

# INTERACTIVE SIMULATIONS OF STATISTICS WITH EXCEL

Santiago-Acosta RD\*, Hernández-Cooper EM and Yescas-Martínez F

*Tecnológico de Monterrey, Mexico*

---

**Abstract:** Generally speaking, students fail to understand many statistical concepts and are not able to use them for finding solutions to real problems. The goal of this work is to determine the impact of interactive simulations for the solution of statistical problems. The proposed method consists of four stages. During the first stage, students tackle several situations through experience and computer simulations. Then, the theory is established and some problems are worked out in class. During the last stage, a conceptual exam is solved by the students. Twenty eight students took part in the present work. The performance of students in this group was compared with the performance of students that belonged to a control group. The results illustrate that students were able to show a significant improvement in their performance through the strategy proposed in this work.

**Keywords:** Interactive simulation, Inferential Statistics, Probability

---

## Introduction

A special interest exists in teaching probability and statistics when dealing with concepts such as: random experimentation, probability, random variables, probability distribution functions among others. The main reason is that several miss conceived ideas about probability appear due to a wrong interpretation of these concepts. The basic concept in this area is known as the probability distribution function of one random variable. This concept is one of the key ideas involved in the study of probability and statistics, since it is related with many subjects related with randomness which must be incorporated in any course on probability and statistics (Batanero, 2001).

Several studies indicate that students deal with great difficulties to understand the concepts related with probability (Benko, 2006). According to Fischbein (1975), intuition is a key aspect for developing probabilistic thinking and generally, our intuition can be misleading. Other studies indicate that there is a conflict between the basic ideas on probability and our own experiences, which is added to difficulties that emerge when learning and using ill-conceived concepts in order to solve problems on probability (Batanero & Sánchez, 2005).

Additionally, most professors tend to teach probability by using a structured scheme where concepts are studied and problems are solved afterwards. We consider that in order to reduce the difficulties that students have when learning probability and statistics, and increase their skills when using probabilistic models, it is necessary to use computational tools that allow simulations of several scenarios of their interest. As pointed out by Batanero and Sánchez (2005), with the help of numerical simulations of random data and using frequency distributions, the understanding of probability distribution functions seems more natural for students. This offers some logic to the mathematical expression defined by the probability function. Additionally, studies show some advantages when using computational tools when performing simulations of stochastic processes (Kuhn et al, 2006).

\*Corresponding Authors' Email: [ruben.dario@tec.mx](mailto:ruben.dario@tec.mx)

The selection of didactical tools represents an important step on curricular design. The strategy that has been adopted in this work, is based on the Activity-Class-Exercise cycle (Dubinsky,2002). The activity part consists of an exploration of concepts through the simulation of several processes, which motivates the study of probability concepts. The lecture part of the cycle, consists of reviewing theoretical models, which are compared with experimental results. During the exercise sessions, some problems in a specific context are solved.

Our proposal is to use a strategy based on the ACE cycle, where experimentation of probabilistic concepts is achieved through interactive simulations in Excel. The subject of discrete probability distributions is selected because it includes alternative ways of learning through simulations, which are useful on complementary subjects that are covered on probability and statistics courses.

## METHODOLOGY

### *Proposal*

The proposed methodology is based on activities in the Excel package, where the following discrete probability functions are explored: Binomial, Hypergeometric, Geometric, Negative Binomial, Poisson and Multinomial. The following didactic sequence for the study of each one of the mentioned distributions is:

- Simulation and exploration of an example related with any of the mentioned distributions.
- Discussion, generalization of the problem and simulations in teams of students with four members.
- Discussions in class about the experience and the theoretical model that describes the results and solutions to examples within a context of interest
- Online evaluation in Microsoft Forms.

As an example, students are asked to toss a coin until they obtain two tails, and they must repeat the experiment 10 times and write down a relative frequency table as shown in table 1. Afterwards, students share their results with the entire group and these results are written in a table. The final step consists of results discussion through teams of students.

*Table 1: Example of the frequency table made by students*

number of tosses	0	1	2	3	4
Relative frequency					

The final step of the first stage consists in a simulation of this process in Excel, which is generalized by using Visual Basic. The problem just described, is analyzed through the code shown in Figure 1.

```

Sub calculabinneg()
Call borrabinneg
k = Cells(3, 2)
p = Cells(2, 2)
num = Cells(4, 2)
For i = 1 To num
r = 0
q = 1
j = 1
While q = 1
If p > Rnd Then
Cells(i + 4, 4 + j) = 1
r = r + 1
If r = k Then
Cells(i + 4, 4) = Cells(4, 4 + j)
q = 0
End If
Else
Cells(i + 4, 4 + j) = 0
End If
j = j + 1
Wend
Next i
End Sub
    
```

Figure 1. Visual Basic Code of a process with a Negative Binomial distribution

The only instructions required are: Cells, While-Wend, If-Then-Else-Endif and For-Next. These instructions are briefly described below:

- Cells (i,j) allows to read and write on the cell located at row i and column j in the Excel sheet.
- While-Wend is an instruction used to generate a cycle that runs as long as the variable q is false
- For-Next is a cycle that is repeated as many times as indicated.
- If-Then-Else-Endif is a control structure that allows to execute a determined process, depending on whether a variable is true or false.

Figure 2 shows the Excel sheet that was obtained. A data section where users are able to fine tune the success probabilities for obtaining a tail, the number of coin tosses and the number of repetitions of the experiment are illustrated. Additionally, the relative frequency tables obtained are shown in Figure 2.

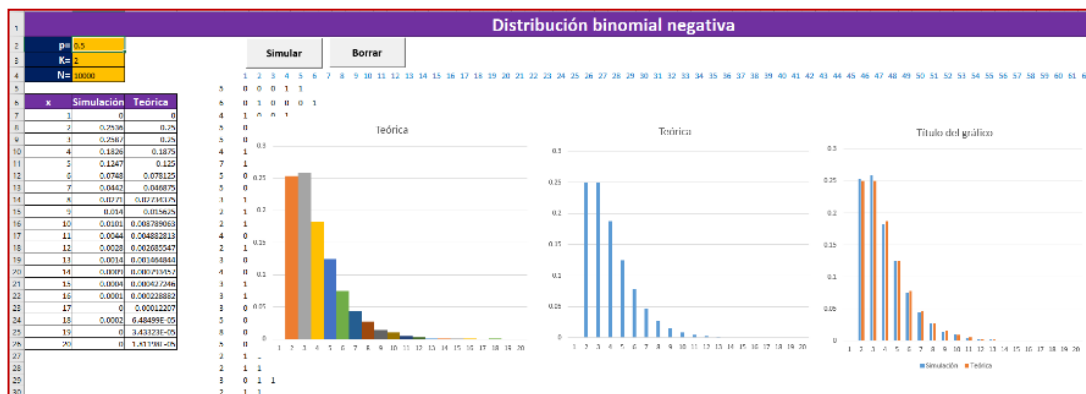


Figure 2: Excel sheet illustrating a process with a Negative Binomial distribution

The third part of the didactic sequence consists of discussions on the model known as the Negative Binomial distribution. Examples on the simulation and practical applications of the distribution are performed. In figure 2, theoretical results are also shown and these are compared with results obtained from the simulations.

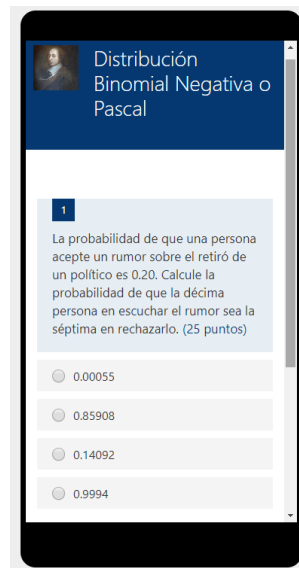


Figure 3. Evaluation of the Negative Binomial distribution performed in Forms-Microsoft.

