействию, а потому рекомендуется для описания асимметричных распределений. В качестве мер рассеяния рекомендуется использовать процентиля (25-й и 75-й, называемые также нижним и верхним квартилями, соответственно используются чаще других), а также размах вариации. Таким образом, переменную «*dohod*» можно описать следующим образом: уровень семейных доходов в выборке варьировал от 1 900 до 59 000 рублей в месяц ($Me = 10\,200$), причем доходы 50 % семей находились в промежутке от 7 500 до 14 300 рублей в месяц. Верхний и нижний квартили также представляют в виде **Q1 и Q3 соответственно**.

Распределение значений переменной «*ves*» в выборке подчинялось закону нормального распределения, а потому может описываться с помощью средней арифметической и стандартного отклонения: $M = 3\,388$ г, $SD = 436$ г. Интервальную оценку популяционной средней можно представить с помощью доверительных интервалов (ДИ). В данном примере масса тела детей, рожденных в срок 37–42 недели у первородящих женщин г. Северодвинска, составляет 3 388 (95 % ДИ: 3 359–3 417) г. Многие авторы представляют выборочные данные в виде $M \pm m$, где M – средняя арифметическая, а m – стандартная ошибка средней величины. Желание представлять среднюю ошибку средней арифметической вместо стандартного отклонения понятно, так как она в \sqrt{n} раз меньше последнего и может маскировать существенный разброс данных вокруг среднего значения, особенно при асимметричных распределениях. Поэтому рекомендуется всегда сначала проверять распределение данных и в случае нормального распределения представлять выборочные данные в виде средней арифметической и стандартного отклонения. Асимметричные распределения лучше описывать с помощью медианы, процентиля и размаха вариации. Помимо медианы можно для описания центральных тенденций использовать моду (M_0). Мода представляет собой наиболее часто встречающееся значение переменной. Помимо самой моды рекомендуется представлять, в какой доле случаев переменная принимает значение, равное моде.

Одним из способов оценки вариабельности признака является расчет коэффициента вариации (coefficient of variation, C_v), который SPSS не рассчитывает, но который легко получить путем деления стандартного отклонения на среднюю арифметическую с последующим умножением результата на 100 %. Данные считаются достаточно однородными при $C_v < 10\%$ [7], однако это разделение достаточно условно. Коэффициент вариации может применяться для сравнения разброса данных, имеющих разные размерности. Для семейного дохода $C_v = 60\%$, в то время как для массы тела новорожденных $C_v = 13\%$, то есть можно говорить о том, что разброс доходов в изучаемой выборке варьирует в несколько

раз сильнее, чем разброс массы тела новорожденных.

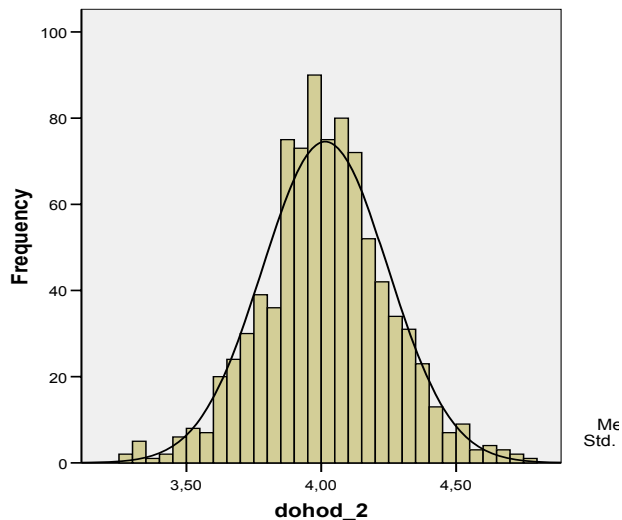
Уже говорилось о том, что тип и распределение данных определяет выбор статистических критериев для проверки гипотез. Одним из необходимых условий применения параметрических критериев является нормальное распределение. Более подробно о необходимых условиях для применения тех или иных критериев будет рассказано в последующих статьях. Однако даже если оригинальные данные не подчиняются закону нормального распределения, их можно трансформировать таким образом, что распределение приблизится к нормальному, а значит, можно будет применять параметрические критерии при условии соблюдения других условий. При правосторонней асимметрии (по мере увеличения асимметрии) чаще всего применяют следующие виды «нормализующей трансформации»: извлечение квадратного корня – \sqrt{x} , логарифмическое преобразование с использованием натурального логарифма – $\text{Ln}(x)$, логарифмическое преобразование с использованием десятичного логарифма – $\text{Log}(x)$, гармоническое преобразование – $1/x$. При применении гармонического преобразования лучше использовать $-1/x$ вместо $1/x$ для сохранения направленности. В противном случае наименьшие и наибольшие значения поменяются местами, что может затруднить интерпретацию результатов некоторых видов анализа. Если значение переменной равно нулю, то для проведения логарифмического или гармонического преобразования следует добавлять к нулю некоторое малое число, например 0,001. При левосторонней асимметрии имеет смысл преобразовывать данные путем их возведения в степень (обычно во вторую или третью). Выбор наиболее подходящего вида трансформации для имеющихся данных определяется методом проб и ошибок, а об успешности преобразования смотрят по графикам, коэффициентам асимметрии и эксцесса и результатам проверки распределения с помощью статистических критериев.

