# Interpreting $p$-values

Jarad Niemi

Iowa State University

August 21, 2019

What is the probability the null hypothesis is true when you see a $p$-value equal to 0.04?

What is the probability the alternative hypothesis is true when you see a $p$-value equal to 0.04?

# Outline

- Bayesians vs Frequentists - not the point of today's talk
- ASA Statement on $p$-values
- STAT 226 hypothesis testing recipe
- Hypothesis testing false dichotomy
- Interpreting a $p$-value through Bayes rule

# Bayesianism and Frequentism

Fundamental difference (IMO):

- Frequentists interpret/define probability as the long-run relative frequency of an event occurring in a series of attempts, i.e.

$$P(A) = \lim_{n \to \infty} \frac{I(A_n)}{n}$$

  where $I(A_n)$ is the indicator that event $A$ occurs in the $n$th attempt.
- Bayesians interpret/define probability as a personal statement about the degree of belief with larger numbers indicating a higher personal belief in an event.

https://stats.stackexchange.com/questions/230097/think-like-a-bayesian-check-like-a-frequentist-what-does-that-mean:

> *Think like a Bayesian, check like a frequentist.*

# ASA Statement on *p*-values

https://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108:

1. *p*-values can indicate how incompatible the data are with a specified statistical model.
2. *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.
3. Scientific conclusions and business or policy decisions should **NOT** be based only on whether a *p*-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A *p*-value, or statistical significance, does **NOT** measure the size of an effect or the importance of a result.
6. By itself, a *p*-value does **NOT** provide a good measure of evidence regarding a model or hypothesis.

Bold-face and capitalization have been added for emphasis; the original article bold-faced these sentences.

# STAT 226 Recipe

Simplified for $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with $H_0 : \mu = \mu_0$:

1. Determine $\mu_0$.
2. Obtain
$$t = \frac{\overline{y} - \mu_0}{s/\sqrt{n}}.$$
3. Find the $p$-value (from JMP).
4. Decision with conclusion
   - If $p$-value is small enough, **reject null hypothesis**.

     Conclude that the data most likely came from a population that has a mean different than $\mu_0$.
   - If $p$-value is not small enough, **fail to reject null hypothesis**.

     The data lack sufficient evidence to conclude that the population mean is different than $\mu_0$.

# False dichotomy

Consider the hypothesis test:

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_A : \mu \neq \mu_0.$$

which implies the scientific question "is the population mean $\mu_0$ or not?"

The false dichotomy is that the only two possibilities are

$$H_0 : Y_i \overset{ind}{\sim} N(m_0, \sigma^2) \qquad \text{versus} \qquad H_A : Y_i \overset{ind}{\sim} N(\mu, \sigma^2), \mu \neq \mu_0.$$

In reality, all model assumptions are wrong including

- independence,
- normality,
- constant variance, and
- mean is $\mu_0$.

We need to evaluate these assumptions before we conclude $\mu \neq m_0$ for the population of interest.

# Interpreting a $p$-value when assumptions are true

Suppose it is true that $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ and we obtain a $p$-value below our pre-determined threshold, e.g. $p = 0.04 < 0.05$. We can use Bayes rule to interpret this $p$-value:

$$P(H_0|p) = \frac{P(p|H_0)P(H_0)}{P(p|H_0)P(H_0) + P(p|H_A)P(H_A)}.$$

- $P(H_0) = 1 - P(H_A)$ is the relative frequency of null hypotheses that are true in the experiments that you conduct.
- $P(p|H_0)$ is the distribution of $p$-values when $H_0$ is true.
- $P(p|H_A)$ is the distribution of $p$-values when $H_A$ is true.

```
shiny::runGitHub('jarad/pvalue')
```

# Relationship to ASA Statement on *p*-values

1. *p*-values can indicate how incompatible the data are with a specified statistical model.
2. *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.
3. Scientific conclusions and business or policy decisions should **NOT** be based only on whether a *p*-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A *p*-value, or statistical significance, does **NOT** measure the size of an effect or the importance of a result.
6. By itself, a *p*-value does **NOT** provide a good measure of evidence regarding a model or hypothesis.

# Main questions

What is the probability the null hypothesis is true when you see a $p$-value equal to 0.04?

What is the probability the alternative hypothesis is true when you see a $p$-value equal to 0.04?

# Summary

Summary

- As typically presented, hypothesis testing presents a false dichotomy of the only two possibilities being that the null model or the alternative model are correct.
- The probability the null hypothesis is true conditional on a particular $p$-value depends on a few unknowable values.

This slides are available

- `https://github.com/jarad/pvalue2019`
- `http://www.jarad.me/research/presentations.html`