



UNIVERSIDAD DE PANAMÁ  
VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO  
PROGRAMA DE MAESTRÍA EN ESTADÍSTICA APLICADA

**INFORME DE PROYECTO DE INTERVENCIÓN**

**FACTORES ACADÉMICOS Y SOCIODEMOGRÁFICOS QUE INFLUYEN EN  
EL RESULTADO DE LA PRUEBA DE ADMISIÓN DE LA FACULTAD DE  
CIENCIAS NATURALES, EXACTAS Y TECNOLOGÍA DE LA UNIVERSIDAD  
DE PANAMÁ, 2023: UN ANÁLISIS MULTIVARIANTE.**

Presentado como requisito para optar  
al grado de Maestro en Estadística  
Aplicada

**EDUVIGIS MERCEDES RODRÍGUEZ**

ASESORA:

MAGISTER CLARA CRUZ

PANAMÁ, REPÚBLICA DE PANAMÁ 2025

**UNIVERSIDAD DE PANAMÁ  
VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO  
FACULTAD DE CIENCIAS NATURALES EXACTAS Y TECNOLOGÍA  
DEPARTAMENTO DE ESTADÍSTICA  
ESCUELA DE ESTADÍSTICA**

**PROGRAMA DE MAESTRÍA EN ESTADÍSTICA APLICADA**

**Tribunal Examinador del Trabajo de Graduación**

Los profesores miembros del tribunal examinador del trabajo del informe de proyecto de intervención, de la estudiante Eduvigis Mercedes Rodríguez, participaron como revisores y evaluadores.

Titulada:

**FACTORES ACADÉMICOS Y SOCIODEMOGRÁFICOS QUE INFLUYEN  
EN EL RESULTADO DE LA PRUEBA DE ADMISIÓN DE LA FACULTAD DE  
CIENCIAS NATURALES, EXACTAS Y TECNOLOGÍA DE LA  
UNIVERSIDAD DE PANAMÁ, 2023: UN ANÁLISIS MULTIVARIANTE.**

Fue sustentada en el Campus Harmodio Arias Madrid de la Universidad de Panamá, aula \_\_\_ el día \_\_\_ de \_\_\_\_\_ del 2025, ante el Jurado constituido por los siguientes docentes:

Presidente: \_\_\_\_\_

Representante de VIP: \_\_\_\_\_

Miembros: \_\_\_\_\_ y \_\_\_\_\_

Habiendo recibido la calificación de \_\_\_\_\_

---

EL SECRETARIO DEL JURADO

Panamá, \_\_\_\_\_ de \_\_\_\_\_ del 2025

## ÍNDICE GENERAL

<b>ÍNDICE DE TABLAS</b> .....	v
<b>ÍNDICE DE FIGURAS</b> .....	vi
<b>DEDICATORIA</b> .....	vii
<b>AGRADECIMIENTO</b> .....	viii
<b>ABREVIATURAS</b> .....	ix
<b>RESUMEN</b> .....	x
<b>ABSTRACT</b> .....	xi
<b>INTRODUCCIÓN</b> .....	1
<b>CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA</b> .....	3
1.1 Antecedentes .....	3
1.1.1 El Análisis Estadístico en las Pruebas de Admisión .....	4
1.2 Formulación del Problema.....	5
1.3 Justificación .....	6
1.4 Objetivos Generales y Específicos.....	7
1.4.1 Objetivo General.....	7
1.4.2 Objetivos Específicos .....	7
1.5 Hipótesis .....	8
<b>CAPÍTULO II. MARCO TEÓRICO</b> .....	9
2.1 Antecedentes Históricos .....	9
2.2 Marco Conceptual .....	13
2.2.1 Rendimiento Académico .....	13
2.2.2 Pruebas de Admisión .....	14
2.2.3 Factores Académicos .....	15
2.2.4 Factores Sociodemográficos .....	15
2.2.5 Relación entre las Variables de Estudio .....	17
2.3 Marco Referencial.....	19

2.4 Marco Contextual.....	21
2.5 Descripción Teórica de las Técnicas Estadísticas Utilizadas.....	23
2.5.1 Regresión Logística Binaria (RLB).....	23
2.5.2 Análisis de Correspondencias Múltiples (ACM).....	26
CAPÍTULO III. MARCO METODOLÓGICO.....	29
3.1 Tipo de Investigación .....	29
3.2 Diseño de la Investigación.....	29
3.3 Definición de Variables .....	31
3.4 Instrumentos Utilizados.....	34
3.5 Población .....	34
3.6 Tratamiento Estadístico .....	35
3.6.1 Depuración de Datos.....	37
3.6.2 Análisis Descriptivo.....	37
3.6.3 Codificación y Preparación de Variables .....	37
3.6.4 Modelado con Regresión Logística Binaria (RLB).....	38
3.6.5 Análisis de Correspondencias Múltiples (ACM).....	39
3.6.6 Validación y reproducibilidad .....	41
3.6.7 Visualizaciones Empleadas .....	42
3.8 Recursos Técnicos y Metodológicos Utilizados .....	44
CAPÍTULO IV. ANÁLISIS DE RESULTADOS .....	45
4.1 Análisis Descriptivo .....	45
4.1.1 Preparación de los Datos y Caracterización de la Cohorte .....	45
4.2 Regresión Logística Binaria (RLB).....	53
4.2.1 Especificación del modelo logístico .....	54
4.2.2 Comprobación de supuestos del modelo.....	54
4.2.3 Modelo inicial e interpretación preliminar.....	55
4.2.4 Modelo reducido y coeficientes finales .....	56

4.2.5 Bondad de ajuste y desempeño del modelo .....	57
4.2.6 Evaluación gráfica del modelo .....	59
4.2.7 Discusión.....	60
4.3 Análisis de Correspondencias Múltiples (ACM) .....	62
4.3.1 Contribuciones y Calidad de Representación (cos2).....	66
4.3.2 Mapas factoriales del Análisis de Correspondencias Múltiples (ACM).....	68
4.4 Síntesis de hallazgos.....	74
4.5 Indicador de Probabilidad de Ingreso Regular (IPIR).....	75
<b>CONCLUSIONES</b> .....	78
<b>RECOMENDACIONES</b> .....	80
<b>REFERENCIAS BIBLIOGRÁFICAS</b> .....	82
<b>ANEXOS</b> .....	87

## ÍNDICE DE TABLAS

Tabla 1. Técnicas multivariantes seleccionadas según enfoque .....	28
Tabla 2. Definiciones conceptuales y operacionales de las variables .....	32
Tabla 3. Resumen de procesos realizados en el tratamiento estadístico .....	36
Tabla 4. Resumen de tipo de gráficos en el tratamiento de los datos .....	42
Tabla 5. Resumen de técnicas utilizadas en el modelado multivariante .....	43
Tabla 6. Participantes de la prueba de admisión, según clasificación de ingreso. Año 2023 .....	46
Tabla 7. Distribución del índice predictivo, según sexo. Año 2023.....	47
Tabla 8. Distribución del índice predictivo, según bachillerato. Año 2023.....	47
Tabla 9. Distribución del índice predictivo, según colegio de procedencia. Año 2023 .....	48
Tabla 10. Distribución del índice predictivo, según grupo de edades. Año 2023 .....	49
Tabla 11. Distribución del índice predictivo, según provincia. Año 2023.....	50
Tabla 12. Estadísticos descriptivos de variables continuas.....	50
Tabla 13. Interpretación de las razones de momios (OR) e intervalos de confianza del 95 % del modelo reducido de regresión logística binaria .....	56
Tabla 14. Razones de momios (OR) e intervalos de confianza del 95 % del modelo reducido de regresión logística binaria .....	57
Tabla 15. Métricas de desempeño del modelo logístico sobre la cohorte .....	58
Tabla 16. Matriz de Burt. Frecuencias conjuntas de las variables cualitativas.....	62
Tabla 17. Resumen de pruebas Chi-cuadrado.....	63
Tabla 18. Inercia y proporción de las primeras dimensiones del ACM .....	64
Tabla 19. Contribuciones y calidad de representación ( $\cos^2$ ) .....	66
Tabla 20. Categorías con mayor contribución a la dimensión 2 .....	67
Tabla 21. Clasificación de niveles de riesgo según el IPIR .....	76

## ÍNDICE DE FIGURAS

Figura 1. Evolución de los métodos multivariantes aplicados en la investigación educativa y social.....	12
Figura 3. Gráfica de los participantes de la prueba de admisión, según clasificación de ingreso. Año 2023.....	46
Figura 4. Diagrama de mosaico del índice predictivo, según colegio de procedencia. Año 2023.....	48
Figura 5. Gráfica de distribución del índice predictivo, según promedio de bachillerato. Año 2025.....	51
Figura 6. Histograma de la distribución del Índice Predictivo (IP). Año 2023.....	52
Figura 7. Gráfica de la curva ROC del modelo logístico reducido.....	59
Figura 8. Gráfica de la curva calibración del modelo logístico reducido.....	60
Figura 9. Gráfica de sedimentación del análisis de correspondencias múltiples...	65
Figura 10. Mapa factorial de categorías (ACM, Dim1–Dim2).....	69
Figura 11. Gráfica de individuos mejor representados en el plano.....	72

## DEDICATORIA

Al culminar esta especialidad, que representa un paso significativo en mi desarrollo profesional en el campo de la Estadística Aplicada, dedico este logro a quienes han sido mi mayor fuente de inspiración y fortaleza.

A nuestro Padre Celestial, quien nos sostiene en medio de un mundo donde la mujer carga con desafíos constantes; a Él, que es nuestra fuente de vida, quien nos levanta cuando sentimos caer, que nos fortalece frente al cansancio y nos recuerda que somos capaces de alcanzar nuestras metas.

A mi familia, especialmente a mi esposo Emiliano González, a mis hijos Nataly Mercedes y Moisés David, por acompañarme con amor, paciencia y apoyo incondicional. A mi madre Luisa, y a mi padre Eric Rodríguez (q. e. p. d.), quien sé que, desde donde esté, se siente orgulloso de cada paso que doy.

A los docentes que nos guiaron con compromiso y generosidad a lo largo de esta especialidad, y que sembraron en mí el deseo de continuar perfeccionándome. A todas las personas que, de una u otra forma, influyeron positivamente en nuestro proceso.

Dedico este trabajo a todas las mujeres y profesionales que luchan por abrirse paso en el mundo académico y técnico. Que este esfuerzo sirva como motivación para quienes aún están en el camino.

## **AGRADECIMIENTO**

Agradezco profundamente a Dios Todopoderoso y Omnipotente por permitirme culminar esta etapa académica y fortalecerme en cada desafío. Su guía y luz han sido mi sostén en este camino.

A mi familia, en especial a mis padres, esposo y a mis hijos, por su amor, apoyo constante y palabras de ánimo que me impulsaron a no abandonar este sueño. A mis amigos, colegas y seres queridos, por motivarme a continuar y celebrar cada pequeño avance.

Extiendo mi gratitud a los docentes y facilitadores de esta especialidad, a quienes compartieron su conocimiento con dedicación, exigencia y empatía, y a todas las personas que directa o indirectamente fueron parte de este proceso de crecimiento y formación profesional.

A cada uno de ustedes, gracias infinitas por haber contribuido a este logro.

## ABREVIATURAS

Abreviatura	Significado
IP	Índice Predictivo (puntaje total de admisión)
PCA	Puntaje de aptitud de la prueba (verbal/numérica)
PCG	Puntaje de conocimientos científicos
ACM	Análisis de correspondencias múltiples
RLB	Regresión logística binaria
Dim	Dimensión del espacio factorial
Inercia	Varianza explicada por dimensión ( $\chi^2/n$ )
CTR	Contribución de la categoría a la dimensión
Coord.	Coordenadas factoriales
AIC	Criterio de información de Akaike
BIC	Criterio de información bayesiano
LL	Log-verosimilitud
2LL (Deviance)	Desviación del modelo
ROC	Curva característica operativa del receptor
AUC	Área bajo la curva ROC
R <sup>2</sup> McF	Pseudo-R <sup>2</sup> de McFadden
Wald	Estadístico de Wald
gl	Grados de libertad
R	Lenguaje y programa libre

*Nota.* Se definen las abreviaturas utilizadas a lo largo de este proyecto de intervención; para mayor detalle, véase el Anexo C.

## RESUMEN

En Panamá, el acceso a la educación superior se ha venido ampliando bajo principios de equidad, transparencia y respaldo en evidencia empírica. En ese marco, y en coherencia con el compromiso institucional, se analizó la cohorte 2023 de la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá con el fin de estimar la relación entre factores académicos y sociodemográficos y la clasificación de ingreso derivada del índice predictivo. Se trabajó con la base institucional completa bajo un enfoque observacional y analítico que integró la caracterización de esta con técnicas multivariantes, como la regresión logística binaria y el análisis de correspondencias múltiples.

Los resultados muestran un patrón académico estable, donde los gráficos factoriales permiten observar estructuras claras que relacionan factores académicos con las condiciones sociodemográficas. Mientras que la regresión aísla el efecto neto de cada predictor sobre la probabilidad de clasificación. Por tratarse de un análisis censal, las conclusiones se limitan estrictamente a la cohorte estudiada.

El uso combinado de métodos multivariantes ofrece evidencia útil para la gestión institucional, ya que permite ajustar los procesos de admisión y orientar de forma temprana los programas de apoyo académico, fortaleciendo la toma de decisiones oportuna, basada en datos.

*Palabras clave:* admisión universitaria; rendimiento académico; factores sociodemográficos; regresión logística; análisis de correspondencias múltiples; educación superior; Panamá.

## ABSTRACT

In Panama, access to higher education has been expanding under principles of equity, transparency, and reliance on empirical evidence. Within this framework, and in line with the institutional commitment, the 2023 cohort of the Faculty of Natural, Exact, and Technological Sciences at the University of Panama was analyzed to estimate the relationship between academic and sociodemographic factors and the admission classification derived from the predictive index. The complete institutional database was used under an observational and analytical approach that integrated its characterization with multivariate techniques such as binary logistic regression and multiple correspondence analysis.

The results reveal a consistent academic pattern, where factorial graphs display clear structures linking academic factors with sociodemographic conditions, while logistic regression isolates the net effect of each predictor on the probability of classification. Since this was a census analysis, the conclusions are strictly limited to the cohort studied.

The combined use of multivariate methods provides useful evidence for institutional management, as it enables adjustments to admission processes and early guidance of academic support programs, thereby strengthening timely and data-driven decision-making.

*Keywords:* university admission; academic performance; sociodemographic factors; logistic regression; multiple correspondence analysis; higher education; Panama

## INTRODUCCIÓN

El sistema panameño ha hecho avances sostenidos en la educación superior, para ampliar el acceso con criterios de equidad, transparencia y sustentación empírica. En coherencia con ese horizonte, la Universidad de Panamá consolida prácticas de mejora en la gobernanza académica informada por datos. En este contexto institucional, comprender cómo se articulan factores académicos y sociodemográficos con los resultados de la prueba de admisión es clave para orientar decisiones y acompañar la progresión del estudiantado desde el ingreso.

Este estudio examina la cohorte 2023 de la Facultad de Ciencias Naturales, Exactas y Tecnología con el objetivo de identificar patrones y asociaciones entre características del aspirante y su clasificación de ingreso. Se trabajó con la base institucional completa bajo un enfoque observacional y analítico.

La estrategia metodológica integró la caracterización descriptiva con dos técnicas multivariantes complementarias entre ellas la regresión logística binaria, para estimar el efecto neto de los predictores sobre la probabilidad de una clasificación favorable, y análisis de correspondencias múltiples, para sintetizar relaciones entre categorías y perfilar estructuras latentes.

La implementación del software libre R y la trazabilidad del flujo analítico favorecen la reproducibilidad y su actualización en proyectos posteriores. Por el carácter censal de la base, la interpretación se circunscribe a la cohorte analizada y los hallazgos derivados del análisis multivariante proporcionan evidencia accionable para la gestión, tal que permiten afinar procedimientos de admisión, focalizar apoyos académicos tempranos

y aportar insumos útiles para la mejora pedagógica y la orientación vocacional en la educación media.

Finalmente, el documento se estructura en cuatro capítulos: el Capítulo I aborda los antecedentes, la formulación del problema, la justificación y los objetivos de la investigación; el Capítulo II desarrolla el marco teórico que sustenta el estudio; el Capítulo III detalla el tratamiento estadístico aplicado; y el Capítulo IV presenta los resultados y la discusión, seguidos de las conclusiones, recomendaciones y los anexos correspondientes.

## **CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA**

### **1.1 Antecedentes**

A lo largo de las décadas, se ha observado un creciente interés por determinar los factores que influyen en el aprendizaje, con el objetivo de implementar metodologías acordes a las tendencias educativas. Los factores sociodemográficos y los institucionales asociados con las dimensiones del rendimiento académico han demostrado ser valiosos en el análisis de investigaciones educativas.

De acuerdo con Santos Álvarez (2013), aunque los estudiantes suelen manejar con solvencia los aspectos teóricos, presentan limitaciones al aplicarlos en contextos prácticos, lo que resalta la necesidad de fortalecer la conexión entre teoría y práctica.

En ese mismo contexto, Garbanzo (2013), manifiesta que los factores socioeconómicos vinculados al desempeño académico universitario todavía carecen de suficiente explicación, particularmente en el contexto de la educación superior pública, que tiene un llamado social con las poblaciones estudiantiles.

Así, los factores como el género, lugar de residencia, dependencia administrativa, la rama del establecimiento educacional de egreso, el promedio de secundaria, pueden ser los puentes entre los egresados de la educación media y la superior.

Pascual (2011) subraya que factores institucionales como la dirección escolar, el clima de convivencia, el liderazgo y la administración de recursos pueden incidir de manera significativa en el rendimiento académico (p. 29).

Estudios previos, entre ellos los de Salagre y Serrano (2003) y Cunningham et al. (2007), muestran que tanto la modalidad de bachillerato como el promedio de secundaria ejercen un efecto importante sobre el desempeño universitario.

Lacarrière (2008) enfatiza la relevancia de la preparación docente y Ríos (2012) resalta la importancia de un entorno institucional positivo, factores que influyen indirectamente en la preparación de los estudiantes para enfrentar los retos de la educación superior.

### **1.1.1. El análisis estadístico en las pruebas de admisión.**

La estadística, como ciencia interdisciplinar, aporta al investigador en el desarrollo de la formulación del problema, la definición y clasificación de los tipos de variables aleatorias que conforman la problematización de la investigación. También proporciona las técnicas adecuadas para la selección de la muestra y el análisis de los datos recopilados, lo que permite la interpretación y la toma de decisiones.

El uso de bases de datos confiables da respuestas a las problemáticas planteadas en diversas investigaciones. En este caso, las pruebas de admisión universitarias, mediante el análisis estadístico, aportarían información sobre múltiples variables de interés, evidenciando cambios significativos o debilidades que podrían abordarse con propuestas de mejora en la praxis del aula, la gestión escolar, entre otras, permitiendo cumplir con los objetivos, misión.

En el desarrollo de la investigación, el análisis multivariante proporciona una variedad de métodos y técnicas. Díaz (2002) explica que el análisis multivariante se aplica a conjuntos de datos con múltiples mediciones tomadas sobre diversos individuos o

unidades de estudio (p. 56). Sin embargo, es relevante tener en cuenta qué técnica se ajusta mejor al contexto, de modo que, al reducir las dimensiones, se puedan identificar los factores asociados al rendimiento académico de los estudiantes. Estos factores, que podrían detectarse mediante técnicas multivariantes, aportarían a las instituciones educativas públicas de la educación media académica, profesional y técnica.

## **1.2 Formulación del Problema**

En Panamá, las pruebas de admisión de las universidades públicas han sido, durante décadas, un instrumento fundamental para evaluar las fortalezas y debilidades académicas de los aspirantes a ingresar a estudios superiores. Los egresados de las dos modalidades de educación media, científico-humanista y técnico-profesional, acceden a estas pruebas con una formación diferenciada, basada en competencias generales que buscan responder a sus intereses, aptitudes y aspiraciones vocacionales para la continuidad de estudios superiores.

Sin embargo, una problemática persistente en el contexto educativo nacional es que una parte significativa de los estudiantes presenta bajo rendimiento académico, alta vulnerabilidad al abandono escolar y poca inclinación a continuar estudios a nivel superior (Fiegehen & Díaz, 2008). Esta situación plantea desafíos tanto para la formación previa como para los mecanismos de acceso a la educación superior.

Analizar estadísticamente, mediante las técnicas multivariantes los factores académicos y sociodemográficos asociados a los resultados de las pruebas de admisión, permitió aportar evidencia objetiva que contribuye a la toma de decisiones. Los hallazgos de esta investigación pueden beneficiar la mejora de prácticas pedagógicas, la gestión

educativa y las estrategias de orientación vocacional en las instituciones de educación media del país.

A partir de esta problemática, se formuló la siguiente pregunta de investigación: ¿Cuáles son los factores académicos y sociodemográficos que influyen en los resultados de la prueba de admisión de la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá?

### **1.3 Justificación**

La presente investigación responde a la necesidad de comprender el rendimiento de los aspirantes mediante el uso de técnicas estadísticas multivariantes. La influencia de los factores académicos y sociodemográficos en los resultados de admisión de la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá, se analizan con el software libre (R). El estudio representa una innovación en el contexto nacional y permite generar evidencia útil para fortalecer la equidad, la calidad y la toma de decisiones en la educación superior.

Los resultados brindan información aplicable para mejorar los procedimientos de admisión, en especial al reconocer perfiles de aspirantes con distinta probabilidad de éxito académico. Además de caracterizar la situación, el estudio sugiere medidas específicas para fortalecer la gestión universitaria.

En consecuencia, este estudio aporta desde una doble perspectiva: empírica, al mostrar la relación entre factores académicos, sociodemográficos, resultados de ingreso y metodológica, al dejar una herramienta de análisis replicable mediante el software libre.

Este enfoque contribuye a evaluar la equidad entre aspirantes y a fortalecer la calidad de la educación superior en Panamá.

#### **1.4 Objetivos Generales y Específicos**

Los objetivos de esta investigación son:

##### **1.4.1 Objetivo general.**

- Analizar los factores académicos y sociodemográficos que influyen en los resultados obtenidos en la prueba de admisión de la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá.

##### **1.4.2 Objetivos específicos.**

- Depurar la base de datos de los estudiantes que realizaron la prueba de admisión.
- Caracterizar las condiciones generales de los estudiantes que realizaron la prueba de admisión en la Universidad de Panamá.
- Identificar los factores que influyen en los resultados de la prueba de admisión en la Universidad de Panamá.
- Determinar la relación entre las características sociodemográficas y el puntaje obtenido en la prueba de admisión.
- Examinar la influencia de los factores académicos en el resultado de la prueba de admisión.

- Identificar cuáles de las variables académicas y sociodemográficas presentan mayor capacidad predictiva sobre el puntaje total de admisión.
- Aplicar técnicas de análisis estadístico multivariante para identificar los factores académicos y sociodemográficos que influyen en los resultados obtenidos en la prueba de admisión.
- Formular recomendaciones basadas en los hallazgos, orientadas a la mejora de los procesos de admisión y la gestión educativa en la Facultad de Ciencias Naturales, Exactas y Tecnología.

### **1.5 Hipótesis**

En este estudio no se formulan hipótesis estadísticas, ya que se trabaja con la totalidad de los datos disponibles del año 2023. Al analizar a la población completa, no se requiere hacer inferencias, sino más bien caracterizar y describir los patrones observados directamente.

