 (9th sept, 2020)

SAMPLE SURVEYS

sample survey

(Def): If the data is ~~eat~~ collected from the part of the population i.e., sample it is known as sample survey.

Eg: Survey of BSC (MSCS) students, survey of higher income group etc.

Principal steps in a sample survey:

1) Objectives of the survey: The first step is to define in clear terms the objectives of the survey. These objectives should be proportional in terms of time, money and manpower.

2) Defining the population to be sampled: The population from which the sample is drawn should be defined clearly.

3) The frame & sampling units: The population is divided into divisions know as sampling units. These units should cover the entire population & must be distinct. The frame is the list (or) map which is used for covering the population. The frame should be upto date.

4) Data to be collected: The data should be collected keeping in view the objectives of the survey. we should not collect extra data & essential data should not be omitted.

5) The questionnaire (or) schedule: The questionnaire contains set of questions which should be ~~fin~~ filled by the respondent. The schedule contains set of questions which should be filled by the interviewer. The questions should be

clear & brief instructions should be given for filling up the questionnaire (or) schedule

(6) Method of collecting Information:

i) Interview Method: In this method, the investigator goes from house to house & collects the information & fills up the schedule. This method is costly but more accurate

ii) Mailed Questionnaired Method: In this method, the questionnaire is mailed to the individuals who will fill it & send it to the required address. This method is cheap but not accurate.

iii) (7) Non-respondents: Sometimes the data cannot be collected for all the sampled units.
Eg: The selected people may not be available at his place or even refuse to give certain information. This is called non-response. Such cases should be handled with care.

(8) Selection of proper sampling Design: A no. of designs are available for the selection of sample. The costs & time should be considered before making a final selection.

(9) Organisation of field work: The investigator should be trained in selecting the sample units, recording the measurements before starting the field work. Supervisory staff should be there for inspection after field work.

(10) Pre-test: It helps in asking the questions effectively & ~~improve the~~ improvement in the questionnaire.

(11) Summary & Analysis of the data:

(a) Scrutiny & editing of the data: The schedule should be thoroughly scrutinised & edited to avoid duplication errors.

(b) Tabulation of data: The data can be tabulated by using hand and machine tabulation. For large scale survey machine tabulation is faster & cheap.

(c) Statistical Analysis: Different methods are available for the analysis of the data. It should be free from errors.

(d) Reporting & conclusions: Finally a detailed report is prepared & the final conclusions are drawn.

(12) Information gained for future surveys:
Any completed survey is helpful in taking lessons from it for designing future surveys. The mistakes committed can be rectified and efficiency can be increased.

(10th sept, 2020)

Principles of sample survey:

1. (a) Principle of Statistical Regularity
(b) Principle of Inertia of large Numbers

(a) It states that large no. of items selected at random from a large group will have the characteristics of the large group. If a sample is selected at random, it will satisfy this principle.

(b) It states that other things being equal, as the sample size increases the results will be more accurate.

Ex: If n is large we get approximately 50% heads & 50% tails.

2. Principle of validity: It means it should enable us to obtain valid tests & estimates about the parameters of the population. The sample obtained by probability sampling satisfy this principle.

3. Principle of Optimisation: It states that getting optimum results in terms of efficiency & cost of the design. It consists in

i) achieving a given level of efficiency at minimum cost

ii) obtaining maximum possible efficiency with given level of cost.

Sampling & Non-Sampling Errors:

i) Sampling Errors: These are present in sample data & these are absent in population data.

These are due to following reasons:

(1) Faulty selection of the sample: If the sample is selected by defective method

Ex: Purposive sampling (or) Judgement sampling we get sampling error.

This can be reduced by selecting sample using random method.

2) Substitution: If difficulties arise in selecting a particular sampling unit, the investigator substitutes a population unit. We get sampling error since the properties of a sample are different from population.

3) Faulty demarcation of sampling units: If the border lines are not properly drawn in agricultural experiments we get sampling error.

4) Constant error due to improper choice of the statistics for estimating the population parameters:

If we want to estimate population variance σ^2 we have to use $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ [unbiased] instead of

$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ [biased]

ii) Non-sampling Errors: These are present in both sample data & population data.

These are due to following reasons;

1) Faulty planning (or) Definitions: Data is not sufficient w.r.t objectives of the survey, Errors in recording the measurements, lack of trained & qualified investigators

2) Response Errors: These are due to following reasons:

i) Response errors may be accidental.

ii) Response errors may be due to prestige bias

iii) Response errors may be due to self interest.

iv) Response errors may be due to bias of interviewer

v) Response errors may be due to failure of respondents memory.

3) Non-response Bias: In this case the respondent may not be at his place or even refuse to give certain information

4) Errors in coverage: If the objectives of the survey are not stated clearly we get error since some items are included which are not to be included & some are deleted which are in the survey.

5) Compiling Errors: If the data are not properly edited & coded we get non sampling error.

6) Publication Errors: If the data are not properly proof readed we get non sampling error.

(11th sept, 2020)

Advantages ~~& disa'~~ of sampling over complete census:

① Less time: There is saving in time & labour since only a part of the population has to be examied and analysed.

② Reduced cost of the survey: The cost of the sample survey is very less as compared to population survey since only a part of the population has to be examined & analysed.

③ Greater Accuracy of Results: The results of a sample survey are more accurate than population survey since in case of a sample we can use latest equipment and trained people for its analysis.

④ Greater scope: sample survey has generally greater scope as compared with population census, since sample survey saves in time and money & it is possible to get more information from a small group of people.

⑤ If the population is too large
Ex: Trees in a jungle we have to apply sampling method only.

⑥ If testing is destructive
Ex: testing of crackers, electric bulbs etc.
In such cases we have to appling sampling method only.

⑦ If the population is hypothetical
Ex: Tossing of coin, etc; we have to apply sampling method only.

Limitations of sampling:

① Proper care should be taken in the planning & execution of sample survey otherwise the results will be in accurate.

② sampling theory require the services of trained & qualified people & latest equipment for its planning, execution & analysis. In the absence of these the results are inaccurate.

Types of sampling

(1) simple random sampling (SRS)
(2) stratified random sampling (STRS)
(3) systematic sampling (SYS)

Purposive ~~(or)~~ Judgement
Probability sampling
Mixed sampling.

* SRS: It is the technique of drawing a sample in such a way that each unit of the population has an equal & independent chance to be in the sample.

(a) SRSWR: If a unit is selected & it is replaced back before drawing the second unit & this procedure is repeated it is known as SRSWR.

Ex: N - total units $P(E_1)=\frac{1}{N}$; $P(E_2)=\frac{1}{N}$; ..... $P(E_r)=\frac{1}{N}$.

(b) SRSWOR: If a unit is selected & it is not replaced back before drawing the second unit & this procedure is repeated it is known as SRSWOR

Ex: N - total units

$P(E_1) = \frac{1}{N}$ Remaining N-1

$P(E_2) = \frac{1}{N-1} = \frac{1}{N-(2-1)}$ Remaing N-2

$P(E_3) = \frac{1}{N-2} = \frac{1}{N-(3-1)}$

$\vdots$

$P(E_r) = \frac{1}{N-(r-1)}$

Th-①

Statement: In SRSWOR, the probability of selecting a specified unit of the population at any given draw is equal to the propability of its being selected at the first draw.

i.e. $P(E_1) = P(E_r) = \frac{1}{N}$ $\left[\begin{array}{l} E_1 = \text{selected at the } 1^{st} \text{ draw} \\ E_r = \text{selected at the } r^{th} \text{ draw} \end{array}\right]$

Proof: Let N = Total no. of population units

Let E_1 = unit selected at the 1^{st} draw

E_2 = Unit selected at the 2^{nd} draw

$\vdots$

E_r = unit selected at the r^{th} draw.

Then $P(E_1) = \frac{1}{N}$, $P(E_2) = \frac{1}{N-1}$

$P(E_r)$ = P[that the specified unit is not selected in any one of the previous $(r-1)$ drawing & then selected at the r^{th} draw]

$= \prod_{i=1}^{r-1} P\{\text{it is not selected at } i^{th} \text{ draw}\} \times P\{\text{it is selected at the } r^{th} \text{ draw given that it is not selected at the previous } (r-1) \text{ draws}\}$ (∵ draws are independent)

By multiplication th of probility. (∵ draws are independent)

$= \prod_{i=1}^{r-1} \left[1 - \frac{1}{N-(i-1)}\right] \times \frac{1}{N-(r-1)}$

$= \prod_{i=1}^{r-1} \left[\frac{N-i}{N-i+1}\right] \times \frac{1}{N-(r-1)}$

$= \frac{\cancel{N-1}}{N} \cdot \frac{\cancel{N-2}}{\cancel{N-1}} \cdot \frac{N-3}{\cancel{N-2}} \cdots \frac{\cancel{N-r+1}}{N-r+2} \times \frac{1}{\cancel{N-r+1}}$

$= \frac{1}{N}$

$\therefore P(E_r) = P(E_1) = \frac{1}{N}$

NOTE: The probability that a specified unit is included in the sample of size n is $= \frac{n}{N}$.

Sampling theory, Time series, Index number & Analysis

*Selection of a SRS:

(a) Lottery method (b) Random Numbers method.

(a) suppose we want to select n students out of N. we assign the numbers 1 to N, one number to each student & we write these numbers on N chits which are of same shape, size, colour etc.

These chits are then put in a bag & thoroughly shuffled & then 'n' chits are drawn one by one. This gives the random sample of size n.

The main draw back is it takes lot of time if the population is large.

(b) Random numbers are the numbers from digits 0, 1, 2, ... 9 which appear with same frequency & independent of each other. If we have to select a sample from a population of size $N (\leq 99)$ then the numbers can be combined two by two to give pairs from 00 to 99. similarly for $N \leq 999$ (or) $N \leq 9999$

It consists of following steps:

i) Identify the N units in the population with the numbers from 1 to N

ii) select at random any page of the random number tables & pick up the numbers in any row (or) column (or) diagonal at random.

iii) The population units corresponding to the numbers selected in step ii) will give the required random sample.

for selecting a samples

Random Number Tables

1. Tippet's Random Numbers Tables
2. fisher & Yates Tables
3. Kendall & Babington Smith's tables
4. Rand corporation tables.

~~Notations~~ & Terminology in SRS:

NOTE: $\left(\sum_{i=1}^{3}(x_i)\right)^2 = x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_2x_3 + 2x_3x_1$

$$= \sum_{i=1}^{3} x_i^2 + 2\sum_{i \neq j = 1} x_i x_j$$

***

$$\left[\sum_{i=1}^{n}(x_i)\right]^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i \neq j = 1}^{n} x_i x_j$$

Notations & Terminology in SRS

N = Population size ; n = sample size

$Y_1, Y_2 \ldots Y_N \rightarrow$ population values

$y_1, y_2 \ldots y_n \rightarrow$ sample values

Population Mean $\bar{Y}_N = \dfrac{Y_1 + Y_2 + \ldots + Y_N}{N} = \sum_{i=1}^{N} \dfrac{Y_i}{N}$ $\quad \sum Y_N$

Sample mean $\bar{y}_n = \dfrac{y_1 + y_2 + \ldots + y_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$

$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{N} a_i y_i$

where $a_i = \begin{cases} 1, & \text{if } i\text{th unit is included in the sample} \\ 0, & \text{if } i\text{th unit is not included in the sample.} \end{cases}$

Population Mean square $= S^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N}\left(Y_i^2 + \bar{Y}_N^2 - 2Y_i\bar{Y}_N\right)\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 + \sum_{i=1}^{N}\bar{Y}_N^2 - 2\bar{Y}_N \sum_{i=1}^{N} Y_i\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 + N\bar{Y}_N^2 - 2\bar{Y}_N(N\bar{Y}_N)\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 + N\bar{Y}_N^2 - 2N\bar{Y}_N^2\right]$$

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\bar{Y}_N^2\right]$$

sample mean square $= s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\left(y_i^2 + \bar{y}_n^2 - 2y_i\bar{y}_n\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n}\bar{y}_n^2 - 2\bar{y}_n \sum_{i=1}^{N} y_i\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 + n\bar{y}_n^2 - 2\bar{y}_n(n\bar{y}_n)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 + n\bar{y}_n^2 - 2n\bar{y}_n^2\right]$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2\right]$$

