Chapter 2

Collection and Presentation of Data

Σ

Definition of Statistics

Statistics is the branch of science that deals with the **collection**, **organization**, **analysis**, **interpretation** and **presentation** of data.

Outline of Chapter 2

- Preliminaries
- Methods of Data Collection
  - Common Methods of Data Collection
  - The Questionnaire
- Sampling
  - Methods of Probability Sampling
  - Methods of Nonprobability Sampling
- **Tabular and Graphical Presentation**
- **The Frequency Distribution**
- **The Stem-and-Leaf Display**

Reference: Chap. 3-5 &12 of Elementary Statistics by ACS

Part 3

Organization and Presentation of Data

Presentation of Data

- We highlight the important information that we were able to extract

- With good data presentation, we can discover, explore and understand possible relationships.

- Poor data presentation will only mislead, deceive, and misinform.

- Always keep in mind your message/point.

Textual Presentation

Textual Presentation of data incorporates important figures in a paragraph of text.

The objective is to **focus the readers' attention to some information** that need particular emphasis as well as to **some important comparisons** and to supplement with a narrative account from a table or a chart.

Textual Presentation

It could also show the summary measures like minimum, maximum, totals and percentages.

We do not need to put all figures in a textual presentation; we just have to select the most important ones that we want to focus on.

When the data become voluminous, the textual presentation is strongly not advised because the presentation becomes almost incomprehensible.

Textual Presentation

Excerpt taken from the business section of the Philippine Star:

"The 30-company Philippine Stock Exchange Index finished down 10.21 points, or 0.5 percent, at 1,921.33, after falling 0.8 percent Tuesday following a seven-day rally that boosted the main index by 6.4 percent. Weighing on the index were losses incurred by Globe Telecom, down 2.4 percent at P830, Ayala Land, off 1.2 percent at P8, and Jollibee, lower by 3.4 percent at P28.50 on profit taking.
"Partly offsetting the market's losses were Ayala Corporation, up 1.4 percent at P7.30, and Bank of the Philippine Islands, which rose one percent to P50 on bargain hunting. Ayala unit Manila Water rose 1.7 percent to P6.10 after the water utility Tuesday posted a 57 percent year-on-year rise in first-quarter net profit.
"All sectoral indicators ended lower, except the oil sub index, which finished higher. Decliners led gainers 46 to 21, while 51 stocks were unchanged."

Note: Focus on important figures and not minute details of the data set. If you wish to present all the figures, use the tabular presentation instead.

Reading Assignment

1. On Tabular Presentation
Read pages 120-138
Guide Questions:
a. What are the different types of tables?
b. What are the different parts of a formal statistical table? What are the functions of each part?
c. What are the basic principles followed in the construction of a formal statistical table?
d. What are the different classifications of summarized data? What are the recommended arrangements of the row captions for each classification?

2. On Graphical Presentation
Read pages 138-163
Guide Questions:
a. What are the general qualities of a good chart?
b. What are the different types of chart?
c. What are the different parts of these charts?
d. What are the pointers that we need to remember in constructing a good chart?

Tabular Presentation

- **Tabular presentation** of data arranges figures in a systematic manner in rows and columns.

- It is the most common method of data presentation.

- Uses include description, comparison, showing relationships among variables of interest

- It can contain frequency distributions, proportions, percentages, totals, averages

- It is concise and usually requires minimal explanations

- Should be simple and easy to understand

Tabular Presentation

1. Leader Work

Leader work has the simplest layout among the three types of tables. It contains no table title or column headings and has no table borders. This table needs an introductory or descriptive statement so that the reader can understand the given figures.

The Population in the Philippines for the Census Years 1975 to 2000 is as follows [a]

1975	42,070,660
1980	48,098,460
1990	60,703,206[b]
1995	68,616,536[b]
2000	76,498,735

a National Statistics Office

b The 1990 and 1995 figures include the household population, homeless population, and Filipinos in Philippines embassies and mission abroad. In addition, the census comprise institutional population found living quarters such as penal institutions, orphanages, hospitals, military camps, etc.

Tabular Presentation

2. Text Tabulation

The format of text tabulation is a little bit more complex than leader work. It already has column headings and table borders so that it is easier to understand than leader work. However it still does not have table title and table number. Thus, it also requires an introductory statement so that the readers can comprehend the given figures. Similar to leader work, we can place additional explanatory statement in the footnote.

