

Summary of Chapter 1



1 Concepts

- population – sample
- primary data – secondary data
- parameter – statistic
- descriptive statistics – inferential statistics
- nonprobability sample – probability sample
- numerical (quantitative) data – categorical (qualitative) data
- time series data – cross sectional data

2 Tools

- stem-and-leaf plot
- frequency distribution and histogram
- scatter diagram
- bar chart, pie chart, and pareto diagram

Concepts

Tools

Example

Home Page

Title Page

◀ ▶

◀ ▶

Page 1 of 8

Go Back

Full Screen

Close

Quit

2.1. Stem-and-leaf display

Suppose that the data are represented by x_1, x_2, \dots, x_n and that each number x_i consists of at least two digits.

To construct a stem-and-leaf plot, each number x_i are divided into two parts:

- (1) a stem, consisting of one or more of the leading digits;
- (2) a leaf, consisting of the remaining digits.

The stem-and-leaf display allows us to quickly determine some important features of the data that are not obvious from the data:

- (1) shape;
- (2) central tendency or middle of the data;
- (3) spread or variability.

Concepts

Tools

Example

Home Page

Title Page

◀◀ ▶▶

◀ ▶

Page 2 of 8

Go Back

Full Screen

Close

Quit

2.2. Histogram

The histogram is a graph of the observed frequencies versus the ring diameter. It represents a visual display of the data in which one may more easily see three properties:

- (1) Shape;
- (2) Location, or central tendency;
- (3) spread or variability.

Several guidelines are helpful in constructing histograms. When the data are numerous, grouping them into bins or cells is very useful. Generally,

1. Use between 4 and 20 bins — often choosing the number of bins approximately equal to the square root of the sample size works well.
2. Make the bins of uniform width.
3. Start the lower limit for the first bin just slightly below the smallest data value.

Note: Grouping the data into bins condenses the original data, so some detail is lost. Thus, when the number of observations is relatively small, or when the observations only take a few values, the histogram may be constructed from a frequency distribution of ungrouped data. Alternatively, a stem-and-leaf display could be used.



Concepts

Tools

Example

Home Page

Title Page



Page 3 of 8

Go Back

Full Screen

Close

Quit

3 Example

The data displayed here represent the electricity cost during the month of July 2003 for a random sample of 50 one-bedroom apartments in a large city.

Raw Data on Utility Charges (\$)

96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

- Place the data into an ordered array.
- Construct a stem-and-leaf display for these data.
- Plot a percentage histogram.

Solution using R:

```
Utility<- c(96,171,202,178,147,102,153,197,127,82, 157,185,90,116,  
172,111,148,213,130,165,141,149,206,175,123,128,144,168,109,167,  
95,163,150,154,130,143,187,166,139,149,108,119,183,151,114,135,  
191,137,129,158) —— save the data in an array
```

```
SortUtility<- sort(Utility) —— sort the array into an ordered array
```

```
SortUtility —— list the ordered array
```

```
stem(Utility) —— function ‘stem()’ create a stem-and-leaf display
```

```
hist(Utility) —— function ‘hist()’ create a histogram
```

