Introduction

The concept of a National Lottery is by no means a new one. In fact, the first recorded public lottery was sponsored by Augustus Caesar in AD 10 to raise funds for the repair of Rome. By 1993 there were 165 lotteries in operation world-wide, but it was not until 19 November 1994 that the first draw of the UK National Lottery took place. Since that date, ticket sales have averaged around £68 million per week, with prize payouts ranging from £10 to £22.5 million.

With such huge fortunes up for grabs, it is thus imperative that the draw is seen to be fair, and to this end the lottery machines and balls are kept under close scrutiny, and regularly checked by the Office of National Weights and Measures. However, despite these checks, it is still the onus of the statistician to say whether or not the numbers selected pass reasonable tests of randomness.

In addition, when looking at lists showing the number of people winning the various prizes each week, there appears to be a huge spread of values. For instance, in Week 9 a staggering 133 people shared the Jackpot prize, against an expected value of 5. It is therefore also interesting to see what the proportion of people winning prizes each week can tell us about the manner in which they select their numbers.

In summary then, this project is mostly concerned with answering the following questions:

- Are the winning numbers chosen by the lottery machines random? If not, why not?
- Are the numbers chosen by gamblers random? If not, in what way are they non-random?

Gameplay Summary

For completeness, the method of play in the UK National Lottery is as follows:

Gamblers pay £1 for each set of six numbers they wish to select from the range of integers $\{1, 2, \dots, 49\}$. They do this by marking their chosen numbers on a playslip, which is laid out as 9 rows of 5 followed by a row of 4. Alternatively, as of Draw 71, a player can choose to buy a "Lucky Dip" ticket, whereby the computer selects a set of six "random" numbers.

A prize is won if at least three of these numbers match the numbers on six balls drawn "randomly" from a set of 49 similarly numbered balls. A seventh ball, the "Bonus Number" is also selected; the Bonus prize going to combinations matching the Bonus number and five of the six main numbers. No single ticket can win more than one prize.

Prize Structure

The prize pool for each weekly draw is 45% of the total week's takings. This is then rounded down to the nearest 10p. Match 3 prizes are usually a fixed £10 prize. After these have been deducted, the remaining prize pool is divided as follows:

Match 6	(Jackpot)	52%
Match 5+	(Bonus)	16%
Match 5		10%
Match 4		22%

If the Jackpot is not won, then the entire Jackpot pool is "rolled over" to the next draw. This can happen a maximum of three times - if it occurs a fourth time the pool is shared amongst the Bonus prize winners. The Jackpot pool can also be boosted by the announcement of a "Super Draw", where the Jackpot pool is artificially topped up from reserve cash. If ever there are no winners of one of the other four prize categories, the prize pool for that prize is added to that of another prize, though this has never happened to date.

In the unlikely event that the total prize fund is insufficient to give all winners £10, the prize fund will be shared equally amongst all winners. This rule came into effect from Draw 71, and does not affect any Rollover / Super Draw prize money. This does not eliminate the possibility of a situation where, say, the Match 4 prize is of a lower value than the Match 3 prize, as would probably happen if the proportion of Match 3 winners were around 4%.

Prize Winning Probabilities

PRIZE	FREQUENCY	PROBABILITY	ODDS
Match 6 (Jackpot)	$1 = f_J$	7.15×10^{-8}	1 in 13,983,816
Match 5+ (Bonus)	$6 = f_B$ (6C5)	4.29×10^{-7}	1 in 2,330,636
Match 5	$252 = f_5$ (6C5 x 42C1)	1.80×10^{-5}	1 in 55,491
Match 4	$13{,}545 = f_4$ (6C4 x 43C2)	9.69×10^{-4}	1 in 1,032
Match 3	$246{,}820 = f_3$ (6C3 x 43C3)	0.0177	1 in 57
Any Prize	$260{,}624 = f_P$	0.0186	1 in 54

OTHER MATCHING	FREQUENCY	PROBABILITY	ODDS
Match 2	$1{,}851{,}150 = f_2$	0.132	1 in 8
Match 1	$5{,}775{,}588 = f_1$	0.413	1 in 2.4
Match 0	$6{,}096{,}454 = f_0$	0.436	1 in 2.3
No Prize	$13{,}723{,}192 = f_N$	0.981	1 in 1.02

§1. Are The Winning Numbers Random?

An independent body, BSI Test, is commissioned by Camelot (the UK National Lottery Operators) to carry out tests on the numbers drawn, both for the "live" draws and under test conditions. Their results are not disclosed. I have therefore selected a variety of tests to perform on the data generated by the "live" draws.

Unless otherwise specified, all calculations and results given in this project refer to the first 100 lottery draws, i.e. from 19 November 1994 through to 12 October 1996

There are three lottery machines, "Arthur", "Guinevere" and "Lancelot", one of which is randomly chosen to select the six winning numbers each week. Similarly, there are eight sets of balls, numbered 1 through 8. Information on which machine and ball set were used each week can be found in Appendix A.

Test 1 - Equality of Marginal Frequencies

There have been a number of occasions when the frequency of occurrence of various numbers has been commented upon. For example, number 44 appearing 24 times while number 39 only occurred 6 times. But is this all that surprising?

It has been pointed out (Joe, 1993) that we need to modify the standard goodness-of-fit statistic $\sum_k \{O_k - E_k\}^2 / E_k$ when we are sampling m balls at a time without replacement. Suppose we are drawing m balls from M over a series of D draws, and further that ball k is drawn a total of X_k times. Then it has been shown that the statistic required reduces to:

$$M(M-1)\left\{\sum_k X_k^2 - m^2 \times D^2/M\right\} \Big/ \{(M-m)\cdot D\cdot m\} = W \text{ (say)},$$

to be compared to a χ^2_{M-1}.

