Descriptive statistics

Descriptive statistics is a discipline quantitatively describing the main properties of a collection of data, e.g. measurement results. Descriptive (also descriptive) statistics try to summarize essential information about the given data with a few numbers and pictures. Descriptive statistics is used by e.g. epidemiology.

Descriptive statistics deals with characters (data) of different nature. Since the character of the character can affect the way in which the statistical file can be described, the data is divided into several **data groups''*, which we can also call* *data scales*. So, for example, for data that has a nominal nature, it does not make sense to describe it using an average or median, only modus can be used as a position measure.

Nominal data

- These are data that are **only descriptive**, they cannot be sorted and if any numbers are assigned to them, it is only to mark them in a certain way.
- These are essentially all epidemiological characteristics (place, time, gender...).
- If we were to take, for example, ethnicity, we have white, black, Hispanic, Asian, Indian...

Ordinal data

- Data that already has a **certain order** (from the English *order* = order).
- For further analysis, it can sometimes be useful to combine the data into categories, e.g. if we code the respondents' education with codes starting from zero (no education) gradually according to the level of education achieved, we can, for example, group bachelor's and master's education for the purposes of analysis.
- An example is the generally known *probability scale* from 1 to 5:

  if, for example, we were to ask in a health questionnaire how many people could be vaccinated if they opened a vaccination station in the neighborhood, we could sort their answers as follows:
  1. he would not go to get vaccinated;
  2. would consider vaccination;
  3. maybe he should go get vaccinated;
  4. he would probably go get vaccinated;
  5. he would certainly go and get vaccinated;

Interval data

- It provides **more information'** than nominal and ordinal scales, specifically, unlike ordinal data, it makes sense to evaluate data not only in the sense of "larger-smaller", but also in the sense of "how much one value is greater than the other".
- Does not have a fixed zero value - the choice of zero is somewhat arbitrary, although there may be good reasons for it.
- Example: IQ scale - The average is 100, this Intelligence Quotient value is the majority of the population. Above-average intelligence has values above 100. Just because someone has an IQ of 0 doesn't mean they have no intelligence, but that the range of this scale is set so that it can't accommodate such a low level of intelligence.'

Ratio data

- The scale on which the ratio data is distributed is often referred to as the ratio scale.
- For this scale, the **zero point is already fixed** and expresses the complete absence of the given value.
- Unlike the interval scale, it also makes sense to talk about how many times one value is larger/smaller than the other.
- Example: weight, mortality indicator, often the number of points from the test.

Links

Related Articles

- Position Measures
- Measures of variability
- Methodology in epidemiology

References

- TIMMRECK, Thomas C.. *An Introduction to Epidemiology.* 3. edition. John and Bartlett Publishers, 2002. pp. 205-210. ISBN 0763700606.

- BENCKO, Vladimír. *Epidemiologie, výukové texty pro studenty 1. LF UK, Praha.* 2. edition. Univerzita Karlova v Praze – Nakladatelství Karolinum, 2002. pp. 16-24. ISBN 80-246-0383-7.