

**IDENTIFICACIÓN DE LA NECESIDAD DE APOYO ACADÉMICO EN LOS  
ESTUDIANTES DE GRADO DÉCIMO DEL COLEGIO SAN FRANCISCO  
JAVIER, ANTES DEL ICFES, UTILIZANDO ALGUNAS PRUEBAS  
ESTADÍSTICAS.**

**MÓNICA DEL CARMEN JURADO ESCUDERO**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
SAN JUAN DE PASTO  
2009**

**IDENTIFICACIÓN DE LA NECESIDAD DE APOYO ACADÉMICO EN LOS  
ESTUDIANTES DE GRADO DÉCIMO DEL COLEGIO SAN FRANCISCO  
JAVIER, ANTES DEL ICFES, UTILIZANDO ALGUNAS PRUEBAS  
ESTADÍSTICAS.**

**MÓNICA DEL CARMEN JURADO ESCUDERO  
CODIGO: 01832237**

**Trabajo de grado para optar al título de Especialista en Estadística**

**DIRECTOR  
JORGE HUMBERTO MAYORGA ALVAREZ**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
SAN JUAN DE PASTO  
2009**

## **FORMATO UNICO PARA ENTREGA DE LOS TRABAJOS DE GRADO**

**IDENTIFICACIÓN DE LA NECESIDAD DE APOYO ACADÉMICO EN LOS ESTUDIANTES DE GRADO DÉCIMO DEL COLEGIO SAN FRANCISCO JAVIER, ANTES DEL ICFES, UTILIZANDO ALGUNAS PRUEBAS ESTADÍSTICAS.**

**IDENTIFICATION OF THE ACADEMIC SUPPORTING NECESSITY ON STUDENTS IN TENTH GRADE FROM SAN FRANCISCO JAVIER HIGH SCHOOL – BEFORE ICFES TEST - BY USING SOME STATISTIC PROOFS.**

### Resumen

El presente trabajo aborda un estudio estadístico, cuyo fundamento es el modelo de regresión logística, por medio del cual se analiza la relación existente entre el desempeño académico de los estudiantes Javerianos desde el grado sexto hasta el grado décimo y los resultados obtenidos en las pruebas de Estado ICFES por parte de los estudiantes en el grado once. En el desarrollo de tal estudio se analizan diferentes modelos para finalmente escoger seis de ellos, uno por cada área del núcleo común, calculando la probabilidad que tiene cada estudiante del grado once de obtener un resultado por debajo del nivel superior. Estos modelos se constituyen en un instrumento objetivo por demás, que nos permite identificar a los estudiantes que necesiten de refuerzo académico en las áreas de conocimiento establecidas en el estudio antes de la prueba ICFES y poder así establecer en el Colegio, actividades preventivas que les permitan adquirir herramientas para mantener o mejorar el nivel en el que se encuentra la Institución actualmente.

### Summary

The following research has been focused on a statistic study based on a logistic regression model as a means to analyze the relationship between academic performance of the students in Javeriano High School - from sixth grade to tenth grade - and the results obtained in the ICFES tests sat by students in eleventh grade.

In the fulfillment of such a study different models are analyzed to choose six of them - each one per each common nucleus – by calculating the probability that

every student in eleventh grade has in order to get a lower result according to the highest level. These models are taken as objective tools which help identify those students who need academic reinforcement in the knowledge areas established on this study before the ICFES test. So that to be able to set up preventive tasks in the high school aimed to get tools to maintain or improve the level in which the institution is currently.

Palabras claves: Regresión logística, refuerzo académico, Actividades preventivas  
Pruebas de estado – Colombia, Examen de admisión a la  
Universidad

Key words: Logistic regresión, Academic reinforcement, Preventive tasks,  
State tests – Colombia, Higher education, University-admission test

**FIRMA DEL DIRECTOR:** \_\_\_\_\_

**Autora: MÓNICA DEL CARMEN JURADO ESCUDERO - 1975**

## TABLA DE CONTENIDO

1. INTRODUCCIÓN	
2. MARCO TEÓRICO	13
2.1. CONTEXTO EDUCATIVO	13
2.2. MODELO ESTADÍSTICO	22
2.2.1. MODELO DE REGRESIÓN LOGÍSTICA (LOGIT)	22
Pruebas de bondad y ajuste	30
Pruebas de ajuste del modelo	30
Pruebas sobre subconjuntos de parámetros	31
Coeficiente de determinación	32
Pruebas individuales de parámetros	33
3. METODOLOGÍA	35
3.1. RECOLECCIÓN DE LA INFORMACIÓN	37
3.2. CONSTRUCCIÓN DEL MODELOS DE REGRESIÓN LOGÍSTICA PARA	37
3.3. NOMENCLATURA DE VARIABLES	37
4. MODELOS OBTENIDOS	38
4.1. MODELO PARA EL ÁREA DE MATEMÁTICAS	38
4.2. MODELO PARA EL ÁREA DE CIENCIAS	60
4.2.1. MODELO BIOLÓGIA	60
4.2.2. MODELO QUÍMICA	83
4.2.3. MODELO FÍSICA	104
4.3. MODELO PARA EL ÁREA DE SOCIALES	125
4.4. MODELO PARA EL ÁREA DE LENGUAJE	147
5. CONCLUSIONES	168
BIBLIOGRAFÍA	169

## LISTA DE TABLAS

TABLA 1. DESCRIPCIÓN DE VARIABLES INDEPENDIENTES	38
TABLA 2. DESCRIPCIÓN DE VARIABLES DEPENDIENTES.	38
TABLA 3. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	39
TABLA 4. ANÁLISIS DE VARIANZA PARA LA REGRESIÓN	40
TABLA 5. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	41
TABLA 6. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	42
TABLA 7. BONDAD DE AJUSTE	44
TABLA 8. VALOR DEL CHI-CUADRADO EN CADA MODELO PRESENTE EN EL ESTUDIO	44
TABLA 9. TABLA DE CONTINGENCIAS PARA LA PRUEBA DE HOSMER Y LEMESHOW	45
TABLA 10. VALOR DEL -2LOG VEROSIMILITUD EN CADA MODELO PRESENTE EN EL ESTUDIO	46
TABLA 11. TABLA DE CLASIFICACIÓN PARA IM	46
TABLA 12. LISTADO DE CASOS MAL CLASIFICADOS	47
TABLA 13. TOTAL DE CASOS MAL CLASIFICADOS EN CADA MODELO PRESENTE EN EL ESTUDIO	50
TABLA 14. RESIDUOS ATÍPICOS PARA IM	51
TABLA 15. CAPACIDAD DE PREDICCIÓN DEL MODELO	52
TABLA 16. MATRIZ DE CORRELACIÓN PARA LOS COEFICIENTES ESTIMADOS	54
TABLA 17. VARIABLES EN EL MODELO	55
TABLA 18. CASOS POSIBLES QUE NECESITAN REFUERZO ACADÉMICO EN MATEMÁTICAS ANTES DE LAS PRUEBAS ICFES	57
TABLA 19. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	61
TABLA 20. ANÁLISIS DE VARIANZA PARA LA REGRESIÓN	62
TABLA 21. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	62
TABLA 22. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	63
TABLA 23. BONDAD DE AJUSTE	66
TABLA 24. VALOR DEL CHI-CUADRADO EN CADA MODELO PRESENTE EN EL ESTUDIO	66
TABLA 25. TABLA DE CONTINGENCIAS PARA LA PRUEBA DE HOSMER Y LEMESHOW	67
TABLA 26. TABLA DE CONTINGENCIAS PARA LA PRUEBA DE HOSMER Y LEMESHOW	67
TABLA 27. VALOR DEL -2LOG VEROSIMILITUD EN CADA MODELO PRESENTE EN EL ESTUDIO	69
TABLA 28. TABLA DE CLASIFICACIÓN PARA IB. MODELO 1	69
TABLA 29. TABLA DE CLASIFICACIÓN PARA IB. MODELO 3	69
TABLA 30. LISTADO DE CASOS MAL CLASIFICADOS	70

TABLA 31. TOTAL DE CASOS MAL CLASIFICADOS EN CADA MODELO PRESENTE EN EL ESTUDIO	73
TABLA 32. RESIDUOS ATÍPICOS PARA IB	74
TABLA 33. CAPACIDAD DE PREDICCIÓN DEL MODELO	75
TABLA 34. MATRIZ DE CORRELACIÓN PARA LOS COEFICIENTES ESTIMADOS	77
TABLA 35. VARIABLES EN EL MODELO	78
TABLA 36. CASOS POSIBLES QUE NECESITAN REFUERZO ACADÉMICO EN BIOLOGÍA ANTES DE LAS PRUEBAS ICFES	79
TABLA 37. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	83
TABLA 38. ANÁLISIS DE VARIANZA PARA LA REGRESIÓN	84
TABLA 39. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	85
TABLA 40. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	86
TABLA 41. BONDAD DE AJUSTE	88
TABLA 42. VALOR DEL CHI-CUADRADO EN CADA MODELO PRESENTE EN EL ESTUDIO	89
TABLA 43. TABLA DE CONTINGENCIAS PARA LA PRUEBA DE HOSMER Y LEMESHOW	89
TABLA 44. VALOR DEL -2LOG VEROSIMILITUD EN CADA MODELO PRESENTE EN EL ESTUDIO	90
TABLA 45. TABLA DE CLASIFICACIÓN PARA IQ	91
TABLA 46. LISTADO DE CASOS MAL CLASIFICADOS	91
TABLA 47. TOTAL DE CASOS MAL CLASIFICADOS EN CADA MODELO PRESENTE EN EL ESTUDIO	95
TABLA 48. RESIDUOS ATÍPICOS PARA IQ	95
TABLA 49. CAPACIDAD DE PREDICCIÓN DEL MODELO	96
TABLA 50. MATRIZ DE CORRELACIÓN PARA LOS COEFICIENTES ESTIMADOS	98
TABLA 51. VARIABLES EN EL MODELO	99
TABLA 52. CASOS POSIBLES QUE NECESITAN REFUERZO ACADÉMICO EN QUÍMICA ANTES DE LAS PRUEBAS ICFES	100
TABLA 53. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	104
TABLA 54. ANÁLISIS DE VARIANZA PARA LA REGRESIÓN	106
TABLA 55. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	106
TABLA 56. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	107
TABLA 57. BONDAD DE AJUSTE	109
TABLA 58. VALOR DEL CHI-CUADRADO EN CADA MODELO PRESENTE EN EL ESTUDIO	110
TABLA 59. TABLA DE CONTINGENCIAS PARA LA PRUEBA DE HOSMER Y LEMESHOW	110
TABLA 60. VALOR DEL -2LOG VEROSIMILITUD EN CADA MODELO PRESENTE EN EL ESTUDIO	111
TABLA 61. TABLA DE CLASIFICACIÓN PARA IF.	112
TABLA 62. LISTADO DE CASOS MAL CLASIFICADOS	112

TABLA 63. TOTAL DE CASOS MAL CLASIFICADOS EN CADA MODELO PRESENTE EN EL ESTUDIO	116
TABLA 64. RESIDUOS ATÍPICOS PARA IF	116
TABLA 65. CAPACIDAD DE PREDICCIÓN DEL MODELO	117
TABLA 66. MATRIZ DE CORRELACIÓN PARA LOS COEFICIENTES ESTIMADOS	119
TABLA 67. VARIABLES EN EL MODELO	120
TABLA 68. CASOS POSIBLES QUE NECESITAN REFUERZO ACADÉMICO EN FÍSICA ANTES DE LAS PRUEBAS ICFES	121
TABLA 69. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	125
TABLA 70. ANÁLISIS DE VARIANZA PARA LA REGRESIÓN	127
TABLA 71. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	127
TABLA 72. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	128
TABLA 73. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	130
TABLA 74. BONDAD DE AJUSTE	131
TABLA 75. VALOR DEL CHI-CUADRADO EN CADA MODELO PRESENTE EN EL ESTUDIO	132
TABLA 76. TABLA DE CONTINGENCIAS PARA LA PRUEBA DE HOSMER Y LEMESHOW	132
TABLA 77. VALOR DEL -2LOG VEROSIMILITUD EN CADA MODELO PRESENTE EN EL ESTUDIO	133
TABLA 78. TABLA DE CLASIFICACIÓN PARA IS	134
TABLA 79. LISTADO DE CASOS MAL CLASIFICADOS	134
TABLA 80. TOTAL DE CASOS MAL CLASIFICADOS EN CADA MODELO PRESENTE EN EL ESTUDIO	138
TABLA 81. RESIDUOS ATÍPICOS PARA IS	138
TABLA 82. CAPACIDAD DE PREDICCIÓN DEL MODELO	139
TABLA 83. MATRIZ DE CORRELACIÓN PARA LOS COEFICIENTES ESTIMADOS	141
TABLA 84. VARIABLES EN EL MODELO	142
TABLA 85. CASOS POSIBLES QUE NECESITAN REFUERZO ACADÉMICO EN SOCIALES ANTES DE LAS PRUEBAS ICFES	143
TABLA 86. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	147
TABLA 87. ANÁLISIS DE VARIANZA PARA LA REGRESIÓN	148
TABLA 88. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	149
TABLA 89. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	150
TABLA 90. RESULTADOS DE LAS ESTIMACIONES DEL MODELO	151
TABLA 91. BONDAD DE AJUSTE	153
TABLA 92. VALOR DEL CHI-CUADRADO EN CADA MODELO PRESENTE EN EL ESTUDIO	153
TABLA 93. TABLA DE CONTINGENCIAS PARA LA PRUEBA DE HOSMER Y LEMESHOW	154
TABLA 94. VALOR DEL -2LOG VEROSIMILITUD EN CADA MODELO PRESENTE EN EL ESTUDIO	155
TABLA 95. TABLA DE CLASIFICACIÓN PARA IL	155

TABLA 96. LISTADO DE CASOS MAL CLASIFICADOS	156
TABLA 97. TOTAL DE CASOS MAL CLASIFICADOS EN CADA MODELO PRESENTE EN EL ESTUDIO	159
TABLA 98. RESIDUOS ATÍPICOS PARA IL	159
TABLA 99. CAPACIDAD DE PREDICCIÓN DEL MODELO	160
TABLA 100. MATRIZ DE CORRELACIÓN PARA LOS COEFICIENTES ESTIMADOS	162
TABLA 101. VARIABLES EN EL MODELO	163
TABLA 102. CASOS POSIBLES QUE NECESITAN REFUERZO ACADÉMICO EN LENGUAJE ANTES DE LAS PRUEBAS ICFES	164

## LISTA DE FIGURAS

FIGURA 1. CAPACIDAD DE PREDICCIÓN DEL MODELO PARA IM	52
FIGURA 2. CAPACIDAD DE PREDICCIÓN DEL MODELO PARA IB	75
FIGURA 3. CAPACIDAD DE PREDICCIÓN DEL MODELO PARA IQ	97
FIGURA 4. CAPACIDAD DE PREDICCIÓN DEL MODELO PARA IF	118
FIGURA 5. CAPACIDAD DE PREDICCIÓN DEL MODELO PARA IS	140
FIGURA 6. CAPACIDAD DE PREDICCIÓN DEL MODELO PARA IL	161

## 1. INTRODUCCIÓN

El principio básico que lidera el Colegio Javeriano, se fundamenta en el MAGUIS IGNACIANO, “Ser más para servir mejor”; bajo este precepto, se busca la excelencia académica y para la consecución de esta misión se desarrollan procesos de aprendizaje acordes con la propuesta educativa de la Compañía de Jesús y los lineamientos generales propuestos por el Ministerio de Educación Nacional.

Uno de los procesos que nos ayudan a velar por esta excelencia académica son las pruebas ICFES, quienes anualmente nos retroalimentan para así poder realizar actividades correctivas en aquellas áreas que lo ameriten; sin embargo, el principal reto en la búsqueda de la excelencia es poder prevenir esas falencias académicas, antes de la medición que realiza el ICFES.

Bajo este contexto, es indispensable poder determinar con anterioridad a la prueba, qué estudiantes tienen una alta probabilidad de obtener resultados que hagan a la institución bajar del nivel en el que se encuentra y poder con ellos implementar algunas actividades de refuerzo académico (acciones preventivas) antes de las pruebas ICFES.

Actualmente en la institución no existe un instrumento que ayude a la identificación de estos estudiantes; normalmente la selección se la realiza, bajo supuestos que se fundamentan en las vivencias personales, en la interacción de docentes y estudiantes en la que generalmente se particularizan situaciones, tendiendo a seleccionar a aquellos estudiantes que únicamente se han destacado por obtener un bajo desempeño académico y de normalización en su vida escolar, aquellos que son minoría y fácil de identificar, específicamente en los dos últimos años

cursados (10° y 11°), ignorando su proceso real, transformando esta actividad en una simple selección carente de objetividad.

En este orden de ideas, el presente trabajo aborda un estudio estadístico, cuyo fundamento es el modelo de regresión logística, por medio del cual se analiza la relación existente entre el desempeño académico de los estudiantes Javerianos desde el grado sexto hasta el grado décimo y los resultados obtenidos en las pruebas de Estado ICFES por parte de los estudiantes cuando están en el grado once. En el desarrollo de tal estudio se analizan diferentes modelos para finalmente escoger seis de ellos, uno por cada área del núcleo común, calculando la probabilidad que tiene cada estudiante del grado once de obtener un resultado por debajo del nivel superior. Estos modelos se constituyen en un instrumento objetivo por demás, que nos permite identificar a los estudiantes que necesiten de refuerzo académico en las áreas de conocimiento establecidas en el estudio antes de la prueba ICFES y poder así establecer en el Colegio, actividades preventivas que les permitan adquirir herramientas para mantener o mejorar el nivel en el que se encuentra la Institución actualmente.

## 2. MARCO TEÓRICO

### 2.1. CONTEXTO EDUCATIVO

La última década se ha caracterizado por los enormes esfuerzos que Colombia ha hecho en materia educativa, teniendo como precepto el mandato constitucional de garantizar a todos los(as) niños(as) y los(as) jóvenes el derecho de recibir una educación de calidad, que les permita vivir e interactuar en la sociedad en igualdad de condiciones, y continuar aprendiendo durante toda la vida.

En concordancia con lo anterior, la Ley 115 de 1994 estableció el objeto y los fines de la educación.

*“Objeto: La educación es un proceso de formación permanente, personal, cultural, y social que se fundamenta en una concepción integral de la persona humana, de su dignidad, de sus derechos y de sus deberes”<sup>1</sup>*

Fines: los cuales son 13 que describen las características que todo estudiante Colombiano debe adquirir a lo largo de su formación educativa, características de formación social, humanista, técnica, artística y científica que permiten la integralidad del ser humano.

Además la ley 115, definió un conjunto de áreas obligatorias y fundamentales del conocimiento y dejó abierta la posibilidad de introducir asignaturas optativas, pertinentes y necesarias de acuerdo con las características locales donde se desarrolla la acción escolar. De la misma manera, la Ley dio

---

<sup>1</sup> Ministerio de educación. Ley General de educación 1994.

autonomía a las instituciones educativas para definir, en el marco de lineamientos curriculares y normas técnicas producidas por el Ministerio de Educación Nacional, su propio Proyecto Educativo Institucional (PEI).

En el centro de la discusión sobre cómo mejorar la calidad está la pregunta ¿qué saberes y competencias deben desarrollar los estudiantes como resultado de su paso por los diferentes grados y ciclos escolares? Por tratarse de educación para todos, el preescolar, la básica y la media deben proporcionar a toda la población estudiantil las mismas oportunidades de aprendizaje y desarrollo individual y social. De allí que sea conveniente contar con pautas o normas comunes, precisas y básicas para estos tres niveles educativos.

Lo anterior motivó al Ministerio de Educación Nacional a desarrollar estándares curriculares<sup>2</sup>, con los cuales busca concretar los lineamientos expedidos, de manera que las instituciones escolares cuenten con una información común para formular sus planes de estudio de acuerdo con sus prioridades educativas establecidas en el PEI.

El punto de partida fueron los lineamientos curriculares producidos por el Ministerio de Educación Nacional, pero su desarrollo se enriqueció tanto con la participación de maestros de diversas regiones y de académicos con la consulta de currículos de otros países, En todo momento se buscó construir unos estándares propios que respondan a nuestra realidad y expectativas, pero sin dar la espalda a los avances y desarrollos que en materia curricular se vienen detectando en diferentes partes del mundo.

Los estándares curriculares son criterios que especifican lo que todos los estudiantes de educación preescolar, básica y media deben saber y ser capaces

---

<sup>2</sup> Ministerio de educación. Estándares para la excelencia educativa.

de hacer en una determinada área y grado. Se traducen en formulaciones claras, universales, precisas y breves, que expresan lo que debe hacerse y cuán bien debe hacerse. Están sujetos a la verificación; por lo tanto, también son referentes para la construcción de sistemas y procesos de evaluación interna y externa, consistentes con las acciones educativas.

La noción de estándar curricular hace referencia a una meta que expresa, en forma observable, lo que el estudiante debe saber, es decir, los conceptos básicos de cada área, así como las competencias, entendidas como el saber hacer, utilizando esos conceptos. La noción de logro, por otra parte, hace referencia al nivel en el cual los estudiantes alcanzan una determinada meta o estándar.

Con los estándares curriculares no se pretende 'uniformar' la educación; con ellos se busca contar con un referente común, que asegure a todos el dominio de conceptos y competencias básicas para vivir en sociedad y participar en ella en igualdad de condiciones. Las instituciones educativas, en el marco de su PEI, son autónomas para elegir sus enfoques y estrategias pedagógicas<sup>3</sup>.

Al establecerse los estándares curriculares, se comienza a hablar de la calidad en la educación. La educación es de calidad cuando todos los niños y jóvenes, independientemente de sus condiciones socioeconómicas y culturales, alcanzan los objetivos propuestos por el sistema educativo, los cuales están establecidos en la Ley General de Educación, y realizan aprendizajes útiles para su vida y para la sociedad. Esto significa desarrollar competencias básicas para:

- Comprender lo que leen, expresarse en forma oral y escrita, calcular y resolver problemas (competencias básicas).
- Convivir con otros, trabajar y decidir en grupo (competencias interpersonales).

---

<sup>3</sup> Ministerio de educación. Estándares para la excelencia educativa.

- Actuar con responsabilidad, integridad y autocontrol, y desarrollar la autoestima (cualidades personales).

Por tanto, el Gobierno Nacional ha implementado una serie de evaluaciones censales de competencias de los alumnos de primaria y de básica secundaria el cual se constituye en el programa estratégico de la actual política educativa del gobierno nacional para mejorar la calidad de la educación básica a partir de la elaboración, por parte de las instituciones educativas, de planes de mejoramiento con base en los resultados de las evaluaciones; estas evaluaciones son conocidas como pruebas SABER, las cuales son aplicadas a estudiantes de grado: 3°,5°,7° y 9°.

El Colegio San Francisco Javier de la ciudad de Pasto, de carácter privado, el cual forma parte de los colegios de la Compañía de Jesús, que se enmarcan en un sistema educativo integrado por una red de instituciones que a partir de su experiencia y de una unidad de orientación, comparten una visión y unas finalidades comunes, la preocupación por encontrar y adoptar los principios y los métodos más eficaces para realizar los objetivos y las funciones educativas que les son propias, buscando la participación activa de la persona en su proceso educativo que supone asumir propósitos de excelencia académica, humana y de servicio a la sociedad; esto implica superar actitudes y posturas pasivas; asumir con responsabilidad y decisión, el proceso educativo por el cual libremente optó y, a partir de la reflexión, la crítica, y la investigación ofrecer respuestas creativas a los desafíos que la vida y el mundo plantean<sup>4</sup>. Esta propuesta educativa se rige en las normas de educación vigentes para el sector público en la prestación del servicio académico y busca primordialmente el desarrollo integral de la persona en aras de la excelencia (ser más para servir mejor, Magis Ignaciano<sup>5</sup>), por lo que es reconocido como una de las mejores instituciones educativas de la región. Como

---

<sup>4</sup> Propuesta educativa de la Compañía de Jesús. fundamentos y practica. Nov 2005

<sup>5</sup> El proceso del Magis Ignaciano y el mejoramiento continuo.

caso particular, busca la excelencia académica, fundamentando el desarrollo de competencias en sus estudiantes y preparándolos para un excelente desempeño en su vida universitaria y profesional.

El Colegio San Francisco Javier, busca el mejoramiento continuo de cada acción, idea y sentimiento que se suscitan en sus estudiantes, par lograr la excelencia integral de los mismos, por esta razón, ha optado, por construir un sistema de gestión de calidad en el servicio prestado, bajo la norma ISO9000:2000, cuyo objeto primordial, es al igual que el Magis, el mejoramiento continuo, establen procesos de evaluación y auto evaluación permanente que permiten identificar, debilidades y por ende, se pueden establecer planes de mejoramientos, acciones preventivas y correctivas en el servicio educativo que se presta a la comunidad de Pasto.

El sistema de gestión de calidad permite el aseguramiento, la evaluación y la mejora permanente de la calidad del servicio ofrecido por una organización a la luz de los fines de la institución y con un conjunto de estándares de desempeño como referencia<sup>6</sup>.

Bajo el criterio de la certificación de calidad según la norma ISO9000:2000, el Colegio San Francisco Javier ha establecido como un indicador de evaluación de la calidad en el servicio educativo, el nivel designado por el ICFES<sup>7</sup>, valoración que se determina por los resultados ICFES<sup>8</sup> de sus estudiantes. El cual, desde marzo de 2000, permite entregar, por alumno, cinco (5) tipos de resultados con 77 datos. A los planteles educativos les agrupa los resultados de sus estudiantes en cinco (5) informativos, que son:

- Puntaje en las pruebas del Núcleo Común y de la prueba interdisciplinar

---

<sup>6</sup> Propuesta educativa de la Compañía de Jesús. fundamentos y practica. Nov 2005

<sup>7</sup> Procedimiento, prestación y planificación del servicio.

<sup>8</sup> Pag. Web del ICFES

- Distribución de frecuencias de los puntajes en las pruebas del núcleo común
- Resultados por grupos de preguntas
- Distribución de los niveles de competencia en las pruebas del núcleo común
- Distribución de frecuencias de los grados de profundización

Son los resultados que los estudiantes obtienen en las pruebas que presentan todos, las cuales configuran el NÚCLEO COMÚN, los que permiten ahora comparaciones. Sin embargo, se ha insistido en que no deben sumarse ni promediarse porque en cada prueba se mide bajo determinados paradigmas y criterios o estándares lo que el estudiante sabe hacer independientemente en las disciplinas que se evalúan.

Para clasificar los planteles educativos entonces se toman los puntajes de sus alumnos en el núcleo común con excepción del idioma extranjero que es obligatorio pero optativo.

La clasificación y ordenación de planteles desde el año 2000, obedece en síntesis, a las siguientes características:

- a) Tener presente que es diferente, en términos de la exigencia de las pruebas, responder a preguntas muy complejas en comparación con otras menos complejas.
- b) Establecer rangos de diferente amplitud en la escala, de tal manera que los rangos más cortos se encuentren en la parte de mayor concentración de alumnos en la distribución y los rangos más largos en la parte de menor concentración (los extremos de la escala)
- c) Definir las categorías de rendimiento para cada área evaluada en las cuales se ubican los planteles en el año 2000 según el análisis de las distribuciones de puntajes en las pruebas del núcleo común tanto a nivel nacional como

departamental. Se determinan 14 categorías, de 1 a 14, siendo la 1 la más baja y la 14 la más alta.

- d) Distribuir los resultados de los estudiantes en 14 rangos de puntuación, para la escala de 0 a 100 puntos, que es utilizada para calificar en cada una de las pruebas. En algunos puntos de la escala cada rango es de 10 puntos y otros de esto depende de la forma en que se concentra la mayoría y se distribuyen los demás en el país y los departamentos.
- e) Determinar el ÍNDICE DE LOGRO con base en la distribución por rangos, el cual indica dónde queda el colegio en cada prueba. Para ello se determina el rango que alcanza cada estudiante y la proporción de estudiantes que tiene el plantel en cada rango o categoría.
- f) Determinar la categoría del plantel, la cual será más alta en cada prueba, según si tiene mayor número de estudiantes en rangos superiores.
- g) Establecer la ubicación definitiva por prueba comparando la distribución de los estudiantes de cada institución con distribuciones ideales para cada una de las categorías, que son un criterio externo a la prueba y a la misma población. Es decir, que la categoría en que quede un colegio en particular no depende de la clasificación de colegios en cada prueba, o lo que es lo mismo, es independiente de la clasificación de los demás planteles.

Con el propósito de mantener la clasificación general de los planteles en siete categorías (**muy superior, superior, alto, medio, bajo, inferior y muy inferior**) y ordenar los colegios, se generó un nuevo índice el cual se produce en una escala de 0 a 1. El índice permite ordenar los colegios y al establecerse 6 puntos de corte se diferencian las 7 categorías de clasificación. Conviene advertir que en esta

ordenación no se establece tampoco diferencia significativa estadísticamente colegio a colegio.

Es importante aclarar que solamente figurarán en la clasificación aquellos planteles educativos que inscribieron en grupo a los estudiantes de su plantel, de acuerdo con los procedimientos establecidos por el ICFES.

Los planteles educativos son considerados como elementos centrales en el desarrollo del proceso educativo y en los que se generan las relaciones e interrelaciones que condicionan la calidad educativa. Sin embargo, pese a la importancia de la escuela, no puede pasarse por alto los condicionamientos que le generan las disposiciones legislativas, las bases presupuestales, las políticas y disposiciones de las autoridades administrativas, el contexto en que se ubica el centro educativo y la extracción socio-económica de la población a la que atiende. Por ello, la valoración de la calidad de los colegios no puede realizarse en forma aislada, sin tomar en consideración los condicionantes a los que el centro educativo se halla sometido.

El ICFES recomienda tener presente que la clasificación de los planteles, con base en los resultados de los alumnos en los exámenes de estado para ingreso a la educación superior, es un indicador necesario pero no suficiente para evaluar la calidad de la educación que imparte un plantel.

Estas clasificaciones en categorías no son el resultado de la comparación entre colegios sino de la ubicación del colegio frente a parámetros teóricos definidos según los principios de la curva normal. Esto quiere decir, que los resultados de los alumnos de cada colegio se comparan con puntos reconocidos del modelo de la curva normal a partir de datos como el promedio, la variabilidad y el número de estudiantes.

## **Sistema de calificación y promoción del estudiante**

La educación, por tratarse de un proceso humano, se encuentra inveteradamente sujeta a transformaciones permanentes mediadas por la necesidad de mejorarla. En esa medida, se recurre con frecuencia a la revisión de teorías, concepciones, métodos, procedimientos e instrumentos, en la perspectiva de buscar una correspondencia adecuada con las necesidades educativas de la población, de acuerdo con las exigencias propias de cada época de desarrollo y según los contextos. Es aceptado también que Colombia requiere con urgencia participar en los cambios que le permitan articularse a los procesos de desarrollo social, científico, tecnológico y productivo del resto de países, con la mira de mejorar y alcanzar los niveles de bienestar y calidad de vida que la población.

