Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра программного обеспечения информационных технологий

Дисциплина: Теория вероятностей и математическая статистика (ТВиМС)

ОТЧЕТ:

По расчетной работе

Выполнил

студент: гр. 751003 Гринчик В.В.

Проверили: Лапицкая Н.В.

Петюкевич Н.С.

Минск 2019

СОДЕРЖАНИЕ

Оглавление

[**1 Постановка задачи** 3](#_Toc9858891)

[**2 Теория и использованные формулы** 4](#_Toc9858892)

[**3 Анализ данных** 9](#_Toc9858893)

[3.1 Дискретная случайная величина 9](#_Toc9858894)

[3.2 Непрерывная случайная величина 15](#_Toc9858895)

[**Вывод** 24](#_Toc9858896)

## **1 Постановка задачи**

В результате статистических наблюдений некоторой совокупности относительно изучаемого признака Х были получены выборочные данные (данные из открытых источников, размер выборки в диапазоне от 100 до 1000, 2 набора – дискретные и непрерывные).

ТРЕБУЕТСЯ:

1. составить дискретный или интервальный ряд распределения частот и частостей случайной величины Х;

2. построить полигон или гистограмму частот;

3. найти эмпирическую функцию распределения признака Х и построить её график;

4. вычислить все возможные числовые характеристики выборочных данных и проанализировать полученные результаты (близость дисперсии к нулю, соотношение между математическим ожиданием и дисперсией, близость оценок средней, моды и медианы, ассиметрия и эксцесс);

5. проверить качество группировки с помощью доли межгрупповой дисперсии в общей дисперсии выборки (правило сложения дисперсий)

6. выдвинуть гипотезу о виде распределения рассматриваемой случайной величины Х. Обосновать выбор вида распределения.

7. Написать аналитическое выражение теоретической функции распределения, найти теоретические (выравнивающие частоты), построить их графики и сравнить теоретические и эмпирические характеристики графически

8. приняв уровень значимости 0.05 по критериям согласия Пирсона, Романовского, Колмогорова, Ястремского, подтвердить или отвергнуть выдвинутую гипотезу о виде распределения.

## **2 Теория и использованные формулы**

В данном разделе приведены основные формулы и определение, согласно которым были произведены расчёты и построены соответствующие графики.

**Частота** - число единиц совокупности с данным значением признака Ni. Сумма Ni-тых должна быть равна количеству всех элементов ряда. Ниже приведены виды частот.

*Частота для дискретного ряда* – подсчет количества исходов, где варианта имела некое значения xi.

*Частота для интервального ряда* – подсчет количества исходов, где варианта оказалось на промежутке (x1…x2).

**Частотность**

mi – частота варианты;

N - объём совокупности.

**Оценка математического ожидания**

x – значения вариант;

mi – частота варианты.

**Формула Стерджесса**:



N – количество элементов в ряду.

h - величина интервала;

xmax – наибольшее значение группового признака;

xmin - наименьшее значение группового признака;

**Математическое ожидание**:

хі  - значение вариант

рі  - соответствующая вероятность

**Дисперсия**:

хі – значение вариант

хср – среднее значение

Мі – соответствующее математическое ожидание

**Оценка дисперсии**

хі – значение вариант

рі - вероятность

**Среднее квадратическое отклонение**:

D– дисперсия

**Мода**: Модой называется наиболее часто встречающееся значение признака у единиц совокупности мода в итерационных рядах.

– нижняя граница модального интервала

- величина интервала

– частота модального интервала

– частота интервала, предшествующего модальному

– частота интервала, следующего за модальным

**Медиана**: Медианой называют значение признака, делящего ранжированный ряд пополам. Медиана есть для дискретного и интервального ряда, что является довольно очевидным фактом.

*Медиана дискретного ряда* – зависит от накопленных частот.

