УДК 519.246.3: 681.3.06

К ПРОВЕРКЕ ОТКЛОНЕНИЯ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ОТ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ В СИСТЕМЕ КОМПЬЮТЕРНОЙ МАТЕМАТИКИ MATLAB

А.В. НЕСТЕРОВ, С.В. НЕСТЕРОВ, Д.А. КОЗАК

Кубанский государственный технологический университет,
350072, Российская Федерация, г. Краснодар, ул. Московская, 2;
электронная почта: briefkasten129@rambler.ru

Исследована возможность проверки отклонения распределения вероятностей от нормального распределения по ГОСТ 5479-2002 с помощью системы MATLAB и ее пакета расширения Statistics Toolbox. Особое внимание уделено автоматизации построения статистических графиков и расчета статистик критериев согласия. Работоспособность предложенных процедур подтверждена при решении тестовых задач.
Ключевые слова: нормальный график, гистограмма, критерии асимметрии и эксцесса, критерий Шапиро-Уилка, критерий Эппса-Палли.

Задача, название которой вынесено в заголовок статьи, является типовой в прикладной статистике. Известно несколько вариантов ее решения. Часть из них рекомендована к применению ГОСТ 5479-2002 [1]. В учебной практике изучают все критерии, предлагаемые этим стандартом. В связи с этим возникает проблема программного обеспечении (ПО) расчета названных статистик. С учебной точки зрения для этого целесообразно использовать один программный продукт. Им может быть не специализированная статистическая система (Statistica, SPSS, Minitab и др.), а универсальная система компьютерной математики (MATLAB, MathCAD, Maple и другие), используемая при изучении спецдисциплин. Такая ситуация складывается в ТАУ при исследовании систем управления статистическими методами [2]. Эффективным средством решения типовых задач ТАУ служит MATLAB и его пакеты расширения. Примеры использования пакета расширения Statistics Toolbox для идентификации электроприводов даны в работах [3-5]. Поэтому естественной выглядит попытка использовать эту систему также для проверки отклонения распределения вероятностей эмпирических данных от нормального по ГОСТ 5479-2002. Это составляет предмет настоящего исследования.

Рассматриваемый стандарт устанавливает следующие методы и виды критериев:

1) графический метод;

2) направленные критерии $\sqrt{b_1}$ и b_2;

3) многонаправленный критерий $\sqrt{b_1}$ и b_2;

4) многосторонние критерии Шапиро-Уилка и Эппса-Палли;

5) совместный критерий (модифицированный критерий Шапиро-Уилка).

Любопытно, что порядок следования критериев, принятый стандартом, отражает трудоемкость их вычисления с помощью MATLAB. То есть, самым простым для пользователя является графический метод. Его основным инструментом служит нормальный вероятностный график (normal probability plot или, короче, normplot). Стандарт приводит пример построения такого графика вручную. Этот пример принят в качестве тестового при исследовании возможностей MATLAB. Проверке на "нормальность" подлежит некоторый вариационный ряд (см. табл.1 в [1]) X_k = 0,2; 0,33; 0,445; 0,49; 0,78; 0,92; 0,95; 0,97; 1,04; 1,71; 2,22; 2,275; 3,65; 7,0; 8,8.

С наибольшей быстротой соответствующий ему нормальный график строят с помощью *M*-функции *normplot* из библиотеки пакета расширения Statistics Toolbox [6] согласно Script 1.

SCRIPT 1:

```
Xk=[0.2 0.33 0.445 0.49 0.78 0.92 0.95 0.97 1.04 1.71 2.22 2.275 3.65 7.0 8.8];
normplot(Xk)
```

Результат показан на рисунке 1 и, как это следует из сравнения, ничем не отличается от нормального графика, полученного в стандарте.

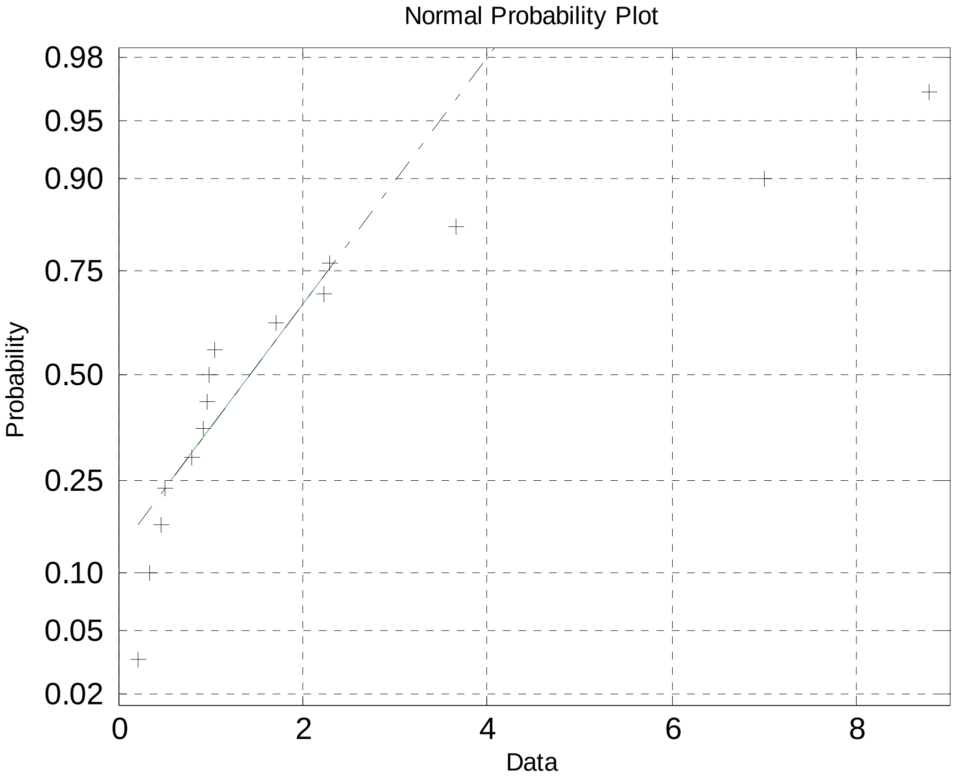

Рисунок 1 – Нормальный вероятностный график

В первом случае для изображения эмпирической функции распределения (ЭФР) использованы маркеры "+". В непосредственной близости от них проведена пунктиром прямая линия, построенная при линейной аппроксимации ЭФР. Сходство двух рассматриваемых графических зависимостей свидетельствует в пользу гипотезы нормального распределения исследуемой выборки x_k [7, 8]. В противном случае различие может указывать на другой вид распределения (логнормальный, экспоненциальный и др.) [8]. Точный ответ дает специальное исследование, но это уже другая задача (подробности см. ГОСТ 5479-2002).

Как показывает рисунок 1, полученная ЭФР существенно отличается от прямой, следовательно, гипотеза нормальности отклоняется.

Заметим в таком случае, что предложенный стандартом пример завершается отрицательной оценкой. Поэтому с учебной точки зрения для полной характеристики графического метода целесообразно привести "положительный" пример нормальной выборки. В связи с этим далее рассмотрен пример 2.10 из [9]. Задан следующий вариационный ряд X_k = 434;

436; 443; 445; 445; 446; 447; 447; 448; 451; 452; 453; 456; 458; 458; 462; 462; 468; 472; 477.

Ему соответствует нормальный график, показанный на рисунке 2.

Ясно видно, что ЭФР хорошо аппроксимируется прямой линией и, значит, нет достаточных оснований для отклонения гипотезы о нормальном распределении.

Этим возможности MATLAB и Statistics Toolbox не исчерпываются: вычислительным процессом можно управлять для достижения каких-либо учебных целей. Так при углубленном изучении графического метода можно решать ту же задачу построения нормального вероятностного графика, не привлекая *M*-функции *normplot*. Очевидно, что этот путь сложнее за счет поэтапного "программировани" задачи, которое позволяет "визуализировать" ее сущность.

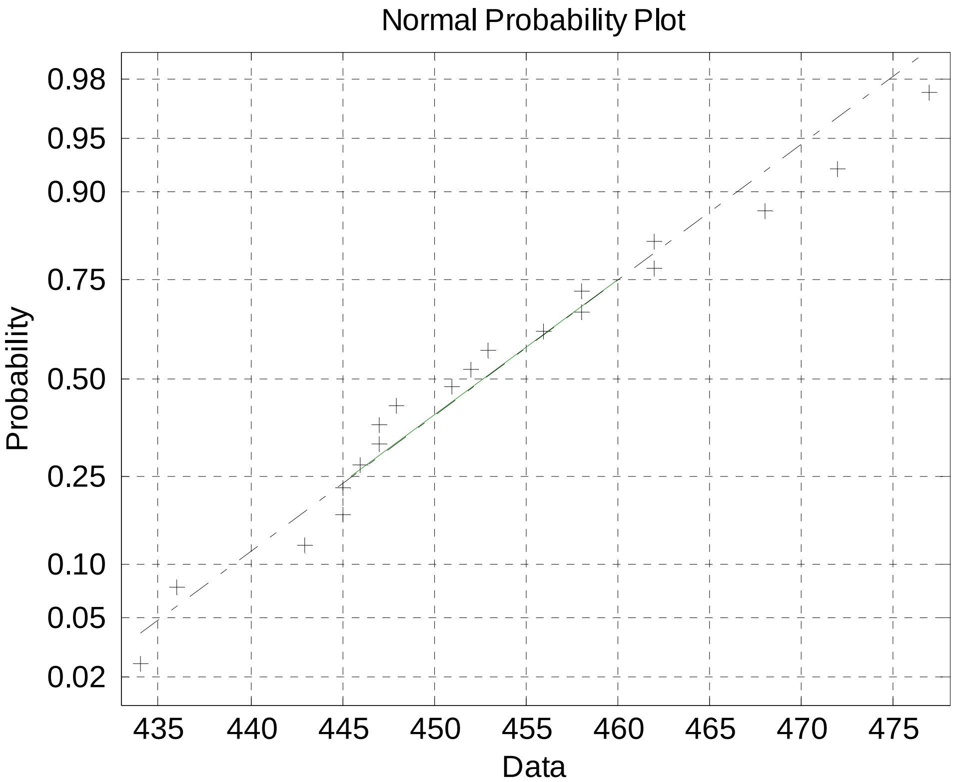

Рисунок 2 – Нормальный вероятностный график (пример 2.10)

