



# Factores propios de la educación media y su relación con el desempeño en el primer curso de Estadística en la Universidad

Erika Minotta Gaviria

Universidad del Valle  
Facultad de Ingeniería, Escuela de Estadística  
Santiago de Cali, Colombia  
2019



# **Factores propios de la educación media y su relación con el desempeño en el primer curso de Estadística en la Universidad**

**Erika Minotta Gaviria**

**199825152**

Trabajo de grado presentado como requisito parcial para optar al título de:

**Estadística**

Director:

Ph.D José Rafael Tovar Cuevas

Universidad del Valle

Facultad de Ingeniería, Escuela de Estadística

Santiago de Cali, Colombia

2019

## Dedicatoria

A mis padres y a mi hermana, por su paciencia,  
comprensión, unión y ejemplo, a Diego  
Villaquirán por su disposición, compañía y  
apoyo incondicional.

# Agradecimientos

Agradezco al profesor José Rafael Tovar, asesor y director de este trabajo, por su idea de investigación, su acompañamiento, compartir su experiencia y enseñanzas, por su disciplina y paciencia durante el proceso.

A los profesores del programa de Estadística y de la Facultad de Salud, por los conocimientos impartidos, su calidad humana, además de sus recomendaciones y buenos consejos que han sido un gran apoyo para alcanzar el logro propuesto.

A la Universidad del Valle, institución que me acogió, es parte importante en la realización de mis proyectos de vida y me ha brindado una nueva oportunidad para concluir este proceso de formación.



## Resumen

La estadística se imparte en diferentes niveles de educación y con el fin de establecer las características o variables que afectan el rendimiento académico de los estudiantes en el primer curso de estadística al que se enfrentan en la Universidad, se plantea el uso de técnicas de estadística clásica y bayesiana para analizar datos obtenidos con la aplicación de una encuesta a estudiantes de una universidad privada en Cali en el segundo semestre de 2017. Como metodología clásica se propuso un modelo de regresión con variable de respuesta binaria utilizando la función de enlace Log-Log Complementario, útil cuando la proporción de éxitos es pequeña, para predecir la probabilidad de que un estudiante repruebe o no el curso de estadística e identificar variables relacionadas con la experiencia previa de los estudiantes en el colegio, que se asocian significativamente con el resultado obtenido en el curso. La proporción de estudiantes que reprobaron la asignatura es del 6.75 %. Las variables explicativas que presentan relación significativa con la probabilidad de reprobar el curso fueron el desempeño alcanzado en cursos de matemáticas en la universidad, la experiencia en cursos de estadística en el bachillerato, conceptos de estadística vistos como parte de un curso de matemáticas en el colegio y haber visto al menos cuatro temas de estadística en el bachillerato. El modelo propuesto propuesto discrimina bien a los estudiantes a un punto de corte 0,65 y arrojó un porcentaje correcto de clasificación superior al 50 %. En la aproximación bayesiana se estima la probabilidad de reprobar el curso caracterizando a los estudiantes por la combinación de las variables significativas planteando la utilización de distribuciones apriori informativas y no informativas. Se encontró que la mayor probabilidad de riesgo de que los estudiantes reprueben el curso de estadística está en personas que no tuvieron un buen desempeño en cursos de matemáticas en la universidad, tuvieron experiencia en cursos de estadística en el bachillerato, no vieron conceptos de estadística como parte del curso de matemáticas y reportan haber visto cuatro o más temas de estadística en el colegio.

**Palabras clave:** enseñanza de la estadística, desempeño académico, Prueba Exacta de Fisher, Prueba Chi-Cuadrado, Modelo con respuesta binaria, enlace Log-Log Complementario, Curva ROC, Capacidad clasificatoria del modelo, Estadística Bayesiana.

# Abstract

The statistics is given in different levels of education and in order to establish the characteristics or variables that affect the academic yield of the students in the first course of statistics in the University, there appears the use of skills of classic and bayesian statistics to analyze information obtained with the application of a survey to students of a private university in Cali in the second semester of 2017. As classic methodology proposed a model to itself with binary response using the Complementary Log-Log link, useful when the successes proportion is small, to predict the probability of which a student should reprove or not the statistics course and to identify variables related to the experience previous to the students in the school, who collaborate significantly with the result obtained in the course. The students' proportion that they reprovred the subject is 6.75 %. The explanatory variables that present significant relation with the probability of reprovring the course were the performance reached in courses of mathematics in the university, the experience in courses of statistics in the high school, concepts of statistics seen like part of a course of mathematics in the school and to have seen at least four statistics topics in the high school. The proposed model discriminates well against the students to a court point 0,65 and a correct percentage of classification superior to 50 %. In the bayesian approach there is estimated the probability of reprovring the course characterizing the students for the combination of the significant variables raising the distribution use apriori informative and not informative. One found that the biggest probability of risk that the students reprove the statistics course is in persons who did not have a good performance in mathematics courses in the university, they had experience in statistics courses in the high school, they did not see concepts of statistics like part of the course of mathematics and they bring to have seen four or more statistics topics in the school.

**Keywords:** statistics education, academic performance, Fisher's Exact Test, Chi-Squared Test, Bayesian Statistics, Models for binary data, Complementary Log-Log link Model, ROC Curve, Classification Capacity of the model.



# Contenido

|                                                       |             |
|-------------------------------------------------------|-------------|
| <b>Resumen</b>                                        | <b>VII</b>  |
| <b>Lista de Figuras</b>                               | <b>XIII</b> |
| <b>Lista de Tablas</b>                                | <b>1</b>    |
| <b>1. Introducción</b>                                | <b>3</b>    |
| <b>2. Planteamiento del problema</b>                  | <b>5</b>    |
| 2.1. Problema de investigación contextual . . . . .   | 7           |
| 2.2. Problema estadístico . . . . .                   | 7           |
| <b>3. Justificación del Proyecto</b>                  | <b>8</b>    |
| <b>4. Objetivos</b>                                   | <b>9</b>    |
| 4.1. Objetivo General . . . . .                       | 9           |
| 4.2. Objetivos Específicos . . . . .                  | 9           |
| <b>5. Revisión de antecedentes</b>                    | <b>10</b>   |
| <b>6. Marco Teórico</b>                               | <b>13</b>   |
| 6.1. Marco Teórico Contextual . . . . .               | 13          |
| 6.2. Marco Teórico Estadístico . . . . .              | 17          |
| 6.2.1. Prueba exacta de Fisher . . . . .              | 18          |
| 6.2.2. Prueba Chi-Cuadrado de Independencia . . . . . | 19          |
| 6.2.3. Modelo Lineal Generalizado . . . . .           | 20          |

|                                                                                                                   |    |
|-------------------------------------------------------------------------------------------------------------------|----|
| 6.2.4. Modelo de Regresión Binaria . . . . .                                                                      | 21 |
| 6.2.5. Enlace log-log complementario . . . . .                                                                    | 23 |
| 6.2.6. Método Newton-Rapshon para el Ajuste del MLG . . . . .                                                     | 24 |
| 6.2.7. Interpretación de parámetros del Modelo con Respuesta Binaria<br>(Enlace Log-Log Complementario) . . . . . | 26 |
| 6.2.8. Prueba de Bondad de ajuste del Modelo . . . . .                                                            | 27 |
| 6.2.9. Criterio de información de Akaike (AIC) . . . . .                                                          | 28 |
| 6.2.10. Prueba de significancia de los parámetros . . . . .                                                       | 29 |
| 6.2.11. Curva ROC . . . . .                                                                                       | 29 |
| 6.2.12. Índice de Youden . . . . .                                                                                | 32 |
| 6.2.13. Capacidad Clasificatoria del Modelo con Respuesta Binaria (Enlace<br>Log-Log Complementario) . . . . .    | 32 |
| 6.2.14. Enfoque Bayesiano . . . . .                                                                               | 34 |
| 6.2.15. Distribución Binomial . . . . .                                                                           | 35 |
| 6.2.16. Distribución Beta . . . . .                                                                               | 36 |
| 6.2.17. Teorema de Bayes . . . . .                                                                                | 36 |
| 6.2.18. Distribución Apriori . . . . .                                                                            | 38 |
| 6.2.19. Distribución Posterior . . . . .                                                                          | 42 |
| 6.2.20. Estimación Puntual . . . . .                                                                              | 44 |
| 6.2.21. Estimación por regiones . . . . .                                                                         | 44 |

## **7. Metodología 45**

|                                                      |    |
|------------------------------------------------------|----|
| 7.1. Fuentes de Información . . . . .                | 45 |
| 7.1.1. Participantes . . . . .                       | 45 |
| 7.1.2. Muestra . . . . .                             | 46 |
| 7.1.3. Instrumento de medición . . . . .             | 46 |
| 7.1.4. Consolidación del archivo de datos . . . . .  | 47 |
| 7.2. Análisis Exploratorio de los Datos . . . . .    | 48 |
| 7.3. Obtención de la variable de respuesta . . . . . | 49 |

|                                                                                                                           |            |
|---------------------------------------------------------------------------------------------------------------------------|------------|
| 7.4. Obtención de las variables explicativas . . . . .                                                                    | 50         |
| 7.4.1. Construcción de Índice de Ansiedad . . . . .                                                                       | 52         |
| 7.4.2. Pruebas de Asociación . . . . .                                                                                    | 55         |
| 7.5. Propuesta de Modelo . . . . .                                                                                        | 55         |
| 7.5.1. Experimento Estadístico . . . . .                                                                                  | 56         |
| 7.5.2. Modelo Estadístico . . . . .                                                                                       | 56         |
| 7.5.3. Evaluación del modelo . . . . .                                                                                    | 58         |
| 7.6. Aproximación Bayesiana . . . . .                                                                                     | 60         |
| 7.6.1. Modelo Estadístico . . . . .                                                                                       | 60         |
| 7.7. Comparación por áreas de conocimiento . . . . .                                                                      | 64         |
| <b>8. Resultados</b>                                                                                                      | <b>65</b>  |
| 8.1. Análisis exploratorio de los datos . . . . .                                                                         | 65         |
| 8.2. Obtención de variables explicativas . . . . .                                                                        | 68         |
| 8.3. Ajuste del Modelo con Respuesta Binaria . . . . .                                                                    | 69         |
| 8.3.1. Capacidad clasificatoria del modelo de respuesta binaria con función<br>de enlace Log-Log Complementario . . . . . | 72         |
| 8.4. Aproximación Bayesiana . . . . .                                                                                     | 75         |
| <b>9. Conclusiones y recomendaciones</b>                                                                                  | <b>81</b>  |
| <b>A. Anexo: Encuesta aplicada</b>                                                                                        | <b>84</b>  |
| <b>B. Anexo: Diccionario de datos</b>                                                                                     | <b>89</b>  |
| <b>C. Anexo: Pruebas de asociación</b>                                                                                    | <b>94</b>  |
| <b>D. Anexo: Análisis Exploratorio de datos</b>                                                                           | <b>97</b>  |
| <b>Bibliografía</b>                                                                                                       | <b>100</b> |

# Lista de Figuras

|                                                           |    |
|-----------------------------------------------------------|----|
| 8-1. Distribución de estudiantes por Facultades . . . . . | 66 |
| 8-2. Distribución según resultado en el curso . . . . .   | 68 |
| 8-3. Curva ROC . . . . .                                  | 73 |



# Lista de Tablas

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| 6-1. Funciones de enlace más utilizadas (Madsen & Thyregod, 2010) . . . . . | 21 |
| 6-2. Tabla de Clasificación . . . . .                                       | 33 |
| 7-1. Estudiantes matriculados por curso de Estadística. . . . .             | 46 |
| 7-2. Escala de la pregunta 15 según combinaciones de respuesta . . . . .    | 53 |
| 7-2. Escala de la pregunta 15 según combinaciones de respuesta . . . . .    | 54 |
| 7-3. Conformación de grupos. . . . .                                        | 61 |
| 7-4. Estimadores . . . . .                                                  | 63 |
| 8-1. Estimaciones Modelo 1 Log-Log Complementario . . . . .                 | 69 |
| 8-2. Estimaciones Modelo 2 Log-Log Complementario . . . . .                 | 70 |
| 8-3. Estadísticas Descriptivas para probabilidades . . . . .                | 71 |
| 8-4. Clasificación . . . . .                                                | 74 |
| 8-5. Elicitación de hiperparámetros . . . . .                               | 76 |
| 8-6. Estimación clásica y Bayesiana . . . . .                               | 76 |
| 8-7. Probabilidades Predictivas Posteriores . . . . .                       | 79 |



# Declaración

Me permito afirmar que he 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 he reconocido. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de Tesis o Trabajo de Grado.

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

Santiago de Cali, 06.12.2018

---

(Erika Minotta Gaviria)



# 1. Introducción

En la actualidad, la estadística ha cobrado bastante importancia (Behar Gutiérrez & Grima Cintas 2001) y temas de esta área se imparten en los diferentes niveles de educación, primaria, secundaria y universitaria, pues las necesidades del mundo actual exigen que las personas tengan una formación básica en estadística.

El proceso de enseñanza – aprendizaje de la estadística, se ha visto afectado por diversos factores, entre los que se puede mencionar el enfoque y forma de enseñanza que ha hecho que las personas en general asocien la estadística a cursos con fuertes componentes matemáticos. Esta situación repercute en el progreso de los cursos porque algunos estudiantes llegan a éstos con preconcepciones que luego se convierten en un obstáculo para el aprendizaje; y en el caso de los profesores se dificulta al abordaje de los estudiantes que tienen diferentes características, diferencias en el nivel de experiencia o acercamiento a la Estadística previamente o con bajo interés o actitudes negativas hacia el curso. Al final de los cursos, como indica León Salazar (2007), los estudiantes no reconocen la verdadera utilidad de los conceptos adquiridos y la forma de aplicarlos en su ejercicio profesional.

Esta preocupación ha abierto líneas de estudio sobre los problemas de enseñanza y aprendizaje de la estadística, una de ellas, para identificar factores que inciden en los resultados académicos de los estudiantes en estos cursos. Este trabajo aborda, para el caso de estudiantes que no están matriculados en carreras de estadística o matemáticas pero que deben cursar al menos un curso como parte de los currículos de sus planes de estudio, el conocimiento y análisis de las relaciones que puedan existir entre la experiencia

previa con cursos de estadística que tienen los estudiantes y su rendimiento académico en el primer curso del área que ven en su pregrado, en una universidad privada de la ciudad de Cali.

Se espera que los resultados obtenidos a partir del desarrollo de este proyecto constituyan un aporte significativo a los procesos de evaluación y revisión de currículos universitarios, considerando que la reflexión sobre los elementos proporcionados por el análisis del rendimiento académico puede contribuir a la mejora de la calidad de la educación. Además, estos resultados pueden ser insumo para otras investigaciones cuya población objeto de estudio sea la misma o con características similares, y que tengan como finalidad la construcción de indicadores de ansiedad frente a los cursos de Estadística.

## 2. Planteamiento del problema

En la educación superior, un gran número de programas de pregrado de diferentes áreas como ciencias naturales, ingenierías, ciencias administrativas, ciencias económicas, ciencias sociales, jurídicas, salud y humanidades, incluyen en sus currículos al menos un curso introductorio de estadística, porque contribuyen a la formación integral y le proporcionan al profesional, herramientas para desenvolverse en entornos en los que en alguna situación requieren la toma de decisiones basadas en indicadores obtenidos a partir de datos numéricos. Como menciona León Salazar (2007), la importancia de la estadística es tal, que en Colombia, al igual que en muchos otros países, forma parte del currículo de la educación secundaria e incluso se incluyen algunos temas en la educación primaria. Las dinámicas actuales, la globalización y los desarrollos tecnológicos exigen un mayor conocimiento en el área.

Sin embargo, muchos estudiantes asocian los cursos de estadística con cursos de desarrollos matemáticos y al finalizarlos no han comprendido su importancia ni reconocen cómo utilizar realmente los conceptos adquiridos para aplicarlos en su ejercicio profesional (León Salazar 2007). Llegan a estos cursos con preconcepciones y actitudes negativas hacia los mismos, que a su vez se asocian con altos niveles de ansiedad al enfrentar el desarrollo del curso. Esta percepción puede obedecer a las experiencias personales del estudiante, ya sea el escaso contacto con el área, la forma en que se les enseñó Estadística (León Salazar 2007) o incluso los comentarios o información tomada de su entorno.

Investigadores de temas relacionados con la enseñanza de la estadística han planteado que,

por ejemplo, en las carreras de áreas de ciencias sociales, los estudiantes muestran poco interés o una actitud negativa hacia la asignatura y esto se traduce en un obstáculo para el aprendizaje como concluye Gil Flores (1999). La actitud que los estudiantes asumen frente a estos cursos puede ser un elemento favorecedor del aprendizaje o, por el contrario, convertirse en una dificultad para finalizar exitosamente el curso. Por tanto, en el proceso de enseñanza - aprendizaje, los profesores encargados de la docencia en esta área ven con preocupación que la predisposición de los alumnos afecte el desarrollo de los cursos.

Por lo mencionado anteriormente y con base en la literatura especializada, los docentes sospechan que las formas de enseñanza tradicionales no están dando los resultados deseados y se ha incrementado su preocupación por mejorar la eficacia de sus tareas docentes. Se ha supuesto que la experiencia formativa con asignaturas de matemáticas y de estadística considerando el número de asignaturas cursadas y las notas obtenidas, es el principal determinante de las actitudes de los alumnos, esperando que a mayor experiencia previa o mayor aprovechamiento de la misma habrá actitudes más positivas y menor ansiedad por parte de los estudiantes. En este sentido, estudios realizados han mostrado correlaciones positivas entre las actitudes hacia la estadística y variables como el nivel de conocimientos previos en el área, habilidades matemáticas básicas o número de cursos con contenido estadístico desarrollados previamente (Roberts & Saxe 1982), que existe correlación entre las actitudes y ansiedad de los estudiantes hacia la estadística y el rendimiento en la materia, pero con una capacidad limitada para la predicción de este último (Márquez 2004). Al parecer, la formación previa en estadística y matemáticas es un antecedente consistente de las actitudes y la ansiedad hacia la estadística, pero esta relación tiende a ser más débil que la encontrada entre las actitudes o la ansiedad con indicadores de la percepción que tienen los estudiantes acerca de sus capacidades en estas materias, según indica Márquez (2004), por lo que ha propuesto que se tenga en cuenta medidas específicas del aspecto del autoconcepto en investigaciones relacionadas con el tema.

Son muchas las variables que afectan el aprendizaje y el rendimiento académico de los

estudiantes, por lo que debe ser de interés conocer cuáles son los factores determinantes y en qué medida se puede incidir para hacer más eficaz el proceso de enseñanza - aprendizaje. Si bien, uno de estos factores es el conocimiento previo, es decir las aptitudes previamente adquiridas por el alumno, no se conoce con total certeza si realmente el conocimiento previo constituye un indicador con gran capacidad predictiva sobre el rendimiento académico en el primer curso de Estadística al que se enfrenta el estudiante en la universidad, dentro del currículo de su programa.

Por tanto, surge la necesidad de resolver el interrogante ¿De qué manera, las características del estudiante respecto a la experiencia adquirida en el colegio, se relacionan con su rendimiento académico en el primer curso de Estadística en la Universidad?

## 2.1. Problema de investigación contextual

Entender cómo se relacionan los factores previos como *experiencia en cursos de matemática y estadística en el nivel de educación media, tipo de colegio de procedencia, contenidos de estadística vistos, momento en el que fueron vistos y percepción*, con el desempeño en el primer curso de Estadística al que se enfrenta un estudiante en su carrera universitaria.

Se desea realizar el análisis para estudiantes de una universidad privada de la ciudad de Cali, matriculados en programas de pregrado de diferentes áreas de conocimiento, que incorporan cursos de Estadística en sus currículos.

## 2.2. Problema estadístico

Establecer un modelo que permita identificar la existencia de posibles asociaciones entre las características del estudiante en relación con la experiencia adquirida en la educación media, previo a los cursos de Estadística en la Universidad. Se consideran variables que afectan directamente el resultado académico en el curso y aquellas que lo evidencian, haciendo referencia específicamente a la experiencia en el colegio (por auto-reporte).

### 3. Justificación del Proyecto

Con este trabajo se espera profundizar y conocer con mayor precisión la problemática que se presenta en el proceso de enseñanza y aprendizaje de la estadística, especialmente con estudiantes de carreras no relacionadas con áreas de Matemáticas y Estadística.

Recientemente a nivel local, no se han realizado estudios que conduzcan a valorar el efecto o influencia de la aproximación a la estadística en la educación media sobre el rendimiento académico en un curso introductorio de Estadística en la educación superior. Abordar esta temática, identificando las posibles asociaciones existentes entre los aspectos mencionados, permite obtener información significativa que contribuya al análisis, por parte de los encargados de impartir docencia en el área de Estadística, respecto a las metodologías empleadas, el abordaje a los estudiantes que presentan diferencias en el nivel de experiencia previo y sus percepciones, entre otros aspectos, para determinar la necesidad o no, de reestructurar los programas de los cursos y el planteamiento de metas u objetivos de la formación básica en estadística. De igual forma, los hallazgos contribuirán al reforzamiento de la importancia de una adecuada enseñanza de Estadística en la educación media, lo que puede ser tenido en cuenta por los organismos administrativos del sistema educativo que trabajan para garantizar una educación con calidad, innovación y pertinencia.

Los resultados, además, podrán servir de insumo para otras investigaciones cuya población objeto de estudio sea la misma o con características similares, y que tengan como finalidad ahondar en la construcción de indicadores de ansiedad frente a los cursos de Estadística.

## **4. Objetivos**

### **4.1. Objetivo General**

Estudiar la relación entre la experiencia que se tiene con tópicos de estadística durante la educación media y el desempeño en el primer curso de estadística en la Universidad.

### **4.2. Objetivos Específicos**

- Proponer un modelo que permita establecer la relación entre las variables asociadas con la experiencia con estadística en el colegio y el desempeño académico de los estudiantes en el primer curso del área en la universidad.
- Estimar la probabilidad de reprobación de un curso de estadística en relación con la experiencia previa con Estadística.
- Comparar por áreas de conocimiento en la universidad, las relaciones encontradas de los factores propios de la educación media.

## 5. Revisión de antecedentes

Temas relacionados con los factores que influyen en el desempeño académico de los estudiantes de educación superior en cursos de Estadística han sido abordados por varios investigadores en estudios nacionales e internacionales. Se presentan en este punto, algunos estudios realizados entre los años 2007 y 2016 que sirven de apoyo al desarrollo del presente estudio, por investigar temas similares utilizando diferentes técnicas estadísticas para dar solución al problema planteado en sus investigaciones.

**Tovar Cuevas et al. (2007)** realizaron un estudio de las preconcepciones de estudiantes de la Pontificia Universidad Javeriana Cali sobre el curso de estadística, que tuvo como propósito indagar sobre la experiencia de los estudiantes en cursos de estadística y matemáticas, sus ideas preconcebidas respecto a los cursos de estadística y su conocimiento previo. Se aplicó un cuestionario de 16 puntos que fue respondido por 239 estudiantes de diferentes escuelas. Se hizo un análisis descriptivo de la muestra de estudiantes para su caracterización con base en la información recogida; se utilizó el método fenomenográfico para el análisis de respuestas abiertas, agrupando las respuestas de los estudiantes en unidades y luego en categorías con las que se obtuvieron distribuciones de frecuencia por facultad, así los resultados eran comparables entre agrupaciones de carreras. Los resultados mostraron que los estudiantes están de acuerdo en la importancia de realizar al menos un curso de estadística en su entrenamiento personal; y que estos cursos deben pensarse y planearse para que se promueva un ambiente dominado por el pensamiento estadístico, en lugar de ser considerados como cursos de cálculo o de matemáticas en los que los estudiantes hacen cálculos sin tener ideas claras sobre el problema que da lugar a éstos.



**Bologna & Vaiman (2013)** analizaron la relación entre las actitudes hacia la Estadística, exposición previa a contenidos estadísticos y los resultados en la asignatura, de la carrera de Psicología. La población estuvo formada por 721 alumnos de primer año de la carrera de Psicología, cohorte 2011, Universidad Nacional de Córdoba, Argentina. Como instrumentos se utilizaron un cuestionario para indagar sobre la existencia de cursos de Estadística en la escuela que debía responderse al momento del primer acceso al aula virtual, durante las dos primeras semanas de clase y la Escala de Actitud hacia la Estadística, calificaciones en el primer parcial de Psicoestadística. Para la escala de actitudes, se realizó un análisis exploratorio para analizar los valores perdidos y atípicos y correlaciones ítem – total para cada ítem. Se hizo un análisis factorial con extracción de factores por componentes principales, distribución de frecuencias para describir la exposición previa a Estadísticas y medidas de tendencia central para los resultados del parcial.

Se hallaron tres componentes para la **actitud: agrado, utilidad y afinidad**, este último es afectivo que resulta de una mezcla de los componentes confianza y ansiedad. Se aplicaron correlaciones bivariadas entre las componentes de la actitud, la exposición previa a contenidos de Estadística y los resultados del primer parcial para medir asociación. Se ajustaron modelos de regresión y se encontró que el componente afectivo es el que con más intensidad se asocia al resultado del parcial, en mayor medida que la exposición a Estadística en la escuela media. Se concluye que la sola exposición a Estadística en la escuela no es suficiente para incidir sobre el resultado del parcial, pero sí para modificar positivamente la afinidad con la estadística y, de esta forma, aportar a mejores niveles de logro.

**Osorno (2015)** investigó sobre los preconceptos de los estudiantes sobre el primer curso de Estadística en la universidad. El propósito planteado fue explorar aspectos personales y de la formación básica relacionados con la actitud y los imaginarios que tienen los estudiantes de carreras diferentes a Estadística frente al primer curso de estadística. Se aplicó un

cuestionario a 816 estudiantes de tres universidades privadas y una pública en Cali. Con el análisis descriptivo univariado se abordaron aspectos de índole personal de los estudiantes a los que se les aplicó el cuestionario, los relacionados con la formación básica de estadística en el colegio. Con el análisis bivariado se presentaron resultados sobre la cantidad, temas vistos y la dificultad por tipos de colegio y también se hizo análisis descriptivo y contrastes de hipótesis para las variables asociadas con el contexto universitario. Se propuso un índice de ansiedad que posteriormente fue dicotomizado para aplicar pruebas de asociación con otras variables de interés. Se identificó que los estudiantes que vieron estadística en el colegio presentan un índice más alto de ansiedad frente al curso universitario de estadística, también los temas de estadística en los que hay relación entre el grado de dificultad y el índice de ansiedad y los más recordados por los estudiantes.