Concepts

Tools

Example

Home Page

Title Page

◀◀ ▶▶

◀ ▶

Page 5 of 8

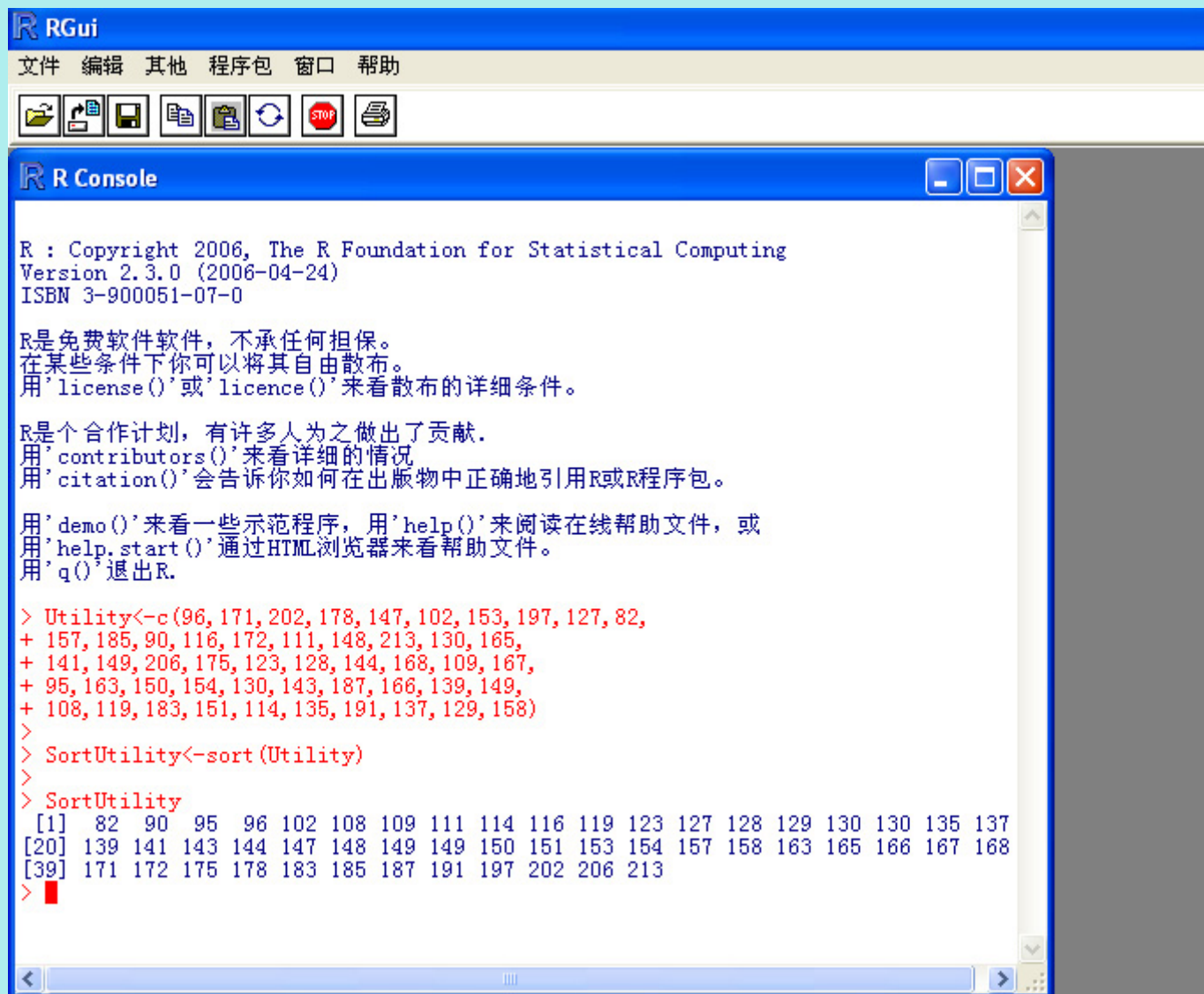
Go Back

Full Screen

Close

Quit

Outputs:



The screenshot shows the RGui application window. The title bar is 'R RGui'. The menu bar includes '文件', '编辑', '其他', '程序包', '窗口', and '帮助'. The toolbar contains icons for file operations and execution. The R Console window is open, displaying the following text:

```
R : Copyright 2006, The R Foundation for Statistical Computing
Version 2.3.0 (2006-04-24)
ISBN 3-900051-07-0

R是免费软件软件，不承担任何担保。
在某些条件下你可以将其自由散布。
用'license()'或'licence()'来看散布的详细条件。

R是个合作计划，有许多人为之做出了贡献。
用'contributors()'来看详细的情况
用'citation()'会告诉你如何在出版物中正确地引用R或R程序包。

用'demo()'来看一些示范程序，用'help()'来阅读在线帮助文件，或
用'help.start()'通过HTML浏览器来看帮助文件。
用'q()'退出R.

> Utility<-c(96,171,202,178,147,102,153,197,127,82,
+ 157,185,90,116,172,111,148,213,130,165,
+ 141,149,206,175,123,128,144,168,109,167,
+ 95,163,150,154,130,143,187,166,139,149,
+ 108,119,183,151,114,135,191,137,129,158)
>
> SortUtility<-sort(Utility)
>
> SortUtility
[1] 82 90 95 96 102 108 109 111 114 116 119 123 127 128 129 130 130 135 137
[20] 139 141 143 144 147 148 149 149 150 151 153 154 157 158 163 165 166 167 168
[39] 171 172 175 178 183 185 187 191 197 202 206 213
>
```



Concepts

Tools

Example

Home Page

Title Page

◀ ▶

◀ ▶

Page 6 of 8

Go Back

Full Screen

Close

Quit

```
> stem(Utility)
```

The decimal point is 1 digit(s) to the right of the |

8		2056
10		2891469
12		378900579
14		1347899013478
16		356781258
18		35717
20		263

```
>  
> █
```

Home Page

Title Page

◀◀ ▶▶

◀ ▶

Page 7 of 8

Go Back

Full Screen

Close

Quit



R : Copyright 2006, The R Foundation for
Version 2.3.0 (2006-04-24)
ISBN 3-900051-07-0

R是免费软件软件，不承担任何担保。
在某些条件下你可以将其自由散布。
用'license()'或'licence()'来看散布的详细

R是个合作计划，有许多人为之做出了贡献。
用'contributors()'来看详细的情况
用'citation()'会告诉你如何在出版物中正确

用'demo()'来看一些示范程序，用'help()'来
用'help.start()'通过HTML浏览器来看帮助文
用'q()'退出R.

```
> Utility<-c(96,171,202,178,147,102,153,  
+ 157,185,90,116,172,111,148,213,130,165,  
+ 141,149,206,175,123,128,144,168,109,16  
+ 95,163,150,154,130,143,187,166,139,149,  
+ 108,119,183,151,114,135,191,137,129,15  
+  
> hist(Utility)  
> █
```

Histogram of Utility

