**Лабораторная забота №1.**

**Статистическое моделирование случайных величин. Точечное оценивание параметров распределения и функций распределения случайных величин.**

**Часть I.**

* Смоделировать выборку из ***п*** независимых наблюдений над случайной величиной ***X***, имеющей нормальный закон распределения с параметрами (***a***,**2**).
* Сгруппировать данные в интервалы и найти интервальный ряд абсолютных частот. Число интервалов группировки выбрать в соответствии с правилом, указанном в Вашем варианте. Вычислить средствами Python сумму абсолютных частот. Построить диаграмму абсолютных частот.
* Сгруппировать данные в интервалы и найти интервальный ряд относительных частот. Вычислить средствами Python сумму относительных частот. Построить диаграмму относительных частот.
* Для визуализации данных:
* Построить гистограмму относительных частот и теоретическую кривую распределения случайной величины ***X*** (причем гистограмму и кривую распределения разместить на одном рисунке).
* Построить гистограмму абсолютных и график теоретической частоты распределения случайной величины ***X*** (причем и гистограмму, и график теоретической частоты разместить на одном рисунке).
* Построить по выборке график эмпирической функции распределения случайной величины ***X*** (кумуляту относительных частот) и график теоретической функции распределения случайной величины ***X***, разместив их на одном рисунке.
* Построить по выборке бокс-плот распределения (“ящик с усами”) и дать ему статистическую интерпретацию. В частности, ответить на вопрос: Каково теоретически ожидаемое число выбросов для выборки данного объема?

Проанализировав результаты п.1 и 2., ответить на следующие **вопросы:**

* Какова Ваша оценка вероятности попадания значений случайной величины ***X*** в k-ый интервал полученной группировки?
* Какой из полученных в п.2 графиков содержит геометрическую интерпретацию этой оценки, покажите ее на графике.
* Какой из полученных в п.2 графиков содержит геометрическую интерпретацию вероятности попадания значений случайной величины ***X*** в k-ый интервал, покажите ее на графике.
* Какова Ваша оценка значения интегральной функции распределения в правой границе k-ого интервала полученной группировки? А чему равно точное значение интегральной функции распределения в правой границе k-ого интервала?
* Используя таблицу значений функции Лапласа, найти вероятность

(вычисления вероятности сделать “вручную” )

P(|X-MX|<q) (\*)

3.1. Найти вероятность (\*), используя встроенные функции Python.

3.2. Найдите по выборке оценку вероятности (\*), т.е. относительную частоту события

{|*X*-M*X*|<*q*}.

3.3. Увеличьте объем выборки ***п*** в 50 раз и повторите п. 2.2. Объясните, как увеличение объема выборки повлияло на точность оценки вероятности (\*).

4.Вычислить по выборке точечные оценки параметров распределения случайной величины ***X***: математического ожидания, медианы, дисперсии, стандартного отклонения, коэффициента ассиметрии, эксцесса двумя способами:

4.1. первый способ –непосредственное применение формул для вычисления точечных оценок;

4.2. второй способ – применение встроенных функций Python, например, describe, std, mean и т.д.

4.3. Увеличьте объем выборки ***п*** в 50 раз и повторите п. 4. 2.

Проанализировав результаты п.4, ответить на следующие **вопросы:**

* Чему равны коэффициенты асимметрии и эксцесса распределения случайной величины ***X***?
* Объясните, как увеличение объема выборки повлияло на точность оценок параметров распределения.
* Какое свойство точечных оценок параметров распределения случайной величины ***X*** объясняет такое поведение оценок параметров распределения в ответ на рост объема выборки?

**Часть II.**

1.Смоделировать выборку из ***п*** независимых наблюдений над случайной величиной ***Y***, имеющей закон распределения, указаный в Вашем варианте.

2. Для визуализации данных построить гистограмму относительных частот и теоретическую кривую распределения случайной величины ***Y*** (причем гистограмму и кривую распределения разместить на одном рисунке).

