QMS102-F16

Lecture 02

Lecture 01 Review

- What statistics is
- How statistics is fundamental to business
- The basic concepts and vocabulary of statistics
- Data Types
- Measurement Scale
- Stem-and-leaf Plot

Lecture 02

Outline:

- Rules and convention to construct a stem-and-Leaf Plot
- How do you construct a Stem-and-Leaf Plot with negative and positive data values
- Interpretation of results using Stem-and-Leaf Plot
- Frequency Distribution

Stem-and-Leaf Displays

- A stem-and-leaf display conveys information about the following aspects of the data:

- identification of a typical or representative value
- extent of spread about the typical value
- presence of any gaps in the data
- extent of symmetry in the distribution of values
- number and location of peaks
- presence of any outlying values

Rules and Convention to Construct a Stem-and-Leaf Plot

Stem Rules

- The number of stems should be from 6 to 13 stems
- The stem values should be consecutive numbers or repeated numbers. The numbers may each be repeated twice or 5 times.
- The stem units must be indicated if the stem is not to be taken at face value.
- There must be at least one leaf associated with the first and last stem.

Rules and Convention to Construct a Stem-and-Leaf Plot

Leaf Rules

- The leaf of each data value is the next single digit after the stem.
  - When the stems are repeated twice, the leaves values 0 to 4 go to the first stem and the leaves values 5 to 9 go to the second stem of the same values. This order is reversed when the stems are negative values.
  - When the stems are repeated 5 times, the leaves values 0 to 1 go to the first stem, values 2 and 3 go to the second stem, values 4 and 5 go to the third stem, values 6 and 7 go to the fourth stem, and the values 8 and 9 go to the fifth stem. This order is reversed when the stems are negative values.

Rules and Convention to Construct a Stem-and-Leaf Plot

Leaf Rules (cont'd)

- There is no rounding off.
- The leaf values are written in ascending order when positive and descending order when negative.
- The leaf values must be evenly spaced.
- No commas or dashes between the numbers are allowed.

Stem Rule 1

- The number of stems should be from 6 to 13 items

- There are 4 stems, which does not meet the requirement of stems in the plot.
- What should you do? To increase the number of stems, you should apply.

Stem	Leaf
1	67788899
2	0012257
3	28
4	2

- There are 7 items, which meets the requirement of stems in the plot.

Stem	Leaf
1	67788899
2	0012257
3	28
4	2
5	233478999
6	00236666888
7	222234447777

Stem Rule 2

- The stem values should be consecutive numbers or repeated numbers. The numbers may each be repeated twice or 5 times.

- Stem values are not consecutive numbers in the plot.

Stem	Leaf
1	67788899
4	0012257
5	28
8	2

- Stem values are consecutive numbers in the plot.

Stem	Leaf
1	67788899
2	0012257
3	28
4	2
5	233478999
6	00236666888
7	222234447777

Stem Rule 2

- The stem values should be consecutive numbers or repeated numbers. The numbers may each be repeated twice or 5 times.
- Each stem value should be repeated twice or 5 times

Stem	Leaf
1	2467788899
2	0012257
3	28

Stem	Leaf
1	24
1	67788899
2	00122
2	57
3	2
3	8

Stem	Leaf
1	133567788899
2	00122578

Stem	Leaf
1	1
1	33
1	5
1	677
1	88899
2	001
2	22
2	5
2	7
2	8

Stem Rule 3

- The stem units must be indicated if the stem is not to be taken at face value
- The stem unit must represent the actual data value.

Stem (10)	Leaf
2	24
3	67788899
4	00122
5	57
6	2
7	8

- Blank means that the stem should be taken at face value

Stem	Leaf
2	24
3	67788899
4	00122
5	57
6	2
7	8

Stem Rule 4

- There must be at least one leaf associated with the first and last stem.

Stem (10)	Leaf
2	24
3	67788899
4	00122
5	57
6	2
7	8

Stem	Leaf
2	24
3	67788899
4	00122
5	57
6	2
7	8

Leaf Rule 1

The leaf for each data value is the next single digit after the stem.

- When the stems are repeated twice, the leaves values 0 to 4 go to the first stem and the leaves values 5 to 9 go to the second stem of the same values.
- This order is reversed when the stems are negative
- When the stems are repeated 5 times. the leaves values 0 and 1 go to the first stem, values 2 and 3 go to the second stem. values 4 and 5 go to the third stem, values 6 and to go to the fourth stem, and values 8 and 9 go to the fifth stem.
- This order is reversed when the stems are negative values.

Leaf Rule 1

- There must be at least one leaf associated with the first and last stem.
- A) Rule for positive stems that are repeated twice

Stem	Leaf
2	24
3	67788899
4	00122
5	57

Stem	Leaf
2	24
2	
3	
3	67788899
4	00122
4	
5	
5	57

  - The positive stem is repeated 2 times.
  - Leaves values 0 to 4 go to the first stem.
  - Leaves values 5 to 9 go to the second stem.
- For the positive stem values, leaves are arranged in ascending order.

Leaf Rule 1

- There must be at least one leaf associated with the first and last stem.
- B) Rule for negative stems that are repeated twice

Stem	Leaf
-5	8887632
-4	9985533320
-3	9664444311
-2	8777753320

Stem	Leaf
-5	88876
-5	32
-4	99855
-4	33320
-3	966
-3	4444311
-2	877775
-2	3320

  - The negative stem is repeated 2 times.
  - Leaves values 5 to 9 go to the first stem.
  - Leaves values 0 to 4 go to the second stem.

