10 | 17

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 unknown population quantities such as population mean, variance e.t.c. often called population parameter or breifly PARAMETERS from a knowledge of corresponding sample qualities such as; sample mean, variance, e.t.c. often called sample statistics or breifly STATISTICS. ~~In general,~~

In general, a study of inferences made concerning a population by use of samples 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 given 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 a 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 ~~the~~ 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 parameters in statistics.

Ex 1:

a) Take a sample (without replacement of any size two) from the above 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 of size four.

SOLUTION

Sample	$\bar{x}$ ($n=2$)	Sample	$\bar{x}$ ($n=4$)
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, 6	4·00
4, 5	4·50	2, 3, 5, 6	4·25
4, 6	5·00	2, 4, 5, 6	4·50
5, 6	5·50	3, 4, 5, 6	4·50

$^{N}C_{n}$ $^{P}C_{s}$ $^{6}C_{2}=15$

EX 2:

Classify the means resulting from samples of each of (a) & (b) of Example 1 and calculate their relative frequency.

Solution

n=2

$\bar{x}$	f	P(x)
1.50	1	1/15
2.00	1	1/15
2.50	2	2/15
3.00	2	2/15
3.50	3	3/15
4.00	2	2/15
4.50	2	2/15
5.00	1	1/15
5.50	1	1/15

n=4

$\bar{x}$	f	P(x)
2.50	1	1/15
2.75	1	1/15
3.00	2	2/15
3.25	2	2/15
3.50	3	3/15
3.75	2	2/15
4.00	2	2/15
4.25	1	1/15
4.50	1	1/15

EX 3:

Calculate the mean and variance of the sample data in fig 1.1.

Solution

x	f	P(x)	fx	$xP(x)$	$x-\bar{x}$	$x-\mu$	$(x-\bar{x})^2$	$(x-\mu)^2$	$f(x-\bar{x})^2$	$f(x-\mu)^2$
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		2.25		2.25/6
6		1/6		6/6		2.5		6.25		6.25/6
				21/6						17.50

Mean $= \mu = \Sigma x P(x) = \frac{21}{6} = 3.50$

Variance $= \sigma^2 = \frac{\Sigma (x-\bar{x})^2 P(x)}{f}$

$= \frac{17.50}{6} = 2.9174$

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

SOLUTION

For $n = 2$

$\bar{x}$	$P(\bar{x})$	$\bar{x}\,P(\bar{x})$	$\bar{x} - 3.5$	$(\bar{x} - 3.5)^2$	$(\bar{x} - 3.5)^2 P(\bar{x})$
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	1.00	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

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

$$\text{Variance} = \Sigma (\bar{x} - 3.50)^2 P(\bar{x})$$

$$= \frac{17.50}{15} = 1.167$$

Following the same method we can calculate the mean and variance for $n = 4$ as

Mean = 3.50 and Variance = 0.2917.

Note the equality of the parent pp mean given in pg 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 estimate of the pp mean. The sample variance calculated above can be derived by using this formular.

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

*pp ⇒ population

where N is the pp size, σ^2 is the pp 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 σ^2/n

DEFINITION: The Square root of Sampling Variance is called the Standard Error. $= \sigma/\sqrt{n}$

N.B

The mean of a sampling distribution is the same as mean of the population i.e the Sampling mean is an UNBIASED value of the pp mean.

SAMPLING AND SAMPLING DISTRIBUTION.

Population and Samples: A major purpose of doing research is to refer or generalize from a Sample to a larger population. This process of inference is accomplished by using statistical methods based on probability. Population is the term used to describe a large set or collection of items that has something in common. The Universe of population Consists of the total collection of items or elements that fall within a Scope of statistical Investigation.

The purpose of defining a statistical population is to provide very explicit limits for the data collection process and for the inferences and conclusion that maybe drawn from the study.

~~A Selected in such away that~~

A Sample on the other hand is a subject of the

PP selected in such a way that it is representative of the larger population. The term population and sample are relative. An aggregate of element which constitute a pp for one purpose may merely be a sample 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 is less expensive than pp. Since a smaller number of items are examined.
3) A study of an entire pp is impossible in most situations hence, sampling may represent the only possible or practicable method to obtain the desired information. For example: In the case of processes such as manufacturing where the universe is conceptually infinite including all features as well as current production. It is not possible to accomplish a complete enumeration of the pp. Also, in destructive sampling of a finite pp. ~~Also in~~ ~~observation~~ It is possible to effect a complete enumeration of the pp but, it will not be practical to do so. For example the production of bulbs.
4) Samples can be selected to reduce heterogenal i.e (very difference).
5) Samples results are often more accurate than the result based on pp, since more time and resources can be spent on 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 resisting statistics. It is this aspect of sampling that permits investigator to make probability statements about observation in a study.

METHODS OF SAMPLING. ✓

Items can be selected from statistical universes of duration in a varities of ways. It is useful to distinguish random from non-random method of Selection. The best way to ensure that a sample will need reliable and valid inferences is to use prob. Samples or random samples in which the prob. of being included in the sample is known for each subject in population. In a nutshell, this is a definition of RANDOM SAMPLING METHOD.

NON RANDOM SAMPLING METHOD ⇒ These are refered to as judgment sampling i.e selection method in which judgment is exercised in deciding which element of a universe is to be included in the sample. The basic reason random sampling is preferable to non random sampling is that in Judgement selection, there is no objective method of measuring the precision or reliability of estimate made from the sample. On the other hand in random Sampling, the precision with which estimate of pp values can be made obtainable from the Sampling value. ~~this is to a~~ this is a very important advantage since random Sampling ~~technologies~~ techniques provide and objective basis for measuring errors due to the Sampling process and for stating the degree of confidence to be placed upon estimate of pp values.

