ТРЕБУЕТСЯ ЛИ ДЛЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ГЕОЛОГИЧЕСКИХ ДАННЫХ ИХ ПОДЧИНЕНИЕ НОРМАЛЬНОМУ (ГАУССОВСКОМУ) РАСПРЕДЕЛЕНИЮ?

Ю. А. Ткачев

Институт геологии Коми НЦ УрО РАН, Сыктывкар

e-mail: tkachev@geo.komisc.ru

Исследовано влияние отклонения распределения случайных величин от нормального, что в геологии наблюдается очень часто, на эффективность применения параметрических критериев проверки статистических гипотез. Установлено сильное преувеличение отрицательного влияния ненормальности распределения на результаты проверки гипотез. Даны рекомендации по применению компьютерного моделирования для проверки статистических гипотез.

Ключевые слова: гауссовское, логнормальное, прямоугольное распределения, проверка статистических гипотез, компьютерное моделирование случайных совокупностей с заданным распределением.

ON THE QUESTION IF GEOLOGICAL DATA NEED TO CONFORM TO NORMAL (GAUSSIAN) DISTRIBUTION FOR STATISTICAL ANALYSIS

Yu. A. Tkachev

Institute of Geology, Komi Scientific Centre, Ural Division of the Russian Academy of Sciences

This research focuses on the effectiveness of the parametric criteria for statistical hypothesis verification when random values deviate from normal distribution, which happens often in geology. According to the research, the negative impact of this deviation on the hypothesis verification results is unreasonably exaggerated. Guidelines are given on how to use computer modelling for statistical hypotheses verification.

Keywords: Gaussian, logarithmically normal, rectangular distribution, statistical hypothesis verification, computer modelling of random sets with given distribution.

Профессиональные математики, специалисты в теории вероятностей и математической статистике знают ограничения, накладываемые на применение методов обработки данных отсутствием сведений о распределении исходных величин или неподчинением их распределения нормальному закону. Один из ведущих специалистов своего времени по применению статистических методов в геологии Д. А. Родионов [7, с. 85] писал: «*Весьма распространенное заблуждение среди геологов-практиков заключается в том, что если распределения случайных величин отличаются от нормального, то выводы, базирующиеся на сравнении средних значений, необоснованны*». Конечно, автор подразумевал, что это утверждение справедливо и для других многочисленных методов проверки статистических гипотез.

Есть и другая крайняя точка зрения — полное отрицание возможности применения статистических методов в геологии. Например, В. Л. Поляков в эпиграф своей статьи [6, с. 68] вынес утверждение: «*Теория и практика традиционной статистики казуистические и логически несостоятельные*».

Прежде чем исследовать поставленный в заголовке вопрос, необходимо пояснить природу статистических выводов. Они никогда не бывают абсолютными, а сопровождаются оценкой вероятности их истинности. Одна из задач математической статистики как раз и заключается в установлении этой вероятности.

Среди многочисленных задач математической статистики обычно выделяют две: 1) проверку статистических гипотез и 2) оценку доверительных интервалов. Покажем, к каким последствиям в этих задачах приводит ошибка, вызванная отклонением исходных данных от «нормального» (гауссовского) закона.

Проверка статистических гипотез

Именно этой процедурой в большинстве случаев заканчивается статистический анализ данных, и по результатам проверки делаются выводы или принимаются решения. Перед проверкой гипотез выбирается *уровень значимости α* , под которым понимается *ошибка первого рода* — *вероятность ошибочного отклонения нулевой*

гипотезы, когда она на самом деле верна.

Для проверки гипотез используют *параметрические критерии* (они же так называемые «статистики»), но таблицы *критических значений* таких критериев даны в руководствах для нормального (гауссовского) и редкого числа некоторых других распределений. В этом и заключается суть проблемы: насколько серьезно мы ошибемся, используя параметрический критерий, если наше распределение — не гауссовское? Казалось бы, в этом случае следует применять непараметрические критерии, свободные от влияния типа распределения исходных величин. Однако это преимущество достигается уменьшением мощности критерия, т. е. уменьшением вероятности отклонения ложной гипотезы. Мощность критерия, как известно, зависит от количества используемой информации. Так, непараметрический критерий сравнения средних двух сопряженных выборок — *критерий знаков* — использует 1 бит информации на каждую пару проб. Например, если компонента больше в первой выборке, то это фиксирует-

ся единицей (1), если во второй — нулем (0). Всего используется n бит.

Параметрический критерий t использует по 2 байта на результат анализа каждой пробы. Всего 2 байта $\times 2n = 16 \text{ бит} \times 2n = 32n \text{ бит}$, т. е. в 32 раза больше. Так как мощность зависит от количества используемой информации, ущерб от её уменьшения часто перекрывает ущерб от необходимости в самом худшем случае задавать меньшую ошибку первого рода.

Теория утверждает, что вероятность ошибки первого рода (отклонение нулевой гипотезы, когда она верна) окажется выше принятой исследователем перед проверкой. Но насколько это «опасно» в практическом отношении? Классиком математической статистики Б. Л. Ван дер Варденом [1] было доказано, что отклонение распределения от нормального закона мало влияет на уровень значимости критерия α , приспособленного для нормального распределения. Даже в худшем случае оно не превысит 2α .

Следует иметь в виду, что уровень значимости α не вычисляется, а произвольно выбирается (назначается) исследователем. Он сам решает, какой уровень значимости для исследуемого вопроса считать приемлемым. При этом принцип выбора таков: чем тяжелее последствия совершения ошибки первого рода, тем меньшую вероятность её совершения необходимо закладывать в проверку гипотезы. В научных исследованиях эта величина, как правило, не регламентируется. Исследователь, обнаруживший интересную закономерность, хочет убедить себя и своих читателей в том, что обнаруженное им — не случайность, и по своему разумению принимает, как правило, $\alpha = 0.05\text{--}0.10$ (5—10 %).

Однако в технологических разработках, в частности в фармакологии, проверка ряда гипотез, например о безопасности лекарства, регламентирована, и уровень значимости принимается равным 1×10^{-6} (1×10^{-4} %). В грандиозных проектах, где есть риск глобальной катастрофы, уровень значимости, вероятно, должен быть еще меньше, предположим 1×10^{-9} , т. е. одна миллиардная.

Доверительные интервалы

Задача об оценке доверительных интервалов может быть преобразована в задачу о проверке гипотез о концах этих интервалов, т. е. привязана к описанному выше случаю.

Допустим, проверяется гипотеза о равенстве содержаний K_2O и Na_2O в некой осадочной толще формации на уровне значимости $\alpha = 0.10$ (10 %). Гипотеза не отклонена. Практически это означает, что если бы мы продолжили серию таких же опробований, то по большинству из них получили бы тот же результат, но в некотором количестве случаев (около 10 %) гипотеза $K_2O = Na_2O$ могла быть отклонена. Эти сведения, несомненно, полезны как для обретения уверенности в научном утверждении, так и для делового человека, собирающегося, например, разрабатывать месторождение, особенно когда проверяется гипотеза о том, что содержание полезного компонента в выемочном блоке не ниже промышленного.

Однако заключение по результатам проверки гипотезы о равенстве средних с помощью параметрического критерия можно оспорить, сославшись на то, что наши первичные данные, возможно, не подчиняются нормальному закону распределения. Против этого можно выдвинуть три возражения.

Во-первых, как мы уже знаем, реальный уровень ошибки первого рода α будет хотя и выше заданного, но все же не выше (по Ван дер Вардену) 2α 0.2 (20 %). Значит, если мы непременно желаем добиться реального уровня значимости не ниже 0.1, то следует задаваться табличным значением α не менее 0.05.

Во-вторых, можно применить непараметрический критерий, не зависящий от закона распределения данных, смирившись с неконтролируемым, возможно значительным увеличением ошибки второго рода — принятии ложной нулевой гипотезы. При этом необходимо иметь в виду, что в геолого-минералогических областях существует много ситуаций, в которых невозможно подобрать непараметрический критерий. Реальная практика такова, что почти всегда используют несколько разных критериев и решения принимают по совокупности проверок.

