



**Factores que afectan la permanencia de los estudiantes en el programa de
subsidio económico de la Universidad del Valle**

**Carolina García Ospina
Nicole Dayana Sepúlveda Sánchez**

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2023

**Factores que afectan la permanencia de los estudiantes en el programa de
subsidio económico de la Universidad del Valle**

**Carolina García Ospina
Nicole Dayana Sepúlveda Sánchez**

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico(a)

Director(a):
Ph.D. Álvaro José Flórez Poveda
Codirector(a):
Ph.D. Luz Adriana Pereira Hoyos

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2023

Agradecimientos

En primer lugar, expresamos nuestro agradecimiento a Dios por habernos permitido alcanzar este punto en nuestro proceso académico y por darnos fuerza en los momentos de debilidad. A nuestro Director Alvaro José Flórez Poveda Ph.D. en Estadística y nuestra Codirectora Luz Adriana Pereira Hoyos Ph.D. en Estadística quienes desde su conocimiento y experiencia nos guiaron a lo largo del camino. De igual manera, nuestros agradecimientos para Alfonso Puchana Rojas, jefe de la Sección de Desarrollo Humano y Oderay Alegría Cobo Coordinadora del Programa de Subsidio Económico de la Universidad del Valle por brindarnos el archivo de datos para el presente estudio.

A mis padres, por su apoyo continuo, sus consejos y la motivación que me dieron a lo largo de mi carrera, a mis hermanas quienes siempre han estado a mi lado, brindándome palabras de aliento y motivación en los momentos en que más lo necesitaba. Finalmente, a mi compañera Nicole, su apoyo incondicional fue fundamental para mi proceso. *Carolina García Ospina.*

A mi hermano, por proponerme y darme la confianza para iniciar la carrera, a mis padres, por apoyarme y motivarme en todo momento, a mi tío, por ser un ejemplo a seguir, a mi novio, por su paciencia y confianza en mi proceso y finalmente a mi compañera Carolina, por su acompañamiento y apoyo en los momentos más difíciles. *Nicole Dayana Sepúlveda Sánchez*

Resumen

La capacidad económica es un factor importante en la vida de un estudiante para poder mantenerse dentro de una institución educativa y suplir tanto las necesidades básicas personales como las académicas, es por esto que las universidades ofrecen diferentes ayudas económicas como lo son los subsidios económicos y becas alimenticias para dar un soporte económico a los estudiantes que les permita culminar su etapa académica.

En este caso particular la Universidad del Valle, ofrece a los estudiantes que cumplen con los requisitos establecidos un subsidio económico el cual se espera que sea invertido en equipo tecnológico para el uso académico, en materiales solicitados y/o suplir necesidades básicas. Sin embargo, hay casos de estudiantes que cuentan con este subsidio económico y terminan perdiéndolo antes del tiempo establecido, es por esto por lo que se quiere estudiar cuáles son los factores que están afectando a los estudiantes en su permanencia dentro del subsidio económico por medio de un modelo multinomial que permite identificar en presencia de cuáles variables es más probable que un estudiante pierda el subsidio económico dentro de los cinco semestres definidos. En el estudio se evidenció que el semestre académico, la ciudad de procedencia y el tipo de régimen al que está adscrito el estudiante son variables que influyen a la hora de que un estudiante deserte del programa del subsidio económico.

Palabras clave: Multinomial, Subsidio, factores.

Abstract

The economic capacity is an important factor in a student's life to be able to sustain themselves within an educational institution and meet both their personal and academic basic needs. That is why universities offer different financial aids such as economic subsidies and meal scholarships to provide financial support to students, allowing them to complete their academic journey.

In this particular case, the University of Valle offers an economic subsidy to students who meet the established requirements. It is expected that this subsidy will be invested in technological equipment for academic use, requested materials, and/or meeting basic needs. However, there are cases of students who receive this economic subsidy but end up losing it before the established time. Therefore, there is a desire to study the factors that are affecting students in their ability to maintain the economic subsidy using a multinomial model that allows identifying which variables are more likely to cause a student to lose the economic subsidy within the defined 5 semesters. The study shows that the academic

semester, the city of origin and the type of system the student is enrolled in are variables that influence the student's decision to drop out of the subsidy program.

Keywords: Multinomial, factors, subsidy.

Contenido

Resumen	VII
Lista de Figuras	XII
Lista de Tablas	XIV
1 Introducción	2
2 Planteamiento del problema	5
3 Objetivos	7
3.1 General	7
3.2 Específicos	7
4 Justificación	8
5 Antecedentes	9
6 Marco teórico	12
6.1 Subsidio económico universitario	12
6.2 Modelo Lineal Generalizado	12
6.3 Modelo de regresión logística multinomial	13
6.3.1 Distribución multinomial	13
6.3.2 Respuestas ordinales: logit acumulativo	13
6.3.3 Supuesto de odds proporcionales	14
6.4 Estimación por máxima verosimilitud	14
6.5 Modelo lineal mixto	15
6.6 Modelo logístico multinomial con intercepto aleatorio	16
6.7 Estimación por máxima verosimilitud	16
6.8 Prueba de hipótesis	17
6.8.1 Prueba de razón de verosimilitud	17
6.9 Selección de variables	18
6.10 Criterios de información	18
6.11 Separación completa	19

7	Metodología	20
7.1	Archivo de datos	20
7.2	Identificación de variables	21
7.2.1	Variable de respuesta	21
7.2.2	Variables explicativas	21
7.3	Análisis exploratorio de los datos	23
7.4	Modelo estadístico planteado	23
7.4.1	Modelo de regresión logística multinomial con efecto aleatorio	23
7.4.2	Modelo de regresión logística multinomial	24
7.4.3	Planteamiento del modelo de regresión logística multinomial	24
7.5	Selección de variables	25
7.5.1	Prueba de razón de verosimilitud	25
7.6	Comprobación del supuesto de odds propocionales	25
8	Resultados	27
8.1	Análisis exploratorio de los datos	27
8.1.1	Análisis univariado de la información personal	27
8.1.2	Análisis univariado de la información académica	28
8.2	Análisis bivariado	29
8.3	Modelo logístico multinomial con efecto aleatorio	33
8.4	Selección de variables	33
8.4.1	Prueba de razón de verosimilitud	34
8.5	Supuesto de odds proporcionales	34
8.6	Modelo reducido ajustado	34
8.6.1	Predicción para casos particulares	36
9	Conclusiones y recomendaciones	38
9.1	Conclusiones	38
9.2	Recomendaciones	39
	Bibliografía	40

Lista de Figuras

7-1	Pasos para la metodología	20
8-1	Relación entre el número de semestres subsidiados y el género	29
8-2	Relación entre el número de semestres subsidiados y ciudad de procedencia .	30
8-3	Relación entre el número de semestres subsidiados y estrato de residencia . .	31
8-4	Relación entre el número de semestres subsidiados y régimen de EPS	31
8-5	Relación entre el número de semestres subsidiados y la Facultad	32
8-6	Probabilidades de recibir el subsidio dado el promedio acumulado y semestre	36
8-7	Probabilidades de recibir el subsidio dado el régimen y la ciudad de procedencia	37

Lista de Tablas

7-1	Descripción de las variables explicativas	22
8-1	Análisis de las variables de información personal	27
8-2	Análisis de las variables de información académica	28
8-3	Prueba de verosimilitud: Mod completo vs Mod con Efecto Aleatorio	33
8-4	Algoritmo de selección: Método Backward	33
8-5	Prueba de verosimilitud: Mod completo vs Mod reducido	34
8-6	Resumen del modelo reducido	35
8-7	Predicción para un caso específico propuesto	36

Abreviaturas

Abreviatura	Término
<i>AIC</i>	Criterio de información de Akaike
<i>ASES</i>	Programa de Acompañamiento y Seguimiento Estudiantil
<i>BIC</i>	Criterio de información bayesiano de Schwarz
<i>GLM</i>	Modelos Lineales Generalizados
<i>MEN</i>	Ministerio de Educación en Colombia
<i>UV</i>	Universidad del Valle
<i>VBU</i>	Vicerrectoría de Bienestar Universitario

Declaración

Nos permitimos afirmar que hemos realizado el presente Trabajo de Grado de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en el propio trabajo. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los hemos reconocido. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de Tesis o Trabajo de Grado.

Igualmente declaramos que los datos utilizados en este trabajo están protegidos por las correspondientes cláusulas de confidencialidad.

Santiago de Cali, 15.12.2023

Carolina García Ospina

Nicole D. Sepúlveda

1 Introducción

La deserción estudiantil en educación superior según el Sistema para la Prevención de la Deserción de la Educación Superior (SPADIES), es el estado de un estudiante que de manera voluntaria o forzosa no registra matrícula por dos o más periodos académicos consecutivos del programa en el que se matriculó; y no se encuentra como graduado, o retirado por motivos disciplinarios. La deserción es el resultado del efecto de diferentes factores como individuales, académicos, institucionales, y socioeconómicos SPADIES (2022).

Lo anterior resulta ser un desafío en la educación superior en Colombia, considerando que en el año 2021, según el Ministerio de Educación Nacional MEN (2021) se registraron tasas de deserción del 25,47 % para la educación universitaria, del 34,01 % para la educación tecnológica, y del 36,54 % para la educación técnico-profesional. Lo anterior revela que alrededor de un tercio de los estudiantes que se inscriben en estos programas educativos no logran completar sus estudios.

El principal factor que se relaciona con la deserción estudiantil son las dificultades económicas. Según el SPADIES (2018) los jóvenes de ingresos inferiores han sido identificados como la población más vulnerable en términos de deserción universitaria, debido a la falta de apoyo financiero. Lo anterior implica que estos estudiantes se encuentren en la difícil situación de no contar con los ingresos suficientes para asumir el sostenimiento de la carrera como la financiación de la matrícula.

Por lo anterior, se considera importante centrarse en factores de prevención mediante el enfoque de la permanencia estudiantil definida por Velázquez y González (2017) como el período del tiempo durante el cual el estudiante se mantiene y cursa su nivel académico hasta que pueda finalizarlo satisfactoriamente. Con relación a esto, González (2002) señala que cuando un estudiante cuenta con un respaldo económico suficiente para sufragar sus gastos académicos se crea un entorno más propicio y atractivo para su asistencia a clases, ya que al disponer de todas las facilidades para acceder a la educación, el estudiante se encuentra en una posición más productiva. Sin embargo, es importante reconocer que cuando un estudiante carece de los recursos suficientes, puede priorizar otras actividades con el fin de cubrir sus necesidades, lo cual puede conllevar a tener un impacto negativo en su rendimiento académico y, en algunos casos, incluso llevar al abandono definitivo de sus estudios.

En este contexto, Velasco (2016) destaca que las universidades suelen ofrecer una variedad de programas socioeconómicos por medio de la dirección de bienestar universitario; los programas incluyen servicios como restaurantes universitarios, auxilios económicos y exoneración de matrículas basadas en recursos económicos, excelencia académica y participación en eventos, entre otros; Sin embargo, señala que la demanda de ayudas económicas rebasa la cantidad de recursos disponibles. Específicamente, la Vicerrectoría de Bienestar Universitario (VBU) de la Universidad del Valle (UV) se preocupa por el bienestar de sus estudiantes y, a través de la Sección de Desarrollo Humano y Promoción Socioeconómica, ofrece ayudas económicas condicionadas para aquellos estudiantes que se encuentren en situaciones de vulnerabilidad. De acuerdo con Valencia (2019) el objetivo es contribuir a la formación integral y al bienestar de todos los miembros de la comunidad universitaria, propiciando espacios adecuados para el desarrollo individual y colectivo.

En relación con esto, surge el Programa de Subsidio Económico, que tiene como objetivo influir positivamente en el rendimiento académico, la permanencia en la universidad y la culminación de los estudios superiores, dado que proporciona un apoyo económico a los estudiantes que cumplan con ciertos requisitos, garantizando así su continuidad en la universidad; Los requisitos para acceder a este beneficio incluyen residir fuera de la ciudad de Cali o en barrios periféricos distantes de la Universidad del Valle (UV), tener una clasificación socioeconómica correspondiente a los estratos 1, 2 y 3, estar matriculado en un programa académico a partir del segundo semestre, contar con un promedio acumulado igual o superior a tres (3.0) y tener matriculado en forma definitiva un mínimo de nueve (9) créditos para las tecnologías y de doce (12) para los programas profesionales. es importante destacar que este beneficio puede ser otorgado por un máximo de cinco (5) períodos académicos durante el programa inicialmente matriculado. Sin embargo, existen circunstancias que podrían llevar a su retiro, como la cancelación del semestre académico, bajo rendimiento académico, presentación de fraude o inexactitud en la documentación proporcionada, o la imposición de una sanción disciplinaria según lo establecido en el reglamento.

En este trabajo de grado, se ha desarrollado una propuesta metodológica para modelar la probabilidad de que un estudiante pierda el beneficio del subsidio económico antes de los cinco semestres establecidos por la VBU, tomando en consideración las condiciones iniciales con las que el estudiante accede al subsidio. Estas pueden ser de naturaleza académica, como el semestre en el que comienzan a recibir el subsidio, el promedio acumulado y la facultad a la que pertenecen. También se tienen en cuenta variables sociodemográficas, como el estrato de residencia y el lugar de procedencia, así como aspectos personales, como el género y el régimen de salud al que están adscritos.

La propuesta metodológica se basa en el uso de modelos de regresión logística multinomial, los cuales son adecuados para analizar situaciones en las que existen más de dos categorías o resultados posibles. Se emplea dicha metodología porque permite estimar la probabilidad de que un estudiante pierda el subsidio antes de culminar el límite de semestres reglamentarios, teniendo en cuenta que las categorías pueden tomar un valor desde 1 hasta 5 semestres subsidiados. De igual manera, se puede determinar cómo distintas características de un estudiante puede influir en la probabilidad de que pierda el subsidio.

El trabajo fue desarrollado a lo largo de 9 capítulos. El primer Capítulo presenta la introducción donde se realiza una contextualización del trabajo. El segundo Capítulo aborda el planteamiento del problema, identificando la problemática a estudiar; en el tercer Capítulo se presentan los objetivos planteados. En el Cuarto capítulo se presenta la justificación de llevar a cabo este trabajo y su relevancia, seguido de esto, en el quinto Capítulo se expone la revisión de literatura, que incluye los antecedentes y estudios previos que se tuvieron en cuenta para la construcción de este. a continuación en el sexto Capítulo se desarrolla el marco teórico, que proporciona información relevante que aporte en la comprensión del trabajo, seguido de esto, en el séptimo Capítulo se describe la metodología desarrollada, con el fin de presentar claridad en los procesos que permitieron el cumplimiento de los objetivos. Posteriormente, en el octavo Capítulo se presentan los resultados obtenidos a partir de la metodología utilizada. Finalmente, en el noveno Capítulo se presentan las conclusiones obtenidas y las recomendaciones para futuros trabajos que se puedan derivar a partir de este.

2 Planteamiento del problema

Uno de los principales problemas que enfrenta el sistema de educación superior colombiano concierne a los altos niveles de deserción académica en el pregrado. Pese a que los últimos años se han caracterizado por aumentos de cobertura e ingreso de estudiantes nuevos, el número de alumnos que logra culminar sus estudios superiores no es alto, dejando entrever que una gran parte de éstos no continúan con sus estudios, principalmente en los primeros semestres. Según estadísticas, de cada cien estudiantes que ingresan a una institución de educación superior cerca de la mitad no logra culminar su ciclo académico y obtener la graduación (Guzman et al., 2009). De igual forma, Gutierrez et al. (2021) señalan que la principal razón para el abandono de los estudios superiores en Colombia es la dificultad económica para costear sus estudios (49,3 %).

La UV presenta en promedio unas tasas de deserción del 40 % en los primeros cinco semestres y una deserción total mayor al 50 % (Ramirez, 2019). Siendo esta una cifra muy aproximada de los promedios que se observan en Colombia y en Latinoamérica, la cual está alrededor del 50 %. De esta forma, la deserción se concentra en los primeros cuatro semestres. De igual forma, Escobar et al. (2006) identificaron que en la UV el factor socioeconómico, psicosocial y afectivo están asociados a la deserción.

En consecuencia de lo anterior, el MEN (2009) señala que el financiamiento del costo de la matrícula y los gastos de sostenimiento de un estudiante son dos de los principales determinantes del abandono de sus estudios en educación superior. Por ello, el Gobierno Nacional ha fortalecido el financiamiento a la demanda a través de un esquema de financiación liderado por el ICETEX, entidad que apoya a estudiantes de todas las instituciones de educación superior con créditos educativos y subsidios para sostenimiento, focalizando la atención en la población con menores ingresos del país. Al dirigir sus esfuerzos hacia aquellos que enfrentan mayores dificultades económicas, el Gobierno Nacional demuestra su compromiso por garantizar igualdad de oportunidades en la educación superior. Al proporcionar apoyo financiero a los estudiantes con menores recursos, se busca reducir las barreras económicas que podrían llevar al abandono de los estudios.

En este caso específico, la UV brinda apoyo económico a estudiantes vulnerables que se encuentran en riesgo de interrumpir su proceso académico, otorgándoles un aporte

financiero semestral durante un máximo de cinco (5) semestres. A pesar del interés de la UV en brindar apoyo a sus estudiantes y promover su permanencia, se ha observado, según las actas de asignación publicadas por la VBU (2022) para los años 2016 a 2022, que aproximadamente el 24 % de los estudiantes pierden este beneficio cada semestre debido a diversas características.

Al emplear modelos de regresión logística multinomial, se busca comprender cómo distintas características de los estudiantes pueden influir en la probabilidad de perder el subsidio antes de culminarlo. Estas características incluyen el promedio acumulado, número de semestres matriculados, la facultad a la que pertenecen, el estrato de residencia, el lugar de procedencia, el género y el régimen de salud. El uso de modelos de regresión logística multinomial permite analizar estas variables simultáneamente y determinar su impacto relativo en la probabilidad de perder el subsidio en cada uno de los cinco semestres.

3 Objetivos

3.1. General

Evaluar los efectos de las características, tales como Ciudad, Estrato, Facultad, Género, entre otras, que afectan en la permanencia de los estudiantes en el programa de subsidio económico de la Universidad del Valle.

3.2. Específicos

- Ajustar un modelo estadístico que permita estimar la probabilidad de que el estudiante pierda el subsidio económico antes de los 5 semestres establecidos.
- Determinar los factores iniciales en los estudiantes que podrían tener un impacto en la eventual pérdida del subsidio económico.

4 Justificación

A pesar del constante interés y compromiso de la UV en brindar apoyo económico a través del programa de subsidio económico estudiantil, según las actas de asignación publicadas por la VBU (2022) cerca del 25 % de los estudiantes que reciben este beneficio lo terminan perdiendo antes de los cinco semestres. Lo anterior se debe a diversos factores fundamentales que inciden en la continuidad y éxito de los estudiantes en el programa de bienestar universitario. Estos factores pueden variar en función de las circunstancias y condiciones individuales de cada estudiante, algunas de las situaciones más comunes que influyen negativamente en la permanencia en el programa incluyen el bajo rendimiento académico, la cancelación del semestre académico, incurrir en una sanción disciplinaria, entre otras.

Dado lo anterior, resulta fundamental identificar los factores que impiden a los estudiantes cumplir la totalidad de periodos del programa, ya que al identificar estas causas, se podrá implementar estrategias y políticas que promuevan una mayor permanencia en el programa y promover el éxito estudiantil en la universidad. Así mismo, la importancia de lo expuesto radica en la posibilidad de llevar a cabo un seguimiento efectivo de los estudiantes desde el programa de subsidios. Este seguimiento permitiría identificar posibles factores de riesgo y, en consecuencia, prevenir la pérdida del subsidio a través de un acompañamiento integral al estudiante. (MEN, 2014)

Analizar la permanencia de los estudiantes en el programa de subsidio económico por medio de un modelo de regresión multinomial permite calcular la probabilidad de que el estudiante pierda el beneficio en 5 semestres o menos. De igual forma, se pueden identificar factores de riesgo para la pérdida del subsidio. Estos pueden ser que su lugar de residencia sea fuera de Cali, estar matriculado en un semestre académico avanzado o que el lugar de vivienda sea de un estrato alto, entre otros.

Dado que los estudiantes que pertenecen a la misma facultad tienen en común muchas características, como clases, profesores, horarios, etc, es posible que la probabilidad de perder el subsidio esté correlacionada. Por esta razón, se quiere considerar la adición de un efecto aleatorio asociado a la facultad a la que pertenece el estudiante, con el objetivo de cuantificar esta correlación en el proceso de ajuste del modelo.

5 Antecedentes

En una investigación llevada a cabo por Garbanzo (2013), se exploraron los factores que influyen en el rendimiento académico de estudiantes universitarios en sus últimos semestres, centrándose en la influencia del nivel socioeconómico. El objetivo principal fue analizar la población de 120 estudiantes que ingresaron a la universidad de Costa Rica en el 2003 y de la cual se dividió en dos grupos, el primer grupo de 60 estudiantes que tenían la beca y el segundo grupo de 60 estudiantes que nunca solicitaron la beca, enfocándose así en encontrar las diferencias entre aquellos beneficiarios del subsidio y aquellos que no solicitaron dicho beneficio. La investigación permitió identificar que el rendimiento académico de los estudiantes es un fenómeno multicausal y complejo, en el que se analizan distintos tipos de componentes, tales como: factores sociales, personales e institucionales, sin embargo, en el trabajo se expresa que existe una diferencia en las distintas variables de mejora e índice de satisfacción para los estudiantes becados y quienes no estuvieron becados. Así pues, se puede concluir que, en el caso de los estudiantes beneficiarios, se percibe una mayor percepción positiva en cuanto a los servicios ofertados por la Universidad y así mismo, una gran diferencia en las condiciones económicas.

En su trabajo de grado, Valencia (2019) abordó una problemática específica al evaluar el programa de Subsidio Económico de la UV, donde el objetivo principal de la investigación fue analizar el efecto del beneficio en la permanencia estudiantil de los beneficiarios que ingresaron a la universidad en el año 2011. El estudio se llevó a cabo siguiendo una metodología evaluativa basada en la experimentación, donde se estableció un grupo de control compuesto por estudiantes que ingresaron en el mismo año, pero que no fueron beneficiados por el apoyo económico del Subsidio. Los estudiantes del grupo de control presentaron características similares a aquellos que sí recibieron el subsidio. Se realizó un seguimiento de los estudiantes hasta el año 2017, con la expectativa de que se hubieran graduado para ese momento y se encontró que el recibir el beneficio no solo económico sino también académico influye de manera positiva sobre el estudiante para mitigar el problema de deserción. El estudio se centró en analizar los aspectos positivos del programa ofrecido por la VBU; Sin embargo, se enuncia que no se presentaron recomendaciones o mejoras específicas para el mismo. Estas recomendaciones son de gran relevancia, ya que permiten al programa analizar si los recursos se están distribuyendo de manera adecuada y realizar ajustes necesarios para su óptimo funcionamiento.

