

## Resumen:

El objetivo de este trabajo es encontrar un modelo estadístico pedagógico de predicción del rendimiento académico de los estudiantes con impacto en el mejoramiento continuo del estudiante, a partir de la información publicada en la literatura especializada, en los últimos 50 años y de la consulta directa a estudiantes universitarios.

Para ello se consideran 36 escalas con sus 180 dimensiones latentes o factores construidas con más de 1000 ítems de respuesta y 18 preguntas adicionales, incluyendo un módulo de variables sociodemográficas constituidas por 15 variables. De todo ese registro de datos se obtienen más de 200 variables y dimensiones que normalmente se utilizan para explicar el rendimiento académico de los estudiantes universitarios a nivel mundial.

Se valoran, analizan y someten a un crítico y extenso ejercicio de validación de evidencia estadística, con el fin de encontrar unas cuantas variables “las más importantes”, que permitan construir un modelo estadístico pedagógico que acerque a una adecuada predicción del rendimiento académico que un estudiante tendrá en el futuro inmediato.

Por lo tanto, el objetivo de esta investigación es estudiar la relación entre cinco indicadores de rendimiento académico y un total de 200 variables, con el propósito de explicar el rendimiento académico de los estudiantes universitarios y conformar un modelo estadístico que apoye la gestión académica y administrativa. Todo ello, para que se logre una mejora continua del proceso de aprendizaje para estudiantes universitarios.

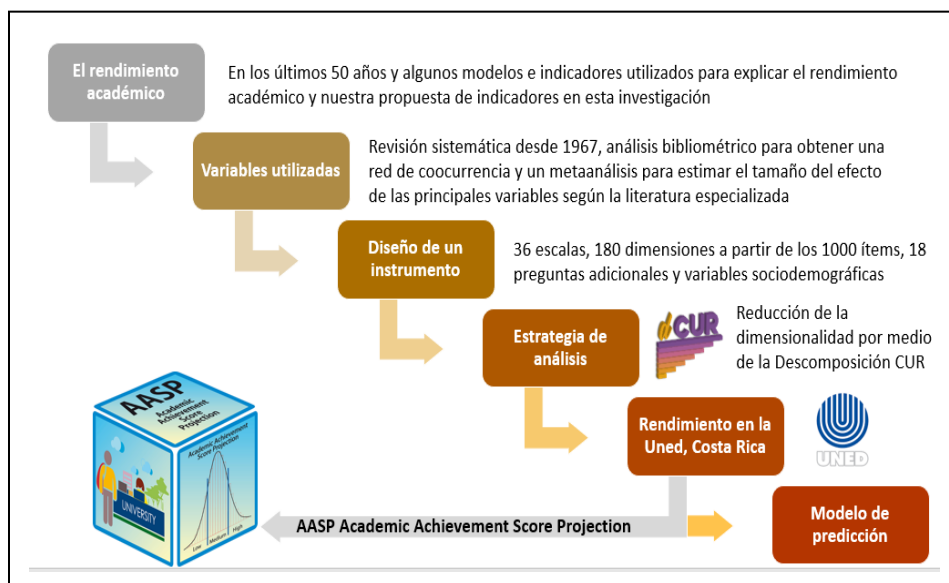
Determinar la relevancia estadística valorada mediante el tamaño del efecto de esas variables al explicar el desempeño de los estudiantes desde un punto de vista estadístico. Paralelamente se brinda una mejora a la técnica multivariante Descomposición de la matriz CUR y finalmente se diseña un modelo estadístico de clasificación de estudiantes que estime el percentil del rendimiento académico, junto con otros indicadores del estado pedagógico actual del estudiante.

Esta investigación se ha desarrollado con acciones concretas descritas en cinco etapas: la primera denominada “Búsqueda de las variables más relevantes para explicar el rendimiento académico”, que enmarca una revisión sistemática de la literatura que dio sustento en su parte inicial al diseño del cuestionario, considera también un análisis bibliométrico y un metaanálisis. Seguidamente se estableció el “Diseño y construcción estadística de cinco indicadores de rendimiento académico” como variables a utilizar para la construcción de los cluster de estudiantes según su nivel de rendimiento académico bajo, medio y alto.



