Research Article
Received: 24.11.2024
Accepted: 30.12.2024
Published: 31.12.2024

Citation: Toprak, E., & Kızılok Kara, E. (2024). Statistical analysis of educational data using copula functions: The case of 2018 PISA. *International Journal of Scholars in Education, 7*(2), 169-184. https://doi.org/10.52134/ueader.1590602

Statistical Analysis of Educational Data Using Copula Functions: The Case of 2018 PISA*

Ebru TOPRAK**, Emel KIZILOK KARA***

Abstract**:** In this study, using 2018 PISA data, the dependency structures between mathematics, science, and reading scores were analyzed with elliptic and archimedean copula functions according to the factors of gender and educational level at home. PISA is an international assessment that measures the education levels of students in science, mathematics, and reading and aims to compare education levels among countries. Copula functions are statistical tools that allow flexible modeling of dependency relationships between variables and provide the most appropriate way of obtaining multivariate distributions. In this study, firstly, models were constructed for data pairs consisting of PISA maths, reading, and science scores according to home education level and gender factors. Then, the copula models that best explain these structures were determined by goodness-of-fit tests, and the copula parameters for the selected models were estimated. Finally, joint and conditional probabilities were calculated for these score pairs in order to evaluate the effect of reading scores on maths and science courses. The study emphasizes the potential use of copula models in educational research and provides new findings on the impact of gender and home education level on PISA performances.
Keywords*:* Dependency, Copula, Conditional probability, Joint probability, PISA achievement scores.

Kopula Fonksiyonları ile Eğitim Verilerinin İstatistiksel Analizi: 2018 PISA Örneği

Öz: Bu çalışmada, 2018 PISA verileri kullanılarak matematik, fen ve okuma puanları arasındaki bağımlılık yapıları, cinsiyet ve evdeki eğitim düzeyi faktörlerine göre eliptik ve arşimedyan kopula fonksiyonlarıyla analiz edilmiştir. PISA, fen, matematik ve okuma alanlarında öğrencilerin eğitim seviyelerini ölçen ve ülkeler arasındaki eğitim düzeylerini karşılaştırmayı amaçlayan uluslararası bir değerlendirmedir. Kopula fonksiyonları ise değişkenler arasındaki bağımlılık ilişkilerini esnek bir şekilde modellemeye olanak tanıyan ve çok değişkenli dağılımların en uygun biçimde elde edilmesini sağlayan istatistiksel araçlardır. Çalışmada, ilk olarak evdeki eğitim düzeyi ve cinsiyet faktörlerine göre PISA matematik, okuma ve fen puanlarından oluşan veri çiftleri için modeller oluşturulmuştur. Daha sonra, bu yapıları en iyi açıklayan kopula modelleri, uyum iyiliği testleri ile belirlenmiş ve seçilen modellere ilişkin kopula parametreleri tahmin edilmiştir. Son olarak, okuma puanının matematik ve fen dersleri üzerindeki etkisini değerlendirebilmek amacıyla bu puan çiftleri için ortak olasılıklar ile koşullu olasılıklar hesaplanmıştır. Çalışma, kopula modellerinin eğitim araştırmalarındaki potansiyel kullanımını vurgulayarak cinsiyet ve evdeki eğitim düzeyinin PISA performansları üzerindeki etkilerine ilişkin yeni bulgular sunmaktadır.
Keywords*:* Bağımlılık, Kopula, Koşullu olasılık, Ortak olasılık, PISA başarı puanları.

---

* This study was produced from the master's thesis prepared by the first author under the supervision of the second author.

** M.Sc. Student, Kırıkkale University, Institute of Science, Department of Statistics, ORCID:0000-0003-4753-5080, ebruttoprak@gmail.com
*** Assoc. Prof. Dr., Kırıkkale University, Faculty of Economics and Administrative Sciences, Department of Actuarial Sciences, ORCID:0000-0001-7580-5709 , emel.kizilok@kku.edu.tr (Corresponding author)

Introduction

PISA, which was first implemented in 2000 and has been implemented every three years to date, and in which Turkey first participated in 2003, is an important international project that allows all participating countries to evaluate their education systems. "Policy makers around the world use PISA results to compare the knowledge and skill levels of students in their own countries with the knowledge and skill levels of students in other countries participating in the project, to set standards for raising the level of education (e.g. average scores achieved by countries, educational outcomes of countries and their capacity to ensure the highest level of equality in educational opportunities) and to identify the strengths and weaknesses of education systems (MEB, 2013; Toprak, 2017). In addition to revealing the status of countries in the measured areas, PISA results also identify periodic progress, regression or stagnation thanks to the fact that the application is carried out every three years. PISA applications conducted by the Organization for Economic Cooperation and Development (OECD) also provide detailed tables comparing the economic development of countries with their educational status (OECD, 2014). Considering the importance that countries attach to PISA applications in order to see their status in the field of education and, accordingly, to correct their deficiencies, examining PISA achievements, which are extremely critical, emerges as an important problem situation. At this point, the subject of the study was determined to examine the relationships between PISA science, mathematics and reading achievements in terms of different variables.

The copula method will be used to model the dependency structures between PISA science, mathematics and reading achievements. Copula is a probabilistic modeling method that is frequently used to model the dependence of multivariate data in fields such as finance, economics and actuarial science. Copulas allow to model the dependence structure of the joint distribution independently of the marginal distributions. It can also explain the dependence structure between variables when the marginals are not normally distributed.

Copula models provide a wide range of applications in different fields such as financial data analysis (Patton, 2013; Kizilok Kara et al. 2022; Kara and Kemaloglu, 2016), hydrology (Favre et al. 2004; Genest and Favre 2007; Kizilok Kara and Yildiz, 2014, Baykal, 2024) and environmental data modeling (Goda 2010; Kwon and Yoon 2017; Kizilok Kara 2017; Bhatti and Do 2019; Nguyen-Huy, 2019), allowing the examination of dependency structures. However, as a result of the research, although there are studies using PISA data in Turkey (Anıl 2009; Kasap et al. 2021; Sarıer 2021), a limited number of studies have been identified in which the dependency structure between achievement scores is modeled with the copula method. The most recent study in this field was conducted by Pala and Sağlam (2019) for 2006-2015 PISA data and does not include an evaluation by demographic characteristics.

The aim of this study is to analyze the dependency structures between math, reading and science scores according to variables such as gender and level of education at home with the help of copula functions using 2018 PISA (URL-1) data. The most appropriate copula functions for each demographic group were determined by goodness-of-fit tests. Both joint and conditional probabilities were calculated through these copula functions, with a particular focus on the probability of students having above average achievement scores. By using up-to-date data and evaluating dependency structures based on demographic variables, the study aims to make a new methodological contribution to the literature.

Elliptical (Gaussian and t-copula) and archimedean (Clayton, Frank, Gumbel and Joe) copula families are used throughout the study. Maximum Pseudo-Likelihood Estimation (MPLE) method is used to estimate the copula parameters. Akaike Information Criterion (AIC), Bayes Information Criterion (BIC), Log-Likelihood Function (LL), and Cramer von Mises (CvM) values were used for goodness-of-fit tests and model selection. The analyses are performed with the 'copula' package in R software written by Hofert et al. (2024) and we refer to key literature sources such as Cherubini et al. (2004), Joe (2014), Nelsen (2006) and Emrechet (2003).

