ENGINEERING STATISTICS

(Lectures)

University of Technology,
Building and Construction Engineering Department
(Undergraduate study)

SAMPLING THEORY

Dr. Maan S. Hassan

Lecturer: Azhar H. Mahdi

2010 – 2011
Link between Population and Sampling:

1.0 SAMPLING DISTRIBUTIONS

Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.

For instance, the mean fill volume of a can (population) is required to be 300 mm.

An engineer takes a random sample of 25 cans and computes the sample average fill volume to be

\[ \bar{x} = 298 \text{ mm} \]

The engineer will probably decide that the population mean is \( \mu = 300 \) mm, even though the sample mean was 298 mm because he or she knows that the sample mean is a reasonable estimate of \( \mu \) and that a sample mean of 298 mm is very likely to occur, even if the true population mean is \( \mu = 300 \) mm.

Test values of \( \bar{x} \) vary both above and below \( \mu = 300 \) mm.

**Definition**

The probability distribution of a statistic is called a sampling distribution.
The sampling distribution of a statistic depends on:

- The distribution of the population,
- The size of the sample, and
- The method of sample selection.

2.0 SAMPLING DISTRIBUTIONS OF MEANS

Suppose that a random sample of size \( n \) is taken from a normal population with mean \( \mu \) and variance \( \sigma^2 \).

Now each observation in this sample, say, \( X_1, X_2, X_3 \ldots X_n \), is a normally and independently distributed random variable with mean \( \mu \) and variance \( \sigma^2 \).

The sample mean:

\[
\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}
\]

has a normal distribution with mean:

\[
\mu_{\bar{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \bar{\mu}
\]

and variance:

\[
\sigma^2_{\bar{X}} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}
\]

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean \( \mu \) and variance \( \sigma^2/2 \).
EXAMPLE 1:

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal.

Find the probability that a random sample of $n=25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $x\bar{=}\bar{X}$ is normal, with mean $\mu_{x\bar{=}} = 100$ ohms and a standard deviation of:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability (shaded area) is shown in the figure below:

![Normal Distribution with Standard Deviation](image)

Standardizing the point $x\bar{=} = 95$ in the Figure. We find that:

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,
\[ P(\bar{X} < 95) = P(Z < -2.5) = 0.0062 \]

3.0 SAMPLING DISTRIBUTIONS OF DIFFERENCES:

For two independent populations, let the first population has mean \( \mu_1 \) and variance \( \sigma_1^2 \) and the second population has mean \( \mu_2 \) and variance \( \sigma_2^2 \). Suppose that both populations are normally distributed. Then, we can say that the sampling distribution of \((\bar{x}_1 - \bar{x}_2)\) is normal with mean:

\[ \mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2 \]

And variance

\[ \sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \]

**Definition:**

If we have two independent populations with means \( \mu_1 \) and \( \mu_2 \) and variances \( \sigma_1^2 \) and \( \sigma_2^2 \) and if \( \bar{x}_1 \) and \( \bar{x}_2 \) are the sample means of two independent random samples of sizes \( n_1 \) and \( n_2 \) from these populations, then the sampling distribution is:

\[ Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \]
EXAMPLE 2:

The effective life of a component used in an engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution.

The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of $n_1 = 16$ components is selected from the “old” process and a random sample of $n_2 = 25$ components is selected from the “improved” process.

What is the probability that the difference in the two sample means $\bar{x}_2 - \bar{x}_1$ is at least 25 hours? Assume that the old and improved processes can be regarded as independent populations.

The distribution of $\bar{x}_1$ is normal with mean $\mu_1 = 5000$ hours and standard deviation

$$\sigma_1/\sqrt{n_1} = 40/\sqrt{16} = 10$$

and the distribution of $\bar{x}_2$ is normal with mean $\mu_2 = 5050$ hours and standard deviation

$$\sigma_2/\sqrt{n_2} = 30/\sqrt{25} = 6$$

Now the distribution of $\bar{x}_2 - \bar{x}_1$ is normal with mean

$$\mu_2 - \mu_1 = 5050 - 5000 = 50$$

and variance

$$\sigma_2^2/n_2^2 + \sigma_1^2/n_1^2 = 6^2 + 10^2 = 136$$

This sampling distribution is shown in the Figure below:

The probability that $\bar{x}_2 - \bar{x}_1 \geq 25$ hours is the shaded portion of the normal distribution in this figure.
So,
\[ z = \frac{25 - 50}{\sqrt{136}} = -2.14 \]
and we find that:
\[ P(\overline{X}_2 - \overline{X}_1 \geq 25) = P(Z \geq -2.14) = 0.9838 \]

**EXERCISES:**

1. PVC pipe is manufactured with a mean diameter of 1.01 inch and a standard deviation of 0.003 inch. Find the probability that a random sample of \( n = 9 \) sections of pipe will have a sample mean diameter greater than 1.009 inch and less than 1.012 inch.

2. A synthetic fiber used in manufacturing carpet has tensile strength that is normally distributed with mean 75.5 psi and standard deviation 3.5 psi. Find the probability that a random sample of \( n = 6 \) fiber specimens will have sample mean tensile strength that exceeds 75.75 psi.

3. A random sample of size \( n_1 = 16 \) is selected from a normal population with a mean of 75 and a standard deviation of 8. A second random sample of size \( n_2 = 9 \) is taken from another normal population with mean 70 and standard deviation 12. Let \( \overline{x}_1 \) and \( \overline{x}_2 \) be the two sample means. Find
   a) The probability that \( \overline{x}_1 - \overline{x}_2 \) exceeds 4
   b) The probability that \( 3.5 \leq \overline{x}_1 - \overline{x}_2 \leq 5.5 \)

4. The elasticity of a polymer is affected by the concentration of a reactant. When low concentration is used, the true mean elasticity is 55, and when high concentration is used the mean elasticity is 60. The standard deviation of elasticity is 4, regardless of concentration. If two random samples of size 16 are taken, find the probability that \( \overline{x}_{\text{high}} - \overline{x}_{\text{low}} \geq 2 \).