

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICAS
ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



PROYECTO DE TITULACIÓN
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:
MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN GESTIÓN DE LA
CALIDAD Y LA PRODUCTIVIDAD

TEMA:

“ANÁLISIS ESTADÍSTICO DE LAS CAUSAS DE LA NO APROBACIÓN DE LOS
ESTUDIANTES DEL CURSO DE INGRESO A UNA ESCUELA SUPERIOR
MILITAR”

AUTOR:

VERÓNICA LOURDES CAMACHO TORRES

Guayaquil – Ecuador

2018

DEDICATORIA

A mis padres por brindarme su amor y apoyo incondicional y ser mi guía en cada una de las actividades y proyectos que decido emprender.

A mi esposo e hija.

AGRADECIMIENTO

A Dios por darme la sabiduría y fortaleza necesaria para culminar esta etapa de mi vida.

A mi esposo por siempre ofrecerme su apoyo tanto en lo académico como en lo personal.

Al director de este Proyecto de Titulación, por haber sido parte del desarrollo de la presente investigación con sus conocimientos y guía.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestos en este Proyecto de Titulación me corresponde exclusivamente; el patrimonio intelectual del mismo, corresponde exclusivamente a la Facultad de Ciencias Naturales y Matemáticas, Departamento de Matemáticas de la Escuela Superior Politécnica del Litoral.

A handwritten signature in blue ink, appearing to read 'Verónica Lourdes Camacho Torres', written over a horizontal line.

Verónica Lourdes Camacho Torres

TRIBUNAL DE GRADUACIÓN



Mgtr. Mario David Solórzano Carvajal

Director



Mgtr. Alex Moreno Salazar

Presidente



Mgtr. Gina Ochoa Jara

Vocal 1



Mgtr. Wendy Plata Alarcón

Vocal 2

ÍNDICE GENERAL

DEDICATORIA.....	i
AGRADECIMIENTO.....	ii
DECLARACIÓN EXPRESA	iii
TRIBUNAL DE GRADUACIÓN	iv
ÍNDICE GENERAL	v
INDICE DE TABLAS	viii
ÍNDICE DE GRÁFICOS.....	x
RESUMEN.....	xi
Capítulo 1	1
1. GENERALIDADES.....	1
1.1 Introducción.....	1
1.2 Planteamiento del Problema	1
1.3 Objetivo General	2
1.4 Objetivos Específicos	2
1.5 Alcance.....	2
1.6 El Proceso de Admisión a la Escuela Militar	2
1.6.1 Requisitos previos.....	2
1.6.2 Flujo del proceso de admisión	3
1.6.2.1 Etapa 1: Recepción de documentos.....	3
1.6.2.2 Etapa 2: Prueba psicológica	3
1.6.2.3 Etapa 3: Prueba académica	3
1.6.2.4 Etapa 4: Prueba física	4
1.6.2.5 Etapa 5: Visita al aspirante	4
1.6.2.6 Etapa 6: Entrevista final.....	4
1.7 Estado del Arte.....	4
Capítulo 2	7
2. MARCO TEORICO.....	7
2.1 Análisis multivariante.....	7

2.1.1	El modelo lineal: regresión lineal múltiple	7
2.1.1.1	Modelos lineales generalizados: regresión logística binomial	8
2.1.1.2	Herramientas de la calidad	10
2.1.1.3	Diagrama de causa y efecto	10
2.1.1.4	Estratificación	10
Capítulo 3	11
3.	ESTUDIO Y RESULTADOS	11
3.1	Estrategia de Estudio y Análisis	11
3.2	Los Datos, Definición de Variables y Visión Preliminar	12
3.2.1	Los datos	12
3.2.2	Definición de variables	12
3.2.3	Visión preliminar.....	13
3.3	Análisis de las Causas de la No Aprobación de los Aspirantes por Etapas .	14
3.3.1	Diagrama de causa y efecto	14
3.3.2	Estratificación.....	15
3.4	Análisis Descriptivo Bivariante de las Etapas del Proceso	18
3.4.1	Etapa 1: Recepción de documentos	18
3.4.2	Etapa 2: Prueba psicológica	19
3.4.3	Etapa 3: Prueba académica	21
3.4.4	Etapa 4: Prueba física.....	22
3.4.5	Etapa 5: Visita al aspirante	23
3.4.6	Etapa 6: Entrevista final	24
3.5	Modelos Multivariantes de Regresión Logística Binomial	25
3.5.1	Modelo logístico global del proceso	25
3.5.2	Modelos logísticos por cada etapa del proceso	27
3.5.2.1	Etapa 1: Recepción de documentos.....	27
3.5.2.2	Etapa 2: Prueba psicológica	30
3.5.2.3	Etapa 3: Prueba académica	32
3.5.2.4	Etapa 4: Prueba física	35

3.5.2.5	Etapa 5: Visita al aspirante	36
3.5.2.6	Etapa 6: Entrevista final.....	38
3.5.3	Predicción de la probabilidad de ingreso de un aspirante dado	40
3.5.3.1	Probabilidad frecuentista global.....	40
3.5.3.2	Probabilidad frecuentista por etapas	40
3.5.3.3	Probabilidad usando el modelo logístico global.....	41
3.5.3.4	Probabilidad usando los modelos logísticos por etapas	41
3.6	Propuesta de Mejoras y de Temas Específicos para Pruebas Académicas	42
3.6.1	Propuesta de mejoras	42
3.6.2	Propuesta de temas específicos para las pruebas académicas	43
Capítulo 4	44
4.	CONCLUSIONES Y RECOMENDACIONES	44
4.1	Conclusiones.....	44
4.2	Recomendaciones.....	44
Referencias	46
Apéndices	48

INDICE DE TABLAS

Tabla 1 Número y porcentaje de aprobados y no aprobados por etapa	13
Tabla 2 Tabla de frecuencias de los no aprobados en etapa 1	17
Tabla 3 Regresión por pasos para el Modelo Global.....	25
Tabla 4 Betas y p-values del modelo global con menor AIC	26
Tabla 5 Betas estimados valores p del modelo global.....	26
Tabla 6 Odds-ratios en el modelo global	27
Tabla 7 Matriz de confusión - modelo global	27
Tabla 8 Regresión por pasos para el modelo de Recepción de Documentos	28
Tabla 9 Betas y p-values del modelo1 con menor AIC	28
Tabla 10 Betas estimados y valores p del modelo en la etapa 1	29
Tabla 11 Odds-ratios en modelo etapa 1	29
Tabla 12 Matriz de confusión - modelo etapa 1	30
Tabla 13 Regresión por pasos para el modelo de Prueba Psicológica	30
Tabla 14 Betas estimados y valores p del modelo en la etapa 2.....	31
Tabla 15 Odds-ratios del modelo en etapa 2.....	31
Tabla 16 Matriz de confusión - modelo etapa 2.....	32
Tabla 17 Regresión por pasos para el modelo de Prueba Académica.....	33
Tabla 18 Betas y p-values en modelo 3 con menor AIC.....	33
Tabla 19 Betas estimados y valores p del modelo en etapa 3	33
Tabla 20 Odds-ratios del modelo en etapa 3.....	34
Tabla 21 Matriz de confusión - modelo etapa 3.....	34
Tabla 22 Regresión por pasos en el modelo de Prueba Física	35
Tabla 23 Betas estimados y valores p del modelo en etapa 4	35
Tabla 24 Odds-ratios del modelo en etapa 4.....	36
Tabla 25 Matriz de confusión - modelo etapa 4.....	36
Tabla 26 Regresión por pasos en el modelo de Visita al Aspirante	37

Tabla 27 Betas estimados y valores p del modelo en etapa 5	37
Tabla 28 Odds-ratios del modelo en etapa 5.....	38
Tabla 29 Matriz de confusión - modelo etapa 5.....	38
Tabla 30 Regresión por pasos en el modelo de Entrevista Final	39
Tabla 31 Beta estimado y valor p del modelo.....	39
Tabla 32 Matriz de confusión - Modelo 6.....	40
Tabla 33 Probabilidad frecuentista por etapas	41
Tabla 34 Probabilidades con modelos logísticos.....	42

ÍNDICE DE GRÁFICOS

Gráfico 1 Porcentaje de aprobados y no aprobados por etapa	14
Gráfico 2 Diagrama de causa y efecto para los no admitidos	15
Gráfico 3 Diagrama de Pareto para las causas principales	16
Gráfico 4 Diagrama de Pareto en recepción de documentos	17
Gráfico 5 Porcentaje de aspirantes en Recepción de Documentos	18
Gráfico 6 Porcentaje de aspirantes en Prueba Psicológica	20
Gráfico 7 Porcentaje de aspirantes en Prueba Académica	21
Gráfico 8 Porcentaje de aspirantes en Prueba Física	22
Gráfico 9 Porcentaje de aspirantes en Visita al Aspirante	23
Gráfico 10 Porcentaje de aspirantes en Entrevista Final	24

RESUMEN

En el Ecuador, cada año se presentan convocatorias para los aspirantes que deseen ingresar a las distintas ramas de una Escuela Superior Militar, y son muchos los jóvenes que se inscriben; sin embargo, son pocos los que logran el objetivo; ya sea por no cumplir con los requisitos de documentación, o no aprobar las distintas pruebas a las que son sometidos los aspirantes.

En este trabajo, se ha realizado un análisis estadístico de las causas que más influyen en la no aprobación de las distintas etapas del proceso de selección a una Escuela Militar, y proponer de manera general, mejoras al sistema de selección. En el capítulo 1 se presenta el problema, los objetivos y el alcance de la investigación; el capítulo 2 trata el marco de referencia para el estudio, y se describe el análisis bivariante, el modelo logístico binomial y el flujo del proceso de admisión. En el capítulo 3 se aplican las técnicas descritas anteriormente a los datos obtenidos y como aplicación de los modelos logísticos, se calcula la probabilidad de ingreso de un aspirante de acuerdo a factores tales como el sexo, edad, región de donde proviene, nivel socioeconómico y tipo de colegio donde estudió. También se proponen algunas mejoras al proceso de selección y temas específicos para algunas pruebas académicas.

Finalmente, en el capítulo 4, constan las conclusiones y recomendaciones del estudio realizado. Los gráficos se realizaron con R y Excel, y la obtención de los modelos logísticos se hicieron utilizando el lenguaje R.

Capítulo 1

1. GENERALIDADES

1.1 Introducción

En el Ecuador, muchos jóvenes recién graduados y otros profesionales, aspiran a ingresar a una Escuela Militar para seguir una carrera como grumete u oficial. Los aspirantes son de sexo masculino y femenino, de todas las regiones del país, de edades entre 18 y 22 años, de todos los niveles socioeconómicos, y provienen tanto de colegios públicos como privados.

Los aspirantes deben cumplir con todos los requisitos exigidos por la institución en cada etapa del proceso de admisión, tanto en la documentación requerida, características antropométricas, pruebas psicológicas, académicas y físicas; sin embargo, la mayoría de quienes inician el proceso no logran culminarlo con éxito. Los cupos disponibles varían cada año de acuerdo a los requerimientos de la institución.

1.2 Planteamiento del Problema

La admisión a los procesos de reclutamiento es de elección libre y voluntaria para los aspirantes que deseen ingresar a las escuelas de formación militar del Ecuador. En el año 2017 cerca de 9.756 aspirantes se han presentado con el fin de ingresar a una Escuela Superior Militar, la institución, realizará una primera clasificación, seleccionando 600 aspirantes de acuerdo con el rendimiento académico en su etapa estudiantil secundaria. Una vez hayan sido seleccionados se les administrará una prueba de carácter académico/físico, si llegan al puntaje requerido esto será aprobado por la junta académica que está conformada por la máxima autoridad que es el comandante General, dos oficiales, y una psicóloga,

Se pretende hacer un estudio estadístico de los aspirantes que no ingresan a las Escuela Superior Militar, por diversas razones entre ellas, la falta de documentación, las características antropométricas del aspirante (estatura mínima 1.60 mujeres, 1.65 varones, peso corporal ideal).

Debido en muchas ocasiones a la falta de conocimiento, el aspirante no logra completar la documentación del aspirante (cédula, acta de grado, certificado de no

haber sido dado de baja de ninguna institución militar, etc.) o no reúnen los requisitos necesarios para la aprobación del ingreso a una Escuela Superior Militar.

Por lo tanto, se plantea determinar qué causas son las que en mayor grado influyen en la no aprobación del curso, y en qué etapa se observa la mayor tasa de reprobación.

1.3 Objetivo General

Analizar las causas de la no aprobación de los estudiantes del curso de ingreso a la Escuela Superior Militar.

1.4 Objetivos Específicos

- Elaborar un análisis estadístico descriptivo de los factores que inciden en la aprobación del curso de ingreso a la Escuela Superior Militar.
- Escoger temas específicos para el desarrollo de las pruebas académicas y pruebas físicas.

1.5 Alcance

Este estudio abarca a las personas mayores de 18 años bachilleres en ciencias, que son aspirantes a ingresar a una Escuela Militar, los profesionales con título de tercer y cuarto nivel, tienen la oportunidad de ingresar como Oficiales Especialistas en el Ecuador. Y se considera información pertinente del año 2017.

1.6 El Proceso de Admisión a la Escuela Militar

El proceso de admisión a la Escuela Militar consta de seis etapas que son: recepción de documentos, prueba psicológica, prueba académica, prueba física, visita al aspirante y entrevista final. Para seguir este proceso, el aspirante debe cumplir ciertos requisitos previos que se detallan en el siguiente numeral.

1.6.1 Requisitos previos

Los requisitos previos para postular como aspirante a la Escuela Militar son los siguientes:

- Ser ecuatoriano por nacimiento.
- Ser bachiller en ciencias generales / fima / bgu (orientación físico matemático) /bachillerato internacional.
- Estatura descalza 160cm mujeres / 165cm hombres.
- Evaluación ene/ser bachiller - (superior a 700)

- Título de bachiller registrado en la página del ministerio de educación
- Mayor de edad hasta los 22 años 0 meses (calculados a la fecha prevista al ingreso de las escuelas de formación –enero del siguiente año).
- Estado civil soltero.
- Acreditar condiciones psicológicas, académicas, médicas y físicas que le permitan someterse al régimen de formación militar.
- No haber sido dado de baja del servicio activo de acuerdo a la reglamentación vigente de las escuelas de formación de las FF.AA., policía nacional y comisión de tránsito del Ecuador.
- No registrar antecedentes penales, y/o procesos judiciales pendientes.

1.6.2 Flujo del proceso de admisión

1.6.2.1 Etapa 1: Recepción de documentos

Para entregar la carpeta el aspirante debe acercarse al Centro de Reclutamiento de la Escuela Superior Militar, con todos los requisitos solicitados anteriormente debe vestir con traje formal y presentar su cédula de identidad, cabe indicar que la entrega de carpeta es un trámite personal y debe asistir únicamente el aspirante.

Se entregará un código que será el comprobante de entrega de carpeta, el mismo que lo identifica como aspirante del proceso y en caso de que la carpeta sea aprobada, deberá presentar el comprobante en la siguiente fase.

1.6.2.2 Etapa 2: Prueba psicológica

El aspirante debe presentarse 30 minutos antes con traje formal a rendir el examen psicológico, que contiene la prueba abstracta que dura 30 minutos y la de personalidad 15 minutos, el aspirante tiene que poseer la cédula de identidad y su código de aspirante. El aspirante deberá estar pendiente de la página web y su cuenta personal, con el fin de verificar si aprobó las pruebas psicotécnicas y estar pendiente de la fecha, hora y lugar para la próxima fase, donde se darán instrucciones específicas al aspirante.

1.6.2.3 Etapa 3: Prueba académica

El aspirante debe presentarse de vestimenta formal a rendir los exámenes académicos: Matemáticas, Física, Química, Realidad Nacional, y Lenguaje, para los cuales se asignan 5 días en los que debe rendir las pruebas. Cada materia tiene una

duración de 2 horas. El aspirante debe estar pendiente de la página web y su cuenta personal, a fin de que verifique si aprobó el examen académico y estar pendiente de la fecha, hora y lugar para la próxima fase, donde se darán instrucciones específicas al aspirante.

1.6.2.4 Etapa 4: Prueba física

El aspirante debe presentarse a la hora signada en traje de deporte (calentador, pantaloneta, short, licra y camiseta), traer su cédula de identidad y su código. El primer día se receptorán las pruebas de campo: trote 3200 m en un tiempo de 12:30 segundos para varones y para las mujeres 15:30 segundos, los abdominales para varones es de 50 en 1:30 segundos y para mujeres 40 en 1:30 segundos y flexiones de codo para hombres 50 en 1:00 minuto y para las mujeres 25 en 1:00 minuto. El segundo día las pruebas de piscina: natación 200 m en 4:30 segundos para varones y 5:30 segundos para mujeres y salto de decisión de 10m tanto para hombres como mujeres.

1.6.2.5 Etapa 5: Visita al aspirante

Una vez aprobada la prueba física, se le realizará una visita al aspirante en su domicilio, por parte de dos oficiales asignados, para constatar la veracidad de la información entregada por el aspirante.

1.6.2.6 Etapa 6: Entrevista final

En la entrevista estará el director de la Escuela Militar, el Jefe de Reclutamiento, el psicólogo, los padres de familia y el aspirante, en la cual las autoridades le realizarán 5 preguntas (2 el director, 2 el jefe de reclutamiento, 1 el psicólogo) en un tiempo de 2 minutos por pregunta.

1.7 Estado del Arte

(Tinisaray, 2015). En la Universidad de Sevilla. Departamento de Estadística e Investigación Operativa ha realizado la Construcción de un modelo para determinar el rendimiento académico de los estudiantes basado en learning analytics (análisis del aprendizaje), mediante el uso de técnicas multivariantes. Recoge una revisión de las técnicas estadísticas aplicadas en la investigación educativa. Se introduce la

metodología de estudio, selección de casos y variables que permiten justificar la elección de los modelos multivariantes.

Se obtienen los resultados del modelo empírico multinivel estimando el modelo jerárquico con 2 y 3 niveles: estudiante (nivel inferior), aula (nivel intermedio) y escuela (nivel superior), utilizando el software Stata/SE 12.0. También se desarrolla un modelo logístico bivariante binario y un modelo logístico bivariante ordinal, los parámetros de los modelos se estiman usando el software R con la plataforma RStudio. Que presentan los resultados, así respecto al modelo multinivel, el de mejor ajuste para el rendimiento académico incluye: Tres covariables del Nivel 2: tasa de repetidores, ciclo y tipo de docente; Ocho variables del Nivel 1: edad, rinde supletorio, repite materia, participa en chat, participa en foro, participa en video colaboración, N° comentarios, N° accesos al LMS; Cuatro interacciones multinivel; La varianza de cinco pendientes del Nivel 1. Los modelos logísticos bivariantes permiten confirmar que las covariables más relevantes son la edad de ingreso a la universidad y la participación activa en línea. Este estudio, al identificar la influencia que ejercen sobre el rendimiento académico las variables consideradas, permite a las instituciones educativas mejorar la focalización de las intervenciones y los servicios de apoyo a estudiantes con mayor riesgo de fracaso académico.

