The final publication is available at

https://library.iated.org/publications/INTED2016

Additional Information

TEACHING BINOMIAL AND POISSON DISTRIBUTIONS IN AN UNDERGRADUATE ENGINEERING COURSE

Carmen Capilla[1]

[1]*Polytechnic University of Valencia (SPAIN)*

Abstract

The teaching of discrete probability distributions is discussed in this paper. Seven groups of students involved in an undergraduate computer science degree are considered. The main goals are teaching students to calculate probabilities by hand and using statistical software. Binomial and Poisson models are introduced, and applied in attribute quality control exercises. The activities and assessments are presented, and the evaluation results are discussed.

Keywords: Discrete probability distributions, quality control, computer science, undergraduate students.

1 INTRODUCTION

The objective of this work is to discuss the teaching of the Binomial and Poisson probability distributions in an undergraduate computer science engineering degree. Seven groups of students are considered (Table 1). They were in the second semester of the degree. Three groups followed the course in the academic year 2013_14, and the other four in 2014_15. In two groups of the first year (groups 1 and 2) and in another two of the second one (groups 4 and 5), the basic concepts of these discrete distributions were explained in two classroom lectures of 90 minutes, by the same lecturer in Spanish. Section 2 describes the teaching methodology. In one group of the first year and in another one of the second one (groups 3 and 7), the classroom lectures were given by another teacher in Valencian. There was another group (6) in 2014_15 with a different teacher for the classroom sessions in Spanish. There was a computer laboratory class related to this subject. It had the same teacher for the seven groups and the number of students involved in them is given in Table 1. These sessions are commented in Section 3. The practice assessment and its results are presented in Section 4.

Table 1. Groups following the computer laboratory class on discrete probability distributions in the academic year 2014_15.

Group	Number students	Academic year	Language
1	41	2013_14	Spanish
2	47	2013_14	Spanish
3	29	2013_14	Valencian
4	45	2014_15	Spanish
5	50	2014_15	Spanish
6	48	2014_15	Spanish
7	32	2014_15	Valencian

2   TEACHING METHODOLOGY IN THE CLASSROOM SESSIONS

The main goals when teaching this unit, was that students learn to calculate specific probabilities and to apply several statistical equations. The concepts were explained in the classroom using power project documents, distributions tables, the list of formulas elaborated by the teachers, solved problems, and a collection of exams completed by students of previous years. The course notes were based on [1-5, 7]. They could consult any on-line source of information, using their personal computers, tablets or mobile phones. First, the binomial distribution theoretical definition was introduced, with some examples of random variables to illustrate it. Its mass probability formula was presented. This model dependence on its parameters (event probability and number of trials), was graphically illustrated using statistical software (Fig. 1).

The binomial equation was applied to an attribute quality control [6] exercise. The acceptance probability of a sampling plan was computed, considering a given sample size, and a criterion to decide its acceptability. This equation application was also illustrated with a dice-rolling example. The formulas to compute the binomial distribution central location and dispersion measures (mean and variance) were introduced, as well as their use with the above mentioned exercises. The approximation of this model with the Gaussian distribution was enunciated. This property was studied in the next unit. Finally the binomial distributions sum was given with its application in a quality control problem.

