CHAPTER 6

# RATIO AND REGRESSION ESTIMATES

**6.1 INTRODUCTION**

In the previous Chapters we have considered the problems associated with the estimation of a single population characteristics. If ancillary or supplementary information is available for each of the units of the population, it can under suitable conditions, be used in several ways to improve the efficiency of the estimators of the variable under investigation (estimand). The supplementary information must be correlated with the variable under study. This method of estimation is usually appreciated and generally gives more precise estimates regarding population total as compared to the simple estimation method. The supplementary information usually referred to as **benchmark variable or Auxiliary Variable.**

The term ratio estimation applies to a widely used class of estimation plan which incorporates the prior information including individual values, which is closely related to the variable under investigation for all units of the population. Such situations arise very frequently in practice, especially where the variables under consideration are the current level of some economic variable (such as retail sales) and the results of a survey giving some previous level of that variable. Ratio estimators can be formed corresponding to any selection plan, but in this Chapter attention will be confined to simple and stratified random sampling. Ikeda-Mizuno-Sen- Lahiri selection procedure in which the first unit is selected with probability proportional to the measure of size and remaining n-1 units with equal probability and without replacement will also be described. This makes the ratio estimator unbiased of population total Y.

Given a population N units (Y1, X1), (Y2, X2), …., (YN, XN) the **population ratio**

 (6.1.1)

where Yi is variable under study Xi is benchmark variable. If (y1, x1), (y2, x2), …., (yn, xn), are the values based on a simple random sample of size n, drawn from a population of N units, the **sample ratio** is defined as:

 (6.1.2)

Where ,

and the **ratio estimator** or **classical(conventional) ratio estimator**, , of the population total Y will be;

 (6.1.3)

where

|  |  |
| --- | --- |
| , |  |

The ratio estimator of the population mean may be written directly using (6.1.3) i.e.

 (6.1.4)

Ratio estimator contains bias and consistent. It is applicable when sample size is large. Ratio estimator has the following properties:

* ratio estimator is generally biased, but the bias decreases as the sample size increases as in case of unbiased estimation and the distribution of r tends to normality (this is a property of consistent estimator), and
* the mean square error of ratio estimator is smaller than that of unbiased estimator, if there is high correlation between benchmark variable and estimand.

**6.2 expectation and mean square error of
ratio and ratio estimator**

In this section expectation and variance (mean square error) of ratio and ratio estimator are derived.

**6.2.1 ratio**

**THEOREM (6.1)**

For large n, ratio estimator is approximately equal to R i.e

 , (6.2.1)

with variance(m. s. e)

  (6.2.2)

\*The Var. indicates mean square error (m. s. e)

### proof

We know that

  (6.2.3)

Since n is large, in the denominator may be taken as then (6.2.3) becomes

  (6.2.4)

Taking expectation of both sides of (6.2.4), we get

 ,

as .

Hence E(r)≈ R◊

In order to obtain mean square error, take the expectation of square of both sides of (6.2.4)

 .

Using the concept of simple random sampling [Theorem (2.2)], we can write

 ◊

**Remarks (i):**

An approximate unbiased variance estimator of (6.2.2) may be written directly using the sample analogy.

  (6.2.5)

If , is not known then (6.2.5) alternatively be (replacing  in the denominator)

 . (6.2.6)

For calculation purpose, the most convenient form of (6.2.5) is

 . (6.2.7)

Rao and Rao (1971) compared (6.2.5) and (6.2.6) and found that (6.2.6) is frequently less biased.

**Remark (ii)**

The expression (6.2.2) may also be written alternatively as

, as 

 , (6.2.8)

 , (6.2.9)

where ρ is the correlation coefficient between X and Y i.e. ρ = 

**Remark (iii)**

The equation (6.2.8) may also be put as

 ,

 . (6.2.10)

where CY, CX are coefficient of variation for Y and X respectively and Cyx are coefficient of covariation.

**6.2.2 Ratio Estimator**

The difference between ratio and ratio estimator is only a multiple of X, which is a constant multiplier. It is clear that in the analysis of sampling error  is related to that of ratio, r. It can be easily proved analogues to (6.2.1) that

  (6.2.11)

In order to obtain the m.s.e. of , (6.2.2), (6.2.8), (6.2.9) and (6.2.10) be multiplied by X2

  (6.2.12)

  (6.2.13)

  (6.2.14)

  (6.2.15)

An approximate variance(m s e) estimator of (6.2.12) is

  (6.2.16)

Note that approximate variance (m. s e) estimator expressions for (6.2.13) and (6.2.14) may be obtained by replacing and r for R, in the respective expressions. An approximate variance estimator of (6.2.15) may be written directly i.e.

  (6.2.17)

For practical purposes the most convenient form for (6,2,16) is

  (6.2.18)

(6.2.3) may be written as [(Hansen, Hurwitz and Madow (1953) Vol. II
Chapter 5)] to the correct to order n-1

  (6.2.19)

An approximate variance estimator of (6.2.19) is

  (6.2.20)

6.3 Standard Error and Confidence limits for
Ratio and Ratio Estimator

The standard error of ratio and ratio estimator respectively is

 , (6.3.1)

and

 . (6.3.2)

If sample is large enough and normal approximation applies then the confidence limits for R and Y may be obtained as

 , (6.3.3)

and

  (6.3.4)

where t is normal deviate chosen to correspondence to confidence probability as defined in Chapter 2.

* 1. **BIAS OF THE RATIO ESTIMATOR**

Since E(r) is approximately equal to R, bias in the ratio estimator exist and may be obtained as

 = = .

Expanding the last term of above expression by Taylor’s expression and neglecting the higher order terms

 

Taking the expectation

 Bias ≈

 =

 =

 = (6.4.1)

 =

 =

 = (6.4.2)

Bias is negative, zero or positive provided ⋛.

(6.4.1) may alternatively be put as

  (6.4.3)

Bias = 0, if R =  which is a condition that regression line y on x is a straight line through the origin.

Hartley and Ross (1954) expressed the bias in the following form:

  (6.4.4)

This can be proved as;

 or 

This does not depend on the size of sample, and does not increase or decrease with the increase or decrease of sample size.

The bias in  may be estimated from (6.4.2) and (6.4.3) by multiplying X.

 Bias  (6.4.5)

and

 Bias  (6.4.6)

* 1. **COMPARISON OF**  AND 

We know that

  (6.2.8)

and

  (2.4.1)

(6.2.8) will be less than (2.4.1) provided

 

or 

or  (6.5.1)

Hence the efficiency of ratio estimator with respect to simple (unbiased) estimate is not only depending on the high correlation but also depends on the coefficient of variation of variables. The benchmark variable must have high coefficient of variation than estimand. Further comparison under a stochastic model will be made in Section 6.9.

