

STATISTICS: CHAPTER 1

DAVID

Remark 0.1. *Statistics is split into two branches: Descriptive and inferential. The first is when the purpose of an investigation is to describe the data that has or will be collected. The latter is to give a conclusion of a much broader group using a smaller one, to infer. This smaller group is a sample while the larger is a population.*

Remark 0.2. *When a characteristic of a person or object changes, it is known as a variable, while the things that don't are known as constants.*

Exercise 0.3. Identify some of the variables and constants in a study comparing the math achievement of tenth grade boys and girls in the southern United States.

Solution 1. Variables: The gender, the achievements. Constants: Location, grade level.

We need some sort of measurement, this determines what operation may be used.

(1) *Nominal Level - This involves whether two objects are similar or dissimilar; example short or non-short, college student or non-college student and so on. This does not focus on comparing two things just if a characteristic is present so it seems. In this case if elements are similar they are given the same numerical number, else different.*

(2) *Ordinal Level - This measurement involves on similar or dissimilar but also on ordering such characteristic. For example, what the success in college student, we would rank them. The numbers are relevant and they must have some ordering of the categories.*

(3) *Interval Level - In here, we do a similar approach to Ordinal Level, except that the difference between assigned number gives some sort of indication. Basically it seems that the numbers given must be relevant. Thus addition and subtraction leads to justifiable conclusions.*

(4) *Ratio Level - this seems to be the same as interval except the numbers now have some structure? 0 must represent none or the beginning. This allows us to use multiplication.*

It is clear that ratio level is much stronger, the level of usefulness is ascending.

Exercise 0.4. Identify the level of measurement (nominal, ordinal, interval, or ratio) most likely to be used to measure the following variables:

- (1) Ice cream flavor
- (2) The speed of five runners in a one-mile race, as measured by the runners' order of finish, first, second, third, and so on.
- (3) Temperature measured in degrees Celsius.
- (4) The annual salary of individuals.

Solution 2. Nominal, Ordinal, Interval, Ratio

39 **Remark 0.5.** *Discrete and continuous variables, one is in integers, the other is in*
 40 *any $x \in I \subseteq \mathbb{R}$.*

41 **Exercise 0.6.** *Let our population consist of all eighth grade students in the United States, and*
 42 *let X represent the region of the country in which the student lives. X is a variable,*
 43 *because there will be different regions for different students. X is not naturally nu-*
 44 *merically valued, but because X represents a finite number of distinct categories, we*
 45 *can assign numbers to these categories in the following simple way: 1 = Northeast,*
 46 *2 = North Central, 3 = South, and 4 = West.*

47 **Solution 3.** Obviously, X is discrete and X is nominal level.

48 *The following is all answers to the book chapter one, along with code.*

49 *10 (a) Look at the dataset, it says \$id, thus id. (b) next is asvmath8, the*
 50 *variable type seems to be Factor w/2 levels "No", "Yes": I believe this is*
 51 *just like python or such. We have Factor? not sure what it means, but i*
 52 *assume it means we have two choices? by the w/2 levels then 1 = "no" and*
 53 *2 = "yes". We can do NES? to see the sort of data legend? Basically it*
 54 *is if they took adv mathematics or not in 8th grade. (c) It is 48, it states*
 55 *it. We can also as the book hints use dim(NELS), im assuming to get its*
 56 *dimension.*

57 *12 This uses Ratio, we could use ordinal and put them into groups.*

58 *I regressed from doing the rest as they were trivial, also the answers are online.*