Санкт-Петербургский
Политехнический университет Петра Великого

**Отчет по лабораторным работам №3-4
по дисциплине
"Математическая статистика"**

Студент: Скворцов Владимир Сергеевич
Преподаватель: Баженов Александр Николаевич

Группа: 5030102/10201

Санкт-Петербург
2024

Содержание

1 Постановка задачи

1.1 Боксплот Тьюки

Сгенерировать выборки размером 20 и 100 элементов. Построить для них боксплот Тьюки.

1.2 Доверительные интервалы для параметров нормального распределения

Сгенерировать выборки размером 20 и 100 элементов. Вычислить параметры положения и рассеяния:

- для нормального распределения,
- для произвольного распределения.

2 Теоретическое обоснование

2.1 Функции распределения

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{1}$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \tag{2}$$

- Распределение Стьюдента $t(x, 0, 3)$ с тремя степенями свободы

$$t(x, 0, 3) = \frac{6\sqrt{3}}{\pi(3 + t^2)^2} \tag{3}$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \tag{4}$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \le \sqrt{3} \\ 0, & |x| > \sqrt{3} \end{cases} \tag{5}$$

2.2 Боксплот Тьюки

Боксплот (англ. box plot) — график, использующихся в описательной статистике, компактно изобрадающий одномерное распределение вероятностей. Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Границами ящика служат первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выброса). Длину «усов» определяют разность первого квартиля и полутора межквартальных расстояний и сумма третьего квартиля и полутора межквартальных расстояний. Формула имеет вид

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1),\ X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \tag{6}$$

где X_1 — нижняя граница уса, X_2 — верхняя граница уса, Q_1 — первый квартиль, Q_3 - третий квартиль. Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков. Выбросами считаются величины , такие что:

$$\left[\begin{array}{l} x < X_1^T \\ x > X_2^T \end{array}\right. \tag{7}$$

2.3 Доверительные интервалы для параметров нормального распределения

Пусть $F_T(x)$ — функция распределения Стьюдента с $n-1$ степенями свободы. Полагаем, что $2F_T(x)-1 = 1-\alpha$, где α — выбранный уровень значимости. Тогда $F_T(x) = 1-\alpha/2$. Пусть $st_{1-\alpha/2}(n-1)$ — квантиль распределения Стьюдента с $n-1$ степенями свободы и порядка $1-\alpha/2$. Тогда получаем

$$P\left(\overline{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \overline{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1-\alpha, \tag{8}$$

что и даст доверительный интервал для m с доверительной вероятностью $\gamma = 1\alpha$ для нормального распределения.

Случайная величина $n\frac{s^2}{\sigma^2}$ распределена по закону χ^2 с $n-1$ степенями свободы. Тогда

$$P\left(\overline{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \overline{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1-\alpha, \tag{9}$$

3 Описание работы

Лабораторные работы выполнены с использованием Python и его сторонних библиотек: `numpy`, `pandas`, `matplotlib`, `seaborn`.

Ссылка на GitHub репозиторий: https://github.com/vladimir-skvortsov/spbstu-mathematical-statistics