##### EXAMPLE 6.1

The population of Greece for 69 Urban areas is known from 1941 population census is X = 13,559 in hundreds. 20 urban areas are selected at random using simple random sampling without replacement and the population for 1941 for these areas where noted from the census record and the present population for these 20 areas were obtained through field work to estimate the present population of 69 areas. Estimate the present population using ratio estimation method and simple unbiased estimating method. Compare the efficiency of these two methods (Raj 1972): Data are given in Table 6.1.

**SOLUTION:**







Table 6.1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | y | x2 | Y2 | Xy |
| 122 | 126 | 14884 | 15876 | 15372 |
| 290 | 257 | 84100 | 66049 | 74530 |
| 70 | 108 | 4900 | 11664 | 7560 |
| 141 | 185 | 19881 | 34225 | 26085 |
| 497 | 421 | 247009 | 177241 | 209237 |
| 130 | 148 | 16900 | 21904 | 19240 |
| 87 | 119 | 7569 | 14161 | 10353 |
| 198 | 338 | 39204 | 114244 | 66924 |
| 347 | 368 | 120409 | 135424 | 127696 |
| 151 | 221 | 22801 | 48841 | 33371 |
| 83 | 121 | 6889 | 14641 | 10043 |
| 153 | 151 | 23409 | 22801 | 23103 |
| 156 | 224 | 24336 | 50176 | 34944 |
| 327 | 410 | 106929 | 168100 | 134070 |
| 304 | 295 | 92416 | 87025 | 89680 |
| 139 | 177 | 19321 | 31329 | 24603 |
| 623 | 790 | 388129 | 624100 | 492170 |
| 150 | 176 | 22500 | 30976 | 26400 |
| 217 | 236 | 47089 | 55696 | 51212 |
| 91 | 129 | 8281 | 16641 | 11739 |
| 4276 | 5000 | 1316956 | 1741114 | 1488332 |

 x = population for 1941

 y = present population

**(i) Ratio estimation method**

 = r . X = (1.16932) (13559) = 15885

= 543874.826

S.E.  = 737.4786

**(ii) Simple estimation method**

= 436922.193

= 2090.364

The efficiency of classical ratio estimator  over simple estimation method  is



**6.6 UNBIASED RATIO ESTIMATOR**

We have seen that under simple random sampling, classical (conventional) ratio estimator is biased. Lahiri (1951) suggested that classical ratio estimator can be made unbiased if the selection procedure is changed. Midzuno (1950) and Sen (1951) proved the same result. Lahiri suggested that the first unit was selected with probability proportional to the aggregate of the size (PPAS) or with probability proportional to , and the remaining *n –* 1units with equal probability and without replacement. Midzuno (1951) simplified this procedure as “the first unit is selected with probability proportional to Xi (measure of size), and the remaining (*n* – 1) units like Lahiri (1951)”. This idea was introduced by Ikeda (1950) – reported by Midzuno (1951). This sampling scheme has striking resemblance to the simple random sampling without replacement. In fact, it may be viewed as a generalization of the simple random sampling when extra information on the population is available.

Let we have a population of N units. The probability that ith unit is first one to be selected and subsequent (n – 1) units with equal probability and without replacement is

  .

The probability that jth unit is first one to be selected and subsequent (n – 1) draws with equal probability and without replacement

 ,

and so on the probability P(s) for the two selections are therefore

 .

Since there are n such selection therefore the probability of the selection of the sample will be

  . (6.6.1)

 . (6.6.2)

The classical ratio estimator is

  (6.1.3)

In order to prove the unbiasedness under Ikeda-Midzuno-Sen-Lahiri scheme, we take the expectation of (6.1.3)

**THEOREM 6.2:**

Classical ratio estimator is unbiased under Ikeda-Midzuno- Sen –Lahiri selection procedure with variance

 (6.6.3)

**PROOF**

Taking the expectation of (6.1.3) we have



Where P(s) is the total probability of the sample.

Putting the value of P(s) from (6.6.1) we will have

  

On simplification we get

   ◊ (6.6.4)

 denotes the sum over all possible samples.

The variance expression of  may be derived as;

 

Var()

Substituting the value of P(s) from (6.6.1)

 .◊ (6.6.3)

Note that the . This is very strong property and will be referred to as **Ratio Estimator Property***.*

**THEOREM 6.3**

The mean of ratio estimator is an unbiased with variance

  (6.6.5)

**PROOF**

For this we can take the expectation (6.1.4), we then have

 (6.1.4)

Taking the expectation of (6.1.4) we get

  

Using (6.6.2) we have

  ◊ (6.6.5)

Proceeding by the same way as before we can derive the variance expression of , i.e.

  (6.6.6)

**THEOREM 6.4**

An unbiased estimator of  is

  (6.6.7)

  (6.6.8)

**PROOF**

It may be proved that . For this

 

 =   (6.6.9)

and

 

  (6.6.10)

Hence



 

Similarly we can show that an unbiased estimator of population total will be

  (6.6.11)

**6.7 HARTLEY-ROSS UNBIASED RATIO ESTIMATOR**

Hartley and Ross (1954) proposed another unbiased ratio-type estimator of R.

**THEOREM 6.5:**

In simple random sampling without replacement an unbiased estimator of R is

 , (6.7.1)

where [average of the ratio] and for large sample the Var(rHR) is

. (6.7.2)

where .

#### PROOF

We know that (using the concept of simple random sampling)

 

 



Therefore

  ◊ (6.7.3)

We know from simple random sampling

 

or



Cov (ri xi) can be estimated by this concept, i.e.



 

so

  (6.7.4)

Taking expectation of (6.7.1) and using (6.7.3) and (6.7.4)



 

Hence 

**Note:** Since  which is not easy to calculate so in large sample survey it is unlikely to be used, through it is an unbiased estimator.

The corresponding unbiased ratio estimator  for population total Y is

  (6.7.5)

  (6.7.6)

Now for large n (6.7.1) takes the form

 

By the law of large number the random variable  coverge in probability to . Hence the limiting distribution of

 

is the same as the distribution of

 

Thus

 

  (6.7.7)

Note that .

From (6.2.9) (ignoring correction factor) and (6.7.7) we have:



where , or



 

We can conclude that Var(r) is less than Var(rHR) if β is closer to R then to . This result was given by Goodman and Hartley (1958). Note that the efficiency of rHR over rc depends upon the nature of the relation between Y and X and R. Also note that it does not provide satisfactory variance estimator, as approximation is applied as in case of classical ratio estimator. Goodman and Hartley (1958) also derived exact formula for variance.

