Case Studies of Inquiry and Integration in Class Sessions

Class Session 3 in the COEP Course "Thinking, Inquiry and Research

Mohanan

Dear friends,

This is a report on Class sesstion 3 of the CoEP course on "Thinking, Inquiry, and Research", with a few additional examples and pieces of exposition that can function as a "report' for colleagues who were not present at the session, simultaneously serving as class notes for the students who were present in class. For me, this is also a set of notes on what can become a chapter in a textbook on a trans-disciplinary introduction to research.

You will recall that class session 1explored the concepts of circles, triangles, rectangles, lines, straight lines, and points in geometry as the basis for a brief experience of engaging with the construction and evaluation of definitions in mathematics. Class session 2 extended the core of the strategy used for the investigation of geometric concepts to the construction and evaluation of the definition of the concept of success in the humanities. The purpose of the switch of conceptual inquiry to from straight lines and points in math to success in human affairs was to sensitize students to the trans-disciplinary threads that integrate inquiry inquiry in apparently diverse domains.

The primary chunk of session 3 was an introduction to observational inquiry as one of the two important strands of scientific inquiry. This was sanwitched between conceptual inquiry into "What is the distinction between solid and liquid?" before observational inquiry, and "What is caste?" after it. Let me report on the slice on solid vs liquid first.

What is Solid?

For the first 20-30 minutes of class session 3, we used the strategies of conceptual inquiry to investigate a physics-chemistry question "What is the distinction between solid and liquid?" The first goal was to demonstrate that we often entertain illusions of knowledge, which, when interrogated, turn out to be ignorance covered in words that we do not understand. (For details, watch the demonstration in the video "What is a solid?" at https://www.youtube.com/watch?v=mL_BdkaHDq8) In the CoEP session, however, we proceeded to explore:

a) every day ('commonsense') concepts embedded in ordinary language on the one hand, and academic concepts on the other (e.g., 'animal' in ordinary English, 'jaanwar' in Hindi vs 'animal' in biology)
b) the every day distinction betwee solid and liquid based on tactile-kinesthetic experience (a congenitally blind person can experience the distinction between solid material and liquid material), in contrast to the textbook definition based on visual experience (shape),
c) the need to reject the textbook definition, and
d) the need to doubt and question textbooks, teachers, other 'authorities', and ourselves.

We left the investigation of the academic concept of solid and liquid for subsequent exploration. [I am currently thinking of using the following scenarios when we come back to the solid-liquid issue. Imagine that you are lying down on a frozen swimming pool when it is completely dark. The body of the water does not yield to the pressure from your body, and you do not sink. You will judge the water to be solid. If on the other hand, the water is not frozen, your body goes through the water body, and you will judge it to be liquid. If the water were replaced by very soft rubber or foam, even though the body of rubber/foam yields to the pressure from your body, you will still judge it to be solid. Why? What is the difference between the way unfrozen water, foam, wheat dough made for making bread, and batter made for making dosa yield to the pressure from your finger? Does unfrozen water behave as solid for waterspiders that walk on water (https://www.youtube.com/watch?v=kx02qT2p5uk) and what are called 'Jesus lizards' run on it(https://www.youtube.com/watch?v=45yabrnryXk)? Does what is called 'oobleck' behave like liquid when you stand on it and behave like solid when you run or jump on it (https://www.youtube.com/watch?v=yHlAcASsf6U)? Please watch these videos: they are incredibly interesting.]

Observational Inquiry

Strands of Scientific Inquiry

After the conceptual inquiry into the distinction between solid and liquid states of matter, we turned to scientific inquiry proper. The core components of scientific inquiry are:

- singular observation (conceptualized as data points in statistical research)
- observational generalizations (conceptualizable as data lines or data regions)
- causal generalizations (observational generalizations supplemented with causal relations) and
- theories (made up of (i) theoretical concepts and (ii) model, laws, or a combination of the two) that explain observational generalizations

The first two bullets cover observational inquiry, and the last one is on theoretical inquiry, with inquiry into causal generalization in the limbo between the two.

Data Points, Scatter Plots, Generalizations, Limits of Variability, and Frequency Distribution

The aim of this part of the class was to initiate students into quantiative and qualitative observational inquiry, such that it can be enhanced with experimental inquiry at a later stage.

