Summary of Chapter 2

1 Measures

- mean, median, mode, geometric mean
- quartiles
- range, interquartile range, variance, standard deviation, coefficient of variation
- skewness, kurtosis
- covariance, correlation coefficient

2 Tools

- box-and-whisker plot
- shape of distribution (symmetric, skewed, peaked)
- empirical rules
- scatter diagram

3 Example

A bank branch located in a commercial district of a city has developed an improved process for serving customers during the 12:00 P.M. to 1:00 P.M. peak lunch period. The waiting time in minutes (defined as the time the customer enters the line to when he or she is served) of all customers during this hour is recorded over a period of one week. A random sample of 15 customers is selected, and the results are as follows:

4.21	5.55	3.02	5.13	2.34	3.54	3.20	4.50
0.38	5.12	6.46	6.19	4.77	6.00	6.10	3.79

a. Compute the arithmetic mean, median, first quartile, and the third quartile.

b. Compute the variance, standard deviation, range, interquartile range, and coefficient of variation.

c. Form the box-and-whisker plot, and describe the shape of the distribution for the waiting time.

1. Solution using formulae:

Mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 4.394$

To calculate following measures, place the data into an ordered array $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$:

0.38 2.34 3.02 3.20 3.54 3.79 4.21 4.50 4.77 5.12 5.13 5.55 6.00 6.10 6.19 6.46

Median:

Step 1: calculate the position: $q_2 = (n+1)/2 = (16+1)/2 = 8.5$

Step 2: find the median: $Q_2 = (X_{(8)} + X_{(9)})/2 = (4.50 + 4.77)/2 = 4.635.$

The first quartile and the third quartile:

Step 1: calculate the positions:

$q_1 = (n+1)/4 = 17/4 = 4.25, \qquad q_3 = 3(n+1)/4 = 12.75$

Step 2: find Q_1 and Q_3:

$Q_1 = X_{(4)} = 3.20, \qquad Q_3 = X_{(13)} = 6.00.$

Variance:

$$\begin{aligned} S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1) \\ &= \frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2) \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}(\sum_{i=1}^{n} X_i)^2\right] = 2.6876, \end{aligned}$$

Standard deviation $S = \sqrt{S^2} = 1.639,$

Range = $X_{\text{max}} - X_{\text{min}} = 6.46 - 0.38 = 6.08,$

Interquartile range IQR = $Q_3 - Q_1$ = 6.00 - 3.20 = 2.80,

Coefficient of variation CV = $\left(\frac{S}{\bar{X}}\right) \times 100\% = \left(\frac{1.639}{4.394}\right) \times 100\% = 37.30\%.$

As Mean $<$ Median, the distribution of the waiting time is skewed to the left.

2. Solution using ‘PHStat2’:

- Step 1: Open ‘Bank1.xls’ in **PHStat2** (see Figure 1).
- Step 2: Select **PHStat** / **Descriptive Statistics** / **Box-and-Whisker Plot** (see Figure 2).
- Step 3: In the Box-and-Whisker dialog box: (see Figure 3)
  - Enter **A1:A17** in the **Raw Data Cell Range** edit box.
  - Select the **Single Group Variable** option button.
  - Enter a title in the **Title** edit box.
  - Select **Five-Number Summary**.
  - Click the **OK** button.

Outputs:

(1) Five-Number Summary (see Figure 4);

(2) Box-and-Whisker plot (see Figure 5).

(3) From the plot we can see that the distribution is left-skewed.

Find mean, variance, standard deviation, and range of the data:

