



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

FACTORES QUE INFLUYEN EN LA PRODUCCIÓN DE PAPA
(SOLANUM TUBEROSUM), ECUADOR 2020

Trabajo de Titulación

Tipo: Proyecto de Investigación

Presentado para obtener el grado académico de:

INGENIERO ESTADÍSTICO

AUTOR: URQUIZO URQUIZO DIEGO VLADIMIR

DIRECTOR: ING. HÉCTOR SALOMÓN MULLO GUAMINGA, Phd.

Riobamba – Ecuador

2022

© 2022, **Diego Vladimir Urquizo Urquizo**

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el derecho de Autor.

Yo, DIEGO VLADIMIR URQUIZO URQUIZO, declaro que el presente trabajo de titulación es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este trabajo de titulación; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 07 de diciembre de 2022

Diego Vladimir Urquizo Urquizo
0605262518

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Titulación, certifica que: El Trabajo de Investigación: **FACTORES QUE INFLUYEN EN LA PRODUCCIÓN DE PAPA (SOLANUM TUBEROSUM), ECUADOR 2020**, realizado por el señor: **DIEGO VLADIMIR URQUIZO URQUIZO**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación. El mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

	Firma	Fecha
Ing. Tania Paulina Morocho Barrionuevo, Mgs. PRESIDENTE DEL TRIBUNAL	 _____	2022-12-07
Ing. Héctor Salomón Mullo Guaminga, Phd. DIRECTOR DEL TRABAJO DE TITULACIÓN	 _____	2022-12-07
Ing. Johanna Enith Aguilar Reyes, Mgs. ASESORA DEL TRABAJO DE TITULACIÓN	 _____	2022-12-07

DEDICATORIA

Este trabajo es dedicado primero a Dios y mi familia, por todo el apoyo brindado durante este proceso de aprendizaje. Con mención especial a mis padres Sebastián y Sara por la inculcación de sus valores, principios y mentalidad de luchar por las metas planteadas además del apoyo económico. A mis hermanos Braulio y Aylin, que con su ejemplo de perseverancia y experiencia han sabido guiarme por el buen camino.

Diego

AGRADECIMIENTO

El más franco agradecimiento a toda la planta docente de la Escuela de Estadística por compartir los conocimientos que han sido necesarios para crecer como profesional y persona. De manera especial agradecer al Ingeniero Héctor Mullo tutor de este trabajo de titulación y docente de algunas de las materias a lo largo de mi trayectoria como estudiante en la institución, a la Ingeniera Johanna Aguilar miembro del trabajo de titulación que ha guiado de buena manera en el camino como estudiante y en la elaboración de este trabajo.

Diego

ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS.....	x
ÍNDICE DE ILUSTRACIONES.....	xi
ÍNDICE DE ANEXOS	xii
RESUMEN.....	xiii
SUMMARY	xiv
INTRODUCCIÓN	1

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL	5
1.1. Bases conceptuales.....	5
<i>1.1.1. La agricultura.....</i>	<i>5</i>
<i>1.1.2. Cultivos Permanentes.....</i>	<i>6</i>
<i>1.1.3. Cultivos Transitorios y barbechos</i>	<i>6</i>
<i>1.1.4. Pastos Cultivados</i>	<i>6</i>
<i>1.1.5. Pastos Naturales</i>	<i>6</i>
1.2. La papa	6
<i>1.2.1. Taxonomía.....</i>	<i>6</i>
<i>1.2.2. Anatomía</i>	<i>7</i>
<i>1.2.3. Origen</i>	<i>8</i>
<i>1.2.4. Factores medioambientales</i>	<i>8</i>
1.2.4.1. Temperatura	8
1.2.4.2. Agua.....	8
1.2.4.3. Suelo	9
1.2.4.4. Viento.....	9
<i>1.2.5. Factores controlables.</i>	<i>9</i>
1.2.5.1. Fertilizantes	9

1.2.5.2.	Fertilizantes Orgánicos	9
1.2.5.3.	Fertilizantes Químicos	10
1.2.5.4.	Fitosanitario	10
1.2.5.5.	Plaguicidas	10
1.2.6.	<i>Producción en el Ecuador</i>	11
1.2.7.	<i>Factores influyentes en la producción de papa presentados los el International Potato Center.</i>	12
1.3.	Teoría estadística	12
1.3.1.	<i>Detección de datos atípicos</i>	12
1.3.2.	<i>Análisis exploratorio de datos (AED)</i>	14
1.3.3.	<i>Análisis de relación o asociación entre variables</i>	15
1.3.3.1.	Coeficiente de correlación de Pearson	15
1.3.3.2.	Test chi-cuadrado de independencia.	15
1.3.4.	<i>Modelo de regresión lineal múltiple</i>	16
1.3.4.1.	Estimación y significancia de parámetros.....	17
1.3.4.2.	Coeficiente de determinación.....	18
1.3.4.3.	Residuos.....	18
1.3.4.4.	Supuestos de un modelo de regresión	19
1.3.5.	<i>Validación del modelo de regresión.</i>	20
1.3.6.	<i>Comparación y selección de modelos</i>	20
1.3.6.1.	AIC.....	21
1.3.6.2.	BIC	21
1.3.7.	<i>Técnicas multivariadas</i>	21
1.3.7.1.	Análisis de correspondencia.....	22
1.3.7.2.	Análisis factorial de datos mixtos (AFDM)	24

CAPITULO II

2.	MARCO METODOLÓGICO.....	28
-----------	--------------------------------	-----------

2.1.	Tipo de la Investigación.	28
2.2.	Diseño de la investigación no experimental.	28
2.2.1.	<i>Localización de estudio</i>	29
2.2.2.	<i>Población de estudio</i>	29
2.2.3.	<i>Método de muestreo</i>	29
2.2.4.	<i>Tamaño de la muestra</i>	30
2.2.5.	<i>Técnica de recolección de datos.</i>	30
2.2.6.	<i>Identificación de variables</i>	31
2.2.7.	<i>Modelo estadístico</i>	31
2.3.	Variables en estudio.	32
2.3.1.	<i>Operacionalización de variables.</i>	32

CAPÍTULO III

3.	MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS.	34
3.1.	Detección de datos atípicos.	34
3.2.	Análisis Exploratorio de Datos (AED)	34
3.3.	Análisis de relación o asociación entre variables.	38
3.3.1.	<i>Correlación de Pearson para variables cuantitativas</i>	38
3.3.2.	<i>Contraste de independencia para variables cualitativas</i>	39
3.4.	Modelo de regresión lineal múltiple	40
3.4.1.	<i>Selección del modelo</i>	42
3.4.1.	<i>Validación del modelo</i>	43
3.4.1.1.	Multicolinealidad-Factor de inflación de varianza (VIF)	43
3.4.1.2.	Supuestos del modelo.	44
3.5.	Técnicas multivariantes	46
3.5.1.	<i>Análisis de correspondencia</i>	46
3.5.2.	<i>Análisis factorial de datos mixtos AFDM</i>	48
3.5.2.1.	Primer análisis factorial de datos mixtos.	48

3.5.2.2.	Segundo análisis factorial de datos mixtos.	51
3.6.	Identificación de los factores influyentes obtenidos con diferentes técnicas estadísticas.	54
3.7.	Comparación de los factores identificados con los factores presentados por el International Potato center.	55

CONCLUSIONES

RECOMENDACIONES

BIBLIOGRAFÍA

ANEXO A: ACCESO A INFORMACIÓN INVESTIGADA

ANEXO B: ALGORITMO PARA DETECCIÓN DE DATOS ATÍPICOS (MATLAB)

ANEXO C: CÓDIGO EN R.

ÍNDICE DE TABLAS

Tabla 1-1: Formato anova para el contraste de hipótesis de los parámetros estimados.....	18
Tabla 1-2: Operacionalización de variables.....	32
Tabla 1-3: Distribución estadística de frecuencia de la variable superficie sembrada.....	34
Tabla 2-3: Distribución estadística de frecuencia de la variable producción de papa	35
Tabla 3-3: Resumen estadístico de las variables cuantitativas.....	36
Tabla 4-3: Resumen de la distribución estadística de frecuencias variables cualitativas.	36
Tabla 5-3: Coeficiente de correlación de Pearson entre las variables cuantitativas.....	38
Tabla 6-3: Valores-p resultantes del contraste chi-cuadrado de independencia.	39
Tabla 7-3: Continuación de la Tabla 3-6.	39
Tabla 8-3: Resumen del modelo de regresión “Modelo 1”, variable respuesta producción.	40
Tabla 9-3: Resumen del modelo de regresión “Modelo 2”, variable respuesta producción.	41
Tabla 10-3: Valores AIC y BIC de los modelos estimados.	42
Tabla 11-3: Factores de inflación de la varianza (VIF) del Modelo 2	43
Tabla 12-3: Coordenadas de las dos primeras dimensiones para la representación gráfica del análisis de correspondencia	46
Tabla 13-3: Coordenadas de las variables cuantitativas en el AFDM	52
Tabla 14-3: Coordenadas de las variables cualitativas en el AFDM.	53
Tabla 15-3: Resumen de los factores influyentes identificados con las técnicas estadísticas....	54
Tabla 16-3: Comparación factores influyentes presentados por el CIP y las identificadas en el estudio.....	55

ÍNDICE DE ILUSTRACIONES

Ilustración 1-1: Superficie de cada categoría del uso agropecuario en el Ecuador 2017-2020... 5	5
Ilustración 2-1: Anatomía de la papa	7
Ilustración 1-2: Localización del estudio.	29
Ilustración 1-3: Histograma de frecuencia de la variable superficie sembrada.	34
Ilustración 2-3: Histograma de frecuencia de la variable producción de papa.....	35
Ilustración 3-3: QQ-plot para analizar la normalidad de los residuos del “modelo 2”.....	44
Ilustración 4-3: Valores absolutos de los residuos estandarizados vs los valores ajustados.	45
Ilustración 5-3: Representación gráfica del análisis de correspondencia de variables involucradas en el estudio.	47
Ilustración 6-3: Grafico de sedimentación del primer AFDM.	48
Ilustración 7-3: Representación de variables con el primer AFDM.	49
Ilustración 8-3: Calidad de representación de las variables con el primer AFDM.....	49
Ilustración 9-3: Contribución de las variables en el primer AFDM.	50
Ilustración 10-3: Grafico de sedimentación del segundo AFDM.....	51
Ilustración 11-3: Representación de variables con el segundo AFDM.	51
Ilustración 12-3: Representación de las variables cuantitativas en el plano factorial.	52
Ilustración 13-3: Representación de las variables cualitativas en el plano factorial.	53

ÍNDICE DE ANEXOS

ANEXO A: ACCESO A INFORMACIÓN INVESTIGADA

ANEXO B: ALGORITMO PARA DETECCIÓN DE DATOS ATÍPICOS (MATLAB).

ANEXO C: CÓDIGO EN R.

RESUMEN

La presente investigación tuvo como objetivo determinar los factores influyentes en la producción de papa en Ecuador usando la información recolectada por el INEC en la Encuesta de Superficie y Producción Agropecuaria Continua 2020 (ESPAC) y comparar con agentes presentados por el International Potato Center (CIP). El estudio fue de tipo no experimental con un alcance exploratorio relacional, se analizó 13 variables, 2 cuantitativas (Superficie sembrada, producción) y 11 cualitativas (semilla usada, afectación en la producción, preparación de suelo, uso de: riego, fertilizantes, fertilizante orgánico, fertilizante químico, plaguicida orgánico, plaguicida químico, fitosanitario, mes de siembra), con el uso de R-Rstudio se realizó: análisis exploratorio de datos(AED), análisis de regresión lineal múltiple(ARLM) y análisis factorial de datos mixtos(AFDM). Los resultados mostraron: producción media de 12Tm³/ha sembrada, 97% de terrenos produjeron entre (0.01-150) Tm³, fuerte correlación entre la producción y superficie sembrada, dependencia entre semilla usada, uso de: fertilizante, plaguicidas, fitosanitario, riego y mes de siembra. Según el ARLM y AFDM concluyeron con mayor significancia que los factores que influyen en la producción son: superficie sembrada, uso de: fertilizantes, plaguicida, fitosanitario, tipo de semilla, afectación en la producción; a comparación de la afirmación del CIP “mes de siembra” fue un factor significativo únicamente cuando el terreno no tiene riego, es decir, que la producción en un terreno con riego no depende del mes de siembra. Los resultados obtenidos garantizaron su confiabilidad y fueron concluyentes al ser muy similares al de los estudios realizados por el CIP, por otro lado, la investigación aportó conocimiento acerca de la problemática estudiada que a partir del análisis realizado y conocimiento logrado se recomendó que los productores/asesores tengan un juicio previo a base de análisis de datos para el asesoramiento y elaboración de planes de mejoramiento de la producción de papa en Ecuador.

Palabras clave: <FACTORES INFLUYENTES>, <PAPA>, <PRODUCCIÓN>, <INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS>, <ANÁLISIS ESTADÍSTICO>, <RSTUDIO>.

0120-DBRA-UPT-2023



SUMMARY

The objective of this research was to determine the influential factors in potato production in Ecuador using the information collected by INEC in the Continuous Agricultural Surface and Production Survey 2020 (ESPAC) and compare with agents presented by the International Potato Center (IPC). The study was of a non-experimental type with a relational exploratory scope, 13 variables were analyzed, 2 quantitative (Area planted, production) and 11 qualitative (seed used, affectation in production, soil preparation, use of: irrigation, fertilizers, organic fertilizer, chemical fertilizer, organic pesticide, chemical pesticide, phytosanitary, planting month), exploratory data analysis (EDA), multiple linear regression analysis (MLRA) and factorial analysis of mixed data (FAMD) were done with the use of R-RStudio. The results showed: average production of 12Tm³/ha planted, 97% of the lands produced between (0.01-150) Tm³, strong correlation between production and planted area, dependence between seed used, use of: fertilizer, pesticides, phytosanitary, irrigation and planting month. According to the MLRA and FAMD concluded with greater significance that the factors that influence production are: planted area, use of: fertilizers, pesticides, phytosanitary products, type of seed, impact on production; Compared to the IPC statement “planting month” was a significant factor only when the land is not irrigated, that is, production in irrigated land does not depend on the planting month. The results obtained guaranteed their reliability and were conclusive as they were very similar to those of the studies performed by the IPC, on the other hand, the research provided knowledge about the problem studied that, based on the analysis done and the knowledge obtained, it was recommended that the producers/ advisors have a prior judgment based on data analysis for advice and preparation of improvement plans for potato production in Ecuador.

Keywords: <INFLUENTIAL FACTORS>, <POTATOES>, <PRODUCTION>, <NATIONAL INSTITUTE OF STATISTICS AND CENSUES>, <STATISTICAL ANALISYS>, <RSTUDIO>.



Edgar Mesias Jaramillo Moyano

0603497397

INTRODUCCIÓN

La agricultura hoy en día es uno de los temas de investigación más estudiados e investigados, por su implicación en distintos ámbitos dentro de un territorio, siendo así, una actividad que conforman las bases del desarrollo de un estado por su gran aporte en áreas de la economía, alimentación y empleo dentro del mismo.

De ahí su importancia que según lo descrito en (Bank, 2018) la agricultura es una actividad que tiene la capacidad de reducir la pobreza hasta en un 80% a nivel mundial.

En Ecuador, como describe (Agricultura, 2019) la importancia de la agricultura radica en que aporta con un 8% a la producción total anual del país (Producto Interno Bruto), además, de generar empleo en el sector rural.

El sector agrícola también contribuye de gran manera al comercio del país, en especial los productos que sirven para la exportación como: el banano, cacao café, papas, entre otros, siendo así también una fuente de ingreso de divisas para la economía de nuestro país.

La papa conocida también como patata de un nombre científico *Solanum tuberosum* es un alimento comestible que contiene una gran cantidad de carbohidratos, este alimento es de origen andino y hoy en día en la mayor parte de la población serrana de Ecuador es un producto indispensable ya sea para usarlo de alimento o como un medio para generar ingresos mediante la comercialización de la misma. Además, que según (Arias, 2010) la papa se ubica entre los cinco cultivos más importantes a nivel mundial luego del trigo, arroz y el maíz.

Sin embargo, el cultivo de este tubérculo no es nada sencillo, debido a la presencia de diferentes factores que influyen en la producción de la misma, siendo estas controlables o naturales las cuales no están claras por lo que con esta investigación se pretende encontrar y agruparlos de manera que ayude al agricultor la maximización en la cosecha de la misma.

Antecedentes

En el año 2020, de acuerdo a (INEC, 2021) en Ecuador se ha destinado 25924 hectáreas para el cultivo de papa, donde se han producido 408313 toneladas métricas (Tm) de dicho producto.

En todo el país los encargados de la producción de papa han sido los pequeños agricultores donde la proporción de su producción ha sido aproximadamente 16 Tm por hectárea, donde en la

mayoría de los casos estos pequeños productores no tienen acceso a diferentes medios para conseguir recursos que ayuden a mejorar su cosecha, además de no saber de qué factores depende el aumento de producción de papa, dificultando así el incremento de ganancia y mejoras de la economía de este grupo de agricultores.

En la serranía ecuatoriana la papa constituye un producto básico que está presente diariamente en las mesas de esta población gracias a su amplia variedad y su facilidad de cultivo.

Según (Manuel Pumisacho, 2002) la papa se ha convertido en uno de los cultivos de mayor valor económico y alimenticio para el país, ya que comprende una importante fuente de ingresos para las familias campesinas de la Sierra y ocupa el quinto lugar en producción después de la caña de azúcar, arroz, banano y yuca.

En Ecuador se ha identificado la existencia de una significativa presencia de terrenos donde se han cultivado la papa, siendo Carchi, Chimborazo, Tungurahua, Cotopaxi y Pichincha las principales provincias representantes de esta producción.

Planteamiento del problema

Enunciado del problema

En la agricultura, específicamente en la serranía ecuatoriana la papa es el principal producto sembrado, además este producto está presente en la mayoría de las mesas, siendo además un producto que ayuda a los ingresos de esta población, por lo que es menester un análisis que determine los factores influyentes en la producción de la papa, lo cual sirva de ayuda para incrementar la cosecha y de la misma manera el incremento de comercialización que ayude en los ingresos de las familias cultivadoras de este producto.

Formulación (Incógnita)

¿Cuáles son los factores que influyen en la producción de papa en el Ecuador basado en la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) 2020?

Justificación

En Ecuador, la papa es uno de los productos más importantes y presentes en el menú de sus habitantes, sumado a que también es una fuente de ingreso para las familias cultivadoras de ello. Este tubérculo es cultivado en alturas desde los 2700 a 3400 msnm, a una temperatura de 10°C a 15°C por lo que es evidente que la producción se desarrolla en la región Sierra del Ecuador, con mayor presencia en las provincias de Carchi, Chimborazo, Tungurahua, Cotopaxi, Pichincha, Bolívar, Cañar, Azuay e Imbabura (Asas, 2013).

La producción de papa es considerada una de las fuentes de ingreso respaldándonos en que según (Asas, 2013) los habitantes del Ecuador destinan aproximadamente 10% de sus recursos para la compra de este tubérculo. La producción de la papa en la región andina del Ecuador reviste singular importancia desde el punto de vista económico, social y cultural (Benitez, 2003).