## CAPÍTULO II. MARCO TEÓRICO

### 2.1 Antecedentes Históricos

La evaluación del rendimiento académico ha sido, desde hace décadas, una herramienta fundamental para la admisión a instituciones de educación superior en todo el mundo. Las pruebas de admisión buscan estimar las habilidades, conocimientos y competencias necesarias para afrontar exitosamente los estudios universitarios, y su implementación ha evolucionado con el tiempo, adaptándose a los contextos educativos, tecnológicos y sociales (Martínez, 2016).

En varios países latinoamericanos se han implementado exámenes estandarizados de ingreso universitario, con el propósito de asegurar criterios de equidad y transparencia en el acceso. Panamá no ha sido la excepción. En la Universidad de Panamá, particularmente en la Facultad de Ciencias Naturales, Exactas y Tecnología, la prueba de admisión se ha consolidado como un instrumento clave para la selección de estudiantes. Esta se compone de secciones que miden la comprensión verbal, el razonamiento numérico y los conocimientos generales (González & Rojas, 2019).

Ante las crecientes demandas institucionales, ha surgido la necesidad de profundizar en el estudio de los factores que inciden en los resultados de estas pruebas. En este sentido, el uso de técnicas estadísticas multivariantes ha permitido explorar patrones de asociación entre variables académicas y sociodemográficas, así como modelar la probabilidad de éxito en la admisión (Hair et al., 2019). Entre estas metodologías destacan la regresión logística binaria (RLB), utilizada para modelar variables dicotómicas mediante la estimación de la probabilidad de ocurrencia de un evento (Hosmer, Lemeshow &

Sturdivant, 2013). En este estudio, la RLB se aplica para predecir la clasificación de los estudiantes en dos categorías de admisión: ingreso regular e ingreso propedéutico. Asimismo, el análisis de correspondencias múltiples (ACM) facilita la exploración gráfica y estructural de relaciones entre variables categóricas, permitiendo identificar perfiles de estudiantes y relaciones entre categorías analíticas (Greenacre & Blasius, 2006).

Investigaciones recientes también evidencian un cambio en la concepción de las pruebas de admisión. García (2020) subraya que los sistemas de ingreso en Centroamérica han transitado de evaluaciones centradas en la memorización hacia enfoques que valoran las competencias de razonamiento verbal, lógico y matemático, con el propósito de garantizar mayor validez en la predicción del rendimiento universitario. Además, plantea que estas pruebas no deben entenderse únicamente como filtros de selección, sino también como instrumentos diagnósticos que identifiquen necesidades de nivelación y apoyen la implementación de cursos propedéuticos.

En el caso de Panamá, este planteamiento cobra especial relevancia, ya que coincide con la necesidad institucional de comprender cómo los factores académicos y sociodemográficos se relacionan con los resultados de admisión universitaria. Estudios nacionales han evidenciado que el rendimiento previo, las condiciones socioeducativas y las desigualdades de origen influyen en los procesos de ingreso y permanencia en la educación superior (MEDUCA, 2023; Contraloría General de la República, 2022). En este contexto, la incorporación de modelos estadísticos rigurosos contribuye al fortalecimiento de la gestión educativa al ofrecer evidencia sólida para la toma de decisiones, permitiendo

identificar factores críticos, orientar recursos y apoyar la formulación de políticas de admisión más equitativas y basadas en datos (OECD, 2021).

Por otra parte, la disponibilidad de herramientas estadísticas ha acompañado la evolución de los enfoques analíticos en la educación superior, impulsando estudios más rigurosos y basados en evidencia. Este desarrollo ha permitido que los procesos de investigación académica avancen hacia metodologías más integrales, apoyadas en la exploración de relaciones complejas entre variables y en la aplicación de criterios cuantitativos para la toma de decisiones.

En particular, el acceso al software libre *R* ha democratizado la posibilidad de aplicar técnicas avanzadas de análisis multivariante que anteriormente estaban restringidas a instituciones con licencias costosas, permitiendo que universidades públicas de América Latina integren metodologías más robustas para evaluar sus procesos educativos (R Core Team, 2023; Pérez-Luño & Martínez-Gómez, 2020). Este avance ha favorecido que instituciones como la Universidad de Panamá fortalezcan los estudios sobre admisión y rendimiento académico mediante herramientas reproducibles, abiertas y accesibles.

En este contexto de ampliación y acceso, las técnicas estadísticas no solo han evolucionado desde el plano teórico, sino también en su implementación práctica mediante el desarrollo de programas especializados.

## Figura 1

*Evolución de los métodos multivariantes aplicados en la investigación educativa y social*

<b>Década de 1960-1970</b>
<ul style="list-style-type: none"><li>• Inicia el desarrollo teórico de técnicas como el Análisis de Componentes Principales (PCA) y el Análisis Factorial (Hotelling, 1933; Jolliffe, 1972).</li></ul>
<b>Años 1980</b>
<ul style="list-style-type: none"><li>• Se populariza el uso de la regresión múltiple y MANOVA en las ciencias sociales, apoyadas por software como SPSS y SAS (Norusis, 1983; Tatsuoka, 1988).</li></ul>
<b>Década de 1990</b>
<ul style="list-style-type: none"><li>• Surgen paquetes como STATA y Minitab que permiten ejecutar modelos más complejos con mayor facilidad (StataCorp, 1992; Minitab, 1995).</li></ul>
<b>Años 2000</b>
<ul style="list-style-type: none"><li>• Aparece AMOS, especializado en modelos de ecuaciones estructurales, y SmartPLS orientado al modelado con muestras pequeñas (Arbuckle, 2005; Ringle, Wende &amp; Will, 2005).</li><li>• Se consolida el uso del Análisis de Correspondencias Múltiples (ACM) en ciencias sociales para analizar datos categóricos (Greenacre &amp; Blasius, 2006).</li></ul>
<b>Década de 2010</b>
<ul style="list-style-type: none"><li>• Se generaliza el uso de R y Python como herramientas de código abierto, incorporando bibliotecas para SEM, PCA y visualización multivariada (R Core Team, 2014; Wickham, 2016).</li></ul>
<b>Década de 2020</b>
<ul style="list-style-type: none"><li>• Se expande el aprendizaje automático (machine learning) y la inteligencia artificial, integrando métodos tradicionales (ACM, RLB, PCA) con redes neuronales y modelos predictivos más complejos (Hastie, Tibshirani &amp; Friedman, 2021).</li></ul>

Nota. Elaboración propia a partir de la sistematización de la literatura académica, basada en Hair, Black, Babin y Anderson (2014); Greenacre y Blasius (2006); Jolliffe (2002); Kline (2016); y Hastie, Tibshirani y Friedman (2021).

La Figura 1 presenta la evolución histórica de los métodos multivariantes en la investigación educativa y social a nivel internacional, destacando el surgimiento de técnicas y software, incluidos desarrollos recientes como el aprendizaje automático y la

inteligencia artificial, que posteriormente fueron incorporados en la región latinoamericana.

## **2.2 Marco Conceptual**

A partir de los antecedentes, se requiere precisar los conceptos centrales que orientan el análisis estadístico, destacando el rendimiento académico como variable clave y los factores sociodemográficos asociados que permiten explicar su relación con los procesos de admisión universitaria. En este apartado se profundiza en dichos conceptos.

### **2.2.1 Rendimiento académico**

El rendimiento académico constituye uno de los constructos más estudiados en el ámbito educativo, pues refleja el nivel de aprendizaje y desempeño de los estudiantes dentro del sistema formal. Tinto (2012) plantea que el rendimiento académico se relaciona directamente con la permanencia y logro de los estudiantes en la universidad, y depende de elementos personales e institucionales.

En el contexto de los procesos de admisión universitaria, el rendimiento académico suele considerarse un predictor significativo del desempeño futuro, ya que refleja la trayectoria previa del estudiante y su capacidad para afrontar exigencias académicas superiores (Tinto, 2012; Pascarella & Terenzini, 2005). Para esta investigación, dicho rendimiento se operacionaliza mediante un índice predictivo que integra el promedio de secundaria, la Prueba de Capacidades Académicas (PCA) en sus componentes verbal y numérico, y la Prueba de Conocimientos Generales (PCG). Este índice sintetiza de manera global el desempeño previo del estudiante en áreas fundamentales y constituye la variable dependiente principal de los modelos analíticos que serán evaluados.

### **2.2.2 Pruebas de admisión**

Las pruebas de admisión son instrumentos estandarizados que permiten evaluar de forma objetiva las habilidades, conocimientos y competencias necesarias para afrontar con éxito los estudios universitarios. Diversas investigaciones han mostrado que este tipo de evaluaciones constituyen predictores relevantes del rendimiento académico inicial, ya que aportan información sobre las capacidades cognitivas y el potencial de aprendizaje del aspirante (Camara & Echternacht, 2000; Kobrin et al., 2008).

Aunque su estructura varía según el contexto institucional, estas pruebas suelen incluir componentes de razonamiento verbal, razonamiento numérico y conocimientos generales. Cada una de estas dimensiones busca medir habilidades asociadas a la comprensión lectora, el pensamiento lógico-matemático y la apropiación de contenidos fundamentales que se consideran necesarios para el desempeño en la educación superior.

En el contexto latinoamericano, González y Rojas (2019) destacan que las pruebas de admisión han servido tanto para la selección de aspirantes como para promover procesos más equitativos y meritocráticos. Estas evaluaciones permiten a las instituciones establecer criterios uniformes de ingreso y valorar de forma más objetiva la preparación previa de quienes aspiran a cursar estudios universitarios.

En la Universidad de Panamá, las pruebas administradas por la Facultad de Ciencias Naturales, Exactas y Tecnología incluyen las dos áreas de la Prueba de Capacidades Académicas (PCA), correspondientes al razonamiento verbal y numérico, así como la Prueba de Conocimientos Generales (PCG). Estas dimensiones reflejan el nivel académico alcanzado al finalizar la educación media y constituyen indicadores significativos del

desempeño inicial en la vida universitaria, al permitir identificar fortalezas y áreas de mejora en el perfil cognitivo de los estudiantes admitidos (Pascarella & Terenzini, 2005).

### **2.2.3 Factores académicos**

Los factores académicos corresponden a un conjunto de variables que describen la trayectoria escolar previa del estudiante y que pueden incidir directamente en su rendimiento futuro. Estos elementos permiten comprender la formación acumulada antes del ingreso a la educación superior y aportan información clave para analizar el desempeño universitario.

Entre los factores más relevantes destacan el promedio general de secundaria y el tipo de bachillerato cursado. Ambos indicadores reflejan no solo la preparación académica del estudiante, sino también el nivel de exposición a determinadas áreas del conocimiento, lo cual puede influir en su adaptación a las exigencias curriculares de la universidad (Horn & Kojaku, 2001).

De acuerdo con García (2020), estos antecedentes no solo evidencian el desarrollo cognitivo alcanzado, sino también competencias transversales como disciplina, responsabilidad y constancia. Tales características se consideran fundamentales para comprender la capacidad del estudiante para sostener un desempeño adecuado durante el primer año universitario y enfrentar los desafíos propios de la educación superior.

### **2.2.4 Factores sociodemográficos**

Los factores sociodemográficos corresponden a un conjunto de características que describen el contexto social, económico, cultural y geográfico de un individuo. Estas variables permiten comprender las condiciones de origen del estudiante y su posible

influencia en los procesos educativos, especialmente en el rendimiento, el acceso y la permanencia en la educación superior (Babbie, 2014; UNESCO, 2020). Entre los factores más comúnmente analizados se encuentran el sexo, el nivel socioeconómico, el lugar de residencia, el tipo de colegio de procedencia y la estructura familiar.

En el caso particular de esta investigación, los factores sociodemográficos seleccionados como: sexo, tipo de colegio (oficial o particular) y provincia de residencia, permiten caracterizar al estudiante a partir de su contexto y explorar de qué manera estas condiciones podrían influir en su desempeño académico inicial. Garbanzo (2013) señala que elementos como el nivel socioeconómico, el entorno familiar y la ubicación geográfica generan desigualdades estructurales que inciden en las oportunidades de éxito educativo, especialmente en los procesos de ingreso a la educación superior.

A nivel regional, investigaciones realizadas en universidades latinoamericanas evidencian que el género y el tipo de institución educativa de procedencia pueden asociarse con diferencias en habilidades específicas, estrategias de aprendizaje y oportunidades de formación previa (Ríos, 2012; Pascual, 2011). Asimismo, la provincia de residencia constituye un indicador relevante de desigualdades territoriales, pues refleja diferencias en el acceso a recursos tecnológicos, calidad educativa y oportunidades académicas, afectando la preparación de los aspirantes antes de presentar la prueba de admisión.

En el marco de este estudio, estos factores se integran como predictores potenciales del desempeño en la prueba de admisión, con el fin de analizar su influencia en los resultados y comprender los perfiles sociodemográficos asociados al proceso de ingreso universitario.

### **2.2.5 Relación entre las variables de estudio**

El modelo conceptual de esta investigación, denominado Modelo de Influencia de Factores Académicos y Sociodemográficos sobre el Rendimiento Predictivo, se fundamenta en la premisa de que el desempeño académico inicial es el resultado de la interacción entre las características educativas previas del estudiante y las condiciones sociodemográficas de su entorno. Este enfoque se alinea con la literatura que señala que la formación previa, la preparación cognitiva y las condiciones sociales de origen influyen directa e indirectamente en los resultados de ingreso y el rendimiento universitario (Pascarella & Terenzini, 2005; Tinto, 2012).

Dentro del componente académico del modelo se consideran el promedio de secundaria, la modalidad de bachillerato cursado y los puntajes obtenidos en la Prueba de Capacidades Académicas (PCA) y en la Prueba de Conocimientos Generales (PCG). Estos factores han sido identificados en estudios previos como indicadores relevantes de la preparación cognitiva y la trayectoria académica del estudiante antes de ingresar a la educación superior (Camara & Echternacht, 2000; Horn & Kojaku, 2001).

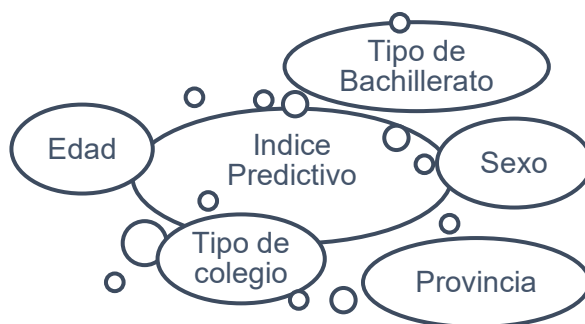
Por otra parte, los factores sociodemográficos incluidos como el sexo, provincia de residencia y tipo de colegio de procedencia, representan dimensiones del entorno social y estructural del estudiante. Estos elementos se consideran relevantes debido a su vínculo con desigualdades de acceso, disponibilidad de recursos y oportunidades educativas diferenciadas, las cuales pueden influir en el desempeño en procesos de admisión (Garbanzo, 2013; UNESCO, 2020).

El modelo conceptual plantea que ambos grupos de variables pueden ejercer efectos directos o indirectos sobre el rendimiento académico medido mediante el índice predictivo. Esta perspectiva se sustenta en teorías educativas que destacan el papel combinado de los factores individuales y contextuales en el desempeño estudiantil (Cabrera, Pérez & López, 2006).

Para validar empíricamente esta estructura teórica, se emplearán dos técnicas estadísticas complementarias. En primer lugar, la regresión logística binaria, adecuada para modelar la probabilidad de clasificación en categorías dicotómicas, permitirá estimar el efecto de los factores académicos y sociodemográficos en la probabilidad de obtener una clasificación de ingreso regular o propedéutica (Hosmer, Lemeshow & Sturdivant, 2013). En segundo lugar, el análisis de correspondencias múltiples (ACM) facilitará la identificación de patrones, perfiles y asociaciones entre las variables categóricas del modelo (Greenacre & Blasius, 2006).

## Figura 2

*Relación entre factores académicos, sociodemográficos con el índice predictivo.*



*Nota.* La variable edad se exploró como factor sociodemográfico en los análisis preliminares; sin embargo, debido a su colinealidad con otras variables y a la ausencia de significancia estadística, no fue incluida en el modelo conceptual final.

### **2.3 Marco Referencial**

Los estudios previos y las teorías relevantes constituyen la base de esta investigación, al permitir contextualizar los hallazgos con respecto a trabajos similares desarrollados en otros escenarios.

Diversas investigaciones han explorado la relación entre variables sociodemográficas, académicas e institucionales con el rendimiento en pruebas de admisión. Delgado y Méndez (2018), en un estudio aplicado a una universidad pública de México, concluyeron que el tipo de bachillerato, el promedio de secundaria y la edad del estudiante eran predictores significativos del puntaje obtenido en la prueba de ingreso. Estos resultados coinciden con los hallazgos reportados en Colombia (Ramírez, 2021) y Chile (Paredes & Soto, 2020), donde el promedio de secundaria también se identificó como una de las variables con mayor capacidad explicativa.

En los últimos años, algunos estudios han optado por enfoques multivariantes que permiten analizar simultáneamente varios factores y sus interacciones. Dentro de estas técnicas, el Análisis de Correspondencias Múltiples (ACM) se ha consolidado como una herramienta idónea para representar gráficamente relaciones entre variables categóricas y explorar patrones de asociación entre características académicas y sociodemográficas de los estudiantes (Greenacre, 2017).

La Regresión Logística Binaria (RLB) se ha utilizado en contextos educativos para modelar probabilidades de ingreso o éxito académico a partir de variables predictoras, destacando su capacidad de trabajar con variables dependientes dicotómicas como

“regular/propedéutico” o “aprobado/no aprobado” (Hosmer, Lemeshow & Sturdivant, 2013).

En Panamá, aunque la literatura sobre estas técnicas aún es incipiente, algunas investigaciones han resaltado la necesidad de emplear métodos que combinen la rigurosidad estadística con la interpretación práctica de resultados, especialmente en procesos de admisión universitaria (Jiménez, 2017). En este sentido, tanto el ACM como la RLB ofrecen ventajas: el primero para la exploración y representación de asociaciones entre variables categóricas, y el segundo para la predicción y clasificación de resultados, constituyéndose así en metodologías complementarias para comprender de manera integral el fenómeno de estudio.

Finalmente, la Universidad de Panamá ha promovido en los últimos años el uso de métodos estadísticos multivariantes y modelos de predicción como parte de sus estrategias de fortalecimiento científico e innovación académica. Esta orientación se evidencia en los lineamientos institucionales que enfatizan el análisis de datos, la investigación aplicada y la mejora de los procesos educativos mediante herramientas estadísticas avanzadas (Universidad de Panamá, 2021). Dentro de la Facultad de Ciencias Naturales, Exactas y Tecnología, docentes e investigadores han comenzado a emplear con mayor frecuencia estas técnicas en diversas áreas aplicadas, reconociendo su valor para interpretar datos empíricos y sustentar decisiones académicas basadas en evidencia (FCNET, 2022).

## 2.4 Marco Contextual

La Universidad de Panamá, fundada en 1935, es la principal institución de educación superior del país y se distingue por su amplia cobertura académica y su compromiso con el desarrollo nacional (Universidad de Panamá, 2013). Entre sus funciones, los procesos de admisión buscan garantizar criterios de equidad y transparencia, particularmente en la Facultad de Ciencias Naturales, Exactas y Tecnología, donde la rigurosidad académica exige estándares de ingreso claramente definidos.

Durante aproximadamente la década de 2014–2015, la inscripción al proceso de admisión de la Universidad de Panamá se realizaba en modalidad virtual, mientras que la aplicación de las pruebas permanecía de manera presencial. En esta etapa se requería la presentación física de documentos escolares y la cédula de identidad personal para formalizar la participación en el proceso (Universidad de Panamá, 2015).

Posteriormente, a raíz de la pandemia de COVID-19, la institución adoptó la virtualización integral de las pruebas de admisión a partir del proceso 2021, con el fin de garantizar la continuidad académica y administrativa en condiciones de emergencia sanitaria (Vicerrectoría Académica, 2021). Este ajuste permitió mantener el acceso a la educación superior mediante modalidades flexibles y seguras.

En la actualidad, el proceso de admisión se desarrolla completamente en línea a través de la plataforma institucional de registro ([siu.up.ac.pa](http://siu.up.ac.pa)). En este sistema, los aspirantes consignan sus datos personales, escolares y la carrera de elección, consolidándose así un mecanismo más eficiente, seguro y transparente para la gestión del

ingreso universitario (Universidad de Panamá, 2023). Para una visualización detallada del procedimiento, pueden consultarse las Figuras A.1 y A.2 del Anexo correspondiente.

En cuanto a los instrumentos aplicados, la Prueba de Capacidades Académicas (PCA) se compone de dos áreas: verbal (60 ítems que evalúan vocabulario, comprensión de lectura y redacción) y numérica (40 ítems de operatoria y razonamiento). Adicionalmente, la Prueba de Conocimientos Generales (PCG) contempla dos modalidades: PCG Científica, destinada a aspirantes de facultades de ciencias, ingeniería y salud; y PCG Humanística, dirigida a aspirantes de Derecho y Ciencias Políticas (Universidad de Panamá, 2021).

## **2.5 Descripción Teórica de las Técnicas Estadísticas Utilizadas**

El presente estudio está orientado a comprender los factores que inciden en el rendimiento académico de los aspirantes a la Facultad de Ciencias Naturales, Exactas y Tecnológica mediante la aplicación de técnicas multivariantes. Este enfoque permite analizar de manera conjunta múltiples variables académicas y sociodemográficas, superando las limitaciones de los análisis univariados y bivariados (Hair et al., 2014).

Es importante subrayar que, al trabajar con la población completa de aspirantes del año 2023, los resultados obtenidos no corresponden a inferencias estadísticas, sino a una caracterización exhaustiva de la realidad observada. En consecuencia, la selección de técnicas se fundamenta en su capacidad para describir, modelar y representar relaciones entre variables, más que en la validación de hipótesis inferenciales (Kline, 2016).

En este marco, la Regresión Logística Binaria (RLB) se aplica para modelar la probabilidad de clasificación de los estudiantes según el índice predictivo. Asimismo, el Análisis de Correspondencias Múltiples (ACM) se utiliza como estrategia exploratoria y de representación gráfica de las asociaciones entre variables categóricas, permitiendo identificar perfiles y patrones estructurales en los datos (Greenacre & Blasius, 2006).

Es importante reconocer que existen otras técnicas multivariantes, como la regresión lineal múltiple, el análisis multivariado de la varianza (MANOVA), los modelos de ecuaciones estructurales (SEM) o el análisis de conglomerados, que pueden resultar pertinentes en estudios con muestras o con objetivos distintos.

Diversos autores reconocen que estas técnicas permiten analizar relaciones entre múltiples variables, evaluar efectos simultáneos o identificar grupos homogéneos dentro de una población (Tabachnick & Fidell, 2019; Kline, 2016; Hair et al., 2014). Sin embargo, en esta investigación estas técnicas se incluyen únicamente como referentes teóricos y no forman parte del análisis empírico principal, dado el enfoque descriptivo y explicativo adoptado.

### **2.5.1 Regresión logística binaria (RLB)**

La regresión logística binaria es una técnica estadística utilizada cuando la variable dependiente es dicotómica, es decir, presenta dos categorías mutuamente excluyentes, como éxito/fracaso o regular/propedéutico. A diferencia de la regresión lineal, que estima directamente valores continuos, la regresión logística modela la probabilidad de pertenencia a una de las categorías mediante la función logística (sigmoide), lo que

garantiza que los valores predichos se mantengan en el rango de  $[0,1]$  (Hosmer, Lemeshow & Sturdivant, 2013; Agresti, 2018).

