

ПРОВЕРКА ГИПОТЕЗ О ХАРАКТЕРЕ ВЕРОЯТНОСТНОГО РАСПРЕДЕЛЕНИЯ С ПОМОЩЬЮ ПРОГРАММЫ STADIA

Д.И. Ермаченко

Л.В. Незамова, ассистент каф.ВМиП

Донецкий национальный технический университет

Darina_home@list.ru

Проверка статистической гипотезы заключается в том, чтобы установить, не противоречат ли данные выборочных наблюдений выдвинутой гипотезе. С этой целью производится количественная оценка степени достоверности предлагаемой гипотезы, которая осуществляется с помощью специально построенного статистического критерия. В данной работе показан пример выполнения проверки гипотез о характере вероятностного распределения с помощью программы STADIA.

The verification of statistical hypothesis consists in that, to set, whether the data of selective supervisions to the pulled out hypothesis do not contradict. To that end quantitative estimation of degree of authenticity of offered hypothesis, which is carried out by the specially built statistical criterion is produced. An example of implementation of verification of hypotheses is shown in the given work about nature of the probabilistic distributing by the STADIA program.

Перевірка статистичної гіпотези полягає в тому, щоб встановити, чи не суперечать дані вибіркових спостережень висунутій гіпотезі. З цією метою виробляється кількісна оцінка ступеня достовірності запропонованої гіпотези, яка здійснюється за допомогою спеціально побудованого статистичного критерію. В даній роботі показаний приклад виконання перевірки гіпотез про характер розподілу імовірності за допомогою програми STADIA.

ВЕРОЯТНОСТНОЕ РАСПРЕДЕЛЕНИЕ, ГИСТОГРАММА, КРИТЕРИЙ СОГЛАСИЯ

Статистический анализ является одним из этапов научных исследований. Статистические программы являются трудоемкими при реализации, их цена часто недоступна индивидуальному пользователю.

Международный рынок использует различные компьютерные технологии обработки и анализа данных – статистические программные продукты (СПП). СПП можно разделить на профессиональные, универсальные и специализированные. Профессиональные пакеты включают большое количество методов анализа (SAS, BMDP). Универсальные пакеты располагают диапазоном стат.методов, не ориентированных на конкретную предметную область (STATISTICA, SPSS, STADIA ...). Специализированные (BIOSTAT, DATASCOPE ...) ориентированы на узкую область анализа данных [1].

Используемая в данной работе СПП STADIA включает наиболее часто применяемые статистические методы, значительная часть которых изучается в

вузовских курсах и описана в учебной литературе.

Первым этапом любого статистического исследования является статистическое наблюдение. Оно сводится к сбору данных о массовых явлениях путем регистрации их признаков. При обработке этих экспериментальных данных иногда возникает необходимость аппроксимировать эмпирическое распределение соответствующим законом распределения. Выбор предполагаемого теоретического закона из возможных альтернативных вариантов на основе выборочной совокупности является предположением. Задача установления теоретического закона распределения случайной величины по эмпирическому распределению требует решения 2 подзадач: определить вид и оценки параметров предполагаемого теоретического закона распределения. Проверка соответствия предполагаемого закона распределения экспериментальным данным выполняется на основании критериев согласия [2]. Выполним проверку статистической гипотезы, чтобы установить, не противоречат ли данные выборочных наблюдений выдвинутой гипотезе о законе распределения. На основании оценки гистограммы закона распределения делается предположение о виде закона распределения, выдвигается нулевая гипотеза. На основании выборки независимых наблюдений выполним проверку статистических гипотез о характере законов распределения генеральных совокупностей с использованием СПП STADIA.

В качестве выборок используем данные табл. 1 и табл. 2.

Таблица 1. Котировка акций (+, -)

50,9	18,7	80,5	35,5	28	32,5	96,8	55	79	69,3
66,7	81,8	108	-4,97	92,9	95,7	103	60,8	-50	6
20	-100	-10	-41,8	67,6	-18,2	-10,2	89,7	31,7	202
94,3	33,9	-52,8	-32,8	58	50	52	75	81,2	30
72	87,5	10	63	86	45	-73,5	69	18	-9,32
45	65	30,8	-8,13						

Таблица 2. % ответов не верно/верно при повторном тестировании (в табл. \approx 2 знака)

0,58	0,49	2,38	4,53	1,22	0,18	0,58	0,18	0,86	0,54	0,52	4,34	0,24
1,92	0,94	0,11	0,25	1,71	0,06	0,14	2,63	0,83	1,62	1,43	0,17	1,53
0,39	0,49	0,11	0,15	0,78	0,04	0,26	0,07	0,66	0,68	0,06	0,05	0,35
0,19	0,05	0,34	0,03	2,94	0,88							

После запуска системы STADIA необходимо ввести выборочные совокупности (ВС). Исходя из требований размещения данных в программе, вводим ВС по столбцам (рис.1). При проверке гипотез необходимо задать уровень значимости. Определим для данной задачи уровень значимости 0,05.