The findings of this study provide an important resource for the formulation of educational policies based on PISA results. Revealing different dependency structures according to demographic variables can contribute to the efforts to provide equal opportunities in education and to shape policies for student achievement.

In the rest of the study, firstly, copula, joint and conditional probability definitions according to copula, elliptic and archimedean copula functions, parameter estimation and model selection are introduced. Then, the models based on PISA data and demographic variables are described. Descriptive statistics, correlation and symmetricity test results of the models are presented. Then, the best copula models selected for the models according to the goodness-of-fit test results and the copula analysis results including parameter estimates are presented. Finally, some joint and conditional probabilities are calculated. The study is concluded with the conclusion section.

Materials and Methods

In this section, the copula, some copula functions, joint probabilities and conditional probabilities according to copula, parameter estimation, goodness of fit tests and model selection criteria are introduced.

Copula Theory

Copulas are mathematical tools used to model the dependence structure between multivariate distributions. They offer the possibility to analyze non-linear dependencies without requiring any assumptions about marginal distributions. Copulas create reliable models even when the data is not normally distributed and provide the flexibility to accurately reflect dependence under different marginal distributions. Thanks to these properties and the ability to calculate joint and conditional probabilities, copulas have a wide range of applications in areas such as finance, insurance, risk management and actuaries.

Copulas are defined by Sklar's Theorem introduced by Sklar (1959). According to this theorem, for a continuous random vector (X, Y) with marginals F and G, two-dimensional joint distribution function $H(x, y)$ is defined with only a copula $C : [0,1]^2 \rightarrow [0,1]$.

$$H(x,y) = P(X \leq x, Y \leq y) = C(F(x), G(y)) \tag{1}$$

On the other hand, the joint survival function $\overline{H}(x,y) = P(X > x, Y > y)$ depending on the copula is defined as

$$\overline{H}(x,y) = \bar{F}(x) + \bar{G}(y) - 1 + C\big(1 - \bar{F}(x), 1 - \bar{G}(y)\big) \tag{2}$$

Here $\bar{F}$ and $\bar{G}$ are the marginal survival functions of the random variables X, and Y, respectively.

On the other hand, conditional probabilities can be defined using joint probability information C and $\bar{C}$ with the marginals F, and G. The conditional probabilities used in the study are given below.

$P(\mathrm{X} > \mathrm{x} | \mathrm{Y} > \mathrm{y})$: The probability that Y the variable is above a certain value, given that X the variable is above a certain value:

$$P(X > x | Y > y) = \frac{P(X>x, Y>y)}{P(Y>y)} = \frac{\overline{H}(x,y)}{\bar{G}(y)} \tag{3}$$

$P(Y > y | X > x)$: The probability that X the variable is above a certain value, given that Y the variable is above a certain value:

$P(Y > y|X > x) = \frac{P(X>x,Y>y)}{P(X>x)} = \frac{\bar{H}(x,y)}{\bar{F}(y)}$ (4)

When the marginal distributions $U(0,1)$ are uniformly distributed, the copula $C(u,v)$ and the copula-based the joint survival function $\bar{C}(u,v)$ with transformations $F(x) = U$ and $G(y) = V$, respectively, are defined as follows (Nelsen, 2006):

$C(u,v) = P(U \leq u, V \leq v) = H(F^{-1}(u), G^{-1}(v))$ (5)

$\bar{C}(u,v) = P(U > u, V > v) = 1 - u - v + C(u,v)$ (6)

The conditional probabilities according to the copula $C(u,v)$ and the joint life function $\bar{C}(u,v)$ are given below:

If we denote by $\bar{C}_v(u)$ the probability that U is above a certain value when V is known to be above a certain value, and similarly, if we denote by $\bar{C}_u(v)$ the probability that V is above a certain value when U is known to be above a certain value, these conditional probabilities can be defined as follows, respectively.

$\bar{C}_v(u) = P(U > u|V > v) = \frac{P(U>u,V>v)}{P(V>v)} = \frac{\bar{C}(u,v)}{\bar{G}(v)}$ (7)

$\bar{C}_u(v) = P(V > v|U > u) = \frac{P(U>u,V>v)}{P(U>u)} = \frac{\bar{C}(u,v)}{\bar{F}(u)}$ (8)

Copulas are classified into several families to model dependency structures in different ways, the most well-known being elliptic and archimedean. copulas. Elliptical copulas (Gaussian and t) are known for their ability to model symmetric dependencies, while archimedean copulas (Clayton, Frank, Gumbel, and Joe) offer flexible dependency structures. The Frank copula offers a symmetric structure that models upper and lower tail dependence, while the Clayton copula has an asymmetric structure, and models negative left tail dependence. The Gumbel copula is asymmetric and is mostly used to model positive right tail dependence. The Joe copula is a copula that models positive dependencies and right tail dependencies (Cherubini et al. 2004, Joe 2014, Nelsen 2006, Embrechts et al. 2003). The mathematical functions for these copulas are given below:

Gaussian copula:

$$C_\theta(u,v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} exp\left(-\frac{s^2 - 2\theta st + s^2}{2(1-\theta^2)}\right) dsdt, \qquad \theta \in [0,1]$$

t copula:

$$C_\theta(u,v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{\phi_v^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} exp\left(-\left(1 + \frac{s^2-2\theta st+t^2}{v(1-\theta^2)}\right)\right) dsdt, -1 < \theta < 1, v > 2$$

Clayton copula:

$$C_\theta(u,v) = \left[max\left(u^{-\theta} + v^{-\theta} - 1; 0\right)\right]^{-1/\theta}, \ \theta \in [-1,\infty) \setminus \{0\}$$

Frank copula:

$$C_\theta(u,v) = -\frac{1}{\theta} ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right), \ \theta \neq 0$$

Gumbel copula:

$$C_\theta(u,v) = exp\left(-\left[(-logu)^\theta + (-logv)^\theta\right]^{1/\theta}\right), \qquad \theta \in [1,\infty)$$

Joe copula:

$$C_\theta(u,v) = 1 - ((1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta)^{(\frac{1}{\theta})}, \qquad \theta \in [1,\infty)$$

***Parameter Estimation*:** Parametric and nonparametric methods are used for copula parameter estimation. Among parametric methods, maximum likelihood estimation (MLE) estimates the parameters of the copula and marginal distributions simultaneously, while the inference function of marginals (IFM) method performs these estimates in two steps: in the first step, the marginal distribution parameters of each variable are estimated independently; in the second step, the parameters of the copula function are estimated using these marginal distributions (Joe and Xu, 1996). Among nonparametric methods, the canonical maximum likelihood (CML) method does not directly estimate marginal distributions but instead estimates copula parameters using empirical distribution functions (Cherubini et al., 2004). The maximum pseudo-likelihood estimation method (MPLE) is a simple and widely used approach that estimates copula parameters using pseudo-observations generated based on the ordering of the data without using empirical distribution functions (Shih and Louis, 1995; Genest et al. 1995).