**6.8 RATIO ESTIMATOR AS MODEL-UNBIASED**

Cochran (1953) has given a best and simple definition of best linear unbiased estimator. Consider all estimators  of Y that are linear functions of sample values yi, that are of the form

  (6.8.1)

where the c does not depend on  though they may a function xi. The choice of the  restricted to those that give unbiased estimation of Y. The estimator with the smallest variance is called **best linear unbiased estimator**. The model is:

  (6.8.2)

where εi are independent of the xi and xi are > 0. The xi (i = 1, 2, …… N) are known. The model is the same that was employed by Cochran (1953), which appears to have been originated by H.F. Smith (1938). Useful references to this model are Cochran (1953, 63, 77), Brewer (1963b), Rao (1966), Sarndal and Wright (1984), Godambe and Joshi (1965), Hanif (1969) Foreman and Brewer (1971), Royall (1970). Royal and Herson (1975),Brewer and Hanif(1983) Brewer, et al (1988), Cassel, et al (1976), Brewer (1979), Isaki and Fuller (1982), Hansen, Madow and Tepping (1983), smith (1991), Sarndal et al (1992), Samiuddin et al (1992)and many others.

Brewer (1963b) defined an unbiased ratio estimator under model (6.8.2). He used the concept of unbiased ness which was different from that given in randomization (design - based) theory. Royall (1970) also used this model. Brewer and Royall regarded an estimator  (estimated population total) is unbiased if  in repeated selections of the finite population and sampled under the model. Under model (6.8.2) Brewer (1963b) proved that the classical ratio estimate was model – unbiased and is best linear unbiased estimator for any sample [random or not] selected solely according to the values of the Xi. This result hold goods if the following line conditions are satisfied;

1. The relation between estimated (yi) and benchmark (xi) is linear and passes though the origin.
2. The Var(yi) about this line is proportional to xi.

**THEOREM 6.6**

Under the model (6.8.2) classical ratio estimator is unbiased with variance

=  (6.8.3)

**PROOF:**

We know that

  (8.6.4)

Using model (6.8.2) we have

 

Since E(εi) = 0 we then have

  (6.8.5)

We also know that

  or  (6.8.6)

Now

 

 (6.8.7)

Therefore we say that  is model unbiased if

 (6.8.8)

The variance expression of , i.e.

  (6.8.9)

 

Using the condition of model we will have:

  (6.8.10)

Using (6.8.2), (6.8.5) and (6.8.9) in (6.8.9), we will have

 (6.8.11)

Let us for simplicity we assume  then (6.8.11) will be:

  (6.8.12)

We can minimize  w.r.t. ci. For this the Langrang’s multiplier will be

 

Differentiating unconditionally with respect to ci, we get.



or 

We know from (6.8.7) that 

or

 

Hence

 



The best linear unbiased estimator , which is a classical (conventional) ratio estimator.

For the derivation of  we proceed as follows:

 

Since  and 

Divide  into sample and non-sample values we have

 

or

 

Squaring and taking the expectation





Substituting the value of Var(xi), we have:



   (6.8.3)

Using all these assumptions a model-unbiased estimator  from the sample may be easily proved as

 . (6.8.13)

Putting this value of  in (6.8.4) a model-unbiased variance estimator is

  (6.8.14)

This model based unbiased estimator is not only superior to  but is the best of a whole class of estimators. For details see Brewer (1963b, 1979), Royall (1970), Royall and Herson (1973) and Samiuddin, Hanif and Asad (1978).

* 1. **COMPARISON  AND  UNDER STOCHASTIC MODEL**

It is an established fact that the choice of a suitable sample plan is central to the design of a sample survey. Sample design can be regarded as comprising separate selection and estimation procedures, but the choices of these are so interdependent that they must be considered together for virtually all purposes. Some times the nature of the sample plan is determined by circumstances, but usually the designer is faced with a choice, and frequently it is obvious which of a number of possible plan will be most efficient in terms of minimum sample error for given cost( or vice versa). Standard sampling theory using imputed values for such quantities as the means, variances, and correlation coefficient of the (finite) population, or strata or clusters within it, can often indicate which design is most efficient. Sometimes, however, this is not so. A well-known example is the comparison between classical ratio estimation using unequal probabilities. To obtain a straight forwarded answer in this case, Cochran (1953) made use of a certain super population (6.8.2) which is intuitively attractive and appears to have some empirical basis. The purpose here is to compare classical ratio estimator and unbiased estimation method of estimation using equal probabilities and using large scale sample result which can be obtained using generalization of model. Comparison for probability proportional to size will be discussed in Chapter 7, 8 and 9. The stochastic model used here for the purpose of comparing efficiencies.

**6.9.1. Unbiased Estimate for Population Total Based on Simple Random Sampling**

We know that:

 

Putting the value of (6.8.2) we get



= (6.9.1)

Also

  (6.9.2)

or

  (6.9.3)

 





 (6.9.4)

as cross product term is equal to 

Now on first term of (6.9.4) using (6.8.2) will be





 (6.9.5)

Similarly

 

 or

  (6.9.6)

Using (6.9.5) and (6.9.6) in (6.9.4) we get:

  (6.9.7)

**6.9.2. Ratio Estimate**

  (6.1.3)

 

  (6.9.8)

Now

 



 (6.9.9)

Now





  (6.9.10)

Comparing (6.9.7) and (6.9.11) we have:

 



So Ratio Estimator will always be more efficient if is positive or



 0r 

Foreman and Brewer (1971) combined the model with the intercept i.e.

 Yi =α +β Xi +€i

With the same assumption given in (6.8.2) and compared various method of estimation. They provided that ratio method of estimation is more efficient than unbiased estimation method provided | α | < | βX | [Foreman and Brewer-1971].

**6.10. RATIO ESTIMATOR IN STRATIFIED RANDOM SAMPLING**

Ratio estimation method may also be applied to stratified random sampling as **stratum by stratum** (Separate Ratio Estimate) and **across stratum** (Combined Ratio Estimate).

The stratum by stratum ratio estimator  is

  (6.10.1)

This is an aggregate over strata to yield an estimate of population aggregate value, Y. Alternatively, a ratio estimator may be expressed directly in terms of r and the population benchmark X. This is called across-stratum ratio estimator and is expressed as:

  (6.10.2)

The m.s.e. of stratum by stratum,  in simple random sampling without replacement may be written in a straightforward manner via the concept of stratified random sampling.

