Use of Evidence: Explanation of Outcome (c)

(Slide 1) This lecture explains the third learning outcome that relates to the concept of evidence.

(Slide 2) Outcome (c) may be stated as follows:

c) Draw inferences from graphs, tables, and other graphical representations of quantitative evidence.

It is a critical thinking skill, belonging specifically to the branch of critical thinking known as "quantitative reasoning." The term "quantitative literacy" is also sometimes used, but the name "quantitative reasoning" makes more sense for us because it emphasizes that what you are learning is a method of thought, of reasoning, and not merely the acquisition of facts. The term "literacy" is ambiguous in a way that "reasoning" is not: "literacy is used to describe a skill, the ability to read and write the written form of a language, but it is also is used in the sense of "cultural literacy" to refer to how well individuals know facts about history, political institutions, things like that.

(Slide 3) In 1994, the Mathematical Association of America developed a general statement of the quantitative literacy goals that mathematics professors believed that all college graduates should possess. Our outcome (c), the one this lecture focuses on, is a slightly rephrased version of the first of their five goals; it is boldfaced in this slide. The second MAA goal asks that students be able to represent mathematical information in a variety of modes. The remaining goals ask that students be able to perform basic mathematical manipulations, using algebra, geometry, and statistics, and then to check the reasonableness of their results, recognizing the limits of each of the methods they use. Thus the final three goals require the use of particular mathematical techniques, and it is reasonable to expect that students will learn these in mathematics courses or in quantitative courses in other disciplines such as economics, physics, or sociology. The second goal is really a handmaiden to the final three: in order to use algebra, geometry, and statistics, a student must have some ability to represent mathematical information symbolically, visually, numerically, and verbally. This skill is generally taught along with goals three to five.

Goal three requires that students be able to use mathematics, in the sense of performing the manipulations to get a solution to a problem. This is a thinking skill, but the thinking involved is largely problem solving. The other goals expect students to reflect on the manipulations done in goal three, but goals four and five depend on the student's ability to perform specific manipulations. Only goal one allows students to think about the use of math where other people have performed the actual work of calculation, and the student is asked to think about (specifically for us, draw inferences from) the calculations and representations done by others. This is the reason that this particular outcome has been selected for this course.

(Slide 4) We discussed in outcome (1c) how inferences may be drawn from texts or arguments. The present outcome just extends that one beyond the use of written texts to a variety of graphical representations of evidence. To understand the present outcome, once you have a grasp of the general concepts of inference and evidence, all you need are an understanding of the various methods of graphical representation that you are likely to encounter in your studies, and an idea of the kinds of inferences that are typically made from such presentations. This lecture will be structured around these two topics.

Types of Graphical Representation of Evidence

(Slide 5) The commonly used Microsoft program Excel allows users to chart numerical data in eleven basic formats: **Column**, **Line**, Pie, Bar, Area, **Scatter**, Stock, Surface, Doughnut, Bubble, and Radar charts. This is a convenient list, and quantitatively literate readers should be familiar with most if not all of these. In the present lecture we will discuss column, line, and scatter charts. Bar charts are basically the same as column charts, just with the data arranged slightly differently.

(Slide 6) Column charts are used to indicate how the value of an item varies across time periods, or to compare values for different items in the same period. The two examples given here compare sales figures – for some reason sales is commonly used as an example of how to draw charts, and we have yielded in this case to that tradition. In any case, the first chart of the current slide shows how the value of sales (in dollars, in the mid-south region) varies across time. In this particular chart we are not told what is being sold, though presumably this would be known by context; for example, it could be total sales by company, or sales of a particular good.

The emphasis, instead, is on variation across time: as in most businesses, the year is divided into four quarters, and values for two years are shown. Consequently, eight values are given (two years times four quarters equals eight quarters of data). This allows comparison in two ways: we can see how sales increase from year to year in any quarter (by comparing red to blue columns in the same quarter), and we can see how sales vary quarter-by-quarter in any given year (by comparing columns of the same color in different quarters).

(Slide 7) To read column charts, find the top of a particular chart and look across to the left axis to read the corresponding value. Thus, in the first quarter sales in the midsouth region, according to the first column, were 30,000 in 2004 and just under 50,000 in 2005. Comparison with the tables that correspond to each chart indicates the advantages and disadvantages of presenting quantitative data graphically. What is the value of sales in quarter one, 2005? Well, it's hard to know exactly from the chart: we know it's a little under 50,000, but we can't tell exactly. From the table to know it's actually 48,000. This illustrates the fact that it's hard or impossible in some cases to read exact values off a chart: that's a disadvantage. An advantage is that's easier in a glance to see approximate comparisons. In the second chart, we see immediately that sales in the northwest are significantly lower than in the three southern and eastern regions, and that they are very approximately half the value of sales in the southeast. The same information can be gotten from the table below, but to make comparisons between regions it's necessary to do some quick math: 8,000 is just under half of 20,000, and 15,000 is a little over half of 22,000.