We began observational inquiry with the concept of ***data points.*** An example of a data point is the observation that KPMohanan is 180cms tall. Another example would the data point that Tara Mohanan is 170cms tall. Along the dimension (= axis / parameter) of human height, we can represent the location of these two points as follows, with the gray line representing the one-dimensional ***data space*** of human height (= length along the vertical dimension), and the blue dots representing the points in this space:

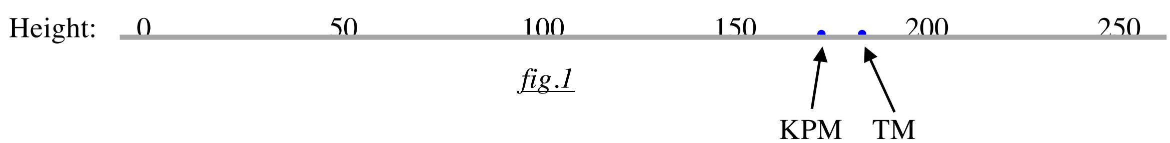


fig.1

If we plot the height of a randomly selected sample of a million human beings as illustrated above, we will have what is called a ***scatter plot*** of dots scattered over a stretch. The dots in the scatter plot constitute a ***region*** in the data space where data points occur, extending from about 50 cms to about 250cms. (The shortest human that we know of is 50 cms, and the tallest about 350cms.)

What this example illustrates is the ***variability*** in the values of (the variable of) height in the human population (the value of human height in centimeters can be 100, 120, 121, 150, 173 etc.), and the ***limits of the variability*** (human height is less than 50 or more than 250). Statistics uses the term ***distribution*** to refer to the region in which data points occur (= are distributted), making a prediction about where the points *cannot occur*.

When we use such a scatter plot of the *sample* of one million humans to conclude that the height of adult human beings ranges from 50 to 250 cms, we are making an ***observational generalization*** about the population of human beings, using a form of inductive reasoning from the sample to the population. In statistics, the relation between the data points and the sample comes under *descriptive statistics,* while the reasoning that extends the properties of a sample of a population as properties of the population comes under *inferential statistics*.

The considerations that cover data points, samples, and population come under ***observational inquiry***, one of the strands of scientific inquiry. This includes not only sensory observation, but also counting, measuring, verbally mediated evidence, statistical inquiry, experimental enquiry, instrumentation, and other research methodologies such as surveys, interviews, case studies, field work, ethnography, and evidence from written texts, covering both qualitative and quantitative data. Some of these topics I hope to pick up at a later stage in the course.

To proceed further, when we compare the number of humans who are about 170cms (about 5 five and a half feet) tall with those who are about 100 cms (a little more than three feet) and 200cms (about six and a half feet) tall, we find that the people who are 100 cms and 200 cms tall are far fewer than those who are 170, and those who are 150 is far more than those who are 170. Such differences in frequency are compared by plotting height along one axis and frequency along another, in a two-dimensional height-frequency space:

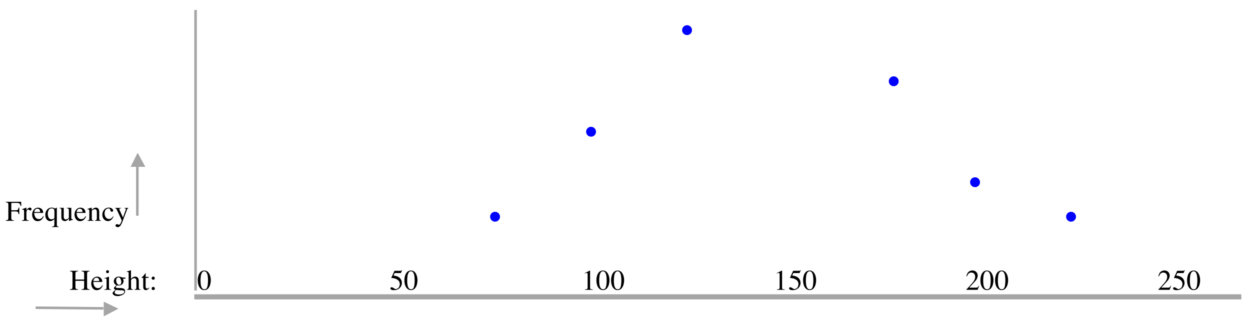


fig.2

The concept of ***average*** refers to the centre of this range (with three different notions of center, namely, *mean, median,* and *mode*.) When we say that the average height of an Indian is about 155 cms (plus or minus 5 cms) we are talking about an observational generalization about the the relative *frequency* of different heights in a population, in contrast the *range* of height in fig.1. The so called 'bell-curve' or 'normal distribution', for instance, is a curve that we get by drawing a line through the data points of frequency distribution. (For details, see the chapter four of Derek Rowntree's *Statistics without Tears* at https://ia601606.us.archive.org/26/items/StatisticsWithoutTears/Rowntree-StatisticsWithoutTears.pdf)

Please bear in mind that figure 2 is a synthesis of statistics and geometry: numerical data viewed from the perspective of space, points, lines, and regions.