La papa se desarrolla de una mejor manera en suelos ricos de materia orgánica, además de la presencia de una considerable humedad, sin embargo, estos años, según datos recolectados por el INEC en las ESPAC, se ha notado una disminución de producción de este cultivo. De ahí su importancia la realización de esta investigación, para identificar a base de un análisis de datos los factores que influyen en la producción de este producto y dar a los agricultores una herramienta más para conocer y prestar mayor atención a dichos factores, aumentando así el rendimiento lo que también traerá como consecuencia una obtención de beneficios económicos mayor, permitiendo mejorar la calidad de vida de personas dedicadas al cultivo de este tubérculo.

OBJETIVOS

Objetivo general

- Identificar mediante técnicas estadísticas univariantes y multivariantes los factores que influyen en la producción de papa en el Ecuador, tomando en cuenta los factores influyentes presentados por el International Potato Center.

Objetivos específicos

- Realizar un análisis estadístico exploratorio de las variables involucradas en la investigación.
- Determinar dependencia bivariado entre las variables relacionadas al cultivo de papa.

- Proponer un modelo de regresión múltiple que explique y prediga la producción de papa en toneladas métricas.
- Describir relaciones existentes entre variables cualitativas relacionadas al cultivo de la papa mediante técnicas estadísticas multivariantes.
- Identificar los factores influyentes obtenidos con técnicas estadísticas.
- Comparar los factores influyentes en la producción de papa identificados en la investigación con los presentados por el International Potato Center.

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL

1.1. Bases conceptuales

1.1.1. La agricultura

La agricultura en el Ecuador cumple un rol fundamental en el ámbito de la economía, donde el uso agropecuario se encuentra dividida en cuatro categorías las que son: Cultivos Permanentes, Cultivos Transitorios y barbechos, Pastos Cultivados y Pastos Naturales.

Observando la gráfica según las cifras del INEC obtenidas en la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) 2020 podemos apreciar la disminución de la superficie en todas las categorías del uso agropecuario, sin embargo, en el año 2020 hubo un incremento en la superficie del uso agropecuario en la categoría de cultivos transitorios y barbechos.

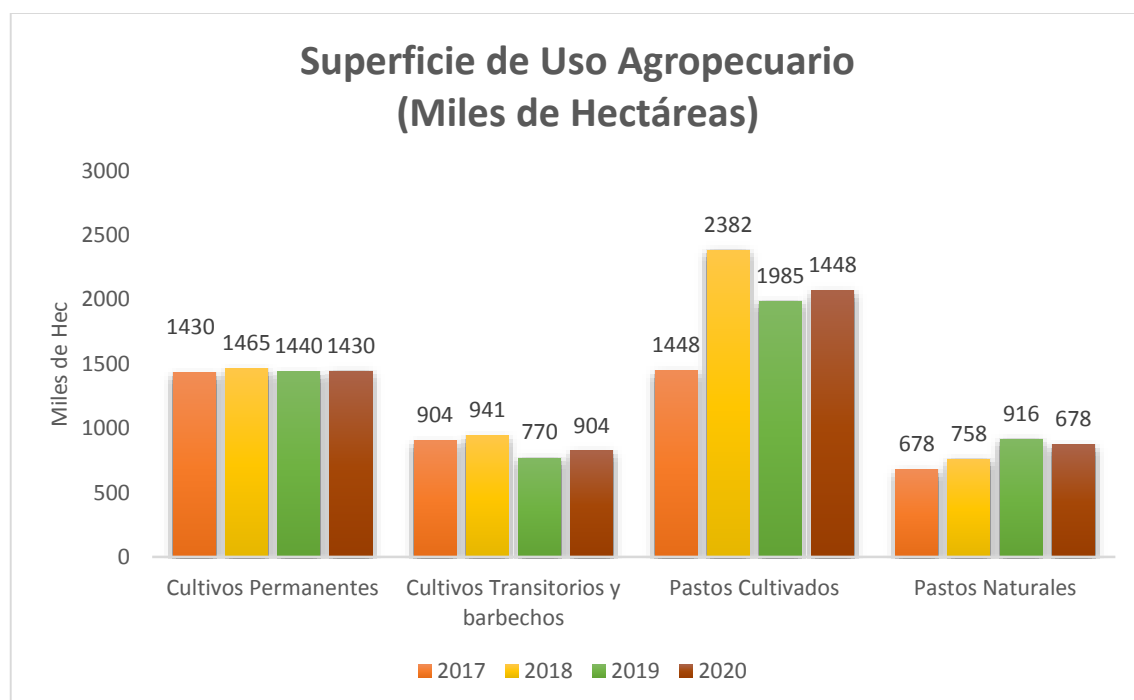


Ilustración 1-1: Superficie de cada categoría del uso agropecuario en el Ecuador 2017-2020.

Realizado por: Urquizo, Diego, 2022.

1.1.2. Cultivos permanentes

Es también conocida como cultivo perenne, como un aspecto principal diferenciador de los cultivos transitorios es que su desarrollo y crecimiento es mucho mayor, además de generar varias cosechas antes de terminar su ciclo de vida. El uso de superficie para el cultivo de esta ha ido disminuyendo a lo largo de los años donde, algunos de estos cultivos representativos son el banano y cacao.

1.1.3. Cultivos transitorios y barbechos

Se caracterizan por tener una sola cosecha los productos que representan a esta categoría son: el maíz, la papa, el arroz, entre otros, para el 2020 a diferencia de los demás cultivos el uso de la superficie agropecuaria para esta categoría ha aumentado en 134 (Miles de hectáreas).

1.1.4. Pastos cultivados

Son grupos de gramíneas y leguminosas, utilizados en su gran mayoría como alimento de animales que cumplen ciertas funciones como: contribución de proteínas, carbohidratos y apoyo al crecimiento. Este pasto también ha sufrido un decremento en relación a los años anteriores.

1.1.5. Pastos naturales

Son pastos específicamente destinados al pastoreo y gracias a la captación del carbono estos pastos soportan las variaciones climáticas, además de ayudar al suelo en la retención del agua obtienen la resistencia suficiente a las sequías. Esta categoría es la más pequeña en el uso de superficie agropecuaria del Ecuador y a pesar de esto sigue en decremento.

1.2. La papa

1.2.1. Taxonomía

La papa ha sido motivo de desacuerdo por muchos años en cuanto a su clasificación taxonómica, donde en 1972 científicos según (Wikimedia, 2003) decidieron como su clasificación como:

Reino: *Plantae*

Familia *Solanales*

Género: *Solanum*

Subgénero: *Potatoe*

Especie: *Solanum tuberosum*

1.2.2. Anatomía

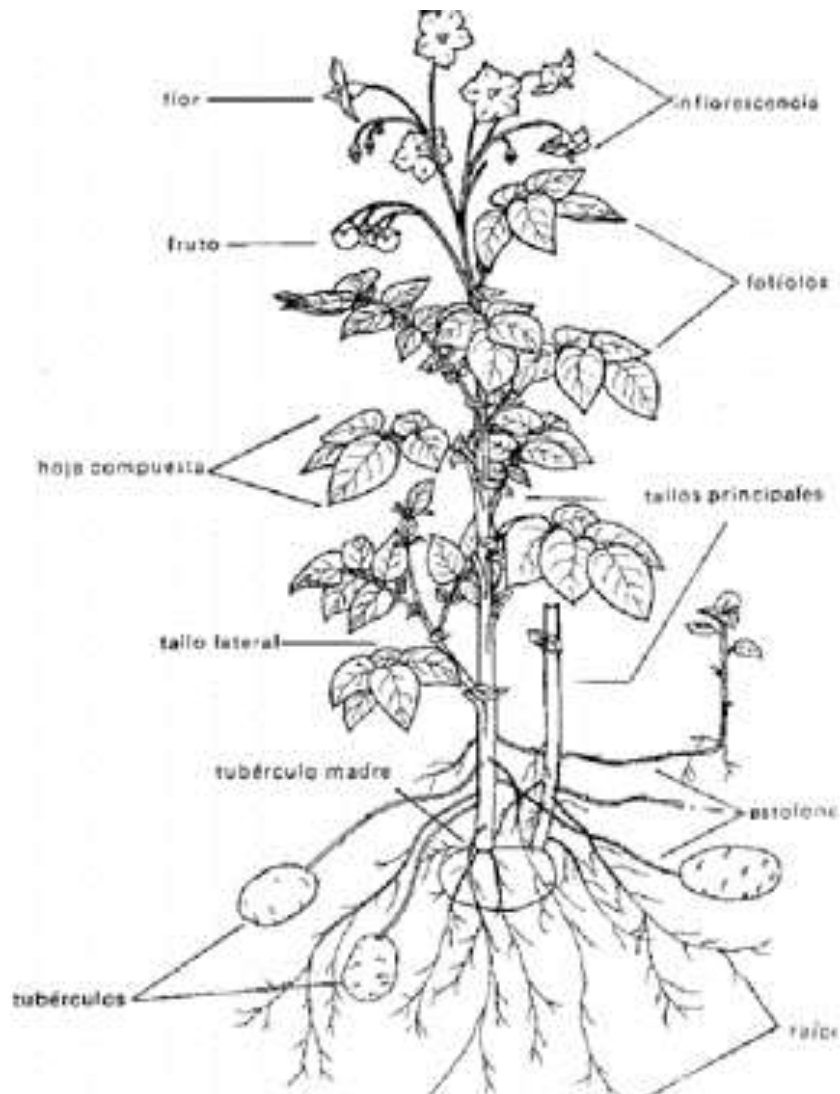


Ilustración 2-1: Anatomía de la papa

Fuente: ECURED, 2016.

Flor: Las flores son de diferentes colores dependiendo la clase de papa y tienen forma pentámera.

Tallo: Su tallo es de forma triangular que a su vez se ramifica, de las cuales el tallo principal es el que crece directamente de la semilla y sus ramificaciones son tallos secundarios. Los tallos secundarios que resultan de la ramificación y están cerca de la semilla se denominan estolón y cumplen la función de un tallo principal.

Hojas: Las hojas tienen forma lanceolada que se ubican de forma espiral en los tallos terciarios.

Raíz: Son delgadas y a partir de ellas aparecen ramificaciones de raíces laterales, y alcanzan rotundidades o longitudes de hasta 50 cm.

Tubérculo: Resulta cuando un estolón engrosa. Es considerado como una parte más del tallo el cual cumple la función de almacenamiento y reserva de alimento además de la reproducción.

Brote: Son formaciones que tienen lugar en los ojos del tubérculo, este es la principal característica para la identificación a la variedad perteneciente. Esto debido a que permite observar partes como tricomas, primordios foliares enclaustrados, lenticelas, radículas y estolones (Ecured, 2016).

1.2.3. Origen

Según (ARGENpapa, 2015) el origen de papa tiene lugar aproximadamente en el año 8000 a. C. en lugares altiplanos de los andes, cerca de los lugares al sur del Perú. Resultando ser junto al maíz unos de los productos clave en la alimentación de este país durante muchos años.

1.2.4. Factores medioambientales

Los factores medioambientales considerados también incontrolables según (ONDARURAL, 2021) son:

1.2.4.1. Temperatura

La papa se puede cultivar en terrenos a alturas de 2400 a 3700 msnm, donde sus temperaturas anuales en promedio están en un intervalo de 6° C y 14° C, además q su precipitación comprende alrededor de 700 a 1200 milímetros anuales de agua por ciclo.

1.2.4.2. Agua

En cuanto al agua este cultivo necesita una humedad uniforme del terreno días antes del sembrío, posterior a esto, es necesario aplicar un riego proporcional, es decir que no sea excesivo peor aún insuficiente. Se debe aplicar el primer riego después de haber brotado el mayor número de plantas,

lo cual, en un cultivo bien conducido, se produce entre los 20 a 30 días después de la siembra. Es recomendable aplicar riego de agua en un lapso de 12 a 15 días o a su vez en intervalos de tiempo donde se observa la necesidad de la plata.

1.2.4.3. Suelo

Los suelos ideales para el cultivo de papa son los llamados negros andinos, el cual debe estar debidamente preparado, mediante el arado, rastrado y nivelado, con una buena cantidad de abono preferencialmente orgánico.

1.2.4.4. Viento

Para la producción de papa es importante establecer una barrera rompevientos alrededor del área sembrado debido a que por su altura de sembrado los vientos son muy fuerte, con el fin de evitar daños en las planta o provocación de la erosión del suelo.

1.2.5. Factores controlables.

1.2.5.1. Fertilizantes

Los fertilizantes según (Rural, 2019) son sustancias o compuestos nutritivos para los vegetales en forma tal que pueden ser absorbidos por las plantas. Son usados para mejorar las características de las plantas, consecuentemente, aumentar la producción, reponer y evitar deficiencias de nutrientes. También según aporta (Finck, 1988) los fertilizantes son sustancias que se aplican directa o indirectamente a las plantas para favorecer su crecimiento, aumentar su producción o mejorar su calidad.

1.2.5.2. Fertilizantes orgánicos

Los fertilizantes orgánicos definidos por (Rural, 2019) son fertilizantes producto de una nula o poca participación humana.

Ejemplos de este fertilizante son el estiércol, humus de lombriz, recortes de hierba, entre otros.

1.2.5.3. Fertilizantes químicos

Por otro lado (Rural, 2019) afirma que los fertilizantes químicos son elaborados por el hombre, tienen distintos orígenes que pueden ser mineral, animal, vegetal o sintético.

Los principales fertilizantes químicos elaborados para la nutrición de la tierra son el nitrógeno, fósforo y potasio.

1.2.5.4. Fitosanitario

Los productos fitosanitarios definidos por (MITECO, 2011) son mezclas químicas que contienen varias sustancias activas y otros ingredientes, cuyo objetivo es la de proteger a los vegetales y sus productos de organismos nocivos. Estos productos contribuyen también al aumento del rendimiento en la agricultura, brindando la seguridad de obtener un buen alimento.

El uso excesivo y descontrolado de fitosanitarios pueden resultar desfavorables para la producción vegetal, produciendo riesgos para humanos, animales y medio ambiente, produciendo así un coste alto para la sociedad.

La función de estos productos es la proteger los cultivos contra organismos nocivos, regular el crecimiento, mejorar la conservación y destruir vegetales indeseables (Navas Cuenca, 2012).

1.2.5.5. Plaguicidas

Los plaguicidas según (OIRSA, 2005) son productos orgánicos y químicos utilizadas para el control de plagas, donde como detalla (Cipotato, 2011) las plagas a controlarse son: insectos, ácaros, hongos, bacterias, virus, etc. El uso de este producto en ocasiones no es indispensable debido a que son reemplazados por otras formas de control. El uso descontrolado de estos productos en especial los químicos, pueden ocasionar daños severos en la salud, medioambiente y la agricultura misma. Algunos de los plaguicidas orgánicos son el sulfato de cobre, el Piretrina y la rotenona, de igual forma, los plaguicidas químicos más usados son: Regent, Malathion, Phyton, entre otros.

1.2.6. Producción en el Ecuador

Por sus características ecológicas el área productora de papa de la sierra según (Valderrama, 1980) se subdivide en tres regiones principales:

La Región Norte que comprende las áreas frías de las provincias de Carchi e Imbabura, la Región Central que comprende las provincias de Pichincha; Cotopaxi Tungurahua, Chimborazo y Bolívar, finalmente la Región Sur esta comprendía por Cañar, Azuay y Loja.

La región norte comprende la producción de un 20% del total, esta región tiene características ecológicas inmejorables además de una tecnología avanzada a su disposición por lo que resulta promedios relativamente altos del rendimiento del cultivo.

La producción en su mayoría está enfocada en las variedades *Cholau Violeta* y *Curipamba* que resultaron de una adaptación de productos colombianos realizados por los antepasados. Por otro lado, la variedad que menos se cultiva en esta región es la Santa Catalina con una proporción de tan solo un 0.5% del total

La distribución de la precipitación permite hacer siembras escalonadas casi durante todo el año, lo cual reduce las fluctuaciones de producción. Esta región ha venido aumentando su participación en la producción nacional en los últimos años.

La Región Central es la más importante en cuanto a volumen de producción, con alrededor de un 40% del total. En esta región la producción al tener diferencias en sus características ecológicas y tecnológicas la producción se enfoca principalmente en las variedades cómo son Chela y Santa Catalina.

La Región Sur produce alrededor de un 40% del total, y también ha venido disminuyendo en importancia debido a poseer las peores condiciones para el cultivo de este producto además de no disponer tecnología actualizada por lo que las variedades producidas en esta región son la *Bolona* y la *Jubalefia*.

Este tubérculo cuenta con una superficie sembrada de aproximadamente 26 mil hectáreas, de donde se producen 400 mil toneladas repartidas en la alimentación de cada familia. En el país, cada persona consume un promedio de 30 Kg de papa al año (INEC, 2020).

Donde instituciones como el MAGAP e INIAP han buscado incentivar la producción, comercialización y consumo de este cultivo a través de GAD's y sectores gastronómicos.

1.2.7. Factores influyentes en la producción de papa presentados los el International Potato Center.

El International Potato Center (Centro internacional de la papa) fue fundada en 1971, este organismo según (Center, 2020) se enfoca en la investigación para el desarrollo con un enfoque en papa, camote y raíces y tubérculos andinos. Ofrece soluciones científicas innovadoras para mejorar el acceso a alimentos nutritivos y asequibles, fomentar el crecimiento de las empresas, empleo e impulsar la resiliencia climática de los sistemas agroalimentarios de raíces y tubérculos.

De acuerdo a las investigaciones realizados por el CIP los principales factores que influyen en la producción de papa según este organismo son:

Tipo de semilla usada.

Factores abióticos (Son factores que afectan y producen pérdidas en el cultivo).

Uso de fertilizantes (Químicos y Orgánicos).

Enfermedades.

Uso de plaguicidas (Químicos y Orgánicos).

Preparación del suelo.

Prácticas culturales (Mes de siembra)

Características agroecológicas y climáticas. (Manuel Pumisacho, 2002)

1.3. Teoría estadística

En la actualidad la estadística se ha convertido en una de las ramas de la matemática que mayor aplicación ha tenido dentro de la investigación por lo que hace posible obtener desarrollos y avances científicos importantes, paquetes accesibles y herramientas que faciliten realizar cálculos.

1.3.1. Detección de datos atípicos

Para producir conclusiones ciertas en una investigación, es conocido que todo depende de los datos que se disponga, donde una de las características necesarias para esta, es que el conjunto de datos que se posee no tenga casos, observaciones o individuos atípicos.

Un caso atípico u outlier según (Hawkings, 1980) es una observación que se desvía mucho de las demás y despierta sospechas de ser generadas por un mecanismo diferente.

La presencia de datos atípicos afecta a mediciones estadísticas numéricas que son muy sensibles a esta cómo son la media, la varianza, correlación etc. Creando así un error al momento de concluir los resultados obtenidos.

Aunque en una investigación es más conveniente deshacernos de estas observaciones, también es importante analizarlos cómo un conglomerado diferente para determinar su comportamiento e identificar si es posible los factores que causan dicha anomalía.

