

FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN



CARRERA PROFESIONAL DE ADMINISTRACIÓN Y NEGOCIOS INTERNACIONALES

TESIS

CALCULAR Y COMPARAR EL COMPORTAMIENTO FUTURO DEL PRECIO DE LA
PAPA, ENTRE LOS PRONÓSTICOS DE LOS MODELOS ARIMA Y GAM EN EL PERÚ, EN
EL PERIODO 2011 – 2022.

Autor: Lady Geraldine Durand Chavez

Asesor: Diego Eliseo Carpio Segura

Para la obtención del grado de:

LICENCIADA EN ADMINISTRACIÓN Y NEGOCIOS INTERNACIONALES

AREQUIPA – PERÚ

2021

DEDICATORIA

La presente tesis está dedicada a Dios por permitirme gozar de salud y poder vivir este momento tan importante de mi formación profesional, en especial a toda mi familia por siempre brindarme su apoyo incondicional y todas aquellas personas que siempre estuvieron a mi lado aportando un granito de arena.

AGRADECIMIENTO

Un agradecimiento especial al profesor Arturo Alatriza Corrales por inspirar la presente tesis, a mi asesor Diego Carpio Segura por su apoyo incondicional y confianza, de la misma manera a todos mis maestros de la Universidad la Salle los llevare siempre en mi corazón, a mi familia y amigos que siempre me adelantaron hasta escribir estas palabras.

Muchas gracias a todos.

INDICÉ

DEDICATORIA	II
AGRADECIMIENTO.....	III
RESUMEN	IX
ABSTRACT.....	X
I. Capítulo: El problema de la investigación científica	11
1.1. Planteamiento del problema de investigación científica	11
1.2. Variables.....	13
1.3. Objetivos.....	13
1.3.1. Objetivo general	13
1.3.2. Objetivos específicos.....	14
1.4. Formulación de problema.....	14
1.4.1. Problema general.....	14
1.4.2. Problemas específicos	14
1.5. Hipótesis	15
1.5.1. Hipótesis nula.....	15
1.5.2. Hipótesis alternativa.....	15
1.6. Justificación.....	15
II. Capítulo: Revisión y fundamentación teórica	17
2.1. Antecedentes.....	17
2.2. Bases teóricas	23
2.2.1. Papa.....	23
2.2.1.1. Consumo de papa.....	23
2.2.1.2. Producción.....	23
2.2.1.3. Precio en chacra.....	24

2.2.1.4.	Precio en el mercado mayorista de Lima.....	25
2.2.1.5.	Estacionalidad.....	26
2.2.1.6.	Variedades	27
2.2.1.7.	Clima	28
2.2.2.	Método ARIMA	29
2.2.2.1.	Definición	29
2.2.2.2.	Proceso estocástico.....	30
2.2.2.3.	Proceso estocástico estacionario.....	30
2.2.2.4.	Función de autocorrelacion	31
2.2.2.5.	Ruido blanco.....	32
2.2.2.6.	Procesos autorregresivos: AR(p).....	32
2.2.2.6.1.	Modelo autorregresivo de orden 1, AR (1).....	32
2.2.2.6.2.	Proceso AR(p)	33
2.2.2.7.	Proceso de medias móviles: MA(q).....	33
2.2.2.8.	Orden de integridad (I)	34
2.2.2.9.	Las cuatro etapas del modelo ARIMA	35
2.2.2.10.	Beneficios del modelo ARIMA.....	35
2.2.3.	Modelo aditivo generalizado (GAM).....	35
2.2.4.	Programación R.....	38
2.2.4.1.	Definición	38
2.2.4.2.	Características.....	39
III.	Capítulo: Método.....	40
3.1.	Tipo de investigación.....	40
3.2.	Población	40
3.3.	Técnica e instrumento.....	40

3.4. Matriz de consistencia	41
Elaboración propia.....	42
3.5. Software estadístico R	43
3.6. Procedimiento.....	43
IV. Capitulo: Resultados y discusiones	46
4.1. Método ARIMA	46
4.1.1. Precios promedios mensuales al por mayor de la papa en el Perú.....	46
4.1.2. Estacionariedad	46
4.1.3. Estimación del mejor modelo ARIMA	47
4.1.3.1. Modelo de autocorrelación y parcial autocorrelación	47
4.1.3.2. Pronostico del modelo ARIMA (2,1,2)	50
4.1.3.3. Comparación entre realidad y pronostico 2021	52
4.1.3.4. Comparación entre la realidad y pronóstico de la papa canchán 2018 - 2020.....	53
4.1.4. Validación	55
4.2. Modelo Aditivo Generalizado	55
4.2.1. Tipo de distribución de probabilidades	55
4.2.2. Generación del mejor modelo GAM.....	56
4.2.3. Inspección del modelo GAM	58
4.2.4. Validación del modelo.....	59
4.2.5. Predicción.....	59
4.3. Comparación de los métodos ARIMA y GAM.	65
Conclusiones	67
Recomendaciones.....	68
Bibliografía.....	69
Anexos.....	74

ÍNDICE DE TABLAS

Tabla 1 Cuadro de variables	13
Tabla 2 Perú: Indicadores de producción de papa (2015 - 2019).....	24
Tabla 3 Variedad de papa por origen, color y uso	27
Tabla 4 Tipos de familias exponenciales	38
Tabla 5 Técnicas e instrumentos.....	41
Tabla 6 Matriz de consistencia	42
Tabla 7 Los coeficientes obtenidos del mejor modelo ARIMA	50
Tabla 8 Pronostico de precio de la papa entre los años 2021 - 2022.....	51
Tabla 9 Comparación entre modelo ARIMA y GAM.....	65

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Consumo per capital de papa 2010 - 2020 por Kg.	23
Ilustración 2 Producción y precio en chacra de la papa Perú 2015 - 2019	25
Ilustración 3 Evolución Mensual del precio (S/) de papa - Lima 2019.	26
Ilustración 4 Estacionalidad de la producción de papa Perú	27
Ilustración 5 Los diferentes climas en el Perú	29
Ilustración 6 Flujograma general del procedimiento metodológico	44
Ilustración 7 Procesamiento de datos en R	45
Ilustración 8 Variación del precio de papa (S/ - Kg) entre los años 2011 al 2020	46
Ilustración 9 Proceso estocástico del precio de papa (S/ - Kg) 2011 – 2020.....	47
Ilustración 10 distribución de los elementos de medias móviles (autocorrelación).....	48
Ilustración 11 La distribución del parcial autocorrelacional para generar los autorregresivos	48
Ilustración 12 Los diferentes modelos ARIMA obtenidos en R	49
Ilustración 13 Pronostico de precio de papa entre los años 2021 - 2022.....	52
Ilustración 14 Comparación entre pronostico y realidad del precio de la papa año 2021.....	52
Ilustración 15 Evaluación del precio de papa canchán por kilogramos del 2008 al 2017	53
Ilustración 16 Pronostico de precio de papa canchan para los años 2018 - 2020	54
Ilustración 17 Comparación entre pronostico y realidad del precio de papa canchan 2018 - 2020	54
Ilustración 18 Validación del modelo ARIMA (2,1,2)	55
Ilustración 19 Histograma de frecuencia del precio de la papa entre 2011 - 2020	56
Ilustración 20 Resultados obtenidos del modelo GAM	57
Ilustración 21 Gráficos obtenidos del Modelo Aditivo Generalizado	58
Ilustración 22 Prueba de bondad del ajuste del modelo	59
Ilustración 23 Predicción del precio de la papa mediante el modelo GAM	60
Ilustración 24 Determinación del mejor modelo ARIMA en el programa R	61
Ilustración 25 Pronostico de la producción de papa entre los años 2021 - 2030.....	62
Ilustración 26 Comparación de producción de papa, entre la predicción – realidad 2021	62
Ilustración 27 Comparación de producción de papa, entre la predicción – realidad 2018 - 2020	63
Ilustración 28 Predicción del precio de la papa utilizando el modelo SARIMA 2021 - 2030	64

RESUMEN

La papa es el cuarto alimento más importante a nivel mundial y el segundo en el Perú, por su alto contenido nutricional, este tubérculo se siembra principalmente en las zonas altoandinas, señalándolo como el termómetro de la pobreza ya que no existe muchas alternativas de cultivo en este sector; es así que ello provoca el principal problema que aqueja al sector, es decir la sobreproducción. Por lo expuesto anteriormente la presente investigación estimó el comportamiento futuro de la papa, a través de los modelos ARIMA y GAM para finalmente compararlos.

El estudio considero los datos mensuales entre el periodo de años 2011 al 2020, obtenidos de las estadísticas del Ministerio de Agricultura y Riego (MINAGRI) y aplicando las metodologías para cada modelo, se determinó que el mejor sistema para ARIMA es (2,1,2), a través de ejecutar la predicción de precios para los próximos 24 meses; así mismo el mejor modelo para GAM es la distribución Gamma, mediante la ejecución de dos tipos de predicciones: (a) La primera solo corriendo el modelo GAM con la función “Predit”, y (b) el segundo sumándole el pronóstico de la variable explicativa de producción mediante el sistema SARIMA((1,0,0)(2,1.1), para los años 2021 al 2030. De esta manera, ambos resultados son de suma utilidad para la planificación y toma de decisiones del sector agrícola.

Palabras claves: GAM, ARIMA, Predicción, Programación R, Estadística, Precios de alimentos.

ABSTRACT

The potato is the fourth most important food worldwide and the second in Peru, due to its high nutritional content, this tuber is planted mainly in the high Andean areas, pointing to it as the thermometer of poverty since there are not many alternatives of cultivation in this sector; this causes the main problem that afflicts the sector, that is, overproduction. Therefore, the present research estimated the future behavior of the potato, through the ARIMA and GAM models to finally compare them.

The study considered the monthly data between the period of years 2011 to 2020, obtained from the statistics of the Ministry of Agriculture and Irrigation (MINAGRI) and applying the methodologies for each model, it was determined that the best system for ARIMA is (2,1,2), through executing the price prediction for the next 24 months; likewise, the best model for GAM is the Gamma distribution, through the execution of two types of predictions: (a) The first only running the GAM model with the "Predit" function, and (b) the second adding the forecast of the explanatory variable of production through the SARIMA system((1,0,0)(2,1.1), for the years 2021 to 2030. In this way, both results are very useful for planning and decision-making in the agricultural sector.

Keywords: GAM, ARIMA, Prediction, R Programming, Statistics, Food Prices

Capítulo: El problema de la investigación científica básica

1.1.Planteamiento del problema de investigación científica

La papa es el cuarto alimento de mayor consumo a nivel internacional por su gran contenido energético, fuente de carbohidratos, proteínas, vitaminas, minerales y bajo en grasas; destacando como uno de los alimentos más importantes en la nutricional mundial. Por otro lado, los principales productores de papa son los países desarrollados con más del 50% de participación en el mercado, ofreciendo tanto papa fresca como procesada, esta última en anteriores años ha presentado un gran crecimiento debido a los cambios que los consumidores vienen adquiriendo como papas congeladas, chips, entre otros. (Ministerio de Agricultura y Riego [MINAGRI], 2020).

Mientras tanto en el Perú es el segundo tubérculo de mayor importancia, solo por debajo del arroz, teniendo una participación del 10% en el Producto Bruto Interno (PBI) para el 2019 según MINAGRI, su valor se debe a dos factores principales: (a) El primero porque es fuente de ingresos de más de 710 mil familias, y (b) el segundo porque no hay ninguna comida en el país inca donde no se incluya la papa. (MINAGRI, 2020).

Actualmente el 45,3% de producción de papa se presenta en la sierra (Apurímac, Arequipa, Ayacucho, Cusco, Moquegua, Puno y Tacna), y el restante en las zonas altas del centro y norte del país; por consiguiente, el 95% de la producción de la papa se concentra en la cordillera y solo un 5% en la costa peruana, Asimismo, las malas campañas de siembra y cosecha del tubérculo perjudican a las personas de escasos recursos, lo cual se manifiesta durante el primer semestre del año con una mayor oferta en comparación a la demanda, dándose precios de venta por debajo de los costos, siendo la producción del 73%, y solo un 27% para el segundo semestre del año; funcionando así como un termómetro medidor de la pobreza (MINAGRI, 2020)

Además, manifiesta dificultades de asociatividad para la compra de insumos, producción y comercialización; debido principalmente a que el 83% está conformado por pequeños productores que cultivan menos de cinco hectáreas, elevándose los costos de siembra, en consecuencia, teniendo un poder de negociación bajo en toda la cadena logística. En cuanto, a los rendimientos de producción son del 15%, siendo uno de los más bajos a nivel América, ocupando el puesto 16 de 21 países; debido principalmente a los cambios de clima que se presentan en las zonas altas,

como la presencia de granizado, sequias y heladas, que influyen en los volúmenes de oferta de la papa; por último, otro factor que también repercute es la falta de tecnología utilizada durante la siembra y cosecha del tubérculo, ya que se siguen empleando técnicas prehispánicas, como el camellones, andenes y concha, etc. (MINAGRI, 2020).

Por lo expuesto anteriormente, el Ministerio de Agricultura y Riego creó el sistema de información de cultivo cuyo objetivo es el brindar información sobre la intención de siembra en las diferentes regiones comparándolas con el promedio histórico de labranza en las últimas cinco campañas, brindando sugerencias del tipo de cultivo a sembrar, así como también la fecha exacta para plantar; sin embargo, estos datos no son comunicados a tiempo a los agricultores de las zonas altas del país. Asimismo, el MINAGRI viene planificando y creando el Instituto Nacional de Innovación Agraria (INIA), con los objetivos de fomentar ideas innovadoras buscando proporcionar información a los agricultores sobre los beneficios de adquirir semillas de calidad, con el propósito de incrementar el ingreso por hectárea en 31% y fomentar el riego tecnificado por ser más resistentes a las plagas o enfermedades; sin embargo, estos proyectos todavía se encuentran en la etapa de difusión de información. (Instituto Nacional de Innovación Agraria [INIA], 2019)

En efecto, las soluciones planteadas por los organismos encargados no son del todo eficientes, principalmente porque el comportamiento del mercado es regulado en su totalidad por la oferta y demanda, es así que en el período de enero a junio existe un exceso de producción presentándose los índices de precios más bajos del año; en consecuencia, los agricultores toman la decisión de disminuir la producción y la calidad de papa. De otro lado, en el siguiente semestre debido a la poca oferta, este valor de venta sube presentándose un ciclo de subidas y bajas en los precios donde el principal afectado es el primer escalón de la cadena productiva de la papa; por consiguiente, no existe una estabilidad que le permita crecer y desarrollarse como empresario, a fin de generar nuevas ideas de procesamiento industrial de papa como se demuestra en los países desarrollados.

Por consiguiente, la presente investigación pretende calcular el comportamiento futuro de la papa, a través de los modelos ARIMA y GAM para finalmente compararlos mediante el procesamiento de los datos en el programa R, que es reconocido como el mejor programa estadístico para una planificación óptima de las actividades de producción de papa en nuestro país.

De otro lado, el estudio trabajó con una base de datos de 120 meses, comprendido en el período de años entre el 2011 y 2020; asimismo, que esta información sea respecto al precio de la papa por kilo al por mayor, recolectado de la base de estadísticas del MINAGRI; y, por último, que se estimara valores futuros por los siguientes 24 meses. Es importante destacar que cada modelo presenta una metodología distinta que será explicada a continuación: (a) el modelo ARIMA debe cumplir con la estacionariedad, determinar la cantidad de autorregresivos y medias móviles necesarios que permitan realizar la predicción más adecuada para la presente investigación y (b) en el modelo GAM, primero se establecerá el tipo de distribución adecuada para el modelo, después se analizará la relación entre variables, luego la ejecución del mejor sistema, y por último realizar el pronóstico.

1.2. Variables

Tabla 1 Cuadro de variables

Tipo de Variables	Variabes	Dimensiones
Variable independiente	Sistema ARIMA y Modelo Aditivo Generalizado (GAM)	- Series de tiempos.
Variable dependiente	El comportamiento futuro del precio.	- Valor de la papa promedio al por mayor. - Datos brindados a nivel Perú. - Información mensual del precio entre los años 2011 al 2020.

Elaboración: Propia

1.3. Objetivos

1.3.1. Objetivo general

Calcular y comparar el comportamiento futuro del precio de la papa, entre los pronósticos de los modelos ARIMA Y GAM en el PERÚ, en el periodo 2011 – 2022.

1.3.2. Objetivos específicos

OE 01. Definir la metodología adecuada para los modelos ARIMA y GAM.

OE 02. Validar los modelos ARIMA y GAM, sobre el comportamiento del precio de la papa en el Perú entre los años 2011- 2020.

OE 03. Predecir el comportamiento del precio de papa para el periodo 2021 al 2022 de los modelos ARIMA y GAM en el Perú.

OE 04. Comparar ambos métodos, determinando las principales ventajas y desventajas.

1.4. Formulación de problema

1.4.1. Problema general

¿Cómo calcular y comparar el comportamiento futuro del precio de la papa, entre los pronósticos de los modelos ARIMA Y GAM en el PERÚ, en el periodo 2011 – 2022?

1.4.2. Problemas específicos

PE 01. ¿Cuál es la metodología adecuada para los modelos ARIMA y GAM?

PE 02. ¿Cómo validar los modelos ARIMA y GAM, sobre el comportamiento del precio de la papa en el Perú entre los años 2011- 2020??

PE 03. ¿Cuál es el procedimiento para predecir el comportamiento del precio de papa para el periodo 2021, entre los modelos ARIMA y GAM en el Perú?

PE 04. ¿Cuáles son las principales ventajas y desventajas que presentan los dos modelos?

1.5. Hipótesis

1.5.1. Hipótesis nula

El Modelo Aditivo Generalizado (GAM) muestra resultados eficientes de pronósticos del precio de papa en comparación con el método ARIMA, fundamentado en la literatura econométrica donde indica que los modelos lineales presentan una mayor efectividad en la explicación entre variables, mientras que el método no estocástico tradicional muestra mejores resultados de predicción en una serie de tiempo.

1.5.2. Hipótesis alternativa

El Modelo Aditivo Generalizado no muestra mejores resultados de pronósticos del precio de papa en comparación con el método ARIMA.

1.6. Justificación

La presente investigación es de suma importancia para los productores de papa en nuestro país, ya que brindará la estimación del comportamiento futuro del precio del tubérculo mediante los modelos ARIMA y GAM; de esta manera, el informe permitirá: (a) Contar con nuevas alternativas de solución a los problemas que presenta este sector; (b) beneficiar principalmente a los pequeños productores, ya que podrán mitigar los riesgos de sobreproducción, falta de poder de negociación; y (c) mejorar la calidad del tubérculo; generando así la sostenibilidad en el largo plazo, y el desarrollo del sector, a través de la incursión en nuevos rubros de procesamiento de papa.

En cuanto al aspecto social, el índice de mano de obra podría presentar crecimiento principalmente en las zonas altas de la región, disminuyendo de esta manera la tasa de pobreza en el sector, es así que ello provocaría un mayor movimiento de la economía. De otro lado, al contar con una mejor planificación de la producción de la papa, esta podría ser exportada o comercializada en mayores cantidades durante los meses de estacionalidad de las potencias en este rubro, pues en el Perú todo el año se siembra y cosecha este tubérculo.

