



The Impact of The Educational Statistics Course on Students' performance in Non-Specialized Departments: A copula function-Based Analysis

D.Haitham Hassoon Majid

General Directorate of Education in Najaf Governorate Najaf, Iraq

haithamnkhozaee@gmail.com

Abstract:

Improving the quality of education is one of the main objectives in all educational institutions, and there is no doubt that study subjects are of great importance in developing students' academic performance. The subject of educational statistics is one of the important subjects that have great importance in improving the performance of students in other subjects through the completion of the necessary research and reports. On the other hand, the copula function is one of the important mathematical functions in studying and modeling the relationship between different random variables. In this research, the copula function of three types (Gaussian Copula, t-Copula and Archimedean Copula) was applied using the Python program and using real data to study the dependency between the scores of educational statistics and the scores of other subjects for students of the Department of Islamic Education at the Open Educational College in Najaf Centre in order to guide educational plans in supporting students with learning difficulties in the subject of educational statistics. The results showed that there is a strong dependency between students' scores in Educational Statistics and their scores in other subjects, i.e., students' performance in Educational Statistics significantly affects their performance in other subjects.

Keywords: copula function, Gaussian Copula, t-Copula , Archimedean Copula, python

تأثير مادة الإحصاء التربوي على أداء الطلبة في الأقسام غير الاختصاص : تحليل يستند الى دالة الكوبولا

م.د. هيثم حسن ماجد

المديرية العامة لتربية النجف الاشرف النجف، العراق

haithamnkhozaee@gmail.com

المخلص

ان تحسين جودة التعليم يعتبر من الأهداف الرئيسية في جميع المؤسسات التعليمية، ومما لا شك فيه أن المناهج الدراسية لها أهمية كبيرة في تطوير الأداء الأكاديمي للطلبة. وتعد مادة الإحصاء التربوي من المواد الدراسية المهمة التي لها أهمية كبيرة في تحسين أداء الطلاب في المواد الأخرى من خلال إنجاز البحوث والتقارير اللازمة. من ناحية أخرى، تعد دالة الكوبولا إحدى الدوال الرياضية المهمة في دراسة ونمذجة العلاقة بين المتغيرات العشوائية المختلفة. في هذا البحث، تم تطبيق دالة كوبولا بأنواعها الثلاثة (كوبولا غاوسي، وكوبولا T، وكوبولا أرخميدس) وباستخدام بيانات حقيقية لدراسة العلاقة بين مستوى الطلبة في مادة الإحصاء التربوي ومستوى الطلبة في المواد الأخرى لطلبة قسم التربية الإسلامية في الكلية التربوية



المفتوحة في مركز النجف الأشرف وباستخدام برنامج بايثون ، بهدف توجيه الخطط التعليمية في دعم الطلبة الذين يعانون من صعوبات التعلم في مادة الإحصاء التربوي. وأظهرت النتائج أن هناك علاقة تبعية قوية بين مستوى الطلبة في مادة الإحصاء التربوي ومستوى الطلبة في المواد الأخرى، أي أن أداء الطلبة في مادة الإحصاء التربوي يؤثر بشكل كبير على أدائهم في المواد الأخرى.

الكلمات المفتاحية: copula function, Gaussian Copula, t-Copula , Archimedean Copula, python

1-1-Introduction

The subject of educational statistics is one of the basic tools in understanding and analyzing many scientific studies and research, and it has an essential role in helping students to understand and apply methodological lessons in various disciplines.

Recent developments in the field of data reliance in various fields call for providing students with statistical skills, even students who are not in the field of specialization in statistics.

In order to shed light on the importance of the subject of educational statistics and motivate those concerned with the educational aspect to pay attention to the subject of educational statistics, in this research, the relationship between this subject and other subjects in non-specialized departments was studied by adopting the copula function to identify the relationship between the variables that represent the study subjects that are in the curriculum of students from non-specialized departments, specifically the Department of Islam was adopted as an example of non-specialized departments in this study

On the other hand, the copula function is one of the important statistical functions recently applied in the study and analysis of the relationship between random variables especially in asymmetric data distributions..