A pesar de tratarse de una técnica multivariada, en este estudio se definirá este apartado cómo el primer paso para empezar los análisis posteriores y no provocar inconvenientes.

Algoritmo para la detección de casos atípicos:

En base a (Cuadras, 2018). Sea $X = (X_1, \dots, X_p)^t$ una distribución estadística p-variante con componentes cuantitativas. Se tiene una distribución estadística unitaria p-variante; $x = \{x_i\}_{i=1}^n$, con n tamaño muestral.

Se debe:

Estandarizar los datos de forma multivariante:

$$Z = S_x^{-\frac{1}{2}}(X - \bar{x}), \tag{1-1}$$

Donde: $S_x^{-\frac{1}{2}}$ es la raíz cuadrada de la matriz inversa de la matriz de varianzas y covarianzas

- a. Tomar $J = 1$ y $z_i^{(1)} = z_i$
- b. Calcular la dirección \mathbf{d}_j con norma unidad, que maximiza el coeficiente de curtosis univariado de los datos proyectados. Llamemos $y_i^{(j)} = \mathbf{d}_j' z_i^{(j)}$ a la proyección de la observación $z_i^{(j)}$ sobre la dirección \mathbf{d}_j , donde las nuevas $y_i^{(j)}$ son de una nueva variable estadística $Y^{(j)}$ escalar unidimensional.

Proyectar los datos sobre un espacio de dimensión $p - j$ definido como el espacio ortogonal a la dirección \mathbf{d}_j . Para ello tomar $Z^{(j+1)} = Z^{(j)}(I - \mathbf{d}_j \mathbf{d}_j')$ hacer $j = j + 1$

- c. Repetir 2 y 3 hasta obtener las p -direcciones d_1, \dots, d_p .
- d. Repetir 2 y 3 pero ahora minimizando la curtosis para obtener otras p -direcciones d_{p+1}, \dots, d_{2p} .
- e. Una vez hecho esto, considerar como sospechosos aquellos puntos que en alguna de estas $2p$ -direcciones están claramente alejadas del resto, es decir, verifican $\frac{|y_i^{(j)} - \text{med}(y^{(j)})|}{\text{MEDA}(y^{(j)})} > 5$ o sea, luego de aplicarles la estandarización robusta a cada una de las $2p$ variables, son mayores a 5.

Una vez hecho esto se retiran los sospechosos y se repite el algoritmo sin ellos, es decir que la estandarización tendrá un nuevo vector de medias y matriz de covarianzas calculados sin sospechosos. Se repite el algoritmo hasta que ya no se detecten más sospechosos o se haya eliminado una proporción de datos prefijada. Una vez que la muestra esté libre de sospechosos se calcula el vector de medias \bar{x}_R y la matriz de covarianzas S_R de los datos no sospechosos, y luego la distancia de Mahalanobis de cada dato sospechoso con respecto al vector de medias \bar{x}_R :

$$d_M^2(x_i, \bar{x}_R) = (x_i - \bar{x}_R)' S_R^{-1} (x_i - \bar{x}_R) \quad (2-1)$$

Si esta distancia es mayor que $p + 3\sqrt{2p}$ entonces el dato es atípico. Si es menor el dato no es atípico y se regresa a la matriz. Finalmente se calcula el vector de medias \bar{x}_f y la matriz de covarianzas S_f a partir de los puntos no atípicos, y éstas serán las estimaciones finales obtenidas de los datos.

1.3.2. *Análisis exploratorio de datos (AED)*

El AED es necesaria en toda investigación de tipo cuantitativa debido a su aportación en la visualización de forma general el comportamiento de cada una de las variables involucradas, por esto, según (Batanero C., 1991) el AED es recomendable para conocer y observar la estructura de los datos a nivel univariante y multivariante de la información, siendo un grupo de herramientas estadísticas que recopilan, analizan e interpretan a los datos, extrayendo toda la información posible de estos, aportando a generar nuevas hipótesis sobre ellos, además que de acuerdo a (Gallardo, 2003) ayuda a determinar las posibles relaciones existentes entre las variables y los individuos de estudio.

Este análisis se basa en extraer y observar la mayor parte de información acerca del comportamiento de un conjunto de datos, además de ser un primer paso que orienta a las posteriores etapas en la ejecución de la investigación y el análisis.

Por lo general se estudia las medidas de tendencia central, medidas de variabilidad, medidas de posición y forma (De la Puente Viedma, 2018).

1.3.3. Análisis de relación o asociación entre variables

Este análisis generalmente se lo realiza mediante pares de variables, existen diferentes métodos de análisis de los cuales en este trabajo se utilizará el coeficiente de correlación de Pearson para variables cuantitativas y el test chi cuadrado de independencia para el caso de las variables cualitativas.

1.3.3.1. Coeficiente de correlación de Pearson

De acuerdo a esta media permite determinar la existencia o no de dependencia o relación lineal entre dos variables.

Está definida cómo:

$$r_{xy} = \frac{cov(x,y)}{var(x)var(y)} \quad (3-1)$$

La que cuenta con las propiedades:

- a. $0 \leq |r_{xy}| \leq 1$
- b. r_{xy} es invariante a transformaciones lineales

Mientras más cercano a 1 es el coeficiente $|r_{xy}|$, mayor asociación existe entre variables y si $|r_{xy}|$ es cercano a 0 entonces no existe relación lineal entre el par de variables.

1.3.3.2. Test chi-cuadrado de independencia.

Cuando se tiene datos con variables de tipo categórico es conveniente representarlas en una tabla de contingencia o tabla de doble entrada. Donde el principal es conocer si las variables involucradas en esta tabla son independientes o no, en otras palabras, si el valor de una variable no depende de la otra.

Una de las maneras de conocer si las variables son independientes es verificando si la frecuencia observada en la celda (i,j) es igual o muy cercano al producto de sus respectivas frecuencias marginales entre el tamaño de la muestra (frecuencia esperada en la celda).

La técnica antes mencionada es el popular test chi-cuadrado de independencia que según (Mendivelso, 2018) como su nombre lo indica sirve para someter a prueba hipótesis referidas a distribuciones de frecuencias. En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula, fue desarrollada por Karl Pearson en el año 1904.

Hipótesis:

$H_0: n_{ij} = \pi_i \pi_j$; las variables son independientes.

$H_1: n_{ij} \neq \pi_i \pi_j$; las variables están relacionadas.

Estadístico

$$\chi_0^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (4-1)$$

χ_0^2 sigue una distribución chi cuadrada con $(c - 1)(f - 1)$ grados de libertad, donde c es el número de categorías de la primera variable y f el número de categorías de la segunda variable.

Región de rechazo

$$P(\chi^2 \geq \chi_0^2) = p - \text{valor}$$

Si $p - \text{valor} > \alpha$ se rechaza H_0 , donde α es el nivel de significancia

1.3.4. Modelo de regresión lineal múltiple

Cómo (Gutierrez Pulido Humberto, 2008) menciona en la vida real se presentan situaciones donde se piensa que una variable Y está influida o relacionada por algunas variables independientes por lo que se siente la necesidad de tomarlas en cuenta para tratar de predecir o explicar el comportamiento de una variable Y . Una de las maneras sencillas de identificar esta relación es mediante una regresión lineal múltiple que está definida como: Sea X_1, X_2, \dots, X_k variables

independientes o regresores, y sea Y una variable de respuesta, entonces el modelo de regresión lineal múltiple estocástica con k variables independientes es el polinomio de primer orden:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (5-1)$$

A esta ecuación se llega mediante la disposición de una matriz de datos $X = (X_1, \dots, X_k)$ y la necesidad de obtener una ecuación en termino matricial igual a $Y = X * b + e$, donde b contiene los valores estimados de los parámetros β y e los errores.

1.3.4.1. Estimación y significancia de parámetros

La estimación de parámetros para una regresión lineal múltiple en su mayoría de veces se lo realiza a través de la estimación por mínimos cuadrados ordinarios (MCO) donde al momento de agregar variables cualitativas como variables explicativas se hace uso de las variables ficticias (Dummy).

La estimación de los parámetros de un modelo de regresión busca obtener los valores de los β_i que minimizan la suma de los errores cuadrados.

Si se tiene un $Y \cong \hat{Y}$ entonces su estimación consiste en encontrar:

$$\hat{Y} = X * b \quad (6-1)$$

donde $b = (X'X)^{-1}X'y$

Si el problema trata de usar variables dummy esta estimación sufre un leve cambio únicamente en la matriz de datos original, donde se incrementarán columnas de datos binarios de cada una de las categorías de las variables cualitativas donde tendrá 1 si la categoría se cumple y 0 si no.

La mejor estimación de los parámetros será cuando el valor de los errores sean los más pequeños posibles:

$$\min \sum e^2 = e'e \quad (7-1)$$

con $e = y - \hat{y}$

Propiedades de los estimadores

- $E(b) = \beta$ y $var(b) = \sigma^2(X'X)^{-1}$
- $E(\varepsilon) = 0$ y $var(\varepsilon) = \sigma^2I$
- $cov(\varepsilon_i, \varepsilon_j) = 0$
- $cov(\varepsilon_i, X_j) = 0$
- $cov(\varepsilon_i, \hat{y}_j) = 0$

La significancia de estos parámetros al ser estimaciones se deben verificar su significancia mediante un contraste la cual trabaja bajo la hipótesis nula: $H_0: \beta_i = 0$

Que generalmente se realiza mediante el uso de una tabla de formato ANOVA con la distribución de Fisher como sigue:

Tabla 1-1: Formato anova para el contraste de hipótesis de los parámetros estimados.

FV	SC	gl	MC	F	$F \sim F_{p, (n-p-1)}$
Regresión	$\hat{y}'\hat{y} = b'X'Xb$	p	$\hat{y}'\hat{y}/p$	MCR/MCE	
Error	$e'e$	$n-p-1$	$e'e/(n-p-1)$		
Total	$y'y$	$n-1$			

Realizado por: Urquizo, Diego, 2022.

1.3.4.2. Coeficiente de determinación

El coeficiente de determinación permite medir la variabilidad explicada por un modelo propuesto en relación a una variable respuesta. Generalmente para que un modelo sea válido el coeficiente de determinación debe ser superior a 0.70. Su cálculo se realiza mediante el estadístico R^2 :

$$R^2 = \frac{(\hat{y}'\hat{y})}{(y'y)} = 1 - \frac{e'e}{y'y} = 1 - \frac{SCE}{SCT} \quad (8-1)$$

1.3.4.3. Residuos

Es la diferencia entre los datos observados y estimados o predichos. Sirven para analizar el error de ajuste de un modelo (Gutierrez Pulido Humberto, 2008).

$$e_i = y_i - \hat{y}_i \quad (9-1)$$

1.3.4.4. Supuestos de un modelo de regresión

• Normalidad

Este supuesto se enfoca que los residuos en un modelo de regresión siguen una distribución normal, para ello existen diferentes técnicas de verificación que pueden ser mediante graficas QQ-Plot o histogramas, o a su vez mediante contrastes de hipótesis Shapiro-Wilk, Kolgorov-Smirnov, Jarque Bera entre otras, siendo la primera y la última las más usadas y confiables. Por otro lado, el no cumplimiento de este supuesto provoca graves consecuencias en el modelo propuesto debido a que anula su validez completamente:

Los contrastes de hipótesis de normalidad trabajan bajo las siguientes hipótesis:

H_0 : Los residuos se ajustan a una distribución normal.

H_1 : Los residuos no se ajustan a una distribución normal.

Estadístico de la prueba Jarque Bera

$$JB = n \left[\frac{A^2}{6} + \frac{(K - 3)^2}{24} \right] \quad (10-1)$$

Donde:

A =coeficiente de asimetría.

K =coeficiente de kurtosis.

• Homocedasticidad

Este supuesto se enfoca en que los residuos de un modelo de regresión presenten variabilidad similar en todo el conjunto de observaciones. Al igual que el supuesto de normalidad, el no cumplimiento de esta, afecta de manera negativa al modelo propuesto haciendo que pierda su validez.

Los contrastes de hipótesis de homocedasticidad trabajan bajo la hipótesis:

H_0 : Existe homocedasticidad en los residuos.

H_1 : No existe homocedasticidad en los residuos.

Para modelos de regresión lineal una opción es el test de Goldfeld-Quandt.

• Independencia

Este supuesto al igual que los anteriores se enfoca en que los residuos sean independientes uno de los otros siendo así también que el no cumplimiento de esta. Afecta de manera negativa al modelo propuesto haciendo que pierda su validez.

Los contrastes de hipótesis de Independencia trabajan bajo la hipótesis:

H_0 : Los residuos son independientes.

H_1 : Los residuos están relacionados.

Estadístico de prueba de Durbin Watson

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (11-1)$$

1.3.5. Validación del modelo de regresión.

Uno de los aspectos importantes a tomar en cuenta en la validación o no de un modelo de regresión lineal es también la identificación de la presencia o no de multicolinealidad entre las variables predictoras. Donde según (Kutner Michael, 2004) para asegurar que no existe problemas de multicolinealidad el coeficiente VIF (Factor de inflación de la Varianza) debe ser menor a 5.

1.3.6. Comparación y selección de modelos

Al momento de seleccionar y decidirse por un modelo que explique de mejor manera la variable respuesta en cuestión depende de ciertos factores los cuales según (Aparicio, 2013, p. 119) Son:

- La significatividad de los predictores
- El coeficiente de determinación ajustado
- El error residual del ajuste s^2
- El estadístico AIC (Akaike Information Criteria)

Sin embargo, en la práctica lo más común para realizar este proceso de selección se toma en base dos criterios: el AIC y el BIC.

1.3.6.1. AIC

El criterio de información de Akaike es basado en la función de verosimilitud que también incluye una penalización que aumenta dado el número de parámetros estimados en el modelo de regresión. Gana el modelo que menor estadístico tenga, el AIC está definido como:

$$AIC = -2l(\hat{\beta}) + 2p \quad (12-1)$$

1.3.6.2. BIC

Existe una versión que también considera el número de datos utilizados en el ajuste, conocido como Schwarz's Bayesian criterion, que se conoce como BIC, y se define:

$$BIC = -2l(\hat{\beta}) + \log(n)p \quad (13-1)$$

Si existen valores pequeños en ambos modelos estos reflejan mejores modelos (Montesinos, 2011, p. 3).

1.3.7. Técnicas multivariadas

El Análisis Multivariante de datos según (Cuadras, 2018) es la parte de la estadística donde se estudia, analiza, representa e interpreta los datos resultantes de la observación de $p > 1$ variables estadísticas sobre una muestra de individuos.

La información multivariante está contenida en una matriz de datos, la cual también puede estar compuesta por información acerca de distancias o similaridades entre los individuos, las distancias o similaridades miden el grado de discrepancia o diferencia entre los individuos.

En matriz de datos cada una de las p variables numéricas observadas se denominan variables univariantes y el conjunto de las p variables forman una variable vectorial o multivariante. Los valores de las p variables se representa en las columnas de una matriz \mathbf{X} de dimensiones $(n \times p)$ conocida como matriz de datos \mathbf{X} .

La matriz de datos \mathbf{X} , se representa:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (14-1)$$

Las técnicas que es de principal interés en esta investigación buscarán la existencia o no de relaciones entre las características de las variables que se posee.

1.3.7.1. Análisis de correspondencia

El análisis de correspondencia según (De la fuente Fernandez, 2011) tienen como objetivo resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible. En este sentido, el análisis de correspondencia es usado para representar datos que se presentan en forma de tablas de contingencia de dos variables de tipo cualitativa.

Además, es posible aplicar en tablas de distancia entre elementos, tablas de proximidad o tablas de preferencia.

El procedimiento consiste en: a partir de una tabla de contingencia N obtener una matriz de correspondencias $P = \frac{1}{n}N$, al cuál se define dos vectores, un vector $r_{1 \times 1}$ con los totales marginales de las filas de P y un vector $c_{1 \times 1}$ con los totales marginales de las columnas de P .

A partir de estos dos vectores se definen las matrices $D_r = \text{diag}(r)$ y $D_c = \text{diag}(c)$, el cual, aplicando la teoría de la correlación canónica, se puede considerar la descomposición singular

$$D_r^{-\frac{1}{2}}(P - rc')D_c^{-1/2} = UD_\lambda V' \quad (15-1)$$

.

Donde D_λ es la matriz diagonal con los autovalores singulares en orden decreciente.

Los vectores que cuantifican las categorías tanto de filas cómo columnas son las columnas de las matrices definidas por $A = D_r^{-1/2}U$ y $B_0 = D_c^{-1/2}$, estas matrices son utilizadas en la representación gráfica conjunta (Cuadras, 2018).

Para determinar la calidad de representación de las variables involucradas se hace uso de la llamada variabilidad geométrica o inercia donde una inercia alta significará que las frecuencias marginales o perfiles de las filas presentan grandes desviaciones del centro de gravedad y una inercia baja significará una desviación del centro de gravedad poca.

La inercia viene definida por:

$$P_m = 100 * \frac{\sum_{k=1}^m \lambda_k^2}{\sum_{k=1}^K \lambda_k^2} \quad (16-1)$$

Donde m es el número de columnas de A que se está utilizando en la representación y K es el número total de categorías de la variable representada en la tabla de contingencia como filas.

El **análisis de correspondencia múltiple** es una generalización del análisis de correspondencia simple y utiliza los mismos principios que el ACS.

Además, que de acuerdo a (Cuadras, 2018) el análisis de correspondencia múltiple es una técnica multivariante que permite representar las categorías de las variables cualitativas involucradas en un estudio, teniendo como finalidad observar y describir las relaciones existentes o no entre ellas.

Esta representación parte de una super-matriz de indicadores Z que en sus columnas contiene variables dicotómicas de cada una de las categorías de las variables cualitativas contenidas en una matriz de datos X.

$$B_u = Z'Z$$

$$F = \frac{B_u}{n} \quad (17-1)$$

La matriz de Burt B_u debido a su naturaleza de obtención es una matriz simétrica, es decir que el análisis de correspondencia aplicado a sus filas, serán lo mismo cuando se aplican a sus columnas.

La matriz F es una equivalencia a la matriz P del ACS y es la matriz el cual será sometida al AC.

Que consiste en la descomposición espectral $D_r^{-\frac{1}{2}}(P - rc')D_c^{-1/2} = UD_\lambda V'$, donde D_c y D_r coinciden, y su representación gráfica se consigue a partir de la representación de las dos primeras columnas de la matriz: $A_0 = B_0 = D_r^{-1/2}U$ de la cual entre mayor sea la inercia: $P_m =$

$100 * \frac{\sum_{k=1}^m \lambda_k^2}{\sum_{k=1}^K \lambda_k^2}$, generalmente de las dos primeras coordenadas, mejor será la calidad de representación de las categorías de las variables, se acepta una inercia igual o superior a un 70% para generalizar las conclusiones obtenidas.

1.3.7.2. *Análisis factorial de datos mixtos (AFDM)*

El nacimiento del AFDM remonta en los años 1979 y 1990 donde Brigitte Escofier sugiere la introducción de variables cuantitativas a un análisis de correspondencia múltiple (ACM) y Gilbert Saporta por otro lado sugiere la introducción de variables cualitativas a un análisis de componentes principales (ACP), donde ambas aportaciones arrojan los mismos resultados, además generando resultados de mayor facilidad de interpretación y mucho potencial a comparación de otras técnicas.