(Torres, 2014). Este artículo sintetiza un estudio realizado en el 2013, con el objetivo de predecir la reprobación en cursos de matemática por parte de los estudiantes de las carreras de Farmacia, Matemática, Ciencias Actuariales, Física y Meteorología de la Universidad de Costa Rica (UCR). Mediante la técnica estadística de regresión logística, aplicada a la cohorte del 2010, se pronosticó la reprobación de los estudiantes de estas carreras en uno de sus cursos introductorios de matemática (Cálculo I para Física y Meteorología, Principios de matemática para Matemática y Ciencias actuariales y ecuaciones diferenciales aplicadas para Farmacia). Para estos modelos se utilizaron, como variables predictivas, los componentes del promedio de admisión a la UCR, el género del estudiante y el porcentaje de ítems correctos en la Prueba de Habilidades Cuantitativas (HC). La variable con mayor importancia en todos los modelos fue HC y los modelos con mejores porcentajes de clasificación

correcta de casos fueron las regresiones logísticas. Para los casos utilizados en la estimación de los modelos de Física-Meteorología, Farmacia y Matemática-Ciencias Actuariales se obtuvo un 89.8%, 73.6% y 93.9% de clasificación correcta, respectivamente.

Capítulo 2

2. MARCO TEORICO

En esta sección se hará una descripción general de los conceptos y definiciones que se van a utilizar en el desarrollo de este estudio.

2.1 Análisis multivariante

Cuando se pretende analizar un fenómeno en el que se consideran más de dos factores simultáneamente se deben usar técnicas multivariantes. Existen muchos métodos multivariantes dependiendo del caso que se trate de estudiar, pero una forma de clasificar a estos métodos es atendiendo al tipo de variables del que se dispone, los métodos de interdependencia consideran a todas las variables o factores con igual importancia; en cambio, en los métodos de la dependencia se destaca uno o más factores como dependientes del resto de factores considerados como independientes (Pérez, 2004), el objetivo principal de estos métodos de la dependencia es encontrar una ecuación que relacione a la(s) variable(s) dependiente(s) con las variables independientes.

Dentro de los métodos multivariantes de la dependencia se encuentran los modelos lineales y los modelos lineales generalizados y, dentro de estos, la regresión logística binomial.

2.1.1 El modelo lineal: regresión lineal múltiple

Un modelo lineal (en el contexto de este proyecto) es una ecuación que relaciona una variable dependiente con una o varias variables independientes de una forma particular. La regresión lineal múltiple es un modelo lineal cuya expresión funcional es la siguiente (Rencher, 2002):

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

En esta expresión y es la variable dependiente o explicada, X_1, X_2, \dots, X_k son las variables explicativas, independientes o exógenas, $\beta_0, \beta_1, \dots, \beta_k$ son los coeficientes de las variables, y ε es el error aleatorio. En la práctica, se dispone de una tabla de datos para n individuos de los cuales se ha obtenido valores para la variable

dependiente y las independientes, y el objetivo es encontrar estimaciones de los coeficientes β y usar la ecuación estimada para predecir el valor de la variable dependiente dados ciertos valores de las variables explicativas.

Sin embargo, deben cumplirse ciertas hipótesis sobre el modelo, los coeficientes, los residuos, y debe validarse el modelo. Para esto, existe una serie de pruebas que permiten establecer si se cumplen las hipótesis o no. En caso que no se cumpla una o varias de las hipótesis sobre las que se basa el modelo de regresión múltiple, debe seguirse lo que la teoría recomienda para solucionar el problema. Existe una amplia literatura respecto a los problemas que se presentan al tratar de obtener una ecuación de regresión múltiple: Multicolinealidad, Heterocedasticidad, Autocorrelación, falta de especificación; además debe cumplirse la hipótesis de normalidad de los residuos del modelo, y el modelo debe explicar en un alto grado la variabilidad de la variable dependiente. Cabe indicar que las variables independientes pueden ser de escala, nominales u ordinales y la dependiente, de escala.

2.1.1.1 Modelos lineales generalizados: regresión logística binomial

Un modelo lineal generalizado tiene la siguiente expresión funcional (Agresti, 2002):

$$g(\mu_i) = \sum_j \beta_j X_{ij} \quad i = 1, 2, \dots, n$$

O de forma explícita:

$$g(\mu) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j$$

Este modelo es similar al de regresión múltiple, pero aquí se identifican específicamente sus componentes: la expresión a la derecha de la igualdad se denomina *predictor lineal*, en el cual las X son las j variables explicativas del modelo y los β son los coeficientes de las variables explicativas, $\mu = E(Y)$ es el valor esperado de la variable aleatoria Y que es la variable dependiente (puede ser binomial, Poisson, entre otras), y la función g que se denomina *función link* que realiza una transformación de μ .

La utilidad de los modelos lineales generalizados radica en que se pueden modelar variables respuesta tipo binomial o Poisson. En el caso que la variable dependiente es binomial, es decir éxito o fracaso, y p la probabilidad de éxito, es decir, $p = P(Y=1)$, el *modelo de regresión logística binomial* tiene la siguiente expresión:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j$$

En esta ecuación, el lado derecho es el predictor lineal, y el izquierdo se conoce como el logaritmo del odds-ratio, también se la puede expresar como $\text{logit}(p)$, es decir, la ecuación puede ser escrita como:

$$\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j$$

De esta ecuación se puede despejar p , la probabilidad de éxito, y la expresión resultante es:

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j}}$$

Cabe indicar que la probabilidad que devuelve el modelo está condicionada a los valores de las variables explicativas.

Este modelo es más flexible en cuanto a los supuestos, comparado con la regresión múltiple, pero para determinar la calidad del modelo deben hacerse ciertas pruebas y calcular algunos indicadores. El modelo nulo es aquel que no contiene ninguna variable explicativa y solo tiene la constante como parámetro β ; el modelo saturado es aquel que tiene tantos parámetros β como número de observaciones tenga la muestra.

Una forma de medir la calidad del modelo obtenido es comparándolo con el modelo nulo y con el modelo saturado y determinar qué tan diferentes son. Si el modelo obtenido es significativamente diferente del modelo nulo y no es significativamente diferente del modelo saturado, entonces se tiene un buen modelo. Esta comparación se realiza en base a un estadístico llamado *deviance* que sigue

una distribución chi-cuadrado (Venables & Ripley, 2002). La forma de hacer estas comparaciones usando el lenguaje R se muestra en los apéndices.

Una forma de medir la proporción de deviance explicada por el modelo es:

$$\frac{\text{Deviance}}{\text{Null Deviance}}$$

Este coeficiente es un valor entre 0 y 1, por lo tanto puede interpretarse como un porcentaje, mientras más cercano a 1 el modelo explica mejor la variabilidad de la respuesta.

Con el modelo obtenido se puede realizar predicciones para determinar si un caso es éxito o fracaso. La proporción de aciertos es una medida relativa del buen ajuste del modelo, pero una mejor apreciación se tiene construyendo la tabla de confusión, que es una tabla que cruza los éxitos y fracasos en los datos, con los éxitos y fracasos predichos por el modelo.

2.1.1.2 Herramientas de la calidad

Un sistema de calidad utiliza ciertas herramientas para gestionar el mejoramiento de la calidad, entre estas herramientas están el diagrama de causa y efecto o diagrama de Ishikawa, y la estratificación.

2.1.1.3 Diagrama de causa y efecto

Un diagrama de causa y efecto o diagrama de espina de pescado, es un diagrama que muestra en un extremo el efecto de un problema de calidad y en su parte superior e inferior las causas posibles de esos problemas y las causas de segundo nivel para las causas de primer nivel.

Este diagrama permite tener una visión global del problema y marca una ruta a seguir mediante la identificación de las causas principales y las de segundo nivel.

2.1.1.4 Estratificación

La estratificación es un procedimiento para clasificar los datos de una variable o factor considerando las categorías de otro factor. Es un análisis bivariante para establecer si las categorías de un factor influyen en las categorías del otro factor y suele representarse por medio de diagramas de barras. Es útil para identificar causas de tercer nivel de un problema específico.

Capítulo 3

3. ESTUDIO Y RESULTADOS

3.1 Estrategia de Estudio y Análisis

Teniendo presente que el aspirante debe aprobar todas las seis etapas del proceso para poder ser admitido en la Escuela Superior Militar, la estrategia a utilizar es la siguiente:

Realizar un análisis descriptivo que incluye una tabla de frecuencias y un diagrama de barras del número de aspirantes que aprueban y no aprueban cada etapa del proceso.

Realizar un análisis de las causas más importantes de la no aprobación del proceso de admisión usando un diagrama de causa y efecto y la estratificación.

Realizar un análisis descriptivo bivariante, de cada etapa del proceso, mediante tablas de contingencia y diagrama de barras, entre el factor que representa la aprobación o no aprobación de cada etapa y los factores: sexo, edad, región del país de donde proviene el aspirante, nivel socioeconómico y el tipo de colegio de donde proviene.

Construir un modelo de regresión logística global sin considerar las etapas intermedias del proceso de admisión, para predecir la probabilidad de ingreso de un aspirante a la Escuela Militar, en función de los factores: sexo, edad, región del país de donde proviene el aspirante, nivel socioeconómico y el tipo de colegio de donde proviene. **Se obtendrá el modelo más adecuado mediante la regresión por pasos usando la instrucción step () del software R.**

Construir un modelo de regresión logística para cada etapa del proceso, que permita predecir la probabilidad de aprobación de un aspirante en dicha etapa, en función de los factores: sexo, edad, región del país de donde proviene el aspirante, nivel socioeconómico y el tipo de colegio de donde proviene. Se obtendrá el modelo más adecuado en cada etapa mediante la regresión por pasos usando la instrucción step () de R.

Realizar la predicción de la probabilidad de ingreso a la Escuela Militar de un aspirante dado, con ciertos niveles de los factores: sexo, edad, región del país de

donde proviene el aspirante, nivel socioeconómico y el tipo de colegio donde estudió, de cuatro formas diferentes:

- Usando el concepto de probabilidad frecuentista global del proceso.
- Usando la probabilidad frecuentista condicional por cada etapa del proceso.
- Usando la probabilidad condicional mediante el modelo logístico global, tomando en cuenta los niveles de los factores del aspirante

Usando la probabilidad condicional mediante los modelos logísticos de cada etapa, tomando en cuenta los niveles de los factores del aspirante.

Con base en los resultados obtenidos, proponer mejoras a los procedimientos en las etapas del proceso de admisión, y temas específicos para las pruebas académicas sobre aspectos determinados del estudio.

3.2 Los Datos, Definición de Variables y Visión Preliminar

En esta sección se utilizará los datos obtenidos de los aspirantes a ingresar a la Escuela Superior Militar, definiendo las variables para cada estudiante.

3.2.1 Los datos

Se dispone de la información en Excel de todos los aspirantes (9.756) a ingresar a la Escuela Militar en las 6 etapas del proceso de admisión 2017, por lo tanto, no es necesario realizar un estudio muestral.

3.2.2 Definición de variables

. Las variables disponibles para cada aspirante son:

código: Número que identifica a cada aspirante

sexo: Sexo del aspirante con los niveles Mujer y Hombre

edad: Edad del estudiante

region: Región de donde proviene el estudiante, con los niveles costa y sierra (este nivel incluye a los aspirantes de la amazonia)

nivel socio: factor ordinal que representa el Nivel socioeconómico del aspirante, con los niveles A, B, C+, C-, D, siendo el nivel A el mejor posicionado socioeconómicamente y el nivel D la última categoría socioeconómica.

colegio: Tipo de colegio de donde proviene, con los niveles Público y Privado

aprueba: este factor indica si el aspirante aprobó o no aprobó la etapa del proceso que se analiza en ese momento; es decir, cada etapa del proceso tiene su propio factor aprueba. Los niveles son: No aprueba y Aprueba.

etapa: Este factor indica en qué etapa del proceso el aspirante quedó fuera, con los niveles Ninguna, Recepción de documentos, Prueba psicológica, Prueba académica, Prueba Física, Visita, Entrevista.

En el apéndice 1 se detalla la codificación en R de la lectura de los datos, la conversión de las variables en factores y se muestra un encabezado de los datos.

3.2.3 Visión preliminar

La Tabla 1 muestra el número de aspirantes que aprueban y no aprueban cada una de las etapas del proceso:

Tabla 0.1

Número y porcentaje de aprobados y no aprobados por etapa

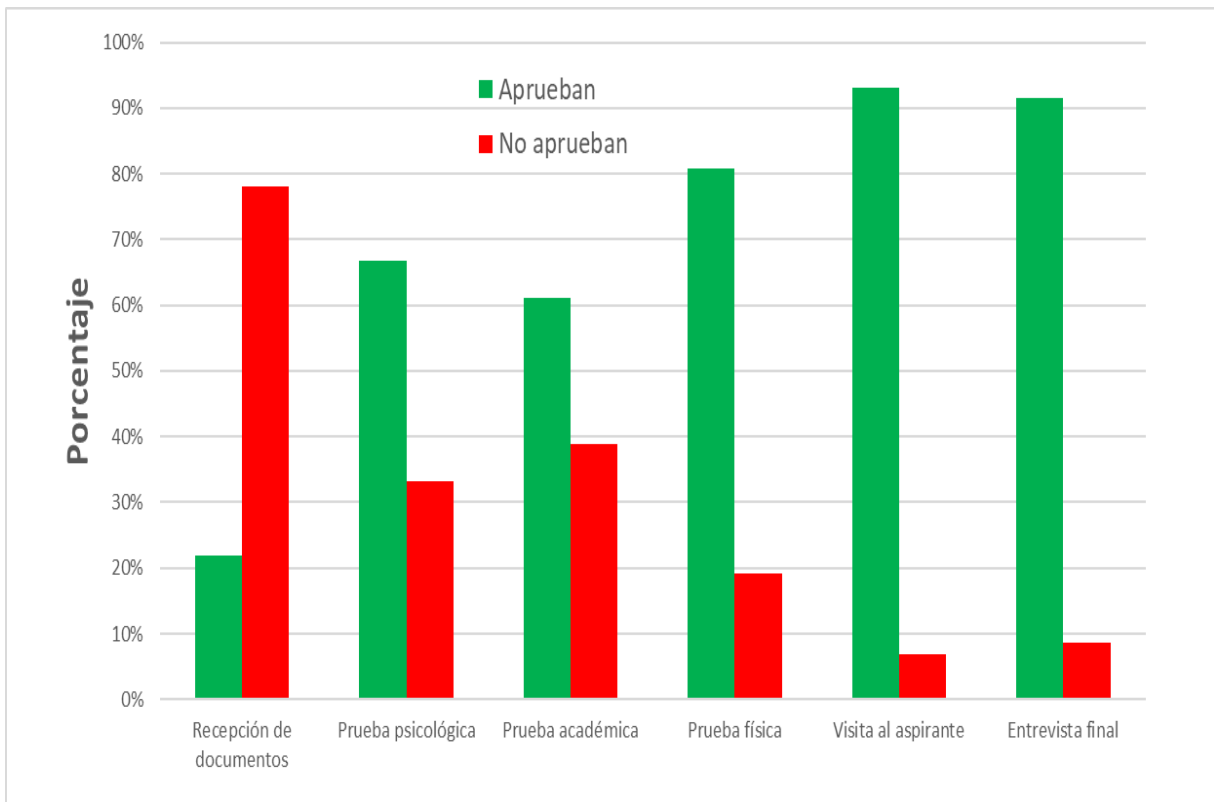
Etapa	Aprueban		No aprueban		Total Aspirantes
	Aspirantes	Porcentaje	Aspirantes	Porcentaje	
Recepción de documentos	2.132	21,9%	7.624	78,1%	9.756
Prueba psicológica	1.425	66,8%	707	33,2%	2.132
Prueba académica	871	61,1%	554	38,9%	1.425
Prueba física	704	80,8%	167	19,2%	871
Visita al aspirante	656	93,2%	48	6,8%	704
Entrevista final	600	91,5%	56	8,5%	656

Fuente: Escuela Militar

Elaborado por: Verónica Camacho Torres

El Gráfico 1 muestra de mejor forma esta información:

Gráfico 1
 Porcentaje de aprobados y no aprobados por etapa



Elaborado por: Verónica Camacho Torres
 Fuente: Escuela Militar

Del Gráfico 1 se puede observar claramente que, en la primera etapa de recepción de documentos, es en donde se presenta la mayor cantidad de aspirantes que no aprueban una etapa (78,1%) y es la única etapa donde la cantidad de reprobados es mayor que la de aprobados

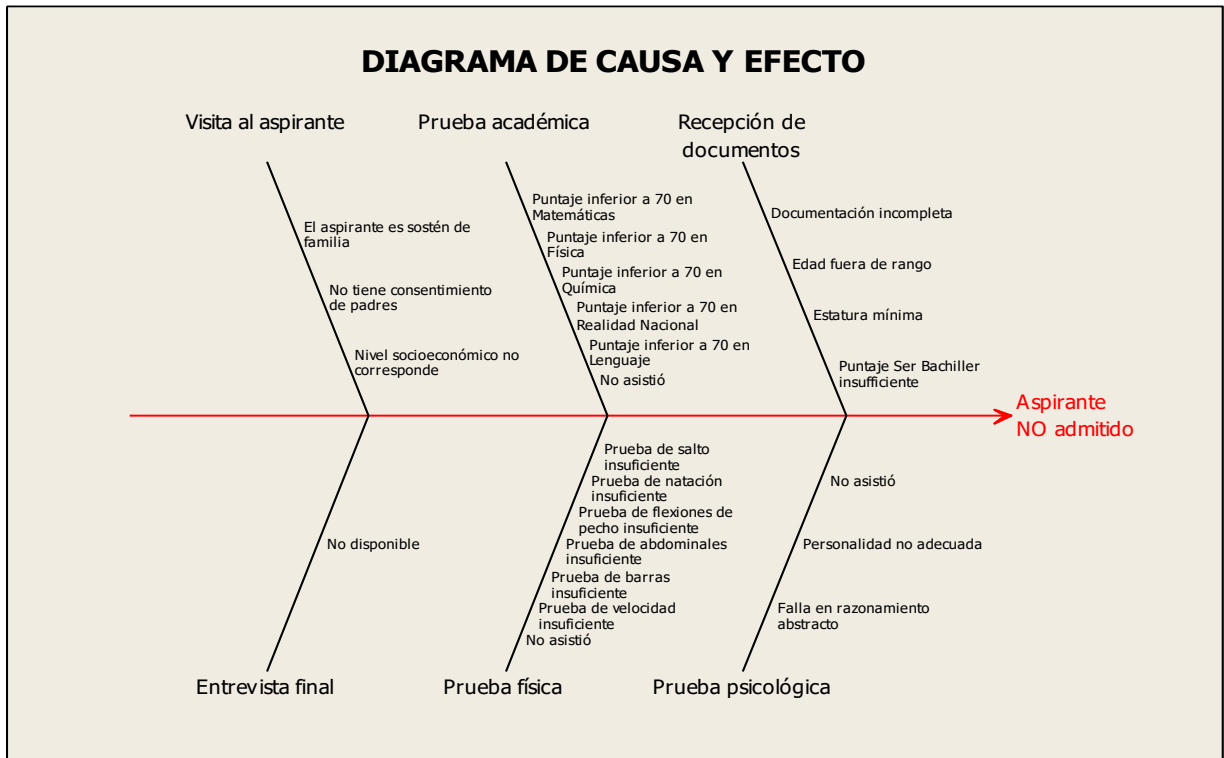
3.3 Análisis de las Causas de la No Aprobación de los Aspirantes por Etapas

Por medio de un diagrama de causa y efecto, diagrama de Pareto se observara las causas principales de los aspirantes no aprobados.

3.3.1 Diagrama de causa y efecto

En el Gráfico 2 se muestra el diagrama de causa y efecto, las causas de primer nivel serían las etapas, y las de segundo nivel son las causas correspondientes a cada etapa del proceso. El código en R se muestra en el apéndice 2.

