Marketing Research and Analysis-II (Application Oriented) Prof. Jogendra Kumar Nayak Department of Management Studies Indian Institute of Technology – Roorkee

Lecture - 6 Marketing Research Process - IV

Welcome friends for lecture series of the course Marketing Research and Analysis II. In the last lecture, we were discussing about some of the measures of the central tendency and why it is called central tendency and what is basically the use of them in the marketing research. So we discussed about things like what is mean, median, and we had started with mode and let us see. So as if you remember mean is nothing but the average. So when you take suppose why it is required, for example please understand a company who is let us say selling biscuits.

For example right now it wants to know whether can it make pack biscuits sell biscuits in a way that it is consumed on a daily or if not biscuit take bread for example. Now that bread would be consumed on a daily basis, so can it make a package for small families where member size of 4 for example, so that the packet remains fresh. Once it is opened, then it is consumed and it is over. So if it can do this, then people need not store unnecessarily and they will not have to pay also more because of large size packet. So to understand this, what it does basically.

(Refer Slide Time: 02:19)

Measures of Central Tendency

It tries to make a study of observing several families of such where 4 members are there, a small nuclear family, and sees what is the basic consumption pattern. Now if it takes for example and sees that the consumption of one family is let us say one is 8 slices, one is 13 slices, one is 11 slices, one is 9 slices, one is 16 slices. So suppose it takes such data. Now it asks you what is the best way of making it. Now let us take the average of it, now average for example 57. So 57 divided by 1, 2, 3, 4, 5, so 57/5 that is equal to, 57 cannot be made.

Either we make it let us say 55 or 60 for example to divide easily. So 11 or 12 slices let us say. So it makes a packet of let us say 11 or 12 slices and it can sell it in the market, that is one way of understanding from a marketing point of view. Similarly the question is when is the company use the median. The median it will use only to understand that in a particular location for example how would I divide. For example let us say this case only. If I arrange it in order how does it look 8, 9, 11, 13, and 16. So what is the middle point, now obviously this is the middle point, so 11.

So now you see here also we got somewhere around 11, here also we are getting 11. Now suppose it finds that this is the middle point, so this middle point cuts the entire line, may be for example this is the line, into two parts basically. So this is basically to understand and tell the marketer well the people, if you look at the consumption this is the middle point of the consumption right. So this sometimes why it is good is because if you see the middle point is not affected by any extreme value.

For example in this case had it not been 16 and something around let us say 14 or let say 13, 14 let us say 14 or if you add another variable suddenly suppose one family is consuming 30, let us say 30 slices, remove this, 30. So if you look at what is happening, 57+30 is 87, 87/6 is suddenly almost let us say 15 slices. Suddenly you see the number of slices has grown up from 11 to 15 almost okay. So this is what is happening, this is the problem with the mean.

The mean is getting affected by extreme values, the extreme values could be very high value or it could be something like only let us say 4 slices, so there also very small or very large value is affecting the mean, but that does not happen in the case of the median because the median had it been if you bring in the 30 also, suppose I take out this one and put in the 30 still my middle point is the 11 only. So that is the advantage of a median.

The median is utilized whenever you see there is a case of some outliers or extreme values present in the data set.

(Refer Slide Time: 05:28)

Mode value: A measure of location recognised by the location of the most frequently occurring value of a set of data

The mode is that value of an observation which occurs most frequently in the data set, that is, the point (or class mark) with the highest frequency.



Now mode. When is the mode being used by the marketer? The mode is for example I say take let us say the company has produced packs of 8 slices, 10 slices, 12 slices, 14 slices, 16 slices. It has done 5 packets and it wants to see how many of these are sold in the entire month. Now this is 100 packets, this has been sold let us say 150 packets, this has been sold 300 packets, this has been sold 160 packets, this has been sold let us say 120 packets.

Now the company imagines well that this is the one which has got the highest frequency, that means large number of people are more interested to buy a packet of 12 slices of bread. So this gives an idea for the company well it is better and suppose you can make it more explicit by understanding which location or which kind of families, that is different part of the data analysis, but at least specifically to understand if I can know that mostly the people are buying bread packet which has got 12 slices in it then this is the highest frequency which is called the mode.