In this study, the pseudo maximum likelihood estimation (MPLE) method expressed by equation (9) with respect to the copula density function $c(u_i, v_i; \theta)$ is used for parameter estimation.

$$\hat{\theta} = arg\max_{\theta} \ell(\theta) = arg\max_{\theta} \sum_{i=1}^{n} log\, c\,(u_i, v_i; \theta) \quad (9)$$

The pseudo-observations are calculated by $u_i = r_i\,/(n+1)$ and $v_i = s\,/(n+1)$. Here r_i= $Rank(x_i)$ and $s_i = Rank(y_i)$ are the ranks of observations, n is the number of observations. Parameter estimation was performed with the MPLE method using the 'fitCopula' function in the "copula" package of the R program (Hofert et al., 2024).

Goodness-of-fit Tests: A goodness-of-fit test is a statistical method used to assess how well a model fits the observed data. This test measures the fit of the model to the data by analyzing the differences between the observed and expected distributions from the model. If the probability distribution of the population is unknown, it may be difficult to accurately represent the population with a traditional probability distribution. Therefore, it is a logical approach to select an appropriate distribution using a large amount of information through various techniques (Shin et al., 2010).

In this context, KS (Kolmogorov-Smirnov), CvM (Cramer-von Mises) and AD (Anderson Darling) tests are widely used. In this study, the CvM test, one of the goodness of fit tests, was used and the gofCopula test of the “copula“ package in the R program implemented with the function (Hofert et al., 2024).

The CvM test is a method that measures the fit between two distributions by evaluating the sum of the squares of the differences between the observed and theoretical distribution functions. To obtain the CvM test statistic, $\{(x_i, y_i)\}_{i=1}^{n}$, the data set is first arranged in order of magnitude. Marginal distributions $F_X(x)$and $F_Y(y)$ empirically estimated and $u_i = \hat{F}_X(x_i)$, $v_i = \hat{F}_Y(y_i)$, $i = 1,2, \dots, n$ so-called pseudo-observations are created.

The CvM test statistic is calculated by integrating the squares of the differences between the empirical copula function $\hat{C}_n(u, v)$ and the theoretical copula function, $C(u, v; \hat{\theta})$ obtained for the pseudo-observations (u_i, v_i) over the entire data set, with the formula $W^2 = n \int_0^1 \int_0^1 \left(C_n(u, v) - C(u, v; \theta)\right)^2 dC(u, v; \theta)$. However, since it is usually difficult to calculate in closed form, the following approximate formula is used for n observations:

$$W^2 = n \cdot \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\hat{C}_n(u_i, v_j) - C(u_i, v_j; \hat{\theta})\right)^2 \quad (10)$$

Model Selection: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are widely used in the literature to determine the most appropriate copula model for the data. In addition, the Copula Information Criterion (CIC) developed by Grønneberg et al. (2014) based on the k-fold cross-validation method is used to select the most appropriate copula model (Kwon and Yoon, 2017). The model with the lowest AIC and BIC values or the highest CIC value is considered the best fit to the data.

In this study, AIC, BIC and CIC criteria were used for the selection of copula models. In the model selection process, the calculation methods of each criterion defined by Akaike (1974), Schwarz (1978) and Grønneberg et al. (2014) are given below, respectively:

$$AIC = -2\mathcal{L} + 2k \tag{11}$$

$$BIC = -2\mathcal{L} + k \log n \tag{12}$$

Here, $\mathcal{L}$ represents the log likelihood value calculated by $\mathcal{L} = \sum_{i=1}^{n} log\ c\left(u_i, v_i, \hat{\theta}\right)$, k represents the number of parameters in the model, and n represents the sample size.

CIC is the most appropriate copula selection method calculated using the k-fold cross-validation method. In this method, the data set is divided into k equal-sized subsets. Each layer is used as the test set in turn, while the remaining $k-1$ layer is used as the training set. The copula model is trained on the training set and the log likelihood value is calculated on the test set. The CIC value is obtained by averaging the log likelihood values obtained for all layers:

$$CIC = -2 \cdot \frac{1}{k} \sum_{j=1}^{k} \mathcal{L}_j + \gamma\left(\hat{\theta}\right) \tag{13}$$

Here, $\mathcal{L}_j$ represents the maximum likelihood value calculated for each layer, and $\gamma\left(\hat{\theta}\right)$ represents the penalty term depending on the complexity of the model. In our study, these calculations were made using the fitCopula (for AIC and BIC) and xvCopula (for CIC) functions of the "copula" package in the R program (Hofert et al., 2024).

PISA Dataset and Models

In the study, the 2018 PISA data from (URL-1) which includes the Mathematics, Reading and Science achievement scores of male and female students classified according to their home education levels, were used. Educational opportunities at home include parental education level, parental involvement in homework, and physical resources for homework. These physical resources can be concretized as the student's own room, books, computer, and internet connection. In addition, the family's level of education, occupation and support for the student can be mentioned as examples of intangible resources. All these resources are expressed by the index of educational opportunities at home. In the current study, the level of educational opportunities at home was analyzed through three subgroups. For example, 2018 PISA data are grouped into the following index ranges:

- Low level group; -4.75 / -2.25
- Intermediate group; -2.24 / 0.25
- High level group; 0.26 / 2.76

For data pairs consisting of PISA mathematics, reading and science scores (1: Mathematics-Reading, 2: Mathematics-Science, 3: Reading-Science), models were created according to variables such as home education levels (1: Low, 2: Medium, 3: High) and gender (M: Male, F: Female) and these model definitions are given in Table 1. Descriptive statistics for each model are presented in Table 2 for male and female students. Table 3 shows the Pearson, Spearman and Kendall correlation values and the symmetry test results. Marginal distribution goodness of fit tests was performed for the mathematics, reading and science variables in each model and the results for male and female students are presented in Tables 4 and 5, respectively.

Table 1
Model definitions

Gender	Education Level	Couple	Model
Male	Low	Math-Reading	Model M11
		Math-Science	Model M12
		Reading-Science	Model M13
	Middle	Math-Reading	Model M21
		Math-Science	Model M22
		Reading-Science	Model M23
	High	Math-Reading	Model M31
		Math-Science	Model M32
		Reading-Science	Model M33
Female	Low	Math-Reading	Model F11
		Math-Science	Model F12
		Reading-Science	Model F13
	Middle	Math-Reading	Model F21
		Math-Science	Model F22
		Reading-Science	Model F23
	High	Math-Reading	Model F31
		Math-Science	Model F32
		Reading-Science	Model F33

Table 2 presents the descriptive statistics of male and female students' math, reading and science achievement by gender and educational level. In general, for both genders, it was observed that achievement averages increased as the level of education increased. While the distribution of achievement was generally symmetrical for male students at lower levels of education, the distribution of achievement was skewed to the right for female students, especially at lower levels of education, meaning that higher achievements were observed less frequently. Female students have higher averages than male students in reading and science achievement at all levels of education, and skewness and kurtosis values are generally close to symmetry. In addition, it is noteworthy that for both genders, achievements show a more homogeneous distribution as the level of education increases, i.e. the standard deviation decreases.