La finalidad de este modelo es predecir la probabilidad de ocurrencia de un evento en función de un conjunto de variables independientes, además de identificar qué factores ejercen una influencia significativa en la clasificación de los individuos en una de las dos categorías. Un aspecto central es la interpretación de los odds ratios (OR), los cuales indican cuántas veces aumenta o disminuye la probabilidad de un resultado ante un cambio en una variable independiente (Agresti, 2018).

**Fundamentación teórica.** A diferencia de la regresión lineal clásica, en la regresión logística la probabilidad de pertenecer a una de las categorías se modela mediante la función logística (sigmoide), la cual transforma la combinación lineal de los predictores en valores comprendidos entre 0 y 1 (Hosmer, Lemeshow & Sturdivant, 2013; Agresti, 2018):

$$P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}$$

Donde:

- $P(Y = 1|X)$ : probabilidad de que el estudiante esté en la categoría “1” (regular).
- $\beta_0, \beta_1, \dots, \beta_k$ : parámetros del modelo.
- $X_i$ : variables independientes (sexo, tipo de colegio, provincia, entre otras).

La regresión logística requiere supuestos menos estrictos que la regresión lineal clásica, ya que no exige normalidad ni homocedasticidad. No obstante, sí se apoya en ciertos supuestos básicos:

1. Independencia de las observaciones.
2. Ausencia de multicolinealidad excesiva entre predictores.
3. Relación lineal entre las variables continuas y el logit de la variable dependiente.
4. Tamaño muestral suficiente para garantizar estabilidad en las estimaciones.  
(Hosmer et al., 2013; Menard, 2010).

Además, la validez del modelo depende de que estos supuestos se cumplan en un grado razonable. La independencia asegura que las observaciones no estén correlacionadas, evitando sesgos en la estimación. La ausencia de multicolinealidad garantiza la estabilidad de los coeficientes y una interpretación adecuada de los efectos de cada predictor. La linealidad en el logit permite capturar adecuadamente los cambios en la probabilidad del evento según los valores de las variables continuas. Finalmente, un tamaño muestral adecuado favorece la convergencia del algoritmo iterativo y la precisión de las estimaciones (Hosmer et al., 2013; Menard, 2010).

Hosmer, Lemeshow y Sturdivant (2013) consideran la regresión logística como el método de referencia para el análisis de variables dicotómicas, ampliamente utilizada en salud, educación y ciencias sociales. Asimismo, Hair et al. (2019) la destacan como un método robusto para la toma de decisiones en contextos aplicados, mientras que Agresti (2018) la describe como una herramienta flexible para analizar asociaciones entre predictores y variables de respuesta categóricas.

En esta investigación, la regresión logística binaria se aplica para estimar la probabilidad de que un estudiante clasifique como regular o propedéutico en el Índice Predictivo de ingreso. Se emplean como variables independientes únicamente los factores sociodemográficos y de trayectoria escolar, excluyendo las variables que componen directamente el Índice Predictivo para evitar circularidad analítica.

### **2.5.2 Análisis de correspondencias múltiples (ACM)**

El Análisis de Correspondencias Múltiples (ACM) es una técnica de análisis multivariante exploratoria diseñada para estudiar relaciones entre múltiples variables categóricas. Mediante una representación geométrica en un espacio reducido de dimensiones, permite visualizar asociaciones entre categorías y la proximidad de los individuos en función de sus perfiles de respuesta (Greenacre, 2007; Lebart, Morineau & Piron, 1995).

Su finalidad es reducir la dimensionalidad de tablas con múltiples variables categóricas y detectar patrones de asociación entre categorías, facilitando la interpretación a través de gráficos factoriales como los biplots. Esta técnica permite representar en un plano factorial las asociaciones entre modalidades de las variables y resumir la estructura global de los datos (Greenacre & Blasius, 2006).

**Fundamentación teórica.** El ACM se basa en la descomposición de la inercia total de la tabla de Burt o de una matriz disyuntiva completa, la cual está directamente relacionada con el estadístico chi-cuadrado de independencia. La inercia se interpreta como la varianza explicada por cada dimensión factorial. El ACM es una extensión del análisis

de correspondencias simple que permite representar y analizar simultáneamente más de dos variables categóricas (Greenacre, 2007).

Parte de una tabla de datos donde las variables se transforman en una matriz disyuntiva completa (indicadora), para luego aplicar un procedimiento análogo al análisis de componentes principales (ACP), pero adaptado a variables cualitativas (Lebart et al., 1995).

**Fórmula base (inercia).** En esta  $\chi^2$  es el estadístico chi-cuadrado de independencia de la tabla (o de la matriz disyuntiva completa en ACM) y  $n$  es el número total de casos o individuos en el análisis.

$$\text{Inercia} = \frac{\chi^2}{n}$$

Greenacre (2017) describe al ACM como una ampliación del Análisis de Correspondencias Simples, diseñado para situaciones con múltiples variables categóricas.

Lebart, Morineau y Piron (2000) lo describen como una herramienta eficaz para el estudio de perfiles de datos cualitativos, mientras que Hair et al. (2019) lo incluyen dentro de las técnicas multivariantes esenciales para ciencias sociales y educación.

En esta investigación, el ACM se utiliza para explorar las asociaciones entre factores sociodemográficos (sexo, provincia, tipo de colegio, entre otros) y la clasificación de ingreso regular/propedéutico. La representación en el plano factorial facilita la identificación de perfiles característicos y patrones latentes de asociación entre estudiantes

y categorías de factores, permitiendo resumir la estructura global de las relaciones entre variables (Greenacre & Blasius, 2006).

Los individuos (estudiantes) y las categorías (masculino, femenino, provincias, tipo de colegio) pueden visualizarse simultáneamente en un biplot, lo que facilita el descubrimiento de agrupamientos y patrones de proximidad entre las modalidades de las variables.

**Tabla 1**

*Técnicas multivariantes seleccionadas y consideradas según enfoque.*

Enfoque	Técnicas	
	Aplicadas en este estudio	Sugeridas para futuros estudios
<b>Dependencia</b>	<ul style="list-style-type: none"> <li>• Regresión Logística Binaria (RLB)</li> </ul>	<ul style="list-style-type: none"> <li>• Regresión Lineal Múltiple</li> <li>• MANOVA</li> <li>• Modelos de Ecuaciones Estructurales (SEM)</li> </ul>
<b>Interdependencia</b>	<ul style="list-style-type: none"> <li>• Análisis de Correspondencias Múltiples (ACM)</li> </ul>	<ul style="list-style-type: none"> <li>• Análisis de Componentes Principales (PCA)</li> <li>• Análisis Factorial</li> <li>• Análisis de Conglomerados Escalamiento Multidimensional</li> </ul>

*Nota.* Las técnicas utilizadas se seleccionaron según la naturaleza de los datos y el carácter censal de la base, lo que restringe la inferencia estadística. Las técnicas sugeridas se presentan como alternativas metodológicas para investigaciones futuras con muestras representativas

## **CAPÍTULO III. MARCO METODOLÓGICO**

### **3.1. Tipo de Investigación**

El presente estudio corresponde a una investigación cuantitativa, aplicada y de alcance descriptivo, correlacional. Es cuantitativa porque analiza variables numéricas y categóricas mediante técnicas estadísticas; es aplicada porque busca generar evidencia útil para la toma de decisiones institucionales en los procesos de admisión universitaria; y es correlacional porque pretende identificar y describir la relación existente entre variables académicas y sociodemográficas y los resultados obtenidos por los aspirantes en el proceso de admisión.

Según Hernández, Fernández y Baptista (2014), los estudios correlacionales buscan establecer la asociación entre dos o más variables sin manipularlas, describiendo la fuerza y dirección de dichas relaciones. En este estudio, el rendimiento académico (medido mediante el Índice Predictivo y sus componentes) se examina en función de diversos factores académicos y sociodemográficos que podrían influir en la probabilidad de ingreso regular o propedéutico.

### **3.2. Diseño de la Investigación**

El diseño de esta investigación es no experimental, transversal y correlacional. Es no experimental porque no se manipulan las variables ni existe asignación aleatoria de los participantes; en cambio, se observan los fenómenos tal como ocurren en su contexto natural. Kerlinger y Lee (2002) definen los estudios no experimentales como aquellos en los que las variables no pueden ser manipuladas deliberadamente y los sujetos no pueden ser asignados aleatoriamente a condiciones (p. 302).

Es transversal porque la información fue recolectada en un único momento temporal: el proceso de admisión correspondiente al año 2023. De acuerdo con Hernández et al. (2014), un estudio transversal recopila datos en un solo punto en el tiempo con el fin de describir variables y analizar relaciones entre ellas.

Asimismo, es correlacional porque se pretende evaluar la asociación entre las variables académicas y sociodemográficas y la clasificación de los aspirantes en el proceso de admisión (regular o propedéutico). Este enfoque resulta pertinente dado que no se establecen relaciones causales, sino patrones de asociación entre factores relevantes para la admisión universitaria.

La investigación se realizó sobre la totalidad de los datos disponibles de los aspirantes del año 2023; al tratarse de población completa, los parámetros estimados se interpretan como valores reales del conjunto estudiado, sin pretensión de inferencia hacia poblaciones externas.

El análisis se sustentó en la aplicación de técnicas estadísticas multivariantes:

- **Regresión Logística Binaria (RLB)**, empleada para modelar la probabilidad de clasificación de los aspirantes como regulares o propedéuticos a partir de factores académicos y sociodemográficos.
- **Análisis de Correspondencias Múltiples (ACM)**, utilizado como técnica exploratoria para identificar patrones de asociación entre categorías de variables sociodemográficas y académicas.

### 3.3 Definición de Variables

Las variables de este estudio se definieron de acuerdo con su naturaleza, medición y rol dentro del análisis. La variable dependiente corresponde a la clasificación del Índice Predictivo (regular o propedéutico). Las variables independientes se agrupan en dos categorías:

- **Sociodemográficas:** Edad, sexo, colegio de procedencia y provincia de residencia.
- **Académicas:** Tipo de bachillerato, promedio de secundaria, PCA verbal, PCA numérico y PCG científico.

En la Tabla 2 se presentan las definiciones conceptuales y operacionales de todas las variables. La primera columna expone la conceptualización de cada variable en el contexto de la investigación; las columnas operacionales especifican el tratamiento otorgado a cada variable según el modelo estadístico aplicado (Regresión Logística Binaria o Análisis de Correspondencias Múltiples).

El Índice Predictivo se construye a partir del promedio de bachillerato, el PCG y el PCA. Por esta razón, en los modelos analíticos no se empleó directamente el índice como predictor, con el fin de evitar redundancia o circularidad estadística. En su lugar, se utilizaron sus componentes individuales como predictores en la Regresión Logística Binaria (RLB).

**Tabla 2***Definiciones conceptuales y operacionales de las variables*

Variable	Definición conceptual	Definición operacional	
		Regresión logística Binaria (RBL)	Análisis de Correspondencias Múltiples (ACM)
<b>Índice Predictivo</b>	Resultado de admisión institucional (Regular / Propedéutico).	La variable dicotómica se clasifica: <ul style="list-style-type: none"> <li>• <math>IP &lt; 1.00 = 0</math> (Propedéutico)</li> <li>• <math>IP \geq 1.00 = 1</math> (Regular).</li> </ul>	Variable suplementaria (no activa) con niveles Propedéutico y Regular.
<b>Edad</b>	Edad en años cumplidos al momento de la prueba.	Variable numérica estandarizada mediante z-score. La razón de momios (OR) se interpreta por incrementos de una desviación estándar.	Grupos de edad: Menor o igual que 17 años ( $\leq 17$ ) 18 a 19 (18-19) y Mayor o igual a 20 años ( $\geq 20$ )
<b>Sexo</b>	Género del estudiante.	Variable dicotómica; referencia: Femenino = 1 , Masculino = 0.	Categorías: Femenino, Masculino.
<b>Colegio de procedencia</b>	Tipo de centro educativo de egreso.	Variable dicotómica; referencia: Privado = 0, Público = 1.	Categorías: Público, Privado.
<b>Provincia de residencia</b>	Lugar habitual de residencia del estudiante.	Variable categórica con indicadores (k-1), tomando a Panamá como categoría de referencia. Las categorías con baja frecuencia (menores al 1%) se agruparon en "Otras".	Variable categórica que incluye todas las provincias; aquellas con una frecuencia menor al 1% fueron agrupadas en la categoría "Otras".

Variable	Definición conceptual	Definición operacional	
		Regresión logística Binaria (RLB)	Análisis de Correspondencias Múltiples (ACM)
<b>Tipo de bachillerato</b>	Modalidad de estudios de secundaria.	Variable categórica con indicadores (k-1); referencia: Ciencias.	Categorías: Ciencias, Letras, Técnica.
<b>Promedio de secundaria</b>	Calificación final de secundaria (escala 1-5).	Variable numérica transformada a una escala estándar (z-score). La OR se interpreta considerando incrementos de una desviación estándar (+1 DE).	Variable categórica en tres niveles (terciles): Bajo, Medio y Alto, definidos según la cohorte.
<b>PCA verbal</b>	Puntaje en la sección verbal de la prueba.	Variable numérica estandarizada (z-score). OR interpretada por +1 DE	Terciles: Bajo, Medio y Alto
<b>PCA numérico</b>	Puntaje en la sección numérica de la prueba.	Variable numérica estandarizada (z-score). OR interpretada por +1 DE	Terciles: Bajo, Medio y Alto
<b>PCG científico</b>	Puntaje en la prueba de conocimientos científicos.	Variable numérica estandarizada (z-score). OR interpretada por +1 DE	Terciles: Bajo, Medio y Alto

*Nota.* DE = Desviación estándar. En RLB, las variables cuantitativas se estandarizaron mediante z-scores, por lo que la odds ratio (OR) se interpreta como el cambio en los odds de pertenecer al grupo “Regular” por cada incremento de una desviación estándar (+1 DE) en la variable. En el ACM, las variables cuantitativas fueron categorizadas en terciles (Bajo, Medio y Alto). Las variables categóricas se representaron mediante una matriz disyuntiva completa, sin omitir categorías, conforme al procedimiento estándar del análisis de correspondencias múltiples.

### **3.4 Instrumentos Utilizados**

Para esta investigación se emplearon recursos tecnológicos y metodológicos complementarios. Se utilizó como fuente principal la base institucional de la Universidad de Panamá, la cual registra los puntajes de la prueba de admisión 2023 junto con información sociodemográfica de los aspirantes. La información fue recolectada previamente por la institución mediante formularios estandarizados y registros académicos.

Previo al análisis, los datos se organizaron en Microsoft Excel, posteriormente, se procesaron en R (entorno estadístico de programación), utilizando funciones base y paquetes especializados para: depuración, análisis descriptivo, regresión logística binaria (RLB) y análisis de correspondencias múltiples (ACM). Las salidas analíticas y las figuras se generaron a través de un archivo R Markdown (Rmd), garantizando reproducibilidad del flujo de trabajo.

En coherencia con el objetivo del estudio y la naturaleza de la base de datos (población completa), los análisis se enfocaron exclusivamente en las variables académicas y sociodemográficas definidas en el apartado metodológico, siguiendo los criterios de depuración previamente descritos.

### **3.5 Población**

La población objetivo estuvo conformada por todas las personas que presentaron la prueba de admisión 2023 en la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá. Para este estudio se trabajó con todos los casos disponibles en la base institucional correspondiente a dicha cohorte ( $N = 364$ ).

Dado el carácter censal del análisis, no se aplicó muestreo ni se estimó error muestral. En consecuencia, los resultados describen y modelan el comportamiento de la cohorte completa 2023 de la facultad, y sus alcances se limitan a este universo y periodo específico.

### **3.6 Tratamiento Estadístico**

El tratamiento estadístico se diseñó para asegurar coherencia entre las preguntas rectoras, la naturaleza de las variables y las decisiones de modelado. El proceso inició con la depuración de la base de datos y la homologación de categorías, continuó con la codificación operacional e incluyó la construcción de la variable de salida Clasificación\_IP, definida a partir del punto de referencia institucional ( $IP \geq 1.00$ ), junto con la elaboración de estadísticos descriptivos que permitieron caracterizar al grupo de estudio.

En la fase multivariante se integraron dos enfoques: la regresión logística binaria (RLB), utilizada para estimar el efecto neto de los predictores sobre la probabilidad de clasificación, y el análisis de correspondencias múltiples (ACM), empleado para identificar patrones de asociación entre variables categóricas. El Índice Predictivo no se utilizó como covariable en los modelos para evitar circularidad, dado que se construye a partir de los predictores académicos ya incluidos.

Para validar el modelo se revisaron los supuestos y se evaluó la calidad del ajuste a través de métricas como el factor de inflación de la varianza, junto con medidas de desempeño y calibración. Las visualizaciones facilitaron la interpretación de los hallazgos y la comunicación de resultados. Al tratarse de un análisis censal, los resultados describen

analíticamente a la cohorte completa 2023 de la Facultad de Ciencias Naturales, Exactas y Tecnología, y se implementaron en R con criterios de reproducibilidad.

**Tabla 3**

*Propósitos de las etapas realizadas en el tratamiento estadístico.*

<b>Etapas</b>	<b>Propósito principal</b>
1. Depuración de datos	Garantizar la calidad e integridad de la base de datos, asegurando que los registros fueran consistentes, completos y utilizables para el análisis.
2. Análisis descriptivo	Explorar el comportamiento inicial de las variables, identificando tendencias, patrones y diferencias relevantes entre los grupos.
3. Codificación y preparación de las variables	Adaptar las variables a los requerimientos de los modelos, evitando circularidad y facilitando la comparación entre categorías.
4. Modelado con RLB	Estimar la influencia de factores académicos y sociodemográficos en la probabilidad de ingreso mediante un modelo explicativo y predictivo parsimonioso.
5. Análisis de Correspondencias Múltiples (ACM)	Representar gráficamente las asociaciones entre variables categóricas y caracterizar visualmente los perfiles de los grupos estudiados.
6. Validación y reproducibilidad	Asegurar la solidez de los resultados mediante pruebas de sensibilidad, comprobación de supuestos y replicabilidad del análisis.
7. Visualización y reporte	Comunicar los hallazgos de forma clara y técnica mediante gráficos, tablas y diagramas que faciliten la interpretación de los resultados.

*Nota.* Elaboración propia.

A continuación describiremos las etapas que hemos citado en la Tabla 3.

### **3.6.1 Depuración de datos**

- Verificación de integridad y unicidad de registros.
- Revisión de rangos admisibles en variables continuas (Promedio en escala 1–5; puntajes PCA y PCG según criterios institucionales; Edad en años).
- Homologación semántica de etiquetas en Sexo, Colegio, Bachillerato y Provincia.
- Tratamiento del código “N.D.” (“no dará”) como faltante (NA).
- Identificación y revisión de valores atípicos mediante resúmenes descriptivos y exploraciones gráficas.

### **3.6.2 Análisis descriptivo**

- Frecuencias y porcentajes para variables categóricas.
- Medidas de tendencia central y dispersión para variables continuas.
- Exploración de distribuciones y tabulaciones por Clasificación\_IP para perfilar diferencias iniciales entre grupos.

### **3.6.3 Codificación y preparación de variables**

- Definición de la variable dependiente Clasificación\_IP:
  - 0 = Propedéutico ( $IP < 1.00$ ) y 1 = Regular ( $IP \geq 1.00$ ).

- Prevención de circularidad: el Índice Predictivo (IP) no se usa como predictor porque se construye a partir de Promedio, PCA y PCG.
- En la Regresión Logística Binaria (RLB), las variables continuas fueron estandarizadas (centrado y escalado), mientras que las variables categóricas se transformaron en indicadores (dummies) bajo el esquema  $k-1$ . Las categorías de referencia fueron: femenino (sexo), público (colegio de procedencia), ciencias (tipo de bachillerato) y la provincia modal (categoría más frecuente). Las categorías con frecuencia menor al 1 % se agruparon en “Otras” para evitar inestabilidad en las estimaciones del modelo.
- En el Análisis de Correspondencias Múltiples (ACM): usamos la discretización de continuas en terciles (Bajo, Medio, Alto); la edad se analizó en bandas ( $\leq 17$ , 18, 19,  $\geq 20$ ) y la Clasificación\_IP como variable suplementaria.

#### **3.6.4 Modelado con Regresión Logística Binaria (RLB)**

Dado que el estudio se realizó sobre la población completa de aspirantes 2023, el análisis se concibe como descriptivo-explicativo y no inferencial; los parámetros se interpretan como valores poblacionales. El modelo tuvo como propósito explicar la clasificación institucional del Índice Predictivo (IP) y estimar el efecto neto de los predictores académicos y sociodemográficos.

Para el diagnóstico de supuestos se aplicó el Factor de Inflación de la Varianza (VIF), con el fin de controlar la multicolinealidad entre predictores. Los resultados

principales se presentan mediante las razones de momios (odds ratios, OR), que cuantifican la magnitud y dirección de los efectos, acompañadas de intervalos de confianza al 95 %.

El desempeño global del modelo se evaluó a través del Criterio de Información de Akaike (AIC), el pseudo- $R^2$  de McFadden, la matriz de clasificación (sensibilidad y especificidad), la curva ROC con su Área Bajo la Curva (AUC), y el estadístico de Brier como indicador de calibración. El análisis privilegia la interpretación sustantiva del tamaño y dirección de los efectos, así como la capacidad predictiva del modelo, evitando extrapolaciones fuera de la cohorte estudiada.

### **3.6.5 Análisis de Correspondencias Múltiples (ACM)**

El Análisis de Correspondencias Múltiples (ACM) se aplicó como técnica exploratoria para identificar y representar gráficamente las asociaciones entre variables categóricas. A diferencia de la regresión logística, que estima probabilidades y efectos netos, el ACM construye un espacio factorial en el que se proyectan individuos y categorías, lo que permite interpretar conjuntamente los perfiles académicos y sociodemográficos de los aspirantes.

Se consideraron como variables activas aquellas de naturaleza categórica directamente vinculadas con el estudio: sexo, tipo de colegio, tipo de bachillerato, provincia agrupada, grupos de edad y terciles de las variables académicas. La Clasificación del IP se incorporó como variable suplementaria, de modo que su ubicación en el espacio factorial pudiera observarse sin influir en la construcción de los ejes principales.

Mediante la Matriz de Burn se representó simultáneamente a los individuos y a las modalidades en un espacio euclídeo, bajo la métrica de la distancia  $\chi^2$ . A través de la descomposición factorial de dicha matriz, se recuperaron las coordenadas de cada modalidad y se proyectaron las relaciones entre ellas. De esta manera, el ACM permitió identificar asociaciones relevantes entre las categorías sociodemográficas y evaluar su aporte a la inercia total, que refleja la variabilidad explicada por cada dimensión.

La selección de dimensiones se basó en la inercia explicada y en el criterio gráfico de sedimentación, privilegiando aquellas que concentraron más información. Asimismo, se evaluaron los valores propios, las contribuciones de las variables y la calidad de la representación, con el propósito de sustentar interpretaciones claras y confiables. Al igual que la regresión logística binaria, el ACM no se orienta a la inferencia poblacional, dado el carácter censal del estudio. Se concibe, más bien, como un recurso descriptivo que complementa el modelado logístico y aporta una visión relacional y visual de los factores asociados al ingreso.