В качестве соответствующего алгоритма можно воспользоваться методическими указаниями к рассмотренному примеру 2.10 [9]. Согласно им в общем случае необходимо:

1) по заданной выборке x_i получить вариационный ряд X_k; для этого предназначена *M*-функция *sort* [6];

2) для каждого члена вариационного ряда вычислить накопленные частости

$$P_k = \frac{k - 0{,}5}{n}, \qquad (1)$$

которые служат оценками вероятности события $X_k < x_i$;

3) вычислить с помощью функции *norminv* [6] соответствующие *p*-квантили нормального распределения *Z* или обратную функцию распределения

$$Z = \Phi^{-1}(P_k); \qquad (2)$$

4) построить нормальный вероятностный график с помощью функции *plot*(x, z).

Результат выполнения указанных действий в соответствии со Script 2 показан на рисунке 3.

SCRIPT 2:

```
Xk=[434 436 443 445 445 446 447 447 448 451 452 453 456 458 458 462 462 468 472 477];
n=length(Xk);
k=1:1:n;
Pk=(k-0.5)/n;
Z=norminv(Pk);
plot(Xk,Z,’*’)
```

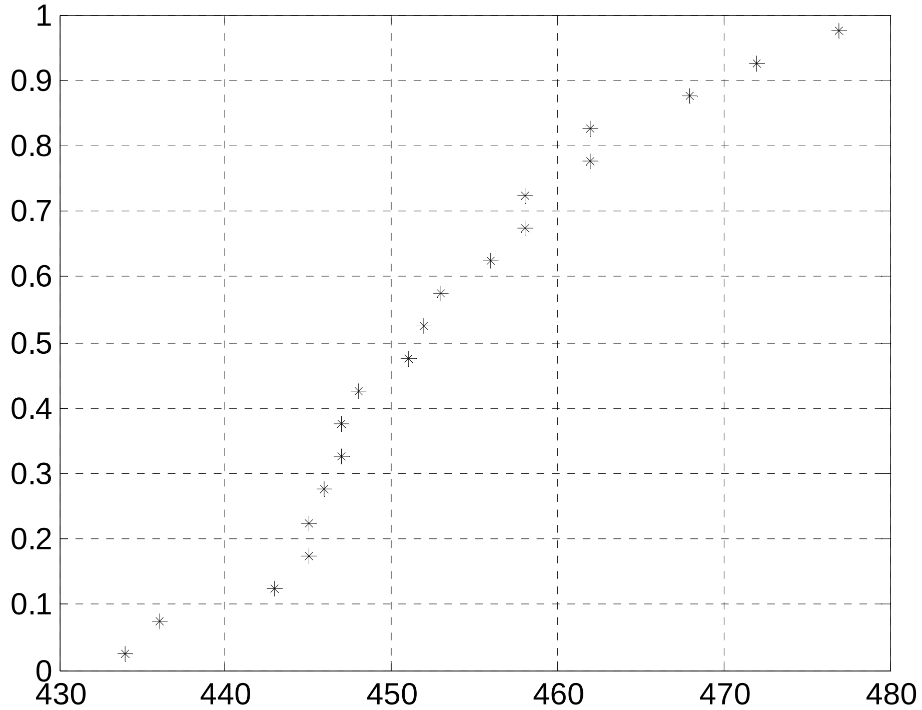

Рисунок 3 – Нормальный вероятностный график

Аппроксимирующую прямую можно построить несколькими способами с помощью разных *M*-функций [6], из которых *lsline* требует наименьших затрат труда пользователя. Результат показан на рисунке 4.

Заметим также, что аппроксимация выполнена методом наименьших квадратов. От "классического" нормального графика с вероятностной сеткой (рис.1) анализируемый рисунок 4 отличается равномерной сеткой. Первая из названных сеток может служить для приближенной выборочной оценки параметров нормального распределения μ и σ, если соответствующая гипотеза принята. В настоящей работе такая задача не ставится. Более того, при необходимости ее можно решить значительно проще и точнее при помощи *M*-функций *mean* и *std* (см. Script 3). В связи с этим пользователь может не затруднять себя построением вероятностной сетки, а принять рисунок 4 как достаточный для искомой оценки нормальности.

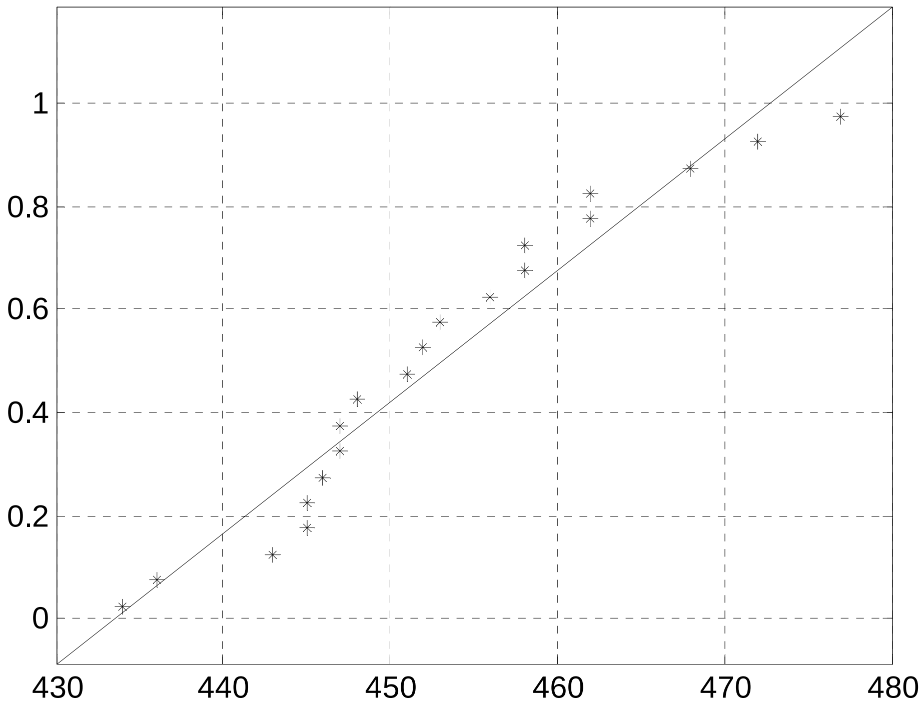

Рисунок 4 – Нормальный вероятностный график

В дидактическом отношении полезным может оказаться рисунок 5, построенный согласно Script 3 (здесь mu и sigma – соответственно среднее значение и СКО исследуемой выборки X_k).

SCRIPT 3:

```
Xk=sort(Xi);
n=length(Xk);
mu=mean(Xk);
sigma=std(Xk);
k=1:1:n;
Pk=(k-0.5)/n;
Z=norminv(Pk,mu,sigma);
Fx=normcdf(Xk, mu,sigma);
plot(Xk,Z,'*',Xk,Fx); grid on
```

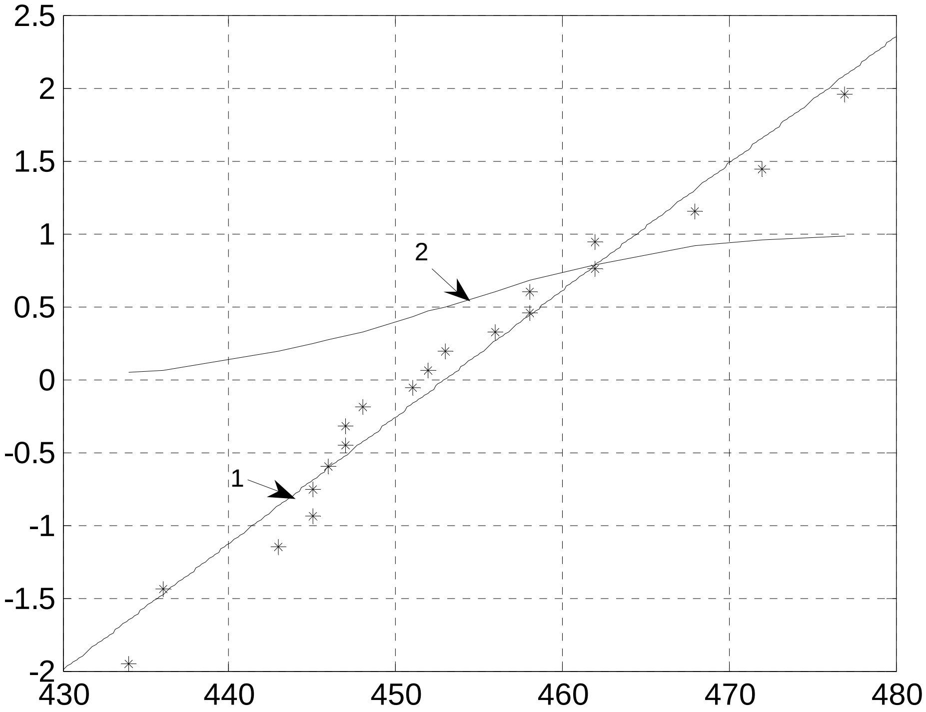

Рисунок 5 – Нормальный вероятностный график

Очевидно, что от предыдущего рисунка он отличается наличием ЭФР в "классическом" виде кумулятивной кривой 2 [7]. Это позволяет сопоставить графики 1 и 2 между собой и увидеть в буквальном смысле результат спрямления кумуляты 2 согласно преобразованию (2).