SAMPLE RANDOM SAMPLING: This is one of the methods of Sampling. It has been that a random Sampling prob. Sample is a simple drawn in such a way that the prob. of inclusion of every elements in the population is known. A ~~Sa~~ 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 ~~of~~ in which each of the 'N' pp element has an equal prob '1/N' of being the one selected, the first drawn, '1/N' of being the one selected, the 2nd one drawn and so on. Until the nth sample has been drawn since there are $^{N}C_{n}$ possible samples of 'n' the prob. that any sample of size 'n' will be drawn $1/{}^{N}C_{n}$ E.g Let $N = 5$; $n = 2$ the possible samples will be $^{5}C_{2} = 10$. Supposing the 5 nos are labled. Lets take 2 possible samples we have.

METHOD OF SIMPLE RANDOM SAMPLING.

1) Drawing Chips from a bowl: If attention is restricted to the most straight foward situation in which the elements are easily identified and can be number. For example, suppose there are 100 students in class and we wish to draw a simple random sampling of 20 of these student without repeating assigning nos from 1-100 to each of the student and place these numbers on physically similar slips of paper which could be placed in a bowl shake the bowl to accomplish a thorough mixing of the slips then draw the sample. The slip is drawn and the number of it recorded. Also, the 2nd is drawn and the no of it is recorded and so on. The student corresponding to these 20 numbers constitute the required simple random sampling.

2) Tables of Random Numbers: If the population size is very large, the above procedure can become quite unwieldy and time consuming. It 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 table are useful. Sample as well. A table of random digits is simple table of digit 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 for each of these strata. A sample from each strata may be obtained by simple random sampling of some other forms of probability sampling eg systematic sampling Combined to to yield an over all estimate of a pp parameter or they may be compared with one another to reveal between strata differences. We shall concentrate on the case where the objective is to obtain an overall best estimate of a pp parameter by combing the result of stratification as compared to simple random sampling is to obtain reduction in sampling error or synonymously an increase in precision. I.e as you reduce your sampling error, you are indirectly increasing precision. The reduction in sampling error is element which are more alike with respect to the characteristics under investigation. Stratification is most effective when the elements within strata as homogenous as possible and difference of elements among strata as as great as possible to minimise differences among elements within strata and to maximise differences among strata.

4) Cluster Sampling:- This is a technique in which the population is sub-divided into groups or clusters, then a probability sample of these clusters is

drawn and studied. The primary aim of cluster Sampling is to accomplish cost Samplings in sample design. Sampling error is reduced in cluster Sampling when the units within clusters are as heterogenous as possible. For Instance, If a Single cluster duplicates all of the heterogeneity, which exists in the population, we would only need to draw this 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 no. In many practical applications, Systematic Sampling is used in place of sample random sampling as a method of Obtaining random selection. In a systematic Sample, every kth elements is drawn from the listed or arranged in some Specified manner. The starting point is selected at random from 1st k element. Eg if a researcher has serially arranged list of farmers and he decide of every 5th name on the list where k = 5, then a starting point is selected at random, the 1st 5 names and thus every 5th name selected after the 1st no. k Sometimes is determined by dividing the no. of items in Sampling frame by the Sample Size. If the members of a population occurs in random or then the result of the Systematic Sampling very Similar to those of Simple random Sampling. However Systematic Sampling should be around whenever there is a periodic or cyclical vary 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 a central one throughout all statistical works. Whenever we have measurement inferences or decision making. The possibility of error is present. There are 2 different types of error which may be present in statistical measurement namely

— Systematic error
— Random error

Systematic error cause a measurement to be incorrect in some systematic ways. They are of a type which persist even when the sample size is increased. They are errors involved in the procedures of a statistical investigation and may occur in the planning stages or during or after the collection process. Systematic error is sometimes called BIASED Among the causes of biased are: faulty design of a questionaire such as or refusers by respondent to provide correct information, incomplete sampling frame, mistakes in planning and processing of data and so on. These errors may be veiwed as arising primarily from inaccuracies or deficiencies in the increasing instrument. These errors can be reduced or completely eliminated.

Random or sampling 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 same definition and procedures are used. These samples mean tend to distribute themselves below and above the true population parameter. The difference between the mean of a particular sampling and the pp mean is said to be a random error or a sampling error. The complete collection of factors which could be explained why the sample mean differed from the pp mean is unknown and the errors decreases on the average

for this reason that a larger sample of observation is observed preferred to a smaller one since sample 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 deviation computed from a population is called parameter. The value of such a measure computed from a sample is called STATISTICS. Thus statistics is derived from a sample and parameter is from population value. For statistics, we use greek letters & Roman letters for population.

FUNCTION :=> This is a value by which every member of one set is assigned to or paired with one member of another set. Eg
Suppose x is a set of students in a seminar and that y is a set of project topics. If 'f' is a rule that assigns to every element x in the set X, a unique element y in the set Y, then 'f' is said to be a function that maps X into Y. On the other hand, a random

Variable is a fn 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 X represent the number of heads 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 "obtain the no. of x in a sample space." and this is a random sample

SAMPLING DISTRIBUTION:

Let us consider how statistic differ from sample to sample. If repeated simple random sampling of the same size are drawn from statistic population. E.g A given statistic such as a mean or proportion will vary from one sample to another. The distribution of means can then be examined to estimate the amount of variation that can be from one sample to another. probability distribution of such a statistic is referred to as a sampling distribution. Thus, we have a sampling distribution of mean, proportion and so on.

SAMPLING DISTRIBUTION OF MEAN: Let f(x) be a prob. distribution of some given pp from which we take a sample of size n. Then, a probability distribution of a statistic $\bar{x}$ is called the sampling distribution of means. We use the following theorems to explain this: