

Modelo de regresión logística para el análisis del riesgo agrícola en la zona Sud del Departamento de Potosí

Location of the Sustainable Development Goals in the Municipality of Potosí: A Strategy for the Future

Delmy Jeanneth Chirinos Cortez

pt.delmy.chirinos.c@upds.net.bo

<https://orcid.org/0000-0003-1128-606X>

Universidad Privada Domingo Savio.

Potosí – Bolivia

Recibido el 12 de agosto de 2021 / Arbitrado el 02 de septiembre de 2021 / Aceptado el 15 de noviembre de 2021 / Publicado el 01 de enero de 2022

RESUMEN

La regresión logística es un modelo estadístico utilizado para predecir el resultado de una variable categórica. El objetivo de este artículo es relacionar una serie de aspectos importantes para la aplicación de un modelo estadístico a la investigación de la gestión del riesgo agrícola. Se orientó en la investigación cuantitativa y el paradigma positivista tipo proyectiva, diseño no experimental, de campo. Utilizando como población la provincia Sud Chichas, Nor Chichas y Modesto Omiste. Se emplearon encuestas, análisis de riesgos. Se utilizaron la tabulación de encuestas mediante el SPSS. Los resultados fueron que las tres provincias sometidas a estudio cuentan con aproximadamente la misma ponderación en cuanto a su número de habitantes y se hace notar que no se toma en cuenta los datos del municipio de Atocha. Se concluye que, la variable dependiente refiere a 0 como la ausencia de la característica que interesa al estudio o llamado también control y se denota con 1 a la característica presente, vale decir el riesgo agrícola.

Palabras clave: Modelo; riesgo; regresión: logística; agrícola.

ABSTRACT

Logistic regression is a statistical model used to predict the outcome of a categorical variable. The objective of this article is to list a series of important aspects for the application of a statistical model to research in agricultural risk management. It was guided by quantitative research and the positivist paradigm. Using the provinces of Sud Chichas, Nor Chichas and Modesto Omiste as population. Research and risk analysis were used. Survey tabulation was used using SPSS. The results were that the three provinces under study have approximately the same weight in terms of number of inhabitants and it is observed that the data from the municipality of Atocha are not taken into account. It is concluded that the dependent variable refers to 0 as the absence of the characteristic that is of interest in the study or also called control and 1 denotes the present characteristic that is, agricultural risk.

Keywords: Model; risk; Logistic regression; agricultural.

INTRODUCCIÓN

El modelo de regresión logística para Ouorou et al., (2021), es una herramienta estadística ampliamente utilizada en diversos campos, incluido el análisis del riesgo agrícola. En el contexto agrícola, este modelo se emplea para predecir y evaluar el riesgo asociado con diferentes variables que pueden afectar la producción y el rendimiento de los cultivos. Es importante acotar que, el riesgo agrícola es una preocupación constante para los productores, ya que están expuestos a una variedad de factores que pueden impactar negativamente en sus cosechas, como condiciones climáticas adversas, plagas, enfermedades y fluctuaciones en los precios de los productos.

Por otra parte, el uso del modelo de regresión logística en el análisis del riesgo agrícola según Llaugel y Fernández (2021), permite a los agricultores y expertos identificar y cuantificar la probabilidad de ocurrencia de eventos adversos, así como comprender mejor las relaciones entre las variables que influyen en dichos eventos. En el contexto del análisis del riesgo agrícola, el modelo de regresión logística se utiliza para predecir la probabilidad de que ocurra un evento específico, como la presencia de una plaga o una enfermedad en un cultivo determinado. Al incorporar variables relevantes, como condiciones climáticas, prácticas agronómicas y características del suelo, el modelo puede proporcionar información valiosa sobre los factores que contribuyen al riesgo y ayudar a los agricultores a tomar medidas preventivas o correctivas.

Dentro de este orden de ideas, la aplicación del modelo de regresión logística en el análisis del riesgo agrícola en opinión de Camacho (2022), ofrece varias ventajas significativas. Por un lado, permite una evaluación cuantitativa y objetiva del riesgo, lo que facilita la toma de decisiones informadas y la implementación de estrategias de gestión del riesgo más efectivas. Además, al identificar las variables clave que influyen en el riesgo agrícola, el modelo puede ayudar a optimizar los recursos y mejorar la resiliencia de las operaciones agrícolas frente a posibles amenazas.

En conclusión, el modelo de regresión logística es una herramienta poderosa y versátil para el análisis del riesgo agrícola. Su capacidad para predecir la probabilidad de eventos adversos e identificar factores de riesgo clave lo convierte en un recurso invaluable para los agricultores y profesionales del sector agrícola. Al integrar este modelo en sus procesos de toma de decisiones, los actores del sector pueden mejorar la gestión del riesgo, aumentar la productividad y garantizar la sostenibilidad a largo plazo de sus operaciones agrícolas.

En habidas cuentas, existen diversos métodos estadísticos que pueden ser empleados para analizar la información recolectada en materia de políticas públicas; entre esos métodos científicos, están los que pertenecen al análisis multivariado. El propósito del análisis multivariado, es lograr que grandes cantidades de datos complicados y complejos sean analizados simultáneamente; es decir, estos métodos tienen como finalidad el de resumir grandes cantidades de datos en pocas variables que se consideran las más representativas. Estas técnicas son empleadas en diversos campos de investigación; tanto del sector público como privado de un país, tocando los diversos ámbitos del saber humano.

En esta perspectiva, estos métodos estadísticos, tienen diversas modalidades de uso y García et al., (2020), han dicho que, dependiendo principalmente del tipo de datos que se desea analizar y tipo de estudio que se desea plantear. En ese sentido, en muchas investigaciones existe la necesidad de efectuar procedimientos con métodos de dependencia para el análisis e interpretación de datos cualitativos y cuantitativos, tomando en cuenta las características propias de la información. En estos casos, el método multivariado de regresión logística puede ser de gran ayuda para realizar este tipo de análisis.

