



Universidad Agraria de La Habana
"Fructuoso Rodríguez Pérez"
Facultad de Medicina Veterinaria



**EMPLEO DE MODELOS AUTOREGRESIVOS (ARIMA) PARA LA
VIGILANCIA SINDRÓMICA EN CRÍAS DE CERDOS EN UNA GRANJA
DE LA PROVINCIA ARTEMISA, CUBA.**



AUTOR

Dr. MVZ Jany Del Pozo Fernández

Tesis presentada en opción al título de Máster en Biomatemática

Mayabeque, 2019



Universidad Agraria de La Habana
"Fructuoso Rodríguez Pérez"
Facultad de Medicina Veterinaria



**MODELOS AUTOREGRESIVOS (ARIMA) PARA LA
VIGILANCIA SINDRÓMICA EN CRÍAS DE CERDOS EN UNA
GRANJA DE LA PROVINCIA ARTEMISA, CUBA.**

AUTOR

Dr. MVZ Jany Del Pozo Fernández

TUTORES

Dr. MV Pastor Alfonso Zamora. DrC

Dr. C Sergio Hernández González

Dra.C Lucía Fernández Chuairey

Tesis presentada en opción al título de Máster en Biomatemática

Mayabeque, 2019

AGRADECIMIENTOS

Ante todo agradezco a la Universidad Agraria de La Habana (UNAH), por forjarme como profesional y apoyarme en mi superación.

A mis tutores el Dr. C. Pastor Alfonso, el Dr. C Sergio Hernández González, y la Dra. C Lucía Fernández Chuairey, quienes con su acertada orientación y sabios conocimientos ayudaron al desarrollo, organización y culminación de esta tesis.

A mi amado hermano por todo el amor que ha brindado a mi vida.

Y a mis adorados padres por el estímulo y las enseñanzas brindadas a través de mi vida.

A mis idolatrados abuelos que me dieron tanto amor, dedicación y paciencia.

A mi novio por su amor y apoyo incondicional.

A mis compañeros , amigos, familiares, en fin a todos los que de una forma u otra hicieron que me creciera ante las dificultades y que siguiera siempre por el camino del saber, a todos, MUCHAS GRACIAS.

RESUMEN

A nivel internacional aumenta el desarrollo alcanzado en la aplicación de nuevos procedimientos para la vigilancia en la salud, los cuales están basados en teorías matemáticas para la detención de señales anormales y toma oportuna de decisiones para evitar consecuencias adversas . El objetivo del presente trabajo fue desarrollar una alternativa metodológica basada en procedimientos estadísticos y empleo de series temporales para la alerta rápida mediante vigilancia sindrómica en la primera etapa de crecimiento de los cerdos en una granja porcina en Cuba. Para ello se analizaron datos correspondientes a las variables nacimientos y mortalidad, en el período (2010-2017), se aplicó Estadística Descriptiva, dócima de comparación (factores mes y año), representación gráfica de las variables y análisis de tendencia , aplicación de la Metodología de (Box y Jenkins, 1978), se probaron diferentes Modelos Autorregresivos (ARIMA), selección del mejor modelo, así como la predicción de valores y determinación del límite superior en caso de mortalidad y límite inferior en la variable número de nacimiento como momentos de alarma en este sistema de producción, Se concluye que bajo las condiciones de estudio el indicador números de nacimiento en un mes no debe ser menor de 109 cerditos y en caso de la mortalidad no debe exceder 11.6 % en el mes ya constituyen una señal de alarma .El desarrollo de procedimientos estadístico y la aplicación de Modelos ARIMA, constituyen una vía eficiente para determinar el momento óptimo de activar una alarma en los sistemas de vigilancia sindrómica.

Palabras Claves: Series temporales, Vigilancia Sindrómica, Modelos Autorregresivos.

SUMMARY

At an international level, the development achieved in the application of new procedures for health surveillance is increasing, which are based on mathematical theories for the arrest of abnormal signals and timely decision making to avoid worse consequences. The objective of the present work was to develop a methodological alternative based on statistical procedures and use of time series for a rapid alert by syndromic surveillance in the first stage of growth of pigs in a pig farm in Cuba. For this, data corresponding to the birth and mortality variables were analyzed, in the period (2010-2017), Descriptive Statistics was applied, comparison test (month and year factors), graphic representation of the variables and trend analysis, application of the Methodology (Box and Jenkins, 1978), different Autoregressive Models (ARIMA), selection of the best model were tested, as well as the prediction of values and determination of the upper limit in case of mortality and lower limit in the variable number of birth as moments of alarm in this production system, It is concluded that under the conditions of study the indicator of birth numbers in a month should not be less than 109 little pigs and in case of mortality should not exceed 11.6% in the month already constitute a signal of alarm. The development of statistical procedures and the application of ARIMA Models, constitute an efficient way to determine the optimal moment of cause an alarm in syndromic surveillance systems.

Key Words: Time Series, Syndromic Surveillance, Autoregressive Models.

INDICE**pág.****INTRODUCCIÓN****.1****CAPÍTULO I. FUNDAMENTOS TEÓRICOS****4**

I.1 Métodos Estadísticos. Aplicaciones más frecuentes en la Sanidad Agropecuaria

I.1.1 Series Temporales. Fundamentos teóricos

4

I.1.2 Series temporales, su empleo en la predicción y detección de patrones anormales en la aplicación de la vigilancia sindrómica

7

I.1.3. Vigilancia sindrómica. Empleo de series temporales en la predicción de valores y detección de patrones anormales

12

I.2 El cerdo en Cuba. Principales factores relacionados con el manejo y la bioseguridad que inciden en la productividad y la salud de la porcina

14**CAPÍTULO II. EMPLEO DE MODELOS AUTOREGRESIVOS (ARIMA) PARA LA VIGILANCIA SINDRÓMICA EN CRÍAS DE CERDOS****16**

II:1 Materiales y Métodos

16

II.2 Resultados y Discusión

20**CONCLUSIONES****53****RECOMENDACIONES****54****REFERENCIAS BIBLIOGRÁFICAS****55**

INTRODUCCIÓN

La Estadística es una de las ramas de la Matemática de mayor universalidad, ya que en su devenir histórico muchos de sus métodos se han desarrollado para resolver situaciones específicas en diferentes esferas del conocimiento y la ciencia (Herrera, 2013).

Morante (2018) señala que la Estadística es el nexo común que presentan la mayoría de las investigaciones científicas en las que interviene el tratamiento de datos y la interpretación y predicción de los mismos. Además, alega que dada la importancia que adquiere, aparece en prácticamente todas las áreas de trabajo, siendo un factor claro en las predicciones y tomas de decisiones a partir de datos observados.

En Cuba, existen experiencias, resultados y criterios sobre las potencialidad del empleo de procedimientos estadísticos, modelación y simulación de proceso y metodologías de estadístico-matemática aplicados a Procesos Agrarios, cuyo uso favorecen las investigaciones científicas y contribuyen a la búsqueda de soluciones y producciones óptimas.

En tal sentido Fernández y col. (2013) hacen referencia al empleo de la Estadística Aplicada a las Ciencias Agropecuarias, así como su contribución en la docencia, investigación y transferencia de conocimiento. La literatura reporta aplicaciones en este campo, asociados a estudios de curvas de crecimiento animal y vegetal, en Metodología de Superficie de Respuesta (MSR) en los cultivos de la caña de azúcar, cítricos, pastos y forrajes, en curvas de lactancia, en estudios ecológico-estadísticos de plagas, en sistemas pastoriles, en estudio de sostenibilidad socioeconómica así como en evaluaciones genéticas para el mejoramiento animal, entre otras investigaciones reportadas por, Fernández (2011, 2017), Del Valle y Guerra (2012), Vázquez y col. (2014) y Rodríguez (2015)

En la actualidad en el área del conocimiento de la salud animal se incluyen líneas de investigación que dan respuesta a las exigencias y desarrollo del país, que consideran estudios sobre nuevas formas de vigilancia epidemiológica con el fin de mejorar la eficiencia en los sistemas de detección de enfermedades, siendo una alternativa la denominada vigilancia sindrómica con el empleo de procedimientos estadísticos, que incluyen como herramientas básicas el empleo de las series temporales: (modelos autoregresivos) avalados por los trabajos de Ramírez y col, (2005), Hoinville, (2013). Dórea y col, (2013) entre otros.

Por todo lo anteriormente expuesto, se propone el siguiente

PROBLEMA CIENTÍFICO:

En Cuba a pesar de los esfuerzos realizados y los estudios que se desarrollan en esta área del conocimiento, aún no se cuenta con: procedimientos estadísticos para el análisis de los indicadores número de nacimientos y mortalidad que permitan activar alarmas en los sistemas de vigilancia sindrómica” en la primera etapa de vida cerdos.

Y la siguiente **HIPÓTESIS** de trabajo:

Los modelos autoregresivos (ARIMA) aplicados a indicadores productivos y de salud en crías porcinas permiten detectar aberraciones que justifiquen intervención.

Según lo expuesto se propone como **OBJETIVO GENERAL:**

Desarrollar modelos autoregresivos (ARIMA) para detectar ocurrencia de aberraciones en indicadores de la primera etapa de desarrollo en cerdos.

Con los siguientes **OBJETIVOS ESPECÍFICOS**:

1. Describir el comportamiento de los indicadores número de nacimientos y mortalidad en la categoría de cría porcina como una forma de vigilancia sindrómica.
2. Desarrollar un modelo para la detección rápida de aberraciones, con el empleo de Series Temporales de datos asociados al número de nacimientos y mortalidad en cerdos en la categoría de crías.

