
The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2016**Volume 4, Pages 414-423****ICEMST 2016: International Conference on Education in Mathematics, Science & Technology****GPSS TOOL FOR STUDENTS IN STATISTICS EDUCATION**

Ecem İREN

Dept. of Computer Engineering, İzmir University

Serdar KORUKOĞLU

Dept. of Computer Engineering, Ege University

ABSTRACT: Today, using of existing software and technologies becomes important in education and training. Academic institutions need various educational software tools which are appropriate for their own academic standards, curriculums and assessment tools. Those tools make lectures more clearly, enjoyable and concretely. Therefore they help students understand content of lectures by showing real results with graphics or tables. In the study, to take attention importance of mentioned tools, it is aimed to give a recommendation for statistics education. Many of statistics tutorials are taught in theoretical form in universities. While exercising statistical problems related to theoretical topic with students, some special tools should be applied for analyzing statistical formulas and their results. One of these tool is GPSS (General Purpose Simulation System) which is a software high powered general purpose computer simulation environment, designed for simulation professionals. It is a comprehensive modeling tool covering both discrete and continuous computer simulation, with an extremely high level of interactivity and visualizability. At this point, GPSS tool is introduced and its benefits are explained in terms of its operations supplied to users in the paper. Also, Kolmogorov-Smirnov normality test is applied and implemented in the paper.

Key words: Statistics education, GPSS Tool, improvements in statistics education

INTRODUCTION

Nowadays, many computer based applications are developed with various purposes. Some of those applications can be taken into consideration in teaching process. Because, applying some software tools in education and training has gained considerable importance. Especially, integrating theoretical and numerical based courses with software tools will be very useful for students in academic institutions. Owing to fundamental concepts and proofs of numerical based courses prevents students from understanding main ideas; significance of using those tools is increasing steadily (Kumar & Kumaresan, 2008). These tools make information learned in the lecture more memorable by providing several functions and visualization capabilities. Beside this, they help lectures become more enjoyable, concretely and meaningful for the students because practicing theoretical formulas or rules in real sense attracts their attention and interests. By this method, students can analyze problems related to course topics and comment out results by obtaining graphics and tables through the tools. Students can explore funny sides of theoretical and symbolic concepts. In this context, statistics education is associated with GPSS tool and it is recommended in the paper. GPSS is an alternative environment to other packages such as Minitab or Excel and suitable for improving special simulation applications. This paper is organized in the following way: The next section introduces GPSS tool and its main functions by giving examples. Third section describes Kolmogorov-Smirnov normality test which we took into account as a method in the paper and then implementation of it is explained with some scenarios through GPSS tool. Fourth section shows results and findings according to the study and last section summarizes our study and points to benefits of both GPSS tool and general usage of software tools in statistics education.

Gpss (General Purpose Simulation System)

GPSS which is a software high powered general purpose computer simulation environment, designed for simulation professionals. It is a comprehensive modeling tool covering both discrete and continuous computer simulation, with an extremely high level of interactivity and visualizability. Student version of GPSS is available

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: Ecem İREN -icemstoffice@gmail.com

and free. This tool provides lots of functionalities for users who desire to make simulations on various subjects. Some of important functionalities are explained below:

FUNCTION: Functions have list members to determine output result. In other words, this command analyses an argument and according to parameter, it chooses a list member. Functions divide into categories based on the usage scenario. D type function can be given as an example for one of these categories. When function works, it randomly generates a value between 1 and 6 depending on probability argument (Figure 1).

```
MY1 FUNCTION RN1,D6
0.05,1/0.15,2/0.40,3/0.80,4/0.90,5/1.0,6
```

Figure 1. D Type Function

TRANSFER: Generally, this command is useful when it is necessary to jump to the location in the program specified by the operand after the comma represented in Figure 2. In the example, when program jumps to OUT block, current transaction terminates.

```
TRANSFER ,OUT
GO1 SAVEVALUE SAY2+,1
OUT TERMINATE 1
```