Th ② statement: In SRSWOR, the sample mean is an unbiased estimate of the population mean i.e., $E(\bar{y}_n) = \bar{Y}_N$

Proof: $\bar{y}_n = \frac{1}{n}\sum_{i=1}^{N} a_i y_i$

$$E(\bar{y}_n) = E\left[\frac{1}{n}\sum_{i=1}^{N} a_i y_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} E(a_i)\, y_i \text{ —— ①}$$

a_i : 1   0

$P(a_i)$: $P(a_i=1)$ $P(a_i=0)$

$$\therefore E(a_i) = 1\times P(a_i=1) + 0\times P(a_i=0)$$

$$= 1\cdot P\left(\begin{matrix}i^{th} \text{ unit is included} \\ \text{in a sample of size } n\end{matrix}\right) + 0\cdot P\left(\begin{matrix}i^{th} \text{ unit is not included} \\ \text{in a sample of size } n\end{matrix}\right)$$

$$= 1\times\frac{n}{N} + 0\left(1-\frac{n}{N}\right)$$

$$= \frac{n}{N} \text{ —— ②}$$

Substituting ② in ①,

we get $E(\bar{y}_n) = \frac{1}{n}\sum_{i=1}^{N}\frac{n}{N}Y_i$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i$$

$$= \bar{Y}_N$$

$$\therefore \boxed{E(\bar{y}_n) = \bar{Y}_N}$$

Th-③ ** Imp

Statement: In SRSWOR, sample mean square is an unbiased estimate of the population mean square, i.e, $E(s^2) = S^2$

Proof: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left\{\frac{1}{n}\sum_{i=1}^{n} y_i\right\}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left\{\sum_{i=1}^{n} y_i^2 + \sum_{i\neq j=1}^{n} y_i y_j\right\}\right]$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} y_i^2\left\{1-\frac{1}{n}\right\} - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j \text{ —— ①}$$

$$E(s^2) = E\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{N} a_i y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{N} a_i a_j Y_i Y_j\right]$$

$$\left[\because \bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{N} a_i Y_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} E(a_i)Y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{N} E(a_i a_j)Y_i Y_j \text{ —— ②}$$

But $E(a_i) = \frac{n}{N}$ —— ③

$$E(a_i a_j) = 1\cdot P(a_i a_j = 1) + 0\cdot P(a_i a_j = 0) \quad \left[\because a_i a_j \text{ takes values 1 \& 0}\right]$$

$$= P(a_i a_j = 1)$$

$$= P[(a_i = 1)\cap(a_j = 1)] \quad \because \boxed{P(A\cap B) = P(A)\cdot P(B|A)}$$

$$= P(a_i=1)\, P(a_j=1/a_i=1)$$

$$= \frac{n}{N}\,\frac{n-1}{N-1} \quad \text{—— (4)}$$

$\because P(a_j=1/a_i=1) = P$ (j^{th} unit is included in the sample given that i^{th} unit is included in the sample)

$$= \frac{n-1}{N-1}$$

Substituting (3) & (4) in (2)

we get

$$E(s^2) = \frac{1}{n}\sum_{i=1}^{N}\frac{n}{N}y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{N}\frac{n(n-1)}{N(N-1)}y_iy_j$$

$$= \frac{1}{N}\sum_{i=1}^{N}y_i^2 - \frac{1}{N(N-1)}\sum_{i\neq j=1}^{N}y_iy_j$$

$$= S^2 \qquad \text{(using (1))}$$

$$\therefore \boxed{E(s^2) = S^2}$$

Th-(4)

Statement: In SRSWOR, the variance of the sample mean is given by $V(\bar{y}_n) = \frac{N-n}{Nn}S^2$

Proof: $V(\bar{y}_n) = E(\bar{y}_n^2) - E(\bar{y}_n)^2$

$$= E(\bar{y}_n^2) - \bar{Y}_N^2 \quad \text{—— (1)}$$

Consider $E(\bar{y}_n^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}y_i\right]^2$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n}y_i^2 + \sum_{i\neq j=1}^{n}y_iy_j\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{N}a_iy_i^2 + \sum_{i\neq j=1}^{N}a_ia_jy_iy_j\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N}E(a_i)y_i^2 + \sum_{i\neq j=1}^{N}E(a_ia_j)y_iy_j\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N}\frac{n}{N}y_i^2 + \sum_{i\neq j=1}^{N}\frac{n(n-1)}{N(N-1)}y_iy_j\right]$$

$$= \frac{1}{nN}\sum_{i=1}^{N}y_i^2 + \frac{n-1}{nN(N-1)}\sum_{i\neq j=1}^{N}y_iy_j \quad \text{—— (2)}$$

we know that

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y}_N)^2$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N}y_i^2 - N\bar{Y}_N^2\right]$$

$$\Rightarrow (N-1)S^2 = \sum_{i=1}^{N}y_i^2 - N\bar{Y}_N^2$$

$$\therefore \sum_{i=1}^{N}y_i^2 = (N-1)S^2 + N\bar{Y}_N^2 \quad \text{—— (3)}$$

Also w.k.t.

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}y_i^2 - \frac{1}{N(N-1)}\sum_{i\neq j=1}^{N}y_iy_j \qquad \text{(from th (3))}$$

$$\Rightarrow N(N-1)S^2 = (N-1)\sum_{i=1}^{N}y_i^2 - \sum_{i\neq j=1}^{N}y_iy_j$$

$$\Rightarrow \sum_{i\neq j=1}^{N}y_iy_j = (N-1)\sum_{i=1}^{N}y_i^2 - N(N-1)S^2$$

$$= (N-1)\left[(N-1)S^2 + N\bar{Y}_N^2\right] - N(N-1)S^2 \qquad \text{(from (3))}$$

$$\sum_{i\neq j=1}^{N} = (N-1)^2S^2 + N(N-1)\bar{Y}_N^2 - N(N-1)S^2 \quad \text{—— (4)}$$

Sub (3) & (4) in (2)

we get.

$$E(\bar{y}_n^2) = \frac{1}{nN}\left[(N-1)S^2 + N\bar{Y}_N^2\right] + \frac{n-1}{nN(N-1)}\left[(N-1)^2S^2 + N(N-1)\bar{Y}_N^2 - N(N-1)S^2\right]$$

$$V(\bar{y}_n^2) = {}^2\frac{1}{nN}\left[(N-1)S^2+N\bar{y}_N^2\right] + \frac{n-1}{nN(N-1)}\left[\begin{array}{c}(N-1)^2S^2+N(N-1)\bar{y}_N^2\\ -N(N-1)S^2\end{array}\right] - \bar{y}_N^2$$

$$= \frac{(N-1)S^2}{nN} + \frac{N\bar{y}_N^2}{nN} + \frac{(n-1)(N-1)^2S^2}{nN(N-1)} + \frac{(n-1)N(N-1)\bar{y}_N^2}{nN(N-1)}$$

$$- \frac{(n-1)(N-1)NS^2}{nN(N-1)} - \bar{y}_N^2$$

$$= \frac{\begin{array}{c}(N-1)^2S^2 + N(N-1)\bar{y}_N^2 + (n-1)(N-1)^2S^2 + (n-1)N(N-1)\bar{y}_N^2 - nN(N-1)\bar{y}_N^2\\ -(n-1)(N-1)NS^2\end{array}}{nN(N-1)}$$

$$= \frac{\cancel{N-1}}{nN(\cancel{N-1})}\left[\begin{array}{c}(N-1)S^2+N\bar{y}_N^2+(n-1)(N-1)S^2+(n-1)N\bar{y}_N^2-(n-1)NS^2\\ -nN\bar{y}_N^2\end{array}\right]$$

$$= \frac{1}{nN}\left[\begin{array}{c}\cancel{NS^2}-\cancel{S^2}+\cancel{N\bar{y}_N^2}+\cancel{nNS^2}-nS^2-\cancel{NS^2}+\cancel{S^2}+\cancel{nN\bar{y}_N^2}-\cancel{N\bar{y}_N^2}\\ -\cancel{nNS^2}+NS^2-\cancel{nN\bar{y}_N^2}\end{array}\right]$$

$$= \frac{1}{nN}\left[NS^2-nS^2\right] \Rightarrow \frac{N-n}{nN}S^2$$

$$\boxed{V(\bar{y}_n^2) = \frac{N-n}{nN}S^2}$$

19th Sept, 2020 (21st)

NOTE: sampling fraction

$f = \frac{n}{N}$ = sampling fraction

$$V(\bar{y}_n) = \frac{N-n}{nN} S^2$$

$$= \frac{S^2}{n}\left[\frac{N-n}{N}\right]$$

$$= \frac{S^2}{n}\left[1-\frac{n}{N}\right]$$

$$\boxed{V(\bar{y}_n) = \frac{S^2}{n}[1-f]}$$

$1-f$ is called finite population correction (FPC)

If N is large i.e. n is small then $f \to 0$

then $1-f \to 1$.

Th-5

St : In SRSWR the variance of the sample mean is given by

$$V(\bar{y}_n) = \left[\frac{N-1}{Nn}\right] S^2$$

Proof : Let, $y_1, y_2, \ldots, y_n$ be a random sample of size n from a population with common pop variance σ^2.

then, $V(\bar{y}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$

$$= \frac{1}{n^2}\left(V\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2}\left(V(y_1) + V(y_2) + \ldots + V(y_n)\right)$$

$$= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \ldots \sigma^2\right)$$

Since, the ~~$\frac{1}{n^2}(n\sigma$~~ draws are independent with common variance σ^2

$$V(\bar{y}_n) = \frac{1}{n^2}(n\sigma^2)$$

$$V(\bar{y}_n) = \frac{\sigma^2}{n}$$

But $\sigma^2 = \frac{1}{N}\left[\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2\right]$

Also, $S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2\right]$

$\Rightarrow N\sigma^2 = (N-1)S^2$

$\Rightarrow \sigma^2 = \frac{(N-1)}{N} S^2$

$\therefore V(\bar{y}_n) = \frac{1}{n}\left(\frac{N-1}{N} S^2\right)$

$$\boxed{\therefore V(\bar{y}_n)_{SRSWR} = \frac{N-1}{Nn} S^2}$$

$$\boxed{\therefore V(\bar{y}_n)_{SRSWR} = \frac{N-1}{Nn} S^2}$$

compaurison of SRSWOR (VS) SRWR

$$V(\bar{y}_n)_{SRSWOR} = \frac{N-n}{Nn} S^2$$

00-10

$$V(\bar{y}_n)_{SRSWR} = \frac{N-1}{Nn} S^2$$

$\because V(\bar{y}_n)_{SRSWOR} < V(\bar{y}_n)_{SRSWR}$

$\therefore$ SRSWOR is more efficient than SRSWR

Merits & Demerits of SRS

Merits:

① Since the sample units are selected at random giving equal preference, the personal bias (error) is completely eliminated

② SRS is more representative of the population as compared to the judgement (or) purposive sampling.

③ Sample mean ($\bar{y}_n$) as an estimate of population mean ($\bar{Y}_n$) becomes more efficient are sample size n increases.

Demerits:

① The selection of SRS requires an upto date frame (list) which is not available & this restricts the use of SRS technique

② Administrative Inconvenience: The sampling units in SRS are widely spread geographically & in such a case the cost of collecting data is more in terms of time & money.

③ Sometimes SRS gives non-random results

Ex: If we draw a random sample of size 13 from a pack of cards, we may get all the cards of the same suit.

④ SRS requires larger sample as compared to STRS (stratified Random sampling) for a given precision.

22nd sept, 2020

Non-Probability, Probability & mixed sampling.

① Non-probability sampling

*Subjective (or) ~~pur~~ purposive (or) judgement sampling

In this method ~~under~~ the sample is selected with definite purpose in view & the choice of the sampling units depends on the judgement of the ~~unit~~ investigator. This method suffers from a draw of favouritism & doesn't represents the population.

