Sampling &Sampling Distribution

SC Agarkar

VNBRIMS, Thane

Sampling Distribution

Population	Sample	Sample Statistic	Sampling Distribution
Water in a river	10 gallon container of water	Mean of Mercury in ppm of water	Sampling distribution of the mean
All professional basketball player	Groups of five players	Median height	Sampling distribution of the median
All parts produced by a manufacturing process	50 parts	Proportion defective	Sampling distribution of the proportion

The Standard Error

- The standard deviation of the distribution of a sample statistic is known as the standard error of the statistic.
- The term standard error is used because it conveys a specific meaning. For example the height of a freshman in a large university. We could take a series of samples and calculate the mean height. We expect a some variability in our observed means. This variability results from sampling error due to chance.

The terminology

When we wish to refer to the	We usethe conventional term
Standard deviation of the distribution of sample means	Standard error of the mean
Standard deviation of the distribution of sample proportions	Standard error of the proportion
Standard deviation of the distribution of sample medians	Standard error of the median
Standard deviation of the distribution of sample ranges	Standard error of the range

Conceptual Basis

- Assume that we are studying the distribution of operating hours of all the filter screens in a large industrial pollution control system before a screen become clogged.
- We will get three distributions: The population distribution, The sample frequency distribution and The sampling distribution of mean. They are different from each other.

Mean and Standard Error for normal distribution

- In case of a normal distribution the sampling distribution has a mean equal to the population mean ($\mu_X = \mu$).
- The sampling distribution has a standard deviation (a standard error) equal to the population standard deviation divided by the square root of the sample size.
- $SE = \sigma \div \sqrt{n}$

Example 1

- A bank calculates that its individual savings accounts are normally distributed with a mean of $ 2,000 and a standard deviation of $ 600. If a bank takes a random sample of 100 accounts, what is the probability the sample mean will lie between $ 1,900 and $ 2,050?
- Standard Error of the mean SE $=\sigma \div \sqrt{n} = 60$
- Z value = Difference $(x - \mu)/SE$
- For x=1900, z = -1,67
- For x= 2050, z = 0.83
- Z value of -1.67 means an area of 0.4525 while the z value of 0.83 means an area of 0.2967.
- If we add these values we get 0.7492 as the total probability that the sample mean will be between $1900 and 2050. This result can also be shown graphically.

Example 2

- Daily sales figures of 40 shopkeepers were calculated and the average sales and S.D. were found to be Rs. 528 and 600 respectively. Is the assertion that daily sales on the average is Rs. 400 only contradicted at 5% level of significance by the sample?
- SE = 94.94
- Z = Difference/SE = 1.348
- The value is less than 1.96 (as per the table at 5% level of significance). Therefore, the hypothesis is accepted.

Sampling from Non Normal Population

- In many cases the data may not be normally distributed or the cases to be considered are very few to be approximated by normal distribution. In such cases above formula may not hold true.
- The population distribution and the sampling distribution of the mean in such cases is different. The later tends to show normal distribution.
- If the sample size is large then the sampling distribution of the mean approaches normality regardless of the shape of population distribution.

Central Limit Theorem

- The mean of the sampling distribution of the mean will equal the population mean and the sample distribution of the mean will approach normality (as the sample size increases).
- The above relationship between the shape of population distribution and the shape of the sampling distribution of the mean is called the Central Limit Theorem.
- It assures that the sampling distribution of the mean approaches normalcy as the sample size increases.

Example

- The distribution of annual earnings of all bank tellers with five years experience is skewed negatively. This distribution has a mean of \$ 15000 and SD of \$2000. If we draw the random sample of 30 tellers what is the probability that their earnings will average more than \$15,750?
- SE $=\sigma \div \sqrt{n} = 365.16$
- $Z = 2.05$
- For the z value of 2.05 area is 0.4798. The area between the right hand tail and assumed average is 0.0202.
- There is slightly more than 2% chance of average earnings being more than \$ 15,750.

Relationship between sample size and Standard error

- From the formula we notice that as n increases SE decreases.
- When n=10 SE= 31.63
- When n=100 then SE=10
- Thus by increasing the sample size tenfold the standard error dropped from 31.63 to 10, one third of the original value.
- Owing to the fact that SE varies inversely with the square root of n, there is a diminishing return in sampling.

