Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

«Волгоградский государственный технический университет»

(ВолгГТУ)

Кафедра: «Высшая математика»

Протокол лабораторной работы №1

по дисциплине «Теория вероятностей, математическая статистика и случайные процессы»

На тему: “Первичная статистическая обработка данных для одномерной генеральной совокупности ”

Подготовили студенты группы ИВТ – 360

ФИО студентов: Князев А. А.

Скориков А.В.

ФИО преподавателя: Андреева М.И.

г. Волгоград, 2014

**Теоретическая часть**

**Вариация признаков в совокупности и её значения**

Обычно полученные наблюдаемые данные представляют собой множество расположенных в беспорядке чисел. Просматривая это множество чисел, трудно выявить какую-либо закономерность их варьирования (изменения). Для изучения закономерностей варьирования значений случайной величины опытные данные подвергают обработке.

**Пример.** Проводились наблюдения над числом Х оценок полученных студентами ВУЗа на экзаменах. Наблюдения в течение часа дали следующие результаты: 3; 4; 3; 5; 4; 2; 2; 4; 4; 3; 5; 2; 4; 5; 4; 3; 4; 3; 3; 4; 4; 2; 2; 5; 5; 4; 5; 2; 3; 4; 4; 3; 4; 5; 2; 5; 5; 4; 3; 3; 4; 2; 4; 4; 5; 4; 3; 5; 3; 5; 4; 4; 5; 4; 4; 5; 4; 5; 5; 5. Здесь число **Х** является **дискретной случайной величиной**, а полученные о ней сведения представляют собой **статистические (наблюдаемые) данные**.

Расположив приведенные выше данные в порядке **неубывания** и **сгруппировав** их так, что в каждой отдельной группе значения случайной величины будут одинаковы, получают **ранжированный ряд** данных наблюдения.

В примере имеем четыре группы со следующими значениями случайной величины:   
2; 3; 4; 5. Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называют **вариантом**, а изменение этого значения **варьированием**.

По характеру вариацииразличают дискретные и непрерывные признаки.

**Дискретные** признаки отличаются друг от друга на некоторую конечную величину, т.е. даны в виде конкретных чисел. (Например, число детей в семье).

**Непрерывные** признаки могут отличаться друг от друга на сколь угодно малую величину и в определенных границах принимать любые значения. (Например, зарплата рабочих, % выполнения).

Варианты обозначают малыми буквами латинского алфавита с соответствующими порядковому номеру группы индексами - **xi**. Число, которое показывает, сколько раз встречается соответствующий вариант в ряде наблюдений называют частотой варианта и обозначают соответственно - **mi**.

Сумма всех частот ряда - **объем выборки** (для примера *n* = 60). Отношение частоты варианта к объему выборки называют **относительной частотой**.

**Статистическим распределением выборки**  называют перечень вариантов и соответствующих им частот или относительных частот.

Статистический ряд частот:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| xi | 2 | 3 | 4 | 5 |
| mi | 8 | 12 | 23 | 17 |

Статистический ряд относительных частот:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| xi | 2 | 3 | 4 | 5 |
| wi | 0.133 | 0.2 | 0.383 | 0.284 |

Контроль: 0.133 + 0.2 + 0.383 + 0.284 = 1

Статистическое распределение можно задать также *в виде последовательности интервалов* и соответствующих им частот (в качестве частоты, соответствующей интервалу, принимают сумму частот, попавших в этот интервал).

Интервал указывает пределы значений варьирующего признака и обозначаются нижней и верхней границами интервала. Такие распределения наиболее распространены в практике статистической работы.

При построении интервальных рядов необходимо прежде всего установить число групп (интервалов). Для этого нужно определить величину интервала (**h**). Для построения вариационного ряда с равными интервалами следует определить **размах вариации (R)**- разность между максимальным и минимальным значением признака: R = x max - x min;

Размах вариации делится на число групп **k**, т.е. .

Число групп приблизительно определяется по формуле Стерджесса k ≈ 1+3,322 lg n,

где **n** - число изучаемых единиц совокупности. Это выражение, почти всегда дробное число, округляем до целого.

Величина интервала должна определяться в соответствии с точностью данных наблюдения: если исходные данные представлены целыми числами, то и величина интервала округляется до ближайшего целого числа.

