Probability and Statistics

Slides Set 2

Cumulative Frequency

Cumulative Frequency Distribution
(less than)

Cumulative Frequency Distribution
(greater than)

Example:

Class Interval	Class Boundaries	Frequency (f)	C.f(<)	c.f(>)
30-39	29.5-39.5	3	3	40
40-49	39.5-49.5	1	4	37
50-59	49.5-59.5	8	12	36
60-69	59.5-69.5	10	22	28
70-79	69.5-79.5	7	29	18
80-89	79.5-89.5	7	36	11
90-99	89.5-99.5	4	40	4
		40		

Equal

Equal

Representation of Cumulative Frequency (<)

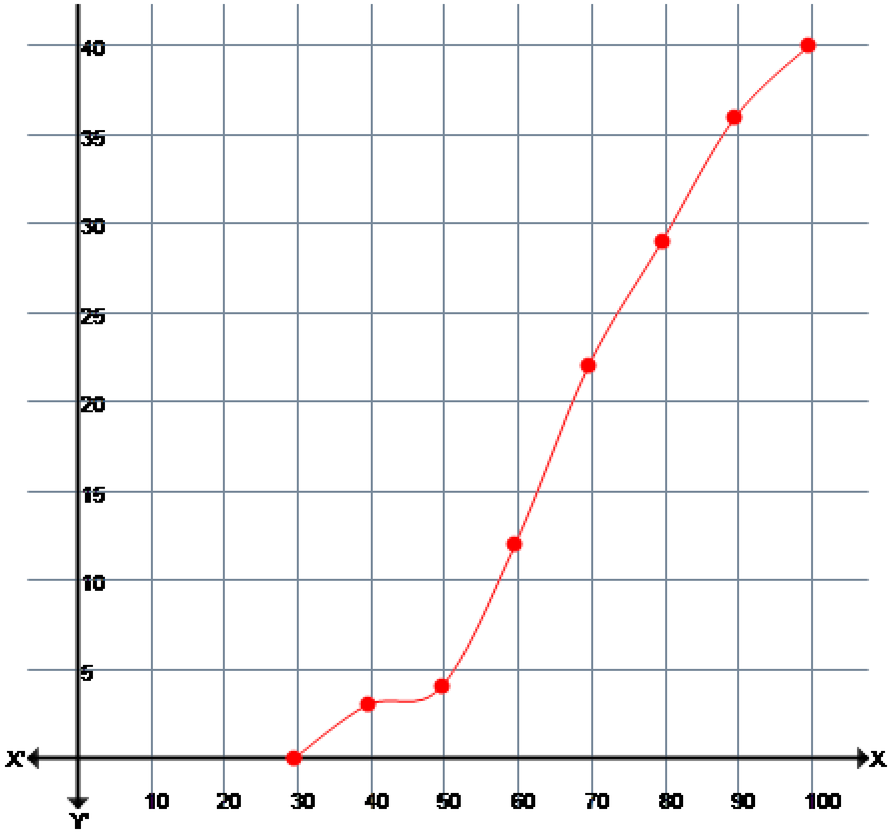
Upper Class Boundaries	C.f(<)
Less than 39.5	3
Less than 49.5	4
Less than 59.5	12
Less than 69.5	22
Less than 79.5	29
Less than 89.5	36
Less than 99.5	40

Representation of Cumulative Frequency (>)

lower Class Boundaries	c.f(>)
More than 29.5	40
More than 39.5	37
More than 49.5	36
More than 59.5	28
More than 69.5	18
More than 79.5	11
More than 89.5	4

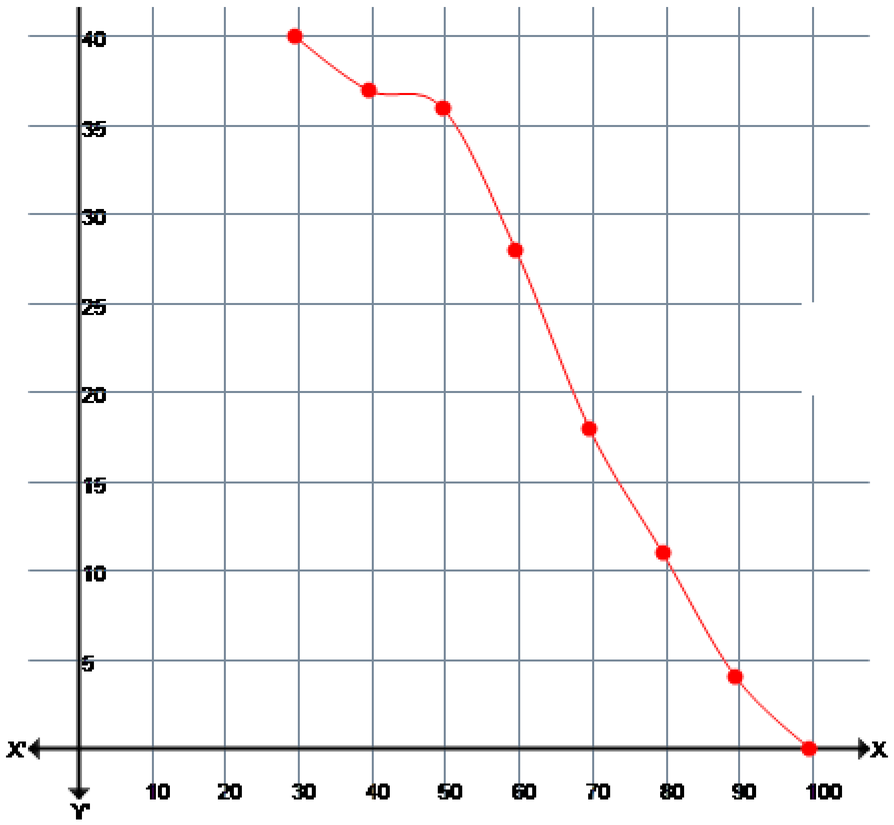
Ogive Curve (<)
(Drawn b/w U.C.B and c.f(<))

