

# UNIVERSIDAD TÉCNICA DE AMBATO



## FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

### MAESTRÍA EN MATEMÁTICA APLICADA

---

**Tema:** “MODELO MATEMÁTICO PARA DETERMINAR LA CALIDAD  
DE SERVICIO EN EL TRANSPORTE PÚBLICO URBANO EN LA  
CIUDAD DE AMBATO”

---

Trabajo de titulación previo a la obtención del grado académico de Magíster en  
Matemática Aplicada

Modalidad de titulación proyecto de desarrollo

**Autor:** Ing. Orlando Ruben Toscano Ramos

**Director:** Dr. Freddy Geovanny Benalcázar Palacios, Mg.

Ambato – Ecuador

2021

## **APROBACIÓN DEL TRABAJO DE TITULACIÓN**

A la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial. El Tribunal receptor de la Defensa del Trabajo de Titulación presidido por la Ingeniera Elsa Pilar Urrutia Urrutia Mg., e integrado por los señores: Ingeniero Ernesto Antonio Ponsot Balaguer Doctor, Ingeniera Clara Augusta Sánchez Benítez Magister, designados por la Unidad Académica de Titulación de Posgrados de la Facultad de Ingeniería en Sistemas Electrónica e Industrial de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: “Modelo matemático para determinar la calidad de servicio en el transporte público urbano en la ciudad de Ambato”, elaborado y presentado por el Ing. Orlando Ruben Toscano Ramos, para optar por el Grado Académico de Magister en Matemática Aplicada; una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la Universidad Técnica de Ambato.

-----  
Ing. Elsa Pilar Urrutia Urrutia, Mg.  
Presidente y Miembro del Tribunal de Defensa

-----  
Ing. Ernesto Antonio Ponsot Balaguer Dr.  
Miembro del Tribunal de Defensa

-----  
Ing. Clara Augusta Sánchez Benítez Mg.  
Miembro del Tribunal de Defensa

## **AUTORÍA DEL TRABAJO DE TITULACIÓN**

La responsabilidad de las opiniones, comentarios y críticas emitidas en el trabajo de titulación presentado con el tema: “Modelo matemático para determinar la calidad de servicio en el transporte público urbano en la ciudad de Ambato”, le corresponde exclusivamente a: Ing. Orlando Ruben Toscano Ramos, autor bajo la dirección de: Dr. Freddy Geovanny Benalcázar Palacios, Mg, director del trabajo de investigación; y el patrimonio intelectual pertenece a la Universidad Técnica de Ambato

-----  
Ing. Orlando Ruben Toscano Ramos

**AUTOR**

-----  
Dr. Freddy Geovanny Benalcázar Palacios, Mg

**DIRECTOR**

## **DERECHOS DE AUTOR**

Autorizo a la Universidad Técnica de Ambato, para que el trabajo de titulación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos de mi trabajo de titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.

-----  
Ing. Orlando Ruben Toscano Ramos  
1803364783

## ÍNDICE GENERAL

CONTENIDO	Pág.
PORTADA .....	i
APROBACIÓN DEL TRABAJO DE TITULACIÓN .....	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN .....	iii
DERECHOS DE AUTOR.....	iv
ÍNDICE DE TABLAS .....	vii
ÍNDICE DE FIGURAS.....	viii
AGRADECIMIENTO .....	ix
DEDICATORIA .....	x
RESUMEN EJECUTIVO .....	xi
EXECUTIVE SUMMARY .....	xiii
CAPÍTULO I .....	1
EL PROBLEMA DE INVESTIGACIÓN .....	1
1.1.    Introducción .....	1
1.2.    Justificación .....	2
1.3.    Objetivos .....	3
1.3.1.    Objetivo General .....	3
1.3.2.    Objetivos Específicos .....	3
CAPÍTULO II.....	4
ANTECEDENTES INVESTIGATIVOS .....	4
2.1.    Estado del Arte .....	4
2.2.    Marco teórico .....	10
2.2.1.    Regresión logística. ....	10
2.2.2.    Regresión logística binomial. ....	10
2.2.3.    Estimación del modelo .....	11
2.2.4.    Odds Ratio.....	13
2.2.5.    Matriz de confusión.....	15
CAPÍTULO III.....	16

MARCO METODOLÓGICO .....	16
3.1. Ubicación .....	16
3.2. Equipos y materiales .....	17
3.3. Tipo de investigación .....	19
3.4. Problema e hipótesis.....	20
3.5. Población o muestra .....	20
3.6. Recolección de información .....	21
3.7. Procesamiento y análisis estadístico .....	21
3.8. Análisis descriptivo de datos .....	22
3.9. Variables respuestas o resultados alcanzados .....	28
CAPÍTULO IV .....	30
RESULTADOS Y DISCUSIONES .....	30
4.1. Análisis de las variables .....	30
4.2. Construcción del modelo.....	33
4.3. Diagnóstico de la multicolinealidad en las variables predictoras .....	34
4.4. Bondad de ajuste del modelo.....	35
4.5. Residuos del modelo .....	36
4.6. Interpretación de los odds ratio (OR) .....	37
4.7. Determinación del punto de corte y la curva Roc .....	39
4.8. Matriz de confusión.....	40
4.9. Predicciones .....	41
CAPÍTULO V .....	44
CONCLUSIONES Y RECOMENDACIONES .....	44
5.1. CONCLUSIONES .....	44
5.2. RECOMENDACIONES .....	45
5.3. BIBLIOGRAFÍA.....	46
5.4. ANEXOS .....	49

## ÍNDICE DE TABLAS

	<u>Pág.</u>
Tabla 1. Rutas de las cooperativas de transporte urbano de Ambato.....	16
Tabla 2. Variables de estudio .....	21
Tabla 3. Variable Tiempo de espera .....	23
Tabla 4. Variable Trato al usuario.....	24
Tabla 5. Variable Estado actual de las unidades .....	25
Tabla 6. Variable Forma de conducción del transportista.....	26
Tabla 7. Variable Calidad de servicio .....	27
Tabla 8. Resultados del Modelo.....	29
Tabla 9. Transformación de los resultados a variables dicotómicas.....	30
Tabla 10. Estructura de la base de datos .....	31
Tabla 11. Resumen del modelo de regresión logística.....	33
Tabla 12. Multicolinealidad para medir si las variables son las adecuadas .....	34
Tabla 13. Estadístico ji-cuadrado.....	35
Tabla 14. Estadísticos para determinar el ajuste del modelo .....	36
Tabla 15. Interpretación de los odds ratio (OR).....	37
Tabla 16. Punto de corte y área bajo la curva Roc .....	39
Tabla 17. Predicción y clasificación .....	42

## ÍNDICE DE FIGURAS

	Pág.
Figura 1. Representación gráfica de la curva Roc.....	15
Figura 2. Matriz de confusión o tabla de contingencia .....	15
Figura 3. Análisis Descriptivo de la variable Tiempo de espera.....	23
Figura 4. Análisis Descriptivo de la variable Trato al usuario.....	24
Figura 5. Análisis Descriptivo de la variable Estado actual de las unidades. ....	25
Figura 6. Análisis Descriptivo de la variable Forma de conducción del transportista. .....	26
Figura 7. Análisis Descriptivo de la variable Calidad del servicio.....	27
Figura 8. Correlación entre las variables de estudio .....	32
Figura 9. Residuos del modelo.....	37
Figura 10. Curva ROC. ....	40
Figura 11. Matriz de confusión. ....	41
Figura 12. Comportamiento de las predicciones.....	43
Figura 13. Ciudad de Ambato. ....	50
Figura 14. Línea de autobuses de Ambato. ....	50
Figura 15. Base de datos binarizada.....	50
Figura 16. Parada el hospital.....	50
Figura 17. Parada en el redondel de Izamba. ....	50
Figura 18. Parada en el mercado mayorista. ....	50
Figura 19. Parada en el mercado modelo. ....	50
Figura 20. Parada en el Moll de los Andes - Municipio. ....	50
Figura 21. Parada en el redondel de Huachi Chico.....	50
Figura 22. Script del modelo.....	50

## **AGRADECIMIENTO**

En primer lugar, agradezco a Dios por guiarme en todo momento y permitirme llegar a esta instancia tan añorada de mi vida.

A mis adorados padres por su infinito amor y apoyo incondicional en todos los momentos, ya que desde pequeño me han inculcado valores y principios que los he sabido llevar, llegando a poder cristalizar este grado de superación personal tan importante para mí.

A mi querida familia por estar pendientes y creer siempre en mí y brindarme en todo momento la mano y las palabras necesarias para seguir adelante.

A todos los docentes de la maestría por proporcionarme sus valiosos conocimientos en las diferentes asignaturas que impartieron en cada módulo, en especial al Dr. Freddy Geovanny Benalcázar Palacios, Mg por su tiempo y guía en este proceso de titulación.

A mis compañeros de maestría con los cuales se formó un buen equipo de trabajo a Daysi y Carlos por su valioso apoyo y aporte.

Ing. Orlando Ruben Toscano Ramos

## **DEDICATORIA**

La presente tesis dedico en primer lugar a Dios por ser siempre mi fortaleza y guía para no desmayar y luchar en la consecución de mis objetivos.

A mis queridos padres Edgar Toscano y Ilda María por su amor, enseñanza, consejos, apoyo total y el estar pendientes en cada momento de mi vida.

A mis hermanos Santy, Liliana e Irene, a mis sobrinitos, mis cuñados por su amor y apoyo incondicional, por ser el motor constante y darme el ánimo para no desmayar y continuar siempre creciendo profesionalmente.

Ing. Orlando Ruben Toscano Ramos

**UNIVERSIDAD TÉCNICA DE AMBATO**  
**FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E**  
**INDUSTRIAL**  
**MAESTRÍA EN MATEMÁTICA APLICADA**

**TEMA:**

“MODELO MATEMÁTICO PARA DETERMINAR LA CALIDAD DE SERVICIO  
EN EL TRANSPORTE PÚBLICO URBANO EN LA CIUDAD DE AMBATO”

**AUTOR:** Ing. Orlando Ruben Toscano Ramos.

**DIRECTOR:** Dr. Freddy Geovanny Benalcázar Palacios, Mg.

**LÍNEA DE INVESTIGACIÓN:** Modelamiento matemático.

**FECHA:**

**RESUMEN EJECUTIVO**

La presente investigación es de tipo cuantitativa, descriptiva y correlacional. El objetivo fue construir un modelo matemático que permita determinar la calidad del servicio del transporte público urbano en la ciudad de Ambato en el periodo 2020 - 2021. Los datos fueron obtenidos mediante la técnica de la encuesta con su instrumento el cuestionario, la misma que se aplicó a 400 usuarios que utilizan el transporte público en la ciudad de Ambato. Para lograr los objetivos propuestos en la investigación se aplicó el modelo de regresión logística binaria, que ayuda a clasificar y predecir la calidad del servicio del transporte público urbano en la ciudad de Ambato.

Las variables involucradas en esta investigación fueron: como variable dependiente la calidad del servicio y como variables predictoras o independientes: el tiempo de espera, el trato al usuario, el estado actual de las unidades y la forma de conducción del transportista. Para el procesamiento y el análisis de los datos se empleó el programa Microsoft Excel 2016 y el software libre RStudio versión 4.0.1.

Se concluye que el modelo de regresión logística construido se ajusta correctamente a las variables predictoras, la frontera de decisión arrojada por el modelo fue de 0,6142 o el 61,42% que sirve para determinar la calidad del servicio en seguro e inseguro. El modelo clasifica correctamente 308 observaciones, de las cuales 205 observaciones

son clasificadas como seguras representando el 51,25% de los datos totales, mientras 103 observaciones fueron clasificadas como inseguras, que equivalen al 25,75%.

Con el análisis efectuado se concluye que el transporte público urbano de la ciudad de Ambato es medianamente seguro, por lo cual las autoridades pertinentes deberían poner énfasis en mejorar la calidad del servicio.

**Descriptores:** Calidad del servicio, correlacional, modelo logit, modelo matemático, regresión logística, transporte público,

**TECHNICAL UNIVERSITY OF AMBATO**  
**FACULTY OF SYSTEMS, ELECTRONIC AND INDUSTRIAL**  
**ENGINEERING**  
**MASTER'S OF APPLIED MATHEMATICS**

**THEME:**

"MATHEMATICAL MODEL TO DETERMINE THE QUALITY OF SERVICE IN  
URBAN PUBLIC TRANSPORTATION IN THE CITY OF AMBATO"

**AUTHOR:** Ing. Orlando Ruben Toscano Ramos

**DIRECTED BY:** Dr. Freddy Geovanny Benalcázar Palacios, Mg.

**RESEARCH LINE:** Mathematical modeling

**DATE:**

**EXECUTIVE SUMMARY**

This research is quantitative, descriptive and correlational. The objective was to build a mathematical model that allows determining the quality of the urban public transport service in the city of Ambato in the period 2020 - 2021. The data were obtained through the survey technique with its instrument the questionnaire, the same as applied to 400 users who use public transport in the city of Ambato. To achieve the objectives proposed in the research, the binary logistic regression model was applied, which helps to classify and predict the quality of the urban public transport service in the city of Ambato

The variables involved in this research were: as a dependent variable the quality of the service and as predictive or independent variables: the waiting time, the treatment of the user, the current state of the units and the way of driving of the carrier. For the processing and analysis of the data, the Microsoft Excel 2016 program and the free software RStudio version 4.0.1 were used.

It is concluded that the constructed logistic regression model correctly adjusts to the predictor variables, the decision frontier yielded by the model was 0.6142 or 61.42%, which is used to determine the quality of the safe and unsafe service. The model correctly classifies 308 observations, of which 205 observations are classified as safe, representing 51.25% of the total data, while 103 observations were classified as unsafe, which is equivalent to 25.75%.

With the analysis carried out, it was determined that the urban public transport of the city of Ambato is moderately safe, for which the pertinent authorities should place emphasis on improving the quality of the service.

**Keywords:** Quality of service, correlational, logit model, mathematical model, logistic regression, public transport.

# CAPÍTULO I

## EL PROBLEMA DE INVESTIGACIÓN

### 1.1. Introducción

Esta investigación se encuentra enfocada en la construcción de un modelo matemático que permita determinar la calidad del servicio del transporte público urbano en la ciudad de Ambato.

Para el desarrollo del modelo matemático se empleó el modelo de regresión logística binaria. Como primer punto se estableció cual es la relación existente entre las variables calidad de servicio, tiempo de espera, trato al usuario, estado actual de las unidades y la forma de conducción del transportista. Posteriormente con el software libre RStudio versión 4.0.1 que utiliza el método de máxima verosimilitud se determinó si los coeficientes del modelo son significativos. A continuación, se aplicaron diferentes pruebas estadísticas para medir la bondad de ajuste del modelo como el estadístico ji-cuadrado, el test de Hosmer y Lemeshow, la prueba de Nagelkerke y test de Cox y Snell. Para la interpretación de los coeficientes del modelo se realizó la transformación de base logarítmica a exponencial y así observar cómo inciden las variables predictoras sobre la variable dependiente. Por último, para clasificar y predecir se encontró el punto de corte, la curva roc y la matriz de confusión con lo cual se determinó que el modelo clasifica correctamente 308 observaciones, de las cuales 205 observaciones son clasificadas como seguras y representan el 51,25% de los datos totales, mientras que 103 observaciones fueron clasificadas como inseguras, que equivalen al 25,75%, 92 observaciones no se clasificaron correctamente debido a que las variables predictoras pueden encontrarse medianamente correlacionadas.