Figure 2. Transfer Command

MATRIX: Matrix entity is an array of elements in GPSS tool and it is applied when users want to store information in the program. Declaration of a matrix and a matrix representation is demonstrated in Figure 3. Matrix is created with 1000 rows and 2 columns.

```
MAT MATRIX ,1000,2
MAT
```

Dim 1	Dim 2	
	1	2
1	6	6
2	2	3
3	2	6
4	4	5
5	4	3

Figure 3. Matrix Entity

TABLE: A table entity is a set of integers used to collect data for a histogram. Each integer item specifies a frequency class in a histogram. Declaration of a table and a table representation is seen in Figure 4. Table has first frequency class with upper limit as 1, each frequency class increases by 1 and number of frequency classes is stated as 7.

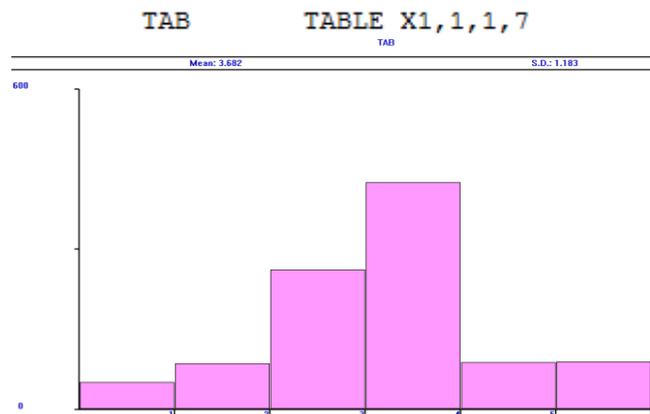


Figure 4. Table Entity

TEST: This command is similar to “if-else” structure. As it is seen from Figure 5, if counter is less than 1000, counter will be incremented by 1, otherwise, program transfers to GO block.

```
FINISH TEST LE X$SAY,1000,GO
SAVEVALUE SAY+,1
```

Figure 5. Table Entity

Also, GPSS provides many advantages for visual education of students which are listed below:

- Students can watch status of all variables of a given statistical problem at different time intervals while simulation is proceeding (Figure 6).

Savevalue	Value	Retry Ch...
 1	1.000	0
 2	58.310	0

Savevalue	Value	Retry Ch...
 1	2.000	0
 2	-7.571	0

Figure 6. Savevalue Windows at Different Time Intervals

- Students can monitor distribution of all values derived from any type of distribution to frequency sets on table entity while simulation is going on (Figure 7 and Figure 8).