Gráfico 2
Diagrama de causa y efecto para los no admitidos

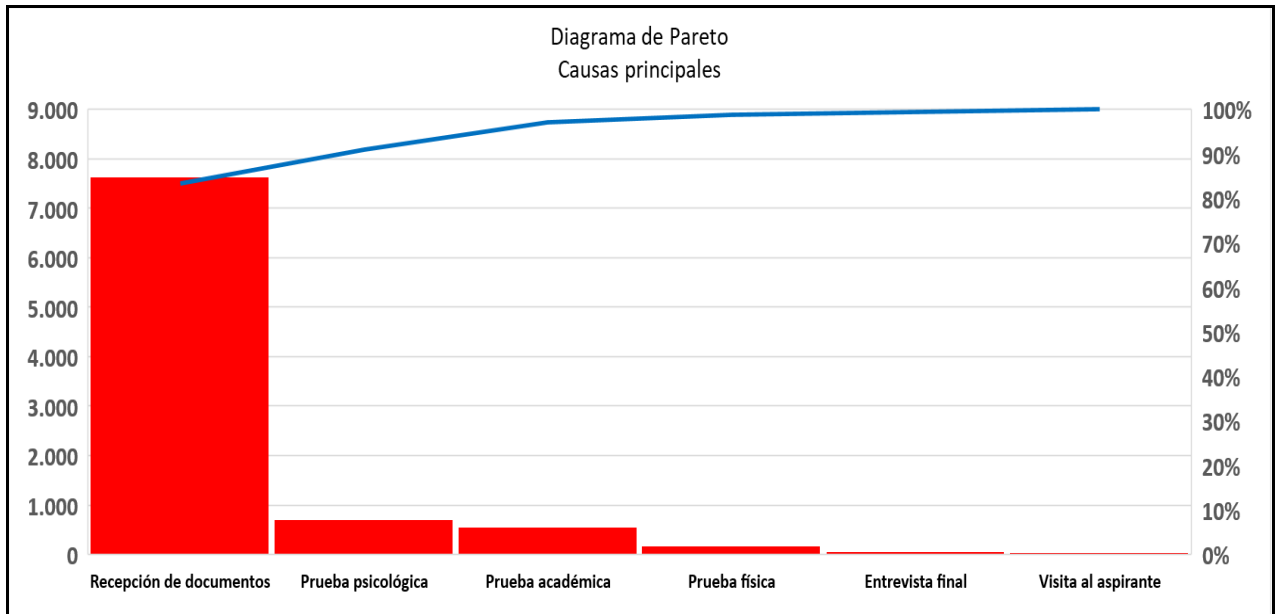


Elaborado por: Verónica Camacho Torres
Fuente: Escuela Militar

3.3.2 Estratificación

En el Gráfico 3 se muestra el diagrama de Pareto para las causas principales o de primer nivel para la no aprobación del aspirante:

Gráfico 3
Diagrama de Pareto para las causas principales



Elaborado por: Verónica Camacho Torres
Fuente: Escuela Militar

Se puede observar que más del 80% de los aspirantes no aprobados se dan en la etapa 1 de recepción de la documentación del aspirante; es decir, esta es la causa principal que hay que analizar mediante la estratificación de segundo nivel.

En el apéndice 3 se explica el código en R para seleccionar a los aspirantes que aprobaron y no aprobaron cada etapa y se han especificado las causas por las que no aprobaron la etapa. Esto es necesario para realizar el diagrama de Pareto de segundo nivel, el análisis bivariante y los modelos logísticos.

La Tabla 2 muestra la distribución de frecuencias para los aspirantes no aprobados en la etapa de recepción de la documentación:

Tabla 0.2

Tabla de frecuencias de los no aprobados en etapa 1

Causa	No aprobados	Porcentaje
Documentación incompleta	2.264	29,7%
Edad fuera de rango	775	10,2%
Estatura mínima	1.870	24,5%
Nota Ser Bachiller	2.715	35,6%
Total	7.624	100%

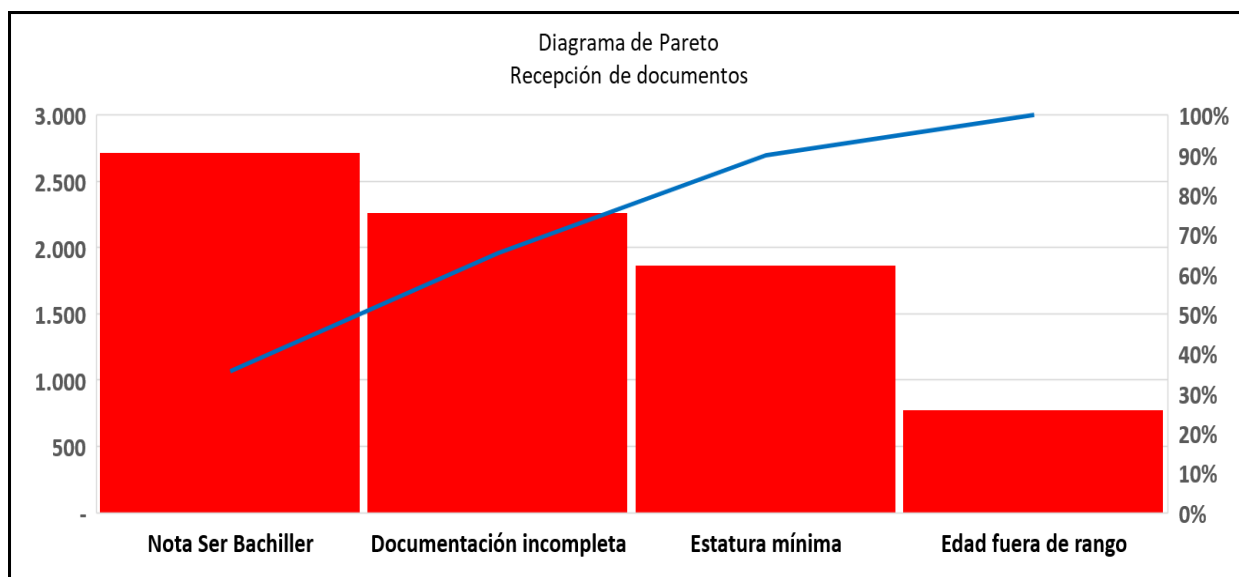
Fuente: Escuela Militar

Elaborado por: Verónica Camacho Torres

El diagrama de Pareto correspondiente a la tabla 2 se muestra en el gráfico 4:

Gráfico 4

Diagrama de Pareto en recepción de documentos



Elaborado por: Verónica Camacho Torres

Fuente: Escuela Militar

Del diagrama se observa que las causas más importantes por las que los aspirantes no aprobaron la etapa 1 son: la calificación no satisfactoria en el examen Ser Bachiller y entrega de documentación incompleta.

En la siguiente sección se realiza un análisis descriptivo bivalente entre el factor que indica si el aspirante aprobó o no aprobó la etapa y los factores: sexo edad, región de procedencia del aspirante, nivel socioeconómico y tipo de colegio de donde

proviene el estudiante, para cada una de las etapas del proceso mediante Tablas de contingencia y diagramas de barras.

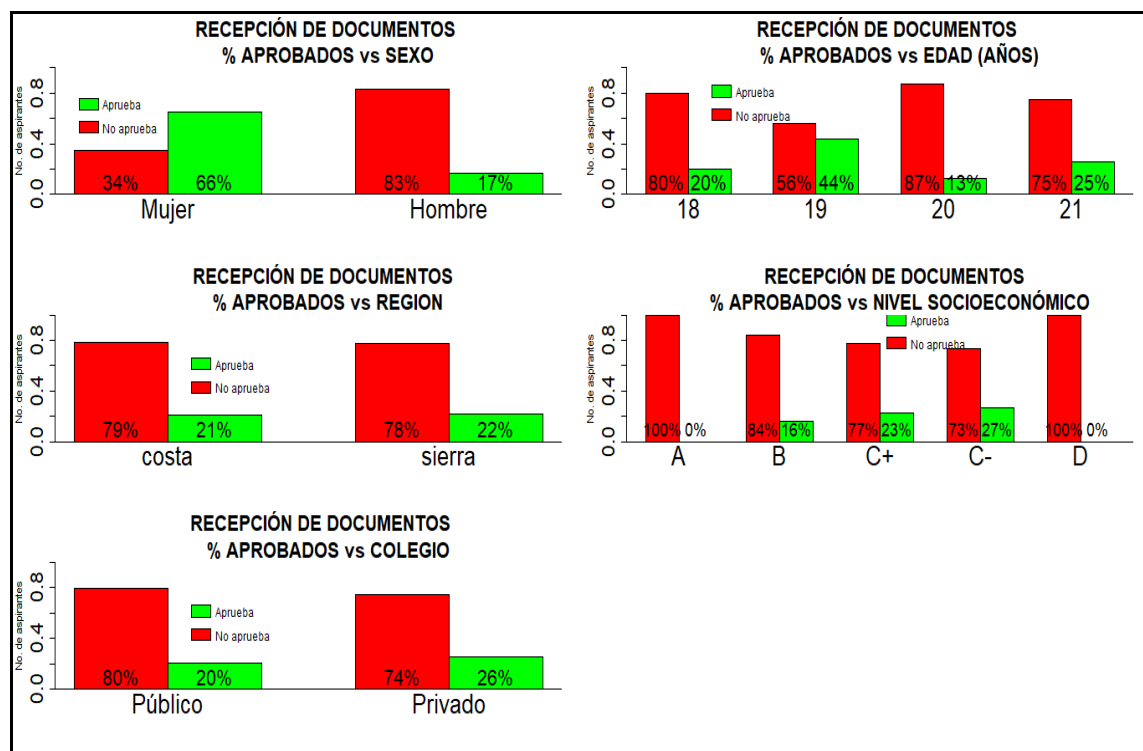
3.4 Análisis Descriptivo Bivariante de las Etapas del Proceso

En el siguiente análisis descriptivo determinaremos las etapas del proceso de ingreso de los aspirantes a la Escuela Superior Militar.

3.4.1 Etapa 1: Recepción de documentos

En el apéndice 4 se presenta el código en R para obtener las Tablas de frecuencias y los diagramas de barras entre el factor que representa si el aspirante aprobó o no aprobó la etapa de recepción de documentos, y los factores propios del aspirante: sexo, edad, región de procedencia del aspirante, nivel socioeconómico y tipo de colegio de donde proviene. El gráfico 5 muestra los diagramas de barras correspondientes: el color rojo de la barra corresponde al número de aspirantes que no aprobaron y el verde al número de aspirantes que si aprobaron la etapa:

Gráfico 5
Porcentaje de aspirantes en Recepción de Documentos



Elaborado por: Verónica Camacho Torres
Fuente: Escuela Militar

De la observación del Gráfico 5 se puede establecer que en esta etapa:

Existen más aspirantes hombres que mujeres, pero de igual manera, son los que más reprueban esta etapa; es decir, el sexo del aspirante está fuertemente relacionado con la no aprobación de esta etapa, sin considerar los otros factores en el análisis. En las mujeres, aunque en menor cantidad, se presenta el resultado contrario, más mujeres aprueban esta etapa que las que reprueban.

Los que tienen 20 años son los que más han reprobado, seguido de los que tienen 18, y son las edades que tienen la mayor cantidad de aspirantes.

Existen más aspirantes de la sierra que de la costa, pero así mismo son los que más reprueban esta etapa.

El mayor número de aspirantes tiene un nivel socioeconómico C+, seguido de C- y de B, y en esa misma secuencia es el número de reprobados en esta etapa. Existen muy pocos aspirantes de los niveles A y D.

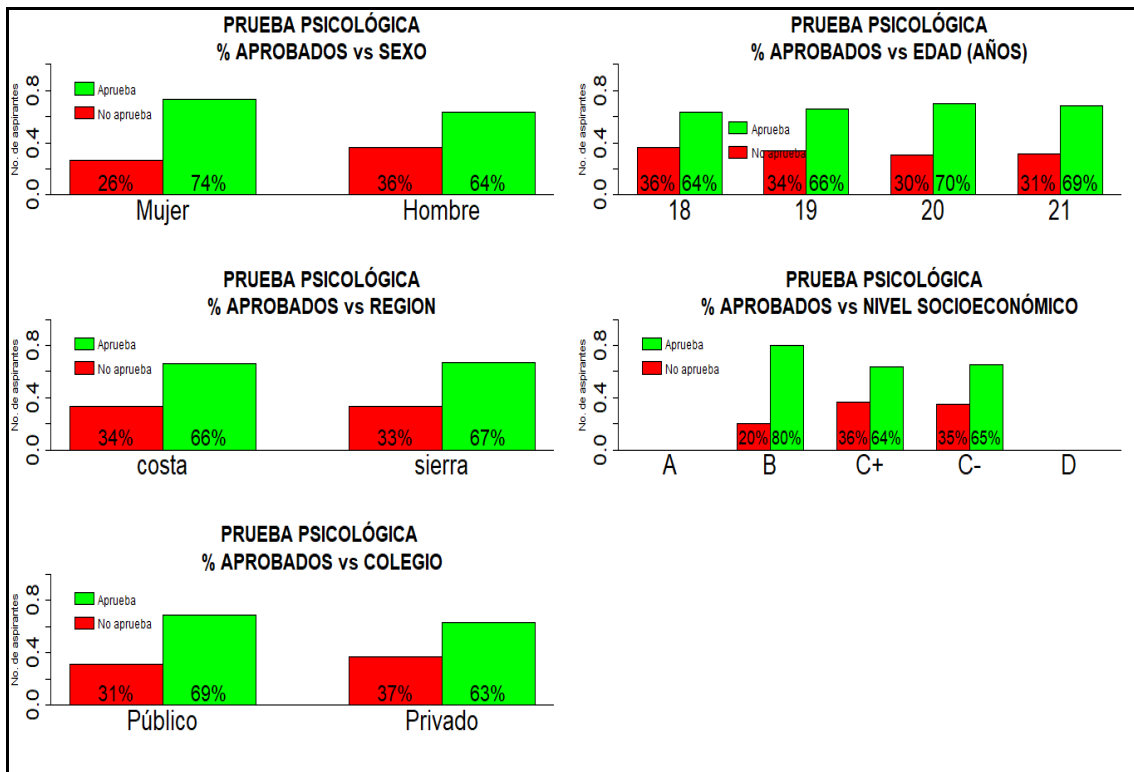
En esta etapa existen mayor cantidad de aspirantes que provienen de colegios públicos que de privados, de igual manera, la mayor cantidad de reprobados son de colegios públicos

Podría decirse que es muy probable que una combinación de niveles de los factores como hombre de 18 o 20 años, de la región sierra, con nivel socioeconómico C+ y que provenga de colegio público, repruebe esta etapa. Esto no es del todo cierto, ya que habría que estudiar si los niveles de los factores interactúan unos con otros, y para esto es necesario utilizar una técnica multivariante que permita realizar ese análisis. Esto se hará más adelante.

3.4.2 Etapa 2: Prueba psicológica

En el apéndice 5 se presenta el código en R para obtener las tablas de frecuencias y los diagramas de barras entre el factor que representa si el aspirante aprobó o no aprobó la etapa de prueba psicológica, y los factores propios del aspirante: sexo, edad, región de procedencia del aspirante, nivel socioeconómico y tipo de colegio de donde proviene. El Gráfico 7 muestra los diagramas de barras correspondientes: el color rojo de la barra corresponde al número de aspirantes que no aprobaron y el verde al número de aspirantes que si aprobaron la etapa:

Gráfico 6
Porcentaje de aspirantes en Prueba Psicológica



Elaborado por: Verónica Camacho Torres
Fuente: Escuela Militar

De la observación del gráfico 6 se puede establecer que en esta etapa:

Se produce en todos los casos el efecto contrario de la etapa anterior, es decir, la cantidad de aspirantes que aprueban es mayor que los que no aprueban.

El factor edad no parece tener un efecto significativo sobre la aprobación de esta etapa, ya que las barras son casi uniformes.

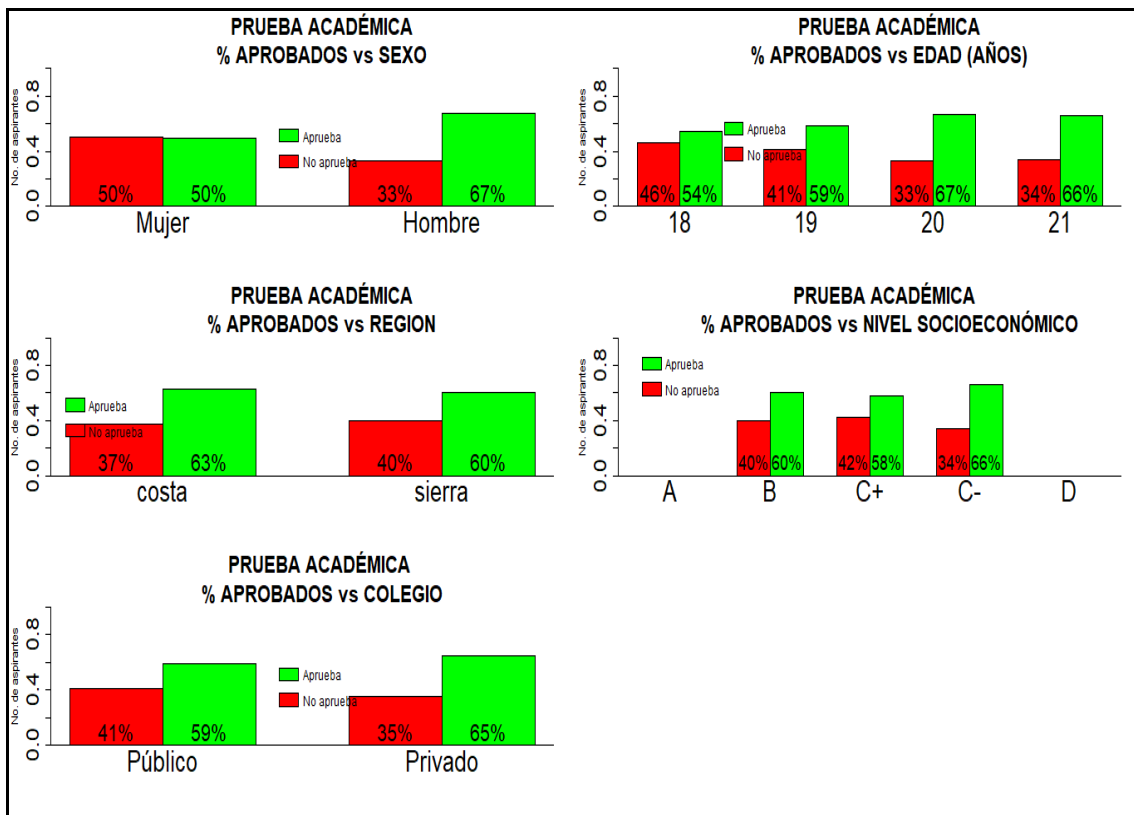
Existen más hombres que aprueban esta etapa, aunque las mujeres que aprueban lo hacen en mayor proporción que los hombres.

Existen más aprobados de la región sierra, y la proporción de aprobados es muy similar en las dos regiones. La mayor cantidad de aprobados tienen un nivel socioeconómico C+ seguidos del C-. La mayor cantidad de aprobados provienen de colegios públicos.