- For the negative stem values, leaves are arranged in descending order.

Leaf Rule 2 and 3

- There is no rounding off.
- The record value 876 with a defined stem value of 8, you should record leaf value 7 next to 8.

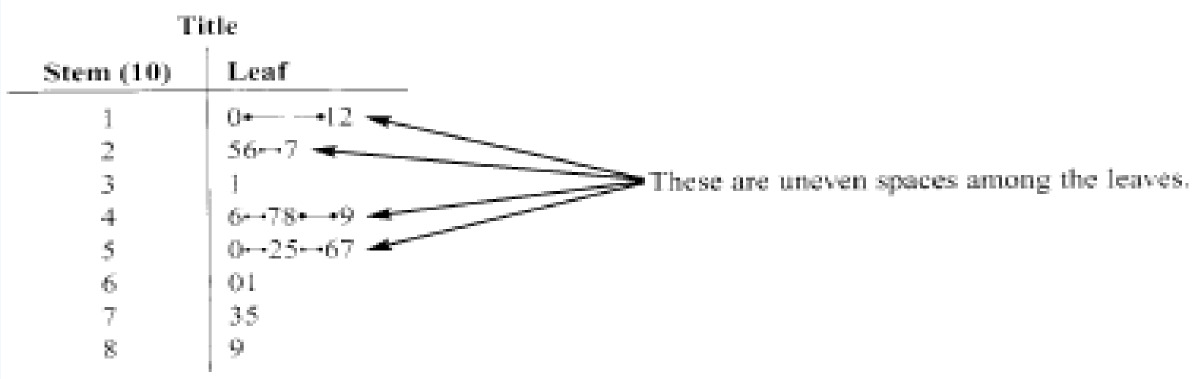
- The leaf values are written in ascending order when positive, and descending order when negative.

Title

Stem (100,00)	Leaf (10,000)	
-0	9 7 5	Corresponds to the negative stem; the leaves are arranged in descending order.
-0	4 4 3 1	
0	0 0 2 4	
0	5 5 6 8 8	
1	2	
1	5 6	
2	1 1 3	Corresponds to the positive stem; the leaves are arranged in ascending order.
2	7 9	

Leaf Rule 4 and 5

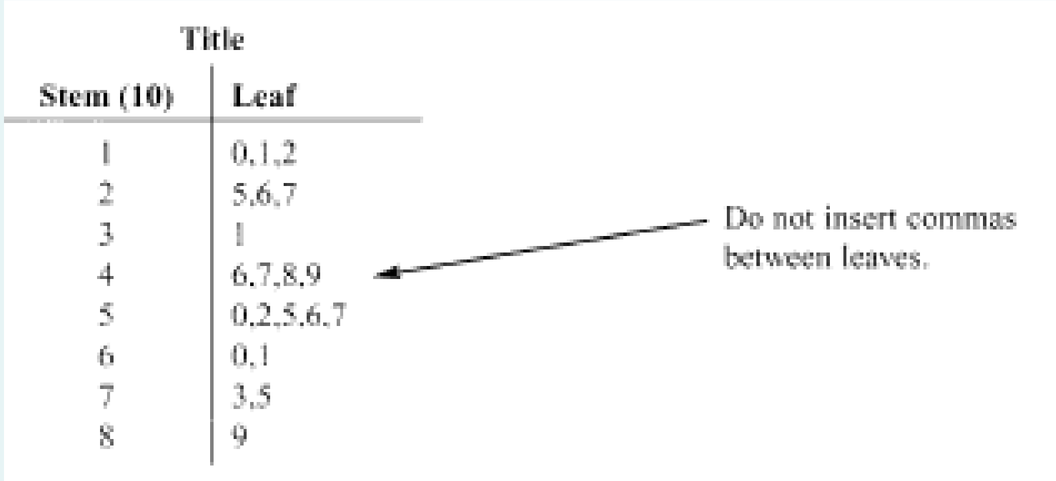
- The leaf values must be evenly spaced.



Title

Stem (10)	Leaf
1	0 12
2	56 7
3	1
4	6 78 9
5	0 25 67
6	01
7	35
8	9

- No commas or dashes between the numbers are allowed.



Title

Stem (10)	Leaf
1	0,1,2
2	5,6,7
3	1
4	6,7,8,9
5	0,2,5,6,7
6	0,1
7	3,5
8	9

How do you construct a Stem-and-leaf plot using data with positive values?

Steps:

- Sort the data in ascending order
- Find maximum and minimum values
- Split the minimum and maximum values into stems and leaves.
- Construct the stem-and-leaf plot.
- Record the leaf values
- **For each step you have to check stem and leaf rules.**

How do you construct a Stem-and-leaf plot using data with negative values?

Steps:

- Sort the data in ascending order
- Find maximum and minimum values
- Split the minimum and maximum values into stems and leaves.
- Construct the stem-and-leaf plot.
- Record the leaf values
- **For each step you have to check stem and leaf rules.**

Interpretation of results using stem-and-leaf plot

- A) Using the stem-and-leaf plot to detect the shape of the data distribution.
- The symmetry of a data distribution can be classified in three ways:
  - l) skewed to the left (the bulk of the data is located on the right side of the distribution; see Graphic A),
  - (2) symmetrical (the bulk of the data is located in the middle of the distribution; see Graphic B) and
  - (3)skewed to the right (the bulk of the data is located on the left side of the distribution; see Graphic C).