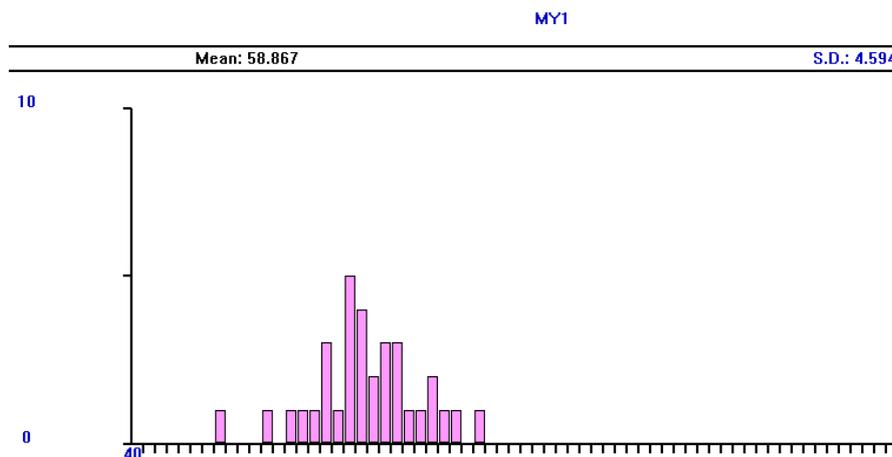


Figure 7. Table -1

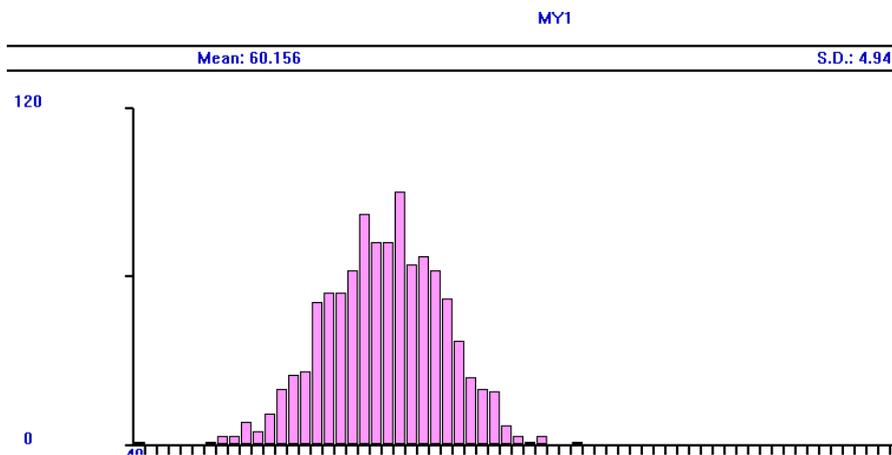


Figure 8. Table -2

- Students can keep track of output values stored in matrix entity during the simulation.
- It provides basic build-in procedures required in statistical problems such as generating Normal, Binomial, Poisson, Exponential, Discrete Uniform distributions with random values. Likewise, it involves some string operations.

METHODS

Some operations which can be computed through GPSS tool and listed above are practiced through a scenario in this part. The scenario is selected as Kolmogorov-Smirnov normality test method which can be realized either theoretically or practically in GPSS environment by the students in the lecture. Kolmogorov-Smirnov normality test is an alternative for goodness of fit test of Chi-Square. Some information about this test technique and an example concerning it is given in detail in the following.

Kolmogorov-Smirnov Normality Test

The *Kolmogorov-Smirnov* test statistic is a well known test statistic used to solve goodness-of fit problems was proposed by Kolmogorov in 1933. Kolmogorov suggested a goodness of fit test for one sample. In 1939, Smirnov who is a Russian mathematician developed a goodness of fit test for two independent samples. Due to similarity between Kolmogorov and Smirnov tests, it is known as Kolmogorov-Smirnov goodness of fit test in practice (Bircan, Karagöz, & Kasapoğlu, 2003). This test technique is nonparametric of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K-S test), or to compare two samples (two-sample K-S test) [2]. An example of deciding whether a sample belongs to normal distribution is explained and then implementation of K-S test is shown.

Example: Let's say that S(x) defines empirical distribution function of x value which is number of values less than x. S(x) is calculated in the following:

Suppose that there is a study with ten observations those are listed below:

95.00 72.97 94.34 102.64 97.91 106.18 102.67 108.89 112.04 99.32

Firstly, observation values are ordered from least to greatest as shown below. Then empirical function result of each value is found in Table 1.

72.97 94.34 95.00 97.91 99.32 102.64 102.67 106.188 108.89 112.04

Table 1. Empirical Function Result of Each Value

X	72.97	94.34	95.00	97.91	99.32	102.64	102.67	106.18	108.89	112.04
S(X)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

Assume that those observation values are taken from a normal distributed cluster whose mean is 100 and variance is 10. Therefore, two hypotheses are claimed in the following:

H₀= The given data follow specified normal distribution with mean and standard deviation 100 and 10, respectively.

H₁= The given data do not follow specified normal distribution with mean and standard deviation 100 and 10, respectively.

Then, for each data value, cumulative distribution function is calculated with formula that is shown in Figure 9. Data values are represented with X. For X=95, cumulative distribution function result is found as shown in Figure 10.

$$P\left[\frac{x-\mu}{\sigma}\right]$$

Figure 9. Cumulative Distribution Function

$$F(95) = P\left[\frac{x-\mu}{\sigma} \leq \frac{95-100}{10}\right] = P(Z \leq -0.4) = 0.341$$

Figure 10. Cumulative Distribution Function for X=95

Similar calculation is done for all data values in the sample and final result table is represented in Table 2.

