How Large Are Your G-Values? Try Gossets Guinnessometrics When a Little p Is Not Enough

314004 Reading Course "Statistische Wissenschaften"

T. Goldbecker, H. Nordemann, M. Paluch

January 18, 2021

Bielefeld University

Table of Contents

Introduction

The Author

- Stephen T. Ziliak
- Economic Statistician and Historian
- Professor of economics at Roosevelt University, Chicago, IL.
- Claims, that he has found "a statistical giant, whose methods, when heeded, could help [...] more or less instantly."
- A lot of his work is based on the "invisible giant"

Figure 1: Stephen T. Ziliak

William S. Gosset (1876-1937)

- Better known as “Student”
- Self-trained statistician, chemist and Head Brewer at Guinness Brewery in Dublin (1899-1937)
- 19 of 21 publications with pseudonym
- Best known for Student’s t-distribution, table and test of significance
- Guinness was largest brewery in the world
- Important to consider costs

Figure 2: William S. Gosset

Motivation

How can William S. Gosset help us today?

Crises

Crisis of replication: cannot repeat an entire study
Crisis of reproducibility: cannot reproduce numerical findings with original data
Crisis of statistical significance: cannot conclude from statistical significance to substantive significance

---

$=$ **Crisis of validity**

Origins of the current crisis

9 reasons for the crisis:

- Attempted falsification of an assumed-to-be true null hypothesis without a loss function
- Reduction of problem to one value (like $p<0.05$ or $t>1.96$)
- Assuming independence of observations
- "One and done"
- Single sample of convenience and arbitrary size
- Irreproducible results
- Big Data is not the solution for everything
- Neglecting prior knowledge
- Binary decisions: significant or not significant

Guinnessometrics

What is Guinnessometrics?

- Name chosen by S. T. Ziliak
- Collection of methods first used by Gosset at Guinness

"Guinnessometrics is [...] an experimental approach to decision-making, which is based on statistical methodologies."[1]

- Experimental: makes use of experiments
- Decision-making: use gained knowledge to back up and finally make a decision
- Statistical methodologies: support with quantitative foundation

[1] Stats+Stories Interview
https://statsandstories.net/methods/so-what-is-guinnessometrics

10 G-Values

- 10 principles of Guinnessometrics
- Not mathematical values, rather a guideline
- Goal: fulfill as many as possible ($\rightsquigarrow$ the larger, the better) with maximum possible strength
- Key: substantive significance emerging from small series of independent and economical experiments
- practical orientation
- little or no need for a p-value or randomized placebo-controlled trial.

Excursion: Sampling Theory

To understand the G-Values better, we (very) briefly introduce the following concepts:

- Observational Study vs. Experimental Study
- Random Sampling and Stratification
- Stratification vs. Blocking
- Randomized controlled trial (RCT)

Source: [Montana State University] and [Unicef]

Observational Study vs. Experimental Study

Observational Study

Observe individuals and measure variables, do not try to influence responses.

Cannot be used to derive cause-and-effect relationships (only correlation).

Experimental Study

Impose one or more treatments on individuals in order to observe their response.

Can be used to derive cause-and-effect relationships.

Random Sampling and Stratification

Random Sample

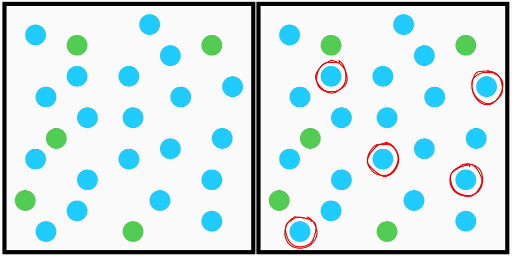
A sample containing individuals randomly chosen from the population of interest.

Types:

- **Simple Random Sample**
  Take randomly from the population as is, each member of the population has the same chance to be selected.
- **Stratified Random Sample**
  Separate population into non-overlapping homogeneous groups called **strata**, then perform simple random sampling within each stratum, finally combine them.

