Statistical inference II

Statistical inference could be defined as the act of drawing inferences about a population or characteristics from information contained in a sample. In this process, if a population value is known, there will be no need to make inferences about them.

Sampling Theory

Sampling theory is a study of relationship of a population and samples drawn from the population. It is of great value in many connections. It is useful in estimation of know population quantities. Useful in estimation of unknown population quantities such as population mean, variance e.t.c. ==Often called population parameter or briefly parameters from a knowledge of corresponding sample qualities such as sample mean, variance, e.t.c. Often called sample statistics or briefly statistics. ==

In general, a study of inferences made concerning a population by use of sample drawn from it together with indications of the accuracy of such inferences using probability theory is called statistical inference.

Sampling Distribution

If we take a sample from a give size of a population and repeat the sampling as many times as possible, calculate the mean of each sample and summarize into a frequency distribution, the resulting distribution is called sampling Dist of the sample mean.

Suppose we have a population
S = (1, 2, 3, 4, 5, 6)
A possible interpretation of the population can be that it is a list of all possible outcomes of throwing a die once. The probability distribution of this population can be as follows.

Fig 1.1 probability of throwing a die once:

x	prob(x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Total	1

Since this is a finite population, we can examine the relationship between the population and sampling distribution. We do this by considering the arithmetic mean as the most important perimeter in statistics.

Ex 1:

a) take a sample (without replacements of any size two) from the both population and calculate the mean and repeat the process until no samples have the same members.

b. Carry out the same process for a sample size of four.

Solution

Sample	x̄(n = z)	Sample	x̄ (n = 4)
1, 2	1.50	1,2,3,4	2.50
1,3	2.00	1,2,3,5	2.75
1,4	2.50	1,2,3,6	3.00
1,5	3.00	1,2,4,5	3.25
1,6	3.50	1,2,4,6	3.50
2,3	2.50	1,2,5,6	3.25
2,4	3.00	1,3,4,5	3.50
2,5	3.50	1,3,4,6	3.75
2,6	4.00	1,3,5,6	4.00
3,4	3.50	1,,4,5,6	3.50
3,,5	4.00	2,3,4,5	3.50
3,6	4.50	2,3,4,5	3.50
4,5	4.50	2,3,5,6	4.25
4,6	5.00	2,4,5,6	4.50
56	5.50	3,4,5,6	4.50

Ex 2:

Classified means resulting from the samples of each of a) and b) of example 1 and calculating their relative frequency.

n=2 n=4

x̄	f	P(x)	x̄	F	P(x)
1.50	1	1/15	2.50	1	1/15
2.00	1	1/15	2.75	1	1/15
2.50	2	2/15	3.00	2	2/15
3.00	2	2/15	3.25	2	2/15
3.50	3	3/15	3.50	3	3/15
4.00	2	2/15	3.75	2	2/15
4.50	2	2/15	4.00	2	2/15
5.00	1	1/15	4.25	1	1/15
5.50	1	1/15	4.50	1	1/15

Ex 3:

Calculate the mean and variants of the sample data in fig1.1

X	F	F(x)	X - x̄	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
X	P(x)	Xp(x)	x-μ	$(x-\mu)^2$	$f(x-\mu)^2$
1	1/6	1/6	-2.5	6.25	6.25/6
2	1/6	2/6	-1.5	2.25	2.25/6
3	1/6	3/6	-0.5	0.25	0.25/6
4	1/6	4/6	0.5	0.25	0.25/6
5	1/6	5/6	1.5	2.5	2.25/6
6	1/6	6/6	2.5	6.25	6.25/6
		21/6			17.50

Mean $= \mu = \sum x\,p(x) \div n$

$= \dfrac{21}{6} = 3.50$

Variance $= \sigma^2 = \dfrac{\sum (x-\bar{x})^2\,P(x)}{f}$

$= \dfrac{17.50}{6} = 2.917$

Similarly, from the sampling distribution showed in example 2 we can calculate the corresponding sample mean and variance for each sampling data.

Solution

For $n = 2$

$\bar{x}$	$P(x)$	$\bar{x}P(x)$	$\bar{x} - 3.5$	$(\bar{x} - 3.5)^2$	$(\bar{x} - 3.5)^2 P(x)$
1.5	1/15	1.5/15	-2.0	4.00	4.0/15
2.0	1/15	2.0/15	-1.5	2.25	2.25/15
2.5	2/15	5.0/15	-1.0	1.00	2.0/15
3.0	2/15	6.0/15	-0.5	0.25	0.50/15
3.5	3/15	10.0/15	0.0	0.00	0.0
4.0	2/15	8.0/15	0.5	2.25	0.50/15
4.5	2/15	9.0/15	1.0	1.00	2.0/15
5.0	1/15	5.0/15	1.5	2.25	2.25/15
5.5	1/15	5.5/15	2.0	4.00	4.0/15
		52.5/15			17.50/15

$$\text{Mean} = \frac{52.50}{15} = 3.50$$

$$\text{Variance} = \sum (\bar{x} - 3.50)^2 P(x) = \frac{17.50}{15} = 1.167$$

Following the same method, we can calculate the mean and variants for n = 4 as mean = 3.50 and variance = 0.2917

Note the quality of the parent's population mean given fig1.1 and the two sampling distribution means as given above. This is not by accident but a property of the method of calculating sample mean.

This property is that the sample mean is an UNBIASED estimates of the population mean. The sampling variants calculated above can be derived by using this formula.

$$\text{Sampling Variance} = \frac{N-n}{N-1} \frac{\sigma^2}{n}$$

$$Samplingvariance = \frac{N-n}{N-1} * \frac{\sigma^2}{n}$$

where N is the population size, σ2 is the population variance and n is the sample size.

In the above example, $n=2$, $N=6$, $\sigma^2 = 2.917$

Sampling variance $= \frac{4}{5}\left(\frac{2.917}{2}\right) = 1.167$

Also for $n=4$, $N=6$, $\sigma^2 = 2.917$

Sampling variance $= \frac{2}{5}\left(\frac{2.917}{4}\right) = 0.2917$

- When N is very large, the formula given for sampling variance above reduces to σ2/2

Definition: The square roots of sampling variance is called the standard error.

$\sigma/\sqrt{n}$

The mean of a sampling distribution is the same as mean of the population. The sampling mean is an unbiased value of the population mean.

sampling and the sampling distribution

Population and samples: a major purpose of doing research is to refer or generalized from a sample to a larger population. This process of inference is accomplished by using a statistical method based on probability. Population is the term used to describe a large set or collection of items that has something in common i.e universe of population consists of the total collection of items or elements that falls within a scope of statistical investigation.

The purpose of defining a statistical population is to provide very explicit limit for the data collection process and for the inference and the conclusion that may be drawn from the study.

A sample on the other hand is a subject of the population selected in such a way that it is representative of the larger population. The term population and sample are relative. An aggregates of element which constitute a population for one purpose may merely the example for another. e.g the average age of students in a class.

Reasons For Sampling.

1. Sample can be studied more simply than population
2. A study of a sample to less expensive than population since smaller number of items are examined.
3. A study of an entire population is impossible in most situations hence, sampling may represent the only possible or practical method to obtain the desired information.
   For example in the case of processing such as manufacturing where the universe is conceptually in finite including all features as well as current production. It is not possible to accomplish a complete enumeration of the population. Also, in destructive sampling of a finite

population. It is possible to affect a complete enumeration of the population but, it will not be practical to do so. For example the production of bulbs.
4. Samples can be selected to reduce heterogenic i.e. very difference.
5. Samples results are often more accurate than the result based on population, since more time and resources can be spent and the people who performed the observation and collection of data.
6. If sample are properly selected, probability method can be used to estimate an error in resistance statistics. It is the aspect of sampling that permit investigator to make probabilities statements about observation in a study.

Methods of Sampling

Items can be selected from statistical universes of duration in a variety of ways. It is useful to distinguish random from non-random method of selection. The best way to ensure that a sample will be reliable and valid inferences is to use probable sample or random sample in which the probability of being included in the sample is known for each subject in population. In a nutshell, this is a definition of random sampling method.

None random sampling method; these are referred to as judgment sampling IE selection method in which judgment is exercise in deciding which elements of a universe is to be included in the sample. The basic reason random sampling is preferable to non random sampling is that in judgment selection, there is no objective method of measuring depreciation or reliability of estimates made from the sample. Under other hands in random sampling, the precision with which estimates of population values can be made obtainable from the sampling value. This is very important to advantage since random something techniques provides an objective basis for measuring errors due to the something process and for stating the degree of confidence to be placed upon the estimates of population values.

Simple random sampling: this is one of the methods of sampling. It has been that a random probability sample is a sample drawn in such a way that the probability of inclusion of every element in the population is known. A simple random sample of size n element is a sample drawn in such a way that every combination of n element has an equal chance of being the sample selected since no practical sampling situation involves sampling without replacement. It is useful to think of this type of sample as one in which each of the n population element has an equal probability 1/n of being the one selected, the first drawn, 1/n of being the one selected, the second one drawn and so on. Until the nth sample has been drawn since there are nCn possible samples of 'n', The probability that any sample of size n will be drawn 1/nCn
E.G Let N=5, n=2 the possible samples will be 5C2 =10. Supposing the 5 number are labeled. Let's take 2 possible samples we have.

Method of Simple Random Sample

1. **Drawing chips from a bowl:** if attention is restricted to the most straightforward situation in which the elements are easily identical and can be numbered. For example, suppose there are hundred students in class and we wish to draw a simple random sampling of 20 of these students without repeating assuring numbers from 1-100 to each of the students and place

these numbers on physically similar slips of paper which could be placed in a bowl shake the ball to accomplish a thorough mixing of the slips then draw the sample. The slope is done and the number of it is recorded. Also, the second is drawn and the number of it is recorded and so on. The students correspond to these 20 numbers constitute the required simple random sampling.

