Confidence Intervals

**Margin of error** – shows how accurate a guess is based on the variability of the estimate.

\[ m \geq z^* \frac{\sigma}{\sqrt{n}} \] (need to know how to solve for \( n \))

**Confidence Interval** –

1. An interval calculated from the data (estimate ± margin of error)
2. Gives the probability that the interval will capture the true parameter value in repeated samples

\[ \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \]

is the expression for a confidence interval with an SRS of size \( n \) having a known population standard deviation and mean. The interval is exact when the population distribution is normal and is approximately correct for large \( n \) in other cases.

**Inference Toolbox** - Conditions for constructing a confidence interval for \( \mu \):

1. Identify the population of interest and the parameter you want to draw conclusions about.
2. Choose the appropriate inference procedure. Verify the conditions for using the selected procedure.
3. If the conditions are met, carry out the inference procedure.
   \( \text{CI} = \text{estimate} \pm \text{margin of error} \)
4. Interpret your results in the context of the problem.

**Example for constructing a confidence interval using your Inference Toolbox:**

(p. 549 10.6) A study of the career paths of hotel general managers sent questionnaires to an SRS of 160 hotels belonging to major US hotel chains. There were 114 responses. The average time these 114 general managers had spent with their current company was 11.78 years. Give 99% confidence interval for the mean number of years general managers of major-chain hotels have spent with their current company.

**Step 1:** The population is all general managers of major-chain hotels. The parameter of interest is the mean number of years general managers of major-chain hotels have spent with their current company.

**Step 2:** Construct a confidence interval (inference procedure). Conditions for constructing a confidence interval are met (1. the data comes from an SRS of the population as stated in the problem, and 2. sampling distribution of \( \bar{x} \) is approximately normal because our sample size = 114 and CLT).

**Step 3:** \( \text{CI} = 11.78 \pm 2.576 \left( \frac{3.2}{\sqrt{114}} \right) = 11.78 \pm 0.77 \) or 11.01 to 12.55 years.
Step 4: We are 99% confident that the true mean of the number of years general managers of major-chain hotels have spent with their current company is between 11.01 and 12.55 years.

Cautions for using \( \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \) for estimating a normal mean:
- The data must be an SRS from the population.
- Because \( \bar{x} \) is strongly influenced by outliers you should search for outliers and try to correct them or justify removing them before computing the interval.
- You must know the standard deviation of the population. (This is not a reality.)
- The margin of error in a confidence interval covers only random sampling errors.

Test of Significance

Test of significance asks two questions:
1. Does the sample result (\( \bar{x} \)) reflect the real parameter of the population (usually \( \mu \))?  
2. Could we get the sample result just by chance?

Null hypothesis (\( H_0 \)) is the statement being tested in a test of significance.

P-Value is the probability, computed assuming that \( H_0 \) is true, that the observed outcome would take a value as extreme or more extreme than that actually observed. The smaller the P-value is, the stronger is the evidence against \( H_0 \) provided by the data.

Statistical significance (\( \alpha \)) is the decisive P-value.

Inference Toolbox - Conditions for significance tests:
1. Identify the population of interest and the parameter you want to draw conclusions about. State the null and alternative hypotheses in words and symbols.
2. Choose the appropriate inference procedure. Verify the conditions for using the selected procedure.
3. If the conditions are met, carry out the inference procedure.  
   i. Calculate the test statistic  
   ii. Find the P-value
4. Interpret your results in the context of the problem.
Example for constructing a significance test using your Inference Toolbox:
(p. 576 10.39) Bottles of a popular cola are supposed to contain 300ml of cola. There is some variation from bottle to bottle because the filling machinery is not perfectly precise. The distribution of the contents is normal with standard deviation of 3ml. An inspector who suspects that the bottler is underfilling measures the contents of six bottles. The results are (299.4, 297.7, 301.0, 298.9, 300.2, 297.0). Is this convincing evidence that the mean contents of cola bottles are less than the advertised 300ml?

**Step 1:** The population is all bottles of a popular cola. The parameter of interest is the mean number of milliliters in a bottle of cola 300ml. $H_0: \mu = 300$ versus $H_a: \mu < 300$.

**Step 2:** Construct a one-sample $z$ test for a population mean (inference procedure). Conditions for constructing a one-sample $z$ test for a population mean are met (1. the data comes from an SRS of the population as stated in the problem, and 2. sampling distribution of $\bar{x}$ comes from a population that is approximately normal as stated in the problem).

**Step 3:** $z = \frac{299.0 - 300}{\frac{3}{\sqrt{6}}} = -.79$ and $P$-value $= .2148$

**Step 4:** This is reasonable variation when the null hypothesis is true, so we don’t reject the null hypothesis.
Type I & Type II Error

- If we reject \( H_o \) when in fact \( H_o \) is true, this is a **Type I error**.
- If we accept \( H_o \) when in fact \( H_o \) is false, this is a **Type II error**.
  - *Step 1:* Write the rule for accepting \( H_o \) in terms of \( \bar{x} \)
  - *Step 2:* Find the probability of accepting \( H_o \) assuming that the alternative is true.

The probability that a fixed level \( \alpha \) significance test will reject \( H_o \) when a particular alternative value of the parameter is true is called the **power** of the test against that alternative. The power of a test against any alternative is 1 minus the probability of a Type II error for that alternative.

<table>
<thead>
<tr>
<th>Decision based on sample</th>
<th>Truth about the population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( H_o ) true</td>
</tr>
<tr>
<td>Reject ( H_o )</td>
<td><strong>Type I Error</strong></td>
</tr>
<tr>
<td>Fail to reject ( H_o )</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>