This mode helps the company to produce more of that and so that the consumers always buy that and that also helps in its production process and selling process, okay the marketing process. So mode is that value of an observation which occurs most frequently in the data set. Now I hope you understood how mode is being used in the marketing perspective right okay. (Refer Slide Time: 07:12)

Example : Frequency Distribution of Sales Per Day

Salo selune (Class introal)	33-36	37-60	61-64	63-68	69-72	72 and above
Number of days (Perguency)	1	٩.	0	+	4	1

Using the data calculate the mode of sales distribution of the units of item during the 20 days period.

Now let us see how do you calculate. Now this is given the frequency distribution of sales per day 53 to 56, sales value 57 to 60, this is X Y Z product may be right, 61 to 64. The ranges are given to you, class interval is given to you, how many days it is being sold at this interval. Now this is 2, 4, 5, 4, 4, 1. Now if you see the highest is here correct, the highest is here so the highest frequency is 5 which is reflecting the class interval of 61 to 64, so maximum time the 61, the people are buying whatever it is.

The sales value is lying in between 61 and 64 okay. How do you calculate mode? Now if you want to use it, you can use it. Now this is how you see, the formula is this one. (Video Starts: 08:02) Mode is calculated by this formula. Now what is this, let us see.

$$M_o = l + (f_m - f_{m-1})h / (2f_m - f_{m-1} - f_{m+1})$$

The formula is l, this l you can see here, l is nothing but the lower limit of the model class interval. Now lower limit in this one in this case when this is your highest frequency, let us say the lower limit is 61 okay, f_{m-1} so what is f_{m-1} this one, frequency of the class preceding the mode class interval, preceding, just preceding.

So which one is preceding, this one is the preceding okay, f_{m+1} what is that, now f_{m+1} is this one we are talking about. The frequency of the class following the mode class, now what is following, the following in is this one correct. So you have preceding, you have following okay. So this one is for the preceding this one is for the following okay. Now using this and f_m obviously is the mode that means of the frequency of the highest one. So this one correct. So now using this what you have done.

One more thing is important. What did you see here. Since the largest frequency corresponds to the class interval of 61 to 64, it is the mode class. So we have l = 61, $f_m = 5$, $f_{m-1} = 4$, $f_{m+1} = 4$ and h = 3. Now what is this 3, let us see. So this difference, the height this difference is the 3 okay. So now this entire value becomes how much, 61+((5-4)/(10-4-4)). Now just look at it if you can see l = 61, 61; f_m is 5, so 5; f_{m-1} is 4, 4 right; $2f_m$ so 2 x 5 right 2 x f_m 10; f_{m-1} is 4; again f_{m+1} is 4.

Now this is when you do so this one is becoming total, 62.5. So you can calculate 65-4, so $61+(1/2) \ge 3$, so this is equal to 62.5 correct okay (**Video Ends: 10:33**).

So now I hope you have understood basically when you talk about mean, median, and mode, generally we talk it from only research process, but we do not connect it many a times with the marketing perspective. So as a marketer, how do you use this mean, median and mode is very important to understand okay.

(Refer Slide Time: 10:55)

Relationship between Mean, Median and Mode

- In a unimodal and symmetrical distribution the values of mean, median, and mode are equal.
- In other words, when these three values are all not equal to each other, the distribution is not symmetrical.
- A distribution that is not symmetrical, but rather has most of its values either to the right or to the left of the mode, is said to be skewed.



What is the relationship. In unimodal, unimodal means, why it is called unimodal, one mode. When there is only one mode and the data is symmetrical, the values of mean, median and mode are equal. In other words, this is what happens in the proper normal distribution this is what we are saying u (mu), median and mode all lie at the same line. In other words when these three values are not equal to each other, the distribution is not symmetric. Now that is very interesting.

When these mean, median and mode are not equal, that means the mean is not equal to mode or any one of them is not equal to the median whatever, in such a condition the distribution is a not symmetrical distribution and it has some of its values either to the right or left of the mode. Now mode, why this is the mode, because this is obviously the highest frequency.

(Refer Slide Time: 11:50)

 For asymmetrical distribution, Karl Pearson has suggested a relationship between these three measures of central tendency as.