Ante esto se generan nuevas políticas de evaluación y promoción donde se elimina la valoración cuantitativa; con el decreto 230 de 2002 en el cual se establece la evaluación integral, sistemática y continua<sup>9</sup> de carácter cualitativo; en un primer momento se crea la escala: I de insuficiente, A de aceptable, B de bueno y E de excelente. Posteriormente se da la escala D de deficiente, I de insuficiente, A de aceptable, S de sobresaliente y E de excelente<sup>10</sup>.

Esta forma de evaluación se convierte en un proceso de carácter subjetivo puesto que la valoración de un estudiante de una u otra manera está sujeta al proceso que desarrolla él mismo y el criterio del docente al respecto. La evaluación por logros, con sus respectivas recuperaciones, ha provocado un descenso en el interés de los alumnos y hasta de los docentes y menor dedicación de parte y parte, por cuanto hay un final más o menos previsible".

---

<sup>9</sup> Ministerio de Educación Nacional. Serie documento de trabajo, .La evaluación en el aula y más allá de ella., Impreandes Presencia, Santafé de Bogotá, D. C., 1997

<sup>10</sup> Finalidades y alcances del Decreto 230 del 11 de febrero de 2002

De acuerdo con la percepción de los ciudadanos, el mayor problema del sistema de evaluación vigente consiste en "calificar a los estudiantes con conceptos y no con notas numéricas, unido al hecho de que el 95% de los alumnos debe ser promovido al siguiente grado. Por tanto, en la actualidad se propone el Plan Decenal de Educación<sup>11</sup>, el cual plantea modificar el Decreto 230 de 2002 y reglamentar el sistema de evaluación y promoción de estudiantes de manera que responda a las metas de calidad, permanencia y cobertura, y la aplicación de pruebas que evalúen el nivel de logro de aprendizajes por competencias.

En el colegio Javeriano, durante los últimos tres años ha implementado en su proceso de evaluación un rango numérico a cada valoración cualitativa propuesta por el MEN, especificada a continuación:

- E: (de 4.6 a 5,0)
- S: (de 4.0 a 4,5)
- A: (de 3,9 a 3.0)
- I: (de 1.0 a 2,9)
- D: (de 0 a 0,9)

Esta escala se utilizó para poder transformar las variables cualitativas de la base de datos a variables cuantitativas.

## **2.2. MODELO ESTADÍSTICO**

### **2.2.1. MODELO DE REGRESIÓN LOGÍSTICA (LOGIT)<sup>12</sup>**

La regresión logística es un instrumento estadístico de análisis multivariado, de uso tanto explicativo como predictivo. Resulta útil su empleo cuando se tiene una variable dependiente dicotómica Y, esto es, una respuesta binaria del tipo 0/1 y un

---

<sup>11</sup> Plan Decenal de Educación

<sup>12</sup> Montgomery (2000).

conjunto de variables predictoras o independientes ( $X_1, X_2, X_3, \dots, X_n$ ), en las cuales no se establece ninguna restricción, pudiendo ser cuantitativas, tanto continuas como discretas, y categóricas, con dos o más modalidades. Este método estudia la relación existente entre las variables mencionadas anteriormente.

El propósito del análisis consiste en:

- \* Predecir la probabilidad de que un evento ocurra
- \* Determinar qué variables pesan más para aumentar o disminuir la probabilidad de que el evento ocurra.

El modelo, estima los coeficientes de tales cambios. Cuanto más coincidan los estados pronosticados con los estados reales de los sujetos, mejor ajustará el modelo.

Una variable dicotómica es aquella que sólo admite dos categorías como respuesta, que definen opciones o características mutuamente excluyentes. Estas variables pueden ser incorporadas en una ecuación de regresión: sus valores sólo pueden variar entre cero y uno ( $Y=SI$  ,  $Y=NO$ ); ( $Y=0$  ,  $Y=1$ ), y sus coeficientes indicarán, en cada caso, cuanto aumentan o disminuyen los “odds” de probabilidad del evento que se procura predecir cuando una de estas variables pasa de cero a uno.

En nuestro caso de estudio, el modelo de regresión logística permite estimar o predecir la probabilidad de que un individuo posea una característica ( $Y=Sobre$  el Nivel Muy Superior,  $Y=Bajo$  el Nivel Muy Superior); en función de una determinada o unas determinadas características individuales ( $X_1, X_2, X_3, \dots, X_n$ ), donde  $X_n$  representa las valoraciones de los estudiantes en las diferentes materias cursadas desde grado sexto hasta décimo. En este caso, para el cálculo de la probabilidad de obtener un puntaje por encima del nivel muy superior en los estudiantes del

grado 11 de los estudiantes del colegio San Francisco Javier se debe tener en cuenta el modelo de regresión logística.

Definida la variable dependiente como la ocurrencia o no de un acontecimiento, el modelo de regresión logística la expresa en términos de probabilidad, utilizando la función logística para estimar la probabilidad de que ocurra el acontecimiento se expresa mediante la siguiente formulación:

$$\pi_i = \frac{e^{(x_i \beta)}}{1 + e^{(x_i \beta)}} = \frac{1}{1 + e^{-(x_i \beta)}}$$

Siendo:

$\pi_i$  la probabilidad de ocurrencia del evento.

$x_i \beta$  la combinación lineal de las variables  $\beta$  regresoras, siendo los coeficientes asociados a las variables independientes.

$i = 1, 2, 3, \dots, n$  Las observaciones realizadas se denotan por medio de una matriz como:

$X'_{in} = (\text{Nota Matemáticas } 6^\circ, \text{ Nota Matemáticas } 7^\circ, \dots, \text{ Nota Matemáticas } 10, \text{ Nota Sociales } 6^\circ, \text{ Nota Sociales } 7^\circ, \dots, \text{ Nota Sociales } 10, \text{ Nota Castellano } 6^\circ, \text{ Nota Castellano } 7^\circ, \dots, \text{ Nota Castellano } 10, \text{ Nota Ciencias } 6^\circ, \text{ Nota Ciencias } 7^\circ, \dots, \text{ Nota Ciencias } 10)_{n \times k}$

Donde, cada fila representará la combinación de la valoración en cada área en cada grado que explican el estado del estudiantes en esta área específica.

Los  $n$  vectores se agrupan en la matriz

$$\begin{bmatrix} y_1 & x_{11} & x_{21} & \dots & x_{1k} \\ y_2 & x_{12} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ y_n & x_{1n} & x_{2n} & \dots & x_{nk} \end{bmatrix}$$

donde los  $y_i$  sólo pueden tomar los valores 0 ó 1.

Los coeficientes asociados a las variables independientes se denotaran como un vector fila:  $\beta'_i = (\beta_1, \beta_2, \dots, \beta_n)_{k \times 1}$

y la variable dependiente  $Y_i$  será:

$$Y_i = \begin{cases} 1 & \text{con probabilidad } P(Y = 0) = \pi \\ 0 & \text{con probabilidad } P(Y = 1) = 1 - \pi \end{cases}$$

es decir,

$$Y_i = \begin{cases} 0, & \text{"Está sobre nivel muy superior"} \\ 1, & \text{"No está sobre nivel muy superior"} \end{cases}$$

Con una probabilidad de  $1 - \pi$  si  $Y_i$  cumple con estar en nivel superior y con una probabilidad de  $\pi$  si  $Y_i$  no cumple con estar en nivel. Adicionalmente se debe cumplir que  $0 \leq \pi_i \leq 1$ .

Si el modelo tiene la forma:

$$y_i = x'_i \beta + \varepsilon_i$$

Con  $x'$  como se la definió anteriormente y  $\varepsilon_i$  el error.

donde  $E(\xi_i)=0$  y el valor esperado de la variable respuesta es:

$$E(y_i) = 1(\pi_i) + 0(1 - \pi_i) = \pi_i$$

Como;

$$y_i = x_i'\beta + \varepsilon_i, \quad E(y_i) = \pi_i, \quad \text{de donde, } x_i'\beta = \pi_i$$

Por tanto, decir que, la respuesta esperada  $E(y_i) = x_i'\beta$  equivale a decir que la probabilidad de que la variable respuesta  $y_i = 1$  tiene un valor de 1.

Así mismo, se puede observar que la variable respuesta es binaria, por tanto, los términos de error  $\varepsilon_i$  sólo pueden tomar dos valores:

$$\begin{aligned} \varepsilon_i &= 1 - x_i'\beta \quad \text{cuando } y_i = 1 \\ \varepsilon_i &= -x_i'\beta \quad \text{cuando } y_i = 0 \end{aligned}$$

Es decir, los errores no son normales y su varianza no es constante:

$$\begin{aligned} \sigma_y^2 &= E(y_i - E(y_i))^2 \\ &= (1 - \pi_i)^2 \pi_i + (0 - \pi_i)^2 (1 - \pi_i) \\ &= \pi_i (1 - \pi_i) \end{aligned}$$

Como forma funcional para  $E(y_i)$ , se usa la función de respuesta logística que garantiza valores en el intervalo (0, 1) y tiene la forma:

$$E(y) = \frac{e^{(x'\beta)}}{1 + e^{(x'\beta)}}$$

Equivalente a :

$$\Pr\{Y_i = 1\} = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_1 - \dots - \beta_m x_m)}$$

Puesto que el modelo anterior no es lineal respecto a las variables independientes, se considera la inversa de la función logística, que es el logit o logaritmo de la odds o ventaja de que un suceso ocurra, definiéndose ésta como el cociente entre la probabilidad de que ocurra un acontecimiento y la probabilidad de que no ocurra, que es su complementaria, como puede observarse en la siguiente expresión:

$$\text{Logit}(\pi) = \text{odds} = \ln \frac{\pi}{1 - \pi}, \quad \text{odds} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi}, \quad \eta = \text{Odds}$$

$$\eta = x_i' \beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi}$$

Con  $i=1,2,3,\dots,n$

Los odds, son la razón entre la probabilidad de que el evento ocurra (este por debajo del nivel muy superior) y aquella que no ocurra (esta en el nivel muy superior); el cálculo de los mismos ayudarán a la interpretación Los coeficientes asociados a las variables independientes.

La formulación anterior facilita la interpretación del modelo y de sus coeficientes, que reflejan, de este modo, el cambio en el logit correspondiente a un cambio unitario en la variable independiente considerada.

La capacidad predictiva del modelo de regresión logística se valora mediante la comparación entre el grupo de pertenencia observado y el pronosticado por el modelo, que clasifica a los individuos en cada grupo definido por la variable dependiente en función de un punto de corte establecido para las probabilidades predichas a partir de los coeficientes estimados y del valor que toman las variables explicativas para cada individuo.

Para estimar los parámetros del predictor  $x_i'\beta$ , se usa el método de máxima verosimilitud. Como cada observación de la muestra sigue la distribución Bernoulli, la distribución de probabilidades de cada observación será:

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, \quad i = 1, 2, \dots, n$$

y cada observación  $y_i$  toma el valor 0 o 1.

Como las observaciones son independientes, la función de verosimilitud será:

$$\begin{aligned} L(y_1, y_2, \dots, y_n, \beta) &= \prod_{i=1}^n f(y_i) \\ &= \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \end{aligned}$$

o por facilidad, el logaritmo de la función de verosimilitud:

$$\begin{aligned} \ln L(y_1, y_2, \dots, y_n, \beta) &= \ln \prod_{i=1}^n f(y_i) \\ &= \sum_{i=1}^n \left[ y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) \right] + \sum_{i=1}^n \ln(1 - \pi_i) \end{aligned}$$

Ahora bien, sea:

$$1 - \pi_i = \left[ 1 + e^{(x_i'\beta)} \right]^{-1} \quad y \quad \eta = \ln \frac{\pi}{1 - \pi} \quad \text{Entonces, el logaritmo de la función de}$$

verosimilitud se puede expresar como:

$$L(\beta) = \ln L(y, \beta) = \sum_{i=1}^n y_i x_i' \beta - \sum_{i=1}^n \ln(1 + e^{(x_i'\beta)})$$

Para estimar el  $\beta_{k \times 1}$  vector que maximice:  $L(\beta)$

$$\frac{\partial L(\beta)}{\partial \beta_i} = 0$$

donde  $i = 1, \dots, n$

Esto es equivalente a:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0$$
$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0 \quad \text{para } i = 1, \dots, p$$

De manera general, se pueden utilizar métodos numéricos para calcular los Estimaciones  $\hat{\beta}$  por máxima verisimilitud, pero resulta más fácil utilizar el método de mínimos cuadrados ponderados generalizados iterados.

De donde se obtendrá el valor estimado del predictor lineal  $\hat{\eta} = x_i' \hat{\beta}$  y el valor esperado del modelo de regresión logística

$$\hat{y} = \hat{\pi} = \frac{e^{\hat{\eta}}}{1 + e^{\hat{\eta}}}$$
$$= \frac{e^{(x' \hat{\beta})}}{1 + e^{(x' \hat{\beta})}}$$
$$= \frac{1}{1 + e^{(-x' \hat{\beta})}}$$

## Pruebas de bondad y ajuste

Las pruebas de hipótesis para el modelo de regresión logística se basan en pruebas de cociente de máxima verosimilitud (procedimiento para muestras grandes). Este método conduce a un estadístico conocido como desviación, que nos ayuda a determinar si el modelo propuesto describe de manera acertada o no la relación entre los datos.

## Pruebas de ajuste del modelo

La desviación del modelo compara el logaritmo de verosimilitud del modelo ajustado con la verosimilitud del modelo saturado, que en este caso es un modelo donde las probabilidades  $\pi(x_i)$  son totalmente irrestrictas por lo que al igualar  $\pi(x_i) = y_i$  se maximizara la verosimilitud.

Función de verosimilitud del modelo ajustado:

$$\ln L(\hat{\beta}) = \sum_{i=1}^n y_i x_i' \hat{\beta}_i - \sum_{i=1}^n \ln[1 + e^{x_i' \hat{\beta}}]$$

La desviación compara el logaritmo de verosimilitud del modelo saturado con el logaritmo de verosimilitud del modelo ajustado. De manera específica, la desviación del modelo se define como:

$$\begin{aligned} \lambda(\hat{\beta}) &= 2 \ln L(\text{modelo saturado}) - 2 \ln L(\hat{\beta}) \\ &= 2[\ell(\text{modelo saturado}) - \ell(\hat{\beta})] \end{aligned}$$

donde  $\ell$  representa el logaritmo de la función de verosimilitud.

Si el modelo de regresión logística es la función correcta para describir el comportamiento de los datos, y  $n$  es un tamaño de muestra grande, entonces la desviación del modelo tendrá una chi-cuadrado con  $n-p$  grados de libertad. Si la desviación del modelo representa valores grandes el modelo no es correcto, mientras que si representa valores pequeños significa que el modelo ajustado que tiene menos parámetros se ajusta casi tan bien como el modelo saturado.

Si  $\lambda(\beta) \leq \chi_{\alpha, n-p}^2$  el modelo ajustado es adecuado

Si  $\lambda(\beta) > \chi_{\alpha, n-p}^2$  el modelo ajustado no es adecuado

Nota: La desviación se considera como en el modelo de regresión lineal en el error de la suma de cuadrados residuales dividido entre la  $\sigma^2$  varianza del error .

### **Pruebas sobre subconjuntos de parámetros**

La desviación también se utiliza para hacer pruebas de hipótesis sobre subconjuntos de parámetros del modelo.

Supóngase que se desea probar:  $H_0 : \beta_2 = 0$

Por consiguiente el modelo completo será:  $\eta = X\beta$  y el modelo reducido será  $\eta = X_1\beta_1$ , y su desviación será  $\lambda(\beta_1)$  y la diferencia en la desviación se definirá como:

$$\lambda(\beta_2 / \beta_1) = \lambda(\beta_1) - \lambda(\beta)$$

donde  $\lambda(\beta_2 / \beta_1)$  tiene una distribución chi-cuadrado con  $r$  ( $r=n-(n-r)(n-p)$ ) grados de libertad.

Si  $\lambda(\beta_2 / \beta_1) \leq \chi_{\alpha,r}^2$  rechaza la hipótesis nula

Si  $\lambda(\beta_2 / \beta_1) > \chi_{\alpha,r}^2$  no rechaza la hipótesis nula

La diferencia de la desviación a veces se conoce como desviación parcial, que es una prueba de cociente de verosimilitud:

$$\frac{L(\hat{\beta}_1)}{L(\hat{\beta})}$$

El estadístico para la prueba de cociente de verosimilitud es igual a

$$\chi^2 = -2 \ln \frac{L(\hat{\beta}_1)}{L(\hat{\beta})}$$

### **Coefficiente de determinación**

Así como el cociente de verosimilitud es un estadístico que permite identificar el grado de ajuste del modelo, Cox y Snell en 1989 introdujeron el concepto de coeficiente de determinación para modelos lineales generalizados. Pretende de igual forma cuantificar mediante un valor comprendido entre 0 y 1 la bondad del ajuste; siendo mayor la variación explicada por el modelo mientras más alto es el valor de esta medida:

$$R^2 = 1 - \left( \frac{L(\hat{\beta}_1)}{L(\hat{\beta})} \right)^{2/n}$$

En 1991 Nagelkerke propone:

$$\bar{R}^2 = \frac{R^2}{\max(R^2)} \quad \text{donde } \max(R^2) = 1 - L(\hat{\beta})^{2/n}$$

### Pruebas individuales de parámetros

Pruebas para coeficientes individuales del modelo.

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Se puede utilizar el método de diferencia de la desviación, pero adicionalmente se puede utilizar el estadístico de Wald, que se basa en la teoría de las Estimaciones de máxima verosimilitud como sigue:

Sea G la matriz de  $p \times p$  de las segundas derivadas parciales de la función logaritmo de verosimilitud, esto es,

$$G_{ij} = \frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_j}, \quad i, j = 0, 1, \dots, k$$

G se llama matriz hessiana. Si los elementos de la hessiana se evalúan en las estimaciones de máxima verosimilitud  $\beta = \hat{\beta}$ , la matriz de covarianza para muestra grande, de los coeficientes de regresión es,  $Var(\hat{\beta}) = \sum = -G(\hat{\beta})^{-1}$ , las raíces cuadradas de los elementos diagonales de esta matriz son los errores

estándar de muestras grandes de los coeficientes de regresión, por lo que el estadístico de prueba de la hipótesis nula es:

$$Z_0 = \frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}}$$

La distribución de referencia para este estadístico es la distribución normal estándar.

Es posible que algunas de las supuestas variables explicativas no sean tales y no tengan ningún efecto sobre la variable respuesta; para poder identificarlas y eliminarlas del modelo, se recurre a la prueba de Wald, la cual se limita a contrastar la hipótesis de nulidad del coeficiente  $\beta_j$  asociado a la variable  $X_j$ :

$H_{0j}$ : "X<sub>j</sub> no influye sobre Y:  $\beta_j = 0$ "

frente a la alternativa:

$H_{1j}$ : "X<sub>j</sub> influye sobre Y:  $\beta_j \neq 0$ ".

El estadístico de prueba utilizado es

$$W_j = \frac{\hat{\beta}_j^2}{s_j^2}$$

que se distribuye como una  $\chi^2$  con 1 grado de libertad cuando la muestra es grande, siendo  $s_j^2$  la varianza de la estimación de  $\beta_j$ . Si  $p\_valor < \alpha$  se rechaza  $H_0$  con un nivel de significancia de  $\alpha$ . Entonces podemos decir que no existe evidencia estadística para rechazar la hipótesis nula, es decir, que la información que se perderá al eliminar la variable  $X_j$  en el siguiente paso no es significativa.

### 3. METODOLOGÍA

El presente trabajo se realizará con base en los resultados académicos de estudiantes egresados y estudiantes de grado 10° del colegio San Francisco Javier de Pasto

La base de datos, la cual está en construcción estará constituida por las valoraciones cualitativas que demuestran el desempeño académico en cada grado y por cada asignatura que mide el ICFES desde el grado sexto hasta el grado once de los estudiantes egresados que conforman las generaciones 2003-2004 hasta la generación 2006-2007 con sus respectivos resultados ICFES.

Existen diferentes criterios de evaluación implementados en el colegio San Francisco Javier de Pasto a lo largo de historia y en especial en los últimos 15 años.

En algunas generaciones dicho proceso de valoración se desarrollo bajo el criterio de valoración cualitativa, utilizando las los conceptos Excelente, Bueno e Insuficiente. Posteriormente se implementa una valoración con Excelente, Sobresaliente, Aceptable, Insuficiente y Deficiente y actualmente se ha adicionando a cada valoración un intervalo de notas numéricas, con el objetivo de volver esa evaluación más objetiva. El presentarse estos cambios implica que la base de datos obtenida tendrá variedad en la información, por lo cual se deben establecer algunos parámetros que permitan unificar estas valoraciones y así obtener datos con notaciones iguales para el estudio.

Los resultados ICFES conformarán una variable dicotómica 1-0, 0 para aquellos resultados que permiten categorizar a la institución en un nivel Muy Superior y 1 para aquellas que hagan que el colegio baje de este nivel.)

Con los datos de la base se pretende:

1. Plantear un modelo de regresión logística permite estimar o predecir la probabilidad de que un estudiante de grado décimo posea una característica (Y= Nivel muy superior, Y=Debajo del nivel muy superior) en función del desempeño académico en su vida escolar.
2. Analizar si existen diferencias significativas entre los estudiantes que se encuentran en el nivel muy superior y aquellos que están por debajo de él, teniendo en cuenta el desempeño académico de los estudiantes, en el caso de que existan, explicar en qué sentido se dan y proporcionar procedimientos de clasificación sistemática en los estudiantes de grado décimo que presentarán el ICFES y determinar la probabilidad de ser un posible caso de obtener un puntaje por debajo del nivel muy superior.
3. Se realizarán estos estudios en cada área de conocimiento que mide el ICFES, se comparará cada modelo y se determinará cual es más adecuado para el pronóstico en cada área. Con los resultados obtenidos se pretende identificar las falencias académicas presentadas en su formación y por tanto, la institución deberá implementar actividades de refuerzo académico que permitan mejorar el resultado ICFES del estudiante de grado décimo que al ser incluido en el estudio, muestra una probabilidad alta de ser un posible caso de obtener un puntaje por debajo del nivel muy superior.

### **3.1. RECOLECCIÓN DE LA INFORMACIÓN**

Para recolectar la información necesaria para el estudio se consultó el historial académico de cada estudiante Javeriano a lo largo de su vida escolar en el Colegio desde grado sexto a décimo en las asignaturas de castellano, matemáticas, sociales y biología, igualmente se consultó en la base de datos que maneja el ICFES referente a los resultados obtenidos por los estudiantes en las áreas evaluadas.

### **3.2. CONSTRUCCIÓN DEL MODELOS DE REGRESIÓN LOGÍSTICA PARA CADA ASIGNATURA**

Se tomó como variables independientes el resultado icfes en cada asignatura denotando 0 cuando el estudiante está sobre el nivel muy superior y 1 por debajo del mismo, y las covariables, las valoraciones académicas. En total hay 19 covariables las cuales pueden estar presentes en el modelo. Posteriormente se buscó un modelo de regresión logística en cada asignatura que ajuste la variable dependiente con las variables independientes. Para realizar esta tarea se utilizaron las versiones académicas de los programas StatGraphics 5.1 y SPSS 15.01.

### **3.3. NOMENCLATURA DE VARIABLES**

Las 19 variables independientes y las 6 asignaturas que evalúa el ICFES, como variables dependientes, se denotan como se indica a continuación. (Tabla 1 y Tabla 2).

**Tabla 1. Descripción de variables independientes**

<b>COVARIABLES</b>	<b>NOMENCLATURA</b>
Matemáticas sexto	M6
Matemáticas séptimo	M7
Matemáticas octavo	M8
Matemáticas noveno	M9
Matemáticas décimo	M10
Biología sexto	B6
Biología séptimo	B7
Biología octavo	B8
Biología noveno	B9
Química – Física	QF
Lenguaje sexto	L6
Lenguaje séptimo	L7
Lenguaje octavo	L8
Lenguaje noveno	L9
Lenguaje décimo	L10
Sociales sexto	S6
Sociales séptimo	S7
Sociales octavo	S8
Sociales noveno	S9

**Tabla 2. Descripción de variables dependientes.**

<b>VARIABLES DEPENDIENTES</b>	<b>NOMENCLATURA</b>
Matemáticas	IM
Biología	IB
Química	IQ
Física	IF
Lenguaje	IL
Sociales	IS

## 4. MODELOS OBTENIDOS

### 4.1. Modelo para el área de Matemáticas

Modelo 1

Se tomó como Variable dependiente IM y como variables independientes: M6, M7, M8, M9, M10, C6, C7, C8, C9, C10, S6, S7, S8, S9, B6, B7, B8, B9, QF.

Para obtener el valor estimado del predictor lineal  $\eta = x\beta$  y el valor esperado del modelo de regresión logística se utilizó el método de mínimos cuadrados ponderados generalizados iterados, obteniendo los resultados siguientes:

**Tabla 3. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	0,477	0,365	0,191
B7	-0,443	0,303	0,144
B8	-0,446	0,376	0,235
B9	1,342	0,299	0,000
QF	-0,333	0,278	0,231
M6	-0,754	0,477	0,114
M7	-0,930	0,522	0,075
M8	-0,086	0,380	0,821
M9	-1,129	0,461	0,014
M10	-0,940	0,353	0,008
S6	0,411	0,504	0,415
S7	-0,575	0,448	0,200
S8	1,165	0,532	0,029
S9	0,687	0,484	0,156
L6	-0,786	0,436	0,072
L7	0,181	0,400	0,652
L8	-0,284	0,422	0,500
L9	-0,015	0,428	0,972

L10	0,233	0,394	0,554
Constante	8,318	2,183	0,000

### Modelo de regresión

$$IM = \frac{e^v}{1 + e^v}$$

Donde  $v$  es  $x'\beta$

$$v = 8,31839 + 0,47692*B6 - 0,442773*B7 - 0,446138*B8 + 1,34159*B9 + 0,232994*L10 - 0,786318*L6 + 0,180645*L7 - 0,284278*L8 - 0,0148987*L9 - 0,940274*M10 - 0,753675*M6 - 0,930227*M7 - 0,0861761*M8 - 1,12861*Mt - 0,332977*QF + 0,411095*S6 - 0,574899*S7 + 1,16528*S8 + 0,686654*S9$$

Se realizó un análisis de varianza para estudiar el ajuste del modelo, del cual se obtuvieron los resultados que se presentan en la tabla 4.

**Tabla 4. Análisis de varianza para la regresión**

Fuente	Desviación	G.I.	Valor P
Modelo	96,603	19	0,0000
Residuos	375,142	324	0,0263
Total	471,745	343	

Como el valor p para el modelo en la tabla anterior es inferior a 0.01, se puede concluir que hay una relación estadísticamente significativa entre las variables de estudio.

Se observó el estadístico de Cox y Snell para cuantificar la bondad del ajuste, en este caso es igual a 24.4%, este valor puede suponer que el modelo explica de manera no tan adecuada la variabilidad del mismo. Sin embargo el porcentaje

ajustado de Nagelkerke, el cual, es más adecuado para comparar modelos con diferentes números de variables independientes es 32.8%, al parecer mejora el ajuste, no obstante, los resultados no son satisfactorios, por lo cual se trata de mejorar el modelo identificando aquellas variables que no aportan significativamente al modelo. Aquí se aplicó el sistema paso a paso hacia atrás obteniendo los siguientes resultados:

## Modelo 2

**Tabla 5. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B7	-0,469	0,286	0,100
B9	1,271	0,275	0,000
L6	-0,633	0,374	0,091
M6	-0,661	0,443	0,136
M7	-1,055	0,465	0,023
M9	-1,158	0,423	0,006
M10	-1,038	0,338	0,002
S8	0,856	0,488	0,080
S9	0,658	0,444	0,138
Constante	8,393	2,058	0,000

El modelo resultante es:

$$v = 8,39287 - 0,469473*B7 + 1,27136*B9 - 0,63318*L6 - 0,661M6 - 1,055M7 - 1,158M9 - 1,038M10 - 0,856S8 + 0,658 S9$$

El modelo es significativo según el análisis de varianza, sin embargo el ajuste baja con respecto al anterior modelo, puesto que los estadísticos de Cox y Snell, y Nagelkerke, son respectivamente del 22,9% y 30,64%

Posteriormente se buscaron otros modelos de regresión logística construyendo una base donde las variables independientes tomaron los valores 1 y 0 (modelo 3), 0 si el estudiante tiene sobresaliente o excelente y 1 si tiene insuficiente o aprobado.

### Modelo 3

**Tabla 6. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	-0,288	0,391	0,462
B7	0,629	0,387	0,104
B8	0,547	0,397	0,168
B9	-1,655	0,416	0,000
QF	-0,067	0,509	0,896
M6	0,349	0,303	0,248
M7	0,465	0,395	0,239
M8	0,002	0,380	0,996
M9	0,564	0,432	0,192
M10	1,926	0,870	0,027
S6	-0,144	0,416	0,729
S7	0,768	0,407	0,059
S8	-0,361	0,452	0,425
S9	-1,450	0,563	0,010
L6	0,322	0,317	0,311
L7	0,166	0,318	0,602
L8	-0,046	0,368	0,901
L9	0,140	0,471	0,766
L10	0,719	0,612	0,240
Constante	-1,896	0,860	0,027

Modelo

$$v = -1,89614 - 0,288099*B6 + 0,628588*B7 + 0,546659*B8 - 1,65469*B9 + 0,718838*L10 + 0,321584*L6 + 0,165729*L7 - 0,0455832*L8 +$$

$$0,140269*L9 + 1,92559*Mat10 + 0,349431*Mat6 + 0,46523*Mat7 + \\ 0,00207158*Mat8 + 0,563671*Mat9 - 0,0668324*QimFis - \\ 0,143998*Sos6 + 0,768465*Sos7 - 0,360656*Sos8 - 1,45017*Sos9$$

Sin embargo, el análisis no mostró mejores resultados, los estadísticos de Cox y Snell, y Nagelkerke, son respectivamente del 21% y 28%.

Al igual que en el modelo inicial, se reducen las variables utilizando el método paso a paso hacia atrás, obteniendo el modelo:

#### **Modelo 4**

$$v = -1,92952 + 0,756407*B7 + 0,482914*B8 - 1,65994*B9 + 0,819635*L10$$

Sin embargo, según el análisis de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente los del 13% y 18%. Lo cual indica poco ajuste e inferior al del modelo inicial.