Este documento está compuesto de cinco capítulos, en el primer capítulo se presenta la justificación y los objetivos de la investigación; en el segundo capítulo se da a conocer el estado del arte y la sustentación teórica más importante para desarrollar el modelo; el tercer capítulo hace referencia a la metodología aplicada; en el cuarto capítulo se presentan los resultados y las discusiones realizadas en base del modelo

desarrollado; en el quinto y último capítulo se da a conocer las conclusiones y recomendaciones a las que se llegó al realizar el presente trabajo de titulación.

## **1.2. Justificación**

El transporte tiene relación muy estrecha con el desarrollo económico del país. Los servicios de transporte son importantes para el desarrollo de la sociedad, ya que hacen posible el acceso a recursos, bienes, insumos, etc. que de otra manera no serían asequibles por razones de distancia [1].

En la actualidad los modelos matemáticos son usados para analizar la relación entre 2 o más variables, por ejemplo, para minimizar efectos negativos y promover el uso de recursos tomados en base a distintos parámetros que tienen relación con la calidad de servicio. Los modelos matemáticos describen la relación entre variables de decisión planteadas, restricciones y objetivos, su presentación es en forma de ecuaciones matemáticas, en donde el modelo ha de representar el sistema real para el cual se desea tomar la mejor decisión [2].

El sistema de transporte urbano del cantón Ambato no cuenta con herramientas tecnológicas en sus unidades para determinar la calidad de servicio que brindan, lo cual se considera necesario y oportuno para conocer el sentir de la población y de los propios prestadores del servicio, y así todas las cooperativas de transporte conozcan lo que piensan los ciudadanos con relación al servicio recibido. En la presente investigación se pretende construir un modelo matemático para determinar la calidad del servicio que brindan las cooperativas de transporte en la ciudad de Ambato.

Ambato se ubica entre los diez cantones más poblados del Ecuador. Su población es atendida por un sistema de transporte público urbano que está conformado por cinco cooperativas, dando cobertura a un total de 22 rutas que permiten el traslado de la ciudadanía en las nueve parroquias urbanas de la ciudad y diez parroquias rurales. En la actualidad el sistema de transporte público urbano en la ciudad, al igual que en la gran mayoría de las ciudades del país, presenta novedades ya sea en la prestación del servicio, atención a los usuarios, estado actual de las unidades, forma de conducción, etc.

### **1.3. Objetivos**

#### **1.3.1. Objetivo General**

Construir un modelo matemático para determinar la calidad de servicio en el transporte público urbano en la ciudad de Ambato.

#### **1.3.2. Objetivos Específicos**

- Aplicar una encuesta para determinar el criterio de los usuarios sobre la calidad de servicio del transporte público urbano en la ciudad de Ambato.
- Determinar el peso de las variables a aplicar en la construcción del modelo matemático.
- Desarrollar el modelo matemático para determinar la calidad de servicio en el transporte público urbano en la ciudad de Ambato.

## CAPÍTULO II

### ANTECEDENTES INVESTIGATIVOS

#### 2.1. Estado del Arte

En este apartado se tomaron en cuenta ciertas investigaciones, las cuales sirvieron como base bibliográfica para el desarrollo del presente trabajo de investigación:

Gamarra & Delgado [1] en su obra titulada “Calidad del servicio de transporte público urbano en la ciudad del Cusco 2014” aplican un modelo de regresión logística usando cuatro variables independientes: el estado actual de las unidades, la forma de manejo del conductor, el tiempo de viaje y el trato del conductor y cobrador, con lo cual llegan a conocer y evaluar la percepción de la calidad del servicio de transporte público. Los investigadores concluyeron que con un modelo logit ajustado correctamente se puede establecer la calidad del servicio de transporte público urbano de la ciudad del Cusco.

De igual manera Pizarro [2] en su trabajo titulado “Factores explicativos de la demanda de Calidad en el servicio de las empresas de transporte público de pasajeros en el Distrito de Laredo 2015” emplea el método de regresión logística binaria y llega a determinar la percepción del usuario sobre el estado físico de los vehículos de las empresas de transporte público del distrito de Laredo. El investigador estableció que la percepción del usuario sobre el estado físico de los vehículos está determinada por tres criterios: condiciones de aseo de la unidad, comodidad y seguridad en los asientos y señalización de la unidad.

Moran [3] en su documento titulado “Análisis de la calidad del servicio de transporte urbano de buses en la ciudad de Esmeraldas desde la óptica del usuario” emplea el método científico, analítico y estadístico, en especial los modelo de regresión son herramientas estadísticas que ayudan a identificar las fortalezas y oportunidades del servicio de transporte urbano de buses en la ciudad de Esmeraldas a partir de la perspectiva del usuario, teniendo como variables: el trato recibido de parte de choferes y cobradores, la comodidad y la limpieza de los buses. El investigador estableció que la calidad del servicio de transporte urbano en Esmeraldas presenta múltiples deficiencias debido a la falta de evaluaciones al servicio en casi 70 años de utilización,

siendo el trato recibido por parte de choferes y cobradores la mayor queja que presentan los usuarios.

Paiva & Yataco [4] en la obra titulada “Relación entre la calidad del servicio logístico y satisfacción de los clientes de un operador logístico de Lima 2019” utilizan un modelo de regresión logit con un enfoque cuantitativo de tipo correlacional y llegaron a determinar que la relación entre la calidad del servicio logístico de los usuarios y la satisfacción de los clientes es un factor fundamental en el servicio público. Los investigadores concluyeron que al mejorar la puntualidad aumentará la credibilidad de los clientes en el servicio logístico y al mismo tiempo constataron que la relación entre la puntualidad y la satisfacción de los clientes es adecuada.

En la publicación de Berrones, Cano, Sánchez & Martínez [5] titulada “Entregas tardías o incorrectas en el autotransporte de carga y su relación con las condiciones laborales de los choferes: Un modelo de regresión logística” aplican una regresión logística a una lista de datos relacionados con el cumplimiento de embarques y logran identificar en qué medida las condiciones laborales de los conductores influyen en el incumplimiento de las entregas a tiempo, y la repercusión en la competitividad de las cadenas de suministro. Los investigadores establecieron que a pesar de que los diferentes grupos de conductores comparten riesgos y exigencias derivadas de su proceso de trabajo, sus condiciones laborales varían de acuerdo a diferentes factores determinados por aspectos tecnológicos, sociales, culturales, económicos y de regulación, tanto en materia laboral como de tránsito.

Peña, Castro & Talaverano [6] en el trabajo titulado “Factores asociados a somnolencia diurna en conductores de transporte público de Lima Metropolitana” utilizan la regresión logística para demostrar que los conductores de transporte público con seis o menos horas de sueño al día presentaron mayor probabilidad de somnolencia diurna, para las otras variables incluidas como la edad mayor a 50 años, grado de instrucción secundaria, grado de instrucción superior, índice de masa corporal y el sistema de rotación de trabajo no pudieron precisar con exactitud si existe mayor o menor probabilidad de somnolencia. Los investigadores concluyeron que aproximadamente una quinta parte de los conductores evaluados presentan somnolencia diurna.

Aguilar, Heald & Guerrero [7] en el documento titulado “Análisis de competitividad y movilidad en destinos turísticos del Estado de Guanajuato a través de un modelo de

regresión logística” aplican un modelo de regresión logística para caracterizar la probabilidad de movilidad entre destinos turísticos tradicionales, consolidados y emergentes en el contexto del estado de Guanajuato, México. Los investigadores establecieron que el modelo de regresión logística clasifica adecuadamente a los datos de la base disponible y determinaron los factores de mayor impacto en la movilidad, buscando de esta forma poder tomar decisiones sobre la competitividad turística del estado.

Mayorga, Ruiz & Aldaz [8] en el trabajo titulado “Percepciones acerca de la contaminación del aire generada por el transporte urbano en Ambato, Ecuador” aplican un modelo de regresión logística sobre la variable dependiente y logran determinar la percepción de los usuarios del transporte urbano referente a la contaminación del aire a partir de una encuesta aplicada a 382 usuarios en cinco sectores estratégicos. Los investigadores concluyeron que el nivel de emisiones ocasionados por el transporte público urbano es alto debido a la utilización del combustible diésel en las unidades de transporte; también mencionan que existe escaso control del cumplimiento de las regulaciones por parte de los contaminadores; además, se comprobó que el aire contaminado genera enfermedades que afectan a niños, mujeres embarazadas y ancianos generando inevitablemente un egreso financiero.

Camarena [9] en su obra titulada “Fidelización de los clientes y la Calidad de Servicio de la Empresa de Transportes América Express de la Ciudad de Trujillo 2018” aplica una encuesta como técnica y un cuestionario para determinar la variable dependiente; para el análisis estadístico emplea un modelo de regresión logística y logra determinar la relación entre la fidelidad de los clientes y la calidad del servicio de la empresa del transporte. El investigador estableció que el nivel de veracidad de los clientes es de un 5% bueno, 33% malo y por consiguiente 62% regular, por lo que dicha empresa debe tener un mejoramiento en las dimensiones (atraer, vender, satisfacer), para pretender establecer un vínculo a largo plazo con sus clientes.

En la publicación de Luque [10] titulada “Proceso de capacitación y calidad de servicio en la empresa de transporte público urbano José María Arguedas Andahuaylas 2015” emplea la regresión logit con un enfoque cuantitativo para examinar los datos de manera científica, de forma numérica y mediante herramientas estadísticas, con lo cual logró determinar la relación que existe entre el proceso de capacitación y la calidad de

servicio en la empresa de transporte público urbano. El investigador estableció que la empresa requiere de una capacitación a su personal para mejorar la calidad del servicio del transporte público urbano.

Mena & Ynchicsana [11] en el documento titulado “Cumplimiento de la normatividad del transporte público y calidad del servicio en la empresa Yllary Qosqo periodo octubre - noviembre Cusco, 2017” utilizan la entrevista y la encuesta para la recolección de la información y para realizar la inferencia emplean los modelos de regresión, con lo cual llegaron a demostrar que la normatividad del transporte público influye en la calidad de servicio, tanto en usuarios como de trabajadores. Los investigadores establecieron que la influencia del cumplimiento de la normatividad del transporte público en la calidad de prestación del servicio, se explica correctamente mediante la regresión logística, ya que según el índice de buen trato al usuario, el 61.25 % consideraron al trato como regular. Esto significa que la empresa debe mejorar en este aspecto. Por otra parte, según el índice de seguridad, se señala que el 59.38% lo considera como regular y la seguridad que brinda la empresa indica que el 46.88% lo considera como regular, al recoger pasajeros en paraderos autorizados.

Barón, Allende & Arena [12] en el trabajo titulado “Modelo de accesibilidad a sistemas de transporte público según la experiencia de usuario en el contexto urbano” emplearon los modelos de regresión logística con 6 variables predictivas para llegar a evaluar la accesibilidad a sistemas de transporte público. Los investigadores llegaron a concluir que el transporte público es una actividad que abarca a todas las actividades urbanas, mejorando las condiciones de acceso existentes y así proponen una infraestructura apropiada para construcciones futuras.

Rugel [13] en la obra titulada “Análisis de la calidad de servicio percibida por el usuario de transporte público de la ciudad de Guayaquil” emplea un modelo logit con un enfoque cuantitativo y cualitativo, con lo cual llegó a determinar los factores que inciden en la percepción sobre el servicio recibido por los usuarios del transporte público urbano masivo de Guayaquil “Metro vía”. El investigador concluyó que el transporte público cubre una de las necesidades más importantes en la vida de las medianas y grandes urbes, el mismo que permite la integración social de la ciudad y las zonas periféricas. El servicio del transporte público será mejor si la comodidad, el

tiempo de espera de los usuarios, la calidad en la atención, la seguridad y la frecuencia de ruta son adecuadas.

Beneitez [14] en su documento titulado “Mejora en la recolección de datos y la modelización de la calidad percibida en sistemas de transporte público” utilizó un modelo matemático de regresión logit para explicar la satisfacción de los clientes que usan el servicio de transporte público, considerando distintos aspectos como el servicio y atención a los usuarios en base a la toma de datos. El investigador concluyó que con un modelo logit correctamente ajustado se logra analizar la calidad del servicio del transporte público, a partir de los datos obtenidos mediante las encuestas, permitiendo estimar modelos que definen aquellos aspectos del servicio que más influyen en la percepción de la calidad de los usuarios, estableciendo prioridades de inversión para los operadores o autoridades del servicio.

Camargo [15] en el trabajo titulado “Caracterización y determinación de las variables que inciden en la accidentalidad en usuarios de moto en Bogotá, por medio de Regresión Logística Múltiple” emplea los modelos de regresión logística multivariada y el método descriptivo, correlacional y el explicativo, con lo cual llegó a determinar las variables que están relacionadas con la accidentalidad de los motociclistas y algunas características de los usuarios de moto en la ciudad de Bogotá. El investigador estableció que con el modelo de regresión logística múltiple se logró identificar las variables que se encuentran relacionadas con la presencia de motocicletas, las mismas que son: haber recibido una multa de tránsito, manejar bajo la influencia de alcohol y el uso diario de la moto para sus actividades cotidianas.

La publicación de Llamuca & Aguilar [16] titulada “Evaluación de la calidad del servicio de transporte urbano en bus de la ciudad de Ambato” utilizó la regresión logística con un enfoque cualitativo y deductivo, con un alcance descriptivo y correlacional, con lo cual llegó a evaluar y explicar la calidad del servicio de transporte público urbano en la ciudad de Ambato. Los investigadores concluyeron que el nivel de satisfacción de los usuarios del servicio de transporte urbano de Ambato está relacionado con las variables: el servicio ofertado, la accesibilidad, la información relativa a la ruta, el tiempo de cumplimiento de los horarios, la atención al cliente, el confort, la seguridad e impacto ambiental. Con inferencia estadística determinaron que

las personas que utilizan el transporte público de Ambato están insatisfechas en una tasa del 60%.

Apaza [17] en su obra titulada “Determinación de modelos matemáticos de caracterización de flujo vehicular para el centro histórico de la ciudad de Juliaca” empleó un modelo de regresión lineal y logística con un enfoque cuantitativo y llegó a diseñar un modelo matemático con el que explicó las condiciones del flujo vehicular del centro histórico de la ciudad de Juliaca. El investigador además estableció que la investigación desarrolla una serie de métodos que tienen la finalidad de analizar, diagnosticar y proponer alternativas de solución al problema del tránsito urbano en la ciudad de Juliaca, concluyendo que el porcentaje de vehículos no motorizados sumados a la estructura monocéntrica de la ciudad tienen una incidencia directa en los factores de caracterización del tránsito vehicular, lo que implica un nivel de movilidad bajo, reducción en la capacidad del límite de usuarios en las unidades de transporte y una calidad del tránsito vehicular bastante deficiente.

