

Comparación de la estructura de dependencia entre variables de préstamos estudiantiles en universidades públicas y privadas mediante cópulas

Gabriel Sanabria Alvarado | Carné: C27184
Diego Alberto Vega Víquez | Carné: C38367
Jeikel Navarro Solís | Carné: C25518
Andy Roberto Peralta Duarte | Carné: C25827

2025-09-19

Introducción

El financiamiento de la educación superior ha sido siempre un aspecto vital para el acceso y la permanencia de los estudiantes en las universidades. En particular, los préstamos estudiantiles son una de las principales fuentes de recursos para cubrir los costos de matrícula y manutención, lo que genera un impacto directo en las condiciones financieras de los estudiantes y sus familias. La dinámica de estos préstamos no solo depende de los montos otorgados, también de la interacción entre múltiples variables, tales como la tasa de interés, el plazo, los ingresos familiares, el tipo de institución y las características socioeconómicas del solicitante.

Bajo el contexto de esta investigación, resulta fundamental comprender cómo se relacionan estas variables en distintos sectores de la educación superior, tanto en las universidades públicas como en las privadas. Las dependencias entre variables financieras suelen presentar comportamientos no lineales y colas pesadas, lo que dificulta su caracterización mediante métodos estadísticos tradicionales basados únicamente en correlaciones lineales. Ante este desafío, las cópulas ofrecen una herramienta robusta para modelar y comparar estructuras de dependencia, al permitir describir con mayor precisión la forma en que interactúan las variables sin restringirse a supuestos de linealidad o normalidad.

El *Federal Family Education Loan (FFEL) Program* fue uno de los principales mecanismos de financiamiento de la educación superior en Estados Unidos hasta su eliminación en 2010. Bajo este esquema, prestamistas privados originaban préstamos estudiantiles con garantía federal, lo que implicaba que el gobierno

asumía parte del riesgo crediticio.

Durante el año académico 2009-2010, el Departamento de Educación implementó medidas extraordinarias para sostener el flujo de crédito en el marco de la crisis financiera de 2008-2009. A través de la *Ensuring Continued Access to Student Loans Act (ECASLA)* se activaron programas de compra y participación de préstamos, mediante los cuales el gobierno adquirió directa o indirectamente alrededor de 60 mil millones de dólares en préstamos FFEL hacia octubre de 2010 (U.S. Department of Education 2010).

La base de datos *2009-2010 Award Year FFEL Volume by School – Award Year Quarterly Activity (04/01/2010-06/30/2010)*”, con fecha de procesamiento 4 de mayo de 2012, forma parte de los reportes administrativos que detallan, por institución, los volúmenes de préstamos otorgados bajo el programa FFEL. Estos registros permitían a las autoridades educativas monitorear el nivel de endeudamiento de los estudiantes según el tipo de institución.

En marzo de 2010 se aprobó la *Health Care and Education Reconciliation Act*, que eliminó el FFEL Program y dispuso que a partir del 1º de julio de 2010 todos los nuevos préstamos federales se otorgaran únicamente bajo el *Federal Direct Loan Program*, en el que el financiamiento proviene directamente del gobierno federal (U.S. Department of Education 2010, 2012).

El estudio de la estructura de dependencia mediante cópulas en los préstamos estudiantiles permite identificar diferencias significativas entre universidades públicas y privadas, lo cual aporta información valiosa para la gestión del riesgo crediticio, la formulación de políticas de financiamiento y la evaluación de la equidad en el acceso a la educación. De esta manera, la comparación de ambos sectores no solo tiene un interés académico y metodológico, sino también un impacto social y económico, al ofrecer evidencia empírica sobre los patrones de endeudamiento y las condiciones bajo las cuales los estudiantes acceden a recursos financieros.

Objetivos

Objetivo General

Analizar y comparar las dependencias estadísticas entre variables clave de los préstamos estudiantiles (como el número de préstamos, monto originado, monto desembolsado y número de beneficiarios) en universidades públicas y privadas, utilizando cópulas, con el fin de identificar patrones estructurales diferenciados en la distribución del crédito educativo.

Objetivos Específicos

- Clasificar y caracterizar las instituciones educativas según su tipo (públicas, privadas y sin fines de lucro), y estimar las distribuciones marginales de las variables de interés relacionadas con los préstamos estudiantiles.

- Aplicar y ajustar modelos de cópulas para capturar y describir la estructura de dependencia entre las variables de préstamos estudiantiles, diferenciando por tipo de institución.
- Comparar de manera gráfica y numérica las cópulas ajustadas entre universidades públicas y privadas, evaluando si las diferencias observadas en las dependencias son estadísticamente significativas e implican cambios en el comportamiento del endeudamiento estudiantil.

Documentación de la base de datos

La base de datos proviene del sitio web oficial del Gobierno de los Estados Unidos <https://catalog.data.gov/dataset/national-student-loan-data-system-722b0> y es de acceso público. Contiene datos acerca de préstamos estudiantiles otorgados por distintas instituciones de educación superior en todo el país entre el 1 de abril de 2010 hasta el 30 de julio del mismo año. La base cuenta con 3793 observaciones, o sea, universidades distintas. Se tienen 4 tipos de préstamos, subsidiados y no subsidiados, según si la responsabilidad de pago la asume el estudiante o el padre del estudiante. Se cuenta con las siguientes variables:

- OPE ID: identificador único de la institución dentro de la Oficina de Educación Postsecundaria, OPE por sus siglas en inglés.
- School: Nombre de la universidad.
- State: Identificador de 2 letras del estado en el que se encuentra.
- ZIP: Código del sistema postal de Estados Unidos.
- School Type: Si la institución es pública o privada.

Además cada tipo de préstamo tiene las siguientes variables:

- Recipients: La cantidad de personas que obtuvieron préstamos.
- # of Loans Originated: El número de préstamos dados por la universidad.
- \$ of Loans Originated: El valor total en dólares estadounidenses de los préstamos dados.
- # of Disbursements: Cantidad de desembolsos.
- \$ of Disbursements: Cantidad de dinero desembolsado.

En el presente trabajo la variable School Type será de particular interés para lograr el objetivo.

Análisis Descriptivo de la Base de Datos

El conjunto de datos cuenta con 3,793 instituciones y 25 variables, entre identificadores, características de las instituciones (estado, tipo, código postal) y métricas asociadas a los distintos programas de préstamos (subsidiados, no subsidiados, Parent PLUS y Grad PLUS).

En términos de valores faltantes, únicamente la variable *Zip Code* presenta 19 registros vacíos, lo cual representa un porcentaje mínimo (0.5%) y no compromete la representatividad del análisis. El resto de las variables se encuentra completo, lo que garantiza la solidez de las inferencias estadísticas.

En el resumen estadístico de las variables numéricas se observa una marcada asimetría:

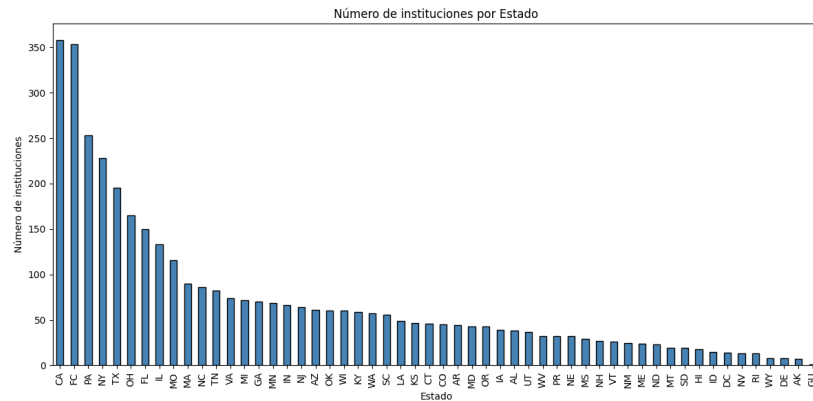
- La mayoría de instituciones tiene volúmenes reducidos de préstamos, pero existen casos extremos con montos muy elevados (ejemplo: hasta 419 millones USD en desembolsos subsidiados y 572 millones USD en desembolsos no subsidiados).
- Los valores de las medianas son mucho menores que los promedios, lo que confirma la presencia de outliers positivos que concentran gran parte del financiamiento.
- Programas como Parent PLUS y Grad PLUS muestran valores promedios reducidos en comparación con subsidiados y no subsidiados, evidenciando su menor participación en términos agregados.

