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

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.
Программа отработала корректно.

4 Результаты

4.1 Нормальное распределение

i	Границы интервалов	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	$[-\infty, -1.23]$	4	0.214	4.3	-0.3	0.02
2	$[-1.23, -0.45]$	4	0.233	4.7	-0.7	0.09
3	$[-0.45, -0.14]$	4	0.102	2.0	2.0	1.89
4	$[-0.14, 0.47]$	4	0.189	3.8	0.2	0.01
5	$[0.47, \infty]$	4	0.263	5.3	-1.3	0.30

Таблица 1: $n = 20$, $\mu = -0.29$, $\sigma = 1.2$

χ^2 наблюдаемое = 2.3, критическое = 6.0
Гипотеза принимается

i	Границы интервалов	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	$[-\infty, -0.91]$	15	0.155	15.5	-0.5	0.01
2	$[-0.91, -0.50]$	14	0.118	11.8	2.2	0.41
3	$[-0.50, -0.13]$	14	0.135	13.5	0.5	0.01
4	$[-0.13, 0.28]$	14	0.167	16.7	-2.7	0.43
5	$[0.28, 0.60]$	14	0.121	12.1	1.9	0.30
6	$[0.60, 1.05]$	14	0.137	13.7	0.3	0.01
7	$[1.05, \infty]$	15	0.167	16.7	-1.7	0.17

Таблица 2: $n = 100$, $\mu = 0.10$, $\sigma = 1.0$

χ^2 наблюдаемое = 1.3, критическое = 9.5
Гипотеза принимается

4.2 Равномерное распределение

i	Границы интервалов	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	$[-\infty, -0.61]$	4	0.199	4.0	0.0	0.00
2	$[-0.61, -0.07]$	4	0.204	4.1	-0.1	0.00
3	$[-0.07, 0.38]$	4	0.196	3.9	0.1	0.00
4	$[0.38, 1.24]$	4	0.286	5.7	-1.7	0.51
5	$[1.24, \infty]$	4	0.116	2.3	1.7	1.22

Таблица 3: $n = 20$, $\mu = -0.29$, $\sigma = 1.2$

χ^2 наблюдаемое = 1.7, критическое = 6.0
Гипотеза принимается

i	Границы интервалов	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	$[-\infty, -1.06]$	15	0.099	9.9	5.1	2.57
2	$[-1.06, -0.35]$	14	0.193	19.3	-5.3	1.48
3	$[-0.35, -0.08]$	14	0.105	10.5	3.5	1.15
4	$[-0.08, 0.55]$	14	0.256	25.6	-11.6	5.28
5	$[0.55, 0.95]$	14	0.138	13.8	0.2	0.00
6	$[0.95, 1.27]$	14	0.083	8.3	5.7	3.92
7	$[1.27, \infty]$	15	0.125	12.5	2.5	0.50

Таблица 4: $n = 100$, $\mu = 0.10$, $\sigma = 1.0$

χ^2 наблюдаемое = 14.9, критическое = 9.5
Гипотеза отвергается

5 Анализ результатов проверки гипотез о нормальности распределения

5.1 Анализ нормально распределённых выборок

Объем выборки $n = 20$:

- Наблюдаемое значение $\chi^2 = 2.3$ меньше критического значения 6.0.
- Гипотеза о нормальности распределения **принимается**, что согласуется с ожиданиями, так как данные были сгенерированы из нормального распределения.
- Несмотря на небольшой объем выборки, критерий корректно не отвергает гипотезу H_0.

Объем выборки $n = 100$:

- Наблюдаемое значение $\chi^2 = 1.3$ значительно меньше критического значения 9.5.
- Гипотеза о нормальности распределения **принимается**.
- Увеличение объема выборки улучшает точность оценки, и критерий χ^2 демонстрирует высокую чувствительность к нормальному распределению.

5.2 Анализ равномерно распределённых выборок

Объем выборки $n = 20$:

- Наблюдаемое значение $\chi^2 = 1.7$ меньше критического значения 6.0.
- Гипотеза о нормальности распределения **принимается**, хотя данные были сгенерированы из равномерного распределения, что является ошибкой II рода.
- Это может быть связано с недостаточным объемом выборки для выявления отклонения от нормальности.

Объем выборки $n = 100$:

- Наблюдаемое значение $\chi^2 = 14.9$ превышает критическое значение 9.5.
- Гипотеза о нормальности распределения **отвергается**, что соответствует ожиданиям, так как данные имеют равномерное распределение.
- Критерий χ^2 демонстрирует хорошую чувствительность к отклонениям от нормальности при увеличении объема выборки.

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 объему данных - для надежных выводов рекомендуется использовать выборки большего объёма.
- Оценки ММП параметров для нормального распределения показывают хорошую точность