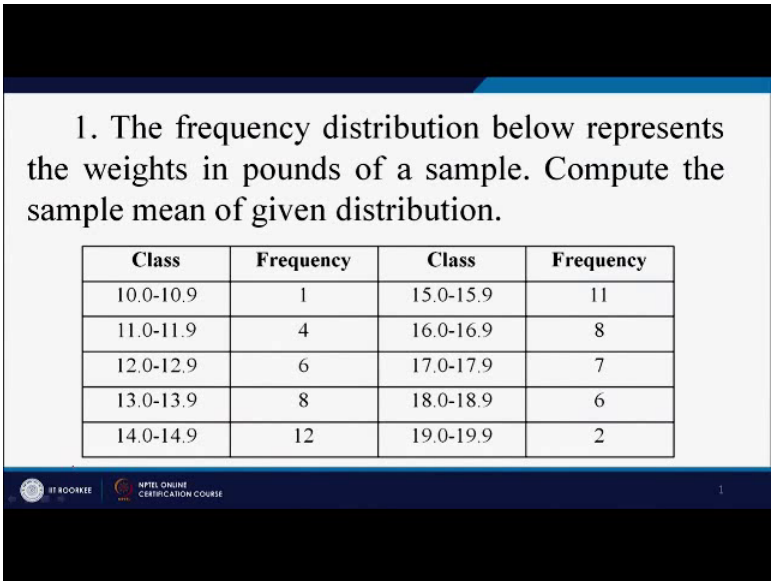
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 09
Applications of Measures of Central Tendency and Measures of Variation

Good afternoon friends as you are aware in previous couple of sessions we did workout couple of methods of measuring Central tendency and couple of methods of measuring dispersion. We are going to see in today's lecture some of the applications of those two. Lets us look at very first slide this is a question where in will been given weights of different items collected over a period of time.

(Refer Slide Time: 01:16)

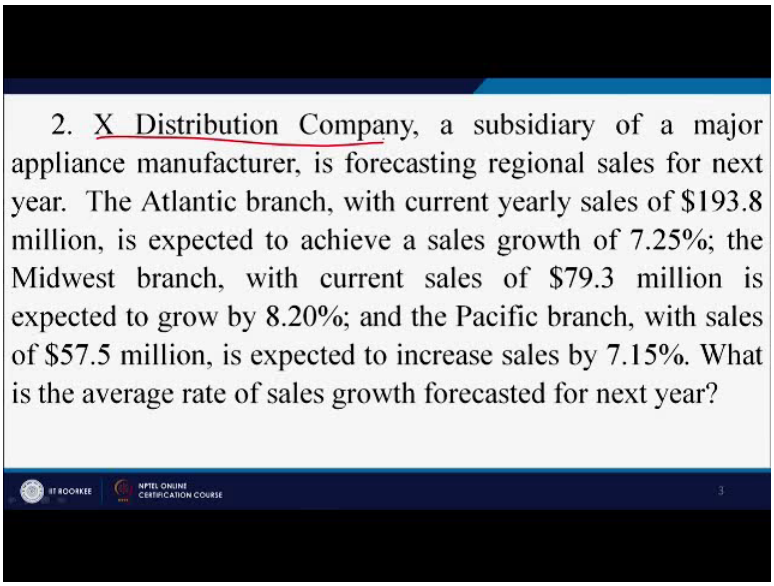


And after putting all of them in classes along with their frequencies we have got this particular frequency table. Now we have to find out mean of this particular question. So, very simple example we have already worked out this type of example in previous couple of sessions. So, it would be very simple what you would be doing first thing the first step is what would be taking midpoint is it not. Midpoint of each of these classes is it not.

So, that would be lets term as x, so x into f multiply both of them take summation of x and f and divided by n right, n is sample size right. So, here there are how many observations you can see

1, 2, 3 no this is not, these are the classes so the end would be sum of all this frequencies all this frequencies and all this frequencies. So, that sum would be 65, so mid points are mentioned over here. So, this f*x would be 988.50 so mean is 15.2077 this is a very simple example of grouped data or ungrouped data. So this was a question on grouped data right because we have arrange data into classes and each class was having different frequencies right this was a case of grouped data.