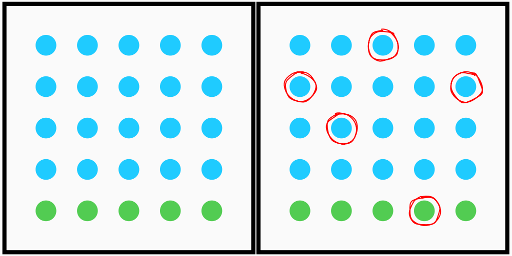
Example

Simple Random Sample



Example

Stratified Random Sample



Stratification vs. Blocking

Stratification

A sampling technique; addresses variables that are given and cannot be controlled (like *sex*, *age*, *ethnicity*, . . .).

Blocking

Part of a study design; addresses variables that can be controlled (e.g. one block gets an experimental drug, the other block a placebo).

- techniques to reduce noise/variance in the data
- variability within each block/stratum lower than variability in entire sample
- estimates of effect more efficient within each block

Randomized controlled trial (RCT)

- experimental form of impact evaluation
- use Simple/Stratified Random Sampling to generate a representative sample from the population of interest
- randomly assign units to two or more groups, e.g. one experimental group which receives the treatment, one control group which does receive a placebo $\rightsquigarrow$ Randomized
- if all groups of same size: *balanced* design
- can now compare effects/results between groups $\rightsquigarrow$ Controlled

G-Values

We now start to introduce the G-Values in reversed order, that is starting with G-10 and ending with G-1.

G-10 Consider The Purpose Of The Inquiry, And Compare With Best Practice

Purpose of a study/experiment?

- Saving lives/money
- Advancing health
- Increasing wealth
- Making beer/medicine
- ...

$\rightsquigarrow$ Falsification of a Null Hypothesis is **not** the purpose of a study/experiment!

G-10 Consider The Purpose Of The Inquiry, And Compare With Best Practice

Best Practice Research

- There is prior interest, else why test?
- Compare novel treatment/variable to best practice/prevailing wisdom, not to assumed-to-be-true H_0 or placebo
- Purpose of experiment has little/nothing to do with significance of H_0
- Gosset's statement about the null test-and-p-value procedure:
  merely something to set up like a coconut to stand until it is hit.

G-10 Example

Test different varieties of cereals to find out which maximizes the yield (and therefore the farmers money).

- Design and repeat small samples of stratified & balanced experiments
- Replace native cereals with improved cereal variety

$$
\begin{array}{lll}
\text{Gain:} & +\ \pounds\, 250.000/\text{year} & \times\ 25\ \text{years} \\
\text{Cost:} & -\ \pounds\, 40.000 & \text{(cumulated)} \\
\hline
\text{Net Gain:} & +\ \pounds\, 6.210.000 &
\end{array}
$$

$\rightsquigarrow$ *Cannot* assess significance of results without weighing its utility and disutility

G-9 Estimate The Stakes (Or Eat Them)

- Purpose of an experiment: put something substantive at stake (*belief, technology, policy variables, ...*)
- Meaning of an experiment are not t- & p-values, but lies in magnitudes of effects (*bees, beer, love, basketball, ...*)
- Estimation of magnitudes of effects should be center of inquiries
- Failure to specify stakes of a hypothesis could be devastating
- Quantities of substantive gain/loss central to Guinnessometrics approach
- Experiment with hops: p-value ≈ 0.07, still led to a huge amount of net gain [Ziliak, 2008]

G-8 Study Correlated Data: Abba, Take a Chance on Me

Currently popular: use completely randomized studies consisting of a treatment and control group.

Is this always the best one can do?

G-8 Study Correlated Data: Abba, Take a Chance on Me

- Complete randomization done when
  - No reasonable grouping suggest itself
  - The investigator knows little or nothing about stratification
  - Cost of being wrong is negligible
- Stratification or blocking can add precision and efficiency to a study (compared to completely randomized studies)
- Stratification can reduce required sample size by 40% or more and is never worse than Randomized Sampling
- Aim of stratified sampling:
  - Increase amount of information per unit of cost
  - Obtain information about certain stratum of interest

G-8 Study Correlated Data: Abba, Take a Chance on Me

So why ABBA?

