



MAKING SENSE OF DATA

Intermediate series

HYPOTHESIS TESTS

Copyright © 2012 by City of Bradford MDC

Sampling distributions
Confidence intervals
Hypothesis tests
The t-distribution

Hypothesis testing is a statistician's way of trying to confirm or deny a claim about a population using data from a sample. This topic gives you the big picture of hypothesis testing as well the details for hypothesis tests for one or two means or proportions. Possible errors are examined that can occur in the process.

Doing a hypothesis test

A *hypothesis test* is a statistical procedure that is designed to test a claim. Typically, the claim is being made about a population parameter (one number that characterises the entire population). Because parameters tend to be unknown quantities, there may be several claims about what their values might be. For example, the claim that 25% (or 0.25) of all women delay having children for career reasons is a claim about the proportion (the *parameter*) of all women (the *population*) who delay having children.

Identifying what you're testing

More specifically, the delay having children claim is that the parameter, the population proportion (p), is equal to 0.25. (This claim is called the *null hypothesis*.) If you wish to test this claim, you're questioning the claim and have a hypothesis of your own (called the *alternative hypothesis*). You may hypothesise, for example, that the actual proportion of women who delay having children is lower than 0.25, based on your observations. Or you may hypothesise that the proportion may be higher than 0.25.

In addition to testing hypotheses about categorical variables, you can also test hypotheses about numerical variables, such as the average commuting time for people working in London or their average household income. In these cases, the parameter of interest is the population average, or mean, (denoted μ). Again, the claim is that this parameter is equal to a certain value, versus some alternative.

Setting up the hypotheses

Every hypothesis test contains two hypotheses. The first hypothesis is called the *null hypothesis*, denoted H_0 . The null hypothesis always states that the population parameter is *equal* to the claimed value. In our example, with the claim that the proportion of women who delay having children is 0.25, the statistical shorthand for the null hypothesis in this case would be as follows: $H_0: \mu = 0.25$.

What's the alternative?

Before conducting a hypothesis test, you need two possible hypotheses — the null hypothesis being one of them. If the null hypothesis is found not to be true, what's your alternative going to be? Actually, three possibilities exist for the second (or alternative) hypothesis, denoted H_1 .

- The population parameter is *not equal* to the claimed value ($H_1: \mu \neq 0.25$)
- The population parameter is *greater than* the claimed value ($H_1: \mu > 0.25$)
- The population parameter is *less than* the claimed value ($H_1: \mu < 0.25$)

Which alternative hypothesis you choose in setting up your hypothesis test depends on what you're interested in concluding, should you have enough evidence to refute the null hypothesis (the claim). For example, if you want to test whether the proportion of women who delay having children is larger than 0.25, you use the greater-than alternative. Your hypotheses for that test would be $H_0: \mu = 0.25$ versus $H_1: \mu > 0.25$.

Knowing which hypothesis is which

How do you know which hypothesis to put in H_0 and which one to put in H_1 ? Typically, the null hypothesis says that nothing new is happening; the previous result is the same now as it was before. In general you set up H_0 and H_1 so that you believe H_0 is true unless the evidence (your data and statistics) show otherwise. And in that case, where you have sufficient evidence you reject H_0 in favor of H_1 . The burden of proof is on the researcher to show sufficient evidence against H_0 before it is rejected.

If H_0 is rejected in favour of H_1 , the researcher can say they found a *statistically significant* result.

Finding sample statistics

The null hypothesis makes a statement about what the population parameter is. You need a measure of how much your results can be expected to change if you took a different sample. The data you collect or source measure that variable of interest, and the statistics you calculate will include the *sample statistic* that most closely estimates the population parameter. For example, if you're testing a claim about the proportion of women who delay having children, you need to calculate the proportion of these women in your sample.

Standardizing the evidence: the test statistic

To interpret your sample statistic, you need to convert it from original units to a *standard score*. This is the number of standard errors that a statistic lies above or below the mean.

When finding a standard score (z) for a sample mean or proportion, you take your statistic, subtract the mean, and divide the result by the standard error. In the case of hypothesis tests, you use the value in H_0 as the mean. (That's because we assume H_0 is true, unless you have enough evidence against it) This standardized version of your statistic is called a *test statistic*, and it's the main component of a hypothesis test.