CAPÍTULO I. FUNDAMENTOS TEÓRICOS

I.1 Métodos Estadísticos. Aplicaciones más frecuentes en la Sanidad Agropecuaria

En las investigaciones Agrarias y muy en especial las asociadas a la Sanidad Agropecuaria, se eleva el desarrollo alcanzado en la aplicación de nuevos procedimientos para la vigilancia en la salud, los cuales están basados en teorías matemáticas para la detención de señales anormales y toma oportuna de decisiones para evitar peores consecuencias.

Dentro de las técnicas actuales se destaca el empleo de series temporales: Modelos Autorregresivos Integrados con Medias Móviles (ARIMA) con la metodología de Box y Jenkins (1976), los que han sido empleados en diferentes ramas de la ciencia, tales como: las Agropecuarias, Médicas, Económicas, entre otras y que serán abordados en el presente trabajo.

I.1.1 Series temporales. Fundamentos teóricos

En las investigaciones científicas en las ciencias agrarias, es común el uso de series temporales que son un apoyo en la solución de problemas complejos, a la búsqueda de predicciones y de soluciones óptimas.

Autores como Guerra y col. (2010) consideran que las series cronológicas o series de tiempo, representan una sucesión de valores $X_1, X_2, X_3, \dots, X_n$, correspondientes a un mismo fenómeno el cuál es observado a intervalos fijos de tiempo.

Analizando el criterio anterior y otros reportados en la literatura, resaltan los aportados por Coutin (2007), quien señala que las series temporales pueden clasificarse, según la forma en que se ofrecen los valores de las mediciones en dos tipos:

- **Continuas:** cuando los valores se ofrecen de forma permanente, de manera tal que cada uno de ellos representa el estado de la variable en un instante, el cual puede ser tan pequeño como teóricamente se quiera suponer.
- **Discretas:** cuando los valores se ofrecen para intervalos de tiempo, generalmente homogéneos y donde representan la magnitud acumulada del estado de la variable durante ese intervalo.

Además alegan que atendiendo al tipo de componente que pueden tener se pueden agrupar en tres grupos: con tendencia y sin estacionalidad, con tendencia y con estacionalidad y sin tendencia y con estacionalidad.

Con posterioridad Vallejo (2012) señala que según los objetivos del análisis de series temporales, éstas se utilizan con fines: descriptivo y/o predictivos. Fundamenta que dentro de las técnicas de estudio de las series temporales, se pueden distinguir esencialmente dos grandes grupos:

1. Los métodos clásicos o de descomposición (métodos no paramétricos)
2. Los modelos Autorregresivos Integrados con Medias Móviles (ARIMA) de series temporales (métodos paramétricos) propuestos por Box y Jenkins (1976) (estos por su importancia y actualidad serán descrito en el próximo epígrafe)

Autores como Guerra y col. (2010) consideran que los llamados métodos clásicos de una serie cronológica presentan varios tipos de movimientos o formas de comportamiento que son llamadas componentes. Las componentes más frecuentes de una serie cronológica son: Tendencia, Estacionalidad, Ciclos y Perturbaciones aleatorias.

- **Tendencia.** La tendencia representa el movimiento a largo plazo de la serie, el cual puede ser considerado como el valor en torno al cual se mueven los demás componentes.
- **Estacionalidad:** Representa el movimiento de forma periódica que tiene lugar para períodos menores a un año, depende de la forma en que se toma la información que puede ser: mensual, trimestral, semestral etc. Los factores que se reflejan en esta componente, tienen un comportamiento más inestable que la tendencia, por esto, resulta importante el estudio de este componente cuando se quieren hacer pronósticos a corto plazo.
- **Ciclos:** Es un movimiento oscilatorio y ondulatorio en la serie cronológica que se diferencia de la estacionalidad, solo en la periodicidad, pues mientras la estacionalidad tiene período de un año, el ciclo no tiene periodicidad fija y siempre tendrá longitud superior a un año. La componente cíclica a pesar de hallarse presente en la mayoría de las series cronológicas no siempre resulta necesaria su inclusión en los análisis. Esta componente es fundamental en la Meteorología y la Sismología, así como en la Medicina y la Biología.
- **Perturbación aleatoria:** Desviación del comportamiento normal por causas aleatorias.

Estas componentes dan lugar a los modelos clásicos de las Series cronológicas:

- Aditivo: $P = T + E + C + A$
- Multiplicativo: $P = T * E * C * A$
- Mixto: $P = T * E + C * A$

Donde "T" es la tendencia, "E" es la estacionalidad, "C" es la ciclicidad y "A" la aleatoriedad.

Espallargas y Solis (2012), argumentan que con el enfoque clásico en el análisis de series temporales más bien se describe el comportamiento del proceso en estudio.

I.1.2 Series temporales, su empleo en la predicción y detección de patrones anormales en la aplicación de la vigilancia sindrómica

Desde hace algunas décadas se ha fortalecido el enfoque de análisis de las series temporales, con el objetivo de realizar pronósticos de valores futuros de la variable analizada y para este caso, los métodos clásicos resultan inoperantes. Además se alega que en general las series temporales se caracterizan por la dependencia que existe entre observaciones sucesivas y por consiguiente una marcada presencia de multicolinealidad.

A partir de la década de los años 70 el enfoque estocástico o aleatorio en el análisis de las series temporales, aportó nuevas herramientas de análisis a partir de la teoría desarrollada por Box y Jenkins (1976). También se puede encontrar con la denominación de Modelos ARIMA (*Autorregressive Integrated Moving Average del inglés*) o Modelos Autorregresivos (AR) Integrados de Medias Móviles.

Autores como Peña (2005); Ramírez y col. (2005), Hoinville y col, (2013), Dórea y col. (2013) entre otros, consideran que Box y Jenkins (1976) desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental y alegan que la metodología consiste en introducir componentes aleatorias a las ecuaciones para que les permita una mayor flexibilidad y se compone de los siguientes modelos:

- *Autorregresivos* [AR(p)].
- *Medias Móviles* [MA(q)].
- *Mixtos* [ARMA(p,q)].

- *Autorregresivos integrados de medias móviles* [ARIMA(p,d,q)S].

Donde

p= número de parámetros de la parte autorregresiva ϕ

d= número de diferencias para estacionarizar en media

q= número de parámetros de la parte de medias móviles θ

s= Periodicidad (sobra).

Los autores antes mencionados señalan que: en Estadística, un Modelo Autorregresivo Integrado de Medias Móviles (ARIMA) es una generalización de Modelo Autorregresivo de Media Móvil (ARMA) y que estos modelos se ajustan a los datos de series temporales para entender mejor los datos o para predecir puntos futuros de la serie. El modelo conocido como un modelo ARIMA (p, d, q), donde p , d y q son enteros iguales o mayores que cero. Además reportan que según la teoría descrita, al modelo ARMA (p, q) generalizado, se le agrega el parámetro d para crear el modelo ARIMA (p, d, q), donde d es un entero positivo, la cual se expresa en esta función

$$Z_t = \partial + U_t + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} - \theta_1 U_{t-1} - \dots - \theta_q U_{t-q}$$

Dónde:

$$U_{t-1} \sim n(0, \sigma_u^2)$$

∂ : Promedio.

U_t : Error aleatorio.

ϕ_{t-j} : Ponderaciones de procesos a tiempos atrás.

Z_{t-j} : Valores autorregresivos.

θ_{t-j} : Ponderaciones de errores a tiempos atrás.

U_{t-j} : Valores de medias móviles.

σ_u^2 : Varianza de errores aleatorios.

Los autores Box y Jenkins (1976) en sus artículos reflejan una metodología para la identificación de modelos ARIMA en series temporales los cuales consisten en seguir diferentes pasos fundamentales para lograr un buen modelo, los mismos consisten en:

1. **Identificación**
2. **Estimación de Parámetros**
3. **Verificación del modelo**
4. **Las predicciones**

Los autores antes mencionados describen detalladamente cada uno de estos pasos, los que se resumen a continuación:

1. Identificación: El objetivo principal es determinar si es necesario realizar transformación a los datos para estacionalizar la serie de tiempo en media y en varianza. La representación gráfica de la serie nos indica si esta es estacionaria o no a través de la función de autocorrelación simple (FAS) y su función de autocorrelación parcial (FAP). En caso que la serie no sea estacionaria, se debe aplicar uno de los siguientes procedimientos

a) Si tiene tendencia: Se aplicará **diferencias regulares** hasta que desaparezca. Normalmente el orden de la diferencia será 1, y raramente será mayor a 3.

b) Si es **heterocedástica**: en este caso se refiere que no tiene varianza constante, por lo que habrá que transformar la serie. Para ello se aplica logaritmo.

c) Si es **estacional**: se tomaran **diferencias estacionales** hasta que desaparezca el patrón que se repite. En la práctica es muy raro tener que aplicar más de una diferencia estacional.

Una vez obtenido el gráfico de la nueva serie transformada se puede intentar deducir su estructura observando la FAS y la FAP. Observando los autocorrelogramas se hace una primera idea de qué modelo subyace a la serie, o al menos de cuáles son los primeros modelos que se deben probar. A partir de ahí se ajusta el modelo $ARIMA(p,d,q)$ según lo que se haya deducido de todos los apartados anteriores. Frecuentemente ocurre que se deba probar varios modelos $ARIMA(p,d,q)$ hasta encontrar el modelo que mejor se ajuste .

