

Ergonomics for Beginners Industrial Design Perspective
Prof. D. Chakrabarti
Department of Design
Indian Institute of Technology, Guwahati

Module No. # 03

Human physical dimension concern

Lecture No. # 14

Statistical Treatment of Data and Percentile Calculations

(Refer Slide Time: 00:30)

Ergonomics for beginners Industrial design perspective		
Modules	Area of discussion	No. of classes
Module 1	Introducing Ergonomics and content details	2
Module 2	Discipline approach Ergonomics/ Human Factors	5
Module 3 Human physical dimension concern 7		
Module 4	Posture and movement	8
Module 5	Behaviour and perception	5
Module 6	Visual issues	2
Module 7	Environments Factors	1
Module 8	Ergonomic design process	4
Module 9	Performance support and design intervention	5
Module 10	Design Ergonomics in India: scope for exploration	1

Welcome to this 14th session of ergonomics for beginners industrial design perspective. Now, today, the last class of module number 3, that is human physical dimension concerns.

(Refer Slide Time: 00:38)

Module 3 Human physical dimension concern 7

Today's session in this module

- Class 8 Human body- structure and function, anthropometrics
- Class 9 Anthropometry: body growth and somatotypes
- Class 10 Anthropometry landmark- Static: stand postures
- Class 11 Anthropometry landmark- Sitting postures
- Class 12 Anthropometry: squatting and cross-legged postures
- Class 13 Measuring technique

Class 14 Statistical treatment of data and percentile calculations

(Refer Slide Time: 00:53)

Anthropometric measuring techniques

Last sessions' brief

A. Measuring circumference with the help of a measuring tape

B. Representative anthropometric measuring technique using anthropometric board, an anthropometric board, in standing posture

C. Representative anthropometric measuring technique using anthropometric rod set, in standing posture

We have discussed the direct measurement procedures using anthropometric rod set, anthropometric board, and specific measuring devices as suitable to measure some dimensions like grip, foot data etc.

That is the today's session of this module will be class 14 that is statistical treatment of data and percentile calculations. In last class, we have discussed that the anthropometric measuring techniques, and before that anthropometric landmarks were discussed. This is specifically required, when we do not have any ready reference data; for design application we may need to generate some data, and how to generate the data? Last class we have discussed. And now, we will be discussing how the statistical treatment can be done, on those collected data.

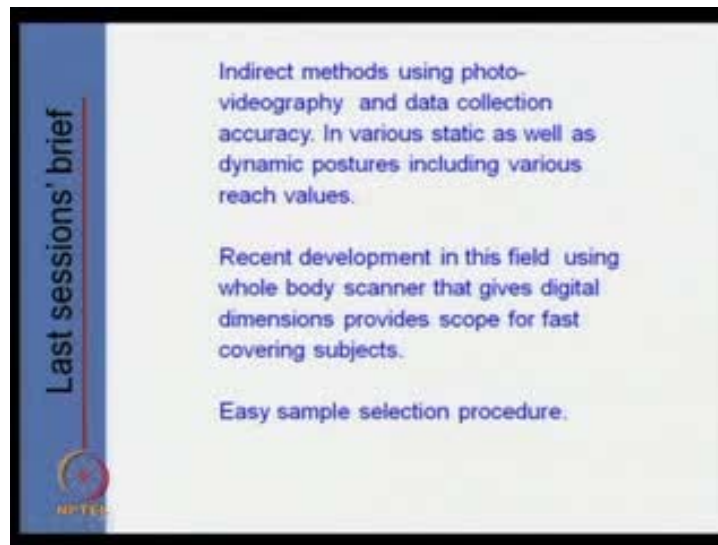
Now, in last class, we have mentioned that the basic anthropometric techniques, in using anthropometric rod set, and other equipment's like a height adjustable tools weight, body weight, and cone for grip diameter, and etcetera. And measuring circumferences with the help of inextensible or non-elastic measuring tape, and then using different grids that is that anthropometric measuring boards are used that can be developed that can be done.

(Refer Slide Time: 02:53)



Now, we have discussed a brief measurement procedure using anthropometric rod set, anthropometric board, and specific measuring devices, as suitable to measure some dimensions like grip, foot etcetera also. Now, that was basically we discussed for standing measurements, sitting that is seat desk posture measurement; and sitting on floor posture that is cross legged squatting; considering static values as well as various dynamic reach values, we have discussed in last classes.

(Refer Slide Time: 03:24)



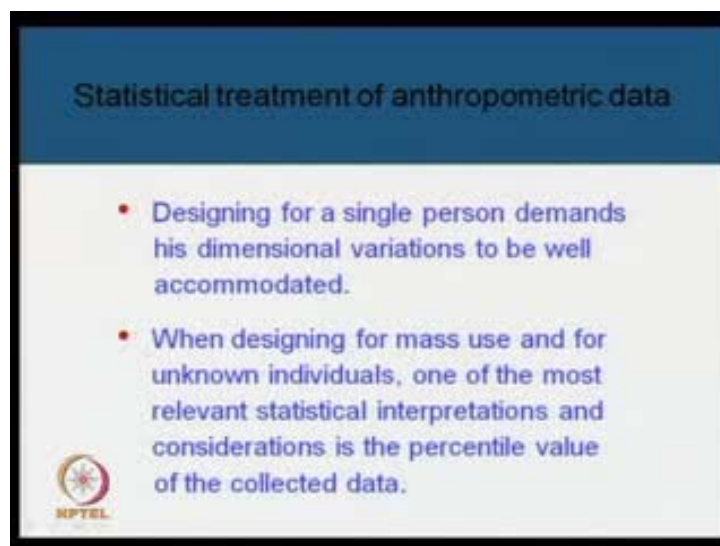
So, now, today and also we have discussed the indirect methods using photography and videography techniques and data collection accuracy system. How to make the data accurate? In various static as well as dynamic postures including various reach values, we have discussed. And then recent development in this field, using whole body scanner that gives digital dimensions, that provides us a scope for fast, covering many subjects; that is for the or anthropometric general survey, it would be very good.

And also we have discussed the easy sampling selection procedure, that is, that if we do not have a very large number of population, and we need to rely on some small samples; then it would be better to take around 20 percent eye estimated smaller size people, 20 percent eye estimated larger size of people, and then rest 60 percent randomly selected; so that, we can make a the both **the** ends of that population, from smaller dimensions to largest dimension, we can consider, we can cover it.

