Figure 1.4. DATA-BASED INVESTIGATING: Racial Differences in IQ

Two matters of statistical interest are raised by the article EM9419 reprinted in this Figure 1.4.

∗ The use of the normal (Gaussian) distribution as a model for the shape of histograms of large data sets of IQ scores – of particular note are the last paragraph of the right-hand column below and the first paragraph overleaf on page 1.10.

∗ The comments, highlighted below in the middle of this page, about the importance of *correct* scientific (or statistical) method in generating, for the Question of interest, an Answer that is likely to be (close to) correct.

– Also noteworthy is that a *wrong* method of investigating (*i.e.*, a deficient Plan) *may* generate a (close to) correct Answer.

EM9419: The Toronto Star, November 13, 1994, page E8

Why IQ testing is junk science

And why it just leads to more junk science – statistical analyses 'proving' whites are smarter than blacks

BY A.K. DEWDNEY
SPECIAL TO THE STAR

WHEN IT COMES to bad science, there's one thing that past and present studies linking race and intelligence have in common: IQ plays a crucial role in all of them.

In 1989, for example, J. Phillipe Rushton, a psychologist at the University of Western Ontario in London, Ont., sparked a media storm when he claimed that Asians, whites, and blacks could be ranked in that order when it came to decreasing cranial capacity, decreasing intelligence and less orderly behaviour.

Perhaps Rushton thought that by enthroning Asians at the high end of his scale he might avoid the controversy raised in 1968 by another psychologist, Arthur Jensen of the University of California at Berkeley. Jensen published a paper claiming that whites were more intelligent than blacks and, moreover, that the trait was inherited. Jensen concluded that special training programs for blacks were useless.

Rushton, who a few months ago published a book called *Race, Evolution, and Behavior* (Transaction Publishers, New Brunswick, N.J.), has recently been joined by two like-minded colleagues, Charles Murray, a fellow of the American Enterprise Institute, and Richard J. Herrnstein, a psychologist late of Harvard. The Murray and Herrnstein book called *The Bell Curve* (The Free Press/Simon & Shuster), is cast in the Jensen-Rushton mold: whites are smarter than blacks and the difference is inherited.

As in previous controversies, media articles and stories responding to these recent books continue to miss the point. The real problem appears to be IQ, the so called "intelligence quotient."

For most of this century the concept of IQ has gripped the popular imagination to the point where most people think it measures something called "intelligence." Without a large sample of IQ test scores, not to mention a pseudo-scientific literature on the subject, none of these authors would have been able to make a case for his racial contentions. In a nutshell, junk science (IQ) just leads to more junk science (race and IQ).

By "junk science" we do not necessarily mean science that produces wrong conclusions but science that does not follow the scientific process. Of course, wrong methods often produce wrong results. But no one will know what the right results are, scientifically speaking, until someone does the right research. In the meantime, we are left with a practice – IQ testing – that has no accompanying theory to speak of and many indications that the results, as such, are wrong. Here is what happened on the road to our present mess.

By 'junk science' we do not necessarily mean science that produces wrong conclusions but science that does not follow the scientific process. Of course, wrong methods often produce wrong results. But no one will know what the right results are, scientifically speaking, until someone does the right research.

The notion that one could test a person for his or her "intelligence" began innocently enough in the late 19th century when the French psychologist, Alfred Binet, developed a test of scholastic potential for school children. Binet simply made up a large number of questions that took the form of an examination but which reflected things that normal adults might be expected to know.

By 1917, the American IQ testing pioneer Lewis H. Terman, a professor of psychology at Stanford University, had developed the now famous Stanford Binet test. In devising his test, Terman followed Binet: he simply made up the questions. There was no underlying theory, no definition of intelligence, not a shred of genuine scientific research to guide him in formulating them. No one really knew what (if anything) "intelligence" was. IQ tests then, as now, were simply made up of whole cloth. There was no underlying theory..... not a shred of genuine scientific research.