The procedure for converting a statistic to a test statistic (z-score):

1. Take your statistic minus the claimed value (given by H_0)
2. Divide by the standard error of the statistic

Your test statistic represents the distance between your actual sample results and the claimed population value, in terms of number of standard errors. If you see that the distance between the claim and the sample statistic is small (in terms of standard errors), your sample isn't far from the claim and your data supports H_0 . However, if the distance is large, your data are showing less evidence to support H_0 .

Weighing the evidence and making decisions: p -values

To test whether a claim is true, you're looking at the test statistic taken from your sample, and seeing whether it supports the claim. To determine this you look at where the test statistic ends up on its corresponding *sampling distribution*. In the case of means or proportions (if certain conditions are met) you use the standard normal (Z) distribution.

If your test statistic is out in the tails of the standard normal distribution, far from 0, it means the results of this sample do not verify the claim, hence we reject H_0 . If the null hypothesis is true, most (about 95%) of the samples will result in test statistics that lie roughly within 2 standard errors of the claim. If H_1 is the *not-equal-to* alternative, any test statistic outside this range will result in H_0 being rejected (see Figure 1)

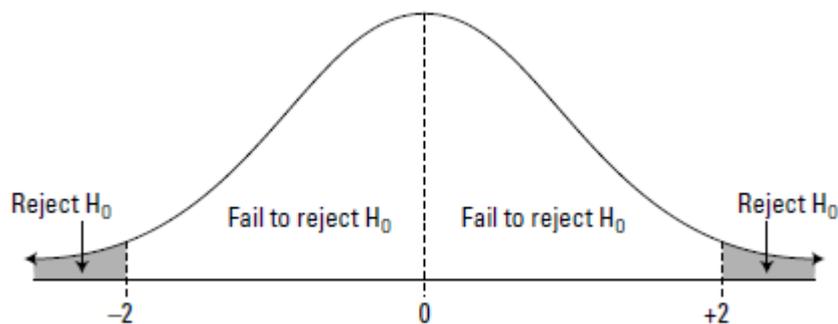


Figure 1: Test statistics and your decision.



If your test statistic is close to 0, you can't reject the claim shown in H_0 . However, this does not mean you accept the claim as truth either. This is because the test statistic is the evidence, and either there is enough evidence to reject H_0 or there isn't. In a hypothesis test we either say "reject H_0 " or "fail to reject H_0 " — we never say "accept H_0 ".

Finding the p-value

To be more specific about your conclusion, you can determine exactly how far out on the standard normal distribution the test statistic falls. In the case of means or proportions, you do this by looking up the test statistic on the standard normal distribution (Z -distribution, see “The normal distribution” topic) and finding the probability of being at that value or beyond it (in the same direction). This p -value measures how likely it was that you would have got your sample results if the null hypothesis were true. The farther out your test statistic is on the tails of the standard normal distribution, the smaller the p -value will be, and the more evidence you have against the null hypothesis being true.

To find the p -value for your test statistic:

1. **Look up the location of your test statistic on the standard normal distribution (see “The Z-table” section)**
2. **Find the percentage chance of being at or beyond that value in the same direction:**
 - a. If H_1 contains a less-than alternative (left tail), find the probability from “The Z-table” that corresponds to your test statistic
 - b. If H_1 contains a greater-than alternative (right tail), find the probability from “The Z-table” that corresponds to your test statistic, and then take 1 minus that value
3. **Double this probability if (and only if) H_1 is the not-equal-to alternative**
This accounts for both the less-than and the greater-than possibilities
4. **Change the probability to a percentage by multiplying by 100**

Interpreting a p-value

To make a proper decision about whether or not to reject H_0 , you determine your cut-off probability for your p -value before doing a hypothesis test; this cut-off is called an alpha level (α). Typical values for α are 0.05 or 0.01. Here is how to interpret your results for any given alpha level:

- ✓ If the p -value is greater than or equal to α , you fail to reject H_0
- ✓ If the p -value is less than α , reject H_0
- ✓ p -values on the borderline (very close to α) are treated as marginal results



When you hear about a result that has been found to be statistically significant, ask for the p -value and make your own decision.

General steps for a hypothesis test

Here is a *summary of the steps* involved in doing a hypothesis test.

1. Set up the null and alternative hypotheses: H_0 and H_1
2. Take a random sample of individuals from the population and calculate the sample statistics (means and standard deviations)
3. Convert the sample statistic to a test statistic by changing it to a standard score (all formulas for test statistics are provided in the next section)
4. Find the p -value for your test statistic
5. Examine your p -value and make your decision

Calculating the test statistic

Testing one population mean

This test is used when the variable is numerical and only one population or group is being studied. The null hypothesis is $H_0: \mu = \mu_0$.

