I. Back to the distribution of sample means and the standard normal distribution

The standard deviation of the distribution of sample means (std. error) is \( \sigma_x = \frac{\sigma}{\sqrt{n}} \).

Just as \( Z = \frac{(X_i - \mu)}{\sigma} \) for an individual datum of a population, \( Z = \frac{\bar{X} - \mu}{\sigma_x} \) for a sample mean of a random sample from a population.

Example: Government statistics in China report that the mean birth weight is 109.9 oz (\( \mu \)) and the standard deviation is 13.6 oz (\( \sigma \)). You want to test a research hypothesis that babies born in Yunan province are heavier at birth than the average for China.

\( H_0: \bar{X} = 109.9 \text{ oz} \) vs. \( H_1: \bar{X} \neq 109.9 \text{ oz} \); set \( \alpha = 0.05 \)

For a random sample of 25 male birth weights from Yunan Province; the mean weight, \( \bar{X} \), is 116.0.

\[
Z = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \quad \text{or} \quad Z = \frac{116.0 - 109.9}{13.6/\sqrt{25}} = 2.24
\]

From the standard normal table, a z score of 2.24 corresponds to a probability of 0.125.

II. Type I and Type II error

<table>
<thead>
<tr>
<th></th>
<th>If ( H_0 ) is true</th>
<th>If ( H_0 ) is false</th>
</tr>
</thead>
<tbody>
<tr>
<td>If ( H_0 ) is rejected</td>
<td>Type I error</td>
<td>No error</td>
</tr>
<tr>
<td>If ( H_0 ) is not rejected</td>
<td>No error</td>
<td>Type II error</td>
</tr>
</tbody>
</table>

Type I error (\( \alpha \)) - falsely reject a null hypothesis when it is true.

Type II error (\( \beta \)) - Fail to reject (falsely accept) a null hypothesis when indeed it is false

Always the case that as you decrease \( \alpha \), you increase \( \beta \) and vice versa
III. One-sided versus two-sided testing.

IV. Confidence Intervals

In the Chinese baby example, let's say the national mean was unknown, but the variance was known (185 oz$^2$; standard deviation 13.6 oz.). In our sample of 25 babies, the sample mean is 116.0 oz. We would like to ask the question, how likely is it that 116.0 lies within some interval with a high degree of confidence, say 95% certainty.

We can ask this because we know the initial population was normally distributed with unknown mean $\mu$, and standard deviation $\sigma=13.6$.

$\mu$ is unknown, but is estimated by $\bar{x} = 116.0$ oz.

The 95% confidence interval for $\mu$ is given by $\mu \pm 1.96\sigma_{\bar{x}}$

$$116.0 \pm 1.96 \frac{13.6}{\sqrt{25}} = 116.0 \pm 5.3$$

We are 95% confident that this interval contains the true mean $\mu$. 
V. Student's \( t \) distribution

While all of the above procedures regarding the normal distribution required knowing the population variance or standard deviation, it may seem strange that we would indeed know something quantitative about the population variance or standard deviation, but know nothing about the mean. *More commonly, neither the population mean or variance is known quantitatively.*

The logical, and indeed accepted, approach to estimating the population mean and the population standard deviation is to calculate the sample mean and sample standard deviation. Replacing \( \sigma \) with \( s \), we get a new variable \( t \):

\[
Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \text{becomes} \quad t = \frac{\bar{X} - \mu}{s / \sqrt{n}}
\]