3.4.3 Etapa 3: Prueba académica

En el apéndice 6 se presenta el código en R para obtener las tablas de frecuencias y los diagramas de barras entre el factor que representa si el aspirante aprobó o no aprobó la etapa de prueba académica, y los factores propios del aspirante: sexo, edad, región de procedencia del aspirante, nivel socioeconómico y tipo de colegio de donde proviene. El gráfico 8 muestra los diagramas de barras correspondientes: el color rojo de la barra corresponde al número de aspirantes que no aprobaron y el verde al número de aspirantes que si aprobaron la etapa:

Gráfico 7
Porcentaje de aspirantes en Prueba Académica



Elaborado por: Verónica Camacho Torres

Fuente: Escuela Militar

De la observación del Gráfico 7 se puede establecer que en esta etapa:

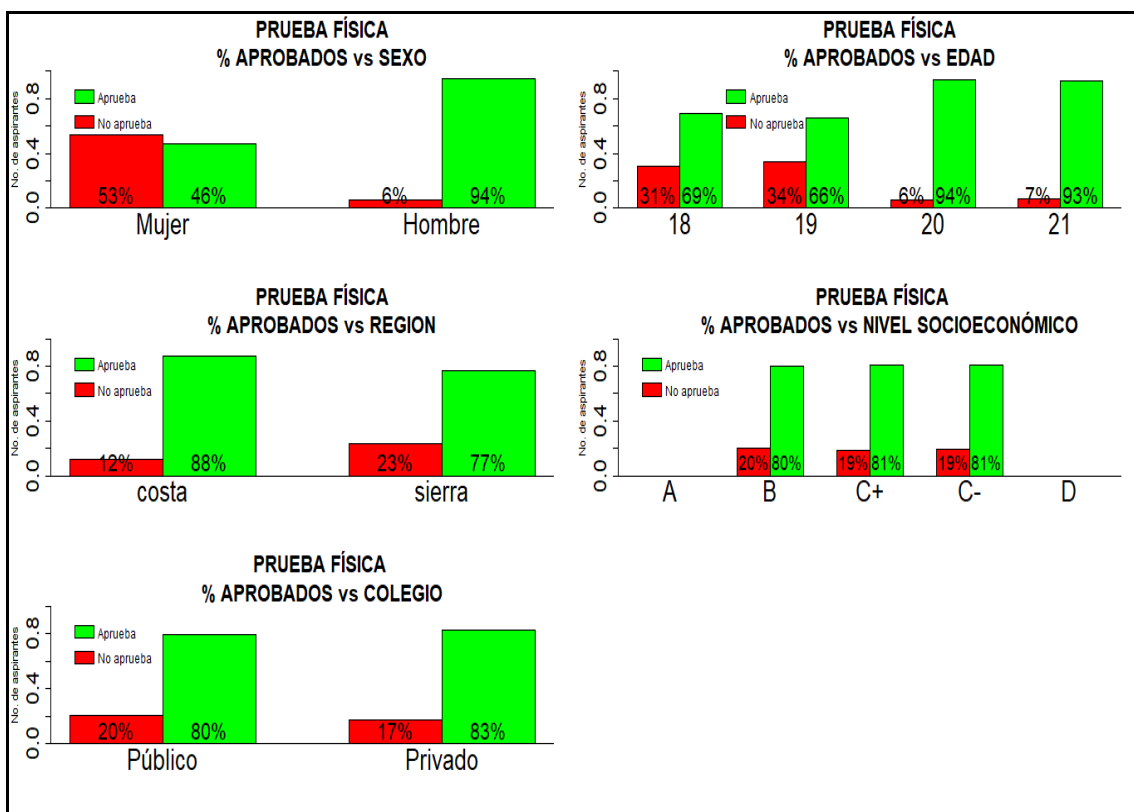
Los aspirantes hombres, los de la región sierra, los de nivel socioeconómico C+ y los que provienen de colegio público son los que más han aprobado esta etapa. La

proporción de mujeres que aprobaron y no aprobaron es prácticamente la misma. La edad no tiene un efecto claro sobre la aprobación de esta etapa.

3.4.4 Etapa 4: Prueba física

En el apéndice 7 se presenta el código en R para obtener las Tablas de frecuencias y los diagramas de barras entre el factor que representa si el aspirante aprobó o no aprobó la etapa de prueba física, y los factores propios del aspirante: sexo, edad, región de procedencia del aspirante, nivel socioeconómico y tipo de colegio de donde proviene. El Gráfico 9 muestra los diagramas de barras correspondientes: el color rojo de la barra corresponde al número de aspirantes que no aprobaron y el verde al número de aspirantes que si aprobaron la etapa:

Gráfico 8
Porcentaje de aspirantes en Prueba Física



Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

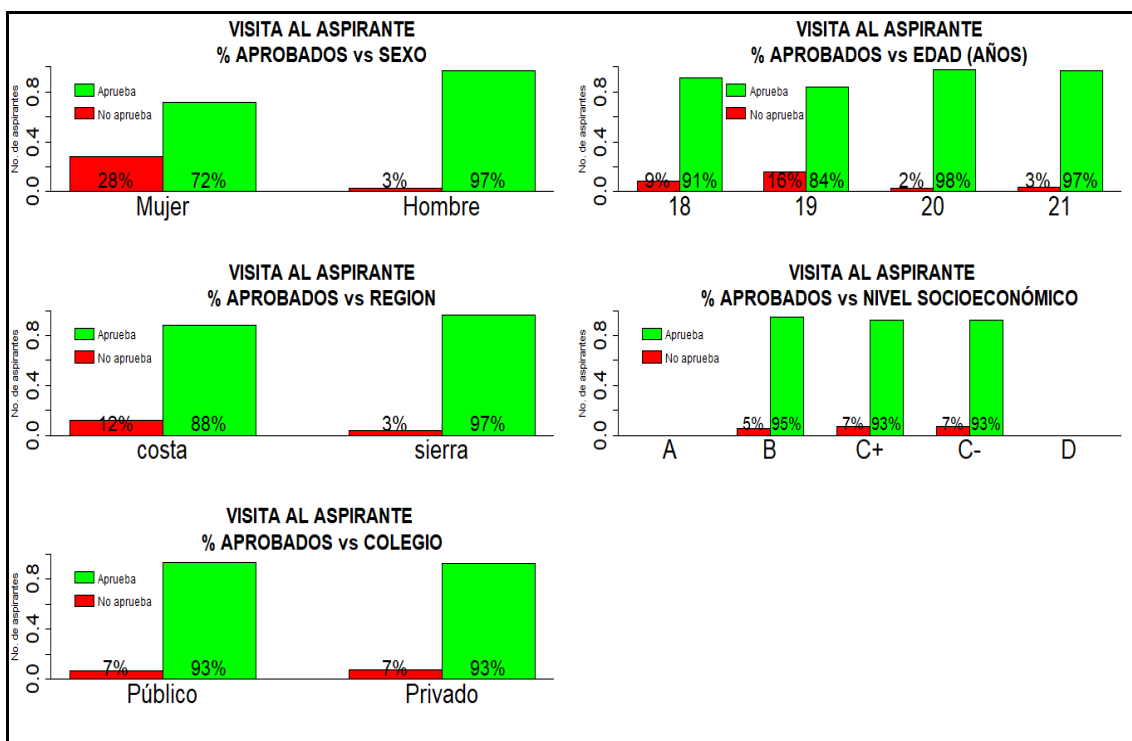
De la observación del gráfico 8 se puede establecer que en esta etapa:

Es mucho mayor la cantidad de hombres que aprueban esta etapa que las mujeres, y los hombres que reprobaban son muy pocos; los aspirantes de 20 y 21 años aprueban en mayor proporción que los de 18 y 19; los aspirantes de la sierra aprueban en mayor cantidad, aunque los de la costa lo hacen en mayor proporción. Los aspirantes de nivel socioeconómico C+ y C- aprueban en mayor cantidad esta etapa; los aspirantes de colegio público aprueban en mayor número.

3.4.5 Etapa 5: Visita al aspirante

En el apéndice 8 se presenta el código en R para obtener las Tablas de frecuencias y los diagramas de barras entre el factor que representa si el aspirante aprobó o no aprobó la etapa de visita al aspirante, y los factores propios del aspirante: sexo, edad, región de procedencia del aspirante, nivel socioeconómico y tipo de colegio de donde proviene. El Gráfico 10 muestra los diagramas de barras correspondientes: el color rojo de la barra corresponde al número de aspirantes que no aprobaron y el verde al número de aspirantes que si aprobaron la etapa:

Gráfico 9
Porcentaje de aspirantes en Visita al Aspirante



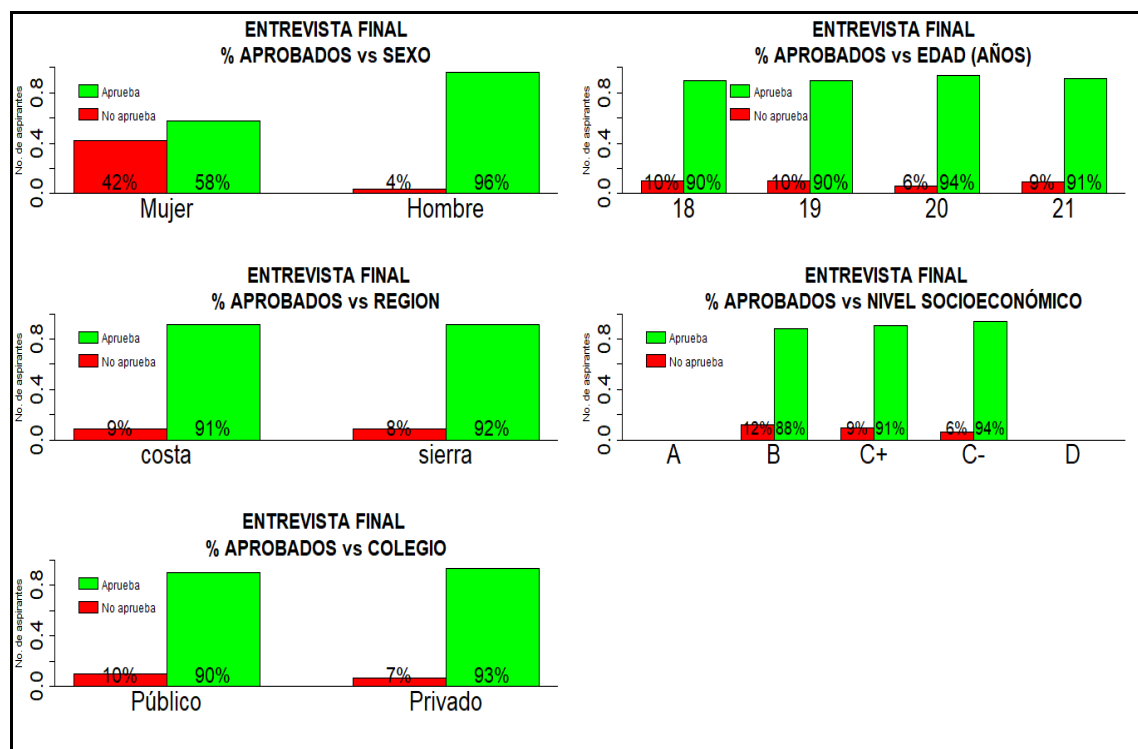
Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De la observación del gráfico 9 se puede establecer que en esta etapa: La proporción de aspirantes aprobados es mucho mayor que los que no aprueban en todos los niveles de todos los 5 factores. Los aspirantes hombres, los de 20 y 21 años, los de la sierra, los de nivel socioeconómico C+ y C-, y los que provienen de colegio público aprueban en mayor número esta etapa.

3.4.6 Etapa 6: Entrevista final

En el apéndice 9 se presenta el código en R para obtener las tablas de frecuencias y los diagramas de barras entre el factor que representa si el aspirante aprobó o no aprobó la etapa de entrevista final, y los factores propios del aspirante: sexo, edad, región de procedencia del aspirante, nivel socioeconómico y tipo de colegio de donde proviene. El Gráfico 11 muestra los diagramas de barras correspondientes: el color rojo de la barra corresponde al número de aspirantes que no aprobaron y el verde al número de aspirantes que si aprobaron la etapa:

Gráfico 10
Porcentaje de aspirantes en Entrevista Final



Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De la observación del Gráfico 10 se puede establecer que en esta etapa: La cantidad de hombres que aprueban esta etapa es mucho mayor que la de mujeres. Los aspirantes hombres, los de 20 y 21 años, los de la sierra, los de nivel socioeconómico C+ y C-, y los que provienen de colegio público aprueban en mayor número esta etapa.

3.5 Modelos Multivariantes de Regresión Logística Binomial

En esta sección se realizará el modelo logístico global, se construirán modelos logísticos por cada una de las etapas del proceso, predicción de la probabilidad de ingreso de un aspirante dado, probabilidad frecuentista global y por etapas como la Recepción de documentos, pruebas psicológicas, pruebas académicas, pruebas físicas, visita al aspirante y entrevista final.

3.5.1 Modelo logístico global del proceso

Una de las formas de estudiar el efecto conjunto de un grupo de factores sobre una respuesta tipo binario es construyendo un modelo de regresión logística binaria para la respuesta en función de los factores considerados en el análisis. En esta sección y en las siguientes se construyen modelos de regresión logística binaria.

En el apéndice 10 se detalla el código en R para obtener el modelo de regresión logística binaria que tenga como variable de respuesta el logit de la probabilidad de ingreso del aspirante a la Escuela Militar y como variables explicativas: sexo, edad, región, nivel socioeconómico y tipo de colegio donde estudió el aspirante. Primero se incluyen todas las variables explicativas, y luego, mediante el mecanismo de regresión por pasos, se obtiene el mejor modelo que incluya a las variables explicativas que son significativas para explicar la variable respuesta, la tabla 3 resume estos pasos:

Tabla 0.3
Regresión por pasos para el Modelo Global

Paso	Factores explicativos	AIC	Deviance
1	sexo, edad, region, nivelsocio, colegio	4.419,6	4.401,6
2	sexo, edad, nivelsocio, colegio	4.417,7	4.401,7
3	sexo, edad, colegio	4.463,0	4.455,0

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El mejor modelo de acuerdo al AIC es el que incluye a las variables sexo, edad, nivelsocio y colegio ya que posee el menor AIC, **sin embargo**, los valores p del modelo construido incluyendo estos 4 factores se muestran en la Tabla 4:

Tabla 0.4
Betas y p-valores del modelo global con menor AIC

Variable	Beta estimado	valor p
sexo-Hombre	0,30779	0,04703
Edad	0,23166	3,93 e-09
colegio-Privado	0,33467	0,00010
nivelsocio B	0,11594	0,999729
nivelsocio C+	-15,86833	0,956149
nivelsocioC-	-0,32183	0,998496
nivelsocio D	-3,65899	0,954782

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De acuerdo a esta Tabla, **la variable nivelsocio no es significativa en ninguno de sus niveles, por lo que se la descarta del modelo**. La tabla 5 muestra las betas estimados y los valores p del modelo que incluye las variables sexo, edad y colegio eliminando el factor nivelsocio:

Tabla 0.5
Betas estimados valores p del modelo global

Variable	Beta estimado	valor p
sexo-Hombre	0,29055	0,06024
Edad	0,22774	0,00000
colegio-Privado	0,33559	0,00013

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El modelo global obtenido tiene la siguiente expresión:

$$\text{logit}(P) = -7.56379 + 0.29055(\text{sexo}) + 0.22774(\text{edad}) + 0.33559(\text{colegio})$$

Los factores sexo, edad y colegio son significativos de acuerdo a sus valores p en el análisis de deviance, **aunque se ha incluido la variable sexo porque de manera práctica está en el límite del nivel de significancia**. El test de la razón de

verosimilitud indica que el modelo difiere significativamente del modelo nulo (p-value 0) y no difiere significativamente del modelo saturado (p-value 1). El R^2 es 0.01191467 que es un valor muy bajo. El AIC de este modelo es 4.463,72.

La tabla 6 muestra una estimación puntual de los odds-ratio para las variables binarias sexo, edad y colegio:

Tabla 0.6
Odds-ratios en el modelo global

Factor	Odds-ratio
Sexo	1,337163
Edad	1,255759
Colegio	1,398765

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De los valores de esta tabla se puede expresar que el odds de los hombres es 33,7% mayor que el de las mujeres; así mismo, el odds de los aspirantes de colegios privados es 39,87% mayor que el de los colegios públicos; y el odds de la edad es 25,57% mayor por cada año de incremento en la edad, manteniendo constante los demás factores.

La tasa de aciertos del modelo global en el pronóstico es del 93,8 %, la Tabla 7 muestra la matriz de confusión del modelo:

Tabla 0.7
Matriz de confusión - modelo global

OBSERVACIONES	PRONÓSTICO	
	Aprueba	No Aprueba
Aprueba	0	600
No Aprueba	0	9.156

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

3.5.2 Modelos logísticos por cada etapa del proceso

3.5.2.1 Etapa 1: Recepción de documentos

En el apéndice 11 se detalla el código en R para obtener el modelo de regresión logística binaria que tenga como variable de respuesta el logit de la probabilidad de

aprobación del aspirante en esta etapa de recepción de documentos y como variables explicativas: sexo, edad, región, nivel socioeconómico y tipo de colegio donde estudió el aspirante. Primero se incluyen todas las variables explicativas, y luego, mediante el mecanismo de regresión por pasos, se obtiene el mejor modelo que incluya a las variables explicativas que son significativas para explicar la variable respuesta, la tabla 8 resume estos pasos:

Tabla 0.8
Regresión por pasos para el modelo de Recepción de Documentos

Paso	Factores explicativos	AIC	Deviance
1	sexo, edad, region, nivelsocio, colegio	8.973,9	8.955,9
2	sexo, edad, nivelsocio, colegio	8.973,0	8.957,0
3	sexo, nivelsocio, colegio	8.972,8	8.958,8
4	sexo, colegio	9.164,7	9.158,7

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El mejor modelo de acuerdo al AIC es el que incluye a las variables sexo, nivelsocio y colegio ya que posee el menor AIC, **sin embargo**, los valores p del modelo construido incluyendo estos 3 factores se muestran en la Tabla 9:

Tabla 0.9
Betas y p-values del modelo1 con menor AIC

Variable	Beta estimado	valor p
sexo-Hombre	-2,29476	2 e-16
colegio-Privado	0,37284	3,5 e-11
nivelsocio B	0,20492	0,999
nivelsocio C+	-16,47429	0,921
nivelsocioC-	-0,41852	0,997
nivelsocio D	-361.346	0,922

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De acuerdo a esta tabla, **la variable nivelsocio no es significativa en ninguno de sus niveles, por lo que se la descarta del modelo**. La Tabla 10 muestra las betas estimados y los valores p del modelo que incluye las variables sexo y colegio eliminando el factor nivelsocio:

Tabla 0.10
 Betas estimados y valores p del modelo en la etapa 1

Variable	Beta estimado	valor p
sexo-Hombre	-2,25901	0,00000
colegio-Privado	0,36582	0,00000

Elaborado por: Verónica Camacho Torres
 Fuente: Escuela Superior

El modelo 1 obtenido tiene la siguiente expresión:

$$\text{logit}(P) = 0,53193 - 2,25901(\text{sexo}) + 0,36582(\text{colegio})$$

Los factores sexo y colegio son significativos de acuerdo a sus valores p en el análisis de deviance. El test de la razón de verosimilitud indica que el modelo difiere significativamente del modelo nulo (p-value 0) y no difiere significativamente del modelo saturado (p-value 1). El R^2 del modelo es 0.1059974 que es un valor bajo. El AIC del modelo 1 es 9.164,7

La Tabla 11 muestra una estimación puntual de los odds-ratio para las variables binarias sexo y colegio:

Tabla 0.11
 Odds-ratios en modelo etapa 1

Factor	Odds-ratio
sexo	0,10445384
colegio	1,44169571

Elaborado por: Verónica Camacho Torres
 Fuente: Escuela Superior

De los valores de esta tabla se puede expresar que el odds de los hombres es el 10% del de las mujeres, es decir, en esta etapa las mujeres son más propensas a aprobar que a desaprobado la etapa que los hombres. El odds de los aspirantes de colegios privados es el 44,1% mayor que el de los aspirantes de colegios públicos, manteniendo constante los demás factores.

