**Algoritmos de Optimización**

**Actividad 1: Asignación Óptima de Horas de Tutoría Basada en Riesgo Académico**

Integrantes:

Cindy Liliana Vargas Duque   
 Luis Angel Vargas Narvaez   
 Jesús Ariel González Bonilla

**Introducción**

Este trabajo aborda la asignación óptima de horas de tutoría para mitigar el riesgo académico en programas universitarios. El problema es relevante porque exige optimizar un recurso escaso; las horas adicionales de acompañamiento, con el fin de maximizar la permanencia estudiantil y la eficiencia institucional. La masificación de la educación superior y la presión por indicadores de retención demandan mecanismos cuantitativos, reproducibles y auditables. En respuesta, se propone un marco formal, apoyado en datos, que prioriza intervenciones considerando el programa, el área disciplinar y el nivel de riesgo.

El objetivo general es minimizar el riesgo académico agregado posterior a la intervención, sujeto a restricciones presupuestales y de capacidad. Los objetivos específicos son: i) construir métricas de riesgo comparables; ii) parametrizar un modelo lineal interpretable; iii) realizar análisis de sensibilidad presupuestal; y iv) derivar precios sombra para apoyar la toma de decisiones estratégicas.

1. **Formular el problema de optimización**

La Corporación Universitaria del Huila – Corhuila gestiona distintos programas y cohortes cuyos estudiantes presentan un desempeño heterogéneo en las áreas de *lectura crítica, competencias ciudadanas, razonamiento cuantitativo e inglés*. En la siguiente gráfica se visualiza la distribución de los puntajes actuales en cada área: de modelos predictivos dependientes de historiales académicos completos.A screen shot of a graph

AI-generated content may be incorrect.

Ilustración 1 - Distribución de puntajes por área

Como se observa, razonamiento cuantitativo (verde) y competencias ciudadanas (naranja) concentran muchos puntajes bajos (picos alrededor de 60–100), lectura crítica (azul) se ubica en un rango intermedio (80–120), mientras que inglés (rojo) presenta valores altos con picos entre 200 y 260. Esta heterogeneidad permite identificar cohortes y áreas con mayor riesgo académico, lo que sustenta la necesidad de redistribuir las horas de tutoría de forma óptima.

La institución dispone de un presupuesto limitado de horas de tutoría. Cada área tiene una capacidad máxima de horas que puede absorber por semestre (por disponibilidad de docentes o restricciones curriculares). Además, se sabe que una hora de tutoría reduce el riesgo en una proporción constante para cada área, a partir de datos históricos o asumidos.

* 1. **Variables de decisión y parámetros**
* (C): conjunto de cohortes.
* (A): conjunto de áreas ({LC,CC,RQ,ING }).
* (x\_{c,a}): **horas de tutoría** asignadas a la cohorte (c) en el área (a) (variables continuas).
* (H): **presupuesto global** de horas disponibles.
* (H\_a): capacidad máxima de horas que el área (a) puede absorber.
* (r\_{c,a}): **riesgo base** de la cohorte (c) en el área (a), calculado a partir de los puntajes observados (mayor riesgo implica menor puntaje).
* (\_a): **coeficiente de efectividad** del área (a), que representa la reducción esperada del riesgo por cada hora de tutoría.
* (H\_c): capacidad máxima de horas que la cohorte (c) puede recibir (opcional).

* 1. **Función objetivo**

Se busca minimizar el **riesgo residual total** posterior a la intervención. Suponiendo que cada hora de tutoría reduce linealmente el riesgo, la función objetivo puede expresarse como:

lo cual equivale a maximizar la reducción agregada del riesgo (\_{c,a} *a, x*{c,a}). Para efectos computacionales se suele trabajar con la versión de maximización.

* 1. **1.4 Restricciones**

1. **Presupuesto global:** la suma de horas asignadas no puede superar (H):
2. **Capacidad por área:** cada área (a) tiene una capacidad máxima (H\_a):
3. **Límite por cohorte (opcional):** si existe un límite (H\_c) para cada cohorte (c):
4. **No negatividad:** las variables deben ser no negativas:
   1. **Clasificación del tipo de problema**

* **Determinista:** los parámetros (r\_{c,a}), (\_a), (H) y (H\_a) se consideran conocidos y fijos para el periodo de análisis.
* **Lineal:** la función objetivo y las restricciones son sumas ponderadas de las variables de decisión.
* **Con restricciones:** existe un presupuesto global y capacidades por área/cohorte.
* **Variables continuas:** las horas pueden asignarse de forma fraccionaria.

