CHAPTER ONE

Scales of Measurements

Introduction

Statistical analysis is based on observations from everyday life. These observations attempt to document the systematic or random changes of defined characteristics or conditions of individual or groups of individuals with similar distinctiveness we called variables. Data collection or gathering of observations require that we make measurements in the form of numerical values. Various measurement scales involve either categorization of events (qualitative measurements) or using numbers to characterize the size of events (quantitative measurements). All statistical analyses require numerical computations. These computations employ a standardized notation system for statistical procedures. One important computational method used for statistical analyses is the summation notation.

In this chapter we will develop the concepts of variables (properties of objects or events), scales of measurements (numerical categorization or quantification of data), and the summation notation ($\sum$).

Variables

A variable is a property (characteristic or condition) of an object or event that changes or can take on different values for different individual.

The weight of individuals is a variable; weight can take on different values for different individual. Height is also another characteristic of an individual that can be measured. In statistics, we also distinguish between different kinds of variables from many perspectives. Some characteristics of an object or event do not vary, but remain the same for every individual; these are called constants.

A constant is a property (characteristic or condition) of an object or event that does not vary but is the same for every individual.

The number of fingers for all normal human infants is ten; this is a constant. Throughout this text and in many statistical textbooks, variables are represented by uppercase letters, for example, X and Y. An individual value of that variable is represented by the letter and a subscript, example, X_3; this might represent the weight of student number three (whose name might be Carl or Susan) enrolled in a specific statistics class of 25 students.

Discrete and Continuous Variables

Discrete variables are variables that take on a small set of possible values or consist of separate, indivisible categories; no values can exist between two neighboring categories. Examples of discrete variables are: 1. days of the week {Sun, Mon, Tue, Wed, Thurs, Fri, Sat or 1, 2, 3, 4, 5, 6, 7}, 2. the number of individuals in a family unit {Mother + Father + Children}, or 3. types of stress levels {high, medium, low}.

Continuous variables are variables that can take on any value or an infinite number of possible values that fall between any two observed values. A continuous variable is divisible into an infinite number of intermediate values or fractions of values. Examples of continuous variables are: 1. weight {1.201 pounds, 1.211, pounds, 1.22134... pounds, etc}, and 2. walking speed {2.41 mph, 3.05 mph}.

Independent and Dependent Variables

Independent variables are those variables controlled or manipulated by a researcher. In behavioral scientific research, the independent variable consists of the two (or more) treatment conditions to which subjects of an experiment are exposed to.

Dependent variables are variables being measured or observed (data or scores) in order to assess the effect of treatment. A researcher may which to study the effect of room temperature (dependent variable) on student's ability to recall (independent variable) a number of items during a memory test.

The changes in the measure of the dependent variable *depend* upon the changes in the independent variable. A researcher may study how hours spent doing homework assignments (independent variable) affects students performance (dependent variable) on an exam.

Since numerical data collection is the basis for recording values or scores from various observations, some knowledge of the scales of numerical measurements used to classify or quantify these values is important. Depending on the classification of numerical data into one or several measurement scales, we are able to determine how such data can me manipulated or summarized mathematically (summed, averaged, divided, etc.) and how the results of such manipulations or summaries are interpreted.

Tony and Variables

Problem 1.1: In an attempt to measure social behavior, a researcher conducts a study to examine the amount of food consumed (caloric intake) of male and female subjects in the presence of a person of the same gender, or a person of the opposite gender.

Tony: “Rose, what are the independent variables in this study?”

Rose: “The *gender of the subjects*, male and female, and the *gender of the other person*; for the researcher will collect data about male and female eating habit or consumption in a social context to make some statement about social behavior.”

Tony: “What is the dependent variable?

Rose: “Social behavior in the form of caloric intake in the presence of same or opposite gender.”

Tony: “Are there any discrete or continuous variables in this study?”

Rose: “Yes, gender can only have a limited number of possible values or classifications {Same gender, Opposite gender} or combinations of four categories if the researcher collects data at the level of gender eating/gender present in the social context {Male/Male, Male/Female, Female/Male, Female/Female}; both of these are examples of discrete variables, since they have a limited set of possible values with no intermediate possibilities.” Caloric intake, on the other hand, is a continuous variable, since its measurement can have an infinite number of possible whole and fractional parts.”

