

Проверка гипотезы нормальности числа падения овсяной муки по малым выборкам

Наталья А. Шмалько	¹	kafedra-tith@yandex.ru	 0000-0002-8939-0265
Игорь А. Никитин	²	nikito.igor@gmail.com	 0000-0002-8988-5911
Дарья А. Велина	²	d.velina@mgutm.ru	 0000-0002-2398-3375
Лилия Ф. Пономарева	³	eponomareva@mgutm.ru	 0000-0002-6645-6356
Сергей Е. Терентьев	⁴	seterentiev@sgsha.ru	 0000-0002-8546-7253

1 Кубанский государственный технологический университет, ул. Московская, 2, г. Краснодар, 350072, Россия

2 Московский государственный университет технологий и управления им. К.Г. Разумовского, ул. Земляной Вал, 73, г. Москва, 109004, Россия

3 Башкирский институт технологий и управления, ул. Смоленская, 34, г. Мелеуз, 453856, Россия

4 Смоленская государственная сельскохозяйственная академия, ул. Большая Советская, 10/2, г. Смоленск, 214000, Россия

Аннотация. Проверка гипотезы нормальности малых выборок требуется для установления принадлежности получаемого эмпирического распределения теоретическому распределению. Условием проверки гипотезы нормальности по совокупности малых независимых выборок является наличие их достаточного числа при одинаковом объеме. В таком случае возможным представляется проверить гипотезу нормальности генеральных совокупностей, из которых были взяты изучаемые выборки при допущении, что параметры этих совокупностей имеют разные значения. При проверке гипотезы нормальности по большому числу малых выборок из каждой выборки случайным образом отбирается по одному значению только первых, вторых и т.д. измерений, допуская, тем самым, упрощение и случайный отбор данных. Объектом данного исследования являются малые выборки числа падения овсяной муки, используемой в хлебопекарном производстве при выработке хлебопекарной продукции. Целью данной работы явилась проверка гипотезы нормальности по малым выборкам эксперимента с использованием непараметрического критерия согласия ω^2 Мизерса каждого из четырех определений числа падения, поскольку округление значений прямых измерений исключает случайный характер величины или ее нормальное распределение в пользу равномерного. Установлено, что при уровне значимости $p = 0,05$ табличное значение $(\omega^2)_{1-p}$ больше вычисленного значения ω^2 для всех четырех определений, отсюда гипотеза нормального распределения малых выборок для всех четырех определений (как случайных величин) числа падения овсяной муки не отклоняется. Полученные в работе результаты согласуются с общепринятыми классическими представлениями о проверке статистической гипотезы о нормальном распределении выборок. Статистический метод обеспечивает достаточную точность изучаемого показателя в технических системах и не требует проведения синтеза статистического критерия для проверки гипотезы нормальности малых выборок.

Ключевые слова: проверка гипотезы, число падения, овсяная мука, малая выборка, непараметрическая статистика.

Testing the hypothesis of the normality of the falling number of oat flour in small samples

Natalya A. Shmal'ko	¹	kafedra-tith@yandex.ru	 0000-0002-8939-0265
Igor A. Nikitin	²	nikito.igor@gmail.com	 0000-0002-8988-5911
Daria A. Velina	²	d.velina@mgutm.ru	 0000-0002-2398-3375
Lilya F. Ponomareva	³	eponomareva@mgutm.ru	 0000-0002-6645-6356
Sergey E. Terentev	⁴	seterentiev@sgsha.ru	 0000-0002-8546-7253

1 Kuban State Technological University, Moskovskaya St., 2, Krasnod, 350072, Russia

2 K.G. Razumovsky Moscow State University of Technologies and Management, Zemlyanoy Val, 73, Moscow, 109004, Russia

3 K.G. Razumovsky Moscow State University of Technologies and Management, Bashkir Institute of Technology and Management, Smolenskaya st., 34, Meleuz, 453856, Russia

4 Smolensk State Agricultural Academy, Bol'shaya Sovetskaya str., 10/2, Smolensk, 214000, Russia

Abstract. Verification of the hypothesis of the normality of small samples is required to establish whether the empirical distribution obtained belongs to the theoretical distribution. The condition for testing the hypothesis of normality for a set of small independent samples is the presence of a sufficient number of them with the same volume. In this case, it is possible to test the hypothesis of the normality of the general aggregates from which the studied samples were taken, assuming that the parameters of these aggregates have different values. When testing the hypothesis of normality for a large number of small samples, only one value of the first, second, etc. measurements is randomly selected from each sample, thereby allowing simplification and random selection of data. The object of this study is small samples of the falling number of oat flour used in bakery production in the development of bakery products. The purpose of this work was to test the hypothesis of normality for small samples of the experiment using the nonparametric criterion of agreement ω^2 of the smallest of each of the four definitions of the incidence number, since rounding the values of direct measurements excludes the random nature of the quantity or its normal distribution in favor of a uniform one. It was found that at a significance level of $p = 0,05$, the table value $(\omega^2)_{1-p}$ is greater than the calculated value of ω^2 for all four definitions, hence the hypothesis of the normal distribution of small samples for all four definitions (as random variables) of the falling number of oat flour does not deviate. The results obtained in this work are consistent with the generally accepted classical concepts of testing the statistical hypothesis of the normal distribution of samples. The statistical method provides sufficient accuracy of the studied indicator in technical systems and does not require the synthesis of a statistical criterion to test the hypothesis of the normality of small sample.

Keywords: hypothesis testing, falling number, oat flour, small sample, nonparametric statistics.