Evaluating this statistic for the first 100 draws gives:

	W	p-value
Main Balls	41.8493	0.721703
Main Balls and Bonus	46.0800	0.551834

which is obviously far from significant.

Eschewing the information about modifying the test statistic, I performed some more conventional chi-squared tests for association between the machines, and the ball sets:

	d.o.f.	Pearson Statistic	p-value
Machines	96	95.143	0.929628
Ball sets	288	253.460	0.505530

So again, there is nothing here to suggest that the numbers which are chosen are affected by the choice of machine or ball set.

Test 2 - Sum of the Numbers

A second simple test that can be carried out involves looking at the sum of the seven winning numbers.

Let x_i (i=1,2,...,m) be the value of the ith ball drawn, and set $S(d) = \sum_{i=1}^{m} x_i(d)$.

Assuming that the balls are drawn randomly, S(d) has mean and variance $\mu = m(M+1)/2$ and $\sigma^2 = m(M+1)(M-m)/12$ respectively, and we can compare these with the sample mean and variance of S(d) $(1 \le d \le D)$. If the latter are denoted as U and V respectively, then the test statistics are $(U - \mu)/\sigma$ as standard normal, and $(D-1)V/\sigma^2$ as $\chi^2{}_{D-1}$.

Performing these tests on the data, we obtain:

	U	$(U-\mu)/\sigma$	p-value
Arthur	173.89	-0.032	0.9744
Guinevere	183.18	0.234	0.8150
Lancelot	171.15	-0.110	0.9124
Overall	176.68	0.048	0.9616

	V	$(D-1)V/\sigma^2$	d.o.f	p-value
Arthur	1347.4	37.397	34	0.315857
Guinevere	1360.4	41.090	37	0.296027
Lancelot	1440.9	30.582	26	0.244184
Overall	1376.9	111.276	99	0.187892

and once again, we have no reason to believe that the numbers drawn are in any way extraordinary.

Test 3 - "D^2"

This test, and the one that follows it, have null distributions which were less easy to calculate analytically. To get round this problem, I wrote a number of programs in C to simulate the statistic a large number of times and judge the extremity of the observed statistic by comparing it to these simulated values. An example of such a program is given in Appendix B.

The D^2 test has test statistic:

$$X = (x_1 - x_5)^2 + (x_2 - x_6)^2 + (x_3 - x_7)^2, \text{ so that } 3 \le X \le 6356.$$

X was simulated D times, where D is the number of occasions on which a particular lottery machine has been used (Arthur 35, Guinevere 38, Lancelot 27), in order to obtain $\overline{X}$. $\overline{X}$ was then evaluated 1000 times for each machine.

Letting U be the sample mean of the 1000 $\overline{X}$ and V the sample variance, we can compare $Z = \dfrac{\overline{X}_{obs} - U}{\sqrt{V}}$ to a standard normal:

	$\overline{X}_{obs}$	U	V	Z	p-value
Arthur	1102	1223.64	17747.02	-0.913	0.3612
Guinevere	1224	1222.43	16592.96	0.012	0.9904
Lancelot	1216	1222.74	23621.25	-0.044	0.9648

Similarly to before, these p-values give no cause for concern.

Test 4 - "Hypersphere"

The hypersphere test has statistic:

$$X = \sum_i \{X_i - 25\}^2, \text{ so } 28 \le X \le 3619$$

and was implemented in much the say way as Test 3, i.e. by computer simulation.

The following data resulted:

	$\overline{X}_{obs}$	U	V	Z	p-value
Arthur	1425.2	1400.30	5731.16	0.329	0.7412
Guinevere	1225.3	1399.57	5370.36	-2.378	0.0174
Lancelot	1480	1399.77	7663.48	0.916	0.3598

The p-value for Guinevere, 1.74%, is quite small, but not unreasonably so. When conducting as many tests as we are here, it seems likely that some will result in slightly extreme values.

Test 5 - "Evens"

This test, as the name suggests, has statistic:

X = "Number of even valued balls", so $0 \le X \le 7$.

Since we are selecting 7 numbers from 49, of which there are 24 even numbers, the number of combinations with X even numbers is given by $\binom{24}{X}\binom{25}{7-X}$:

X	Combinations with X even numbers
0	480 700
1	4 250 400
2	14 663 880
3	25 603 600
4	24 439 800
5	12 751 200
6	3 364 900
7	346 104
Total (49C7)	85 900 584

The observed values of X are as follows:

	Occurrences			
X	Arthur	Guinevere	Lancelot	Overall
0-2	5	7	3	15
3	15	7	11	33
4	10	16	5	31
5-7	5	8	8	21
Totals	35	38	27	100

which give the following results when compared to a χ^2_3:

	Pearson Statistic	p-value
Arthur	3.50098	0.320635
Guinevere	4.50413	0.211922
Lancelot	5.13493	0.162179
Overall	3.29010	0.349023

Once again, there is no evidence that the numbers being drawn are non-random.

Test 6 - Contiguity

This test examines the maximum number of contiguous values amongst the order statistics of the seven selected numbers:

$$X = \text{"Maximum number of contiguous values", so } X \subset \{0,2,3,4,5,6,7\}.$$

The total number of combinations for each value of X is as follows:

X	Combinations with X maximum contiguous values
0	32 224 114
2	46 698 344
3	6 367 956
4	569 492
5	38 829
6	1 806
7	43
Total (49C7)	85 900 584

The observed values of X are as follows:

	Occurrences			
X	Arthur	Guinevere	Lancelot	Overall
0	10	20	10	40
2	19	16	12	47
3-7	6	2	5	13
Totals	35	38	27	100

which give the following results when compared to a χ^2_2:

	Pearson Statistic	p-value
Arthur	4.25095	0.119376
Guinevere	3.74831	0.153485
Lancelot	4.08173	0.129916
Overall	4.08951	0.129412

Yet again, these p-values fail to suggest that the draws are anything but random.

