

Описательная статистика

1.1	Выполнение лабораторной работы	1
1.2	Описание лабораторной работы	1
1.3	Рекомендации	2

2.1 Выполнение лабораторной работы

В результаты выполнения работы участником (командой) должны быть предоставлены: код, используемый при выполнении заданий лабораторной работы, а также отчет о ходе выполнения работы. Отчет должен содержать:

1. Титульную информацию: для каждого участника должны быть указаны ФИО, ИСУ, практический поток.
2. Ход выполнения работы: *краткая* информация о выполненных шагах.
3. Основную часть: описание каждого шага, промежуточные результаты и их анализ.
4. Заключение: приоритетные выводы по результатам выполнения всей лабораторной работы.

2.2 Описание лабораторной работы

1. Выберите два распределения (одно дискретное, другое непрерывное). Сгенерируйте, используя библиотеки, выборку из $N = 1000$ значений для каждого из распределений.
2. Для каждой выборки:
 - (a) Рассчитайте основные описательные статистики:
 - Квартили: Q_1 , Q_2 , Q_3 .
 - Меры центральной тенденции: выборочное среднее, медиана, мода.
 - Меры вариабельности: размах выборки, интерквартильный размах, дисперсия, стандартное отклонение, коэффициент вариации, среднее абсолютное отклонение.
 - Меры формы распределения: коэффициенты асимметрии, эксцесса.
 - Первые 5 начальных и центральных моментов (допускается использование смещенных характеристик).
 - (b) Постройте графики:
 - Эмпирическую функцию распределения eCDF с наложением ее на теоретическую CDF;
 - Эмпирическую плотность распределения (гистограмму) ePDF для непрерывного распределения с наложением ее на теоретическую PDF;Постройте несколько вариантов гистограммы: с автоматическим выбором количества бинов, с выбором по одной из известных формул и подберите вручную какое-то количество, которое покажется визуально оптимальным.

- Эмпирическую функцию вероятности (многоугольник вероятности) для дискретного распределения с наложением ее на теоретическую функцию вероятности;
 - Boxplot с выделением выбросов (по правилу $1.5 \cdot \text{IQR}$).
- (с) Исследуйте устойчивость характеристик. Рассмотрите выборку из непрерывного распределения и
- добавьте к выборке 5% выбросов (значения, которые лежат в области низкой вероятности появления),
 - пересчитайте статистики из пункта (а) и сравните с исходными
 - постройте графики изменения мер вариабельности при постепенном добавлении выбросов (0%, 1%, ..., 5%). Возьмите не менее 10 процентных пунктов.
 - Сделайте выводы об устойчивости этих характеристик.

2.3 Рекомендации

- Во время подготовки всех таблиц и графиков, предполагайте, что это ваш опорный описательный конспект по статистическому исследованию данных, к которому вы будете регулярно возвращаться при построении моделей машинного обучения в дальнейшем.
- Для генерации данных и подсчета статистических характеристик, пользуйтесь средствами любой из библиотек — NumPy, SciPy.stats, pandas и т.д.
- По желанию потренируйтесь в создании визуальных материалов, используя не только библиотеку matplotlib, но и plotly, seaborn и другие. Плюс в карму обеспечен.
- Для генерации данных используйте следующие распределения:
 - **Дискретные** (через `numpy.random` или `scipy.stats`):
 - * Бернулли: p (вероятность успеха), `scipy.stats.bernoulli`
 - * Биномиальное: n (число испытаний), p (вероятность успеха), `numpy.random.binomial`
 - * Пуассона: λ (среднее), `numpy.random.poisson`
 - * Геометрическое: p (вероятность успеха), `scipy.stats.geom`
 - * Гипергеометрическое: M (общее объектов), n (успешных), N (выборка), `scipy.stats.hypergeom`
 - * Отрицательное биномиальное: r (число успехов), p (вероятность), `scipy.stats.nbinom`
 - **Непрерывные** (через `numpy.random` или `scipy.stats`):
 - * Нормальное: μ (loc), σ (scale), `numpy.random.normal`
 - * Экспоненциальное: λ^{-1} (scale), `numpy.random.exponential`
 - * Хи-квадрат: k (степени свободы), `scipy.stats.chi2`
 - * Стьюдента (t): k (степени свободы), `scipy.stats.t`
 - * Фишера (F): d_1, d_2 (степени свободы), `scipy.stats.f`
 - * Логнормальное: μ (в логарифме), σ (в логарифме), `scipy.stats.lognorm`