En su trabajo de grado López (2021) estuvo interesada en determinar el impacto que tienen los programas ofrecidos por la VBU, tales como los subsidios económicos y las becas de alimentación sobre el rendimiento académico y la permanencia de los estudiantes pertenecientes a las sedes de Cali quienes ingresaron en el año 2012. Para lograr su objetivo, la autora hace uso de un método estadístico llamado “Propensity Score Matching”, que consiste en reforzar los argumentos sobre la causalidad de las relaciones entre las variables planteadas de tipo dummy, continuas y categóricas, entre las cuales se encuentra el estrato socioeconómico, el nivel educativo de la madre, el grupo étnico, el número de hijos, entre otras. El método apoya en la disminución del impacto del sesgo de selección que surge debido a la imposibilidad de asignar aleatoriamente a los individuos a un grupo de control. Los resultados obtenidos permiten concluir que los estudiantes beneficiarios de los programas ofrecidos por la VBU obtienen un mejor rendimiento académico y logran graduarse en el tiempo establecido por la Universidad.

Por otra parte Gonzalez and Martinez (2022) realizaron un estudio para conocer cuáles son los factores que influyen en la deserción de los estudiantes admitidos por Condición de Excepción y el programa Ser Pilo Paga III de la UV en el año 2017 y durante su proceso de acompañamiento en el programa ASES (Acompañamiento y Seguimiento Estudiantil), la metodología planteada para el cumplimiento de este fue un modelo logístico en donde tuvieron en cuenta la facultad como efecto aleatorio, dado que refleja las características específicas de cada una y su influencia en la probabilidad de perder el subsidio. Los resultados que obtuvieron señalan que la condición de admisión de los estudiantes no es significativa, es decir, el comportamiento de deserción entre ambos grupos que tomaron de estudio eran similares. También resaltan que los factores que más inciden en la deserción estudiantil son: la inasistencia a los encuentros socio-educativos entre el estudiante y el monitor, el número de bajos rendimientos académicos y el sexo del estudiante, debido a que un estudiante del sexo masculino tiene 2.6 veces más probabilidades de desertar que una estudiante de sexo femenino.

Escobar et al. (2006) presentan una visión general del comportamiento sobre la permanencia, deserción y la graduación universitaria en la UV para las cohortes que ingresaron desde 1994 hasta 2001, la metodología que se aplicó fueron las funciones de supervivencia de los estudiantes mediante el método no paramétrico de Kaplan-Meier y el riesgo de supervivencia utilizado se asignó mediante la distribución Weibull. El autor resalta que los apoyos de la VBU permiten favorecer la permanencia prolongada en los estudios universitarios, además, menciona que las variables de ingreso de los hogares y el estrato socioeconómico no se identificaron como características riesgosas para desertar de un programa académico. En conclusión la investigación permitió asociar que en la universidad, al igual que en otros contextos nacionales e internacionales, la deserción es un fenómeno que presenta un mayor peso sobre las personas de sexo masculino, dado que la graduación se asocia con un mayor

porcentaje a las personas de sexo femenino.

6 Marco teórico

6.1. Subsidio económico universitario

El subsidio económico está reglamentado por la Resolución No 091 del 13 de diciembre de 2004 del Consejo Superior de la UV, es un apoyo que brinda la Universidad a los estudiantes regulares de pregrado de las sedes de Cali, cuya situación de vulnerabilidad económica pone en riesgo su continuidad académica. Este auxilio consiste en un aporte mensual del 30 % del Salario mínimo legal mensual vigente (SMMLV), que se entrega en el transcurso del semestre y aplica máximo para cinco períodos académicos.

Los apoyos económicos están orientados a estudiantes que forman parte de grupos de inclusión, como las comunidades étnicas y las personas con discapacidad y/o con condiciones socioeconómicas desfavorables de los estratos socioeconómicos iguales o inferiores a tres y preferiblemente residentes fuera del perímetro urbano de la institución. Por lo que se estaría priorizando la población que históricamente ha tenido mayor dificultad para acceder a la educación superior, entre los cuales se destacan indígenas, afrodescendientes, población diversa, madres cabeza de familia, población víctima del conflicto, reinsertados-reincorporados, campesinos, entre otros, con un enfoque interseccional de los casos.

El proceso inicia con la convocatoria semestral a todos los estudiantes pertenecientes a esta clasificación socioeconómica, para que realicen las solicitudes; luego un equipo de profesionales en trabajo social realizan las entrevistas y visitas domiciliarias, seguido de lo anterior, estudian y analizan los casos para finalmente realizar la asignación del beneficio, teniendo en cuenta que un cuerpo colegiado conformado por el Vicerrector de Bienestar Universitario, el jefe de la sección de desarrollo humano, el coordinador del proyecto y dos representantes estudiantiles son los encargados de tomar las decisiones.

6.2. Modelo Lineal Generalizado

El modelo lineal generalizado (GLM por sus siglas en inglés) es una generalización de los modelos lineales que permiten modelar variables de respuesta cuya distribución de probabilidad pertenece a la familia exponencial, la motivación de su desarrollo radica en

que en muchas ocasiones las respuestas no siguen una distribución normal, como es el caso de variables que representan conteos o proporciones. De esta manera, el GLM es apto para abordar esos casos ya que tiene la capacidad de ajustarse a diversas distribuciones de respuesta Montgomery et al. (2022).

Sean $Y_1, Y_2, Y_3, \dots, Y_n$ variables aleatorias independientes. De acuerdo con Dobson and Barnett (2018), el GLM asume que Y_i sigue una distribución que pertenece a la familia exponencial, esto es:

$$f(y; \theta) = \exp[y(b(\theta)) + c(\theta) + b(y)]$$

donde, $c(\theta)$ y $b(y)$ son funciones conocidas, $b(\theta)$ el parámetro natural de la distribución o función canónica.

Dado esto, se está interesado en modelar la media μ_i por medio de un vector de predictores, siendo así:

$$g(\mu_i) = x_i^T \beta$$

Donde x_i es el vector de predictores correspondiente a la i -ésima observación, β es el vector de coeficientes y g es monótona y diferenciable que representa la función de enlace.

6.3. Modelo de regresión logística multinomial

6.3.1. Distribución multinomial

Según Agresti (2012), se denota el número de categorías de respuesta como c . Para el sujeto i , sea π_{ik} la probabilidad de respuesta en la categoría k , con $\sum_{k=1}^c \pi_{ik} = 1$. La elección de la categoría es el resultado de un solo ensayo multinomial. Sea $y_i = (y_{i1}, \dots, y_{ic})$ el ensayo multinomial para el sujeto i , $i = 1, \dots, N$, donde $y_{ik} = 1$ cuando la respuesta está en la categoría k y $y_{ik} = 0$ en caso contrario. Entonces $\sum_{k=1}^c y_{ik} = 1$, y la distribución de probabilidad multinomial para ese sujeto es

$$p(y_{i1}, \dots, y_{ic}) = \pi_{i1}^{y_{i1}} \dots \pi_{ic}^{y_{ic}}.$$

6.3.2. Respuestas ordinales: logit acumulativo

Según Agresti (2012), cuando los modelos tienen variables de respuesta ordinal presentan tendencias monótonas, en la que las respuestas caen en categorías más altas o más bajas a

medida que aumenta el valor de una variable explicativa. Este tipo de modelos son más parsimoniosos que los modelos con respuesta nominal, ya que potencialmente requieren menos parámetros.

6.3.3. Supuesto de odds proporcionales

Sea y_i la categoría de respuesta para el sujeto i . Es decir, $y_{ik} = k$ significa que $y_{ik} = 1$ y $y_{ij} = 0$ para $t \neq k$ para los c indicadores multinomiales. Con el fin de usar el orden de categorías, los modelos se expresan en términos de probabilidades acumuladas en la siguiente ecuación.

$$P(y_i \leq k) = \pi_{i1} + \dots + \pi_{ik}, \quad k = 1, \dots, c.$$

Los logits acumulativos son logits de estas probabilidades acumulativas como se presenta en la ecuación (6-1)

$$\text{logit}[P(y_i \leq k)] = \log \frac{P(y_i \leq k)}{1 - P(y_i \leq k)}, \quad k = 1, \dots, c - 1 \quad (6-1)$$

Así pues, la ecuación del modelo logit acumulado está dada por:

$$\text{logit}[P(y_i \leq k)] = \alpha_k + x_i^T \beta, \quad k = 1, \dots, c - 1 \quad (6-2)$$

Donde cada logit acumulativo tiene su propio intercepto (α_k) y los interceptos son crecientes en k , mostrándose los términos de cada uno por sí mismo porque dependen de k , en cambio los efectos no. Además, este modelo establece que los efectos de β de la variable explicativa son los mismos para cada logit acumulativo.

6.4. Estimación por máxima verosimilitud

De acuerdo con Agresti (2012), se puede ajustar modelos de regresión logística multinomial por medio de la estimación de máxima verosimilitud partiendo de la función (6-3)

$$L(\alpha_1, \dots, \alpha_c, \beta) = \left(\prod_{i=1}^n \left[\prod_{k=1}^c \pi_k(x_i)^{y_{ik}} \right] \right) = \prod_{i=1}^n \left\{ \prod_{k=1}^c \left[\frac{\exp(\alpha_k + x_i^T \beta)}{1 + \exp(\alpha_k + x_i^T \beta)} - \frac{\exp(\alpha_{k-1} + x_i^T \beta)}{1 + \exp(\alpha_{k-1} + x_i^T \beta)} \right]^{y_{ik}} \right\} \quad (6-3)$$

donde;

- c : Representa el número total de categorías de la variable de respuesta.
- k : Representa el valor que puede tomar cada categoría, es decir $1, 2 \dots, c$.

Al verse como una función de (α_k, β) se puede maximizar para obtener las estimaciones de máxima verosimilitud utilizando la puntuación de Fisher (McCullagh and Nelder, 1989) o el método de Newton-Raphson.

6.5. Modelo lineal mixto

De acuerdo con Montgomery et al. (2022), el enfoque de efectos aleatorios para ampliar el modelo de regresión lineal univariable a situaciones con datos que están correlacionados o agrupados se fundamenta en la suposición de que, en el caso de cada individuo o sujeto, es posible representar la respuesta a través de un modelo de regresión lineal. Sin embargo, en este enfoque, los coeficientes de regresión se consideran específicos para cada grupo o también llamado cluster al que pertenece el individuo.

El modelo lineal mixto general parte del supuesto de que el vector de respuestas Y_i relacionado con el i -ésimo grupo cumple con lo siguiente:

$$Y_i | b_i \sim N(X_i\beta + Z_ib_i, \Sigma_i) \quad (6-4)$$

donde; $b_i \sim N(0, D)$

Para matrices de diseño conocidas X_i y Z_i con dimensiones $n_i \times p$ y $n_i \times q$ respectivamente, un vector β de p -dimensiones que contiene coeficientes de regresión desconocidos, un vector de b_i de q -dimensiones que contiene coeficientes de regresión específicos para cada grupo y se asume que siguen una distribución normal multivariada q -dimensional con una media 0 y una matriz de covarianza D , junto con una matriz de covarianza desconocida Σ_i .

Dado que el modelo contiene efectos aleatorios da lugar al término “modelos mixtos”, que marginalmente se define con la siguiente la distribución:

$$Y_i \sim N(X_i\beta, Z_i^T D Z_i + \Sigma_i) \quad (6-5)$$

Donde los componentes en β se denominan “efectos fijos” y los componentes en b_i se denominan “efectos aleatorios”.

