Санкт-Петербургский
Политехнический университет Петра Великого

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

Построение гистограм различных вероятностных распределений и получение оценок положения и оценок рассеяния соответсвующих распределений

Студент: Белоус Фёдор Васильевич
Преподаватель: Баженов Александр Николаевич

Группа: 5030102/10101

Санкт-Петербург
2024

Содержание

1 Постановка задачи

1.1 Бокс-плот Тьюки

Для 5 распределений:

- Нормальное распределение $N(x, 0, 1)$
- распределение Коши $C(x, 0, 1)$
- Распределение Стьюдента $t(x, 0, 3)$ с тремя степенями свободы
- Распределение Пуассона $P(k, 10)$
- Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

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

1.2 Доверительные интервалы

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

- для нормального распределения [1, с. 457-458],
- для произвольного распределения [1, с. 461-462]

Представить результаты графически.

2 Теоретическое обоснование

2.1 Бокс-плот 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

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1),\ X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \tag{1}$$

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:

$$\left[\begin{array}{l} x < X_1^T \\ x > X_2^T \end{array}\right. \tag{2}$$

2.2 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

$$P\left(\overline{x}-\frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}<m<\overline{x}+\frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right)=1-\alpha, \tag{3}$$

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

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}}<\sigma<\frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right)=1-\alpha \tag{4}$$

3 Описание работы

Лабораторные работы выполнены на языке программирования Python. С использованием сторонних библиотек `numpy`, `matplotlib`, `pandas`, `IPython`, `seaborn` были построены графики box-plot для различных случайных распределений. Ссылка на GitHub репозиторий: https://github.com/feodorrussia/Mathematical-statistics/tree/master/Lab_2

4 Результаты

4.1 Гистограммы и графики плотности распределения

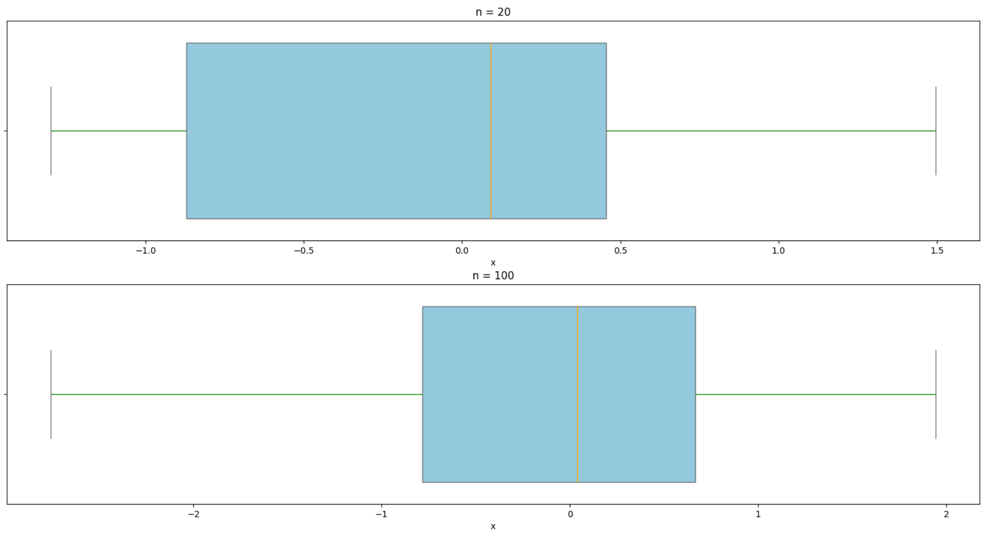

Рис. 1: Box plot для нормального распределения (1)