Por otro lado, Camacho (2022), manifiesta que, la regresión logística es un método multivariante que

permite estimar la relación existente entre una variable dependiente no métrica en particular dicotómica y un conjunto de variables independientes métricas pero dicotomizadas y por supuesto las variables no métricas y tiene como finalidad modelar la probabilidad de que una unidad experimental caiga en un grupo particular, con base en la información medida en la propia unidad, mientras que para Llaugel y Fernández (2021), es apropiada cuando la variable dependiente es categórica, es decir, nominal o no métrica y las variables independientes son cuantitativas o cualitativas. En otros casos, la variable dependiente consta de dos o más grupos o clasificaciones, por ejemplo: sucedió, no sucedió, éxito o fracaso y una clasificación de tres grupos puede comprender clasificaciones como: alta, media y baja.

Se considera la variable dicotómica de recurrencia (se presentó una denuncia de desastre agrícola más de una vez en una comunidad) y ocurrencia (se presentó una denuncia de desastre agrícola en la comunidad del departamento de Potosí). Dicha variable, por su naturaleza misma forma parte de los indicadores indispensables para construir el Plan Nacional de alerta temprana, en base a la ocurrencia y recurrencia de factores climáticos, factores sociodemográficos y productivos de gran interés y a la vez complejidad para los gobiernos en el mundo.

Se consideró trabajar con datos de la Zona Sud del departamento de Potosí, contando para ello con registros administrativos oficiales de 4 de las estaciones meteorológicas y datos Poblacionales y de agrícolas de los últimos censos nacionales tanto de población y vivienda como del censo agropecuario, esta información permitió un análisis representativo.

Este estudio tiene por propósito analizar la probabilidad el riesgo agrícola en la Zona Sud del departamento de Potosí, para ello se ha considerado información desde el año 2013 al 2017 de manera anual, el estudio sirve para la toma de decisiones y planificación de las acciones propias del Plan Nacional de alerta temprana.

Considérese que el Sistema de Alerta Temprana “Es el conjunto de capacidades necesarias para generar y difundir información de alerta que sea oportuna y significativa, con el fin de permitir que las personas, las comunidades y las organizaciones amenazadas se preparen y actúen de forma apropiada y con suficiente tiempo de anticipación para reducir la posibilidad de que se produzcan pérdidas o daños” (EIRD, 2009 pág. 34).

El presente artículo tiene como objetivo relacionar una serie de aspectos importantes para la aplicación de un modelo estadístico a la investigación de la gestión del riesgo agrícola. Aunque el análisis multivariante tiene sus raíces en la estadística univariante y bivariante, la extensión al dominio multivariante introduce conceptos y cuestiones adicionales, que van desde el valor teórico, hasta las escalas de medida utilizadas, los errores de medición, los resultados estadísticos de las pruebas de significación y los intervalos de confianza.

Es importante acotar que, la utilización de un modelo multivariante conlleva la elaboración de un plan de investigación bien definido que incluye los objetivos analíticos en términos conceptuales, la selección del método, la evaluación de los supuestos básicos de dicho método y sus técnicas, la estimación del modelo y su interpretación, para finalizar con la aplicación de las técnicas de validación para determinar la estabilidad de los resultados obtenidos.

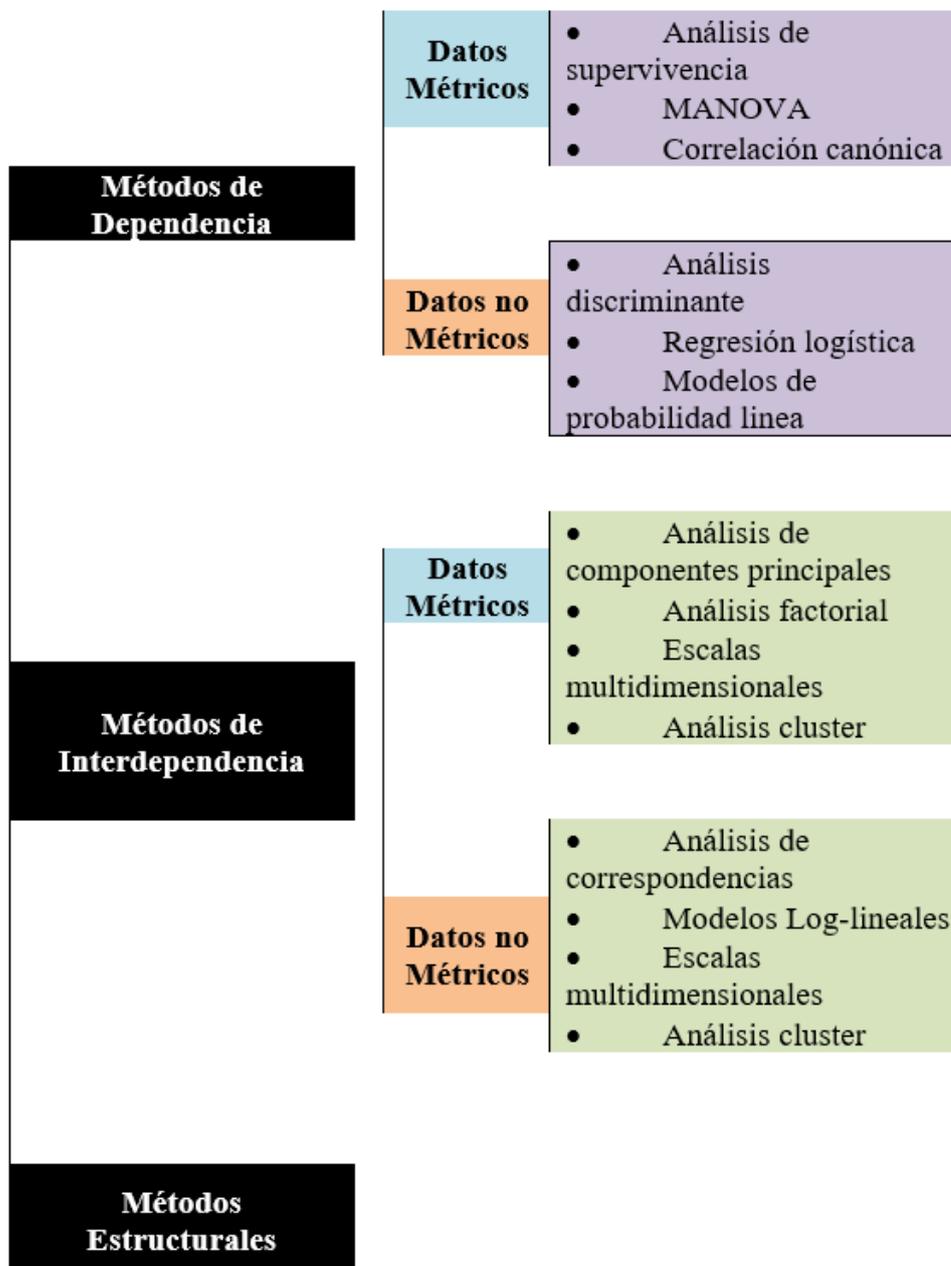
En este mismo orden, el análisis simultáneo de múltiples características de dichos individuos, objetos o entes es la característica clave del análisis multivariante, ello porque dicha simultaneidad en el análisis permite analizar la interrelación existente entre todas las variables aleatorias que representan dichas características. De lo anterior se deduce que el carácter multivariante de los métodos englobados bajo la denominación de análisis multivariante, no reside tanto en la multiplicidad de variables consideradas en

