

Министерство образования и науки Российской Федерации  
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ

**Компьютерные технологии анализа  
данных и исследования статистических  
закономерностей:  
проверка гипотезы о виде распределения**

Методические указания  
к выполнению курсовых проектов  
для студентов V-го курса ФПМИ  
по направлению 010400.68  
дневного отделения

Новосибирск, 2013

Методические указания предназначены для студентов, выполняющих курсовые проекты по курсу "Компьютерные технологии анализа данных и исследования статистических закономерностей" в первом семестре (направление 010400.68 – Прикладная математика и информатика, магистерская программа – *Математическое и программное обеспечение информационных технологий моделирования и анализа данных*). Указания содержат необходимые сведения для выполнения курсового проекта, порядок выполнения, структуру оформления пояснительной записки и примерное содержание её разделов, варианты заданий.

Составители: доктор техн. наук, проф. *Б.Ю. Лемешко*,  
канд. техн. наук, доц. *С.Н. Постовалов*,  
канд. техн. наук, доц. *Е.В. Чимитова*

Работа подготовлена на кафедре  
прикладной математики

## Тема «Проверка гипотезы о виде распределения».

### 1. Цель выполнения работы

Знакомство с современными тенденциями развития аппарата прикладной математической статистики и состоянием программного обеспечения задач статистического анализа. Освоение методов статистического моделирования как средства исследования и развития аппарата прикладной математической статистики. Исследование особенностей методов проверки статистических гипотез. Закрепление навыков проведения самостоятельных исследований.

Курсовой проект (КП) носит исследовательский характер. Варианты заданий могут быть связаны с известными методами или критериями, изложение которых в литературных источниках не позволяет однозначно охарактеризовать их свойства по сравнению с аналогами, либо остается неясной картина корректности статистических выводов в случае нарушения стандартных предположений. Задание может быть связано с исследованием свойств оценок или критериев в нестандартных условиях, связанных с конкретной областью приложения, либо в заданных нестандартных условиях.

Предпочтение отдается вариантам заданий, ориентированным на развитие аппарата прикладной математической статистики и создание программного обеспечения задач статистического анализа.

В процессе выполнения КП необходимо ознакомиться с историческими аспектами, состоянием и тенденциями развития в соответствующем разделе прикладной математической статистики, предпосылками, определившими интерес и потребности к исследованиям в данном направлении.

Охарактеризовать место и роль исследуемого метода или критерия в развитии прикладных методов статистического анализа, актуальность проводимых исследований. Обязательно наличие библиографического обзора работ, связанных со становлением и развитием соответствующего метода.

В процессе выполнения КП должны быть описаны предпосылки (условия), выполнение которых обеспечивает корректность статистических выводов при использовании данного метода или критерия.

Выполнение экспериментальной части КП может предусматривать:

- самостоятельную разработку некоторого программного обеспечения;
- развитие существующего программного обеспечения;
- использование различных математических пакетов и программных систем, например, для исследования реальных свойств классических методов и критериев статистического анализа в условиях нарушения стандартных предположений.

При анализе результатов численных экспериментов целесообразно использование доступных математических пакетов, программных систем статистического анализа и средств графической визуализации.

Тематика КП может быть связана с анализом и развитием статистических методов, статистических критериев, характером использования методов и критериев в статистических пакетах, с корректностью применения статистических методов в приложениях и в программном обеспечении.

## 2. Методические указания

### 2.1 Статистические гипотезы и критерии их проверки

*Статистической гипотезой* называется любое утверждение о виде или свойствах распределения наблюдаемых в эксперименте случайных величин. Обычно статистические гипотезы делят на следующие виды: *однородности*, если имеется две или более выборок случайных величин; *независимости*, если имеется выборка многомерной случайной величины; *случайности*, если есть предположения о наличии в последовательности наблюдений систематических изменений; *о виде распределения*, если есть предположения о законе распределения случайной величины (рисунок 1).

Проверка статистической гипотезы состоит в том, чтобы сформулировать такое правило, которое позволило бы по результатам проведенных наблюдений принять или отклонить гипотезу. Правило, согласно которому гипотеза принимается или отвергается, называется *критерием* проверки статистической гипотезы.

С проверкой статистических гипотез связывают ошибки двух типов. *Ошибкой первого рода* называют событие, когда верная проверяемая гипотеза отвергается критерием. *Ошибкой второго рода* называют событие, когда неверная проверяемая гипотеза принимается критерием. Вероятности ошибок первого и второго рода обозначают  $\alpha$  и  $\beta$ , соответственно. Вероятность ошибки второго рода зависит от выдвигаемой конкурирующей гипотезы. Вероятность отклонения ложной проверяемой гипотезы, т.е. принятия правильного решения в пользу конкурирующей, называется *мощностью*, и она равна  $1 - \beta$ . Вероятность ошибки первого рода также называют *уровнем значимости* критерия.