En consecuencia, se trata de un problema de **programación lineal continua** con múltiples restricciones que puede resolverse mediante el **método simplex**.

1. **Seleccionar el método de resolución**

A screenshot of a computer program

AI-generated content may be incorrect.

La naturaleza lineal del problema, junto con la ausencia de términos cuadráticos o lógicos, sugiere el uso de Programación Lineal (PL). El método simplex (o sus variantes dual simplex/HiGHS) es particularmente adecuado porque:

* Eficiencia: explota la estructura de los vértices factibles y garantiza encontrar la solución óptima en un número finito de iteraciones.
* Interpretabilidad: permite obtener precios sombra o multiplicadores duales que cuantifican el beneficio marginal de relajar cada restricción (ej. incrementar el presupuesto de horas o la capacidad de una determinada área).
* Estabilidad numérica: para problemas de tamaño moderado como el de las cohortes de Corhuila, los algoritmos simplex implementados en PuLP y HiGHS proporcionan soluciones robustas y reproducibles.

Se descarta el uso de gradiente descendente porque este se diseñó para optimizar funciones continuas no necesariamente sujetas a restricciones estrictas, y en este caso se necesita satisfacer exactamente las cotas de capacidad y presupuesto. Además, el simplex ofrece una interpretación dual que facilita la toma de decisiones estratégicas por parte de la institución.

El método de resolución seleccionado es Simplex a través de la interfaz PuLP con backend HiGHS/CBC. La elección se fundamenta en: (i) estructura puramente lineal, (ii) necesidad de obtener precios sombra (derivados del óptimo dual) y (iii) tamaño moderado del problema que hace innecesario un enfoque basado en gradiente continuo general o metaheurísticas. Se descarta gradiente descendente por carecer de explotación estructural de vértices y por la potencial dificultad de satisfacer restricciones con igual precisión sin un manejador dedicado. Se menciona que si en futuras extensiones se introducen costos enteros o funciones piecewise no lineales se evaluaría Branch-and-Bound o reformulaciones lineales adicionales.

A screenshot of a computer

AI-generated content may be incorrect.

1. **Resolver el problema usando Python**

La resolución se realizó empleando **Python 3.11**, aprovechando varias bibliotecas:

* **pandas** para cargar y preparar los datos de estudiantes, calcular puntajes promedio por cohorte y clasificar el nivel de riesgo.
* **PuLP** como interfaz de modelado para construir y resolver el problema de programación lineal; se utilizó el solver **HiGHS** por su velocidad y acceso a información dual.
* **matplotlib** y **seaborn** para generar visualizaciones, incluyendo la distribución de puntajes (**Figura 1**) y gráficos comparativos antes y después de la asignación óptima.

El proceso de implementación siguió los siguientes pasos:

**Carga y limpieza de datos:** se importó la base de datos institucional (por ejemplo, data.xlsx) con información de programas, cohortes y puntajes en las cuatro áreas. Se imputaron valores faltantes y se normalizaron los puntajes para construir un indicador de riesgo (r\_{c,a}) inversamente proporcional al rendimiento.  
**Definición de parámetros:** a partir de la base se estimaron el presupuesto global (H), las capacidades (H\_a), y se definieron los coeficientes de efectividad (\_a) (basados en evidencia empírica o supuestos razonables).  
**Modelado del PL:** usando PuLP se definió una variable continua x[(c,a)] para cada cohorte y área. Se codificó la función objetivo de maximización (*a x*{c,a}) y se añadieron las restricciones de capacidad y presupuesto.  
**Resolución y análisis de resultados:** se resolvió el modelo con HiGHS, obteniendo las horas óptimas x[(c,a)] y calculando la reducción de riesgo por cohorte y área. Se extrajeron los precios sombra de las restricciones y se construyeron tablas y gráficos que muestran la asignación de horas y el riesgo residual.  
**Análisis de sensibilidad:** se ejecutaron simulaciones variando el presupuesto (H) para estudiar los retornos marginales y detectar el punto de rendimientos decrecientes.

