

Topic 16 - Sampling Distributions

Statistics for Managers

June 3, 1999

Sampling

For a finite population, if every member in a population is observed (interviewed in case the members are people) then the collection of observed values is a census. However, most finite populations are so large, a census is prohibitively costly.

Sampling the population allows one to infer the values of the populations to any practical degree of accuracy with far less cost than a census.

Statistical inference is based on Scientific Sampling, meaning that every possible sample has a known probability of being taken.

Generally, the mechanism for sampling is such that the probability of observing any sample can be ascertained.

E.g. suppose a population consists of six unknown values: $x_1, x_2, x_3, x_4, x_5, x_6$. One could sample scientifically by making sure that each member has an equal probability of being in the sample. Start by making a list of population ID's (1,2,3,4,5,6), or a frame in statistical terminology.

A sample of size one, $n = 1$, is taken by observing the of the events $\{ID = 1\}, \{ID = 2\}$, etc. with equal probability, i.e. $1/6$.

A sample of size two, $n = 2$, is taken by observing one of the events $\{ID = 1, ID = 2\}, \{ID = 1, ID = 3\}$, etc. with equal probability, i.e.

$$1/\binom{6}{2} = 1/15.$$

A sample in which every population member has an equal probability of appearing in the sample is called a simple random sample.

Sampling Inference

For large or infinite populations, one assumes a certain family of the population probability function with unknown parameters, thus a sample is taken in order to estimate the parameters of the probability distribution.

“Bell Shaped” populations are useful for describing many finite and infinite populations. Thus, it is sufficient to estimate the mean and standard deviation of the population. The sample mean and the sample standard deviation are widely used estimators of the population mean and standard deviation.

An aggregate value computed from a sample is called a sample statistic. The sample mean and standard deviation are examples of sample statistics.

The distribution of values that a sample statistic obtains due to sampling from a population is called a sampling distribution.

Example

Assume we have a six valued population of errors in one physical dimension of a manufactured part. Let the sample size be two, $n = 2$.

Population

ID	Unknown Error Value
1	-4
2	-2
3	-1
4	1
5	2
6	4

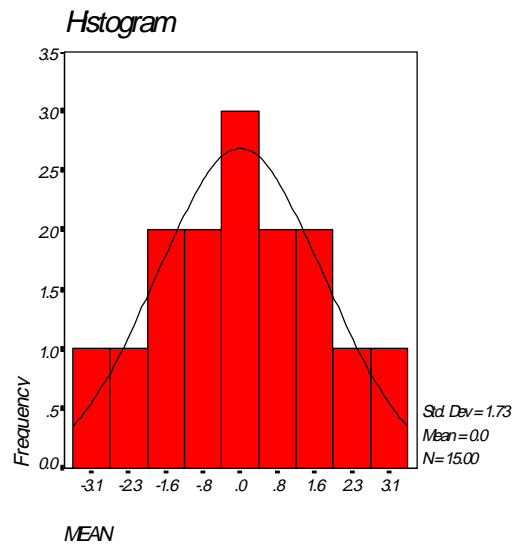
Unknown Population Parameters

$$\mu = 0, \sigma = \sqrt{8.4}$$

All Possible Samples, n = 2

ID Combo	y ₁	y ₂	Mean	VAR	SD
1, 2	-4.00	-2.00	-3.00	2.00	1.41
1, 3	-4.00	-1.00	-2.50	4.50	2.12
1, 4	-4.00	1.00	-1.50	12.50	3.54
1, 5	-4.00	2.00	-1.00	18.00	4.24
1, 6	-4.00	4.00	0.00	32.00	5.66
2, 3	-2.00	-1.00	-1.50	0.50	0.71
2, 4	-2.00	1.00	-0.50	4.50	2.12
2, 5	-2.00	2.00	0.00	8.00	2.83
2, 6	-2.00	4.00	1.00	18.00	4.24
3, 4	-1.00	1.00	0.00	2.00	1.41
3, 5	-1.00	2.00	0.50	4.50	2.12
3, 6	-1.00	4.00	1.50	12.50	3.54
4, 5	1.00	2.00	1.50	0.50	0.71
4, 6	1.00	4.00	2.50	4.50	2.12
5, 6	2.00	4.00	3.00	2.00	1.41

Sampling Distribution of the Sample Mean



Unbiased Sample Statistics

A statistic is said to be an unbiased estimate of a population parameter if the mean of the sampling distribution is equal to the population parameter.