В-третьих, надо использовать как можно больше косвенных данных, коррелирующихся с прямыми. Эффект от этого будет двойным: а) увеличится число степеней свободы, а это важный параметр при проверке гипотезы, б) распределение значений критерия приблизится к нормальному, так как большинство используемых критериев суть суммы или функции от

сумм исходных данных, а суммы и средние быстро сойдутся к нормальному закону.

Компьютерное моделирование распределений — универсальный метод проверки статистических гипотез

Все три указанных приема могут оказаться ненужными, если в корне изменить саму концепцию проверки статистических гипотез, не связывая себя заранее никакими табулированными «теоретическими» распределениями, а изучая конкретное выборочное распределение с помощью компьютерного моделирования.

Современная настольная вычислительная техника позволяет за считанные секунды моделировать огромные массивы случайных величин, подчиняющихся различным распределениям, в том числе и таким, плотности вероятности которых не имеют аналитического выражения.

Для моделирования широкого спектра правоасимметричных распределений можно воспользоваться композицией нормального и логнормального распределений. Случайная величина такого распределения равна сумме двух компонент — нормальной с заданными характеристиками μ_x и σ_x и логнормальной с параметрами μ_{lgx} и σ_{lgx} . Такие логнормальные и комбинированные распределения весьма широко распространены в геологии. Им подчиняются содержания малых элементов в горных породах и рудах. Результаты спектральных анализов также близки к ним ввиду того, что их погрешности имеют логнормальное распределение.

Таким образом, проверка гипотез компьютерным моделированием с помощью любых критериев, в том числе соответствующих по смыслу параметрическим, может производиться для выборок, случайные величины в которых подчиняются любому распределению. Но при этом должны использоваться не готовые таблицы критических значений критериев, а величины, рассчитанные для данного конкретного случая, т. е. для наблюдаемых видов распределений в каждом конкретном случае. При современной вычислительной технике сделать это нетрудно. В настоящее время нет никаких препятствий для проверки статистических гипотез о случайных величинах и их параметрах.

рах, *распределенных как угодно*. Пока разработанная нами методика проверки гипотез с применением компьютерного моделирования доступна только пользователям, владеющим программированием, и не доведена до сервисного уровня.

Анализ «обоснований»
«невозможности»
применения математической
статистики
для обработки негауссовских
распределений

Рассмотрим типичный пример наскоков на теорию вероятностей и математическую статистику вообще, на её применение в геологии, в частности на обильные поучения «господам-товарищам геологам», преподаваемые оскорбительным тоном (см. статью В. Л. Полякова [6]). В одном из нравоучительных пассажей он пишет, что все процедуры математической статистики якобы разработаны только для гауссовских распределений, а они в геологии практически не встречаются, так как согласно центральной предельной теореме для этого требуется 40000 (сорок тысяч) образцов.

Для обоснования приводится ссылка на расчеты О. С. Ивашева-Мусатова: «... согласно предельной центральной теореме, для обеспечения 0.01%-ной точности измерения требуется 40 000 дат... О чем, уважаемые товарищи-господа геологи, необходимо знать точно» [6, с. 71]. Ссылка в цитате дана на учебное пособие О. С. Ивашева-Мусатова «Теория вероятностей и математическая статистика» (М.: Наука, 1979, 256 с.).

Обращение к численным примерам обязывает к строгой формулировке задачи. Увы, в работе В. Л. Полякова [6] даже не упоминается, что означает: «... для обеспечения 0.01 %-ной точности измерения». Задана ошибка измерения (0.01 %), но не указано, ни какая это ошибка, ни какова процедура измерения, ни каков исходный материал. Понятие «0.01 %-ная точность» весьма неоднозначное. Неясно, имеется ли в виду относительная погрешность, составляющая 0.01 % от измеряемой величины, или абсолютная, например 0.01 % химического элемента в породе. Не указано даже, что 0.01 % — это предельная, средняя арифметическая или средняя квадратическая ошибка.