Finally, during the fourth stage, an evaluation is performed through Forms-Microsoft, as illustrated in figure 3.

### **Classroom Experience**

The simulations of discrete probability distributions in Excel were performed during the fall semester of 2018. Twenty-eight engineering students took part on this course at Tecnológico de Monterrey-Campus Estado de México. Students worked in teams constituted by four members. The work performed by the students was based on a Visual Basic workshop and its application to Excel. A computational lab was used during the first two stages proposed in this work. Six sessions of one and a half hours were dedicated for this purpose. The reports were analyzed, using a rubric, and a test related with discrete distributions was applied. Finally, a survey on the learning environment was applied to the students. The performance of the control and experimental groups of students was evaluated through this test.

### **Results and Discussion**

This section is dedicated to discuss some results obtained in this work. Reports of six activities related with the Binomial, Hypergeometric, Geometric, Pascal, Poisson and Multinomial distributions are illustrated in figure 4. The rubric used considered five aspects: expertise in the use of Excel and Visual-Basic, discrete distribution modelling, generalization of models and results on the evaluation of concepts management., On the one hand, the results show that students were able to develop

abilities in the use of Excel and Visual-Basic. On the other hand, there is a lack of deep analysis in the relation between the theoretical model and the results obtained from the simulations.

The test applied to students considered ten questions related with the discrete distribution. Figure 5 shows the results of this test. The arithmetic mean on each question for the experimental group is slightly higher than the mean obtained by the control group. The performance of the control group is slightly better on questions associated with the Poisson distribution. This result may be related with the small amount of time spent on this topic by the experimental group. The results are consistent with those reported by Batanero and Sánchez (2005), regarding ill-conceived ideas in probability theory. The experimental group were able to reduce these errors through simulation, while the control group maintained the conventional didactic strategy.

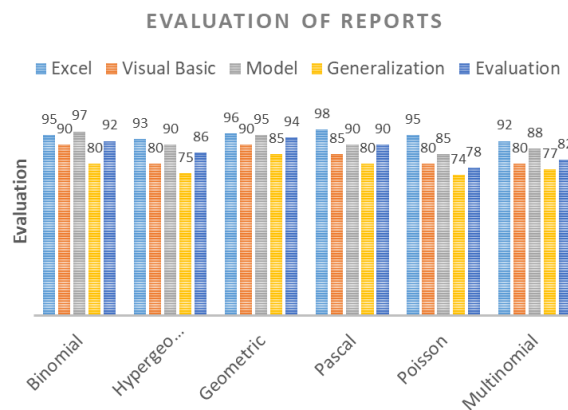


Figure 4. Results obtained from reports on interactive simulations.

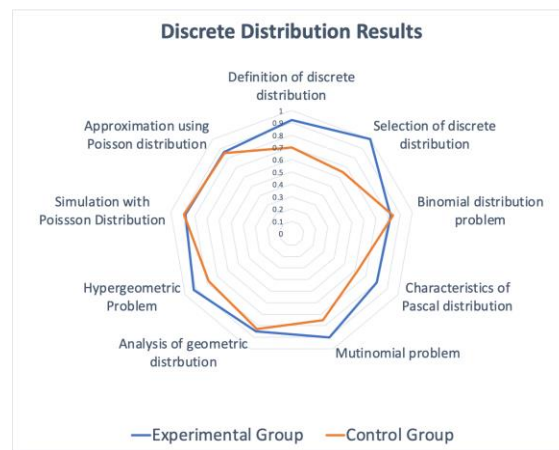


Figure 5. Results from the discrete distributions test.

The results taken from the survey on the student's opinion about these type of activities are shown in figure 6.

