Hypothesis Testing of Single Mean and Single Proportion: Assumptions*

Susan Dean
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

When you perform a **hypothesis test of a single population mean** μ using a **Student's-t distribution** (often called a t-test), there are fundamental assumptions that need to be met in order for the test to work properly. Your data should be a **simple random sample** that comes from a population that is approximately **normally distributed**. You use the sample **standard deviation** to approximate the population standard deviation. (Note that if the sample size is sufficiently large, a t-test will work even if the population is not approximately normally distributed).

When you perform a **hypothesis test of a single population mean** μ using a normal distribution (often called a z-test), you take a simple random sample from the population. The population you are testing is normally distributed or your sample size is sufficiently large. You know the value of the population standard deviation.

When you perform a **hypothesis test of a single population proportion** p, you take a simple random sample from the population. You must meet the conditions for a **binomial distribution** which are there are a certain number n of independent trials, the outcomes of any trial are success or failure, and each trial has the same probability of a success p. The shape of the binomial distribution needs to be similar to the shape of the normal distribution. To ensure this, the quantities np and nq must both be greater than five ($np > 5$ and $nq > 5$). Then the binomial distribution of sample (estimated) proportion can be approximated by the normal distribution with $\mu = p$ and $\sigma = \sqrt{\frac{p \cdot q}{n}}$. Remember that $q = 1 - p$.

Glossary

Definition 1: Binomial Distribution
A discrete random variable (RV) which arises from Bernoulli trials. There are a fixed number, n, of independent trials. "Independent" means that the result of any trial (for example, trial 1) does not affect the results of the following trials, and all trials are conducted under the same conditions. Under these circumstances the binomial RV X is defined as the number of successes in n trials. The notation is: $X \sim B(n, p)$. The mean is $\mu = \text{np}$ and the standard deviation is $\sigma = \sqrt{\text{npq}}$. The probability of exactly x successes in n trials is $P(X = x) = \binom{n}{x} p^x q^{n-x}$.

Definition 2: Normal Distribution
A continuous random variable (RV) with pdf f(x) = $\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)2}/2\sigma^2$, where μ is the mean of the distribution and σ is the standard deviation. Notation: $X \sim N(\mu, \sigma)$. If $\mu = 0$ and $\sigma = 1$, the

*Version 1.16: Jun 12, 2012 8:33 am -0500
†http://creativecommons.org/licenses/by/3.0/

RV is called **the standard normal distribution**.

Definition 3: Standard Deviation
A number that is equal to the square root of the variance and measures how far data values are from their mean. Notation: s for sample standard deviation and σ for population standard deviation.

Definition 4: Student-t Distribution
Investigated and reported by William S. Gossett in 1908 and published under the pseudonym Student. The major characteristics of the random variable (RV) are:

- It is continuous and assumes any real values.
- The pdf is symmetrical about its mean of zero. However, it is more spread out and flatter at the apex than the normal distribution.
- It approaches the standard normal distribution as n gets larger.
- There is a "family" of t distributions: every representative of the family is completely defined by the number of degrees of freedom which is one less than the number of data.