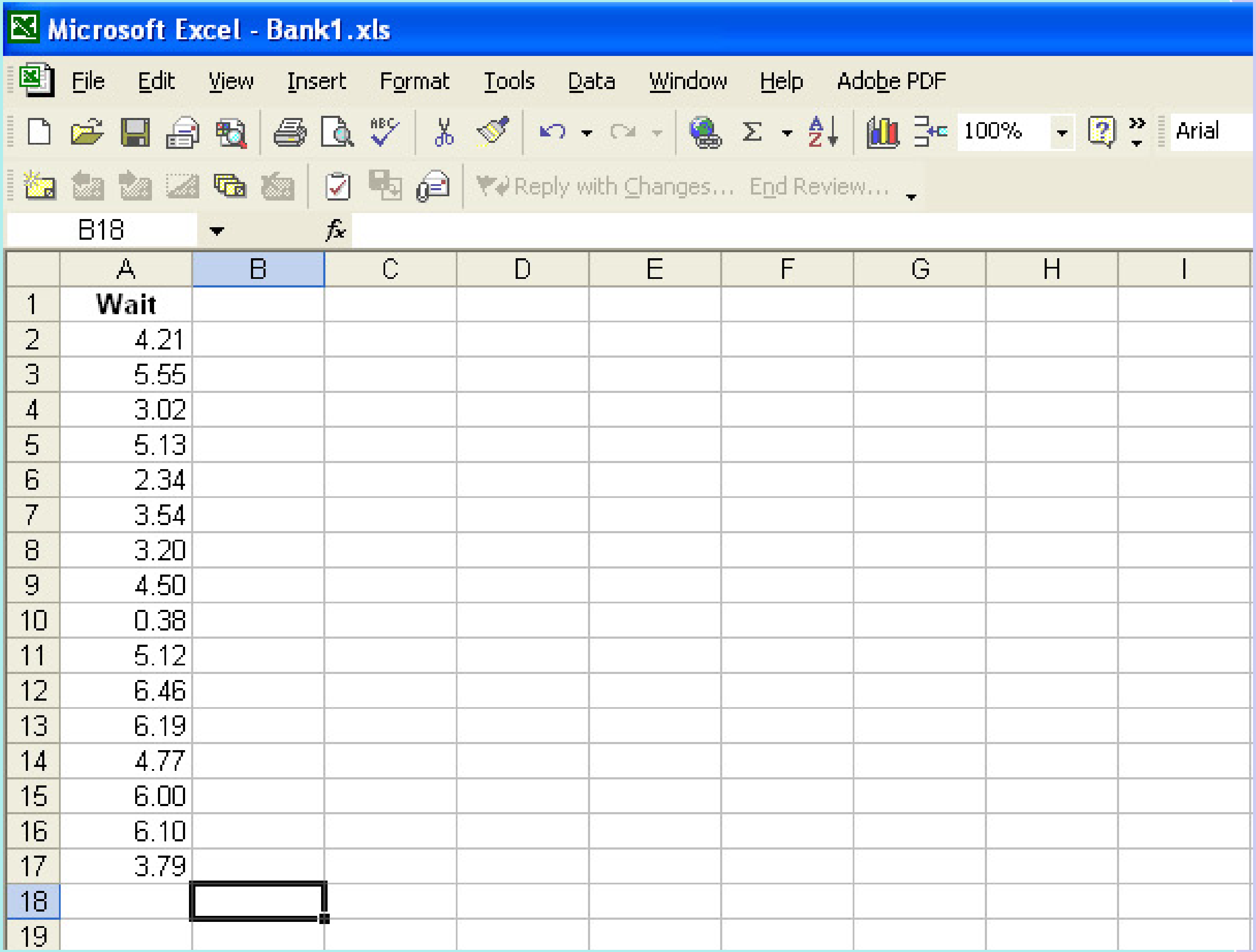
- Select **Tools** / **Data Analysis** / **Descriptive Statistics** (see Figures 6 and 7).
- In the Descriptive Statistics dialog box: (see Figure 8)
  - Enter **A1:A17** in the **Input Range** edit box.
  - Select the **Labels in First Row**.
  - Select **Summary Statistics**.
  - Click the **OK** button.

Outputs:

Minimum	0.380	Range	6.080
Q_1	3.200	Interquartile Range	2.800
Mean	4.394	Std.	1.639
Median	4.635	Variance	2.686
Q_3	6.000	CV	37.30%
Maximum	6.460		

where

$$CV = \left(\frac{S}{\bar{X}}\right)100\% = \frac{1.639}{4.395} \times 100\% = 37.29\%.$$



	A	B	C	D	E	F	G	H	I
1	**Wait**								
2	4.21								
3	5.55								
4	3.02								
5	5.13								
6	2.34								
7	3.54								
8	3.20								
9	4.50								
10	0.38								
11	5.12								
12	6.46								
13	6.19								
14	4.77								
15	6.00								
16	6.10								
17	3.79								
18									
19									

Figure 1.