Silva & Torres [18] en su documento titulado “La calidad del servicio de transporte urbano en la ciudad de Cuenca” utilizó la encuesta para la recolección de la información empleando los modelos de regresión, con lo cual llegó a determinar la percepción de la calidad del servicio de transporte público urbano de la ciudad de Cuenca, siendo las variables empleadas para el desarrollo del modelo: la atención al cliente, tiempo de viaje, trato al usuario, forma de manejo, capacitación al conductor y servicio ofertado. Los investigadores concluyen que al utilizar los modelos de regresión se llega a determinar la calidad del servicio de transporte urbano incluyendo capacitaciones al conductor sobre el trato, mejoras de las paradas con ubicación estratégica y el mantenimiento de las unidades para evitar contaminación.

Vivas [19] en el trabajo titulado “Calidad del servicio en el transporte público de la Cooperativa Frontera Norte de la ciudad de Tulcán” utilizó una metodología basada en los modelos logit con un enfoque cualitativo y descriptivo, con lo cual llegó a determinar el nivel de calidad del servicio. El investigador estableció que con un modelo logit se logra determinar las expectativas y las percepciones de los usuarios del cantón Tulcán sobre la calidad del servicio del transporte público.

En la publicación de Navarro [20] titulada “La gestión de calidad del servicio y atención al cliente en la micro y pequeñas empresas del rubro transporte público buses

ruta Sullana - Piura del distrito de Sullana en el año 2018” se empleó la regresión logit con un enfoque descriptivo y se llegó a determinar varias características de la gestión de la calidad del servicio y de la atención al cliente en el transporte público. El investigador concluye que el modelo explica adecuadamente la gestión y la atención al cliente en el servicio del transporte público.

## 2.2. Marco teórico

### 2.2.1. Regresión logística.

La regresión logística permite estimar la probabilidad de una variable cualitativa en función de una o más variables cuantitativas o cualitativas. Esta regresión se clasifica en regresión logística binomial y multinomial. La primera de ellas es cuando una cualidad puede tomar dos modalidades y por lo general es la más utilizada, mientras que la multinomial se utiliza cuando una cualidad podría tomar más de dos modalidades diferentes, exhaustivas y mutuamente excluyentes; la diferencia de ambas está en cómo se realiza el planteamiento del problema que ayuda a explicar, predecir y clasificar una variable de estudio [21].

### 2.2.2. Regresión logística binomial.

La regresión logística binomial es un conjunto de herramientas estadísticas que permite comprobar la hipótesis sobre una variable dependiente y es la variante que corresponde al caso en que se valora la contribución de diferentes factores en la ocurrencia de un evento simple. Comúnmente esta regresión se utiliza cuando la variable dependiente es binomial, es decir contiene dos categorías también llamadas dicotómicas. Un modelo de regresión logística no predice el valor de la variable dependiente Y si no que directamente estima la probabilidad que ocurra Y, dadas las variables predictoras  $X_1, X_2, \dots, X_n$  [21].

$P(Y = 1   X_1, X_2, \dots, X_n) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$	(1)
--	-----

donde:

P es la probabilidad y

Y es la variable dependiente o respuesta.

$X_i$  corresponde a las variables predictoras.

$\beta_i$  representan los coeficientes estimados del modelo.

### 2.2.3. Estimación del modelo

Estimar un modelo no es otra cosa que aproximar los valores de sus parámetros, realizar las pruebas de significancia sobre las variables predictoras y determinar la bondad de ajuste del modelo mediante herramientas estadísticas.

**Prueba de significancia.** Son utilizadas para comparar variables entre varias muestras. Se emplea el nivel de significación alfa ( $\alpha$ ), que representa la probabilidad de rechazar la hipótesis nula cuando es verdadera. Un nivel de significancia de 0.05 indica un riesgo de 5% de concluir que existe una diferencia cuando en realidad no hay ninguna diferencia. [22].

#### **Bondad de ajuste en los modelos de regresión logística.**

Una de las pruebas estadísticas utilizadas para medir la bondad del ajuste de un modelo de regresión logística es el estadístico ji-cuadrado.

#### **Estadístico ji-cuadrado.**

Este estadístico compara las frecuencias observadas y esperadas de una muestra en un escenario binomial y su ecuación es:

$$X^2 = \sum_{j=1}^J \frac{(y_j - n_j p_j)^2}{n_j p_j (1 - p_j)} \quad (2)$$

donde

$X^2$  es el estadístico ji- cuadrado

$y_i$  = variables predichas

La hipótesis nula será rechazada para un nivel de significancia  $\alpha = 5\%$  cuando el valor crítico encontrado sea superior al valor crítico prefijado de la prueba según los grados de libertad establecidos [22].

Otras pruebas que ayudan a determinar la bondad de ajuste de un modelo de regresión logística son:

### Prueba de Hosmer y Lemeshow

Esta medida sirve para evaluar la bondad del ajuste de un modelo de regresión logística, es decir, si el p valor del modelo es mayor al nivel de significancia  $\alpha = 5\%$  o el estadístico de prueba pertenece al intervalo de confianza del 95%, el modelo está bien ajustado [23].

Los valores de la muestra se agrupan en g clases y los valores observados y esperados se calculan mediante la función de máxima verosimilitud, el estadístico de Hosmer y Lemeshow se define mediante la siguiente ecuación:

$$HL = \sum_{i=1}^g \frac{(Obs1_i - Esp1_i)^2}{Esp1} + \sum_{i=1}^g \frac{(Obs0_i - Esp0_i)^2}{Esp0} \quad (3)$$

donde:

HL es el estadístico de Hosmer y Lemeshow.

g clases de los grupos.

Obsk son los valores observados de  $Y=k$ ,  $k=0,1$ .

Espk son los valores esperados de  $Y=k$ ,  $k=0,1$ .

### Prueba de Cox y Snell

La prueba de Cox y Snell es un coeficiente de determinación que se usa para estimar la proporción de varianza de la variable dependiente explicada por las variables predictoras. Esta prueba se basa en la comparación del log de la verosimilitud (LL) para un modelo de línea base, el rango de oscilación esta entre 0 y 1 [24].

Si el p valor del modelo es superior al nivel de significancia establecido de  $\alpha = 5\%$  el modelo está bien ajustado. En conclusión, la medida de Cox-Snell está definida mediante la siguiente expresión [25].

$$Rcs = \exp(\beta \cdot X_i) \cdot h_o(t_i) \quad (4)$$

donde:

Rcs es el estadístico de Cox y Senll

$\beta$  es el parámetro.

$X_i$  es la variable predictora

$h_0(t_i)$  es la tasa de riesgo de línea base

### **Prueba de Nagelkerke**

Esta medida es una versión corregida de la medida de Cox y Snell ya que tiene un valor máximo inferior a 1. La prueba de Nagelkerke mejora la escala del estadístico para cubrir el intervalo total de 0 a 1 [24].

Si el valor de  $R^2$  de Nagelkerke tiene un valor de 0 se concluye que no existe ningún efecto de variación. Por el contrario, si la medida de  $R^2$  de Nagelkerke tiene un valor cercano a 1, se determina que existe una variación entre las variables [26].

#### **2.2.4. Odds Ratio**

Para interpretar los coeficientes estimados del modelo es necesario transformar las variables predictoras; para esto se aplica el logaritmo natural a la ecuación principal de los modelos de regresión logística y así encontrar la razón de posibilidades, llamada también como odds ratio, es decir para poder interpretar adecuadamente las variables predictoras que forman el modelo es necesario que su resultado este en forma exponencial. Esta transformación es presentada en las siguientes ecuaciones [27].

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (5)$$

En esta ecuación no es posible interpretar los coeficientes debido a que se encuentran en forma logarítmica.

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k} \quad (6)$$

Mediante la ecuación (6) ya es posible interpretar los variables predictoras debido a que los coeficientes de las variables se encuentran presentadas de forma exponencial y su valor puede estar entre  $-\infty$  al  $+\infty$

### **Punto de corte y la Curva Roc.**

El punto de corte o frontera de decisión establece la separación entre la probabilidad de ocurrencia del éxito y el fracaso. Para la selección del punto de corte óptimo se sugiere verificar el comportamiento de algunos parámetros como:

Número de falsos positivos (FP), que son los errores tipo I o el número de predicciones incorrectas de clase positiva.

Número de falsos negativos (FN), que son los errores tipo II o el número de predicciones incorrectas de clase negativa.

Sensibilidad (S), es la proporción de la cantidad de falsos positivos que fueron identificados por el algoritmo.

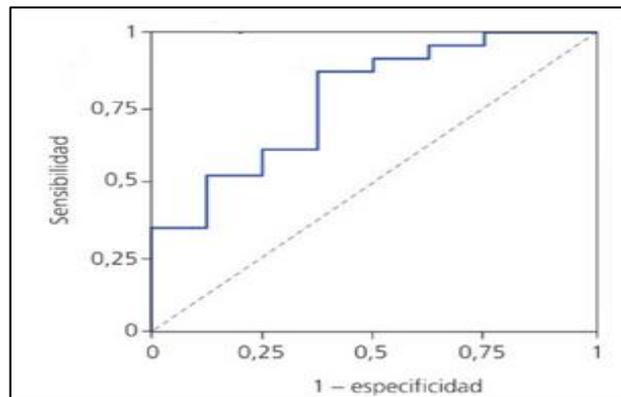
Especificidad (E), se trata de los cantidad de casos positivos verdaderos que el algoritmo ha clasificado correctamente.

Índice de Youden, es una prueba estadística que indica el rendimiento de una prueba de diagnóstico en forma dicotómica. [28].

Los puntos de corte proporcionan un equilibrio en la relación de especificidad y sensibilidad, lo cual ayuda a tomar una decisión acerca de cuál es el punto de corte óptimo y generalmente su valor es el codo de la curva. Los valores de sensibilidad y especificidad son utilizados como los ejes del sistema de coordenadas para la obtención de la curva de Roc. [29]

Una curva Roc es la representación gráfica de la probabilidad que se obtiene a partir de los resultados positivos verdaderos llamados especificidad y los resultados falsos positivos denominados sensibilidad.

La gráfica de la curva roc en función de la especificidad y la sensibilidad es:



**Figura 1.** Representación gráfica de la curva Roc.

**Fuente:** Principios de Bioquímica Clínica y Patología Molecular.

### 2.2.5. Matriz de confusión

También llamada tabla de contingencia, se encuentra formada por una partición de los elementos en función de su clase y su predicción. Es una herramienta que indica el trabajo de un algoritmo de aprendizaje supervisado, en la que se puede observar que cada fila indica los aciertos o fallas que tiene un modelo de regresión logística binaria, mientras que cada columna indica el número de predicciones de cada clase, de esa manera los valores de los verdaderos positivos representa una clasificación adecuada, los valores de los verdaderos negativos indican una clasificación inadecuada, mientras los valores de los falsos positivos y falsos negativos simbolizan los valores que el modelo no logra clasificar correctamente [30].



**Figura 2.** Matriz de confusión o tabla de contingencia

**Fuente:** Inteligencia artificial avanzada.

## CAPÍTULO III

### MARCO METODOLÓGICO

#### 3.1. Ubicación

##### Historia de Ambato

Ambato es la capital de la provincia del Tungurahua y se encuentra situada a 2600 metros de altitud. Esta provincia está conformada por 9 cantones: Ambato, Píllaro, Pelileo, Patate, Tisaleo, Quero, Mocha, Cevallos y Baños. La ciudad de Ambato es conocida como la tierra de los tres Juanes por ser la cuna de Juan León Mera, Juan Montalvo y Juan Benigno Vela, pero algunas personas también la denominan como el Jardín del Ecuador por ser una tierra de flores y frutas [31].

##### Transporte de Ambato

Ambato cuenta con una red amplia de carreteras que se conecta con las poblaciones cercanas de su provincia en especial con Baños. Además la provincia está conformada por una amplia red de cooperativas de transporte distribuidas en 5 urbanas y 3 rurales, que abarcan todos los rincones de su territorio y con los que se conecta fácilmente con el centro y las zonas comerciales [31].

Las cooperativas de transporte público urbano de la ciudad de Ambato realizan el recorrido por varias rutas, presentadas en la tabla 1.

**Tabla 1.** Rutas de las cooperativas de transporte urbano de Ambato

<b>Cooperativas de transporte</b>	<b>Líneas de ruta</b>
Jerpazol	Circula por la línea 21 y 22.
Libertadores	Operará por la línea 1, 2, 3, 4 y 5.
Tungurahua	Circula por la línea 6, 7, 8, 9, 10, 11, 12 y 13
Unión Ambateña	Circula por la línea 14, 15, 16, 17 y 18.
Vía Flores	Realiza su recorrido por la línea 19 y 20.

**Fuente:** Gad Municipal de la ciudad de Ambato.

## **Paradas de autobuses de Ambato**

Las paradas más populares y frecuentes para los usuarios del transporte público son:

El ex redondel de Izamba. Es una de las paradas que con más frecuencia utilizan los ciudadanos de la ciudad de Ambato, porque es la parada que da inicio a las rutas de las diferentes líneas de transporte público.

Mercado Modelo. Ubicada en el centro de la ciudad junto al mercado homónimo.

Parque 12 de noviembre. Es una de las principales paradas, donde se observan a diario aglomeraciones de personas por ser el epicentro de la ciudad.

Redondel de la policía. Está ubicada en la vía a Huachi Chico y al ser el redondel que une distintas vías de los sectores de la ciudad, se ha ido convirtiendo en una de las más populares.

Parada del Mall de los Andes. Esta parada es la más recurrida por las personas por encontrarse junto al centro comercial de la ciudad de Ambato,

Parada del mercado Mayorista. En esta parada no solo existe la participación de los habitantes de la ciudad, sino también de personas de otras provincias, por encontrarse junto a donde se desarrollan las actividades más comerciales de la ciudad en el mercado mayorista.

### **3.2. Equipos y materiales**

Para desarrollar el modelo matemático se empleó un modelo de regresión logística, con lo cual se logra determinar la calidad del servicio del transporte público de la ciudad de Ambato. En el modelo de regresión logística para construir y verificar el comportamiento del modelo se deben seguir los siguientes pasos:

1. Medir la correlación entre las variables. Consiste en determinar la relación que existe entre la variable dependiente y las variables independientes, para ello se usa un análisis gráfico de correlación.
2. Determinar la multicolinealidad y la significancia de las variables predictoras. La significancia consiste en verificar que las variables que forman el modelo sean

significativas, mientras que en la multicolinealidad se demostrará que las variables predictoras no estén correlacionadas.

3. Evaluar el ajuste del modelo. Esta etapa consiste en verificar si el modelo se ajusta adecuadamente a las variables y sirve para determinar la variable dependiente, para ello se emplea diferentes pruebas estadísticas (Ji-cuadrada, Hosmer y Lemeshow, Cox y Snell y la de Nagelkerke).
4. Interpretar los odds ratio (OR). Consiste en aplicar el logaritmo natural a la ecuación de la probabilidad y así interpretar la razón par (odds ratio) que se expresará en forma exponencial.
5. Determinar el punto de corte y la curva Roc. Consiste en encontrar la frontera de decisión y su representación gráfica mediante la curva Roc.
6. Interpretar la matriz de confusión. Mediante la matriz de confusión se muestran los valores que el modelo predice de forma correcta e incorrecta y el valor de las proporciones que no logra clasificar.
7. Predicción de la probabilidad con el modelo. En esta etapa se determina si la variable dependiente es éxito o fracaso.

