

Міністерство освіти і науки України
Національний університет «Запорізька політехніка»

Кафедра програмних засобів

ЗВІТ

Дисципліна «Емпіричні методи в інформаційних технологіях»

Робота №1

Тема «Статистичний аналіз і первинна обробка даних»

Виконав варіант 19

Студент КНТ-122

Онищенко О. А.

Прийняли

Викладач

Леоценко С. Д.

2024

Мета	3
Завдання	3
Результати	4
<i>Файли</i>	4
<i>Код</i>	9
Гістограма.....	10
Висновки	12
Питання	12
<i>Робота з даними у Statgraphics</i>	12
<i>Призначення і використання різних типів вікон</i>	12
<i>Загальні й унікальні властивості Statgraphics</i>	12
<i>Що таке квантіль, мода, медіана?</i>	13
<i>Що таке ряди розподілу?</i>	13
<i>Які характеристики розподілу ви знаєте?</i>	13
<i>Як будується гістограма?</i>	13
<i>Що таке таблиця частот?</i>	14
<i>На що впливає ширина інтервалу?</i>	14

Мета

Ознайомитися з можливостями пакетів статистичної обробки даних. Навчитися використовувати пакети статистичної обробки даних для первинного аналізу даних.

Завдання

Дослідити можливості програмного пакету опрацювання даних Statgraphics та мови програмування R.

Ознайомитися із загальними положеннями теорії ймовірностей та статистичної обробки даних.

Згенерувати стовпець на основі наступної інформації: $N = \text{Var} * 10$, $m = \text{Var}$, $s = \text{Var} / 10$, де Var - номер варіанта, N - кількість дослідів, m - математичне сподівання, s - дисперсія. Для пакету Statgraphics використати функцію `Rnormal(N, m, s)`.

Зберегти отриману вибірку у форматі *.txt.

Побудувати гістограму, таблицю частот, отримати описові статистики ряду розподілу.

Зберегти результати аналізу у форматі *.rtf.

Побудувати гістограму, таблицю частот, отримати описові статистики ряду розподілу з використанням внутрішніх функцій мови R або реалізувати функції мовою R самостійно.

Зробити висновок.

Оформити звіт.

Відповісти на контрольні питання.

Результати

Файли

stats.rtf

```
> frequency_table
dataset
13.6715527254682 14.3423502256557 15.0156090558024 15.5478702856091
1
1
1
1
15.7062927073258 15.7121874975163 15.7266105493884 16.0240136937913
1
1
1
1
16.0781625438537 16.1294922525253 16.138646039035 16.3411398226211
1
1
1
1
16.3580017591321 16.3694268773802 16.4301059705818 16.4356104407392
1
1
1
1
16.5852084191763 16.6177080699823 16.6533202449359 16.6642102372106
1
1
1
1
16.7074944021144 16.7216395296906 16.7309106872835 16.8002569939164
1
1
1
1
16.8762608928066 17.007998333176 17.0705808015429 17.1039741239526
1
1
1
1
17.1210726584239 17.139803796742 17.1728690301367 17.1846445224719
1
1
1
1
17.2023328269303 17.2051601746505 17.2428677586766 17.3644532268203
1
1
1
1
17.4564076171184 17.5710041829767 17.6370365382266 17.6371841788347
1
1
1
1
17.7041640066529 17.7384847202997 17.7624500829161 17.8033189291496
1
1
1
1
17.9387956095401 18.0199465987062 18.0322805879592 18.0777796500814
1
1
1
1
18.1114808495007 18.1946526058476 18.2465175009881 18.3161584918358
1
1
1
1
18.3177496424126 18.3208492552421 18.3294271995463 18.3685564832589
1
1
1
1
18.3726654399489 18.3732852726742 18.4342348763924 18.4378224289117
1
1
1
1
18.4867065327959 18.4998978572655 18.520002162031 18.5223049442725
1
1
1
1
18.5306771500907 18.5433033384619 18.573815513851 18.6885117309515
1
1
1
1
18.7036012166944 18.7771423039222 18.7957614539543 18.811568707614
1
1
1
1
18.8640882889928 18.8674782712331 18.8728751250295 18.8819841368853
1
1
1
1
18.9805311624323 18.9900063327416 18.9906210367805 19.0303574270705
1
1
1
1
19.0337910950775 19.040487776151 19.0407203983941 19.0815306124239
1
1
1
1
19.0911894541195 19.0955056037989 19.1318941462682 19.1575597717942
1
1
1
1
19.1852132828716 19.2208899698206 19.2298826069008 19.2368893199574
1
1
1
1
19.2528723298761 19.2761277423803 19.2863924110721 19.2899323283774
1
1
1
1
```