el estudio, como en la consideración conjunta e interrelacionada de las mismas. Ahora bien, el objetivo del análisis multivariante es permitir a quien lo aplique tomar decisiones óptimas en el contexto en que se encuentre, teniendo en cuenta la información que aporta el conjunto de datos analizado.

Clasificación de los métodos multivariante

El análisis multivariante según Barreiro y Castro (2023), es mediante la descripción de los principales procedimientos que engloba a los métodos dependientes, interdependientes y estructurales, estos autores presentan la clasificación mediante la figura 1:

Figura 1. Clasificación de los métodos multivariante



Fuente: Barreiro y Castro (2023)

Métodos de dependencia:

Análisis de regresión: Supóngase que en el problema objeto de estudio se tiene una variable dependiente (Y) y una colección de variables independientes que se suponen explicativas de la anterior bajo una forma lineal o linealizable (X_1, X_2, \dots, X_n)

Si tanto la variable a explicar cómo las variables “explicativas” son cuantitativas y los datos relativos a las últimas conforman una matriz de rango completo, la técnica que, entre otras cosas, proporciona la relación lineal de Y con X_1, X_2, \dots, X_n , o en otros términos, que permite predecir los cambios en el valor de en respuesta a los cambios en los valores de X_1, X_2, \dots, X_n , o, lo que es lo mismo, la “explicación” del comportamiento de la variable de nuestro interés mediante la información suministrada por una serie de variables de las que se supone depende linealmente, se denomina regresión múltiple. El modelo de regresión múltiple viene dado por $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + e_i$ o, en términos matriciales, $Y = X\beta + e$, donde X es una matriz de cantidades conocidas y de rango completo y la inclusión del término de error se justifica por la omisión en el modelo de variables explicativas relevantes o errores de medida.

Métodos de independencia:

Análisis de componentes principales ACP: el análisis de componentes principales examina las relaciones entre un conjunto de p variables correlacionadas y las transforma en un nuevo conjunto de variables incorreladas denominadas componentes principales. Estas nuevas variables son combinaciones lineales de las originales y se derivan en orden de importancia, de tal manera que la primera componente principal recoge, de la variación total de los datos originales, la mayor parte posible. Y así sucesivamente. Esta técnica es originaria de K. Pearson (1901) y fue desarrollada posteriormente por Hotelling (1933, 1936). Su objetivo fundamental es ver si unas pocas componentes recogen la mayor parte de la variación de los datos originales. Si es así, se puede argüir que la dimensionalidad del problema no es p sino inferior a p. En la práctica no siempre es fácil la identificación de las componentes principales por lo que su principal uso recae en la reducción de la dimensionalidad de los datos para simplificar posteriores análisis. Por ejemplo, es una manera muy útil de encontrar agrupaciones en los datos cuando estos vienen caracterizados por un elevado número de variables.

Métodos estructurales:

Los modelos de ecuaciones estructurales permiten describir, gráfica y analíticamente, las relaciones que se cree que existen entre las variables observables y las no observables, teniendo en cuenta la dirección de cada una de tales relaciones. A partir de la información muestral, se pueden estimar tales relaciones y juzgar su importancia, de tal manera que se puede simplificar el diagrama inicial representativo de las posibles relaciones hasta obtener un modelo parsimonioso.

Un aspecto clave en la estimación de cualquier tipo de modelo estadístico es su sujeción al planteamiento de una teoría debidamente asentada en el área de conocimiento en que se esté trabajando. Este requisito es especialmente importante en un área de modelización tan flexible como ésta. Por ello, se debe prestar una gran atención a la especificación e identificación del modelo. El primer aspecto se refiere al correcto planteamiento del sistema de ecuaciones en función de la teoría subyacente (cumplimiento de supuestos básicos, definición de algunos parámetros como fijos y otros como libres o estimables); el segundo aspecto tiene que ver con que la cantidad de información disponible sea suficiente para tener una estimación única de los parámetros libres, más de una o ninguna.