The distribution of cellular subscribers per telephone operator as of December 2003 is as follows:

Telephone Operator	No. of Subscribers
SMART	10,080,112
GLOBE TELECOM*	8,800,000
PILTEL	2,867,085
EXTELCOM	29,896
Total	22,509,560

*Islacom subscribers are included in the figures of Globe
Source: National Telecommunications Commission

Tabular Presentation

3. Formal Statistical Table

The **formal statistical table** is the most complete type of table since it has all the different and essential parts of a table like table number, table title, head note, box head, stub head, column headings, and so on.

It **can be a stand-alone table since it does not need any accompanying texts** and it could be easily understood on its own.

Table 4.1.
The Parts of a Formal Statistical Table

panel   TABLE NO. – Title of the Table   heading
(head note)

Stubhead	Spanner head			Spanner Head			Total
	Column Head	Column Head	Column Head	Column Head	Column Head	Column Head	
CENTER HEAD				Cell			Cell
Total line caption							
Row caption	Cell	Cell	Cell	Cell	Cell	Cell	Cell ◄ Line
Row caption				Cell			
Row caption				Cell			
Row caption				Cell			
CENTER HEAD				Cell			
Total line caption							
Row caption				Cell			
Row caption				Cell			
Row caption				Cell			
Row caption				Cell			

b l o c k (left side)

field (right side)

¹Footnote
Source note:

Example 4.4: We now show the parts of an actual formal statistical table.

table no.

heading → Table 4.2. Total Number of Families, Total and Average Annual Family Income and Expenditure by Income Class and Urban/Rural Classification: 2000 ← title
(in thousand pesos) ····> head note
spanner head

stub head ···
column head ↓

Income Class	Total no. of Families	Income		Expenditure	
		Total	Average	Total	Average
				column heads	
URBAN center head					
All income classes	7,489,853	1,535,250,064	205	1,234,285,343	165
Under ₱10,000	6,840	56,801	8	59,693	9
10,000 - 19,999	50,326	812,883	16	908,440	18
20,000 - 29,999	126,398	3,209,473	25	3,389,398	27
30,000 - 39,999	206,629	7,286,107	35	7,620,222	37
40,000 - 49,999	309,445	13,998,124	45	14,125,894	46
50,000 - 59,999	356,091	19,611,853	55	19,066,504	54
60,000 - 79,999	742,181	51,834,409	70	49,559,978	67
80,000 - 99,999	761,548	68,497,516	90	63,900,511	84
100,000 -149,999	1,522,669	187,488,838	123	168,737,118	111
150,000 -249,999	1,718,997	331,582,321	193	278,565,606	162
250,000 -499,999	1,236,285	419,348,347	339	326,570,009	264
500,000 and over	452,445	431,577,392	954	301,781,968	667
RURAL center head					
All income classes	7,779,802	664,181,810	85	567,561,086	73
Under ₱10,000	28,716	228,558	8	274,307	10
10,000 - 19,999	278,686	4,464,313	16	4,938,941	18
20,000 - 29,999	710,253	18,069,507	25	19,345,217	27
30,000 - 39,999	963,913	33,850,702	35	35,173,323	36
40,000 - 49,999	1,079,063	48,440,786	45	49,071,045	45
50,000 - 59,999	840,034	46,175,765	55	45,048,374	54
60,000 - 79,999	1,241,038	85,641,067	69	80,449,749	65
80,000 - 99,999	734,733	65,568,120	89	58,302,544	79
100,000 -149,999	908,391	109,295,569	120	93,958,409	103
150,000 -249,999	663,196	125,644,796	189	97,837,987	148
250,000 -499,999	292,148	96,124,602	329	66,663,544	228
500,000 and over	39,631	30,678,324	774	16,497,643	416

Note: Details may not add up to totals due to rounding. ·····> footnote
Source: National Statistics Office. ···> source note

Tabular Presentation

3. Formal Statistical Table

Heading consists of the table number, title and head note. It is located on top of the table of figures.
i. **Table number** is the number that identifies the position of the table in a sequence.
ii. **Table title** states in telegraphic form of the subject, data classification, and place and period covered by the figures in the table.
iii. **Head** note appears below the title but above the top cross rule of the table and provides additional information about the table.

