

Министерство образования и науки Российской Федерации  
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ

**Компьютерные технологии анализа  
данных и исследования статистических  
закономерностей:  
исследование мощности критериев  
проверки статистических гипотез**

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

Новосибирск, 2012

Методические указания предназначены для студентов, выполняющих курсовые проекты по курсу "Компьютерные технологии анализа данных и исследования статистических закономерностей" в третьем 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 «Исследование мощности критериев проверки статистических гипотез».**

### **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 оценивания и критериев проверки статистических гипотез. Исследование статистических свойств оценок. Исследование мощности критериев относительно различных конкурирующих гипотез. Сравнительный анализ мощности групп критериев, ориентированных на проверку аналогичных гипотез. Закрепление навыков проведения самостоятельных исследований.

### **2. Методические указания**

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.

Расчетная часть КП должна предусматривать вычислительные эксперименты, направленные, например, на проверку выполнения асимптотических свойств оценок, на проверку соответствия распределений статистик критериев предельным (или асимптотическим) в зависимости от

объемов выборок и при выполнении предпосылок применения соответствующего критерия. Вычислительные эксперименты могут быть, направлены на уточнение области использования метода или критерия при исследовании последних в условиях нарушения стандартных предположений, обуславливающих корректность статистических выводов.

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.

### **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 текста 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 метода, алгоритма или критерия. Четко сформулировать предпосылки, выполнение которых обуславливает корректность применения соответствующего метода или 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. Наметить план вычислительных экспериментов по исследованию мощности критериев согласия: сформулировать несколько альтернативных гипотез (не менее двух), объемы выборок, доли цензурирования и т.п.

4. Выбрать доступные и наиболее оптимальные средства для проведения вычислительных экспериментов при необходимости, разработав требуемое программное обеспечение.

5. Провести намеченные вычислительные (имитационные) эксперименты. Оценить точность моделирования. Удостовериться в корректности результатов исследований. Проанализировать результаты экспериментов, сравнить с имеющимися теоретическими результатами.

6. Сформулировать выводы по работе. Оформить пояснительную записку по работе.

## Варианты заданий

№ варианта	Критерии	Уровень сложности
1	1,2,3,4	17
2	5,6	17

3	1,7	20
4	8,9	18
5	9,10	18
6	11,12,13	18
7	12,13,14	18
8	13,14,15	18
9	14,15,16	18
10	17,18,19	18
11	18,19,20	18
12	19,20,21	18
13	20,21,22	18
14	21,22,23	18
15	22,23,24	18
16	23,24,25	18
17	24,25,26	18
18	27,28,29	20
19	28,29,30	20
20	29,30,31	20
21	32,33,34	19
22	35,36,37	19
23	36,37,38	19
24	37,38,39	19
25	38,39,40	19
26	39,40,41	19
27	35,37,39	19
28	36,38,40	19
29	42,43,44	20
30	43,44,45	20

№	Критерий	Источник	Закон распределения $\xi$	Дополнительные исследования
<b>Критерии согласия</b>				
1	Колмогорова	[1]	Нормальное, Экспоненциальное, Коши	Статистика Колмогорова с поправкой Большева $\frac{6nD_n + 1}{6\sqrt{n}}$
2	Смирнова	[1]	Логистическое, Вейбулла, Коши	
3	Крамера-Мизеса-Смирнова	[1]	Лапласа, Экспоненциальное, Коши	
4	Андерсона-Дарлинга	[1]	Нормальное, Рэлея, Коши	
5	$\chi^2$ Пирсона (простая гипотеза)	[1]	Нормальное, Вейбулла, Коши	Число интервалов группирования 2, 3, 5, 7, 10 АОГ, РВГ
6	$\chi^2$ Рао-Робсона-Никулина (сложная гипотеза)	[1]	Нормальное, Экспоненциальное	Число интервалов группирования 2, 3, 5, 7, 10 АОГ, РВГ
7	Колмогорова для цензурированных справа или слева выборок (простая гипотеза)	[2]	Экспоненциальное, Рэлея	Степени цензурирования 5, 10, 20, 30, 40, 50, 60, 70, 80 Цензурирование I типа, II типа, слева, справа
8	Модифицированный медианный критерий	[3], с. 225	Нормальное, Экспоненциальное, Коши	
9	Модифицированный критерий Колмогорова-Смирнова	[3], с. 225	Логистическое, Вейбулла, Коши	
10	Модифицированный вероятностный критерий	[3], с. 226	Нормальное, Рэлея, Коши	
11	Шермана	[3], с.319	Равномерный	
12	Морана	[3], с.320	Равномерный	
13	Ченга-Спиринга).	[3], с.322	Равномерный	
14	Саркади-Косика).	[3], с.323	Равномерный	
15	Хегази-Грина	[3], с.326	Равномерный	
16	Гринвуда-Кэсенберри-Миллера	[3], с.320	Равномерный	
17	Большева	[3], [4]	Экспоненциальное	
18	Гнеденко	[5] с. 1812	Экспоненциальное	
19	Харриса	[5] с. 1813	Экспоненциальное	
20	Холландера-Прошана	[5] с. 1814	Экспоненциальное	
21	Гини	[5] с. 1815	Экспоненциальное	
22	Эпштейна	[5] с. 1816	Экспоненциальное	
23	Кокса-Оукса	[6] с. 33	Экспоненциальное	
24	Эпса-Палли	[6] с. 36	Экспоненциальное	
25	Ватсона	[3], с.282, с.222	Экспоненциальное	
26	Купера	[3], с.282, с. 223	Экспоненциальное	
27	t-критерий Стьюдента при известных дисперсиях	[1]	Нормальное, Экспоненциальное, Коши	рассмотреть разные комбинации объемов выборок
28	t-критерий Стьюдента	[1]	Нормальное,	рассмотреть разные комбинации