6.6. Modelo logístico multinomial con intercepto aleatorio

Según Verbeke et al. (2009) el modelo logístico multinomial con intercepto aleatorio con el uso de la función logit está dado por:

$$\text{logit}[P(Y_{ij} \leq k | X_i, Z_i)] = \alpha_k + x_{ij}^T \beta + z_{ij}^T b_i, \quad k = 1, \dots, c - 1, \quad (6-6)$$

Donde,

X_i y Z_i son las matrices de diseño habituales para los efectos fijos y efectos aleatorios, α_k es el intercepto aleatorio asociado a la categoría k , respectivamente, x_{ij} y z_{ij} son las filas correspondientes a la j -ésima ocasión de medición, β y b_i son los vectores habituales de parámetros fijos y aleatorios.

6.7. Estimación por máxima verosimilitud

Los modelos de efectos aleatorios se pueden ajustar mediante la maximización de la verosimilitud marginal, que se obtiene integrando los efectos aleatorios donde ϕ representa los parámetros de dispersión o variabilidad asociados con la distribución de los efectos aleatorios.

Según Verbeke et al. (2009), la función de densidad marginal del grupo i se convierte en la siguiente forma:

$$f_i(y_i | \beta, D, \phi) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij} | b_i, \beta, \phi) f(b_i | D) db_i \quad (6-7)$$

A partir de esto, se deriva la función de verosimilitud marginal:

$$L(\beta, D, \phi) = \prod_{i=1}^N f_i(y_i | \beta, D, \phi) = \prod_{i=1}^N \int f_{ij}(y_{ij} | b_i, \beta, \phi) f(b_i | D) db_i \quad (6-8)$$

Al maximizar la ecuación (6-8) se presenta un desafío principal debido a la presencia de N integrales sobre los efectos aleatorios, que son q -dimensionales, representados como b_i . En algunos casos, como se muestra en la ecuación (6-5), es posible resolver estas integrales de manera analítica. Sin embargo, en general, no existen expresiones analíticas disponibles para las integrales en la ecuación (6-8), lo que requiere la aplicación de métodos de aproximación numérica. Entre estos métodos se incluyen la cuadratura gaussiana, la cuasi-verosimilitud penalizada (PQL) y la cuasi-verosimilitud marginal (MQL), como se describe en Verbeke et al. (2009). Estos métodos son utilizados para aproximar las integrales y permitir la estimación de los parámetros del modelo de efectos aleatorios.

6.8. Prueba de hipótesis

6.8.1. Prueba de razón de verosimilitud

La prueba de razón de verosimilitud busca evaluar la bondad de ajuste de dos modelos que se comparan entre sí. De esta manera, Hernández and López (2022) exponen que el objetivo es comparar si el modelo completo y el modelo reducido exhiben diferencias estadísticamente significativas. Esta comparación se puede llevar a cabo con una prueba de razón de verosimilitud, donde θ se refiere a los parámetros del modelo que se están evaluando. Suponiendo que:

$$H_0 : \theta \in \Theta_0 \text{ vs } H_a : \theta \in \Theta$$

En el caso de la prueba, el estadístico de prueba se calcularía de la siguiente manera:

$$LR = -2 \log \left(\frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} \right) \quad (6-9)$$

Por lo general, la prueba de razón de verosimilitud se formula en función de los valores de log-verosimilitud del modelo de la siguiente manera:

$$LR = -2(\log L(\hat{\Theta}_0) - \log L(\hat{\Theta})) \quad (6-10)$$

De acuerdo con Hernández and López (2022), el estadístico de prueba, sigue una distribución $LR \sim \chi^2_{p-p_0}$, donde p representa el número de parámetros en el modelo completo y p_0 el número de parámetros asumiendo que H_0 es verdadera. En este contexto, $\hat{\Theta}_0$ es el vector de parámetros estimados bajo la suposición de que H_0 es verdadera y $\hat{\Theta}$ representa el vector de estimaciones de los parámetros del modelo.

■ Prueba para Efectos Aleatorios

En el caso de la prueba para comparar un modelo sin efectos aleatorios contra uno que sí tiene efectos aleatorios, de acuerdo con Verbeke and Molenberghs (2000) se tiene que:

$$H_0 : d = 0 \text{ vs } H_a : d > 0 \quad d = \text{Varianza de los efectos aleatorios}$$

La distribución nula asintótica de $-2 \ln \lambda_N$ es una mezcla de χ^2_1 y χ^2_0 con pesos iguales de 0.5, donde la distribución χ^2_0 asigna toda la probabilidad al valor 0.

6.9. Selección de variables

Cuando se dispone de un conjunto grande de variables explicativas, es válido preguntarse si se deben agregar todas, y en caso de que no se deba, cuáles se deben agregar y que otras se deberían excluir. En términos generales, de acuerdo a lo que refiere González (2015) los métodos de selección de variables se dedican a abordar la problemática de construir y seleccionar un modelo. Al agregar cada vez más variables al modelo, el ajuste de los datos mejora, sin embargo, es posible que la precisión individual de los coeficientes disminuya al obtenerse una mayor varianza y por consecuencia se produce un sobreajuste en la función de regresión estimada. El otro escenario es incluir menos variables en el modelo, en este caso las varianzas se reducirían pero puede haber sesgos en las estimaciones. Por lo tanto, el objetivo de los métodos de selección de variables es encontrar un modelo que tenga un buen ajuste con los datos y al mismo tiempo que permita encontrar un equilibrio entre bondad de ajuste y sencillez.

Según González (2015) el objetivo de los algoritmos de selección de variables es seleccionar secuencialmente el mejor modelo, pero en cada paso incluir o ir excluyendo variables explicativas hasta dar cumplimiento con algún criterio determinado. Los algoritmos más conocidos son:

- **Método Forward Selection:** La selección hacia adelante parte del modelo nulo que no tiene covariables, en cada paso del algoritmo, se agrega una nueva variable al modelo actual en función de algún criterio hasta que no se agreguen más variables, es decir, se introduce la variable más significativa hasta que el criterio se cumpla.
- **Método Backward Selection:** La selección hacia atrás parte del modelo que incluya todas las covariables candidatas, en cada paso del algoritmo, se va eliminando las covariables una a una hasta que el criterio se cumpla.
- **Método Stepwise Selection:** La selección por segmentos parte de seguir la misma metodología de la selección hacia adelante, con la diferencia de que en la de segmentos se evalúa los candidatos que ya se habían ingresado previamente, es decir, una covariable que ya se encuentre definida en el modelo puede ser eliminada si así lo considera el paso posterior.

6.10. Criterios de información

El criterio de información de Akaike (AIC) de acuerdo con Correa and Salazar (2016), se considera un criterio paramétrico comparativo entre los modelos y no una prueba de hipótesis sobre el ajuste de un modelo, así pues este criterio representa una herramienta para la selección de modelos posibles.

Se define como:

$$AIC = -2\log(L(\Theta)) + 2r$$

donde r es el número de parámetros en el modelo y $L(\Theta)$ es el valor de la verosimilitud obtenida por el modelo estimado.

En conclusión, al comparar distintos modelos, el modelo asociado al menor AIC se considerará como el mejor.

El criterio de información bayesiano de Schwarz (BIC) es un criterio paramétrico para la selección de modelos al igual que el AIC, de acuerdo con Correa and Salazar (2016) se indica que es considerado como el mejor modelo comparado con otros, aquel que presente menor valor de BIC.

Se define como:

$$BIC = -2\log(L(\Theta)) + r \log n$$

donde r es el número de parámetros en el modelo y n es el número de observaciones que fueron usadas en el proceso de estimación.

6.11. Separación completa

De acuerdo con Agresti (2015), se dice que la separación completa en modelos de regresión logística ocurre cuando hay un hiperplano que puede dividir el espacio de valores de las variables explicativas, de tal manera que en un lado todas las observaciones tienen un valor de la variable respuesta igual a 0, mientras que en el otro lado todas las observaciones tienen un valor de la variable respuesta igual a 1. Esto implica que existe una discriminación perfecta, ya que podemos predecir de manera precisa los resultados de la muestra con solo conocer los valores de las variables explicativas. De esta manera, Agresti (2015) menciona que en la práctica, se puede sospechar de la separación completa cuando la ecuación de predicción ajustada predice de manera perfecta los resultados de la variable respuesta para todo el conjunto de datos. Es decir, la probabilidad estimada $\hat{\pi}_i$ es igual a 1, cuando $y_i = 1$, y $\hat{\pi}_i$ es = 0.0. Otro indicio es que el valor máximo de log-verosimilitud es igual a 0, también porque los errores estándar pueden parecer anormalmente grandes, lo cual puede ser otra señal de advertencia.

7 Metodología

En este capítulo se aborda la metodología presentada en la Figura 7-1, con el fin de dar cumplimiento a los objetivos establecidos anteriormente. En primer lugar, se discute sobre el archivo de datos proporcionado para el desarrollo de esta trabajo, para posteriormente aplicar un modelo logístico multinomial que permita identificar los factores que afectan la permanencia en el programa de subsidio económico de la UV.

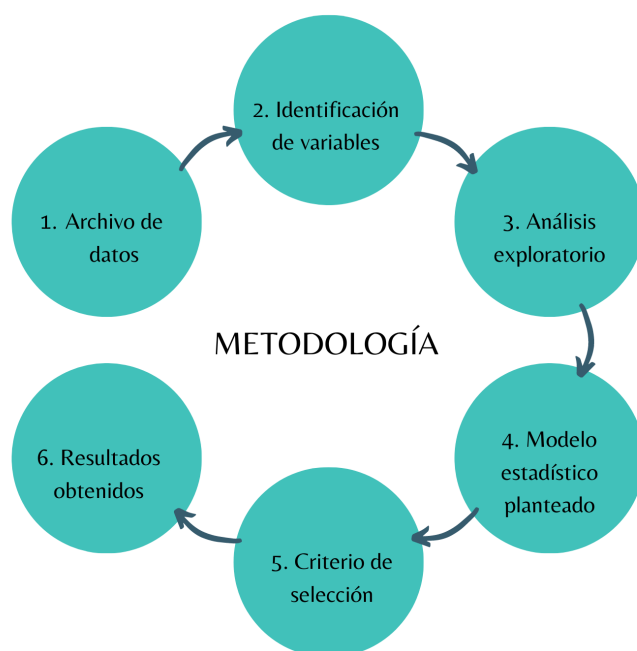


Figura 7-1: Pasos para la metodología

7.1. Archivo de datos

Para llevar a cabo la investigación, se utilizó una base de datos proporcionada por la oficina de Desarrollo Humano y Promoción Socioeconómica de la VBU. La cual comprende registros de 130 estudiantes admitidos al programa de subsidio económico en el primer semestre del 2019, matriculados en programas de pregrado y tecnológicos, quienes fueron entrevistados de manera presencial por la trabajadora social del programa.

Con el objetivo de determinar el número de semestres que el estudiante recibe el subsidio hasta perderlo, se les realizó un seguimiento durante cinco períodos académicos. Lo anterior, teniendo en cuenta que el semestre indica el número de semestres que el estudiante ha matriculado en la universidad. Esto implica que si un estudiante es beneficiario del subsidio y ha matriculado nueve semestres, es porque existe la posibilidad de que pueda completar la totalidad del beneficio, ya que el programa realiza un análisis detallado de cada caso y admite a los estudiantes que identifica que están cursando asignaturas que corresponden a semestres inferiores al quinto.