*Медиана в интервальном ряду*

– нижняя граница интервала, который содержит медиану

- величина интервала

– частота медианного интервала

– сумма частот

– сумма накопленных частот интервалов, предшествующих медианному

**Размах вариации:**

хmax– максимальное значение

хmin–минимальное значение

**Коэффициент вариации:**

хср – среднее значение

– среднее квадратическое отклонение

**Асимметрия:**

– момент третьего порядка

– среднее квадратическое отклонение

**Эксцесс:**

– момент четвертого порядка

– среднее квадратическое отклонение

**Центральный момент m-порядка:**

– частота

хі – значение

хср – среднее значение

**Критерий согласия Пирсона**:



mi – эмпирическая частость

m’i – теоретическая частость

**Критерий Романовского:**



 - наблюдаемое значение

k – количество степеней свободы

**Критерий Ястремского**:



k – число степеней свободы

n – число групп

Z = 0.6, при n < 20

**Теорема сложения дисперсий**

******

- общая дисперсия

- средняя из групповых дисперсий

- межгрупповая дисперсия

**Групповая средняя:**

xi – очередное значение в группе

n – объём группы

**Общая средняя:**

k – количество групп

ni – объём очередной группы

xji – очередное значение из группы ni

**Групповая дисперсия:**

xi – очередное значение в группе

n – объём группы

xгр.ср. – групповая средняя данной группы

**Внутригрупповая дисперсия:**

k – количество групп

Dгрi – групповая дисперсия очередной группы

ni – объём очередной группы

**Межгрупповая дисперсия:**

k – количество групп

ni – объём очередной группы

xсрi – групповая средняя очередной группы

xср – общая средняя

## **3 Анализ данных**

## 3.1 Дискретная случайная величина

В качестве исходного дискретного ряда была взята статистика направления ветра в городе Минск с 01.02.2019 по 31.05.2019 (1 - Север, 2 - Северо-Восток, 3 - Восток, 4 - Юго-Восток, 5 - Юг, 6 - Юго-Запад, 7 - Запад, 8 - Северо-Запад). В результате ряд состоит из 113 значений. Ниже на рисунке 1 приведены скриншоты части таблицы с исходным рядом.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Все значения | 8 | 7 | 1 | 2 | 1 | 3 | 4 |

Рисунок 1 – Часть таблицы дискретного ряда

1) Далее на основе полученных данных была построена таблица частот и частостей, показанная на рисунке 2 и рисунке 3, где mi – обозначает, сколько раз встречается xi, а wi – отношение mi общему количеству элементов ряда n = 119.



Рисунок 2 – Таблица частот



Рисунок 3 – Таблица частостей

2) По данным из вышеприведённой таблицы были построены полигон частот и полигон относительных частот, которые представлены на рисунке 4 и рисунке 5.

Рисунок 4 – Полигон частот

Рисунок 5 – Полигон относительных частот (частостей)

3) Эмпирическая функция распределения составляется на основе значений вариант и накопленных ими частот, как видно на рисунке 6. На рисунке 7 также приведён график эмпирической функции.

Рисунок 6 – Эмпирическая функция распределения

Рисунок 7 – График эмпирической функции распределения

4) Для исходных данных были вычислены все возможные числовые характеристики: математическое ожидание, оценка математического ожидания, дисперсия, оценка дисперсии, среднее квадратическое отклонение, среднее линейное отклонение, мода, медиана, размах вариации, коэффициент вариации, асимметрия и эксцесс. Для нахождения асимметрии и эксцесса также необходимо было дополнительно вычислить центральные моменты 3 и 4 порядков соответственно. Результаты вычислений представлены на рисунке 8. Формулы, по которым производились вычисления, представлены в разделе «Используемые формулы».



Рисунок 8 – Расчеты числовых характеристик

После выполнения расчётов мы можем сделать некоторые выводы по следующим пунктам:

* близость дисперсии к нулю: полученное нами значение дисперсии равно 4.95, что является небольшим числом, а это означает, что значения случайной величины сравнительно близки друг к другу;
* асимметрия: полученное нами отрицательное значение показывает левостороннюю асимметрию;
* эксцесс: полученное нами отрицательное значение показывает пологое распределение нашего ряда.

