



**Lic. en Ciencias del Comportamiento**

***Evaluación del riesgo de consumo de drogas según  
características particulares de individuos***

**Integrantes**

Ignacia Auguste  
Trinidad Morán  
Florencia Reverendo

**Asignatura**

Ciencia de Datos

**Profesores**

María Noelia Romero  
Ignacio Spiousas

**Fecha de entrega**

7 de diciembre del 2024

**Tutorial**

Nro 1, Hs: 19:00

**Semestre y año de presentación**

Primavera 2024

## **1. Introducción**

Hoy en día el problema de las adicciones atraviesa distintos grupos sociales en todo el mundo. La prevalencia del acceso a y el abuso de sustancias da crecimiento al número de adictos a nivel global, desafío cada vez más ineludible para los sistemas de salud que demanda implementación de estrategias para su mitigación (Peacock et al., 2018; Johnson, 1975).

Existen múltiples factores ambientales y biológicos que pueden influir en el consumo de sustancias adictivas, al igual que factores psicológicos como los rasgos de personalidad y problemas de salud mental que interactúan con el desarrollo y mantenimiento del abuso de sustancias (Belfiore *et al.*, 2024) . Investigaciones previas han analizado la relación entre los rasgos de personalidad del modelo de los cinco grandes y el consumo de drogas (Five Factor Model; Goldberg, 1992; McCrae y John, 1992).

En el presente trabajo, se buscará responder la siguiente pregunta: ¿Qué características de personalidad y demográficas permiten detectar factores de riesgo de consumo de drogas, tanto legales como ilegales? Para ello, se utilizará el modelo de riesgos proporcionales de Cox para evaluar el impacto relativo de ciertas variables de interés sobre la probabilidad de que un individuo se convierta en usuario de una droga específica, basándonos en el historial de consumo hasta el momento de la observación.

Integrar el uso de modelos predictivos para identificar factores relevantes en el consumo de drogas podría tener un gran impacto sobre las políticas públicas, para el diseño y la implementación de programas de prevención y rehabilitación de individuos propensos al consumo (Chy & Buadi, 2024).

## **Literatura previa**

En la literatura, se ha profundizado sobre la naturaleza de los factores que impactan en el consumo de sustancias. En particular, ya se han utilizado las características demográficas y el *Five Factor Model* (Goldberg, 1992; McCrae y John, 1992) como variables predictoras del consumo de drogas en adolescentes. La presente propuesta de investigación busca profundizar sobre la implementación de modelos de aprendizaje automático, para evaluar el desempeño y poder predictivo al explorar la relación entre perfiles de personalidad, impulsividad, búsqueda de sensaciones y factores demográficos en el consumo de drogas..

En el trabajo de Nesa, Shaha y Yoon (2022) se propuso predecir el crimen juvenil causado por la adicción a sustancias utilizando técnicas de *machine learning*. El estudio de Beltz y Weigard (2019) propuso tres modelos para analizar neuroimágenes con enfoques en el uso de sustancias de adolescentes. Estas investigaciones nos fueron útiles para seleccionar nuestro método de análisis. El modelo de proporción de riesgo de Cox no ha sido mencionado en la literatura por lo que no sólo se estaría llenando un hueco en la literatura, sino que también se estaría proponiendo otro marco donde se analizan los factores de riesgo.

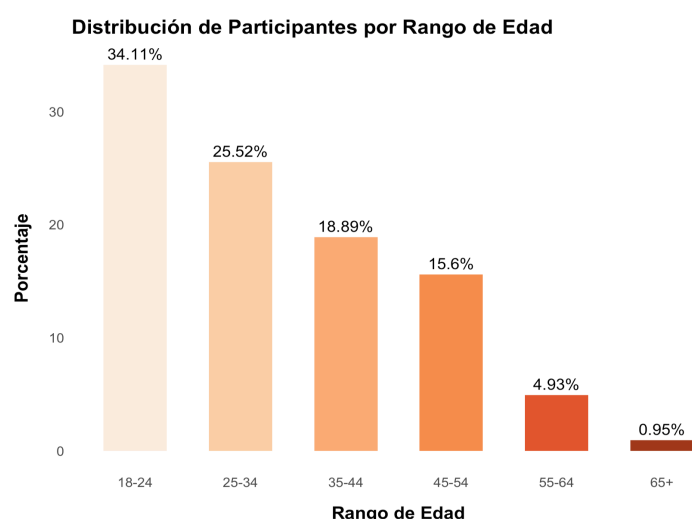
Asimismo, ha sido ampliamente estudiado qué rasgos de personalidad del *Five Factor Model* (Goldberg, 1992; McCrae y John, 1992) tienen los adictos. Unos ejemplos son las investigaciones de Terracciano et al. (2008) y de Andreassen et al. (2013). En estas dos, se demostró que había rasgos de personalidad en común entre los adictos.

