

**ANALYSIS OF THE FACTORS THAT INFLUENCE  
THE STUDENTS' PERFORMANCE OF UNIFORM HIGH  
SCHOOL AND TECHNICAL VOCATIONAL SCHOOL  
IN THE SUBJECT OF STATISTICS  
(STATISTICAL DATA PROCESSING METHODS: REGRESSION  
ANALYSIS AND STRUCTURAL EQUATION MODELING)**

By

*Evangellos Papakonstantinou\**, *Vassilios Gialamas\*\**, *Dimitris Karageorgos\*\*\**

\*Dr of Statistics and Educational Research (National Kapodistrian University of Athens)

\*\*Assistant professor of Statistics at the National Kapodistrian University of Athens

\*\*\*Assistant professor of Statistics and Educational Research  
at the National Kapodistrian University of Athens

**Abstract**

In the present research, an attempt was made to find the factors that determine the level of achievement of the students in the subject of Statistics. This subject is taught in the second (upper) level of Secondary Education (S.E.), viz. in the third (last) grade of Uniform High School (U.H.S.) and in the first grade of the Technical Vocational School -T.V.S. (in the field of the Economy and Management). Furthermore, the achievement of the students in Statistics was examined according to subject matter. Based on the findings of this research, some proposals have been formulated for the better instruction of the subject.

In Greece, S.E. is comprised of the first level which corresponds to the Middle or Junior High School and the second level which includes the U.H.S. and the T.V.S. The length of study for each level is three years.

**1. Introduction**

The present research was based on sample data obtained in April 2001 in schools in both urban centers and in rural regions of Greece, given the fact that there has never been any similar study neither in the Uniform High School (U.H.S.) nor in the Technical Vocational School (T.V.S.). The lack of such studies is expected, to some degree, since the subject of Statistics is an area of instruction that has been taught during the last two to three decades at the level of S.E. (Secondary Education).

The present research has made an effort to avoid a quest of a fragmentary

nature of possible ways and means that could potentially improve the instruction of the subject. On the contrary, it went through a holistic investigation of the issue, looking for the factors that influence the performance of the students of the U.H.S. and T.V.S. in Statistics. Moreover, it examined the performance of the students in the various unit subjects of Statistics and, based on the findings of the research, it proceeds in the formulation of recommendations with the goal of improvement in student scores in the U.H.S. and T.V.S. in Statistics. For the technical processing of the research data the method of Regression Analysis (R. A.), in the version of Standardized Multiple Regression Model was used. In addition, the newer technique of Structural Equation Modeling (S. E. M.), was used, which validated and extended the findings of R. A.

## **2. Reasons for Conducting the Present Research**

There has not been to date a widely conducted study that examines the influence of many and of diverse character factors in the student performance of the U.H.S. and T.V.I. in Statistics. The research to date involved the study of one or two factors relative to the influence that those exert on the performance of the students in the subject of Statistics. As an example, some researchers have analyzed the influence that has on the students' performance on Statistics, the study of this object with the use of P/C (Graham A., 1999, Bratton, G., 1999, Barret G., 1999). Other researchers have studied the influence of the efficient knowledge of Mathematics, that have the students, on the better understanding of Statistics by the them (Burril G. et al., 1992). Let us keep in mind, that Statistics forms one of the scientific fields of Mathematics (N.C.T.M. 2003). Other researchers have studied the influence of the students' opinion, about the practical usefulness of the Statistics, on their performance on this subject (Talsma G., 1999, Takis S., 1999, Du- Feu C.,1999). Also, the influence of the social factors on the performance student on Statistics have analyzed by many researchers (Kontogiannopoulos G. et al. 2000, Hodgson T. et al. 1998).

An additional reason for conducting this research was the limited experience that exists at the level of U.H.S. and T.V.S. relative to the instruction of Statistics as compared to the instruction of other subject matters.

One further reasons for conducting the present research was the desire to come up with documented aid to facilitate the educational endeavor of the instructions in Statistics. Let us keep in mind that most instructional techniques and theories are based, to a great extent, on the so called classical subjects such as History, Literature, Mathematics e. t. c.

Finally one additional reason of conducting an in depth examination of the present issue was our confirmation (being educators who have as a profession the responsible instruction of this subject in Secondary Education -S.E.) of the existence of difficulties that face the instructors and the instructed relative to the teaching of Statistics.

### **3. The Framework of the Study**

Within the framework of conducting this research and in both cases of the U.H.S. and T.V.S. only the cognitive domain was studied, the other two (effective domain and psychomotor one) being outside the scope of the present research were not examined. The questions of the questionnaire belong to four categories: i. Knowledge, ii. Comprehension, iii. Applications, and iv. Advanced level questions (Analysis, Synthesis and Evaluations).

The questions cover the subject units of Statistics that are taught in the last grade of U.H.S. namely 1. Introduction and Data Presentation, 2. Measures of Central Tendency, 3. Measures of Variation, 4. Regression, and 5. Probability Theory. The questionnaire presented to the students included 50 questions, 10 in each unit. Similarly, questions were given to the instructors of the students. The answers to all these questionnaires constituted the raw material of the research. In the case of the T.V.S., where the students are taught the above subject units except that of Probability Theory, the questionnaire included 40 questions only.

