

**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD
DEL CUSCO**

**ESCUELA DE POSTGRADO
MAESTRÍA EN ESTADÍSTICA**



**APLICACIÓN DE REGRESION LOGISTICA MULTINOMIAL EN
EL ESTUDIO DE PREFERENCIAS DE LOS ESTUDIANTES DE
LAS UNIVERSIDADES DEL CUSCO, 2019**

TESIS PRESENTADO POR:

**BR. DAVID RODRIGUEZ
QUISPE**

**PARA OPTAR EL GRADO
ACADÉMICO DE MAESTRO EN
ESTADÍSTICA**

ASESOR:

**DR. CLETO DE LA TORRE
DUEÑAS**

CUSCO – PERÚ

2021

DEDICATORIA

A mis padres quienes me dieron vida, educación, apoyo y consejos. A mis compañeros de estudio, a mis maestros y amigos y principalmente a mi amigo Abraham, quienes sin su ayuda nunca hubiera podido hacer esta **tesis**. A todos ellos se los agradezco desde el fondo de mi alma. Para todos ellos hago esta **dedicatoria**

AGRADECIMIENTO

A mi asesor de tesis Dr. Cleto de la Torre Dueñas por su generosidad y predisposición al brindarme la oportunidad de recurrir a su capacidad y experiencia en un marco de confianza, afecto y amistad, fundamentales para la concreción de este trabajo.

También expreso mi agradecimiento a todos los docentes de la maestría de Estadística de la Universidad de San Antonio Abad del Cusco por sus enseñanzas en mi formación profesional.

David R.

RESUMEN

El presente trabajo de investigación tiene por finalidad determinar los factores asociados a las preferencias de los estudiantes de las universidades del Cusco, mediante la regresión logística multinomial. Donde se perciben las preferencias como lo que la persona quería ser en su vida adulta o profesional teniendo en cuenta diversos aspectos, más allá del mero interés y como resultado de la comparación entre distintas opciones, que implicaría medición y evaluación para conocer si las preferencias de los sujetos en un momento determinado, se concentran o dispersan en áreas y grupos vocacionales. El estudio de investigación es nivel básico, alcance descriptivo – correlacional, enfoque cuantitativo y de diseño no experimental transversal. Para el estudio se aplicó encuestas a los estudiantes de quinto de secundaria de la ciudad del Cusco, obteniendo como conclusión que los factores asociados a la elección de una universidad en la ciudad del Cusco, son las siguientes: Estudios se realizan sin interrupciones por tomas de local, el ingreso económico, el prestigio de la universidad, el equipamiento de los laboratorios de la universidad, colegio del que proviene, influencia de profesores de colegio y academia, además un factor asociado es si las escuelas profesionales están acreditadas, y si está licenciada la universidad.

Palabras Claves: Preferencia, Regresión Logística.

ABSTRACT

The purpose of this research work is to determine the factors associated with the preferences of the students of the universities of Cusco, through multinomial logistic regression. Preferences were understood as what the person wanted to be in their adult or professional life taking into account various aspects, beyond mere interest and as a result of the comparison between different options which would involve measurement and evaluation to know if the preferences of the subjects at a given time, are concentrated or dispersed in areas and vocational groups. The study corresponds to a quantitative research approach, with descriptive scope and a non-experimental transversal design. For the study, surveys were applied to fifth-year high school students in the city of Cusco, obtaining as a conclusion that the factors associated with the election of a university in the city of Cusco are the following: Studies are carried out without interruptions by taking local, the economic income, the prestige of the university, the equipment of the laboratories of the university, school from which it comes, influence of professors of college and academy, in addition an associated an associated factor is whether the professional schools are accredited, and if the university is licensed college.

Keywords: Preference, Logistic Regression

INTRODUCCIÓN

El modelo de regresión logística multinomial o también conocido como modelo con respuesta politémica, es una generalización del modelo de regresión logístico binomial (Mc Cullagh, 1989) en el que se desea estimar la probabilidad de que el individuo presente o no un evento específico, dado un conjunto de variables que explican características particulares de los individuos. En el caso del modelo multinomial, la variable endógena tiene más de dos alternativas a considerar como posibles respuestas, por lo cual la distribución de probabilidad adecuada para modelar este fenómeno es la distribución multinomial. Se debe tener en cuenta que la regresión logística multinomial difiere de la regresión logística condicional y ordinal, pues en la regresión condicional las variables explicativas hacen referencia a atributos de las alternativas, variando sus valores para cada una de ellas, mientras que pueden o no variar para cada individuo, además solo se estima un vector de parámetros, a diferencia de el caso multinomial en el que hay tantos vectores como categorías menos una. Por otra parte en la regresión ordinal, las estimaciones no se pueden realizar con los modelos mencionados anteriormente, ya que esta tiene como particularidad el uso de la información adicional suministrado por el orden de las categorías de la variable respuesta. Ahora bien, la variable respuesta del modelo de regresión logística multinomial es una variable aleatoria con distribución multinomial, que se puede considerar como el número de éxitos en cada una de las g categorías que se presentan en n ensayos independientes

La regresión logística multinomial se conoce por una variedad de otros nombres, incluyendo regresión multiclase LR, la regresión multinomial, función SoftMax regression, Logit multinomial, clasificador de máxima entropía (MaxEnt)

Este trabajo de tesis está organizado en cuatro capítulos de la siguiente forma:

El capítulo I comprende planteamiento del problema, situación del problema, justificación de la investigación, objetivos de la investigación.

El capítulo II comprende Marco teórico conceptual, bases teóricas de regresión logística multinomial, métodos de estimación por máxima verosimilitud, antecedentes de la investigación, hipótesis y operacionalización de variables.

El capítulo III aborda la metodología, tipo y diseño de investigación, población de estudio, selección de muestra, técnicas de recolección de datos e información, confiabilidad y validez de instrumento.

En el capítulo IV se muestra resultados, descripción de la muestra, selección de variables, partición de la muestra, modelo de regresión logística multinomial y factores asociados a la elección de una universidad, criterio de ajuste del modelo global, interpretación del modelo, tasa de buena clasificación, perfil de los estudiantes según universidad pronosticada, conclusiones.

ÍNDICE

I. CAPITULO I	12
PLANTEAMIENTO DEL PROBLEMA	12
1.1. SITUACIÓN PROBLEMÁTICA.....	12
1.2. FORMULACIÓN DEL PROBLEMA	13
1.2.1. Problema general.....	13
1.2.2. Problemas específicos.....	13
1.3. JUSTIFICACION DE LA INVESTIGACION.....	14
1.4. OBJETIVOS DE LA INVESTIGACION	14
1.4.1. Objetivo general.....	14
1.4.2. Objetivos específicos.....	14
II. CAPITULO II	15
MARCO TEORICO CONCEPTUAL	15
2.1. BASES TEORICAS DE REGRESIÓN LOGÍSTICA MULTINOMIAL.....	15
2.1.1. FORMULACIÓN DEL MODELO	16
2.1.2. Interpretabilidad del modelo.....	22
2.1.3. Mas criterios a tener en cuenta sobre las variables de estudio.	24
2.2. MÉTODO DE ESTIMACIÓN. ESTIMACIÓN DE MÁXIMA VEROSIMILITUD.	26
2.3. MARCO CONCEPTUAL	33
2.4. ESTADO DE ARTE DEL TRABAJO DE INVESTIGACIÓN	34
2.5. PREFERENCIAS VOCACIONALES	36
2.6. ANTECEDENTES DE LA INVESTIGACIÓN	38
2.6.1. Antecedentes internacionales	38
2.6.2. Antecedentes nacionales	38
2.7. HIPÓTESIS.....	40
2.7.1. Hipótesis general.....	40
2.7.2. Hipótesis específicas	40
2.8. IDENTIFICACIÓN DE VARIABLES E INDICADORES.....	40
2.9. OPERACIONALIZACIÓN DE VARIABLES.....	41
III. CAPITULO III	42
METODOLOGÍA	42
3.1. NIVEL DE INVESTIGACIÓN	42
3.2. TIPO DE INVESTIGACIÓN	42
3.3. DISEÑO DE INVESTIGACIÓN	42
3.4. UNIDAD DE ANÁLISIS.....	43
3.5. POBLACIÓN DE ESTUDIO	43

3.6.	SELECCIÓN DE MUESTRA	43
3.7.	TÉCNICAS DE RECOLECCIÓN DE DATOS E INFORMACIÓN	44
3.8.	CONFIABILIDAD Y VALIDEZ DE INSTRUMENTO	44
3.8.1.	Validación de expertos.....	44
3.8.2.	Análisis de fiabilidad del instrumento.....	45
3.9.	ANÁLISIS E INTERPRETACIÓN DE LA INFORMACION.....	47
IV.	CAPITULO IV	48
	RESULTADOS	48
4.1.	DESCRIPCIÓN DE LA MUESTRA	48
4.2.	SELECCIÓN DE VARIABLES.....	50
4.2.	PARTICIÓN DE LA MUESTRA SOBRE LA VARIABLE DEPENDIENTE Y FACTORES CUALITATIVOS.....	52
4.3.	MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL. FACTORES ASOCIADOS A LA ELECCIÓN DE UNA UNIVERSIDAD.	52
4.4.	CRITERIO DE AJUSTE DEL MODELO GLOBAL.....	54
4.5.	CRITERIOS DE BONDAD DE AJUSTE DEL MODELO	54
4.6.	CALIDAD DEL AJUSTE DEL MODELO	55
4.7.	INTERPRETACIÓN DEL MODELO.....	55
4.8.	TASA DE CLASIFICACIÓN	58
4.9.	PERFIL DE LOS ESTUDIANTES SEGÚN UNIVERSIDAD PRONOSTICADA.....	59
	CONCLUSIONES	66
	RECOMENDACIONES	68
	BIBLIOGRAFIA	69
	WEBGRAFIA	70
	ANEXOS	71
	MATRIZ DE CONSISTENCIA.....	71
	INSTRUMENTO	73
	REGISTRO FOTOGRAFICO.....	74

ÍNDICE DE TABLAS

Tabla N°1: Datos con estructura de regresión logística.....	27
Tabla N°2: Datos con estructura de regresión logística multinomial.....	28
Tabla N°3: Operacionalización de los factores asociados.....	41
Tabla N°4: Cuantificación de la validación por expertos.	44
Tabla N°5: interpretación del Coeficiente de confiabilidad.....	45
Tabla N°6: Confiabilidad del instrumento.	46
Tabla N°7: Plan de análisis de datos.	47
Tabla N°8: Descripción de la muestra por sexo, tipo de colegio y provincia.....	48
Tabla N°9: Carrera profesional a la cual desean postular los estudiantes.	49
Tabla N°10: Motivos para elegir una carrera profesional.	50
Tabla N°11: Pruebas de la razón de verosimilitud.....	51
Tabla N°12: Resumen de procesamiento de casos.....	52
Tabla N°13: Selección de variables para el modelo.....	53
Tabla N°14: Criterios de ajuste del modelo.....	54
Tabla N°15: Bondad de ajuste.	54
Tabla N°16: Pseudo R cuadrado.	55
Tabla N°18: Tasa de clasificación por el modelo logístico multinomial.....	58
Tabla N°19: Perfil de los estudiantes según universidad pronosticada e ingreso económico. ...	59
Tabla N°20: Perfil de los estudiantes por prestigio de la universidad según universidad pronosticada.	60
Tabla N°21: Perfil de los estudiantes por los estudios se realizan sin interrupciones por tomas de local según universidad pronosticada.....	61
Tabla N°22: Perfil de los estudiantes por equipamiento de laboratorios de la universidad según universidad pronosticada.....	62
Tabla N°23: Perfil de los estudiantes por influencia de profesores de colegio y academia según universidad pronosticada.....	63

ÍNDICE DE FIGURAS

Figura 1: categoría de respuesta pronosticada de ingreso económico.....	60
Figura 2: Categoría de respuesta pronosticada de prestigio de la universidad	61
Figura 3: Categoría de respuesta pronosticada de los estudios que realizan sin interrupciones por tomas de local.....	62
Figura 4: Categoría de respuesta pronosticada de equipamiento de laboratorios de la universidad.....	63
Figura 5: Categoría de respuesta pronosticada de influencia de profesores de colegio y academia.....	64

CAPITULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. SITUACIÓN PROBLEMÁTICA

En estos últimos años han aparecido otras universidades como la UAC, UTEA, Continental y Alas Peruanas, que han hecho estas universidades que el mercado de elección sea muy grande, sin embargo, es importante enfatizar que aspectos o factores personales e institucionales hacen que prioricen en la decisión de una elección de una universidad que puede ser la UNSAAC, UAC, otras Univ. Para este fin es importante técnicas estadísticas que nos permitan modelar, pronosticar una de estas técnicas es Regresión Logística Multinomial. Esta técnica permite primero determinar los factores asociados o influyentes en la elección de la carrera o por otro lado permite predecir si un estudiante postula a una universidad.

Las preferencias vocacionales como las variables que influyen en las decisiones de los preadolescentes sobre su dedicación futura, que en la mayoría de los casos se manifiestan en la edad adulta mediante la realización de una profesión.

Se entendían las preferencias como lo que la persona quería ser en su vida adulta o profesional teniendo en cuenta diversos aspectos, más allá del mero interés y como resultado de la comparación entre distintas opciones (Rivas, 1998), que implicaría medición y evaluación para conocer si las preferencias de los sujetos en un momento determinado, se concentran o dispersan en áreas y grupos vocacionales.

Según Pulso Perú (2019) sobre las preferencias de los peruanos acerca de la educación en el país. Los resultados mostraron que la educación privada es la

preferida por la población, el 60% de los encuestados considera que es mejor un colegio privado que uno del Estado. Además, el 55% dijo preferir las universidades privadas.

