

Discussion: P-values, Alternatives to P-Values, and Statistical Models

David Zucker

Department of Statistics and Data Science
Hebrew University of Jerusalem

P-values

- P-values have been around for about 100 years
- In the last 5-10 years, there has been strong debate on utility of p-values
(quotes in Francq's presentation)

P-values (2)

- The p-value is the probability of occurrence of the observed result (or a more extreme one) if the null hypothesis is true.
- Thus, it is a measure of the strength of the evidence against the null hypothesis
- As such, it is an indirect measure of the strength of the evidence in favor of the alternative hypothesis

Alternative Measures of Strength of Evidence

- s-values
- Bayesian posterior probability
- Likelihood ratio / Bayes factor

$$s = -\log_2 p, \quad p = \left(\frac{1}{2}\right)^s$$

$$p = 0.05 \Leftrightarrow s = 4.3$$

$$p = 0.01 \Leftrightarrow s = 6.6$$

Statistical Significance

- What do we call significant?
- $p=0.048$ versus $p=0.052$
- Does it make sense to have a uniform standard for all of science and everything else?

Individual Success Probability (ISP)

$$\text{ISP} = P(X^{(n+1)} \leq c) \quad (\text{or } \geq) \quad (\text{for example, } c = 140)$$

$X^{(n+1)}$ = outcome value for a new patient

If $X \sim N(\mu, \sigma^2)$, then

$$\text{ISP} = \Phi\left(\frac{c - \mu}{\sigma}\right)$$

ISP vs P-value

- ISP is not a measure of the strength of the evidence: it is a point estimate of a feature of the distribution of the outcome
- Hence, ISP is not a competitor to the p-value

Strength of Evidence

- A measure of the strength of evidence is essential in statistics
 - How loudly are the data speaking?
 - What is the degree of uncertainty in our results?

Some Issues With ISP

- Assumes that the X of a new patient has the same distribution as the X of the patients in the current study
- Distributional assumptions
- How to choose the cutoff?

Significance Revisited

- “The chances of getting the observed result if there were no treatment effect is 5%” – this is a statement of fact.
 - $p=0.05$ is a shorthand version of this statement
- “The data provide statistically significant evidence that the treatment works” – this is an expression of a judgment.

Structure of a Scientific Paper

1. Introduction
2. Methods
3. Results
4. Discussion

A Suggestion

- Keep the word “significant” out of the Results section.
- Reserve “significant” and similar terms for the Discussion section.

Statistical Models

- Truth vs Approximation

Statistical Models – Purposes of Models

- Compact summary of data
 - Goodness of fit vs. simplicity
- Inference about the behavior of the system
 - Is A related to B?

Interpretative H0 vs. H1

- Amygdala data: $r = 0.56$; $p = 0.048$,
- 95% confidence interval [0.01; 0.85] for ρ

“Perceived stress was associated with amygdalar activity.”

vs.

“It appears that there is some association between stress and amygdalar activity, but the association may be extremely weak.”

$$H_0 : \rho \leq \rho^* \text{ vs. } H_1 : \rho > \rho^*$$

Statistical Models

- Not all statistics is model-based
 - Example: Finite population / survey sampling – here there is a truth

Further Remarks

- Model checking
- Robustness
- Sensitivity analysis

Be cautious!