### **4. Sampling Technique and Tests used by the Research**

#### **I. Sampling Technique**

In order to materialize the research, a sample size of 1195 students was obtained with the method of stratified random sampling from the U.H.S. of the urban complex of Athens-Piraeus and from the geographical region of Peloponnese<sup>1</sup>.

The social strata of the areas that were sampled were taken into account in order to make further comparisons between urban centers and rural areas. Similarly a smaller sample of 427 students of T.V.S. was obtained from the same administrative regions, since the number of students matriculated in these schools is significantly smaller than those of the U.H.S.

The sample size, given the personal nature of the research and the conditions in Greece is considered acceptable.

## II. The Tests used for the Realization of the Research

The tests that were used in the realization of the research were checked in order to verify if they are suitable and reliable to be given to the students. According to the internationally acceptable standards, the following criteria and indices were used (Kubiwzyn and Borich, 2000, Kassotakis, 1999).

**TABLE 1**  
The Results of the Indices and Criteria as the Reliability of the Tests

	<b>Indices &amp; Criteria (I. &amp; C)</b>	<b>Results of the (I. &amp; C)</b>	<b>Summary description of the results. (In accordance with the provisions of international standards set by similar studies)</b>
<b>1</b>	<b>Discrimination index (D)</b>	$D < 3$	The questions were appropriate, in order to distinguish above average students from average and below average ones. All the questions have $D > 3$ .
<b>2</b>	<b>Difficulty index</b>	$0,2 < D.I. < 0,8$	The questions had a Difficulty Index (D.I) between 0,2 and 0,8,
<b>3</b>	<b>Reliability index (D.I.)</b>	$r^2 = 0,91$	The employment of S.H.M. <sup>2</sup> and the subsequent calculation of the $r^2$ showed a high credibility degree <sup>3</sup> .
<b>4</b>	<b>Criterion related validity</b>	$r^2 = 0,90$	We have a high correlation (0,90) among students who had received high (low) scores at the pilot test and at the final one respectively.
<b>5</b>	<b>Content validity</b>	See tables 2 & 3	See the following table (2 & 3) of specifications <sup>4</sup> .
<b>6</b>	<b>Marks adjustment</b>	There's not the element of luck	All relative audits did not reveal any significant differences with regards to the element of luck.

**TABLE 2**  
The Table of Specifications - The Case of the High School

Thematic Sections	Questions				
	Knowledge	Comprehension	Applications	High level questions	Total
Introduction & Data Presentation	2	3	3	2	10
Measures of Central Tendency	2	3	3	2	10
Measures of Variation	2	3	3	2	10
Regression	2	3	3	2	10
Probabilities	2	3	3	2	10
<b>Total</b>	10	15	15	10	50

**TABLE 3**  
The Table of Specifications - The Case of the Technical School

Thematic Sections	Questions				
	Knowledge	Comprehension	Applications	High level questions	Total
Introduction & Data Presentation	2	3	3	2	10
Measures of Central Tendency	2	3	3	2	10
Measures of Variation	2	3	3	2	10
Regression	2	3	3	2	10
<b>Total</b>	8	12	12	8	40

Tables 2 and 3 show the subject units of Statistics that are taught to the students of U.H.S. and T.V.S. respectively. In both cases, every subject unit is represented by two questions of Knowledge, three comprehension, three application and two High Level questions (analysis, synthesis and evaluation).

## 5. Techniques used in the processing of the Data of the Research

a. Regression Analysis (R.A.). This technique was used in version of Standardized Multiple Regression Model (S. M. R. M) because: S. M. R. M is devoid of the influence of metric system. Therefore, we can make reliable comparisons with other similar research in which different measurement systems were used, ii. In models with many explanatory variables  $X_i$ , like here, the existence of many intermediate calculations and the round off the decimal numbers (especially after the inversion of matrix  $XX$  that contains the values of variables  $X_i$ ) frequently lead to results that diverge from the actual<sup>5</sup>. We should remember that the use of S.M.R.M. avoids the possibility of such numerical complications. When we have a S.M.R. Model- that the  $b_i$  (regression coefficient) shows the alteration that will occur at the  $S$  (standard deviation) of the dependent  $Y$ , from the alteration of a point at the  $S_i$  of an explanatory variable  $X_j$ ,

b. Structural Equation Modeling (S.E.M.). This method of statistical data processing is a new technique which is not yet used widely by researchers compared with other techniques of Statistical processing of data (like, for example, the R.A). A brief presentation of the advantages and possibilities of S.E.M. will follow later in the presentation of the results obtained.

## 6. The Case of Uniform High School (U.H.S.)

### 6.1 The Results obtained by the Use of Standardized Multiple Regression Model (S. M. R. M.)

Following the examination of 25 candidate explanatory variables by using Regression Analysis, a regression function of the form  $Y_1 = g(X_1, b)$  was obtained with only seven explanatory variables (where  $X_1$  is the variables vector and  $b$  is the parameters vector). The selection of the candidate variables examined was made based on some few previous research reports (they are listed in paragraph 2) that examined in a fragmented way the influence of one or two factors in the scores of students in Statistics. In addition, we included some other variables that were deemed influential in the score of student in Statistics, such as the Student's overall performance from last year, the desire of students to continue their studies, the existence of a home library, the gender of

the students, the tuition in Statistics etc. In Greece, the tuition is an important factor in educational achievement, especially in Science and Mathematics (Verdis, 2202).

With respect to the necessary requirements for the application of R.A., the relevant examinations showed that these requirements are present. More specifically, there were not the phenomenon of the Heteroscedacity, the errors  $e_i (\hat{y}_i - y_i)$  follow the normal distribution, no significant degree of Multicollinearity was observed among the explanatory variables. The number of sample data was much bigger (1195) than the minimum required according to international research standards (ten times the number of the explanatory variables, Draper 1977). The resulting regression model follows below

$$\hat{Y} = 0,229X_1 + 0,340X_2 + 0,127X_3 + 0,242X_4 + 0,114X_5 + 0,098X_6 - 0,084X_7$$

**TABLE 4**

The Regression Analysis Results

A Summary Presentation of the Regression Analysis Results							
Case: U.H.S.							
Statistical data processing Method: The method of the Least Squares. Model: SMRM							
Dependent Variable (Y): Performance in the Subject of Statistics							
Explanatory Variables		Unstandard Coefficient Beta	Standard Coefficient Beta	t Statistic <sup>a</sup>	Significance Value	R <sup>2</sup> Change	Sign. F Change
Constant (b <sub>0</sub> )		-4,529		-9,835			
X <sub>1</sub>	Student's performance Maths from the last year	0,343	0,229	6,901	0,000	0,432	0,000
X <sub>2</sub>	Hours studying Statistics	1,197	0,340	16,544	0,000	0,147	0,000

*continues*

X <sub>3</sub>	Supplementary tuition in Statistics	0,636	0,127	6,685	0,000	0,021	0,000
X <sub>4</sub>	Student's overall performance from last year	0,384	0,242	7,506	0,000	0,013	0,000
X <sub>5</sub>	Desire of the student to continue studies	0,491	0,114	6,317	0,000	0,014	0,000
X <sub>6</sub>	Parents' educational level	0,179	0,098	5,564	0,000	0,012	0,000
X <sub>7</sub>	Student's place of residence <sup>7</sup>	-1,356	-0,084	-4,732	0,000	-0,006	0,000
- $R^2 = 0,646$ and							
- $R^2_{adj} = 0,644$ .							
- $F_{7, 1195} = 309,761$ , $p < 0,0005$							

The above regression model<sup>8</sup> interprets, satisfactorily, the variance of the dependent variable "Students' performance at the Statistics test" (Y)<sup>9</sup>, since the  $R^2_{adj}$  (coefficient of determination)<sup>10</sup> is equal with 0,64 ( $0 < R^2_{adj} < 1$ )<sup>11</sup>. In this model, the greatest effect on the students' performance at the Statistics test is exerted by their previous year's grade in Mathematics ( $b_i = 0,229$ ) and the number of hours allocated each week for the studying of Statistics ( $b_i = 0,340$ ). The table 3 shows the effects that are exerted by the rest explanatory variables on the variable Y.

Also, the above table shows that, out of the (7) seven explanatory variables which interpret the dependent variable (Y) "Students' performance at the Statistics test", four have social denotations, "Supplementary tuition in Statistics", "Parents' educational level", "Desire of the student to continue studies" and the variable "Student's place of residence".

#### Student performance according to subject unit

The overall performance of the students in the Uniform High School (U.H.S.) in the test taken in Statistics was approximately 12,14 points on the 20 point grading scale used in the S.E. in Greece. The breakdown according to subject unit is as follows: a. First unit (Introduction, statistical data collection



and presentation) 12,2. Second unit (Measures of Central Tendency) 13,6 points. 3. Third unit (Dispersion Measures) 12,2 points. 4. Fourth unit (Regression) 11,2 points and fifth unit (Probabilities) 11,5 points.

## 6.2 The Case of S.E.M.

The technique of Structural Equation Modeling (S. E. M.) is a combination of Explanatory Factor Analysis and Regression Analysis (J. B. Ulman, 1996). Compared to R. A., the S. E. M. comprises a much more recent technique of statistical processing of data. S. E. M., if we disregard the disadvantage of having some complexity (e.g., compared to R. A.) as a technique and the special attention it requires in the interpretation of the results, has certain significant advantages. More specifically, S. E. M. allows us to investigate a variable Y with the help of additional variables of our own invention called latent variables which are related to the other observed variables (which are called indicator variables). Furthermore a variable which in one context is independent in another one (context) may be dependent. This implies that in this formulation we may have a series of independent variables (exogenous) and another series of dependent variables (endogenous) (Hair et al. 1998, Arbuckle et al. 1996). All these basic properties of S. E. M. and others that are not mentioned here enable us to perform a more complete and in depth investigation of the issue under study. Many researchers recommend the utilization of the S. E. M. technique for the processing of statistical data in the field of educational research (Alimissis, 1995, Hair, 1998).

In the case of U.H.S. in addition to the original dependent variable "Performance in Statistics" that was examined using R. A, we considered as internal (endogenous) variables those of "Students' Performance in Mathematics from the Last Year" and "Hours studying per week". This was because the last two variables, as R. A. showed, were the most significant ones in the interpretation of the behavior of the factor under examination and we sought further study of the phenomenon based on some reasonable assumptions mentioned later. Moreover the explanatory variable "Desire of the student to continue their studies" was included in the group of internal (endogenous/variables because it was deemed that this variable may be influenced by the students' interests and the time spent in the study of the field of Mathematics in general.

In order to achieve an even more complete investigation of the issue under examination we constructed a latent (construct) variable "Overall Study of Mathematics". The latent (construct) variable the "Overall Study of Mathe-

matics" has the variable "Hours studying Statistics per week" as indicator one. The construct (latent) variable "Overall Study of Mathematics" expresses the student's overall study at the wider field of Mathematics (Algebra, Statistics, Geometry etc.).

In this way we achieve an in depth investigation of the relationship between variables "Student's performance in Mathematics from the last year" and "Hours Studying Statistics per Week". We accept that the latent variable "Overall Study of Mathematics" that involves the more general studies of the students in the wider field of Mathematics interprets to the degree of 81% ( $r=0,9$ ) the "Hours Studying Statistics per Week". Therefore, we conclude that a student who studied Mathematics in general for many hours last year, the same student will be studying Statistics for many hours this year. The table below lists the endogenous and exogenous variables that together comprise the S.E.M. model.

**TABLE 5**  
The variables of the SEM model

Endogenous variables	Exogenous variables
Performance in Statistics	Student's place of residence
Student's performance in Mathematics	Parents' educational level
Desire of the student to continue studies	Tutoring in Statistics
Hours studying Statistics	Student's overall performance
	Error "Statistics Test"
	Error - "Statistics' Usefulness"
	Error -"Performance in Mathematics"
	Overall Study of Mathematics
	Error -" Overall Study of Mathematics"

As we notice on the above table (table 5) the total number of the model's variables is 13 (thirteen). There are eight observed variables (four endogenous variables and four exogenous ones), which are the first four in order of appearance at the two columns of table 3. Moreover, there are five no unobserved exogenous variables, which appear last at the second column of above table, four out of which express the respective errors and the remaining one is the latent variable "Overall Study of Mathematics".

**TABLE 6**  
Standardized Regression Weights

Dependent Variables	←	Independent Variables	b
Student's performance in Mathematics from the last year	←	Student's overall performance from the last year	0,776
Student's performance in Mathematics from the last year	←	Overall Study of Mathematics	0,142
Desire of the student to continue studies	←	Student's performance in Mathematics from the last year	0,142
Desire of the student to continue studies	←	Overall Study of Mathematics	0,194
Desire of the student to continue studies	←	Student's place of residence	-0,075
Performance in Statistics	←	Student's place of residence	-0,082
Performance in Statistics	←	Parents' educational level	0,089
Performance in Statistics	←	Desire of the student to continue studies	0,100
Performance in Statistics	←	Student's performance in Mathematics from the last year	0,195
Performance in Statistics	←	Student's overall performance from the last year	0,230
Performance in Statistics	←	Tutoring in Statistics	0,096
Performance in Statistics	←	Overall Study of Mathematics	0,417

**TABLE 7**  
Exogenous Variables Correlations

Exogenous variables		Correlations
Student's overall performance from the last year	Overall Study of Mathematics	0,473
Student's place of	Overall Study of Mathematics	-0,094
Parents' educational level	Student's place of residence	-0,147
Parents' educational level	Tutoring in Statistics	0,107
Supplementary tuition in Statistics	Overall performance of last year	0,243

*continues*

Exogenous variables		Correlations
Parents' educational level	Overall Study of Mathematics	0,145
Tutoring in Statistics	Overall Study of Mathematics	0,410
Tutoring in Statistics	Student's place of residence	-0,170
Parents' educational level	Students' overall performance from the last year	0,054

Moreover, the relevant requirements that need to be satisfied for the application of S. E. M. are in place, so that the sound application of the model can be achieved<sup>12</sup>.