Необходимость последнего объясняют тем, что сходство расположения точек ЭФР с прямой линией "на глаз" установить легче, чем при "естественной" форме кумулятивной кривой [7]. Иллюстрацией сказанного может служить рисунок 6, на котором изображены две кумулятивные кривые, построенные согласно Script 4.

SCRIPT 4:

```
Xk=sort(Xi);
n=length(Xk);
mu=mean(Xk);
sigma=std(Xk);
k=1:1:n;
Pk=(k-0.5)/n;
Fx=normcdf(Xk, mu,sigma);
```

plot(Xk,Fx,Xk,Pk); grid on

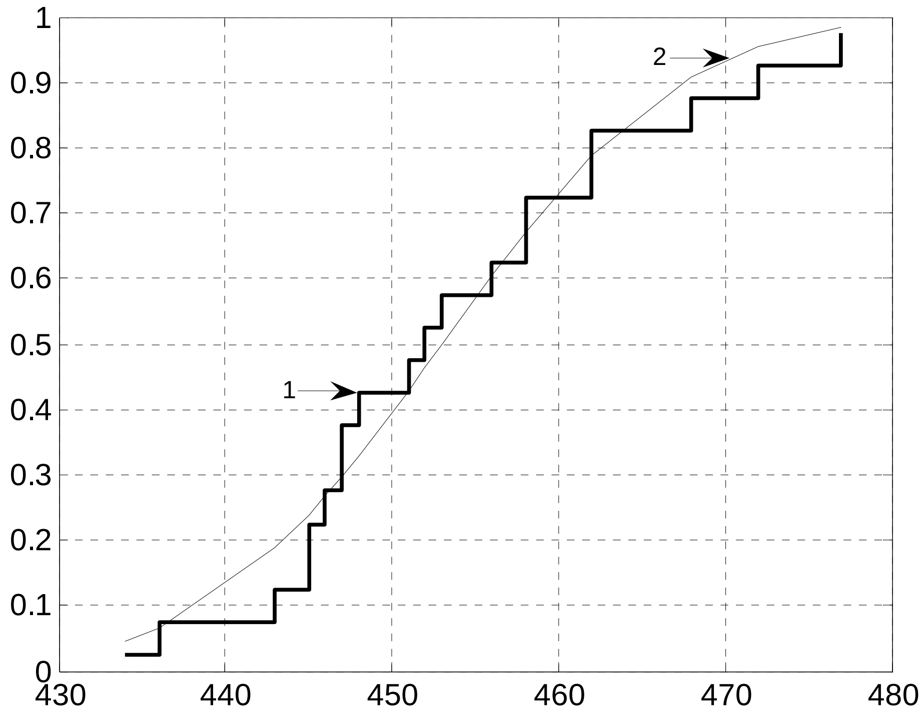

Рисунок 6 – Кумулятивная кривая

Понятно, что кривая 1 – эмпирическая функция распределения $\widetilde{\Phi}(x)$, а кривая 2 – теоретическая Φ(x), показанная также на рисунке 5.

Других известных графических средств изображения экспериментальных данных, и которые используют для проверки нормальности, ГОСТ 5479-2002 не упоминает. Между тем, построение в MATLAB полигона частот и гистограммы не вызывает никаких затруднений. Более того, *M*-функция *histfit* из библиотеки Statistics Toolbox позволяет пользователю построить гистограмму так же быстро и легко, как и нормальный график; сравни Script 1 и Script 5.

SCRIPT 5:

```
m=1+3.322*log10(n);
m=round(m);
histfit(Xk,m)
```

В первой строке последнего по формуле Стерджеса рассчитывается количество интервалов *m*, а затем во второй округляется до целого значения.

На рисунках 7 и 8 изображены гистограммы распределения случайных величин из двух тестовых примеров, рассмотренных выше. Выводы, которые следуют из анализа гистограмм, полностью совпадают с результатами анализа нормальных графиков. В учебном процессе гистограмму можно использовать для проверки выводов, сделанных с помощью нормального графика и других критериев согласия.

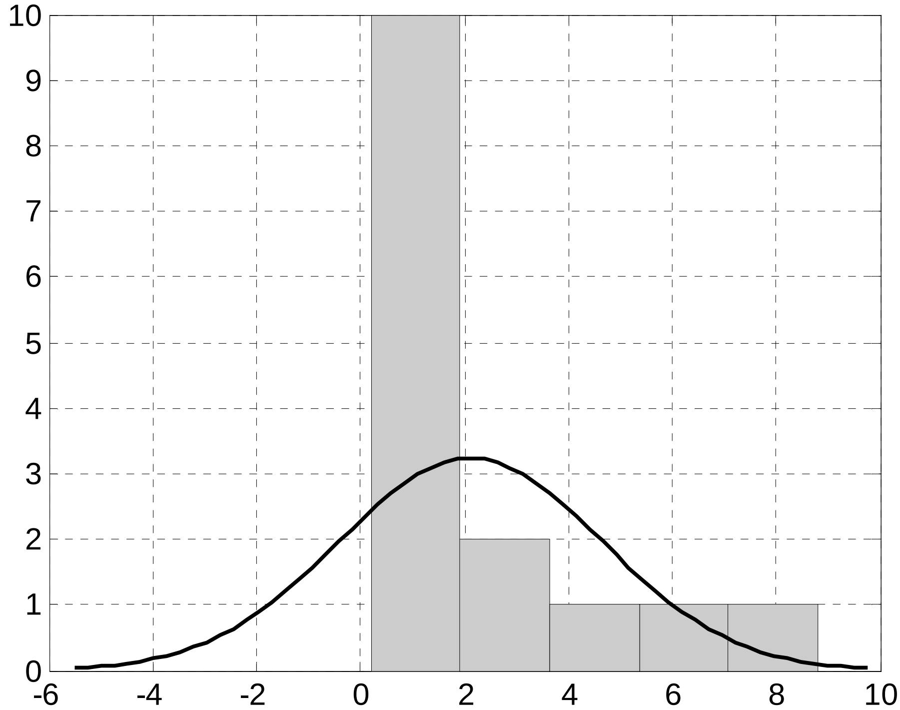

Рисунок 7 – Гистограмма

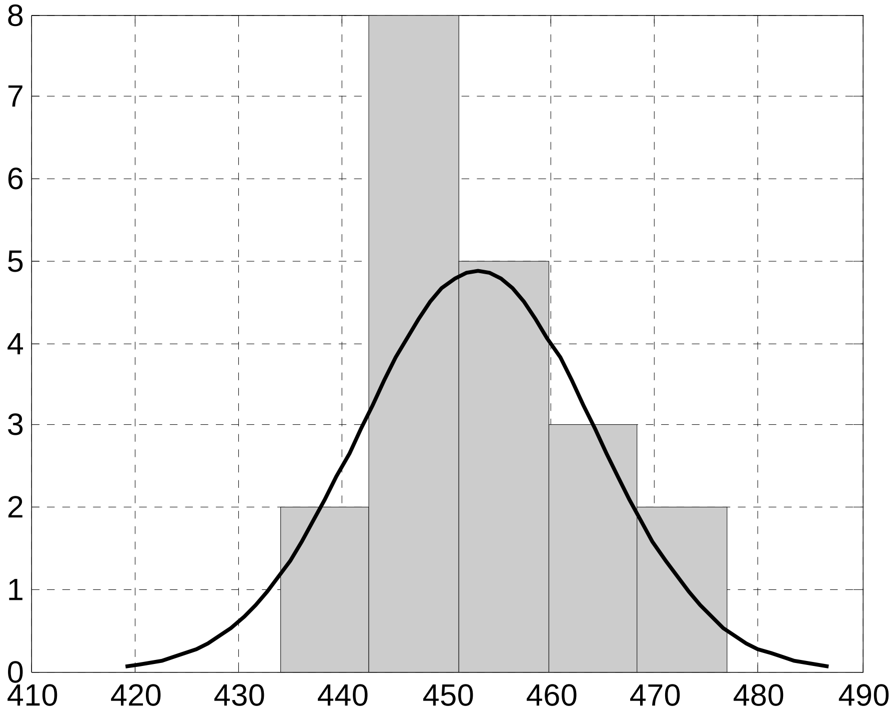

Рисунок 8 – Гистограмма

На каждую гистограмму для сравнения наложен график плотности нормального распределения (PDF). Этот прием с дидактической точки зрения раскрывает смысл понятий асимметрии и эксцесса и их связь с проверкой гипотезы нормальности распределения. В учебной практике допустима приближенная оценка отклонения от нормальности сопоставлением гистограммы с графиком PDF. Однако такая оценка всегда будет субъективной и грубой по сравнению с аналитической, позволяющей выразить количественно асимметрию (*skewness*) и эксцесс (*kurtosis*) эмпирического распределения. ГОСТ 5479-2000 предлагает направленные и многонаправленный (совместный) критерии отклонения от нормальности. Статистики критериев асимметричности и эксцесса (кривизны по терминологии ГОСТ 5479-2000) соответственно рассчитывают по следующим формулам

