

Лабораторная работа №1.

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

Часть I.

1. Смоделировать выборку из n независимых наблюдений над случайной величиной X , имеющей нормальный закон распределения с параметрами (a, σ^2) .
 - 1.1. Сгруппировать данные в интервалы и найти интервальный ряд абсолютных частот. Число интервалов группировки выбрать в соответствии с правилом, указанным в Вашем варианте. Вычислить средствами Python сумму абсолютных частот. Построить диаграмму абсолютных частот.
 - 1.2. Сгруппировать данные в интервалы и найти интервальный ряд относительных частот. Вычислить средствами Python сумму относительных частот. Построить диаграмму относительных частот.
2. Для визуализации данных:
 - 2.1. Построить гистограмму относительных частот и теоретическую кривую распределения случайной величины X (причем гистограмму и кривую распределения разместить на одном рисунке).
 - 2.2. Построить гистограмму абсолютных и график теоретической частоты распределения случайной величины X (причем и гистограмму, и график теоретической частоты разместить на одном рисунке).
 - 2.3. 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.
 - 2.4. 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**:

- A) Какова Ваша оценка вероятности попадания значений случайной величины X в k -ый интервал полученной группировки?
 - B) Какой из полученных в п.2 графиков содержит геометрическую интерпретацию этой оценки, покажите ее на графике.
 - C) Какой из полученных в п.2 графиков содержит геометрическую интерпретацию вероятности попадания значений случайной величины X в k -ый интервал, покажите ее на графике.
 - D) 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?
3. Используя таблицу значений функции Лапласа, найти вероятность (вычисления вероятности сделать “вручную”)

$$P(|X-MX|<q\sigma) \quad (*)$$

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

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

$$\{|X-MX| < q\sigma\}.$$

3.3. Увеличьте объем выборки n 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 асимметрии, 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 n в 50 раз и повторите п. 4. 2.

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

- А) Чему равны коэффициенты асимметрии и эксцесса распределения случайной величины X ?
- В) Объясните, как увеличение объема выборки повлияло на точность оценок параметров распределения.
- С) 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 n независимых наблюдений над случайной величиной Y , имеющей закон распределения, указанный 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:

- А) Какова Ваша оценка вероятности попадания значений случайной величины Y в k -ый интервал полученной группировки?
- В) 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”)

$$P(|Y - M Y| < q\sigma) \quad (*)$$

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

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

$$\{ |Y - M Y| < q\sigma \}.$$

2.3. Увеличьте объем выборки n 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 асимметрии, эксцесса, применяя встроенных функций Python, например, describe, std, mean и т.д.

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

ВАРИАНТЫ ЗАДАНИЙ

<u>Вариант</u>	параметры ($a; \sigma^2$)	n	q	k	Правило для расчета числа интервалов группировки	Закон распределения случайной величины Y
<u>1</u>	(1;4)	80	2	3	Стерджеса	Логнормальное со математическим ожиданием, равным 1 и стандартным отклонением, равным 2
<u>2</u>	(-1;1)	100	1	4	Скотта	Хи-квадрат с числом степеней свободы, равным 7
<u>3</u>	(-3;9)	70	1,5	5	Фридмана - Диакониса	Экспоненциальное с математическим ожиданием, равным 3
<u>4</u>	(3;16)	90	1,3	3	Стерджеса	F-распределение с числом степеней свободы $k_1=5$ и $k_2=10$
<u>5</u>	(-2;9)	110	1,25	4	Скотта	t-распределение Стьюдента с числом степеней свободы $k=7$
<u>6</u>	(-1;4)	120	1,4	5	Фридмана Диакониса	Равномерное распределение на отрезке (1,2)
<u>7</u>	(0;9)	60	1,75	3	Стерджеса	Треугольное на отрезке (-2, 4) и модой, равной 1
<u>8</u>	(0;4)	70	0,5	4	Скотта	Логистическое с параметрами

						масштаба и сдвига 0 и 1 соответственно
<u>9</u>	(-4;16)	80	0,75	3	Фридмана Диакониса	Релея с модой, равной 2.
<u>10</u>	(-2;9)	90	1,6	3	Стерджеса	Распределение Парето на промежутке $(3, \infty)$ с параметром формы, равным 4
<u>11</u>	(-1;1)	100	2,1	4	Скотта	Логнормальное со средним, равным 5 и стандартным отклонением, равным 1
<u>12</u>	(3;4)	70	1,45	5	Фридмана Диакониса	t-распределение Стьюдента с числом степеней свободы $k=4$
<u>13</u>	(0;25)	80	1,55	3	Стерджеса	Экспоненциальное с математическим ожиданием, равным 7
<u>14</u>	(4;9)	90	1,35	5	Скотта	Треугольное на отрезке $(0, 5)$ и модой, равной 3
<u>15</u>	(2;16)	140	1,2	2	Фридмана Диакониса	Равномерное распределение на отрезке $(-1,3)$
<u>16</u>	(-5;9)	150	1,6	5	Стерджеса	t -распределение Стьюдента с числом степеней свободы $k=6$
<u>17</u>	(10;1)	100	1,65	2	Скотта	Распределение Парето на промежутке $(1, \infty)$ с параметром формы, равным 2
<u>18</u>	(-2;1)	110	0,8	3	Фридмана Диакониса	Хи-квадрат с числом степеней свободы, равным 5
<u>19</u>	(-3;9)	125	0,85	4	Стерджеса	Релея с модой, равной 1.
<u>20</u>	(-5;4)	145	1,95	5	Скотта	Экспоненциальное с математическим ожиданием, равным 5
<u>21</u>	(7;25)	160	1,1	6	Фридмана Диакониса	Распределение Парето на промежутке $(4, \infty)$ с параметром формы, равным 5

<u>22</u>	$(-1;16)$	150	1,2	4	Стерджеса	Треугольное на отрезке $(2, 5)$ и модой, равной 3
<u>23</u>	$(2;25)$	190	1,9	6	Скотта	Логистическое с параметрами масштаба и сдвига 1 и 3 соответственно
<u>24</u>	$(0;16)$	120	2,3	5	Фридмана-Диакониса	F-распределение с числом степеней свободы $k_1=6$ и $k_2=5$
<u>25</u>	$(-3;9)$	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.