Chapter 23 - Inference about Means

August 24, 2010

In Chapter 19-22, we learned how to do confidence intervals and test hypothesis for proportions. In this chapter we will do the same for means.

The Central Limit Theorem (Chap. 18) gave us the sampling distribution for means as Normal with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

- How do you find the population standard deviation, σ?
- We will use the sample standard deviation, s, which is the estimate for σ to get the Standard Error of $SE(\bar{x}) = \frac{s}{\sqrt{n}}$.

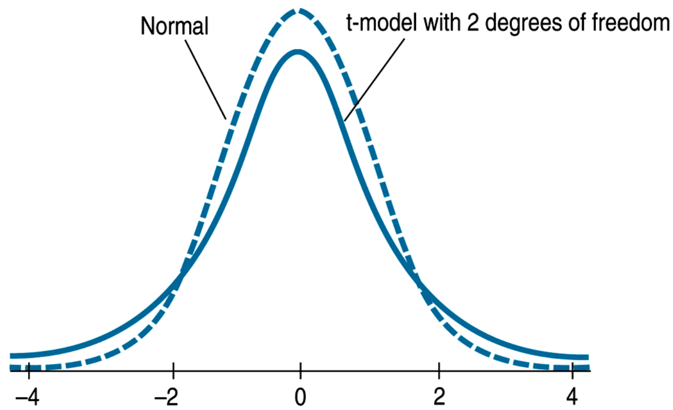
What sort of distribution can be used? The distribution is no longer Normal since the Standard Error introduces extra variation from s.

- William S. Gosset found the sampling model while working at Guinness Brewery in Dublin, Ireland.
- The model Gosset found is commonly referred to as the Student's *t* distribution.
- This model is actually a family of related distributions that depend on a parameter known as degrees of freedom (df).
- A sampling distribution model for means, when the conditions are met, the standardized sample mean

  $$t = \frac{\bar{x} - \mu}{SE(\bar{x})} \quad \text{where} \quad SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

  follows a Student's *t* distribution.
- Correcting for the extra variation causes this model to give us a larger margin of error which will make our intervals wider and our P-values will be larger than from the Normal Model.

- Student's t-models are unimodal, symmetric, and bell shaped like the Normal, but t-models with only a few degrees of freedom have fatter tails than the Normal.



- As the degrees of freedom increase, the t-models look more like the Normal. The t-model with infinite (∞) degrees of freedom is the Normal.

Assumptions and Conditions

- Randomization Condition: The data comes from a random sample or randomized experiment.

- 10% Condition: When sample is drawn without replacement, the sample should be no more than 10% of the population.

- Nearly Normal Condition: The data come from a distribution that is unimodal and symmetric.
    - The smaller the sample size (n<15 or so), the more closely the data should follow a Normal model.
    - For moderate sample sizes (in between 15 and 40 or so), the t works well as long as the data are unimodal and symmetric.
    - For larger sample sizes, the t models are safe to use even if the data are skewed.

One Sample t-Interval for the Mean:

$$\bar{x} \pm t_{n-1}^{*} \frac{s}{\sqrt{n}}$$

One Sample t-Test for the Mean:

We test the hypothesis $\mathrm{H}_0 : \mu = \mu_0$ using the test statistic

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

and the Student's t model with n-1 df.