19.2946987465326	19.2949878911993	19.3247026572366	19.3579723591974
1	1	1	1
19.4189662336999	19.4240632738918	19.4326947450671	19.5055762401786
1	1	1	1
19.5405440596866	19.5488916471676	19.5566152821429	19.5903754285079
1	1	1	1
19.5958334476441	19.6390711384913	19.647131261848	19.6593534799528
1	1	1	1
19.6624125593478	19.6762068354891	19.707216476616	19.7178369062617
1	1	1	1
19.7226406140181	19.73605178009	19.7708678094232	19.7766699492596
1	1	1	1
19.7767914751404	19.7850223750251	19.7929062588424	19.8515730520766
1	1	1	1
19.9350981035798	20.0030416550156	20.0038005431154	20.0535941516939
1	1	1	1
20.0719225485817	20.078146589799	20.0788553596097	20.140206585534
1	1	1	1
20.2119046526501	20.2265613722399	20.2651156620327	20.3442009657781
1	1	1	1
20.3856835523974	20.4016667942954	20.4108517429317	20.4387009750966
1	1	1	1
20.5285262290446	20.5926502952856	20.6875972649464	20.7228173430738
1	1	1	1
20.7554720937043	20.8004374649324	20.8125177842668	20.8412435202877
1	1	1	1
20.87422214614	20.8767314569886	20.9372438710269	20.9462509319623
1	1	1	1
20.9766731770953	20.9889793281436	21.0181015274964	21.0188647563558
1	1	1	1
21.0263306171014	21.0779734032326	21.0854656806625	21.1196862775299
1	1	1	1
21.1371766396281	21.1866887925435	21.2018586112711	21.2780082970732
1	1	1	1
21.3059893594986	21.4384920893888	21.4570414586767	21.4761203779459
1	1	1	1
21.5491836380158	21.5954044124987	21.7729988388447	21.8752857596358
1	1	1	1
21.9117017373998	22.0174719693907	22.0616142754607	22.1292286050755
1	1	1	1
22.2143730061064	22.3262220700742	22.4071251011494	22.6960588073051
1	1	1	1
22.8416933003323	22.9206935983798	22.9494065338436	23.004185197952
1	1	1	1
23.2038043526357	23.3428361672514	23.4616669726454	23.5878289308928
1	1	1	1
23.6803809037772	23.9123294723199		
1	1		

```
> mean_value
[1] 19.28143
```

```
> median_value
[1] 19.28816
```

```
> mode_value
[1] 20.4387
```

```
> qunatiles
      0%      25%      50%      75%     100%
13.67155 18.08620 19.28816 20.66386 23.91233
```

