المؤتمر العلمي الدولي الرابع عشر
لجمعية الرياضيات العراقية والمنعقد تحت شعار
الإبداع يلتقي بالتحديات من اجل التقدم العلمي والتكنولوجي
للمدة 4 - 5 اب 2024
دمشق - سورية

On Mathematical Expectation of the Sample Variance in Simple Sampling Technique

Yasser Al Zaim
Department of Statistics, Faculty of Science, Damascus University, Syria
Email: y.alzaim@damascusuniversity.edu.sy

Abstract

This paper is a brief study focusing on the behavior of the mathematical expectation of sample variance in two different situations: sampling without replacement and sampling with replacement. Formally, we show that when sampling is with replacement, there exists a crucial difference between the two situations, namely, distinct samples and indistinct samples.
Keywords. finite population, simple random sampling, sample variance, unbiased estimator

1. Introduction

It is really obvious today that studying the entire population is not an approach to solve challenging problems in statistic or the other related sciences.
The suitable way to deal with the big and complex data that emerges in different studies, is to draw samples from the studied population. Since the drawn samples are random, one should study the behavior of sample quantities.
Many references deal with studying the sample variance s^2 in two cases, drawing with replacement and drawing without replacement. But, in the case of drawing with replacement, investigating the behavior of $\mathrm{E}\left[s^2\right]$ when the chosen samples are distinct or indistinct are omitted from those references, and this is what we will talk about in this paper.

2. The unbiased estimator of population variance

Let $x_1, x_2, \cdots, x_n$ denote a simple sample characteristics that were drawn from a finite population with elements $y_1, y_2, \cdots, y_N$, with $n,, N$. Denote by $|\Omega|$ to the number of samples of size n drawn from N. That is,

$$|\Omega| = \begin{cases} \binom{N}{n}; \textit{drawing without replacement} \\ \binom{N+n-1}{n}; \textit{drawing indistinct samples with replacement} \\ N^n; \textit{drawing distinct samples with replacement} \end{cases}$$

If $\bar{x}$ is the sample mean, then its variance is

$$\mathrm{V}(\bar{x}) = \frac{N-n}{N-1}\frac{\sigma^2}{n},$$

when the sampling is without replacement, and it is

$$\mathrm{V}(\bar{x}) = \frac{\sigma^2}{n},$$

when sampling with replacement.

It is well known that $\bar{x}$ is an unbiased estimator of $\bar{y}$ whatever the sampling is. As a result, we have that

$$\mathrm{E}\left[\sum_{i=1}^{n} x_i\right] = \frac{n}{N}\sum_{i=1}^{N} y_i$$

The variance of y_i's is defined as

$$\sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N},$$

and it is called the complete variance. We take also

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

as the variance of y_i's. Similarly, the variance of x_i's is defined as

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

Theorem 2.1. For a simple random sample and the drawing is without replacement, s^2 is an unbiased estimator of S^2.

Proof. We follow [1], P.26.

$$\mathrm{E}\left[s^2\right] = \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}(x_i \pm \bar{y} - \bar{x})^2\right]$$

$$= \frac{1}{n-1}\mathrm{E}\sum_{i=1}^{n}\left[(x_i - \bar{y})^2 - 2(x_i - \bar{y})(\bar{x} - \bar{y}) + (\bar{x} - \bar{y})^2\right]$$

$$= \frac{1}{n-1}\mathrm{E}\sum_{i=1}^{n}(x_i - \bar{y})^2 - \frac{n}{n-1}\, E\left[\bar{x} - \bar{y}\right]^2$$

$$= \frac{1}{n-1}\mathrm{E}\sum_{i=1}^{n}(x_i - \bar{y})^2 - \frac{n}{n-1}\mathrm{V}(\bar{x})$$

The last statement using the previous results becomes

$$\mathrm{E}\left[s^2\right] = \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\frac{N-n}{N-1}\frac{\sigma^2}{n}$$

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

Theorem 2.2. For a simple random sample and the drawing is with replacement, s^2 is an unbiased estimator of σ^2.

Proof. The proof begins by repeating the same steps as in Theorem 2.1, but with some changes we find that

$$\mathrm{E}\left[s^2\right] = \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\frac{\sigma^2}{n} = \sigma^2 \square$$

It is obvious in the Theorem 2.2 that knowing whether if the drawn samples are distinct or indistinct, does not play any role in the proof. This idea plays the center role in the following section.

3. Simulation Study

In this section, we present different scenarios to show that when sampling is with replacement and the samples are distinct, $\mathrm{E}\left[s^2\right] = \sigma^2$. The *R language* codes are available by contacting the author.

Assume that a rural population is distributed in five villages as shown in Table 3.1.

Table 3.1 *Residents distribution in five villages.*

village	a	b	c	d	e
residents	700	500	550	625	600

First Scenario

Suppose that the object in this scenario is to draw, without replacement, simple samples of size n=3. Obviously, $|\Omega| = 10$ .Table 3.2 contains the possible simple samples ω_k along with their means $\bar{x}_k$ and variances s_k^2, where $k = 1, 2, \ldots, 10$.

Table 3.2 *Possible chosen samples when sampling is without replacement.*

k	1	2	3	4	5	6	7	8	9	10
ω_k	abc	abd	abe	bcd	bce	cde	acd	ace	ade	bde
$\bar{x}_k$	583.33	608.33	600	625	616.67	641.67	558.33	550	575	591.67
s_k^2	10833.33	10208.33	10000	5625	5833.33	2708.33	3958.33	2500	4375	1458.33

The results in Table 3.2 show that

$$\mathrm{E}\left[s^2\right]=\sum_{k=1}^{|\Omega|}\frac{s_k^2}{|\Omega|}=\frac{s_1^2+s_2^2+\ldots+s_{10}^2}{10}=5750$$

$$S^2=\sum_{i=1}^{N}\frac{\left(y_i-\bar{y}\right)^2}{N-1}=\frac{(700-595)^2+\ldots+(600-595)^2}{4}=5750$$

and this is consistent with Theorem 2.1.