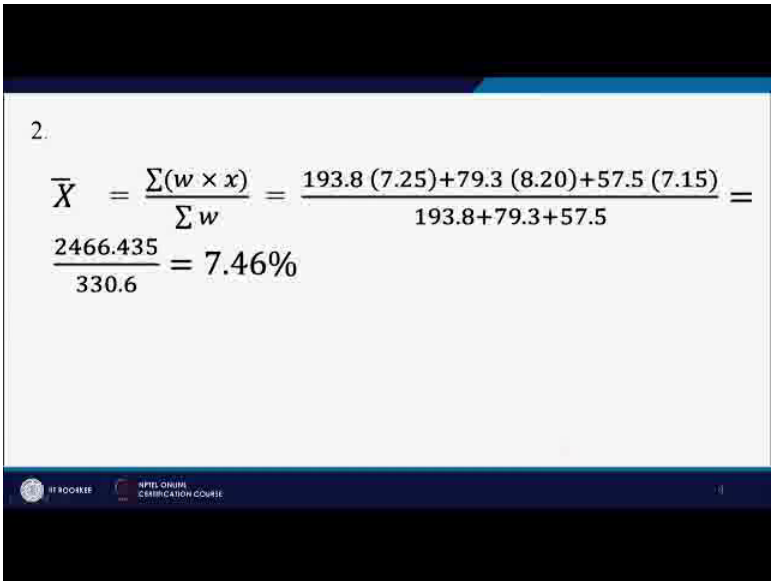
(Refer Slide Time: 03:04)



Let us look at another example there is a company called X distribution company subsidiary of a major appliance manufacturers is forecasting sales of its product for the next year. So, it has got three branches Atlanta branch then it has got Midwest branch and Pacific branch. The Atlantic branch which currently which has got current sales of this much million is expected to achieve a sales growth of 7.25%. Similarly this branch has got this sales much and expected to grow at this rate.

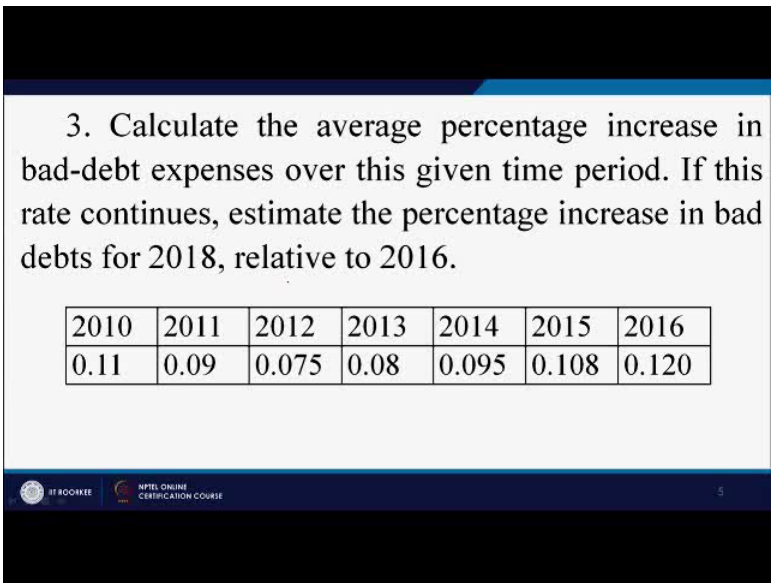
Similarly Pacific branch expected growth rate is 7.15 and sales is this, so we have to find out what is the average rate of sales growth forecast for the next year. So, this is an example of what kind do we have to calculate arithmetic mean or geometric mean or some other mean, just think for a while. So, if you answer is arithmetic mean that is wrong answer, if it is Geometric mean that is also wrong answer right.

(Refer Slide Time: 04:24)



So, the correct answer for this question is weighted average mean. So, what you have to do we have been given sales of these different branches and along with their expected growth rate. So, just multiply sales with their respective growth rates and take sum of these values right, sum of all these three sales right this is summation right. So, you will get 7.46 the growth rate for the next year sales right. So, this was again a very simple question.

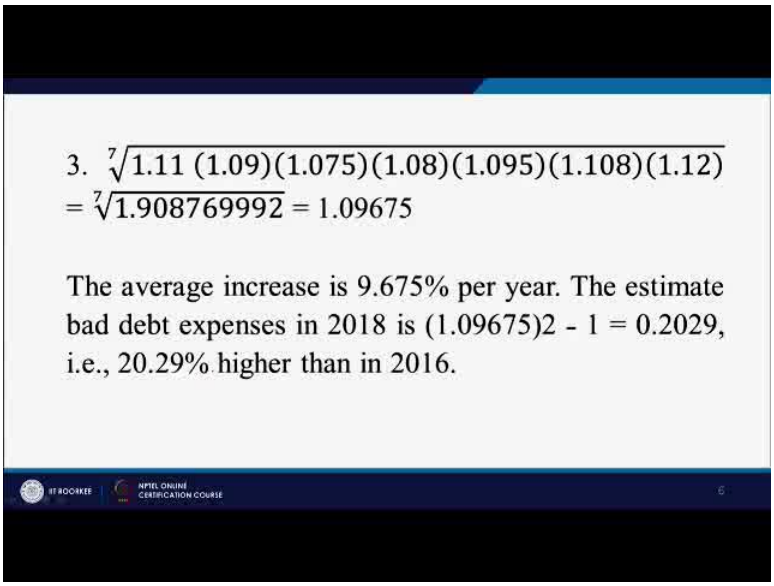
(Refer Slide Time: 05:04)