$$\left|\sqrt{b_1}\right| = \frac{m_3}{m_2^{3/2}}; \quad (3)$$

$$b_2 = \frac{m_4}{m_2^2}, \quad (4)$$

где m_j – выборочный центральный момент j-го порядка (j = 2, 3 и 4),

$$m_j = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^j \ . \tag{5}$$

Statistics Toolbox поддерживает расчет названных статистик с помощью *M*-функций *skewness*, *kurtosis* и *moment* [6, 10]. В дидактическом отношении последняя функция отличается тем, что требует от пользователя "программировать" вычислительный процесс согласно формулам (5), (3) и (4) и тем самым самостоятельно реализовывать на практике метод моментов [8, 10]. Напротив, функции *skewness* и *kurtosis* освобождают пользователя от какого-либо программирования задачи. Эффективность рассматриваемых функций легко проверить при решении тестовых задач, в качестве которых могут служить примеры, предлагаемые ГОСТ 5479-2000.

Так, пример 1 иллюстрирует применение критерия асимметричности, статистика которого рассчитывается по формуле (3). Соответствующие моменты второго и третьего порядков равны

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{\text{ср}})^2 \ , \tag{6}$$

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{\text{ср}})^3 \ . \tag{7}$$

При использовании *M*-функции *moment* последовательность вычисления эмпирической асимметрии, обозначенной для удобства $A = \left|\sqrt{b_1}\right|$, будет следующей

SCRIPT 6:

```
Xk=[1.25 1.35 1.40 … 5.10];
n=length(Xk);
m2=moment(Xk,2)
      m2=0.9379
m3=moment(Xk,3)
```

```
    m3=0.2546
A=m3/m2^(3/2)
    A=0.2802
```

M-функция *skewness* возвращает точно такой результат, но требует от пользователя минимальных затрат труда (см. Script 7).

SCRIPT 7:

```
A=skewness(Xk,1)
   A=0.2802
```

Как всякое выборочное значение, которое является случайной величиной, коэффициент асимметрии может отличаться от своего математического ожидания, равного нулю в случае нормального распределения. Так в рассматриваемом примере A = 0,2802. Для оценки значимости названного отклонения разработано несколько критериев согласия. Самый "экономный" из них предлагает ГОСТ 5479-2000. От пользователя стандарт не требует никаких вычислений. При этом критерий согласия выражается неравенством $A < A_{\text{кр}}$ (здесь $A_{\text{кр}}$ – критическое значение, определяемое по таблице 8 [1]). Таким образом, если указанное неравенство выполняется, то эмпирическое распределение можно считать нормальным. Для рассматриваемого примера 1 критическое значение коэффициента асимметрии $A_{\text{кр}} = 0{,}53$ (α = 0,05) больше вычисленного A = 0,2802, что позволяет не отклонять нулевую гипотезу нормального распределения. Иллюстрацией к сделанному выводу может служить рисунок 9.

Пример 2 показывает применение критерия эксцесса, статистику которого вычисляют по формуле (4). Соответственно момент второго порядка рассчитывают по формуле (6), а момент четвертого порядка – по формуле

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{\text{ср}})^4\,. \qquad (8)$$

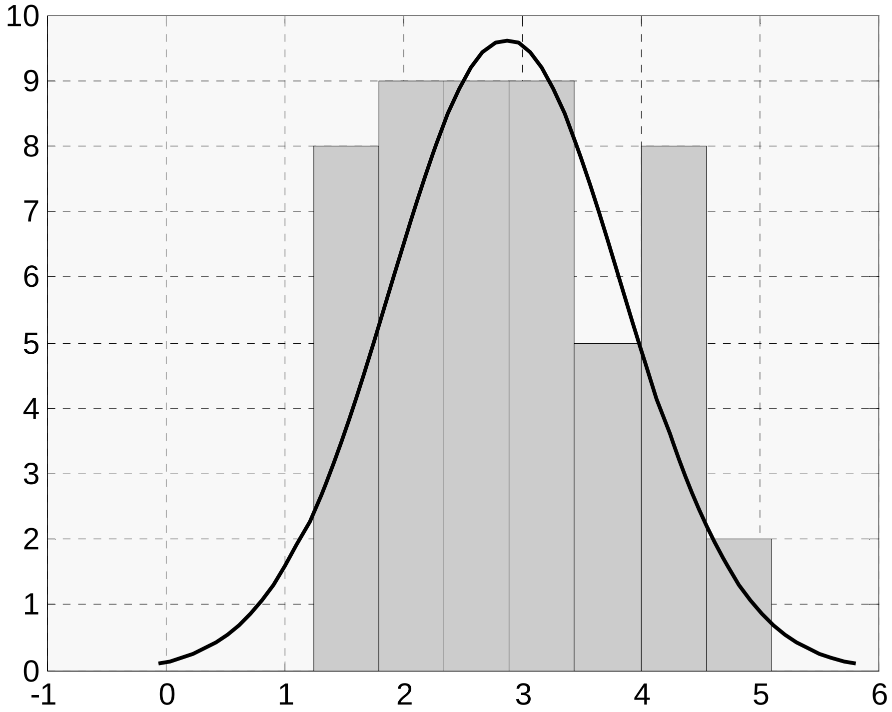

Рисунок 9 – Гистограмма (пример 1)

При обращении к *M*-функции *moment* последовательность вычисления эмпирического эксцесса, обозначенного для удобства $E = b_2$, будет следующей

SCRIPT 8:

```
Xi=[9.5  14.4  10.2 … 6.0];
Xk=sort(Xi);
n=length(Xk);
m2=moment(Xk,2)
    m2=37.9964
m4=moment(Xk,4)
    m4=7.0980e+003
E=m4/m2^2
    E=4.9165
```

M-функция *kurtosis* возвращает точно такой результат при минимальных затратах труда пользователя, см. Script 9.

SCRIPT 9:

```
E=kurtosis(Xk,1)
    E=4.9165
```

Полученное значение отличается от математического ожидания, равного 3 в случае нормального распределения. Для оценки значимости такой разницы ГОСТ 5479-2000 предлагает простейший критерий согласия, который выражается неравенством $E < E_{\text{кр}}$ (здесь $E_{\text{кр}}$ – критическое значение, определяемое по таблице 9 [1]). Эмпирическое распределение можно считать нормальным, если указанное неравенство выполняется. Для рассматриваемого примера 2 критическое значение коэффициента эксцесса $E_{\text{кр}} = 3{,}99$ (α = 0.05) меньше вычисленного E = 4,9165, поэтому нулевую гипотезу о нормальном распределении отклоняют. Рисунок 10 иллюстрирует сделанный вывод.

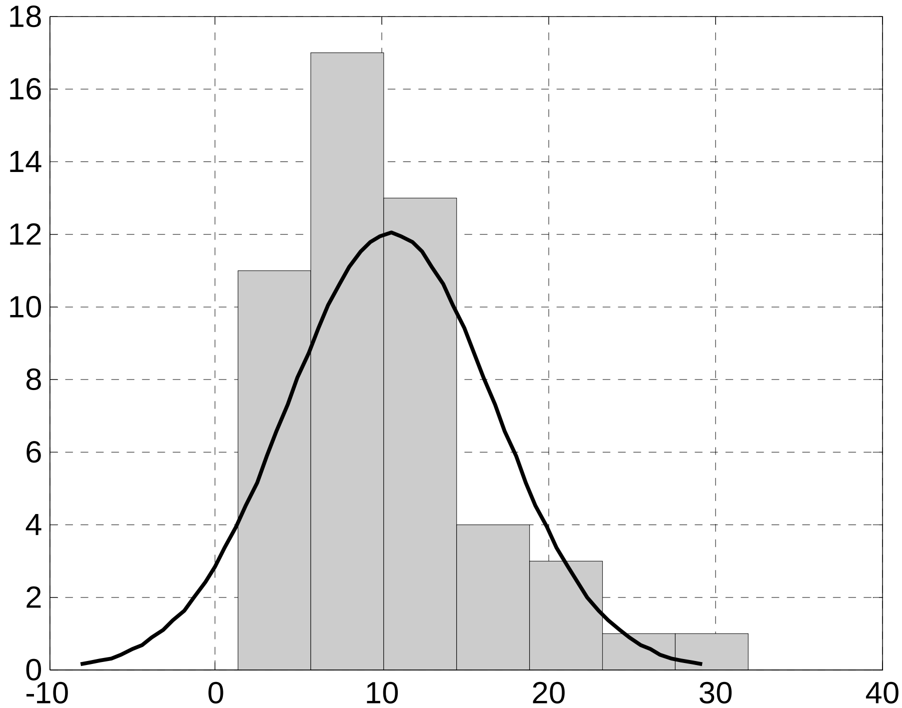

Рисунок 10 – Гистограмма (пример 2)

Рассмотренные выше порознь статистики направленных критериев $\sqrt{b_1}$ и b_2 являются одновременно статистиками совместного критерия [1]. Ясно, что при использовании последнего расчет названных статистик следует выполнять согласно Script 6 и Script 8 или Script 7 и Script 9.