Ex: If the investigator wants to show the standard of living of hyderabad is increased, he collects data from hitech city, banjara hills & ignore low lying & slum areas.

② Probability sampling:

It is the scientific method of selecting samples according to some laws of chance in which each unit in the population has some definite pre-assigned probability of being selected in the sample.

The different types of probability sampling are

ⓐ where each unit has an equal chance of being selected.

ⓑ sampling units have diff probabilities of being selected.

ⓒ The probability of selection of a unit is ~~propo~~ propotional to the sample size.

③ Mixed sampling:

If the samples are selected paitly according to some laws of chance & partly according to a fixed sampling rule they are known as mixed samples & the technique is known as mixed sampling

23rd sept, 2020

problem

① A population consits of 6 values 1, 2, 3, 4, 5, 6 draw all random samples of size 2, using SRSWR & SRSWOR Also verify.

(a) SRSWR

$$E(\bar{y}_n) = \bar{Y}_N \; ; \; E(s^2) = \sigma^2 \; ; \; V(\bar{y}_n) = \frac{N-1}{Nn} S^2$$

(b) SRSWOR

$$E(\bar{y}_n) = \bar{Y}_N \; ; \; E(s^2) = S^2 \; ; \; V(\bar{y}_n) = \frac{N-n}{Nn} S^2$$

Sol $N = 6, n = 2$

$$\bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y}_N)^2 = \frac{1}{6-1}\left((1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2\right)$$

$$= \frac{1}{5}\left[6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25\right]$$

$$= \frac{17.5}{5}$$

$$= 3.5$$

$S^2 = 3.5$

$$\sigma^2 = \frac{1}{N}\Sigma(y_i - \bar{Y}_N)^2 = \frac{1}{6}(17.5) = 2.9166.$$

(a) SRSWR

no. of samples $= N^n = 6^2 = 36$

sample no.	sample values	$\bar{y}_n$	$(\bar{y}_n - \bar{Y}_N)^2$	s^2	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$
1	(1,1)	1	6.25	0	$=1((1-1)^2+(1-1)^2=0$
2	(1,2)	1.5	4	0.5	0.5
3	(1,3)	2	2.25	2	~~2.25~~ 2
4	(1,4)	2.5	1	4.5	4.5
5	(1,5)	3	0.25	8	~~6.25~~ 8
6	(1,6)	3.5	0	12.5	~~30.25~~ 12.5
7	(2,1)	1.5	4	0.5	~~36~~ 0.5
8	(2,2)	2	2.25	0	0
9	(2,3)	2.5	1	0.5	0.5
10	(2,4)	3	0.25	2	2
11	(2,5)	3.5	0	4.5	4.5
12	(2,6)	4	0.25	8	8
13	(3,1)	2	2.25	2	2
14	(3,2)	2.5	1	0.5	0.5
15	(3,3)	3	0.25	0	0
16	(3,4)	3.5	0	0.5	1
17	(3,5)	4	0.25	2	2
18	(3,6)	4.5	1	4.5	4.5
19	(4,1)	2.5	1	4.5	4.5
20	(4,2)	3	0.25	2	2
21	(4,3)	3.5	0	0.5	0.5
22	(4,4)	4	0.25	0	0

S.No.	S.values	$\bar{y}_n$	$(\bar{y}_n - \bar{Y}_N)^2$	s^2	$s^2 = \frac{1}{n-1}\Sigma(y_i - \bar{y}_n)^2$
23	(4,5)	4.5	1	0.5	0.5
24	(4,6)	5	2.25	2	2
25	(5,1)	3	0.25	8	8
26	(5,2)	3.5	0	4.5	4.5
27	(5,3)	4	0.25	2	2
28	(5,4)	4.5	1	0.5	0.5
29	(5,5)	5	2.25	0	0
30	(5,6)	5.5	4	0.5	0.5
31	(6,1)	3.5	0	12.5	12.5
32	(6,2)	4	0.25	8	8
33	(6,3)	4.5	1	4.5	4.5
34	(6,4)	5	2.25	2	2
35	(6,5)	5.5	4	0.5	0.5
36	(6,6)	6	6.25	0	0
		126	52.5	105	105

verification

$$E(\bar{y}_n) = \frac{126}{36} = 3.5 = \bar{Y}_N$$

$$E(s^2) = \frac{105}{36} = 2.916 = \sigma^2$$

$$V(\bar{y}_n) = \frac{52.5}{36} = 1.458$$

$$\left[\frac{N-1}{Nn}\right] s^2 = \left[\frac{6-1}{6(2)}\right] 3.5 = \left[\frac{5}{12}\right](3.5) = 1.458$$

In SRSWOR $E(\bar{y}_n) = \bar{Y}_N$; $E(s^2) = \sigma^2$; $V(\bar{y}_n) = \left[\frac{N-1}{nN}\right] s^2$

(b) SRSWOR

No. of samples are $N_{C_n} = 6C_2 = 15$

S.No.	S values	$\bar{y}_n$	$(\bar{y}_n - \bar{Y}_N)^2$	s^2
1	(1,2)	1.5	4	0.5
2	(1,3)	2	2.25	2
3	(1,4)	2.5	1	4.5
4	(1,5)	3	0.25	8
5	(1,6)	3.5	0	12.5
6	(2,3)	2.5	1	0.5
7	(2,4)	3	0.25	2
8	(2,5)	3.5	0	4.5
9	(2,6)	4	0.25	8
10	(3,4)	3.5	0	0.5
11	(3,5)	4	0.25	2
12	(3,6)	4.5	1	4.5
13	(4,5)	4.5	1	0.5
14	(4,6)	5	2.25	2
15	(5,6)	5.5	4	0.5
		52.5	~~17.5~~ 17.25	52.5

Verification

* $E(\bar{y}_n) = \frac{52.5}{15} = 3.5 = \bar{Y}_N$

* $E(s^2) = \frac{52.5}{15} = 3.5 = S^2$

* $V(\bar{y}_n) = \frac{17.25}{15} = 1.1\ ;\ \left(\frac{N-n}{Nn}\right)S^2 = \left[\frac{6-2}{6(2)}\right]3.5 = 1.16$

~~25th~~

UNIT II

25th Sept 2020

(11) Stratified Random Sampling

The sampling technique which will reduce S^2 the population heterogeneity is known as stratified sampling. Stratification means division into groups. The population is divided into various groups such that units with in each group are homogeneous & the group means are different. The population consisting of N units is divided into K homogeneous mutually disjoint subgroups (nothing in common) known as strate of sizes $N_1, N_2 \ldots N_K$ such that $N = \sum_{i=1}^{K} N_i$. If a SRS of size n_i $(i=1,2,\ldots K)$ is drawn from each of the stratum respectively such that $n = \sum_{i=1}^{K} n_i$. The sample is known as STRS of size n & the technique is called STRS.

The criterion which enables us to classify various sampling units into diff strate is known as stratifying factor (S.F) Ex: age, income, area.

Advantages of STRS

(1) More representative: In SRS some groups have more representation, some have less representation & some have zero rep. But in STRS all the groups have some representation & it is treated as most efficient system of sampling.

(2) ~~Ex~~ Greater Accuracy: STRS is more accurate than SRS.

(3) Administrative Convenience: As compared with SRS, stratified samples are concentrated more geographically & also time & work involved in collecting ~~da~~ the data is very less.

(4) Suppose a population consists of i) literates & illiterates.

ii) People living in hostels, hospitals.

In such a case we use STRS technique only.

Notations & Terminology:

Let k = no. of strata

N_i = The no. of sampling units in the i^{th} stratum $(i = 1, 2, \dots k)$

$N = \sum_{i=1}^{k} N_i$ = total no. of sampling units in the population

n_i = The no. of sampling units selected with SRSWOR from the i^{th} stratum

$n = \sum_{i=1}^{k} n_i$ = total sample size from all the strata

Let y_{ij} $(i = 1, 2, \dots k;\ j = 1, 2, \dots N_i)$ be the value of the j^{th} unit in the i^{th} stratum.

$$\begin{array}{c|c} \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1N_1} \\ y_{21} & y_{22} & \dots & y_{2N_2} \\ \vdots & & & \\ y_{i1} & y_{i2} & \dots & y_{iN_i} \\ \vdots & \vdots & & \vdots \\ y_{k1} & y_{k2} & \dots & y_{kN_k} \end{pmatrix} & \begin{matrix} N_1 \\ N_2 \\ \vdots \\ N_i \\ \vdots \\ N_k \\ \hline N \end{matrix} \end{array}$$

$\bar{Y}_{N_i}$ = population mean of i^{th} stratum

$$= \frac{y_{i1} + y_{i2} + \dots + y_{iN_i}}{N_i} = \sum_{j=1}^{N_i} y_{ij}/N_i$$

$$\bar{Y}_N = \text{population mean} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{N_i} y_{ij}}{N} = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{Y}_{N_i}$$

$$= \sum_{i=1}^{k} (N_i/N)\bar{Y}_{N_i} = \sum_{i=1}^{k} p_i \bar{Y}_{N_i}$$

where $p_i = N_i/N$ = weight of i^{th} stratum

S_i^2 = population mean square for the i^{th} stratum

$$\boxed{S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(y_{ij} - \bar{Y}_{N_i})^2} \quad (i = 1, 2, \dots k)$$

y_{ij} = value of j^{th} sampled unit from i^{th} stratum

$\bar{y}_{n_i}$ = sample mean of i^{th} stratum

s_i^2 = sample mean square of the i^{th} stratum

$$= \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{n_i})^2 \quad (i = 1, 2, \dots k)$$

Estimate of $\bar{Y}_N$ ***

$$\boxed{\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{y}_{n_i} = \sum_{i=1}^{k} p_i \bar{y}_{n_i}}$$

Th 1 **statement:** $\bar{y}_{st}$ is an unbiased estimate of the population mean $\bar{Y}_N$ i.e., $E(\bar{y}_{st}) = \bar{Y}_N$

proof: $\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{y}_{n_i}$

$$E(\bar{y}_{st}) = E\left[\frac{1}{N}\sum_{i=1}^{k} N_i \bar{y}_{n_i}\right] = \frac{1}{N}\sum_{i=1}^{k} N_i E(\bar{y}_{n_i}) \quad \text{--- (1)}$$

since the sample in each of the stratum is a SRS

we have $E(\bar{y}_n) = \bar{Y}_N$

$$\Rightarrow E(\bar{y}_{n_i}) = \bar{Y}_{N_i} \quad \text{--- (2)}$$

Substituting (2) in (1)

we get, $E(\bar{y}_{st}) = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{Y}_{N_i} = \bar{Y}_N$

$$\boxed{\therefore E(\bar{y}_{st}) = \bar{Y}_N}$$

Th ②

Statement : $V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i(N_i-n_i)\frac{S_i^2}{n_i}$

$= \sum_{i=1}^{k} P_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2$

Proof: $\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{y}_{n_i}$

$V(\bar{y}_{st}) = V\left[\frac{1}{N}\sum_{i=1}^{k} N_i \bar{y}_{n_i}\right] = \frac{1}{N^2}\sum_{i=1}^{k} N_i^2 V(\bar{y}_{n_i})$ —— ①

Since the sample in each of the stratum is SRS,

we have

$V(\bar{y}_n) = \frac{N-n}{Nn} S^2$

$\Rightarrow V(\bar{y}_{n_i}) = \frac{N_i-n_i}{N_i n_i} S_i^2$ —— ②

Sub ② in ①, we get

$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i^2 \frac{(N_i-n_i)}{N_i n_i} S_i^2$

$= \frac{1}{N^2}\sum_{i=1}^{k} N_i(N_i-n_i)\frac{S_i^2}{n_i}$

(or) $V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k}\frac{N_i^2}{N_i}\left(\frac{N_i-n_i}{n_i}\right)S_i^2$

$= \sum_{i=1}^{k}\left(\frac{N_i^2}{N^2}\right)\left[\frac{N_i-n_i}{N_i n_i}\right]S_i^2$

$= \sum_{i=1}^{k} P_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2$

Allocation of sample size in STRS

Allocation of the sample sizes for different strata is done in two ways

(a) Proportional allocation

(b) Optimum allocation

(a) **Proportional Allocation**: Allocation of n_i's to various strata is called proportional if the sample fraction is constant for each stratum i.e.,

