2025-11-30

Statistical Inference

Bhashkar Paudyal

Abstract

This course was taught by Prof. Dhiraj Giri. It teaches statistical Inference.

Contents

1 Sampling and Some Sampling Distributions

1.1 Sampling Distribution

Definition 1.1.1 (Sampling Distribution)

Suppose you have a population from which you take every possible sample of n entities's attribute denoted by $S = \{S_1, S_2, S_3, \ldots\}$ where S_i is the tuple of values of attributes of the n entities in the i^{th} sample $s.t.$ $S_i = (v_1, v_2, v_3, \ldots, v_n)$ where v_i is the value of the attribute of the i^{th} entity in S_i. Then the distribution of some summary statistics of each sample gives the sampling distribution of the sample statistics.

ⓘ **Note**

A sample is a subset taken from the whole population.

Example. Take the values $1, 2, 3, 4, 5$. Here, $\mu = 3, s = \sqrt{2}$

Let's take every possible sample of size 2 from this set.
{1,2}, {1,3}, {1,4}, {1,5}, {2,3}, {2,4}, {2,5}, {3,4}, {3,5}, {4,5}

Taking the mean of each sample,
1.5, 2, 2.5, 3, 2.5, 3, 3.5, 3.5, 4, 4.5

The frequency distribution of this sample mean is the sampling distribution of sample mean.

x	f
1.5	1
2	1
2.5	2
3	2
3.5	2
4	1
4.5	1

Table 1: Frequency Distribution of sample mean

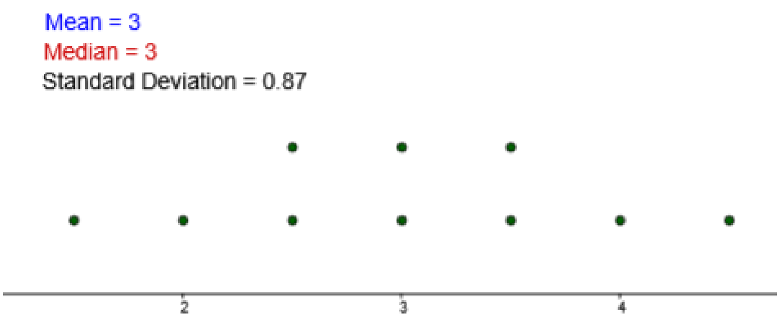

Figure 1: Distribution of the sample mean

Important

In the above example, the sample statistics was mean, but it could have been any other sample statistics.

We used sample mean as the sample statistics so, the frequency distribution in Table 1 is called the **sampling distribution of the sample mean**.

1.2 Recalling Normal Distribution

Normal Distribution: If $X \sim N(\mu, \sigma^2)$ then the PDF of X is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1}$$

and the CDF is:

$$F(x) = \int_a^b f(x)dx \tag{2}$$

Standard Normal Distribution: If $Z \sim N(0, 1)$ then the PDF of Z is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tag{3}$$

Note

The Normal distribution is a family of curve, so it is not unique. While the standard normal distribution is a normal distribution curve so it is a single unique curve. This makes it easy to pre-compute the integrals of the CDF and store it in a table known as **z-table**.

We will be dealing with **z-table** and standard normal distribution. If a variable is not standard normal, we will standardize it.

For standardization, use the following variable transformation.

$$Z = \frac{X - \mu}{\sigma} \tag{4}$$

1.3 Central Limit Theorem

Theorem 1.3.1 (Central Limit Theorem)

The Central Limit Theorem (CLT) states that if you take sufficiently large samples from any population, the distribution of the sample means will be approximately a normal distribution, regardless of the original population's distribution.

Important

Here, if $\overline{X} \sim N\left(\mu, \sigma^2_{\overline{X}}\right)$ is a random variable that models the mean of a sample. To standardize this we have to use the following variable transformation:

$$Z_{\overline{X}} = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma/\sqrt{n}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{5}$$

where,
$\overline{X}$ = sample mean random variable
$\boldsymbol{\mu_{\overline{X}} = \mu}$ **= mean of sample means = population mean**
$\sigma_{\overline{X}} = \sigma/\sqrt{n}$ = standard error

1.4 Finite Population Correction Factor

The finite population correction factor (FPC) is a correction factor used to correct the standard error when sampling where the sample size is less than 5% of the total population.

Sampling Distribution of finite population

The sampling distribution of sample mean, $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\frac{N-n}{n-1}\right)$ where,
μ = population mean
σ = population standard deviation

1.5 Sampling Distribution of Sample Proportion

1.5.1 Proportion

Suppose there are 100 students in a class. If 30 students like maths, 40 like science and 30 like english. Then the proportion of liking maths : science : english = 3:4:3 That means, 30% of them maths, 30% like english and 40% like maths. This proportion can be written as a number from 0 to 1 like 0.3, 0.3, and 0.4.

Sample Proportion can be thought of relative frequency in the frequency distribution in Table 1.

x	f	p = f/$\sum f$
1.5	1	0.1
2	1	0.1
2.5	2	0.2
3	2	0.2
3.5	2	0.2
4	1	0.1
4.5	1	0.1

1.5.2 Sampling Distribution of Sample Proportion

Consider the random variable, $p \sim N\left(P, \frac{PQ}{n}\right) = N\left(\mu_p, \sigma_p^2\right)$

Important

The variable transformation is with the following formula:

$$Z_p = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \tag{6}$$

Here, P is the true population proportion, $Q = 1 - P$.

1.6 Sampling Distribution of Difference Between Sample Mean

This distribution can be modeled by

$$\overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}\right) \tag{7}$$

where,

$\mu_{\overline{X_1} - \overline{X_2} = \mu_1 - \mu_2}$

$\sigma_{\overline{X_1} - \overline{X_2} = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}$

1.7 Sampling Distribution of Difference Between Sample Proportion

This distribution can be modeled by

$$p_1 - p_2 \sim N\left(P_1 - P_2, \frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right) \tag{8}$$

where,

$\mu_{p_1 - p_2} = P_1 - P_2$

$\sigma_{p_1 - p_2} = \frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}$