(Refer Slide Time: 05:06)



(Refer Slide Time: 05:49)

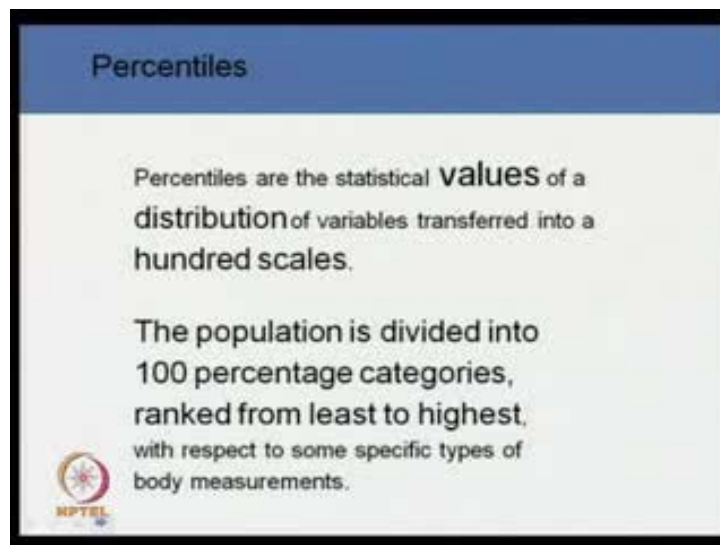


So, with all this the today's topic is the statistical analysis of anthropometric data and the percentile value, means, that when we design something for a single persons use, then we can measure him directly; and according to those measurements, we can make a design for his use, that is exclusively for his use say - a chair, a table or a bed, like that; **but when**, which we can say that designing for a single person demands his dimensional variations to be well accommodated in a design. But when designing for a mass use and for unknown individuals use, which we do not know who will be the intended user, one of the most relevant statistical interpretations and considerations in this would be the

percentile value of the collected data; means, when we are making a chair, then the popliteal height, means, the opposite side of the knee that is inner thigh that inner that portion of knee joint or popliteal height.

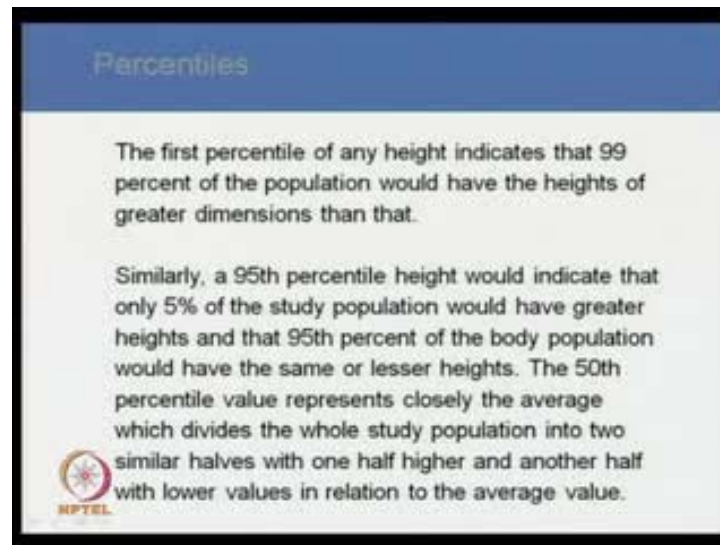
So, we can say that the chair height should be at per popliteal height, but this popliteal height from whom we get? A short height people, large height people or an average height people; so that, most of the people falling within this means small size to largest size everybody can get comfort.

(Refer Slide Time: 07:39)



So, in this situation we need a specific value that is that percentile value of the body dimensions. Now, what is the percentile value, and how to calculate, and etcetera that we will discuss now.

(Refer Slide Time: 08:15)



Now, what is a percentile? The percentiles are the statistical values of a distribution of variables transferred into a hundred scales. The population is divided into 100 percentage categories, ranked from least to highest, with respect to some specific types of body measurements. Now, the first percentile of any height indicates that 99 percent of the population would have the heights greater dimensions than, that means, if we measure some people say 1000 people, we measure their heights; now, the minimum height or the lowest height what we observed? If we consider it as a one in a hundred scale, one and the highest value is the hundred in that scale, and rest all the data are arranged within this one to hundred scale then we can say that **the** what would be **the one person**, if we consider make a door with one percentile height or **the** this population means what is happen? The only one percent of this population, those height is within that limit will be able to pass through without any problem.

Now, if we consider **the nine**, the 50 percentile means the total population will be divided into two halves; **means**, 50 percentile means below this 50 percentile around 50 percent of the body height those who have lower height than that they will feel no problem, they can pass through that door, but those who have higher value than that they will have some problem, they may need to bend their heads to cross that door.

If we see that 99th percentile data of that height, means, what is happen? 99 percent of the population will have lower height than that only 1 percent, means, 99 above till 100; only 1 percent population will be on the higher side means over that data.

Now, one question comes normally for design use, **in many**, everywhere, it is said that - consider the data from 5th percentile to 95th percentile, why not 1 percentile? Why not 100th percentile always? It is said that - from below the 5th percentile value these are almost free data, and above that height, above the 95th percentile height is the tallest kind of thing, that is also free data. So, suppose we want to make a bed for hostel purpose hostel cot, for that if we measure all the students, then few students may have very short height, and few student may have very tall height. Now, if we say that - we should make a cot where all the students can sleep without any problem; then ideally we can say that - take the 100th percentile value with the tallest person's height of that population, we measured and make a cot; so that, everybody falling below this height, they will not have any problem. Now, what is happen is from 95th to 100? This very few people will be there, so if we make all the cots with 100 percentiles height data, then there will be wastage of material resource; so, for that this 95 percentile is normally considered as the higher value, and the rest, if it is required to ensure safety kind of thing, then the maximum thing is necessary, and then allowance maybe given like.

If we do not want someone should touch a top on overhead hanging some object, no one should touch, at that very time, we should consider the upper maximum arm reaches, maximum length is the 100 percentile length kind of thing; so that, below that height or everybody all the people, they will not get touching to that dangerously hanging thing, in that and then, someone may raise their to, and can extend their arm reach, and can reach it; in that cases, some extra allowances to be added in this dimension.

So, these are the some of the considerations that we need to fit; another thing, suppose in one, in certain cases everybody should get touch, to operate some handle or switch like that, like any control over for any control operation, in that case, it is that everybody should get touch to that, in that case, we will try to consider the lower value of this arm dimension; so that, higher person, higher valued people they will not have any problem. So, in these cases, we should use the lower percentile value, normally it maybe 5th percentile; why? Now, below the 5th percentile very less population will be there,


representation will be there, and they may touch with extension of hand means considering the peak data.

(Refer Slide Time: 16:02)

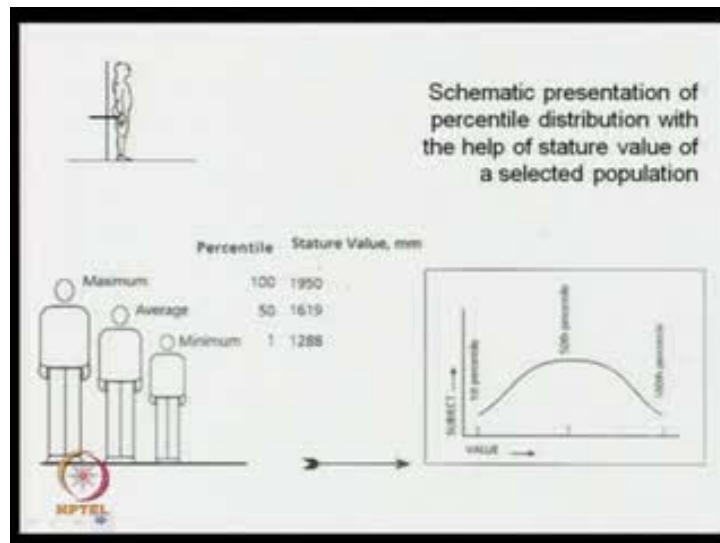
Percentiles

The first percentile of any height indicates that 99 percent of the population would have the heights of greater dimensions than that.