$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_i}{N_i} = \dots = \frac{n_k}{N_k} = \frac{\sum n_i}{N_i} = \frac{n}{N} = C$

$\Rightarrow \frac{n_i}{N_i} = C = \frac{n}{N}$

$\Rightarrow n_i = \left(\frac{n}{N}\right)N_i$

$\Rightarrow n_i \propto N_i \quad (i=1,2,\dots k)$

$V(\bar{y}_{st})$ Under proportional allocation:

$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i(N_i-n_i)\frac{S_i^2}{n_i}$

$= \frac{1}{N^2}\sum_{i=1}^{k}\left(\frac{N_i^2}{n_i}-N_i\right)S_i^2$

$= \frac{1}{N^2}\left[\sum_{i=1}^{k}\frac{N_i^2 S_i^2}{n_i} - \sum_{i=1}^{k} N_i S_i^2\right]$

$= \frac{1}{N^2}\sum_{i=1}^{k}\frac{N_i^2 S_i^2}{n_i} - \frac{1}{N^2}\sum_{i=1}^{k} N_i S_i^2$

$= \frac{1}{N^2}\sum_{i=1}^{k}\frac{N_i^2 S_i^2}{\frac{n}{N}N_i} - \frac{1}{N}\sum_{i=1}^{k}\frac{N_i}{N}S_i^2 \quad \left[\because n_i = \frac{n}{N}N_i\right]$

$= \frac{1}{n}\sum_{i=1}^{k}\frac{N_i}{N}S_i^2 - \frac{1}{N}\sum_{i=1}^{k}\frac{N_i}{N}S_i^2$

$$= \frac{1}{n}\sum_{i=1}^{K} p_i S_i^2 - \frac{1}{N}\sum_{i=1}^{K} p_i S_i^2 \quad (\because p_i = N_i/N)$$

$$V(\bar{y}_{st})_{prop} = \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{K} p_i S_i^2$$

$$\therefore V(\bar{y}_{st})_p = \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{K} p_i S_i^2$$

Max or min

(b) **Optimum Allocation**: Allocation of n_i's to various strata is called optimum if

i) $V(\bar{y}_{st})$ is minimum for fixed n (efficiency will be more)

ii) $V(\bar{y}_{st})$ is minimum for fixed cost C

iii) Total cost C is minimum for fixed value of $V(\bar{y}_{st})$

Cost function: It is given by

$$C = a + \sum_{i=1}^{K} C_i n_i$$

where a is the over head cost & C_i is the cost per unit in the i^{th} stratum.

Th-③ ***

Statement: $V(\bar{y}_{st})$ is minimum for fixed total size of the sample (n) if $n_i \propto N_i S_i$

Proof: $V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{K} N_i(N_i-n_i)\frac{S_i^2}{n_i}$

The problem is to minimise $V(\bar{y}_{st})$ subject to the condition $\sum_{i=1}^{K} n_i = n$ (fixed).

*** This is same as minimising

$$\phi = V(\bar{y}_{st}) + \lambda\left[\sum_{i=1}^{K} n_i - n\right]$$

where λ is un-know as Lagranges multiplier

$$\phi = \frac{1}{N^2}\sum_{i=1}^{K} N_i(N_i-n_i)\frac{S_i^2}{n_i} + \lambda\left[\sum_{i=1}^{K} n_i - n\right]$$

from the principle of maxima & minima

$f'=0$ & $f''>0$

By Partial Diff

$\frac{\partial f}{\partial x}$

$$\frac{\partial \phi}{\partial n_i} = 0 \quad \& \quad \frac{\partial^2 \phi}{\partial n_i^2} > 0$$

$$\phi = \frac{1}{N^2}\sum_{i=1}^{K}\left(\frac{N_i^2 S_i^2}{n_i} - N_i S_i^2\right) + \lambda\left[\sum_{i=1}^{K} n_i - n\right]$$

$$= \frac{1}{N^2}\sum_{i=1}^{K}\frac{N_i^2 S_i^2}{n_i} - \frac{1}{N^2}\sum_{i=1}^{K} N_i S_i^2 + \lambda\left[\sum_{i=1}^{K} n_i - n\right]$$

$$\frac{\partial \phi}{\partial n_i} = 0 = \frac{1}{N^2}\left[N_i^2 S_i^2\right]\left[\frac{-1}{n_i^2}\right] + \lambda(1)$$

$$\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}$$

$$= 0 = \frac{-N_i^2 S_i^2}{N^2 n_i^2} + \lambda$$

Not required will be >0 only

$$\text{Also } \frac{\partial^2 \phi}{\partial n_i^2} = -\frac{N_i^2 S_i^2}{N^2}\left[-2 n_i^{-3}\right]$$

$$= \frac{2 N_i^2 S_i^2}{N^2 n_i^3} > 0$$

$$\Rightarrow \frac{N_i^2 S_i^2}{N^2 n_i^2} = \lambda \Rightarrow n_i^2 = \frac{N_i^2 S_i^2}{N^2 \lambda} \Rightarrow \boxed{n_i = \frac{N_i S_i}{N\sqrt{\lambda}}} \quad \text{— ①}$$

To find λ, we sum. Eq^n ① over i from 1 to K, we get

$$\sum_{i=1}^{K} n_i = \frac{1}{N\sqrt{\lambda}}\sum_{i=1}^{K} N_i S_i$$

$$\Rightarrow \sqrt{\lambda} = \frac{1}{Nn}\sum_{i=1}^{K} N_i S_i \quad (\because \sum n_i = n)$$

Sub λ in Eq^n ①, we get

$$n_i = \frac{N_i S_i}{\cancel{N}\frac{1}{\cancel{N}n}\sum N_i S_i} \Rightarrow \frac{n}{\sum N_i S_i} N_i S_i$$

$$\Rightarrow n_i = C\, N_i S_i \quad \left[\because C = \frac{n}{\sum N_i S_i} = \text{const}\right]$$

$$\Rightarrow \boxed{n_i \propto N_i S_i}$$

This is known as Neyman's formula for optimum allocation greater the value of $N_i S_i \Rightarrow$ greater sample size (n_i)

$V(\bar{y}_{st})$ under optimum allocation:

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i(N_i-n_i)\frac{S_i^2}{n_i}$$

$$= \frac{1}{N^2}\sum_{i=1}^{k} N_i\left(\frac{N_i}{n_i}-\frac{n_i}{n_i}\right)S_i^2$$

$$= \frac{1}{N^2}\sum_{i=1}^{k} N_i\left(\frac{N_i}{n_i}-1\right)S_i^2$$

$$= \frac{1}{N^2}\sum_{i=1}^{k} N_i\left[\frac{N_i}{\frac{nN_iS_i}{\Sigma N_iS_i}}-1\right]S_i^2 \quad \left(\because n_i = \frac{nN_iS_i}{\Sigma N_iS_i}\right) \text{ from (iii)}$$

$$= \frac{1}{N^2}\sum_{i=1}^{k} N_i\left\{\frac{\Sigma N_iS_i}{nS_i}-1\right\}S_i^2$$

$$= \frac{1}{N^2}\sum_{i=1}^{k}\left\{\frac{N_i}{n\cancel{S_i}}\cancel{S_i^2}\,\Sigma N_iS_i - N_iS_i^2\right\}$$

$$= \frac{1}{N^2}\frac{\left(\sum_{i=1}^{k} N_iS_i\right)^2}{n} - \frac{1}{N^2}\sum_{i=1}^{k} N_iS_i^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{k} p_iS_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} p_iS_i^2 \quad \left(\because p_i = \frac{N_i}{N}\right)$$

$$\therefore V(\bar{y}_{st})_{opt} = \frac{1}{n}\left(\sum_{i=1}^{k} p_iS_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} p_iS_i^2$$

Th-4 — 3rd Oct, 2020

Statement: In STRS with given cost function $C = a + \sum_{i=1}^{k} c_i \cdot n_i$, $V(\bar{y}_{st})$ is minimum if $n_i \propto \frac{N_iS_i}{\sqrt{C_i}}$

Proof: $V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i(N_i-n_i)\frac{S_i^2}{n_i}$

The problem is to minimise $V(\bar{y}_{st})$ subject to the condition $C = a + \sum_{i=1}^{k} c_in_i$ (fixed)

This is same as minimising

$$\psi = V(\bar{y}_{st}) + \lambda\left(\sum_{i=1}^{k} c_in_i + a - C\right)$$

where λ is unknown Legrangee's multiplyer.

$$\psi = \left[\frac{1}{N^2}\sum_{i=1}^{k} N_i(N_i-n_i)\frac{S_i^2}{n_i}\right] + \lambda\left(\sum_{i=1}^{k} c_in_i + a - C\right)$$

~~From the principle of maxima & mini~~

$$= \frac{1}{N^2}\sum_{i=1}^{k}\left[\frac{N_i^2S_i^2}{n_i} - \frac{N_i\cancel{n_i}S_i^2}{\cancel{n_i}}\right] + \lambda\left(\sum_{i=1}^{k} c_in_i + a - C\right)$$

$$\psi = \frac{1}{N^2}\sum_{i=1}^{k}\frac{N_i^2S_i^2}{n_i} - \frac{1}{N^2}\sum_{i=1}^{k} N_iS_i^2 + \lambda\left[\sum_{i=1}^{k} c_in_i + a - C\right]$$

from the principle of maxima & minima

$$\frac{\partial\psi}{\partial n_i} = 0 \quad \& \quad \frac{\partial^2\psi}{\partial n_i^2} > 0$$

$$\frac{\partial\psi}{\partial n_i} = 0 = \frac{1}{N^2}(N_i^2S_i^2)\left[-\frac{1}{n_i^2}\right] - 0 + \lambda\left[c_i - 0 + 0\right]$$

$$= -\frac{N_i^2S_i^2}{N^2n_i^2} + \lambda c_i$$

$$\Rightarrow -\frac{N_i^2S_i^2}{N^2n_i^2} + \lambda c_i = 0 \Rightarrow \lambda c_i = \frac{N_i^2S_i^2}{N^2n_i^2}$$

$$\Rightarrow n_i^2 = \frac{N_i^2S_i^2}{N^2(\lambda c_i)}$$

$$\Rightarrow n_i = \frac{N_iS_i}{N\sqrt{\lambda c_i}} \quad \text{--- (1)}$$

To find λ, we take summation eq^n ① over i from 1 to K we get

$$\sum_{i=1}^{K} n_i = \frac{1}{N\sqrt{\lambda}} \sum_{i=1}^{K} \frac{N_i S_i}{\sqrt{C_i}}$$

$$n = \frac{1}{N\sqrt{\lambda}} \sum_{i=1}^{K} \frac{N_i S_i}{\sqrt{C_i}} \qquad \left(\because n = \sum_{i=1}^{K} n_i\right)$$

$$\sqrt{\lambda} = \frac{1}{Nn} \sum_{i=1}^{K} \frac{N_i S_i}{\sqrt{C_i}} \quad \text{— ②}$$

Sub λ in eq^n ①

$$n_i = \frac{N_i S_i}{\cancel{N}\sqrt{C_i}\left[\frac{1}{\cancel{N}n} \sum_{i=1}^{K} \frac{N_i S_i}{\sqrt{C_i}}\right]}$$

$$= \frac{n N_i S_i}{\sqrt{C_i}\left[\sum_{i=1}^{K} \frac{N_i S_i}{\sqrt{C_i}}\right]}$$

$$n_i = C \frac{N_i S_i}{\sqrt{C_i}}$$

where $C = \dfrac{n}{\sum_{i=1}^{K} \frac{N_i S_i}{\sqrt{C}}} = \text{const}$

finally,

$$\boxed{n_i \propto \frac{N_i S_i}{\sqrt{C_i}}}$$

Conclusion: n_i is large if, N_i is large
S_i is large
& C_i is small

Comparison of STRS with SRS

(a) Proportional Allocation (vs) SRS

$$V(\bar{y}_{st})_p = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{K} p_i S_i^2 \; ; \; S_i^2 = \frac{1}{N_i - 1} \sum_{i=1}^{N_i} (y_{ij} - \bar{Y}_{Ni})^2$$

$$V(\bar{y}_n)_{SRSWOR} = \frac{N-n}{Nn} S^2 = \left(\frac{1}{n} - \frac{1}{N}\right) S^2$$

where $S^2 = \dfrac{1}{N-1} \sum_{i=1}^{K} \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_N)^2$.