Finalmente se realizó el estudio con una base donde las variables independientes son los promedios en cada asignatura, el estudio se hizo, con un modelo completo (modelo 5) y luego se eliminaron las variables menos significativas (modelo 6); los resultados arrojados por el software no indican un buen ajuste en ninguno de los dos casos, puesto que los de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente de 13% y 17% en el primer caso (modelo 5) y del 17% y 19% en el modelo reducido (modelo 6).

Del estudio anterior se puede deducir que el ajuste del modelo que describe la relación entre IM y el rendimiento académico es bajo y el modelo que mejor explica esta relación es el modelo 1, modelo al que se le realizan diferentes pruebas para analizar la habilidad del modelo para efectos de clasificación.

### Test de Bondad de ajuste Chi-cuadrado

Se aplica la Prueba de Bondad de Ajuste Chi Cuadrado para determinar la habilidad de clasificación del modelo.

**Tabla 7. Bondad de ajuste**

Clase	Intervalo logia	n	VERDADERO		FALSO	
			Observado	Esperado	Observado	Esperado
1	menor que -0,721246	69	13,0	12,1422	56,0	56,8578
2	-0,721246 a 0,105599	69	33,0	29,9268	36,0	39,0732
3	0,105599 a 0,598834	69	36,0	40,9336	33,0	28,0664
4	0,598834 a 1,4578	69	49,0	49,9909	20,0	19,0091
5	1,4578 o mayor	68	62,0	60,0065	6,0	7,99351
Total		344	193,0		151,0	

Chi-cuadrado = 2,72739 con 3 g.l. Valor P = 0,43559

Dado que el Valor P es superior a 0.10, no hay razón para descartar al modelo como medio de clasificación.

Al realizar el anterior estudio con los otros modelos, se obtuvieron los resultados que se indican en la tabla 8.

**Tabla 8. Valor del Chi-cuadrado en cada modelo presente en el estudio**

MODELO	Chi-cuadrado	Valor P
1	2,72739	0,43559
2	1,48213	0,686398
3	2,83283	0,41812
4	5,09802	0,164756
5	1,37216	0,71207
6	0,377263	0,944898

Como se puede observar todos los modelos presentan una buena habilidad clasificatoria.

### Prueba de Hosmer-Lemeshow

**Tabla 9. Tabla de contingencias para la prueba de Hosmer y Lemeshow**

	IM = 0,00		IM = 1,00		Total
	Observado	Esperado	Observado	Esperado	Observado
1	31	30,490	3	3,510	34
2	24	25,695	10	8,305	34
3	17	21,107	17	12,893	34
4	20	18,166	15	16,834	35
5	19	15,132	15	18,868	34
6	14	12,694	20	21,306	34
7	14	11,031	20	22,969	34
8	5	8,117	29	25,883	34
9	5	5,061	27	26,939	32
10	2	3,508	37	35,492	39

Chi-cuadrado = 8,497 con 8 G.L. Valor P = 0,386

Dado que el Valor P de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo, decir, el modelo se ajusta adecuadamente a los datos.

### Doble logaritmo del estadístico de Verosimilitud

Al comparar el valor del  $-2\log$  verosimilitud cuando el modelo no tiene variables independientes con el valor del  $-2\log$  verosimilitud cuando se introducen estas variables al modelo se obtuvo lo siguiente.

-2log verosimilitud sin variables independientes: 471,745

-2log verosimilitud con variables independientes: 375,142

Como el valor del -2log verosimilitud se redujo, el ajuste es bueno, más no ideal porque esta reducción no tiende a cero.

Al aplicar este estadístico en los otros modelos observamos los resultados que se indican en la tabla 10

**Tabla 10. Valor del -2log verosimilitud en cada modelo presente en el estudio**

MODELO	-2log verosimilitud. (A)	-2log verosimilitud. (D)
1	471,745	375,142
2		382,4116451
3		389,6721083
4		420,2451178
5		423,3810966
6		425,0060923

Los resultados arrojados por el software en cada modelo indican un buen ajuste al modelo logístico, sin embargo las reducciones son menores en los modelos del 2 al 6.

**Tabla 11. Tabla de clasificación para IM**

Observado		Pronosticado		
		IM		Porcentaje correcto
		0	1	
IM	0	87	64	57,6
	1	40	153	79,3
Porcentaje global				69,8

La ecuación del modelo ya diseñada nos proporciona una probabilidad que nos permite predecir a partir de ella para cada sujeto un valor de Y (Y predicho), tal que si  $P(Y=1|X) \leq 0.5$  entonces Y pred= 1, y si  $P(Y=1|X) > 0.5$  entonces Y pred=0

Al observar la lista de clasificación del modelo se contabilizan 104 casos mal clasificados de 344 casos, correspondientes al 30,2%, los cuales se indican a continuación:

**Tabla 12. Listado de casos Mal Clasificados**

Caso	Observado IM	Pronosticado	Grupo Pronosticado
19	1	0,383916738	0
32	1	0,477901882	0
35	1	0,323896413	0
41	0	0,857045853	1
49	1	0,409210314	0
52	0	0,809335466	1
69	1	0,35083023	0
82	1	0,307676916	0
83	0	0,699545176	1
88	0	0,54105712	1
94	0	0,820987529	1
95	1	0,453134583	0
97	0	0,628821094	1
98	1	0,335142593	0
99	1	0,158282993	0
101	0	0,818813837	1
104	1	0,246602115	0
105	1	0,33181445	0
111	1	0,277616634	0
116	0	0,592122685	1
119	1	0,403148727	0
121	0	0,779882823	1

122	0	0,64667227	1
126	0	0,891861386	1
130	0	0,617692564	1
132	0	0,528691618	1
133	1	0,335421683	0
137	1	0,367692781	0
139	1	0,414634195	0
141	1	0,162848008	0
150	1	0,479805802	0
153	1	0,452367146	0
156	0	0,67022881	1
158	1	0,492035941	0
160	0	0,6611719	1
163	1	0,186328064	0
164	1	0,071759075	0
165	1	0,443023345	0
169	1	0,405627187	0
170	0	0,52365419	1
171	1	0,224236147	0
173	0	0,510580745	1
174	1	0,36799163	0
175	0	0,568817022	1
176	0	0,533889255	1
182	1	0,344742515	0
183	1	0,310556362	0
184	0	0,625978774	1
187	0	0,749759596	1
188	0	0,697397482	1
194	0	0,555513449	1
195	0	0,628821094	1
198	0	0,549838841	1
200	1	0,379821435	0
207	0	0,567484685	1
211	0	0,789877545	1
212	1	0,413070923	0

213	0	0,628821094	1
214	0	0,591654598	1
215	1	0,499161822	0
222	0	0,513907149	1
223	0	0,688445743	1
227	0	0,604145315	1
230	0	0,698407508	1
233	0	0,588218726	1
234	0	0,521562677	1
235	1	0,365777899	0
236	1	0,338976693	0
237	0	0,537001369	1
242	0	0,628821094	1
249	0	0,628821094	1
255	0	0,646773678	1
256	0	0,651699373	1
261	0	0,680437152	1
262	1	0,444848575	0
263	0	0,694841622	1
264	0	0,663251868	1
265	0	0,628821094	1
274	1	0,208292688	0
275	1	0,200781241	0
278	0	0,528130552	1
281	0	0,628821094	1
284	0	0,919608524	1
285	0	0,528130552	1
287	1	0,432168528	0
288	0	0,628821094	1
297	0	0,528130552	1
298	1	0,379447596	0
299	0	0,528130552	1
300	0	0,628821094	1
301	1	0,467958474	0
308	0	0,528130552	1

312	0	0,614240292	1
315	0	0,766977332	1
320	1	0,200068288	0
322	0	0,819928295	1
323	0	0,628821094	1
324	0	0,694252416	1
325	0	0,534736797	1
327	0	0,672310692	1
328	0	0,628821094	1
329	0	0,52365419	1
334	0	0,544301392	1
342	0	0,7984273	1

Al realizar este mismo estudio para cada modelo encontramos los resultados de la tabla 13.

**Tabla 13. Total de casos mal clasificados en cada modelo presente en el estudio**

MODELO	CASOS MAL CLASIFICADOS
1	104
2	120
3	109
4	124
5	112
6	119

## Residuos atípicos

**Tabla 14. Residuos atípicos para IM**

Fila	Y	Predicho Y	Residuo	Residuo Pearson	Desviación Residuo
41	0	0,857046	-0,857046	-2,45	-1,97
52	0	0,809335	-0,809335	-2,06	-1,82
94	0	0,820988	-0,820988	-2,14	-1,85
99	1	0,158283	0,841717	2,31	1,92
101	0	0,818814	-0,818814	-2,13	-1,85
126	0	0,891861	-0,891861	-2,87	-2,11
141	1	0,162848	0,837152	2,27	1,91
163	1	0,186328	0,813672	2,09	1,83
164	1	0,0717591	0,928241	3,60	2,30
284	0	0,919609	-0,919609	-3,38	-2,25
322	0	0,819928	-0,819928	-2,13	-1,85

La tabla de residuos atípicos lista todas las observaciones que tienen residuo Pearson o desviación de residuo superiores a 2.0 en valor absoluto. Estos residuos estandarizados miden cuántas desviaciones estándar de cada valor observado de IM proceden del modelo ajustado. En este caso, hay 11 residuos estandarizados superiores a 2.0, y 2 superiores a 3.0.

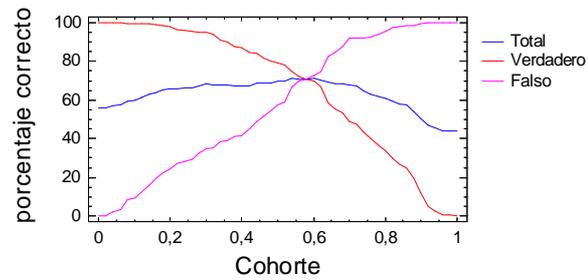
Al estudiar con detalle estos individuos se observa que aquellos que están por ENCIMA del nivel muy superior (0), su nivel académico es BAJO; y aquellos que están por DEBAJO del nivel muy superior (1), presentan un rendimiento académico BUENO en su vida escolar. Por lo tanto, es conveniente mantenerlos en el estudio puesto que son casos reales que no presentan errores al levantamiento ni digitación de la información ya que se confrontó la base de datos con la información de los libros de registro.

## PREDICCIÓN DEL MODELO

Tabla 15. Capacidad de Predicción del Modelo

Cohorte	VERDADERO	FALSO	Total
0,0	100,00	0,00	56,10
0,05	100,00	3,31	57,56
0,1	99,48	9,27	59,88
0,15	99,48	17,22	63,37
0,2	97,93	24,50	65,70
0,25	95,34	29,14	66,28
0,3	94,82	34,44	68,31
0,35	90,67	39,07	68,02
0,4	87,05	41,72	67,15
0,45	82,90	50,99	68,90
0,5	79,27	57,62	69,77
0,55	73,06	68,87	71,22
0,6	69,95	72,85	71,22
0,65	57,51	84,11	69,19
0,7	49,22	92,05	68,02
0,75	40,93	92,72	63,66
0,8	33,68	95,36	60,76
0,85	25,39	98,01	57,27
0,9	11,92	99,34	50,29
0,95	1,04	100,00	44,48
1,0	0,00	100,00	43,90

Figura 1. Capacidad de Predicción del Modelo para IM



En la tabla y el gráfico se observa la capacidad de predicción del modelo 1. Si el valor predicho es más grande que el punto de cohorte, la respuesta se predice como VERDADERA. Si el valor predicho es inferior a o igual al punto de cohorte, la respuesta se predice para ser FALSO. La tabla muestra el porcentaje de datos observados predichos correctamente a diferentes puntos de cohorte. En nuestro caso con el punto de cohorte utilizado 0,5, el 79,27% de todas las respuestas VERDADERAS se predijeron correctamente, mientras que el 57,62% de todas las respuestas FALSAS se predijeron correctamente, para un total de 69,77% casos clasificados correctamente.

**Tabla 16. Matriz de correlación para los coeficientes estimados**

	Cons	B6	B7	B8	B9	QF	M6	M7	M8	M9	M10	S6	S7	S8	S9	L6	L7	L8	L9	L10
Cons	1																			
B6	0,032	1																		
B7	0,13	0,009	1																	
B8	-0,24	-0,12	-0,03	1																
B9	-0,07	0,151	-0,03	0,084	1															
QF	0,051	-0,04	-0,01	0,026	-0,2	1														
M6	-0,23	-0,05	-0,07	0,062	0,1	0,009	1													
M7	-0,25	-0,19	-0,08	0,01	-0,09	-0,02	-0,08	1												
M8	-0,09	-0,03	-0,14	-0,03	-0,08	-0,17	-0,09	-0,14	1											
M9	-0,18	-0,03	-0,03	-0,09	0,036	0,005	0,125	-0,11	-0,18	1										
M10	-0,28	-0,01	-0,06	0,013	-0,06	-0,15	-0,13	0,054	-0,02	0,034	1									
S6	-0,01	-0,17	0,01	-0,07	-0,07	-0,05	-0,16	0,086	-0	-0,13	-0,04	1								
S7	0,035	-0,21	-0,1	-0,02	-0,03	0,032	-0,14	-0,11	0,014	-0	0,06	-0,17	1							
S8	-0,05	0,139	0,082	-0,2	-0,02	-0,03	-0,1	-0,17	-0,01	-0,17	-0,02	-0,01	-0,15	1						
S9	-0,14	0,039	-0,19	-0,08	-0,09	-0,12	-0,07	-0,05	0,187	-0,04	-0,01	-0,04	-0,01	-0,15	1					
L6	-0,07	-0,23	-0,07	0,007	-0,08	-0,08	-0,18	0,112	0,1	0,007	0,111	-0,21	0,009	-0,15	0,02	1				
L7	-0,06	-0,11	-0,13	-0,04	-0	-0,01	-0,09	-0,05	0,027	-0,01	0,04	-0,12	-0,13	0,122	-0,04	-0,14	1			
L8	0,157	0,097	0,071	-0,03	-0,06	0,026	0,117	-0,22	-0,21	-0	-0,18	-0,1	-0,03	-0,11	-0,09	-0,15	-0,2	1		
L9	0,045	-0,04	-0,07	0,021	-0,15	0,047	0,044	0,077	-0,15	-0,12	-0,09	-0,04	-0,06	-0,12	-0,12	-0,03	-0,01	0,008	1	
L10	-0,12	-0,05	-0,1	0,016	-0,07	-0,04	-0,05	0,065	0,073	-0,08	-0,03	0,089	0,006	-0	-0,13	-0,03	-0,04	-0,12	-0,32	1

Esta tabla muestra las correlaciones estimadas entre los coeficientes en el modelo ajustado. Estas correlaciones pueden utilizarse para detectar la presencia de serias multicorrelaciones, es decir, correlación entre las variables pronosticadas. En este caso, no hay ninguna correlación con valores absolutos superiores a 0.5.

## INTERPRETACION DEL MODELO

Como se puede observar en el estudio anterior, el modelo 1 presenta un adecuado ajuste, las variables que ayudan a predecir si un estudiante de grado décimo del Colegio Javeriano estará por encima o debajo del nivel muy superior en las pruebas de estado ICFES en la asignatura de matemáticas son todas las materias de matemáticas, castellano, sociales y naturales, que cursó desde grado sexto a grado décimo en la institución.

A continuación se indica el valor de las probabilidades que tiene cada variable de estudio en el modelo.

**Tabla 17. Variables en el modelo**

Asignatura	B	Exp(B)
B6	0,477	1,611
B7	-0,443	0,642
B8	-0,446	0,640
B9	1,342	3,825
QF	-0,333	0,717
M6	-0,754	0,471
M7	-0,930	0,394
M8	-0,086	0,917
M9	-1,129	0,323
M10	-0,940	0,391

Asignatura	B	Exp(B)
S6	0,411	1,508
S7	-0,575	0,563
S8	1,165	3,207
S9	0,687	1,987
L6	-0,786	0,456
L7	0,181	1,198
L8	-0,284	0,753
L9	-0,015	0,985
L10	0,233	1,262
Constante	8,318	4098,568

De lo anterior se puede decir que el valor de 0,477 que se asocia a la variable B6, incrementa el 161% la posibilidad de estar por debajo del nivel muy superior en el área de matemáticas, puesto que  $e^{0,477} = 1,61$

De igual manera se pueden leer todos los coeficientes que acompañan a cada variable de estudio.

Como se puede observar, se evidencia la importancia de las asignaturas de sociales en los niveles de sexto, octavo y noveno, que realizan un aporte alto para el cálculo de esta probabilidad en el área de matemáticas, la cual debe tenerse en cuenta en el desarrollo de los conceptos académicos en adelante; de igual manera se presentan las asignaturas de biología en grados sexto y noveno y lenguaje en los niveles de sexto y décimo. Estas dos áreas son razonables puesto que en ellas se desarrollan además de la competencia lectora, que es propio del lenguaje, la habilidad de desarrollar conceptos matemáticos en biología, el saber hacer en contexto.

De igual manera, se puede observar que asignaturas propias del área como matemáticas en los diferentes niveles, realizan un aporte bajo para el cálculo de esta probabilidad,

El modelo seleccionado (modelo 1) ayudará a detectar a aquellos estudiantes de grado décimo que requieren de un refuerzo académico antes de presentar la prueba ICFES. A continuación se indica la probabilidad de que cada estudiante de grado décimo pueda estar por debajo del nivel muy superior. (Tabla 18).

**Tabla 18. Casos posibles que necesitan refuerzo académico en matemáticas antes de las pruebas ICFES**

Caso	B6	B7	B8	B9	QF	L6	L7	L8	L9	L10	M6	M7	M8	M9	M10	S6	S7	S8	S9		P(y=1)
1	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	-0,154	0,4615
2	4,6	4,6	4,3	3,5	3,5	4,6	4,6	4,3	3,5	3,5	4,6	4,3	3,5	3,5	3,5	4,6	4,6	4,3	3,5	-1,507	0,1814
3	3,5	3,5	3,5	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	4,6	3,5	4,3	0,595	0,6445
4	4,6	4,3	4,3	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,3	3,5	3,5	3,5	3,5	4,6	4,6	4,3	3,5	-0,458	0,3874
5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	2	3,5	2	4,3	3,5	3,5	3,208	0,9611
6	4,6	4,3	4,3	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,6	3,5	3,5	4,3	4,3	4,3	4,6	4,3	4,3	-0,469	0,3849
7	4,3	3,5	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	3,5	4,3	4,3	4,3	0,000	0,5001
8	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,067	0,5168
9	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	-1,922	0,1277
10	3,5	2	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	4,3	3,5	3,5	2,153	0,896
11	4,6	4,6	4,3	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	-1,788	0,1434
12	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	4,3	4,3	3,5	0,551	0,6343
13	4,3	3,5	3,5	4,3	3,5	3,5	3,5	3,5	4,3	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	3,178	0,96
14	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	-0,434	0,3932
15	4,6	4,3	4,6	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,6	4,3	4,3	4,3	4,6	4,6	4,6	4,3	-1,770	0,1455
16	4,3	4,3	4,3	4,6	4,6	4,3	4,3	3,5	4,6	4,6	4,6	4,3	3,5	4,6	4,6	4,3	4,3	4,3	4,6	-1,230	0,2262
17	3,5	3,5	3,5	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	3,5	4,3	4,3	4,3	0,045	0,5112
18	3,5	4,3	3,5	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	4,3	4,3	3,5	4,3	1,612	0,8336
19	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,3	4,3	-1,901	0,13
20	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,6	4,3	4,3	-1,593	0,1689
21	4,6	4,3	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	-1,789	0,1432
22	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	-1,283	0,217
23	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,3	4,6	-2,271	0,0935
24	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,6	4,3	4,3	-1,622	0,165
25	3,5	3,5	2	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,6	3,5	3,5	2,886	0,9471
26	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	4,3	3,5	3,5	1,844	0,8634
27	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,3	4,6	4,6	4,6	-2,045	0,1146
28	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,6	4,6	4,3	4,3	0,556	0,6356
29	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,6	4,3	4,3	4,3	3,5	4,6	4,3	4,3	-2,119	0,1072
30	3,5	3,5	2	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,6	3,5	3,5	2,886	0,9471
31	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,731	0,6751
32	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	-0,154	0,4615
33	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	3,5	3,5	4,3	4,3	4,3	4,3	1,085	0,7474

34	4,6	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	3,5	4,3	4,3	4,3	0,003	0,5007
35	4,3	3,5	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	4,3	4,3	1,077	0,7459
36	3,5	2	2	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	2	3,5	2	2	4,3	4,3	3,5	2,985	0,9519
37	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	3,5	1,328	0,7906
38	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,067	0,5168
39	4,6	4,3	4,3	3,5	3,5	4,6	4,6	4,3	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,3	4,6	4,3	3,5	-1,566	0,1727
40	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,6	4,6	4,3	-1,102	0,2493
41	4,6	4,3	4,3	3,5	3,5	4,6	4,6	3,5	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,6	4,3	4,3	3,5	-1,043	0,2605
42	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	4,3	3,5	3,5	-0,002	0,4996
43	3,5	4,3	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	2	3,5	4,3	3,5	3,5	3,148	0,9588
44	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,067	0,5168
45	3,5	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,6	4,3	3,5	1,136	0,757
46	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	3,5	3,5	4,6	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	-1,286	0,2165
47	4,6	4,3	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,6	3,5	3,5	-0,564	0,3626
48	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,6	4,3	4,3	-1,648	0,1614
49	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,3	4,6	4,3	4,3	-0,783	0,3136
50	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	4,6	3,5	3,5	1,588	0,8303
51	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	3,5	3,5	4,3	4,6	4,3	3,5	-1,159	0,2389
52	4,6	4,6	4,3	4,3	4,3	4,6	4,6	4,6	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,3	4,3	4,3	-1,639	0,1626
53	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	2	3,5	3,5	4,6	3,5	3,5	-1,028	0,2635
54	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	2	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,860	0,7026
55	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	4,3	3,5	3,5	1,192	0,767
56	3,5	3,5	3,5	2	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,3	3,5	3,5	0,377	0,593
57	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	0,527	0,6288
58	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2,304	0,9092
59	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,3	4,6	4,3	4,6	-2,394	0,0836
60	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,6	4,6	4,3	4,3	0,556	0,6356
61	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,6	4,3	4,3	4,3	3,5	4,6	3,5	3,5	-3,601	0,0266
62	3,5	3,5	2	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,3	3,5	3,5	3,058	0,9551
63	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,6	4,6	4,6	2,596	0,9306
64	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	4,6	4,3	4,3	1,155	0,7604
65	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	3,5	3,5	4,3	4,6	4,3	4,3	0,912	0,7135
66	4,6	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	3,5	4,6	3,5	3,5	-1,651	0,1609
67	4,3	3,5	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	-0,405	0,4001
68	3,5	2	2	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	2	3,5	2	2	4,3	3,5	3,5	2,053	0,8862
69	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	4,3	1,878	0,8673
70	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	1,549	0,8247
71	4,6	4,3	4,3	3,5	3,5	4,6	4,6	4,3	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	-0,845	0,3005

72	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,3	3,5	-1,829	0,1384
73	4,6	4,3	4,3	3,5	3,5	4,6	4,6	3,5	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,6	4,3	4,3	3,5	-1,043	0,2605	
74	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	4,3	3,5	3,5	-0,002	0,4996	
75	3,5	4,3	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	2	3,5	4,6	4,3	3,5	3,907	0,9803	
76	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,6	4,6	4,3	1,726	0,8489	
77	3,5	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	3,5	1,309	0,7873	
78	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	3,5	3,5	4,6	4,3	3,5	3,5	3,5	4,3	4,3	3,5	3,5	-2,219	0,0981	
79	4,6	4,3	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	-0,391	0,4034	
80	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,3	3,5	3,5	-2,957	0,0494	

## CONCLUSIONES DEL MODELO

El modelo de regresión logística obtenido, nos indica un ajuste adecuado de las variables de estudio y la variable IM.

Las pruebas de bondad y ajuste que ilustraron el grado de clasificación del modelo, con las pruebas Chi cuadrado, Hosmer-Lemeshow y el estadístico de verosimilitud del modelo indican un adecuado ajuste de clasificación de los datos observados.

El estadístico de verosimilitud indica un adecuado ajuste del modelo, sin embargo se observa que la disminución de este valor al introducir las variables independientes es mínimo, pero se evidencia reducción de este valor al comparar este dato con el que arroja el estadístico cuando realiza el cálculo sin las introducir las variables independientes.

El modelo seleccionado, tiene un 69.8% de casos bien clasificados y el 30.2% corresponde a los casos mal clasificados.

Al realizar el estudio de multicorrelaciones del modelo se observa ausencia de correlaciones entre las variables de estudio.

A pesar de que todas las variables del modelo son significativas, las variables de mayor aporte en el modelo son: S8, S9, S6, L7, L10, B9, B6.

El pronóstico que realiza el modelo, indica que cuanto mejor es el record académico de los estudiantes la probabilidad de estar por debajo del nivel muy superior en el IM disminuye; no obstante, se evidencian casos de estudiantes que presentan un rendimiento académico adecuado, sin embargo, los resultados en las pruebas de estado en matemáticas, muestra una probabilidad alta de estar por debajo del nivel muy superior. También se observan casos donde algunos estudiantes que tiene un desempeño académico bajo, muestra una probabilidad baja de estar por debajo del nivel muy superior. Estos dos sucesos describen la realidad que año tras año envuelve este sistema de evaluación.

## **4.2. Modelo para el área de Ciencias**

### **4.2.1. Modelo Biología**

Modelo Inicial:

Se tomó como Variable dependiente ICFES Biología y como variables independientes: M6, M7, M8, M9, M10, C6, C7, C8, C9, C10, S6, S7, S8, S9, B6, B7, B8, B9, QF.

Para obtener el valor estimado del predictor lineal  $\eta = x\beta$  y el valor esperado del modelo de regresión logística se utilizó el método de mínimos cuadrados ponderados generalizados iterados, obteniendo los resultados siguientes:

**Tabla 19. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	-0,155	0,337	0,645
B7	-0,411	0,282	0,145
B8	-0,312	0,348	0,370
B9	-0,219	0,262	0,404
QF	0,140	0,255	0,583
M6	0,299	0,447	0,503
M7	-0,380	0,493	0,441
M8	-0,897	0,403	0,026
M9	-0,332	0,410	0,418
M10	-0,041	0,299	0,890
S6	-0,312	0,476	0,512
S7	-0,367	0,425	0,387
S8	0,979	0,481	0,042
S9	0,102	0,440	0,816
L6	0,078	0,408	0,848
L7	0,128	0,365	0,725
L8	0,013	0,401	0,975
L9	-1,116	0,455	0,014
L10	-0,085	0,399	0,832
Constante	10,630	2,252	0,000

**Modelo de regresión**

$$IB = \frac{e^v}{1 + e^v}$$

Donde  $v$  es  $x'\beta$

$$v = 10,6295 - 0,154845*B6 - 0,411015*B7 - 0,311501*B8 - 0,218561*B9 - 0,0847331*L10 + 0,0782534*L6 + 0,128361*L7 + 0,0127122*L8 - 1,11629*L9 - 0,0414022*M10 + 0,299256*M6 - 0,379767*M7 - 0,896534*M8 - 0,3315*M9 + 0,140263*QF - 0,312382*S6 - 0,367261*S7 + 0,979117*S8 + 0,102485*S9$$

Se realizó un análisis de varianza para estudiar e ajuste del modelo, del cual se obtuvieron los resultados que se presentan en la tabla 20.

**Tabla 20. Análisis de varianza para la regresión**

Fuente	Desviación	G.I.	Valor P
Modelo	60,8553	19	0,0000
Residuos	415,088	324	0,0005
Total	475,943	343	

Como el valor p para el modelo en la tabla anterior es inferior a 0.01, se puede concluir que hay una relación estadísticamente significativa entre las variables de estudio.

Se observó el estadístico de Cox y Snell para cuantificar la bondad del ajuste, en este caso es igual a 16.21%, este valor puede suponer que el modelo explica de manera no tan adecuada la variabilidad del mismo. Sin embargo el porcentaje ajustado de Nagelkerke, el cual, es más adecuado para comparar modelos con diferentes números de variables independientes es 21,63%, al parecer mejora el ajuste, no obstante, los resultados no son satisfactorios, por lo cual se trata de mejorar el modelo identificando aquellas variables que no aportan significativamente al modelo. Aquí se aplicó el sistema paso a paso hacia atrás obteniendo los siguientes resultados:

Modelo 2

**Tabla 21. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B7	-0,356	0,263	0,176
B8	-0,337	0,341	0,322
B9	-0,195	0,244	0,423
M7	-0,347	0,467	0,457

M8	-0,898	0,392	0,022
M9	-0,345	0,399	0,386
S7	-0,390	0,372	0,295
S8	1,030	0,457	0,024
L9	-1,130	0,418	0,007
Constante	10,950	2,001	0,000

El modelo resultante es:

$$v = 10,950 - 0,356*B7 - 0,337*B8 - 0,195*B9 - 1,130*L9 - 0,347*M7 - 0,898*M8 - 0,345*M9 - 0,390*S7 + 1,030*S8$$

El modelo es significativo según el análisis de varianza, sin embargo el ajuste baja con respecto al anterior modelo, puesto que los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 15,8% y 21,1%

Posteriormente se buscaron otros modelos de regresión logística construyendo una base donde las variables independientes tomaron los valores 1 y 0 (modelo 3), 0 si el estudiante tiene sobresaliente o excelente y 1 si tiene insuficiente o aprobado.

Modelo 3

**Tabla 22. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	0,072	0,372	0,846
B7	0,327	0,353	0,354
B8	0,468	0,384	0,223
B9	0,436	0,341	0,201

QF	-0,326	0,448	0,467
M6	-0,104	0,300	0,728
M7	0,277	0,368	0,451
M8	0,573	0,355	0,107
M9	0,498	0,405	0,219
M10	0,602	0,764	0,431
S6	0,336	0,391	0,390
S7	0,438	0,383	0,253
S8	-0,457	0,410	0,265
S9	-0,775	0,465	0,095
L6	-0,181	0,305	0,553
L7	0,261	0,298	0,381
L8	-0,387	0,345	0,261
L9	1,212	0,448	0,007
L10	0,260	0,565	0,646
Constante	-2,780	0,824	0,001

#### Modelo

$$v = -2,77954 + 0,0722461*B6 + 0,327236*B7 + 0,468035*B8 + 0,436198*B9 + 0,259748*L10 - 0,180706*L6 + 0,260652*L7 - 0,387395*L8 + 1,2118*L9 + 0,601939*M10 - 0,104334*M6 + 0,277191*M7 + 0,572552*M8 + 0,497633*M9 - 0,325775*QF + 0,336006*S6 + 0,437836*S7 - 0,456931*S8 - 0,775437*S9$$

En este modelo, el análisis mostró mejores resultados que el anterior pero casi similar al modelo 1, los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 15,6% y 20,86%.