Por último, la distribución por tipo de institución revela que:

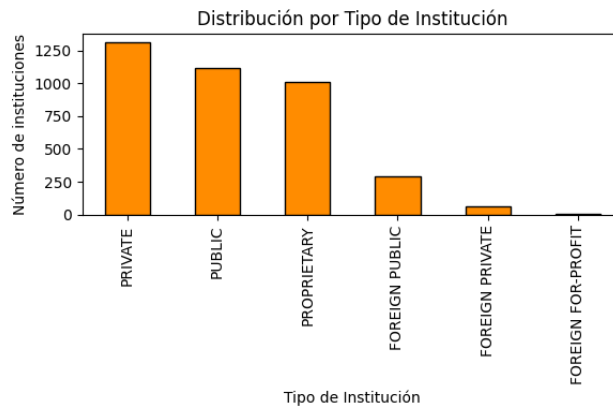
- Privadas (1,311) y públicas (1,119) dominan el panorama.
- Les siguen las propietarias (1,011), lo cual destaca la presencia significativa de instituciones con fines de lucro en el mercado educativo.
- Existen también instituciones extranjeras (288 públicas y 60 privadas) y un pequeño número de for-profit internacionales (4 casos), lo que subraya el carácter global del sistema de préstamos estudiantiles.

En conjunto, estas estadísticas iniciales evidencian que el dataset es lo suficientemente amplio y detallado para explorar diferencias en el acceso y financiamiento entre estados, tipos de institución y modalidades de préstamo.

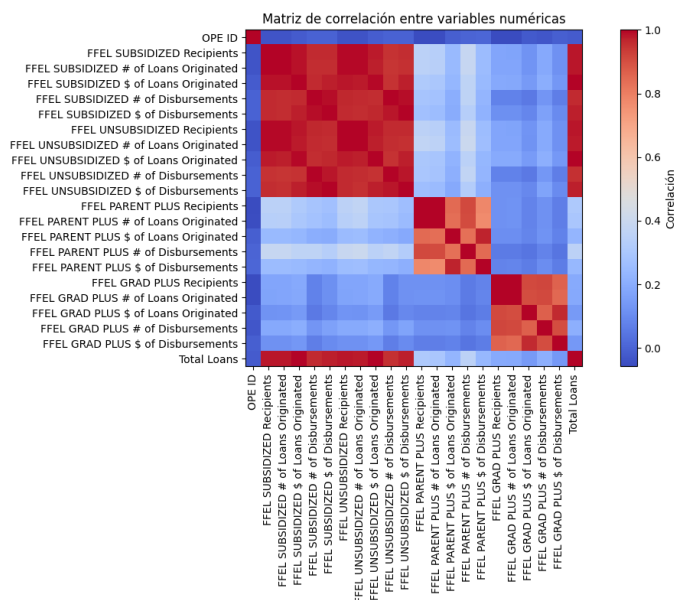
La distribución de instituciones por estado muestra una fuerte concentración en algunos territorios como California, Florida, Pensilvania y Nueva York, que destacan por su tamaño poblacional y densidad universitaria. Este patrón sugiere que la oferta educativa se concentra en estados con mayor población estudiantil potencial y una tradición consolidada de educación superior. En contraste, estados pequeños como Wyoming, Delaware, Alaska y Guam cuentan con muy pocas instituciones reportadas.



Cuando observamos la distribución por tipo de institución, los resultados reflejan un predominio de universidades privadas, seguidas por las públicas y en menor medida las propietarias. Esta composición resalta la relevancia del sector privado en el sistema universitario estadounidense, aunque también evidencia la presencia significativa de instituciones públicas que garantizan acceso masivo a la educación.



Finalmente, la matriz de correlación de variables numéricas permite identificar relaciones estrechas entre el número de receptores, el monto originado y el monto desembolsado en las diferentes modalidades de préstamos (subsidiados, no subsidiados, Parent PLUS y Grad PLUS). Las correlaciones altas confirman que las dinámicas de originación y desembolso están fuertemente vinculadas, lo que implica que los estados y tipos de institución con mayor capacidad de originación también concentran mayores desembolsos.