Test 7 - Minimum Separation

It has been proven (Haigh, 1995) that there are $\binom{n-(r-1)(k-1)}{k}-\binom{n-r(k-1)}{k}$ ways of selecting k distinct integers from {1,2,...,n} such that the minimum separation between successive ordered choices is exactly r.

Considering only the six main balls, and letting X be the minimum separation as defined above, the total number of combinations for each value of X is as follows:

X	Combinations with minimum separation X
1	6 924 764
2	3 796 429
3	1 917 719
4	869 884
5	340 424
6	107 464
7	24 129
8	2 919
9	84
Total (49C6)	13 983 816

The observed values of X are as follows:

	Occurrences			
X	Arthur	Guinevere	Lancelot	Overall
1	21	14	15	50
2	10	17	9	36
3-9	4	7	3	14
Totals	35	38	27	100

which give the following results when compared to a χ^2_2:

	Pearson Statistic	p-value
Arthur	2.9277	0.9041785
Guinevere	5.9559	0.050897
Lancelot	2.3072	0.315499

Unlike previous tests, the "overall" statistic for this test can take a higher degree of freedom, due to the large number of draws the data covers. The 3-9 category for X can be split into a 3 category containing 5 occurrences, and a 4-9 category with 9 occurrences. Using this information and comparing to a χ^2_3 gives a p-value of 0.037286, which is quite small. As with test 4, it appears to be Guinevere that has generated the most anomalous data.

Test 8 - Gaps

So far, all the tests I have conducted have looked at aspects of randomness within one draw. I therefore thought it important to now consider the independence between draws. As was noted earlier, there was much public comment when, for example, ball 44 appeared in the winning selection every week for a month in Summer 1996, whereas ball 39 failed to appear for over a year! This test should hopefully put such observations into context.

To investigate this matter, I noted down the number of draws between successive appearances of each of the 49 balls. The lengths of these 600 gaps were then tabulated, to give the "observed" values below. The "expected" values shown were estimated using a computer program (see Appendix B) to simulate 100 000 sets of 100 lottery draws.

Gap Size	Observed	Expected	Gap Size	Observed	Expected
1	73	78.74059	14	16	12.65385
2	59	68.46286	15	13	11.00096
3	68	59.49242	16	8	9.53213
4	46	51.72983	17	9	8.27456
5	50	44.97890	18	7	7.20359
6	31	39.02480	19	7	6.23539
7	42	33.96111	20	0	5.41470
8	30	29.48650	21/22	7	8.76635
9	28	25.61190	23/24	8	6.60495
10	22	22.26740	25-27	8	6.95702
11	18	19.32632	28-31	5	5.64584
12	20	16.81033	≥32	4	7.23014
13	21	14.58756	Total	600	600

This data has a χ^2_{24} statistic of 20.837, which gives a p-value of 0.648306. Hence there is no evidence to contradict the hypothesis that the draws are independent.

Summary

The eight tests detailed previously were selected to test various aspects of randomness regarding the "live" data generated by the three lottery machines. As such, there is no substantial evidence to suggest that the draws are at all unfair (although the Guinevere data produces some slightly dubious results, and it may be interesting to reperform these tests at a later date.)

In many of the tests, it has been necessary to combine two or more extreme categories, reducing the number of degrees of freedom accordingly in order to perform a chi-squared goodness-of-fit test. In most cases, it will take several hundred more draws before this can be avoided.

§2. Randomness of Gamble Choice ?

Given, then, that the winning numbers chosen by the lottery machines are random, our attention turns now to the selections made by those playing the National Lottery. As was mentioned in the introduction to this project, there appears to be a huge variation in the number of tickets winning prizes. But is this an unreasonable spread of values?

The Poisson Model

Suppose we assume that the N selections made in a particular week are random, and let the total number of different tickets by $F = \binom{49}{6} = 13,983,816$. Then using the notation given in the table on page 4, the number of winners in a particular category will follow a Poisson distribution with mean $\mu = Nf_i/F$. This model should be an excellent approximation, due to the high volume of sales and possible choices.

However, it seems that this model simply does not hold. For example, if we assume that around 60 million tickets are bought each week (an underestimate), then we would expect around 4 tickets to share the Jackpot prize each week, and on only 2% of occasions would we expect there to be no Jackpot winner. Yet in the first 100 draws there have been 17 such occurrences!

If we approximate this Poisson model to a Normal distribution, then we would expect the values of $Z = (X - \mu)/\sqrt{\mu}$ (where X is the number of people winning the relevant prize) to lie in the range (-2,2) on 95% of occasions. But this does not occur. If we consider, for example, the total number of winners across all prize categories, then we discover that for the 100 draws in question Z takes a mean absolute value of 165.3. In the case of Week 9, mentioned in the introduction, Z reaches 704.450! Only on one occasion (Week 60) was Z at all reasonable (1.146).