Graphic A: Skewed to the Left

Stem	Leaf
10	2
11	1 2
12	3 4 4
13	5 5 6 7 8
14	3 3 5 6 7 7 7
15	2 2 3 3 4 4 5 5 6 7 8
16	2 3 4 5 6 6 7 8 8 8 9 9 9
17	1 5 5 6
18	0 1

Graphic B: Symmetrical

Stem	Leaf
10	2
11	1 2
12	3 4 4
13	1 5 5 6 7 8 9
14	3 3 5 6 7 7 7 8 8 9
15	2 2 3 3 4 4
16	2 3 4 5
17	1 5 5
18	0 1

Graphic C: Skewed to the Right

Stem	Leaf
10	2
11	1 2 3 5 6 6 6 7 7 8
12	2 3 4 5 6 6 7 8 8 9 9 9
13	0 0 4 4 5 5 6 7
14	3 3 5 6 7 7
15	2 2 3 3
16	3 4 4
17	1 5 8
18	1 2

Interpretation of results using stem-and-leaf plot

- B) Using the stem-and-leaf plot to extract information
- Example:
- A random sample of employees was taken from a technology company, and the hours of overtime claimed per month were and recorded and summarized in the following stem-and-leaf plot.

Stem (10)	Leaf
10	2
11	1 2 3 5 6 6 6 7 7 8
12	2 3 4 5 6 6 7 8 8 8 9 9 9
13	0 0 4 4 5 5 6 7
14	3 3 5 6 7 7
15	2 2 3 3
16	3 4 4
17	1 5 8
18	1 2

Interpretation of results using stem-and-leaf plot cont'd

- How many employees are being sampled?

Each leaf represents one data value of an employee. Therefore, the sample size is equal to the number of leaves. There are 50 leaves, which means there are 50 employees in this sample.

- What percent of employees claimed at least 152 hours per month?

"At least 152" includes the value 152. By counting the values including 152 and greater than 152, you get 12 values. These values are 152, 152, 153, 153, 163, 164, 164, 171, 175, 178, 181, and 182. Each employee has one overtime value. Therefore, 12 values represents 12 employees. Thus, you have (12/50)*100=24% of employees claimed at least 152 hours per month.

- What percent of employees claimed less than 130 hours per month?

"Less than 130" does not include the value 130. By counting the values less than 130 and not including 130, you get 24 values. These values are 102, 111, 112, 113, 115, 116, 116, 116, 117, 117, 118, 122, 123, 124, 125, 126, 126, 127, 128, 128, 128, 129, 129, and 129. Each employee has one overtime value. Therefore, 24 values represents 24 employees. Thus, you have (24/50)*100=48% of employees claimed less than 130 hours per month.

Interpretation of results using stem-and-leaf plot cont'd

- What percent of employees claimed more than 163 hours per month?

"More than 163" does not include the value 163. By counting the values more than 163 and not including 163, you get 7 values. These values are 164, 164, 171, 175, 178, 181, and 182. Each employee has one overtime value. Therefore, 7 values represents 7 employees. Thus, you have (7/50)*100 = 14% of employees claimed more than 163 hours per month.

- What percent of employees claimed at most 115 hours per month?

"At most 115 hours" includes the value 115. By counting the values including 115 and less than 115, you get 5 values. These values are 102, 111, 112, 113, and 115. Each employee has one overtime value. Therefore, 5 values represents 5 employees. Thus, you have (5/50)*100=10% of employees claimed at most 115 hours per month.

Frequency Distribution

- The **frequency distribution** is a summary table in which the data are arranged into numerically ordered classes.
- General Guidelines for Constructing a Frequency Distribution
- 1. Number of classes: You should not have too few or too many classes. For the questions you will solve in this course you should use 5 to l0 classes in your final constructed frequency distribution.
- 2. Notation for indicating classes: There are several possible notations that are used to designate the classes. In this course we will use the "and under" notation(e.g., 260 and under 270). The numerical values used in designating the classes when using the "and under“ notation are called boundaries.

Frequency Distribution cont'd

- 3. Close-ended classes: A close-ended class has both a lower and an upper boundary. An open-ended class is missing one of these boundaries. For example, the class" l00 and over“ is an open-ended class because there is no upper boundary. You should avoid using open-ended classes in this course.
- 4. Class width (denoted as CW): A class that has an upper and a lower boundary has a finite class width. The class width is determined by the following formula:

  Class Width = Upper boundary - Lower boundary

- It is best to select a class width that is an "easy" number to work with. The recommended class widths are shown in the following Table

Frequency Distribution cont'd

- Table: A list of "Nice" numbers

For narrower classes			For wider classes	
Divide by 10			Multiply by 10	
0.01	0.1	1	10	100
0.02	0.2	2	20	200
0.025	0.25 (if data has at least 2 decimals)	2.5(If the data has at least one decimal)	25	250
0.05	0.5	5	50	500

Note: In this course, when you are asked to construct a frequency distribution, all classes must have the same class width.

Frequency Distribution cont'd

- 5. No gaps between classes: Classes are mutually exclusive and are not overlapping. In other words, every possible value must fall into exactly one class. There must be no gaps between the classes (i.e.. the upper boundary of one class will also be the lower boundary of the next class).
- 6. All classes must have the same class width: When working with continuous data, the upper boundary of one class and the lower boundary of the following class reference the same value. Since the boundary value should fall into only one class, the wording" and under" can be used to clarify that the upper boundaries are not included in the classes. To keep it simple, use equal class width for each class.

Frequency Distribution cont'd

- 7. The first and last class must contain frequencies: The lowest value (or minimum value) is in the first class, and the highest value (or maximum value) is in the last class.
- 8. Boundaries: There are several guidelines that should be used to determine the actual numerical values of the boundaries.
  - a) The boundaries should look like the data, that is, have the same number of decimal places as the data. If the data have one decimal point, then the boundaries should have one decimal point. If the data are whole numbers then the boundaries should be whole numbers. The reason for having this rule is to retain the information about the type of quantitative data. When we grouped the data, we lost the information about the nature of the data. With the grouped data in the form of a frequency distribution, we only know how many values (i.e. frequency) fall in each class.
  - b) Each boundary should be a multiple of the chosen class width.

Pick a Nice Number for the class width

- "Nice“ numbers are listed in Slide 26. Always start off with this list of 4 numbers; 1, 2, 2.5, 5.
- If you want a larger class width, you can multiple each number in the list of 1,2,2.5,5 by 10 and you will have 10, 20, 25, 50.
- If you still want a larger class width, multiply again by 10 and you have 100, 200, 250, 500.
- If you want a smaller class width, divide each number in the list of 1, 2, 2.5, 5 by l0 and you have 0.1, 0.20, 0.25, 0.5.
- If you still want a smaller class width, divide again by l0 and you have 0.01, 0.02, 0.025, 0.05.

Determine Class Width

- To figure out the class width in your frequency distribution, you should estimate the class width using the following formula with the highest value (or maximum value) and lowest value (or minimum value) in a data set. Divide the difference by the minimum number of classes specified by the guide line. The minimum number of classes is 5. The value you obtained is called the "estimated class width" because this value rarely turns out to be 1, 2, 2.5 or 5 (the "nice" numbers).

$$Estimated\ class\ width = \frac{Highest\ value\ - Lowest\ value}{Minimum\ number\ of\ classes\ recommended}$$

$$Estimated\ class\ width = \frac{Highest\ value\ - Lowest\ value}{5}$$

NOTE: Whichever class width you choose to construct your frequency distribution, you must always check that the number of classes is between 5 and 10 classes.

Determine Class Width

Example: Given the lowest value 643 and the highest value 1028, calculate the estimated class width using the formula.

Minimum value = 643 and maximum value = 1028

$$Estimated\ class\ width = \frac{1028 - 743}{5} = 77$$

Start with this list	Multiply by 10	Multiply by 10 again
1	10	100
2	20	200
2.5	25	250
5	50	500

77 is between 50 and 100; thus you will pick either 50 or 100 to construct your classes.

Determine Class Width cont'd

- The first class should contain lowest data value and the last class should contain the highest data value.
- To obtain the first class lower boundary, we have to find the greatest multiple of the class width that is less than or equal to the minimum data value.

1. CW=50. The lowest boundary is 600 which is multiple of 50 and the first class contains the minimum data value. So the classes:

    600 and under 650

    650 and under 700

    700 and under 750

    750 and under 800

    800 and under 850

    850 and under 900

    900 and under 950

    950 and under 1000

    1000 and under 1050

Determine Class Width cont'd

2. CW=100. The lowest boundary is 600 which is multiple if 100 and the first class contains the minimum data value, 643. So the classes:

600 and under 700

700 and under 800

800 and under 900

900 and under 1000

1000 and under 1100

There are 5 classes for CW=100 and 9 classes for CW=50. Both resulting frequency distributions are correct.

Determine Class Width

Example: A manufacturer of insulation randomly selects 20 winter days and records the daily high temperature in degrees F.

24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27

Minimum value=12 and maximum value = 58

$$Estimated\ class\ width = \frac{58-12}{5} = 9.2$$

Start with this list	Multiply by 10
1	10
2	20
2.5	25
5	50

9.2 is between 5 and 10. So the class width can be either 5 or 10

Determine Class Width cont'd

1. CW=5. The lowest boundary is 5 which is multiple if 5 and the first class contains the minimum data value. So the classes:

    5 and under 10

    10 and under 15

    15 and under 20

    20 and under 25

    25 and under 30

    30 and under 35

    35 and under 40

    40 and under 45

    45 and under 50

    50 and under 55

    55 and under 60

Determine Class Width cont'd

2. CW=10. The lowest boundary is 10 which is multiple if 10 and the first class contains the minimum data value. So the classes:

10 and under 20

20 and under 30

30 and under 40

40 and under 50

50 and under 60

There are 11 classes for CW=5 and 5 classes for CW=10. Only for CW=10, the resulting frequency distribution is correct.

Issues about picking the correct class width

Question: Can I only use a nice number with one decimal place if my data only have one decimal place?

Answer:

- The answer is yes; for example, you may use the following nice numbers for your class width: 0.1, 0.2, 0.5, and 2.5. You may not use 0.25 because the data have one decimal place. However, you can use 0.25 if the data have two decimal places.

- Note: You have to check that the number of classes is between 5 and 10.

Organizing Numerical Data: Frequency Distribution Example

Example: A manufacturer of insulation randomly selects 20 winter days and records the daily high temperature in degrees F.

24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27

- Solution:
- Sort raw data in ascending order:
  12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
- Find max value: 58, min value: 12, range: **58 - 12 = 46**
- $Estimated\ class\ width = \frac{58-12}{5} = 9.2$
- For CW=10
  - **Class 1: 10 and under 20**
  - **Class 2: 20 and under 30**
  - **Class 3: 30 and under 40**
  - **Class 4: 40 and under 50**
  - **Class 5: 50 and under 60**

Organizing Numerical Data: Frequency Distribution Example

Count observations & assign to classes

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Frequency (f)
10 and under 20	3
20 and under 30	6
30 and under 40	5
40 and under 50	4
50 and under 60	2
Total	20

Relative frequency or percentage distributions

Consider the previous example. Find relative frequency or percentage distribution.

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Frequency (f)
10 and under 20	3
20 and under 30	6
30 and under 40	5
40 and under 50	4
50 and under 60	2
Total	20

Organizing Numerical Data: Relative & Percent Frequency Distribution Example

$$\textit{Relative frequency}(rf) = \frac{\textit{Frequency}\ (f)}{n}$$

$$\textit{Percentage of frequency}(\%) = \left(\frac{\textit{Frequency}\ (f)}{n}\right) \times 100$$

Class	Frequency	Relative Frequency	Percentage
10 and under 20	3	.15	15
20 and under 30	6	.30	30
30 and under 40	5	.25	25
40 and under 50	4	.20	20
50 and under 60	2	.10	10
Total	20	1.00	100

Organizing Numerical Data: Relative & Percent Frequency Distribution Example

Class	Frequency	Relative Frequency	Percentage
10 but less than 20	3	.15	15%
20 but less than 30	6	.30	30%
30 but less than 40	5	.25	25%
40 but less than 50	4	.20	20%
50 but less than 60	2	.10	10%
Total	20	1.00	100%

Relative Frequency = Frequency / Total, e.g. 0.10 = 2 / 20

Organizing Numerical Data: Cumulative Frequency Distribution Example

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10 but less than 20	3	15%	3	15%
20 but less than 30	6	30%	9	45%
30 but less than 40	5	25%	14	70%
40 but less than 50	4	20%	18	90%
50 but less than 60	2	10%	20	100%
Total	20	100	20	100%

Organizing Numerical Data: Cumulative Frequency Distribution Example

Class	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10 but less than 20	3	15%	3	15%
20 but less than 30	6	30%	9	45%
30 but less than 40	5	25%	14	70%
40 but less than 50	4	20%	18	90%
50 but less than 60	2	10%	20	100%
Total	20	100	20	100%

Cumulative Percentage = Cumulative Frequency / Total * 100 e.g. 45% = 100*9/20

Visualizing Numerical Data: The Histogram

- A vertical bar chart of the data in a frequency distribution is called a **histogram.**
- In a histogram there are no gaps between adjacent bars.
- The **class boundaries** (or **class midpoints**) are shown on the horizontal axis.
- The vertical axis is either **frequency, relative frequency,** or **percentage**.
- The height of the bars represent the frequency, relative frequency, or percentage.

Why Use a Frequency Distribution?

- It condenses the raw data into a more useful form
- It allows for a quick visual interpretation of the data
- It enables the determination of the major characteristics of the data set including where the data are concentrated / clustered

Organizing Numerical Data: Frequency Distribution Example

Data in ordered array:

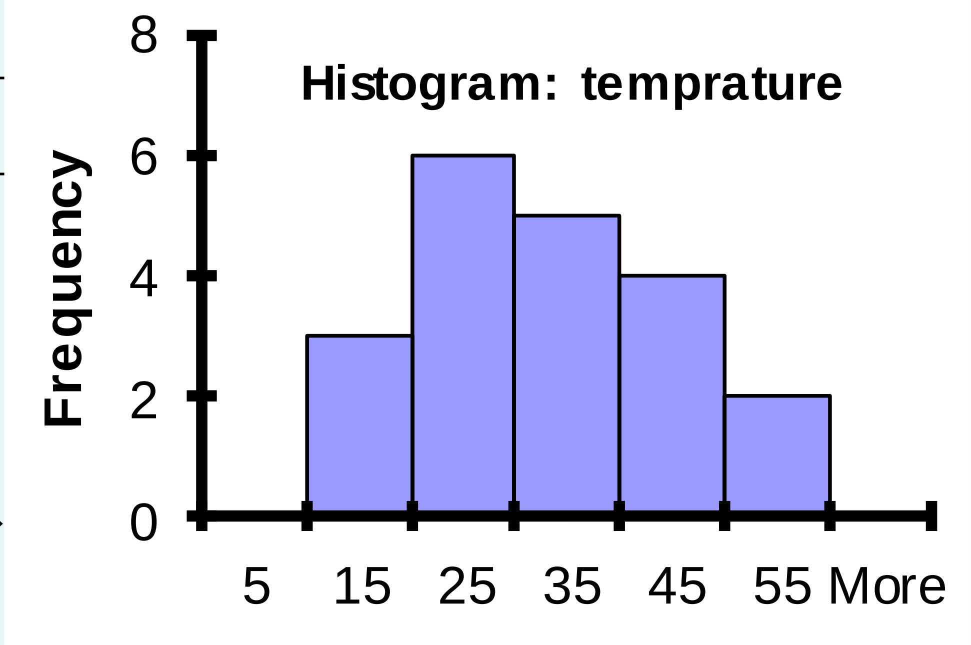
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Midpoints	Frequency
10 and under 20	15	3
20 and under 30	25	6
30 and under 40	35	5
40 and under 50	45	4
50 and under 60	55	2
Total		20