> Mean - Mode = 3 (Mean - Median) Or Mode = 3 Median - 2 Mean

You see Karl Pearson has suggested relationship. Karl Pearson, the very famous Karl Pearson's correlation and all you have understood, you must have heard. So he gives a formula that

Mean - Mode = 3(Mean - Median)

Since Mean and Median are very close, it is a middle point and it is also very close to this, so he has given a formula Mean – Mode = 3(Mean – Median) or you can use it,

$$Mode = 3(Median) - 2(Mean)$$

okay, you can use this same formula and find out this one.

(Video Starts: 12:27) Now look at this diagram and it will be more clearer to you. In the first case, this is a symmetrical case as you can see. Mean, median, mode are all lying in the same place, but if you look at this graph, now this looks like it is slightly tilted correct. So here what has happened, the highest peak, the highest is at this point, so this is my mode correct. This is the highest point frequency, so this is my mode.

The middle point which cuts it into equal parts is the median and this is the this part right, but what has happened is most of the data are moving towards, you need to understand that we have a 0 point and this is the positive and this is my negative. So when my tail is more

towards this positive side, then I am saying it is skewed to the right that means when the mean is moving more towards the positive direction, this positive right, plus this is +1, +2, +3, +4, + infinity, this is minus infinity side.

So this is skewed to the right. Similarly if you look at this, this tail is more towards the left. So then minus value, minus side, negative side, so that is why it is called as skewed to the left. People try to remember it many a times I have seen when I ask students they try to rethink oh what is positively skewed, sometimes they get confused. No, you should only be understanding this much, that the tail is moving to which direction, is the tail tapered toward, moving towards the negative side or positive side.

If it is moving towards the positive side, then it is positively tailed distribution, it is right tailed or positive tailed or it is a left tailed if it is moving to the negative side.

Now let us look this is very important, you should be very very clear because your data analysis gets affected by these things. If most of the values of observation and distribution fall to the right of the mode as shown in figure B, this is the figure B, it is said to be skewed to the right and positively skewed. Distributions that are skewed right contain a few unusually large values of observations. In this case what happens, mode remains under the peak, obviously highest frequency.

So let us see this one, for example, this as I showed here and in this case this is my mode okay. The median and the mean move to the right. The median the middle point and the average, they are moving towards the right for a right tailed distribution. Now is it not true, let us see, they are moving towards the right, so they are shifting okay. In such a condition, the order of magnitude looks like this, mean is greater than median, is greater than mode right. So the mean is the largest value, then the median, and then the mode.

If you look at the other side, the other side which is skewed to the left, in such a condition what is happening. If the distribution is skewed to the left or negatively skewed, the lower magnitude are concentrated more to the left of the mode. So left of the mode look this side okay, the mode is again under the peak as usual, mode does not change, but the median and mean now move to the left okay. So the order of the magnitude now becomes mode is the

highest, then mode is greatest mode, then median right and then followed by the mean okay (Video Ends: 15:56).

This is how it looks. So here, the mean is become the smallest and in the positively skewed the mean was the highest, it was the lowest in the left skewed, and positive skewed it was the highest.

(Refer Slide Time: 16:17)

KEY POINT

In general, for a single-peaked skewed distribution (nonsymmetrical), the median is preferred to the mean for measuring location because it is neither influenced by the frequency of occurrence of a single observation value as mode nor it is affected by extreme values.



Key point now this is very important for a single-peaked skewed distribution which is that means nonsymmetrical. The median is preferred to the mean, why, what is the reason the median is preferred because as I explained in the bread slice case. Because the median is the middle point, it is not affected by any extreme value so because it is neither influenced by the frequency of occurrence of a single observation as mode nor it is affected by extreme values that is why in such a condition the median is the best thing.

You should be very clear with it. I hope you are understanding the clear meaning of these things mean, median, and mode and its connection with the real-life situations.

(Refer Slide Time: 16:55)

COMPARISON BETWEEN MEASURES OF CENTRAL TENDENCY

Comparison between the measures of central tendency.

(Refer Slide Time: 17:00)

Choice of Method

The choice of which method to use for describing a distribution of values of observations in a data set is not always easy. The choice to use any one of these three is mainly guided by their characteristics.

The characteristics of these three differ from each other with regard to three factors:

- Presence of outlier data values
- Shape of the frequency distribution of data values
 Status of theoretical development

Choice of method, how do you decide. The choice of method to use for describing a distribution of values of observations, please understand when you are talking about distribution of values means we are generally saying when we are trying to create a plot of data and see how the random data is distributed, so that is what we are interested. The distribution of data is what is interest for the researcher.

He is trying to see how is the distribution distributed, is it symmetric, is it skewed to only one side, may be to the right, to the left or whatever it is right. How do you do that. So when you are having such a problem, the characteristics of these three differ from each other, so what is saying.

The choice to use any one of these 3 means, these 3 is mean, median, mode, is mainly guided by the characteristics. Now what is the characteristics of these 3. Presence of outliers, as I said if outliers are present, then the mean is the first thing that will get affected, so you cannot use them. Shape of the distribution from that you can also understand and status of the theoretical development.

(Refer Slide Time: 18:09)

The Presence of Outlier Data Values

- The data values that differ in a big way from the other values in a data set are known as *outliers* (either very small or very high values)
- The median is not sensitive to outlier values because its value depend only
 on the number of observations and the value always lies in the middle of
 the ordered set of values, whereas mean, which is calculated using all data
 values is sensitive to the outlier values in a data set
- Obviously, smaller the number of observations in a data set, greater the influence of any outliers on the mean
- The median is said to be resistant to the presence of outlier data values, but the mean is not

Let us see what does it mean. Now if there are outliers that means very small or very high values, median is not sensitive, you have understood right. Since it only depends on the number of observations and the values lie in the middle whereas mean is calculated using all data which is sensitive, obviously smaller the number of observations in a data set, greater the influence of any outlier. If you have large data set, may be the effect of outliers would be nullified, would be decreased, but if you have a smaller data set, it is still is more dangerous. Median is said to be resistant to the presence of outliers, first thing okay.

(Refer Slide Time: 18:45)

Shape of Frequency Distribution

- In general, the median is preferred to the mean as a way of measuring location for single peaked, skewed distributions
- One of the reasons is that it satisfies the criterion that the sum of absolute difference (i.e., absolute error of judgment) of median from values in the data set is minimum, that is, Σ |x - Med |= min
- In other words, the smallest sum of the absolute errors is associated with the median value is the data set as compared to either mean or mode
- When data is multi-modal, there is no single measure of central location and the mode can vary dramatically from one sample to another, particularly when dealing with small samples

Shape. In general, median is preferred to the mean as a way of measuring location for single peaked, skewed distributions, means single peaked but a skewed distribution. One of the reasons is that it is the positive or negatively skewed distribution, it satisfies the criterion that the sum of the absolute, see this term is extremely important, because you need to connect it not only here but everywhere absolute difference, now what is this absolute difference, that is the absolute error of judgment of median from values in the data set is minimum, why because it is at the middle point.

So x, x is any variable minus the median is always the minimum ((x - Med) = min). This why I am saying is important because also when you draw regression lines, when you are talking about variance, you are basically trying to see how much the data is moving from a particular point, may be that is the mean or something, but how much it is moving, and this movement is very important in real life because also used to understand when you have large data pools and you see something, we say well I want to know how much does it move from the mean.

Then only I can understand whether should I be going for this particular product or not. If it is not it is moving sufficiently from the mean, that means it might be having a different requirement, it is a different product altogether. In other words, the smallest sum of the absolute errors is associated with a median value is a data set as compared to either mean or mode. Now when data is multimodal that means what, there are now two frequencies or more than two frequency, higher frequencies, that means two modes, not one mode but more than one mode.

There is no single measure of central location and the mode can vary dramatically from one sample to another, particularly when dealing with small samples.

