Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное

образовательное учреждение высшего образования

«Воронежский государственный лесотехнический

университет имени Г.Ф. Морозова»

Кафедра автоматизации производственных процессов

Лабораторная работа №2

по дисциплине «Технологии обработки информации»

«Описательная статистика»

Выполнил ст-т гр. ИС2–191–ОБ Голубятников И.С.

Проверил: к.т.н. доц. Мещерякова А.А.

Воронеж 2021

Цель работы: получение навыков расчета в среде Scilab параметров центральной тенденции и вариации.

Теоретическая часть

Мода – наиболее часто встречающееся значение в выборке, наборе данных. В случае, если данные сгруппированы и построено распределение частот, модой является значение, имеющее наибольшую частоту.

Медиана определяется как серединное значение выборки, или значение, выше и ниже которого располагается одинаковое число наблюдений. Для нахождения медианы обязательно упорядочить данные. Медиана является точной серединой выборки. Обозначается Me и определяется по-разному для выборок с четным и нечетным числом элементов. Для нечетного количества наблюдений медиана есть наблюдение с номером (n+1)/2. Для четного количества наблюдений медиана вычисляется как среднее значение наблюдений с номерами n/2 и (n+2)/2 В случае нечетного количества наблюдений медиана есть просто середина выборки, выше и ниже которой располагается одинаковое количество наблюдений.

Среднее определяется как среднее арифметическое выборки, то есть как сумма всех значений выборки, деленная на ее объем. Следуя определению, будем находить среднее значение по формуле:

.

где - сумма всех значений выборки,

n - объем выборки.

Взвешенное среднее – это среднее значение, получаемое при объединении нескольких групп наблюдений.

Размах – это разница между наибольшим и наименьшим значениями. Для нахождения размаха прежде рекомендуется упорядочить данные в порядке возрастания. Можно записать размах с помощью формулы:

*.*

Квартили – это значения, которые делят вариационный ряд на четыре равные по объему части. Квартильный размах (Inter Quartile Range - IRQ) – это разница между третьим и первым квартилями. Таких значений должно быть три: первая, вторая и третья квартиль соответственно. Для начала данные следует упорядочить. Затем отыскивается медиана, которая является вторым квартилем по определению. После этого находятся первый и третий квартили. Существует несколько вариантов формального определения квартилей.

Квартильный размах находится по формуле:

.

Дисперсия для набора данных или выборки – это среднее арифметическое квадратов отклонений значений от их среднего.

Дисперсия обозначается . Основная формула (по определению) для нахождения дисперсии:

.

Стандартное отклонение – квадратный корень из дисперсии выборки.

Обозначается s и вычисляется по формуле:

.

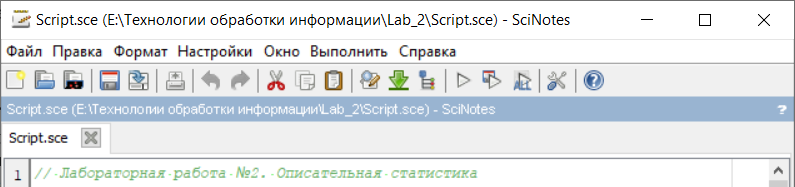
Коэффициент вариации вычисляется как отношение стандартного отклонения к среднему значению выборки.

Формула для коэффициента вариации:

.

Задания

1. Создать скрипт для выполнения работы и сохранить его. В первой строке написать комментарий, содержащий название работы:



2. Сгенерировать последовательность случайных чисел размером 40:



3. Преобразовать выборку в целые числа, умножив каждый элемент на 100 и округлив до целого значения:



--> sample\_A

sample\_A =

column 1 to 12

2. 8. 0. 3. 7. 6. 8. 7. 9. 1. 6. 7.

column 13 to 24

7. 2. 5. 2. 2. 2. 9. 7. 3. 9. 2. 3.

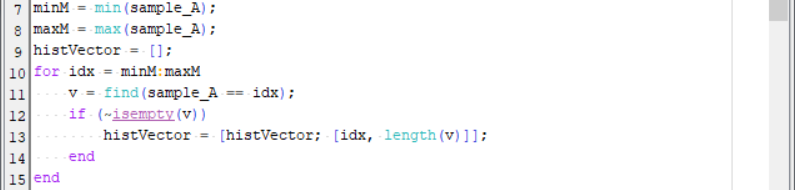
column 25 to 36

4. 3. 6. 5. 3. 6. 5. 4. 3. 6. 4. 9.

column 37 to 40

1. 5. 3. 4.