Un fragmento representativo del código utilizado para construir el modelo es el siguiente:

import pulp

# Crear el modelo de maximización

model = pulp.LpProblem("Asignacion\_Horas", pulp.LpMaximize)

# Variables de decisión x[(c,a)]

x = {(c, a): pulp.LpVariable(name=f"x\_{c}\_{a}", lowBound=0)

     for c in cohortes for a in areas}

# Función objetivo: maximizar la reducción de riesgo

model += pulp.lpSum(alpha[a] \* x[(c, a)] for c in cohortes for a in areas)

# Restricciones de capacidad por área

for a in areas:

    model += pulp.lpSum(x[(c, a)] for c in cohortes) <= H\_area[a], f"capacidad\_{a}"

# Restricción de presupuesto global

model += pulp.lpSum(x[(c, a)] for c in cohortes for a in areas) <= H\_total, "presupuesto"

# Resolver el modelo con HiGHS

model.solve(pulp.PULP\_CBC\_CMD(msg=False))

# Recuperar soluciones óptimas

soluciones = {(c, a): x[(c, a)].value() for c in cohortes for a in areas}

A screenshot of a computer screen

AI-generated content may be incorrect.

1. **Análisis e interpretación de resultados**

Los resultados cuantifican la reducción del riesgo agregado y exhiben la redistribución eficiente de horas enfocada en áreas con mayor pendiente de mitigación. Las figuras muestran comparativas de riesgo inicial versus residual y barras de asignación de horas. El análisis de sensibilidad evidencia cómo las reducciones marginales decrecen tras un umbral presupuestal (punto RD), sugiriendo que expansiones adicionales del recurso producirían retornos cada vez menores. Los precios sombra indican qué restricciones limitan el óptimo y priorizan inversiones: un precio sombra positivo para el presupuesto total implica valor en ampliar horas; precios sombra en capacidades de áreas revelan cuellos de botella específicos. La tabla de resultados y el resumen cuantitativo integran métricas de reducción absoluta y porcentual, junto con identificación de restricciones activas.

A graph with a line and a blue line

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

1. **Reducción significativa del riesgo:** la suma del riesgo residual (*{c,a} (r*{c,a} - *a x*{c,a})) disminuyó en un **porcentaje considerable** respecto al escenario sin tutoría, especialmente en *razonamiento cuantitativo* y *competencias ciudadanas*.
2. **Asignación focalizada:** las cohortes con puntajes más bajos recibieron más horas, mientras que *inglés*, con un mejor desempeño base, absorbió menos tutorías. Esto coincide con la distribución observada en la Figura 1.
3. **Precios sombra y cuellos de botella:** los multiplicadores duales asociados a las restricciones de capacidad por área señalaron dónde las horas adicionales tendrían mayor impacto. Un precio sombra alto en la restricción de *razonamiento cuantitativo* sugiere que aumentar su capacidad produciría mejoras significativas.
4. **Sensibilidad al presupuesto:** el análisis de sensibilidad mostró un **punto de rendimientos decrecientes**; más allá de cierto número de horas, la reducción marginal de riesgo por hora adicional se vuelve pequeña. Esto orienta a la institución sobre cuánto expandir el recurso antes de que deje de ser costo-efectivo.

**Conclusiones**

* La programación lineal resulta eficaz y transparente para asignar horas de tutoría según riesgo; en corhuila produjo una distribución interpretable y alineada con las necesidades de las cohortes, con potencial de mejorar la permanencia.
* El análisis dual entregó precios sombra útiles para decisiones de inversión, por ejemplo priorizar contratación en áreas con mayor valor marginal.
* limitaciones: supuestos de linealidad, ausencia de tratamiento explícito de la incertidumbre, calibración manual de coeficientes (k, α) y posibles interacciones no lineales no modeladas.
* Líneas futuras: estimar empíricamente las efectividades (métodos causales/a-b testing), agregar costos diferenciados y variables enteras (modelos mixtos), y extender a programación estocástica o robusta.
* Integrar el modelo con pipelines de aprendizaje automático para actualizar parámetros y cerrar el ciclo de retroalimentación; pese a las limitaciones, ofrece una base cuantitativa sólida para priorizar recursos.

**Referencias**

PuLP Community (2023). PuLP: Linear Programming in Python.

Python Software Foundation (2024). Python Language Reference.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering.