Assignment: What are data?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Unit 3**: In this assignment, you will practice with (the type of) variables and their attributes that make up data. | | | | |
|  | 1. [Units of analysis and units of observation](https://vimeo.com/user29453510/review/138071440/80ba1d0d2c) 2. [Measurement level of Variables](https://vimeo.com/user29453510/review/136584711/e01302269e) 3. [Variables as constructions](https://player.vimeo.com/external/233436888.hd.mp4?s=49803c81b9dd5245ccd830d8cb1d5994ff64db28&profile_id=175) 4. [(Broad format) data matrix](https://player.vimeo.com/external/233304476.hd.mp4?s=235a9bc8b20b3ecc690d08720a3d70e0e5a5a409&profile_id=175) 5. [Missing values](https://player.vimeo.com/external/233456604.hd.mp4?s=6de8a13da127e17c96c3c0be5e056151fe440146&profile_id=175) |  | Unit (of analysis); Unit of observation; Variable (attributes / values); Dichotomy (dummy variable); Nominal measure; Interval measure; Ordinal measure; Ratio measure | | |
|  | Babbie, p. 97-105  Babbie, p. 180-184 |  | Form groups of 4 (max), face each other. |  | 90 minutes |
|  | Read this assignment carefully and answer the questions before coming to class.  Bring either a print or digital version of your answers to the lecture. | | | | |

*Make sure you installed the program SPSS on your laptop before this session.*

1. In social research, one studies *variables* that are measured on certain *units*. In the next sentences, please indicate the unit and the variable(s) that are mentioned:

|  |  |  |
| --- | --- | --- |
|  | Units | Variable(s) |
| *In attachment research, the behaviour of a child when the parent walks away is often studied.* | Children | the behaviour of a child,  the parent walking away (or not) |
| *In the Stanford prison experiment, researchers tried to investigate the effect of assigned power (roles) on aspects of behaviour.* | Subjects (individuals) | Assigned power  Behavior (unspecified) |
| *In the criticized Milgram’s Experiments, the level of obedience was studied among individuals.* | Individuals | Level of obedience |
| *An experiment conducted by Solomon Asch gave insight to the fact that most individuals conform to the behaviour of a group.* | individuals | Conform (yes or no) to the behaviour of a group |
| *Ivan Pavlov discovered, in his early experiments, that dogs produce different levels of salivary secretions in response to different foods.* | **D**ogs | Levels of salivary sections  types of foods |

1. **Units of observation** do not have to be identical to the **units of analysis**. While we use the units of observation to collect data, the units of analysis are the units we are interested in. Indicate the units of analysis and the units of observation for the following examples. Note, that units of observation and units of analysis can be identical.

|  |  |  |
| --- | --- | --- |
| **Example** | **Unit of analysis** | **Unit of observation** |
| A researcher wants to investigate the general attitude towards immigrants in Enschede. Therefore, she randomly selects citizens of Enschede to participate in a survey. | Individuals living in Enschede | Individuals living in Enschede |
| A researcher wants to study if certain company characteristics can explain why some companies have a higher percentage of employees on sick leave than other companies. For this, the researcher holds a survey among HR managers of a sample of companies. | Companies | HR managers |
| A researcher wants to compare the average educational level of different groups of ethnic minorities. For this, the researcher uses aggregated data of the CBS (Central Bureau of Statistics) on educational attainment. | Groups of ethnic minorities | Individuals |

1. Failing to distinguish between units of analysis and the units of observation can lead to problems. In general, inference should not be made to lower levels of aggregation. Doing that is referred to as the **ecological fallacy**. On the basis of the following three statements, formulate opposite conclusions about individual level behaviour on the basis of aggregate data. Make clear why the hypothesis may be true.

|  |  |  |
| --- | --- | --- |
|  | *Alternative 1* | *Alternative 2* |
| *In cities with a higher percentages of student residents, more beer per capita is consumed than in cities with a lower percentage of student residents.*  What is the effect of being a student on drinking beer? | *Students drink more beer than non-students.* | students are causing so much nuisance, the non-studying residents have to drink more beer, just to be able to sleep at night, when the students roam the cities. |
| *In cities with a higher percentage of immigrants, support for populist parties is higher than in cities with a low level of immigrants.*  What is the effect of (not) being an immigrant on voting for a populist party? | *Migrants vote for populist parties* | *The more migrants in a city, the more the others vote for populist parties* |
| *(US 1900s) In cities with a high percentage of immigrants, the level of literacy is higher.*  What is the relationship between being an immigrant (again: US 1900s) in literacy? | *Immigrants are more literate* | *Immigrants move to places with high levels of economic growth and these are generally places with a highly literate population* |

1. The values or attributes of variables have to be **complete** and **mutually exclusive**. Indicate for the following examples whether they fulfil these requirements. If an aspect is missing, make clear how you would do that.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Complete** | **Mutually exclusive** | **What would have been better?** |
| Age, measured in years | Yes | Yes |  |
| BMI | Yes | Yes |  |
| Level of happiness, measured on a scale from 1 (not happy at all) to 7 (very happy) | Yes | Yes |  |
| Level of education you followed:  primary education, lower vocational, secondary education, higher education, other | Yes | No, it is possible that you’ve followed several | **Highest** level of education, measured in the following categories: primary education, secondary education, higher education, other |
| Weight,  3-50 kg,  50-100 kg,  >100 kg,  >150 kg | No, < 3 kg is missing | No,  50 kg in two categories  + > 150 and > 100 overlap | Weight, measures in the categories: 0-50 kg, 51-100 kg, 101 - 150 kg, >150 kg |

1. There are five different types of measurement levels: dichotomous, nominal (sometimes called ‘qualitative’), ordinal, interval and ratio. Indicate for the following example variables, which type of measurement level you would choose.

|  |  |
| --- | --- |
| **Variable** | **Level or measurement** |
| Marital Status | Nominal (no order in widowed / divorced) |
| Favourite book | Nominal |
| Temperature in Celsius | Interval |
| Satisfaction, measured on a scale from 1-7 | Ordinal (although, often used as interval) |
| Driver’s license? Yes/No | Dichotomous |
| Student number | Nominal |
| Length, in centimetres | Ratio |

1. Sometimes we cannot assign a value or attribute to an individual, although the variable we are using is complete and has mutually exclusive categories? Why is that?

Because sometimes data are missing, because of item non-response, or because of filter questions.

7.