Гипотезу, которую мы проверяем, будем называть *основной* или *нулевой* гипотезой, и будем всегда обозначать  $H_0$ . Альтернативные или *конкурирующие* гипотезы будем обозначать  $H_1, H_2, \dots, H_m$ .

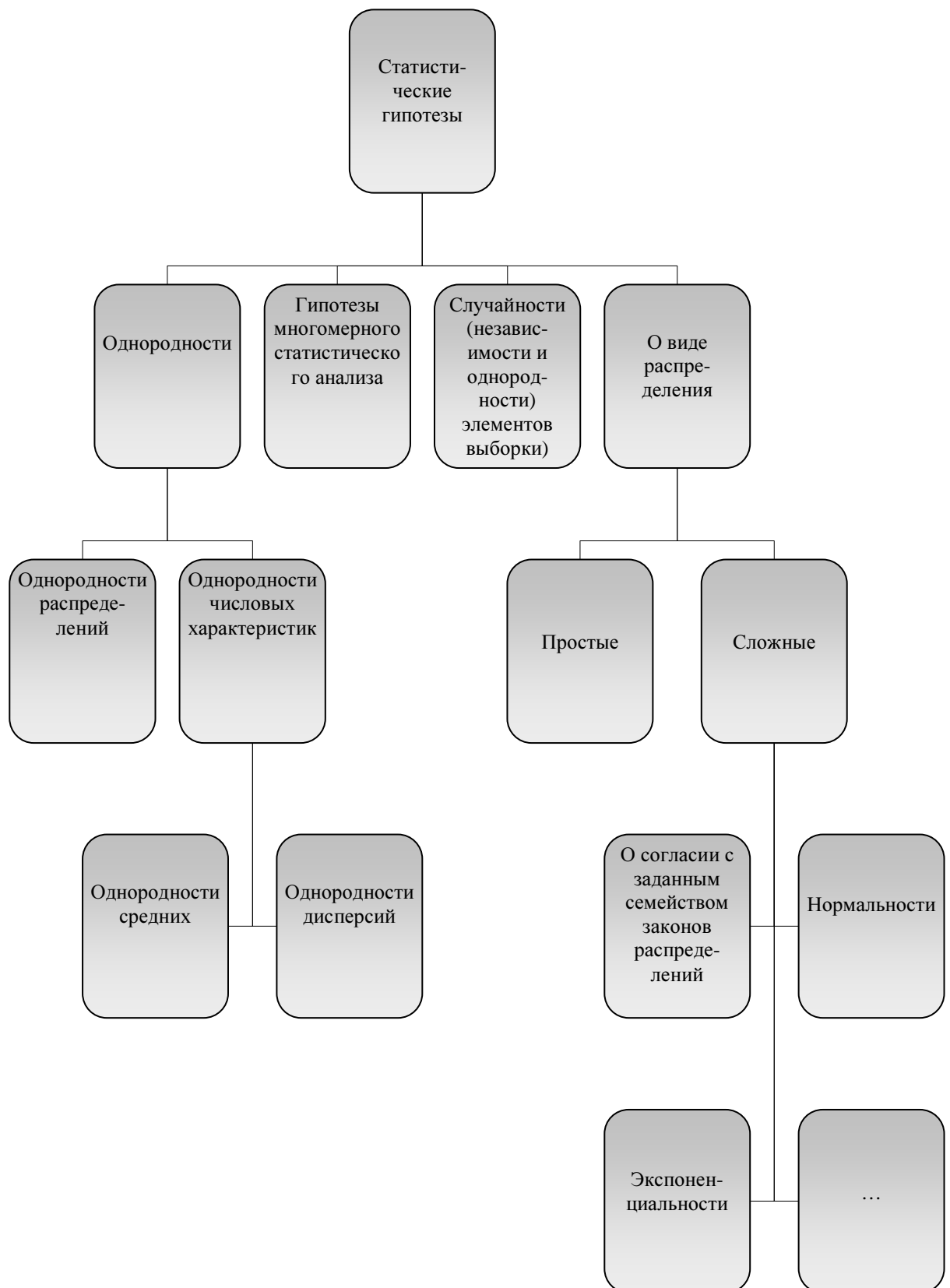


Рисунок 1 – Классификация статистических гипотез

Любой критерий проверки статистической гипотезы разбивает выборочное пространство на доверительную область  $\mathcal{H}_0$  и критическую

область  $\mathcal{X}_1$ . При попадании выборки в критическую область гипотеза отвергается, а при попадании в доверительную область – принимается. Чаще всего такое разбиение производится с помощью одномерной *статистики* – функции от выборки, поэтому критическая и доверительная область формулируются уже как подмножества множества вещественных чисел.