When a scientific method goes wrong from the very start, it may become extremely difficult to repair later. Too many people make their living this way. Too many other people believe the results. The small proportion of the psychological community that still supports the IQ concept has erected a rampart of pseudoscientific theory to bolster its position.

Owing perhaps to the complete absence of theoretical foundations, the IQ field took on a curious circularity. On being asked what "intelligence" was, for example, the Harvard psychologist Edmund Boring remarked, "Intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test." It gets worse.

IQ testers and theorists defend their ground by claiming, for starters, that IQ tests must measure something natural because the distribution of IQ scores follows the famed normal distribution. This is the famous bell-shaped curve that many natural measurements seem to follow. The central hump of the curve reflects a large number of individuals clustered around the average, the tails of the curve reflect the smaller number of individuals that measure much larger or smaller than average.

(*continued overleaf*)

But those who frame IQ tests anticipate the bell-shaped curve by the painstaking elimination of questions that lead to non-normal distributions. IQ tests follow the normal distribution because they're made that way. It's obviously much easier to make a quack than a duck.

Another circularity enters when testers "validate" their tests by comparing score distributions of different tests or by comparing test scores with the academic achievement of those tested. Who would be surprised to discover that the correlations are relatively high?

Over most of this century, a plethora of IQ tests have developed and the notion of IQ as an innate quality, not to say a heritable one, has permeated our culture. Someone who does well in an ordinary test has earned a high grade but someone who does well in an IQ test has got a high IQ.

Very few people realize, however, that IQ scores are very plastic. They have been raised by 20 or 30 points for black inner city children, for example, by training them in the ways of the white-oriented educational system. Throughout their checkered history, IQ tests have been used not only by educators anxious to measure academic potential but by those who would establish critical differences between races. Thorough critiques by the renowned geneticist R.C. Lewontin, the biologist Stephen Jay Gould, and many other scientists have only drawn the IQ wagons into a tighter circle.

The favorite defence of Jensen, Rushton and the others is a rather arcane statistical method known as factor analysis They claim that a factor called g, short for general intelligence, emerges from the analysis when applied to large numbers of IQ scores. But to point out that emergent factors depend very heavily on some arbitrary decisions made at the start of the analysis seems to do little good. The racial analysts simply forge ahead with grinding statistical analyses that impress psychologists far more than they impress scientists.

Although IQ is a scientifically meaningless measure, those who still worry about the implications of differing average scores for races may take some comfort in the bell-shaped curve itself. When scientists talk about within-group differences versus between-group differences, they refer to two overlapping bell-shaped curves. Consider two such curves that reflect the distribution of test scores for a large group of blacks and a large group of whites. According to Rushton, the black bell centres on 85 while the white bell centres on 100.

Forgetting for the moment the well-known cultural biases of IQ tests, note that the two bell curves overlap almost completely. Although the white curve is higher on the right and the black curve is higher on the left, you can construct a kind of "IQ buddy" system. The vast majority of whites, for example, could each be assigned a black "buddy" who has exactly the same measured IQ. At the low end of the scale a white may have two or more black IQ buddies. At the high end of the scale a black may have two or more white buddies.

This simple observation carries an important message. The work of Jensen, Rushton, Murray, Herrnstein and others who persist in the largely pointless study of racial statistics cannot apply to individuals. For this very reason, it would be impossible to formulate any public policy based on their ideas. Who, after all, would want to see their IQ buddy get a raw deal?

We may continue to give people tests for scholastic aptitude but let's not call them "intelligence" tests and let us not pretend that science has a workable theory of intelligence.

A. K. Dewdney is a mathematician and associate professor of computer science at the University of Western Ontario. His book on pathological science will be published next year.

The article EM9419 reprinted overleaf on page 1.9 and above is also used in Figure 5.17 of the STAT 220 Course Materials and in Statistical Highlight #101.