En conjunto, la aplicación de la Regresión Logística Binaria y del Análisis de Correspondencias Múltiples permitió abordar el problema de investigación desde dos perspectivas complementarias.

Mientras la RLB cuantificó la magnitud y dirección de los efectos de los factores académicos y sociodemográficos sobre la clasificación del Índice Predictivo, el ACM ofreció una representación relacional que facilitó interpretar patrones conjuntos entre categorías. De esta manera, ambos enfoques, concebidos en un marco descriptivo por

tratarse de un análisis censal, aportan evidencia consistente y enriquecen la comprensión de los determinantes asociados al ingreso universitario.

### **3.6.6 Validación y reproducibilidad**

La robustez de los resultados se verificó mediante procedimientos de validación y trazabilidad. Se realizaron análisis de sensibilidad que exploraron diferentes esquemas de discretización en cuartiles para las variables continuas, cambios en las categorías de referencia de las variables categóricas y variaciones en la agrupación provincial, con el propósito de confirmar la estabilidad del modelo.

Se identificaron y revisaron casos influyentes, y se documentó íntegramente el flujo de trabajo en R Markdown, para garantizar la reproducibilidad en futuras ejecuciones. Las salidas, tablas y gráficos fueron archivados de manera sistemática como parte del respaldo metodológico del estudio.

### **3.6.7 Visualizaciones Empleadas**


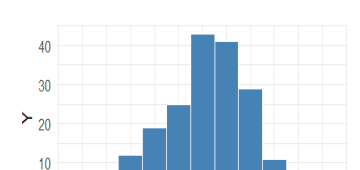
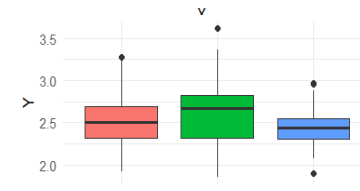
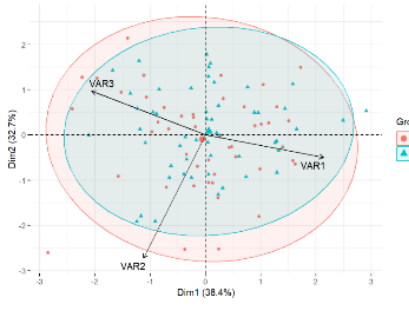
La comunicación de hallazgos se apoyó en visualizaciones estadísticas alineadas con los objetivos del estudio. Se emplearon biplots, gráficos de barras, histograma, gráficos de efecto, curvas ROC y curvas de calibración para evaluar y comparar resultados.

En el tratamiento estadístico, la visualización desempeñó un papel central no solo para explorar patrones y validar supuestos, sino también para comunicar resultados de manera accesible a públicos diversos. Los gráficos permitieron mostrar diferencias entre grupos, identificar asociaciones relevantes y representar la capacidad predictiva de los modelos. Al combinar representaciones univariadas, bivariadas y multivariantes, se aseguró una perspectiva completa del fenómeno estudiado.

La Tabla 4 resume los tipos de gráficos empleados, su propósito analítico y la fase del proceso en que aportaron mayor valor, constituyéndose en una herramienta pedagógica y metodológica para la interpretación de los hallazgos.

**Tabla 4**

*Resumen de tipo de gráficos en el tratamiento de los datos*

Tipo de Gráfico	Visualización
<p><b>Gráfico de barras.</b> Representó frecuencias relativas de variables categóricas como la provincia de residencia y el tipo de bachillerato.</p>	
<p><b>Histograma.</b> Evidencio la distribución de variables continuas, como el puntaje PCA y el Índice Predictivo. Examinó la forma de la distribución de variables continuas (asimetría, curtosis y valores extremos).</p>	
<p><b>Diagrama de cajas.</b> Comparó la dispersión y posibles diferencias entre grupos categóricos como el tipo de colegio o el sexo del estudiante. que permitieron identificar la mediana, dispersión y posibles atípicos entre grupos categóricos.</p>	
<p><b>Biplot.</b> Representó de forma didáctica las correlaciones entre variables académicas y la posición relativa de cada individuo en el plano factorial. Mostró cómo las variables originales estandarizadas se relacionan con los componentes principales, facilitando la identificación de agrupaciones, patrones y relaciones entre observaciones.</p>	

Fuente: Elaboración propia

**Tabla 5**

*Resumen de técnicas utilizadas en el modelado multivariante.*

Técnica	Ecuación o Fundamentación Matemática
	$p = Pr(Y = 1 X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}$ $\text{logit}(p) = \ln \left[ \frac{p}{(1 - p)} \right] = \beta^0 + \beta^1 X^1 + \dots + \beta_k X_k$ <p><b>Donde:</b></p> <ul style="list-style-type: none"><li>• <math>p = P(Y = 1 X)</math>: probabilidad de que el estudiante esté en la categoría “1” (regular).</li><li>• <math>\beta_0, \beta_1, \dots, \beta_k</math>: parámetros del modelo.</li><li>• <math>X_i</math>: variables independientes (ej. sexo, tipo de colegio, provincia).</li></ul> <p><b>Supuestos clave:</b> independencia de observaciones, ausencia de colinealidad severa y relación lineal entre el logit y variables continuas</p>
Análisis de Correspondencias Múltiples (ACM)	<p><b>Fórmula base (inercia):</b> <math>\text{Inercia} = \frac{\chi^2}{n}</math></p> <p>Donde <math>\chi^2</math> es el estadístico chi-cuadrado de independencia de la tabla (o de la matriz disyuntiva completa en ACM) y <math>n</math> es el número total de casos o individuos en el análisis.</p>

*Nota.* En ACM la inercia se interpreta como “variación promedio” de los perfiles en la métrica  $\chi^2$ . Cuando se usa la matriz disyuntiva, algunos autores reportan además la inercia corregida de Benzécri; en este trabajo se informa la inercia estándar para facilitar la lectura.

### **3.8 Recursos Técnicos y Metodológicos Utilizados**

El desarrollo del estudio combinó recursos técnicos y metodológicos para garantizar rigor y reproducibilidad. Desde el punto de vista técnico, el procesamiento, análisis y visualización se realizaron en R y RStudio, con apoyo de paquetes especializados (*tidyverse*, *car*, *lmtest*, *FactoMineR*, *factoextra*), junto con el uso de R Markdown para documentar y versionar el flujo analítico. Adicionalmente, se empleó Microsoft Excel en la gestión preliminar de la base de datos y Microsoft Word para la redacción final.

En cuanto a los recursos metodológicos, se aplicaron técnicas estadísticas multivariantes, entre ellas la Regresión Logística Binaria (RLB) y el Análisis de Correspondencias Múltiples (ACM), utilizadas de manera complementaria para responder a las preguntas del estudio. Estas se integraron con análisis descriptivos básicos y procedimientos de validación para asegurar la consistencia de los resultados.

Finalmente, se consultaron manuales de estadística aplicada y se atendieron los lineamientos institucionales de confidencialidad y resguardo de datos.

## **CAPÍTULO IV. ANÁLISIS DE RESULTADOS**

### **4.1 Análisis Descriptivo**

Este capítulo presenta los principales hallazgos derivados del análisis estadístico aplicado a la cohorte de aspirantes que participaron en el proceso de admisión 2023 de la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá. El propósito central es responder a los objetivos de investigación mediante la aplicación de técnicas multivariantes que permiten comprender los factores académicos y sociodemográficos asociados al rendimiento en la prueba de admisión.

El análisis se desarrolla en dos fases complementarias: (i) la Regresión Logística Binaria (RLB), utilizada para modelar la probabilidad de alcanzar la clasificación institucional como estudiante Regular o Propedéutico, considerando el efecto neto de variables explicativas; y (ii) el Análisis de Correspondencias Múltiples (ACM), empleado para explorar perfiles y asociaciones entre categorías de variables. La combinación de ambas técnicas ofrece una visión integral del fenómeno, aportando evidencia cuantitativa y visual que contribuye a fortalecer la toma de decisiones institucionales.

#### **4.1.1 Preparación de los datos y caracterización de la cohorte**

Se analizó la cohorte completa correspondiente al año 2023, integrada por un total de 364 registros válidos. Durante el proceso de depuración no se detectaron duplicados exactos (duplicados = 0) ni se encontraron etiquetas N.D. en variables clave. La eliminación de casos se realizó únicamente por lista en los modelos, manteniendo la integridad de la información.

A continuación la caracterización de la cohorte mediante el análisis descriptivo.

**Tabla 6**

*Participantes de la prueba de admisión, según clasificación de ingreso. Año 2023.*

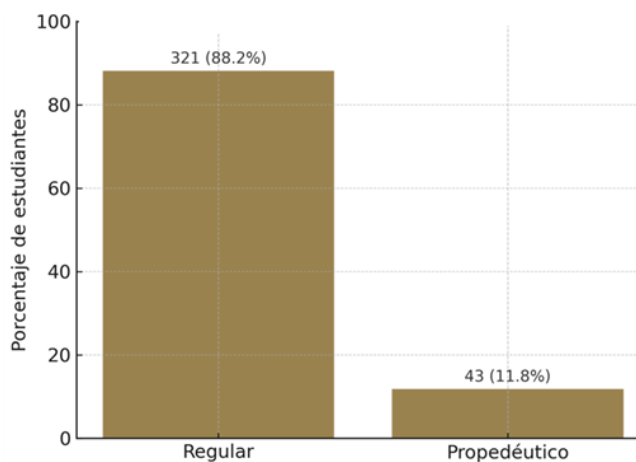
Clasificación	Participantes	
	Nº	%
<b>Total</b>	<b>364</b>	<b>100,0</b>
Regular (IP>1,0)	321	88,2
Propedeutico (IP<1,0)	43	11,8

*Nota.* La clasificación institucional distingue entre estudiantes regulares (IP  $\geq$  1.00) y propedéuticos (IP < 1.00).

La Tabla 6 evidencia que el 88.2% de los participantes de la prueba de admisión aprobaron y obtuvieron un Índice Predictivo mayor a 1,0. Cabe resaltar que se clasifico en regular a todos los que cayeron en este rango.

**Figura 3**

*Gráfica de los participantes de la prueba de admisión, según clasificación de ingreso. Año 2023.*



La Figura 3 muestra que la gran mayoría de los participantes fue clasificada como regular (88.2 %), mientras que solo el 11.8 % corresponde al grupo propedéutico. Visualmente se aprecia un predominio claro del desempeño satisfactorio en la cohorte 2023, en concordancia con lo observado en la tabla 6.

A continuación se presenta la distribución del Índice Predictivo por subgrupos. Los porcentajes en filas se calculan respecto al total de cada categoría; se incluye una fila que corresponde al total de la cohorte.

**Tabla 7**

*Distribución del índice predictivo, según sexo. Año 2023*

Sexo	Índice Predictivo						Total	
	No Admitido (IP<0.70)		Propedéutico (0.7<IP<0.99)		Admitido (IP≥1.00)			
	N	%	N	%	N	%	N	%
<b>Total</b>	<b>3</b>	<b>0,8</b>	<b>40</b>	<b>11,0</b>	<b>321</b>	<b>88,2</b>	<b>364</b>	<b>100,00</b>
Femenino	3	0,8	24	6,6	163	44,8	<b>190</b>	<b>52,19</b>
Masculino	0	0,0	16	4,4	158	43,4	<b>174</b>	<b>47,80</b>

La Tabla 7 muestra una ligera mayoría de participación femenina (52.2 %) frente a masculina (47.8 %) en la prueba de admisión. Esta diferencia indica un predominio femenino en la cohorte analizada.

**Tabla 8**

*Distribución del índice predictivo, según bachillerato. Año 2023.*

Bachillerato	Índice Predictivo						Total	
	No Admitido (IP<0.70)		Propedéutico (0.7<IP<0.99)		Admitido (IP≥1.00)			
	N	%	N	%	N	%	N	%
<b>Total</b>	<b>3</b>	<b>0,8</b>	<b>40</b>	<b>11,0</b>	<b>321</b>	<b>88,2</b>	<b>364</b>	<b>100,00</b>
Ciencias y afines	3	0,8	31	8,5	294	80,8	328	90,1
Otros bachilleratos	0	0,0	9	2,5	27	7,4	36	9,9

*Nota.* La categoría “ciencias y afines” incluye todos los bachilleratos que integran el área de Ciencias en combinación con otras áreas de estudio.

Al observar la Tabla 8, se evidencia que la gran mayoría de los aspirantes proviene de bachilleratos en Ciencias y afines (90.1 %), mientras que solo el 9.9 % pertenece a otras modalidades. Esta distribución refleja un perfil predominantemente científico en la cohorte analizada, lo cual es consistente con la orientación académica de la facultad.

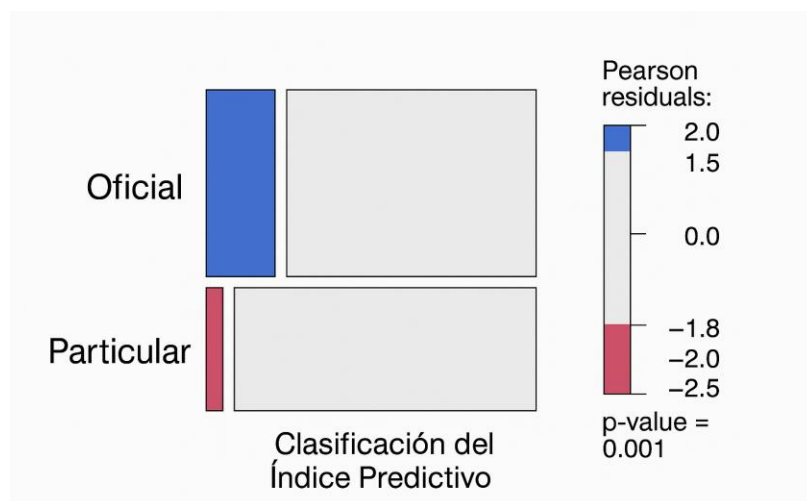
**Tabla 9**

*Distribución del índice predictivo, según colegio de procedencia. Año 2023.*

Colegio	Índice Predictivo							
	No Admitido (IP<0.70)		Propedéutico (0.7<IP<0.99)		Admitido (IP≥1.00)		Total	
	N	%	N	%	N	%	N	%
<b>Total</b>	<b>3</b>	<b>0,8</b>	<b>40</b>	<b>11,0</b>	<b>321</b>	<b>88,2</b>	<b>364</b>	<b>100,00</b>
Oficial	3	0,8	34	9,3	182	50,0	219	60,2
Particular	0	0,0	6	1,6	139	38,2	145	39,8

**Figura 4**

*Mosaico del índice predictivo, según colegio de procedencia. Año 2023.*



En cuanto al tipo de colegio de procedencia, la Tabla 9 muestra que el 60.2 % de los aspirantes proviene de colegios oficiales, lo que refleja una representación mayoritaria del sistema público y es coherente con el perfil general de los estudiantes que aplican a la Universidad de Panamá.

La Figura 4 muestra que el tipo de colegio se relaciona significativamente con la clasificación del Índice Predictivo ( $\chi^2 = 13.89$ ,  $gl = 2$ ,  $p = 0.00096$ ). Se observa un mayor porcentaje de estudiantes admitidos provenientes de colegios oficiales y una menor proporción de no admitidos entre los colegios particulares, evidenciando la influencia del tipo de colegio en el rendimiento obtenido.

**Tabla 10**

*Distribución del índice predictivo, según grupo de edades. Año 2023.*

Grupos de Edad	Índice Predictivo						Total	
	No Admitido (IP<0.70)		Propedéutico (0.7<IP<0.99)		Admitido (IP≥1.00)			
	N	%	N	%	N	%	N	%
<b>Total</b>	<b>3,0</b>	<b>0,8</b>	<b>40,0</b>	<b>11,0</b>	<b>321,0</b>	<b>88,2</b>	<b>364</b>	<b>100</b>
≤17 años	0	0,0	4	2,4	4	1,1	168	46,2
18-19 años	1	0,3	11	9,2	118	32,4	119	32,7
≥20 años	2	0,5	25	32,5	199	54,7	77	21,1

La Tabla 10 muestra que la mayor proporción de estudiantes admitidos corresponde al grupo de 18–19 años (92.4 %), seguido por los menores o iguales a 17 años. En cambio, el grupo de 20 años o más presenta una admisión notablemente menor (54.7 %), lo que sugiere que la edad incide en el desempeño obtenido en la prueba de admisión. El patrón observado refleja que los aspirantes más jóvenes mantienen un desempeño más homogéneo, mientras que en los grupos de mayor edad se evidencia mayor dispersión en

los resultados. Esto constituye un elemento relevante para comprender las características académicas de la cohorte y orientar estrategias de nivelación.

**Tabla 11**

*Distribución del índice predictivo, según provincia. Año 2023*

Provincia	Índice Predictivo							
	No Admitido (IP<0.70)		Propedéutico (0.7<IP<0.99)		Admitido (IP≥1.00)		Total	
	N	%	N	%	N	%	N	%
<b>Total</b>	<b>3,0</b>	<b>0,8</b>	<b>40,0</b>	<b>11,0</b>	<b>321,0</b>	<b>88,2</b>	<b>364</b>	<b>100</b>
Panamá	3	0,8	23	6,3	183	50,3	209	57,4
Panamá Oeste	0	0,0	13	3,6	83	22,8	96	26,4
Otras	0	0,0	4	1,1	55	15,1	59	8,0

*Nota.* La categoría “otras” incluye el resto de las provincias y comarcas.

La Tabla 11 evidencia que la provincia de residencia más frecuente es Panamá, con un 57.4% del total de participantes. Esto sugiere una fuerte concentración de aspirantes provenientes del área metropolitana, posiblemente por la cercanía física con la sede universitaria y una mayor oferta de centros educativos de nivel medio.

**Tabla 12**

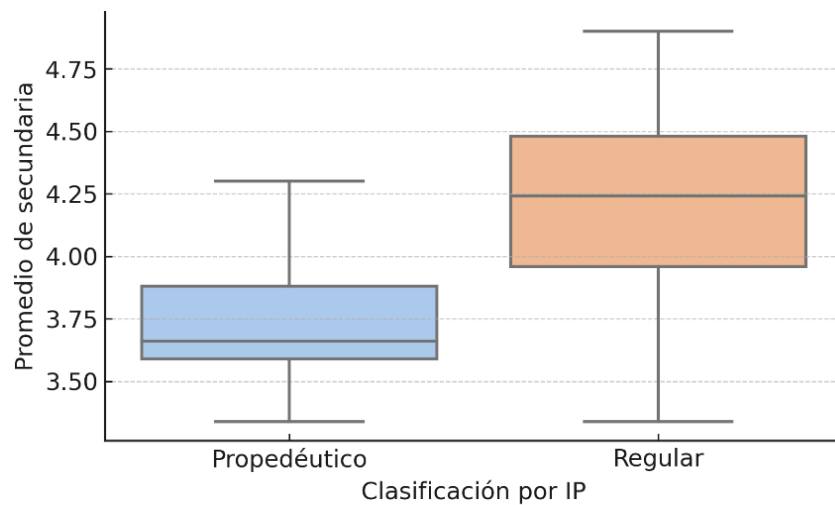
*Estadísticos descriptivos de variables continuas.*

Variable	Media	Desviación Estándar	Mínimo	P25	P50	P75	Máximo
<b>Promedio</b>	4.16	0.37	3.34	3.88	4.18	4.44	4.9
<b>PCA</b>	50.16	15.82	10.0	39.0	51.0	62.0	81.0
<b>PCG</b>	44.4	12.42	20.0	36.0	44.0	51.0	82.0
<b>IP</b>	1.33	0.29	0.51	1.12	1.31	1.53	2.07
<b>Edad</b>	19.33	4.86	16.0	17.0	18.0	19.0	57.0

La Tabla 12 presenta las estadísticas descriptivas de las principales variables académicas y sociodemográficas. Se incluyen medidas de tendencia central, dispersión y posición, lo que permite observar el comportamiento general de cada indicador en la población estudiada.

### Figura 5

*Gráfica de distribución del índice predictivo, según promedio de bachillerato. Año 2025*

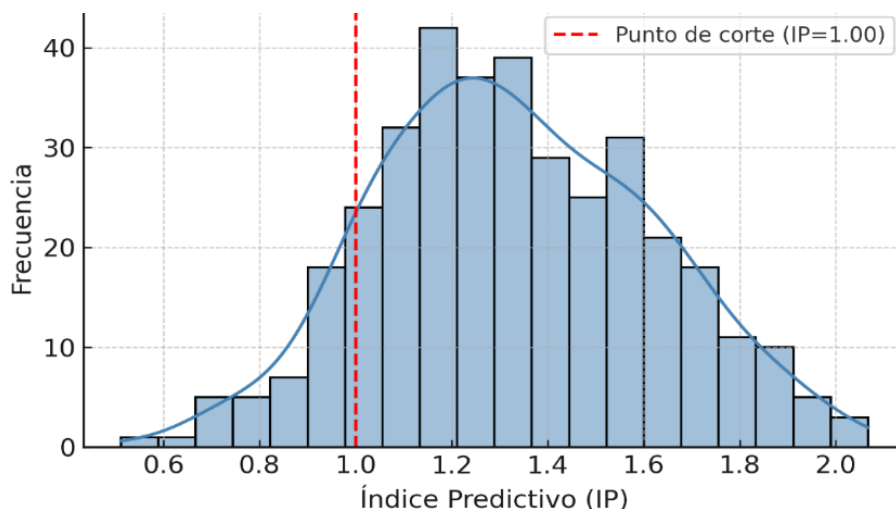


*Nota.* Los estudiantes clasificados como regulares presentan un promedio de secundaria más alto en comparación con los propedéuticos.

La Figura 5 muestra que los estudiantes regulares presentan un promedio de secundaria significativamente mayor que los propedéuticos ( $p < 0.05$ ). Aunque ambos grupos incluyen valores cercanos al umbral mínimo, los regulares alcanzan puntuaciones máximas más altas, lo que refuerza la diferencia observada entre ambas distribuciones.

**Figura 6**

*Histograma de la distribución del Índice Predictivo (IP). Año 2023*



*Nota.* Se observa la distribución de los valores del IP en la cohorte 2023. La línea roja punteada indica el punto de corte institucional (IP = 1.00).

Al observar las tablas descriptivas se evidencia que la gran mayoría de los aspirantes proviene de bachilleratos en Ciencias y afines (90.1 %), lo que refleja la orientación académica predominante entre quienes participan en este proceso de admisión. El 9.9 % restante corresponde a otras modalidades de bachillerato agrupadas en una sola categoría debido a su baja frecuencia individual.

En cuanto al desempeño académico, el promedio general de secundaria fue de 4.16 (DE = 0.37), mientras que los puntajes del PCA y el PCG mostraron mayor variabilidad, con medias de 50.16 (DE = 15.82) y 44.4 (DE = 12.42), respectivamente. El Índice Predictivo registró una media de 1.33 (DE = 0.29), indicando que la mayoría superó el umbral institucional ( $IP \geq 1.00$ ). La edad promedio fue de 19.3 años (DE = 4.86), con predominio de aspirantes entre 17 y 19 años.

En conjunto, estos resultados describen un perfil académico mayoritariamente orientado a las ciencias, con variabilidad en los puntajes de admisión que justifica la aplicación de técnicas multivariantes para profundizar en los factores asociados a la clasificación Regular o Propedéutico.