Entre las bases teóricas en este artículo, se encuentra el análisis de regresión logística, según Camacho (2022), es una herramienta valiosa para comprender y predecir eventos categóricos, como el riesgo agrícola, al proporcionar información sobre las relaciones entre variables y estimar probabilidades de ocurrencia. Su aplicación permite tomar decisiones informadas y gestionar eficazmente situaciones donde la variable dependiente es discreta o categórica.

Así mismo, el análisis de regresión logística presenta unas características específicas que según Álvarez et al, (2021), son: el modelo estadístico específico, se utiliza cuando la variable dependiente es categórica, como la presencia o ausencia de un evento. Igualmente, la probabilidad como resultado, aquí se estima la probabilidad de que ocurra un evento en función de las variables independientes. También una de las características es la función logística, esta emplea la función logística para transformar la salida en una probabilidad entre 0 y 1 y por último los coeficientes logarítmicos, es decir, los coeficientes de regresión representan el cambio en el logaritmo de las probabilidades.

Por su parte, Llaugel y Fernández (2021), manifiesta que los objetivos del análisis de regresión logística, están compuesto por la predicción de probabilidades, el cual se encarga de determinar la probabilidad de ocurrencia de un evento en función de las variables productoras, otro objetivo es la identificación de factores significativos, donde se identifica qué variables independientes influyen significativamente en la variable dependiente. Así también habla de la evaluación del riesgo, es decir del análisis y cuantificación del riesgo asociado con diferentes variables en un contexto específico, como el riesgo agrícola. Y finalmente, la clasificación binaria, adecuado para problemas de clasificación binaria con dos posibles resultados.

El análisis de regresión logística tiene la misma estrategia del análisis de regresión múltiple, el cual se diferencia esencialmente del análisis de regresión logística porque la variable independiente es métrica; en la práctica el uso de ambas técnicas tiene mucha semejanza, aunque sus enfoques matemáticos sean diferentes. La variable dependiente o respuesta no es continua si no discreta, generalmente toma valores de 1, 0. Las variables explicativas pueden ser cuantitativas o cualitativas y la ecuación del modelo no es una función lineal de partida, si os

La regresión logística como técnica multivariada exploratoria y predictiva

La regresión logística como método estadístico, es apropiado cuando la variable dependiente es categórica y las variables independientes son o no métricas, ha sido reconocida su capacidad y poderío en el análisis de datos en diversas áreas científica como, la medicina, la agroecología, la economía, las ciencias sociales, etc. permite la predicción de pertenencia de la unidad de análisis a uno de los grupos pre establecidos.

Regresión logística simple

Este modelo tiene la forma de:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{para } i = 1, 2, \dots, n \quad (1.1)$$

De esto se deduce que:

$$\text{Si } y = 1, \quad \varepsilon_i = 1 - \beta_0 - \beta_1 X_i \quad (1.2)$$

$$\text{Si } y = 0, \quad \varepsilon_i = -\beta_0 - \beta_1 X_i \quad (1.3)$$

Por tanto ε_i no puede tener distribución normal debido a que toma valores discretos, el modelo de regresión lineal simple, no es aplicable para el caso de variables respuesta de tipo dicotómico. En el análisis de regresión lineal simple, el punto principal del proceso de estimaciones del modelo es un gráfico de dispersión de la variable respuesta versus la regresora, pero este gráfico resulta limitado cuando hay solo dos valores posibles para la variable respuesta por tanto se debe usar otros gráficos, estos resultan de la suavización de los valores de la variable respuesta, representando después los valores de la variable respuesta versus la regresora

$$\text{Sea: } \pi(x) = E(y/x) \quad (1.4)$$

Que presenta la media condicional de $y=1$ dado x donde $\pi(x)$, 0 y 1 dado a x , representa la probabilidad de que ocurra $y=1$, ciertamente tenga relación lineal dentro el rango de la variable regresora. De la relación $\pi(x)$ y x para valores intermedios de x . se espera una relación curvilínea. Para cualquier valor grande de x , $\pi(x)$ tomará valores cercanos a 1 y para valores cercanos de x , $\pi(x)$ tomara valores cercanos a cero. Curva en forma de S o signoide que tiene las propiedades requeridas para $\pi(x)$ y que tiene las propiedades de una función de distribución de probabilidades acumuladas.

Para esta probabilidad se usa la función de distribución acumulada de la distribución logística dado por:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1.5)$$

Tiene, cuando $\beta_0 < 0$ y cuando $\beta_1 > 0$ además este modelo toma valores en el intervalo $[0, 1]$. Cuando $P[Y = 1] = 0.5$ el valor de x es: $-\frac{\beta_0}{\beta_1}$ que brinda información muy útil. Una transformación de $\pi(x)$ que es lo central del estudio de la regresión logística es la transformación Logit, transformación que se define en términos $\pi(x)$ como:

$$g(x) = \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x \quad (1.6)$$

Lo importante de esta transformación es que tiene muchas propiedades semejantes al modelo de la regresión lineal simple, por ejemplo, son lineales sus parámetros, puede ser continua y puede tomar cualquier valor real dependiente de x . Para el modelo de la regresión lineal simple, la variable respuesta se expresa como:

$$y = E(y|x) + \varepsilon \quad (1.7)$$

Para la variable respuesta dicotómica la expresamos como:

$$y = \pi(x) + \varepsilon \quad (1.8)$$

Entonces, que ocurre con este modelo.

