Group – 9
March 2024
MA324

Copula Structure

Statistical Interference and Multivariate Analysis

Rayala Naga Pragna Sri (210123050)
Pranavi Sudireddy (210123045)
Kodudula Ashish Reddy (210123033)
Kancharla Abhijeet Goud (210123031)

Central corneal thickness in diabetic and non-diabetic patients

We took the reference from an Original Research Article, Corneal thickness and endothelial cell density in diabetic and nondiabetic patients: a hospital based comparative study and the result was given from their study that, the mean Central corneal thickness in diabetic group was significantly higher (518.40±28.13 µm) compared to control group (490.14±24.31 µm) ($p<0.001$).

The data was analyzed and represented using mean±Sd for representing quantitative data. Independent samples 't' test was used for comparing the two groups. In the present study, out of 200 study subjects, 100 subjects were diabetics and 100 subjects were nondiabetics each.

Now, from the data of corneal thickness of 100 diabetic patients (stored in a txt file named x_data.txt) and of 100 non-diabetic patients (y_data.txt file), we are going to proceed with our copula construction.

Marginal Distribution: We create two instances of the 'GuassianKDE' class from the Copulas library to model the marginal distributions of the variables x and y. These kernel density estimators (KDEs) are used to estimate the probability density function of each variable.

Fit Marginal Distributions: We fit the marginal distributions to the observed data (x_data and y_data). This process estimates the parameters of the KDEs to best represent the marginal distributions of the data.

Transform Data: We transform the original data points (x_data and y_data) to their corresponding cumulative distribution function (CDF) values. This transformation maps the original data to a uniform distribution between 0 and 1.

Performed parameter estimation for a **Clayton copula** using **maximum likelihood estimation (MLE)** and then generated synthetic data using the estimated copula model. The negative log-likelihood function is minimized to find the optimal parameter value, and the resulting copula model is used to generate synthetic data points that capture the dependence structure between the variables.

Parameter Estimation of Clayton Copula

In this section, we describe the process of estimating the parameter of a Clayton copula using maximum likelihood estimation (MLE) and different optimization methods. The Clayton copula is a widely used model for capturing dependence between variables, particularly in finance and risk management applications

The Clayton Copula

The Clayton copula is

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

For our applications $0 < \theta < \infty$ so this can be simplified to

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta} \qquad (0,\infty)$$

We began by defining a negative log-likelihood function (neg_log_likelihood) tailored for the Clayton copula. This function takes the parameter theta as input, fits a Clayton copula to the transformed data using the given theta, and computes the negative log-likelihood of the data under the copula model.

Initial Guess for Parameter Theta: We provide an initial guess for the parameter theta. This initial value is used as the starting point for the optimization algorithm to find the maximum likelihood estimate (MLE) of the parameter theta.

Estimate Copula Parameter using MLE: We use the minimize function from the SciPy library to minimize the negative log-likelihood function and estimate the parameter theta.

Obtain Estimated Parameter Theta: We extract the estimated value of the parameter theta from the optimization result and assign it to the variable estimated_theta.

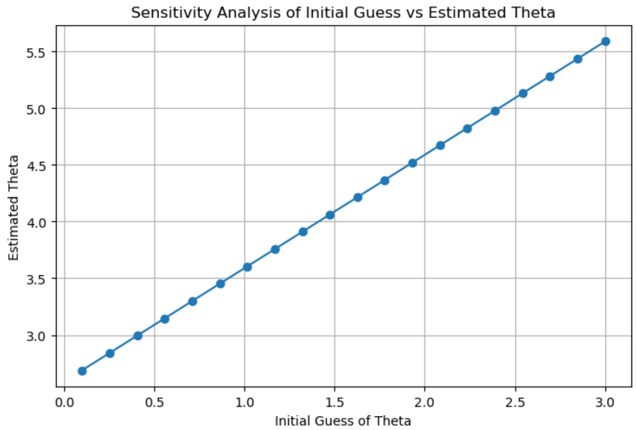
Analysis of Estimated Theta using MLE for different initial guesses of theta:

For initial guess of theta = 1.0
Estimated theta using MLE: 3.58

For initial guess of theta = 2.0
Estimated theta using MLE: 4.58

For initial guess of theta = 0.1
Estimated theta using MLE: 2.68

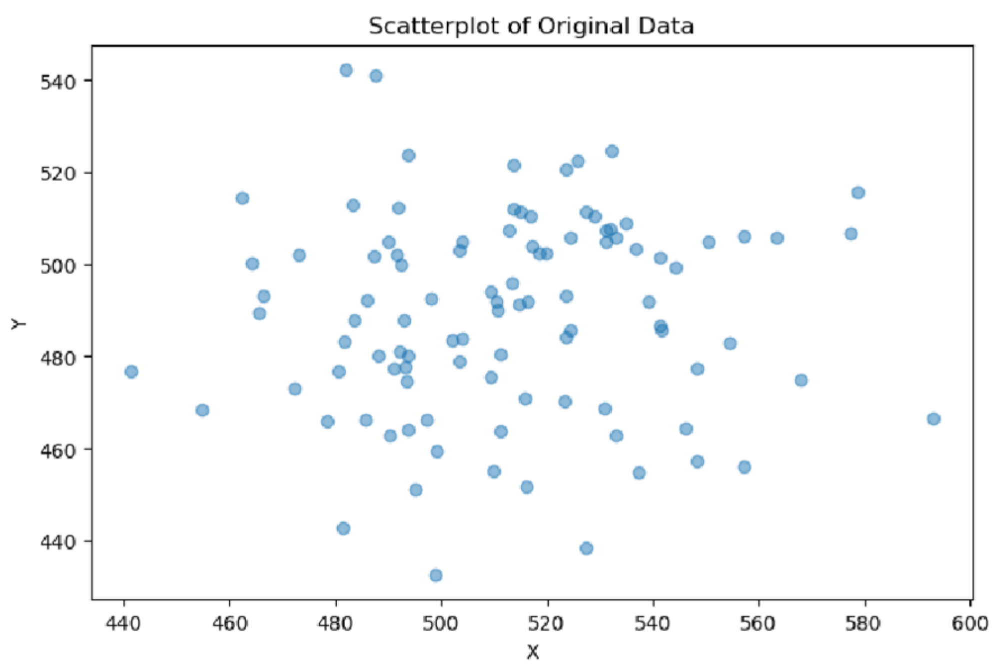
For initial guess of theta = 0.5
Estimated theta using MLE: 3.087