Box head consists of spanner heads and columns heads.
i. **Spanner** head is a caption or label describing two or more column heads.
ii. **Column** head is a label that describes the figures in a column.
iii. **Panel** is a set of column heads under the same spanner head.

Stub consists of row captions, center head, and stub head. It is located at the left side of the table.
i. **Row caption** is a label that describes the figures in a row.
ii. **Center head** is a label describing a set of row captions.
iii. **Stub head** is a caption or label that describes all of the center heads and row captions. It is located on the first row.
iv. **Block** is a set of row captions under the same center head.

Graphical Presentation

- The graph or statistical chart is a very powerful tool in presenting data.

- It is an important medium of communication because we can create a pictorial representation of the numerical figures found in tables without showing too many figures.

- We construct graphs not only for presentation purposes but also as an **initial step in analysis**.

- The graph, as a tool for analysis, can exhibit possible associations among the variables and can facilitate the comparison of different groups.

- It can also reveal trends over time.

Graphical Presentation

1. Line Chart
2. Column Chart
3. Horizontal Bar Chart
4. Pie Chart
5. Pictograph
6. Statistical Maps

Next meeting, we will critique and improve some tables and graphs.

Organization of Data

Definition 5.1.
Raw data are data in their original form.

Definition 5.2.

The **array** is an ordered arrangement of data according to magnitude.

Note:

• Raw data have not been organized in any manner and observations are recorded in the order observed.

• An array may be in ascending or descending order.

• An array is also called sorted data or ordered data.

Example of Raw Data

Table C.1. Final Grades of a Sample of Statistics 101 Students

(page 613)

Table C.1 Final Grades of a Sample of Stat 101 Students					
79	76	62	92	61	64
85	88	85	87	86	73
58	78	77	92	77	80
68	72	60	85	65	78
76	81	61	94	40	84
68	89	60	85	54	77
73	86	79	88	58	74
73	70	43	97	64	56
80	77	82	91	65	75
64	85	60	94	62	53
79	76	58	92	73	66
71	81	75	97	70	49
68	85	82	65	82	45
81	63	85	80	59	63
86	53	71	72	51	77
62	90	80	86	65	51
71	82	52	95	87	73
63	80	83	90	73	51
69	75	88	95	82	44
61	45	56	81	88	76
50	78	80	69	57	58
67	71	51	97	78	59
59	74	63	83	70	70
78	86	63	85	81	80
81	80	70	45	74	77
64	73	80	67	70	77
86	78	87	97	81	73
71	58	94	61	70	71
76	65	69	68	56	69
73	88	85	96	64	65
71	84	93	63	82	64
84	67	65	72	78	80
85	70	80	76	79	71
76	78	87	87	83	72
87	45	69	56	79	54

Example of an Array

Table 5.1. Array of Final Grades (page 165)

Table 5.1 Array of Final Grades of a Sample of Statistics 101 Students

50	78	80	69	57	58
58	78	77	92	77	80
59	74	63	83	70	70
61	45	56	81	88	76
62	90	80	86	65	51
63	80	83	90	73	51
64	85	60	94	62	53
64	73	80	67	70	77
67	71	51	97	78	59
68	72	60	85	65	78
68	89	60	85	54	77
68	85	82	65	82	45
69	75	88	95	82	44
71	81	75	97	70	49
71	82	52	95	87	73
71	58	94	61	70	71
71	84	93	63	82	64
73	86	79	88	58	74
73	70	43	97	64	56
73	88	85	96	64	65
76	81	61	94	40	84
76	65	69	68	56	69
76	78	87	87	83	72
78	86	63	85	81	80
79	76	62	92	61	64
79	76	58	92	73	66
80	77	82	91	65	75
81	63	85	80	59	63
81	80	70	45	74	77
84	67	65	72	78	80
85	88	85	87	86	73
85	70	80	76	79	71
86	53	71	72	51	77
86	78	87	97	81	73
87	45	69	56	79	54

Illustration using MS Excel

Preliminary Step: Create an Excel file containing your data.

1. Highlight cells containing the data. If there are column labels, include them.

2. Click Data > Sort (OR Home > Sort & Filter > Custom Sort)

3. Fill up the dialogue box. Identify the sorting variable in the "Sort by" box. Click desired arrangement (ascending or descending). Click "My data has headers" if there are column labels.