Al igual que en el modelo inicial, se reducen las variables utilizando el método paso a paso hacia atrás, obteniendo el modelo 4.

Modelo 4

$$v = -2,269 + 0,446*B7 + 0,462*B8 + 0,332*B9 + 0,602*M8 + 0,514*M9 + 0,532*S7 - 0,475*S8 - 0,626*S9 + 1,231*L9$$

Sin embargo, según el análisis de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente los del 14,6% y 19,5%. Lo cual indica poco ajuste e inferior al del modelo inicial.

Finalmente se realizó el estudio con una base donde las variables independientes son los promedios en cada asignatura, se hizo el análisis con un modelo completo (modelo 5) y luego se eliminaron las variables menos significativas (modelo 6); los resultados arrojados por el software no indican un buen ajuste en ninguno de los dos casos, puesto que los de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente de 10,9% y 14,5% en el primer caso (modelo 5) y del 10,5% y 14% en el modelo reducido (modelo 6).

Del estudio anterior se puede deducir que el ajuste del modelo que describe la relación entre IB y el rendimiento académico es bajo y los modelos que mejor explican esta relación son los modelos 1 y 3, modelos a los cuales se les realizan otras pruebas para analizar la habilidad del modelo para efectos de clasificación.

### **Test de Bondad de ajuste Chi-cuadrado**

Se aplica la Prueba de Bondad de Ajuste Chi Cuadrado para evaluar la habilidad de clasificación del modelo.

Modelo 1

**Tabla 23. Bondad de ajuste**

Clase	Intervalo Logia	n	VERDADERO		FALSO	
			Observado	Esperado	Observado	Esperado
1	menor que -0,519084	69	16,0	14,0708	53,0	54,9292
2	-0,519084 a 0,16621	69	33,0	31,5434	36,0	37,4566
3	0,16621 a 0,425664	69	41,0	39,9308	28,0	29,0692
4	0,425664 a 0,681799	69	42,0	43,6103	27,0	25,3897
5	0,681799 o mayor	69	49,0	51,8446	19,0	16,1554
		68				
Total		344	181,0		163,0	

Chi-cuadrado = 1,34263 con 3 g.l. Valor P = 0,719032

Dado que el Valor P es superior a 0.10, no hay razón para descartar al modelo como medio de clasificación.

Al realizar el anterior estudio con los otros modelos, se obtuvieron los resultados que se indican en la tabla 24

**Tabla 24. Valor del Chi-cuadrado en cada modelo presente en el estudio**

MODELO	Chi-cuadrado	Valor P
1	1,34263	0,719032
2	2,0032	0,367291
3	4,27924	0,232844
4	2,62219	0,269525
5	0,239287	0,971011
6	1,71711	0,633134

Como se puede observar todos los modelos presentan una buena adecuación.

## Prueba de Hosmer-Lemeshow

### Modelo 1

**Tabla 25. Tabla de contingencias para la prueba de Hosmer y Lemeshow**

	IB = 0,00		IB = 1,00		Total
	Observado	Esperado	Observado	Esperado	
Paso 1	31	29,899	3	4,101	34
1 2	21	24,404	13	9,596	34
3	20	19,994	14	14,006	34
4	17	17,168	17	16,832	34
5	17	14,498	16	18,502	33
6	9	13,911	25	20,089	34
7	10	8,870	13	14,130	23
8	15	12,524	19	21,476	34
9	11	11,469	23	22,531	34
10	12	10,264	38	39,736	50

Chi-cuadrado = 7,134 con 8 G.L. Valor P = 0,522

Dado que el Valor P de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo, es decir el modelo se ajusta adecuadamente a los datos.

### Modelo 3

**Tabla 26. Tabla de contingencias para la prueba de Hosmer y Lemeshow**

	IB = 0		IB = 1		Total
	Observado	Esperado	Observado	Esperado	
1	31	30,199	3	3,801	34
2	20	24,906	14	9,094	34
3	22	19,884	12	14,116	34
4	20	17,036	14	16,964	34

5	15	14,759	19	19,241	34
6	12	13,163	22	20,837	34
7	4	5,171	11	9,829	15
8	21	20,868	44	44,132	65
9	13	10,123	21	23,877	34
10	4	5,891	22	20,109	26

Chi-cuadrado = 7,909 con 8 G.L. Valor P = 0, ,442

Dado que el Valor P de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo, es decir el modelo se ajusta adecuadamente a los datos, sin embargo, existe mejor adecuación de los datos en el modelo1.

### **Doble logaritmo del estadístico de Verosimilitud**

**AI** compara el valor del  $-2\log$  verosimilitud cuando el modelo no tiene variables independientes con el valor del  $-2\log$  verosimilitud cuando se introducen estas variables al modelo se obtuvo lo siguiente.

$-2\log$  verosimilitud sin variables independientes: 475,943

$-2\log$  verosimilitud con variables independientes: 415,0876835

Como el valor del  $-2\log$  verosimilitud se redujo, el ajuste es bueno, más no ideal porque esta reducción no tiende a cero.

Al aplicar este estadístico en los otros modelos observamos los resultados que se indican en la tabla 27.

**Tabla 27. Valor del -2log verosimilitud en cada modelo presente en el estudio**

MODELO	-2log verosimilitud. (A)	-2log verosimilitud. (D)
1	475,943	415,0876835
2		416,6404230983
3	474,212	415,9232657
4		436,1632080075
5	475,722	434,749883779
6		437,0851716299

Los resultados arrojados por el software en cada modelo indican un buen ajuste al modelo logístico, los modelos 1 y 3 son los que presentan mayor disminución, sin embargo, las reducciones son menores en cada modelo.

**Tabla 28. Tabla de clasificación para IB. Modelo 1**

Observado		Pronosticado		
		IB		Porcentaje correcto
		0,00	1,00	
Paso 1	IB 0,00	81	82	49,7
	1,00	39	142	78,5
Porcentaje global				64,8

**Tabla 29. Tabla de clasificación para IB. Modelo 3**

Observado		Pronosticado		
		IB		Porcentaje correcto
		0	1	
CN	0	85	77	52,5
	1	37	145	79,7
Porcentaje global				66,9

La ecuación del modelo ya diseñada nos proporciona una probabilidad que nos permite predecir a partir de ella para cada sujeto un valor de Y (Y predicho), tal que si  $P(Y=1|X) \leq 0.5$  entonces Y pred= 1, y si  $P(Y=1|X) > 0.5$  entonces Y pred=0

Al observar la lista de clasificación del modelo1 se contabilizan 121 casos mal clasificados de 344 casos, correspondientes al 38,9%, y en el modelo 3 se contabilizan 114 casos, que corresponden al 32,58%. Teniendo en cuenta esta última información el modelo que mejor ajusta los datos es el modelo 3. A continuación se indican los casos mal clasificados en este modelo:

**Tabla 30. Listado de casos Mal Clasificados**

Caso	Observado	Pronosticado	Grupo pronosticado
2	0	0,7	1
4	0	0,6	1
9	0	0,5	1
18	0	0,6	1
21	1	0,5	0
24	1	0,1	0
29	0	0,5	1
34	0	0,6	1
38	1	0,2	0
43	0	0,6	1
45	0	0,6	1
47	0	0,6	1
54	0	0,6	1
56	1	0,4	0
58	1	0,1	0
60	1	0,5	0
63	0	0,6	1
67	1	0,2	0
69	0	0,6	1
73	0	0,7	1
76	0	0,5	1

78	0	0,6	1
79	1	0,4	0
83	0	0,6	1
86	1	0,3	0
87	0	0,7	1
91	0	0,7	1
94	0	0,7	1
96	0	0,7	1
97	0	0,7	1
101	0	0,5	1
105	1	0,4	0
108	0	0,6	1
110	0	0,5	1
111	1	0,5	0
117	0	0,7	1
121	0	0,7	1
122	0	0,6	1
123	0	0,7	1
124	0	0,7	1
125	0	0,7	1
130	0	0,5	1
131	0	0,7	1
132	0	0,6	1
133	0	0,6	1
134	1	0,3	0
136	0	0,7	1
138	0	0,7	1
141	1	0,3	0
145	0	0,7	1
150	0	0,5	1
156	0	0,6	1
157	0	0,6	1
162	0	0,6	1
163	0	0,7	1
169	0	0,5	1

175	0	0,6	1
176	0	0,5	1
177	1	0,1	0
178	1	0,2	0
180	1	0,5	0
181	0	0,6	1
183	1	0,5	0
184	0	0,6	1
187	0	0,6	1
194	0	0,7	1
195	0	0,7	1
199	1	0,3	0
206	1	0,5	0
209	1	0,5	0
212	1	0,4	0
213	0	0,7	1
216	0	0,7	1
217	0	0,6	1
222	0	0,7	1
223	0	0,7	1
227	0	0,6	1
233	0	0,6	1
235	0	0,5	1
240	1	0,4	0
241	0	0,7	1
244	1	0,5	0
247	0	0,7	1
248	0	0,7	1
250	1	0,3	0
253	1	0,5	0
254	1	0,3	0
255	0	0,7	1
258	1	0,3	0
264	1	0,4	0
267	1	0,3	0

268	1	0,3	0
271	0	0,7	1
281	0	0,7	1
285	0	0,7	1
287	1	0,4	0
288	0	0,7	1
291	1	0,4	0
297	0	0,7	1
299	0	0,7	1
307	1	0,2	0
311	1	0,4	0
313	0	0,7	1
316	1	0,5	0
318	1	0,4	0
320	0	0,6	1
324	0	0,7	1
328	0	0,7	1
329	0	0,7	1
334	0	0,7	1
336	0	0,7	1
338	0	0,7	1
341	1	0,3	0
343	0	0,7	1

Al realizar este mismo estudio para cada modelo encontramos los resultados de la tabla 31

**Tabla 31. Total de casos mal clasificados en cada modelo presente en el estudio**

MODELO	CASOS MAL CLASIFICADOS
1	121
2	119
3	114

4	119
5	126
6	124

## Residuos atípicos

**Tabla 32. Residuos atípicos para IB**

Fila	Y	Predicho Y	Residuo	Residuo Pearson	Desviación Residuo
24	1	0,13922	0,86078	2,49	1,99
38	1	0,16283	0,83717	2,27	1,91
58	1	0,149103	0,850897	2,39	1,95
67	1	0,178813	0,821187	2,14	1,86
177	1	0,142319	0,857681	2,45	1,97

La tabla de residuos atípicos lista todas las observaciones que tienen residuo Pearson o desviación de residuo superiores a 2.0 en valor absoluto. Estos residuos estandarizados miden cuántas desviaciones estándar de cada valor observado de IB proceden del modelo ajustado. En este caso, hay 11 residuos estandarizados superiores a 2.0, y 2 superiores a 3.0.

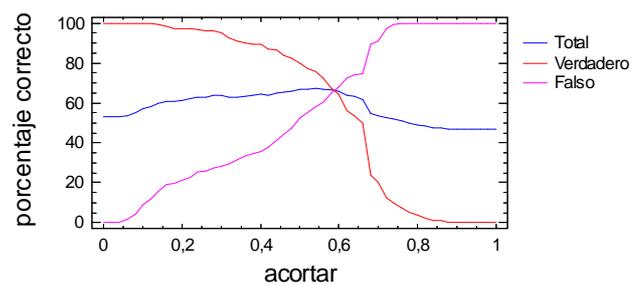
Al estudiar con detalle estos individuos se observa que aquellos que están por ENCIMA del nivel muy superior (0), su nivel académico es BAJO; y aquellos que están por DEBAJO del nivel muy superior (1), presentan un rendimiento académico BUENO en su vida escolar. Por lo tanto, es conveniente mantenerlos en el estudio puesto que son casos reales que no presentan errores al levantamiento ni digitación de la información ya que se confrontó la base de datos con la información de los libros de registro y la percepción de profesores antiguos de la institución sobre la veracidad de la información.

## PREDICCIÓN DEL MODELO

Tabla 33. Capacidad de Predicción del Modelo

Cohorte	VERDADERO	FALSO	Total
0,0	100,00	0,00	52,91
0,05	100,00	0,00	52,91
0,1	100,00	8,64	56,98
0,15	98,35	16,67	59,88
0,2	97,25	20,99	61,34
0,25	96,15	25,31	62,79
0,3	95,05	28,40	63,66
0,35	90,66	31,48	62,79
0,4	89,56	35,80	64,24
0,45	85,16	42,59	65,12
0,5	79,67	52,47	66,86
0,55	73,63	58,64	66,57
0,6	64,29	67,90	65,99
0,65	51,10	74,69	62,21
0,7	20,33	91,36	53,78
0,75	8,24	100,00	51,45
0,8	3,85	100,00	49,13
0,85	1,10	100,00	47,67
0,9	0,00	100,00	47,09
0,95	0,00	100,00	47,09
1,0	0,00	100,00	47,09

Figura 2. Capacidad de Predicción del Modelo para IB



En la tabla y el gráfico se observa la capacidad de predicción del modelo 1. Si el valor predicho es más grande que el punto de cohorte, la respuesta se predice como VERDADERA. Si el valor predicho es inferior a o igual al punto de cohorte, la respuesta se predice para ser FALSO. La tabla muestra el porcentaje de datos observados predichos correctamente a diferentes puntos de cohorte. En nuestro caso con el punto de cohorte utilizado 0,5, el 79,67% de todas las respuestas VERDADERAS se predijeron correctamente, mientras que el 52,47% de todas las respuestas FALSAS se predijeron correctamente, para un total de 66,86% casos clasificados correctamente.

**Tabla 34. Matriz de correlación para los coeficientes estimados**

	Cons	B6	B7	B8	B9	QF	M6	M7	M8	M9	M10	S6	S7	S8	S9	L6	L7	L8	L9	L10
Cons	1																			
B6	-0,071	1																		
B7	0,142	0,003	1																	
B8	-0,25	-0,08	-0,08	1																
B9	-0,134	0,081	-0	0,129	1															
QF	-0,123	-0,04	0,053	0,002	-0,32	1														
M6	0,022	-0,07	-0,16	0,022	0,094	-0,02	1													
M7	-0,052	-0,19	-0,05	0,005	0,002	-0,04	-0,09	1												
M8	-0,131	-0,07	-0,16	0,076	0,04	-0,02	0,016	-0,07	1											
M9	-0,026	0,023	0,01	-0,15	0,078	-0,11	0,098	-0,14	-0,29	1										
M10	-0,72	-0	-0,2	0,014	0,04	-0,03	-0,17	0,034	0,029	-0,07	1									
S6	0,057	-0,24	0,078	-0,09	-0	-0,03	-0,16	0,098	-0,09	-0,04	-0,07	1								
S7	-0,088	-0,22	-0,17	-0,04	0,003	0,065	-0,03	-0,15	0,096	-0,07	0,014	-0,15	1							
S8	0,004	0,089	0,107	-0,1	-0,13	-0,04	-0,06	-0,16	-0,13	-0,06	-0,02	-0,04	-0,17	1						
S9	-0,005	-0	-0,19	-0,09	-0,25	0,018	-0,01	0,038	0,066	-0,08	-0,02	-0,18	0,052	-0,13	1					
L6	0,029	-0,12	-0,12	-0,01	-0,01	-0,08	-0,14	0,042	0,061	-0,05	0,136	-0,2	-0,04	-0,1	0,028	1				
L7	-0,01	-0,12	-0,17	0,048	0	0,021	-0,08	0,037	-0,02	0,012	0,104	-0,08	-0,05	0,027	-0,08	-0,17	1			
L8	0,102	0,1	0,084	-0,11	-0,02	-0,07	0,09	-0,27	-0,13	-0	0,002	-0,16	-0,16	-0,03	0,069	-0,03	-0,22	1		
L9	-0,037	-0,06	-0,02	0,048	-0,1	0,078	-0,01	0,059	0,001	-0,15	-0,05	0,033	0,064	-0,1	-0,14	-0,05	0,016	-0,14	1	
L10	-0,134	0,027	-0,09	0,01	0,015	-0,28	0,055	-0,04	-0,02	0,08	-0,09	0,087	0,018	-0,04	-0,22	-0,04	-0,04	-0,08	-0,35	1

Esta tabla muestra las correlaciones estimadas entre los coeficientes en el modelo ajustado. Estas correlaciones pueden utilizarse para detectar la presencia de serias multicorrelaciones, es decir, correlación entre las variables pronosticadas. En este caso, no hay ninguna correlación con valores absolutos superiores a 0.5.

## INTERPRETACION DEL MODELO

Como se puede observar en el estudio anterior, el modelo 3 presenta un adecuado ajuste, las variables que ayudan a predecir si un estudiante de grado décimo del colegio Javeriano estará por encima o debajo del nivel muy superior en las pruebas de estado ICFES en la asignatura de Biología son todas las materias de matemáticas, castellano, sociales y naturales, que cursó desde grado sexto a grado décimo en la institución codificadas 1 si su rendimiento fue insuficiente o aceptable y 0 si su rendimiento fue excelente o sobresaliente.

A continuación se indican el valor de las probabilidades que tiene cada variable de estudio en el modelo.

**Tabla 35. Variables en el modelo**

	B	Exp(B)
B6	0,072	1,075
B7	0,327	1,387
B8	0,468	1,597
B9	0,436	1,547
QF	-0,326	0,722
M6	-0,104	0,901
M7	0,277	1,319
M8	0,573	1,773
M9	0,498	1,645
M10	0,602	1,826

	B	Exp(B)
S6	0,336	1,399
S7	0,438	1,549
S8	-0,457	0,633
S9	-0,775	0,461
L6	-0,181	0,835
L7	0,261	1,298
L8	-0,387	0,679
L9	1,212	3,36
L10	0,26	1,297
Constante	-2,78	0,062

De lo anterior se puede decir que el valor de 0,072 que se asocia a la variable B6, incrementa el 107% la posibilidad de estar por debajo del nivel muy superior en el área de BIOLOGÍA, puesto que  $e^{0,072} = 1,075$

De igual manera se pueden leer todos los coeficientes que acompañan a cada variable de estudio.

Como se puede observar, las variables QF, M6, S8, S9, L6 y L8 establecen un aporte bajo al cálculo de esta probabilidad, mientras que B8, B9, M8, M9, M10 y S7 aporta considerablemente a este cálculo.

El modelo seleccionado (modelo 3) ayudará a detectar a aquellos estudiantes de grado décimo que requieren de un refuerzo académico antes de presentar la prueba ICFES. A continuación se indica la probabilidad de que cada estudiante de grado décimo pueda estar por debajo del nivel muy superior. (Tabla 36)

**Tabla 36. Casos posibles que necesitan refuerzo académico en biología antes de las pruebas ICFES**

Caso	B6	B7	B8	B9	QF	L6	L7	L8	L9	L10	M6	M7	M8	M9	M10	S6	S7	S8	S9		P(y=1)
1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0,343206	0,585
1	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	1	-0,300882	0,4253
3	1	1	1	0	0	1	1	0	0	0	0	1	1	1	1	1	0	1	0	-0,003687	0,4991
4	0	0	0	1	1	0	0	0	1	1	0	1	1	1	1	0	0	0	1	-0,023691	0,4941
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
6	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	-2,317192	0,0897
7	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	-1,759161	0,1469
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
11	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	1,187532	0,7663
13	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1	0	0	0	-0,436506	0,3926

14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,281907	0,0926
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
16	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	-2,594383	0,0695
17	1	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0	-0,646328	0,3438
18	1	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0	0	1	0	-1,097038	0,2503
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	-2,206988	0,0991
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
16	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,238872	0,5594
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
18	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	-1,432164	0,1928
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	-2,443534	0,0799
30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
31	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0,343206	0,585
33	0	0	0	1	1	0	0	0	0	0	1	1	0	1	1	0	0	0	0	-1,396688	0,1983
34	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	-1,774497	0,145
35	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	-2,460153	0,0787
36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0,7680491	0,6831
37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0,4320431	0,6064
38	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
39	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	1	-0,873434	0,2945
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
41	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	-1,260829	0,2208
41	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	-0,261434	0,435
43	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	-0,016118	0,496
44	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
45	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	-1,547573	0,1754
46	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	1	-0,300882	0,4253
47	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0,092342	0,5231
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
49	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	-0,995966	0,2697
50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
51	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	1	-0,873434	0,2945
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
53	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0,4154521	0,6024
54	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0,4154521	0,6024
55	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
56	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772

57	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,7489541	0,679
58	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,676708	0,663
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,77954	0,0584
60	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	-1,432164	0,1928
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	-3,675902	0,0247
61	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
63	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1,5434861	0,824
64	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	1,575574	0,8286
65	0	0	0	1	1	0	0	0	0	0	1	1	0	1	1	0	0	0	0	-1,396688	0,1983
66	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	1	-3,006865	0,0471
67	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	-3,692521	0,0243
68	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,3111181	0,5772
69	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1,2074801	0,7699
70	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1,5434861	0,824
71	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	-0,097997	0,4755
71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-3,554977	0,0278
73	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	-1,260829	0,2208
74	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	-0,261434	0,435
75	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0,4408131	0,6085
76	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1,5434861	0,824
77	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	-1,547573	0,1754
78	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	1	1	-0,757813	0,3191
79	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0,092342	0,5231
80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	-4,011908	0,0178

## CONCLUSIONES DEL MODELO

El modelo de regresión logística obtenido, nos indica un ajuste adecuado de las variables de estudio y la variable IB.

Las pruebas de bondad y ajuste que ilustraron el grado de clasificación del modelo, con las pruebas Chi cuadrado, Hosmer-Lemeshow y el estadístico de verosimilitud del modelo indicaron un adecuado ajuste de clasificación de los datos observados.

El estadístico de verosimilitud indica un adecuado ajuste del modelo, sin embargo se observa que la disminución de este valor al introducir las variables independientes es mínimo, pero se evidencia reducción de este valor al comparar este dato con el que arroja el estadístico cuando realiza el cálculo sin las introducir las variables independientes.

El modelo seleccionado, tiene un 669% de casos bien clasificados y el 32.59% corresponde a los casos mal clasificados.

Al realizar el estudio de multicorrelaciones del modelo se observa ausencia de correlaciones entre las variables de estudio.

A pesar de que todas las variables del modelo son significativas, las variables de menor aporte en el modelo son: QF, M6, S8, S9, L6 y L8.

El pronóstico que realiza el modelo indica que cuanto mejor es el record académico de los estudiantes la probabilidad de estar por debajo del nivel muy superior en el IB disminuye; no obstante, se evidencian casos de estudiantes que presentan un rendimiento académico adecuado, sin embargo, los resultados en las pruebas de estado en biología, muestra una probabilidad alta de estar por debajo del nivel muy superior. También se observan casos donde algunos estudiantes que tiene un desempeño académico bajo, muestra una probabilidad baja de estar por debajo del nivel muy superior. Estos dos sucesos describen la realidad que año tras año envuelve este sistema de evaluación.

#### 4.2.2. Modelo Química

Modelo Inicial:

Se tomó como Variable dependiente IQ y como variables independientes: M6, M7, M8, M9, M10, C6, C7, C8, C9, C10, S6, S7, S8, S9, B6, B7, B8, B9, QF.

Para obtener el valor estimado del predictor lineal  $\eta = x_i\beta$  y el valor esperado del modelo de regresión logística se utilizó el método de mínimos cuadrados ponderados generalizados iterados, obteniendo los resultados siguientes:

**Tabla 37. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	-0,224	0,329	0,496
B7	-0,486	0,274	0,076
B8	-0,345	0,344	0,315
B9	0,114	0,250	0,648
QF	-0,171	0,249	0,492
M6	-0,443	0,432	0,305
M7	0,277	0,474	0,559
M8	-0,128	0,349	0,715
M9	-0,279	0,393	0,478
M10	-0,069	0,290	0,812
S6	0,119	0,458	0,795
S7	-0,136	0,415	0,742
S8	-0,008	0,453	0,985
S9	0,364	0,431	0,399
L6	0,078	0,395	0,843
L7	-0,034	0,358	0,925
L8	0,091	0,379	0,809
L9	-1,036	0,426	0,015
L10	-0,187	0,386	0,629
Constante	9,135	2,169	0,000

Los datos obtenidos indican el ajuste del modelo que describe la relación entre IQ y el rendimiento académico conformado por las 19 variables antes mencionadas.

### Modelo de regresión

$$IQ = \frac{e^v}{1 + e^v}$$

Donde  $v$  es  $x'\beta$

$$v = 9,13532 - 0,186679*L10 + 0,0782511*L6 - 0,03371*L7 + 0,091443*L8 - 1,0358*L9 - 0,0692019*M10 - 0,443494*M6 + 0,276761*M7 - 0,127515*M8 - 0,27926*M9 - 0,171208*QF + 0,118937*S6 - 0,136368*S7 - 0,00837214*S8 + 0,363593*S9 - 0,224303*B6 - 0,486246*B7 - 0,345169*B8 + 0,114228*B9$$

Se realizó un análisis de varianza para estudiar el ajuste del modelo, del cual se obtuvieron los resultados que se presentan en la tabla 38.

**Tabla 38. Análisis de varianza para la regresión**

Fuente	Desviación	G.I.	Valor P
Modelo	43,8997	19	0,0010
Residuos	432,974	324	0,0005
Total	476,874	343	

Como el valor p para el modelo en la tabla anterior es inferior a 0.01, se puede concluir que hay una relación estadísticamente significativa entre las variables de estudio.

Se observó el estadístico de Cox y Snell para cuantificar la bondad del ajuste, en este caso es igual a 12%, este valor puede suponer que el modelo explica de manera no tan adecuada la variabilidad del mismo. Sin embargo el porcentaje ajustado de Nagelkerke, el cual, es más adecuado para comparar modelos con

diferentes números de variables independientes es 16%, al parecer mejora el ajuste, no obstante, los resultados no son satisfactorios, por lo cual se trata de mejorar el modelo identificando aquellas variables que no aportan significativamente al modelo. Aquí se aplicó el sistema paso a paso hacia atrás obteniendo los siguientes resultados:

Modelo 2

**Tabla 39. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
L9	-1,070	0,386	0,006
S9	0,410	0,404	0,310
M6	-0,453	0,391	0,247
M7	0,252	0,426	0,555
M9	-0,305	0,378	0,420
B6	-0,223	0,291	0,444
B7	-0,521	0,264	0,048
B8	-0,346	0,335	0,302
QF	-0,161	0,229	0,483
Constante	8,886	1,978	0,000

El modelo resultante es:

$$v = 8,886 - 0,223 * B6 - 0,521 * B7 - 0,346 * B8 - 1,070 * L9 - 0,161 * QF - 0,453 * M6 - 0,252 * M7 - 0,305 * M9 + 0,410 S9$$

El modelo es significativo según el análisis de varianza, sin embargo el ajuste baja con respecto al anterior modelo, puesto que los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 11,8% y 15,7%

Posteriormente se buscaron otros modelos de regresión logística construyendo una base donde las variables independientes tomaron los valores 1 y 0 (modelo 3), 0 si el estudiante tiene sobresaliente o excelente y 1 si tiene insuficiente o aprobado.

### Modelo 3

**Tabla 40. Resultados de las estimaciones del modelo**

Variabes	Estimación $\beta$	Error Estandar	Valor P
B6	0,44069656	0,36491142	0,22717002
B7	0,50473826	0,35352119	0,15336514
B8	0,45207904	0,38129734	0,23576693
B9	0,11871832	0,33283369	0,72132418
QF	-0,61425236	0,43900771	0,16175795
M6	0,19058078	0,28965914	0,51057125
M7	-0,4969479	0,37345133	0,18329108
M8	0,04702365	0,35674191	0,89513126
M9	0,22913713	0,40176981	0,56846106
M10	1,07406958	0,84318751	0,20272702
S6	-0,06194381	0,38279277	0,87144702
S7	0,25939434	0,37733336	0,49180501
S8	0,42840108	0,3917536	0,27415359
S9	-0,82049263	0,45608093	0,07201763
L6	-0,08536912	0,29805222	0,77455389
L7	0,19655458	0,29356982	0,50315576
L8	-0,04083378	0,33383147	0,90264674
L9	1,01848915	0,43317703	0,01871259
L10	0,4396544	0,55397441	0,42740699
Constante	-2,72910206	0,87584801	0,00183351

## Modelo

$$\begin{aligned}v = & -2,7291 + 0,440697*B6 + 0,504738*B7 + 0,452079*B8 + 0,118718*B9 + \\ & 0,439654*L10 - 0,0853691*L6 + 0,196555*L7 - 0,0408338*L8 + \\ & 1,01849*L9 + 1,07407*M10 + 0,190581*M6 - 0,496948*M7 + \\ & 0,0470236*M8 + 0,229137*M9 - 0,614252*QF - 0,0619438*S6 + \\ & 0,259394*S7 + 0,428401*S8 - 0,820493*S9\end{aligned}$$

En este modelo, el análisis mostró mejores resultados que el anterior pero casi similar al modelo 1, los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 12,8% y 17%.

Al igual que en el modelo inicial, se redujeron las variables utilizando el método paso a paso hacia atrás, Sin embargo, según el análisis de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente los del 12,1% y 16,1%. Lo cual indica poco ajuste y es casi similar a los resultados de los modelos anteriores.

Finalmente se realizó el estudio con una base donde las variables independientes son los promedios en cada asignatura, se hizo el análisis con un modelo completo (modelo 5) y luego se eliminaron las variables menos significativas (modelo 6); los resultados arrojados por el software no indican un mejor ajuste en ninguno de los dos casos, puesto que los de estadísticos de Cox y Snell y Nagelkerke, respectivamente son 9% y 11,9% para el modelo 5 y 8% y 11,4%, para el modelo 6.

Del estudio anterior se puede deducir que el ajuste del modelo que describe la relación entre IQ y el rendimiento académico es bajo y el modelo que mejor explica esta relación es el modelo 3, modelo al cual se le realizara otras pruebas para analizar la habilidad del modelo para efectos de clasificación.

## Test de Bondad de ajuste Chi-cuadrado

Se aplica la Prueba de Bondad de Ajuste Chi Cuadrado para evaluar la habilidad de clasificación del modelo.

Modelo 1

**Tabla 41. Bondad de ajuste**

Clase	Logia Intervalo	n	VERDADERO		FALSO	
			Observado	Esperado	Observado	Esperado
1	menor que -0,696606	69	16,0	14,5829	40,0	39,6787
2	-0,696606 a 0,0668883	69	29,0	29,3213	53,0	54,4171
3	0,0668883 a 0,431877	74	39,0	42,2599	35,0	31,7401
4	0,431877 a 0,550595	85	56,0	53,5709	29,0	31,4291
5	0,550595 o mayor	47	33,0	33,2649	14,0	13,7351
Total		344	173,0		171,0	

Chi-cuadrado = 1,07212 con 3 g.l. Valor P = 0,783807

Dado que el Valor P es superior a 0.10, no hay razón para descartar al modelo como medio de clasificación.