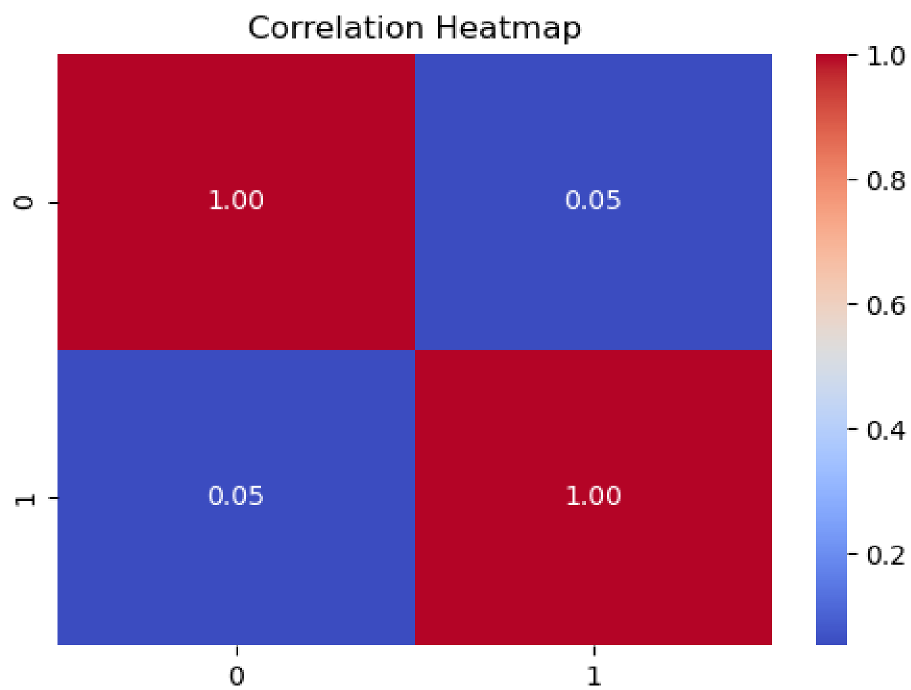


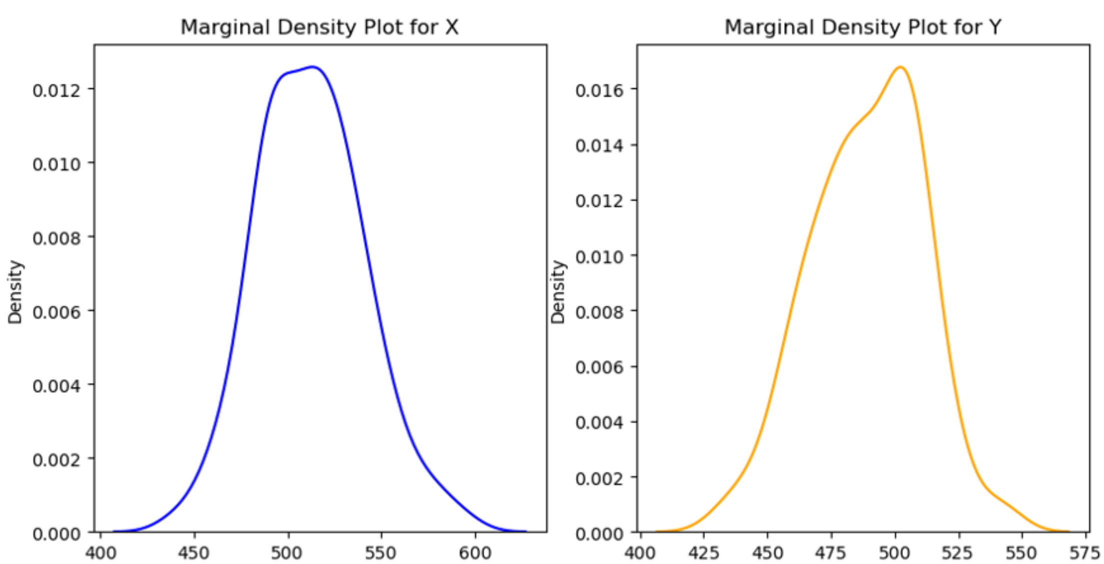
Fit Copula with Estimated Parameter: We create a new Clayton copula object with the estimated parameter theta and fit it to the transformed data (transformed_x and transformed_y).