 (6.10.3)

  (6.10.4)

An approximate variance estimator of (6.10.3) and (6.10.4) are

 (6.10.5)

  (6.10.6)

The m. s. e. of  is

(6.10.7)

  (6.10.8)

An approximate variance estimator of (6.10.7) and (6.10.8)



(6.10.9)

and

 (6.10.10)

respectively.

The difference of (6.10.4) and (6.10.8) is worth noting

 



 (6.10.11)

Now the second term on the right hand side is usually smaller under the situation when the ratio estimation is applicable, it is obvious that the first term is always positive. This indicates that  may be smaller than that of . The bias of the stratum by stratum ratio estimator is the sum of the biases associated with the estimate of each stratum, a decreasing function of nh­­ while the bias of the across stratum ratio estimator is a similar decreasing function of n. The general conclusion that could be drawn from the above is that stratum by stratum ratio estimator is recommended only when the sample size is large within each stratum and when the sample size is small, across stratum method is recommended for applications.

A ratio estimate based on a stratified sample can be more efficient than one based on a simple random sampling based on the same sample sizes except for stratum-by-stratum ratio estimates based on very small samples. Significance gains in the efficiency of ratio estimates can result, however, if sampling units are stratified on the basis of a suitable measure of size other than that used as the ratio estimation bench mark variable.

Stratum by stratum ratio estimator is usually more precise than the corresponding across-stratum ratio estimator except when stratum sample sizes are small. A cross-stratum ratio estimators must necessarily be used when the population bench mark X is available but not the stratum bench marks Xh.

The advantage of estimating across strata is purely to reduce the bias introduced by ratio estimator. The bias is appreciable only when the sample sizes within a stratum are very small. The more the variation in Rh, the greater the gain for use of stratum by stratum estimators.

# Example 6.2

A pilot survey was conducted to estimate the total number of orange trees in Sargodha districts. The district was divided into two stratum according to the information provided by Revenue circle. A simple random sample was used to select the sample. Following information are available:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Stra-tum** | **Total number of village** | **Sample village** | **Total area under orange trees (Acres)** | **Area under orange trees (Acres)** | **Total number of trees** |
| 1 | 632 | 8 | 25672 | 15.2, 3.2, 4.5, 10.8, 5.6, 9.8, 15.9, 21.6 | 940, 380, 415, 860, 410, 640, 815, 1120 |
| 2 | 730 | 10 | 20490 | 7.3, 1.2, 5.3, 2.1, 8.5, 7.8, 10.50, 2.8, 3.6, 5.1 | 480, 50, 230, 115, 500, 374, 516, 315, 315, 330 |

Estimate the total trees in 1362 villages using separate and combined ratio estimate and compare the efficiency of these two methods.

Solution;

 X1 = 25672, X2 = 20490, X = 46162 N1 = 632, N2 = 730, N = 1362

 n1 = 8, n2 = 10, n = 18

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stratum | Nh | Xh | nh |  |  | rh |  |  |
| 1 | 632 | 25672 | 8 | 10.825 | 697.5 | 64.43 | 41.071 | 77728.6 |
| 2 | 730 | 20490 | 10 | 5.42 | 322.5 | 59.50 | 9.291 | 24716.1 |
|  | 1362 | 46162 | 18 |  |  |  |  |  |

= 496.51, xst = 7.93,  = 6.4086, = 278.798, = 3.048

rc = 62.54, = 157.214 ρ1 = 0.9567, ρ2 = 0.888

Total number of trees

(i) Separate Ratio Estimate () =  = 2873347

(ii) Combined Ratio Estimate  =  = 2890276

(iii) = using equation (6.10.6) = 1744094849

(iv)  = 41762.36

(v) = using equation (6.10.10) = 1116831797

(vi) = 33419.033

The efficiency of  to 

= 

**6.11. JACK KNIFE AND BOOTSTRAP METHODS**

* + 1. **Jack Knife**

Jack Knife methods are developed by Quenouille (1956) and Turkey (1958) for celebrating standard, error bias and confidence limits when there is no reliable information of the model available.

Suppose a sample of observations ….. from a population with unknown parameter , Suppose, the estimation of  is based on the sample. The probability distribution is unknown. If sample size is large, then the random variable  is approximately normal with mean zero and various 1 where  and 

E (T) and Var(T) may be expanded as

E(T) =  and

Var(T)= 

Where  and, i=1,2….. are unknown.

One method is to divide the sample in to two equal sub-samples and to calculate  and  for  the bias and variance in  are

Bias =

Var(T) = 

 Thus Bias = 

and  to the order .

In order to improve upon these bias and variance, Quenouille (1956) and Tukey (1958) and many others provide a technique called Jack knife.

Consider  to be a relevant estimation of  based on a sample of size n.

We define  to be another estimation of  based on n – 1 observations by excluding ith observation. Then, we define a new estimator

 

 Where 

We find

 

This leads to 

Where  then 

is known as Jack Knife estimation of the Variance of .

The Bias is estimated as

 Bias = 

The Jack Knife method provides confidence limit for  as



Where is the point of the standard normal distribution, and



For example, consider  then

reduces to



**Example 6.3**

Suppose a sample of size 5 is taken from a population with finite mean 4. The observations are 2, 3, 6, 5, 4.



Let all possible sample of size *n-1* when one is excluded

2,3,6,5 excluding 4 16/4=4

2,3,6,4 excluding 5 15/4=3.75

2,3,5,4 excluding 6 14/4=3.50

2,6,5,4 17/4=4.25

3,6,5,4 18/4=4.50



Bias = 0 this is so as we drew all possible samples of size 4.



**Example 6.4**

Suppose a sample of size 5 is drawn from a population with mean  and Variance  . Suppose  is to be estimated. The observations are

 2, 3, 6, 5, 4

**Solution.** Let the  be estimated by 

The calculations are made to find Jack Knife bias and Variance.

S.No.    

1 4 4.0 0 0

2 5 3.75 1.0 1

3 6 3.50 2.0 1

4 3 4.25 -1.0 1

5 2 4.50 -2.0 4

 = 0.0 10.0

Where









 5

**Example 6.5:**

 Suppose *r* is the condition for pairs of 

Table 1

j    

1

2

3

4

5

X denotes scores on Test A

Y denotes scores on Test B

r denotes the sample correlation itself



 Correlation coefficient when  removed from sample





Now we find r = 0.776 and =1.0352



**Example 6.6**

Suppose pairs of observations are collected, where

 Y denotes average score on Test A

 Z denotes average score on Test B

j     

1 52 3.4 0.89 5(0.78-0.89)=-1.62 1.43 -1.12

2 64 3.3 0.76 0.18 1.00 045

3 56 2.8 0.76 0.29 0.98 .71

4 58 3.3 0.78 0.00 1.04 0

5 67 3.4 0.78 0.63 0.93 1.46

6 58 3.1 0.78 -0..6 1.05 0.14

7 -0.11 1.06 -.29

.