Table 2
Descriptive statistic

Gender	Education Level	Success	N	Mean	Median	Mode	Std. Deviation	Skewness	Kurtosis
Male	Low	Maths	153	365.3851	376.5970	151.8840	99.2378	-0.0397	-0.5918
		Reading	153	373.6318	366.3450	198.9440	86.9523	0.3074	-0.5166
		Science	153	399.9808	395.8160	218.3790	85.2644	0.1899	-0.3253
	Middle	Maths	1635	446.7775	446.9170	470.1040	84.2605	0.1052	-0.3052
		Reading	1635	441.3712	441.5230	334.9190	85.5481	0.0447	-0.4112
		Science	1635	453.7162	451.1460	345.5830	81.6772	0.1292	-0.3308
	High	Maths	1706	470.5440	467.2435	396.7150	86.8013	0.2305	-0.2555
		Reading	1706	466.5770	464.0395	489.7070	88.7855	0.1343	-0.3960
		Science	1706	476.8362	472.8860	454.6140	84.7532	0.1374	-0.2948
Female	Low	Maths	101	358.3634	360.5450	198.9500	87.9623	0.0004	-0.8263
		Reading	101	399.1779	393.6810	264.3040	71.6977	0.3299	-0.4532
		Science	101	412.4590	398.4060	244.1080	78.4219	0.3445	-0.7898
	Middle	Maths	1700	441.6859	435.2780	389.0830	79.2538	0.3627	0.1074
		Reading	1700	467.9501	465.8625	432.6020	78.8890	0.1209	-0.2779
		Science	1700	461.2697	458.4480	490.3950	76.2567	0.1412	-0.2752
	High	Maths	1595	465.7769	462.6160	385.5930	87.0589	0.1939	-0.1907
		Reading	1595	493.9962	493.5980	412.7390	84.8008	-0.0265	-0.1984

	Science	1595	488.1876	489.7940	417.9890	81.6587	-0.0608	-0.3811

The relationships between male and female students' mathematics, reading and science achievements were analyzed using Pearson, Spearman and Kendall correlation methods and symmetry tests were performed using the ' exchTest ' function in the "copula" package in the R software. According to the results given in Table 3, strong positive relationships were found in both genders. Especially for female students, the relationship between reading and science achievements was found to be quite high. It was observed that the relationships between achievements became stronger as the level of education increased. The p values of all correlation tests were below 0.05 and these relationships were found to be statistically significant. In addition, according to the symmetry test results, the models support the symmetric distribution assumption (p>0.05).

Table 3
Correlation values and symmetry test values

Models	Pearson *	Spearman *	Kendall *	Symmetry Tests (p- value **)
Model M11	0.8280	0.8412	0.6483	0.012218 (0.6798)
Model M12	0.8417	0.8435	0.6488	0.007646 (0.9745)
Model M13	0.8323	0.8435	0.6452	0.018625 (0.2243)
Model M21	0.8094	0.8070	0.6082	0.010212 (0.9316)
Model M22	0.8260	0.8244	0.6271	0.021749 (0.0854)
Model M23	0.8669	0.8684	0.6790	0.015981 (0.1863)
Model M31	0.8191	0.8169	0.6214	0.018632 (0.2133)
Model M32	0.8372	0.8328	0.6379	0.010334 (0.8776)
Model M33	0.8687	0.8639	0.6763	0.008717 (0.9436)
Model F11	0.7947	0.8083	0.6040	0.013136 (0.6748)
Model F12	0.7652	0.7534	0.5489	0.015293 (0.6259)
Model F13	0.8549	0.8545	0.6749	0.020782 (0.1084)
Model F21	0.7975	0.7887	0.5928	0.016194 (0.4201)
Model F22	0.8210	0.8156	0.6185	0.010656 (0.9006)
Model F23	0.8524	0.8480	0.6554	0.010287 (0.8846)
Model F31	0.8191	0.8184	0.6226	0.014972 (0.5230)
Model F32	0.8373	0.8380	0.6442	0.010941 (0.8107)
Model F33	0.8741	0.8696	0.6836	0.018021 (0.0774)

(*) All correlations are significant($p < 0.05$), (**) all models are symmetric ($p > 0.05$).

Table 4, and Table 5 present the goodness of fit results of the marginal distribution for male and female students' achievement. Here, the best-fitting marginal distributions for each category are marked with (*).

Table 4
Marginal distribution goodness of fit results for male students

Education Level	Success	Marginal Distribution	KS	p value	AIC	Par 1	Par 2
Low	Math	Normal	0.0835	0.2361	1844.0327	365.3851	98.9129
		Log-Normal	0.1021	0.0824	1857.1806	5.8603	0.2943
		Gamma	0.0869	0.1983	1849.4386	12.5384	0.0343
		Weibull *	**0.0769**	**0.3260**	**1841.9317**	4.1684	402.5597
	Read	Normal	**0.0540**	**0.7636**	1803.5920	373.6318	86.6677
		Log-Normal	0.0694	0.4535	1800.2014	5.8959	0.2358
		Gamma*	0.0627	0.5846	**1799.1500**	18.7380	0.0501
		Weibull	0.0653	0.5308	1808.4101	4.6584	408.1857
	Science	Normal *	**0.0475**	**0.8804**	1797.5934	399.9808	84.9853
		Log-Normal	0.0616	0.6077	1799.6863	5.9681	0.2190

		Gamma *	0.0581	0.6793	**1796.8978**	21.3366	0.0535
		Weibull	0.0706	0.4303	1802.8045	5.1087	434.4287
Middle	Math	Normal *	**0.0170**	**0.7310**	**19141.8254**	446.7775	84.2347
		Log-Normal	0.0428	0.0050	19180.9615	6.0836	0.1944
		Gamma	0.0327	0.0601	19149.8782	27.1391	0.0608
		Weibull	0.0401	0.0105	19203.0397	5.7729	481.6767
	Read	Normal *	**0.0192**	**0.5811**	**19191.4163**	441.3712	85.5219
		Log-Normal	0.0425	0.0054	19254.4915	6.0702	0.2015
		Gamma	0.0316	0.0763	19213.5396	25.6945	0.0581
		Weibull	0.0319	0.0715	19230.8552	5.6932	476.5935
	Science	Normal *	**0.0279**	**0.1576**	**19040.0030**	453.7162	81.6522
		Log-Normal	0.0293	0.1208	19069.4545	6.1008	0.1846
		Gamma	0.0806	0.0000	19258.5373	48.2940	0.1062
		Weibull	0.0406	0.0090	19112.4624	6.0264	487.9299
High	Math	Normal	0.0243	0.2643	20074.2931	470.5440	86.7758
		Log-Normal	0.0268	0.1732	20073.2200	6.1366	0.1876
		Gamma *	**0.0165**	**0.7410**	**20055.7739**	28.7082	0.0611
		Weibull	0.0464	0.0013	20171.7918	5.7964	506.8684
	Read	Normal *	**0.0260**	**0.1975**	**20151.4134**	466.5770	88.7595
		Log-Normal	0.0304	0.0856	20181.6387	6.1267	0.1955
		Gamma	0.0804	0.0000	20427.1537	45.1324	0.0967
		Weibull	0.0439	0.0028	20216.3691	5.7242	503.4106
	Science	Normal *	**0.0259**	**0.2015**	**19992.8236**	476.8362	84.7284
		Log-Normal	0.0268	0.1733	20020.3232	6.1509	0.1821
		Gamma	0.0666	0.0000	20216.8212	49.5142	0.1039
		Weibull	0.0440	0.0027	20073.8805	6.0859	512.4590

