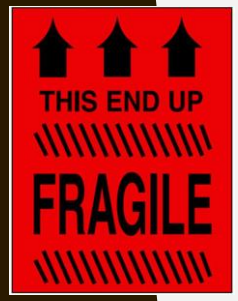
MAXIMIZE G-VALUES

BY STEPHEN T. ZILIAK

TRY GOSSET'S GUINNESSOMETRICS

G-10 *CONSIDER THE PURPOSE & COMPARE WITH BEST PRACTICE*



- Falsification of a null hypothesis is <u>not</u> the main purpose of the experiment or observational study
- Making money or beer or medicine - ideally more and better than the competition and best practice – is
- Estimating the importance of your coefficient relative to results reported by others, is
- As the **2016 ASA Statement on Statistical Significance** makes clear, merely falsifying a null hypothesis with a qualitative yes/no, exists/does not exist, significant/not significant answer, is not itself significant science, and should be eschewed

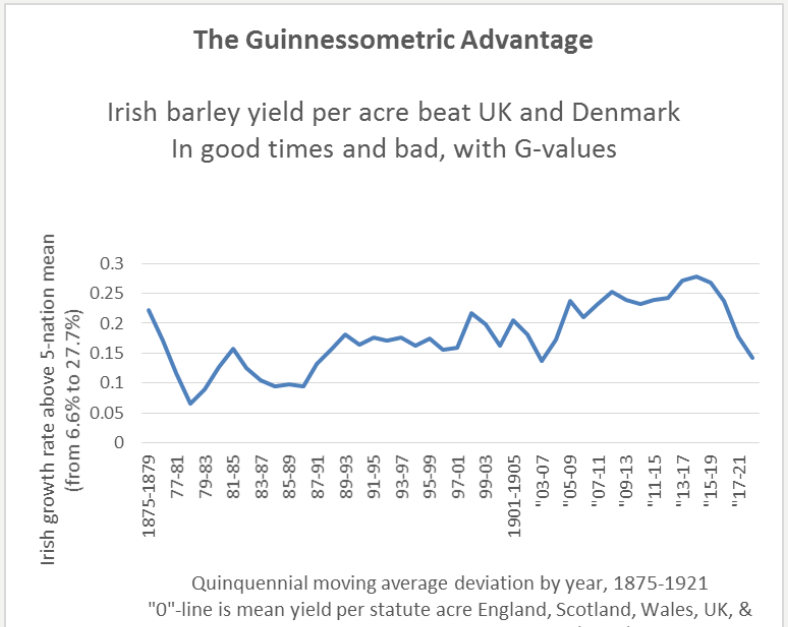
Gosset (1938, Biometrika): "**I personally choose the method which is most likely to be profitable** when designing the experiment rather than use Prof. Fisher's system of a posteriori choice which has always seemed to me to savour rather too much of 'heads I win, tails you lose'"

G-9 *ESTIMATE THE STAKES (OR EAT THEM)*

- Estimation of magnitudes of effects, and demonstrations of their substantive meaning, should be the center of most inquiries. Failure to specify the stakes of a hypothesis is the first step toward eating them (gulp)

<u>This</u>:

Or <u>This</u>:



G-8 *STUDY CORRELATED DATA: "ABBA," TAKE A CHANCE ON ME*

- Most regression models assume "i.i.d." error terms – independently and identically distributed. Yet most data in the social and life sciences are correlated by systematic, non-random effects—not independent

- Gosset solved the problem of correlated soil plots with the "ABBA" layout, maximizing the correlation of paired differences between the As and Bs with a perfectly balanced chiasmic arrangement

G-7 *MINIMIZE "REAL ERROR" W/ 3 "R"S: REPRESENT, REPLICATE, REPRODUCE*

- A test of significance on a single set of data is nearly valueless
- Fisher's *p*, Student's *t*, and other tests should only be used when there is actual repetition of the experiment. "One and done" is scientism, not scientific
- Random error is not equal to real error, and is usually smaller and less important than the sum of non-random errors. Measurement error, confounding, specification error, and bias of the auspices, are frequently larger in all the testing sciences, agronomy to medicine
- Guinnessometrics minimizes real error by repeating trials on stratified and balanced yet independent experimental units, controlling as much as possible for local fixed effects.

G-6 *ECONOMIZE WITH "LESS IS MORE": SMALL SAMPLES OF INDEPENDENT EXPERIMENTS*

- Small sample analysis and distribution theory has an economic origin and foundation: changing inputs to the beer on the large scale (for Guinness, enormous global scale) is risky, with more than money at stake

- Smaller samples, as Gosset showed in decades of barley and hops experimentation, does not mean "less than", and Big Data is in any case not the solution for many problems

- Gosset's first "z"-table began at $n=2$ (Student 1908, Biometrika; Ziliak 2008)

G-5 *KEEP YOUR EYES ON THE SIZE MATTERS/HOW MUCH? QUESTION*

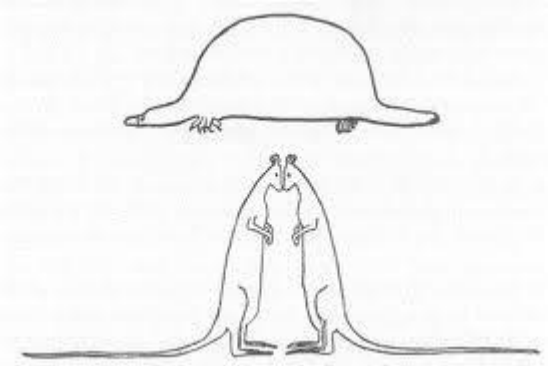
- There will be distractions but the expected loss and/or profit functions rule, or should. Are regression coefficients or differences between means large or small? Compared to what? How do you know?

- For example, said Gosset (1904, Guinness Laboratory Reports):

  "It might be maintained that malt extract should be [estimated] within .5 of the true result with a probability of 10 to 1"

- See Ziliak's and McCloskey's (2008) numerous examples of Oomph vs Precision

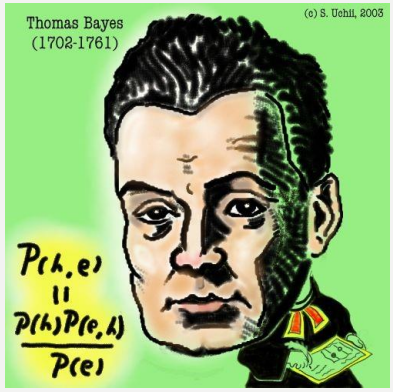
G-4 *VISUALIZE*



- Parameter uncertainty is not the same thing as model uncertainty
- Remember Stephen Jay Gould's "The Median isn't the Message" (1985)
- Does the result hit you between the eyes? Does the study show magnitudes of effects across the entire distribution? Advances in visualization software continue to outstrip advances in statistical modeling, making more visualization a no brainer.

G-3 *CONSIDER POSTERIORS & PRIORS TOO ("IT PAYS TO GO BAYES")*

- The sample on hand is rarely the only thing that is "known". Subject matter expertise is an important prior input to statistical design and affects analysis of posterior results



- Gosset at Guinness was wise to keep quality assurance metrics and bottom line profit at the center of his inquiry
- How does prior information fit into the story and evidence? Advances in Bayesian computing software make it easier and easier to do a Bayesian analysis, merging prior and posterior information, values, and knowledge

G-2 *COOPERATE UP, DOWN, AND ACROSS (NETWORKS AND VALUE CHAINS)*

- For example, where would brewers be today without the continued cooperation of farmers? Perhaps back on the farm and not at the brewery making beer
- Statistical science is social, and cooperation helps. Guinness financed a large share of modern statistical theory, and not only by supporting Gosset and other brewers with academic sabbaticals (Ziliak 2018)
- Still, the 120 year old Carlsberg-Guinness relationship is "barley" noticed

G-1 *ANSWER THE BREWER'S ORIGINAL QUESTION ("HOW SHOULD YOU SET THE ODDS?")*

- No bright-line rule of statistical significance can answer the brewer's question
- As Gosset told Karl Pearson (1905):

> "When I first reported on the subject [of "The Application of the 'Law of Error' to the Work of the Brewery" (1904)], I thought that perhaps there might be some degree of probability which is conventionally treated as sufficient in such work as ours and I advised that some outside authority should be consulted as to what certainty is required to aim at in large scale work. However it would appear that in such work as ours *the degree of certainty to be aimed at* must depend on the *pecuniary advantage to be gained by following the result of the experiment, compared with the increased cost of the new method, if any, and the cost of each experiment.*"

HAIKU

Little p-value

What are you trying to say

Of significance?

SOURCE AND COPYRIGHT PAGE

Stephen T. Ziliak, "How Large are Your G-Values? Try Gosset's Guinnessometrics When a Little 'p' is Not Enough," *The American Statistician* (forthcoming Fall 2018, special issue on Statistical Inference and Scientific Method in the World Beyond P < 0.05). This paper was presented at Hacktoberfest 2018, a joint meeting of the Chicago R-User and R-Ladies Group, Haymarket Brewery, Oct. 16, 2018 Chicago

Ziliak is Professor of Economics at Roosevelt University and Conjoint Professor of Business and Law at the University of Newcastle (Australia)

Email: sziliak@Roosevelt.edu Website: https://blogs.Roosevelt.edu/sziliak