**Vilà Baños & Rubio Hurtado (2016)** midieron las actitudes de los estudiantes de Pedagogía de la Universidad de Barcelona hacia la Estadística, a través de la Escala de Actitudes hacia la Estadística (EAE) de Auzmendi (1992). Se realizó un análisis factorial y de consistencia interna del instrumento, un análisis descriptivo de los datos y pruebas de contraste para observar diferencias según variables relevantes en la teoría (conocimiento previo de Estadística); y un análisis de clusters para identificar tipologías de sujetos que reúnen una serie de características comunes y presentan un perfil diferenciador de actitud hacia la Estadística. Se evidenciaron niveles neutros-bajos de actitud general hacia la materia, pero también se identificaron grupos de estudiantes con perfiles diferenciados: un grupo con una actitud desfavorable, no necesariamente ansioso, con apenas conocimientos previos de Estadística y muy pocas habilidades numéricas; otro grupo caracterizado por una actitud positiva, al que le gusta la Estadística, pero que muestra preocupación y ansiedad, posee mayor índice de conocimientos previos y habilidades numéricas; y un tercer grupo que destaca sobre todo por la ansiedad ante la Estadística, con menos conocimientos previos y menos habilidades de cálculo. Dada la potencial relación entre actitud y logro académico, las conclusiones apuntan hacia la necesidad de explorar estrategias de enseñanza adaptativa para los diferentes grupos.

## **6. Marco Teórico**

### **6.1. Marco Teórico Contextual**

En esta sección, se presentan conceptos teóricos que serán de utilidad en el desarrollo del presente trabajo. En la primera parte se exponen términos y conceptos relacionados con el proceso de enseñanza de la estadística para contextualizar el problema de estudio. En la segunda parte se describen las técnicas estadísticas utilizadas en las diferentes etapas del análisis de los datos, para ejecutar la fase exploratoria, el ajuste de modelos de regresión y la comparación por áreas de conocimiento, para dar solución al problema planteado.

Batanero (2000*b*) afirma que la comprensión de un concepto no puede reducirse a conocer las definiciones y propiedades, sino a reconocer los problemas donde debe emplearse el concepto, notaciones y palabras con que se denota y en general todas sus representaciones, habilidad operatoria en los diferentes algoritmos y procedimientos relacionados con el concepto y la capacidad de argumentar y justificar propiedades, relaciones y soluciones de problemas.

Batanero (2004) piensa que la problemática educativa radica en que la incorporación de la estadística desde la escuela todavía no es un hecho. Aunque los currículos de educación primaria y secundaria la incluyen, los profesores suelen dejar el tema para el final del programa y frecuentemente lo omiten. Los alumnos llegan a la universidad sin los conocimientos básicos y es preciso comenzar el programa repitiendo los contenidos de estadística descriptiva y cálculo de probabilidades que debieron haber visto en la escuela.

Los profesores universitarios, que deben tratar de llegar a la inferencia estadística, que es una parte que les resultará muy útil a los estudiantes, deben acelerar las explicaciones, suprimir actividades prácticas y gran parte de demostraciones o razonamientos que podrían llevar al alumno a comprender mejor la metodología estadística. El alumno no puede asimilar el contenido en un tiempo tan limitado y solo consigue un aprendizaje memorístico que no podrán aplicar en su vida profesional. Estos problemas se agravan por la masificación de los cursos y la falta de recursos como laboratorios de informática o profesores ayudantes, que faciliten una enseñanza más aplicada a la estadística. No es extraño que los alumnos estén desmotivados y la estadística termine siendo una de las asignaturas menos populares para los estudiantes.

Batanero menciona además, que los alumnos no sólo aprenden estadística en las horas lectivas ni en los centros de enseñanza. No solo se encuentra información estadística en la prensa y medios de comunicación o en textos de otras asignaturas, sino que la internet ha empezado a modificar las relaciones docentes ya sea con participación voluntaria o no de los profesores. Los profesores en los diferentes niveles educativos han de aceptar la rapidez del cambio tecnológico e involucrarse si quieren guiar la educación estadística y crear una verdadera cultura estadística en la sociedad.

Behar et al. (2011) exponen que el objetivo de la educación se puede mirar en términos de un cambio conductual. En un curso de estadística se piensa en cambiar las actitudes y deseos, las habilidades y los conocimientos, en el sentido de propiciar en el estudiante un crecimiento que le permita aprovechar los principios y las herramientas estadísticas en la solución de los problemas que se le presentarán en su vida profesional futura.

En el proceso de aprendizaje y crecimiento, de acuerdo con varios descriptores, se deben lograr cambios en las áreas de conocimiento (lógica), habilidades (física) y actitudes y deseos (emocional). En algunos casos se desarrolla la parte emocional en el sentido que el profesor transmite a los estudiantes el aprecio y compromiso que tienen con lo que están enseñando,

a su vez que en el proceso el estudiante va adquiriendo habilidades y destrezas basados en el conocimiento adquirido, pero si no hay lugar a la práctica, a que el estudiante confronte sus conocimientos y habilidades con problemas reales, no se completa el ciclo de aprendizaje.

En lo que respecta a la motivación, Behar y Ojeda (1995) han planteado que el motivo para estudiar una materia en particular puede ser de diferente naturaleza y ocupar diferentes escalas o grados con relación al aprendizaje. Una razón para estudiar estadística puede deberse a que aparece en el plan de estudios y es indispensable aprobar esos créditos. Lo que puede ser el caso de estudiantes de ingeniería, biología, administración, economía, entre otros. En esta situación, estudiar no tiene como propósito aprender, si no aprobar. El estudiante se inclina por ejercitarse para aprobar y no para resolver sus propias inquietudes, no le preocupa incorporar para sí mismo conceptos que le permitan abordar con éxito situaciones en su vida profesional. En este caso el motivo no propicia la actitud de aprender estadística. Incluso, puede haber predisposición a no poner su mejor esfuerzo en las tareas y actividades que el profesor plantea por una animadversión a las “matemáticas” o a “los números”. Y si deben cursar mas de un curso de estadística, esto puede resultar en una predisposición negativa. Una motivación de alto nivel se da cuando se reconoce el valor de la utilidad e importancia de lo que se estudia, cuando sabe que lo que va a aprender le servirá en su vida profesional, entonces hay una actitud distinta hacia el aprendizaje. Poco frecuente en estudiantes de otras carreras frente a la materia de estadística. Por lo que se ha recomendado ilustrar a los estudiantes sobre cómo la estadística puede ser una herramienta tan valiosa como los propios conocimientos de la disciplina que estudian, que sepan por qué les conviene aprender y para qué les va a servir concretamente en su vida profesional. Y de manera general, con respecto a los contenidos, una recomendación es realizar una caracterización general de las necesidades de cada profesión para que los cursos de servicio se den considerando los conceptos claves.

La problemática de la educación estadística para no estadísticos es compleja, pero se puede ver de otra forma si se la conceptualiza desde la perspectiva de la dificultad en el aprendizaje

de los conceptos, principios y procedimientos de la metodología estadística. El aspecto primario es la falta de una actitud positiva y motivación adecuada para el aprendizaje. El razonamiento estadístico no es probabilidad y matemática, ni tampoco importan mucho los procedimientos aritméticos y de cálculo. Es más importante, contextualizar la metodología estadística en el proceso e investigación particularizado a la disciplina para la cual se enseña. Además, hablar un lenguaje en los términos de los problemas a los que se enfrentará el estudiante cuando sea un profesional en la disciplina de su interés, permitirá que valore en su justa medida la estadística por su importancia y utilidad como una importante herramienta. A quienes enseñan, recomienda considerar los enfoques del aprendizaje participativo, entre los cuales se destaca la enseñanza basada en proyectos estudiantiles, además de actividades de aprendizaje, seminarios estudiantiles, ejercicios grupales con datos reales, conferencias, lecciones dictadas por el profesor, sin olvidar la importancia de incluir la computadora y los paquetes estadísticos de manera responsable (Behar et al. (2011). Si los aspectos mencionados, incorporados a los cursos de estadística, propician el pensamiento estadístico (Behar Gutiérrez & Grima Cintas 2011), resultando en un cambio de actitud, el estudiante podrá integrar a su lógica elementos esenciales para enfrentar una problemática donde haya que considerar la variabilidad y la incertidumbre o identificar si requiere de ayuda experta, contando con la capacidad para entender y seguir indicaciones.

Behar Gutiérrez & Grima Cintas (2001) plantean que no todos los estudiantes aprenden de la misma manera y que cada uno tiene su propio estilo de aprendizaje que hace que tenga más afinidad con unas estrategias que con otras, razón por la cual es conveniente involucrar una amplia gama de actividades que hagan robusto el proceso a los estilos de aprendizaje. Además, existen diferentes niveles en los objetivos de enseñanza, que pueden lograrse escogiendo las estrategias adecuadas a cada uno de ellos, que varían en complejidad. Los objetivos de mayor nivel solo podrán alcanzarse con un aprendizaje activo, donde el estudiante es el protagonista y disfruta con el aprendizaje.

Los estudiantes antes de tomar el curso tienen sus propias concepciones sobre la aleatoriedad

y sus actitudes frente a la incertidumbre no siempre son acertadas, lo cual se constituye en una limitante del proceso de aprendizaje. En este sentido es conveniente que el profesor explore las preconcepciones que tienen los estudiantes para tomarlas en cuenta en su estrategia, pues el estudiante, para aceptar lo nuevo, tiene que convencerse que lo viejo no funciona. El manejo del proceso de enseñanza-aprendizaje de la estadística puede enfocarse bajo la filosofía del mejoramiento continuo de la calidad, planeando todas las fases, empezando por el currículo, definiendo unas metas claras, ejecutando el proceso planeado, verificando la consecución de los objetivos y con base en ello, retroalimentando el proceso con acciones correctivas. Para aplicar esta filosofía, es necesario tener una medida clara del logro en términos del aprendizaje, la cual constituye la evaluación, pues no se puede mejorar lo que no se puede medir. La investigación sobre instrumentos fiables de evaluación es uno de los pilares y prioridades de la investigación en educación estadística, pues de esta manera, se avanzaría en el grado de generalización de los resultados de investigaciones empíricas. Con instrumentos bien definidos de evaluación podrían medirse más objetivamente los impactos producidos por cambios en los métodos, estrategias, o medios de aprendizaje (Behar Gutiérrez & Grima Cintas 2001).

Los conceptos y reflexiones planteados por los autores sobre la relación existente entre la forma como se conduce el proceso de enseñanza-aprendizaje de la estadística y las metas que se aspiran lograr con el mismo, además del enfoque de los cursos introductorios de Estadística, la problemática de la enseñanza en esta área, avances hechos en este campo y algunas opciones que pueden ser abordadas para mejorar el proceso, permiten contextualizar el problema de estudio y son de utilidad para comprender y analizar las respuestas emitidas por los estudiantes a través del instrumento de medición aplicado, así como para identificar relevancias en las variables a considerar en el presente estudio.

## **6.2. Marco Teórico Estadístico**

De acuerdo con la naturaleza de las variables consideradas para estudio, el indicador del rendimiento académico de los estudiantes cuyos valores de respuesta son binarios y la

inclusión de otras variables numéricas y categóricas, se tomará como base para la obtención de resultados, los planteamientos sobre el modelo lineal generalizado y modelos de regresión binaria propuestos por Agresti (2015), Dobson & Barnett (2008) y Fox (2015).

### 6.2.1. Prueba exacta de Fisher

Esta prueba, propuesta en el año 1934, es útil para hacer inferencia en casos de distribuciones exactas cuando el tamaño de muestra es pequeño, para analizar si existe dependencia estadísticamente significativa entre dos variables dicotómicas. Una distribución no depende de parámetros desconocidos que resultan del condicionamiento sobre los totales marginales de la tabla de contingencia. Estos no suelen ser fijados naturalmente. Para muestreo independiente binomial en las dos filas, solo los totales marginales de la fila son fijos. Bajo la  $H_0$ : independencia, condicionada en ambos grupos de totales marginales producen la distribución hipergeométrica (Agresti 2002)

$$p(t) = P(n_{11} = t) = \frac{\binom{n_{1+}}{t} \binom{n_{2+}}{n_{+1}-t}}{\binom{n}{n_{+1}}} \quad (6-1)$$

En (6-1) se expresa la distribución de  $n_{ij}$  solo en términos de  $n_{11}$ . Dados los totales marginales,  $n_{11}$  determina los conteos de las otras tres celdas. El rango de posibles valores para  $n_{11}$  es  $m_- \leq n_{11} \leq m_+$ , donde  $m_- = \max(0, n_{1+} + n_{+1} - n)$  y  $m_+ = \min(n_{1+}, n_{+1})$ .

Para tablas 2x2, la independencia es equivalente al odds ratio  $\theta = 1$ . Asumiendo  $H_0 : \theta = 1$ , las dos variables son independientes y  $H_1 : \theta > 1$ , que existe asociación positiva entre las dos variables. El valor-p es igual a  $P(n_{11} \geq t)$ , donde  $t_0$  denota el valor observado de  $n_{11}$ , hace referencia a la probabilidad de la cola derecha de la distribución hipergeométrica. Se rechaza la  $H_0$  cuando el valor-p es menor que el nivel de significancia ( $\alpha$ ), es decir, con un nivel de confianza  $(1 - \alpha)$  las dos variables están asociadas positivamente.



### 6.2.2. Prueba Chi-Cuadrado de Independencia

La prueba Chi-cuadrado de independencia es una prueba que mide la diferencia entre una distribución de frecuencias observadas y esperadas, sirve para determinar si dos variables categóricas están asociadas. Se considera la  $H_0$  que las probabilidades de las celdas en una tabla de contingencia de dos vías son iguales a ciertos valores fijos  $\{\pi_{ij}\}$ . Para una muestra de tamaño  $n$  con conteo de celdas  $\{n_{ij}\}$ , los valores  $\{\mu_{ij} = n\pi_{ij}\}$  son las frecuencias esperadas. Aquí  $\mu_{ij} = E(n_{ij})$  bajo  $H_0$ . Para determinar si los datos contradicen la  $H_0$ , se compara  $\{n_{ij}\}$  con  $\{\pi_{ij}\}$ , si la  $H_0$  es verdadera, los valores fijos deben estar más cerca de los valores de las frecuencias esperadas en cada celda (Agresti 1996).

En tablas de contingencia de dos vías con probabilidades conjuntas  $\{\pi_{ij}\}$  para variables de respuesta, la hipótesis nula de independencia estadística es

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}; \text{ para todo } i \text{ y } j.$$

Para probar  $H_0$ , se identifica  $\mu_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j}$  como la frecuencia esperada de  $n_{ij}$  asumiendo independencia. Por lo general,  $\pi_{i+}$  y  $\pi_{+j}$  son desconocidos, así como el valor esperado. Para estimar las frecuencias esperadas, las proporciones muestrales por las probabilidades marginales desconocidas, dando

$$\hat{\mu}_{ij} = n\hat{\pi}_{i+}\hat{\pi}_{+j} = n \left( \frac{n_{i+}}{n} \right) \left( \frac{n_{+j}}{n} \right) = \frac{n_{i+}n_{+j}}{n} \quad (6-2)$$

basados en los totales marginales de la fila  $n_{i+}$  y los totales marginales de la columna  $n_{+j}$ . Para la prueba de independencia en tablas de contingencia  $I \times J$ ,  $X^2$  es igual a (Agresti 2002)

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \quad (6-3)$$

Con grados de libertad  $df = (I - 1)(J - 1)$  para tamaños de muestra grandes. Se rechaza la  $H_0$ , si el valor-p de la prueba es menor al nivel de significancia ( $\alpha$ ), es decir, con un nivel de confianza  $(1 - \alpha)$  existe asociación entre las dos variables.

### 6.2.3. Modelo Lineal Generalizado

Los Modelos Lineales Generalizados (MLG) fueron introducidos por Nelder and Wedderburn (1972) y son una extensión de los modelos lineales que permiten utilizar datos con distribuciones no normales que resultan ser miembros de la familia exponencial, entre los cuales se encuentran la distribución Normal, Poisson, Binomial, Exponencial y Gamma (Montgomery et al. 2006).

De forma similar al modelo lineal con distribución normal, el MLG usa una combinación lineal de variables para predecir una respuesta, por lo tanto, el modelo puede escribirse como:

$$Y = X\beta + \varepsilon$$

Donde la variable respuesta es determinada por un componente lineal  $X\beta$  que es el valor esperado de la respuesta ( $\mu$ ), mas el término aleatorio.

Un Modelo Lineal Generalizado es especificado así, (Madsen & Thyregod 2010):

- Distribución / Función de varianza: Especificación de la distribución o función de la varianza  $V(\mu)$ . La función explica la varianza en función del valor medio para una determinada familia de distribuciones.
- Función de enlace: Especificación de la función de enlace  $g(\cdot)$ , que describe la relación entre el predictor lineal  $\eta_i$  y el valor medio del parámetro  $\mu_i = E[Y_i]$ . Donde  $g(\mu)$  es una función conocida, monótona y diferenciable de  $\eta$ . Así que,  $\mu = g^{-1}(\eta)$
- Predictor lineal: Especificación de la dependencia lineal dada por  $g(\mu) = \eta = (x)^T\beta$ , donde el parámetro  $\beta$  es estimado mediante el método de máxima verosimilitud.

Si la distribución de probabilidad de la variable aleatoria  $Y$  depende de un solo parámetro  $\theta$ , pertenece a la familia exponencial si la función de densidad se puede expresar así (Dobson

& Barnett 2008),

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)] \quad (6-4)$$

Donde  $a(y) = y$  es la forma canónica,  $b(\theta)$  es el parámetro natural de la distribución y  $d(y)$  una constante.

En la tabla **6-1** se pueden observar las funciones de enlace más utilizadas dependiendo del tipo de distribución:

**Tabla 6-1.:** Funciones de enlace más utilizadas (Madsen & Thyregod, 2010)

| Nombre                 | Función de enlace $\eta = g(\mu)$    | $\mu = g^{-1}(\eta)$              |
|------------------------|--------------------------------------|-----------------------------------|
| Identidad              | $g(\mu)$                             | $\eta$                            |
| Logarítmica            | $\log(\mu)$                          | $\exp(\eta)$                      |
| Logit                  | $\log\left(\frac{\mu}{1-\mu}\right)$ | $\frac{\exp(\eta)}{1+\exp(\eta)}$ |
| Recíproca              | $\frac{1}{\mu}$                      | $\frac{1}{\eta}$                  |
| Potencia               | $\mu^k$                              | $\eta^{1/k}$                      |
| Raíz cuadrada          | $\sqrt{\mu}$                         | $\eta^2$                          |
| Probit                 | $\Phi^{-1}(\mu)$                     | $\Phi(\eta)$                      |
| Log-Log                | $\log[-\log(\mu)]$                   | $\exp[-\exp(\eta)]$               |
| Log-Log Complementaria | $\log[-\log(1-\mu)]$                 | $1 - \exp[-\exp(\eta)]$           |

#### 6.2.4. Modelo de Regresión Binaria

Se considera una serie de eventos binarios llamados, donde la respuesta  $Y$  puede tomar uno de dos valores posibles, codificados generalmente como 1 para éxito o categoría de interés y 0 para éxito y fracaso. La distribución de  $Y$  es una Bernoulli, en la que

$$P(Y_i = 0) = 1 - \pi_i ; P(Y_i = 1) = \pi_i$$

Si además se observan una serie de covariables continuas o de tipo factor, el objetivo del análisis con el modelo lineal generalizado es predecir la probabilidad asociada al éxito (o al fracaso), en función de dichas covariables (Martínez Mayoral & Morales Socuéllamos 2001).

Suponiendo que la variable aleatoria  $y$  es el número de éxitos en  $n$  ensayos independientes en los que la probabilidad de éxito,  $\pi$ , es la misma en todos los ensayos (Dobson & Barnett 2008). Entonces  $Y$  tiene una distribución binomial,  $y \sim \text{Bin}(n, \pi)$ , distribución que se utiliza en los MLG para este tipo de datos. Aquí las observaciones son el número de éxitos obtenidos en  $n$  ensayos, cada uno con probabilidad de ocurrencia  $\pi$ , el parámetro de interés  $\pi$  es la proporción de éxitos, luego, la función de densidad de probabilidad está dada por

$$f(y; \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$$

Y se reescribe

$$\begin{aligned} f(y; \pi) &= \exp \left[ y \log \pi - y \log(1 - \pi) + n \log(1 - \pi) + \log \binom{n}{y} \right] \\ &= \exp \left[ y \log \left( \frac{\pi}{1 - \pi} \right) + n \log(1 - \pi) + \log \binom{n}{y} \right] \end{aligned} \quad (6-5)$$

La función 6-5 es de la forma 6-4 con  $b(\pi) = \log \pi - \log(1 - \pi) = \log [\pi/(1 - \pi)]$ . La media y la varianza de la distribución binomial son

$$E(y) = n\pi ; \text{Var}(y) = n\pi(1 - \pi)$$

El modelo lineal generalizado con un componente aleatorio binomial (para datos con respuesta binaria) se denomina **Modelo de Probabilidad Lineal** y se expresa:

$$\pi(x) = \beta_0 + \beta_1 X$$

Este modelo usado en aplicaciones prácticas, tiene la desventaja que aunque  $\pi$  es una probabilidad, es decir, está restringido al dominio  $[0, 1]$ , los valores que ajusta pueden ser menores que cero o mayores a uno, afectando la predicción para determinados valores de las variables explicativas. Por tanto, para garantizar que el rango de valores esté restringido al

intervalo  $[0, 1]$ , se debe hacer una transformación a  $\pi$  que se llama *función de enlace*  $g(\pi)$  usando la función de distribución acumulada (Dobson & Barnett 2008).

$$\pi = \int_{-\infty}^t f(s)ds$$

donde  $f(s) \geq 0$  y  $\int_{-\infty}^{\infty} f(s)ds = 1$ . La función de probabilidad  $f(s)$  es llamada **distribución de tolerancia**.

Según McCullagh & Nelder (1989), las principales funciones de enlace  $g(\pi)$  utilizadas para los modelos lineales generalizados con respuesta binaria son:

1. Logit o función logística

$$g_1(\pi) = \log \left[ \frac{\pi}{1 - \pi} \right]$$

2. Probit o función normal inversa

$$g_2(\pi) = \Phi^{-1}(\pi)$$

3. Función Log-Log complementaria

$$g_3(\pi) = \log [-\log(1 - \pi)]$$

Las funciones logit y probit son adecuadas cuando la distribución es simétrica, cuando la distribución no es simétrica y la probabilidad de ocurrencia de un evento es muy baja o muy grande, la función de enlace más conveniente es la Log-Log complementaria (McCullagh & Nelder 1989).

### 6.2.5. Enlace log-log complementario

La función de tolerancia es la función de densidad de la distribución del valor extremo:

$$f(s) = \beta_2 \exp \{(\beta_1 + \beta_2 s) - \exp(\beta_1 + \beta_2 s)\}$$

El enlace Log-Log complementario sale de esa función:

$$\pi = \int_{-\infty}^x f(s)ds = 1 - \exp\{-\exp(\beta_1 + \beta_2 s)\}$$

$$g(\pi) = \text{cloglog}(\pi) = \log(-\log(1 - \pi)) = \beta_1 + \beta_2 x \quad (6-6)$$

### 6.2.6. Método Newton-Raphson para el Ajuste del MLG

Entre los métodos más utilizados para ajustar el modelo lineal generalizado, está el procedimiento iterativo Newton-Raphson que converge rápidamente y se basa en la secuencia iterativa a partir de las estimaciones por mínimos cuadrados ponderados iterativos (Dobson & Barnett 2008).

Considere las variables aleatorias independientes  $y_1, \dots, y_n$  que satisfacen las propiedades del modelo lineal generalizado. Se desea estimar los parámetros  $\beta$ , que se relacionan con las variables  $y_i$ , la  $E(y_i) = \mu_i$  y  $g(\mu_i) = x_i^T \beta$ .

