Data Article

Exploration of UK Lotto results classified into two periods

Hilary I. Okagbue,⁎ Muminu O. Adamu, Pelumi E. Oguntunde, Abiodun A. Opanuga, Manoj K. Rastogi

Department of Mathematics, Covenant University, Ota, Nigeria
Department of Mathematics, University of Lagos, Akoka, Lagos, Nigeria
National Institute of Pharmaceutical Education and Research, Hajipur 844102, India

Abstract

United Kingdom Lotto results are obtained from urn containing some numbers of which six winning numbers and one bonus are drawn at each draw event. There is always a need from prospective players for analysis that can aid them in increasing their chances of winning. In this paper, historical data of the United Kingdom Lotto results were grouped into two periods (19/11/1994–7/10/2015 and 10/10/2015–10/5/2017). The classification was as a result of increase of the lotto numbers from 49 to 59. Exploratory statistical and mathematical tools were used to obtain new patterns of winning numbers. The data can provide insights on the random nature and distribution of the winning numbers and help prospective players in increasing their chances of winning the lotto.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specification Table

| Subject area | Decision Science |
|--------------|-----------------|
| More specific subject area | Lottery Statistics/Gambling Theory |
| Type of data | Table |

⁎ Corresponding author.
E-mail address: hilary.okagbue@covenantuniversity.edu.ng (H.I. Okagbue).

http://dx.doi.org/10.1016/j.dib.2017.07.037
2352–3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
How data was acquired
The data was retrieved from www.lottery.co.uk [1]

Data format
Processed data from November 19, 1994 to May 10, 2017

Experimental factors
Data refined from the results archived in www.lottery.co.uk, only the single cases were considered

Experimental features
Statistical analysis, digital root analysis.

Data source location
United Kingdom

Data accessibility
All the data are in this data article.

Software
Microsoft Excel and Minitab 17 Statistical Software

Value of the data

- The data analysis provides a different approach of classifying winning numbers of the UK lotto results [1–6].
- The data analysis can be extended to winning pairs and triples.
- The use of digital root provides another avenue for studying probabilities of winning [7,8].
- Discovery of new patterns can encourage more players thereby improving the economic conditions and welfare of the country [9–11].
- The data can be useful for educational purposes and gambling researchers, number theorists, lotto operators, statisticians, journalists and so on.
- The method and analysis can be replicated for other lotto game results.

1. Data

The data for this study has been analysed to a certain extent, archived and updated at each draw in [1]. This data article contains data generated from different approach other than what was contained in [1] and it is publicly available. The data was on gathered on draw by draw basis. The data is divided into two periods; period A: when the lotto numbers are from one to forty nine (19/11/1994–7/10/ 2015) and period B: when the lotto numbers are from one to fifty nine (10/10/2015–10/5/2017). The draws for periods A and B are 2065 and 166 respectively. The data obtained for periods A and B when the winning numbers are classified using certain number criteria are shown in Tables 1–4. The fre-

| Numbers | Period A | Period B |
|---------|----------|----------|
| 1–10    | 2488     | 157      |
| 11–20   | 2423     | 174      |
| 21–30   | 2577     | 165      |
| 31–40   | 2601     | 181      |
| 41–50   | 2301     | 157      |
| 51–59   |          | 162      |

The most single winning numbers from period A corresponds to 31–40 and the least corresponds to 41–50. Understandingly, the last class contains only 9 numbers for period A. Currently, from the analysis, prospective players with numbers 31–40 and 11–20 has more frequency than other classes.
Table 2
The lotto single winning numbers classified in multiples.

| Multiples | Period A | Period B |
|-----------|----------|----------|
| 2         | 6036     | 499      |
| 3         | 4093     | 294      |
| 4         | 2993     | 226      |
| 5         | 2293     | 167      |
| 6         | 2053     | 145      |
| 7         | 1761     | 130      |
| 8         | 1511     | 111      |
| 9         | 1277     | 94       |
| 10        | 1026     | 82       |

Remark: The frequency of occurrence decreases with increasing multiples of number for both periods.

Table 3
The lotto single winning numbers classified in odd and even numbers.

| Numbers | Period A | Period B |
|---------|----------|----------|
| Even    | 6036     | 499      |
| Odd     | 6354     | 497      |
| Total   | 12,390   | 996      |

Remark: More single odd winning numbers were drawn in period A. However, almost the same frequency was drawn for both even and odd single winning numbers in period B. Chi-square tests and t-tests may not be useful in confirmation the result since the possible winning numbers are more than the even numbers by one.

Table 4
The lotto single winning numbers classified in prime and non-prime numbers.

| Numbers | Period A | Period B |
|---------|----------|----------|
| Prime   | 3770     | 307      |
| Non-prime | 8620   | 689      |
| Total   | 12,390   | 996      |

Remark: Prime numbers appeared in 27% and 31% of all the single winning numbers in periods A and B respectively.