Quantitative and Qualitative Generalizations, and Invariance

The above examples of the observational genearalizations on the limits and frequencies of the variability of human height are ***quantitative***, that is to say, they can be expressed as lines or regions provided by the space of numbers. Observational generalizations on the height, weight, volume in the population of adult human beings is expressible in terms of numbers, so these are quantitative observational generalizations. In contrast, the statement that the the natural colour of the iris of the eye in the human population can be black, gray, brown, green, or blue, but not yellow, orange, red, pink, purple, or white is a statement of the limits of variability that is not expressed in terms of numbers, and hence is a ***qualitative*** generalization. The following propositions express qualitative generalizations:

The natural colour of the human hair can be black, gray, white, brown, red, or yellow but not green, blue, or purple.

The natural skin colour of humans can be dark brown, light brown, yellowish, or pink, but not jet black, blue, green or purple.

These are generalizations that express the limits of variability of a variable trait. In contrast, the following generalizations express ***invariance***, absence of variability:

The colour of the human blood is red. (There are no humans with black, white, green or blue blood.)

Every human being has neocortex.

No human beings have wings.

Every adult human being has exactly one heart.

A word of caution. The mere existence of numbers does not make something quantitative. A human being's date of birth, telephone number, and pin code may be given as numbers, but these numbers do not express quantiies and do not lend themselves to quantiative generalizations. (It makes no sense to say that the average pin code number indian houses is such and such. Instead of using people names, we can identify humans using their identify card number, but that does not make the information quantitative.)

The generalization that every adult human being has exactly one heart and the generalization that every adult human has exactly two lungs use numbers, but they are not statistical generalizations. Are they quantitative generalizations?

Two-Dimensional Generalizations

Let us turn to a different kind of observational generalizations. While figure 1 has a one-dimensional height space, figure 2 has a two dimensional height-frequency space. Another kind of two dimensional space is involves a *systematic relation* between two variables, say, height and weight. Let us suppose KPMohanan's weight is 90kg. We could represent this as a data point on the unidimensional weight-space, as we did in figure 1. Alternatively, we could use a two-dimensional height-weight space, a in figure 3:

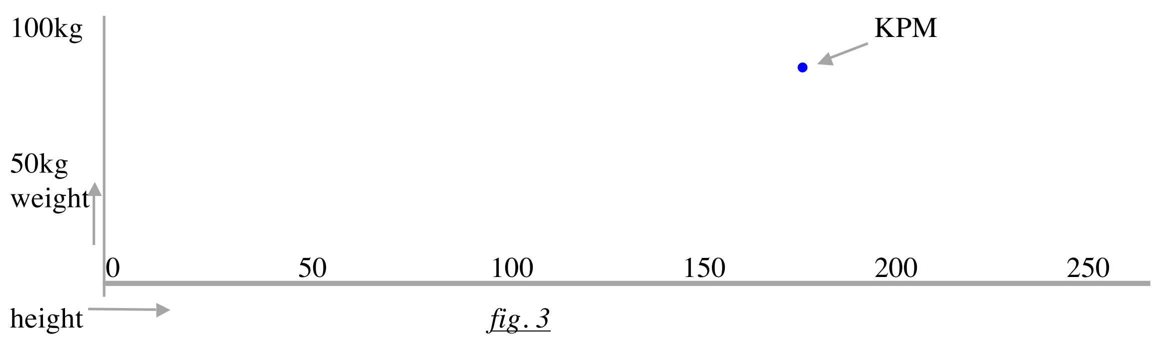


fig. 3

The data point in fig 3 is in a two dimensional space. Hence, to specify its location, unlike the case in figure 1, we need two pieces of information, namely, height and weight. (Please keep track of the connection between statistics and geometry.) Now, if we take a random sample of adult humans, measure their height and weight, and represent the information on each human as a point in a height-weight space, we will see something like the following:

(figure given in a separate file called 'Scatter Plot')

fig. 4

What such a scatter plot shows is that the relation between height and weight is not random: *all else being equal, as one increases, the other tends to increase*. This doesn't mean that in *every case*, if A's height is greater than B's height, A's weight is also greater. What it means is that if if A's height is greater than B's height, then it is highly probable that A's weight is also greater. In such a case, we say that there is a ***correlation*** between height and weight among human adults, a correlation being a systematic relation between two variables x and y, such that if we know the value of x, we can predict the value of y within a certain range.