Доверительная область включает такие значения статистики критерия, при которых гипотеза принимается, а критическая область – значения, при которых гипотеза отвергается. Кроме того, вероятность попадания выборки (статистики критерия) в критическую область, когда гипотеза верна, по определению равна вероятности ошибки первого рода, а вероятность попадания выборки (статистики критерия) в доверительную область, когда гипотеза не верна, равна вероятности ошибки второго рода.

Как правило, встречаются три вида критических областей для статистики критерия:

- правосторонняя критическая область  $(t_\alpha, \infty)$ ;
- левосторонняя  $(-\infty, t_\alpha)$ ;
- двусторонняя  $(-\infty, t_{\alpha_1}) \cup (t_{\alpha_2}, \infty)$ ,  $\alpha_1 + \alpha_2 = \alpha$ .

## 2.2 Вычисление достигаемого уровня значимости

Достижимый уровень значимости (p-value) определяется как вероятность попадания статистики критерия:

- в область  $(S(X_n), \infty)$ , если критическая область правосторонняя;
  - в область  $(-\infty, S(X_n))$ , если критическая область левосторонняя;
- где  $S(X_n)$  - вычисленное значение статистики по реализации выборки.

Гипотеза отвергается, если достигаемый уровень значимости оказывается меньше заданной вероятности ошибки первого рода. Достоинство процедуры проверки гипотезы с использованием p-value в том, что не нужно заранее фиксировать уровень значимости и определять критическую область для значений статистики критерия. Кроме того, p-value характеризует “степень уверенности” в принимаемом решении, т.е. чем меньше p-value, тем больше оснований для отвержения основной гипотезы.

Если критическая область двусторонняя, то однозначного способа вычисления достигаемого уровня значимости нет. Например, можно использовать такой способ:

1. Пусть  $p = P\{S \in (S(X_n), \infty) | H_0\}$ .
2. Тогда p-value =  $2 \cdot \min(p, 1 - p)$ .
3. Гипотеза отвергается, если p-value  $< \alpha$ .

Достижимый уровень значимости является случайной величиной, определенной на интервале  $[0,1]$ , на основании которой делается статистический вывод о принятии гипотезы. Чем ближе значение  $p$ -value к 1, тем больше оснований для принятия гипотезы, чем ближе значение  $p$ -value к 0, тем больше оснований для отвержения гипотезы. Однако следует помнить о следующем важном замечании относительно  $p$ -value.

Когда основная гипотеза ложна, то  $p$ -value будет стремиться к 0 с ростом объема наблюдаемой выборки. Однако, когда основная гипотеза истинна,  $p$ -value не стремится к 1, а распределено равномерно на интервале  $[0,1]$ .

Вычисление  $p$ -value относительно просто, когда известно теоретическое распределение статистики критерия при справедливости основной гипотезы. Однако возможны ситуации, когда:

- неизвестен аналитический вид закона распределения статистики;
- известен только асимптотический закон распределения статистики;
- закон распределения статистики меняется от объема выборки, от метода оценивания параметров, от процента цензурирования и т.п.

В таких ситуациях для вычисления  $p$ -value можно эффективно применить метод Монте-Карло (алгоритм 1). Очень важным для корректного вычисления  $p$ -value (с требуемой точностью) является вопрос о правильном выборе числа повторений  $N$ , который будет рассмотрен в пункте 2.3.

**Алгоритм 1.** Вычисление достигаемого уровня значимости статистического критерия методом Монте-Карло.

*Входные данные:* гипотеза  $H_0$ , выборка  $X_n$ , количество повторений  $N$ , функция вычисления статистики  $S(X_n)$ .

*Действия.*

1. Вычислить  $S = S(X_n)$  – статистику критерия по выборке.
2. Установить  $m = 0$ .
3. Сгенерировать выборку  $Y_n$  при верной гипотезе  $H_0$ .
4. Вычислить значение  $S(Y_n)$ .
5. Если критическая область правосторонняя и  $S(Y_n) > S(X_n)$ , то  $m = m + 1$ .
6. Повторять шаги 3-5  $N$  раз.

*Выходные данные:* оценка достигаемого уровня значимости ( $p$ -value) равна

- $\hat{p} = \frac{m}{N}$  для правосторонней критической области;
- $\hat{p} = 1 - \frac{m}{N}$  для левосторонней критической области;

- $\hat{p} = 2 \cdot \min\left(\frac{m}{N}, 1 - \frac{m}{N}\right)$  для двусторонней критической области.