Scales of Measurements

Measurement is defined as the assignment of numbers to objects or events according to prescribe rules. Once assigned, these numbers have certain properties of which we must be aware of as we perform arithmetic or mathematics operations.

Nominal scale is the assignment of numbers for the sole purpose of **differentiating** one object from another. Joe's book locker is labeled 50 which differentiates his from Sam's whose locker is labeled 80. Assigning "1" for *female* and "2" for *male* in order to categorize gender in a survey is a nominal assignment of a number. Nominal assignments are not subjected to arithmetic manipulations.

Ordinal scale is the assignment of numbers for the purpose of **differentiating** between objects as well as showing the **direction** of the difference between them. The ranking of objects or events is a good example of an ordinal scale. On a survey questionnaire one may assign "1" for *Low Sociability*, "2" for *Average Sociability*, and "3" for *High Sociability*. We can now use *more than* or *less than* terms to compare numbers. One cannot say that a person with an *Average Sociability* assignment of "2" is twice as sociable as a person with an assignment of "1".

Interval scale is the assignment of numbers to **differentiate** and assess the **amount** of the difference between objects or events in **equal intervals**. A good example of an interval scale is measurement of temperature on the Fahrenheit (F) or Celsius (C) scale. We can say that a temperature increase from 20^0 to 40^0 F is twice as much increase as from 50^0 to 60^0 F. But we cannot say that 40^0 F is twice as hot as 20^0 F.

Ratio Scale has all the characteristics of the **interval** scale plus an **absolute value** point. An absolute value point allows us to make statements involving ratios of two numerical observations, such as “twice as long” or “half as fast”. The zero point for the Fahrenheit temperature scale has an arbitrary zero point; therefore, we cannot say that 80^0 F is twice as warn as 40^0 F. If it takes Joe 6 minutes to run a mile and Sam takes 12 minutes to run a mile, we can say that Sam is twice as fast as Joe because 0 minutes is an absolute value point; this assignment belongs to a ratio scale.

Most physical scales such as time, length, and weight are ratio scales, but very few behavioral measurements are of this type. In SPSS-format data files interval and ratio variables can be classified under the **Variable View** as *Scale* measurement, nominal variables as *Nominal*, and ordinal variables as *Ordinal* measurements, respectively. Numeric variables without defined value labels and more than a specified number of unique values are set to scale.

Nominal scale: Numbers used only to distinguish among objects or events

Ordinal scale: Numbers used only to place objects or events in order (rank objects)

Interval scale: Number scale on which equal intervals between objects represent equal differences in magnitude – differences are meaningful

Ratio scale: An interval scale with a true absolute zero point – ratios of numbers are meaningful or represents ratios of magnitude

Tony and Scales of Measurements

Problem 1.2: Indicate whether the tasks described would involve nominal, ordinal, interval, or ratio scaling.

a. Measuring the time (to nearest second) for a 1-mile run.

b. Assigning room numbers at a new motel.

c. Measuring Square footage in classrooms at an elementary school.

d. Assigning grades on a multiple choice test taken by a class in Canadian History.

e. Ranking a French poodle as the best of breed at a dog show.

f. Explaining the Fahrenheit temperature scale.

g. Choosing the best cherry pie from 5 entries at the country fair.

h. Distributing automobile license plate numbers at a motor vehicle office.

Tony: "How can these tasks be classified into the appropriate scales of measurements and what are rationales of such classifications?"

Rose: "I would use the definitions of scales of measurements and try to determine whether the data being manipulated by each task is trying to distinguish its uniqueness, rank or order it, involves an interval scale where the magnitude of differences are meaningful or, involves an interval scale where ratios are meaningful."

(a) Ratio scale, since the ratio of time is meaningful (I can say that one time is faster than another)
(b) Nominal scale, since room number is only for the purpose of distinquishing between rooms.
(c) Ratio scale, since I can say from the ratio of square footages of one room and another that one is twice as large as the other, etc.
(d) Interval scale, since grade perfromance or acievement tests are such (there are no absolute zeros for these scales, but differences are meaningful)
(e) Ordinal scale, since ranking or ordering is important here
(f) Interval scale, since magnitude of difference is meaningful and there is not absolute zero for this measurment unit ("0" degree F is different from but relative to "0" degree C)
(g) Ordinal scale, since this involve ranking (or preference of what is best) only, and ,
(h) Nominal scale, since numbers on license places are for sole purpose of distinquishing between automobiles.