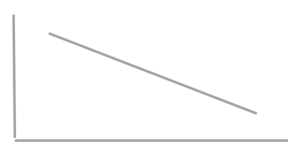
The standard practice in the visual representation of quantitative correlations is to draw a line through the center of the scatterplot (if there is a centre where the points are maximally clustered together, as in the case of the average in the case of the distribution of values in a single variable.)

fig. 5 (**figure given in a separate file called 'scatter plot 1'**)

The correlation between human height and weight is only approximate. This is a correlation in science. In contrast, the correlation in mathematics between the diameter and circumverence of a circle (familiar to us as the formula that says that the circumference of a circle is pi times its diameter) is an exact correlation: given the value of the length of the diameter of a given circle, we can predict the length of its circumference exactly. The correlation between the area of an equilateral triangle and the length of its sides is also an exact correlation.

The generalizaations on the height and weight of human beings, the diameter and circumference of circles, and the area and lengths of equilateral triangles, are examples of ***positive correlation***, where the value of one variable increases when the other increases. In contrast, the observational generalization known as Boyle's law, which says that as the pressure of a given body of gas increases, its volume decreases, expresses a ***negative correation*** between pressure and volume: as one increases, the other decreases.

positive correlation

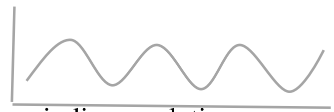


negative correlation

fig. 6 (to be completed)

The examples we have examined so far are ***linear correlations***, that is, correlations that are representable as straight lines. Not all correlations are linear. For instance, consider the correlation between age and height in the human population. Beginning from birth, as age increases, height also increases, but when a person reaches age 20 or so, height stops increasing even when age keeps increasing. This is a curved line, not a straight line. In the literature, it is called a ***non-linear correlation***. (Many people use the word 'line' to mean 'straight line', so in that usage 'non-linear' means 'not a straight line'.)

Some non-linear correlations can be ***periodic***. That is to say, as the value of one variable increases, the other keeps alternating between increasing and decreasing. An example of a periodic correlation is that between time and temperature in any location on the earth. If we measure the temperare at a place at 5am 6am, 7am, 8am, and so on, we find that the temperature steadily increases and reaches a peak in the afternoon, after which it keeps going down and reaches a minimum after midnight. And then it starts going up again. This is the *daily cycle* (24 hour cycle) of temperature-time correlation. If we calculate the mean temperature of each day for over a 10,000 days we find that the mean temperature keeps going up for a certain period, reaches a peak (summer) and then keeps going down, reaching a minimum (winter), after which it goes up again, in a *yearly cycle* (365 day cycle) of temperature-time correlation.



periodic correlation

Similar cycles exist for the duration of daylight, which keeps going up and down in a yearly cycle, the longest day being in summer and the shortest day being in winter. This also means there is a correlation between temperature and length of daylight. Finally, as the length of daylight increases, the length of night decreases, illustrating a negative correlation between the two cycles.

Periodicity, the phenomenon of the repetition of events along time, is a fundamental property of nature. It is found in the position of the moon, the sun, the plants, and the stars, in the waxing and waning of the moon. in the rainy seasons, in the seasonal changes in fruits and flowers, in the beating of the human heart, in the movement of a simple pendulum, and in the pulsing of atoms.

To turn to qualitative correlations, consider the following ones:

Living organisms which produce chlorophyl do not have vertebrae.
Living organisms which have red blood have vertebrae.
Organisms which have vertebrae have eyes.
Organisms which have six legs have compound eyes.

The attributes of producing chlorophyl and having vertebrae have two values each: produces chlorphyl, does not produce chlorophyl, has a vertebra, does not have a vertebra. Of these the combinations [produce chlorphyl + not-have vertebra], [not-produce cholorphyl + not-have vertebra], and [not-produce chlorophyl + not-have vertebra] exist, but the combination [produce chlorophyl + have vertebra] does not exist. If we represent this in a space in which each dimension has two values, then what it says is that no organism can occur in the location [produce chlorophyl + have vertebra]. Similar remarks apply to the rest.