In the second part of the discrete distributions unit, the Poisson model definition was explained with two examples related to the computer science engineering school (CSES). The plot (Fig. 2) and formula of the mass probability function, and the distribution mean and variance were introduced. The equation application was illustrated with two examples: the first one related to the CSES parking, and the second one to daily traffic mortal accidents in the Spanish roads (considering only days from Monday to Friday). The calculations with a plot of the cumulative probability curves were explained. Two exercises (CSES parking and quality control) were solved using this plot. The sum of independent Poisson variables was tackled in the last part of this distribution introduction, with a computation of the probability of a certain number of traffic mortal accidents from Monday to Friday. In both distributions, the problems probabilities were also obtained using statistical software.
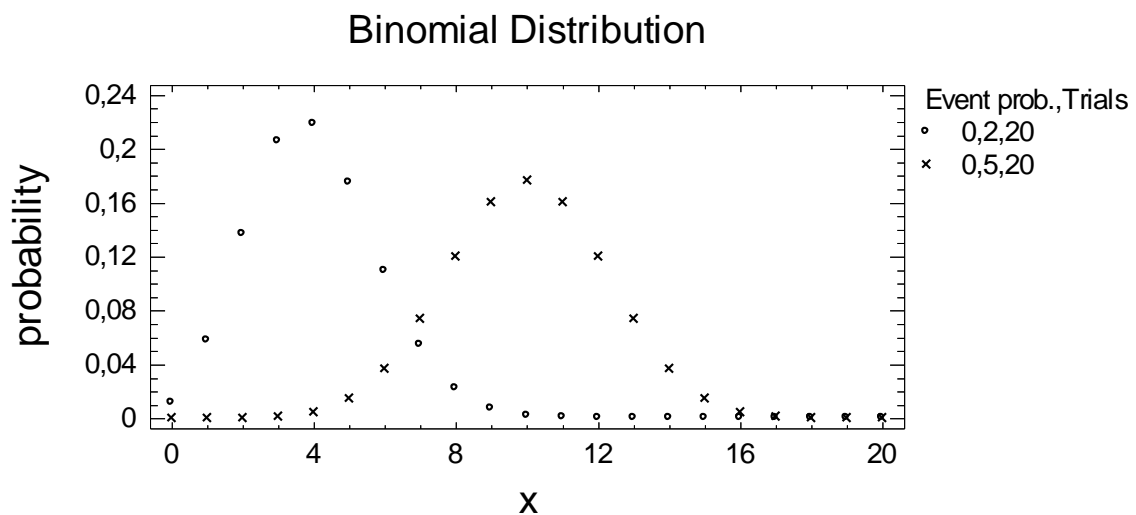

Figure 1. Mass probability plot of Binomial models with event probabilities 20% and 50%, and number of trials 20.
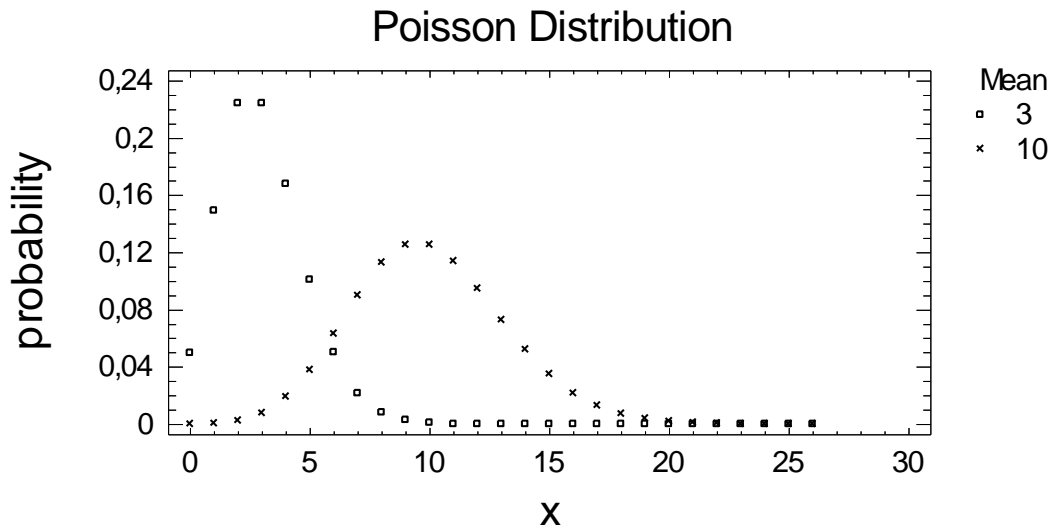
Poisson Distribution

Figure 2. Mass probability plot of Poisson model with means 3 and 10.

There was a written assessment of the classroom lectures after the concepts and methods explanation. Students solved problems with a collaborative or non-collaborative approach, and individually presented consensus or non-consensus answers. They could consult any bibliography during the exercise. In the academic year 2013_14, the two groups' assessment consisted of two exercises. The first one was an attribute quality control problem (number of defective units in a hypothetic LCD screens production process). The aim was to evaluate students' application skill of the binomial cumulative probability, mean and standard deviation formulas. The second exercise posed three probability computations on the number of bugs corrected by CRC codes, when sending packages through the net. It was based on the last exercise of the computer laboratory session [8]. The answer had to be found using the Poisson cumulative probability plot, the mass probability formula and the Poisson variables sum property. The task had to be completed in a maximum time of 1 hour. In the 2014_15 course, one group was evaluated with the same binomial exercise in the classroom. 36 students completed the assessment (class attendance 81,8%). The other group was evaluated with a homework assignment (the same two exercises of previous year) due for next week. 96,1% of students completed this assignment. Table 2 and 3 contain the marks frequencies, means and the standard deviations.

