SMGroup

SM Journal of Biometrics & Biostatistics

Research Article

# Deviation Test: Comparison of Degree of Randomness of the Tables of Random Numbers due to Tippet, Fisher & Yates, Kendall & Smith and Rand Corporation

**Dhritikesh Chakrabarty***

*Department of Statistics, Handique Girls' College, India*

**Article Information**

Received date: Jul 26, 2017
Accepted date: Aug 14, 2017
Published date: Aug 30, 2017

***Corresponding author**

Dhritikesh Chakrabarty, Department of Statistics, Handique Girls' College, Guwahati, Assam, India,
Email(s): dhritikesh.c@rediffmail.com, dhritikeshchakrabarty@gmail.com



**Keywords** Random numbers table; Tippet, Fisher & Yates; Kendall & Smith R and Corporation; testing of randomness; Deviation test

## Abstract

The randomness of each of the four tables of random numbers namely (1) Tippet's Random Numbers Table, (2) Fisher & Yates Random Numbers Table, (3) Kendall and Smith's Random Numbers Table and (4) Random Numbers Table due to Rand Corporation has been examined and a comparison of the merits of them has been studied with respect to the degree of randomness. Deviation test (based on t statistic) has been applied in examining the randomness of each of the four tables. This paper describes the testing of randomness of the four random numbers tables and a comparison of the degree of randomness of them.

## Introduction

Drawing of random sample has been found to be the vital or basic work in almost every branch of experimental sciences. The scientific method of selecting a random sample consists of the use of random numbers table. Several tables of random numbers have already been constructed by the renowned scientists. Some of them (in chronological order) are due to *Tippett* (1927), *Mahalanobis* (1934), *Kendall* & *Smith* (1938, 1939), *Fisher* & *Yates* (1938), *Hald* (1952), *Royo* & *Ferrer* (1954), *RAND Corporation* (1955), *Quenouille* (1959), *Moses* & *Oakford* (1963), *Rao*, *Mitra* & *Matthai* (1966), *Snedecor* and *Cochran* (1967), *Rohlf* & *Sokal* (1969), *Manfred* (1971), *Hill* & *Hill* (1977) etc. Among these tables, the four tables namely [1-15].

(1) Tippet's Random Numbers Table that consists of 10,400 four-digit numbers,

(2) Fisher & Yates Random Numbers Table that comprises 7500 two-digit numbers,

(3) Kendall and Smith's Random Numbers Table [3] &

(4) Random Numbers Table by Rand Corporation [7] are widely used in drawing of simple random sample (with or without replacement) from a population.

Fisher & Yates obtained the random numbers from the $10^{th}$ to $19^{th}$ digits of A.S. Thompson's 20-figure logarithmic tables. In choosing from those digits, an element of randomness was introduced by using playing cards for the selection of half pages of the tables and of a column between $10^{th}$ to $19^{th}$ and finally for allotting these digits to the $50^{th}$ place in a block. In this case, the question arises whether the method applied in selecting the numbers has made the numbers random. This creates the necessity of determining the degree of randomness of the random numbers table constructed by Fisher and. Similarly, there is necessity of examining the degree of randomness of the other tables of random numbers. In the mean time, one study has been made on examining the degree of randomness of the four tables of random numbers due to Tippet, Fisher & Yates, Kendall & Smith and Rand Corporation respectively [16]. However, the study was done by the application of frequency test (based on chi-square statistic). The findings obtained in this study will be more trustable if the same findings are obtained in another study on the same objective with the same data but by the application of another suitable test statistic. Accordingly, an attempt has here been made on the same study by another suitable test statistic. The randomness of each of the four tables of random numbers namely (1) Tippet's Random Numbers Table, (2) Fisher & Yates Random Numbers Table, (3) Kendall and Smith's Random Numbers Table and (4) Random Numbers Table due to Rand Corporation has been examined and a comparison of the merits of them has been studied with respect to the degree of randomness. Deviation test (based on $t$ statistic) has been applied in examining the randomness of each of the four tables. This paper describes the testing of randomness of the four random numbers tables and a comparison of the degree of randomness of them.