(Slide 8) Charts take advantage of the fact that humans are designed to judge relative sizes and distances in space, and represent numerical information in ways that leverage this ability. Some people prefer charts, other people tables, and sometimes information will be presented in both formats. But charts can present information in less space, and if one wishes to present general trends in data, charts are often to be preferred. On the other hand, if one wants to present raw data which the reader can use in calculations, then it's generally necessary to use a table format.

(Slide 9) This slide illustrates an exception to one of the generalizations made on the previous slide. If a chart just shows a small number of data points, it may be convenient to indicate the exact quantities of each data point. If there are too many data points, then doing this makes the presentation too complex for the reader to digest, but there are times that this is the best way of representing data. If the values of each data point are indicated, then the reader can both compare relative quantities visually, rather than by having to do mathematical calculations as he would in a chart, but he can also get the exact values should he need them.

(slide 10) Line and column charts can often be used interchangeably. In such cases, line charts simply indicate the values shown at the top of each column. This is illustrated in the present slide in two ways. Above, the same data are represented in both column and line formats. You can see from the line chart, for example, as from the column chart, that values for the first quarter are 30,000 and just under 50,000 in the two years represented. The fact that the dots on the line chart represent values corresponding to the tops of the columns in a column chart is illustrated in the chart at the bottom of the slide, where a column is overlaid over a portion of the line chart.

(Slide 11) One advantage to line charts over column charts is that they can represent continuous as well as discrete data. This slide illustrates line charts representing continuous and discrete values: the former gives the position of a puck at different times during a three second interval in a hockey match, whereas the latter is the same line chart we saw in the last slide. The movement of an object over time is continuous in the sense that at any time it has a distinct position, no matter how finely we break time down into intervals. It's not like it disappears at one moment, only to appear at a later period in a different location. It's always somewhere. Sales, in contrast, are discrete quantities: at some moment a sale is being made, at other times not. It's not like the sale of a loaf of bread is always somewhere. Sales are events: sometimes they occur, other times not. Hockey pucks, in contrast, are objects, and not the kind of objects that tend to come into being and then disappear, only to reappear elsewhere.

(Slide 12) Now, to draw the line chart for the hockey puck, we took locations at quarter-second intervals, and we picked numbers to make the line look fairly smooth. Instead, we could have smoothed the line out so that the line between data points was not straight, but rather curved. But drawing it as we did preserves for readers of the chart the knowledge that measurements of position were taken at intervals, and consequently we don't know exactly where the puck was located between measurements. This is illustrated in the first version of the chart, where the actual data points are indicated.

By the way, the measurement of position in units is relative to a starting point. Imagine that we have ruled on the ice something similar to the divisions given on a yard stick, with the puck beginning, at the beginning of measurement (or time equals zero) at the zero point of the yard stick. Then we are in a position to measure relative displacement from its starting position in whatever units are given by our yardstick, for example feet.

(Slide 13) Scatter plots are similar to line charts that show individual data points, with the difference that they do not show a line connecting the points, but rather just plot the data points themselves. Compare the line chart illustrating the movement of a hockey puck over time, which you've already seen so it's reproduced here in a smaller size at the bottom of the screen, to the larger scatter plot shown above it. This scatter plot show the average GRE scores of students in different intended majors, charting performance on the writing section of the test against performance on the quantitative section. Since the hockey puck is a single object following a given trajectory over time, it has only one position at any given time, so no point is directly above another point – if two points were stacked one over another, this would indicate that a single puck was in two different places simultaneously, a physical impossibility. In contrast, there's nothing to keep students in two majors from having the same average quantitative score on the GRE, as for example Psychology and Higher Education intended majors seem to do.

(Slide 14) In fact, the difficulty in drawing a line connecting the points on the scatter plot isn't the key to the difference between scatter and line charts. In a scatter plot, the values of each data point are independent of each other. In the case of the GRE exam, people deciding to major in physics are, on the whole, different from the people intending to major in English (there might be the occasional student who applies to graduate school in both fields, but these are so few in number as to be irrelevant to the overall averages), so the value of the one point has nothing to do with the value of the other. In contrast, the position of a hockey puck at a given time (let's call it time equals t+1) is <u>not</u> independent of its position at an earlier time (say at time equals t). Where the puck is at t constrains where it can be at t+1.

So while column and line charts are largely interchangeable (you can redraw one as the other), this really isn't the case with scatter plots. Certainly in practice it would be possible to plot the data points for a line chart and not draw the connecting line, in which case what you get would look like a scatter plot. But as a practical matter, scatter plots are drawn for different reasons, and different kinds of evidence, than line charts. It is also possible to draw lines connecting the data points in a scatter plot, or, as in the case of this example, to draw a line that indicates the general correlation between variables. In this case, the line indicates that, across majors and in general terms, students who score higher on the quantitative section tend to score lower on the writing section, and vice versa, with the slope of the line indicating the general (inverse) correlation between these two variables.

