Math 122
Introduction to Statistics
ONE MEAN T TEST

CENTRAL LIMIT THEOREM

VARIANT ON CENTRAL LIMIT THEOREM

Recall that one version of the Central Limit Theorem is:

Central Limit Theorem: Suppose x is a random variable with mean μ and standard deviation σ and that $\bar{x}$ is the mean of simple random samples of size n of values of x. If x is normal, or if $n > 30$, then the distribution of

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

is approximately standard normal.

To directly apply the Central Limit Theorem, we need to know σ (which is unlikely). If our population is not normal, then we also need a large sample ($n > 30$). There are times when a large sample is inconvenient (for example, when estimating repair costs of crashing a luxury vehicle). A new type of distribution was discovered by William Gosset (working for Guinness Brewery) to allow better approximation of the sampling distribution of the mean when σ is not known. Thus we have a variation on the Central Limit Theorem:

Central Limit Theorem Variant: Suppose x is a random variable with mean μ and that $\bar{x}$ is the mean of simple random samples of size n of values of x. If x is essentially normal then the distribution of

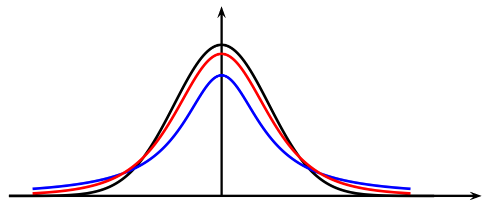
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\bar{x} - \mu}{s}\sqrt{n}$$

is approximately a Student t distribution with $n - 1$ degrees of freedom.

This version still requires some normality in the distribution of x, but that is a light requirement. The theorem works well with distributions that are approximately symmetric and bell shaped.

STUDENT'S t DISTRIBUTION

Student's t distribution is actually a family of distributions which vary by sample size – or by *degree of freedom*. Each t distribution (like a standard normal or z distribution) has mean 0 but a standard deviation greater than 1. The density curve for a t distribution is a bell curve whose central height is determined by its degree of freedom.



Since these distributions have standard deviations larger than 1, the density curves are spread out more than a standard normal distribution. The central peak of each density curve is below the density curve for the standard normal distribution. The larger the number of degrees of freedom, the closer the t distribution is to a standard normal distribution.

Degrees of Freedom: If the mean $\bar{x}$ of a sample of n values is known, and if we arbitrarily specify $n-1$ of the values, then we can always select an n^{th} value to obtain the desired mean. To describe this situation, we say that there are $n-1$ degrees of freedom. We abbreviate degress of freedom as df. For this section, $df = n-1$.

Probabilities: For a random variable x with a t Distribution and df degrees of freedom,

$$P(a < x < b) = tcdf(a, b, df).$$

HYPOTHESIS TEST WITH STUDENT'S t Distribution

We can perfom hypothesis tests for means using a Student t distribution to test claims about means. These claims will all have the form of one of these:

$$\mu = a \text{ or } \mu \leq a \text{ or } \mu \geq a \text{ or } \mu \neq a \text{ or } \mu < a \text{ or } \mu > a.$$

Our Null Hypothesis H_0 and Alternative Hypothesis will always be paired according to:

H_0	H_1
$\mu \leq a$	$\mu > a$
$\mu = a$	$\mu \neq a$
$\mu \geq a$	$\mu < a$

Example: Bob believes that college students do not get enough sleep. One day, he asked 100 of his friends how much sleep they had the night before. His friends slept an average of 5.37 hours that night with a standard deviation of 2.87 hours. Treat this sample as a simple random sample and use this information to test the claim that college students on average get less than 6 hours of sleep each night.

1. **Parameters:** Let x be the number of hours slept each night by a college student. Let μ be the average number of hours of sleep each night by college students. Let $n = 100$. Let $\bar{x}$ be the average hours of sleep of a simple random sample of 100 college students, and let s be the standard deviation of such a sample. Our observed value of $\bar{x}$ is 5.37 hours, and our observed value of s is 2.87 hours.

2. **Symbolic Claim:** The claim is $\mu < 6$.

3. **Opposite of the Claim:** The opposite of the claim is $\mu \geq 6$.

4. H_0 **and** H_1**:** H_0 is $\mu \geq 6$ and H_1 is $\mu < 6$.

5. **Significance Level:** We will default to a significance level of $\alpha = 0.05$.

6. **Distribution and Test Statistic:** Since $n > 30$, we treat the sample mean $\bar{x}$ as having a t distribution with $df = n - 1 = 99$ degrees of freedom and mean $\mu = 6$ (the mean is from H_0). Our test statistic is the t-score of our observed value of x:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{5.37 - 6}{2.87/\sqrt{100}} = -2.1951.$$

7. P**-value:** Since H_1 involves $<$, this is a left-tailed test. Then

$$P = P(t < -2.1951) = tcdf(-9, -2.1951, 99) = 0.0152.$$

8. **Formal Conclusion:** Since $P < \alpha$, we reject H_0 (and support H_1).

9. **Final Conclusion:** Since our claim is the same as H_1, we first rephrase our formal conclusion to refer to H_1. Since we are rejecting H_0, we are supporting H_1. This means that we are supporting our claim. Our conclusion is:

   > There is enough sample evidence to support the claim that college students on average get less than 6 hours of sleep per night.

Interpretation: College students seem to get less than six hours of sleep.