Table 2. Cumulative Distribution Function Results of All Values

X	F(X)
72.97	P(X<=72.97) = P(Z<= -2.54) = 0.005
94.34	P(X<=94.34) = P(Z<= -0.47) = 0.318
95.00	P(X<=95.00) = P(Z<= -0.40) = 0.341
97.91	P(X<=97.91) = P(Z<= -0.12) = 0.450
99.32	P(X<=99.32) = P(Z<= 0.01) = 0.504
102.64	P(X<=102.64) = P(Z<= 0.33) = 0.630
102.67	P(X<=102.67) = P(Z<= 0.33) = 0.631

106.18	$P(X \leq 106.18) = P(Z \leq 0.67) = 0.751$
108.89	$P(X \leq 108.89) = P(Z \leq 0.94) = 0.826$
112.04	$P(X \leq 112.04) = P(Z \leq 1.24) = 0.893$

If H_0 hypothesis is true, for all values of X, F(X) and S(X) results should be similar. Otherwise, at least for some values of X, there will be a big difference between F(X) and S(X). Absolute value (D) of the greatest difference between F(X) and S(X) is identified as the test statistics and is computed as shown in Figure 8. Also, all differences related to sample data values are illustrated in Table 3.

$$D = \max [F(X) - S(X)]$$

Figure 8. Calculation of Test Statistics

Table 3. Cumulative Distribution Function Results of All Values

X	F(X)	S(X)	F(X)-S(X)	[F(X)-S(X)]
72.97	0.005	0.10	-0.094	0.094
94.34	0.318	0.20	0.118	0.118
95.00	0.341	0.30	0.041	0.041
97.91	0.450	0.40	0.050	0.050
99.32	0.504	0.50	0.004	0.004
102.64	0.630	0.60	0.030	0.030
102.67	0.631	0.70	-0.068	0.068
106.18	0.751	0.80	-0.048	0.048
108.89	0.826	0.90	-0.073	0.073
112.04	0.893	1.00	-0.106	0.106

When last column of Table 3 is examined, it is seen that the largest absolute value is 0.118. Then, that value is compared with the critical value exists in Kolmogorov Smirnov Test Statistics Table. According to the table, for $\alpha=0.05$ (significance level) and $n=10$ (sample size), the critical value is equal to **0.409**. Due to 0.118 is less than 0.409, H_0 hypothesis can be accepted in this way. Therefore, it can be concluded that there is enough support to say that sample is selected from normal distributed cluster whose mean and standard deviation 100 and 10, respectively (Satici, 2014).

Kolmogorov-Smirnov Normality Test Implementation with GPSS Tool

Implementation of K-S test in GPSS tool is based on four methods. Firstly, pure normally distributed sample is tested, then mixture of a normal distribution and a triangular distribution is analyzed, afterwards Weibull distributed sample is considered and finally exponentially distributed sample is examined.

Generally, in each method, following steps are taken into account:

- Generate values randomly for creating a sample
- Find mean and standard deviation of the sample
- Sort the sample in ascending order
- Obtain the empirical and cumulative distribution function of each value
- Obtain all differences between cumulative and empirical functions in the sample
- Select maximum difference and compare with critical value found in K-S table

While writing the program, various procedures are applied to obtain some results. For example; in order to decide if sample distribution is normal, DECIDENORMAL procedure is created (Figure 11).

```

PROCEDURE DECIDENORMAL (KVALUE)

BEGIN

    TEMPORARY INDEX1,MAXVALUE;
    MAXVALUE=DIFFERENCE [1,1];
    INDEX1=2;
    WHILE (INDEX1<=1000) DO
    BEGIN
        IF (MAXVALUE<DIFFERENCE [INDEX1,1]) THEN BEGIN
            MAXVALUE=DIFFERENCE [INDEX1,1];
        END;
        INDEX1=INDEX1+1;
    END;

    IF (MAXVALUE<=KVALUE) THEN BEGIN
        RETURN 1;
    END;
    ELSE RETURN 0;

END;

```

Figure 11. DECIDENORMAL Procedure

Moreover, in order to find cumulative distribution function of each value, ZRESULTVALUE procedure is benefited (Figure 12).

```

PROCEDURE ZRESULTVALUE (A)
BEGIN

    TEMPORARY RSQRT2PI,NO1,NO2,NO3,NO4,NO5,K,CND;
    NO1=0.31938153;
    NO2=-0.356563782;
    NO3=1.781477937;
    NO4=-1.821255978;
    NO5=1.330274429;
    RSQRT2PI=0.398942280401432;
    K= 1.0 / (1.0 + 0.2316419 # ABS (A));
    CND = RSQRT2PI # EXP (-0.5 # A # A) # (K # (NO1 + K # (NO2 + (K # (NO3 + K # (NO4 + K # NO5))))));
    IF (A>0) THEN BEGIN
        CND= 1.0 - CND;
    END;

    RETURN CND;

END;

```

Figure 12. ZRESULTVALUE Procedure

RESULTS and FINDINGS

In this part, all experiments using implementation are evaluated with sample size 1000. Aim of making test with large sample size is to observe power of K-S test because when sample size increases, reliability of test results becomes stronger. Also, it is considered that it will be difficult to represent tables of small sample size. Observations are divided into four categories to measure effectiveness of K-S normality test as mentioned previous part.

Step1: Testing values generated from normal distribution in the implementation

When testing is done with mean 60 (μ) and standard deviation 25 (σ), sample distribution can be seen with GPSS tables in different time intervals through implemented simulation program (Figure 13). K-S simulation gives valid result for this sample by providing a message to the user shown in Figure 14. In addition, program has been designed to produce some result matrixes which store Z value and cumulative distribution function result depending on that Z value of each item in the sample (Table 4 and Table 5). It can be observed from Table 6, the greatest difference (test statistics) is found as 0.018 and it is less than critical value (0.043) so it can be said that sample is normally distributed.

MY1

Mean: 60.156

S.D.: 4.940

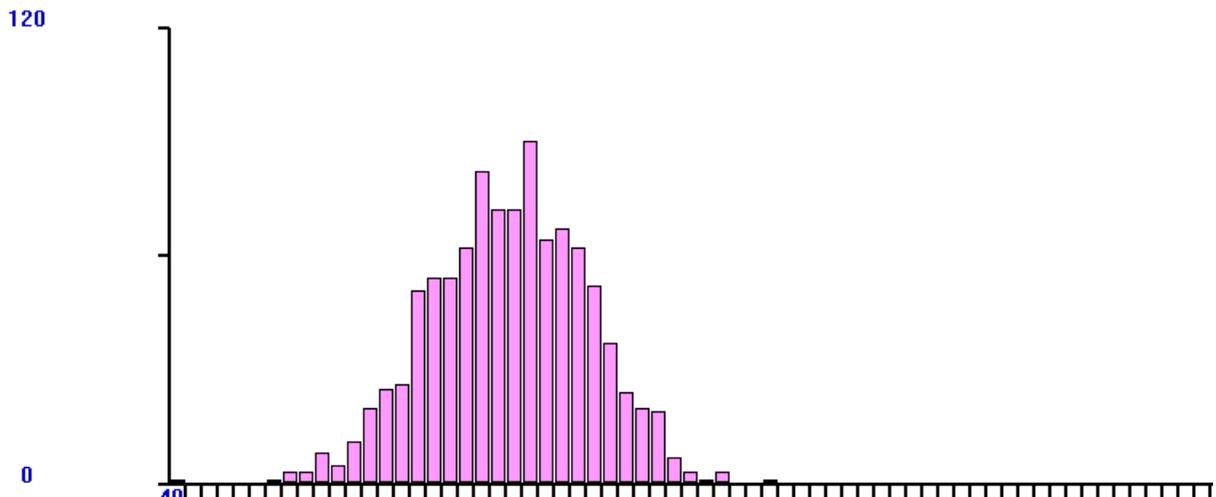


Figure 13. Experimental Table

1
"NORMAL DISTRIBUTION"