$$(N-1)S^2 = \sum_{i=1}^{K} \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_N)^2$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{N_i} \left[y_{ij} - \bar{Y}_{Ni} + \bar{Y}_{Ni} - \bar{Y}_N\right]^2$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_{Ni})^2 + \sum_{i=1}^{K} \sum_{j=1}^{N_i} (\bar{Y}_{Ni} - \bar{Y}_N)^2 + 2 \sum_{i=1}^{K} \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_{Ni})(\bar{Y}_{Ni} - \bar{Y}_N)$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_{Ni})^2 + \sum_{i=1}^{K} N_i (\bar{Y}_{Ni} - \bar{Y}_N)^2 + 0$$

$$(N-1)S^2 = \sum_{i=1}^{K} \left[(N_i - 1) S_i^2\right] + \sum_{i=1}^{K} N_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

If $N - 1 \cong N$; $N_i - 1 \cong N_i$, we get

$$N S^2 = \sum_{i=1}^{K} N_i S_i^2 + \sum_{i=1}^{K} N_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$\Rightarrow S^2 = \sum_{i=1}^{K} \frac{N_i}{N} S^2 + \sum_{i=1}^{K} \frac{N_i}{N} (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$= \sum_{i=1}^{K} p_i S_i^2 + \sum_{i=1}^{K} p_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$\therefore V(\bar{y}_n)_{SRSWOR} = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\sum_{i=1}^{K} p_i S_i^2 + \sum_{i=1}^{K} p_i (\bar{Y}_{Ni} - \bar{Y}_N)^2\right]$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{K} p_i S_i^2 + \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{K} p_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$= V(\bar{y}_{st})_p + \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{K} p_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$\Rightarrow V(\bar{y}_n)_{SRSWOR} \geq V(\bar{y}_{st})_p \qquad \left[\because \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{K} p_i (\bar{Y}_{Ni} - \bar{Y}_N)^2 \geq 0\right]$$

$\Rightarrow$ STRS under proportional allocation is more effecient than SRS.

(b) Neyman's Optimum Allocation (vs) proportional Allocation

$$V(\bar{y}_{st})_{opt} = \frac{1}{n}\left(\sum_{i=1}^{K} p_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{K} p_i S_i^2$$

$$V(\bar{y}_{st})_P = \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{K} p_i S_i^2$$

$$\therefore V(\bar{y}_{st})_P - V(\bar{y}_{st})_{opt} = \left[\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{K} p_i S_i^2\right] - \left[\frac{1}{n}\left(\sum_{i=1}^{K} p_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{K} p_i S_i^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{K} p_i S_i^2 - \cancel{\frac{1}{N}\sum_{i=1}^{K} p_i S_i^2} - \frac{1}{n}\left(\sum_{i=1}^{K} p_i S_i\right)^2 + \cancel{\frac{1}{N}\sum_{i=1}^{K} p_i S_i^2}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{K} p_i S_i^2 - \left(\sum_{i=1}^{K} p_i S_i\right)^2\right]$$

$\Rightarrow V(\bar{y}_{st})_P - V(\bar{y}_{st})_{opt} \geq 0$ $\quad (\because$ RHS is $\geq 0)$

$\Rightarrow V(\bar{y}_{st})_P \geq V(\bar{y}_{st})_{opt}$

$\Rightarrow V(\bar{y}_{st})_{opt} \leq V(\bar{y}_{st})_P$

$\Rightarrow$ STRS under Neyman's optimum allocation is more efficient than STRS under proportional allocation.

(C) Neyman's Opt (vs) SRS

$$V(\bar{y}_{st})_{opt} = \frac{1}{n}\left(\sum_{i=1}^{K} p_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{K} p_i S_i^2$$

$$V(\bar{y}_n)_R = \left(\frac{1}{n}-\frac{1}{N}\right)S^2$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)\left[\sum_{i=1}^{K} p_i S_i^2 + \sum_{i=1}^{K} p_i\left(\bar{Y}_{Ni}-\bar{Y}_N\right)^2\right]$$

$$V(\bar{y}_n)_R - V(\bar{y}_{st})_{opt} = \left\{\left(\frac{1}{n}-\frac{1}{N}\right)\left[\sum_{i=1}^{K} p_i S_i^2 + \sum_{i=1}^{K} p_i\left(\bar{Y}_{Ni}-\bar{Y}_N\right)^2\right]\right\}$$

$$- \left\{\frac{1}{n}\left(\sum_{i=1}^{K} p_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{K} p_i S_i^2\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{K} p_i S_i^2 + \frac{1}{n}\sum_{i=1}^{K} p_i\left(\bar{Y}_{Ni}-\bar{Y}_N\right)^2 - \cancel{\frac{1}{N}\sum_{i=1}^{K} p_i S_i^2} - \frac{1}{N}\sum_{i=1}^{K} p_i\left(\bar{Y}_{Ni}-\bar{Y}_N\right)^2$$

$$- \frac{1}{n}\left(\sum_{i=1}^{K} p_i S_i\right)^2 + \cancel{\frac{1}{N}\sum_{i=1}^{K} p_i S_i^2}$$

$\Rightarrow V(\bar{y}_n)_R - V(\bar{y}_{st})_{opt} \geq 0$ $\quad [\because$ RHS is $\geq 0]$

$\Rightarrow V(\bar{y}_n)_R \geq V(\bar{y}_{st})_{opt}$

$\Rightarrow V(\bar{y}_{st})_{opt} \leq V(\bar{y}_n)_R$

$\therefore$ STRS under optimum allocation is more efficient than SRS

Finally (a) $V(\bar{y}_{st})_P \leq V(\bar{y}_n)_R$

(b) $V(\bar{y}_{st})_{opt} \leq V(\bar{y}_{st})_P$

(c) $V(\bar{y}_{st})_{opt} \leq V(\bar{y}_n)_R$

$\Rightarrow V(\bar{y}_{st})_{opt} \leq V(\bar{y}_{st})_P \leq V(\bar{y}_n)_R$

III. Systematic Sampling (SYS)

It consits in selecting only the first unit at random & the rest being selected according to some predetermined pattern involving regular spacing of units.

Let N = Population size ; n = sample size

then $N = nK \Rightarrow K = N/n$ — Integer

K = sampling internal.

SYS consists in drawing a random number i from 1 to K & selecting the unit corresponding to this number & every K^{th} unit subsequently. Thus, the SYS of size n will consits of the units $i, i+K, i+2K, \ldots, i+(n-1)K$

'i' is called the random start & it ~~dep~~ determines the entire sample.

Problems on Comparison of STRS with SRS

① A sample of 30 students is to be drawn from a population of 300 students from two colleges A & B.

College	N_i	$\bar{Y}_{Ni}$	E_i
A	200	300	10
B	100	60	40

Draw sample sizes using proportional & optimum allocations. Also calculate variances under proportional & optimum allocations & compare with SRSWOR.

Sol ⓐ Proportional allocation -

$$n_i = \frac{n}{N}(N_i)$$

$$V(\bar{y}_{st})_p = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{K} p_i S_i^2 \; ; \; p_i = \frac{N_i}{N}$$

$N = 300$; $n = 30$

$$E_i^2 = \frac{1}{N_i}\sum_{j=1}^{N_i}(y_{ij} - \bar{Y}_{Ni})^2 \; ; \; S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(y_{ij} - \bar{Y}_{Ni})^2$$

$K = 2$,

$$n_1 = \frac{30}{300}(200) = 20 \; ; \; n_2 = \frac{30}{300}(100) = 10$$

from ε_i^2 & S_i^2

$$S_i^2 = \frac{1}{N_i - 1} N_i \varepsilon_i^2 \qquad p_i = \frac{N_i}{N}$$

$$S_1^2 = \frac{1}{200-1}(200)[10]^2$$

$$S_1^2 = 100.5025$$

$$S_1 = \sqrt{100.5025}$$

$$S_1 = 10.02509$$

$$S_2^2 = \frac{1}{100-1}(100)(40)^2$$

$$S_2^2 = 1605.3511$$

$$S_2 = \sqrt{1605.3511}$$

$$S_2 = 40.0668$$

$$p_1 = \frac{200}{300} = \frac{2}{3}$$

$$p_2 = \frac{100}{300} = \frac{1}{3}$$

$$v(\bar{y}_{st})p = \left(\frac{1}{30} - \frac{1}{300}\right)\left\{\frac{2}{3}(100.5025) + \frac{1}{3}(1605.3511)\right\}$$

$$= \left[\frac{1}{30} - \frac{1}{300}\right](67.0016 + 535.1170)$$

$$= \left[\frac{1}{30} - \frac{1}{300}\right](602.1186)$$

$$= 20.0706 - 2.0070$$

$$\boxed{v(\bar{y}_{st})p = 18.0636.}$$

(b) optimum allocation

sample size n_i is given by.

$$n_i = \frac{n(N_i S_i)}{\sum_{i=1}^{K} N_i S_i} \quad ; \quad V(\bar{y}_{st})_{opt} = \frac{1}{n}\left(\sum_{i=1}^{K} p_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{K} p_i S_i^2$$

$$n_1 = \frac{30[200(10.025)]}{200(10.025)+100(40.066)}$$

$$= \frac{60150}{6011.6}$$

$$n_1 = 10.005 \cong 10$$

$$n_2 = \frac{30[100(40.066)]}{200(10.025)+100(40.066)}$$

$$= \frac{120198}{6011.6}$$

$$n_2 = 19.9999 \cong 20$$

$$V(\bar{y}_{st})_{opt} = \frac{1}{n}\left(\sum_{i=1}^{k} p_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} p_i S_i^2$$

$$= \frac{1}{30}\left[\frac{2}{3}(10.025) + \frac{1}{3}(40.066)\right]^2 - \frac{1}{300}\left[\frac{2}{3}(100.5025) + \frac{1}{3}(1605.351)\right]$$

$$= \frac{1}{30}\left[6.6833 + 13.3553\right]^2 - \frac{1}{300}\left[67.0016 + 535.1166\right]$$

$$= \frac{1}{30}(20.0386)^2 - \frac{1}{300}(602.1227)$$

$$= \frac{401.5454}{30} - \frac{602.1227}{300}$$

$$= 13.3848 - 2.0070$$

$$V(\bar{y}_{st})_{opt} = 11.3778$$

(ii) Sys Notations & terminology

8th Dec, 2020

→ Let y_{ij} denote the observation on the j^{th} unit of the $(i=1,2,\dots K)$ $(j=1,2,\dots n)$

→ $\bar{y}_i$ is mean of the i^{th} systematic sample & $\boxed{\bar{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}}$

→ S^2 is the population mean sq $\boxed{S^2 = \frac{1}{N-1}\sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij}-\bar{y})^2}$

on simplification. $\boxed{S^2 = \frac{1}{nK-1}\sum_{i=1}^{K}\sum_{j=1}^{n}(y_{ij}-\bar{y})^2}$

→ $\bar{y}_{..}$ is the population mean $\boxed{\bar{y}_{..} = \frac{1}{nK}\sum_{i=1}^{K}\sum_{j=1}^{n} y_{ij}}$

10th Dec, 2020

Merits & Demerits of Systematic sampling.

Merits:

1) Systematic sampling is operationally more conveint than simple random sampling. (or) stratified random sampling. Time & work involve is also very ~~als~~ less.

2) Systematic sampling may be more efficient than, simple random sampling provided the list from which the sample units of drawn.

Demerits:

* Systematic sampling has a no of limitations & ~~and~~ disadvantage as given below:

1) The main disadvantage of systematic sampling is that systematic samples not in general random samples in the requirement in merit to ~~the~~ is rearely fullfill

2) If ~~so~~ 'N' is nd multiple of 'n' then

i) actually sample size is diff from that required.

ii) sample mean is not an unbaised estimate of the population mean.

iii) In systematic sampling it is not possible to find an unbiased estimate of population variance.