El análisis factorial de datos mixtos tal y como explica (Pagés, 2004) consiste en introducir simultáneamente variables cuantitativas y cualitativas como elementos activos de un análisis factorial común el cual tiene como objetivo encontrar relaciones entre grupos de variables dando como resultados variables latentes llamadas también factores. Permitiendo así explorar la asociación entre todas las variables, tanto cuantitativas como cualitativas.

Esta metodología transforma las variables cuantitativas en cualitativas, descompone su intervalo de variación en clases y sometiendo la tabla homogénea resultante a un análisis de ACM. Para aplicar esta técnica es recomendable tener un tamaño muestral relativamente grande, generalmente mayor a 100. En este sentido (Cuadras, 2018) define como un método multivariante que pretende expresar p variables observables como combinación lineal de m variables latentes, que se denominan factores.

Una de las ventajas de esta técnica es que permite el análisis de datos respetando su naturaleza de medida, teniendo como su fuerte el estudio de similaridad de los individuos y la relación entre variables. (Zubcoff, 2017)

Otra de las principales aportaciones de esta técnica es que estudia la similitud entre individuos de la muestra, es decir, identifica la relación existente entre variables en estudio y las similitudes existentes entre grupos de forman los factores. Además, según (Fernandez & Calvo, 2002) la función del AFDM es generar indicadores de las inercias entre individuos y posee una gama amplia de visualización gráfica de resultados.

Metodología del AFDM

A partir de la disposición de I individuos, a cada i se le asigna un peso p_i tal que $\sum p_i = 1$, donde generalmente se asume que todos los individuos tienen el mismo peso resultando así que $p_i = \frac{1}{I} \forall i$.

La descripción de individuos viene dada por:

- K_1 variables cuantitativas estandarizadas $\{k = 1, K_1\}$, estandarizadas debido a que involucran dos tipos de variables.
- Q variables cualitativas $\{q = 1, Q\}$, la q -ésima variable presenta las categorías $Kq\{kq = 1, Kq\}$, el número total de categorías es $\sum_q k_q = k_2$ y se denota con p_{kq} la proporción de individuos que poseen la categoría k_q .

De las cuales el número total de variables viene definido como $K = K_1 + K_2$.

Representación de variables.

La representación gráfica es una de las cosas de mayor interés cuando se realiza un análisis factorial, ya que con la ayuda de esta se podrá observar de manera más clara las interdependencias existentes entre las variables en estudio.

Esta representación consiste en: se establece R^I como un espacio generado en función de I , donde este espacio esta denotado como D y es la matriz diagonal con los pesos de los individuos.

$$D(i, j) = \begin{cases} 0 & \text{si } i \neq j \\ p_i & \text{si } i = j \end{cases} \quad (18-1)$$

Y por lo anterior dicho los pesos suelen ser constantes por lo tanto la construcción matricial está definida cómo:

$$D = \left(\frac{1}{I}\right) I_d \quad (19-1)$$

Con $I = \text{número de individuos}$ y I_d matriz identidad de orden $d = i$.

Las variables cuantitativas al igual que en un ACP se encuentran representadas por un vector, mientras que las variables cualitativas son similares al de un ACM, donde la nube N_q representa las q variables cualitativas, dicha nube genera subespacios E_q que es el conjunto de funciones centradas constantes en las clases de la partición definida por q con dimensiones $K_q - 1$.

Inercia (Variabilidad explicada) y calidad de representación

El AFDM debido a que es un tema extenso y la investigación no se enfoca a la explicación de esta metodología solo se procederá a realizarlo haciendo uso del lenguaje de programación R.

Sin embargo, la obtención de la **inercia para determinar la calidad de representación** se enfoca en una de las propiedades del ACM, donde la inercia proyectada de N_q sobre una variable centrada es igual a la razón de correlación cuadrada entre q e y , denotado como $\eta^2(q, y)$. Donde a partir de esto se busca la dirección v de R^I que maximice la inercia proyectada por la nube N_k , el criterio a maximizar es:

$$\sum_{k \in K_1} r^2(k, v) + \sum_{q \in Q} \eta^2(q, v) \quad (20-1)$$

Con el conocimiento de que las k variables cuantitativas están estandarizadas, la proyección de la variable k en v es igual a $\cos(\theta_{kv}) = r(k, v)$ donde θ_{kv} es el ángulo entre los vectores k y v , de igual manera en variables cualitativas al saber que v está centrada entonces $\eta^2(q, v) = \cos^2(\theta_{qv})$, donde θ_{qv} es el ángulo entre v y su proyección en E_q . Dando como resultado el criterio:

$$\sum_{k \in K_1} \cos^2(\theta_{kv}) + \sum_{q \in Q} \cos^2(\theta_{qv}) \quad (21-1)$$

La calidad de representación producida por cada una de las dimensiones en términos generales se mide mediante el \cos^2 , para una variable dada la suma del \cos^2 en todas las dimensiones es igual

a uno, si una variable está bien representada solo por dos dimensiones la suma del \cos^2 de estas dos dimensiones serán cercano a uno. Por otro lado, la contribución de cada una de las variables se mide en porcentajes tomando en cuenta los valores propios de cada una de las dimensiones que se consideren y sus respectivas contribuciones (generalmente $1/I$) (Pagés, 2004). Donde es deseable que este criterio sea lo más cercano a 1 posible para decir que las variables tienen una buena calidad de representación.

Implementación en R.

En R los resultados de interés se los puede obtener con el uso del paquete FactorMineR donde:

La inercia o varianza total explicada por cada una de las dimensiones se identifica en la observación un gráfico de sedimentación, que en el eje de las abscisas contiene las dimensiones y el eje de las ordenadas representa el porcentaje de varianza explicada por cada una de las mismas.

La calidad de representación se determina en los gráficos de barra de los \cos^2 , donde si la barra es más cercana a 1, su calidad de representación es buena.

R facilita además la visualización de las contribuciones de cada una de las variables en las 2 dimensiones del plano, lo que ayuda a realizar un segundo AFDM, donde se use únicamente las variables que más contribuyen y conseguir resultados más eficientes.

Finalmente, la representación de las variables se los realiza en el plano factorial lo que ayudara a determinar relaciones existentes entre las variables en estudio.

CAPITULO II

2. MARCO METODOLÓGICO

2.1. Tipo de la investigación.

De acuerdo a la intención de la investigación y la naturaleza de los objetivos formulados, el presente trabajo reúne las condiciones que la realizan una investigación mixta ya que las variables en estudio son cuantitativas y cualitativas, según el objetivo es aplicada, en razón que para su desarrollo teórico conceptual se apoyará en conocimientos previos del International Potato Center, según el nivel de profundización en el objeto de estudio en primera instancia será descriptiva, luego explicativa y finalmente correlacionada de acuerdo a la finalidad de misma, según la manipulación de variables no experimental por que la matriz de información provienen de una fuente secundaria y los datos usados no serán manipulados ni sufrirán cambios, según el tipo de inferencia será inductiva ya que se pretende conocer los factores influyentes en la producción de papa a través de la información centrada en el ESPAC Ecuador 2020, según el periodo temporal transversal ya que la información está tomada dentro de un periodo de tiempo que es 2020.

2.2. Diseño de la investigación no experimental.

La presenta investigación debido a la naturaleza de sus objetivos formulados, reúne condiciones que la realizan una investigación mixta ya que las variables en estudio son cuantitativas y cualitativas, según la manipulación de variables será una investigación no experimental porque las matrices de información provienen de una fuente secundaria y los datos usados no serán manipulados ni sufrirán cambios.

2.2.1. Localización de estudio



Ilustración 1-2: Localización del estudio.

Fuente: (Wikimedia, 2008)

La investigación se realizará en Ecuador, en especial en las provincias de la región sierra.

2.2.2. Población de estudio

La población de investigación de la ESPAC 2020 tal y como está descrita en (INEC, 2020) corresponde a todos aquellos terrenos con explotación agropecuaria en toda la superficie continental del Ecuador.

2.2.3. Método de muestreo

El INEC a partir del 2002 para llevar a cabo la ESPAC emplea la metodología de muestreo de marcos múltiples (MMM) la cual resulta de la combinación entre muestreo de marco de áreas (MMA) y muestreo con el marco de lista (MML), esto con el fin de seleccionar unidades de

investigación para recolectar información de las unidades de producción agropecuaria (UPA). (INEC, 2015)

El muestreo de marco de áreas (MMA) es un procedimiento estadístico que consiste en dividir la superficie total de un país en grandes áreas sin superposición ni omisión denominadas Unidades Primarias de Muestreo (UPM). Por otro lado, el muestreo de marco de lista investiga las unidades de producción agropecuarias (UPAS) grandes: más de 100 Ha, las sectoriales y especiales, previamente separados de la población objetivo de investigación (INEC, 2013), de estos dos tipos de muestreo según (Guamán Daquilema & Mullo Guaminga, 2018) el primer método realiza la segmentación de la superficie total del estado por estratos en base a la intensidad productiva de cada área del país y el segundo método sirve para constatar las principales explotaciones dedicadas a determinados cultivos, los cuales son tomados en cuenta para de tal forma mejorar la calidad de estimaciones estadísticas.

2.2.4. *Tamaño de la muestra*

La muestra está conformada por los terrenos que participaron en la ESPAC y que a su vez produjeron papa en el año 2020 dando como resultado 1243 casos de investigación, sin tomar en cuenta el factor de expansión. Donde los puntos principales de recolección de información fueron: Tungurahua con un 22.7% de casos de terrenos que participaron en la encuesta, Chimborazo con un 19.7%, Cotopaxi con un 15%, Carchi con un 9.3%, Azuay con un 8.7% y Pichincha con un 8.1%. (INEC, 2020)

2.2.5. *Técnica de recolección de datos.*

La técnica de recolección de datos realizado por el INEC es mediante encuestadores de campo, donde esta información es ingresada mediante un aplicativo de ingreso de datos ESPAC, el cual es un sistema que permite obtener una base de datos por capítulos debidamente estructurada permitiendo la verificación de información, revisión de rendimientos por tipos de cultivos y densidad de siembra y validaciones pertinentes a cada capítulo, dando como resultado archivos en Excel o SPSS (INEC, 2020).

Para esta investigación la información proviene de una fuente secundaria de acceso público específicamente de los datos recolectados en la encuesta ESPAC año 2020 disponible en la página web sobre estadísticas agropecuarias del INEC.

2.2.6. Identificación de variables

Para realizar esta investigación se usó la base de datos ESPAC 2020, en concreto la sección de cultivos transitorios y dentro de ella a los individuos (terrenos) productores de papa que participaron en dicha encuesta, se tomaron en cuenta 13 variables, 10 variables cualitativas y 3 variables cuantitativas.

Las cuales se nombran a continuación:

- **Cuantitativas**

1. Superficie sembrada.
2. Cantidad cosechada.

- **Cualitativas**

1. Semilla Usada
2. Afectación en la producción.
3. Preparación de suelo.
4. Uso de riego.
5. Uso de fertilizantes.
6. Uso de fertilizante orgánico.
7. Uso de fertilizantes químico.
8. Uso de Fitosanitario.
9. Plaguicida orgánico.
10. Plaguicida químico.
11. Mes de siembra.

2.2.7. Modelo estadístico

Los modelos estadísticos a desarrollarse según la información poseída serán técnicas multivariantes. Además, se propondrá un modelo de regresión para estimar la producción de papa en toneladas métricas y para validar el modelo ejecutado, se probará supuestos de normalidad (Jarque Bera) Homocedasticidad (Bartlett) e independencia (Durbin Watson)

2.3. Variables en estudio.

2.3.1. Operacionalización de variables.

Tabla 1-2: Operacionalización de variables

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
ct_k510ha	Superficie Sembrada.	Cuantitativa	Razón	$[0, +\infty[$
ct_prod	Producción en Toneladas métricas.	Cuantitativa	Razón	$[0, +\infty[$
ct_k506	Tipo de semilla usada.	Cualitativa	Nominal	<ul style="list-style-type: none"> • Común • Mejorada • Certificada • Híbrida nacional • Híbrida internacional
ct_afecta_prod	Afectación en la producción.	Cualitativa	Nominal	<ul style="list-style-type: none"> • Sequia • Plagas • Inundación • Semilla • Practicas inadecuadas • Edad de la plantación
ct_presueuti	Preparación de suelo	Cualitativa	Nominal	<ul style="list-style-type: none"> • Tractor • Yunta • Motocultor • Azadón
ct_riego	Uso de riego.	Cualitativa	Nominal	<ul style="list-style-type: none"> • Si • No

ct_aferti	Uso de fertilizante	Cualitativa	Nominal	•Si •No
ct_forg	Uso de fertilizante orgánico.	Cualitativa	Nominal	•Si •No
ct_fqui	Uso de fertilizante químico	Cualitativa	Nominal	•Si •No
ct_afito	Uso de fitosanitario	Cualitativa	Nominal	•Si •No
ct_porg	Uso de plaguicida orgánico	Cualitativa	Nominal	•Si •No
ct_pqui	Uso de plaguicida químico	Cualitativa	Nominal	•Si •No
ct_messiem_int	Mes de siembra	Cualitativa	Nominal	•Meses del año

Realizado por: Urquizo, Diego, 2022.

CAPÍTULO III

3. MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS.

3.1. Detección de datos atípicos.

Después de la ejecución del algoritmo en el software matlab para la detección de datos atípicos en la base de datos que se posee, se verificó que ningún caso sospechoso es una observación atípica, por lo que los resultados, análisis y conclusiones que siguen son totalmente confiables.

3.2. Análisis exploratorio de datos (AED)

Tabla 1-3: Distribución estadística de frecuencia de la variable superficie sembrada(ct_k510ha)

Límite Inferior	Límite Superior	Marca de clase	n_i	N_i	f_i	F_i
0.01	8.01	4.01	1217	1243	97.91	97.91
9.01	17	13.01	19	1243	1.53	99.44
18	26	22	2	1243	0.16	99.6
27	35	31	2	1243	0.16	99.76
54	62	58	1	1243	0.08	99.84
72	80	76	1	1243	0.08	99.92
99	107	103	1	1243	0.08	100

Realizado por: Urquizo, Diego, 2022.

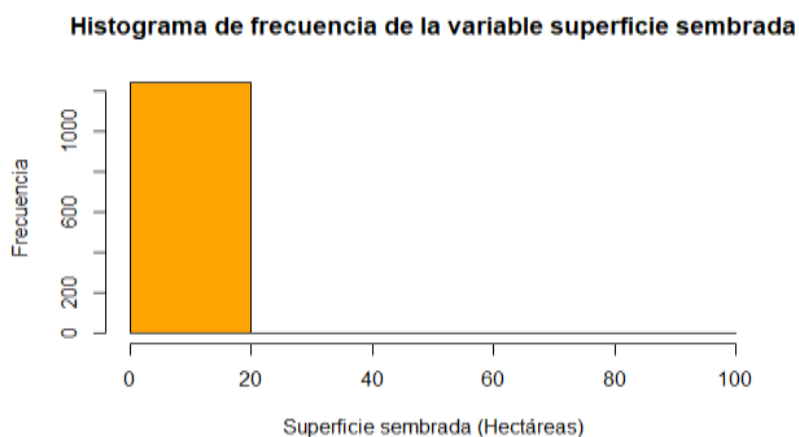


Ilustración 1-3: Histograma de frecuencia de la variable superficie sembrada.

Realizado por: Urquizo, Diego, 2022.

En el 97.91% de los casos la superficie sembrada está entre 0.01 y 8.01 hectáreas, mientras que en el 1.53% de los casos la superficie sembrada está entre 9.01 y 17 hectáreas.

Tabla 2-3: Distribución estadística de frecuencia de la variable producción de papa (ct_prod)

Límite Inferior	Limite Superior	Marca de clase	n_i	N_i	f_i (%)	F_i (%)
0.01	159	79.50	1215	1243	97.75	97.75
160	319	239.50	18	1243	1.45	99.2
320	479	399.50	5	1243	0.4	99.6
480	639	559.50	1	1243	0.08	99.68
960	1119	1039.50	2	1243	0.16	99.84
1120	1279	1199.50	1	1243	0.08	99.92
1760	1919	1839.50	1	1243	0.08	100

Realizado por: Urquizo, Diego, 2022.

Histograma de frecuencia de la variable producción de papa

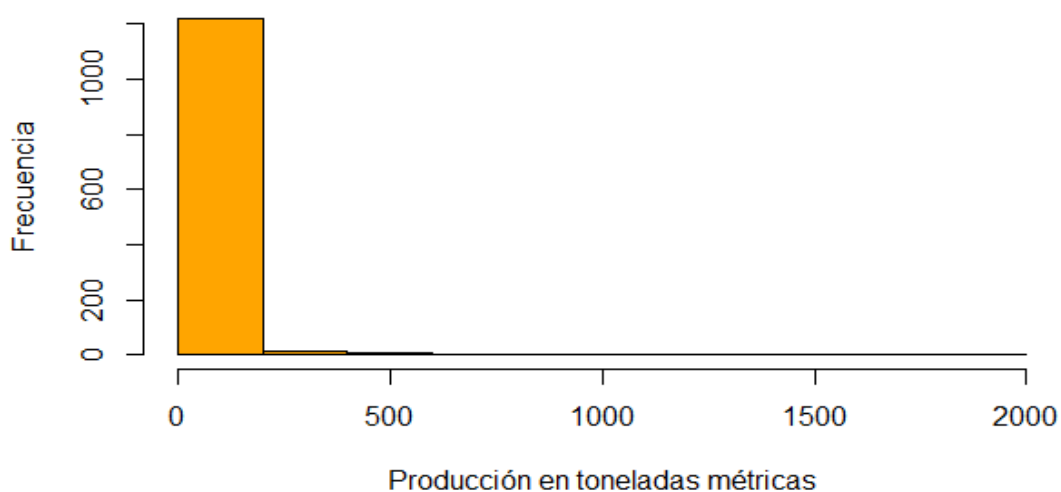


Ilustración 2-3: Histograma de frecuencia de la variable producción de papa.

Realizado por: Urquizo, Diego, 2022.

Según datos de la ESPAC 2020, el 97.98% de los terrenos produjeron entre 0.01 y 159 toneladas métricas de papa, por otro lado, el 1.45% de terrenos produjeron entre 160 y 319 toneladas métricas.

Tabla 3-3: Resumen estadístico de las variables cuantitativas.

Medidas		Superficie sembrada (ct_k510ha)	Producción (ct_prod)
Medidas de tendencia central	Media	1.01	17.27
	Mediana	0.25	1.45
	Moda	0.5	0.45
Medidas de posición	25%	0.1	0.45
	50%	0.25	1.45
	75%	0.5	5.45
Medidas de dispersión	Varianza	20.65	8083.51
	Desviación estándar	4.54	89.91
	Coefficiente de variación	4.52	5.2
Medidas de forma	Coefficiente de asimetría	14.62	13.3
	Coefficiente de curtosis	265.28	222.34

Realizado por: Urquizo, Diego, 2022.