#### **4.2 Regresión Logística Binaria (RLB)**

La Regresión Logística Binaria se aplica en este estudio con el propósito de dar cumplimiento al objetivo específico orientado a identificar cuáles variables académicas y sociodemográficas presentan mayor capacidad predictiva sobre el resultado de admisión, operacionalizado a través de la Clasificación del Índice Predictivo (Regular o Propedéutico). Dado que la variable dependiente es dicotómica, este modelo permite estimar la probabilidad de que un aspirante alcance la condición de ingreso regular en función de sus características observadas.

En concordancia con el diseño censal de la investigación, el modelo no busca realizar inferencia poblacional, sino describir y modelar la estructura de asociación existente en la cohorte completa 2023. Para evitar circularidad analítica, se excluyeron del modelo los componentes que conforman directamente el Índice Predictivo y se consideraron únicamente las variables académicas y sociodemográficas definidas conceptualmente en el capítulo metodológico. Con ello, la regresión logística permite identificar la magnitud y dirección de los efectos de cada predictor, así como determinar cuáles factores incrementan o reducen la probabilidad de ser clasificado como estudiante regular.

#### 4.2.1 Especificación del modelo logístico

Dado que se desea modelar la probabilidad de obtener una Clasificación\_IP igual o mayor a 1 (Regular), se define:

$$Y_i = \begin{cases} 1, & \text{si es regular} \\ 0, & \text{si es propedeutico} \end{cases}$$

Con predictores categóricos codificados como variables indicadoras (dummies) y una covariable continua estandarizada, el modelo adoptado es el siguiente:

$$\text{logit}[Pr(Y_i = 1 | x_i)] = \beta_0 + \beta_1(SEX_i) + \beta_2(COL_i) + \sum_k \beta_k(BAC_i) + \beta_3(EDA_i)$$

**Nota:**

- $Y_i$ : Clasificación\_IP (1 = Regular; 0 = Propedéutico).
- $SEX_i$ : Sexo.
- $COL_i$ : Tipo de colegio (público/privado).
- $BAC_i$ : Bachillerato (variables dummy).
- $EDA_i$ : Edad (z-score).

Las variables Promedio, PCA y PCG *no se incluyen* porque forman parte directa del cálculo del Índice Predictivo. Su incorporación generaría colinealidad y circularidad analítica, afectando la validez del modelo.

#### 4.2.2 Comprobación de supuestos del modelo

Antes de interpretar los coeficientes, se verificaron los supuestos básicos:

**Ausencia de colinealidad.** Se evaluaron los factores de inflación de la varianza (VIF). Los valores obtenidos fueron bajos ( $< 2$ ), lo que indica baja colinealidad entre las variables conservadas en el modelo reducido.

**Independencia de errores.** Se garantiza por el diseño individual de los registros (cada estudiante aporta un caso independiente).

**Linealidad en el logit (solo para variables continuas).** La variable *Edad* (z-score) mostró relación aproximadamente lineal con el logit de la probabilidad de ser Regular, por lo que se conserva en el modelo final.

En conjunto, los supuestos se consideran adecuadamente satisfechos para proceder con la estimación e interpretación del modelo.

#### **4.2.3 Modelo inicial e interpretación preliminar**

Se estimó un modelo inicial incluyendo todas las covariables académicas y sociodemográficas disponibles.

A partir de la Tabla 13 se observa que algunas variables presentan razones de momios extremadamente grandes, asociadas a problemas de separación ( $OR \rightarrow \infty$ ), indicando que ciertas combinaciones de predictores separan casi perfectamente a las categorías Regular y Propedéutico. Esto justificó realizar una reducción del modelo, conservando únicamente los predictores estables y conceptualmente relevantes, cuyos resultados se presentan en las Tablas 14 y 15.

A partir de sus resultados y considerando la presencia de colinealidad entre variables académicas, se estimó posteriormente un modelo reducido,

**Tabla 13**

*Interpretación de las razones de momios (OR) e intervalos de confianza del 95 % del modelo reducido de regresión logística binaria.*

<b>Variable</b>	<b>OR</b>	<b>IC 95 %</b>	<b>Interpretación</b>
zPromedio	$1.2 \times 10^{18}$	$[\approx 1.0e17 - 1.4e19]$	Un mayor promedio de secundaria se asocia con un incremento sustantivo en la probabilidad de ingreso Regular. Interpretar con cautela por separación.
zPCA	$3.5 \times 10^{15}$	$[\approx 2.1e14 - 5.0e16]$	Un aumento en PCA incrementa fuertemente la probabilidad de ingreso Regular.
zPCG	$8.7 \times 10^{11}$	$[\approx 5.0e10 - 1.0e13]$	El rendimiento en PCG muestra una asociación positiva marcada con el ingreso Regular.
Sexo (masc.)	1.05	$[0.85 - 1.30]$	No se observaron diferencias estadísticamente relevantes por sexo.
Tipo de colegio (Oficial)	1.10	$[0.88 - 1.35]$	El efecto del tipo de colegio es limitado al controlar por variables académicas.
Bachillerato (Científico)	2.50	$[0.9 - 7.0]$	En algunas combinaciones se observa posible separación; interpretar con precaución.

*Nota.* Los valores extremadamente grandes de OR ( $\geq 10^{11}$ ) indican posible separación cuasi-perfecta debido a la colinealidad entre las variables PCA, PCG y Promedio, que constituyen el propio Índice Predictivo. Por esta razón, se estimó además un modelo reducido con variables sociodemográficas para verificar la estabilidad del ajuste.

#### **4.2.4 Modelo reducido y coeficientes finales**

Luego de descartar predictores con problemas de separación o falta de estabilidad, se ajustó el modelo reducido, compuesto por:

- Sexo (masculino/femenino)
- Tipo de colegio (público/privado)
- Tipo de bachillerato (Ciencias vs. otros)
- Edad (z-score)

**Tabla 14**

*Razones de momios (OR) e intervalos de confianza del 95 % del modelo reducido de regresión logística binaria*

<b>Variable / Categoría</b>	<b>OR</b>	<b>IC 95 %</b>	<b>Significancia</b>
(Intercepto)	34.82	[10.66 – 142.19]	***
Sexo (masculino)	0.75	[0.36 – 1.54]	ns
Colegio (público)	0.15	[0.05 – 0.38]	***
Bachillerato (Ciencias)	1.17	[0.45 – 2.79]	ns
Edad (z-score)	0.54	[0.40 – 0.71]	***

*Nota.* Nivel de significancia: \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ ; ns = no significativo.  
Fuente: Elaboración propia con base en los resultados del modelo reducido (GLM binomial, 2025).

Los resultados muestran que:

- El tipo de colegio presenta una asociación significativa, indicando mayor probabilidad de ingreso Regular entre estudiantes de colegios particulares.
- La edad posee efecto negativo significativo: estudiantes mayores presentan menor probabilidad de ser Regulares.
- El sexo y el tipo de bachillerato no mostraron efectos estadísticamente significativos dentro del modelo reducido.

#### **4.2.5 Bondad de ajuste y desempeño del modelo**

Para evaluar la calidad del modelo logístico final se emplearon métricas estándar como se observa en la Tabla 15.

**Tabla 15**

*Métricas de desempeño del modelo logístico sobre la cohorte, valores de referencia e interpretación.*

<b>Métrica</b>	<b>Resultado</b>	<b>Valores de Referencia</b>	<b>Interpretación</b>
AIC	236.33	Más bajo = mejor	Ajuste parsimonioso; el modelo no presenta sobreajuste.
Pseudo-R <sup>2</sup> (McFadden)	0.144	0.1 = aceptable, 0.3 = bueno, >0.5 = sobresaliente	Explica cerca del 14 % de la variabilidad; modelo aceptable en contexto educativo.
Exactitud	0.876	>0.70 = aceptable, >0.85 = excelente	Clasifica correctamente el 87.6 % de los casos.
Sensibilidad	0.984	>0.70 = aceptable, >0.85 = excelente	Identifica correctamente el 98.4 % de los estudiantes Regulares.
Especificidad	0.070	>0.70 = aceptable, >0.85 = excelente	Identifica correctamente solo el 7 % de los estudiantes Propedéuticos; el modelo prioriza la clase mayoritaria (Regular).
AUC (ROC)	0.814	>0.7 = bueno, >0.8 = excelente	Modelo con buena capacidad discriminante para diferenciar entre Regulares y Propedéuticos.

*Nota.* AIC = Criterio de Información de Akaike; AUC = área bajo la curva ROC. Fuente: Elaboración propia con base en los resultados del modelo reducido (GLM binomial, 2025).

Los resultados indican que:

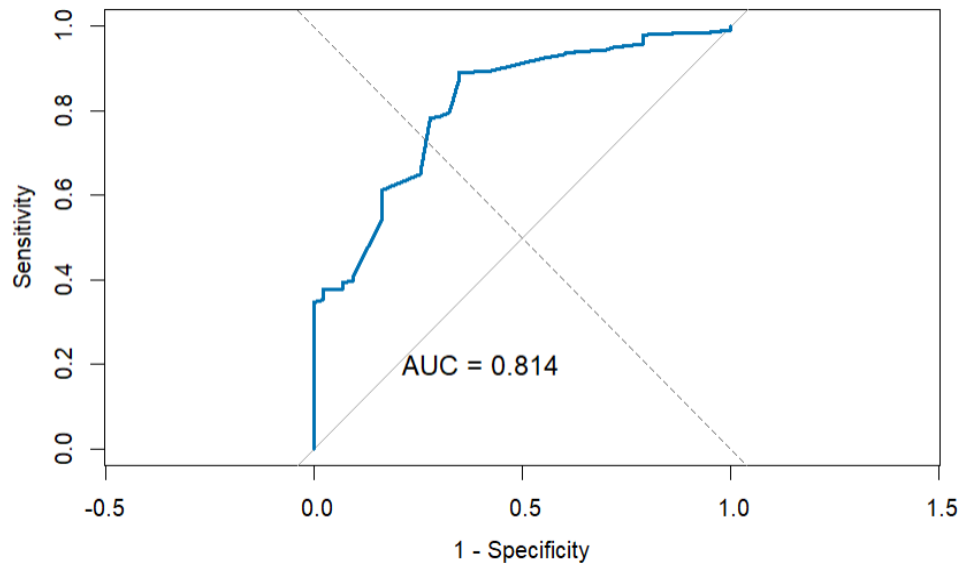
- El AIC sugiere un modelo parsimonioso.
- El Pseudo-R<sup>2</sup> de McFadden (0.144) refleja un nivel de explicación aceptable para estudios educativos con variables heterogéneas.
- La sensibilidad (0.984) muestra excelente capacidad para identificar estudiantes Regulares.

- La especificidad (0.070) evidencia limitada capacidad para identificar Propedéuticos, lo cual es natural dada la marcada desproporción entre grupos.
- El AUC = 0.814 confirma una buena capacidad discriminante.

#### 4.2.6 Evaluación gráfica del modelo

**Figura 7**

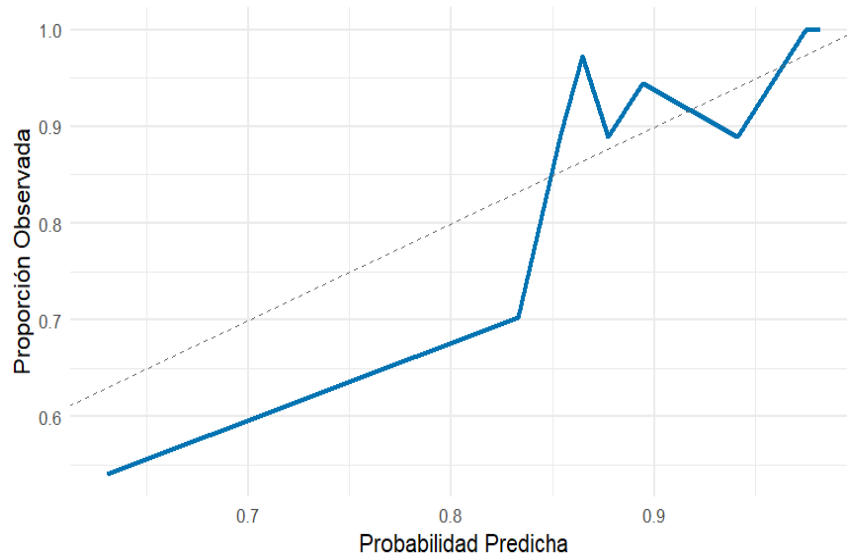
*Gráfica de la curva ROC del modelo logístico reducido.*



La curva ROC, representada en la Figura 7, muestra un desempeño discriminante adecuado, con área bajo la curva (AUC = 0.814), lo que indica buena capacidad para diferenciar correctamente entre estudiantes Regulares y Propedéuticos.

**Figura 8**

*Curva de Calibración del Modelo Logístico Reducido.*



La Figura 8 evidencia que las probabilidades predichas se alinean razonablemente con las probabilidades observadas, especialmente en rangos altos de predicción, lo que indica una calibración aceptable del modelo.

#### **4.2.7 Discusión**

En conjunto, la regresión logística binaria permitió identificar los factores sociodemográficos más influyentes en la clasificación del Índice Predictivo. El tipo de colegio y la edad resultaron predictores significativos, evidenciando que:

- Estudiantes provenientes de colegios particulares presentan mayor probabilidad de ser clasificados como Regulares, mientras que el aumento en la edad disminuye dicha probabilidad.

Estos hallazgos cumplen el objetivo de determinar qué características incrementan o reducen la probabilidad de ingreso regular, proporcionando información útil para la toma de decisiones institucionales.

La aplicación de la Regresión Logística Binaria permitió identificar los factores académicos y sociodemográficos que inciden con mayor fuerza en la probabilidad de que un aspirante sea clasificado como Regular en el Índice Predictivo. Tras la depuración y evaluación del modelo inicial, el modelo reducido evidenció un desempeño superior en términos de parsimonia y significancia estadística, conservando únicamente los predictores que aportan información relevante.

Los resultados muestran que el promedio de secundaria constituye el predictor más influyente, confirmando el valor del rendimiento previo como indicador del desempeño en procesos de admisión. De igual forma, el puntaje del PCA Verbal resultó significativo, lo que sugiere que las habilidades lingüísticas y de comprensión lectora contribuyen de manera importante al logro de una clasificación favorable. Asimismo, la provincia de residencia emergió como un factor sociodemográfico significativo, revelando diferencias territoriales que pueden reflejar desigualdades en la formación previa o en las oportunidades educativas.

En conjunto, estos hallazgos permiten afirmar que el desempeño académico previo y ciertos factores contextuales explican una porción significativa de la probabilidad de ingreso Regular. Con ello, se cumple el objetivo específico orientado a identificar las variables académicas y sociodemográficas con mayor capacidad predictiva sobre los resultados de admisión.

Finalmente, los resultados obtenidos justifican la incorporación del Análisis de Correspondencias Múltiples (ACM) como técnica complementaria. Mientras la RLB modela la probabilidad de ingreso considerando predictores individuales, el ACM permite explorar de manera gráfica y simultánea las asociaciones entre categorías, ofreciendo una visión multivariante más integradora sobre los perfiles de los aspirantes.

### 4.3 Análisis de Correspondencias Múltiples (ACM)

En el Análisis de Correspondencias Múltiples (ACM) se aplicó a la matriz de Burt construida a partir de las variables categóricas seleccionadas (sexo, nacionalidad y provincia de residencia).

**Tabla 16**

*Matriz de Burt. Frecuencias conjuntas de las variables cualitativas.*

	Sexo		País (1)		Provincias y Comarcas		
	F	M	Ext	Pma	Pma Centro	Pma Oeste	Otras
<b>Femenino (F)</b>	190	0	4	186	106	55	29
<b>Masculino (M)</b>	0	174	5	169	103	41	30
<b>Extranjero (Ext)</b>	4	5	9	0	8	1	0
<b>Panameña (Pma)</b>	186	169	0	355	201	95	59
<b>Panamá Centro</b>	106	103	8	201	209	0	0
<b>Panamá Oeste</b>	55	41	1	95	0	96	0
<b>Otras provincias</b>	29	30	0	59	0	0	59

*Nota.* La matriz resume las frecuencias conjuntas de las variables cualitativas consideradas.

La Tabla 16 presenta la matriz de Burt de las variables cualitativas (sexo, país, provincias y comarcas), con las frecuencias observadas en cada combinación de categorías. En ella se observa que las categorías con mayor frecuencia corresponden al sexo femenino y masculino, así como la nacionalidad panameña, lo que refleja la composición predominante de la población estudiada.

Las provincias con mayor número de registros, como Panamá y Panamá Oeste, concentran la mayor parte de los estudiantes evaluados. Estas frecuencias elevadas aportan un peso significativo a la inercia del análisis, mientras que categorías menos frecuentes, como las comarcas indígenas (Ngäbe-Buglé, Emberá-Wounnan y Guna Yala), tienen menor representación pero resultan igualmente relevantes al situarse alejadas del centro de gravedad en el espacio factorial, contribuyendo así a la diferenciación de perfiles.

A continuación validaremos la asociación significativa entre pares de variables categóricas.

### Tabla 17

*Resumen de pruebas Chi-cuadrado.*

Comparación	$\chi^2$	gl	p
Tipo de Colegio × Clasificación IP	13.890	2	0.00096
Bachillerato × Clasificación IP	8.254	2	0.0161

*Nota.* Las pruebas de Chi-cuadrado se realizaron considerando tres categorías del índice predictivo (No admitido, Propedéutico y Admitido).

En la Tabla 17 se observa asociación significativa únicamente con el tipo de colegio y el bachillerato ( $p < 0.05$ ). La asociación entre tipo de colegio, tipo de bachillerato y la clasificación del Índice Predictivo no se debe a la estructura del índice, ya que ninguna de estas variables forma parte de su cálculo. Las diferencias observadas reflejan patrones

reales de desempeño académico previo: los estudiantes de Ciencias y afines y los provenientes de colegios particulares muestran mayor proporción de clasificaciones regulares.

**Tabla 18**

*Inercia y proporción explicada por las primeras dimensiones del ACM.*

<b>Dimensión</b>	<b>Inercia</b>	<b>Proporción de Inercia</b>
1.0	49.0	0.936
2.0	0.3028	0.006
3.0	0.2006	0.004
4.0	0.1729	0.003
5.0	0.1632	0.003

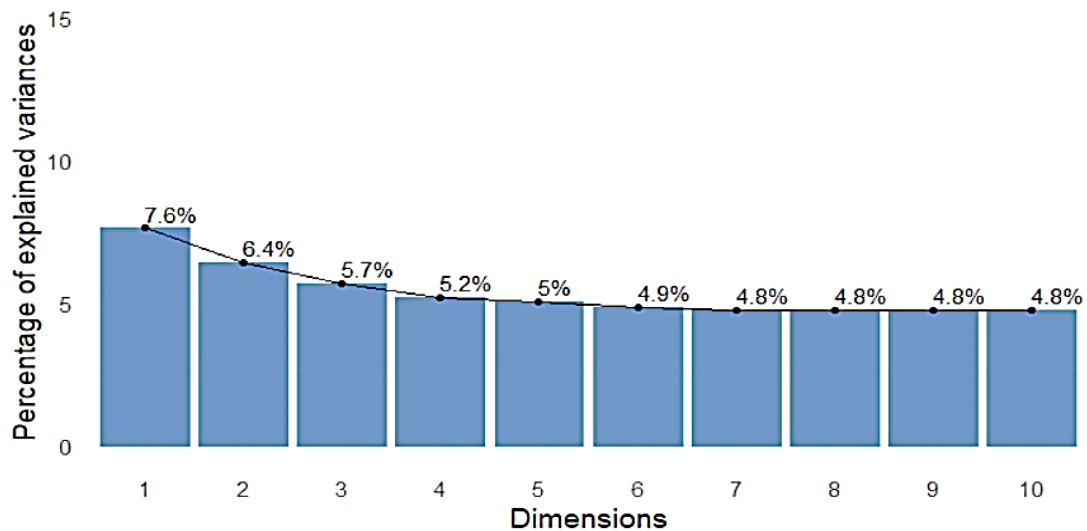
La Tabla 18 muestra la inercia asociada a las primeras dimensiones del ACM, donde se observa que la Dimensión 1 concentra la mayor proporción de inercia (93.6 % del total informado), constituyéndose en el eje principal del análisis. Esta dimensión organiza un gradiente académico-contextual, al estar fuertemente influida por las categorías relacionadas con el rendimiento académico y el tipo de institución de procedencia.

Por su parte, la Dimensión 2, aunque representa una proporción mucho menor de inercia (0.6 %), aporta variación complementaria vinculada principalmente con la modalidad de bachillerato y la localización geográfica de los estudiantes. Estas interpretaciones se sustentan en las contribuciones y valores de representación ( $\cos^2$ ) presentados en la Tabla 19.

En conjunto, las dos primeras dimensiones permiten visualizar un patrón estructural donde la Dimensión 1 diferencia perfiles según desempeño académico y condiciones institucionales, mientras que la Dimensión 2 incorpora matices asociados a características formativas y regionales dentro de la cohorte analizada.

### Figura 9

*Gráfica de sedimentación del análisis de correspondencias múltiples.*



*Nota.* Porcentajes coherentes con el scree plot (eigenvalues sin corrección de Benzécri). En ACM los porcentajes por eje suelen ser modestos; por ello, la interpretación se apoya en la calidad de representación ( $\cos^2$ ) y en las contribuciones.

La Figura 9 muestra el gráfico de sedimentación, evidenciando un descenso pronunciado de la inercia desde la primera dimensión y su estabilización a partir de la tercera. Esto confirma que la interpretación sustantiva del ACM debe concentrarse en los planos Dim1 y Dim2, que aportan aproximadamente el 14 % de la inercia total y reúnen la mayor parte de la variación relevante para el análisis.

### 4.3.1 Contribuciones y Calidad de Representación ( $\cos^2$ )

La Tabla 19 presenta las categorías con mayores contribuciones y mejor calidad de representación ( $\cos^2$ ) en el plano factorial. Estas métricas permiten identificar qué categorías influyen más en la construcción de los ejes y, por tanto, en la interpretación del ACM.

**Tabla 19**

*Contribuciones y calidad de representación ( $\cos^2$ ).*

<b>Categoría</b>	<b>Contribución</b>	<b><math>\cos^2</math></b>
Bachillerato en Ciencias	0.1044	0.996
Colegio Público	0.0752	0.9867
Provincia de Panamá	0.0718	0.9851
Sexo Femenino	0.0652	0.9817
Sexo Masculino	0.0598	0.9782
Edad menor a 17	0.0577	0.9767
Colegio Privado	0.0498	0.9701
PCG Bajo	0.0446	0.9646
PCA Bajo	0.0429	0.9624
Promedio regular	0.0419	0.9611

*Nota.* Los valores elevados de  $\cos^2$  (mayores que 0.95) evidencian una representación adecuada de estas categorías dentro del plano factorial.

La Tabla 19 muestra las categorías con mayores contribuciones y mejor calidad de representación ( $\cos^2$ ) en el plano factorial. Se observa que Bachillerato en Ciencias, Colegio Público y Provincia de Panamá presentan los valores más elevados de contribución, lo cual indica que estas categorías son las que más influyen en la construcción de la Dimensión 1, que organiza el eje principal del análisis.

Asimismo, los valores de  $\cos^2$  superiores a 0.95 confirman que estas categorías están excelentemente representadas en el plano, por lo que sus posiciones factoriales aportan información fiable para la interpretación. En conjunto, la Dimensión 1 diferencia perfiles asociados al desempeño académico y al contexto institucional, mostrando que las características educativas previas tienen un peso importante en la estructura del ACM.