Согласно выше изложенному, решим первую подзадачу – определим вид предполагаемого теоретического закона распределения. Для построения графических объектов используем внутреннее меню программы, которое позволяет построить графики данных по категориям: научные, деловые и т.д.

	x10	x3	x7	x6	x7	x8	x7	x1
6	32,5	0,177						
7	96,8	0,575						
8	55	0,178						
9	79	0,864						
10	69,3	0,538						
11	66,7	0,516						
12	81,8	4,34						
13	108	0,234						
14	-4,97	1,92						

Рисунок 1 - Таблица исходных данных

Построим для каждой выборки (x10 и x3) гистограммы. Выполним действия по пунктам меню *Статистика->Гистограмма/нормальность*. В окне “Анализ переменной” выделим переменную, например x10. В новом окне “Гистограмма” программа предлагает рассчитанные значения: число интервалов, левая и правая граница интервалов. Пользователь имеет возможность вносить коррективы в предложенные значения. После подтверждения введенных значений, пользователю будут выведены на экран результаты расчетов (рис.3) и построена гистограмма (рис.2).

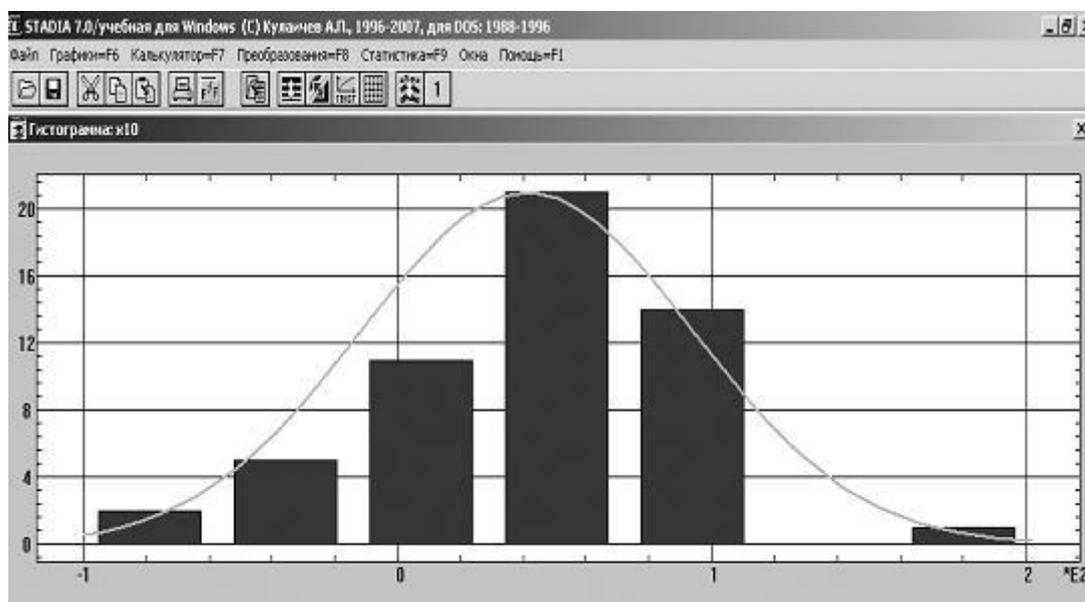


Рисунок 2 – Гистограмма выборки x10

По виду гистограммы можно сделать предположение (принять нулевую гипотезу) о том, что совокупность распределена по нормальному закону.

Одним из критериев согласия, используемых при решении вопроса о принятии нулевой гипотезы является критерий Пирсона. Для принятия нулевой гипотезы должно выполняться равенство $\chi^2 < \chi_{кр}^2$.