Для примера рассмотрим логарифмическое преобразование переменной «*dohod*» с использованием десятичного логарифма. Для проведения преобразования в SPSS следует в меню **Transform** выбрать вариант **Compute**. В левом верхнем углу появится пустое окно, в которое нужно ввести имя новой переменной (для нашего примера назовем ее «*dohod_2*»), затем из списка функций в правом нижнем окне следует выбрать интересующую функцию (для данного примера **Arithmetic**, затем **LG10**) и перенести ее в правое верхнее окно путем нажатия на кнопку «▶». Затем из списка переменных в столбце слева следует выбрать преобразуемую переменную и перенести ее с помощью кнопки «▶» в правое верхнее окно. Можно записать $\text{LG10}(\text{dohod})$ в правом верхнем окне вручную. Запуск преобразования осуществляется кнопкой «ОК», после чего в окне SPSS появится новая переменная, распределение которой необходимо проверить по алгоритму, рассмотренному выше.

Описательная статистика для переменной «dohod_2» представлена в табл. 5.

Таблица 5

Описательная статистика для переменной «dohod_2»
Histogram



Преобразование уменьшило показатель асимметрии с 2,196 до 0,011, а показатель эксцесса с 7,480 до 0,536. Кроме того, медиана и средняя арифметическая после преобразования стали приблизительно равны. Однако достаточно ли этого для того, чтобы считать, что данные подчиняются закону нормального распределения? Необходимо провести проверку распределения с помощью статистических критериев и графически. Согласно критерию Shapiro-Wilk распределение данных даже после преобразования отличается от нормального (табл. 6).

Таблица 6

Результат проверки распределения переменной «dohod_2» с помощью статистических критериев
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
dohod_2	,046	869	,000	,995	869	,003

a. Lilliefors Significance Correction

В то же время и гистограмма (рис. 5), и квантильная диаграмма (рис. 6) показывают, что распределение, полученное после трансформации, близко к нормальному.

В такой ситуации можно, несмотря на результат применения критерия Shapiro-Wilk, считать распределение близким к нормальному и применять параметрические методы статистики, но только если соблюдаются прочие условия.

Возведение 10 в степень, равную значению средней арифметической, полученной в результате преобразования с помощью десятичного логарифма (в нашем примере 4,0149), позволит получить величину, называемую средней геометрической. Использование средней геометрической предпочтительнее при

описании распределений с сильной правосторонней асимметрией. Если в нашем примере средняя арифметическая для семейного дохода (11 967 рублей) превышает среднюю геометрическую ($10^{4.0149} = 10\ 349$ рублей) «всего» на 14 %, то для вариантов: 100, 100, 1 000, 1 000, 10 000, 1 000 000 эти две средние будут различаться в 78 раз!