Third one is calculate the average percentage increase in the bad that expensive over this given time period. So time period is from 2010 to 2016 if this rate continues estimate the percentage increase in bad debt for 2018 relative to 2016 this is question. So, what we have to apply here which mean Arithmetic mean, weighted average mean or geometric mean. So, it would be case

of geometric mean. So, how many observations are there 1,2,3,4,5,6,7 this is n =7 right. So, you have got these different rates so will go to previous slideshow so, .11, .09, .075 so this is 1.09 not 1.9 ok.

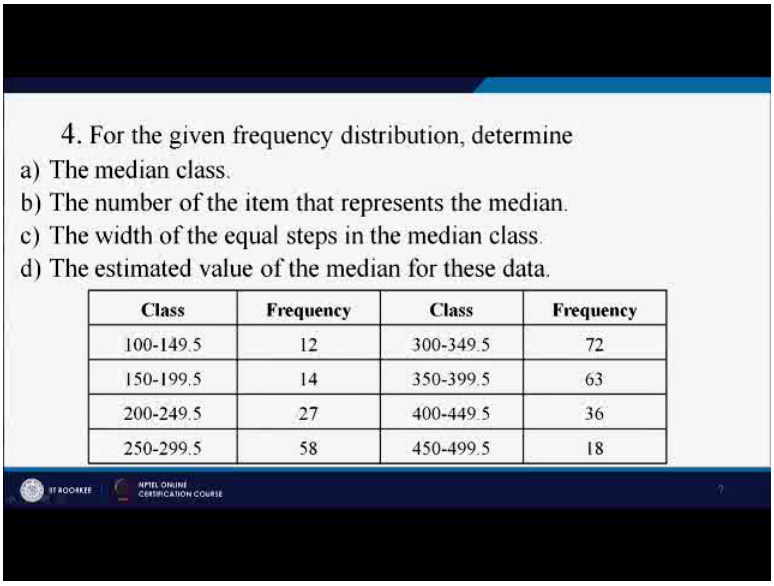
(Refer Slide Time: 06:16)



So, similarly you can see, all these are you can just multiply these values by 100 right. So, you will get percentage value, so in this case now all these values are all right. So, answer to this question is 1.09 ok. So the average increase rate is 9.675 per year. How did we get this, this value right multiplied by 100, 9.675% per year. The estimated bad debt expenses in 2018 is this value 1.09 and square of this right is not multiplication the square of this -1.

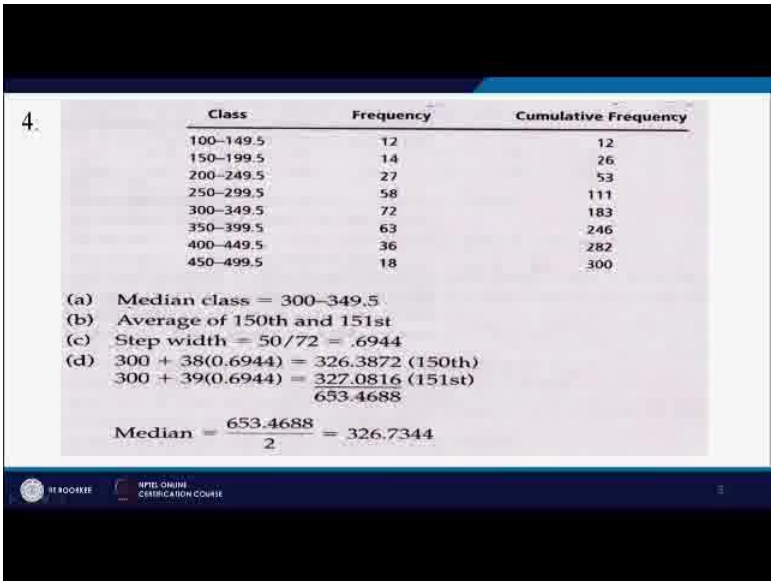
So, you will get 20.29 as the answer to this particular question that would be the estimated bad debt in 2018.

(Refer Slide Time: 07:29)



Let us move on to next question for the given frequency distribution find out median class, the number of items that represents the median class. So, how many items are there, the width of the equal steps in median class, the estimated value of the median for this set of data for this particular table. So, what would be the median class how to find out median class?

