Tests of Significance

Stephen B. Holt, Ph.D.

Rockefeller College of Public Affairs & Policy
University at Albany State University of New York

April 4, 2022

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
    - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
    - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.
    - Pearson's "r" coefficient - Measure of covariance; describes the strength and direction of the relationship between two variables.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
    - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.
    - Pearson's "r" coefficient - Measure of covariance; describes the strength and direction of the relationship between two variables.
    - Tests of significance

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
    - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.
    - Pearson's "r" coefficient - Measure of covariance; describes the strength and direction of the relationship between two variables.
    - Tests of significance
5. Organize and report results.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
    - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.
    - Pearson's "r" coefficient - Measure of covariance; describes the strength and direction of the relationship between two variables.
    - Tests of significance
5. Organize and report results.
    - Pie and bar graphs - Depict the distribution of a categorical variable.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
    - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.
    - Pearson's "r" coefficient - Measure of covariance; describes the strength and direction of the relationship between two variables.
    - Tests of significance
5. Organize and report results.
    - Pie and bar graphs - Depict the distribution of a categorical variable.
    - Histogram - Depict the distribution of a quantitative variable.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
   - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
   - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.
   - Pearson's "r" coefficient - Measure of covariance; describes the strength and direction of the relationship between two variables.
   - Tests of significance
5. Organize and report results.
   - Pie and bar graphs - Depict the distribution of a categorical variable.
   - Histogram - Depict the distribution of a quantitative variable.
   - Scatterplot - Depict the relationship between two quantitative variables.

Basic Process

Most policy research involves deceptively simple steps:

1. Define the question you would like answered.
2. State hypotheses about the answer to the question.
3. Collect data that can answer the question.
4. Calculate measures to test hypotheses put forward about the relationship of interest.
    - Mean, Median - Measures of central tendency; describes value of X or Y in a typical case.
    - Quartiles, Standard deviation - Measures of spread; describes range of values in sample/population and measures deviations from the typical case.
    - Pearson's "r" coefficient - Measure of covariance; describes the strength and direction of the relationship between two variables.
    - Tests of significance
5. Organize and report results.
    - Pie and bar graphs - Depict the distribution of a categorical variable.
    - Histogram - Depict the distribution of a quantitative variable.
    - Scatterplot - Depict the relationship between two quantitative variables.
    - Two-way table - Joint distribution of two categorical variables.

Reasoning of Significance Tests

- We have seen that the properties of the sampling distribution of $\overline{x}$ help us estimate a range of likely values for population mean μ
  - Centered on μ
  - Normal distribution with a narrower measure of spread than the population
- Example: You are in charge of ensuring safe streets. You randomly sample speeds of drivers on 4 parts of a main avenue.
- The average speed in your sample was 48 mph. Obviously, we cannot expect every section of the avenue to have the same travel speeds. Thus,
  - Is the somewhat higher speed in your sample due to chance variation?
  - Is it evidence that the city should consider more aggressive enforcement or changes to the streetscape?

Stating Hypotheses

A test of statistical significance tests a specific hypothesis using sample data to decide on the validity of the hypothesis.

In statistics, a hypothesis is an assumption or a theory about the characteristics of one of more variables in one or more populations.

Example: What you want to know: Does the street need more attention for safety reasons?

That same question reframed statistically: Is the population mean μ for the distribution of speeds traveled on the road equal to 35 mph (i.e., the speed limit)?

Stating Hypotheses

The null hypothesis is a very specific statement about a parameter of the population(s). It is labeled H_0.

The alternative hypothesis is a more general statement about a parameter of the population(s) that is exclusive of the null. It is labeled H_a.

Example: Travel speeds on main avenue:

H_0: $\mu = 35mph$ (μ is the average speed of travelers on the road) H_a: $\mu \neq 35mph$ (μ is either larger or smaller)

One-sided and Two-sided Tests

- A two-tail or two-sided test of the population mean has these null and alternative hypotheses:
  - $H_0 : \mu =$ [a specific number] $H_a : \mu \neq$ [a specific number]
- A one-tail or one-sided test of a population mean has these null and alternative hypotheses:
  - $H_0 : \mu =$ [a specific number] $H_a : \mu <$ [a specific number]
  - $H_0 : \mu =$ [a specific number] $H_a : \mu >$ [a specific number]

The FDA tests whether a generic drug has an absorption extent similar to the known absorption extent of the brand-name drug it is copying. Higher or lower absorption would both be problematic, thus we test:
$H_0 : \mu_{generic} = \mu_{brand}$ $H_a : \mu_{generic} \neq \mu_{brand}$ two-sided

How to Choose?

What determines the choice of a one-sided versus a two-sided test is what we know about the problem before we perform a test of statistical significance.
Example: A health advocacy group tests whether the mean nicotine content of a brand of cigarettes is greater than the advertised value of 1.4 mg.
Here, the health advocacy group suspects that cigarette manufacturers sell cigarettes with a nicotine content higher than what they advertise in order to better addict consumers to their products and maintain revenues. Thus, this is a one-sided test: $H_0 : \mu = 1.4\text{mg}$ $H_a : \mu > 1.4\text{mg}$
It is important to make that choice before performing the test or else you could make a choice of "convenience" or fall into circular logic.
In practice, **we want to exercise caution** - a two-sided t-test will thus be preferred in most instances.

The P-Value

The speed of drivers in your city has a known standard deviation of 10 mph.
H_0: $\mu = 35 mph$ versus H_a: $\mu \neq 35 mph$
Tests of statistical significance quantify the chance of obtaining a particular random sample result if the null hypothesis were true. This quantity is the **P-value**.
This is a way of assessing the "believability" of the null hypothesis, given the evidence provided by a random sample.

Interpreting The P-Value

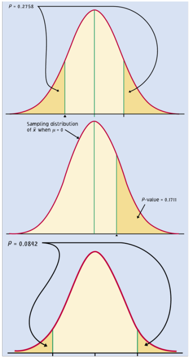
With a small p-value we reject H_0. The true property of the population is significantly different from what was stated in H_0.
Thus, small P-values are strong evidence AGAINST H_0
But how small is small...?

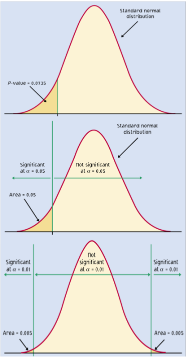
Interpreting The P-Value

P = 0.2758

P = 0.1711

P = 0.0892





P = 0.0735

P = 0.05

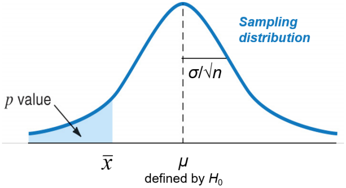
P = 0.01

When the shaded area becomes very small, the probability of drawing such a sample at random gets very slim. Oftentimes, a P-value of 0.05 or less is considered significant: The phenomenon observed is unlikely to be entirely due to chance event from the random sampling.

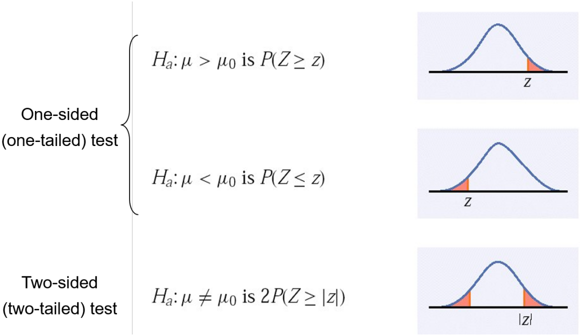
Tests for a Population Mean

The p-value is the area under the sampling distribution for values at least as extreme, in the direction of H_a, as that of our random sample.
Again, we first calculate a z-value and then use a z-table:

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$



P-value in one-sided and two-sided tests



To calculate the P-value for a two-sided test, use the symmetry of the normal curve. Find the P-value for a one-sided test and double it.

Does the street need attention for speeding?

- H_0: $\mu = 35mph$ versus H_a: $\mu \neq 35mph$
- What is the probability of drawing a random sample such as yours if H_0 is true?
  $\overline{x} = 48mph$ $\sigma = 10mph$ $n = 4$
  $$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \frac{48 - 35}{\frac{10}{\sqrt{4}}} \rightarrow 2.4$$
  From a z-table, the area under the standard normal curve to the left of z is 0.9918.
  To the right, this would be 1 - 0.9918 or 0.0082.
  For a two-sided test, we would multiply by 2 (2×0.0082) for a p-value of 0.0164.
  The probability of getting a random sample average this far above μ is so low that we can safely reject H_0.
  We would conclude that the street does need some safety attention.

Steps for Tests of Significance

1. State the null hypotheses H_0 and the alternative hypothesis H_a.
2. Calculate value of the test statistic.
3. Determine the P-value for the observed data.
4. State a conclusion.

The significance level: α

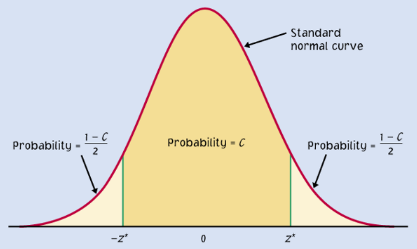
The significance level, α, is the largest P-value tolerated for rejecting a true null hypothesis (how much evidence against H_0 we require). This value is decided arbitrarily before conducting the test.

- If the P-value is equal to or less than α ($P \leq \alpha$), then we reject H_0.
- If the P-value is greater than α ($P > \alpha$), then we fail to reject H_0.

Example: The speed sample p-value was 0.0164. If α had been set to 1%, we would fail to reject the null and the p-value would be insignificant. If α had been set to 5%, we would reject the null and the p-value would be significant.

Confidence intervals and Inference

Because a two-sided test is symmetrical, you can also use a confidence interval to test a two-sided hypothesis. In a two-sided test, C = 1 - α.



Example: $\sigma = 10$ mph: H_0: $\mu = 35mph$ versus H_a: $\mu \neq 35mph$
Sample average 48 mph. 95% CI for μ = 48 mph $\pm$
$1.96 \times \frac{10}{\sqrt{4}} \rightarrow 48mph = \pm 9.8mph$
35 mph is not in the 95% CI (38.2 to 57.8 mph). Thus, we reject H_0.