Agri. Res. J. Kerala, 1979, 17 (2) 240—247

ESTIMATION OF PARAMETERS UNDER RANDOM DETERMINATIONS ON EACH UNIT OF A RANDOM SAMPLE

A, INDIRADEVI and P, U, SURENDRAN

College of Veterinary and Animal Science, Mannuthy, Kerala.

When several determinations could be made on unit of a sample, es in most of the agricultural sample surveys, Singh and Gupta (1978) have described a sampling procedure and recommended methods of obtaining an unbiased estimate of the population mean and an estimate of its variance. In establishing these results the authors have assumed that the determinations obey normal probability law and have made use of generating functions. In the present article assuming the same sampling procedure as recommended by Singh and Gupta we have obtained estimates of mean and its variance under more general conditions. The procedure suggested is also simple.

Materials and Methods

The sampling procedure described by Singh and Gupta is as follows. Consider a population of N units out of which a random sample is selected for experimentation. Let y be the character under study. On every unit selected j $(j \leq s)$ determinations of y are made. Assuming that j is a random variable we wish to obtain an unbiased estimate of the population mean and an estimate of the variance of this estimate. For example, estimating the total production of crops when harvesting is spread over a long period of about 2—3 months and involves a number of random pickings, estimating the mean effect of a treatment on diseased animals as measured by observations on certain characteristic the observations being taken till the first sign of recovery, estimating the true mean performance of students with the help of a random number of tests on each etc.

Let Y_{ik} be the k^{th} observation made on the i^{th} unit selected i = 1, 2, .. n and k — 1, 2,. . . . . j_i where j_i is the number of observations on the i^{th} unit selected and is a random value of j.

If we treat the number of determinations on a unit as possibly infinite, we may take the model $Y_{ik} = Y_i + \epsilon_{ik}$ Where Y_i is the true and unknown measurement of the i^{th} unit and ϵ_{ik}'s are random errors, following independent distributions with mean O and variance σ_ϵ^2 When $Y_i = \mu$ and ϵ_{ik}'s are normal we obtain the sampling situation described by Singh Gupta.

Consider the estimator $\bar{\bar{Y}}$ the mean of the sample means $\bar{Y}_i$'s given by

$$\bar{\bar{Y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{i.} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{j_i} \sum_{k=1}^{j_i} Y_{ik}$$

Theorem

The expected value and variance of the estimator $\bar{\bar{Y}}$ are given by

$$E(\bar{\bar{Y}}) = \mu \text{ and } V(\bar{\bar{Y}}) = \frac{1}{n} E\left(\frac{1}{j}\right) \sigma^2$$

Proof

$$E(\bar{\bar{Y}}) = \frac{1}{n} E_j E_{Y/j} \left\{ \sum_{i=1}^{n} \sum_{k=1}^{j_i} \frac{y_{ik}}{j_i} \right\} = \frac{1}{n} E_j \left(\sum_{i=1}^{n} \mu \right) = \mu$$

$\text{Var}(\bar{\bar{Y}})$

$$= \frac{1}{n^2} E_j \left\{ \sum_{i=1}^{n} \frac{V_y}{j^2} \left(\sum_{k=1}^{j} Y_{ik} \right) \right\} + \frac{1}{n^2} V_j \left\{ E_y \left\{ \sum_{i=1}^{n} \frac{1}{j} \sum_{k=1}^{j} Y_{ik} \right\} \right\}$$

$$= \frac{1}{n} E\left(\frac{1}{j}\right) \sigma^2 + 0 = \frac{1}{n} E\left(\frac{1}{j}\right) \sigma^2$$

To get an estimate of σ^2, if $Q = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{j_i} \frac{(Y_{ik} - \bar{Y}_i)^2}{j_i - 1}$

$$E(Q) = \sigma^2$$

$$\therefore \quad \text{Est Var}(\bar{\bar{Y}}) = \frac{1}{n} E\left(\frac{1}{j}\right) Q$$

These results hold in the general case when j follows any probability distribution. However when j follows some of the known probability distributions $E(1/j)$ may be actually evaluated and substituted, in the above expressions to get the estimates. Thus tables of $E(1/j)$ have been worked out for the following distributions.