Al realizar el anterior estudio con los otros modelos, se obtuvieron los resultados que se indican en la tabla 42.

**Tabla 42. Valor del Chi-cuadrado en cada modelo presente en el estudio**

MODELO	Chi-cuadrado	Valor P
1	3,54421	0,31507
2	5,03053	0,169572
3	1,07212	0,783807
4	0,688913	0,708606
5	0,762761	0,85835
6	0,795932	0,850438

Como se puede observar todos los modelos presentan una buena adecuación.

**Tabla 43. Tabla de contingencias para la prueba de Hosmer y Lemeshow**

		Q = 0		Q = 1		Total
		Observado	Esperado	Observado	Esperado	
Paso	1	29	29,547	5	4,453	34
1	2	23	24,202	11	9,798	34
	3	22	21,246	12	12,754	34
	4	19	18,133	15	15,867	34
	5	18	15,495	16	18,505	34
	6	15	14,052	19	19,948	34
	7	8	10,816	20	17,184	28
	8	23	23,772	42	41,228	65
	9	14	13,735	33	33,265	47

Chi-cuadrado = 2,539 con 7 g.l. Valor P = 0,924

Dado que el Valor P de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo decir el modelo se ajusta adecuadamente a los datos.

### Doble logaritmo del estadístico de Verosimilitud

Al compara el valor del  $-2\log$  verosimilitud cuando el modelo no tiene variables independientes con el valor del  $-2\log$  verosimilitud cuando se introducen estas variables al modelo se obtuvo lo siguiente.

$-2\log$  verosimilitud sin variables independientes: 476,8736322528

$-2\log$  verosimilitud con variables independientes: 441,1521318515

Como el valor del  $-2\log$  verosimilitud se redujo, el ajuste es bueno, más no ideal porque esta reducción no tiende a cero.

Al aplicar este estadístico en los otros modelos observamos los resultados que se indican en la tabla 44

**Tabla 44. Valor del  $-2\log$  verosimilitud en cada modelo presente en el estudio**

MODELO	$-2\log$ verosimilitud. (A)	$-2\log$ verosimilitud. (D)
1	476,8736323	432,9739432
2		433,8321941714
3		432,9739432106
4		432,5008676788
5		444,5869508
6		446,0604190899

Los resultados arrojados por el software en cada modelo indican un buen ajuste al modelo logístico, los modelos 1,3 y 4 son los que presenta mayor disminución, sin embargo, las reducciones son menores en cada modelo.

**Tabla 45. Tabla de clasificación para IQ**

Observado		Pronosticado		
		Q		Porcentaje correcto
		0	1	
Q	0	91	80	53,2
	1	40	133	76,9
Porcentaje global				65,1

La ecuación del modelo ya diseñada nos proporciona una probabilidad que nos permite predecir a partir de ella para cada sujeto un valor de Y (Y predicho), tal que si  $P(Y=1|X) \leq 0.5$  entonces  $Y_{pred} = 1$ , y si  $P(Y=1|X) > 0.5$  entonces  $Y_{pred} = 0$

Al observar la lista de clasificación del modelo3 se contabilizan 121 casos mal clasificados de 344 casos, correspondientes al 35,1%. Teniendo en cuenta esta información el modelo que mejor ajusta los datos es el modelo 3. A continuación se indican los casos mal clasificados en este modelo:

**Tabla 46. Listado de casos Mal Clasificados**

Caso	Observado	Pronosticado	Grupo pronosticado
1	0	0,79211781	1
2	0	0,63210489	1
4	0	0,60632176	1
6	1	0,41064761	0
8	0	0,5405274	1
13	1	0,3617596	0
25	0	0,63427367	1
26	0	0,55856059	1
34	0	0,54077421	1
38	1	0,31815647	0
47	0	0,57949275	1
51	0	0,67661175	1

54	0	0,52959322	1
55	0	0,57949275	1
56	1	0,35914445	0
57	0	0,58759708	1
58	1	0,36085016	0
60	1	0,31972274	0
63	0	0,64689823	1
66	1	0,17260723	0
68	0	0,77770977	1
70	0	0,78311527	1
76	1	0,46139367	0
79	1	0,42245979	0
83	0	0,69689742	1
86	1	0,30737014	0
87	0	0,58904393	1
88	0	0,52907621	1
91	0	0,63427367	1
94	0	0,58904393	1
97	0	0,63427367	1
99	1	0,30355373	0
106	1	0,34505399	0
108	0	0,60032686	1
110	1	0,48212802	0
116	0	0,61166821	1
118	0	0,58759708	1
120	0	0,63427367	1
121	0	0,52744704	1
122	0	0,57707387	1
124	0	0,69222931	1
128	0	0,63427367	1
130	0	0,62329829	1
131	0	0,63427367	1
133	0	0,52303673	1
135	1	0,46239673	0
136	0	0,63427367	1

137	1	0,47490299	0
138	0	0,63427367	1
139	0	0,53082624	1
141	1	0,27086943	0
145	0	0,63427367	1
147	0	0,51287029	1
148	0	0,63427367	1
150	0	0,55667565	1
153	1	0,15250404	0
162	0	0,56003303	1
165	1	0,38906077	0
166	0	0,55089651	1
167	0	0,56188825	1
174	1	0,37597688	0
177	1	0,32056994	0
178	1	0,15973132	0
179	1	0,28636699	0
180	0	0,53679159	1
181	1	0,43046909	0
182	1	0,34216588	0
183	1	0,4189802	0
186	0	0,5114195	1
187	0	0,52460914	1
195	0	0,63427367	1
196	0	0,66584043	1
197	0	0,64529902	1
199	1	0,36987511	0
200	0	0,74807596	1
202	0	0,66396314	1
206	0	0,53652473	1
218	0	0,60811837	1
221	0	0,63427367	1
222	0	0,55089651	1
225	1	0,47834956	0
227	0	0,5860963	1

234	1	0,29623558	0
235	1	0,39809043	0
237	1	0,45143671	0
238	0	0,58860695	1
241	0	0,63427367	1
243	1	0,47700223	0
249	0	0,63427367	1
250	1	0,30633076	0
251	0	0,66302369	1
254	1	0,3376401	0
255	0	0,63427367	1
257	0	0,57967956	1
258	1	0,28247868	0
261	0	0,58759708	1
264	1	0,49247338	0
265	0	0,63427367	1
268	1	0,40200725	0
269	0	0,63427367	1
271	0	0,63427367	1
285	0	0,58904393	1
287	1	0,29027344	0
288	0	0,63427367	1
289	0	0,53275952	1
291	1	0,17030608	0
292	1	0,45217971	0
297	0	0,58904393	1
299	0	0,58904393	1
307	1	0,10341953	0
313	0	0,65384062	1
317	0	0,63427367	1
320	0	0,58759708	1
323	0	0,63427367	1
334	0	0,63427367	1
336	0	0,63427367	1
337	0	0,60954107	1

338	0	0,60954107	1
342	0	0,55170531	1
343	0	0,65384062	1

Al realizar este mismo estudio para cada modelo encontramos los resultados de la tabla 47

**Tabla 47. Total de casos mal clasificados en cada modelo presente en el estudio**

MODELO	CASOS MAL CLASIFICADOS
1	129
2	130
3	120
4	125
5	132
6	132

### Residuos atípicos

**Tabla 48. Residuos atípicos para IQ**

Fila	Y	Predicho Y	Residuo	Residuo Pearson	Desviación Residuo
63	1	0,172607	0,827393	2,19	1,87
153	1	0,152504	0,847496	2,36	1,94
178	1	0,159731	0,840269	2,29	1,92
291	1	0,170306	0,829694	2,21	1,88
307	1	0,10342	0,89658	2,94	2,13

La tabla de residuos atípicos lista todas las observaciones que tienen residuo Pearson o desviación de residuo superiores a 2.0 en valor absoluto. Estos residuos estandarizados miden cuántas desviaciones estándar de cada valor observado de IQ proceden del modelo ajustado. En este caso, hay 5 residuos estandarizados superiores a 2.0, y 2 superiores a 3.0.

Al estudiar con detalle estos individuos se observa que aquellos que están por ENCIMA del nivel muy superior (0), su nivel académico es BAJO; y aquellos que están por DEBAJO del nivel muy superior (1), presentan un rendimiento académico BUENO en su vida escolar. Por lo tanto, es conveniente mantenerlos en el estudio puesto que son casos reales que no presentan errores al levantamiento ni digitación de la información ya que se confrontó la base de datos con la información de los libros de registro y la percepción de profesores antiguos de la institución sobre la veracidad de la información.

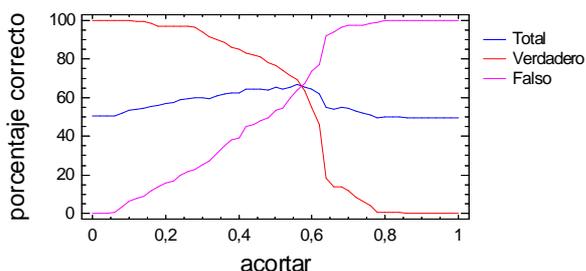
## PREDICCIÓN DEL MODELO

**Tabla 49. Capacidad de Predicción del Modelo**

Cohorte	VERDADERO	FALSO	Total
0,0	100,00	0,00	50,29
0,05	100,00	0,58	50,58
0,1	100,00	6,43	53,49
0,15	99,42	11,11	55,52
0,2	97,11	15,79	56,69
0,25	97,11	19,88	58,72
0,3	94,22	25,15	59,88
0,35	89,02	32,16	60,76
0,4	84,97	39,18	62,21

0,45	82,08	46,78	64,53
0,5	76,88	53,22	65,12
0,55	70,52	60,82	65,70
0,6	54,91	73,68	64,24
0,65	16,76	92,98	54,65
0,7	12,14	97,66	54,65
0,75	4,62	98,25	51,16
0,8	0,58	100,00	50,00
0,85	0,00	100,00	49,71
0,9	0,00	100,00	49,71
0,95	0,00	100,00	49,71
1,0	0,00	100,00	49,71

**Figura 3. Capacidad de Predicción del Modelo para IQ**



En la tabla y el gráfico se observa la capacidad de predicción del modelo 1. Si el valor predicho es más grande que el punto de cohorte, la respuesta se predice como VERDADERA. Si el valor predicho es inferior a o igual al punto de cohorte, la respuesta se predice para ser FALSO. La tabla muestra el porcentaje de datos observados predichos correctamente a diferentes puntos de cohorte. En nuestro caso con el punto de cohorte utilizado 0,5, el 76,88% de todas las respuestas VERDADERAS se predijeron correctamente, mientras que el 53,22% de todas las respuestas FALSAS se predijeron correctamente, para un total de 65,12% casos clasificados correctamente.

**Tabla 50. Matriz de correlación para los coeficientes estimados**

	Cons	B6	B7	B8	B9	QF	M6	M7	M8	M9	M10	S6	S7	S8	S9	L6	L7	L8	L9	L10
Co	1																			
B6	-0,09	1																		
B7	0,12	0,00	1																	
B8	-0,21	-0,08	-0,06	1																
B9	-0,11	0,09	-0,03	0,12	1															
QF	-0,07	-0,08	0,07	0,00	-0,29	1														
M6	0,03	-0,07	-0,14	0,02	0,10	-0,03	1													
M7	0,00	-0,19	-0,08	0,01	0,02	-0,02	-0,08	1												
M8	-0,08	-0,07	-0,17	0,07	0,03	-0,01	0,03	-0,07	1											
M9	0,00	0,02	0,00	-0,14	0,07	-0,10	0,10	-0,17	-0,31	1										
M10	-0,78	0,02	-0,16	-0,03	0,04	-0,06	-0,16	0,01	0,02	-0,07	1									
S6	0,08	-0,24	0,05	-0,09	-0,02	0,00	-0,15	0,08	-0,11	-0,06	-0,06	1								
S7	-0,08	-0,21	-0,17	-0,05	-0,01	0,05	-0,04	-0,17	0,08	-0,06	0,02	-0,15	1							
S8	-0,06	0,09	0,12	-0,08	-0,14	-0,06	-0,07	-0,16	-0,11	-0,02	0,01	-0,03	-0,15	1						
S9	-0,01	-0,02	-0,22	-0,08	-0,22	0,01	-0,01	0,03	0,08	-0,06	-0,01	-0,18	0,08	-0,14	1					
L6	0,01	-0,12	-0,11	-0,01	-0,01	-0,07	-0,16	0,06	0,07	-0,05	0,13	-0,19	-0,05	-0,11	0,05	1				
L7	0,00	-0,12	-0,16	0,04	0,00	0,02	-0,07	0,02	-0,03	0,01	0,09	-0,09	-0,04	0,04	-0,08	-0,17	1			
L8	0,07	0,10	0,09	-0,11	-0,01	-0,06	0,08	-0,26	-0,12	0,02	0,01	-0,13	-0,15	-0,05	0,07	-0,04	-0,22	1		
L9	-0,02	-0,03	0,00	0,06	-0,12	0,06	0,02	0,02	-0,06	-0,19	-0,04	0,00	0,05	-0,06	-0,16	-0,04	0,00	-0,10	1	
L10	-0,13	0,06	-0,10	0,01	-0,02	-0,29	0,05	-0,05	-0,04	0,08	-0,08	0,08	0,02	-0,01	-0,22	-0,05	-0,02	-0,10	-0,31	1

Esta tabla muestra las correlaciones estimadas entre los coeficientes en el modelo ajustado. Estas correlaciones pueden utilizarse para detectar la presencia de serias multicorrelaciones, es decir, correlación entre las variables pronosticadas. En este caso, no hay ninguna correlación con valores absolutos superiores a 0.5.

## INTERPRETACION DEL MODELO

Como se puede observar en el estudio anterior, el modelo 3 presenta un adecuado ajuste, las variables que ayudan a predecir si un estudiante de grado décimo del colegio Javeriano estará por encima o debajo del nivel muy superior en las pruebas de estado ICFES en la asignatura de Química son todas las materias de matemáticas, castellano, sociales y naturales, que cursó desde grado sexto a grado décimo en la institución codificadas 1 si su rendimiento fue insuficiente o aceptable y 0 si su rendimiento fue excelente o sobresaliente.

A continuación se indican el valor de las probabilidades que tiene cada variable de estudio en el modelo.

**Tabla 51. Variables en el modelo**

	B	Exp(B)
B6	0,44069656	1,55378915
B7	0,50473826	1,65655188
B8	0,45207904	1,57157616
B9	0,11871832	1,12605269
QF	-0,61425236	0,54104525
M6	0,19058078	1,20995211
M7	-0,4969479	0,60838468
M8	0,04702365	1,0481468
M9	0,22913713	1,25751447
M10	1,07406958	2,92726803

	B	Exp(B)
S6	-0,06194381	0,9399357
S7	0,25939434	1,29614482
S8	0,42840108	1,53480154
S9	-0,82049263	0,44021474
L6	-0,08536912	0,91817331
L7	0,19655458	1,21720175
L8	-0,04083378	0,95998868
L9	1,01848915	2,76900804
L10	0,4396544	1,55217069
Con		

De lo anterior se puede decir que el valor de 0,44 que se asocia a la variable B6, incrementa el 155% la posibilidad de estar por debajo del nivel muy superior en el área de QUIMICA, puesto que  $e^{0,44} = 1,55$

De igual manera se pueden leer todos los coeficientes que acompañan a cada variable de estudio.

Como se puede observar, las variables QF, M7, S6, S9, L6 y L8 establecen un aporte bajo al cálculo de esta probabilidad, mientras que M10 y S9 aporta considerablemente a este cálculo.

El modelo seleccionado (modelo 3) ayudará a detectar a aquellos estudiantes de grado décimo que requieren de un refuerzo académico antes de presentar la prueba ICFES. A continuación se indica la probabilidad de que cada estudiante de grado décimo pueda estar por debajo del nivel muy superior.

**Tabla 52. Casos posibles que necesitan refuerzo académico en química antes de las pruebas ICFES**

Caso	B6	B7	B8	B9	QF	L6	L7	L8	L9	L10	M6	M7	M8	M9	M10	S6	S7	S8	S9		P(y=1)
1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	-0,3401	0,416
1	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	1	-1,2368	0,225
3	1	1	1	0	0	1	1	0	0	0	0	1	1	1	1	1	0	1	0	-0,0007	0,5
4	0	0	0	1	1	0	0	0	1	1	0	1	1	1	1	0	0	0	1	-1,7337	0,15
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
6	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	-3,2199	0,038
7	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	-2,6721	0,065
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
11	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	-0,7276	0,326

13	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1	0	0	0	-1,3343	0,208
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	-2,5	0,076
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
16	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	-2,7229	0,062
17	1	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	-1,7323	0,15
18	1	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0	0	1	-0,8671	0,296
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	-2,6821	0,064
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0,2912	0,572
16	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	-0,1495	0,463
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
18	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	-2,9499	0,05
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-2,791	0,058
30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0,2912	0,572
31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0,2912	0,572
31	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	-0,3401	0,416
33	0	0	0	1	1	0	0	0	0	0	1	1	0	1	1	0	0	0	-2,2278	0,097
34	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	-3,3263	0,035
35	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	-2,6161	0,068
36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-0,1372	0,466
37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	-0,0753	0,481
38	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0,2912	0,572
39	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	-1,2838	0,217
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
41	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	-1,3246	0,21
41	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0,2442	0,561
43	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	-0,2135	0,447
44	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0,2912	0,572
45	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	-2,63	0,067
46	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	-1,2368	0,225
47	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	-0,5689	0,361
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061

49	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	-2,8312	0,056
50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
51	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	1	-1,2838	0,217
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
53	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0,1006	0,525
54	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0,1006	0,525
55	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
56	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
57	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,5506	0,634
58	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,1099	0,527
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2,7291	0,061
60	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	-2,9499	0,05
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	-3,1831	0,04
61	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
63	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0,6833	0,664
64	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0,052	0,513
65	0	0	0	1	1	0	0	0	0	0	1	1	0	1	1	0	0	0	0	-2,2278	0,097
66	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	1	-3,7184	0,024
67	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	-3,0082	0,047
68	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,2912	0,572
69	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0,7452	0,678
70	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0,6833	0,664
71	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	-0,4633	0,386
71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-3,5496	0,028
73	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	-1,3246	0,21
74	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0,2442	0,561
75	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	-0,6419	0,345
76	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0,6833	0,664
77	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	-2,63	0,067
78	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	1	1	-0,8084	0,308
79	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	-0,5689	0,361
80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	-3,1212	0,042

## CONCLUSIONES DEL MODELO

El modelo de regresión logística obtenido, esta conformado por todas la variables independientes en estudio ditcotomisadas.

El modelo de regresión logística obtenido, nos indica un ajuste adecuado de las variables de estudio y la variable **IQ**.

Las pruebas de bondad y ajuste que ilustraron el grado de clasificación del modelo, con las pruebas Chi cuadrado, Hosmer -Lemeshow y el estadístico de verosimilitud del modelo indicaron un adecuado ajuste de clasificación de los datos observados.

El estadístico de verosimilitud indica un adecuado ajuste del modelo, sin embargo se observe que la disminución de este valor al introducir las variables independientes es mínimo, pero se evidencia reducción de este valor al comparar este dato con el que arroja el estadístico cuando realiza el cálculo sin introducir las variables independientes.

El modelo seleccionado. tiene un 651% de casos bien clasificados y el 351% corresponde a los casos mal clasificados.

Al realizar el estudio de multicorrelaciones del modelo se observa ausencia de correlaciones entre las variables de estudio.

A pesar de que todas las variables del modelo son significativas, las variables OF. M7, S6, S9, L6 y L8 establecen un aporte *bajo al* cálculo de *la* probabilidad que tiene un estudiante Javeriano de décimo grado, de obtener un puntaje por debajo del nivel muy superior, mientras que M10 y S9 aporta considerablemente a este cálculo.

El pronóstico que realiza el modelo indica que cuanto mejor es el record

académico de los estudiantes la probabilidad de estar por *debajo del* nivel muy superior en el LQ disminuye, no obstante, se evidencian casos de estudiantes que presentan un rendimiento académico adecuado, sin embargo, los resultados en las pruebas de estado en química, muestra una probabilidad alta de estar por debajo del nivel muy superior. También se observan casos donde algunos estudiantes que tiene un desempeño académico bajo, muestra una probabilidad baja de estar por debajo del nivel muy superior. Estos dos sucesos describen la realidad que año tras año envuelve este sistema de evaluación.

#### 4.2.3. Modelo Física

Modelo Inicial:

Se tomó como Variable dependiente IF y como variables independientes: M6, M7, M8, M9, M10, C6, C7, C8, C9, C10, S6, S7, S8, S9, B6, B7, B8, B9, QF.

Para obtener el valor estimado del predictor lineal  $\eta = x_i\beta$  y el valor esperado del modelo de regresión logística se utilizó el método de mínimos cuadrados ponderados generalizados iterados, obteniendo los resultados siguientes:

**Tabla 53. Resultados de las estimaciones del modelo**

Variabes	Estimación $\beta$	Error Estandar	Valor P
B6	-0,3515	0,34087	0,302499
B7	0,00753	0,28196	0,978688
B8	-0,1656	0,35457	0,640408
B9	-0,2379	0,2635	0,366523
QF	-0,0071	0,25569	0,977763
M6	-0,31	0,44226	0,48329

M7	-0,7312	0,50892	0,150764
M8	-1,3456	0,43307	0,001889
M9	-0,7814	0,43503	0,072447
M10	0,1705	0,29139	0,558458
S6	-0,3389	0,47967	0,47982
S7	0,22013	0,42392	0,60357
S8	1,12206	0,48636	0,021051
S9	0,88512	0,45473	0,0516
L6	0,84385	0,41925	0,04414
L7	-0,0218	0,37064	0,953124
L8	-0,3091	0,40886	0,4496
L9	-0,4884	0,42201	0,247142
L10	-0,4351	0,40208	0,27922
Constante	8,25671	2,18313	0,000156

Los datos obtenidos indican el ajuste del modelo que describe la relación entre IF y el rendimiento académico conformado por las 19 variables antes mencionadas.

### Modelo de regresión

$$IF = \frac{e^v}{1 + e^v}$$

Donde  $v$  es  $x'\beta$

$$v = 8,25671 - 0,351466*B6 + 0,0075323*B7 - 0,165629*B8 - 0,237942*B9 - 0,435079*L10 + 0,843846*L6 - 0,021788*L7 - 0,309128*L8 - 0,488397*L9 + 0,170503*M10 - 0,310031*M6 - 0,731229*M7 - 1,34561*M8 - 0,781445*M9 - 0,00712684*QF - 0,338929*S6 + 0,22013*S7 + 1,12206*S8 + 0,88512*S9$$

Se realizó un análisis de varianza para estudiar e ajuste del modelo, del cual se obtuvieron los resultados que se presentan en la tabla 54.

**Tabla 54. Análisis de varianza para la regresión**

Fuente	Desviación	G.I.	Valor P
Modelo	63,4988	19	0,0000
Residuos	413,096	324	0,0006
Total	476,595	343	

Como el valor p para el modelo en la tabla anterior es inferior a 0.01, se puede concluir que hay una relación estadísticamente significativa entre las variables de estudio.

Se observó el estadístico de Cox y Snell para cuantificar la bondad del ajuste, en este caso es igual a 16,8%, este valor puede suponer que el modelo explica de manera no tan adecuada la variabilidad del mismo. Sin embargo el porcentaje ajustado de Nagelkerke, el cual, es más adecuado para comparar modelos con diferentes números de variables independientes es 22,5%, al parecer mejora el ajuste, no obstante, los resultados no son satisfactorios, por lo cual se trata de mejorar el modelo identificando aquellas variables que no aportan significativamente al modelo. Aquí se aplicó el sistema paso a paso hacia atrás obteniendo los siguientes resultados:

Modelo 2

**Tabla 55. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
M7	-0,92097	0,471099	0,05059
M8	-1,40366	0,415683	0,000733
M9	-0,75995	0,414584	0,066795
S6	-0,52885	0,431754	0,220614
S8	1,08562	0,461594	0,018678
S9	0,75665	0,43093	0,079112
L6	0,63584	0,374083	0,08918

L9	-0,52895	0,410404	0,197449
L10	-0,4998	0,391444	0,20167
Constante	7,75577	1,920908	5,4E-05

El modelo resultante es:

$$v = 7,75577 - 0,499798*L10 + 0,635844*L6 + 0,75665*S9 + 1,08562*S8 - 0,52885*S6 - 0,75995*M9 - 1,40366*M8 - 0,92097*M7 - 0,528951*L9$$

El modelo es significativo según el análisis de varianza, sin embargo el ajuste baja con respecto al anterior modelo, puesto que los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 16,1% y 21,5%.

Posteriormente se buscaron otros modelos de regresión logística construyendo una base donde las variables independientes tomaron los valores 1 y 0 (modelo 3), 0 si el estudiante tiene sobresaliente o excelente y 1 si tiene insuficiente o aprobado.

Modelo 3

**Tabla 56. Resultados de las estimaciones del modelo**

Variabes	Estimación $\beta$	Error Estandar	Valor P
B6	0,4698	0,36745	0,20104
B7	0,2865	0,35937	0,42536
B8	0,1528	0,38699	0,69299
B9	0,2006	0,33704	0,55172
QF	0,258	0,43905	0,55681
M6	0,2168	0,29574	0,46352
M7	0,3519	0,3673	0,338

M8	0,9206	0,35594	0,0097
M9	0,6922	0,41222	0,09312
M10	-0,4393	0,66818	0,51092
S6	0,3445	0,39272	0,38044
S7	-0,5052	0,39569	0,20169
S8	-0,4508	0,41091	0,27256
S9	-0,6587	0,45559	0,14824
L6	-0,6566	0,30698	0,03243
L7	-0,017	0,29963	0,95467
L8	0,576	0,33894	0,08925
L9	0,2911	0,43664	0,505
L10	-0,3349	0,54581	0,53954
Constante	-1,2415	0,66345	0,0613

#### Modelo

$$v = -1,24154 + 0,469822*B6 + 0,286473*B7 + 0,152784*B8 + 0,200603*B9 - 0,334864*L10 - 0,656631*L6 - 0,0170316*L7 + 0,575983*L8 + 0,291084*L9 - 0,439267*M10 + 0,216795*M6 + 0,351919*M7 + 0,920632*M8 + 0,692185*M9 + 0,25798*QF + 0,344453*S6 - 0,505202*S7 - 0,450846*S8 - 0,658674*S9$$

En este modelo, el análisis mostró mejores resultados que el anterior pero casi similar al modelo 1, los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 13,2% y 17,7%.

Al igual que en el modelo inicial, se reducen las variables utilizando el método paso a paso hacia atrás, obteniendo el modelo:

$$v = -1,37276 + 0,519713*Bio6 - 0,568554*L6 + 0,321*M7 + 0,918020*M8 + 0,378690*Soc6 - 0,485473*Soc7 - 0,568232*Soc9 + 0,538165*L8 + 0,68662*M9 +$$

Sin embargo, según el análisis de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente los del 12,6 % y 16,4%. Lo cual indica poco ajuste e inferior al del modelo inicial.

Finalmente se realizó el estudio con una base donde las variables independientes son los promedios en cada asignatura, se hizo el análisis con un modelo completo (modelo 5) en el cual los estadísticos de Cox y Snell y Nagelkerke, son respectivamente 10% y 13,4%, posteriormente se eliminaron las variables menos significativas con el método paso a paso hacia atrás (modelo 6); en el cual los estadísticos de Cox y Snell y Nagelkerke, son respectivamente 9,4% y 12,5%.

Del estudio anterior se puede deducir que el ajuste del modelo que describe la relación entre IF y el rendimiento académico es bajo y el modelo que mejor explica esta relación es el modelo 1, modelo al cual se le realizan otras pruebas para analizar la habilidad del modelo para efectos de clasificación.

### Test de Bondad de ajuste Chi-cuadrado

Se aplica la Prueba de Bondad de Ajuste Chi Cuadrado para evaluar la habilidad de clasificación del modelo.

**Tabla 57. Bondad de ajuste**

Clase	Intervalo Logia	n	VERDADERO		FALSO	
			Observado	Esperado	Observado	Esperado
1	menor que -0,664287	69	11,0	14,0146	58,0	54,9854
2	-0,664287 a 0,0392032	69	34,0	29,9957	35,0	39,0043
3	0,0392032 a 0,295594	73	36,0	40,1415	37,0	32,8585
4	0,295594 a 0,650992	65	41,0	39,6878	24,0	25,3122
5	0,650992 o mayor	68	55,0	53,1604	13,0	14,8396
Total		344	177,0		167,0	

Chi-cuadrado = 3,11177 con 3 g.l. Valor P = 0,374707

Dado que el Valor P es superior a 0.10, no hay razón para descartar al modelo como medio de clasificación.

Al realizar el anterior estudio con los otros modelos, se obtuvieron los resultados que se indican en la tabla 58.

**Tabla 58. Valor del Chi-cuadrado en cada modelo presente en el estudio**

MODELO	Chi-cuadrado	Valor P
1	3,11177	0,374707
2	0,564655	0,754027
3	2,1437	0,543118
4	1,142	0,76694
5	4,15206	0,245502
6	4,63482	0,200572

Como se puede observar todos los modelos presentan una buena adecuación.

**Tabla 59. Tabla de contingencias para la prueba de Hosmer y Lemeshow**

	IF = ,00		IF = 1,00		Total
	Observado	Esperado	Observado	Esperado	
1	30	30,008	4	3,992	34
2	27	24,317	7	9,683	34
3	14	20,970	20	13,030	34
4	20	17,714	14	16,286	34
5	19	16,171	15	17,829	34
6	20	17,668	21	23,332	41
7	11	13,927	23	20,073	34
8	13	12,398	21	21,602	34
9	9	9,966	25	24,034	34
10	4	3,860	27	27,140	31

Chi-cuadrado = 10,411 con 3 g.l. Valor P = 0,237

Dado que el Valor P de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo decir el modelo se ajusta adecuadamente a los datos.

### **Doble logaritmo del estadístico de Verosimilitud**

Al compara el valor del  $-2\log$  verosimilitud cuando el modelo no tiene variables independientes con el valor del  $-2\log$  verosimilitud cuando se introducen estas variables al modelo se obtuvo lo siguiente.

$-2\log$  verosimilitud sin variables independientes: 476,5945216

$-2\log$  verosimilitud con variables independientes: 413,0957045

Como el valor del  $-2\log$  verosimilitud se redujo, el ajuste es bueno, más no ideal porque esta reducción no tiende a cero.

Al aplicar este estadístico en los otros modelos observamos los resultados que se indican en la tabla 60.

