M6S2 - P-values

Professor Jarad Niemi

STAT 226 - Iowa State University

October 24, 2018

Outline

- Review of statistical hypotheses
  - Null vs alternative
  - One-sided vs two-sided
- Pvalues
  - test statistic
  - as or more extreme
  - interpretation

Statistical hypotheses

Most statistical hypotheses are statements about a population parameters.

For example, for a population mean μ, we could have the following null hypothesis with a two-sided alternative hypothesis:

$$H_0 : \mu = 0 \quad \text{versus} \quad H_a : \mu \neq 0$$

Or we could have the following null hypothesis with a one-sided alternative

$$H_0 : \mu = 98.6 \quad \text{versus} \quad H_a : \mu > 98.6$$

or, equivalently

$$H_0 : \mu lle 98.6 \quad \text{versus} \quad H_a : \mu > 98.6$$

P-values

Definition

A test statistic is a summary statistic that you use to make a statement about a hypothesis. A p-value is the (frequency) probability of obtaining a test statistic as or more extreme than you observed if the null hypothesis (model) is true.

We will discuss the following phrases one at a time

- if the null hypothesis (model) is true,
- test statistic,
- as or more extreme than you observed, and
- (frequency) probability.

Null hypothesis (model)

Recall that we have a null hypothesis, e.g.

$$H_0 : \mu = m_0$$

for some known value m_0, e.g. 0. But we also have statistical assumptions, e.g.

$$X_i \stackrel{iid}{\sim} N(\mu, \sigma^2).$$

Thus, the statement if the null hypothesis (model) is true means that we assume

$$X_i \stackrel{iid}{\sim} N(m_0, \sigma^2).$$

If the null hypothesis is true, we have

$$\frac{\overline{X} - m_0}{S/\sqrt{n}} \sim t_{n-1}$$

where $\overline{X}$ is the sample mean and S is the sample standard deviation.

ACT scores example

The mean composite score on the ACT among the students at Iowa State University is 24. We wish to know whether the average composite ACT score for business majors is different from the average for the University. We sample 100 business majors and calculate an average score of 26 with a standard deviation of 4.

Let X_i be the composite ACT score for student i who is a business major at Iowa State University with $E[X_i] = \mu$.

What is the null hypothesis? The null hypothesis is

$$H_0 : \mu = 24.$$

What is the null hypothesis model?

$$X_i \stackrel{iid}{\sim} N(24, \sigma^2).$$

Test statistic

Let $X_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$. The following are all summary statistics:

- sample mean ($\overline{X}$),
- sample median (Q2),
- sample standard deviation (S),
- sample variance (S^2),
- min, max, range, Q1, Q3, interquartile range, etc.

The test statistic ... you observed is just the actual value you calculate from your sample, e.g. the observed sample mean ($\overline{x}$), the observed sample standard deviation (s), etc.

We will be primarily interested in the t-statistic:

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}.$$

ACT scores example

The mean composite score on the ACT among the students at Iowa State University is 24. We wish to know whether the average composite ACT score for business majors is different from the average for the University. We sample 100 business majors and calculate an average score of 26 with a standard deviation of 4.

What is the observed sample mean?

$$\overline{x} = 26$$

What is the observed sample standard deviation?

$$s = 4$$

What is the t-statistic when the null hypothesis is true?

$$t = \frac{26 - 24}{4/\sqrt{100}} = 5$$

As or more extreme

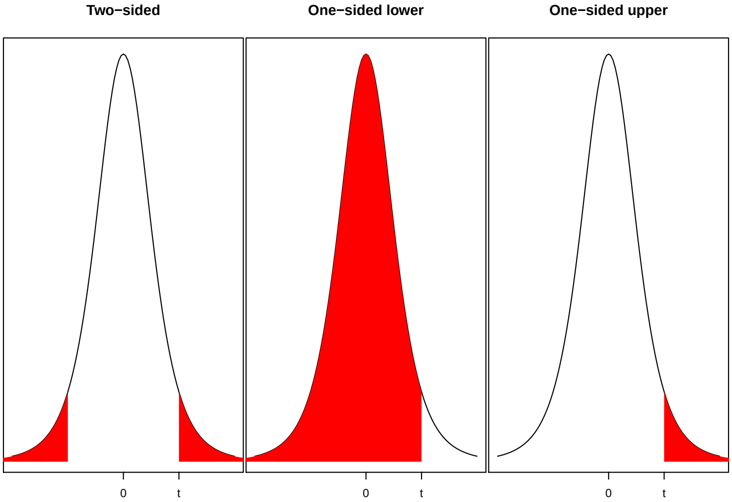
The sampling distribution of the t-statistic is a t-distribution with $n-1$ degrees of freedom where n is the sample size, i.e.