1. When j is distributed equally integers $< S$ such that

$$P(j = x) = \frac{1}{S} \; ; \; x = 1, 2 \ldots S$$

Table 1 E (1/j) for the truncated uniform type discrete distribution

$$P(j = x) = \frac{1}{S} \; ; \; x = 1, 2 \ldots\ldots S$$

S	E (1/j)	S	E (1/j)	S	E (1/j)	S	E (1/j)
1	1	7	0.37,041	13	0.24,463	19	0.18,651
2	0.75,000	8	0.33,973	14	0.23,226	20	0.17,968
3	0.61,111	9	0.31,433	15	0.22,094	21	0.17,339
4	0.52,083	10	0.29,293	16	0.21,104	22	0.16,758
5	0.45,667	11	0.27,454	17	0.20,208	23	0.16,218
						24	0.15,716
6	0.40,833	12	0.25,860	18	0.19.393	25	0.15,247

Table 1 gives E (1/j) for S from 1 to 25

2. When j has the truncated binomial distribution

$$P(j = x) = \frac{1}{1 - q^{s}} \binom{S}{x} p^{x} q^{s-x} \quad ; \quad x = 1, 2 \ldots S$$

Table 2 gives E (1/j) for S from 1 to 15

3. When j has the truncated geometric distribution

$$P(j = x) = \frac{1}{1 - q^{s}} \; q^{x-1} \; p \; ; \; x = 1, 2 \ldots S$$

Table 3 gives E (1/j) for S from 1 to 15

4. When j has the truncated Poisson distribution

$$P(j = x) = \frac{1}{\sum_{x=1}^{s} \frac{m^{x}}{x!}} \times \frac{m^{x}}{x!} \; ; \; x = 1, 2 \ldots S$$

Table 4 gives E (1/j) for S from 1 to 10

The truncation was done at 0 to avoid non-observation of selected units.

For the estimator $\hat{\mu}$ of Singh and Gupta given by

$$\hat{\mu} = \frac{1}{nm} \frac{f(m)}{f(m)} \sum_{i=1}^{n} Y_i . \quad \text{Where } f(m) . \sum_{J=0}^{S} \frac{m^{i}}{J!}$$

$$\mathrm{Var}\,(\hat{\mu}) = \frac{1}{n\,m^2}\left[\frac{f(m)}{f'(m)}\right]^2 E(j)\,\sigma^2$$

Hence $\bar{\bar{Y}}$ is superior to $\hat{\mu}$ if

$$\frac{1}{m^2}\left[\frac{f(m)}{f'(m)}\right]^2 E(j) > E\left(\frac{1}{j}\right)$$

Results and Discussion

From the tables it may be seen that the estimator $\bar{\bar{Y}}$ is superior to $\hat{\mu}$ of Singh and Gupta for small values of m (less than *2).* Also, the estimator $\bar{\bar{Y}}$ and its variance are derived for general situations as ϵ_{ik}'s are not restricted to be normal.

These results may be extended to the cases where we wish to estimate the true mean of a finite population of N units. For example if Y_{ik} represents the score of the i^{th} individual in the k^{th} test and if *y,* represents his true score we might wish to estimate the true mean score of N individuals of whom n are sampled, the decision being made on random number of tests on each. Another example is when the data of crop for n selected fields out of N might be available for different randomly selected seasons and we wish to estimate the true mean or total yield of the population of N fields. Here $\bar{\bar{Y}}$ is an unbiased estimator for the parameter μ where

$$\mu = \frac{1}{N}(y_1 + Y_2 + \ldots\ldots + y_N)$$

Also if $\sigma^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(y - \mu)^2$

$$\mathrm{Var}\,(\bar{\bar{y}}) = \frac{N-n}{N\,n}\,\sigma^2 + \frac{1}{n}\,E(1/j)\,\sigma_e^2$$

$$\text{If } Q = \frac{1}{n}\sum_{k=1}^{n}\sum_{k=1}^{j_i}\frac{(y_{ik} - \bar{y}_i)^2}{j_j - 1}$$

$$R = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{\bar{y}})^2$$

$$\mathrm{Est\ Var}\,(\bar{\bar{y}}) = \frac{N-n}{N\,n}\,R + \frac{Q}{n}\,E(1/j)$$

Table 2 E(1/j) for the truncated binomial $P(j = x) = \frac{1}{1-q^s}\binom{s}{x} p^x q^{s-x}$, $x = 1, 2 \ldots s$

S/P	0.1	0,2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	1	1	1	1	1	1	1	1	1
2	0.97,368	0.04,444	0.91,176	0.87.500	0.83,333	0.78,571	0.73.071	0.66,667	0.59,091
3	0.94,760	0.89,071	0.82,880	0.76,190	0.69,048	0.61,538	0.53,837	0.51,075	0.39,189
4	0.92,216	083,898	0.75,158	0.66,176	0.57,222	0.48,645	0.40,843	0 34,188	0.28,915
5	0.89,646	0.78,768	0.68,056	0.57,474	0.47,688	0.39,156	0.32,183	0.26,829	0.22,891
6	0.87,220	0.74,023	0.61,583	0.50,026	0.40,132	0.32,233	0.27,314	0.22,030	0.18,956
7	0.84,788	0.69,523	0.55,734	0.43,693	0.34,194	0.27,147	0.22,173	0.18,691	0.16,181
8	0.82,401	0.65,274	0.50,491	0.38,436	0.29,530	0,23,348	0.19,151	0.16,238	0.14.801
9	0.80,005	0.61,238	0.46,811	0.34,021	0.25,847	0.20,447	0.16,856	0.14,359	0.12,038
10	0.77,768	0.57,506	0.41,673	0.30,327	0.22,911	0.18,177	0.15,057	0.12,872	0.11,252
11	0.75,525	0.5X916	0.38,010	0.27,243	0.20,541	0.16,381	0,13,608	0.11,665	0.10,216
12	0.73,334	0.50,591	0.34,780	0 24,655	0.18,601	0.14.891	0.12,416	0.10,666	0.09,355
13	0.71,194	0.47,705	0.31,939	0.29,122	0.16,992	0.13,643	0.11,452	0.09,825	0.08,628
14	0,69,105	0.44,872	0.29,460	0.20,619	0.15,638	0.12,600	0.10,568	0.90,108	0.08,006
15	0.67,436	0.42,415	0.27,245	0.19,036	0.14,486	0.11,212	0.09,837	0.08,488	0.07,467

Table 3 $E(1/j)$ for the truncated geometric distribution $P(x) = \frac{1}{1-q^s} q^{x-1} p$, $x = 1, 2, \ldots, s$

S	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	1	1	1	1	1	1	1	1	1
2	0.76,316	0.77,778	0.79,412	0.81,250	0.83,333	0.85,714	0.88,462	0.91,667	0.95,465
3	0.63,469	0.66,120	0.69,100	0.72,449	0.76,190	0.80,342	0.84,892	0.89.516	0.94,895
4	0.55,314	0.58,988	0.63,130	0.67,739	0.72,778	0.78,161	0.83,751	0,89,369	0.94,832
5	0 49,656	0.54,238	0.59,396	0.65,056	0.71,075	0.77,258	0.83,389	0.89,281	0.94,825
6	0.45,499	0.50,900	0.56,955	0.63,477	0.70,212	0.76,885	0.83,275	0.89,263	0,94,825
7	0.42,319	0.48,472	0.55,313	0.62,549	0.69,771	0.76,731	0.83,240	0.89,259	0.94,825
8	0.39,815	0,46,659	0.54,185	0.61,962	0.69,551	0.76,667	0.83,229	0.89,258	0.94,825
9	0.37,798	0.45,281	0.53,414	0.61,617	0.69,432	0.76,641	0.83,225	0.89,258	0.94,825
10	0.36,144	0.44,220	0.52,873	0.61,408	0.69,374	0.76,631	0.83,225	0.89,258	0,94,825
11	0.34,769	0.43,395	0.52,495	0.61,281	0.69,345	0.76,627	0.83,225	0.89,258	0.94,825
12	0 33,849	0.42,748	0.52,229	0.61,204	0.69,330	0.76,625	0 83,225	0.89,258	0.94,825
13	0.32,632	0.42,233	0.52,042	0.61,157	0.69,322	0.76,624	0.83,225	0.89,258	0.94,825
14	0.31,792	0.41,835	0.51,911	0.61,129	0.69,319	0.76,624	0 83.225	0.89,258	0.94,825
15	0.31,068	041,514	0.51,818	0.61,112	0.69,317	0 76,624	0.83,224	0.89,258	0.94,825

Table 4 E (1/j) for the truncated Poisson distribution P (j = x)

$$= \frac{m^x}{x!\, f_s(m)} \quad \text{and} \quad f_s(m) = \sum_{x=1}^{s} \frac{m^x}{x!}, \quad x = 1, 2, \ldots\ldots S$$