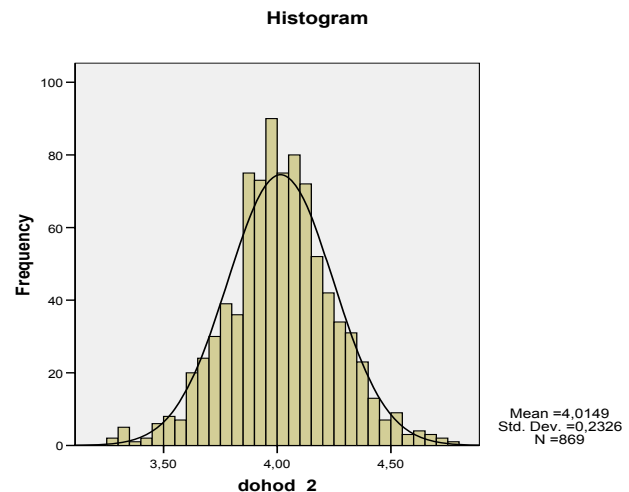


Рис. 5. Гистограмма переменной «dohod_2»

Normal Q-Q Plot of dohod_2

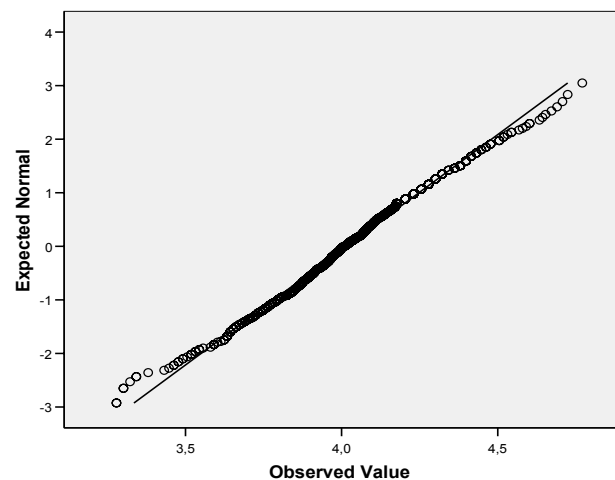


Рис. 6. Квантильная диаграмма переменной «dohod_2»

Результаты проверки распределения с помощью статистических критериев всегда следует интерпретировать с осторожностью, так как они чувствительны к объемам выборок. Вероятность получения статистически значимых различий при проверке распределения при одинаковом отклонении фактического распределения от нормального при $n = 1\ 000$ значительно выше, чем, скажем, при $n = 30$. Некоторые исследователи [10] рекомендуют всегда считать распределение отличающимся от нормального при $n < 30$. При условии $30 < n < 100$, если статистические критерии покажут отклонение распределения от нормального ($p < 0,05$), следует считать, что распределение отличается от нормального, если графики

и значения асимметрии и эксцесса не свидетельствуют об обратном. При условии $n \geq 100$, если нулевую гипотезу о соответствии распределения нормальному отклонить нельзя ($p > 0,05$), распределение считают нормальным, если графики и значения асимметрии и эксцесса не говорят о противоположном. Для условного соответствия распределения нормальному допускается нахождение показателей асимметрии и эксцесса в интервале от -1 до 1 [10], хотя встречается и более консервативный подход, согласно которому допускаются значения асимметрии и эксцесса от $-0,5$ до $0,5$ [3]. Автор придерживается мнения, что всегда следует проверять распределение несколькими способами, из которых оценка квантильной диаграммы представляется наиболее информативным.

Описание качественных данных представляет меньше проблем для исследователя. Для графического представления используются столбиковые, составные столбиковые и секторные диаграммы. Обычно достаточно представить частоты, с которыми те или иные значения качественных признаков встречаются в выборке. Предпочтительнее интервальная, чем точечная, оценка частот, то есть с указанием (обычно 95 %) доверительных интервалов. SPSS автоматически не рассчитывает доверительные интервалы для пропорций, поэтому для определения нижнего предела 95 % доверительного интервала можно использовать формулу: $p - 1,96\sqrt{p(100 - p)/n}$, а для верхнего — $p + 1,96\sqrt{p(100 - p)/n}$, где p — пропорция, а n — объем выборки. Следует помнить, что указанная формула дает адекватную оценку доверительных интервалов только при выполнении следующего условия: произведения $n \times p$ и $(n - 1) \times p$ должны иметь значение не менее 5, а по мнению более консервативных исследователей, не менее 10. В противном случае следует рассчитывать точный доверительный интервал с помощью других программ для статистической обработки данных.