La tasa de aciertos del modelo 1 en el pronóstico es del 81,4 %, la Tabla 12 muestra la matriz de confusión del modelo:

Tabla 0.12
Matriz de confusión - modelo etapa 1

OBSERVACIONES	PRONÓSTICO	
	Aprueba	No Aprueba
Aprueba	676	1.456
No Aprueba	358	7.266

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

3.5.2.2 Etapa 2: Prueba psicológica

En el apéndice 12 se detalla el código en R para obtener el modelo de regresión logística binaria que tenga como variable de respuesta el logit de la probabilidad de aprobación del aspirante en esta etapa de prueba psicológica y como variables explicativas: sexo, edad, región, nivel socioeconómico y tipo de colegio donde estudió el aspirante. Primero se incluyen todas las variables explicativas, y luego, mediante el mecanismo de regresión por pasos, se obtiene el mejor modelo que incluya a las variables explicativas que son significativas para explicar la variable respuesta, la tabla 13 resume estos pasos.

Tabla 0.13
Regresión por pasos para el modelo de Prueba Psicológica

Paso	Factores explicativos	AIC	Deviance
1	sexo, edad, region, nivelsocio, colegio	2.654,0	2.648,0
2	sexo, edad, nivelsocio, colegio	2.653,0	2.641,0

Elaborado por: Verónica Camacho Torres
Escuela Superior

La Tabla 14 muestra los betas estimados y los valores p de las variables incluidas en el mejor modelo:

Tabla 0.14
Betas estimados y valores p del modelo en la etapa 2

Variable	Beta estimado	valor p
sexo-Hombre	-0,48380	0,00000
edad	0,10660	0,01140
nivelsocio-C+	-0,52740	0,00000
nivelsocio-C-	0,36434	0,00001
colegioPrivado	-0,26487	0,00628

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El modelo 2 obtenido, tiene la siguiente expresión:

$$\text{logit}(P) = -0,77655 - 0,48380(\text{sexo}) + 0,10660(\text{edad}) - 0,52740(\text{nivelsocioC+}) \\ + 0,36434(\text{nivelsocioC-}) - 0,26487(\text{colegio})$$

Los factores sexo, edad, nivel socioeconómico y colegio son significativos de acuerdo a sus valores p en el análisis de deviance. El test de la razón de verosimilitud indica que el modelo difiere significativamente del modelo nulo (p-value 0) y también difiere significativamente del modelo saturado (p-value 0), lo que implica que hay ciertas variables adicionales que no han sido consideradas en el modelo y deben ser incluidas. El R^2 es de 0.02510099

La Tabla 15 muestra una estimación puntual de los odds-ratio para las variables binarias sexo y colegio:

Tabla 0.15
Odds-ratios del modelo en etapa 2

Factor	Odds-ratio
sexo	0,616436
colegio	0,767306
edad	1,112489
nivelsocio-C+	0,590137
nivelsocio-C-	1,439564

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De esta Tabla se puede expresar que el odds de los hombres es el 61,64% del de las mujeres; y el odds de los aspirantes de colegios privados es el 76,73% del de los colegios públicos: el odds de la edad se incrementa en el 11,25% por cada año de incremento en la edad manteniendo constante los demás factores. Con relación a la variable ordinal nivel socioeconómico, se tiene que los niveles son significativos son C+ y C-, el odds del nivel C+ es el 59% con respecto al nivel base que es el B, ya que no existen aspirantes en el nivel A y E en esta etapa. De la misma forma, el odds del nivel C- es el 43,95% mayor que el odds del nivel B. Esto sugiere que el nivel C- es el que tiene una mayor proporción a a probar que a desaprobado la etapa 2 que los demás niveles socioeconómicos.

La tasa de aciertos del modelo 2 en el pronóstico es del 66,8 %, la Tabla 16 muestra la matriz de confusión del modelo:

Tabla 0.16
Matriz de confusión - modelo etapa 2

OBSERVACIONES	PRONÓSTICO	
	Aprueba	No Aprueba
Aprueba	1.425	0
No Aprueba	707	0

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

3.5.2.3 Etapa 3: Prueba académica

En el apéndice 13 se detalla el código en R para obtener el modelo de regresión logística binaria que tenga como variable de respuesta el logit de la probabilidad de aprobación del aspirante en esta etapa de prueba académica y como variables explicativas: sexo, edad, región, nivel socioeconómico y tipo de colegio donde estudió el aspirante. Primero se incluyen todas las variables explicativas, y luego, mediante el mecanismo de regresión por pasos, se obtiene el mejor modelo que incluya a las variables explicativas que son significativas para explicar la variable respuesta, la tabla 17 resume estos pasos:

Tabla 0.17
Regresión por pasos para el modelo de Prueba Académica

Paso	Factores explicativos	AIC	Deviance
1	sexo, edad, región, nivelsocio, colegio	1.860,3	1.846,3
2	sexo, edad, nivelsocio, colegio	1.858,8	1.846,8
3	sexo, edad, nivelsocio	1.858,8	1.848,8
4	sexo, edad	1.860,7	1.854,7

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El mejor modelo de acuerdo al AIC es el que incluye a las variables sexo, edad, nivelsocio y colegio ya que posee el menor AIC, **sin embargo**, los valores p del modelo construido incluyendo estos 3 factores se muestran en la tabla 18:

Tabla 0.18
Betas y p-values en modelo 3 con menor AIC

Variable	Beta estimado	valor p
sexo-Hombre	0,66517	8,23 e-9
edad	0,14153	0,00498
nivelsocio-B	0,14629	0,19098
nivelsocio-C+	0,1550	0,09396
colegioPrivado	0,1673	0,16011

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

Se puede ver de esta Tabla que **los factores nivelsocio y colegio no son significativos, por esta razón la función step () de R elimina primero el factor colegio y luego el factor nivelsocio**. La tabla 19 muestra los betas estimados y los valores p de las variables sexo y edad correspondientes al mejor modelo:

Tabla 0.19
Betas estimados y valores p del modelo en etapa 3

Variable	Beta estimado	valor p
sexo-Hombre	0,67881	0,00000
edad	0,14749	0,00331

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El modelo 3 obtenido tiene la siguiente expresión:

$$\text{logit}(P) = -2,8425 + 0,6788(\text{sexo}) + 0,1475(\text{edad})$$

Los factores sexo y edad son significativos de acuerdo a sus valores p en el análisis de deviance. El test de la razón de verosimilitud indica que el modelo difiere significativamente del modelo nulo (p-value 0) y también difiere significativamente del modelo saturado (p-value 0), es decir, deben incluirse en el modelo factores adicionales que no han sido considerados en el estudio. El R^2 es de 0.02605566 que es un valor muy bajo. El AIC del modelo 3 es 1.860,7.

La Tabla 20 muestra una estimación puntual de los odds-ratio para las variables binarias sexo y edad:

Tabla 0.20
Odds-ratios del modelo en etapa 3

Factor	Odds-ratio
sexo	1,97153
edad	1,158922

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El odds-ratio de la variable binaria sexo es 1,9715, lo que significa que el odds de los hombres es 97,15% mayor que el de las mujeres, y el odds de la edad se incrementa en el 15,88% por cada año de incremento en la edad, manteniendo los demás factores constantes.

La tasa de aciertos del modelo 3 en el pronóstico es del 60,35 %, la Tabla 21 muestra la matriz de confusión del modelo:

Tabla 0.21
Matriz de confusión - modelo etapa 3

OBSERVACIONES	PRONÓSTICO	
	Aprueba	No Aprueba
Aprueba	702	169
No Aprueba	396	158

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

3.5.2.4 Etapa 4: Prueba física

En el apéndice 14 se detalla el código en R para obtener el modelo de regresión logística binaria que tenga como variable de respuesta el logit de la probabilidad de aprobación del aspirante en esta etapa de prueba física y como variables explicativas: sexo, edad, región, nivel socioeconómico y tipo de colegio donde estudió el aspirante. Primero se incluyen todas las variables explicativas, y luego, mediante el mecanismo de regresión por pasos, se obtiene el mejor modelo que incluya a las variables explicativas que son significativas para explicar la variable respuesta, la tabla 22 resume estos pasos:

Tabla 0.22
Regresión por pasos en el modelo de Prueba Física

Paso	Factores explicativos	AIC	Deviance
1	sexo, edad, región, nivelsocio, colegio	566,2	552,2
2	sexo, edad, región, colegio	563,1	553,1
3	sexo, edad, región	561,7	553,7

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

La Tabla 23 muestra los betas estimados y los valores p de las variables incluidas en el mejor modelo:

Tabla 0.23
Betas estimados y valores p del modelo en etapa 4

Variable	Beta estimado	valor p
sexo-Hombre	2,91820	0,00000
Edad	0,66840	0,00000
región-Sierra	-0,88040	0,00024

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El modelo 4 obtenido tiene la siguiente expresión:

$$\text{logit}(P) = -12,3873 + 2,9182(\text{sexo}) + 0,6684(\text{edad}) - 0,8804(\text{region})$$

Los factores sexo, edad y región son significativos de acuerdo a sus valores p en el análisis de deviance. El test de la razón de verosimilitud indica que el modelo difiere significativamente del modelo nulo (p-value 0) y no difiere significativamente del modelo saturado (p-value 1). El R^2 es 0.3496225.

La Tabla 24 muestra los valores estimados de los odds-ratios para las variables binarias sexo y region:

Tabla 0.24
Odds-ratios del modelo en etapa 4

Factor	Odds-ratio
Sexo	18,50794
Edad	1,951113
región	0,414617

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De la Tabla anterior se puede indicar que el odds de los hombres es 18,50 veces el odds de las mujeres; así mismo, el odds de los aspirantes de la sierra es el 41,46% de los aspirantes de la costa; el odds de la edad se incrementa en el 95,11% por cada año de incremento en la edad, manteniendo constantes los demás factores.

La tasa de aciertos del modelo 4 en el pronóstico es del 85,9 %, la tabla 25 muestra la matriz de confusión del modelo:

Tabla 0.25
Matriz de confusión - modelo etapa 4

OBSERVACIONES	PRONÓSTICO	
	Aprueba	No Aprueba
Aprueba	657	47
No Aprueba	76	91

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

3.5.2.5 Etapa 5: Visita al aspirante

En el apéndice 15 se detalla el código en R para obtener el modelo de regresión logística binaria que tenga como variable de respuesta el logit de la probabilidad de aprobación del aspirante en esta etapa de visita al aspirante y como variables

explicativas: sexo, edad, región, nivel socioeconómico y tipo de colegio donde estudió el aspirante. Primero se incluyen todas las variables explicativas, y luego, mediante el mecanismo de regresión por pasos, se obtiene el mejor modelo que incluya a las variables explicativas que son significativas para explicar la variable respuesta, la tabla 26 resume estos pasos:

Tabla 0.26
Regresión por pasos en el modelo de Visita al Aspirante

Paso	Factores explicativos	AIC	Deviance
1	sexo, edad, región, nivel socio, colegio	262,6	248,6
2	sexo, edad, región, colegio	260,0	250,0
3	sexo, edad, región	258,0	250,0

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

La tabla 27 muestra los betas estimados y los valores p de las variables incluidas en el mejor modelo:

Tabla 0.27
Betas estimados y valores p del modelo en etapa 5

Variable	Beta estimado	valor p
sexo-Hombre	2,74610	0,00000
Edad	0,49350	0,00184
región-Sierra	1,49810	0,00003

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El modelo 5 obtenido tiene la siguiente expresión:

$$\text{logit}(P) = -9,3503 + 2,7461(\text{sexo}) + 0,4935(\text{edad}) + 1,4981(\text{region})$$

Los factores sexo, edad y región son significativos de acuerdo a sus valores p en el análisis de deviance. El test de la razón de verosimilitud indica que el modelo difiere significativamente del modelo nulo (p-value 0) y no es diferente significativamente del modelo saturado (p-value 1). El R^2 es 0.286687.

La Tabla 28 muestra los valores estimados de los odds-ratios para las variables binarias sexo y región:

Tabla 0.28
Odds-ratios del modelo en etapa 5

Factor	Odds-ratio
Sexo	15,58174
Edad	1,638039
Región	4,473182

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

De los valores de la tabla se puede expresar que el odds de los hombres es 15,58 veces el odds de las mujeres; así mismo, el odds de los aspirantes de la sierra es 4,47 veces el odds de los aspirantes de la costa; el odds de la edad se incrementa en el 63,38% por cada año de aumento en la edad, manteniendo los demás factores constantes.

La tasa de aciertos del modelo 5 en el pronóstico es del 93,2 %, la tabla 29 muestra la matriz de confusión del modelo:

Tabla 0.29
Matriz de confusión - modelo etapa 5

OBSERVACIONES	PRONÓSTICO	
	Aprueba	No Aprueba
Aprueba	650	6
No Aprueba	42	6

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

3.5.2.6 Etapa 6: Entrevista final

En el apéndice 16 se detalla el código en R para obtener el modelo de regresión logística binaria que tenga como variable de respuesta el logit de la probabilidad de aprobación del aspirante en esta etapa de entrevista final y como variables explicativas: sexo, edad, región, nivel socioeconómico y tipo de colegio donde estudió el aspirante. Primero se incluyen todas las variables explicativas, y luego, mediante el mecanismo de regresión por pasos, se obtiene el mejor modelo que incluya a las variables explicativas que son significativas para explicar la variable respuesta, la tabla 30 muestra estos pasos:

Tabla 0.30
Regresión por pasos en el modelo de Entrevista Final

Paso	Factores explicativos	AIC	Deviance
1	sexo, edad, región, nivelsocio, colegio	382,8	288,8
2	sexo, región, nivelsocio, colegio	301,1	289,1
3	sexo, región, colegio	299,3	291,3
4	sexo, colegio	297,4	291,4
5	Sexo	297,1	293,1

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

La Tabla 31 muestra los betas estimados y los valores p de las variables incluidas en el mejor modelo:

Tabla 0.31
Beta estimado y valor p del modelo

Variable	Beta estimado	valor p
sexo-Hombre	2,95320	0,00000

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

El modelo 6 obtenido tiene la siguiente expresión incluyendo el intercepto:

$$\text{logit}(P) = 0,3159 + 2,9532(\text{sexo})$$

El factor sexo es significativo de acuerdo a su valor p en el análisis de deviance. El test de la razón de verosimilitud indica que el modelo difiere significativamente del modelo nulo (p-value 0) y no difiere significativamente del modelo saturado (p-value 1). El R^2 es 0.2340864. El AIC del modelo 6 es 297,1.

El odds-ratio de la variable binaria sexo es 19,17, lo que indica que el odds de los hombres es 19,17 veces el odds de las aspirantes mujeres, manteniendo constantes los demás factores.

La tasa de aciertos del modelo 6 en el pronóstico es del 91,5 %, la Tabla 32 muestra la matriz de confusión del modelo:

Tabla 0.32
Matriz de confusión - Modelo 6

OBSERVACIONES	PRONÓSTICO	
	Aprueba	No Aprueba
Aprueba	600	0
No Aprueba	56	0

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

3.5.3 Predicción de la probabilidad de ingreso de un aspirante dado

A manera de aplicación de lo desarrollado en los numerales anteriores, se va a calcular la probabilidad de que un aspirante ingrese a la Escuela superior de cuatro maneras: calculando la probabilidad frecuentista del proceso global, calculando la probabilidad frecuentista por etapas sin considerar los niveles de los factores del aspirante, usando el modelo logístico global y mediante la probabilidad condicional con los modelos logísticos de cada etapa; estos dos últimos considerando los niveles de los factores del aspirante. Los niveles de los factores para el aspirante se supondrán en sexo = “Hombre”, edad = 21 años, región = sierra, nivel socioeconómico = C+ y colegio = “Privado”

3.5.3.1 Probabilidad frecuentista global

De acuerdo a este concepto se cuentan el número de aspirantes aprobados (600) y el número total de aspirantes que iniciaron el proceso de admisión (9.756) sin condicionar a ningún factor del aspirante. Realizando el cálculo se obtiene:

$$P(\text{ingresar}) = \frac{600}{9.756} = 0.0615006$$

Es decir, sin condicionar ningún factor específico del aspirante, su probabilidad de ingresar a la Escuela militar es del 0.06150062 y dado que es menor que 0,5, el aspirante no ingresaría a la Escuela Militar

3.5.3.2 Probabilidad frecuentista por etapas

En la tabla 33 se detalla, para cada etapa, el número de aspirantes que aprueba la etapa, el número total de aspirantes al inicio de la etapa y la probabilidad frecuentista de aprobar dicha etapa:

Tabla 0.33
Probabilidad frecuentista por etapas

Etapas	Aprobados	Total	Probabilidad
Recepción de documentos	2.132	9.756	0,218532185
Prueba psicológica	1.425	2.132	0,668386492
Prueba académica	871	1.425	0,611228070
Prueba física	704	871	0,808266361
Visita al aspirante	656	704	0,931818182
Entrevista final	600	656	0,914634146

Fuente: Escuela Militar

Elaborado por: Verónica Camacho Torres

La probabilidad de ingresar, sin considerar los niveles de los factores del aspirante y tomando en cuenta la secuencia de etapas, se obtiene multiplicando los resultados de las probabilidades de la tabla 33, el resultado es 0,0615006, que es el mismo resultado que se obtuvo con la probabilidad frecuentista global; es decir, calcular la probabilidad de esta manera no produce un resultado distinto.

3.5.3.3 Probabilidad usando el modelo logístico global

En el apéndice 17 se muestra el código en R para predecir la probabilidad del aspirante utilizando el modelo logístico global, y para este modelo sólo se incluyen los factores sexo, edad y colegio. La probabilidad obtenida sin considerar lo que pasa en las etapas intermedias es 0.1038559, que es un poco mayor que el valor encontrado con el concepto frecuentista.

3.5.3.4 Probabilidad usando los modelos logísticos por etapas

En el apéndice 17 se muestra el código en R para encontrar la probabilidad en cada etapa del proceso utilizando los modelos logísticos obtenidos anteriormente. Un resumen de los resultados se muestra en la Tabla 34:

Tabla 0.34
Probabilidades con modelos logísticos

Etapas	Probabilidad
Recepción de documentos	0,2040347
Prueba psicológica	0,6025080
Prueba académica	0,7178009
Prueba física	0,9755855
Visita al aspirante	0,9948184
Entrevista final	0,9633508

Elaborado por: Verónica Camacho Torres
Fuente: Escuela Superior

La probabilidad de ingresar considerando los factores del aspirante y la secuencia de etapas se obtiene multiplicando los resultados de las probabilidades de la tabla 34, el resultado es 0.08250202.

3.6 Propuesta de Mejoras y de Temas Específicos para Pruebas Académicas

Mediante el estudio realizado se propone mejoras en cada una de las etapas para el proceso de admisión y los temas específicos de los aspirantes a una Escuela Superior Militar.