### 2.3 Определение количества повторений при оценивании вероятности наступления некоторого события

Пусть требуется вычислить вероятность  $p$  появления некоторого случайного события  $A$ . В каждой из  $N$  реализаций процесса количество наступлений события  $A$  является случайной величиной  $\xi$ , принимающей значение  $x_1=1$  с вероятностью  $p$ , и значение  $x_2=0$  с вероятностью  $1-p$ . Случайная величина  $\xi$  подчиняется распределению Бернулли с вероятностью успеха  $p$  и имеет математическое ожидание  $p$  и дисперсию  $p(1-p)$ .

В качестве оценки для искомой вероятности  $p$  принимается частота  $m/N$  наступлений события  $A$  при  $N$  реализациях

$$\frac{m}{N} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (1)$$

где  $x_i$  – количество наступлений события  $A$  в реализации с номером  $i$ .

В силу центральной предельной теоремы теории вероятностей частота при достаточно больших  $N$  имеет распределение, близкое к нормальному:

$$\sqrt{N} \frac{m/N - M\xi}{\sqrt{D\xi}} = \sqrt{N} \frac{m/N - p}{\sqrt{p(1-p)}} \rightarrow \eta \in N(0,1).$$

Отсюда

$$P\left\{\left|\sqrt{N} \frac{m/N - p}{\sqrt{p(1-p)}}\right| < t_\gamma\right\} = \Phi(t_\gamma) - \Phi(-t_\gamma) = 2\Phi(t_\gamma) - 1 = \gamma,$$

где  $t_\gamma = \Phi^{-1}\left(\frac{\gamma+1}{2}\right)$  – квантиль стандартного нормального распределения,

$$P\left\{\left|m/N - p\right| < t_\gamma \frac{\sqrt{p(1-p)}}{\sqrt{N}}\right\} = \gamma.$$

Таким образом, погрешность моделирования  $p$  с доверительной вероятностью  $\gamma$  равна

$$\varepsilon = t_\gamma \frac{\sqrt{p(1-p)}}{\sqrt{N}}. \quad (2)$$

Отсюда количество повторений, для которого погрешность моделирования не превышает  $\varepsilon$ , равно



$$N = \left[ t_{\gamma}^2 \frac{p(1-p)}{\varepsilon^2} \right] + 1, \quad (3)$$

где  $[\cdot]$  означает целую часть, так как количество повторений должно быть натуральным числом.

Погрешность моделирования зависит от величины  $p$  и достигает максимального значения при  $p = 0,5$ . При  $p \rightarrow 0$  и при  $p \rightarrow 1$  погрешность стремится к нулю. Для значений  $p$  близких к нулю или единице имеет смысл рассматривать *относительную* погрешность моделирования

$$\delta = \frac{\varepsilon}{p(1-p)} \quad (4)$$

Величина  $\delta_0 = \frac{\varepsilon}{p}$  показывает, во сколько раз погрешность моделирования превосходит вероятность  $p$ , близкую к 0. Величина  $\delta_1 = \frac{\varepsilon}{1-p}$  показывает, во сколько раз погрешность моделирования превосходит вероятность  $1-p$ , когда  $p$  близко к 1. Формула (4) обобщает эти две ситуации.

Тогда формулу (3) можно переписать для относительной погрешности в виде

$$N = \left[ t_{\gamma}^2 \frac{1}{p(1-p)\delta^2} \right] + 1. \quad (5)$$

### 3. Структура пояснительной записки к КП

Рекомендуемый объем записки не должен превышать 30-35 страниц текста.

Примерная структура пояснительной записки имеет следующий вид.

#### 1. Введение

Во введении указывается цель работы, кратко характеризуется место и значение соответствующего метода в аппарате прикладной математической статистики.

#### 2. Постановка задачи

Постановка задачи включает алгоритм или последовательность действий, связанную с применением соответствующего метода или критерия, основные соотношения, а также формулировку предпосылок, обуславливающих область корректного использования метода.

#### 3. Аналитический обзор

Аналитический обзор характеризует состояние соответствующей области на момент появления (создания) метода (критерия). Он включает указания на ключевые работы, связанных с возникновением,

развитием и актуальным применением соответствующих методов в приложениях.

#### **4. Результаты исследований**

Приводятся результаты исследований (численных экспериментов) автора КП. Из теста записки должно быть однозначно понятно, с какой целью проводились эксперименты, что для этого было сделано, какие средства использовались, какова точность экспериментов, что было выявлено в результате исследований (что подтвердилось, что не подтвердилось, почему).