Frequency Distribution

Definition 5.3.
The **frequency distribution** is a way of summarizing data by showing the number of observations that belong in the different categories or classes.

Notes:
- This is also called **grouped data**.
- The classes may be distinct values/qualitative categories or the classes may be intervals of values of the variable. If the classes are distinct values then the frequency distribution is called **single value grouping**. If the classes are intervals of values then the frequency distribution is called **grouping by class intervals**.

Example: Single Value Grouping

Suppose we have data on the number of children of 50 currently married women using any modern contraceptive method. Construct its frequency distribution.

0	0	1	2	2	2	3	3	4	4
0	0	1	2	2	3	3	3	4	4
0	1	1	2	2	3	3	3	4	4
0	1	1	2	2	3	3	3	4	5
0	1	1	2	2	3	3	3	4	5

We use single value grouping since there are only a few distinct values in the data set.

Example: Single Value Grouping

Distribution of Married Women Using any Modern Contraceptive
by Number of Children

Number of Children	Number of Married Women
0	7
1	8
2	11
3	14
4	8
5	2

Example: Grouping by Class Intervals

Table 5.4 a

Final Grade	No. of Students
40 - 49	8
50 59	23
60 - 69	42
70 - 79	62
80 - 89	58
90 - 99	17
Total	210

Table 5.4 b

Final Grade	No. of Students
40 - 46	7
47 53	9
54 - 60	18
61 - 67	30
68 - 74	41
75 - 81	48
82 - 88	39
89 - 95	13
96 - 102	5
Total	210

Table 5.4 c

Final Grade	No. of Students
40 - 45	7
46 51	6
52 - 57	10
58 - 63	24
64 - 69	26
70 - 75	35
76 - 81	45
82 - 87	34
88 - 93	13
94 - 99	10
Total	210

Reading Assignment

- Steps in Constructing the Frequency Distribution Table (FDT) using Equal Class Sizes (page 175-176)

- Handout (Google Drive)

- Usually, we use statistical software to generate frequency distribution.

Remarks

When we present the frequency distribution in a formal statistical table, always put the appropriate column labels to describe the data. Thus, instead of using lower and upper class limits and f, we use the variable of interest and identify the units being counted.

Frequency Distribution of Weights of Children
(in pounds)

Weight	No. of Children
35.00 - 39.99	7
40.00 - 44.99	6
45.00 - 49.99	5
50.00 - 54.99	1
55.00 - 59.99	3
60.00 - 64.99	6
65.00 - 69.99	2
Total	30

Remarks

- Whenever possible, all classes should be of the same size.
  - If the classes sizes are not equal, it becomes difficult to tell whether the differences in class frequencies result mainly from differences in the concentration of items or from differences in the class sizes.
- However, there are certain instances when it is not practical to use equal class sizes. One such case is **when the distribution is badly skewed**.

Remarks

Suppose the salaries range from 20,000 to 1,000,000 but 95% of observations are less than 80,000. If all class sizes are equal to 100,000 so that the first class is 0 − 99,999 then the FDT is almost useless. There will be around 10 classes but 95% belong in the first class. No information would be provided about the distribution of salaries of the 95% with salaries from 1 − 100,000.

For such distributions, unequal class intervals are generally used. For instance, equal class sizes of say 10,000 might be used for the range wherein most of the salaries fall, after which the size might increase to say 100,000. Another option is to use an open class interval to account for the remainder of salaries that are not included in the classification.

Illustration: MS Excel

How to tally frequencies easily in Excel?