Los estudiantes que egresan de las instituciones educativas del nivel secundario de la región Cusco tienen diferentes opciones para la formación superior, entre estas las universidades: Nacional San Antonio Abad del Cusco, Andina del Cusco, Tecnológica de los Andes y la Universidad Continental, la preferencia por una de ellas se debe a diversos factores personales e institucionales. En el presente trabajo se analizará dichos factores que predisponen su elección por una de ellas.

1.2. FORMULACIÓN DEL PROBLEMA

1.2.1. Problema general

¿Cuáles son los factores asociados a las preferencias de los estudiantes de las universidades del Cusco, obtenida mediante la regresión logística multinomial?

1.2.2. Problemas específicos

- a) ¿Los factores tangibles están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019?
- b) ¿Los factores académicos están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019?
- c) ¿La imagen institucional es un factor asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019?

1.3. JUSTIFICACIÓN DE LA INVESTIGACIÓN

El presente trabajo de investigación se justifica debido a la necesidad de identificar los factores asociados a las preferencias de los estudiantes de las universidades del Cusco; estos resultados servirán de base a las universidades para implementar acciones de marketing y políticas institucionales de mejoramiento de calidad de servicios educativos.

1.4. OBJETIVOS DE LA INVESTIGACIÓN

1.4.1. Objetivo general

Determinar los factores asociados a las preferencias de los estudiantes de las universidades del Cusco, mediante la regresión logística multinomial.

1.4.2. Objetivos específicos

- a) Describir los factores tangibles están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019.
- b) Determinar los factores académicos que están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019.
- c) Establecer si la imagen institucional es un factor asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019.

CAPITULO II

MARCO TEÓRICO CONCEPTUAL

2.1. BASES TEORICAS DE REGRESIÓN LOGÍSTICA MULTINOMIAL

MODELO LOGÍSTICO MULTINOMIAL

El objetivo principal del Modelo estadístico es conocer la relación entre una variable dependiente cualitativa con más de dos respuestas (multinomial) y una o más variables explicativas independientes (o covariables) las cuales pueden ser cualitativas o cuantitativas.

REGRESIÓN LOGÍSTICA MULTINOMIAL

La regresión logística multinomial es utilizada en modelos con variable dependiente de tipo nominal con más de dos categorías (politómica) y es una extensión multivariante de la regresión logística binaria clásica. Las variables independientes pueden ser tanto cuantitativas y/o cualitativas (categóricas o mayormente factores)

Para estos modelos se analizan eligiendo como categoría de referencia a una de las variables de respuesta y se modelan varias ecuaciones simultáneamente, una para cada una de las categorías respecto a la de referencia. (Gonzales, 2015).

2.1.1. FORMULACIÓN DEL MODELO

Para iniciar con formulación del modelo de la regresión logística se catalogan a la variable dependiente con valores como 0 y 1, por lo consiguiente se espera que como resultado la media de la variable sea representada por la proporción de casos que ocurren en una de sus dos categorías (en el caso binomial) o en una de sus múltiples categorías (en el caso multinomial). El valor predicho de la probabilidad por el modelo según la categoría puede ser interpretado como la probabilidad de que un caso caiga en esa categoría según libro de (Scott Menard, 2002).

Un modelo lineal no se ajusta adecuadamente a variables binomiales, dado que los valores predichos de la variable dependiente con este modelo, pueden tomar valores imposibles de probabilidad mayores que 1 o menores que 0, a pesar de que los valores observados estén entre 0 y 1. El mismo escenario se extiende a variables multinomiales. (Roque, 2018)

(Dueñas, s.f.) El mejor modelo que linealiza la relación entre variable dependiente categórica e independiente es el modelo logit, construido a través de regresión logística.

En una variable dependiente binomial $Y : \{0 ; 1\}$, si se conoce la probabilidad de pertenecer a una clase ($Y = 0$), se puede conocer la probabilidad de pertenecer a la otra clase ($Y=1$), es decir:

$$P(Y = 1) = [1 - P(Y = 0)] \quad (1.1)$$

(Gómez Hernández & Palacios Arias, 2013) Se puede tratar de aplicar el modelo lineal de probabilidad, expresado como:

$$P(Y = 1) = \beta_0 + \beta_1 X \quad (1.2)$$

Donde:

$P(Y = 1)$, es la probabilidad asociada a la variable predicha (Y),

X , es la variable independiente o predictora,

β , son los parámetros de la población a ser estimados.

(Dueñas, s.f.) Este modelo de probabilidad presenta el problema de no linealidad, con valores Predichos que pueden ser menores que cero o mayores que uno.

Un paso a la solución del problema a reemplazar la probabilidad [$P(y = 1)$] con el Odds ($Y = 1$), el cual se expresa como:

$$Odds(Y = 1) = \frac{P(Y = 1)}{[1 - P(Y = 1)]} \quad (1.3)$$

Como se deduce de la Ecuación 1.3, el Odds puede variar entre 0 y $+\infty$, para valores de P entre 0 y 1, por lo que valores de Odds menores que 0 generen valores imposibles de P (el valor de una variable puede ser cualquier cifra entre $-\infty$ y $+\infty$, sin embargo, en este modelo está restringida a valores entre 0 y $+\infty$). Según (Roque, 2018) para evitar lo anterior se requiere otra transformación, conocida como *logit*(Y), que se expresa:

$$logit(Y) = \ln\left(\frac{P(Y=1)}{[1-P(Y=1)]}\right) \quad (1.4)$$

El logit varía entre $-\infty$ y $+\infty$, cuando el Odds varia entre 0 y $+\infty$, para valores de la probabilidad P varía entre 0 y 1, por lo que se elimina el problema de que la probabilidad estimada puede exceder el máximo o el mínimo posible.

Así la ecuación de la relación entre la variable dependiente y las variables independientes es la siguiente:

$$\text{logit}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q \quad (1.5)$$

El logit de la variable dependiente se puede convertir a Odds mediante la ecuación:

$$\text{Odds}(Y = 1) = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q)} \quad (1.6)$$

Se puede encontrar la probabilidad $P(Y = 1)$ igualando y despejando dicha probabilidad, las ecuaciones 1.3 y 1.6, llegando a la expresión:

$$P(Y = 1) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q)}} \quad (1.7)$$

(Gómez Hernández & Palacios Arias, 2013) Cuando la variable dependiente tiene más de dos categorías, se aplica la versión multinomial de la regresión logística. Para una variable dependiente con k categorías, la regresión requiere k-1 ecuaciones logísticas, una para cada categoría, en relación a otra categoría tomada como referencia. La relación se representa por las siguientes funciones:

$$\begin{aligned} g_1(x) &= \ln \left[\frac{P(Y = 1|X)}{P(Y = k|X)} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1q}x_q \\ &= X' B'_1 \end{aligned} \quad (1.8)$$

$$\begin{aligned} g_2(x) &= \ln \left[\frac{P(Y = 2|X)}{P(Y = k|X)} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2q}x_q \\ &= X' B'_2 \end{aligned} \quad (1.9)$$

⋮

$$g_{k-1}(x) = \ln \left[\frac{P(Y = k-1|X)}{P(Y = k|X)} \right] = \beta_{(k-1)0} + \beta_{(k-1)1}x_1 + \beta_{(k-1)2}x_2 + \dots + \beta_{(k-1)q}x_q$$

$$= X' B'_{(k-1)} \quad (1.10)$$

Donde:

$g_i(x)$, es la función logit de la categoría i contra la categoría de referencia.

X , es el vector de variables independientes, $X = (x_0, x_1, \dots, x_q)'$ con $x_0 = 1$

β_{ij} , es el vector de coeficientes, estimado para la categoría i y la variable j .

Veamos la formulación de estos modelos de forma general.

Consideremos una variable de respuesta politómica Y con más de dos categorías de respuestas que denotaremos por Y_1, Y_2, \dots, Y_k .

(Ari, 2016) pretende explicar la probabilidad de cada categoría de respuesta en función de un conjunto de covariables $X = \{x_1, x_2, \dots, x_q\}$ observadas. Es decir, ajustar un modelo de la forma $\pi_j = P(Y = Y_j | X = x)$, $\forall j = 1, \dots, k$. Para cada vector x de valores observados de las variables explicativas X .

Para Osorio, Ospina, & Lenis, (2009) afirman que, para el caso de una variable de respuesta binaria, la distribución condicionada a cada combinación de valores observados de las covariables sigue una distribución de Bernoulli. Cuando la variable de respuesta es politómica, la distribución de Bernoulli se convierte en una distribución multinomial de parámetros igual a las probabilidades de cada una de las categorías de respuesta. Es decir, $(Y|X = x) \rightarrow M(n; \pi_1, \dots, \pi_k)$, siendo $\sum_{j=1}^k \pi_j = 1$.

Así que, para obtener un modelo lineal, obtendremos $\binom{k}{2}$ transformaciones logit para comparar cada par de categorías de la variable respuesta, que sería de este tipo:

$$\ln \left[\frac{\frac{\pi_i}{\pi_i + \pi_j}}{\frac{\pi_j}{\pi_i + \pi_j}} \right] = \ln \left[\frac{\pi_i}{\pi_j} \right], \quad \forall i, j = 1, \dots, k (i \neq j) \quad (1.11)$$

La ecuación 1.11 representa el logaritmo de la ventaja de respuesta Y_i frente a Y_j condicionado a las observaciones de las variables independientes que caen en uno de ambos niveles. Pero para construir el modelo logit de respuesta multinomial sería suficiente considerar $(k - 1)$ transformaciones logit básicas, definidas con respecto a una categoría de referencia o base. (Dueñas, s.f.)

Para Dueñas, (s.f.) tomando como categoría de referencia la última Y_k . Así las transformaciones logit generalizadas se definen como

$\text{Logit}_j(x) = \ln \left[\frac{\pi_j}{\pi_k} \right]$, $\forall j = 1, \dots, k - 1$, siendo $\text{logit}_j(x)$ el logaritmo de la ventaja de respuesta Y_j dado que las observaciones de las variables independientes caen en la categoría Y_j o en la Y_k .

(Roque, 2018) El modelo lineal para cada una de las transformaciones logit generalizado, para q variables explicativas, es de la siguiente forma:

$$\text{logit}_j(x) = \sum_{i=0}^q \beta_{ij} X_i = x' \beta_j, \quad \forall j = 1, \dots, k - 1 \quad (1.12)$$

Para cada vector de valores observados de las variables explicativas $x = (x_0, x_1, x_2, \dots, x_q)'$ con $x_0 = 1$ y $\beta_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{qj})'$ el vector de parámetros asociado a la categoría Y_j .

Para las probabilidades de respuesta, podemos escribir el modelo de la siguiente forma:

$$\pi_j = \frac{e^{(\sum_{i=0}^q \beta_{ij} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}}, \quad \forall j = 1, \dots, k-1 \quad (1.13)$$

Entonces para encontrar la probabilidad de respuesta para la categoría tomada como referencia π_k , se ocupa la propiedad que $\sum_{j=1}^k \pi_j = 1$.

Por lo tanto

$$\pi_k = 1 - \sum_{j=1}^{k-1} \pi_j$$

$$\pi_k = 1 - \left(\frac{e^{(\sum_{i=0}^q \beta_{i1} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} + \frac{e^{(\sum_{i=0}^q \beta_{i2} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} + \dots + \frac{e^{(\sum_{i=0}^q \beta_{i(k-1)} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} \right)$$

$$\begin{aligned} \pi_k &= \frac{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} \\ &\quad - \left(\frac{e^{(\sum_{i=0}^q \beta_{i1} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} + \frac{e^{(\sum_{i=0}^q \beta_{i2} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} + \dots \right. \\ &\quad \left. + \frac{e^{(\sum_{i=0}^q \beta_{i(k-1)} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} \right) \end{aligned}$$

$$\pi_k = \frac{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)} - e^{(\sum_{i=0}^q \beta_{i1} X_i)} - e^{(\sum_{i=0}^q \beta_{i2} X_i)} - \dots - e^{(\sum_{i=0}^q \beta_{i(k-1)} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}}$$

$$\pi_k = \frac{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)} - \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}}$$

Eliminando términos semejantes en la diferencia, se tiene que:

$$\pi_k = \frac{1}{1 + \sum_{j=1}^{k-1} e^{(\sum_{i=0}^q \beta_{ij} X_i)}} \quad (1.14)$$

2.1.2. Interpretabilidad del modelo.

(Roque, 2018) nos muestra enseguida, la interpretación de los parámetros del modelo, pero distinguiendo los casos según de qué tipo son las variables explicativas, cuantitativas o cualitativas.

➤ Una variable predictora cuantitativa X.

Para (Roque, 2018) en el modelo (ecuación 1.15) tenemos solo una única covariable cuantitativa X , el modelo para cada valor observado x de la variable X viene dada por:

$$\text{logit}_j(x) = \beta_{0j} + \beta_{1j}x, \quad \forall j = 1, \dots, k - 1 \quad (1.15)$$

Y si desarrollamos la exponencial de los parámetros β_{ij} asociados a cada categoría de la variable dependiente, que se interpreta en términos de cocientes de ventajas (odd_ratio):

$$\text{odds ratio}_j(\Delta X = 1) = \frac{\frac{\pi_j(x+1)}{\pi_k(x+1)}}{\frac{\pi_j}{\pi_k}} = \frac{e^{(\beta_{0j} + \beta_{1j}(x+1))}}{e^{(\beta_{0j} + \beta_{1j}(x))}} = e^{\beta_{1j}},$$
$$\forall j = 1, \dots, k - 1 \quad (1.16)$$

$\text{odds ratio}_j(\Delta X = 1)$, es el cociente de ventajas de respuesta Y_j frente a la última categoría Y_k cuando aumenta en una unidad la variable X . (Dueñas, s.f.)

