Math 593 - Worksheet on Hypothesis Testing for the Mean and Proportions

Name: ____________________

We now arrive at the notion of hypothesis testing (HT) for the mean (we will revisit HT's in the context of proportions later in this worksheet). The setting for now is restricted to the case where the mean is assumed normally distributed with known standard deviation σ. In this topic, the goal is to determine (to within a prescribed acceptable probability of being wrong), whether or not the sample mean one obtains supports some claim about the population mean, or does not support the claim.

1. **Definition:** *Hypothesis Testing* is the mathematical process by which a quantitative conjecture about a parameter is assessed for reasonableness.

2. For the hypothesis testing done in this class (and much of it in general), several objects are required:
   (a) An assumption about the underlying probability distribution of the RV whose parameter is being tested.
   (b) A *test statistic,* computed from data (as well as, perhaps, assumed theoretical information)
   (c) A table of values of the underlying probability distribution against which to compare the test statistic

3. **Example:** (Do not yet attempt to answer this question) In an advertisement, a pizza shop claims that its mean delivery time is less than 30 minutes. A random selection of 36 delivery times yields a sample mean of 28.5 minutes and a standard deviation of 3.5 minutes. Does this provide sufficient evidence to support the claim?

   This is exactly the type of question on can answer using a hypothesis test. But beware! a hypothesis test *never* provides a definitive confirmation or rebuttal to the claim being tested. The *only way to determine a parameter is to acquire data on all the observational units in the population.* And *even then* the answer is certain to be definitive only at the time the data was gathered - in particular, large populations of living or otherwise dynamic things may change sufficiently rapidly that one's conclusions may be fleeting.

   Worse yet, if, during the data-gathering process, the population may have changed, the conclusion may not even be definitively valid during that time. Consider the length of time during which the US Census Bureau gathers data; for example March 1 - April 1, 2010 for the upcoming census. One can be certain that the US population has both gained and lost observational units, rendering *logically* impossible any chance of determining any parameter for the US population *at any particular point in time.*

4. We are thus left with only the possibility of probabilistic conclusions about parameters. Recall the confidence interval you calculated for Problem 30 in the previous worksheet. You will have found that the 95% CI for the mean time is $(67.1, 76.7)$. This interval denotes a range of values which has a probability of $.95$ of containing the true mean which is unknown. It would be unreasonable, to within a threshold of 5% error, to make a claim that the true population mean equals μ_0 unless $\mu_0 \in (67.1, 76.7)$. This is the essential idea of hypothesis testing, but the conclusions of HT may be a bit broader. Read on.