$$\begin{aligned} \mathcal{S} \quad y = 1, \quad \varepsilon_i = 1 - \pi(x) \quad & \text{y tiene probabilidad } \pi(x) \\ \mathcal{S} \quad y = 0, \quad \varepsilon_i = -\pi(x) \quad & \text{y tiene probabilidad } 1 - \pi(x) \end{aligned}$$

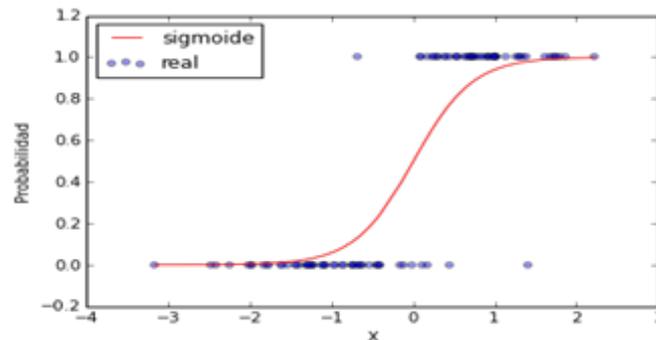
Entonces ε_i tiene una distribución binomial con media cero y varianza $\pi(x)[1 - \pi(x)]$ Por tanto la distribución condicional de la variable respuesta tiene distribución de probabilidades binomial con media $\pi(x)$ El lado izquierdo se llama también logaritmo de ODDS ratio o razón de ventaja a favor de éxito o

también razón de probabilidad de $y = 1$ contra $y = 0$ específicamente.

$$ODDS\ RATIO = \frac{\pi(X)}{1 - \pi(X)} \quad (1.9)$$

Se denota en el modelo logístico como (ODDS) a la ocurrencia de eventos, esta razón se define como el cociente de la probabilidad de que el evento ocurra a la probabilidad de que el evento no ocurra.

Gráfico 1. Gráfica de la regresión logística



En general, la gráfica de una función de distribución es, una función de distribución es una función real de variable real: (Rojo, J. M, Pág. 5)

$F: \mathbb{R} \rightarrow \mathbb{R}$

De forma que verifica:

Está acotada en el intervalo $[0,1]$

$0 \leq F(x) \leq 1 \forall x$

Es monótona no decreciente:

$1 \leq x \leq x \Rightarrow () () 1 \leq F x \leq F x$

Y, además, está definida en todo \mathbb{R} , tomando los siguientes valores:

$F(-\infty) = 0$

$F(+\infty) = 1$

Riesgo

Es la probabilidad de ocurrencia de uno o varios eventos adversos que pudiesen ocasionar lesiones o pérdida de vidas, afectación a propiedades, a la producción de alimentos y el medio ambiente, y la detención de la actividad económica en un lugar y periodo de exposición determinado. Por otro lado, para Andrade y Arteaga (2021), Los riesgos son eventos o situaciones que pueden tener consecuencias negativas para las personas, empresas y sociedad en general. Estos pueden ser causados por diversas razones como errores humanos, desastres naturales, accidentes de tránsito entre otros.

Riesgo agrícola

La Organización para la Cooperación y el Desarrollo Económicos (OCDE), 2023 y el Departamento de Agricultura de Estados Unidos (USDA) advierten sobre los riesgos en la agricultura en sus respectivas páginas web. La preocupación está justificada. Como apuntan desde la oficina norteamericana, “el riesgo

es un aspecto importante del negocio de la agricultura”, pues incertidumbres referentes a las condiciones meteorológicas, los rendimientos, los precios, las políticas gubernamentales, los mercados globales y otros factores pueden provocar fallas en los ingresos agrícolas. Para el USDA, “la gestión del riesgo consiste en elegir entre alternativas que reducen los efectos financieros que pueden derivarse de tales incertidumbres”. Organización para la Cooperación y el Desarrollo Económicos (OCDE) y el Departamento de Agricultura de Estados Unidos (USDA), (Website: www.ceigram.upm.es, consultado 2019). Así, desde el Departamento norteamericano distinguen cinco tipos generales de riesgo: riesgo de producción, riesgo de precio o mercado, riesgo financiero, riesgo institucional y riesgo personal o humano. El que compete al estudio es el riesgo de producción, que tiene que ver con la incertidumbre que provoca el crecimiento de los cultivos y el ganado. Dentro de este tipo de riesgo, factores como el tiempo, las enfermedades o las plagas pueden provocar mermas en la calidad y la cantidad de las cosechas.

MÉTODO

Este artículo tiene como objetivo relacionar una serie de aspectos importantes para aplicar un modelo estadístico en la investigación de la gestión del riesgo agrícola. Para ello, se adoptó un enfoque cuantitativo y se propuso un modelo estadístico basado en el enfoque proyectivo descrito por Hurtado (2006).

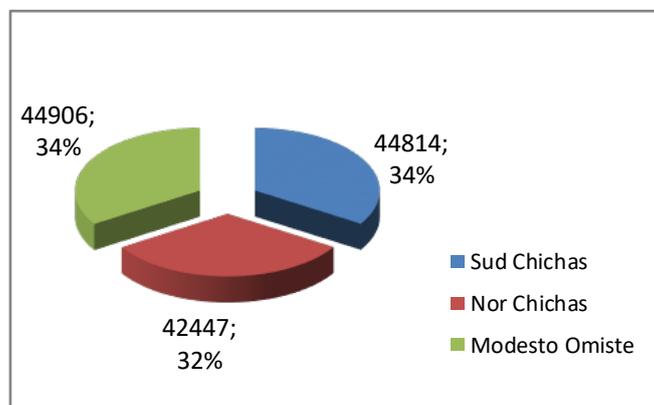
El diseño de la investigación se clasificó como no experimental, ya que no se manipularon deliberadamente las variables y se observaron los fenómenos en su entorno natural para su posterior análisis, siguiendo la definición de Hernández et al. (2014). La población seleccionada para el estudio fue la Zona Sud del Departamento de Potosí, Bolivia, específicamente las provincias Sud Chichas, Nor Chichas y Modesto Omiste.

