Лабораторная работа по Математической статистике № 1
«Первичная обработка выборки»

Задание 1. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных по биномиальному закону с параметрами ***n*** и ***p***:

$p_k = C_n^k \cdot p^k \cdot q^{n-k}$, $k = 0,1,...,n$. $n = 5 + V \bmod 19$, $p = 0,2 + 0,005V$.

Задание 2. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных по геометрическому закону с параметром ***p***:

$p_k = q^k \cdot p$, $k = 0,1,...$ $p = 0,2 + 0,005V$.

Задание 3. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных по закону Пуассона с параметром $\boldsymbol{\lambda}$:

$p_k = \dfrac{\lambda^k}{k!} e^{-\lambda}$, $k = 0,1,...$ $\lambda = 0,5 + 0,01 \cdot V$.

Задание 4. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных равномерно на множестве $\{0,...,n-1\}$:

$p_k = \dfrac{1}{n}$, $k = 0,...,(n-1)$. $n = 5 + V \bmod 28$.

Задание 5. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных по гипергеометрическому закону с параметрами ***M***, ***K***, ***m***:

$p_k = \dfrac{C_K^k \cdot C_{M-K}^{m-k}}{C_M^m}$, $k = 0,1,...,m$. $m = 5 + V \bmod 7$, $M = 2m + 3$, $K = m + 1 + V \bmod 2$.

Следуя **Указаниям** для всех выборок в Заданиях 1-5

построить:

1) статистический ряд;

2) полигон относительных частот;

3) график эмпирической функции распределения;

найти:

1) выборочное среднее;

2) выборочную дисперсию;

3) выборочное среднее квадратическое отклонение;

4) выборочную моду;

5) выборочную медиану;

6) выборочный коэффициент асимметрии;

7) выборочный коэффициент эксцесса;

составить таблицы:

1) сравнения относительных частот и теоретических вероятностей;

2) сравнения рассчитанных характеристик с теоретическими значениями.

Задание 6. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных по нормальному закону с параметрами $a=(-1)^V \cdot 0{,}05 \cdot V$ и σ^2, где $\sigma = 0{,}005 \cdot V + 1$.

Задание 7. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных по показательному закону с параметром $\lambda = 2 + (-1)^V \cdot 0{,}01 \cdot V$.

Задание 8. Получить выборку, сгенерировав 200 псевдослучайных чисел, распределенных равномерно на отрезке $[a,b]$, где $a=(-1)^V \cdot 0{,}02 \cdot V$, $b = a + 6$.

Следуя **Указаниям** для всех выборок в Заданиях 6-8

построить:

1) интервальный ряд и ассоциированный статистический ряд;

2) гистограмму относительных частот;

3) график эмпирической функции распределения;

найти:

1) выборочное среднее;

2) выборочную дисперсию с поправкой Шеппарда;

3) выборочное среднее квадратическое отклонение;

4) выборочную моду;

5) выборочную медиану;

6) выборочный коэффициент асимметрии;

7) выборочный коэффициент эксцесса;

составить таблицы:

1) сравнения относительных частот и теоретических вероятностей попадания в интервалы;

2) сравнения рассчитанных характеристик с теоретическими значениями.

V – номер варианта.

Результаты вычислений приводить в отчете с точностью до ***0,00001.***

Указания

Для получения выборок нужно использовать соответствующие функции применяемого языка программирования.

В Заданиях 1-5 полученную выборку $\{x_1, x_2, x_3, ..., x_N\}$ упорядочить по возрастанию, определить частоты n_i и относительные частоты (частости) w_i, построить статистический ряд вида:

x_i	n_i	w_i	s_i
x_1^*	n_1	w_1	s_1
x_2^*	n_2	w_2	s_2
…	…	…	…
x_m^*	n_m	w_m	s_m
	$\sum_{i=1}^{m} n_i$	$\sum_{i=1}^{m} w_i$	-

$x_i^* < x_j^*$ при $i < j$, n_i – частота x_i^* (число значений x_i^*, встречающихся в выборке), $\sum_{i=1}^{m} n_i = N$;

w_i – относительная частота (частость) значения x_i^*, $w_i = \dfrac{n_i}{N}$, $\sum_{i=1}^{m} w_i = 1$; $s_k = \sum_{j=1}^{k} w_j$,

$s_1 = w_1, s_m = 1$.

Полигон относительных частот – ломаная линия, соединяющая последовательно точки с координатами $(0, \tilde{w}_0), (1, \tilde{w}_1), ..., (M, \tilde{w}_M)$, где $M = x_m^* = \max\{x_i^* : 1 \le i \le m\}$; $\tilde{w}_j = w_i$, если существует такое x_i^*, что $j = x_i^*$, и $\tilde{w}_j = 0$ в противном случае.

Эмпирическая функция распределения

$$F_N^Э(x) = \sum_{x_i^* \le x} w_i = \begin{cases} 0, & x < x_1^*, \\ w_1, & x_1^* \le x < x_2^*, \\ w_1 + w_2, & x_2^* \le x < x_3^*, \\ w_1 + w_2 + w_3, & x_3^* \le x < x_4^*, \\ \dots\dots\dots\dots\dots\dots, & \\ 1, & x \ge x_m^*. \end{cases}$$

Выборочное среднее

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{m} x_i^* \cdot n_i = \sum_{i=1}^{m} x_i^* \cdot w_i .$$

Выборочный момент k-ого порядка (выборочный k-ый момент)

$$\overline{\mu}_k = \sum_{i=1}^{m} (x_i^*)^k \cdot w_i , \ \overline{\mu}_1 = \overline{x} .$$

Выборочная дисперсия

$$D_B = \sum_{i=1}^{m} (x_i^* - \overline{x})^2 \cdot w_i = \overline{\mu}_2 - (\overline{\mu}_1)^2 .$$

Выборочный центральный момент k-ого порядка (выборочный центральный k-ый момент)

$\bar{\mu}_k^0=\sum_{i=1}^{m}(x_i^*-\bar{x})^k\cdot w_i$, $\bar{\mu}_1^0=0$, $\bar{\mu}_2^0=D_B$,

$\bar{\mu}_3^0=\bar{\mu}_3-3\bar{\mu}_2\bar{\mu}_1+2(\bar{\mu}_1)^3$,

$\bar{\mu}_4^0=\bar{\mu}_4-4\bar{\mu}_3\bar{\mu}_1+6\bar{\mu}_2(\bar{\mu}_1)^2-3(\bar{\mu}_1)^4$.

Выборочное среднее квадратическое отклонение

$\bar{\sigma}=\sqrt{D_B}$

Выборочная мода $\bar{M}_0$

$\bar{M}_0=\{x_i^* \mid n_i=\max n_k\}$, если $n_i=\max n_k>n_j,\ i\neq j$;

если $n_i=n_{i+1}=...=n_{i+j}=\max n_k$, то $\bar{M}_0=\frac{1}{2}(x_i^*+x_{i+j}^*)$,

если $n_i=n_j=\max n_k>n_l$, $i<l<j$, то $\bar{M}_0$ – не существует.

Выборочная медиана

$$\bar{M}_e=\begin{cases} x_i^*, & F_N^Э(x_{i-1}^*)<0{,}5<F_N^Э(x_i^*), \\ \frac{1}{2}(x_i^*+x_{i+1}^*), & F_N^Э(x_i^*)=0{,}5. \end{cases}$$

Выборочный коэффициент асимметрии $\bar{\gamma}_1=\dfrac{\bar{\mu}_3^0}{\bar{\sigma}^3}$.

Выборочный коэффициент эксцесса $\bar{\gamma}_2=\dfrac{\bar{\mu}_4^0}{\bar{\sigma}^4}-3$.

В Заданиях 6-8 полученную выборку $\{x_1, x_2, x_3, ..., x_N\}$ упорядочить по возрастанию, определить интервалы $[a_0,a_1],(a_1,a_2],\dots,(a_{m-1},a_m]$; $a_0=x_{(1)}$, $a_m=x_{(N)}$, $a_k-a_{k-1}=\dfrac{a_m-a_0}{m}$, $k=1,...,m$, где число интервалов находится по формуле Стерджеса $m=1+[\log_2 N]$.

Интервальный ряд оформляется в виде:

Интервалы	n_i	w_i
$[a_0,a_1]$	n_1	w_1
$(a_1,a_2]$	n_2	w_2
…	…	…
$(a_{m-1},a_m]$	n_m	w_m
	$\sum_{i=1}^{m}n_i$	$\sum_{i=1}^{m}w_i$

n_i – число значений, попавших в i-ый интервал; w_i – относительная частота попадания в i-ый интервал, $w_i = \dfrac{n_i}{N}$.

Ассоциированный статистический ряд:

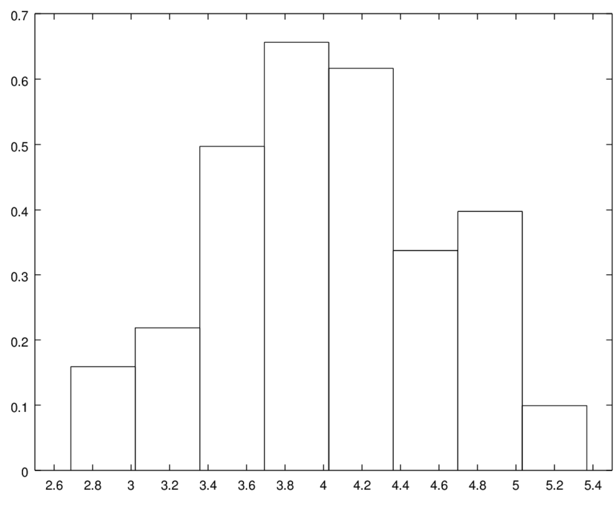
x_i^*	n_i	w_i
x_1^*	n_1	w_1
x_2^*	n_2	w_2
…	…	…
x_m^*	n_m	w_m
	$\sum_{i=1}^{m} n_i$	$\sum_{i=1}^{m} w_i$

где $x_i^* = \dfrac{a_{i-1}+a_i}{2}$ – середина интервала $(a_{i-1}, a_i]$.

