



SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY

COURSE NAME :19CS407 DATA ANALYTICS WITH R
II YEAR /IV SEMESTER

Unit 1- Introduction

Topic : Descriptive Statistics - Scale Types





DESCRIPTIVE STATISTICS





Descriptive Statistics in R



01

Summarizing your Data

02

Getting Average Measures

03

Cumulative Measures in R

04

*Row & Column
Summary Functions in R*



Descriptive Statistics



- ✓ During the time of the Roman emperor Caesar Augustus, an edict to survey the population was issued. It was expected that everyone was covered: the whole population.
- ✓ It is like that in all studies we do about a population, whatever the population is. It can be the people in a country, the employees of an organization, the animals of a zoo, the cars of an institution, the R&D institutions in a country, all the nails produced in a given machine, and so on.
- ✓ But in many situations it is difficult or even impossible to survey all the population. For instance, to collect all the nails ever produced in a machine is typically impossible in practice.



Descriptive Statistics

- ✓ By analyzing a subset of the population it is possible to estimate in a quantified way particular values for the population
- ✓ An example would be the proportion of votes intended for a given party. Generalizing the knowledge obtained from a sample to all of a population is called statistical inference (or induction), since it involves inferring information about the population from a sample

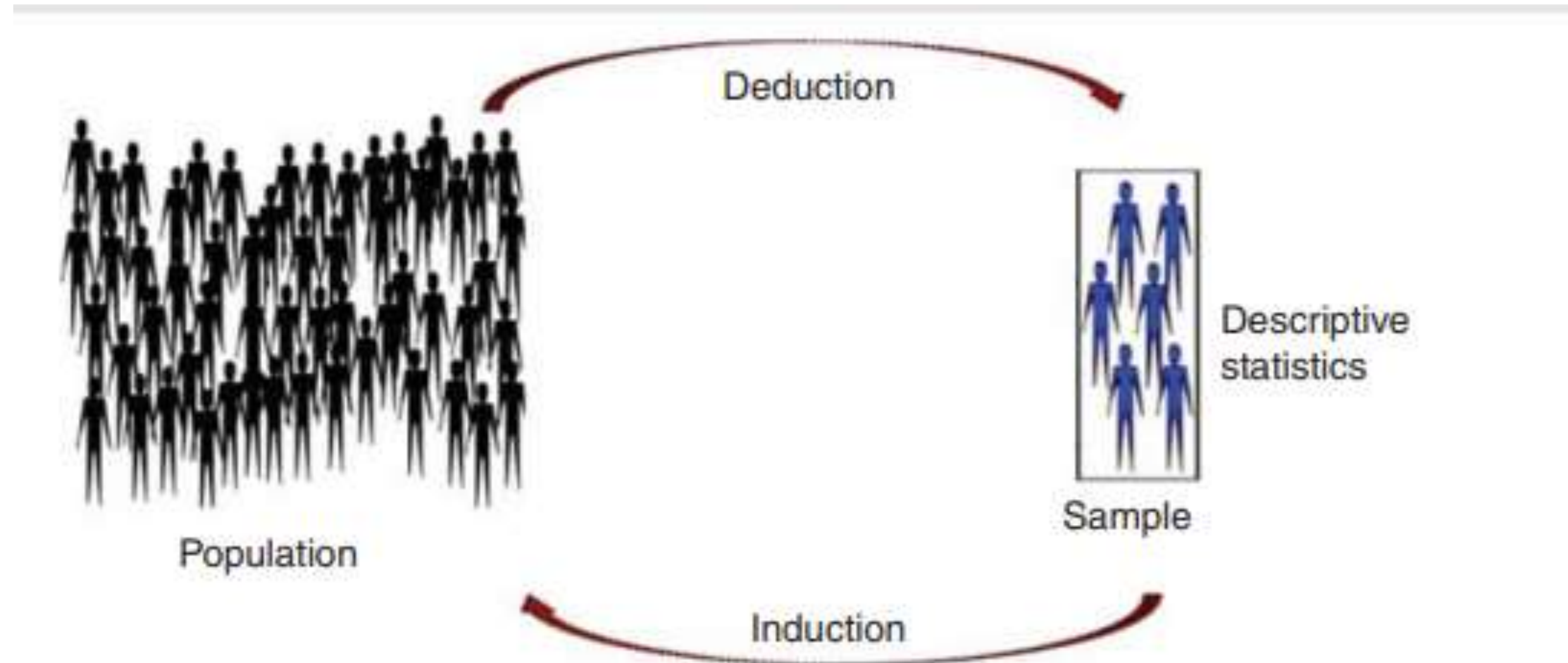


Descriptive Statistics

- ✓ While induction generalizes from the sample to the population, deduction particularizes from the population to the sample.
- ✓ For instance, a deductive problem would be as follows. Given the population of an university, what is the probability of selecting people from two different continents in a random sample of size 10?
- ✓ In other words, knowing the population, the goal is to deduce the nature of a sample of size 10. Yes, probabilities are about deduction