Table 5
The lotto single winning numbers classified in digital roots 1–9.

| Digital root | Lotto numbers |
|--------------|---------------|
| 1            | 1 10 19 28 37 46 55 |
| 2            | 2 11 20 29 38 47 56 |
| 3            | 3 12 21 30 39 48 57 |
| 4            | 4 13 22 31 40 49 58 |
| 5            | 5 14 23 32 41 50 59 |
| 6            | 6 15 24 33 42 51 60 |
| 7            | 7 16 25 34 43 52 61 |
| 8            | 8 17 26 35 44 53 62 |
| 9            | 9 18 27 36 45 54 63 |
frequency distribution of the lotto winning numbers when they are classified according to their digital roots is shown in Table 6 and the various lotto numbers that constitute each digital root are listed in Table 5. This article also introduces the use of the frequencies of digital root in chi-square tests. Finally, simulated data showed the uniformity, randomness and non-normality of occurrence of winning numbers in UK lotto game.

2. Methods and materials

Various aspects of statistical, mathematical and psychological analysis of lottery have been considered [12–18].

2.1. Digital root

This is the sum of digits of a studied number until a single digit number is the final outcome [19–22]. Digital roots often reveal hidden patterns of distributions as seen in [23–25]. This can be applied to lotto to reveal hidden patterns of distribution of winning numbers. The complete list of numbers grouped under their respective digital roots and is shown in Table 5. The digital root of the single winning numbers for periods A and B is shown in Table 6.

2.2. Chi-square test of independence

The Pearson chi-square test is conducted to determine whether the observed values conform to theoretical expectations. The expected frequencies in the chi-square test of independence follow the uniform distribution. Details on chi-square test and other tests can be found in [26–31]. This paper introduces the use of frequency obtained from the digital roots of number instead of all the numbers in chi-square test of independence. This approach was compared with the traditional procedure using the frequency data in totality. The results of the Chi-square tests for periods A and B using Table 6 are shown in Tables 7 and 9 while the decision rule based on different confidence intervals are shown in Tables 8 and 10.

The expected value was obtained from Table 6 by the sum of all the values under the column (Period A) divided by 9.

The statistical hypothesis is stated; null hypothesis imply independence while the alternative imply otherwise.

\[ \chi_{cal} < \chi_{sig} \] Accept the null hypothesis (independence);

\[ \chi_{cal} > \chi_{sig} \] Accept the alternative hypothesis (association);

\[ \chi_{cal} = 2.593494056 \] (From Table 7).

Table 6

| Digital root | Period A | Period B |
|--------------|----------|----------|
| 1            | 1467     | 127      |
| 2            | 1513     | 128      |
| 3            | 1532     | 113      |
| 4            | 1526     | 126      |
| 5            | 1254     | 119      |
| 6            | 1284     | 87       |
| 7            | 1258     | 107      |
| 8            | 1279     | 94       |
| 9            | 1277     | 82       |

Remark: The importance of the digital roots is that it makes use of all the observations unlike what was obtained in Table 2, where the multiples excluded the prime numbers.
The decision rule for the different level of significance of the chi-square test for period A is shown in Table 8.

The expected value was obtained from Table 6 by the sum of all the values under the column (Period B) divided by 9.

The statistical hypothesis is stated; null hypothesis imply independence while the alternative imply otherwise.

$X_{cal} \leq X_{sig}$ Accept the null hypothesis (independence);

$X_{cal} > X_{sig}$ Accept the alternative hypothesis (association);

$X_{cal} = 8.758306$ (From Table 9).

| Number | Observed | Expected | Residual | Statistic |
|--------|----------|----------|----------|-----------|
| 1      | 1467     | 1517.22  | −50.22   | 1.662282596 |
| 2      | 1513     | 1517.22  | −4.22    | 0.01173752   |
| 3      | 1532     | 1517.22  | 14.78    | 0.143979383  |
| 4      | 1526     | 1517.22  | 8.78     | 0.05080898   |
| 5      | 1254     | 1264.22  | −10.22   | 0.082618848  |
| 6      | 1284     | 1264.22  | 19.78    | 0.309478097  |
| 7      | 1258     | 1264.22  | −6.22    | 0.030602585  |
| 8      | 1279     | 1264.22  | 14.78    | 0.172793027  |
| 9      | 1277     | 1264.22  | 12.78    | 0.12919302   |

Table 8
Decision rule for the chi-square test for period A.

| $\alpha$ | 0.995 | 0.99 | 0.975 | 0.95 | 0.90 | 0.10 | 0.05 | 0.025 | 0.010 |
|-----------|-------|------|-------|------|------|------|------|-------|-------|
| Decision  | Reject $H_0$ | Reject $H_0$ | Reject $H_0$ | Accept $H_0$ | Accept $H_0$ | Accept $H_0$ | Accept $H_0$ | Accept $H_0$ | Accept $H_0$ |

