USING PERCEPTIONS OF RANDOMNESS TO TEACH STATISTICS TO FIRST YEAR STATISTCS STUDENTS IN SOUTH AFRICA

JI de Wet
Free State University, South Africa
dewetji.blms@mail.uovs.ac.za

The perception of randomness is very important in the statistical science. It is therefore of critical importance that statistics students master the concept of randomness. It is usually difficult to change deep rooted perceptions of randomness. Many statistics students in South Africa are so-called historically disadvantaged students. They are classified as such because they generally received poor mathematics teaching at school level compared to other statistics students. The lottery game can be regarded as an ideal opportunity to teach randomness and other statistical concepts. Three data sets are compared: The lotto winning numbers, random numbers generated and the random numbers generated by two groups of first year statistics students.

INTRODUCTION

In any lottery game n winning numbers are selected from N possible numbers. In South Africa 6 numbers from 49 numbers 1,2,……,49 are selected using a mechanical device. The method consists of the selection of 6 numbered balls, one after another, without replacement from a glass box containing 49 numbered balls. It is generally accepted that the method is fully random. This means that any of the ${}^{49}C_6$ combinations of 6 numbers has the same chance of being selected.

The lottery game can be regarded as an ideal opportunity to teach randomness and other statistical concepts. The reason is that playing the lotto is very popular in South Africa, especially amongst students.

The results of Boland and Pawitan (1999) were used in this study and applied to the South African situation.

THE EXPERIMENT

First year statistics students were asked to choose 6 numbers *at random in any order*, from the 49 numbers 1, 2,….., 49 as if they were playing the Lotto. This was done during a class test to ensure that the results are objective. A total of 1054 students participated. The student data comprises two sets, namely a total of 334 Afrikaans speaking students and a total of 710 English speaking students. The majority (90% plus) of the " historically disadvantaged students" fall in the latter group. The data were cleaned from "illegal" choices. For example, zero's, numbers repeated, and numbers bigger than 49. From the Afrikaans group, 97% were correctly completed against 93.7% from the English group. This resulted in 335 Afrikaans students and 665 English students.

The results were compared with 456 available winning numbers of the Lotto since its inception in March 2000 until 5 October 2005 and with a simulated data set of 500 Uniformly distributed numbers.

The results can be used to teach several statistical concepts, e.g., *sampling distribution of the mean, goodness-of –fit tests,* the *Central Limit Theorem, simulation,* etc.

RESULTS

The data were analyzed using Microsoft Excel. This is software readily available to all students and proved easy to use.

MEANS OF SUCCESSIVE CHOICES

The following figures present histograms for the means of successive choices of the four datasets.

