

We start with a **null hypothesis** H_0 , which we'll assume to be true until we have evidence to the contrary. Exactly what constitutes contrary evidence is determined by our choice of **alternative hypothesis** H_1 . We **reject** H_0 in favor of H_1 if an appropriate test statistic falls in a **critical region**. If the test statistic does not fall in the critical region, then we fail to reject H_0 .

Null hypotheses come in two kinds: **simple** in which the value of the parameter is exactly specified (e.g. $\mu = \mu_0$) and **complex** in which the parameter has a range of possible values (e.g. $\mu \leq \mu_0$). As suggested by the name, simple null hypotheses are easier to work with. A clever choice of alternative hypothesis can make a simple null hypothesis appropriate in many situations that might seem to require a complex null hypothesis. For example, $H_0 : \mu \leq \mu_0$ can be tested against $H_1 : \mu \not\leq \mu_0$ using the simple null hypothesis $H_0 : \mu = \mu_0$ (tested against $H_1 : \mu > \mu_0$).

1. Let x_1, x_2, \dots, x_n be a random sample from a population with mean μ . We will use the sample mean \bar{x} to test the null hypothesis $H_0 : \mu = \mu_0$ against different alternative hypotheses. Describe the critical region for each alternative hypothesis (i.e. which values of \bar{x} would lead us to reject H_0 in favor of H_1).

- $H_1 : \mu \neq \mu_0$.
- $H_1 : \mu > \mu_0$.
- $H_1 : \mu < \mu_0$.

When testing a hypothesis, there are two kinds of mistake:

- **Type I error** is rejecting H_0 when it is true;
- **Type II error** is failing to reject H_0 when it is false.

The probability of a type I error is α (this is the same α we've been dealing with). The probability of a type II error is β . Usually you should choose the largest acceptable value for α since this will minimize β . The critical region is chosen so that the test statistic lands in the critical region with probability α when H_0 is true. It may also be useful to find the **P-value** (or **observed significance level**) of your data: this is the smallest value for α that leads you to reject H_0 with your data.

Example. In class we tested my hypothesis that the mean height of a Gonzaga undergraduate is 68 inches. Formally, we tested $H_0 : \mu = 68$ against $H_1 : \mu \neq 68$. We assumed that our population was normally distributed and that the class constituted a random sample. The sample data gave $\bar{x} = 68.6$ and $s^2 = 7.2$ with a sample size of $n = 20$ (disagreement with your calculations may be the result of transcription errors). Our test statistic was

$$t = \frac{68.6 - 68}{2.6833/\sqrt{20}} \approx 1$$

The P -value of the test is $2P(T \geq 1) = 2(0.165) = 0.33$ (where T has a t -distribution with 19 df; we multiply by 2 because this is a two-tailed test and \bar{x} being large or small constitutes evidence against H_0). Different disciplines have different standards for what P -value is evidence against H_0 , but generally a P -value of 0.1 is considered significant (but maybe not conclusive).

Using a random sample of size n from an infinite population, x_1, x_2, \dots, x_n , we have the following test statistics.

For **tests about the mean** ($H_0 : \mu = \mu_0$) test statistics are:

- $z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ for samples from a population with a known variance σ^2 (all sample sizes if the population is normal, otherwise just for large samples);
- $z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ for large samples;
- $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ for small samples ($n \leq 30$ or as big as your t -table goes) from a normally distributed population.

For **tests about a population proportion** ($H_0 : \theta = \theta_0$) we call use the sample proportion $\hat{\Theta}$ or the sample total $X = n\hat{\Theta}$ and the test statistic is:

- x (X is binomial with parameters n and θ_0 , works best for small samples);
- $z = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} = \frac{x - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$ for large samples (both $n\theta_0 \geq 10$ and $n(1-\theta_0) \geq 10$).

For tests about the variance ($H_0 : \sigma^2 = \sigma_0^2$) the test statistic is

- $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ for samples from a normally distributed population.

For tests about the difference of two means ($H_0 : \mu_1 - \mu_2 = \delta$) the test statistics are:

- $z = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ for samples from populations with a known variances σ_1^2 and σ_2^2 (all sample sizes if the populations are normal, otherwise just for large samples);
- $z = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ for large samples;
- $t = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{s_p^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ for small samples from normally distributed populations with the same variance.

$$\text{Recall } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

For tests about the ratio of two variances ($H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$) the test statistic is

- $\frac{s_1^2}{s_2^2}$ ($\frac{s_1^2}{s_2^2}$ has an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom).

Problems in which hypotheses aren't explicitly stated should be completed in two steps:

- What are your null and alternative hypotheses? (Does your null hypothesis make sense as a default assumption?)
- Use the data to test H_0 against H_1 .

2. Newly purchased tires are supposed to be filled to a pressure of 30 lb/in². Let μ denote the true average pressure. We wish to test the hypothesis $H_0 : \mu = 30$ against the alternate hypothesis $H_a : \mu \neq 30$. Find the P -value of each measurement of the sample mean and sample standard deviation of a random sample of size $n = 100$.

- $\bar{x} = 28.2$, $s = 8$
- $\bar{x} = 28.2$, $s = 4$
- $\bar{x} = 30.6$, $s = 4$

3. The article "Analysis of Reserve and Regular Bottlings: Why Pay for a Difference Only the Critics Claim to Notice?" reported on an experiment to determine if wine tasters could correctly distinguish between reserve and regular versions of a wine. In each trial tasters were given 4 indistinguishable containers of wine, two of which contained the regular version and two of which contained the reserve version of the wine. The taster then selected 3 of the containers, tasted them, and was asked to identify which one of the 3 was different from the other 2. In 855 trials, 346 resulted in correct distinctions. Does this provide compelling evidence that wine tasters can distinguish between regular and reserve wines?

4. The sample average unrestrained compressive strength for 45 specimens of a particular type of brick was 3107 psi, and the sample standard deviation was 188 psi.

- Does the data strongly indicate that the true average unrestrained compressive strength is less than the design value of 3200? Test using $\alpha = 0.001$. What does it mean to use such a small value for α ?
- Is this strong evidence that $\sigma < 200$ psi?

5. Suppose a survey of 948 American college football fans finds that 497 prefer playoffs to the bowl system. Is this strong evidence that a majority of fans prefer playoffs?

6. It is known that roughly $\frac{2}{3}$ of all people have a dominant right foot and $\frac{2}{3}$ have a dominant right eye. Do people also kiss to the right? The article "Human Behavior: Adult Persistence of Head-Turning Asymmetry" reported that in a random sample of 124 kissing couples, 80 of the couples tended to lean more to the right than left. Does this result suggest that more than half of all couples lean right when kissing? Does this result provide evidence against the hypothesis that $\frac{2}{3}$ of all kissing couples lean right?

7. Minor surgery on horses under field conditions requires a reliable short-term anesthetic. The article “A Field Trial of Ketamine Anesthesia in the Horse” reports that for a sample of 73 horses to which ketamine was administered the average lateral recumbency time was 18.86 minutes with a standard deviation of 8.6 minutes.

- a) Does this data suggest that the true average lateral recumbency time is less than 20 minutes?
- b) Does this data suggest that the true average lateral recumbency time is more than 15 minutes?
- c) Does this data suggest that the true variance of lateral recumbency time is more than 64 minutes?

8. A random sample of 10 Black Angus steers is weighed, giving a sample standard deviation of 238 pounds. Test $H_0 : \sigma = 250$ against $H_1 : \sigma \neq 250$.