На примере построения этих 3-х кривых, зная формулы других замечательных кривых, можно дополнять и развивать программу.

Использование данного материала расширяет кругозор интересующихся математикой и усиливает познавательный интерес к ней, развивает пространственное представление и мышление. Применение замечательных кривых математики широко распространено в производстве, строительстве, военном деле.

Список использованных источников

1. А. И. Маркушевич Замечательные кривые М., 1978 г., 48 стр. с илл.
2. Вольф Д. OpenGL 4. Язык шейдеров. Книга рецептов/ пер. с англ. А. Н. Киселева – М.: ДМК Пресс, 2015. – 368 стр. с илл.
3. Выгодский М.Я. Справочник по высшей математике. Изд-во «Наука». М. 1977 г.
4. Штейнгауз Г. Математический калейдоскоп: Пер. с польского. – М.: Наука. Главная редакция физико-математической литературы, 1981 г.

ИЗУЧЕНИЕ НЕКОТОРЫХ ПОСТ-ТЕСТОВ МНОЖЕСТВЕННЫХ СРАВНЕНИЙ И ИХ РАБОТОСПОСОБНОСТИ ПРИ НЕНОРМАЛЬНОСТИ ВЫБОРОК МЕТОДОМ МОНТЕ-КАРЛО

Авторы: Калиничева Диана, Радзиевская Вячеслава студентки 4 курса Университета «Дубна» филиала «Протвино» г. Протвино Московской области

Научный руководитель: к. ф-м н., доцент Масликов Александр Альбертович

Аннотация.

Методом Монте-Карло симулируется использование тестов, выполняемых при множественных сравнениях на выборках малого объёма из искаженных (по сравнению с нормальным) распределений. Вычисляются эмпирические эффективные вероятности ошибок 1-го рода для дисперсионного анализа, пост-тестов множественных сравнений Тьюки и процедуры Бенжамини-Хохберга.

Annotation.

The Monte Carlo method simulates the use of tests performed in multiple comparisons on small-volume samples from distorted (compared to normal) distributions. Empirical effective probabilities of errors of the 1st kind are calculated for the analysis of variance, Tukey's post-tests of multiple comparisons and Benjamini-Hochberg's procedure.

Ключевые слова: Математическая статистика, множественные сравнения, дисперсионный анализ, апостериорные тесты.

Keywords: Mathematical statistics, multiple comparisons, ANOVA, post-hoc tests.

Условием надёжной работы большинства классических статистических тестов является нормальное (Гауссово) распределение исследуемых генеральных совокупностей и их однородность (гомогенность) по дисперсиям [1]. В то же время во многих отраслях знания (например, в медицине, в фармакологии, в токсикологии) при математико-статистических исследованиях зачастую приходится иметь дело с выборками небольшого объема (5—10 элементов), для которых проблематично гарантированно установить нормальность и/или однородность дисперсий. Поэтому представляется интересным изучить поведение традиционных статистических критериев на ненормальных распределениях и выборках малого объема. В качестве инструмента мы использовали программы, написанные в пакете Wolfram Mathematica (WM). В основе исследования лежит метод Монте-Карло, т.е. идея в том, чтобы генерировать случайные выборки из различных распределений и применять к ним исследуемые тесты. В WM есть возможность подключить генератор псевдослучайных чисел “Mersenne Twister”, который мы и использовали. Вихрь Мерсенна - это генератор сдвигового регистра обобщенной обратной связи с огромным периодом ($2^{19937}-1$), гарантирующим высочайшую степень случайности. Пакет WM позволяет формировать компактные программы для генерации огромного числа случайных выборок из большого набора библиотечных распределений. Мы вполне можем использовать количество генераций порядка 10^5—10^6 на каждый набор параметров.