Попытаемся установить, какую «точность» каких «измерений» мог

иметь в виду В. Л. Поляков. Предположим, что это погрешность среднего значения, выведенного из 40000 анализов некоторой совокупности данных. В таком случае необходимо было снабдить свое утверждение разъяснением процедуры опробования, ибо *эта погрешность определяется из погрешности единичного анализа (аналитической погрешности) и разброса значений оцениваемого параметра в анализируемой выборке* (т. е. природного, геохимического среднего квадратического отклонения). Без указания этих параметров задача не определена и не имеет решения.

В геологической практике выборка формируется совокупностью образцов, каждый из которых анализируется несколько раз. В более сложных случаях выборка имеет многоуровневую иерархическую структуру. Как её составить, чтобы получить максимальную точность оценки при ограниченных ресурсах, — эти вопросы составляют особую область математической статистики — *выборочный метод* [4, 5, 8, 9]. И здесь для правильной постановки задачи требуется учесть ряд параметров, характеризующих как примененные аналитические методы, так и исходный материал. Например, пусть предварительно установлено, что измеряемое свойство в объекте по пробам предполагаемого веса и формы (форма и вес пробы тоже имеют значение!) колеблется со средним квадратическим отклонением σ_a ; среднеквадратическая погрешность единичного измерения равна σ_a . Только после этого можно поставить вопрос, каково должно быть число образцов из объекта m и число повторностей анализов каждого образца n , чтобы средняя квадратическая погрешность оценки свойства в объекте не превышала α .

Ответ заключается в решении неравенства

$$\sqrt{\frac{\sigma_a^2}{m \cdot n} + \frac{\sigma_a^2}{m}} < \alpha.$$

Примем для определенности, что геохимическое среднеквадратическое отклонение в десять раз больше аналитической погрешности, что весьма реально: $\sigma_a = 10 \sigma_a$. Тогда

$$\sqrt{\frac{\sigma_a^2}{m \cdot n} + \frac{100\sigma_a^2}{m}} < \alpha \rightarrow \sigma_a \sqrt{\frac{1}{m \cdot n} + \frac{100n}{m \cdot n}} <$$

$$< \alpha \rightarrow \frac{1}{m \cdot n} + \frac{100}{m} < \frac{\alpha^2}{\sigma_a^2}.$$

Пусть $m = 10$, $n = 10$ (десять образцов, проанализированных по 10 раз каждый), тогда

$$\frac{1}{100} + \frac{100}{10} < \frac{\alpha^2}{\sigma_a^2} \rightarrow \frac{1+1000}{100} < \frac{\alpha^2}{\sigma_a^2} \rightarrow \approx$$

$$\approx 10 < \frac{\alpha^2}{\sigma_a^2} \rightarrow \sqrt{10} < \frac{\alpha}{\sigma_a}.$$

Вывод: если допустимая ошибка оценки в $\sqrt{10}$ (т. е. в 3.16) раз больше аналитической погрешности, а геохимическая (природная) изменчивость в 10 раз больше аналитической погрешности, то наших десяти образцов, проанализированных по 10 раз, вполне достаточно. Значения реальные, но сильно отличаются от пугающих «40 000 дат», которые «господам-товарищам геологам *надо знать точно*».

Соответствует ли поставленный В. Л. Поляковым вопрос существу расчетов, на которые он ссылается? Обращаясь к О. С. Ивашеву-Мусатову. После долгих поисков обнаруживаем, что такого автора среди специалистов математической статистики нет. Наконец устанавливаем, что существует Ивашев-Мусатов [3], и с трудом (В. Л. Поляков не утруждает себя указанием цитированных страниц) находим нужное место. В нём речь идет о том, сколько необходимо независимых случайных слагаемых с произвольными распределениями, чтобы их сумма следовала гауссовскому распределению. Нет ни слова ни о каких-либо измерениях, ни об обеспечении какой-либо точности этих измерений. Насколько цифры, которыми оперирует В. Л. Поляков, соответствуют тем, которые использует в своем примере О. С. Ивашев-Мусатов? Здесь характерная для В. Л. Полякова небрежность сыграла с ним злую шутку. У О. С. Ивашева-Мусатова в примере было задано $\alpha = 0.01$, а в цитированном В. Л. Поляковым утверждении она превратилась в 0.01 %, т. е. в сто раз меньше!