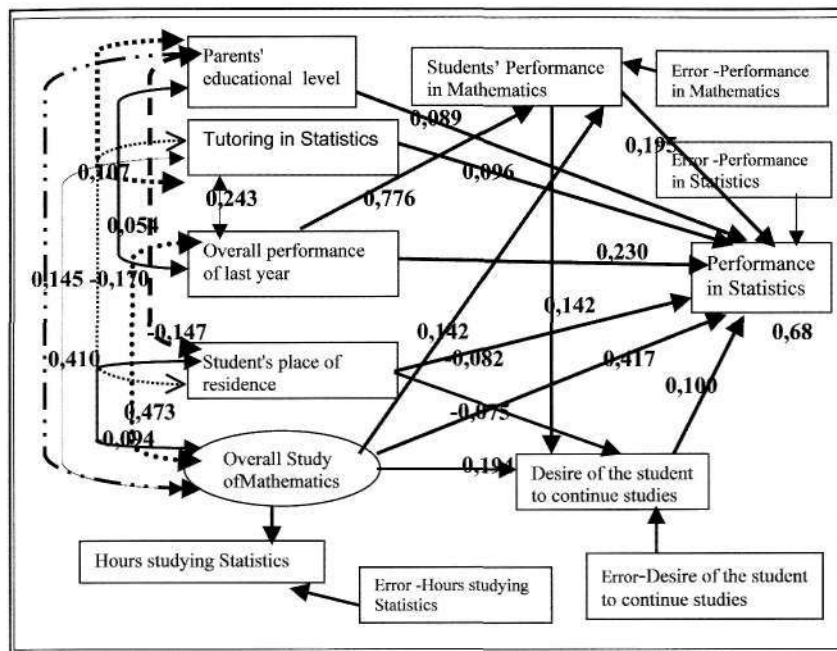
By examining the diagram of the SEM (High School) and table 6 we see, that out of all exogenous variables, the larger influence to the endogenous variables "Students' Performance in Mathematics from the last year" (path coefficient, 0,776) and "Performance in Statistics" (0,230) is exerted by the variable "Students' overall performance from the last year". Furthermore the latent variable "Overall study of Mathematics" exerts a significant influence on the endogenous variable "Performance in Statistics" (0,417) and a less significant influence on the endogenous variable "Students' Performance in Mathematics from the last year" (0,142). The rest of the influences that are exerted on the endogenous variables by the exogenous are of a lesser degree.

With regard to the endogenous variable that exerts influences on other endogenous, something that is possible (Draper N. & Smith H. 1997), it is noteworthy that there is influence of the endogenous variable "Students' Performance in Mathematics from the last year" on the endogenous one "Performance in Statistics" (0,195) and on the endogenous variable "Desire of the student to continue studies" (0,142).

Table 7 above and the S. E. M. diagram show that the largest correlations of the exogenous variables<sup>13</sup> is between variables "Overall study of Mathematics" and "Students' overall performance from the last year" (0,473) and between variables "Tutoring in Statistics" and "Overall Study of Mathematics" (0,410) and "Tutoring in Statistics" and "Students' overall performance from the last year" (0,243) and so on.

In general, we can say that the S. E. M. model that was implemented sufficiently interprets the behavior of the variable under study "Performance in Statistics", since the value of  $R^2$  (coefficient of determination  $-R^2 = 0,68$ ) is high

**DIAGRAM 1**  
The Diagram of the SEM (High School)



enough, being close enough (and slightly higher) to that obtained by R. A. ( $R^2 = 0.64$ ). This proves that S. E. M. yielded results that confirm and enhance those obtained by R. A.

## 7. The Case of the Technical Vocational Schools (V.T.S.)

### 7.1 The Results obtained by Standardized Multiple Regression Model (S.M.R.M.)

Working in a similar way as that used for U.H.S., we reached the construction of a regression model for the T.V.S., which is:

$$\hat{Y} = 0,305X_1 + 0,282X_2 + 0,295X_3 + 0,133X_4 + 0,085X_5 + 0,072X_6$$

Naturally, the model satisfies all the necessary conditions that the application of regression analysis requires and were elaborated in the case of U.H.S.

**TABLE 8**  
Presentation of the Regression Analysis Results

Case: V.T.S.							
Statistical data processing Method: The method of the Least Squares. Model: SMRM							
Dependent Variable (Y): Performance in the Subject of Statistics							
Explanatory Variables		Unstandard Coefficient Beta	Standard Coefficient Beta	t Statistics	Significance Value	R <sup>2</sup> Change	Sign. F Change
Constant (b <sub>0</sub> )		-2,145		-6,585	0,000		
X <sub>1</sub>	Students' Performance in Mathematics from the last year	0,184	0,305	8,872	0,000	0,420	0,000
X <sub>2</sub>	Hours studying Statistics per Week	0,344	0,282	9,019	0,000	0,150	0,000
X <sub>3</sub>	Student's overall performance from the last year	0,253	0,295	8,715	0,000	0,092	0,000
X <sub>4</sub>	Tutoring in Statistics	0,245	0,133	4,051	0,000	0,013	0,000
X <sub>5</sub>	Statistics' Usefulness	0,095	0,085	2,777	0,006	0,005	0,009
X <sub>6</sub>	Parents' educational level	0,048	0,072	2,357	0,019	0,004	0,019
- R <sup>2</sup> = 0,646 and R <sup>2</sup> <sub>adj.</sub> = 0,644							
- F <sub>6, 427</sub> = 151,725, p < 0,0005							

As we can see, out of all independent variables, the ones that exert the largest influence on the investigated variable Y (Performance in Statistics) are: The variable "Students' performance in Mathematics" (0,305), the variable "Student's overall performance from the last year" (0,295), variable "Hours studying Statistics per week" (0,282) and so on. Moreover the resulting R<sup>2</sup> = 0,68 is quite satisfactory.

### Student performance according to subject unit

At the outset, the overall performance of the first year students in the T.V.S. (in the field of Economy and Management) in the test taken in Statistics was approximately 9,5 points on the 20 point grading scale used in the S.E. in Greece. The breakdown according to subject unit is as follows

1. First unit (Introduction, statistical data collection and presentation) 10 points
- 2 (10,2). Second unit (Measures of Central Tendency) 11,2 points (the highest average of all units).
3. Third unit (Dispersion Measures) 8,8 points and
4. Fourth unit (Regression) 8 points.