➤ Más de una variable predictora cuantitativa.

Para el modelo logit generalizado múltiple, los cocientes de ventajas se definen incrementando una de las variables y manteniendo fijas las demás. (Gómez Hernández & Palacios Arias, 2013)

$$odds\ ratio_j(\Delta X = 1|X_s = x_s, s \neq r) = \frac{\frac{P(Y = Y_j|X_r = x_r + 1, X_s = x_s, s \neq r)}{P(Y = Y_k|X_r = x_r + 1, X_s = x_s, s \neq r)}}{\frac{P(Y = Y_j|X_r = x_r, X_s = x_s, s \neq r)}{P(Y = Y_k|X_r = x_r, X_s = x_s, s \neq r)}}$$

$$= e^{\beta_{rj}}, \quad \forall j = 1, \dots, k-1, \quad s = 1, \dots, q, \quad r = 1, \dots, q, \quad s \neq r$$

Para Gómez Hernández & Palacios Arias, (2013) siendo los $odds\ ratio_j(\Delta X = 1|X_s = x_s, s \neq r)$ el cociente de ventajas de respuesta Y_j frente a la última categoría Y_k cuando aumenta en una unidad la variable X_r y las demás se mantiene fijas.

➤ **Variables predictoras categóricas.**

Además, (Roque, 2018) nos dice si pasara de incluir en el modelo variables independientes categóricas, se introducen mediante sus variables del diseño asociados (variables dummies).

Supongamos que tenemos la variable categórica A con categorías A_1, A_2, \dots, A_p . Si de esta variable realizamos la transformación a variables de diseño mediante el método parcial que asigna un uno a la variable asociada a cada categoría y un cero al resto, y tomando como categoría de referencia la primera, obtenemos $p-1$ variables que las denotaremos como X_m^A ($m = 2, \dots, p$).

(Gómez Hernández & Palacios Arias, 2013) nos indican que el modelo de regresión logística multinomial generalizado que obtendremos seguirá siendo un modelo lineal, como en los casos anteriores, para cada logit generalizado en función de esas variables de diseño procedentes de la variable X y viene dado por:

$$logit_{ij}(x) = \ln \left[\frac{\pi_{ij}}{\pi_{ik}} \right] = \beta_{oj} + \sum_{m=2}^p \tau_{mj}^A X_{lm}^A ; \quad l = 1, \dots, p; j = 1, \dots, k-1 \quad (1.17)$$

Siendo $\pi_{ij} = p(Y = Y_j|A = A_l)$, la probabilidad de respuesta Y_j en la categoría A_l .

Dado una observación para un individuo en específico y sabiendo que solo puede pertenecer a una de las categorías de las variables de diseño X_m^A ($m = 2, \dots, p$). Se puede definir el modelo como:

$$\text{logit}_{ij}(x) = \beta_{oj} + \tau_{ij}, \text{ para cualquier } l = 1, \dots, k - 1 \quad (1.18)$$

Siendo $\tau_{ij} = 0$, $\forall j = 1, \dots, k - 1$ ya que es tomada como referencia.

Este modelo en términos de cocientes de ventajas viene dado por:

$$\text{odds ratio}_{(l1)j} = \frac{\frac{\pi_{lj}}{\pi_{lk}}}{\frac{\pi_{1j}}{\pi_{1k}}} = \frac{e^{(\beta_{oj} + \tau_{ij})}}{e^{(\beta_{oj})}} = e^{(\tau_{ij})}, \quad \forall j = 1, \dots, k - 1, \forall l = 2, \dots, p,$$

que es el cociente de respuesta Y_j frente a la última categoría Y_k para la categoría A_l de A respecto a la primera categoría A_1 .

2.1.3. Mas criterios a tener en cuenta sobre las variables de estudio.

El proceso para escoger el conjunto de variables predictoras que se incluyen en el modelo, toma algunos puntos a tener en cuenta:

(Roque, 2018) nos sugiere insertar todas aquellas variables que se consideran importante para el modelo, independientemente de si se ha demostrado o no significancia estadística en un análisis univariado previo, ya que puede conducir a dejar de incluir en el modelo covariables con una débil asociación a la variable dependiente en solitario, pero que podrían demostrar ser fuerte predictores de la misma al tomarlas en conjunto con el resto de covariables. Aunque se aconseja incluir todas las variables que aparentemente tienen relación con la variable dependiente. También es de tener presente que debemos de conseguir un modelo que sea lo más reducido posible que explique los datos (principio de parsimonia), y que además sea congruente e interpretable.

Además (Roque, 2018) nos indica que es importante que para obtener el mejor modelo, se puede recurrir a métodos de selección paso a paso, mediante inclusión hacia adelante, o por eliminación hacia atrás, o a la selección de variables por mejores subconjuntos de covariables. Estos métodos se encuentran implementados en la mayoría de los paquetes estadísticos y se describen posteriormente.

(Gómez Hernández & Palacios Arias, 2013) nos indica que debemos de tener en cuenta para elegir el número de covariables a incluir en un modelo de regresión logística es el tamaño muestral. Ya que modelos excesivamente grandes para muestras con tamaños muestrales relativamente pequeños podrían provocar errores estándar grandes o coeficientes estimados falsamente muy elevados (sobreajuste). Por lo que se suele recomendar, que por cada covariable se cuenta con un número mínimo de 10 individuos por cada categoría de la variable dependiente con menor representación.

Además (Roque, 2018) nos indica que debemos tener en cuenta de los modelos de regresión logística, es la inclusión de factores de interacción, para estudiar como la asociación de dos o más covariables puede influir en la variable dependiente. Estas interacciones pueden ser de primer orden (tomadas las covariables dos a dos o de mayor orden, pero estas últimas suelen ser de difícil interpretación). Las interacciones se incluyen siempre que sean interpretables y tengan significado desde el punto de vista de la investigación. Si en un modelo se incluye una interacción de dos o más covariables, estas deben de estar incluidas también en el modelo de forma aislada.

2.2. MÉTODO DE ESTIMACIÓN. ESTIMACIÓN DE MÁXIMA VEROSIMILITUD.

(Ari, 2016) indica que para la estimación de los coeficientes del modelo y de sus errores estándar se utiliza la estimación por máxima verosimilitud, es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente y proporcionada por los datos de la muestra. Al contrario de lo que ocurre con la estimación de los coeficientes de regresión lineal múltiple que se utiliza el método de los mínimos cuadrados, los cálculos para las estimaciones de los coeficientes de la regresión logística multinomial no son directos, hay que llevar a cabo métodos iterativos, como por ejemplo el método de Newton-Raphson.

El **método de Newton-Raphson** es un método iterativo con el que podemos encontrar aproximaciones de soluciones de ecuaciones no lineales. El método parte de un valor inicial que se introduce en una expresión relacionada con la ecuación, obteniendo así un resultado. Ese resultado se introduce en la misma expresión, obteniendo un nuevo resultado, y así sucesivamente. Si la elección del valor inicial es buena, cada vez que introducimos uno de los resultados obtenidos en esa expresión (es decir, cada vez que realizamos una *iteración* del método) el método nos proporciona una aproximación a la solución real (Rodríguez López, 2017)

Al aplicar estos métodos además de obtener las estimaciones de los coeficientes de regresión, se obtienen sus errores estándar y las covarianzas entre las covariables del modelo. A continuación, describimos el método de estimación de máxima verosimilitud para el cálculo de los coeficientes de nuestro modelo de regresión logística multinomial. (Li, Bioucas-Dias, & Plaza, 2012)

Primero veamos un ejemplo de cómo es la estructura de datos de un modelo logístico multinomial.

Tabla N°1: Datos con estructura de regresión logística

Unidad	Categoría de respuesta Y						m_i
	1			2			
i	Y_1	X_{11}	X_{12}	Y_2	X_{21}	X_{22}	
1	1	1	1	2	1	1	4
2	2	1	2	2	1	2	6
3	1	2	1	1	2	1	2
4	0	2	2	1	2	2	2

De lo anterior se puede observar que de manera general si disponemos de una muestra aleatoria de tamaño N con n combinaciones diferentes de valores de las variables explicativas X_1, X_2, \dots, X_q . Denotemos a cada combinación de valores de las variables explicativas por $x_i = (x_{i0}, x_{i1}, \dots, x_{iq})'$ con $x_{i0} = 1 \forall i = 1, \dots, n$. En cada una de estas combinaciones se tiene una muestra aleatoria de $j > 0$ observaciones independientes de la variable de respuesta politémica Y , de entre las cuales denotamos por y_{ij} al número de observaciones que caen en la categoría de respuesta $Y_j \forall j = 1, \dots, k$.

Así se verifica que $\sum_{j=1}^k y_{ij} = m_i$ y $\sum_{i=1}^n m_i = N$.

Además que; $y_{ik} = m_i - \sum_{j=1}^{k-1} y_{ij}$ y $\pi_{ik} = 1 - \sum_{j=1}^{k-1} \pi_{ij}$

Tal como se puede observar en la **Tabla N°2** siguiente:

Tabla N°2: Datos con estructura de regresión logística multinomial

	CATEGORIA									
	1				...	k-1				
<i>Unid</i>	Y_i	X_{11}	...	X_{1q}	...	Y_{k-1}	$X_{(k-1)1}$...	$X_{(k-1)q}$	m_i
1	y_{11}	X_{111}	...	X_{11q}	...	$y_{1(k-1)}$	$X_{1(k-1)1}$...	$X_{1(k-1)q}$	$\sum_{j=1}^k y_{1j} = m_1$
.				
<i>i</i>	y_{i1}	X_{i11}	...	X_{i1q}	...	$y_{i(k-1)}$	$X_{i(k-1)1}$...	$X_{i(k-1)q}$	$\sum_{j=1}^k y_{ij} = m_i$
.				
<i>n</i>	y_{n1}	X_{n11}	...	X_{n1q}	...	$y_{n(k-1)}$	$X_{n(k-1)1}$...	$X_{n(k-1)q}$	$\sum_{j=1}^k y_{nj} = m_n$

Los vectores $(y_{i1}, \dots, y_{i(k-1)})' \quad \forall i = 1, \dots, n$, siguen una distribución de probabilidad multinomial independiente $M(m_i; \pi_{i1}, \dots, \pi_{i(k-1)})$, siendo $\pi_{ij} = p(Y = Y_j | X = x_i)$ y verificando que $\sum_{j=1}^k \pi_{ij} = 1$.

Su función de densidad viene dada por:

$$f(y; \pi) = \frac{m_i!}{y_{i1}! \dots y_{ik}!} \pi_{i1}^{y_{i1}} \dots \pi_{ik}^{y_{ik}} \quad (1.19)$$

Esta misma función se puede reescribir aplicando logaritmo y exponencial, quedando de la siguiente forma.

$$f(y; \pi) = \exp \left\{ \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \pi_{i1}^{y_{i1}} \dots \pi_{ik}^{y_{ik}} \right) \right\}$$

Aplicando propiedad de logaritmos y sabiendo que:

$$y_{ik} = m_i - \sum_{j=1}^{k-1} y_{ij}$$

Se tiene:

$$f(y; \pi) = \exp \left\{ \sum_{j=1}^{k-1} y_{ij} \ln(\pi_{ij}) + \left(m_i - \sum_{j=1}^{k-1} y_{ij} \right) \ln(\pi_{ik}) + \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\}$$

$$f(y; \pi) = \exp \left\{ \sum_{j=1}^{k-1} y_{ij} \ln(\pi_{ij}) + m_i \ln(\pi_{ik}) - \sum_{j=1}^{k-1} y_{ij} \ln(\pi_{ik}) + \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\}$$

$$f(y; \pi) = \exp \left\{ \sum_{j=1}^{k-1} y_{ij} \ln \left(\frac{\pi_{ij}}{\pi_{ik}} \right) + m_i \ln(\pi_{ik}) + \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\} \quad (1.20)$$

Donde:

$$\pi_{ik} = 1 - \sum_{j=1}^{k-1} \pi_{ij} \quad e \quad y_{ik} = m_i - \sum_{j=1}^{k-1} y_{ij} \quad y \quad E[y_{ij}] = \mu_{ij} = m_i \pi_{ij}$$

$$C(y_1, \dots, y_{k-1}) = \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \right)$$

$$\text{Con } \theta_{ij} = \ln \left(\frac{\pi_{ij}}{\pi_{ik}} \right) = \beta_{0j} + \sum_{p=1}^q \beta_{pj} x_{ipj}. \quad \text{Para } j = 1, \dots, k-1 \quad (1.21)$$

Además aplicando exponencial a θ_{ij} se tiene que:

$$e^{\theta_{ij}} = \frac{\pi_{ij}}{\pi_{ik}}$$

$$\Rightarrow \pi_{ij} = \pi_{ik} e^{\theta_{ij}} \quad (1.22)$$

También, dado que $\sum_{j=1}^k \pi_{ij} = 1$, se tiene:

$$1 - \pi_{ik} = \sum_{j=1}^{k-1} \pi_{ij} \quad (1.23)$$

Por (1.22)

$$\sum_{j=1}^{k-1} \pi_{ij} = \pi_{ik} \sum_{j=1}^{k-1} e^{\theta_{ij}} \quad (1.24)$$

Por transitividad dado (1.23) y (1.24)

$$\begin{aligned} 1 - \pi_{ik} &= \pi_{ik} \sum_{j=1}^{k-1} e^{\theta_{ij}} \\ \Leftrightarrow \left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right) \pi_{ik} &= 1 \\ \Leftrightarrow \pi_{ik} &= \frac{1}{\left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right)} \quad (1.25) \end{aligned}$$

Por tanto, de (1.22) y (1.25)

$$\pi_{ij} = \frac{1}{\left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right)} e^{\theta_{ij}} \quad (1.26)$$

Siendo θ_{ij} función de parámetros $\beta^{i's}$ definido como en (1.21)

$$Y \quad b(\theta_{i1}, \dots, \theta_{ik-1}) = -m_i \ln(\pi_{ik}) == -m_i \ln \left(\frac{1}{\left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right)} \right)$$

$$b(\theta_{i1}, \dots, \theta_{ik-1}) = -m_i \ln(1) + m_i \ln \left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right) = m_i \ln \left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right)$$

Como los θ_{ij} están en función de los $\beta^{i's}$ pero los π_{ij} están en función de los θ_{ij} .