.

.

15

r = 0.78, = 1.035

r Scale 

T Scale  = = -.098



Bias = 

**6.11.2 Influence Function Explained in Robust Estimation**

….. from  t is an estimator of , we want to find Bias, standard error and 95% C.I

 is unknown and moments of T is not possible

Suppose, we find

E(T) =  and

Var(T)=  -----🡪 **(1)**

As for large n  follow *N(0,1)*  randomly.

 Now split the data randomly in to two equal parts. Calculate estimator and  of  and estimate Bias and Variance of **(1)** by 

Let  denote the sample statistic based on a random sample of size *n*.

 Denotes the sample statistics based on *n-1* values excludes.

Let

 Where







Where 

 is known as Jack Knife estimation of the Variance of 

**6.11.3. Bootstrap**

 A new recent & difficult approved bootstrap method. Given *n* sample values we define empirical frequency function  which has the Values  at each 

 at 

 = 0 elsewhere

The bootstrap estimate of sampling variance of a statistic by selecting a sample of size *n* with replacement from the population estimator by and computing  from this sample;

This is done M times, going  for j = 1, 2, 3, ….. M

The bootstrap estimator is



and estimator of variance



**6.12 SOME OTHER RATIO ESTIMATORS**

In this Section, some ratio estimators are presented some of which are unbiased and some are approximately unbiased.

**(a) Hartely and Ross Unbiased Ratio Estimator:**

Hartley and Ross (1954) presented unbiased ratio type estimator population total Y as

 , (6.7.5)

where

 

**(b) Mickey’s Ratio Estimator:**

Mickey (1959) has given an unbiased ratio of estimator. In this estimator the sample units are divided into k groups of size m each such that n = mk. An unbiased ratio estimator of population total Y is

, (6.12.1)

where

 classical ratio estimator computed from the sample after omitting the jth groups for n = 2,  reduced to .

**(c) Quenouille-Durbin Ratio Estimator:**

Quenouille (1956) has given a ratio estimator

 (6.12.2)

Cochran (1963) used modified from of  which is simplier and more convenient

 (6.12.3)

**(d) Ratio Estimator Based on Group Sample Means:**

, (6.12.4)

where

 .

 reduces to  if n = 2

Tin (1965) used modified form of 

 (6.12.5)

In fact he used  in place of T in (6.11.4).

**(e) Pascual’s (1961) Ratio Estimator:**

Pascual (1961) has given another type of ratio estimator i.e.

 (6.12.6)

7

**(f) Beale-Tin Ratio Estimator:**

These estimators are

 (6.12.7)

and

 (6.12.8)

Rao (1969) conducted empirical study to compare the efficiencies of these ratio estimators. He considered sample size n = 2 , 4, 6, 8,12 from the 20 natural and artificial populations. There is no definite conclusion which estimator is superior to another. It depends upon the nature of the population and also on sample size. In the example 6.3 and 6.4 given below we have only explained how the calculation for different estimators can be worked out.

**Example 6.7**

A sample of 15 villages is drawn from a population of 50 villages and the following information for 15 villages is recorded xi, denotes the population for the previous year and yi denotes the population of the current year under estimate. Using this information estimate the total population of 50 villages. Compare this estimate with simple random sample. Also find the standard error in each case. [The true total,
ΣXi = 4746 and ΣYi = 5431].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi |  | yi |  |  |  |  |
|  |  |  |  |  |  |  |
| 65 |  | 79 |  | 1.215 |  |  |
| 73 | 70.33 | 96 | 85.00 | 1.315 | 1.111 |  |
| 73 |  | 80 |  | 1.096 |  |  |
|  |  |  |  |  |  |  |
| 102 |  | 104 |  | 1.020 |  |  |
| 74 | 88.00 | 85 | 96.33 | 1.149 | 1.133 |  |
| 88 |  | 100 |  | 1.136 |  |  |
|  |  |  |  |  |  |  |
| 76 |  | 96 |  | 1.236 |  |  |
| 73 | 85.00 | 85 | 98.33 | 1.164 | 1.142 | 1.141 |
| 106 |  | 114 |  | 1.075 |  |  |
|  |  |  |  |  |  |  |
| 97 |  | 116 |  | 1.196 |  |  |
| 112 | 104.33 | 121 | 115.33 | 1.080 | 1.156 |  |
| 104 |  | 109 |  | 1.048 |  |  |
|  |  |  |  |  |  |  |
| 116 |  | 126 |  | 1.086 |  |  |
| 136 | 119.67 | 150 | 131.00 | 1.103 | 1.163 |  |
| 107 |  | 117 |  | 1.093 |  |  |

# SOLUTION

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | = | 1402 |  | = | 1578 |  | = | 17.002 |
|  | = | 137018 |  | = | 171378 |  | = | 1.133 |
|  | = | 93.467 |  | = | 105.2 | r | = | 1.126 |
|  | = | 152941 |  | = | 0.962 |  |  |  |
|  | = | 398.516 |  | = | 1.132 |  |  |  |
|  | = | 363.373 |  | = | 358.16 |  |  |  |

The estimated values of the populations using the estimators are given below:

|  |  |  |
| --- | --- | --- |
| Estimators |  | Estimated values |
|  |  |  |
| Classical Ratio |  | 5344 |
| Hartley – Ross |  | 5342 |
| Mickey’s |  | 5360 |
| Quenouille-Durbin |  | 5144 |
| Cochran |  | 5059 |
| Grouped Sample Mean |  | 5339 |
| Tin |  | 5336 |
| Beale-Tin |  | 5355 |

**6.13 regression estimate**

When the relation between variable under study and benchmark variable is linear and does not pass through the origin, we use another method of estimation, called **regression estimate**. Ratio and regression estimates are identical when the regression line passes through the origin. In fact ratio estimate is a special case of regression estimate.

If (y1, x1) (y2, x2), . . . , (yn, xn) are the sample values, based on simple random samples, drawn from a population of N units, then the classical regression estimator  of population mean  is defined as:

  (6.13.1)

where byx is the regression coefficient y on x and is

  (6.13.2)

or

 (6.13.3)

6.14 EXPECTATION, VARIANCE AND VARIANCE ESTIMATOR
OF REGRESSION ESTIMATE

**THEOREM (6.8)**

A simple random sample of size n is drawn from a population of size N, the sample mean based on regression estimate is an unbiased of population mean i.e.