#### **5. Выводы**

На основании п.4. приводится краткая формулировка ключевых результатов численных экспериментов. Формулируются общие выводы, в которых приводится краткая характеристика исследуемого метода (критерия), области его применения, даются рекомендации по использованию.

#### **6. Список использованных источников**

Приводится список источников, отмеченных в библиографическом обзоре ключевых работ по теме КП, включая Интернет-издания, а также работ, использованных в связи с проведенными экспериментами.

#### **4. Порядок выполнения**

1. Сформулировать постановку задачи, связанной с применением исследуемого критерия. Четко сформулировать предпосылки, выполнение которых обуславливает корректность применения соответствующего критерия.

2. Выполнить библиографический обзор ключевых работ, связанных с возникновением, развитием и актуальным применением соответствующих методов в приложениях. Основной упор сделать на использование ресурсов научной электронной библиотеки (<http://elibrary.ru/defaultx.asp>), информационно-поисковых систем (<http://scholar.google.ru/>), свободной энциклопедии (<http://www.wikipedia.org/>), сайтов научных журналов. Провести исторический анализ, характеризующий состояние соответствующей области на момент появления (создания) метода (критерия).

3. Разработать программное обеспечение для проверки гипотезы о виде распределения в соответствии с вариантом задания.

Входные данные для проверки гипотезы:

- выборка (формат исходных файлов с выборками должен соответствовать программе ISW);
- уровень значимости критерия (вероятность ошибки первого рода);
- количество повторений в методе Монте-Карло (или погрешность моделирования).

Результаты проверки гипотезы должны включать:

- статистику критерия;
- достигаемый уровень значимости (p-value), вычисленный по предельному распределению статистики (если есть);

- достигаемый уровень значимости (p-value), вычисленный по методу Монте-Карло;
- результат проверки гипотезы (отклоняется или нет).
- 4. Провести тестирование разработанного программного обеспечения по выборкам, смоделированным в соответствии с основной гипотезой.
- 5. Если критерий имеется в ISW, то сравнить полученные результаты в п.4 с результатами работы программы ISW.
- 6. Провести проверку гипотезы на реальных данных, с указанием источника данных.
- 7. Сформулировать выводы по работе. Оформить пояснительную записку по работе.

### **Варианты заданий**

1. Проверка гипотезы о виде распределения по критерию Колмогорова
2. Проверка гипотезы о виде распределения по критерию Смирнова
3. Проверка гипотезы о виде распределения по критерию Крамера-Мизеса-Смирнова
4. Проверка гипотезы о виде распределения по критерию Андерсона-Дарлинга
5. Проверка гипотезы о виде распределения по критерию  $\chi^2$  Пирсона
6. Проверка гипотезы о виде распределения по критерию Рао-Робсона-Никулина
7. Проверка гипотезы о виде распределения по критерию отношения правдоподобия
8. Проверка гипотезы о виде распределения по критерию Реньи
9. Проверка гипотезы о виде распределения по модифицированному медианному критерию
10. Проверка гипотезы о виде распределения по модифицированному критерию Колмогорова-Смирнова
11. Проверка гипотезы о виде распределения по модифицированному вероятностному критерию
12. Проверка гипотезы о виде распределения по критерию Жанга  $Z_k$
13. Проверка гипотезы о виде распределения по критерию Жанга  $Z_s$
14. Проверка гипотезы о виде распределения по критерию Жанга  $Z_a$
15. Проверка гипотезы равномерности по критерию Шермана
16. Проверка гипотезы равномерности по критерию Морана
17. Проверка гипотезы равномерности по критерию Ченга-Спиринга.
18. Проверка гипотезы равномерности по критерию Саркади-Косика.
19. Проверка гипотезы равномерности по критерию Хегази-Грина
20. Проверка гипотезы равномерности по критерию Гринвуда-Кэсенберри-Миллера
21. Проверка гипотезы экспоненциальности по критерию Большева
22. Проверка гипотезы экспоненциальности по критерию Гнеденко
23. Проверка гипотезы экспоненциальности по критерию Харриса

