

Санкт-Петербургский Политехнический Университет Петра
Великого
Физико-механический институт

Высшая школа прикладной математики и вычислительной
физики

Лабораторная работа №5
по дисциплине “Математическая статистика”

Обучающаяся:

А.Д. Балакшина
(группа 5030102/20101)

Преподаватель:

А.Н. Баженов

Санкт-Петербург

2025

Содержание

1	Формулировка задания	3
2	Используемые формулы	3
3	Выполнение работы	4
4	Результаты	4
4.1	Результаты для выборки: Нормальное распределение $N=100$	4
4.2	Результаты для выборки: Нормальное распределение $N=20$	4
4.3	Результаты для выборки: Равномерное распределение $N=100$	5
4.4	Результаты для выборки: Равномерное распределение $N=20$	5
5	Анализ результатов проверки гипотез о нормальности распределения	6
5.1	Анализ нормально распределённых выборок	6
5.2	Анализ равномерно распределённых выборок	6
6	Общие выводы	7

1 Формулировка задания

1. Сгенерировать выборку объёмом 100 элементов для нормального распределения $N(x, 0, 1)$.
2. По сгенерированной выборке оценить параметры μ и σ нормального закона методом максимального правдоподобия.
3. В качестве основной гипотезы H_0 считать, что сгенерированное распределение имеет вид $N(x, \hat{\mu}, \hat{\sigma})$.
4. Проверить основную гипотезу, используя критерий согласия χ^2 с уровнем значимости $\alpha = 0.05$.
5. Исследовать точность (чувствительность) критерия χ^2 :
 - Сгенерировать выборки равномерного распределения объёмом 20, 100 элементов, нормального распределения объёмом 20 элементов.
 - Проверить их на нормальность.

2 Используемые формулы

1. **Функция плотности нормального распределения:**

$$N(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

2. **Статистика критерия χ^2 :**

$$\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

где:

- n_i – наблюдаемые частоты,
- p_i – теоретические вероятности,
- n – объём выборки,
- k – количество интервалов.

3. **Квантиль распределения χ^2 :**

$$\chi^2_{1-\alpha}(k-1)$$

где:

- α – уровень значимости,
- $k-1$ – степени свободы.

3 Выполнение работы

Лабораторная работа выполнена на языке программирования Python 3.12 с использованием библиотек numpy, scipy, pandas.
Программа отработала корректно.

4 Результаты

4.1 Результаты для выборки: Нормальное распределение N=100

Оценки ММП параметров нормального распределения: $\hat{\mu} = 0.061$, $\hat{\sigma} = 0.885$

Интервал	Наблюдаемая частота	Ожидаемая частота	$(O - E)^2/E$
(-inf, -0.80]	14	16.7	0.427
(-0.80, -0.32]	21	16.7	1.127
(-0.32, 0.06]	15	16.7	0.167
(0.06, 0.44]	18	16.7	0.107
(0.44, 0.92]	17	16.7	0.007
(0.92, inf]	15	16.7	0.167

Таблица 1: Таблица частот для проверки нормальности выборки (Нормальное распределение N=100)

Статистика χ^2 : 2.002

Критическое значение: 7.815

Вывод: Не отвергаем H_0

4.2 Результаты для выборки: Нормальное распределение N=20

Оценки ММП параметров нормального распределения: $\hat{\mu} = -0.088$, $\hat{\sigma} = 0.841$

Интервал	Наблюдаемая частота	Ожидаемая частота	$(O - E)^2/E$
(-inf, -0.90]	5	3.3	0.833
(-0.90, 0.27]	6	10.0	1.600
(0.27, inf]	9	6.7	0.817

Таблица 2: Таблица частот для проверки нормальности выборки (Нормальное распределение N=20)

Статистика χ^2 : 3.250

Критическое значение: 11.070

Вывод: Не отвергаем H_0

4.3 Результаты для выборки: Равномерное распределение N=100

Оценки ММП параметров нормального распределения: $\hat{\mu} = 0.537$, $\hat{\sigma} = 0.296$

Интервал	Наблюдаемая частота	Ожидаемая частота	$(O - E)^2/E$
(-inf, 0.25]	21	16.7	1.127
(0.25, 0.41]	14	16.7	0.427
(0.41, 0.54]	8	16.7	4.507
(0.54, 0.66]	21	16.7	1.127
(0.66, 0.82]	14	16.7	0.427
(0.82, inf]	22	16.7	1.707

Таблица 3: Таблица частот для проверки нормальности выборки (Равномерное распределение N=100)

Статистика χ^2 : 9.322

Критическое значение: 7.815

Вывод: Отвергаем H_0

4.4 Результаты для выборки: Равномерное распределение N=20

Оценки ММП параметров нормального распределения: $\hat{\mu} = 0.550$, $\hat{\sigma} = 0.233$

Интервал	Наблюдаемая частота	Ожидаемая частота	$(O - E)^2/E$
(-inf, 0.45]	10	6.7	1.667
(0.45, 0.78]	5	10.0	2.500
(0.78, inf]	5	3.3	0.833

Таблица 4: Таблица частот для проверки нормальности выборки (Равномерное распределение N=20)

Статистика χ^2 : 5.000

Критическое значение: 11.070

Вывод: Не отвергаем H_0

5 Анализ результатов проверки гипотез о нормальности распределения

5.1 Анализ нормально распределённых выборок

Для выборки из нормального распределения объемом $N = 100$ получены следующие результаты:

- Оценки параметров: $\hat{\mu} = 0.061$, $\hat{\sigma} = 0.885$ близки к теоретическим значениям (0 и 1 соответственно)
- Значение статистики $\chi^2 = 2.000$ существенно меньше критического значения 7.815
- Наблюдаемые частоты хорошо согласуются с ожидаемыми (максимальное расхождение 1.127)

Вывод о нормальности распределения подтверждается, гипотеза H_0 не отвергается.

Для выборки $N = 20$:

- Оценки параметров: $\hat{\mu} = -0.088$, $\hat{\sigma} = 0.841$ также близки к теоретическим
- Статистика $\chi^2 = 3.250$ меньше критического значения 11.070
- Несмотря на малое количество наблюдений, критерий показал адекватные результаты

Гипотеза о нормальности не отвергается, хотя для малых выборок мощность критерия снижается.

5.2 Анализ равномерно распределённых выборок

Для равномерного распределения ($N = 100$):

- Оценки $\hat{\mu} = 0.537$, $\hat{\sigma} = 0.296$ отличаются от параметров нормального распределения
- Статистика $\chi^2 = 9.320$ превышает критическое значение 7.815
- Заметные расхождения в интервалах $(0.41, 0.54]$ и $(0.82, +\infty)$

Критерий корректно отвергает гипотезу H_0 о нормальности.

Для малой выборки ($N = 20$):

- Статистика $\chi^2 = 5.000$ меньше критического значения 11.070
- Несмотря на равномерное распределение, критерий не смог его обнаружить

- Это демонстрирует низкую мощность критерия χ^2 для малых выборок

Гипотеза H_0 не отвергается, что является ошибкой II рода.

6 Общие выводы

- Критерий χ^2 хорошо работает для больших выборок ($N = 100$), правильно идентифицируя как нормальные, так и ненормальные распределения
- Для малых выборок ($N = 20$) мощность критерия недостаточна, что приводит к ошибкам II рода. Критерий чувствителен к объёму данных - для надёжных выводов рекомендуется использовать выборки большего объёма.
- Оценки ММП параметров для нормального распределения показывают хорошую точность