11th Dec, 2020

(Th) ***

Statement: If the population consist of linear trend then

P.T. $V(\bar{y}_{st}) \leq V(\bar{y}_{sys}) \leq V(\bar{y}_n)_R$

Proof: Let us suppose that the population has linear trend given by the model.

$$y_i = i \quad (i=1,2,\ldots N)$$

Now, $$\sum_{i=1}^{N} y_i = \sum_{i=1}^{N} i$$

$$= 1+2+\ldots+N$$

$$= \frac{N(N+1)}{2}$$

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} i^2$$

$$= 1^2+2^2+\ldots+N^2$$

$$= \frac{N(N+1)(2N+1)}{6}$$

$$\bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$= \frac{1}{\cancel{N}}\frac{\cancel{N}(N+1)}{2} = \frac{N+1}{2}$$

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N}(y_i-\bar{y}_N)^2\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\bar{y}_N^2\right]$$

$$= \frac{1}{N-1}\left[\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4}\right]$$

$$= \frac{1}{N-1}\left[\frac{N(N+1)}{2}\left(\frac{2N+1}{3} - \frac{N+1}{2}\right)\right]$$

$$= \frac{1}{N-1} \cdot \frac{N(N+1)}{2}\left[\frac{4N+2-3N-3}{6}\right]$$

$$= \frac{1}{\cancel{N-1}} \cdot \frac{N(N+1)}{2}\left[\frac{\cancel{N-1}}{6}\right]$$

$$= \frac{N(N+1)}{12}$$

$$\text{var}(\bar{y}_n)_R = \left[\frac{1}{n} - \frac{1}{N}\right] S^2$$

$$= \frac{N-n}{Nn} S^2$$

$$= \frac{nK-n}{nKn} \cdot \frac{N(N+1)}{12}$$

$$= \frac{\cancel{n}(K-1)}{n^{\cancel{2}} K} \cdot \frac{\cancel{nK}(nK+1)}{12}$$

$$\text{var}(\bar{y}_n)_R = \frac{(K-1)(nK+1)}{12}.$$

$$\text{var}(\bar{y}_{st}) = \frac{K-1}{n^2 K} \sum_{j=1}^{n} S_j^2$$

w.k.t $S^2 = \frac{N(N+1)}{12}$

for population of N units

since, j th stratum consist of K units

we have, $S_i^2 = \frac{K(K+1)}{12}$

$$V(\bar{y}_{st}) = \frac{K-1}{n^2 K} \sum_{j=1}^{n} \frac{K(K+1)}{12}$$

$$= \frac{K-1}{n^2 \cancel{K}} \cdot \frac{\cancel{K}(K+1)}{12} \cancel{n}.$$

$$= \frac{K^2 - 1}{12n}$$

for finding out $V(\bar{y}_{sys})$ we have

$\bar{y}_i$ = mean of the value of i^{th} sample

= mean of the i^{th} row.

$$= \frac{1}{n} \sum_{j=1}^{n} y_{ij}$$

$$= \frac{1}{n} [i + (i+k) + (i+2k) + \dots + i + (n-1)k]$$

$$= \frac{1}{n} [i + i + \dots + i + k + 2k + \dots + (n-1)k]$$

$$= \frac{1}{n} [ni + k(1 + 2 + \dots + (n-1)k]$$

$$y_i = \frac{1}{n} [ni + (1 + 2 + \dots + (n-1)k]$$

$$= \frac{1}{n} \left[ni + \frac{(n-1)(n)}{2} k \right]$$

$$y_{i.} = i + \frac{n-1}{2} k$$

Also,

$$\bar{y}_{..} = \bar{y}_N = \frac{N+1}{2} = \frac{nk+1}{2}$$

$$\bar{y}_{i.} - \bar{y}_{..} = i + \frac{n-1}{2} k - \frac{nk+1}{2}$$

$$= i + \frac{\cancel{nk} - k - \cancel{nk} - 1}{2}$$

$$= i + \left[\frac{-(k+1)}{2} \right]$$

$$\bar{y}_i - \bar{y}_{..} = i - \frac{k+1}{2}$$

$$var(\bar{y}_{sys}) = \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_i - \bar{y}_{..})^2$$

$$= \frac{1}{k} \sum_{i=1}^{k} \left[i - \frac{k+1}{2} \right]^2$$

$$= \frac{1}{k} \sum_{i=1}^{k} \left[i^2 + \left(\frac{k+1}{2}\right)^2 - 2i \frac{k+1}{2} \right]$$

$$= \frac{1}{k} \left[\sum_{i=1}^{k} i^2 + \sum_{i=1}^{k} \left(\frac{k+1}{2}\right)^2 - 2 \frac{k+1}{2} \sum i \right]$$

$$var(\bar{y}_{sys}) = \frac{1}{k} \left[\frac{k(k+1)(2k+1)}{6} + \frac{k(k+1)^2}{4} - 2 \frac{k+1}{2} \cdot \frac{k(k+1)}{2} \right]$$

$$= \frac{(K+1)(2K+1)}{6} + \frac{(K+1)^2}{4} - \frac{(K+1)^2}{2}$$

$$= \frac{K+1}{2}\left[\frac{2K+1}{3} + \frac{K+1}{2} - K+1\right]$$

$$= \frac{K+1}{2}\left[\frac{4K+2+3K+3-6K-6}{6}\right]$$

$$= \frac{K+1}{2}\left[\frac{K-1}{6}\right]$$

$$\text{var}(\bar{y}_{sys}) = \frac{K^2-1}{12}$$

$$\text{var}(\bar{y}_{st}) : \text{var}(\bar{y}_{sys}) : \text{var}(\bar{y}_n)_R$$

$$\frac{K^2-1}{12n} : \frac{K^2-1}{12} : \frac{(K-1)(nK+1)}{12}$$

$$\Rightarrow \frac{1}{n} : 1 : n$$

$$\Rightarrow V(\bar{y}_{st}) \leq V(\bar{y}_{sys}) \leq V(\bar{y}_n)_R$$

18th Dec, 2020

Measurement of seasonal variations:

Seasonal variations can be measured by using the following four methods:

1. Simple average method
2. Ratio to Trend method
3. Ratio to Moving average method.
4. Link relative method.

① Simple average method: This is the simplest of ~~the~~ all the methods of measuring seasonal variations & consists the following steps

→ Arrange the data by years & months (or quarters if quarterly data are given)

→ Calculate the average for each month for monthly data or each quarter for quarterly data.

→ Calculate the average of monthly averages i.e., $\bar{x} = \frac{1}{12}\sum_{i=1}^{12} \bar{x}_i$

→ Seasonal ~~indices~~ indices for diff months are ~~eptain~~ obtained by expressing monthly averages as percentage of $\bar{x}$. Seasonal index for i^{th} month $= \frac{\bar{x}_i}{\bar{x}} \times 100$

Merits & Demerits:

1. This is the simplest method of measuring seasonal variations.
2. This method is based on the basic assumption that the data do not contain any trend & cyclic component which is not true in general.

Problems

① The data below give the avg quarterly prices of a commodity for four year.

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1980	40.3	44.8	46.0	48.0
1981	50.1	53.1	55.3	59.5
1982	47.2	50.1	52.1	55.2
1983	55.4	59.0	61.6	65.3

Calculate the seasonal variation indices.

Sol Assuming that the trend is absent in the above data, the difference in the averages of various quarters (if there is any) will be due to seasonal changes.

Year	1st Q	2nd Q	3rd Q	4th Q
1980	40.3	44.8	46.0	48.0
1981	30.1	53.1	55.3	59.5
1982	47.2	50.1	52.1	55.2
1983	55.4	59.0	61.6	65.3
Total	193.0	207.0	215.0	228.0
Avg.	48.25	51.75	53.75	57.0

The average of averages $= \frac{48.25+51.75+53.75+57.0}{4}$

$= 52.69$

Seasonal Index for 1st Q $= \frac{48.25}{52.69} \times 100 = 91.57$

S.I for 2nd Q $= \frac{51.75}{52.69} \times 100 = 98.21$

S.I for 3rd Q $= \frac{53.75}{52.69} \times 100 = 102.01$

S.I for 4th Q $= \frac{57}{52.69} \times 100 = 108.18$

② Use the method of monthly avgs to determine the monthly indices for the following data of production of a commodity for the years 1981, 1982, 1983:

Sol

Month	Production in lakhs of tonnes 1981	1982	1983	Total	Monthly avg	Seasonal Index
Jan	12	15	16	43	14.33	104.886
Feb	11	14	15	40	13.33	97.566
Mar	10	13	14	37	12.33	90.247
April	14	16	16	46	15.33	112.205
May	15	16	15	46	15.33	112.205
June	15	15	17	47	15.66	114.620
July	16	17	16	49	16.33	119.524
Aug	13	12	13	38	12.66	92.662
Sept	11	13	10	34	11.33	82.928
Oct	10	12	10	32	10.66	78.024
Nov.	12	13	11	36	12.00	87.832
Dec	15	14	15	44	14.66	107.301
Total				492	163.95	1200
Avgs				41	13.6625	100

② Ratio to trend Method:

21st Dec, 2020

This method is an improvement over the simple avg method & is based on the assumption that seasonal variation for any given month is constant factor of the trend the measurement of seasonal variations by this method consists the following steps.

→ obtain the trend values by the least squares method by fitting a mathematics curve or a straight line or a second-degree parabola

→ express the original data as the % of trend values assuming the multiplicative model there % will contain the seasonal cyclic & irregular components.

→ The cyclic & irregular components by averaging the percentages for diff months (quaters) if the data are monthly (quarterly) thus leaving us with indices of seasonal variations.

→ finally these indices obtained in step 3 or adjested to a total of 1200 for montly data or 400 per quarterly data by multiplying them throughout a constant K. where $K = \frac{400}{\text{Total of indices}}$

or $K = \frac{1200}{\text{total of indices}}$

Merits & demerits:

1. The main advantage of this method over ratio to moving avg if we can obtain the trend values for each seasons.

1. If the series exibit cyclical swings the trend values obtained by the least squares method can never follow the actual data as clearly as 12 months moving average.

Problems

① Calculate seasonal variation for following data by using ratio to trend method.

Year	1st Q	2nd Q	3rd Q	4th Q
1979	30	40	36	34
1980	34	52	50	44
1981	40	58	54	48
1982	54	76	68	62
1983	80	92	86	82.

Soln First of all we determine the trend values for the yearly averages(Y) by fitting a linear trend by the method of least squares.

Year (X)	Yearly total	Yearly avg(Y)	U=X-1972	UY	U²	Trend value
1979	140	35	-2	-70	4	32
1980	180	45	-1	-45	1	44
1981	200	50	0	0	0	56
1982	260	65	1	65	1	68
1983	340	85	2	170	4	80
Total		280	0	120	10	

For the line $Y = a + bU$, the normal eqn for estimating 'a' & 'b' are

$$\left.\begin{aligned} \Sigma Y &= na + b\Sigma U \\ \Sigma UY &= a\Sigma U + b\Sigma U^2 \end{aligned}\right\} \Rightarrow \quad a = \frac{\Sigma Y}{n} = \frac{280}{5} = 56, \quad b = \frac{\Sigma UY}{\Sigma U^2} = \frac{120}{10} = 12$$

∴ Trend line is $Y = 56 + 12U$

$U = -2 \Rightarrow \hat{Y} = 56 - 24 = 32$

$U = -1 \Rightarrow \hat{Y} = 56 - 12 = 44$

Similarly other trend values as given in the above table can be obtained.

Yearly increment in the trend value $= b \Rightarrow 12$

Quarterly increment $= \frac{12}{4} = 3$

Next we determine the quarterly trend values as follows:

For 1979, the trend value for the middle quarter, i.e., half of second quarter & half of third quarter, is 32 & since the quarterly increment in value is 3, we obtain the trend values for second & 3rd quarters of 1979 as 32 − 1.5 i.e., 30.5 & ~~33.5~~ 32 + 1.5 i.e., 33.5 respectively & consequently the trend values for 1st quarter is 30.5 − 3 = 27.5 & 4th Quarter is 33.5 + 3 = 36.5, Similarly we can get the trend values for other years as given in the following table.