0. Install **Analysis ToolPak** add-in.
1. Enter data, one column per variable (include column label).
2. In another column, enter the upper class limits.
3. Click Tools/Data Analysis/Histogram/OK.
4. Fill up dialogue box. Identify cells containing the data in Input Range box. Identify cells containing the upper class limits in Bin Range box. Click Labels then OK.
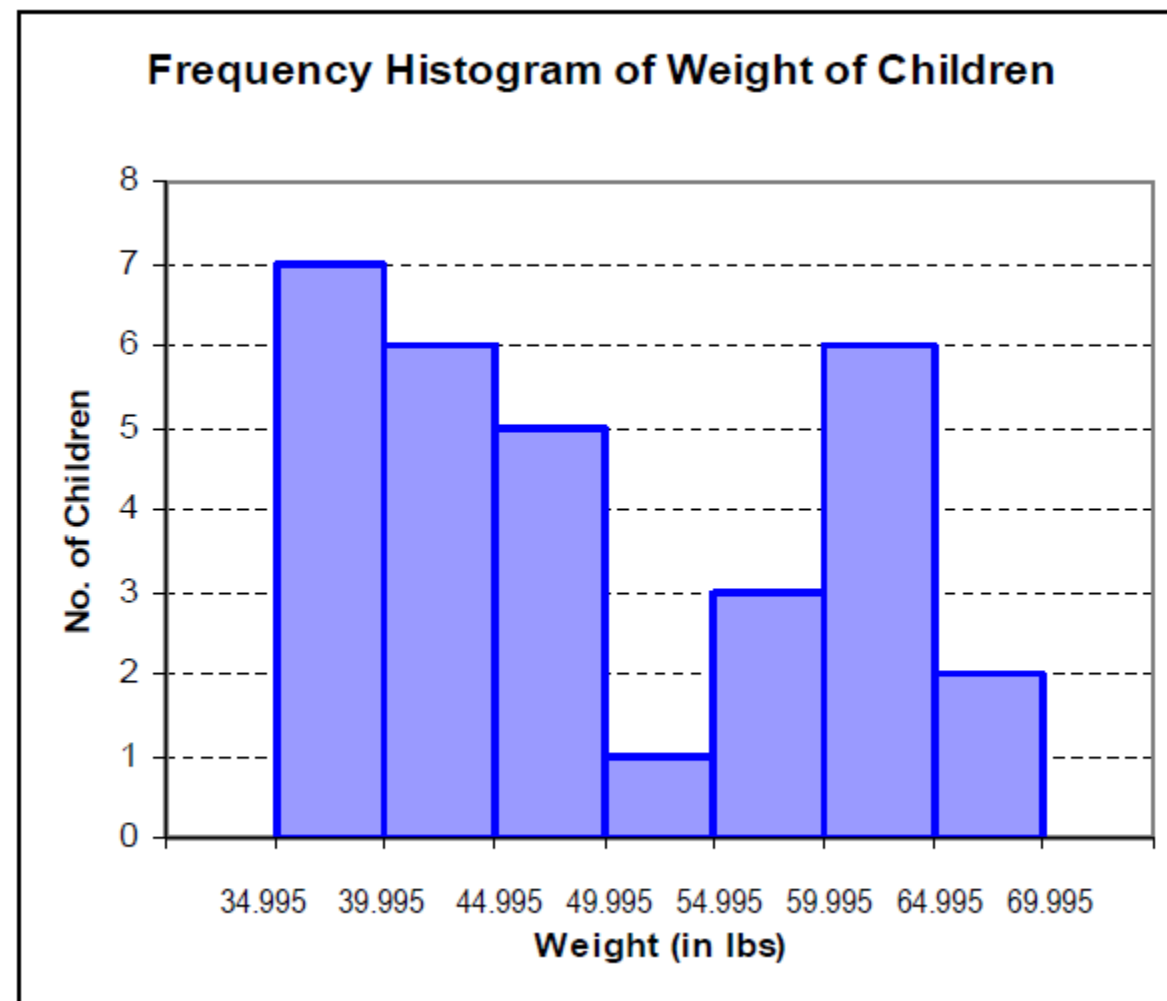
Graphical Presentation of the FDT

1. Frequency Histogram

- The frequency histogram shows the **overall picture of the distribution** of the observed values in the dataset.

- It displays the **class boundaries on the horizontal axis** and the **class frequencies on the vertical axis**.

- The frequency histogram **shows the shape** of the distribution.

- The **area under the frequency histogram** corresponds to the **total number of observations**.

- The tallest vertical bar shows the frequency of the class interval with the largest class frequency.

Graphical Presentation of the FDT

1. Frequency Histogram

LCB	UCB	f
34.995	39.995	7
39.995	44.995	6
44.995	49.995	5
49.995	54.995	1
54.995	59.995	3
59.995	64.995	6
64.995	69.995	2

Frequency Histogram of Weight of Children

Graphical Presentation of the FDT

2. Relative Frequency/ RF Percentage Histogram

- The RF or RFP histogram displays the **class boundaries on the horizontal axis** and the **relative frequencies or RFPs of the class intervals on the vertical axis**.

