

Universidad EAFIT

Programa de Economía

Ciencia de Datos

Modelado de Recaudos Tributarios en Mercados Legalizados

Un Análisis Predictivo de la Distribución de Cannabis en Colombia

Autores:

Santiago Tupaz^{1,2}

Moisés Quintero¹

Vanessa Rodrigues¹

¹ Universidad EAFIT, Programa de Economía, Medellín, Colombia

² Universidad Nacional de Colombia, Programa de Estadística, Sede Medellín

Profesora Supervisora:

Paula María Almonacid Hurtado

Universidad EAFIT

Medellín, Colombia

10 de febrero de 2026

Resumen

Este artículo examina cómo podría proyectarse la estructura del recaudo tributario del Estado colombiano bajo distintos escenarios de legalización y regulación de la distribución de marihuana. Utilizando técnicas de ciencia de datos y análisis econométrico, construimos un modelo predictivo que integra variables sociodemográficas de encuestas nacionales con supuestos de regulación tributaria. Los resultados proporcionan evidencia cuantitativa para la formulación de políticas públicas en el contexto de la reforma tributaria y regulación de mercados alternativos en Colombia.

Palabras clave: recaudo tributario, legalización de cannabis, ciencia de datos, política fiscal, modelado de escenarios

Índice

1. Introducción	3
1.1. Contexto Global y Relevancia	3
1.2. Pregunta de Investigación	3
1.3. Importancia y Motivación	3
2. Marco Teórico y Evidencia Empírica	4
2.1. Experiencias Internacionales	4
2.2. Supuestos del Análisis	4
3. Metodología	5
3.1. Datos y Fuentes de Información	5
3.2. Variable Objetivo	5
3.3. Estrategia Analítica	6
3.4. Enfoque de Modelado	6
3.5. Escenarios de Simulación	7
3.6. Escenarios Simulados	7
4. Datos: Primera Descripción	7
4.1. Caracterización del Dataset	7
4.2. Interpretación de Datos Faltantes	7
5. Resultados Preliminares	8
6. Conclusiones	8

1. Introducción

1.1. Contexto Global y Relevancia

La legalización de sustancias controladas representa una transformación significativa en la política pública, con profundas implicaciones fiscales, sociales y económicas. A nivel mundial, más de 40 países han legalizado cannabis en alguna forma, generando ingresos tributarios considerables. En Estados Unidos, jurisdicciones como Colorado y Washington han recaudado más de 4,000 millones de dólares desde la legalización, demostrando el potencial fiscal de estos mercados regulados.

Colombia, por su parte, enfrenta un contexto único: históricamente vinculada a economías ilícitas relacionadas con drogas, la regulación de la distribución de marihuana representa una oportunidad para transformar estructuras económicas informales en fuentes formales de recaudo. La creciente consideración de marcos regulatorios alternativos abre la puerta a un análisis cuantitativo de escenarios tributarios.

1.2. Pregunta de Investigación

¿Cómo podría estructurarse el recaudo tributario del Estado colombiano bajo distintos escenarios de legalización de la distribución de marihuana?

La “estructura del recaudo tributario” se define como la participación porcentual de cada fuente de ingreso (impuestos a la renta, IVA, impuestos específicos, etc.) dentro del total recaudado por el Estado en un período determinado. Esta pregunta es fundamental porque va más allá de cuantificar ingresos adicionales para comprender cómo se reconfigurarían las proporciones del ingreso fiscal total bajo diferentes marcos regulatorios.

1.3. Importancia y Motivación

Desde una perspectiva fiscal, la legalización introduciría impuestos específicos al cannabis que constituirían una fuente de recaudo identificable y relevante. La evidencia internacional (Hansen et al., 2020; Dills et al., 2021) muestra que estos ingresos no son marginales ni temporales, sino que tienden a consolidarse a través del tiempo.

