

So when do we use which hypothesis test, expanded.
(This is so much easier to explain in a classroom. Grr.)

(By the way, this helps with deciding which-confidence-intervals-when, too.)

Alright, here goes. **First, decide how many samples were actually taken for the study.** Be careful with this because sometimes claims are based on old samples that aren't actually part of the current study. For us, the answer to this is either **1 sample or 2 samples**.

Next, decide what kind of data it is.

If the data is "either or" (e.g. "do they prefer peanuts or almonds?") or yes/no (e.g. "did you vote in the last election?") then we are looking at proportions. This type of data is **binary data**, that is either 0 or 1.

The other possibility is that the **data is continuous**. e.g. "how many miles do travel to get to work?", finding someone's pulse, or scores on an exam.

Now we have 4 categories based on those 2 answers.

1 sample, continuous → either z-test for pop. mean, t-test for pop. mean, or nothing fits.
(see below for figuring which of these 3 apply.)

2 sample, continuous → 2 sample z-test for pop. mean if appropriate conditions are met.

1 sample, binary data → 1 sample z-test for proportions if appropriate conditions are met.

2 sample, binary data → 2 sample z-test for proportions if appropriate conditions are met.

Note for proportions, we always use a z-test (with appropriate conditions). This is because binary data is pretty close normal as our sample size gets large.

When to use a z-test:

1. If the sample came from a normal distribution and the population standard deviation is *known*.
2. If the sample came from a normal distribution and the sample size is bigger than 29 and the population standard deviation is *unknown*, then we may substitute the sample standard deviation for the population standard deviation.
3. If the sample is large enough that \bar{X} is close to being normally distributed. For this class, we'll say that $n > 29$ is good enough, although you would want to examine this carefully before making real-world decisions. If population SD is known, we use that, but if it is unknown, we use the sample SD.

When to use a t-test:

1. If the sample came from a normal distribution, the population SD is unknown, and the sample size is less than 30.

Consult a statistician or start doing some research if:

1. \bar{X} isn't almost normally distributed for your given sample size. If you can't answer that, you probably want to check with someone.