Como se mencionó anteriormente, el promedio acumulado es un criterio para la asignación del subsidio, sin embargo, en la base de datos se encuentran estudiantes de primer semestre, por lo tanto, se decide omitir a estos del análisis y finalmente, la base de datos se reduce a 123 estudiantes.

7.2. Identificación de variables

7.2.1. Variable de respuesta

Dado que el interés radica en conocer la probabilidad de que un estudiante pierda el subsidio económico antes de 5 semestres debido a ciertas características, se toma como variable de respuesta el número de semestres que el estudiante recibe el subsidio. Esta variable toma valores entre 1-5, de manera que si toma el valor de 5 el estudiante recibió el subsidio por completo.

7.2.2. Variables explicativas

Las variables que se tomaron para el estudio fueron la ciudad de procedencia del estudiante, el estrato socioeconómico del lugar de residencia del estudiante, la facultad a la que se encuentra inscrito, el género con el que se identifica el estudiante, el promedio acumulado, el semestre en el que se encuentra matriculado al ingresar al programa y finalmente el régimen de EPS al que se encuentra afiliado.

La base de datos contenía otras variables adicionales, como la condición de víctima de violencia, discapacidad, tipo de jornada, si el estudiante estaba inscrito en dos o más carreras, pertenencia a una etnia, y el número de bajos rendimientos académicos. Sin embargo, estas variables no se pudieron incorporar en el análisis dado que se presenta una notable diferencia en las proporciones entre las categorías.

Tabla 7-1: Descripción de las variables explicativas

Variable	Descripción	Categoría	Porcentaje	Tipo de variable
Ciudad	Ciudad de procedencia del estudiante	Cali Otro lugar	51.2 % 48.8 %	categorica
Estrato	Estrato del lugar de residencia del estudiante	1 2 3 4	26.0 % 49.5 % 21.9 % 2.4 %	categorica
Facultad	Facultad a la que pertenecen los estudiantes	Artes Integradas C. de la Administración C. Naturales y Exactas C. Sociales y Económicas Educación y Pedagogía Humanidades Ingeniería Psicología Salud	14.6 % 5.6 % 5.6 % 2.4 % 10.5 % 17.0 % 21.1 % 2.4 % 20.3 %	categorica
Género	Género con el que se identifica el estudiante	Femenino Masculino	70.7 % 29.3 %	categorica
Promedio	Promedio acumulado del estudiante al ingresar al programa	[3.1 - 4.9]		numérica
Semestre	Semestre que cursa el estudiante al ingresar al programa	[2 - 9]		numérica
Régimen	Régimen de EPS al cual está afiliado el estudiante	Subsidiado Contributivo	62.6 % 37.4 %	categorica

7.3. Análisis exploratorio de los datos

Inicialmente se realizó un análisis descriptivo univariado para las variables de interés propuestas en la Tabla 7-1 con el cual se espera tener una vista inicial del comportamiento estadístico de los datos y un análisis descriptivo bivariado para observar la relación entre cada una de las variables explicativas con el número de semestres que el estudiante recibe el subsidio, con el fin de determinar posibles características que afecten en la permanencia de los estudiantes en el subsidio.

7.4. Modelo estadístico planteado

7.4.1. Modelo de regresión logística multinomial con efecto aleatorio

Con el objetivo de tener en cuenta la variabilidad que existe entre los estudiantes de una facultad a otra, se propone incluir la variable “facultad” como un efecto aleatorio. A pesar de que algunas facultades presentan baja frecuencia de estudiantes subsidiados, se decide no agruparlas debido a que no comparten las mismas características académicas, por ejemplo, algunas facultades presentan mayor cantidad de asignaturas, mayor exigencia en el número de créditos, horarios más extensos, entre otras. Es por esto, que en el modelo se fija cada facultad como un factor que causa un riesgo mayor en la pérdida del subsidio económico.

Para evaluar el modelo de regresión logística multinomial con efecto aleatorio se establece inicialmente el modelo con efectos fijos, en el que se determina que todas las covariables tienen el mismo efecto sobre la pérdida del subsidio y no se incluye la covariable facultad porque es la que se desea comprobar si tiene un efecto aleatorio, tomando la ecuación inicialmente presentada.

Después de definir el modelo con efectos fijos, se procede a plantear el modelo con la variable facultad como efecto aleatorio, quedando de la siguiente manera.

$$\text{Logit}[P(y_{ij} \leq k | u_j)] = \alpha_k + \beta_1 \cdot \text{Semestre}_{ij} + \beta_2 \cdot \text{Régimen}_{ij} + \beta_3 \cdot \text{Estrato}_{2ij} + \beta_4 \cdot \text{Estrato}_{3ij} + \beta_5 \cdot \text{Estrato}_{4ij} + \beta_6 \cdot \text{Género}_{ij} + \beta_7 \cdot \text{Ciudad}_{ij} + \beta_8 \cdot \text{Promedio}_{ij} + u_j$$

Para $k = 1, 2, \dots, 5$

donde:

- y_{ij} : Número de semestres subsidiados del i-ésimo estudiante de la j-ésima Facultad.
- $u_j \sim N(0, d)$: Representa el efecto aleatorio asociado a la j-ésima Facultad.
- α_k : Intercepto aleatorio asociado a la categoría k

- $Semestre_{ij}$: Número de semestres matriculados por el estudiante.
- $Regimen_{ij}$: 1 si el estudiante pertenece al régimen subsidiado y 0 si pertenece al régimen contributivo.
- $Estrato_{2ij}$: 1 si el estrato es 2 y 0 si es lo contrario.
- $Estrato_{3ij}$: 1 si el estrato es 3 y 0 si es lo contrario.
- $Estrato_{4ij}$: 1 si el estrato es 4 y 0 si es lo contrario.
- $Género_{ij}$: 1 si el estudiante es de género masculino y 0 si es de género femenino.
- $Ciudad_{ij}$: 1 si la ciudad es diferente de Cali y 0 si es Cali.
- $Promedio_{ij}$: Promedio acumulado del estudiante al ingresar al programa de subsidios.

Para comprobar si el efecto aleatorio de la variable es significativo en el modelo, se realiza una prueba de razón de verosimilitud con el fin de verificar si la varianza de los efectos aleatorios (d) es igual a cero. La hipótesis se plantea de la siguiente manera:

$$H_0 : d = 0 \quad H_a : d > 0$$

En donde la hipótesis nula H_0 es que la varianza de los efectos aleatorios es igual a cero, mientras que la hipótesis alternativa H_1 es que la varianza es mayor que cero. Por lo tanto, se busca evidencia en contra de H_0 para determinar si el efecto aleatorio es significativo.

7.4.2. Modelo de regresión logística multinomial

Con el objetivo de modelar la probabilidad de que un estudiante beneficiado pierda el subsidio económico en cinco semestres o menos, se utiliza una extensión de los modelos lineales generalizados. En este caso, un modelo de regresión logística multinomial con función de enlace “logit”.

7.4.3. Planteamiento del modelo de regresión logística multinomial

Para el planteamiento del modelo se tuvo en cuenta todas las variables explicativas que se encontraban disponibles, obteniendo así un modelo completo al que más adelante se le realizará selección de variables para obtener un modelo óptimo.

De tal manera que el modelo propuesto es:

$$Logit[P(y_i \leq k)] = \alpha_k + \beta_1 \cdot Semestre_i + \beta_2 \cdot Régimen_i + \beta_3 \cdot Estrato_{2i} + \beta_4 \cdot Estrato_{3i} +$$

$$+\beta_5 \cdot Estrato_{4i} + \beta_6 \cdot Género_i + \beta_7 \cdot Ciudad_i + \beta_8 \cdot Promedio_i$$

$$i = 1, 2, \dots, 123, \quad k = 1, 2, \dots, 5.$$

7.5. Selección de variables

Para el ajuste del modelo es importante elegir las covariables adecuadas que ayuden a predecir la pérdida del subsidio económico, dado que si no se incluyen covariables que son importantes se corre el riesgo de presentar sesgo, mientras que si se incluyen covariables insignificantes se eleva la varianza de las estimaciones. También es relevante tener en cuenta que no se deben incluir todas las variables porque puede generar problemas de multicolinealidad. Es por esto que se opta por realizar la selección de variables mediante algoritmos de selección con un criterio específico.

Esta elección se fundamenta en la consideración de diversos criterios de selección, tales como Backward, Forward y Stepwise, en donde se evaluaron estos algoritmos utilizando el criterio de información de Akaike (AIC). En este caso, se ha optado por utilizar el algoritmo de selección Backward, que permite inicialmente incluir todas las covariables posibles, y luego eliminar de manera iterativa aquellas que menos contribuyen a la predicción de la pérdida del subsidio económico.

La elección final de Backward se basa en la observación de que este método presentaba un valor de AIC más bajo en una de las iteraciones del proceso de selección del modelo. Este enfoque proporciona un equilibrio entre la inclusión inicial de todas las covariables y la eliminación iterativa basada en el rendimiento del modelo.

7.5.1. Prueba de razón de verosimilitud

Para determinar si el modelo obtenido mediante la selección de variables proporciona el mismo ajuste que el modelo completo, se lleva a cabo una prueba de razón de verosimilitud con un nivel de significancia de $\alpha = 0,05$.

7.6. Comprobación del supuesto de odds proporcionales

La prueba de supuestos de odds proporcionales se utiliza para evaluar la validez del supuesto de proporcionalidad de odds en el modelo logístico multinomial. Este supuesto de acuerdo con Kleinbaum and Klein (2002) implica que los efectos de las variables explicativas en las diferentes categorías de la variable respuesta siguen una relación constante de odds a medida que las variables explicativas cambian.

Se validará el supuesto de odds proporcionales por medio de una prueba de razón de verosimilitud, en donde se plantean las siguientes hipótesis:

H_0 : Los odds proporcionales son iguales entre las categorías de la variable respuesta.

H_a : Los odds proporcionales son diferentes entre al menos dos categorías de la variable respuesta.

8 Resultados

8.1. Análisis exploratorio de los datos

8.1.1. Análisis univariado de la información personal

Tabla 8-1: Análisis de las variables de información personal

Variable	Categoría	Frecuencia	Porcentaje
Género	Femenino	87	70.7 %
	Masculino	36	29.3 %
Ciudad	Cali	63	51.2 %
	Otra Ciudad	60	48.8 %
Estrato	1	32	26 %
	2	61	49.6 %
	3	27	22 %
	4	3	2.4 %
Régimen de la EPS	Subsidiado	77	62.6 %
	Contributivo	46	37.4 %

En la tabla 8-1 se puede observar un resumen de la información personal de los estudiantes, donde se observa que la proporción de mujeres es mucho mayor que la de hombres, además se evidencia que el porcentaje de personas que provienen de lugares distintos a la ciudad de Cali es casi la misma proporción de estudiantes que son de la ciudad. En cuanto a estrato socioeconómico se evidencia que el mayor porcentaje pertenecen a estratos 1 y 2, seguido de estratos 3 y 4 y finalmente, en la mayoría de estudiantes se encuentra que están adscritos al régimen de EPS subsidiado

8.1.2. Análisis univariado de la información académica

Tabla 8-2: Análisis de las variables de información académica

Variable	Categoría	Frecuencia	Porcentaje
Número de semestres que el estudiante recibe el subsidio	1	5	4.1 %
	2	19	15.4 %
	3	6	4.9 %
	4	30	24.4 %
	5	63	51.2 %
Facultad	Artes Integradas	18	14.6 %
	Ciencias de la Administración	7	5.7 %
	Ciencias Naturales y Exactas	7	5.7 %
	Ciencias Sociales y Económicas	3	2.4 %
	Educación y Pedagogía	13	10.6 %
	Humanidades	21	17.1 %
	Ingeniería	26	21.1 %
	Psicología	3	2.4 %
	Salud	25	20.3 %
Semestre matriculado	2	3	2.4 %
	3	32	26 %
	4	28	22.8 %
	5	14	11.4 %
	6	15	12.2 %
	7	10	8.1 %
	8	14	11.4 %
	9	7	5.7 %

En la tabla 8-2 se puede observar un resumen de la información académica de los estudiantes, donde se observa que aproximadamente la mitad de los estudiantes (48.8 %) no reciben la

totalidad de los subsidios. Además, solo el 19.5 % se beneficia del subsidio solamente por un año (dos semestres). Con respecto a la facultad, se evidencia que la mayor proporción de estudiantes pertenecen a ingeniería y humanidades con un 21.1 % y 17.1 % respectivamente. En relación con los semestres matriculados, casi la mitad de los estudiantes reciben el subsidio entre el tercer y el cuarto semestre y hay una cantidad muy baja en el segundo y noveno semestre.

8.2. Análisis bivariado

El primer análisis bivariado se presenta en la Figura 8-1 el cual es con la variable “número de semestres subsidiados” y “género” en el que se identificó que la frecuencia es mayor en los estudiantes que culminaron los 5 semestres subsidiados para ambos géneros, en donde hubo un 51.72 % de mujeres y un 50 % de hombres, al disminuir un semestre se evidencia que la cantidad de estudiantes que tuvieron el subsidio se reduce casi a la mitad para ambos géneros y así progresivamente hasta llegar a 1 semestre subsidiado en donde se encuentra la más baja frecuencia de estudiantes. Además, se evidencia que la proporción para cada género es casi la misma, por lo tanto no se identifica una diferencia entre el número de semestres subsidiados de acuerdo al género.

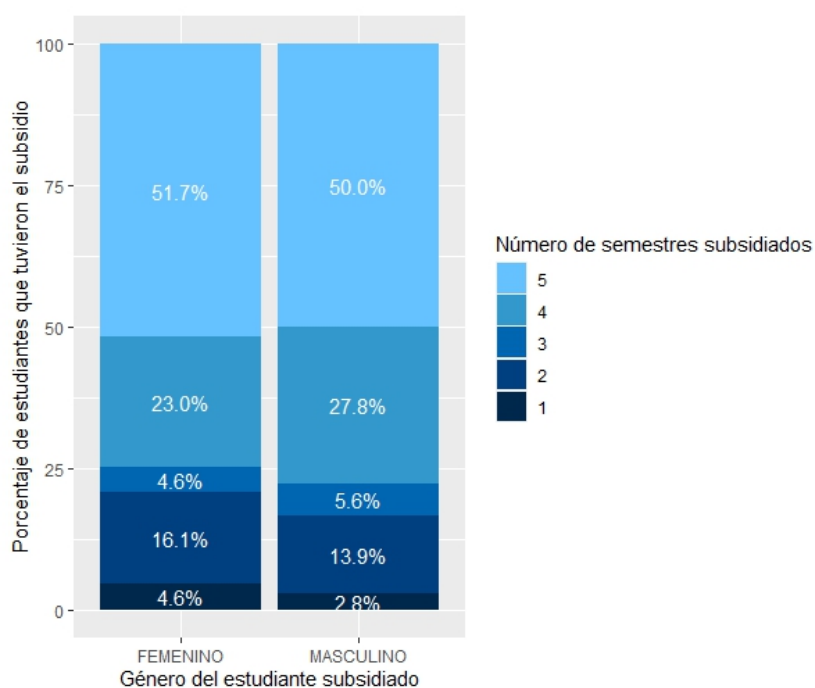


Figura 8-1: Relación entre el número de semestres subsidiados y el género

En cuanto a la ciudad de procedencia de los estudiantes subsidiados se identifica por medio de la Figura 8-2 que sin discriminar el lugar, hay una mayor cantidad de estudiantes que

logran culminar los cinco semestres con el subsidio, se evidenció también que el número de estudiantes que pierden el subsidio después de recibirlo durante un semestre es en menor frecuencia para ambas categorías. Finalmente, se evidencia que el 22.2 % de los estudiantes cuyo lugar de procedencia es de Cali se benefician del subsidio solamente por un año (dos semestres), comparado con los estudiantes de otro lugar distinto a Cali que representa un 16.6 % de los estudiantes.

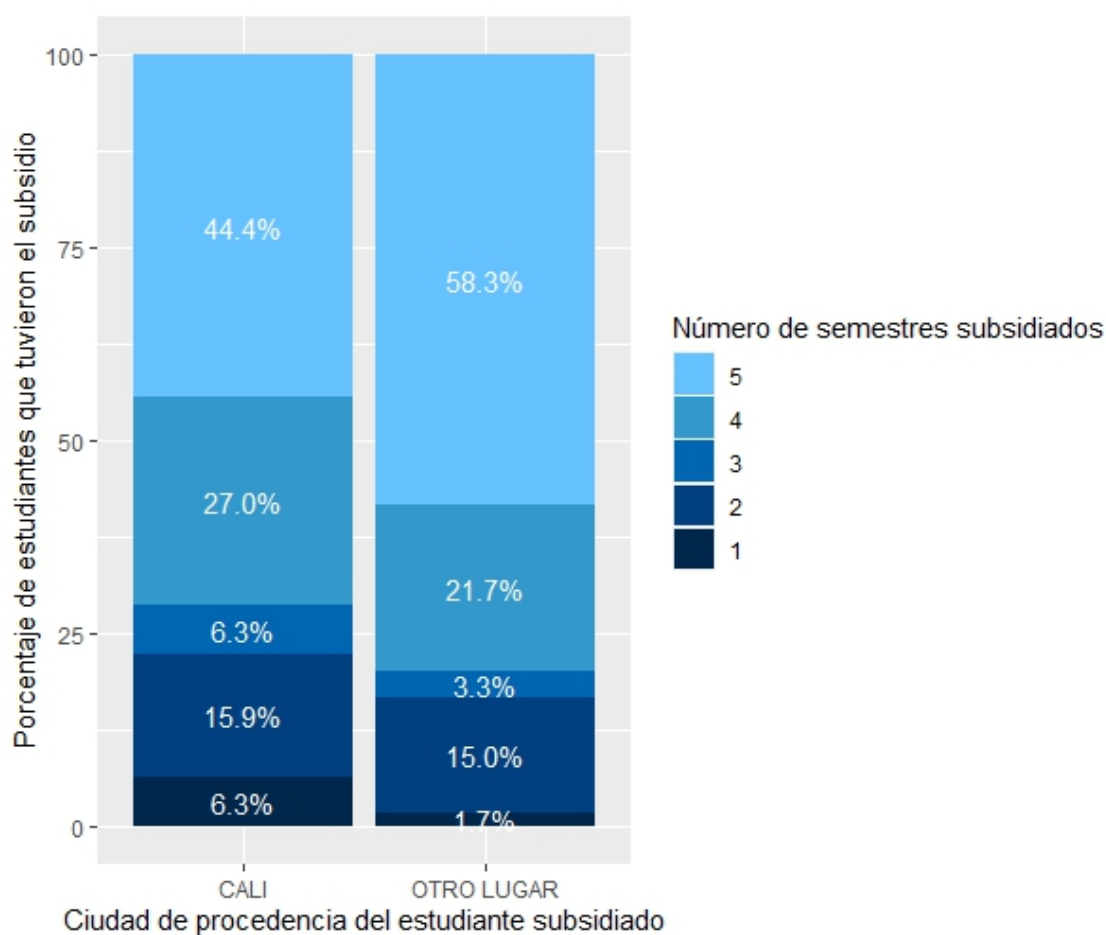


Figura 8-2: Relación entre el número de semestres subsidiados y ciudad de procedencia

De acuerdo con la Figura 8-3 se evidencia que en el estrato 1 es mayor la cantidad de estudiantes que logran culminar los 5 semestres subsidiados (21) en comparación a los que lo pierden antes de tiempo (11). Además, se identifica que en el caso de los estratos 1-3, la frecuencia de estudiantes que pierden el subsidio económico antes de 2 semestres subsidiados es baja, presentando casos entre 1 y 2 estudiantes.

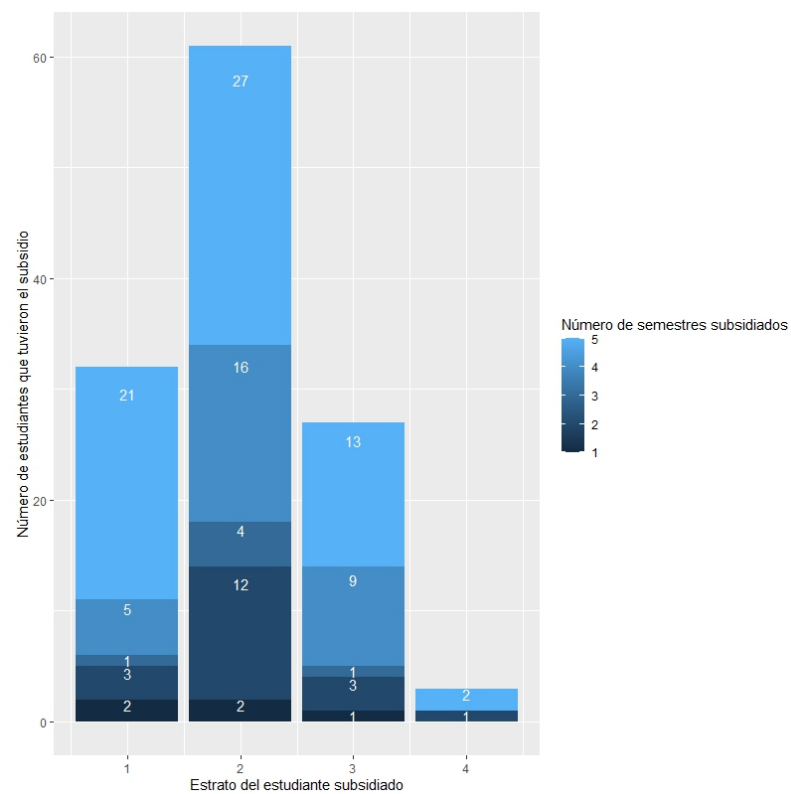


Figura 8-3: Relación entre el número de semestres subsidiados y estrato de residencia

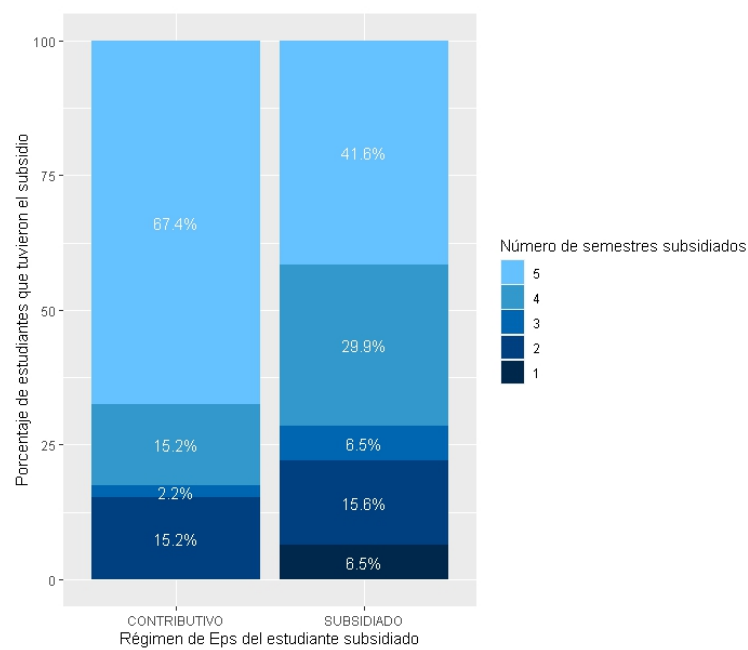


Figura 8-4: Relación entre el número de semestres subsidiados y régimen de EPS

En la Figura 8-4, se observa que la distribución del número de semestres subsidiados varía según el régimen de EPS. Por ejemplo, para los estudiantes que recibieron el subsidio solo durante un semestre, el 100 % pertenecía al régimen subsidiado. Por otro lado, en el régimen contributivo se observa que es mayor la proporción de estudiantes que logran culminar los 5 semestres subsidiados (67.4 %) a comparación de los que no (32.6 %), mientras que en el régimen subsidiado es mayor la proporción de estudiantes que no logran culminar exitosamente los 5 semestres subsidiados (58.4 %) a comparación de los que sí (41.6 %).