4 Результаты

4.1 Гистограммы и графики плотности распределения

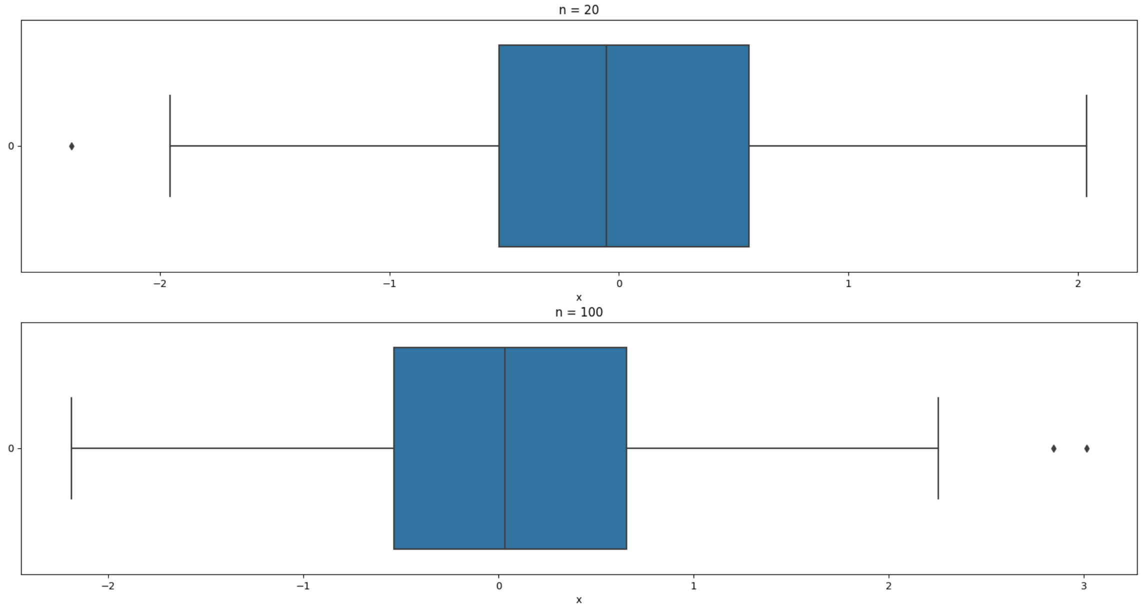

Рис. 1: Нормальное распределение (1)

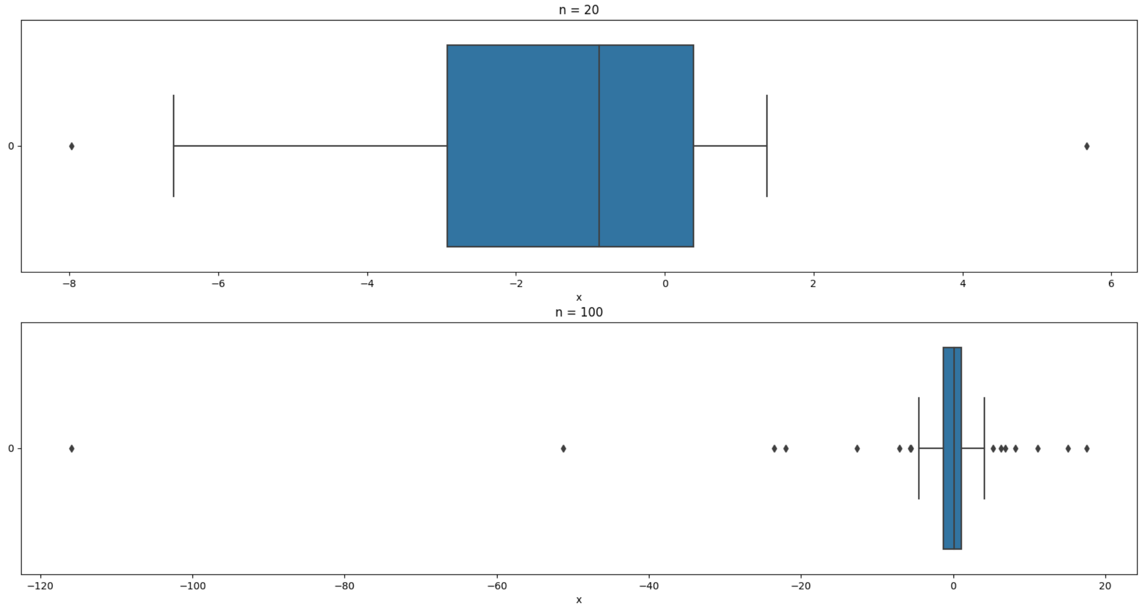

Рис. 2: Распределение Коши (2)

