Teaching Intro Stat in a World Beyond "p<0.05"

Catherine Case
catherine.case@uga.edu

Motivation

- "Statistical education will require major changes at all levels to move to a post 'p<0.05' world."
  - Wasserstein, Scherm, and Lazar (2019)
- "Don't is not enough."
  - Not for users of statistics *or* teachers of statistics
- What strategies can we use to *replace* a black-and-white view of statistical significance?

Backward Design

- Backward Design (Wiggins & McTighe, 1998) is a method of curriculum planning with three stages

Outcomes

Based on reforms suggested in the editorial "Moving to a World Beyond $p<0.05$" Wasserstein, Scherm, and Lazar (2019)

18 outcomes (divided into five categories) that are achievable in an introductory statistics class.

I. P-values
II. Errors and Power
III. Confidence Intervals
IV. Effect sizes
V. The Research Process

Outcomes

P-values

- 1A: Interpret p-values as continuous probabilities.
  - Avoid common misinterpretations of p-values. For example, a p-value does not provide the probability that the null hypothesis is true.

Despite the limitations of *p*-values (as noted in Principles 5 and 6 of the ASA statement), however, we are not recommending that the calculation and use of continuous *p*-values be discontinued. Where *p*-values are used, they should be reported as continuous quantities (e.g., $p = 0.08$). They should also be described in language stating what the value means in the scientific context. We believe that a reasonable prerequisite for reporting any *p*-value is the ability to interpret it appropriately. We say more about this in Section 3.3.

Outcomes

Confidence intervals

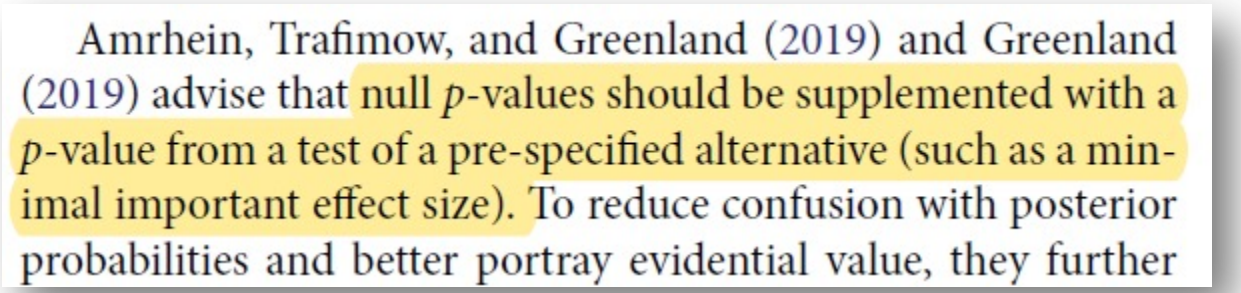
- 3C: Interpret a confidence interval in context and consider whether the upper and lower limits have different practical implications.
- 3D: Recognize a confidence interval as an estimate subject to error, and distinguish between random and non-random errors.

(Greenland 2019), which is described in Section 3.2. As noted in Section 3.1, you might present a confidence interval. Sound practices in the interpretation of confidence intervals include (1) discussing both the upper and lower limits and whether they have different practical implications, (2) paying no particular attention to whether the interval includes the null value, and (3) remembering that an interval is itself an estimate subject to error and generally provides only a rough indication of uncertainty given that all of the assumptions used to create it are correct and, thus, for example, does not "rule out" values outside the interval.

Outcomes

Effect sizes

- 4C: Calculate a p-value from a test of a pre-specified alternative, such as a minimal important effect size.



Amrhein, Trafimow, and Greenland (2019) and Greenland (2019) advise that null *p*-values should be supplemented with a *p*-value from a test of a pre-specified alternative (such as a minimal important effect size). To reduce confusion with posterior probabilities and better portray evidential value, they further

Outcomes

The Research Process

- 5C. Explain why different p-values in replication studies do not always imply inconsistent results.