**Tabla 60. Valor del  $-2\log$  verosimilitud en cada modelo presente en el estudio**

MODELO	$-2\log$ verosimilitud. (A)	$-2\log$ verosimilitud. (D)
1	476,5945216	413,0957045
2		416,1901360453
3		427,5608007
4		431,5619599494
5		440,0309214
6		442,6114514585

Los resultados arrojados por el software en cada modelo indican un buen ajuste al modelo logístico, el modelo 1 es el que presenta mayor disminución, sin embargo, las reducciones son menores en cada modelo.

**Tabla 61. Tabla de clasificación para IF.**

Observado			Pronosticado		
			IF		Porcentaje correcto
			0,00	1,00	
Paso 1	IF	0,00	86	81	51,5
		1,00	37	140	79,1
Porcentaje global					65,7

La ecuación del modelo ya diseñada nos proporciona una probabilidad que nos permite predecir a partir de ella para cada sujeto un valor de Y (Y predicho), tal que si  $P(Y=1|X) \leq 0.5$  entonces  $Y_{pred} = 1$ , y si  $P(Y=1|X) > 0.5$  entonces  $Y_{pred} = 0$

Al observar la lista de clasificación del modelo1 se contabilizan 118 casos mal clasificados de 344 casos, correspondientes al 34,3%.

**Tabla 62. Listado de casos Mal Clasificados**

Caso	Observado	Pronosticado IF	Grupo pronosticado
1	0	0,72805758	1
2	0	0,51968861	1
4	0	0,50846371	1
9	0	0,77000199	1
13	1	0,34074466	0
21	0	0,62703796	1
23	0	0,50947782	1

24	1	0,39542064	0
27	1	0,46911191	0
29	0	0,58617711	1
30	0	0,50748402	1
36	1	0,35091772	0
38	1	0,33532413	0
42	0	0,62021353	1
49	1	0,46234838	0
50	0	0,6572339	1
52	1	0,49452095	0
54	0	0,75566695	1
55	0	0,61916138	1
58	1	0,36289126	0
59	0	0,50979954	1
63	1	0,4117648	0
65	1	0,46303111	0
66	1	0,0950075	0
68	0	0,6273054	1
71	0	0,65027403	1
74	0	0,52726861	1
76	0	0,58977596	1
77	0	0,62198301	1
80	1	0,41451714	0
84	0	0,59109951	1
85	1	0,40372817	0
86	1	0,40509097	0
91	0	0,57336506	1
95	1	0,25778481	0
97	0	0,57336506	1
99	1	0,21888742	0
103	0	0,51150326	1
104	0	0,59597475	1
108	0	0,58173866	1
110	1	0,39894866	0
113	1	0,17239992	0

120	0	0,57336506	1
121	0	0,53514876	1
122	0	0,59616348	1
124	0	0,56789873	1
125	0	0,60157419	1
127	1	0,39067146	0
131	0	0,57336506	1
133	0	0,61086198	1
134	1	0,24305163	0
136	0	0,57336506	1
137	0	0,50665015	1
138	0	0,57597801	1
141	1	0,16835457	0
142	0	0,68129066	1
145	0	0,51262983	1
146	0	0,57304871	1
149	0	0,61949214	1
151	0	0,54188104	1
157	0	0,53030755	1
163	0	0,73473963	1
169	0	0,5955844	1
172	0	0,53122851	1
175	0	0,7183373	1
176	0	0,52533298	1
180	0	0,50796982	1
183	1	0,4154131	0
186	0	0,55059878	1
187	0	0,62486894	1
188	0	0,50451718	1
190	1	0,36285634	0
192	0	0,51311442	1
195	0	0,57336506	1
197	1	0,40191449	0
201	0	0,60332808	1
203	1	0,33358918	0

204	0	0,51467716	1
206	1	0,35625188	0
213	0	0,57336506	1
214	0	0,53135215	1
215	1	0,26475553	0
216	0	0,70170077	1
217	0	0,57304871	1
218	0	0,76849957	1
221	0	0,50995863	1
225	1	0,25872794	0
228	0	0,56750678	1
230	0	0,6260848	1
236	0	0,62639101	1
237	0	0,57008506	1
238	0	0,74767713	1
241	0	0,82297668	1
244	1	0,35121431	0
247	0	0,57336506	1
248	0	0,74575092	1
252	1	0,17940978	0
253	1	0,38782261	0
256	0	0,57043117	1
258	1	0,39313765	0
267	1	0,42610692	0
268	1	0,48691341	0
271	0	0,57336506	1
278	0	0,53122851	1
285	0	0,53122851	1
288	0	0,57336506	1
303	1	0,46591391	0
308	0	0,53122851	1
311	0	0,57662456	1
312	1	0,42371524	0
315	0	0,570599	1
321	0	0,90333233	1

322	0	0,51053434	1
323	0	0,57336506	1
335	0	0,53523815	1
337	0	0,64317983	1
339	1	0,35163878	0
342	0	0,88016322	1

Al realizar este mismo estudio para cada modelo encontramos los resultados de la tabla 63

**Tabla 63. Total de casos mal clasificados en cada modelo presente en el estudio**

MODELO	CASOS MAL CLASIFICADOS
1	118
2	123
3	121
4	123
5	132
6	126

### Residuos atípicos

**Tabla 64. Residuos atípicos para IF**

Fila	Y	Predicho Y	Residuo	Residuo Pearson	Desviación Residuo
66	1	0,0950075	0,904992	3,09	2,17
113	1	0,1724	0,8276	2,19	1,88
141	1	0,168355	0,831645	2,22	1,89
241	0	0,17941	-0,8229	-2,16	-1,86
252	1	0,142319	0,82059	2,14	1,85
321	0	0,903332	-0,903332	-3,06	-2,16
342	0	0,880163	-0,880163	-2,71	-2,06

La tabla de residuos atípicos lista todas las observaciones que tienen residuo Pearson o desviación de residuo superiores a 2.0 en valor absoluto. Estos residuos estandarizados miden cuántas desviaciones estándar de cada valor observado de IF proceden del modelo ajustado. En este caso, hay 7 residuos estandarizados superiores a 2.0, y 2 superiores a 3.0.

Al estudiar con detalle estos individuos se observa que aquellos que están por ENCIMA del nivel muy superior (0), su nivel académico es BAJO; y aquellos que están por DEBAJO del nivel muy superior (1), presentan un rendimiento académico BUENO en su vida escolar. Por lo tanto, es conveniente mantenerlos en el estudio puesto que son casos reales que no presentan errores al levantamiento ni digitación de la información ya que se confrontó la base de datos con la información de los libros de registro.

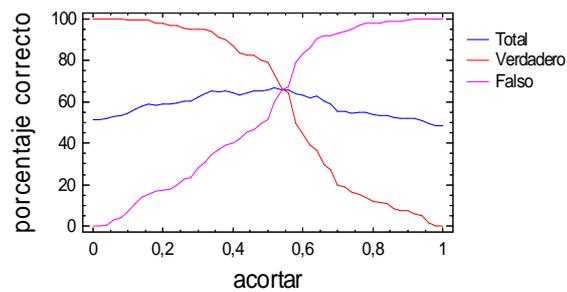
## PREDICCIÓN DEL MODELO

**Tabla 65. Capacidad de Predicción del Modelo**

Cohorte	VERDADERO	FALSO	Total
0,0	100,00	0,00	51,45
0,05	100,00	1,80	52,33
0,1	99,44	7,19	54,65
0,15	99,44	13,77	57,85
0,2	97,74	17,37	58,72
0,25	96,61	20,36	59,59
0,3	94,92	28,14	62,50
0,35	93,22	37,13	65,99
0,4	87,01	40,12	64,24
0,45	82,49	46,11	64,83
0,5	79,10	51,50	65,70
0,55	64,97	66,47	65,70

0,6	44,63	83,23	63,37
0,65	32,77	91,02	61,05
0,7	19,77	92,81	55,23
0,75	16,38	96,41	55,23
0,8	11,86	98,20	53,78
0,85	9,60	98,80	52,91
0,9	7,34	99,40	52,03
0,95	2,82	100,00	50,00
1,0	0,00	100,00	48,55

**Figura 4. Capacidad de Predicción del Modelo para IF**



En la tabla y el gráfico se observa la capacidad de predicción del modelo 1. Si el valor predicho es más grande que el punto de cohorte, la respuesta se predice como VERDADERA. Si el valor predicho es inferior a o igual al punto de cohorte, la respuesta se predice para ser FALSO. La tabla muestra el porcentaje de datos observados predichos correctamente a diferentes puntos de cohorte. En nuestro caso con el punto de cohorte utilizado 0,5, el 79,10% de todas las respuestas VERDADERAS se predijeron correctamente, mientras que el 51,50% de todas las respuestas FALSAS se predijeron correctamente, para un total de 65,70% casos clasificados correctamente.

**Tabla 66. Matriz de correlación para los coeficientes estimados**

	Cons	B6	B7	B8	B9	QF	M6	M7	M8	M9	M10	S6	S7	S8	S9	L6	L7	L8	L9	L10
Con	1																			
B6	-0,02	1																		
B7	0,14	0,03	1																	
B8	-0,25	-0,07	-0,06	1																
B9	-0,19	0,11	0,01	0,12	1															
QF	0,09	-0,04	-0,01	0,04	-0,19	1														
M6	-0,26	-0,04	-0,09	0,04	0,15	0,01	1													
M7	-0,22	-0,15	-0,06	0,00	-0,03	0,00	-0,13	1												
M8	-0,27	-0,03	-0,11	0,03	0,02	-0,13	0,01	-0,10	1											
M9	-0,12	0,01	-0,02	-0,18	0,11	-0,06	0,12	-0,15	-0,19	1										
M10	-0,16	0,03	-0,12	0,00	-0,01	-0,23	-0,12	0,01	0,02	-0,04	1									
S6	0,02	-0,18	-0,01	-0,05	-0,07	-0,03	-0,19	0,05	-0,06	-0,02	-0,04	1								
S7	0,06	-0,20	-0,13	-0,03	0,00	0,06	-0,11	-0,12	0,00	-0,03	0,03	-0,17	1							
S8	-0,01	0,12	0,06	-0,15	-0,13	-0,04	-0,05	-0,15	-0,12	-0,14	0,02	-0,05	-0,15	1						
S9	-0,12	-0,04	-0,16	-0,06	-0,16	-0,08	-0,07	0,01	0,07	-0,04	0,01	-0,07	0,04	-0,13	1					
L6	0,00	-0,22	-0,08	0,00	-0,03	-0,10	-0,19	0,04	0,07	-0,10	0,09	-0,21	-0,02	-0,07	0,08	1				
L7	-0,07	-0,09	-0,09	-0,05	-0,02	-0,04	-0,09	0,02	0,02	0,01	0,02	-0,08	-0,15	0,08	-0,04	-0,16	1			
L8	0,15	0,11	0,06	-0,03	-0,01	-0,01	0,11	-0,24	-0,13	-0,03	-0,11	-0,13	-0,03	-0,10	-0,12	-0,13	-0,20	1		
L9	-0,01	-0,06	-0,05	0,03	-0,15	0,04	0,03	0,05	-0,04	-0,10	-0,12	0,00	-0,05	-0,15	-0,10	-0,02	-0,03	-0,04	1	
L10	-0,18	-0,01	-0,09	0,01	-0,09	-0,04	-0,03	0,04	0,00	0,01	0,00	0,06	-0,02	0,03	-0,16	-0,05	0,00	-0,10	-0,31	1

Esta tabla muestra las correlaciones estimadas entre los coeficientes en el modelo ajustado. Estas correlaciones pueden utilizarse para detectar la presencia de serias multicorrelaciones, es decir, correlación entre las variables pronosticadas. En este caso, no hay ninguna correlación con valores absolutos superiores a 0.5.

## INTERPRETACION DEL MODELO

Como se puede observar en el estudio anterior, el modelo 1 presenta un adecuado ajuste, las variables que ayudan a predecir si un estudiante de grado décimo del colegio Javeriano estará por encima o debajo del nivel muy superior en las pruebas de estado ICFES en la asignatura de Física son todas las materias de matemáticas, castellano, sociales y naturales, que cursó desde grado sexto a grado décimo en la institución.

A continuación se indican el valor de las probabilidades que tiene cada variable de estudio en el modelo.

**Tabla 67. Variables en el modelo**

	B	Exp(B)
B6	-0,351	0,704
B7	0,008	1,008
B8	-0,166	0,847
B9	-0,238	0,788
QF	-0,007	0,993
M6	-0,310	0,733
M7	-0,731	0,481
M8	-1,346	0,260
M9	-0,781	0,458
M10	0,171	1,186

	B	Exp(B)
S6	-0,339	0,713
S7	0,220	1,246
S8	1,122	3,071
S9	0,885	2,423
L6	0,844	2,325
L7	-0,022	0,978
L8	-0,309	0,734
L9	-0,488	0,614
L10	-0,435	0,647
Constante	8,257	3853,408

De lo anterior se puede decir que el valor de  $-0,351$  que se asocia a la variable B6, incrementa el 70,4 posibilidad de estar por debajo del nivel muy superior en el área de QUIMICA, puesto que  $e^{-0,351} = 0,704$

De igual manera se pueden leer todos los coeficientes que acompañan a cada variable de estudio.

Como se puede observar, las variables M10, S7, S9 y L10 establecen un aporte alto al cálculo de esta probabilidad, mientras que M8, M9, M7 aporta considerablemente poco a este cálculo.

El modelo seleccionado (modelo 1 ayudará a detectar a aquellos estudiantes de grado décimo que requieren de un refuerzo académico antes de presentar la prueba ICFES. A continuación se indica la probabilidad de que cada estudiante de grado décimo pueda estar por debajo del nivel muy superior.

**Tabla 68. Casos posibles que necesitan refuerzo académico en física antes de las pruebas ICFES**

Caso	B6	B7	B8	B9	QF	L6	L7	L8	L9	L10	M6	M7	M8	M9	M10	S6	S7	S8	S9		P(y=1)
1	4,3	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	3,5	3,5	-0,058	0,4856
2	4,6	5	4,3	3,5	4	4,6	5	4	3,5	3,5	5	4	3,5	3,5	3,5	5	4,6	4,3	3,5	0,2827	0,5702
3	3,5	4	3,5	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	3,5	3,5	4	4,6	3,5	4,3	-0,184	0,454
4	4,6	4	4,3	3,5	4	4,6	4	4	3,5	3,5	4	4	3,5	3,5	3,5	5	4,6	4,3	3,5	0,9649	0,7241
5	3,5	4	2	3,5	4	3,5	4	4	3,5	3,5	4	2	3,5	2	3,5	2	4,3	3,5	3,5	3,4975	0,9706
6	4,6	4	4,3	4,3	4	4,3	4	4	4,3	4,3	5	4	3,5	4,3	4,3	4	4,6	4,3	4,3	0,2523	0,5627
7	4,3	4	4,3	4,3	4	3,5	4	4	4,3	4,3	4	4	4,3	4,3	4,3	4	4,3	4,3	4,3	-1,332	0,2089
8	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	3,5	3,5	0,4717	0,6158
9	4,6	5	4,6	4,6	5	4,6	5	5	4,6	4,6	5	5	4,6	4,6	4,6	5	4,6	4,6	4,6	-2,206	0,0992
10	3,5	2	3,5	3,5	4	3,5	2	4	3,5	3,5	4	4	3,5	2	3,5	4	4,3	3,5	3,5	1,6652	0,8409
11	4,6	5	4,3	4,6	5	4,6	5	5	4,6	4,6	5	5	4,6	4,6	4,6	5	4,6	4,6	4,6	-2,157	0,1037
12	4,3	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	4,3	3,5	0,5928	0,644

13	4,3	4	3,5	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	3,5	3,5	4	4,3	4,3	4,3	0,8671	0,7041
14	4,6	4	4,3	4,3	4	4,3	4	4	4,3	4,3	5	4	4,3	3,5	4,3	4	4,3	4,3	4,3	-1,097	0,2502
15	4,6	4	4,6	4,3	4	4,6	4	5	4,3	4,3	5	5	4,3	4,3	4,3	5	4,6	4,6	4,3	-1,53	0,178
16	4,3	4	4,3	4,6	5	4,3	4	4	4,6	4,6	5	4	3,5	4,6	4,6	4	4,3	4,3	4,6	-0,562	0,3632
17	3,5	4	3,5	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	4,3	4,3	4	4,3	4,3	4,3	0,1585	0,5395
18	3,5	4	3,5	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	3,5	3,5	4	4,3	3,5	4,3	-0,262	0,4349
19	4,6	4	4,3	4,3	4	4,3	4	4	4,3	4,3	4	5	4,3	4,6	4,3	4	4,6	4,3	4,3	-2,017	0,1174
20	4,3	4	4,3	4,3	4	4,6	4	4	4,3	4,3	4	4	3,5	4,3	4,3	4	4,6	4,3	4,3	-0,128	0,4679
21	4,6	4	4,6	4,6	5	4,6	5	5	4,6	4,6	5	5	4,6	4,6	4,6	5	4,6	4,6	4,6	-2,209	0,099
22	4,6	4	4,3	4,3	4	4,3	4	4	4,3	4,3	4	4	4,3	4,3	4,3	4	4,6	4,3	4,3	-1,564	0,1731
23	4,6	5	4,6	4,6	5	4,6	5	5	4,6	4,6	5	5	4,6	4,6	4,6	5	4,6	4,3	4,6	-2,543	0,0729
24	4,6	4	4,3	4,3	4	4,6	4	4	4,3	4,3	5	4	4,3	4,3	4,3	5	4,6	4,3	4,3	-1,505	0,1817
25	3,5	4	2	3,5	2	3,5	4	4	3,5	3,5	4	4	2	2	3,5	4	4,6	3,5	3,5	3,9874	0,9818
26	4,3	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	2	3,5	3,5	3,5	4	4,3	3,5	3,5	1,2874	0,7837
27	4,6	5	4,6	4,6	5	4,6	5	5	4,6	4,6	5	5	4,6	4,6	4,6	4	4,6	4,6	4,6	-2,105	0,1086
28	4,6	4	4,3	4,3	4	4,3	4	4	4,3	4,3	4	4	3,5	3,5	4,3	5	4,6	4,3	4,3	0,6214	0,6505
29	4,3	4	4,3	4,3	4	4,3	4	5	4,3	4,3	4	5	4,3	4,3	4,3	4	4,6	4,3	4,3	-1,499	0,1826
30	3,5	4	2	3,5	2	3,5	4	4	3,5	3,5	4	4	2	2	3,5	4	4,6	3,5	3,5	3,9874	0,9818
31	3,5	2	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	3,5	3,5	0,4604	0,6131
32	4,3	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	3,5	3,5	-0,058	0,4856
33	4,6	4	4,3	3,5	4	4,3	4	4	4,3	4,3	4	4	4,3	3,5	3,5	4	4,3	4,3	4,3	-0,112	0,4721
34	4,6	4	4,3	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	4,3	4,3	4	4,3	4,3	4,3	-0,372	0,4081
35	4,3	4	4,3	4,3	4	3,5	4	4	4,3	4,3	4	4	4,3	4,3	4,3	4	4,3	4,3	4,3	-1,372	0,2023
36	3,5	2	2	3,5	4	3,5	2	4	3,5	3,5	4	4	2	3,5	2	2	4,3	4,3	3,5	3,9102	0,9804
37	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	4,3	3,5	1,0982	0,7499
38	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	3,5	3,5	0,4717	0,6158
39	4,6	4	4,3	3,5	4	4,6	5	4	3,5	3,5	5	4	4,3	3,5	3,5	4	4,6	4,3	3,5	-0,694	0,3331
40	4,3	4	4,3	4,3	4	4,3	4	4	4,3	4,3	4	4	4,6	4,3	4,3	4	4,6	4,6	4,3	-1,525	0,1787
41	4,6	4	4,3	3,5	4	4,6	5	4	3,5	3,5	5	4	4,3	3,5	3,5	5	4,3	4,3	3,5	-0,615	0,351
42	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	4,3	3,5	3,5	4	4,3	3,5	3,5	-0,605	0,3532
43	3,5	4	3,5	3,5	2	3,5	4	4	3,5	3,5	4	2	2	3,5	2	4	4,3	3,5	3,5	3,3479	0,966
44	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	3,5	3,5	0,4717	0,6158
45	3,5	4	4,3	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	3,5	3,5	4	4,6	4,3	3,5	0,1325	0,5331
46	4,3	4	4,3	3,5	4	4,3	4	4	3,5	3,5	5	4	3,5	3,5	3,5	4	4,3	4,3	3,5	0,1749	0,5436
47	4,6	4	3,5	3,5	4	4,3	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,6	3,5	3,5	0,8322	0,6968
48	4,6	4	4,3	4,3	4	4,6	4	4	4,3	4,3	5	4	4,6	4,3	4,3	5	4,6	4,3	4,3	-1,909	0,1291
49	4,3	4	4,3	3,5	4	4,3	4	4	4,3	4,3	4	4	3,5	3,5	4,3	4	4,6	4,3	4,3	1,0189	0,7348
50	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	2	3,5	4	4,6	3,5	3,5	1,7099	0,8468

51	4,6	4	4,3	3,5	4	4,3	4	4	3,5	3,5	4	4	4,3	3,5	3,5	4	4,6	4,3	3,5	-0,848	0,2998
52	4,6	5	4,3	4,3	4	4,6	5	5	4,3	4,3	5	4	4,6	4,3	4,3	5	4,3	4,3	4,3	-2,072	0,1119
53	3,5	4	3,5	2	4	3,5	4	4	3,5	3,5	4	4	3,5	2	3,5	4	4,6	3,5	3,5	1,8188	0,8604
54	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	2	3,5	3,5	3,5	4	4,3	3,5	3,5	1,3205	0,7893
55	3,5	4	3,5	3,5	4	3,5	2	4	3,5	3,5	4	2	3,5	3,5	3,5	4	4,3	3,5	3,5	1,6012	0,8322
56	3,5	4	3,5	2	2	3,5	4	4	3,5	3,5	4	4	2	2	3,5	4	4,3	3,5	3,5	4,0299	0,9825
57	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	3,5	3,5	3,5	0,2956	0,5734
58	4,3	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	2	3,5	3,5	3,5	4	3,5	3,5	3,5	1,1113	0,7524
59	4,6	5	4,6	4,6	5	4,6	5	5	4,6	4,6	5	5	4,6	4,6	4,6	4	4,6	4,3	4,6	-2,441	0,0801
60	4,6	4	4,3	4,3	4	4,3	4	4	4,3	4,3	4	4	3,5	3,5	4,3	5	4,6	4,3	4,3	0,6214	0,6505
61	4,3	4	4,3	4,3	4	4,3	4	5	4,3	4,3	4	5	4,3	4,3	4,3	4	4,6	3,5	3,5	-3,105	0,0429
62	3,5	4	2	3,5	2	3,5	4	4	3,5	3,5	4	4	2	2	3,5	4	4,3	3,5	3,5	3,9214	0,9806
63	3,5	2	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,6	4,6	4,6	2,7343	0,939
64	4,3	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,6	4,3	4,3	1,6143	0,834
65	4,6	4	4,3	3,5	4	4,3	4	4	4,3	4,3	4	4	4,3	3,5	3,5	4	4,6	4,3	4,3	-0,046	0,4886
66	4,6	4	4,3	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	4,3	4,3	4	4,6	3,5	3,5	-1,912	0,1288
67	4,3	4	4,3	4,3	4	3,5	4	4	4,3	4,3	4	4	4,3	4,3	4,3	4	4,3	3,5	3,5	-2,978	0,0484
68	3,5	2	2	3,5	4	3,5	2	4	3,5	3,5	4	4	2	3,5	2	2	4,3	3,5	3,5	3,0126	0,9531
69	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	4,3	4,3	1,8063	0,8589
70	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	4,3	4,3	2,0774	0,8887
71	4,6	4	4,3	3,5	4	4,6	5	4	3,5	3,5	5	4	4,3	3,5	3,5	4	4,3	4,3	4,3	-0,052	0,4869
72	4,3	4	4,3	4,3	4	4,3	4	4	4,3	4,3	4	4	4,6	4,3	4,3	4	4,3	4,3	3,5	-2,636	0,0669
73	4,6	4	4,3	3,5	4	4,6	5	4	3,5	3,5	5	4	4,3	3,5	3,5	5	4,3	4,3	3,5	-0,615	0,351
74	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	4,3	3,5	3,5	4	4,3	3,5	3,5	-0,605	0,3532
75	3,5	4	3,5	3,5	2	3,5	4	4	3,5	3,5	4	2	2	3,5	2	4	4,6	4,3	3,5	4,3116	0,9868
76	3,5	4	3,5	3,5	4	3,5	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,6	4,6	4,3	2,4801	0,9227
77	3,5	4	4,3	4,3	4	3,5	4	4	4,3	4,3	4	4	3,5	3,5	3,5	4	4,3	4,3	3,5	0,0664	0,5166
78	4,3	4	4,3	3,5	4	4,3	4	4	3,5	3,5	5	4	3,5	3,5	3,5	4	4,3	3,5	3,5	-0,723	0,3268
79	4,6	4	3,5	3,5	4	4,3	4	4	3,5	3,5	4	4	3,5	3,5	3,5	4	4,3	3,5	3,5	0,7662	0,6827
80	4,6	4	4,3	4,3	4	4,6	4	4	4,3	4,3	5	4	4,6	4,3	4,3	5	4,3	3,5	3,5	-3,581	0,0271

## CONCLUSIONES DEL MODELO

El modelo de regresión logística escogido esta conformado por todas las variables independientes del estudio.

El modelo de regresión logística obtenido, nos indica un ajuste adecuado de las variables de estudio y la variable IF.

Las pruebas de bondad *y* ajuste que ilustraron el grado de clasificación del modelo, con las pruebas Chi cuadrado, Hosmer-Lemeshow y el estadístico de verosimilitud del modelo indicaron un adecuado ajuste de clasificación de los datos observados.

El estadístico de verosimilitud indica un adecuado ajuste del modelo, sin embargo se observa que la de este *valor* al introducir las variables independientes es mínimo, pero se evidencia reducción de este valor al comparar este dato con el que arroja el estadístico cuando realiza el cálculo sin las introducir las variables independientes.

El modelo seleccionado, tiene un 65.7% de casos bien clasificados y el 34.3% corresponde a los casos mal clasificados.

Al realizar el estudio de multicorrelaciones del modelo se observa ausencia de correlaciones entre las variables de estudio.

A pesar de que todas las variables del modelo son significativas, las variables M10, S7, S9 y L **10** establecen *un* aporte *alto al* cálculo *de* la probabilidad que tiene un estudiante Javeriano de décimo grado, de obtener un puntaje por debajo del nivel muy superior.

El pronóstico que realiza el modelo indica que cuanto mejor es el record académico de los estudiantes la probabilidad de estar por debajo del nivel muy superior en el IF disminuye; no obstante, se evidencian casos de estudiantes que presentan un rendimiento académico adecuado, sin embargo, los resultados en las pruebas de estado en matemáticas, muestra una probabilidad alta de estar por

debajo del nivel muy superior. También se observan casos donde algunos estudiantes que tiene un desempeño académico bajo, muestra una probabilidad baja de estar por debajo del nivel muy superior. Estos dos sucesos describen la realidad que año tras año envuelve este sistema de evaluación.

### 4.3. Modelo para el área de Sociales

#### Modelo 1

Se tomó como Variable dependiente IS Matemáticas y como variables independientes: M6, M7, M8, M9, M10, C6, C7, C8, C9, C10, S6, S7, S8, S9, B6, B7, B8, B9, QF.

Para obtener el valor estimado del predictor lineal  $\eta = x\beta$  y el valor esperado del modelo de regresión logística se utilizó el método de mínimos cuadrados ponderados generalizados iterados, obteniendo los resultados siguientes:

**Tabla 69. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	-0,29730618	0,33987672	0,38171151
B7	-0,99299346	0,28226768	0,00043495
B8	0,64535096	0,35083259	0,06584402
B9	0,25862111	0,25049527	0,30186647
QF	0,37198552	0,24704706	0,13213718
M6	-0,47140391	0,43692631	0,28062811
M7	0,76247437	0,48900765	0,11894245
M8	0,85931768	0,35126937	0,01443224
M9	-0,27975123	0,39085463	0,47415043

M10	-0,16041517	0,29238048	0,58324421
S6	0,16503938	0,46186044	0,72084025
S7	0,44444307	0,42330734	0,2937503
S8	-0,69051456	0,45474965	0,12890102
S9	-0,45497471	0,4389458	0,2999611
L6	-0,30342614	0,40411805	0,45275172
L7	0,22884237	0,36236438	0,52769693
L8	-0,60116063	0,38527603	0,11868015
L9	-0,54405452	0,37841678	0,15051565
L10	0,0106945	0,37310381	0,97713289
Constante	4,23145017	1,92044186	0,02756845

Los datos obtenidos indican el ajuste del modelo que describe la relación entre IS y el rendimiento académico conformado por las 19 variables antes mencionadas.

### Modelo de regresión

$$IS = \frac{e^v}{1 + e^v}$$

Donde  $v$  es  $x'\beta$

$$v = 4,23145 + 0,0106945*L10 - 0,303426*L6 + 0,228842*L7 - 0,601161*L8 - 0,544055*L9 - 0,160415*M10 - 0,471404*M6 + 0,762474*M7 + 0,859318*M8 - 0,279751*M9 + 0,371986*QF + 0,165039*S6 + 0,444443*S7 - 0,690515*S8 - 0,454975*S9 - 0,297306*B6 - 0,992993*B7 + 0,645351*B8 + 0,258621*B9$$

Se realizó un análisis de varianza para estudiar el ajuste del modelo, del cual se obtuvieron los resultados que se presentan en la tabla 70.

**Tabla 70. Análisis de varianza para la regresión**

Fuente	Desviación	G.I.	Valor P
Modelo	47,225	19	0,0003
Residuos	426,679	324	0,0001
Total	473,904	343	

Como el valor p para el modelo en la tabla anterior es inferior a 0.01, se puede concluir que hay una relación estadísticamente significativa entre las variables de estudio.

Se observó el estadístico de Cox y Snell para cuantificar la bondad del ajuste, en este caso es igual a 12.8%, este valor puede suponer que el modelo explica de manera no tan adecuada la variabilidad del mismo. Sin embargo el porcentaje ajustado de Nagelkerke, el cual, es más adecuado para comparar modelos con diferentes números de variables independientes es 17.2%, al parecer mejora el ajuste, no obstante, los resultados no son satisfactorios, por lo cual se trata de mejorar el modelo identificando aquellas variables que no aportan significativamente al modelo. Aquí se aplicó el sistema paso a paso hacia atrás obteniendo los siguientes resultados:

Modelo 2

**Tabla 71. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B7	-1,06006689	0,2612951	4,9713E-05
B8	0,47593745	0,32814064	0,14694508
QF	0,30711564	0,22616527	0,17448726
M7	0,61802834	0,4379107	0,15815283
M8	0,87900545	0,3380921	0,00932511
S8	-0,73684034	0,42807171	0,08519618

L8	-0,61250793	0,34659515	0,07719234
L9	-0,60578928	0,34034758	0,0750901
Constante	3,08069437	1,60479002	0,05489746

El modelo resultante es:

$$v = 3,080 - 1,060 * B7 + 0,47593 * B8 + 0,3071 * QF + 0,618 * M7 + 0,8790 * M8 - 0,7368 * S8 - 0,612 * L8 - 0,605 * L9$$

El modelo es significativo según el análisis de varianza, sin embargo el ajuste baja con respecto al anterior modelo, puesto que los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 10,9% y 14,6%

Posteriormente se buscaron otros modelos de regresión logística construyendo una base donde las variables independientes tomaron los valores 1 y 0 (modelo 3), 0 si el estudiante tiene sobresaliente o excelente y 1 si tiene insuficiente o aprobado.