data.txt

```
"x"
"1" 20.4387009750966
"2" 19.0911894541195
"3" 18.2465175009881
"4" 19.3247026572366
"5" 18.573815513851
"6" 19.5405440596866
"7" 17.4564076171184
"8" 16.4301059705818
"9" 20.0038005431154
"10" 21.5491836380158
"11" 20.8767314569886
"12" 19.73605178009
"13" 19.2899323283774
"14" 21.1866887925435
"15" 21.9117017373998
"16" 18.0777796500814
"17" 19.4326947450671
"18" 19.2528723298761
"19" 16.6533202449359
"20" 19.7929062588424
"21" 16.6642102372106
"22" 16.0240136937913
"23" 20.0030416550156
"24" 18.6885117309515
"25" 17.7624500829161
"26" 19.5055762401786
"27" 19.2208899698206
"28" 20.9372438710269
"29" 17.2051601746505
"30" 17.139803796742
"31" 22.0174719693907
"32" 13.6715527254682
"33" 19.2368893199574
"34" 19.7226406140181
"35" 21.2780082970732
"36" 19.8515730520766
"37" 18.811568707614
"38" 20.7228173430738
"39" 18.8674782712331
"40" 15.5478702856091
"41" 20.87422214614
"42" 20.4016667942954
"43" 19.0303574270705
"44" 16.3411398226211
"45" 18.4998978572655
"46" 17.1846445224719
"47" 20.078146589799
"48" 20.4108517429317
```

"49" 19.2761277423803
"50" 17.2023328269303
"51" 19.0955056037989
"52" 19.1852132828716
"53" 20.8004374649324
"54" 17.6370365382266
"55" 18.9805311624323
"56" 16.138646039035
"57" 17.7041640066529
"58" 15.7062927073258
"59" 18.4342348763924
"60" 21.1196862775299
"61" 17.6371841788347
"62" 18.7957614539543
"63" 18.0199465987062
"64" 21.4570414586767
"65" 22.9206935983798
"66" 19.9350981035798
"67" 17.007998333176
"68" 19.2946987465326
"69" 16.3694268773802
"70" 18.520002162031
"71" 19.2298826069008
"72" 19.1575597717942
"73" 23.2038043526357
"74" 23.004185197952
"75" 19.647131261848
"76" 18.3294271995463
"77" 19.7766699492596
"78" 17.0705808015429
"79" 22.0616142754607
"80" 18.3161584918358
"81" 22.8416933003323
"82" 18.7036012166944
"83" 19.2949878911993
"84" 17.1210726584239
"85" 23.3428361672514
"86" 21.0779734032326
"87" 16.5852084191763
"88" 23.6803809037772
"89" 19.707216476616
"90" 18.1114808495007
"91" 20.0535941516939
"92" 21.4761203779459
"93" 18.5223049442725
"94" 23.9123294723199
"95" 19.7850223750251
"96" 21.3059893594986
"97" 19.6593534799528
"98" 20.9889793281436
"99" 16.7216395296906
"100" 20.9766731770953
"101" 20.2651156620327
"102" 16.7074944021144
"103" 19.4189662336999
"104" 22.3262220700742
"105" 21.0181015274964
"106" 19.7178369062617

"107" 19.6624125593478
"108" 19.4240632738918
"109" 20.0788553596097
"110" 15.7266105493884
"111" 16.0781625438537
"112" 20.8412435202877
"113" 21.0188647563558
"114" 15.0156090558024
"115" 19.7708678094232
"116" 18.9906210367805
"117" 21.0854656806625
"118" 20.2119046526501
"119" 22.6960588073051
"120" 19.5566152821429
"121" 19.7767914751404
"122" 18.1946526058476
"123" 14.3423502256557
"124" 23.4616669726454
"125" 21.5954044124987
"126" 17.8033189291496
"127" 21.4384920893888
"128" 19.0337910950775
"129" 20.140206585534
"130" 20.3442009657781
"131" 18.9900063327416
"132" 16.3580017591321
"133" 19.5488916471676
"134" 16.8762608928066
"135" 19.1318941462682
"136" 17.2428677586766
"137" 20.5285262290446
"138" 18.3208492552421
"139" 19.5903754285079
"140" 19.040487776151
"141" 21.0263306171014
"142" 19.6762068354891
"143" 16.7309106872835
"144" 19.0815306124239
"145" 18.3685564832589
"146" 17.1039741239526
"147" 18.8819841368853
"148" 20.9462509319623
"149" 18.8728751250295
"150" 18.3177496424126
"151" 17.5710041829767
"152" 18.3726654399489
"153" 18.7771423039222
"154" 19.6390711384913
"155" 21.1371766396281
"156" 21.8752857596358
"157" 19.2863924110721
"158" 17.3644532268203
"159" 17.7384847202997
"160" 20.3856835523974
"161" 22.9494065338436
"162" 20.7554720937043
"163" 18.4378224289117
"164" 20.8125177842668


```
"165" 16.6177080699823
"166" 15.7121874975163
"167" 22.1292286050755
"168" 22.2143730061064
"169" 17.1728690301367
"170" 19.3579723591974
"171" 18.0322805879592
"172" 20.2265613722399
"173" 19.0407203983941
"174" 21.2018586112711
"175" 18.8640882889928
"176" 21.7729988388447
"177" 18.3732852726742
"178" 20.5926502952856
"179" 16.8002569939164
"180" 22.4071251011494
"181" 20.0719225485817
"182" 23.5878289308928
"183" 19.5958334476441
"184" 18.4867065327959
"185" 17.9387956095401
"186" 16.1294922525253
"187" 16.4356104407392
"188" 20.6875972649464
"189" 18.5306771500907
"190" 18.5433033384619
```