Se tiene:

$$\pi_{ij} = \pi_{ij}(\beta), i = 1, \dots, n, \quad j = 1, \dots, k - 1$$

Y la función de verosimilitud para (1.20) viene dada por

$$\begin{aligned}
L(\beta; y_1, \dots, y_n) &= \\
&= \prod_{i=1}^n \exp \left\{ \sum_{j=1}^{k-1} y_{ij} \ln \left(\frac{\pi_{ij}(\beta)}{\pi_{ik}(\beta)} \right) \right. \\
&\quad \left. + m_i \ln(\pi_{ik}(\beta)) + \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\} \quad (1.27)
\end{aligned}$$

La función de log-verosimilitud es:

$$\begin{aligned}
l &= l(\beta; y_1, \dots, y_n) \\
&= \sum_{i=1}^n \left\{ \sum_{j=1}^{k-1} y_{ij} \ln \left(\frac{\pi_{ij}(\beta)}{\pi_{ik}(\beta)} \right) + m_i \ln(\pi_{ik}(\beta)) + \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\} \quad (1.28)
\end{aligned}$$

$$\begin{aligned}
l &= l(\beta; y_1, \dots, y_n) \\
&= \sum_{i=1}^n \left\{ \sum_{j=1}^{k-1} y_{ij} (\mathbf{X}_{ij} \boldsymbol{\beta}_j) \right. \\
&\quad \left. - m_i \ln \left(1 + \sum_{r=1}^{k-1} \exp\{\mathbf{X}_{ir} \boldsymbol{\beta}_r\} \right) + \ln \left(\frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\} \quad (1.29)
\end{aligned}$$

Para obtener las ecuaciones de verosimilitud se tiene que derivar (1.28) ó (1.29) con respecto al parámetro de interés e igualar a cero, teniendo en cuenta que:

$$\theta_{ij} = \ln \left(\frac{\pi_{ij}}{\pi_{ik}} \right) = \beta_{0j} + \sum_{p=1}^q \beta_{pj} x_{ipj}$$

Entonces se deducen las ecuaciones de verosimilitud para los parámetros respectivos y definido π_{ij} como (1.26)

$$\begin{aligned}
0 = U_{jo} &= \frac{\partial l}{\partial \beta_{oj}} \sum_{i=1}^n \left(Y_{ij} - m_i \frac{\exp\{\mathbf{X}_{ij}\boldsymbol{\beta}_j\}}{1 + \sum_{r=1}^{k-1} \exp\{\mathbf{X}_{ir}\boldsymbol{\beta}_r\}} \right) \\
&= \sum_{i=1}^n (y_{ij} - m_i \pi_{ij}) \quad (1.30)
\end{aligned}$$

$$j = 1, \dots, k - 1.$$

$$\begin{aligned}
0 = U_{jp} &= \frac{\partial l}{\partial \beta_{pj}} \sum_{i=1}^n X_{ijp} \left(Y_{ij} - m_i \frac{\exp\{\mathbf{X}_{ij}\boldsymbol{\beta}_j\}}{1 + \sum_{r=1}^{k-1} \exp\{\mathbf{X}_{ir}\boldsymbol{\beta}_r\}} \right) \\
&= \sum_{i=1}^n X_{ijp} (Y_{ij} - m_i \pi_{ij}) \quad (1.31)
\end{aligned}$$

$$j = 1, \dots, k - 1, \quad p = 1, \dots, q.$$

(Li, Bioucas-Dias, & Plaza, 2012) indica que para obtener los estimadores de máxima verosimilitud hay que resolver $k - 1$ sistemas de $q + 1$ ecuaciones no lineales, que no se pueden resolver explícitamente. Es necesario acudir a métodos numéricos para obtener aproximaciones de $\boldsymbol{\beta}_j = (\beta_{oj}, \dots, \beta_{qj})$.

Los métodos numéricos se encargan de adaptar métodos matemáticos en variable continua a un sistema de representación de información discreto. En ocasiones, también se emplean métodos numéricos cuando no es posible obtener una solución analítica de un problema, o cuando el volumen de cálculos es demasiado grande para hacer un tratamiento manual. (Dueñas, s.f.)

2.3. MARCO CONCEPTUAL

REGRESIÓN

El estudio de la regresión en la parte estadística y es un proceso estadístico para estimar las relaciones entre variables. Incluye varias técnicas para el modelado y estudio de diversas variables, cuando la atención se centra en la relación entre una variable dependiente y una o más variables independientes (o predictoras). Más específicamente, el análisis de regresión ayuda a entender cómo el valor de la variable dependiente varía al cambiar el valor de una de las variables independientes, manteniendo el valor de las otras variables independientes fijas. Más comúnmente, el análisis de regresión estima la esperanza condicional de la variable dependiente dadas las variables independientes - es decir, el valor promedio de la variable dependiente cuando se fijan las variables independientes.

REGRESIÓN LOGÍSTICA

Los modelos de regresión logística son modelos estadísticos en los que la intención es determinar la relación entre una variable dependiente cualitativa, dicotómica (regresión logística binaria o binomial) o con más de dos categorías (regresión logística multinomial) y entre variables explicativas independientes, que pueden ser cualitativas o cuantitativas (Fernández San Martín, 2004).

CALIDAD DE ENSEÑANZA

Es aquello vinculado a la educación y es el proceso de enseñanza y aprendizaje que permite instruir a una persona.

2.4. ESTADO DE ARTE DEL TRABAJO DE INVESTIGACIÓN

➤ Modelo De Regresión General

$$g(y) = f(x) + \varepsilon$$

Si g es una función identidad y $f(x)$ es una función lineal

entonces $\boxed{y = \beta_0 + \sum \beta_i x_i + \varepsilon}$ M.R.L.M.

Donde: $y \sim Normal$, y : variable cuantitativa ,

x : variable cuantitativa

- Qué pasa si $y \sim binomial$, $y \rightarrow \begin{cases} 1: \text{exito (postula)} \\ 0: \text{fracaso(nopostula)} \end{cases}$

entonces g va ser una función exponencial

$$\boxed{g(y) = \frac{e^{\beta_0 + \sum \beta_i x_i + \varepsilon}}{1 + e^{\beta_0 + \sum \beta_i x_i + \varepsilon}}}$$
 M.R.L.B

$$P(y = 1)$$

Es un modelo que cuantifica la probabilidad: $P(y = 0)$

- Ahora si y tiene varias categorías $y \rightarrow \begin{cases} A: \text{UNSAAC} \\ B: \text{UAC} \\ C: \text{OTROS} \end{cases}$

entonces $y \sim Multinomial(n, P_1, P_2, \dots, P_k)$

$$\boxed{g_j(y) = \frac{e^{\beta_0 + \sum \beta_i x_i + \varepsilon}}{1 + e^{\beta_0 + \sum \beta_i x_i + \varepsilon}}}$$
 M.R.L.M

Donde: y : variable cualitativa , x_1, x_2, \dots, x_m $\begin{cases} \text{cualitativas} \\ \text{cuantitativas} \end{cases}$

Ya teniendo el modelo ahora como obtenemos los parámetros

$\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_m$ mediante el método de Máxima De Verosimilitud

- Modelo estimado:
$$\widehat{\pi}_j = \frac{e^{\widehat{\beta}_0 + \sum \widehat{\beta}_i x_i + \varepsilon}}{1 + e^{\widehat{\beta}_0 + \sum \widehat{\beta}_i x_i + \varepsilon}}$$
, que se obtiene a

partir de los datos.

Donde:

$$\begin{cases} \widehat{\pi}_1 : \text{probabilidad de que postule a la UNSAAC} \\ \widehat{\pi}_2 : \text{probabilidad de que postule a la UAC} \\ \widehat{\pi}_3 : \text{probabilidad de que postule a OTROS} \end{cases}$$

además $\widehat{\pi}_1 + \widehat{\pi}_2 + \widehat{\pi}_3 = 1$

- Entonces a partir del modelo obtenemos los valores de $\widehat{\pi}_j$:

Esto es lo que sucede

Estudiantes	x_1 x_2 x_3 \dots x_m	entonces en base a estos datos se va estimar:			\hat{y}
		$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	
1	-----	0.50	0.30	0.20	UNSAAC
2	-----	0.20	0.70	0.10	UAC
⋮		-----			⋮
⋮			⋮		⋮
n	-----				⋮

Entonces agrupamos a los que postulan a la UNSAAC, UAC, Otros y también vemos que perfil tienen, que Características son las que mas Interesan para el modelo.

2.5. PREFERENCIAS VOCACIONALES

Plantea las preferencias vocacionales como áreas ocupacionales y estereotipos profesionales y su medida por medio de actividades, habilidades, ocupaciones y autoevaluaciones y además son un constructo teórico que permite organizar los datos de una persona, y comprender en qué se diferencia de otra en su personalidad, intereses y conductas. Permite establecer la disposición o propensión a actuar de determinada manera de las personas, y “se pueden medir a través de los inventarios de intereses y son expresiones de la personalidad”. Holland (1997)

Las consideran como el interés por los perfiles profesionales y el mejor indicador del desarrollo vocacional, y las definen como el grado de gusto o atracción de la persona por las distintas profesiones. Rivas (1995)

Para Gil Beltrán (2005) las preferencias son una de las variables fundamentales en la orientación y asesoramiento vocacional y Rocabert (1987, 2006) las considera uno de los elementos básicos de pronóstico en la elección vocacional. Fernández; Peña; Viñuela, y Torio (2007) consideran las preferencias como un aspecto muy importante en los procesos de orientación escolar y de toma de decisiones académica y profesional.

Mira y López (1947) definen la elección vocacional como el proceso que consiste en elegir la carrera, profesión u oficio que mejor conviene a un individuo de acuerdo a sus aptitudes y las posibilidades que le ofrece el medio. Súper (1957) considera la elección vocacional como la concreción de las preferencias de la persona en una ocupación elegida (maxi decisión) a la que se atribuye características personales y se llega después de muchas mini decisiones, que

están influidas por distintos determinantes. Álvarez (1991) la define básicamente por el proceso por el que el individuo valora y actualiza sus preferencias, lo que conducirá a formular libremente su decisión personal en relación a sus planes y proyectos. Para Méndez (1998) la elección académica profesional es un elemento previo a la toma de decisiones profesionales a partir de una exploración de la personalidad y de las opciones profesionales y del mundo laboral, se formula una propuesta de decisión personal (elección individual) que conduce a la toma de decisiones y a la transición para la vida laboral y adulta.

Rivas (2003) define la elección vocacional como la concreción en la resolución del problema vocacional que tiene como puntos de partida de su realidad: los deseos o expectativas, es decir, la conciencia, el querer y el conocimiento de sí mismo y de su entorno

Según Elejabeitia (1995) y Fundación de Ayuda a la Drogadicción (FAD, 2007) la elección vocacional, de oficio o carrera profesional que realiza el alumnado es el resultado de un conjunto de factores y de condicionantes sociales, así como de esfuerzos, expectativas e interés personal, que concurren a lo largo de un proceso que se desarrolla en el tiempo y en los lugares en que transcurre la vida de quienes toman dicha decisión.

2.6. ANTECEDENTES DE LA INVESTIGACIÓN

2.6.1. Antecedentes internacionales

Mamani (2018) realizó el trabajo sobre factores que influyen en la elección de una universidad en los estudiantes jóvenes y adultos del Perú, cuyo objetivo fue el estudio de los factores que influyen en la elección de una universidad en los estudiantes jóvenes y adultos del Perú; donde la información contribuye a obtener datos relevantes sobre el proceso de toma de decisiones de los estudiantes que contribuirá a comprender comportamientos futuros que cada vez son más selectivos y exigentes. Asimismo, el número de postulaciones y matrículas van en crecimiento en las universidades privadas y públicas del Perú. Determino que los factores institucionales, factores sociales y factores demográficos repercuten en la elección de la universidad; los mismos que entran como una propuesta de modelo de análisis en el presente estudio.

Esta propuesta permitirá brindar herramientas a los gerentes de marketing educativo del Perú para que enfoquen sus estrategias en mejorar estos factores que son tan importantes para el estudiante y las universidades; permitiendo atraer nuevos estudiantes y recomendando a otros la universidad.

2.6.2. Antecedentes nacionales

Lozano y Chafloque (2017) desarrollaron el trabajo: Factores determinantes en la elección de la carrera de gestión de recursos humanos en una universidad privada de Perú. Cuyo objetivo fue determinar los factores determinantes en la elección de carrera de gestión de recursos humanos en una universidad privada

de Perú. El método empleado fue diseño descriptivo basado en data primaria de estudiantes de la carrera de recursos humanos, mediante la aplicación de un cuestionario bajo un formato de auto reporte. Los resultados muestran que los factores que influyeron en los estudiantes del primer al octavo ciclo en su decisión para estudiar la carrera fueron la infraestructura, ubicación geográfica y bolsa de trabajo. Mientras que en los estudiantes que cursan los últimos ciclos, la infraestructura y la malla curricular son los principales factores que influyeron en la elección de la carrera

2.7. HIPÓTESIS

2.7.1. Hipótesis general

Los factores asociados a las preferencias de los estudiantes de las universidades del Cusco, mediante la regresión logística multinomial son académicos y de imagen institucional.