Modelo 3

**Tabla 72. Resultados de las estimaciones del modelo**

Variabes	Estimación $\beta$	Error Estandar	Valor P
B6	0,381	0,372	0,306
B7	1,392	0,365	0
B8	-0,759	0,393	0,054
B9	-0,193	0,337	0,566
QF	-0,607	0,434	0,163
M6	0,306	0,292	0,294

M7	-0,629	0,387	0,104
M8	-1,465	0,41	0
M9	0,541	0,415	0,192
M10	0,456	0,725	0,53
S6	-0,012	0,392	0,976
S7	-0,042	0,39	0,915
S8	0,645	0,4	0,107
S9	0,106	0,444	0,812
L6	0,06	0,3	0,841
L7	-0,214	0,304	0,481
L8	0,469	0,341	0,168
L9	0,34	0,423	0,422
L10	0,326	0,534	0,541
Constante	-0,646	0,681	0,343

Modelo

$$v = -0,646 + 0,381*B6 + 1,392*B7 - 0,759*B8 - 0,193*B9 + 0,06*L6 - 0,214*L7 + 0,469*L8 + 0,34*L9 + 0,326*L10 + 0,306*M6 - 0,629*M7 - 1,465*M8 + 0,541*M9 + 0,456*M10 - 0,607*QF - 0,012*S6 - 0,042*S7 + 0,645*S8 + 0,106*S9$$

Sin embargo, el análisis no mostró mejores resultados, los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 13,5% y 18%.

Al igual que en el modelo inicial, se reducen las variables utilizando el método paso a paso hacia atrás, obteniendo el modelo 4.

Modelo 4

**Tabla 73. Resultados de las estimaciones del modelo**

Variabes	Estimación $\beta$	Error Estandar	Valor P
B7	1,49560688	0,32820609	5,1911E-06
B8	-0,66232044	0,37740165	0,07926813
QF	-0,46024978	0,37492983	0,21961126
M6	0,37293221	0,2686141	0,16502842
M7	-0,55267747	0,36666876	0,13173542
M8	-1,36603709	0,39023104	0,00046424
M9	0,65136152	0,39884501	0,10244389
S8	0,69761122	0,37319377	0,06158105
L8	0,45942543	0,31085108	0,13941851
Constante	-0,11438664	0,4207094	0,78570668

$$v = -0,114 + 1,4956*B7 - 0,662*B8 + 0,4594*L8 - 0,460*QF + 0,372*M6 - 0,552*M7 - 1,36*M8 + 0,651*M9 + 0,6976 *S8$$

Sin embargo, según el análisis de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente los del 12,4% y 16,6%. Lo cual indica poco ajuste y similar al del modelo 1.

Finalmente se realizó el estudio con una base donde las variables independientes son los promedios en cada asignatura, se hizo el análisis con un modelo completo (modelo 5) y luego se eliminaron las variables menos significativas (modelo 6); los resultados arrojados por el software no indican un buen ajuste en ninguno de los dos casos, puesto que los de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente de 4% y 5.4% en el primer caso (modelo 5) y del 3,5% y 4,6% en el modelo reducido (modelo 6).

Del estudio anterior se puede deducir que el ajuste del modelo que describe la relación entre IS y el rendimiento académico es bajo y el modelo que mejor explica esta relación es el modelo 1, modelo al que se le realizan otras pruebas para analizar la habilidad del modelo para efectos de clasificación.

### Test de Bondad de ajuste Chi-cuadrado

Se aplica la Prueba de Bondad de Ajuste Chi Cuadrado para evaluar la habilidad de clasificación del modelo.

**Tabla 74. Bondad de ajuste**

Clase	Intervalo Logia	n	VERDADERO		FALSO	
			Observado	Esperado	Observado	Esperado
1	menor que -0,533383	69	18,0	17,9593	51,0	51,0407
2	-0,533383 a 0,14351	69	30,0	32,5239	39,0	36,4761
3	0,14351 a 0,55914	93	57,0	55,6089	36,0	37,3911
4	0,55914 a 0,813203	45	29,0	29,8221	16,0	15,1779
5	0,813203 o mayor	68	54,0	52,0858	14,0	15,9142
Total		344	188,0		156,0	

Chi-cuadrado = 0,82498 con 3 g.l. Valor P = 0,843482

Dado que el Valor P es superior a 0.10, no hay razón para descartar al modelo como medio de clasificación.

Al realizar el anterior estudio con los otros modelos, se obtuvieron los resultados que se indican en la tabla 75.

**Tabla 75. Valor del Chi-cuadrado en cada modelo presente en el estudio**

MODELO	Chi-cuadrado	Valor P
1	0,82498	0,843482
2	5,61489	0,131924
3	2,83167	0,41831
4	1,79098	0,616897
5	12,3664	0,006227
6	4,73328	0,192399

Como se puede observar todos los modelos presentan una buena adecuación.

**Tabla 76. Tabla de contingencias para la prueba de Hosmer y Lemeshow**

	IS = 0,00		IS = 1,00		Total
	Observado	Esperado	Observado	Esperado	Observado
1	26	27,3511557	8	6,64884426	34
2	24	23,0593065	10	10,9406935	34
3	23	19,4920904	11	14,5079096	34
4	17	17,1500636	18	17,8499364	35
5	12	14,9035643	22	19,0964357	34
6	16	13,4927709	18	20,5072291	34
7	10	12,3272622	24	21,6727378	34
8	13	11,383406	21	22,616594	34
9	6	9,67016576	28	24,3298342	34
10	9	7,1702147	28	29,8297853	37

Chi-cuadrado = 7,282 con 3 g.l. Valor P = 0,506

Dado que el Valor P de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo decir el modelo se ajusta adecuadamente a los datos

### Doble logaritmo del estadístico de Verosimilitud

Al compara el valor del  $-2\log$  verosimilitud cuando el modelo no tiene variables independientes con el valor del  $-2\log$  verosimilitud cuando se introducen estas variables al modelo se obtuvo lo siguiente.

$-2\log$  verosimilitud sin variables independientes: 473,904

$-2\log$  verosimilitud con variables independientes: 426,67

Como el valor del  $-2\log$  verosimilitud se redujo, el ajuste es bueno, más no ideal porque esta reducción no tiende a cero.

Al aplicar este estadístico en los otros modelos observamos los resultados que se indican en la tabla 77.

**Tabla 77. Valor del  $-2\log$  verosimilitud en cada modelo presente en el estudio**

MODELO	$-2\log$ verosimilitud. (A)	$-2\log$ verosimilitud. (D)
1	473,904	426,6791887463
2		434,1322184168
3		427,116105543
4		428,4147220191
5		459,8374709489
6		461,8040497232

Los resultados arrojados por el software en cada modelo indican un buen ajuste al modelo logístico, sin embargo las reducciones son menores en los modelos del 2 al 6.

**Tabla 78. Tabla de clasificación para IS**

Observado		Pronosticado		
		IS		Porcentaje correcto
		0,00	1,00	0,00
IS	0,00	78	78	50,0
	1,00	30	158	84,0
Porcentaje global				68,6

La ecuación del modelo ya diseñada nos proporciona una probabilidad que nos permite predecir a partir de ella para cada sujeto un valor de Y (Y predicho), tal que si  $P(Y=1|X) \leq 0.5$  entonces  $Y \text{ pred} = 1$ , y si  $P(Y=1|X) > 0.5$  entonces  $Y \text{ pred} = 0$

Al observar la lista de clasificación del modelo se contabilizan 108 casos mal clasificados de 344 casos, correspondientes al 31,4% , los cuales se indican a continuación:

**Tabla 79. Listado de casos Mal Clasificados**

Caso	Observado	Pronosticado	Grupo pronosticado
13	0	0,63415692	1
17	0	0,58197635	1
18	0	0,60278255	1
19	1	0,24881237	0
21	0	0,52113188	1
22	0	0,60552668	1
23	1	0,44504577	0
30	0	0,74043198	1
38	1	0,27789548	0
41	0	0,66516101	1
46	0	0,50459847	1
52	0	0,82038581	1
54	0	0,67985482	1

58	1	0,34312119	0
60	1	0,45720907	0
61	0	0,66156054	1
63	0	0,67286641	1
71	1	0,37663974	0
72	0	0,64414187	1
76	0	0,71001226	1
82	1	0,44032615	0
87	0	0,69876492	1
89	0	0,58692609	1
94	0	0,63192647	1
95	1	0,14057775	0
96	0	0,94268689	1
104	1	0,40601927	0
108	0	0,55931739	1
110	1	0,45225467	0
111	0	0,60355509	1
117	1	0,20137422	0
118	0	0,71618701	1
119	1	0,33180392	0
123	0	0,76440689	1
124	0	0,6927917	1
132	0	0,75107688	1
133	0	0,65333486	1
135	1	0,39956216	0
136	0	0,63625347	1
138	0	0,50029036	1
140	0	0,68476875	1
143	1	0,24622302	0
144	1	0,47473539	0
147	0	0,80601085	1
158	0	0,81017157	1
159	1	0,20577469	0
162	0	0,64206791	1
166	0	0,52375524	1

167	0	0,58238819	1
168	0	0,64591368	1
169	1	0,35050701	0
174	1	0,35234172	0
175	0	0,66127054	1
176	0	0,80137161	1
177	1	0,40658408	0
181	0	0,50272166	1
197	0	0,59159817	1
199	0	0,65264626	1
200	0	0,65290215	1
204	1	0,20335958	0
208	0	0,5570827	1
209	1	0,45830653	0
215	0	0,66454695	1
217	0	0,53171356	1
218	0	0,61706228	1
219	0	0,50058696	1
228	0	0,77740941	1
231	1	0,35384701	0
233	1	0,29916108	0
239	0	0,5876756	1
245	1	0,3078455	0
246	0	0,51491175	1
248	1	0,36023696	0
255	0	0,69812358	1
257	0	0,56252435	1
261	1	0,35343515	0
262	0	0,65400689	1
263	0	0,5135406	1
264	0	0,59194402	1
265	0	0,63625347	1
266	0	0,50029036	1
270	0	0,57441015	1
271	0	0,63625347	1

272	0	0,53319689	1
273	1	0,20337654	0
278	0	0,57441015	1
281	0	0,63625347	1
282	0	0,6724371	1
285	0	0,57441015	1
288	0	0,63625347	1
297	0	0,57441015	1
299	0	0,57441015	1
303	0	0,87609076	1
304	1	0,45557596	0
308	0	0,57441015	1
312	1	0,26432938	0
313	0	0,59682551	1
317	0	0,59489807	1
320	0	0,57372644	1
323	0	0,63625347	1
327	0	0,54762564	1
328	0	0,63625347	1
329	0	0,59682551	1
330	0	0,54334444	1
331	1	0,44247817	0
334	0	0,63253279	1
336	0	0,63625347	1
338	0	0,53319689	1

Al realizar este mismo estudio para cada modelo encontramos los resultados de la tabla 80.

**Tabla 80. Total de casos mal clasificados en cada modelo presente en el estudio**

MODELO	CASOS MAL CLASIFICADOS
1	108
2	123
3	113
4	115
5	138
6	134

### Residuos atípicos

**Tabla 81. Residuos atípicos para IS**

Fila	Y	Predicho Y	Residuo	Residuo Pearson	Desviación Residuo
52	0,0	0,820386	-0,820386	-2,14	-1,85
95	1,0	0,140578	0,859422	2,47	1,98
96	0,0	0,942687	-0,942687	-4,06	-2,39
147	0,0	0,806011	-0,806011	-2,04	-1,81
158	0,0	0,810172	-0,810172	-2,07	-1,82
176	0,0	0,801372	-0,801372	-2,01	-1,80
303	0,0	0,876091	-0,876091	-2,66	-2,04

La tabla de residuos atípicos lista todas las observaciones que tienen residuo Pearson o desviación de residuo superiores a 2.0 en valor absoluto. Estos residuos estandarizados miden cuántas desviaciones estándar de cada valor observado de IM proceden del modelo ajustado. En este caso, hay 11 residuos estandarizados superiores a 2.0, y 2 superiores a 3.0.

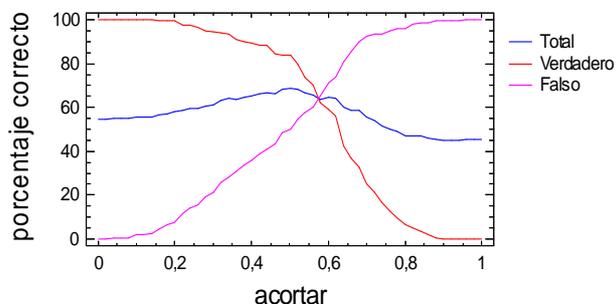
Al estudiar con detalle estos individuos se observa que aquellos que están por ENCIMA del nivel muy superior (0), su nivel académico es BAJO; y aquellos que están por DEBAJO del nivel muy superior (1), presentan un rendimiento académico BUENO en su vida escolar. Por lo tanto, es conveniente mantenerlos en el estudio puesto que son casos reales que no presentan errores al levantamiento ni digitación de la información ya que se confrontó la base de datos con la información de los libros de registro.

### PREDICCIÓN DEL MODELO

**Tabla 82. Capacidad de Predicción del Modelo**

<i>Cohorte</i>	<i>VERDADERO</i>	<i>FALSO</i>	<i>Total</i>
0,0	100,00	0,00	54,65
0,05	100,00	0,64	54,94
0,1	100,00	1,92	55,52
0,15	99,47	2,56	55,52
0,2	99,47	7,69	57,85
0,25	96,28	14,74	59,30
0,3	94,68	21,15	61,34
0,35	93,09	30,13	64,53
0,4	89,36	35,90	65,12
0,45	86,70	42,95	66,86
0,5	84,04	50,00	68,60
0,55	72,34	58,97	66,28
0,6	59,04	71,15	64,53
0,65	39,36	82,69	59,01
0,7	25,00	92,31	55,52
0,75	14,89	94,23	50,87
0,8	6,38	96,15	47,09
0,85	3,72	98,72	46,80
0,9	0,00	99,36	45,06
0,95	0,00	100,00	45,35
1,0	0,00	100,00	45,35

**Figura 5. Capacidad de Predicción del Modelo para IS**



En la tabla y el gráfico se observa la capacidad de predicción del modelo 1. Si el valor predicho es más grande que el punto de cohorte, la respuesta se predice como VERDADERA. Si el valor predicho es inferior a o igual al punto de cohorte, la repuesta se predice para ser FALSO. La tabla muestra el porcentaje de datos observados predichos correctamente a diferentes puntos de cohorte. En nuestro caso con el punto de cohorte utilizado 0,5, el 84,04% de todas las respuestas VERDADERAS se predijeron correctamente, mientras que el 502% de todas las respuestas FALSAS se predijeron correctamente, para un total de 68,60% casos clasificados correctamente.

**Tabla 83. Matriz de correlación para los coeficientes estimados**

	Cons	B6	B7	B8	B9	QF	M6	M7	M8	M9	M10	S6	S7	S8	S9	L6	L7	L8	L9	L10
Cons	1																			
B6	-0	1																		
B7	0,1	0,1	1																	
B8	-0,2	-0,1	-0,1	1																
B9	-0,1	0,1	-0	0,1	1															
QF	0,1	-0,1	-0	0	-0,2	1														
M6	-0,2	-0	-0,1	0	0,1	0	1													
M7	-0,2	-0,2	-0,1	0	0	-0	-0,2	1												
M8	-0,1	-0,1	-0,2	-0	-0,1	-0,1	-0,1	-0,1	1											
M9	-0,1	-0	0	-0,2	0,1	-0,1	0,1	-0,2	-0,1	1										
M10	-0,2	0	-0,1	-0	-0	-0,2	-0,1	0	-0	-0	1									
S6	-0	-0,2	-0	-0	-0	-0	-0,2	0,1	-0	-0,1	-0	1								
S7	0	-0,2	-0,1	0	-0	0,1	-0,1	-0,2	0	-0	0,1	-0,2	1							
S8	-0,1	0,1	0,1	-0,2	-0,1	-0	-0	-0,2	-0	-0,1	0	-0,1	-0,1	1						
S9	-0,2	-0	-0,2	-0,1	-0,2	-0,1	-0	0	0,1	0	-0	-0,1	0,1	-0,2	1					
L6	-0	-0,2	-0	-0	-0	-0,1	-0,2	0,1	0,1	-0,1	0,1	-0,2	-0	-0,1	0,1	1				
L7	-0,1	-0,1	-0,1	-0	-0	-0	-0,1	-0	0	-0	0	-0,1	-0,1	0,1	-0,1	-0,2	1			
L8	0,2	0,1	0,1	-0	-0,1	-0	0,1	-0,3	-0,2	-0,1	-0,2	-0,1	-0	-0,1	-0,1	-0,1	-0,2	1		
L9	0	-0	0	0	-0,1	-0	0	0	-0,2	-0,1	-0,1	-0	-0,1	-0,1	-0,1	-0,1	-0	0	1	
L10	-0,2	-0	-0,1	0	-0,1	-0	-0,1	0	0,1	-0,1	0	0,1	0	-0	-0,1	-0	-0	-0,1	-0,3	1

Esta tabla muestra las correlaciones estimadas entre los coeficientes en el modelo ajustado. Estas correlaciones pueden utilizarse para detectar la presencia de serias multicorrelaciones, es decir, correlación entre las variables pronosticadas. En este caso, no hay ninguna correlación con valores absolutos superiores a 0.5.

## INTERPRETACION DEL MODELO

Como se puede observar en el estudio anterior, el modelo 1 presenta un adecuado ajuste, las variables que ayudan a predecir si un estudiante de grado décimo del Colegio Javeriano estará por encima o debajo del nivel muy superior en las pruebas de estado ICFES en la asignatura de Sociales son todas las materias de matemáticas, castellano, sociales y naturales, que cursó desde grado sexto a grado décimo en la institución.

A continuación se indica el valor de las probabilidades que tiene cada variable de estudio en el modelo.

**Tabla 84. Variables en el modelo**

Asignatura	B	Exp(B)
B6	-0,3	0,74
B7	-0,99	0,37
B8	0,65	1,91
B9	0,26	1,3
QF	0,37	1,45
M6	-0,47	0,62
M7	0,76	2,14
M8	0,86	2,36
M9	-0,28	0,76
M10	-0,16	0,85

Asignatura	B	Exp(B)
S6	0,17	1,18
S7	0,44	1,56
S8	-0,69	0,5
S9	-0,45	0,63
L6	-0,3	0,74
L7	0,23	1,26
L8	-0,6	0,55
L9	-0,54	0,58
L10	0,01	1,01
Constante	4,23	68,8

De lo anterior se puede decir que el valor de  $-0,3$  que se asocia a la variable B6, incrementa el 74% la posibilidad de estar por debajo del nivel muy superior en el área de matemáticas, puesto que  $e^{-0,3} = 0,74$

De igual manera se pueden leer todos los coeficientes que acompañan a cada variable de estudio.

Como se puede observar, se evidencia la importancia de las asignaturas B8,B9,QF, M7,M8,S6,S7,L7,L10 las cuales realizan un aporte alto para el cálculo de esta probabilidad en el área de sociales, mientras que las otras asignaturas aportan en un porcentaje bajo.

El modelo seleccionado (modelo 1) ayudará a detectar a aquellos estudiantes de grado décimo que requieren de un refuerzo académico antes de presentar la prueba ICFES. A continuación se indica la probabilidad de que cada estudiante de grado décimo pueda estar por debajo del nivel muy superior.

**Tabla 85. Casos posibles que necesitan refuerzo académico en sociales antes de las pruebas ICFES**

Caso	B6	B7	B8	B9	QF	L6	L7	L8	L9	L10	M6	M7	M8	M9	M10	S6	S7	S8	S9		P(y=1)
1	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,300	0,574
1	4,6	4,6	4,3	3,5	3,5	4,6	4,6	4,3	3,5	3,5	4,6	4,3	3,5	3,5	3,5	4,6	4,6	4,3	3,5	-0,697	0,332
3	3,5	3,5	3,5	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	4,6	3,5	4,3	-0,096	0,476
4	4,6	4,3	4,3	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,3	3,5	3,5	3,5	3,5	4,6	4,6	4,3	3,5	-0,937	0,282
5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	2	3,5	2	4,3	3,5	3,5	-1,025	0,264
6	4,6	4,3	4,3	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,6	3,5	3,5	4,3	4,3	4,3	4,6	4,3	4,3	-1,194	0,233
7	4,3	3,5	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	3,5	4,3	4,3	4,3	-0,168	0,458
8	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,915	0,714
9	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	-0,595	0,355
10	3,5	2	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	4,3	3,5	3,5	2,481	0,923
11	4,6	4,6	4,3	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	-0,789	0,312
11	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	4,3	4,3	3,5	-0,734	0,324
13	4,3	3,5	3,5	4,3	3,5	3,5	3,5	3,5	4,3	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	-0,459	0,387

14	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	-0,287	0,429
15	4,6	4,3	4,6	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,6	4,3	4,3	4,3	4,6	4,6	4,6	4,3	-0,384	0,405
16	4,3	4,3	4,3	4,6	4,6	4,3	4,3	3,5	4,6	4,6	4,6	4,3	3,5	4,6	4,6	4,3	4,3	4,3	4,6	-0,868	0,296
17	3,5	3,5	3,5	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	3,5	4,3	4,3	4,3	-1,134	0,243
18	3,5	4,3	3,5	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	4,3	4,3	3,5	4,3	-0,812	0,307
19	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,3	4,3	-0,091	0,477
10	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,6	4,3	4,3	-0,925	0,284
11	4,6	4,3	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	-0,297	0,426
11	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	-0,236	0,441
13	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,3	4,6	-0,388	0,404
14	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,6	4,3	4,3	-0,419	0,397
15	3,5	3,5	2	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,6	3,5	3,5	-1,347	0,206
16	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	4,3	3,5	3,5	-0,467	0,385
17	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,3	4,6	4,6	4,6	-0,645	0,344
18	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,6	4,6	4,3	4,3	-1,260	0,221
19	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,6	4,3	4,3	4,3	3,5	4,6	4,3	4,3	-0,231	0,443
30	3,5	3,5	2	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,6	3,5	3,5	-1,347	0,206
31	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	2,404	0,917
31	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,300	0,574
33	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	3,5	3,5	4,3	4,3	4,3	4,3	-0,755	0,32
34	4,6	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	3,5	4,3	4,3	4,3	-1,556	0,174
35	4,3	3,5	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	4,3	4,3	0,524	0,628
36	3,5	2	2	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	2	3,5	2	2	4,3	4,3	3,5	-0,755	0,32
37	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	0,494	0,621
38	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,915	0,714
39	4,6	4,3	4,3	3,5	3,5	4,6	4,6	4,3	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,3	4,6	4,3	3,5	0,238	0,559
40	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4	4,6	4,6	4,3	-0,096	0,476
41	4,6	4,3	4,3	3,5	3,5	4,6	4,6	3,5	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,6	4,3	4,3	3,5	0,636	0,654
41	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	4,3	3,5	3,5	1,602	0,832
43	3,5	4,3	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	2	3,5	4,3	3,5	3,5	-2,630	0,067
44	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	0,915	0,714
45	3,5	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,6	4,3	3,5	-0,483	0,381
46	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	3,5	3,5	4,6	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	-0,471	0,384
47	4,6	4,3	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,6	3,5	3,5	-0,316	0,422
48	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,6	4,3	4,3	-0,161	0,46
49	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,3	4,6	4,3	4,3	-1,427	0,193
50	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	4,6	3,5	3,5	1,468	0,813
51	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4	3,5	3,5	4	4,3	4,3	3,5	3,5	4,3	4,6	4,3	3,5	0,402	0,599
51	4,6	4,6	4,3	4,3	4,3	4,6	4,6	4,6	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,3	4,3	4,3	-0,704	0,331
53	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	2	3,5	3,5	4,6	3,5	3,5	0,703	0,669
54	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	2	3,5	3,5	3,5	3,5	4,3	3,5	3,5	-0,606	0,353
55	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	4,3	3,5	3,5	-0,572	0,361
56	3,5	3,5	3,5	2	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,3	3,5	3,5	-0,901	0,289
57	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	0,559	0,636
58	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	-0,822	0,305
59	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,6	4,3	4,6	4,3	4,6	-0,437	0,392
60	4,6	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,6	4,6	4,3	4,3	-1,260	0,221
61	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,6	4,3	4,3	4,3	3,5	4,6	3,5	3,5	0,686	0,665
61	3,5	3,5	2	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	3,5	4,3	3,5	3,5	-1,481	0,185

63	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,6	4,6	4,6	1,277	0,782
64	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,6	4,3	4,3	-0,483	0,381
65	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	3,5	3,5	4,3	4,6	4,3	4,3	-0,621	0,349	
66	4,6	4,3	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	3,5	4,6	3,5	3,5	-0,506	0,376		
67	4,3	3,5	4,3	4,3	4,3	3,5	4,3	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	4,3	3,5	3,5	1,440	0,809		
68	3,5	2	2	3,5	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	2	3,5	2	2	4,3	3,5	3,5	-0,203	0,449		
69	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	4,3	0,130	0,533	
70	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	4,3	-0,002	0,5	
71	4,6	4,3	4,3	3,5	3,5	4,6	4,6	4,3	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,3	4,3	4,3	4,3	-0,259	0,436		
71	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4	4,3	4,3	3,5	0,342	0,585		
73	4,6	4,3	4,3	3,5	3,5	4,6	4,6	3,5	3,5	3,5	4,6	4,3	4,3	3,5	3,5	4,6	4,3	4,3	3,5	0,636	0,654		
74	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	3,5	4,3	3,5	3,5	1,602	0,832	
75	3,5	4,3	3,5	3,5	2	3,5	3,5	3,5	3,5	3,5	3,5	2	2	3,5	2	3,5	4,6	4,3	3,5	-3,049	0,045		
76	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,6	4,6	4,3	-0,076	0,481		
77	3,5	4,3	4,3	4,3	3,5	3,5	3,5	4,3	4,3	4,3	3,5	3,5	3,5	3,5	3,5	3,5	4,3	4,3	3,5	-0,617	0,351		
78	4,3	4,3	4,3	3,5	3,5	4,3	4,3	4,3	3,5	3,5	4,6	4,3	3,5	3,5	3,5	4,3	4,3	3,5	3,5	0,082	0,52		
79	4,6	4,3	3,5	3,5	3,5	4,3	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,5	4,3	3,5	3,5	-0,449	0,389		
80	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,3	4,3	4,3	4,6	4,3	4,6	4,3	4,3	4,6	4,3	3,5	3,5	0,622	0,651		

## CONCLUSIONES DEL MODELO

El modelo de regresión logística escogido esta conformado por todas las variables independientes del estudio.

El modelo de regresión logística obtenido, nos indica un ajuste adecuado de las variables de estudio y la variable IS.

Las pruebas de bondad y ajuste que ilustraron el grado de clasificación del modelo, con las pruebas Chi cuadrado, Hosmer-Lemeshow y el estadístico de verosimilitud del modelo indicaron un adecuado ajuste de clasificación de los datos observados.

El estadístico de verosimilitud indica un adecuado ajuste del modelo, sin embargo se observa que la disminución de este valor al introducir las variables independientes es mínimo, pero se evidencia reducción de este valor al comparar

este dato con el que arroja el estadístico cuando realiza el cálculo sin las introducir las variables independientes.

El modelo seleccionado, tiene un 68.6% de casos bien clasificados y el 31.4% corresponde a los casos *mal* clasificados.

Al realizar el estudio de multicorrelaciones del modelo se observa ausencia de correlaciones entre las variables de estudio.

A pesar de que todas las variables del modelo son significativas, las variables B8,B9,QF, M7,M8,S6,S7,L7 y L10 establecen *un* aporte alto al cálculo de la probabilidad que tiene un estudiante Javeriano de décimo grado, de obtener un puntaje por debajo del nivel muy superior.

El pronóstico que realiza el modelo indica que cuanto mejor es el record académico de los estudiantes la probabilidad de estar por debajo del nivel muy superior en el IS disminuye no obstante, se evidencian casos de estudiantes que presentan un rendimiento académico adecuado, sin embargo, los resultados en las pruebas de estado en sociales, muestra una probabilidad alta de estar por debajo del nivel muy superior. También se observan casos donde algunos estudiantes que tiene un desempeño académico bajo, muestra una probabilidad baja de estar por debajo del nivel muy superior. Estos dos sucesos describen la realidad que año tras año envuelve este sistema de evaluación.

#### 4.4. Modelo para el área de Lenguaje

Modelo 1 (Inicial):

Se tomó como Variable dependiente IL y como variables independientes: M6, M7, M8, M9, M10, C6, C7, C8, C9, C10, S6, S7, S8, S9, B6, B7, B8, B9, QF.

Para obtener el valor estimado del predictor lineal  $\eta = x\beta$  y el valor esperado del modelo de regresión logística se utilizó el método de mínimos cuadrados ponderados generalizados iterados, obteniendo los resultados siguientes:

**Tabla 86. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	0,35241338	0,35273309	0,31774934
B7	-0,15596881	0,28660956	0,58631372
B8	-0,23730501	0,34970132	0,49739551
B9	-0,65076551	0,26129848	0,01275612
QF	-0,32458228	0,27727738	0,2417576
M6	0,561813	0,45680833	0,21874732
M7	-0,94197144	0,50893025	0,06418608
M8	-0,25948757	0,38702096	0,50255555
M9	0,05675862	0,40581499	0,888768
M10	-0,15101832	0,32572149	0,6429039
S6	-0,79118606	0,48239686	0,10098138
S7	1,10282183	0,45028399	0,0143186
S8	0,01093503	0,4727914	0,98154764
S9	-0,65144637	0,47316937	0,16858273
L6	-0,24540975	0,41835752	0,55747041
L7	-0,09763001	0,37131413	0,79260373
L8	-0,34790883	0,42271719	0,41049107

L9	-0,58725951	0,41156397	0,1536095
L10	0,44704372	0,37699373	0,23569652
Constante	10,9693735	2,2252588	8,2448E-07

Los datos obtenidos indican el ajuste del modelo que describe la relación entre IL y el rendimiento académico conformado por las 19 variables antes mencionadas.