Análisis descriptivo diferenciado por tipo de institución

Para profundizar en las diferencias estructurales entre instituciones, se calcularon estadísticas diferenciadas por **tipo de universidad**. El interés radica en observar cómo varían los montos, el número de receptores y los patrones de desembolso en universidades públicas, privadas, propietarias y extranjeras.

School Type	FFEL SUBSIDIZED Recipients			FFEL SUBSIDIZED \$ of Loans Originated		
	mean	median	max	mean	median	max
FOREIGN FOR-PROFIT	776.250000	653.0	1326	4.529914e+06	4623499.0	6822160
FOREIGN PRIVATE	19.933333	2.0	544	1.405527e+05	8500.0	4116315
FOREIGN PUBLIC	1.697917	0.0	42	1.044460e+04	0.0	197709
PRIVATE	100.358505	17.0	11124	3.518584e+05	49560.0	42844944
PROPRIETARY	232.600396	37.0	42121	7.975982e+05	110856.0	166178989

Los resultados muestran que las instituciones propietarias (for-profit) concentran en promedio mayores montos en préstamos subsidiados y no subsidiados, mientras que las privadas tradicionales lideran en los programas Grad PLUS, donde captan más del 59% del total nacional. Las universidades públicas, aunque de menor participación en los préstamos más especializados, presentan medianas más elevadas, lo que indica una distribución menos sesgada.

School Type	FFEL SUBSIDIZED \$ of Loans Originated	FFEL UNSUBSIDIZED \$ of Loans
FOREIGN FOR-PROFIT	1.085188	1.136940
FOREIGN PRIVATE	0.505063	0.513743
FOREIGN PUBLIC	0.180152	0.177051
PRIVATE	27.626487	28.476998
PROPRIETARY	48.293688	49.587992
PUBLIC	22.309422	20.107276

El cálculo de la participación relativa confirma que las instituciones propietarias dominan en los préstamos subsidiados y no subsidiados, mientras que las privadas sin fines de lucro concentran la mayor parte del financiamiento Grad PLUS, lo que refuerza la hipótesis de que el tipo de institución condiciona el acceso a determinadas modalidades de crédito.

School Type	Ratio_Subsidized	Ratio_Unsubsidized
FOREIGN FOR-PROFIT	1.480766	1.481532
FOREIGN PRIVATE	4.415618	3.919582
FOREIGN PUBLIC	5.001230	6.053892
PRIVATE	2.489295	2.209379
PROPRIETARY	2.515551	2.567925
PUBLIC	1.823189	1.674309

En cuanto a la eficiencia en el desembolso, los valores promedio muestran que los programas en universidades extranjeras presentan ratios muy elevados, lo que implica que casi la totalidad del crédito originado se desembolsa. En contraste, las universidades públicas y privadas en EE. UU. mantienen ratios más moderados, cercanos a 2, lo que refleja procesos de desembolso más regulados.

	School	State	School Type	FFEL
7	ALABAMA AGRICULTURAL & MECHANICAL UNIVERSITY	AL	PUBLIC	133889
8	FAULKNER UNIVERSITY	AL	PRIVATE	234195
17	UNIVERSITY OF WEST ALABAMA	AL	PUBLIC	205179
26	TROY UNIVERSITY	AL	PUBLIC	725656
28	UNIVERSITY OF SOUTH ALABAMA	AL	PUBLIC	222113
40	REMINGTON COLLEGE - MOBILE CAMPUS	AL	PRIVATE	237936
44	COLUMBIA SOUTHERN UNIVERSITY	AL	PROPRIETARY	165638
97	UNIVERSITY OF ARIZONA (THE)	AZ	PUBLIC	335255
102	UNIVERSAL TECHNICAL INSTITUTE	AZ	PROPRIETARY	854625
111	UNIVERSITY OF PHOENIX	AZ	PROPRIETARY	166178

El análisis de outliers evidencia que un reducido grupo de universidades —entre ellas la University of Phoenix y la Universal Technical Institute— concentra montos extraordinarios, llegando incluso a más de 223 millones de dólares en préstamos no subsidiados. Esta alta concentración es crucial para el análisis de dependencia, pues estos casos extremos pueden distorsionar las correlaciones tradicionales y justificar el uso de cópulas para modelar colas pesadas y no linealidades.