Para cada  $y_i$  la función de log-verosimilitud es:

$$l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i) \quad (6-7)$$

donde las funciones  $b, c$  y  $d$  se encuentran definidos en la ecuación 6-4. Se definen el valor esperado y la varianza de  $a(y) = y$ ,

$$E(y_i) = -c'(\theta_i)/b'(\theta_i) \quad (6-8)$$

$$\text{Var}(y_i) = [b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)] / [b'(\theta_i)]^3 \quad (6-9)$$

La función de log-verosimilitud para todos los  $y_i$  es:

$$\iota = \sum_{i=1}^n \iota_i = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i) \quad (6-10)$$

Para obtener el estimador máximo verosímil para el parámetro  $\beta$ , se requiere:

$$\frac{\partial \ell}{\partial \beta_j} = U_j = \sum \left[ \frac{\partial \ell}{\partial \beta_j} \right] = \sum_{i=1}^n \left[ \frac{\partial \ell_i}{\partial \theta_i}, \frac{\partial \theta_i}{\partial \mu_i}, \frac{\partial \mu_i}{\partial \beta_j} \right] \quad (6-11)$$

Se aplica la regla de la cadena para la diferenciación de cada uno de los términos de la ecuación 6-11, obteniendo las siguientes expresiones:

$$\begin{aligned} \frac{\partial \ell_i}{\partial \theta_i} &= b'(\theta_i)(y_i - \mu_i) \\ \frac{\partial \theta_i}{\partial \mu_i} &= \frac{1}{b'(\theta_i) \text{Var}(y_i)} \\ \frac{\partial \mu_i}{\partial \beta_j} &= \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \end{aligned}$$

Finalmente, se obtiene la función Score  $U_j$  a partir de la expresión  $g(\mu_i) = x_i^T \beta = \eta_i$ , tal que:

$$U_j = \sum_{i=1}^n \left[ \frac{(y_i - \mu_i)}{\text{Var}(y_i)} x_{ij} \left( \frac{\partial \mu_i}{\partial \eta_i} \right) \right] \quad (6-12)$$

A partir de 6-12, se constituye la matriz de información  $\mathfrak{J}$  y se denota de la siguiente forma:

$$\begin{aligned} \mathfrak{J}_{jk} &= E \left\{ \sum_{i=1}^n \left[ \frac{(y_i - \mu_i)}{\text{Var}(y_i)} x_{ij} \left( \frac{\partial \mu_i}{\partial \eta_i} \right) \right] \sum_{l=1}^n \left[ \frac{(y_l - \mu_l)}{\text{Var}(y_l)} x_{lk} \left( \frac{\partial \mu_l}{\partial \eta_l} \right) \right] \right\} \\ &= \sum_{i=1}^n \frac{E[(y_i - \mu_i)^2] x_{ij} x_{ik}}{[\text{Var}(y_i)^2]} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \end{aligned} \quad (6-13)$$

La  $E[(y_i - \mu_i)(y_l - \mu_l)] = 0$  para  $i \neq l$  cuando las  $y_i$  son independientes y  $E[(y_i - \mu_i)]^2 = \text{Var}(y_i)$ , por lo anterior 6-13 queda simplificada de la siguiente forma:

$$\mathfrak{J}_{jk} = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{Var}(y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad (6-14)$$

La expresión 6-14 se puede escribir como:  $\mathfrak{J} = X^T W X$ , donde  $W$  es una matriz diagonal de dimensión  $n \times n$  con elementos,

$$w_{ii} = \frac{1}{Var(y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

Teniendo en cuenta la estimación de ecuaciones  $\theta^m = \theta^{(m-1)} + \frac{U^{(m-1)}}{\mathfrak{I}^{(m-1)}}$  descrito por Dobson & Barnett (2008), el método se generaliza en:

$$b^m = b^{(m-1)} + [\mathfrak{I}^{(m-1)}]^{-1} U^{(m-1)} \quad (6-15)$$

donde  $b^m$  es el vector de parámetros a estimar,  $\beta_1, \dots, \beta_p$ , en la iteración m-ésima,  $[\mathfrak{I}^{(m-1)}]^{-1}$  es la inversa de la matriz de información con elementos  $\mathfrak{I}_{jk}$  dada la expresión 6-14 y  $U^{(m-1)}$  es el vector de elementales dado en la expresión 6-12, todos evaluados en  $b^{(m-1)}$ .

Según Montgomery et al. (2006) el procedimiento de iteración continúa hasta la convergencia, es decir:

$$\left| b_i^m - b_i^{(m-1)} \right| < \delta \quad (6-16)$$

con  $i = 1, 2, \dots, p$ , donde  $\delta$  es un número pequeño, por ejemplo  $1,0 \times 10^{-6}$ . En cada iteración se debe evaluar la bondad de ajuste para asegurar que se haya obtenido una reducción en su valor.

### 6.2.7. Interpretación de parámetros del Modelo con Respuesta Binaria (Enlace Log-Log Complementario)

#### Odds Ratio

El odds asociado a un evento se define como la razón entre la probabilidad de que el evento ocurra y la probabilidad de que no ocurra, es decir, es un número que expresa cuanto más posible es que se produzca el evento frente a que no se produzca (Silva & Barroso 2004).



Sea el evento  $E$ , la probabilidad de que ocurra el evento es  $P(E)$  y el odds correspondiente es  $O(E)$ , luego se tiene:

$$O(E) = \frac{P(E)}{1 - P(E)} \quad (6-17)$$

El *odds ratio* es la razón entre el *odds* correspondiente a un evento bajo cierta condición y el *odds* correspondiente al mismo evento bajo otra condición. El *odds ratio*  $OR$  es tal que:

$$OR = \frac{\frac{P_F(E)}{1 - P_F(E)}}{\frac{P_{\bar{F}}(E)}{1 - P_{\bar{F}}(E)}} \quad (6-18)$$

En el modelo con función de enlace Log-Log Complementario se modela linealmente el logaritmo del odds ratio de esta forma (Silva & Barroso 2004):

$$\ln(OR) = \beta(X_1 - X_2)$$

Luego,

$$OR = e^{\beta(X_1 - X_2)}$$

En el planteamiento de un modelo con función de enlace Log-Log Complementario el cociente de probabilidades *odds ratio* tiene las siguientes propiedades (Silva & Barroso 2004):

- Cuando  $OR > 1$ , La variable es un factor de riesgo.
- Cuando  $OR < 1$ , La variable es un factor de protección.
- Cuando  $OR = 1$ , La variable no presenta influencia.

### 6.2.8. Prueba de Bondad de ajuste del Modelo

Con el fin de verificar si un modelo de regresión ajustado representa adecuadamente el proceso que se estudia y por tanto, compatible con los datos utilizados, se hace uso de la **Razón de Verosimilitud** (Silva & Barroso 2004).

La *desviación del modelo* compara el logaritmo de la verosimilitud del modelo ajustado con el logaritmo de la verosimilitud de un modelo saturado, es decir, el modelo que tiene exactamente  $n$  parámetros y se ajusta perfectamente a los datos de la muestra (Montgomery et al. 2006). En forma específica, la desviación del modelo se define así:

$$\lambda(\beta) = 2\ln L(\text{modelo saturado}) - 2\ln L(\hat{\beta}) \quad (6-19)$$

Si el tamaño de la muestra es grande, la desviación del modelo tiene aproximadamente una distribución Chi-Cuadrado, con  $n - p$  grados de libertad. Valores grandes de la desviación del modelo indican que el modelo no es correcto, mientras que valores pequeños indican que el modelo ajustado se ajusta bien a los datos. Los criterios de prueba son los siguientes (Montgomery et al. 2006):

- Si  $\lambda(\beta) \leq \chi^2_{\alpha, n-p}$ , se concluye que el modelo ajustado es adecuado.
- Si  $\lambda(\beta) > \chi^2_{\alpha, n-p}$ , se concluye que el modelo ajustado no es adecuado.

### 6.2.9. Criterio de información de Akaike (AIC)

El criterio de información de Akaike (An Information Criterion, AIC) (Akaike 1974) proporciona un método simple y objetivo que selecciona el modelo más adecuado para caracterizar los datos experimentales, el que menos información pierde de todos los modelos considerados. Se define como:

$$AIC = -2\log(\mathcal{L}(\hat{\theta})) + 2K \quad (6-20)$$

donde  $\log(\mathcal{L}(\hat{\theta}))$  es el logaritmo de la máxima verosimilitud, que permite determinar los valores de los parámetros libres de un modelo estadístico, y  $K$  es el número de parámetros libres del modelo.

Esta expresión proporciona una estimación de la distancia entre el modelo y el mecanismo que realmente genera los datos observados, que es desconocido y en algunos casos imposible de caracterizar. Como la estimación se hace en función de los datos experimentales, esta

distancia es siempre relativa y dependiente del conjunto de datos experimentales. El valor de AIC es útil para comparar modelos similares con distintos grados de complejidad o modelos iguales (con las mismas variables) pero con funciones de enlace distintas. El menor valor de AIC indica que o bien el modelo se ajusta mejor a los datos experimentales o que es menos complejo. Este criterio cuantifica simultáneamente la precisión y sencillez del modelo (Martinez et al. 2009).

### 6.2.10. Prueba de significancia de los parámetros

Una vez estimado el modelo, es de interés contrastar si determinada variable explicativa tiene un efecto significativo sobre la respuesta o no. Para esto se puede utilizar la Prueba de Wald.

#### Prueba de Wald

El estadístico de Wald permite contrastar si un coeficiente estimado  $\beta$  es significativamente distinto de cero. El estadístico de prueba para la hipótesis nula en

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

es

$$Z_0 = \frac{\hat{\beta}_j}{\sqrt{Var(\hat{\beta}_j)}} \sim N(0, 1)$$

La distribución de referencia para este estadístico es la distribución normal estándar. Otras pruebas elevan al cuadrado el estadístico de Wald  $Z_0$  y lo comparan con una distribución Chi-Cuadrado con un grado de libertad (Montgomery et al. 2006).

### 6.2.11. Curva ROC

La curva ROC es una herramienta estadística utilizada para clasificar la capacidad discriminante de una prueba diagnóstica dicotómica. Es una prueba basada en una variable de decisión, cuyo objetivo es clasificar a los individuos de una población en dos grupos: uno

que presente un evento de interés y otro que no. Esta capacidad discriminante está sujeta al valor umbral elegido de entre todos los posibles resultados de la variable de decisión, es decir, la variable por cuyo resultado se clasifica a cada individuo en un grupo u otro. La curva es el gráfico resultante de representar, para cada corte elegido, las medidas de sensibilidad y especificidad de la prueba diagnóstica.

Por un lado, la sensibilidad cuantifica la proporción de individuos que presenta el evento de interés y que son clasificados por la prueba como portadores de dicho evento. Por otro lado, la especificidad cuantifica la proporción de individuos que no lo presentan y son clasificados por la prueba como tal (Valle Benavides 2017).

La **curva ROC poblacional** representa  $1 - \text{especificidad}$  frente a la sensibilidad para cada punto de corte( $c$ ) en la escala de resultados de la prueba en estudio. Esto es,  $Y = f(x)$ , donde (Valle Benavides 2017):

$$ROC(c) = \begin{cases} y = S(c) \\ x = 1 - E(c) \end{cases} \quad (6-21)$$

Dada la dificultad de obtener datos poblacionales, se puede aproximar la expresión 6-21 por la **curva ROC muestral**, que representa la *fracción de falsos positivos(FFP)* en abscisas frente a la *fracción de verdaderos positivos(FVP)* en ordenadas (Valle Benavides 2017).

$$ROC_p(c) = \begin{cases} y = FVP(c) \\ x = FFP(c) \end{cases} \quad (6-22)$$

Ya que en ambos ejes se tienen probabilidades, la curva ROC tanto muestral como poblacional, estará contenida en el cuadro  $[0,1] \times [0,1]$ . Además, se considera que los enfermos tienen valores de  $x$ , en general, mayores que los sanos. Por esta razón, la curva estaría contenida en el triángulo:  $\{(x, y) \mid 0 \leq x \leq y \leq 1\}$  (Valle Benavides 2017).

Dados los resultados de la prueba, no siempre se sabe que distribución siguen o bajo qué parámetros. Esta situación conduce a tener curvas escalonadas (*Método no paramétrico*) o curvas suaves suponiendo la distribución previamente (*Método paramétrico*) (Valle Benavides 2017).

### Método No Paramétrico

El método no paramétrico para la construcción de la curva ROC no hace suposición alguna sobre la distribución de los resultados de la prueba en ambos grupos. Debido a esta característica, este método tiene la propiedad de robustez (Valle Benavides 2017).

### Método Empírico

En el método empírico se hace uso de las *funciones de distribución empíricas*:  $\hat{F}_{nE}$  y  $\hat{F}_{nS}$  de los individuos que presentan el evento y los que no, respectivamente, para la construcción de la curva ROC (Valle Benavides 2017).

Dada una muestra aleatoria simple  $x_1, \dots, x_n$  asociada a la variable aleatoria  $x$  con función de distribución  $F$ , se define la función de distribución empírica asociada a la muestra como (Valle Benavides 2017):

$$\begin{aligned} \hat{F}_n : \mathbb{R} &\rightarrow [0, 1] \\ x \rightarrow \hat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i(x) \end{aligned}$$

donde:

$$\varepsilon_i(x) = \begin{cases} 1 & \text{si } x_i \leq x \\ 0 & \text{si } x_i > x \end{cases} \quad (6-23)$$

En este caso, las expresiones de ambas funciones de distribución empíricas son:

$$\begin{aligned} \hat{F}_{nE}(x) &= \frac{1}{nE} \sum_{i=1}^{nE} \varepsilon_i(x) \\ \hat{F}_{nS}(x) &= \frac{1}{nS} \sum_{i=1}^{nS} \varepsilon_i(x) \end{aligned}$$

siendo  $x_1, \dots, x_{nE}$  los individuos que presentan el evento en la muestra y  $x_1, \dots, x_{nS}$  los individuos que no presentan el evento en la muestra.

Se define la curva ROC empírica como la curva construída uniendo los  $(1 - \hat{F}_{nS}(x), 1 - \hat{F}_{nE}(x))$  consecutivos.

$$\widehat{ROC}_n(t) = 1 - \hat{F}_{nE}(1 - \hat{F}_{nS}(x))$$

siendo  $n$  el tamaño de la muestra.

### 6.2.12. Índice de Youden

El Índice de Youden es una medida de exactitud que expresa la diferencia entre las respuestas positivas correctas y las respuestas positivas incorrectas, es decir, los positivos de la prueba en ausencia del evento de interés (Valle Benavides 2017):

$$\gamma = Pr(y = 1 \mid D = 1) + Pr(y = 1 \mid D = 0)$$

que equivale a,

$$\gamma = S + (1 - E)$$

Y el estimador se puede escribir como sigue:

$$\hat{\gamma} = FVP - FFP \tag{6-24}$$

El índice 6-24 toma valores en el intervalo  $[0, 1]$ , los valores cercanos a 0 indican una prueba con poca capacidad discriminatoria y los valores cercanos a 1 corresponden a una prueba perfecta pues significaría que  $S = 1$  y  $E = 1$  (Valle Benavides 2017).

### 6.2.13. Capacidad Clasificatoria del Modelo con Respuesta Binaria (Enlace Log-Log Complementario)

Para validar los resultados obtenidos con el modelo de predicción de respuesta binaria (enlace Log-Log Complementario) se puede construir una tabla de clasificación en la que se

muestran los casos bien clasificados en la diagonal principal y los casos mal clasificados en la segunda diagonal. Se establece un valor para el punto de corte  $p_0$  y si la  $p_i$  obtenida para el  $i$ -ésimo sujeto es inferior al  $p_0$ , la probabilidad de éxito se considera tan baja para ese sujeto que se le pronostica “No éxito”, y si la  $p_i$  obtenida para el  $i$ -ésimo sujeto es superior al  $p_0$ , la probabilidad de éxito se considera tan alta para ese sujeto que se le pronostica “éxito”. Esto permite crear un mecanismo de clasificación en función del resultado probabilístico.

En consecuencia de aplicar este mecanismo, se producen falsos positivos (se predice el éxito, pero realmente no lo obtuvieron), y falsos negativos (se predice como no éxito cuando realmente si lo obtuvieron).

Por lo anterior, para cada  $p_0$  se puede formar una configuración con la que pueden clasificarse los datos de una muestra de acuerdo con los resultados reales y predichos (Ayçaguer 1994). Esto se muestra en la tabla **6-2**:

**Tabla 6-2.:** Tabla de Clasificación

| Predicho por modelo      | Observado |           | Total |
|--------------------------|-----------|-----------|-------|
|                          | Éxito=1   | Fracaso=0 |       |
| Éxito ( $P > P_0$ )      | a         | b         | a+b   |
| Fracaso ( $P \leq P_0$ ) | c         | d         | c+d   |
| Total                    | a+c       | b+d       | n     |

donde:

- a: Es el número de éxitos bien clasificados por el modelo.
- b: Es el número de fracasos mal clasificados por el modelo.
- c: Es el número de éxitos mal clasificados por el modelo.
- d: Es el número de fracasos bien clasificados por el modelo.

Luego,

- Verdaderos positivos (VP): (sensibilidad) predicción de éxito cuando realmente obtuvo éxito.
- Verdaderos fallos (VF): (especificidad) predicción de fracaso cuando realmente obtuvo fracaso.
- Falsos positivos (FP): predicción de éxito cuando realmente obtuvo fracaso.
- Falsos fallos (FF): predicción de fracaso cuando realmente obtuvo éxito.
- Porcentaje correcto de clasificación (P.C.C): probabilidad de acertar en la predicción para todos los individuos.

$$VP = \frac{a}{a + c} \quad VF = \frac{d}{b + d} \quad FP = \frac{b}{b + d} \quad FF = \frac{c}{a + c}$$

#### 6.2.14. Enfoque Bayesiano

Las conclusiones estadísticas bayesianas sobre un parámetro  $\theta$ , o datos no observados  $\tilde{y}$ , se hacen en términos de las declaraciones de probabilidad. Estas declaraciones de probabilidad están condicionadas al valor observado de  $y$ , y se escriben como  $p(\theta|y)$  o  $p(\tilde{y}|y)$ . También se condiciona implícitamente los valores conocidos de cualquier covariable  $x$ . En el nivel de condicionamiento básico sobre los datos observados es que la inferencia bayesiana se aparta del enfoque a la inferencia estadística clásica, que se basa en una evaluación retrospectiva del procedimiento utilizado para estimar  $\theta$  (o  $\tilde{y}$ ) sobre la distribución de los posibles valores de  $y$  condicionales al verdadero valor desconocido de  $\theta$ . A pesar de esta diferencia, en muchos análisis simples, resultan conclusiones similares a los resultados en los dos enfoques a la inferencia estadística. Sin embargo, los análisis obtenidos que utilizan métodos bayesianos pueden ser extendidos a problemas más complejos. (Gelman et al. 2013)

Desde el enfoque bayesiano, una cuantificación de la subjetividad puede emitir juicios de probabilidad sobre una hipótesis  $\theta$  (conocimiento a priori) y expresar el grado de convicción (probabilidad subjetiva) al respecto, tanto antes de observar los datos como después de observarlos. El teorema de Bayes en su formulación expresa:



$$P(\theta|datos) = \frac{P(datos|\theta)P(\theta)}{P(datos)}$$

La probabilidad a priori de una hipótesis  $P(\theta)$ , se ve trasformada en una probabilidad a posteriori  $P(\theta|datos)$  una vez se incluye la evidencia que los datos aportan, es decir, el teorema de Bayes devuelve la percepción probabilística original actualizada por los datos. Expresa la probabilidad a priori  $P(\theta)$  como un número, sin embargo, la convicción inicial se puede expresar mediante una distribución de probabilidad.

### 6.2.15. Distribución Binomial

Suponiendo que se realizan  $n$  ensayos independientes, cada uno de los cuales resulta en un “éxito” con probabilidad  $\theta$  y en un “fallo” con probabilidad  $(1 - \theta)$ . Si  $X$  representa la cantidad de éxitos que ocurren en las  $n$  pruebas, entonces se dice que  $X$  se distribuye Binomial con parámetros  $(n, \theta)$  (Ross 2014).

La función de probabilidad de masa de una variable aleatoria binomial que tiene parámetros  $(n, \theta)$  viene dada por

$$f(X|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad (6-25)$$

Donde  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

Se tiene que

- $x$ : número de éxitos, con  $x = 0, 1, 2, \dots, n$ .
- $\theta$ : parámetro a estimar y representa la probabilidad de éxito, con  $0 < \theta < 1$ .
- $n$ : número de ensayos de Bernoulli, con  $n = 1, 2, \dots$ .
- $P(B_i|A)$ : Probabilidad posterior.

La función de verosimilitud de los datos está dada por

$$L(\theta | x_1, x_2, \dots, x_r) = \prod_{j=1}^r \binom{n_j}{x_j} \theta^{x_j} (1 - \theta)^{n_j - x_j}$$

### 6.2.16. Distribución Beta

Una variable aleatoria  $x$  con campo de variación  $[0,1]$  sigue una distribución Beta con parámetros  $\alpha, \beta$  (ambos positivos) cuando su función de densidad es:

$$x \sim \text{Beta}(\alpha, \beta)$$

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad x \in (0, 1) \quad (6-26)$$

Dado que su campo de variación es el espacio  $[0,1]$ , la distribución beta se utiliza a menudo como modelo de distribución de probabilidades o proporciones. Se usa en la estimación bayesiana de proporciones y probabilidades como distribución conjugada a una verosimilitud binomial. Cabe mencionar que una distribución Beta(1,1) coincide con una distribución uniforme en el intervalo  $[0,1], U([0,1])$  (De Lejarza 2012).

Su media está dada por

$$\mu = \frac{\alpha}{\alpha + \beta}$$

La varianza será:

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Y la moda se obtiene maximizando la función de densidad:

$$\mu_0 = \frac{\alpha - 1}{\alpha + \beta - 2}$$

### 6.2.17. Teorema de Bayes

Sean  $B_1, B_2, \dots, B_k$  eventos mutuamente excluyentes y exhaustivos. Para cualquier evento nuevo  $A$ , tenemos

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^k P(A|B_i)P(B_i)}$$

donde

- $P(A|B_i)$ : Verosimilitud (Probabilidad observada).
- $P(B_i)$ : Apriori.
- $P(A)$ : Probabilidad total.
- $P(B_i|A)$ : Probabilidad posterior.

### Teorema de Bayes para variables aleatorias

Sean  $X$  y  $\theta$  variables aleatorias con funciones de probabilidad  $f(x|\theta)$  y  $g(\theta)$ .

$$g(\theta|x) = \frac{f(x|\theta)g(\theta)}{\int_{\Theta} f(x|\theta)g(\theta)d\theta}$$

donde

- $X$ : Datos (escalar, vector o matriz).
- $\theta$ : Parámetros desconocido (escalar, vector o matriz).
- $f(x_1, x_2, \dots, x_n|\theta)$ : Verosimilitud de los datos dado el parámetro (desconocido)  $\theta$ .
- $g(\theta)$ : Distribución a priori de  $\theta$ .

Por el teorema anterior se obtiene la **distribución posterior**. La inferencia bayesiana se deriva de esta distribución. En la práctica, el denominador de la expresión anterior no necesita ser calculado en general, y la regla de Bayes se escribe como (Correa Morales et al. 2018)

$$g(\theta|x_1, x_2, \dots, x_n) \propto f(x_1, x_2, \dots, x_n|\theta)g(\theta) \quad (6-27)$$

Solo se necesita conocer la distribución posterior hasta una constante de normalización. En ocasiones se puede identificar la distribución posterior de  $\theta$  observando el numerador. El teorema de Bayes lo que hace es actualizar el modelo con la nueva información

$$\begin{aligned} g(\theta|x_1) &\propto f(x_1|\theta)g(\theta) \\ g(\theta|x_1, x_2) &\propto f(x_2|\theta)f(x_1|\theta)g(\theta) \\ &\propto f(x_2|\theta)g(\theta|x_1) \end{aligned}$$

Tanto en la estadística clásica como en la inferencia bayesiana, hay un parámetro poblacional de interés que se requiere estimar. La inferencia bayesiana modela la probabilidad de obtener diferentes valores de  $X$ , bajo diferentes valores de los parámetros. La diferencia con la estadística clásica, es que la inferencia bayesiana considera el parámetro como una variable aleatoria. Esto conduce a una aproximación diferente para modelar el problema y la inferencia sobre el parámetro. La inferencia bayesiana está basada en la probabilidad del parámetro dado los datos, en lugar de la distribución de los datos dado el parámetro.

### 6.2.18. Distribución Apriori

La distribución apriori mide el grado de conocimiento inicial que se tiene de los parámetros de estudio. Si se tiene un conocimiento previo sobre los parámetros, este se traduce en una distribución apriori. Es posible plantear tantas distribuciones apriori como estados iniciales de conocimiento existan y los diferentes resultados obtenidos en la distribución a posteriori bajo cada uno de los enfoques, adquieren importancia en relación con la convicción que se tenga sobre cada estado inicial.

Una distribución apriori puede ser:

- **Propia:** Es una distribución que asigna pesos no negativos y que suman o integran hasta uno, a todos los valores posibles del parámetro. Una distribución propia satisface las condiciones de función de densidad de probabilidad (Correa Morales et al. 2018).
- **Impropia:** Es una distribución que suma o integra a un valor diferente de uno. Si el valor es finito, entonces la distribución impropia induce una distribución propia normalizando la función. Si el valor es infinito, entonces la distribución funciona como ponderación o herramienta técnica para llegar a una distribución a posteriori (Correa Morales et al. 2018).
- **Apriori Informativa:** Es una distribución de probabilidad que refleja el conocimiento que se tiene sobre el parámetro antes de obtener la muestra de datos experimentales.

- **Apriori No Informativa:** Es una distribución que expresa el grado de desconocimiento frente al estado generador de los datos.

Para la elección de una distribución **apriori no informativa**, se cuentan con algunos criterios.

### Principio de la Razón de Insuficiencia de Laplace

Si no se tiene información de un especialista y el espacio parametral es finito, se puede utilizar una distribución apriori uniforme para reflejar ignorancia total, asumiendo que todos los valores posibles de  $\theta$  son equiprobables.

$$g(\theta) \propto 1 \quad 0 \leq \theta \leq 1$$

### Distribución Beta(1,1)

Una distribución Beta(1,1) coincide con una uniforme en  $[0,1]$  ya que su función de densidad degenera en la constante unidad:

$$x \sim \text{Beta}(1, 1) \implies f(x) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} x^{1-1} (1-x)^{1-1} = 1$$

Este resultado además permite utilizar la distribución uniforme como a priori no informativa en la estimación bayesiana de proporciones.

### Método de Jeffreys

La distribución apriori de Jeffreys satisface la propiedad local de uniformidad para distribuciones apriori no informativas. Está basada en la matriz de información de Fisher (Correa Morales et al. 2018).

Sea  $f(x|\theta)$  la densidad de  $x$  dado  $\theta$ . La información de Fisher es definida como

$$I(\theta) = -E \left[ \frac{\partial^2 \log(f(x|\theta))}{\partial \theta^2} \right]$$

Si  $\theta$  es un vector de  $p$  componentes, entonces

$$I(\vec{\theta}) = -E \left[ \frac{\partial^2 \log(f(x | \vec{\theta}))}{\partial \theta_i \partial \theta_j} \right]_{pxp}$$

y entonces  $I(\vec{\theta})$  será una matriz de dimensión  $pxp$ . Luego la apriori de Jeffreys se define como

$$g(\theta) = \sqrt{I(\theta)}$$

Esta distribución es impropia para muchos modelos, sin embargo, es propia para algunos (Correa Morales et al. 2018).