Figure 14. Simulation Message

Table 4. Matrix of Z Values

ZMATRIX

Dim 1	Dim 2
	1
1	-4.116
2	-3.023
3	-2.817
4	-2.768
5	-2.679
~	~

Table 5. Matrix of Cumulative Distribution Function Results

ZRESULTMATRIX

Dim 1	Dim 2
	1
1	.000
2	.001
3	.002
4	.002

Table 6. Test Statistics for Normally Distributed Sample

- MATRIX WINDOW

DIFFERENCE

Dim 1	Dim 2
	1
555	.018
556	.018
557	.018
558	.017
559	.016

Step 2: Testing numbers generated from mixture of a normal distribution and a triangular distribution in the implementation

In this phase, simulation generates a sample from mixture of a normal and a triangular distribution. While normal distribution has mean 30 (π) and standard deviation 7 (σ), triangular distribution has minimum value 30, maximum value 100 and mode value 90. Result table is expected to appear like in Figure 15 and it is clear that normality is disrupted with triangular distribution. At the end of the simulation, experimental table resembles to the expected table (Figure 16). K-S simulation gives reject result for this sample by providing the message to the user shown in Figure 17. It can be noticed from Table 7, the greatest difference is found as 0.157 and it is greater than critical value (0.043) so it can be said that sample is not normally distributed.