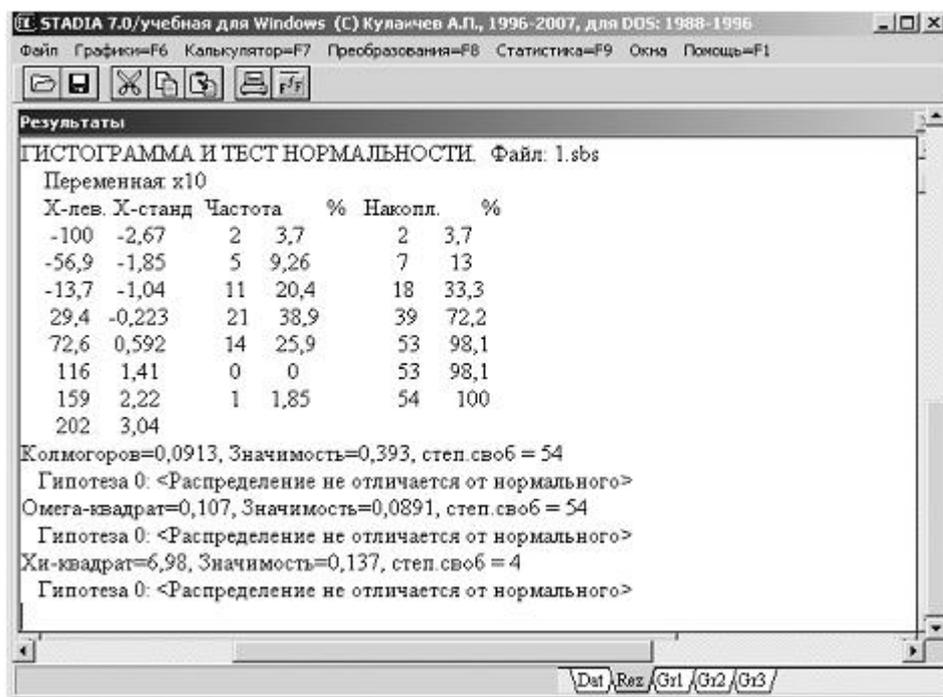


Рисунок 3 – Результаты расчетов

В расчетах, для каждого интервала указаны: X-лев. – левая граница интервала в исходных единицах; X-станд. – левая граница интервала; Частота – число выборочных значений, попавших в интервал; % - относительная частота; Накопл – накопленное число выборочных значений; % - относительная накопленная частота. На экран выводятся статистики критериев согласия: - Колмогорова, омега-квадрат, Пирсона. Для каждого критерия согласия, на указанным уровнем значимости, выводится сообщение о принятии нулевой или альтернативной гипотезы. С учетом принятого в задаче уровня значимости 0,05 значение статистики χ^2 не превышает критического значения меры расхождения, что свидетельствует о согласовании данных с гипотезой о нормальном распределении.

Вторая подзадача в решении рассматриваемого вопроса – определение оценок параметров распределения (рис 4).

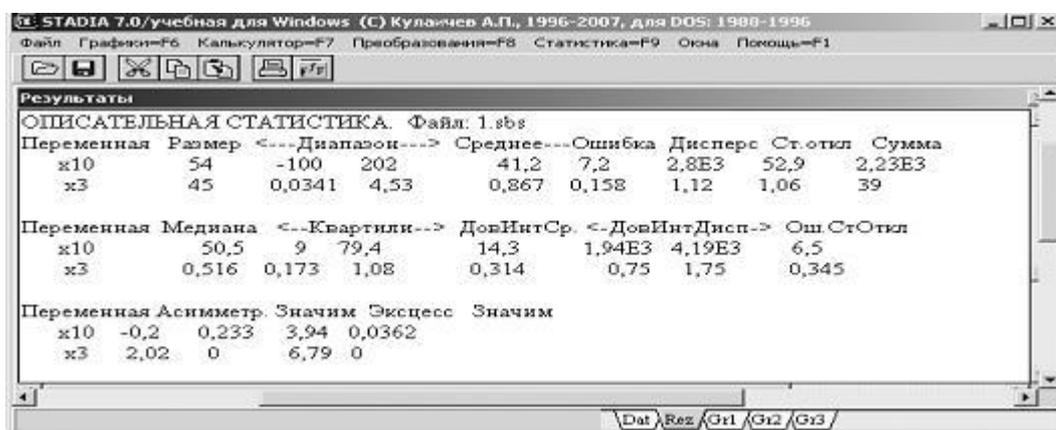


Рисунок 4 – Оценки параметров

Оценки параметров получили для выборки x_{10} и x_3 . Для получения этих значений выполнили действия по пунктам меню *Статистика->Описательная статистика*. Для распределения x_3 значимость моментов 3-го и 4-го порядка равна 0. Построим гистограмму распределения x_3 , чтобы определить вид предполагаемого теоретического закона распределения (рис. 5).

