Министерство науки и высшего образования РФ

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образовани**я

**«Томский государственный университет систем управления**

**и радиоэлектроники» (ТУСУР)**

Кафедра автоматизированной обработки информации (АОИ)

**ОПИСАТЕЛЬНАЯ СТАТИСТИКА**

Отчет по лабораторной работе № 1

По дисциплине «Теория вероятностей и

математическая статистика»

Вариант №2

Выполнил:

Студент гр.429-3

Бабец А. А.

Принял:

Доцент каф. Математики

Лугина Н.Э.

Томск 2021

Анализируемые данные – прирост численности городского населения за 2016 год. Объём выборки равен 41. Для работы данные были отсортированы, составлен вариационный ряд. Данные продемонстрированы в приложении 1.

а) Среднее арифметическое. Обозначение: . Формула:

,

где *n* – объём выборки, *xi*– варианта, *nj* – частота варианты, *k* – число элементов в статистическом ряду. Среднее арифметическое – статистический аналог математического ожидания. По вариационному ряду находим среднее арифметическое с помощью функции СРЗНАЧ(). Таким образом, среднее арифметическое равно 6140.

Выборочная дисперсия. Обозначение: . Формула(так как *n*>30):

,

где *n* – объём выборки, *k* – число различных вариант выборки. Используя вариационный ряд, находим дисперсию по выборке. Её значение равно 564620295,8.

Выборочное среднеквадратическое отклонение. Обозначение: . Формула:

.

Таким образом, оно равно 10084,27016

Найдено и исключено одно аномальное значение, не попадающее в интервал . Объём выборки стал равен 40.

б) В математической статистике мода() определяется по выборке и является вариантой с наибольшей частотой. В данной выборке мода отсутствует, так как все варианты уникальны. Следовательно, выборка полимодальна.

Медиана делит выборку на две части: первая половина вариант меньше медианы, вторая – больше. Медиана найдена с помощью функции МЕДИАНА и равна -1078.

Размах выборки. Обозначение: *R*. Формула:

,

где *x(n)* – наибольшая варианта выборки, *x(1)* – наименьшая варианта выборки. Таким образом, размах выборки равен 77988. Размах можно оценить с помощью СКО по формуле и равно 12998. СКО, найденное с помощью размаха, сильно отличается от СКО, найденного ранее и является грубой оценкой.

в) Можно найти три числа *q1*, *q2*, *q3*, которые делят выборку аналогично медиане на 4 равные части. Это квартили. Число *q2* совпадает с медианой. Число *q1* – нижняя квартиль. Число *q3* – верхняя квартиль. Квартили *q1* и *q3* найдены с помощью функции КВАРТИЛЬ и соответственно равны 3462 и 3911.

Все данные из предшествующих пунктов занесены в таблицу в ячейки под соответствующими подписями. См. приложение 1.

г) Для получения сгруппированного статистического ряда необходимо разделить диапазон значений выборки на интервалы и найти количество попадающих в интервалы значений. Количество интервалов рассчитаем по формуле Старджеса . Оно равно приблизительно 6,28. Таким образом получаем 6 интервалов. Длину интервалов получаем, поделив размах выборки на количество интервалов . Таким образом получаем длину интервалов, равную 12405.

Формулы математического ожидание и дисперсии отличаются от вышеупомянутых тем, что вместо *xj* используются середины и частота интервалов. Так математическое ожидание равно 29051,47855, а дисперсия 30440. Неточность обоснована использованием середин интервалов.

д) Гистограмма построена по данным сгруппированного статистического ряда с помощью мастера диаграмм. Модальный интервал(с наибольшей частотой) [-6775:5630). Середина этого интервала - -572,5. С модой сравнить не можем, так как выборка полимодальная.

е) Сгруппированный статистический ряд из накопленных частот представлен в приложении 1. Его диаграмма является графиком эмпирической функции распределения F. Медианным интервалом является интервал [-6775:5630). Его середина - -572,5, отличается от ранее найденной медианы, равной -1078. Таким образом, медианный интервал действительно содержит медиану выборки, однако она не является его серединой.

Далее идёт приложение 1.