Далее можно определить границы всех интервалов ряда распределения. Нижнюю границу I-го интервала можно принять равной минимальному значению признака.

При построении интервальных рядов для непрерывных признаков имеет место совпадение верхних границ предшествующих интервалов и нижних границ следующих за ними интервалом. В какой интервал относить единицы совокупности.

Если приведенный вариационный ряд с неравными интервалами, то для правильного представления о характере распределения необходимо рассчитать абсолютную и относительную плотности распределения.

*Абсолютная плотность:* ; *Относительная плотность:*

Эти показатели необходимы для преобразования интервалов изменения оценки данных, собранных по различным совокупностям и по разному обработанных.

**Графическое представление вариационного ряда**

Графическое изображение рядов расширения облегчает их анализ и позволяет судить о форме распределения. Для графического изображения дискретного ряда применяют **полигон распределения***.* На оси абсцисс отмечают точки, соответствующие величине варианты признака. Из них восстанавливаются перпендикуляры, высота которых - частости этих вариантов. Вершины перпендикуляров соединяются отрезками прямых. Крайние вершины соединяются с точками на оси абсцисс, отстоящими на одно деление от xmax и xmin.

Для графического изображения интервальных вариационных рядов применяется **гистограмма***.*

Она строится так, что на оси абсцисс откладываются равные отрезки, которые соответствуют величине интервалов вариационного ряда. На отрезках строят прямоугольники, площади которых пропорциональны частотам (частостям) интервала.

|  |  |
| --- | --- |
| Гистограмма относительных частот | Полигон относительных частот |
| C:\Users\Ю\Desktop\гистограмма.png |  |

Гистограмма может быть преобразована в полигон распределения, для чего середины верхних сторон прямоугольников соединяют отрезками прямых. Две крайние точки прямоугольников замыкаются по оси абсцисс на середины интервалов, в которых частоты равны 0.

При увеличении числа наблюдений совокупности увеличивается число групп интервального ряда, что соответственно приводит к уменьшению величины интервала. При этом ломанная линия будет иметь тенденцию превращения в плавную кривую, которую называют **кривой распределения***.* Она характеризует в обобщенном виде вариацию признака и распределение частот внутри однокачественной совокупности.

В ряде случаев для изображения вариационных рядов используется кумулятивная кривая (*кумулянта*).

При построении кумулянты интервального ряда распределения нижней границе первого интервала соответствует частота, равная 0, а верхней границе - вся частота данного интервала. Верхней границе второго интервала соответствует накопленная частота, равная сумме частот первых двух интервалов и т.д.

|  |
| --- |
| Кумулятивная кривая |
|  |

Изображение вариационного ряда в виде кумулянты особенно удобно при сравнении вариационных рядов, а так же в экономических исследованиях, в частности для анализа концентрации производства.

Помимо этого, для интервальных вариационных рядов при необходимости вычисляют **эмпирическую функцию распределения**. Её значения (накопленные частости) представляет собой сумму накопления значений элементов ряда относительных частот т.е.

Собственно сама эмпирическая функция распределения также имеет график. Она является разрывной для любой случайной величины *Х*, т.к. результаты наблюдений дискретны.

|  |
| --- |
| График эмпирической функции распределения |
|  |

**Числовые характеристики статистического распределения: выборочное среднее, оценки дисперсии.**

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины.

**Выборочной средней** называется среднее арифметическое значение случайной величины, принимаемых в выборке. Оно служит для оценки математического ожидания исследуемой случайной величины.

Выборочную среднюю находим по формуле:

где – середина частичного интервала.

**Выборочную дисперсию** ищем как центральный момент второго порядка по формуле:

Так же, для выборочной дисперсии справедлива формула:

Примечание. При *n* < 30 рекомендуется пользоваться формулой «исправленной» дисперсии:

Тогда выборочное среднее квадратическое отклонение будет:

**Практическая часть**

1. **Описание программы**

Данная программа позволяет без особого труда производить первичную статистическую обработку данных. Её функционал позволяет пользователю работать именно так, как ему удобно, а не так как этого требует от него сама программа. Этот фактор значительно облегчает и ускоряет процесс взаимодействия пользователя и программы.

Но, дабы не говорить об объекте абстрактно давайте ознакомимся с конкретными функциональными моментами.

