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 $H_1 : \mu \nleq \mu_0$ using the simple null hypothesis $H_0 : \mu = \mu_0$ (tested against $H_1 : \mu > \mu_0$).

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

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

The probability of a type I error is α; 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.

Using a random sample of size n from an infinite population, $x_1, x_2, \ldots, x_n$, we have the following test statistics.
For **tests about the mean** ($H_0 : \mu = \mu_0$) test statistics are:

- $z = \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ for samples from a population with a known variance σ^2 (all sample sizes if the population is normal, otherwise just for large samples);
- $z = \dfrac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ for large samples;
- $t = \dfrac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ for small samples ($n \leq 30$ or as big as your t-table goes) from a normally distributed population ($n - 1 df$).

For **tests about a population proportion** ($H_0 : \theta = \theta_0$) we can 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);
- $z = \dfrac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} = \dfrac{x - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$ 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

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

For **tests about the difference of two means** ($H_0 : \mu_1 - \mu_2 = \delta_0$) the test statistics are:

- $z = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ for samples from populations with a known variances (all sample sizes if the populations are normal, otherwise just for large samples);
- $z = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ for large samples;
- $t = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ for small samples from normally distributed populations with the same variance ($n_1 + n_2 - 2$ df).

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

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

- $\dfrac{s_1^2}{s_2^2}$ ($\frac{S_1^2}{S_2^2}$ 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 three steps:

a) State your null and alternative hypotheses. (Does your null hypothesis make sense as a default assumption?)

b) Use the data to test H_0 against H_1, testing either at a specified significance level or giving a P-value.

c) State your conclusion clearly.

1. I have heard that the average height of an American man is 70 inches. Test this hypothesis using the height data you contributed (assume that the population is normally distributed and that the class constitutes a random sample). The data give $n = 28$, $\overline{x} = 72.08036$, and $s = 2.191158$. Calculate the P-value of this result (remember to be clear about your hypotheses).

2. 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?

3. 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.

a) 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?

b) Is this strong evidence that $\sigma < 200$ psi?

4. 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.

a) Does this result suggest that more than half of all couples lean right when kissing?

b) Does this result provide evidence against the hypothesis that $\frac{2}{3}$ of all kissing couples lean right?

5. Are sons taller than their fathers? There are two ways to use your heights data to answer the question:

1. Test $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 > 0$ where μ_1 is the mean height of sons and μ_2 is the mean height of fathers;

2. Test $H_0 : \mu = 0$ against $H_1 : \mu > 0$ where μ is the mean difference between the height of a son and the height of his father.

a) Use method 1 to answer the question: n_1, $\overline{x}_1$, and s_1 are given in problem 1; $n_2 = 34$, $\overline{x}_2 = 70.82353$, and $s_2 = 2.633935$.

b) In order to use the pooled estimator to work, we must assume that the variance of the sons' heights is equal to the variance of fathers' heights. Test $H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = 1$ against $H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$ at significance level 0.05. It may help to know that $f_{.025,27,33} = 2.054075$ and $f_{.975,27,33} = 0.4738242$.

c) Use method 2 to answer the question: $n = 27$, $\overline{x} = 1.046296$, $s = 3.008032$.