(Refer Slide Time: 20:43)

The Status of Theoretical Development

 Although the three measures of central tendency—Mean, Median, and Mode, satisfy different mathematical criteria but the objective of any statistical analysis in *inferential statistics* is always to minimize the *sum of* squared deviations(centrs)) altern from these measures to every value in the data set

So shape of the distribution and the outliers, now the third thing they are affecting the study. The third is also similarly having an intension, this is what I was telling in the last slide also. The intention of any researcher is to see that there is the minimum sum of the squared deviations. Now what are the squared deviations. They are the error terms. Error does not mean actually error. Please understand error actually is the unexplained part. So when you are talking about this unexplained, the squared deviations if it is very high, that means what, our unexplained variance is more.

So we need to minimize so least square regression or variance whatever you talk about. So the criterion of the sum of squared deviations is also called least, this is what I am talking about, least squares criterion. Since arithmetic mean satisfies the least squares criterion, it is mathematically consistent with several techniques of statistical inference, that is since these central tendency measures, for example the mean, median, mode and especially the mean also many a times is they follow this consistent, they are consistent with this least square criteria

That is why mostly you see that the mean is largely utilized for any data analysis tools as the basic platform upon which other techniques are built up on. (Video Starts: 22:19) Now coming to something called the interquartile range. Now what is this interquartile range, why should you. The word quarter you must have heard. What is this, first let see. It is a measure of variability based on dividing a data set into quartiles. Now why it is important. The values

The criterion of the sum of squared deviations is also called *least squares* criterion. Since arithmetic mean satisfies the least squares criterion, it is mathematically consistent with several techniques of statistical inference

that divide each part are called the first, second, and third quartiles. For example just remember first quarter, second quarter.

A company has a first quarter report, second quarter report right, second quarter report, third quarter report. So it has divided basically into several quarters. How do you measure this quarter and why it is important because this quartile range helps in measuring your dispersion, is measuring your variability. So how do you measure. Suppose there are numbers given to you. First, put the numbers in orders, 1, 2, 5, you can see they are all put in order, so 1 being the lowest 27 being the highest and they are all in order. Find the median.

So what is the median. How many are there, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11. So if 11 points are there, so which is the middle. Now (11+1)/2 = 6, so the 6th one is 1, 2, 3, 4, 5, 6, so this is my median. Group the numbers above and below the median. So this numbers are grouped, these numbers are grouped, so there are 1, 2, this is 2, this is 3, this is find Q₁ and Q₃. So this is Q₁ in this case, which is the Q₁.

Now 1, 2, 3, 4, 5. There are 5 values, so the middle point is this one and in this one, the middle point is 18. So right now $Q_1 = 5$ and $Q_3 = 18$. Subtract Q_1 from Q_3 . So if you divide or by whatever you do so you get this number. So this number in fact in real life what does it mean. (Video Ends: 24:05) It tells you what is the difference of may be the sales between the third quarter and the first quarter.

So if the sales difference is very high, then either there is a seasonality factor or there is some change in consumer behavior, change in some technology that has come up in the market, may be there is a substitute that has emerged which is new and it has changed the entire sales process, something has happened in the market, and now research has to be done what is the reason, why there is so much of variability.

(Refer Slide Time: 24:43)

Standard deviation

 a quantity expressing by how much the members of a group differ from the mean value for the group.



Similarly standard deviation is another important measure that you need to understand. A quantity expressing by how much the members of a group differ from the mean value for the group. Now this is population standard deviation, this is sample standard deviation. So you need to understand. Population standard deviation when we are taking the observed value minus the population mean. So when we are trying to see how much is the data set or the values in the data set, different or moving away from the population mean. So,

$$\sigma = \sqrt{(\sum (x_i - u)^2/n)}$$

is my standard deviation. If you remove this square root and you just square this one, then it is called as variance, nothing more than that. In the sample case what is happening, instead of u (mu) which you do not have, the population mean you do not have, you take the samples mean only. So

$$\mathsf{s} = \sqrt{(\Sigma(\mathbf{x}_i - \overline{\mathbf{x}})^2 / n - 1)},$$

So this is giving me my standard deviation okay (**Refer Slide Time: 25:49**)



After this measures you have understood, I have briefed you this measures so that you are clear when you are dealing with data analysis what is this quartile mean, what is this deviation mean and all, and how they are important or they are applicable in the marketing research, now coming to the last step of the marketing research process now report preparation and presentation.

