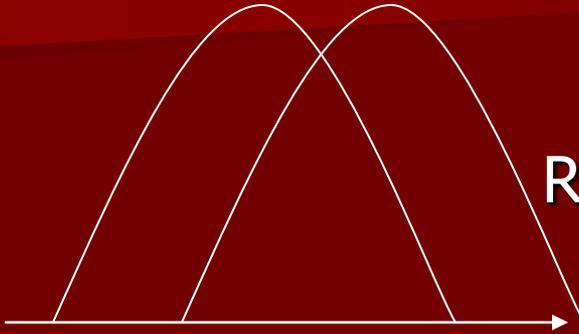
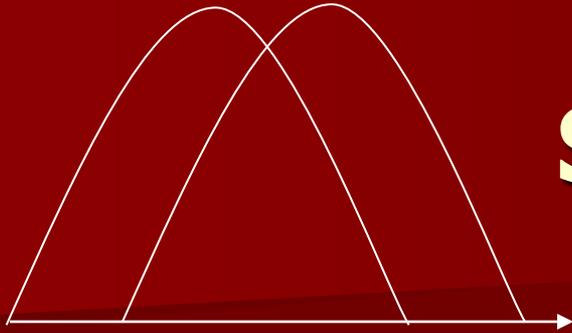


Size Matters:



The Standard Error of
Regressions in the American
Economic Review

By Stephen T. Ziliak and Deirdre N. McCloskey



Size Matters

**Presented by Stephen T. Ziliak
at a plenary session of the 2004 American
Economic Association meetings, San Diego.**

Chair: Kenneth J. Arrow

Respondent: Deirdre N. McCloskey

**Organizer: Morris Altman, Association for Social
Economics and Editor, *Journal of Socio-Economics*
(Special issue on statistical significance, v. 33,
2004) .**

Discussants & Commentators:

**Morris Altman, Graham Elliott, Sir Clive Granger,
Fiona Fidler, Gerd Gigerenzer, Joel Horowitz,
Edward Leamer, Peter Lunt, Nathan Berg, Geoff
Cumming, Bruce Thompson, Neil Thomasan, Mark
Burgman, Anthony O'Brien, Erik Thorbecke,
Jeffrey Wooldridge, and Arnold Zellner.**

“The Standard Error of Regressions” (1996) gave the intellectual history of significance testing:

Since Edgeworth (1885), statistical significance is not the same as policy or scientific significance—what we call economic significance. Importance.

A *statistically* significant difference may not be economically important.

And a statistically insignificant difference *may be* economically important.

**Significance testing is old.
Laplace stamped his name on the method
(though not the term) in 1773**



**In 1885 F.Y. Edgeworth coined the very term
“statistical significance,”
and circulated a definition:**

It's about sampling error only!

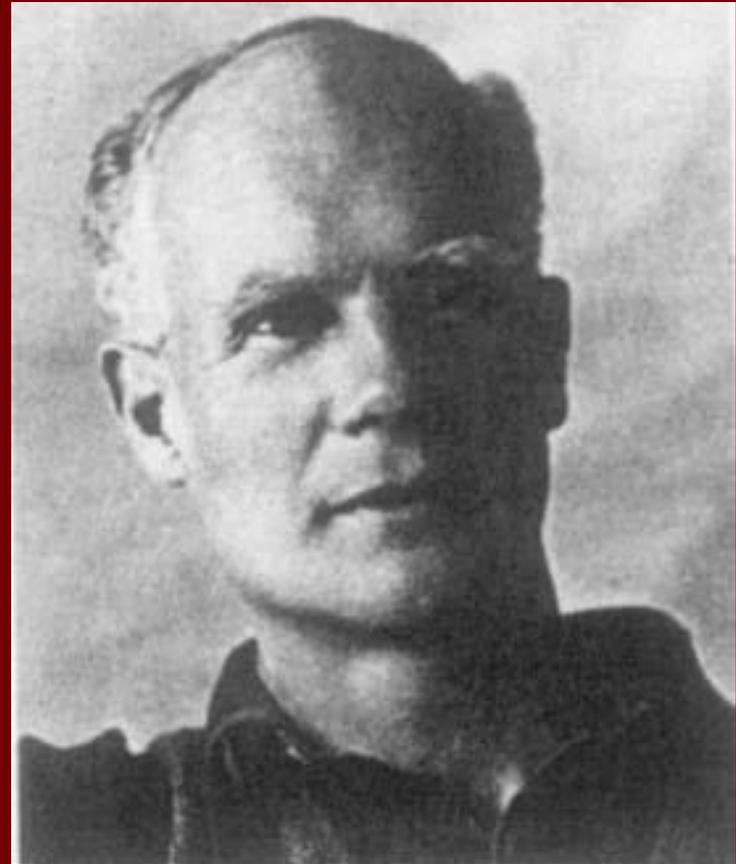
What's the chance I'll see
this difference again, under similar
experimental conditions?