Для цитирования

Шмалько Н.А., Никитин И.А., Велина Д.А., Пономарева Л.Ф., Терентьев С.Е. Проверка гипотезы нормальности числа падения овсяной муки по малым выборкам // Вестник ВГУИТ. 2023. Т. 85. № 1. С. 118–126. doi:10.20914/2310-1202-2023-1-118-126

For citation

Shmal'ko N.A., Nikitin I.A., Velina D.A., Ponomareva L.F., Terentev S.E. Testing the hypothesis of the normality of the falling number of oat flour in small samples. Vestnik VGUIT [Proceedings of VSUET]. 2023. vol. 85. no. 1. pp. 118–126. (in Russian). doi:10.20914/2310-1202-2023-1-118-126

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License

Введение

При проведении прямых измерений требуется статистическая обработка их результатов с установлением принадлежности получаемого эмпирического распределения теоретическому распределению. Условием проверки гипотезы нормальности по совокупности малых независимых выборок является наличие их достаточно большого числа при одном и том же объеме. В таком случае можно проверить «...гипотезу нормальности генеральных совокупностей, из которых были взяты изучаемые выборки при допущении, что параметры этих совокупностей имеют разные значения» [1]. К изучению принимается относительное отклонение, равное

$$\tau = \frac{x_{ik} - \bar{x}_k}{S_k} \quad (1)$$

где x_{ik} – i -й элемент k -й выборки; $\bar{x}_k$, S_k – среднее и среднеквадратичное отклонение k -й выборки.

Следует учитывать, что распределение величины τ не зависит от параметров генеральной совокупности: количества выборок и величины стандартного отклонения, а зависит только от объема выборки. Плотность вероятности величины τ выражается следующим образом:

$$f(\tau) = \begin{cases} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{f+1}} \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)} \left(1 - \frac{\tau^2}{f+1}\right)^{\frac{f-2}{2}} & \text{при } |\tau| < \sqrt{f+1}; \\ 0 & \text{при } |\tau| \geq \sqrt{f+1}, \end{cases} \quad (2)$$

при $|\tau| < \sqrt{f+1}$;
0 при $|\tau| \geq \sqrt{f+1}$,

где f – число степеней свободы, определяемое как уменьшенный на два объем выборки (число измеряемых определений). При прогнозировании расхождения эмпирического распределения с теоретическим полагают, что уже «...при четырех повторностях измерений в малой выборке относительные отклонения в отдельных выборках в изучаемой совокупности могут подчиняться равномерному распределению, если исходные совокупности нормальны» [1]. Если число выборок достаточно велико, то данное условие позволит выполнить проверку гипотезы нормальности. Обычно при проверке гипотезы нормальности по большому числу малых выборок из каждой выборки случайным образом отбирается по одному значению только первых, вторых и т. д. измерений, допуская, тем самым, упрощение и случайный отбор данных. Например, при объеме выборки в четыре измерения плотность вероятности величины τ приобретает вид:

$$f(\tau) = \frac{1}{2\sqrt{3}} \text{ при } |\tau| < \sqrt{3}. \quad (3)$$

В рекомендациях [2] изложены правила проверки опытного распределения с теоретическим при проверке простых и сложных гипотез путем применения критериев согласия Колмогорова, Смирнова и ω^2 Мизерса. Вначале определяются со сложностью проверяемой гипотезы, а в самом процессе проверки согласия по выборке вычисляют значение статистики используемого критерия. Считается, что «...распределения статистик непараметрических критериев как при простых, так и сложных гипотезах очень быстро сходятся к предельным законам», не допуская ошибок при анализе данных с числом малых выборок более $15 \div 20$. Малый объем выборок также позволяет «...успешно различать близкие гипотезы», что представляется удобным при разработке выборочного плана.

Объектом данного исследования являются малые выборки числа падения овсяной муки, используемой в хлебопекарном производстве при выработке хлебопекарной продукции повышенной пищевой ценности [3], в том числе при составлении композитных смесей с различными видами муки [4, 5]. Как показывают исследования, «...овсяная мука, как и другие виды муки из крупяных культур, отличается от традиционной хлебопекарной муки специфическими технологическими свойствами, в том числе отсутствием способности образовывать клейковину. Автолитическая активность подобного сырья в силу специальной технологии получения крупы и удаления анатомических частей определяется не столь активностью амилолитических ферментов, расщепляющих крахмал до низкомолекулярных декстринов и олигосахаридов, сколько степенью автолиза крахмала и накоплением водорастворимых веществ при прогреве водномучных суспензий. Качество крупяной продукции может определяться на основе инструментальной оценки проб муки стандартным методом числа падения, реализуемым в условиях внутризаводского контроля. В качестве улучшителей амилолитической активности муки из крупяных культур следует рассматривать зерновые продукты и их композиции, определяющие получение феноменологической модели прогнозирования варочных свойств готового продукта» [6].

Целью данной работы явилась проверка гипотезы нормальности числа падения овсяной муки по малым выборкам эксперимента с использованием непараметрического критерия согласия ω^2 Мизерса, поскольку, в отличие от широко применяемого критерия χ^2 Пирсона, рассматриваемый критерий «...основывается на непосредственно полученных в эксперименте негруппированных значениях случайной величины» [1].

Материалы и методы

Материалом исследования являлся образец овсяной муки, соответствующий требованиям СТО 53548590-019-2013 (ООО «ХЛЕБЗЕРНО-ПРОДУКТ», г. Таганрог, Ростовская область). Число падения овсяной муки измеряли стандартным методом ГОСТ 27676-88 [7]: «...сущность которого заключается в определении времени свободного падения шток-мешалки в клейстеризованной водно-мучной суспензии». За окончательный результат числа падения принимают среднее арифметическое результатов параллельного определения двух навесок, допускаемое расхождение между которыми не должно превышать 10% от их средней арифметической величины. При превышении допускаемого расхождения определение повторяют. Вычисления проводят до первого десятичного знака с последующим округлением результата до целого числа.

В данном стандарте приводятся правила округления повторных результатов измерений, производимых двухканальной измерительной системой прибора типа ПЧП. Так, в *примере 1* результат определения по первой навеске составляет 150 с, по второй – 160 с. Среднее арифметическое значение – 155 с. Допускаемое расхождение от этого среднего арифметического значения составляет 15,5 с. Фактическое расхождение между результатами параллельного определения двух навесок составляет 10 с, что не превышает допускаемого расхождения между ними. Среднее арифметическое значение (155 с) принимают за окончательный результат определения числа падения.

При контрольном (повторном) определении числа падения допускаемое расхождение между контрольным (повторным) и первоначальным определением не должно превышать 10% от их средней арифметической величины. При контрольном (повторном) определении за окончательный результат принимают результат первоначального определения, если расхождение между результатами контрольного (повторного) и первоначального определений не превышает допускаемого значения; если расхождение превышает допускаемое значение, за окончательный результат принимают результат контрольного (повторного) определения.

