TESTS OF HYPOTHESES

As was mentioned earlier, sometimes we cannot survey or test all persons or objects; therefore, we have to take a sample. From the results of analysis from the sample data, we can predict the results from the population. Some questions that one may want to answer are

1. Are unmarried workers more likely to be absent from work than married workers?
2. Are the sixth graders in a certain school significantly less skilled in their mathematical abilities than the average student in the district?
3. In Fall 1996, did students in Math 163-01 score the same on the exam as students in Math 163-02?
4. Is there any difference between the strengths of steel wire produced by the XY Company and Bob’s Wire Company?
5. A hospital spokesperson claims that the average daily room charge for a specific procedure is $622. Can we reject this claim?

WHAT IS A HYPOTHESIS?

- **Hypothesis:** A statement about the value of a population parameter developed for the purpose of testing.
- **Examples of hypotheses, or statements, made about a population parameter are:**
  - The mean monthly income from all sources for systems analysts is $3,625.
  - Twenty percent of all juvenile offenders ultimately are caught and sentenced to prison.

- **Hypothesis testing:** A procedure, based on sample evidence and probability theory, used to determine whether the hypothesis is a reasonable statement and should not be rejected, or is unreasonable and should be rejected.

| Following is a five-step procedure for testing a hypothesis. |

5 STEPS IN THE HYPOTHESIS TESTING PROCEDURE

1. State the null hypothesis and the alternate hypothesis.
   - Null Hypothesis – statement about the value of a population parameter.
   - Alternate Hypothesis – statement that is accepted if evidence proves null hypothesis to be false.
2. **Select the appropriate test statistic and level of significance.**
When testing a **hypothesis of a proportion**, we use the z-statistic or z-test and the formula

\[ z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \]

When testing a **hypothesis of a mean**, we use the z-statistic or we use the t-statistic according to the following conditions.

**If the population standard deviation, \( \sigma \), is known and either the data is normally distributed or the sample size \( n > 30 \),** we use the normal distribution (z-statistic).

**When the population standard deviation, \( \sigma \), is unknown and either the data is normally distributed or the sample size is greater than 30 \( n > 30 \),** we use the t-distribution (t-statistic).

A traditional guideline for choosing the **level of significance** is as follows: (a) the 0.10 level for political polling, (b) the 0.05 level for consumer research projects, and (c) the 0.01 level for quality assurance work.

3. **State the decision rules.**
The decision rules state the conditions under which the null hypothesis will be **accepted or rejected**. The **critical value** for the test-statistic is determined by the level of significance. The critical value is the value that divides the non-reject region from the reject region.

4. **Compute the appropriate test statistic and make the decision.**
When we use the **z-statistic**, we use the formula

\[ z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \]

When we use the **t-statistic**, we use the formula

\[ t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \]

**Compare the computed test statistic with critical value.** If the computed value is within the rejection region(s), we reject the null hypothesis; otherwise, we do not reject the null hypothesis.

5. **Interpret the decision.**
Based on the decision in Step 4, we **state a conclusion in the context of the original problem.**
EXAMPLE: The average score of all sixth graders in school District A on a math aptitude exam is 75 with a standard deviation of 8.1. A random sample of 100 students in one school was taken. The mean score of these 100 students was 71. Does this indicate that the students of this school are significantly less skilled in their mathematical abilities than the average student in the district? (Use a 5% level of significance.)

SOLUTION: In this problem, we know the mean and standard deviation for the population, \( \mu = 75 \) and \( \sigma = 8.1 \) (all sixth graders in district A). We also know the mean for the sample of 100 students in a certain school in district A. Thus, we are testing the sample mean against the population mean with a population standard deviation (\( \sigma \) is known).

This is a large sample because \( n > 30 \) which is usually used to determine whether to use large or small sample techniques. Since \( \sigma \) is known and \( n > 30 \), we use the z-test that is based on the normal curve or normal distribution.

Step 1. State the null hypothesis (contains =, \( \geq \), or \( \leq \)) and alternate hypothesis (usually contains “not”). Think of the statement “Does this indicate that the students of this school are significantly less skilled in their mathematical abilities than the average student in the district?”

From “…students of this school are significantly less skilled…,” we write the alternate hypothesis as \( H_1: \mu < 75 \)

\[ H_0: \mu \geq 75 \quad H_1: \mu < 75 \]

Step 2. Select a level of significance.

Stated in the problem as 5% or \( \alpha = 0.05 \)

Step 3. Identify the statistical test to use.

Use z-test because \( \sigma \) is known and the sample (\( n=100 \)) is a large sample (\( n > 30 \)).

\[
z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}
\]

Recall that in the normal curve, \( Z=0 \) corresponds to the mean. \( Z=1, 2, 3 \) represent 1, 2, and 3 standard deviations above the mean; the negatives are below the mean.
Step 4. Formulate a decision rule.

Since the alternate hypothesis states \( \mu < 75 \), this is a one-tailed test to the left. For \( \alpha = 0.05 \), we find \( z \) in the normal curve table that gives a probability of 0.05 to the left of \( z \). This means the negative of the \( z \) value (critical value) corresponding to a table value of 0.5 - 0.05 = 0.45 or \( z = -1.645 \).

That is \( P(z < -1.645) = 0.05 \). Because 0.4500 is exactly half way between 0.4495 and 0.4505, we get half way between 1.640 and 1.650 to get \( z = 1.645 \). Since 71 is to the left of 75, we have \( z = -1.645 \). That is \( P(z < -1.645) = 0.05 \).

[NOTE: This is not a scaled drawing.]

\[
\begin{array}{c|c|c}
X & 71 & 75 \\
Z & -1.645 & 0 \\
\end{array}
\]

Thus, we reject the null hypothesis if \( z < -1.645 \). And accept the alternate hypothesis that the students in the school sampled are less skilled in math aptitude than those in district A.

Step 5. Take a sample; arrive at a decision.

The sample of 100 students have been tested and found that their mean score was 71. Using the statistical test (z-test) identified in Step 3 compute the test statistic by the formula from Step 3.

\[
z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{71 - 75}{8.1/\sqrt{100}} = -4.938
\]

Since the computed \( z = -4.938 < -1.645 \) (critical \( z \) value), we reject the null hypothesis that the students in the school are not less skilled in mathematical ability. Thus, we conclude that the sixth graders in the school are less skilled in mathematical ability than the sixth graders in District A.

The following problem is presented for students to work:

A sample of 250 married workers showed 22 missed more than 5 days last year for any reason. A sample of 300 unmarried workers showed 35 missed more than 5 days. Use the 5% level of significance to test and answer the question: Are unmarried workers more likely to be absent from work than married workers?