Por último, debido a la insuficiente información sobre el pronóstico de precios de productos agrícolas en el Perú, con esta investigación se buscará dar el primer paso para que las diferentes carreras, universidades, investigadores, entre otros, puedan realizar este tipo de estudios que enriquezca más los conocimientos, tanto de aspectos cualitativos como cuantitativos, y brindando mejores alternativas a los agricultores y al Estado.

II. Capítulo: Revisión y fundamentación teórica

2.1. Antecedentes

Existen pocas investigaciones que estudien el comportamiento futuro de los precios de papa, específicamente en la ciudad de Arequipa, igualmente, no se ha realizado un diseño donde se obtenga el mejor pronóstico entre diferentes modelos de predicción. Sin embargo, a nivel nacional encontramos la publicación de Laurente (2017) sobre la aplicación del modelo ARIMA para la producción de la papa en la región de Puno. De otro lado, en el ámbito internacional existe una gran variedad de pronóstico de precios para los diferentes productos agrícolas, con una sola variable encontramos ARIMA o SARIMA y los multivariados que van desde los Modelos Aditivos Generalizados (GAM) hasta la aplicación de redes neuronales artificiales, siendo considerados para este estudio el modelo ARIMA y GAM. A continuación, se presentan los antecedentes más importantes y considerados para la realización del presente estudio:

Antecedente N°1:

Laurente, L. (2017): *Aplicación del modelo ARIMA para la producción de la papa en la región de Puno-Perú.*

El estudio permite conocer la modelación y proyección de la producción de papa en la región Puno, debido a la suma importancia que tiene para combatir la pobreza en esta región que señala un 24.2% según el año de estudio, los años toman en consideración que fueron desde 1950 al 2017.

Este pronóstico se realizó por el modelo Arima, basado en Criterio de Información de Akaike (AIC) y Criterio de Información de Schwarz (SC) se seleccionó el mejor modelo ARIMA de Box y Jenkins (1976).

Los resultados muestran que el modelo ARIMA (3, 0, 4) es el más apropiado para capturar el comportamiento y la predicción de la producción de papa, siempre cuando se trabaje con información anual, permitiendo de esta manera planificar la oferta del producto y generar mejores políticas en el sector agrícola.

Antecedente N°2

Sanchez, D. (2018): Modelo ARIMA para el pronóstico de la producción de cacao en el Perú 2012 al 2018 - Perú.

El estudio determinó un modelo que explique mejor el pronóstico de producción mensual del cacao, bajo el período de los años 2012-2017, y prediciendo desde agosto del 2017 hasta julio 2018.

Se realizó mediante el modelo de Box Jenkins, el cual definió las características, tendencias, estacionalidad y el futuro de la variable estudiada. Los resultados determinaron que era mejor utilizar el modelo SARIMA, debido que no se presentó estacionariedad sino estacional, cumplió con la normalidad, aleatoriedad de los errores, la homogeneidad de las varianzas de los errores por White y la independencia de los errores por Box Pierce. Por tanto, se logró obtener un error entre el 4% y 6% aproximadamente, encontrándose dentro de los intervalos de confianza. Para el mes de agosto del 2017 la producción será de 13,104 mil toneladas, 11,298 mil toneladas para septiembre, 10,374 mil toneladas para octubre, 10,890 mil toneladas para noviembre y 11,247 mil toneladas para diciembre.

Antecedente N°3

Lizarazo.C. (2016): Diseño de un instrumento de cobertura sobre el comportamiento del precio de la papa diacol capiro, para el pequeño y mediano productor, a partir del mejor pronóstico entre un modelo SARIMA y una RNA - Colombia.

El antecedente diseñó un instrumento financiero que permitiera garantizar al pequeño y mediano productor, la venta de sus cosechas, en consecuencia, se desarrollaron en dos modelos de predicción, el de SARIMA y el de una Red Neuronal Artificial. Las principales variables utilizadas en el estudio fueron el historial mensual de precios de papa de los últimos cinco años y la meteorología, a fin de determinar el rendimiento de producción cuando se presenta los fenómenos del niño y la niña. Los resultados obtenidos mediante el modelo SARIMA permitió explicar el comportamiento de la variación de los precios, valorado por un árbol binomial, teniendo un nivel de significancia del 0.0669 con un error del 0.0006. En cuanto a la aplicación de la RNA por medio de la técnica estadística de Análisis de Componentes Principales no se pudo completar el estudio.

Aunque no se logró concluir con el estudio de la Red Neuronal Artificial, cumplieron con el objetivo principal el de diseñar un instrumento de cobertura con el modelo SARIMA, considerando la técnica de arborización de mucha relevancia ya que permite conocer los precios del producto en diferentes escenarios y de esta manera poder realizar contratos a futuro asegurando una utilidad fija para el agricultor y que esta no se encuentre en riesgo por la variación de precios.

Antecedente N°4

Ruiz & Barrios. (2019): Análisis del precio de la manzana mediante un modelo SARIMA – México.

El estudio generó información del comportamiento de los precios de la manzana, mediante el análisis de series de tiempo durante los doce meses, en el modelo ARIMA, tomándose una muestra desde 1998 al 2017; para un mejor ajuste de los datos se transformó la serie en logaritmos. Al generar el modelo, obtuvo un margen de error del 2%, al comparar los datos reales con los conseguidos en el ARIMA, garantizados pronósticos relativamente buenos, el precio para julio del 2018 sería de \$31.99 cuatro puntos porcentuales más que el mismo mes del año 2017, distribuyéndose de forma normal los residuos.

Antecedente N°5

Rojalin, Debdutt, Sankar & Kanta. (2019): Pronostico del precio del maní mediante el modelo de series de tiempo - India.

Propuso un modelo ARIMA que permita predecir el precio del maní de manera mensual desde enero a diciembre del 2019, tomando datos desde 2004 hasta el 2018, con los criterios de MAPE y RMSE, considerando la mejor opción el ajuste a la variable permitiendo obtener mejores resultados. Se determinó que el precio en el año 2019 será estable durante todos los meses de esta manera el agricultor podrá tomar una mejor decisión en la presiembra y post siembra que le permita vender de manera rápida su producto, el estudio también busca generar interés en el estado para generar mejores leyes los cuales sean beneficiosos para los agricultores y por ende para toda la población.

Antecedente N°6

Broz & Viego. (2014): Predicción de precios de productos de Pinus SPP. con modelos ARIMA - Argentina.

El estudio evaluó la capacidad predictiva del modelo ARIMA en cuanto a los precios mensuales de multiproductos forestales (pulpable, aserrable fino, grueso y laminable) entre los años 2002 al 2013, obteniendo resultados al mes del siguiente, ya que el análisis funciona bien el corto plazo porque se pierde capacidad de pronóstico en el largo plazo.

Los resultados obtenidos son de satisfacción ya que se encuentran en un margen de error del 1% al 2%, pero se presentan ciclos estacionarios que deberían ser estudiados a mayor profundidad para perfeccionar el proyecto, debido a que en algunos periodos el error alcanzaba del 7% al 12%, específicamente cuando el precio se encontraba al alza. Por ello el estudio propone establecer políticas de registro sistemático y oportuno de los precios.

Antecedente N°7

García, Romero & Mayorga. (2015): Modelado del sistema logístico de la cadena productiva de la papa empleando dinámica de sistemas - Colombia.

El estudio explicó el comportamiento de la papa en las diferentes etapas de su producción y comercialización, permitiendo generar políticas para una mayor competitividad de la cadena desde la óptica del transporte. Las variables planteadas en el proyecto fueron las ventas, costos de transportes, costos de producción, la tasa de inversión, la oferta y demanda; los cuales fueron procesados en el diagrama de Forrester en un periodo de 60 meses.

Igualmente, el estudio indica que, al aumentar la utilidad, como resultado de una buena planificación y control de la distribución de la papa, permitirá contar con mayor inversión en tecnología, mejores ambientes de la conservación de la papa haciendo más atractiva en el mercado y generando una mayor demanda.

Antecedente N°8

Minaya. (2014): Análisis de la rentabilidad en la producción de papa blanca comercial en las regiones de Huánuco y Lima – Perú.

El estudio evaluó la rentabilidad de los productos de la papa blanca comercial en las regiones de Huánuco y Lima, empleando el método de simulación estocástica de Monte Carlo, la metodología considera los precios en chacra, los rendimientos obtenidos hectáreas y los costos de producción.

Asimismo, el estudio concluye que en Lima se tiene una rentabilidad superior en 8 veces más que en Huánuco; de esta manera, es más probable que se registren pérdidas económicas para los agricultores en esta última zona debido a los precios y rendimientos por hectárea, además de costos de producción, que incluye maquinaria, mano de obra, semillas entre otros. En otros términos, no es nada productivo cultivar papa en la ciudad de los caballeros de León mostrando mayores costos que ganancias.

Antecedente N°9

*Norza. (2011): Relaciones ontogénicas y espacio-temporales en la dieta del calamar gigante (*Dosidicus gigas*) en Perú, utilizando un Modelo Aditivo Generalizado, Lima – Perú.*

El estudio determinó que la dieta del calamar depende de la talla, tiempo, hora, temperatura y distribución espacial del animal, realizando el estudio a 4,178 calamares, la talla de la especie fue de 14 a 112 cm, en 2 a 299 millas náuticas de la costa entre los años 2004 al 2009, utilizando el Modelo Aditivo Generalizado.

Comprobando que la dieta es de gran variabilidad, según el modelo GAM se presenta significancia entre la variable predictora y respuesta, demostrándose que la llanura estomacal del calamar depende de la edad siendo los jóvenes los que consumen más alimentos, especialmente durante la noche por estar lejos de la costa y no se muestran cambios en cuanto a la temperatura.

Antecedente N°10

Aguilar. (2019): Modelos Aditivos Generalizados - España.

El estudio comprobó si la implementación del Modelo Aditivo Generalizado benefició en la toma de decisiones para la comercialización de las criptomonedas, teniendo como variable de respuesta, el precio de las criptomonedas; y como variables explicativas, cantidad, tasa de cambio tiempo y lagprecio.

Los resultados reflejaron que al ser este mercado muy volátil se tuvieron que realizar diferentes modelos siendo el que tuvo mayor fiabilidad el de precio \sim lagprecio + tiempo, a pesar que al comienzo del día refleja un alza de la moneda y al término de este disminuye; cabe destacar que en muchos de los casos no se refleja ese resultado. También se evidenció la importancia de capturar información en tiempo real y de obtener respuesta en el mismo instante.

Antecedente N°11

Santana & Mateos. (2014) El arte de programar en R: un lenguaje para la estadística - México.

El estudio explicó detalladamente el manejo del programa R, desde cómo crear una función, las diferencias entre número enteros, reales, logaritmos, qué tipo de tablas se acepta vectores, matrices, los diferentes gráficos que se obtiene según al estudio que se realiza.

Se trata de un estudio de suma importancia debido a que se consultó diferentes inquietudes que se tuvieron al momento de pasar a la práctica y poder llegar a los resultados del proyecto planteados en esta tesis.

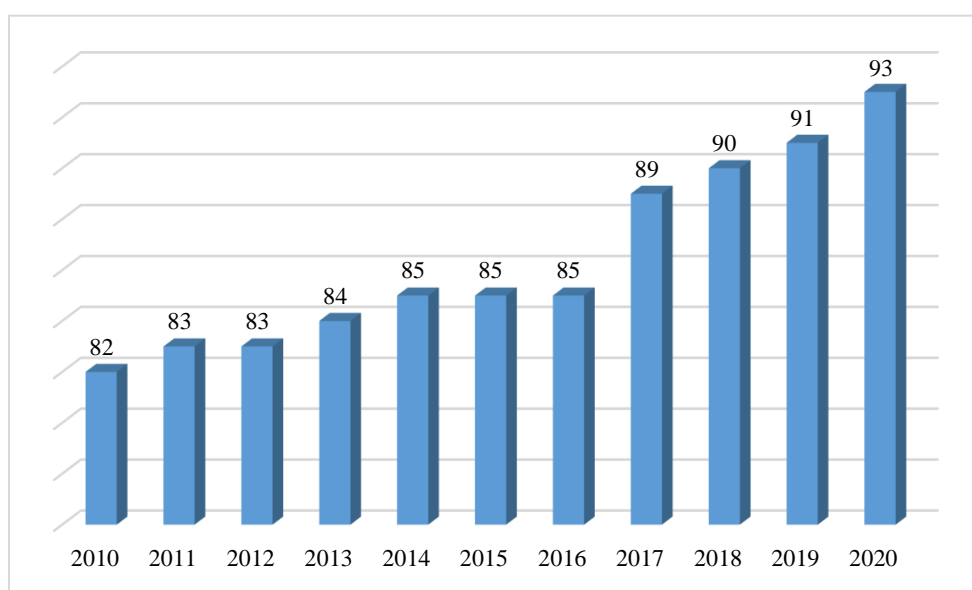
2.2. Bases teóricas

2.2.1. Papa

2.2.1.1. Consumo de papa

El Perú es el primer consumidor de papa a nivel Latinoamérica, ya que en la dieta de los peruanos no puede faltar el arroz y papa, anualmente se consume entre 54 y 90 kilos por persona, respectivamente. Por otro lado, es el principal ingrediente de la gastronomía nacional tanto de restaurantes y pollería; según cifras del Instituto Nacional de Estadística e Informática (INEI) en diciembre del 2019, este último representa el 40% del mercado nacional. (MINAGRI, 2020)

Ilustración 1 Consumo per capital de papa 2010 - 2020 por Kg.



Fuente: MINAGRI.

2.2.1.2. Producción

Los índices de cosecha, producción y rendimiento presentaron un aumento progresivo durante el período 2015-2019 (ver Tabla 2), de esta manera, en la sierra se muestra la mayor cultivación del tubérculo con un 54% del total del mercado, sin embargo, la costa presenta el mayor rendimiento en todo el Perú, debido al manejo del cultivo bajo seco y bajo riego respectivamente.

Tabla 2 Perú: Indicadores de producción de papa (2015 - 2019)

Año	Superficie de cosecha (Miles Ha.)	Producción (Miles To.)	Rendimiento (Kg/Ha.)
2015	316.5	4,715.9	14,899
2016	310.7	4,514.2	14,529
2017	310.4	4,776.3	15,388
2018	325.6	5,131.5	15,761
2019	330.0	5,331.1	16,156
Tasa de crecimiento promedio anual (2015 - 2019)			
	1.20%	2.20%	1.00%

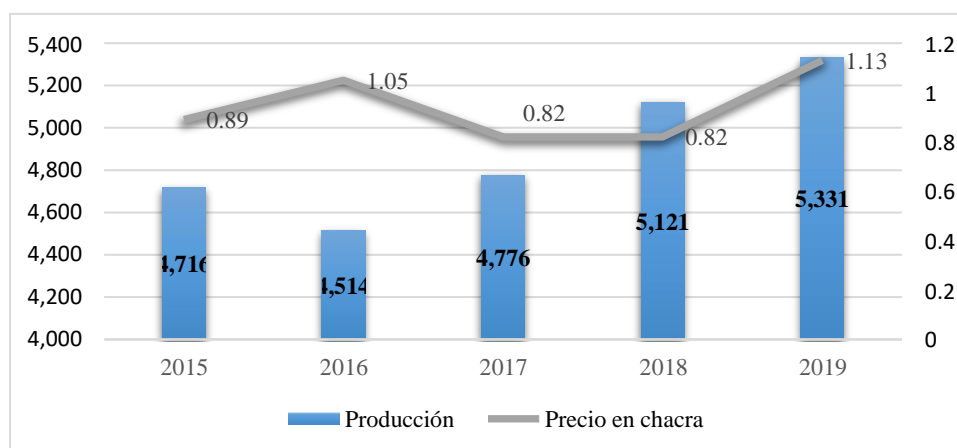
Fuente: MINAGRI – DGESEP.

Estos indicadores permiten generar 110,000 puestos de trabajo permanente principalmente en las zonas más altas de los 3,000 metros sobre el nivel del mar, no obstante, a nivel continental se ocupa el puesto dieciséis en cuanto al rendimiento de las cosechas, a pesar que nos encontramos en la tercera posición de producción, debido a que el rendimiento de la cosecha depende del uso adecuado de los fertilizantes, abonos, sistema de riego, etc. (MINAGRI, 2020)

2.2.1.3. Precio en chacra

Los precios entre los años 2015 al 2019 tuvieron un crecimiento anual del 5%, para el primer año fue de 0.89 soles por kilo presentando un crecimiento del 17% en comparación con el siguiente año debido principalmente a la disminución de la producción, para los años 2016 y 2017 se mantuvo en S/0.82 a pesar que la oferta siguió aumentando en un 7%; para lo cual, en el último periodo analizado, el precio subió en 37 puntos porcentuales teniendo la mayor producción de estos 5 años analizados. (MINAGRI, 2020)

Ilustración 2 Producción y precio en chacra de la papa Perú 2015 - 2019

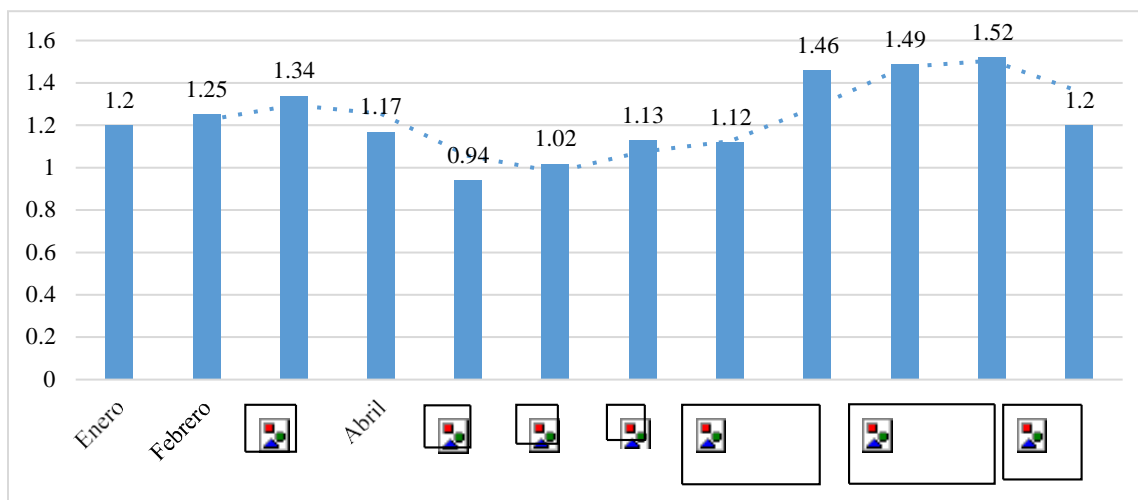


Fuente: MINAGRI – DGESEP – DEA.

2.2.1.4. Precio en el mercado mayorista de Lima

La mayor cosecha de papa se realiza durante los primeros seis meses del año siendo del 73% de la producción; mientras que el 27% restante se realiza durante el último semestre en promedio. Ello genera una sobreproducción de los meses de enero a junio, generando niveles de precios que oscilan entre 0.27 a 0.18 soles por kilo, para los años 2018 y 2019, respectivamente, no logrando cubrir los costos de producción; en consecuencia, se genera una correlación negativa; es así que esta situación se produce relativamente durante todos los años siendo el más afectado el primer eslabón de la cadena. Cabe señalar, que esta situación se ha visto afectada aún más, debido a la presencia del Covid – 19, que ha conllevado que los restaurantes, pollerías y comidas rápidas dejen de operar. (MINAGRI, 2020)

Ilustración 3 Evolución Mensual del precio (S/) de papa - Lima 2019.