В *примере 2* результат первоначального определения составляет 150 с, контрольного (повторного) – 170 с. Среднее арифметическое значение – 160 с. Допускаемое расхождение от этого среднего значения составляет 16 с. Фактическое расхождение составляет 16 с.

Фактическое расхождение между результатами первоначального и контрольного (повторного) определений составляет 20 с, что превышает допускаемое расхождение. За окончательный результат определения числа падения принимают результат контрольного (повторного) определения – 170 с.

В *примере 3* результат первоначального определения составляет 150 с, контрольного (повторного) – 160 с. Среднее арифметическое значение – 155 с. Допускаемое расхождение составляет 15,5 с. Фактическое расхождение между результатами первоначального и контрольного (повторного) определений составляет 10 с, что не превышает допускаемого расхождения. За окончательный результат принимают результат первоначального определения – 150 с.

Округление результатов определения проводят следующим образом: если первая из отбрасываемых цифр равна или больше 5, то последнюю сохраняемую цифру увеличивают на единицу; если меньше 5, то ее оставляют без изменения [7].

Очевидно, что при таком подходе проведение многократных (повторных) измерений для серии опытов и соблюдение правил округления приводит к получению ожидаемых одинаковых (и/или) повторных результатов от одного определения к другому, случайный характер которых либо исключается, либо носит характер равномерного распределения. В связи с этим, для изучения числа падения как случайной величины была получена совокупность малых выборок для проверки гипотезы о том, что функция распределения такой случайной величины есть $F(x)$, для чего необходимо было построить эмпирическую функцию распределения $F_n(x)$ для сравнения с ней путем применения непараметрического критерия согласия ω^2 Мизерса

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x) \quad (4)$$

предполагая наличие у $F(x)$ производной, т. е. плотности вероятности, имеем выражение

$$\partial F(x) = F'(x) dx = f(x) dx \quad (5)$$

Алгоритм применения непараметрического критерия согласия ω^2 Мизерса при проверке гипотезы нормальности по совокупности малых выборок заключался в следующем. Вначале выборки преобразовывали в вариационный ряд $x_1 < x_2 < x_3 < \dots < x_n$ и разбивали всю область интегрирования на интервалы $(-\infty, x_1)$, (x_1, x_2) , ..., (x_{n-1}, x_n) , $(x_n, +\infty)$.

Принимая во внимание выражение (4), получали

$$\omega^2 = \int_{-\infty}^{x_1} [0 - F(x)]^2 \partial F + \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} \left[\frac{k}{n} - F(x) \right]^2 \partial F +$$

$$+ \int_{x_n}^{+\infty} [1 - F(x)]^2 \partial F.$$

$$\int_{-\infty}^{x_1} F^2(x) \partial F = \frac{F^3(x)}{3} \Big|_{-\infty}^{x_1} = \frac{F^3(x_1)}{3};$$

$$\int_{x_k}^{x_{k+1}} \left[\frac{k}{n} - F(x) \right]^2 \partial F = \left[\frac{F(x) - \frac{k}{n}}{3} \right]^3 \Big|_{x=x_k}^{x=x_{k+1}} =$$

$$= \frac{\left[F(x_{k+1}) - \frac{k}{n} \right]^3}{3} - \frac{\left[F(x_k) - \frac{k}{n} \right]^3}{3}$$

$$\int_{x_n}^{+\infty} [1 - F(x)]^2 \partial F = \frac{[1 - F(x_n)]^3}{3}$$

После преобразования выражение (7) имело вид

$$\omega^2 = \frac{F^3(x_1)}{3} + \sum_{k=1}^{n-1} \frac{\left[F(x_{k+1}) - \frac{k}{n} \right]^3}{3} -$$

$$- \sum_{k=1}^{n-1} \frac{\left[F(x_k) - \frac{k}{n} \right]^3}{3} + \frac{[1 - F(x_n)]^3}{3}$$

Объединив члены, зависящие от $F(x_k)$ с данными $k = 1, 2, \dots, n$, находящиеся в двух суммах (8), получили

$$\omega^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^n \left[F(x_k) - \frac{2k-1}{2n} \right]^2 \quad (9)$$

Равенство (9) показывает, каким образом критерий ω^2 Мизерса зависит от отдельных членов вариационного ряда. Критические значения для величины $n\omega^2$, т. е. квантили $(n\omega^2)_{1-p}$ приведены в таблице 1.

Таблица 1.
Квантили распределения $n\omega^2$ [1]
Table 1.
Quantiles of the distribution $n\omega^2$ [1]

p	$(n\omega^2)_{1-p}$	p	$(n\omega^2)_{1-p}$
0,5	0,1184	0,05	0,4614
0,4	0,1467	0,03	0,5489
0,3	0,1843	0,02	0,6198
0,2	0,2412	0,01	0,7435
0,1	0,3473	0,001	1,1679

Если вычисленное значение $n\omega^2$ меньше табличного $(n\omega^2)_{1-p}$, то гипотеза о совпадении теоретического закона распределения $F(x)$ с эмпирическим (выборочным) $F_n(x)$ не отвергается, но при $n\omega^2$ больше табличного $(n\omega^2)_{1-p}$ гипотеза отклоняется. Отмечается, что распределение $n\omega^2$ значительно быстрее, чем χ^2 , сходится к предельному закону, особенно

в области больших значений ω^2 , существенных для вероятностной оценки.

Результаты

Выборки опытных данных числа падения овсяной муки сведены в таблицу 2. На первом этапе исследования выбирали из результатов четырех параллельных определений первое значение показателя числа падения x_1 и вычисляли для каждого из 12 выборок величину относительного отклонения τ по формуле (1), результаты расчетов приведены в таблице 3. Далее следовала проверка гипотезы нормальности распределения выборок при помощи непараметрического критерия согласия ω^2 Мизерса, для чего расположили полученные значения τ в вариационный ряд, и для каждого элемента ряда определяли значения эмпирической функции распределения $F_n(\tau)$ по формуле

$$F_n(\tau) = \frac{2k-1}{2n} \quad (10)$$

где $n = 12$, $k = 1, 2, \dots, n$ и значения теоретической функции $F(\tau)$ по формуле (3).

Результаты расчета теоретической и эмпирической функций распределения для первого определения выборки сведены в таблице 4, где значения τ расположены в порядке возрастания.