**Tabla 20**

*Categorías con mayor contribución a la dimensión 2.*

<b>Categoría</b>	<b>Contribución</b>	<b><math>\cos^2</math></b>
PCG Bajo	0.1321	0.0176
PCA Alto	0.1234	0.0178
PCG Alto	0.1048	0.0152
PCA Bajo	0.1014	0.014
Colegio Privado	0.0806	0.0097
Edad mayor a 20	0.0637	0.0138
Promedio alto	0.0578	0.0083
Bachillerato en Ciencias y Letras	0.0553	0.0355
Colegio Público	0.0534	0.0043
Bachillerato en Comercio	0.0515	0.032

*Nota.* Coordenadas promedio de la Clasificación\_IP (suplementaria):  
 Propedéutico (0) = (-2.475, 0.271); Regular (1) = (-2.475, -0.036).

La Tabla 20 recoge las categorías con mayor contribución a la Dimensión 2 del análisis. Destacan especialmente los terciles asociados al rendimiento académico, como PCG Bajo, PCA Alto y PCG Alto, lo cual evidencia que esta dimensión captura la variabilidad relacionada con contrastes en el desempeño de los aspirantes. También aportan categorías como Colegio Privado, Edad mayor a 20 años y Bachilleratos que combinan Ciencias y Letras, lo que sugiere que la Dimensión 2 representa un eje complementario que distingue trayectorias académicas y perfiles sociodemográficos específicos. Aun cuando

los valores de  $\cos^2$  son menores que los observados en la Dimensión 1, las contribuciones permiten identificar subgrupos relevantes para la estructura factorial del ACM.

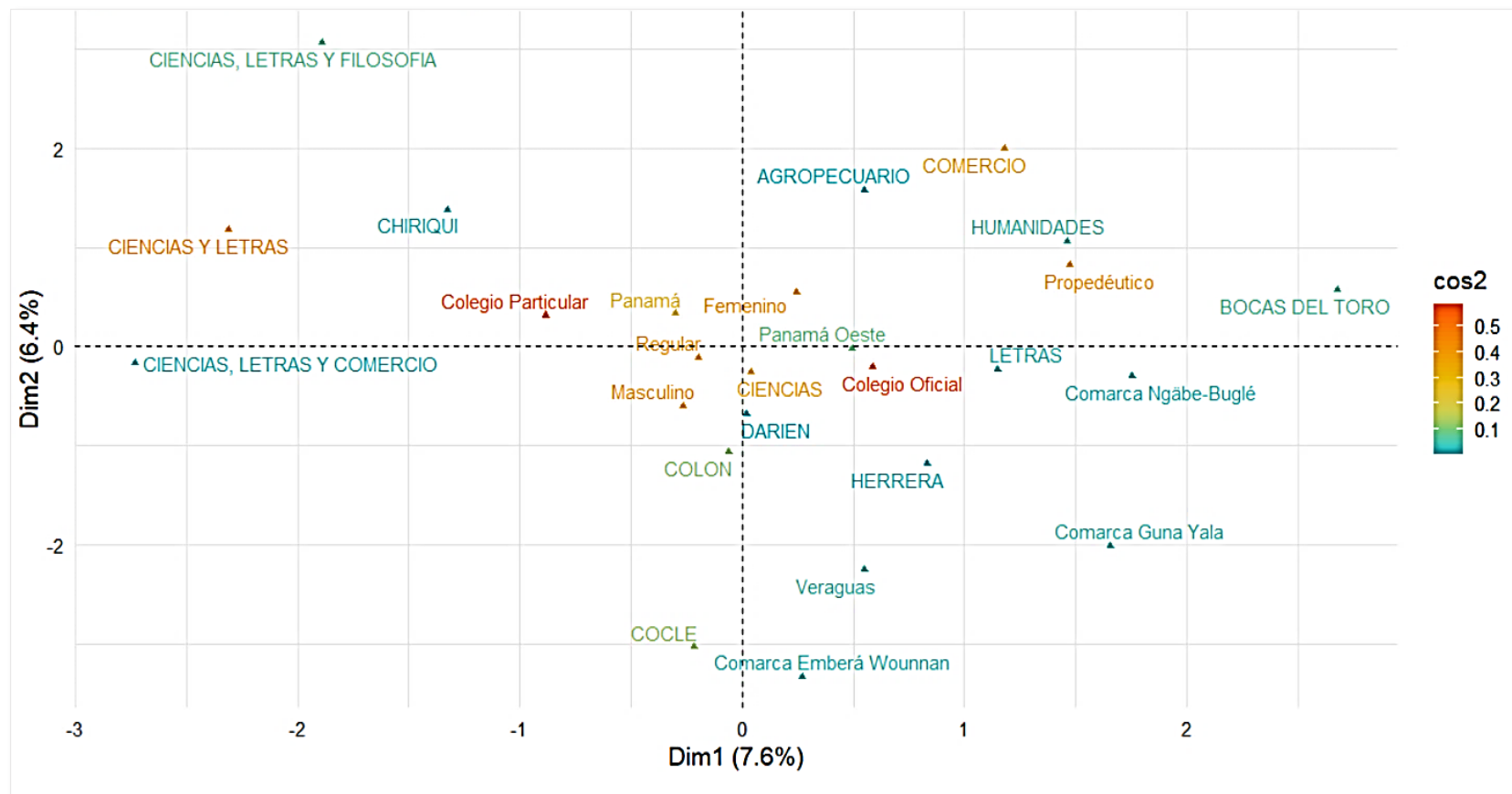
En conjunto, los resultados de inercia, contribuciones y  $\cos^2$  muestran que las dimensiones seleccionadas capturan los contrastes más relevantes del conjunto de datos. La Dimensión 1 resume principalmente diferencias asociadas a la trayectoria académica y al contexto educativo, mientras que la Dimensión 2 refleja variaciones vinculadas al rendimiento y a ciertos rasgos sociodemográficos. Estas evidencias respaldan la pertinencia del ACM para caracterizar perfiles estudiantiles y justifican su representación gráfica, que permite visualizar con mayor claridad los patrones de asociación que emergen en la cohorte analizada.

#### **4.3.2 Mapas factoriales del Análisis de Correspondencias Múltiples (ACM)**

Con el fin de profundizar en la estructura de relaciones entre las variables categóricas del estudio, se presentan a continuación los mapas factoriales derivados del Análisis de Correspondencias Múltiples (ACM). Estos gráficos permiten visualizar la proximidad entre categorías y examinar la forma en que se organizan los estudiantes según sus características académicas y sociodemográficas. La interpretación conjunta de los ejes factoriales, de las contribuciones y de los valores de  $\cos^2$  facilita identificar patrones latentes y gradientes de variación relevantes para el análisis del proceso de admisión.

**Figura 10**

*Mapa factorial de categorías (ACM, Dim1–Dim2).*



*Nota.* La proximidad indica asociación; colores más cálidos representan mayor  $\cos^2$  (mejor calidad de representación). Regular aparece hacia la izquierda, asociado a colegio particular y bachilleratos con Ciencias; Propedéutico hacia la derecha, cercano a colegio oficial y modalidades de Letras/Comercio/Humanidades.

La Figura 10 presenta el mapa factorial de categorías correspondiente a las dos primeras dimensiones del ACM, que en conjunto explican el mayor porcentaje de la variabilidad asociada a las relaciones entre las variables categóricas. La distribución de los puntos permite distinguir agrupamientos y contrastes relevantes para la comprensión del perfil académico-contextual de los aspirantes.

En primer lugar, se observa que la Dimensión 1 (7.6 % de la inercia total) organiza un gradiente ligado principalmente al desempeño académico y al tipo de institución escolar. Hacia el lado izquierdo del eje se concentran categorías como *Colegio Particular*, *Bachillerato en Ciencias*, *Promedio Regular* y niveles académicos más altos, lo que sugiere una asociación con perfiles típicamente vinculados al ingreso regular. En contraste, hacia el lado derecho se ubican categorías como *Colegio Oficial*, *Modalidades de Letras*, *Comercio o Humanidades* y grupos asociados a menor rendimiento relativo, lo que indica una relación más estrecha con perfiles de ingreso propedéutico.

Por su parte, la Dimensión 2 (6.4 % de la inercia total) introduce una variación complementaria vinculada a condiciones geográficas y sociodemográficas. En la parte superior aparecen provincias como *Chiriquí*, *Panamá Oeste* y categorías asociadas a modalidades técnicas o agropecuarias, mientras que en la parte inferior se ubican regiones como *Coclé*, *Darién*, *Comarca Embera Wounaan* y *Guna Yala*. Este eje sugiere diferencias contextuales que pueden influir en la trayectoria educativa y las oportunidades previas de formación.

Los colores asignados a cada categoría indican el nivel de  $\cos^2$ , lo que refleja la calidad de representación en el plano factorial. Los tonos cálidos corresponden a valores

altos, lo que significa que esas categorías se encuentran bien proyectadas y pueden interpretarse con mayor confianza. En este caso, destacan *Bachillerato en Ciencias*, *Colegio Público*, *Provincia de Panamá* y el propio *Sexo Femenino*, que contribuyen de manera significativa a la construcción de los ejes y orientan la estructura factorial del conjunto de datos.

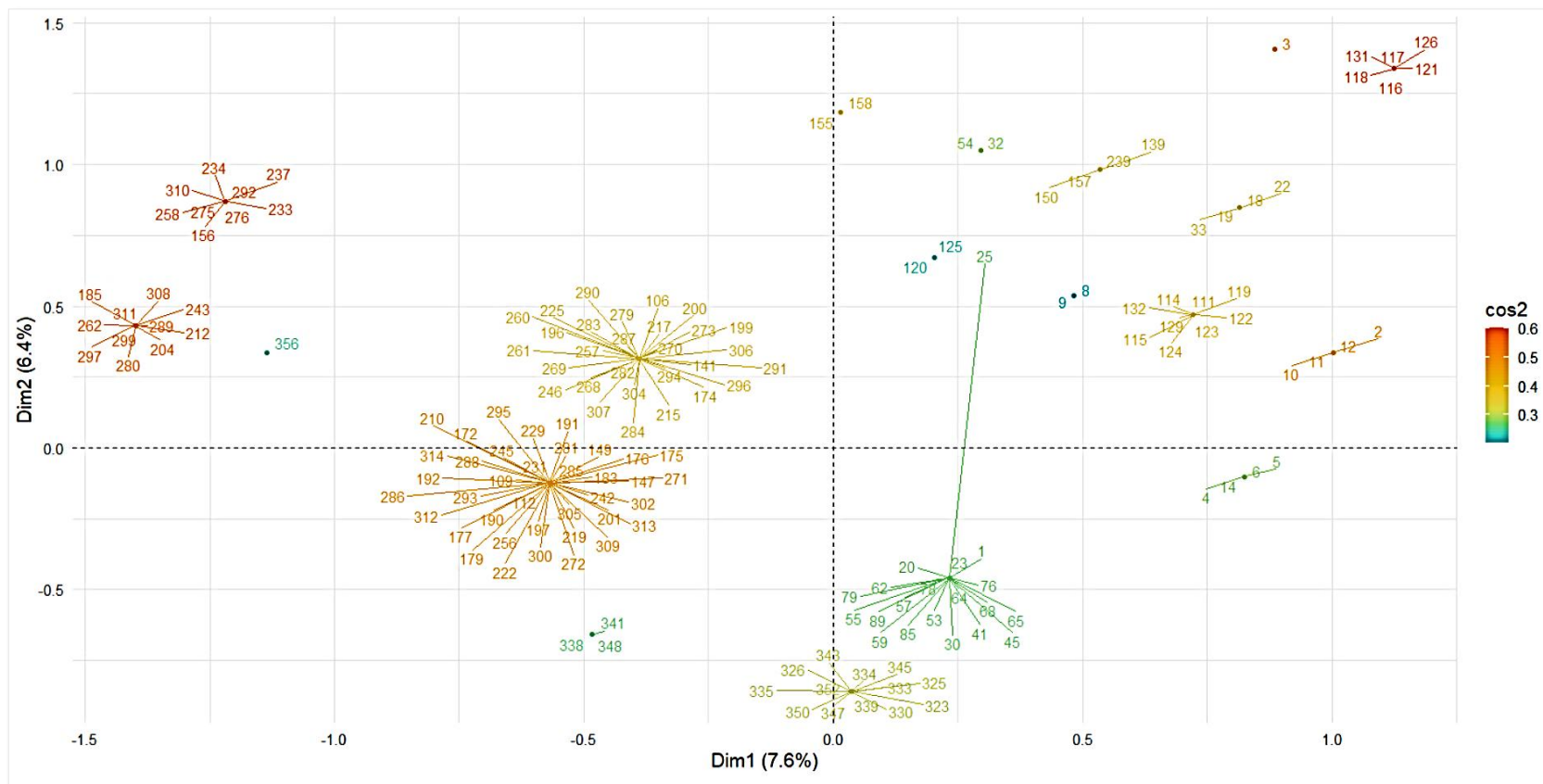
En síntesis, la Figura 10 revela dos patrones centrales:

- Un gradiente académico asociado al rendimiento y al tipo de institución (Dimensión 1).
- Un gradiente contextual vinculado a localización geográfica y modalidades educativas (Dimensión 2).

La comprensión de estos patrones se enriquece al examinar la distribución de los individuos en el plano factorial, lo que permite verificar si los perfiles analizados a nivel de categorías se reflejan también en la organización de los estudiantes. Para ello, se presenta la Figura 11, que muestra únicamente a los individuos con mayor calidad de representación ( $\cos^2$  elevados) y resalta los segmentos que explican su contribución relativa a cada dimensión.

**Figura 11**

*Gráfica de individuos mejor representados en el plano (ACM, Dim1–Dim2).*



**Nota.** Se presentan únicamente los individuos con mayor calidad de representación ( $\cos^2$ ) en el plano factorial Dim1–Dim2. La longitud y orientación de los segmentos reflejan la contribución relativa de cada individuo a las dimensiones mostradas, permitiendo identificar agrupamientos y diferenciaciones dentro de la cohorte analizada.

La Figura 11 presenta la distribución de los individuos con los mayores valores de  $\cos^2$ , es decir, aquellos que se encuentran mejor representados en el plano definido por las dos primeras dimensiones. Su visualización permite comprobar la coherencia entre la estructura de categorías observada en la Figura 10 y el comportamiento de los estudiantes individualmente considerados.

En el cuadrante superior izquierdo se concentra un grupo de individuos asociados a perfiles académicos favorables, típicos de estudiantes con promedios altos, procedentes de bachilleratos científicos y con trayectoria en instituciones particulares. Esta ubicación coincide con la posición de esas categorías en el mapa factorial anterior, lo que confirma la consistencia de la estructura identificada.

En contraste, en el cuadrante derecho se observa un grupo diferenciado de individuos vinculados a modalidades humanísticas, comerciales o letras, así como a colegios oficiales y categorías asociadas a menor rendimiento relativo. Este patrón reafirma el gradiente académico-contextual detectado en la Dimensión 1 y evidencia la existencia de subgrupos claramente definidos dentro de la cohorte.

Asimismo, se identifican segmentos más dispersos en la zona inferior del plano, relacionados con estudiantes provenientes de provincias con características socioeducativas distintas, lo que refuerza la interpretación de la Dimensión 2 como un eje de variación geográfica y contextual.

La lectura conjunta de las Figuras 10 y 11 confirma la presencia de patrones estructurales coherentes entre las categorías y los individuos del estudio. Los gradientes identificados permiten comprender la manera en que factores académicos y

sociodemográficos se articulan para generar perfiles diferenciados dentro del proceso de admisión. La combinación de los mapas factoriales de categorías e individuos proporciona una visión integral que fortalece la interpretación del ACM y sustenta la validez de los resultados obtenidos en este análisis.

#### **4.4 Síntesis de hallazgos**

La regresión logística binaria (RLB) evidencia un gradiente académico hacia la clasificación *Regular*, de manera que la probabilidad de aprobación aumenta conforme incrementan el Promedio, el PCA y el PCG. Paralelamente, el análisis de correspondencias múltiples (ACM) organiza combinaciones de categorías académicas y sociodemográficas asociadas a cada grupo de Clasificación\_IP, cumpliendo con los objetivos de descripción, explicación y perfilamiento planteados en el estudio.

Los resultados de ambos métodos convergen en un mismo patrón: la mayor fortaleza académica se asocia sistemáticamente con la clasificación *Regular*, mientras que perfiles académicos y contextuales menos favorables se relacionan con la condición Propedéutica. Esta convergencia respalda la utilidad de ambos enfoques para orientar estrategias de nivelación y apoyos tempranos.

El modelo RLB especificado con Promedio, PCA y PCG reproduce la regla institucional de clasificación y explica adecuadamente el patrón observado en la cohorte. El modelo presenta alto desempeño predictivo, tanto en discriminación como en calibración ( $AUC \approx 0.999$ ;  $Brier \approx 0.007$ ), lo que respalda su uso para fines operativos.

De manera complementaria, el ACM sintetiza la variación principal en el plano Dim1–Dim2, donde las categorías asociadas a mayores valores de Promedio, PCA y PCG

se agrupan con la clasificación Regular. Esta coherencia entre el modelo predictivo y el modelo exploratorio refuerza la validez interna del análisis: mientras la RLB cuantifica los efectos de los predictores, el ACM proporciona una visión estructural y contextual de los perfiles estudiantiles.

En conjunto, los hallazgos fortalecen la validez del modelo institucional y aportan evidencia operativa para focalizar tempranamente apoyos académicos, orientar acciones de acompañamiento y mejorar la toma de decisiones en el proceso de admisión.

#### **4.5 Indicador de Probabilidad de Ingreso Regular (IPIR)**

Al culminar este proyecto, se propone un indicador operativo derivado de la Regresión Logística Binaria (RLB), denominado Indicador de Probabilidad de Ingreso Regular (IPIR). Este indicador expresa, en una escala de 0 a 100, la probabilidad individual de clasificar como Regular y facilita la interpretación de resultados por estudiante, grupo o institución. Su uso permitiría focalizar apoyos académicos tempranos, orientar acciones de acompañamiento y mejorar la planificación de cupos.

##### **4.5.1 Modelo base del IPIR**

El modelo logístico para la cohorte 2023 predice la clasificación del Índice Predictivo a partir de sus tres componentes académicos: Promedio, PCG y PCA. Se expresa como:  $Clasificación_{IP} \sim PROMEDIO + PCG + PCA$

La probabilidad individual de ser clasificado como Regular para el estudiante  $i$  se calcula

como: 
$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 PROMEDIO_i + \beta_2 PCG_i + \beta_3 PCA_i)}}$$

donde  $\beta_0, \beta_1, \beta_2, \beta_3$  son los coeficientes estimados del modelo logístico.

#### 4.5.2 Cálculo del indicador

El IPIR se define como:  $IPIR_i = 100 \times p_i$

Por complemento, se puede calcular el Indicador de Riesgo Propedéutico (IRP) como:  $IRP_i = 100 - IPIR_i$

**Clasificación sugerida de bandas.** Con base en la distribución de probabilidades estimadas en la cohorte 2023, se sugieren tres bandas de referencia inicial:

**Tabla 21**

#### Clasificación de niveles de riesgo según el IPIR

Rango IPIR	Nivel de Riesgo	Interpretación	Acción Sugerida
< 50	Alto riesgo	Probabilidad muy baja de éxito académico.	Refuerzo inmediato y curso propedéutico.
51 – 70	Riesgo moderado	Aprobado mínimo con debilidades claras.	Seguimiento cercano y apoyo focalizado.
71 – 80	Riesgo medio	Desempeño aceptable pero con fragilidades.	Tutorías y actividades de nivelación.
≥ 81	Bajo riesgo	Desempeño alto y bases sólidas.	Alta probabilidad de éxito, se recomienda seguimiento estándar.

*Nota.* Elaboración propia. Estas bandas pueden ajustarse empíricamente en función de los percentiles observados en cada cohorte.

#### 4.4.3 Usos operativos del IPIR

**Alerta temprana.** Identificar estudiantes con IPIR bajo y priorizar acompañamiento académico.

**Planificación de cupos.** Estimar el número esperado de estudiantes Regulares en cada cohorte.

**Focalización de apoyo.** Asignar recursos académicos según el perfil de riesgo estimado.

**Monitoreo institucional.** Evaluar el impacto de programas propedéuticos mediante variaciones del IPIR entre cohortes.

El IPIR complementa los modelos estadísticos aplicados al ofrecer una medida operativa de interpretación directa. Su construcción basada en la RLB garantiza coherencia con los patrones observados en la cohorte 2023, permitiendo identificar, de forma individual, la probabilidad de éxito académico según el índice de admisión. Asimismo, proporciona evidencia práctica para la toma de decisiones institucionales y sugiere un mecanismo replicable para futuras cohortes con posibilidad de recalibración. El código R para la estimación se incluye en el Anexo B.

## CONCLUSIONES

El análisis multivariante de la cohorte 2023 permitió comprender de manera integral la configuración de la clasificación de ingreso en la Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá. Al integrar información académica y sociodemográfica en un mismo marco, la Regresión Logística Binaria (RLB) evidenció el efecto neto de los predictores académicos sobre la probabilidad de clasificación Regular, mientras que el Análisis de Correspondencias Múltiples (ACM) mostró gráficamente las asociaciones entre categorías y perfiles de aspirantes. La convergencia de ambas técnicas refuerza la validez interna del estudio y genera insumos prácticos para la gestión universitaria. De manera específica, los resultados indican que:

1. Los modelos parsimoniosos, sustentados en el núcleo académico (Promedio, PCA y PCG), logran un desempeño adecuado sin añadir complejidad innecesaria.
2. La RLB confirma que estos tres componentes resultaron estadísticamente significativos ( $p < 0.05$ ) en la estimación de la probabilidad de clasificación Regular, mientras que el ACM permitió visualizar perfiles coherentes con la misma lógica.
3. El análisis de tablas de contingencia y pruebas de Chi-cuadrado evidenció asociaciones significativas entre ciertas variables sociodemográficas, aunque con menor peso discriminatorio frente a los factores académicos.
4. El Índice Predictivo, como medida compuesta derivada de Promedio, PCA y PCG, no debe usarse como predictor adicional en modelos explicativos; en cambio, se

consolida como herramienta clasificatoria y operativa para anticipar cuántos aspirantes requerirán curso propedéutico.

5. Las variables sociodemográficas, entre ellas el tipo de colegio y la provincia, aportan contexto al perfil de los aspirantes, pero su influencia en la clasificación es marginal frente al rendimiento académico.
6. Los estudiantes de colegios particulares constituyen un grupo minoritario, presentan proporcionalmente más casos clasificados como regulares, mientras que los egresados de colegios oficiales concentran una mayor proporción de índice menor a uno. Este patrón evidencia diferencias en el rendimiento académico previo asociadas al tipo de formación secundaria.
7. Al trabajar con censo y no con muestra, los hallazgos describen y explican a la totalidad de la cohorte 2023 sin pretensión de inferencia poblacional, pero con alto valor diagnóstico e institucional.

En síntesis, los hallazgos confirman que el desempeño académico previo y los resultados en las pruebas institucionales constituyen los principales determinantes de la clasificación de ingreso. Los perfiles identificados permiten orientar con oportunidad los apoyos propedéuticos y las estrategias de nivelación temprana, fortaleciendo la planificación académica y la equidad en el acceso universitario.