В отличие от критериев асимметрии и эксцесса критерии Шапиро-Уилка (W-критерий) и критерий Эппса-Палли (T_{EP}-критерий) не поддерживаются пакетом Statistics Toolbox. Расчет статистик названных критериев возможен непосредственно в MATLAB, разнообразие встроенных функций которого допускает несколько вариантов автоматизации расчетов.

ГОСТ 5479-2000 представляет W-статистику критерия Шапиро-Уилка в следующем виде

$$W = \frac{S^2}{nm_2}, \tag{9}$$

где n – размер вариационного ряда; m_2 – выборочный центральный момент второго порядка, см. (6).

Сумму S вычисляют по формуле

$$S = \sum_{k=1}^{n/2} [a_k (x_{n+1-k} - x_k)],$$

где k – индекс, принимающий значения 1, 2,…, $n/2$ при четном n и 1, 2,…, $(n-2)/2$ при нечетном n; a_k – коэффициенты, значения которых определяют "вручную" по таблице 10 [1] в соответствии с n и k.

Попытка автоматизации расчета статистики W-критерия по формуле (9) сталкивается с проблемой определения коэффициентов a_k. Эту задачу традиционно решают с помощью аппроксимирующих зависимостей, заменяющих собой таблицы. Наиболее известна аппроксимация Казакявичуса [11]. В этом случае используют упрощенный критерий Шапиро-Уилка, статистику которого вычисляют по формуле

$$W_1 = \left(1 - \frac{0{,}6695}{n^{0{,}6518}}\right) \cdot \frac{S^2}{B}. \tag{10}$$

Если $W_1 > 1$, то нулевую гипотезу нормального распределения случайной величины x отклоняют на уровне значимости α = 0,05. Величины S^2 и B представляют собой суммы следующего вида

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - X_{\text{ср}})^2;$$

$$B = \frac{1}{n-1}\left[\sum_{k=1}^{m} a_k (X_{n-k+1} - X_k)\right]^2,$$

где X_k – элементы вариационного ряда; $X_{\text{ср}}$ – среднее арифметическое значение; X_{n-k+1} – элементы вариационного ряда, расположенные в порядке их убывания, на что и указывает индекс $n - k + 1$.

Коэффициенты a_0, a_k и новую переменную z_k (k = 1, 2,…, m = n/2) вычисляют по аппроксимированным зависимостям [11]

$$a_0 = \frac{0{,}899}{(n-2{,}4)^{0{,}4162}} - 0{,}02;$$

$$a_k = a_0\left[z_k + \frac{1483}{(3-z_k)^{10{,}845}} + \frac{71{,}61\cdot 10^{-10}}{(1.1 - z_k)^{8{,}26}}\right];$$

$$z_k = \frac{n-2k+1}{n-0{,}5}.$$

Таким образом, искомой является статистика W_1, а все названные величины S^2, B, a_0, a_k и переменные X_{n-k+1} и z_k являются промежуточными. Они могут потребоваться только для отладки ПО и сверки результатов вычислений в MATLAB с данными примера 4 [1], используемого далее в качестве тестового.

Как видно, алгоритм расчета W_1 предельно прост. Его реализация в MATLAB не намного сложнее. Несколько усложняет всю процедуру только получение вектора X_{n-k+1}. Его образуют элементы вариационного ряда, то есть элементы вектора X_k, но расположенные в обратном порядке по сравнению с последним. Об этом уже было сказано выше. Это значит, что вектор X_{n-k+1} является зеркальным отражением вектора X_k. На практике в MATLAB такую операцию над векторами осуществляют с помощью функции *fliplr* [12-19], что позволяет исключить более сложную переиндексацию векторов. Вдобавок к этому целесообразно переименовать вектор X_{n-k+1}, заменив сложное обозначение более простым *X*. Заметим также, что сумма *B* содержит *m* слагаемых. Поэтому такой же размерностью характеризуются рассмотренные векторы $X(1:m)$ и $X_k(1:m)$.

В итоге, с учетом принятых обозначений для вычисления приближенного значения статистики Шапиро-Уилка W_1 в режиме командной строки MATLAB достаточно выполнить следующие действия согласно SCRIPT 10. В качестве тестового служит пример 4 из ГОСТ 5479-2000.

SCRIPT 10:

```
Xk=[520 556 561 … 963 1056 1074];
n=length(Xk);
m=n/2;
Xcp=mean(Xk);
S2=sum((Xk-Xcp).^2);
X=fliplr(Xk);
k=1:m;
zk=(n-2.*k+1)/(n-0.5);
a0=0.899/(n-2.4)^0.4162-0.02;
ak=a0.*(zk+1483./(3-zk).^10.845+71.6*10^-10./(1.1-zk).^8.26);
B=sum(ak.*(X(1:m)-Xk(1:m))).^2;
W1=(1-0.6695/n^0.6518)*S2/B
W1=
    0.9587
```

Следовательно, гипотеза нормального распределения не отклоняется, т.к. $W_1 < 1$. Этот вывод подтверждается также наглядно гистограммой и нормальным графиком, изображенными соответственно на рисунках 11 и 12.

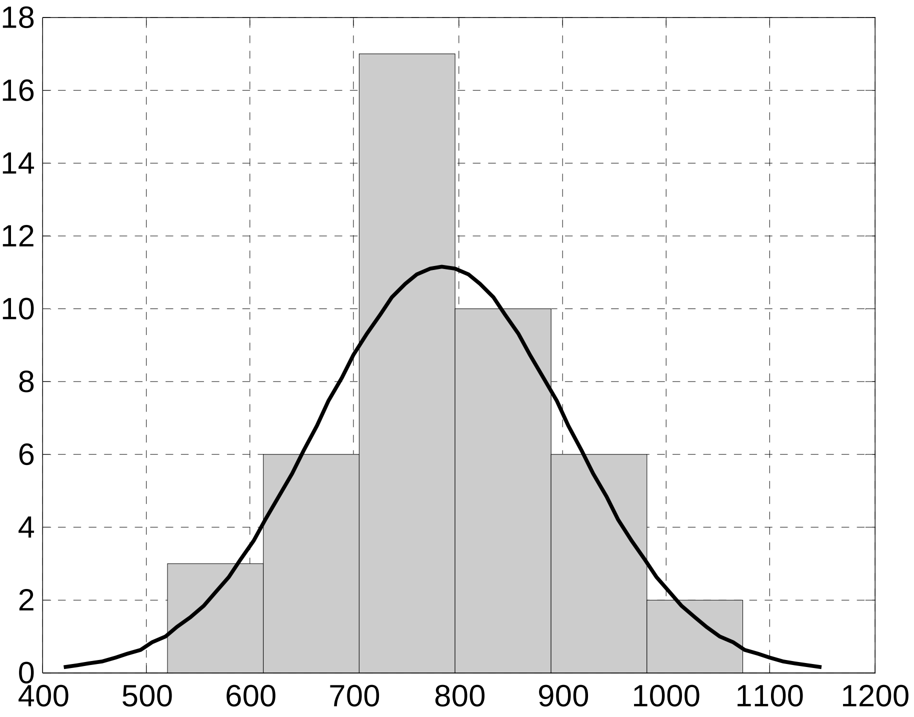

Рисунок 11 – Гистограмма (пример 4 [1])

Последний из критериев, рекомендованных ГОСТ 5479-2002, – критерий Эппса-Палли – с вычислительной точки зрения является самым сложным. Статистику этого критерия рассчитывают по формуле

$$T_{EP} = 1 + \frac{n}{\sqrt{3}} + \frac{2}{n}\sum_{k=2}^{n}\sum_{j=1}^{k-1}\exp\left\{\frac{-(X_j - X_k)^2}{2m_2}\right\} - \sqrt{2}\sum_{j=1}^{n}\exp\left\{\frac{-(X_j - X_{\text{ср}})^2}{4m_2}\right\},$$

применение которой показано на примере 6 [1]. Далее эта задача рассматривается как тестовая с сохранением оригинальных обозначений. По условию задачи необходимо проверить гипотезу нормального распределения случайной величины *x*, представленной вариационным рядом $X_j =$ 4,9; 5,0; 6,5; 10,9; 11,0; 11,4; 12,7; 13,1; 14,0; 14,5. В стандарте предложено T_{EP} вычислять по частям, обозначив слагаемые следующим образом

$$A = \sqrt{2}\sum_{j=1}^{n} \exp\left\{\frac{-(X_j - X_{\text{ср}})^2}{4m_2}\right\};$$

$$B = \frac{2}{n}\sum_{k=2}^{n}\sum_{j=1}^{k-1} \exp\left\{\frac{-(X_j - X_k)^2}{2m_2}\right\}.$$