В данном исследовании мы остановили своё внимание на пост-тестах множественных сравнений, т.е. тестах, которые выполняются после того как, например, дисперсионный анализ (ANOVA) определил наличие каких-то различий [2, 3]. Назначение этой группы тестов – определить, какие именно группы статистически значимо отличаются между собой. Разработка этой группы тестов обусловлена катастрофическим падением мощности парного t-сравнения при увеличении множественности сравнений (так называемый эффект поправки Бонферони). Мы приведём результаты по исследованию апостериорного теста множественных сравнений Тьюки (Tukey) [4] (они остаются справедливыми и для его модификации Ньюмена-Кейлса (Newman-Keuls)), а также по пост-тесту процедуре Бенжамини-Хохберга (Benjamini-Hochberg) [5]. В основе этих тестов лежит t-критерий Стьюдента сравнения средних, опирающийся на известную t-статистику [2, 3].

t-распределение Стьюдента - это непрерывное одномерное распределение с одним параметром - количеством степеней свободы. Пусть $Y_0, Y_1, ..., Y_n$ — независимые стандартные нормальные случайные величины (матожидание=0, дисперсия=1). Тогда распределение случайной величины *t*, где $t = \frac{Y_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}}$ называется распределением Стьюдента с *n* степенями свободы (*df*=*n*).

Плотность вероятности этого распределения $f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}}$, где Г(...)- гамма функция Эйлера. Отметим, что эта формула допускает и дробное число степеней свободы.

t-критерий может применяться в двух модификациях: 1) в варианте статистически одинаковых дисперсий групп; 2) в варианте статистически различных дисперсий. В первом варианте число степеней свободы в t-статистике $df = n + m - 2$, где *n* и *m* – объёмы сравниваемых выборок. Во втором варианте число степеней свободы становится дробным числом и вычисляется по весьма сложной формуле. Для процедуры Бенжамини-Хохберга мы тестируем оба варианта t-критерия.

Отклонения распределений генеральных совокупностей от нормального Гауссового распределения можно описывать параметрами: эксцесс (коэффициент островершинности *E*) и асимметрия (*A*). Для нормального распределения *E* =3, а *A*=0. На первом шаге в качестве модели искажённого нормального распределения мы взяли распределение χ^2 с числом степеней свободы *df* = 4. Это распределение является частным случаем Гамма распределения и описывается функцией плотности вероятности:

$$f_{\chi^2(n)} = \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \tag{17}$$

Параметры данного распределения: матожидание $\mu = n$,дисперсия $\sigma^2 = 2n$, коэффициент асимметрии $\sqrt{8/n}$, эксцесс 3+12/n, где n – число степеней свободы. В нашем случае χ^2 распределения с df = 4 имеем: $\mu = 4$, $\sigma^2 = 8$, коэффициент асимметрии $A = \sqrt{2}$, эксцесс E = 6.

Исследования методом Монте-Карло дисперсионного анализа и пост-теста Тьюки на нормальных выборках подтверждают правильность общего подхода. Т.е. дают эмпирическую вероятность ошибки 1-го рода 0.05 при заложенном уровне значимости 5%. При этом наблюдаются ситуации, когда ANOVA указывает на наличие различий, а тест Тьюки не в состоянии их определить. Это следствие меньшей мощности пост-тестов по сравнению с дисперсионным анализом.

Далее, мы исказили исследуемое распределение, взяв χ^2 с числом степеней свободы df = 4. Обсчитали это распределение, меняя объем выборки, методом Монте-Карло с числом испытаний 20000 (для повышения точности планируется довести число испытаний до 10^6). Число сравниваемых выборок взяли 4, и, кроме того, исследовали случай, когда асимметрия 2-х из 4-х выборок направлена в отрицательную сторону. Далее применялся дисперсионный анализ (ANOVA) и пост тест Тьюки (эти же результаты справедливы и для пост-теста Ньюмена-Кейлса (NK)). Результаты этого модельного исследования представлены в Таблице 1. Мы видим, что искажение нормальности распределения не сказывается катастрофическим образом на достоверности ANOVA и пост-тестов. Более того, при одинаковой направленности асимметрии наблюдается даже уменьшение вероятности ошибки 1-го рода. Только в случае разнонаправленных асимметрий эта ошибка незначительно возрастает.