3.6.1 Propuesta de mejoras

Con base en los resultados obtenidos anteriormente, se procede a proponer mejoras a los procedimientos en cada etapa del proceso:

En la etapa 1 de recepción de documentos, se propone realizar una preinscripción personal del aspirante, en la cual sólo se pedirá su cédula de identidad, su nota del examen Ser Bachiller de la página del SENESCYT, y se medirá su estatura. Esto permitirá que en esta fase se verifique sólo esta documentación y automáticamente descartar a quienes no cumplan con estos requisitos; de esta manera, se evita tener que procesar abundante documentación de todos los aspirantes al inicio del proceso. Una vez terminada esta etapa de preinscripción, se continúa con la inscripción de los aspirantes que aprobaron la preinscripción.

En la etapa 2 de la prueba psicológica, no se proponen mejoras, debido a que se miden aspectos de personalidad y abstracción del aspirante.

En la etapa 3 de las pruebas académicas, se propone temas específicos para los contenidos a evaluar en las pruebas, debido a que los temas que se evalúan son muy generales y abarcan muchos subtemas. Adicionalmente se propone publicar con varios meses de anticipación estos contenidos, para darle al aspirante el tiempo para que pueda revisar estos temas.

En la etapa 4 de pruebas físicas, se propone discriminar el tipo de pruebas para hombres y mujeres, ya que en la actualidad ambos realizan las mismas pruebas, aunque con estándares de medición diferentes.

Para las etapas 5 y 6, de visita al aspirante y entrevista final, no se tienen propuestas de mejora.

3.6.2 Propuesta de temas específicos para las pruebas académicas

De acuerdo a un sondeo realizado a algunos aspirantes sobre las pruebas académicas, consideran que el temario es muy extenso para el tiempo que transcurre entre la fecha publicación del temario y la fecha para rendir la prueba. Por esta razón, en el apéndice 18 se proponen temarios específicos de las materias de matemáticas, física y química, que son las materias con temarios más extensos. Esto debe ser tomado como una propuesta y solamente eso.

Capítulo 4

4. CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

Como conclusiones del presente estudio se tienen las siguientes:

La etapa de recepción de documentos es la que presenta la mayor cantidad de aspirantes que desapruban el proceso y supera en gran medida a las demás etapas del proceso; por lo tanto, es la etapa a la que debe darse mayor atención. La calificación del examen Ser Bachiller del aspirante y la documentación incompleta, son las causas más importantes de la no aprobación del aspirante en esta etapa.

Los modelos logísticos encontrados tienen un coeficiente R^2 muy pequeño, lo que podría significar que el modelo tendría interacción entre los factores del aspirante, que el modelo no sea lineal, o que existen factores influyentes en la variable respuesta que no han sido considerados. Por lo tanto, las probabilidades estimadas con los modelos obtenidos deben mirarse con reserva.

El factor sexo es un factor que influye mucho en la respuesta, tanto en la descripción bivariante, como en todos los modelos logísticos multivariante. Esto quiere decir que el sexo es determinante al momento de explicar si el aspirante aprueba o no una etapa del proceso.

4.2 Recomendaciones

Como recomendaciones para un estudio posterior podemos anotar:

Se estudiará las causas de tercer nivel para la documentación incompleta, es decir, qué documentación es la que en mayor grado no presentan los aspirantes, para poder proponer mejoras al proceso.

Se construirá los modelos logísticos que incluyan factores adicionales a los utilizados en este estudio y determinar su ajuste a los datos.

Se construirá los modelos logísticos probit con los factores actuando independientemente y de igual forma construir modelos logísticos logit y probit con interacción entre los factores.

Referencias

- Agresti, A. (2002). *Categorical Data Analysis*. New Jersey: John Wiley & Sons Inc.
- Freund, J., Miller, I., & Miller, M. (2000). *Estadística Matemática con Aplicaciones*. México: Pearson Educación.
- Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos - Aplicaciones con SPSS*. Madrid: Pearson Educación S.A.
- Rencher, A. (2002). *Methods of Multivariate Analysis*. New York: John Wiley & Sons Inc.
- Venables, W., & Ripley, B. (2002). *Modern Applied Statistics with S*. New York: Springer.
- Wackerly, D., Mendenhall, W., & Scheaffer, R. (2010). *Estadística Matemática con Aplicaciones*. México D.F.: Cengage Learning Editores S.A.

Apéndices

Apéndice 1

Se cargan las librerías:

```
library(openxlsx)
library(qcc)
```

La lectura de los datos y una vista del encabezado de los mismos se muestran a continuación:

```
datos<-read.xlsx(file.choose(),sheet=1,colNames=T)
head(datos)
##          codigo  sexo  edad  region  nivelsocio  colegio  etapa  causa
## 1             1     1    18       1           3         1     1     2
## 2             2     0    18       1           3         0     1     2
## 3             3     1    18       0           4         0     1     2
## 4             4     1    20       1           4         0     1     3
## 5             5     1    19       0           2         1     0     0
## 6             6     1    20       0           3         1     1     1
0
```

Con las siguientes instrucciones se leen los datos y se transforma en factores las variables sexo, region, nivelsocio (ordinal), etapa y colegio. Las variables codigo y causa se mantienen numéricas:

```
datos<-read.xlsx(file.choose(),sheet=1,colNames=T)
codigo<-datos$codigo
sexo<-factor(datos$sexo,labels=c("Mujer","Hombre"))
edad<-datos$edad
region<-factor(datos$region,labels=c("costa","sierra"))
nivelsocio<-ordered(datos$nivelsocio, labels = c("A","B","C+","C-","D"))
colegio<-factor(datos$colegio, labels = c("Público", "Privado"))
etapa<-factor(datos$etapa, labels = c("Ninguna", "Recepción de documentos",
"Prueba psicológica", "Prueba académica", "Prueba física", "Visita",
"Entrevista"))
causa<-datos$causa
```

Se crea el factor "ingresa" que indica si el aspirante fue admitido o no en la Escuela militar:

```
ingresa<-ifelse(etapa=="Ninguna", 1, 0)
ingresa<-factor(ingresa, labels = c("No ingresa", "Ingresa"))
```

Se crea la base de datos con las variables transformadas y se muestra un encabezado:

```
datos<-data.frame(codigo, sexo, edad, region, nivelsocio, colegio, etapa,
causa,
ingresa)
head(datos)
```

```
##   codigo  sexo edad region nivelsocio colegio          etapa
## 1     1 Hombre  18 sierra          C+ Privado Recepción de documentos
## 2     2  Mujer  18 sierra          C+ Público Recepción de documentos
## 3     3 Hombre  18 costa           C- Público Recepción de documentos
## 4     4 Hombre  20 sierra          C- Público Recepción de documentos
## 5     5 Hombre  19 costa            B Privado           Ninguna
## 6     6 Hombre  20 costa          C+ Privado Recepción de documentos
##   causa   ingresa
## 1     2 No ingresa
## 2     2 No ingresa
## 3     2 No ingresa
## 4     3 No ingresa
## 5     0  Ingresa
## 6     0 No ingresa
```

Apéndice 2

A continuación, se muestra el código para el diagrama de causa y efecto de la no aprobación en las etapas del proceso:

```
x1<-c("Documentación incompleta", "Edad fuera de rango", "Estatura mínima",  
"Nota Ser Bachiller")  
x2<-c("Razonamiento abstracto", "Personalidad", "No asistió")  
x3<-c("Matemáticas", "Física", "Lenguaje", "Química", "Realidad nacional",  
"No asistió")  
x4<-c("Natación", "Salto", "Flexiones de pecho", "Abdominales", "Barras",  
"Velocidad", "No asistió")  
x5<-c("Grupo socioeconómico NO", "Consentimiento familiar", "Sostén de  
familia")  
x6<-c("No disponible")  
causas<-list(DOCUMENTOS=x1, PRUEBA_PSILOGICA=x2, PRUEBA_ACADEMICA=x3,  
PRUEBA_FISICA=x4, VISITA=x5, ENTREVISTA=x6)  
efecto<-"No admitido"  
cause.and.effect(cause = causas, effect = efecto, title = "DIAGRAMA DE  
CAUSA Y EFECTO", cex = c(1.1, 0,9, 1.3))
```

Apéndice 3

Las siguientes líneas crean el dataframe etapa_1 que consta de todos los aspirantes al inicio del proceso, y se crea el factor “aprueba” que indica si el aspirante aprobó esta etapa o no:

```
etapa_1<-datos
etapa_1$aprueba<-ifelse(etapa_1$etapa=="Recepción de documentos", 0, 1)
etapa_1$aprueba<-factor(etapa_1$aprueba, labels = c("No aprueba",
"Aprueba"))
```

Se crea el dataframe "NA_etapa1" que contiene la información de los aspirantes que no aprobaron esta etapa y se transforma la variable causa en un factor que contiene las causas por las que el aspirante no aprobó esta etapa:

```
NA_etapa1<-subset(etapa_1, aprueba=="No aprueba")
NA_etapa1$causa<-factor(NA_etapa1$causa, labels = c("Documentación",
"Edad", "Estatura", "Nota Ser Bachiller"))
```

Las siguientes líneas crean el dataframe etapa_2 que consta de todos los aspirantes que aprobaron la etapa 1, y se crea el factor "aprueba" que indica si el aspirante aprobó esta etapa de prueba psicológica:

```
etapa_2<-subset(etapa_1, aprueba=="Aprueba")
etapa_2$aprueba<-ifelse(etapa_2$etapa=="Prueba psicológica", 0, 1)
etapa_2$aprueba<-factor(etapa_2$aprueba, labels = c("No aprueba",
"Aprueba"))
```

Se crea el dataframe "NA_etapa2" que contiene la información de los aspirantes que no aprobaron esta etapa y se transforma la variable causa en un factor que contiene las causas por las que el aspirante no aprobó esta etapa:

```
NA_etapa2<-subset(etapa_2, aprueba=="No aprueba")
NA_etapa2$causa<-factor(NA_etapa2$causa, labels = c("Razonamiento
abstracto", "Personalidad", "No asistió"))
```

Las siguientes líneas crean el dataframe etapa_3 que consta de todos los aspirantes que aprobaron la etapa 2, y se crea el factor "aprueba" que indica si el aspirante aprobó esta etapa de prueba académica:

```
etapa_3<-subset(etapa_2, aprueba=="Aprueba")
etapa_3$aprueba<-ifelse(etapa_3$etapa=="Prueba académica", 0, 1)
etapa_3$aprueba<-factor(etapa_3$aprueba, labels = c("No aprueba",
"Aprueba"))
```

Se crea el dataframe "NA_etapa3" que contiene la información de los aspirantes que no aprobaron esta etapa y se transforma la variable causa en un factor que contiene las causas por las que el aspirante no aprobó esta etapa:

```
NA_etapa3<-subset(etapa_3, aprueba=="No aprueba")
NA_etapa3$causa<-factor(NA_etapa3$causa, labels = c("Matemáticas",
"Física", "Lenguaje", "Química", "Realidad nacional", "No asistió"))
```

Las siguientes líneas crean el dataframe etapa_4 que consta de todos los aspirantes que aprobaron la etapa 3, y se crea el factor "aprueba" que indica si el aspirante aprobó esta etapa actual de prueba física:

```
etapa_4<-subset(etapa_3, aprueba=="Aprueba")
etapa_4$aprueba<-ifelse(etapa_4$etapa=="Prueba física", 0, 1)
etapa_4$aprueba<-factor(etapa_4$aprueba, labels = c("No aprueba",
"Aprueba"))
```

Se crea el dataframe "NA_etapa4" que contiene la información de los aspirantes que no aprobaron esta etapa y se transforma la variable causa en un factor que contiene las causas por las que el aspirante no aprobó esta etapa:

```
NA_etapa4<-subset(etapa_4, aprueba=="No aprueba")
NA_etapa4$causa<-factor(NA_etapa4$causa, labels = c("Natación", "Salto",
"Flexiones de pecho", "Abdominales", "Barras", "Velocidad", "No asistió" ))
```

Las siguientes líneas crean el dataframe etapa_5 que consta de todos los aspirantes que aprobaron la etapa 4, y se crea el factor "aprueba" que indica si el aspirante aprobó esta etapa actual de la visita en su domicilio:

```
etapa_5<-subset(etapa_4, aprueba=="Aprueba")
etapa_5$aprueba<-ifelse(etapa_5$etapa=="Visita", 0, 1)
etapa_5$aprueba<-factor(etapa_5$aprueba, labels = c("No aprueba",
"Aprueba"))
```

Se crea el dataframe "NA_etapa5" que contiene la información de los aspirantes que no aprobaron esta etapa y se transforma la variable causa en un factor que contiene las causas por las que el aspirante no aprobó esta etapa:

```
NA_etapa5<-subset(etapa_5, aprueba=="No aprueba")
NA_etapa5$causa<-factor(NA_etapa5$causa, labels = c("Grupo socioeconómico
NO", "Consentimiento familiar", "Sostén de familia"))
```

Las siguientes líneas crean el dataframe etapa_6 que consta de todos los aspirantes que aprobaron la etapa 5, y se crea el factor "aprueba" que indica si el aspirante aprobó esta etapa actual de la entrevista final:

```
etapa_6<-subset(etapa_5, aprueba=="Aprueba")
etapa_6$aprueba<-ifelse(etapa_6$etapa=="Entrevista", 0, 1)
etapa_6$aprueba<-factor(etapa_6$aprueba, labels = c("No aprueba",
"Aprueba"))
```

Se crea el dataframe "NA_etapa6" que contiene la información de los aspirantes que no aprobaron esta etapa y se transforma la variable causa en un factor que contiene las causas por las que el aspirante no aprobó esta etapa:

```
NA_etapa6<-subset(etapa_6, aprueba=="No aprueba")
NA_etapa6$causa<-factor(NA_etapa6$causa, labels =c("No disponible"))
```

Se obtiene la tabla de frecuencias de segundo nivel para las causas más importantes de la no aprobación de los aspirantes en la etapa 1 de recepción de la documentación del aspirante:


```
causas_nivel_2<-table(NA_etapa1$causa)
```

Apéndice 4

A continuación, se crean las tablas de frecuencias y de proporciones entre el factor aprueba en la etapa 1 de Recepción de Documentos y los factores sexo, edad, region, nivelsocio y colegio:

```
tabla1_sexo<-table(etapa_1$aprueba, etapa_1$sexo)
tabla1_sexo
##
##           Mujer Hombre
## No aprueba  358   7266
## Aprueba     676   1456
tabla1_sexo_prop<-prop.table(tabla1_sexo, 2)
tabla1_sexo_prop
##
##           Mujer   Hombre
## No aprueba 0.3462282 0.8330658
## Aprueba    0.6537718 0.1669342
tabla1_edad<-table(etapa_1$aprueba, etapa_1$edad)
tabla1_edad
##
##           18   19   20   21
## No aprueba 2287  758 3105 1474
## Aprueba    581  594  459  498
tabla1_edad_prop<-prop.table(tabla1_edad, 2)
tabla1_edad_prop
##
##           18           19           20           21
## No aprueba 0.7974198 0.5606509 0.8712121 0.7474645
## Aprueba    0.2025802 0.4393491 0.1287879 0.2525355
tabla1_region<-table(etapa_1$aprueba, etapa_1$region)
tabla1_region
##
##           costa sierra
## No aprueba 2884   4740
## Aprueba    783   1349
tabla1_region_prop<-prop.table(tabla1_region, 2)
tabla1_region_prop
##
##           costa   sierra
## No aprueba 0.7864740 0.7784529
## Aprueba    0.2135260 0.2215471
tabla1_nivelsocio<-table(etapa_1$aprueba, etapa_1$nivelsocio)
tabla1_nivelsocio
##
##           A    B    C+   C-   D
## No aprueba  80 1789 3506 2091  158
## Aprueba     0  345 1026  761   0
tabla1_nivelsocio_prop<-prop.table(tabla1_nivelsocio, 2)
tabla1_nivelsocio_prop
```

```
##
##           A           B           C+           C-           D
## No aprueba 1.0000000 0.8383318 0.7736099 0.7331697 1.0000000
## Aprueba    0.0000000 0.1616682 0.2263901 0.2668303 0.0000000
tabla1_colegio<-table(etapa_1$aprueba, etapa_1$colegio)
tabla1_colegio
##
##           Público Privado
## No aprueba   5446   2178
## Aprueba     1379    753
tabla1_colegio_prop<-prop.table(tabla1_colegio, 2)
tabla1_colegio_prop
##
##           Público   Privado
## No aprueba 0.7979487 0.7430911
## Aprueba   0.2020513 0.2569089
```

Los diagramas de barras correspondientes, en los que se muestran en color verde el número de aspirantes que aprobaron la etapa y en rojo los que no aprobaron, se muestran con el siguiente código:

```
par(mfrow=c(3,2))
BP1_sexo<-barplot(tabla1_sexo, beside = T, main = "      RECEPCIÓN DE
DOCUMENTOS
      APROBADOS vs SEXO", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP1_sexo, labels=as.character(round(tabla1_sexo_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP1_edad<-barplot(tabla1_edad, beside = T, main = "      RECEPCIÓN DE
DOCUMENTOS
      APROBADOS vs EDAD (AÑOS)", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend(3, 3500, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP1_edad, labels=as.character(round(tabla1_edad_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP1_region<-barplot(tabla1_region, beside = T, main = "      RECEPCIÓN DE
DOCUMENTOS
      APROBADOS vs REGION", col=c("red", "green"), ylab="No. de
aspirantes",
```

```

        cex.main=2, cex.names = 2.5)
legend(0.7, 5500, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP1_region, labels=as.character(round(tabla1_region_prop, 2)), c(2,2),
     cex = 2, pos = 3)
abline(h=0)

BP1_nivel socio<-barplot(tabla1_nivel socio, beside = T, main = "
RECEPCIÓN DE DOCUMENTOS
      APROBADOS vs NIVEL SOCIOECONÓMICO", col=c("red", "green"),
ylab="No. de aspirantes",
      cex.main=2, cex.names = 2.5)
legend(0, 4000, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP1_nivel socio, labels=as.character(round(tabla1_nivel socio_prop, 2)),
     c(2,2), cex = 2, pos = 3)
abline(h=0)

BP1_colegio<-barplot(tabla1_colegio, beside = T, main = "      RECEPCIÓN DE
DOCUMENTOS
      APROBADOS vs COLEGIO", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend(4.5, 6000, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP1_colegio, labels=as.character(round(tabla1_colegio_prop, 2)),
     c(2,2), cex = 2, pos = 3)
abline(h=0)