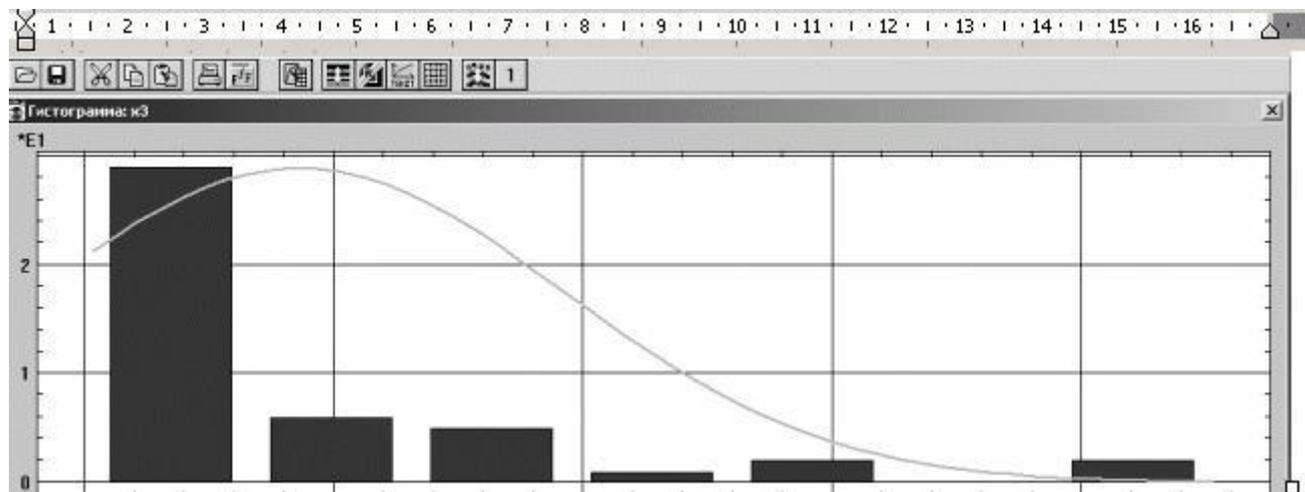


Рисунок 5 – Гистограмма выборки x_3

По виду гистограммы можно сделать предположение (принять нулевую гипотезу) о том, что совокупность распределена по экспоненциальному закону.

Для проверки этой гипотезы выполним действия по пунктам меню *Статистика->Согласие распределений*. Из меню выбора теоретических распределений выбираем пункт 2=экспоненциальное.

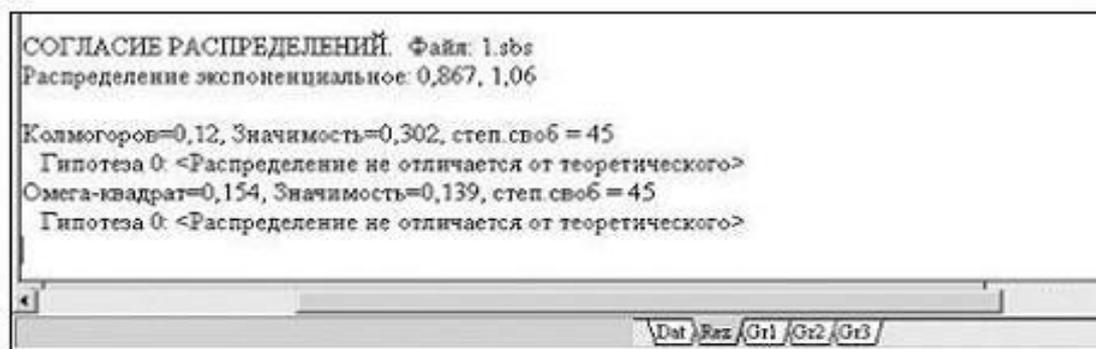


Рисунок 6 – Гистограмма выборки x_3

Литература

1. Халафян А.А. STATISTICA 6. Статистический анализ данных. – М: ООО “Бином-Пресс”, 2008 г. – 512с.
2. Четыркин Е.М., Калихман И.Л. Вероятность и статистика – М: Финансы и статистика, 1982 г. – 356.
3. Реннер А.Г., Зиновьева О.А., Аралбаева Г.Г. Проверка гипотез о характере распределения. Оренбург, 2002. – 45 с.