Los equipos y materiales utilizados fueron:

Los programas y herramientas estadísticas utilizadas son:

Google drive se utilizó para subir la encuesta y para receptar las respuestas. Se creó un documento en Excel para poder tabular con mayor facilidad.

Microsoft Excel 2016 se utilizó para la tabulación de las encuestas y para la transformación de las variables dicotómicas tipo carácter a variables numéricas.

Software RStudio versión 4.0.1 se utilizó para realizar el código informático del modelo de regresión logística, también permitió efectuar el análisis estadístico, logrando de esa manera predecir la probabilidad de la calidad de transporte en la ciudad de Ambato.

Usando Microsoft Word 2016 se realizó la encuesta y se procedió a la escritura del proyecto de titulación.

Los equipos utilizados fueron:

Un computador portátil Hp con un procesador Intel (R) Core (TM) i5 de octava generación, con una RAM de 8GB.

Un celular Huawei p30 life con un procesador Hisilicon kirin 710, con una memoria RAM de 4GB.

### **3.3. Tipo de investigación**

La presente investigación se basó en un estudio de análisis tipo cuantitativo, descriptivo y correlacional.

#### **Investigación cuantitativa**

En la investigación cuantitativa se realiza la recolección de los datos mediante mediciones numéricas y análisis estadísticos con el fin de determinar el comportamiento de las variables [32].

Se aplicó inicialmente la investigación cuantitativa porque se elaboró un cuestionario de 10 preguntas cerradas de opción múltiple, de orden cualitativo, con la finalidad de obtener información necesaria sobre la calidad del transporte público en la ciudad de Ambato, y de esa manera a través del modelo construido, predecir el comportamiento sobre la calidad del servicio de transporte público.

#### **Investigación descriptiva**

La investigación descriptiva busca explicar las propiedades, características y los perfiles de personas, comunidades, procesos o cualquier otro fenómeno que se encuentre sometido a un análisis. También se podría decir que esta investigación pretende medir o recolectar información de forma independiente o conjunta de una variable que se busca estudiar [32].

La investigación es de tipo descriptivo porque ayudó a especificar las características y propiedades sobre la calidad del servicio del transporte urbano en la ciudad de Ambato, mediante la aplicación del modelo construido.

## **Investigación correlacional**

La investigación correlacional tiene como objetivo conocer la relación que existe entre dos o más variables de una muestra en particular y así evaluar el grado de asociación que existe entre ellas, para después cuantificarla, analizarla y establecer vinculaciones en dicha correlación. [32].

La investigación correlacional sirvió para comparar la relación que existe entre las variables de estudio como: Calidad del servicio, tiempo de espera, trato al usuario, estado actual de las unidades de transporte y forma de conducción del transportista. Una vez conocido el grado de correlación entre las variables se busca crear un modelo matemático que determine cómo es la calidad del servicio del transporte público urbano en la ciudad de Ambato.

### **3.4. Problema e hipótesis**

#### **Problema**

¿Cómo se determina la calidad del servicio del transporte público urbano de la ciudad de Ambato al utilizar diversas variables como: el tiempo de espera, el trato al usuario, el estado actual de las unidades y la forma de conducción del transportista?

#### **Hipótesis**

Las variables como el tiempo de espera, trato al usuario, estado actual de las unidades y forma de conducción del transportista son las adecuadas para generar un modelo de regresión logística que explica y describe adecuadamente a la variable calidad del servicio del transporte público urbano en la ciudad de Ambato.

### **3.5. Población o muestra**

La población de estudio estuvo conformada por 165185 habitantes de la ciudad de Ambato comprendida entre mujeres y hombres de 10 años en adelante, que pueden utilizar el transporte público de la ciudad en sus distintas paradas de autobuses.

Para establecer el tamaño de la muestra se aplicó la siguiente fórmula estadística con un nivel de significancia  $\alpha = 5\%$ . [3]

$$n = \frac{N}{E^2(N - 1) + 1}$$

Donde:

N = población

n = tamaño de la muestra

E = error admisible de la muestra (5%)

$$n = \frac{165185}{[(0.05^2)(165185)] + 1}$$

$$n = \frac{165185}{413.96}$$

$$n = 400$$

La muestra estuvo conformada de 400 personas de las cuales fueron: 175 mujeres adultas, 150 hombres adultos y 75 adolescentes.

### 3.6. Recolección de información

Para la recolección de la información se aplicó la técnica de la encuesta y como instrumento un cuestionario que contiene 10 ítems con preguntas de opción múltiples y cerradas. De los 10 ítems se utilizaron 5 preguntas para la construcción del modelo, presentadas en la tabla 2.

**Tabla 2.** Variables de estudio

<b>Variables independientes</b>	<b>Variable dependiente</b>
Tiempo de espera	Calidad del servicio
Trato al usuario	
Estado actual de las unidades	
Forma de conducción del transportista	

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

### 3.7. Procesamiento y análisis estadístico

Para procesar la información obtenida mediante la técnica de la encuesta con su instrumento el cuestionario se utilizaron los programas estadísticos:

Microsoft Excel 2016. Con este programa se logró sintetizar y analizar las 400 encuestas aplicadas a diferentes personas de la ciudad de Ambato. También sirvió para seleccionar las variables que contiene el modelo matemático.

Software RStudio versión 4.0.1. Con esta herramienta estadística se determinó la correlación que existe entre las variables de estudio. También sirvió para desarrollar el código informático que ayuda a determinar la calidad del servicio del transporte público urbano de la ciudad de Ambato.

El análisis estadístico se realizó mediante una regresión logística.

Lo primero que se realizó fue determinar si las variables están correlacionadas; al comprobar que están moderadamente relacionadas entre sí, se procedió a la construcción de un modelo que determine la calidad del servicio del transporte público urbano de la ciudad de Ambato.

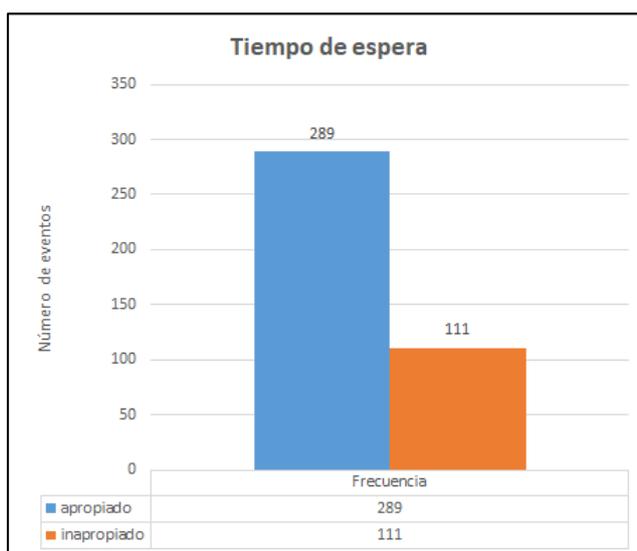
Para verificar si los parámetros del modelo son significativos se utilizó la distribución normal con un nivel significancia de  $\alpha = 5\%$ . También se emplearon las pruebas estadísticas: Ji-cuadrada, Hosmer y Lemeshow, Cox y Snell y la de Nagelkerke que verifican si el modelo de regresión logística es el adecuado.

Una vez comprobado que el modelo se ajusta adecuadamente a los datos se procedió a interpretar los valores de la razón de posibilidades (odds ratios) de cada variable predictora; para ello se aplicó el logaritmo natural a la ecuación de la probabilidad y así se determinó la ecuación de la probabilidad en forma exponencial, con lo cual se pudo interpretar el exponente de cada variable predictora. Por último, se encontró el punto de corte con el cual se logró determinar que el modelo de regresión logística clasifica adecuadamente 308 datos de un total de 400 con un valor de 0.8170 del área bajo la curva Roc.

### **3.8. Análisis descriptivo de datos**

A continuación se describen los resultados de la encuesta de la calidad de servicio de transporte urbano, en gráficos de proporciones y tablas de frecuencia de las preguntas que tienen relación con las variables de estudio independientes, a saber: tiempo de espera, trato al usuario, estado actual de las unidades y forma de conducción del transportista, así como la variable dependiente calidad del servicio.

### Variable Independiente: Tiempo de espera



**Figura 3.** Análisis Descriptivo de la variable Tiempo de espera.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 3 se presentan los resultados de la pregunta relacionada con la variable tiempo de espera, en la que se preguntó si siendo un usuario frecuente de una línea de servicio de transporte público urbano como considera que son los horarios de servicio. Lo consideran apropiado 289 personas, mientras que inapropiado un total de 111, dando como conclusión que en su mayoría los usuarios se encuentran satisfechos con el tiempo de espera del transporte público urbano.

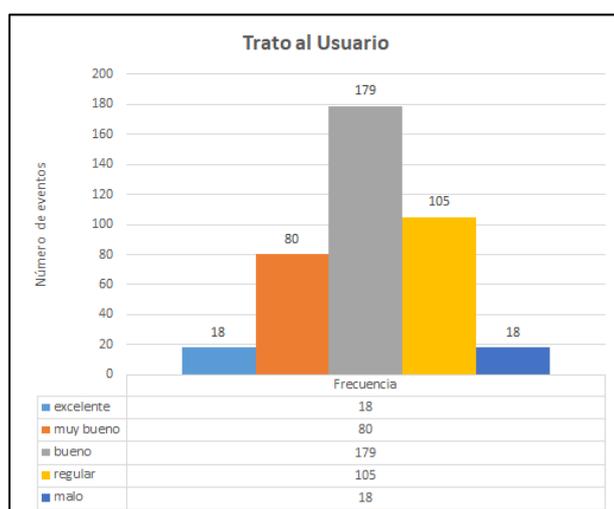
**Tabla 3.** Variable Tiempo de espera

Alternativas	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa acumulada
apropiado	289	289	0,72	0,72
inapropiado	111	400	0,28	1

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 3 se presentan las frecuencias de la variable tiempo de espera. Se concluye que 72 % consideran como apropiado el tiempo de espera, es decir, un porcentaje alto los usuarios se encuentran satisfechos con el tiempo de espera del transporte público urbano.

## Variable Independiente: Trato al usuario



**Figura 4.** Análisis Descriptivo de la variable Trato al usuario.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 4 se presentan los resultados de la pregunta relacionada con la variable trato al usuario en la que se le preguntó cómo considera el trato que ha recibido por parte de las personas a cargo de la unidad de transporte (conductor-ayudante), a lo cual respondieron que es excelente 18 personas, muy bueno 80, bueno 179, regular 105 y malo 18 personas, dando como resultado que en su mayoría dicen que el trato al usuarios es bueno es decir hay amabilidad en la atención.

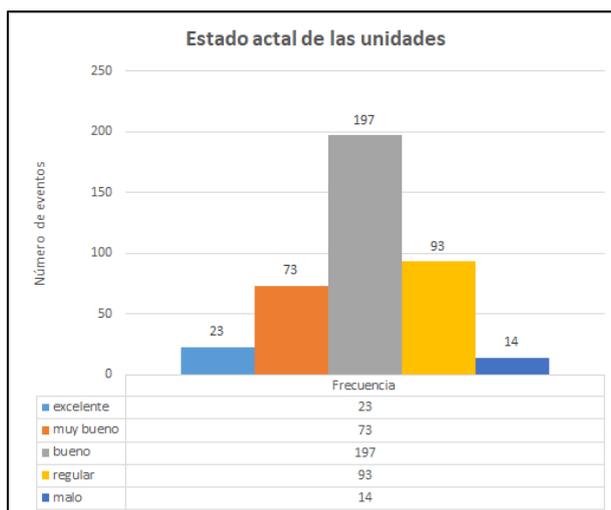
**Tabla 4.** Variable Trato al usuario

Alternativas	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa acumulada
excelente	18	18	0,05	0,05
muy bueno	80	98	0,20	0,25
bueno	179	277	0,45	0,69
regular	105	382	0,26	0,96
malo	18	400	0,05	1

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 4 se presenta la tabla de frecuencias de la variable trato al usuario, dejando como conclusión que en un 45 % lo considera como bueno, es decir, casi la mitad de la población encuestada considera que ha recibido un buen trato por parte del conductor o ayudante.

## Variable Independiente: Estado actual de las unidades



**Figura 5.** Análisis Descriptivo de la variable Estado actual de las unidades.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 5 se presentan los resultados de la pregunta relacionada con la variable estado actual de las unidades de servicio. En este caso se preguntó considerando las unidades de transporte que con mayor frecuencia utiliza, como considera la apariencia en asientos, piso y entorno, a lo cual respondieron que es excelente 23 personas, muy bueno 73, bueno 197, regular 93 y malo 14 personas, dejando como conclusión que en su mayoría los encuestados consideran como bueno el estado, es decir, la mitad de la población encuestada considera que el estado actual de las unidades de servicio está en buen estado.

**Tabla 5.** Variable Estado actual de las unidades

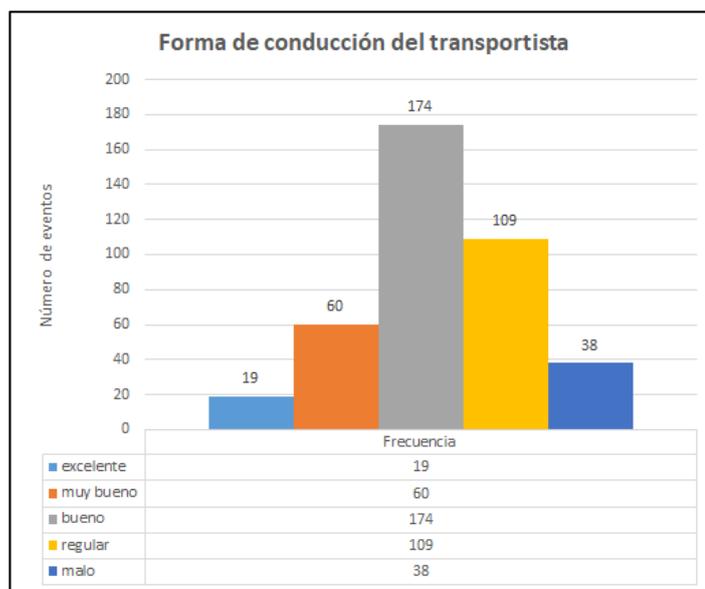
Alternativas	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa acumulada
excelente	23	23	0,06	0,06
muy bueno	73	96	0,18	0,24
bueno	197	293	0,49	0,73
regular	93	386	0,23	0,97
malo	14	400	0,04	1

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 5 se presenta la tabla de frecuencias de la variable estado actual de las unidades, dejando como conclusión que un 49 % lo considera como bueno, es decir,

la mayoría de los encuestados considera que el estado actual de las unidades cumple con las normativas de confort y seguridad.