Table 5

Marginal distribution goodness of fit results for female students

Education Level	Success	Marginal Distribution	KS	p value	AIC	Par 1	Par2
Low	Math	Normal	0.0986	0.2797	1193.9560	358.3634	87.5257
		Log-Normal	0.1039	0.2256	1198.0386	5.8497	0.2573
		Gamma	0.0994	0.2710	1195.2476	15.4475	0.0431
		Weibull *	**0.0902**	**0.3835**	**1193.0293**	4.6249	392.4949
	Read	Normal	0.0837	0.4781	1152.6573	399.1779	71.3419
		Log-Normal *	**0.0482**	**0.9730**	**1149.9515**	5.9734	0.1792
		Gamma	0.0575	0.8924	1149.9803	31.1312	0.0781
		Weibull	0.1062	0.2046	1159.4555	5.9500	429.4407
	Science	Normal	0.1385	0.0414	1170.7654	412.4590	78.0327
		Log-Normal *	**0.1118**	**0.1603**	**1166.7381**	6.0043	0.1888
		Gamma	0.1200	0.1090	1167.2267	28.3195	0.0687
		Weibull	0.1415	0.0350	1176.3761	5.6768	445.2812
Middle	Math	Normal	0.0359	0.0253	19694.4179	441.6859	79.2305
		Log-Normal *	**0.0147**	**0.8536**	**19660.1867**	6.0745	0.1805
		Gamma	0.0542	0.0001	19788.1485	45.1288	0.1021
		Weibull	0.0585	0.0000	19839.7410	5.7899	475.1238
	Read	Normal *	**0.0196**	**0.5339**	**19678.7320**	467.9501	78.8658
		Log-Normal	0.0287	0.1221	19706.3077	6.1338	0.1724
		Gamma	0.0514	0.0003	19777.1576	47.2835	0.1010
		Weibull	0.0438	0.0029	19764.9994	6.4092	501.4310
	Science	Normal *	**0.0224**	**0.3607**	**19563.3487**	461.2697	76.2343
		Log-Normal	0.0251	0.2325	19581.8719	6.1200	0.1685
		Gamma	0.0575	0.0000	19674.4995	50.6932	0.1098

		Weibull	0.0496	0.0005	19658.6931	6.5032	493.7424
High	Math	Normal	0.0223	0.4086	18777.8205	465.7769	87.0316
		Log-Normal	0.0283	0.1540	18792.2675	6.1258	0.1911
		Gamma *	**0.0204**	**0.5235**	**18769.9071**	28.3919	0.0611
		Weibull	0.0427	0.0059	18861.8263	5.7373	502.0037
	Read	Normal *	**0.0181**	**0.6715**	**18693.9868**	493.9962	84.7742
		Log-Normal	0.0419	0.0074	18769.8000	6.1871	0.1785
		Gamma	0.0298	0.1186	18728.3738	32.8852	0.0665
		Weibull	0.0338	0.0528	18739.6598	6.4171	529.5292
	Science	Normal *	**0.0191**	**0.6028**	**18573.5433**	488.1876	81.6331
		Log-Normal	0.0445	0.0036	18647.5229	6.1761	0.1737
		Gamma	0.0346	0.0442	18608.4186	34.5749	0.0708
		Weibull	0.0324	0.0709	18603.9610	6.6643	522.6050

Copula Analysis of PISA Data

In this part of the study, it was aimed to determine the joint distribution functions of the dependency structures of 2018 PISA mathematics, reading and science scores according to demographic variables such as gender and education level at home by using copula functions for each model created in Table 1. Appropriate copula selections were made by considering the smallest Cramer von Mises (CvM), smallest AIC, BIC and largest CIC values and the results are given in Table 6. Here, the best-fitting copulas for each model are marked with (*).

Table 6

Selection of the best copula model.

		Goodness of Fit Test		Information Criteria		
Model	Copula	CvM	p value	AIC	BIC	CIC
Model M11	Normal	0.0309	0.0235	-160.5268	-157.4964	78.6394
	t	0.0347	0.0125	-160.0805	-157.0500	77.3422
	Clayton	0.0759	0.0005	-101.7045	-98.6741	8.9800
	Frank*	**0.0167**	0.2223	**-173.6919**	**-170.6615**	**84.7366**
	Gumbel	0.0354	0.0235	-160.6966	-157.6661	79.1489
	Joe	0.1283	0.0005	-138.0549	-135.0245	61.8950
Model M12	Normal*	**0.0143**	0.4461	**-177.0821**	**-174.0517**	**87.4671**
	t	0.0282	0.0395	-165.0112	-161.9808	77.2770
	Clayton	0.0528	0.0085	-113.5975	-110.5670	24.5576
	Frank	0.0162	0.2702	-176.3328	-173.3024	84.9246
	Gumbel	0.0305	0.0385	-167.5535	-164.5230	78.9948
	Joe	0.1303	0.0005	-139.6985	-136.6681	64.6237
Model M13	Normal	0.0378	0.0115	-162.0242	-158.9937	78.4387
	t	0.0494	0.0015	-157.1345	-154.1040	76.5346
	Clayton	0.0911	0.0005	-96.2035	-93.1731	2.0860
	Frank*	**0.0207**	0.1204	**-171.3872**	**-168.3567**	**86.0487**
	Gumbel	0.0328	0.0355	-166.7235	-163.6931	82.6460
	Joe	0.0916	0.0005	-149.1639	-146.1335	72.6009
Model M21	Normal*	**0.0319**	0.0225	**-1704.2463**	**-1698.8469**	**846.9480**
	t	0.1139	0.0005	-1612.1597	-1606.7603	803.9543
	Clayton	0.5274	0.0005	-1154.9323	-1149.5329	359.7287
	Frank	0.0631	0.0005	-1632.4727	-1627.0733	814.6000
	Gumbel	0.1345	0.0005	-1603.3696	-1597.9702	799.0946
	Joe	0.9881	0.0005	-1313.1456	-1307.7462	649.9319
Model M22	Normal*	**0.0436**	0.0035	**-1823.8232**	**-1818.4238**	**908.8047**
	t	0.1310	0.0005	-1732.7807	-1727.3813	864.9755
	Clayton	0.5309	0.0005	-1201.8948	-1196.4954	326.8904
	Frank	0.0598	0.0005	-1758.4100	-1753.0106	879.3316
	Gumbel	0.1323	0.0005	-1747.1018	-1741.7024	870.3753