Descriptive Statistics





Descriptive Statistics



- ✓ The ways we have to describe and visualize data are usually categorized according to the number of attributes we are considering.
- ✓ The analysis of single attributes is called univariate analysis, for pairs of attributes it is bivariate analysis, and for groups of more than two attributes it is multivariate analysis.



Scale Types



- ✓ The name of the contact, the maximum temperature registered last week in their town, their weight, height and gender, together with the information on how good their company is.
- ✓ There are two large families of scale types: **qualitative** and **quantitative**.
- ✓ Qualitative scales categorize data in a nominal or ordinal way. Nominal data cannot be ordered according to how big or small a certain characteristic is. But ordinal data can



Scale Types



Table 2.1 Data set of our private list of contacts with weight and height.

Friend	Max temp (°C)	Weight (kg)	Height (cm)	Gender	Company
Andrew	25	77	175	M	Good
Bernhard	31	110	195	M	Good
Carolina	15	70	172	F	Bad
Dennis	20	85	180	M	Good
Eve	10	65	168	F	Bad
Fred	12	75	173	M	Good
Gwyneth	16	75	180	F	Bad
Hayden	26	63	165	F	Bad
Irene	15	55	158	F	Bad
James	21	66	163	M	Good
Kevin	30	95	190	M	Bad
Lea	13	72	172	F	Good
Marcus	8	83	185	F	Bad
Nigel	12	115	192	M	Good



Scale Types



- ✓ **Example** : The name of the contact is expressed on a nominal scale, while the information on how good their company is can be expressed on an ordinal scale because we can define an order of magnitude, ranging from good to bad.
- ✓ Good expresses a higher level of fellowship than bad. This notion of magnitude does not exist in the names.
- ✓ There are two types of scale for quantitative data: **absolute** (ratios) and **relative** (intervals). The difference between them is that in absolute scales there is an absolute zero while in relative scales there is no absolute zero.



Scale Types



- ✓ **Example** : When the attribute “height” is zero it means there is no height.
- ✓ This is also true for the weight. But for the temperature, when we have 0°C it does not mean there is no temperature. When we talk about weight, we can say that Bernhard weighs twice as much as Irene, but we cannot say that the maximum temperature last week in Dennis’ home town was twice that in Eve’s.
- ✓ This is why we usually use a change in temperature to characterize how the temperature varied in a given day instead of a ratio



Scale Types



- ✓ **As an example**, consider the attribute “weight” expressed in an absolute scale in kilograms. We can convert it to any other scale Which values should be assigned to the hyper-parameters?
- ✓ **Relative:** The weight can be converted to a relative scale by, for instance, subtracting a value of 10. The old zero becomes -10 and the new zero is the old 10. That means that the new zero does not mean anymore that there is no weight. The new 80kg is no longer twice the new 40kg. Try to figure out why.



Scale Types



- ✓ **Ordinal:** We can define, for instance, levels of fatness: “fat” when the weight is larger than 80kg, “normal” when the weight is larger than 65kg but less than or equal to 80kg, and “thin” when the weight is less than or equal to 65kg.
- ✓ With this classification, we still have the possibility to define groups of people as being more or less fat. Why have we chosen 65 and 80kg as the levels of fatness? There was no special reason, but there is a rationale



Scale Types



- ✓ **Nominal:** We can transform the previous classification – fat, normal and thin – into B, A and C, respectively. With such a classification it is not possible to order the contacts according to how fat they are, because B, A and C do not quantify anything.



Scale Types



- ✓ **Be attentive:** an attribute can be expressed as a number but the scale type does not have to be quantitative. It can be ordinal or even nominal. Think about a card you have with a numeric code. What kind of quantitative information does it contain?
- ✓ The answer is “nothing”: it is just a key. Its value may eventually express how old the card is but, typically, nothing more than that.
- ✓ If it was a code with letters it would contain the same information



Assessment 1



To classify the data set in different values





References



1. João Moreira, Andre Carvalho, Tomás Horvath – “A General Introduction to Data Analytics” – Wiley -2018

Thank You