Similarly, a 95th percentile height would indicate that only 5% of the study population would have greater heights and that 95th percent of the body population would have the same or lesser heights. The 50th percentile value represents closely the average which divides the whole study population into two similar halves with one half higher and another half with lower values in relation to the average value.



(Refer Slide Time: 16:48)



So, these are the some common application for percentile value; now, how to calculate this percentile value from a raw data that we are going to discuss. Now, let us see in this slide whatever we discussed, let us see the slides for specifically is mentioned here for reference. The first percentile of any height indicates that 99 percent of the population would have the heights of greater dimensions than that. Similarly, a 95th percentile

height would indicate that only 5 percent of the study population would have greater heights, and that 95th percent of the body population would have the same or lesser heights. The 50th percentile value represents closely the average which divides the whole study population into two similar halves with one half higher and another half with lower values in relation to the average value; that is the 50th percentile value.

(Refer Slide Time: 17:07)

Original data collected, in mm:									
1877	1635	1674	1710	1530	1584	1747	1634	1560	1398
1596	1716	1396*	1494	1603	1490	1735	1619	1660	1623
1803	1840	1561	1585	1670	1832	1690	1640	1911	1564
1444	1470	1437	1541	1641	1475	1621	1711	1632	1613
1770	1710	1737	1656	1681	1516	1673	1628	1600	1940**

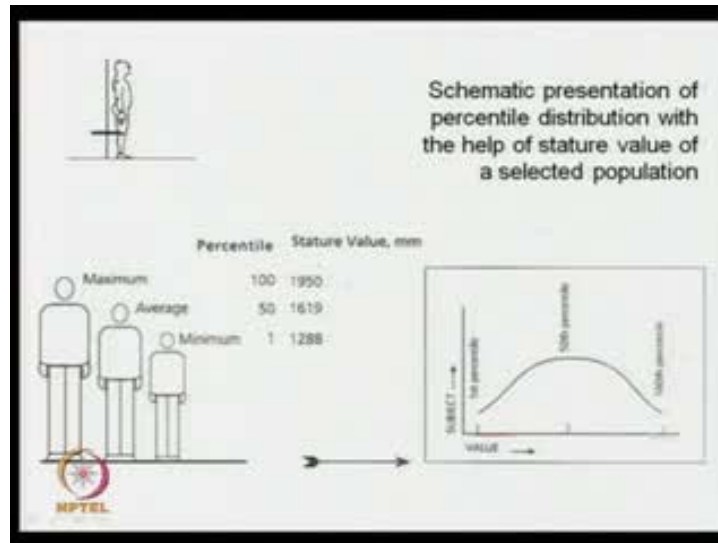
Stage I Identification of two end values.
 The lowest score* is 1396 and the highest score** is 1940

Now, see **the** this schematic presentation of percentile distribution, with the help of stature value of a selected population. Now, suppose if we measure, if we see this is a diagram, some data, 5th body height data, we have taken here for discussion purpose; these are the data original data collected in millimeter.

Now, here the minimum value is that 1396 millimeter - total body height, and maximum value we got suppose 1940 millimeter; so, 50 percent's we have measured, and these are **the** their data.

Now, suppose **the** 1 percentile, suppose we did that 1288 millimeter like this, this is a minimum per height; then 50th percentile is closely average or if our sample is proper then average and 50th percentile will match together, otherwise it is not the same if our sample size is not proper.

(Refer Slide Time: 18:42)



So, average how much 1619 millimeter, and the maximum is the 100 percentile is 1950 millimeter. Now, if we transfer this, in this diagram, in this graph, that suppose, this is the value in millimeter, so what means, and this **the** subjects; so, 1 percent data is here, 50th percentile here, and 100 percent like this, so the total graph is coming like this; if it is possible, this total bell shape curve is possible if our sample size is proper.

(Refer Slide Time: 19:07)

Statistical analysis of data for percentile calculation

- The calculation process is described here is an example, taking the stature values of a group of 50 individuals.
- This might help designers in the compilation of special data collected by them for special purposes, where ready sources are not available.
- It will also help them understand the method.

Below is a reference data set to get acquainted with the calculation procedure.

Now, the calculations: the calculation process is described here is an example taking, the stature values of body height of a group of 50 individuals. This might help designers in

the compilation of special data collection collected by them for special purposes, where ready sources are not available. It will also help them understand the method. Below is a reference data of data set to get acquainted with the calculation process.

(Refer Slide Time: 19:52)

Original data collected, in mm:

1877	1635	1674	1710	1530	1584	1747	1634	1560	1398
1596	1716	1396*	1494	1603	1490	1735	1619	1660	1623
1803	1840	1561	1585	1670	1832	1690	1640	1911	1564
1444	1470	1437	1541	1641	1475	1621	1711	1632	1613
1770	1710	1737	1656	1681	1516	1673	1628	1600	1940**

Stage I Identification of two end values.
The lowest score* is 1396 and the highest score** is 1940

(Refer Slide Time: 20:40)

Stage II Arrangement of classes

The difference between the highest and the lowest value, $1940 - 1396 = 544$ may be divided into manageable blocks, say if 50 individuals are in this group then, $544 / 50 = 10.88$ - 11 class of each class interval having 50 scores.

For ease of calculation a round figure of 350 instead of 1396 may be considered as the lowest value of the lower class.

Starting from 1350 with class intervals of 50, classes should be arranged till the uppermost class attends the highest score obtained in the original data.

Here we get 12 classes as the following class distribution.

The stage-I: first stage is that to identification of the two end values, that is the lowest score is this one, and the highest score is this one, for stage-I. Stage-II is that arrangement of classes, means, in from, this lowest value and highest value if we arrange, then total 5th value; now, from the highest value to lowest value, if we minus it

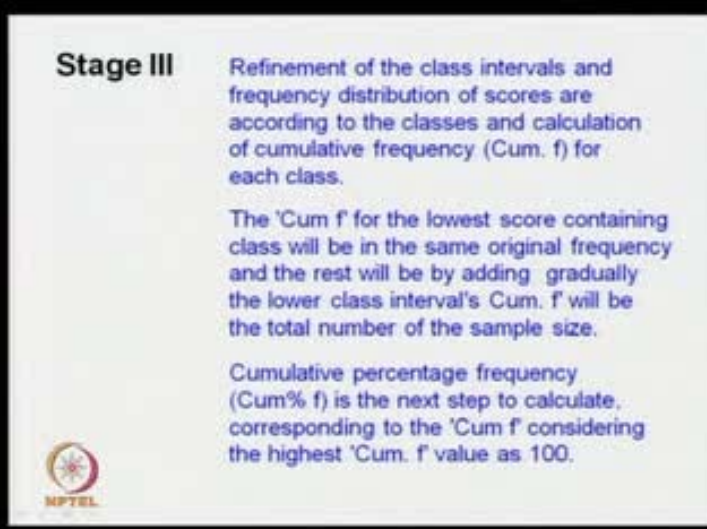
from the highest value to lowest value, it comes 544 millimeter, maybe divided into manageable blocks say - if 50 individuals are in this group, then we can say that to from highest to lowest, these much value is in between divided by 50, so 10.88 classes we can have, means, we can say that 11, class of 11 of each class interval having 50 scores.

(Refer Slide Time: 22:10)

Classification of groups with each class intervals of 50 scores	
1900-1949	Highest class interval containing the highest score 1940
1850-1899	
1800-1849	
1750-1799	
1700-1749	
1650-1699	
1600-1649	
1550-1599	
1500-1549	
1450-1499	
1400-1449	
1350-1399	Lowest class interval containing the lowest score 1396

For ease of calculation, a round figure of 350 instead of 1396, 1396 instead that to make it a round figure we are considering 1350 instead of 1396, means, 1350 we have taken it as a lower class. So, starting from 1350 with class intervals of 50, classes should be arranged till the uppermost class attend, the highest score obtained in the original data; here, we get 12 classes as the following class distribution like this, so minimum we are taking 1350, and 50 data, then next another 50, another 50, another 50, like this way we are crossing like this; so, like this way the topmost grade class 1900, 1949, in this the highest value,, whatever we are originally got 1940 lies in this, and the lowest value that is 1396 lies here. So, this is called calculation of groups with each class interval of 50 scores, we have considered arbitrarily these 50 scores.

(Refer Slide Time: 22:59)




Stage III

Refinement of the class intervals and frequency distribution of scores are according to the classes and calculation of cumulative frequency (Cum. f) for each class.

The 'Cum f for the lowest score containing class will be in the same original frequency and the rest will be by adding gradually the lower class interval's Cum. f will be the total number of the sample size.

Cumulative percentage frequency (Cum% f) is the next step to calculate, corresponding to the 'Cum f considering the highest 'Cum. f value as 100.



Now, the stage number 3: refinement of the class intervals, what is the refinement of class interval? That is that refinement of the class intervals and frequency distribution of scores are according to the classes and calculations of calculative or cumulative frequency - that is **Cum** Cum f - for each class. The cumulative frequency for the lowest score containing class will be the same original frequency and the rest will be by adding gradually the lower class interval's cumulative frequency will be the total number of the sample size. Cumulative percentage frequency is the next step to calculate, corresponding to the cumulative frequency considering the highest cumulative frequency value as 100. Now, we see the graph.

(Refer Slide Time: 24:02)

Show here, what this is, the class interval number 1, class interval 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, so total 12 class intervals; so, within this class interval, how many scores we have? How many tally? So, here if we go back to the original class, from here we can say that there will be 2, from this we can find out that, 2 scores are falling within this. So, like this we have to consider that here 2 frequencies tally are here; so, counting this how many scores are falling within this 2, here a 4, in this 3, like that we found it, the tally number; means, the original scores falling within this classes.

Now, means the frequency in these groups are 2, 2, 4, 3, 6, 1, 2, 7, 7, 1, 3, 2, 1, 2 like that; now, the cumulative frequency is that, suppose in this group, in this class, only 2 are there, in this class 2, but if we consider that 1, 4, 4, 9, this value, below this value, total 4 are there, this 2 from the below and 2 of this class so total 4, so this 4 will be below this; so, this is called by adding one after another, it is the cumulative frequency like that the total 50 here will be below this. Now, this is the cumulative frequency by adding one after another, means, then this we have considered 50 percentile, this 50 individuals.

(Refer Slide Time: 27:19)


Stage IV If we look at the bar diagram of this distribution, then if any individual score is between 1399 and 1400, it is not counted either in class 1st or in 2nd.

To accommodate this in the proper group, the mid-1 point between the upper value of the lower class and lower value of the next higher class may be considered as the true limit of these two class intervals.

Hence, the true lowest value of the class 1400 - 1449, would be 1349.5.

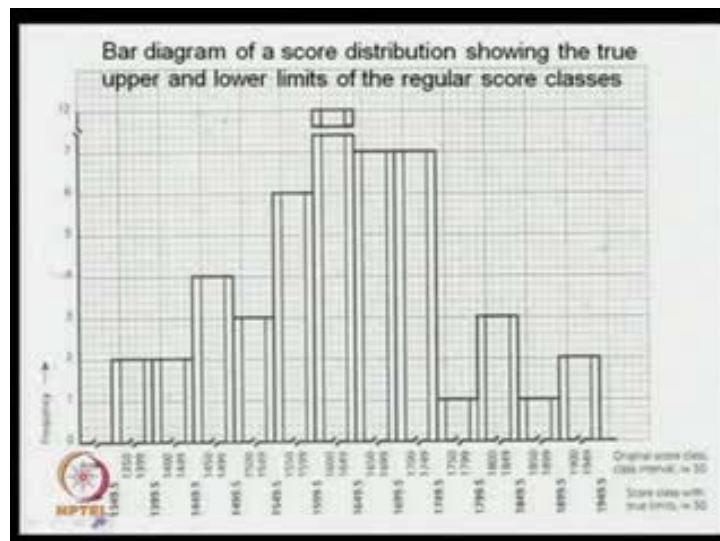
The same would also be the true highest value of the class 1350 - 1399.

If there is no possibility of its happening in manner or if that minute consideration is not required, this step may be avoided in calculation.



Now, the 50 data we got it now for percentile, we are transferring this to a 100 scale; 100 scales, means, what is happen? Then if the top it should be 100 means, in this case, it is just 2 multiplications. So, this cumulative percentage frequency would be 50 will be 100, 40 into 2 is the 96, like that, so 47 into 2 is 94, like that maybe 11 into 2 is 22 like that way, 2 into 2 equal to 4 something like that, this is the cumulative frequency; from this cumulative frequency and this is our main calculation starts.

(Refer Slide Time: 27:32)



Now, if we look at the bar diagram next; this is the examples, now if we see the bar diagram here, then what it says? Now, our class intervals where 1350 to 1399, next class interval was 1400 to 1449 like that. Now, if we see this one, then we can say that in... Now, we are making these classes in a graph.

Now, in this is the graph where the lowest value is 1350 and highest is that 1390; so, next group starts from 1400 and higher value of this class is 1449, but if we make a graph, then there this portion remains blank which data from here 1350 to 1449, this portion remains vacant for that purpose; like this way, it is distributed for that a specific case has been considered that, consider from what is the gap here? In this case, from 1399 to 1400 is there, so divide it by 2, because it is this for this dimension, one-dimension one value is divide distributed between these 2; so, if we divide by 2 that is 0.5. So, this specific group if we start 0.5 below, means, the score class it starts with the true limit of this group is, the true limit will be 1349.5 instead of 1350, and upper value of this group would be 1399.5, means, adding 0.5. The next one, just below minus 0.5 it will be 1399.5 instead of 1400; means, this point is being shared by both the groups. So, what is happen? All the groups like I said, then there will be a fixed, so no gaps will be there.

(Refer Slide Time: 30:41)

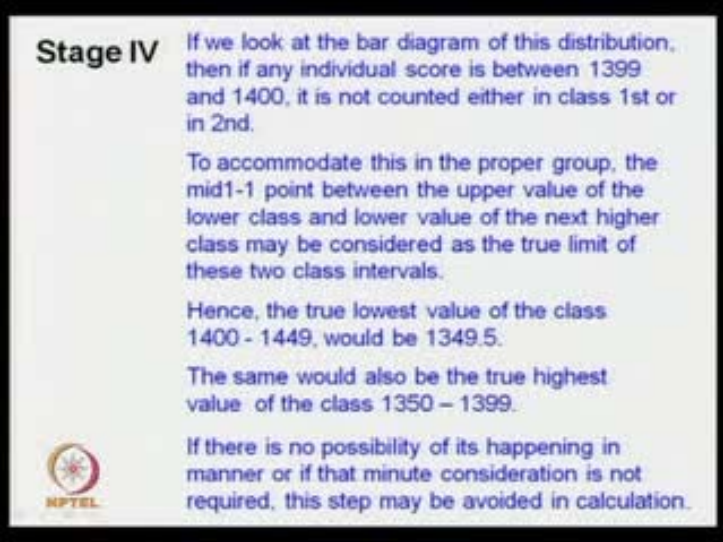
Stage V: The calculation

True lowest values	Classes with class interval, $i = 50$	True highest values	Frequency, f	Cum. f .	Cum. % f .
1899.5	1900-1949	1949.5	2	50	100
1849.5	1850-1899	1899.5	1	48	96
1799.5	1800-1849	1849.5	3	47	94
1749.5	1750-1799	1799.5	1	44	88
1699.5	1700-1749	1749.5	2	43	86
1649.5	1650-1699	1699.5	2	36	72
1599.5	1600-1649	1649.5	12	29	58
1549.5	1550-1599	1599.5	6	17	34
1499.5	1500-1549	1549.5	3	11	22
1449.5	1450-1499	1499.5	4	8	16
1399.5	1400-1449	1449.5	2	4	8
1349.5	1350-1399	1399.5	2	2	4
			N = 50		

So, the true value of the classes is 1349.5 to 1399.5, the next 1449 like that, and the scores are there, as per within this 2 scores, within this 4 scores, within this 6 scores like

this; after this, the main calculation part comes, the calculation is that we have to that we see here; so, these were the original group class, we mentioned.

(Refer Slide Time: 31:05)




Stage IV If we look at the bar diagram of this distribution, then if any individual score is between 1399 and 1400, it is not counted either in class 1st or in 2nd.

To accommodate this in the proper group, the mid-1 point between the upper value of the lower class and lower value of the next higher class may be considered as the true limit of these two class intervals.

Hence, the true lowest value of the class 1400 - 1449, would be 1349.5.

The same would also be the true highest value of the class 1350 - 1399.

If there is no possibility of its happening in manner or if that minute consideration is not required, this step may be avoided in calculation.



Now, if we from or this if we read this slide what it says - if we look at the bar diagram that we have seen right now or the distribution, then if any individual score is between 1399 and 1400, it is not counted either in class 1st or in 2nd. To accommodate this in the proper group, the mid value that 1 - 1 point between the upper value of the lower class and lower value of the next higher class maybe considered as the true limit of these two class intervals; that just now we have seen through this graph.


Hence, the true lowest value of the class second 1400 - 1449, would be 1349.5, like this; the same value also would be true highest value of this class. If there is no possibility of its happening in manner or if that minute consideration is not required, these step maybe avoided in percentile calculation also.

(Refer Slide Time: 32:59)

Stage IV

To sort out the true lower value and higher value of any class, this would be 0.5 less than the original lower value and 0.5 more than the original higher value.

By this it could be said that an individual score of 1399.4 would be counted in the true class of 1349.5, and 1399.7 would be in the class of 1399.5 - 1449.5.




To sort out the true lower value and higher value of any class, this could be 0.5 less than the original lower value and 0.5 more than the original higher value, is the true value. By this it could be said that an individual score of 1399.4 would be continuing, will be counted in the true class of 1349.5, and 1399.7 would be in the class of this.

(Refer Slide Time: 33:48)

Stage V: The calculation

True lowest values	Classes with class interval, $i = 50$	True highest values	Frequency, f	Cum. f	Cum. % f
1899.5	1800-1849	1949.5	2	50	100
1849.5	1750-1799	1899.5	1	48	96
1799.5	1700-1749	1849.5	3	47	94
1749.5	1750-1799	1799.5	1	44	88
1699.5	1700-1749	1749.5	2	43	86
1649.5	1650-1699	1699.5	2	36	72
1599.5	1600-1649	1649.5	12	29	58
1549.5	1550-1599	1599.5	8	17	34
1499.5	1500-1549	1549.5	3	11	22
1449.5	1450-1499	1499.5	4	8	16
1399.5	1400-1449	1449.5	2	4	8
1349.5	1350-1399	1399.5	2	2	4

$N = 50$



Now, the calculation, first I will discuss this table, and then we will go for **the** its description. Now, the thing is that, this is the original class minus 0.5 plus 0.5, this is the true value like this, is a true value, it is a true higher value; this is the original frequency

as we have seen earlier to in these class 2, 2, 4 like this way; cumulative frequency the same thing, now whatever we got in adding this 2 plus 2 equal 2 plus 0, if this a to the 4, 2 plus 2 equal to 4 like this way, and 4 plus 4 equal to 8 like that, by this way adding we are coming to 50; now with this 50, this is the cumulative frequency is there.

Now, we would like to see, suppose i , we want to measure or to calculate 25th percentile value, what will be the 25th percentile value of this 50 percents data? We can say that 50, 20 percentile, so with this cumulative frequency the 25 is falling within this, means, below this true limit there are 22 percent people, above this there will be other people, and 25th are within this true value somewhere. How to find i ? It like, still 22, this is the upper value, then from this value to this value, what are 50 scores are there, originally 6 percents are here, frequency is 6; so, 50 by 6.

(Refer Slide Time: 36:50)

Percentile calculation by arithmetic process

P25, the 25th percentile value corresponding to 'Cum% f falls in the class of 1550 -1599, i.e. the true class of 1549.5 - 1599.5 and correspondingly to the 'Cum% f, 25th percent position of the cum f, will be the 25% of the total N, $50 = 12.5$ th 'cum f rank score.

Below 1549.5 score limit there are 11 person say, 22% of the population.

To get the 25th percentile figure of this score we have to go to the score that corresponds to the 12.5 cumulative corresponds to the 12.5 cumulative frequency levels, that means, $12.5 - 11 = 1.5$ f score in this class.

The class 1549.5 - 1599.5 has a frequency of six. Each frequency carries $50 / 6 = 8.33$ score value.

Hence, adding 12.50 to the lowest value of this class 1549.5, $1549.5 + 12.5 = 1562$ mm would be the 25th percentile score of this population group 75% of this survey of the population mentioned above will have larger stature than this value and the rest will lie below this level.

NPTEL

(Refer Slide Time: 37:14)

Stage V: The calculation

True lowest values	Classes with class interval, $i = 50$	True highest values	Frequency, f	Cum. f .	Cum. % f .
1899.5	1900-1949	1949.5	2	50	100
1849.5	1850-1899	1899.5	1	48	96
1799.5	1800-1849	1849.5	3	47	94
1749.5	1750-1799	1799.5	1	44	88
1699.5	1700-1749	1749.5	2	43	86
1649.5	1650-1699	1699.5	7	38	72
1599.5	1600-1649	1649.5	12	29	58
1549.5	1550-1599	1599.5	6	17	34
1499.5	1500-1549	1549.5	3	11	22
1449.5	1450-1499	1499.5	4	8	16
1399.5	1400-1449	1449.5	2	4	8
1349.5	1350-1399	1399.5	2	2	4
			N = 50		

Handwritten notes: 25/2 = 12.5 (top right), 12.5 (next to Cum. f. 17), 11 (next to Cum. f. 11)

Now, if we see the calculation part here. Percentile 25, the 25th percentile value corresponding cumulative percentage frequency falls in the class of 1550 - 1599, in this case; in this the true class of 1549.5- 1599.5 and correspondingly to the cumulative percentage frequency, 25th percent position of the cumulative frequency, will be the 25th percent of the total number in that is 12.5th cumulative frequency rank; means, we want 25, in this case what is happen? We have taken the double, double means into 2, like this have you got it from cumulative frequency, cumulative percentage frequency. So, 25 if we require, then 25 divided by 2 equal to 12.5; so, in this original cumulative frequency, it would be it would be 12.5 score.

So, till 11 **is** this is there, so from 11 to 12.5 how far we should go? That is that below 1549.5 score limit there are 11 persons, 22 percent of the population. To get 25th percentile figure of this score we have to go to the score that corresponds to the 12.5 cumulative corresponds to the 12.5 cumulative frequency levels, that means, 12.5 minus 11 **is equal to** 1.5 frequency score to the class; means, here is 11, so we have to go to 12.5 means 1.5, we have cross here; so, within this 50, there are total 6, so 50 by 6 for one score, **for one frequency how much it crosses, then into 1.5 where it matches**. Here, that will be the 25th percentile value, means, the class this has a frequency of 6. Each frequency carries 50 score, whether in one class divided by 6 is equal to 8.33 score value, for each frequency.


Hence, adding 12.50, means, if this is for 1, then 1.5 multiplied to this, then it becomes 12.50; so, with the lowest value here, with the lowest value, this one below this 11 are there, then this one plus 12.50 whatever is coming from here, if we add it means 1.45.5 plus 12.5, you have to 1562 millimeter it would be the 25th percentile score of this population group, where 75 percent of this survey of the population mentioned above or the first, will be larger stature than this value and the rest will be the below this value; so like, this all the percentile figures we can calculate accordingly like.

(Refer Slide Time: 41:15)

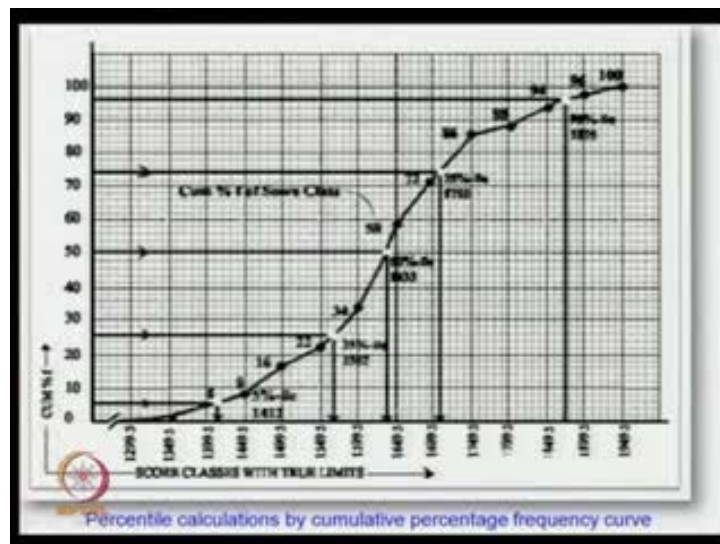
Percentile calculation by using graph

A curve with Y axis of cumulative percentage frequency and an X axis of the true limits of the score group as shown in below Fig. can also be used as a graphical method of getting the required percentile values.

Drawing a horizontal line from the required percentile point of Y axis to the curve and from that meeting a vertical line to the x axis would give a corresponding score and that would be the respective percentile value.



(Refer Slide Time: 41:22)



Now, this is by a calculation, we have done arithmetic calculation. Now, by a graph, in this graph, we can say that suppose we are measuring a graph, this is the true cumulative, the true class values, in this class values, this is the cumulative percentile percentage frequency; the frequency we are putting adding like this way. So if we want to see that where the 8, this is 4, this is 8, this is 16, like that as per the, now if we want to have a 25 th mark in this graph which is 25 th, then where it falls this.

So, count the measurement here, and then make a this each row is the... So, what is from this interval to this interval? 50 data, 50 is there, so 1, 2, 3, 4, 5, so means, each dimension is a 10 th; so, then what would be corresponding figure matches here, that would be the percentile.

(Refer Slide Time: 43:05)

Percentile calculation by using formula

$$P_p = I + \frac{PN - F}{f_p} \times i$$

P_p = required percentile rank
I = lowest value of the class intervals
PN = cumulative frequency in relation to the P_p point
F = cumulative frequency below the lowest value of P_p class
f_p = actual frequency of P_p class
i = class interval score

Examples:

$P_5 = 1399.5 + (2.5 - 2) \div 2 \times 50 = 1412$ [PN = 5% of total N, 50 = 2.5]
 $P_{10} = 1449.5 + (5 - 4) \div 4 \times 50 = 1462$ [PN = 10% of total N, 50 = 5]
 $P_{25} = 1549.5 + (12.5 - 11) \div 6 \times 50 = 1562$ [PN = 25% of total N, 50 = 12.5]
 $P_{50} = 599.5 + (25 - 17) \div 12 \times 50 = 1632.8 \approx 1633$ [PN = 50% of total N, 50 = 25]
 $P_{75} = 1699.5 + (37.5 - 36) \div 7 \times 50 = 1710$ [PN = 75% of total N, 50 = 37.5]
 $P_{95} = 1849.5 + (47.5 - 47) \div 1 \times 50 = 1874.5 \approx 1875$ [PN = 95% of total N, 50 = 47.5]

the shorter the class interval, the more accurate the results.

So, if we require 50 percentile like this way, if we require 75 th, 2 percentile, 3 percentile, it is like this way. So, it is a graphical system to measure that so percentile calculations by cumulative percentage frequency curve, another is that. Now, if we see whatever the mathematical calculations we have, arithmetic calculation we have discussed, the formula would be that preferred percentile rank P p equal to I plus P N minus F divided f p into I; P p means required percentile rank; if we require 5th percentile, then P p would be 5 p 5. Then lowest value of the class interval, if we go to this, in the this is the five, means, it is here; so, this is the lowest percentile, lowest interval of the class, lowest interval of the class plus to 2.5, means, this is that cumulative

frequency in relation to the P_p point, means, we want 5. So, in that table we have calculated 2 multiplications from the cumulative frequency to the percentage frequency; so, 5 divided by 2 to 0.5.

So, below that class only two are there, so point five level, we have to, and divided by f_p ; means, f is that cumulative frequency below the lowest value of P_p class, and f_p is that actual frequency the P_p class, in our discussion, it is 50; so, the actual frequency to the class, is that, it is number 2, in this case, in this actually **the 2**.

Now, then into I , I means class interval score; in our case, we have taken the class interval of 50 scores, so like this way, so 1412 would be a percentile 5 value, in this distribution; so, accordingly p percentile 10, 25, 50, etcetera will be one can calculate it.


(Refer Slide Time: 45:29)

Mean, Median and Percentiles

The mean is the best measure of the central tendency of the distribution. In the above it is 1637.5

The median is the midpoint of a score arrangement from the lowest to the highest order, which would be the 50th percentile figure. In the above case, it is 1632.8, and its graphical representation is shown in Fig below.

Roughly, for the purpose of a survey of anthropometric data, the subject number is vast and hence, in references, when addressing the 50th percentile value, the term 'average' is also used.

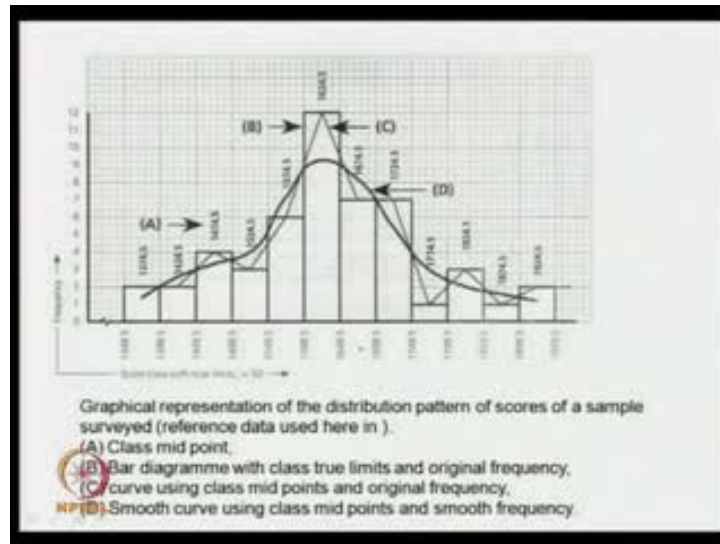
 MPTEL

Now, the shorter the class interval, the more accurate will be the results. Now, Mean, Median, and Percentiles: the mean is the best is average, is the best measure of the central tendency of the distribution. In the above scores, 50 scores individuals, we have measured value comes to 1637.5, but as per our percentile 50th percentile concern, it comes to 1632.8 means 1633 millimeter, but here we can see that average is different, this may be due to non-appropriate sample size.

So, wherever this possibility is there, it would be better to take, mean, 5th percentile value, 50th percentile value for this. The median is the mid-point of a score arrangement

from the lowest to the highest order, which would be the 50th percentile figure. In the above case, whatever data concern we are discussing, it is 1632.8, and its graphical presentation is shown here below. So, roughly, for the purpose of a survey of anthropometric data, the subject number is vast, and hence, the references, when addressing the 50th percentile value, the term average is also used; this normally for anthropometric survey a vast population size is considered

(Refer Slide Time: 47:25)



Now, in this case, the graphical representation if we see, now earlier what is happened? We have seen this graph, this bar diagram, we have seen. Now, the centre points, if we do like, then we will get the curvature; we will get the curvature like this.

So, **the** in the graphical presentation, of this distribution pattern of scores, of a sample survey, that reference data used here, in A is the class mid-point; these are the class mid points, means, minus it divided by 2 that comes the mid-point. Bar diagram with class true limits, and original frequency is the B, this structure; the blocks are the bar diagram with true class intervals. C curve, this is the C curve, this one, C curve using class mid points and original frequency; and this is the D curve, a smooth curve using class mid points and smoothening the frequency.

(Refer Slide Time: 48:59)

Smoothing a frequency distribution

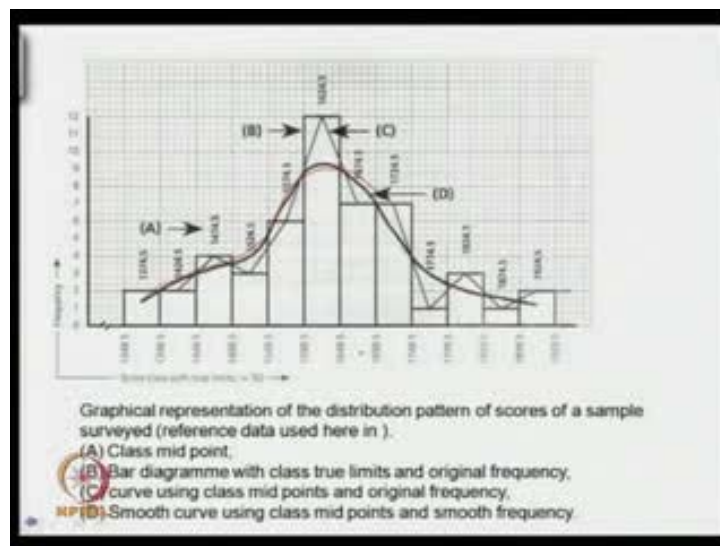
True lowest values	Classes with class interval, $i = 50$	True highest values	Group rank	Original frequency	Smooth frequency*
1948.5	1950-1999	1999.5	0	0	
1898.5	1900-1949	1949.5	12	2	1.25
1848.5	1850-1899	1899.5	11	1	1.25
1798.5	1800-1849	1849.5	10	2	2.00
1748.5	1750-1799	1799.5	9	1	3.00
1698.5	1700-1749	1749.5	8	7	5.50
1648.5	1650-1699	1699.5	7	7	8.25
1598.5	1600-1649	1649.5	6	12	9.25
1548.5	1550-1599	1599.5	5	6	6.75
1498.5	1500-1549	1549.5	4	3	4.00
1448.5	1450-1499	1499.5	3	4	2.75
1398.5	1400-1449	1449.5	2	2	2.50
1348.5	1350-1399	1399.5	1	2	1.50
1298.5	1300-1349	1349.5	0	0	

$N = 50$

*Calculation for smooth frequency of a corresponding group:

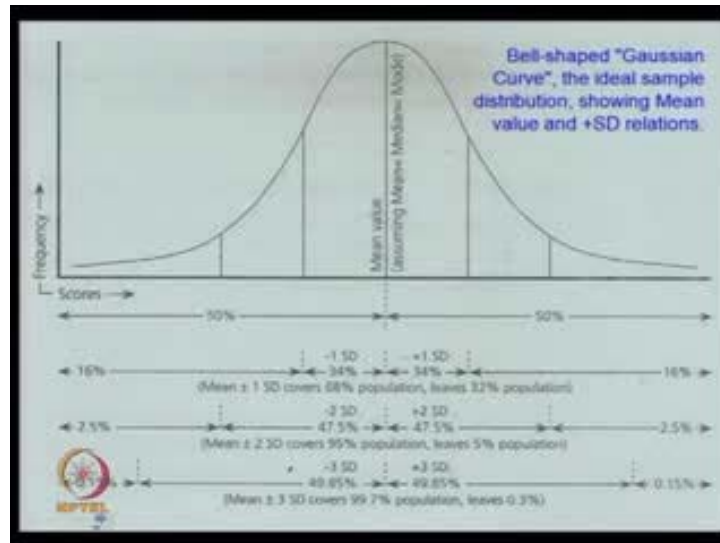
$$\frac{2 \times \text{Frequency of the original group} + \text{Frequency of the group just above the original group} + \text{Frequency of the group just below the original group}}{4}$$