Таблица 1. Вероятности ошибок 1-го рода для ANOVA и пост-теста Тьюки.
I – сонаправленные асимметрии, II – разнонаправленные асимметрии.

	Объем	5	6	7	8	9	10
II	ANOVA	0,04395	0,04785	0,0453	0,04575	0,0444	0,0484
	Tukey (NK)	0,04355	0,0462	0,0445	0,0443	0,0436	0,0465
II	ANOVA	0,06025	0,0593	0,05755	0,05615	0,0548	0,0547
I	Tukey (NK)	0,0596	0,05775	0,05535	0,05775	0,05465	0,0542

Перейдём к исследованию процедуры Бенжамини-Хохберга. Приведём результаты тестирования процедуры Бенжамини-Хохберга на нормальных выборках различного объёма. Процедура организована таким образом, что при обнаружении негомогенности дисперсий используется t-сравнение по 2-му варианту, а при однородных дисперсиях по 1-му. Результаты исследований пост-теста процедуры Бенжамини-Хохберга для различного количества и объёма выборок приводим в Таблице 2. Количество симуляций здесь составляет 10^5.

Таблица 2 – Вероятности ошибок 1-го рода для процедуры Бенжамини-Хохберга.

Кол-во элементов \ Кол-во выборок	4	5	6	7
6	0.04467	0.04232	0.04185	0.04079
7	0.04483	0.04415	0.0424	0.04156
8	0.04607	0.04364	0.04230	0.04212

Далее мы планируем исказить нормальность распределения в одном и в обеих направлениях и получить для искаженных выборок результаты аналогичные Таблице 2. Наблюдаем, что эмпирическая вероятность ошибки 1-го рода укладывается в декларируемый уровень значимости и имеется тенденция к понижению этой вероятности с ростом количества сравниваемых выборок.

Представляется полезным итерационным методом найти уровни значимости, соответствующие статистически наблюдаемым ошибкам 1-го рода для используемых «ненормальных» распределений. Имеет смысл провести аналогичные исследования для распределений с другими значениями параметров и другими малыми объёмами выборок. Также мы планируем аналогичным методом исследовать влияние «ненормальности» на мощность критериев.

Список использованных источников

1. Гмурман В.Е. Теория вероятностей и математическая статистика – М., Высшая школа, 2003;
2. С.Гланц. Медико-биологическая статистика. Пер. с англ. – М.,Практика,1998 – 459 с.;
3. Лапач С.Н., Чубенко А.В., Бабич П.Н., Статистические методы в медико-биологических исследованиях с использованием Excel. – 2-е изд. перераб. и доп. – К.:МОРИОН, 2001. – 408 с.
4. Тьюки, Джон (1949). "Сравнение индивидуальных средних в Дисперсионном анализе". Биометрия. 5 (2): 99–114.
5. Y. Benjamini, Y. Hochberg. «Controlling the Discovery Rate : a Rractical and Powerful Approach to Multiple Testing». Journal of the Royal Statistical Society.Series B (Methodological), Vol.57, No. 1 (1995), pp 289-300;

РАЗРАБОТКА АЛГОРИТМА ШИФРОВАНИЯ НА ОСНОВЕ ASCII КОДА

Автор: Никифоров Иван Сергеевич, студент 1 курса филиала «Протвино» Государственного университета «Дубна»

Научный руководитель: Кульман Татьяна Николаевна, к.т.н., доцент Государственного университета «Дубна» (филиал Протвино)

Аннотация

На примере реального проекта рассматривается алгоритм шифрования на основе ASCII кода. Описывается и показывается работа алгоритма

Annotation

An ASCII code-based encryption algorithm is considered as an example of a real project. The algorithm is described and shown

Ключевые слова: алгоритм, шифрование, ASCII

Keywords: algorithm, encryption, ASCII