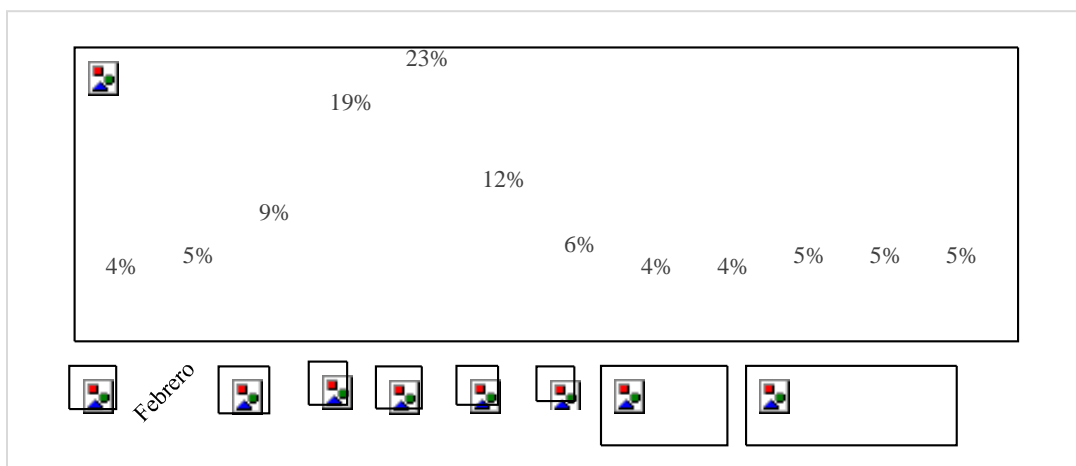
Fuente: DGESEP – SISAP – MINAGRI.

2.2.1.5. Estacionalidad

En el Perú todo el año se siembra papa debido a los diferentes climas que se cuenta en el país, pero es durante los meses de agosto y noviembre que se muestra la mayor estacionalidad con el 73% de participación aprovechándose la temporada de lluvia que son favorables para el cultivo de la papa. (MMINAGRI, 2020)

En cuanto a la cosecha, de la misma manera se presenta todo el año, pero se tiene una mayor oferta durante los meses de enero a junio que representa el 72% del total, siendo el mes de mayo, el período que reporta el mayor volumen de producción, específicamente en la región Puno logrando una concentración del 94% de cosecha durante los meses de marzo a mayo. (MINAGRI, 2020)

Ilustración 4 Estacionalidad de la producción de papa Perú



Fuente: MINAGRI – DGESEP – DEA

2.2.1.6. Variedades

Se dividieron en tres diferentes grupos, según su origen, color externo de la papa y uso.

Tabla 3 Variedad de papa por origen, color y uso

Criterios	Grupos	Características
Por su origen	Nativas	Harinosas: se consumen cocidas.
	Modernas	Menos harinosas se consumen cocidas o fritas.
Por su color	Blancas	Cascara blanquecina, crema o cremosa.
	de color	Cascara color rojo, rojizo, morado o bicolor.
Por su uso	Amargas	Para elaborar chuño, moraya o tunta.
	Amarillas	Sopas, papilla o puré.
	Industriales	Hojuelas (chips), para frita y otros

Fuente: La papa: producción, transformación y comercialización – Egústiza. B.R.

2.2.1.6.1. Papas nativas

Se siembra principalmente en la sierra a más de 3000 metros sobre el nivel de mar, obteniéndose más de 2500 tipos de variedades, cuenta con grandes beneficios por ser un buen antiinflamatorio, ayuda a combatir los problemas digestivos y son ricos en carbohidrato. (MINAGRI, 2020)

2.2.1.6.2. Papas modernas

Este tipo de papa resulta del cruce de dos o más papas, también son conocidas como mejoradas, híbridas o blancas, producción de este tipo 75 especies entre ellas tenemos a la cica, mariva, yungay, canchán, andina, etc. (MINAGRI, 2020)

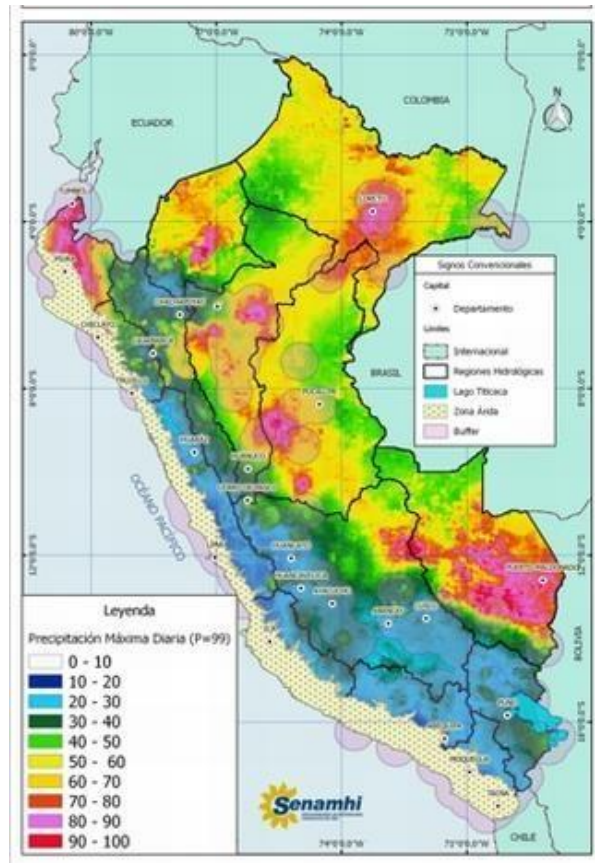
2.2.1.6.3. Papas industriales

Presentan un alto contenido de sólido (materia seca), bajo en azúcares, sembrándose principalmente en las zonas altas, las cuales son utilizadas para fritura, pre cocido o concentrado, puré y conservas, entre las principales especies encontramos a la desértica, capiro, maria bonita, tacna, primavera, única, etc. (MINAGRI, 2020)

2.2.1.7. Clima

El valor óptimo del clima para una buena campaña de producción de papa oscila entre los 15° a 20° C, ya que a una menor temperatura de 15° afectaría a la calidad del tubérculo; mientras a una temperatura mayor de los 28° afecta al crecimiento del tubérculo. De otro lado, el brillo solar requerido por estas plantas debe ser de 10 a 12 horas y la humedad debe ser menor al 80%. (MINAGRI, 2017)

Ilustración 5 Los diferentes climas en el Perú



Fuente: SENAMHI

2.2.2. Método ARIMA

2.2.2.1. Definición

La palabra ARIMA significa Promedio Móvil Integrado Autorregresivo por sus siglas en inglés, busca identificar, estimar y diagnosticar dinámica de series donde el tiempo juega un rol muy importante, la principal ventaja de esta técnica que no necesita diferentes datos, sino trabajando con valores pasados, llamado también como modelo univariado. (Arce & Mahia, 2003)

2.2.2.2. Proceso estocástico

Es un conjunto de observaciones aleatorias distintas que dependen de un argumento llamado tiempo, buscando describir la estructura probabilística para poder predecir y por último, el ordenamiento de las sucesiones es importante porque si no alteraría las características de la serie de tiempo. (Gonzales, 2009)

Siendo el proceso aleatorio representado por Y_t donde:

T = puede tomar cualquier número entre $-\infty$ y $+\infty$, representando el subíndice t

t = Paso del tiempo, por lo cual, para caracterizar un modelo estocástico se debe conocer la media y varianza de cada Y_t y la covarianza para variables diferentes a T . (Arce & Mahia, 2003)

2.2.2.3. Proceso estocástico estacionario

La teoría de predicción nos indica que se aprende de los comportamientos pasados pudiendo de esta manera poder determinar el futuro, para lo cual en los procesos estocásticos deben existir estabilidad, ya que se si las series de tiempo son inestables no se pueden utilizar para pronosticar. Según Gonzales (2009), la estacionariedad se puede caracterizar por la función de distribución o los momentos del proceso, los cuales son explicados a continuación:

2.2.2.3.1. Estacionariedad estricta: se presenta cuando se cumple la siguiente formula:

$$F[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}] = F[Y_{t_1+k}, Y_{t_2+k}, \dots, Y_{t_n+k}] \quad \forall (t_1, t_2, \dots, t_n) \text{ y } k$$

Es decir, la función de distribución de n variables aleatorias no se altera al desplazarse k periodos en el tiempo.

2.2.2.3.2. Estacionariedad en covarianza: el proceso estocástico Y_t , se cumple las siguientes condiciones:

Las variables aleatorias tienen idéntica la media y es finita.

$$E(Y_t) = \mu < \infty, \quad \forall t$$

μ = El valor esperado de la variable.

∞ = Infinito.

$\forall t$ = Para todos los tiempos.

Todas las variables tienen igual varianza y es finita en lo largo del tiempo.

$$V(Y_t) = E[Y_t - \mu]^2 = \sigma_y^2 < \infty, \quad \forall t$$

σ = Desviación típica de un conjunto de datos.

μ = El valor esperado de la variable.

∞ = Infinito.

$\forall t$ = Para todos los tiempos.

La autocovarianza solo es determinada por los números de periodos k de variables independientemente del tiempo en el que este.

$$cov(Y_t Y_s) = E[Y_t - \mu][Y_s - \mu] = \gamma_{|t-s|} = \gamma_k < \infty, \quad \forall k$$

μ = El valor esperado de la variable.

∞ = Infinito.

$\forall t$ = Para todos los tiempos.

Γ = Transformación integral de una variable real a una compleja.

S = Variable compleja.

2.2.2.4. Función de autocorrelacion

El coeficiente de autocorrelacion de periodos K en un proceso estocástico estacionario que mide la relación lineal entre las variables, siendo la esencial herramienta para conocer la estructura dinámica del modelo. Por otro lado, los correlogramas debe presentar decrecimiento hacia cero mientras se aumenta los periodos K . (Gonzales, 2009)

$$PK = \frac{cov(Y_t Y_{t+k})}{\sqrt{V(Y_t)}\sqrt{V(Y_{t+k})}} = \frac{\gamma_k}{\sqrt{\gamma_0 \gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

K = Periodos.

γ = Calculo de distribución de probabilidad.

t = Tiempo.

cov = Covarianza.

2.2.2.5. Ruido blanco

Según Gonzales (2009) este modelo es el más fácil del proceso estocástico, siendo una secuencia de variables aleatorias de medias cero, varianzas constante y covarianzas nulas, presentando una dependencia temporal entre sus variables, es decir lo que sucedió en el paso influye en el presente y este proceso es importante para la elaboración del modelo ARIMA.

$$t = 0, \pm 1, \pm 2, \dots :$$

$$E(a_t) = 0, \forall t \quad V(a_t) = \sigma^2, \forall t \quad Cov(a_t a_s) = 0, \forall t \neq s$$

∞ = Infinito.

$\forall t$ = Para todos los tiempos.

S = Variable compleja.

σ = Desviación típica de un conjunto de datos.

2.2.2.6. Procesos autorregresivos: AR(p)

Se da este término cuando las variables endógenas en un periodo t es expresada por las observaciones de ella misma determinados en tiempos anteriores y teniendo un margen de error. (Arce & Mahia, 2003)

Las condiciones generales según Gonzales (2009):

- El proceso no sea anticipado es decir que el presente no dependa del futuro, donde el valor futuro de Y o innovación a .
- Sea invertible es decir que depende de su pasado el ahora, es por ello que Y_{t-k} en Y_t debe ir reduciendo mientras nos acercamos al presente.

2.2.2.6.1. Modelo autorregresivo de orden 1, AR (1)

$$Y_t = \phi Y_{t-1} + a_t$$

Y_{t-1} = Esta determinado únicamente por su valor pasado

a_t = Perturbación contemporánea

ϕ = su valor es 1.61803398874989..., decimales infinito no periódicos.

Es una versión del modelo lineal generalizado restringido, así como también de las medias móviles en donde los infinitos coeficientes dependen de un solo parámetro. (Gonzales, 2009)

2.2.2.6.2. Proceso AR(p)

Según Gonzales (2009) el proceso autorregreso en p, indica que Y_T en función a su retardo hasta $t-p$ y una innovación contemporánea.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$$

Y_{T-1} = Esta determinado únicamente por su valor pasado y sucesivamente.

P = número de valores pasados.

a_t = Perturbación contemporánea

ϕ = su valor es 1.61803398874989..., decimales infinito no periódicos.

2.2.2.7. Proceso de medias móviles: MA(q)

Explica el valor de una determinada variable en tu tiempo y termino independiente, establecido por una secuencia de errores en un periodo precedente ponderados convenientemente. Es considerado también este modelo de corto plazo, ya que solo la perturbación es de un solo periodo. (Arce & Mahia, 2003)

2.2.2.7.1. Proceso MA (1)

Este modelo no es anticipante porque el presente no se influenciado por el pasado, para que sea invertible es necesario que el método autorregresivo sea convergente, teniendo que ser $|\theta| < 1$ teniendo que poner restricciones sobre el parámetro θ . (Gonzales, 2009)

$$Y_T = a_t - \theta a_{t-1}$$

a_t = Perturbación contemporánea.

Θ = Valor razonable.

2.2.2.7.2. Proceso MA (q)

Se deben incluir más de un retardo en el modelo, la memoria aumenta y la estructura de la dinámica por el método aumenta su riqueza. (Gonzales, 2009)

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Θ = Valor razonable.

q = Número de valores pasados.

a_t = Perturbación contemporánea.

2.2.2.8. Orden de integridad (I)

Es el número de diferencias que hay que tomar al proceso para conseguir la estacionariedad en media o también podemos decir las raíces unitarias del proceso. Las series temporales que encontramos con recurrencia son I (0) e I (1). (Gonzales, 2009)

$$\Phi_p(L)\Delta^d Y_t = \delta + \Theta_q(L)a_t$$

$\Phi_p(L)$ = Polinomio autorregresivo estacionario.

$\Theta_q(L)$ = Invertible de medias móviles.

δ = Cambio en la estadística de bondad de ajuste de la desviación.

Δ^d = La cantidad de cambio en una variable.

2.2.2.9. Las cuatro etapas del modelo ARIMA

- *Identificación:* Se utiliza los datos de proyecto en investigación para determinar cómo se generó la serie, intentando sugerir un subgrupo que permita interpretar mejor el ARIMA(p,d,q), de la misma manera conocer el orden del proceso (p,d,q) y conocer la cantidad de modelos que genero la serie. (Gonzales, 2009)
- *Estimación:* Usar de forma eficiente los datos de esta manera se realiza las inferencias sobre los parámetros condicionados, asegurando que el método utilizado es el apropiado. (Gonzales, 2009)
- *Validación:* Se realiza una comparación entre los modelos para estar seguros que los datos se están ajustando debidamente y si esto no seda saber cuáles son los problemas que se presenta para mejorar el análisis. (Gonzales, 2009)
- *Predicción:* Determinar cuál es el mejor modelo predictivo y conocer su capacidad de pronóstico. (Gonzales, 2009)

2.2.2.10. Beneficios del modelo ARIMA

Según Hernández (2015) entre las principales ventajas encontramos:

- Se puede utilizar de las diferentes ciencias como marketing, biología, geografía, etc.
- Realiza ideales predicciones en el corto plazo, cuando se cumple adecuadamente con los parámetros.
- Necesita pocos recursos para realizar el modelo y sencillo de ejecutar.
- Sirven de referencia para otros modelos de pronóstico.
- Derivan de una familia de métodos y no es solo un método.

2.2.3. Modelo aditivo generalizado (GAM)

Es una extensión del modelo lineal, teniendo una característica particular donde podemos verificar la contribución de cada variable predictora en la variable respuesta mediante suavizadores de covariables. Esta dada por la siguiente ecuación. (Alegre, 2011)

$$Y_i = \beta_0 + f(X_{ij}) + \varepsilon_i, \text{ donde } \varepsilon_i \overset{a}{\sim} N(0, \sigma^2)$$

Y_i = Variable respuesta.

β_0 = Parámetros.

$f(X_{ij})$ = Variable predictora mediante suavizadores.

ε_i = Residuo.

El modelo GAM cuenta con la función que relaciona la variable respuesta con la explicativa mediante suavizadores, de esta manera $f(X_i)$ se halla de la siguiente manera según Alegre (2011)

$$f(X_i) = \sum_{j=1}^p \beta_j \times b_j(X_i)$$

Σ = Sumatoria.

β = Parámetros.

$b_j(X_i)$ = Suavizadores.

Existen varias clases de suavizadores utilizados en el GAM, pero el más utilizados es el spline regresivo penalizado justamente como el nombre lo indica es dividido en dos partes, la primera es la regresión siendo ajustado por el método de mínimos cuadrados permitiendo determinar la cercanía del pronóstico a la realidad, está dada por la siguiente ecuación. (Wood y Augustin 2002).

$$\|Y - X\beta\|^2 = \lambda \int f''(x)^2 dx$$

β = Parámetros.

$\|Y - X\beta\|^2$ = Ajuste de la curva mediante la suma de mínimos cuadrados.

λ = Igual a la media y varianza.

$\lambda \int f''(x)^2 dx$ = Spline penalizado.

λ = Parámetro de suavizado.

f = Función de suavizado.

La fórmula de los mínimos cuadrados en la siguiente:

$$S = \sum_{j=1}^n (Y_j - u)^2 = \|Y - u\|^2 = \|Y - X\beta\|^2$$

j = Numero de observaciones.

$\| \quad \|$ = Debido a la norma Euclideana.

Y = Formato de datos en formato vector.

β = Parámetros.

X = Suavizadores.

Σ = Sumatoria.

Como se puede observar en las anteriores formular las ecuaciones buscan determinar los spline penalizados y la suavización, por lo tanto, debe haber equilibrio entre ambas curvas, prefiriendo utilizar los grados de libertad donde va desde 0 al infinito, mientras más grande sea el número obtenido indica que es menos lineal la suavización. (Wood y Augustin 2002).

Además, el modelo GAM pertenece a la familia exponencial siendo el enlace entre la variable respuesta y explicativa, se puede elegir según la base de datos a estudiar pudiendo ser continua, binaria, normal, etc. Teniendo como distribución Gaussiana, Poisson y Binomial negativa, este tipo de análisis es de suma importancia en las investigaciones exploratorias, predictivas, explicativas y también para determinar la relación entre los efectos covariables. (Alegre, 2011)

Tabla 4 Tipos de familias exponenciales

Familia	Función de enlace	Tipo de datos
Guassian	Identity	Residuales homogéneos y normales, los valores pueden ser positivo o negativos
Gamma	Inverse, identity y log	Numero reales (mediciones), los numero deben ser positivos.
Poisson	Log, identity y sqrt	Residuales no normales, variable respuesta con números enteros (conteos).
Binomial	logit	Variable respuesta binaria.
Binomial negativa	Identity, log y sqrt	Residuales no normales, variable respuesta con números enteros (conteos), teniendo excesiva sobredispersión.