There is further evidence, were it needed, that this Poisson model fails spectacularly: Consider the correlations between the number of people winning the various prizes. If the combinations being sold were truly random, then these correlations would be approximately zero (since no ticket can win more than one prize). The tables below show that this is not the case here:

	Bonus	Match 5	Match 4	Match 3
Jackpot	0.782	0.868	0.617	0.371
Bonus		0.805	0.692	0.510
Match 5			0.878	0.639
Match 4				0.890

If we eliminate Weeks 9 and 70, both of which had an extraordinarily large number of Jackpot winners, the table becomes:

	Bonus	Match 5	Match 4	Match 3
Jackpot	0.394	0.672	0.494	0.341
Bonus		0.561	0.548	0.467
Match 5			0.884	0.688
Match 4				0.904

Factors Influencing Gamblers' Choice of Combinations

It seems certain, then, that the combinations chosen by lottery players are far from random. So in what ways are they non-random? Camelot, the organisers of the UK National Lottery, refuse to reveal any data regarding which individual numbers, or combinations of numbers are chosen. The only information I could glean was that "research into the matter showed the public tend to use birthdays or special dates."

This commercial secrecy is in stark contrast to similar foreign lotteries, where the marginal distributions of each integer are regularly published. Perhaps oddly, this dissemination of knowledge has not greatly affected the numbers selected in subsequent draws.

Over the past few months, I have taken every opportunity to question people on their choice of lottery numbers, and have found that dates and ages do indeed occur very often. But not all combinations chosen are premeditated in such a fashion - many punters select their numbers at the moment of sale, and here the layout of the lottery playslip plays an important role. It has been speculated that such spontaneous purchases give rise to combinations which are far too evenly spread, that avoid numbers in the outer columns, that avoid contiguous values, etc. I intend in this section to investigate the validity of such hypotheses.

As a result of the reluctance by Camelot to disclose information on which numbers are chosen, the only accurate details we have for each draw are the number of people who chose the Jackpot combination. Consequently, weeks in which the number of Bonus prize winners differs most greatly from six times the number of Jackpot winners should be quite revealing. On many occasions the Jackpot is not won, so a ratio of Jackpot and Bonus winners is not always define. Hence, if J and B are the number of Jackpot and Bonus prize winners respectively, set $R = (B+6)/(J+1)$.

Now, for the first 100 draws, R takes a median value of 7.4, suggesting that for most draws the number of bonus winners is greater than we would expect, given the number of jackpot winners. R has been particularly large in Weeks 72 (38.0), 76 (34.0), 13 (29.0), 62 (27.0) and 15 (25.5). On most of these occasions, the Bonus number fills a relatively large gap between two main numbers. The exception is Week 62, where the Bonus number joined number 5 as the only two winning values less than 23.

R has been at its lowest in Weeks 70 (1.052), 31 (1.750), 9 (1.881), 14 (2.200) and 42 (2.375). There is certainly less of a common factor here, although in some cases (70, 9, 14, 42), the Bonus number was particularly high, and hence perhaps less likely to be selected when birthdates are used.

A - Minimum Separation

It has been suggested that gamblers' choices of combination are more evenly spread than a random selection would produce. One method of checking this is using the notion of minimum separation, mentioned earlier in Test 7.

We can do this by comparing the total number of tickets winning the Jackpot when the minimum separation takes various values to the expected number of winners if random selections were being made (Poisson model). Note that over all 100 weeks there were 526 Jackpot winners, against an expected 486.

Minimum separation	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
1	50	127	245.09	0.518
2	36	120	172.67	0.695
3	5	46	24.26	1.896
4	4	150	20.75	7.228
5	3	19	14.14	1.344
6	1	7	4.65	1.505
7	1	57	4.93	11.556

Using Spearman's Rank Correlation to compare the minimum separation to the ratio $\left(\frac{A}{E}\right)$, we get a p-value of around 10%, suggesting that there is a slight dependence upon minimum separation.

On the one occasion when the minimum separation was ≥ 7 (an event occurring with probability of less than 0.2%), over 11 times as many tickets as expected won the Jackpot. It would therefore appear that selections being chosen <u>are</u> too evenly spread to be random.

B - Maximum Separation

If the minimum separation between successive winning numbers affects the number of jackpot winners, then it seems likely that the maximum separation also will. The logic being that if people are selecting a fairly even spread of numbers, the maximum separation between them will not be that great.

Using a similar method to before:

Maximum separation	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
10	4	199	19.27	10.329
11	3	6	12.55	0.478
12	7	39	35.22	1.107
13	5	18	23.76	0.758
14	11	61	52.92	1.153
15	9	36	40.07	0.898
16	5	27	22.84	1.182
17	12	33	60.56	0.545
18	9	18	43.40	0.415
19	7	8	34.38	0.233
20	6	11	29.96	0.367
21	3	6	15.67	0.383
22	1	2	5.45	0.367
23	3	16	12.13	1.058
24	4	5	17.22	0.290
26	4	5	19.48	0.257
27	2	12	9.57	1.254
28	2	8	9.85	0.812
29	2	15	14.8	1.017
30	1	1	4.45	0.225

When we consider each and every value the maximum separation has taken, as above, there does not appear to be much link with the number of Jackpot winners. But suppose we consider as a group those cases when the maximum separation is greater than 15 (i.e. three rows of the lottery playslip):

Maximum separation	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
≤ 15	39	359	187.78	1.953
>15	61	167	302.71	0.552

So it would appear that the maximum separation of winning numbers does play some role in determining the number of Jackpot winners, though admittedly not to the extent that the minimum separation does.

C - Maximum Number of Contiguous Values

Continuing to examine the theory that players are selecting combinations with a very even spread of values, I have looked now at those weeks when two or more of the winning numbers were contiguous. It is worth noting that such situations are expected to occur on 49.5% of draws, and this has roughly been the case for the 100 draws so far.