24. Проверка гипотезы экспоненциальности по критерию Холландера-Прошана
25. Проверка гипотезы экспоненциальности по критерию Гини
26. Проверка гипотезы экспоненциальности по критерию Эпштейна
27. Проверка гипотезы экспоненциальности по критерию Кокса-Оукса
28. Проверка гипотезы экспоненциальности по критерию Эппса-Палли
29. Проверка гипотезы экспоненциальности по критерию Ватсона
30. Проверка гипотезы экспоненциальности по критерию Купера
31. Проверка гипотезы экспоненциальности по критерию Дешпанде
32. Проверка гипотезы экспоненциальности по критерию Клара
33. Проверка гипотезы экспоненциальности по критерию Барингхауса-Хенце
34. Проверка гипотезы экспоненциальности по критерию Хенце
35. Проверка гипотезы экспоненциальности по критерию Хенце-Мейнтаниса
36. Проверка гипотезы нормальности по критерию Фросини
37. Проверка гипотезы нормальности по критерию Хегази-Грина
38. Проверка гипотезы нормальности по критерию Гири
39. Проверка гипотезы нормальности по критерию Дэвида-Хартли-Пирсона
40. Проверка гипотезы нормальности по критерию Шпигельхальтера
41. Проверка гипотезы нормальности по критерию Шапиро-Уилка
42. Проверка гипотезы нормальности по критерию Ройстона
43. Проверка гипотезы нормальности по критерию Эппса-Палли,
44. Проверка гипотезы нормальности по критерию Д'Агостино
45. Проверка гипотезы о распределении Вейбулла-Гнеденко по критерию Майкла
46. Проверка гипотезы о распределении Вейбулла-Гнеденко по критерию Шапиро-Уилка
47. Проверка гипотезы о распределении Вейбулла-Гнеденко по критерию Тайку-Синга
48. Проверка гипотезы о распределении Вейбулла-Гнеденко по критерию Майка-Фертига-Шуйера

## Литература

1. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : [монография] / Б. Ю. Лемешко [и др.]. - Новосибирск, 2011. - 887 с. : ил., табл.
2. Кобзарь А.И. Прикладная математическая статистика для инженеров и научных работников. – М.: Физматлит, 2006. – 816 с.

## Приложение. Формат выборки программы ISW

Система работает только с одномерными выборками, хранящимися в файлах с расширением “**dat**”. Файл можно создать с помощью любого текстового редактора, либо сгенерировать программно по заданному формату.

### 1.4.1.1. Формат входных данных

Первая строка файла содержит название выборки, в этой строке может быть произвольная информация, но мы рекомендуем вводить в этой строке источник этой выборки, информацию о случайной величине, условия проведения эксперимента. Информация из этой строки используется при построении графиков.

Во второй строке файла содержится информация о типе выборки. Остальные строки содержат информацию в зависимости от типа выборки.

#### • Тип выборки 0. Точечная выборка

Точечная выборка объемом  $n$  наблюдений имеет следующий формат:

<название выборки>

0 n

<наблюдение 1>

<наблюдение 2>

...

<наблюдение n>

#### • Тип выборки 1. Интервальная выборка с абсолютной и относительной погрешностью

Интервальная выборка объемом  $n$  наблюдений с абсолютной погрешностью  $a$  и относительной погрешностью  $r$  имеет следующий формат:

<название выборки>

1 n a r

<наблюдение 1>

<наблюдение 2>

...

<наблюдение n>

- **Тип выборки 2. Частично группированная выборка**

Частично группированная выборка из  $n$  точечных наблюдений и  $k$  интервальных наблюдений имеет формат:

<название выборки>

2 k n

<n\_1> <n\_2> ... <n\_k>

<x\_1> <x\_2> ... <x\_{k-1}>

<наблюдение 1>

<наблюдение 2>

...

<наблюдение n>

где <n\_i> - количество наблюдений в i-м интервале

и <x\_i> - i-я граничная точка

- **Тип выборки 3. Группированная выборка**

Группированная выборка  $k$  интервальных наблюдений имеет формат:

<название выборки>

3 k

<n\_1> <n\_2> ... <n\_k>

<x\_1> <x\_2> ... <x\_{k-1}>

где <n\_i> - количество наблюдений в i-м интервале и <x\_i> - i-я граничная точка

- **Тип выборки 4. Цензурированная слева выборка I-го типа**

Цензурированная выборка из  $n$  точечных наблюдений и интервала цензурирования *слева* имеет формат:

<название выборки>

4 n

<n\_c>

<x\_c>

<наблюдение 1>

<наблюдение 2>

...

<наблюдение n>

где <n\_c> - количество наблюдений в интервале цензурирования

и <x\_c> - точка цензурирования

- **Тип выборки 5. Цензурированная справа выборка I-го типа**

Цензурированная выборка из  $n$  точечных наблюдений и интервала цензурирования *справа* имеет формат:

<название выборки>

5 n

<n\_c>

<x\_c>

<наблюдение 1>

<наблюдение 2>

...