2.7.2. Hipótesis específicas

- a) Los factores tangibles asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019 son la infraestructura y el equipamiento
- b) Los factores académicos asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019 son la formación del docente y la investigación
- c) La imagen institucional es un factor asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019

2.8. IDENTIFICACIÓN DE VARIABLES E INDICADORES

Variable de estudio: Factores asociados a las preferencias de los estudiantes por las universidades

2.9. OPERACIONALIZACIÓN DE VARIABLES

Tabla N°3: Operacionalización de los factores asociados

Variables	Dimensiones	Indicadores
Factores asociados a la preferencia	Académicos	-Formación del docente -Investigación -Calidad de enseñanza
	Tangibles	Equipamiento Infraestructura
	Imagen	Marca Posicionamiento
	Social	Amigos
	Económicos	Ingreso económico

Fuente: Elaborado por el tesista en base al soporte teórico

CAPITULO III

METODOLOGÍA

3.1. NIVEL DE INVESTIGACIÓN.

Es básica porque lo que se está haciendo es profundizar los conocimientos en el tema de estudio.

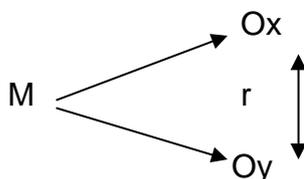
3.2. TIPO DE INVESTIGACIÓN.

Es descriptivo - correlacional por qué se va detallar cada una de los caracteres o los perfiles importantes de la variable y correlacional porque vamos a ver cómo es la relación que presentan entre las variables que hemos tomado. (Sánchez y Reyes, 2006).

3.3. DISEÑO DE INVESTIGACIÓN

Esta “investigación no experimental es la que se realiza sin manipular deliberadamente variables. Es decir, se trata de estudios donde no se varia en forma intencional las variables independientes para ver su efecto sobre otras variables” y transversal por que se registra la información en un solo momento y en un tiempo único según Hernández Sampieri, R. (2014).

El diagrama que se utilizará es:



Donde:

M = Es la muestra

O= Son las observaciones

x= Factores asociados a la preferencia del estudiante por la Universidad

y= Preferencia del estudiante por la Universidad

r= La relación entre las variables

3.4. UNIDAD DE ANÁLISIS

La unidad de análisis, estudiantes del quinto de secundario de la región Cusco

3.5. POBLACIÓN DE ESTUDIO

Todos los estudiantes que están en quinto de secundaria en la región Cusco 2019

3.6. SELECCIÓN DE MUESTRA

En el presente trabajo se ha desarrollado un muestreo probabilístico, para esto se ha elegido una población de 6000 estudiantes de todos los estudiantes que están en quinto de secundaria en la región cusco, para el tamaño de muestra se obtiene a partir de la siguiente relación:

$$n \geq \frac{Z_{(1-\alpha/2)}^2 * Npq}{(N-1)(\varepsilon)^2 + Z_{(1-\alpha/2)}^2 * pq}$$

Donde:

n: Tamaño de muestra buscado N=6000: Tamaño de la población.

p=0.50: Probabilidad de éxito q=0.50: Probabilidad de fracaso

$\varepsilon = 5\% = 0.05$: error del estudio.

Reemplazando los valores, se tiene:

$$n \geq \frac{(1,96)^2 * 6000 * 0,50 * 0,50}{(6000-1)(0,05)^2 + (1,96)^2 * 0,50 * 0,50} \geq 361.100$$

en base a la relación que se obtiene, como investigador se decidió considerar 385 estudiantes para el estudio.

3.7. TÉCNICAS DE RECOLECCIÓN DE DATOS E INFORMACIÓN

Se hará uso de la técnica de la encuesta, por considerar sus cualidades de obtener los datos con mayor confianza de las unidades de investigación y su estructura permite acceder a ella, mediante un cuestionario previamente diseñado con las preguntas que permitan obtener la información requerida, el cual el encuestador le entregará al encuestado para que lo llene, sobre la variable de interés.

El instrumento a utilizar para el recojo de datos en el presente trabajo de investigación es el cuestionario con una escala de valoración que ha sido diseñados a partir de los indicadores que provienen de las dimensiones de la investigación.

3.8. CONFIABILIDAD Y VALIDEZ DE INSTRUMENTO

3.8.1. Validación de expertos

Este instrumento para su validez será sometido al proceso de validación de expertos; y la ficha de validación del instrumento

A continuación, se describe la cuantificación de la validación por los expertos.

Tabla N°4: Cuantificación de la validación por expertos.

Experto	Validación
Dra. Karla Zelmira Aparicio Arenas	95%
Mgt. Samuel Mocrovejo Delgado	90%

FUENTE: Elaboración Propia

3.8.2. Análisis de fiabilidad del instrumento

El instrumento seleccionado será elaborado por el investigador tomando en cuenta las variables e indicadores.

En el caso del cálculo del índice o coeficiente de confiabilidad, que cuanto más cerca este a la unidad el instrumento presentará muy alta confiabilidad, pero bastará que el instrumento presente alta confiabilidad, solo que en este caso la no cercanía a la unidad es indicador que podría presentar datos heterogéneos que nos podría llevar a cometer ciertos errores, pero para evitar estos errores recurriremos a los estadísticos de prueba correspondientes para demostrar la hipótesis del investigador. Para tomar una decisión más acertada interpretaremos el valor del coeficiente de confiabilidad según la tabla siguiente.

Tabla N°5: interpretación del Coeficiente de confiabilidad.

Rangos	Magnitud
0.81 a 1.00	Muy alta
0.61 a 0.80	Alta
0.41 a 0.60	Moderada
0.21 a 0.40	Baja
0.01 a 0.20	Muy baja

Fuente: Elaboración Propia

Para calcular el valor de α , utilizaremos la fórmula:

Mediante la varianza de los ítems

$$\alpha = \frac{K}{K - 1} \left| 1 - \frac{\sum V_i}{V_t} \right|$$

α = Alfa de Cronbach

K = Número de Ítems

V_i = Varianza de cada ítem

V_t = Varianza total

Para tal efecto de la confiabilidad se someterá a la prueba respectiva mediante la varianza de los ítems a ambos instrumentos aplicados, para las variables de estudio.

Para la evaluación de la confiabilidad de los instrumentos se aplicó una prueba piloto, obteniendo los siguientes resultados de alfa de Cronbach.

Tabla N°6: Confiabilidad del instrumento.

Instrumento	Valor de Alfa de Cronbach.	Interpretación
Cuestionario de factores asociados a la preferencia de los estudiantes	0,95	Confiabilidad Muy alta

FUENTE: Elaboración propia.

De acuerdo al resultado de alfa de Cronbach de 0,95 el instrumento es alta mente confiable, por lo tanto, se puede aplicar para realizar el presente estudio.

3.9. ANÁLISIS E INTERPRETACIÓN DE LA INFORMACION.

Tabla N°7: Plan de análisis de datos.

HIPÓTESIS A SER PROBADA	Los factores asociados a las preferencias de los estudiantes de las universidades del Cusco son académicos y de imagen institucional
HIPÓTESIS NULA	Los factores académicos y de imagen institucional no están asociados a las preferencias de los estudiantes por las universidades del Cusco
HIPÓTESIS ALTERNA	Los factores asociados a las preferencias de los estudiantes de las universidades del Cusco son académicos y de imagen institucional
NIVEL DE SIGNIFICANCIA	5%
PRUEBA ESTADÍSTICA	Regresión Logística multinomial
REGLA DE DECISIÓN	Si $P < 0.05$ se acepta H_a Si $P > 0.05$ Se acepta H_0

FUENTE: Elaboración propia.

CAPITULO IV

RESULTADOS

Mediante el paquete estadístico SPSS versión 26 se ha obtenido los resultados que se muestra en las siguientes tablas, con la finalidad de determinar los factores asociados a las preferencias de los estudiantes que acabaron y estar por culminar estudios de educación secundaria sobre las universidades del Cusco, mediante la regresión logística multinomial, para ello se ha encuestado a 385 estudiantes mediante el instrumento que ha sido validado por juicio de expertos además de una validación interna en la cual se obtuvo 0.776 de alpha de Cronbach.

4.1. DESCRIPCIÓN DE LA MUESTRA

Tabla N°8: Descripción de la muestra por sexo, tipo de colegio y provincia.

		frecuencia	%
Sexo	Masculino	211	54,8%
	Femenino	174	45,2%
Colegio	Particular	218	56,6%
	Nacional	167	43,4%
Provincia	Cusco	270	70,1%
	La convención	25	6,5%
	Canchis	40	10,4%
	Anta	14	3,6%
	Urubamba	10	2,6%
	Paucartambo	9	2,3%
	Calca	2	0,5%
	Acomayo	5	1,3%
	Canas	5	1,3%
	Quispicanchis	4	1,0%
	Chumbivilcas	1	0,3%

Se puede observar que los alumnos de sexo masculino, son un poco más de la mitad (54.8%), además el 56.6% de los estudiantes es de colegio particular, y el resto son de colegios nacional, nos es mucha la diferencia. Por tanto, la

muestra está casi equitativa por sexo y tipo de colegio del cual proceden. Sin embargo, el 70.1% de los estudiantes entrevistados provienen de la provincia de Cusco, seguido de un 10.4% que proviene de Canchis.

Tabla N°9: Carrera profesional a la cual desean postular los estudiantes.

		Recuento	% de N tablas
Carrera profesional	Ing. Civil	57	14,8%
	Ing. Geológica	11	2,9%
	Ing. Electrónica	7	1,8%
	Arquitectura	26	6,8%
	Ing. Mecatrónica	2	0,5%
	Ing. Mecánica	5	1,3%
	Ing. Sistemas	9	2,3%
	Medicina	114	29,6%
	Ing. Minas	7	1,8%
	Psicología	14	3,6%
	Odontología	4	1,0%
	Ing. Biomédica	1	0,3%
	Biología	11	2,9%
	Economía	11	2,9%
	Contabilidad	6	1,6%
	Farmacia	5	1,3%
	Enfermería	1	0,3%
	Administración	13	3,4%
	Derecho	44	11,4%
	Ing. Industrial	2	0,5%
	Marketing	1	0,3%
	Estomatología	2	0,5%
	Educación	9	2,3%
	Cs. Comunicación	2	0,5%
	Historia	1	0,3%
	Filosofía	1	0,3%
	Arqueología	3	0,8%
	Ing. Eléctrica	4	1,0%
	Turismo	3	0,8%
	Antropología	1	0,3%
	Ing. Ambiental	4	1,0%
	Ing. Química	2	0,5%
Ing. Agronomía	2	0,5%	

Se observa en la tabla 9, que el 29,6% de los estudiantes entrevistados quieren postular a Medicina Humana, seguido de un 14.8% de estudiantes que desea postular a la carrera profesional de Ing. Civil. Y 11.4% de los estudiantes desea postular a la carrera de derecho.

En la tabla N° 10 se muestra los motivos por los cuales un estudiante elige una carrera profesional, y el 48.8% elige porque tiene una motivación de ser profesional en esa área, seguido de un 26.8% que elige por que la carrera tiene buena perspectiva para trabajo.

Tabla N°10: Motivos para elegir una carrera profesional.

		Recuento	% de N tablas
Motivo para elegir dicha carrera profesional	Buena perspectiva para trabajo	103	26,8%
	Influencia padres	20	5,2%
	Influencia amigos	2	0,5%
	Prestigio social carrera	41	10,6%
	Motivación Profesional	188	48,8%
	Falta de capacidad	3	0,8%
	Otros	28	7,3%

4.2. SELECCIÓN DE VARIABLES

Las variables que no aportan al modelo logístico multinomial serán identificadas, mediante pruebas de razón de verosimilitud con la estadística chi cuadrado, por tanto las variables que no aportan al modelo de regresión logística multinomial son las siguientes: “Imagen de la universidad”, “Calidad de formación de la universidad”, “Mayor posibilidad de ingreso a la universidad”, “Falta de medios económicos”, “Poca exigencia de estudios”, “Por qué tus familiares estudiaron en dicha universidad”, estas tres variables que a continuación mencionaremos , son variables que a un 10% de significación sí podrían considerarse como

variables que aportan al modelo, pero como trabajaremos a un 5% (Buena perspectiva de encontrar trabajo, Infraestructura de la universidad y Motivo para elegir dicha carrera profesional) como se muestra en la tabla N°11.

Tabla N°11: Pruebas de la razón de verosimilitud.

Efecto	Criterios de ajuste de modelo	Pruebas de la razón de verosimilitud		
	LL-2	Chi-cuadrado	gl	Sig.
Intersección	199,054 ^a	,000	0	.
Ingreso económico	226,864	27,811	2	,000
Imagen de la universidad	200,430	1,376	2	,502
Prestigio de la universidad	216,710	17,656	2	,000
Calidad de formación de la universidad	203,093	4,039	2	,133
Está acredita la universidad	212,501	13,447	2	,001
Está licenciada la universidad	210,272	11,218	2	,004
Mayor posibilidad de ingreso a la universidad	201,368	2,314	2	,314
Falta de medios económicos	199,903	,849	2	,654
Buena perspectiva de encontrar trabajo	203,779	4,725	2	,094
Poca exigencia de estudios	201,587	2,533	2	,282
Buena perspectiva salariales al egresar	205,123	6,070	2	,048
Por indicación de tus padres	207,983	8,929	2	,012
Por qué tus amigos estudian en esa universidad	205,259	6,205	2	,045
Por qué tus familiares estudiaron en dicha universidad	202,070	3,017	2	,221
Por influencia de profesores de colegio y academia	216,210	17,156	2	,000
Infraestructura de la universidad	204,278	5,224	2	,073
Equipamiento de laboratorios de la universidad	209,028	9,975	2	,007
Los estudios se realizan sin interrupciones por tomas de local	229,690	30,636	2	,000
Sexo	199,135	,082	2	,960
Colegio	212,504	13,451	2	,001
Motivo para elegir dicha carrera profesional	218,358	19,304	12	,081

Nota: Logaritmo de la verosimilitud -2 = LL-2

Observación

Todas las variables en las cuales el sig. >0.05 no serán considerados en el estudio debido a que mediante las pruebas de verosimilitud estas variables no son significativas, y su participación no sirve para discriminar entre los que postularan a una universidad u otra universidad.

