W06
Graphical Resresentaiton of Experimental Data

Frequency Distributions and Their Graphs

Frequency Distribution

- A table that shows **classes** or **intervals** of data with a count of the number of entries in each class.
- The **frequency, *f*,** of a class is the number of data entries in the class.

Class width 6 – 1 = 5

Class	Frequency, f
1 – 5	5
6 – 10	8
11 – 15	6
16 – 20	8
21 – 25	5
26 – 30	4

Lower class limits

Upper class limits

Rules:

1. Decide on the number of classes.
   - Usually between 5 and 20; otherwise, it may be difficult to detect any patterns.
2. Find the class width.
   - Determine the range of the data.
   - Divide the range by the number of classes.
   - *Round up to the next convenient number.*
3. Find the class limits.
   - You can use the minimum data entry as the lower limit of the first class.
   - Find the remaining lower limits (add the class width to the lower limit of the preceding class).
   - Find the upper limit of the first class. Remember that classes cannot overlap.
   - Find the remaining upper class limits.
4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency f for each class.

Frequency Distributions and Their Graphs

Example: Constructing a Frequency Distribution

The following sample data set lists the number of minutes 50 Internet subscribers spent on the Internet during their most recent session. Construct a frequency distribution that has seven classes.

50 40 41 17 11 7 22 44 28 21 19 23 37 51 54 42 86
41 78 56 72 56 17 7 69 30 80 56 29 33 46 31 39 20
18 29 34 59 73 77 36 39 30 62 54 67 39 31 53 44

1. Number of classes = 7 (given)
2. Find the class width

$$\frac{\max - \min}{\#\text{classes}} = \frac{86-7}{7} \approx 11.29$$

Round up to 12

3. Use 7 (minimum value) as first lower limit. Add the class width of 12 to get the lower limit of the next class.

   7 + 12 = 19

   Find the remaining lower limits.

The upper limit of the first class is 18 (one less than the lower limit of the second class).

Class width = 12

Lower limit	Upper limit
7	
19	
31	
43	
55	
67	
79	

Frequency Distributions and Their Graphs

Example: Constructing a Frequency Distribution

4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency f for each class.

Determining the Midpoint

$$\frac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}$$

Class	Tally	Frequency, f
7 – 18	𝍸 I	6
19 – 30	𝍸 𝍸	10
31 – 42	𝍸 𝍸 III	13
43 – 54	𝍸 III	8
55 – 66	𝍸	5
67 – 78	𝍸 I	6
79 – 90	II	2

$\Sigma f = 50$

Class	Midpoint	Frequency, f
7 – 18	$\frac{7+18}{2} = 12.5$	6
19 – 30	$\frac{19+30}{2} = 24.5$	10
31 – 42	$\frac{31+42}{2} = 36.5$	13

Class width = 12

Frequency Distributions and Their Graphs

Example: Constructing a Frequency Distribution

Relative Frequency of a class
Portion or percentage of the data that falls in a particular class

$$\text{relative frequency} = \frac{\text{class frequency}}{\text{Sample size}} = \frac{f}{n}$$

Class	Frequency, f	Relative Frequency
7 – 18	6	$\frac{6}{50} = 0.12$
19 – 30	10	$\frac{10}{50} = 0.20$
31 – 42	13	$\frac{13}{50} = 0.26$

Cumulative frequency of a class: The sum of the frequency for that class and all previous classes.

Class	Frequency, f	Cumulative frequency
7 – 18	6	6
19 – 30	10	16
31 – 42	13	29

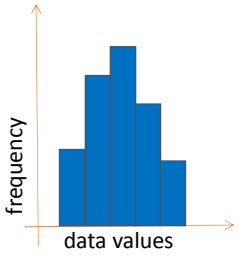
Class	Frequency	Midpoint	Relative frequency	Cumulative frequency
7 – 18	6	12.5	0.12	6
19 – 30	10	24.5	0.20	16
31 – 42	13	36.5	0.26	29
43 – 54	8	48.5	0.16	37
55 – 66	5	60.5	0.10	42
67 – 78	6	72.5	0.12	48
79 – 90	2	84.5	0.04	50

$\Sigma f = 50$ $\quad \Sigma \frac{f}{n} = 1$

Frequency Distributions and Their Graphs

Frequency Histogram

- A bar graph that represents the frequency distribution.
- The horizontal scale is quantitative and measures the data values.
- The vertical scale measures the frequencies of the classes.
- Consecutive bars must touch.



Class boundaries
The numbers that separate classes without forming gaps between them.