Fuente: Instituto de Ciencias Antonia Brack

2.2.3.1. Beneficios de utilizar el modelo GAM

- No se necesita mostrar independencia entre variables.
- La distribución normal no es de relevancia en este tipo de modelo.
- Es un medio para calcular funciones no lineales.
- Capacidad predictiva mejor

2.2.4. Programación R

2.2.4.1. Definición

R es un lenguaje de programación de código libre siendo su entorno el análisis estadístico con gráficos de alta calidad, por lo cual los métodos numéricos aplicados son de manera sencilla y efectiva, compatible con Linux, Windows, MacOS, (Santana & Farfán, 2014)

2.2.4.2. Características

Según Santana & Farfán (2014) se presentan cuatro características importantes:

- El sistema base de R, que puedes bajar de CRAN y en todo lo demás, la programación y las funciones de software constan de paquetes modulares.
- Los gráficos en R son muy buenas y mejor que la gran mayoría de paquetes estadísticos comerciales.
- Es una herramienta de suma importancia para el trabajo interactivo, pero también un lenguaje poderoso para la elaboración de nuevas herramientas de programación.
- Por último, al ser de código abierto permite adaptarlos a nuestras necesidades, de la misma manera mejorar el programa y compartirlo con el público en general.

III. Capítulo: Método

3.1. Tipo de investigación

El presente estudio busca realizar la estimación del precio futuro de la papa, mediante los métodos estadísticos ARIMA y el Modelo Aditivo Generalizado (GAM), para luego ser comparados; en consecuencia, el proyecto presenta un enfoque cuantitativo, cuya meta será demostrar la hipótesis a través del cálculo de los métodos elegidos bajo los lineamientos de validez y confiabilidad por la comunidad científica. Del mismo modo, esta investigación es de conocimiento básica porque diagnosticó el problema de la sobreproducción de papa, que genera distorsión en el mercado perjudicándose principalmente a los productores.

Por último, el nivel de investigación es descriptiva longitudinal retrospectiva, ya que las variables estudiadas corresponden a períodos del pasado; por consiguiente, se utilizará la técnica data secundaria, otra importante característica es que se cuenta con un amplio compendio de teoría sobre series de tiempo, pero con muy escasas investigaciones empíricas; en consecuencia, el diseño del estudio es no experimental, ya que no se tomó datos de los fenómenos naturales de la poblacional a estudiar.

3.2. Población

La población de la presente investigación se compone por el precio promedio al por mayor de la papa reportados mensual, determinando los años a estudiar según los procedimientos de fines estadísticos, tomándose 120 datos entre los años 2011 - 2020, estos datos fueron extraídos de la siguiente página del Ministerio de Desarrollo Agrario y Riesgo-MINAGRI (<http://sistemas.midagri.gob.pe/sisap/portal2/ciudades/>)

3.3. Técnica e instrumento.

La siguiente investigación por ser de enfoque cuantitativa posee una gran cantidad de instrumentos de recolección de datos, por lo cual en la presente investigación se utilizará una guía de información documental considerando las variables de precio y producción de papa.

Tabla 5 Técnicas e instrumentos

Información a ser recolectada	Fuentes de información	Técnica de recopilación de la información	Instrumento de recopilación de información
Precio promedio al por mayor de la papa en el Perú.	Se refiere al término medio obtenido de los diferentes precios según el tipo de papa, esta información secundaria se extraerá del Sistema de Información de Abastecimiento y Precios (SISAP).	Técnica de investigación documental de procesamiento.	Ficha de procesamiento de datos electrónicos. Se realizará el registro de los últimos 10 años del precio promedio por mayor de la papa en periodos mensuales.

Elaboración propia

3.4. Matriz de consistencia

En el siguiente cuadro (ver Tabla 6) se resume los elementos básicos del proyecto de investigación, permitiendo tener un mejor panorama.

Tabla 6 Matriz de consistencia

			Operacionalización		
Problema	Objetivos	Hipótesis	Variables	Indicadores	Metodología, Técnicas e instrumentos
<p>Problema principal ¿Cómo calcular y comparar el comportamiento futuro del precio de la papa, entre los pronósticos de los modelos ARIMA Y GAM en el PERÚ, en el periodo 2011 – 2022?</p> <p>Problemas específicos PE 01. ¿Cuál es la metodología adecuada para los modelos ARIMA y GAM? PE 02. ¿Cómo validar los modelos ARIMA y GAM, sobre el comportamiento del precio de la papa en el Perú entre los años 2011- 2020? PE 03. ¿Cuál es el procedimiento para predecir el comportamiento del precio de papa para el periodo 2021, entre los modelos ARIMA y GAM en el Perú? PE 04. ¿Cuáles son las principales ventajas y desventajas que presentan los dos modelos?</p>	<p>Objetivo general Calcular y comparar el comportamiento futuro del precio de la papa, entre los pronósticos de los modelos ARIMA Y GAM en el PERÚ, en el periodo 2011 – 2022.</p> <p>Objetivos específicos OE 01. Definir la metodología adecuada para los modelos ARIMA y GAM. OE 02. Validar los modelos ARIMA y GAM, sobre el comportamiento del precio de la papa en el Perú entre los años 2011-2020. OE 03. Predecir el comportamiento del precio de papa para el periodo 2021 al 2022 de los modelos ARIMA y GAM en el Perú. OE 04. Comparar ambos métodos, determinando las principales ventajas y desventajas.</p>	<p>Hipótesis nula El Modelo Aditivo Generalizado muestra resultados eficientes de pronósticos del precio de papa en comparación con el método ARIMA, fundamentado en la literatura econométrica donde indica que los modelos lineales presentan una mayor efectividad en la explicación entre variables, mientras que el método no estocástico tradicional muestra mejores resultados de predicción en una serie de tiempo.</p> <p>Hipótesis alternativa El Modelo Aditivo Generalizado no muestra mejores resultados de pronósticos del precio de papa en comparación con el método ARIMA.</p>	<p>Variable independiente Sistema ARIMA y Modelo Aditivo Generalizado (GAM)</p> <p>Variable dependiente El comportamiento futuro del precio de la papa.</p>	<p>Para la variable independiente: Validación del modelo.</p> <p>Para la variable dependiente: a) Cantidad de cosecha de papa. b) Costo de transporte, insecticidas y semilla. c) Mano de obra (operarios en el campo).</p>	<p>Tipo de investigación Investigación básica: Se diagnosticó el problema de la sobreproducción de papa los cuales genera distorsión en el mercado perjudicándose principalmente a los productores.</p> <p>Diseño de investigación Descriptiva longitudinal retrospectiva: Las variables son estudiadas de años pasados utilizándose la técnica data secundaria.</p> <p>Población Esta conformo por el precio promedio al por mayor de la papa reportados mensual, determinando los años a estudiar según los procedimientos de fines estadísticos, tomándose 120 datos entre los años 2011 - 2020.</p> <p>Técnica Técnica de investigación documental de procesamiento.</p>

Elaboración propia

3.5. Software estadístico R

Es la plataforma más utilizada por la comunidad científica para realizar análisis estadísticos siendo de código abierto debido que se encuentra registrado en el Licencia Publica General(GPL) por lo cual, permite solucionar los problemas de programación de manera rápida, también cuenta con paquetes de métodos muy innovadores y diversos para censos, series de tiempos o cualquier proceso estadístico. (Mas, 2018)

La aplicación en el siguiente trabajo se realizará de la siguiente manera:

- Se instaló y activan los diferentes paquetes necesarios para la aplicación del método ARIMA y GAM.
- Seguidamente se procede a descargar la tabla de datos tanto de precio como de producción.
- Después se convertirá en una tabla de series de tiempo.
- Procediendo a realizar las diferentes asunciones teóricas requeridas para el método ARIMA y GAM.
- Luego de contar con los requisitos necesarios, se procede ejecutar las técnicas estudiadas por la presente investigación.
- Para terminar, corresponde ejecutar la predicción en ambas metodologías.

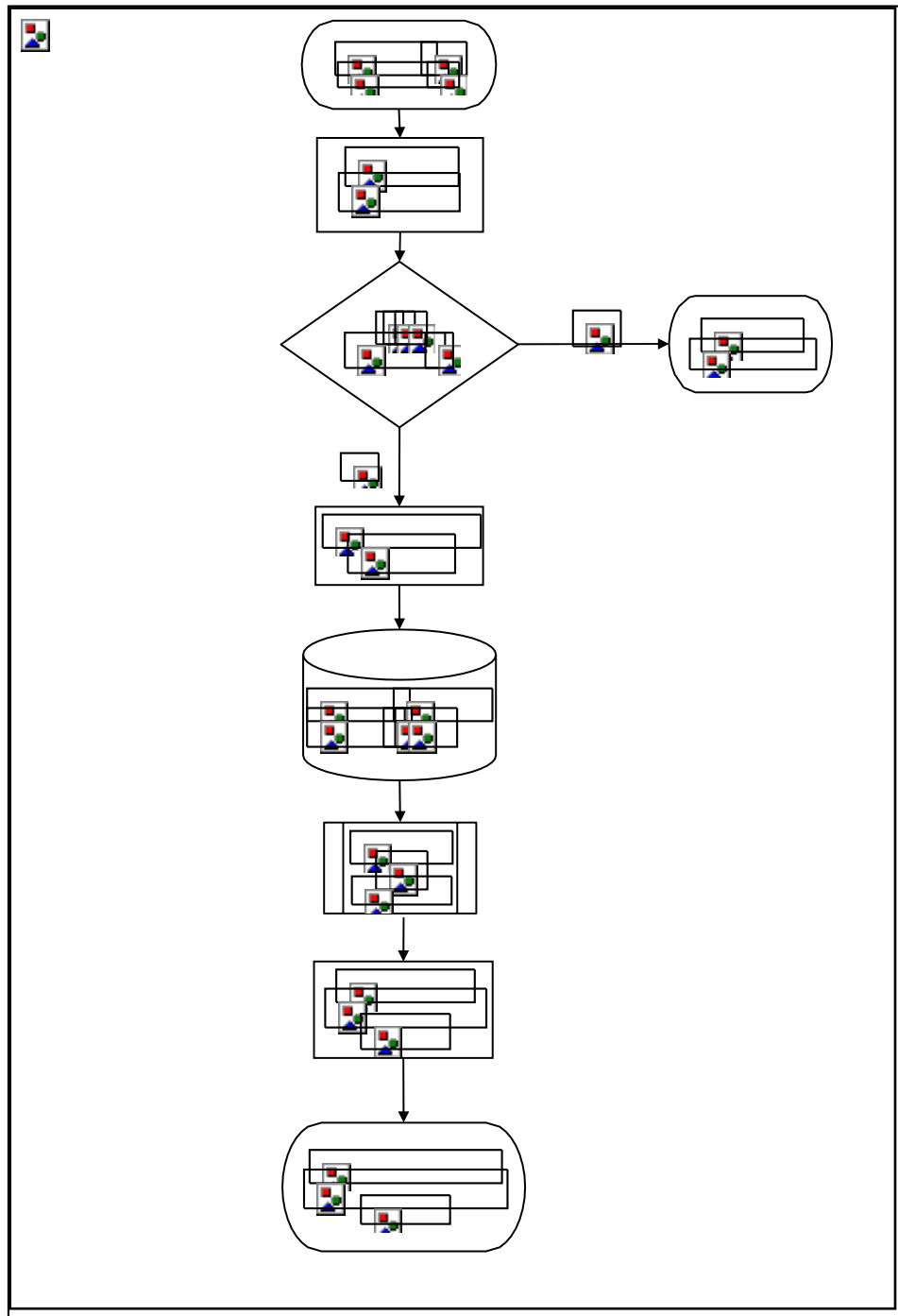
3.6. Procedimiento

El procedimiento metodológico de la investigación es el siguiente:

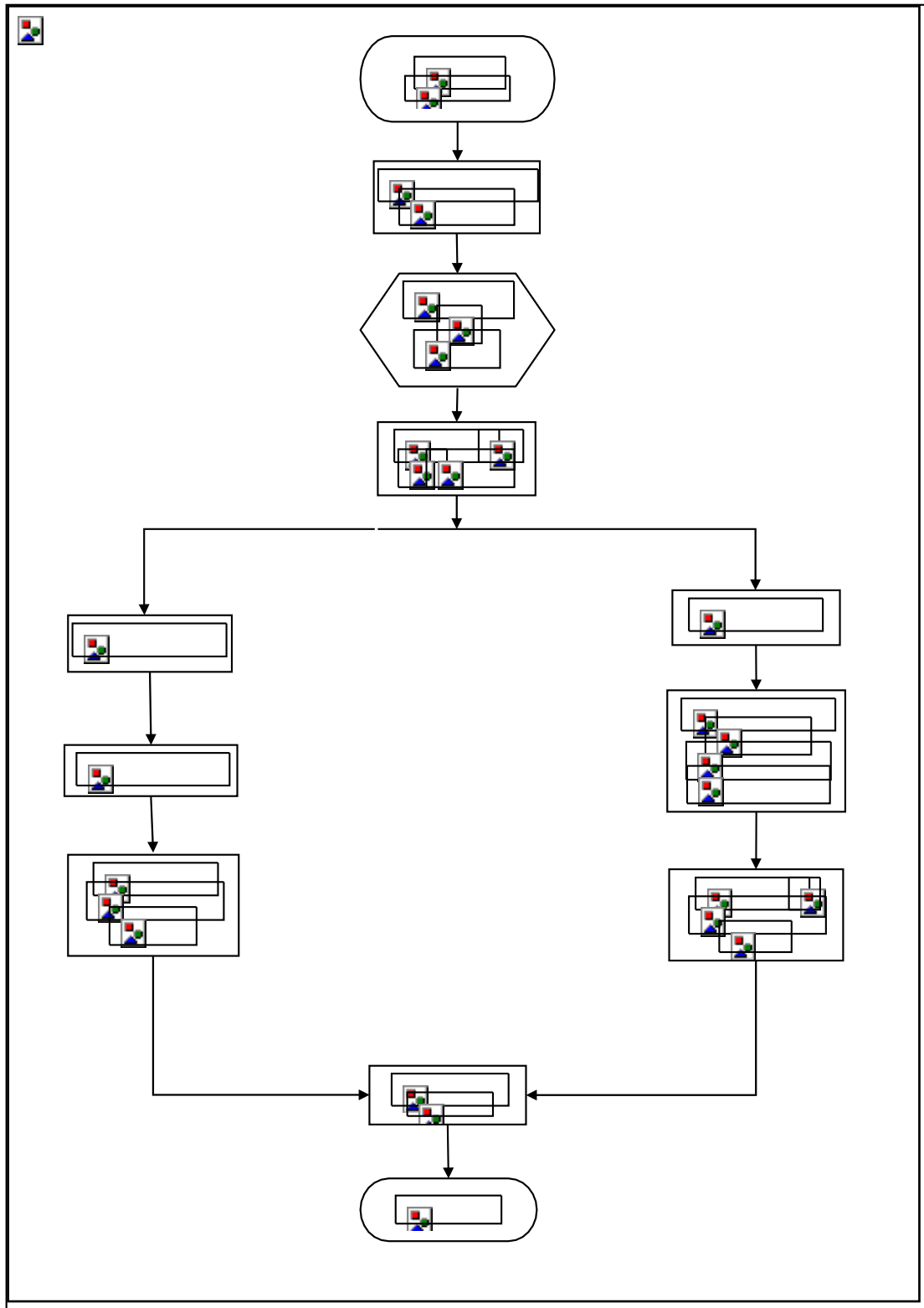
- Se organizó la información para conocer el estado del arte de la investigación, la bibliografía y la situación que se presenta actualmente sobre la problemática en la producción de papa, la planificación de cosechas realizado por la entidad encargada y las principales soluciones planteadas por otros investigadores de pregrado y posgrado en el Perú y a nivel internacional.
- Definición de la población la cual, es de 120 datos tomados mensualmente entre los años 2011 al 2020 del precio promedio al por mayor de la papa, también de los métodos a utilizar los cuales son: ARIMA y GAM.
- Estructurar los datos según a los requerimientos establecidos por cada técnica a utilizar en esta investigación.

- Procesamiento de los datos en el software R el cual es explicado con mayor detalle en la Ilustración N° 7.

Ilustración 6 Flujoograma general del procedimiento metodológico



Elaboración propia

Ilustración 7 Procesamiento de datos en R

Elaboración propia

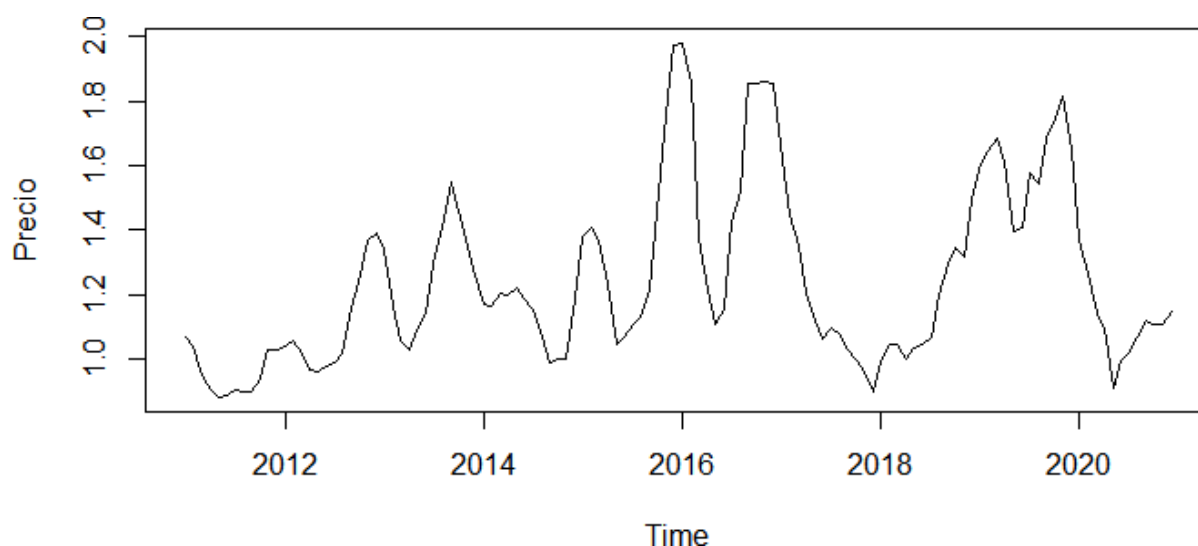
IV. Capítulo: Resultados y discusiones

4.1. Método ARIMA

4.1.1. Precios promedios mensuales al por mayor de la papa en el Perú

En la siguiente grafico (ver Ilustración 8) se puede observar los diferentes precios promedio que se obtuvo de la papa entre los años 2011 al 2020, donde las 4 primeras temporadas el monto máximo fue de 1.4 soles, el periodo 2016 tuvo el valor más alto de toda la serie llegando casi a los S/2.00, siendo para el último año analizado uno de los importes más bajos el cual oscila en un 1.20 soles.