Indeed, when it comes to reproducibility, Amrhein, Trafimow, and Greenland (2019) demand that we **be modest** in our expectations. "An important role for statistics in research is the summary and accumulation of information," they say. "If replications do not find the same results, this is not necessarily a crisis, but is part of a natural process by which science evolves. The goal of scientific methodology should be to direct this evolution toward ever more accurate descriptions of the world and how it works, not toward ever more publication of inferences, conclusions, or decisions."

Outcomes: P-values

A. Interpret p-values as continuous probabilities.
   - Avoid common misinterpretations.

B. Use p-values to describe the strength of evidence against a stated hypothesis.
   - Explain why a small p-value provides strong evidence.

C. Predict how changes to the effect size or sample size will impact the p-value.

D. Understand what others mean when they use the term *statistically significant*.
   - Recognize common misunderstandings associated with this term.

Assessment: P-values

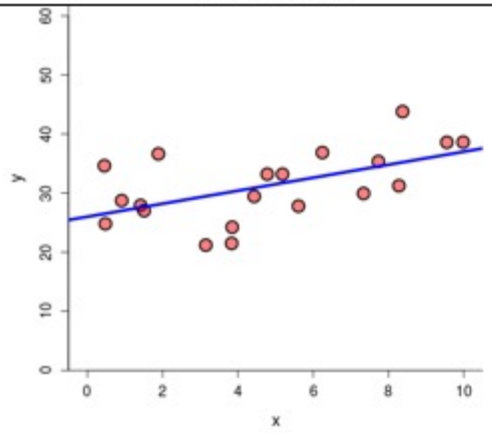
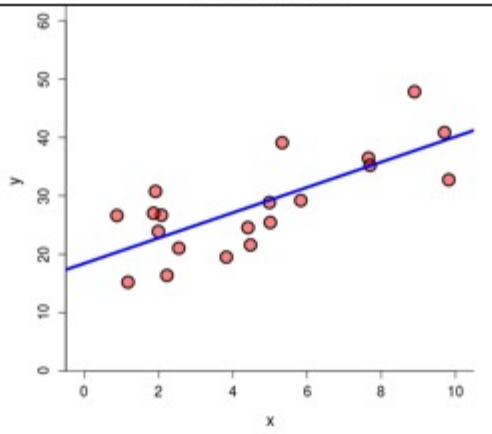
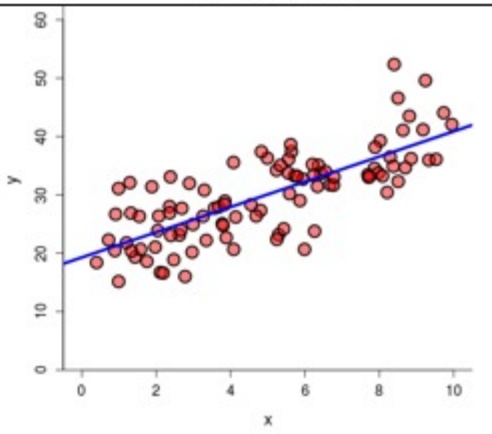
In 2015, Airbnb reported that 63% of its listings are for the entire home. Recently a sample of 250 listings from Asheville, NC was collected, and 182 out of 250 listings were for the entire home.

- The two-sided p-value is 0.0013. Interpret this p-value as a probability.
  - It is the probability of what, assuming what?

- Does this data provide strong evidence that the percentage of Airbnb listings that offer entire homes is higher than 63% in Asheville?
  - For this question, it is not enough to say yes/no because the p-value is large/small. Explain why the size of the p-value supports your answer.

Assessment: P-values

- You are conducting a hypothesis test of $H_0: \beta = 0$ vs. $H_A: \beta \neq 0$. Which of the datasets below would have the smallest p-value?