Table 2. Frequency tabulation for the binomial distribution assessment of group 4 (academic year 2014_15).

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		4	0,1111	4	0,1111
1	0,0	1,0	0,5	0	0,0000	4	0,1111
2	1,0	2,0	1,5	2	0,0556	6	0,1667
3	2,0	3,0	2,5	0	0,0000	6	0,1667
4	3,0	4,0	3,5	0	0,0000	6	0,1667
5	4,0	5,0	4,5	0	0,0000	6	0,1667
6	5,0	6,0	5,5	0	0,0000	6	0,1667
7	6,0	7,0	6,5	2	0,0556	8	0,2222
8	7,0	8,0	7,5	0	0,0000	8	0,2222
9	8,0	9,0	8,5	6	0,1667	14	0,3889
10	9,0	10,0	9,5	22	0,6111	36	1,0000

Mean = 8,04167   Standard deviation = 3,47722

Exercises feedback was provided to the students by giving them their corrections, followed by explanations of the solving procedure, and comments on the most frequent errors and misunderstandings. In these explanations students participated by writing the exercises solutions on the blackboard, and posing questions on them.

Table 3. Frequency tabulation for the discrete probability distributions assignment of group 5 (academic year 2014_15).

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		10	0,2041	10	0,2041
1	0,0	1,0	0,5	0	0,0000	10	0,2041
2	1,0	2,0	1,5	1	0,0204	11	0,2245
3	2,0	3,0	2,5	0	0,0000	11	0,2245
4	3,0	4,0	3,5	0	0,0000	11	0,2245
5	4,0	5,0	4,5	1	0,0204	12	0,2449
6	5,0	6,0	5,5	0	0,0000	12	0,2449
7	6,0	7,0	6,5	2	0,0408	14	0,2857
8	7,0	8,0	7,5	6	0,1224	20	0,4082
9	8,0	9,0	8,5	4	0,0816	24	0,4898
10	9,0	10,0	9,5	25	0,5102	49	1,0000

Mean = 7,20408   Standard deviation = 3,97639

Fig. 3 shows the multiple box-and-whisker plot of both groups marks. In this same year (2014_15), the discrete distribution subject was also assessed with a problem in the midterm exam, proposed by another teacher of the same subject. The question was related to a quality control problem (sample size computation), in a communication network context. The evaluation goals were to assess students' individual skills applying the Poisson distribution, and its cumulative probabilities plot, and using probabilistic and statistical terminology to explain the solving procedure. Descriptive statistical analysis of the two groups' exam question marks, is given in Table 4. The Man-Whitney W test is in Table 5, indicating that there is significant difference between the medians. The distributions comparison is included in Table 6 (Kolmogorov-Smirnov test). The quantile plot of both distributions is in Fig. 4.
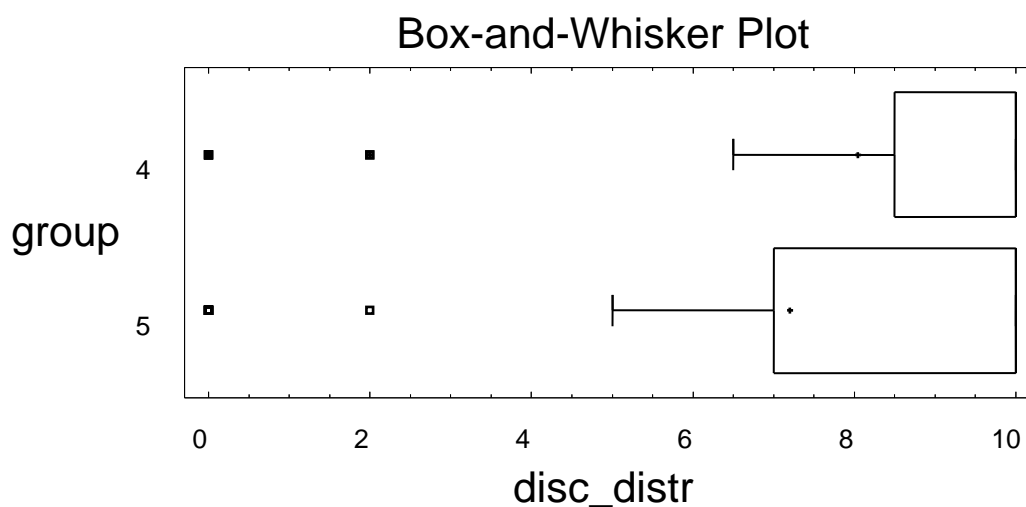