(Refer Slide Time: 49:48)



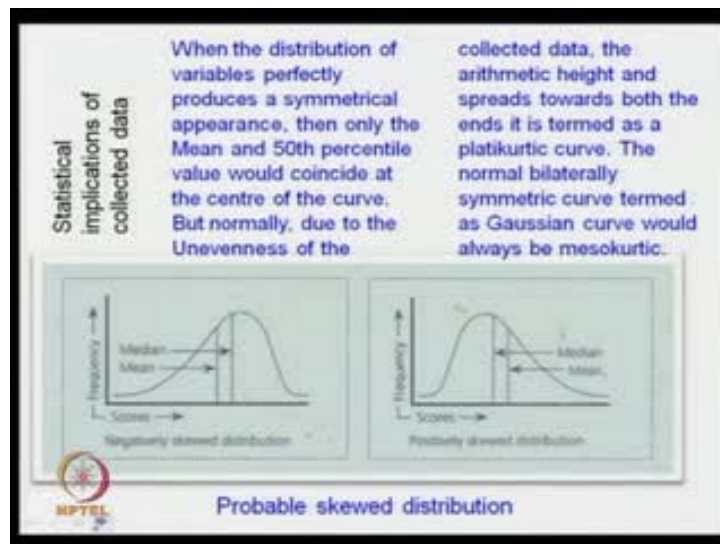
How to make the smoothing of the curve? True class intervals, then see that in this class interval, what was that group ranks, and then original frequency, and then by adding. **the** So, these system the calculations, calculation for smooth frequency of a corresponding group is that two times the original frequency of the original group plus frequency of the group, just above the original group plus divided by the frequency of the group, just below the original group, if you see and divided by 4, this will be the smoothing points; by these, from with using this details, the smoothing of the curve is made and shown here.

(Refer Slide Time: 49:59)



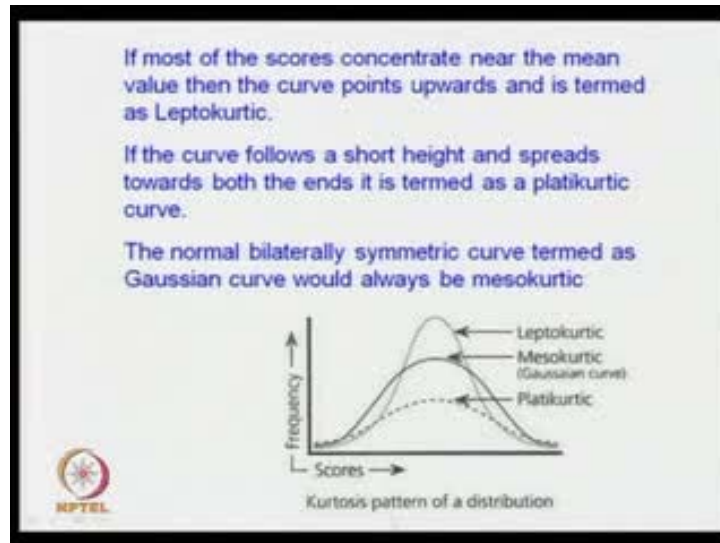
Now, the standard deviation and sketch: these are the standard deviation of a diagram, that in any statistical book or reference book, this will be, these are given there. So, this is the bell shaped curve, the Gaussian curve, the ideal sample distribution showing, mean, values and standard and plus minus S D relationships; so, like this, so from this mean value, if we have 1 S D minus and 1 S D plus, then it calculates 34 plus 34 is 68 percent of the population, will lie within this range, and then rest like accordingly this chart it can be seen.

(Refer Slide Time: 50:57)



Now, in certain cases, if we see that the curves are in different structures; so, this cuteness it shows that if our subjects sampling is not proper, then these type of curves are possible.

(Refer Slide Time: 51:18)



Now, if most of the source is concerned near the mean value, then it is termed as a leptokurtic curve. If the curve follows a short height and spreads towards the both the ends it is termed as platikurtic curve. The normal bilaterally symmetric curve termed as Gaussian curve, and would always be mesokurtic.

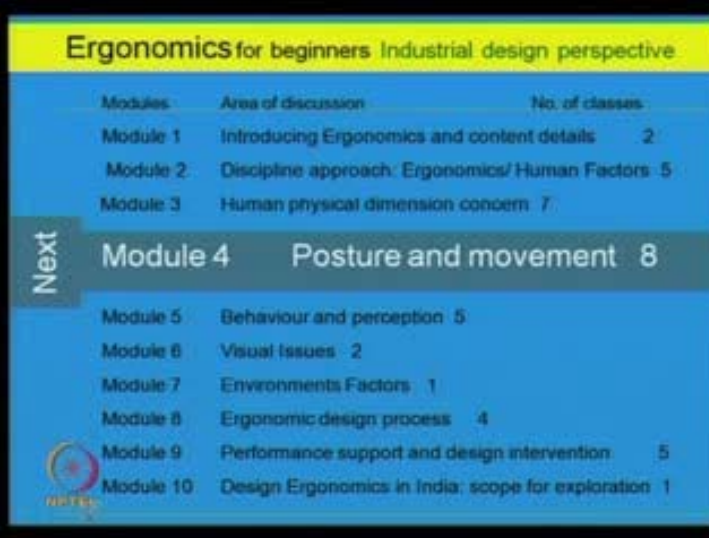
(Refer Slide Time: 51:47)

Anthropometric data are normally presented in percentile figures with range of minimum and maximum values, and sometimes with Mean \pm SD understands the data distribution pattern in a particular study population group. It helps in selecting the required data of design relevance.

NPTEL

Anthropometric data are normally presented in percentile figures with range of minimum and maximum values, and sometimes with mean plus minus SD to understand the data distribution pattern in a particular study population group. It helps in selecting the required data of design relevance.

(Refer Slide Time: 52:09)



Modules	Area of discussion	No. of classes
Module 1	Introducing Ergonomics and content details	2
Module 2	Discipline approach: Ergonomics/Human Factors	5
Module 3	Human physical dimension concern	7
Next	Module 4 Posture and movement	8
Module 5	Behaviour and perception	5
Module 6	Visual issues	2
Module 7	Environments Factors	1
Module 8	Ergonomic design process	4
Module 9	Performance support and design intervention	5
Module 10	Design Ergonomics in India: scope for exploration	1

Now, with this we can say that - today we are completing the module number 3 that is human physical dimension concern, and the next a new module we will start as a module 4 that posture and movement concern. So, with this we are concluding today's session; so, next session we will go for posture and movement and its design relevances.

Thank you very much.