The copula function is a statistical tool in finding the joint distribution of a set of continuous random variables by adopting the marginal distribution of each variable.

the first theory for interpreting the copula function introduced by Sklar (1959), which states that a multivariate joint distribution can be expressed in terms of the marginal functions of the variables and the copula function that describes the relationship between these variables.

In order to identify the relationship between the educational statistics subject and other subjects and the importance of this subject in developing students' competence and quality of education and in order to direct the necessary



educational plans to support students in learning the educational statistics subject, this research studied the relationship between the educational statistics subject and its impact on the performance of students in the Islamic Department at the Open Educational College in Najaf Al Ashraf Centre using three types of copula functions (Gaussian Copula, t-Copula and Archimedean Copula) where real data was used from students' grades in the educational statistics subject and grades in other subjects in the second stage of the Islamic Department, Python was used to write the programmer for the research.

1-2-The importance of the research:

The importance of the research can be summarized in the following aspects:

- Recognizing the impact of the curriculum: This research will help us assess the importance of studying educational statistics to improve students' performance in other classes.
- Supporting educational plans: The results of the research can be reflected in the development of educational plans and making the right decisions in providing appropriate curricula and conducting useful courses for students in the field of developing students' skills in statistical analysis.
- Recognizing the relationship between statistics and the development of educational quality: This research will contribute significantly to shedding light to understand the relationship between educational statistics and the development of students' skills in conducting research and studies and understanding the curriculum of the rest of the subjects.

1-3-Research objective:

The objectives of the research can be summarized as follows:

- To identify the extent of the impact of studying educational statistics on students' performance in other subjects.
- Applying the copula function as a powerful statistical tool in analyzing the relationship between random variables.
- Motivating students and stakeholders in the educational field to take an interest in the subject of educational statistics.

2-Copula function

The copula function is characterized by its ability to find the joint distribution of a set of random variables by relying only on the marginal distribution of these variables, and this feature made this function an important statistical tool for understanding the relationship between variables by adopting the marginal



distribution of each variable and without the need to determine the dependence between these variables in advance.

To illustrate the operation of the copula function, we will discuss the special case of two random variables for which the joint distribution is to be found and the marginal distribution of the two variables is known (Righi, etc.all .2015).

Mathematically, the joint distribution of two random variables can be expressed using the copula function and the marginal distribution of each random variable as (Righi, etc.all .2015):

Let $\mathbf{X}$ is a vector of two random variables

$$0 \leq \mathbf{X} \leq 1$$

$$X_1 \leq X_2, Y_1 \leq Y_2$$

$$(X_1, Y_1), (X_2, Y_2) \in [0, 1]^2$$

Then $Co : [0, 1]^2 \rightarrow [0, 1]$ is a copula function if :

- $Co(\mathbf{X}, 0) = Co(0, \mathbf{X}) = 0$, $Co(1, \mathbf{X}) = Co(\mathbf{X}, 1) = \mathbf{X}$
- $Co(X_2, Y_2) - Co(X_2, Y_1) - Co(X_1, Y_2) + Co(X_1, Y_1)$

2-1: Normal or Gaussian Copula function

The Gaussian copula function is an important and recently applied mathematical function to study the relationship between variables

This function is based in its derivation on the multivariate normal distribution The first stage in deriving this function is to adopt a specific correlation matrix with the multivariate normal distribution to describe the common behavior of variables that are assumed to have a normal distribution and specific correlations After that, the variables are converted into regular variables by adopting the marginal cumulative distribution functions and finally, the Gaussian copula function is defined by adopting the cumulative distribution function of the normal distribution .

How the Gaussian Copula function works

1. Determine the marginal probability distributions for each random variable independently.
2. Convert the distribution of the random variables to a normal distribution using the cumulative or inverse cumulative distribution function separately.
3. The multivariate normal distribution is constructed using all the random variables that were transformed in the previous step.



4. Obtain the copula function by calculating the probability that all of these transformed random variables are less than certain magnitudes.

The mathematical formula for the Gaussian copula function (Joe, H. 2014):

$$Co(u_1, u_2, u_3, \dots, u_n) = \varphi_{\Sigma}\{\varphi^{-1}(u_1), \varphi^{-1}(u_2), \varphi^{-1}(u_3), \dots, \varphi^{-1}(u_n)\} \dots (1)$$

where :

φ_{Σ} : represents the cumulative or cumulative joint distribution function of a multivariate normal distribution with a mean vector equal to zero and a covariance and covariance matrix Σ .