Con el fin de alcanzar el objetivo establecido, se llevaron a cabo encuestas y análisis de riesgos. Las encuestas se tabularon utilizando el software SPSS. Además, se realizó una confrontación de factores y una calificación ponderada, así como un análisis de maquinarias.

RESULTADOS

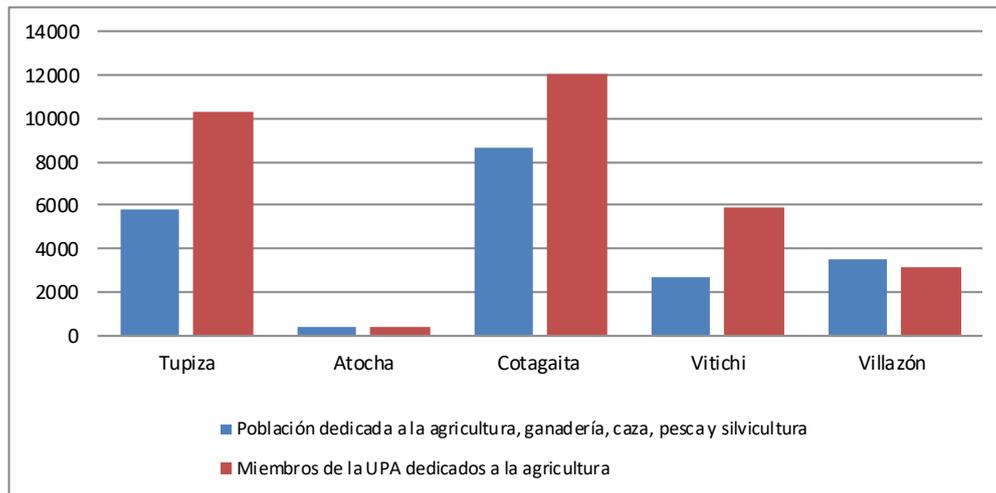
A continuación, en el gráfico 2, inicia la descripción de los resultados de la investigación:

Gráfico 2. Habitantes por provincia, zona sud Potosí



El gráfico señala que las tres provincias sometidas a estudio cuentan con aproximadamente la misma ponderación en cuanto a su número de habitantes y se hace notar que no se toma en cuenta los datos del municipio de Atocha correspondiente a la provincia Sud Chichas, por no contar con datos registrados tanto climáticos como de reportes de desastres agrícolas en los años comprendidos entre el 2013 y 2017. Por otro lado, se evidencia su poca incidencia en el sector agrícola.

Gráfico 3. Población dedicada a la agropecuaria y miembros de la UPA dedicada a la agricultura



El gráfico muestra que el municipio de Cotagaita involucra poblacionalmente y en número de miembros de cada Unidad productiva Agropecuaria a una mayor cantidad de personas, le sigue Tupiza con población cercana a las 6.000 personas, le sigue Vitichi con más de 5.000 personas dedicadas a la actividad agropecuaria y con miembros dentro de la UPA que superan las 2.000, Villazón no supera a los 4.000 habitantes dedicados a la agropecuaria y sus UPA están cercanas a las 3.000. Finalmente, destaca el insipiente aporte a la agricultura por parte del municipio de Atocha, municipio que queda fuera del alcance del presente estudio.

Del gráfico se puede determinar que la provincia Nor Chichas compuesta por los municipios de Cotagaita y Vitichi cuenta con mayor población dedicada a la agropecuaria al igual que el número de sus unidades productivas dedicadas exclusivamente a la agricultura. Por el contrario, Sud Chichas, presenta un caso extremo que señala una escasa población e unidades productivas dedicadas a la agricultura, Modesto Omiste por sus características topográficas y climáticas muestra una actividad disminuida en cuanto a población e unidades productivas dedicadas a la agricultura. La tabla 1 corresponde a los Eventos adversos en Bolivia:

Tabla 1. Bolivia: Eventos adversos de origen natural reportados, según tipo de evento (En número de reportes)

Descripción	20002 - 2013	%
Total	13.00.3	100%
inundación	5.444	42%
Sequía	2.212	17%
Helada	2.249	17%
Granizada	2.211	17%
Deslizamiento, Mazamorra	278	2%
Viento Huracanado	246	2%
Incendio	321	2%
Sismo	25	0%
Plaga	15	0%
Tormenta Eléctrica	1	0%
Contaminación	1	0%

Fuente: Instituto Nacional de Estadística (2013)

Se encontró que, los eventos más frecuentes en Bolivia son las inundaciones que representan el 42% de los registrados entre el 2002 y el 2013, le sigue con un 17% las sequías, con un 17% también están presentes las heladas, que son más propias de la zona andina del país, el restante 6% corresponde a los deslizamientos, vientos huracanados, incendios. Esto se conoció en base a los reportes del entonces Viceministerio de Defensa Civil – INE que las pérdidas por el fenómeno El Niño en los años 1983 – 1984 representó el 15,3% del PIB.

El Niño en los años 2015-2016 ha afectado a 72.000 familias y 30.000 Ha. de diferentes cultivos. Las inundaciones en los años 2013 y 2014 causaron pérdidas por más de \$US 111 millones. Esta información, nos muestra la incidencia económica que tiene la agricultura en la evolución del PIB, siendo por tanto imperativo atender la gestión de riesgo agrícola.

El pronóstico del tiempo (determinístico y limitado en su horizonte temporal) se separa conceptualmente del pronóstico del clima (necesariamente probabilístico). Entendemos por pronóstico del clima, la predicción de los valores medios, dispersión y valores extremos de variables meteorológicas en escalas de tiempo mayores, vale decir, meses, estaciones, años, etc.