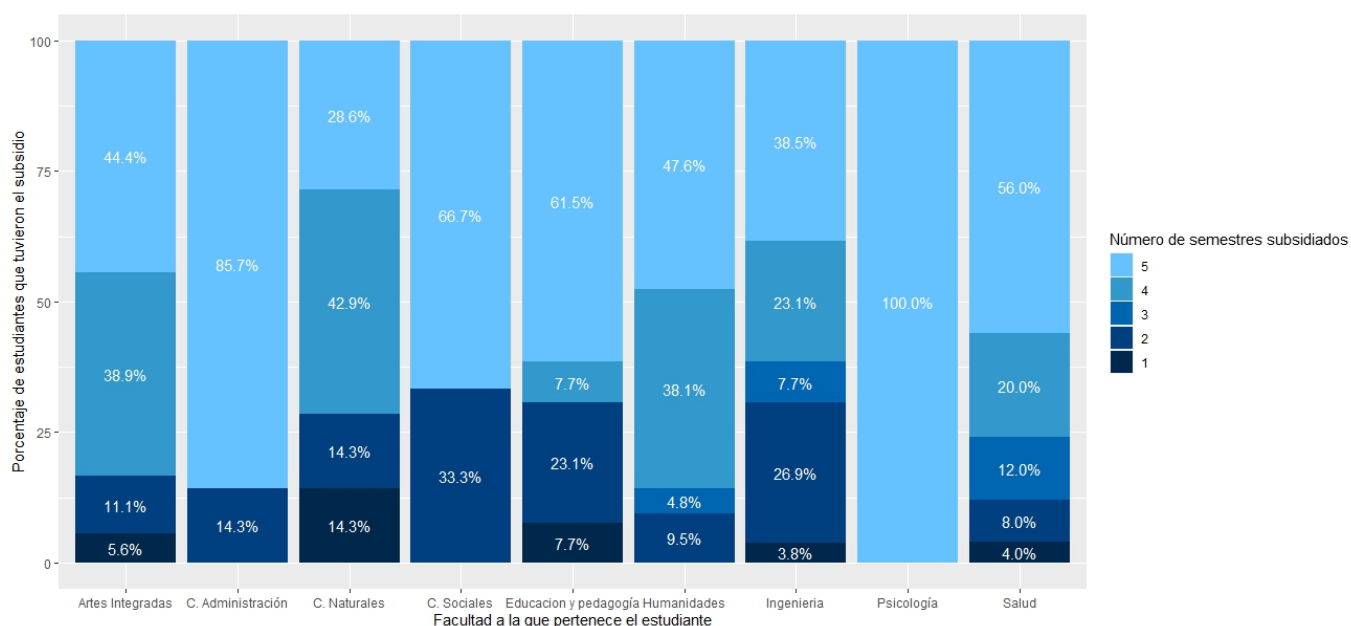


Figura 8-5: Relación entre el número de semestres subsidiados y la Facultad

En la Figura 8-5, se observa la distribución del número de semestres que el estudiante recibe el subsidio respecto a la facultad a la que pertenece, se identifica una variación significativa en la proporción de estudiantes que lograron culminar los cinco semestres del subsidio. En la Facultad de Ingeniería la cual presenta mayor frecuencia de estudiantes, se identifica una mayor proporción de beneficiarios que no logran culminar los 5 semestres subsidiados(61.5 %), al igual que los estudiantes que pertenecen a la Facultad de Humanidades en donde el 47.6 % culminan, comparado al 52.4 % de estudiantes que lo pierden antes de los 5 semestres. En el caso de la Facultad de Salud que es la que presenta la segunda frecuencia más alta, se identifica que pasa lo contrario a Ingeniería y Humanidades, en esta facultad la mayoría de estudiantes logran culminar los 5 semestres subsidiados representando el 56 %.

8.3. Modelo logístico multinomial con efecto aleatorio

Se estima el modelo logístico multinomial con efecto aleatorio asociado a la facultad, tal como se menciona en la metodología para realizar una prueba de razón de verosimilitud y así evaluar si el efecto aleatorio es significativo, es decir, $d > 0$.

Tabla 8-3: Prueba de verosimilitud: Mod completo vs Mod con Efecto Aleatorio

Modelo	LogLike	Valor-p
Con Efecto Aleatorio	-140.12	
Completo	-140.13	0.5

Los resultados de la tabla **8-3** sugieren que no hay diferencia significativa entre el modelo con efectos aleatorios y el modelo sin efectos aleatorios en términos de ajuste a los datos. Dado que se identificó que el efecto aleatorio de facultad no era significativo en el modelo, se procede a no considerarla en el modelo final y por lo tanto se hace uso de un modelo de regresión logística multinomial sin tener en cuenta la variable “Facultad”.

8.4. Selección de variables

En la Tabla **8-4** se observa los distintos modelos generados por el algoritmo Backward, junto con el valor correspondiente de AIC. El modelo 3, que incluye las covariables de “semestre”, “régimen”, “ciudad” y “promedio” es el que presenta el valor más bajo de AIC (303.42) comparado con el valor del modelo completo que es (316.92). Por lo tanto, el modelo seleccionado para este estudio es el modelo 3, que incluye las covariables que considera más relevantes.

Tabla 8-4: Algoritmo de selección: Método Backward

	Modelo	AIC
1	Semestre + Régimen + Estrato + Género + Ciudad + Promedio	316.92
2	Semestre + Régimen + Género + Ciudad + Promedio	305.32
3	Semestre + Régimen + Ciudad + Promedio	303.42

De acuerdo con el algoritmo de selección utilizado, se determina que el modelo más adecuado es el siguiente:

$$\text{Logit}[P(y_i \leq k)] = \alpha_k + \beta_1 \cdot \text{semestre}_i + \beta_2 \cdot \text{régimen}_i + \beta_3 \cdot \text{ciudad}_i + \beta_4 \cdot \text{promedio}_i$$

8.4.1. Prueba de razón de verosimilitud

Considerando el modelo completo y el modelo reducido que no incluye las variables de Estrato y Género, se procede a realizar una prueba de razón de verosimilitud con el objetivo de evaluar si las covariables que se están eliminando del modelo contribuyen de manera significativa a la explicación a través del análisis de varianza.

$$H_0 : \beta_3 = \beta_4 = 0 \text{ vs } H_a : \beta_i \neq 0 \text{ } i = 3, 4$$

Tabla 8-5: Prueba de verosimilitud: Mod completo vs Mod reducido

Modelo	LogLike	Valor-p
Reducido	-140.58	
Completo	-140.13	0.9235

Al comparar el modelo completo con el modelo reducido, se concluye que las variables Estrato y Género carecen de significancia cuando ya se incluyen otras variables en el modelo, como el Semestre, Régimen, Ciudad y Promedio. Esto implica que el modelo reducido ofrece un nivel de ajuste equivalente al del modelo completo. Por lo tanto, se determina que es apropiado utilizar el modelo reducido.

8.5. Supuesto de odds proporcionales

Durante el proceso de ajuste del modelo sin asumir proporcionalidad de odds, se identificaron problemas de convergencia debido a la presencia de separación completa. Este problema se puede deber a que hay categorías con una cantidad muy baja o nula de estudiantes para ciertos números de semestres que reciben el subsidio, por ejemplo, en la covariable régimen no hay suficientes observaciones para cada número de semestres en los que el estudiante recibió el subsidio. Con esto, el modelo puede tener dificultades para estimar los efectos de esas covariables debido a la falta de información suficiente. Teniendo en cuenta lo anterior, se ajusta el modelo bajo la suposición de odds proporcionales.

8.6. Modelo reducido ajustado

Una vez seleccionadas las covariables por medio de los algoritmos de selección y definidas las covariables a utilizar se presenta en la Tabla 8-6 las estimaciones del modelo logístico multinomial.

Tabla 8-6: Resumen del modelo reducido

Coeficientes	Parm	Estimación	Std Error	Z-Value	P-Value	OR
Intercepto 1	α_1	-2.53599	2.09769	-1.209	0.22668	
Intercepto 2	α_2	-0.73136	2.06077	-0.355	0.72267	
Intercepto 3	α_3	-0.41885	2.05901	-0.203	0.83881	
Intercepto 4	α_4	0.89571	2.05935	0.435	0.66360	
Semestre	β_1	0.31469	0.09274	3.393	0.00069	1.3698337
Promedio	β_2	-0.70916	0.48272	-1.469	0.14181	0.4920583
Ciudad	β_3	-0.72777	0.37296	-1.951	0.05101	0.4829839
Régimen	β_4	0.96044	0.39848	2.410	0.01594	2.6128456

En la Tabla **8-6** se pueden interpretar los coeficientes del modelo a partir de los odds ratio (OR), es decir, determinar la probabilidad relativa de que un estudiante pierda el beneficio antes del quinto semestre subsidiado según ciertas características. También se identificaron variables con un valor p menor a 0.05, lo que indica una significancia estadística en relación con la probabilidad de perder el subsidio antes de culminarlo.

Del resultado anterior se observa que, manteniendo las demás covariables constantes:

- Para el semestre como el coeficiente es positivo, la probabilidad de perder el subsidio antes del tiempo acordado aumenta con el semestre matriculado en el que lo adquieren.
- En cuanto al coeficiente de la ciudad de procedencia de los estudiantes, la probabilidad de perder el subsidio antes del tiempo acordado disminuye cuando la ciudad es diferente de Cali.
- Con relación al régimen de Eps al que se encuentra adscrito el estudiante, se evidencia con el coeficiente positivo que la probabilidad de perder el subsidio aumenta en los estudiantes cuyo régimen es subsidiado.
- Para el promedio de los estudiantes se evidencia que, la probabilidad de perder el subsidio antes de culminar los cinco semestres reglamentarios disminuye a medida que el promedio acumulado sea más alto.

8.6.1. Predicción para casos particulares

En el presente estudio, se realizó un modelo de regresión multinomial para determinar los factores asociadas a la pérdida del subsidio económico. Con el fin de evaluar las probabilidades de pertenecer a cada categoría de la variable de respuesta, se realizaron predicciones específicas para casos particulares. En este caso se considera un estudiante que recibe su primera asignación del subsidio estando en el segundo semestre matriculado, con un promedio acumulado de 3.6, su lugar de procedencia es la ciudad de Cali y el régimen al que se encuentra adscrito es contributivo; Para este caso particular se obtuvieron las siguientes probabilidades:

Tabla 8-7: Predicción para un caso específico propuesto

	1	2	3	4	5
1	0.01143494	0.05424951	0.02198418	0.1758269	0.7365045

En la Tabla 8-7 se observa la probabilidad de que el estudiante con las características especificadas anteriormente pierda el subsidio en cada uno de los cinco semestres posibles. De esta manera se observa que, existe una alta probabilidad de que el estudiante complete los cinco períodos con el beneficio.

