2nd Material Subject: Descriptive Statistics

Undergraduate of Telecommunication Engineering

MUH1F3 - PROBABILITY AND STATISTICS

Telkom University
Center of eLearning & Open Education Telkom University
Jl. Telekomunikasi No.1, Bandung - Indonesia
http://www.telkomuniversity.ac.id

Lecturer: Nor Kumalasari Caecar Pratiwi, S.T., M.T. (caecarnkcp@telkomuniversity.ac.id)

السلام عليكم ورحمة الله وبركاته

WELCOME

TABLE OF CONTENTS:

LEARNING OBJECTIVES:

After careful study of this chapter, student should be able to do the following:

1. **Construct and interpret visual data displays, including the stem-and-leaf display, and the box plot**
2. **Explain how to use box plots and other data displays to visually compare two or more samples of data**

STEM LEAF DIAGRAM

A stem-and-leaf diagram is a good way to obtain an informative visual display of a data-set $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \cdots, \mathbf{x_n}$ where each number $\mathbf{x_i}$ consists of at least two digits. To construct a stem-and leaf diagram, use the following steps.

1. **Divide each number x_i into two parts: a stem, consisting of one or more of the leading digits, and a leaf, consisting of the remaining digit.**
2. **List the stem values in a vertical column.**
3. **Record the leaf for each observation beside its stem.**
4. **Write the units for stems and leaves on the display.**

STEM LEAF DIAGRAM

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Figure 1: Waiting Time Data (in μ**s**) of 80 Users of Internet Service Users from PT.Cyberjaya Providers

STEM LEAF DIAGRAM

To illustrate the construction of a stem and leaf diagram in Figure 1, we will select as stem values the numbers **7**, **8**, **9**, $\cdots$, **24**. The resulting stem-and-leaf diagram is presented in Figure 2. The last column in the diagram is a frequency count of the number of leaves associated with each stem. Furthermore, the waiting time are distributed approximately symmetrically about the central value. The stem-and-leaf diagram enables us to determine quickly some important features of the data that were not immediately obvious in the original display in Figure 1.

STEM LEAF DIAGRAM

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Stem: Tens and hundreds digits (psi); Leaf: Ones digits (psi).

Figure 2: The Stem and Leaf for Waiting Time Data (in μ**s**) of 80 Users PT.Cyberjaya

STEM LEAF DIAGRAM

Stem	Leaf
6	1 3 4 5 5 6
7	0 1 1 3 5 7 8 8 9
8	1 3 4 4 7 8 8
9	2 3 5

(*a*)

Stem	Leaf
6L	1 3 4
6U	5 5 6
7L	0 1 1 3
7U	5 7 8 8 9
8L	1 3 4 4
8U	7 8 8
9L	2 3
9U	5

(*b*)

Stem	Leaf
6z	1
6t	3
6f	4 5 5
6s	6
6e	
7z	0 1 1
7t	3
7f	5
7s	7
7e	8 8 9
8z	1
8t	3
8f	4 4
8s	7
8e	8 8
9z	
9t	2 3
9f	5
9s	
9e	

(*c*)

Figure 3: Various Type of Stem and Leaf Diagram

STEM LEAF DIAGRAM

Figure 3 is the stem and leaf diagram for 25 observations on batch yields from a chemical process. In Figure 3(a), we have used 6, 7, 8, and 9 as the stems. This results in too few stems, and the stem-and-leaf diagram does not provide much information about the data. In Figure 3(b), we have divided each stem into two parts (Lower and Upper), resulting in a display that more adequately displays the data. Figure 3(c) illustrates a stem and leaf display with each stem divided into five (z, t, f, s, e) parts. There are too many stems in this plot, resulting in a display that does not tell us much about the shape of the data.

BOX WHISKER PLOT

The **Box Whisker Plot** is a graphical display that simultaneously describes several important features of a data set, such as center, spread, departure from symmetry, and identification of unusual observations or outliers. A box plot displays the three quartiles, the minimum, and the maximum of the data on a rectangular box, aligned either horizontally or vertically. The box encloses the interquartile range with the left (or lower) edge at the first quartile $\mathbf{Q_1}$, and the right (or upper) edge at the third quartile $\mathbf{Q_3}$.

$$\mathbf{Q_i} = \mathbf{Data}\frac{\mathbf{i(n+1)}}{\mathbf{4}} \tag{1}$$

$$\mathbf{IQR} = \mathbf{Q_3} - \mathbf{Q_1} \tag{2}$$

$$\mathbf{ONE\ STEP} = \mathbf{1.5 \cdot IQR} \tag{3}$$

$$\mathbf{TWO\ STEP} = \mathbf{3 \cdot IQR} \tag{4}$$

BOX WHISKER PLOT

$$\textbf{UPPER OUTER FENCES (UOF)} = \textbf{Q}_3 + \textbf{TWO STEP} \tag{5}$$

$$\textbf{UPPER INNER FENCES (UIF)} = \textbf{Q}_3 + \textbf{ONE STEP} \tag{6}$$

$$\textbf{LOWER OUTER FENCES (LOF)} = \textbf{Q}_1 - \textbf{TWO STEP} \tag{7}$$

$$\textbf{LOWER INNER FENCES (LIF)} = \textbf{Q}_1 - \textbf{ONE STEP} \tag{8}$$

A data is said to be a **Mild Outlier** if:

$$\textbf{LOF} < \textbf{X} \leq \textbf{LIF} \quad \text{or} \quad \textbf{UIF} \leq \textbf{X} < \textbf{UOF} \tag{9}$$