OPEN ACCESS

ISSN: 2573-5470

## The Test Statistic Used

A test statistic is a standardized value that is calculated from sample data during a hypothesis-testing process. A test most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known is *t* -test which is based on *t* statistic. The statistic *t* is the ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error. The *t* statistic is a measure of how extreme a statistical estimate is and this statistic is computed by subtracting the hypothesized value from the statistical estimate and then dividing by the estimated standard error. The *t*-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis [17-25].

A *t*-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistics (under certain conditions) follow a Student's *t* distribution.

The t-statistic was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Dublin, Ireland. "Student" was his pen name [26,17,18,27,28,].

Let *S* be a statistic (estimator) of parameter ì with E(*S*) = ì .

Then a statistic *t* defined by

$$t = \frac{S - E(S)}{SE(S)}$$

where *SE*(*S*) is the standard error of the statistics *S* with õ degrees of freedom is known as *t* statistic with õ degrees of freedom which is used to test a null hypothesis which is equivalent to the null hypothesis that

$E(S) = \mu$

and if the null hypothesis is true then the *t* statistic follows a probability distribution whose probability density function is given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where õ is the number of degrees of freedom and Ã is the gamma function [29-31].

This may also be expressed as

$$f(t) = \frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Where B is the Beta function.

The probability density function is symmetric, and its overall shape resembles the bell shape of a normally distributed variable with mean 0 and variance 1, except that it is a bit lower and wider.

As the number of degrees of freedom grows, the *t*-distribution approaches the normal distribution with mean 0 and variance 1.

Thus in order to test a null hypothesis by *t* statistic, one is required to search for a statistic *S* based on available data and then to find its mathematical expectation and standard error so that *t* statistic can be constructed for testing the null hypothesis.

## T Statistic for the Current Study

Let us consider the Fisher & Yates random numbers table.

Let *d* be the variable denoting the deviation (amount of deviation) of observed number of occurrences from the theoretical number of occurrences of a digit.

Let

$d_i = d_i\ (N)$

be the deviation of the observed number of occurrences of the digit *i* from its theoretical number of occurrences among *N* occurrences of the 10 digits in the table ($i$ = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9).

Then among the 10 deviations

$d_0, d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9$

any nine can assume independent values.

If the occurrences of the 10 digits are random then $d_i = 0$

*for i* = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

in the ideal situation.

However due to chance error, $d_i$ may assume non-zero value.

Thus, $d_i$'s chance errors but not assignable error if the occurrences of the 10 digits in the set of the *N* occurrences.

The chance variables are independently & identically distributed. *N* (0, σ) variables.

Thus, testing of randomness of occurrences of the 10 digits in the table is equivalent to testing the hypothesis $H_o$ that E(di)=0 for *i* = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Now, let

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

Then

$$t = \frac{\bar{d} - E(\bar{d})}{S.E(\bar{d})} \sim t_{n-2}$$

Now,

$$E(\bar{d}) = \frac{1}{n}\sum E(d_i) = 0$$ , when $H_o$ is true.

Also,

Variance $(\bar{d}) = \frac{\sigma^2}{n}$ , $\sigma^2$ is unknown.

However unbiased estimate of $\sigma^2$ is given by

$$S^2 = \frac{1}{n-1}\sum(d_i - \bar{d})^2 = \frac{1}{n-1}\left[\sum d_i^2 - \frac{\left(\sum d_i\right)^2}{n}\right]$$

Which implies that an unbiased estimator of Variance of $\bar{d} = s^2/\eta$

and automatically an unbiased estimator of Standard Deviation of $\bar{d}$ is $s/\sqrt{\eta}$ .

Therefore, statistic *t* for testing $H_o$ becomes

$t = \frac{\bar{d}}{s/\sqrt{n}}$ Which follows *t* distribution with *(n-2)* degrees of freedom. When $H_o$ is true

Thus, the null hypothesis

$H_0: E(d_i) = 0$

is rejected at the significance level *α* if the calculated value of t is found to be exceeding its corresponding theoretical value that corresponds to the level of significance *α*.

However, if the absolute value of deviation $d_i$ is considered then the null hypothesis to be tested in this case will be

$H_0: E(d_i) = 0$

against the alternative hypothesis

$H_1: E(d_i) > 0$

In this case, the null hypothesis

$H_0: E(d_i) = 0$

is rejected at the significance level $\frac{\alpha}{2}$ if the calculated value of *t* is found to be exceeding its corresponding theoretical value that corresponds to the level of significance $\frac{\alpha}{2}$

This statistic can be applied to test the randomness of the whole table or of any part of the table.