	Joe	0.9494	0.0005	-1456.9277	-1451.5283	726.1706
Model M23	Normal*	**0.0402**	0.0025	**-2219.2454**	**-2213.8460**	**1100.8888**
	t	0.1061	0.0005	-2142.4721	-2137.0727	1065.1773
	Clayton	0.4923	0.0005	-1531.4791	-1526.0797	458.0981
	Frank	0.0644	0.0005	-2160.5108	-2155.1114	1083.6490
	Gumbel	0.1264	0.0005	-2114.8745	-2109.4751	1057.1608
	Joe	1.0130	0.0005	-1742.2796	-1736.8802	859.4011
Model M31	Normal*	**0.0437**	0.0025	**-1836.4099**	**-1830.9679**	**916.3304**
	t	0.1230	0.0005	-1752.5438	-1747.1019	871.7316
	Clayton	0.5844	0.0005	-1239.5138	-1234.0719	353.8382
	Frank	0.0750	0.0005	-1793.9180	-1788.4761	895.0620
	Gumbel	0.1397	0.0005	-1732.8884	-1727.4465	860.4503
	Joe	1.0547	0.0005	-1419.0104	-1413.5685	707.7315
Model M32	Normal*	**0.0734**	0.0005	**-1980.6022**	**-1975.1603**	**985.0178**
	t	0.1606	0.0005	-1872.5831	-1867.1412	931.3925
	Clayton	0.7262	0.0005	-1254.5368	-1249.0949	282.3220
	Frank	0.1011	0.0005	-1914.4780	-1909.0361	956.8899
	Gumbel	0.0886	0.0005	-1916.6963	-1911.2544	951.8412
	Joe	0.7803	0.0005	-1625.6206	-1620.1787	807.3084
Model M33	Normal*	**0.0501**	0.0015	**-2342.8707**	**-2337.4288**	**1166.8050**
	t	0.1079	0.0005	-2256.5547	-2251.1128	1120.1798
	Clayton	0.6404	0.0005	-1602.5468	-1597.1049	488.8919
	Frank	0.1047	0.0005	-2227.4291	-2221.9872	1113.0263
	Gumbel	0.0723	0.0005	-2252.1064	-2246.6645	1122.4463
	Joe	0.7975	0.0005	-1885.9234	-1880.4815	938.5881

Table 6 (Continued) Selection of the best copula model

		Goodness of Fit Test		Information Criteria		
Model	Copula	CvM	p value	AIC	BIC	CIC
Model F11	Normal	0.0279	0.1064	-94.3829	-91.7678	45.8349
	t	0.0395	0.0305	-88.1103	-85.4952	42.8937
	Clayton	0.0673	0.0035	-68.5421	-65.9269	11.8895
	Frank*	**0.0233**	0.1234	**-96.1168**	**-93.5017**	**48.8754**
	Gumbel	0.0372	0.0485	-85.6661	-83.0510	43.2265
	Joe	0.1033	0.0015	-68.3828	-65.7677	26.4905
Model F12	Normal*	**0.0180**	0.3881	**-81.4144**	**-78.7993**	**39.4696**
	t	0.0339	0.0794	-73.4414	-70.8262	34.2042
	Clayton	0.0517	0.0155	-55.2510	-52.6358	10.5081
	Frank	0.0157	0.4890	-76.5508	-73.9357	37.5761
	Gumbel	0.0292	0.1134	-74.9429	-72.3278	36.6713
	Joe	0.0842	0.0025	-61.7298	-59.1147	29.9567
Model F13	Normal	0.0294	0.0514	-123.6421	-121.0270	60.3359
	t	0.0347	0.0255	-122.1683	-119.5532	58.7985
	Clayton	0.0486	0.0095	-98.0322	-95.4171	22.8606
	Frank*	**0.0150**	0.3541	**-127.3674**	**-124.7523**	**63.3642**
	Gumbel	0.0410	0.0215	-113.6253	-111.0102	57.0905
	Joe	0.1292	0.0005	-89.1367	-86.5216	41.3561
Model F21	Normal*	**0.0313**	0.0195	**-1684.5886**	**-1679.1502**	**843.0044**
	t	0.0860	0.0005	-1631.2507	-1625.8123	810.4349
	Clayton	0.6100	0.0005	-1129.4387	-1124.0003	352.4272
	Frank	0.0829	0.0005	-1589.7045	-1584.2661	794.8592
	Gumbel	0.0823	0.0005	-1648.4645	-1643.0261	822.9915
	Joe	0.7999	0.0005	-1388.1743	-1382.7359	689.1320
Model F22	Normal*	**0.0291**	0.0275	**-1865.3743**	**-1859.9360**	**932.8761**
	t	0.1029	0.0005	-1780.1917	-1774.7533	888.8107
	Clayton	0.5881	0.0005	-1243.1867	-1237.7484	366.7113
	Frank	0.0768	0.0005	-1764.1615	-1758.7231	882.3307
	Gumbel	0.0959	0.0005	-1793.7086	-1788.2702	892.1150
	Joe	0.8738	0.0005	-1500.1892	-1494.7508	746.4985

Model F23	Normal*	**0.0277**	0.0315	**-2169.4330**	**-2163.9946**	**1085.8801**
	t	0.0889	0.0005	-2094.2331	-2088.7947	1043.9665
	Clayton	0.5286	0.0005	-1536.9721	-1531.5338	516.9028
	Frank	0.0765	0.0005	-2045.9703	-2040.5319	1021.3401
	Gumbel	0.1144	0.0005	-2055.0322	-2049.5938	1023.0629
	Joe	0.9980	0.0005	-1689.3811	-1683.9427	838.3634
Model F31	Normal*	**0.0316**	0.0165	**-1766.7548**	**-1761.3802**	**884.1718**
	t	0.1052	0.0005	-1694.9972	-1689.6226	844.0886
	Clayton	0.4469	0.0005	-1298.0983	-1292.7237	471.7959
	Frank	0.0496	0.0005	-1684.2674	-1678.8928	841.1828
	Gumbel	0.1844	0.0005	-1618.2542	-1612.8796	805.9786
	Joe	1.2101	0.0005	-1284.6458	-1279.2712	638.3470
Model F32	Normal*	**0.0581**	0.0005	**-1908.8210**	**-1903.4463**	**948.0684**
	t	0.1266	0.0005	-1830.2134	-1824.8387	908.2109
	Clayton	0.6349	0.0005	-1273.3662	-1267.9915	367.3689
	Frank	0.0817	0.0005	-1835.8852	-1830.5106	916.6340
	Gumbel	0.0862	0.0005	-1838.5946	-1833.2200	912.5179
	Joe	0.7790	0.0005	-1543.1495	-1537.7748	765.9583
Model F33	Normal*	**0.0207**	0.0834	**-2285.3918**	**-2280.0172**	**1138.6953**
	t	0.0680	0.0005	-2208.7831	-2203.4085	1097.5980
	Clayton	0.4367	0.0005	-1700.1341	-1694.7595	630.9208
	Frank	0.0810	0.0005	-2142.4474	-2137.0727	1070.7635
	Gumbel	0.1233	0.0005	-2117.8157	-2112.4410	1053.3135
	Joe	1.0604	0.0005	-1705.0189	-1699.6442	846.1195

The parameter estimates (MPLE) and standard error values for the copulas selected for each model are presented in Table 7.

