Common Statistical Misinterpretations

Part I

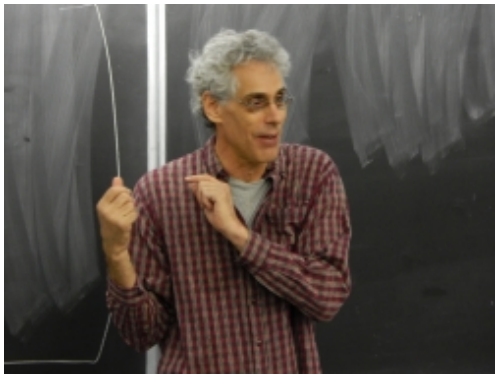
Yadong Lu

Univeristy of California, Irvine

December 3, 2017

Table of Contents

Author:



Sander Greenland (born January 16, 1951) is an American statistician and epidemiologist known for his contributions to epidemiologic methods, meta-analysis, Bayesian inference and causal inference, among other topics.

Educated at the University of California, Berkeley and the University of California, Los Angeles (UCLA), he has held a professorship in epidemiology at UCLA School of Public Health since 1989, and additionally a professorship in statistics at UCLA College of Letters and Science since 1999.

Motivation and Methods:

Motivation

- Epidemic of shortcut definitions and interpretations that are simple, but wrong, disatrously.
- Misinterpretation and abuse: statistical tests, confidence intervals, statistical power.
- Some journal bans all statistical tests and confidence intervals

Motivation and Methods:

Motivation

- Epidemic of shortcut definitions and interpretations that are simple, but wrong, disatrously.
- Misinterpretation and abuse: statistical tests, confidence intervals, statistical power.
- Some journal bans all statistical tests and confidence intervals

Methods

- Explain the meaning of significance tests, statistical power in a general way.
- Review 25 common misconceptions

Understanding Statistical Model, Hypothesis and Tests

Statistical Model

- Mathematical representation of data variability
- A set of assumptions is embodied in a model.

Understanding Statistical Model, Hypothesis and Tests

Statistical Model

- Mathematical representation of data variability
- A set of assumptions is embodied in a model.

Problems:

- Models incorporate unrealistic or unjustified assumptions.
- Hard to define scope of the model. (overfitting)
- Model in a compressed form: lead to unremarked assumptions (Think about poisson model)

Understanding Statistical Model, Hypothesis and Tests

Statistical Hypothesis Tests

- Determine whether a particular effect has a specific size.
- Null hypothesis: target effect size has "null" value or zero effect.
- Can also test whether the effect size fall in a specific range.
- Focusing *only* on null hypothesis test: lead to misunderstanding. Remeber we can test all kinds of target hypothesis rather just the "null" hypothesis.

Uncertainty, probability, and statistical significance

Refined Goal of Statistical Analysis

Evaluation of certainty or uncertainty regarding the size of an effect.

Uncertainty, probability, and statistical significance

Refined Goal of Statistical Analysis

Evaluation of certainty or uncertainty regarding the size of an effect.

Statistics vs. ML

- Machine learning care less about the evaluation of the effect.
- Goal: use the effect (association) to produce prediction(MLP), or generation(Generative Adversarial Net).

Uncertainty, probability, and statistical significance

Probability: in Frequentist World

- In frequentist: "Probability" refers not to hypothesis, but to quantities that are hypothetical frequencies of data patterns under an assumed statistical model
- Think about "likelihood" : refer to the probability of the data given parameters, **not** the probability of the parameter taking on a given value.

Uncertainty, probability, and statistical significance

P Value

- The probability, under the null hypothesis, of obtaining a result equal to or more extreme than what was actually observed.
- It is a hypothetical frequency, also known as the "observed significance level" for the test hypothesis

Uncertainty, probability, and statistical significance

P Value

- The probability, under the null hypothesis, of obtaining a result equal to or more extreme than what was actually observed.
- It is a hypothetical frequency, also known as the "observed significance level" for the test hypothesis

Traditional Definition:

Focusing on null hypothesis, treating all other assumptions to be correct.

Uncertainty, probability, and statistical significance

P Value

- The probability, under the null hypothesis, of obtaining a result equal to or more extreme than what was actually observed.
- It is a hypothetical frequency, also known as the "observed significance level" for the test hypothesis

Traditional Definition:

Focusing on null hypothesis, treating all other assumptions to be correct.

A More General View

P value is a statistical summary of the compatibility between the observed data and what we would predict or expect to see if we knew the entire statistical model (all the assumptions used to compute the P value) were correct.

Uncertainty, probability, and statistical significance

Interpretation:

- It is true that the smaller the P value, the more unusual the data would be if *every* single assumption were correct; but a very small P value *does not* tell us which assumption is incorrect.

Uncertainty, probability, and statistical significance

Interpretation:

- It is true that the smaller the P value, the more unusual the data would be if *every* single assumption were correct; but a very small P value *does not* tell us which assumption is incorrect.
- For example, the P value may be very small because the targeted hypothesis is false; but it may instead (or in addition) be very small because the some assumptions were violated

Uncertainty, probability, and statistical significance

Interpretation:

- It is true that the smaller the P value, the more unusual the data would be if *every* single assumption were correct; but a very small P value *does not* tell us which assumption is incorrect.
- For example, the P value may be very small because the targeted hypothesis is false; but it may instead (or in addition) be very small because the some assumptions were violated
- Conversely, a large P value indicates only that the data are not unusual under the model, but does not imply that the model or any aspect of it (such as the targeted hypothesis) is correct;

Uncertainty, probability, and statistical significance

Interpretation(contd):

- The general definition of a P value may help one to understand why statistical tests tell us much less than what many think they do.
- Need assurance of every other assumptions in the model is correct - an assurance that is lacking in far too many studies.

Uncertainty, probability, and statistical significance

Interpretation(contd):

- The general definition of a P value may help one to understand why statistical tests tell us much less than what many think they do.
- Need assurance of every other assumptions in the model is correct - an assurance that is lacking in far too many studies.

P Value Significance Level:

- Significance level (α level): often used to refer to the "cut-off", fixed, part of study design.
- P value: a number computed from the data.

Moving from Tests to Estimations

- we can vary the test hypothesis (or effect size) to see how P value changes. (eg: we can test $H_0 : \mu = 0, 0.5, 1$ etc)

Moving from Tests to Estimations

- we can vary the test hypothesis (or effect size) to see how P value changes. (eg: we can test $H_0 : \mu = 0, 0.5, 1$ etc)
- Effect sizes whose test produced $P > 0.05$ will define of range of sizes that would be considered more compatible with data than sizes outside the range. (this determines a range for μ)
- The range corresponds to 95% confidence interval.

Moving from Tests to Estimations

- we can vary the test hypothesis (or effect size) to see how P value changes. (eg: we can test $H_0 : \mu = 0, 0.5, 1$ etc)
- Effect sizes whose test produced $P > 0.05$ will define of range of sizes that would be considered more compatible with data than sizes outside the range. (this determines a range for μ)
- The range corresponds to 95% confidence interval.
- Confidence interval can be thus be seen as a convenient way of summarizing the results of hypothesis tests for many effect sizes.

$$P(D_2|p,q) = \sum_{k_1=0}^{30} \sum_{k_2=0}^{30-k_1} \binom{100}{k_1} [q(1-p)]^{k_1} \binom{100-k_1}{k_2} [pq]^{k_2} \binom{100-k_1}{30-k_1}$$

Thank you!

$$P(D_2|p,q) = \sum_{k_1=0}^{30} \sum_{k_2=0}^{30-k_1} \binom{100}{k_1} [q($$