  (6.14.1)

with variance

  (6.14.2)

###### PROOF

We know that

  (6.13.1)

Taking the expectation and using the concept of simple random sample, we have

 

Hence

 ◊

Note that the unbiasedness does not depend on the value of byx.

The variance of  is

 

Substituting the value of  from (6.12.1) in the above expression and re-arranging the terms





 (6.14.2)

Since  is an unbiased estimator for all values of byx, now the question arises, what should be the best value of byx so that  be minimum. For this finding the partial differentiation of (6.14.2) w.r.t. byx and equating to zero we have

  (6.14.3)

or

  (6.14.4)

where βYX is population regression coefficient.

Using (6.14.4) in (6.14.2) and on simplification, we get

  (6.14.5)

which takes minimum value.

The estimated population total

  (6.14.6)

and variance of  is

  (6.14.7)

and

  (6.14.8)

If we compare (6.2.14) and (6.14.8), we get:

 

and say

 iff  < 0

or

  (6.14.9)

both become identical if  which is the optimum value of byx as it minimizes the . Comparing simple random sampling with regression estimate:

  (6.14.10)

Hence  is more efficient than . Both are identical if  = 0.

**6.14.1 Unbiased Variance Estimator**

An unbiased variance estimator of (6.13.2) may be written in a straight forward way as

  (6.14.11)

For computation purposes the most convenient for of (6.13.3) is

  (6.14.12)

An unbiased estimator if (6.14.5) [using the concept of simple random sampling] is

  (6.14.13)

An unbiased variance estimator of (6.14.7) and (6.14.8) are

  (6.14.14)

and

  (6.15.15)

respectively.

**6.15 STANDARD ERROR AND CONFIDENCE LIMITS OF  and **

The standard error of  and  is

 , (6.15.1)

and

 . (6.15.2)

The confidence limits of  and  is

  (6.15.3)

and

  (6.15.4)

# Example 6.8

From the data given in example 6.1 estimate total number of persons using the method of regression estimate and find the standard error of your estimate. Compare your estimated values with the one that obtained from ratio estimate.

# Solution

We know that



Total number of persons







We can easily compare, andas

|  |  |  |  |
| --- | --- | --- | --- |
|  | Regression | Ratio estimate | Simple estimate (SRS) |
| Estimated total | 16007 | 15885 | 17250 |
| Estimated variance | 485031.4398 | 543874.826 | 4396622.193 |
| Standard error | 696.442 | 737.4786 | 2090.364 |

We can see that:

 <  < 

**6.16 REGRESSION ESTIMATE AS MODEL-UNBIASED**

# THEOREM 6.9

In a finite population values Yi (i = 1, 2, . . . . . N) are randomly drawn from super population in which

 yi = α + β xi + εi (6.16.1)

where E(εi) = 0 = E = 0 and  for fixed values of xi then  is model unbiased for any sample size and the variance of  is

  (6.16.2)

**PROOF**

We know that

   (6.16.3)

Substituting the value of yi from the model (6.16.1) and on simplification we get

  (6.16.4)

Since  is unbiased estimator of  which is zero. Thus  is distributed normally about zero mean is repeated samples. This is of the order  as standard error of sample covariance is of the order ,  is of the order unity. Therefore, (b - β) is of the order .

**THEOREM 6.9**

For a simple random sampling the mean of the regression estimator under (6.16.1) of the regression estimate is unbiased i.e.  with variance

  (6.16.5)

**PROOF:**

We know from the model (6.16.1)

 

or

1

we also know that

 

or

 

Substituting the value of  from the model and (b – B) from (6.16.4) and on simplification

  (6.16.6)

Taking expectation

 

Since

 , therefore 

Hence  is model unbiased.◊

Also squaring (6.16.6) and taking expectation under the model.

 

 

Therefore the 

  (6.16.2)

**6.16.1 Unbiased Variance Estimator**

For unbiased variance estimator we know that

 

and 

From this

 

or

 

We have proved that b - β is of the order . Hence we left

 

or

 

Since , Hence

 

is model unbiased.

**6.17 DIFFERENCE ESTIMATOR**

We have already seen in Section (6.8) that ratio estimator is best linear unbiased estimator if the relation between estimand (Yi) and benchmark (Xi) variables are linear and passes through the origin, i.e. Yi - kXi = 0, where k is constant. In practical life such type of relation is not always possible. There might be a situation when the relation between Yi and Xi of the form Yi - k Xi = a, where a is constant. We wish to estimate  and . When xi and yi are correlated then estimator  may be improved by introducing a factor known as difference **factor**. If it is assumed that there is a unit charge in y when a unit charge is made in x, then simple difference function may be introduced as;

  (6.17.1)

It is further assumed that x and y have equal variances. Let we define a more generalized from of (6.17.1) as

 , (6.17.2)

where  and k are known.

**THEOREM 6.9**

The difference estimator (6.17.1) is an unbiased estimator of population total Y with variance is:

  (6.17.3)

**PROOF**

Taking the expectation of (6.17.2)

 

Using the concept given in the simple regression on can write



Taking the square and expectation both sides



 (6.17.3)

An optimal value of k can be obtained by minimizing (6.17.3). For this we find partial differentiations w.r.t. k and equating to zero.

  = 0 (6.17.4)

Therefore, 

Putting the value of k in (6.17.3), we will have

 

If k = R, then

 

which is variance of  for ratio estimator.

If K = 0, then 

An unbiased estimator of (6.17.3) is

  (6.17.5)

The difference estimator is superior to  if [from (7.17.3) and (3.4.1)]

 

or

 

Therefore k lies between 0 and 2 β or k is outside the range,  is superior to . For details Zarkovic (1956).

**6.18 BIAS OF REGRESSION ESTIMATE**

If we use a simple random sample then linear regression estimate is biased i.e., the bias of the regression estimate is trivial and decreases as size of the sample increases. The bias comes in  by two reasons.

1. β is estimated by the ratio of  to that of 
2. It involves the product of b and which are estimates.

**THEOREM 6.10**

The bias in regression line using simple random selection proceed is .

**PROOF**

We know that in the regression estimates processes,  are involved. Suppose

 

and

 

where

 

Putting these values in (6.12.1) we have

 

Further

 

Since E(ε) = E(ε1) = 0.Therefore,

  (6.18.2)

If  then regression estimate reduces to ratio estimate then bias in ratio estimates takes the form

  (6.18.1)

6.19 REGRESSION ESTIMATE IN STRATIFIED RANDOM SAMPLING

We know that

  (6.12.1)

For hth stratum

  (6.19.1)

The regression estimate for estimated total  for separate (stratum by stratum) is

  (6.19.2)

with variance [may be written in straight forward way]

 (6.19.3)

 (6.19.4)

If bh = βh, then

  (6.19.5)

Similarly combined (across stratum) regression estimate is defined as

  (6.19.6)

bc is combined regression coefficient

The variance expression of  may also be written in a usual way i.e.