Desde una perspectiva económica, anticipar cómo se reconfiguraría la estructura del recaudo resulta clave para:

- Evaluar sostenibilidad fiscal bajo diferentes diseños tributarios
- Optimizar la asignación del gasto público
- Diseñar estructuras impositivas eficientes
- Contribuir al debate de política pública con evidencia cuantitativa

2. Marco Teórico y Evidencia Empírica

2.1. Experiencias Internacionales

Hansen, Miller y Weber (2020) estudian los efectos fiscales de la legalización en Washington y Oregón, encontrando que:

- La apertura de mercados legales genera ajustes inmediatos en ventas formales
- Existe consumo transfronterizo significativo: entre 8.1 % y 11.5 % del recaudo de Washington provenía de consumidores no residentes
- La legalización crea nuevas fuentes de recaudo tributario que no existían bajo prohibición

Dills, Goffard y Miron (2021) analizan los ingresos fiscales en múltiples estados, observando que:

- Los impuestos específicos al cannabis generan incrementos persistentes y significativos en recaudos estatales
- Estos ingresos continúan creciendo en años posteriores a la legalización
- La magnitud de los ingresos justifica su inclusión explícita en la estructura tributaria

2.2. Supuestos del Análisis

El estudio se fundamenta en los siguientes supuestos:

1. **Shock estructural:** La legalización incrementa el volumen de transacciones visibles en el mercado formal, absorbiendo parte del consumo que canalizaba la informalidad
2. **Movilidad de demanda:** El consumo no depende únicamente del mercado interno sino que responde a diferencias regulatorias y de precios
3. **Impuesto específico identificable:** La legalización introduce un tributo específico al cannabis que constituye una fuente relevante y consolidable
4. **Relación precio-demanda:** La tasa impositiva efectiva influye directamente en el tamaño del mercado legal y en el nivel de recaudo

3. Metodología

3.1. Datos y Fuentes de Información

Base de Datos Principal: Encuesta Nacional de Consumo de Sustancias (DANE)

- Observaciones actuales: 169,346 individuos con características sociodemográficas
- Capítulos disponibles en la encuesta DANE a explorar para seleccionar variables relevantes
- Proceso de descubrimiento de datos para identificar variables que se acoplen a la investigación

Variables Complementarias por Integrar:

- **DIAN:** Datos de recaudo tributario por fuente
- **DANE:** Variables macroeconómicas (PIB, inflación, consumo de hogares)
- **Múltiples capítulos de la encuesta:** Por explorar para segmentación de mercado

3.2. Variable Objetivo

Unidad de Análisis: Colombia en períodos anuales o semestrales, con enfoque macroeconómico.

Definición Formal: La variable objetivo es la **estructura del recaudo tributario**, definida como la participación porcentual de cada fuente tributaria dentro del total recaudado por el Estado en un período determinado:

$$Y_{k,t} = \frac{\text{Recaudo de la fuente } k \text{ en período } t}{\text{Recaudo total en período } t} \quad (1)$$

Características:

- **Variable composicional:** Las participaciones suman 1 en cada período
- **Acotada:** $Y_{k,t} \in [0, 1]$ para cada fuente
- **Dependiente del escenario:** El impuesto específico al cannabis aparece solo en escenarios de legalización
- **Evolución temporal:** Permite analizar cómo se reconfigura la estructura del recaudo a través del tiempo

3.3. Estrategia Analítica

El análisis se estructura en fases:

1. **Exploración de datos:** Conocimiento profundo de la naturaleza de los datos disponibles
 - Revisión de múltiples capítulos de la encuesta DANE
 - Selección de variables que se acoplen a la pregunta de investigación
 - Análisis de calidad, completitud y patrones en los datos
2. **Procesamiento:** Limpieza, normalización y tratamiento de valores faltantes
3. **Análisis exploratorio:** Estadísticas descriptivas, segmentación demográfica, patrones de comportamiento
4. **Modelado:** Desarrollo de modelo de machine learning a ser especificado según naturaleza de los datos
 - Enfoque adaptativo: El algoritmo específico dependerá de características de los datos
 - Benchmark: Probit ordenado con zeros inflados como modelo de referencia
5. **Simulación de escenarios:** Proyecciones fiscales bajo diferentes marcos regulatorios
6. **Validación:** Evaluación del desempeño predictivo fuera de muestra