Таблица 2.
Выборки опытных данных числа падения овсяной муки

Table 2.
Sampling of experimental data of the falling number of oat flour

№.	x_1	x_2	x_3	x_4
1	336	342	345	336
2	342	340	334	340
3	336	341	341	342
4	332	343	332	351
5	342	346	339	348
6	347	348	342	349
7	332	333	336	334
8	335	341	336	334
9	347	353	339	335
10	337	340	333	333
11	338	335	330	331
12	342	334	332	339

Таблица 3.
Расчет относительного отклонения для первого определения выборки

Table 3.
Calculation of the relative deviation for the first sample definition

№.	$\bar{X}$	S_k	$x_1 - \bar{X}$	τ_1
1	339,75	3,897	-3,75	-0,962
2	339,00	3,000	+3,00	+1,000
3	340,00	2,345	-4,00	-1,706
4	339,50	8,016	-7,50	-0,936
5	343,75	3,491	-1,75	-0,501
6	346,50	2,693	+0,50	+0,186
7	333,75	1,479	-1,75	-1,183
8	336,50	2,693	-1,50	-0,557
9	343,50	6,982	+3,50	+0,501
10	335,75	2,947	+1,25	+0,424
11	333,50	3,202	+4,50	+1,405
12	336,75	3,961	+5,25	+1,325

Таблица 4.

Расчет теоретической и эмпирической функций распределения для первого определения выборки

Table 4.

Calculation of theoretical and empirical distribution functions for the first determination of the sample

№	τ_1	$F(\tau_1)$	$F_n(\tau_1)$	$F(\tau_1) - F_n(\tau_1)$	$[F(\tau_1) - F_n(\tau_1)]^2$
1	-0,962	+ 0,0087	+ 0,0417	- 0,0330	0,001089
2	+ 1,000	+ 0,1830	+ 0,1250	+ 0,0580	0,003364
3	- 1,706	+ 0,2567	+ 0,2083	+ 0,0484	0,002343
4	- 0,936	+ 0,2654	+ 0,2917	- 0,0263	0,000692
5	- 0,501	+ 0,3917	+ 0,3750	+ 0,0167	0,000279
6	+ 0,186	+ 0,3554	+ 0,4583	- 0,1029	0,010588
7	- 1,183	+ 0,5537	+ 0,5417	+ 0,0120	0,000144
8	- 0,557	+ 0,6224	+ 0,6250	- 0,0026	0,000007
9	+ 0,501	+ 0,6446	+ 0,7083	- 0,0637	0,004058
10	+ 0,424	+ 0,7887	+ 0,7917	- 0,0030	0,000009
11	+ 1,405	+ 0,8825	+ 0,8750	+ 0,0075	0,000056
12	+ 1,325	+ 0,8750	+ 0,9056	- 0,0306	0,000936
Σ					0,023565

Вычисления производили, принимая во внимание, что $f(\tau) = 0$ при $\tau < -\sqrt{3}$, откуда имеем

$$F(\tau) = \int_{-\sqrt{3}}^{\tau_k} \frac{1}{2\sqrt{3}} d\tau = \frac{1}{2\sqrt{3}}(\tau_k + \sqrt{3}), k = 1, 2, \dots, n \quad (11)$$

Отсюда, при $\tau_1 = -1,706$

$$F(-1,706) = \frac{-1,706 + \sqrt{3}}{2\sqrt{3}} = \frac{-1,706 + 1,732}{2 \cdot 1,732} = 0,0087.$$

Значение эмпирической функции определяли по формуле (10), при $\tau = \tau_1$ получили

$$F_n(\tau_1) = \frac{2 \times 1 - 1}{2 \times 12} = \frac{1}{24} = 0,0417.$$

Расчет статистики $n\omega^2$ производили по упрощенной формуле

$$n\omega^2 = \frac{1}{12n} + \sum_{k=1}^n \left[F(\tau_k) - \frac{2k-1}{2n} \right]^2 \quad (12)$$

Подставив значение $n = 12$ и полученную из таблицы 4 сумму $[F(\tau_1) - F_n(\tau_1)]^2$ получили значение статистики $n\omega^2$

$$n\omega^2 = \frac{1}{12 \times 2} + 0,023565 = 0,03051 \text{ (для первого определения).}$$

Для уровня значимости $p = 0,05$ табличное значение $(n\omega^2)_{1-p} = 0,4614$, а вычисленное значение $n\omega^2$ (0,03051) меньше табличного. Следовательно, гипотеза нормального распределения малых выборок по первому определению числа падения овсяной муки не отклоняется. Для проверки данного утверждения в отношении второго, третьего и четвертого определений произвели дополнительные этапы расчетов, результаты которых приведены в таблицах 5–10.

Таблица 5.

Расчет относительного отклонения для второго определения выборки

Table 5.

Calculation of the relative deviation for the second sample definition

№ выборки Sample No.	$\bar{X}$	S_k	$x_3 - \bar{X}$	τ_3
1	339,75	3,897	+ 2,25	+ 0,577
2	339,00	3,000	+ 1,00	+ 0,333
3	340,00	2,345	+ 1,00	+ 0,426
4	339,50	8,016	+ 3,50	+ 0,437
5	343,75	3,491	+ 2,25	+ 0,645
6	346,50	2,693	+ 1,50	+ 0,557
7	333,75	1,479	- 0,75	- 0,507
8	336,50	2,693	+ 4,50	+ 1,671
9	343,50	6,982	+ 9,50	+ 1,361
10	335,75	2,947	+ 4,25	+ 1,442
11	333,50	3,202	- 0,50	- 0,156
12	336,75	3,961	- 2,75	- 0,694

Таблица 6.

Расчет теоретической и эмпирической функций распределения для второго определения выборки

Table 6.