### Apriori de Jeffreys para un Modelo de Probabilidad Binomial

Sea  $X$  una variable aleatoria que se distribuye Binomial  $X \sim Bin(n, \theta)$

$$f(x | \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

Para elicitar una distribución apriori no informativa para el parámetro  $\theta$ , se obtiene la información de Fisher

$$I(\theta) = -E \left\{ \frac{\partial^2}{\partial \theta^2} \left[ \ln \binom{n}{x} \theta^x (1 - \theta)^{n-x} \right] \right\}$$

$$I(\theta) = \frac{n}{\theta} + \frac{n}{1 - \theta}$$

Con la información de Fisher se aproxima la distribución apriori

$$g(\theta) = \sqrt{\frac{n}{\theta(1 - \theta)}}$$

$$g(\theta) = \sqrt{n} \theta^{-1/2} \theta^{-1/2}$$

$$g(\theta) \propto \theta^{-1/2} \theta^{-1/2}$$

Se aproxima a una distribución Beta con parámetros  $\alpha = \frac{1}{2}$  y  $\beta = \frac{1}{2}$

$$g(\theta) = \frac{\Gamma\left(\frac{1}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right)} \theta^{\frac{1}{2} - 1} (1 - \theta)^{\frac{1}{2} - 1} \text{ con } \theta > 0 \quad (6-28)$$

En cuanto a la distribución **apriori informativa**, se cuentan con algunos criterios.

### Apriori Conjugada

Suponiendo que una distribución apriori pertenece a cierta clase o familia de distribuciones paramétricas  $D$ ; entonces se dice que esa distribución es conjugada respecto a la verosimilitud si la distribución aposteriori también pertenece a la clase  $D$ .

Dicho de otra forma, si  $F$  es una clase de distribuciones de muestreo  $p(y|\theta)$ , y  $P$  es una clase de distribuciones apriori para  $\theta$ , entonces la clase  $P$  es una familia conjugada para  $F$  si  $P(\theta|y)$  pertenece a la clase  $P$  para todo  $f(y|\theta) \in F$  y  $P(\theta) \in P$  (Gelman et al. 2013). La distribución inicial domina a la función de verosimilitud y  $P(\theta|y)$  tiene la misma forma que  $P(\theta)$ , con los parámetros actualizados por la información muestral.

El uso de distribuciones conjugadas puede simplificar el procedimiento de integración requerido para la marginalización, ya que al pertenecer la apriori y la aposteriori a la misma familia, el proceso de actualización de parámetros se simplifica (Correa Morales et al. 2018).

La distribución  $Beta(\alpha, \beta)$  es la distribución previa conjugada para verosimilitudes *Bernoulli*, *Binomial*, *Binomial Negativa*.

### Elicitación de Hiperparámetros de la Distribución Beta

Se define  $\theta$  como la proporción de interés que es una cantidad aleatoria con distribución de probabilidad  $g(\theta)$  que se desconoce. Una distribución que modela bien  $g(\theta)$  es la distribución Beta con hiperparámetros  $a$  y  $b$ . Según la metodología propuesta por Tovar Cuevas (2012), se puede usar la información que se dispone del parámetro para determinar un intervalo de probabilidad que en  $(1 - \alpha)\%$  de los valores de  $\theta_0$  caigan dentro del intervalo  $(\theta_1, \theta_2)$ .

Luego, el centro del intervalo  $\theta_0$  se puede aproximar a la media de la distribución beta (Tovar Cuevas 2012),

$$\theta_0 = \frac{\theta_1 + \theta_2}{2} \cong \frac{\alpha}{\alpha + \beta} \quad (6-29)$$

Con los dos percentiles de la distribución apriori y una aproximación del valor esperado, se puede hacer uso de la desigualdad de Chebyshev's para obtener la varianza (Tovar Cuevas 2012)

$$\begin{aligned}
 P(|\theta - E(\theta)| \geq k\sigma) &\leq \frac{1}{k^2} = \alpha \\
 P(|\theta - E(\theta)|^2 \geq k^2\sigma^2) &\leq \alpha \\
 P(\alpha[\theta - \theta_0]^2 \geq \sigma^2) &\leq \alpha
 \end{aligned} \tag{6-30}$$

La varianza es una función de la probabilidad apriori establecida al intervalo que contiene  $\theta$  y la distancia entre  $\theta_0$  y un percentil de la distribución. Si  $\theta$  es reemplazado por uno de los valores conocidos  $\theta_1$  ó  $\theta_2$  en la ecuación 6-30, se puede obtener un valor aproximado para la varianza de la distribución apriori Beta (Tovar Cuevas 2012),

$$\sigma^2 \leq [\theta_1 - \theta_0]^2 \cong \frac{ab}{(a+b)^2(a+b+1)} \tag{6-31}$$

Resolviendo el sistema de las dos ecuaciones 6-29 y 6-31, se encuentran los valores de  $a$  y  $b$  que son los hiperparámetros de la distribución apriori Beta (Tovar Cuevas 2012).

$$\begin{aligned}
 w &= \frac{\theta_0}{1 - \theta_0} \\
 a &= wb \\
 b &= \frac{w - [(w+1)^2\sigma^2]}{(w^3 + 3w^2 + 3w + 1)\sigma^2}
 \end{aligned}$$

### 6.2.19. Distribución Posterior

La distribución posterior es la descripción de las creencias del investigador tras la observación de los datos. Una vez conocidos los hiperparámetros de las distribuciones de probabilidad que describen el comportamiento de la cantidad aleatoria de interés (antes de observar los datos), se obtiene la forma analítica y gráfica de la distribución de probabilidades actualizada por la información contenida en los datos, a esto se le denomina distribución posterior (Tovar Cuevas 2015).



Utilizar el **Principio de Laplace** permite obtener una distribución posterior Beta, que es la distribución previa conjugada para una verosimilitud binomial y queda expresada como  $(\theta | x) \sim \text{Beta}(x + 1, n - x + 1)$

$$g(\theta | x) = \frac{\Gamma(n + 2)}{\Gamma(x + 1) \Gamma(n - x + 1)} \theta^{x+1-1} (1 - \theta)^{n-x+1-1} \text{ con } x > 0, n > 0 \quad (6-32)$$

En el caso de elicitar la distribución **apriori de Jeffreys** para un modelo Binomial, que tiene distribución  $\text{Beta}(\frac{1}{2}, \frac{1}{2})$  como se indica en 6-28, la distribución posterior será  $(\theta | x) \sim \text{Beta}(x + \frac{1}{2}, n - x + \frac{1}{2})$

$$g(\theta | x) = \frac{\Gamma(1 + n)}{\Gamma(x + \frac{1}{2}) \Gamma(n - x + \frac{1}{2})} \theta^{x+\frac{1}{2}-1} (1 - \theta)^{n-x+\frac{1}{2}-1} \text{ con } \alpha, \beta > 0 \quad (6-33)$$

Utilizando como **apriori informativa la distribución Beta que es conjugada para la distribución Binomial**, se actualizan las creencias sobre  $\theta$  aplicando el teorema de Bayes, así se obtiene la distribución posterior que es proporcional a una distribución Beta con parámetros  $x + \alpha$  y  $n - x + \beta$

$$\begin{aligned} g(\theta|x) &\propto \theta^x (1 - \theta)^{n-x} * \text{theta}^{\alpha-1} (1 - \theta)^{\beta-1} \\ &\propto \text{theta}^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1} \end{aligned}$$

$$\theta|x \sim \text{Beta}(x + \alpha, n - x + \beta)$$

$$g(\theta|x, \alpha, \beta) = \frac{\Gamma(n + \alpha + \beta)}{\Gamma(x + \alpha) \Gamma(n - x + \beta)} \theta^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1} \text{ con } \alpha, \beta > 0 \quad (6-34)$$

La media, moda y varianza de la distribución posterior están dados por

$$\begin{aligned} E(\theta) &= \frac{x + \alpha}{n + \alpha + \beta} \\ M_o(\theta) &= \frac{x + \alpha - 1}{n + \alpha + \beta - 2} \\ Var(\theta) &= \frac{(x + \alpha)(n - x + \beta)}{(n + \alpha + \beta)^2 (n + \alpha + \beta + 1)} \end{aligned}$$

### 6.2.20. Estimación Puntual

Dada una distribución sobre un parámetro particular  $\theta$ , se requiere seleccionar un mecanismo para escoger un “buen” estimador  $\hat{\theta}$  (Correa Morales et al. 2018). Dado que generalmente no se tiene una respuesta totalmente cierta en el problema de estimación, se debe de alguna forma medir el error que se comete en la estimación. Luego, se debe seleccionar un procedimiento o función de decisión, que describe la forma de tomar una decisión una vez un resultado muestral ha sido obtenido.

Por tanto, se definen las siguientes Funciones de pérdida:

- **Función de pérdida cuadrática:** La media de la distribución posterior es el estimador bayesiano bajo una función de pérdida cuadrática.
- **Función de pérdida lineal:** La mediana es el estimador bayesiano cuando la función de pérdida es el valor absoluto.
- **Función de pérdida uno a uno:** La moda es el estimador bayesiano cuando la función de pérdida es uno a uno o escalonada. Este estimador también es llamado estimador máximo-aposteriori (MAP).

### 6.2.21. Estimación por regiones

Una región de credibilidad del  $(1 - \alpha)\%$  es un intervalo formado por dos valores  $\theta(t_1 \text{ y } t_2)$  tales que

$$RC_{(1-\alpha)\%} = \int_{t_1}^{t_2} f(\theta | x) d\theta = (1 - \alpha)$$

donde,  $f(\theta | x)$  es la distribución posterior

$(1 - \alpha)$  es el porcentaje de los valores de  $\theta$  posibles.

## 7. Metodología

Para el alcance de los objetivos propuestos se plantea una metodología que está dividida en cuatro etapas, la primera orientada a la conformación y consolidación de la base de datos, la segunda corresponde al análisis exploratorio con un análisis bivariado de los datos; la tercera, al planteamiento de un modelo estadístico para valorar el efecto de la aproximación con la estadística en la educación media sobre el rendimiento académico en el curso de estadística en la universidad y finalmente, comparaciones entre grupos de estudiantes de acuerdo con el área de conocimiento en el que se encuentran matriculados.

### 7.1. Fuentes de Información

#### 7.1.1. Participantes

Estudiantes de una universidad privada de la ciudad de Cali, pertenecientes a 23 programas académicos de seis facultades, 148 estudiantes de Ciencias Administrativas y Económicas, 8 de Ciencias de la Educación, 67 de Ciencias de la Salud, 43 de Ciencias Naturales, 41 de Derecho y Ciencias Sociales y 98 estudiantes de Ingeniería. Todos matriculados en el primer curso de Estadística ofrecido dentro de sus programas de pregrado, que fueron informados de los objetivos del estudio y aceptaron diligenciar el instrumento de manera voluntaria e individual. En la tabla **7-1** se muestran los cursos y el número de estudiantes matriculados en los mismos.

**Tabla 7-1.:** Estudiantes matriculados por curso de Estadística.

| <b>Curso</b>                                  | <b>No. Estudiantes</b> |
|-----------------------------------------------|------------------------|
| Bioestadística                                | 63                     |
| Bioestadística y Fundamentos de Epidemiología | 67                     |
| Estadística y Probabilidad Básica             | 43                     |
| Fundamentos de Probabilidad y Estadística     | 11                     |
| Teoría de Probabilidades                      | 221                    |
| <b>Total</b>                                  | <b>405</b>             |

### 7.1.2. Muestra

La muestra estuvo conformada por 405 estudiantes. Respondieron de manera individual una encuesta en forma escrita que se aplicó entre la primera y segunda semana después de iniciado el semestre.

### 7.1.3. Instrumento de medición

Corresponde a una encuesta que tuvo origen en el año 2007, posteriormente revisada, validada y modificada con base en los resultados obtenidos en otros estudios y las respuestas de los encuestados. En esta ocasión, la encuesta busca obtener información acerca de las expectativas de los estudiantes frente al curso que van a iniciar y también sobre sus conocimientos en la materia y su experiencia frente a ese tipo de conocimiento.

La encuesta aplicada está conformada por 22 ítems. Los seis primeros recogen información sobre edad, género y código de estudiante, además de la carrera, materia y grupo matriculado. En adelante, son 16 preguntas cerradas. Seis preguntas indagan sobre la experiencia con matemáticas y estadística que ha tenido el estudiante en los diferentes niveles de formación por los que ha pasado, opinión sobre el grado de dificultad en los temas vistos y recoge información sobre el estamento del colegio del que se graduó de bachillerato. Las siguientes tres preguntas capturan opiniones sobre la necesidad e importancia de los cursos de

estadística y la forma de estudio o dedicación para los mismos. Cinco preguntas sobre ideas, percepciones e imaginarios respecto al curso de estadística y el grado de dificultad que espera del mismo. También se captura información sobre condiciones laborales con dos preguntas. En el Anexo A aparece una copia del instrumento aplicado para obtener los datos en campo.

#### **7.1.4. Consolidación del archivo de datos**

Los datos recolectados con el cuestionario aplicado al iniciar el segundo semestre del año 2017 fueron consolidados en un archivo de datos completando 405 registros y se construyó un diccionario de datos que describe las características de las variables que son objeto de estudio y la codificación asignada para el registro de los datos. Ver en la sección de Anexos.

#### **Depuración del archivo de datos**

En el diseño de la encuesta, en la pregunta No. 2 sobre experiencia y desempeño en los cursos de estadística en diferentes niveles de formación, en el caso en que la respuesta del encuestado fuera afirmativa sobre tener experiencia, debía calificar su desempeño. Si respondió afirmativamente sobre la experiencia pero no calificó su desempeño, se registró como “No dato” con código 8888. Si por el contrario, respondió negativamente a la experiencia y sin embargo marcó alguna de las calificaciones sobre el desempeño, la respuesta se registró como “No aplica” con código 99. Esta medida también se aplicó para la pregunta No. 4.

Para la pregunta No. 3 que dice:

*Los conceptos de estadística que estudió en el bachillerato los tomó:*

- a) Dentro de un curso de solo estadística*
- b) Como parte del curso de matemáticas*
- c) Dentro de otros cursos de ciencias (Química, Física, etc.)*

No se especificó claramente si la pregunta debía tener una o más respuestas, considerando las opciones de respuesta, es válido que se haya presentado mas de una de las situaciones

planteadas, por tanto, en la tabulación se registró cada opción como un campo con respuesta binaria. Se registraron de manera similar las respuestas para las preguntas No. 5, 11 y 15, que si fueron diseñadas con respuesta múltiple.

En la pregunta No. 5 se podía marcar varias opciones de las cinco posibles, siempre que en la pregunta 4 hubiese indicado que alguno de los temas tuvo grado de dificultad “Difícil” o “Muy difícil”. Sin embargo, hubo estudiantes que no les pareció que los temas vistos clasificaran como “Difícil” o “Muy difícil” y sin embargo, respondieron tener alguna razón por la cual decir que fue difícil algún tema que recordó haber visto en el colegio. En este caso, estas respuestas se registraron como “No aplica”; si la pregunta si aplicaba, es decir, respondió que el tema que recordado le fue “Difícil” o “Muy difícil”, pero no contestó ninguna opción en la pregunta No. 5, entonces se registró cada razón como “No dato”.

Para la pregunta 8 en la que se indicó que se debía marcar una respuesta y se presentaron casos en que se encontraron dos opciones marcadas, se determinó seleccionar al azar una de las respuestas para ser registradas.

El campo código es de gran importancia como dato de enlace, ya que se incluyen en la base de datos, otros campos que almacenan información respecto al desempeño académico del estudiante en el curso de estadística, como son las calificaciones.

## **7.2. Análisis Exploratorio de los Datos**

Con el análisis exploratorio se hizo un acercamiento a los datos para caracterizar la muestra objeto de estudio con variables como sexo, edad, estamento del colegio en el que cada estudiante obtuvo el grado de bachiller y facultad en la que está matriculado en la universidad. Se determina el nivel de experiencia que tuvo el estudiante durante su educación media previo a su ingreso a la universidad y en qué nivel de escolaridad la obtuvo, además la relevancia, es decir, si los conceptos adquiridos los vio en curso específico de estadística o como parte de otros cursos de matemáticas o ciencias.

Con el análisis bivariado se buscó establecer las posibles relaciones entre los factores propios de la educación media y realizar comparaciones entre los programas de pregrado de diferentes áreas de conocimiento (facultades). Se realizó un análisis por tipo de colegio para la percepción frente al desempeño en matemáticas y estadística, temas vistos y grados de dificultad referidos. Las variables mencionadas también se cruzan con el resultado obtenido en el curso de estadística cursado en la Universidad en el periodo de estudio.

### **Eliminación de registros**

Para el análisis de datos, fue necesario eliminar algunos registros por las razones que se mencionan a continuación:

- 21 registros no contaban con la calificación final obtenida en el curso de estadística.
- 16 registros que corresponden a estudiantes que no clasificaron para la construcción del índice de ansiedad por respuestas faltantes en el cuestionario.

Con base en lo anterior, se contó finalmente con 368 registros efectivos para las siguientes etapas de análisis.

## **7.3. Obtención de la variable de respuesta**

Para obtener la variable de respuesta, se tomó como referencia los indicadores de rendimiento académico, que para el caso corresponden a las calificaciones reportadas por los profesores de cada curso y corresponden a:

- Primer examen parcial
- Segundo examen parcial
- Nota final de cada estudiante en el curso de estadística.

Para el tratamiento de los datos registrados sobre el rendimiento académico, específicamente para la nota final del curso, se tiene en cuenta la escala de calificaciones establecida

por la universidad para obtener una variable de respuesta de estructura binaria con dos posibles valores, clasificando los registros en las categorías de “Aprobación” y “No Aprobación” del curso, es decir, las calificaciones iguales o superiores a tres punto cero (3.0) serán consideradas como aprobatorias y codificadas con (0) y las calificaciones inferiores a tres punto cero (3.0) serán consideradas como reprobatorias y codificadas con (1).

Sea  $X$  la calificación final obtenida por cada estudiante en el curso de Estadística. Donde  $X \in [0, 5]$ . Luego, se puede construir la variable aleatoria  $Y$  tal que:

$$Y = \begin{cases} 1 & \text{si } X < 3 \\ 0 & \text{si } X \geq 3 \end{cases} \quad (7-1)$$

Cabe mencionar que fueron pocos los casos de éxito (curso reprobado) encontrados, pues solamente 25 estudiantes reprobaron el curso (6.8 %).

## 7.4. Obtención de las variables explicativas

Para obtener las variables explicativas, se consideraron los ítems o preguntas que pueden tener alguna relación con la variable de respuesta. Se omitió la utilización de algunas variables con ayuda de un experto, por considerar que no tenían relación de forma coherente con la variable respuesta, de acuerdo con los objetivos de estudio, quedando 27 covariables. Algunas variables fueron recategorizadas y dicotomizadas para facilitar el procedimiento de ajuste del modelo de regresión binaria.

Las covariables consideradas son:

- Facultad
- Carrera
- Materia o curso de estadística matriculado
- Semana de avance del semestre



- Edad: 0-(17-22 años) y 1-(23-28 años)
- Género: 0-Femenino y 1-Masculino
- Desempeño en matemáticas en primaria, bachillerato y universidad: 1-Bueno (Excelente, Muy bueno y bueno), 0-No Bueno (Regular y Malo)
- Experiencia en cursos de estadística en Primaria, Bachillerato y Universidad: 1-Si y 0-No
- Conceptos vistos dentro de un curso de Estadística: 1-Si y 0-No
- Conceptos vistos como parte del curso de matemáticas: 1-Si y 0-No
- Conceptos vistos dentro de otros cursos de ciencias: 1-Si y 0-No
- Temas de estadística vistos en el colegio: Para cada uno de los siete temas indicados (Tablas de frecuencias, Gráficos de barras e Histogramas, Media o promedio, mediana, moda, Rango, varianza, desviación estándar, Permutaciones, Combinaciones, Reglas de probabilidad) 1-Si y 0-No
- Temas de estadística vistos en el colegio: Se construye a partir de la pregunta No. 4, registrando 1-Si (si marca cuatro o más temas) y 0-No (si marca menos de cuatro temas)
- Tipo de colegio del que se graduó de bachillerato: 1-Oficial y 0-Privado.
- Considera como necesarios e importantes los cursos de estadística incluidos en la carrera 1-Si y 0-No
- Trabajo remunerado 1-Si y 0-No
- Índice de ansiedad

### 7.4.1. Construcción de Índice de Ansiedad

Como parte de las variables explicativas se construyó una variable a partir de la propuesta de Índice de Ansiedad hecha por Osorno (2015), usando una aproximación basada en la Heurística. Para esto se tomaron tres preguntas de la encuesta que conforman los factores determinantes del nivel de ansiedad y los valores según las respuestas consignadas en los cuestionarios.

Las preguntas del cuestionario, consideradas para evaluar el nivel de ansiedad que presenta el estudiante frente al curso que va a iniciar son:

14. *¿Cómo espera que sea el grado de dificultad del curso de estadística que matriculó?:*  
a.-Muy alto b.-Alto c.-Bajo d.-Muy bajo
15. *¿Cuáles de los siguientes enunciados, podrían ser una razón que explique su respuesta a la pregunta anterior?:*  
a) *Tengo buenos conocimientos en matemáticas*  
b) *Estoy interesado y tengo una buena actitud hacia la carrera*  
c) *Tengo confianza porque tuve buenos resultados en los cursos previos*  
d) *Me parecen difíciles las materias que utilizan las matemáticas*  
e) *No tengo el suficiente tiempo para dedicarle a la materia*  
f) *No he visto nada de estadística anteriormente*
16. *¿Cómo se siente ahora frente al curso de estadística que va a comenzar? Marque con una X el recuadro que más identifique su respuesta*

| <b>Preocupado:</b>                                                     | <b>Tenso:</b>                                                            | <b>Tranquilo:</b>                                                   | <b>Muy tranquilo:</b>                                  |
|------------------------------------------------------------------------|--------------------------------------------------------------------------|---------------------------------------------------------------------|--------------------------------------------------------|
| Porque va a ser un curso muy difícil y es muy probable que lo repruebe | Porque tendrá que esforzarse mucho para poder aprobar con la nota mínima | Porque cuenta con las herramientas necesarias para aprobar el curso | Porque es un curso que no le demandará mayor esfuerzo. |

Las valoraciones establecidas para la pregunta No. 15 se colocaron de manera heurística apoyándose en la opinión de una psicóloga experta en educación.

A la pregunta No. 14 (grado de dificultad del curso de estadística) se le asigna una escala de 1 a 2 y se codifica en una nueva variable denominada **Factor1**:

$$BAJO \text{ o } MUY \text{ BAJO (respuesta c o d)} \rightarrow BAJO \rightarrow 1$$

$$ALTO \text{ o } MUY \text{ ALTO (respuesta a o b)} \rightarrow ALTO \rightarrow 2$$

La pregunta No. 15 sobre las razones que explican el grado de dificultad, es cerrada y tiene múltiples respuestas. Las opciones a, b y c son razones positivas (+) para que el estudiante explique su percepción de grado de dificultad “Bajo o Muy Bajo”. Las opciones d, e y f son razones negativas (-) para explicar que el curso de estadística va a tener un grado de dificultad “Alto o Muy Alto”. Por tanto, las razones positivas serán equivalentes a índices de ansiedad bajo y las razones negativas equivalentes a índices de ansiedad alto. A las posibles combinaciones de respuestas se asigna un valor en una escala de 1 a 7 relacionando, 1 con un índice de ansiedad bajo y 7 con un índice de ansiedad alto. En la tabla 7-2 se presenta la asignación de los valores de la escala para la pregunta No. 15 como resultado de las posibles combinaciones de respuestas que se registran en la nueva variable **Factor2**.

**Tabla 7-2.:** Escala de la pregunta 15 según combinaciones de respuesta

| Razón de dificultad | Combinaciones de respuestas                          | Valor asignado |
|---------------------|------------------------------------------------------|----------------|
| +                   | a, b, c                                              | 3              |
| ++                  | ab, ac, bc                                           | 2              |
| +++                 | abc                                                  | 1              |
| + -                 | ad, bd, cd, ae, be, ce, af, bf, cf                   | 4              |
| ++ -                | abd, acd, bcd, abe, ace, bce, abf, acf, bcf          | 4              |
| +++ -               | abcd, abce, abcf                                     | 4              |
| + - -               | ade, bde, cde, adf, bdf, cdf, aef, bef, cef          | 4              |
| ++ - -              | abde, acde, bcde, abdf, acdf, bcdf, abef, acef, bcef | 4              |

**Tabla 7-2.:** Escala de la pregunta 15 según combinaciones de respuesta

| Razón de dificultad | Combinaciones de respuestas | Valor asignado |
|---------------------|-----------------------------|----------------|
| +++—                | abcde, abcdf, abcef         | 4              |
| +—                  | ade, bde, cde               | 4              |
| ++—                 | abde, acde, bcde            | 4              |
| +++—                | abcde                       | 4              |
| -                   | d, e, f                     | 5              |
| —                   | de, df, ef                  | 6              |
| —                   | def                         | 7              |

A la pregunta No. 16 (sentimiento frente al curso a comenzar) se le asigna una escala de 1 a 2 y se codifica en una nueva variable denominada **Factor3** relacionando a 1 con un índice de ansiedad bajo y 2 con un índice de ansiedad alto:

$$\text{Tranquilo o muy Tranquilo} \rightarrow \text{Tranquilo} \rightarrow 1$$

$$\text{Preocupado o Tenso} \rightarrow \text{Tenso} \rightarrow 2$$

Luego, se obtiene el Índice de Ansiedad 7-2 con la sumatoria de los valores obtenidos en los Factores 1, 2 y 3.

$$\text{Índice de Ansiedad (Z)} = \sum_{i=1}^3 \text{Factor}_i \quad (7-2)$$

Esta nueva variable toma valores entre 3 y 11, siendo 3 el más bajo nivel de ansiedad (1 en la pregunta 14, 1 en la pregunta 15 y 1 de la pregunta 16) y 11 el más alto nivel de ansiedad (2 en la pregunta 14, 7 en la pregunta 15 y 2 en la pregunta 16).

Estos valores del Índice de Ansiedad se transforman en una variable dicotómica. Para valores menores o iguales a 7 se considera que hay Ansiedad Manejable. Para valores entre 8 y 11

se considera que hay Ansiedad Importante, como se indica en 7-3.

$$Z = \begin{cases} 0 & \text{si } 3 \leq Factor_i \leq 7 \quad \text{Ansiedad Manejable (AM)} \\ 1 & \text{si } 8 \leq Factor_i \leq 11 \quad \text{Ansiedad Importante (AI)} \end{cases} \quad (7-3)$$

#### 7.4.2. Pruebas de Asociación

Con base en la lista de variables mencionadas, para las variables categóricas se construyeron arreglos en tablas de contingencia para establecer asociaciones utilizando la prueba Chi-Cuadrado y la prueba exacta de Fisher para los casos en que la prueba Chi-Cuadrado no se pudo realizar, asumiendo un nivel de significancia del 5% ( $\alpha=0.05$ ); lo anterior, para evaluar posibles relaciones existentes entre los ítems o preguntas seleccionadas como variables explicativas y la variable de respuesta (resultado en el curso de estadística), así identificar algunos factores que pueden incidir en el riesgo de reprobar el curso de estadística.