**TABLE 9**

Students' Performance according to Subject Unit in U.H.S and T.V.S.

	<b>Subject Units</b>	<b>U.H.S.</b>	<b>T.V.S.</b>
1	Introduction, Statistical Data Collection and Presentation	12,2	10,2
2	Measures of Central Tendency	13,6	11,2
3	Dispersion Measures	12,2	8,8
4	Regression	11,2	8
5	Probabilities	11,5	-
	Student's overall performance in Statistics test	12,14	9,5

## **7.2 Presentation of S.E.M. Results**

As we can see in the correspondig tables (10 & 11) and the associated S.E.M. diagram (2), the largest influence is exerted by the exogenous variable "Students' Overall Performance from the Last Year" on the endogenous variable "Students' Performance in Mathematics from the Last Year" (0,374). In addition, the same exogenous variable exerts considerable influence on variable "Performance in Statistics" (0,250). Large influence is also exerted by variable "Students' Performance in Mathematics from the Last Year" on variable "Performance in Statistics" (0,334). The same exogenous variable, "Students' Performance in Mathematics from the Last Year", exerts considerable influence, also, on endogenous variable "Usefulness of Statistics" (0,307). The exogenous (latent) variable "Overall study of Mathematics" exerts influence on both variables "Performance in Statistics" (0,265) and "Students' Performance in Math-

ematics from the Last Year” (0,125) and on variable “Usefulness of Statistics” (0,247) and so on.

**TABLE 10**  
Standardized Regression Weights

Dependent Variables	←	Independent Variables	$b_i$
Student's performance in Mathematics from the last year	←	Student's overall performance from the last year	0,374
Student's performance in Mathematics from the last year	←	Overall Study of Mathematics	0,125
Student's performance in Mathematics from the last year	←	Parents' educational level	0,139
Student's performance in Mathematics from the last year	←	Supplementary tuition in Statistics	0,203
Statistics' degree of usefulness	←	Student's performance in Mathematics from the last year	0,307
Statistics' degree of usefulness	←	Overall Study of Mathematics	0,247
Performance in Statistics	←	Parents' educational level	0,074
Performance in Statistics	←	Statistics' degree of usefulness	0,101
Performance in Statistics	←	Student's performance in Mathematics of last year	0,334
Performance in Statistics	←	Student's overall performance from the last year	0,250
Performance in Statistics	←	Supplementary tuition in Statistics	0,137
Performance in Statistics	←	Overall Study of Mathematics	0,265

**TABLE 11**  
Exogenous Variables Correlations

Variables		Correlations
Parents' educational level	Supplementary tuition in Statistics	0,332
Supplementary tuition in Statistics	Student's overall performance from the last year	0,460
Parents' educational level	Overall Study of Mathematics	0,315

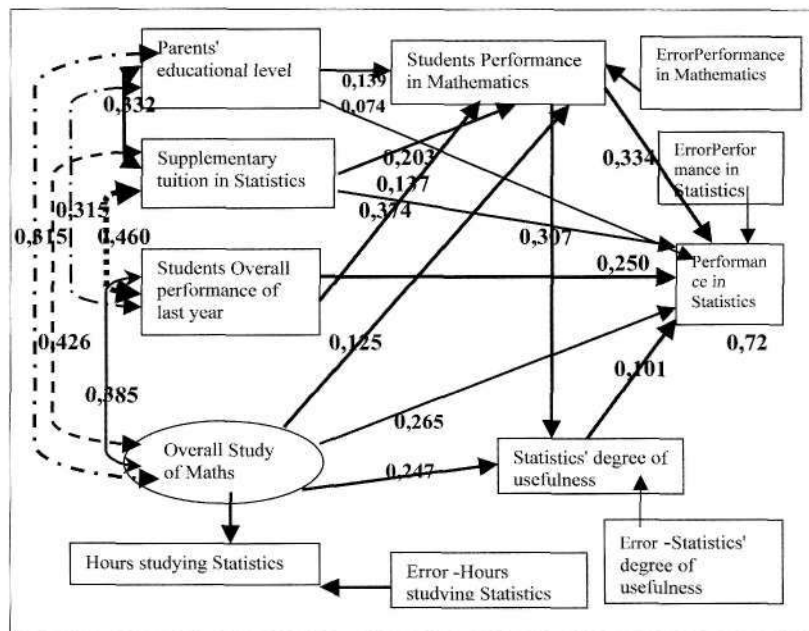
*continues*



Variables		Correlations
Supplementary tuition in Statistics	Overall Study of Mathematics	0,426
Student's overall performance	Overall Study of Mathematics	0,385
Parents' educational level	Student's overall performance from the last year	0,315

With regard to the correlation among the exogenous variables (Table 11), the large ones appear to be between variables "Student's overall performance from the last year" and "Tutoring in Statistics" (0,460), as well as, between variables "General study of Mathematics" and "Tutoring in Statistics" (0,426). In addition high correlation is observed between variables "General Study of Mathematics" and "Student's overall performance from the last year" (0,385) as well as between variables "Tutoring in Statistics" and "Parents' Educational Level" (0,332) and so on.