This statistic can similarly be applied in testing of the randomness of the other three tables namely Tippet's Random Numbers Table, Kendall & Smith's Random Numbers Table & Random Numbers Table due to Rand Corporation.

## Findings Obtained

Observed values of *t* statistic obtained from the tables of random numbers due to Tippet, Fisher & Yates, Kendall & Smith and Rand Corporation have been shown in table 1.1, table 1.2, table 1.3 and table 1.4 respectively as presented below.

From table 1.1 it is found, on comparing the observed values with the corresponding theoretical values of *t*, that the lack of randomness of Tippet's Random Numbers Table can be treated to be highly significant (i.e. significant at both 5% level of significance & 1% level of significance) except the four parts corresponding to the four sets of trials specifically 1st 2000, 10th 2000, 17th 2000 and last 1600 trials. However, the lack of randomness in these four parts of the table is not insignificant but simply significant (i.e. significant at 5% level of significance).

From table 1.2 it is found, on comparing the observed values with the corresponding theoretical values of *t*, that the lack of randomness of Fisher & Yates shows Random Numbers Table can be treated to be highly significant (i.e. significant at both 5% level of significance & 1% level of significance) except the two parts corresponding to the two sets of 11th & 13th 1000 trials. However, the lack of randomness in these two parts of the table is not insignificant but simply significant (i.e. significant at 5% level of significance).

From table 1.3 it is found, on comparing the observed values with the corresponding theoretical values of *t*, that the lack of randomness of Kendall & Smith's Random Numbers able can be treated to be highly significant (i.e. significant at both 5% level of significance & 1% level of significance) except the part corresponding to the set of 5th 2000 trials. However, the lack of randomness in this part of the table is not insignificant but simply significant (i.e. significant at 5% level of significance).

From table 1.4 it is found, on comparing the observed values with the corresponding theoretical values of *t*, that the lack of randomness of Rand Corporation Random Numbers Table can be treated to be

**Table 1.1:** Observed value of *t* statistic obtained from Tippet's Random Numbers Table.

| Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic |
|---|---|---|---|---|---|
| 1st 2000 | 2.65 | 8th 2000 | 4.14 | 15th 2000 | 7.48 |
| 2nd 2000 | 3.56 | 9th 2000 | 5.37 | 16th 2000 | 4.19 |
| 3rd 2000 | 4.15 | 10th 2000 | 3.31 | 17th 2000 | 2.8 |
| 4th 2000 | 3.57 | 11th 2000 | 4.42 | 18th 2000 | 4.04 |
| 5th 2000 | 3.99 | 12th 2000 | 5.43 | 19th 2000 | 5.44 |
| 6th 2000 | 4.14 | 13th 2000 | 3.64 | 20th 2000 | 3.93 |
| 7th 2000 | 3.99 | 14th 2000 | 7.48 | Last 1600 | 2.83 |

**Table 1.2:** Observed value of *t* Statistic obtained from Fisher & Yates Random Numbers Table.

| Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic |
|---|---|---|---|---|---|
| 1st 1000 | 6.17 | 6th 1000 | 5.06 | 11th 1000 | 3.22 |
| 2nd 1000 | 5.56 | 7th 1000 | 4.44 | 12th 1000 | 4.39 |
| 3rd 1000 | 4.54 | 8th 1000 | 3.72 | 13th 1000 | 2.37 |
| 4th 1000 | 3.99 | 9th 1000 | 3.81 | 14th 1000 | 3.49 |
| 5th 1000 | 3.83 | 10th 1000 | 4.93 | 15th 1000 | 5.44 |