**Variable Independiente: Forma de conducción del transportista**



**Figura 6.** Análisis Descriptivo de la variable Forma de conducción del transportista.  
**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 6 se presentan los resultados de la pregunta relacionada con la variable forma de conducción del transportista, en la que se preguntó al usuario en su experiencia general como considera la forma de conducir de los conductores del transporte público urbano. Respondieron excelente 19, muy bueno 60, bueno 174, regular 109 y malo 38, dejando como conclusión que casi la mitad de la población encuestada considera que la forma de conducción del transportista es aceptable.

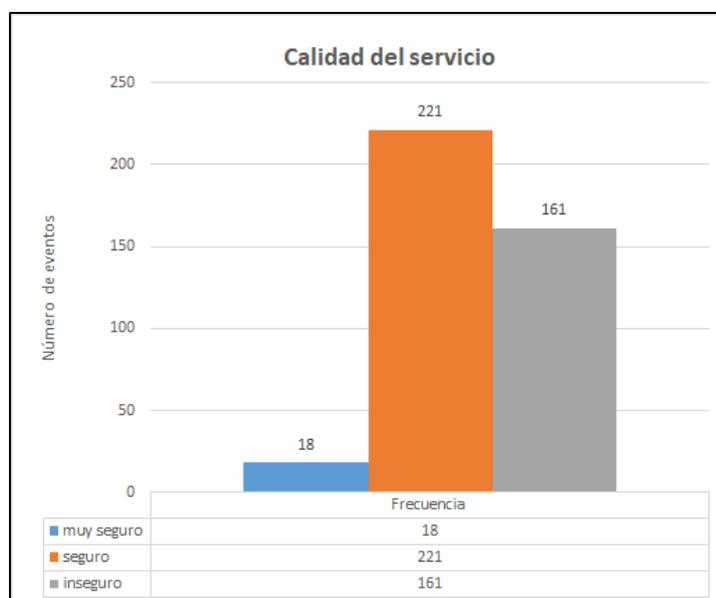
**Tabla 6.** Variable Forma de conducción del transportista

Alternativas	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa acumulada
excelente	19	19	0,05	0,05
muy bueno	60	79	0,15	0,20
bueno	174	253	0,44	0,63
regular	109	362	0,27	0,91
malo	38	400	0,10	1

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 6 se presenta la tabla de frecuencias de la variable forma de conducción del transportista. Se concluye que un 44 % la considera buena, es decir que la mayoría de población encuestada considera que los conductores manejan de manera correcta y respetan las señales de tránsito.

**Variable Dependiente: Calidad del servicio**



**Figura 7.** Análisis Descriptivo de la variable Calidad del servicio.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 7 se presentan los resultados de la pregunta relacionada con la variable calidad del servicio, en la que se le preguntó al usuario en su experiencia general como como se siente al utilizar el servicio de transporte público urbano. A ello respondieron que es muy seguro 18 personas, seguro 221 e inseguro 161 personas, dejando como conclusión que más de la mitad de la población encuestada considera al servicio de transporte público como seguro.

**Tabla 7.** Variable Calidad de servicio

Alternativas	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa acumulada
muy seguro	18	18	0,05	0,05
seguro	221	239	0,55	0,60
inseguro	161	400	0,40	1

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 7 se presentan las frecuencias de la variable dependiente calidad de servicio. Se concluye que 55 % lo considera como seguro, es decir, para la mayoría de los encuestados el servicio brindado por las unidades de transporte público urbano es de calidad.

### **3.9. Variables respuestas o resultados alcanzados**

Con la información de la encuesta aplicada, en el programa Microsoft Excel se construyó una base de datos digital, ya que es una herramienta que permite resumir, calcular y analizar gran cantidad de registros de una manera fácil e intuitiva. Dicha base de datos contiene 5 ítems preseleccionados, que son utilizados como variables para el desarrollo del modelo matemático. Como variable dependiente se tiene a la calidad del servicio y como variables independientes: El tiempo de espera, trato al usuario, estado actual de las unidades de transporte y la forma de conducción del transportista.

Para desarrollar el modelo se determinó la importancia de las variables mediante el gráfico de análisis de correlación, en donde la variable trato al usuario es la más importante con el 50%, luego está la forma de conducción del transportista con el 45%, le sigue el estado actual del vehículo con el 37% y la variable menos significativa es el tiempo de espera con el 33%. Estas variables predictivas se encuentran medianamente relacionadas con la calidad del servicio. Una vez determinada la importancia de las variables independientes y su relación con la variable dependiente se desarrolló un modelo de regresión logística que determina la calidad del servicio del transporte público urbano de la ciudad de Ambato

Para verificar si el modelo desarrollado se ajusta adecuadamente a los datos se utilizaron cuatro pruebas estadísticas como: Ji-cuadrada, Hosmer y Lemeshow, Cox y Snell y la de Nagelkerke. Una vez comprobado que el modelo cumple con las pruebas de ajuste y de adecuación, se procedió a determinar el punto de corte el cual tiene un valor del 60,42%, con lo cual se puede decir que si un dato es superior al valor del punto de corte entonces la calidad del servicio del transporte es seguro, pero si es inferior a dicho valor del punto de corte la calidad del servicio es inseguro.

En la siguiente tabla se presentan los resultados sobre la calidad del servicio del transporte público:

**Tabla 8.** Resultados del Modelo

	Verdadero positivo	Falso positivo
Verdadero positivo	205	34
Falso positivo	58	103
Clasificación del Modelo		
Clasificaciones correctas	308	77%
Clasificaciones incorrectas	92	23%

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 8 se observa que el modelo clasifica el 77% de los datos de forma correcta y el 23% de forma incorrecta. También se observa el resumen de la matriz de confusión en donde se determinó que 205 datos pertenecen a una calidad de servicio del transporte segura, 103 datos arrojaron que la calidad del servicio del transporte es insegura y 92 datos el modelo no logra clasificarlos.

## CAPÍTULO IV

### RESULTADOS Y DISCUSIONES

En este apartado se presentan los resultados obtenidos del procesamiento de la encuesta aplicada a 400 personas que utilizan el transporte público urbano de la ciudad de Ambato en las paradas más frecuentes y populares, con la finalidad de dar respuesta a los objetivos.

#### 4.1. Análisis de las variables

Lo primero que se realizó fue la transformación de los valores de las variables de la encuesta las mismas que contenían preguntas de opción múltiple y fueron transformadas a variables dicotómicas. Y así obtener las variables adecuadas para la construcción del modelo.

**Tabla 9.** Transformación de los resultados a variables dicotómicas

<b>variables</b>	<b>Tiempo de espera</b>	<b>Trato al usuario</b>	<b>Estado actual de las unidades de transporte</b>	<b>Forma de conducción del transportista</b>	<b>Calidad del servicio</b>
Valor real	Apropiado	Adecuado	Bueno	Apropiada	Seguro
Valor Interno	1	1	1	1	1
Valor real	Inapropiado	Inadecuado	Malo	Inapropiada	Inseguro
Valor Interno	0	0	0	0	0

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 9 se observa las variables ya transformadas a dicotómicas. La variable dependiente o variable de estudio es la calidad del servicio, la misma que tiene por opciones seguro representado con el 1 e inseguro por el 0. Mientras que los valores de las variables independientes como el tiempo de espera, trato al usuario, estado actual de las unidades de transporte y forma de conducción del transportista fueron

transformadas a resultados apropiados, adecuado, bueno son representados por 1 y los resultados de inapropiado, inadecuado, malo se simbolizan mediante 0.

Posteriormente se realizó la tabulación de las encuestas, donde se presenta las variables para la construcción del modelo, esta tabulación es presentada en la siguiente tabla:

**Tabla 10.** Estructura de la base de datos

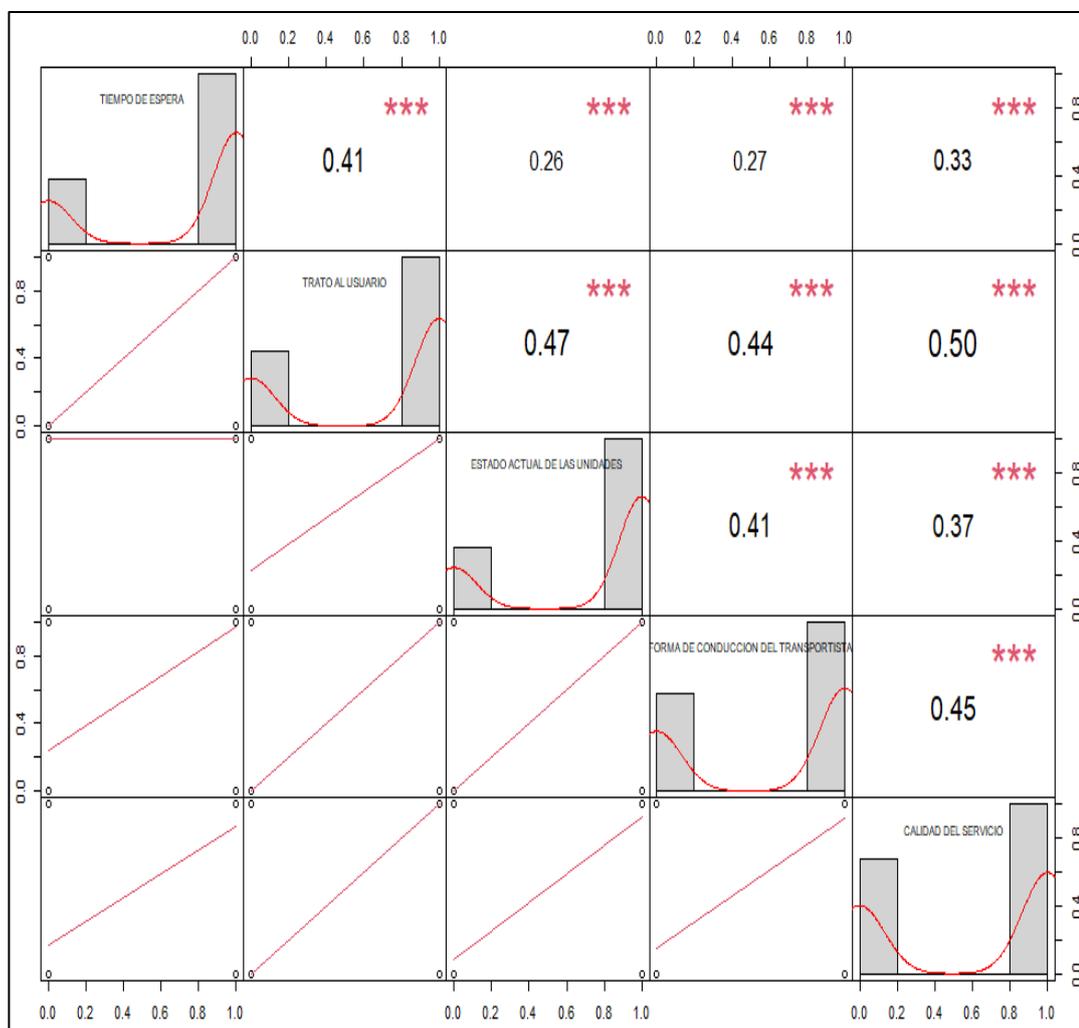
<b>Tiempo de espera</b>	<b>Trato al usuario</b>	<b>Estado actual de las unidades</b>	<b>Forma de conducción del transportista</b>	<b>Calidad del servicio</b>
núm. [1:400]	núm. [1:400]	núm. [1:400]	núm. [1:400]	núm. [1:400]
1	0	0	0	0
0	1	1	1	0
1	1	0	0	1
1	1	0	0	1
1	1	1	1	1
1	0	1	0	0
1	1	1	1	0
1	1	1	0	1
0	0	0	0	0
0	0	0	0	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 10 se presentan la estructura de la base de datos según la encuesta aplicada, donde se observa que todas las variables constan de 400 datos, las mismas que son de tipo numérico y están codificadas con los valores 1 y 0.

Antes de construir el modelo de regresión logística es necesario determinar la relación que existe entre las variables. El tipo de relaciones entre variables puede clasificarse

de la siguiente manera: según la dirección puede ser positiva o negativa y según la fuerza puede ser fuerte si es superior a 70%, moderada si esta entre 25% y 70% y débil cuando es inferior a 25%. La correlación entre las variables es presentada en la siguiente figura:



**Figura 8.** Correlación entre las variables de estudio

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 8 se presenta la correlación entre las variables. Se centra en la variable de estudio, que en este caso es la calidad del servicio del transporte público y se observa que esta variable se encuentra moderadamente relacionada con las otras variables. La calidad del servicio está relacionada en un 33% con el tiempo de espera, un 50% con el trato al usuario, un 37% con el estado actual de la unidad de transporte y un 45% con la forma de conducción del transportista. También se visualiza que existe una

relación positiva, por lo que al mejorar una variable predictora también se mejorará la variable respuesta.

#### 4.2. Construcción del modelo

Para el tratamiento y la construcción del modelo matemático se emplearon los modelos de regresión logística con el fin de determinar la calidad del servicio del transporte público urbano de la ciudad de Ambato a partir del tiempo de espera, el trato al usuario, el estado actual de las unidades de transporte y la forma de conducción del transportista. En la siguiente tabla se presenta el resumen del modelo de regresión logística desarrollado en el software libre RStudio versión 4.0.1 que utiliza el método de máxima verosimilitud:

**Tabla 11.** Resumen del modelo de regresión logística

<b>MODELO DE REGRESIÓN LOGÍSTICA</b>				
Min	1Q	Median	3Q	Max
-1,9626	-0,5848	0,5613	0,5613	2,2403
<b>Coefficientes</b>				
	Estimación	Error estándar	Estadístico z	Probabilidad de  z
Intercepto	-2,4248	0,3379	-7,176	7,15e-13 ***
Tiempo de espera	0,7454	0,2903	2,567	0,0103 *
Trato al usuario	1,5300	0,2979	5,137	2,80e-07 ***
Estado actual de las unidades	0,6121	0,3093	1,979	0,0479 *
Forma de conducción de transportista	1,3056	0,2669	4,891	1,00e-06 ***
<b>Nivel de significancia:</b> 0 ‘****’ 0,001 ‘**’ 0,01 ‘*’ 0,05 ‘.’ 0,1 ‘ ’ 1				
Desviación nula: 539,21 en 399 grados de libertad				
Desviación residual: 392,11 en 395 grados de libertad				
<b>Criterio de información AKAIKE</b> AIC: 402,11				

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 11 se presentan las características más importantes del modelo. Para determinar si las variables que forman parte del modelo son significativas se utiliza el valor crítico de la distribución normal que tiene un valor de 1,96 y el nivel de significancia preestablecido  $\alpha = 5\%$ . Se observa que los valores absolutos del estadístico  $z$  son superiores al valor crítico del 1,96, por tal razón las variables utilizadas son las adecuadas para la construcción del modelo; para corroborar lo expuesto con el estadístico  $z$  también se utiliza la probabilidad del estadístico  $|z|$  donde se visualiza que todas las probabilidades son inferiores al nivel de significancia preestablecido del 0,05, evidenciado de esta manera que las variables que utiliza el modelo son significativas y sirven para explicar y describir a la variable de estudio.