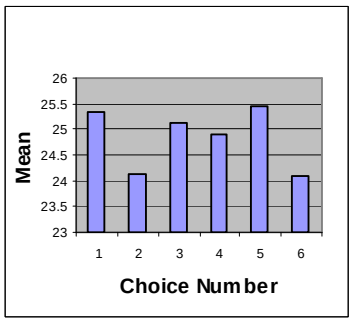

Figure 1: Means of successive choices: Lotto data

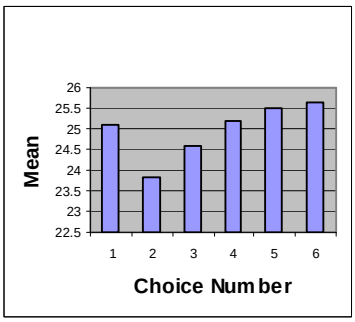

Figure 2: Means of successive choices: Simulated data

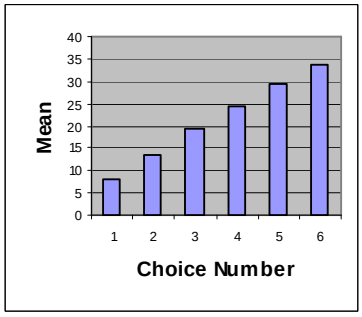

Figure 3: Means of successive choices Student data: Afrikaans

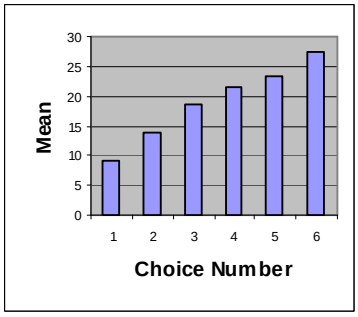

Figure 4: Means of successive choices Student data: English

In the case of truly randomly selected numbers there should be no trend in the means of successive choices. From Figures 1 and 2 randomness is obvious, which is not the case in Figures 3 and 4. Here an obvious trend exists showing a tendency to select smaller numbers first. Similar results were found by Boland and Pawitan (1999).The data can also be used to compare the means of successive choices using proper statistical tests in each of the datasets.

FREQUENCY DISTRIBUTIONS OF NUMBERS SELECTED

The frequency distributions of the numbers selected are shown in the figures below. The horizontal line on each graph indicates the expected frequency if the selections are truly random. The expected frequencies are: Lotto winning numbers: 55.8; Simulated numbers: 61.2; Students (Afrikaans): 41.0; Students (English): 81.4.

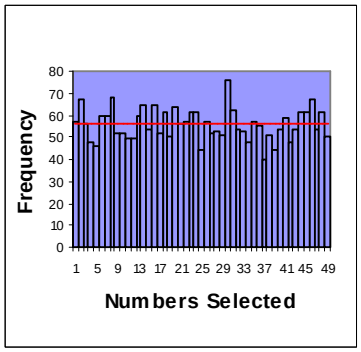

Figure 5: Frequency of numbers Selected: Lotto data

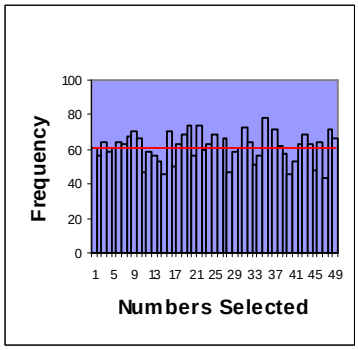

Figure 6: Frequency of numbers Selected: Simulated data

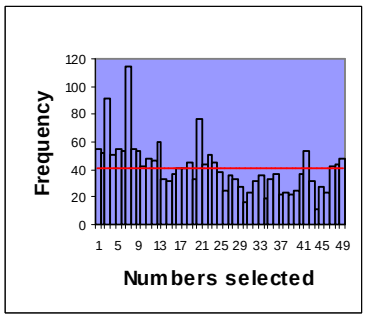

Figure 7: Frequency of numbers Selected: Students Afrikaans

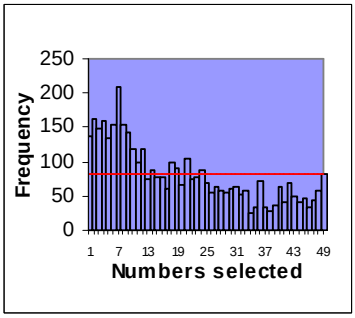

Figure 8: Frequency of numbers Selected: Students: English

From Figures 5 and 6 it is clear that in the case of the Lotto and simulated data no preference for specific numbers exists. In the case of Figures 7 and 8 (student data) it is obvious that there is a tendency to select smaller numbers. It is even more obvious in Figure 8 (Student data: English). It seems that the English group's understanding of randomness is somewhat weaker than the Afrikaans group.

DISTRIBUTION OF THE SAMPLE MEANS

The distribution of the sample means can also be studied using the results of the Lotto experiment. Although the sample size is relative small $(n = 6)$ there were a large number of samples. Let $x = (x_1, x_2, ...x_n)$ represent **Error! Objects cannot be created from editing field codes.**numbers in the sequence they have been selected from $\{1.2,.....,n\}$. Then, if the selections are truly random, $E(\bar{x}) = \frac{N+1}{2}$ and $Var(\bar{x}) = \frac{(N+1)(N-n)}{12n}$.