Calculations for seasonal variations

	Trend values				Trend Eliminated values (Given values as % of trend values)			
Year	1st Q	2nd Q	3rd Q	4th Q	1st Q	2nd Q	3rd Q	4th Q
1979	27.5	30.5	33.5	36.5	109.1	131.1	107.5	93.1
1980	39.5	42.5	45.5	48.5	86.1	122.4	109.9	90.7
1981	51.5	54.5	57.5	60.5	77.7	106.4	93.9	70.3
1982	63.5	66.5	69.5	72.5	85.0	114.3	97.8	85.5
1983	75.5	78.5	81.5	84.5	106.0	117.1	105.5	97.0
				Total	436.9	591.3	514.6	445.6
				Average (A.M)	92.78	118.26	102.92	89.12
				(Adjusted) Seasonal Index	92.0	117.4	102.1	88.4.

(3) Ratio to Moving average Method:

The method of getting seasonal indices by moving averages involves the following steps.

* Calculate the 12 months centered moving averages of the data these moving avg values will gives estimates of the combined effect of trend & cyclic variations.
* Express the original data (except for 6 months in the begining & 6 months at the end) as percentages of the centered moving averages these percentages consists seasonal & irregular variations.

$$\boxed{\% \text{ of Moving avg} = \frac{\text{Original value}}{\text{moving avg value}} \times 100}$$

* The irregular or random variations are wipedout by averaging these indices either arthematic mean or median can be used for averaging.
* finally, we need to adjust to a total of 1200 for monthly data or 400 for quarterly data by multiplying them throughout by a constant K. where K is equal to $\frac{1200}{\text{total of indices}}$ or $\frac{400}{\text{total of indices}}$

$$K = \frac{1200}{\text{total of indices}} \quad (or) \quad \frac{400}{\text{total of indices}}$$

Merits & demerits

* Of all the methods of measuring seasonal variations the ratio to the moving avg method is the most satisfying avg flexible & widely used method.
* This method thus not utilize the complete data for example in the case of 12 month moving averages seasonal indices can't be obtained for the first 6 months & for the last 6 months.

22nd Dec, 2020

Problems

(1) calculate seasonal indices by the ratio to moving avg method from the following data.

Quarter \ Year	1980	1981	1982	1983
Q_1	75	86	90	100
Q_2	60	65	72	78
Q_3	54	63	66	72
Q_4	59	80	85	93.

Sol Assuming multiplicative model of time series, the trend (Moving Avg) values are eliminated by expressing the given values (V_i) as a percentage of trend values & are given in the last column of the above table.

Calculations for moving Averages

Year-Quarter (1)		Price (2)	4-Quarter moving Total (3)	Sum of 2 4-Q moving totals (4)	4-Quarter Moving Avg (4÷8) (5)	Ratio to Moving avg 2÷5×100 (6)	U_i - M.A (7)
1980	I	75					
	II	60	248				
	III	50	259	507	63.375	85.2971	-9.375
	IV	59		523	65.375	90.2485	-6.375
1981	I	86	267	537	67.125	128.1192	18.875
	II	65	273	567	70.875	91.7108	-5.875
	III	63	294	592	74.000	85.1351	-11.000
	IV	80	298	603	75.375	106.1360	-6.625
1982	I	90	305	613	76.625	117.4551	13.375
	II	72	308	621	77.625	92.7536	-5.625
	III	66	313	636	79.500	83.0189	-13.500
	IV	85	323	652	81.500	104.2945	3.500
1983	I	100	329	664	83.000	120.4819	17.000
	II	78	335		~~84~~		
	III	72	343	678	84.750	92.0354	-6.750
	IV	93					

Computation of Seasonal Indices

Year	Trend Eliminated values I Q	II Q	III Q	IV Q	
1980	-	-	85.2071	90.2485	
1981	128.1192	91.7108	85.1351	106.1360	
1982	117.4551	92.7536	83.0189	104.2945	
1983	120.4819	92.0354	-	-	
Total	366.0562	276.4998	253.3611	300.6790	
Avg (A.M.) (S.I)	122.0187	92.1666	84.4537	100.226	Total = 398.865
Adjusted seasonal less Indices	122.3603	92.4246	84.6902	100.5066	399.985 ≅ 400

The seasonal indices obtained as avg (A.M) above are adjusted to a total of 400, by multiplying each of them by a constant factor

$$K = \frac{400}{\text{Sum of seasonal Indices}} = \frac{400}{398.865} = 1.0028.$$

(4) Link Relative method.

*Link relative for any month = $\dfrac{\text{current month value}}{\text{previous month value}} \times 100$

The steps involved in this method are as follows.

Step1: Translate the original data into link relatives as explained above.

Step2: As in the case of ratio to trend method avg the link relatives for each month (quarter) if the data are monthly (quarterly)

Step3: Convert the avg link relatives into chain relatives of (CR) on the base of the 1st season chain relative for any season is obtained on multiplying the link relative of that season by the chain relative of the preceding season & dividing by 100. Thus, for monthly data the chain relative for 1st season i.e., for January is taken to be hundered.

chain relative for Feb = $\frac{\text{LR of feb} \times \text{CR of Jan}}{100}$

& so on CR for dec = $\frac{\text{LR of Dec} \times \text{CR of Nov}}{100}$

Now by taking this dec value as a base a new CR for January can be obtained as $\frac{\text{LR of Jan} \times \text{CR of dec}}{100}$

usually this will not be 100 due to trend & so, we have to adjust the CR for trend.

Step4: This adjustment is done by subtracting a correction factor from each CR & correction factor.

$$d = \frac{1}{12} \times \text{new CR of Jan} - 100$$

Assuming linear trend the correction factor for feb, mar, ..., dec is $d, 2d, \dots, 11d$ respectively.

Step5: Finally, adjust the corrected CR to total 1200 by expressing the corrected chain relatives as percentage of their arthematic mean.

Merits & Demerits

1. This method utilizes data more completely than moving avg method. There is only one less link relative while a 12month moving avg results in ciel of 6 months at each end.

2. This method is effective only if the growth is of constant amount or constant rate.

29th Dec, 2020

Problems

1. The data below gives the avg quarterly prices of a commodity for 5 yrs. Calculate the seasonal variation indices by the method of link relatives.

Quarter \ Year	1979	1980	1981	1982	1983
I	30	35	31	31	34
II	26	28	29	31	36
III	22	22	28	25	26
IV	31	36	32	35	33.

Sol calculations for seasonal Indices by the method of link relatives.

Year	Link Relatives			
	1st Q	2nd Q	3rd Q	4th Q
1979	—	86.7	84.5	140.8
1980	112.9	80.0	78.8	153.5
1981	86.1	83.5	96.8	114.3
1982	90.9	100.0	80.7	140.0
1983	97.1	105.9	72.2	120.9

Arithematic avg	(393.0)/4	(446.1)/5 = 93.22	442.7/5 = 82.54	(685.7)/5
Chain Relatives	$=\frac{98.25}{100}$	$\frac{100\times 93.22}{100}$	$\frac{93.22\times 82.54}{100} = 76.950$	$=137.14$ $\frac{76.05\times 137.14}{100} = 105.4$
Ajusted chain Relatives	100	92.345	75.200	109.779
Seasonal Indices	108.02	99.75	81.23	111.00

Explanation of steps:

The 2nd chain relative for 1st Q = $\frac{105.4\times 98.25}{100} = 103.5$

Correction factor $d = \frac{1}{4}(103.5-100.0) = \frac{3.5}{4} = 0.875$

Adjusted Chain relatives are obtained by subtracing 0.875, 2×0.875, 3×0.875 from the chain relatives of 2nd, 3rd & 4th Q. respectivey.

Avg of adjusted chain Relatives $= \frac{100+92.345+75.200+102.775}{4}$

$= 92.58$

Seasonal variation Index for any quarter.

$= \frac{\text{Adjeusted C.R.}}{92.58} \times 100$

Seasonal indices have been obtained in the above table.

Index Numbers Unit-3 Stats-5 30th Dec

DATE: / / PAGE NO.

Introduction:

~~Def~~

Index numbers: are indicators which reflect changes over a specified period of time in prices of diff ~~combo~~ comodities, industrial production sales imports & exports and cost of living etc.

→ In economics Index numbers ~~mesure~~ measure the pressure of economic behaviour under rightly termed as economic barometers. (The Index numbers are also called as economic barometers)

Def: Index numbers are statistical devices designed to measure the relative changes in the level of a phenomenon with respect to time geographical location (or) other characteristics such as income, profesion etc.

*** Imp

Problems involved in construction of Index numbers:

Following are the some of the imp problems involved in the construction of index numbers.

① The purpose of Index numbers:

* An index number which is properly defined for a purpose can be most useful and powerful tool otherwise it can be equally misleading and dangerous. Thus, the first & foremost problem is to determine the purpose index number without which it is not possible to follow the steps ~~its~~ in its construction.

For example: If the purpose of index number is to measure the changes in the production of steel, the problem of selection of comodities is automatically selected.

② Selection of Comodities:

* Having defined the purpose of index numbers select only those comodeties which are relevant to the index for example if the purpose of an index is to measure the cost of living of poor people we should select only those comodeities or items utilized by persons belonging to this group. Selection of comodeties like cosmetics & other luxury goods like cars, AC's etc ~~may~~ will be ~~absoutally~~ absolutaly useless

③ Data for index numbers:

The data ually the set of prices & quantity consumed of the selected comodities for diff time periods or places etc. The data should be collected from the relaiable sources such as standard trade journals, official publications etc.

For example; for the construction of retail price index numbers the price quotations should be collected from retail outlets ~~as~~ super market, kirana shops etc not from wholesale ~~detoues~~ dealers.

④ Selection of base period:

The period with which the comparison of relative changes in the level of a phenomenon are made is termed as base period. And the index for this period is always taken as 100. The following are the basic ~~crietia~~ criteria for the choice of the base period.

i) The base period must be a normal period i.e., a period free from all abnormalities like floods, earth quakes, labour strikes etc.

ii) The base period should not be too distinct from the given period.

⑤ Type of average to be used:

* Index numbers are also known as special type of averages the following avgs are used in the construction of index numbers.

i) Arthematic mean

ii) Geometric mean

iii) Median.

* Geometric mean is the best avg in the construction of index numbers but mostly arthematic mean is used in the construction of index numbers.

⑥ Selection of weight:

* Generally, various commodities like wheat, rice, kerosene, oil, cloth etc included in the index are not of equal importance proper weights should be attached to them to take into account their relative importance.

⑦ Selection of formula:

For the available information a suitable formula should be selected for the construction of index numbers.

31st Dec, 2020

DATE: / /
PAGE NO.

Calculation of index numbers:

Index numbers are calculated by using the following methods.

① Un weighted index numbers

② weighted index numbers

i) Unweighted index numbers:

↳ Simple aggregative ~~numb~~ method:

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

where P_{01} = price index in current year 1 with respect to base year 0.

p_1 = current year price.

p_0 = base year price.

