Describing Variables Numerically

Overview

- The previous lecture focused on how we can describe our sample graphically

- To save space, we can also use numerical values to describe our sample

- These fall into three categories:
  i. Central tendency (i.e. mean, median and mode)
  ii. Spread (i.e. range and inter quartile range)
  iii. Variability (i.e. variance)

Arithmetic Mean

- This relates to the value obtained when the scores on a variable are added together and then divided by the number of scores

- This is the everyday concept of the 'average'

- Five students receive the following marks on a statistics test; 35, 80, 55, 55 and 60

- The mean is calculated by summing these scores (i.e. 35+80+55+55+60 = 285)

- You then divide by the number of scores (i.e. 285/5 = 57)

- The mean test score is therefore 57

Median

- This relates to the central score once the values have been arranged in numerical order

- To calculate the median of our test scores (35, 80, 55, 55 and 60), we first arrange the scores: 35, 55, 55, 60 and 80

- The middle score (or the 3rd score as we have five cases) is 55 and is therefore the median

- Note that if one has an even number of values then the mean of the two central values is taken as an estimate of the median

Mode

- The mode relates to the value which occurs most frequently

- To calculate the mode of our test scores (35, 80, 55, 55 and 60), we simply count how often each test score is achieved

- We can see that the scores of 35, 60 and 80 all occur once whilst two people scored 55

- Our mode value for these tests scores is therefore 55

- Note that the mode is the only measure of central tendency which can be applied to nominal data

Range

- Another way of describing a sample is to report the range

- This relates to the difference between the lowest and the highest value in the sample

- A sample of 8 people are asked 'How old were you when you passed your driving test?'

- Responses: 17, 18, 19, 21, 22, 25, 35 and 41

- As the lowest value is 17 and the highest value is 41 then the range is 24 years

Inter-Quartile Range

- You will see from the previous example that the range is very much influenced by the extreme scores

- Some researchers therefore prefer to use the inter-quartile range which considers the middle 50% of the cases

- You remove the lowest and highest 25% of the sample; in our case this is 17, 18, 35 and 41

- This leaves 19, 21, 22 and 25 and hence the inter-quartile range is between 19–25 = 6 years

Variance

- The third kind of descriptive statistic focuses on the average amount that each case in the sample differs from the mean
- It is an indication of the variability of your data
- You can work out the variance by calculating:
  - The mean
  - How much each case differs from this mean
  - Squaring each of these deviations
  - Summing these squared deviations
  - Dividing the result by the number of cases
- For example, imagine you ask 5 of your friends how long they spend online each week
- Responses (hours) 1, 5, 8, 16 and 20

Example of Calculating Variance

Hours	Deviation from mean (10)	Deviation Squared
1	-9	81
5	-5	25
8	-2	4
16	6	36
20	10	100
Total	0	246
Variance = 246/5 = 49.2		

Conclusion

- We can describe our sample in terms of central tendency, spread or variability
- The mean is the 'average' value
- The median is the 'middle' value
- The mode occurs most frequently
- The range is the difference between the lowest and the highest value
- The inter-quartile range is the range of the central 50% of the sample
- Variance assesses the variability of the data