- Gosset performed different experiments in soil
- Soil nowhere really uniform; varies from inch to inch
- Variation not random $\rightsquigarrow$ methods for combining errors based on randomness not really applicable
- Called his field layout ABBA; mirror pattern of the layout with A's (treatments) and B's (controls)
- Reflects the use of blocking and stratification techniques

A	A	A	A	A	A	B	B
A	A	A	A	A	A	B	A
A	A	A	A	A	B	A	B
A	A	A	A	A	B	B	A
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

$\longrightarrow$

soil fertility

A	B	B	A	A	B	B	A
A	B	B	A	A	B	B	A
A	B	B	A	A	B	B	A
A	B	B	A	A	B	B	A
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

$\longrightarrow$

soil fertility

G-7 Minimize “Real Error” With The 3 Rs: Represent, Replicate, Reproduce

Reproducibility

A study is reproducible if you can take the original data and the computer code and reproduce all of the numerical findings from the study.

Replicability

The act of repeating an entire study, independently of the original investigator and without the use of the original data.

Representation

Coverage and stratification of all systematic sources of fluctuation.

G-7 Example

We now explore the meaning of the 3 R's in the context of the following question:

> *In what soil and under what conditions of weather does a new barley*[2] *variety outperform an old barley variety in terms of quality?*

[2] Gerste

G-7 Example

- Perform experiments on a single farm X
  - Single data set $\rightsquigarrow$ significance test nearly valueless
  - Question answered: *In soils like at the farm X the new variety did not have greater quality than the old one.*
- Next question: *What is soil like "at the farm X"*?
  - Conduct experiments at other farms, find differences

G-7 Example

To answer the original question, Gosset did the following:

- Conduct not only one experiment, but several ones simultaneously at 10 farms scattered across the country (Replication)
- For insurance, each farmer plants a replicate locally $\rightsquigarrow$ 20 replications in one year (Reproducibility)
- Use representative samples of the farms, correlate with the characteristics of the soil and weather (Representation)

G-7 Minimize "Real Error" With The 3 Rs: Represent, Replicate, Reproduce

Problems with the current approach broadly used and established in sciences:

- Most statistical studies in various sciences perform a single randomized controlled trial using a single sample $\rightsquigarrow$ "one and done".
- Example: one regression model on one data set followed by several tests of significance using NHST and p-values ("p-hacking")
- Treating single samples as if they were repeated samples
- Sterling survey (1959): probability of *replication* decreases with the level of statistical significance
  $\rightsquigarrow$ the lower the p value, the lower the likelihood the study will be replicated

G-7 Minimize "Real Error" With The 3 Rs: Represent, Replicate, Reproduce

To overcome the aforementioned problems, Gosset noted the following and relied on the 3 R's:

- A test of significance on a single data set is nearly valueless
- One should perform t-tests and p-tests only when experiments are repeated several times using new data sets
- Experimental hypothesis confirmed in 19 out of 20 farms
  real error: result of 20th farm
- To minimize the real error, repeat experiments on several stratified, balanced and independent experimental units

G-6 Economize With “Less is More”: Small Samples of Independent Experiments

Due to economic reasons, large sample sizes are not always possible. Gosset also had this problem in the brewery:

- Despite different barley varieties, it should always be ensured that the malt extract is always in the neighbourhood of 133 saccharine
- Errors in the saccharine level can be expensive
- So he had the following problem:
    - For economic reasons, he would like to take as few malt extracts as possible (smallest possible sample size)
    - Nevertheless, he would like to say to a certain probability that the malt extract should be within 0.5-interval of the true result

G-6 Economize With “Less is More”: Small Samples of Independent Experiments

How to choose this probability?

- Gosset rejected artificial rules about significance
- So the odds should depend on:
  - the degree of accuracy which the nature of the experiment allows
  - the importance of the study

In our example:

- *With a probability of 10:1 the estimated malt extract should be in the 0.5-interval around the true result.*

G-6 Economize With "Less is More": Small Samples of Independent Experiments

Positive correlation

Larger sample size $\rightsquigarrow$ Smaller error of the estimates

- Gosset became aware of this correlation, which was largely unexplored at the time, during an experiment:
    - He compared the saccharine content in malt extracts from the experimental brewery with those from the main brewery
- Moreover, he claimed that with sufficient sample size, almost anything can be statistically significant ($\nRightarrow$ substantive important)
- On the other hand, studies with small sample sizes can be substantive important. ($\nRightarrow$ statistically significant)