De acuerdo a la Tabla 3-3, se observa que las dos variables cuantitativas involucradas en la investigación siguen una distribución con un coeficiente de asimetría positiva, lo que indica, que la mayor proporción de datos se encuentran hacia la izquierda de su media. En cuanto a la curtosis también se determina que las variables siguen una distribución leptocúrtica.

Observando además las medidas de dispersión de ambas variables, se puede indicar que la variable con mayor dispersión es la producción con un coeficiente de variación de 5.2% en comparación a la variable superficie sembrada que tiene un coeficiente de variación de 4.52.

Tabla 4-3: Resumen de la distribución estadística de frecuencias para las variables cualitativas.

Variable cualitativa	Categoría	n _i	f _i (%)
Tipo de semilla usada (ct_k506)	Común	1098	88.33
	Mejorada	131	10.54
	Certificada	13	1.05
	Híbrida nacional	1	0.08
	Híbrida internacional	0	0.00

Afectación en la producción (ct_afecta_prod)	Sequía	262	21.08
	Plagas	319	25.66
	Inundación	8	0.64
	Semilla	2	0.16
	Practicadas inadecuadas	39	3.14
	Edad de la plantación	613	49.32
Preparación de suelo (ct_presueuti)	Tractor	741	59.61
	Yunta	7	0.56
	Motocultor	133	10.70
	Azadón	323	25.99
	Ninguno	39	3.14
Uso de riego (ct_riego)	Si	527	42.40
	No	716	57.60
Uso de fertilizante (ct_aferti)	Si	1124	90.43
	No	119	9.57
Uso de fertilizante orgánico (ct_forg)	Si	666	53.58
	No	577	46.42
Uso de fertilizante químico (ct_fqui)	Si	861	69.27
	No	382	30.73
Uso de fitosanitario (ct_afito)	Si	899	72.33
	No	344	27.67
Uso de plaguicidas orgánicos (ct_porg)	Si	20	1.61
	No	1223	98.39
Uso de plaguicidas químicos (ct_pqui)	Si	894	71.92
	No	349	28.08
Mes de Siembra (ct_messiem_int)	Enero	213	17.14
	Febrero	146	10.75
	Marzo	144	10.58
	Abril	126	10.14
	Mayo	92	7.40
	Junio	61	4.91
	Julio	54	4.34
	Agosto	162	13.03
	Septiembre	58	4.67
Octubre	40	3.22	

Noviembre	33	2.65
Diciembre	114	11.17

Realizado por: Urquiza, Diego, 2022.

Basándose en la tabla 3-4, se observa que la semilla de mayor uso en la siembra de papa es la común en un 88.33% de los casos seguido por la semilla mejorada con un 10.54%, la principal afectación de la producción que generaría pérdidas en la producción es la edad de plantación con un 49.32%, seguidos por las plagas (25.66%) y sequías (21,08%), la preparación del suelo para la siembra el 59.61% lo realiza con tractor seguido por un 25.99% que lo hace con azadón, la mayoría de los terrenos (57.6%) no hacen uso del riego, por otro lado el 90.43% de los sembríos hacen uso de fertilizantes, siendo así también que el 53.58% usa fertilizantes orgánicos y el 69.27% también usan fertilizantes químicos, el 72.33% de los sembríos hacen uso de productos de fitosanitario, en relación a los plaguicidas, el 98.39% de los sembríos no están haciendo uso de los productos orgánicos y el 71.92% usan los plaguicidas químicos. Los meses de siembra recomendados por el International Potato Center son agosto, diciembre y enero, donde según la información que se posee se verifica que las personas que siembran la papa si toman en cuenta esta recomendación, teniendo así que el 17.14% de los terrenos fueron sembrado en el mes de enero, seguido por los meses agosto (13.03%) y diciembre (11.17%).

3.3. Análisis de relación o asociación entre variables.

3.3.1. Correlación de Pearson para variables cuantitativas

Tabla 5-3: Coeficiente de correlación de Pearson entre las variables cuantitativas.

Variable	Producción (ct_prod)	Superficie Sembrada (ct_k510ha)
Producción (ct_prod)	1.00	0.86
Superficie Sembrada (ct_k510ha)	0.86	1.00

Realizado por: Urquiza, Diego, 2022.

En base a la tabla 3-5, se tiene que el coeficiente de correlación entre las variables producción y superficie sembrada es positiva alta. Por lo que se puede decir que estas dos variables si están relacionadas.

3.3.2. Contraste de independencia para variables cualitativas

Para determinar la existencia o no de relación entre variables cualitativas se realiza el contraste de hipótesis chi-cuadrado de independencia. La cual está definida bajo la hipótesis nula:

H_0 : Las variables son independientes.

H_1 : Las variables están relacionadas.

Tabla 6-3: Valores-p resultantes del contraste chi-cuadrado de independencia.

Variables	ct_k506	ct_afecta_prod	ct_presueuti	ct_riego	ct_aferti	ct_forg
ct_k506	0.00	-	-	-	-	-
ct_afecta_prod	0.07	0.00	-	-	-	-
ct_presueuti	0.17	0.01	0.00	-	-	-
ct_riego	0.15	0.03	0.00	0.00	-	-
ct_aferti	0.68	0.00	0.00	0.00	0.00	-
ct_forg	0.00	0.02	0.00	0.00	0.00	0.00
ct_fqui	0.00	0.00	0.00	0.75	0.00	0.00
ct_afito	0.01	0.00	0.00	0.23	0.00	0.02
ct_porg	0.00	0.79	0.81	0.02	1.00	0.21
ct_pqui	0.01	0.00	0.00	0.34	0.00	0.03
ct_messiem_int	0.00	0.01	0.00	0.02	0.53	0.00

Realizado por: Urquizo, Diego, 2022.

Tabla 7-3: Continuación de la Tabla 3-6.

Variables	ct_fqui	ct_afito	ct_porg	ct_pqui	ct_messiem_int
ct_k506	-	-	-	-	-
ct_afecta_prod	-	-	-	-	-
ct_presueuti	-	-	-	-	-
ct_riego	-	-	-	-	-
ct_aferti	-	-	-	-	-
ct_forg	-	-	-	-	-
ct_fqui	0.00	-	-	-	-
ct_afito	0.00	0.00	-	-	-
ct_porg	0.20	0.01	0.00	-	-
ct_pqui	0.00	0.00	0.12	0.00	-
ct_messiem_int	0.02	0.00	0.03	0.00	0.00

Realizado por: Urquizo, Diego, 2022.

De acuerdo a las tablas 3-6 y 3-7 podemos observar las asociaciones entre variables de las cuales resaltan que el tipo de semilla sembrada está relacionada con el uso de fertilizante, plaguicida, tanto químico como orgánico y el uso de fitosanitario, también se nota que la variable afectación de producto tiene una fuerte relación con el tipo de preparación de suelo, uso de fertilizantes, plaguicidas (químico) y mes de siembra, finalmente, el tipo de preparación de suelo se observa que está ligado al uso de riego.

3.4. Modelo de regresión lineal múltiple

La estimación de un primer modelo se realizará utilizando todas las variables involucradas en el estudio para posteriormente depurarlo haciendo uso solo de las variables que resultan ser significativas, para incluir variables cualitativas al modelo se usará variables ficticias llamadas también variables Dummy.

Tabla 8-3: Resumen del modelo de regresión completo “Modelo 1”, variable respuesta producción.

Variables	Coefficiente de regresión	Error estándar	t value	Valor-p
Intercepto	152.715	15.851	9.635	0.000
ct_k510ha (Superficie sembrada)	12.183	0.426	28.596	0.000
Semilla (Mejorada)	19.626	6.322	3.104	0.002
Semilla (Certificada)	-19.730	18.107	-1.090	0.276
Semilla (Híbrida Nacional)	-12.009	63.425	-0.189	0.850
Semilla (Híbrida Internacional)	2.441	5.410	0.451	0.652
Afec_prod(Plagas)	0.196	22.991	0.009	0.993
Afec_prod(Inundación)	2.825	45.258	0.062	0.950
Afec_prod(Semilla)	-22.004	11.140	-1.975	0.048
Afec_prod(Practicas inadecuadas)	4.407	4.759	0.926	0.355
Afec_prod(Edad de la plantación)	-10.430	24.235	-0.430	0.667
Pre_suelo(Yunta)	-9.591	6.153	-1.559	0.119
Pre_suelo(Motocultor)	-7.059	4.615	-1.530	0.126
Pre_suelo(Azadon)	-14.708	11.043	-1.332	0.183
Uso_riego(No)	-15.720	8.762	-1.794	0.073

Uso_fertilizante(No)	14.759	4.545	3.247	0.001
Uso_fert_org(No)	6.032	5.579	1.081	0.280
Uso_fert_qui(No)	57.019	29.301	1.946	0.052
Uso_fitosanitario(No)	-147.123	15.205	-9.676	0.000
Uso_plag_org(No)	-56.396	29.104	-1.938	0.053
Uso_plag_qui(No)	-13.673	6.835	-2.000	0.046
Mes_siembra(febrero)	-11.703	6.897	-1.697	0.090
Mes_siembra(marzo)	-13.178	7.157	-1.841	0.066
Mes_siembra(abril)	-26.305	7.937	-3.314	0.001
Mes_siembra(abril)	-10.754	9.238	-1.164	0.245
Mes_siembra(mayo)	-14.036	9.723	-1.444	0.149
Mes_siembra(junio)	-17.302	6.775	-2.554	0.011
Mes_siembra(julio)	-14.604	9.526	-1.533	0.126
Mes_siembra(agosto)	-19.527	11.063	-1.765	0.078
Mes_siembra(noviembre)	-11.870	12.023	-0.987	0.324
Mes_siembra(diciembre)	-0.738	7.724	-0.096	0.924
Observaciones				1243
R_cuadrado				0.78
R_cuadrado_ajustado				0.73
Error Std. Residual			63.11 (df=1211)	
Estadístico F-Fisher			42.24 (df=(31, 1211))	
Valor-p				<2.2e-16

Realizado por: Urquizo, Diego, 2022.

Cómo se puede observar en la tabla 3-8, solo algunas de las variables predictoras son significativas al momento de predecir la producción de papa, por lo que, para obtener un modelo, retiramos las variables que no son significativas, quedándonos así un modelo como sigue:

Tabla 9-3: Resumen del modelo de regresión “Modelo 2” con variables significativas del “Modelo 1”.

Variab	C	E	t	V
Intercepto	131.332	14.599	8.996	0.000
Superficie sembrada	12.332	0.420	29.396	0.000
Semilla (Mejorada)	20.812	6.183	3.366	0.001
Afec_prod(Practicas Inadecuadas)	-28.053	10.328	-2.716	0.007
Uso_fert_org (No)	10.117	3.671	2.756	0.006

Uso_plag_org (No)	-137.162	14.508	-9.454	0.000
Mes_siembra (enero)	12.936	4.884	2.649	0.008
Mes_siembra (mayo)	-13.225	6.941	-1.905	0.049
Mes_siembra (diciembre)	13.652	6.598	2.069	0.039
Observaciones				1243
R_cuadrado				0.75
R_cuadrado_ajustado				0.71
Error Std. Residual			62.89 (df=1234)	
Estadístico F-Fisher			163.1 (df=(8, 1234))	
Valor-p				<2.2e-16

Realizado por: Urquizo, Diego, 2022.

3.4.1. Selección del modelo

La selección del modelo se realiza tomando en base los coeficientes AIC y BIC de la tabla

Tabla 10-3: Valores AIC y BIC de los modelos estimados.

	Modelo 1	Modelo 2
AIC	13865.31	13833.91
BIC	14034.45	13885.17

Realizado por: Urquizo, Diego, 2022.

Cómo se observa, el modelo con menor número de variables predictoras “Modelo 2” tiene menor valor tanto en el coeficiente AIC cómo en el BIC, por lo que determinamos que este es el mejor modelo.

Ecuación del modelo

$$\begin{aligned}
 \text{Producción} = & 131.332 + 12.332(\text{Superficie Sembrada}) \\
 & + 20.812(\text{Semilla usada} = \text{Mejorada}) \\
 & - 28.053(\text{Afectación en la producción} = \text{practicadas inadecuadas}) \\
 & + 10.117(\text{fertilizante organico} = \text{No}) \\
 & - 137.16(\text{plaguicida organico} = \text{No}) + 12.936(\text{Mes siembra} = \text{enero}) \\
 & - 13.225(\text{Mes siembra} = \text{mayo}) + 13.652(\text{Mes siembra} = \text{diciembre})
 \end{aligned}$$

Interpretación:

De acuerdo al modelo mejorado se puede observar que los factores que influyen en la producción de papa son la superficie de terreno que se siembra, el tipo de semilla que se usa, las practicas del manejo de cultivo, el uso de fertilizante, plaguicidas y el mes de siembra.

Además, con un coeficiente de determinación (0.75) lo cual es mayor a 0.7 se puede decir que el modelo si explica de buena manera la producción de papa en toneladas métricas.

Donde, al incrementar 1 hectárea en la superficie sembrada se espera incrementar la producción en 12.33 Tm, usando una semilla mejorada se espera aumentar la producción también en 20.81 Tm, al realizar prácticas inadecuadas de cultivo produce un decremento esperado en la producción de 20.053 Tm de producción, al no usar fertilizante orgánico que es lo mismo decir usar fertilizante químico se espera incrementar la producción en 10.11 Tm, por otro lado, al no usar plaguicidas tanto orgánico se espera un gran decremento de 137.16 Tm, finalmente, en cuanto al mes de siembra según el modelo es recomendable realizar entre los meses de diciembre y enero ya que la producción aumentará en 12.93 y 13.65 Tm respectivamente, y si se realiza en mayo se espera tener un decremento en la producción de 13.22 Tm.

3.4.1. Validación del modelo

La validación del modelo se realizará al modelo 2.

3.4.1.1. Multicolinealidad-Factor de inflación de varianza (VIF)

Tabla 11-3: Factores de inflación de la varianza (VIF) del Modelo 2

Variables	Vif
Superficie sembrada	1.142
Semilla (Mejorada)	1.133
Afec_prod(Practicadas Inadecuadas)	1.019
Uso_fert_org(No)	1.053
Uso_plag_org(No)	1.047
Mes_siembra(enero)	1.065
Mes_siembra(mayo)	1.038
Mes_siembra(diciembre)	1.140

Realizado por: Urquizo, Diego, 2022.

Según la tabla 3-11, al observar los valores del VIF se puede determinar que en el modelo no existe problemas de multicolinealidad.

3.4.1.2. Supuestos del modelo

Normalidad

Se observa en el gráfico 3-3 que los residuos se aproximan de buena manera a la recta, además, de acuerdo al test de Jarque Bera al tener un valor-p de 0.12 se puede indicar que los residuos si se aproximan a una distribución normal.

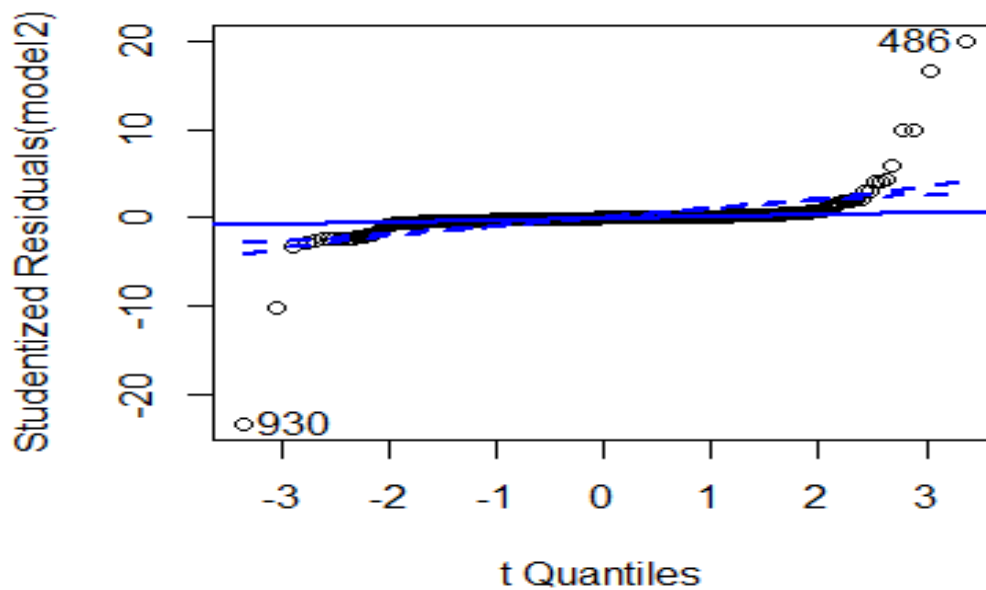


Ilustración 3-3: QQ-plot para analizar la normalidad de los residuos del “modelo 2”.

Realizado por: Urquiza, Diego, 2022.

Homocedasticidad

De acuerdo al gráfico 3-4 se observa si existe homocedasticidad en los residuos, además, según el contraste de Goldfeld-Quandt se afirma que los residuos no tienen problemas de heterocedasticidad con un valor-p igual a 0.99.

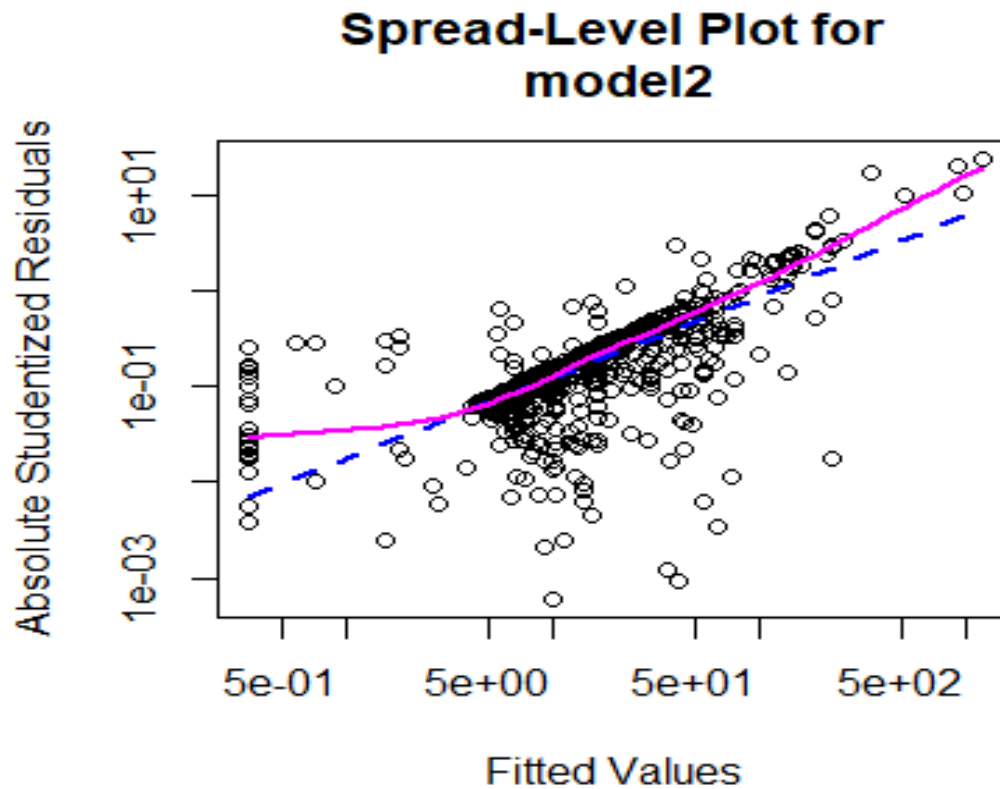


Ilustración 4-3: Valores absolutos de los residuos estandarizados contra los valores ajustados.