	при неизвестных, но равных дисперсиях		Экспоненциальное, Коши	объемов выборок
29	t-критерий Стьюдента при неизвестных и неравных дисперсиях	[1]	Нормальное, Экспоненциальное, Коши	рассмотреть разные комбинации объемов выборок отличие дисперсий в 2х выборках в 2 раза, в 5 раз, в 10 раз
30	Уилкоксона	[1]	Нормальное, Экспоненциальное, Коши	рассмотреть разные комбинации объемов выборок рассмотреть также случай, когда дисперсии не равны (отличие в 2 раза, в 5 раз, в 10 раз)
31	Манна-Уитни	[1]	Нормальное, Экспоненциальное, Коши	рассмотреть разные комбинации объемов выборок рассмотреть также случай, когда дисперсии не равны (отличие в 2 раза, в 5 раз, в 10 раз)
<b>Критерии однородности распределений</b>				
32	Смирнова	[1]	Нормальное, Экспоненциальное, Коши	рассмотреть разные комбинации объемов выборок
33	Лемана-Розенблатта	[1]	Нормальное, Экспоненциальное, Коши	рассмотреть разные комбинации объемов выборок
34	Катценбайссера-Хакля	[3] с. 228	Логистическое, Вейбулла, Коши	рассмотреть разные комбинации объемов выборок
<b>Критерии однородности дисперсий (характеристик рассеяния)</b>				
35	Левене	[1]	Нормальное, Лапласа, Вейбулла	выборочное среднее, выборочная медиана, усеченное среднее рассмотреть разные комбинации объемов выборок
36	Бартлетта	[1]	Нормальное, Макс. значений, Экспоненциальное	рассмотреть разные комбинации объемов выборок
37	Ансари-Бредли	[1]	Нормальное, Рэля, Коши	рассмотреть разные комбинации объемов выборок
38	Муда	[1]	Нормальное, Максвелла, Коши	рассмотреть разные комбинации объемов выборок
39	Сижела-Тьюки	[1]	Лапласа, Экспоненциальное Коши,	рассмотреть разные комбинации объемов выборок
40	Кейпена	[1]	Логистическое, Экспоненциальное, Коши	рассмотреть разные комбинации объемов выборок
41	Клотца	[1]	Нормальное, Вейбулла, Коши	рассмотреть разные комбинации объемов выборок
<b>Критерии выявления тренда</b>				
42	Сериальный критерий Шведа-Эйзенхарта	[3], с.621		
43	Критерий автокорреляции Кенюя	[3], с. 622		
44	Критерий Блума-Кифера-Розенблатта	[3], с.623		
45	Критерий Гёфдинга	[3], с.628		



## 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. Лемешко Б.Ю., Чимитова Е.В., Плешкова Т.А. Проверка простых и сложных гипотез о согласии по цензурированным выборкам // Научный вестник НГТУ. - 2010. - № 4(41). – С.13-28.  
[http://ami.nstu.ru/~headrd/seminar/publik\\_html/N\\_vestnik\\_2010.pdf](http://ami.nstu.ru/~headrd/seminar/publik_html/N_vestnik_2010.pdf)
3. Кобзарь А.И. Прикладная математическая статистика для инженеров и научных работников. – М.: Физматлит, 2006. – 816 с.
4. Большев Л.Н. К вопросу о проверке «показательности». Вероятность и ее применения С. 542-544. (есть в электронном виде)
5. Ascher S. A survey of tests for exponentiality. Communications in Statistics - Theory and Methods, 1811-1825 (есть в электронном виде)
6. Henze N. and Meintanis S.G. Recent and classical tests for exponentiality: a partial review with comparisons. Metrika (2005) 61: 29–45 (есть в электронном виде)