What does $P = 0.035$ mean?

Michael Stob
Calvin University

June ~~2021~~ 2022

P-Value as Evidence

$P = 0.035$ is strong evidence against H_0.

AP Statistics Rubric: Convincing statistical evidence

Fisher Significance Tests

A P-value is a continuous measure of the evidence against H_0. The smaller the P-value, the greater the strength of the evidence.

What is the *P*-value?

The probability that

If the null hypothesis H_0 is true

the test statistic would have a value

as extreme or more extreme

than the observed value

$$P = \text{Prob}(T \geq t_0 \mid H_0)$$

Sherlock Holmes

When you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth.

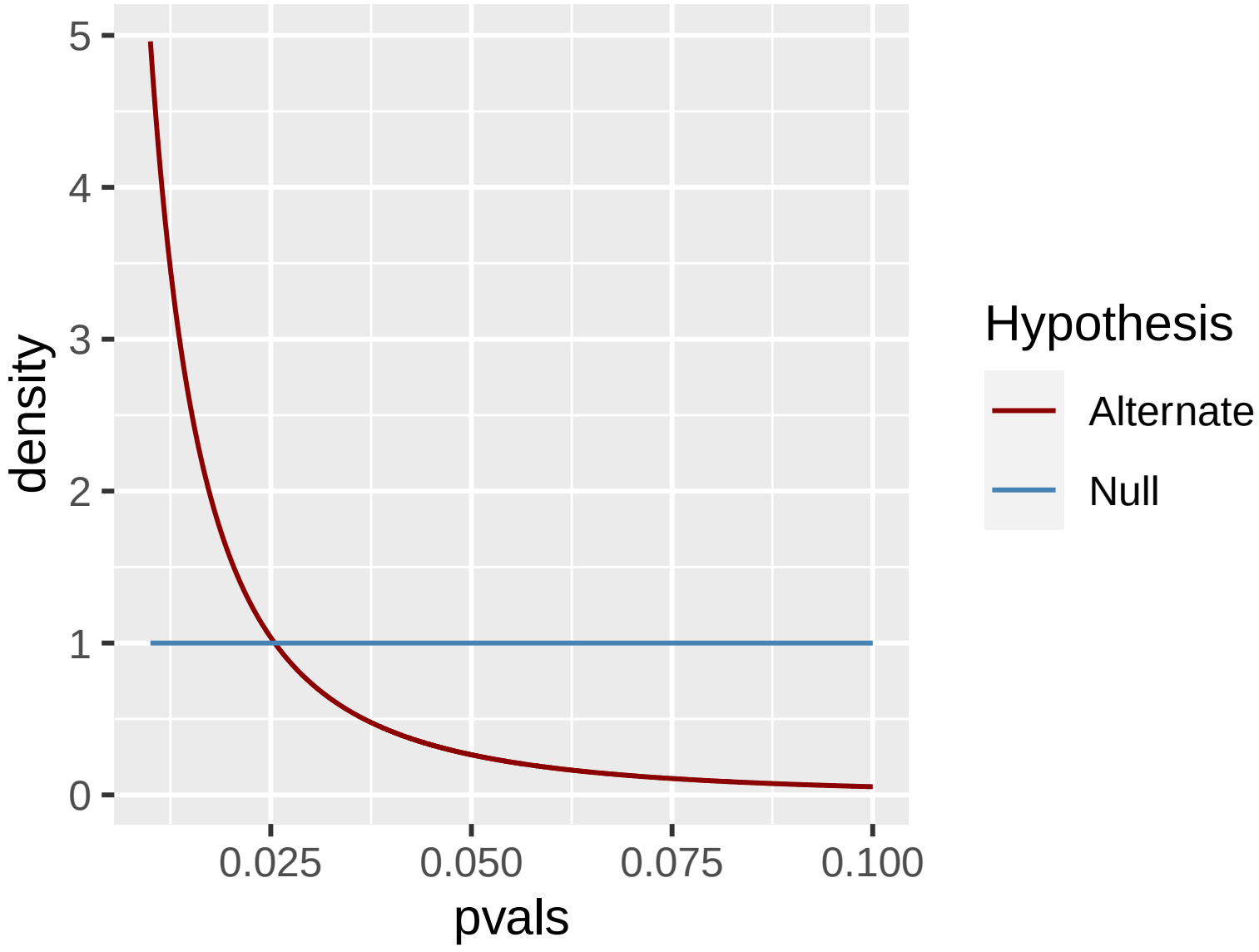
H_0 against alternatives

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_A &: \quad \mu > \mu_0 \end{aligned}$$

Suppose that H_0 is false and $\dfrac{\mu - \mu_0}{\sigma} = 0.75$.

How likely is it that one would see a p-value like 0.035 if $n = 30$?