$$T_{n-1} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

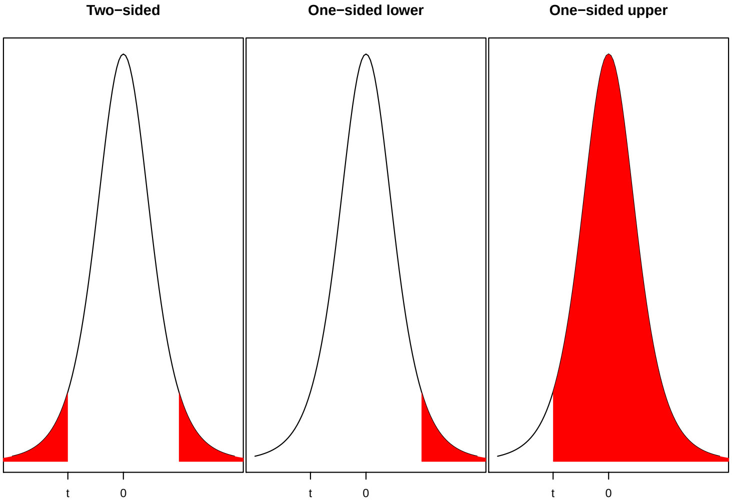
If the null hypothesis is true, replace μ with m_0. The as or more extreme regions depend on the alternative hypothesis. When the null hypothesis is true and for an observed t-statistic t and a random variable T_{n-1} that has a t-distribution with $n-1$ degrees of freedom, the as or more extreme regions are

- $H_a : \mu \neq m_0$ implies the region $T_{n-1} < -|t|$ or $T_{n-1} > |t|$,
- $H_a : \mu < m_0$ implies the region $T_{n-1} < t$, and
- $H_a : \mu > m_0$ implies the region $T_{n-1} > t$.

Graphical depiction of as or more extreme regions



Graphical depiction of as or more extreme regions



(Frequency) Probability

The probability of being as or more extreme than you observed is just the areas depicted on the previous slides. In particular,

- $H_a : \mu \neq m_0$ implies the probability

$$\begin{aligned} P(T_{n-1} < -|t| \text{ or } T_{n-1} > |t|) &= P(T_{n-1} < -|t|) + P(T_{n-1} > |t|)) \\ &= 2P(T_{n-1} > |t|) \end{aligned}$$

- $H_a : \mu < m_0$ implies the probability

$$P(T_{n-1} < t) = P(T_{n-1} > -t)$$

- $H_a : \mu > m_0$ implies the probability

$$P(T_{n-1} > t).$$

P-values (based on the population mean)

Definition

A p-value is the (frequency) probability of obtaining a test statistic as or more extreme than you observed if the null hypothesis (model) is true.

So for the null hypothesis $H_0 : \mu = m_0$, calculate

$$t = \frac{\overline{x} - m_0}{s/\sqrt{n}}$$

and find the appropriate probability:

- $H_a : \mu \neq m_0$ implies p-value $= 2P(T_{n-1} > \ |t|)$,
- $H_a : \mu < m_0$ implies p-value $= \ P(T_{n-1} > -t)$, and
- $H_a : \mu > m_0$ implies p-value $= \ P(T_{n-1} > \ \ t)$.

ACT scores example

The mean composite score on the ACT among the students at Iowa State University is 24. We wish to know whether the average composite ACT score for business majors is different from the average for the University. We sample 100 business majors and calculate an average score of 26 with a standard deviation of 4.

Let X_i be the composite ACT score for student i who is a business major at Iowa State University. Assume $X_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$.

- Null hypothesis $H_0 : \mu = 24$
- Alternative hypothesis $H_a : \mu \neq 24$
- t-statistic:

$$t = \frac{26 - 24}{4/\sqrt{100}} = 5$$

- p-value:

$$\begin{array}{rll} 2P(T_{n-1} > |t|) & = 2P(T_{99} > 5) & \\ & < 2P(T_{80} > 5) & < 2P(T_{80} > 3.416) \\ & = 2 \cdot 0.0005 & = 0.001 \end{array}$$

Interpretation

In 2016, the American Statistical Association published Statement on p-values that states the following principles:

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.

Interpretation (cont.)

The null hypothesis model is

$$X_i \stackrel{iid}{\sim} N(m_0, \sigma^2)$$

for some specificed value m_0.

P-values can indicate how incompatible the data are with [the null hypothesis] model.

The smaller the p-value the larger the incompatibility of the data with the null hypothesis model.

Thus, a small p-value indicates the null hypothesis model is likely not correct. But there are many assumptions in this model that may be wrong, e.g.

- independence,
- identically distributed,
- normality,
- mean is m_0, and
- constant variance.

The p-value doesn't tell us which one is wrong.