The examples we have considered so far have been of one-dimensional space (e.g., limits of variability and frequency distribution of the values of height in the human population) and two dimensional space (e.g., correlation between age and height in the human population.) Now, this should not give the impression that there are no generalizations in a multidimensional space. The correlation holding among the area of a rectangle and the lengths of its sides calls for a three dimensional space: area = length x breadth. The generalization that says the those who have peptic ulcers have a combination of a specific gene called TLR1 and the bacterium called H pylori (see Gene Linked to Ulcer-Causing Bacterial Infections at http://www.everydayhealth.com/ulcer/gene-linked-to-ulcer-causing-bacterial-infections-6086.aspx) also calls for a three dimensional space of peptic ulcers, TLR1 and H pylori. Thus, generalization (whether mathematical or scientific, whether observational or theoreticsl) reside in an *n*-dimensional space, where *n* can be one, two, or more.

In the preceding discussion, we went through a number of examples illustrating different types of observational generalizations. Many of these were discussed in the class on 9 August, but some were not.

The Challenge of Explaining Observational Generalizations

Towards the end of the session, I touched upon the challenge of constructing theoretical explanations for the observational generalizations. Having pointed out the periodic correlations of the daily cycle of temperature and the yearly cycle of temperature, how do we explain them?

The general consensus in the class was that the explanations for such phenomena are to be sought in a theory of the solar system, the two competing theories being the geocentric theory and the heliocentric theory. The geocentric (geo = earth) holds that

- the earth is the center of the universe,
- the earth is stationary, and
- the sun, the moon, and the planet revolve around the earth. (The earth is not a planet.)

In contrast, the heliocentric theory (helio = sun) holds that

- the sun is the center of the solar system,
- the earth, along with the other planets, revolve around the sun, and
- the earth rotates on an axis,
- the axis of the earth's rotation is tilted to the perpendicular to the plane of its revolution.

Based on one or more propositions of the heliocentric theory, we can construct explanations for the daily and yearly cycles of temperature on the earth. In the class discussion, students came up with three competing claims: The daily cycle of temperature can be explained in terms of

A. the distance between the sun and the location on the earth (without appealing to the angle of incidence of sun's rays)
B. the angle of incidence of sun's rays (without appealing to the distance between the sun and the location on the earth)
C. a combination of the distance between the sun and the location on the earth and the angle of incidence of sun's rays.

Their homework assignment was to work in groups of two, and submit an explanation in terms of any one of these positions, the day before the next class.

The best of our science textooks, even those produced by the NCERT, give the impression that doing science is a matter of using one's physical eyes (observing) and doing things with one's physical hands (executing experiments) supplemented by physical instruments, typically in a laboratory. While these are indeed important aspects of physical, biological, and human sciences, there is far more to scientific inquiry than using one's physical eyes and physical hands. Like any other form of rational inquiry, scientific inquiry also calls for the use of the mind: noticing, reflecting, reasoning, evaluating, imagining, and judging. The mental faculties needed for these aspects of scientific inquiry are best found in experiment design (as distinct from the experiment execution in the practicals classes), and the construction, justification and evaluation of theories. It is crucial that our syllabi, textbooks and examinations pay attention to those aspects of scientific inquiry that call for a training of the mind, and not merely that of the body.

A Couple of Additional notes that you don't need to read, but …

In figure 1, we modeled the parameter of height as one-dimensional space (represented as a gray line), and a partucular height of a person as a point (represented as a blue dot). This way of thinking comes from geometry. If you see these concepts in terms of algebra, you will think of the parameter as a variable (that which varies) and a particar height of a person as a value of that variable for the person.

You are going to find different words to refer to the same idea in different domains. Academics tend to use different words to refer to the same concept, and use the same word to refer to two distinct concepts. This is unfortunate, but that is the way the world is, so we need to get use to it.

Here are a few pointers:

The terms variable and value are also used in statistical inquiry. If we use the algebraic symbol h to refer to human height, for instance, the value of h for KPM would be 170. In the language of analytic geometry, what the gray line represents would be called an *axis*. The terms variable and constant, attribute and attribute value, feature and feature value, parameter and value of the parameter, etc. express the same core meanings in different domains. (The term 'trait' is sometimes used to refer to the variable, and sometimes to refer to the value, as in green and blue as traits, or as values of the trait of colour.)

Note also that the term 'scatter plot' is used in statistics typically in the context of a two dimensional data space (see, for instance, the wikipedia entry on scatter plots at http://en.wikipedia.org/wiki/Scatter_plot)), but I am generalizing it to *n*-dimensional data space which includes uni-dimensional data space as well.