Ilustración 8 Variación del precio de papa (S/ - Kg) entre los años 2011 al 2020



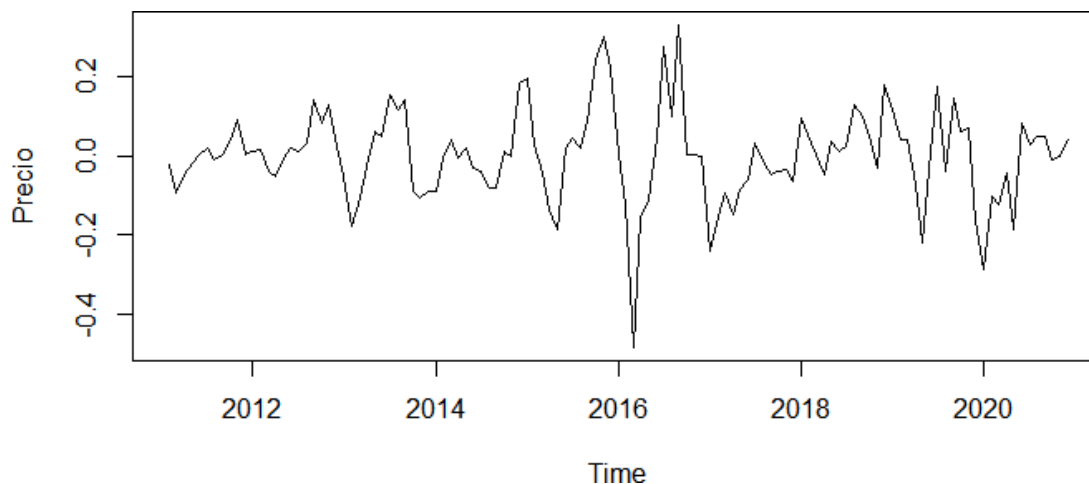
Elaboración: Propia

4.1.2. Estacionariedad

El primer paso para realizar un modelo ARIMA es verificar la estacionalidad de los datos el cual, consiste en que todos los numero estén sobre una misma media, observando en la ilustración 8 no se presenta este requisito debiendo transformarlos ya sea con logaritmos o diferencias, en esta investigación se trabajó mejor con el ultimo termino mencionado.

Se realizó la primera diferencia I (1), apreciándose en la ilustración 9 la existencia de estacionariedad, en efecto los datos sobre el precio de la papa durante los años 2011 al 2020 oscilan en la pendiente cero, para contar con mayor credibilidad se ejecutó la prueba de Dickey y Fuller (1979) donde el proyecto brindo un resultado de $p - \text{valor} = 0.01$ siendo menor a 0.05, es decir significativo en estacionaria uno.

Ilustración 9 Proceso estocástico del precio de papa (S/ - Kg) 2011 – 2020



Elaboración: Propia

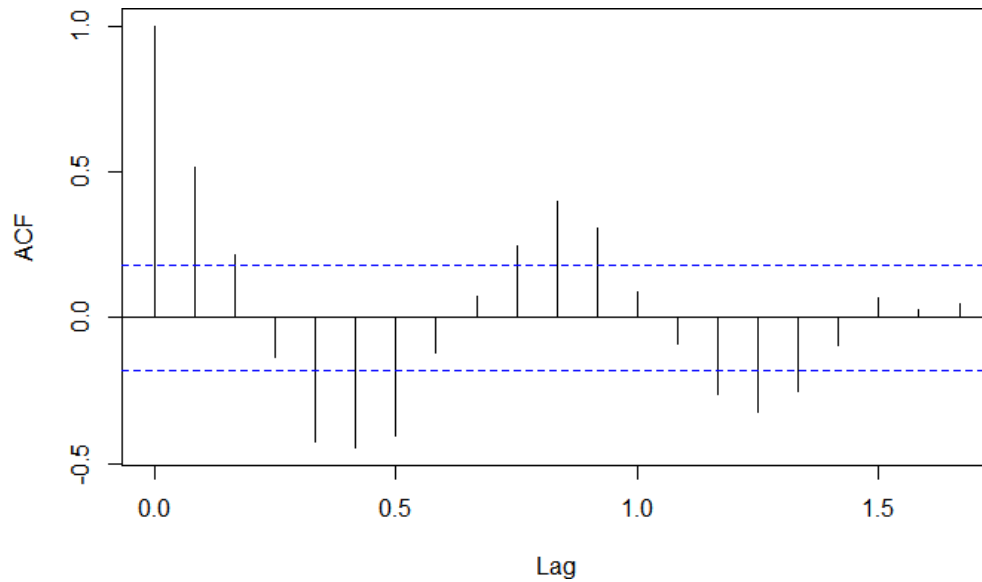
4.1.3. Estimación del mejor modelo ARIMA

4.1.3.1. Modelo de autocorrelación y parcial autocorrelación

Permite conocer el número de autorregresivos y también las medias móviles necesaria para determinar el mejor modelo ARIMA.

La cantidad de medias móviles necesarias en el modelo se obtuvo realizando un sistema de autocorrelación el cual se aprecia en la ilustración 10, distinguiendo que los conjuntos de barras de los primeros datos sobresalen de las líneas punzocortantes azules determinando el número de medias móviles igual a (2).

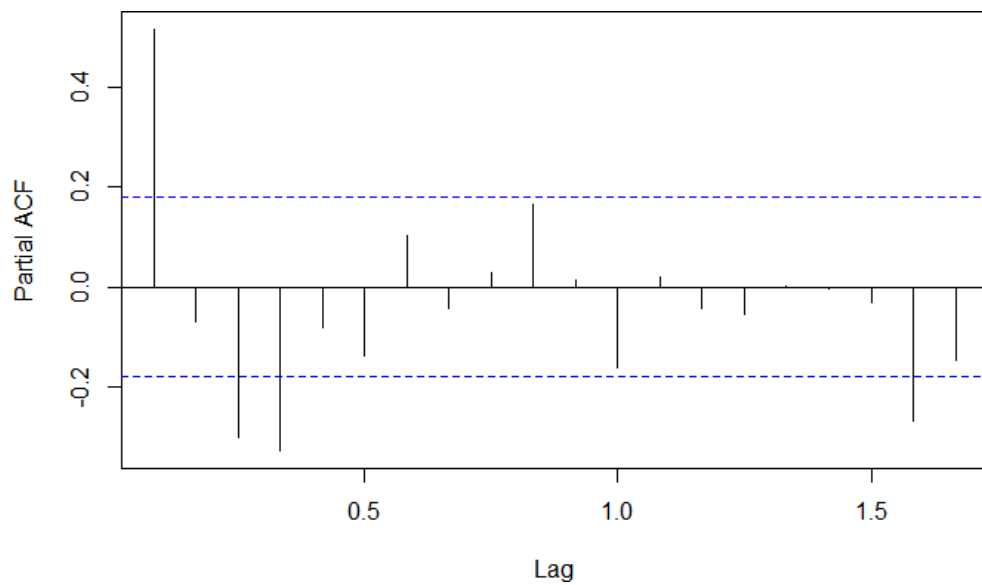
Ilustración 10 distribución de los elementos de medias móviles (autocorrelación)



Elaboración: Propia

Por otro lado, se determinó la cantidad necesaria de autorregresivos, mediante el modelo de parcial autocorrelacional siguiendo la misma dinámica presentada anteriormente, verificando en la ilustración 11, generar (2) autorregresivos para obtener un buen sistema.

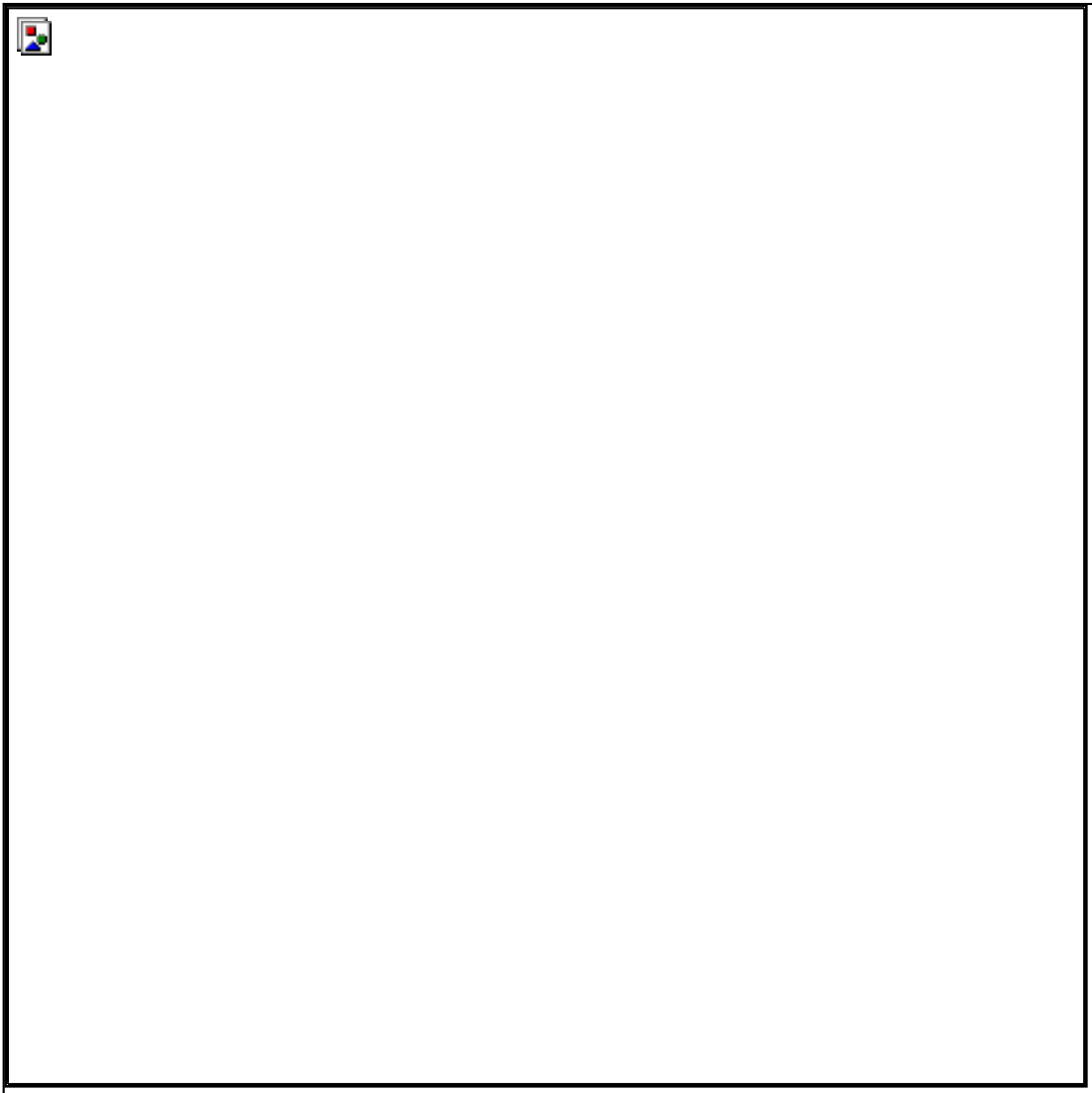
Ilustración 11 La distribución del parcial autocorrelacional para generar los autorregresivos



Elaboración: Propia

Otra manera más rápida y sencilla de conocer el número necesario de autorregresivo $AR(p)$, medias móviles $MA(q)$ y diferenciación también llamado como orden de integridad (d) es a través del paquete “auto.arima” encontrándolo en el programa R, mostrándonos (ver Ilustración 12) varias interacciones entre los tres componentes principales del modelo, para determinar la mejor elección, se elige el sistema que se acerca al cero, obteniendo el $ARIMA(2,1,2)$ corroborando la información anterior presentada.

Ilustración 12 Los diferentes modelos ARIMA obtenidos en R



Elaboración: Propia

4.1.3.2. Pronostico del modelo ARIMA (2,1,2)

Comenzaremos corriendo el modelo con la función que tiene el mismo nombre ARIMA, permitiendo visualizar los coeficientes a utilizar por cada media móvil y autorregresivo para obtener el pronóstico de los precios. (ver Tabla 7)

Tabla 7 Los coeficientes obtenidos del mejor modelo ARIMA

ar1	ar2	ma1	ma2
1.5114	-0.8739	-1.1303	0.5215
0.0628	0.0605	0.1131	0.121

Elaboración: Propia

Luego se procedió a realizar el pronóstico para el año 2021 – 2022 (ver Tabla 8), el cual nos brinda tres tipos de información, la primera proporciona el precio futuro de los siguientes 24 meses, las siguientes dos columnas para el 80% y 95% de confianza respectivamente el valor mínimo y máximo que puede variar el monto por cada periodo.

Tabla 8 Pronostico de precio de la papa entre los años 2021 - 2022

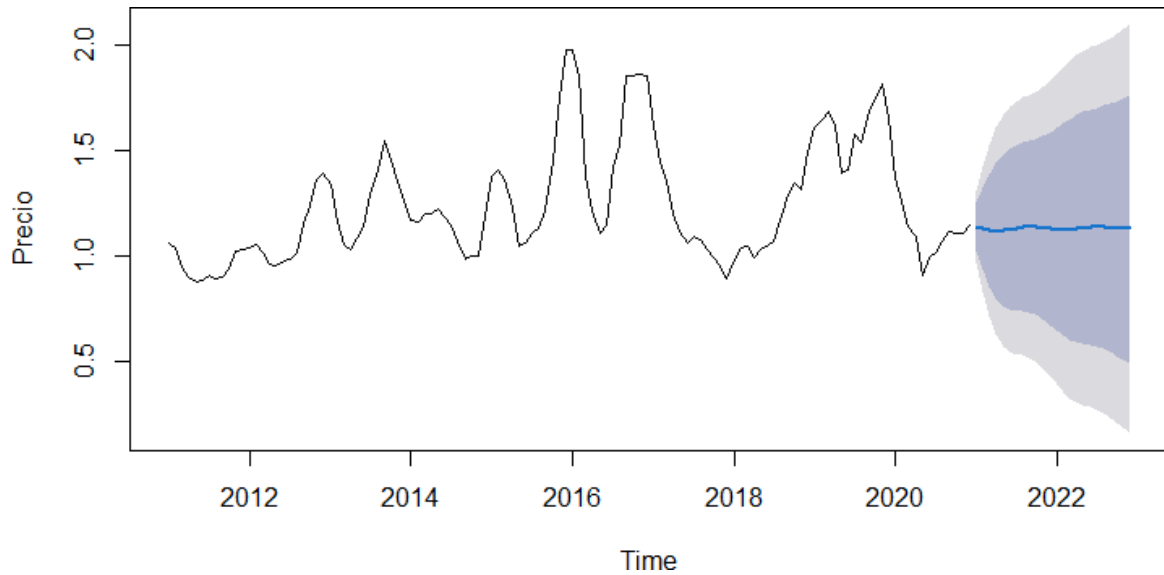
Fecha	Predicción	Mínimo al 80%	Máximo al 80%	Mínimo al 95%	Máximo al 95%
Ene-21	1.14	1.03	1.25	0.97	1.31
Feb-21	1.13	0.94	1.32	0.83	1.42
Mar-21	1.12	0.86	1.38	0.72	1.52
Abr-21	1.12	0.8	1.44	0.63	1.61
May-21	1.12	0.76	1.48	0.57	1.67
Jun-21	1.13	0.75	1.51	0.55	1.71
Jul-21	1.13	0.74	1.53	0.54	1.73
Ago-21	1.14	0.74	1.54	0.53	1.75
Set-21	1.14	0.73	1.55	0.52	1.76
Oct-21	1.14	0.72	1.56	0.50	1.78
Nov-21	1.13	0.70	1.57	0.47	1.80
Dic-21	1.13	0.67	1.59	0.43	1.83
Ene-22	1.13	0.65	1.61	0.39	1.86
Feb-22	1.12	0.62	1.63	0.35	1.90
Mar-22	1.13	0.60	1.65	0.33	1.93
Abr-22	1.13	0.59	1.67	0.31	1.95
May-22	1.13	0.58	1.68	0.29	1.97
Jun-22	1.14	0.58	1.69	0.28	1.99
Jul-22	1.14	0.57	1.70	0.27	2.00
Ago-22	1.14	0.56	1.71	0.25	2.02
Set-22	1.13	0.54	1.72	0.23	2.03
Oct-22	1.13	0.53	1.73	0.21	2.05
Nov-22	1.13	0.51	1.75	0.18	2.08
Dic-22	1.13	0.49	1.76	0.16	2.10

Elaboración: Propia

Entre los principales resultados que muestra la tabla 8 tenemos que el precio mínimo sería de 0.49 y máximo de 1.76 soles para nivel de confianza del 80%, mientras que el valor de 0.16 es el más bajo y la mayor tasación de 2.10 soles con una confianza de 95%, podemos indicar que mientras mayor sea el nivel de confianza el valor mínimo disminuye considerablemente del mismo modo para el incremento, teniendo una relación directa.

Además, en la ilustración 13 muestra un gráfico histórico del precio de la papa desde el 2011 al 2022, donde los últimos dos años representan al pronóstico presentando una pendiente más recta en comparación a los años pasados, pero el sombreado azul indica el mínimo y máximo valor por el cual podría fluctuar el precio.

Ilustración 13 Pronostico de precio de papa entre los años 2021 - 2022

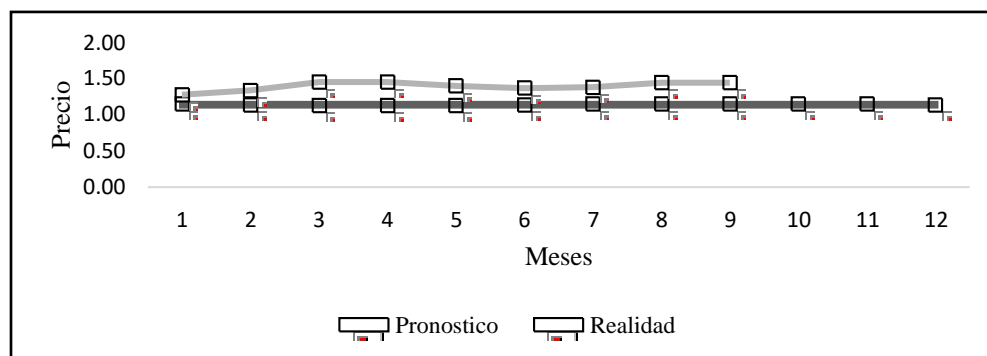


Elaboración: Propia

4.1.3.3. Comparación entre realidad y pronóstico 2021

Se realizó la comparación entre el pronóstico y los precios obtenidos por el SISAP donde se presentan variaciones en el precio por 0.25 céntimos de soles en promedio. También podemos observar en el gráfico (ver Ilustración 14) que la predicción representa similitud con una línea recta, frente a los cambios que se dan en el valor real de la papa durante los meses de enero a septiembre del 2021.

Ilustración 14 Comparación entre pronóstico y realidad del precio de la papa año 2021



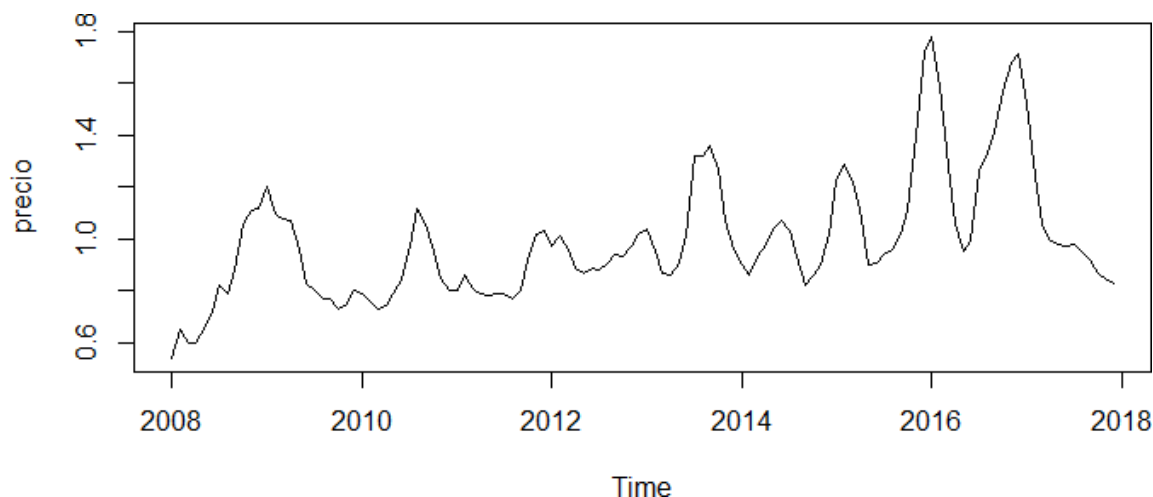
Elaboración: Propia

4.1.3.4. Comparación entre la realidad y pronóstico de la papa canchán 2018 - 2020

A consecuencia de lo expuesto anteriormente se decidió realizar otro modelo ARIMA para una de las variedades más consumidas en el Perú la papa canchán según el Instituto Nacional de Estadística e Informática (INEI), con la intención de corroborar la información obtenida en la ilustración anterior sobre la comparación con los datos pronosticados y reales o tal vez mostrar la pérdida de veracidad por la generalización del precio de la papa.