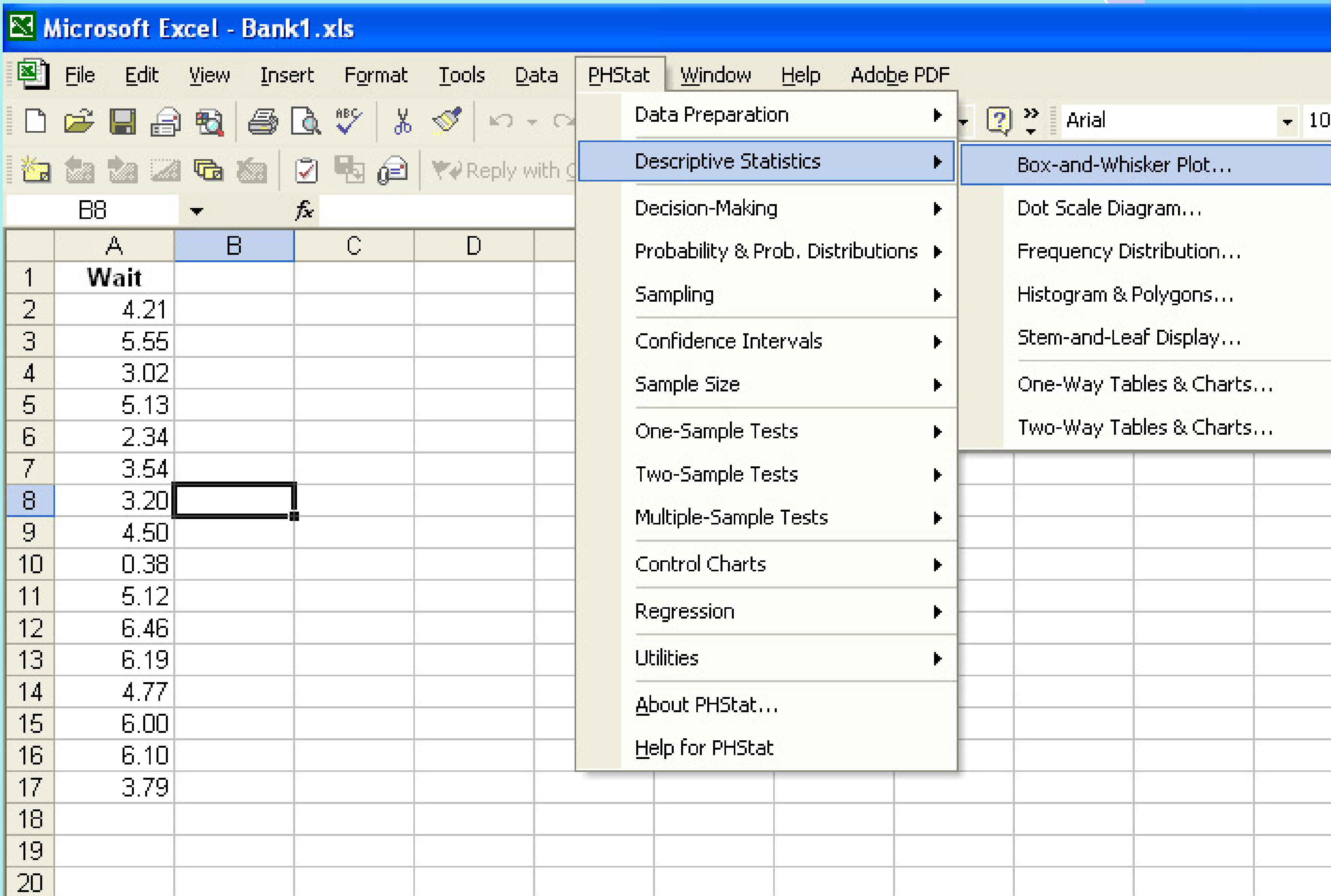

Microsoft Excel - Bank1.xls
File
Edit
View
Insert
Format
Tools
Data
PHStat
Window
Help
Adobe PDF
Reply with
Arial
10
B8
fx
A
B
C
D
Wait
4.21
5.55
3.02
5.13
2.34
3.54
3.20
4.50
0.38
5.12
6.46
6.19
4.77
6.00
6.10
3.79
Data Preparation
Descriptive Statistics
Decision-Making
Probability & Prob. Distributions
Sampling
Confidence Intervals
Sample Size
One-Sample Tests
Two-Sample Tests
Multiple-Sample Tests
Control Charts
Regression
Utilities
About PHStat...
Help for PHStat
Box-and-Whisker Plot...
Dot Scale Diagram...
Frequency Distribution...
Histogram & Polygons...
Stem-and-Leaf Display...
One-Way Tables & Charts...
Two-Way Tables & Charts...

Figure 2.