Maximum number of contiguous values	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
0	50	399	241.41	1.653*
2	40	92	188.85	0.487
3	10	35	56.24	0.622

Again, it appears that this criterion does affect the number of Jackpot winners. Note, though, that excluding the aberrant Week 9 reduces (*) to 1.125, much closer to the overall ratio of 1.108.

Suppose now that we consider again the layout of the playslip. It is possible that players subconciously only recognise numbers as being contiguous when they are on the same row of the playslip. i.e. that they are not so reluctant to select, say, both 20 and 21. So if we discount such occurrences as being contiguous, then the above table becomes:

Maximum number of contiguous values (same playslip row)	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
0	57	422	272.34	1.550
2	38	93	183.03	0.508
3	5	11	31.13	0.353

which gives results more suggestive of a direct correlation.

D - Two Further Playslip-Related Criterion

The layout of the lottery playslip has been the subject of much speculation. It has often been rumoured that several thousand participants regularly choose combinations such as {1, 2, 3, 4, 5, 6}, {1, 6, 11, 16, 21, 26} etc. Unfortunately the credibility of such claims will only be known if such winning combinations do arise.

However, enough data has arisen to investigate various other propositions based on the playslip layout, and here we look at two such ideas.

Do lottery participants have a tendency to select combinations such that:

1. Each number lies in a different row?

2. No number lies in one of the two outermost columns?

Considering these in turn:

Proposition 1	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
True	17	275	82.29	3.342
False	83	251	404.21	0.621

Proposition 2	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
True	5	197	28.40	6.936
False	85	329	458.10	0.718

Which, once again, suggests that there is some validity in these propositions. Along with the previous three factors we have considered, there appears to be overwhelming evidence to suggest that the combinations being chosen tend to be affected greatly by the layout of the lottery playslip, and are far more evenly spread than random selection would yield.

E - Premeditated Combinations

We return now to examining combinations that arise not as a result of the playslip layout, but due to players selecting numbers based on dates or "lucky" numbers. Is there any evidence to suggest that this has had an effect on the proportion of winning tickets?

Consider first, then, the "birthday factor". If a large proportion of numbers are based on birthdates, then we would expect numbers in the ranges [1,12] and [1,31] to be selected often. In a similar manner to before, the following two tables consider the number of winning balls bearing values within these ranges.

Number of winning balls ≤ 12	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
0	21	60	103.29	0.581
1	34	229	163.34	1.402
2	28	172	131.59	1.307
3	14	39	72.26	0.540
4	3	26	16.02	1.623

Number of winning balls ≤ 31	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
1	5	9	24.47	0.368
2	9	16	44.36	0.361
3	19	182	93.36	1.949
4	41	219	203.77	1.075
5	25	95	115.25	0.824
6	1	5	5.28	0.946

The first table indicates little that is notable. On the occasions when not one of the six winning numbers has been under 13, the proportion of Jackpot winners has been somewhat small, but there is no clear pattern.

The second table is slightly more informative, suggesting that gamblers are loathe to select three or more numbers greater than 31 in their combinations.

It has also been posited that people are sceptical of selecting numbers in the forties. The following table looks at the number of winning balls over 39, and the effect, if any, that this has on the number of Jackpot winners:

Number of winning balls ≥ 40	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
0	11	27	53.24	0.507
1	54	398	257.25	1.547
2	23	86	115.23	0.746
3	8	9	41.91	0.215
4	4	6	18.88	0.318

It would certainly seem that there is a reluctance to pick a large number of values above 39. However, the appearance of exactly one "yellow" ball in the winning combination seems the preferred situation for many gamblers - once again suggesting an evenly spread selection.

Finally, I couldn't resist investigating whether the appearance of the numbers 7 and 13, traditionally thought of as being lucky and unlucky respectively, have any effect. This may not be as ridiculous as it sounds - in various overseas lotteries for which the marginal distribution of the integers is published, the number 7 is regularly selected considerably more than any other number.

Winning combination includes 7	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
Yes	13	199	63.06	3.156*
No	87	327	423.44	0.772

Winning combination includes 13	Frequency of occurrence	Actual number winning Jackpot (A)	Expected number winning Jackpot (E)	Ratio $\left(\frac{A}{E}\right)$
Yes	11	27	57.33	0.471
No	89	499	429.16	1.163

Here we have faint evidence to suggest that these historic superstitions are still acted upon, although the ratio (*) is extremely heavily influenced by the Week 9 data.

F - The "Lucky Dip" Influence

As of Draw 71, it has been possible to purchase "Lucky Dip" tickets, whereby the lottery computer "randomly" selects a combination. Between Draws 71 and 100, 7.95% of lottery tickets were bought in such a manner. Assuming that the Lucky Dips are random, we would expect this to have a somewhat "calming" effect on the proportion of winners, especially in the higher prize categories. The line of reasoning behind this is that the Lucky Dips should pick some of the combinations that humans are reluctant to, reducing the likelihood of a Rollover. Similarly, the Lucky Dips should be more uniformly distributed across all combinations, and hence there should be slightly fewer occasions on which a very large number of people win the top prizes.

It would appear that this has been the case. Over Weeks 1-70 there were 13 Rollovers, compared to just 4 over Weeks 71-100 (i.e. 19% of occasions compared to 13%). Similarly, between Weeks 1-70 10 or more tickets shared the Jackpot on 5 occasions, but on only one following the introduction of Lucky Dip (7.1% to 3.3%).

I decided to pursue this line of investigation further, to see if there is any evidence that the variance of the proportion of people winning prizes has altered since the introduction of Lucky Dip. The following table shows the sample standard deviations of the proportion of winning tickets in each prize category both Before and After the introduction of Lucky Dip. the ratio of these two values is then compared to an $F_{69,29}$ distribution.