Степень близости к гауссовскому распределению О. С. Ивашев-Мусатов характеризует средним квадратическим отклонением слагаемых суммы a от математического ожидания идеального нормального распределения и вероятностью этого отклонения в долях сигмы s . В примере № 2 на стр. 153 автор решает эту задачу для $a = 0.01$ и $P = 0.96$. Расчет числа необходимых слагаемых n у О. С. Ивашева-Мусатова ведется по формуле

$$\Phi\left(\frac{a}{\sigma} \sqrt{n}\right) > 0.48,$$

Таблица 1

Результаты обработки моделированных данных

Границы интервалов в долях сигмы	Теоретическая вероятность попадания в t -сигмовый интервал	Доля оценок среднего из n в выборках из гауссовской (Γ) и прямоугольно распределенной совокупности (Π)									
		$n = 2$		5		10		50		100	
		Γ	Π	Γ	Π	Γ	Π	Γ	Π	Γ	Π
± 0.50	0.3829	0.2602	0.3210	0.3204	0.3432	0.3608	0.3638	0.3742	0.3686	0.3766	0.3820
± 1.00	0.6827	0.4644	0.5156	0.5756	0.6158	0.6296	0.6372	0.6738	0.6588	0.6688	0.6810
± 1.50	0.8664	0.6214	0.6420	0.7498	0.7794	0.8090	0.8044	0.8542	0.8524	0.8618	0.8608
± 2.00	0.9545	0.7386	0.7390	0.8562	0.8638	0.9082	0.9056	0.9458	0.9414	0.9510	0.9488
± 2.50	0.9876	0.8090	0.8078	0.9148	0.9154	0.9612	0.9534	0.9846	0.9836	0.9864	0.9840
± 3.00	0.9973	0.8624	0.8566	0.9506	0.9468	0.9838	0.9798	0.9958	0.9950	0.9962	0.9956
± 3.50	0.9995	0.9026	0.8948	0.9730	0.9664	0.9922	0.9908	0.9990	0.9982	0.9992	0.9988
± 4.00	0.9999	0.9368	0.9296	0.9850	0.9796	0.9966	0.9952	0.9998	0.9994	0.9998	0.9998
± 4.50	0.9999	0.9638	0.9558	0.9916	0.9894	0.9986	0.9980	1.0000	1.0000	1.0000	1.0000
± 5.00	0.9999..	0.9832	0.9822	0.9960	0.9970	0.9994	0.9998	1.0000	1.0000	1.0000	1.0000
± 5.50	0.99999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

где $\Phi(x) = \int_0^x \varphi(t) dt$ при, повторим,

$\sigma = 1$, $a = 0.01$, т. е. при $a\sqrt{n} = 2.06$, что дает $n = 206^2 \approx 40\,000$. Подставив вместо ивашев-мусатовской погрешности 0.01 в сто раз меньшую поляковскую (0.01%, что должно дать $n = 20600^2 \approx 400\,000\,000 = 4 \cdot 10^8$ образцов), обнаруживаем, что В. Л. Поляков, поучая «товарищей-господ геологов» знать «точно» число дат, должен был написать 400 000 000!

В. Л. Поляков забывает, что в цитированном им примере [3] используются в качестве исходных данных две величины: $P = 0.96$ — доверительная вероятность появления погрешности в 0.01 и $\sigma = 1$ — среднее квадратическое отклонение, полученное после суммирования 40 000 дат. Из закона сложения дисперсий суммы независимых слагаемых в лучшем случае сле-

дует, что $\sigma_0 = \frac{1}{\sqrt{40000}} = 1/200$. Эти

две величины в схему В. Л. Полякова никак не встраиваются.