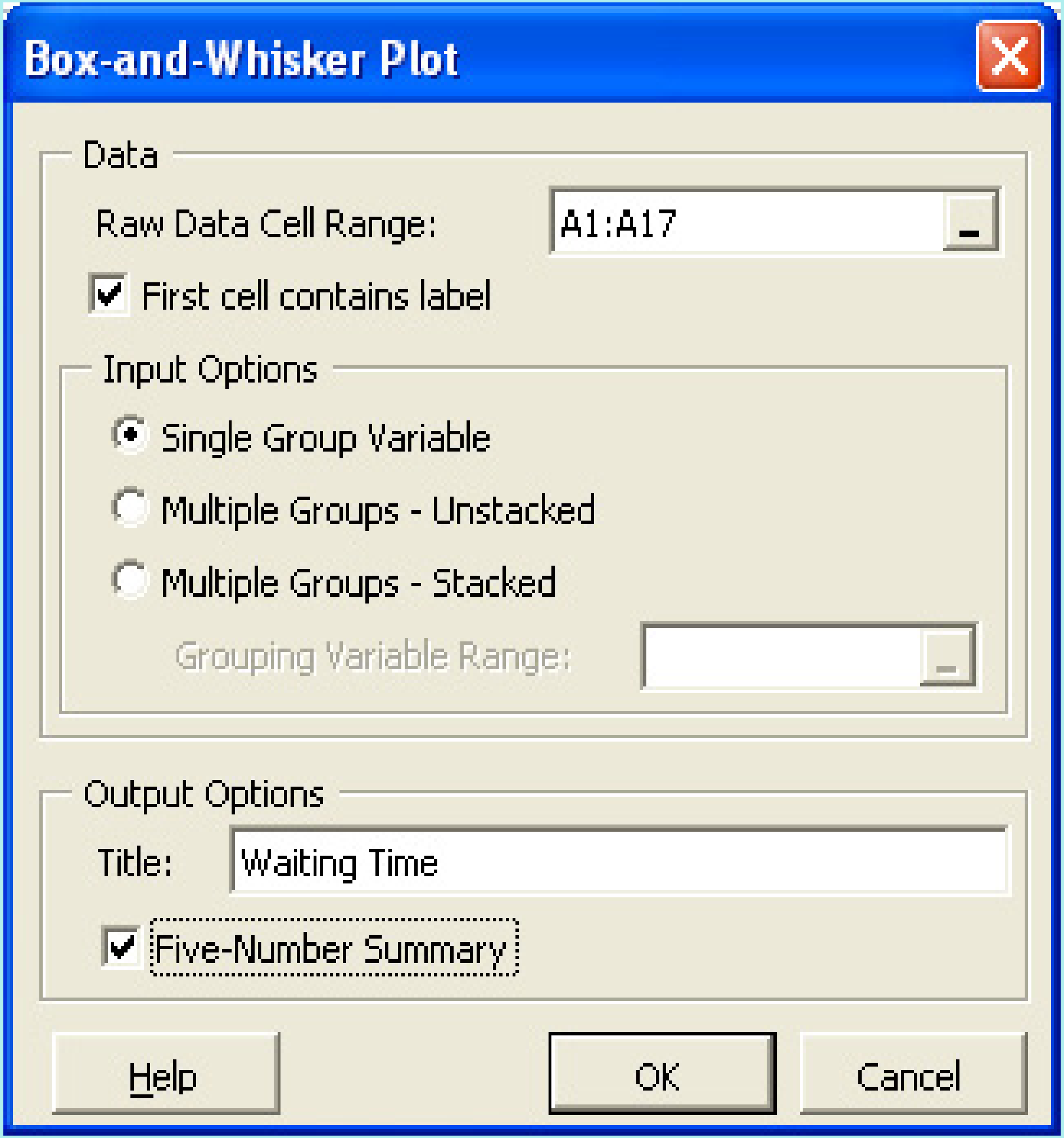
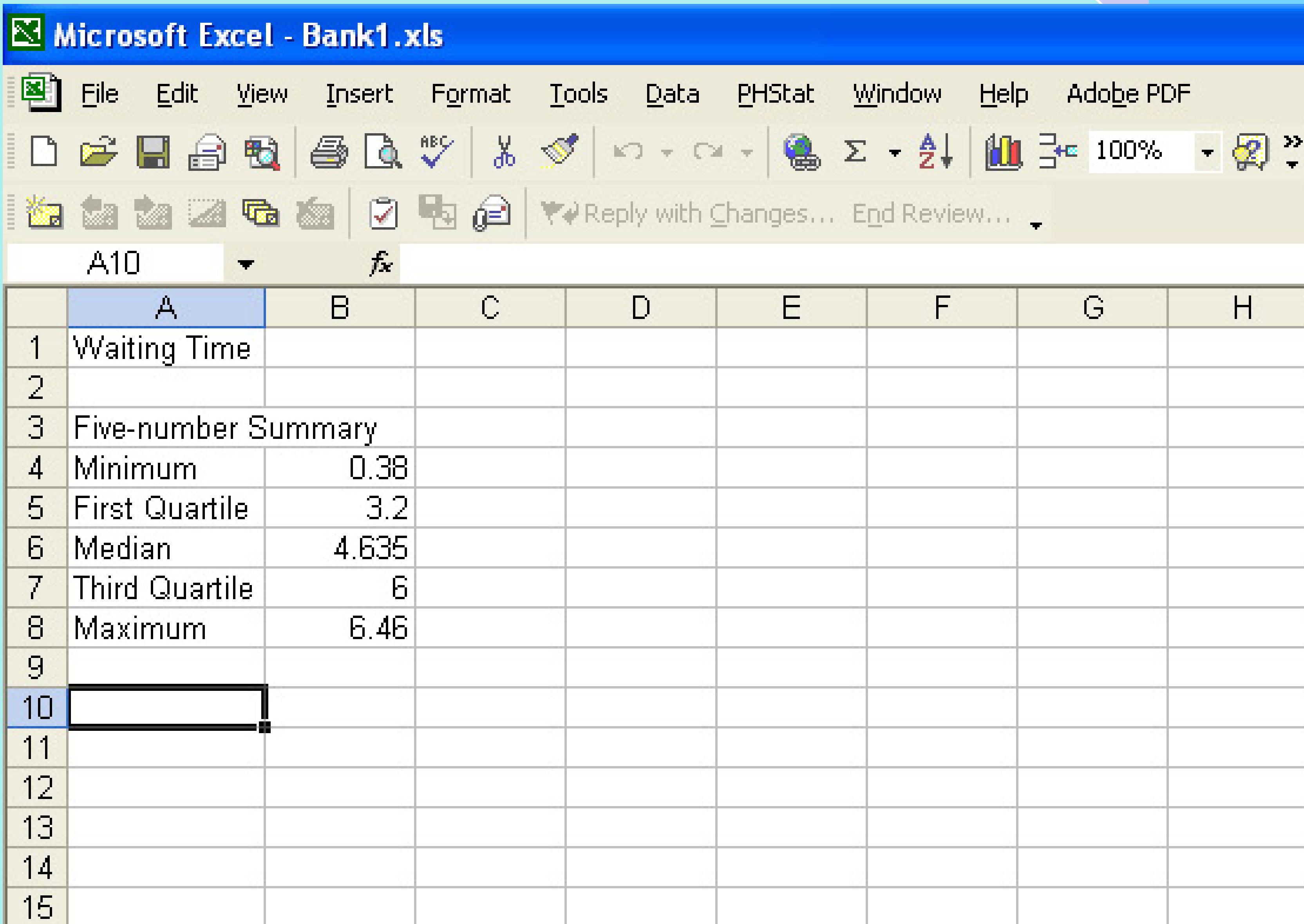