4.2. PARTICIÓN DE LA MUESTRA SOBRE LA VARIABLE DEPENDIENTE Y FACTORES CUALITATIVOS

Tabla N°12: Resumen de procesamiento de casos.

		N	Porcentaje marginal
Universidad	UNSAAC	313	81,5%
	UAC	47	12,2%
	OTROS	24	6,3%
Sexo	Masculino	211	54,9%
	Femenino	173	45,1%
Colegio	Particular	217	56,5%
	Nacional	167	43,5%

De la tabla N°12. Se observa que 81.5% de los estudiantes, decidieron elegir a la UNSAAC como la universidad donde desean realizar o continuar sus estudios superiores. Seguido de un 12.2% quienes eligen a la UAC y solamente el 6.3% eligen a las otras universidades que existen. Mientras que con respecto al género el 54.9% es de sexo masculino, y 45.1% es de sexo femenino. Del total de estudiantes entrevistados el 56.5% era de colegios particulares y el restante de colegios nacionales.

4.3. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL. FACTORES ASOCIADOS A LA ELECCIÓN DE UNA UNIVERSIDAD.

Como resultado final y para alcanzar el objetivo principal de esta aplicación, se realiza un ajuste de un modelo de regresión logística multinomial, aplicando lo

explicado en el marco teórico. Se empieza estimando el modelo con las variables explicativas anteriormente mencionadas, y teniendo en cuenta que, para cada variable independiente categórica, se generan las variables indicadoras correspondientes.

Se seleccionó las variables con el método de “paso a paso hacia adelante” y con el criterio de contrastar la entrada basándose en la significación del Estadístico de puntuación y contrasta la eliminación basándose en la probabilidad del Estadístico de la razón de verosimilitud, que se basa en estimaciones de la máxima verosimilitud parcial.

Tabla N°13: Selección de variables para el modelo.

Modelo	Efecto(s)	Criterios de ajuste LL -2	Pruebas de selección de efecto		
			Chi-cuadrado	g.l.	Sig.
0	Intersección	458,510	.		
1	Los estudios se realizan sin interrupciones por tomas de local	380,846	77,664	2	,000
2	Ingreso económico	337,814	43,032	2	,000
3	Prestigio de la universidad	317,538	20,276	2	,000
4	Equipamiento de laboratorios de la universidad	298,296	19,242	2	,000
5	Colegio	279,415	18,881	2	,000
6	Por influencia de profesores de colegio y academia	268,288	11,127	2	,004
7	Está acredita la universidad	260,437	7,850	2	,020
8	Está licenciada la universidad	252,579	7,858	2	,020

Se puede observar que todas las variables son significativas y tienen presencia dentro del modelo logístico multinomial. Todos presentan un *sig.* < 0.05, por tanto las variables que se detallan en la tabla N°13 participaron en el presente modelo con la finalidad de encontrar los factores asociados a la elección de una universidad en la ciudad del Cusco.

4.4. CRITERIO DE AJUSTE DEL MODELO GLOBAL

Tabla N°14: Criterios de ajuste del modelo.

Modelo	Criterios de ajuste	Pruebas de la razón de verosimilitud		
	LL - 2	Chi-cuadrado	gl	Sig.
Sólo intersección	458,510			
Final	252,579	205,931	16	,000

Se observa que el modelo final es significativo ($\text{sig.} < 0.05$), por tanto, el modelo logístico multinomial es el adecuado para la variable nominal elección de la universidad el cual cuenta con 3 categorías nominales, con las variables predictoras ya comentadas en la tabla 13.

4.5. CRITERIOS DE BONDAD DE AJUSTE DEL MODELO

Tabla N°15: Bondad de ajuste.

	Chi-cuadrado	gl	Sig.
Pearson	1445,488	748	,000
Desvianza	252,579	748	1,000

Según Pearson se tiene una $\chi^2 = 1445,488$ con un $\text{sig} = 0,000$ por tanto el modelo es el adecuado para el conjunto de variables.

Pero según la desvianza $\chi^2 = 252,579$ con un $\text{sig} = 1,000$ no se rechaza la hipótesis, Por tanto, el modelo estimado es aceptable para estimar las probabilidades de pertenencia de las categorías de la variable dependiente.

4.6. CALIDAD DEL AJUSTE DEL MODELO

Para medir la calidad del ajuste del modelo se utiliza como se vio en el Capítulo I, los coeficientes **Pseudo-R2 de Mc-Fadden, de Cox-Snell y de Nagelkerke**. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo final y del modelo inicial con solo la constante.

Tabla N°16: Pseudo R cuadrado.

Cox y Snell	,415
Nagelkerke	,596
McFadden	,449

Para Cox y Snell el 41.5% de la variabilidad de elección de la universidad de preferencia es explicada por las variables consideradas. Mientras que para Nagelkerke es de 59.6% y para Mc Fadden es de 44.9%.

4.7. INTERPRETACIÓN DEL MODELO

Cuando interpretamos los Odds_ratios de cada variable, se asume que el resto de variables independientes se mantienen fijas. Interpretamos cada una de las variables independientes entre las distintas preferencias de universidades tomando como referencia a “Otra Universidad”.

El odds_ratio cambia cuando la i-ésima variable explicativa regresora se incrementa en una unidad, si:

$\beta_i > 0$ significa que el odds_ratio se incrementa

$\beta_i < 0$ significa que el odds_ratio decrece

$\beta_i = 0$ significa que el factor es igual a uno, lo cual hace que el odds_ratio no varía. En la columna $\exp(\beta)$ de la tabla 17 se observa que:

Tabla N°17: Estimaciones de parámetro

Universidad		B	Desv.	Wald	gl	Sig.	Exp(B)	IC. 95% Exp(B)	
			Error					LI	LS
UNSAAC	Intersección	8,772	1,902	21,274	1	,000			
	[Colegio =1]	-2,347	,643	13,329	1	,000	,096	,027	,337
	[Colegio =2]	0 ^b	.	.	0
	Ingreso económico	,000	,000	17,359	1	,000	1,000	1,000	1,000
	Prestigio de la universidad	1,552	,338	21,121	1	,000	4,723	2,436	9,156
	Está acredita la universidad	-,861	,364	5,597	1	,018	,423	,207	,863
	Está licenciada la universidad	,211	,348	,367	1	,545	1,234	,624	2,440
	Por influencia de profesores de colegio y academia	,479	,187	6,541	1	,011	1,615	1,119	2,332
	Equipamiento de laboratorios de la universidad	-1,191	,299	15,878	1	,000	,304	,169	,546
	Los estudios se realizan sin interrupciones por tomas de local	-1,098	,255	18,506	1	,000	,333	,202	,550
OTROS	Intersección	-,745	2,646	,079	1	,778			
	[Colegio =1]	-1,846	,740	6,215	1	,013	,158	,037	,674
	[Colegio =2]	0 ^b	.	.	0
	Ingreso económico	,000	,000	,756	1	,385	1,000	1,000	1,000
	Prestigio de la universidad	1,644	,509	10,423	1	,001	5,174	1,908	14,036
	Está acredita la universidad	,123	,505	,060	1	,807	1,131	,421	3,042
	Está licenciada la universidad	-,689	,382	3,255	1	,071	,502	,237	1,061
	Por influencia de profesores de colegio y academia	,108	,236	,211	1	,646	1,114	,702	1,768
	Equipamiento de laboratorios de la universidad	-,818	,349	5,505	1	,019	,441	,223	,874
Los estudios se realizan sin interrupciones por tomas de local	,035	,328	,011	1	,916	1,035	,545	1,968	

a. La categoría de referencia es: UAC. b. Este parámetro está establecido en cero porque es redundante.

“Universidad Nacional San Antonio Abad del Cusco con referencia a la Universidad Andina del Cusco”

En la tabla 17 se presentan las estimaciones de los parámetros del modelo para la “Universidad Nacional San Antonio Abad del Cusco”, existe 4.7 veces mayor probabilidad de elegir a la UNSAAC que a la universidad Andina del Cusco cuando se mencione sobre el prestigio de la Universidad; además, existe 1,234 veces mayor probabilidad de elegir la UNSAAC con respecto a la UAC cuando se consulta si la universidad esta licenciada.

Los alumnos tienen 1.615 veces mayor probabilidad de elegir a la UNSAAC, que, a la UAC con respecto a la influencia de profesores de colegio y academia, es decir cuánto más el docente motive a los estudiantes a postular a una universidad, mayor será la probabilidad de que el alumno elija a la UNSAAC como casa de estudios superiores.

Cuando al estudiante se le pregunta por seguir estudios superiores y que estos estudios se realicen sin interrupciones por tomas de local ellos prefieren con 0.667 veces mayor probabilidad a la UAC que a la UNSAAC, además el alumno está convencido que la UAC tiene mejor equipamiento de laboratorios y por ello el alumno elegiría a la UAC con 0.696 veces mayor probabilidad de elegir a la UAC que a la UNSAAC.

Un alumno que es de colegio particular tiene 0.904 más veces mayor probabilidad de elegir a la UAC que a la UNSAAC.

“Otras universidades del Cusco con referencia a la Universidad Andina del Cusco”

En la tabla 17 se presentan las estimaciones de los parámetros del modelo para “otras universidades del cusco”, existe 5.174 veces mayor probabilidad de elegir a otra Universidad que a la universidad Andina del Cusco cuando se mencione

sobre el prestigio de la Universidad; además, existe 0.50 veces mayor probabilidad de elegir otra universidad que a la UAC cuando se consulta si la universidad esta licenciada.

Los alumnos tienen 1.114 veces mayor probabilidad de elegir a la UNSAAC, que, a la UAC con respecto a la influencia de profesores de colegio y academia, es decir cuánto más el docente motive a los estudiantes a postular a una universidad, mayor será la probabilidad de que el alumno elija otra universidad antes que a la UAC.

Cuando al estudiante se le pregunta por seguir estudios superiores y que estos estudios se realicen sin interrupciones por tomas de local ellos prefieren con 1.035 veces mayor probabilidad otra universidad que a la UAC.

4.8. TASA DE CLASIFICACIÓN

Tabla N°18: Tasa de clasificación por el modelo logístico multinomial

Observado	Pronosticado			Porcentaje correcto
	UNSAAC	UAC	OTROS	
UNSAAC	304	5	4	97,1%
UAC	17	30	0	63,8%
OTROS	13	7	4	16,7%
Porcentaje global	87,0%	10,9%	2,1%	88,0%

Del total de estudiantes que elige a la UNSAAC bajo el modelo de regresión logística multinomial el 97.15 % de estos estudiantes fueron clasificados como UNSAAC, mientras que 63.8 % de estudiantes que eligen UAC han sido bien clasificados, y 16.7% de los que eligen otras universidades han sido también clasificados de esta manera. En general el modelo clasifico un 88% de casos de manera correcta, está de más decir que el modelo clasifica de mejor manera a estudiantes que eligen a la UNSAAC para continuar estudios superiores.

4.9. PERFIL DE LOS ESTUDIANTES SEGÚN UNIVERSIDAD PRONOSTICADA

Tabla N°19: Perfil de los estudiantes según universidad pronosticada e ingreso económico.

	Universidad	UNSAAC	UAC	OTROS	
Ingreso económico	Media	3427,95	7268,89	15750,00	
	95% de intervalo de confianza para la media	Límite inferior	3180,60	5586,22	3073,08
		Límite superior	3675,30	8951,56	28426,92
	Media recortada al 5%	3191,27	6644,44	14555,56	
	Mediana	3000,00	6000,00	11000,00	
	Varianza	5,249,166,185	31,369,010,101	229,928,571,429	
	Desv. Desviación	2,291,106	5,600,804	15,163,396	
	Mínimo	600	1000	3000	
	Máximo	16000	30000	50000	
	Rango	15400	29000	47000	
	Rango intercuartil	2075	5750	14750	
	Asimetría	2,012	2,142	1,961	
	Curtosis	6,247	5,817	4,354	

Los alumnos que eligen la UNSAAC tienen padres con ingreso económico promedio de 3427.95, mientras que los alumnos que eligen la UAC tienen ingresos económicos promedio de 7268.89, y los alumnos que eligen otras universidades tienen ingresos económicos altos. En general personas con ingresos bajos eligen a la UNSAAC, y alumnos de recursos medios-altos eligen a la UAC; además aquellos alumnos que tienen ingresos más altos eligen otras universidades distintas a las ya mencionadas anteriormente. Todo lo descrito se puede visualizar en la siguiente figura.