Problem 1.3: Classify the following types of data as nominal, ordinal, interval, or ratio.

a. Fourth graders at an elementary school who have not had chicken pox.

b. Number of items correct for 27 students on a mathematics aptitude test.

c. Time required depressing the brake pedal in a driving simulator after consuming 3 ounces of alcohol in 30 minutes.

d. Tags with pet identification numbers distributed after rabies vaccination.

e. Daily high temperatures for the month of August in degrees Celsius.

f. Position at the finish line for the first four runners in a 1500-meter race.

g. List of heights of 52 Colorado mountain peaks that are over 14,000 feet.

Tony: "How would you classify the above data?"

Rose: "By assessing if the data is a label, ranking of objects, magnitude of differences are meaningful, or the ratio of the data type is meaningful."

(a) Nominal scale, since trying to indentify an object, first graders
(b) Interval scale, difference between numbers are meaningful
(c) Ratio scale, since time or speed (response per 30 minutes time interval) is such that ratios are meaningful; faster or slower statements invloves ratios
(d) Nominal scale, since number is used for identification or lable
(e) Interval scale, since temperature scales or interval scales without absolute zeros
(f) Ordinal scale, since this involve ranking (relative positions; first, second, etc.), and ,
(g) Ratio scale, since ratios of heights are meaningful (twice as tall, 3 times as high as, etc.)

Summation Notation

One of the most common symbols used in statistical analysis computations is the uppercase Greek letter **sigma** (∑). This notation for summation means to "add up" or "the sum of". So ∑*X* reads, "Sum of all *X*". The precise notation for summing *n* values of *X* is

$$\sum_{i}^{n} X_i$$

Given X = **{**1, 2, 5, 10, 12**}**, find ∑*X*.

Note. "**{ }**" is a container used in set notation that lists all the elements or possible values of a variable; this listing can contain the entire population or a subset (or sample) of it.

Answer:

$$\sum_{i}^{n} X_i = \sum_{1}^{5} X_i = 1 + 2 + 5 + 10 + 12 = 30$$

Sigma (∑): Symbol indicating summation

The summation sign or symbol, ∑, is often used with other mathematical operations, such as multiplication or squaring. To obtain the correct answer, it is important that different mathematical operations be done in the correct sequence or order of operations. Below is a listing of the correct order when performing mathematical operations.

Order of Mathematical Operations

1. Any calculation within a parenthesis is done first.
2. Squaring *(*or other exponents like X^5*)* is done second.
3. Multiplying and/or dividing is done next (in order from left to right).

4. Summation using the symbol, ∑ is done next.
5. Finally, additions and/or substations are done.

Tony and Summation Notation

Problem 1.4: Given the following data set for *X* and *Y* variables, compute the following summation operations: (a) $\sum X$, (b) $\sum X^2$, (c) $\sum XY$, (d) $\sum(X-Y)$, (e) $\sum(X-Y)^2$.

X	*Y*
7	5
3	1
8	6
6	4
4	3

Tony: “What is the best way to do statistical computations involving summations?”

Rose: “If the dataset(s) is small, a computational table is advised, but for large dataset, you can use a spread sheet like MS Excel or a statistical program to write the summation formulas for each summation operation. I will illustrate a use of a computation table for finding all the summations of above.”

X *(A)*	***Y*** *(B)*	***X²*** *(A)²*	***XY*** *(A)(B)*	***X-Y*** *(C)* *(A) – (B)*	**(*X-Y*)².** *(C)²*
7	**5**	49	35	2	4
3	**1**	9	3	2	4
8	**6**	64	48	2	4
6	**4**	36	24	2	4
4	**3**	16	12	1	1
$\sum X$	$\sum Y$	$\sum X^2$	$\sum XY$	$\sum (X-Y)$	$\sum (X-Y)^2$
28	19	***174***	***122***	***9***	***17***

Answers (a) $\sum X$ = ***29***, (b) $\sum X^2$= **174** (often called *sum of square*)

(c) $\sum XY$ = ***122*** (often called *sum of product square*), (d) $\sum(X-Y)$ = **9**,

(e) $\sum(X-Y)^2$.= **17** (often called *sum of difference squared*)

Note. That $(\sum X)^2 = (28)^2$ **= 784** and $\sum X^2 \neq (\sum X)^2$