	Jackpot	Bonus	Match 5	Match 4	Match 3	All Prizes
S_B	2.447×10^{-7}	4.713×10^{-7}	1.253×10^{-5}	3.122×10^{-4}	3.225×10^{-3}	3.519×10^{-3}
S_A	4.312×10^{-8}	2.582×10^{-7}	6.518×10^{-6}	2.524×10^{-4}	2.932×10^{-3}	3.182×10^{-3}
S^2_B / S^2_A	32.204	3.332	3.696	1.530	1.210	1.223
p-values	< 0.000005	0.000306	0.000112	0.103	0.289	0.278

The extremely small p-values for the larger prizes suggest that the variance of the proportion of winners has decreased since Lucky Dip was inaugurated. If the proportion of tickets purchased using the Lucky Dip method continues to rise, then we can expect this trend to continue, resulting in fewer rollovers (i.e. fewer abnormally large jackpots), but also fewer occasions when many people match all six balls (abnormally small jackpots).

Conclusions

At the beginning of this project, I posed two questions: Firstly, are the winning numbers chosen by the lottery machines random? It appears so. There certainly has not been any evidence over the course of the first 100 draws to suggest anything to the contrary.

Secondly, are the numbers chosen by gamblers random? As a whole, no. It appears that the major factor affecting the non-random selection of combinations is the physical layout of the playslip. In particular, players have a tendency to select a very evenly spread combination, rarely choosing adjacent values or "clusters" of numbers. Other factors which influence the choices to a lesser degree are a general reluctance to choose more than one "high" value, and the use of birthdays and other dates to generate combinations.

Given that a large number of combinations are selected in a non-random manner, than a gambler wishing to achieve the largest possible jackpot prize (should they win!) must try to select a combination unlikely to have been chosen by someone else. While it is, of course, impossible to know if you have chosen a unique combination, the chances of doing so can be improved by randomly selecting numbers, but then rejecting the combination if you have reason to believe it may be particularly popular. By purchasing a number of tickets in such a way, it has been suggested that it may be possible to <u>expect</u> a positive mean return, as opposed to merely dreaming of one.

Acknowledgements

I am grateful to Camelot, especially the National Lottery Line, who have kindly answered my various enquiries and mailed me the various data I have requested.

Much additional data, especially regarding the lottery balls and machines, was obtained from the lottery website managed by Richard K. Lloyd, who has also answered some of my queries on these matters.

The combinatorial techniques required to evaluate the tables of Maximum Contiguous Values on page 10 were explained to me by my project supervisor, Professor Ball.

Most of the statistical calculations were carried out using the software packages Minitab and Microsoft Excel.

I am also indebted to the many people who revealed their lottery combinations to me, along with the often bizarre explanations behind them.

References

Haigh, J (1995)	"Inferring Gamblers' Choice of Combinations in the National Lottery." I.M.A. Vol 31 Numbers 9/10 132-136
Haigh, J (1996)	"The Statistics of the National Lottery."
Joe, H (1993)	"Tests of Uniformity for Sets of Lotto Numbers." Statistics and Probability Letters 16 181-188

	Equipment Used		Draw Numbers							Number of tickets winning prizes						
Week	Machine	Ball Set							B	Jackpot	Bonus	Match 5	Match 4	Match 3	Total	Ticket Sales
1	Guinevere	1	3	5	14	22	30	44	10	7	39	2,139	76,731	1,073,695	1,152,611	48,965,792
2	Arthur	2	6	12	15	16	31	44	37	4	10	649	40,072	802,871	843,606	47,943,564
3	Arthur	2	11	17	21	29	30	40	31	0	8	693	43,792	843,672	888,165	48,263,533
4	Guinevere	2	26	35	38	43	47	49	28	1	10	473	28,244	657,025	685,753	61,495,564
5	Arthur	1	3	5	9	13	14	38	30	2	14	831	60,325	1,151,351	1,212,523	54,673,700
6	Arthur	2	2	3	27	29	39	44	6	1	7	365	35,321	987,567	1,023,261	55,234,642
7	Arthur	2	9	17	32	36	42	44	16	1	6	1,277	80,231	1,137,786	1,219,301	53,293,047
8	Arthur	2	2	5	21	22	25	32	46	0	7	601	46,339	978,819	1,025,766	57,529,347
9	Guinevere	1	7	17	23	32	38	42	48	133	246	6,660	165,738	1,932,744	2,105,521	69,846,979
10	Guinevere	4	6	16	20	30	31	47	4	7	22	865	50,393	903,050	954,337	61,157,534
11	Arthur	3	4	16	25	26	31	43	21	4	16	1,304	58,223	993,762	1,053,309	61,290,932
12	Arthur	2	1	7	37	38	42	46	20	1	4	487	27,456	807,216	835,164	62,170,560
13	Arthur	5	15	18	29	35	38	48	5	1	52	1,086	59,456	1,043,806	1,104,401	61,721,793
14	Guinevere	4	16	19	21	29	36	45	43	9	16	1,195	59,750	1,036,040	1,097,010	60,985,534
15	Guinevere	7	5	8	10	18	31	33	28	1	45	1,135	74,241	1,401,587	1,477,009	61,301,422
16	Arthur	8	11	12	17	26	36	42	13	1	7	846	51,172	1,064,943	1,116,969	61,302,793
17	Guinevere	7	2	13	22	27	29	46	36	2	46	1,402	80,473	1,304,496	1,386,419	67,688,637
18	Guinevere	4	9	18	19	24	31	41	21	2	25	1,210	71,061	1,277,813	1,350,111	62,479,486
19	Guinevere	2	4	17	41	42	44	49	24	0	8	251	30,079	845,813	876,151	62,227,682
20	Lancelot	6	22	25	30	32	41	43	29	2	11	642	50,717	1,103,456	1,154,828	76,269,508
21	Lancelot	8	14	17	22	24	42	47	34	3	37	760	62,896	1,316,087	1,379,783	63,753,399
22	Arthur	7	1	4	6	23	26	49	8	4	27	864	49,354	1,007,003	1,057,252	64,128,388
23	Arthur	8	8	18	20	33	36	38	46	0	17	961	64,270	1,203,869	1,269,117	63,309,485
24	Arthur	6	9	15	22	31	34	48	23	14	33	2,563	83,324	1,341,437	1,427,371	74,379,636
25	Lancelot	4	5	14	17	35	43	48	22	3	88	1,047	55,022	1,035,673	1,091,833	62,725,409
26	Arthur	2	7	16	25	26	28	41	19	3	28	982	62,065	1,189,896	1,252,974	62,879,549
27	Lancelot	6	15	16	17	28	32	46	22	0	7	568	43,642	982,790	1,027,007	62,298,146
28	Arthur	8	12	13	25	37	44	45	9	3	24	881	66,499	1,363,268	1,430,675	74,771,757
29	Arthur	4	1	21	29	31	32	40	27	0	11	517	40,623	931,366	972,517	64,826,761
30	Guinevere	1	12	15	26	44	46	49	14	1	6	535	37,157	890,537	928,236	72,201,984
31	Guinevere	2	27	30	33	38	40	48	2	7	8	751	45,108	976,283	1,022,157	66,006,436
32	Arthur	2	5	15	21	42	43	45	20	1	7	566	46,912	999,505	1,046,991	64,505,245
33	Arthur	8	5	7	8	25	44	48	3	0	11	709	53,187	1,172,809	1,226,716	63,881,931
34	Arthur	4	1	3	11	14	20	40	45	1	16	968	60,492	1,142,270	1,203,747	73,414,008