Finalmente, en la Figura 8-6 está representado graficamente las probabilidades del promedio acumulado y semestre y en la Figura 8-7 las probabilidades del régimen y la ciudad de procedencia del estudiante, estos efectos en la probabilidad de que un estudiante reciba el subsidio en un número determinado de semestres para diferentes valores de las covariables mencionadas.

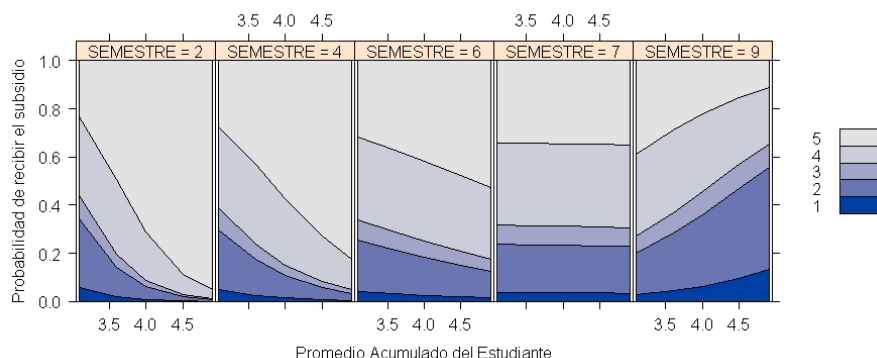


Figura 8-6: Probabilidades de recibir el subsidio dado el promedio acumulado y semestre

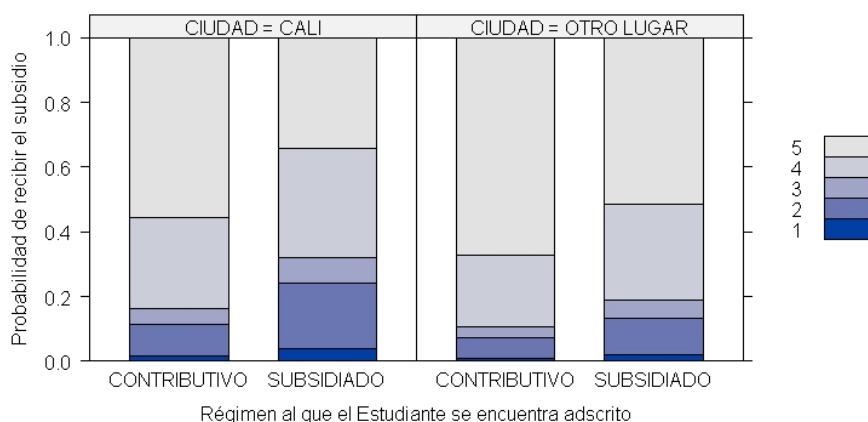


Figura 8-7: Probabilidades de recibir el subsidio dado el régimen y la ciudad de procedencia

De acuerdo a lo anterior, se puede observar que para la situación de recibir el subsidio durante cinco semestres, la mayor probabilidad ocurre en la combinación de promedio acumulado de 4.9 y estar en el segundo semestre matriculado. La probabilidad para esto fue de 0.9523, lo que significa que hay una probabilidad alta de que un estudiante con las características mencionadas reciba el subsidio por 5 semestres. Lo anterior se puede observar en la Figura 8-6 donde los estudiantes que empezaron a recibir el subsidio en el semestre 2 tienen una mayor oportunidad de ser subsidiados más semestres que los que empiezan a recibirlo en semestres posteriores. Por otra parte, se identificó que para la situación de recibir el subsidio durante cinco semestres, la mayor probabilidad ocurre en la combinación de régimen contributivo y lugar de procedencia distinto a Cali con un valor de 0.67. En la Figura 8-7 se puede evidenciar que los estudiantes adscritos al régimen contributivo tienen más probabilidades de recibir el subsidio durante más semestres.

9 Conclusiones y recomendaciones

9.1. Conclusiones

El presente estudio se enfocó en analizar la pérdida del subsidio entre los estudiantes admitidos en el primer semestre del 2019 del programa de subsidio económico de la UV, utilizando un modelo de regresión logística multinomial. La investigación reviste una gran importancia, ya que permite conocer las características claves de los estudiantes con mayor probabilidad de perder el subsidio y así las trabajadoras sociales puedan identificar los estudiantes en riesgo, con el objetivo de brindar un acompañamiento continuo durante los cinco semestres y garantizar la permanencia exitosa de los estudiantes, reduciendo la frecuencia de estudiantes que pierden el beneficio.

En el modelo se identificó que el semestre académico, la ciudad de procedencia y el tipo de régimen al que está adscrito el estudiante son variables que influyen a la hora de que un estudiante pierda el subsidio económico.

Con el modelo ajustado se examinaron las probabilidades de que un estudiante pierda el subsidio económico antes de culminarlo por completo debido a ciertas características, de esta manera, se encontró que los estudiantes que reciben el subsidio en semestres superiores tienen más probabilidades de perderlo antes del tiempo acordado, así como los estudiantes adscritos al régimen subsidiado, en comparación con los estudiantes que pertenecen a una eps con régimen contributivo. Por el contrario, los estudiantes cuyo lugar de procedencia es distinto de Cali tienen menores probabilidades de perder el subsidio en comparación con los estudiantes provenientes de Cali. Finalmente, aunque la covariable de promedio académico, fue considerada en la exploración inicial y en el modelo final, esta variable no demostró una asociación significativa con la pérdida del subsidio económico.

Los resultados mencionados anteriormente permiten identificar al programa de subsidios algunas características de los estudiantes que afectan en la permanencia del programa de subsidio económico, para sugerirles analizar los requisitos actuales para la asignación del beneficio, con el fin de considerar variables relevantes que ayuden a identificar los estudiantes que tienen asignado el subsidio y presentan alto riesgo de perderlo, para así brindar el acompañamiento pertinente y disminuir el riesgo de pérdida.

9.2. Recomendaciones

Se recomienda al programa de subsidios empezar a considerar otras características de los estudiantes, tales como: el nivel educativo y la situación laboral de los padres, el número de integrantes del núcleo familiar y los ingresos familiares. De esta manera, se podría incluir estas nuevas variables en el modelo y determinar si tienen influencia en la probabilidad de que un estudiante pierda el programa de subsidios.

Finalmente, considerando que el modelo sin asumir proporcionalidad de odds presentó problemas de convergencia, se recomienda para estudios posteriores la consideración de un modelo con verosimilitud penalizada. Dicho modelo se puede ver según Correa and Valencia (2011) como un método para introducir cierto grado tolerable de sesgo a cambio de la reducción en la variabilidad de las estimaciones de los parámetros, lo que dará un mejor ajuste a los datos. Lo anterior teniendo en cuenta que el problema de la separación se presenta en la investigación dada su muestra pequeña y los datos dispersos. Con lo expuesto anteriormente, se podría mejorar el ajuste del modelo, otorgando resultados más precisos en la evaluación de las características iniciales relacionados con la pérdida del subsidio económico.

Bibliografía

- Agresti, A. (2012), *Categorical Data Analysis*, Wiley.
- Agresti, A. (2015), *Foundations of Linear and Generalized Linear Models*, John Wiley Sons.
- Correa, J. and Salazar, J. (2016), *Introducción a los modelos mixtos*, Escuela de estadística, Facultad de Ciencias. Universidad Nacional de Colombia.
- Correa, J. and Valencia, M. (2011), *La separación en regresión logística, una solución y aplicación*, Revista Facultad Nacional de Salud Pública’.
- Dobson, A. and Barnett, A. (2018), *An Introduction to Generalized Linear Models - Fourth Edition*, CRC Press Company.
- Escobar, J., Largo, E. and Perez, C. (2006), *Factores asociados a la admisión, deserción y permanencia estudiantil en la universidad del valle (1994 â 2006)*, CIDSE. Universidad del Valle. Colombia.
- Garbanzo, G. (2013), *Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica.*, Revista Electrónica Educare.
- González, A. (2002), *Reflexiones sobre los factores que influyen en la deserción escolar del adolescente*, Universidad Pedagógica Nacional.
- González, A. (2015), *Selección de variables: Una revisión de métodos existentes*, Universidad da Coruña.
- Gonzalez, N. and Martinez, J. (2022), *Factores asociados a la deserción estudiantil para los estudiantes pertenecientes al Programa ASES en el año 2017*, Universidad del Valle.
- Gutierrez, M., Reyes, N., Ortiz, M., Escobar, J. and Robles, S. (2021), ‘Juventudes colombianas 2021. preocupaciones, intereses y creencias’, *Fundación SM Colombia* .
- Guzman, C., Durán, D., Franco, J., Castano, E., Gallon, S., Gómez, K. and Vásquez, J. (2009), *Deserción estudiantil en la educación superior Colombiana*, Ministerio de Educación Nacional.
- Hernández, F. and López, J. (2022), *Modelos Mixtos con R*.

Kleinbaum, D. and Klein, M. (2002), *Logistic regression- A Self-Learning text*, Springer.

López, M. (2021), *Impacto de los programas de Bienestar Universitario en el rendimiento académico y la permanencia de los estudiantes de la Universidad Del Valle.*, Universidad del Valle.

McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models. 2nd Edition*, Chapman and Halls, Journal of Applied Mathematics and Physics.

MEN (2009), *Deserción Estudiantil*, Revolución Educativa’.

MEN (2014), ‘Estrategias para la permanencia en educación superior : Experiencias significativas’, *Ministerio de Educación Nacional*’0.

MEN (2021), ‘Estadísticas de deserción y permanencia en educación superior spadies 3.0 - indicadores 2021’, *Centro de Investigaciones y Documentación Socioeconómica CIDSE* .

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2022), *Introduction to linear regression analysis*, John Wiley Sons.

Ramirez, S. (2019), ‘Herramienta moodle como plataforma tecnológica para la detección temprana del riesgo de abandono en la universidad del valle’, *Universidad del Valle* .

SPADIES (2018), ‘Sistema para la prevención y análisis de la deserción en las instituciones de educación superior: Consultas básicas (2012)’, *Sistema para la Prevención y Análisis de la Deserción en las instituciones de Educación Superior*. .

SPADIES (2022), ‘Glosario’, *Ministerio de educación Nacional*’.

Valencia, A. (2019), *Impacto del Programa Subsidio Económico en la Permanencia de los Estudiantes de Univalle*. Universidad del Valle, Universidad del Valle.

VBU (2022), ‘Subsidios y becas de alimentación- actas.’, *Universidad del Valle* .

URL: <https://vicebienestar.univalle.edu.co/index.php/2015-11-13-19-24-45/2015-11-13-19-26-22/actas>

Velasco, I. (2016), ‘Análisis de las causas de deserción universitaria’, *Universidad Nacional Abierta y a Distancia (UNAD)* .

Velázquez y González, M. A. (2017), ‘Factores asociados a la permanencia de estudiantes universitarios: caso uamm-uat’, *Revista de la Educación Superior* .

Verbeke, G., Fitzmaurice, G., Davidian, M. and Molenberghs, G. (2009), *Longitudinal Data Analysis*, CRC Press Company.

Verbeke, G. and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, CRC Press Company.