```

Apéndice 5

A continuación, se crean las tablas de frecuencias y proporciones entre el factor aprueba de la etapa 2 Prueba Psicológica y los factores sexo, edad, region, nivelsocio y colegio:

```
tabla2_sexo<-table(etapa_2$aprueba, etapa_2$sexo)
tabla2_sexo
##
##           Mujer Hombre
## No aprueba  178   529
## Aprueba    498   927
tabla2_sexo_prop<-prop.table(tabla2_sexo, 2)
tabla2_sexo_prop
##
##           Mujer   Hombre
## No aprueba 0.2633136 0.3633242
## Aprueba    0.7366864 0.6366758
tabla2_edad<-table(etapa_2$aprueba, etapa_2$edad)
tabla2_edad
##
##           18  19  20  21
## No aprueba 211 201 139 156
## Aprueba   370 393 320 342
tabla2_edad_prop<-prop.table(tabla2_edad, 2)
tabla2_edad_prop
##
##           18      19      20      21
## No aprueba 0.3631670 0.3383838 0.3028322 0.3132530
## Aprueba    0.6368330 0.6616162 0.6971678 0.6867470
tabla2_region<-table(etapa_2$aprueba, etapa_2$region)
tabla2_region
##
##           costa sierra
## No aprueba  263   444
## Aprueba    520   905
tabla2_region_prop<-prop.table(tabla2_region, 2)
tabla2_region_prop
##
##           costa   sierra
## No aprueba 0.3358876 0.3291327
## Aprueba    0.6641124 0.6708673
tabla2_nivelsocio<-table(etapa_2$aprueba, etapa_2$nivelsocio)
tabla2_nivelsocio
##
##           A   B  C+  C-  D
## No aprueba 0  70 373 264  0
## Aprueba    0 275 653 497  0
```

```

tabla2_nivel socio_prop<-prop.table(tabla2_nivel socio, 2)
tabla2_nivel socio_prop
##
##           A           B           C+           C- D
## No aprueba 0.2028986 0.3635478 0.3469120
## Aprueba    0.7971014 0.6364522 0.6530880
tabla2_colegio<-table(etapa_2$aprueba, etapa_2$colegio)
tabla2_colegio
##
##           Público Privado
## No aprueba 430      277
## Aprueba    949      476
tabla2_colegio_prop<-prop.table(tabla2_colegio, 2)
tabla2_colegio_prop
##
##           Público Privado
## No aprueba 0.3118202 0.3678619
## Aprueba    0.6881798 0.6321381

```

Los diagramas de barras correspondientes, en los que se muestran en color verde el número de aspirantes que aprobaron la etapa y en rojo los que no aprobaron, se muestran con el siguiente código:

```

par(mfrow=c(3,2))
BP2_sexo<-barplot(tabla2_sexo, beside = T, main = "           PRUEBA
PSICOLÓGICA
      APROBADOS vs SEXO", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP2_sexo, labels=as.character(round(tabla2_sexo_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP2_edad<-barplot(tabla2_edad, beside = T, main = "           PRUEBA
PSICOLÓGICA
      APROBADOS vs EDAD (AÑOS)", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend(8.6, 460, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP2_edad, labels=as.character(round(tabla2_edad_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP2_region<-barplot(tabla2_region, beside = T, main = "           PRUEBA
PSICOLÓGICA
      APROBADOS vs REGION", col=c("red", "green"), ylab="No. de
aspirantes",

```

```

        cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP2_region, labels=as.character(round(tabla2_region_prop, 2)), c(2,2),
     cex = 2, pos = 3)
abline(h=0)

BP2_nivel socio<-barplot(tabla2_nivel socio, beside = T, main = "          PRUEBA
PSICOLÓGICA
      APROBADOS vs NIVEL SOCIOECONÓMICO", col=c("red", "green"),
ylab="No. de aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP2_nivel socio, labels=as.character(round(tabla2_nivel socio_prop, 2)),
     c(2,2), cex = 2, pos = 3)
abline(h=0)

BP2_colegio<-barplot(tabla2_colegio, beside = T, main = "          PRUEBA
PSICOLÓGICA
      APROBADOS vs COLEGIO", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend(0.7, 1100, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP2_colegio, labels=as.character(round(tabla2_colegio_prop, 2)),
     c(2,2), cex = 2, pos = 3)
abline(h=0)

```

Apéndice 6

A continuación, se crean las tablas de frecuencias y proporciones entre el factor aprueba de la etapa 3 Prueba Académica y los factores sexo, edad, region, nivelsocio y colegio:

```
tabla3_sexo<-table(etapa_3$aprueba, etapa_3$sexo)
tabla3_sexo
##
##           Mujer Hombre
## No aprueba  250   304
## Aprueba    248   623
tabla3_sexo_prop<-prop.table(tabla3_sexo, 2)
tabla3_sexo_prop
##
##           Mujer   Hombre
## No aprueba 0.5020080 0.3279396
## Aprueba    0.4979920 0.6720604
tabla3_edad<-table(etapa_3$aprueba, etapa_3$edad)
tabla3_edad
##
##           18  19  20  21
## No aprueba 170 162 106 116
## Aprueba    200 231 214 226
tabla3_edad_prop<-prop.table(tabla3_edad, 2)
tabla3_edad_prop
##
##           18      19      20      21
## No aprueba 0.4594595 0.4122137 0.3312500 0.3391813
## Aprueba    0.5405405 0.5877863 0.6687500 0.6608187
tabla3_region<-table(etapa_3$aprueba, etapa_3$region)
tabla3_region
##
##           costa sierra
## No aprueba  193   361
## Aprueba     327   544
tabla3_region_prop<-prop.table(tabla3_region, 2)
tabla3_region_prop
##
##           costa   sierra
## No aprueba 0.3711538 0.3988950
## Aprueba    0.6288462 0.6011050
tabla3_nivelsocio<-table(etapa_3$aprueba, etapa_3$nivelsocio)
tabla3_nivelsocio
##
##           A    B  C+  C-  D
## No aprueba  0 109 275 170  0
## Aprueba     0 166 378 327  0
```



```

tabla3_nivel socio_prop<-prop.table(tabla3_nivel socio, 2)
tabla3_nivel socio_prop
##
##           A           B           C+           C- D
## No aprueba 0.3963636 0.4211332 0.3420523
## Aprueba    0.6036364 0.5788668 0.6579477
tabla3_colegio<-table(etapa_3$aprueba, etapa_3$colegio)
tabla3_colegio
##
##           Público Privado
## No aprueba   386   168
## Aprueba     563   308
tabla3_colegio_prop<-prop.table(tabla3_colegio, 2)
tabla3_colegio_prop
##
##           Público Privado
## No aprueba 0.4067439 0.3529412
## Aprueba    0.5932561 0.6470588

```

Los diagramas de barras correspondientes, en los que se muestran en color verde el número de aspirantes que aprobaron la etapa y en rojo los que no aprobaron, se muestran con el siguiente código:

```

BP3_sexo<-barplot(tabla3_sexo, beside = T, main = "          PRUEBA ACADÉMICA
          APROBADOS vs SEXO", col=c("red", "green"), ylab="No. de
aspirantes",
                cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
       fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP3_sexo, labels=as.character(round(tabla3_sexo_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP3_edad<-barplot(tabla3_edad, beside = T, main = "          PRUEBA ACADÉMICA
          APROBADOS vs EDAD (AÑOS)", col=c("red", "green"), ylab="No. de
aspirantes",
                cex.main=2, cex.names = 2.5)
legend(8.55, 250, legend=c("Aprueba", "No aprueba"),
       fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP3_edad, labels=as.character(round(tabla3_edad_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP3_region<-barplot(tabla3_region, beside = T, main = "          PRUEBA
ACADÉMICA
          APROBADOS vs REGION", col=c("red", "green"), ylab="No. de
aspirantes",
                cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
       fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)

```

```

text(BP3_region, labels=as.character(round(tabla3_region_prop, 2)), c(2,2),
cex = 2, pos = 3)
abline(h=0)

BP3_nivelsocio<-barplot(tabla3_nivelsocio, beside = T, main = "      PRUEBA
ACADÉMICA
      APROBADOS vs NIVEL SOCIOECONÓMICO", col=c("red", "green"),
ylab="No. de aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP3_nivelsocio, labels=as.character(round(tabla3_nivelsocio_prop, 2)),
c(2,2), cex = 2, pos = 3)
abline(h=0)

BP3_colegio<-barplot(tabla3_colegio, beside = T, main = "      PRUEBA
ACADÉMICA
      APROBADOS vs COLEGIO", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend(5, 650, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP3_colegio, labels=as.character(round(tabla3_colegio_prop, 2)),
c(2,2), cex = 2, pos = 3)
abline(h=0)

```

Apéndice 7

A continuación, se crean las tablas de frecuencias y proporciones entre el factor aprueba de la etapa 4 Prueba Física y los factores sexo, edad, region, nivelsocio y colegio:

```
tabla4_sexo<-table(etapa_4$aprueba, etapa_4$sexo)
tabla4_sexo
##
##           Mujer Hombre
## No aprueba  132    35
## Aprueba    116   588
tabla4_sexo_prop<-prop.table(tabla4_sexo, 2)
tabla4_sexo_prop
##
##           Mujer    Hombre
## No aprueba 0.53225806 0.05617978
## Aprueba    0.46774194 0.94382022
tabla4_edad<-table(etapa_4$aprueba, etapa_4$edad)
tabla4_edad
##
##           18  19  20  21
## No aprueba  61  78  13  15
## Aprueba    139 153 201 211
tabla4_edad_prop<-prop.table(tabla4_edad, 2)
tabla4_edad_prop
##
##           18          19          20          21
## No aprueba 0.30500000 0.33766234 0.06074766 0.06637168
## Aprueba    0.69500000 0.66233766 0.93925234 0.93362832
tabla4_region<-table(etapa_4$aprueba, etapa_4$region)
tabla4_region
##
##           costa sierra
## No aprueba  40    127
## Aprueba    287   417
tabla4_region_prop<-prop.table(tabla4_region, 2)
tabla4_region_prop
##
##           costa    sierra
## No aprueba 0.1223242 0.2334559
## Aprueba    0.8776758 0.7665441
tabla4_nivelsocio<-table(etapa_4$aprueba, etapa_4$nivelsocio)
tabla4_nivelsocio
##
##           A    B  C+  C-  D
## No aprueba  0  33  71  63  0
## Aprueba    0 133 307 264  0
```

```

tabla4_nivel socio_prop<-prop.table(tabla4_nivel socio, 2)
tabla4_nivel socio_prop
##
##           A           B           C+           C- D
## No aprueba 0.1987952 0.1878307 0.1926606
## Aprueba    0.8012048 0.8121693 0.8073394
tabla4_colegio<-table(etapa_4$aprueba, etapa_4$colegio)
tabla4_colegio
##
##           Público Privado
## No aprueba    115     52
## Aprueba       448    256
tabla4_colegio_prop<-prop.table(tabla4_colegio, 2)
tabla4_colegio_prop
##
##           Público Privado
## No aprueba 0.2042629 0.1688312
## Aprueba    0.7957371 0.8311688

```

Los diagramas de barras correspondientes, en los que se muestran en color verde el número de aspirantes que aprobaron la etapa y en rojo los que no aprobaron, se muestran con el siguiente código:

```

par(mfrow=c(3,2))
BP4_sexo<-barplot(tabla4_sexo, beside = T, main = "           PRUEBA FÍSICA
                APROBADOS vs SEXO", col=c("red", "green"), ylab="No. de
aspirantes",
                cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
       fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP4_sexo, labels=as.character(round(tabla4_sexo_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP4_edad<-barplot(tabla4_edad, beside = T, main = "           PRUEBA FÍSICA
                APROBADOS vs EDAD", col=c("red", "green"), ylab="No. de aspirantes"
,
                cex.main=2, cex.names = 2.5)
legend(0.5, 240, legend=c("Aprueba", "No aprueba"),
       fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP4_edad, labels=as.character(round(tabla4_edad_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP4_region<-barplot(tabla4_region, beside = T, main = "           PRUEBA
FÍSICA
                APROBADOS vs REGION", col=c("red", "green"), ylab="No. de
aspirantes",
                cex.main=2, cex.names = 2.5)

```

```

legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP4_region, labels=as.character(round(tabla4_region_prop, 2)), c(2,2),
cex = 2, pos = 3)
abline(h=0)

BP4_nivelsocio<-barplot(tabla4_nivelsocio, beside = T, main = "      PRUEBA
FÍSICA
      APROBADOS vs NIVEL SOCIOECONÓMICO", col=c("red", "green"),
ylab="No. de aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP4_nivelsocio, labels=as.character(round(tabla4_nivelsocio_prop, 2)),
c(2,2), cex = 2, pos = 3)
abline(h=0)

BP4_colegio<-barplot(tabla4_colegio, beside = T, main = "      PRUEBA
FÍSICA
      APROBADOS vs COLEGIO", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP4_colegio, labels=as.character(round(tabla4_colegio_prop, 2)),
c(2,2), cex = 2, pos = 3)
abline(h=0)

```

Apéndice 8

A continuación, se crean las tablas de frecuencias y proporciones entre el factor aprueba de la etapa 5 Visita al Aspirante y los factores sexo, edad, region, nivelsocio y colegio:

```
tabla5_sexo<-table(etapa_5$aprueba, etapa_5$sexo)
tabla5_sexo
##
##           Mujer Hombre
## No aprueba   33    15
## Aprueba     83   573
tabla5_sexo_prop<-prop.table(tabla5_sexo, 2)
tabla5_sexo_prop
##
##           Mujer   Hombre
## No aprueba 0.2844828 0.0255102
## Aprueba    0.7155172 0.9744898
tabla5_edad<-table(etapa_5$aprueba, etapa_5$edad)
tabla5_edad
##
##           18  19  20  21
## No aprueba 12  24   5   7
## Aprueba  127 129 196 204
tabla5_edad_prop<-prop.table(tabla5_edad, 2)
tabla5_edad_prop
##
##           18           19           20           21
## No aprueba 0.08633094 0.15686275 0.02487562 0.03317536
## Aprueba    0.91366906 0.84313725 0.97512438 0.96682464
tabla5_region<-table(etapa_5$aprueba, etapa_5$region)
tabla5_region
##
##           costa sierra
## No aprueba   34    14
## Aprueba    253   403
tabla5_region_prop<-prop.table(tabla5_region, 2)
tabla5_region_prop
##
##           costa   sierra
## No aprueba 0.11846690 0.03357314
## Aprueba    0.88153310 0.96642686
tabla5_nivelsocio<-table(etapa_5$aprueba, etapa_5$nivelsocio)
tabla5_nivelsocio
##
##           A   B  C+  C-  D
## No aprueba 0   7  22  19  0
## Aprueba    0 126 285 245  0
```

```

tabla5_nivel socio_prop<-prop.table(tabla5_nivel socio, 2)
tabla5_nivel socio_prop
##
##           A           B           C+           C- D
## No aprueba 0.05263158 0.07166124 0.07196970
## Aprueba    0.94736842 0.92833876 0.92803030
tabla5_colegio<-table(etapa_5$aprueba, etapa_5$colegio)
tabla5_colegio
##
##           Público Privado
## No aprueba    30    18
## Aprueba      418   238
tabla5_colegio_prop<-prop.table(tabla5_colegio, 2)
tabla5_colegio_prop
##
##           Público Privado
## No aprueba 0.06696429 0.07031250
## Aprueba    0.93303571 0.92968750

```

Los diagramas de barras correspondientes, en los que se muestran en color verde el número de aspirantes que aprobaron la etapa y en rojo los que no aprobaron, se muestran con el siguiente código:

```

par(mfrow=c(3,2))
BP5_sexo<-barplot(tabla5_sexo, beside = T, main = "          VISITA AL
ASPIRANTE
          APROBADOS vs SEXO", col=c("red", "green"), ylab="No. de
aspirantes",
          cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
       fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP5_sexo, labels=as.character(round(tabla5_sexo_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP5_edad<-barplot(tabla5_edad, beside = T, main = "          VISITA AL
ASPIRANTE
          APROBADOS vs EDAD (AÑOS)", col=c("red", "green"), ylab="No. de
aspirantes",
          cex.main=2, cex.names = 2.5)
legend(0.5, 240, legend=c("Aprueba", "No aprueba"),
       fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP5_edad, labels=as.character(round(tabla5_edad_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP5_region<-barplot(tabla5_region, beside = T, main = "          VISITA AL
ASPIRANTE
          APROBADOS vs REGION", col=c("red", "green"), ylab="No. de
aspirantes",

```

```

        cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP5_region, labels=as.character(round(tabla5_region_prop, 2)), c(2,2),
     cex = 2, pos = 3)
abline(h=0)

BP5_nivel socio<-barplot(tabla5_nivel socio, beside = T, main = "      VISITA
AL ASPIRANTE
      APROBADOS vs NIVEL SOCIOECONÓMICO", col=c("red", "green"),
ylab="No. de aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP5_nivel socio, labels=as.character(round(tabla5_nivel socio_prop, 2)),
     c(2,2), cex = 2, pos = 3)
abline(h=0)

BP5_colegio<-barplot(tabla5_colegio, beside = T, main = "      VISITA AL
ASPIRANTE
      APROBADOS vs COLEGIO", col=c("red", "green"), ylab="No. de
aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP5_colegio, labels=as.character(round(tabla5_colegio_prop, 2)),
     c(2,2), cex = 2, pos = 3)
abline(h=0)

```


Apéndice 9

A continuación, se crean las tablas de frecuencias y proporciones entre el factor aprueba de la etapa 6 Entrevista Final y los factores sexo, edad, region, nivelsocio y colegio:

```
tabla6_sexo<-table(etapa_6$aprueba, etapa_6$sexo)
tabla6_sexo
##
##           Mujer Hombre
## No aprueba   35     21
## Aprueba     48    552
tabla6_sexo_prop<-prop.table(tabla6_sexo, 2)
tabla6_sexo_prop
##
##           Mujer     Hombre
## No aprueba 0.42168675 0.03664921
## Aprueba    0.57831325 0.96335079
tabla6_edad<-table(etapa_6$aprueba, etapa_6$edad)
tabla6_edad
##
##           18  19  20  21
## No aprueba 13  13  12  18
## Aprueba   114 116 184 186
tabla6_edad_prop<-prop.table(tabla6_edad, 2)
tabla6_edad_prop
##
##           18           19           20           21
## No aprueba 0.10236220 0.10077519 0.06122449 0.08823529
## Aprueba    0.89763780 0.89922481 0.93877551 0.91176471
tabla6_region<-table(etapa_6$aprueba, etapa_6$region)
tabla6_region
##
##           costa sierra
## No aprueba   22    34
## Aprueba     231   369
tabla6_region_prop<-prop.table(tabla6_region, 2)
tabla6_region_prop
##
##           costa     sierra
## No aprueba 0.08695652 0.08436725
## Aprueba    0.91304348 0.91563275
tabla6_nivelsocio<-table(etapa_6$aprueba, etapa_6$nivelsocio)
tabla6_nivelsocio
##
##           A    B  C+  C-  D
## No aprueba 0   15  26  15  0
## Aprueba    0  111 259 230  0
```

```

tabla6_nivel socio_prop<-prop.table(tabla6_nivel socio, 2)
tabla6_nivel socio_prop
##
##           A           B           C+           C- D
## No aprueba 0.11904762 0.09122807 0.06122449
## Aprueba    0.88095238 0.90877193 0.93877551
tabla6_colegio<-table(etapa_6$aprueba, etapa_6$colegio)
tabla6_colegio
##
##           Público Privado
## No aprueba    40    16
## Aprueba      378   222
tabla6_colegio_prop<-prop.table(tabla6_colegio, 2)
tabla6_colegio_prop
##
##           Público   Privado
## No aprueba 0.09569378 0.06722689
## Aprueba    0.90430622 0.93277311

```

Los diagramas de barras correspondientes, en los que se muestran en color verde el número de aspirantes que aprobaron la etapa y en rojo los que no aprobaron, se muestran con el siguiente código:

```

par(mfrow=c(3,2))
BP6_sexo<-barplot(tabla6_sexo, beside = T, main = "      ENTREVISTA FINAL
      APROBADOS vs SEXO", col=c("red", "green"), ylab="No. de
      aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP6_sexo, labels=as.character(round(tabla6_sexo_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP6_edad<-barplot(tabla6_edad, beside = T, main = "      ENTREVISTA FINAL
      APROBADOS vs EDAD (AÑOS)", col=c("red", "green"), ylab="No. de
      aspirantes",
      cex.main=2, cex.names = 2.5)
legend(0.5, 210, legend=c("Aprueba", "No aprueba"),
      fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP6_edad, labels=as.character(round(tabla6_edad_prop, 2)), c(2,2), cex
= 2, pos = 3)
abline(h=0)

BP6_region<-barplot(tabla6_region, beside = T, main = "      ENTREVISTA
      FINAL
      APROBADOS vs REGION", col=c("red", "green"), ylab="No. de
      aspirantes",
      cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),