<наблюдение n>

где <n\_c> - количество наблюдений в интервале цензурирования

и <x\_c> - точка цензурирования

- **Тип выборки 6. Цензурированная с двух сторон выборка I-го типа**

Цензурированная выборка из  $n$  точечных наблюдений и интервалов цензурирования слева и справа имеет формат:

<название выборки>

6 n

<n\_l><n\_r>

<x\_l><x\_r>

<наблюдение 1>

<наблюдение 2>

...

<наблюдение n>

где <n\_l> - количество наблюдений в интервале цензурирования слева

и <n\_r> - количество наблюдений в интервале цензурирования справа

и <x\_l> - точка цензурирования слева

и <x\_r> - точка цензурирования справа

- **Тип выборки 10. Интервальная выборка**

Интервальная выборка из  $n$  интервальных наблюдений

<название выборки>

10 n

<a\_1> <b\_1>

<a\_2> <b\_2>

<a\_3> <b\_3>

....

<a\_n> <b\_n>

где <a\_i> - левая граница интервального наблюдения

и <b\_i> - правая граница интервального наблюдения.

#### **1.4.1.2. Создание выборки в текстовом редакторе**

Рассмотрим пример, как можно создать выборку с использованием текстового редактора, например Notepad (Блокнот).

## Пример

А) Время ремиссии (в неделях) 42 пациентов с острой лейкемией было приведено в отчете [5] о клинических испытаниях препарата б-терсарторигине (б-МР). Каждый пациент случайным образом получал б-МР или плацебо. Изучение было закончено через один год.

Были получены следующие выборки, в неделях:

Выборка с применением плацебо (21 пациент) содержит следующие наблюдения: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23.

Выборка с применением препарата б-МР (21 пациент): 6, 6, 6, 7, 10, 13, 16, 22, 23, 7, 10, 11, 13, 19, 20, 24, 27, 33, 35, 37, 42.


Создать выборки «Выборка плацебо.dat» и «Выборка б-МР.dat» в текстовом редакторе Notepad.

Б) В таблице приведено распределение толщины 12 000 бобов.

Толщина, мм	До 7.00	7.00-7.25	7.25-7.5	7.5-7.75	7.75-8.00	8.00-8.25	8.25-8.5	8.5-8.75
Количество бобов	32	103	239	624	1187	1650	1883	1930
Толщина, мм	8.75-9.00	9.00-9.25	9.25-9.5	9.25-9.75	9.75-10.00	10.00-10.25	10.25-10.5	Свыше 10.5
Количество бобов	1638	1130	737	427	221	110	57	32

Создать выборку «Толщина бобов.dat» в текстовом редакторе Notepad.

Так как все наблюдения выборки с применением плацебо являются точками, то тип первой выборки – точечный. Чтобы ввести эту выборку, открываем в программе Notepad новый файл и вводим в него данные, как показано на рис. 1.2. Затем сохраняем этот файл, например, с именем «Выборка плацебо.dat». Аналогично вводим вторую выборку «Выборка б-МР.dat». Выборка с толщиной бобов является группированной, поэтому вводим граничные точки и количества по формату «3» (рис. 1.3).

Теперь можно открыть эти выборки в системе. Для этого выбираем в меню **Файл** пункт **Открыть**. Открывается стандартное окно Windows выбора файла. Допускается выбрать не один файл, а несколько, используя клавиши <Ctrl> или <Shift> (рис. 1.4). Список открытых выборок можно посмотреть по кнопке  на вкладке **Выборки** (рис. П1).