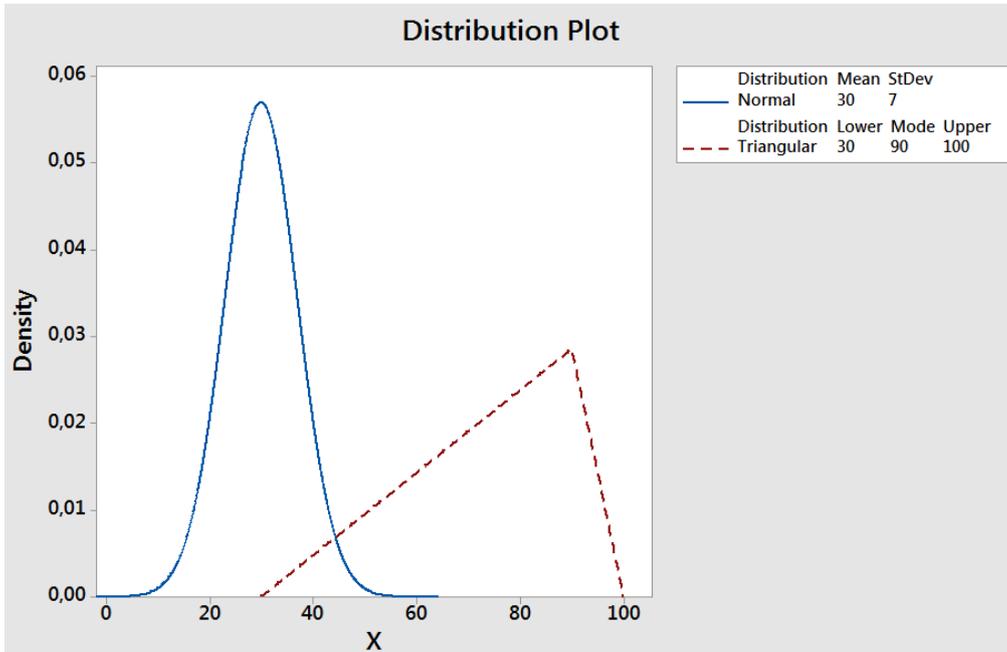


Figure 15. Expected Result Table

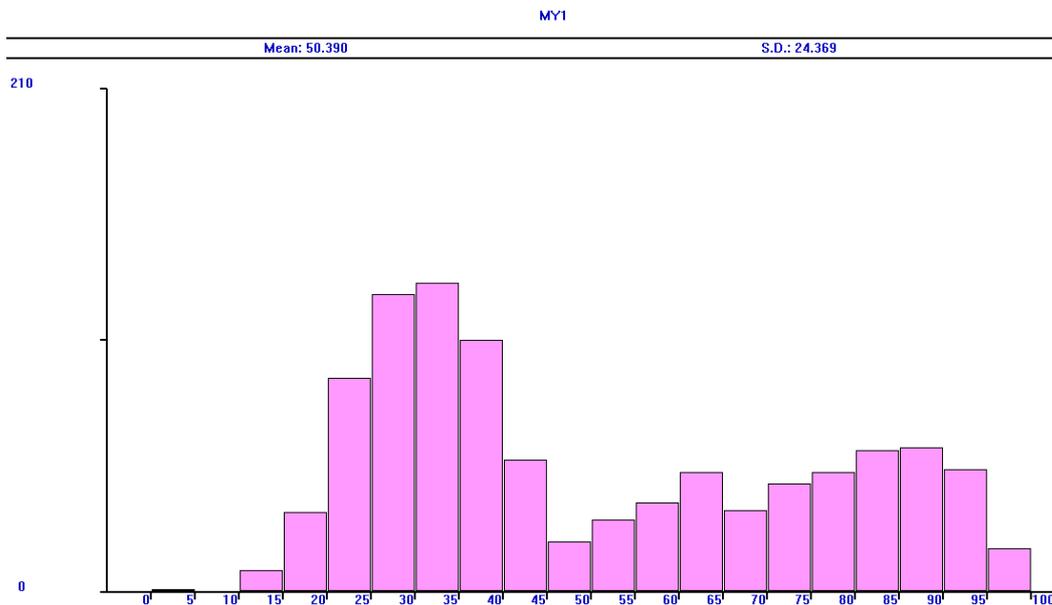


Figure 16. Experimental Table

Dim 1	Dim 2
1	1
	"NOT NORMAL DISTRIBUTION"

Figure 17. Simulation Message

Table 7. Test Statistics for Mixture Distribution

DIFFERENCE

Dim 1	Dim 2
	1
490	.157
491	.153
492	.154
493	.153

Step 3: Testing numbers generated from Weibull distribution in the implementation

In this step, simulation generates a sample from Weibull distribution with shape, locate and scale parameters as 0, 25 and 2, respectively. At the end of the simulation, sample distribution is illustrated with GPSS table (Figure 19). K-S test gives reject result for the sample by giving the message to the user shown in Figure 18. It can be noticed from Table 8, the greatest difference is found as 0.062 and it is greater than critical value (0.043) so it is proven that sample is not normally distributed.

Table 8. Test Statistics for Weibull Distribution

DIFFERENCE

Dim 1	Dim 2
	1
453	.062
454	.060
455	.060
456	.056

RESULTMAT

Dim 2
1
"NOT NORMAL DISTRIBUTION"