**DIAGRAM 2**  
The Diagram of the SEM (Technical School)



As a conclusion it can be said that the Structural Equation Modeling that was applied in the case of Technical Vocational School interprets to a high

degree the variation of the Y (scores in Statistics) since we obtain  $R^2 = 0,72$ . Additionally we observe that value of  $R^2$  obtained by S. E. M. is sufficiently close to that obtained by the more classical method of the Regression Analysis ( $R^2 = 0,68$ ) that was originally applied and it is even slightly larger. This is, therefore, an additional validation and confirmation of the results obtained by Regression Analysis as it was in the case of Uniform High School.

## 8. Conclusions

In both the Uniform High School (U.H.S.) and in the first grade of the Technical Vocational School -T.V.S., the research unequivocally demonstrate the large degree of influence that Mathematics exerts on the Performance of the Students in the Statistics test. This finding leads to the conclusion that at the beginning of the school year it is advisable for the instructor to administer a diagnostic test over the Mathematics material that is relevant and necessary for the understanding of the Statistics subject matter instruction for the rest of the year. Moreover it brings to the forefront the familiar subject of sufficient instruction in Mathematics (in Secondary Education), given that the Mathematics is the Foundation of other subjects (e. g., Physics, Economics).

The discovery of this research that the time that students dedicate in the study of this subject (Statistics), as well as, the general performance of the students in the previous year influence their performance in Statistics, brings to the forefront once again the known problem of finding ways to convince students to embrace books and school in general. Moreover, several variables of social nature made a difference, namely the parents' level of education, any tutoring that students may have undergone in Statistics, their region of residence and the opinion the students might have formed about the usefulness of Statistics. Based on these last observations it is evident that the role of "Human factor" is defining in matters of education. This is the role of the teacher whose duty calls him to find solutions, to the extent possible, to issues of social character that touch upon the educational process.

Moreover we are of the opinion that it is advisable to emphasize the multi subject and multi sided approach for the learning (Gnostic) objective of Statistics<sup>14</sup>. Statistics, given its many and varied applications, is found in other instructional fields like physics, Economics, Sociology and others. This fact creates the conditions for instructional synergy of these subjects and Statistics. Thus the students are aided in better understanding of its usefulness which is an explanatory variable, as we saw, of their performance in the subject of Statistics.

## Notes

1. The region area of Peloponnese included in our sample, in order to make comparisons (as the students' performance in Statistics) between urban centers and rural areas.

2. The abbreviation S.H.M. means: Split Half Method. The S.H.M. is the most applicable method to these researches -as the present one (Kassotakis, 1980, Kubiszyn, T. & Borich, G. 2000).

3. The  $r^2$  symbolizes the known correlation ratio of Pearson ( $0 < r^2 < 1$ ).

4. The table of Specification helps us to better division and presentation of the subject units of Statistics in connection with the categories of the test questions.

5. We Have  $X'Xb = X'Y$  and  $b = (X'X)^{-1} \cdot X'Y$ . Also,  $X'X = r_{xx}$  and  $X'Y = r_{yx}$ . Consequently, it's  $b = (r_{xx})^{-1} \cdot r_{yx}$ . Where, Y the vector of the Y prices, b the vector of the model's parameters and X the matrix of the Xi prices.

6. A big absolute t ( usually  $t > /2/$ ) and a small Sig (p) suggest that a predictor variable is having a large impact on the variable Y.

7. The negative sign (-) shows that the rural areas' students are inferiors to these ones (however, the relative degree is smallest) of the urban centers.

8. The full presentation of the Regression Model (R.M.) includes the term  $e_i$ , which is the difference between the real prices ( $y_i$ ) of the dependent variable Y and the corresponding prices ( $\hat{y}_i$ ) that gives the R.M.

9. In the case of U.H.S. -and the corresponding of V.T.S.- the significance values of the F Statistic are smallest, smaller than 0,05, as it is advisable, and the explanatory variables do a good job explaining the variation in the dependent variable. Generally, the F prices of this study, are very good (and both the cases -U.H.S. & V.T.S.).

10. The  $R^2$ , the coefficient of determination, shows the variation's degree of the dependent variable that is interpreted by the regression model.

11. The  $R^2_{adj} = 1 - (1 - R^2) \cdot n / (n-p)$ ,  $n$  = the sample's size,  $p$  = the number of the regression parameters,  $b_0$  = the regression constant term.

12. The results of the indicative indices are within limits (Ulman, J.B.,1998). Specifically, we have: RMSEA = 0,032 < 0,08. NFI = 0,999 > 0,90. TLI = 0,998 > 0,90. CFI = 1 > 0,90. RFI = 0,996, PCLOSE = 0,908 and so on.

13. This table doesn't show the direction of the influence between the variables. The table informs us, only, as the interaction's degree between two variables (Ulman, J.B., 1998, Hair, J. F. et al.,1998).

14. The last years the education ministry emphasizes on the multi subject approach for learning objectives that are taught in the secondary (and the primary) education of Greece.