	Equipment Used		Draw Numbers							Number of tickets winning prizes						
Week	Machine	Ball Set							B	Jackpot	Bonus	Match 5	Match 4	Match 3	Total	Ticket Sales
35	Arthur	6	1	4	20	31	41	43	38	4	7	552	35,123	795,456	831,142	65,607,043
36	Guinevere	2	2	3	21	22	23	40	24	4	6	732	39,705	952,001	992,448	63,924,721
37	Arthur	7	28	34	41	45	46	49	11	3	6	565	32,007	684,405	716,986	63,812,748
38	Guinevere	8	1	8	25	30	35	45	15	2	63	751	49,225	880,947	930,988	63,405,879
39	Guinevere	4	11	25	28	33	34	47	48	0	6	556	42,929	950,309	993,800	63,525,142
40	Guinevere	8	5	8	23	24	28	48	19	5	30	1,160	82,044	1,588,909	1,672,148	74,575,649
41	Arthur	2	16	18	21	27	38	41	26	4	10	1,190	65,700	1,243,473	1,310,377	65,833,755
42	Guinevere	7	1	15	22	28	40	49	44	7	13	1,226	55,631	1,015,058	1,071,935	65,050,696
43	Guinevere	4	2	12	20	22	41	45	47	0	4	1,011	59,830	1,045,213	1,106,058	64,604,276
44	Arthur	2	2	10	14	25	37	41	5	8	41	1,314	62,929	1,186,903	1,251,195	76,663,008
45	Guinevere	4	5	10	19	24	34	46	28	10	29	1,120	60,622	1,126,075	1,187,856	66,212,779
46	Lancelot	8	10	11	29	32	33	40	16	1	24	705	48,644	1,041,765	1,091,139	65,805,302
47	Guinevere	7	10	22	28	30	36	37	45	3	11	676	55,827	1,170,017	1,226,534	65,608,291
48	Lancelot	8	4	5	9	25	30	47	17	5	32	969	56,689	1,143,504	1,201,199	65,828,800
49	Guinevere	2	2	6	17	19	21	47	5	3	33	1,243	67,572	1,233,455	1,302,306	65,759,630
50	Lancelot	6	7	16	27	33	35	44	5	5	58	1,924	88,896	1,434,454	1,525,337	65,944,108
51	Lancelot	7	6	14	18	27	44	48	1	3	35	1,655	81,956	1,335,232	1,418,881	65,551,611
52	Guinevere	3	7	10	23	28	30	48	3	5	36	1,319	70,944	1,326,485	1,398,789	65,805,646
53	Lancelot	5	4	7	18	33	45	48	1	20	51	2,052	87,665	1,364,652	1,454,440	66,693,414
54	Guinevere	6	16	23	28	30	42	46	45	3	7	764	49,805	1,035,679	1,086,258	65,820,282
55	Lancelot	2	15	16	19	26	35	46	7	1	35	693	44,730	903,662	949,121	65,437,794
56	Lancelot	3	5	11	12	26	29	33	20	3	36	1,310	77,667	1,367,335	1,446,351	65,060,440
57	Lancelot	6	7	8	23	28	35	49	10	1	23	914	66,768	1,344,104	1,411,810	64,973,661
58	Lancelot	2	6	11	34	40	47	49	16	0	5	441	33,322	800,660	834,428	67,920,715
59	Guinevere	4	6	32	39	42	43	45	36	0	15	570	52,156	1,176,218	1,228,959	78,428,855
60	Lancelot	7	2	3	4	13	42	44	24	3	53	1,524	100,140	2,282,389	2,384,109	127,824,795
61	Guinevere	6	21	29	31	32	34	48	25	0	12	455	39,551	956,338	996,356	74,832,559
62	Lancelot	3	5	23	25	30	33	37	3	0	21	1,084	81,069	1,728,211	1,810,385	86,753,064
63	Guinevere	8	16	17	38	41	42	43	28	4	60	1,713	79,694	1,628,683	1,710,154	106,159,714
64	Lancelot	2	2	9	22	26	32	44	40	6	38	2,010	100,276	1,640,117	1,742,447	78,125,931
65	Lancelot	7	4	11	14	15	28	42	6	6	30	1,745	88,945	1,609,965	1,700,691	75,496,220
66	Lancelot	4	4	14	15	16	18	22	33	5	16	1,243	72,891	1,340,880	1,415,035	73,903,189
67	Guinevere	8	2	5	7	24	35	44	30	4	19	1,385	81,659	1,606,795	1,689,862	73,962,371
68	Guinevere	3	9	11	12	24	41	45	6	2	19	956	63,770	1,277,338	1,342,085	73,026,110