G-6 Economize With “Less is More”: Small Samples of Independent Experiments

Back to our example from the beginning:

- Gosset calculated the probabilities of meeting the required precision for different sample sizes:
  - Odds in favour of smaller error than 0.5 are with **2** observations **4:1**
  - Odds in favour of smaller error than 0.5 are with **3** observations **7:1**
  - Odds in favour of smaller error than 0.5 are with **4** observations **12:1**
  - Odds in favour of smaller error than 0.5 are with **5** observations **19:1**

  $\leadsto$ Gosset could take the mean of the at least last four determinations to get the accuracy he required

G-6 Economize With “Less is More”: Small Samples of Independent Experiments

Gosset succeeded in finding a method with which he could evaluate the behaviour of population parameters with small samples.

- In this example, scarcity was the main factor leading to the small sample analysis
    - Large sample sizes are too expensive
    - *“Big Data is in any case not the solution for many problems”*

$\rightsquigarrow$ A small sample analysis is an economic and scientific choice

G-5 Keep Your Eyes on The Size Matters/How Much? Questions

The fifth G-value says that one should always orient oneself to one's loss-function

- How do our regression coefficients behave?
- Small effect sizes are also important

We will explain this in more detail with an example.

G-5 The “flutamide”-example

- There is often no agreement on the effect sizes of drugs
- In 1995, however, 10 authors of independent and randomized clinical trials agreed on the effect size of the drug “flutamide”
    - Flutamide was then being researched for use in prostate cancer
    - The cancer epidemiologists agreed on a noninvasive 12-20% increase in survival

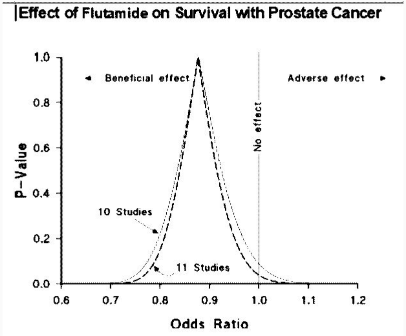
G-5 The "flutamide"-example

- In 1998 Eisenberg published an eleventh, the biggest so far, study dealing with this issue
  - The effect size found was similar,
  - but they did not consider it significant due to a p-value of 0.14
    $\rightsquigarrow$ *"no clinically meaningful improvement"*
- Thereupon Kenneth Rothman and other researchers looked at all 11 studies
  - The pooled data of the first ten studies result in a OR of 0.88[3]
  - Using Eisenberg's data, they calculated a similar OR of 0.87
    $\rightsquigarrow$ Both studies show the same 12% benefit in the odds

[3] The odds ration (OR) is calculated as follows: $OR = \frac{died_{flutamid}}{survived_{flutamid}} / \frac{died_{placebo}}{survived_{placebo}}$
$OR < 1$ is indicating a positive effect.

G-5 The "flutamide"-example

- The p-value function allows to measure the compatibility of the data with each effect size. (instead of just H_0)
- This enables us to achieve the desired quantitative assessment of the data
- Rothman compared the function of the data from the first ten studies with with the resulting function when taking the data from all eleven studies



The graph shows the *P*-value function for the first 10 trials of flutamide (solid line), and the updated function that includes the 1998 trial by Eisenberger *et al.* (dashed line). Although Eisenberger *et al.* interpreted their data as showing no benefit for flutamide based on a lack of statistical significance ($P = 0.14$), their study reinforces the findings of the previous 10 studies, leading to a more precise estimate of a beneficial effect.