Table 7
Parameter estimation results for selected copula models

Gender	Education Level	Couple	Model	Copula	Estimate	Standard Error
Male	Low	Math-Reading	Model E11	Frank	9.1478	0.9950
		Math-Science	Model E12	Normal	0.8389	0.0211
		Reading-Science	Model E13	Frank	8.9777	1.0773
	Middle	Math-Reading	Model E21	Normal	0.8062	0.0077
		Math-Science	Model E22	Normal	0.8214	0.0071
		Reading-Science	Model E23	Normal	0.8630	0.0055
	High	Math-Reading	Model E31	Normal	0.8134	0.0066
		Math-Science	Model E32	Normal	0.8301	0.0067
		Reading-Science	Model E33	Normal	0.8653	0.0054
Female	Low	Math-Reading	Model K11	Frank	7.8731	1.1864
		Math-Science	Model K12	Normal	0.7657	0.0456
		Reading-Science	Model K13	Frank	10.4114	1.1583
	Middle	Math-Reading	Model K21	Normal	0.7946	0.0081
		Math-Science	Model K22	Normal	0.8177	0.0073
		Reading-Science	Model K23	Normal	0.8503	0.0064
	High	Math-Reading	Model K31	Normal	0.8199	0.0072
		Math-Science	Model K32	Normal	0.8368	0.0069
		Reading-Science	Model K33	Normal	0.8739	0.0054

Joint and conditional probabilities for PISA achievement scores

In this section, joint and conditional probabilities are calculated for the 2018 PISA Mathematics, Reading and Science achievement scores of male and female students classified according to their education levels at home, using the most appropriate marginal distributions and copula models determined in the previous sections. Here, for each model defined in Table 1, the averages of the scores were taken as threshold values and the probabilities of success above these threshold values were evaluated. The results obtained are presented in Table 8.

The best-fit marginal distribution parameters for male and female students' Mathematics, Reading and Science scores are given in Tables 4 and 5, respectively, and the copula goodness of fit results are given in Table 6. The mean values presented in Table 2 were used as threshold values. In addition, the probability of students who achieved above average success in a particular course to achieve the same success in other courses was analyzed using conditional probability methods using equations (3) and (4).

In order to examine the effect of reading on mathematics and science achievement, some joint and conditional probabilities were analyzed together. Table 8 presents the joint and conditional probabilities calculated based on the selected copula models for mathematics-reading, mathematics-science and reading-science pairs. In the calculation of these probabilities, U (0,1) transformations based on marginal information were applied and the results were expressed with the notations $\bar{C}_v(u)$ and $\bar{C}_u(v)$ defined in equations (6), (7) and (8), respectively.

The results obtained show how the probabilities of students' succeeding in certain courses can change with conditional probabilities. For example, in model M31, the probability of being successful in both math and reading is 38.67%. In comparison, the probability that a student who is successful in reading is also successful in math is calculated as 77.34% using the conditional probability information $\bar{C}_v(u)$. The F31 model's probabilities are 38.74% and 77.47%, respectively.

Table 8
Joint and conditional probabilities for selected copula models.

Models	Selected copulas	Estimate	Threshold values		$\bar{C}(u,v)$	$\bar{C}_v(u)$	$\bar{C}_u(v)$
Model M11	Frank	9.1478	$\bar{x}_m = 365.3851$	$\bar{x}_r = 373.6318$	0.4153	**0.8817**	0.8097
Model M12	Normal	0.8389	$\bar{x}_m = 365.3851$	$\bar{x}_s = 399.9808$	0.3959	0.8499	0.7720
Model M13	Frank	8.9777	$\bar{x}_r = 373.6318$	$\bar{x}_s = 399.9808$	0.3925	0.8425	**0.8333**
Model M21	Normal	0.8062	$\bar{x}_m = 446.9170$	$\bar{x}_r = 441.5230$	0.3986	**0.7982**	0.7982
Model M22	Normal	0.8214	$\bar{x}_m = 446.9170$	$\bar{x}_s = 451.1460$	0.4092	0.7984	0.8195
Model M23	Normal	0.8630	$\bar{x}_r = 441.5230$	$\bar{x}_s = 451.1460$	0.4215	0.8223	**0.8442**
Model M31	Normal	0.8134	$\bar{x}_m = 470.5440$	$\bar{x}_r = 466.5770$	**0.3867**	**0.7734**	0.8191
Model M32	Normal	0.8301	$\bar{x}_m = 470.5440$	$\bar{x}_s = 476.8362$	0.3913	0.7826	0.8289
Model M33	Normal	0.8653	$\bar{x}_r = 466.5770$	$\bar{x}_s = 476.8362$	**0.4164**	0.8329	**0.8329**
Model F11	Frank	7.8731	$\bar{x}_m = 358.3634$	$\bar{x}_r = 399.1779$	0.4030	**0.8678**	0.7771
Model F12	Normal	0.7657	$\bar{x}_m = 358.3634$	$\bar{x}_s = 412.4590$	0.3772	0.8157	0.7272
Model F13	Frank	10.4114	$\bar{x}_r = 399.1779$	$\bar{x}_s = 412.4590$	0.3974	0.8594	**0.8556**
Model F21	Normal	0.7946	$\bar{x}_m = 441.6859$	$\bar{x}_r = 467.9501$	0.3776	**0.7551**	0.8129
Model F22	Normal	0.8177	$\bar{x}_m = 441.6859$	$\bar{x}_s = 461.2697$	0.3837	0.7674	0.8261
Model F23	Normal	0.8503	$\bar{x}_r = 467.9501$	$\bar{x}_s = 461.2697$	0.4118	0.8236	**0.8236**
Model F31	Normal	0.8199	$\bar{x}_m = 465.7769$	$\bar{x}_r = 493.5980$	**0.3874**	**0.7747**	0.8241
Model F32	Normal	0.8368	$\bar{x}_m = 465.7769$	$\bar{x}_s = 489.7940$	0.3921	0.7842	0.8342
Model F33	Normal	0.8739	$\bar{x}_r = 493.5980$	$\bar{x}_s = 489.7940$	**0.4192**	0.8384	**0.8384**

Similarly, in model M33, the probability of being successful in both reading and science is 41.64%, whereas the probability of a student who is successful in reading is also successful in science, 83.29% was obtained with conditional probability information $\bar{C}_u(v)$. A similar situation was observed in the F33 model, where the probabilities were calculated as 41.92% and 83.84%, respectively.

According to these results, female students with higher levels of education at home have higher joint and conditional probabilities of success than male students Especially when the reading achievement of girls with higher levels of education is considered, they are more likely than boys to be successful in both math and science. However, this situation shows a reverse trend at other education levels.