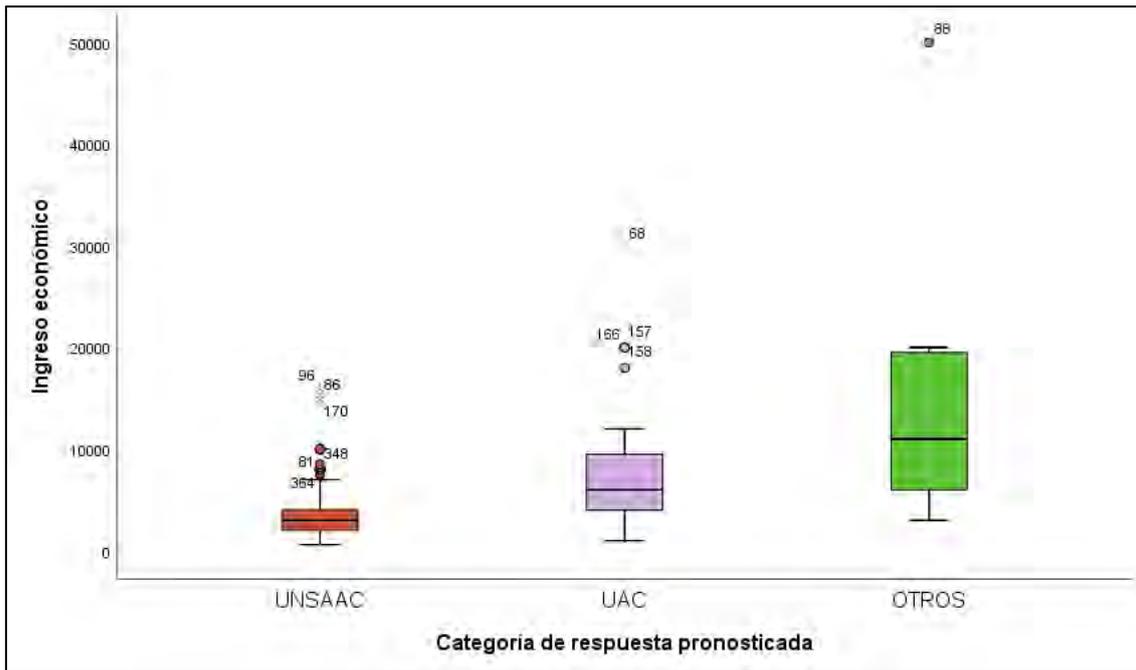


Figura 1: categoría de respuesta pronosticada de ingreso económico

Tabla N°20: Perfil de los estudiantes por prestigio de la universidad según universidad pronosticada.

		Universidad	UNSAAC	UAC	OTROS
Prestigio de la universidad	Media		4,39	3,87	4,38
	95% de intervalo de confianza para la media	Límite inferior	4,30	3,58	3,61
		Límite superior	4,47	4,16	5,14
	Media recortada al 5%		4,48	3,96	4,42
	Mediana		5,00	4,00	5,00
	Varianza		,642	,936	,839
	Desv. Desviación		,802	,968	,916
	Mínimo		1	1	3
	Máximo		5	5	5
	Rango		4	4	2
	Rango intercuartil		1	0	2
	Asimetría		-1,659	-1,297	-,999
	Curtosis		3,549	2,109	-1,039

Los alumnos eligen a la UNSAAC por que creen que tienen altos niveles de prestigio, mientras que los que eligen a la UAC creen que el nivel de prestigio de esta universidad es media, y los que eligen a otras universidades piensan que

esas universidades tienen prestigio de medio a alto. Todo lo descrito se puede visualizar en la siguiente figura.

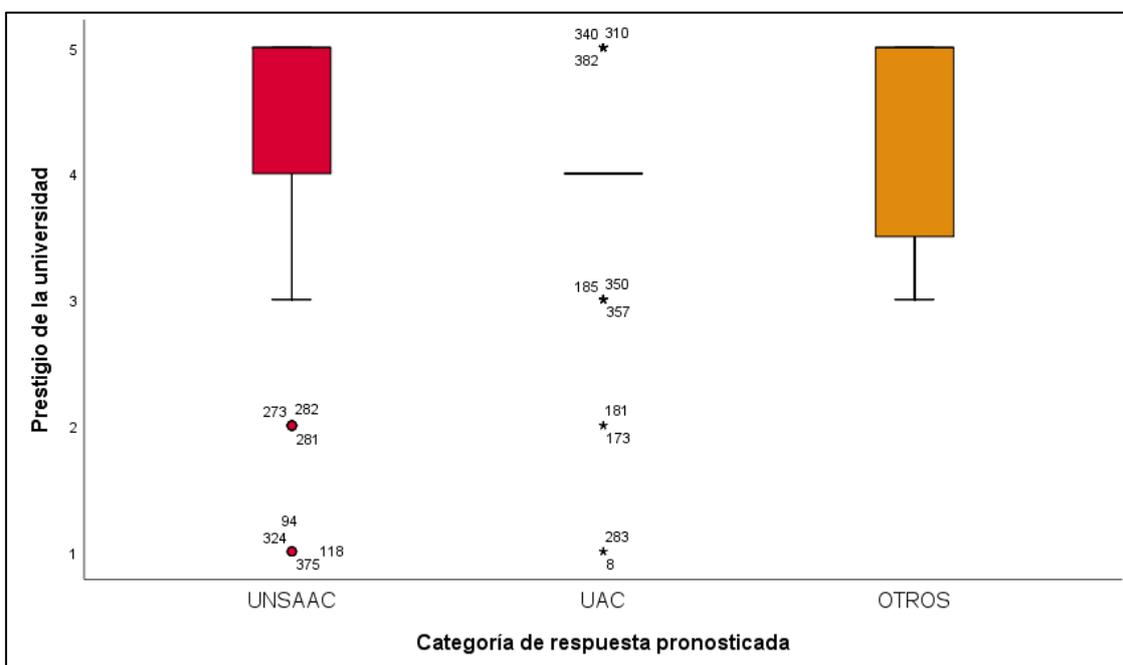


Figura 2: Categoría de respuesta pronosticada de prestigio de la universidad

Tabla N°21: Perfil de los estudiantes por los estudios se realizan sin interrupciones por tomas de local según universidad pronosticada.

Universidad		UNSAAC	UAC	OTROS	
Los estudios se realizan sin interrupciones por tomas de local	Media	2,86	4,51	3,63	
	95% de intervalo de confianza para la media	Límite inferior	2,73	4,30	2,74
		Límite superior	2,98	4,72	4,51
	Media recortada al 5%	2,84	4,59	3,64	
	Mediana	3,00	5,00	3,50	
	Varianza	1,361	,483	1,125	
	Desv. Desviación	1,166	,695	1,061	
	Mínimo	1	2	2	
	Máximo	5	5	5	
	Rango	4	3	3	
	Rango intercuartil	2	1	2	
	Asimetría	,106	-1,530	,045	
Curtosis	-,771	2,679	-,940		

Los alumnos creen que una universidad donde se puede realizar sus estudios sin interrupciones es la UAC, seguido de los alumnos que eligen a otras

universidades en donde podrían realizar sus estudios sin interrupciones, y los que eligen a la UNSAAC saben que sus estudios se verán interrumpidos por tomas de local.

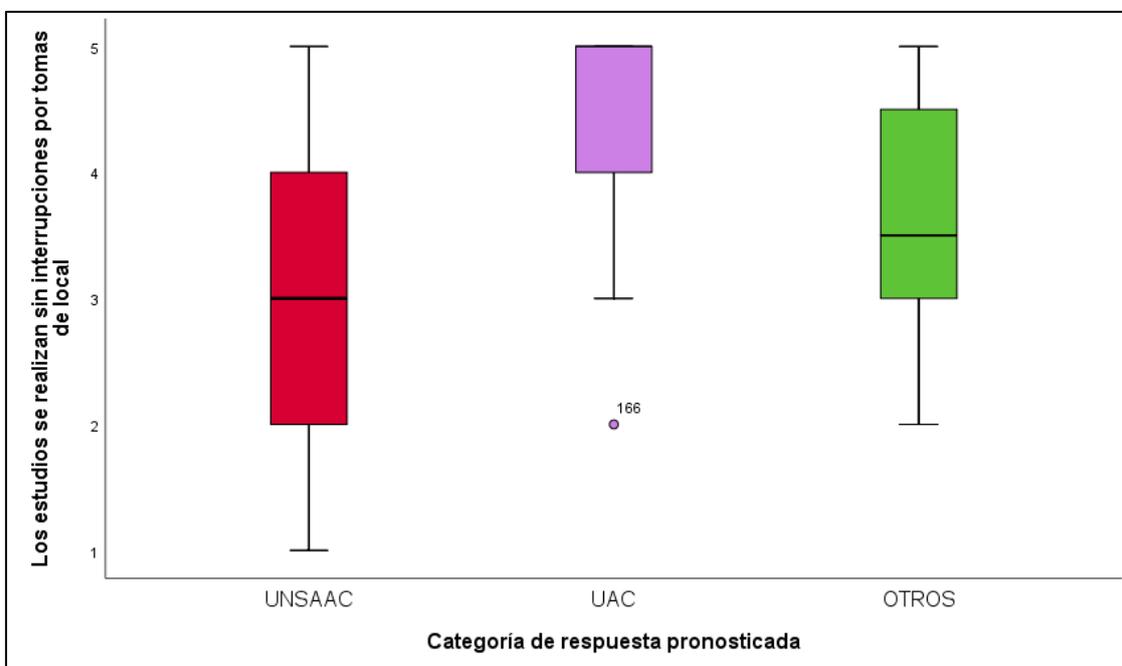


Figura 3: Categoría de respuesta pronosticada de los estudios que realizan sin interrupciones por tomas de local

Tabla N°22: Perfil de los estudiantes por equipamiento de laboratorios de la universidad según universidad pronosticada.

Universidad		UNSAAC	UAC	OTROS	
Equipamiento de laboratorios de la universidad	Media	3,52	4,64	3,75	
	95% de intervalo de confianza para la media	Límite inferior	3,41	4,47	2,59
		Límite superior	3,64	4,82	4,91
	Media recortada al 5%	3,58	4,71	3,83	
	Mediana	4,00	5,00	4,00	
	Varianza	1,102	,325	1,929	
	Desv. Desviación	1,050	,570	1,389	
	Mínimo	1	3	1	
	Máximo	5	5	5	
	Rango	4	2	4	
	Rango intercuartil	1	1	2	
	Asimetría	-,624	-1,374	-1,120	
	Curtosis	-,101	1,017	1,106	

Los alumnos que eligen la UAC perciben que el equipamiento de laboratorios es mejor que el de la UNSAAC y en cambio su pensamiento de los alumnos que eligen otras universidades son más dispersos en su percepción sobre sus laboratorios.

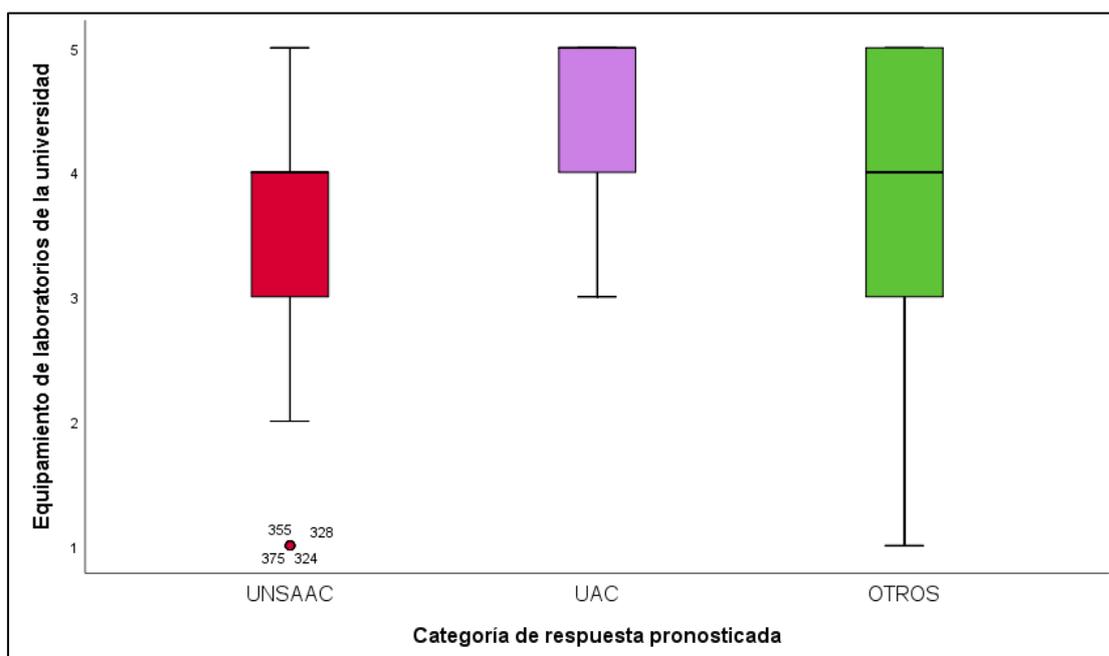


Figura 4: Categoría de respuesta pronosticada de equipamiento de laboratorios de la universidad

Tabla N°23: Perfil de los estudiantes por influencia de profesores de colegio y academia según universidad pronosticada.

Universidad		UNSAAC	UAC	OTROS	
Por influencia de profesores de colegio y academia	Media	2,96	2,22	1,75	
	95% de intervalo de confianza para la media	Límite inferior	2,83	1,90	,88
		Límite superior	3,10	2,55	2,62
	Media recortada al 5%	2,96	2,17	1,67	
	Mediana	3,00	2,00	1,50	
	Varianza	1,521	1,177	1,071	
	Desv. Desviación	1,233	1,085	1,035	
	Mínimo	1	1	1	
	Máximo	5	5	4	
	Rango	4	4	3	
	Rango intercuartil	2	2	1	
	Asimetría	-,116	,428	1,675	
	Curtosis	-1,021	-,618	3,136	

Cuanto los docentes más motiven a los estudiantes a postular a una universidad, los alumnos elegirán a la UNSAAC para continuar estudios superiores, seguido de la UAC y por último otras universidades.