Realizado por: Urquizo, Diego, 2022.

Independencia

El valor-p del estadístico del test de Durbin-Watson en base a los residuos del “modelo 2” es de 0.21, por lo tanto, no se rechaza la hipótesis nula y se puede decir que los residuos no están autocorrelacionados, es decir, cumplen el supuesto de la independencia.

En cuanto a la validación del modelo, a partir del conocimiento de que cumple con los tres supuestos indispensables de un modelo de regresión, además al no tener problemas de multicolinealidad y posee un coeficiente de determinación superior al 0.7, podemos decir que el modelo es el adecuado para estimar la producción de papa en el Ecuador, sumando también que los coeficientes de regresión significativos encontrados en la elaboración del modelo consecuentemente son también factores que influyen en la producción.

3.5. Técnicas multivariantes

3.5.1. Análisis de correspondencia

Mediante el análisis de correspondencia, específicamente en su representación gráfica podemos establecer u observar diferentes relaciones entre los niveles de cada una de las variables categóricas involucradas en el estudio, los cuales se detallan a continuación.

Tabla 12-3: Coordenadas de las 2 primeras dimensiones para la representación gráfica del análisis de correspondencia

Niveles de cada Variable	Dimensión 1 (61.3%)	Dimensión 2 (8.7%)
Común	-0.152	0.372
Mejorada	0.985	-3.237
Certificada	2.671	1.721
Hibrida_nacional	2.938	-7.285
Hibrida_internacional	0.701	-0.358
Tractor	-0.013	-2.903
Yunta	-0.654	0.747
Motocultor	-1.030	1.044
Azadon	-2.564	-3.874
Riego_si	0.122	0.779
Riego_no	-0.089	-0.574
Fert_org_si	-0.392	1.506
Fert_org_no	0.452	-1.739
Fert_qui_si	0.873	-0.123
Fert_qui_no	-1.968	0.278
Fitosaitario_si	1.043	0.316
Fitosaitario_no	-2.725	-0.827
Pla_org_si	1.589	4.888
Pla_org_no	-0.026	-0.080
Pla_qui_si	1.050	0.312
Pla_qui_no	-2.691	-0.799
Ene	0.227	1.177
Feb	0.154	1.123
Mar	-0.295	1.072
Abr	0.016	1.407

May	-0.420	0.699
Jun	-0.437	-0.375
Jul	-0.781	-0.368
Ago	0.724	-1.480
Sep	-0.461	-1.384
Oct	-1.412	0.336
Nov	-1.694	0.480
Dic	0.868	-4.186

Realizado por: Urquizo, Diego, 2022.

De acuerdo a la tabla 3-12 se sabe que la primera dimensión explica un 61.3% de la variabilidad total mientras que la segunda dimensión explica un 8.3%, sumado da un total del 70%, lo que indica que su representación gráfica en 2 dimensiones es totalmente confiable para observar el comportamiento de las variables involucradas.

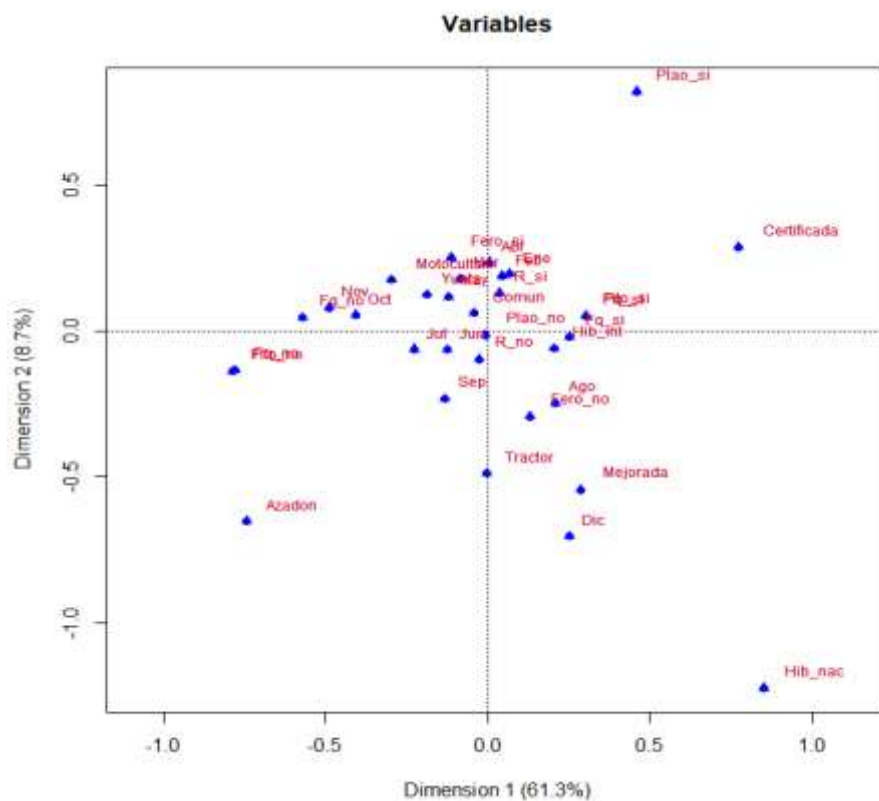


Ilustración 5-3: Representación gráfica del análisis de correspondencia de variables involucradas en el estudio.

Realizado por: Urquizo, Diego, 2022.

Basándose en la tabla 3-12 y observando la gráfica 3-5 se determina que: Los terrenos donde el sembrío de papas se hizo en primeros 5 meses del año en una alta frecuencia tienen agua para el riego, además, se usaron el tipo de semilla común, se aplicaron fertilizantes orgánicos y la preparación del suelo se lo realizó mediante la yunta. En cuanto a sembríos donde se usó semilla híbrida internacional se observa que hubo una alta frecuencia de terrenos donde se usó fertilizante químico y plaguicida químico.

3.5.2. Análisis factorial de datos mixtos AFDM

3.5.2.1. Primer análisis factorial de datos mixtos.

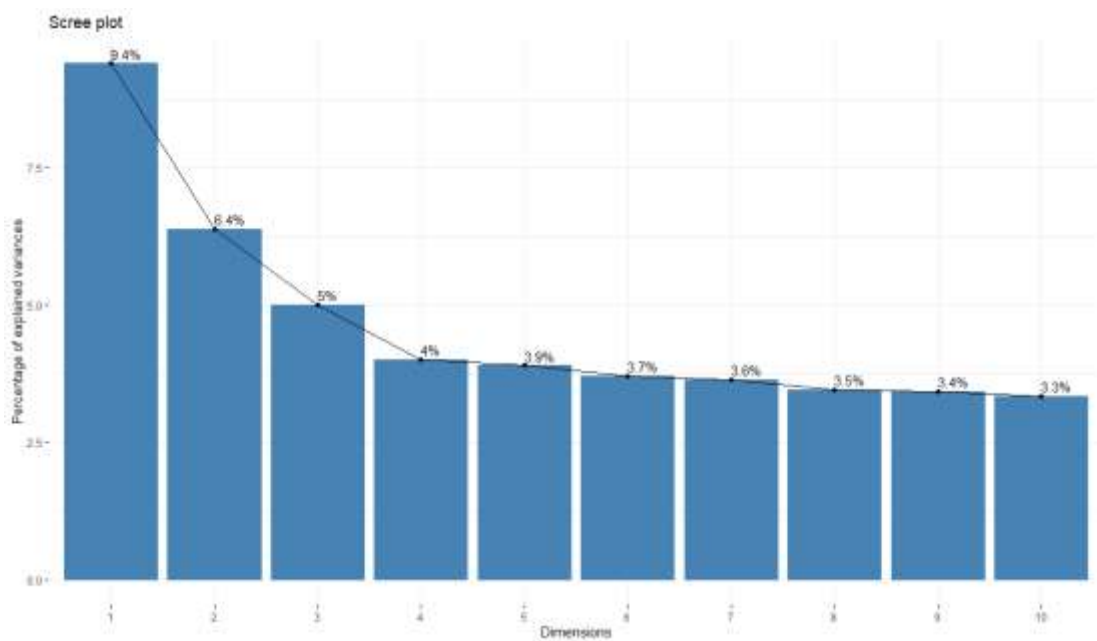


Ilustración 6-3: Grafico de sedimentación del primer AFDM.

Realizado por: Urquiza, Diego, 2022.

Según el grafico 3-6 la primera dimensión explica el 9.4% de la variabilidad total, la segunda dimensión explica el 6.4% sumando así tan solo el 15.8% lo cual es muy bajo.

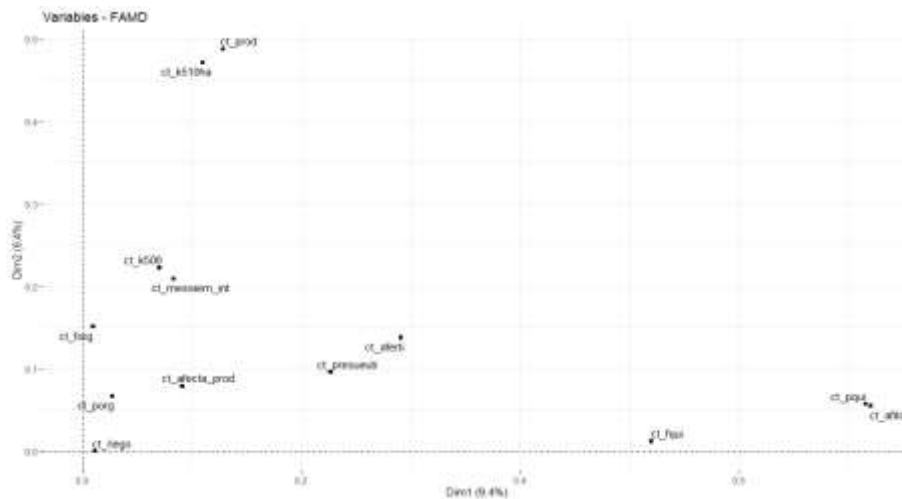


Ilustración 7-3: Representación de variables con el primer AFDM.

Realizado por: Urquizo, Diego, 2022.

El grafico 3-7 muestra que las variables uso de fertilizantes químicos, uso de plaguicidas químicos, uso de fitosanitario y uso de fertilizante tienen mayor correlación con la primera dimensión, mientras que las variables cantidad producida y superficie cosechada tienen mayor correlación con la segunda dimensión. Mientras que el resto presentan baja correlación con las primeras 2 coordenadas.

Estudio de la calidad de representación y contribuciones de las variables en las dos primeras dimensiones del AFDM.

Calidad de representación:

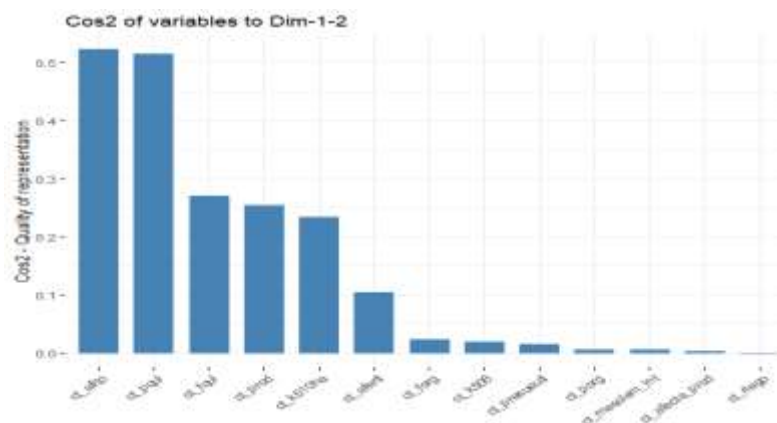


Ilustración 8-3: Calidad de representación de las variables con el primer AFDM.

Realizado por: Urquizo, Diego, 2022.

Las variables uso de fitosanitario, uso de plaguicidas quimicos estan representadas de una forma moderada esto debido a que la suma del \cos^2 de las 2 primeras dimensiones sobrepasan el 0.5, mientras que las variables uso de fertilizante quimico, produccion, superficie sembrada y uso de fertilizante estan representadas de una forma aceptable. (ver grafico 3-8).

Contribución de las variables.

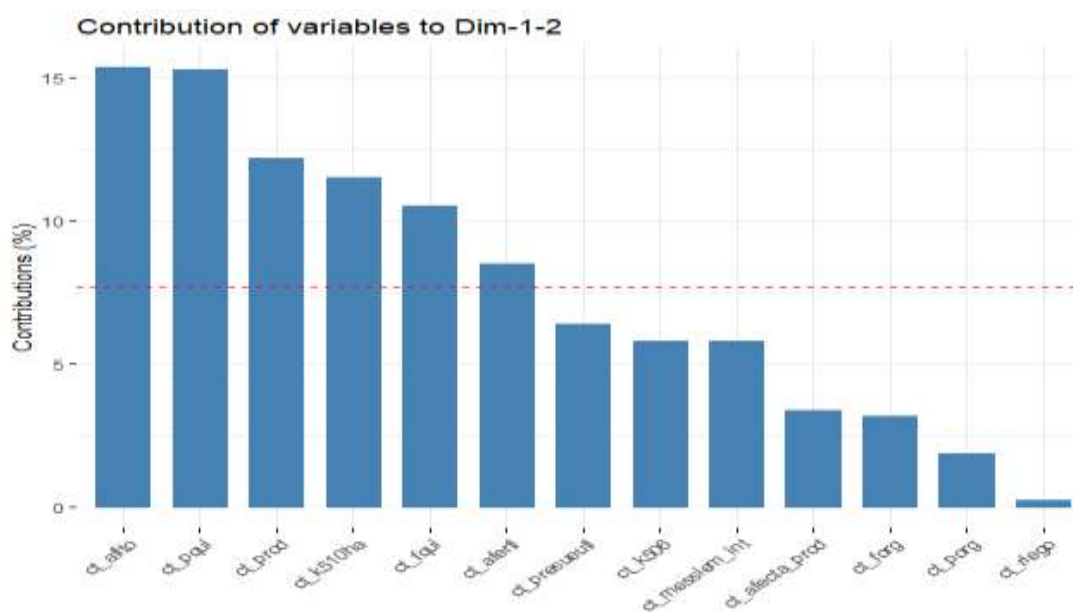


Ilustración 9-3: Contribución de las variables en el primer AFDM.

Realizado por: Urquiza, Diego, 2022.

Como se ve en el grafico 3-9 las variables que mayor contribución tienen en las dos primeras dimensiones del AFDM son: uso de fitosanitario, uso de plaguicida organico, produccion, superficie sembrada, uso de fertilizantes quimicos y uso de fertilizantes.

A consecuencia del analisis de la representación y contribucion de las variables en las dos primeras dimensiones del primer AFDM se procede a realizar un segundo AFDM tomando en cuenta solo las variables que mayor contribucion tienen, el cual mejorará el porcentaje de la variabilidad explicada así como sus interpretaciones.

3.5.2.2. Segundo analisis factorial de datos mixtos.

Este analisis solo tomara en cuenta las variables: uso de fitosanitario, uso de plaguicida organico, produccion, superficie sembrada, uso de fertilizantes quimicos y uso de fertilizantes.

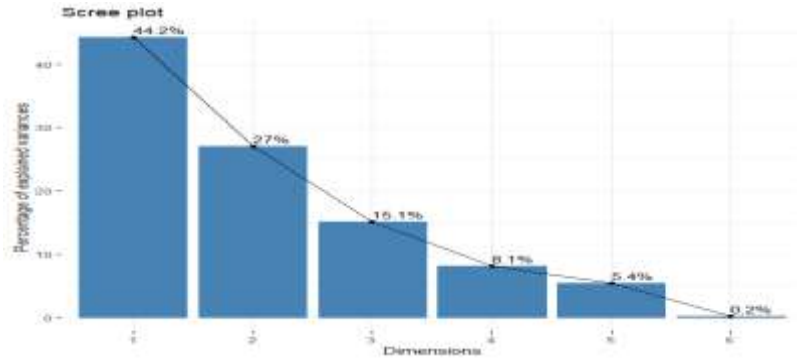


Ilustración 10-3: Grafico de sedimentación del segundo AFDM.

Realizado por: Urquizo, Diego, 2022.

Con el segundo AFDM se observa que el porcentaje de variabilidad total explicada por las dos primeras dimensiones son de un 44.2% y 27% respectivamente, dando como suma un total de 71.2% de variabilidad total explicada (Ver ilustración 10-3).

Con este porcentaje total lo cual ya es adecuado para realizar interpretaciones se observa que las variables uso de fertilizante, uso de fertilizante químico, uso de plaguicida químico y uso de fitosanitario tienen mayor correlación con la primera dimensión mientras que la producción en toneladas métricas y superficie sembrada tienen mayor correlación con la segunda dimensión (Ver ilustración 11-3).

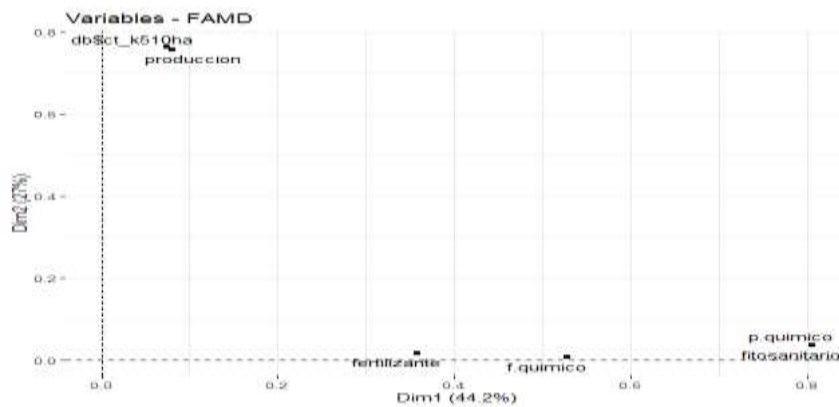


Ilustración 11-3: Representación de variables con el segundo AFDM.

Realizado por: Urquizo, Diego, 2022.

Análisis de las variables cuantitativas

Tabla 13-3: Coordenadas de las variables cuantitativas en el AFDM

Variables	Dimensión 1	Dimensión 2
Superficie sembrada	0.2724	0.8742
Producción	0.2836	0.8701

Realizado por: Urquizo, Diego, 2022.