3.2 Second Scenario

Now, assume that indistinct simple samples of size n=3 are to be drawn with replacement. As a result, $|\Omega|=35$. The results listed in Table 3.3 show that

$$\mathrm{E}\left[s^2\right]=\sum_{k=1}^{|\Omega|}\frac{s_k^2}{|\Omega|}=\frac{s_1^2+\ldots+s_{35}^2}{35}=3833.33$$

$$\sigma^2=\sum_{i=1}^{N}\frac{\left(y_i-\bar{y}\right)^2}{N}=\frac{N-1}{N}S^2=4600$$

which seems to be a contradiction with Theorem 2.2.

Table 3.3 *Possible chosen indistinct samples and sampling is with replacement. The $\bar{x}_k$ and s_k^2 are recorded in parentheses.*

k	$\omega_k(\bar{x}_k, s_k^2)$	k	$\omega_k(\bar{x}_k, s_k^2)$	k	$\omega_k(\bar{x}_k, s_k^2)$	k	$\omega_k(\bar{x}_k, s_k^2)$
1	$bbb(500, 0)$	11	$bed(575, 4375)$	21	$ced(591.67, 1458.33)$	31	$eaa(666.67, 3333.33)$
2	$bbc(516.67, 833.33)$	12	$bea(600, 10000)$	22	$cea(616.67, 5833.33)$	32	$ddd(625, 0)$
3	$bbe(533.33, 3333.33)$	13	$bdd(583.33, 5208.33)$	23	$cdd(600, 1875)$	33	$dda(650, 1875)$
4	$bbd(541.67, 5208.33)$	14	$bda(608.33, 10208.33)$	24	$cda(625, 5625)$	34	$daa(675, 1875)$
5	$bba(566.67, 13333.33)$	15	$baa(633.33, 13333.33)$	25	$caa(650, 7500)$	35	$aaa(700, 0)$
6	$bcc(533.33, 833.33)$	16	$ccc(550, 0)$	26	$eee(600, 0)$		
7	$bce(550, 82500)$	17	$cce(566.67, 833.33)$	27	$eed(608.33, 208.33)$		
8	$bcd(558.33, 3958.33)$	18	$ccd(575, 1875)$	28	$eea(633.33, 3333.33)$		
9	$bca(583.33, 10833.33)$	19	$cca(600, 7500)$	29	$edd(616.67, 208.33)$		
10	$bee(566.67, 3333.33)$	20	$cee(583.33, 833.33)$	30	$eda(641.67, 2708.33)$		

3.3 Third Scenario

In this scenario, we consider drawing with replacement all distinct simple samples of size n=3, hence, $|\Omega|=125$. Table 3.4 reflects all possible simple samples, and as a result we have that

$$\mathrm{E}\left[s^2\right]=\frac{s_1^2+s_2^2+\ldots+s_{125}^2}{125}=\frac{575000}{125}=4600$$

which coincides with Theorem 2.2.

Table 3.4 *Possible chosen distinct samples with replacement.*

k	ω_k	k	ω_k	k	ω_k	k	ω_k	k	ω_k	k	ω_k	k	ω_k
1	aaa	21	aea	41	bda	61	cca	81	dba	101	eaa	121	eea
2	aab	22	aeb	42	bdb	62	ccb	82	dbb	102	eab	122	eeb
3	aac	23	aec	43	bdc	63	ccc	83	"dbc	103	eac	123	eec
4	aad	24	aed	44	bdd	64	ccd	84	dbd	104	ead	124	eed
5	aae	25	aee	45	bde	65	cce	85	dbe	105	eae	125	eee
6	aba	26	baa	46	bea	66	cda	86	dca	106	eba		
7	abb	27	bab	47	beb	67	cdb	87	dcb	107	ebb		
8	abc	28	bac	48	bec	68	cdc	88	dcc	108	ebc		
9	abd	29	bad	49	bed	69	cdd	89	dcd	109	ebd		
10	abe	30	bae	50	bee	70	cde	90	dce	110	ebe		
11	aca	31	bba	51	caa	71	cea	91	dda	111	eca		
12	acb	32	bbb	52	cab	72	ceb	92	ddb	112	ecb		
13	acc	33	bbc	53	cac	73	cec	93	ddc	113	ecc		
14	acd	34	bbd	54	cad	74	ced	94	ddd	114	ecd		
15	ace	35	bbe	55	cae	75	cee	95	dde	115	ece		
16	ada	36	bca	56	cba	76	daa	96	dea	116	eda		
17	adb	37	bcb	57	cbb	77	dab	97	deb	117	edb		
18	adc	38	bcc	58	cbc	78	dac	98	dec	118	edc		
19	add	39	bcd	59	cbd	79	dad	99	ded	119	edd		
20	ade	40	bce	60	cbe	80	dae	100	dee	120	ede		

1. Conclusion

This paper has clarified the effect of drawing simple samples methodology on the mathematical expectation of the sample variance. We have seen, through simulation study, that the sample variance is an unbiased estimator of the population variance only when sampling without replacement and with replacement for distinct samples.

2. References

[1] W. Cochran, *Sampling Techniques, third ed.* John Wiley and Sons, 1977.