

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Омский государственный технический университет»

Факультет информационных технологий и компьютерных систем
Кафедра «Прикладная математика и фундаментальная информатика»

Лабораторная работа
по дисциплине «Теория вероятностей и математическая статистика»

Студента	Курпенова Куата Ибраимовича
	<small>фамилия, имя, отчество полностью</small>
Курс	3, группа ФИТ-222
Направление	02.03.02 Прикладная математика и фундаментальная информатика
	<small>код, наименование</small>
Руководитель	доц., канд. тех. наук
	<small>должность, ученая степень, звание</small>
	Болдовская Т. Е.
	<small>фамилия, инициалы</small>
Выполнил	
	<small>дата, подпись студента</small>
Проверил	
	<small>дата, подпись руководителя</small>

Проверка гипотезы о законе распределения генеральной совокупности

Данные из выборки

-71, -44, -58, -49, -53, -72, -59, -59, -62, -78, -62, -47, -46, -74, -62, -52, -45, -66, -48, -59, -44, -56, -58, -34, -55, -57, -51, -59, -58, -53, -67, -38, -90, -68, -36, -53, -35, -50, -74, -87, -50, -72, -50, -64, -51, -72, -52, -67, -58, -38, -79, -54, -37, -91, -62, -73, -62, -37, -61, -60, -64, -71, -64, -63, -72, -51, -47, -47, -49, -60, -74, -42, -45, -52, -67, -50, -64, -65, -58, -21, -67, -58, -36, -78, -63, -59, -59, -77, -77, -54, -69, -40, -41, -52, -79, -29, -59, -27, -59, -23, -63, -70, -58, -74, -67, -58, -65, -68, -59, -46, -58, -73, -58, -50, -72, -35, -57, -55, -68, -79, -58, -56, -62, -63, -39, -66, -70, -58, -60, -64, -54, -68, -69, -50, -30, -67, -71, -44, -84, -44, -71, -68, -86, -70, -44, -55, -41, -74, -48, -71, -65, -50, -46, -53, -49, -66, -69, -30, -59, -59, -39, -72, -65, -74, -79, -70, -48, -45, -69, -56, -54, -60.

Вариационный ряд

-91, -90, -87, -86, -84, -79, -79, -79, -79, -78, -78, -77, -77, -74, -74, -74, -74, -74, -73, -73, -72, -72, -72, -72, -72, -72, -71, -71, -71, -71, -71, -70, -70, -70, -70, -69, -69, -69, -69, -68, -68, -68, -68, -68, -67, -67, -67, -67, -67, -67, -66, -66, -66, -65, -65, -65, -65, -64, -64, -64, -64, -64, -63, -63, -63, -63, -62, -62, -62, -62, -62, -61, -60, -60, -60, -60, -59, -59, -59, -59, -59, -59, -59, -59, -59, -59, -58, -58, -58, -58, -58, -58, -58, -58, -58, -58, -58, -58, -57, -57, -56, -56, -56, -55, -55, -55, -54, -54, -54, -54, -53, -53, -53, -53, -52, -52, -52, -52, -51, -51, -51, -50, -50, -50, -50, -50, -50, -50, -49, -49, -49, -48, -48, -48, -47, -47, -47, -46, -46, -46, -45, -45, -45, -44, -44, -44, -44, -44, -44, -42, -41, -41, -40, -39, -39, -38, -38, -37, -37, -36, -36, -35, -35, -34, -30, -30, -29, -27, -23, -21.

Интервальный статистический ряд

Количество интервалов в интервальном статистическом ряде рассчитывается по формуле Стёрджеса:

$$n = 1 + 3.322 \cdot \log_{10} N$$

В нашем случае формула принимает вид:

$$1 + 3.322 \cdot \log_{10} 172 = 9$$

Полигон и гистограмма относительных частот

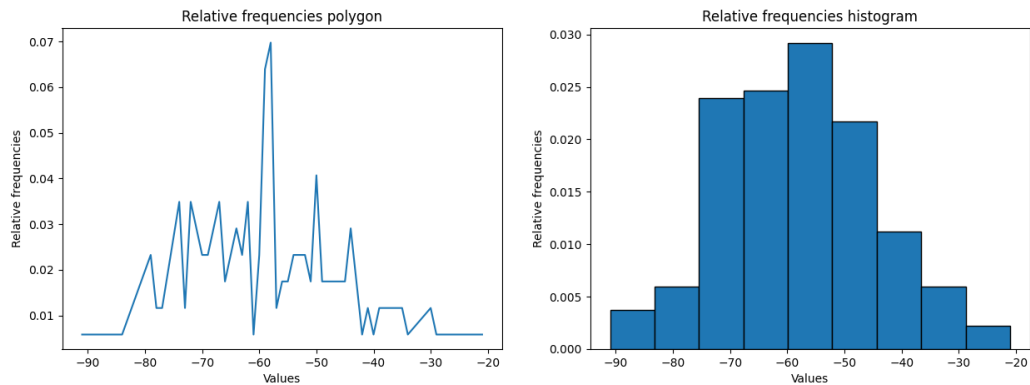


Рисунок 1 – Графики полигона и гистограммы относительных частот

График эмпирической функции распределения

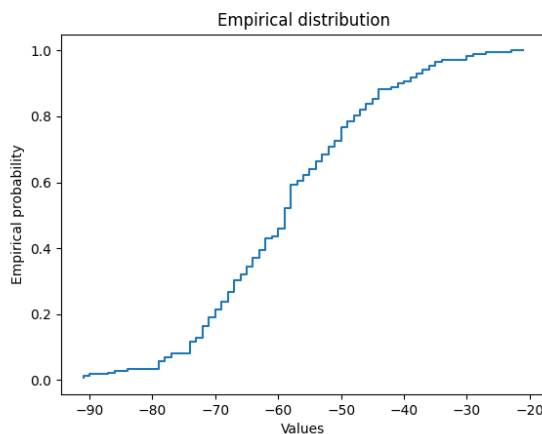


Рисунок 2 – График эмпирической функции распределения

Числовые характеристики выборки

Формула для нахождения выборочного среднего:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Формула для нахождения исправленной выборочной дисперсии:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Мода и медиана находятся в программном коде без формул.

Формула для нахождения эксцесса:

$$E(X) = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - 3$$

Формула для нахождения асимметрии:

$$As = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3$$

```
[+] Sample average: -58.2907
[+] Corrected sample variance: 177.9852
[+] Mode: -58.0000
[+] Median: -59.0000
[+] Kurtosis: -0.0316
[+] Skewness: 0.2517
```

Листинг 1 – Значения параметров выборки

Гипотеза о распределении генеральной совокупности

Я предполагаю, что распределение является нормальным, так как гистограмма относительных частот похожа на нормальное распределение, а график эмпирического распределения похож на сигмоиду.

Оценка параметров нормального распределения генеральной совокупности

Формула оценки среднего:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

Формула оценки стандартного отклонения:

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2}$$

Теоретические аналоги функции

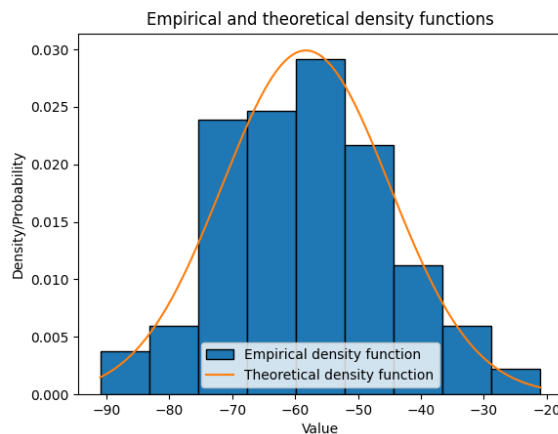


Рисунок 3 – График теоретического аналога плотности распределения

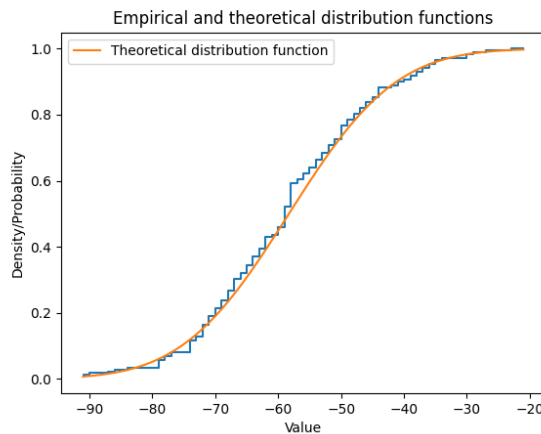


Рисунок 4 – График теоретического аналога эмпирической функции

Выполнение правила «трёх сигма»

Правило трёх сигма (или правило трёх стандартных отклонений) используется в статистике для описания распределения данных в нормальном распределении. Оно утверждает, что примерно 68% данных находятся в пределах одного стандартного отклонения от среднего, около 95% — в пределах

двух стандартных отклонений, и примерно 99.7% — в пределах трёх стандартных отклонений.

В моём случае границы интервалов имеют значения: -98.3140 и -18.2674.

Проверка по критерию Пирсона

Критерий Пирсона, также известный как χ^2 (хи-квадрат) тест, используется в статистике для проверки гипотез о распределении категориальных данных. Основная цель этого критерия — определить, насколько наблюдаемые частоты (или количество случаев) в различных категориях соответствуют ожидаемым частотам, которые предполагаются в соответствии с некоторой теоретической моделью.

Формула для расчёта статистики хи-квадрат:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

[+] Statistic: 5.7050

[+] P-value: 0.6802

Листинг 2 – Результат проверки по критерию Пирсона

Так как $P - value$ меньше уровня значимости, делаем вывод, что нулевая гипотеза верна.

Проверка по критерию Колмогорова

Критерий Колмогорова (или тест Колмогорова-Смирнова) — это непараметрический статистический тест, который используется для проверки гипотез о распределении данных. Он позволяет оценить, насколько хорошо наблюдаемые данные соответствуют предполагаемому теоретическому распределению, или сравнить два эмпирических распределения.

Формула для статистики теста Колмогорова:

$$D = \max_x |F_n(x) - F(x)|$$

[+] Statistic: 0.0785

[+] P-value: 0.2273

Листинг 3 – Результат проверки по критерию Колмогорова

Так как $P - value$ меньше уровня значимости, делаем вывод, что нулевая гипотеза верна.

Построение доверительного интервала

Формула для построения доверительного интервала:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Формула для построения доверительного интервала при неизвестной дисперсии:

$$\bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$$

В моём случае уровень значимости равен 0.05 и границы интервала равны -60.2845 и -56.2969.