Figure 3. Multiple box-and-whisker plot of the two groups marks for the academic year 2014_15, in the discrete probability distributions assessments.

Table 4. Summary Statistics for the exam question marks (academic year 2014_15)

```
                    group=4         group=5
--------------------------------------------------------------
Count                44              50
Average             3,29545         4,92
Median              3,0             5,0
Variance            11,6084         18,0751
Standard deviation  3,4071          4,25148
Minimum             0,0             0,0
Maximum             10,0            10,0
Range               10,0            10,0
Lower quartile      0,0             0,0
Upper quartile      6,0             9,0
Interquartile range 6,0             9,0
Stnd. skewness      1,30352         -0,0969571
Stnd. kurtosis      -1,64957        -2,59944
--------------------------------------------------------------
```

Table 5. Comparison of medians for the exam question mark

Median of group 4: 3,0
Median of group 5: 5,0

Mann-Whitney (Wilcoxon) W test to compare medians

  Null hypothesis: median1 = median2
  Alt. hypothesis: median1 NE median2

  Average rank of sample 1: 41,7727
  Average rank of sample 2: 52,54
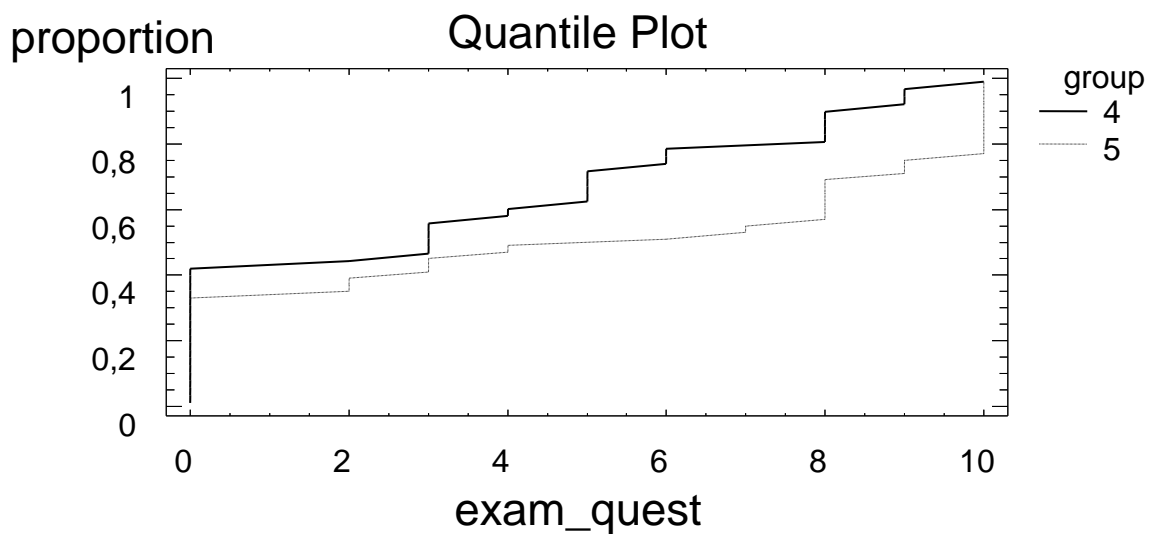
  W = 1352,0   P-value = 0,0490976

Figure 4. Quantile plot of the marks distributions (academic year 2014_15)

Table 6. Kolmogorov-Smirnov Test for the exam question mark

Estimated overall statistic DN = 0,431818
Two-sided large sample K-S statistic = 2,08905
Approximate P value = 0,000323895