В следующем выпуске будет рассматриваться сравнение данных с различными распределениями для двух независимых групп.

Список литературы

1. Банержи А. Медицинская статистика понятным языком: вводный курс / А. Банержи. — М.: Практическая медицина, 2007. — 287 с.
2. Власов В. В. Эпидемиология: учебное пособие для вузов / В. В. Власов. — М.: ГЭОТАР-МЕД, 2004. — 464 с.
3. Жижин К. С. Медицинская статистика: учебное пособие / К. С. Жижин. — Ростов н/Д: Феникс, 2007. — 160 с.
4. Наследов А. Д. SPSS: Компьютерный анализ данных в

психологии и социальных науках / А. Д. Наследов. — СПб.: Питер, 2007. — 416 с.

5. Петри А. Наглядная статистика в медицине / А. Петри, К. Сэбин. — М.: ГЭОТАР-МЕД, 2003. — 144 с.

6. Сергиенко В. И. Математическая статистика в клинических исследованиях / В. И. Сергиенко, И. Б. Бондарева. — М.: ГЭОТАР-МЕД, 2001. — 256 с.

7. Сырцова Л. Е. Основы эпидемиологии и статистического анализа в общественном здоровье и управлении здравоохранением: учебное пособие для ординаторов и аспирантов / Л. Е. Сырцова, И. И. Косаговская, М. М. Авксентьева. — М.: ММА им. И. М. Сеченова, 2003. — 91 с.

8. Таганов Д. SPSS: Статистический анализ в маркетинговых исследованиях / Д. Таганов. — СПб.: Питер, 2005. — 192 с.

9. Флетчер Р. Клиническая эпидемиология: основы доказательной медицины / Р. Флетчер, С. Флетчер, Э. Вагнер. — М.: МедиаСфера, 1998. — 345 с.

10. Chang Y. H. Biostatistics 101: Data presentation / Y. H. Chang // Singapore Medical Journal. — 2003. — N 6. — P. 280–285.

11. Grjibovski A. M. Social variations in fetal growth in Northwest Russia: an analysis of medical records / A. M. Grjibovski, L. O. Bygren, B. Svartbo, P. Magnus // Annals of Epidemiology. — 2003. — N 9. — С. 599–605.

DATA TYPES, CONTROL OF DISTRIBUTION AND DESCRIPTIVE STATISTICS

A. M. Grjibovski

National Institute of Public Health, Oslo, Norway

In the article, some aspects of data primary processing and methods of descriptive statistics, the main task of which is description of maximum compact data received with minimum losses during studies have been considered. However before beginning of description of available data, one should determine their type and distribution, as different data types are described with different methods of descriptive statistics. Besides, determination of data types and distribution will help in future to choose the most appropriate statistical criterion for control of hypotheses. The stated materials give general information about data primary processing and they should attract readers of «Human Ecology» to reading of specialized literature before beginning of work with future publications.

Key words: data types, distribution, descriptive statistics.

Контактная информация:

Гржибовский Андрей Мечиславович — старший советник Национального института общественного здоровья, г. Осло, Норвегия

Тел.: +47 22042392, +47 45268913; e-mail: angr@fhi.no

Статья поступила 23.11.2007 г.

Вниманию авторов

С января 2008 г. вводятся новые единые требования к рукописям, представляемым в научно-практический журнал «Экология человека». Цель введения указанных требований — повышение качества журнала, его цитируемости отечественными и зарубежными исследователями. Ознакомиться с новыми требованиями можно на сайте www.nsmu.ru/pauka_sgmtu/rio/.