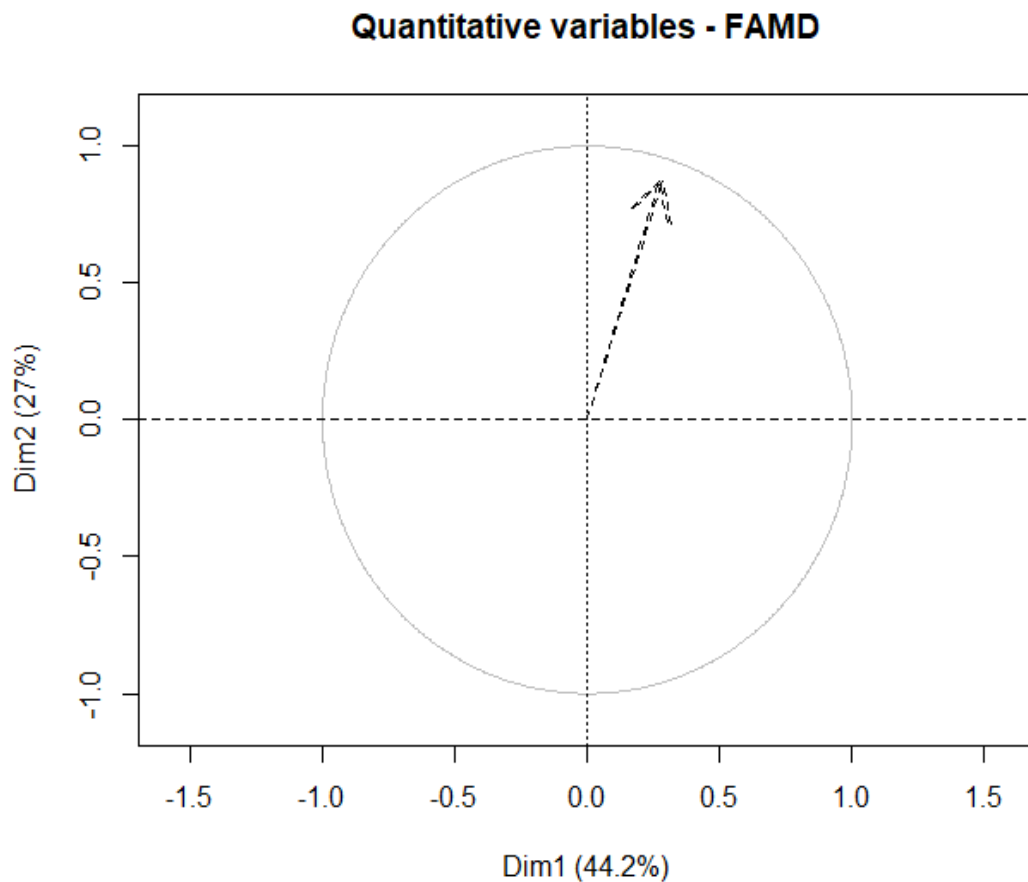


Ilustración 12-3: Representación de las variables cuantitativas en el plano factorial.

Realizado por: Urquizo, Diego, 2022.

La variable producción y superficie sembrada están correlacionadas entre sí y con la segunda dimensión, esto ya que ambos vectores apuntan hacia la misma dirección lo que indica que a mayor superficie cosechada mayor es la producción de papa (Ver gráfico 3-12).

Análisis de las variables cualitativas

Tabla 14-3: Coordenadas de las variables cualitativas en el AFDM.

Variables	Dimensión 1	Dimensión 2
Uso fitosanitario (Si)	0.9043	0.1508
Uso fitosanitario (No)	-2.3633	-0.3940
Uso plaguicida químico (Si)	0.9137	0.1523
Uso plaguicida químico (No)	-2.3406	-0.3901
Uso fertilizante químico (Si)	0.7888	0.0728
Uso fertilizante químico (No)	-1.7779	-0.1642
Uso fertilizante (Si)	0.3172	0.0543
Uso fertilizante (No)	-2.9965	-0.5133

Realizado por: Urquizo, Diego, 2022.

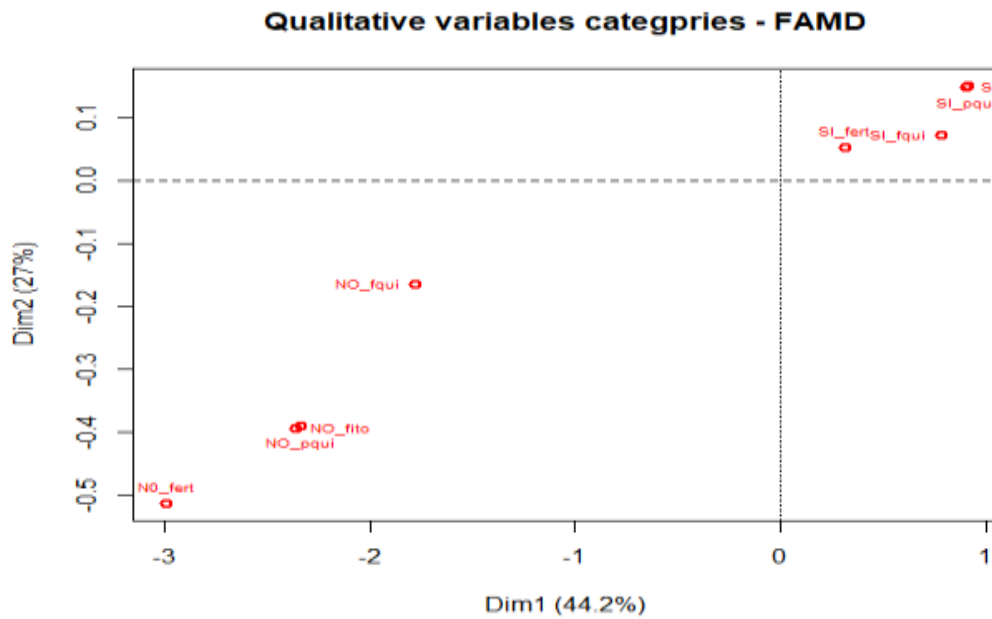


Ilustración 13-3: Representación de las variables cualitativas en el plano factorial.

Realizado por: Urquizo, Diego, 2022.

Cómo se puede evidenciar en la tabla 3-14 y grafico 3-13 el no usar fitosanitario está relacionado al no uso de los demás suplementos que son el fertilizante, plaguicida químico y

fertilizante químico, y viceversa, es decir, el uso de uno de estos suplementos también está relacionado con el uso de los demás.

En resumen, a partir de la observación de las contribuciones de las variables con el primer AFDM, para el segundo análisis se redujo el número de variables a estudiar solamente a seis variables donde el porcentaje de variabilidad explicada es moderada, con el segundo AFDM se observó la presencia de 2 factores que influyen en la producción de papa, donde el primer factor está compuesto por las variables: uso de fitosanitario, uso de fertilizantes, uso de plaguicida químico y uso de fertilizante orgánico, por otro lado el segundo factor que influye en la producción de papa está comprendida por la superficie sembrada.

Si se compara con el análisis bivariado y el análisis factorial, los resultados son similares, ya que en ambos casos se detectó la relación entre la superficie sembrada y la producción, además de la dependencia entre variables uso de fertilizante, uso de fitosanitario y uso de plaguicidas.

3.6. Identificación de los factores influyentes obtenidos con diferentes técnicas estadísticas.

A partir de los análisis realizados previamente se resume las variables influyentes en la producción de papa detectados, en la siguiente tabla:

Tabla 15-3: Resumen de los factores influyentes identificados con las técnicas estadísticas.

Técnicas	ARM	AFDM
Variables influyentes	Superficie sembrada	Superficie sembrada
	Semilla usada	Uso de fertilizante químico
	Afectación en la producción	Uso de plaguicida
	Uso de fertilizante	Uso de fertilizante químico
	Uso de plaguicida	
	Mes de siembra	

Realizado por: Urquiza, Diego, 2022.

Según la tabla 3-15 podemos observar que los factores relacionados o que influyen en la producción de papa son: Superficie sembrada, semilla usada, afectación en la producción, uso de fertilizante, uso de fertilizante químico, uso de plaguicida y mes de siembra.

3.7. Comparación de los factores identificados con los factores presentados por el International Potato center.

Tabla 16-3: Comparación factores influyentes presentados por el CIP y las identificadas en el estudio.

Factores influyentes en la producción de la papa según:	
International Potato Center	Datos de la ESPAC-2020
Tipo de semilla usada.	Superficie sembrada
Factores abióticos (Son factores que afectan y producen perdidas en el cultivo).	Semilla usada
Uso de fertilizantes (Químicos y Orgánicos).	Afectación en la producción
Enfermedades.	Uso de fertilizante
Uso de plaguicidas (Químicos y Orgánicos).	Uso de plaguicida
Preparación del suelo.	Mes de siembra
Prácticas culturales (Mes de siembra)	
Características agroecológicas y climáticas.	

Realizado por: Urquizo, Diego, 2022.

Según la tabla 3-16 se observa que los factores influyentes que se identificó haciendo uso datos de la ESPAC-2020 también están contenidas en las que presenta el CIP, por lo tanto, además de las investigaciones previas realizadas por este organismo también se confirma que estos factores si influyen en la producción de papa haciendo uso técnicas de análisis de datos.

CONCLUSIONES

- En base al análisis exploratorio de datos, se pudo determinar que la mayoría de terrenos donde se sembró papa comprende una superficie entre 0.01 y 8.01 hectáreas, donde se produjo en mayor frecuencia entre 0.01 a 159 toneladas métricas de papa. Además, los cultivos de papa presentan características con altas frecuencias donde la semilla usada es de tipo común, la mayor afectación en la producción es producida por la edad de plantación y plagas, la preparación del suelo lo realizan mediante tractor, no poseen riego, si usan fertilizantes tanto químicos como orgánicos, también usan fitosanitario, no usan plaguicidas orgánicos, pero si plaguicidas químicos, finalmente el mes donde mayormente siembran son entre los meses diciembre-enero y agosto.
- En cuanto a las relaciones entre variables se detectó que la producción de papa tiene una correlación alta con la superficie sembrada, además se resalta la dependencia entre el tipo de semilla usada con el uso de fertilizantes, uso de plaguicida y uso de fitosanitario, de la misma manera la afectación en la producción está relacionada con el tipo de preparación del suelo y mes de siembra, finalmente la preparación del suelo también está relacionada a la disposición o no del riego.
- De acuerdo al modelo de regresión la cual explica un 75% de variabilidad se obtuvo que las variables que influyen significativamente en la producción de papa son la superficie sembrada, el tipo de semilla usada, la afectación en la producción es uso de fertilizante, plaguicida y mes de siembra. Donde el modelo propuesto es:
$$\text{Producción} = 131.332 + 12.332(\text{Superficie Sembrada}) + 20.812(\text{Semilla usada} = \text{Mejorada}) - 28.053(\text{Afectación en la producción} = \text{practicas inadecuadas}) + 10.117(\text{fertilizante organico} = \text{No}) - 137.16(\text{plaguicida organico} = \text{No}) + 12.936(\text{Mes siembra} = \text{enero}) - 13.225(\text{Mes siembra} = \text{mayo}) + 13.652(\text{Mes siembra} = \text{diciembre}).$$
 El cual cumple con todos los supuestos necesarios para su validez.
- Según el análisis de correspondencia se observa que las categorías de las variables cualitativas que tienen relación están descritas como: la existencia de una alta frecuencia donde si el sembrío de papa se hizo en los primeros 5 meses del año estos terrenos tienen riego, se uso un tipo de semilla común, se aplicaron fertilizantes orgánicos y la preparación del suelo se realizó mediante la yunta, por otro lado también se determina una alta frecuencia de que si se usó semilla híbrida internacional se usaron fertilizantes y plaguicidas químicos.
- De acuerdo al AFDM se determina la presencia de dos factores que influyen en la producción de papa, el primer factor está compuesto por: uso de fitosanitario, uso de

fertilizantes, uso de plaguicida químico y uso de fertilizante orgánico, y el segundo factor que influye en la producción de papa está comprendida por la superficie sembrada.

- En cuanto a la comparación de los factores influyentes identificadas en este estudio y los presentados por en CIP, se concluye que mediante el análisis de datos se encontró factores influyentes idénticos al mencionado por dicho organismo siendo estas mayormente factores controlables.

RECOMENDACIONES

- Para un mejor análisis de datos se recomienda a la institución responsable en recolectar datos aumentar información relacionadas a características de tipo cuantitativas relacionadas al cultivo.
- Al cultivar de papa se recomienda tomar una buena decisión acerca del tipo de semilla a usar, puesto que es uno de los principales factores que influyen en su producción, además, del uso de fertilizantes y plaguicidas.
- Debido a la existencia de dependencia entre factores como el mes de siembra, uso de riego y el tipo de semilla es necesario enfocar las capacitaciones realizadas por instituciones como el MAGAP al mejoramiento del sistema de riego, políticas de adquisición y distribución de semillas buscando las mejores alternativas para cada temporada del año.
- La técnica de análisis factorial de datos mixtos (AFDM) usada en esta investigación, aunque presente mayor ventaja frente a técnicas clásicas, no son revisadas dentro de la malla curricular por lo es recomendable incluir la enseñanza de esta dentro de la carrera de Estadística en la ESPOCH.
- Es necesario ampliar la información tales como las enfermedades sufridas por el cultivo, características climáticas, etc. Que ayudaría a identificación y confirmación con investigaciones previas realizadas acerca de los factores que influyen en la siembra de este producto.

BIBLIOGRAFÍA

AGRICULTURA, M. *Agricultura, la base de la economía y alimentación - Ministerio de Agricultura y Ganadería.* [En línea], 2019. <https://www.agricultura.gob.ec/agricultura-la-base-de-la-economia-y-la-alimentacion/#:~:text=En%20Ecuador%2C%20este%20sector%20aporta,la%20pobreza%20en%20el%20campo..>

APARICIO, J. *Modelos lineales Aplicados en R.* Universidad Miguel Hernández, 2013, pp. 45.

ARGENPAPA. *Historia de la papa.* [En línea], 2015. <https://www.argenpapa.com.ar/info/la-papa/>.

ARIAS, K. *Efecto de la fertilización química, orgánica y combinada sobre el rendimiento de la papa variedad Granola.* Maracay : Red Instituto Nacional de Investigaciones Agrícolas de Venezuela, 2010, pp. 12.

ASAS, E. *Evaluación agronómica del cultivo de papa (solanum tuberosum) c.v. "fripapa" a la aplicación de tres abonos completos.* [En línea], 2019. <https://repositorio.uta.edu.ec/bitstream/123456789/6532/1/Tesis-69%20%20Ingeniería%20Agronómica%20-CD%20210.pdf>

BANK, W. *Agricultura y alimentos.* [En línea], 2018. <https://www.bancomundial.org/es/topic/agriculture/overview>.

BATANERO, GODINO y ESTEPA, A. *Análisis exploratorio de datos: Sus posibilidades en la enseñanza Secundaria.* [En línea], Enero de 1991. <http://www.ugr.es/~batanero/pages/ARTICULOS/anaexplora.pdf>.

BENITEZ, J. *Alternativas de comercialización de papa y cebolla colorada.* Quito: Print & Promo, 2010. pág. 77.

CENTER, I. *Generalidades del Centro Internacional de la PAPA.* [En línea], 2020. <https://cipotato.org/es/about/>.

CIPOTATO. *Uso de plaguicidas - Inventario de Tecnologías e Información para el Cultivo de Papa en el Ecuador.* [En línea], 2011. [https://cipotato.org/papaenecuador/uso-de-plaguicidas/#:~:text=Los%20plaguicidas%20son%20productos%20qu%C3%ADmicos,malezas\)%20que%20afectan%20los%20cultivos..](https://cipotato.org/papaenecuador/uso-de-plaguicidas/#:~:text=Los%20plaguicidas%20son%20productos%20qu%C3%ADmicos,malezas)%20que%20afectan%20los%20cultivos..)

CUADRAS, C. *Nuevos métodos de análisis multivariante.* Barcelona : CMC Editions, 2018, pp. 48.

DE LA FUENTE, S. *Análisis de Correspondencias Simples y Múltiples; Universidad Autónoma de Madrid.* [En línea], 2011. <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>.

DE LA PUENTE, C. *Estadística descriptiva e inferencial.* Madrid : IDT, 2018.

ECURED. *Planta de papa.* [En línea], 2016. https://www.ecured.cu/images/c/c7/Planta_de_papa.jpg.

FERNANDEZ, ABASCAL y CALVO. *Análisis factorial múltiple como técnicas de estudio de la estabilidad de los resultados de un análisis de componentes principales.* [En línea], 2002. <https://upcommons.upc.edu/handle/2099/4175>.

FINCK, A. *Fertilizantes y fertilización: fundamentos y métodos para la fertilización de los cultivos.* Barcelona : Editorial Reverté, 1988. pág. 454.

GALLARDO, S. *Análisis exploratorio de datos.* [En línea], 2003. <https://ciberconta.unizar.es/leccion/aed/ead.pdf>.

GUAMÁN DAQUILEMA, Segundo Eduardo & MULLO GUAMINGA, Héctor Salomón. *Análisis estadístico multivariante para el estudio de los factores que influyen en la producción del plátano en el Ecuador, Periodo 2014-2016* [En línea] (Trabajo de titulación) (Ingeniería) Escuela Superior Politécnica de Chimborazo, Chimborazo, Ecuador. 2018. pp. 14-20. <http://dspace.espace.edu.ec/bitstream/123456789/8969/1/226T0043.pdf>.

GUTIERREZ PULIDO, Humbreto & SALAZAR, Román de la Vara. *Analisis y Diseño de Experimentos. Regresión Lineal Multiple.* BOGOTÁ : McGrawHill, 2008, pp. 57.

HAWKINGS, D. *Identification of Outliers.* [En línea], 1980. <https://link.springer.com/book/10.1007/978-94-015-3994-4>.

INEC. *Encuesta de Superficie y Producción Agropecuaria Continua; Instituto Nacional de Estadística y Censo.* [En línea], 2020. <https://www.ecuadorencifras.gob.ec/estadisticas-agropecuarias-2/>.

INEC. *Encuesta de Superficie y Producción Agropecuaria Continua - Metodología; ; Instituto Nacional de Estadística y Censo.* [En línea], 2020. https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_agropecuarias/espac/espac-2020/Metodologia%20ESPAC%202020.pdf.

INEC. *Estadísticas Agropecuarias; Instituto Nacional de Estadística y Censo.* [En línea], 2021. <https://www.ecuadorencifras.gob.ec/estadisticas-agropecuarias-2/>.

INEC. *Metodología de la encuesta de superficie y producción agropecuaria continua (ESPAC); Instituto Nacional de Estadística y Censo.* [En línea], 2015. <https://www.ecuadorencifras.gob.ec/wp-content/uploads/downloads/2016/10/Metodolog%C3%ADa-de-la-Encuesta-de-Superficie-y-Producci%C3%B3n-Agropecuaria-Continua-2014.pdf>.

INEC. *Uso de Plaguicidas en la Agricultura; Instituto Nacional de Estadística y Censo.* [En línea], 2013. https://www.ecuadorencifras.gob.ec/documentos/web-inec/Encuestas_Ambientales/plaguicidas/Plaguicidas-2013/Presentacion_resultados_principales-Uso_de_Plaguicidas_en_la_Agricultura_2013.pdf.

KUTNER, NACHTSHEIM & NETER. *Applied Linear Regression Models.* Cuarta. s.l. : McGraw-Hill Irwin, 2004.

MANUEL PUMISACHO & SHERWOOD. *El cultivo de la papa en el ecuador.* [En línea], 2002. <https://cipotato.org/wp->

content/uploads/Documentacion%20PDF/Pumisacho%20y%20Sherwood%20Cultivo%20de%20Papa%20en%20Ecuador.pdf.