(Slide 15) Reading charts is straightforward. You can focus on particular data points, or on the general trend in the data. For example, the chart of GRE scores is likely going to be used to compare the performance of students in different fields. Philosophy departments frequently brag about the GRE scores of their majors, on the assumption that something in the philosophy curriculum is likely responsible for this level of performance. It is perhaps notable that Philosophy majors score significantly higher on the writing portion of the exam than English majors, and also that they score higher on the quantitative section than any liberal arts major other than Economics. Similarly, it may be interesting that Physics majors score higher on the writing section than engineers, though this may be less surprising since physics is an arts and science field, whereas engineering is a professional major, so physics majors may tend to take more courses in fields like history, English, and philosophy that require substantial writing. In any case, these are some of the inferences that one may reasonably draw from the scatter plot.

Types of Inference from Graphical Representations

(Slide 16) The previous slide illustrated one strategy for reading the GRE chart: we assumed that science and math majors would do better on the quantitative reasoning section, and liberal arts majors would do better on the verbal reasoning section, because coursework in the sciences is heavily math-based and in liberal arts courses students tend to write more – this is all outside knowledge that was brought to bear on reading the chart, because it seemed relevant to particular sorts of questions one might use such a chart for. And it concluded by making comparisons between disciplines, focusing on those near the upper end of the two test sections, and on disciplines like English that could be anticipated as preparing students well for a particular test.

Note that the conclusions drawn from the scatter plot are partially determined by the way the data are presented, as well as by the questions brought to bear by the reader. Thus, the scatter plot helps one quickly see relative values of different data points (for example, physics majors score higher than economics majors on the quantitative section, but about the same on the verbal, as economics majors), but one can draw inferences more easily from the table given on this slide about the absolute scores, because if numbers were added to the scatter plot it would be much too hard to read. Thus, one notices that while philosophy and physics majors score well on both sections of the test compared to other liberal arts or science majors, in absolute terms philosophy majors score more nearly the same on both sections, while physics majors score much more highly on the quantitative section. One also notes that of two social science disciplines that emphasize quantitative skills in some courses, anthropology and history, anthropology students score better on the quantitative section.

(Slide 17) Further conclusions may also be draw from these comparisons, but would again require more outside information. For example, is scoring a 700 on the verbal section as good, somehow, as scoring a 700 on the quantitative section? This is an easy question to ask when one sees that the top science majors score much higher on the quantitative section than liberal arts majors do on the verbal section. But it is hard to answer. How does one compare equal numbers across different tests? ETS, the producer of the GRE, may provide some guidance based on the statistical measures they use to develop the test, but again, the reader will need to also bring further questions and assumptions to bear in interpreting what they say.

(Slide 18) We thus see from this one example that we can read quantitative information in several ways from charts and tables. We can read particular values for particular data points, or we can compare different values, or we can describe the general trend in the data, for example in the line chart we can describe the shape of the line, and how if varies along different segments. While reading particular values doesn't require any interpretation – so "reading" is really the operative verb here – comparing values and describing the shape of the line are slightly more complex skills. Still, little outside knowledge needs to be brought to bear other than basic mathematical knowledge if we wish to make a claim like "the puck moves more than one unit between t=0.5 and t=1.5." But strictly speaking this is probably an inference from the data. You might argue that we can "read" (quote unquote) the change in location directly off the chart by identifying the value on the y-axis corresponding to t=0.5 and t=1.5, and naming the distance between the points. Of course, to get the difference in position requires a mathematical calculation, a subtraction, and so this is a different sort of "reading" than simply getting one value from the chart. But in this case the difference is fairly subtle, and we probably shouldn't worry about whether an inference is required.

(Slide 19) We use the term "reading" somewhat metaphorically, in any case, when describing what we do with charts. Normally we speak of reading written texts, and these conform to conventions that guide us in how we read them. For example, English texts are read left to right, top to bottom, and the meaning of any given sentence evolves in our mind as we read the sentence. Charts and tables don't work the same way. There's no preferred way of reading them: they are simply ways of organizing information. Put differently, the sentences in texts make claims, and reading texts means, initially at least, understanding these claims. No claims are made by charts, in contrast. To get claims out of charts we need to perform inferences ourselves.

(Slide 20) In any case, regardless of where we draw the line between reading and inferring from charts, we can make several general claims about how inferences from charts work. The simplest sort of inference makes claims about relations between data points; we may wish to categorize this as simply reading the chart, or as drawing of simple inferences, but in any case it is a skill that requires little outside knowledge other than the ability to understand the subject of the chart (what the GRE test is, for example) and certain mathematical knowledge (such as the ability to perform subtraction).

More complex inferences require additional information not found in the chart, such as knowledge about how different academic disciplines work, or how tests are constructed. And strictly speaking, all inferences require us to ask questions of the chart, like how far does the puck move in a given second, or what do we make of the fact that philosophy majors score higher than English majors. We may think of some simple questions as in some sense being embedded in the way the chart is designed, and hence attribute them more to the chart-maker than to the chart-reader, but regardless, every inference answers some question brought to the data.

In short, we can think of inferences from charts as requiring three elements: information presented in the chart, outside information, and questions that the inference is designed to answer.