Scientific American Apr 1995 pp. 32-33

THE ANALYTICAL ECONOMIST

The Insignificance of Statistical Significance

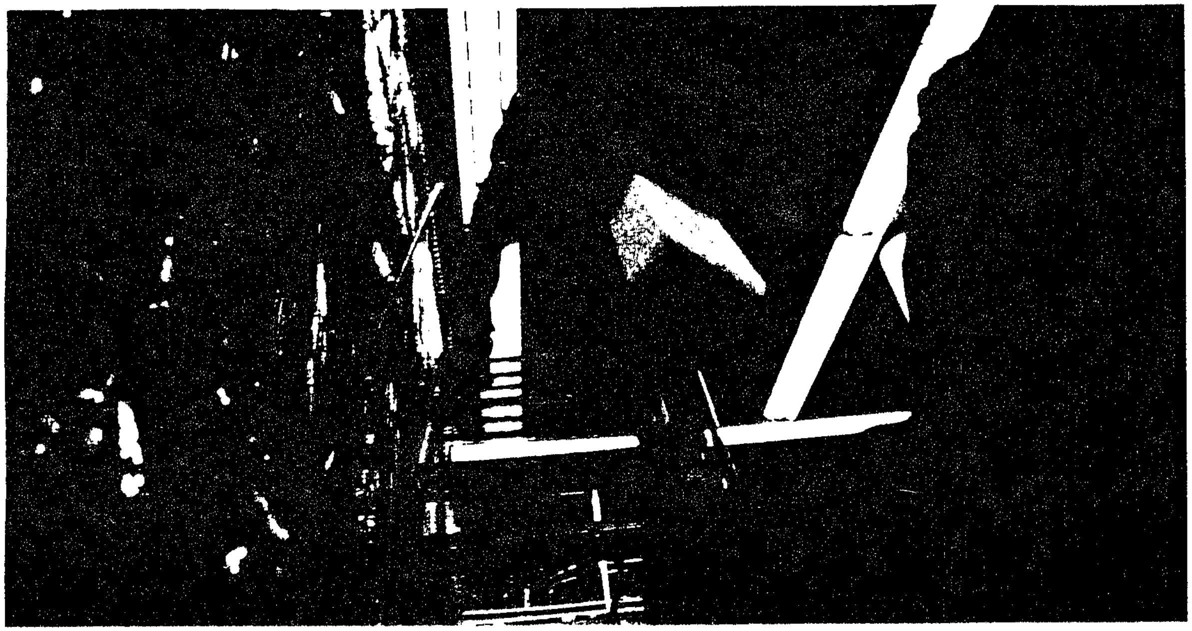
Economists, astrophysicists, sociologists, geologists as well as some medical researchers spend a lot of time looking at experiments that God has already performed. If God had not arranged things so that some stars were young and some were old, the astrophysicists would not know much about stellar evolution. Likewise, if God had not arranged things so that the minimum wage varied relative to the average wage for unskilled labor from decade to decade and state to state, economists would have a hard time convincing anyone that the minimum wage puts poor people out of work.

Economists and astrophysicists come to their knowledge by finding regularities of some kind in the world; one crucial part of their task is figuring out whether particular correlations point to an important law or to the fickle hand of coincidence. As a matter of fact, economists are having a hard time convincing people that the minimum wage contributes to unemployment because recent studies show no "statistically significant" effect on jobs. When Congress takes the issue up later this year, the livelihoods of thousands of people could hang in the balance.

But just what does that phrase mean, and what does it have to do with the debate? Go back two centuries, to Pierre Simon, Marquis de Laplace, the first person to apply the notion of statistical significance to a serious scientific problem. In 1773 Laplace wanted to know where comets came from. He reasoned that if they originated inside the solar system, they would orbit in the same plane as the planets, whereas if they came from the far reaches of space, their paths would have no correlation with those of bodies circling the sun. Laplace checked the motions of the last 12 comets to be discovered and firmly rejected the hypothesis that comets came from inside the solar system. If the comets were of local origin, one might by chance travel at some weird angle to the plane. But the odds of getting two anomalies would be lower, of three lower yet—the probability, so to speak, of rolling snake eyes three times in a row. This was a very smart idea.

In the succeeding two centuries, statisticians have refined Laplace's simple notion into "statistical significance" and developed an arsenal of formulas for determining whether the phenomena that researchers observe are caused by sampling error (accidentally picking unrepresentative subjects) or "real" effects. The gold standard for most studies is the "95 percent confidence level," which indicates odds of only one in 20 that a result arises from chance. Economists use it to test whether the minimum wage has a "significant" effect on employment. Medical researchers use it to decide whether half an aspirin a day keeps the cardiologist away.

Gradually, however, it has dawned on a few scientists that something is screwy. An obvious problem is that with so many people doing so many studies, some of them are going to run into that one-in-20 chance of believing in a mirage. The converse mistake is more subtle: scientists care about whether a result is statistically significant, but they should care much more about whether



MARK PETERSON SABA

LOST JOBS? Some economists say low-paid workers will be fired if the minimum wage rises; others claim the evidence is statistically insignificant.

it is meaningful—whether it has, to use a technical term, oomph.

Sadly, many scientists have started thinking that statistical significance measures oomph. If an answer meets the 95 percent confidence criteria, it must be important; if it doesn't, it isn't.

The clearest refutation of this notion came in the study that established the lifesaving effect of aspirin in men who had already had a heart attack. Researchers stopped the experiment before their numbers reached "real" statistical significance because the effect of a mere half an aspirin a day was so obvious that they considered it unethical to go on giving placebos to anyone.

Is this messy state of affairs Laplace's fault? He was right about comets because the relevant scale for measuring the oomph in orbits was obvious. Furthermore, a sample of a dozen could yield results that were scientifically as well as statistically significant. But the scale for measuring the effects of aspirin or of changes in the minimum wage is not so clear: you may get statistically impeccable answers that make little difference to anyone or "insignificant" ones that are absolutely crucial.

That conundrum is sharpest now in the debate among economists about the minimum wage. David Card and Alan B. Krueger of Princeton University have used tests of statistical significance to argue there is no convincing evidence that the minimum wage has a strong effect. Most other economists disagree, both because their theory tells them otherwise and because they think Card and Krueger are asking for too much certainty. But because both sides are muddled about the difference between oomph and statistical significance, the disagreement is not likely to get resolved in time to help Congress. Depending on what legislators decide, many poor people (not to mention teenagers on summer vacation) might lose their jobs. Ironically, even if they do, economic samplers may not be able to prove how many jobs were lost or that the minimum wage really had an effect.

DONALD N. McCLOSKEY is professor of economics and history at the University of Iowa.