Hypothesis testing NAME:

This worksheet will work us through understanding the ideas we will use for hypothesis tests. We will use our knowledge of normal distributions.

We will use the following poll results from Gallup. The poll, which was conducted on March 26-28, 2004, asked 1,000 American adults the following questions. The sample percentages are provided.

Question	Sample percentage
Is the cost of gasoline a crisis for the US?	13% YES
Is the cost of gasoline a major problem for the US?	56% YES
Do the recent price increases reflect a more permanent change in the cost of gasoline?	55% YES

(source: www.gallup.com)

1. Let's concentrate on the first question. Recall, the population percentage is the **true** percent of American adults who feel that the cost of gasoline is a crisis for the US. Our sample tries to estimate this percent.

Imagine many different samples of 1,000 people. Most of the samples will yield a sample percentage (13% in our case) that is near the true population percentage. Some sample results will be further from the truth than others. Some sample results will be less than the population percentage; some will be greater than it. A few samples will yield results that are drastically different than the true population percentage; some much less and some much more. This indicates that sample proportions are normally distributed.

Picture a list of numbers, each being the sample percentage for a possible sample. The mean of these numbers, in theory, is equal to the true population percentage. The standard deviation of this list of numbers is equal to $\sqrt{\frac{p(1-p)}{n}}$ where p is the true population percentage and n is the sample size.

We are going to use this information and our knowledge of normal distributions to explore these samples.

2. Assume the **true percentage of American adults** who feel that gas prices are a crisis for the US is 15%. The sample percentages (for all possible samples) are normally distributed with a mean of 15% and a standard deviation of $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.15(1-.15)}{1000}} = .011 = 1.1\%$.

Answer the following questions.

a.) Draw the appropriate normal curve. Label the mean and plus or minus one, two, and three standard deviations from the mean.

b.) What is the probability that we sample and find 13% or less of our sample feel that gas prices are a crisis for the US? Draw a quick sketch of the normal curve with the mean, and 13% marked. Shade the area in which we are interested. (This tells us the likelihood of obtaining a sample with a sample percentage equal to or less than ours, assuming the population percentage is 15%.)

c.) What is the probability that we sample and find 20% or more of our sample feel that gas prices are a crisis for the US? Draw a quick sketch of the normal curve with the mean, and 20% marked. Shade the area in which we are interested. (The standard score will be off the charts. What does this mean?)

d.) What is the probability of obtaining a sample that differs from the true population percentage by **less** than 3 percentage points? Draw a quick sketch of the normal curve with the mean, and the endpoints of the desired interval marked. Shade the area in which we are interested.

e.) What is the probability of obtaining a sample that differs from the true population percentage by **more** than 3 percentage points? Draw a quick sketch of the normal curve with the mean, and the endpoints of the desired interval marked. Shade the area in which we are interested.

3. Let's say an oil company claims that 75% of the American people feel that gas prices are a major problem or worse. They are trying to convince Congress to subsidize the oil companies so they can lower gas prices. Does our sample provide evidence of their claim? Let's investigate this in parts.

a.) Assume they are right and that the true population percentage of Americans who feel the gas prices are a major problem or worse is 75%. Use the formula given below to find the standard deviation of the sampling percentages. (The mean of the sampling percentages is assumed to be 75%.)

$$s = \sqrt{\frac{p(1-p)}{n}}$$

b.) Using the Gallup sample given on the first page, what is the sample percentage of people who feel that gas prices are a major problem or worse? (Notice this combines the first two results in the table.)

c.) Assume the oil companies' claim is correct, that the true percentage who feel gas prices are a major problem or worse is 75%. What is the probability that we would get a sample result equal to or less than our own? Draw the appropriate normal curve with the mean and our sample percentage marked. Also, shade the area in which we are interested. Find the standard score and the probability we want.

d.) Given the answer to part c, does our sample result support the oil companies' claim? In other words, if they correctly found the true population percentage to be 75%, wouldn't our sample give a result that is relatively near that number? Would the probability of obtaining a sample such as ours be so low?

Basically, when we get a sample result that has a very unlikely chance of occurring, there are two possible reasons. One, either the sample is flawed. Or, two, the assumption that we know the true population percentage is incorrect.

We will assume all samples are performed correctly and the problem lies with our population assumption. That is the key idea behind hypothesis testing. We will investigate this further with additional worksheets.