Figure 18. Simulation Message

MY1

Mean: 22.367	S.D.: 11.889
--------------	--------------

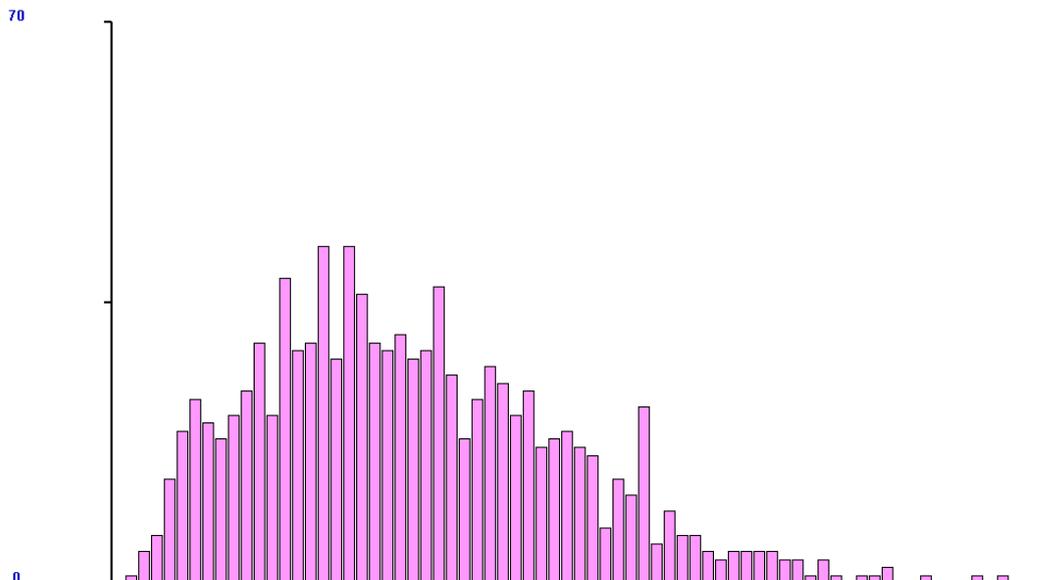


Figure 19. Experimental Table

CONCLUSION

In this study, at first, GPSS Tool and its primary operations are presented. After that, Kolmogorov-Smirnov normality test is introduced and working principle of it is consolidated with an example. Subsequently,

implementation of K-S test on GPSS is clarified and then some test cases are evaluated with specific parameters to validate and verify effectiveness of developed program. In each case, program gives valid result and demonstration of the program is strengthened with tables, matrixes and output messages. Processing statistics courses in this manner provides many advantages to both students and teachers. These advantages are listed below:

- Students can make practices on data analysis by handling data operations and documentation of results. In addition, software tools enable them to construct and compare statistical methods and give a chance for simulating special designed applications and displaying statistical results on their own tables or graphs (Biehler, 1997).
- Teachers can learn new things for the curriculum content of the lectures and this situation brings a new approximation to operation of lectures in classes. Consequently, teachers should be adapted and encouraged to use software tools for their students (Biehler, 1997).

From the reasons described above, usage of such tools should be extended and maybe some informative activities can be arranged for the teachers on this subject to increase attention.

RECOMMENDATIONS

GPSS tool is highly recommended for statistics education in academic foundations on account of the reasons referred in previous part and the opportunities it includes. When GPSS is compared to other statistical packets,

- It is free, easy to learn and use.
- Can be preferred effectively and easily to a wide range of problems designed with specific objectives.
- It is suitable for queue systems related with scenarios about service time of a system to its customers (Ündeğer, 2008).
- Most important one is that students could improve their programming and algorithm knowledge with the aid of GPSS by carrying out their user-story simulation projects.

REFERENCES

- Bircan, H., Karagöz, Y., Kasapoğlu, Y. (2003). Ki Kare ve Kolmogorov Smirnov Uygunluk Testlerinin Simülasyon ile Elde Edilen Veriler Üzerinde Karşılaştırılması. *C.Ü. İktisadi ve İdari Bilimler Dergisi*, 4(1), 69-80
- Kumar, A., Kumaresan, S. (2008). Use of Mathematical Software for Teaching and Learning Mathematics. *ICME 11 Proceedings*, 373-388
- Romeu, J. L. (2003). Kolmogorov-Smirnov: A Goodness of Fit Test for Small Samples. *Selected Topics in Assurance Related Technologies*, 10(6), https://src.alionscience.com/pdf/K_STest.pdf
- Satıcı, B. (2014). Kolmogorov-Smirnov ve Lilliefors Normallik Testleri. Diyarbakır: Dicle Üniversitesi Tıp Fakültesi
- Ündeğer, Ç. (2008). General Purpose Simulation System (GPSS) (Part 7), Ankara: Bilkent Üniversitesi Bilgisayar Mühendisliği Bölümü