

Лабораторная работа №2 по математической статистике

Тема: Расчет несмещенной дисперсии, математического ожидания и проверка соответствия закона распределения эмпирическим данным

Цель работы

- Научиться рассчитывать основные статистические параметры (математическое ожидание и несмещенную дисперсию) на основе данных.
- Освоить построение доверительных интервалов для математического ожидания и дисперсии.
- Изучить методы выравнивания статистических рядов для подбора теоретического распределения.
- Проверить гипотезы о соответствии эмпирического распределения данным теоретической модели на основе критериев согласия.
- Развить навыки программирования на Python с использованием Jupyter Notebook.
- Реализовать методы анализа данных и статистической проверки гипотез без использования специализированных библиотек для обработки данных (Scikit-learn, SciPy и т.п.).

Задачи

1. Анализ данных:

- Рассчитать математическое ожидание и несмещенную дисперсию для предоставленных наборов данных.
- Реализовать функции расчета этих параметров самостоятельно, без использования готовых библиотек.

2. Построение доверительных интервалов:

- На основе рассчитанных параметров построить доверительные интервалы для:
 - Математического ожидания.
 - Дисперсии.
- Указать методы расчета интервалов и аргументировать выбор.

3. Подбор теоретического распределения:

- Провести анализ статистических рядов для каждого набора данных и подобрать теоретическое распределение.
- Рассмотреть стандартные распределения, такие как:
 - Геометрическое
 - Биномиальное
 - Распределение Пуассона

- Равномерное распределение
- Нормальное распределение
- Экспоненциальное распределение
- Использовать собственные реализации функций для генерации данных этих распределений (с использованием библиотеки NumPy).

4. Проверка гипотез:

- Проверить гипотезы о соответствии эмпирических данных теоретическому распределению.
- Использовать критерий согласия Колмогорова-Смирнова.
- Рассчитать и интерпретировать p-value.
- (Дополнительные баллы) Проверить другие критерии согласия, такие как критерий χ^2 (хи-квадрат).

5. Документация и выводы:

- Подробно описать каждый этап работы, включая расчет математического ожидания, дисперсии, построение доверительных интервалов и проверку гипотез.
- Сделать вывод о том, какое теоретическое распределение наиболее точно описывает каждый набор данных.

Описание данных

В предоставленных наборах данных содержатся числовые значения, сгенерированные из различных распределений. Каждое наблюдение представляет собой реализацию случайной величины. Наборы данных неизвестного происхождения (генератор указанного распределения или эмпирические данные).

Основная цель работы заключается в том, чтобы:

- Проанализировать данные,
- Подобрать подходящее теоретическое распределение,
- Проверить гипотезы о принадлежности данных к выбранному распределению.

Требования к выполнению работы

• Оборудование и ПО:

- Python (рекомендуется последняя версия).
- Jupyter Notebook.

• Язык программирования:

- Python.

• Инструменты:

- Графики строить с использованием matplotlib.
- Генерацию данных для теоретических распределений осуществлять с использованием numpy.
- Методы обработки и анализа данных реализовывать вручную:

- не использовать Scikit-learn, SciPy, statsmodels или аналоги.

- **Обязательные элементы:**

- Подсчет математического ожидания и дисперсии.
- Вычисление доверительных интервалов.
- Построение графического сравнения эмпирических и теоретических распределений.
- Проведение проверки гипотез с расчетом p-value.
- Описание каждого этапа выполнения и выводы.

Пошаговая инструкция

1. Подготовка окружения:

- Убедитесь, что Python и Jupyter Notebook корректно установлены.

2. Импорт данных:

- Загрузите предоставленные файлы данных в Jupyter Notebook.
- Проверьте корректность загрузки и структуру данных.
- Реализуйте обработку ошибок, таких как отсутствие файла или некорректный формат данных.

3. Анализ данных:

- Напишите функции для расчета следующих статистических параметров:
 - Математического ожидания (среднего).
 - Несмещенной дисперсии.
- Проверьте правильность расчетов на небольших тестовых наборах (например, для образцов из 5–10 наблюдений).

4. Построение доверительных интервалов:

- Определите формулы для построения доверительных интервалов для:
 - Математического ожидания (на основе центральной предельной теоремы или распределения Стюдента).
 - Дисперсии (при нормальном распределении данных).
- Реализуйте функции для их построения.
- Графически отобразите результаты.

5. Генерация теоретических распределений:

- Определите подходящее теоретическое распределение для каждого набора данных.
- Реализуйте функции генерации данных для стандартных распределений (геометрическое, биномиальное, Пуассона, равномерное, нормальное, экспоненциальное).
- Постройте графики теоретических распределений в сравнении с гистограммами эмпирических данных.

6. Проверка гипотез:

- Для каждого набора данных проверьте гипотезу о соответствии количественных данных выбранному закону распределения.
- Реализуйте основные процедуры:
 - Критерий Колмогорова-Смирнова для проверки гипотезы.
 - Расчет p-value.
 - Дополнительно (опционально) – реализуйте метод χ^2 (хи-квадрат) для проверки.

7. Сопоставление и анализ:

- Графически сопоставьте эмпирические данные с теоретическими распределениями.
- Сделайте выводы о соответствии данных теоретическим моделям.

8. Документация и выводы:

- Оформите все шаги работы в виде читаемого отчета внутри Jupyter Notebook.
- Подпишите графики (легенды, оси, заголовки).
- Итоговый отчет должен содержать четкие выводы и пояснения.

Рекомендации

• Чистота кода:

- Используйте функции для повторяющихся операций.
- Оформляйте код в соответствии с PEP8.
- Добавляйте комментарии к коду и текстовые описания шагов.

• Визуализация:

- Используйте цветовые схемы и легенды для повышения информативности графиков.
- Добавляйте основные статистические параметры на график (например, среднее или доверительные интервалы).

• Прогресс выполнения:

- Часто сохраняйте промежуточные результаты и сравнивайте их с ожидаемыми.

Критерии оценки лабораторной работы

1. Корректность выполнения задания (40%):

- Точные расчеты математического ожидания, дисперсии и доверительных интервалов.
- Корректная проверка гипотез на основе критериями согласия.
- Реализация методов без готовых функций из библиотек.

2. Качество визуализации (20%):

- Четкость, читаемость и информативность графиков.
- Отображение сравнения эмпирических и теоретических распределений.
- Текстовые описания функций и выводов.

3. Структура и чистота кода (20%):

- Использование функций для повторяющихся задач.
- Соблюдение PEP8.
- Понятность, структурированность и минимизация избыточных частей кода.

4. Документация и выводы (10%):

- Подробные описания шагов выполнения работы.
- Логичные и четкие выводы в конце.

5. Творческий подход и дополнительная проверка (10%):

- Использование нескольких методов проверки гипотез.
- Инициативное добавление улучшений (например, выбор методов оптимизации).

Удачи в выполнении лабораторной работы!