  (6.19.7)

If bc = βc then (6.19.7) gives minimum variance

  (6.19.8)

From (6.19.5) and (6.19.9) we have

 

 

 

  (6.19.9)

The R.H.S. of (6.19.9) is positive. Hence

 

This is true unless βh is the same for all the strata.

In fact no hard and fast rule can be given for the efficiency over each other. However, only rough idea can be made. If regression is linear in all the strata and there is small variation in byhXh then across stratum regression estimate is advisable for application but if there is a large variation byhXh from stratum to stratum then stratum by stratum regression estimate can be recommended for application. If the regression is non-linear, across stratum regression estimate is always better for practical application. Some further remarks regarding these are as:

1. Separate regression estimates is appropriate when true regression coefficient βh vary markedly from stratum to stratum.
2. Combined regression estimate is appropriate when βh is presumed to be the same in all the strata.
3. Separate regression estimate is more close to bias, when sample are small within the strata and variances have larger contribution from sampling error in the regression coefficient.
4. In combine regression estimate variance is inflated if the population regression coefficient differs from stratum to stratum.

##### Example 6.7

From the data given in Example 6.2 estimate the total number of trees by using separate and combined regression method of estimation and calculate the variance for each case.

##### Solution

From the data given in Example 6.2, we have

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| b1 | = | 41.6511 | N1 | = | 632 | N | = | 1362 |
| b2 | = | 47.6659 | N2 | = | 730 | n | = | 18 |
| bc | = | 49.2153 | n1 | = | 8 | n2 | = | 10 |
|  |  |  |  |  |  |  |  |  |
|  | = | 496.51 |  | = | 7.93 | ρ1 | = | 0.957 |
| ρ2 | = | 0.884 |  | = | 6.4087 |  | = | 2.995 |
| sy1 | = | 278.798 |  | = | 157.214 |  | = | 40.62 |
|  | = | 28.068 |  |  |  |  |  |  |

(a) Estimation of total trees

 

 

 

 

Comparison of may be seem in the following table:

|  |  |  |
| --- | --- | --- |
|  | RATIO | REGRESSION |
|  | Separate  | Combined  | Separate | Combined  |
| Estimated Total | 2873347 | 2890276 | 2248616 | 2416579 |
| Estimated Variance | 1744094849 | 1116831797 | 798140486.2 | 736067487.7 |
| Standard Error | 41762.36 | 33419.033 | 28251.3803 | 27130.5637 |

**exercises**

6.1 From the following population draw all possible samples of size 2 and calculate . Also find the bias in case of ratio estimate.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Population Units | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| YI | 1 | 3 | 4 | 6 | 7 | 8 | 10 | 11 |
| XI | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

6.2 A sample of 34 villages was selected from a population of 170 villages to estimate the areas under Wheat in Sahiwal District during 1985. The total cultivated area under wheat during 1981 was 20,820 acres. It was found that for 1983 it was 21288. Estimate the area under wheat for 1985 sing 1981 and 1983 as bench mark variable. Compare these two estimated number. Use classical ratio estimate. Compute the standard error for both the cases.

|  |
| --- |
| Area Under Wheat (In Acres) |
| S.No. | 1981(x1) | 1983(x2) | 1985(y) |  | S.No. | 1981(x1) | 1983(x2) | 1985(y) |
| 1 | 401 | 70 | 50 |  | 18 | 186 | 45 | 27 |
| 2 | 630 | 163 | 149 |  | 19 | 1767 | 564 | 515 |
| 3 | 1194 | 284 | 284 |  | 20 | 604 | 238 | 249 |
| 4 | 1170 | 440 | 381 |  | 21 | 700 | 92 | 85 |
| 5 | 1065 | 250 | 278 |  | 22 | 524 | 247 | 221 |
| 6 | 827 | 125 | 111 |  | 23 | 571 | 134 | 133 |
| 7 | 1737 | 558 | 634 |  | 24 | 962 | 131 | 133 |
| 8 | 1060 | 254 | 278 |  | 25 | 407 | 129 | 103 |
| 9 | 360 | 101 | 112 |  | 26 | 715 | 190 | 175 |
| 10 | 946 | 359 | 355 |  | 27 | 845 | 363 | 335 |
| 11 | 4170 | 109 | 99 |  | 28 | 1016 | 235 | 219 |
| 12 | 1625 | 481 | 498 |  | 29 | 184 | 73 | 62 |
| 13 | 827 | 125 | 111 |  | 30 | 282 | 62 | 79 |
| 14 | 96 | 5 | 6 |  | 31 | 194 | 71 | 60 |
| 15 | 1304 | 427 | 339 |  | 32 | 439 | 137 | 100 |
| 16 | 377 | 78 | 80 |  | 33 | 854 | 196 | 141 |
| 17 | 259 | 75 | 105 |  | 34 | 820 | 255 | 263 |

6.3 From Question 6.2, estimate the area under wheat for 1985 using regression method and using 1981 as bench mark variables. Also compute standard error under both cases.

6.4 The number of cows in milk enumerated (y) from a random sample of 20 villages from a tehsil having 84 villages, as also the corresponding census figures (x) in the previous year, are given below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| villages | y | x |  | villages | y | x |  |
| 1 | 237 | 155 |  | 11 | 813 | 616 |  |
| 2 | 1060 | 583 |  | 12 | 666 | 576 |  |
| 3 | 405 | 205 |  | 13 | 681 | 540 |  |
| 4 | 1085 | 738 |  | 14 | 2743 | 2242 |  |
| 5 | 666 | 526 |  | 15 | 1228 | 940 |  |
| 6 | 542 | 284 |  | 16 | 472 | 387 |  |
| 7 | 1337 | 758 |  | 17 | 643 | 675 |  |
| 8 | 1166 | 681 |  | 18 | 180 | 220 |  |
| 9 | 399 | 143 |  | 19 | 583 | 654 |  |
| 10 | 228 | 111 |  | 20 | 1195 | 1787 |  |

 Given that the census estimate f the number of cows in milk in the tehsil was 74488, estimate the number of cows in milk in the current year with and without using the census information and compare the efficiencies of the estimates.