(Refer slide time: 26:11)

In any report or presentation, get right to the point in why these results matter to the firm. The tendency for beginners is to present all the answers and all the data, whether relevant or not. The role of the analyst is to boil, filter, and distill, so the audience receives the pure distillation.

Just after you have done your data analysis, you have understood how to analyze the data, then there is something called the report or the presentation which you need to do and you need to do it right. Although I have taken an entire session in the last lecture series so, but I need to just brief you. In any report or presentation once you have done with the other things, you have to get right to the point in why these results matter to the firm or organization or individual, anybody.

The tendency for beginners is to present all the answers and all the data whether relevant or not, this is a general tendency. If we tend to write everything and anything, the role of the analyst is to boil, filter, and distill so the audience receives the pure distillation, thus the only filtered thing is to be given to the management may be, so that it can help in taking a decision.

(Refer Slide Time: 27:09)

Marketing Research Report

A factual message that transmits research results, vital recommendations, conclusions, and other important information to the client, who in turn bases his or her decision making on the content of the report

So what is this report? This report is a factual message that transmits research results, vital recommendations, conclusions and other important information to the client. When you are writing a publication, a research paper also, please keep it in mind do not write everything because the scientific community understands them. You need not be putting in everything, you just interpret the data and write the report and use the data in such a method that they feel it interesting to read.

(Refer Slide Time: 27:34)

Importance of Marketing Research Report

- The client bases his or her decision making on the contents of the report.
- The marketing research report is the product that represents the efforts of the marketing research team, and it may be the only part of the project that the client will see.
- The time and effort expended in the research process are wasted if the report does not communicate effectively.

What is the importance of market research report. It says the client bases his or her decision on the content of the report. Obviously when we even apply for a proposal or we make a report, the client's major interest is to see this report because if they have funded us for any project, they are interested to see what is the final result out of this research which we have funded. It is the product that represents the efforts of the marketing research team and it may be the only the part of the project that the client will see, other things they are not interested.

Time and effort expended in the research process are wasted if the report does not communicate effectively. So I will tell you everything during the lecture series one by one in this lecture series how to even write your values. You suppose you have done a t-test or z-test or any test and then how do you write, how should you write, sometimes people are not very clear how should they write the research values basically. So this is how you should organize.

(Refer Slide Time: 28:38)

Organizing the Written Report

- Marketing research reports are tailored to specific audiences and purposes, and you must consider both in all phases of the research process, including planning the report.
- · Must consider questions such as:
 - What is your purpose?
 - Who is the audience?
 - What are your audience's interests, values, concerns?

Marketing research reports are tailored to specific audiences and purposes and you must consider both in all phases of the research process including planning and report. So you should know who your audience is, must consider questions such as what is your purpose, who is the audience, what are your audience's interests, values, concerns.

(Refer Slide Time: 28:56)

Presenting Your Research Orally

The purpose of an oral presentation is to succinctly present the research information and to provide an opportunity for questions and discussion.

The purpose of suppose it is an oral presentation is to succinctly present the research information and to provide an opportunity for questions and discussion. Your report your presentation should be able to bring in some thoughts into the minds of the audience. She should be able to bring in some questions which they feel they want an answer for.

(Refer Slide Time: 29:27)

Presenting Your Research Orally

- To be prepared follow these steps:
 - Identify and analyze your audience
 - Find out the expectations your audience has for your presentation.
 - Determine the key points your audience needs to hear.
 - Outline the key points so you can easily refer to them.
 - Present your points clearly and succinctly.

To be prepared for a research presentation oral, you should identify and analyze your audience, this is the first thing. Please never talk something which an audience does not understand or it is not interesting for him or her. Find out the expectations they have from your presentations. Determine the key point your audience needs to hear. Outline the key points clearly. Present your points in a very clear manner.

Well this is one case which I had brought, but may be because of shortage of time today, I am trying to close it down, I hope this lecture today was very clear to you in understanding the measures of central tendency, their real life use, and finally I just briefed you and I had not gone in details, but briefed you about how to make a presentation and even make a report so that the audience enjoys it, and it may be your work that you have written is published in a good journal in a good place so that it brings you some laurels to you. Thank you very much.