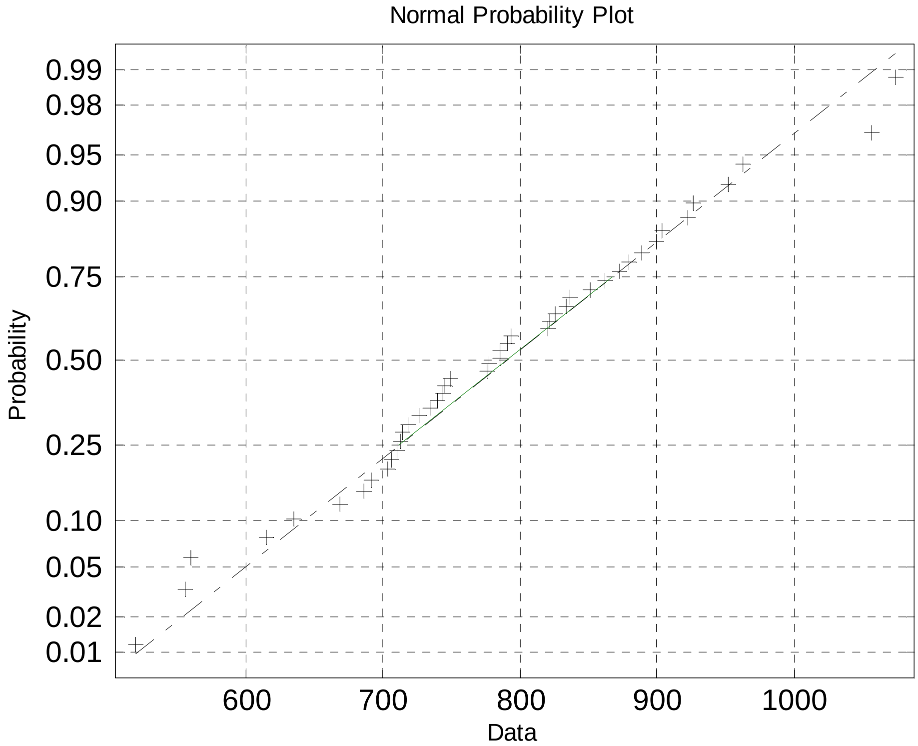

Рисунок 12 – Нормальный график (пример 4 [1])

Процедура расчета *A* настолько проста, что не требует никаких пояснений. Соответственно реализующий ее SCRIPT 11 приведен ниже без комментариев.

SCRIPT 11:

```
Xj=[4.9 5.0 6.5 10.9 11.0 11.4 12.7 13.1 14.0 14.5];
Xcp=mean(Xj);
n=length(Xj);
m2=sum((Xj-Xcp).^2)/n;
A=sqrt(2)*sum(exp(-((Xj-Xcp).^2)/(4*m2)));
A =
        11.2794
```

Вычисление B, представляющего собой "сумму в сумме", может быть произведено двумя способами. Первый вариант заключается в программировании задачи с помощью команд цикла for…end [17, 20-22]. Второй вариант использует функции матричного исчисления и по вычислительным возможностям предпочтительнее первого [16, 18, 19, 23]. Предлагается вначале вычислить разности $X_j - X_k$, совокупность которых образует матрицу D. В рассматриваемом примере она имеет следующий вид

j \ k	2	3	4	5	6	7	8	9	10
1	$x_1 - x_2$	$x_1 - x_3$	$x_1 - x_4$	$x_1 - x_5$	$x_1 - x_6$	$x_1 - x_7$	$x_1 - x_8$	$x_1 - x_9$	$x_1 - x_{10}$
2	–	$x_2 - x_3$	$x_2 - x_4$	$x_2 - x_5$	$x_2 - x_6$	$x_2 - x_7$	$x_2 - x_8$	$x_2 - x_9$	$x_2 - x_{10}$
3	–	–	$x_3 - x_4$	$x_3 - x_5$	$x_3 - x_6$	$x_3 - x_7$	$x_3 - x_8$	$x_3 - x_9$	$x_3 - x_{10}$
4	–	–	–	$x_4 - x_5$	$x_4 - x_6$	$x_4 - x_7$	$x_4 - x_8$	$x_4 - x_9$	$x_4 - x_{10}$
5	–	–	–	–	$x_5 - x_6$	$x_5 - x_7$	$x_5 - x_8$	$x_5 - x_9$	$x_5 - x_{10}$
6	–	–	–	–	–	$x_6 - x_7$	$x_6 - x_8$	$x_6 - x_9$	$x_6 - x_{10}$
7	–	–	–	–	–	–	$x_7 - x_8$	$x_7 - x_9$	$x_7 - x_{10}$
8	–	–	–	–	–	–	–	$x_8 - x_9$	$x_8 - x_{10}$
9	–	–	–	–	–	–	–	–	$x_9 - x_{10}$

Поскольку $k > j$, матрица D всегда будет верхней треугольной. Очевидно, что к этому типу принадлежат матрицы X_j и X_k. При этом индексы принимают значения j = 1, 2,…, k - 1 и k = 2, 3,…, n. Это значит, что элементами матрицы X_j являются первые девять элементов вектора X_j в Script 11. Для удобства работы в MATLAB их можно представить вектором X1 так, что X1=Xj (1:n – 1). Аналогично X2=Xj (2:n), что делает понятным получение X_k. SCRIPT 12 и SCRIPT 13 показывают последовательность формирования матриц X_j и X_k.

SCRIPT 12:

```
X1=Xi(1:n-1);
M1=diag(X1);
E=ones(n-1,n-1);
Xj1=M1*E;
Xj=triu(Xj1);
```

SCRIPT 13:

```
X2=Xi(2:n);
M2=diag(X2);
Xk1=E*M2;
Xk=triu(Xk1);
```

Далее согласно SCRIPT 14 вычисляются элементы матрицы $D = X_j - X_k$, возводятся во вторую степень и т.д. Результатом является квадратная матрица $B1$, которая затем при помощи функции *triu* приводится к треугольной $B2$. После этого рассчитываются внутренняя S_j и внешняя (общая) S_k суммы. Первая есть сумма элементов каждого из девяти столбцов матрицы 2, а последняя равна их сумме $S_k = 23{,}9865$. В завершение находят слагаемое B = 4,7973.

SCRIPT 14:

```
D=Xj-Xk;
B1=exp(-D.^2/(2*m2));
B2=triu(B1);
Sj=sum(B2)
Sj =
0.9996 1.80972 0.8916 1.8528 2.6923 3.0455 3.8952 4.1573 4.7350
Sk=sum(Sj)
```

Sk =

23.9865

B=2/n*Sk

B =

4.7973

Остается только найти статистику критерия Эппса-Палли. Согласно SCRIPT 15 искомое значение равно $T_{EP} = 0{,}2914$ и в точности совпадает со значением, указанным в ГОСТ 5479-2002.

SCRIPT 15:

```
Tep=1+n/sqrt(3)+B-A
Tep=
        0.2914
```

Следовательно, гипотеза нормального распределения не отклоняется, т.к. $T_{EP} = 0{,}2914$ меньше критического значения $T_{KP} = 0{,}357$ на уровне значимости α = 0,05 [1]. Этот вывод подтверждается также наглядно гистограммой и нормальным графиком, изображенными соответственно на рисунках 13 и 14.

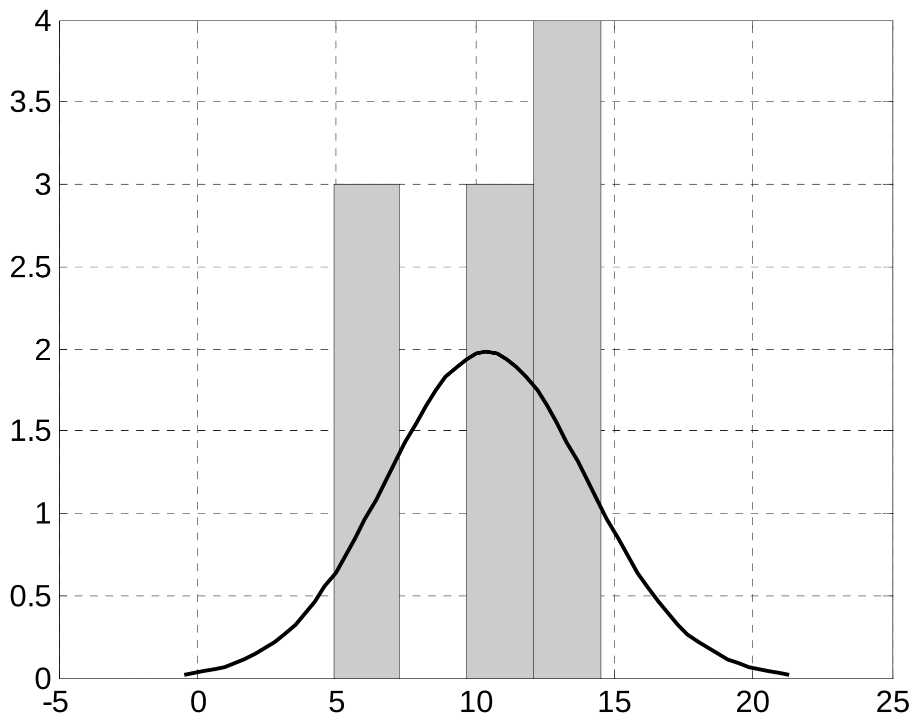

Рисунок 13 – Гистограмма (пример 6 [1])