	Equipment Used		Draw Numbers							Number of tickets winning prizes						
Week	Machine	Ball Set							B	Jackpot	Bonus	Match 5	Match 4	Match 3	Total	Ticket Sales
69	Guinevere	3	14	16	29	30	37	45	7	8	85	1,292	64,739	1,123,801	1,189,925	70,279,408
70	Guinevere	8	2	12	19	28	38	48	45	57	55	3,515	105,624	1,505,073	1,614,324	68,975,891
71	Guinevere	7	5	7	14	18	30	43	28	9	98	1,962	100,195	1,596,881	1,699,145	69,251,936
72	Arthur	6	12	26	27	28	37	49	43	0	32	700	52,940	1,132,669	1,186,341	69,098,713
73	Arthur	8	1	4	6	14	17	38	9	5	40	1,262	80,700	1,486,719	1,568,726	79,821,776
74	Guinevere	6	23	38	40	44	47	49	12	3	16	933	45,190	920,936	967,078	69,652,526
75	Arthur	7	9	28	29	31	40	48	23	2	20	914	55,687	1,127,020	1,183,643	69,363,752
76	Arthur	2	4	6	11	18	31	48	41	0	28	1,405	75,194	1,326,950	1,403,577	68,877,731
77	Arthur	4	6	25	26	33	34	47	49	2	15	909	58,291	1,173,710	1,232,927	79,882,225
78	Arthur	8	7	10	12	22	34	48	11	8	16	1,766	84,392	1,442,183	1,528,365	70,174,699
79	Guinevere	6	4	12	13	33	40	46	41	1	12	1,138	63,001	1,165,145	1,229,297	69,687,983
80	Arthur	7	8	20	26	34	42	43	25	5	23	1,950	73,917	1,164,449	1,240,344	69,781,658
81	Arthur	2	24	35	36	37	39	45	20	2	16	757	44,324	851,243	896,342	68,733,675
82	Arthur	4	11	15	17	25	32	46	29	2	23	1,139	64,621	1,153,534	1,219,319	67,580,316
83	Arthur	7	13	18	25	44	46	47	34	0	13	545	42,343	966,777	1,009,678	66,461,361
84	Lancelot	7	3	4	7	11	17	40	20	9	42	1,834	112,922	1,784,431	1,899,238	77,736,654
85	Lancelot	6	4	17	27	34	35	46	7	2	45	1,002	56,588	1,083,529	1,141,166	68,626,062
86	Lancelot	8	13	26	43	44	45	47	36	0	11	546	44,783	933,653	978,993	68,289,582
87	Arthur	1	5	10	11	12	41	42	2	12	41	1,486	71,433	1,459,553	1,532,525	78,487,718
88	Arthur	3	6	14	20	25	34	44	45	7	19	1,077	59,971	1,143,996	1,205,070	68,102,609
89	Guinevere	2	2	13	19	21	32	45	9	4	59	1,656	69,564	1,220,187	1,291,470	68,196,294
90	Lancelot	4	13	17	26	28	31	36	44	4	18	840	57,219	1,182,744	1,240,825	67,828,482
91	Lancelot	7	3	23	36	38	41	45	44	2	9	717	49,356	1,008,409	1,058,493	67,503,030
92	Lancelot	5	2	28	33	39	42	44	46	1	18	1,023	66,807	1,285,279	1,353,128	66,784,509
93	Guinevere	1	8	11	14	18	33	44	34	7	31	2,040	98,072	1,579,736	1,679,886	67,864,651
94	Guinevere	3	3	5	14	27	44	47	43	7	17	1,687	86,927	1,404,876	1,493,514	67,640,227
95	Guinevere	6	5	13	15	18	32	44	41	5	35	1,679	88,862	1,482,834	1,573,415	67,704,249
96	Lancelot	3	2	9	10	11	38	48	1	5	12	889	57,998	1,169,254	1,228,158	67,769,246
97	Guinevere	7	7	8	12	30	35	41	47	1	17	903	53,547	1,138,159	1,192,627	68,459,299
98	Guinevere	3	19	23	26	31	36	39	3	4	15	765	48,449	1,045,121	1,094,354	68,644,687
99	Lancelot	2	6	9	25	45	47	48	14	2	23	679	45,430	1,008,996	1,055,130	68,915,135
100	Lancelot	1	15	16	25	30	39	45	14	1	17	846	56,382	1,099,160	1,156,406	70,065,289
TOTALS										526	2,724	114,166	6,272,214	118,283,999	124,673,629	6,803,103,754