Kod

```
v <- 19
n <- v * 10
m <- v
s <- v / 10

# generate dataset
dataset <- rnorm(n, m, s)

# save data into txt file
write.table(dataset, "E:\\University\\31EM\\11\\data.txt")

# get histogram
hist(dataset, col="green")

# get frequency table
frequency_table <- table(dataset)
# get mean
mean_value <- mean(dataset)
# get median
median_value <- median(dataset)
# get mode
mode_value <- mode(dataset)
# get qunatiles
qunatiles <- quantile(dataset)
```

```
# write data into rtf file  
sink("E:\\University\\31EM\\11\\stats.rtf")
```

```
frequency_table  
mean_value  
median_value  
mode_value  
qunatiles
```

Гистограма

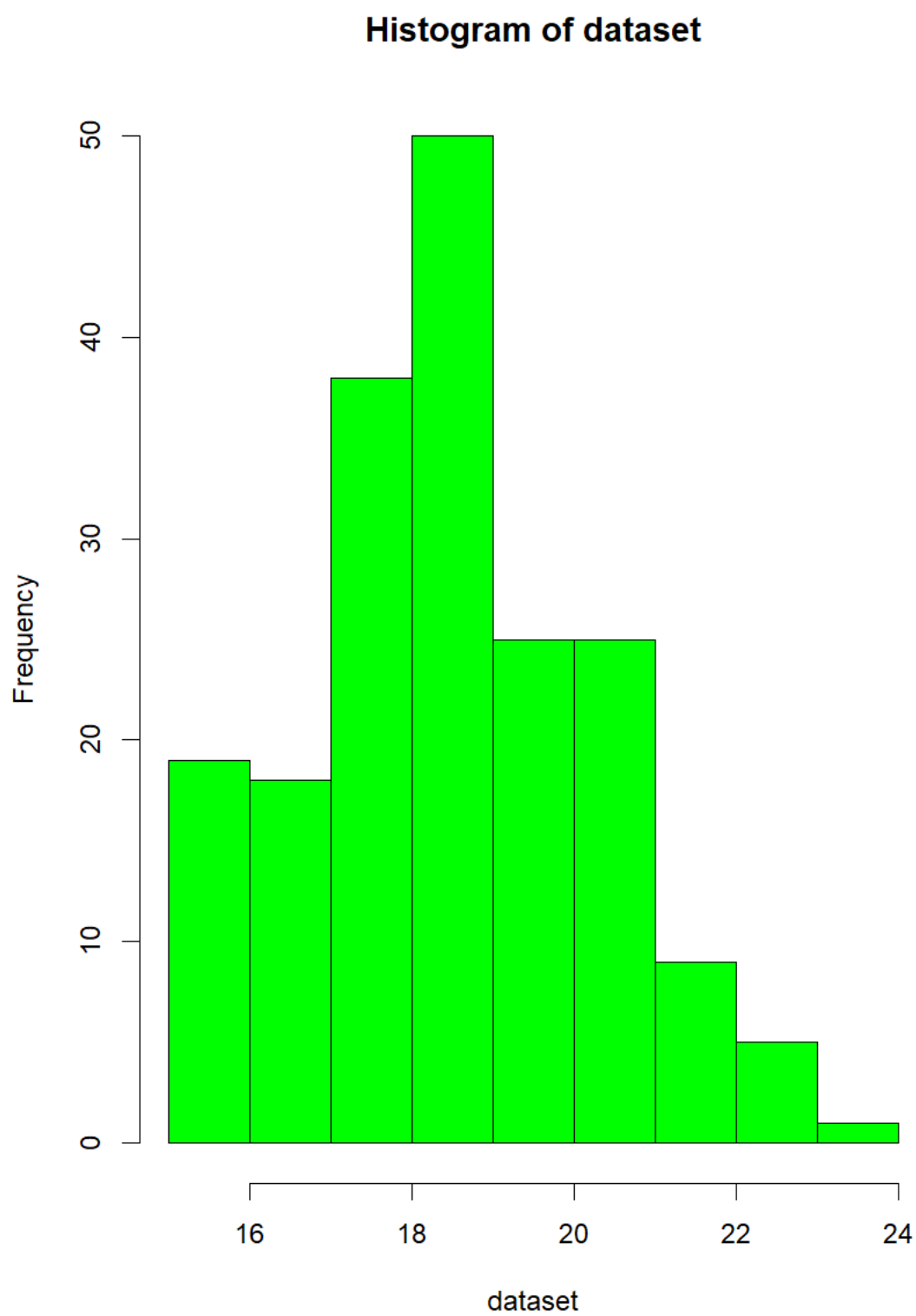


Рисунок 1.1 – Гістограма вибірки

Висновки

Аналіз проведено з використанням вбудованих методів.

Питання

Робота з даними у Statgraphics

Перше треба згенерувати дані функцією `Rnormal(12, 77, 3)`. 12 це кількість значень, 77 це математичне сподівання, 3 це дисперсія. Тоді треба обрати пункт Describe → Numeric Data → One Variable Analysis, обрати abular Options, вибрати перші 4 пункти опцій.

Призначення і використання різних типів вікон

Describe: статистичні методи аналізу: одна чи кілька змінних; процедури підбору розподілів, засоби табуляції даних.

Compare: засоби порівняння двох чи більше вибірок, процедури дисперсійного аналізу.