О. С. Ивашев-Мусатов пишет: «...даже для не очень большой точности (два знака) и не очень большой надежности (всего 96 %) нужно уже очень много слагаемых [40 000]. Правда, надо иметь в виду, что связь, устанавливаемая неравенством [Чебышева], очень грубая, а получающиеся из нее оценки [числа образцов] сильно завышены», (курсив мой. — Ю. Т.) [3, с. 150]. Жаль, что В. Л. Поляков не дочитал до этого места из учебного пособия О. С. Ивашева-Мусатова, иначе он снизил бы проповеднический градус и не настаивал бы на том, что надо не просто знать эти 40000, а «знать точно», и, может быть, понял, что ошибся в $(100)^2 = 10000$ раз!

Для иллюстрации абсурдности рекомендаций из этой статьи [6] и показа действительных отклонений результатов статистической обработки негауссовски распределенных данных от гауссовских мы провели компью-

терное моделирование выборок из двух распределений — гауссовского и сильно отличающегося от него прямоугольного. Исследованные выборки имели численность n ($n = 2, 3, 5, 10, 50, 100$). Выборка каждой численности моделировалась 5000 раз, так что статистики в выборках ($\bar{x}$, S_x^2 , $S_{\bar{x}}^2$ и др.) можно было считать хорошими оценками их в генеральных совокупностях.

По модельным выборкам непосредственным подсчетом определялась доля случаев оценок (из 5000), заключенных в интервалах $\pm t\sigma$ ($t=0.5...5.5$), и сравнивалась с теоретической вероятностью в гауссовском распределении (табл. 1).

Результаты моделирования, представленные в табл. 1 и 2, показали, что отклонения долей оценок средних по выборкам в t -сигмовых интервалах от теоретической вероятности для нормального распределения существенны только при объемах выборок в 2, 3 и 5

Таблица 2

Отклонения долей в t -сигмовых интервалах в случайных выборках объема n относительно теоретических вероятностей гауссовского распределения

t -сигмовые интервалы	Теоретические значения для гауссов. распредел.	Отклонения долей в t -сигмовых интервалах в модельных совокупностях с гауссовским (Γ) и прямоугольным (Π) распределением											
		$n = 2$		3		5		10		50		100	
		Γ	Π	Γ	Π	Γ	Π	Γ	Π	Γ	Π	Γ	Π
± 0.50	0.3829	0.3206	0.1618	0.2475	0.1075	0.1634	0.1039	0.0579	0.0501	0.0229	0.0376	0.0167	0.0026
± 1.00	0.6827	0.3196	0.2446	0.2420	0.1599	0.1567	0.0978	0.0776	0.0665	0.0129	0.0348	0.0202	0.0023
± 1.50	0.8664	0.2827	0.2590	0.2073	0.1768	0.1346	0.1004	0.0662	0.0715	0.0141	0.0161	0.0053	0.0064
± 2.00	0.9545	0.2261	0.2257	0.1695	0.1609	0.1028	0.0949	0.0484	0.0511	0.0090	0.0136	0.0035	0.0058
± 2.50	0.9876	0.1806	0.1818	0.1304	0.1322	0.0735	0.0729	0.0265	0.0344	0.0028	0.0038	0.0010	0.0034
± 3.00	0.9973	0.1352	0.1409	0.0924	0.0932	0.0467	0.0505	0.0134	0.0174	0.0014	0.0022	0.0010	0.0016
± 3.50	0.9995	0.0968	0.1046	0.0642	0.0664	0.0264	0.0330	0.0072	0.0086	0.0004	0.0012	0.0002	0.0006
± 4.00	0.9999	0.0630	0.0702	0.0406	0.0446	0.0148	0.0202	0.0032	0.0046	0.0000	0.0004	0.0000	0.0000
± 4.50	0.9999	0.0361	0.0441	0.0251	0.0285	0.0083	0.0105	0.0013	0.0019	0.0000	0.0000	0.0000	0.0000
± 5.00	0.9999	0.0167	0.0177	0.0113	0.0129	0.0039	0.0029	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000
± 5.50	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Среднее из всех t -сигмовых интервалов		0.1525	0.1318	0.1118	0.0893	0.0664	0.0533	0.0274	0.0278	0.0057	0.0099	0.0043	0.0021

Отклонения частот от гауссовских вероятностей в t -сигмовых интервалах при нормальном (кружки) и прямоугольном (треугольники) распределении в зависимости от объема выборки. Данные из табл. 2

единиц (см. табл. 1). При выборке в 10 единиц этими отклонениями практически можно пренебречь. Только полу- и 1-сигмовые интервалы в модельных выборках оказываются уже теоретических гауссовских на 6 %.