A data is said to be a **Extreme Outlier** if:

$$\textbf{X} \leq \textbf{LOF} \quad \text{or} \quad \textbf{X} \geq \textbf{UOF} \tag{10}$$

BOX WHISKER PLOT

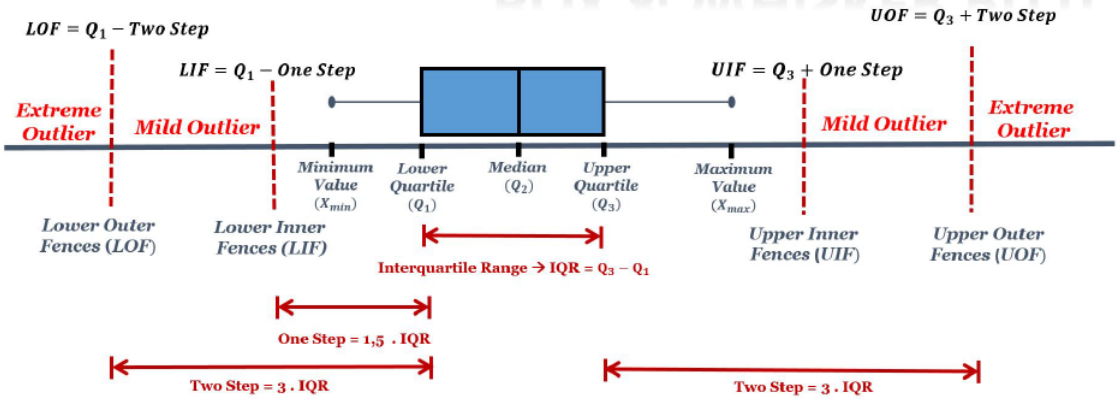

Figure 4: Description of a Box Whisker Plot

BOX WHISKER PLOT

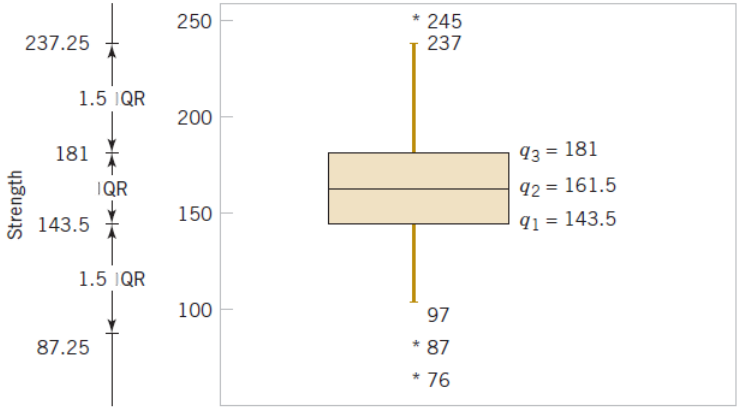

Figure 5: Box Whisker Plot for Waiting Time Data (in μ**s**) of 80 Users of PT.Cyberjaya Users

EXAMPLE

Following are the result of exams for 20 Linear Algebra students:

12	1	51	52	54	54	59	58	90	69
65	59	51	56	52	54	57	53	45	29

Construct and interpret visual data displays with stem-and-leaf display and the box whisker plot!

Answer:

- Step 1: Sorting data from the smallest to the biggest

1	12	29	45	51	51	52	52	53	54
54	54	56	57	58	59	59	65	69	90

- Step 2: Each $\mathbf{x_i}$ value is divided into two parts, the first digit becomes Stem and the second digit become Leaf.

Stem	Leaf	Freqeuncy Cumulative
0	1	1
1	2	2
2	9	3
3	-	3
4	5	4
5	1 1 2 2 3 4 4 4 6 7 8 9 9	17
6	5 9	19
7	-	19
8	-	19
9	0	20

Figure 6: Stem Leaf Diagram

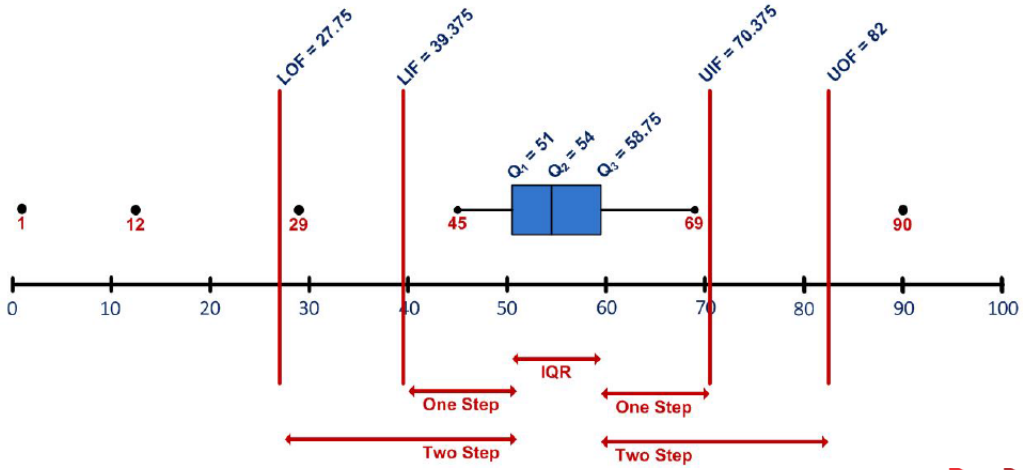

Figure 7: Box and Whisker Plot

From the Box and Whisker Plot, The data has outliers value,

- **Mild Outlier** is 29 (**3rdData**)
- **Extreme Outlier** are: 1, 12 and 90 (the **1st**, **2nd** and **20st** data).

Thank You