3   COMPUTER LABORATORY ACTIVITIES

The seven groups of students had a 90 minute computer laboratory session to practice the discrete distributions concepts and formulas. Each group was divided in two subgroups. They worked in teams of two or three people. Students could solve the questions using any bibliography source, including the teaching material which was in the web page of the university. In the first part of the practice, attribute quality control problems were explained (two in the Spanish speaking groups [8] and four in the Valencian speaking ones [9]). These problems were similar to the quality control exercises commented in Section 2, and required the application of the Binomial and Poisson models. Sample sizes were obtained for required acceptability/rejection criteria and probabilities. Computations were made by hand (binomial distribution) and with the plot of the Poisson cumulative probability curves.

In the next practice activity, students used statistical software to plot the binomial mass and cumulative distribution functions. They computed several probability values on different numbers of defective units, produced in a diskettes factory. They used the mass (Fig. 5) and cumulative probabilities (Fig. 6) plots, and values (Table 7) given by the software. Next question implied to open the inverse cumulative distribution function panel (Table 8).
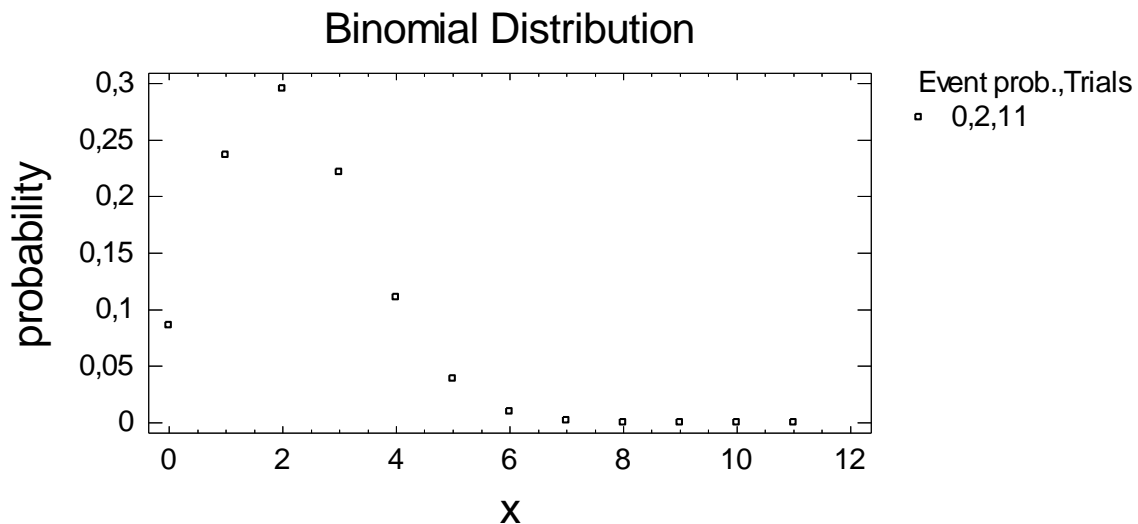

Figure 5. Mass probability function of the number of defective diskettes in a sample of size 11, obtained from a factory production with 20% of defective units.
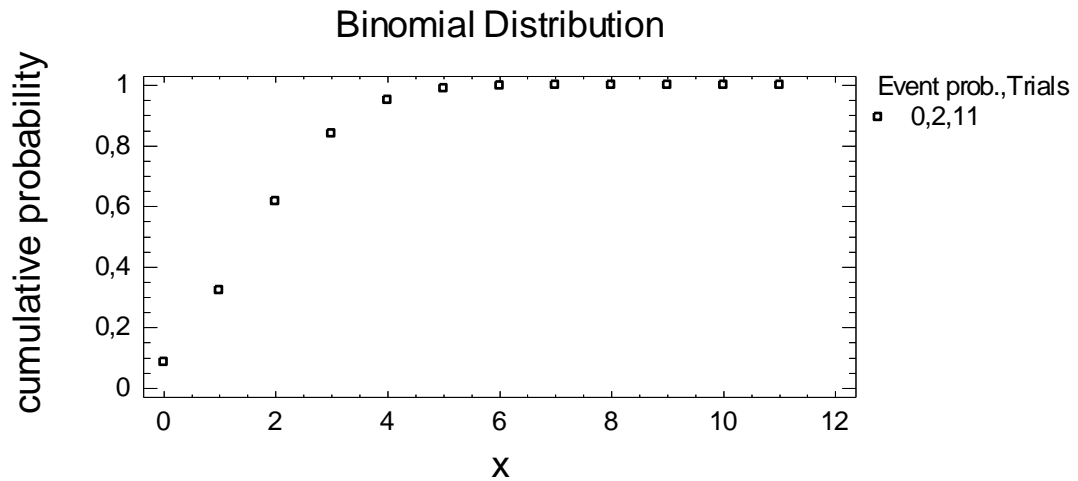
Binomial Distribution