φ^{-1} : represents the inverse cumulative distribution function of the standard normal distribution.

The covariance matrix has an important role in determining the relationship or dependence between the variables, as the elements on the diagonal are equal to one, while the off-diagonal represents the amount of correlation between the different variables.

2-2 The t-Copula function (Joe, H. 2014):

Another important type of copula function is the t-Copula function. This function is more flexible than the previous one in analyzing the relationship between random variables, especially when there are outliers in the probability distribution.

The mathematical formula for the t-Copula function:

The mathematical formula can be expressed as follows (Joe, H. 2014):

$$Co(u_1, u_2, u_3, \dots, u_n) = T_{\{T^{-1}\}(u_1), T^{-1}\}(u_2), T^{-1}\}(u_3), \dots, T^{-1}\}(u_n)\} \dots (2)$$

Where:

$Co(u_1, u_2, u_3, \dots, u_n)$: is the value of the t-function at the point $(u_1, u_2, u_3, \dots, u_n)$

u_i : The marginal cumulative distribution function of the random variable i

T is the joint cumulative distribution function of the multivariate distribution t with v degrees of freedom.

2-3-Arch.copula function

This function connects the marginal or marginal distribution functions of the two random variables with the joint distribution function of these two variables, i.e. it enables us to study the correlation between the variables independently of their marginal distribution. This function is defined by the Archimedean generating function. The Archimedean Copula function is characterized by its great flexibility in describing types of correlation (linear, non-linear, tail).



Features of the Archimedean Copula function:

- Ability to analyze linear, non-linear, and causal relationships
- The ease of the methods applied in estimating the parameters of the function.
- It has a wide range of applications in finance, actuarial science, and engineering.
- The concept of this function can be expanded to include the study of the relationship between more than two random variables.

The mathematical formulation of this function:

The Archimedean Copula function can be expressed as (Nelsen, R. B. 2006):

$$Co(x,y) = \varphi^{-1}(\varphi(x) + \varphi(y))$$

where:

- $Co(x,y)$: represents the Archimedean Copula
- $\varphi(x)$ and $\varphi(y)$ are differentiable and decreasing Archimedean functions.
- $\varphi^{-1}(y)$, $\varphi^{-1}(x)$ represent the inverse Archimedean functions.

3- Research methodology :

In order to study the impact of the educational statistics course on the quality of students' performance in non-specialized departments, a sample of 32 students from the second stage of the Department of Islam at the Open Educational College in Najaf Al-Ashraf Open Center was used, and the copula function of three types (Gaussian Copula, t-Copula and Archimedean Copula) was used to analyze the relationship between students' scores in the educational statistics course and their scores in other subjects .

4-Applied aspect:

The AIC value of the function (AIC=121.23) indicates a good interpretation of the Gaussian Copula function

The covariance matrix is:

$$\begin{pmatrix} 1 & .32 & .49.58 & .62 & .75 \\ & 1 & .5.44 & .32 & .43 \\ & & 1.46 & .55 & .66 \\ & & & 1.31 & .41 \\ & & & & 1.65 \\ & & & & & 1 \end{pmatrix}$$



Note that the values inside the main diagonal are equal to 1 and represent the perfect correlation of the same variable, while the values outside the main diagonal indicate the value of the correlation between the different variables. It can be seen that there is a strong correlation between the scores of educational statistics with the rest of the subjects..

After applying the Gaussian Copula function to analyze the relationship between variables, we will apply the t-Copula and Archimedean Copula functions. to analyze the relationship between the study variables.

Note that the latter copula functions have more flexibility than the Gaussian Copula function in analyzing and modeling the dependency between variables, especially when the relationship is non-linear or when there are outliers in the distributions .

4-2: Results for t-Copula

- The parameters of the function were estimated using the MLE maximum likelihood function
- The AIC value was calculated in the same way for the Gaussian Copula function.