Figure 3: p-value-functions [Rothman et al., 1999]

G-5 The “flutamide”-example

The data from Eisenberg’s eleventh study only reinforce previously published conclusions

- The p-value function is narrower
  ⇝ estimate is more precise
  ⇝ Even clearer evidence of the positive effect of flutamide

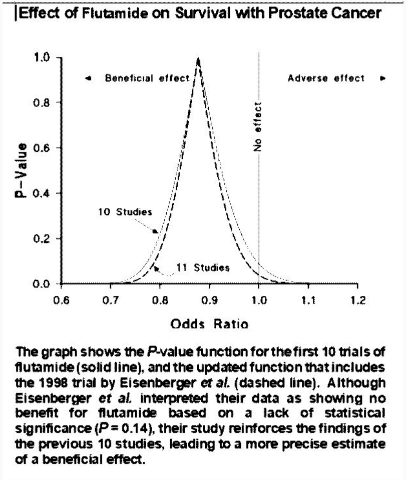


Figure 3: p-value-functions [Rothman et al., 1999]

G-5 Keep Your Eyes on The Size Matters/How Much? Questions

What do we learn from this example?

- Concentration on quantitative measurement of effects instead of significance tests
- *Keep your eyes on the size*
  - Every minimal effect can be important

$\rightsquigarrow$ At the end our loss-function should rule.

G-4 Visualize

- So we see that visualisations can often add some value
- This will also be the subject of G-4
- We introduce G-4 by an example

G-4 Example

- This example consists of 4 data sets:

I.		II.		III.		IV.	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.1	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

G-4 Example

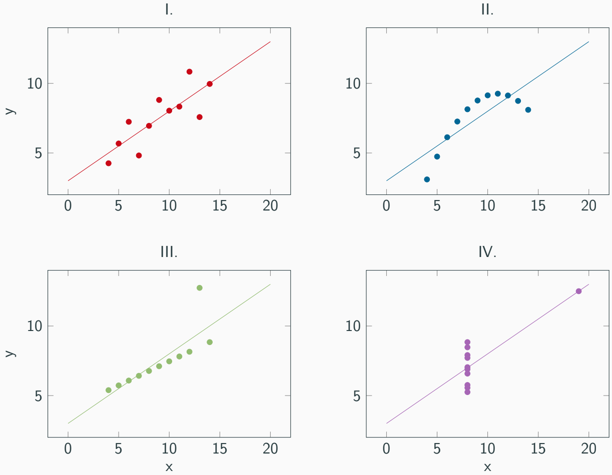
- First steps are now often the calculation of certain values for the four groups:

	I.	II.	III.	IV.
$\bar{x}$	9	9	9	9
$\bar{y}$	7.5	7.5	7.5	7.5
$sd(x)$	3.32	3.32	3.32	3.32
$sd(y)$	2.03	2.03	2.03	2.03
$cor(x, y)$	0.816	0.816	0.816	0.817
linear regression line	y = 3.00 + 0.500x	y = 3.00 + 0.500x	y = 3.00 + 0.500x	y = 3.00 + 0.500x

- According to these key numbers, the four data sets appear to be almost identical
- Often, the analysis of some statisticians already ends here

G-4 Example

A visualisation can often represent the data sets much better:



This example is known as Anscombe's quartet [Anscombe, 1973].

G-4 Visualize

Pay attention to the actual distribution of the data.

- Model uncertainty is not the same thing as parameter uncertainty
- We want to know how the entire distribution looks like. Including...
    - the magnitudes of relationships,
    - the precision,
    - the variation of the variance,
    - ...

G-3 Consider Posteriors And Priors Too (It Pays To Go Bayes)

Reminder: Bayes

$$p(B \mid A) = \frac{p(A \mid B) \cdot p(B)}{p(A)}$$

- $p(B \mid A) \rightsquigarrow$ "posterior"
- $p(A \mid B) \rightsquigarrow$ "likelihood"
- $p(A) \rightsquigarrow$ "evidence"
- $p(B) \rightsquigarrow$ "prior"

Reminder: Bayes Factor

$$BF = \frac{\text{likelihood of the data given } H_1}{\text{likelihood of the data given } H_0}$$

G-3 Consider Posteriors And Priors Too (It Pays To Go Bayes)

$$\underbrace{\frac{p(M_1 \mid \mathcal{D})}{p(M_0 \mid \mathcal{D})}}_{\text{posterior odds}} = \underbrace{\frac{p(\mathcal{D} \mid M_1)}{p(\mathcal{D} \mid M_0)}}_{\text{Bayes factor}} \cdot \underbrace{\frac{p(M_1)}{p(M_0)}}_{\text{prior odds}}$$

- Bayes factor easily computed (Usually H_1 is given)
- Use new evidence combined with prior knowledge to compute the posterior probability that a hypothesis is true
- With the help of the posterior odds it is possible to compute the posterior probability that a null hypothesis is true relative to an explicit alternative
  $\rightsquigarrow$ Classical tests of significance are not able to do that

G-3 Consider Posteriors And Priors Too (It Pays To Go Bayes)

- The data is often not the only thing we know
- Why should we start every time from the beginning?
- Through bayesian methods we can include this available information

G-2 Cooperate Up, Down, And Across (Networks And Value Chains)

"Statistical science is social and cooperation helps."