## RECOMENDACIONES

1. **Institucionalizar el análisis multivariante en admisión.** Realizar anualmente estudios con RLB y ACM para anticipar el número de estudiantes propedéuticos y planificar cupos, recursos y estrategias de nivelación.
2. **Focalizar apoyos según perfiles de riesgo.** Usar las probabilidades estimadas por la RLB y los perfiles del ACM para diseñar programas de refuerzo en áreas críticas (Promedio, PCA o PCG).
3. **Evitar circularidad en los criterios.** Tomar decisiones directamente con los componentes del Índice Predictivo y no con el índice compuesto como predictor en modelos explicativos.
4. **Vincular resultados con la planificación académica.** Traducir los hallazgos en acciones curriculares y pedagógicas antes del inicio del semestre.
5. **Fortalecer el uso pedagógico en la Escuela de Estadística.** Incorporar el caso como práctica aplicada en cursos, fortaleciendo competencias en modelado y análisis con datos reales.
6. **Consolidar la práctica con R Markdown.** Mantener actualizado el flujo metodológico en un documento reproducible que sirva como estándar institucional.
7. **Crear un servicio de apoyo estadístico interno.** Entre la Escuela de Estadística y la Dirección de Admisión, establecer una unidad de asesoría que brinde evidencia objetiva y oportuna para decisiones estratégicas.

8. **Línea de desarrollo futuro.** Como proyección institucional, podría considerarse la exploración del Índice Predictivo de Ingreso Regular (IPIR) como herramienta diagnóstica en plataformas digitales internas. Esta iniciativa no forma parte de las recomendaciones del presente estudio ni responde a los objetivos planteados, pero podría, en una fase posterior, aportar información complementaria para estrategias de orientación académica y nivelación temprana.

## REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A. (2018). *Statistical methods for the social sciences* (5th ed.). Pearson.
- Arbuckle, J. (2005). *AMOS 7.0 User's Guide*. SPSS Inc.
- Babbie, E. (2014). *The practice of social research* (14th ed.). Cengage Learning.
- Cabrera, A. F., Pérez, P., & López, L. (2006). *Pathways to student success in higher education*. Routledge.
- Camara, W. J., & Echternacht, G. (2000). *The SAT I and high school grades: Utility in predicting success in college*. College Board Research Report.
- Cunningham, C., Rodríguez, P., & Martínez, L. (2007). Factores asociados al rendimiento académico en la educación superior: Un análisis multivariado. *Revista Iberoamericana de Educación Superior*, 3(1), 55–72.
- Delgado, A., & Méndez, L. (2018). Análisis de los factores asociados al rendimiento académico en pruebas de admisión universitaria. *Revista Mexicana de Educación Superior*, 8(2), 45–62.
- Díaz, V. (2002). *Introducción al análisis multivariado: Métodos y aplicaciones*. Editorial Universitaria.
- FCNET. (2022). *Informe de investigación y producción académica 2022*. Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá.
- Fiegehen, L., & Díaz, C. (2008). *Desafíos de la educación superior en América Latina: Rendimiento, equidad y acceso*. Fondo de Cultura Económica.
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.
- Garbanzo, G. (2013). Factores asociados al rendimiento académico en la educación superior. *Revista Educación*, 37(1), 1–27.
- García, F. (2020). *Factores asociados al rendimiento académico en la educación media*. Editorial Académica Universitaria.
- García, M. (2020). Tendencias en la evaluación del ingreso a la educación superior en Centroamérica. *Cuadernos de Investigación Educativa*, 35(3), 76–91.
- González, J., & Rojas, P. (2019). Aplicación de pruebas estandarizadas en procesos de admisión universitaria. *Revista de Ciencias de la Educación*, 27(2), 113–130.

- González, M., & Rojas, A. (2019). Evaluación y equidad en los procesos de admisión universitaria en América Latina. *Revista Latinoamericana de Estudios Educativos*, 49(2), 87–112.
- Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). Chapman & Hall/CRC.
- Greenacre, M. (2017). *Correspondence analysis in practice* (3rd ed.). Chapman & Hall/CRC.
- Greenacre, M., & Blasius, J. (Eds.). (2006). *Multiple correspondence analysis and related methods*. Chapman & Hall/CRC.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning* (2nd ed.). Springer.
- Horn, L., & Kojaku, L. (2001). *High school academic curriculum and the persistence path through college*. U.S. Department of Education.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.
- Jiménez, R. (2017). Evaluación de predictores del desempeño académico en estudiantes de primer ingreso. *Revista Panameña de Investigación Educativa*, 5(1), 24–39.
- Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. *Applied Statistics*, 21(2), 160–173.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065), 20150202.
- Kerlinger, F. N., & Lee, H. B. (2002). *Investigación del comportamiento: Métodos de investigación en ciencias sociales* (4.<sup>a</sup> ed.). McGraw-Hill.
- Kline, R. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT for predicting first-year college grade point average. College Board Research Report.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill Irwin.
- Lacarrière, M. (2008). Formación docente y calidad educativa en América Latina: Retos y perspectivas. *Revista Latinoamericana de Educación*, 8(2), 33–48.
- Lebart, L., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Dunod.
- Lebart, L., Morineau, A., & Piron, M. (2000). *Análisis exploratorio de datos multidimensionales*. Dunod.
- Martínez, R. (2016). Evaluación educativa en América Latina: Avances y desafíos. *Revista Latinoamericana de Educación*, 12(1), 45–59.
- Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). SAGE Publications.
- Minitab. (1995). *Minitab reference manual*.
- Montgomery, D. C. (2017). *Design and analysis of experiments* (9th ed.). Wiley.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.
- Norusis, M. J. (1983). *SPSS Statistical Procedures*. McGraw-Hill.
- Paredes, A., & Soto, V. (2020). Factores académicos asociados al rendimiento en pruebas de selección universitaria en Chile. *Educación y Sociedad*, 39(2), 89–102.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (Vol. 2). Jossey-Bass.
- Pascual, J. (2011). Factores que influyen en el rendimiento académico universitario: Un estudio de caso. *Revista Iberoamericana de Educación*, 55(1), 1–12.
- Pascual, M. (2011). Influencia del entorno educativo en el desempeño universitario. *Revista de Investigación Educativa*, 29(2), 233–249.
- Pérez-Luño, A., & Martínez-Gómez, M. (2020). Uso de software estadístico libre en la educación superior. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 28, 45–57.
- R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation.

- R Core Team. (2023). R: A language and environment for statistical computing (Version 4.3) [Computer software]. R Foundation.
- R Core Team. (2025). R: A language and environment for statistical computing (Version 4.4.1) [Computer software]. R Foundation.
- Ramírez, L. (2021). Variables predictoras del desempeño en pruebas Saber 11° en Colombia. *Revista Colombiana de Educación*, 81(1), 105–123.
- Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS*. Hamburg University.
- Ríos, J. (2012). Clima organizacional educativo: Una propuesta teórico-metodológica para su análisis. Plaza y Valdés.
- Ríos, L. (2012). Desigualdades educativas y género en la educación superior latinoamericana. *Revista Iberoamericana de Educación Superior*, 3(8), 45–63.
- Salagre, J., & Serrano, R. (2003). Influencia del tipo de bachillerato en el rendimiento académico universitario. *Revista de Educación*, 330, 305–328.
- Santos Álvarez, J. M. (2013). La formación del pensamiento práctico: Una tarea esencial de la educación superior. Universidad de Salamanca.
- StataCorp. (1992). *Stata statistical software: Release 3*.
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Lawrence Erlbaum Associates.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Tatsuoka, M. M. (1988). *Multivariate analysis*. Macmillan.
- Tinto, V. (2012). *Completing college: Rethinking institutional action*. University of Chicago Press.
- UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education*. United Nations Educational, Scientific and Cultural Organization.
- Universidad de Panamá. (2013). *Universidad de Panamá: 1935–2013. Historia, misión y compromiso con el desarrollo nacional*. Editorial Universitaria.
- Universidad de Panamá. (2015). *Manual de procedimientos de admisión 2015–2016*. Dirección General de Admisión.
- Universidad de Panamá. (2021). Innovación estadística en la investigación universitaria: El uso de biplots como herramienta multivariante. *Revista UP Investiga*, 4(2), 55–70.

- Universidad de Panamá. (2021). Plan estratégico de desarrollo institucional 2021–2026. Vicerrectoría de Planificación.
- Universidad de Panamá. (2023). Guía del proceso de admisión 2023. Sistema de Información Universitaria (SIU).
- Vicerrectoría Académica. (2021). Informe de continuidad académica en modalidad virtual durante la pandemia por COVID-19. Universidad de Panamá.
- Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer.

## ANEXOS

### **Anexo A. Clasificación institucional y perfil de variables**

**A.1** Clasificación institucional de los aspirantes según el Índice Predictivo

**A.2** Perfil rápido de variables

### **Anexo B. Scripts en RMarkdown**

**B.1** Implementación de los modelos estadísticos en R Markdown

**B.3** Script de asociación categórica

**B.2** Código R – Indicador (IPIR)

### **Anexo C. Significado de parámetros y estimadores**

**C.1** Regresión Logística Binaria

**C.2** Análisis de Correspondencias Múltiples

**C.3** Glosario

### **Anexo D. Capturas del Proceso de Admisión**

### **Anexo E. Síntesis metodológica y aplicaciones de técnicas multivariantes**

**Tabla E1** *Clasificación general de las técnicas multivariantes según tipo de enfoque y propósito del análisis*

**Tabla E2** *Requisitos estadísticos y condiciones de aplicación de las técnicas multivariantes*

**Tabla E3** *Aplicabilidad de las técnicas según tipo de población y naturaleza de los datos*

**Tabla E4** *Clasificación de las técnicas multivariantes según tipo y naturaleza del diseño de investigación*

**Tabla E5** *Tipos de estudios más utilizados en el ámbito educativo con poblaciones completas y sus objetivos principales*

## ANEXO A

### CLASIFICACIÓN INSTITUCIONAL Y PERFIL DE VARIABLES

#### A.1 Clasificación institucional de los aspirantes según el Índice Predictivo<sup>1</sup>

En el proceso de admisión de la Universidad de Panamá, la clasificación de los aspirantes se establece en función de su desempeño en las pruebas institucionales y del valor obtenido en el Índice Predictivo (IP), el cual integra tres componentes fundamentales: el Promedio de Bachillerato, el Puntaje de Conocimientos Académicos (PCA) y el Puntaje de Conocimientos Generales (PCG). Este indicador permite estimar la probabilidad de éxito académico del estudiante al ingresar a la educación superior.

Con base en los criterios institucionales de la Facultad de Ciencias Naturales, Exactas y Tecnología, los aspirantes, luego de la pandemia Covid19 hasta el 2023, se clasificaban en tres categorías: Regular, Propedéutico y No admitido. Esta categorización refleja el nivel de preparación del estudiante y determina su condición académica inicial dentro del sistema universitario. A continuación, se describen las categorías establecidas:

**Estudiante Regular (Ingreso Directo):** corresponde a los aspirantes que alcanzan un Índice Predictivo igual o superior a 1.00, lo que indica que poseen las competencias académicas necesarias para iniciar de forma inmediata los estudios de licenciatura. Estos estudiantes se matriculan directamente en el primer semestre de su carrera sin necesidad de nivelación previa.

---

<sup>1</sup> Fuente: Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá (2023).

**Estudiante Propedéutico (Ingreso Condicional):** incluye a quienes presentan un Índice Predictivo inferior a 1.00 pero igual o superior a 0.70. Este grupo requiere cursar un programa propedéutico o nivelatorio con el propósito de fortalecer conocimientos básicos antes de incorporarse plenamente al plan regular de estudios. Su ingreso es condicional hasta culminar satisfactoriamente el curso de nivelación.

- **No Admitido:** se refiere a los aspirantes cuyo Índice Predictivo es menor de 0.70, por lo que no cumplen con los requisitos mínimos de ingreso establecidos por la Facultad. En este caso, el estudiante no es admitido en el periodo correspondiente, aunque puede optar por volver a aplicar en futuras convocatorias.

Esta clasificación fue utilizada como referencia metodológica en la definición de la variable dependiente del modelo de Regresión Logística Binaria desarrollado en el presente estudio. Es relevante mencionar que actualmente la clasificación no es utilizada

## A.2 Perfil rápido de variables

SEDE: categorical | n=364, missing=0, unique=1

FACULTAD: categorical | n=364, missing=0, unique=1

COD\_ESCUELA: categorical | n=364, missing=0, unique=5

COD\_CARRERA: numeric | n=364, missing=0, unique=8

ANO\_LECTIVO: categorical | n=364, missing=0, unique=1

SEXO: categorical | n=364, missing=0, unique=2

EDAD: numeric | n=364, missing=0, unique=26

NACIONALIDAD: categorical | n=364, missing=0, unique=2

PROMEDIO\_DE\_SECUNDARIA: numeric | n=364, missing=0, unique=126

PCA: numeric | n=364, missing=0, unique=68

PCG: numeric | n=364, missing=0, unique=59

INDICE\_PREDICTIVO: numeric | n=364, missing=0, unique=362

PROVINCIA\_RESIDENCIA: categorical | n=364, missing=0, unique=12

DIRECCION\_RESIDENCIA: categorical | n=364, missing=0, unique=335

BACHILLERATO: categorical | n=364, missing=0, unique=8

TIPO\_COLEGIO: categorical | n=364, missing=0, unique=2

## ANEXO B

### SCRIPTS EN RMARKDOWN

#### B.1 Implementación de los modelos estadísticos en R Markdown

##### *# Reproducibilidad*

```
set.seed(1234)
options(stringsAsFactors = FALSE)
```

##### *# Paquetes*

```
library(readxl); library(dplyr); library(tidyr)
library(janitor); library(stringr)
library(FactoMineR); library(factoextra)
library(broom); library(pROC); library(logistf)
library(ggplot2)
```

##### *# Rutas*

```
ruta_excel <- "datos_admision.xlsx"
hoja <- "BASE DE DATOS"
dir.create("salidas_cap4", showWarnings = FALSE)
```

```
# -----
```

##### *# 1) Carga y preparación*

```
# -----
```

```
raw <- read_excel(ruta_excel, sheet = hoja) |> clean_names()
datos <- raw |>
  transmute(
    sexo = str_to_title(sexo),
    edad = suppressWarnings(as.numeric(edad)),
    promedio = suppressWarnings(as.numeric(promedio_de_secundaria)),
    pca = suppressWarnings(as.numeric(pca)),
    pcg = suppressWarnings(as.numeric(pcg)),
    ip = suppressWarnings(as.numeric(indice_predictivo)),
    provincia = str_to_title(provincia_residencia),
    bachiller = str_to_title(bachillerato),
    colegio = case_when(
      str_detect(str_to_lower(tipo_colegio), "oficial|público") ~ "Público",
      TRUE ~ "Privado"
    )
  ) |>
  mutate(
    clasificacion_ip = if_else(ip >= 1, "Regular", "Propedéutico"),
    ip_band = cut(ip, breaks = c(-Inf, .70, .99, 1.99, Inf),
```

```

        labels = c("Muy bajo (<0.70)", "Bajo (0.70–0.99)",
                  "Medio (1.00–1.99)", "Alto (≥2.00)"),
edad_band = cut(edad, breaks = c(-Inf, 17, 18, 19, Inf),
                labels = c("≤17", "18", "19", "≥20"))
)

# Provincias raras → 'Otras' (umbral 1%)
umbral <- max(1, ceiling(0.01 * nrow(datos)))
prov_freq <- table(datos$provincia)
raras <- names(prov_freq[prov_freq < umbral])
datos <- datos |> mutate(provincia_grp = if_else(provincia %in% raras, "Otras",
provincia))

# -----
# 2) Descriptivos y tablas
# -----
# Distribución Regular vs Propedéutico
dist_ip <- datos |>
  count(clasificacion_ip, name = "Frecuencia") |>
  mutate(Porcentaje = round(100 * Frecuencia / sum(Frecuencia), 1))
write.csv(dist_ip, "salidas_cap4/tabla_dist_ip.csv", row.names = FALSE)

# Helper: tabla de bandas por subgrupo
tab_bandas <- function(df, grupo) {
  lvls <- c(sort(unique(df[[grupo]])))
  bandas <- levels(df$ip_band)
  out <- lapply(lvls, function(v) {
    sub <- df[df[[grupo]] == v, ]
    tot <- nrow(sub)
    ctas <- table(sub$ip_band)[bandas]; ctas[is.na(ctas)] <- 0
    tibble(Grupo = v,
           Banda = bandas,
           N = as.integer(ctas),
           Porc = round(100 * as.numeric(ctas) / tot, 1))
  }) |> bind_rows()
  out
}
write.csv(tab_bandas(datos, "sexo"), "salidas_cap4/ip_por_sexo.csv",
row.names = FALSE)
write.csv(tab_bandas(datos, "colegio"), "salidas_cap4/ip_por_colegio.csv",
row.names = FALSE)
write.csv(tab_bandas(datos, "bachiller"), "salidas_cap4/ip_por_bachiller.csv",
row.names = FALSE)
write.csv(tab_bandas(datos, "edad_band"), "salidas_cap4/ip_por_edad.csv", row.names
= FALSE)

```

```
write.csv(tab_bandas(datos, "provincia_grp"), "salidas_cap4/ip_por_provincia.csv",
row.names = FALSE)
```

### **# Descriptivos continuas**

```
desc <- datos |> summarise(
  n = n(),
  media_prom = mean(promedio, na.rm = TRUE), sd_prom = sd(promedio, na.rm =
TRUE),
  media_pca = mean(pca, na.rm = TRUE), sd_pca = sd(pca, na.rm = TRUE),
  media_pcg = mean(pcg, na.rm = TRUE), sd_pcg = sd(pcg, na.rm = TRUE),
  media_ip = mean(ip, na.rm = TRUE), sd_ip = sd(ip, na.rm = TRUE),
  media_edad = mean(edad, na.rm = TRUE), sd_edad = sd(edad, na.rm = TRUE)
)
write.csv(desc, "salidas_cap4/descriptivos_continuas.csv", row.names = FALSE)
```

```
# -----
```

### **# 3) ACM (FactoMineR)**

```
# -----
```

```
datos_acm <- datos |>
  transmute(
    sexo, colegio, bachiller, provincia_grp, edad_band,
    prom_3 = cut(promedio, breaks = quantile(promedio, probs = c(0,1/3,2/3,1), na.rm =
TRUE),
    include.lowest = TRUE, labels = c("Bajo", "Medio", "Alto")),
    pca_3 = cut(pca, breaks = quantile(pca, probs = c(0,1/3,2/3,1), na.rm = TRUE),
    include.lowest = TRUE, labels = c("Bajo", "Medio", "Alto")),
    pcg_3 = cut(pcg, breaks = quantile(pcg, probs = c(0,1/3,2/3,1), na.rm = TRUE),
    include.lowest = TRUE, labels = c("Bajo", "Medio", "Alto"))
  )
res_mca <- MCA(datos_acm, graph = FALSE)
```

### **# Inercia**

```
write.csv(res_mca$eig, "salidas_cap4/acm_inercia.csv")
```

### **# Coordinadas/Contribuciones/Cos2**

```
write.csv(res_mca$var$coord, "salidas_cap4/acm_coord.csv")
write.csv(res_mca$var$contrib, "salidas_cap4/acm_contrib.csv")
write.csv(res_mca$var$cos2, "salidas_cap4/acm_cos2.csv")
```

### **# Figuras ACM**

```
p1 <- fviz_screplot(res_mca) + theme_minimal()
ggsave("salidas_cap4/acm_scee.png", p1, width = 7, height = 4, dpi = 300)
p2 <- fviz_mca_biplot(res_mca, repel = TRUE, ggtheme = theme_minimal())
ggsave("salidas_cap4/acm_biplot_cat.png", p2, width = 7, height = 4.5, dpi = 300)
```

```

# -----
# 4) RLB (parsimoniosa)
# -----
dat_rlb <- datos |>
  mutate(
    zEdad = scale(edad)[,1], zPROM = scale(promedio)[,1],
    zPCA = scale(pca)[,1], zPCG = scale(pcg)[,1],
    y = if_else(clasificacion_ip == "Regular", 1L, 0L)
  ) |>
  drop_na(zEdad, zPROM, zPCA, zPCG, sexo, colegio, bachiller, y)
m <- glm(y ~ zEdad + zPROM + zPCA + zPCG + sexo + colegio + bachiller,
  family = binomial(link = "logit"), data = dat_rlb)
or_tab <- tidy(m, conf.int = TRUE, exponentiate = TRUE) |>
  mutate(term = recode(term,
    "(Intercept)"="Intercepto", "zEdad"="Edad (+1 DE)", "zPROM"="Promedio (+1 DE)",
    "zPCA"="PCA (+1 DE)", "zPCG"="PCG (+1 DE)")) |>
  select(Término = term, OR = estimate, `CI 95% (inf)` = conf.low, `CI 95% (sup)` =
  conf.high)
write.csv(or_tab, "salidas_cap4/rlb_or.csv", row.names = FALSE)
# Métricas
prob <- predict(m, type = "response")
obs <- dat_rlb$y
roc <- pROC::roc(obs, prob)
cm <- table(Obs = obs, Pred = as.integer(prob >= 0.5))
acc <- sum(diag(cm))/sum(cm)
sens <- cm["1", "1"]/sum(cm["1",])
esp <- cm["0", "0"]/sum(cm["0",])
met <- data.frame(
  AIC = AIC(m),
  McFadden = 1 - logLik(m)/logLik(update(m, . ~ 1)),
  Exactitud = acc, Sensibilidad = sens, Especificidad = esp,
  AUC = as.numeric(roc$auc)
)
write.csv(met, "salidas_cap4/rlb_metricas.csv", row.names = FALSE)
# (Opcional) Firth si incluye territorio: logistf(...)
# -----
# 5) Reproducibilidad
# -----
sink("salidas_cap4/sessionInfo.txt"); print(sessionInfo()); sink()}

# --- Prueba Chi-cuadrado de independencia ---
# Cargar librerías
library(dplyr)

```

```

# Tabla de contingencia entre variables categóricas
tabla_chi <- table(datos_admision$Clasificacion_IP, datos_admision$Provincia)

# Visualización de la tabla
tabla_chi

# Aplicación de la prueba Chi-cuadrado
chi_result <- chisq.test(tabla_chi)

# Resultados principales
chi_result

# Mostrar residuos estandarizados (útiles para interpretación)
chi_result$stdres
# Interpretación:
# H0: No existe asociación entre las variables (independencia)
# H1: Existe asociación entre las variables
# Si p-valor < 0.05 → se rechaza H0 → las variables están asociadas.

# --- Construcción de la Matriz de Burt ---

# Cargar librería necesaria
library(FactoMineR)

# Seleccionar solo las variables categóricas para el ACM
vars_categ <- datos_admision %>%
  select(Sexo, Provincia, Colegio, Clasificacion_IP)

# Generar la Matriz de Burt
burt_matrix <- Burt(vars_categ)

# Visualizar las primeras filas de la matriz
head(burt_matrix)

# Dimensiones
dim(burt_matrix)

# Guardar la matriz como archivo para incluir en los anexos
write.csv(burt_matrix, "Matriz_Burt.csv", row.names = FALSE)

# Nota:
# La Matriz de Burt contiene en su diagonal las tablas de frecuencias de cada variable
# y fuera de la diagonal las tablas de contingencia entre pares de variables.