В выборках численностью 50 и 100 единиц отклонения наблюдаются только с третьего знака после десятичной точки. Например, теоретическая вероятность для широко употребляющегося трехсигмового интервала составляет 0.9973, т. е. за пределами этого интервала оказываются три случая из тысячи, а для прямоугольного распределения в выборке 100 единиц — 0.9956, т. е. четыре случая из тысячи. Может быть, в медицине и фармакологии такое отличие существенно, но не в геологии.

Любопытно, что ширина t -сигмовых интервалов в «эмпирических» (модельных) выборках из гауссовских распределений отличается от теоретической для того же гауссовского распределения даже меньше, чем в выборках многих других распределений, в частности прямоугольного.

В табл. 2 и на рисунке приведены результаты расчетов отклонений частот средних значений в выборках от 2 до 100 единиц в t -сигмовых интервалах от вероятностей для прямоугольного и гауссовского распределений. Из этих материалов видно, что отклонения эмпирических частот от теоретических при гауссовском

распределении существенны только для выборок объемом 2 и 3.

Литература

1. Ван дер Варден Б. Л. Математическая статистика. М.: ИЛ, 1960. 434 с.
2. Зайцев Г. Н. Математическая статистика в экспериментальной ботанике. М.: Наука, 1984. 424 с.
3. Ивашев-Мусатов О. С. Теория вероятностей и математическая статистика. М.: Наука, Главная редакция физико-математической литературы. 1979, 256 с.
4. Кокрен У. Методы выборочного исследования. М.: Статистика, 1976. 440 с.
5. Михок Г., Урсяну В. Выборочный метод и статистическое оценивание. М.: Финансы и статистика, 1982. 245 с.
6. Поляков В. Л. О статистике и условиях её использования в геологии // Уральский геологический журнал, 2010. № 5 (77). С. 68—76.
7. Родионов Д. А. Статистические решения в геологии. М.: Недра, 1981. 231 с.
8. Ткачев Ю. А., Кетрис М. П. Методы оптимизации структуры опробования при геолого-геохимических исследованиях и разведке месторождений (на примере залежей нефти). Сыктывкар, 1978. 51 с. (Коми филиал АН СССР. Вып. 14 / Серия препр. сообщ. «Научные рекомендации — народному хозяйству»).
9. Шварц Г. Выборочный метод: Руководство по применению статистических методов оценивания. М.: Статистика, 1978. 213 с.

Рецензент д. г.-м. н. Я. Э. Юдович

Институт геологии
Коми научного центра
УрО РАН
сообщает, что

ГОДИЧНАЯ СЕССИЯ

состоится
19 февраля 2013 г.

в конференц-зале
Института геологии
(ул. Первомайская, 54, каб.
520).

Начало в 10:00

Доклады:

Институт геологии в 2012 г.
академик А. М. Асхабов

Биоразнообразие
и этапность в развитии
палеозойских кораллов Rugosa
д. г.-м. н. В. С. Цыганко

Основные этапы раннедокем-
брийской эволюции
приуральской части
Восточно-Европейского
кратона: метаморфизм
и палеогеодинамические
следствия
д. г.-м. н. А. М. Пыстин

Морфология алмазов
уральского типа
д. г.-м. н. В. И. Ракин

Минералого-технологические
особенности жильного кварца
Урала, Прибайкалья
и Карелии
д. г.-м. н. С. К. Кузнецов,
к. г.-м. н. В. П. Лютоев,
к. г.-м. н. Н. В. Сокерина,
к. г.-м. н. Ю. В. Глухов,
к. г.-м. н. С. Н. Шанина,
П. П. Юхтанов