“But for the purpose of science,”
Edgeworth said to Jevons,

“the discovery of a difference in
condition, a difference of 3 per cent and
much less may well be important.”



Neyman & Pearson emphasized economic significance in their famous paper on Type I and Type II error



Listen:

“Is it more serious to convict an innocent man or to acquit a guilty?”

That will depend on the consequences of the error;

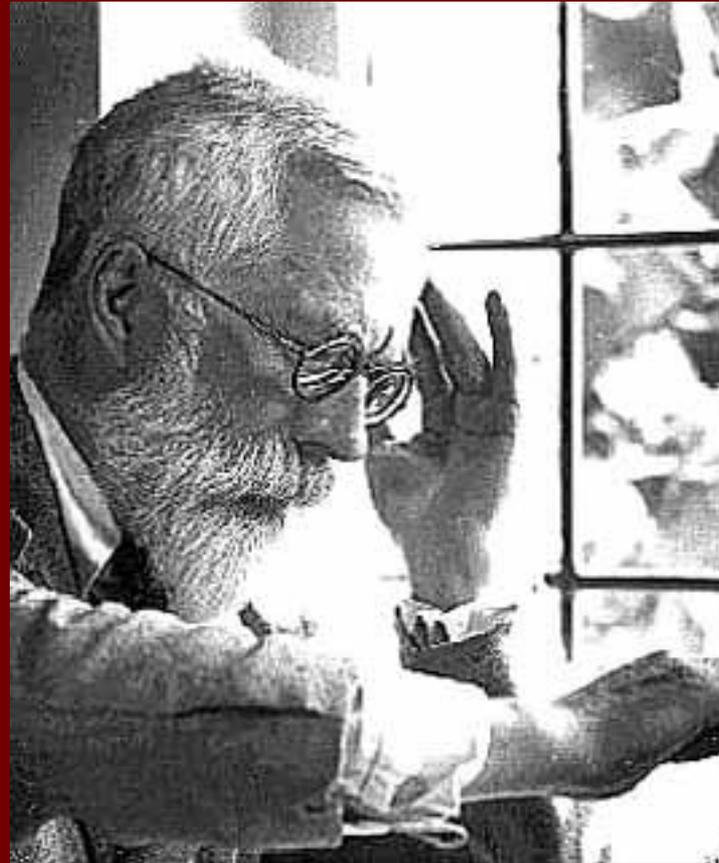
Is the punishment death or fine; what is the danger to the community of released criminals; what are the current ethical views on punishment?

From the point of view of mathematical theory all that we can do is show how the risk of errors may be controlled and minimised.

The use of these statistical tools in any given case, in determining just how the balance should be struck, must be left to the investigator”

(Neyman and Pearson 1933, p. 296).

This is R. A. Fisher.
He created a lot of headaches when he
insisted on the 5% level of significance.



Yet 70% of the authors in the *AER* did not understand the difference between economic and statistical significance

Of the 182 full-length papers published in the 1980s in the *AER*, 70% made a costly mistake:

Example: Illinois Employment Experiment

Example: Social Welfare in Boston

What was Actual Practice in the 1990s?

We applied the 19-item Questionnaire to all the full-length papers published in the *AER* in the 1990s and using a test of statistical significance.

The full selection,
January 1980 to December 1999,
amounts to 369 papers*

It's the population.

Selections from the 19-Item Questionnaire

Does the paper . . .

2. Report units and descriptive statistics for each variable so that the reader can determine “how large is large”?
7. At its first use, consider statistical significance to be one among other criteria of importance?
13. Discuss the scientific conversation within which a coefficient would be considered large or small?
16. Consider more than statistical significance decisive in an empirical argument?

“Yes” is Good!

TABLE 2

The Economic Significance of the *American Economic Review*
Has in Some Regards Improved
(Measured by Net Percentage Difference, 1980s to 1990s)

Survey Questions	Percent Yes in 1990s	Improvement since 1980s
Does the Paper . . .		
2. Report units and descriptive statistics for each variable so that the reader can determine "how large is large"?	66.3	+34.0
13. Discuss the scientific conversation within which a coefficient would be considered large or small?	53.5	+25.5
17. Do a "simulation" to determine whether the coefficients are reasonable?	36.2	+19.4

TABLE 3**. . . But the Essential Confusion of Statistical and Economic Significance is Getting Worse**

Survey Questions	Percent Yes in 1990s	Decline since 1980s
Does the Paper . . .		
10. Eschew asterisk econometrics , the ranking of coefficients according to the absolute value of the test statistic?	31.0	-43.7
11. Eschew sign econometrics , remarking on the sign but not the size of the coefficient?	21.9	-24.8
1. Use a small number of observations, such that . . .	71.1	-14.6

. . . But the Essential Confusion of Statistical and Economic Significance is Getting Worse

Does the Paper . . .	Percent Yes in 1990s	Decline since 1980s
15. Use other criteria of importance besides statistical significance after the crescendo?	27.8	-12.9
16. Consider more than statistical significance decisive in an empirical argument?	20.9	-8.8
7. At its first use, consider statistical significance to be one among other criteria of importance?	39.6	-7.7

What Should We Do?

Robert Solow says “economists should try very hard to be scientific with a small s .”

Yet the authors of the *AER* are trying very hard to be scientific with a small t

(or a large t , if that’s the way the null is set up).

Our Policy Conclusion Is This:

- That the profession adopt the standards of economic significance

Of Gosset aka “Student,” and of Edgeworth, Neyman, Egon Pearson, Wald, Jeffreys, DeGroot, Keynes, Savage, Kruskal, Zellner, Leamer, Goldberger, Granger, Arrow, Solow, Friedman, Tullock, Griliches, Schelling, and others.

- The standards are summarized in our 19-item questionnaire.
It’s one model for a referee report.

Referees should say, “No size? no significance!”

- We should ask, Does the paper strive mainly to measure and interpret economic significance?

If not, *reject*—reject the paper, that is.

- At minimum we should ask, “Is a D+ in hypothesis testing good enough for the official journal of the AEA?”

Steve Ziliak, Deirdre McCloskey, and Janie



“Size Matters” was published
in a special issue of the
Journal of Socio-Economics

TABLE 2
The Economic Significance of the *American Economic Review*
Has in Some Regards Improved
(Measured by Net Percentage Difference, 1980-1999)

Survey Question	Percent Yes in 1990s	Net Improvement since 1980s
Does the paper . . .		
5. Pay attention to the details of the units of measurement, and to the limitations of the data?	81.0	+36.5
2. Report descriptive statistics for regression variables?	66.4	+34.0
9. Examine the power function? ^a	45.5	+28.8
13. Discuss the scientific conversation within which a coefficient would be judged large or small?	54.0	+26.0
18. In the conclusions, distinguish between statistical and economic significance?	52.6	+22.5
17. Do a simulation to determine whether the coefficients are reasonable?	35.0	+21.8

TABLE 3

. . . But the Essential Confusion of Statistical and Economic Significance
is Getting Worse
(Measured by Net Percentage Difference, 1980-1999)

Survey Question	Percent Yes in 1990s	Net Decline since 1980s
Does the paper . . .		
14. Avoid choosing variables for inclusion solely on the basis of statistical significance?	25.5	-42.6
10. Eschew "asterisk econometrics," the ranking of coefficients according to the absolute value of the test statistic?	32.8	-41.9
11. Eschew "sign econometrics," remarking on the sign but not the size of the coefficient?	19.0	-27.7
1. Use a small number of observations, such that statistically significant differences are not found merely by choosing a very large sample?	67.9	-17.8
4. Test the null hypotheses that the authors said were the ones of interest?	83.9	-13.4
15. Use other criteria of importance besides statistical significance after the crescendo?	28.5	-12.2
16. Consider more than statistical significance decisive in an empirical argument?	18.2	-11.5
7. At its first use, consider statistical significance to be one among other criteria of importance?	36.5	-10.8