3.4. Enfoque de Modelado

Dada la naturaleza composicional de la variable objetivo y la necesidad de comprender primero los datos, adoptamos un enfoque flexible:

Modelo Benchmark: Probit ordenado con zeros inflados

- Maneja variables acotadas entre 0 y 1
- Permite capturar características especiales de datos composicionales
- Proporciona punto de referencia para evaluar modelos alternativos

Modelos de Machine Learning:

- Se especificarán después del análisis exploratorio profundo
- Flexibilidad para adaptarse a patrones específicos en los datos
- Potencial para capturar no-linealidades y interacciones complejas

3.5. Escenarios de Simulación

Se evaluarán proyecciones fiscales bajo los siguientes marcos:

1. **Escenario Base:** Prohibición mantenida (estructura tributaria actual)
2. **Escenario 1:** Legalización amplia con IVA estándar (19 %)
3. **Escenario 2:** Legalización con impuesto especial (25 %)
4. **Escenario 3:** Regulación parcial con múltiples tramos impositivos

3.6. Escenarios Simulados

Se evalúan tres escenarios de política:

1. **Escenario Base:** Prohibición (estructura tributaria actual)
2. **Escenario 1:** Legalización amplia con IVA estándar (19 %)
3. **Escenario 2:** Legalización con impuesto especial (25 %)
4. **Escenario 3:** Regulación parcial con múltiples tramos tributarios

4. Datos: Primera Descripción

4.1. Caracterización del Dataset

El análisis se basa en una encuesta nacional de 169,346 individuos con información sociodemográfica:

Variable	Observaciones	Tasa de Respuesta
Sexo	169,346	100 %
Edad	169,346	100 %
Parentesco	169,346	100 %
Consentimiento	49,756	29.4 %
Resultado	49,756	29.4 %

4.2. Interpretación de Datos Faltantes

Las variables “Consentimiento”, “Resultado” y “Seleccionada” presentan 70.6 % de valores faltantes. Esta pauta sugiere que:

- Potencialmente corresponde a un submuestreo intencional dentro de la encuesta

- Los 49,756 casos con respuesta constituyen una submuestra válida
- No necesariamente representa pérdida de información sino un diseño muestral estratificado

La completitud del 100 % en variables demográficas (sexo, edad, parentesco) proporciona una base robusta para segmentación de mercado.

5. Resultados Preliminares

Los análisis detallados de predicción, simulación de escenarios y proyecciones fiscales se encuentran en desarrollo. Los próximos pasos incluyen:

- Estimación de modelos de regresión fraccionaria
- Cálculo de elasticidades precio de demanda
- Simulación de ingresos bajo diferentes tasas impositivas
- Análisis de sensibilidad a parámetros clave

6. Conclusiones

Este trabajo busca contribuir al debate sobre regulación de mercados de drogas en Colombia con herramientas cuantitativas rigurosas. La pregunta que motiva el análisis—cómo podría estructurarse el recaudo tributario bajo legalización—es fundamental para la formulación informada de política pública.

La integración de datos microeconómicos de encuestas nacionales con simulaciones de escenarios fiscales permite generar proyecciones coherentes y técnicamente robustas, aunque necesariamente contrafactuales. El valor de este ejercicio radica en ofrecer un marco analítico que complemente el debate actual, que suele estar dominado por consideraciones morales o de orden público.

Referencias

- Hansen, B., Miller, K., & Weber, C. (2020). Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana. *Journal of Public Economics*, 187, 104159.
- Dills, A., Goffard, S., & Miron, J. (2021). The effect of marijuana legalization on state tax revenue. *Journal of Law and Economics*, 64(3), 423–451.