(Refer Slide Time: 08:09)

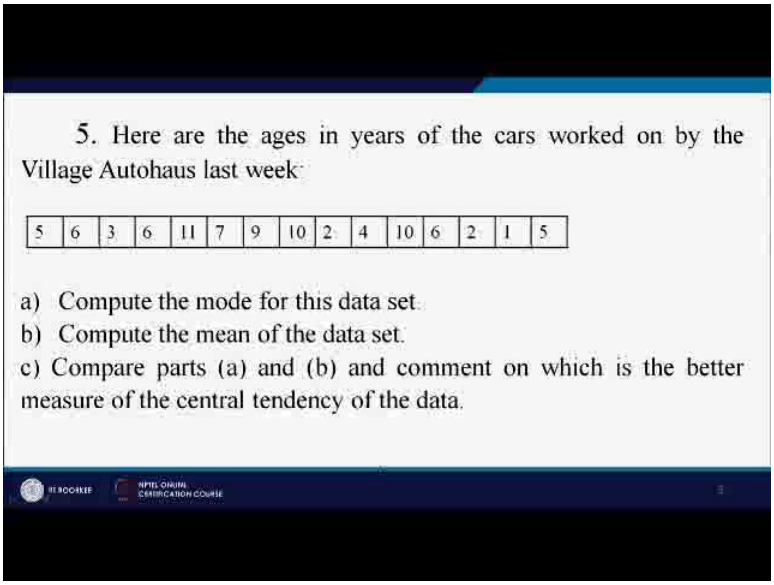


It is just take the summation of all the frequencies. So, this is summation of all the frequencies 300, so this is cumulative frequency right. Take cumulative frequencies are 300, the median class would be what 300 by 2 right. So, this is 150. 150 is where? Here right. So this is your median class. Average of 158 and 151st, in fact what you can do is let say if you have got odd number of data points or odd data set having odd numbers.

Then take the middle one for example if there are three data point 3, 4 and 5 to this would be the median right isn't it. So, it is basically the formula is (n + 1)/ 2 right to this 3 + 1, 4/2 this second position would be median, so, here also you can find out either it would not be any much difference either you divided by 2 are just (300 + 1) / 2, 300.5.

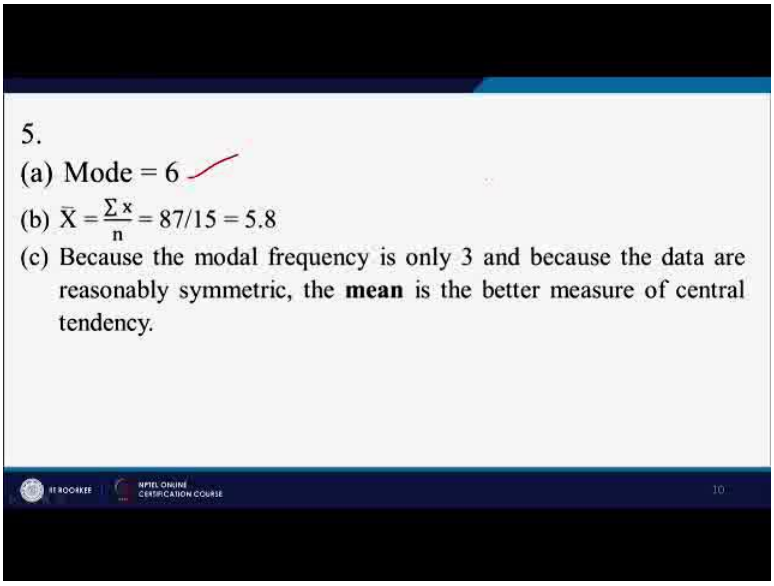
So, that is somewhere here it is divided by 2, 301 / 2 so this would be 150.55 so that that comes over here. So, this is your average of 150^{th} and 151^{st} right, this what I have said. Step width is, so this class interval is 50, right and this interval is also 50, you just take it as 50. So, 50 divided by 72 that would be the step width. Finally, you want to calculate median, so what you want? You want 150^{th} point and 151^{st} point right. So this are those two points. So, median would be this.

(Refer Slide Time: 10:50)



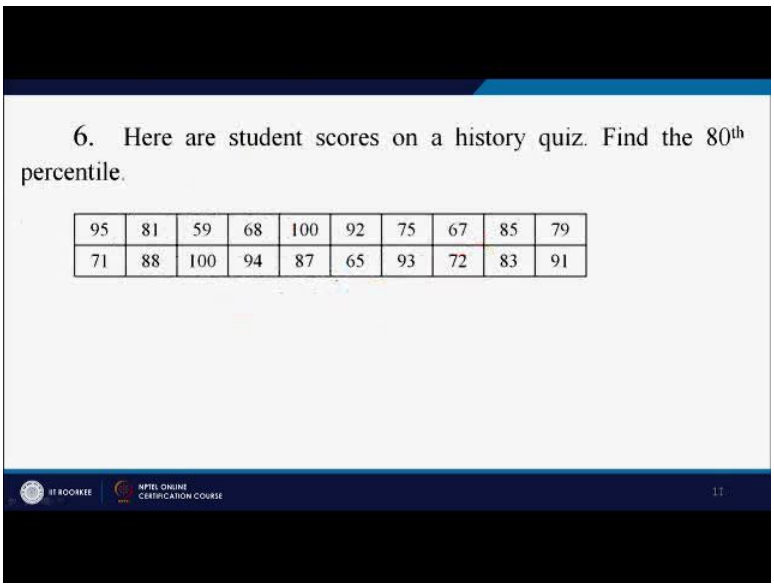
So, another very simple question let us look at one more case here are the ages in years of the cars worked owned by the village autohaus last week. Compute the mode of this data set so how to compute you just arrange them in ascending or descending order and find out which number has got maximum frequency. Compute the mean of the dataset, compare these two parts compare these two answers and comment on which is the better measure of Central tendency for this particular dataset. So, let us calculate so mode is 6, why 6 because it is coming three times 1, 2, 3.

(Refer Slide Time: 11:42)



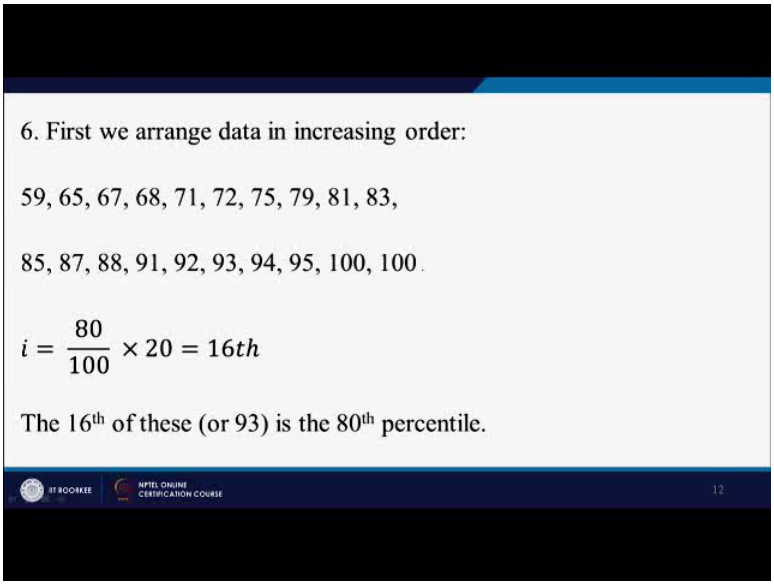
So, mode is 6, mean is 5.8 now which is better out of these two. So, because here the model frequency is only three it is not coming for large number of times and if you look at this data it is fairly dispersed data it is not a concentrated data right so you can that mean is a better measure because the data is reasonable symmetric. So, which is the final answer? Mean is the final answer here.

(Refer Slide Time: 12:26)



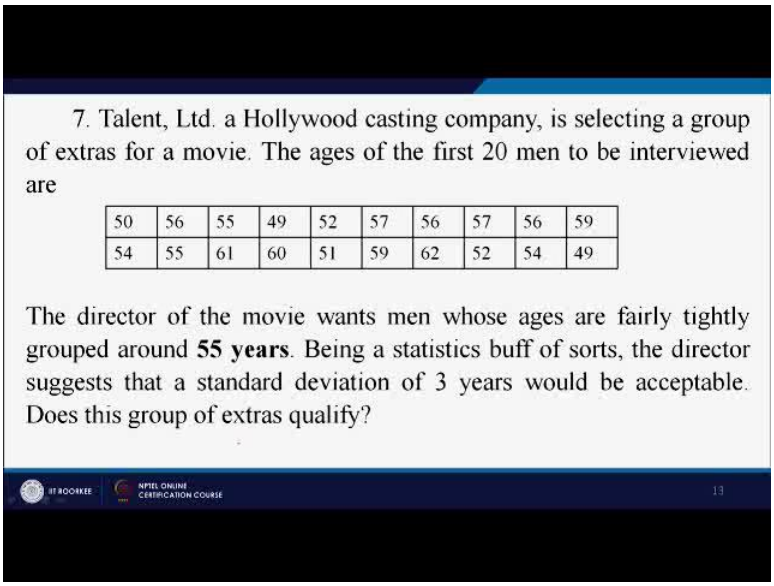
Here are student's scores on history quiz find the 80th percentile. So, how many data points first of all 1 2 3 4 5 6 7 8 9 10 and 10, 20 so total 20 data points that we have to find out what 80th percentile, so, 22 * .8 right so 16, 16th value would be the answer 16th data point would be answer right. So, that would be 93.

(Refer Slide Time: 13:03)



In fact what you should do is, you should arrange data first of all right. In fact in this case this 16th position is of 93 that is why we are getting this answer. So, 93 is the answer to this question.

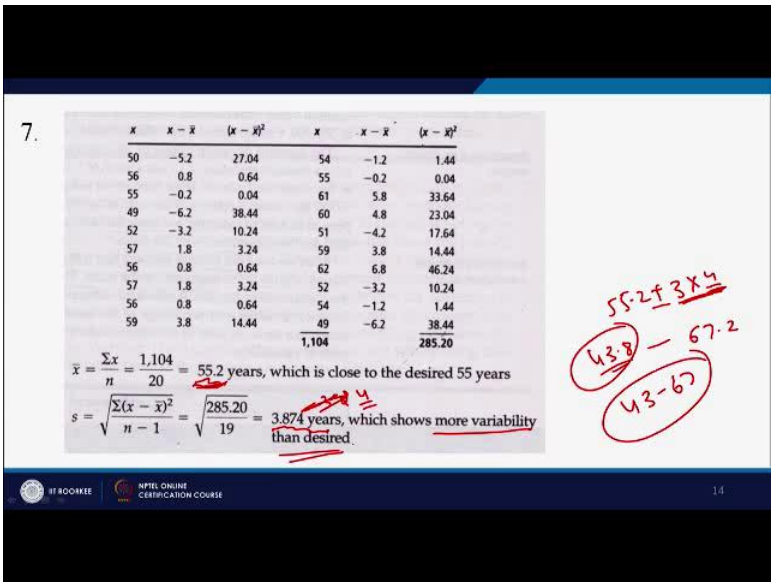
(Refer Slide Time: 13:28)



So, let us look at next question, Talent Limited Hollywood casting company is selecting a group of extras for a movie. The ages of the first 20 men to be interviewed are these are 20 these are ages of different first 20 men. The director of the movie wants man whose ages are fairly tightly grouped around 55 years. So, those people would be selected by the director who have got age nearby 55. If now being a student of statistics the director suggest that a standard deviation of 3 years would be acceptable.

So he is ready to even accept those men who fall in mean ±3 standard deviations. Thus these group extras qualify. So, we have to take a decision over here. So, thus this group of extras qualify so what you should do first of all it is very simple I think you should you should try to find out mean first is it not and then standard deviation and then you can think of what would be the range in ±3 standard deviation.

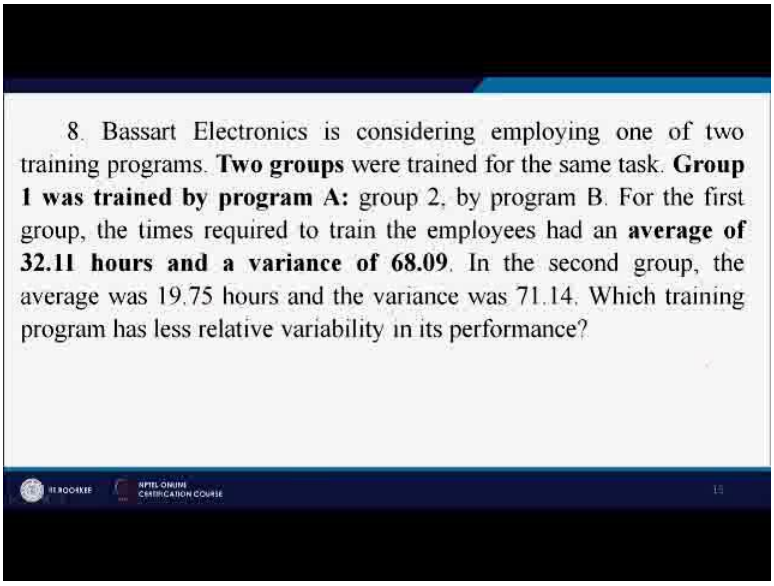
(Refer Slide Time: 15:10)



So, let look at this so simple one is first of all you just take a mean which is 55.2 right so how did you take mean you just the summation of all this divided by 20 so this 55.2. So, mean is, ok right 55 because the director is looking for those people who got age 55 right. But let us look at standard deviation, for standard deviation you need to do all this calculation right. So, standard deviation here is 3.87 let us take it 3.9 for simplicity or let us take it 4.