4. Найти моду распределения. для этого нужно определить, какие значения встречаются в последовательности, и сколько раз встречается каждое из этих значений:



histVector =

0. 2.

1. 1.

2. 6.

3. 7.

4. 4.

5. 4.

6. 5.

7. 5.

8. 2.

9. 4.

5. Найти медиану и среднее распределения:



--> median\_A

median\_A =

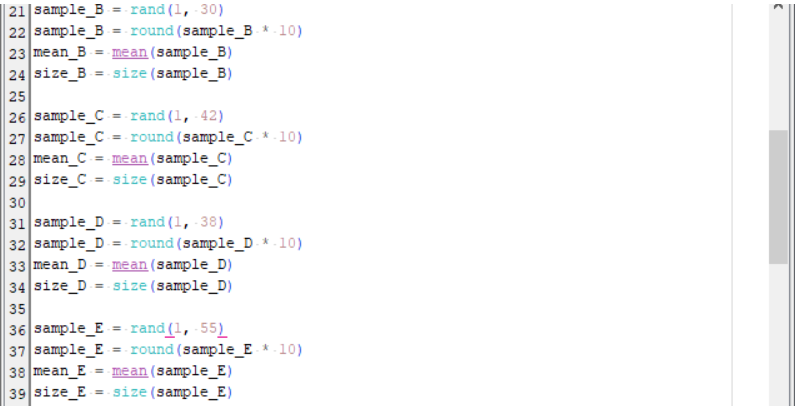
4.5

--> mean\_A

mean\_A =

4.675.

6. Сформировать четыре дополнительные последовательности произвольного объема от 30 до 60: sample\_В, sample\_С, sample\_D и sample\_E.



--> sample\_B

column 1 to 12

3. 1. 8. 2. 1. 7. 2. 7. 8. 4. 4. 9.

column 13 to 24

1. 2. 6. 6. 7. 9. 5. 3. 4. 9. 9. 3.

column 25 to 30

4. 7. 3. 5. 3. 5.

sample\_C

sample\_C =

column 1 to 12

5. 1. 2. 6. 8. 0. 7. 2. 4. 8. 6. 5.

column 13 to 23

2. 8. 1. 3. 9. 8. 5. 10. 6. 10. 1.

column 24 to 35

7. 4. 6. 9. 1. 8. 9. 6. 6. 8. 1. 6.

column 36 to 42

1. 7. 3. 5. 10. 7. 0.

--> sample\_D

sample\_D =

column 1 to 12

6. 3. 3. 6. 1. 6. 7. 3. 0. 5. 4. 2.

column 13 to 24

5. 4. 3. 1. 6. 3. 7. 5. 3. 7. 1. 4.

column 25 to 35

7. 9. 2. 4. 10. 5. 5. 6. 6. 5. 8.

column 36 to 38

8. 10. 8.

-> sample\_E

sample\_E =

column 1 to 12

4. 2. 9. 1. 5. 4. 0. 5. 8. 6. 2. 0.

column 13 to 24

8. 1. 9. 0. 2. 5. 7. 9. 2. 6. 3. 4.

column 25 to 36

9. 8. 8. 3. 4. 4. 7. 8. 4. 8. 5. 1.

column 37 to 47

10. 2. 0. 8. 1. 10. 7. 2. 5. 8. 8.

column 48 to 55

1. 8. 3. 3. 10. 9. 2. 9.

7. Вычислить среднее и размер получившихся последовательностей,

воспользовавшись для этого функциями mean и size . Найти произведения этих

величин для каждой последовательности (А-Е).

--> mean\_B

mean\_B =

4.9

--> size\_B

size\_B =

1. 30.

--> mean\_C

mean\_C =

5.2619048

--> size\_C

size\_C =

1. 42.

--> mean\_D

mean\_D =

4.9473684

--> size\_D

size\_D =

1. 38.

--> mean\_E

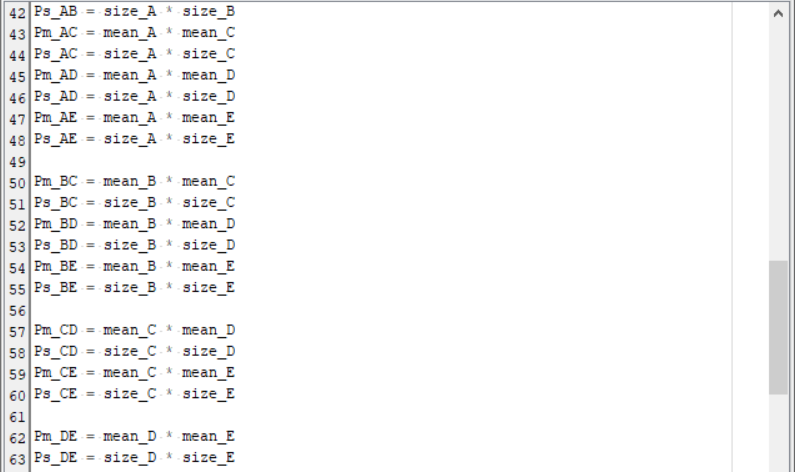
mean\_E =

5.0363636

--> size\_E

size\_E =

1. 55.



8. Найти общее число наблюдений, сложив размеры всех последовательностей.

--> size\_A + size\_B + size\_C + size\_D + size\_E

ans = 5. 205.

9. Найти взвешенное среднее получившихся последовательностей.

-->(mean\_A + mean\_B + mean\_C + mean\_D + mean\_E) / 5

ans = 4.9641274

10. Вычислить размах выборки А. Использовать функции нахождения максимального и минимального элементов min и max .

--> maxM - minM

ans = 9.

minM = min(sample\_A);

maxM = max(sample\_A);

11. Найти квартильный размах (функция iqr ) выборки А.

--> iqr\_A = iqr(sample\_A)

iqr\_A = 4.

12. Вычислить дисперсию (функция variance), стандартное отклонение и коэффициент вариации выборки A.

--> disp\_A = variance(sample\_A)

disp\_A = 6.5326923

--> SO\_A = sqrt(disp\_A)

SO\_A = 2.5559132

--> CV\_A = SO\_A / mean\_A

CV\_A = 0.5467194

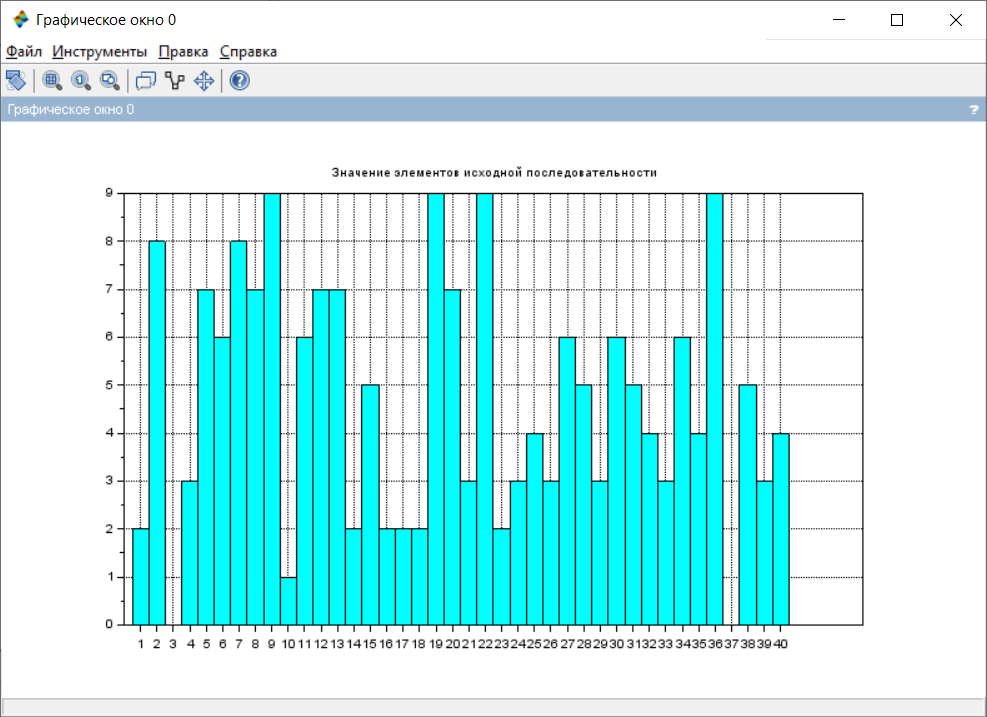
13. Построить гистограмму исходной последовательности А:

--> clf();

--> bar(sample\_A, 1, "cyan");

--> xtitle('Значение элементов исходной последовательности')

--> xgrid();



Вывод: создал скрипт для выполнения работы, сгенерировал последовательность случайных чисел и округлил до их до целых значений, нашел моду, медиану и среднее распрделения A. Сформировал четыре дополнительные последовательности B, C, D, E и вычислил их среднее и размер. Нашел общее число наблюдений и взвешенное среднее получившихся последовательностей. Вычислил размах, квартильный размах, дисперсию стандартное отклонение и коэффициент вариации выборки А. Построил гистограмму исходной последовательности A.