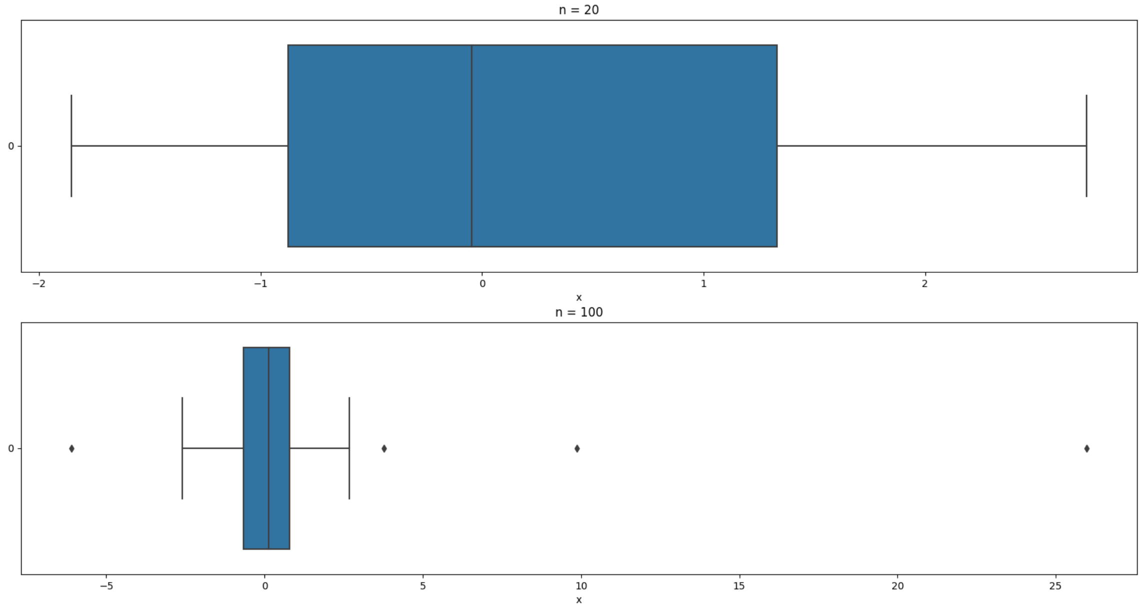

Рис. 3: Распределение Стьюдента (3)

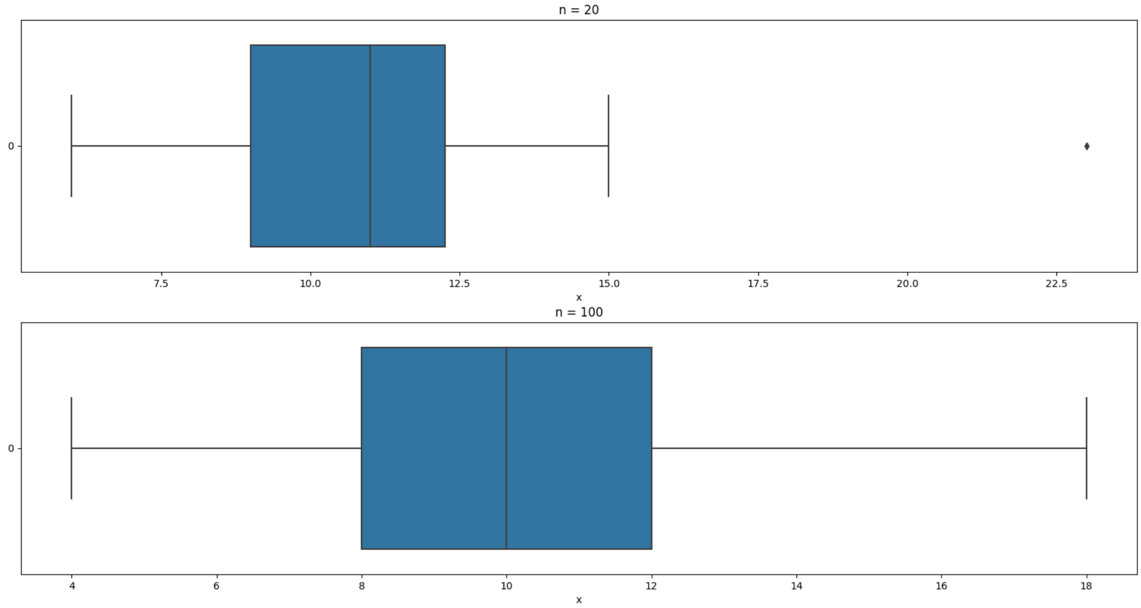

Рис. 4: Распределение Пуассона (4)