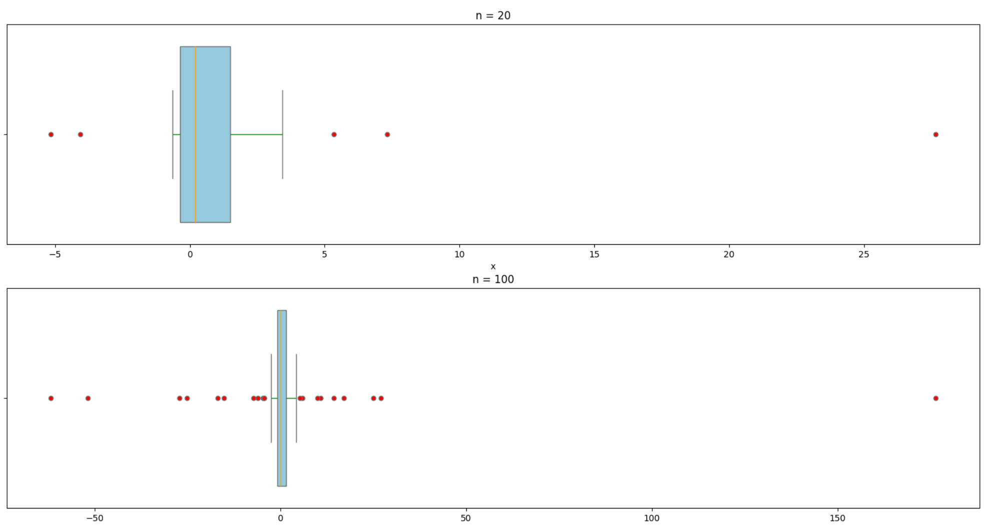

Рис. 2: Box plot для распределения Коши (2)

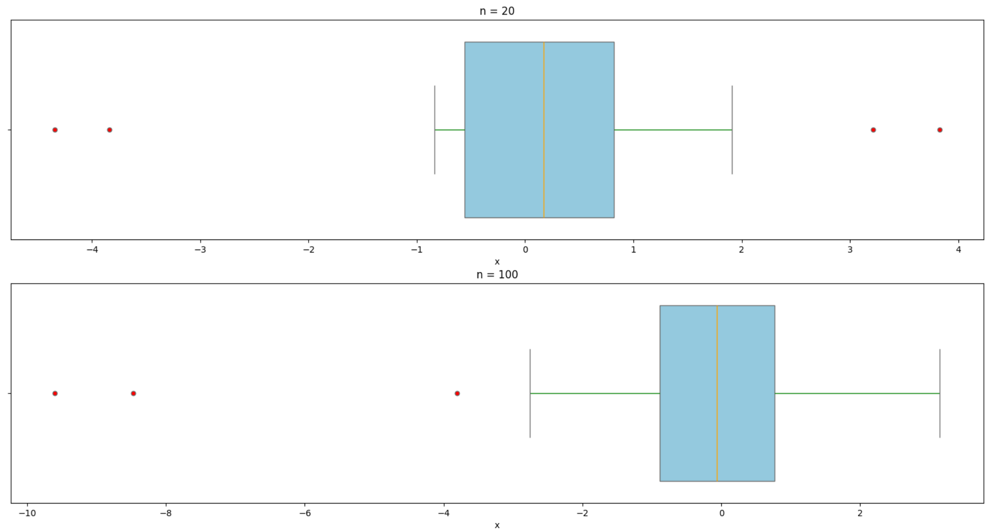

Рис. 3: Box plot для распределения Стьюдента (3)