#### 4.3. Diagnóstico de la multicolinealidad en las variables predictoras

La multicolinealidad es una propiedad muy importante en los modelos de regresión. Es una condición que ocurre cuando las variables predictoras incluidas en un modelo se encuentran fuertemente correlacionadas entre sí. Para determinar la multicolinealidad se examinan los factores de inflación de la varianza (FIV). Con el análisis de los FIV se logra medir qué tanto aumenta la varianza de un coeficiente de regresión estimado. Si los resultados de los FIV son mayores a 5 se concluye que existe multicolinealidad en las variables predictoras y no se debe considerar dichas variables para la creación del modelo.

**Tabla 12.** Multicolinealidad para medir si las variables son las adecuadas

<b>Multicolinealidad entre las variables del modelo</b>	
<b>Variables del modelo</b>	<b>FIV</b>
Tiempo de espera	1,069567
Trato al usuario	1,168711
Estado actual de las unidades de transporte	1,137894
Forma de conducción de transportista	1,075165

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 12 se presentan los factores de inflación de la varianza (FIV) y se observan los valores de FIV de cada variable predictora que forman el modelo de regresión logística; estos valores son inferiores a 5, por tal razón estas variables no presentan multicolinealidad y son las adecuadas para la construcción del modelo logístico. Con

lo expuesto se concluye que estas variables si ayudan a determinar la calidad del servicio del transporte público urbano en la ciudad de Ambato.

#### 4.4. Bondad de ajuste del modelo

Para determinar la adecuación y el nivel de significancia del modelo de regresión logística se aplican las siguientes pruebas estadísticas: la ji-cuadrado, la de Hosmer y Lemeshow, la de Cox y Snell y la de Nagelkerke.

La prueba más utilizada para ver si el modelo es significativo es el estadístico ji-cuadrado, que utiliza un nivel de significancia de  $\alpha = 5\%$  y un contraste de hipótesis.

Ho: El modelo no es significativo ( $\beta_1 = \beta_2 = \dots, \beta_k = 0$ )

Ha: El modelo es significativo (Al menos un coeficiente  $\beta$  es diferente de 0)

Criterio de decisión

Si el  $p.value \geq \alpha$  se acepta la hipótesis nula

Si el  $p.value < \alpha$  se rechaza la hipótesis nula y se acepta la hipótesis alternativa

**Tabla 13.** Estadístico ji-cuadrado

Estadístico Ji cuadrado	
Valor de $R^2$	p.value de $R^2$
147,1007	$8,510006 \times 10^{-31}$

**Fuente** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 13 se presenta el estadístico ji-cuadrado donde se observa el valor crítico encontrado y el p.value de  $R^2$ . A partir del resultado del p.value del estadístico ji-cuadrado se rechaza la hipótesis nula porque el p.value es de  $8,510006 \times 10^{-31}$  y es menor al nivel de significancia del 0,05 por lo tanto, se acepta la hipótesis alternativa concluyendo que el modelo de regresión logística es significativo y existe al menos un coeficiente  $\beta$  que es diferente de 0.

Para determinar la bondad de ajuste del modelo se utilizan tres pruebas estadísticas: la de Hosmer y Lemeshow, la de Cox y Snell y la de Nagelkerke que utilizan un nivel de significancia de  $\alpha = 5\%$  y un contraste de hipótesis.

Ho: El modelo está bien ajustado.

Ha: El modelo no está bien ajustado.

Criterio de decisión

Si el  $p.value \geq \alpha$  se acepta la hipótesis nula

Si el  $p.value < \alpha$  se rechaza la hipótesis nula y se acepta la hipótesis alternativa

**Tabla 14.** Estadísticos para determinar el ajuste del modelo

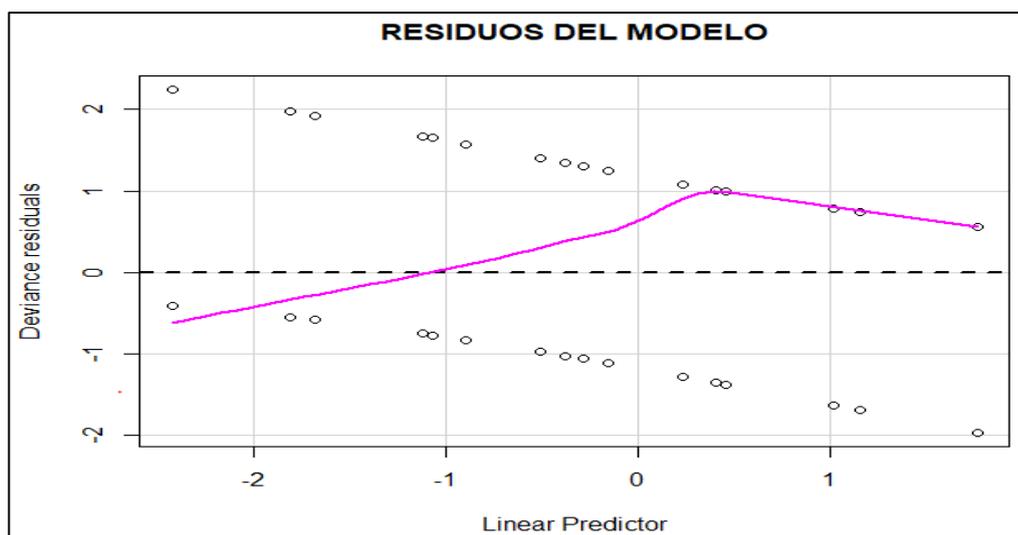
<b>Pruebas estadísticas</b>	<b>p.value de los estadísticos</b>
Hosmer y Lemeshow	0,272808
Cox y Snell	0,307711
Nagelkerke	0,4156869

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 14 se observan tres pruebas estadísticas que ayudan a determinar si el modelo está bien ajustado, al observar que los valores del p.value de las tres herramientas estadísticas son mayores que el nivel de significancia del 0,05 se acepta la hipótesis nula, por lo cual se concluye que el modelo está ajustado correctamente y es adecuado para determinar la calidad del servicio del transporte público urbano en la ciudad de Ambato.

#### **4.5. Residuos del modelo**

Los residuos del modelo son el resultado de la resta entre los datos reales y los datos pronosticados. En la siguiente figura se presenta los residuos del modelo



**Figura 9.** Residuos del modelo.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 9 se presenta el comportamiento de los residuos del modelo. Allí se verifica que dichos residuos no muestran una tendencia que se aleje demasiado del 0, por lo cual se concluye que el modelo se ajusta correctamente a los datos estudiados.

#### 4.6. Interpretación de los odds ratio (OR)

En un modelo de regresión logística no se pueden interpretar los coeficientes debido a que se encuentran relacionados con un logaritmo de la razón Odd. Para una interpretación adecuada de los coeficientes se debe aplicar el logaritmo natural y así obtener una ecuación lineal. En la siguiente tabla se presentan los Odds ratio (OR) de cada una de las variables predictoras.

**Tabla 15.** Interpretación de los odds ratio (OR)

Variabes del modelo	Coefficientes	Razón Odd
Intercepto	-2,4248	0,08849889
Tiempo de espera	0,7454	2,10726683
Trato al usuario	1,5300	4,61813561
Estado actual de las unidades	0,6121	1,84425506
Forma de conducción del transportista	1,3056	3,69007630

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 15 se presenta el valor de los odds ratio de cada variable predictor, se observa que todos los resultados son positivos, por tal razón al incrementar una

variable predictora existirá la probabilidad que se incremente la variable respuesta. La variable tiempo de espera (apropiado – inapropiado) tiene un valor de 2,10726683, por lo cual la probabilidad de éxito que el servicio del transporte sea seguro es el doble cuando el tiempo es apropiado. Con respecto a la variable trato al usuario (adecuado – inadecuado) tiene un resultado de 4,6181356, por tal razón la probabilidad de éxito que la calidad del servicio sea seguro es cuatro veces mayor si el trato al usuario es adecuado. Mientras la variable estado actual de las unidades (bueno – malo) tiene un valor de 1,84425506 por lo que al aumentar una unidad de transporte en buen estado la probabilidad de éxito que la calidad del servicio sea seguro se incrementa en el 84,42%. Por último, la variable forma de conducción del transportista (apropiada – inapropiada) tiene un resultado de 3,69007630, por tal razón la probabilidad de éxito que la calidad del servicio del transporte sea segura es tres veces mayor si la forma de conducir del transportista es adecuada.

Con las variables estimadas y comprobadas que son las adecuadas para la construcción del modelo, lo siguiente es formar la ecuación que determinará la calidad del servicio del transporte público urbano de la ciudad de Ambato.

A partir de la ecuación 1 el modelo de regresión logística se expresa de la siguiente manera:

$$P (Y_i = 1/X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)}}$$

donde

P es la probabilidad de ocurrencia según las variables.

$Y_i$  es la variable dependiente o respuesta.

$X_i$  son las variables predictoras.

$\beta$  representa los coeficientes estimados del modelo

Reemplazando los coeficientes estimados del modelo según la tabla 10, la ecuación que determina la calidad del servicio de transporte público urbano de la ciudad de Ambato se expresa de la siguiente forma.

$$P (Y_i = 1/X_i) = \frac{1}{1 + e^{-(2,4248 + 0,7454X_1 + 1,5300X_2 + 0,6121X_3 + 1,3056X_4)}}$$

#### 4.7. Determinación del punto de corte y la curva Roc

El punto de corte o frontera de decisión determina la probabilidad de éxito o fracaso de un evento de estudio.

En este caso si una probabilidad es superior a la frontera de decisión la calidad del servicio del transporte es seguro, pero si es inferior la calidad del servicio será insegura. Otra forma para determinar si el modelo clasifica correctamente es mediante el área bajo la curva Roc, el valor del área debe ser superior a 0,70 para que el modelo clasifique adecuadamente los datos. En la siguiente tabla se presenta tanto el punto de corte como el área bajo la curva.

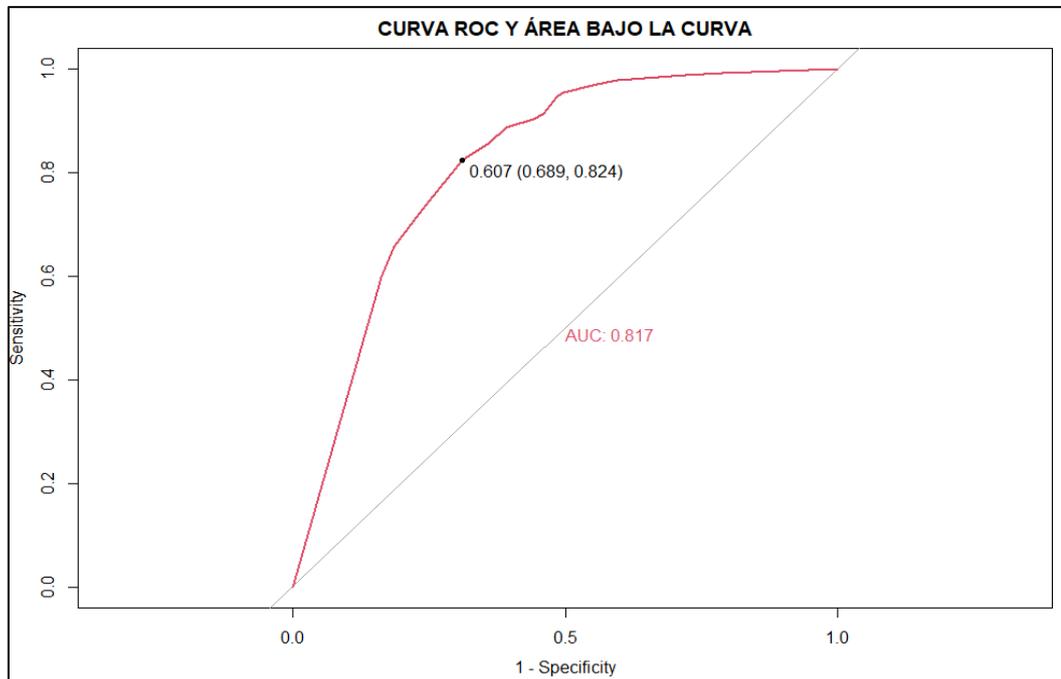
**Tabla 16.** Punto de corte y área bajo la curva Roc

Punto de corte	Clasificación	Área bajo la curva
0,6042	0,77	0,8170
60,43%	77%	81,70%

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 16 se presenta el valor del punto de corte, el porcentaje de clasificación correcta de los datos y el área bajo la curva. El punto de corte tiene un valor del 0,6042 o del 60,42%, el mismo que servirá como frontera de decisión para determinar si la calidad del servicio es seguro o inseguro, si un dato es superior a 0,6042 se concluye que la calidad del servicio es seguro, pero si es inferior a dicho valor la calidad será insegura. También se visualiza el porcentaje de clasificación del modelo el mismo que tiene un valor de 0,770 o el 77%, por lo cual este modelo clasifica 308 datos de forma correcta de los 400 datos. Por último, el área bajo la curva Roc tiene un valor de 0,8170 y es superior 0,70 por tal razón se concluye que el modelo desarrollado es adecuado para determinar la calidad del servicio de transporte público urbano en la ciudad de Ambato.

Una vez conocida la frontera de decisión y el área de la curva Roc. Lo siguiente es interpretar el comportamiento de la curva ROC. Esta curva representa el poder discriminante de un modelo logit, para ello utiliza un intervalo de [0; 1]. Si el resultado de la curva Roc es cercano a 1 se concluye que el modelo logit es el adecuado para clasificar eventos binarios, para ello se utiliza un punto de referencia del 0,70.