Тематика и специализация журнала. Эколого-физиологические основы жизнедеятельности человека. Экология природных и социальных катастроф. Воспроизводство населения и демографические процессы. Социальная политика и общественное здоровье. Журнал представляет междисциплинарное издание научно-теоретической и практической ориентации, направленное на публикацию оригинальных исследований, обзоров, сообщений, отчетов о конференциях, рецензий по актуальным вопросам экологии человека. Имеет следующие рубрики: Экологическая безопасность; Медицинская экология; Ментальная экология; Экология детства; Экологическая морфология; Экологическая физиология; Эндоекология; Социальная экология; Экология образования; Экология труда; Окружающая среда; Экология Баренц-региона; Экология судьбы и др. Ориентирован на широкий круг научной общественности, практических врачей, биологов, экологов.

Подготовка материалов. Статьи присылаются в редакцию в одном экземпляре на дискете 3,5" — 1,4 Мб и двух распечатанных на бумаге экземплярах, причем авторам рекомендуется проверить соответствие текста на дискете распечатанному варианту статьи. К материалам статьи прилагается сопроводительное письмо на имя главного редактора журнала по следующей форме:

1. Рекомендация учреждения, в котором выполнялась данная работа (за подписью проректора по науке либо руководителя учреждения).
2. Фамилия, имя, отчество, ученая степень, звание и должность автора, ответственного за дальнейшую переписку с редакцией.
3. Название учреждения, где работает автор.
4. Почтовый адрес для переписки с указанием почтового индекса.
5. Телефон, факс, адрес электронной почты автора, ответственного за переписку.
6. Фамилии и инициалы всех соавторов.
7. Полное название рукописи, направляемой в редакцию.
8. Количество страниц текста, количество таблиц и рисунков.

9. Указание, для какой рубрики предназначена работа.

10. Указание, является ли работа диссертационным исследованием (кандидатским либо докторским).

11. Заверение редакции в том, что материалы, представляемые в данной статье, не были опубликованы в другом печатном издании. Если материал ранее публиковался, в новой статье должны быть ссылки на предыдущие работы.

12. Дата отправления рукописи.

13. Подписи всех авторов.

Электронный вариант должен быть подготовлен в формате Word 6.0 для Windows 3.1 или Word 7.0 для Windows 95.

Текст статьи необходимо отпечатать на одной стороне листов формата А4 через 1,5–2 интервала, с полями со всех сторон по 2,5 см и обязательной нумерацией страниц. Количество знаков на странице не должно превышать 1 800 (т. е. 30 строк по 60 знаков, включая знаки препинания и пробелы). Таблицы и рисунки (графики, фотографии), а также подписи к ним размещаются на отдельных страницах в конце статьи. Приблизительное расположение иллюстративного материала в тексте указывается на полях с правой стороны.

Структура статей. Название работы должно быть кратким (не более 120 знаков), точно отражающим содержание статьи. Названий в форме вопросительных предложений, а также названий, смысл которых можно прочесть неоднозначно, следует избегать. Следует использовать только стандартные сокращения (аббревиатуры). Не следует применять сокращения в названии статьи. Полный термин, вместо которого вводится сокращение, должен предшествовать первому применению данного сокращения в тексте. Под названием помещаются инициалы и фамилии авторов, затем указываются полное название учреждения, город.

Резюме должно представлять собой краткое, но вместе с тем максимально информативное содержание научной публикации. Его объем не должен превышать 150 слов. В резюме кратко должны быть изложены предпосылки и цели исследования (1–2 предложения), основные методы, включая тип исследования, создание выборки и основные аналитические методы (4–5 предложений), основные результаты с их цифровым выражением и уровнями статистической значимости (3–4 предложения) и основные выводы (1–2 предложения). Следует отмечать новые и важные аспекты исследования. Резюме — единственная часть статьи, которая доступна в электронном формате для широкого круга читателей, поэтому в обязанность

авторов входит обеспечение точного соответствия резюме содержанию всей работы. Резюме для оригинальных исследований должно иметь структурированный вид: цель, методы, результаты, выводы. Резюме представляется на русском и английском языках.