Distribution of p-values, $n = 30$, $d = 0.75$

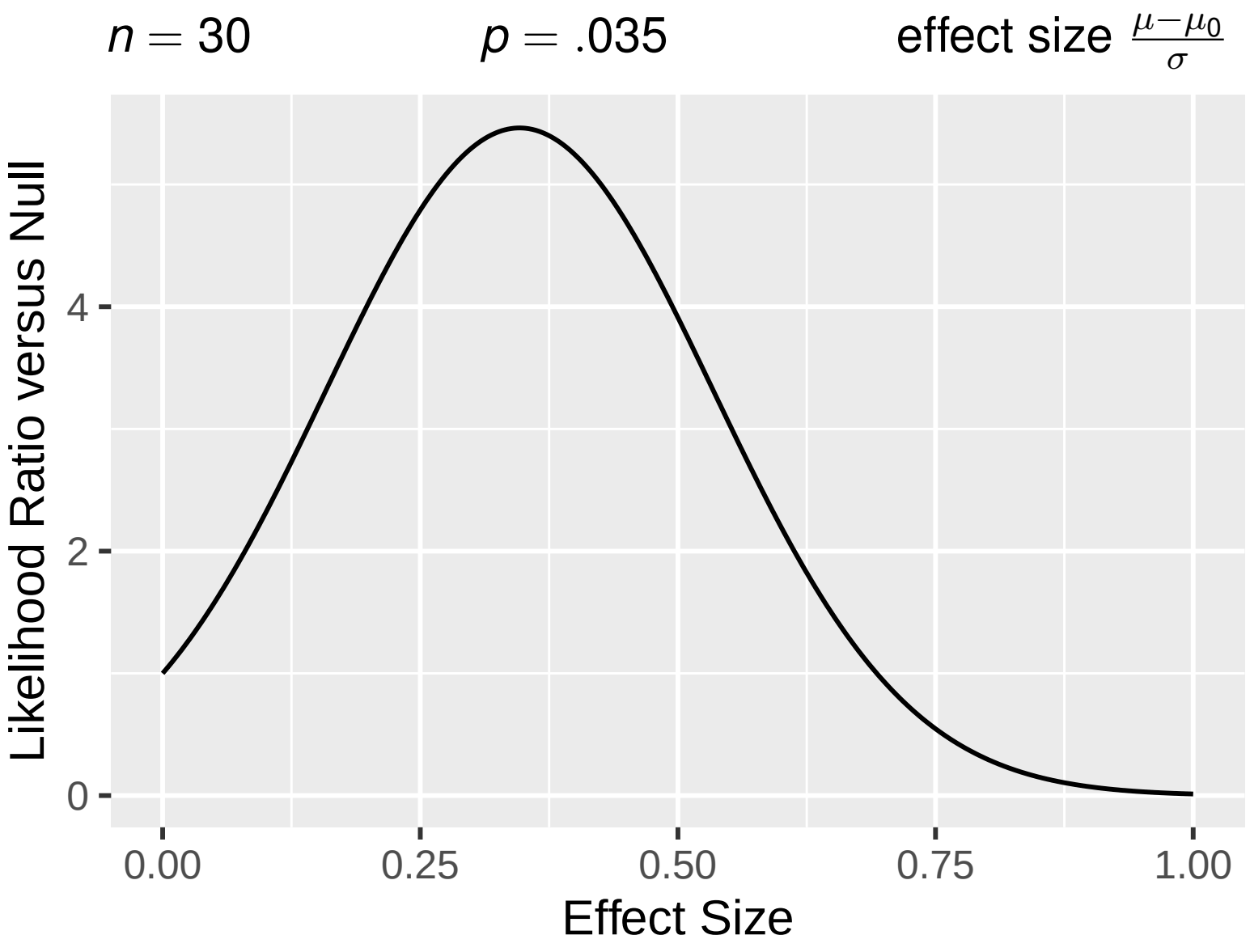


Relative Likelihood

$$\frac{\text{Prob}(P \approx 0.035 \mid \text{alternate})}{\text{Prob}(P \approx 0.035 \mid H_0 \text{ is true})} = 0.55$$

$P = 0.035$ is much more likely to occur if the null is true than if this alternate is true.

Cherry-picking?



Neyman-Pearson Hypothesis Tests

Is a rule that makes a decision to "reject" H_0 or not such that it rejects H_0 falsely with probability α (often .05).

The NP test does not mention P or the word evidence.

F vs. NP

Fisher: Results in a *P*-value which is a measure of strength of evidence against H_0 in these data.

Neyman-Pearson: Results in a decision made by a process that has an error-rate of α (in the long run).

The Mashup and Confusion

Rule: Reject H_0 if $P < \alpha$.

Throws away the value of P and only notices whether $P < \alpha$. ($P = 0.035$ is the same as $P = 0.00001$.)

P is NOT the probability of a Type-I error and has nothing to do with α!

Misinterpreting the definition

A low *p*-value means:

These or more extreme data are unlikely given that H_0 is true.

NOT

H_0 is unlikely to be true given these data.

ASA Statement on *P*-Values, Principle 2

Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133.

> *P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.*

ASA Statement on *P*-values, Principle 3

Scientific conclusions and business or policy decisions should not be based only on whether a P-value passes a specific threshold.

There is very little difference between $P = .051$ and $P = .049$ in what we can conclude about the null hypothesis.

The difference between statistically significant and insignificant is insignificant (Gelman and Loken).

$P = 0.05$ is an obvious superstition. Nothing magical happens at the boundary. (Richard McElreath)

ASA Statement on *P*-values, Principle 6

By itself, a P–value does not provide a good measure of evidence regarding a model or hypothesis.

A *P*-value near 0.05 taken by itself offers only weak evidence against the null hypothesis (Johnson)

Jaynes

Our job is not to follow blindly a rule which would prove correct 95 percent of the time in the long run; there are an infinite number of radically different rules, all with this property. Our job is to draw the conclusions that are most likely to be right in the specific case at hand.