

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Economic Affairs SECO



The nature of data

Viktoria Lets (NLE, SAFOSO, MSP)



Types of data:

- routine data collection which includes:
- laboratory submissions,
- disease surveillance programmes,
- industry- or farm/ bureau based data recording systems and abattoirs.
- <u>structured data collection</u> has been found to provide a more effective way for regular monitoring of disease / production.



Types of data on dairy sector:

- type of dairy husbandry system,
- breed,
- age,
- parity,
- lactation stage,
- udder and milk abnormalities (injuries, blindness, tick infestation and indurations, swelling, milk clots, abnormal secretion),
- results of laboratory analysis,
- etc.



Organizing Data

Whether you are conducting routine surveillance, investigating an outbreak, or conducting a study, you must first compile information in an organized manner.

A <u>variable</u> can be any characteristic that differs from person to person, such as height, sex, smallpox vaccination status, or physical activity pattern.

The <u>value</u> of a variable is the number or descriptor that applies to a particular person, such as female and never vaccinated.



Types of variables

<u>Categorical</u> variables are also known as discrete or qualitative variables. Categorical variables can be further categorized as either <u>nominal</u>, <u>ordinal</u> or <u>dichotomous</u>.

Continuous variables are also known as <u>quantitative</u> variables. Continuous variables can be further categorized as either <u>interval</u> or <u>ratio</u> variables.



Categorical Variables

Categorical variables are also known as discrete or qualitative variables:

Nominal – variables that have two or more categories, but which do not have an intrinsic order.

For example, we could classify types of enterprises into distinct categories such as farm, privet house hold, agriculture enterprises. So "type of enterprises" is a nominal variable with 3 categories called farm, privet house hold, agriculture enterprises.

<u>Ordinal</u> - variables that have two or more categories just like nominal variables only the categories can also be ordered or ranked.

For example, the age of people – young, adult, old.

<u>Dichotomous</u> – nominal variables which have only two categories or levels.

For example, if we were looking at sex of animals, we would categorize animal as either "male" or "female". Other example - disease status — "yes or no". This is an example of a dichotomous variable.



Continuous Variables

Continuous variables are also known as quantitative variables:

<u>Interval</u> variables are variables for which their central characteristic is that they can be measured along a continuum and they have a numerical value (for example, temperature measured in degrees Celsius or Fahrenheit. So the difference between 20C and 30C is the same as 30C to 40C).

Ratio variables are interval variables, but with the added condition that 0 (zero) of the measurement indicates that there is none of that variable. Other examples of ratio variables include height, mass, distance and many more. The name "ratio" reflects the fact that you can use the ratio of measurements. So, for example, a distance of ten metres is twice the distance of 5 metres.



Quiz:

For each of the variables listed below, identify what type of variable it is?

- A Nominal
- B Ordinal
- C Interval
- D Ratio

_ Date of diagnosis
_ Farm of animal
_ Age
_ Sex
Highest alanine aminotransferase (ALT)



Quiz:

For each of the variables listed below, identify what type of variable it is?

- A Nominal
- B Ordinal
- C Interval
- D Ratio
- ___C__ Date of diagnosis
- A Town of residence
- __**D**__ Age
- ___**A**___ Sex
- ___D__ Highest alanine aminotransferase (ALT)



Diagnostic tests in Epidemiology

- In order to calculate measures of disease frequency, we must be able to distinguish between those who have disease and those who do not
- Diagnostic tools are often utilized to determine disease status
- Clinical examinations, laboratory tests, and biological markers are thought to yield more reliable information, in general, than questionnaires for providing information about the disease or risk factor of interest.
- But is this really true, and how can we tell?



- It is desirable to quantify the relationship between diagnostic test result and "true disease status"
- Validity (accuracy):
- Indication of the extent to which an investigation or measurement conforms to the truth (closeness of test result and true clinical state);
- The capacity of a diagnostic technique to measure what it is supposed to measure;
- <u>Specificity and Sensitivity</u> are indicators of the validity of a test.
- Low validity increases probability of "misclassification"



- Reliability (consistency):
- Degree of fluctation of a test series results based on the same sample around a central measurement (true value);
- Repeatability (precision): consistency of the performance of an instrument when used by the same observer over different time periods (intra-observer reliability);
- Reproducibility: consistency of the performance of an instrument when used by different observers on the same samples (inter-observer reliability)



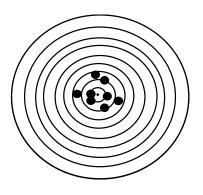
• Bias:

- any systemic error in the design, conduct or analysis of a study that renders results <u>invalid</u>.

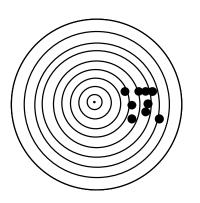
- systematic deviation from the "true" clinical state.



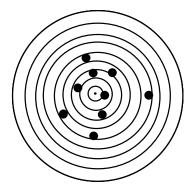
Repeatability (precision), Accuracy and Bias



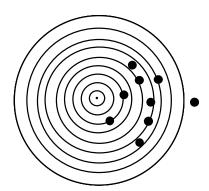
-Relatively precise and accurate



Relatively precise but biased and not accurate



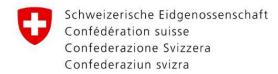
Imprecise and not accurate



Imprecise, not accurate and biased







Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Economic Affairs SECO



Thanks