- It represents the relative frequency of each class by a vertical bar whose height is equal to the relative frequency of the class.

- **The shape of the relative frequency histogram and frequency histogram are the same.**

Graphical Presentation of the FDT

3. Frequency Polygon

The frequency polygon is the **line graph of the FDT**.

We plot the frequencies against the corresponding class marks then connect the points by straight lines.

Since this is a polygon, we need to close the chart by putting an additional class mark at both ends of the horizontal axis and bring down the line to the horizontal axis at the midpoints of the additional class.

Graphical Presentation of the FDT

3. Frequency Polygon

The frequency polygon **also shows the shape** of the data distribution.

The advantage of the frequency polygon over the frequency histogram is that **we may draw and compare two or more frequency distributions**.

The advantage of the frequency histogram is that it is easier to interpret since the boundaries of the class intervals are clearly displayed as the boundaries of the bars.

From the frequency histogram, we can easily construct the frequency polygon by connecting the midpoints of the adjoining bars and then putting two additional classes at both ends to close it.

The area under the frequency polygon is the same as the area under the frequency histogram.
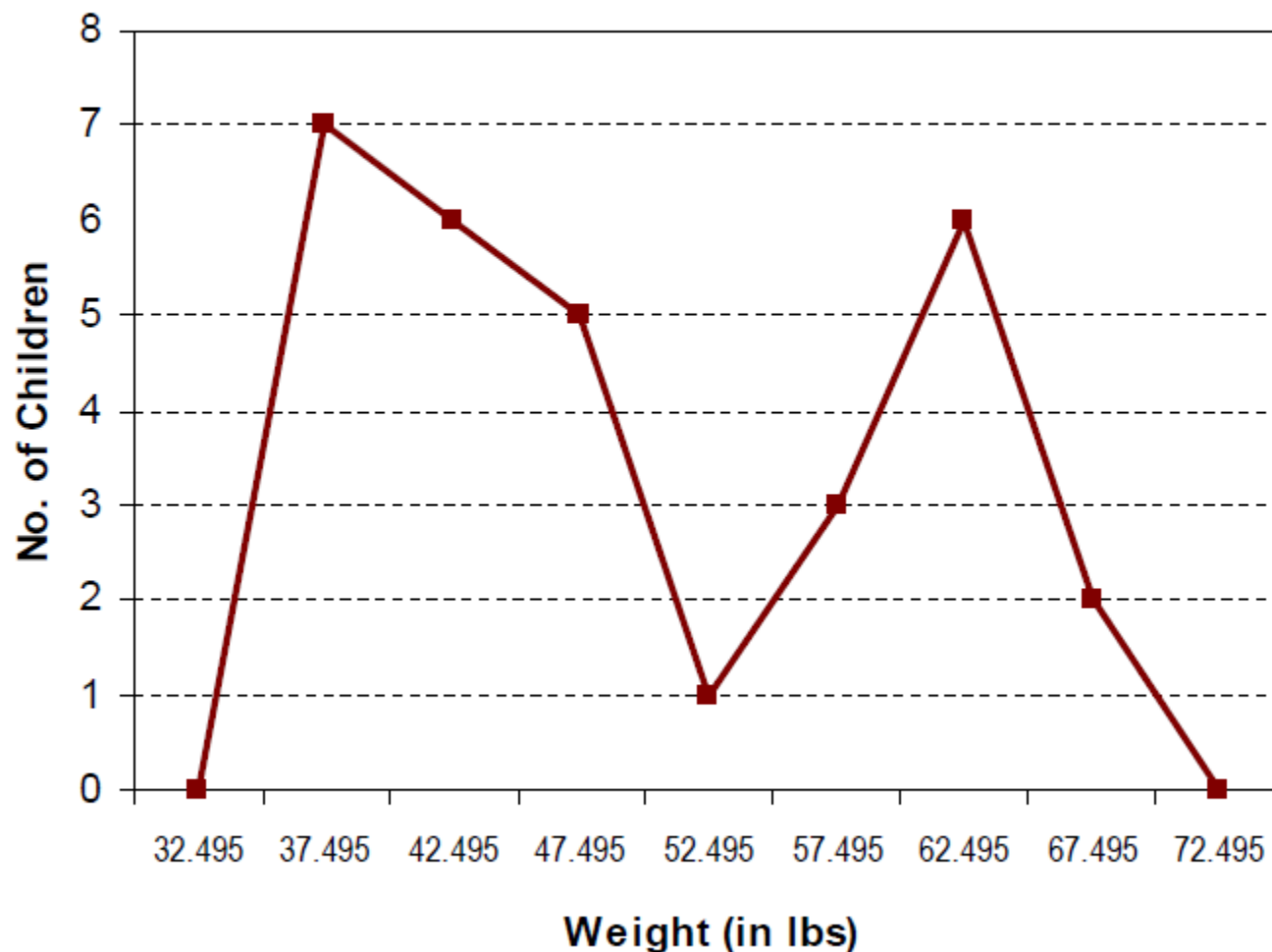
Both areas represent the total number of observations.

Graphical Presentation of the FDT

3. Frequency Polygon

CM	No. of children
32.495	0
37.495	7
42.495	6
47.495	5
52.495	1
57.495	3
62.495	6
67.495	2
72.495	0

Frequency Polygon of Weight of Children

Graphical Presentation of the FDT

4. Ogives

The ogive is the plot of the cumulative frequency distribution. This graphical representation is used when we need to determine the number of observations below or above a particular class boundary.

The **less than ogive** is the plot of the less than cumulative frequencies against the upper class boundaries. On the other hand, the **greater than ogive** is the plot of the greater than cumulative frequencies against the lower class boundaries. Connect the successive points by straight lines.
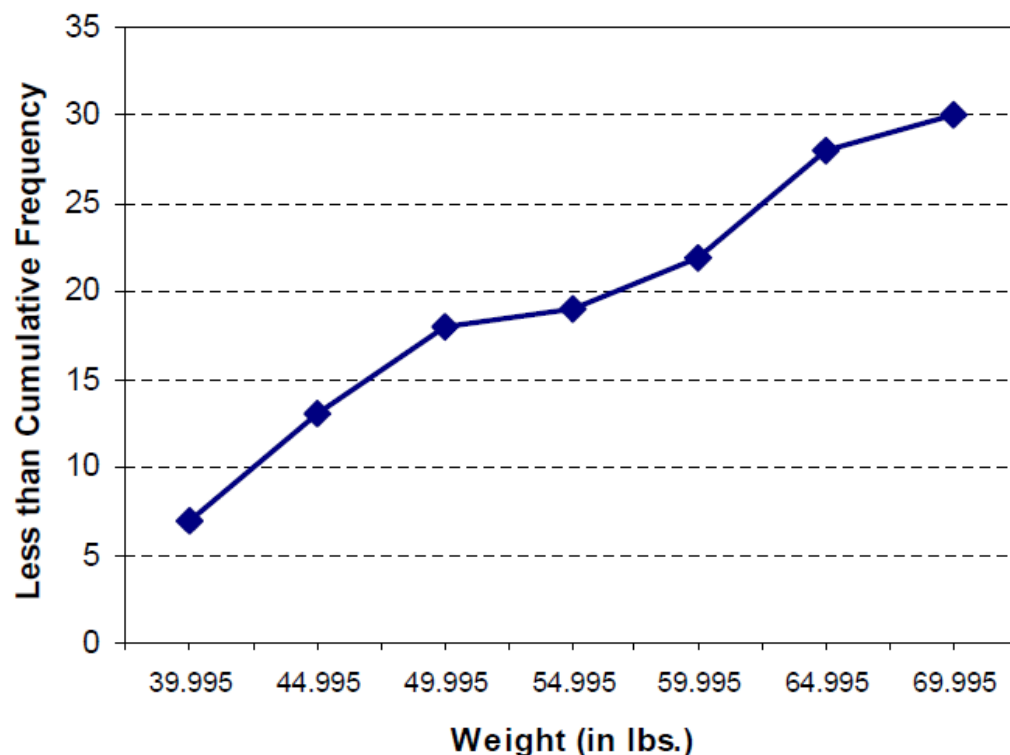
If we superimpose the less than and greater than ogives, the point of intersection gives us the value of the median. The median divides the ordered observations into two equal parts.

Graphical Presentation of the FDT

4. Ogives