Box-and-Whisker Plot
Data
Raw Data Cell Range:
A1:A17
First cell contains label
Input Options
Single Group Variable
Multiple Groups - Unstacked
Multiple Groups - Stacked
Grouping Variable Range:
Output Options
Title:
Waiting Time
Five-Number Summary
Help
OK
Cancel

Figure 3.



Waiting Time	
Five-number Summary	
Minimum	0.38
First Quartile	3.2
Median	4.635
Third Quartile	6
Maximum	6.46

Figure 4

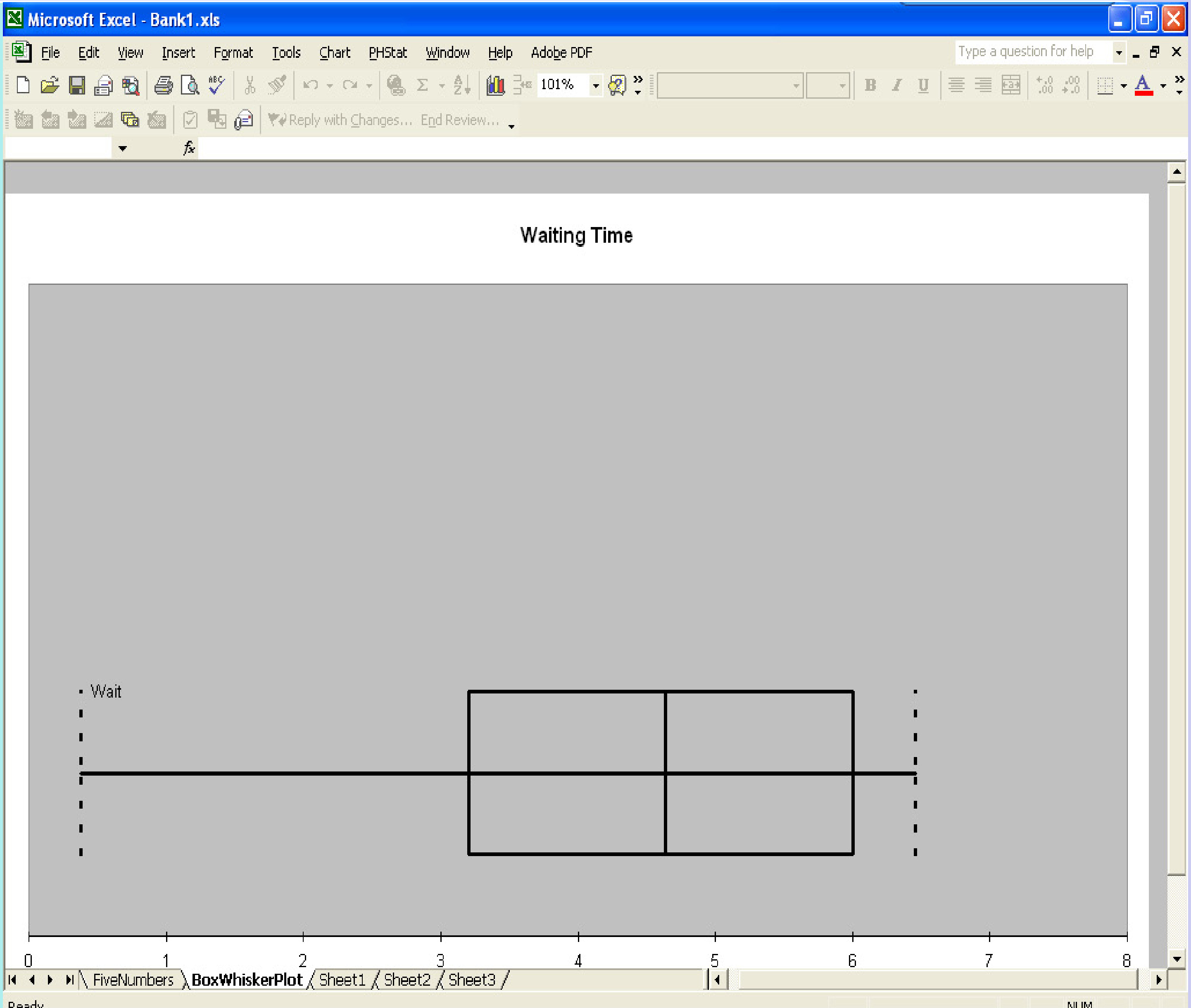
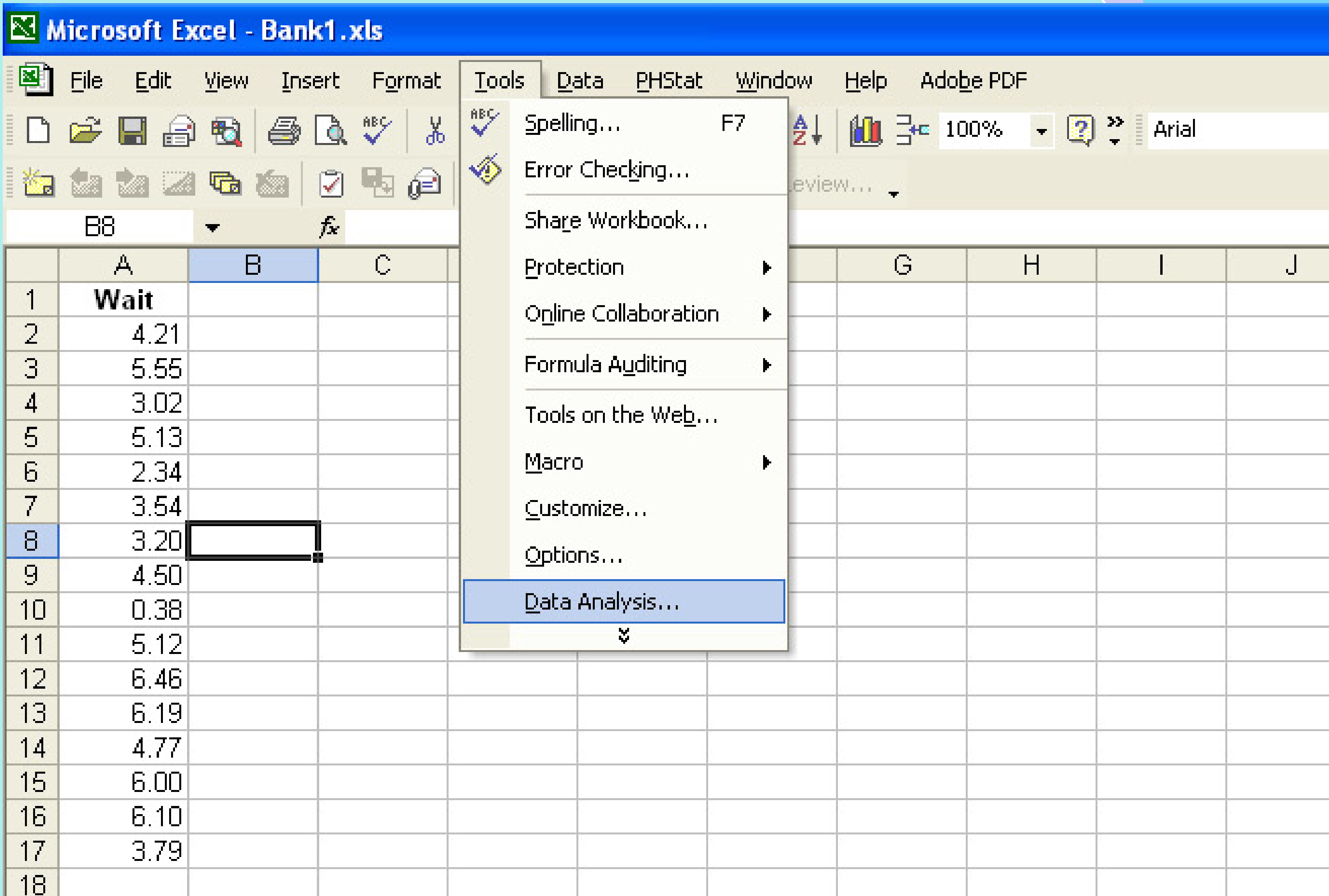