## References

- Alimissis D. (1994), *Causes of the failure of Secondary School Students in physics. Department: Philosophy, Pedagogical and Psychology. Unpublished Doctoral Dissertation, National Kapodistrian University of Athens, Greece.*
- Arbucke J. & Worthke W. (1999), *Amos 4.0 User's Guide*, USA :Small Waters corp.
- Barkatsas E. et al. (1998), *Students' Mathematics Performance and their Attitude toward the Learning of Mathematics: An Attempt to Explore Their Relationship. Nordic studies in mathematics education cl*, pp. 23-48.
- Benton J. (1988), *How did they get that. Journal Mathematics (statistics) Teacher*, v. 81, n. 6, pp. 470-476.
- Barret Gloria (1999), *Investigating Distributions of Sample Means on the Graphing Calculator. Journal Citation: Mathematics-Teacher*, v. 92, n. 8, pp. 744-747.
- Bratton George (1999), *The role of Technology in Introductory Statistics Classes. Journal Citation: Mathematics Teacher*, v. 92, n. 8, pp. 666-669.
- Burrill Gail et al. (1992), *Data Analysis and Statistics across the Curriculum. Curriculum and Evaluation Standard for School Mathematics Addenda Series. National Council of Teachers of Mathematics.*
- Draper N. & Smith, H. (1997), *Applied regression analysis*. Greece, Athens: Papazissis SA.
- du-Feu Chris (1999), *A Sort of Statistics Lesson. Journal Citation: Teaching Statistics*, v. 21, n. 1, pp. 8-10.
- Flavor J. (1999), *On approximate calculations. Journal citation: Teaching Statistics*, v. 21, n. 3, pp. 84-87.
- Friedman, Hershey, et al. (1999). *Teaching statistics using humorous anecdotes"*, *Journal: Mathematics teacher*, v. 92, n. 4, pp. 305-308.
- Godino J. D. et al. (2000). *The statistical consultancy workshop as a pedagogical. Training researchers in the use of statistics*, pp. 339-353, International Association for Statistical Education - International Statistical Institute - Tokyo.
- Graham Alan (1999), *Statistical Nuggets with a Graphics Calculator. Journal Citation: Teaching Statistics*, v. 21, n. 1, pp. 24-27.
- Hubbarb R., *Teaching Statistics with Minitab, Journal Citation: Australian Mathematics Teacher*, v. 48, p. 1012.
- Hair J. F. et al. (1998) *Multivariate Data Analysis*, N.J.: Prentice Hall.
- Hodgson, T. & Borkowski, J. (1998). *Why stratify? Journal citation: Teaching Statistics*, v. 20, n. 3, pp. 66-67.
- Kassotakis, M., (1980), *The valuation of the Students Performance*. Athens.

- Kontogiannopoulou-Polidoridou et al. (2000), Detecting the Students' Performance in the Education of Greece. Metaichmio. Athens.
- Keith Z. T., (1993), *latent Variable Structural Equation Models: L. I. S. R. E. L in Special Education Research, Remedial and Special Education*, v. 14, n. 6, pp. 36-46.
- Kubiszyn, T. & Borich, G. (2000). Educational Testing and Measurement. (6th Ed.). USA : J. Willey & Sons.
- Lossif, G. (1999). *The graphics calculator as a teaching aid in statistics. Journal citation: Teaching statistics* v. 21, n. 2, pp. 45-48.
- Moore A. D., (1995), *Structural Equation Modeling in special Education Research, Remedial and Special Education* v. 16, n. 3, pp. 178-183.
- National Council of Teachers of Mathematics (2000), "Principles and Standards for School Mathematics".
- Paas et al. (1992), *Training Strategies for attaining transfer of Problem- Solving skill in Statistics: A Cognitive Load Approach, Journal of Educational Psychology*, v. 84, n. 4, pp. 249-234, USA.
- Papoulis A. (1991), Probability, Random Variables and Stochastic Process. Mc. Grow -Hill, Inc.
- Schults J., et al. (1990). *Integrating Statistics into a course on functions", Journal citation: Mathematics Teacher*; v. 83, n. 8, pp. 612-616.
- Schwertman N. (1999). *Discovering an optimal property of the Median", Journal citation: Mathematics teacher*; v. 92, n. 8, pp. 692-703.
- Starkweather & Kendall. (1999). *Technological literacy: The rationale and structure for the study of technology in our schools. Journal, Contemporary Educational Systems, Nov., pp. 177-215.*
- Takis Sandra (1999), *Titanic: A Statistical Exploration. Journal Citation: Mathematics Teacher*, v. 92, n. 8, pp. 660-664.
- Talsma Gary (1999). *Data Analysis and Baseball. Journal Citation: Mathematics Teachers*, v. 92, n. 8, pp. 738-742.
- Tatsuro, M. (1986), *Probability and Statistics Teaching in Japanese Senior High School. Journal of educational study in mathematics*, v. 5, p.p. 105-117.
- Ulman J. B. (1998), Structural Equation Modeling, in the book:Using Multivariate Statistics, Tabacnick et al., Wesley 3<sup>rd</sup> Ed.
- Verdis Athanasios (2002). School Effectiveness Researches for Educational Evaluation in Greece. Institute of Education. Doctoral Dissertation, London.