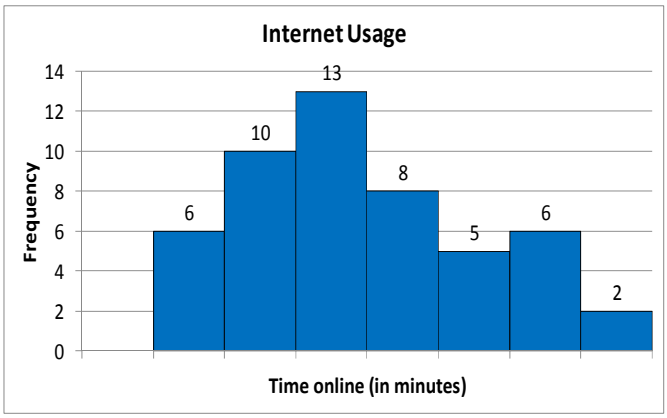
- The distance from the upper limit of the first class to the lower limit of the second class is 19 – 18 = 1.
- Half this distance is 0.5.
- First class lower boundary = 7 – 0.5 = 6.5
- First class upper boundary = 18 + 0.5 = 18.5

Class	Class Boundaries	Frequency, f
7 – 18	6.5 – 18.5	6
19 – 30		10
31 – 42		13

Frequency Distributions and Their Graphs

Example ctd. Class Boundaries

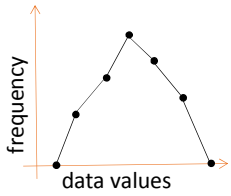
Class	Class boundaries	Midpoint	Frequency, f
7 – 18	6.5 – 18.5	12.5	6
19 – 30	18.5 – 30.5	24.5	10
31 – 42	30.5 – 42.5	36.5	13
43 – 54	42.5 – 54.5	48.5	8
55 – 66	54.5 – 66.5	60.5	5
67 – 78	66.5 – 78.5	72.5	6
79 – 90	78.5 – 90.5	84.5	2



Frequency Distributions and Their Graphs

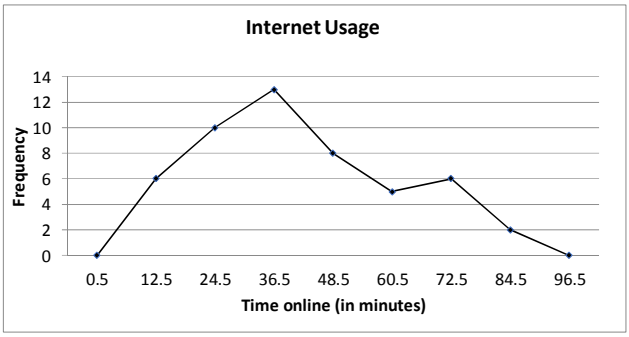
Frequency Polygon

- A line graph that emphasizes the continuous change in frequencies.



Construct a frequency polygon for the Internet usage frequency distribution.

Class	Midpoint	Frequency, f
7 – 18	12.5	6
19 – 30	24.5	10
31 – 42	36.5	13
43 – 54	48.5	8
55 – 66	60.5	5
67 – 78	72.5	6
79 – 90	84.5	2



The graph should begin and end on the horizontal axis, so extend the left side to one class width before the first class midpoint and extend the right side to one class width after the last class midpoint.

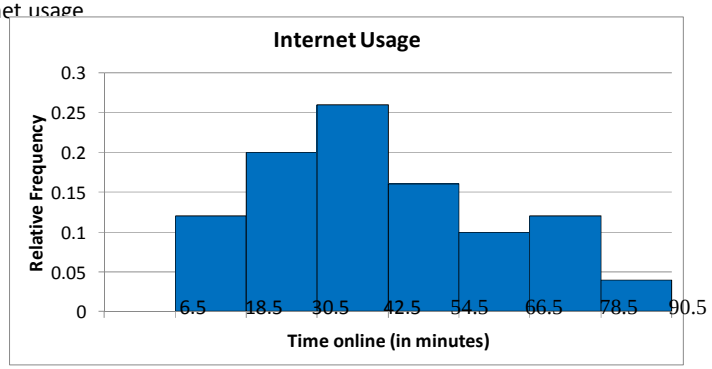
Frequency Distributions and Their Graphs

Relative Frequency Histogram

- Has the same shape and the same horizontal scale as the corresponding frequency histogram.
- The vertical scale measures the **relative frequencies**, not frequencies.

Construct a relative frequency histogram for the Internet usage frequency distribution.