2. Estimación de Parámetros Después de la elección del modelo se pasa a buscar el valor de los parámetros a través de técnica de estimaciones no lineales, con el objetivo que el modelo represente apropiadamente a la serie. Una de las técnicas para la realización de lo anteriormente mencionado es el método de Gauss-Newton

3. Verificación del modelo: antes de la utilización del modelo con fines de predicciones, se debe someter a pruebas partiendo de que todo modelo es erróneo puesto que son representaciones de la realidad. La forma de aprobar el modelo es a través de la validación de los supuestos que los mismos sirven para identificar, estimar y verificar el comportamiento de un modelo $ARIMA$ que ha sido optimizado, estos son:

Parsimonioso: un modelo se dice que es parsimonioso si se ajusta a la serie de forma adecuada sin usar coeficientes innecesarios. Por ejemplo, si un modelo $AR(1)$ y un modelo $AR(2)$ se comportan de forma prácticamente idéntica, elegiremos el modelo $AR(1)$ ya que así tendremos que estimar un coeficiente menos. El principio de parsimonia es importante porque, en la práctica, un modelo parsimonioso suele generar mejores predicciones. La idea de la parsimonia nos da una fuerte

orientación práctica a la hora de modelar e identificar un modelo ARIMA. Así, no tendremos que buscar el proceso ARIMA que realmente genera la serie temporal, sino que nos conformaremos con encontrar un modelo que se aproxime correctamente,

Estacionario: otra condición de gran importancia para lograr un buen modelo ARIMA es que la serie sea estacionaria. Asumir que una serie sea estacionaria permite desarrollar un marco de trabajo bastante simple y usar herramientas estadísticas de muestreo de gran potencia. Así, si la media de un proceso es constante, se podrá usar N observaciones para estimarla, mientras que sería mucho más complicado si la media no fuese estacionaria. Las técnicas más comunes para conocer si una serie temporal es estacionaria son los contrastes de existencia de raíces unidad. Si una serie no es estacionaria podemos modificar dicha serie para convertirla en estacionaria.

Los residuos: son ruido blanco: esta proposición es muy importante a la hora de verificar un modelo ARIMA, una vez se han realizado las etapas de identificación y ajuste. La hipótesis crítica es la de incorrelación, para comprobar esta hipótesis se utilizan distintos métodos de inferencia estadística (típicamente contrastes t y chi cuadrado) aplicados a cada coeficiente de la función de autocorrelación y a la función de autocorrelación parcial.

4. Las predicciones: Aunque el modelo haya sido ajustado y prediga el pasado de una forma suficientemente correcta, lo que realmente se requiere de cualquier modelo de predicción es que realice predicciones satisfactorias. La evaluación de un modelo según este criterio se debe realizar mediante el uso durante un período de prueba o de verificación. La predicción es el fin último y primordial del análisis univariante de series temporales, Intuitivamente, el mejor predictor posible será "el que menos se equivoca" o, en términos estadísticos, aquel que minimiza el error cuadrático medio respecto a otro potencial predictor alternativo

Es importante resaltar que cada uno de estos pasos con una adecuada ejecución del modelo ARIMA permite la estabilización de la varianza debido a la aplicación del operador de diferencia, logrando así tener un modelo más simple y preciso. Ya logrado esto se procede a elegir el mejor modelo, para ello existen estimadores de calidad relativa de los modelos estadísticos, entre ellos el error cuadrático medio (MS) y el criterio de información Akaike (AIC).

En la práctica autores como Lombardo y col. (2004) refieren que el uso de múltiples algoritmos en paralelo se ha explorado a través del uso de reglas de decisión y la utilización conjunta de los resultados de diferentes algoritmos, o mediante el uso de pruebas de bondad de ajuste que permitan decidir cuándo cambiar entre los algoritmos. Por otra parte MacNab (2011) y Lawson (2013) consideran que otro aspecto que ha favorecido el empleo de estos procedimientos ha sido gracias al desarrollo computacional de hardware y el software cada vez más rápidos y potentes.

En la actualidad se cuenta con valiosos resultados relativos al empleo de estos procedimientos por ejemplo en el área de la salud pública en estudios asociados a la vigilancia de enfermedades transmisibles, para la detección precoz del comportamiento anormal de eventos de salud y la posibilidad de preparar intervenciones oportunas que garantizan a las autoridades sanitarias la reducción de las consecuencias, en ocasiones letales, derivadas de enfermedades y otros daños, también en estudios asociados al análisis de la tasa general de desempleo abierto realizado en el Estado de Veracruz, de igual forma en la última década se incrementan investigaciones asociadas a la sanidad agropecuaria y formas de vigilancia epidemiológica. Todas estas aplicaciones es están avaladas en los trabajos de: Ramírez y col. (2005); Cautín (2007); Jackson y col. (2007); Beltran-Alcrudo (2009); ramirez y col. (2011); Hoinville y col. (2013); Dórea y col. (2013).