Por lo tanto, la literatura previa demuestra la relevancia de comprender con qué factores es posible predecir el uso de drogas y que este análisis predictivo es posible realizarlo con métodos de *machine learning*, como el modelo de proporción de Cox. Además, se ha mostrado que las personas adictas tienen rasgos de personalidad en común. Entonces, si usáramos estos rasgos en común como predictores, sería posible predecir si una persona podría o no ser adicta.

## 2. Base de datos

La base de datos a utilizar fue descargada a través de la página UC Irvine Machine Learning Repository, provista por investigadores de la Universidad de *Leicester* (Fehrman et al., 2016). Cuenta con observaciones pertenecientes a 1885 individuos mayores a 18 años recolectadas entre los años 2011 y 2012, a través de una encuesta anónima en línea. Para cada participante se conocen 12 atributos: cinco scores pertenecientes a rasgos de personalidad según el *Five Factor Model* (Goldberg, 1992; McCrae y John, 1992) que contemplan extroversión, apertura, neuroticismo, amabilidad y responsabilidad; el resultado de la escala BIS-11 de Impulsividad; el resultado de la escala ImpSS de búsqueda de sensaciones; y datos demográficos (edad, género, nivel educativo, país y etnia). Como variables relacionadas al consumo de droga se pregunta por el uso de: alcohol, anfetaminas, nitrito de amilo, benzodiacepinas, cannabis, chocolate, cocaína, cafeína, crack, éxtasis, heroína, ketamina, “legal highs”, LSD, metadona, hongos, nicotina y abuso de sustancias volátiles (VSA). También se incluyó una droga ficticia llamada Semeron para filtrar individuos que respondieran positivamente sin criterio. Cada una de estas variables está codificada de manera categórica según el último consumo: "Nunca usado", "Usado hace más de una década", "Usado en la última década", "Usado en el último año", "Usado en el último mes", "Usado en la última semana" y "Usado en el último día".

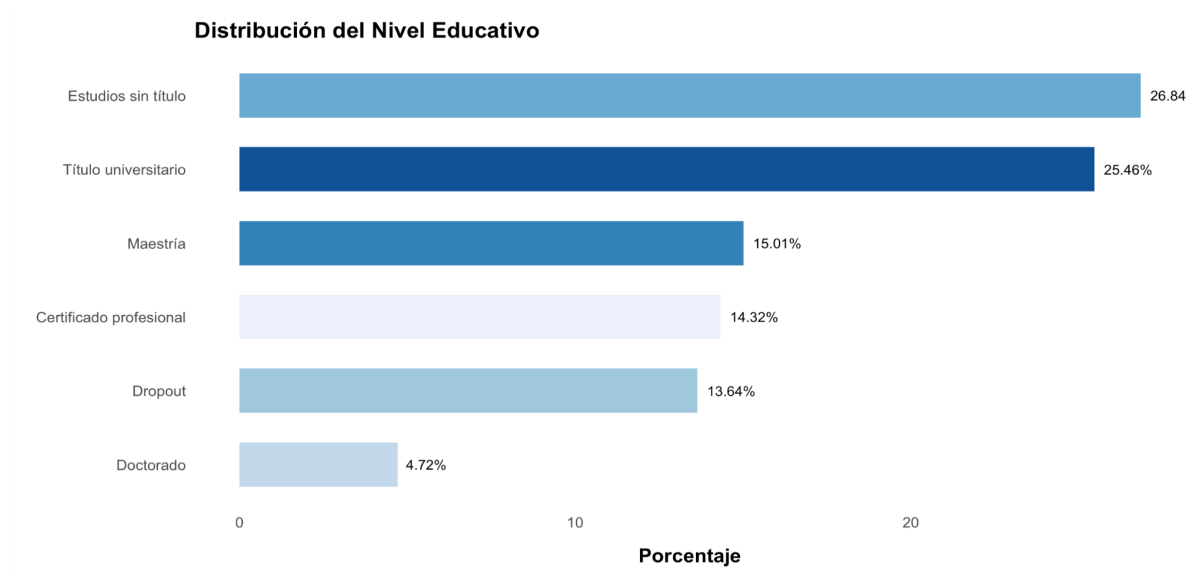
La muestra tiene una media de edad de, aproximadamente, 33 años, con un desvío estándar de 12 años y se distribuye de manera descendente entre los 18 y los 65 años o más.



La distribución para cada rango de edad se expone en la **Figura 1**.

