Verification of Several Important Theorems in Simple Random Sampling Using R Software

Chaobing He

School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China

Email: chaobing5@163.com

Abstract: This paper considers the verification of several important theorems in simple random sampling using R software. First several important theorems in simple random sampling are introduced systematically. Then computer program for the verification of these theorems is written using R. According to these R codes, the paper verifies these theorems. The output proves that the R codes are very practical and effective.

Keywords: simple random sampling; sample mean; sample variance; sample covariance; mathematical expectation

Mathematics Subject Classification: 62D05

1. Introduction

Sampling survey is a kind of non-comprehensive survey. Sampling survey is a statistical analysis method that extracts some actual data from the population according to the principle of randomness and uses probability estimation method to calculate the corresponding quantity index of the population according to the sample data. Sampling survey can be divided into probability sampling and non-probability sampling. Probability sampling is based on the principles of probability theory and mathematical statistics to select samples from the general population of investigation and research according to the principle of randomness, and estimate and infer some characteristics of the general population from the quantity, and the possible errors can be controlled from the sense of probability. Probabilistic sampling is conventionally called a sampling survey.

Simple random sampling is a sampling method that randomly selects n units from the total N units as samples so that the probability of each possible sample being selected is equal. Simple random sampling is one of the most basic sampling methods and is the basis of other sampling methods. The outstanding feature of this method is simple and intuitive. When the sampling box is complete, samples can be directly extracted from it. Since the probability of selection is the same, it is convenient to estimate the target quantity with sample statistics and calculate the sampling error. Therefore, the study of simple random sampling is particularly important. There are a large number of literature on simple random sampling including [1–15] among others. In recent years, computers have played a very important role in the verification of mathematical theorems. The theorems in simple random sampling theory have been rigorously proved mathematically. In this paper, we write a computer program using R language to verify several important theorems in simple random sampling.

The rest of this paper is organized as follows. In Section 2, we present a description of several important theorems in simple random sampling. In Section 3, computer program for the verification of these theorems is written using R. Finally, we summarize and conclude the paper in Section 4.

2. Several important theorems in simple random sampling

Suppose there are N basic units in the population, and $Y_1, Y_2, \cdots, Y_n$ are the values of each basic unit. The variable values of n units of the sample are $y_1, y_2, \cdots, y_n$ respectively.

Let Y_i be the main eigenvalue of unit i, X_i be the other eigenvalue of unit i, and x_i be the sample value corresponding to y_i.

The symbols $\overline{Y}$, $\overline{y}$, S^2, s^2, S_{yx}, s_{yx} and f denote population mean, sample mean, population variance, sample variance, population covariance, sample covariance and sampling fraction respectively. The symbols are as follows.

$$\bar{Y} = \sum_{i=1}^{N} Y_i, \quad \bar{y} = \sum_{i=1}^{n} y_i, \quad f = n/N,$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X}), \quad s_{yx} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}).$$

The symbols $E(\cdot)$, $Var(\cdot)$ and $Cov(\cdot,\cdot)$ denote computing the mathematical expectation, variance and covariance respectively.

Deftnition 1 *From the population of N units, n units are extracted in a batch at a time, so that any unit is equally likely to be selected, any combination of n different units is equally likely to be selected, and this sampling is called simple random sampling.*

Theorem 1 *For simple random sampling, $E(\bar{y}) = \bar{Y}$.*

Theorem 2 *For simple random sampling, $Var(\bar{y}) = \frac{1-f}{n} S^2$.*

Theorem 3 *For simple random sampling, $Cov(\bar{y}, \bar{x}) = \frac{1-f}{n} S_{xy}$.*

Theorem 4 *For simple random sampling, $E(s^2) = S^2$.*

Theorem 5 *For simple random sampling, $E(s_{xy}) = S_{xy}$.*

3. R codes for verification of theorems

In the verification process, we let $N = 5$, $n = 2$, Y_is be 0, 1, 3, 5, 6, and X_is be 5, 3, 8, 4, 9.

The R codes for verification of Theorem 1 are as follows.

```
> N=5
> n=2
> f=n/N
```

```
> Y=c(0,1,3,5,6)
> X=c(5,3,8,4,9)
> m=combn(N,n)
> m
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    1    1    1    1    2    2    2    3    3     4
[2,]    2    3    4    5    3    4    5    4    5     5
> Ybar=mean(Y)
> Sq=var(Y)
> Y[m]
 [1] 0 1 0 3 0 5 0 6 1 3 1 5 1 6 3 5 3 6 5 6
> Ycom=matrix(Y[m],nrow=n)
> Ycom
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    0    0    0    0    1    1    1    3    3     5
[2,]    1    3    5    6    3    5    6    5    6     6
> Ybarcom=apply(Ycom,2,mean)
> Ybarcom
 [1] 0.5 1.5 2.5 3.0 2.0 3.0 3.5 4.0 4.5 5.5
> a=mean(Ybarcom)
> data.frame(target=Ybar,verification=a)
  target verification
1      3            3
```

The output of the last line shows that the verification value is equal to the target value. So Theorem 1 is verified successfully.