MENDIVELSO, F. *Prueba Chi-Cuadrado de independencia aplicada a tablas 2xN*. [En línea], 2018. https://www.researchgate.net/publication/327733869_Prueba_Chi-Cuadrado_de_independencia_aplicada_a_tablas_2xN.

MITECO. *Fitosanitarios*. [En línea], 2011. <https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/productos-quimicos/fitosanitarios/>.

MONTESINOS, A. *Estudio del AIC y BIC en la selección de modelos de vida con datos censurados*. Centro de Investigación en Matemáticas A.C., s.l. : 2011.

NAVAS, E. *Aplicación de productos fitosanitarios*. Málaga : ICB, 2012. pág. 503.

OIRSA. *Uso y manejo seguro de plaguicidas en chile habanero*. San Salvador : s.n., 2005.

ONDARURAL. *Manual: Producción orgánica de la papa*. [En línea], 2021. <https://ondarural.org/manual-produccion-organica-de-la-papa/>.

PAGÉS, J. *Analyse factorielle de données mistes*. [En línea], 2004. http://www.numdam.org/item/?id=RSA_2004__52_4_93_0.

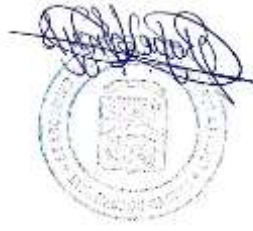
RURAL, S. *¿Qué es y para qué sirve el fertilizante?* [En línea], 2019. <https://www.gob.mx/agricultura/articulos/que-es-y-para-que-sirve-el-fertilizante>.

VALDERRAMA, M. *Producción y Utilización de la papa en el Ecuador*. [En línea], 1980. <http://cipotato.org/wp-content/uploads/2014/09/an20888.pdf>.

WIKIMEDIA. *Región Sierra*. [En línea], 2008. https://es.wikipedia.org/wiki/Región_Sierra.

WIKIMEDIA. *Solanum tuberosum*. [En línea], 2003. https://es.wikipedia.org/wiki/Solanum_tuberosum.

ZUBCOFF, J. *FAMD - Factor analysis of mixed data.* [En línea], 2017.
<https://rua.ua.es/dspace/bitstream/10045/72567/3/Analisis-multivariante-FAMD.pdf>.



ANEXO A: ACCESO A INFORMACIÓN INVESTIGADA

Para la obtención y acceso a los datos usados, dirigirse al enlace:

<https://www.ecuadorencifras.gob.ec/estadisticas-agropecuarias-2/>

ANEXO B: ALGORITMO PARA DETECCIÓN DE DATOS ATÍPICOS (MATLAB).

Función

```
function [idx,dm,mm,Ss] = kur_rcex(x,mode)
%
% [idx,dm,mean,Cov] = kur_rcex(x,mode)
%
% Inputs: observations, x, (a matrix having a row for each observation)
%         mode, 0 use p maxizing and p minimizing directions
%             iterate until no outliers are detected (default)
%             >0 use p maxizing and p minimizing directions
%             iterate "mode" times
%             <0 use only maximizing directions
%             iterate "mode" times
% Outputs: idx, zero/one vector with ones in the positions of the
%          suspected outliers
%          dm, robust Mahalanobis distances
%          mean, robust mean estimator
%          Ss, robust covariance matrix estimator
%
% Daniel Pena / Francisco J Prieto 13/3/01
[n,p] = size(x);
if (n > 750)|(p > 50),
    disp('Data set too large for this code');
    return
end

if nargin < 2,
    mode = 0;
end
```

```

maxit = Inf;
if mode < 0,
    maxit = abs(mode);
    mode = -1;
elseif mode > 0,
    maxit = mode;
    mode = 1;
end

% Initialization of parameters
%% Cutoff points for projections
ctf1 = 3.05 + 0.324*p;
ctf2 = exp(1.522-0.268*log(p));
ctf3 = 3.5;
ctf4 = 3.0;

%% ji2 percentiles (0.01)
ji2a = [ 6.64 9.21 11.35 13.28 15.09 16.81 18.48 20.09 21.67 23.21 ...
        24.73 26.22 27.69 29.14 30.58 32.00 33.41 34.81 36.20 37.57 ];

%% minimum numbers of observations
lmn1 = max(floor(0.5*(n+p+1)),2*p);
lmn2 = min(2*p,max(p+2,floor(0.25*n)));

% Removing suspect outliers
xx0 = [x (1:n)'];
xx = xx0;
nn = n;
it = 1;
while (it <= maxit),
    xx1 = xx(:,1:p);
    V = kur_nw(xx1,mode);
    pr = xx1*V;
    md = median(pr);
    prn = abs(pr - ones(nn,1)*md);
    mad = median(prn);
end

```



```

tt = prn./(ones(nn,1)*mad);

if p > 2,
    ctfo = ctfl - (0:(p-2))*0.75*(ctfl-ctf2)/(p-2);
else
    ctfo = ctfl;
end
if mode >= 0,
    tctf = ones(nn,1)*[ ctfo ctf2 ctf3*ones(1,p-1) ctfl ];
else
    tctf = ones(nn,1)*[ ctfo ctf2 ];
end
taux = tt./tctf;
t = max(taux)';

in = find(t > 1);
nn = length(in);
if nn == 0,
    break
end
inn = find(t <= 1);
nu = length(inn);

if nu < lmn2,
    [v,ix] = sort(t);
    ixx = ix(1:lmn2);
    xx = xx(ixx,:);
    break
end
xx = xx(inn,:);
[nn,pp] = size(xx);
if nn <= lmn2,
    break
end
it = it + 1;
end

```

```

%% Extracting the indices of the outliers
idx = ones(n,1);
[nn,pp] = size(xx);
idx(xx(:,pp)) = zeros(nn,1);

%% recheck observations and relabel them
%% Cutoffs for Mahalanobis distances
if p <= 20,
    ctf = ji2a(p);
else
    ctf = (2.33 + sqrt(2*p-1))^2/2;
end

%% Mahalanobis distances using center and scale estimators
%% based on good observations
sidx = sum(idx);
s1 = find(idx);
s2 = find(idx == 0);
if sidx > 0,
    xx1 = xx0(s1,:);
    xx1r = xx1(:,1:p);
end
xx2r = xx0(s2,1:p);

mm = mean(xx2r);
Ss = cov(xx2r);
if sidx > 0,
    aux1 = xx1r - ones(sidx,1)*mm;
    dd = Ss\aux1';
    dms = sum((aux1.*dd'));
end

%% Ensure that at least lmn1 observations are considered good
ado = sidx + lmn1 - n;
if ado > 0,

```

```

[dmss,idms] = sort(dms);
idm1 = idms(1:ado);
ido = xx0(s1,pp);
s3 = ido(idm1);
idx(s3) = zeros(ado,1);

sidx = sum(idx);
s1 = find(idx);
s2 = find(idx == 0);
if sidx > 0,
    xx1 = xx0(s1,:);
    xx1r = xx1(:,1:p);
end
xx2r = xx0(s2,1:p);

mm = mean(xx2r);
Ss = cov(xx2r);
if sidx > 0,
    aux1 = xx1r - ones(sidx,1)*mm;
    dd = Ss\aux1';
    dms = sum((aux1.*dd'));
end
end

%% Check remaining observations and relabel if appropriate
while sidx > 0,

s1 = find(dms <= ctf);
s2 = length(s1);
if s2 == 0,
    break
end
s3 = xx1(s1,pp);
idx(s3) = zeros(s2,1);
sidx = sum(idx);

```

```

s1 = find(idx);
s2 = find(idx == 0);
if sidx > 0,
    xx1 = xx0(s1,:);
    xx1r = xx1(:,1:p);
end
xx2r = xx0(s2,1:p);

mm = mean(xx2r);
Ss = cov(xx2r);
if sidx > 0,
    aux1 = xx1r - ones(sidx,1)*mm;
    dd = Ss\aux1';
    dms = sum((aux1.*dd'));
end

end

```

% Values to be returned

```

aux1 = x - ones(n,1)*mm;
dd = Ss\aux1';
dms = sum((aux1.*dd'));

```

```
dm = sqrt(dms);
```

Aplicación de la función

```
[idx,dm,mm,Ss] = kur_rcex(X,0)
```

ANEXO C: CÓDIGO EN R.

```

#AED
#Cargar las librerías
library("e1071")
library("prettyR")
library("modeest")
library("sjPlot")

```

```

library("haven")
library("car")
library("openxlsx")
library("lme4")
library("psych")
library("GGally")
library("dplyr")

#Lectura de la base de datos
db<-read_sav(file = "Data/Data TESIS.sav")
names(db)
db$ct_k506<-as.factor(db$ct_k506)
db$ct_afecta_prod<-as.factor(db$ct_afecta_prod)
db$ct_presueuti<-as.factor(db$ct_presueuti)
db$ct_riego<-as.factor(db$ct_riego)
db$ct_aferti<-as.factor(db$ct_aferti)
db$ct_forg<-as.factor(db$ct_forg)
db$ct_fqui<-as.factor(db$ct_fqui)
db$ct_afito<-as.factor(db$ct_afito)
db$ct_porg<-as.factor(db$ct_porg)
db$ct_pqui<-as.factor(db$ct_pqui)
db$ct_messiem_int<-as.factor(db$ct_messiem_int)

# str(bd1)

#VARIABLES CUANTITATIVAS
#Variable eestadística ct_k510ha="Superficie Sembrada"
tb.fs<-sjmisc::frq(db$ct_k510ha,auto.grp=nclass.Sturges(db$ct_k510ha))
tb.fs
write.csv(tb.fs,file="Export/1.tb_fs.csv")
nrow(db)
hist(db$ct_k510ha,breaks = 7,col = "orange",main="Histograma de frecuencia de la variable su
perficie sembrada",
      xlab="Superficie sembrada (Hectáreas)",ylab = "Frecuencia")

#Variable eestadística ct_prod="produccion"
tabla.def<-sjmisc::frq(db$ct_prod,auto.grp=nclass.Sturges(db$ct_prod))
tabla.def
write.csv(tabla.def,file="Export/tabla.df1.csv")

```

```

nrow(db)
hist(db$ct_prod,breaks = 7,col = "orange",main="Histograma de frecuencia de la variable producción de papa",
      xlab="Producción en toneladas métricas",ylab = "Frecuencia")

#Resumén estadístico de las variables cuantitativas
#Tendencia central
moda <- function(x) {
  return(as.numeric(names(which.max(table(x)))))
}
media<-cbind(mean(db$ct_k510ha),mean(db$ct_prod))
mediana<-cbind(median(db$ct_k510ha),median(db$ct_prod))
modas<-cbind(moda(db$ct_k510ha),moda(db$ct_prod))

#Posición
q1<-cbind(quantile(db$ct_k510ha,0.25),quantile(db$ct_prod,0.25))
q2<-cbind(quantile(db$ct_k510ha,0.5),quantile(db$ct_prod,0.5))
q3<-cbind(quantile(db$ct_k510ha,0.75),quantile(db$ct_prod,0.75))

#Dispersion
vari<-cbind(var(db$ct_k510ha),var(db$ct_prod))
des<-cbind(sd(db$ct_k510ha),sd(db$ct_prod))
Cv<-des/media

#Forma
as<-cbind(skewness(db$ct_k510ha),skewness(db$ct_prod))
kur<-cbind(kurtosis(db$ct_k510ha),kurtosis(db$ct_prod))
rvc<-round(rbind(media,mediana,modas,q1,q2,q3,vari,des,Cv,as,kur),2)
write.csv(rvc,file="Export/3.resv.cuanti.csv")

#Variables cualitativas
sjmisc::frq(db$ct_k506)
sjmisc::frq(db$ct_afecta_prod)
sjmisc::frq(db$ct_presueuti)
sjmisc::frq(db$ct_riego)
sjmisc::frq(db$ct_aferti)
sjmisc::frq(db$ct_forg)
sjmisc::frq(db$ct_fqui)
sjmisc::frq(db$ct_afito)
sjmisc::frq(db$ct_porg)

```

```

sjmisc::frq(db$ct_pqui)
sjmisc::frq(db$ct_messiem_int)
#Análisis de relación o asociación entre variables
#Variables Cuantitativas
cor.test(db$ct_k510ha,db$ct_prod)
#Variables cualitativas
chisq.test(db$ct_pqui,db$ct_messiem_int)
#Modelo de leregresion multiple
model1<-lm(ct_prod~.,data=db)
summary(model1)
md1<-summary(model1)$coefficients

mejorada <- ifelse(db$ct_k506==2,1,0)

prc_inad <- ifelse (db$ct_afecta_prod==5, 1,0)

nforg <- ifelse(db$ct_forg==2,1,0)

nporg <- ifelse(db$ct_porg==2,1,0)

enero <- ifelse(db$ct_messiem_int==1,1,0)
mayo <- ifelse(db$ct_messiem_int==5, 1,0)
diciembre <- ifelse(db$ct_messiem_int==12,1,0)

model2<-lm(db$ct_prod~db$ct_k510ha+mejorada+prc_inad+nforg+nporg+enero+mayo+dicie
mbre)
summary(model2)
md2<-summary(model2)$coefficients

write.xlsx(md1,file = "Export/6. Resumen modelo 1.xlsx")
write.xlsx(md2,file = "Export/7. Resumen modelo 2.xlsx")

vif(model2)
par(mfrow=c(2,2))
plot(model2)

```

```
#Selección del mejor modelo
```

```
AIC(model1)
```

```
AIC(model2)
```

```
BIC(model1)
```

```
BIC(model2)
```

```
#Supuestos del modelo
```

```
qqPlot(model2,grid = F)
```

```
jarque.bera.test(residuals(model2))
```

```
gqtest(model2)
```

```
dwtest(model2)
```

```
#Análisis de correspondencia
```

```
dev.off()
```

```
install.packages("ca")
```

```
library(ca)
```

```
library(GDAtools)
```

```
newdb<-db[-c(1,2,4,7)]
```

```
head(newdb)
```

```
disj<-apply(newdb,2,dichotom)
```

```
disj1<-cbind(disj[[1]],disj[[2]],disj[[3]],disj[[4]],disj[[5]],disj[[6]],  
            disj[[7]],disj[[8]],disj[[9]])
```

```
burt<-as.matrix(t(disj1)) %*% as.matrix(disj1)
```

```
nombres<-c("Comun","Mejorada","Certificada","Hib_nac","Hib_int","Tractor",  
           "Yunta","Motocultor","Azadon","R_si","R_no","Fero_si",  
           "Fero_no","Fq_si","Fq_no","Fito_si","Fito_no","Plao_si","Plao_no","Pq_si",  
           "Pq_no","Ene","Feb","Mar","Abr","May","Jun","Jul","Ago","Sep","Oct",  
           "Nov","Dic")
```

```
colnames(burt)<-nombres
```

```
rownames(burt)<-colnames(burt)
```

```
cordenadas<-ca(burt)$rowcoord[,1:2]
```

```
write.xlsx(data.frame(cordenadas,rownames(cordenadas)),file = "Export/8. Analisis correspo  
ndencia.xlsx")
```

```
plot(ca(burt),col = c("blue"),xlab="Dimension 1 (61.3%)",ylab="Dimension 2 (8.7%)",main="
```



```

Variables")
#Análisis factorial de datos mixtos.
library("FactoMineR")
library("factoextra")
library("corrplot")
library("dplyr")
names(db)

#Primer análisis factorial de datos mixtos
res.famd<-FAMD(db,graph=F,ncp=17)
print(res.famd)
# Gráfico de sedimentación
fviz_screplot(res.famd,addlabels=TRUE)
# Resultados del AFDM para las variables
var<-get_famd_var(res.famd)
var
# Gráfico de variables
fviz_famd_var(res.famd,"var",col.var="black",shape.var=15,
              repel=TRUE)
# Diagrama de barras del cos2 de variables en las dos primeras dimensiones
fviz_cos2(res.famd,choice="var",axes=1:2)
# Cos2: calidad de representación en el mapa de factores
fviz_famd_var(res.famd,"var",col.var="cos2",
              gradient.cols=c("green","blue","red"),
              repel=T)
# Diagrama de barras de las contribuciones de las variables en las dos primeras dimensiones
fviz_contrib(res.famd,choice="var",axes=1:2,top=14)
# Contrib: contribuciones de las variables en el mapa de factores
fviz_famd_var(res.famd,"var",col.var="contrib",
              gradient.cols=c("green","blue","red"),
              repel=T)
#Recodificar las variables para el segundo AFDM
db.f<-as.data.frame(db$ct_k510ha)
db.f$produccion<-db$ct_prod
db.f$fitosanitario<-recode(db$ct_afito,'1'="SI_fito",'2'="NO_fito")
db.f$p.quimico<-recode(db$ct_pqui,"1"="SI_pqui","2"="NO_pqui")

```

```

db.f$quimico<-recode(db$ct_fqui,"1"="SI_fqui","2"="NO_fqui")
db.f$fertilizante<-recode(db$ct_aferti,"1"="SI_fert","2"="NO_fert")

#Segundo análisis factorial de datos mixtos
res.famd<-FAMD(db.f,graph=F,ncp=17)
print(res.famd)
# Gráfico de sedimentación
fviz_screplot(res.famd,addlabels=TRUE)
# Resultados del AFDM para las variables
var<-get_famd_var(res.famd)
var
# Gráfico de variables
fviz_famd_var(res.famd,"var",col.var="black",shape.var=15,
              repel=TRUE)
# Diagrama de barras del cos2 de variables en las dos primeras dimensiones
fviz_cos2(res.famd,choice="var",axes=1:2)
# Cos2: calidad de representación en el mapa de factores
fviz_famd_var(res.famd,"var",col.var="cos2",
              gradient.cols=c("green","blue","red"),
              repel=T)
# Diagrama de barras de las contribuciones de las variables en las dos primeras dimensiones
fviz_contrib(res.famd,choice="var",axes=1:2,top=14)
# Contrib: contribuciones de las variables en el mapa de factores
fviz_famd_var(res.famd,"var",col.var="contrib",
              gradient.cols=c("green","blue","red"),
              repel=T)
# Análisis de las variables cuantitativas
quanti.var<-get_famd_var(res.famd,"quanti.var")
quanti.var
# Representar las variables cuantitativas (círculo de correlaciones)
fviz_famd_var(res.famd,"quanti.var",repel=T,
              col.var="black")

coor1<-res.famd$quanti.var$coord[,c(1,2)]
coor1[,1]<-coor1[,1]*(-1)

```

```
# Análisis de las variables cualitativas
quali.var<-get_famd_var(res.famd,"quali.var")
quali.var
#####
# visualizar variables cualitativas
# Representar las categorías de las variables cualitativas en el mapa factorial
fviz_famd_var(res.famd,"quali.var",
               col.var="black")
```



esPOCH

Dirección de Bibliotecas y
Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y
DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 26 / 01 / 2023

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: Diego Vladimir Urquizo Urquizo
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Estadística
Título a optar: Ingeniero Estadístico
f. Analista de Biblioteca responsable: Ing. Rafael Inty Salto Hidalgo



0120-DBRA-UPT-2023