Now mean is 55.2 ± 3 standard deviations this your standard deviation so this becomes what is 12 right. Now so this is 67.2 and this again 12 so 43, - 43.8, something like that so broadly you can say 43 to 67 ok. So, what is the conclusion? Now this standard deviation shows more variability then desired because this is the variability 43 to 67 had it been around let us say 51 to 59 just within 1 Sigma limited we would have considered this group. But here the variability is more right so will not consider this particular group. This group what is the answer? This group does not qualify for extras right.

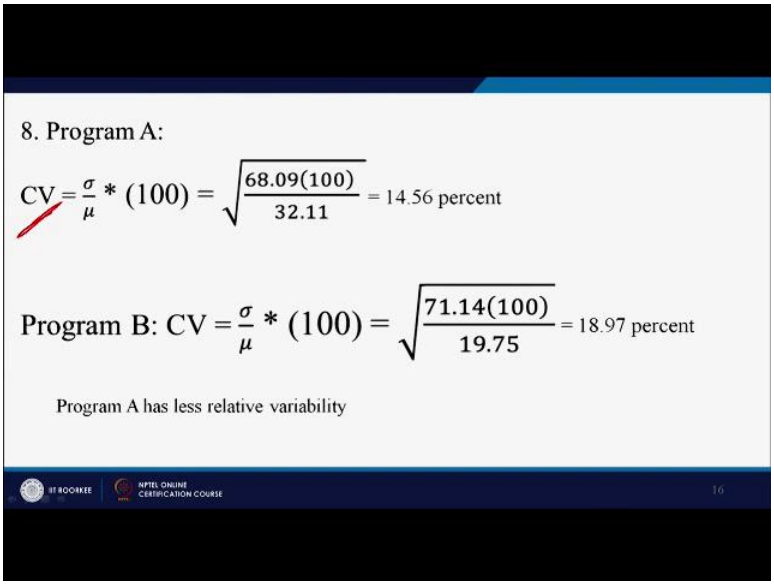
(Refer Slide Time: 16:59)



So, let us look at one more example, two groups in Bassart Electronic is considering employing one of two training programs so there are two groups group 1 and group 2. So, group one was trained by program A and group 2 was trained by program B and the total timing of training was noted. So, for the first group the average timing was 32.11 hours and variance was this, so this was the mean and this was the variance.

In second group the average was this and the variance was this so which training program has less variability. So, this question is what about think little bit. What we have to do here we have to find out which group has got less variability. So, how do we find out variability what are different measures? What are different measures of variability how we get range we got standard deviation without variance and coefficient co-variation right.

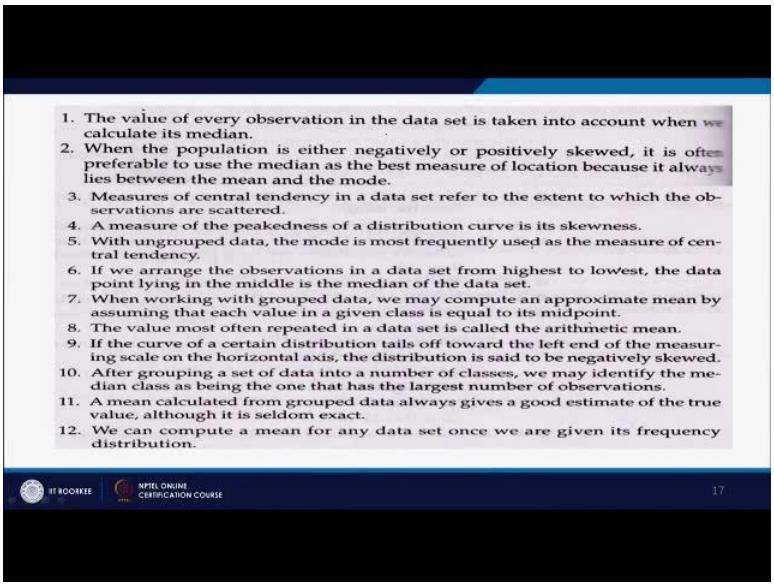
(Refer Slide Time: 18:42)



So, this is basically its coefficient of variation is Sigma by Mu divided by Sigma by μ*100 so it is the ratio of standard deviation to mean right. So, this is 14.56% and coefficient of variation for program B is this. So, where variation is less in case of program A. Program A has got less variability

Let us look at some the statements which would make your fundamentals clearer so you need to and read theoretical part very clearly but when you work on exercises your concepts become more clear. So, let us look at these couple of questions.