**Table 1.3:** Observed value of *t* statistic obtained from Kendall & Smith's Random Numbers Table.

| Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic |
|---|---|---|---|---|---|
| 1st 2000 | 3.50 | 9th 2000 | 3.54 | 17th 2000 | 4.82 |
| 2nd 2000 | 4.38 | 10th 2000 | 4.30 | 18th 2000 | 3.62 |
| 3rd 2000 | 3.59 | 11th 2000 | 5.12 | 19th 2000 | 3.78 |
| 4th 2000 | 3.64 | 12th 2000 | 4.30 | 20th 2000 | 3.83 |
| 5th 2000 | 3.12 | 13th 2000 | 3.77 | 21st 2000 | **5.76** |
| 6th 2000 | 4.85 | 14th 2000 | 4.48 | 22nd 2000 | 3.76 |
| 7th 2000 | 4.06 | 15th 2000 | 4.77 | 23rd 2000 | 3.83 |
| 8th 2000 | 3.57 | 16th 2000 | 4.71 | Last 1700 | 6.91 |

**Table 1.4:** Observed value of *t* statistic obtained from Rand Corporation Random Numbers Table.

| Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic | Trials | Observed value of *t* statistic |
|---|---|---|---|---|---|
| 1st 2000 | 4.04 | 10th 2000 | 5.53 | 19th 2000 | 3.91 |
| 2nd 2000 | 4.21 | 11th 2000 | 4.11 | 20th 2000 | 3.12 |
| 3rd 2000 | 3.21 | 12th 2000 | 3.60 | 21st 2000 | 5.75 |
| 4th 2000 | 4.00 | 13th 2000 | 4.65 | 22nd 2000 | 4.32 |
| 5th 2000 | 5.00 | 14th 2000 | 4.85 | 23rd 2000 | 2.25 |
| 6th 2000 | 6.91 | 15th 2000 | 3.91 | 24th 2000 | 3.96 |
| 7th 2000 | 3.28 | 16th 2000 | 4.80 | 25th 2000 | 3.14 |
| 8th 2000 | 7.27 | 17th 2000 | 4.63 | | |
| 9th 2000 | 3.59 | 18th 2000 | 4.56 | | |

**Table 2.1:** Ranks of the four tables of random numbers as per deviation test obtained in the current study.

| Name of the Random Numbers Table | Maximum value of *t* statistic | Rank with respect to the Lack of Randomness (as per deviation test) | Rank with respect to the Degree of Randomness (as per deviation test) |
|---|---|---|---|
| Due to Tippet | 7.48 | 1 | 4 |
| Due to Fisher & Yates | 6.17 | 3 | 2 |
| Due to Kendall & Smith | 5.76 | 4 | 1 |
| Due to Rand Corporation | 6.91 | 2 | 3 |

**Table 2.2:** Ranks of the four tables of random numbers as per frequency test obtained in the study done by *Chakrabarty & Sarmah* (2017).

| Name of the Random Numbers Table | Rank with respect to the Lack of Randomness (as per deviation test) | Rank with respect to the Degree of Randomness (as per deviation test) |
|---|---|---|
| Due to Tippet | 2 | 3 |
| Due to Fisher & Yates | 1 | 4 |
| Due to Kendall & Smith | 3 | 2 |
| Due to Rand Corporation | 4 | 1 |

highly significant (i.e. significant at both 5% level of significance & 1% level of significance) except the five parts corresponding to the five sets specifically 3rd, 7th, 20th, 23rd & 25th sets of 2000 trials. However, the lack of randomness in these five parts of the table is not insignificant but simply significant (i.e. significant at 5% level of significance).

## Conclusion

The findings, obtained in this study, imply that the degree of the lack of randomness is highest (in other words, the degree of randomness is lowest) in the Fisher & Yates Random Numbers Table among the four tables of random numbers examined. The four tables can be ranked with respect to the degree of randomness as follows:

*Chakrabarty & Sarmah* (2017) [16] have already made the same study. However, they have applied frequency test (based on chi-square statistic) instead of deviation test (based on *t* statistic) applied here. The findings obtained in that study have been shown in table 2.1.

It is observed that the findings obtained in the two studies are not same. This leads to the necessity of searching for the reason(s) of the difference between the findings in the two studies.

Moreover, one problem for researcher at this stage is to make attempt of constructing of random numbers table with more degree of randomness than that of the existing ones.

## References


1. Tippett LHC. "Random Sampling Numbers". Tracts for Computers no. 15, Cambridge University Press, Cambridge, England. 1927.
2. Mahalanobis PC. "Tables of random samples from a normal population". Sankya. 1934; 1: 289-328.
3. Kendall MG, Smith BB. "Randomness and Random Sampling Numbers". Jour Roy Stat Soc. 1938; 101: 147- 166.

4. Kendall MG, Smith BB. "A Table of Random Sampling Numbers". Tracts for Computers no. 24, Cambridge University Press, Cambridge, England. 1939; 101: 147-166.

5. Fisher RA, Yates F. "Statistical Tables for Biological, Agricultural and Medical Research".6th Edition (1982), Longman Group Limited, England. 1938; 37-38 & 134-139.

6. Hald A. "Table of random numbers". In: A Hald: Statistical Tables and Formulas. Wiley. 1952

7. Royo J, Ferrer S. "Tables of Random Numbers Obtained from Numbers in the Spanish National Lottery". Trabajos de Estadistica. 1954; 5: 247-256.

8. Rand Corporation. "A Million Random Digits". Free Press, Glenoe, III. 1955.

9. Quenouille MH. "Tables of Random Observations from Standard Distributions". Biometrika. 1959; 46: 178-202.

10. Moses EL, Oakford VR. "Tables of Random Permutations". George Allen, Unwin. 1963.

11. Rao CR, Mitra SK, Matthai A. "Random Numbers and Permutations". Statistical Publishing Society, Calcutta. 1966.

12. Snedecor GW, Cochran WG. "Statistical Methods". Iowa State University Press Ames Iowa 6th Edition. 1967.

13. Rohlf FJ, Sokal RR. "Ten Thousand Random Digits". In: Rohlf & Sokal: Statistical Tables. Freeman. 1969.

14. Manfred Mohr. "Le Petit Livre de Nombres au Hasard". Édition d'artiste, Paris. 1971.

15. Hill ID, Hill PA. "Tables of Random Times". U.K. 1977.

16. Chakrabarty Dhritikesh, Sarmah Brajendra Kanta. "Comparison of Degree of Randomness of the Tables of Random Numbers due to Tippet, Fisher & Yates, Kendall & Smith and Rand Corporation". Journal of Reliability and Statistical Studies. 2017; 10: 27-42.

17. Pearson ES, Wishart John. "Student's' Collected Papers. Biometrika Office, London. 1942.

18. Fisher Box Joan. "Guinness, Gosset, Fisher, and Small Samples". Statistical Science. 1987; 2: 45-52.

19. Martin Bland. "An Introduction to Medical Statistics". Oxford University Press. 1995.

20. David HA, Gunnink Jason L. "The Paired t Test Under Artificial Pairing". The American Statistician. 1997; 51: 9-12.

21. Zimmerman Donald W. "A Note on Interpretation of the Paired-Samples t Test". Journal of Educational and Behavioral Statistics. 1997; 22: 349-360.

22. Sawilowsky Shlomo S. "Misconceptions Leading to Choosing the t Test Over The Wilcoxon Mann–Whitney Test for Shift in Location Parameter". Journal of Modern Applied Statistical Methods. 2005; 4: 598-600.

23. Rice John A. "Mathematical Statistics and Data Analysis". Third Edition, Duxbury Advanced. 2006.

24. Zabell SL. "On Student's 1908 Article "The Probable Error of a Mean". Journal of the American Statistical Association. 2008; 103: 1-7.

25. Derrick B, Toher D, White P. "How to compare the means of two samples that include paired observations and independent observations: A companion to Derrick, Russ, Toher and White (2017)". The Quantitative Methods for Psychology. 2017; 13: 120-126.

26. Hotelling H. "British Statistics and Statisticians Today". Journal of the American Statistical Association. 1930; 25: 186-190.

27. Raju TN. "William Sealy Gosset and William A. Silverman: Two "students" of science". Pediatrics. 2005; 116: 732-5.

28. Dodge Yadolah . "The Concise Encyclopedia of Statistics". Springer Science & Business Media. 2008.

29. Fisher RA. "Application of Student's t-distribution". Metron. 1925; 5: 90-104.

30. Senn S, Richardson W. "The first t-test". Statistics in Medicine. 1994; 13: 785-803.

31. Pfanzagl J, Sheynin O. "A forerunner of the t-distribution (Studies in the history of probability and statistics XLIV)". Biometrika. 1996; 83: 891-898.