```

```

        fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP6_region, labels=as.character(round(tabla6_region_prop, 2)), c(2,2),
cex = 2, pos = 3)
abline(h=0)

BP6_nivelsocio<-barplot(tabla6_nivelsocio, beside = T, main = "
ENTREVISTA FINAL
    APROBADOS vs NIVEL SOCIOECONÓMICO", col=c("red", "green"),
ylab="No. de aspirantes",
    cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
    fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP6_nivelsocio, labels=as.character(round(tabla6_nivelsocio_prop, 2)),
c(2,2), cex = 2, pos = 3)
abline(h=0)

BP6_colegio<-barplot(tabla6_colegio, beside = T, main = "          ENTREVISTA
FINAL
    APROBADOS vs COLEGIO", col=c("red", "green"), ylab="No. de
aspirantes",
    cex.main=2, cex.names = 2.5)
legend("topleft", legend=c("Aprueba", "No aprueba"),
    fill=c("green", "red") , cex = 1.1, bty = "n", x.intersp = 0.2)
text(BP6_colegio, labels=as.character(round(tabla6_colegio_prop, 2)),
c(2,2), cex = 2, pos = 3)
abline(h=0)

```

Apéndice 10

En principio se obtiene un modelo logístico binomial con el $\text{logit}(p)$ como variable dependiente, donde p es la probabilidad condicional de ingresar a la Escuela Militar, dados los valores de los factores de predicción que son: sexo, edad, region, nivel socio y colegio. A continuación, se aplica la regresión por pasos para encontrar las variables predictoras que más influyen en el factor respuesta:

```
modelo<-glm(ingresa ~ sexo + edad + region + nivel socio + colegio, family = binomial(link = logit), data = datos)
summary(modelo)
summary(step(modelo))
```

Luego de aplicar la regresión por pasos se puede observar que los factores que son significativos en el factor respuesta son sexo, edad y colegio. Se construye el modelo global con estos tres factores, la matriz de confusión y se calcula la tasa de aciertos en el pronóstico del modelo global:

```
modeloglobal<-glm(ingresa ~ sexo + edad + colegio, family = binomial(link = logit), data = datos)
summary(modeloglobal)
probabilidad<-predict(modeloglobal, type = "response")
pronosticoglobal<-ifelse(probabilidad<0.5, "No ingresa", "Ingresa")
table(ingresa, pronosticoglobal)
aciertos<-ifelse(ingresa==pronosticoglobal, 1,0)
tasa_aciertos<-sum(aciertos)/length(aciertos)
tasa_aciertos
## [1] 0.9384994
anova(modeloglobal, test="Chisq")
R2<-1-(modeloglobal$deviance/modeloglobal>null.deviance)
R2
## [1] 0.01191467
1-pchisq(modeloglobal>null.deviance-modeloglobal$deviance, modeloglobal$df.null-modeloglobal$df.residual) # Comparando con el modelo nulo
## [1] 1.287126e-11
1-pchisq(modeloglobal$deviance, modeloglobal$df.residual) # Comparando con el modelo saturado
## [1] 1
```

Apéndice 11

En principio se obtiene un modelo logístico binomial con el logit(P) como variable dependiente, donde p es la probabilidad condicional del aprobar la etapa dados los valores de los factores de predicción que son: sexo, edad, region, nivelsocio y colegio. A continuación, se aplica la regresión por pasos para encontrar las variables predictoras que más influyen en el factor respuesta:

```
modelo<-glm(aprueba ~ sexo + edad + region + nivelsocio + colegio, family = binomial(link = logit), data = etapa_1)
summary(modelo)
summary(step(modelo))
```

Luego de aplicar la regresión por pasos se puede observar que los factores que son significativos en la respuesta de esta etapa son sexo y colegio. Se construye el modelo 1 (por etapa 1) con estos dos factores, la matriz de confusión y se calcula la tasa de aciertos en el pronóstico del modelo 1:

```
modelo1<-glm(aprueba ~ sexo + colegio, family = binomial(link = logit), data = etapa_1)
summary(modelo1)
anova(modelo1, test="Chisq")
R2<-1-(modelo1$deviance/modelo1>null.deviance)
R2
## [1] 0.1059974
1-pchisq(modelo1>null.deviance-modelo1$deviance, modelo1$df.null-modelo1$df.residual) # Comparando con el modelo nulo
## [1] 0
1-pchisq(modelo1$deviance, modelo1$df.residual) # Comparando con el modelo saturado
## [1] 0.9999929
```

Apéndice 12

En principio se obtiene un modelo logístico binomial con el logit(P) como variable dependiente, donde p es la probabilidad condicional del aprobar la etapa dados los valores de los factores de predicción que son: sexo, edad, region, nivelsocio y colegio. A continuación, se aplica la regresión por pasos para encontrar las variables predictoras que más influyen en el factor respuesta:

```
modelo<-glm(aprueba ~ sexo + edad + region + nivelsocio + colegio, family =  
binomial(link = logit), data = etapa_2)  
summary(modelo)  
summary(step(modelo))
```

Luego de aplicar la regresión por pasos se puede observar que los factores que son significativos en la respuesta de esta etapa son: sexo, edad, nivelsocio y colegio. Se construye el modelo 2 (por etapa 2) con estos cuatro factores, la matriz de confusión y se calcula la tasa de aciertos en el pronóstico del modelo 2:

```
modelo2<-glm(aprueba ~ sexo + edad + nivelsocio + colegio, family =  
binomial(link = logit), data = etapa_2)  
summary(modelo2)  
aciertos<-ifelse(etapa_2$aprueba==pronostico2, 1, 0)  
tasa_aciertos<-sum(aciertos)/length(aciertos)  
tasa_aciertos  
## [1] 0.6683865  
anova(modelo2, test="Chisq")  
R2<-1-(modelo2$deviance/modelo2>null.deviance)  
R2  
## [1] 0.02510099  
1-pchisq(modelo2>null.deviance-modelo2$deviance, modelo2$df.null-  
modelo2$df.residual) # Comparando con el modelo nulo  
## [1] 2.672307e-13  
1-pchisq(modelo2$deviance, modelo2$df.residual) # Comparando con el  
modelo saturado  
## [1] 1.011413e-13
```

Apéndice 13

En principio se obtiene un modelo logístico binomial con el logit(P) como variable dependiente, donde p es la probabilidad condicional del aprobar la etapa dados los valores de los factores de predicción que son: sexo, edad, region, nivelsocio y colegio. A continuación, se aplica la regresión por pasos para encontrar las variables predictoras que más influyen en el factor respuesta:

```
modelo<-glm(aprueba ~ sexo + edad + region + nivelsocio + colegio, family =  
binomial(link = logit), data = etapa_3 )  
summary(modelo)  
summary(step(modelo))
```

Luego de aplicar la regresión por pasos se puede observar que los factores que son significativos en la respuesta de esta etapa son: sexo y edad. Se construye el modelo 3 (por etapa 3) con estos dos factores, la matriz de confusión y se calcula la tasa de aciertos en el pronóstico del modelo 3:

```
modelo3<-glm(aprueba ~ sexo + edad , family = binomial(link = logit), data  
=  
etapa_3)  
summary(modelo3)  
pronostico<-predict(modelo3, type = "response")  
pronostico3<-ifelse(pronostico<0.5, "No aprueba", "Aprueba")  
table(etapa_3$aprueba, pronostico3)  
aciertos<-ifelse(etapa_3$aprueba==pronostico3, 1, 0)  
tasa_aciertos<-sum(aciertos)/length(aciertos)  
tasa_aciertos  
## [1] 0.6035088  
anova(modelo3,test="Chisq")  
R2<-1-(modelo3$deviance/modelo3>null.deviance)  
R2  
## [1] 0.02605566  
1-pchisq(modelo3>null.deviance-modelo3$deviance, modelo3$df.null-  
modelo3$df.residual) # Comparando con el modelo nulo  
## [1] 1.680001e-11  
1-pchisq(modelo3$deviance, modelo3$df.residual) # Comparando con el  
modelo saturado  
## [1] 5.662137e-14
```

Apéndice 14

En principio se obtiene un modelo logístico binomial con el logit(P) como variable dependiente, donde p es la probabilidad condicional del aprobar la etapa dados los valores de los factores de predicción que son: sexo, edad, region, nivelsocio y colegio. A continuación, se aplica la regresión por pasos para encontrar las variables predictoras que más influyen en el factor respuesta:

```
modelo<-glm(aprueba ~ sexo + edad + region + nivelsocio + colegio, family =  
binomial(link = logit), data = etapa_4 )  
summary(modelo)  
summary(step(modelo))
```

Luego de aplicar la regresión por pasos se puede observar que los factores que son significativos en la respuesta de esta etapa son sexo, edad y región. Se construye el modelo 4 (por etapa 4) con estos tres factores, la matriz de confusión y se calcula la tasa de aciertos en el pronóstico del modelo 4:

```
modelo4<-glm(aprueba ~ sexo + edad + region, family = binomial(link =  
logit), data = etapa_4)  
summary(modelo4)  
pronostico<-predict(modelo4, type = "response")  
pronostico4<-ifelse(pronostico<0.5, "No aprueba", "Aprueba")  
table(etapa_4$aprueba, pronostico4)  
##  
##  
##          No aprueba          76  
## Aprueba    657          47          91  
aciertos<-ifelse(etapa_4$aprueba==pronostico4, 1, 0)  
tasa_aciertos<-sum(aciertos)/length(aciertos)  
tasa_aciertos  
## [1] 0.858783  
anova(modelo4, test="Chisq")  
R2<-1-(modelo4$deviance/modelo4>null.deviance)  
R2  
## [1] 0.3496225  
1-pchisq(modelo4>null.deviance-modelo4$deviance, modelo4$df.null-  
modelo4$df.residual) # Comparando con el modelo nulo  
## [1] 0  
1-pchisq(modelo4$deviance, modelo4$df.residual) # Comparando con el  
modelo saturado  
## [1] 1
```


Apéndice 15

En principio se obtiene un modelo logístico binomial con el logit(P) como variable dependiente, donde p es la probabilidad condicional del aprobar la etapa dados los valores de los factores de predicción que son: sexo, edad, region, nivelsocio y colegio como. A continuación, se aplica la regresión por pasos para encontrar las variables predictoras que más influyen en el factor respuesta:

```
modelo<-glm(aprueba ~ sexo + edad + region + nivelsocio + colegio, family = binomial(link = logit), data = etapa_5)
summary(modelo)
summary(step(modelo))
```

Luego de aplicar la regresión por pasos se puede observar que los factores que son significativos en la respuesta de esta etapa son sexo, edad y región. Se construye el modelo 5 (por etapa 5) con estos tres factores, la matriz de confusión y se calcula la tasa de aciertos en el pronóstico del modelo 5:

```
modelo5<-glm(aprueba ~ sexo + edad + region, family = binomial(link = logit), data = etapa_5)
summary(modelo5)
pronostico<-predict(modelo5, type = "response")
pronostico5<-ifelse(pronostico<0.5, "No aprueba", "Aprueba")
table(etapa_5$aprueba, pronostico5)
aciertos<-ifelse(etapa_5$aprueba==pronostico5, 1, 0)
tasa_aciertos<-sum(aciertos)/length(aciertos)
tasa_aciertos
## [1] 0.9318182
anova(modelo5, test="Chisq")
R2<-1-(modelo5$deviance/modelo5>null.deviance)
R2
## [1] 0.286687
1-pchisq(modelo5>null.deviance-modelo5$deviance, modelo5$df.null-modelo5$df.residual) # Comparando con el modelo nulo
## [1] 0
1-pchisq(modelo5$deviance, modelo5$df.residual) # Comparando con el modelo saturado
## [1] 1
```

Apéndice 16

En principio se obtiene un modelo logístico binomial con el logit(P) como variable dependiente, donde p es la probabilidad condicional del aprobar la etapa dados los valores de los factores de predicción que son: sexo, edad, region, nivelsocio y colegio. A continuación, se aplica la regresión por pasos para encontrar las variables predictoras que más influyen en el factor respuesta:

```
modelo<-glm(aprueba ~ sexo + edad + region + nivelsocio + colegio, family =  
binomial(link = logit), data = etapa_6 )  
summary(modelo)
```

Luego de aplicar la regresión por pasos se puede observar que el factor que es significativo en la respuesta de esta etapa es sexo. Se construye el modelo 6 (por etapa 6) con este factor, la matriz de confusión y se calcula la tasa de aciertos en el pronóstico del modelo 6:

```
modelo6<-glm(aprueba ~ sexo , family = binomial(link = logit), data =  
etapa_6)  
summary(modelo6)  
pronostico<-predict(modelo6, type = "response")  
pronostico6<-ifelse(pronostico<0.5, "No aprueba", "Aprueba")  
table(etapa_6$aprueba, pronostico6)  
aciertos<-ifelse(etapa_6$aprueba==pronostico6, 1, 0)  
tasa_aciertos<-sum(aciertos)/length(aciertos)  
tasa_aciertos  
## [1] 0.9146341  
anova(modelo6, test="Chisq")  
R2<-1-(modelo6$deviance/modelo6>null.deviance)  
R2  
## [1] 0.2340864  
1-pchisq(modelo6>null.deviance-modelo6$deviance, modelo6$df.null-  
modelo6$df.residual) # Comparando con el modelo nulo  
## [1] 0  
1-pchisq(modelo6$deviance, modelo6$df.residual) # Comparando con el  
modelo saturado  
## [1] 1
```

Apéndice 17

Se crea el dataframe aspirante con los valores de los factores indicados:

```
aspirante<-data.frame(sexo=c("Hombre"), edad= c(21), region=c("sierra"),  
nivel socio=c("C+"), colegio=c("Privado"))
```

Puesto que el modelo global tiene como factores significativos a sexo, edad y colegio, creamos el aspirante global sólo con esos factores y realizamos la predicción de la probabilidad de ingresar con el modelo global condicionado a esos valores:

```
aspirante_global<-aspirante[, c("sexo", "edad", "colegio")]  
probabilidad_global<-predict.glm(modeloglobal, newdata = aspirante_global,  
type="response")  
probabilidad_global  
## 1  
## 0.1038559
```

El valor de la probabilidad condicional por etapas de ingresar a la Escuela Militar se calcula mediante el producto de las probabilidades condicionales de aprobar cada una de las 6 etapas:

```
probabilidad_etapas<-  
probabilidad_1*probabilidad_2*probabilidad_3*probabilidad_4*probabilidad_5*  
probabilidad_6  
probabilidad_etapas  
## 1  
## 0.08250202
```

Apéndice 18

• Temario de Matemáticas

FUNDAMENTOS DE ALGEBRA.

- 1.1 Números reales
- 1.2 Sucesiones y progresiones
- 1.3 Potenciación y radicación
- 1.4 Expresiones algebraicas.
- 1.5 Ecuaciones
- 1.6 Inecuaciones

FUNCIONES

- 2.1 Conceptos y definiciones de funciones
- 2.2 Clasificación de las funciones
- 2.3 Graficas de funciones: dominio y recorrido
- 2.4 Funciones algebraicas:
- 2.5 Funciones especiales
- 2.6 Funciones Trigonométricas

TRIGONOMETRÌA

- 3.1 Identidades trigonométricas
- 3.2 Ecuaciones trigonométricas
- 3.3 Trigonometría de triángulos rectángulos
- 3.4 Trigonometría de triángulos en general

VECTORES BIDIMENSIONALES Y TRIDIMENSIONALES

- 4.1 Vectores bidimensionales
- 4.2 Vectores tridimensionales
- 4.3 Suma y resta de vectores
- 4.4 Producto punto
- 4.5 Producto cruz

MATRICES Y SISTEMAS DE ECUACIONES E INECUACIONES

- 5.1 Matrices y determinantes.
- 5.2 Algebra de matrices
- 5.3 Matriz inversa
- 5.4 Sistemas de ecuaciones con varias incógnitas
- 5.5 Sistemas de ecuaciones no lineales
- 5.6 Sistemas de inecuaciones

GEOMETRÌA ANALÍTICA

- 6.1 Rectas en el plano

6.2 Circunferencias

6.3 Parábolas

6.4 Elipses

6.5 Hipérbolas

NÚMEROS COMPLEJOS

7.1 Definiciones de números imaginarios y complejos

7.2 Potencias de números imaginarios

7.3 Operaciones de números complejos

7.4 Representación geométrica de un numero complejo

7.5 Notación de Euler

GEOMETRÍA PLANA Y DEL ESPACIO

8.1 Rectas en el plano

8.2 Ángulos: clasificación.

8.3 Semejanza y congruencia (Teorema de Thales)

8.4 Polígonos

8.5 Rectas y planos en el espacio

8.6 Cuerpos geométricos

8.7 Circunferencia y círculo

8.8 Sólidos de revolución

LÍMITES

9.1 Definición de límites

9.2 Propiedades de los límites

9.3 Cálculo de límites numérica y gráficamente

9.4 Cálculo de límites algebraicamente

9.5 Cálculo de límites al infinito

• Temario de Física

MEDICIÓN

1.1 Análisis dimensional.

1.2 Conversión de unidades.

1.3 Cifras significativas.

VECTORES

2.1 Vectores en \mathbb{R}^2

2.2 Vectores en \mathbb{R}^3

CINEMÁTICA.

Cinemática en una dimensión.

3.1 Marcos de referencia.

3.2 Desplazamiento, velocidad y rapidez, velocidad instantánea. (MRU)

3.3 Aceleración. (MRUV)

3.4 Caída libre.

3.5 Diagramas de movimiento.

Cinemática en dos dimensiones.

3.6 Movimiento de proyectiles.

3.7 Velocidad relativa.

3.8 Movimiento circular uniforme.

LEYES DEL MOVIMIENTO.

4.1 Fuerza.

4.2 Primera ley de Newton.

4.3 Sistemas de referencia inercial.

4.4 Segunda ley de Newton.

4.5 Tercera ley de Newton

TRABAJO Y ENERGIA.

5.1 Trabajo.

5.2 Tipos de Energía

5.3 Teorema del trabajo y la energía.

5.+4 Fuerzas conservativas y no conservativas.

5.5 Trabajo de la fricción.

5.6 Energía mecánica y su conservación.

5.7 Potencia.

CAMPOS ELECTRICOS.

6.1 Propiedades de las cargas eléctricas.

6.2 Aislantes y conductores.

6.3 Carga inducida.

6.4 Ley de Coulomb.

6.5 Campo eléctrico.

6.6 Líneas de campo eléctrico.

.

• Temario de Química

1 MATERIA-ENERGÍA Y SUS CAMBIOS

2 ESTRUCTURA ATÓMICA

3 CLASIFICACIÓN PERIÓDICA DE LOS ELEMENTOS.

4 ENLACES QUÍMICOS

5 NOMENCLATURA DE COMPUESTOS INORGÁNICOS

6 CÁLCULOS ESTEQUIOMÉTRICOS

- 7 REACCIONES QUÍMICAS
- 8 LEYES DE LOS GASES
- 9 SOLUCIONES
- 10 EQUILIBRIO ÁCIDO-BASE