6.5 A sample survey for the study of yield of orange was conducted. From 146 villages a simple random sample of 13 village was selected. The data is given as:

|  |  |  |
| --- | --- | --- |
| S.No. of villages | Total No. of orange trees | Area under orange orchards (in acres) |
| 1 | 492 | 4.80 |
| 2 | 1008 | 5.99 |
| 3 | 714 | 4.27 |
| 4 | 1265 | 8.43 |
| 5 | 1889 | 14.39 |
| 6 | 784 | 6.53 |
| 7 | 294 | 1.88 |
| 8 | 798 | 6.35 |
| 9 | 780 | 6.58 |
| 10 | 619 | 9.18 |
| 11 | 403 | 2.00 |
| 12 | 467 | 2.20 |
| 13 | 197 | 1.00 |

 Given the total area under orange orchards of 146 villages in 354.78 acres, estimate the total number of orange trees in the Tehsil along with its standard error using the area under orange orchards as the auxiliary variate. Discuss the efficiency of your estimate with the one which does not make any use of the information on the auxiliary variate.

6.6 For estimating the total cattle population, a random sample, wr, of 24 villages was selected from the total 1238 villages. The number of cattle obtained in the survey is given below for each sample village, together with the corresponding census figures relating to a previous period.

|  |  |  |  |
| --- | --- | --- | --- |
| S.No. of Villages | Number of cattle | S.No. of Villages | Number of cattle |
| Census | Survey | Census | Survey |
| 1 | 623 | 654 | 13 | 706 | 707 |
| 2 | 690 | 696 | 14 | 1795 | 1890 |
| 3 | 534 | 530 | 15 | 1406 | 1123 |
| 4 | 293 | 315 | 16 | 118 | 115 |
| 5 | 69 | 78 | 17 | 330 | 375 |
| 6 | 842 | 640 | 18 | 218 | 212 |
| 7 | 475 | 692 | 19 | 160 | 147 |
| 8 | 371 | 292 | 20 | 210 | 297 |
| 9 | 161 | 210 | 21 | 262 | 401 |
| 10 | 298 | 555 | 22 | 262 | 401 |
| 11 | 2045 | 2110 | 23 | 185 | 199 |
| 12 | 1069 | 592 | 24 | 574 | 564 |

 Compare the efficiency of the regression estimator with the ratio estimator. It is given that the number of cattle for the previous period of 1238 villages is 680,900.

6.7 An eye estimate of the fruit weight (xi) on each tree in an orchard having 150 trees was made. The total weight X was found to be 16,600 K.G. A random sample of 10 trees was taken and actual weight of fruit ‘yi’ along with eye estimate were as

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Actual weight (yi) | 67 | 58 | 71 | 39 | 41 | 42 | 39 | 51 | 46 | 61 |
| Eye estimate weight | 67 | 55 | 78 | 40 | 45 | 46 | 40 | 56 | 48 | 59 |

1. Estimate the total actual fruit weight Y by taking simple difference estimator  and find the variance of .
2. Estimate the total actual fruit weight Y by taking regression estimator  and find the variance of .

6.8 In an experimental study in a large paddy field, the weight of grain plus straw (x) and the grain yield (y ) are obtained for each of the large number of sampling units located at random over a field. The following data were obtained:

 

 Compare the efficiencies of ratio method of estimation and that of unbiased estimators where cx, cy are the coefficient of variation of x and y respectively and cyx are coefficient of coveriation.

6.9 A sample of 34 villages were selected from a population of 140 villages for estimating the area under wheat. The following data are given:

 

 x denotes the area under wheat for 1973 and y denotes the area for 1975. Estimate the area under wheat for 1975 by ratio estimation method and compare it with unbiased estimate. The total area under wheat cultivation for 1973 was 21288 acres.

6.10 A trained investigator makes an eye estimate of the area each parcel in a commune containing 200 parcles. This exercise produces an area of 1,160 stemmas. The areas are actually measured on the basis of a random sample of 20 parcles with the following results.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parcel | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Estimated area | 4.8 | 5.8 | 6.0 | 5.9 | 7.6 | 6.7 | 4.7 | 5.8 | 4.4 | 5.2 |
| Actual Area | 4.5 | 5.3 | 5.8 | 6.1 | 7.1 | 6.7 | 4.2 | 5.7 | 3.9 | 5.0 |

 Estimate the total area in the commune using ratio estimate, regression estimate methods and compare the efficiency of these two methods with simple method of estimation.

6.11 From artificial population of strata 2; draw all possible samples of size 2 and find 

|  |  |
| --- | --- |
| Stratum 1 | Stratum 2 |
| X1i | Y1I | X2i | Y2i |
| 1 | 3 | 5 | 6 |
| 2 | 5 | 6 | 8 |
| 3 | 6 | 7 | 9 |
| 4 | 7 | 8 | 12 |

6.12 The following data were collected in a Pilot survey to estimate the production of fresh fruits in 3 districts of Sind.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Stratum No. | Total number of villages | Total area (acres) | Area under orchards | Total number of trees |
| 1 | 985 | 11253 | 10.63, 9.90, 1.45, 3.38, 5.17, 10.35 | 747, 719, 78, 201, 311, 488 |
| 2 | 2196 | 25115 | 14.66, 2.61, 4.35, 9.87, 2.42, 5.60, 4.70, 36.75 | 580, 103, 316, 739, 196, 235, 212, 1646 |
| 3 | 1020 | 18870 | 11.60, 5.29, 7.49, 7.29, 8.00, 1.20, 11.50, 1.70, 2.01, 7.96, 23.15 | 488, 227, 374, 491, 449, 50, 47, 879, 115, 115 |

 Estimated total number of trees by using

1. Separate ratio estimate.
2. Combined ratio estimate.
3. Separate regression estimate.
4. Combined regression estimate.

 Compare their efficiencies with stratified random sampling and simple random sampling.

6.13 If the regression of y on x is linear E(y/x) = ax + b then show that in simple random sample  give smaller variance than

 

6.14 Define ratio estimator for estimating the population total of character y and derive an expression for the standard error of the estimator. State the conditions under which the classical ratio estimator is best linear unbiased estimate.

6.15 If the coefficient of variation of bench mark variable x is more than twice the coefficient of variation of the variable y, then show that in large samples with simple random sampling, the classical ratio estimator  is less precise than unbiased estimator .

6.16 In sample simple random sampling without replacement the classical ratio is  obtain the exact expression for the variance of .

6.17 Values of y and x are measured for each unit in a simple random sample to estimate population ratio , which of the following estimators would you recommend to estimate R. (i) Always use  (ii) Always use  (iii) either use  or  depending on the conditions (given  is known) give reason for your choice.

6.18 Using the difference estimator

 

 and regression estimator

 

 an estimator is defined

 

 Show that

 ,

 

 and

 

 Hence or otherwise, prove that

 