The formula for the test statistic for *one population mean* is

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

where

Z = the z-score on the standard normal distribution

\bar{x} = the sample mean

μ_0 = the claimed value of the population mean

σ / \sqrt{n} = the standard error

Testing one population proportion

This test is used when the variable is categorical and only one population is being studied. The test is looking at the proportion (p) of individuals in the population who have a certain characteristic — for example, the proportion of people who wear glasses. The null hypothesis is $H_0: p = p_0$.

The formula for the test statistic for a *single proportion* is

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where

\hat{p} = the sample proportion

p_0 = the claimed number for the population proportion

$\sqrt{\frac{p_0(1-p_0)}{n}}$ = the standard error

Comparing two population means

This test is used when the variable is numerical and two populations or groups are being compared. Two separate random samples need to be selected, one from each population, in order to collect the data needed for this test. The null hypothesis is that the two population means are the same; in other words, that their difference is equal to 0. The null hypothesis is $H_0: \mu_x - \mu_y = 0$, where μ_x is the mean of the first population, and μ_y is the mean of the second population.

The test statistic *comparing two means* is

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}}$$

where

\bar{x} = the sample mean for x -values
 \bar{y} = the sample mean for y -values

$\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}$ = the standard error



If either of the sample sizes is small (generally less than 30), you use the *t-distribution* with $n_1 + n_2 - 2$ degrees of freedom instead of the standard normal distribution when figuring out the p -value.

Testing the mean difference: paired data

This test is used when the variable is numerical, and the individuals in the sample are either paired up in some way or the same people are used twice (for example, using a pre-test and post-test).

Testing paired data amounts to testing one population mean, where the null hypothesis is that the mean (of the paired differences) is 0, and the alternative hypothesis is that the mean (of the paired differences) is > 0 ; < 0 , or $\neq 0$. The null hypothesis is $H_0: \mu_d = 0$, where μ_d is the population mean of all paired differences. (The d in the subscript reminds you that you're working with the paired differences.)

The formula for the test statistic for *paired differences* is

$$\frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$$

where

\bar{d} = the mean of paired differences

s_d / \sqrt{n} = the standard error

Remember that $\mu_d = 0$ if H_0 is true.

Testing two population proportions

This test is used when the variable is categorical (for example, support/oppose an opinion) and you're interested in the proportion of individuals with a certain characteristic — for example, the proportion of voters. In this case, two populations or groups are being compared (such as the proportion of female versus male voters).

In order to conduct this test, two separate random samples need to be selected, one from each population. The null hypothesis is that the two population proportions are the same; in other words, that their difference is equal to 0. The null hypothesis is $H_0: p_1 - p_2 = 0$, where p_1 is the proportion from the first population, and p_2 is the proportion from the second population.

The formula for the test statistic comparing *two proportions* is

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where

\hat{p} = is the *pooled sample proportion*¹

$\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ = is the standard error

¹ aka the proportion of all individuals from the combined samples that have the characteristic of interest

Errors in hypothesis testing

A *Type I error* is the rejection of a true null hypothesis; whereas a *Type II error* is the failure to reject a false null hypothesis. Figure 2 shows all possible decisions from a hypothesis test.

	Null hypothesis (H_0) is true	Null hypothesis (H_0) is false
Reject null hypothesis	Type I error False positive	Correct outcome True positive
Fail to reject null hypothesis	Correct outcome True negative	Type II error False negative

Figure 2: Truth/falseness of the null hypothesis and test outcomes

A false alarm: Type I error

A Type I error, also known as an error of the first kind, occurs when the null hypothesis (H_0) is true, but is rejected. It is asserting something that is absent, or a “false alarm”. A Type I error may be compared with a so called false positive (a result that indicates that a given condition is present when it actually is not present) in tests where a single condition is tested for. A Type I error happens when we believe a falsehood.

A missed detection: Type II error

A Type II error, also known as an error of the second kind, occurs when the null hypothesis is false, but erroneously fails to be rejected. It is failing to assert what is present, or a “missed detection”. A Type II error may be compared with a so-called false negative (where an actual 'hit' was disregarded by the test and seen as a 'miss') in a test checking for a single condition with a definitive result of true or false. A Type II error happens when we fail to believe a truth.