Marco Teórico

Dependencia entre variables y límites de la correlación lineal

En finanzas y gestión de riesgo, las relaciones entre variables suelen ser no lineales, asimétricas y con colas pesadas. La correlación de Pearson captura solo dependencia lineal y puede ser poco informativa ante no linealidades o valores extremos. Por ello se recomiendan medidas basadas en rangos—como τ de Kendall y ρ de Spearman—y enfoques que separan las marginales de la dependencia (cópulas) (McNeil, Frey, y Embrechts 2015; Embrechts, Lindskog, y McNeil 2003; Nelsen 2006).

Cópulas y el teorema de Sklar

Una cópula es una función de distribución conjunta con marginales uniformes en $[0, 1]$ que permite “ensamblar” marginales univariadas para formar una distribución multivariada (Nelsen 2006). El Teorema de Sklar establece que para cualquier distribución multivariada H con marginales F_1, \dots, F_d existe una cópula C tal que

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

y, recíprocamente, C y las marginales determinan H (Sklar 1959). Esta descomposición posibilita modelar por separado (i) las marginales (p. ej., lognormal o gamma para montos; Poisson/NegBin para conteos) y (ii) la estructura de dependencia (Joe 2014).

Familias de cópulas y dependencia de cola

Entre las familias más utilizadas se encuentran:

Elípticas: gaussiana (sin dependencia de cola) y t (colas pesadas y dependencia de cola superior e inferior positivas cuando ν es finito) (McNeil, Frey, y Embrechts 2015).

Archimedeanas: Clayton (cola inferior fuerte), Gumbel (cola superior fuerte), Frank (simétrica sin dependencia de cola) y Joe (cola superior marcada) (Nelsen 2006; Joe 2014).

La dependencia de cola superior y inferior se definen como

$$\lambda_U = \lim_{x \rightarrow 1} \Pr(U_2 > u | U_1 > u),$$

$$\lambda_L = \lim_{u \rightarrow 0} \Pr(U_2 \leq u | U_1 \leq u)$$

Estas métricas son clave en riesgo crediticio al capturar la probabilidad de movimientos extremos (Embrechts, Lindskog, y McNeil 2003).

Medidas de dependencia robustas

Para comparar sectores (público, privado sin fines de lucro y privado con fines de lucro) conviene emplear τ de Kendall y ρ de Spearman, menos sensibles a escalas y atípicos que Pearson y, además, directamente vinculadas a parámetros de varias cópulas (Nelsen 2006; Joe 2014).

De acuerdo con (Rosenblatt 1952; Nelsen 2006) el proceso de estimación de cópulas generalmente sigue los siguientes pasos:

1. **Ajustar las distribuciones marginales** adecuadas para cada variable (por ejemplo, montos o conteos).
2. **Transformar los datos** mediante la *probability integral transform* (PIT) para obtener variables con distribución uniforme $U(0, 1)$.
3. **Estimar los parámetros de la cópula** sobre los datos transformados.

Este procedimiento permite desacoplar la estructura marginal de la dependencia subyacente (Genest, Ghoudi, y Rivest 1995). En la práctica, es común utilizar esquemas de estimación por verosimilitud en dos etapas, como el método de *Inference Functions for Margins* (IFM) o la pseudo-máxima verosimilitud, empleando funciones de distribución acumulada (CDF) empíricas o paramétricas. Estos métodos son consistentes bajo condiciones generales (Joe 2014).

Selección de modelos y pruebas de ajuste

La comparación entre familias de cópulas se apoya en el *Criterio de Información de Akaike* (AIC) y el *Criterio de Información Bayesiano* (BIC) (Akaike 1974; Schwarz 1978). Para evaluar el ajuste se recomiendan pruebas basadas en el empirical copula process (estadísticos tipo Cramér–von Mises o Kolmogorov–Smirnov) con bootstrap, que muestran buen poder frente a alternativas plausibles en finanzas (Genest, Rémillard, y Beaudoin 2009; Joe 2014).