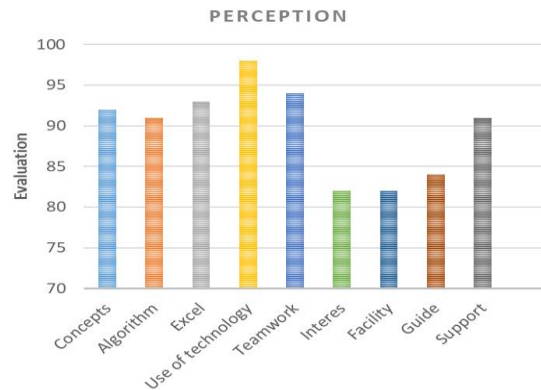


Figure 6. Student's perception on the course.

According to the results shown in figure 6, most students consider that the way in which the model and the solution is obtained through the algorithm, is simple (Algorithm); however, they need a wider guidance to develop the application (guide). Additionally, they consider that the use of Excel is appealing, even when the activities developed are very complex. According to the students in this survey, the use of technology helps to improve their understanding on concepts related with discrete distribution.

## Conclusions

Students apparently have some difficulties when performing simulations in Excel. We considered that files containing the simulations should be provided to the students as a first approach. However, this option was not considered in the long run because students were not proficient in the use of these files. For this reason, we decided to offer a workshop on the use of Visual Basic and its applications with Excel. The workshop allowed students to understand instructions in simple simulations during class and even the more complicated simulations that were performed outside the classroom. On the one hand, students were able to develop their programming skills. On the other hand, the students could gain a better understanding of probability related concepts and were able to build complex simulations in a simple manner. The students presented their works in a Project Divulcation Workshop organized by the Science Department, which shows a major improvement on the student's knowledge.

In general, students were able to understand the differences between several types of discrete distributions and to apply them with fewer difficulties in different contexts. Additionally, students were able to conclude that randomized experiments constitute an appealing experience that helps to understand probabilistic connections. The students were able to use theoretical models independently, and to build computer-based models in order to explore simple and complex problems. Results illustrate that students were able to construct meaningful representations. Complex problems related with discrete distributions were explored and successfully solved by students. These results may indicate that the process of modeling, performing computer simulations, analyzing and discussing, leads to a greater confidence in and a deeper understanding of methods to estimate probabilities. Additionally, with the ACE methodology, students feel more motivated and seem more attracted to the idea of gaining deeper understanding of probability theory. Finally, ACE cycles, offer an alternative to organize probability and statistics courses.

In this work, we focused on the potential of technological tools usage to improve the learning experience in probability courses. We also considered more general goals, concerning the use of technology in more complicated topics. Recently, we began the study of interactive simulations to improve the learning experience related with Statistical Inference (confidence intervals and hypothesis test). Complementary results related to this work, are expected to improve the teaching and learning experience in Probability and Statistics courses.

### **Acknowledgements**

The authors would like to acknowledge the financial support of Writing Lab, TecLabs, Tecnológico de Monterrey, Mexico, in the production of this work.

### **References**

- Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Universidad de Granada.
- Batanero, C., & Sanchez, E. (2005). What is the Nature of High School Students' Conceptions and Misconceptions About Probability? In *Exploring probability in school* (pp.241-266). Boston, MA: Springer.
- Benko, P. (2006). *Study of the development of students' ideas in probability*. Rutgers: NJ: New Jersey University.
- Dubinsky, E. (2002). Reflective abstraction in advanced mathematical thinking. In *Advanced mathematical thinking* (pp. 95-126). Dordrecht, Holland: Springer.
- Fischbein, E. (1975). *The intuitive source of probability thinking in children*. Dordrecht, Holland: Springer.
- Kuhn, M., Hoppe, U., Lingnau, A., & Wichmann, A. (2006). Computational modelling and simulation fostering new approaches in learning probability. *Innovations in Education and Teaching International*, 43(2), 183-194.