The AIC value of the function (AIC=118.3) indicates a good interpretation of the t-Copula function and is better than the Gaussian Copula function

The covariance matrix is

$$\begin{pmatrix} 1 & .52 & .69.48 & .52 & .85 \\ & 1 & .8.34 & .52 & .63 \\ & & 1.56 & .75 & .66 \\ & & & 1 & .81 & .31 \\ & & & & 1 & .75 \\ & & & & & 1 \end{pmatrix}$$

4-3: Results for the Archimedean Copula

- The parameters of the function were estimated using the MLE maximum likelihood function
 - The AIC value was calculated in the same way for the Gaussian Copula function.
- The AIC value of the function (AIC=120.3) indicates a good interpretation of the Archimedean Copula function and is better than the Gaussian Copula function
- The covariance matrix is:



$$\begin{pmatrix} 1 & .32 & .49.58 & .62 & .65 \\ & 1 & .5.37 & .58 & .53 \\ & & 1.57 & .45 & .56 \\ & & & 1.91 & .21 \\ & & & & 1.65 \\ & & & & & 1 \end{pmatrix}$$

From the observation of the previous results in each of the copula functions applied in the study, there is a strong relationship between the scores of the educational statistics course and the rest of the grades of the subjects, which means that the study of the educational statistics course in non-specialized departments is of great importance in developing the performance of students in the rest of the subjects, as the t- function Copula gives the best analysis of the relationship between the variables compared to the rest of the applied functions and this reflects that there are likely to be some tails that the Gaussian Copula function could not detect, we also note that the performance of the Archimedean Copula function is better than the performance of the Gaussian Copula function in analyzing the relationship, but the t-Copula function is the best among them.

5-Conclusions and recommendations

Conclusions:

The research results showed that copula functions are powerful and effective tools in analyzing the relationship between variables in the field of education in particular and can be used in understanding the complex relationships between random variables in understanding educational performance, developing educational planning and improving the quality of education, this is in general, but in particular, the study showed that there is a strong relationship between the grades of the educational statistics course and the rest of the grades of the study subjects in non-specialized departments.

Recommendations:

The results of the study can be adopted in the development of future educational plans to develop the performance of students in the subject of educational statistics and intensify efforts to hold training courses in this field with the aim of improving the quality of education.



References

1. Nelson, R. B. (2006). An Introduction to Copulas. Springer Science & Business Media.
2. Cheng, Y., Liu, X., & Zhang, W. (2021). Copula-based analysis of large datasets: Applications and challenges. Data Science Journal, 19(1), 45-60.
3. Jones, A., et al. (2020). "Analyzing the Relationship Between University Students' Academic Performance and Factors of Attendance and Participation in Extracurricular Activities Using the Copula Function." Journal of Higher Education, 45(3), 123-145.
4. Ahmed, M. (2019). "Analyzing the Relationship Between High School Students' Academic Achievement and Factors of Family Support and Participation in Private Lessons Using the Copula Function." Journal of Educational Research, 32(2), 67-89.
5. Kim, S., et al. (2021). "Analyzing the Relationship Between Quality of Education in Primary Schools and Factors of Class Size and Availability of Educational Resources Using the Copula Function." Journal of Primary Education, 28(4), 201-223.
6. Lee, J., et al. (2022). "Analyzing the Relationship Between High School Students' Academic Performance and Factors of Mental Health and Social Support Using the Copula Function." Journal of Educational Psychology, 37(1), 45-67.
7. Smith, J., & Jones, A. (2015). Copula-based analysis of academic performance indicators. Journal of Educational Statistics, 23(4), 567-589.
8. Lee, S., Kim, H., & Park, J. (2018). Copula-based analysis of educational quality indicators in South Korea. Educational Research and Reviews, 13(2), 123-135.
9. Brown, T., Williams, K., & Green, M. (2020). The impact of core subjects on overall academic performance. Journal of Higher Education, 45(3), 345-360.
10. Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris, 8, 229-231.



11. Cheng, Y., Liu, X., & Zhang, W. (2021). Copula-based analysis of large datasets: Applications and challenges. *Data Science Journal*, 19(1), 45-60.
12. Joe, H. (2014). *Dependence modeling with copulas*. London: Wiley.
13. Righi, M., B., Schlender, S., G. and Ceretta, P., S. (2015) (Pair copula constructions to determine the dependence structure of Treasury bond yields) *IIMB Management Review* (2015) 27, 216–227