Al modelar simultáneamente varias variables (p. ej., número de préstamos, monto originado, monto desembolsado, número de beneficiarios), las *pair-copula constructions* (PCC) o *vine cópulas* (*C-vine*/*D-vine*) permiten construir dependencias multivariadas flexibles a partir de cópulas bivariadas (Bedford y Cooke 2002; Aas et al. 2009).

Las cópulas se han consolidado en gestión de riesgos para modelar pérdidas conjuntas, defaults correlacionados y colas conjuntas (Embrechts, Lindskog, y McNeil 2003; McNeil, Frey, y Embrechts 2015). En crédito, la cópula gaussiana

popularizó el modelado de correlaciones de default en carteras (Li 2000), mientras que la cópula t y varias archimedeanas han destacado por capturar eventos extremos conjuntos. En el contexto de préstamos estudiantiles, estos marcos permiten contrastar si la estructura de dependencia entre variables operativas (montos, conteos, beneficiarios) difiere por tipo de institución, aportando evidencia para la gestión del riesgo, el diseño de políticas y la equidad en el acceso.

Referencias

- Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, y Henrik Bakken. 2009. «Pair-copula Constructions of Multiple Dependence». *Insurance: Mathematics and Economics* 44 (2): 182-98. <https://doi.org/10.1016/j.insmatheco.2008.07.003>.
- Akaike, Hirotugu. 1974. «A New Look at the Statistical Model Identification». *IEEE Transactions on Automatic Control* 19 (6): 716-23. <https://doi.org/10.1109/TAC.1974.1100705>.
- Bedford, Tim, y Roger M. Cooke. 2002. «Vines—A New Graphical Model for Dependent Random Variables». *The Annals of Statistics* 30 (4): 1031-68. <https://doi.org/10.1214/aos/1031689016>.
- Embrechts, Paul, Filip Lindskog, y Alexander McNeil. 2003. «Modelling Dependence with Copulas and Applications to Risk Management». En *Handbook of Heavy Tailed Distributions in Finance*, editado por Svetlozar T. Rachev, 329-84. Elsevier. [https://doi.org/10.1016/S0927-0507\(03\)10005-2](https://doi.org/10.1016/S0927-0507(03)10005-2).
- Genest, Christian, Kilani Ghoudi, y Louis-Paul Rivest. 1995. «A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions». *Biometrika* 82 (3): 543-52. <https://doi.org/10.1093/biomet/82.3.543>.
- Genest, Christian, Bruno Rémillard, y David Beaudoin. 2009. «Goodness-of-fit Tests for Copulas: A Review and a Power Study». *Insurance: Mathematics and Economics* 44 (2): 199-213. <https://doi.org/10.1016/j.insmatheco.2008.10.005>.
- Joe, Harry. 2014. *Dependence Modeling with Copulas*. Boca Raton: Chapman & Hall/CRC.
- Li, David X. 2000. «On Default Correlation: A Copula Function Approach». *The Journal of Fixed Income* 9 (4): 43-54. <https://doi.org/10.3905/jfi.2000.319257>.
- McNeil, Alexander J., Rüdiger Frey, y Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised. Princeton: Princeton University Press.
- Nelsen, Roger B. 2006. *An Introduction to Copulas*. 2.^a ed. New York: Springer. <https://doi.org/10.1007/0-387-28678-0>.
- Rosenblatt, Murray. 1952. «Remarks on a Multivariate Transformation». *The Annals of Mathematical Statistics* 23 (3): 470-72. <https://doi.org/10.1214/aoms/1177729394>.
- Schwarz, Gideon. 1978. «Estimating the Dimension of a Model». *The Annals of Statistics* 6 (2): 461-64. <https://doi.org/10.1214/aos/1176344136>.

- Sklar, Abe. 1959. «Fonctions de répartition à n dimensions et leurs marges». *Publications de l'Institut de Statistique de l'Université de Paris* 8: 229-31.
- U.S. Department of Education. 2010. «Federal Student Aid Annual Report 2010». Washington, D.C.: U.S. Department of Education; https://studentaid.gov/sites/default/files/fsa_annual_report_2009.pdf.
- . 2012. «Federal Family Education Loan Program: Annual Reports and Data». Federal Student Aid; <https://studentaid.gov>.