Tabla 2. Resumen de procesamiento de casos

Casos sin ponderara	N	Porcentaje
Incluido en el análisis	68	100,0
Casos seleccionados	Casos perdidos	0
	Total	68
Casos no seleccionados	0	,0
Total	68	100,0

a. Si la ponderación está en vigor, consulte la tabla de clasificación para el número total de casos.

La tabla muestra que son 68 las comunidades analizadas, aclarando que corresponden a las provincias Nor Lipez, Sud Lipez y Modesto Omiste que corresponden a la zona sud del departamento de Potosí. No existen casos perdidos o datos faltantes.

Tabla 3. Codificación de Variable Dependiente

Valor original	Valor interno
Control	0
Riesgo Agrícola	1

La variable dependiente refiere a 0 como la ausencia de la característica que interesa al estudio o llamado también control y se denota con 1 a la característica presente que interesa analizar, vale decir el riesgo agrícola.

Tabla 4. Codificaciones de variables categóricas

		Frecuencia	Codificación de parámetro (1)
% Perdido superficie aproximado	menor al valor promedio	25	1,000
	mayor al valor promedio	43	,000
Ingreso Familiar (Bs.)	menor al valor promedio	48	1,000
	> al valor promedio	20	,000
Helada	ocurrente	34	1,000
	recurrente	34	,000
Sequía	ocurrente	62	1,000
	recurrente	6	,000
Haba	ocurrente	63	1,000
	recurrente	5	,000
Papa	ocurrente	44	1,000
	recurrente	24	,000
Hortalizas	ocurrente	65	1,000
	recurrente	3	,000
Frutales	ocurrente	51	1,000
	recurrente	17	,000
Maíz	ocurrente	21	1,000
	recurrente	47	,000
Población (hab)	menor al valor promedio	49	1,000
		19	,000

Se puede ver en esta tabla la frecuencia absoluta de cada valor, siendo dicotómica la característica estudiada para cada variable referida a algún factor que incide en el riesgo agrícola en la zona sud del departamento de Potosí. Por ejemplo, se verifica que 25 observaciones han perdido en porcentaje

aproximado de superficie cultivada, menos del valor promedio calculado, en cambio son 43 comunidades las que no han perdido en porcentaje igual o mayor al valor promedio de dicha variable. El ingreso familiar promedio menor a su media aritmética es de 48 comunidades y en cambio son 20 las comunidades que registran un ingreso familiar que en promedio es igual o mayor a la media aritmética.

En cuanto al registro de heladas se verifica que se han distribuido en exactamente a la mitad cada caso, siendo 34 las comunidades que han reportado heladas y otras 34 comunidades de la zona sud que reporta heladas, pero de manera recurrente. Existen 62 comunidades que presentaron sequía, pero han sido 6 las comunidades que han tenido este evento climático de manera recurrente.

El reporte de daño a su producción de haba por una sola vez representa a 63 comunidades, en cambio han sido 5 las comunidades que de manera reiterativa o recurrente han reportada la afectación a este cultivo. Son 44 comunidades que reportan daño en la papa por una sola vez en este periodo estudiado, sin embargo 24 comunidades señalaron que la afectación a la papa ha sido reiterativa.

En cuanto a las hortalizas el daño registrado por una única vez ha sido establecido en 65 comunidades, sin embargo 3 comunidades reportan recurrente la afectación en su producción de hortalizas. Acerca de los frutales se observa que el hecho fue denunciado por 51 comunidades, pero solo 17 comunidades tienen este problema de manera recurrente. El reporte de daño a la producción de maíz se dio en 21 comunidades, pero fueron 47 las comunidades que tuvieron este problema de manera recurrente. En cuanto a la población, la tabla muestra que 49 comunidades tienen en promedio una población inferior al promedio establecido para la zona sud del departamento de Potosí.

DISCUSIÓN

El método multivariado de la regresión logística ha permitido explorar y pronosticar para la población estudiada, además de caracterizar el comportamiento del riesgo agrícola en la zona sud del departamento de Potosí, se logró conocer como respuesta científica la probabilidad del riesgo agrícola, así mismo, el uso importante del modelo es clasificar en forma global a las comunidades de las provincias Nor Chichas, Sud Chichas y Modesto Omiste en una de las 2 alternativas: tiene riesgo agrícola versus no tienen riesgo agrícola.

Lo establecido, se obtiene sustituyendo los valores de las variables independientes en el modelo ajustado, lo que produce la probabilidad estimada, para predecir si la comunidad está en la alternativa 0 o 1, el software SPSS v.22 automáticamente emplea un punto de corte de 0,5. Esto significa que la comunidad con probabilidad estimada $< 0,5$ se clasifica como (ocurrencia-ausencia) = 0 (Control) en cambio se clasifica $> 0,5$ si la probabilidad estimada señala (recurrencia-presencia) = 1 (riesgo agrícola) en la zona sud del departamento de Potosí, que es la que comprende este estudio. En resumen, el modelo final ha sido clasificado correctamente 90 % de las comunidades con riesgo agrícola y un 50 % de las comunidades que no evidencian riesgo agrícola. El porcentaje global de las comunidades clasificados correctamente es de 79,4 %.

De las 10 variables independientes, consideradas inicialmente para el modelo de regresión logística, se obtuvo el modelo ajustado luego del tratamiento bajo el método de Wald, quedando para el modelo solo los coeficientes que pasaron la prueba de significancia de sus coeficientes individuales con un nivel de confianza del 95%.