Microsoft Excel - Bank1.xls
Waiting Time
Wait
0
1
2
3
4
5
6
7
8
FiveNumbers
BoxWhiskerPlot
Sheet1
Sheet2
Sheet3
Ready
NUM

Figure 5.



Microsoft Excel - Bank1.xls
File Edit View Insert Format Tools Data PHStat Window Help Adobe PDF
100%
Arial
B8
Spelling... F7
Error Checking...
Share Workbook...
Protection
Online Collaboration
Formula Auditing
Tools on the Web...
Macro
Customize...
Options...
Data Analysis...
A B C G H I J
1 Wait
2 4.21
3 5.55
4 3.02
5 5.13
6 2.34
7 3.54
8 3.20
9 4.50
10 0.38
11 5.12
12 6.46
13 6.19
14 4.77
15 6.00
16 6.10
17 3.79
18

Figure 6.

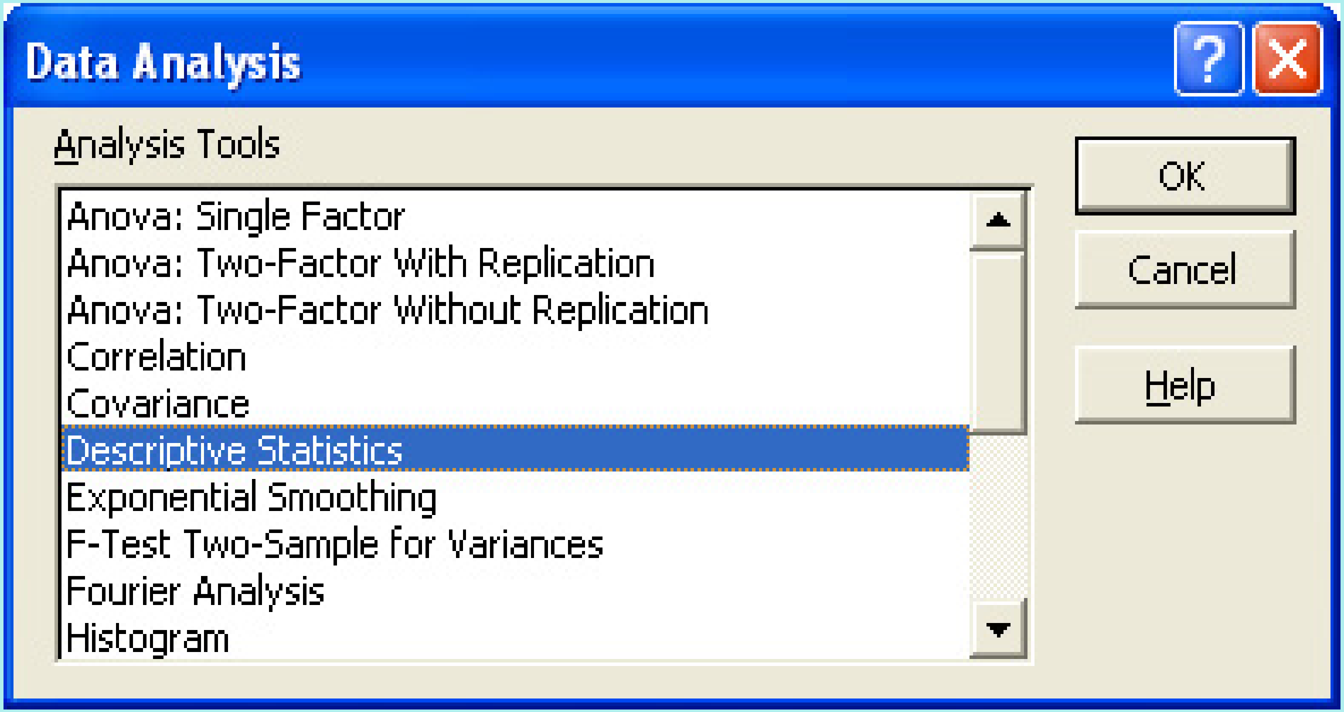

Data Analysis
Analysis Tools
Anova: Single Factor
Anova: Two-Factor With Replication
Anova: Two-Factor Without Replication
Correlation
Covariance
Descriptive Statistics
Exponential Smoothing
F-Test Two-Sample for Variances
Fourier Analysis
Histogram
OK
Cancel
Help

Figure 7.