Para ésta y las siguientes etapas de análisis, la proporción de éxito se modifica en razón a la presencia de registros y campos con datos faltantes que no son tenidos en cuenta para la propuesta de los modelos, trabajando con un total de 237 registros.

### 7.5. Propuesta de Modelo

Una vez establecidas las variables indicadoras del rendimiento académico que determinan la variable de respuesta, dado que se busca estimar la proporción de estudiantes que reprueban el curso de estadística, se plantea de un modelo que permita establecer el efecto de las variables asociadas con la experiencia con estadística en el colegio que inciden sobre el desempeño académico de los estudiantes en el primer curso del área en la universidad, con las covariables que resultan estadísticamente significativas en la etapa de obtención de las variables explicativas.

### 7.5.1. Experimento Estadístico

Seleccionar un estudiante matriculado en un curso de estadística, aplicarle la encuesta, tomar el registro de sus calificaciones obtenidas y observar si reprobó o no el curso.

### 7.5.2. Modelo Estadístico

Sea el **evento de interés**  $A$  tal que:

$$A = \{\text{El estudiante reprobó el curso de estadística}\}$$

Sea  $Y$  la **variable aleatoria** que define el evento  $A$  tal que:

$$Y = \begin{cases} 1 & \text{si ocurre } A \\ 0 & \text{si ocurre } A' \end{cases} \quad (7-4)$$

Luego,  $Y \sim \text{Bernoulli}(\theta)$

Donde  $\theta = P(Y=1)$ , que es la probabilidad de que el estudiante repruebe el curso.

Dado que  $Y$  es binaria, es posible asociar que el comportamiento natural de la variable  $Y$  puede ser modelado usando la distribución de Bernoulli de probabilidades, es decir:

$$P(Y = y) = \theta^y (1 - \theta)^{1-y} \quad \forall y \in \{0, 1\} \quad \theta \in (0, 1) \quad (7-5)$$

Sea también  $X_i$  la variable que define la presencia de las características, tal que:

$$X_i = \begin{cases} 1 & \text{si el estudiante tiene la característica} \\ 0 & \text{si el estudiante no tiene la característica} \end{cases} \quad (7-6)$$

Donde  $i = 1, \dots, k$ . Siendo  $k$  el número de covariables incluídas en el modelo.

Debido a que en la muestra estudiada, se observó una proporción pequeña de estudiantes que reprobaron el curso (16/237 el 6,8 %), se ajustó un modelo de regresión binaria con enlace log-log complementario, de acuerdo con Guzmán et al. (2014), por ser un procedimiento útil para obtener información sobre asociaciones entre variables con respuesta binaria a nivel multivariado en casos en que no es posible asumir una distribución logística simétrica para

el comportamiento de las probabilidades de éxito en la población de interés.

Luego, se hace una reparametrización de  $\theta$  utilizando la función de enlace Log-Log Complementario:

$$\begin{aligned} g(\theta) &= \log(-\log(1 - P(Y = 1 \mid X_1, \dots, X_4))) \\ g(\theta) &= \log(-\log(1 - \theta)) \end{aligned}$$

El modelo final queda especificado de la siguiente manera:

$$\begin{aligned} \log(-\log(1 - \theta)) &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \\ g(\theta) &= \beta_0 + \sum_{i=1}^4 \beta_i X_i \end{aligned} \quad (7-7)$$

Y aplicando exponencial a la ecuación 7-7, finalmente se obtiene

$$\begin{aligned} e^{\log(-\log(1-\theta))} &= e^{\beta_0 + \sum_{i=1}^4 \beta_i X_i} \\ -\log(1 - \theta) &= e^{\beta_0 + \sum_{i=1}^4 \beta_i X_i} \\ \log(1 - \theta) &= -e^{\beta_0 + \sum_{i=1}^4 \beta_i X_i} \\ \widehat{(1 - \theta)} &= e^{-e^{\beta_0 + \sum_{i=1}^4 \beta_i X_i}} \end{aligned} \quad (7-8)$$

Donde  $\widehat{(1 - \theta)}$  equivale a  $P(Y = 1)$ .

El parámetro de regresión  $\hat{\beta}$  se calcula a través del método iterativo de Newton-Raphson. Al ajustar el modelo con la función de enlace Log-Log Complementario es necesario tener presente que toma como referente la asimetría que presenta la variable de respuesta, y que modela la probabilidad de no éxito, es decir, la probabilidad de que el estudiante haya aprobado el curso, por lo que se realiza una recategorización de la variable de respuesta, de forma que:

Cuando  $Y$  toma el valor de 0, la probabilidad es  $P(Y = 0) = \theta$  y

Cuando  $Y$  toma el valor de 1, la probabilidad es  $P(Y = 1) = 1 - \theta$

Por lo tanto,

$\theta$  = Probabilidad de que estudiante apruebe el curso.

$1 - \theta$  = Probabilidad de que estudiante repruebe el curso.

Se realizó el proceso para obtención de dos modelos:

**Modelo 1:** incluyendo las variables que resultaron significativas con las pruebas de asociación, obteniendo

$$\widehat{(1 - \theta)} = e^{-e^{\beta_0 + \sum_{i=1}^2 \beta_i X_i}}$$

**Modelo 2:** construido a partir del modelo original incluyendo todas las variables consideradas para análisis aplicando la técnica de Stepwise, que resulta en:

$$\widehat{(1 - \theta)} = e^{-e^{\beta_0 + \sum_{i=1}^4 \beta_i X_i}}$$

Como criterio de decisión se tuvo en cuenta el Criterio de Información de Akaike (AIC), por tanto, se escogió el modelo con AIC más pequeño y con este modelo se hizo la aproximación bayesiana.

### 7.5.3. Evaluación del modelo

Para evaluar la capacidad clasificatoria del modelo de respuesta binaria con función de enlace Log-Log Complementario, se obtuvo una muestra de validación, ya que la cantidad de estudiantes que reprobaron el curso de estadística fue de 16, es pequeña en comparación con los estudiantes que aprobaron el curso, 221 estudiantes. Por tanto, para obtener la muestra de validación se tomó una muestra sin reemplazamiento del total de individuos que aprobaron el curso de tamaño y se suma a ésta la cantidad de estudiantes que reprobaron



el curso, garantizando la representatividad y asimetría en la distribución de la variable de interés.

Posteriormente se estableció un punto de corte a través de la curva ROC no paramétrica, utilizando la medida de exactitud Índice de Youden para las probabilidades de pronóstico, ya que no se podía asumir un punto de corte igual a 0,5 como sucede cuando se utiliza una función de enlace Logit o Probit. En el caso de estudio que se trata de un modelo asimétrico, se debe buscar un punto de corte a partir del cual, la probabilidad sea lo suficientemente alta para predecir a un nuevo individuo que va a reprobar el curso de estadística. El punto de corte obtenido fue de 0,65.

Con base en este punto de corte, se categorizan las probabilidades obtenidas por el modelo, construyendo una tabla cruzada como se indica en la tabla **6-2**, con el fin de evaluar la capacidad clasificatoria del modelo. Para poder formar la tabla de clasificación se aplicó la técnica de remuestreo de Bootstrap, remuestreando 1000 veces el procedimiento.

Los componentes de la tabla son:

- a: Estudiantes que reprobaron el curso con predicción de haber reprobado el curso.
- b: Estudiantes que aprobaron el curso con predicción de haber reprobado el curso.
- c: Estudiantes que reprobaron el curso sin predicción de haber reprobado el curso.
- d: Estudiantes que aprobaron el curso sin predicción de haber reprobado el curso.

Y se procedió al calcular:

- Correctos Clasificados como Reprobados (CCR): probabilidad de predecir éxito (Reprobar el curso de estadística) dado que el estudiante realmente reprobó el curso.
- Correctos Clasificados como No Reprobados (CCNR): probabilidad de predecir no éxito (Aprobar el curso de estadística) dado que el estudiante realmente aprobó el curso.

- Falsos Reprobados (FR): probabilidad de predecir éxito (Reprobar el curso de estadística) dado que el estudiante realmente aprobó el curso.
- Falsos No Reprobados (FNR): probabilidad de predecir no éxito (Aprobar el curso de estadística) dado que el estudiante realmente reprobó el curso.
- Porcentaje correcto de clasificación (P.C.C): probabilidad de acertar en la predicción para todos los estudiantes.

## 7.6. Aproximación Bayesiana

En esta instancia se plantea un modelo estadístico asociado al problema de estudio propuesto en la sección 7.5, tomando en cuenta la información de cuatro covariables con respuesta binaria que resultaron ser significativas y aportan información acerca de su educación previa:

$$\begin{aligned}
 X_1 &= \begin{cases} 0 ; \text{ si el desempeño del estudiante en cursos de matemáticas en la Universidad fue bueno} \\ 1 ; \text{ si el desempeño del estudiante en cursos de matemáticas en la Universidad no fue bueno} \end{cases} \\
 X_2 &= \begin{cases} 0 ; \text{ Si el estudiante no tuvo experiencia en cursos de estadística en Bachillerato} \\ 1 ; \text{ Si el estudiante tuvo experiencia en cursos de estadística en Bachillerato} \end{cases} \\
 X_3 &= \begin{cases} 0 ; \text{ Si el estudiante no vio conceptos de estadística como parte del curso de matemáticas} \\ 1 ; \text{ Si el estudiante vio conceptos de estadística como parte del curso de matemáticas} \end{cases} \\
 X_4 &= \begin{cases} 0 ; \text{ Si el estudiante tenía menos de 4 temas de estadística vistos} \\ 1 ; \text{ Si el estudiante tenía 4 o mas temas de estadística vistos} \end{cases}
 \end{aligned}$$

Se quiere establecer la combinación de covariables que estimen la mayor proporción de estudiantes que reprueban el curso de estadística, por tanto, se estima  $\theta_j = P(Y = 1|x_1, \dots, x_4), \forall j = 1, \dots, 16$ . Para esto se conforman grupos binomiales con las  $2^4 = 16$  posibles combinaciones de opciones de respuesta de las covariables, como se muestran en la tabla **7-3**.

### 7.6.1. Modelo Estadístico

Se define el **evento**  $B$  tal que:

**Tabla 7-3.:** Conformación de grupos.

| Combinaciones | Covariables |       |       |       |
|---------------|-------------|-------|-------|-------|
|               | $X_1$       | $X_2$ | $X_3$ | $X_4$ |
| 1             | 0           | 0     | 0     | 0     |
| 2             | 1           | 0     | 0     | 0     |
| 3             | 0           | 1     | 0     | 0     |
| 4             | 0           | 0     | 1     | 0     |
| 5             | 0           | 0     | 0     | 1     |
| 6             | 1           | 1     | 0     | 0     |
| 7             | 1           | 0     | 1     | 0     |
| 8             | 1           | 0     | 0     | 1     |
| 9             | 0           | 1     | 1     | 0     |
| 10            | 0           | 1     | 0     | 1     |
| 11            | 0           | 0     | 1     | 1     |
| 12            | 1           | 1     | 1     | 0     |
| 13            | 1           | 0     | 1     | 1     |
| 14            | 1           | 1     | 0     | 1     |
| 15            | 0           | 1     | 1     | 1     |
| 16            | 1           | 1     | 1     | 1     |

$$B = \{\text{El estudiante del } j\text{-ésimo grupo reprobó el curso de estadística}\}$$

Sea  $Y_{ij}$  la **variable aleatoria** que define el evento  $B$  tal que:

$$Y_{ij} = \begin{cases} 1 & \text{si ocurre } B \\ 0 & \text{si ocurre } B' \end{cases} \quad (7-9)$$

Donde  $i = 1, 2, \dots, n_j$   $j = 1, 2, \dots, 16$  grupos.

En consecuencia,  $Y_{ij} \sim \text{Bernoulli}(\theta_j)$

Donde  $\theta_j = P(Y_{ij} = 1 | x_1, \dots, x_4)$ , que es la probabilidad de que el estudiante del  $j$ -ésimo grupo repuebe el curso, dada la combinación de respuestas a las cuatro covariables.

$$P(Y_{ij} = y_{ij}) = \theta_j^{y_{ij}} (1 - \theta_j)^{1-y_{ij}} \quad \forall y_{ij} \in \{0, 1\} \quad \theta_j \in (0, 1) \quad (7-10)$$

Una vez definidos los grupos con las  $j$  combinaciones de las categorías de cada covariable,

se define la variable  $G_j$  que hace referencia al número de estudiantes que reprueban el curso de estadística dentro del grupo  $j$ .

$$G_j = \sum_{i=1}^{n_j} Y_j \sim \text{Binomial}(n_j, \theta_j)$$

donde,

$n_j$  = Al total de estudiantes en el grupo  $j$

$\theta_j$  = La probabilidad de que un estudiante repruebe el curso de estadística si tiene la  $j$ -ésima combinación de las covariables.

La función de masa asociada es

$$P(G_j = g_j | \theta_j) = \binom{n_j}{g_j} \theta_j^{g_j} (1 - \theta_j)^{n_j - g_j}, \quad g_j \in \{0, 1, 2, \dots, n_j\} \quad \theta_j \in (0, 1)$$

Y la función de máxima verosimilitud está dada por

$$L(G_j = g_j | \theta_j) = \binom{n_j}{g_j} \theta_j^{g_j} (1 - \theta_j)^{n_j - g_j}$$

La modelación se realiza por los grupos definidos según la tabla **7-3**.

### Máxima Verosimilitud

Como estimación clásica se utilizó el método de máxima verosimilitud, tomando como valor del parámetro aquel que maximiza la probabilidad de que ocurra la muestra observada. Para el caso se tiene una observación binomial por cada combinación, por tanto  $\mathcal{L}(\theta_j | G_j = g_j) = f(G_j = g_j | \theta_j)$ , entonces el estimador está dado por:

$$\hat{\theta}_j = \frac{g_j}{n_j}$$

El estimador  $\hat{\theta}_j$  corresponde a la proporción de estudiantes que reprueban el curso de estadística en el grupo  $j$ . Luego, se realizó la estimación bayesiana, utilizando distribuciones de probabilidad a priori no informativas y conjugadas, como alternativa debido al bajo número de éxitos observados en los grupos que se conformaron.

### Principio de Laplace

Como no se cuenta con información sobre el comportamiento de la cantidad aleatoria de interés  $\theta_j$  se asume a priori que todos los valores de  $\theta_j$  tienen la misma probabilidad de ocurrir en el espacio  $(0,1)$ , por lo tanto, se asume una distribución Uniforme  $(0,1)$ . Se encontró la distribución posterior por el principio de Laplace, con el que se define la distribución a priori de  $\theta_j$  como  $\pi(\theta_j) = \text{Beta}(1, 1)$ . La distribución posterior está dada por:

$$f(\theta_j \mid G_j = g_j) = \text{Beta}(g_j + 1, n_j - g_j + 1)$$

### Apriori Conjugada

Por el método de distribuciones a priori conjugadas se obtiene que la verosimilitud de una observación binomial está dada por  $\mathcal{L}(\theta_j \mid G_j = g_j) = f(G_j = g_j \mid \theta_j)$ . Luego, la función de máxima verosimilitud toma la forma del núcleo de la distribución beta y la distribución a priori conjugada puede ser  $\pi(\theta_j) = \text{Beta}(a, b)$  entonces la distribución posterior es:

$$f(\theta_j \mid G_j = g_j) = \text{Beta}(g_j + a, n_j - g_j + b)$$

Para encontrar los hiperparámetros  $a$  y  $b$  de la distribución a priori conjugada se hizo uso del método de elicitación propuesto por Tovar Cuevas (2012).

En la tabla 7-4 se resumen los estimadores correspondientes a la probabilidad de que un estudiante repruebe el curso de estadística en cada uno de los métodos propuestos.

**Tabla 7-4.:** Estimadores

| Método    | MV                | LaPlace                   | Jeffrey                     | Conjugada                     |
|-----------|-------------------|---------------------------|-----------------------------|-------------------------------|
| Estimador | $\frac{g_j}{n_j}$ | $\frac{g_j + 1}{n_j + 2}$ | $\frac{g_j + 1/2}{n_j + 1}$ | $\frac{g_j + a}{n_j + b + a}$ |

### Distribución predictiva

Dado que es de interés predecir la probabilidad a priori y posterior de que en caso de realizar un nuevo estudio con el mismo tamaño de muestra, en cada grupo haya por lo menos un estudiante que haya reprobado el curso de estadística una vez que se tome otra muestra, se tiene:

- Probabilidad Predictiva a priori

$$P(G_j \geq 1) = 1 - P(G_j = 0) = 1 - \frac{\Gamma(\alpha + \beta)\Gamma(n_j + \beta)}{\Gamma(\beta)\Gamma(\alpha + \beta + n_j)}$$

- Probabilidad Predictiva posterior

$$P(G_j \geq 1) = 1 - P(G_j = 0) = 1 - \frac{\Gamma(n_j + \alpha + \beta)\Gamma(m_j + n_j - g_j + \beta)}{\Gamma(n_j - g_j + \beta)\Gamma(m_j + n_j + \alpha + \beta)}$$

## 7.7. Comparación por áreas de conocimiento

Con base en los modelos ajustados y las asociaciones encontradas, se procedió a realizar comparaciones por áreas de conocimiento, que en este caso se hace referencia a seis facultades de las que hacen parte los estudiantes participantes en el estudio:

- Ciencias Administrativas y Económicas
- Ciencias de la Educación
- Ciencias de la Salud
- Ciencias Naturales
- Derecho y Ciencias Sociales
- Ingeniería.

## 8. Resultados

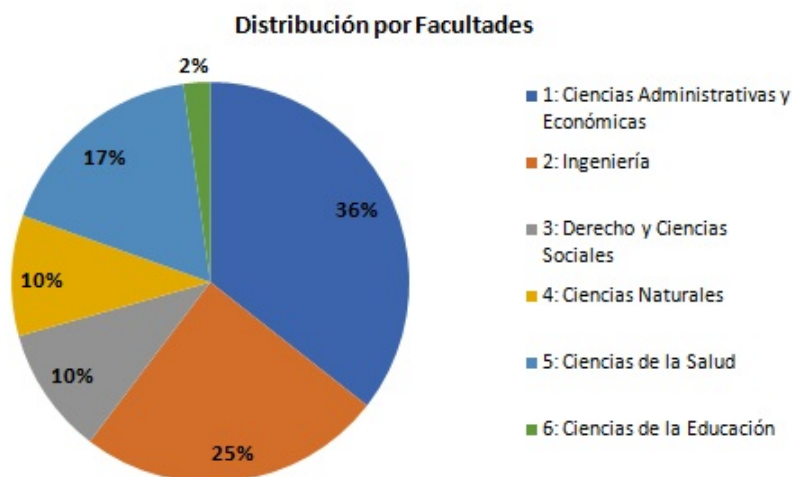
El capítulo iniciará con el análisis exploratorio de las variables de interés. Luego se obtendrán las variables explicativas, se ajustará el modelo con respuesta binaria utilizando el enlace Log-Log Complementario y se presentará la capacidad clasificatoria del modelo. Posteriormente se darán a conocer los resultados de la aproximación bayesiana y la probabilidad predictiva de Bayes, finalizando con el análisis de los resultados encontrados para las diferentes facultades a las que están adscritos los estudiantes.

### 8.1. Análisis exploratorio de los datos

Con base en los registros efectivos, se consideraron 368 estudiantes para el análisis. Entre ellos se observó una edad mínima de 17 y máxima de 28 años con un promedio de edad de 19 años, 358 (97.3 %) tienen edades entre los 17 y 22 años y 10 (2.7 %) están entre los 23 y 28 años. Entre los participantes 180 (48.9 %) son mujeres y 188 (51.1 %) son hombres.

Como se observa en la figura **8-1**, la facultad con mayor número de estudiantes participantes en el estudio fue la de Ciencias Administrativas y Económicas 131 (36 %), en segundo lugar estuvieron 91 (25 %) estudiantes de la Facultad de Ingeniería, en tercer lugar 64 (17 %) estudiantes de Ciencias de la Salud y, la Facultad con menor participación fue la de Ciencias de la Educación con 8 estudiantes que representan el 17 % de la muestra.

En cuanto al colegio de procedencia, se encontró que 229 (62.2 %) finalizaron sus estudios de bachillerato en un colegio privado y 139 (37.8 %) lo hicieron en un colegio oficial.



**Figura 8-1.:** Distribución de estudiantes por Facultades

Mientras adelantan sus estudios de pregrado solo el 19 % tienen al mismo tiempo un trabajo remunerado.

Los estudiantes en su mayoría 325 (88.3 %) manifiestan haber tenido experiencia con Estadística en el bachillerato y 279 de ellos (85.8 %) calificaron como Bueno su desempeño en los cursos de estadística vistos en este nivel de estudios. 106 (28.8 %) estudiantes afirman haber visto Estadística en educación primaria y de ellos, 94 (88.7 %) con un buen desempeño en los cursos. En cuanto a su desempeño en matemáticas 346 (94 %) reportan haber tenido un buen desempeño en esta área en primaria y 320 (87 %) indican lo mismo en bachillerato. Ya para el nivel de educación superior la proporción de buen desempeño baja a 199 (54.1 %) estudiantes.

Respecto a la forma en que tomó los conceptos de estadística vistos en bachillerato, 180 (48.9 %) afirman haberlos visto dentro de un curso de sólo estadística, 176 (47.8 %) indican que fue como parte del curso de matemáticas y 35 (9.5 %) dicen que fue dentro de otros cursos de ciencias como Química, Física, entre otros. Al indagar por conceptos específicos de estadística, se encontró que los conceptos con mayor proporción de estudiantes que los vieron

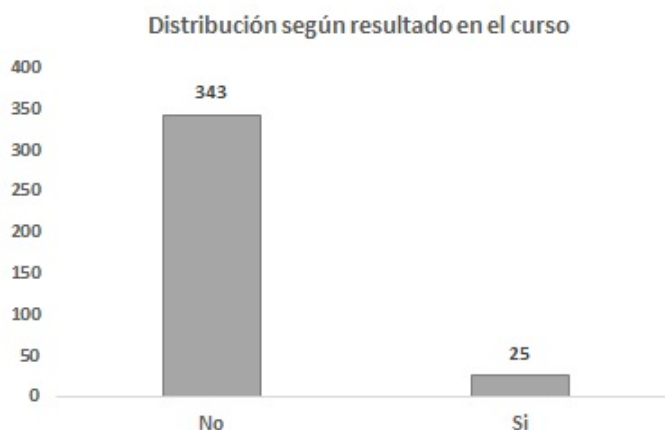


en bachillerato fueron Tablas de frecuencias 299 (81.3 %), Gráficos de barras e Histogramas 333 (90.5 %), Media o promedio, mediana y moda 348 (94.6 %) y Rango, varianza y desviación estándar 228 (62 %). Entre este grupo de conceptos, el mayor grado de dificultad fue atribuido a los temas Rango, varianza y desviación estándar con el 19.3 %. Menos del 46 % de los estudiantes afirmaron haber visto en el colegio los conceptos de Permutaciones, Combinaciones y algunas reglas de probabilidad. Entre los estudiantes que manifestaron haber tenido dificultad con algunos temas vistos en bachillerato, en mayor proporción 159 (43.2 %) consideran que la principal causa es que no le entendían la explicación al profesor, en segundo lugar 70 estudiantes (19 %) afirman que los temas incluían operaciones matemáticas muy difíciles y 41 estudiantes (11.1 %) asocian la dificultad para entender los temas porque los problemas tratados no tenían relación con la vida real.

Entre los estudiantes que tuvieron experiencia con cursos de estadística en el bachillerato, el 54.8 % adquirieron los conceptos estadísticos dentro de un curso de sólo estadística, 45.8 % dentro de un curso de matemáticas y el 62.5 % provienen de colegios privados. Se observa que en los colegios privados se imparten mas conceptos de estadística que en los colegios oficiales, 49 % dentro de cursos de sólo estadística, 47.6 % como parte de un curso de matemáticas y 9.2 % dentro de otros cursos de ciencias.

El grupo de personas incluídas en el estudio pudo ser clasificado en dos grupos, de acuerdo con el índice de ansiedad construido a partir de algunas preguntas del cuestionario. 236 (64.1 %) de los estudiantes presentan un nivel de **ansiedad manejable** y 132 (35.9 %) un nivel de **ansiedad importante**.

En cuanto a los resultados obtenidos por los estudiantes en el curso de Estadística, como se aprecia en la figura **8-2**, 25 (6.8 %) reprobaron y entre los estudiantes que aprobaron, 233 (63.3 %) obtuvieron un nivel bajo, es decir, una calificación final menor o igual a 3.7. De los estudiantes que reprobaron, todos tienen edades menores o iguales a los 22 años, el 88 % tenían algún grado de experiencia con estadística en el bachillerato, el 56 % finalizaron



**Figura 8-2.:** Distribución según resultado en el curso

el bachillerato en colegios privados, el 60 % son hombres. El 52 % cursan programas de la Facultad de Ciencias Administrativas y Económicas, el 64 % cursaron la asignatura Teoría de Probabilidades. El 16 % desarrollan algún trabajo remunerado y el 36 % presentan un nivel de ansiedad importante.

## 8.2. Obtención de variables explicativas

Se aplicó la prueba Chi-Cuadrado y la prueba exacta de Fisher con un nivel de significancia del 5 % ( $\alpha = 0,05$ ) con el fin de analizar la relación entre el rendimiento académico en el curso de estadística y otras 27 variables relacionadas con la experiencia previa con estadística. La distribución conjunta de las covariables y las variables de respuesta, así como los valores-p resultantes de las pruebas de asociación aplicadas, se pueden apreciar en el Anexo C.