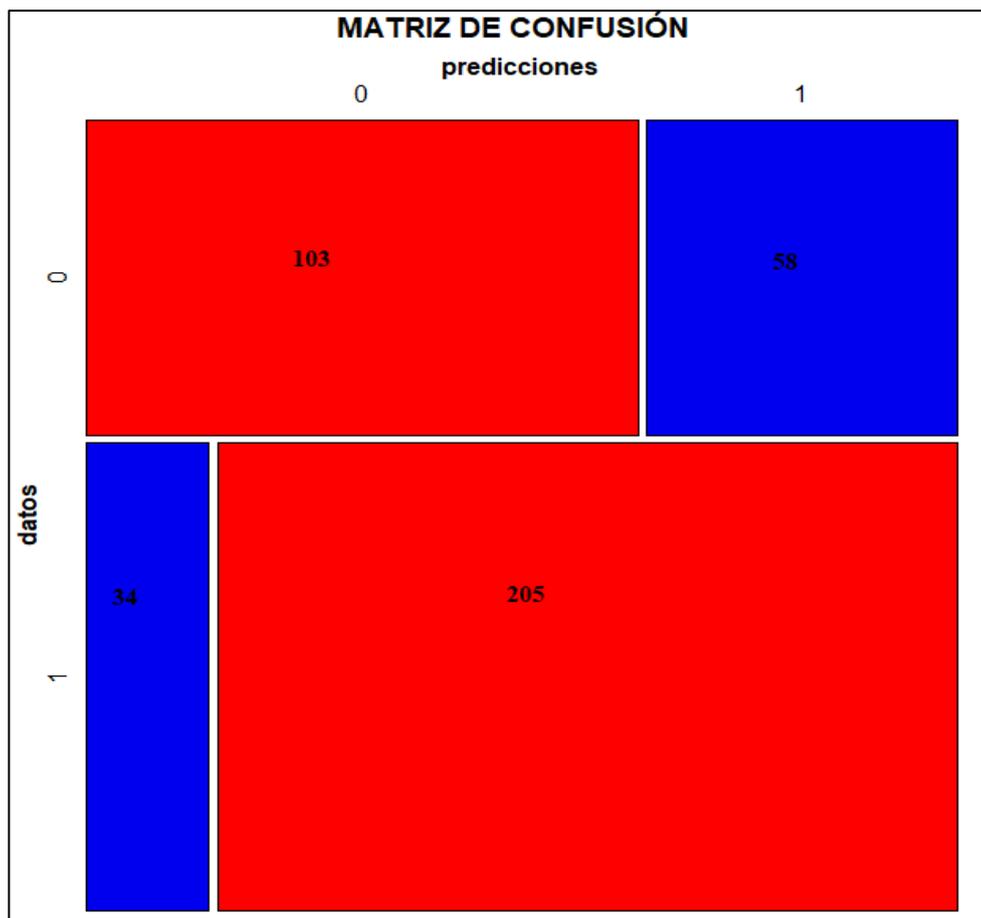
**Figura 10.** Curva ROC.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 10 se presenta el gráfico de la curva Roc, en donde la línea de color rojo representa a la curva Roc, también se visualiza el punto de corte y el área bajo la curva Roc. Mediante este gráfico se observa que la curva Roc está acercándose más al 1 y su valor es de 0,817 el cual es superior a 0,70 por lo cual se concluye que el modelo de regresión logística desarrollado es el adecuado para clasificar, explicar y determinar la calidad del servicio del transporte público urbano en la ciudad de Ambato

#### 4.8. Matriz de confusión

Con el modelo ajustado correctamente lo siguiente es determinar la matriz de confusión, la misma que es una herramienta que permite observar el desempeño de un algoritmo desarrollado mediante un modelo de regresión. En la siguiente figura se presenta la matriz de confusión según las predicciones y los datos pertenecientes a la variable dependiente.



**Figura 11.** Matriz de confusión.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 11 se presenta la matriz de confusión, en la cual se observa que el modelo clasifica correctamente 308 datos e incorrectamente 92. De los datos que clasifica adecuadamente 205 datos son clasificados como exitosos, los mismos que representan a la calidad del servicio del transporte que es seguro y 103 datos son clasificados como fracaso y pertenecen a la calidad del servicio de transporte que es inseguro.

#### 4.9. Predicciones

Una vez determinado que el modelo de regresión logística es adecuado para clasificar, explicar y predecir los datos, se procede a determinar si la calidad del servicio de transporte público urbano en la ciudad de Ambato es segura o insegura. En la siguiente tabla se presentan los resultados alcanzados con el modelo logit.

**Tabla 17.** Predicción y clasificación

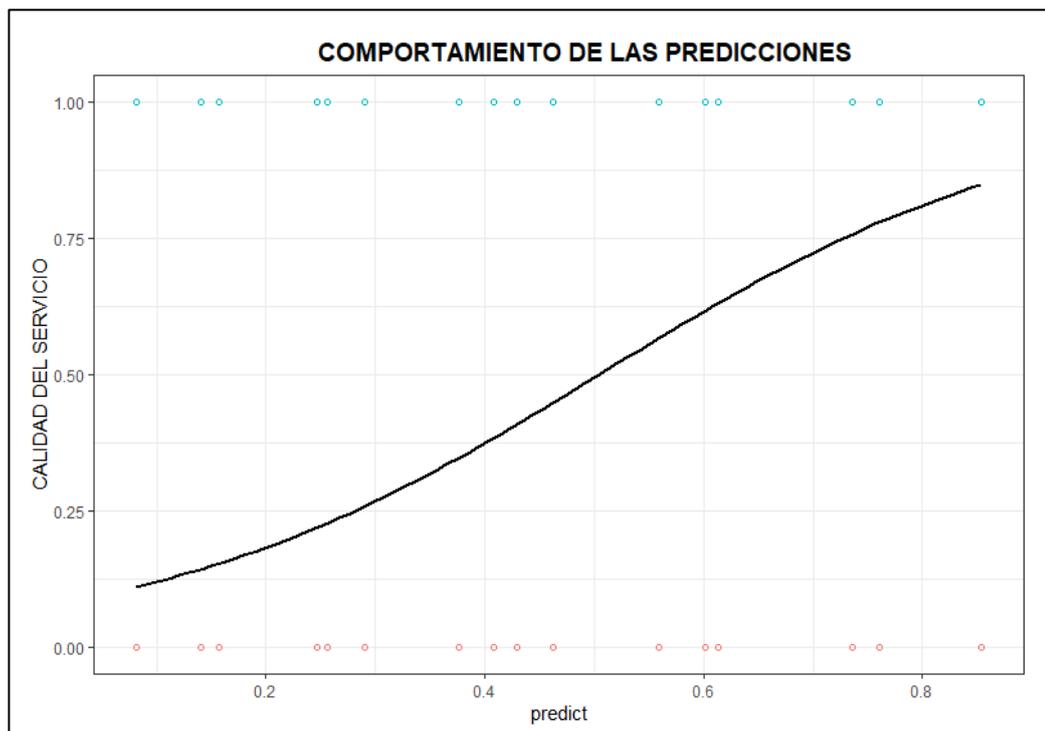
Tiempo de espera	Trato al usuario	Estado actual de las unidades	Forma de conducción del transportista	Calidad del servicio	Predicciones de Probabilidad	Clasificación
núm.	núm.	núm.	núm.	núm.	núm.	Cual.
[1:400]	[1:400]	[1:400]	[1:400]	[1:400]	[1:400]	[1:400]
1	0	0	0	0	0,15717844	Inseguro
0	1	1	1	0	0,73554650	Seguro
1	1	0	0	1	0,46272368	Inseguro
1	1	0	0	1	0,46272368	Inseguro
1	1	1	1	1	0,85425113	Seguro
1	0	1	0	0	0,25591725	Inseguro
1	1	1	1	0	0,85425113	Seguro
1	1	1	0	1	0,61365283	Inseguro
0	0	0	0	0	0,08130361	Inseguro
0	0	0	0	0	0,08130361	Inseguro
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>	<b>n</b>

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la tabla 17 se presenta una muestra de las clasificaciones y predicciones de las probabilidades alcanzadas con el modelo de regresión logística según las 4 variables predictoras. Al comparar cada probabilidad obtenida a través del modelo con el punto de corte que tiene un valor de 0,6142, se logra determinar la calidad del servicio de transporte urbano de la ciudad de Ambato. Se observa que la predicción del dato 1 es de 0,1571 y al comparar con el punto de corte se observa que es inferior a dicho valor, por lo cual la calidad del servicio para la encuesta 1 es insegura, pero si se toma la probabilidad de predicción del dato 5 cuyo valor es de 0,8542 y se compara con el punto de corte, se observa que es superior a dicho valor, por tal razón la calidad del servicio es seguro. Esta comparación se realiza para cada probabilidad de predicción

encontrada con el modelo según las 400 encuestas. A partir de las probabilidades de predicción del modelo y la matriz de confusión se observa que 205 datos son clasificados como seguros y 103 datos son clasificados como inseguros, los demás datos no se logran clasificar mediante el uso del modelo debido a que las variables predictoras pueden encontrarse mediante correlacionadas.

Conocida la clasificación y las probabilidades de predicciones también se presenta la siguiente figura para determinar el comportamiento de las predicciones del modelo.



**Figura 12.** Comportamiento de las predicciones.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

En la figura 12 se presenta el comportamiento de las probabilidades de predicciones alcanzadas con el modelo de regresión logística desarrollado. Según la tabla anterior se obtuvo 205 datos seguros y 103 datos inseguros, por lo que el modelo clasifica el 51,25% de las encuestas aplicadas a los usuarios del transporte público de la ciudad de Ambato como seguros y el 25,75% de las encuestas según los criterios presentados por los usuarios establecen que el servicio de transporte es inseguro.

## CAPÍTULO V

### CONCLUSIONES Y RECOMENDACIONES

#### 5.1. CONCLUSIONES

Las conclusiones a las que se llegó con el desarrollo del presente trabajo de investigación son:

- Se determinó un tamaño de la muestra para la obtención de la información requerida para el desarrollo del trabajo de 400 entrevistados. Se aplicó la encuesta a personas seleccionadas aleatoriamente entre los usuarios del transporte público urbano de la ciudad de Ambato, la cual se basó en un cuestionario de 10 ítems con preguntas de opción múltiple cerradas, logrando así obtener el criterio que tienen sobre la calidad del servicio que reciben del transporte público urbano.
- Mediante el análisis de correlación y de significancia de las variables que forman parte del modelo se logró determinar el peso de las variables independientes del modelo: tiempo de espera, trato al usuario, estado actual de las unidades de transporte y la forma de conducción del transportista. Con el gráfico de correlación se evidenció que todas las variables están relacionadas moderadamente, y los pesos determinados de las variables son: trato al usuario con el 50%, forma de conducción del transportista con el 45%, estado actual de las unidades con 37% y por último el tiempo de espera con el 33%.

Utilizando los modelos de regresión logística se desarrolló un modelo matemático que explica, clasifica, predice y determina la calidad del servicio del transporte público urbano de la ciudad de Ambato, el mismo que se expresa de la siguiente

$$\text{manera } P(Y_i = 1/X_i) = \frac{1}{1 + e^{-(-2,4248 + 0,7454X_1 + 1,5300X_2 + 0,6121X_3 + 1,3056X_4)}}$$

- Para comprobar el funcionamiento del modelo se utilizaron pruebas estadísticas como el ji-cuadrado, Hosmer y Lemeshow, Cox y Snell y la Nagelkerke, con las cuales se logró identificar que todas las variables predictoras que construyen el modelo son significativas. Mediante el modelo desarrollado se logró determinar la calidad del servicio del transporte público urbano de la ciudad de Ambato basado en la información recopilada a través de las encuestas aplicadas, logrando determinar que el 51,25% de las personas consideran que este transporte es seguro y 25,75% de los usuarios consideran que es inseguro.

## 5.2. RECOMENDACIONES

- Para mejorar la calidad del servicio del transporte público se recomienda que las autoridades del Gad Municipal de la ciudad de Ambato se preocupen y evalúen constantemente a las cooperativas que brindan el servicio de transporte ya que los usuarios presentan el criterio que el transporte es medianamente seguro.
- Se recomienda a las autoridades de las cooperativas de transporte público urbano ejecuten capacitaciones permanentes a sus choferes sobre el trato al usuario y la forma de conducción. También deberían realizar las gestiones necesarias para lograr contar con unidades de transporte en buen estado.
- Se sugiere emplear las herramientas estadísticas como: la ji cuadrada que determina la significancia de las variables y del modelo. Las pruebas de Hosmer y Lemeshow, la de Cox y Snell y la de Nagelkerke ayudan a determinar si el modelo se encuentra ajustado correctamente, con estas herramientas estadísticas se logra crear modelos de regresión logística con una mayor capacidad predictora.
- Se recomienda utilizar los modelos de regresión logística porque emplean variables numéricas, categóricas y cualitativas para realizar las predicciones y clasificaciones de los valores de una variable de estudio.

### 5.3. BIBLIOGRAFÍA

- [1] B. Gamarra y J. L. Delgado, Artists, *Calidad del Servicio de Transporte Urbano en la ciudad de Cusco 2014*. [Art]. Universidad Nacional de San Antonio Abad del Cusco, 2014.
- [2] L. Pizarro, Artist, *Factores explicativos de la Demanda de Calidad en el servicio de las empresas de transporte público de pasajeros en el distrito de Laredo 2015*. [Art]. Universidad Nacional de Trujillo, 2015.
- [3] G. Moran, Artist, *Análisis a la calidad del servicio de transporte urbano de buses en la ciudad de Esmeraldas desde la óptica del usuario*. [Art]. Pontificia Universidad Católica del Ecuador , 2016.
- [4] J. P. Ramirez y R. Yataco, Artists, *Relación entre la calidad del servicio logístico y satisfacción de los clientes de un operador logístico de lima 2019*. [Art]. Universidad San Ignacio de Loyola, 2019.
- [5] L. D. Berrones, P. Cano, D. Sánchez y J. MARTINEZ, «Entregas tardías o incorrectas en el autotransporte de carga y su relación con las condiciones laborales de los choferes: Un modelo de regresión logística,» *Nova Scientia*, vol. 10, nº 20, pp. 413-441, 2018.
- [6] K. P. Prado, J. R. d. Castro y A. Talaverano, «Factores asociados a somnolencia diurna en conductores de transporte público de Lima Metropolitana,» *Peru Med Exp Salud Publica*, vol. 36, nº 4, pp. 629-635, 2019.
- [7] E. Aguilar, J. Heald y R. Guerrero, «Análisis de competitividad y movilidad en destinos turísticos del Estado de Guanajuato a través de un modelo de regresión logística,» vol. 17, nº 1, pp. 105-122, 2017.
- [8] C. Mayorga, M. Ruiz y D. Aldas, «Percepciones acerca de la contaminación del aire generada por el transporte urbano en Ambato, Ecuador,» *Revista Espacios*, vol. 41, nº 17, p. 11, 2020.
- [9] K. F. Camarena, Artist, *Fidelización de los clientes y la Calidad de Servicio de la Empresa de Transportes América Express de la Ciudad de Trujillo,2018*. [Art]. Univeridad Cesar Vallejo, 2018.
- [10] N. V. Luque, Artist, *Proceso de capacitación y calidad de servicio en la empresa de transporte público urbano "José María Arguedas" Andahuaylas 2015*. [Art]. Universidad Nacional José María Arguedas, 2015.

- [11] S. M. Mena y S. Ynchicsana, Artists, *Cumplimiento de la normatividad del transporte público y calidad del servicio en la empresa Yllary Qosqo periodo octubre - noviembre Cusco, 2017*. [Art]. Universidad Peruana Austral del Cusco, 2017.
- [12] D. G. A. A. P. A. G. N. Barón, «Modelo de accesibilidad a sistemas de transporte público según la experiencia de usuario en el contexto urbano,» *Proyección*, vol. 11, nº 21, pp. 80-105, 2017.
- [13] N. L. Rugel, Artist, *Análisis de la calidad de servicio percibida por el usuario de transporte público de la ciudad de Guayaquil*. [Art]. Universidad Politécnica Salesiana, 2019.
- [14] E. E. Beneitez, Artist, *Mejora en la recolección de datos y la modelización de la calidad percibida en sistemas de transporte público*. [Art]. Universidad de Cantabria, 2020.
- [15] T. C. Guerra, Artist, *Caracterización y determinación de las variables que inciden en la accidentalidad en usuarios de moto en Bogotá, por medio de regresión logística múltiple*. [Art]. Univeridad Los Libertadores, 2019.
- [16] G. J. A. J. I. Llamuca, «Evaluación de la calidad del servicio de transporte urbano en bus de la ciudad de Ambato,» *Visionario Digital*, vol. 3, nº 2, pp. 26-46, 2019.
- [17] R. Apaza, Artist, *Determinación de modelos matemáticos de caracterización de flujo vehicular para el centro histórico de la ciudad de Juliaca*. [Art]. Universidad Nacional del Altiplano, 2017.
- [18] C. T. S. N.A. Silva, Artist, *La calidad del servicio de transporte urbano en la ciudad de Cuenca*. [Art]. Universidad Salesiana de Cuenca, 2017.
- [19] D. Vivas, Artist, *Calidad del servicio en el transporte público de la Cooperativa Frontera Norte de la ciudad de Tulcán*. [Art]. Universidad Politécnica Estatal del Carchi, 2018.
- [20] M. Navarro, Artist, *La gestión de calidad del servicio y atención al cliente en la micro y pequeñas empresas del rubro transporte público buses ruta Sullana - Piura del distrito de Sullana, año 2018*. [Art]. Universidad Catolica Los Angeles de Chimbote, 2018.
- [21] E. U. Joaquin Aldás, Analisis multivariante aplicado con R, Madrid: Editorial SA, 2017.
- [22] M. Triola, Estadística, México: Pearson, 2006.
- [23] F. Barahona y M. Marva, Una introducción a los conceptos de estadística, España: Lulu, 2016.
- [24] E. B. Morell, Bioestadística básica para investigadores con SPSS, España - Madrid, España : Ediciones ECOE, 2014.

- [25] A. J. L. Montoya, Artist, *Comparación de dos modelos de regresión en fiabilidad*. [Art]. Universidad de Granada, 2011.
- [26] A. Peskova, *Sujetos predominales en el español porteño*, Boston: Pearson , 2015.
- [27] E. Taucher, *Bioestadística*, Santiago de Chile: Editorial Universitaria , 1999.
- [28] L. Fuentes, «Metodología para la elección de punto de corte óptimo para dicotomizar covariables continuas,» *Rev Cubana Genet Comunit* , vol. 3, nº 1, p. 7, 2013.
- [29] A. Rial y J. Varela, *Estadística práctica para la investigación en ciencias de la salud*, La Coruña: Netbiblo, 2014.
- [30] R. Benítez, G. Escudero, S. Kanaan y D. Masip, *Inteligencia artificial avanzada*, Barcelona: Editorial UOC, 2015.
- [31] Ecu Red, Ambato, «EcuRed de Ambato,» 08 06 2019. [En línea]. Available: [https://www.ecured.cu/Ambato\\_\(Ecuador\)#Descripci.C3.B3n](https://www.ecured.cu/Ambato_(Ecuador)#Descripci.C3.B3n).
- [32] S. Hernández, c. Fernandez y P. Baptista, *Metodología de la Investigación*, Sexta ed., México: McGRAW-HILL / Interamericana Editores, S.A. DE C.V., 2014.
- [33] J. Cerda y L. Cifuentes, «Uso de curvas roc en la investigación clínica,» *Revista chilena de infectología*, vol. 29, nº 2, p. 11, 2012.

## 5.4. ANEXOS

### Anexo 1

#### ENCUESTA DE LA CALIDAD DEL SERVICIO DE TRASPORTE URBANO

Dirección de correo electrónico: .....

Encuesta para determinar el nivel de satisfacción de los usuarios con respecto al uso del servicio de transporte público urbano en la ciudad de Ambato

Marque con una (x) donde corresponda

#### I. ASPECTOS GENERALES

##### 1. GENERO

MASCULINO..... FEMENINO.....

##### 2. EDAD

De 10 a 19 años	
De 20 a 35 años	
De 36 años y más	

#### II. ASPECTOS ESPECÍFICOS

1. ¿Cuál línea de transporte público urbano ocupa por lo general para su desplazamiento?:

Jerpazol	
Libertadores	
Tungurahua	
Unión	
Vía Flores	

2. ¿Cuál es la principal razón por la que utiliza el servicio de transporte público urbano?

Para ir a estudiar	
Para ir de compras	
Para ir a trabajar	
Para ir a pasear	
Para hacer diligencias	
Otra	

### III. TIEMPO DE ESPERA

Siendo usted un usuario frecuente de una línea de servicio de transporte público urbano considera que los horarios de servicio son:

Apropiados	
Inapropiados	

### IV. TRATO AL USUARIO

Considera usted que el trato que ha recibido por parte de las personas a cargo de la unidad de transporte (conductor-ayudante), fue:

Excelente	
Bueno	
Regular	
Malo	

### V. ESTADO ACTUAL DEL SERVICIO

Considerando las unidades de transporte de la línea o compañía que con mayor frecuencia utiliza diariamente diría usted que la apariencia en asientos, piso y entorno fue:

Excelente	
Buena	
Regular	
Mala	

VI. FORMA DE CONDUCCIÓN DEL TRANSPORTISTA

En su experiencia general, considera usted que la forma de conducir de los conductores del transporte público urbano es:

Excelente	
Buena	
Regular	
Mala	

VII. TIEMPO DE RUTA

En términos generales, considera usted que el tiempo que demora el servicio de transporte en cumplir el recorrido hasta llegar a su destino es

Apropiado	
Inapropiado	

VIII. CALIDAD DEL SERVICIO

En su experiencia general, como se siente al utilizar el servicio de transporte público urbano:

Seguro (a)	
Inseguro (a)	

## Anexo 2

### Ciudad de Ambato



**Figura 13.** Ciudad de Ambato.  
**Fuente:** Gad Municipal de Ambato.

### Rutas de transporte de la ciudad de Ambato



**Figura 14.** Línea de autobuses de Ambato.  
**Fuente:** Gad Municipal de Ambato.

### Anexo 3

#### Base de datos binarizada

	B	C	D	E	F
1					
2	TIEMPO DE ESPERA	TRATO AL USUARIO	ESTADO ACTUAL DE LAS UNIDADES	FORMA DE CONDUCCION DEL	CALIDAD DEL SERVICIO
3	1	0	0	0	0
4	0	1	1	1	0
5	1	1	0	0	1
6	1	1	0	0	1
7	1	1	1	1	1
8	1	0	1	0	0
9	1	1	1	1	0
10	1	1	1	0	1
11	0	0	0	0	0
12	0	0	0	0	0
13	0	0	1	1	1
14	1	1	1	1	1
15	0	0	0	0	0
16	1	0	0	0	0
17	1	1	1	0	0
18	1	1	1	1	1
19	0	1	1	1	1
20	1	1	1	1	1
21	1	1	1	1	1
22	0	0	1	0	0
391	1	1	1	1	1
392	1	1	1	1	1
393	0	0	1	1	1
394	0	1	1	1	1
395	0	0	0	0	0
396	1	1	1	0	1
397	0	1	1	1	1
398	1	0	1	1	1
399	1	1	1	1	0
400	1	1	1	1	1
401	1	1	1	0	0
402	1	1	1	1	1

**Figura 15.** Base de datos binarizada.

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.

## Anexo 4

### Paradas más frecuentes de Ambato

#### PARADA JUNTO AL HOSPITAL



**Figura 16.** Parada el hospital.

**Fuente:** Tomada por el autor en la ciudad de Ambato en el 2020.

#### PARADA EN EL REDONDEL DE IZAMBA



**Figura 17.** Parada en el redondel de Izamba.

**Fuente:** Tomada por el autor en la ciudad de Ambato en el 2020.

## PARADA EN EL MERCADO MAYORISTA



**Figura 18.** Parada en el mercado mayorista.

**Fuente:** Tomada por el autor en la ciudad de Ambato en el 2020.

## PARADA EN EL MERCADO MODELO



**Figura 19.** Parada en el mercado modelo.

**Fuente:** Tomada por el autor en la ciudad de Ambato en el 2020.

## PARADA EN EL MOLL DE LOS ANDES – MUNICIPIO



**Figura 20.** Parada en el Moll de los Andes - Municipio.

**Fuente:** Tomada por el autor en la ciudad de Ambato en el 2020.

## PARADA EN EL REDONDEL DE HUACHI CHICO



**Figura 21.** Parada en el redondel de Huachi Chico.

**Fuente:** Tomada por el autor en la ciudad de Ambato en el 2020.

## Anexo 5

### Script del Modelo Matemático

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
calidad.R* x predict x transporte x
Source on Save Run Source
1 #LIBRERIAS A UTILIZAR
2 library(ggplot2)
3 library(GGally)
4 library(PerformanceAnalytics)
5 library(dplyr)
6 library(mlogit)
7 library(corrplot)
8 library(readxl)
9 library(vcd)
10 library(pROC)
11 library(MASS)
12 library(visdat)
13 library(caret)
14 library(car)
15 library(class)
16 library(e1071)
17 library(Informationvalue)
18
19 attach(transporte)
20 names(transporte)
21 str(transporte)
22
23 #CORRELACIÓN ENTRE VARIABLES
24 chart.correlation(transporte)
25
26 #CREACIÓN DEL MODELO
27 modelo <- glm(`CALIDAD DEL SERVICIO`~ `TIEMPO DE ESPERA` + `TRATO AL USUARIO`
28 + `ESTADO ACTUAL DE LAS UNIDADES`+
29 + `FORMA DE CONDUCCION DEL TRANSPORTISTA`,
30 data = transporte,family = binomial(link = "logit"))
31 #RESUMEN DEL MODELO
32 summary(modelo)
33
34 #SIGNIFICANCIA DE LAS VARIABLES PREDICTORAS
35 # Al ubicarse en el resumen del modelo en el valor absoluto de z todos son
36 # mayores a 1,96 por tal razón todas las variables predictoras son significativas,
37 # así también se corrobora mediante la probabilidad de z en donde todas las
38 # probabilidades son menores al nivel de significancia del 0,05.
39
40 #MULTICOLINEALIDAD
41 vif(modelo)
42
43 # al observar que los resultados de las variables son menores a 5 se concluye
44 # que no existe multicolinealidad por lo cual las variables son las adecuadas
45 # para la construcción del modelo
46
47 #PRUEBA DE SIGNIFICANCIA
48 #Prueba de significancia para medir si el modelo es el adecuado
49 #Ho:El modelo no es significativo ( $\beta_1 = \beta_2 = \dots, \beta_k = 0$ ); si p.value  $\geq \alpha$ 
50 #Ha:El modelo es significativo (Al menos un coeficiente  $\beta \neq 0$ ); si p-value  $< \alpha$ 
51
52 #ESTADISTICO JI.CUADRADO
53 #valor critico encontrado del estadístico
54 with(modelo, null.deviance-deviance)
55 #p.value del estadístico
56 with(modelo, pchisq(null.deviance-deviance,df.null-
57 df.residual, lower.tail = FALSE ))
58 # como el p.value es de 8,510006e-31 y es menor al 0,05 se concluye que
59 #el modelo es significativo
60
61 # PRUEBAS PARA MEDIR LA BONDA DE AJUSTE
62 #Ho: El modelo esta bien ajustado; si p.value  $\geq \alpha$ 
63 #Ha: El modelo no esta bien ajustado; si p-value  $< \alpha$ 
64
```

```

65 #MEDIDA DE HOSMER Y LEMESHOW
66 H1 <- (modelo$null.deviance-modelo$deviance)/modelo$null.deviance
67 H1
68 #MEDIDA DE COX Y SENLL'S
69 Cs <- 1- exp((modelo$deviance- modelo$null.deviance)/400)
70 Cs
71 #MEDIDA DE NEGELKERKER
72 Ne<- Cs/(1- exp(-(modelo$null.deviance)/400))
73 Ne
74 # como los p.value de las pruebas son;0,2728; 0,3077;0,4156 y sus valores son
75 #mayores a 0,05 se concluye que el modelo esta bien ajustado.
76
77 #RESIDUOS DEL MODELO
78 residualPlots(modelo, type ="deviance") +title("RESIDUOS DEL MODELO")
79
80 #INTERPRETACION DE LOS ODDS RATIO
81 #interpretacion de los parametros a traves de odds rations o del
82 #impacto de la probabilidad
83 exp(coefficients(modelo))
84 # se transforma las variables predictoras a forma exponencial
85
86 # EVALUACION DE LA CAPACIDAD PREDICTIVA
87 #comparar con las observaciones las predicciones
88 transporte$pred<- modelo$fitted.values
89 predict <- data.frame(modelo$fitted.values)
90 # punto de corte en porcentaje
91 punto<- optimalCutoff(transporte$`CALIDAD DEL SERVICIO`, predict,
92                       optimiseFor = "Both")
93 punto
94 #es la frontera de desicion para posteriores observaciones
95
96 #PORCENTAJE DE CLASIFICACION EN LAS OBSERVACIONES
97 observaciones<- 1 - misClassError(transporte$`CALIDAD DEL SERVICIO`,
98                                   predict, threshold = punto)
99 observaciones
100
101 # el modelo clasifica en un 77% a los datos
102 # comparacion entre la prediccion y los datos
103
104 # MATRIZ DE CONFUSIÓN
105 predicciones <- ifelse(test = modelo$fitted.values > 0.5, yes = 1, no = 0)
106 matriz_confusion <- table(modelo$model$`CALIDAD DEL SERVICIO`, predicciones,
107                            dnn = c("datos", "predicciones"))
108 matriz_confusion
109 mosaic(matriz_confusion, shade = T, colorize = T,
110         gp = gpar(fill = matrix(c("red", "blue2", "blue2", "red"), 2, 2),))
111 #mediante la matriz de confusión se observa que en las encuestas y predicciones
112 #que dicen que es apropiado es 205 pero 103 no son apropiadas y que 92 datos
113 #no fueron clasificados correctamente
114
115 #CURVA ROC Y AREA DE PREDICCIÓN
116 predict<- predict(modelo, type = "response")
117 predict
118 roc <- PROC::roc(transporte$`CALIDAD DEL SERVICIO`,predict, auc = TRUE, c1 = TRUE)
119 print(roc)
120 plot.roc(roc,legacy.axes= TRUE, print.thres = "best", print.auc = TRUE,
121          auc.polygon = FALSE, max.auc.polygon = FALSE, auc.polygon.col= "red",
122          col = 2, grid = FALSE, main ="COMPORTAMIENTO DE LAS PREDICCIONES")
123 #mediante este grafico se determina que el modelo clasifica de manera correcta en
124 #0.8175 que al realizarlo de manera aleatoria
125
126 # GRAFICO DEL MODELO
127 ggplot(data = transporte, aes(x = predict, y = `CALIDAD DEL SERVICIO`)) +
128   geom_point(aes(color = as.factor(`CALIDAD DEL SERVICIO`)), shape = 1) +
129   geom_smooth(method = "glm", method.args = list(family = "binomial"),
130             color = "black", se = FALSE) + theme_bw() +
131   theme(legend.position = "none")
132

```

**Figura 22.** Script del modelo

**Fuente:** Elaborado por el autor en base a los datos obtenidos de la encuesta aplicada.