↳ Simple average of price relative method:

using arthematic mean,

$$P_{01} = \frac{\Sigma P}{n}$$

where, $P = \frac{p_1}{p_0} \times 100$

P is price relative

using Geometric mean,

$$P_{01} = AL\left[\frac{\Sigma \log P}{n}\right]$$

where, $P = \frac{p_1}{p_0} \times 100$

ii) weighted index numbers:

↳ weighted aggregative method:

$$P_{01} = \frac{\Sigma p_1 w}{\Sigma p_0 w} \times 100$$

~~→ lash pear method~~

→ **Laspeyre's method:**

If base year quantities as takes as weights i.e., $w = q_0$

then we get laspeyre index number.

$$P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

→ **Paache's price index method:**

~~If the~~ current year quantities are taken as weight then we will get paache's index number.

$$P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

→ **Drobish & Bowley price index number:**

This formula is the arithmatic mean of Laspeyre & Paache's price index numbers. i.e.,

$$P_{01}^{DB} = \frac{L+P}{2}$$

→ **Fisher's Price Index number:**

This formula is the geometric mean of laspeyre & paache's index numbers i.e.,

$$P_{01}^{F} = \sqrt{L \times P}$$

$$P_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

↳ Marshall Edgeworth price index number:

If the arithematic mean of base year & current year quantities are taken as weights then we get marshall edgeworth index number.

i.e., $P_{01}^{ME} = \dfrac{\Sigma p_1 \left(\frac{q_0+q_1}{2}\right)}{\Sigma p_0 \left(\frac{q_0+q_1}{2}\right)} \times 100$

$= \dfrac{\Sigma p_1 (q_0+q_1)}{\Sigma p_0 (q_0+q_1)} \times 100$

↳ walsch method:

By taking geometric mean of base year & current year quantities we will get walsch index number.

i.e., $P_{01}^{W} = \dfrac{\Sigma p_1 \sqrt{q_0 q_1}}{\Sigma p_0 \sqrt{q_0 q_1}} \times 100$

↳ Kelly's method:

In kelly's method we will consider fixed quantities as weights,

$P_{01}^{K} = \dfrac{\Sigma p_1 q}{\Sigma p_0 q} \times 100.$

Problems

① The table below relates to the daily pay of the wage earners on a company's pay roll:

	1978		April 1983	
	Number	Total pay	Number	Total pay
~~Mean~~ Men aged 21 & over	350	2500	300	4200
women aged 18 & over	400	1600	1200	8000
Youth & boys	150	450	100	560
Girls	100	250	400	1540
	1000	41800	2000	14300

Construct an index of daily earnings based on 1978 as base showing the rise of earnings for all employees as one figure.

Soln Regarding number of wage earners as quantities and the weekly wages per labourer as prices we are given the figure q_1 & p_0q_0 for 1978 & the values q_1 & p_1q_1 for 1983. The following table can be easily completed.

q_0	p_0	q_1	p_1	p_0q_0	p_0q_1	p_1q_1	p_1q_0
350	7.14	300	14.00	2500	2142	4200	4900
400	4.00	1200	6.67	1600	4800	8000	2668
150	3.00	100	5.60	450	300	560	840
100	2.50	400	3.85	250	1000	1540	384
Total				4800	8242	14300	8793

$$\cancel{P_{0i}^{La} = \frac{\Sigma P_{ij} q_{0j}}{\Sigma P_{0i} q_{0j}} \times 100 = \frac{8793}{4800}}$$

$$P_{0i}^{La} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100 = \frac{8793}{4800} \times 100 = 183$$

$$P_{0i}^{Pa} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times 100 = \frac{14300}{8242} \times 100 = 173.5$$

$$P_{0i}^{F} = \sqrt{P_{0i}^{La} \times P_{0i}^{Pa}} = \sqrt{183 \times 173.5} = 178.3.$$

(2) Given the data.

	Commodities A	B
P_0	1	1
q_0	10	5
P_1	2	X
q_1	5	2

where P & q respectively stand for price & quantity & subscripts stand for time period. find X, if the ratio b/w Laspeyres's (L) & Paasche's (P) index number is,
L : P :: 28 : 27.

Sol Calculations for Laspeyre's & Paasche's Indices.

commodities	P_0	q_0	P_1	q_1	P_1q_0	P_0q_0	P_1q_1	P_0q_1
A	1	10	2	5	20	10	20	5
B	1	5	X	2	5X	5	2X	2
Total					20+5X	15	10+2X	7

$$P_{01}^{La} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100 = \frac{20+5X}{15} \times 100$$

$$P_{01}^{Pa} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times 100 = \frac{10+2X}{7} \times 100$$

Now $$\frac{P_{01}^{La}}{P_{01}^{Pa}} = \frac{\frac{(20+5x)}{15}}{\frac{(10+2x)}{7}} = \frac{28}{27}$$

$$\Rightarrow \frac{20+5x}{15} \times \frac{7}{10+2x} = \frac{28}{27}$$

$$\Rightarrow \frac{(20+5x)\cancel{\times 7}}{15(10+2x)} = \frac{28}{27}$$

$$\Rightarrow (140+35x)27 = (150+30x)28$$

$$\Rightarrow 3780+945x = 4200+840x$$

$$\Rightarrow 4200-3780 = 945x-840x$$

$$\Rightarrow 420 = 105x$$

$$\Rightarrow x = \frac{420}{105}$$

$$\boxed{x=4}$$

Sat
2nd, Jan, 2021

Imp ***

Good

~~Tests of consistency of Index number.~~

Criteria of Good Index number:

Following are the criteria of Good index number

① unit test (UT)
② Time Reversal test. (TRT)
③ Factor Reversal test (FRT)
④ Circular test - (CT)

↳ Unit test: According to UT, the Index number formula should be independent of units
unit test is satisfied by all index numbers

↳ Time Reversal test: This test was proposed by Irving Fisher. According to time reversal test

$\boxed{P_{01} \times P_{10} = 1}$

Time Reversal test is satisfied by the following index numbers.

① simple aggregative method
② simple Geometric mean of price relative method.
③ weighted aggregative method
④ weighted Geometric mean of price relative method
⑤ fisher's method.
⑥ Marshall Edgeworth method.
⑦ walsch method
⑧ Kelly's method.

↳ Factor Reversal test: This test is also proposed by Irving Fisher. According to factor reversal test. The product of price index ~~&~~ quantity index is equal to the value index i.e.,

$$P_{01} \times Q_{01} = V_{01} \Rightarrow \Sigma P_1 q_1 / \Sigma P_0 q_0$$

DATE: / /
PAGE NO.

Factor reversal test is satisfied by only Fisher's Index number.

↳ Circular test: Circular test is the extension or generalization of time reversal test.

For four years 0, 1, 2, 3 according to circular test $\boxed{P_{01} \times P_{12} \times P_{23} \times P_{30} = 1}$

circular test is satisfied by the following three formulas:

1. simple aggregative method
2. simple Geometric mean of price relative method.
3. kelly's method. (or) weighted aggregative method with fixed weights

* Show that Fisher's Index number satisfies time reversal test & factor reversal test:

(or)

S.T Fisher Index number is an ideal index number.

Proof: Fisher's price index is given by

$$P_{01}^{F} = \sqrt{L \times P}$$

$$= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$P_{01}^{F} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$P_{01}^{F} \times P_{10}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$= \sqrt{\frac{\cancel{\Sigma p_1 q_0}}{\cancel{\Sigma p_0 q_0}} \times \frac{\cancel{\Sigma p_1 q_1}}{\cancel{\Sigma p_0 q_1}} \times \frac{\cancel{\Sigma p_0 q_1}}{\cancel{\Sigma p_1 q_1}} \times \frac{\cancel{\Sigma p_0 q_0}}{\cancel{\Sigma p_1 q_0}}}$$

$$= \sqrt{1} = 1$$

$$P_{01}^{F} \times P_{10}^{F} = 1$$

$\therefore$ Fisher's index number satisfes TRT.

Factor reversal test

$$P_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$Q_{01}^{F} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$P_{01}^{F} \times Q_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$= \sqrt{\frac{\cancel{\Sigma p_1 q_0}}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\cancel{\Sigma p_0 q_1}} \times \frac{\cancel{\Sigma p_0 q_1}}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\cancel{\Sigma p_1 q_0}}}$$

$$= \sqrt{\left(\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}\right)^2} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = V_{01}$$

$\therefore$ Fisher's I.N satisfes FRT

$\therefore$ Fisher's I.N is an ideal I.N.

Cost of living Index Number
(or)
Consumer price Index Number
(or)
Retail price Index Number

Cost of living IN are constructed to study the effect of changes in the prices of a basket of Goods & services on the purchasing power of a particular class of people during current period as compared with some base period.

***Main steps in the Constructions of cost of living IN:**

[Scope & coverage]

① It is to specify the particular population group or the class per which the IN is intended. Such as industrial workers, Govt. employe, low income or middle income class people etc. together with a well defined geographical region such as a city or an industrial area (or) a particular locality in a city. As far as possible the class should form a homogeneous group of people with respect to income.

② Family budget enquiry

Having decided about the class of people explained in step1 we conduct a sample family budget enquiry i.e., we select a sample of families from the class of people for home the index is intended & scrutinise their budgets in detail.

The enquiry should cover a reasonably adequate no. of families & should be conducted during a normal period i.e., a period free from economic boom (or) dipression.

The purpose of the enquiry is to determine the amount an avg the family spents on different items of consumption.

***Calculation of cost of living IN:**

Cost of living IN are constructed by using the following two methods.

① Aggregate Expenditure Method.

In this method cost of living IN are constructed by using the following formula.

$$\boxed{CLI = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100}$$

that,

$$= \frac{\text{total expenditure in current year}}{\text{total expenditure in base year}} \times 100$$

② Family budget Method.

$$\boxed{CLI = \frac{\Sigma PW}{\Sigma W}}$$

where $P = \frac{P_1}{P_0} \times 100$ & $W = P_0 q_0$

***Uses or Applications of cost of living IN:**

① Cost of living INs are used to find purchasing power of Money & real wages.

$$P.P. \text{ of money} = \frac{100}{CLI}$$

$$\text{Real wage} = \frac{\text{Money wage}}{CLI} \times 100$$

② Cost of living Indices are used for the regulation of DA (Dearness Alowance)

③ These indices are also used for deflation of income & value series in National accounts.

④ Cost of living IN are used in wage contracts.

4th Jan, 2021

Base Shifting, Splicing & Deflating of Index numbers

***Base shifting:** shifting of base year from one year to another year is called base-shifting.

Base shifting is done under the following two terms

① The base period is too distinct (or) too old from the current year

② To compare two different series of index numbers with different base years

we express the both series of index numbers with a common base

$$\text{recast Index} = \frac{\text{Index no. of old base year}}{\text{Index no. of new base year}} \times 100$$

***Splicing of Index numbers:**

In order to obtain continuity in the comparison of two or more overlapping series of index numbers we combine (or) splice them into a single continuous series.

Splicing means combaining two series of index numbers to form into a single series of Index numbers.

Splicing of Index numbers is possible if the following conditions are satisfied

① All such series of index numbers must be constructed with same ~~commots~~ commodities

② All such series of index numbers must be constructed with different base periods.

③ All such series of INs must contain atleast one overlapping year.

5th Jan, 2021

* Forward splicing: If the new series is spliced with old series. Then the splicing is called forward splicing.

In forward splicing the multiplication factor is $\frac{a_K}{100}$

where a_K is old series I.no. of the overlapping year.

* Backward splicing: If the old series is spliced with new series. Then the splicing is called backward splicing.

In backward splicing the multiplication factor is $\frac{100}{a_K}$

Deflating of Index Numbers:

Deflating means making allowance for the effect of changing Price levels. The increase in the prices of consumer goods for a class of people over a period of years means a reduction in the purchasing power for the class.

For example, the increase in price of a particular commodity from rupees x (Rs. x) in base year a to ruppes 2x (Rs. 2x) in current year 'b' implies that in current year b. a person can buy only half the amount of ~~commots~~ ~~comm~~ commodity

The Idea of purchasing power of money (or) a measure of the real wages for a class of people is obtained on deflating the wage series by ~~dividing~~ dividing each item by an approximate

$$\text{real wage} = \frac{\text{money wage}}{C \cdot L \cdot I} \times 100$$

The real income is also known as deflated income.

Chain base Index numbers:

The various formulae discussed so far base period is fixed at some perivous period. In some cases the prices will increase in a less span of time i.e., ~~with~~ within 1 year in such cases fixed base index numbers are not ~~suita~~ suitable in these situations we use chain base index numbers.

Construction of chain based Index numbers:

Chain base index numbers are constructed by using the following 2 steps.

Step 1: Calculate link relatives (L.R)

$$L.R = \frac{\text{current year value}}{\text{previous year value}} \times 100$$

Step 2: Calculate chain base index numbers. (C.B.I)

$$C.B.I = \frac{\cancel{\text{current link}}\ \text{current year L.R} \times \text{Previous year C.B.I}}{100}$$

For 1^{st} year the C.B.I will be taken as 100