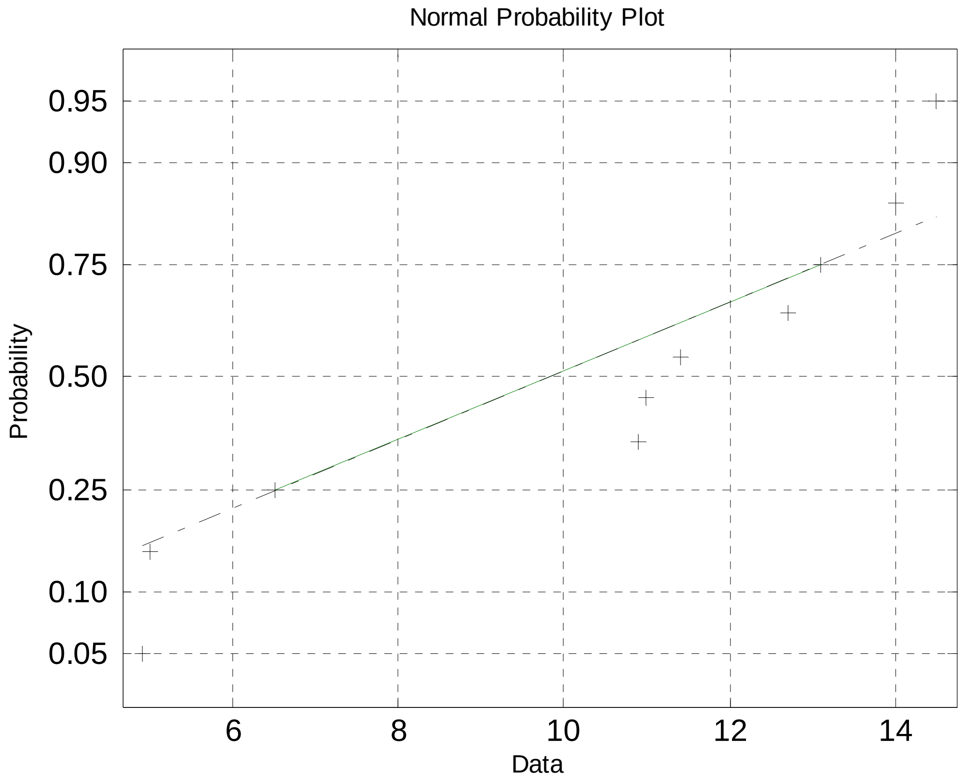

Рисунок 12 – Нормальный график (пример 6 [1])

В итоге, исследователю для расчета статистики критерия Эппса-Палли достаточно только скопировать на свой ПК SCRIPT 11-15, отредактировать в SCRIPT 11 первую строку, заменив выборку тестовой задачи реальными данными, и нажать клавишу Enter.

Таким образом, сформулированная в начале статьи цель достигнута. С помощью функций Statistics Toolbox и MATLAB автоматизированы предложенные ГОСТ 5479-2002 графические и аналитические методы проверки гипотезы нормальности распределения вероятностей эмпирических данных. Вместе с тем, в задачи исследования не входил сравнительный анализ названных методов и критериев согласия. Однако примеры, избранные в качестве тестовых задач, позволяют пользователю MATLAB сделать это самостоятельно.

ЛИТЕРАТУРА

1. ГОСТ Р ИСО 5479-2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения. – М.: ИПК Изд-во стандартов. – 31 с.

2. Нестеров А.В. Теория автоматического управления: учеб. пособие для студентов вузов / А.В. Нестеров, С.В. Нестеров; Куб. гос. технол. ун-т. Краснодар, 2006. – 191 с.

3. Нестеров А.В., Нестеров С.В. Экспериментальное исследование акустической характеристики свободно выбегающего ротора / Межвуз. сб. науч. ст.: Машиностроение. – Краснодар: Издательский дом Юг, 2011. – Вып. 4. С. 75 – 81.

4. Нестеров А.В., Нестеров С.В., Михайлов А.В. Определение постоянной времени механической части электропривода // Современное машиностроение. Наука и образование: Матер. 3-й Междунар. науч.-практ. конф. / Под ред. М.М. Радкевича и А.Н. Евграфова. – СПб.: Изд-во Политехн. ун-та, 2013. – С. 797-803.

5. Нестеров А.В., Нестеров С.В., Манаков К.М. Об измерительной схеме модифицированного метода свободного выбега // Измерения в современном мире-2013: сб. науч. тр. 4-й междунар. науч.-практ. конф. – СПб.: Изд-во Политехн. ун-та, 2013. – С. 54-57.

6. Сергеев А.Г., Латышев М.В., Мищенко З.В. Математическое моделирование задач метрологии и сертификации в MATLAB. – Владимир: Изд-во Владимир. Гос. ун-та, 2003. – 314 с.

7. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – 3-е изд., перераб. и доп.: М.: ИНФРА-М, 2003. – 544 с.

8. Хан Г., Шапиро С. Статистические модели в инженерных задачах. – М.: Мир, 1969. – 395 с.

9. Степнов М.Н. Статистические методы обработки результатов механических испытаний: Справочник: 2-е изд., испр. и доп. – М.: Машиностроение, 2005. – 400 с.

10. Иглин С.П. Теория вероятностей и математическая статистика. – Харьков: НТУ "ХПИ", 2006. – 612 с.

11. Казакявичус К.А. Приближенные формулы для статистической обработки результатов механических испытаний // Заводская лаборатория. 1988. – Т. 54., № 12. – С. 82-85.

12. Нестеров А.В., Нестеров С.В. Об информационном аспекте применения системы MATLAB для расчета типовых статистик // Информатика: проблемы, методология, технологии: Матер. XIII Междунар. науч.-метод. конф. Т. 2. – Воронеж: ИПЦ Воронеж. гос. ун-та, 2013. – С. 427-430.

13. Нестеров А.В., Нестеров С.В. Проверка нормальности распределения эмпирических данных по критерию Шапиро-Уилка в Matlab // Инновации в технологиях и образовании: сб. ст. участников VI Междунар. науч.-практ. конф. «Инновации в технологиях и образовании»: в 4 частях. – Ч. 2. – Белово: Изд-во филиала КузГТУ в г. Белово, Россия; Изд-во ун-та «Св. Кирилла и Св. Мефодия», Велико Тырново, Болгария, 2013. – С. 211-214.

14. Нестеров А.В., Нестеров С.В. О вычислительных особенностях оценки качества продукции хлебопекарного и кондитерского производства статистическими методами в системе компьютерной математики MATLAB // Хлебобулочные, кондитерские и макаронные изделия XXI века: Матер. III Междунар. науч.-практ. конф. – Краснодар: Изд. ФГБОУ ВПО "КубГТУ", 2013. – С. 124-128.

15. Нестеров А.В., Нестеров С.В. Статистический анализ случайных отклонений технологических параметров бурения нефтяных и газовых скважин \\ Нефть и газ Западной Сибири: матер. междунар. науч.-техн. конф., посвященной 50-летию Тюменского индустриального ин-та Т. 2. – Тюмень: ТюмГНГУ, 2013. – С. 28-31.

16. Нестеров А.В., Нестеров С.В., Нестерова Д.А. Проверка отклонения распределения вероятностей от нормального распределения по критерию Шапиро-Уилка в MATLAB // Некоторые вопросы математики и ее приложений: Сб. науч. тр. / под ред. Е.М. Малеко. Вып. 1. – Магнитогорск: Магнитогорск: Изд-во Магнитогорск. гос. техн. ун-та им. Г.И. Носова, 2013. – С. 41-45.

17. Нестеров А.В., Нестеров С.В., Нестерова Д.А. К оценке качества продукции статистическими методами в системе компьютерной математики MATLAB // Управление качеством: матер. Междунар. науч.-практ. конф. – М.: Изд. ун-та Машиностроения, 2013. – С. 51-56.

18. Нестеров А.В., Нестеров С.В. О вычислительных особенностях статистического анализа данных в системе MATLAB при контроле качества продуктов питания // Инновации в индустрии питания и сервисе: Электронный сб. матер. I Междунар. науч.-практ. конф., посвященной 30-летию кафедры технологии и организации питания. – Краснодар: Изд. КубГТУ, 2014. – С. 188-190.

19. Козак Д.А., Нестеров А.В., Нестеров С.В. Об использовании статистических методов оценки безопасности пищевых продуктов и сырья на предприятиях агропромышленного комплекса // Инновационные технологии в сфере питания, сервиса и торговли: Матер. очно-заочной науч.-практ. конф. – Екатеринбург: Изд-во Урал. гос. экон. ун-та, 2014. – С. 192-195.

20. Нестеров А.В., Нестеров С.В. О статистическом анализе эмпирических данных в системе Matlab при контроле показателей качества товарной нефти // Интеграция науки и образования в вузах нефтегазового профиля – 2014: матер. Междунар. науч.-методич. конф. / редкол.: Н.Г. Евдокимова и др. – Уфа: РИЦ УГНТУ, 2014. – С 185-187.

21. Нестеров А.В., Нестеров С.В. Об особенностях статистической оценки безопасности пищевых продуктов средствами MATLAB: Современные проблемы здорового питания. Инновации и традиции: сб. стат. и докл. междунар. науч.-практ. конф. – Барнаул: Алт. гос. техн. ун-т им. И.И. Ползунова, 2014. – С. 12-14.

22. Нестеров А.В., Нестеров С.В. О статистическом анализе данных в системе MATLAB при оценке безопасности мясных и рыбных продуктов // Современные проблемы качества и безопасности продуктов питания в свете требований технического регламента таможенного союза: сборник матер.

междунар. науч.-практ. интернет-конф. – Краснодар: изд. КубГТУ, 2014. – С. 133-135.