Los resultados indican que las variables que presentan asociación significativa con el **resultado final en el curso de estadística** son **haber visto conceptos de estadística como parte del curso de matemáticas** y **el desempeño alcanzado en cursos de**

**matemáticas en la universidad**, puesto que los valores-p arrojados por las pruebas resultaron inferiores al nivel de significancia 0.05.

Las demás covariables analizadas, arrojaron valores superiores al nivel de significancia 0.05, por lo que sugieren que hay evidencia estadística suficiente para no rechazar la hipótesis nula de no asociación entre las variables, en otros términos, las variables de respuesta al parecer, no presentan relación alguna con las covariables.

### 8.3. Ajuste del Modelo con Respuesta Binaria

Para estimar los valores de los coeficientes se ajustó un modelo de respuesta binaria con función de enlace Log-Log Complementario con las dos variables que presentaron asociación. Con ayuda del Software estadístico R, se estimaron los coeficientes del modelo propuesto obteniendo:

$$\ln[-\ln(1 - \theta)] = -2,6762 + 0,9585X_1 - 1,3887X_3 \quad (8-1)$$

Donde  $X_1$  corresponde al desempeño en cursos de matemáticas en la Universidad y  $X_3$  a si el estudiante vio conceptos de estadística como parte del curso de matemáticas.

Con base en este modelo, se presentan los intervalos de confianza para cada parámetro incluido, los valores correspondientes al estadístico de Wald para evaluar la significancia, el Odds-Ratio para cada parámetro y su respectivo intervalo de confianza.

**Tabla 8-1.:** Estimaciones Modelo 1 Log-Log Complementario

| Parámetro  | $\hat{\beta}$ | $EE(\hat{\beta})$ | Wald   | Valor-p  | $IC_{95\%}(\hat{\beta})$ | $exp(\hat{\beta})$ | $IC_{95\%}(exp(\hat{\beta}))$ |
|------------|---------------|-------------------|--------|----------|--------------------------|--------------------|-------------------------------|
| Intercepto | -2,6762       | 0,4331            | -6,180 | 6,42e-10 | [-3,5251 ; -1,8273]      | 0,0688             | [0,0294 ; 0,1608]             |
| $X_1$      | 0,9585        | 0,518             | 1,851  | 0.0642   | [-0,0568 ; 1,9738]       | 2,6078             | [0,9448 ; 7,1978]             |
| $X_3$      | -1,3887       | 0,6428            | -2,160 | 0.0307   | [-2,6486 ; -0,1288]      | 0,2494             | [0,0708 ; 0,8791]             |

Los resultados obtenidos dejan ver que la variable  $X_1$  resulta significativa solo con un  $\alpha = 10\%$  y la variable  $X_3$  significativa con un  $\alpha = 5\%$ , se obtiene un AIC de **112.83**.

Con el fin de encontrar un buen modelo, se procedió a eliminar del modelo la variable  $X_1$  y se estimaron nuevamente los coeficientes del modelo, con lo que se obtuvo que la variable  $X_3$  sigue siendo significativa con un  $\alpha = 5\%$ , para este modelo simple se obtuvo un AIC de **114.4**, superior al modelo con las dos covariables; teniendo en cuenta que un modelo con un criterio de información de Akaike (AIC) menor presenta mejores estimaciones, se eligió en primera instancia el modelo para estimar el resultado final en el curso de estadística con las dos covariables **desempeño en cursos de matemáticas en la Universidad y conceptos de estadística vistos como parte del curso de matemáticas**.

Sin embargo, con la finalidad de encontrar el mejor modelo que pueda realizar las mejores estimaciones de las probabilidades de éxito se decidió utilizar la metodología de selección de variables StepWise, que tiene como fundamento el principio de parsimonia, que logra encontrar el mejor modelo con la menor cantidad de covariables. Se eligió para este caso el mejor modelo partiendo del modelo más complejo que es el que contempla las 27 covariables. Después de aplicar el Stepwise y discutir con el especialista sobre las variables seleccionadas se obtuvo el siguiente modelo:

$$\ln[-\ln(1 - \theta)] = -2,0712 + 1,1829X_1 - 1,8297X_2 - 1,6804X_3 + 1,2676X_4 \quad (8-2)$$

Del cual se obtienen las estimaciones que se muestran en la tabla **8-2**.

**Tabla 8-2.:** Estimaciones Modelo 2 Log-Log Complementario

| Parámetro  | $\hat{\beta}$ | $EE(\hat{\beta})$ | Wald   | Valor-p | $IC_{95\%}(\hat{\beta})$ | $exp(\hat{\beta})$ | $IC_{95\%}(exp(\hat{\beta}))$ |
|------------|---------------|-------------------|--------|---------|--------------------------|--------------------|-------------------------------|
| Intercepto | -2,0712       | 0,9160            | -2,261 | 0,0238  | [-3,8666 ; -0,2758]      | 0,1260             | [0,0209 ; 0,7589]             |
| $X_1$      | 1,1829        | 0,5307            | 2,229  | 0,0258  | [0,1427 ; 2,2231]        | 3,2638             | [1,1534 ; 9,2357]             |
| $X_2$      | -1,8297       | 0,7160            | -2,556 | 0,0106  | [-3,2331 ; -0,4263]      | 0,1605             | [0,0394 ; 0,6529]             |
| $X_3$      | -1,6804       | 0,6975            | -2,409 | 0,016   | [-3,0475 ; -0,3133]      | 0,1863             | [0,0475 ; 0,7310]             |
| $X_4$      | 1,2676        | 0,6841            | 1,853  | 0,0639  | [-0,0732 ; 2,6084]       | 3,5523             | [0,9294 ; 13,5778]            |

En este modelo encontrado por la metodología StepWise se tienen en cuenta dos covariables

adicionales a las incluídas en el primer modelo planteado, **haber tenido experiencia en cursos de Estadística en Bachillerato y haber visto cuatro o más temas de estadística en el Bachillerato**, con la particularidad que el AIC correspondiente es de **109,52**, que es un poco menor al encontrado con el modelo que incluye dos covariables, lo que indica que este es un mejor modelo. Los coeficientes de las cuatro variables resultaron ser estadísticamente significativos con un  $\alpha = 5\%$  según el valor-p asociado el estadístico de Wald y los intervalos del 95 % de confianza para las mismas no contienen el cero. La variable que corresponde a haber visto cuatro o más temas de estadística en el bachillerato resultó significativa con un  $\alpha = 10\%$ .

Con base en los Odds-Ratio calculados que se muestran en la tabla **8-2**, se puede concluir que si un estudiante no tuvo un buen desempeño en cursos de matemáticas en la Universidad, tiene entre 1.15 y 9.24 veces más posibilidades de reprobar el curso de estadística, dado que el intervalo de confianza *Odds-Ratio*, es  $> 1$ , esta variable afecta positivamente la probabilidad de reprobar. Quien no tuvo experiencia con cursos de estadística en el bachillerato, tiene más posibilidades de reprobar en 6.23 que quien si tuvo experiencia. Si el estudiante no vio conceptos de estadística como parte del curso de matemáticas en el bachillerato, tiene más posibilidades de reprobar en 5.37 que quien si los vio de esta forma. Estas dos características aumentan la probabilidad de aprobar, ya que el intervalo de confianza *Odds-Ratio*, es  $< 1$ . Y si el estudiante vio cuatro o más temas de estadística en el colegio, tiene entre 0.93 y 13.58 más posibilidades de reprobar el primer curso de estadística en la universidad.

**Tabla 8-3.:** Estadísticas Descriptivas para probabilidades

| Predicción Probabilidad | Mínimo   | Mediana   | Máximo    |
|-------------------------|----------|-----------|-----------|
| $\hat{\theta}$          | 0.633504 | 0.6556026 | 0.8841633 |

Aplicando exponencial al modelo 8-2 se calcularon las probabilidades para todos los estudiantes. Con base en los resultados presentados en la tabla **8-3** se concluye que la probabilidad de que un estudiante no repruebe el curso de estadística en la universidad, haya tenido un buen desempeño en matemáticas en la universidad, no haya tenido experiencia

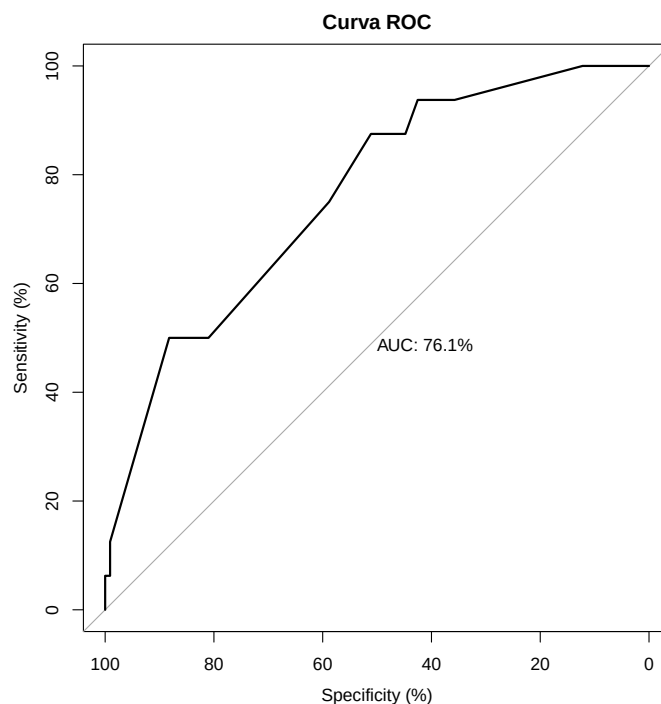
con cursos de estadística en el bachillerato, no haya visto conceptos de estadística como parte del curso de matemáticas y haya visto menos de cuatro temas de estadística en el bachillerato es del 63.35 %. Por el contrario, la probabilidad de que un estudiante repruebe el curso de estadística en la universidad, no haya tenido un buen desempeño en matemáticas en la universidad, haya tenido experiencia con cursos de estadística en el bachillerato, haya visto conceptos de estadística como parte del curso de matemáticas y haya visto al menos cuatro temas de estadística en el bachillerato es del 88.42 %. El 50 % de los estudiantes tienen una probabilidad superior al 65 % de reprobado el curso de estadística.

### **8.3.1. Capacidad clasificatoria del modelo de respuesta binaria con función de enlace Log-Log Complementario**

Con el fin de probar que tan bien está prediciendo la probabilidad de que un estudiante repruebe o no el primer curso de estadística en la universidad, con respecto a las covariables que resultaron significativas en el modelo encontrado, y considerando que el modelo presenta asimetría, se construyó la curva ROC no paramétrica, que indica que tan bien está clasificando el modelo a un estudiante como reprobado cuando éste realmente reprueba la materia y cuando el modelo indica que el estudiante aprueba el curso cuando realmente aprobó la materia.

La curva ROC encontrada como se puede ver en la figura **8-3** tiene un área bajo la curva de **76.15 %**, en el escenario ideal, el área bajo la curva es igual al 100 % pero esto en la práctica es poco probable que suceda, por lo tanto el porcentaje encontrado bajo la curva es medianamente bueno. Como complemento al análisis de la curva ROC se calculó el Accuracy Ratio (AR) para analizar la capacidad que tiene el modelo de discriminar a estudiantes que reprueban la materia y a estudiantes que aprueban la materia, este indicador según Altman, Haldeman and Narayanan (1977) señalan que un modelo será adecuado cuando el indicador AR es superior al 50 %.

$$AR = (ROC - 0,5) * 2 = (0,7615 - 0,5) * 2 = 0,523 = 52,3 \%$$



**Figura 8-3.:** Curva ROC

Según los cálculos realizados el AR de **52.3 %** indica que el modelo es adecuado para discriminar cuando el estudiante reprueba o aprueba el curso.

Una de las medidas de bondad utilizadas en modelos de elección discreta es predecir los valores de la variable de respuesta, donde esta tomará el valor unitario cuando la probabilidad de ocurrencia estimada por el modelo sea mayor a un determinado valor  $K$ , en este caso, clasificará como reprobado, en el caso contrario tomará el valor de cero y se clasificará como aprobado. El valor de  $K$  es determinado como el máximo punto alcanzado por la curva ROC, este máximo se considera como punto de corte a la hora de clasificar y según los resultados encontrados en la curva ROC, el  $K$  para este caso es de **0.65**.

Para establecer si la bondad del modelo es adecuada, es de interés conocer las proporciones de correctos clasificados como reprobados, correctos clasificados como no reprobados, tasa de falsos reprobados y la tasa de falsos no reprobados, que se obtienen de la tabla **8-4** de

**Tabla 8-4.: Clasificación**

|                 | Reprobó | Aprobó | Total |
|-----------------|---------|--------|-------|
| $P_i > 0,65$    | 14      | 108    | 122   |
| $P_i \leq 0,65$ | 2       | 113    | 115   |
| Total           | 16      | 221    | 237   |

clasificación del modelo.

#### **Correctos Clasificados como Reprobados (CCR)**

$$\text{CCR} = Pr(P_i > 0,65/Y_i = 1) = \frac{14}{14+2} = 0,875$$

#### **Correctos Clasificados como No Reprobados (CCNR)**

$$\text{CCNR} = Pr(P_i \leq 0,65/Y_i = 0) = \frac{113}{108+113} = 0,5113$$

#### **Tasa de Falsos Reprobados (FR)**

$$\text{FR} = Pr(P_i > 0,65/Y_i = 0) = \frac{108}{108+113} = 0,4887$$

#### **Tasa de Falsos No Reprobados (FNR)**

$$\text{FNR} = Pr(P_i \leq 0,65/Y_i = 1) = \frac{2}{14+2} = 0,125$$

La probabilidad de clasificar correctamente un estudiante que reprueba el curso es del **87.5 %**, lo que resulta adecuado, ya que se puede decir que de 100 estudiantes que pierden el curso, el modelo clasifica correctamente como reprobados aproximadamente a 88 estudiantes. La probabilidad de clasificar correctamente a un estudiante que aprobó el curso es del **51.13 %**, es decir, el modelo estará clasificando correctamente como aprobados de 100 estudiantes a aproximadamente 51 de ellos.

La tasa de falsos reprobados que indica que la probabilidad de clasificar un estudiante como reprobado cuando este realmente aprobó el curso es del **48,87 %**, es decir, de 100 estudiantes que aprobaron el curso, el modelo esta clasificando erróneamente a aproximadamente 49



estudiantes. La probabilidad de clasificar un estudiante como aprobado cuando este realmente reprobó el curso (Tasa de falsos no reprobados) es del **12,5 %**, es decir, de 100 estudiantes que reprobaron el curso, aproximadamente 13 de ellos el modelo los clasifica como que aprobaron el curso. Se presenta un porcentaje correcto de clasificación (**P.C.C**) del 53.6 %, es decir, la probabilidad de acertar en la predicción para todos los estudiantes es del 53.6 %.

De otro lado, el **valor predictivo positivo (VPP)** obtenido indica que la probabilidad de que un estudiante repruebe el curso dado que haya sido pronosticado como reprobado es del 10 %. El **valor predictivo negativo (VPN)** sugiere que la probabilidad de que un estudiante apruebe el curso dado que haya sido pronosticado como aprobado es del 98 %, por tanto, luego de ser pronosticado como no reprobado, existe aun una probabilidad del 2 % de reprobado el curso de estadística. El valor predictivo positivo es bajo, ya que al haber un mayor número de estudiantes que aprueban, se incrementa el número de falsos reprobados. El 10 % del 51 % pronosticado como reprobados van a estar bien clasificados.

## 8.4. Aproximación Bayesiana

Con el interés en estimar la probabilidad de que un estudiante repruebe el curso de estadística, se caracterizaron a los individuos por las cuatro covariables que aportaron información acerca de su educación previa.

La elicitación de los hiperparámetros se realizó obteniendo como información a priori para el procedimiento, el valor mínimo y máximo de los valores ajustados obtenidos a partir del modelo (Método Empírico de Bayes), por tanto, los hiperparámetros  $a$  y  $b$  son los mismos valores para las 16 combinaciones como información previa y se presentan en la tabla **8-5**. Los resultados obtenidos de las estimaciones realizadas bajo técnicas de la estadística clásica y bayesiana se presentan en la tabla **8-6**.

Se puede observar que bajo la estadística clásica las estimaciones son muy pequeñas, siendo aproximadamente cero la mayoría de las mismas, ya que no se presentan éxitos

**Tabla 8-5.:** Elicitación de hiperparámetros

| Límite inferior | Límite superior | $\theta_0$ | $a$      | $b$      |
|-----------------|-----------------|------------|----------|----------|
| 0.633504        | 0.8841633       | 0.7588337  | 176.0616 | 55.95446 |

**Tabla 8-6.:** Estimación clásica y Bayesiana

| Grupo k | Total | Reprobar | Aprobar | MV                   | Estimaciones        |                     |                     |
|---------|-------|----------|---------|----------------------|---------------------|---------------------|---------------------|
|         |       |          |         |                      | LaPlace             | Jeffreys            | Conjugada           |
| 1       | 1     | 0        | 1       | 0.000 (0;0)          | 0.333 (0.013;0.842) | 0.250 (0.000;0.853) | 0.756 (0.699;0.808) |
| 2       | 1     | 0        | 1       | 0.000 (0;0)          | 0.333 (0.013;0.842) | 0.250 (0.000;0.853) | 0.756 (0.699;0.808) |
| 3       | 15    | 0        | 15      | 0.000 (0;0)          | 0.059 (0.002;0.206) | 0.031 (0.000;0.152) | 0.713 (0.655;0.767) |
| 4       | 6     | 1        | 5       | 0.167 (-0.131;0.465) | 0.250 (0.037;0.579) | 0.214 (0.019;0.558) | 0.744 (0.687;0.797) |
| 5       | 1     | 0        | 1       | 0.000 (0;0)          | 0.333 (0.013;0.842) | 0.250 (0.000;0.853) | 0.756 (0.699;0.808) |
| 6       | 19    | 2        | 17      | 0.105 (-0.033;0.243) | 0.143 (0.032;0.317) | 0.125 (0.023;0.297) | 0.709 (0.652;0.764) |
| 7       | 8     | 0        | 8       | 0.000 (0;0)          | 0.100 (0.003;0.336) | 0.056 (0.000;0.262) | 0.734 (0.676;0.787) |
| 8       | 1     | 1        | 0       | 1.000 (1;1)          | 0.667 (0.158;0.987) | 0.750 (0.147;1.000) | 0.760 (0.703;0.812) |
| 9       | 12    | 0        | 12      | 0.000 (0;0)          | 0.071 (0.002;0.247) | 0.038 (0.000;0.185) | 0.722 (0.664;0.776) |
| 10      | 53    | 4        | 49      | 0.075 (0.0041;0.146) | 0.091 (0.031;0.179) | 0.083 (0.026;0.170) | 0.632 (0.575;0.687) |
| 11      | 7     | 0        | 7       | 0.000 (0;0)          | 0.111 (0.003;0.369) | 0.062 (0.000;0.292) | 0.737 (0.679;0.790) |
| 12      | 15    | 0        | 15      | 0.000 (0;0)          | 0.059 (0.002;0.206) | 0.031 (0.000;0.152) | 0.713 (0.655;0.767) |
| 13      | 1     | 1        | 0       | 1.000 (1;1)          | 0.667 (0.158;0.987) | 0.750 (0.147;1.000) | 0.760 (0.703;0.812) |
| 14      | 30    | 6        | 24      | 0.200 (0.057;0.343)  | 0.219 (0.096;0.375) | 0.210 (0.088;0.367) | 0.695 (0.638;0.749) |
| 15      | 53    | 1        | 52      | 0.019 (-0.018;0.056) | 0.036 (0.005;0.099) | 0.028 (0.002;0.085) | 0.621 (0.564;0.677) |
| 16      | 14    | 0        | 14      | 0.000 (0;0)          | 0.062 (0.002;0.218) | 0.033 (0.000;0.162) | 0.716 (0.658;0.770) |

en determinadas combinaciones; esto sucede con los grupos 1, 2 y 5 que tienen un solo individuo que aprobó el curso. Luego, en los grupos 3, 7, 9, 11, 12 y 16, aunque aumenta la muestra sin éxitos, al estar tan cerca de los límites del espacio paramétrico, el método de máxima verosimilitud no consigue hacer estimación, no logra tener convergencia.

Por el contrario, cuando se tiene presente una información a priori como la no informativa por el método de LaPlace y Jeffreys, se mejora la estimación en los casos donde se cuenta con cero éxitos y las regiones de credibilidad no abarcan valores inferiores a cero. Comparando la estimación de LaPlace y Jeffreys se tiene que, por el método de Jeffreys se presentan estimaciones más pequeñas, que dan idea de lo importante que resulta considerar la inclusión de información subjetiva contenida en una distribución a priori. A su vez, se aprecia que

las estimaciones con información a priori informativa (información de un experto o de otros estudios realizados) son más altas y dejan ver que la apriori tiene un impacto fuerte sobre la estimación.

Es importante anotar que los intervalos de confianza para el estimador clásico presentan límites negativos, lo que indica que debido a los pocos éxitos en cada combinación es necesario aplicar la técnica bayesiana, las regiones de credibilidad para las estimaciones bayesianas presentan una longitud muy pequeña en comparación a los clásicos. La longitud de las regiones de Credibilidad para los estimadores bayesianos con una a priori conjugada tienen la menor longitud en comparación con los otros.

Respecto a los resultados obtenidos para cada uno de los grupos, se encontró que:

- De 100 estudiantes cuyo desempeño en cursos de matemáticas en la universidad no fue bueno, no tuvieron experiencia en cursos de estadística en bachillerato, tuvieron conceptos de estadística vistos como parte del curso de matemáticas y que vieron cuatro o más temas de estadística en el bachillerato, correspondiente al **grupo 13**, se espera que aproximadamente 67 de ellos reprueben el curso de estadística en la universidad.
- Para un grupo de estudiantes cuyo desempeño en cursos de matemáticas en la universidad fue bueno, que tuvieron experiencia en cursos de estadística en bachillerato, que no tuvieron conceptos vistos como parte del curso de matemáticas y que tengan cuatro o más temas de estadística vistos, correspondiente a la **combinación 10**, se espera que de 100 de ellos, el curso aproximadamente sea reprobado por nueve.
- Estudiantes que se caracterizaron por no tener un buen desempeño en el curso de matemáticas en la universidad, no tuvieron experiencia en cursos de estadística en el bachillerato, no tuvieron conceptos de estadística vistos como parte del curso de matemáticas y reconocen haber visto cuatro o más temas de estadística en el bachillerato (**grupo 8**), se espera que de 100 estudiantes que tengan estas características 67 reprueben el curso de estadística.

- En un grupo de estudiantes que no tuvo un buen desempeño en cursos de matemáticas en la universidad, tuvieron experiencia en cursos de estadística en el bachillerato, no tuvieron conceptos de estadística vistos como parte del curso de matemáticas e identifican haber visto cuatro o más temas de estadística en el colegio, (**grupo 14**), de 100 de estos estudiantes, se espera que solamente 22 de ellos reprueben el curso de estadística.
- En un grupo de estudiantes con buen desempeño en cursos de matemáticas en la universidad, que tuvieron experiencia en cursos de estadística en el bachillerato, vieron conceptos de estadística como parte del curso de matemáticas e identifican haber visto cuatro o más temas de estadística en el colegio, (**grupo 15**), se espera que de 100 de estudiantes con estas características, solamente 4 de ellos reprueben el curso de estadística.

Como se ha podido percibir, el hecho que un estudiante tenga experiencia en cursos de estadística en el bachillerato, aparentemente influye en su resultado en el curso de estadística en la universidad.

Luego, en caso de realizarse un nuevo estudio con igual tamaño de muestra, es deseable conocer cuál es la probabilidad de que se encuentre un número mayor o igual de estudiantes que reprueban el curso de estadística en la universidad condicionado a la combinación de covariables. Con este fin se obtuvieron las probabilidades predictivas posteriores utilizando la distribución apriori no informativa de Laplace.

De acuerdo con la información contenida en la tabla **8-7**, con un nuevo estudio hay 99% de probabilidad de que los estudiantes reprueben el curso de estadística dada la combinación 14, que corresponde a estudiantes que no tuvieron un buen desempeño en cursos de matemáticas en la universidad, tuvieron experiencia en cursos de estadística en el bachillerato, no tuvieron conceptos de estadística vistos como parte del curso de matemáticas y reportan haber visto cuatro o más temas de estadística en el colegio. Sin

**Tabla 8-7.:** Probabilidades Predictivas Posteriores

| Grupo                                                                                                                          | $P(Y \geq y)$ |
|--------------------------------------------------------------------------------------------------------------------------------|---------------|
| 1: Desempeño matemát. Univ (B), No estadística bachillerato, No conceptos estad en matemáticas, Menos de cuatro temas vistos   | 0.3333333     |
| 2: Desempeño matemát. Univ (NB), No estadística bachillerato, No conceptos estad en matemáticas, Menos de cuatro temas vistos  | 0.3333333     |
| 3: Desempeño matemát. Univ (B), Si estadística bachillerato, No conceptos estad en matemáticas, Menos de cuatro temas vistos   | 0.4838710     |
| 4: Desempeño matemát. Univ (B), No estadística bachillerato, Si conceptos estad en matemáticas, Menos de cuatro temas vistos   | 0.7307692     |
| 5: Desempeño matemát. Univ (B), No estadística bachillerato, No conceptos estad en matemáticas, Cuatro o más temas vistos      | 0.3333333     |
| 6: Desempeño matemát. Univ (NB), Si estadística bachillerato, No conceptos estad en matemáticas, Menos de cuatro temas vistos  | 0.8752599     |
| 7: Desempeño matemát. Univ (NB), No estadística bachillerato, Si conceptos estad en matemáticas, Menos de cuatro temas vistos  | 0.4705882     |
| 8: Desempeño matemát. Univ (NB), No estadística bachillerato, No conceptos estad en matemáticas, Cuatro o más temas vistos     | 0.6666667     |
| 9: Desempeño matemát. Univ (B), Si estadística bachillerato, Si conceptos estad en matemáticas, Menos de cuatro temas vistos   | 0.4800000     |
| 10: Desempeño matemát. Univ (B), Si estadística bachillerato, No conceptos estad en matemáticas, Cuatro o más temas vistos     | 0.9702516     |
| 11: Desempeño matemát. Univ (B), No estadística bachillerato, Si conceptos estad en matemáticas, Cuatro o más temas vistos     | 0.4666667     |
| 12: Desempeño matemát. Univ (NB), Si estadística bachillerato, Si conceptos estad en matemáticas, Menos de cuatro temas vistos | 0.4838710     |
| 13: Desempeño matemát. Univ (NB), No estadística bachillerato, Si conceptos estad en matemáticas, Cuatro o más temas vistos    | 0.6666667     |
| 14: Desempeño matemát. Univ (NB), Si estadística bachillerato, No conceptos estad en matemáticas, Cuatro o más temas vistos    | 0.9939726     |
| 15: Desempeño matemát. Univ (B), Si estadística bachillerato, Si conceptos estad en matemáticas, Cuatro o más temas vistos     | 0.7476636     |
| 16: Desempeño matemát. Univ (NB), Si estadística bachillerato, Si conceptos estad en matemáticas, Cuatro o más temas vistos    | 0.4827586     |

Abreviaturas.(B):Bueno,(NB):No Bueno

embargo, la probabilidad de que los estudiantes reprobren el curso de estadística en la Universidad es del 67 % si son personas que no tuvieron un buen desempeño en cursos de matemáticas en la universidad, sin experiencia en cursos de estadística en el bachillerato, que tuvieron conceptos de estadística vistos como parte del curso de matemáticas y reportan haber visto cuatro o más temas de estadística en el colegio como se indica en la combinación 13.