```

```

# --- Visualización de la Matriz de Burt ---
library(ggplot2)

# Convertir en formato largo para graficar
burt_long <- as.data.frame(as.table(as.matrix(burt_matrix)))
ggplot(burt_long, aes(Var1, Var2, fill = Freq)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "steelblue") +
  theme_minimal() +
  labs(title = "Matriz de Burt - Visualización de asociaciones",
       x = "", y = "")

```

## B.2 Código R – Indicador (IPIR)

### *#Modelo y probabilidad*

```

# Cargar datos
# datos_admision <- readxl::read_xlsx("C:/ruta/datos_admision.xlsx")

# Modelo parsimonioso
m <- glm(Clasificacion_IP ~ PROMEDIO + PCG + PCA,
        data = datos_admision, family = binomial())

```

### *# Probabilidad e indicador*

```

p <- predict(m, type = "response")
IPIR <- round(100 * p) # 0–100
IRP <- 100 - IPIR # Riesgo de propedéutico
Bandas de acción
corte_alto <- 40
corte_medio <- 60
categoria <- cut(IPIR,
                breaks = c(-1, corte_alto, corte_medio, 100),
                labels = c("Alto apoyo", "Apoyo medio", "Apoyo bajo"))
table(categoria)

```

### *#Planificación y monitore-Total esperado de Regulares (promedio de p)*

```

total_esperado <- sum(p)
promedio_IPIR <- mean(IPIR)

```

### *# Resumen por grupo (ejemplo: colegio)*

```

library(dplyr)
resumen_colegio <- datos_admision %>%
  mutate(p = p, IPIR = IPIR) %>%
  group_by(COLEF = COLEGIO) %>%
  summarise(
    n = n(),

```

```
Regular_observado = mean(Clasificacion_IP),  
Regular_esperado = mean(p),  
Prom_IPIR = mean(IPIR)  
)  
resumen_colegio
```

**ANEXO C**  
**SIGNIFICADO DE PARÁMETROS Y ESTIMADORES**

**C.1 Regresión Logística Binaria**

<b>Símbolo/abreviatura</b>	<b>Significado</b>
$\beta$	Coefficiente del modelo (escala logit)
SE	Error estándar del coeficiente
$z$	Estadístico $z = \frac{\beta}{SE}$
OR	Razón de momios = $\exp(\beta)$
IC95%	Intervalo de confianza del 95 %
$p$	$p - valor$
$\pi(x)$	Probabilidad estimada por el modelo
logit( $p$ )	$\ln\left(\frac{p}{1-p}\right)$

**C.2 Análisis de Correspondencias Múltiples**

<b>Símbolo/abreviatura</b>	<b>Significado</b>
eigen / valor propio	Autovalor asociado a la dimensión
Contrib. (%)	Aporte de una categoría a la dimensión
$\cos^2$	Calidad de representación de un punto

### C.3 Glosario

1. **AIC (Akaike Information Criterion):** Criterio de información de Akaike; compara modelos penalizando la complejidad. Valores más bajos indican mejor equilibrio ajuste–parsimonia.
2. **AUC (Área bajo la curva ROC):** Medida de capacidad discriminativa de un clasificador.  $AUC = 0.5$  (sin discriminación) y  $AUC = 1$  (perfecta).
3. **Brier score (Puntaje de Brier):** Error cuadrático medio de probabilidades predichas para un resultado binario; evalúa calibración (0 = perfecto).
4. **Calibración (diagrama/curva de calibración):** Concordancia entre probabilidades predichas y frecuencias observadas del evento.
5. **Coseno al cuadrado ( $\cos^2$ ):** Calidad de representación de una categoría/variable en el plano factorial  $\left(\frac{ACP}{ACM}\right)$ ; valores cercanos a 1 indican buena representación.
6. **CTR (Contribución):** Aporte de una categoría/variable a la inercia de una dimensión en ACM; por dimensión suma 100 %.
7. **Deviance ( $-2 \text{ Log-Likelihood}$ ):** Medida de ajuste en modelos  $\frac{GLM}{RLB}$ ; valores menores implican mejor ajuste (comparables entre modelos anidados).
8. **Dimensión (en ACM):** Eje latente que resume asociación entre categorías; cada una explica un porcentaje de inercia.
9. **Especificidad (TNR):** Proporción de verdaderos negativos:  $\frac{VN}{VN + FP}$ .
10. **IC 95 % (Intervalo de confianza del 95 %):** Rango de incertidumbre reportado para un parámetro o efecto.

11. **Inercia (en ACM):** Varianza/chi-cuadrado normalizada explicada por las dimensiones del mapa factorial.
12. **Logit:** Transformación logarítmica  $\log\left(\frac{p}{1-p}\right)$  usada como función de enlace en RLB.
13. **Log-verosimilitud (Log-Likelihood, LL):** Suma de probabilidades logarítmicas del modelo; mayor (menos negativa) indica mejor ajuste.
14. **Matriz de confusión:** Tabla  $2 \times 2$  con VP, FP, VN y FN para evaluar clasificación.
15. **Multicolinealidad:** Correlación alta entre predictores; infla varianzas de las estimaciones.
16. **Odds (razón de probabilidades):** Definida como  $odds = \frac{p}{1-p}$ .
17. **Odds ratio (OR):** Razón de odds entre dos niveles; en RLB es  $\exp(\beta)$ .  $OR > 1$  aumenta las odds del evento;  $OR < 1$  las reduce.
18. **Parsimonia (criterio de):** Preferencia por el modelo más simple con desempeño comparable (p. ej., por  $\frac{AIC}{AUC}$ ).
19. **Prevalencia:** Proporción de casos positivos en la muestra o población analizada.
20. **Pseudo- $R^2$  (McFadden, Nagelkerke):** Análogos de  $R^2$  para RLB; valores mayores indican mejor ajuste relativo.
21. **ROC (Curva Característica Operativa del Receptor):** Curva de sensibilidad frente a 1 menos especificidad a lo largo de umbrales; su área es el AUC.
22. **Sensibilidad (TPR/Recall):** Proporción de verdaderos positivos:  $\frac{VP}{VP + FN}$ .
23. **Umbral de clasificación (cut-off):** Valor de probabilidad a partir del cual se asigna la clase positiva; afecta sensibilidad y especificidad.

24. **Valor predictivo negativo (VPN):**  $\frac{VN}{VN + FN}$ ; probabilidad de ser negativo dado un test negativo.
25. **Valor predictivo positivo ( $\frac{VPP}{PPV}$ ):**  $VP/(VP + FP)$ ; probabilidad de ser positivo dado un test positivo.
26. **VIF (Variance Inflation Factor):** Factor de inflación de la varianza para diagnosticar multicolinealidad; valores mayores de 5 a 10 sugieren problema.

## ANEXO D CAPTURAS DEL PROCESO DE ADMISIÓN

### Figura D1

*Formulario de registro en línea para aspirantes de primer ingreso (2023)*

The screenshot shows a web form titled "Formulario de Registro para Aspirantes". The form is divided into several sections:

- Datos Generales:** Includes fields for "Tipo de Identificación", "Categoría de Aspirante", "Nacionalidad", "Primer y Segundo Nombre", "Apellido", "Fecha de Nacimiento", "Tipo de Documento", and "Número de Documento".
- Números Telefónicos:** Fields for "Tipo de contacto", "Número de Teléfono", "Tipo de contacto", and "Número de contacto".
- Correo Electrónico:** Fields for "Tipo de correo", "Correo", "Tipo de correo", and "Correo".
- Dirección:** Fields for "Tipo de Dirección", "Provincia", "Distrito", "Calle/Callejón", and "Barrio".
- Carrera a elegir como primera opción:** Fields for "Tipo de Carrera", "Carrera", "Facultad", "Escuela", and "Carrera".

At the bottom, there is a "Guardar" button and a "Documentación Requerida (Ver Requisitos)" section with a "Subir archivo" button.

### Figura D2

*Portal de inicio de sesión del sistema de admisión de la Universidad de Panamá.*

Sistema de Primer ingreso a de la Universidad de Panamá

## Inicio

The login portal consists of three input fields and a button:

- Cédula:** Input field with a person icon.
- Correo Electrónico:** Input field with an envelope icon.
- # Número de Inscripción:** Input field with a lock icon.
- Ingresar:** A blue button to submit the login information.

## Figura D3

*Página principal de la Dirección General de Admisión de la Universidad de Panamá.*

**INFORMACIÓN DE INTERÉS – PROCESO DE ADMISIÓN**  
**¡BIENVENIDO A LA MEJOR UNIVERSIDAD DEL PAÍS Y CON MÁS TRAYECTORIA!!**

**¿En qué consiste nuestro proceso de admisión?**

La Dirección General de Admisión tiene la misión de ser la puerta de entrada a los estudios universitarios y así lograr la calidad del recurso principal de la Universidad de Panamá, sus estudiantes.

En los tiempos actuales donde los grandes medios de la información y comunicación son una realidad, consideramos propicio incorporarnos en esta vía para brindarles la información necesaria para realizar el proceso de admisión a la Universidad de Panamá, formarse como profesionales que ayuden al enriquecimiento de nuestra nación.

<b>Carreras según bachiller</b> 	<b>Calendario del Proceso de Admisión</b> 	<b>Requisitos para el Proceso de Admisión</b> 
<b>Pruebas de Admisión</b> 	<b>Pasos para la Inscripción Vía Web</b> 1 2 3 4	<b>Requisitos para Extranjeros</b> 
<b>Forma de Pago</b> 	<b>Temarios de Exámenes</b> 	
<b>Oferta Académica</b> 		

## Figura D4

*Documento oficial sobre la implementación de pruebas de admisión virtuales.*

### UNIVERSIDAD DE PANAMÁ DIRECCIÓN DE ADMISIÓN

#### PRUEBAS DE ADMISIÓN VIRTUALES

La Universidad de Panamá comenzó a realizar pruebas virtuales a partir del Proceso de Admisión de 2021 debido al impacto de la pandemia COVID-19. Esta medida fue necesaria para adaptarse rápidamente y garantizar la continuidad de los procesos académicos y administrativos.

Por otro lado, debido a las restricciones de distanciamiento social y movilidad impuestas por la pandemia se hizo imperativo adoptar modalidades virtuales en el proceso enseñanza – aprendizaje, asegurando así un entorno seguro para todos los involucrados y minimizando el riesgo de propagación del virus. A partir de entonces la Dirección General de Admisión ha potenciado los esfuerzos para asegurar un servicio de calidad. De igual manera el número de aspirantes a primer ingreso ha aumentado significativamente con el paso de los años y el ahorro de papel se hace imperante.

#### **Prueba de Capacidades Académicas (P.C.A.):**

- Esta prueba se aplica a todos los estudiantes y como su nombre lo dice mide las capacidades, habilidades y destrezas necesarias para el aprendizaje efectivo en la Universidad.
- La prueba se divide en dos áreas: Verbal y Numérica. El área verbal se divide en tres categorías que son: Léxico (vocabulario), comprensión de lectura y redacción. El área numérica se divide en dos categorías que son: operatoria y razonamiento. La prueba tiene en total 100 preguntas. Del área verbal son 60 y del área numérica 40 preguntas. Cada una tiene un valor de un punto.

#### **Prueba de Conocimientos Generales (P.C.G.):**

- Esta prueba mide los conocimientos específicos que requiere todo estudiantes que desea ingresar a ciertas carreras.

Actualmente se aplican 2 pruebas de conocimientos generales, a saber:

**PCG Científica:** para los aspirantes a las facultades de: Ciencias Agropecuarias, Ciencias Naturales, Exactas y Tecnología, Enfermería, Facultad de Informática, Electrónica y Comunicación (Ing. Mecatrónica, Técnico en Seguridad Informática, Lic. en Ing. Informática y Lic. en Ing. Electrónica y Comunicación), Psicología, Farmacia, Ingeniería, Medicina, Medicina Veterinaria y Odontología (Doctor Cirujano Dental).

**PCG Humanística:** para los aspirantes a la Facultad de Derecho y Ciencias Políticas.

## Figura D5

### *Anuncio oficial sobre las fases de las pruebas de admisión*



UNIVERSIDAD DE PANAMÁ  
VICERRECTORÍA ACADÉMICA  
DIRECCIÓN GENERAL DE ADMISIÓN  
Y  
VICERRECTORÍA DE ASUNTOS ESTUDIANTILES  
DIRECCIÓN DE INVESTIGACIÓN Y ORIENTACIÓN PSICOLÓGICA

### Fases de Admisión 2023

- FASE 1** Inscripción al Proceso de Admisión  
Del viernes 15 de julio al viernes 14 de octubre de 2022
- FASE 2** Aplicación de pruebas
- Pruebas Psicológicas**  
Del lunes 25 de julio al viernes 25 de noviembre de 2022
- Prueba de Capacidades Académicas**  
Del lunes 24 al viernes 28 de octubre de 2022
- Prueba de Conocimientos Generales**  
(Científica, Derecho y Policía)  
Del lunes 14 al viernes 18 de noviembre de 2022
- FASE 3** Entrega de Resultados a los estudiantes  
Del martes 10 al viernes 13 de enero de 2023



#### Observaciones:

1. Panameños que estudian en cualquier colegio a nivel Nacional, ingresan por la plataforma **EXCLUSIVA PARA PANAMEÑOS** (<https://siu.up.ac.pa/>).
2. Los panameños graduados en el extranjero, los extranjeros que vienen de escuelas en el extranjero y los extranjeros que estudian en Panamá, deben **ingresar primero** a la plataforma SIREX (<https://sirex.up.ac.pa/>) para iniciar su trámite, **del 15 de julio al 30 de septiembre del 2022**. **Posteriormente** ingresan a la plataforma (<https://siu.up.ac.pa/>) para inscribirse al proceso de admisión de la Universidad de Panamá.

**COSTO B/.30.00**

Aprobado por el Consejo Académico en su Reunión N°4-22, celebrada el día 27 de abril de 2022

Año 2022: "45 años de los Tratados Torrijos Carter"  
Ciudad Universitaria Octavio Méndez Pereira  
Tél.: 523-5406/5412/5413/5414 admision.general@up.ac.pa

## ANEXO E

### SÍNTESIS METODOLÓGICA Y APLICACIONES DE TÉCNICAS MULTIVARIANTES.

El presente anexo presenta una síntesis metodológica de las técnicas multivariantes aplicadas en el estudio, organizadas en tablas que resumen los enfoques, supuestos, condiciones, tipos de población, diseños de investigación y aplicaciones educativas más frecuentes.

**Tabla E1**

*Clasificación general de las técnicas multivariantes según tipo de enfoque y propósito del análisis.*

<b>Enfoque</b>	<b>Técnica</b>	<b>Requiere variable respuesta</b>	<b>Propósito</b>	<b>Uso en este estudio</b>	<b>Potencial uso en estudios futuros</b>
<b>Dependencia</b>	Regresión logística binaria	Sí	Predicción y clasificación	✓	Aplicable a poblaciones o muestras
<b>Interdependencia</b>	Análisis de correspondencias múltiples (ACM)	No	Identificación de asociaciones entre categorías	✓	Recomendado para explorar estructuras categóricas
<b>Dependencia</b>	Modelos de ecuaciones estructurales (SEM)	Sí	Relaciones causales entre constructos	Teórico	Requiere muestra representativa y supuestos de normalidad
<b>Interdependencia</b>	Análisis de componentes principales (PCA)	No	Reducción dimensional	Teórico	Requiere homocedasticidad y normalidad multivariante
<b>Dependencia</b>	MANOVA	Sí	Comparación de medias multivariantes	Teórico	Requiere normalidad multivariante y matrices de covarianza homogéneas

**Tabla E2**

*Requisitos estadísticos y condiciones de aplicación de las técnicas multivariantes.*

<b>Técnica</b>	<b>Tipo de método</b>	<b>Supuestos principales</b>	<b>Tipo de variable</b>	<b>Requiere muestra o población</b>	<b>Comentarios técnicos</b>
<b>Regresión logística binaria</b>	Paramétrico	No requiere normalidad, pero sí independencia entre observaciones	Dependiente dicotómica, independientes numéricas o categóricas	Población o muestra	Ideal para modelar clasificaciones binarias
<b>ACM</b>	No paramétrico	No requiere homocedasticidad ni normalidad	Categóricas	Población o muestra	Basado en frecuencias; usa matriz de Burt
<b>SEM</b>	Paramétrico	Normalidad multivariante, ausencia de colinealidad, tamaño muestral amplio	Latentes y observadas	Muestra	Permite modelar relaciones complejas
<b>PCA</b>	Paramétrico	Linealidad, normalidad multivariante	Cuantitativas	Muestra	Basado en correlaciones o covarianzas
<b>MANOVA</b>	Paramétrico	Homocedasticidad, normalidad multivariante	Cuantitativas	Muestra	Compara grupos sobre varias dependientes

*Nota:* Presenta los supuestos principales de cada técnica, el tipo de variable, la naturaleza del método (paramétrico o no paramétrico) y si incorpora pruebas de hipótesis.

**Tabla E3***Aplicabilidad de las técnicas según tipo de población y naturaleza de los datos.*

<b>Tipo de población</b>	<b>Tipo de variables</b>	<b>Técnicas aplicables</b>	<b>Justificación</b>	<b>Observaciones</b>
<b>Población completa</b>	Catóricas	Regresión logística binaria, ACM	No se requiere inferencia, se analizan asociaciones y clasificaciones reales	Ideal para estudios censales (como el de 2023)
<b>Población completa</b>	Cuantitativas	PCA descriptivo	Permite síntesis y patrones, sin necesidad de pruebas de hipótesis	Útil para reducción de dimensiones
<b>Muestra</b>	Catóricas	Chi-cuadrado, ACM, SEM	Permite inferir relaciones entre variables	Requiere representatividad y tamaño adecuado
<b>Muestra</b>	Cuantitativas	MANOVA, PCA, Regresión múltiple	Permite contrastar hipótesis o predecir	Requiere normalidad y homocedasticidad
<b>Población o muestra mixta</b>	Mixtas (numéricas y catóricas)	SEM, modelos mixtos, árboles de decisión	Admite variables de distinta naturaleza	Puede combinar enfoques descriptivos y predictivos

*Nota:* Resume qué técnicas son adecuadas para poblaciones completas o muestras, según el tipo de variable y el objetivo del análisis

**Tabla E4***Clasificación de las técnicas multivariantes según tipo y naturaleza del diseño de investigación.*

<b>Técnica</b>	<b>Tipo de diseño más frecuente</b>	<b>Finalidad del análisis</b>	<b>Naturaleza del diseño</b>	<b>Nivel de inferencia</b>	<b>Comentarios interpretativos</b>
<b>Análisis de correspondencias múltiples (ACM)</b>	Exploratorio	Identificar asociaciones entre categorías	Transversal	Descriptivo	Permite descubrir patrones sin hipótesis previa; útil para censos o cortes únicos
<b>Regresión logística binaria</b>	Explicativo	Analizar influencia de variables sobre una clasificación	Transversal o longitudinal	Inferencial	Modela relaciones de dependencia; puede aplicarse a datos de corte o seguimiento
<b>Análisis de componentes principales (PCA)</b>	Exploratorio	Reducir dimensiones y sintetizar información	Transversal	Descriptivo	No requiere hipótesis; se usa para encontrar estructuras latentes en datos cuantitativos
<b>MANOVA</b>	Explicativo	Comparar medias multivariantes entre grupos	Transversal	Inferencial	Permite contrastar hipótesis sobre diferencias grupales en múltiples variables dependientes
<b>Modelos de ecuaciones estructurales (SEM)</b>	Explicativo	Validar relaciones causales entre variables latentes	Transversal o longitudinal	Inferencial	Admite evaluación de trayectorias causales; puede extenderse a modelos de panel

Continuación...

**Tabla E4**

*Clasificación de las técnicas multivariantes según tipo y naturaleza del diseño de investigación.*

<b>Técnica</b>	<b>Tipo de diseño más frecuente</b>	<b>Finalidad del análisis</b>	<b>Naturaleza del diseño</b>	<b>Nivel de inferencia</b>	<b>Comentarios interpretativos</b>
<b>Chi-cuadrado de independencia</b>	Descriptivo / explicativo	Examinar asociación entre variables categóricas	Transversal	Inferencial	Útil para analizar relaciones en tablas de contingencia sin supuestos de distribución
<b>Árboles de decisión (CART, CHAID)</b>	Exploratorio / predictivo	Clasificar o predecir a partir de reglas jerárquicas	Transversal o longitudinal	Inferencial	Combinan exploración y predicción; útiles en bases de datos extensas
<b>Análisis de conglomerados (Cluster Analysis)</b>	Exploratorio	Agrupar casos según similitud multivariante	Transversal	Descriptivo	Identifica tipologías o perfiles; no requiere hipótesis previa
<b>Modelos de series temporales / ARIMA / VAR</b>	Explicativo / predictivo	Analizar evolución de variables a lo largo del tiempo	Longitudinal	Inferencial	Requiere estructura temporal; útil en estudios de seguimiento o tendencias

**Tabla E5**

*Tipos de estudios más utilizados en el ámbito educativo con poblaciones completas y sus objetivos principales.*

<b>Generalidades</b>	<b>Tipos de estudios más utilizados en el ámbito educativo</b>				
	<b>Diagnóstico Educativo (Presente estudio)</b>	<b>Correlacional</b>	<b>Explicativo / Causal</b>	<b>Descriptivo Censal</b>	<b>Evaluativo</b>
<b>Objetivo general</b>	Describir el estado actual de un proceso educativo y sus factores asociados.	Analizar la relación o grado de asociación entre variables.	Determinar causas o efectos de las variables sobre un fenómeno.	Caracterizar una población sin inferencia estadística.	Evaluar el impacto o eficacia de programas educativos.
<b>Tipo de población</b>	Población total (universo completo).	Muestra o población.	Muestra representativa.	Población completa.	Muestra o subpoblaciones.
<b>Temporalidad</b>	Transversal.	Transversal.	Transversal o longitudinal.	Transversal.	Longitudinal o comparativa.
<b>Finalidad analítica</b>	Descriptiva y explicativa.	Descriptiva y asociativa.	Explicativa y predictiva.	Descriptiva.	Explicativa y valorativa.
<b>Tipo de variables</b>	Categorías y continuas.	Cuantitativas o mixtas.	Dependiente e independientes.	Categorías y continuas.	Cuantitativas y cualitativas.

Continuación...

<b>Tipos de estudios más utilizados en el ámbito educativo</b>					
<b>Generalidades</b>	<b>Diagnóstico Educativo (Presente estudio)</b>	<b>Correlacional</b>	<b>Explicativo / Causal</b>	<b>Descriptivo Censal</b>	<b>Evaluativo</b>
<b>Técnicas estadísticas comunes</b>	Regresión logística binaria, ACM.	Correlación, regresión lineal, SEM.	MANOVA, regresión múltiple, SEM.	Estadística descriptiva, ACM, PCA.	ANOVA, MANOVA, modelos mixtos.
<b>Uso de pruebas de hipótesis</b>	No aplica (población total).	Sí, para medir fuerza y significancia de asociación.	Sí, para contrastar efectos causales.	No aplica.	Sí, para verificar diferencias antes y después.
<b>Ejemplo de aplicación</b>	Análisis del proceso de admisión universitario.	Relación entre motivación y rendimiento académico.	Influencia del entorno socioeconómico en logros educativos.	Caracterización del perfil estudiantil de una institución.	Evaluación de impacto de un programa de tutorías.

*Nota.* Los estudios indicados corresponden a los enfoques más comunes en investigaciones educativas aplicadas. Su elección depende del objetivo general, tipo de población y naturaleza de las variables. En estudios censales, el análisis suele ser descriptivo o explicativo sin pruebas inferenciales.