1. Ввод данных можно осуществлять четырьмя различными способами: считыванием таблицы для анализа из файла, заполнением этой таблице вручную (если это необходимо), путем заполнения вручную таблицы интервального статистического ряда частот и также путем ручного ввода данных в таблицу интервального статистического ряда частостей. При любом из предложенных способов ввода данных программа сможет достоверно распознать и обработать полученные данные.

2. Весь функционал программы представлен в одном небольшом окне, что делает его достаточно удобным, поскольку пользователю не придется теряться в ряде окон дабы найти необходимое ему в данный момент.

3. При необходимости детально изучить, какие либо из графиков пользователь может воспользоваться соответствующей кнопкой «Лупы» в левом нижнем углу под каждым из графиков.

Таким образом, программа является простым, и в то же время удобным средством расчета.

1. **Работа программы**

С учетом внутреннего (программного) и внешнего (пользовательского) выполнения программы приложение работает следующим образом:

* Ввод данных экспериментов:

1. Выбор файла с данными (значения в файле должны быть представлены через пробел)
2. Ввод данных экспериментов в таблицу вручную
3. Ввод данных в таблицу «Интервальный статистический ряд частот»
4. Ввод данных в таблицу «Интервальный статистический ряд частостей»

* Ввод данных для расчета
* Если использовались пункты 1 и 2, ввести количество интервалов
* Если использовались пункты 3 и 4, ввести сумму частот
* Расчет данных – нажать кнопку «Вычислить»
* Программа осуществляет расчет данных и отрисовку графиков
* Просмотр графиков - нажать кнопку с изображением «Лупы» под интересующим графиком
* Программа осуществляет отрисовку выбранного графика в отдельном окне

С учетом внутреннего (программного) и внешнего (пользовательского) выполнения программы приложение работает следующим образом:

* выбирается один из типов вычисления значений комбинаторики;
* необходимые данные вводятся в поля ввода;
* посредством нажатия соответствующей клавиши осуществляется запуск вычисления;
* происходит вычисление значения (в соответствии с выбранным типом вычислений);
* выводятся данные.

1. **Примеры работы программы**

**Вариант 7**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 40,1 | 40,5 | 40,3 | 40,7 | 40,6 | 40,2 | 40,6 | 40,4 | 40,8 | 40,7 |
| 40,3 | 40,1 | 40,7 | 40,3 | 40,6 | 40,5 | 40,8 | 40,6 | 40,6 | 40,4 |
| 40,5 | 40,6 | 40,1 | 40,6 | 40,2 | 40,5 | 40,3 | 40,5 | 40,4 | 40,6 |
| 40,5 | 40,3 | 40,7 | 40,5 | 40,9 | 40,8 | 40,5 | 40,6 | 40,7 | 40,6 |
| 40,6 | 40,2 | 40,6 | 40,6 | 40,5 | 40,3 | 40,8 | 40,4 | 40,5 | 40,4 |
| 40,4 | 40,7 | 40,6 | 40,2 | 40,6 | 40,9 | 40,3 | 40,7 | 40,6 | 40,6 |
| 40,7 | 40,6 | 40,5 | 40,8 | 40,5 | 40,6 | 40,8 | 41,0 | 40,4 | 41,0 |
| 40,6 | 40,4 | 40,8 | 40,2 | 40,9 | 40,4 | 40,5 | 40,3 | 40,5 | 40,7 |
| 40,7 | 40,5 | 40,5 | 40,8 | 40,3 | 40,9 | 40,6 | 41,0 | 40,5 | 40,4 |
| 40,6 | 40,7 | 40,8 | 40,5 | 40,8 | 40,6 | 40,6 | 40,5 | 40,5 | 40,6 |

|  |  |
| --- | --- |
|  |  |
|  |  |

**Интервальный статистический ряд частот**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Xi | [40.1 ; 40.25) | [40.25 ; 40.4) | [40.4 ; 40.55) | [40.55 ; 40.7) | [40.7 ; 40.85) | [40.85 ; 41.0] |
| Mi | 8 | 9 | 30 | 25 | 21 | 7 |

**Интервальный статистический ряд частостей**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Xi | [40.1 ; 40.25) | [40.25 ; 40.4) | [40.4 ; 40.55) | [40.55 ; 40.7) | [40.7 ; 40.85) | [40.85 ; 41.0] |
| Wi | 0.08 | 0.09 | 0.3 | 0.25 | 0.21 | 0.07 |