Figure 6. Cumulative probability function of the number of defective diskettes in a sample of size 11, obtained from a factory production with 20% of defective units.

They also applied software to calculate two Poisson distribution probabilities on the number of errors corrected by CRC codes, when sending packages through the net. In this case they only had to open the cumulative distribution function panel (Table 9). In the last question, students had to use the Poisson distributions sum property, to make another probability computation with the same information of the CRC codes example.

Table 7. Probabilities computations for the diskettes quality control problem.

Cumulative Distribution

Distribution: Binomial

Lower Tail Area (<)

Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,0				
2	0,322123				
3	0,617402				
4	0,838861				
5	0,949591				

Probability Mass (=)

Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,0858993				
2	0,295279				
3	0,221459				
4	0,11073				
5	0,0387554				

Upper Tail Area (>)

Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,914101				
2	0,382598				
3	0,161139				
4	0,0504092				
5	0,0116539				

Table 8. Inverse cumulative distribution function panel with a computation posed in relation to the diskettes quality control problem.

Inverse CDF

Distribution: Binomial

CDF	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0,95	5				

Table 9. Probability computations proposed to explain the software options for the Poisson distribution, using the corrected bugs by CRC codes example.

Cumulative Distribution

Distribution: Poisson

Lower Tail Area (<)

Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,0				
5	0,947347				

Probability Mass (=)

Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,135335				
5	0,0360894				

Upper Tail Area (>)

Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,864665				
5	0,0165636				

4 LABORATORY CLASS ASSESSMENT

The evaluation of discrete distributions in the computer laboratory, was at the end of the session. It lasted from 30 to 45 minutes. Each team had to present a consensus written answer to an attribute quality control problem in a database system of a big documentation center. It was proposed by another lecturer. They could consult any type of bibliography. They were asked to compute a sample size using both the binomial and Poisson models. For the first distribution, computations were made with the formulas, and the Poisson one was also applied with the formulas and a cumulative probability plot

The practice evaluation is analyzed to study differences between groups and years. The marks summary statistics are in Table 10. Fig. 7 plots its multiple box-and-whisker diagram, and Fig. 8 represents its variance component plot.

Table 10. Summary statistics of the assessment for the seven groups.

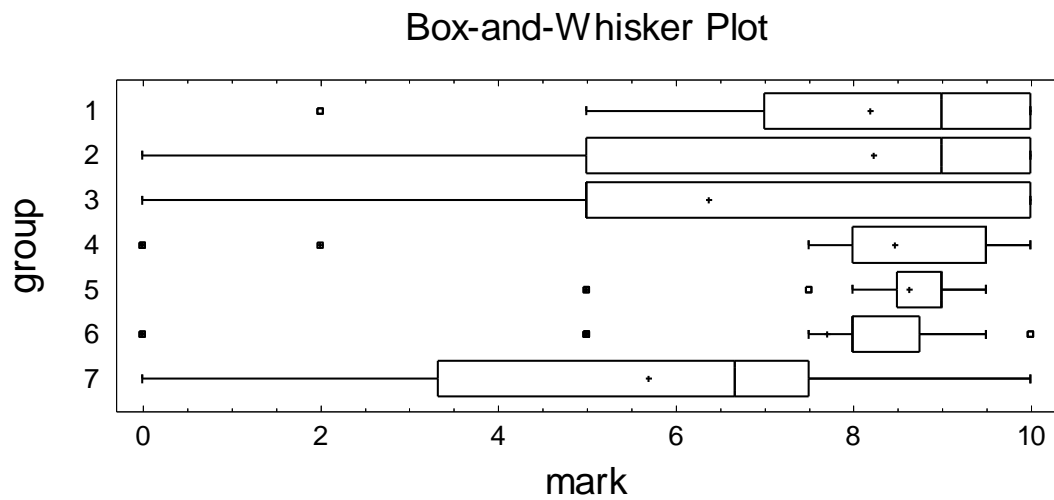| Group	Average	Median	Coefficient of variation	Standardized skewness	Standardized kurtosis
1	8,19512	9,0	26,6252%	-3,66243	1,86486
2	8,23404	9	29,4464%	-3,71586	1,66779
3	6,37931	5	41,3497%	0,540879	-0,395218
4	8,47778	9,5	26,2842%	-8,60927	13,1435
5	8,64	9	10,1952%	-8,83115	15,1536
6	7,70833	8	25,5578%	-7,72555	12,2124
7	5,70312	6,67	50,9065%	-0,895023	0,9251