(Refer Slide Time: 19:38)



The value of every observation in data set is taken into account when we calculate its median it is true or false? It is false we do not take each and every variable when we go for calculating median generally we take each and every variable when we go for standard error. Let us look at second question when the population is either negatively or positively skewed it is often preferably preferable to use the median is the best measure of location because it is always lies between mean and median.

So, I will tell you once again you have got you can have a perfectly bell-shaped distribution which is normal distribution right this known as symmetric distribution. You may have a situation where distribution is like this so this is your point all these are negative values and this side all are positive values right. So, this is a negatively skewed distribution. So, whenever there is a skewed distribution or non symmetric distribution median is the best measure, so this is true.

Let us look at third point, Measures of Central tendency in a data set referred to the extent to which the observations are scattered is it? Do we measure thickness of the data using Central tendency and what it is? Is it the scatteredness? Is central tendency or something else? So, this is false right because Central tendency is something which gets measured by mean, mode and median right, so this is false right.

A measure of the thickness of a distribution curve is skewness, no it is not skewness it is known as kurtosis so this is also false. With un-grouped data the mode is most frequently used as the measure of Central tendency no this is false. With un-grouped data the mode is most frequently used as measure of Central tendency no we do not know what is the nature of the data right, whether the data are normally distributed or non normally distributed.

Sixth if we arrange the observation in a data set from highest to lowest the data point lying in the middle is the median of the dataset, is it correct is it not? So, the middle the 50th percentile point is the median right. Let us look at next one when working with grouped data we may compute an approximate mean by assuming that each value given class is equal to its midpoint, yeah this is true right. Seventh is true.

The value most often repeated in dataset is called arithmetic mean no it is called mode, so it is false. Let us look at next one if the curve of a certain distribution tails of towards the left end of the measuring scale on horizontal axis the distribution is said to be negatively skewed. So, this is just, I just talked about right. So, this is negative side this tail is towards negative side so this is true right.

So, this negative distribution then what would be the positive distribution, it would be like this is it not, so this is horizontal axis will not touch the distribution right. Next one after grouping a set of data into number of classes we made it we may identify the median class is being the one that has the largest number of observation. No, this completely false. Median class will not get determined by the largest number of observation but by what where the cumulative frequency is equal to n + 1 by the number of data points by this is; I will show you one example.

We just did one example like this yeah just see if you compare this question with that statement then infact this would have been the median class but it has happened just by chance it is not necessary that this would be the median class. But this median class is determined by how? This is (300 + 1) / 2 ok. So, let us move on to next question number 11, a mean calculated from grouped data always gives a good estimate of the true value although it is seldom exact, yeah it is it is true.

If you have got large number of data then it is possible that the data would be normal and in most of the cases when data or normal mean is the best measure right. Let us look it next question we can computer mean for any data set once we are given its frequency distribution, is it possible. We can compute mean of any data set once we are given its frequency distribution, it is necessary to frequency distribution for calculating mean, no this is false. If we have un-grouped data we can easily calculate mean.

Next question and we will look at few questions from this particular slide. Next one is this. The mode is always found at highest point of a graph of a data distribution is it like that? Yeah it is like isn't it? So, let us say data 1, 2 and 3 so how would you get this is 1, this is 2 and this is 3

right with the highest point right. The number of elements in a population is denoted by n, No it is not denoted by n but we denote population by N, so this is false.

15th for data array with 50 observations for a data array with 50 observations the median will be the value of 25th observation in the array, is it 25th observation or what? No it would be (25+1)/2 right sources 25.1 right so it would be the mean of 25th and 26th observations right so is false. Extreme values in dataset have a strong effect on median is it correct? Actually one of the advantages of median is not affected by extreme value, so this is, 16^{th} is false.

The difference between largest and smallest observation in a data set is called geometric mean, no, it is called range. No it is false. The dispersion of data set gives insight into reliability of the measure of Central tendency, it is true so you should calculate are you should find out Central tendency but apart from Central tendency if you have got the information about variability then it becomes reliable measure. So, 18th is true right.

The standard deviation is equal to the square root of the variance, is it correct. The standard deviation is equal to the square root of the variance, yes. Standard deviation is equal to the square is equal to sigma is equal to underroot of what? Variance right, ok, so if you got variants just take under root of it, it becomes standard deviation. If you got standard deviation just take square of it right it becomes variance right.

Just take the last question for today's class the difference between the highest and lowest observation in data set is called interquartile range, is it, it is completely false, why? The difference between highest and lowest observation is called range right not the interquartile range. So, with this let me stop here in today session we have seen couple of questions on those topics for which we had classes in last two sessions. So, in next class we will continue solving this particular exercise, thank you very much.