When testing hypotheses using a sample $x_1, x_2, \ldots, x_n$ we use the fact that the sampling distribution $\bar{x}$ is known. But what is it? We'll see in this project that if the parent population is normal with mean μ and variance σ^2 and sample variance s^2 then

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

is normally distributed with mean 0 and variance 1. Whereas

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

is not normally distributed, but rather has the Student t-distribution with $n-1$ degrees of freedom. This distribution is very similar to the normal, although it has longer tails. For large n the two are indistinguishable.

For this project we will do simulations to investigate. There are two functions for doing the simulation that need to be downloaded. Type this exactly

```
> f="http://www.math.csi.cuny.edu/verzani/R/make.z.R"
> source(url(f))
```

Now there are two functions `make.z` and `make.t`.

1 Distribution of z

The distribution z is normal with mean 0 and variance 1. We can check this by running the following command

```
> hist(make.z(n=1))
```

This will draw a histogram of 500 random samples from the distribution when $n = 1$. That is the parent distribution as then $\bar{x} = x_1$.

Notice this is normal as we start with normal data.

Now compare to the histogram for different values of n.

1.1 For $n = 5, 15$ and 50 make histograms as above. Are they similar? Are they identical? Try to draw them on a sheet of paper to be handed in.

1.2 You can compare shapes with side-by-side boxplots. The following command will make them for $n - 5, 15$ and 50.

```
> boxplot(make.z(5),make.z(15),make.z(50))
```

Are they similar? Are they identical? Try to draw them on a sheet of paper to be handed in.

2 The distribution of t.

Notice the only difference between t and z is the replacement of the constant σ^2 with the random s^2 (it depends on the sample). When s^2 is lower than σ^2 the denominator is smaller and the value of t is greater. Thus, the t distribution has a chance of larger values. This is especially true when n is small.

2.1 Make a histogram of the t for $n = 5, 15$ and 50. This is done, for example, with

```
> hist(make.t(5))
```

Are they similar? Are they identical? Try to draw them on a sheet of paper to be handed in.

2.2 You can compare to the normal by comparing boxplots. For $n = 5, 50$ compare boxplots with a command like

```
boxplot(make.t(5),make.z(5))
```

Are they similar? Are they identical? Try to draw them on a sheet of paper to be handed in.

2.3 Make three boxplots for $n = 5, 15$ and 50. Are they similar? Are they identical? Try to draw them on a sheet of paper to be handed in.

2.1 The central limit theorem

The central limit theorem says that even if the x_i are not normal, the distribution of z is approximately normal if n is large enough.

We illustrate this with the exponential distribution. You can get an idea of what it looks like by letting $n = 1$ and adding the `"exp"` argument as follows

```
> hist(make.z(1,"exp"))
```

You should see a curve which decreases.

2.4 Plot histograms for $n = 5, 15$ and 50. Do any look bell-shaped?

2.5 Make a 3 boxplots on the same graph with $n = 5, 15$ and 50. Are they all similar, are they different? How?