Análogamente primero se visualizarán los precios mensuales entre los años 2008 al 2017 con el propósito de examinar los tres siguientes periodos, observando (ver Ilustración 15) que no se presenta estacionariedad.

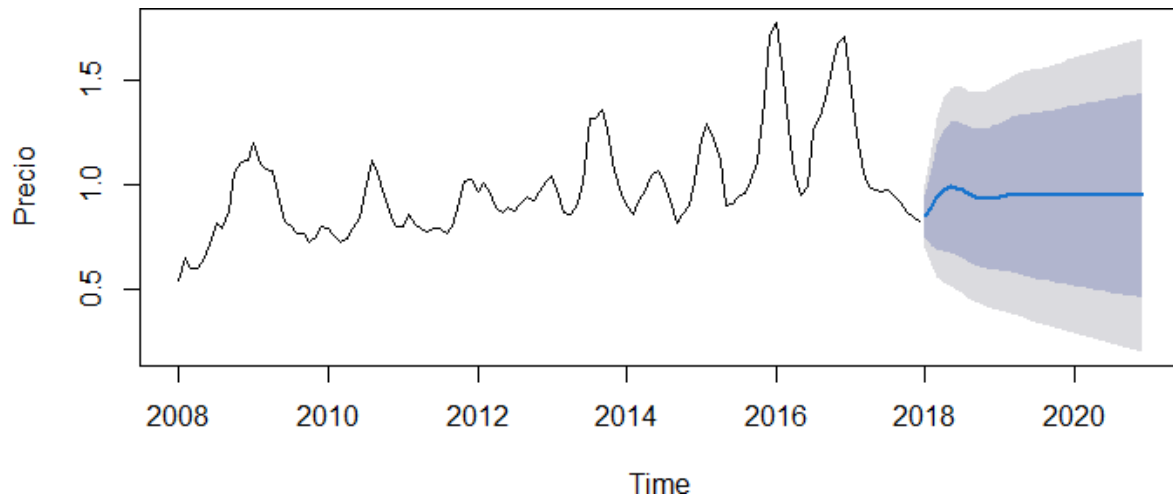
Ilustración 15 Evaluación del precio de papa canchán por kilogramos del 2008 al 2017



Elaboración: Propia

Luego se procedió con la determinación del sistema más adecuado que permita conocer los precios futuros de la papa canchán por kilo al por mayor, siendo el ARIMA (2,1,1) a utilizar para la predicción de los años 2018 al 2020, pudiendo divisar (ver Ilustración 16) una línea recta para los precios pronosticados según la línea azul.

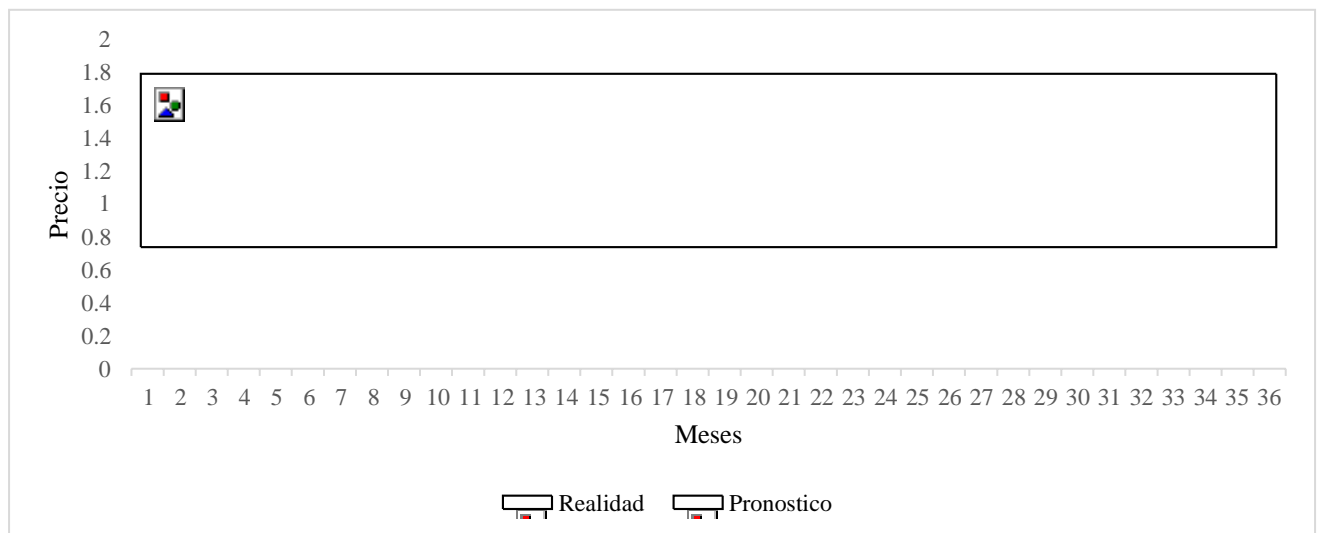
Ilustración 16 Pronostico de precio de papa canchan para los años 2018 - 2020



Elaboración: Propia

Continuando con el análisis se realizó la comparación entre los valores futuros obtenidos y los precios presentados entre los años 2018 al 2020, obteniendo como predicción una línea recta frente a los crecimientos y disminuciones que se dieron en esos 36 meses. (ver Ilustración 17)

Ilustración 17 Comparación entre pronóstico y realidad del precio de papa canchan 2018 - 2020

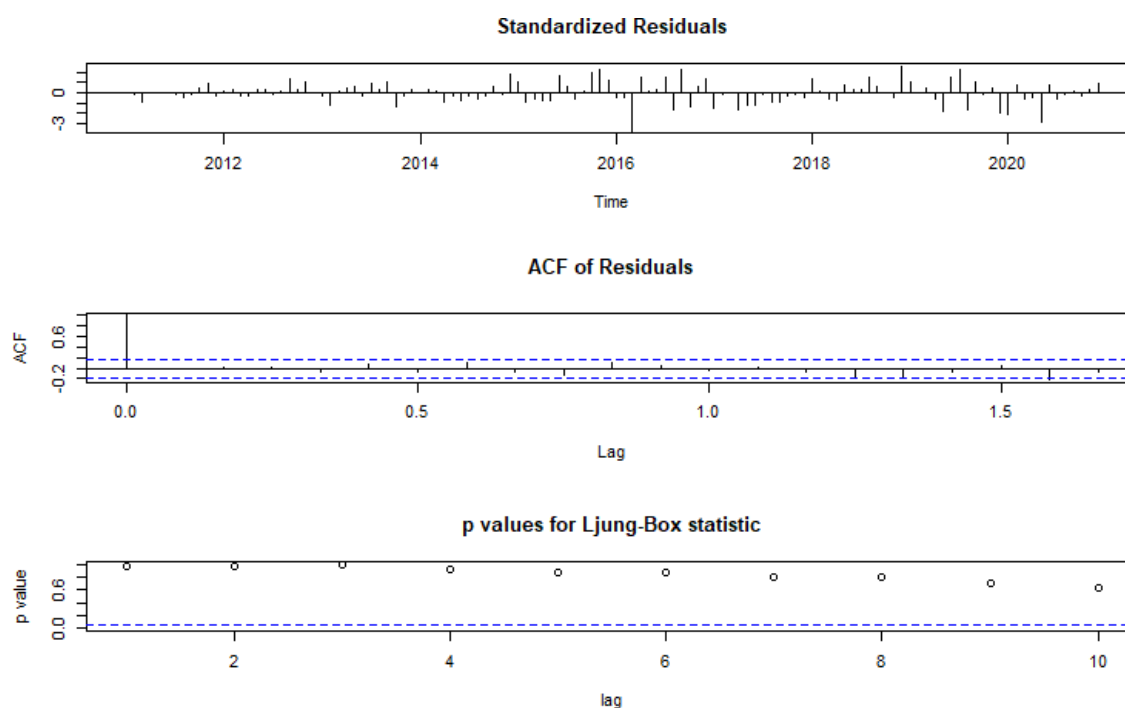


Elaboración: Propia

4.1.4. Validación

Se efectuó la validación del modelo a través de la prueba Ljung y Box (1978) obteniendo como resultado el ruido blanco el cual es significativo si es mayor a 0.05 manifestando que la media es igual a cero, la varianza constante y el error estándar no está seriamente correlacionado. Por lo tanto, en la presente investigación se alcanzó un resultado del 0.97 indicando que el modelo se ajusta notablemente, apreciándose en el siguiente grafico (ver Ilustración 18), donde se observa que se cumplen con todos los parámetros establecidos líneas más arriba del presenta párrafo.

Ilustración 18 Validación del modelo ARIMA (2,1,2)



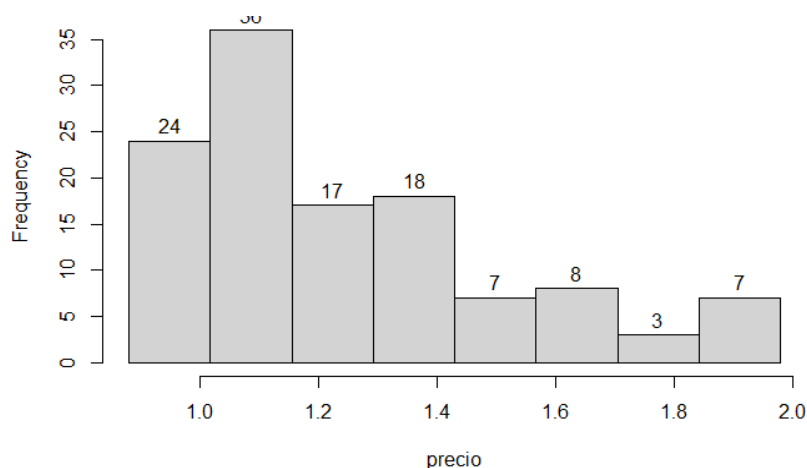
Elaboración: Propia

4.2. Modelo Aditivo Generalizado

4.2.1. Tipo de distribución de probabilidades

En el Modelo Aditivo Generalizado primero se debe establecer el tipo de familia exponencial que permita en la presente investigación enlazar las variables respuestas con sus predictores, determinando los precios mínimos y máximos entre los años 2011 al 2020 para después contar la cantidad de unidades por cada valor, los cuales van desde un S/ 1.00 a S/ 2.00s obteniendo el siguiente histograma (ver Ilustración 19).

Ilustración 19 Histograma de frecuencia del precio de la papa entre 2011 - 2020

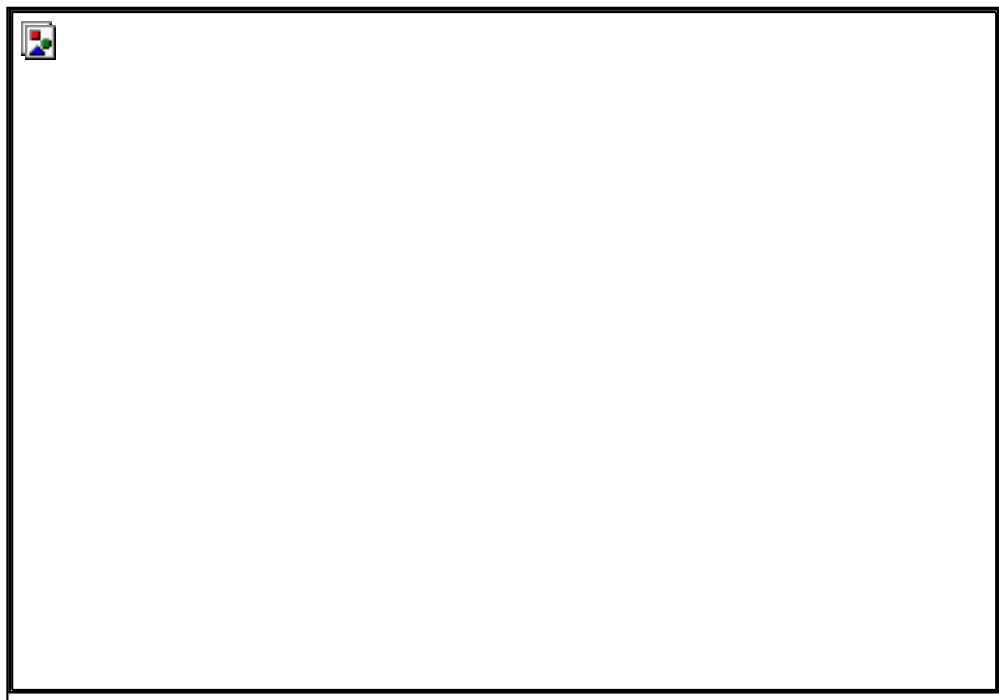


Elaboración: Propia

A continuación, se observa la inclinación del histograma (ver Ilustración 19). a la izquierda es decir que la asimetría es positiva y los datos son continuos, por lo tanto, el tipo de familia a utilizar en esta investigación es Gamma ya que cumple con todos los parámetros expuestos en este párrafo.

4.2.2. Generación del mejor modelo GAM

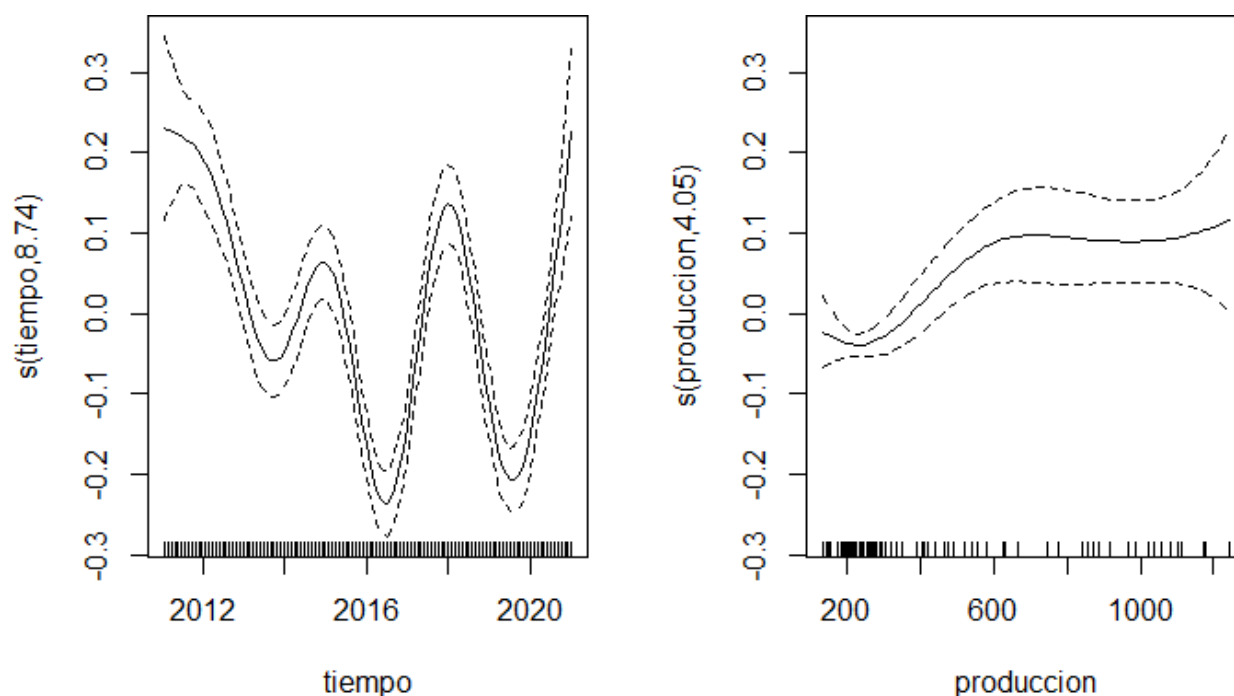
Después de establecer la familia exponencial, corresponde procesar el modelo en el programa R a través de la librería “*gam*” con las variables precio y tiempo, al obtener los resultados no era del todo significativo el sistema para lo cual, se decidió introducir un componente más la producción mensual entre los años 2011 al 2020 (ver Anexo 2), quedando de la siguiente manera la formula “ $precio \sim s(tiempo) + s(producción)$ ”, donde el intercepto es menor a 0.05 al igual que las variables explicativas (tiempo y producción) siendo significativos para el estudio y por último, los grados de libertad efectivos son de 8.74 y 4.09 respectivamente indicando que la suavización no es lineal por ser mayores a uno. (ver Ilustración 20)

Ilustración 20 Resultados obtenidos del modelo GAM

Elaboración: Propia

A continuación, se graficó (ver Ilustración 21) el modelo GAM, donde se presenta dos imágenes de tiempo y producción. Al observar el gráfico de la izquierda el valor mensual de los años 2011 y 2012 empieza a descender, en cambio en los periodos posteriores hasta el 2015 es ascendente pero su mayor incremento es entre el 2017 al 2019; entonces la relación precio y tiempo manifiesta fluctuación. Por otro lado, la relación con producción es positiva durante todos los años 2011 al 2020 presentando crecimiento prologando a pesar de una pequeña disminución en los últimos años y los puntos que se presentan en ambos gráficos es la oscilación manifestada para ambas variables.