**Figura 1.** Distribución de edad en la muestra.

La distribución por género es equilibrada, con un 49.97% de mujeres y un 50.03% de hombres. En cuanto al nivel educativo, el mayor porcentaje tiene estudios sin título (26.84%), seguidos por título universitario (25.46%). Un 13.64% de la muestra abandonó los estudios a las 18 años o antes, un 14.32% posee un certificado profesional o diploma, el 15.01% un título de maestría y el 4.72% un doctorado (**Figura 2**).



**Figura 2.** Distribución del nivel educativo de la muestra.

Casi el total de la muestra pertenece a países angloparlantes. Geográficamente, el 55.38% de los individuos reportaron residir en el Reino Unido, el 29.55% en los Estados Unidos, el 4.62% en Canadá, el 2.86% en Australia, el 1.06% en Irlanda, el 0.27% en Nueva Zelanda, y el 6.26% reporta vivir en otros países. Étnicamente, el 91.25% de los participantes se identifica como blanco, el 1.75% como negro, el 1.38% como asiático, el 3.34% como “otro”, y el 2.28% como una combinación de orígenes étnicos.

### 3. Metodología

Para abordar el problema de predecir sujetos *at risk* de ser consumidores de droga, el evento de interés fue definido como la transición de un individuo de la categoría de no

usuario a usuario de una droga específica. Este enfoque permite estudiar cómo las características del individuo afectan la probabilidad de consumo. En este contexto, se considerará que el evento ya había ocurrido para aquellos individuos que ya eran usuarios de la droga en el momento de la observación.

El tiempo en el modelo será conceptualizado a partir de las categorías de recencia de uso, las cuales se trató como un proxy temporal. Estas categorías serán transformadas en valores numéricos utilizando una escala exponencial para representar ordinalmente el paso del tiempo desde el primer uso hasta el momento de la observación. Con el fin de que la categoría "Nunca" tenga un valor de 0 y que las diferencias entre las categorías sean no lineales hemos decidido transformar las variables de la siguiente manera:  $\text{valor categoría} = e^{\text{categoría} - 1}$ . Los valores iniciales de las categorías son los siguientes: "Nunca": 0; "Hace más de una década": 1; "Última década": 2; "Último año": 3; "Último mes": 4; "Última semana": 5, "Último día": 6. De esta manera los nuevos valores de las categorías que utilizaremos para nuestro análisis son los siguientes: "Nunca": 0; "Hace más de una década": 1.718; "Última década": 6.389; "Último año": 19.085; "Último mes": 54.598; "Última semana": 147.413, "Último día": 402.428. Este enfoque nos permitirá otorgarle más peso a las categorías más recientes, que reflejan un consumo más frecuente.

Como variables predictoras se incluirán las siguientes:

- Rasgos de personalidad: puntuaciones en los cinco grandes rasgos de personalidad (NEO-FFI-R: neuroticismo, extraversión, apertura a la experiencia, amabilidad y responsabilidad), impulsividad (BIS-11) y búsqueda de sensaciones (ImpSS).
- Variables demográficas: edad, género, nivel educativo, país de residencia y etnicidad.
- Historial de consumo de otras sustancias: el consumo de otras drogas o la exposición a entornos sociales relacionados con el consumo.

Se ajustará un modelo de Cox de manera independiente para cada una de las drogas consideradas en la base de datos. De esta manera se evaluará la proporcionalidad de los riesgos utilizando métodos gráficos y pruebas estadísticas para asegurar que las covariables mantuvieron efectos constantes a lo largo del tiempo.

La validez del modelo se verificará mediante un enfoque de validación cruzada, dividiendo los datos en conjuntos de entrenamiento y prueba. A través de este método evaluaremos la estabilidad de los coeficientes estimados y la capacidad predictiva del modelo.

Los resultados se presentarán mediante gráficos de curvas de supervivencia, estratificados por características clave (por ejemplo, niveles altos vs bajos de neuroticismo, género, etc.). Estas curvas permitirán visualizar cómo cambian las probabilidades de que una persona se convierta en usuaria de una droga a lo largo del tiempo, según su perfil de personalidad y otras características relevantes.

#### **4. Conclusiones y limitaciones**

Si lográramos implementar el modelo de riesgos proporcionales de Cox con un marco temporal adecuado y datos estructurados correctamente, se podrían obtener resultados significativos que permitirían evaluar cómo las variables predictoras impactan en el riesgo de consumo de drogas. El modelo sería capaz de identificar cuáles variables tienen un impacto significativo en el riesgo de consumo. Factores como los rasgos de personalidad (por ejemplo, la impulsividad o la búsqueda de sensaciones), las características demográficas (como la edad, el género y el nivel educativo) y otros antecedentes personales aparecerían como variables relevantes.

Si se contara con un marco temporal robusto, se podrían analizar los patrones de riesgo de consumo a lo largo del tiempo. Esto permitiría observar cómo ciertos factores

influyen en momentos específicos de la vida, brindando una visión más completa del fenómeno.