- In the Guinness context, the collaboration with the Irish Department of Agriculture is explicitly mentioned
- Cross-cutting collaborations of statisticians with other scientists can have several advantages:
    - What is the best practice? (compare G-10)
    - Determination of the prior (compare G-3)
    - ...

G-1 Answer The Brewers Original Question ("How Should You Set The Odds?")

> *The degree of certainty to be aimed at must depend on the pecuniary advantage*[4] *to be gained by following the result of the experiment, compared with the increased cost of the new method*

- Gosset mainly looked for financial advantages
- Depending on this potential financial benefit, the degree of certainty to be aimed at was determined
- We do not want binary decisions which are based only on significance
- **Answer the original question!**

[4] finanzieller Vorteil

Conclusion

Summary

- Do not attempt to falsify a null hypothesis
- Consider magnitude of effect
- Use small samples of independent experiments
- Visualize
- Use loss-function
- Use prior and analyse effects on posterior
- Cooperate

Scientific Classification

According to Google Scholar, this article has 9 citations (as of January 18, 2021) and we had access to 7 of them.

Unfortunately, in none of these papers the method was used.

Summary of the key points:

- Use method to design future studies, especially answering the questions "How much is the variable affected?" and "Why does this effect size matter?" [Woolet and Whitman, 2020]
- Do not base decision solely on statistical significance [Valeria and Felipe, 2019] [Abu and Goldberg, 2020]
- Summary and discussion of the risks of significance testing [Abu and Goldberg, 2020]
- Statistical inference is not equal to scientific inference [Wasserstein et al., 2019]

Conclusion i

- Guinnessometrics supplies useful approaches which are worth taking a look at even 80 years after Gosset's death.
- Whether or not the 10 G-Values do solve the crisis of validity depends on the realization, since values are broad.
- Nevertheless, this paper makes clear, that we do not want to know, if a value is statistically significant but the pecuniary advantage, number of lives saved, jobs gained, racism abolished, health crisis averted...

Conclusion ii

"No one could plausibly claim that the 10 G-Values are the end-all, be-all of statistical science. Only that for a great variety of business, medical, and scientific purposes, the Guinnessometric approach to data and decision-making answers far more questions, and far more satisfactorily, than the conventional, unrepeated observational study or placebo controlled RCT judged by the level of a p-value."

Thank You for your attention!

References i

H. O. Abu and R. J. Goldberg. Make scientific reading great and more meaningful again: Reappraisal of the traditional p value in modern-day clinical research and practice. *The American Journal of Medicine*, 133(1):4 – 6, 2020. ISSN 0002-9343. doi: https://doi.org/10.1016/j.amjmed.2019.05.054. URL http://www.sciencedirect.com/science/article/pii/S0002934319305364.

F. J. Anscombe. Graphs in statistical analysis. *The American Statistician*, 27: 17–21, 1973.

Montana State University. Sampling and experimental design. https://math.montana.edu/parker/courses/STAT401/Chapter1.3-1.5.pdf.

K. J. Rothman, E. S. Johnson, and D. S. Sugano. Is flutamide effective in patients with bilateral orchiectomy? *The Lancet*, 353:1184, 1999.

References ii

Unicef. Randomized controlled trials (rcts). https://www.unicef-irc.org/KM/IE/impact_7.php.

D.-R. Valeria and P.-V. Felipe. El informe estadístico: Discusión sobre la significancia científica. 2019.

R. L. Wasserstein, A. L. Schirm, and N. A. Lazar. Moving to a world beyond p¡ 0.05, 2019.

J. Woolet and T. Whitman. Pyrogenic organic matter effects on soil bacterial community composition. *Soil Biology and Biochemistry*, 141:107678, 2020.

S. T. Ziliak. Retrospectives: Guinnessometrics: The economic foundation of "student's" t. *Journal of Economic Perspectives*, 22(4):199–216, December 2008. doi: 10.1257/jep.22.4.199. URL https://www.aeaweb.org/articles?id=10.1257/jep.22.4.199.

Discussion

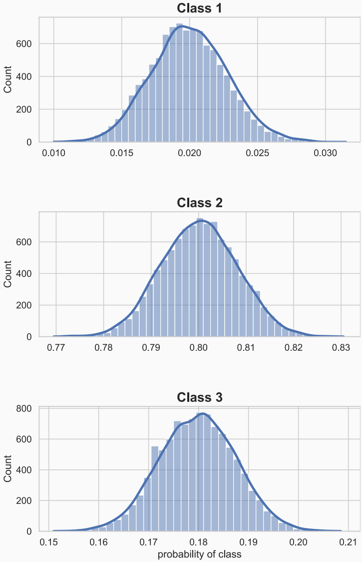
Appendix

Excursion Example: Distribution of Class Proportions

Using **Simple Random Sample**

- population: 10.000 observations with class 1 (2%), class 2 (80%), class 3 (18%)
- draw sample of size 2.000 (20%)
- compute proportions of each class within that sample
- repeat 10.000 times, draw distribution of class proportions

Using **Stratified Random Sampling**, the class proportions of each class of that sample corresponds to the class proportions of the population (per construction!)



G-9 Estimate The Stakes (Or Eat Them)

- Magnitudes of gains/losses expressed with expected loss function
- However: expected value not reliable guide for action (e.g. to measure human behaviour)
- Better: favor expected *utility* over expected value

G-9: Example

Flip a coin (Frank 2017):

- (G1) win \$100 if tails, lose \$0.50 if heads
- (G2) win \$200 if tails, lose \$100 if heads
- (G3) win \$20,000 if tails, lose \$10,000 if heads

Expected values:

- $E(\text{G1}) = (\$100) \times 0.5 + (-\$0.50) \times 0.5 = \$49.75$
- $E(\text{G2}) = (\$200) \times 0.5 + (-\$100) \times 0.5 = \$50$
- $E(\text{G3}) = (\$20{,}000) \times 0.5 + (-\$10{,}000) \times 0.5 = \$5{,}000$

G-9: Example

- G1 accepted by many, followed by G2 and G3
- G2 and G3 have same payoff structure (win twice as much as lose)
- Though G3 has highest expected value, but many would reject it
- Simple reason: losing $10,000 would be tragic

G-9: Example

Same gamble as before:

- (G1) win \$100 if tails, lose \$0.50 if heads
- (G2) win \$200 if tails, lose \$100 if heads
- (G3) win \$20,000 if tails, lose \$10,000 if heads

Expected utility (EU) using $U(x) = \sqrt{x}$, initial amount of money: \$10,000

- $EU(\text{G1}) = (\sqrt{\$10{,}100}) \times 0.5 + (\sqrt{\$9{,}999.50}) \times 0.5 = 100.248$ utils
- $EU(\text{G2}) = (\sqrt{\$10{,}200}) \times 0.5 + (\sqrt{\$9{,}900}) \times 0.5 = 99.999$ utils
- $EU(\text{G3}) = (\sqrt{\$30{,}000}) \times 0.5 + (\sqrt{\$0}) \times 0.5 = 86.603$ utils

G-9 Estimate The Stakes (Or Eat Them)

- Expected utility has problems/limitations (as does the expected value)
- but can account for attitudes towards risk
- Quantities of substantive gain/loss central to Guinnessometrics approach

Advantages

- Not dependent on one single value
- Keep in mind the real interest
- Practical approach
- Useful guideline

Disadvantages

- Values are rather broad
- Leaves open questions:
    - When is a G-Value fulfilled?
    - Are all values equally important? Do they all count the same?
    - Is there a minimum amount of G-Values that must be fulfilled?
    - Can you compare studies by G-Values?
- There is nothing really "new"
- Generality of the method
- Some literature only available in Guinness Archives