Под резюме помещается подзаголовок «Ключевые слова», а после него — от 3 до 10 ключевых слов, отражающих проблемы, изучаемые в ходе исследования. Ключевые слова также представляются на русском и английском языках. Для англоязычной их версии желательно использовать термины из списка медицинских предметных заголовков (MeSH, Medical Subject Headings), используемых в Index Medicus (www.pubmed.com).

В статье целесообразно соблюдать следующий порядок изложения: заглавие, авторы, учреждение, резюме, ключевые слова, введение, методика, результаты исследования, обсуждение результатов, список литературы, резюме на английском языке с ключевыми словами. На отдельных страницах представляются таблицы, рисунки и подписи к рисункам. В разделе «Методика» обязательно указываются сведения о статистической обработке экспериментального или клинического материала. Не допускаются сокращения слов, кроме принятых Комитетом стандартов. Единицы измерения даются в соответствии с Международной системой единиц СИ.

Объем рукописей. Объем рукописи обзора не должен превышать 16 страниц машинописного текста (не включая таблицы, список литературы, подписи к рисункам и резюме на английском языке). Объем рукописи оригинальной статьи не должен превышать 12 страниц машинописного текста; кратких сообщений — 5 страниц; отчетов о конференциях — 3 страниц; рецензий на книги — 3 страниц.

Иллюстрации. Графики, схемы и рисунки должны быть выполнены четко черной тушью на плотной белой бумаге или методом ксерокопирования без перегибов на изображении, в электронном виде они могут быть представлены в форматах Excel или CorelDraw. Фотографии должны быть контрастными, в электронном виде представлены в формате TIFF. Графики, схемы, фотографии должны быть представлены в расчете на печать в черно-белом виде. В подписях под рисунками должны быть сделаны объяснения значений всех кривых, букв, цифр и прочих условных обозначений. Все графы в таблицах должны иметь

заголовки. Сокращение слов в таблицах не допускается. Повторять одни и те же данные в тексте, на рисунках и в таблицах не следует.

Литература. Список литературы должен представлять полное библиографическое описание цитируемых работ в соответствии с ГОСТ 7.1–2003.

Например:

Бойко Е. Р. Некоторые закономерности метаболических перестроек у человека на Крайнем Севере / Е. Р. Бойко // Физиология человека. — 1996. — № 4. — С. 122–129.

Грацианская Л. Н. Состояние здоровья электросварщиков судостроительной промышленности / Л. Н. Грацианская, И. М. Суворов, М. Л. Хаймович и др. // Вопросы профессиональной патологии в машиностроительной промышленности. — М., 1977. — С. 4–8.

Спивак Е. М. Клинико-патогенетические варианты и основы формирования вегетативной дисфункции в раннем и дошкольном возрасте : автореф. дис. ... д-ра мед. наук / Спивак Евгений Маркович. — Н. Новгород, 1993. — 45 с.

Необходимо, чтобы цитируемые источники соответствовали списку литературы. Ссылки на литературные источники в тексте статьи, в рисунках и таблицах обозначаются арабскими цифрами в квадратных скобках [1, 2, 3 ...]. Количество источников не ограничено. Сокращения названий журналов должны соответствовать общепринятому стандарту.

Рецензирование. Статьи, поступившие в редакцию, обязательно рецензируются. Если у рецензента возникают вопросы, статья возвращается на доработку. Редакция оставляет за собой право внесения редакторских изменений в текст, не искажающих смысла статьи.

Реклама. Журнал публикует рекламу по профилю журнала в виде отдельных рекламных модулей на 3-й стороне обложки (полноцветная печать), статей, содержащих коммерческую информацию по профилю журнала, с указанием «публикуется на правах рекламы». Размещение рекламы в журнале платное.

Адрес редакции: 163000, г. Архангельск, пр. Троицкий, д. 51, Северный государственный медицинский университет, редакция журнала «Экология человека»; тел.: (8182) 20-65-63; факс: (8182) 20-61-90; e-mail: gio@nsmu.ru