### Modelo de regresión

$$IL = \frac{e^v}{1 + e^v}$$

Donde  $v$  es  $x'\beta$

$$v = 10,9694 + 0,447044*L10 - 0,24541*L6 - 0,09763*L7 - 0,347909*L8 - 0,58726*L9 - 0,151018*M10 + 0,561813*M6 - 0,941971*M7 - 0,259488*M8 + 0,0567586*M9 - 0,324582*QF - 0,791186*S6 + 1,10282*S7 + 0,010935*S8 - 0,651446*S9 + 0,352413*B6 - 0,155969*B7 - 0,237305*B8 - 0,650766*B9$$

Se realizó un análisis de varianza para estudiar el ajuste del modelo, del cual se obtuvieron los resultados que se presentan en la tabla 87.

**Tabla 87. Análisis de varianza para la regresión**

Fuente	Desviación	G.l.	Valor P
Modelo	67,1192	19	0,0000
Residuos	401,254	324	0,0022
Total	468,373	343	

Como el valor p para el modelo en la tabla anterior es inferior a 0.01, se puede concluir que hay una relación estadísticamente significativa entre las variables de estudio.

Se observó el estadístico de Cox y Snell para cuantificar la bondad del ajuste, en este caso es igual a 17.7%, este valor puede suponer que el modelo explica de manera no tan adecuada la variabilidad del mismo. Sin embargo el porcentaje ajustado de Nagelkerke, el cual, es más adecuado para comparar modelos con diferentes números de variables independientes es 23.8%, al parecer mejora el ajuste, no obstante, los resultados no son satisfactorios, por lo cual se trata de mejorar el modelo identificando aquellas variables que no aportan significativamente al modelo. Aquí se aplicó el sistema paso a paso hacia atrás obteniendo los siguientes resultados:

Modelo 2

**Tabla 88. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B9	-0,69907737	0,25197732	0,00553094
QF	-0,37092181	0,26385204	0,15978509
M7	-0,89056692	0,46236235	0,05408898
S6	-0,68775493	0,41916133	0,1008414
S7	1,17557183	0,41411126	0,0045286
S9	-0,68228364	0,44669387	0,12665967
L8	-0,49943958	0,39702782	0,20841126
L9	-0,65677858	0,38867833	0,09107118
L10	0,46325089	0,37257866	0,21373373
Constante	10,7628013	1,92142724	2,1255E-08

El modelo resultante es:

$$v = 10,7628 - 0,6990 \cdot B9 - 0,3709 \cdot QF + 0,3071 \cdot QF - 0,8905 \cdot M7 - 0,6877 \cdot S6 + 1,1755 \cdot S7 - 0,6822 \cdot S9 - 0,4994 \cdot L8 - 0,6567 \cdot L9 + 0,4632 \cdot L10$$

El modelo es significativo según el análisis de varianza, sin embargo el ajuste baja un poco con respecto al anterior modelo, puesto que los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 16,8 y 22,5%

Posteriormente se buscaron otros modelos de regresión logística construyendo una base donde las variables independientes tomaron los valores 1 y 0 (modelo 3), 0 si el estudiante tiene sobresaliente o excelente y 1 si tiene insuficiente o aprobado.

Modelo 3

**Tabla 89. Resultados de las estimaciones del modelo**

Variabes	Estimación $\beta$	Error Estandar	Valor P
B6	-0,28807437	0,37890391	0,44708518
B7	-0,16796486	0,35830334	0,63922805
B8	0,46396986	0,37880226	0,2206378
B9	1,18503404	0,3297309	0,00032571
QF	-0,25996445	0,44262217	0,55698336
M6	-0,13206831	0,30377048	0,66373411
M7	0,35887795	0,37493009	0,33847322
M8	-0,05325257	0,36873027	0,88516766
M9	0,31817294	0,40840936	0,4359479
M10	0,14397668	0,73496393	0,84469137
S6	0,31230265	0,39518529	0,42937073
S7	-0,30417332	0,3931541	0,43912328
S8	-0,09624701	0,3996084	0,8096688
S9	0,09510323	0,43333686	0,82628612
L6	-0,13661817	0,31042123	0,65986052
L7	0,46917285	0,30506052	0,12405682

L8	0,09137775	0,340124	0,78819126
L9	0,48875861	0,4104344	0,23371936
L10	0,99990139	0,55676137	0,07250652
Constante	-2,44881689	0,77929656	0,00167602

Modelo

$$v = -2,44881 - 0,288*B6 - 0,1679*B7 + 0,4639*B8 + 1,1850*B9 - 0,259*QF - 0,1320*M6 + 0,3588*M7 + 0,3181*M9 + 0,1439*M10 - 0,0532*M8 + 0,31230*S6 - 0,3041*S7 - 0,096*S8 + 0,0951*S9 - 0,13661*I6 + 0,4691*I7 + 0,09137*I8 + 0,4887*I9 + 0,9999*I10$$

Sin embargo, el análisis no mostró mejores resultados, los estadísticos de Cox y Snell y Nagelkerke, son respectivamente del 16,2% y 21,7%.

Al igual que en el modelo inicial, se redujo las variables utilizando el método paso a paso hacia atrás, obteniendo el modelo

Modelo 4

**Tabla 90. Resultados de las estimaciones del modelo**

Variables	Estimación $\beta$	Error Estandar	Valor P
B6	-0,2770951	0,35119256	0,43010507
B8	0,49050326	0,36778695	0,18231485
B9	1,14794261	0,29524874	0,00010105
M7	0,32369701	0,34657982	0,35031661
M9	0,31161676	0,37023744	0,3999742
S7	-0,29567421	0,36553348	0,41858183
L7	0,45351981	0,27469194	0,09873604
L9	0,5032154	0,40010019	0,20849178

L10	0,88196889	0,50370337	0,07995128
Constante	-2,44491618	0,53539358	4,9575E-06

Modelo

$$v = -2,4449 - 0,277*B6 + 0,4905*B8 + 1,1479*B9 + 0,881*L10 + 0,5032*L9 + 0,4535*L7 + 0,3236*M7 + 0,31161*M9 - 0,2956 *S8$$

Sin embargo, según el análisis de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente los del 15,7% y 21,1%. Lo cual indica poco ajuste.

Finalmente se realizó el estudio con una base donde las variables independientes son los promedios en cada asignatura, se hizo el análisis con un modelo completo (modelo 5) y luego se eliminó las variables menos significativas (modelo 6); los resultados arrojados por el software no indican un buen ajuste en ninguno de los dos casos, puesto que los de estadísticos de Cox y Snell y Nagelkerke, que son respectivamente de 10% y 13,5% en el primer caso (modelo 5) y del 9,3% y 12,6% en el modelo reducido (modelo 6).

Del estudio anterior se puede deducir que el ajuste del modelo que describe la relación entre IL y el rendimiento académico es bajo y el modelo que mejor explica esta relación es el modelo 4, modelo al que se le realizan otras pruebas para analizar la habilidad del modelo para efectos de clasificación.

### **Test de Bondad de ajuste Chi-cuadrado**

Se aplica la Prueba de Bondad de Ajuste Chi Cuadrado para evaluar la habilidad de clasificación del modelo.

**Tabla 91. Bondad de ajuste**

Clase	Intervalo Logia	n	VERDADERO		FALSO	
			Observado	Esperado	Observado	Esperado
1	menor que -0,424418	69	15,0	16,6476	54,0	52,3524
2	-0,424418 a 0,399827	69	36,0	34,8134	33,0	34,1866
3	0,399827 a 0,853347	69	50,0	45,1914	19,0	23,8086
4	0,853347 a 1,09478	124	88,0	92,1918	36,0	31,8082
5	1,09478 o mayor	13	10,0	10,1559	3,0	2,84413
Total		344	199,0		145,0	

Chi-cuadrado = 2,53333 con 3 g.l. Valor P = 0,469294

Dado que el Valor P es superior a 0.10, no hay razón para descartar al modelo como medio de clasificación.

Al realizar el anterior estudio con los otros modelos, se obtuvieron los resultados que se indican en la tabla 92.

**Tabla 92. Valor del Chi-cuadrado en cada modelo presente en el estudio**

MODELO	Chi-cuadrado	Valor P
1	2,86197	0,413396
2	1,35909	0,715148
3	1,54542	0,671826
4	2,53333	0,469294
5	1,43824	0,696594
6	1,33879	0,512017

Como se puede observar todos los modelos presentan una buena adecuación.

## Prueba de Hosmer-Lemeshow

**Tabla 93. Tabla de contingencias para la prueba de Hosmer y Lemeshow**

	L = 0		L = 1		Total
	Observado	Esperado	Observado	Esperado	
1	29	29,319	5	4,681	34
2	25	22,429	9	11,571	34
3	18	18,967	17	16,033	35
4	15	15,824	20	19,176	35
5	4	6,941	15	12,059	19
6	12	12,764	25	24,236	37
7	12	10,088	22	23,912	34
8	30	28,668	86	87,332	116

Chi-cuadrado = 3,708 con 6 g.l. Valor P = 0, 716

Dado que el Valor P de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo, es decir, el modelo se ajusta adecuadamente a los datos.

### **Doble logaritmo del estadístico de Verosimilitud**

Al compara el valor del  $-2\log$  verosimilitud cuando el modelo no tiene variables independientes con el valor del  $-2\log$  verosimilitud cuando se introducen estas variables al modelo se obtuvo lo siguiente.

$-2\log$  verosimilitud sin variables independientes: 468,373

$-2\log$  verosimilitud con variables independientes: 409,53

Como el valor del  $-2\log$  verosimilitud se redujo, el ajuste es bueno, más no ideal porque esta reducción no tiende a cero.

Al aplicar este estadístico en los otros modelos observamos los resultados que se indican en la tabla 94.

**Tabla 94. Valor del  $-2\log$  verosimilitud en cada modelo presente en el estudio**

MODELO	$-2\log$ verosimilitud. (A)	$-2\log$ verosimilitud. (D)
1	468,373	401,2541824581
2		405,2822208266
3		407,7033097384
4		409,5353739056
5		432,0338886678
6		434,6282371243

Los resultados arrojados por el software en cada modelo indican un buen ajuste al modelo logístico, sin embargo las reducciones son menores en todos los modelos.

#### Tabla de clasificación

**Tabla 95. Tabla de clasificación para IL**

Observado	Pronosticado		
	L		Porcentaje correcto
	0	1	0
L 0	72	73	49,7
1	31	168	84,4
Porcentaje global			69,8

La ecuación del modelo ya diseñada nos proporciona una probabilidad que nos permite predecir a partir de ella para cada sujeto un valor de Y (Y predicho), tal que si  $P(Y=1|X) \leq 0.5$  entonces  $Y \text{ pred} = 1$ , y si  $P(Y=1|X) > 0.5$  entonces  $Y \text{ pred} = 0$

Al observar la lista de clasificación del modelo se contabilizan 104 casos mal clasificados de 344 casos, correspondientes al 30,2%, los cuales se indican a continuación:

**Tabla 96. Listado de casos Mal Clasificados**

Caso	Observado	Pronosticado	Grupo pronosticado
6	1	0,44285353	0
11	0	0,65503788	1
12	1	0,37597115	0
25	0	0,74928043	1
30	1	0,48671204	0
38	1	0,09695652	0
40	0	0,74928043	1
42	1	0,44285353	0
45	1	0,48671204	0
47	1	0,37597115	0
48	1	0,48671204	0
49	0	0,74928043	1
50	1	0,48671204	0
51	1	0,40688412	0
56	0	0,53445606	1
59	1	0,48671204	0
60	1	0,1918012	0
65	1	0,37597115	0
68	1	0,48671204	0
69	0	0,71470651	1
75	0	0,74928043	1
77	1	0,48671204	0
78	1	0,40688412	0
80	1	0,15285915	0
85	0	0,68636109	1
89	0	0,71144471	1
99	0	0,51886569	1

101	0	0,64719606	1
110	0	0,65503788	1
113	1	0,21056613	0
122	0	0,65503788	1
123	0	0,74928043	1
124	0	0,55300246	1
125	1	0,32492844	0
127	0	0,74928043	1
130	0	0,74928043	1
132	0	0,57872984	1
134	1	0,40980269	0
144	0	0,64372366	1
149	0	0,53445606	1
150	0	0,58167219	1
151	0	0,80066447	1
159	1	0,32214627	0
161	1	0,20594573	0
162	1	0,48671204	0
163	0	0,74928043	1
166	0	0,65503788	1
168	0	0,52203839	1
175	0	0,65503788	1
176	0	0,53761758	1
180	0	0,74627359	1
181	0	0,70895842	1
186	0	0,70895842	1
187	0	0,64663375	1
188	0	0,77101084	1
195	0	0,74928043	1
196	0	0,57483822	1
197	0	0,64443285	1
198	0	0,74928043	1
200	0	0,68375474	1
202	0	0,56120559	1
206	0	0,5260148	1

209	1	0,44456942	0
216	0	0,65503788	1
218	0	0,65503788	1
220	1	0,30864316	0
234	1	0,37875086	0
237	0	0,70710606	1
238	0	0,57872984	1
247	0	0,74928043	1
248	0	0,74928043	1
249	0	0,74928043	1
255	0	0,74928043	1
256	0	0,65503788	1
259	0	0,70895842	1
261	0	0,65503788	1
264	1	0,39546002	0
265	0	0,74928043	1
267	1	0,4598711	0
270	0	0,74928043	1
271	0	0,74928043	1
277	0	0,57264679	1
282	0	0,68375474	1
283	0	0,65503788	1
287	0	0,77101084	1
288	0	0,74928043	1
297	0	0,74928043	1
298	0	0,71847965	1
299	0	0,74928043	1
304	0	0,74928043	1
308	0	0,74928043	1
309	0	0,52203839	1
311	0	0,65503788	1
312	1	0,36438553	0
313	0	0,74928043	1
318	0	0,53064464	1
320	0	0,65503788	1

323	0	0,74928043	1
325	1	0,32866586	0
327	0	0,71470651	1
334	0	0,74928043	1
335	0	0,74928043	1
338	0	0,74928043	1
341	1	0,45731502	0

Al realizar este mismo estudio para cada modelo encontramos los resultados de la tabla

**Tabla 97. Total de casos mal clasificados en cada modelo presente en el estudio**

MODELO	CASOS MAL CLASIFICADOS
1	106
2	110
3	104
4	104
5	128
6	134

### Residuos atípicos

**Tabla 98. Residuos atípicos para IL**

Fila	Y	Predicho Y	Residuo	Pearson Residuo	Desviación Residuo
38	1	0,096956	0,903043	3,05	2,16
60	1	0,191801	0,808199	2,05	1,82
80	1	0,152859	0,847141	2,35	1,94
151	0	0,800664	-0,800664	-2,00	-1,80

La tabla de residuos atípicos lista todas las observaciones que tienen residuo Pearson o desviación de residuo superiores a 2.0 en valor absoluto. Estos residuos estandarizados miden cuántas desviaciones estándar de cada valor observado de ICFESM proceden del modelo ajustado. En este caso, hay 4 residuos estandarizados superiores a 2.0, y 2 superiores a 3.0.

Al estudiar con detalle estos individuos se observa que aquellos que están por ENCIMA del nivel muy superior (0), su nivel académico es BAJO; y aquellos que están por DEBAJO del nivel muy superior (1), presentan un rendimiento académico BUENO en su vida escolar. Por lo tanto, es conveniente mantenerlos en el estudio puesto que son casos reales que no presentan errores al levantamiento ni digitación de la información ya que se confrontó la base de datos con la información de los libros de registro y la percepción de profesores antiguos de la institución sobre la veracidad de la información.

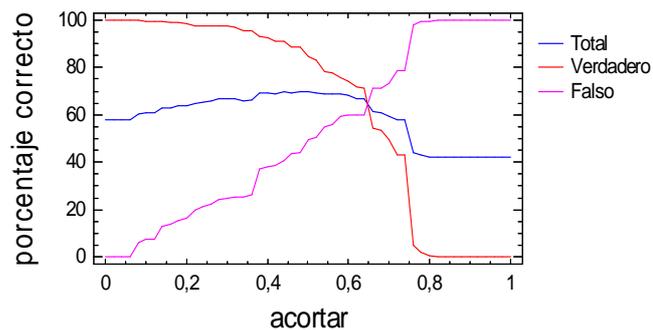
## PREDICCIÓN DEL MODELO

**Tabla 99. Capacidad de Predicción del Modelo**

<i>Cohorte</i>	<i>VERDADERO</i>	<i>FALSO</i>	<i>Total</i>
0,0	100,00	0,00	57,85
0,05	100,00	0,00	57,85
0,1	99,50	7,59	60,76
0,15	99,50	13,10	63,08
0,2	98,49	16,55	63,95
0,25	97,49	21,38	65,41
0,3	97,49	24,83	66,86
0,35	95,48	25,52	65,99
0,4	92,46	37,93	69,48
0,45	89,45	42,07	69,48
0,5	84,42	49,66	69,77

0,55	78,39	55,17	68,60
0,6	74,37	60,00	68,31
0,65	66,83	62,76	65,12
0,7	49,75	73,10	59,59
0,75	5,03	97,93	44,19
0,8	0,50	99,31	42,15
0,85	0,00	100,00	42,15
0,9	0,00	100,00	42,15
0,95	0,00	100,00	42,15
1,0	0,00	100,00	42,15

**Figura 6. Capacidad de Predicción del Modelo para IL**



En la tabla y el gráfico se observa la capacidad de predicción del modelo 1. Si el valor predicho es más grande que el punto de cohorte, la respuesta se predice como VERDADERA. Si el valor predicho es inferior a o igual al punto de cohorte, la respuesta se predice para ser FALSO. La tabla muestra el porcentaje de datos observados predichos correctamente a diferentes puntos de cohorte. En nuestro caso con el punto de cohorte utilizado 0,5, el 84,4% de todas las respuestas VERDADERAS se predijeron correctamente, mientras que el 49,66% de todas las respuestas FALSAS se predijeron correctamente, para un total de 69,77% casos clasificados correctamente.

**Tabla 100. Matriz de correlación para los coeficientes estimados**

	Cons	B6	B8	B9	M7	M9	S7	L7	L9	L10
Cons	1									
B6	-0,089	1								
B8	-0,385	-0,104	1							
B9	-0,317	0,078	0,098	1						
M7	-0,052	-0,17	-0,046	-0,003	1					
M9	-0,158	-0,054	-0,158	0,023	-0,246	1				
S7	-0,022	-0,288	-0,137	-0,043	-0,27	-0,083	1			
L7	0,126	-0,224	-0,022	-0,011	-0,033	-0,032	-0,177	1		
L9	-0,038	-0,093	0,026	-0,144	-0,011	-0,226	0,001	-0,067	1	
L10	-0,492	4E-04	-0,021	-0,116	-0,039	0,015	-0,043	-0,08	-0,41	1

Esta tabla muestra las correlaciones estimadas entre los coeficientes en el modelo ajustado. Estas correlaciones pueden utilizarse para detectar la presencia de serias multicorrelaciones, es decir, correlación entre las variables pronosticadas. En este caso, no hay ninguna correlación con valores absolutos superiores a 0.5.

## INTERPRETACION DEL MODELO

Como se puede observar en el estudio anterior, el modelo 4 presenta un adecuado ajuste, las variables que ayudan a predecir si un estudiante de grado décimo del Colegio Javeriano estará por encima o debajo del nivel muy superior en las pruebas de estado ICFES en la asignatura de lenguaje son B6, B8, B9, M7,

M9,S7,L7,L9 y L10 , codificadas 0 si el estudiante tiene sobresaliente o excelente y 1 si tiene insuficiente o aprobado.

A continuación se indica el valor de las probabilidades que tienen cada variable de estudio en el modelo.

**Tabla 101. Variables en el modelo**

	B	Exp(B)
B6	-0,2770951	0,75798241
B8	0,49050326	1,6331379
B9	1,14794261	3,15170196
M7	0,32369701	1,38222844
M9	0,31161676	1,36563122
S7	-0,29567421	0,74402979
L7	0,45351981	1,57384207
L9	0,5032154	1,6540311
L10	0,88196889	2,41565118
Constante	-2,44491618	0,0867334

De lo anterior se puede decir que el valor de  $-0,2770951$  que se asocia a la variable B6, incrementa el 75,7% la posibilidad de estar por debajo del nivel muy superior en el área de matemáticas, puesto que  $e^{-0,2770951} = 0,757$

De igual manera se pueden leer todos los coeficientes que acompañan a cada variable de estudio.

Como se puede observar, se evidencia la importancia de las asignaturas B9 y L10 las cuales realizan un aporte significativamente alto para el cálculo de esta probabilidad en el área de castellano, mientras que el aporte de las asignaturas B6 y S9 es menor. Las otras asignaturas realizan un aporte alto al modelo de manera similar.

El modelo seleccionado (modelo 4) ayudará a detectar a aquellos estudiantes de grado décimo que requieren de un refuerzo académico antes de presentar la prueba ICFES. A continuación se indica la probabilidad de que cada estudiante de grado décimo pueda estar por debajo del nivel muy superior. (Tabla 102).

**Tabla 102. Casos posibles que necesitan refuerzo académico en lenguaje antes de las pruebas ICFES**

Caso	B6	B8	GO	M7	MO	L7	LO	LIC	S7		P(y1)
1	0	1	1	1	1	1	1	1	0	1,7	0,8
2	0	0	1	0	1	0	1	1	0	0,4	0,6
3	1	1	0	1	1	1	0	0	0	-1	0,2
4	0	0	1	1	1	0	1	1	0	0,7	0,7
5	1	1	1	1	1	1	1	1	0	1,4	0,8
6	0	0	0	1	0	0	0	0	0	-2	0,1
7	0	0	0	1	0	1	0	0	0	-2	0,2
8	1	1	1	1	1	1	1	1	0	1,4	0,8
9	0	0	0	0	0	0	0	0	0	-2	0,1
10	1	1	1	1	1	1	1	1	0	1,4	0,8
11	0	0	0	0	0	0	0	0	0	-2	0,1
12	0	1	1	1	1	1	1	1	0	1,7	0,8
13	0	1	0	1	1	1	0	0	0	-1	0,3
14	0	0	0	0	1	0	0	0	0	-2	0,1
15	0	0	0	0	0	0	0	0	0	-2	0,1
16	0	0	0	0	0	0	0	0	0	-2	0,1
17	1	1	0	1	0	1	0	0	0	-1	0,2
18	1	1	0	1	1	1	0	0	0	-1	0,2
19	0	0	0	0	0	0	0	0	0	-2	0,1
20	0	0	0	0	0	0	0	0	0	-2	0,1
21	0	0	0	0	0	0	0	0	0	-2	0,1
22	0	0	0	0	0	0	0	0	0	-2	0,1
23	0	0	0	0	0	0	0	0	0	-2	0,1
24	0	0	0	0	0	0	0	0	0	-2	0,1

25	1	1	1	1	1	1	1	1	0	1,4	0,8
26	0	1	1	1	1	1	1	1	0	1,7	0,8
27	0	0	0	0	0	0	0	0	0	-2	0,1
28	0	0	0	1	1	0	0	0	0	-2	0,1
29	0	0	0	0	0	0	0	0	0	-2	0,1
30	1	1	1	1	1	1	1	1	0	1,4	0,8
31	1	1	1	1	1	1	1	1	0	1,4	0,8
32	0	1	1	1	1	1	1	1	0	1,7	0,8
33	0	0	1	1	1	0	0	0	0	-1	0,3
34	0	0	0	1	0	0	0	0	0	-2	0,1
35	0	0	0	1	0	0	0	0	0	-2	0,1
36	1	1	1	1	1	1	1	1	0	1,4	0,8
37	1	1	1	1	1	1	1	1	0	1,4	0,8
38	1	1	1	1	1	1	1	1	0	1,4	0,8
39	0	0	1	0	1	0	1	1	0	0,4	0,6
40	0	0	0	0	0	0	0	0	0	-2	0,1
41	0	0	1	0	1	0	1	1	0	0,4	0,6
42	1	1	1	1	1	1	1	1	0	1,4	0,8
43	1	1	1	1	1	1	1	1	0	1,4	0,8
44	1	1	1	1	1	1	1	1	0	1,4	0,8
45	1	0	0	1	1	1	0	0	0	-2	0,2
46	0	0	1	0	1	0	1	1	0	0,4	0,6
47	0	1	1	1	1	1	1	1	0	1,7	0,8
48	0	0	0	0	0	0	0	0	0	-2	0,1
49	0	0	1	1	1	0	0	0	0	-1	0,3
50	1	1	1	1	1	1	1	1	0	1,4	0,8
51	0	0	1	0	1	0	1	1	0	0,4	0,6
52	0	0	0	0	0	0	0	0	0	-2	0,1
53	1	1	1	1	1	1	1	1	0	1,4	0,8
54	1	1	1	1	1	1	1	1	0	1,4	0,8
55	1	1	1	1	1	1	1	1	0	1,4	0,8
56	1	1	1	1	1	1	1	1	0	1,4	0,8
57	1	1	1	1	1	1	1	1	1	1,1	0,7
58	0	1	1	1	1	1	1	1	1	1,4	0,8
59	0	0	0	0	0	0	0	0	0	-2	0,1

60	0	0	0	1	1	0	0	0	0	-2	0,1
61	0	0	0	0	0	0	0	0	0	-2	0,1
62	1	1	1	1	1	1	1	1	0	1,4	0,8
63	1	1	1	1	1	1	1	1	0	1,4	0,8
64	0	1	1	1	1	1	1	1	0	1,7	0,8
65	0	0	1	1	1	0	0	0	0	-1	0,3
66	0	0	0	1	0	0	0	0	0	-2	0,1
67	0	0	0	1	0	0	0	0	0	-2	0,1
68	1	1	1	1	1	1	1	1	0	1,4	0,8
69	1	1	1	1	1	1	1	1	0	1,4	0,8
70	1	1	1	1	1	1	1	1	0	1,4	0,8
71	0	0	1	0	1	0	1	1	0	0,4	0,6
72	0	0	0	0	0	0	0	0	0	-2	0,1
73	0	0	1	0	1	0	1	1	0	0,4	0,6
74	1	1	1	1	1	1	1	1	0	1,4	0,8
75	1	1	1	1	1	1	1	1	0	1,4	0,8
76	1	1	1	1	1	1	1	1	0	1,4	0,8
77	1	0	0	1	1	1	0	0	0	-2	0,2
78	0	0	1	0	1	0	1	1	0	0,4	0,6
79	0	1	1	1	1	1	1	1	0	1,7	0,8
80	0	0	0	0	0	0	0	0	0	-2	0,1

## CONCLUSIONES DEL MODELO

El modelo de regresión logística obtenido, esta conformado por las variables B6, B8,B9, M7,M9,S7.L7,L9,L10, estas variables independientes de estudio están dicotomisadas.

El modelo de regresión logística obtenido, nos indica un ajuste adecuado de las variables de estudio y la variable IL.

Las pruebas de bondad **y** ajuste que ilustraron el grado de clasificación del modelo, con las pruebas Chi cuadrado, Hosmer-Lemeshow y el estadístico de

verosimilitud del modelo indicaron un adecuado ajuste de clasificación de los datos observados.

El estadístico de verosimilitud indica un adecuado ajuste del modelo, sin embargo se observa que la de este *valor* al introducir las variables independientes es mínimo, pero se evidencia reducción de este valor al comparar este dato con el que arroja el estadístico cuando realiza el cálculo sin las introducir las variables independientes.

El modelo seleccionado, tiene un 69.8% de casos bien clasificados y el 30.2% corresponde a los casos mal clasificados.

Al realizar el estudio de multicorrelaciones del modelo se observa ausencia de correlaciones entre las variables de estudio.

A pesar de que todas las variables del modelo son significativas, las variables B9 y L10 establecen *un aporte alto al cálculo de* la probabilidad que tiene un estudiante Javeriano de décimo grado, de obtener un puntaje por debajo del nivel muy superior.

El pronóstico que realiza el modelo indica que cuanto mejor es el record académico de los estudiantes la probabilidad de estar por debajo del nivel muy superior en el IL disminuye; no obstante, se evidencian casos de estudiantes que presentan un rendimiento académico adecuado, sin embargo, los resultados en las pruebas de estado en lenguaje, muestra una probabilidad alta de estar por debajo del nivel muy superior. También se observan casos donde algunos estudiantes que tiene un desempeño académico bajo, muestra una probabilidad baja de estar por debajo del nivel muy superior. Estos dos sucesos describen la realidad que año tras año envuelve este sistema de evaluación.

## 5. CONCLUSIONES

En el estudio se presentan seis modelos estadísticos distintos que explican la relación del desempeño académico de los estudiantes de grado décimo y sus resultados ICFES en cada área evaluada.

Los modelos estudiados no presentan problemas de multicorrelaciones.

Las pruebas de bondad y ajuste que ilustraron el grado de clasificación del modelo, con las pruebas Chi cuadrado, Hosmer-Lemeshow y el estadístico de Likelihood del modelo, indican un adecuado ajuste de clasificación de los datos.

Los estadísticos de verosimilitud indican un adecuado ajuste de los modelos, sin embargo se observa en cada uno de ellos, que la disminución de este valor al introducir las variables independientes es mínimo, pero se evidencia reducción de este valor al comparar este dato con el que arroja e) estadístico cuando realiza el calculo sin las introducir las variables independientes.

Los modelos seleccionados, están entre el 65.1% y 69.8% de casos bien clasificados y el 35.1% y 30.2% corresponde a *los casos mal* clasificados.

Los modelos estadísticos que mejor predicen ajuste tienen, son los modelos para el área de Matemáticas y Lenguaje.

## BIBLIOGRAFÍA

Ministerio de Educación (1994). *Ley General de educación*.

Ministerio de Educación. *Estándares para la excelencia educativa*.

Finalidades y alcances del Decreto 230 de febrero 11 de 2002

Colegio San Francisco Javieriano (2003). *Procedimiento, prestación y planificación del servicio*.

Propuesta educativa de la Compañía de Jesús. (2005). *Fundamentos y práctica*.

El proceso del Magis Ignaciano y el mejoramiento continuo.

Montgomery, Douglas C. P (2000). Introduction to linear regression analysis. John Wiley & Sons.

Wiley – Interscience Publication, Second Edition. Applied Logistic Regression. David Hosmer & Satnley Lemeshow.

Ediciones Díaz de Santos S.A, Excursión a la Regresión Logística en Ciencias de la Salud. Luís Carlos Silva Ayçaguer

Fuentes Adicionales:

Instituto Colombiano para el Fomento de la Educación Superior

<http://www.icfes.gov.co>

Colegio San Francisco Javier

<http://www.javeriano.edu.co>