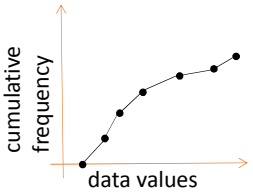
Class	Class boundaries	Frequency, f	Relative frequency
7 – 18	6.5 – 18.5	6	0.12
19 – 30	18.5 – 30.5	10	0.20
31 – 42	30.5 – 42.5	13	0.26
43 – 54	42.5 – 54.5	8	0.16
55 – 66	54.5 – 66.5	5	0.10
67 – 78	66.5 – 78.5	6	0.12
79 – 90	78.5 – 90.5	2	0.04



Frequency Distributions and Their Graphs

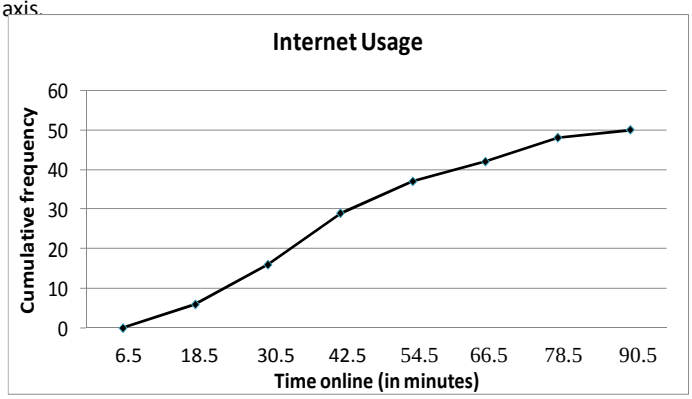
Cumulative Frequency Graph or Ogive

- A line graph that displays the cumulative frequency of each class at its upper class boundary.
- The upper boundaries are marked on the horizontal axis.
- The cumulative frequencies are marked on the vertical axis.



Construct an ogive for the Internet usage frequency distribution.

Class	Class boundaries	Frequency,	Cumulative frequency
7 – 18	6.5 – 18.5	6	6
19 – 30	18.5 – 30.5	10	16
31 – 42	30.5 – 42.5	13	29
43 – 54	42.5 – 54.5	8	37
55 – 66	54.5 – 66.5	5	42
67 – 78	66.5 – 78.5	6	48
79 – 90	78.5 – 90.5	2	50



What graphical representation provides

The distribution of a variable describes the values the variable takes and how often it takes each value

What are we looking for?

Lots of data distributed randomly

- Shape (symetric or not? Unimodal or skewed?)
- Center (where is mean, median or mode in this distribution)
- Spread
- Outliers

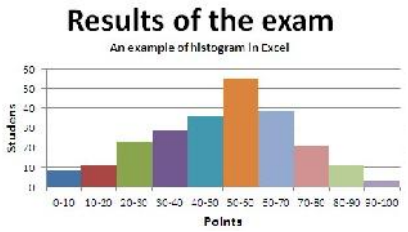
What are the altenatives?

- Stem-and-leaf plots
  a. Displays actual values of all observations
  b. Good for small amounts of data
- Histograms
  a. Displays only summary information
  b. Used for large amounts of data

Stemplot of Data Set

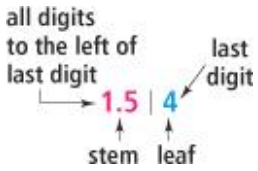
0	4 6
1	2 4 8
2	
3	3 4 4 5 5 7 8
4	2 2 5
5	0 1 8
6	8
7	2