Relate: процедури простого або множинного або поліноміального аналізу дисперсій.

Quality Control: процедури аналізу та оцінки контролю якості.

Загальні й унікальні властивості Statgraphics

Пакет містить багато засобів статистичної обробки даних. Розрахований на фахівців, має більше двохсот п'ятидесяти процедур обробки даних для дисперсійного аналізу, описової статистики, контролю

якості, багатомірного аналізу, планування есперименту, лінійного програмування, тощо.

Що таке квантіль, мода, медіана?

Квантіль – рівний шматок даних.

Мода – найчастіший елемент.

Медіана – середній елемент.

Що таке ряди розподілу?

Згенерований набір даних.

Які характеристики розподілу ви знаєте?

Математичне сподівання – середнє значення.

Мода – найчастіший елемент.

Медіана – середній елемент.

Дисперсія – як далеко значення можуть відходити від медіани.

Розмах – різниця між найбільшим і найменшим значеннями.

Асиметрія – наскільки гістограма відхиляється від центру.

Ексцес – наскільки гістограма гостра.

Як будується гістограма?

По горизонталі показує значення ряду розподілу, по вертикалі як часто кожне з'являється. Тоді графічно показує як часто кожне значення з'являється відносно інших.

Що таке таблиця частот?

Показує як часто кожне значення з'являється.

На що впливає ширина інтервалу?

На те наскільки пласкою буде гістограма.