Se puede afirmar en base a los datos analizados que el perfil del riesgo agrícola en la zona sud del

departamento de Potosí está influido en gran manera por los reportes de daño a frutales y por el reporte de heladas, estas serían las causas del riesgo agrícola y por tanto en ello podemos intervenir o incidir en la realidad agrícola.

Dentro de la construcción del modelo, se desestima la incorporación de las variables que detalla la tabla, pues no aportan de acuerdo a criterio teórico a la comprensión del fenómeno estudiado.

CONCLUSIONES

El modelo de regresión logística tiene un basamento teórico importante y adecuado para el análisis del riesgo agrícola. La regresión logística ofrece un instrumento técnico estadístico útil a corto plazo para establecer la existencia de riesgo agrícola en la zona sur del departamento de Potosí.

El tratamiento de datos estadísticos de riesgo agrícola es esencial para contribuir en la construcción del Sistema de Alerta Temprana. El método de regresión logística permitió el conocimiento en términos probabilísticos del riesgo agrícola en la zona sud del departamento de Potosí.

Con los resultados de la regresión logística, mediante el estadístico de Wald para el riesgo agrícola, se determinó que las variables más influyentes son: “Reporte de heladas” y “Daño a los frutales”, teniendo estas variables un p-valor asociado al estadístico de Wald menor al 5%.

Las variables con poca influencia en el modelo son: la población, el ingreso familiar por mes, los reportes de sequía, el daño al haba, el daño a la papa, el daño al maíz y el daño a las hortalizas, así como el porcentaje de daño en la superficie cultivada.

Los datos son altamente orientadores y favorecen medidas resilientes y nos preparan para la construcción de un sistema de alerta temprana. En estudios futuros de la zona geográfica se deberá tomar especial interés a los eventos climáticos de la helada y considerar que el producto agrícola importante en este modelo son los frutales.

Se prueba que la regresión logística ha permitido el tratamiento de datos cualitativos y ha presentado la posibilidad de reducir el costo del tratamiento de información en base a las variables más determinantes identificados en el modelo de regresión logística.

Los resultados de este trabajo ayudan a que las entidades gubernamentales observen qué se puede hacer para que las comunidades se anticipen al riesgo agrícola y por tanto su vulnerabilidad en diferentes factores como en su economía familiar que luego afectará al PIB nacional. Es evidente que son diversos y complejas las variables que influyen en el riesgo agrícola por lo que este trabajo es el inicio de un abordaje más profundo en el ámbito del análisis multivariado.

REFERENCIAS

- Álvarez, C. Herrera, M., Herrera, E., y Araya, G. (2021). Investigation of Variables that Determine Levels of Self-Esteem in Central American Schoolchildren: A Binary Logistic Regression Model. *MHSalud*, 18(2). 1-22. DOI: <https://doi.org/10.15359/mhs.18-2.2>
- Andrade, G., y Arteaga, M. (2021). Riesgos laborales: Causas y consecuencias en tiempos de Covid-19. *Revista Dominio de las Ciencias*, 7(3). 247-265. <http://dominiodelasciencias.com/ojs/index.php/es/index>
- Barreiro M., y Useche, L. (2023). Métodos multivariantes para la construcción de índices compuestos en la gestión cultural: una revisión de la literatura. *Minerva Journal*, 9(10). 31-39. <https://doi.org/10.47460/>

minerva.v4i10.93

- Camacho, J. (2022). Análisis de regresión logística aplicado al modelamiento espacial de las causas de deforestación en el departamento del Guaviare durante el periodo 2005-2020 y proyección de escenarios de deforestación a 2030." Cuadernos de Geografía: Revista Colombiana de Geografía 31(2). 255-280. <https://doi.org/10.15446/rcdg.v31n2.98012>
- Departamento de Agricultura de Estados Unidos. (21 de abril de 2019). El Departamento de Agricultura de los Estados Unidos (USDA) Extiende su Iniciativa del Strike Force para impulsar el Crecimiento Rural y Oportunidades para Construir Economías Rurales. <https://www.usda.gov/media/news-archives/informacion-y-servicios-del-united-states-department-agriculture-usda-por-sus>
- García, R. García, G. González, D., y García, R. (2020). Modelo de regresión logística para estimar la dependencia según la escala de Lawton y Brody. Medicina de Familia. SEMERGEN, 36(7). 365-371. <https://www.elsevier.es/es-revista-medicina-familia-semergen-40-articulo-modelo-regresion-logistica-estimar-dependencia-S1138359310001322>
- Hernández, R., Fernández, C. y Baptista, P. (2014). Metodología de la investigación. México: Mc Graw Hill.
- Hurtado, J. (2006). El Proyecto de Investigación. Metodología de la investigación Holística. 4ª Edición ampliada. Quirón, Sypal, Caracas. 168 pp. <https://www.geocities.ws/inesotilia28873/fase3/t3m3.html>
- Llaugel, F., y Fernández, A. (2021). Evaluación del uso de modelos de regresión logística para el diagnóstico de instituciones financieras. Revista Ciencia y Sociedad, 36(4). 590-627. file:///Evaluacion_del_uso_de_modelos_de_regresion_logisti.pdf
- Organización para la Cooperación y el Desarrollo Económicos. (06 de julio 2023). Informe OCDE-FAO Perspectivas agrícolas 2023-2032 describe las principales tendencias en la producción, el consumo y el comercio. <https://www.fao.org/venezuela/noticias/detail-events/fr/c/1644151/>
- Ouorou, M. Aleixo, J., Caraciolo, R., y Álvarez, D. (2021). Modelos de Regresión Logística Multinomial Ordinal y Redes Neuronales Artificiales para la clasificación de madera aserrada. Revista Forestal Mesoamericana Kurú, 18(43).29-40. <file:///Dialnet ModelosDeRegresionLogisticaMultinomialOrdinalYRede-8131067.pdf>