Visualizing Numerical Data: The Histogram

Class	Frequency	Relative Frequency	Percentage
10 and under 20	3	.15	15
20 and under 30	6	.30	30
30 and under 40	5	.25	25
40 and under 50	4	.20	20
50 and under 60	2	.10	10
Total	20	1.00	100

(In a percentage histogram the vertical axis would be defined to show the percentage of observations per class)



Frequency Distributions: Some Tips

- Different class boundaries may provide different pictures for the same data (especially for smaller data sets)

- Shifts in data concentration may show up when different class boundaries are chosen

- As the size of the data set increases, the impact of alterations in the selection of class boundaries is greatly reduced

- When comparing two or more groups with different sample sizes, you must use either a relative frequency or a percentage distribution

Visualizing Numerical Data: The Polygon

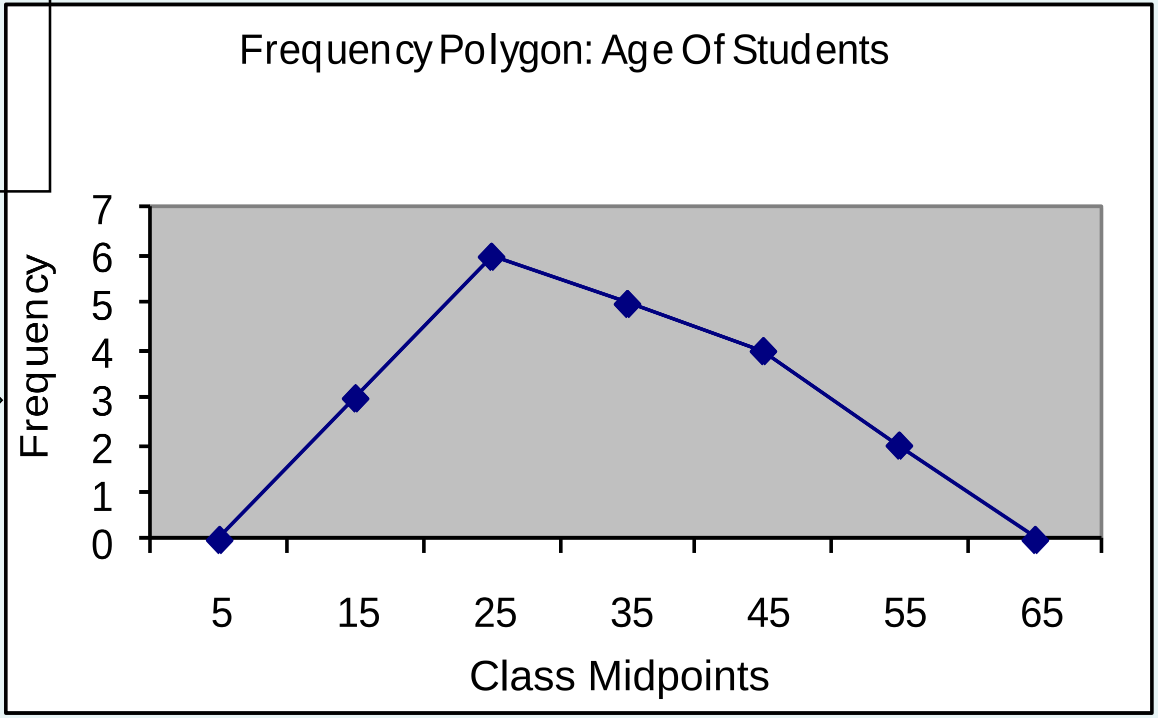
- A **percentage polygon** is formed by having the midpoint of each class represent the data in that class and then connecting the sequence of midpoints at their respective class percentages.

- The **cumulative percentage polygon,** or **ogive,** displays the variable of interest along the *X* axis, and the cumulative percentages along the *Y* axis.

- Useful when there are two or more groups to compare.

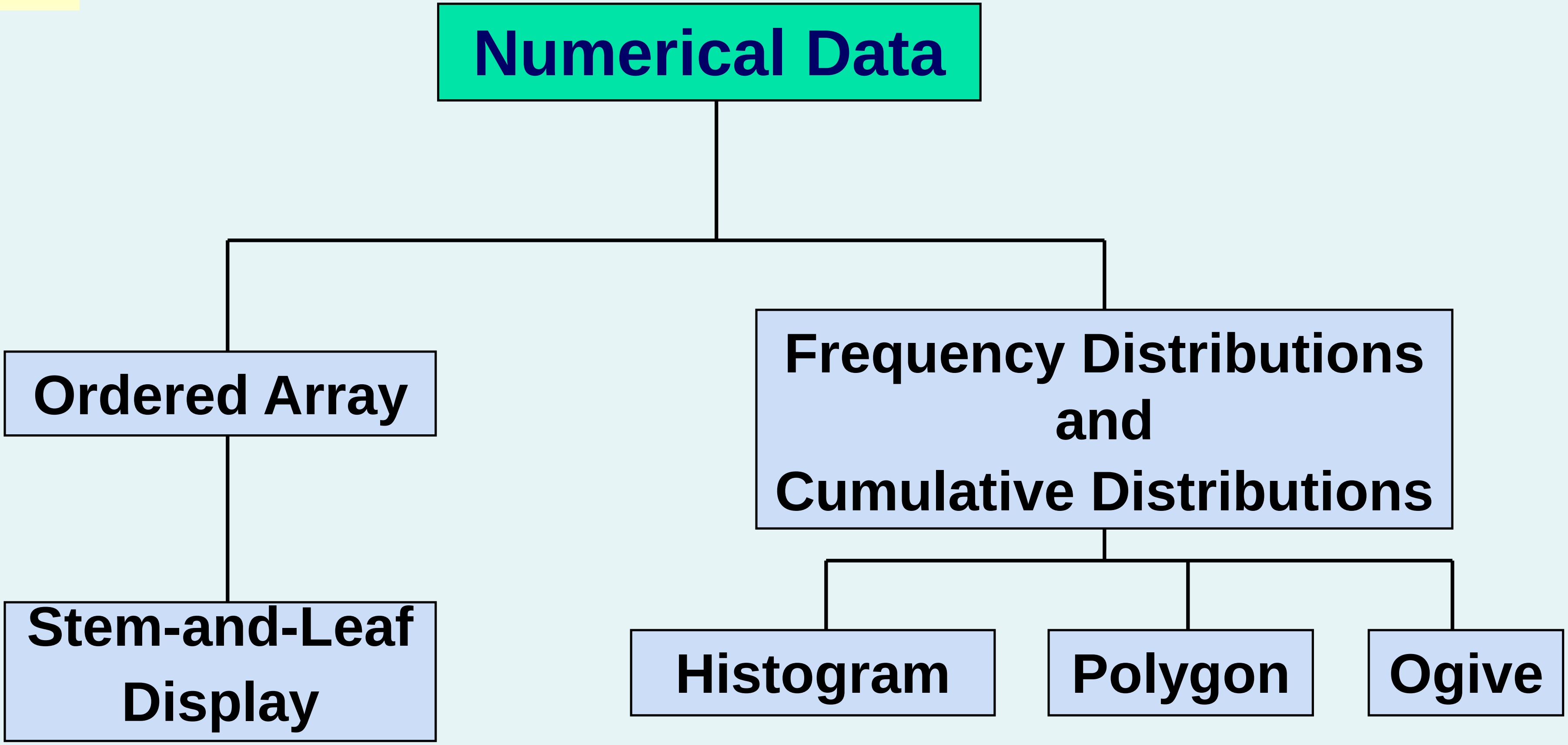
Visualizing Numerical Data: The Frequency Polygon

Class	Class Midpoint	Frequency
10 but less than 20	15	3
20 but less than 30	25	6
30 but less than 40	35	5
40 but less than 50	45	4
50 but less than 60	55	2



(In a percentage polygon the vertical axis would be defined to show the percentage of observations per class)

Visualizing Numerical Data By Using Graphical Displays



Graphical Techniques for Interval Data

- **Example :** Providing information concerning the monthly bills of new subscribers in the first month after signing on with a telephone company.
  - Collect data
  - Prepare a frequency distribution
  - Draw a histogram