Upper Class Boundaries	C.f(<)
Less than 39.5	3
Less than 49.5	4
Less than 59.5	12
Less than 69.5	22
Less than 79.5	29
Less than 89.5	36
Less than 99.5	40

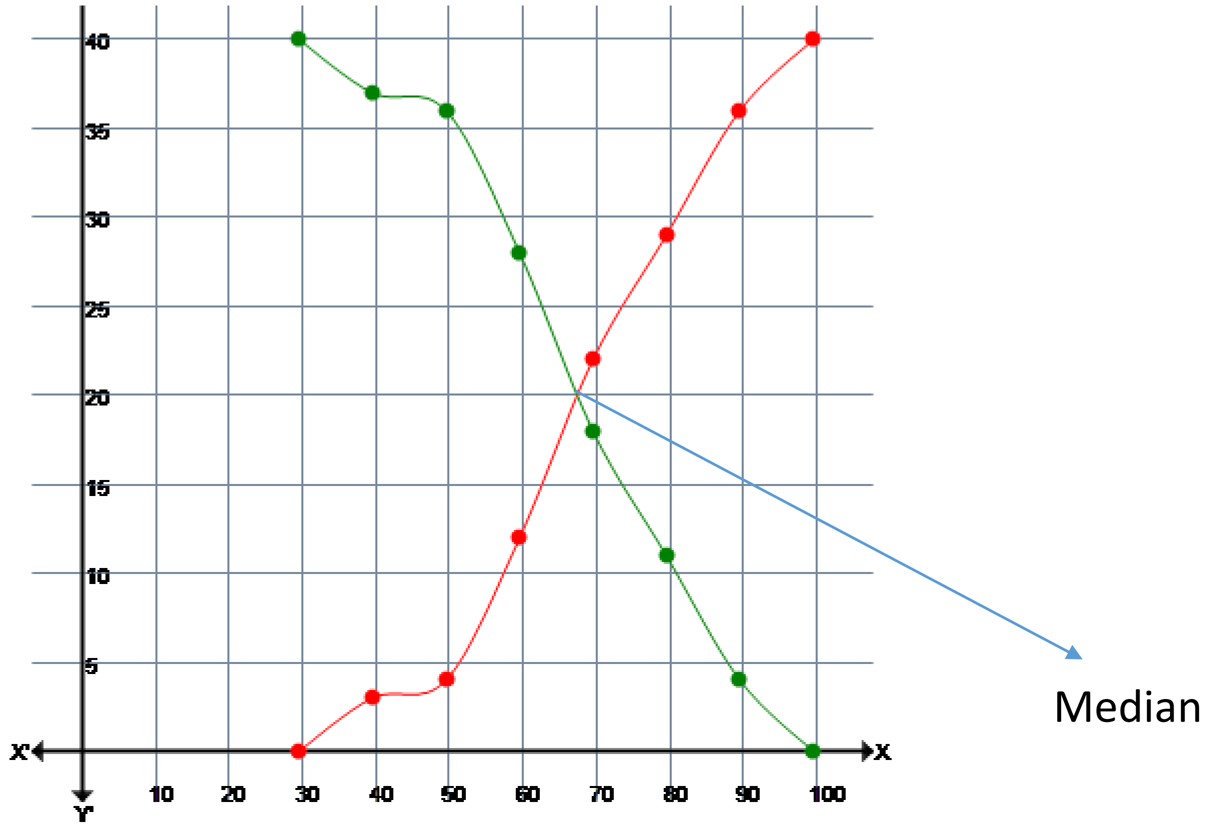


Ogive Curve (>)
(Drawn b/w L.C.B and c.f(>))

lower Class Boundaries	c.f(>)
More than 29.5	40
More than 39.5	37
More than 49.5	36
More than 59.5	28
More than 69.5	18
More than 79.5	11
More than 89.5	4



Ogive Curve (< and > combined)



Measure of Centers/Central Tendency

3.1 Measures of Center

Descriptive measures that indicate where the center or most typical value of a data set lies are called **measures of central tendency** or, more simply, **measures of center.** Measures of center are often called *averages.*

In this section, we discuss the three most important measures of center: the *mean, median,* and *mode.* The mean and median apply only to quantitative data, whereas the mode can be used with either quantitative or qualitative (categorical) data.

- Mean (ideal central tendency)
- Median
- Mode

The Mean

The most commonly used measure of center is the *mean*. When people speak of taking an average, they are most often referring to the mean.

Mean of a Data Set

The **mean** of a data set is the sum of the observations divided by the number of observations.

Example

The Mean

Weekly Salaries Professor Hassett spent one summer working for a small mathematical consulting firm. The firm employed a few senior consultants, who made

between \$800 and \$1050 per week; a few junior consultants, who made between \$400 and \$450 per week; and several clerical workers, who made \$300 per week.

The firm required more employees during the first half of the summer than the second half. Tables 3.1 and 3.2 list typical weekly earnings for the two halves of the summer. Find the mean of each of the two data sets.

TABLE 3.1
Data Set I

\$300	300	300	940	300
300	400	300	400	
450	800	450	1050	

TABLE 3.2
Data Set II

\$300	300	940	450	400
400	300	300	1050	300

Solution As we see from Table 3.1, Data Set I has 13 observations. The sum of those observations is \$6290, so

$$\text{Mean of Data Set I} = \frac{\$6290}{13} = \$483.85 \text{ (rounded to the nearest cent).}$$

Similarly,

$$\text{Mean of Data Set II} = \frac{\$4740}{10} = \$474.00.$$

Interpretation The employees who worked in the first half of the summer earned more, on average (a mean salary of \$483.85), than those who worked in the second half (a mean salary of \$474.00).

Example (Mean of single-value grouping type data)

x	f
0	11
1	23
2	3
3	12
4	10
5	11
6	21

Formula : mean= $\frac{\sum fx}{\sum f}$

x	f	fx
0	11	0
1	23	23
2	3	6
3	12	36
4	10	40
5	11	55
6	21	126
	91	286

$$\sum f$$

$$\sum fx$$

Formula : mean= $\frac{\sum fx}{\sum f} = \frac{286}{91} = 3.142$

Example (Mean of grouped data)

Class Interval	Frequency (f)
30-39	3
40-49	1
50-59	8
60-69	10
70-79	7
80-89	7
90-99	4
	40

Formula : mean$= \frac{\sum fx}{\sum f}$

Class Interval	Frequency (f)	midpoint (x)	fx
30-39	3	34.5	103.5
40-49	1	44.5	44.5
50-59	8	54.5	436
60-69	10	64.5	645
70-79	7	74.5	521.5
80-89	7	84.5	591.5
90-99	4	94.5	378
	40		2720

$$\sum f \qquad \sum fx$$

Formula : mean$= \frac{\sum fx}{\sum f} = \frac{2720}{40} = 68$

The Median

Another frequently used measure of center is the median. Essentially, the *median* of a data set is the number that divides the bottom 50% of the data from the top 50%. A more precise definition of the median follows.

Median of a Data Set

Arrange the data in increasing order.

- If the number of observations is odd, then the **median** is the observation exactly in the middle of the ordered list.
- If the number of observations is even, then the **median** is the mean of the two middle observations in the ordered list.

In both cases, if we let n denote the number of observations, then the median is at position $(n + 1)/2$ in the ordered list.

Example

The Median

Weekly Salaries Consider again the two sets of salary data shown in Tables 3.1 and 3.2. Determine the median of each of the two data sets.

TABLE 3.1
Data Set I

$300	300	300	940	300
300	400	300	400	
450	800	450	1050	

TABLE 3.2
Data Set II

$300	300	940	450	400
400	300	300	1050	300

Solution To find the median of Data Set I, we first arrange the data in increasing order:

300 300 300 300 300 300 **400** 400 450 450 800 940 1050

The number of observations is 13, so $(n + 1)/2 = (13 + 1)/2 = 7$. Consequently, the median is the seventh observation in the ordered list, which is 400 (shown in boldface).

To find the median of Data Set II, we first arrange the data in increasing order:

300 300 300 300 **300** **400** 400 450 940 1050

The number of observations is 10, so $(n + 1)/2 = (10 + 1)/2 = 5.5$. Consequently, the median is halfway between the fifth and sixth observations (shown in boldface) in the ordered list, which is 350.