I.1.3. Vigilancia sindrómica. Empleo de series temporales en la predicción de valores y detección de patrones anormales

A nivel mundial se desarrollan nuevas formas de vigilancia epidemiológica con el fin de mejorar la eficiencia en los sistemas de detección de enfermedades. En este sentido, Ramírez y col. (2005) expresan que el desarrollo de nuevas formas de vigilancia implica la incorporación de medidas de control adecuadas y el incremento en las capacidades diagnósticas y tecnológicas dentro de las que se incluye el empleo de herramientas estadísticas. Una alternativa para enfrentar este reto es mediante la **vigilancia sindrómica**

Se han dado diversos criterios sobre el concepto de vigilancia sindrómica, Hoinville y col. (2013) la define como: la colecta, medición, procesamiento, comparación, análisis sistemático y diseminación oportuna de datos relacionados con la salud y el bienestar animal en poblaciones definidas. Además, señala que la **vigilancia sindrómica** está basada en el empleo de datos relativos a la salud (signos clínicos u otros) que puedan anteceder o sustituir un diagnóstico formal y señalar con suficiente probabilidad de ocurrencia un caso o un brote de enfermedad, de forma que garantice una respuesta oportuna.

En este sentido Ward y Carpenter (2000), German y col. (2001) y Hasler y col. (2011) plantearon que la vigilancia sindrómica tiene como objetivo fundamental disminuir el tiempo desde el momento en que se realiza el diagnóstico hasta que se lanza la alerta epidemiológica; lo cual favorece la aplicación de acciones oportunas para la reducción de la morbilidad y mortalidad.

Los estudios más relevantes sobre la vigilancia sindrómica, se ha aplicado principalmente en la investigación de la salud pública. Sin embargo, en los últimos años ha habido una tendencia creciente hacia la aplicación de estos métodos para los datos de vigilancia de la salud de los animales según criterios de Vourc'h y col. (2006) , Van Metre y col. (2009) y Katz y col. (2011).

En este sentido el autor Beltran-Alcrudo (2009) expresan que a nivel internacional se eleva el desarrollo alcanzado en la aplicación de nuevos procedimientos y tecnologías que se aplican a la vigilancia sindrómica en la salud. El mismo propone sistemas de alerta temprana similares para detectar circulación de cepas de influenza aviar baja patogenicidad en granjas de ponedoras comerciales, a partir del monitoreo de datos de curvas de puestas y mortalidad .

Dórea y col. (2013) consideran que el proceso para determinar un brote comienza con la identificación de un número aberrante de casos notificados (sospechosos o confirmados) de una enfermedad particular en un tiempo y espacio determinado. Estos algoritmos varían en la forma en que calculan un límite de detección, pero la existencia de dicho umbral para la generación de una alarma es una característica común entre todos los métodos de detección, que van desde simples umbrales fijos a la aplicación de métodos estadísticos complejos

Por otra parte Alba y col. (2015) alegan que la combinación de métodos de análisis de series temporales tradicionales con otros métodos novedosos pueden ayudar, no sólo a diseñar sistemas más eficientes de vigilancia sindrómica, sino también a evaluar el impacto de las medidas de intervención de salud aplicadas o los cambios que se producen durante períodos de tiempo prolongados.

La producción porcina no está expensa de sufrir brotes de enfermedades que afecta a la misma y por ello es necesario el estudio y actualización de sistemas de vigilancia que ayuden a mantener el sistema de producción

I.2 El cerdo en Cuba. Principales factores relacionados con el manejo y la bioseguridad que inciden en la productividad y la salud de la porcina.

Los cerdos están entre las especies domésticas más idóneos para ser explotados en cualquier medio (FAO, 2002). Dada su característica prolífera y su capacidad de adaptación a diferentes condiciones de manejo y alimentación no convencional, estos animales se convierten, cada vez más, en la principal fuente de proteína

animal para el consumo humano (Delgado, 2003). En Cuba, el Ministerio de la Agricultura brinda en la actualidad una atención especial al desarrollo de la actividad porcina tanto en la pequeña y mediana escala, como en los sistemas de producción intensiva (Pires y col. 2007).

La producción intensiva de cerdos descansa sobre dos pilares fundamentales: la selección genética y las prácticas de manejo y alimentación que se establecen. La intensidad del proceso de selección genética favorece, entre otros factores deseados, la alta prolificidad, rápido crecimiento y capacidad para engorde, pero determina una baja resistencia a las enfermedades. Dentro de las diferentes acciones del manejo, el destete precoz, el poco tiempo de descanso de las instalaciones entre ciclos productivos, el hacinamiento y las fallas en el saneamiento, condicionan la persistencia de microorganismos enteropatógenos en el ambiente (ACPA, 2008).

En este sentido Quiles y Mendez (2004) plantean que la especie porcina se caracteriza por presentar un porcentaje de mortalidad neonatal muy elevado en comparación con otras especies como la bovina, ovina o equina, constituyendo aproximadamente del 10 al 15% de los lechones nacidos vivos, a pesar de contar la porcinocultura con una de las más modernas tecnologías en Producción Animal.