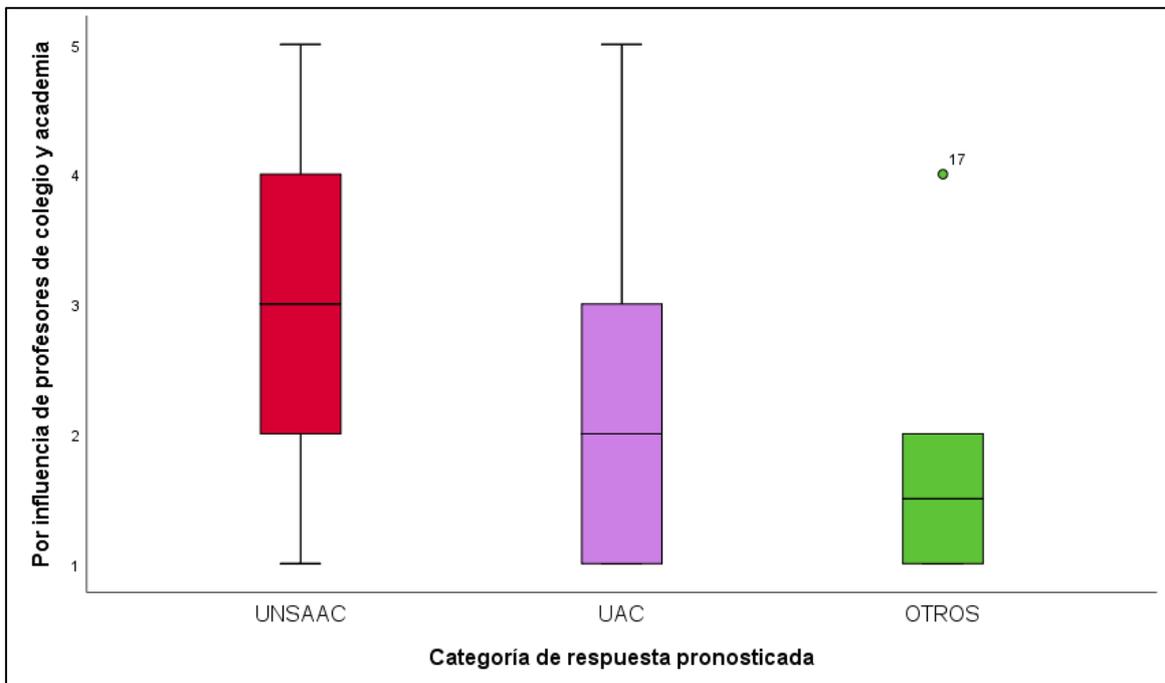


Figura 5: Categoría de respuesta pronosticada de influencia de profesores de colegio y academia

DISCUSIÓN

Hallazgos

Los estudios se realizan sin interrupciones por tomas de local, el ingreso económico, el prestigio de la universidad, el equipamiento de los laboratorios de la universidad, colegio del que proviene, influencia de profesores de colegio y academia, acreditación y licenciamiento de la universidad son factores asociados a la elección de la Universidad.

Comparación con la Literatura.

Mamani (2018); Determino que los factores institucionales, factores sociales y factores demográficos repercuten en la elección de la universidad; los mismos que entran como una propuesta de modelo de análisis en el presente estudio. Estos factores son similares a los obtenidos en nuestro estudio, debido a que también se establece el ingreso económico y los problemas sociales como factor de elección de la Universidad.

En un estudio realizado anteriormente por lozano, haciendo una encuesta en la universidad lozano y Chafloque (2017) muestran que los factores que influyeron en los estudiantes del primer al octavo ciclo en su decisión para estudiar la carrera fueron la infraestructura, ubicación geográfica y bolsa de trabajo. Mientras que en los estudiantes que cursan los últimos ciclos, la infraestructura y la malla curricular son los principales factores que influyeron en la elección de la carrera. Este estudio coincide con algunos factores considerados en nuestro estudio como infraestructura, ingreso económico.

CONCLUSIONES

1. De acuerdo a lo que se analizó en este trabajo de investigación se llega a la siguiente conclusión que los factores asociados a la elección de una universidad en el Cusco, son las siguientes: Los estudios se realizan sin interrupciones por tomas de local, el ingreso económico, el prestigio de la universidad, el equipamiento de los laboratorios de la universidad, colegio del que proviene, influencia de profesores de colegio y academia, además un factor asociado es si está acreditada la universidad, y si está licenciada la universidad.
2. De acuerdo a los resultados que se obtuvo en el trabajo de investigación, Cuando al estudiante se le pregunta por seguir estudios superiores y que estos estudios se realicen sin interrupciones por tomas de local ellos prefieren con 0.677 veces mayor probabilidad a la UAC que a la UNSAAC, además el alumno está convencido que la UAC tiene mejor equipamiento de laboratorios y por ello el alumno elegiría a la UAC con 0.696 veces mayor probabilidad de elegir a la UAC que a la UNSAAC; Un alumno que es de colegio particular tiene 0.904 más probabilidad de elegir a la UAC que a la UNSAAC.
3. De acuerdo a los resultados que se obtuvo en el trabajo de investigación, un alumno tiene 4.7 veces mayor probabilidad de elegir a la UNSAAC que a la universidad Andina del Cusco cuando se mencione sobre el prestigio de la Universidad; además, 1.234 veces mayor probabilidad de elegir la UNSAAC con respecto a la UAC cuando se consulta si la universidad esta licenciada.

4. De acuerdo a los resultados que se obtuvo en el trabajo de investigación los alumnos que eligen a las universidades es por la formación académica y por las investigaciones que realizan además en general el modelo logístico multinomial clasifico un 88% de casos de manera correcta, además el modelo clasifica de mejor manera a estudiantes que eligen a la UNSAAC para continuar estudios superiores

RECOMENDACIONES

[1] A las autoridades universitarias de la ciudad del cusco recomendarles utilizar Modelos Estadísticos de minería de datos para diagnosticar los factores determinantes de elección de escuela profesional de los estudiantes de nivel secundario.

[2] Al Rector de la UNSAAC implementar adecuadamente los laboratorios y mejorar la infraestructura afín de generar mayores niveles de posicionamiento en los estudiantes de nivel secundario de la ciudad del cusco.

[3] A los docentes de la UNSAAC generar nuevas alternativas de reclamo que no perjudiquen el normal desarrollo de las clases a fin de mejorar la imagen bien ganada de la UNSAAC.

BIBLIOGRAFÍA

Ari, Erkan. (2016). Using Multinomial Logistic Regression to Examine the Relationship Between Children's Work Status and Demographic Characteristics

Andersen. E.B.(1990). Introduction to the statistical. analysis of Categorical data. New York, Estados Unidos.

Alvarez, Rafael (1995). Estadística multivariante y no paramétrica con SPSS. Aplicación a las ciencias de la salud. Diaz de Santos S.A Madrid, España.

Abraira, V.(1996). Métodos multivariados en bioestadística. Ed. Centro de estudios Ramón Areces. Madrid, España.

Dueñas, M. (2015). Modelos de respuesta discreta en R y aplicación con datos reales.

Hosmer, DW. Lemeshow,S.(1989). Applied logistic Regression. John Wiley & Sons. New York, U.S.A.

Hosmer, DW. Bofin AM. (2008) Multinomial goodness-of-fit tests for logistic regresión models. New York.

Pando Fernández, V., & Fernández San Martín, R. (2004). Regresión logística multinomial, 000, 323–327.

Silva, A.(1990). Excursión a la regresión logística en Ciencias de la salud. Madrid, España.

WEBGRAFÍA

Ari, Erkan. (2016). Using Multinomial Logistic Regression.

https://www.researchgate.net/publication/311908200_Using_Multinomial_Logistic_Regression_to_Examine_the_Relationship_Between_Children's_Work_Status_and_Demographic_Characteristics

Gómez Hernández, Palacios Arias, Darwin Ernesto (2013) Modelación logística multinomial.

<https://ri.ues.edu.sv/id/eprint/9807>

León y Gambará, 2005; Rubio, 2006 Las preferencias profesionales y vocacionales del alumnado de secundaria.

<https://digibug.ugr.es/bitstream/handle/10481/5543/18751362.pdf?sequence=1>

Li, Bioucas-Dias, A Plaza, 2012

https://scholar.google.com/citations?view_op=list_works&hl=es&hl=es&user=nhSyPiEAAAAJ

Pulso Perú (2019) sobre las preferencias de los peruanos acerca de la educación en el país

<https://pqs.pe/actualidad/economia/peruanos-escoje-educacion-privada-sobre-publica/>

Roque M.J., 2018. "Modelos De Regresión Logística Multinomial De La Calidad.

http://repositorio.unap.edu.pe/bitstream/handle/UNAP/7755/Roque_Cruz_Maria_Juaquina.pdf?sequence=1&isAllowed=y

Scott Menard, 2002. Applied Logistic Regression Analysis, Volumen 106; Volumen 2002

https://books.google.com.pe/books?id=EAI1QmUUsbUC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

ANEXOS

MATRIZ DE CONSISTENCIA

TITULO: APLICACIÓN DE REGRESION LOGISTICA MULTINOMIAL EN EL ESTUDIO DE PREFERENCIAS DE LOS ESTUDIANTES DE LAS UNIVERSIDADES DEL CUSCO, 2019

<p>PROBLEMA GENERAL</p> <p>Cuáles son los factores asociados a las preferencias de los estudiantes de las universidades del Cusco, obtenida mediante la regresión logística multinomial</p> <p>PROBLEMAS ESPECÍFICOS</p> <p>a) ¿Los factores tangibles están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019?</p> <p>b) ¿Los factores académicos están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019?</p> <p>c) ¿La imagen institucional es un factor asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019?</p>	<p>OBJETIVO GENERAL</p> <p>Determinar los factores asociados a las preferencias de los estudiantes de las universidades del Cusco, mediante la regresión logística multinomial</p> <p>OBJETIVOS ESPECIFICOS</p> <p>a) Describir los factores tangibles están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019</p> <p>b) Determinar los factores académicos que están asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019</p> <p>c) Establecer si la imagen institucional es un factor asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019</p>	<p>HIPÓTESIS GENERAL</p> <p>Los factores asociados a las preferencias de los estudiantes de las universidades del Cusco, mediante la regresión logística multinomial son académicos y de imagen institucional</p> <p>HIPÓTESIS ESPECÍFICOS</p> <p>a) Los factores tangibles asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019 son la infraestructura y el equipamiento</p> <p>b) Los factores académicos asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019 son la formación del docente y la investigación</p> <p>c) La imagen institucional es un factor asociados a las preferencias de los estudiantes de las universidades del Cusco, 2019.</p>	<p>VARIABLE INDEPENDIENTE:</p> <p>Factores asociados a las preferencias de los estudiantes.</p> <p>VARIABLE DEPENDIENTE:</p> <p>Universidades</p>
---	---	---	---

JUSTIFICACIÓN	METODOLOGÍA	UNIVERSO POBLACIÓN Y MUESTRA
<p>El presente trabajo de investigación se justifica debido a la necesidad de identificar los factores asociados a las preferencias de los estudiantes de las universidades del Cusco; estos resultados servirán de base a las universidades para implementar acciones de marketing y políticas institucionales de mejoramiento de calidad de servicio educativo.</p>	<p>El presente estudio está enmarcado en el tipo de investigación descriptivo - correlacional por qué se va detallar cada una de los caracteres o los perfiles importantes de la variable y correlacional porque vamos a ver cómo es la relación que presentan entre las variables que hemos tomado.</p> <p>Diseño de investigación es no experimental transversal.</p>	<p>POBLACIÓN: Población finita</p> <p>MUESTRA: 385 estudiantes</p> <p>MÉTODO ESTADÍSTICO</p> <p>Se utilizó Regresión logística multinomial</p>

INSTRUMENTO

ENCUESTA DE PREFERENCIA POR LA UNIVERSIDAD

1. Sexo: Masculino () Femenino ()
2. Colegio donde realizó sus estudios : Particular () Nacional ()
3. Provincia donde realizó sus estudios.....
4. Ingreso económico familiar en soles.....
5. En qué Universidad Realizara sus estudios superiores. Marque Con X
 - a) Universidad Nacional San Antonio Abad del Cusco (UNSAAC) ()
 - b) Universidad Andina del Cusco (UAC) ()
 - c) Otros (). Especifique.....
6. De los siguientes motivos de elección. Marque su nivel de acuerdo con respecto a la universidad seleccionada.

Nro.	Motivo de elección de la Universidad	Totalmente En desacuerdo	En desacuerdo	Indiferente	De acuerdo	Totalmente de acuerdo
1	Imagen de la Universidad					
2	Prestigio de la Universidad					
3	Calidad de formación de la Universidad					
4	Está acreditada la universidad					
5	Esta licenciada la universidad					
6	Mayor posibilidad de ingreso a la Universidad					
7	Falta de medios económicos					
8	Poca exigencia de estudios					
9	Buena perspectiva salariales al egresar					
10	Por indicación de tus padres					
11	Por qué tus amigos estudian en esa universidad					
12	Porqué tus familiares estudiaron en dicha universidad					
13	Por influencia de profesores de colegio y academia					
14	Infraestructura de la universidad					
15	Equipamiento de laboratorios de la Universidad					
16	Los estudios se realizan sin interrupciones por tomas de local					

7. ¿Qué carrera profesional estudiaras?
.....
- 8.Cuál es el motivo para elegir dicha carrera profesional. Marque con X
 - a) Buena perspectiva de encontrar trabajo ()
 - b) Influencia de los padres ()
 - c) Influencia de los amigos ()
 - d) Prestigio social de la carrera ()
 - e) Motivación profesional ()
 - f) Falta de capacidades para cursar otra carrera ()
 - g) Otros () Especifique

REGISTRO FOTOGRAFICO