Ilustración 21 Gráficos obtenidos del Modelo Aditivo Generalizado

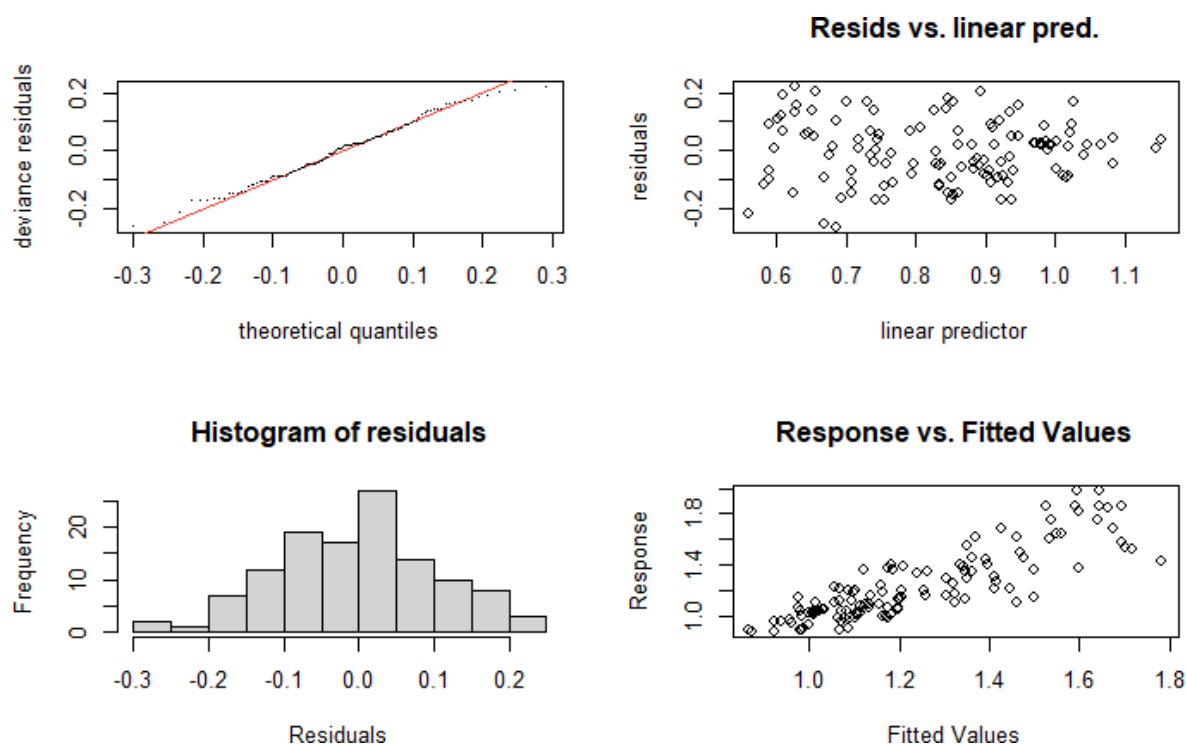


Elaboración: Propia

4.2.3. Inspección del modelo GAM

En los anteriores párrafos demostramos que la relación entre el precio y las variables explicativas tiempo y producción es significativa, ahora bien, inspeccionaremos el modelo donde la desviación de residuales muestra relación entre las variables observadas frente aquellas variables predichas presentando una previsibilidad buena, también manifiesta homocedasticidad es decir que la varianza de los errores es constante en el modelo y tiempo en la mayoría de datos, el histograma de los residuales es asimétrica indicando la existencia de una relación lineal y por último, los residuales frente a los ajustes muestran una varianza constante debido a que después del cero las variables se mueven en ambas direcciones. (ver Ilustración 22)

Ilustración 22 Prueba de bondad del ajuste del modelo



Elaboración: Propia

4.2.4. Validación del modelo

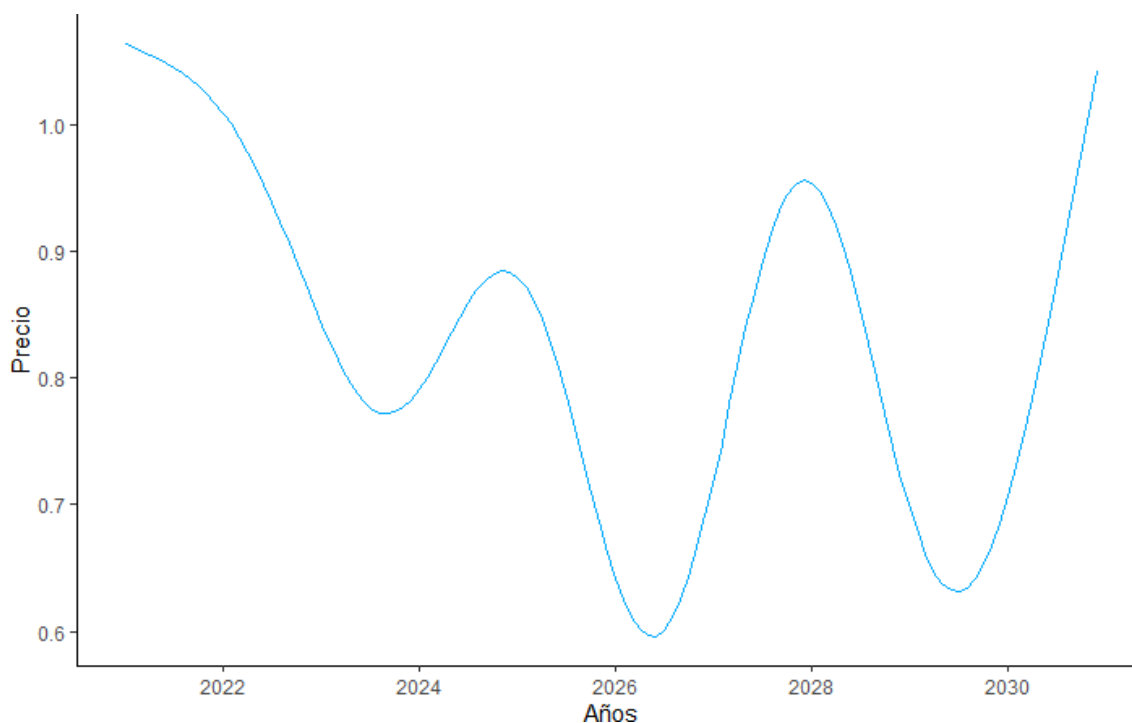
La validación del modelo GAM en la presente investigación está determinado por el R^2 ajustado el cual, permite conocer la relación entre las variables respuestas y explicativas mostrando una interacción entre ambas variables del 69.50% pudiéndose visualizar en la ilustración 20, además, la desviación explicada nos indica que la media está comprometida en un 75%.

4.2.5. Predicción

El pronóstico obtenido en un modelo GAM es por la misma cantidad de unidades estudiadas, por lo tanto, los años de predicción de papa será para los siguientes 120 meses es decir desde el 2021 al 2030, para una mejor interpretación de los resultados se realizaron dos métodos:

El primero es determinado de manera sencilla debido a que en el programa R utilizamos la función "Predit" donde insertamos los resultados obtenidos en la generación del modelo GAM, el cual automáticamente realiza la predicción para los siguientes 10 años. (ver Ilustración 23)

Ilustración 23 Predicción del precio de la papa mediante el modelo GAM



Elaboración: Propia

Después de observar la ilustración 23 muestra que en los tres primeros años hay una caída del precio de la papa al por mayor, pero en los dos siguientes periodos presente una recuperación relativa monetaria del tubérculo, en general el valor económico de la papa es fluctuante durante los 10 años de pronóstico, por otro lado, cabe resaltar para los meses del 2030 presenta un aumento que sobre pasa el S/1.00, pudiendo especular que se debe al cambio climático, el aumento de la población y la reducción de las áreas verdes.

En la segunda predicción se buscó realizar más complejo el sistema para lo cual, se incluyó el pronóstico de producción mensual durante los años 2011 al 2030 a través de un SARIMA, que a continuación será explicado:

Para empezar, se ordenó los datos de la producción entre los años 2011 al 2020 mensualmente, se utilizó la función “auto.arima” para conocer la cantidad de autoregresivos, diferencias y las medias móviles necesarias para determinar el mejor modelo. (ver Ilustración 24).

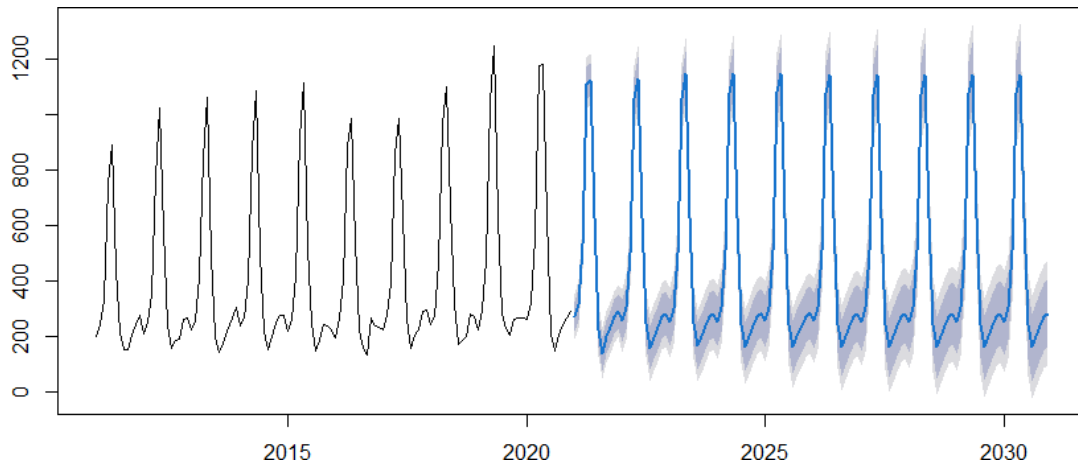
Ilustración 24 Determinación del mejor modelo ARIMA en el programa R



Elaboración: Propia

Para luego realizar la predicción de la producción siguiendo los mismos pasos que el modelo ARIMA, determinando utilizar en este caso un SARIMA (1,0,0) (2,1,1), debido principalmente a que presenta estacionalidad. Para estar seguro que el modelo es el adecuado se realizó la prueba de Ljung-Box en el cual se obtuvo un p-valor de 0.47 siendo significativo ya que es mayor a 0.05, procediendo a realizar el pronóstico del modelo SARIMA para los 120 meses de los años 2021 al 2030 (ver Ilustración 25).

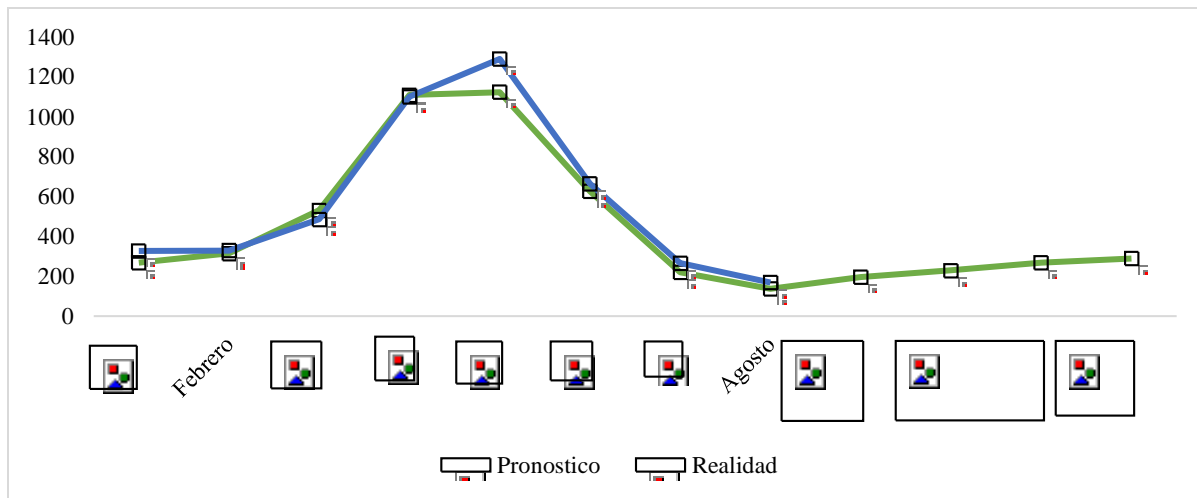
Ilustración 25 Pronostico de la producción de papa entre los años 2021 - 2030



Elaboración: Propia

Como se puede observar en la ilustración 25 la predicción presenta oscilaciones muy parecidas a los años anteriores, para una mejor comprensión del grafico se comparó los meses de enero hasta agosto del 2021 donde el pronóstico es muy acertado frente a la realidad de la producción de papa (ver Ilustración 26).

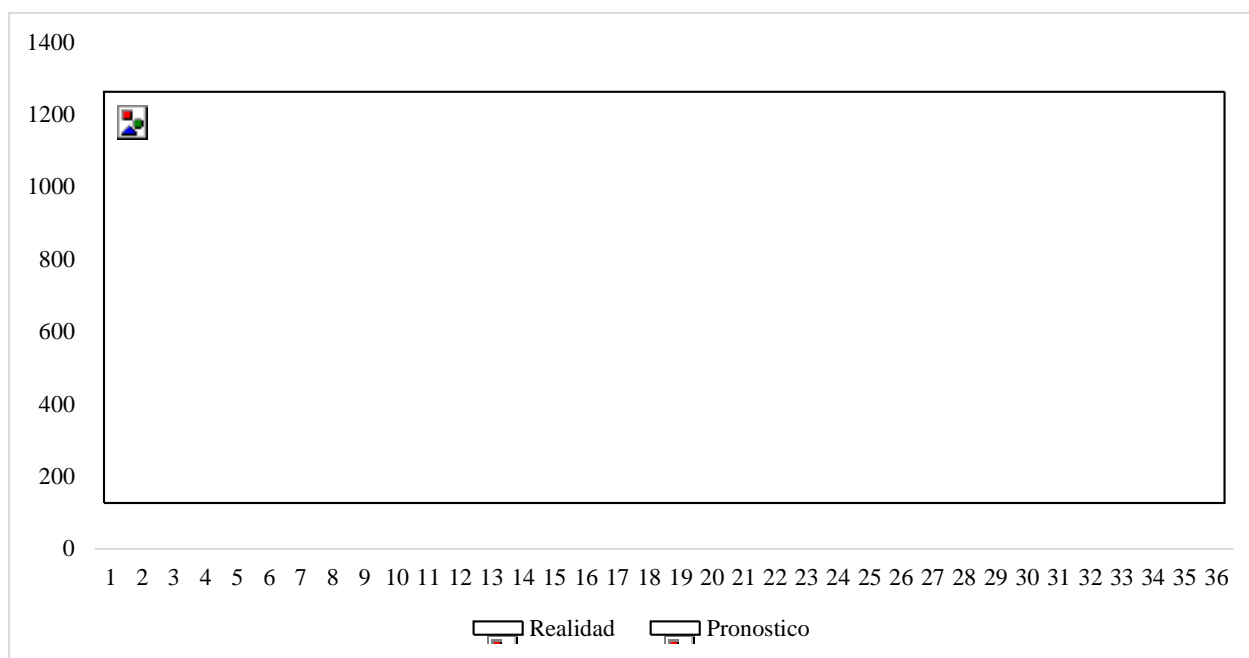
Ilustración 26 Comparación de producción de papa, entre la predicción – realidad 2021



Elaboración: Propia

De la misma manera se realizó la comparación para los años 2018 al 2020, por lo cual se utilizaron datos de producción a nivel nacional para los años 2008 al 2017. Encontrando en la ilustración 27 que el comportamiento predicho presenta mucha similitud frente a la realidad que se presentó durante los periodos pronosticados.

Ilustración 27 Comparación de producción de papa, entre la predicción – realidad 2018 - 2020

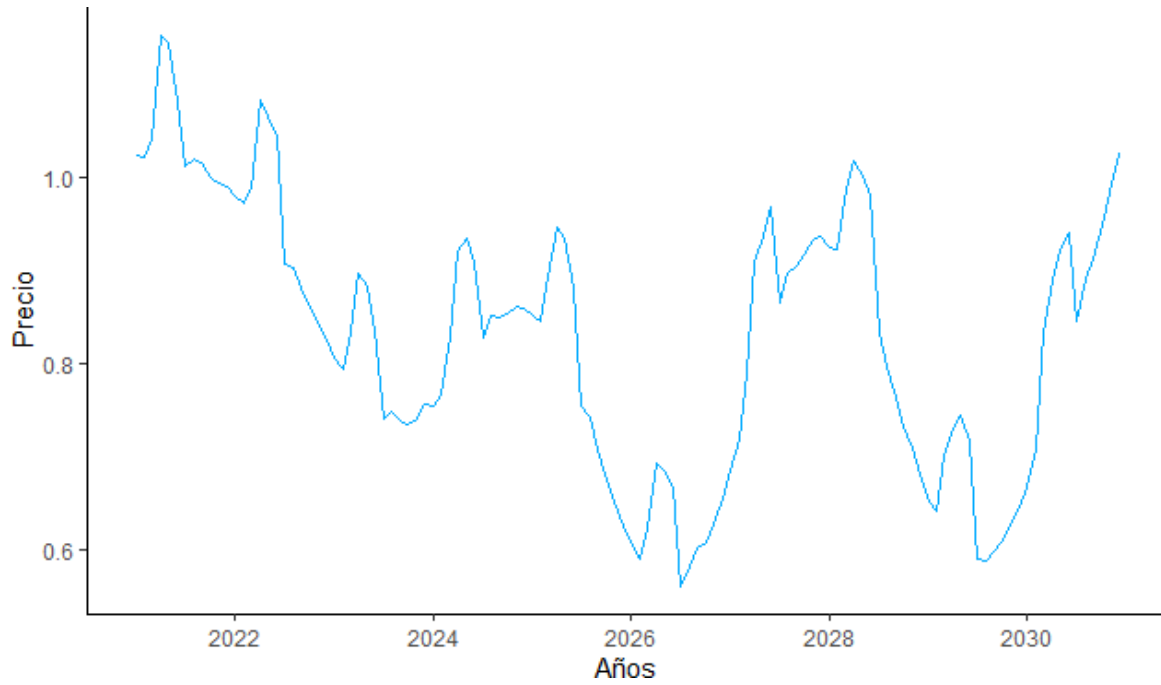


Elaboración: Propia

Después de conocer la producción futura entre los años 2021 al 2030 se procedió a realizar el pronóstico del precio de la papa de la misma manera con la librería “Predit” en el programa R donde las variables explicativas son el tiempo y la nueva data de producción, para poder obtener una mejor predicción que se acerque a la realidad.

En la ilustración 28 comienza mostrando el precio de la papa por encía del S/1.00 para el año 2021, pero en los siguientes periodos la disminución del valor monetario es graduable hasta inicios del 2024 llegando a S/0.70 por kilo de papa, sin embargo, en los dos próximos periodos presenta recuperación del precio, pero en el 2026 se manifiesta el valor mínimo de toda la seria siendo menor a 0.60 soles y para terminar el último año se da un crecimiento graduable hasta pasar el S/. 1.00.