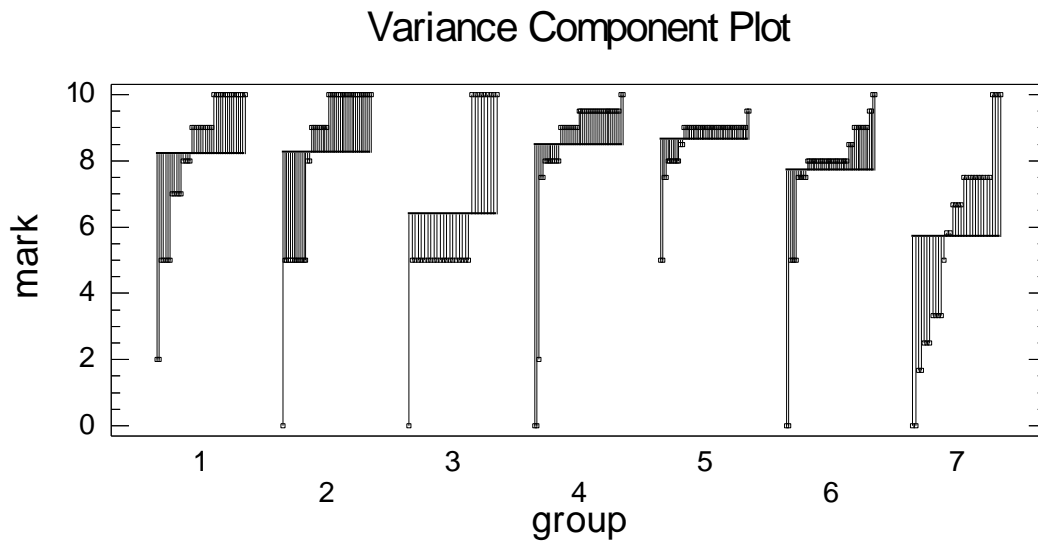**Figure 7.** Multiple box-and-whisker plot of the assessment marks by group.

Figure 8. Variance component plot of the assessment marks by group.

REFERENCES

[1] Romero, R., Zúnica, L. (2013). Métodos estadísticos para ingenieros. Valencia, Spain. Polytechnic University Editorial.

[2] Box, G. E. P., Hunter. W.G., Hunter Stuart J.   (2008). Estadística para investigadores: introducción al diseño de experimentos, análisis de datos y construcción de modelos. Reverté.

[3] Jain, R. (1991). The art of computer systems performance analysis: Techniques for experimental design, measurement, simulation, and modeling. John Wiley & Sons.

[4] Montgomery, D.C., Runger, G.C. (1996) Probabilidad y estadística aplicadas a la ingeniería. McGraw-Hill.

[5] Peña, D. (1986-1995). Estadística, modelos y métodos Vol. 1. Alianza Universidad.

[6] Burr, I.W. (1976). Statistical quality control methods. Marcel Dekker, Inc.

[7] Vázquez, E. (2013). Unidad didáctica 4. Distribuciones de probabilidad-Parte 1. Creative Commons. http://creativecommons.org/licences/by-nc-sa/2.5/es.

[8] Vázquez, E. (2012). Práctica 3. Distribuciones discretas. Creative Commons license. http://creativecommons.org/licences/by-nc-sa/2.5/es.

[9] Alcover, R. (2014) Pràctica 3. Distribucions discretes. Creative Commons license. http://creativecommons.org/licences/by-nc-sa/2.5/es.