SDM - IE

Group Members

1. Rishabh Tiwari, 57
2. Yoshit Verma, 58
3. Vedanti Wadekar, 59
4. Nidhi Worah, 60

Title Of Project

Application of central limit theorem in cholesterol

Theoretical Background

This report will begin with a brief introduction, accompanied through the evaluation, and give up
with recommendations for in addition analysing. The evaluation will introduce the central limit theorem in cholesterol, explain exclusive kinds of central limit theorem and give examples of its applications in cholesterol.

Cholesterol molecules are transported in blood by large macromolecular assemblies (illustrated below) called lipoproteins that are really a conglomerate of molecules including apolipoproteins, phospholipids, cholesterol, and cholesterol esters. This macromolecular carrier particles make it possible to transport lipid molecules in blood, which is essentially an aqueous system.

Different classes of these lipid transport carriers can be separated (fractionated)based on their density and where they layer out when spun in a centrifuge. High density lipoprotein cholesterol (HDL) is sometimes referred to as the "good cholesterol," because higher concentrations of HDL in blood are associated with a lower risk of coronary heart disease. In contrast, high concentrations of low density lipoprotein cholesterol (LDL) are associated with an increased risk of coronary
heart disease. The illustration on the right outlines how total cholesterol levels are classified in terms of risk, and how the levels of LDL and HDL fractions provide additional information regarding risk.

Literature Survey

According to the central limit theorem, the means of a random sample of size, n, from a population with mean, μ, and variance, σ^2, distribute normally with mean, μ, and variance, σ^2/n. Using the central limit theorem, a variety of parametric tests have been developed under assumptions about the parameters that determine the population probability distribution. Compared to non-parametric tests, which do not require any assumptions about the population probability distribution, parametric tests produce more accurate and precise estimates with higher statistical powers. However, many medical researchers use parametric tests to present their data without knowledge of the contribution of the central limit theorem to the development of such tests. Thus, this review presents the basic concepts of the central limit theorem and its role in

Central Limit Theorem Statement

The central limit theorem states that whenever a random sample of size n is taken from any distribution with mean and variance, then the sample mean will be approximately normally distributed with mean and variance. The larger the value of the sample size, the better the approximation to the normal.

This theorem allows us to use a sample to make inferences about a population because it states that if n is sufficiently large, the sampling distribution will be approximately normal no matter what the population looks like

Assumptions of Central Limit Theorem

- The sample should be drawn randomly
- The samples drawn should be independent of each other
- The sample size shouldn't exceed 10% of the population
- The sample size should be sufficiently large ($n \geq 30$)

Formula

Sample mean -

$$\mu_{\bar{x}} = \mu$$

Sample Standard Deviation -

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}$$

The Normal Distribution Variate -

$$z = (\bar{x} - \mu_{\bar{x}})/\sigma_{\bar{x}}$$

Cholesterol Molecules

Cholesterol molecules are transported in blood by large macromolecular assemblies (illustrated below) called lipoproteins. This macromolecular carrier particles make it possible to transport lipid molecules in blood, which is essentially an aqueous system.

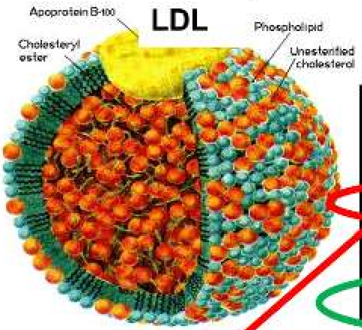
Majorly, there are 2 types of lipid transport carriers -

- High density lipoprotein cholesterol(HDL) - good cholestrol
- Low density lipoprotein cholesterol(LDL) - bad cholestrol

The illustration on the right outlines how total cholesterol levels are classified in terms of risk, and how the levels of LDL and HDL fractions provide additional information regarding risk.

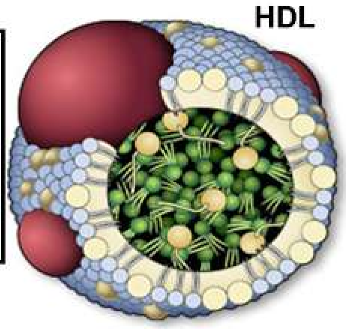
Total Cholesterol (mg/deciliter)

- Desirable: <200
- Borderline High: 200-239
- High: 240+



Total Fractionated into Lipoprotein Carriers

- Chylomicrons
- VLDL (very low density)
- LDL (low density)
- IDL (intermediate density)
- HDL (high density)



LDL (mg/deciliter)	
<100	Optimal
100-129	Above optimal
130-159	Borderline high
160-189	High
>190	Very high

HDL Concentration (mg/deciliter)	
<40	Low (High risk)
40-50	Average risk
60+	High (Low risk)

Example 1:

Data from a study was found that subjects over age 50 had a mean HDL of 54 and a standard deviation of 17. Suppose a physician has 40 patients over age 50 and wants to determine the probability that the mean HDL cholesterol for this sample of 40 men is 60 mg/dl or more (i.e., low risk).

Answer:

The population mean is 54, but the question is what is the probability that the sample mean will be >60?

In general,

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

the standard deviation of the sample mean is

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

Therefore, the formula to standardize a sample mean is:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

And in this case:

$$Z = \frac{60 - 54}{17 / \sqrt{40}} = \frac{6}{2.7} = \mathbf{2.22}$$

P(Z > 2.22) can be looked up in the standard normal distribution table, and because we want the probability that P(Z > 2.22), we compute is as P(Z > 2.22) = 1 - 0.9868 = 0.0132.

Therefore, the probability that the mean HDL in these 40 patients will exceed 60 is 1.32%.

Example 2:

What is the probability that the mean HDL cholesterol among these 40 patients is less than 50 in example 1?

Answer:

In general,

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

the standard deviation of the sample mean is

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

Therefore, the formula to standardize a sample mean is:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

And in this case:

$$Z = \frac{50 - 54}{17 / \sqrt{40}} = \frac{-4}{2.7} = -1.48$$

From the standard normal distribution table P(Z<-1.48)=0.0694.

Therefore, the probability that the mean HDL among these 40 patients will be less that 50 is 6.94%.

Conclusion:

the sampling distributions of the mean for smaller sample sizes are much broader. For small sample sizes, it's not unusual for sample means to be further away from the actual population mean. You obtain less precise estimates.

In closing, understanding the central limit theorem is crucial when it comes to trusting the validity of your results and assessing the precision of your estimates. Use large sample sizes to satisfy the normality assumption even when your data are non normally distributed and to obtain more precise estimates.