Por lo que es de suma importancia conocer todos aquellos aspectos relacionados con la mortalidad en la categoría cría, para poder manipular y atender a los lechones en las mejores condiciones y tomar las medidas necesarias para evitar pérdidas por concepto de mortalidad en este período crítico de la vida productiva del cerdo (Lazo y col. 2011).

CAPÍTULO II. EMPLEO DE MODELOS AUTOREGRESIVOS (ARIMA) PARA LA VIGILANCIA SINDRÓMICA EN CRÍAS DE CERDOS

II.1 MATERIALES Y MÉTODOS

La investigación se realizó con datos correspondientes a una granja porcina ubicada en la Provincia de Artemisa (figura II.1) que labora bajo condiciones de explotación comercial.

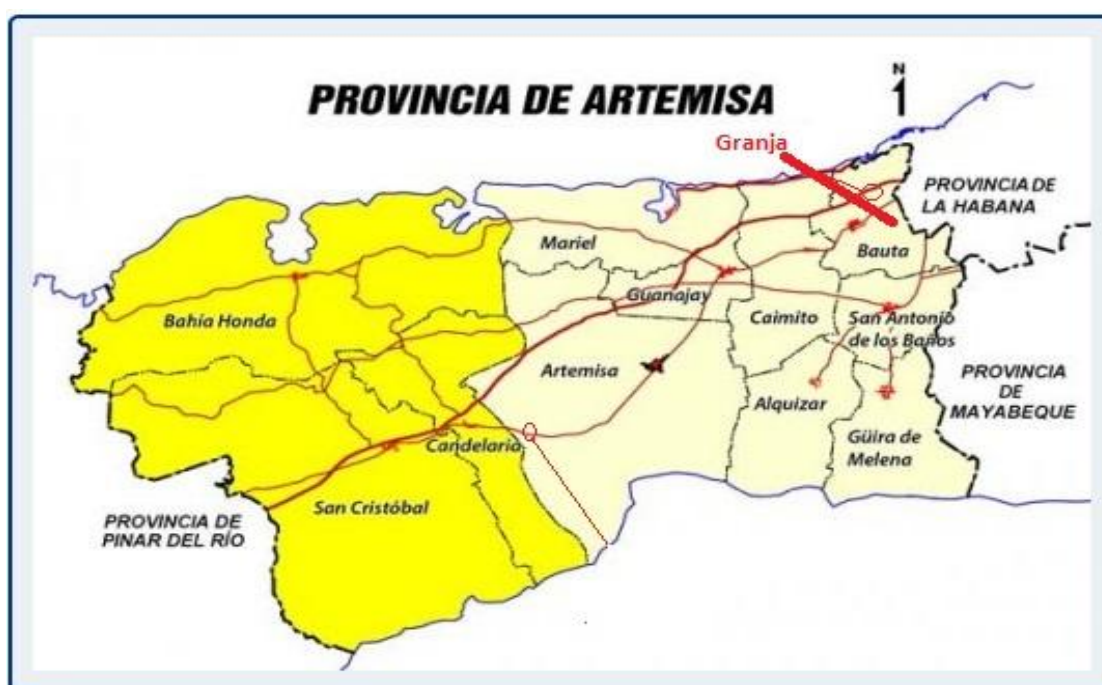


Figura II.1 -Localización Geográfica de la granja porcina.

La unidad de producción realiza rutinas de manejo necesarias pertenecientes a la etapa en estudio, las mismas son ineludibles para lograr un buen sistema de producción, las cuales cumplen lo establecido según en Manual de Procedimientos Técnicos para la Crianza Porcina, CUBA MINAG (2016).

Caracterización y descripción de los datos seleccionados

Se cuentan con registros mensuales durante ocho años (período 2010-2017) de los indicadores:

- Número de Nacimiento
- Mortalidad

Se concibió enfocarse en una primera etapa (desde el nacimiento hasta el destete) Los datos fueron recopilados manualmente por el personal de trabajo de la unidad, los mismo fueron capturados en una hoja de cálculo de Excel.

Números de Nacimiento

Se contó con un total de 96 observaciones número de nacimiento mensual. La suma del número de nacimiento de todos los meses indica un total de 29 324 nacimientos en los ocho años correspondientes al período, con distribuciones por año como muestra la Figura II.2.