The sample mean and the sample variance are unbiased estimates of the population mean and variance:

$$E(\bar{Y}) = \mu$$
$$E(s^2) = \sigma^2$$

E.g.: For the six valued population example:

$$E(\bar{Y}) = 0$$
$$E(s^2) = 8.4$$

Standard Error of the Mean

For random samples of size n taken from a population having mean μ and standard deviation σ , the sampling distribution of \bar{Y} has

Mean:

$$E(\bar{Y}) = \mu$$

Standard Deviation (called Standard Error of the Mean)

(Finite population size N , Sample without replacement)

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

(Infinite population)

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

E.g.: Six valued population

$$E(\bar{Y}) = 0$$

$$\sigma_{\bar{Y}} = \sqrt{8.4} = 2.898$$

Sample without replacement (finite population size 6)

$$\sigma_{\bar{Y}} = \frac{\sqrt{8.4}}{\sqrt{2}} \sqrt{\frac{6-2}{6-1}} = 1.833$$

Central Limit Theorem

In case the population is infinite or large, the sampling distribution of the sample mean is known to be of only one kind of probability distribution: Normal

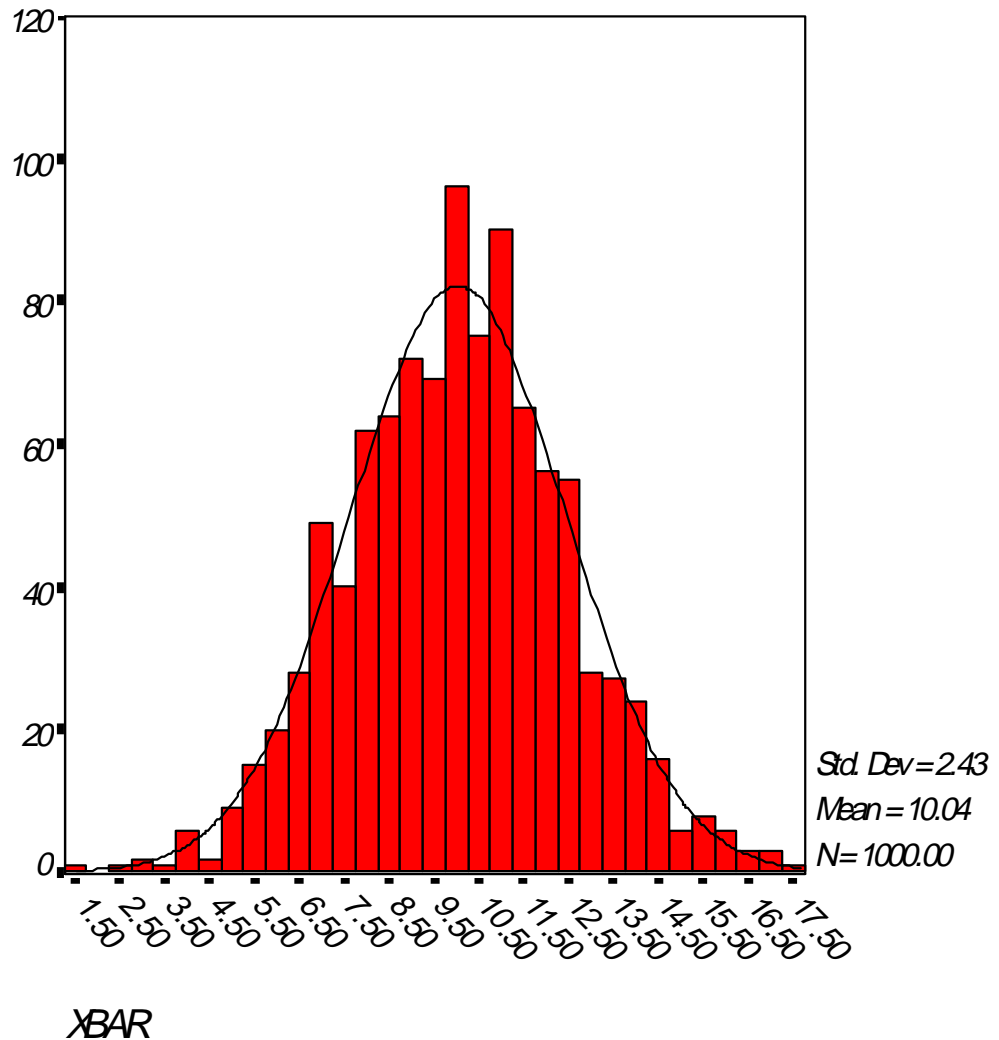


Figure 1 One-Thousand Samples from Population with Mean 10 and SD 20.

Example 1.

Based on the central limit theorem, what is the probability that the sample mean will fall within 5 units of the true population mean with a sample of size $n = 64$ and the population standard deviation is $\sigma = 20$

$$z = \frac{(\mu - 5) - \mu}{20/\sqrt{64}} = -2$$

$$z = \frac{(\mu + 5) - \mu}{20/\sqrt{64}} = +2$$

Therefore, the probability is about 95%.