The R codes for verification of Theorem 2 are as follows.

```
> Vy=(1-f)/n*Sq
> d=mean((Ybarcom-Ybar)^2)
> data.frame(target=Vy,verification=d)
  target verification
```

```
1    1.95         1.95
```

The output of the last line shows that the verification value is equal to the target value. So Theorem 2 is verified successfully.

The R codes for verification of Theorem 3 are as follows.

```
> Xbar=mean(X)
>  Vxy=(1-f)/n*cov(Y,X)
>  Xcom=matrix(X[c(m)],nrow=n)
>  Xbarcom=apply(Xcom,2,mean)
>  e=mean((Xbarcom-Xbar)*(Ybarcom-Ybar))
> data.frame(target=Vxy,verification=e)
  target verification
1    1.05         1.05
```

The output of the last line shows that the verification value is equal to the target value. So Theorem 3 is verified successfully.

The R codes for verification of Theorem 4 are as follows.

```
>  SYsqu=apply(Ycom,2,var)
> b=mean(SYsqu)
> data.frame(target=Sq,verification=b)
  target verification
1     6.5          6.5
```

The output of the last line shows that the verification value is equal to the target value. So Theorem 4 is verified successfully.

The R codes for verification of Theorem 5 are as follows.

```
> fu=function(vec) vec-mean(vec)
>  Sdy=apply(Ycom,2,fu)
>  Sdx=apply(Xcom,2,fu)
> Sm=Sdy*Sdx
>  Sn=apply(Sm,2,sum)/(n-1)
```

```
> g=mean(Sn)
> data.frame(target=cov(Y,X),verification=g)
  target verification
1    3.5          3.5
```

The output of the last line shows that the verification value is equal to the target value. So Theorem 5 is verified successfully.

In summary, these five theorems are verified successfully using R software.

4. Conclusions

In this paper, we consider the computer verification of several important theorems in simple random sampling. The computer program for the verification is written using R. According to these R codes, the paper verifies these theorems. The output proves that the R codes are very practical and effective.

Reference

[1] H. O. Hartley, A. Ross, Unbiased ratio estimators, Nature 174 (4423) (1954) 270–271.

[2] H. A. David, H. Hartley, E. S. Pearson, The distribution of the ratio, in a single normal sample, of range to standard deviation, Biometrika 41 (3/4) (1954) 482–493.

[3] X. Meng, Scalable simple random sampling and stratified sampling, in: International Conference on Machine Learning, PMLR, 2013, pp. 531–539.

[4] C. Kadilar, H. Cingi, Ratio estimators in simple random sampling, Applied mathematics and computation 151 (3) (2004) 893–902.

[5] S. Gupta, J. Shabbir, On improvement in estimating the population mean in simple random sampling, Journal of Applied Statistics 35 (5) (2008) 559–566.

[6] C. Kadilar, H. Cingi, Improvement in estimating the population mean in simple random sampling, Applied Mathematics Letters 19 (1) (2006) 75–79.

[7] L. Goldstein, G. Reinert, Stein's method and the zero bias transformation with application to simple random sampling, The Annals of Applied Probability 7 (4) (1997) 935–952.

[8] S. Singh, Simple random sampling, in: Advanced Sampling Theory with Applications, Springer, 2003, pp. 71–136.

[9] P. West, Simple random sampling of individual items in the absence of a sampling frame that lists the individuals, New Zealand Journal of Forestry Science 46 (1) (2016) 1–7.

[10] C. Kadilar, H. Cingi, Improvement in variance estimation in simple random sampling, Communications in Statistics—Theory and Methods 36 (11) (2007) 2075–2081.

[11] P. Pathak, On simple random sampling with replacement, Sankhyā: The Indian Journal of Statistics, Series A (1962) 287–302.

[12] S. Mitra, P. Pathak, The nature of simple random sampling, The Annals of Statistics (1984) 1536–1542.

[13] S. Gupta, J. Shabbir, Variance estimation in simple random sampling using auxiliary information, Hacettepe Journal of mathematics and Statistics 37 (1) (2008) 57–67.

[14] H. P. Singh, R. S. Solanki, A new procedure for variance estimation in simple random sampling using auxiliary information, Statistical papers 54 (2) (2013) 479–497.

[15] W. G. Cumberland, R. Royall, Does simple random sampling provide adequate balance?, Journal of the Royal Statistical Society: Series B (Methodological) 50 (1) (1988) 118–124.