5) Хоть полученные значения математического ожидания и дисперсии не равны, они близки по значение, из-за чего будет рассматривать вид распределения Пуассона. Таким образом функция распределения Пуассона будет выглядеть следующим образом:

, где λ – средняя взвешенная величина.

Теоретические (выравнивающие) частоты находятся по формуле:

, где N – число испытаний. В качестве используется функция Пуассона, где .



Рисунок 9 – Расчет Пуассоновского распределения

Исходя из полученных значений, был построен график функции распределения для рассматриваемого Пуассоновского распределения. Данный график представлен на рисунке 10.

Рисунок 10 – График функции для теоретического распределения

График, представленный на рисунке 10, имеет определенные отличия от графика на рисунке 7, возможно выбранное теоретическое распределение будет опровергнуто.

6) По критерию согласия Пирсона были посчитаны χ2 наблюдаемое и χ2 критическое (с использованием соответствующих формул). На основании полученных результатов был сделан следующий вывод: Поскольку неравенство χ2 наблюдаемое ≤ χ2 критическое не выполняется, то выдвинутая гипотеза отвергается.

Критерий согласия Пирсона высчитывается по формуле  и сравнивается по таблице с критическим значением , значением где d – число степеней свободы d = n – r - 1 = 8 – 1 – 1 = 6. (где n – число уникальных вариант, r – число параметров у выбранного теоретического распределения). Вычисленный критерий меньше критического значения, значит, гипотеза не опровергается и она может быть принята. Критерий Романовского высчитывается по формуле . Если вычисленное значение меньше 3, следовательно, гипотеза снова подтверждается.

Как видно из рисунка 8, ни в одном из критериев не выполнились условия подтверждения гипотезы.



Рисунок 11 – Критерии Пирсона и Романовского

## 3.2 Непрерывная случайная величина

В качестве исходного непрерывного ряда были взяты результаты замеров температуры в городе Минск с 01.02.2019 по 31.05.2019. На рисунке 12 показана таблица значений.



Рисунок 12 – Данные температуры в Минске

Далее было произведено разбиение на интервалы в соответствии с формулой Стерджесса рисунок 13. Таблица интервалов приведена ниже на рисунке 14.



Рисунок 13 – Формулы Стерджесса для данных о погоде



Рисунок 14 – Интервальный ряд данных

На следующем шаге, проанализировав интервалы разбиения, был построена плотности вероятности, график которой приведён на рисунке 15.

Интервал

Рисунок 15 – Гистограмма частости данных о погоде

Эмпирическая функция распределения составляется на основе значений вариант и накопленных ими частот. Составленная функция предоставлена на рисунке 16. На основе полученной функции был построен график, представленный на рисунке 17.

Рисунок 16 – Эмпирическая функция распределения

Рисунок 17 – График эмпирической функции распределения

Исходя из внешнего вида гистограммы частот и эмпирической функции распределения, было сформировано предположение, что ряд распределён по нормальному закону распределения.

Для исходных данных были вычислены все возможные числовые характеристики: математическое ожидание, оценка математического ожидания, дисперсия, оценка дисперсии, среднее квадратическое отклонение, среднее линейное отклонение, мода, медиана, размах вариации, коэффициент вариации, асимметрия и эксцесс. Для нахождения асимметрии и эксцесса также необходимо было дополнительно вычислить центральные моменты 3 и 4 порядков соответственно. Результаты вычислений представлены на рисунке 18. Формулы, по которым производились вычисления, представлены в разделе «Используемые формулы».