**Эмпирическая функция распределения**

**Выборочная средняя**

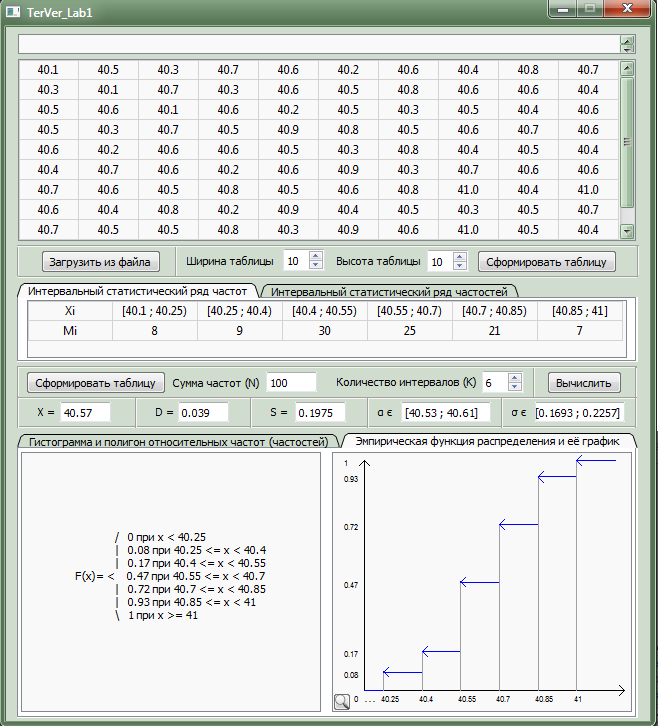
= 40.577

**Выборочная дисперсия**

=

**Выборочное среднее квадратическое отклонение**

**Результат работы программы**



**Вариант 15**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.7 | 3.6 | 3.7 | 3.3 | 3.7 | 3.5 | 3.8 | 3.6 | 3.7 | 3.8 |
| 3.1 | 3.2 | 3.1 | 3.9 | 3.4 | 3.3 | 3.5 | 3.4 | 3.5 | 3.8 |
| 3.1 | 3.8 | 3.7 | 3.3 | 3.4 | 3.5 | 3.3 | 3.6 | 4.0 | 3.6 |
| 3.6 | 3.1 | 3.6 | 3.5 | 4.0 | 3.8 | 3.4 | 3.5 | 3.8 | 3.6 |
| 3.5 | 3.7 | 3.2 | 3.7 | 3.5 | 3.3 | 3.7 | 3.4 | 3.5 | 3.8 |
| 3.7 | 3.6 | 3.5 | 3.3 | 3.8 | 3.5 | 3.5 | 3.8 | 3.6 | 3.6 |
| 3.6 | 3.2 | 3.6 | 3.6 | 3.3 | 3.7 | 3.3 | 3.6 | 3.4 | 3.7 |
| 3.5 | 3.8 | 3.2 | 3.5 | 3.5 | 3.4 | 3.5 | 3.5 | 3.7 | 3.8 |
| 3.8 | 3.5 | 3.2 | 3.5 | 3.6 | 3.6 | 3.4 | 3.9 | 3.5 | 3.7 |
| 3.6 | 3.6 | 3.6 | 3.8 | 3.3 | 3.4 | 3.4 | 3.7 | 3.8 | 3.8 |

|  |  |
| --- | --- |
|  |  |

**Интервальный статистический ряд частот**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Xi | [3.1 ; 3.25) | [3.25 ; 3.4) | [3.4 ; 3.55) | [3.55 ; 3.7) | [3.7 ; 3.85) | [3.85 ; 4.0] |
| Mi | 9 | 9 | 30 | 19 | 29 | 4 |

**Интервальный статистический ряд частостей**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Xi | [3.1 ; 3.25) | [3.25 ; 3.4) | [3.4 ; 3.55) | [3.55 ; 3.7) | [3.7 ; 3.85) | [3.85 ; 4.0] |
| Wi | 0.09 | 0.09 | 0.3 | 0.19 | 0.29 | 0.04 |

**Эмпирическая функция распределения**

**Выборочная средняя**

= 3.578

**Выборочная дисперсия**

=

**Выборочное среднее квадратическое отклонение**

**Результат работы программы**