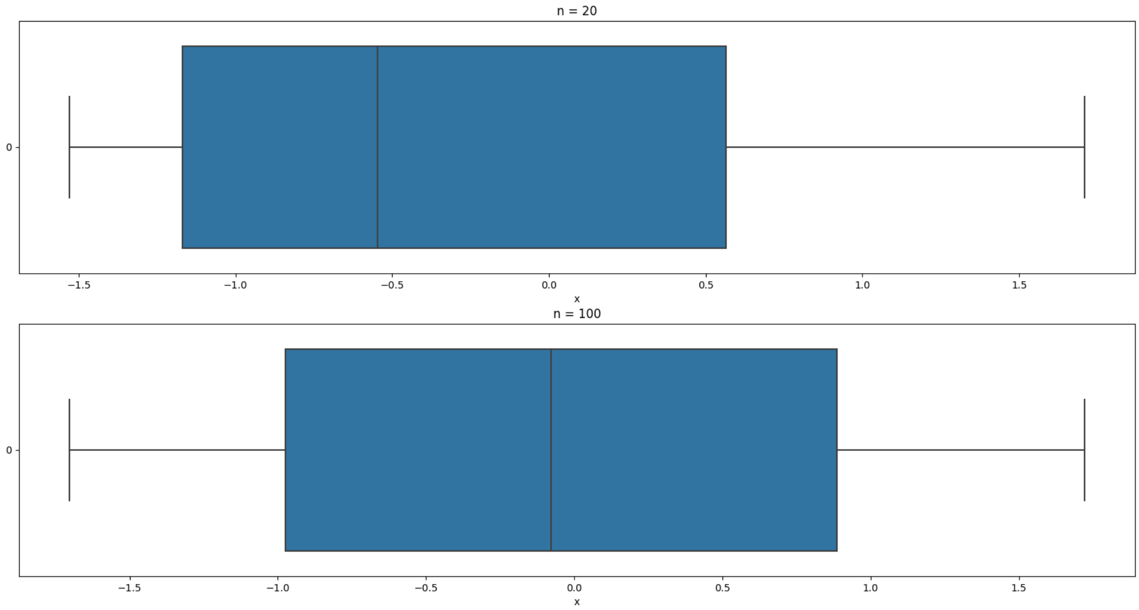

Рис. 5: Равномерное распределение (5)

4.2 Доверительные интервалы для параметров распределений

n = 20	m	σ
	$-0.72 < \mathrm{m} < 0.15$	$0.72 < \sigma < 1.35$
n = 100	m	σ
	$-0.31 < \mathrm{m} < 0.08$	$0.86 < \sigma < 1.14$

Таблица 1: Доверительные интервалы для параметров нормального распределения (1)

n = 20	m	σ
	$0.56 < m < 1.24$	$0.56 < \sigma < 1.06$
n = 100	m	σ
	$0.90 < m < 1.30$	$0.88 < \sigma < 1.17$

Таблица 2: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

5 Выводы

В процессе выполнения лабораторной работы был проведен анализ пяти уникальных распределений: нормальное, Коши, Стьюдента, Пуассона и равномерное. Были сгенерированы выборки разных объемов для каждого из них - 10, 50 и 1000 элементов. Были созданы гистограммы каждого распределения и нанесены на них графики плотности соответствующих распределений, что облегчило наглядное сопоставление формы распределения выборок с их теоретическими аналогами. Были также рассчитаны разные показатели положения и рассеяния для каждой выборки, включая выборочную среднюю величину, медиану, полусумму крайних элементов выборки, полусумму квартилей и усеченное среднее. Использовалась стандартная формула для оценки дисперсии.

На основании полученных данных были сделаны следующие выводы:

1. В случае нормального распределения, оценки показателей положения и рассеяния становятся ближе к их теоретическим значениям по мере увеличения размера выборки.

2. Для распределения Коши показатели положения и рассеяния менее стабильны и могут сильно отличаться от теоретических даже при больших размерах выборки.

3. Распределение Стьюдента при небольших размерах выборки также демонстрирует определенную нестабильность оценок, однако с увеличением размера выборки результаты становятся более точными.

4. Для распределения Пуассона и равномерного распределения, оценки показателей положения и рассеяния кажутся стабильными при любом объеме выборки.

5. В общем, выборочное среднее является наиболее чувствительным к экстремальным значениям по сравнению с медианой, особенно в меньших выборках. Однако с увеличением размера выборки, влияние этих экстремальных значений на среднее значение уменьшается. В то же время, медиана обычно более устойчива к выбросам и мало варьирует с изменением размера выборки.

6. Медиана является чувствительной к типу распределения: в нормальном и распределении Стьюдента медиана равна среднему, в распределении Коши она дает надежные, устойчивые к выбросам оценки, в Пуассоновском приближается к среднему, и в равномерном равна половине суммы минимального и максимального значений.