Note: $\alpha$ denotes the level of significance

| Number | Observed | Expected | Residual | Statistic |
|--------|----------|----------|----------|-----------|
| 1      | 127      | 118.17   | 8.83     | 0.659802826 |
| 2      | 128      | 118.17   | 9.83     | 0.817710925 |
| 3      | 113      | 118.17   | −5.17    | 0.226190234 |
| 4      | 126      | 118.17   | 7.83     | 0.518819497 |
| 5      | 119      | 118.17   | 0.83     | 0.005829737 |
| 6      | 87       | 101.28   | −14.28   | 2.013412322 |
| 7      | 107      | 101.28   | 5.72     | 0.323048973 |
| 8      | 94       | 101.28   | −7.28    | 0.52328594  |
| 9      | 82       | 101.28   | −19.28   | 3.670205371 |

Table 10
Decision rule for the chi-square test for period B.

| $\alpha$ | 0.995 | 0.99 | 0.975 | 0.95 | 0.90 | 0.10 | 0.05 | 0.025 | 0.010 |
|-----------|-------|------|-------|------|------|------|------|-------|-------|
| Decision  | Reject $H_0$ | Reject $H_0$ | Reject $H_0$ | Reject $H_0$ | Accept $H_0$ | Accept $H_0$ | Accept $H_0$ | Accept $H_0$ | Accept $H_0$ |

Note: $H_0$ denotes the null hypothesis (observations are random)
The decision rule for the different level of significance of the chi-square test for period B is shown in Table 10.

The basis for the statistical decision is that the calculated chi-square statistic is compared with the one tabulated at different degrees of freedom. This revealed that the distribution of the winning numbers of UK lotto is purely random especially at high confidence intervals. This has shown that the UK lotto game is fair.

2.3. Simulation analysis

Monte Carlo simulation was used to generate 20,000 simulated results using the discrete uniform distributions for periods A and B. The results are shown as histograms in Figs. 1 and 2.

The simulation results revealed the uniformity in frequency distributions of the lotto numbers and hence the winning numbers does not appear to cluster around any specific value. However, the extreme values 1, 49 and 59 seem to deviate from uniformity. This is one of the major drawbacks of Monte Carlo simulation used to generate those results.
Acknowledgement

This work is a benefit of sponsorship from Centre for Research, Innovation and Discovery, Covenant University, Ota, Nigeria.

Transparency document. Supplementary material

Transparency document data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.07.037.

References

[1] (www.lottery.co.uk). (Accessed 10 May 2017).
[2] T.V. Wang, R.J.D. Potter van Loon, M.J. van den Assem, D. van Dolder, Number preferences in lotteries, Judgm. Decis. Mak. 11 (3) (2016) 243–259.
[3] J. Haigh, The statistics of the national lottery, J. R. Stat. Soc. Ser. A: Stat. Soc. 160 (2) (1997) 187–206.
[4] N. Henze, A statistical and probabilistic analysis of popular lottery tickets, Stat. Neerl. 51 (2) (1997) 155–163.
[5] P.C. Moore, The development of the UK National Lottery: 1992–96, J. R. Stat. Soc. Ser. A: Stat. Soc. 160 (2) (1997) 169–185.
[6] S.R. Searle, Winning probabilities of lotto in the United States, Chance 11 (1) (1998) 20–24.
[7] J. Simon, An analysis of the distribution of combinations chosen by UK national lottery players, J. Risk Uncertain. 17 (3) (1998) 243–277.
[8] P. Pugh, P. Webley, Adolescent participation in the UK national lottery games, J. Adolesc. 23 (1) (2000) 1–11.
[9] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[10] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[11] L. Farrell, E. Morgenroth, I. Walker, A time series analysis of UK lottery sales: long and short run price elasticities, Oxf. Bull. Econ. Stat. 61 (4) (1999) 513–526.
[12] H. Stern, T.M. Cover, Maximum entropy and the lottery, J. R. Stat. Soc. Ser. A. Stat. Soc. 160 (2) (1997) 169–185.
[13] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[14] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[15] L. Farrell, E. Morgenroth, I. Walker, A time series analysis of UK lottery sales: long and short run price elasticities, Oxf. Bull. Econ. Stat. 61 (4) (1999) 513–526.
[16] H. Stern, T.M. Cover, Maximum entropy and the lottery, J. R. Stat. Soc. Ser. A. Stat. Soc. 160 (2) (1997) 169–185.
[17] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[18] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[19] L. Farrell, E. Morgenroth, I. Walker, A time series analysis of UK lottery sales: long and short run price elasticities, Oxf. Bull. Econ. Stat. 61 (4) (1999) 513–526.
[20] H. Stern, T.M. Cover, Maximum entropy and the lottery, J. R. Stat. Soc. Ser. A. Stat. Soc. 160 (2) (1997) 169–185.
[21] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[22] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[23] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[24] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[25] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[26] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[27] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[28] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[29] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[30] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[31] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[32] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[33] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[34] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[35] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[36] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.
[37] L. Farrell, I. Walker, The welfare effects of lotto: evidence from the UK, J. Public Econ. 72 (1) (1999) 99–120.
[38] D. Forrest, O.D. Guiley, R. Simmons, Elasticity of demand for UK National Lottery tickets, Natl. Tax J. 53 (4) (2000) 853–863.