Microsoft Excel - Bank1.xls

	A	B
1	*Wait*	
2		
3	Mean	4.39375
4	Standard Error	0.409847406
5	Median	4.635
6	Mode	#N/A
7	Standard Deviation	1.639389622
8	Sample Variance	2.687598333
9	Kurtosis	0.839455678
10	Skewness	-0.905354269
11	Range	6.08
12	Minimum	0.38
13	Maximum	6.46
14	Sum	70.3
15	Count	16

Figure 8.

3. Solution using R:

```
> WT<-c(4.21,5.55,3.02,5.13,2.34,3.54,3.20,
+ 4.50,0.38,5.12,6.46,6.19,4.77,6.00,6.10,3.79)

> summary(WT)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.380   3.455   4.635   4.394   5.662   6.460

> var(WT)
[1] 2.687598

> sd(WT)
[1] 1.639390

> boxplot(WT, horizontal=TRUE)
```

According to above results,

Range = Max - Min = 6.460 - 0.380 = 6.080

IQR = $Q_3 - Q_1$ = 5.662 - 3.455 = 2.207

CV = $\left(\frac{S}{\bar{X}}\right) \times 100\% = (1.639/4.394) \times 100\% = 37.30\%.$

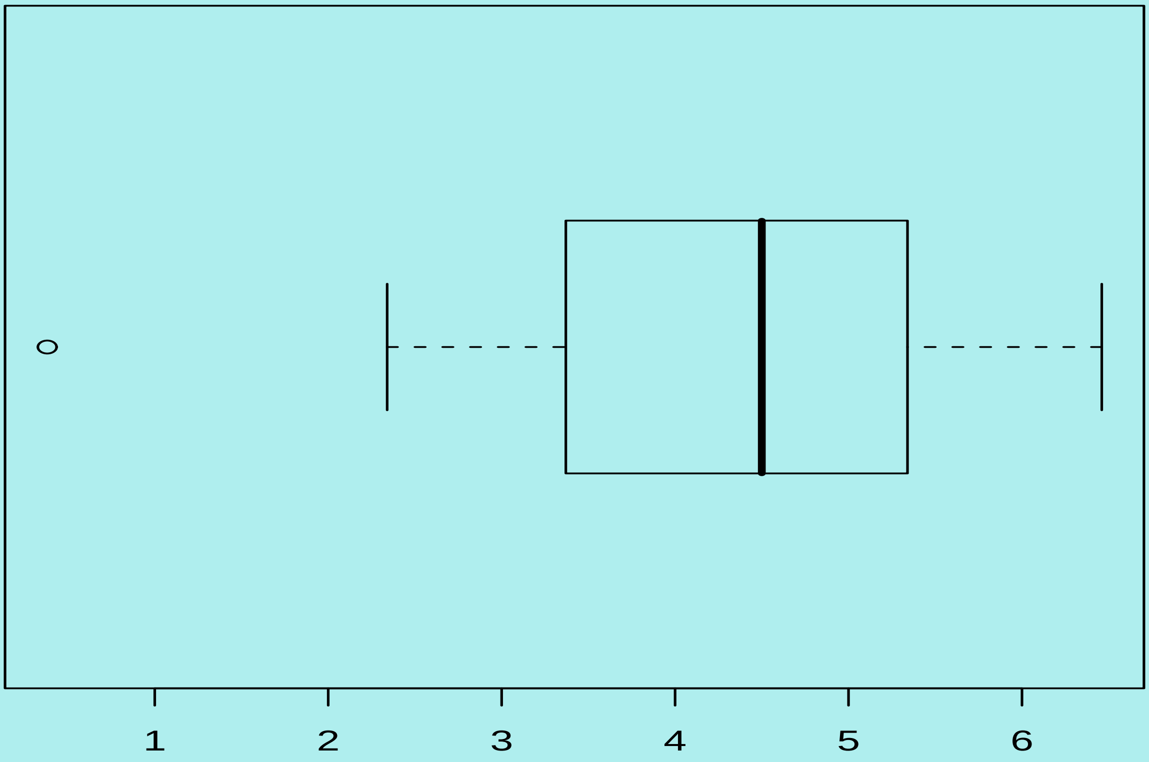


Figure 9 From the plot, we can see that the distribution is left-skewed.