Эмпирическая функция распределения:

$$F_N^{Э}(x;x_1,x_2,...,x_N) = \sum_{x_k \le x} \frac{1}{N} = \begin{cases} 0, & x < x_{(1)}, \\ \frac{1}{N}, & x_{(1)} \le x < x_{(2)}, \\ \frac{2}{N}, & x_{(2)} \le x < x_{(3)}, \\ \frac{3}{N}, & x_{(3)} \le x < x_{(4)}, \\ \dots\dots\dots\dots\dots, \\ 1, & x \ge x_{(N)}. \end{cases}$$

Образец гистограммы относительных частот



Площадь i-ого столбца гистограммы равна w_i , а высота $\dfrac{w_i}{h}$.

Выборочное среднее:

$\overline{x} = \frac{1}{N}\sum_{i=1}^{m} x_i^* \cdot n_i = \sum_{i=1}^{m} x_i^* \cdot w_i$.

Выборочная дисперсия с поправкой Шеппарда:

$s_B^2 = \sum_{i=1}^{m} (x_i^* - \overline{x})^2 \cdot w_i - \frac{h^2}{12}$, где $h = (a_m - a_0)/m$.

Выборочное среднее квадратичное отклонение:

$\tilde{\sigma} = \sqrt{s_B^2}$.

Выборочная мода:

Если модальный интервал (на котором высота гистограммы максимально) один, то

$$\bar{M}_0 = a_k + h\frac{w_{k+1} - w_k}{2w_{k+1} - w_k - w_{k+2}}$$

где a_k – левая граница модального интервала (a_k, a_{k+1});

a_{k+1} – правая граница модального интервала (a_k, a_{k+1});

w_{k+1} – относительная частота на модальном интервале;

w_k, w_{k+2} – относительные частоты интервалов слева и справа от модального интервала.

Если модальных интервалов несколько и все они идут подряд

(т.е. интервалы $(a_k, a_{k+1}), \ldots (a_{k+l-1}, a_{k+l})$ – все модальные), то

$$\bar{M}_0 = a_k + l \cdot h \cdot \frac{w_{k+l} - w_k}{2w_{k+l} - w_k - w_{k+l+1}} .$$

Если между модальными интервалами находятся немодальные, то считаем, что выборочной моды не существует.

Выборочная медиана:

$\bar{M}_e = a_{k-1} + \frac{h}{w_k}\left(\frac{1}{2} - \sum_{i=1}^{k-1} w_i\right)$, если $\sum_{i=1}^{k-1} w_i < \frac{1}{2} < \sum_{i=1}^{k} w_i$;

$\bar{M}_e = a_k$, если $\sum_{i=1}^{k} w_i = \frac{1}{2}$.

Выборочный момент k-ого порядка: $\bar{\mu}_k = \overline{x^k} = \sum_{i=1}^{m} (x_i^*)^k \cdot w_i$, $\bar{\mu}_1 = \overline{x}$.

Выборочный центральный момент k-ого порядка:

$\bar{\mu}_k^0 = \sum_{i=1}^{m} (x_i^* - \overline{x})^k \cdot w_i$, $\bar{\mu}_1^0 = 0$, $\bar{\mu}_2^0 = D_B = \bar{\mu}_2 - (\bar{\mu}_1)^2$.

Выборочный коэффициент асимметрии: $\bar{\gamma}_1 = \dfrac{\bar{\mu}_3^0}{\bar{\sigma}^3}$.

Выборочный коэффициент эксцесса: $\bar{\gamma}_2 = \dfrac{\bar{\mu}_4^0}{\bar{\sigma}^4} - 3$.

Графики гистограмм относительных частот и эмпирических функций распределения должны иметь масштаб, обеспечивающий удобный просмотр; точки деления на горизонтальной оси должны быть расположены через 1 или 2, на вертикальной оси – через 0,05.

В разделе отчета **Краткие теоретические сведения** должна быть сначала информация о выборках и случайных величинах из Заданиях 1-5: определения статистического ряда, полигона относительных частот, эмпирической функций распределения, формулы для расчета выборочного среднего, выборочной дисперсии, выборочной моды, выборочной медианы, выборочного коэффициента асимметрии, выборочного коэффициента эксцесса. Для каждого распределения из Заданиях 1-5 нужно привести выражения для вероятностей ряда распределения, а также для математического ожидания (среднего значения), дисперсии, среднего квадратичного отклонения, моды, медианы, коэффициента асимметрии, коэффициента эксцесса.
Для Заданий 6-8 нужно привести: определения интервального ряда, гистограммы относительных частот, формулы для выборочных значений дисперсии с поправкой Шеппарда, моды, медианы, выражения для плотности, функции распределения, математического ожидания, дисперсии, среднего квадратичного отклонения, моды, медианы, коэффициента асимметрии, коэффициента эксцесса соответствующих случайных величин.

В этом разделе должны быть описаны функции языка программирования, которые использованы в программе расчета.

В разделе отчета **Результаты расчетов** для каждого задания должен быть отдельный подраздел, в котором нужно привести номер варианта и значения параметров соответствующего распределения; таблицу 20*10 (20 строк, 10 столбцов) с выборкой псевдослучайных чисел $\{x_1, x_2, x_3, ..., x_N\}$, полученных в соответствии с заданным законом распределения; . таблицу 20*10 (20 строк, 10 столбцов) с полученными псевдослучайными числами, упорядоченными по возрастанию $\{x_{(1)}, x_{(2)}, x_{(3)}, ..., x_{(N)}\}$, $x_{(1)} < x_{(2)} < x_{(3)} < ... < x_{(N)}$.

Для Заданий 1-5 должны быть
– статистический ряд;
– полигон относительных частот;
– график эмпирической функции распределения.

На каждый график полигона относительных частот нужно наложить выделенный красным цветом график полигона соответствующих теоретических вероятностей (ломаной линии, соединяющей последовательно точки с координатами $(0, p_0), (1, p_1), (2, p_2), ..., (k, p_k), ...$).

Графики полигонов относительных частот и эмпирических функций распределения должны иметь масштаб, обеспечивающий удобный просмотр; точки деления на горизонтальной оси должны расположены через 1 или 2, на вертикальной оси – через 0,05.

Для Заданий 6-8 должны быть

– интервальный ряд;

– ассоциированный статистический ряд;

– гистограмма относительных частот;

– график эмпирической функции распределения.

Для Заданий 1-5 таблица сравнения относительных частот и теоретических вероятностей должна иметь вид

Название распределения

j	$\tilde{w}_j$	p_j	$\lvert\tilde{w}_j - p_j\rvert$
0	$\tilde{w}_0$	p_0	$\lvert\tilde{w}_0 - p_0\rvert$
1	$\tilde{w}_1$	p_1	$\lvert\tilde{w}_1 - p_1\rvert$
…	…	…	…
M	$\tilde{w}_M$	p_M	$\lvert\tilde{w}_M - p_M\rvert$
	$\sum_{i=0}^{M} \tilde{w}_j$	$\sum_{i=0}^{M} p_j$	$\Delta_{\max}$

где $M = x_m^* = \max\{x_i^* : 1 \le i \le m\}$; p_j – теоретическая вероятность значения j, рассчитанная по формулам данного распределения; $\tilde{w}_j = w_i$, если существует такое x_i^*, что $j = x_i^*$, и $\tilde{w}_j = 0$ в противном случае; $\Delta_{\max} = \max\{|\tilde{w}_j - p_j| : 0 \le j \le M\}$;

Для Заданий 6-8 таблица сравнения относительных частот и теоретических вероятностей должна иметь вид

Название распределения

Интервал	w_i	p_i	$\lvert w_i - p_i\rvert$
$[a_0, a_1]$			
$(a_1, a_2]$			
…			
$[a_{m-1}, a_m]$			
	$\sum_{i=1}^{m} w_i$	$\sum_{i=1}^{m} p_i$	$\Delta_{\max}$

где w_i – относительная частота попадания в i-ый интервал, p_i – теоретическая вероятность попадания в i-ый интервал, $\Delta_{\max} = \max\{|w_i - p_i|,\ i = 1,\dots,m\}$.

Таблица сравнения рассчитанных характеристик с теоретическими значениями для всех Заданий должна иметь вид

Название распределения

Название показателя	Эксперимен-тальное значение	Теоретическое значение	Абсолютное отклонение	Относительное отклонение
Выборочное среднее				
Выборочная дисперсия				
Выборочное среднее квадратичное отклонение				
Выборочная мода				
Выборочная медиана				
Выборочный коэффициент асимметрии				
Выборочный коэффициент эксцесса				

Абсолютное отклонение – модуль разности экспериментального и теоретического значений.
Относительное отклонение – отношение абсолютного отклонения к теоретическому значению, если теоретическое значение равно нулю, то записывается ‘ – ‘.