El 50 % de los estudiantes que reprobaron el curso dada la combinación 14, son estudiantes de la Facultad de Ciencias Administrativas y Económicas, siendo esta la proporción más alta en la clasificación por combinación de covariables. Solamente un estudiante de la

Facultad de Ciencias Naturales clasificó en la combinación ocho que indica que no tuvo un buen desempeño en cursos de matemáticas en la universidad, no tuvo experiencia en cursos de estadística en el bachillerato, no tuvo conceptos de estadística vistos como parte del curso de matemáticas y reporta haber visto cuatro o más temas de estadística en el colegio. De igual manera, solo un estudiante de la Facultad de Ciencias Administrativas y Económicas clasificó en la combinación 13 que son personas que no tuvieron un buen desempeño en cursos de matemáticas en la universidad, sin experiencia en cursos de estadística en el bachillerato, que tuvieron conceptos de estadística vistos como parte del curso de matemáticas y reportan haber visto cuatro o más temas de estadística en el colegio. En ambos casos, los estudiantes reprobaron el curso de estadística.

Estudiantes de la Facultad de Ciencias de la Educación clasificaron en las combinaciones 12 y 15, pero ninguno reprobó el curso. En estas combinaciones la probabilidad de reprobación es baja y tienen en común haber visto estadística en bachillerato y haber tomado conceptos de estadística como parte de un curso de matemáticas.

Estudiantes de la Facultad de Ciencias de la Salud clasificaron en las combinaciones 6, 10 y 15 que presentan bajas estimaciones en la probabilidad de reprobación y coinciden en haber tenido experiencia con estadística en el bachillerato.

## 9. Conclusiones y recomendaciones

La utilización de un modelo de regresión binaria con enlace log-log complementario permitió establecer la relación entre variables asociadas con la experiencia con estadística en el colegio y el desempeño académico de los estudiantes en el primer curso de estadística en la universidad, aún con la limitación en cuanto al bajo número de datos correspondientes a estudiantes que reprobaron la asignatura, que para el caso fue una proporción del 6.75 %. Las características o variables explicativas que presentan relación significativa con la probabilidad de reprobación del curso de estadística fueron el desempeño alcanzado en cursos de matemáticas en la universidad, la experiencia en cursos de estadística en el bachillerato, conceptos de estadística vistos como parte de un curso de matemáticas en el colegio y haber visto al menos cuatro temas de estadística en el bachillerato.

Se encontró que el nivel de ansiedad que manifiesta el estudiante, establecido de manera heurística, no resultó ser una covariable influyente de manera significativa en su desempeño en el curso de estadística a nivel universitario; si bien el 36 % de los estudiantes que reprobaron el curso presentaban un índice de ansiedad importante, esta variable considerada para estudio, no resultó ser determinante en el resultado en el primer curso de estadística en la universidad. Esto puede deberse a la construcción misma del índice. La ansiedad es un aspecto de importancia, puede ser una consecuencia o un factor asociado a otras características como la experiencia previa y puede ser evaluada de otra manera.

El modelo planteado tiene una probabilidad del 87.5 % de clasificar correctamente a los estudiantes que reprueban el curso y del 51,1 % de clasificar correctamente al estudiante

que aprueba. El modelo discrimina bien a los estudiantes a un punto de corte 0,65, ya que arroja un porcentaje correcto de clasificación superior al 50 %.

Los estudiantes que presentan una probabilidad mayor de reprobar el primer curso de estadística que cursan en la Universidad, son aquellos cuyo desempeño en cursos de matemáticas en la universidad no fue bueno, no tuvieron experiencia en cursos de estadística en bachillerato, tuvieron conceptos de estadística vistos como parte del curso de matemáticas y vieron cuatro o más temas de estadística en el bachillerato.

La mayor probabilidad de riesgo de que los estudiantes reprueben el curso de estadística está en personas que no tuvieron un buen desempeño en cursos de matemáticas en la universidad, tuvieron experiencia en cursos de estadística en el bachillerato, no tuvieron conceptos de estadística vistos como parte del curso de matemáticas y reportan haber visto cuatro o más temas de estadística en el colegio, de acuerdo con las probabilidades predictivas obtenidas a partir de una distribución a priori por el Principio de Insuficiencia de Laplace. Seis estudiantes de la muestra que reprobaron el curso de estadística, cumplen con las condiciones mencionadas y de ellos, tres pertenecen a la facultad de Ciencias Administrativas y Económicas.

En la aproximación bayesiana, las estimaciones obtenidas a partir de una distribución apriori informativa (conjugada) fueron altas, lo que deja ver que hay un impacto muy fuerte de la apriori en la estimación. Es de anotar la importancia de disponer de información de un experto o de datos de otros estudios para la elicitación de hiperparámetros, con el objeto de corroborar o rebatir esta apreciación.

Con este estudio se percibe que el hecho de que un estudiante tenga experiencia en cursos de estadística en el bachillerato, influye en su resultado en el curso de estadística, por tanto, se sugiere una reflexión sobre la importancia de que el profesional en Estadística tenga mayor participación impartiendo la enseñanza de Estadística en los colegios, de forma que



se promueva realmente el pensamiento estadístico y contribuya positivamente al proceso de aprendizaje mejorando la experiencia vivida por el estudiante.

Una limitación del estudio fue la aplicación del cuestionario, si bien el propósito inicial era aplicarlo en el primer día del curso, por inconvenientes en la logística, no fue posible lograr esto con todos los encuestados, puesto que a algunos se les aplicó la encuesta hasta dos o tres semanas después del inicio del curso. Esta situación podría generar un sesgo en las respuestas de algunos estudiantes. Sin embargo, se esperó que esto no fuera relevante si se considera que si hay ansiedad o preocupación presentes, estos sentimientos no necesariamente desaparecen en corto tiempo, pero si podría suceder que la percepción no fuese la misma que al inicio del curso.

Finalmente, los resultados arrojados por el modelo dejan una inquietud en cuanto al hecho que no se esperaría que ver diferentes temas de estadística en el colegio representara una desventaja, al estar asociado con la probabilidad de reprobar posteriormente un curso de estadística en un nivel superior; por esta razón sería recomendable estudiar esta asociación en futuros estudios que permitan analizar con mayor profundidad las posibles razones que originan esta relación.

## **A. Anexo: Encuesta aplicada**

UNIVERSIDAD ICESI – CALI  
ESCUELA DE CIENCIAS DE LA EDUCACIÓN  
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA  
ÁREA DE ESTADÍSTICA

Señor estudiante: con la presente encuesta, los profesores del área de Estadística buscamos obtener información acerca de sus expectativas frente al curso que va a iniciar como también sobre sus conocimientos en la materia y su experiencia frente a este tipo de conocimiento. El uso de sus respuestas es absolutamente confidencial y para fines exclusivamente académicos. Por favor responda a la totalidad de las preguntas con la mayor sinceridad posible, agradecemos su colaboración.

Por favor diligencie el formulario con bolígrafo.

Código \_\_\_\_\_  
Carrera \_\_\_\_\_  
Materia \_\_\_\_\_  
Grupo matriculado \_\_\_\_\_  
Edad \_\_\_\_\_  
Género: M \_\_\_\_ F \_\_\_\_

Marque con X el desempeño alcanzado en los cursos de matemáticas

| Nivel previo de formación | Desempeño |           |       |         |      |
|---------------------------|-----------|-----------|-------|---------|------|
|                           | Excelente | Muy bueno | Bueno | Regular | Malo |
| Primaria                  |           |           |       |         |      |
| Bachillerato              |           |           |       |         |      |
| Universidad (si lo tiene) |           |           |       |         |      |

Marque con X su experiencia y desempeño en los cursos de estadística vistos antes por usted

| Nivel previo de formación  | Experiencia? |    | Si tiene experiencia, señale el número de cursos en cada nivel y su calificación para el desempeño del mismo |           |           |       |         |      |
|----------------------------|--------------|----|--------------------------------------------------------------------------------------------------------------|-----------|-----------|-------|---------|------|
|                            | Si           | No | # cursos                                                                                                     | Desempeño |           |       |         |      |
|                            |              |    |                                                                                                              | Excelente | Muy bueno | Bueno | Regular | Malo |
| Primaria                   |              |    |                                                                                                              |           |           |       |         |      |
| Bachillerato               |              |    |                                                                                                              |           |           |       |         |      |
| Otra carrera o institución |              |    |                                                                                                              |           |           |       |         |      |
| Carrera actual (repitente) |              |    |                                                                                                              |           |           |       |         |      |

Los conceptos de estadística que vio en el bachillerato los tomó:

- Dentro de un curso de solo estadística \_\_\_\_
- Como parte del curso de matemáticas \_\_\_\_
- Dentro de otros cursos de ciencias (Química, Física, etc.) \_\_\_\_

Si éste es el primer curso de estadística que matricula en la universidad y estudió temas de estadística en el colegio. En la lista de conceptos de estadística que aparece en la columna izquierda, marque los que recuerde haber visto en el colegio y marque también nivel de dificultad que usted les asoció en su curso.

| Conceptos                            | Visto en el colegio |    | Grado de dificultad |         |       |           |             |
|--------------------------------------|---------------------|----|---------------------|---------|-------|-----------|-------------|
|                                      | Si                  | no | Muy difícil         | Difícil | Fácil | Muy fácil | No recuerdo |
| Tablas de frecuencias                |                     |    |                     |         |       |           |             |
| Gráficos de barras e Histogramas     |                     |    |                     |         |       |           |             |
| Media o promedio, mediana, moda      |                     |    |                     |         |       |           |             |
| Rango, varianza, desviación estándar |                     |    |                     |         |       |           |             |
| Permutaciones                        |                     |    |                     |         |       |           |             |
| Combinaciones                        |                     |    |                     |         |       |           |             |
| Algunas reglas de probabilidad       |                     |    |                     |         |       |           |             |

Si considera que entender alguno de estos temas fue muy difícil o difícil, ¿cuál o cuáles de las siguientes causas cree que pueden haber sido la razón? (puede marcar varias)

- Los temas incluían operaciones matemáticas muy difíciles \_\_\_\_
- Los problemas tratados no tenían relación con la vida real \_\_\_\_
- No le entendía la explicación al profesor \_\_\_\_
- No prestaba atención porque no le gustan las matemáticas \_\_\_\_
- Resolver estos temas demanda mucho tiempo y prefería hacer otras cosas \_\_\_\_

Usted se graduó del bachillerato en un colegio de tipo:

- Oficial \_\_\_\_
- Privado \_\_\_\_

En el currículo de la carrera que está estudiando, se ha incluido 1 o 2 cursos de estadística, ¿considera que incluirlos es necesario e importante? Sí \_\_\_\_ No \_\_\_\_

¿Cuál de los siguientes enunciados cree usted que podría ser la razón más importante para explicar su respuesta a la pregunta anterior? (Marque solo una opción)

- En el mundo laboral actual, el trabajo interdisciplinario y el manejo e interpretación de datos facilita la comunicación entre profesionales. \_\_\_\_
- Un aspecto importante en el ejercicio de la profesión es la toma de decisiones. \_\_\_\_
- Tomar buenas decisiones requiere de evidencia determinada a partir de la recolección de datos. \_\_\_\_
- Cuando esté ejerciendo la profesión y requiera de algún análisis estadístico, sólo será suficiente contratar a un profesional de la estadística. \_\_\_\_
- El enfoque que quiero dar a mi vida profesional, no requiere de conocimientos en estadística. \_\_\_\_
- La universidad considera que es importante incluir estadística en el currículo, pero para mí no lo es. \_\_\_\_
- Pienso que es importante, pero en realidad no sé por qué. \_\_\_\_

Para afrontar con éxito el curso que está por comenzar ¿cómo será su forma de estudiar, en cuanto al tiempo de trabajo fuera de clase que le piensa dedicar?

- a. Estudiaré solo unas pocas horas antes de cada examen. \_\_\_\_
- b. Le dedicaré máximo una hora a la semana a esta materia. \_\_\_\_
- c. Le dedicaré al menos una hora a la semana a esta materia. \_\_\_\_
- d. Le dedicaré al menos 2 horas a la semana a esta materia. \_\_\_\_
- e. La materia no requiere que le dedique tiempo adicional para aprobarla. \_\_\_\_

¿Ha averiguado a amigos o estudiantes de semestres posteriores, acerca de cómo es el curso de estadística que va a iniciar? Sí \_\_\_\_ No \_\_\_\_

Si su respuesta anterior es afirmativa, algunos de los comentarios hechos por sus amigos o estudiantes de semestres posteriores fueron: (puede marcar más de una opción)

- a. La materia es fácil y no requiere de mucho trabajo para aprobarla. \_\_\_\_
- b. Es suficiente estudiar unas pocas horas antes de cada examen. \_\_\_\_
- c. No se le entiende a los profesores y es necesario buscar otros tutores. \_\_\_\_
- d. La materia requiere de mucho tiempo de dedicación y mucho esfuerzo adicional para aprobarla. \_\_\_\_
- e. Es necesario dedicar tiempo adicional suficiente para entender los temas y hacer los ejercicios propuestos. \_\_\_\_
- f. Con asistir a las tutorías de los profesores es suficiente. \_\_\_\_

¿Tiene en la actualidad algún tipo de trabajo por el cual recibe remuneración? Sí \_\_\_\_ No \_\_\_\_

En caso de estar trabajando, usted está realizando:

- a. Monitoria en la universidad. \_\_\_\_
- b. Trabajo todas las noches entre 2 y 8 horas. \_\_\_\_
- c. Trabajo entre 2 y 8 horas al día en el negocio de la familia o propio. \_\_\_\_
- d. Trabajo entre 2 y 8 horas al día en negocio o empresa ajeno. \_\_\_\_
- e. Trabajo solo los fines de semana en el negocio de la familia o propio. \_\_\_\_
- f. Trabajo solo los fines de semana en negocio o empresa ajeno. \_\_\_\_

¿Cómo espera que sea el grado de dificultad del curso de estadística que matriculó?

- a. Muy alto. \_\_\_\_ b. Alto. \_\_\_\_ c. Bajo. \_\_\_\_ d. Muy bajo \_\_\_\_

Cuáles de los siguientes enunciados, podría ser una razón que explique su respuesta a la pregunta anterior.

- a. Tengo buenos conocimientos en matemáticas. \_\_\_\_
- b. Estoy interesado y tengo una buena actitud hacia la carrera. \_\_\_\_
- c. Tengo confianza porque tuve buenos resultados en los cursos previos. \_\_\_\_
- d. Me parecen difíciles las materias que utilizan las matemáticas. \_\_\_\_
- e. No tengo el suficiente tiempo para dedicarle a la materia. \_\_\_\_
- f. No he visto nada de estadística anteriormente. \_\_\_\_

¿Cómo se siente ahora frente al curso de estadística que va a comenzar? (marque con una X el recuadro que más identifique su respuesta)

|                                                                                           |                                                                                  |                                                                                       |                                                                             |
|-------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>Preocupado:</b> porque va a ser un curso muy difícil y es muy probable que lo repruebe | <b>Tenso:</b> porque tendrá que esforzarse mucho para aprobar con la nota mínima | <b>Tranquilo:</b> porque cuenta con las herramientas necesarias para aprobar el curso | <b>Muy tranquilo:</b> porque es un curso que no le demandará mayor esfuerzo |
|-------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|

## **B. Anexo: Diccionario de datos**

**Diccionario de datos**

| Numero | Variable         | Descripción                                                                 | Categorías                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Tipo                  |
|--------|------------------|-----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| 1      | Id               | Identificador encuesta                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Cuantitativa-Discreta |
| 2      | Universidad      | Nombre Universidad                                                          | 1: Universidad Icesi                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Cualitativa-Nominal   |
| 3      | Código           | Código del estudiante                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Cualitativa-Nominal   |
| 4      | Facultad         | Nombre Facultad                                                             | 1: Ciencias Administrativas y Económicas<br>2: Ingeniería<br>3: Derecho y Ciencias Sociales<br>4: Ciencias Naturales<br>5: Ciencias de la Salud<br>6: Ciencias de la Educación                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal   |
| 5      | Carrera          | Carrera en la que está matriculado cada estudiante                          | 1: Administración de Empresas<br>2: Contaduría Pública y Finanzas Internacionales<br>3: Economía y Negocios Internacionales<br>4: Mercadeo Internacional y Publicidad<br>5: Diseño de Medios Interactivos<br>6: Diseño Industrial<br>7: Ingeniería Bioquímica<br>8: Ingeniería de Sistemas<br>9: Ingeniería Industrial<br>10: Ingeniería Telemática<br>11: Ciencia Política<br>12: Derecho<br>13: Música<br>14: Sociología<br>15: Biología<br>16: Química con énfasis en Bioquímica<br>17: Química Farmacéutica<br>18: Medicina<br>19: Licenciatura en Artes<br>20: Licenciatura en Ciencias Sociales<br>21: Licenciatura en Lenguas Extranjeras<br>22: Licenciatura en Literatura y Lengua Castellana<br>23: Psicología | Cualitativa-Nominal   |
| 6      | Materia          | Curso de Estadística matriculado                                            | 08131: Teoría de Probabilidades<br>08290: Fundamentos de Probabilidad y Estadística<br>08293: Estadística y Probabilidad Básica<br>08295: Bioestadística<br>08299: Bioestadística y Fundamentos de Epidemiología                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Cualitativa-Nominal   |
| 7      | Grupo            | Grupo matriculado                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Cualitativa-Nominal   |
| 8      | Semana           | Semana del semestre en la que se realizó la encuesta                        | 1: semana 1<br>2: semana 2<br>3: semana 3                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Cualitativa-Ordinal   |
| 9      | edad             | Edad del estudiante                                                         | 0: 17 - 22 años<br>1: 23 - 28 años                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Cualitativa-Nominal   |
| 10     | Género           | Genero del estudiante                                                       | 0: Femenino<br>1: Masculino                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Cualitativa-Nominal   |
| 11     | PrimariaMat      | Desempeño en cursos de matemáticas en Primaria                              | 0: 1-Excelente, 2-Muy Bueno, 3-Bueno<br>1: 4-Regular, 5-Malo                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Cualitativa-Ordinal   |
| 12     | BachillerMat     | Desempeño en cursos de matemáticas en Bachillerato                          | 0: 1-Excelente, 2-Muy Bueno, 3-Bueno<br>1: 4-Regular, 5-Malo                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Cualitativa-Ordinal   |
| 13     | UniversidadMat   | Desempeño en cursos de matemáticas en la Universidad                        | 0: 1-Excelente, 2-Muy Bueno, 3-Bueno<br>1: 4-Regular, 5-Malo                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Cualitativa-Ordinal   |
| 14     | PrimEstExp       | Experiencia en cursos de estadística en Primaria                            | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal   |
| 15     | PrimEstcursos    | Número de cursos de estadística vistos en Primaria                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Cualitativa-Nominal   |
| 16     | PrimEstcalific   | Desempeño en cursos de estadística vistos en Primaria                       | 0: 1-Excelente, 2-Muy Bueno, 3-Bueno<br>1: 4-Regular, 5-Malo                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Cualitativa-Ordinal   |
| 17     | BachiEstExp      | Experiencia en cursos de estadística en Bachillerato                        | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal   |
| 18     | BachiEstcursos   | Número de cursos de estadística vistos en Bachillerato                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Cualitativa-Nominal   |
| 19     | BachiEstcalific  | Desempeño en cursos de estadística vistos en Bachillerato                   | 0: 1-Excelente, 2-Muy Bueno, 3-Bueno<br>1: 4-Regular, 5-Malo                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Cualitativa-Ordinal   |
| 20     | OtraCEstExp      | Experiencia en cursos de estadística en otra carrera o institución          | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal   |
| 21     | OtraCEstcursos   | Número de cursos de estadística vistos en otra carrera o institución        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Cualitativa-Nominal   |
| 22     | OtraCEstcalific  | Desempeño en cursos de estadística vistos en otra carrera o institución     | 0: 1-Excelente, 2-Muy Bueno, 3-Bueno<br>1: 4-Regular, 5-Malo                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Cualitativa-Ordinal   |
| 23     | CarrActEstExp    | Experiencia en cursos de estadística en la carrera actual si es repitente   | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal   |
| 24     | CarrActEstcursos | Número de cursos de estadística vistos en la carrera actual si es repitente |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Cualitativa-Nominal   |



**Diccionario de datos**

| Numero | Variable                 | Descripción                                                                    | Categorías                                                                                                                                                                                                                                                                                                                                | Tipo                |
|--------|--------------------------|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 25     | CarrActEstcalific        | Desempeño en cursos de estadística vistos en la carrera actual si es repitente | 0: 1-Excelente, 2-Muy Bueno, 3-Bueno<br>1: 4-Regular, 5-Malo                                                                                                                                                                                                                                                                              | Cualitativa-Ordinal |
|        | Conceptos                | Forma en que tomó los conceptos de estadística vistos en bachillerato          | 0: Dentro de un curso de solo estadística<br>2: Como parte del curso de matemáticas<br>3: Dentro de otros cursos de ciencias (Química, Física, etc.)                                                                                                                                                                                      | Cualitativa-Nominal |
| 26     | Concestad                | Conceptos vistos dentro de un curso de solo Estadística                        | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 27     | Concemat                 | Conceptos vistos como parte del curso de matemáticas                           | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 28     | Conceotros               | Conceptos vistos dentro de otros cursos de ciencias (Química, Física, etc.)    | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 29     | TablaFrecVistoColegio    | Vio en el colegio el tema Tablas de frecuencias                                | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 30     | TablaFrecGradoDif        | Grado de dificultad en el tema Tablas de frecuencias                           | 0: 3-Fácil, 4-Muy fácil<br>1: 1-Muy difícil, 2-Difícil<br>NA: No recuerdo                                                                                                                                                                                                                                                                 | Cualitativa-Ordinal |
| 31     | GrafBarrHistVistoColegio | Vio en el colegio el tema Gráficos de barras e Histogramas                     | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 32     | GrafBarrHistGradoDif     | Grado de dificultad en el tema Gráficos de barras e Histogramas                | 0: 3-Fácil, 4-Muy fácil<br>1: 1-Muy difícil, 2-Difícil<br>NA: No recuerdo                                                                                                                                                                                                                                                                 | Cualitativa-Ordinal |
| 33     | MeMedMoVistoColegio      | Vio en el colegio el tema Media o promedio, mediana, moda                      | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 34     | MeMedMoGradoDif          | Grado de dificultad en el tema Media o promedio, mediana, moda                 | 0: 3-Fácil, 4-Muy fácil<br>1: 1-Muy difícil, 2-Difícil<br>NA: No recuerdo                                                                                                                                                                                                                                                                 | Cualitativa-Ordinal |
| 35     | RanVarDesVistoColegio    | Vio en el colegio el tema Rango, varianza, desviación estándar                 | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 36     | RanVarDesGradoDif        | Grado de dificultad en el tema Rango, varianza, desviación estándar            | 0: 3-Fácil, 4-Muy fácil<br>1: 1-Muy difícil, 2-Difícil<br>NA: No recuerdo                                                                                                                                                                                                                                                                 | Cualitativa-Ordinal |
| 37     | PermuVistoColegio        | Vio en el colegio el tema Permutaciones                                        | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 38     | PermuGradoDif            | Grado de dificultad en el tema Permutaciones                                   | 0: 3-Fácil, 4-Muy fácil<br>1: 1-Muy difícil, 2-Difícil<br>NA: No recuerdo                                                                                                                                                                                                                                                                 | Cualitativa-Ordinal |
| 39     | CombiVistoColegio        | Vio en el colegio el tema Combinaciones                                        | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 40     | CombiGradoDif            | Grado de dificultad en el tema Combinaciones                                   | 0: 3-Fácil, 4-Muy fácil<br>1: 1-Muy difícil, 2-Difícil<br>NA: No recuerdo                                                                                                                                                                                                                                                                 | Cualitativa-Ordinal |
| 41     | RegProbVistoColegio      | Vio en el colegio el tema Algunas reglas de probabilidad                       | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 42     | RegprobGradoDif          | Grado de dificultad en el tema reglas de probabilidad                          | 0: 3-Fácil, 4-Muy fácil<br>1: 1-Muy difícil, 2-Difícil<br>NA: No recuerdo                                                                                                                                                                                                                                                                 | Cualitativa-Ordinal |
| 43     | Temasvistos              | Conteo del número de conceptos o temas vistos                                  | 0: Menos de 4 temas vistos<br>1: 4 o más temas vistos                                                                                                                                                                                                                                                                                     | Cualitativa-Nominal |
|        | Causadifcil              | Razones que originaron la dificultad en los temas                              | 1: Los temas incluían operaciones matemáticas muy difíciles<br>2: Los problemas tratados no tenían relación con la vida real<br>3: No le entendía la explicación al profesor<br>4: No prestaba atención porque no le gustan las matemáticas<br>5: Resolver estos temas demanda mucho tiempo y prefería hacer otras cosas<br>99: No aplica | Cualitativa-Nominal |
| 44     | Causadifcila             | Los temas incluían operaciones matemáticas muy difíciles                       | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 45     | Causadifcilb             | Los problemas tratados no tenían relación con la vida real                     | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 46     | Causadifcile             | No le entendía la explicación al profesor                                      | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |

**Diccionario de datos**

| Numero | Variable          | Descripción                                                                                                               | Categorías                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Tipo                |
|--------|-------------------|---------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 47     | Causadifild       | No prestaba atención porque no le gustan las matemáticas                                                                  | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 48     | Causadifile       | Resolver estos temas demanda mucho tiempo y prefería hacer otras cosas                                                    | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 49     | Colegio           | Tipo de colegio                                                                                                           | 0: Privado<br>1: Oficial                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Cualitativa-Nominal |
| 50     | Cursosnecesimport | Considera que los cursos de estadística incluidos en la carrera que está estudiando son necesarios e importantes          | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 51     | Razonimport       | Razón más importante para considerar necesario que se incluya cursos de estadística en su carrera                         | 1: En el mundo laboral actual, el trabajo interdisciplinario y el manejo e interpretación de datos facilita la comunicación entre profesionales<br>2: Un aspecto importante en el ejercicio de la profesión es la toma de decisiones<br>3: Tomar buenas decisiones requiere de evidencia determinada a partir de la recolección de datos<br>4: Cuando esté ejerciendo la profesión y requiera de algún análisis estadístico, sólo será suficiente contratar a un profesional de la estadística<br>5: El enfoque que quiero dar a mi vida profesional, no requiere de conocimientos en estadística<br>6: La universidad considera que es importante incluir estadística en el currículo, pero para mí no lo es<br>7: Pienso que es importante, pero en realidad no sé por qué | Cualitativa-Nominal |
| 52     | FormaEstudio      | Cómo será su forma de estudiar, en cuanto al tiempo de trabajo fuera de clase que le piensa dedicar                       | 1: Estudiaré solo unas pocas horas antes de cada examen<br>2: Le dedicaré máximo una hora a la semana a esta materia<br>3: Le dedicaré al menos una hora a la semana a esta materia<br>4: Le dedicaré al menos 2 horas a la semana a esta materia<br>5: La materia no requiere que le dedique tiempo adicional para aprobarla                                                                                                                                                                                                                                                                                                                                                                                                                                                | Cualitativa-Nominal |
| 53     | AveriguaCurso     | Ha averiguado a amigos o estudiantes de semestres posteriores, acerca de cómo es el curso de estadística que va a iniciar | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
|        | Comentarios       | Algunos de los comentarios hechos por sus amigos o estudiantes de semestres posteriores fueron                            | a: La materia es fácil y no requiere de mucho trabajo para aprobarla<br>b: Es suficiente estudiar unas pocas horas antes de cada examen<br>c: No se le entiende a los profesores y es necesario buscar otros tutores<br>d: La materia requiere de mucho tiempo de dedicación y mucho esfuerzo adicional para aprobarla<br>e: Es necesario dedicar tiempo adicional suficiente para entender los temas y hacer los ejercicios propuestos<br>f: Con asistir a las tutorías de los profesores es suficiente                                                                                                                                                                                                                                                                     | Cualitativa-Nominal |
| 54     | Comentarioa       | La materia es fácil y no requiere de mucho trabajo para aprobarla                                                         | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 55     | Comentariob       | Es suficiente estudiar unas pocas horas antes de cada examen                                                              | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 56     | Comentarioc       | No se le entiende a los profesores y es necesario buscar otros tutores                                                    | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 57     | Comentariod       | La materia requiere de mucho tiempo de dedicación y mucho esfuerzo adicional para aprobarla                               | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 58     | Comentarioe       | Es necesario dedicar tiempo adicional suficiente para entender los temas y hacer los ejercicios propuestos                | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |
| 59     | Comentariof       | Con asistir a las tutorías de los profesores es suficiente                                                                | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Cualitativa-Nominal |

**Diccionario de datos**

| Numero | Variable         | Descripción                                                                       | Categorías                                                                                                                                                                                                                                                                                                                                                                        | Tipo                |
|--------|------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 60     | TrabajoRemun     | Tiene en la actualidad algún tipo de trabajo por el cual recibe remuneración      | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                    | Cualitativa-Nominal |
| 61     | ModalidadTrabajo | En caso de estar trabajando, está realizando                                      | 1: Monitoria en la universidad<br>2: Trabajo todas las noches entre 2 y 8 horas<br>3: Trabajo entre 2 y 8 horas al día en el negocio de la familia o propio<br>4: Trabajo entre 2 y 8 horas al día en negocio o empresa ajeno<br>5: Trabajo solo los fines de semana en el negocio de la familia o propio<br>6: Trabajo solo los fines de semana en negocio o empresa ajeno       | Cualitativa-Nominal |
| 62     | GradoDificEsp    | Cómo espera que sea el grado de dificultad del curso de estadística que matriculó | 1: Muy alto<br>2: Alto<br>3: Bajo<br>4: Muy bajo                                                                                                                                                                                                                                                                                                                                  | Cualitativa-Ordinal |
|        | ExplicGradoDific | Cuáles enunciados, podrían ser razón para explicar el grado de dificultad         | 1: Tengo buenos conocimientos en matemáticas<br>2: Estoy interesado y tengo una buena actitud hacia la carrera<br>3: Tengo confianza porque tuve buenos resultados en los cursos previos<br>4: Me parecen difíciles las materias que utilizan las matemáticas<br>5: No tengo el suficiente tiempo para dedicarle a la materia<br>6: No he visto nada de estadística anteriormente | Cualitativa-Nominal |
| 63     | Enunciadoa       | Tengo buenos conocimientos en matemáticas                                         | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                    | Cualitativa-Nominal |
| 64     | Enunciadob       | Estoy interesado y tengo una buena actitud hacia la carrera                       | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                    | Cualitativa-Nominal |
| 65     | Enunciadoc       | Tengo confianza porque tuve buenos resultados en los cursos previos               | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                    | Cualitativa-Nominal |
| 66     | Enunciadod       | Me parecen difíciles las materias que utilizan las matemáticas                    | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                    | Cualitativa-Nominal |
| 67     | Enunciadoe       | No tengo el suficiente tiempo para dedicarle a la materia                         | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                    | Cualitativa-Nominal |
| 68     | Enunciadof       | No he visto nada de estadística anteriormente                                     | 0: No<br>1: Si                                                                                                                                                                                                                                                                                                                                                                    | Cualitativa-Nominal |
| 69     | SentimCurso      | Cómo se siente ahora frente al curso de estadística que va a comenzar             | 1: Preocupado<br>2: Tenso<br>3: Tranquilo<br>4: Muy tranquilo                                                                                                                                                                                                                                                                                                                     | Cualitativa-Nominal |
| 70     | Parcial1         | Nota del primer parcial                                                           | 0: Aprobó<br>1: Reprobó                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal |
| 71     | Parcial2         | Nota del segundo parcial                                                          | 0: Aprobó<br>1: Reprobó                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal |
| 72     | Notafinal        | Nota final en el curso                                                            | 0: Aprobó<br>1: Reprobó                                                                                                                                                                                                                                                                                                                                                           | Cualitativa-Nominal |
| 73     | Notafinal2       | Nivel obtenido en el curso                                                        | 0: Alto (nota superior a 3,7)<br>1: Bajo (nota igual o inferior a 3,7)                                                                                                                                                                                                                                                                                                            | Cualitativa-Nominal |
| 74     | IN               | Índice de Ansiedad                                                                | 0: Ansiedad Manejable (AM)<br>1: Ansiedad Importante (AI)                                                                                                                                                                                                                                                                                                                         | Cualitativa-Nominal |
|        |                  |                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                   |                     |

## **C. Anexo: Pruebas de asociación**

PRUEBAS DE ASOCIACIÓN  
NOTAFIN - Nota final en el curso

| Variables                |           | Nota final en el curso de Estadística |        |             |       | CHI-CUADRADO<br>Valor-p | FISHER<br>Valor-p |
|--------------------------|-----------|---------------------------------------|--------|-------------|-------|-------------------------|-------------------|
|                          |           | Aprobó                                |        | Reprobó     |       |                         |                   |
|                          |           | Estudiantes                           | %      | Estudiantes | %     |                         |                   |
| Facultad                 | 1         | 72                                    | 91,1%  | 7           | 8,9%  | ,818                    | 0,812             |
|                          | 2         | 55                                    | 96,5%  | 2           | 3,5%  |                         |                   |
|                          | 3         | 21                                    | 91,3%  | 2           | 8,7%  |                         |                   |
|                          | 4         | 23                                    | 95,8%  | 1           | 4,2%  |                         |                   |
|                          | 5         | 47                                    | 92,2%  | 4           | 7,8%  |                         |                   |
|                          | 6         | 3                                     | 100,0% | 0           | 0,0%  |                         |                   |
| Carrera                  | 1         | 20                                    | 87,0%  | 3           | 13,0% | ,987                    | 0,9859            |
|                          | 2         | 5                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 3         | 35                                    | 89,7%  | 4           | 10,3% |                         |                   |
|                          | 4         | 12                                    | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 6         | 3                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 7         | 12                                    | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 8         | 12                                    | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 9         | 25                                    | 92,6%  | 2           | 7,4%  |                         |                   |
|                          | 10        | 3                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 11        | 8                                     | 88,9%  | 1           | 11,1% |                         |                   |
|                          | 12        | 4                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 14        | 1                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 15        | 5                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 16        | 2                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 17        | 16                                    | 94,1%  | 1           | 5,9%  |                         |                   |
|                          | 18        | 47                                    | 92,2%  | 4           | 7,8%  |                         |                   |
|                          | 19        | 1                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 21        | 2                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 23        | 8                                     | 88,9%  | 1           | 11,1% |                         |                   |
| Materia                  | 08131     | 113                                   | 92,6%  | 9           | 7,4%  | ,791                    | 0,8385            |
|                          | 08290     | 6                                     | 100,0% | 0           | 0,0%  |                         |                   |
|                          | 08293     | 20                                    | 90,9%  | 2           | 9,1%  |                         |                   |
|                          | 08295     | 35                                    | 97,2%  | 1           | 2,8%  |                         |                   |
|                          | 08299     | 47                                    | 92,2%  | 4           | 7,8%  |                         |                   |
| Semana                   | 1         | 76                                    | 96,2%  | 3           | 3,8%  | ,438                    | 0,4356            |
|                          | 2         | 124                                   | 91,9%  | 11          | 8,1%  |                         |                   |
|                          | 3         | 21                                    | 91,3%  | 2           | 8,7%  |                         |                   |
| Edad                     | 17 - 22   | 215                                   | 93,1%  | 16          | 6,9%  | 1                       | 1,000             |
|                          | 23 - 28   | 6                                     | 100,0% | 0           | 0,0%  |                         |                   |
| Genero                   | Femenino  | 102                                   | 95,3%  | 5           | 4,7%  | ,3699                   | ,304              |
|                          | Masculino | 119                                   | 91,5%  | 11          | 8,5%  |                         |                   |
| PrimariaMat              | Bueno     | 210                                   | 92,9%  |             | 7,1%  | ,7653                   | 1,000             |
|                          | No Bueno  | 11                                    | 100,0% | 0           | 0,0%  |                         |                   |
| BachillerMat             | Bueno     | 193                                   | 93,7%  | 13          | 6,3%  | ,486                    | ,447              |
|                          | No Bueno  | 28                                    | 90,3%  | 3           | 9,7%  |                         |                   |
| UniversidadMat           | Bueno     | 142                                   | 95,9%  | 6           | 4,1%  | ,033                    | ,058              |
|                          | No Bueno  | 79                                    | 88,8%  | 10          | 11,2% |                         |                   |
| PrimEstExp               | No        | 163                                   | 94,8%  | 9           | 5,2%  | ,130                    | ,149              |
|                          | Si        | 58                                    | 89,2%  | 7           | 10,8% |                         |                   |
| BachiEstExp              | No        | 23                                    | 88,5%  | 3           | 11,5% | ,302                    | ,396              |
|                          | Si        | 198                                   | 93,8%  | 13          | 6,2%  |                         |                   |
| OtraCEstExp              | No        | 209                                   | 93,3%  | 15          | 6,7%  | ,889                    | ,607              |
|                          | Si        | 12                                    | 92,3%  | 1           | 7,7%  |                         |                   |
| Concestad                | No        | 109                                   | 96,5%  | 4           | 3,5%  | ,060                    | ,072              |
|                          | Si        | 112                                   | 90,3%  | 12          | 9,7%  |                         |                   |
| Concemat                 | No        | 108                                   | 89,3%  | 13          | 10,7% | ,012                    | ,01762            |
|                          | Si        | 113                                   | 97,4%  | 3           | 2,6%  |                         |                   |
| Conceotros               | No        | 199                                   | 93,0%  | 15          | 7,0%  | ,629                    | 1,000             |
|                          | Si        | 22                                    | 95,7%  | 1           | 4,3%  |                         |                   |
| TablaFrecVistoColegio    | No        | 42                                    | 97,7%  | 1           | 2,3%  | ,201                    | ,317              |
|                          | Si        | 179                                   | 92,3%  | 15          | 7,7%  |                         |                   |
| GrafBarrHistVistoColegio | No        | 20                                    | 100,0% | 0           | 0,0%  | ,209                    | ,374              |
|                          | Si        | 201                                   | 92,6%  | 16          | 7,4%  |                         |                   |
| MeMedMoVistoColegio      | No        | 8                                     | 100,0% | 0           | 0,0%  | ,439                    | 1,000             |
|                          | Si        | 213                                   | 93,0%  | 16          | 7,0%  |                         |                   |

## PRUEBAS DE ASOCIACIÓN

**NOTAFIN - Nota final en el curso**

| Variables     |                     | Nota final en el curso de Estadística |       |             |      | CHI-CUADRADO | FISHER |
|---------------|---------------------|---------------------------------------|-------|-------------|------|--------------|--------|
|               |                     | Aprobó                                |       | Reprobó     |      |              |        |
|               |                     | Estudiantes                           | %     | Estudiantes | %    |              |        |
| RanVarDesVis  | No                  | 84                                    | 95,5% | 4           | 4,5% | ,298         | ,423   |
| toColegio     | Si                  | 137                                   | 91,9% | 12          | 8,1% |              |        |
| PermuVistoCol | No                  | 128                                   | 91,4% | 12          | 8,6% | ,180         | ,201   |
| egio          | Si                  | 93                                    | 95,9% | 4           | 4,1% |              |        |
| CombiVistoCol | No                  | 121                                   | 91,7% | 11          | 8,3% | ,276         | ,310   |
| egio          | Si                  | 100                                   | 95,2% | 5           | 4,8% |              |        |
| RegProbVisto  | No                  | 124                                   | 93,2% | 9           | 6,8% | ,991         | 1,000  |
| Colegio       | Si                  | 97                                    | 93,3% | 7           | 6,7% |              |        |
| Temasvistos   | #temas <4           | 74                                    | 96,1% | 3           | 3,9% | ,224         | ,279   |
|               | #temas >=4          | 147                                   | 91,9% | 13          | 8,1% |              |        |
| Colegio       | Privado             | 132                                   | 94,3% | 8           | 5,7% | ,445         | ,445   |
|               | Oficial             | 89                                    | 91,8% | 8           | 8,2% |              |        |
| Cursosnecesi  | No                  | 11                                    | 91,7% | 1           | 8,3% | ,823         | ,577   |
| mport         | Si                  | 210                                   | 93,3% | 15          | 6,7% |              |        |
| TrabajoRemun  | No                  | 185                                   | 92,5% | 15          | 7,5% | ,285         | ,478   |
|               | Si                  | 36                                    | 97,3% | 1           | 2,7% |              |        |
| IN            | Ansiedad Manejable  | 147                                   | 94,2% | 9           | 5,8% | ,403         | ,422   |
|               | Ansiedad Importante | 74                                    | 91,4% | 7           | 8,6% |              |        |

## **D. Anexo: Análisis Exploratorio de datos**

**Tabla D-1: Asignaturas Matriculadas**

| Asignatura                                    | Frecuencia | %            |
|-----------------------------------------------|------------|--------------|
| Teoría de Probabilidades                      | 195        | 53,0         |
| Fundamentos de Probabilidad y Estadística     | 14         | 3,8          |
| Estadística y Probabilidad Básica             | 41         | 11,1         |
| Bioestadística                                | 54         | 14,7         |
| Bioestadística y Fundamentos de Epidemiología | 64         | 17,4         |
| Total                                         | <b>368</b> | <b>100,0</b> |

**Tabla D-2: Características de los estudiantes**

|                    |              | Frecuencia | %    |
|--------------------|--------------|------------|------|
| Edad               | 17 - 22 años | 358        | 97,3 |
|                    | 23 - 28 años | 10         | 2,7  |
| Sexo               | Femenino     | 180        | 48,9 |
|                    | Masculino    | 188        | 51,1 |
| Colegio            | Privado      | 229        | 62,2 |
|                    | Oficial      | 139        | 37,8 |
| Trabajo Remunerado | No           | 296        | 80,4 |
|                    | Si           | 70         | 19,0 |
| Ansiedad           | Manejable    | 236        | 64,1 |
|                    | Importante   | 132        | 35,9 |

**Tabla D-3: Experiencia con Estadística**

|                             |    |            |      | Buen desempeño |      |
|-----------------------------|----|------------|------|----------------|------|
| Experiencia con Estadística |    | Frecuencia | %    | Frecuencia     | %    |
| Primaria                    | No | 253        | 68,8 | 94             | 88,7 |
|                             | Si | 106        | 28,8 |                |      |
| Bachillerato                | No | 41         | 11,1 | 279            | 85,8 |
|                             | Si | 325        | 88,3 |                |      |
| Otra Carrera                | No | 265        | 72,0 | 11             | 73,3 |
|                             | Si | 15         | 4,1  |                |      |

**Tabla D-4: Desempeño en cursos de Matemáticas**

| Desempeño en Matemáticas |          | Frecuencia | %           |
|--------------------------|----------|------------|-------------|
| Primaria                 | Bueno    | 346        | <b>94,0</b> |
|                          | No Bueno | 21         | 5,7         |
| Bachillerato             | Bueno    | 320        | <b>87,0</b> |
|                          | No Bueno | 48         | 13,0        |
| Universidad              | Bueno    | 199        | <b>54,1</b> |
|                          | No Bueno | 122        | 33,2        |



**Tabla D-5: Forma en que tomó los conceptos de estadística vistos en bachillerato**

|                          |    | Frecuencia | %    |
|--------------------------|----|------------|------|
| Curso de Estadística     | No | 176        | 47,8 |
|                          | Si | 180        | 48,9 |
| Curso de Matemáticas     | No | 180        | 48,9 |
|                          | Si | 176        | 47,8 |
| Otros cursos de Ciencias | No | 321        | 87,2 |
|                          | Si | 35         | 9,5  |

**Tabla D-6: Temas de Estadística vistos en bachillerato**

|                                      |            |      | Dificultad |      |
|--------------------------------------|------------|------|------------|------|
| Tema                                 | Frecuencia | %    | Frecuencia | %    |
| Tablas de frecuencias                | 299        | 81,3 | 23         | 6,3  |
| Gráficos de barras e Histogramas     | 333        | 90,5 | 20         | 5,4  |
| Media o promedio, mediana, moda      | 348        | 94,6 | 26         | 7,1  |
| Rango, varianza, desviación estándar | 228        | 62,0 | 71         | 19,3 |
| Permutaciones                        | 156        | 42,4 | 81         | 22,0 |
| Combinaciones                        | 169        | 45,9 | 88         | 23,9 |
| Algunas reglas de probabilidad       | 152        | 41,3 | 52         | 14,1 |

**Tabla D-7: Posibles causas que originaron dificultad**

| Causa                                                                     | Frecuencia | %    |
|---------------------------------------------------------------------------|------------|------|
| 1: Los temas incluían operaciones matemáticas muy difíciles               | 70         | 19,0 |
| 2: Los problemas tratados no tenían relación con la vida real             | 41         | 11,1 |
| 3: No le entendía la explicación al profesor                              | 159        | 43,2 |
| 4: No prestaba atención porque no le gustan las matemáticas               | 13         | 3,5  |
| 5: Resolver estos temas demanda mucho tiempo y prefería hacer otras cosas | 23         | 6,3  |

# Bibliografía

- Agresti, A. (1996), 'An introduction to categorical data analysis, 1996'.
- Agresti, A. (2002), *Categorical Data Analysis*, John Wiley.
- Agresti, A. (2015), *Foundations of linear and generalized linear models*, John Wiley & Sons.
- Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE transactions on automatic control* **19**(6), 716–723.
- Ayçaguer, L. C. S. (1994), *Excursión a la regresión logística en ciencias de la salud*, Ediciones Díaz de Santos.
- Batanero, C. (2000a), '¿hacia dónde va la educación estadística?', *Blaix* **15**(2), 13.
- Batanero, C. (2000b), 'Significado y comprensión de las medidas de posición central', *Uno. Revista de Didáctica de las Matemáticas* **25**, 41–58.
- Batanero, C. (2004), 'Los retos de la cultura estadística', *Yupana* **1**(1), 27–37.
- Behar Gutiérrez, R. & Grima Cintas, P. (2001), 'Mil y una dimensiones del aprendizaje de la estadística', *Estadística Española* **43**(148), 189–207.
- Behar Gutiérrez, R. & Grima Cintas, P. (2011), 'La estadística en la educación superior ¿formamos pensamiento estadístico?', *INGENIERÍA Y COMPETITIVIDAD* **5**(2), 84–90.  
**URL:** [http://praxis.univalle.edu.co/index.php/ingenieria\\_y\\_competitividad/article/view/2299](http://praxis.univalle.edu.co/index.php/ingenieria_y_competitividad/article/view/2299)
- Behar, R., Ojeda Ramírez, M. M. et al. (2011), 'El problema de la educación estadística: perspectivas desde el aprendizaje'.

- Bologna, E. L. & Vaiman, M. (2013), 'Actitudes, experiencia previa y nivel de logro en estadística en la carrera de psicología', *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* pp. 91–103.
- Castro R., A. (2017), 'Factores de riesgo para contraer infecciones de transmisión sexual: Una aproximación usando métodos propios del paradigma bayesiano', Universidad del Valle.
- Chavez M., V. F. (2017), 'Estimación de la proporción de muertes y de los tiempos hasta ocurrir la misma por cáncer de mama a partir del resultado de tres marcadores moleculares', Universidad del Valle.
- Cordoba C., L. A. (2015), 'Modelo de regresion binaria para estimar la probabilidad de recurrencia-persistencia en pacientes con cancer diferenciado de tiroides un ano despues del tratamiento', Universidad del Valle.
- Correa Morales, J. C., Causil, B. & Javier, C. (2018), *Introducción a la estadística bayesiana: notas de clase*, Instituto Tecnológico Metropolitano.
- De Lejarza, J. M. (2012), 'Modelos de probabilidad ii'.
- Dobson, A. J. & Barnett, A. (2008), *An introduction to generalized linear models*, CRC press.
- Fox, J. (2015), *Applied regression analysis and generalized linear models*, Sage Publications.
- Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian data analysis*, Chapman and Hall/CRC.
- Gil Flores, J. (1999), 'Actitudes hacia la estadística. incidencia de las variables sexo y formación previa', *Revista española de pedagogía* pp. 567–589.
- Guzmán, G., Sánchez de Guzmán, G. & Tovar, J. (2014), 'Relación entre la presencia de tiroiditis linfocítica con la gravedad y recurrencia/persistencia del carcinoma diferenciado de tiroides', *Revista argentina de endocrinología y metabolismo* **51**(4), 169–177.
- León Salazar, J. J. (2007), 'Enseñando estadística a futuros ingenieros', *Scientia et Technica* **13**(34).

- Madsen, H. & Thyregod, P. (2010), *Introduction to general and generalized linear models*, CRC Press.
- Martínez, D. R., Julio, L. A., Cabaleiro, J. C., Pena, T. F., Rivera, F. F. & Blanco, V. (2009), 'El criterio de información de akaike en la obtención de modelos estadísticos de rendimiento', *XX Jornadas de Paralelismo*.
- Martínez Mayoral, M. A. & Morales Socuélamos, J. (2001), *Modelos lineales generalizados*, Universidad Miguel Hernández.
- McCullagh, P. & Nelder, J. A. (1989), *Generalized linear models*, Vol. 37, CRC press.
- Montgomery, D., Peck, E. & Vining, G. G. (2006), 'Introducción al análisis de regresión lineal', *Compañía Editorial Continental. Tercera edición. México*.
- Márquez, J. C. (2004), 'Una revisión de las evidencias de fiabilidad y validez de los cuestionarios de actitudes y ansiedad hacia la estadística', *Statistics Education Research Journal* **3**(1), 5–28.
- Osorno, S. (2015), 'Preconceptos de los estudiantes sobre el primer curso de estadística en la universidad', Universidad del Valle.
- Roberts, D. M. & Saxe, J. E. (1982), 'Validity of a statistics attitude survey: A follow-up study', *Educational and psychological measurement* **42**(3), 907–912.
- Ross, S. M. (2014), *Introduction to probability models*, Academic press.
- Silva, L. & Barroso, I. (2004), 'Regresión logística, volumen 27 de cuadernos de estadística series', *La Muralla, Hespérides*.
- Tovar Cuevas, J. R. (2012), 'Eliciting beta prior distributions for binomial sampling', *Rev. Bras. Biom* **30**(1), 159–172.
- Tovar Cuevas, J. R. (2015), 'Inferencia bayesiana e investigación en salud: un caso de aplicación en diagnóstico clínico', *Revista Médica de Risaralda* **21**(1).

- Tovar Cuevas, J. R., Castillo Sánchez, H. & Marín, M. d. P. (2007), 'Preconcepciones de estudiantes de la pontificia universidad javeriana cali sobre el curso de estadística', *Pensamiento Psicológico* **3**(9).
- Valle Benavides, A. R. d. (2017), 'Curvas roc (receiver-operating-characteristic) y sus aplicaciones'.
- Vilà Baños, R. & Rubio Hurtado, M. (2016), 'Actitudes hacia la estadística en el alumnado del grado de pedagogía de la universidad de barcelona', *REDU. Revista de Docencia Universitaria* **14**(1), 131–150.
- URL:** <https://polipapers.upv.es/index.php/REDU/article/view/5766>