

Tutorial 1:
Introduction to Data Science

What is Data? - Solutions

Lovnesh Bhardwaj

September 16, 2024

Exercises

1. **Identify Statistical Units**

The statistical units in the table provided are the students Alice and Bob. These units are described by the following features:

- Name
- Age
- Major

2. **Differentiate Data Types**

The following features are classified as:

- Height of buildings in Switzerland: Continuous (Numerical)
- Eye Color of students at USI: Nominal (Categorical)
- Number of Siblings of Data Science students: Discrete (Numerical)
- Favorite Food of University Professors: Nominal (Categorical)
- Pixel brightness value on digital camera: Continuous (Numerical, subject to truncation)
- Final ranking of 24 teams in UEFA Euro: Ordinal (Categorical)
- Survival time of breast cancer patients: Continuous (Numerical, subject to censoring)

A note for the students: We had a brief discussion during the class about why we would want rankings to be ordinal instead of discrete, one argument for them being ordinal can be the fact that rankings provide information about the relative positions (1st, 2nd, 3rd, etc.), but do not offer a measure of how much better or worse one rank is compared to

another. In an ordinal ranking, you know the order but not the magnitude of difference between the positions. For example, if Germany ranked 12th in the Euros, Switzerland ranked 11th, and Italy ranked 13th, this tells you their relative order but not how much better Switzerland was compared to Germany or Italy. In contrast, a discrete scale would involve meaningful, fixed differences between values. However, rankings do not inherently convey such fixed differences, which is why they are categorised as ordinal.

3. Create Data Records

A data table for three employees is as follows:

Employee ID	Name	Department	Salary (CHF)
101	John	Data Science	90,000
102	Jane	HR	80,000
103	Max	Marketing	85,000

4. Create Data Record from Frequency Table

Based on the frequency table:

ID	Major (Bachelor)	Gender	Status
1	Data Science	Male	Admitted
2	Computer Science	Female	Denied
3	Computer Science	Female	Admitted
4	Data Science	Female	Denied
5	Computer Science	Male	Denied
6	Data Science	Female	Admitted
⋮	⋮	⋮	⋮
146	Computer Science	Male	Admitted

5. Create Your Own Data Record

This exercise is up to how you interpret and analyse the picture. One idea is that each picture can be one statistical unit. The features can include the number of people in a painting, the colour palette of the picture, the aspect ratio, etc.