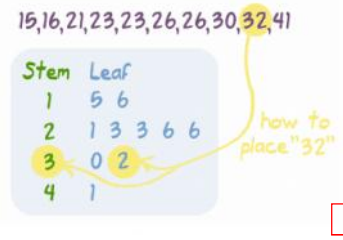
Key: 1|0 = 10



Stem and Leaf Plot

Stem and Leaf Basics





Decimal Between Stem and Leaf

12.3, 12.5, 13.0

Becomes

12	3, 5
13	0

Key: 12 | 3 = 12.3 units

Decimal in the Stem

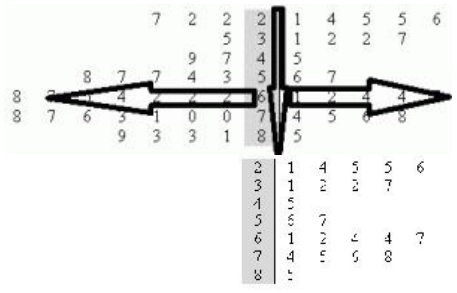
1.23, 1.25, 1.30

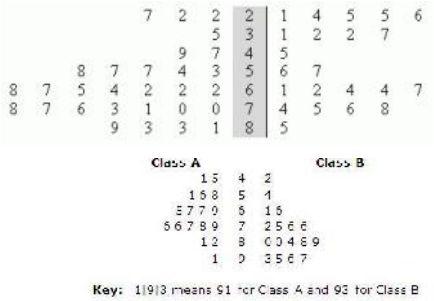
Becomes

1.2	3, 5
1.3	0

Key: 1.2 | 3 = 1.23 units

Back to back Stemplot





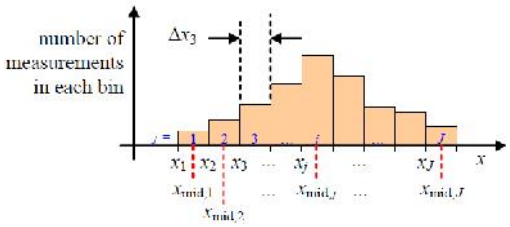
Histograms

Histogram

Histograms are bar graphs and height of each bar shows number of the individuals that has a value within a particular class.

- This particular class is define with a sub-range in distribution and named as bin and showed on x axis.
- The y axis " the height of the bin" shows the frequency of the indivuals in bins.

1. A histogram is constructed by dividing up the **n** measurements of a sample **into J bins** or intervals (also called classes)
   rule of thumb is 5-20 bins.
1. such that for the first bin (j = 1), $x_1 < x < x_2$, thesecond bin (j = 2), $x_2 < x < x_3$, etc.*x*mid,*j* is the middle value of x in bin j. For example,*x*mid,2= (*x*2 + *x*3)/2.
2. Afterwards, a bar plot is made of the frequency(also called the class frequency) which is the number of measurements in each bin.



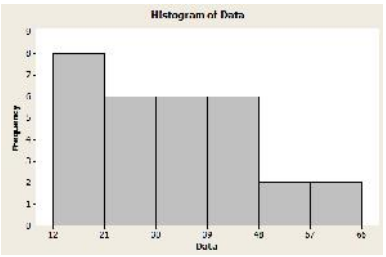
12	16	27	31	40	42
14	20	27	32	40	51
14	20	27	33	40	56
14	21	29	34	40	60
19	23	31	36	40	62

Range=65-12=53

Choose to have 6 groups

Bin width $\frac{53}{6} = 8.8$

round this up to **9**



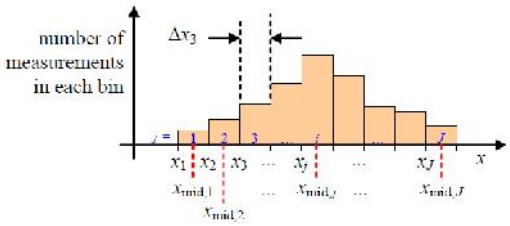
Histograms

Histogram

Histograms are bar graphs and height of each bar shows number of the individuals that has a value within a particular class.

- This particular class is define with a sub-range in distribution and named as bin and showed on x axis.
- The y axis " the height of the bin" shows the frequency of the indivuals in bins.

1. A histogram is constructed by dividing up the **n** measurements of a sample **into J bins** or intervals (also called classes)
   rule of thumb is 5-20 bins.
1. such that for the first bin (j = 1), $x_1 < x < x_2$, thesecond bin (j = 2), $x_2 < x < x_3$, etc.*x*mid,*j* is the middle value of x in bin j. For example,*x*mid,2= (*x*2 + *x*3)/2.
2. Afterwards, a bar plot is made of the frequency(also called the class frequency) which is the number of measurements in each bin.



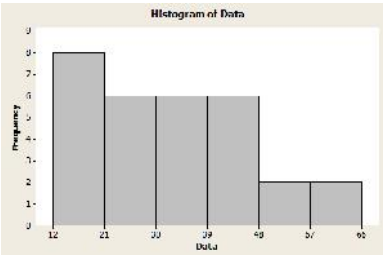
12	16	27	31	40	42
14	20	27	32	40	51
14	20	27	33	40	56
14	21	29	34	40	60
19	23	31	36	40	62

Range=65-12=53

Choose to have 6 groups

Bin width $\frac{53}{6} = 8.8$

round this up to **9**



IKCU\Mechatronics Engineering\MEE104 Measurement and Evaluation Techniques Dr.Levent ÇETİN

