

Introduction to Statistics



1

Statistics

- Statistics is the science of gathering, organizing, analyzing and drawing conclusions from numerical data



2

How Do We Obtain These Numbers? Measurement



3

Units of Statistical Analysis

- Any entities that our data describe (aka units of observation, or cases)
 - Individual people
 - Schools, universities, organizations
 - Geographical areas
 - Countries
 - Entire world at different points in time



"Man, they can't do anything right.
Half the countries are underdeveloped.
The other half are overdeveloped."

4

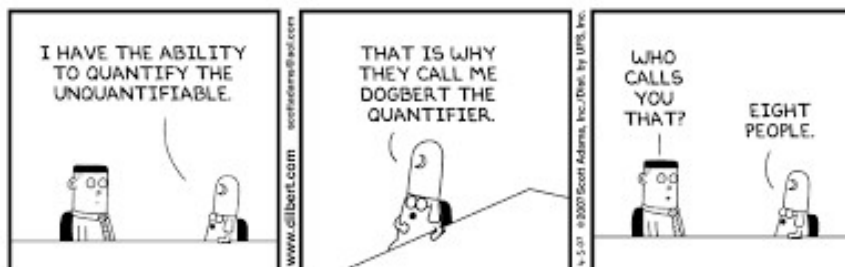
Variables

- We characterize each unit of analysis by a number of traits and attributes = variables
- Variable is any characteristic that can have different values (at least 2)
- Examples: race, income, education

5

Quantitative vs Qualitative Variables

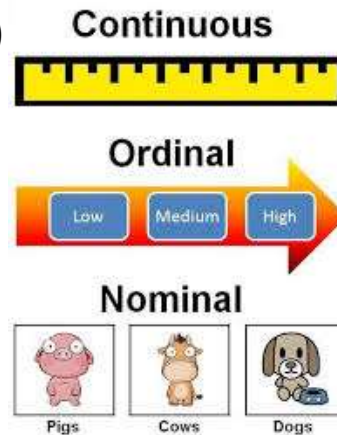
- Variables can be quantitative (measured in numbers) and qualitative (measured in words).
- Statistics are used to analyze quantitative data (or qualitative data that have been quantified)



6

Variables: Levels of Measurement

- Continuous (interval or ratio)
- Ordinal
- Nominal (or categorical)



7

Nominal (or categorical)

- Values vary in quality but not the amount
- Examples: nationality, religion, occupation
- Can be represented by numbers, but still not quantitative (e.g., student ID values)
- Categories should be exhaustive (nothing left out) and exclusive (no overlap)
- Dichotomies (yes/no variables) are always nominal

8

Ordinal

- Numbers denote order, ascending or descending
- Distances between numbers not defined
- Examples: agreement scales, approval scales



9

WebDonuts.com

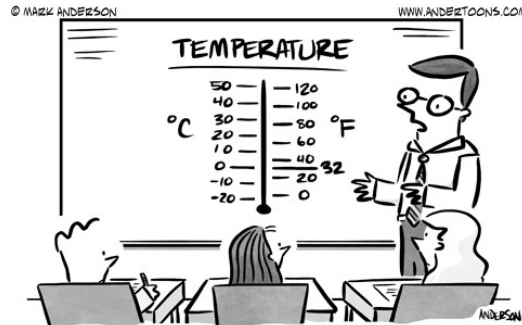


"And with 10 being the highest, you're sure you're only at a 6?"

10

Interval

- Distances between numbers are meaningful
- Difference between 1 and 2 is the same as between 10 and 11
- Example: Fahrenheit temperature scale



"I don't think Fahrenheit is trying to show off, it's just different."

11

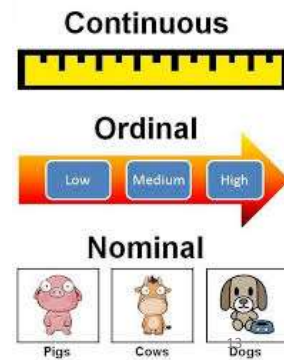
Ratio

- Ratio: has a natural zero point (= total lack of)
- Examples: age, weight, income
- That 0 value may never occur in the data (e.g., one can't weigh 0 lbs)

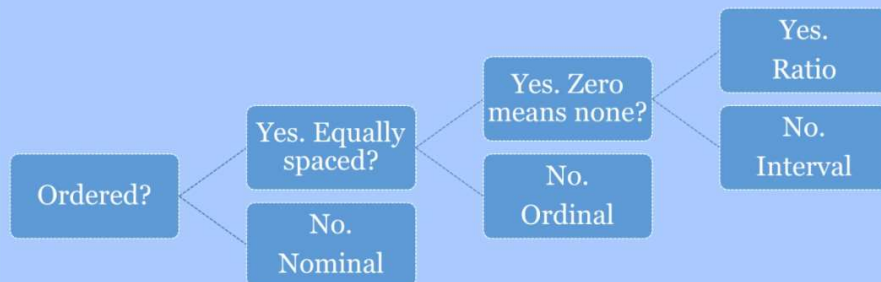


Interval + Ratio

- Interval and ratio variables are analyzed and treated the same in statistics
- Jointly, they are called continuous or scale



LEVEL OF MEASUREMENT DECISION TREE



Survey of Undergrad Students at BC

Survey ID	Age	GPA	State of origin	Satisfaction with major
1	19	3.2	NJ	Very satisfied
2	22	2.5	MA	Neither
3	20	4.0	MA	Somewhat satisfied
4	23	2.2	CA	Somewhat satisfied
5	21	3.7	TN	Very satisfied

Which ones are variables in this dataset? Levels of measurement?

- Year in school (freshman, sophomore, junior, senior)
- Hours per week spent studying
- Major
- Number of students at the university
- Whether or not each student is employed while studying
- Number of roommates
- Whether or not the university is located in the US
- Self-rated stress level (scale 1–100)
- Whether women are more likely to major in Sociology than men
- Whether or not each student owns a car

15