Lecture 26
Sampling Distributions

1 Introduction

In this lecture, we focous on sampling from distributions or populations and study the important quantities (parameters) as the sample mean and sample variance which will be of vital importance in future chapters. Sometimes, it would be difficult to study the parameters present in a population because of its size and convinience. It is necessary to take a sample from the population under study and then try to infer or study about the population parameters. It may be noted that if the population is so small then we can study the population as a whole. Statistical properties suggest that sample is a true reflection of the population. Next we would like to define some useful definitions, examples and properties of a population and a sample.

Definition 1 *A set consists of the totality of observations with which we are concerned, whether the size (number) is finite or infinite is called a* **population.**

Definition 2 *Any subset of a population is called a* **sample.**

Definition 3 *Let $X_1, X_2, \ldots, X_n$ be n independent random variables, each having the same probability distribution with pmf/pdf $f(x)$. Define $X_1, X_2, \ldots, X_n$ to be a random sample of size n from the population $f(x)$ and write the joint distribution as*

$$f(x_1, x_2, \ldots x_n) = f(x_1)f(x_2)\ldots f(x_n).$$

In other words, a random sample consists of independent and identically distributed (iid) random variables.

Example 1 *The sample mean $\overline{X}$, sample variance S^2, median M, mode etc are examples of a statistic.*

Definition 4 *Any function of the random variables constituting a random sample is called a statistic. Let $X_1, X_2, \ldots, X_n$ be iid random variables and further let $x_1, x_2, \ldots, x_n$ be be the observed values of the sample. Then we define the sample mean and sample variance as follows*

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2\right].$$

where $\overline{x}$ and s^2 are realizations of $\overline{X}$ and S^2 respectively. The positive square root of the sample varience is called the sample standard deviation. Student have studied these satistics in the first Chapter 1.

Definition 5 *The sample median is defined as*

$$m = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)}, & \text{if } n \text{ is even.} \end{cases} \tag{1.1}$$

The observations may be arranged in ascending or desending order to find out the middle terms or median depending upon n is even or odd. The sample median is also a location measure that shows the middle value of the sample. The sample MODE is the value of the sample that occurs most often.

Question-2:-The lengths of time, in minutes, that 10 patients waited in a doctor?s office before receiving treatment were recorded as follows: 5, 11, 9, 5, 10, 15, 6, 10, 5, and 10. Treating the data as a random sample, find (a) the mean; (b) the median; (c) the mode.
Answer:-$\overline{x} = 8.6$, $m = 9.5min$ and MODE is 5 and 10 minutes.
Question-12:-The tar contents of 8 brands of cigarettes selected at random from the latest list released by the Federal Trade Commission are as follows: 7.3, 8.6, 10.4, 16.1, 12.2, 15.1, 14.5, and 9.3 milligrams. Calculate (a) the mean; (b) the variance.
Answer:-$\overline{x} = 11.69mg$, and

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] = 10.776mg.$$

Assignments
Students are advised to work out the following problems as assignments.
Page-257, Q-3, 5, 7 10.

Definition 6 *The distribution of a statistic is called the* **sampling distribution.**

2 Sampling Distributions of $\overline{X}$, S^2 and the Central Limit Theorem

In order to know the distributions of $\overline{X}$, S^2, we need the following theorem and few sampling distributions..

Theorem 2.1 (The Central Limit Theorem) *Let $X_1, X_2, \ldots, X_n$ be a sequence of iid random variables taken from a population with mean μ and variance σ^2, then the limiting form of the distribution of*

$$Z = \sqrt{n}\left(\frac{\overline{X} - \mu}{\sigma}\right),$$

as $n \to \infty$, is the standard normal distribution $N(0, 1)$.

As a consequence of the above theorem, the following conclusion can be derived. Let $X_1, X_2, \ldots, X_k$ be an iid Binomial random variables with mean np and variance npq, then the random variable $Z = \sqrt{k}\left(\frac{\overline{X}-np}{\sqrt{npq}}\right)$, has a standard normal distribution for large k. The above theorem can be applied to any probability distribution, may be continous or discrete.

Example 2 *An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean* 800 *hours and standard deviation of* 40 *hours. Find the probability that a random sample of* 16 *bulbs will have an average life of less than* 775 *hours.*

According to the question, we need to compute $P(\overline{X} \leq 775)$. Applying Central Limit Theorem, we get

$$z = \sqrt{n}\left(\frac{\overline{x} - \mu}{\sigma}\right) = 4\left(\frac{775 - 800}{40}\right) = -2.5.$$

Hence $P(\overline{X} < 775) = P(Z < -2.5) = 0.0062$. (Use standard normal tables.)

Question-17:- If all possible samples of size 16 are drawn from a normal population with mean 50 and standard deviation 5, what is the probability that the sample mean $\overline{X}$ will fall in the interval from $\mu_{\overline{X}} - 1.9\sigma_{\overline{X}}$ to $\mu_{\overline{X}} - 0.4\sigma_{\overline{X}}$? Assume that the sample means can be measured to any degree of accuracy.

Answer:- We have to find $P(\mu_{\overline{X}} - 1.9\sigma_{\overline{X}} < \overline{X} < \mu_{\overline{X}} - 0.4\sigma_{\overline{X}})$. Substituting the values of $\mu_{\overline{X}} = 50$ and $\sigma_{\overline{X}} = \frac{5}{4}$, then we have

$$P(\mu_{\overline{X}} - 1.9\sigma_{\overline{X}} < \overline{X} < \mu_{\overline{X}} - 0.4\sigma_{\overline{X}}) = P(-1.9 < Z < -0.4) = P(Z < -0.4) - P(Z < -1.9).$$

Using standard normal tables we get the required probability $= 0.3446 - 0.0287 = 0.3159$.

Question-23:- The random variable X, representing the number of cherries in a cherry puff, has the following probability distribution $P(X = 4) = 0.2$, $P(X = 5) = 0.4$, $P(X = 6) = 0.3$, $P(X = 7) = 0.1$, then find the mean $\mu_{\overline{X}}$ and variance $\sigma^2_{\overline{X}}$ of the mean $\overline{X}$ for random samples of 36 cherry puffs.

Answer:- Direct calculation gives $\mu = 5.3$ and $\sigma^2 = 0.81$. Hence $\mu_{\overline{X}} = 5.3$ and $\sigma^2_{\overline{X}} = \frac{0.81}{36} = 0.0225$.

Assignment:-Q-19, 20, 24