|  |  |
| --- | --- |
| Числовые характеристики выборочных данных | |
| Математическое ожидание (M(X)) | 8,101101501 |
| |  | | --- | |  | | 136,82 |
| |  | | --- | |  | | 65,62784552 |
| |  | | --- | | Оценка математического ожидания | | 8,101101501 |
| Дисперсия (D(X)) | 71,19 |
| Среднее квадратическое отклонение (X) | 8,437692274 |
| |  | | --- | | Среднее линейное отклонение | | 7,33935056 |
| Оценка дисперсии | 71,19465091 |
| Мода (Мо) | 2 |
| Медиана (Ме) | 6 |
| Размах вариации ( R) | 32 |
| Коэффициент вариации (V(X)) | 1,041548766 |
| Центральный момент 3-го порядка | 192,003096 |
| Центральный момент 4-го порядка | 9079,338418 |
| Ассиметрия | 0,319622382 |
| Эксцесс | -1,208736509 |



Рисунок 18 – Числовые характеристики

Исходя из предположений о том, что полученное распределение данных о температуре – нормальное распределение, то функция распределения будет выглядеть следующим образом:

*,* где µ - математическое ожидание, σ – среднеквадратическое отклонение, .

Mi' = φ = =

На рисунке 19 представлены расчёты по выбранному распределению.



Рисунок 19 – Таблица расчетов экспоненциального распределения

Исходя из полученных значений, был построен график функции распределения для рассматриваемого нормального распределения. Данный график представлен на рисунке 20.

Рисунок 20 – График функции теоретического распределения

График, представленный на рисунке 20, определенно имеет сходство с графиком на рисунке 17, возможно выбранное теоретическое распределение будет подтверждено.

Для подтверждения гипотезы был найден наблюдаемый критерий согласия Пирсона. Формула нахождения критерия находится в разделе «Использованные формулы». Найденное число степеней свободы и принятый уровень значимости 0.05 был найден критический критерий. После сравнения, наблюдаемого и критического критериев между собой, мы не можем принять выдвинутую ранее гипотезу, так как наблюдаемый критерий больше критического. Также был посчитан критерий Романовского, где мы получили критерий R > 3, из-за чего критерий отвергается. Результаты подсчета показаны на рисунке 21.



Рисунок 21 – Критерий Пирсона, Романовского для экспоненциального распределения

На примере данных о погоде проведена проверка качества разбиения на интервалы. Для этого используем правило сложения дисперсий:

******

На основе этих данных рассчитываем общую ****** и межгрупповую дисперсию****** , среднюю из внутригрупповых дисперсий***.***



Рисунок 22 – Таблица проверки выполнения формулы

Отношение межгрупповой дисперсии к общей составило 98%, а также сумма межгрупповой и средней внутригрупповой дисперсии приблизительно равна общей дисперсии, следовательно, разбиение на группы проведено успешно.

## **Вывод**

В ходе проделанной работы были найдены следующие данные: данные замеров температуры в городе Минске в качестве непрерывной случайной величины и данные о направлении ветра в городе Минске. Эти данные были загружены в Excel для дальнейшей работы с ними.

Для непрерывных случайных величин было произведено разбиение на интервалы, построены гистограммы частостей, найдены эмпирическую функцию распределения и построены её графики, вычислены всевозможные числовые характеристики, проведён анализ качества разбиения на группы с помощью правила сложения дисперсий; выдвинуто предположение о виде распределения, найдены критерии согласия Пирсона и Романовского, благодаря которым опровергнули выдвинутый вид распределения.

Для дискретных случайных величин была сформирована таблицу частот и частостей, на основе которых построен полигон относительных частот, найдена эмпирическая функция распределения и простроен её график, вычислены всевозможные числовые характеристики, произведена проверка на пуассоновское распределение. Можно заметить, что в распределении Пуассона дисперсия равна математическому ожиданию (среднему взвешенному). В данном случае величины не равны между собой, но очень близки друг к другу. Приведённые значения могут означать только одно - это не распределение Пуассона. Но тем не менее была проверена теория распределения Пуассона, которая не оправдалась.