Calculation of theoretical and empirical distribution functions for the second determination of the sample

№ выборки Sample No.	τ_3	$F(\tau_3)$	$F_n(\tau_3)$	$F(\tau_3) - F_n(\tau_3)$	$[F(\tau_3) - F_n(\tau_3)]^2$
1	- 0,694	+ 0,2997	+ 0,0417	+ 0,2580	0,066564
2	- 0,507	+ 0,3536	+ 0,1250	+ 0,2286	0,052258
3	- 0,156	+ 0,4550	+ 0,2083	+ 0,2467	0,060861
4	+ 0,333	+ 0,5961	+ 0,2917	+ 0,3044	0,092659
5	+ 0,426	+ 0,6230	+ 0,3750	+ 0,2480	0,061504
6	+ 0,437	+ 0,6262	+ 0,4583	+ 0,1679	0,028224
7	+ 0,557	+ 0,6608	+ 0,5417	+ 0,1191	0,014161
8	+ 0,557	+ 0,6666	+ 0,6250	+ 0,0416	0,001731
9	+ 0,645	+ 0,6862	+ 0,7083	- 0,0221	0,000488
10	+ 1,361	+ 0,8929	+ 0,7917	+ 0,1012	0,010241
11	+ 1,442	+ 0,9163	+ 0,8750	+ 0,0413	0,001706
12	+ 1,671	+ 0,9824	+ 0,9056	+ 0,0768	0,005898
Σ					0,396295

$$n\omega^2 = \frac{1}{12 \cdot 12} + 0,396295$$

$$= 0,4032 \text{ (для второго определения).}$$

Для уровня значимости $p = 0,05$ табличное значение $(n\omega^2)_{1-p} = 0,4614$, а вычисленное

значение $n\omega^2$ (0,4032), так же, как и для первого определения, меньше табличного. Отсюда, гипотеза нормального распределения малых выборок по второму определению числа падения овсяной муки не отклоняется.

Таблица 7

Расчет относительного отклонения для третьего определения выборки

Table 7

Calculation of the relative deviation for the third sample definition

№ выборки Sample No.	$\bar{x}$	S_k	$x_3 - \bar{x}$	τ_3
1	339,75	3,897	+ 5,25	+ 1,347
2	339,00	3,000	- 5,00	- 1,667
3	340,00	2,345	+ 1,00	+ 0,426
4	339,50	8,016	- 7,50	- 0,936
5	343,75	3,491	- 4,75	- 1,361
6	346,50	2,693	- 4,50	- 1,671
7	333,75	1,479	+ 2,25	+ 1,521
8	336,50	2,693	- 0,50	- 0,186
9	343,50	6,982	- 4,50	- 0,645
10	335,75	2,947	- 2,75	- 0,933
11	333,50	3,202	- 3,50	- 1,093
12	336,75	3,961	- 4,75	- 1,199

Таблица 8.

Расчет теоретической и эмпирической функций распределения для третьего определения выборки

Table 8.

Calculation of theoretical and empirical distribution functions for the third determination of the sample

№ выборки Sample No.	τ_3	$F(\tau_3)$	$F_n(\tau_3)$	$F(\tau_3) - F_n(\tau_3)$	$[F(\tau_3) - F_n(\tau_3)]^2$
1	- 1,671	+ 0,0176	+ 0,0417	- 0,0241	0,000581
2	- 1,677	+ 0,0188	+ 0,1250	- 0,1062	0,011278
3	- 1,361	+ 0,1071	+ 0,2083	- 0,1012	0,010241
4	- 1,199	+ 0,1539	+ 0,2917	- 0,1378	0,018989
5	- 1,093	+ 0,1845	+ 0,3750	- 0,1905	0,036290
6	- 0,936	+ 0,2298	+ 0,4583	- 0,2285	0,052212
7	- 0,933	+ 0,2307	+ 0,5417	- 0,3110	0,096721
8	- 0,645	+ 0,3138	+ 0,6250	- 0,3112	0,096845
9	- 0,186	+ 0,4463	+ 0,7083	- 0,2620	0,068644
10	+ 0,426	+ 0,6230	+ 0,7917	- 0,1687	0,028460
11	+ 1,347	+ 0,8889	+ 0,8750	+ 0,0139	0,000193
12	+ 1,521	+ 0,9391	+ 0,9056	+ 0,0335	0,001122
Σ					0,421576

Таблица 9.

Расчет относительного отклонения для четвертого определения выборки

Table 9.

Calculation of the relative deviation for the fourth sample definition

№ выборки Sample No.	$\bar{x}$	S_k	$x_4 - \bar{x}$	τ_4
1	339,75	3,897	- 3,75	- 0,962
2	339,00	3,000	+ 1,00	+ 0,333
3	340,00	2,345	+ 2,00	+ 0,853
4	339,50	8,016	+ 11,50	+ 1,435
5	343,75	3,491	+ 4,25	+ 1,217
6	346,50	2,693	+ 2,50	+ 0,928
7	333,75	1,479	+ 0,25	+ 0,169
8	336,50	2,693	- 2,50	- 0,928
9	343,50	6,982	+ 8,50	+ 1,217
10	335,75	2,947	- 2,75	- 0,933
11	333,50	3,202	- 2,50	- 0,781
12	336,75	3,961	+ 2,25	+ 0,568

$$n\omega^2 = \frac{1}{12 \cdot 12} + 0,421576$$

$$= 0,4285 \text{ (для третьего определения).}$$

Для уровня значимости $p = 0,05$ табличное значение $(n\omega^2)_{1-p} = 0,4614$, а вычисленное значение $n\omega^2$ (0,4285), так же, как и для первого и второго определений, меньше табличного. Отсюда, гипотеза нормального распределения малых выборок по третьему определению числа падения овсяной муки также не отклоняется.

Таблица 10.

Расчет теоретической и эмпирической функций распределения для четвертого определения выборки

Table 10.

Calculation of theoretical and empirical distribution functions for the fourth determination of the sample

№ выборки Sample No.	τ_3	$F(\tau_3)$	$F_n(\tau_3)$	$F(\tau_3) - F_n(\tau_3)$	$[F(\tau_3) - F_n(\tau_3)]^2$
1	-0,962	+0,2223	+0,0417	+0,1806	0,032616
2	-0,933	+0,2307	+0,1250	+0,1057	0,011173
3	-0,928	+0,2321	+0,2083	+0,0238	0,000566
4	-0,781	+0,2745	+0,2917	-0,0172	0,000296
5	+0,169	+0,5488	+0,3750	+0,1738	0,030206
6	+0,333	+0,5961	+0,4583	+0,1378	0,018989
7	+0,568	+0,6640	+0,5417	+0,1223	0,014957
8	+0,853	+0,7462	+0,6250	+0,1213	0,014714
9	+0,928	+0,7679	+0,7083	+0,0596	0,003552
10	+1,217	+0,8513	+0,7917	+0,0596	0,003552
11	+1,217	+0,8513	+0,8750	-0,0237	0,000562
12	+1,435	+0,9143	+0,9056	+0,0087	0,000076
Σ					0,131259

$$n\omega^2 = \frac{1}{12 \cdot 12} + 0,131259 = 0,1382 \text{ (для четвертого определения).}$$

Для уровня значимости $p = 0,05$ табличное значение $(n\omega^2)_{1-p} = 0,4614$, а вычисленное значение $n\omega^2$ (0,1382), так же как и для всех предыдущих определений, меньше табличного. Отсюда, гипотеза нормального распределения малых выборок по четвертому определению числа падения овсяной муки также не отклоняется.

В пользу полученных результатов свидетельствует формулировка другого критерия – критерия на отклонение от нормального распределения, служащего для проверки отклонения распределения вероятностей от нормального распределения. В стандарте ГОСТ Р ИСО 5479–2002 [8] «...критерий на отклонение от нормального распределения имеет нулевую гипотезу в том случае, когда выборка из n независимых наблюдений подчиняется одному и тому же нормальному распределению». Но такой критерий для проверки гипотезы используется преимущественно при необходимости сопоставления $\bar{x}$ выборочного с $\bar{x}$ теоретическим, полученным при помощи t -критерия, т. е. параметрического критерия. В противном случае, проверке подлежит простая гипотеза с задаваемой функцией распределения, с которой наблюдают согласие с наблюдаемой выборкой.

Обсуждение

Полученные в работе результаты согласуются с общепринятыми классическими представлениями о проверке статистической гипотезы о нормальном распределении выборок. Указанные в литературе методы синтеза новых статистических критериев для проверки гипотез малых выборок с помощью аналитических и нейросетевых подходов не лишены ряда недостатков. Например, в работе [9] приводится решение задачи по проверке гипотезы нормального распределения исследуемых данных на выборках объемом от 16 до 25 опытов, адаптированных

для осуществления многокритериального анализа. Однако, параллельно с ростом числа объединяемых статистических критериев (до 210) растут взаимные корреляционные связи между новыми синтезированными критериями, что ограничивает возможности указанного метода. Снижение корреляционной сцепленности искусственных нейронов добиваются объединением классических критериев не более трех с настройкой квантователей давать равные вероятности ошибок первого и второго рода [10]. В следующей работе [11] показана возможность оптимизации проверки гипотез нормальности или равномерности распределений малых выборок путем перехода от применения бинарных искусственных нейронов к троичным нейронам при совместном использовании пяти классических статистических критериев. Однако ограничением данного подхода служит использование синтетических статистических критериев при малых выборках в 16 опытов. Другой подход статистического анализа малых выборок путем синтеза новых статистических критериев включает операцию дифференцирования случайных данных малой выборки. Вероятность появления ошибок первого и второго родов снижается до 0,075, что допускается только для ряда приложений нейросетевой биометрии [12]. В данной работе вычисления произведены при $p = 0,05$, что обеспечивает достаточную точность изучаемого показателя (числа падения) в технических системах и не требует проведения синтеза статистического критерия для проверки гипотезы нормальности малых выборок.

Заключение

При проверке гипотезы нормальности малых выборок полагались на мнение специалистов, что уже при четырех повторностях измерений в малой выборке относительные отклонения в отдельных выборках в изучаемой совокупности могут подчиняться равномерному распределению, если исходные совокупности нормальны. Если число выборок достаточно

для анализа, то данное условие позволит выполнить проверку гипотезы нормальности. Как правило, при проверке гипотезы нормальности по большому числу малых выборок из каждой выборки случайным образом отбирается по одному значению только первых, вторых и т. д. измерений, допуская, тем самым, упрощение и случайный отбор данных. Отсюда, целью исследования явилась проверка гипотезы нормальности малых выборок каждого из четырех определений числа падения, поскольку округление значений прямых измерений исключает случайный характер величины или ее нормальное распределение в пользу равномерного. В данном случае, проверке подлежала простая гипотеза с задаваемой функцией распределения, с которой наблюдают согласие с наблюдаемой выборкой

при использовании непараметрического критерия согласия ω^2 Мизерса. При уровне значимости $p = 0,05$ табличное значение $(n\omega^2)_{1-p}$ больше вычисленного значения $n\omega^2$ для всех четырех определений. Поэтому гипотеза нормального распределения малых выборок для всех четырех определений (как случайных величин) числа падения овсяной муки не отклоняется.

Благодарности

Исследования выполнялись с использованием оборудования ЦПК «Исследовательский центр пищевых и химических технологий» КубГТУ (СКР_3111), развитие которого поддерживается Министерством науки и высшего образования РФ (Соглашение № 075–15–2021–679).