## TESIS DOCTORAL

Modelo estadístico pedagógico para la toma de decisiones con impacto en el mejoramiento continuo del rendimiento de los estudiantes universitarios, basado en los métodos de selección CUR



La tercera etapa corresponde a una “Caracterización psicográfica y académica de los estudiantes de la Universidad Estatal a Distancia, Costa Rica con las variables relevantes” para continuar posteriormente con la etapa 4 en la que se realiza la “Construcción del modelo estadístico pedagógico” mediante la determinación de funciones discriminantes. Finalmente, la etapa 5 corresponde a la “Creación del modelo estadístico en una aplicación móvil para estudiantes” disponible de manera inmediata para los estudiantes universitarios.

El análisis exhaustivo, tanto de las herramientas estadísticas utilizadas, como de los artículos que tratan el rendimiento académico, nos ha permitido llegar a las siguientes conclusiones:

1. No hay unanimidad en la forma de evaluar el rendimiento académico, lo cual complica los estudios comparativos.
2. Se han identificado más de 270 variables que los investigadores han explorado, para explicar el rendimiento académico, en los últimos 50 años.
3. Los hallazgos de los diferentes investigadores no son concordantes en cuanto a cuáles de estas variables están significativamente asociadas con el rendimiento académico y cuáles no; probablemente porque las herramientas estadísticas utilizadas no capturan ni las interrelaciones ni las interacciones entre las variables predictoras.
4. La búsqueda de variables latentes para reducir la alta dimensionalidad del problema tampoco es un tema resuelto en la literatura estadística especializada. Hay una gran controversia a la hora de seleccionar la técnica más adecuada.
5. Nuestra contribución a nivel teórico se centra en la Descomposición CUR, técnica de muy reciente aparición, aplicable al análisis de matrices de altas dimensiones, la cual centra su atención en la búsqueda de las variables relevantes. Se propone un algoritmo que consiste en la selección de las mejores columnas y filas según la distribución de probabilidad de los puntajes leverage, sobre las  $k$  componentes utilizadas para su cálculo. Se ajusta la distribución empírica de los puntajes leverage a una distribución teórica existente para cada uno de los  $k$  componentes, con los que se puede calcular los puntajes leverage ( $k < N$ ), escogiendo aquellas columnas y filas en las que se acumule esa probabilidad.
6. Se ha creado el software denominado Dinamic CUR que dinamiza la búsqueda de las mejores columnas y filas; aquéllas que minimizan la diferencia entre la matriz original y la matriz aproximada CUR.
7. La propuesta teórica y el software desarrollado nos ha permitido identificar las variables más relevantes con el rendimiento académico según la descomposición CUR, estas son:

La autoeficacia percibida, la autorregulación del aprendizaje, la actitud que tienen los estudiantes hacia los estudios, la motivación académica, la constancia, el estilo de aprendizaje, la administración del tiempo, la opinión sobre el sistema de evaluación de los aprendizajes, el uso de los recursos didácticos que ofrece la universidad, son algunas de las más relevantes. El promedio de horas que le dedican a los estudios por semana, la realización de actividad física y la escolaridad de los padres son variables significativas, pero no las más relevantes, al contrario de lo que se ha creído hasta ahora. El clima de clase y el número de estudiantes en las clases son dos aspectos externos no inherentes al estudiante, pero también con alta influencia en el rendimiento académico.



En un proceso de análisis y depuración de variables para encontrar las variables de mayor fortaleza estadística para predecir el rendimiento académico, se sometieron al análisis discriminante las variables identificadas del metaanálisis, de la descomposición CUR y del análisis de variancia univariado, mejorando de un total de 200 variables a 84 variables importantes y terminando con un modelo de 28 variables altamente significativas.

8. Con estas 28 variables y dimensiones ha sido posible construir y desarrollar un modelo de predicción del percentil esperado del rendimiento académico, así como una estimación puntual y por intervalo del 95% de los cinco indicadores de rendimiento académico definidos en esta investigación.
9. Este modelo ha sido implementado en una APP, denominada “Academic Achievement Score Projection, AASP”, manejable desde los móviles, que permitirá a los estudiantes predecir su rendimiento académico y las variables con mayor impacto, controlando esas variables podrán modificar su rendimiento.