Thus, in the case of random selections of 6 from 49 we expect: $\bar{x} \sim N(25.5, 5.46^2)$.

The distribution of the sample means, with the expected normal distribution, for random selections, for the different data sets, are given below:

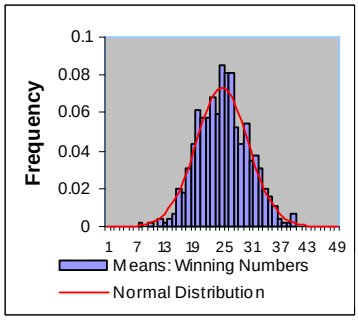

Figure 9: Distribution of Sample Means: Lotto data

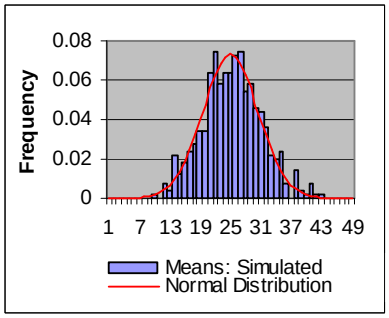

Figure 10: Distribution of Sample Means: Simulated data

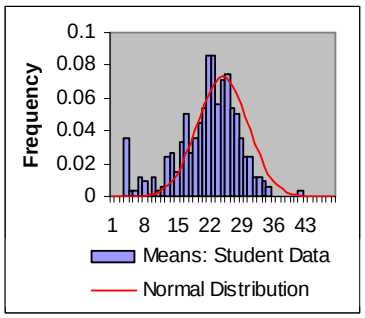

Figure 11: Distribution of Sample Means: Students: Afrikaans

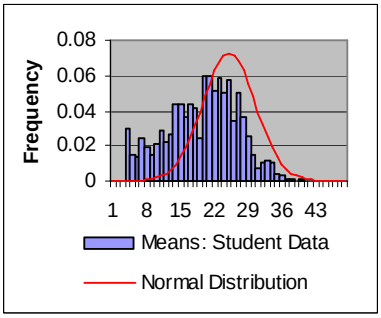

Figure 12: Distribution of Sample Means: Students: English

From the above figures it is clear that the distribution of sample means in the case of the Lotto data and the Simulated data is fairly normal, which is not the case for the student data.

TESTS FOR UNIFORM MARGINAL DISTRIBUTIONS.

The asymptotic distribution, under the hypothesis of true randomness, of

$X^2 = \frac{N-1}{N-m}\sum\frac{(O_i - E)}{E}$ is χ^2 with $N-1$ degrees of freedom (Joe, 1993).

The results of our experiment appear in Table 1.

Table 1: Testing for Uniform marginals

Data Set	Lotto	Simulation	Afrikaans	English
n	456	500	335	665
X^2	48.31	60.32	444.79	1178.81

With $\alpha = 0.05$ we have $\chi^2_{48,0.95} = 65.17$. The hypothesis of a Uniform marginal distribution is clearly rejected in the case of the student data. This result confirms the earlier results.

It is interesting to discuss the results with the students. Especially the tendency to select smaller numbers. This can be used to teach the concept of randomness, bearing in mind the student's perceptions.

SUMMARY

- It is useful to use Lotto data to determine students' perceptions of randomness.
- This type of data catches the imagination of students.
- The data can be used to teach a variety of elementary statistical concepts.
- The outcome of our experiment is that first year South African students of statistics do not really know what is meant by randomness. It was even more so in the case of historically disadvantaged students.
- There was no reason to doubt the randomness of the Lotto in South Africa.

REFERENCES

Boland, P. J. and Pawitan, Y. (1999). Trying to be random in selecting numbers for Lotto. *Journal of Statistics Education*, 7(3).

Joe, H. (1993). Tests of uniformity for sets of lotto numbers. *Statistics and Probability Letters*, 16, 181-188.