

Data types

Statistical data, i.e. information about one element of a statistical file , can be divided into several groups according to their nature. This division is of great importance, depending on the type of data one can choose, for example, a suitable descriptive statistic (e.g. location measures) or a suitable statistical test for hypothesis testing.

Nominal data

Nominal data are such data that only have the meaning of a certain quality, which is why they are sometimes referred to as qualitative data. There is usually only a finite set of options to choose from. Individual values are incomparable, that is, data cannot be ordered, there is no such thing as "size". A typical example of nominal data is e.g. blood group , family status or perhaps a DNA sequence , i.e. some category (that is why we speak of categorical data). Alternative (dichotomous) data are a special group, they only have the values YES and NO.

In some cases, coding of individual values with numbers is used, but these numbers must be viewed only as symbols. This is because while it is convenient to use some system when coding, it is ultimately irrelevant what value is assigned to which numerical code.

Nominal data cannot be compared and therefore only their frequency in the statistical file can be evaluated. A useful measure of position is the mode .

Ordinal data

Like data, ordinal data represents a nominal selection from some number of possibilities. A significant difference from nominal data is that ordering can be introduced in a natural way, and in each pair of values it is easy to determine which value is larger and which is smaller. It is thanks to this quantification that from ordinal data we are now talking about qualitative data.

A typical example of ordinal data can be, for example, the highest level of education. It is clear that ordinal data do not allow assessing the "distance" of individual categories or values.

Interval data

Interval data are such data in which it makes sense to also evaluate the distances between individual categories or values. An example of such data can be, for example, the temperature measured in the Celsius temperature scale. The temperature difference of 10 °C is always the same regardless of the specific starting temperature. However, interval data does not include a ratio of values; e.g. the statement that the temperature has increased 3x by increasing the temperature from 10 °C to 30 °C is completely wrong from a physical point of view.

Proportional data

Proportional (proportional) data represent data for which ratios of individual values are already defined. This includes actually all physical quantities defined in accordance with SI . Another characteristic of them is that they have a clearly defined absolute zero (ie thermodynamic absolute zero, zero distance, zero mass, etc.).

The importance of distinguishing between ordinal, interval and ratio data lies primarily in the fact that when handling such data, one must always keep in mind the limitations arising from their nature.

Ratio and interval data are usually continuous, their values can change smoothly in a certain interval. On the contrary, ordinal data and even more so nominal data are usually discrete, they acquire only a certain finite number of possible values.

In some cases, continuous data needs to be converted to discrete. The usual way is to simply divide the possible values into intervals, with each interval representing one category. A typical example is dividing the height into intervals of several centimeters.

Links

- ws:Typ data

Related articles

- Statistical hypothesis testing
- Position measures
- Measures of variability

References

-
-

External links

-