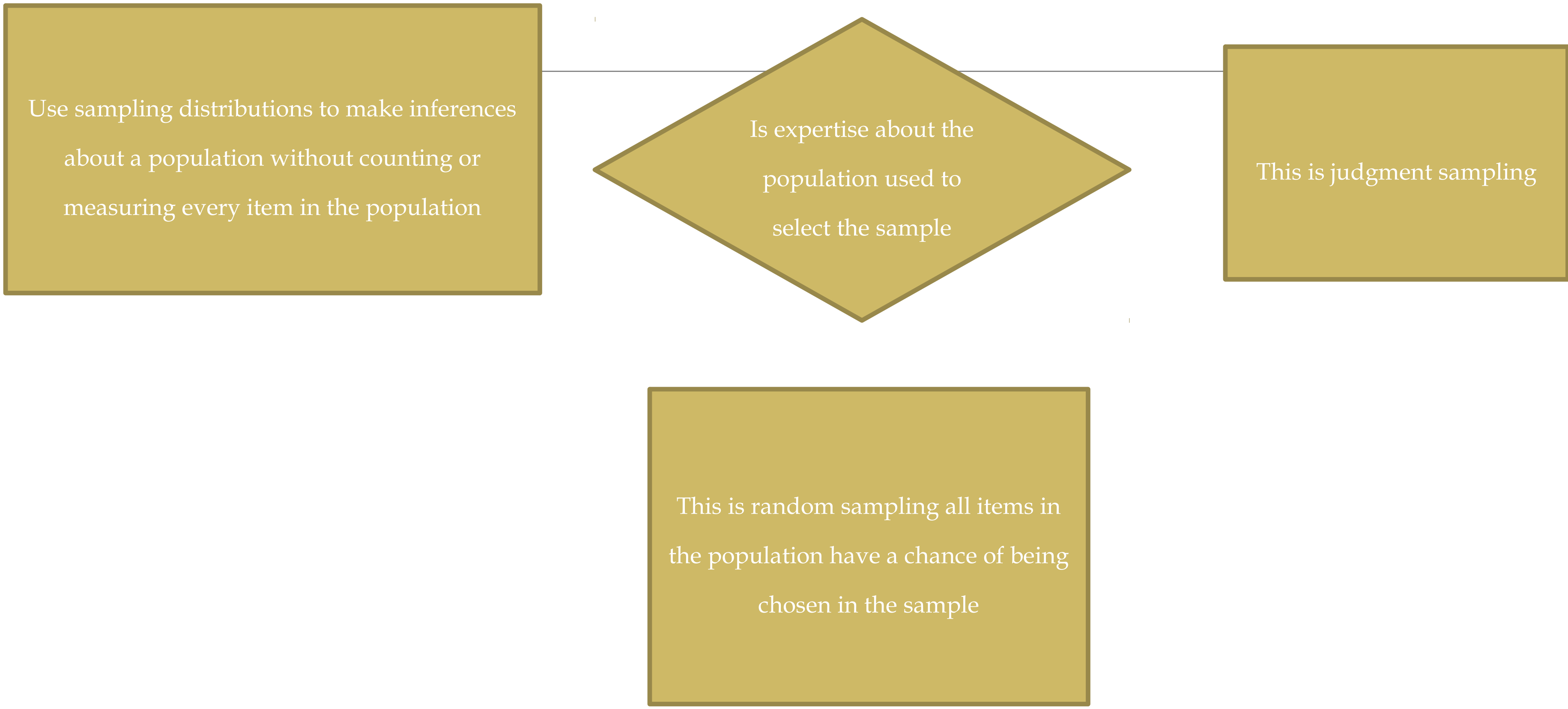
Standard error of the mean for finite population

- Many of the decision makers use finite population. In this case the formula for standard error is
- $SE=\sigma/n *\sqrt{N-n/N-1}$
- The second term in the above formula is called as the finite population multiplier.

Example

- There are 20 textile mills experiencingexcessivelabourturnover. The study indicates that SD of the distribution of annual turnover is 75 employees. If 5 mills are chosen then
- SE= 33.54 * 0.888 = 29.8
- In cases in which population is very large in relation to the size of the sample, the finite population multiplier is close to 1.

Flow chart


Use sampling distributions to make inferences about a population without counting or measuring every item in the population
Is expertise about the population used to select the sample
This is judgment sampling
This is random sampling all items in the population have a chance of being chosen in the sample

Thank you, Would monkey's decision depend on sampling distribution?