Литература

- 1 Ахназарова С.Л., Кафаров В.В. Оптимизация эксперимента в химии и химической технологии. М.: Высшая школа, 1978. 319 с.
- 2 Р 50.1.037–2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. М.: ИПК стандартов, 2002. 66 с.
- 3 Сокол Н.В., Казарцева А.Т., Санжаровская Н.С. Влияние на хлебопекарные свойства смесей пшеничной муки с продуктами переработки овса // Труды Кубанского государственного аграрного университета, 2014. № 49. С. 162–168.
- 4 Вершинина О.Л., Зернаева Е.А., Бондаренко А.Н. Влияние мучной композитной смеси на хлебопекарные свойства ржаной муки // Известия высших учебных заведений. Пищевая технология. 2018. № 2–3 (262–263). С. 44–46. doi: 10.26297/0579–3009.2018.2–3.11
- 5 Буховец В.А., Ефимова Д.В., Давыдова Л.В. Разработка технологии производства хлебобулочных изделий повышенной пищевой ценности // Техника и технология пищевых производств. 2019. Т. 49. №. 2. С. 193–200.
- 6 Шмалько Н.А. Автолитическая активность муки из крупяных культур // Научные труды КубГТУ. 2019. № 59. С. 29–40.
- 7 ГОСТ 27676–88. Зерно и продукты его переработки. Метод определения числа падения. М.: Стандартиформ, 2009. 5 с.
- 8 ГОСТ Р ИСО 5479–2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения. М.: ИПК стандартов, 2002. 31 с.
- 9 Иванов А.И., Куприянов Е.Н. Синтез новых более мощных статистических критериев через мультипликативное объединение классических критериев Фрощини и Мурота-Такеучи с критерием Херста для проверки гипотезы малых выборок // Надежность. 2022. № 1. Т. 22. С. 52–55. doi: 10.21683/1729–2646–2022–22–1–52–55
- 10 Лукин В.С. Сравнение мощности обычной и логарифмической форм статистических критериев среднего гармонического при использовании для проверки гипотезы нормального распределения данных малой выборки // Известия высших учебных заведений. Поволжский регион. Технические науки. 2020. № 4 (56). С. 19–26. doi: 10.21685/2072–3059–2020–4–2
- 11 Иванов А.И., Савинов К.Н., Еременко Р.В. Эффект перехода от применения бинарных искусственных нейронов к троичным нейронам при совместном использовании пяти классических статистических критериев проверки гипотез нормальности или равномерности распределений малых выборок // Вестник Пермского университета. Математика. Механика. Информатика. 2022. № 3 (58). С. 59–67. doi: 10.17072/1993–0550–2022–3–59–67
- 12 Иванов А.И., Малыгин А.Ю., Полковникова С.А. Новый статистический критерий большой мощности, полученный дифференцированием случайных данных малой выборки // Известия высших учебных заведений. Поволжский регион. Технические науки. 2021. № 3 (59). С. 67–74. doi: 10.21685/2072–3059–2021–3–7
- 13 Jarman K.H. Beyond basic statistics: Tips, tricks, and techniques every data analyst should know. John Wiley & Sons, 2015.
- 14 Lee M.H., Lee J.S., Lee J.Y., Kim Y.H. et al. Uncertainty analysis of a GHG emission model output using the block bootstrap and Monte Carlo simulation // Sustainability. 2017. V. 9. №. 9. P. 1522. doi: 10.3390/su9091522
- 15 Ribeiro L.F.O., Vitoria E.L.D., Soprani Júnior G.G., Chen P. et al. Impact of Operational Parameters on Droplet Distribution Using an Unmanned Aerial Vehicle in a Papaya Orchard // Agronomy. 2023. V. 13. №. 4. P. 1138. doi: 10.3390/agronomy13041138
- 16 Yeater K.M., Duke S.E., Riedell W.E. Multivariate analysis: greater insights into complex systems // Agronomy Journal. 2015. V. 107. №. 2. P. 799–810. doi: 10.2134/agronj14.0017
- 17 Munkholm L. J., Heck R. J., Deen B. Long-term rotation and tillage effects on soil structure and crop yield // Soil and Tillage Research. 2013. V. 127. P. 85–91. doi: 10.1016/j.still.2012.02.007
- 18 Tsitko I., Wiik-Miettinen F., Mattila O., Rosa-Sibakov N. et al. A small in vitro fermentation model for screening the gut microbiota effects of different fiber preparations // International journal of molecular sciences. 2019. V. 20. №. 8. P. 1925. doi:10.3390/ijms20081925
- 19 Tosh S.M. Review of human studies investigating the post-prandial blood-glucose lowering ability of oat and barley food products // European journal of clinical nutrition. 2013. V. 67. №. 4. P. 310–317. doi: 10.1038/ejcn.2013.25
- 20 Carranza-García M., García-Gutiérrez J., Riquelme J.C. A framework for evaluating land use and land cover classification using convolutional neural networks // Remote Sensing. 2019. V. 11. №. 3. P. 274. doi: 10.3390/rs11030274

References

- 1 Akhnazarova S.L., Kafarov V.V. Optimization of experiment in chemistry and chemical technology. Moscow, Higher School, 1978. 319 p. (in Russian).
- 2 Р 50.1.037–2002. Applied statistics. Rules for verifying the agreement of the experimental distribution with the theoretical one. Part II. Nonparametric criteria. M., IPK standards, 2002. 66 p. (in Russian).
- 3 Sokol N.V., Kazartseva A.T., Sanzharovskaya N.S. Influence on baking properties of wheat flour mixtures with oat processing products. Proceedings of the Kuban State Agrarian University, 2014. no. 49. pp. 162–168. (in Russian).