In general, the results show that students' success in one subject can positively affect their probability of success in other subjects. In other words, a student who is known to be successful in one subject is more likely to be successful in other subjects, while the probability of being successful in both subjects is lower. While these probabilities are expected to increase as the level of education at home increased, in some cases it has been observed that a student who is known to be successful in reading is less likely to be successful in math or science. These findings suggest that home education may have a significant effect on student achievement, but that this effect is not constant and consistent in all cases.

The study allows for an examination of the dependency structures between achievement scores stratified by educational levels at home and a more detailed assessment of the probabilities of students achieving above average. It also reveals how these probabilities change according to different levels of education. In this context, dependency models and copula functions allow not only to calculate the probabilities of above-average achievement, but also to analyze how these probabilities change conditionally.

Conclusion

This study examined how the dependency structures between students' achievement in mathematics, reading, and science vary with demographic factors such as gender and level of education at home, using 2018 PISA data. Analyses with copula functions identified the most appropriate models to accurately predict achievement probabilities.

In particular, the dependency between PISA achievement scores of male and female students according to their level of education at home was examined using copula functions. First, various models were created for PISA achievement scores according to these demographic characteristics and the best copulas were determined for each model created with the CvM goodness of fit test, and the selected functions were supported by AIC, BIC, and CIC criteria. Then, some joint probabilities were obtained using the parameter estimates made with the pseudo-maximum likelihood (MPLE) method. Here, the averages for each model were taken as thresholds and the probabilities of the achievement scores exceeding these thresholds were considered as success. In addition, conditional probabilities were calculated to see the effect of reading on math and science courses.

The results show that the level of education at home has a partial effect on students' likelihood of success, but this effect is not consistent in all cases. Female students with higher levels of education were found to be more likely than male students to be successful in mathematics based on their success in reading.

As a result, the findings of the study show that copula models are an effective tool in understanding the dependent factors affecting student achievement and can be used effectively in educational analyses. These analyses constitute an important basis for examining the effects of other demographic variables, other than home education, on achievement and for better guiding students' educational processes.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Anıl, D. (2009). Uluslararası öğrenci başarılarını değerlendirme programı (PISA)'nda Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler. *Eğitim ve Bilim*, *34*(152), 87-100.

Baykal, T. (2024). Joint frequency analysis of streamflow and sediment amount with copula functions in the Kızılırmak Basin, Turkey. *Natural Hazards*, 1-20.

Bhatti, M. I., & Do, H. Q. (2019). Recent development in copula and its applications to the energy, forestry and environmental sciences. International *Journal of Hydrogen Energy*, *44*(36), 19453-19473.

Cherubini U, Luciano E, Vecchiato W (2004). Copula methods in finance. *John Wiley and Sons*, 289, New York.

Embrechts, P., McNeil, A., & Straumann, D. (2003). Copulas: A Review. *Journal of Multivariate Analysis*, *88*(2), 327-344.

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water resources research*, *40*(1).

Genest, C., Ghoudi, K. and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Printed in Great Britain*, *82*, 543-552.

Genest, C. Favre, A-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, *12*(4), 347–368.

Goda, K. (2010). Statistical modeling of joint probability distribution using copula: application to peak and permanent displacement seismic demands. *Structural Safety*, *32*(2), 112-123.

Grønneberg, S., and Hjort, N.L. (2014) The copula information criteria. *Scandinavian Journal of Statistics*, *41*, 436–459.

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2024). Copula: multivariate dependence with copulas, *R package version*, *1*, 1-4, https://CRAN.R-project.org/package=copula

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models, *Department of Statistics*, *University of British Columbia*.

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Kara, E. K., & Kemaloglu, S. A. (2016). Portfolio optimization of dynamic copula models for dependent financial data using change point approach. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, *65*(2), 175-188.

Kizilok Kara, E, & Yıldız, O. (2014). Bivariate analysis of precipitation and runoff in the Hirfanlı Dam Basin, Turkey, using Copulas. *İstatistik: Journal of the Turkish Statistical Association*, *7*(3), 63-70.

Kizilok Kara, E. (2017). The Earthquake Risk Analysis Based on Copula Models for Turkey. Sigma: *Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi*, *35*(2).

Kasap, Y., Doğan, N. & Koçak, C. (2021). PISA 2018'de okuduğunu anlama başarısını yordayan değişkenlerin veri madenciliği ile belirlenmesi. *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, *19*(4), 241-258.

Kızılok Kara, E., Açık Kemaloğlu, S., Evkaya, Ö.O. (2022). Modeling Currency Exchange Data with Asymmetric Copula Functions. In: Terzioğlu, M.K. (Eds), *Advances in Econometrics, Operational Research, Data Science and Actuarial Studies. Contributions to Economics*. Springer, Cham. https://doi.org/10.1007/978-3-030-85254-2_4

Kwon, T., & Yoon, S. (2017). Analysis of extreme wind speed and precipitation using copula. *Journal of the Korean Data and Information Science Society*, *28*(4), 797-810.

MEB, (2013). PISA 2012 Ulusal ön raporu http://pisa.meb.gov.tr/wp-content/uploads/2013/12/pisa2012-ulusal-on-raporu.pdf, (Erişim tarihi: 10 Temmuz 2022).

Nelsen, R.B. (2006). *An Introduction to Copulas*. 2nd edition, Springer, New York.

Nguyen-Huy, T., Deo, R. C., Mushtaq, S., Kath, J., & Khan, S. (2019). Copula statistical models for analyzing stochastic dependencies of systemic drought risk and potential adaptation strategies. *Stochastic Environmental Research and Risk Assessment*, *33*(3), 779-799.

OECD, (2014). PISA 2012 results in focus. https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf, (Erişim tarihi: 14 Temmuz 2016).

Pala, M. ve Sağlam, F. (2019). PISA fen, matematik ve okuma puanları arasındaki bağımlılık yapısının kapula ile modellenmesi. *Selçuk Üniversitesi Fen Fakültesi Fen Dergisi*, *45*(2), 149-162.

Patton, A. (2013). Copula methods for forecasting multivariate time series. *Handbook of economic forecasting*, *2*, 899-960.

Sarıer, Y. (2021). PISA uygulamalarında Türkiye'nin performansı ve öğrenci başarısını yordayan değişkenler. *Türkiye Sosyal Araştırmalar Dergisi*, *25*(3), 905-926.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, *51*, 1384-1399.

Shin, H. J., Sung, K. M., & Heo, J. H. (2010). Derivation of modified Anderson-Darling test statistics and power test for the Gumbel distribution. *Journal of Korea Water Resources Association*, *43*(9), 813-822.

Sklar, A. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Publ Inst Statist Univ*, *8*, 229-231.

Toprak, E. (2017). *Yapay sinir ağı, karar ağaçları ve ayırma analizi yöntemleri ile PISA 2012 matematik başarılarının sınıflandırılma performanslarının karşılaştırılması*. (Doctoral Thesis). Hacettepe University, Ankara

URL-1: http://www.oecd.org/pisa/data/ (Access Date: 03.07.2022).