	S \ m	0.5	1	2	3	4	5
f_1 (m)	1	1	1	1	1	1	1
E (1/j)		1	1	1	1	1	1
f_2 (m)	2	0.6,250	1,5,000	4.0,000	7.5,000	12.0,000	17.5,000
E (1/j)		0.8,101	0.8,333	0,7,500	0.7,000	0.6,667	0.64,028
f_3 (m)	3	0.6,453	1.6,667	5.3,333	12.0,000	22.6,667	38.3,333
E (1/j)		0,7,947	0.7,833	0.6,458	0.5,625	0.5,098	0.4,746
f_4 (m)	4	0.6,484	1.7,084	6.0,000	15.3,750	33.3,334	64.3,750
E (1/j)		0.7,926	0.7,703	0.6.018	0.4,939	0,4,267	0.3,833
f_5 (m)	5	0.6,487	1.7,167	6.2,667	17.4,000	41.8,667	90.4,167
E (1/j)		0.7,923	0,7,676	0.5,847	0.4,597	0.3.805	0.3,308
f_6 (m)	6	0.6,487	1.7,181	6.3.556	18.4,121	47.5,556	112.1,181
E (1/j)		0.7,923	0.7,671	0,5,789	0.4,436	0.3.549	0.2,990
f_7 (m)	7	0.6,487	1.7,183	6.3,810	18.8,460	50.8,064	127.6,191
E (1/j)		0.7,923	0.7,670	0.5,771	0.4,367	0.3,413	0.2,801
f_8 (m)	8	0.6.487	1.7,183	6.3.873	19.0,087	52.4,318	137.3,072
E (1/j)		0.7,923	0.7,670	0.5,767	0.4,340	0.3,346	0.2,691
f_9 (m)	9	0.6,487	1.7,183	6.3,887	19.0,629	53.1,542	142.6,895
E (1/j)		0.7,923	0.7,670	0.5,766	0.4,331	0.3.316	0.2,632
f_{10} (m)	10	0.6,487	1 7,183	6.3,890	19.0,645	53.4,432	145.3,806
E (1/j)		0.7,923	0.7,670	0.5,765	0.4,331	0.3,303	0.2,602

Sometimes the number of possible obsetvations per unit may be limited. Without assumptions of any particular model for y_{ik}, here we have an crdinary

two-stage sampling procedure. If $M_i = M$ for all i, $\bar{\bar{y}}$ is unbiased for $\mu =$

$$\frac{1}{NM} \sum_{i=1}^{N} \sum_{k=1}^{M} y_{ik} \text{ and}$$

$$\text{Est} \quad \text{Var}(\bar{\bar{y}}) = \frac{N-n}{Nn} R + \frac{1}{N} \left\{ E(1/j) - \frac{1}{M} \right\} \frac{Q}{J}$$

If Mi 's are different $\bar{\bar{y}}$ is no longer an unbiased estimate of the population mean. However, it may be made unbiased by selecting the sampling units with probability proportional to their sizes Mj . We may also adopt sampling with replacement in the second stage of selection and thereby the number of possible observations may be treated as infinite.

Summary

When there are several determinations per unit of a sample and j, the number of determinations is a random variable, the mean of the means per unit can be taken to be an unbiased estimate of the population mean. The variance of the estimate and the estimate of this variance is derived under general conditions. Since the expressions involve E (1/j) in particular situations where E (1/j) may be calculated and tabled, they may be evaluated with the help of the tables given. These results may also be extended to other situations suitably.

സംഗ്രഹം

ഒരു സാമ്പിളിലെ ഓരോ യൂണിററിന്മേലുമുള്ള നിരീക്ഷണങ്ങളുടെ സംഖ്യ ഒരു യാദൃശ്ചികചരം ആയിരുന്നാൽ യൂണിററ് ffiacu^fflmgss സമാന്തര മാധ്യം സമഷ്ടി മാധ്യത്തിൻെറ അഭിനതി ഇല്ലാത്ത ആകലം ആയി എടുക്കാം. ആകലത്തിൻെറ മാനകവിചലനവും മററും പൊതുവായിട്ടുള്ള പരിതസ്ഥിതികളിൽ കണ്ടുപിടിക്കാവുന്നതാണ്. ചില പ്രത്യേക പരിതസ്ഥിതികളിൽ അവയുടെ മൂല്യം അവശ്യംവേണ്ടി ഗണിച്ച പട്ടികകളിൽ നിന്ന് കണക്കാക്കാവുന്നതാണ്. കുറച്ചുകൂടി വിപുലമായ ചുററുപാടുകളിലും ഈ ആകലം ഉപയോഗിക്കാവുന്നതാണ്.

Acknowledgement

The authors are greatful to Dr. P. G. Nair, Dean, College of Veterinary and Animal Sciences, Mannuthy, Trichur for his kind interest in this work, The authors greatefully acknowledge the help rendered by Smty. T. K. Indira Bai, Technical Assistant, Departmept of Statistics in the preparation of the tables.

REFERENCE

Singh M and Gupta, V. K. 1978. 'On the Estimation of mean parameters with random number of obbservation on each unit', *I, A. S. R. I. Tech, Rep. No,* 7/78.

(M. S. Received: 24-11-1978)