3.Построить по выборке бокс-плот распределения (“ящик с усами”) случайной величины ***Y*** и дать ему статистическую интерпретацию. В частности, ответить на вопрос:

Каково теоретически ожидаемое число выбросов для выборки данного объема?

Проанализировав результаты п.1 и 2, ответить на следующие вопросы:

* Какова Ваша оценка вероятности попадания значений случайной величины ***Y*** в k-ый интервал полученной группировки?
* Какова вероятность попадания значений случайной величины ***Y*** в k-ый интервал полученной группировки?

2. Используя теоретические знания о функции распределения случайной величины, найти вероятность (вычисления вероятности сделать “вручную” )

P(|***Y*** -M ***Y*** |< *q*) (\*)

2.1. Найти вероятность (\*), используя встроенные функции Python.

2.2. Найдите по выборке оценку вероятности (\*), т.е. относительную частоту события

{ ***Y*** -M ***Y*** |<q}.

2.3. Увеличьте объем выборки ***п*** в 50 раз и повторите п. 2.2. Объясните, как увеличение объема выборки повлияло на точность оценки вероятности (\*).

3. Вычислить по выборке точечные оценки параметров распределения случайной величины ***X***: математического ожидания, медианы, дисперсии, стандартного отклонения, коэффициента ассиметрии, эксцесса, применяя встроенных функций Python, например, describe, std, mean и т.д.

Чему равны коэффициенты ассиметрии и эксцесса распределения случайной величины ***Y***? Объясните, как увеличение объема выборки повлияло на точность оценок параметров распределения.

**ВАРИАНТЫ ЗАДАНИЙ**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Вариант** | **параметры**  (***a***;**2**) | ***n*** | ***q*** | **k** | **Правило для расчета числа интервалов группировки** | **Закон распределения случайной величины *Y*** |
| 14 | (4;9) | 90 | 1,35 | 5 | Скотта | Треугольное на отрезке (0, 5) и модой, равной 3 |
| 15 | (2;16) | 140 | 1,2 | 2 | Фридмана Диакониса | Равномерное распределение  на отрезке (-1,3) |
| 16 | (-5;9) | 150 | 1,6 | 5 | Стерджеса | t -распределение Стьюдента с числом степеней свободы k=6 |
| 17 | (10;1) | 100 | 1,65 | 2 | Скотта | Распределение Парето на промежутке (1, ) с параметром формы, равным 2 |
| 18 | (-2;1) | 110 | 0,8 | 3 | Фридмана Диакониса | Хи-квадрат с числом степеней свободы, равным 5 |
| 19 | (-3;9) | 125 | 0,85 | 4 | Стерджеса | Релея с модой, равной 1. |
| 20 | (-5;4) | 145 | 1,95 | 5 | Скотта | Экспоненциальное с математическим ожиданием, равным 5 |
| 21 | (7;25) | 160 | 1,1 | 6 | Фридмана Диакониса | Распределение Парето на промежутке (4, ) с параметром формы, равным 5 |
| 22 | (-1;16) | 150 | 1,2 | 4 | Стерджеса | Треугольное на отрезке (2, 5) и модой, равной 3 |
| 23 | (2;25) | 190 | 1,9 | 6 | Скотта | Логистическое с параметрами масштаба и сдвига 1 3и 3 соответственно |
| 24 | (0;16) | 120 | 2,3 | 5 | Фридмана Диакониса | F-распределение с числом степеней свободы k1=6 и k2=5 |
| 25 | (-3;9) | 95 | 2,5 | 4 | Скотта | Логистическое с параметрами масштаба и сдвига 0 и 2 соответственно |

***Указание.***

***Использовать библиотеки***

math, numpy, pandas, statistics, scipy.stats , seaborn, matplotlib.pyplot и др.

В частности, изучить функции из категории summary statistics из библиотеки scipy.stats и

функции для разыгрывания распределений с заданным законом из библиотеки np.random.