Ilustración 28 Predicción del precio de la papa utilizando el modelo SARIMA 2021 - 2030

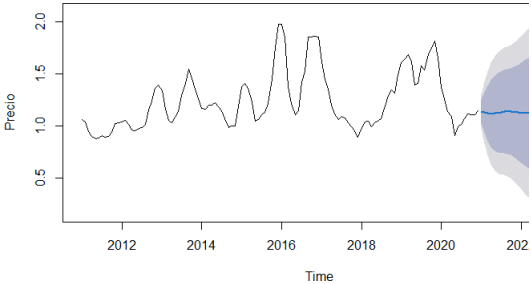
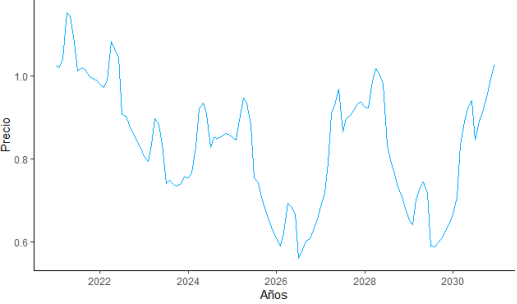


Elaboración: Propia

4.3. Comparación de los métodos ARIMA y GAM.

En el siguiente cuadro (ver Tabla 9). se presenta de una manera más dinámica ambos modelos:

Tabla 9 Comparación entre modelo ARIMA y GAM

Modelo ARIMA	Modelo GAM
	
<ul style="list-style-type: none"> - El modelo ARIMA presenta un pronóstico lineal frente a los datos obtenidos como referencia en el pasado. - Las predicciones se pueden realizar en el corto y largo plazos. - Solo es necesario la variable respuesta, ya que se trabaja con el tiempo pasado, para determinar el futuro. - Sigue parámetros lineales, ya que es una extensión de este mismo. - Genera automáticamente los diferentes modelos que se pueden trabajar hasta alcanzar el mejor de ellos. 	<ul style="list-style-type: none"> - El Modelo Aditivo Generalizado trabaja con suavizadores que permite obtener curvas que se asemejan a la realidad. - En su mayoría, las predicciones son de largo plazo. - Este tipo de método se puede interpretar mejor, a través del uso y análisis de gráficos. - El modelo trabaja con distintas variables para la obtención de mejores predicciones, debido a la interacción entre ellas que se asemejan a la realidad. - El pronóstico para los 120 meses comprendidos en el período de años entre el 2021 y 2030,

-
- Adaptable a cualquier tipo de investigación.
 - La predicción para el precio de la papa se asemeja a una línea recta para los 24 meses comprendidos en el período de años entre el 2021 y 2022.
 - El valor futuro del precio promedio de la papa es S/1.13, siendo su variación de un a dos centavos para los dos años pronosticados.
 - Una de las principales ventajas del modelo ARIMA, en cuanto a la predicción, es que ofrece valores mínimos y máximo de esta misma, en consecuencia, se contaría con un rango, donde podría oscilar el precio de la papa.
- presenta un gráfico que demuestra la variación constante del precio, también conocido como fluctuación.
 - En los dos primeros períodos pronosticados, se puede observar que el precio es mayor a S/1.00, pero aproximadamente durante el primer semestre del año 2021 llega a costar el kilo de papa al por mayor S/ 1.70, para luego presentar un decrecimiento constante hasta llegar el precio 0.90 céntimos durante los siguientes meses hasta el 2022.
 - La metodología utilizada para la predicción de precios no determina los valores mínimo y máximo para esta misma.
-

Elaboración: Propia

Conclusiones

Conclusión 1: El modelo ARIMA permite obtener la proyección del precio de la papa mediante los diferentes parámetros lineales que se deben cumplir, y trabajando de manera sencilla y rápida. Sin embargo, los pronósticos obtenidos no se asemejan a la realidad ya que sigue un patrón recto donde no se aprecia claramente la disminución y subida del índice de precio.

Conclusión 2: El pronóstico obtenido mediante la técnica del modelo GAM fue de mayor precisión en comparación con el ARIMA, mostrando: (a) Las fluctuaciones del precio en los diferentes meses para el período de años 2021-2030, (b) descenso desde el período de años 2021-2024, llegando hasta 0.80 céntimos de soles el kilo; sin embargo, su mayor quiebre es el año 2026, siendo el precio inferior a los S/.0.60; para que posteriormente, presentará su mayor crecimiento en el año 2026, siendo el precio mayor a un S/. 1.00, aproximadamente.

Conclusión 3: Al comparar ambos modelos podemos dar a conocer que los mejores pronósticos se obtuvieron con el Modelo Aditivo Generalizado corroborando la hipótesis nula e indicando que los modelos lineales son más precisos al explicar la relación entre variables, ya que los otros métodos modelan sus curvas según las predicciones que presentan las variables.

Conclusión 4: Los años determinados para el presente estudio, fueron del período comprendido entre el 2011 al 2020, logrando obtener 240 datos mensuales entre las variables de producción y precio; de esta manera, estos fueron considerados bajo la aplicación de procedimientos con fines estadísticos.

Conclusión 5: La validación del modelo ARIMA fue determinado por el p-valor, que fue de 0.97 mayor a 0.05 permitiendo saber que el método es significativo; por otro lado, para el caso del modelo GAM, su efectividad fue comprobada con el R^2 ajustado, indicando que la variable respuesta es explicada en un 69.50% por la producción y el tiempo.

Conclusión 6: El presente estudio permitió poder realizar el método SARIMA para conocer la producción futura de la papa, obteniendo un índice de pronóstico, en comparación con la realidad y aumentando la confiabilidad de la segunda predicción del modelo GAM. Por otro lado, la mayor producción de papa corresponde al mes de marzo, con más de 1,200 mil toneladas; mientras que la menor cosecha corresponde al mes de agosto llegando a alcanzar como máximo 200 mil toneladas.

Recomendaciones

Recomendación 1: Realizar el modelo GAM considerando una mayor cantidad de variables, como pueden ser el precio de insumo para siembra de la papa, el clima de las diferentes regiones, entre otros.

Recomendación 2: Tomar en consideración el presente estudio consultando y analizando las fuentes primarias procedentes de cualquier región del Perú, con el propósito de conocer la situación real y las principales dificultades que se presentan en estas regiones.

Recomendación 3: Ejecutar el nuevo estudio con los alcances realizados en los anteriores párrafos y bajo el contexto posterior de la pandemia, a fin de comparar los pronósticos de la presente investigación.

Bibliografía

- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., ... Helios, D. R. (2020). *Basic Functions for Power Analysis Description*. 1–22.
- Wood, S. N., & Augustin, N. H. (2002). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, *157*(2–3), 157–177. [https://doi.org/10.1016/S0304-3800\(02\)00193-X](https://doi.org/10.1016/S0304-3800(02)00193-X)
- Davis, M. L., Woodard, D. A., Pesto, M., & Toya, R. E. (1977). Attempts to purify hemopoietic stem cell enrichment in bone marrow by use of glass wool filtration. *Experimental Hematology*, *5*(4), 310–318.
- Hastie, T. J., & Tibshirani, R. J. (2017). Generalized additive models. *Generalized Additive Models*, pp. 1–335. <https://doi.org/10.1201/9780203753781>
- Del, V., Calcinación, M. D. E., La, E. N., Para, T., El, O., Para, T., ... Suelo, D. E. L. (2019). *La Molina Molina*. 10–12.
- Aguilar Barreiro, P. (2019). Modelos aditivos generalizados. *Departamento de Estadística e Investigación Operativa*, 90.
- González-Casimiro, M. P. (2009). Análisis de series temporales: Modelos ARIMA. In *Universidad del País Vasco*.
- Gomes, F. C. (1989). Os modelos Arima e a abordagem de Box-Jenkins uma aplicação na previsão do IBOVESPA a curtíssimo prazo. *Revista de Administração de Empresas*, *29*(2), 63–70. <https://doi.org/10.1590/s0034-75901989000200006>
- Olivero, N. (1997). Modelos Arima. *Revista Científica Ciencia Médica*, (1), 23–30.
- Makridakis, S., & Hibon, M. (1997). ARMA models and the Box-Jenkins methodology. *Journal of Forecasting*, *16*(3), 147–163. [https://doi.org/10.1002/\(SICI\)1099-131X\(199705\)16:3<147::AID-FOR652>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X)

- Mendoza, N. V. R., & Piedade, S. M. D. S. (2019). Generalized Additive Mixed Models With P-Splines Applied To Sugarcane Production in the State of São Paulo. *Revista Brasileira De Biometria*, 37(1), 17–31. <https://doi.org/10.28951/rbb.v37i1.339>
- David, B., Sánchez Sánchez, A., Sánchez Sánchez, B., Alexander, D., Chafloque Viteri, L., & Albert, S. (2018). *Universidad Nacional De Trujillo Facultad De Ciencias Físicas Y Matemáticas Escuela Academico Profesional De Estadística "Modelo Arima Para El Pronóstico De La Producción De Cacao*.
- Santana, J. S., & Farfán, E. M. (2014). El Arte de programar en R. *Instituto Mexicano de Tecnología Del Agua*, 182. Retrieved from https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf
- Análisis espacial con R : Usa R como un Sistema de Información Geográfica Jean-François Mas*.(n.d.).
- Minaya, C. (2015). *Analysis of the Profitability in the Production of Commercial White*. 76(2), 369–375.
- Laurente, L. F., & Laurente, F. (2019). Aplicación del modelo ARIMA para la producción de la papa en la región de Puno-Perú. *Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales*, 6, 11. Retrieved from http://www.scielo.org.bo/pdf/riiarn/v6n1/v6n1_a06.pdf
- García, J., Romero, F., & Mayorga, O. (2014). Modelado del sistema logístico de la cadena productiva de la papa empleando dinámica de sistemas. *AVANCES Investigación En Ingeniería*, 11(2), 86–93.
- González, V., Alegret, M., Betancourt, J., Bioestadística, E. De, Ernesto, C., Guevara, C., ... Titular, P. (2011). Curvas RECEIVER Operating Characteristic y Matrices de Confusión en la Elaboración de Escalas Diagnósticas. *Resumen / Abstract*, 7, 1–10.
- Broz, D. R., & Viego, V. N. (2014). Predicción de precios de productos de Pinus spp.con modelos ARIMA. *Madera y Bosques*, 20(1), 37–46. <https://doi.org/10.21829/myb.2014.201174>

- Pani, R., & Biswal, S. K. (2019). Weekly price forecasting of cotton using time series model. *Journal of Advanced Research in Dynamical and Control Systems*, 11(10 Special Issue), 236–240. <https://doi.org/10.5373/JARDCS/V11SP10/20192796>
- Kubben, P., Dumontier, M., & Dekker, A. (2018). Fundamentals of clinical data science. In *Fundamentals of Clinical Data Science*. <https://doi.org/10.1007/978-3-319-99713-1>
- Ruiz Hernández, J. A., Barrios Puente, G., & Gómez Gómez, A. A. (2019). Análisis del precio de la manzana mediante un modelo SARIMA. *Revista Mexicana de Ciencias Agrícolas*, 10(2), 225–237. <https://doi.org/10.29312/remexca.v10i2.509>
- Henao, J. D. V., Morales, Y. O., & Cardona, C. J. F. (2011). Análisis y predicción de series de tiempo en mercados de energía usando el lenguaje R. *DYNA (Colombia)*, 78(165), 287–296.
- PNIA. (2019). Formulación de la política nacional de innovación agraria, el plan nacional de innovación, y planes de los programas macro regionales, nacionales y transversales, así como la definición e implementación de una estructura organizacional modernizada. 1–400. Retrieved from <https://www.pnia.gob.pe/post/2662/>
- Cambios del sector papa en el Peru en la ultima decada: Los aportes del proyecto Innovacion y Competitividad de la Papa (INCOPA). (2011). In *Cambios del sector papa en el Peru en la ultima decada: Los aportes del proyecto Innovacion y Competitividad de la Papa (INCOPA)*. <https://doi.org/10.4160/9789290604112>
- INIA. (2012). Papa INIA 303 - Canchán. *Ministerio de Agricultura*, 1o.
- Agroideas. (2019). *Catálogo de Mejora Tecnológica para la Competitividad Agraria: Programa AGROIDEAS*. 1–39. Retrieved from <https://drive.google.com/file/d/1lwnMj-zkAgEeiEHxSs8I4T3LGsY4VMRd/view>
- Libertad, L. (n.d.). *Requerimientos Agroclimáticos del cultivo de papa*.
- Comercial, A., & Al, O. (2016). *GUIA “Modelo para el establecimiento de los actores de la cadena productiva” Junio 2016. AGRARIOS DEL SECTOR AGRICULTURA Y RIEGO*. (2018). 1–44.

- INIA, (Ministerio de agricultura y riego). (2016). *Guía metodológica para la transferencia de tecnología agraria*. 195. Retrieved from <http://repositorio.inia.gob.pe/handle/inia/479>
- Ministerio de Agricultura y Riego. (2021). MOC Marco orientador de cultivos 2021. *Ministerio de Agricultura y Riego*, 148, 162.
- Pol, G. D. E., & Agrarias, T. (2020). *De Papa Fresca E Importaciones De Papa Prefrita*. Retrieved from <https://cdn.www.gob.pe/uploads/document/file/1267493/Producción de papa fresca e importaciones de papa prefrita congelada.pdf>
- Aracil, J., & Gordillo, F. (1997). *Dinamica de Sistemas.pdf* (pp. 10–109). pp. 10–109.
- El, P., Mediano, P. Y., Mejor, A. P. D. E. L., & Economía, M. (2016). *comportamiento del precio de la papa diacol capiro , pronóstico entre un modelo SARIMA y una RNA CLAUDIA HAYDEE LIZARAZO RENGIFO Trabajo de grado presentado como requisito para optar al título en : Ingeniería Financiera Ing . Financiero Carlos Mario Garcí*.
- Pol, G. D. E., & Agrarias, T. (2017). *Papa : Características de la Producción Nacional y de la Comercialización en Lima Metropolitana*.
- Vanlehn, K., Wetzel, J., Grover, S., & Van De Sande, B. (2017). Learning How to Construct Models of Dynamic Systems: An Initial Evaluation of the Dragoon Intelligent Tutoring System. *IEEE Transactions on Learning Technologies*, 10(2), 154–167.
<https://doi.org/10.1109/TLT.2016.2514422>
- Mora Flores, W. (2016). Cómo utilizar R en métodos numéricos. *Revista Digital: Matemática, Educación e Internet*, 16(1), 1–74. <https://doi.org/10.18845/rdmei.v16i1.2480>
- Intelectual, P., & Agust, S. (2013). “*Cinco modalidades de aplicación de Confidor, para el control de Russeliana solanicola Tuthil en plan nacional de cultivos (Campaña Agrícola 2018-2019) I*.” (2019). 1–293.
- Doctorado, F. (n.d.). *EQUIVALENTE Y SU IMPLEMENTACIÓN EN EL ENTORNO*. Familiar, A. (n.d.). *Recomendaciones de Política Recomendaciones de Política. No Title*. (2011).

Donato, J. de M., Dormido, S., & Morilla, F. (2005). Fundamentos de la dinámica de sistemas y Modelos de dinámica de sistemas en epidemiología. *Manual Dinámica Sistemas*, 69.

Retrieved from

http://www.proyectosame.com/ds_documentos/manual_dinamica_sistemas.pdf

Anexos

Anexo 1

Ficha Electrónica

Título: Producción de la papa entre los años 2011 al 2020 en miles.

Fecha: Septiembre 2021

Autor: Lady Geraldine Durand Chavez

Página electrónica: MINAGRI

Meses/Años	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Enero	196.58	206.02	223.41	234.45	215.23	193.40	223.85	241.92	222.96	258.76
Febrero	242.95	251.80	255.00	264.33	259.18	262.80	275.02	271.23	291.08	320.80
Marzo	333.08	350.10	411.67	403.15	419.35	403.28	387.36	466.28	474.62	556.13
Abril	747.55	775.81	843.00	869.30	917.75	885.90	854.07	966.31	1039.37	1172.57
Mayo	887.22	1023.13	1058.69	1083.60	1113.67	984.61	986.39	1100.09	1244.79	1179.86
Junio	438.66	578.24	487.57	492.21	517.50	540.71	629.64	666.94	624.90	624.54
Julio	210.06	240.03	195.99	218.13	244.37	223.16	259.33	316.73	279.03	205.73
Agosto	149.50	155.16	142.39	151.37	144.66	154.03	154.79	171.54	224.74	145.71
Septiembre	150.51	181.29	173.63	196.33	185.58	133.49	200.92	185.12	202.20	203.41
Octubre	204.35	189.02	216.92	245.54	240.20	264.18	221.60	197.96	258.79	243.15
Noviembre	239.40	259.86	266.61	272.91	235.76	236.08	289.20	277.10	263.84	265.58
Diciembre	272.59	264.27	302.38	273.68	222.67	232.59	294.12	270.32	262.91	290.79

Meses/Años	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Enero	196.58	206.02	223.41	234.45	215.23	193.40	223.85	241.92	222.96	258.76
Febrero	242.95	251.80	255.00	264.33	259.18	262.80	275.02	271.23	291.08	320.80
Marzo	333.08	350.10	411.67	403.15	419.35	403.28	387.36	466.28	474.62	556.13
Abril	747.55	775.81	843.00	869.30	917.75	885.90	854.07	966.31	1039.37	1172.57
Mayo	887.22	1023.13	1058.69	1083.60	1113.67	984.61	986.39	1100.09	1244.79	1179.86
Junio	438.66	578.24	487.57	492.21	517.50	540.71	629.64	666.94	624.90	624.54
Julio	210.06	240.03	195.99	218.13	244.37	223.16	259.33	316.73	279.03	205.73
Agosto	149.50	155.16	142.39	151.37	144.66	154.03	154.79	171.54	224.74	145.71
Septiembre	150.51	181.29	173.63	196.33	185.58	133.49	200.92	185.12	202.20	203.41
Octubre	204.35	189.02	216.92	245.54	240.20	264.18	221.60	197.96	258.79	243.15
Noviembre	239.40	259.86	266.61	272.91	235.76	236.08	289.20	277.10	263.84	265.58
Diciembre	272.59	264.27	302.38	273.68	222.67	232.59	294.12	270.32	262.91	290.79

Anexo 2

Ficha Electrónica

Título: Precio de la papa entre los años
2011 al 2020.

Fecha: Septiembre 2021

Autor: Lady Geraldine Durand Chavez

Página electrónica: MINAGRI

Meses/Años	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Enero	1.07	1.04	1.34	1.17	1.38	1.98	1.61	0.99	1.61	1.37
Febrero	1.04	1.05	1.16	1.16	1.41	1.86	1.44	1.04	1.64	1.26
Marzo	0.95	1.01	1.06	1.20	1.37	1.37	1.35	1.04	1.68	1.14
Abril	0.90	0.96	1.03	1.20	1.23	1.22	1.20	1.00	1.61	1.10
Mayo	0.88	0.96	1.09	1.22	1.05	1.11	1.12	1.03	1.39	0.91
Junio	0.88	0.98	1.14	1.19	1.06	1.15	1.06	1.04	1.41	0.99
Julio	0.90	0.99	1.29	1.15	1.11	1.42	1.09	1.07	1.58	1.02
Agosto	0.90	1.02	1.41	1.07	1.13	1.52	1.08	1.20	1.54	1.07
Septiembre	0.90	1.16	1.55	0.99	1.22	1.85	1.03	1.30	1.69	1.12
Octubre	0.94	1.24	1.46	1.00	1.46	1.85	0.99	1.34	1.75	1.11
Noviembre	1.03	1.37	1.35	1.00	1.76	1.86	0.96	1.31	1.82	1.11
Diciembre	1.03	1.39	1.26	1.18	1.97	1.86	0.89	1.49	1.65	1.15

Meses/Años	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Enero	1.07	1.04	1.34	1.17	1.38	1.98	1.61	0.99	1.61	1.37
Febrero	1.04	1.05	1.16	1.16	1.41	1.86	1.44	1.04	1.64	1.26
Marzo	0.95	1.01	1.06	1.20	1.37	1.37	1.35	1.04	1.68	1.14
Abril	0.90	0.96	1.03	1.20	1.23	1.22	1.20	1.00	1.61	1.10
Mayo	0.88	0.96	1.09	1.22	1.05	1.11	1.12	1.03	1.39	0.91
Junio	0.88	0.98	1.14	1.19	1.06	1.15	1.06	1.04	1.41	0.99
Julio	0.90	0.99	1.29	1.15	1.11	1.42	1.09	1.07	1.58	1.02
Agosto	0.90	1.02	1.41	1.07	1.13	1.52	1.08	1.20	1.54	1.07
Septiembre	0.90	1.16	1.55	0.99	1.22	1.85	1.03	1.30	1.69	1.12
Octubre	0.94	1.24	1.46	1.00	1.46	1.85	0.99	1.34	1.75	1.11
Noviembre	1.03	1.37	1.35	1.00	1.76	1.86	0.96	1.31	1.82	1.11
Diciembre	1.03	1.39	1.26	1.18	1.97	1.86	0.89	1.49	1.65	1.15