- 4 Vershinina O.L., Zemaeva E.A., Bondarenko A.N. The influence of a flour composite mixture on the baking properties of rye flour. *Izvestia of higher educational institutions. Food technology*. 2018. no. 2-3 (262-263). pp. 44-46. doi: 10.26297/0579-3009.2018.2.3.11 (in Russian).
- 5 Bukhovets V.A., Efimova D.V., Davydova L.V. Development of technology for the production of bakery products of increased nutritional value. *Technique and technology of food production*. 2019. vol. 49. no. 2. pp. 193-200. (in Russian).
- 6 Shmal'ko N.A. Autolytic activity of flour from cereals. *Scientific works of KubSTU*. 2019. no. S9. pp. 29-40. (in Russian).
- 7 GOST 27676-88. Grain and its processed products. Method for determining the number of falls. M., Standartinform, 2009. 5 p. (in Russian).
- 8 GOST R ISO 5479-2002. Statistical methods. Checking the deviation of the probability distribution from the normal distribution. M., IPK Standards, 2002. 31 p. (in Russian).
- 9 Ivanov A.I., Kupriyanov E.N. Synthesis of new more powerful statistical criteria through multiplicative combination of classical Frocini and Murota-Takeuchi criteria with the Hurst criterion for testing the hypothesis of small samples. *Reliability*. 2022. no. 1. vol. 22. pp. 52-55. doi: 10.21683/1729-2646-2022-22-1-52-55 (in Russian).
- 10 Lukin V.S. Comparison of the power of the usual and logarithmic forms of statistical criteria of the harmonic mean when used to test the hypothesis of the normal distribution of small sample data. *Izvestiya higher educational institutions. Volga region. Technical sciences*. 2020. no. 4 (56). pp. 19-26. doi: 10.21685/2072-3059-2020-4-2 (in Russian).
- 11 Ivanov A.I., Savinov K.N., Eremenko R.V. The effect of transition from the use of binary artificial neurons to ternary neurons with the combined use of five classical statistical criteria for testing hypotheses of normality or uniformity of distributions of small samples. *Bulletin of Perm University. Mathematics. Mechanics. Computer science*. 2022. no. 3 (58). pp. 59-67. doi: 10.17072/1993-0550-2022-3-59-67 (in Russian).
- 12 Ivanov A.I., Malygin A.Yu., Polkovnikova S.A. A new statistical criterion of high power obtained by differentiating random data of a small sample. *Izvestia of higher educational institutions. Volga region. Technical sciences*. 2021. no. 3 (59). pp. 67-74. doi: 10.21685/2072-3059-2021-3-7 (in Russian).
- 13 Jarman K.H. *Beyond basic statistics: Tips, tricks, and techniques every data analyst should know*. John Wiley & Sons, 2015.
- 14 Lee M.H., Lee J.S., Lee J.Y., Kim Y.H. et al. Uncertainty analysis of a GHG emission model output using the block bootstrap and Monte Carlo simulation. *Sustainability*. 2017. vol. 9. no. 9. pp. 1522. doi: 10.3390/su9091522
- 15 Ribeiro L.F.O., Vitoria E.L.D., Soprani Júnior G.G., Chen P. et al. Impact of Operational Parameters on Droplet Distribution Using an Unmanned Aerial Vehicle in a Papaya Orchard. *Agronomy*. 2023. vol. 13. no. 4. pp. 1138. doi: 10.3390/agronomy13041138
- 16 Yeater K.M., Duke S.E., Riedell W.E. Multivariate analysis: greater insights into complex systems. *Agronomy Journal*. 2015. vol. 107. no. 2. pp. 799-810. doi: 10.2134/agronj14.0017
- 17 Munkholm L. J., Heck R. J., Deen B. Long-term rotation and tillage effects on soil structure and crop yield. *Soil and Tillage Research*. 2013. vol. 127. pp. 85-91. doi: 10.1016/j.still.2012.02.007
- 18 Tsitko I., Wiik-Miettinen F., Mattila O., Rosa-Sibakov N. et al. A small in vitro fermentation model for screening the gut microbiota effects of different fiber preparations. *International journal of molecular sciences*. 2019. vol. 20. no. 8. pp. 1925. doi: 10.3390/ijms20081925
- 19 Tosh S.M. Review of human studies investigating the post-prandial blood-glucose lowering ability of oat and barley food products. *European journal of clinical nutrition*. 2013. vol. 67. no. 4. pp. 310-317. doi: 10.1038/ejcn.2013.25
- 20 Carranza-García M., García-Gutiérrez J., Riquelme J.C. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*. 2019. vol. 11. no. 3. pp. 274. doi: 10.3390/rs11030274

Сведения об авторах

Наталья А. Шмалько к.т.н., доцент, кафедра пищевой инженерии, Кубанский государственный технологический университет, ул. Московская, 2, г. Краснодар, 350072, Россия, kafedra-tith@yandex.ru

<https://orcid.org/0000-0002-8939-0265>

Игорь А. Никитин д.т.н., доцент, зав. кафедрой биотехнологий пищевых продуктов из растительного и животного сырья, Московский государственный университет технологий и управления им. К.Г. Разумовского, ул. Земляной Вал, 73, г. Москва, 109004, Россия, nikito.igor@gmail.com

<https://orcid.org/0000-0002-8988-5911>

Дарья А. Велина младший научный сотрудник, кафедра биотехнологий пищевых продуктов из растительного и животного сырья, Московский государственный университет технологий и управления им. К.Г. Разумовского, ул. Земляной Вал, 73, г. Москва, 109004, Россия, d.velina@mgutn.ru

<https://orcid.org/0000-0002-2398-3375>

Лилия Ф. Пономарева к.б.н., доцент, зав. кафедрой технологии пищевых производств, Московский государственный университет технологий и управления им. К.Г. Разумовского, ул. Смоленская, 34, г. Мелеуз, 453856, Россия, eponomareva@mgutn.ru

<https://orcid.org/0000-0002-6645-6356>

Сергей Е. Терентьев к.с.-х.н., врио ректора, Смоленская государственная сельскохозяйственная академия, ул. Большая Советская, 10/2, г. Смоленск, 214000, Россия, seterentiev@sgsha.ru

<https://orcid.org/0000-0002-8546-7253>

Вклад авторов

Все авторы в равной степени принимали участие в написании рукописи и несут ответственность за плагиат

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Information about authors

Natalya A. Shmal'ko Cand. Sci. (Engin.), associate professor, food engineering department, Kuban State Technological University, Moskovskaya St., 2, Krasnodar, 350072, Russia, kafedra-tith@yandex.ru

<https://orcid.org/0000-0002-8939-0265>

Igor A. Nikitin Dr. Sci. (Engin.), associate professor, head of department of biotechnology of food products from plant and animal raw materials, K.G. Razumovsky Moscow State University of technologies and management, Zemlyanoy Val, 73, Moscow, 109004, Russia, nikito.igor@gmail.com

<https://orcid.org/0000-0002-8988-5911>

Daria A. Velina junior researcher, biotechnology of food products from plant and animal raw materials department, K.G. Razumovsky Moscow State University of technologies and management, Zemlyanoy Val, 73, Moscow, 109004, Russia, d.velina@mgutn.ru

<https://orcid.org/0000-0002-2398-3375>

Lilya F. Ponomareva Cand. Sci. (Biology), associate professor, head of department of food production technologies, K.G. Razumovsky Moscow State University of technologies and management, Smolenskaya st., 34, Meleuz, 453856, Russia, eponomareva@mgutn.ru

<https://orcid.org/0000-0002-6645-6356>

Sergey E. Terentiev Cand. Sci. (Agric.), interim rector, Smolensk State Agricultural Academy, Bol'shaya Sovetskaya str., 10/2, Smolensk, 214000, Russia, seterentiev@sgsha.ru

<https://orcid.org/0000-0002-8546-7253>

Contribution

All authors are equally involved in the writing of the manuscript and are responsible for plagiarism

Conflict of interest

The authors declare no conflict of interest.

Поступила 25/12/2022	После редакции 25/01/2023	Принята в печать 20/02/2023
Received 25/12/2022	Accepted in revised 25/01/2023	Accepted 20/02/2023