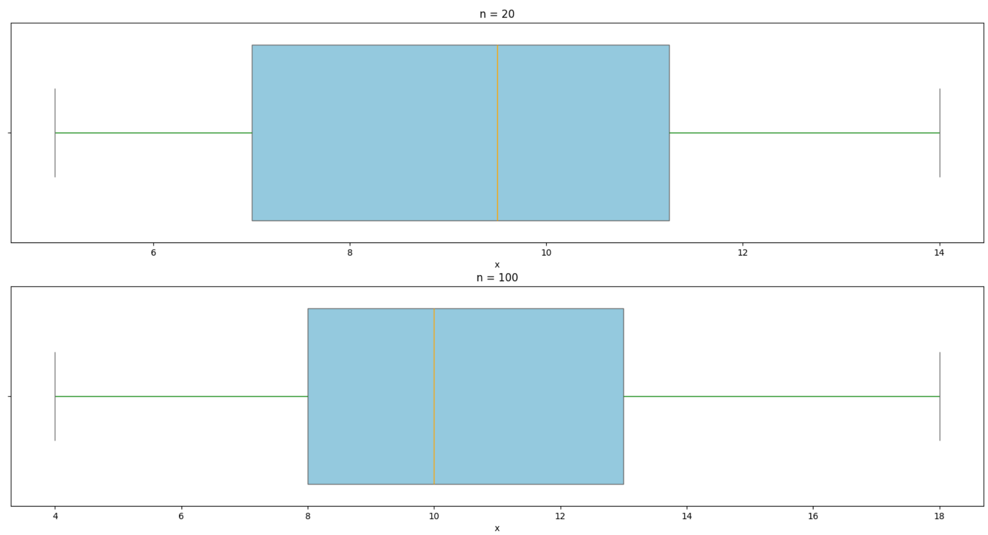

Рис. 4: Box plot для распределения Пуассона (4)

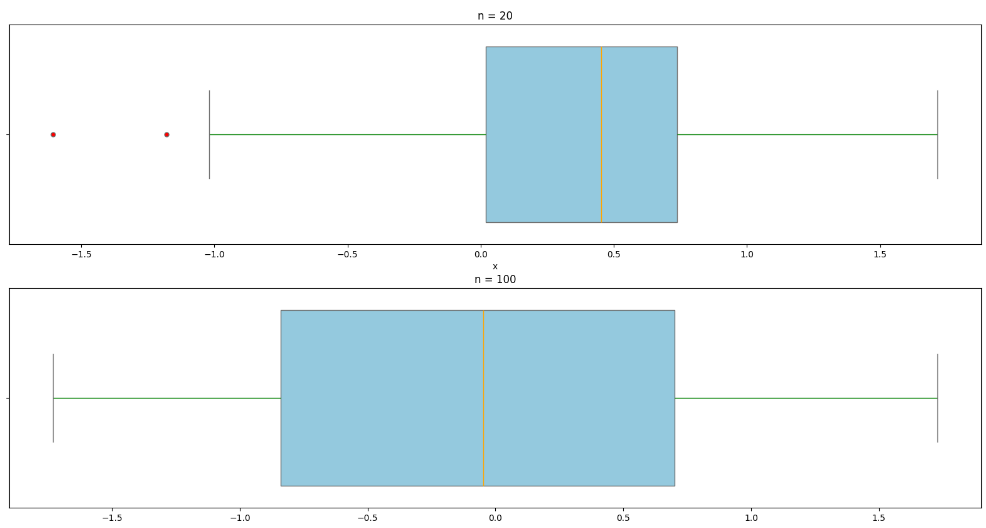

Рис. 5: Box plot для равномерного распределения (5)

4.2 Доверительные интервалы для параметров распределений

n = 20	m (15)	σ (16)
	-0.61 < m < 0.39	0.82 < σ < 1.55
n = 100	m	σ
	-0.28 < m < 0.10	0.86 < σ < 1.13

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

n = 20	m (15)	σ (16)
	$0.16 < m < 0.95$	$0.29 < \sigma < 0.34$
n = 100	m	σ
	$0.32 < m < 0.69$	$0.26 < \sigma < 0.30$

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

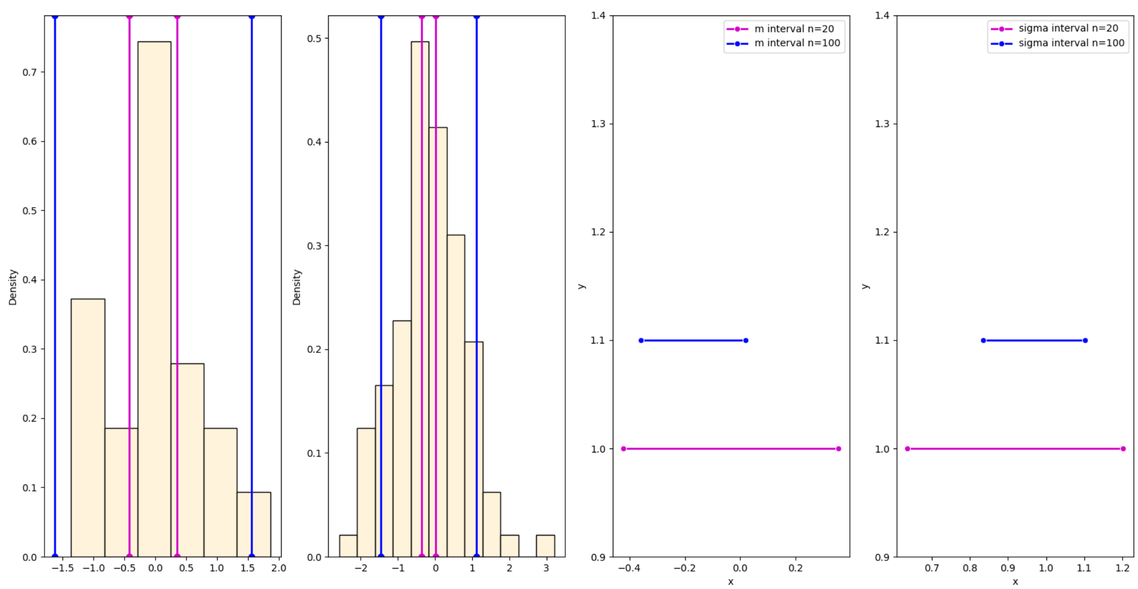

Рис. 6: Гистограммы и оценки для параметров нормального распределения (1)

5 Выводы

По результатам выполнения лабораторной работы были сгенерированы выборки размером 20 и 100 элементов и построены для них боксплоты Тьюки.

Боксплот предоставляет наглядное представление основных статистических характеристик выборки, включая медиану, квартили, межквартальный размах и выбросы.Такие графические представления позволяют исследователям легче выявлять различия в форме и разбросе данных, а также обнаруживать наличие выбросов, что важно при интерпретации статистических данных. Результаты анализа графиков позволяют визуально сопоставить распределение данных для двух различных выборок. Значительное улучшение в проработке метрик наблюдается при увеличении размера выборки до 100 элементов, так как это повышает точность оценок параметров распределения.
Нормальное распределение:

Боксплот для нормального распределения показывает, что медиана совпадает с центром ящика, что характерно для симметричного распределения. Усы боксплота примерно одинаковой длины, что говорит о равномерном разбросе значений.
Распределение Коши:

Боксплот для распределения Коши характеризуется длинными усами и наличием выбросов, что указывает на тяжелые хвосты распределения. Медиана смещена к нижнему квартилю из-за экстремальных значений.
Распределение Пуассона:

Боксплот для распределения Пуассона показывает, что медиана и центр ящика смещены в сторону более низких значений, что характерно для распределения с дискретными значениями. Нет длинных усов, так как значения дискретны и не имеют хвостов.
Равномерное распределение:

Боксплот для равномерного распределения характеризуется усами одинаковой длины, что указывает на равномерный разброс данных. Отсутствие выбросов свидетельствует о том, что все значения находятся в пределах усов.

Также в ходе выполнения лабораторной работы были сгенерированы две выборки размерами 20 и 100 элементов для нормального и произвольного распределения. Затем для каждой из них были вычислены параметры распределения: среднее значение и дисперсия.

Результаты, представленные графически, демонстрируют, что количество элементов в выборке влияет на точность оценок параметров. Более большое количество наблюдений (т.е. 100 элементов) приводит к более точным и стабильным оценкам среднего и дисперсии, как для нормального, так и для произвольного распределения. Для выборки с меньшим количеством элементов (20 элементов) оценки могут сильно варьироваться в зависимости от конкретной выборки, что также наглядно отображено на графиках.

Таким образом, лабораторная работа позволила изучить различные статистические методы анализа данных, а также научиться использовать программные средства для визуализации и обработки статистических данных.