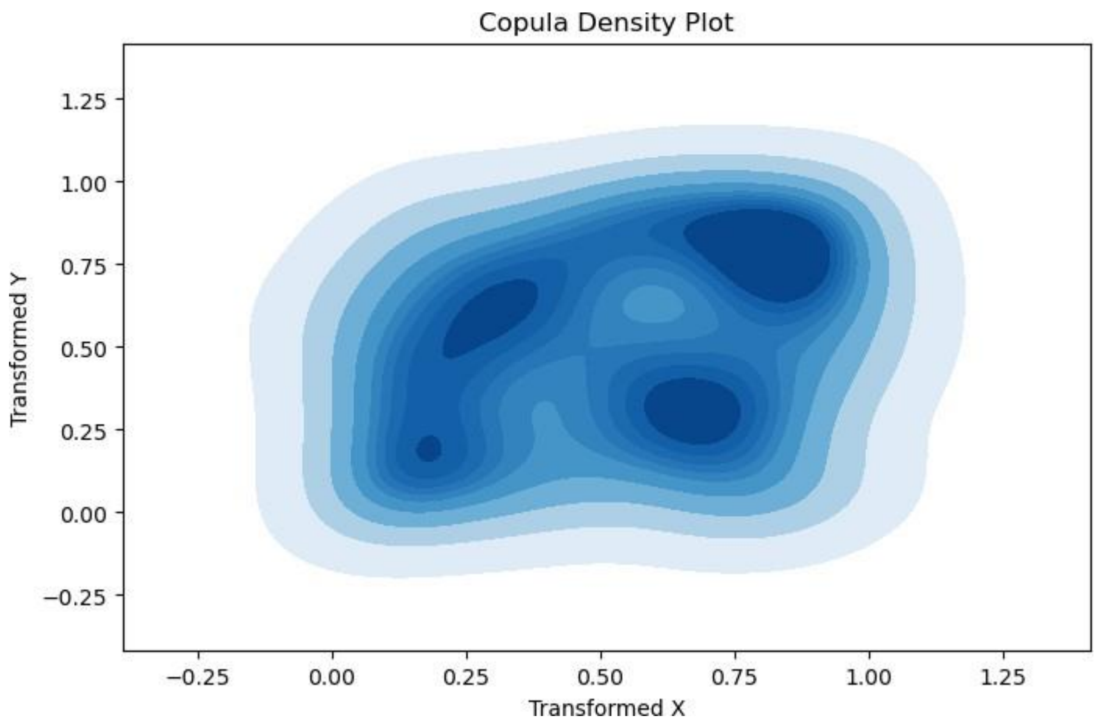
Generate Synthetic Data using Copula: Finally, we use the fitted copula model to generate synthetic data points (synthetic_data) by sampling from the copula distribution.

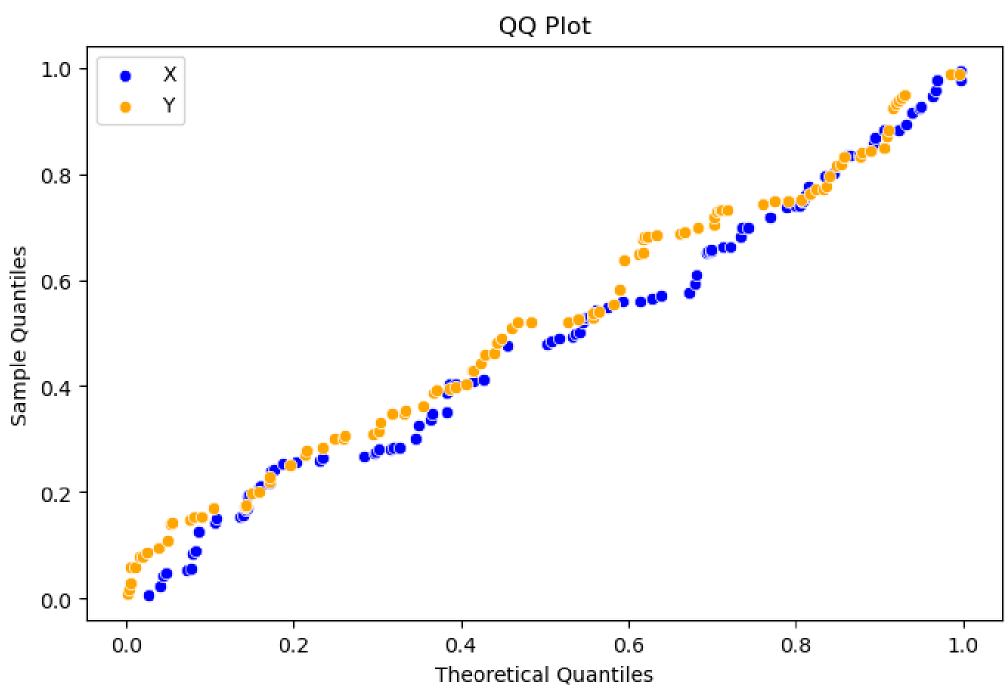
Results:



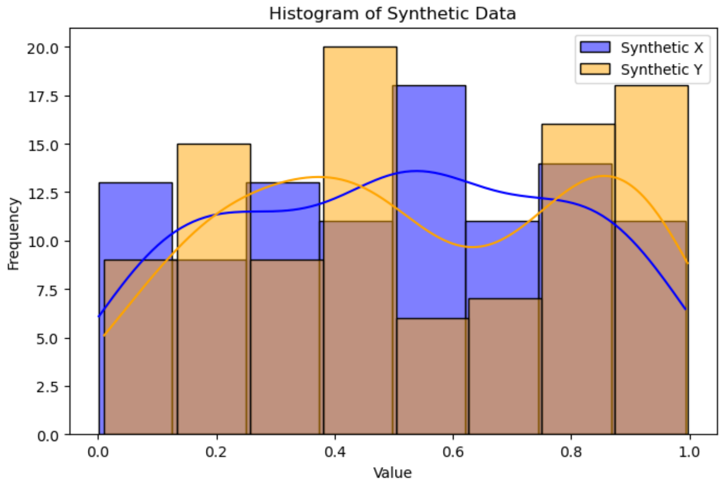

Correlation Heatmap
1.00
0.05
0.05
1.00
0
1
0
1
1.0
0.8
0.6
0.4
0.2


Marginal Density Plot for X
Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
400
450
500
550
600
Marginal Density Plot for Y
Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
400
425
450
475
500
525
550
575

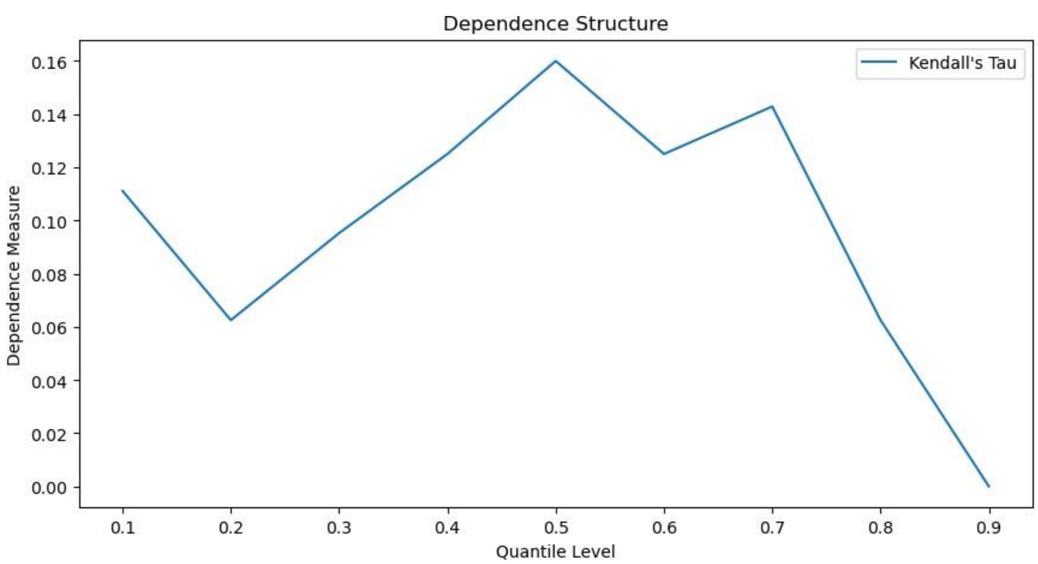

Copula Density Plot
Transformed Y
Transformed X
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.25
0.00
0.25
0.50
0.75
1.00
1.25

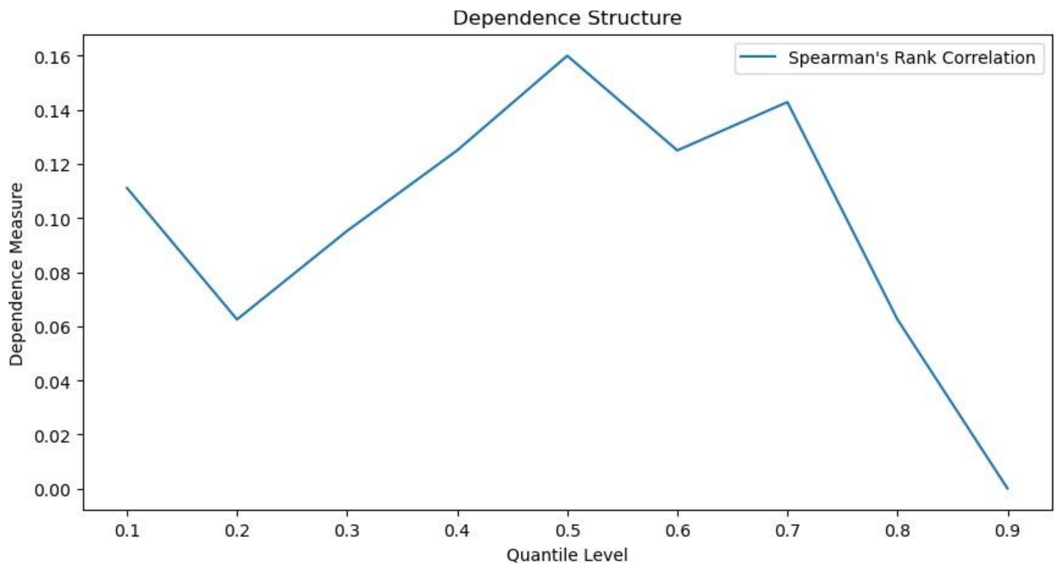

QQ Plot
X
Y
Sample Quantiles
Theoretical Quantiles
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Using synthetic data generated from a Clayton copula,

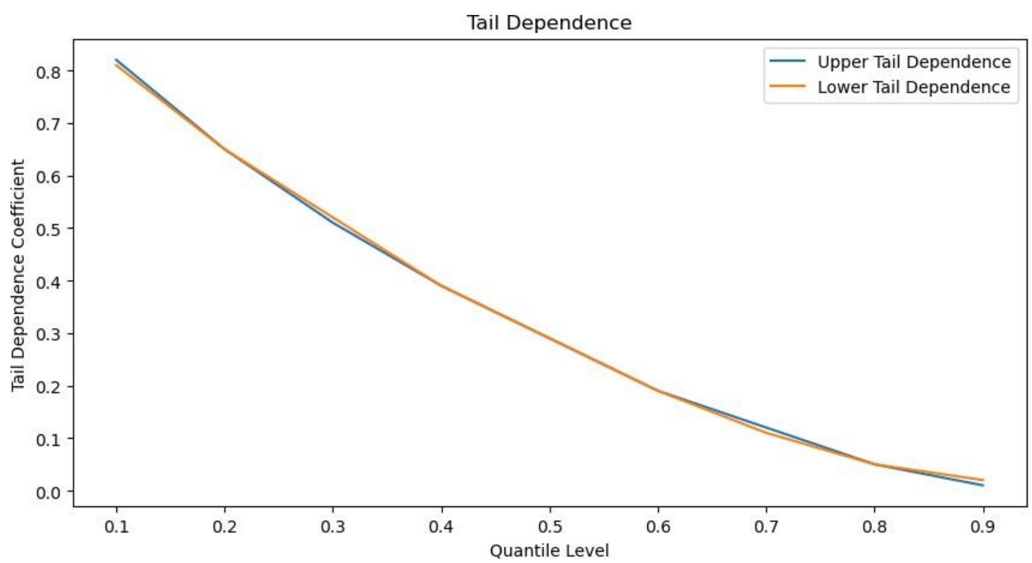


Kendall's Tau: 0.11353535353535356
Spearman's Rank Correlation: 0.18390639063906392
Pearson Correlation Coefficient: 0.17411731903318448
Copula exhibits upper tail dependence
Upper tail dependence coefficient 0.01
Copula exhibits lower tail dependence
Lower tail dependence coefficient 0.02





Dependence Structure
Spearman's Rank Correlation
Dependence Measure
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Quantile Level



Tail Dependence
Upper Tail Dependence
Lower Tail Dependence
Tail Dependence Coefficient
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Quantile Level