Esperaríamos que permitiera comparar los riesgos entre distintos subgrupos de individuos. Por ejemplo, sería posible analizar cómo varía el riesgo de consumo entre personas de diferentes niveles educativos o grupos de edad, lo que facilitaría el diseño de estrategias preventivas específicas.

Además, a partir de las características de cada individuo, el modelo podría generar perfiles personalizados de riesgo, lo que sería especialmente útil para identificar casos con mayor probabilidad de consumo y priorizar intervenciones.

Una de las principales limitaciones de este proyecto radica en que no contamos con datos longitudinales que permitan realizar un seguimiento de los individuos con medidas repetidas a lo largo del tiempo. Esto nos impide observar directamente la variable que queremos predecir. En su lugar, solo podemos estimar la probabilidad de que ocurra un evento en un período de tiempo, lo que dificulta validar las predicciones realizadas por el modelo.

Adicionalmente, el modelo de Cox asume que el efecto de las covariables sobre el riesgo es constante a lo largo del tiempo. Sin embargo, sabemos que factores como el paso del tiempo y la edad tienen una influencia significativa sobre la predisposición al consumo, y esta variación no se captura de manera adecuada bajo este supuesto. Además, el modelo no incorpora variables ambientales ni genéticas, las cuales sabemos que tienen una influencia considerable en los patrones de consumo.

Otro desafío relevante es la colinealidad entre algunas variables predictoras, que puede afectar la estabilidad y confiabilidad de las estimaciones del modelo. La presencia de



relaciones fuertes entre las covariables dificulta la interpretación de los resultados y compromete la capacidad del modelo para identificar contribuciones independientes al riesgo.

Finalmente, la ausencia de un marco temporal claro complica significativamente el análisis. Si bien el modelo de Cox se basa en la medición del tiempo hasta un evento, no contamos con datos específicos de duración en esta implementación, lo que limita la precisión y utilidad del enfoque en este contexto.

## 5. Referencias bibliográficas

- Andreassen, C. S., Griffiths, M. D., Gjertsen, S. R., Krossbakken, E., Kvam, S., & Pallesen, S. (2013). The relationships between behavioral addictions and the five-factor model of personality. *Journal of Behavioral Addictions*, 2(2), 90–99.  
doi:10.1556/jba.2.2013.003
- Belfiore, C., Galofaro, V., Cotroneo, D., Lopis, A., Tringali, I., Denaro, V., & Casu, M. (2024). A Multi-Level Analysis of Biological, Social, and Psychological Determinants of Substance Use Disorder and Co-Occurring Mental Health Outcomes. *Psychoactives*. <https://doi.org/10.3390/psychoactives3020013>.
- Beltz, A. M. & Weigard, A. (2019). Methodological Advances in Leveraging Neuroimaging Datasets in Adolescent Substance Use Research. *Current Addiction Reports*, 6, 495-503. <https://doi.org/10.1007/s40429-019-00275-x>
- Chy, M., & Buadi, O. (2024). Role of Machine Learning in Policy Making and Evaluation. *International Journal of Innovative Science and Research Technology (IJISRT)*.  
<https://doi.org/10.38124/ijisrt/ijisrt24oct687>.
- Fehrman, E., Egan, V. & Mirkes, E. (2016). *Drug Consumption (Quantified)*. Recuperado el 7 de diciembre de 2024 de <https://archive.ics.uci.edu/dataset/373/drug+consumption+quantified>.

- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26–42. <https://doi.org/10.1037/1040-3590.4.1.26>
- Johnson, B. D. (1975). *Understanding British addiction statistics*. United Nations. Office on Drugs and Crime. Recuperado el 7 de diciembre de 2024 de: [https://www.unodc.org/unodc/en/data-and-analysis/bulletin/bulletin\\_1975-01-01\\_1\\_page006.html](https://www.unodc.org/unodc/en/data-and-analysis/bulletin/bulletin_1975-01-01_1_page006.html)
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175–215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- Nesa, M., Shaha, T. R. & Yoon, Y. (2022). Prediction of juvenile crime in Bangladesh due to drug addiction using machine learning and explainable AI techniques. *Journal of Computational Social Science*, 5, 1467-1487. <https://doi.org/10.1007/s42001-022-00175-7>
- Peacock, A., Leung, J., Larney, S., Colledge, S., Hickman, M., Rehm, J., ... Degenhardt, L. (2018). Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. *Addiction*, 113(10), 1905-1926. doi:10.1111/add.14234
- Terracciano, A., Löckenhoff, C. E., Crum, R. M., Bienvenu, O. J., & Costa, P. T. (2008). Five-Factor Model personality profiles of drug users. *BMC Psychiatry*, 8(1). doi:10.1186/1471-244x-8-22