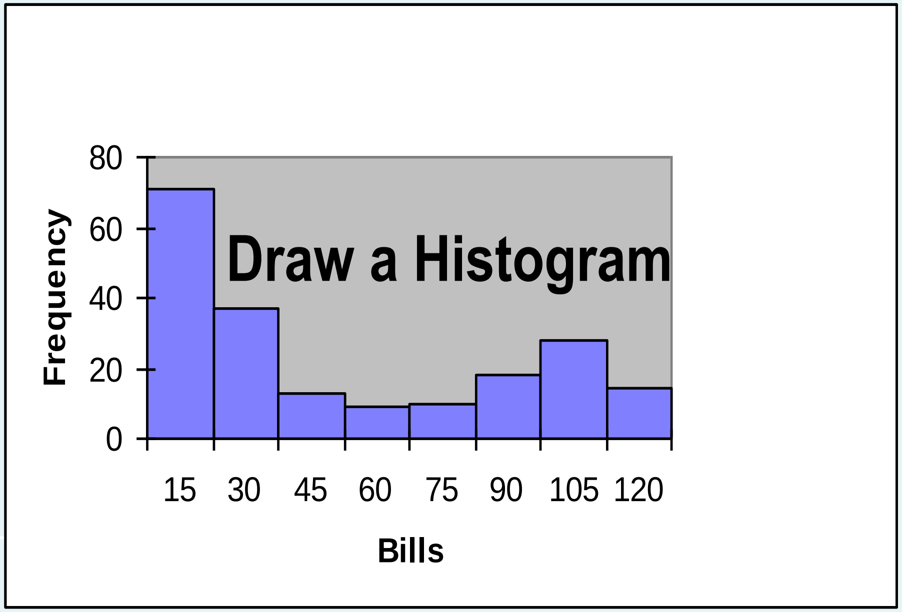
Example : Providing information

Collect data

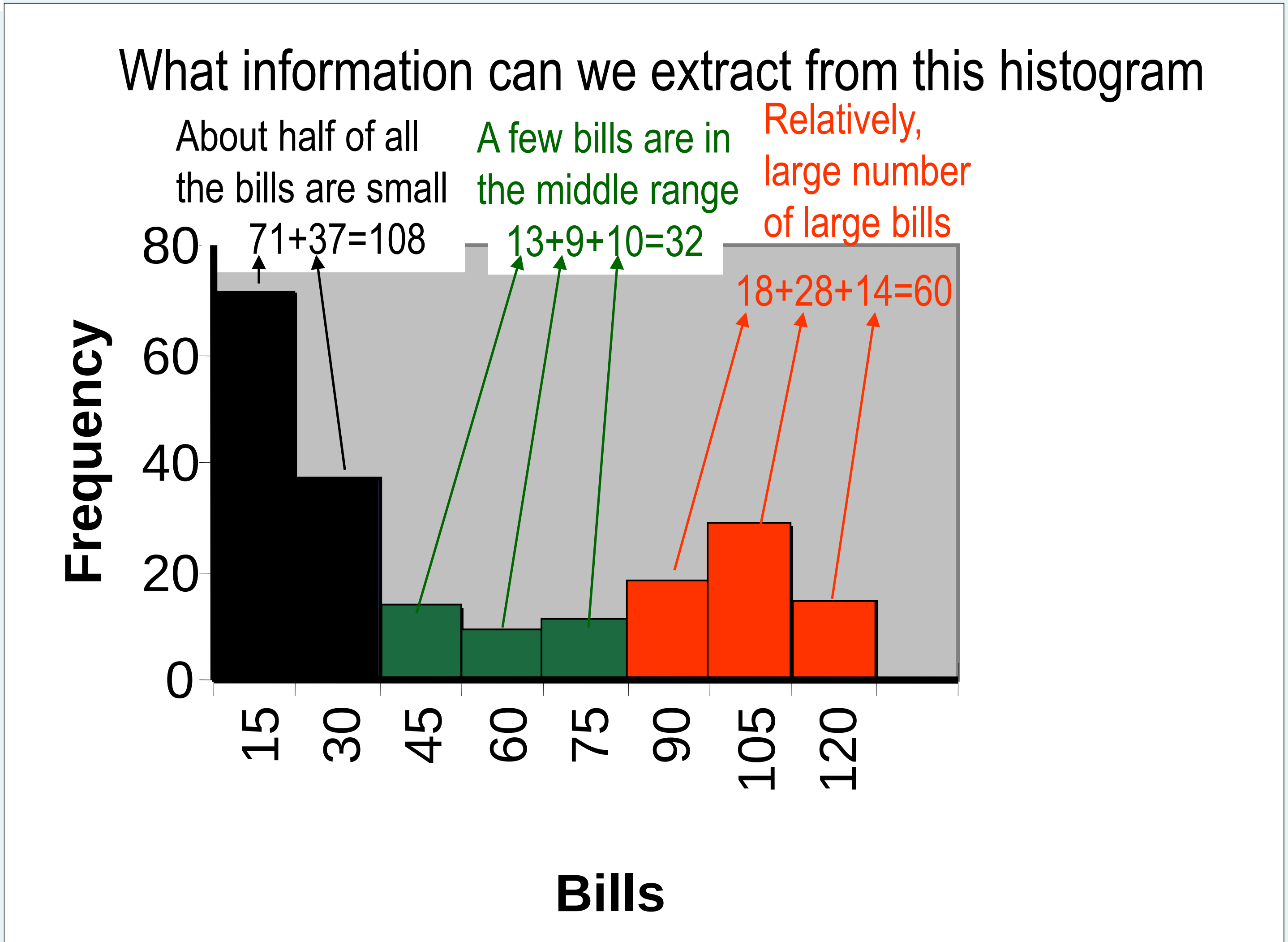
Bills
42.19
38.45
29.23
89.35
118.04
110.46
0.00
72.88
83.05
.
.

(There are 200 data points

Bin	*Frequency*
15	71
30	37
45	13
60	9
75	10
90	18
105	28
120	14



Example : Providing information



Relative Frequency Histograms Process

- Divide the range of the data into **5-10 subintervals** of equal length
- Calculate the **approximate width** of the subinterval as range/number of subintervals
- Round the approximate width up to a convenient value
- Use the method of **left inclusion**, including the left endpoint but not the right, in your tally

Relative Frequency Histograms Process, Continued

- Create a **statistical table** including the subintervals, their frequencies, and relative frequencies

- Draw the **relative frequency histogram**, plotting the subintervals on the horizontal axis and the relative frequencies on the vertical axis

The height of a bar represents:

- The **proportion** of measurements falling in that class or subinterval
- The **probability** that a single measurement, drawn at random from the set, will belong to that class or subinterval

Relative Frequency Histograms

- If the data are discrete, one class might be assigned for each integer value taken on by the data

- A large number of integer values may need to be grouped into classes