Dataset A	Dataset B	Dataset C
		
Sample size = 20 Slope = 1.10, correlation = 0.56	Sample size = 20 Slope = 2.17, correlation = 0.74	Sample size = 100 Slope = 2.17, correlation = 0.74

Assessment: P-values

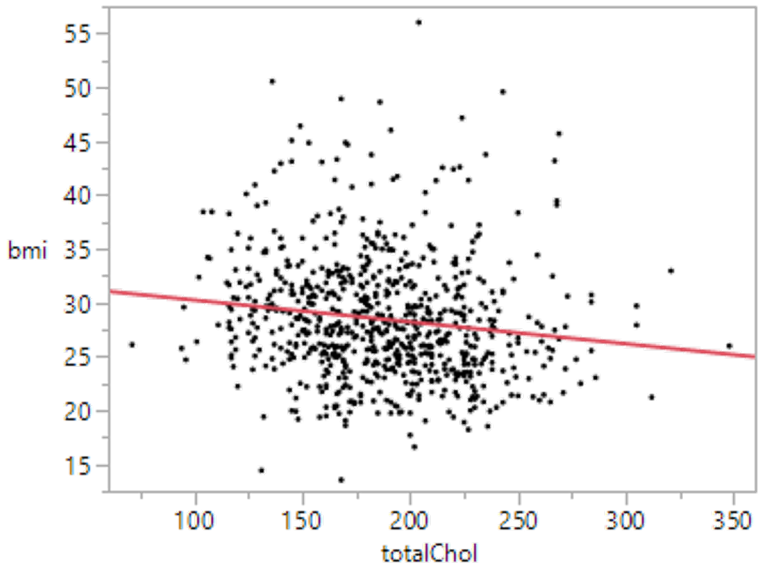
- "One study published in the *Annals of Internal Medicine* journal studied 521,330 people in 10 European countries over 16 years and found a statistically significant lower mortality rate among coffee drinkers."
  - What does "statistically significant" mean in this context?

Learning experiences: P-values

- Introduce high standards for statistical communication through formative assessment.
- Explore examples that challenge common misunderstandings.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	32.26809	1.003113	32.17	<.0001*
totalChol	-0.020267	0.005257	-3.86	0.0001*



Learning experiences: P-values

- You may have heard that bacon causes cancer or that eating bacon is as bad as smoking.
- Is the link between processed meat and cancer **statistically significant**?
  - What does that mean?
  - How is it presented in the article?
- Is the link between processed meat and cancer **practically important**?
  - This one is more subjective.
  - What information do you need in order to decide?

Learning experiences: P-values

Yes, coffee drinkers seem to live longer. But don't get too excited.

The question isn't if, it's how much.

BY SARA CHODOSH JULY 11, 2017

HEALTH

Science Isn't Broken

It's just a hell of a lot harder than we give it credit for.

By Christie Aschwanden
Graphics by Ritchie King
Filed under Scientific Method
Published Aug. 19, 2015

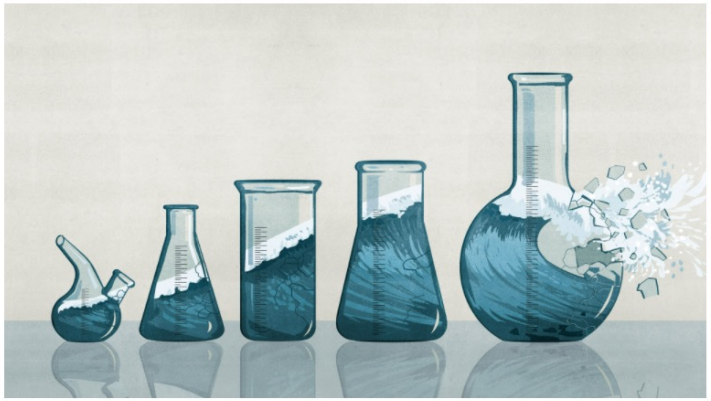
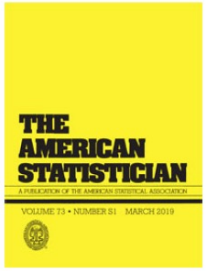


ILLUSTRATION BY SHOUT



The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Moving to a World Beyond "$p < 0.05$"

Teaching Intro Stat in a World Beyond "p<0.05"

Catherine Case
catherine.case@uga.edu