2. **Tables of random numbers;** if the population size is very large, they are both procedure can become quite unwieldy and time consuming. It's may even introduce BIASED. If the slips are not thoroughly mixed the tables of random digits can be used to select the required samples. These tables are useful. Samples as a table of random digits is simple table of digits which have been generated by a random process.
3. **Stratified random sampling:** in stratified random sampling the population is classified into actually exclusive sub groups or strata and probability samples are drawn independently from each of these strata. A sample from each strata maybe obtained by simple random sampling of some other forms of probability sampling for example systems can be combined to yield and over all estimates of a perimeter or then be compared with one another to reveal between strata differences. We shall concentrate on the case where the objective is to obtain an overall estimate of a population parameter by combining the results of stratification as compared to simple random sampling is to obtain reduction in sampling error or synonymously and increase in precision which is as you reduce your sampling error, you are indirectly increasing precision. The reduction in something error is element which are more alike with respect to the other characteristics under investigation stratification is most effective when the element within strata are homogeneous as possible and difference of element amongst strata are as great as possible to minimize differences among elements within strata until maximize differences amongst strata.
4. **Cluster sampling:** this is a technique in which the population is subdivided into groups or clusters, then a probability sample of this clusters is drawn and studied. The primary aim of clusters sampling is to I complete cost sampling in sample design. Something error is reduced in clusters sampling when the units within clusters are has heterogeneous as possible. For instance, if a single cluster duplicates all of the heterogeneity, which exists in the population, we will only need to draw these cluster into our sample to have a good description of the population.
5. **Systematic sampling;** you select kth elements of the population until you have all the sample in the population k can be any number. In many practical applications, systematic sampling is used in place of sample random sampling as a method of obtaining random selection. In a systematic example, every kth elements is drawn from the listed or arranged in some specified manner. The starting point is selected at random from the first k elements. Example if a researcher has serial arranged list of farmers and he de pls of every 5th name on the list where k, then is starting points Is selected at random, the first five names and then every 5th name selected after the first number k. Sometimes it's determined by dividing the numbers of items in sampling frame by the sample size. If the mean of a population occurs in random or then the results of the systematic sampling very similar to those of sample random sampling. However systematic sampling should be around whenever there is a periodic or cyclical very in the order of the population elements.

Definition of Terms

Sampling frame/Frame: a listing of all the elements in the population is called the sampling frame or simply frame

Types of Error

The concept of error is essential one throughout all statistical works. Whenever we have measurement inferences or decision making the possibility of error is presented. There are two different types of error which may be present in statistical measurements namely:

- Systematic error
- Random error

Systematic error cause a measurement to be incorrect in some systematic ways. They are over type which persist even when the sample size is increased. They are errors involved in the procedure of a statistical investigation and may occur in the planning stages or during or after the collection process.

Systematic errors is sometimes called biased. Among the causes of biased are:
Faulty design of a questionnaire such as or refuses by respondent to provide correct information, incomplete sampling frame, mistakes in planning and processing of data and so on. This errors may be viewed as arising primarily from inaccuracies or deficiencies in the increasing instrument. These areas can be reduced or completely eliminated.

Random or something errors arise from operation of a large number of uncontrolled facts called chance. For example, if repeated random samples of the same size are drawn from a population with replacement a particular statistics such as the mean will differ from sample to sample even if the sample definition and procedures are used. These samples mean tend to distribute themselves below and above this population parameter. The difference between the mean of a particular sampling and of population mean is said to be a random error. The complete collection of factors which could be explained why the sample mean deferred from the PV mean is unknown random error decreases on the average.

For this reason that a larger sample of observation is preferred to a smaller one since sampling errors are smaller for larger samples. The results are more reliable and more precise.

Parameter: the value of a measure such as the mean, a median or a standard derivation completed from a population is called parameter. The value of such a measure computed from a sample is called statistics this It's derived from a sample and parameter is derived from population value. For statistics we use Greek letters and Roman letters for population

Function: this is a value by which every member of one set is assigned to apparent with one member of another set. EG suppose x is a set of six students in a seminar and that's y is the set of project topics. If cf It's a rule that's assigned to every element x in the set x a unique elements y in the set y, then f is said to be a f2 that maps x into y. On the other hand, a random variable is a f2 which assigns numerical value to the different outcomes defined by a sample space Don't forget that in the above example the rule is choose a seminar topic which is the random variable.

Example 2: suppose a coin is tossed twice so that the sample space is x, let's represent the number of head s which comes up, so the random variable is how the number of x obtained in two tosses of a fair coin. The rule in this case is to obtain a number of x in a sample space. And this is a random sample.

Sampling Distribution

Let us consider how statistics differ from sample to sample. If repeated, simple random sampling of the same size are drawn from statistical population. Example a given statistic such as a mean or proportion Will vary from one sample to another. The distribution of mean can be examined to estimate a amount of variation that can be from one sample to another. Probability distribution of such a statistic referred 2 as a sampling distribution. Thus, we have a sampling distribution of mean, proportion and so on.

Sampling Distribution of Mean

Let f(x) be the probability distribution of some given population from which we take a sample of size n. Then the probability distribution of a statistical is called the sampling distribution of mean. We use the following theorems to explain this.