Interpretation Again, the analysis shows that the employees who worked in the first half of the summer tended to earn more (a median salary of $400) than those who worked in the second half (a median salary of $350).

Example (Median of single-value grouping type data)

x	f
0	11
1	23
2	3
3	12
4	10
5	11
6	21
	91

$$\sum f$$

For finding median, compute $\frac{\sum f}{2} = \frac{91}{2} = 45.5^{th}\ value$

x	f	c.f(<)
0	11	11
1	23	34
2	3	37
3	12	49
4	10	59
5	11	70
6	21	91
	91	

Median class

Median

Example (Median of grouped data)

Class Interval	Class Boundaries	Frequency (f)
30-39	29.5-39.5	3
40-49	39.5-49.5	1
50-59	49.5-59.5	8
60-69	59.5-69.5	10
70-79	69.5-79.5	7
80-89	79.5-89.5	7
90-99	89.5-99.5	4
		40

Now, use formula median $= l + \frac{h}{f}\left(\frac{\sum f}{2} - c.f(<)\right)$

Where

$$l = L.C.B\ of median\ class$$
$$h = width\ of median\ class$$
$$f = frequency\ of median\ class$$
$$c, f(<) = c.f(<)\ of\ previous\ class\ to\ median\ class$$

For finding median, first compute $\frac{\sum f}{2} = \frac{40}{2} = 20^{th}\ \ value$

Class Interval	Class Boundaries	Frequency (f)	c.f(<)
30-39	29.5-39.5	3	3
40-49	39.5-49.5	1	4
50-59	49.5-59.5	8	12
60-69	59.5-69.5	10	22
70-79	69.5-79.5	7	29
80-89	79.5-89.5	7	36
90-99	89.5-99.5	4	40
		40	

Median class

Formula : median $= l + \frac{h}{f}\left(\frac{\sum f}{2} - c.f(<)\right)$

$$= 59.5 + \frac{10}{10}\left(\frac{40}{2} - 12\right)$$
$$= 67.5\ ans$$

The Mode

The final measure of center that we discuss here is the *mode*.

Mode of a Data Set

Find the frequency of each value in the data set.

- If no value occurs more than once, then the data set has *no mode*.
- Otherwise, any value that occurs with the greatest frequency is a **mode** of the data set.

Example

The Mode

Weekly Salaries Determine the mode(s) of each of the two sets of salary data given in Tables 3.1 and 3.2 on page 91.

TABLE 3.1
Data Set I

$300	300	300	940	300
300	400	300	400	
450	800	450	1050	

TABLE 3.2
Data Set II

$300	300	940	450	400
400	300	300	1050	300

Solution Referring to Table 3.1, we obtain the frequency of each value in Data Set I, as shown in Table 3.3. From Table 3.3, we see that the greatest frequency is 6, and that 300 is the only value that occurs with that frequency. So the mode is $300.

Proceeding in the same way, we find that, for Data Set II, the greatest frequency is 5 and that 300 is the only value that occurs with that frequency. So the mode is $300.

Interpretation The most frequent salary was $300 both for the employees who worked in the first half of the summer and those who worked in the second half.

Example (Mode of single-value grouping type data)

x	f
0	11
1	23
2	3
3	12
4	10
5	11
6	21
	91

For finding mode, Mark the class in which highest frequency occur

x	f
0	11
1	23
2	3
3	12
4	10
5	11
6	21
	91

Modal class

Mode

Example (Mode of grouped data)

Class Interval	Class Boundaries	Frequency (f)
30-39	29.5-39.5	3
40-49	39.5-49.5	1
50-59	49.5-59.5	8
60-69	59.5-69.5	10
70-79	69.5-79.5	7
80-89	79.5-89.5	7
90-99	89.5-99.5	4
		40

For finding mode, first mark the class in which highest frequency occur

Class Interval	Class Boundaries	Frequency (f)	c.f(<)
30-39	29.5-39.5	3	3
40-49	39.5-49.5	1	4
50-59	49.5-59.5	8	12
60-69	59.5-69.5	10	22
70-79	69.5-79.5	7	29
80-89	79.5-89.5	7	36
90-99	89.5-99.5	4	40
		40	

Modal class

Now, use formula mode $= l + h \times \frac{(f_m - f_1)}{2f_m - f_1 - f_2}$

Where

$$l = L.C.B\ of modal\ class$$
$$h = width\ of modal\ class$$
$$f_m = frequency\ of modal\ class$$
$$f_1 = frequency\ of\ previous\ class\ to\ modal\ class$$
$$f_2 = frequency\ of\ next\ class\ to\ modal\ class$$

Formula : mode$= l + h \times \frac{(f_m - f_1)}{2f_m - f_1 - f_2}$

$$= 59.5 + 10\left(\frac{10-8}{2(10)-8-7}\right)$$
$$= 63.5\ ans$$