23. Нестеров А.В., Нестеров С.В., Нестерова Д.А. К статистической оценке качества пищевых продуктов // Перспективные технологии производства продукции из сырья животного и растительного происхождения: Матер. междунар. науч.-техн. Интернет-конф. – Краснодар: ФГБОУ ВПО "КубГТУ", 2013. – С. 116-119.

REFERENCES

1. GOST R ISO 5479-2002. Statisticheskie metody. Proverka otklonenija raspredelenija verojatnostej ot normal'nogo raspredelenija. – M.: IPK Izd-vo standartov. – 31 s.

2. Nesterov A.V., Nesterov S.V. Teorija avtomaticheskogo upravlenija: ucheb. posobie dlja studentov vuzov / Kub. gos. tehnol. un-t. Krasnodar, 2006. – 191 s.

3. Nesterov A.V., Nesterov S.V. Jeksperimental'noe issledovanie akusticheskoj harakteristiki svobodno vybegajushhego rotora / Mezhvuz. sb. nauch. st.: Mashinostroenie. – Krasnodar: Izdatel'skij dom Jug, 2011. – Vyp. 4. S. 75 – 81.

4. Nesterov A.V., Nesterov S.V., Mihajlov A.V. Opredelenie postojannoj vremeni mehanicheskoj chasti jelektroprivoda // Sovremennoe mashinostroenie. Nauka i obrazovanie: Mater. 3-j Mezhdunar. nauch.-prakt. konf. / Pod red. M.M. Radkevicha i A.N. Evgrafova. – SPb.: Izd-vo Politehn. un-ta, 2013. – S. 797-803.

5. Nesterov A.V., Nesterov S.V., Manakov K.M. Ob izmeritel'noj sheme modificirovannogo metoda svobodnogo vybega // Izmerenija v sovremennom mire-2013: sb. nauch. tr. 4-j mezhdunar. nauch.-prakt. konf. – SPb.: Izd-vo Politehn. un-ta, 2013. – S. 54-57.

6. Sergeev A.G., Latyshev M.V., Mishhenko Z.V. Matematicheskoe modelirovanie zadach metrologii i sertifikacii v MATLAB. – Vladimir: Izd-vo Vladimir. Gos. un-ta, 2003. – 314 s.

7. Tjurin Ju.N., Makarov A.A. Analiz dannyh na komp'jutere. – 3-e izd., pererab. i dop.: M.: INFRA-M, 2003. – 544 s.

8. Han G., Shapiro S. Statisticheskie modeli v inzhenernyh zadachah. – M.: Mir, 1969. – 395 s.

9. Stepnov M.N. Statisticheskie metody obrabotki rezul'tatov mehanicheskih ispytanij: Spravochnik: 2-e izd., ispr. i dop. – M.: Mashinostroenie, 2005. – 400 s.

10. Iglin S.P. Teorija verojatnostej i matematicheskaja statistika. – Har'kov: NTU "HPI", 2006. – 612 s.

11. Kazakjavichus K.A. Priblizhennye formuly dlja statisticheskoj obrabotki rezul'tatov mehanicheskih ispytanij // Zavodskaja laboratorija. 1988. – T. 54., № 12. – S. 82-85.

12. Nesterov A.V., Nesterov S.V. Ob informacionnom aspekte primenenija sistemy MATLAB dlja rascheta tipovyh statistik // Informatika: problemy, metodologija, tehnologii: Mater. XIII Mezhdunar. nauch.-metod. konf. T. 2. – Voronezh: IPC Voronezh. gos. un-ta, 2013. – S. 427-430.

13. Nesterov A.V., Nesterov S.V. Proverka normal'nosti raspredelenija jempiricheskih dannyh po kriteriju Shapiro-Uilka v Matlab // Innovacii v tehnologijah i obrazovanii: sb. st. uchastnikov VI Mezhdunar. nauch.-prakt. konf. «Innovacii v tehnologijah i obrazovanii»: v 4 chastjah. – Ch. 2. – Belovo: Izd-vo filiala KuzGTU v g. Belovo, Rossija; Izd-vo un-ta «Sv. Kirilla i Sv. Mefodija», Veliko Tyrnovo, Bolgarija, 2013. – S. 211-214.

14. Nesterov A.V., Nesterov S.V. O vychislitel'nyh osobennostjah ocenki kachestva produkcii hlebopekarnogo i konditerskogo proizvodstva statisticheskimi metodami v sisteme komp'juternoj matematiki MATLAB // Hlebobulochnye, konditerskie i makaronnye izdelija XXI veka: Mater. III Mezhdunar. nauch.-prakt. konf. – Krasnodar: Izd. FGBOU VPO "KubGTU", 2013. – S. 124-128.

15. Nesterov A.V., Nesterov S.V. Statisticheskij analiz sluchajnyh otklonenij tehnologicheskih parametrov burenija neftjanyh i gazovyh skvazhin \\ Neft' i gaz Zapadnoj Sibiri: mater. mezhdunar. nauch.-tehn. konf., posvjashhennoj 50-letiju Tjumenskogo industrial'nogo in-ta T. 2. – Tjumen': TjumGNGU, 2013. – S. 28-31.

16. Nesterov A.V., Nesterov S.V., Nesterova D.A. Proverka otklonenija raspredelenija verojatnostej ot normal'nogo raspredelenija po kriteriju Shapiro-Uilka

v MATLAB // Nekotorye voprosy matematiki i ee prilozhenij: Sb. nauch. tr. / pod red. E.M. Maleko. Vyp. 1. – Magnitogorsk: Magnitogorsk: Izd-vo Magnitogorsk. gos. tehn. un-ta im. G.I. Nosova, 2013. – S. 41-45.

17. Nesterov A.V., Nesterov S.V., Nesterova D.A. K ocenke kachestva produkcii statisticheskimi metodami v sisteme komp'juternoj matematiki MATLAB // Upravlenie kachestvom: mater. Mezhdunar. nauch.-prakt. konf. – M.: Izd. un-ta Mashinostroenija, 2013. – S. 51-56.

18. Nesterov A.V., Nesterov S.V. O vychislitel'nyh osobennostjah statisticheskogo analiza dannyh v sisteme MATLAB pri kontrole kachestva produktov pitanija // Innovacii v industrii pitanija i servise: Jelektronnyj sb. mater. I Mezhdunar. nauch.-prakt. konf., posvjashhennoj 30-letiju kafedry tehnologii i organizacii pitanija. – Krasnodar: Izd. KubGTU, 2014. – S. 188-190.

19. Kozak D.A., Nesterov A.V., Nesterov S.V. Ob ispol'zovanii statisticheskih metodov ocenki bezopasnosti pishhevyh produktov i syr'ja na predprijatijah agropromyshlennogo kompleksa // Innovacionnye tehnologii v sfere pitanija, servisa i torgovli: Mater. ochno-zaochnoj nauch.-prakt. konf. – Ekaterinburg: Izd-vo Ural. gos. jekon. un-ta, 2014. – S. 192-195.

20. Nesterov A.V., Nesterov S.V. O statisticheskom analize jempiricheskih dannyh v sisteme Matlab pri kontrole pokazatelej kachestva tovarnoj nefti // Integracija nauki i obrazovanija v vuzah neftegazovogo profilja – 2014: mater. Mezhdunar. nauch.-metodich. konf. / redkol.: N.G. Evdokimova i dr. – Ufa: RIC UGNTU, 2014. – C 185-187.

21. Nesterov A.V., Nesterov S.V. Ob osobennostjah statisticheskoj ocenki bezopasnosti pishhevyh produktov sredstvami MATLAB: Sovremennye problemy zdorovogo pitanija. Innovacii i tradicii: sb. stat. i dokl. mezhdunar. nauch.-prakt. konf. – Barnaul: Alt. gos. tehn. un-t im. I.I. Polzunova, 2014. – S. 12-14.

22. Nesterov A.V., Nesterov S.V. O statisticheskom analize dannyh v sisteme MATLAB pri ocenke bezopasnosti mjasnyh i rybnyh produktov // Sovremennye problemy kachestva i bezopasnosti produktov pitanija v svete trebovanij

tehnicheskogo reglamenta tamozhennogo sojuza: sbornik mater. mezhdunar. nauch.-prakt. internet-konf. – Krasnodar: izd. KubGTU, 2014. – S. 133-135.

23. Nesterov A.V., Nesterov S.V., Nesterova D.A. K statisticheskoj ocenke kachestva pishhevyh produktov // Perspektivnye tehnologii proizvodstva produkcii iz syr'ja zhivotnogo i rastitel'nogo proishozhdenija: Mater. mezhdunar. nauch.-tehn. Internet-konf. – Krasnodar: FGBOU VPO "KubGTU", 2013. – S. 116-119.

THE CHECKING OF A DEVIATION TO DISTRIBUTE PROBABILITIES FROM NORMAL DISTRIBUTION IN MATLAB SYSTEM

A.V. NESTEROV, S.V. NESTEROV, D.A. KOZAK

Kuban State Technological University,
2, Moskovskaya st., Krasnodar, Russian Federation, 350072;
e-mail: briefkasten129@rambler.ru

Possibility to check a deviation of distribution of probabilities from normal distribution in accordance with GOST 5479-2002 by means of MATLAB system and its package of the Statistics Toolbox expansion is investigated. The special attention is paid to automation to creation statistical schedules and calculation the statistician of consent criteria. Operability of the offered procedures is confirmed during the solution of test tasks.

Key words: normal probability plot, histogram, criteria of asymmetry and excess, Shapiro-Wilk test, Epps-Pally test.