Исследование распределений статистик и мощности критериев однородности в случае больших массивов данных

**Оглавление**

[**Введение** 3](#_Toc514019298)

[**1.** **Критерии проверки однородности законов распределения** 5](#_Toc514019299)

[**1.1.** **Общая постановка** 5](#_Toc514019300)

[**1.2.** **Критерий Смирнова** 5](#_Toc514019301)

[**1.3.** **Критерий Лемана-Розенблатта** 5](#_Toc514019302)

[**1.4.** **Критерий Андерсона-Дарлинга** 5](#_Toc514019303)

[**2.** **Исследование распределения статистик** 5](#_Toc514019304)

[**3.** **Исследование мощностей критериев** 5](#_Toc514019305)

[**Заключение** 5](#_Toc514019306)

[**Список литературы** 5](#_Toc514019307)

[**Приложение А. Программные модули** 5](#_Toc514019308)

# **Введение**

В прикладных исследованиях довольно часто возникает необходимость выяснить, имеют ли различия генеральные совокупности, из которых взяты две независимые выборки. В математической статистике данная задача формулируется как проверка гипотезы об однородности законов распределения вероятностей. Необходимость проверки данных гипотез появляется в различных ситуациях, когда хотят удостовериться в неизменности (или напротив в изменении) статистических свойств некоторого объекта или процесса после целенаправленного изменения фактора или факторов (методики, технологии и т.д.), неявным образом влияющих на исследуемый объект. Иными словами, проверяется изменение или наоборот сохранение статистических показателей объекта или процесса до некоторого оказанного воздействия и после с течением времени. Например, надо выяснить, влияет ли способ упаковки некоторых деталей на заводе на их потребительские качества через год после хранения. Или другой пример применения исследований однородности: в маркетинге важно выделить сегменты потребительского рынка.

В случае если установлена однородность двух выборок, то вполне вероятно группировка сегментов, из которых они взяты, в один. В последующем это позволит воплотить в жизнь по отношению к ним схожую рекламную политику (проводить одни и те же маркетинговые  процедуры и т.п.). В случае если же установлено отличие, то поведение потребителей в двух сегментах различно, объединять эти сегменты невозможно, и могут понадобиться различные рекламные компании, своя для каждого из этих сегментов.

На практике чаще всего приходится иметь дело с данными ограниченной точности. Зачастую, это целые числа, или данные с одним, двумя знаками после запятой. При больших объемах выборок, количество повторений в выборках тоже становится большим. Становится интересно, можно ли руководствоваться данными по исследованию критериев однородности для таких выборок. Подчиняются ли статистики критериев предельным распределениям, и при каких объемах выборок можно реально пользоваться этими предельными распределениями статистик критериев. Исследования распределений статистик и мощностей критериев однородности подробно рассматривались в работах [1, 2]. Таким образом, целью данной работы является исследование критериев однородности на данных ограниченной точности. Исследования проводятся с использованием метода компьютерного моделирования.

# **Критерии проверки однородности законов распределения**

## **Общая постановка**

С необходимостью решения задач проверки гипотез о принадлежно­сти двух выборок случайных величин одной и той же генеральной совокуп­ности постоянно сталкиваются при анализе случайных ошибок средств из­мерений, при статистическом управлении качеством процессов. Такая задача естественно возникает при поверке средств измерений, когда пытаются убе­диться в том, что закон распределения случайных ошибок измерений не пре­терпел существенных изменений по истечении некоторого интервала вре­мени.

Задача проверки однородности двух выборок формулируется следующим образом. Пусть имеются две упорядоченные по не убыванию выборки размером  и  :

 и .

Для определенности обычно полагают, что , но это совсем необязательно. Проверяется гипо­теза о том, что две выборки извлечены из одной и той же генеральной сово­купности, т. е. :  при любом .

## **Критерий Смирнова**

## **Критерий Лемана-Розенблатта**

## **Критерий Андерсона-Дарлинга**

# **Исследование распределения статистик**

# **Исследование мощностей критериев**

# **Заключение**

# **Список литературы**

1. Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению / Б.Ю. Лемешко. – М: ИНФРА–М, 2016. – 207 с.
2. Лемешко Б. Ю. О сходимости распределений статистик и мощности критериев однородности Смирнова и Лемана–Розенблатта / Б. Ю. Лемешко, С. Б. Лемешко // Измерительная техника. – 2005. – № 12. – С. 9–14.
3. Большев Л. Н. Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. – М. : Наука, 1983. – 416 с.

# **Приложение А. Программные модули**