MATH 124 Spring 2005

Lecture: 2
Date: Feb 3, 2005

Skills you should acquire from this lecture:

- Definitions of population and sample and knowledge of how they are used in statistics.
- Familiarity with summation notation.

Related readings in the textbook:

Math 124 Lecture 2

Where are we at and where are we heading?

Last time we talked about data and how statistics was about collecting, organizing and understanding data. Today we will expand on these issues and introduce some notation that we will use for the rest of the class.

Two different types of statistical analysis

- **Exploratory Data Analysis** sometimes called *descriptive statistics* is about investigating the values of the measurements of the variables in your dataset. This could involve examining each variable by itself or looking at relationships between variables.
- **Confirmatory Data Analysis** sometimes called *inferential statistics* is the about using information from the sample to draw conclusions about the population from which it is drawn.

What will we do in this class?

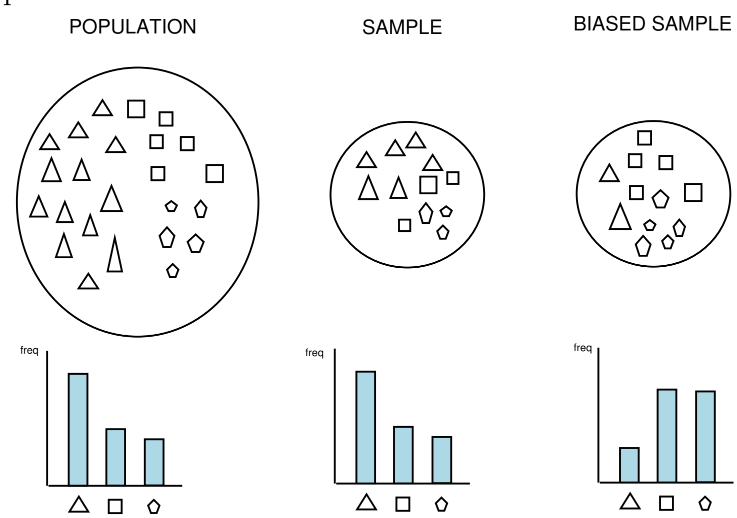
We will investigate both Exploratory Data Analysis (EDA) and Confirmatory Data Analysis. We will spend the next sequence of lectures exploring topics in EDA and leave CDA for later in the semester.

Samples and Populations

Next, lets define these two terms:

- **Population** a well defined set of set of objects/units/individuals that we are interested in studying eg All adults in the USA, all widgets produced at a factory in a month, all bottles filled in a day at a soft drink company
- **Sample** a subset of the the population eg the first 1000 people selected by randomly choosing from valid social security numbers, every 10th widget off the assembly line, the bottles filled in the first hour of the day

A study of all the elements of a population is called a *census*. In an ideal world for anything which we wanted to study we could carryout a census, unfortunately this is almost never possible. Why? Cost or time prohibitive or perhaps our measurement method is destructive. So instead we usually look at data based upon samples chosen from the population. One property that we would like is that our sample reflect the population from which it is drawn so that any conclusions we make based on the sample can be generalized to the population. We would say that a sample that does not reflect the general population as biased.

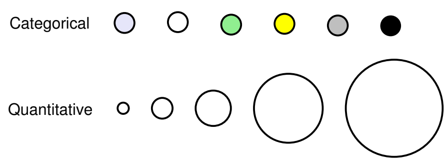


Variables

- **variable** a characteristic which may change between different objects in a population.

  Categorical Variable a variable that places each individual into one of a specified number of categories. This may or may not be ordered. eg high or low, eye color, gender

  Quantitative Variable a variable which takes numerical values. Something for which we could add or average. eg heights, volumes, weights


Categorical
Quantitative

Notation

When talking about variables in statistics we like to simplify our discussion by using symbols. Specifically we use letters (eg "x", "y", "z", ...) to represent data

(later in the class we will use upper case letters to denote Random Variables, but don't worry if you don't know what these mean just yet). In addition we use a subscript on the letter to represent the observation number.

eg1 Suppose x = "Height"

then x_5 would be the 5th data value recorded for height

x_{15} would be height for 15th observation

x_i would be height for ith observation eg2

x	y
3	5
1	2
6	6
10	4
7	8

$x_1 = 3$

$y_4 = 4$

$(x_2, y_2) = (1, 2)$

Why do we do this? So that we can easily write formulas.

Summation Notation

Very often in statistics we want to add up a set of similar things. For example, add 10 observations of the variable x. We could write this as

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}$$

Kind of long isn't it? Now ~~i~~magine instead of just 10 terms it was 100's or 1000's of terms. It would take a long time to write this out.

However, instead of doing it out the long way we use a mathematical short hand which we call summation notation:

$$\sum_{i=1}^{10} x_i \quad \text{is short for} \quad x_1 + x_2 + \cdots + x_9 + x_{10}$$

(this is the value we should substitute for i in the first term of the expansion)

(Σ is the greek symbol capital sigma it can be read in this context as "sum of")

(This should be the value we substitute for i in the last term of the expansion)

Note 1. from the first value to the last value we keep incrementing by 1.

Note 2. If the summation symbol appears without lower and upper bounds then it means to sum over all possible values of i.

Examples

x	y	z
10	-1	2
12	0	3
15	1	5
17	2	7
19	-2	11
21	0	13

As self exercise find:

$$\sum x_i \,,\quad \sum_{i=2}^{4} x_i \,,\quad \sum_{i=1}^{6} y_i^2 \,,\quad \left(\sum y_i\right)^2 \,,$$

$$\left(\sum_{i=1}^{4} x_i\right)\left(\sum_{i=1}^{4} y_i\right),\quad \sum_{i=1}^{4} x_i y_i \,,\quad \sum_{i=1}^{6} (y_i + z_i) \,,\quad \sum_{i=1}^{6} z_i^{\,y_i}$$

Extended solutions

$$\sum x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$
$$= 10 + 12 + 15 + 17 + 19 + 21 = 94$$

$$\sum_{i=2}^{4} x_i = x_2 + x_3 + x_4 = 12 + 15 + 17 = 44$$

$$\sum_{i=1}^{6} y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 + y_6^2$$
$$= (-1)^2 + 0^2 + (1)^2 + 2^2 + (-2)^2 + 0^2$$
$$= 10$$

$$\left(\sum y_i\right)^2$$
$$= (y_1 + y_2 + \cdots + y_6)^2$$
$$= (-1 + 0 + 1 + 2 + -2 + 0)^2$$
$$= 0^2 = 0$$

$$\left(\sum_{i=1}^{4} x_i\right)\left(\sum_{i=1}^{4} y_i\right) = (x_1 + x_2 + x_3 + x_4)(y_1 + y_2 + y_3 + y_4)$$
$$= (10 + 12 + 15 + 17)(-1 + 0 + 1 + 2)$$
$$= (54)(2) = 108$$

~~Note in general~~

$$\sum_{i=1}^{4} x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4$$
$$= (-1)(10) + (12)(0) + (15)(1) + (17)(2)$$
$$= -10 + 0 + 15 + 34 = 39$$

Note that in general $\sum x_i y_i \neq \sum x_i \sum y_i$

$$\sum_{i=1}^{6} (y_i + z_i) = (y_1+z_1) + (y_2+z_2) + \cdots + (y_6+z_6)$$

$$= (-1+2) + (0+3) + (1+5) + (2+7)$$

$$+ (-2+11) + (0+13)$$

$$= +1 + 3 + 6 + 9 + 9 + 13$$

$$= 41$$

$$\sum_{i=1}^{6} z_i^{y_i} = z_1^{y_1} + z_2^{y_2} + \cdots + z_6^{y_6}$$

$$= 2^{-1} + 3^{0} + 5^{1} + 7^{2} + 11^{-2} + 13^{0}$$

$$= \frac{1}{2} + 1 + 5 + 49 + \frac{1}{121} + 1$$

$$= 56\frac{122}{242}$$