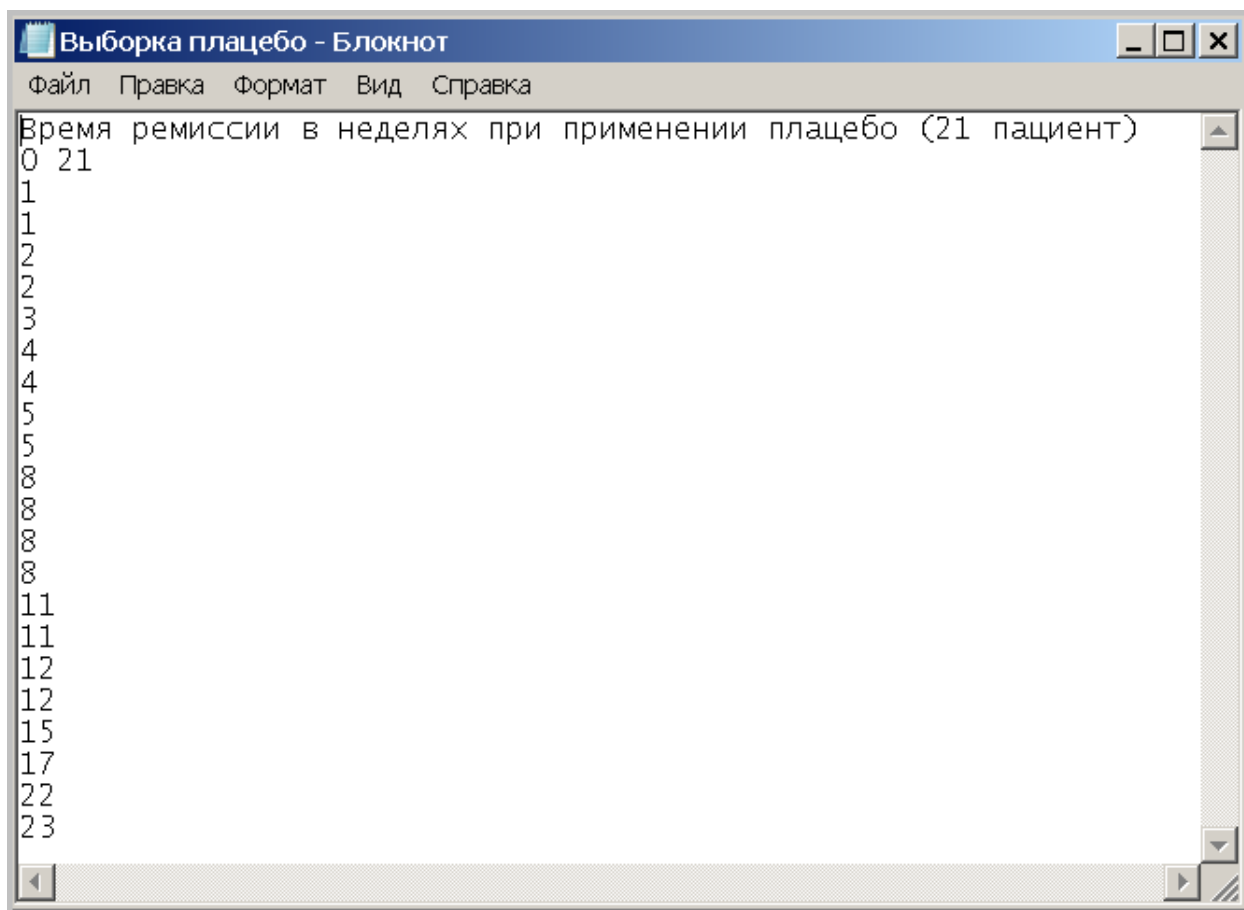


Рис. П1. Создание выборки «Выборка плацебо» в текстовом редакторе

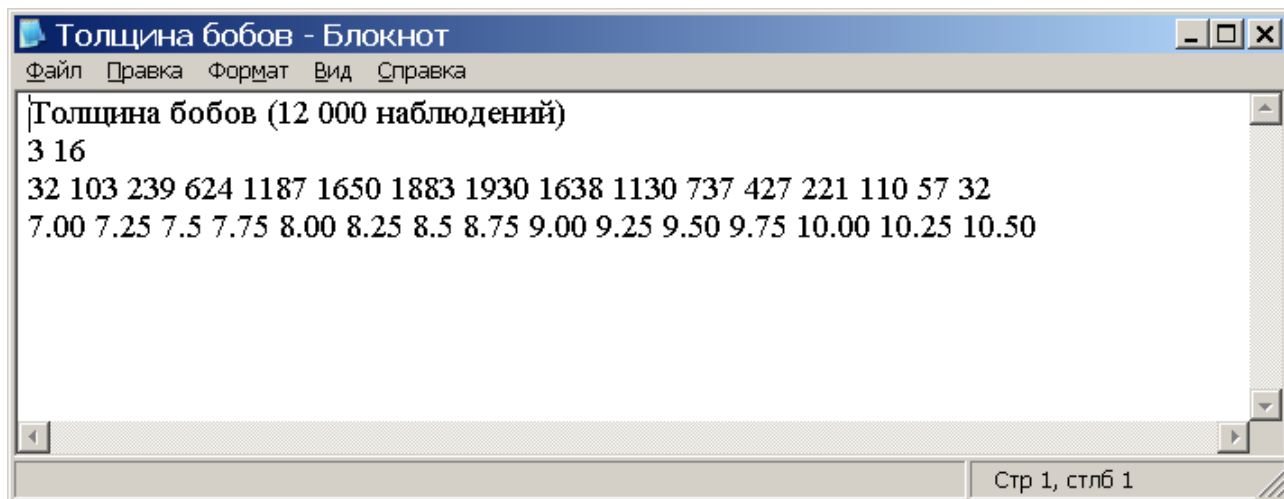


Рис. П2. Создание группированной выборки