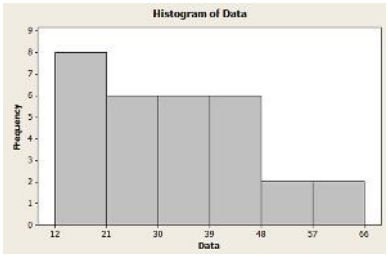
Histogram

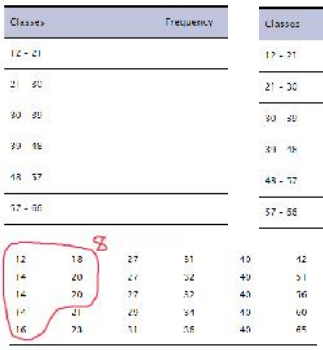
An example

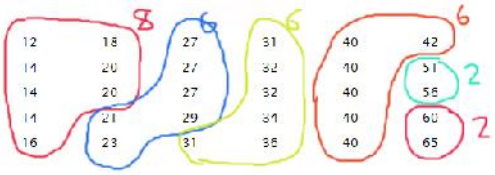
12	18	27	31	40	42
14	20	27	32	40	51
14	20	27	32	40	56
14	21	29	34	40	60
16	23	31	36	40	65

Bin width= 9 / Number of bins=6

Classes	Frequency
12 - 21	8
21 - 30	6
30 - 39	6
39 - 48	6
48 - 57	2
57 - 66	2



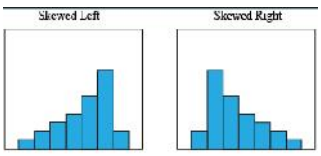




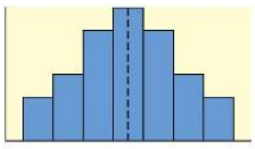
Histogram Shape

Skewness of Data

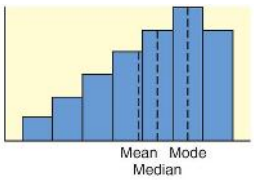
Sometimes the data is not symetric and this shows that the data is not purely random. The measure of the asymetry of the data is skewness.



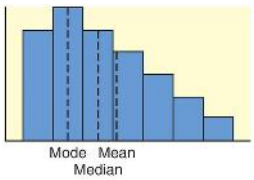
The mean, median and mode are all measures of the center of a set of data. The skewness of the data can be determined by how these quantities are related to one another.



In general, when a data distribution is mount-shaped symmetrical, the values for the mean, median, and mode are the same or almost the same.



For skewed-left distributions, the mean is less than the median and the median is less than the mode.



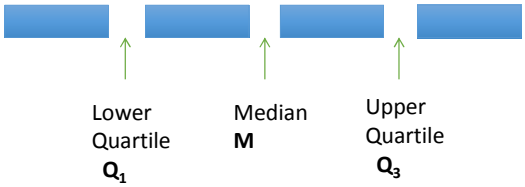
For skewed-right distributions, the mode is the smallest value, the median is the next largest, and the mean is the largest.

Box and whiskers plot

Another display that is helpful for reflecting properties of a sample is **the box and-whisker plot**. This plot encloses the interquartile range of the data in a box that has the median displayed within. The interquartile range has as its extremes the 75th percentile (upper quartile) and the 25th percentile (lower quartile).
In addition to the box, "whiskers" extend, showing extreme observations in the sample. For reasonably large samples, the display shows center of location, variability, and the degree of asymmetry.

Quartile

Quartiles in statistics are values that divide your data into quarters. A quartile divides a sorted (least to greatest) data set into 4 equal parts, so that each part represents ¼ of the data set.



25% of all the data has a value less than or equal to Q_1
50% of all the data has a value less than or equal to **M**
75% of all the data has a value less than or equal to Q_3
50% of all the data lies between Q_1 and Q_3

Example

3, 4, 5, 6, 6, 7, 8, 9, 9, 10, 11

- There are 11 data items
- The **median** is the 6th item. So **M=7**.
- The **lower quartile** is the 3rd item. (It is the middle of the lower half.) Q_1**=5**
- The **upper quartile** is the 9th item. (It is the middle of the upper half.) Q_3**=9**

Interquartile Range (IQR): The difference between the third and first quartiles. IQR = $Q_3 - Q_1$ =9-5=4

Box-and-whisker plot

Requires (**five-number summary**):

- Minimum entry
- First quartile Q_1
- Median Q_2
- Third quartile Q_3
- Maximum entry

Plot rules:

1. Construct a horizontal scale that spans the range of the data.
2. Plot the five numbers above the horizontal scale.
3. Draw a box above the horizontal scale from Q_1 to Q_3 and draw a vertical line in the box at Q_2.
4. Draw whiskers from the box to the minimum and maximum entries.

Example:
The test scores of 15 employees enrolled in a CPR training course are listed. Find the first, second, and third quartiles of the test scores.

13 9 18 15 14 21 7 10 11 20 5 18 37 16 17

Min = 5 Q1 = 10 Q2 = 15 Q3 = 18 Max = 37