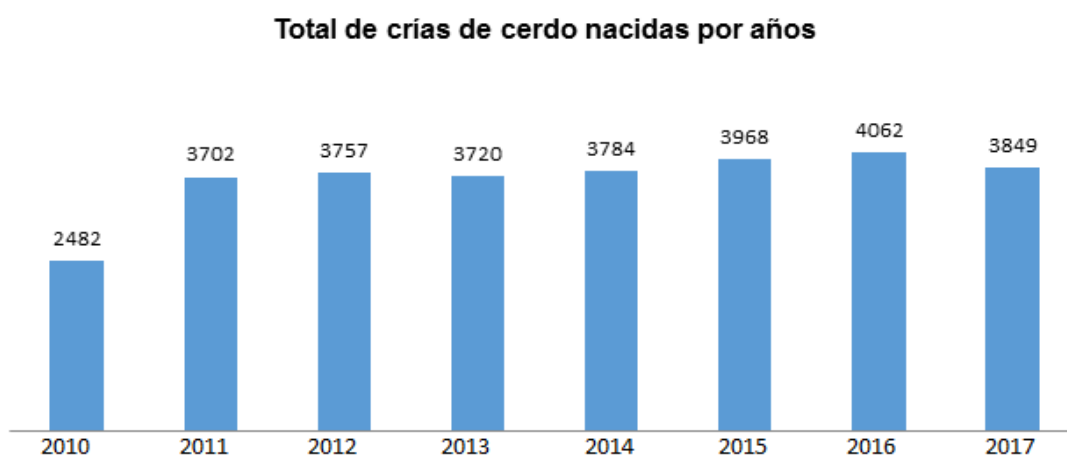


Figura II.2. Total de crías de cerdos nacidas por años período (2010 - 2017).

Mortalidad

Se contó con un total de 96 observaciones del indicador de mortalidad, cada año se analizó por mes el total de por ciento de mortalidad alzado en el mismo. Y un segundo período (2013-2017) un total de cinco años y 60 observaciones mensuales el cual se utilizó ambos períodos para trabajar en el capítulo de series de tiempo.

Con la finalidad de analizar el comportamiento productivo y de salud en la primera etapa (desde el nacimiento hasta el destete) en la unidad porcina en estudio, se realizó una estadística descriptiva (variables número de nacimiento y mortalidad por año) que incluyó medidas de tendencia central, posición y de dispersión. Se graficaron los registros de ambas variables de cada año por meses para ver el comportamiento de las mismas y poder observar los momentos bruscos de caída en caso del número de nacimiento y el aumento en el caso de la mortalidad.

Para hacer un análisis más exhaustivo y determinar si existe diferencia estadísticamente significativa entre la cantidad de nacimientos promedios y en mortalidad (teniendo en cuenta el factor año y el factor mes) se aplicó el Análisis de Varianza Simple (ANOVA). Y en caso necesario se aplicó pruebas de comparación Múltiple (Duncan, 1955).

Posteriormente y con el objetivo de construcción del modelo ARIMA se aplicó la Metodología de (Box y Jenkins, 1978) que incluye desde Identificación, Estimación de Parámetros, Verificación del modelo hasta las predicciones. Una caracterización de estos pasos se describen como:

En la **Identificación y Representación gráfica de la serie**, se realizó una exploración de datos y su representación gráfica, con el objetivo de apreciar los detalles más relevantes de la serie así como Aleatoriedad, Tendencia, Periodicidad, Outliers, y la estacionalidad de la serie (en caso de necesario se le

aplicó transformaciones a la serie, para estacionarizar en varianza, el filtro de logaritmo y para estacionarizar en media el filtro diferenciación).

Para la **Estimación de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF)**. El primero se realizó para determinar el valor q (medias móviles) mientras el segundo su finalidad fue conocer el valor p (autorregresivo) del modelo. Así como observar más detenidamente si existe periodicidad en los datos.

En el caso **Obtención del mejor modelo ARIMA** un criterio básico de bondad de ajuste fue el **(criterio de MS)**, posteriormente se procedió a la verificación del modelo, donde el mismo debe cumplir los supuestos ser Admisible, Parsimonioso y Estable y verificar que los residuos se distribuyan normalmente, con media cero y varianza constante.

Por último se realizó la **Predicción de valores** donde una vez ajustado el modelo se pronosticó el número de nacimientos de cerdos en la granja para dos meses siguientes al último registro Diciembre 2017, para establecer los niveles bajos en el número de nacimiento se tuvo en cuenta el límite inferior predicho mensuales para determinar el valor de alarma.

Al analizar la serie mortalidad (**Identificación y Representación gráfica de la serie**) se procedió a seleccionar un subciclo ya que al inicio del período de la serie original (2010-2017) tuvo gran variabilidad con puntos aberrantes, por ello se analizó y se obtuvo una nueva serie recortada (2013-2017), la cual no considerara los meses que pertenecen a los tres primeros años.

Para determinar el valor representativo como alerta se procedió luego de pronosticar los límites inferiores del número de nacimientos, se halló el promedio de los mismos y se procedió a establecer un único valor como alarma. En caso de la variable mortalidad se tuvo en cuenta el límite superior de ambas series analizadas (serie original y la serie recortada)

Para el procesamiento de los datos, se utilizaron las aplicaciones estadísticas, Statgraphics Plus Centurion versión XV.II Marca registrada por StarPoint Technologies. 2015, Statística INC (2004) y el Microsoft Office Excel, Botman (1999).

II.2 RESULTADOS Y DISCUSIÓN

Nacimiento

En la tabla II.1, se observa que el número de nacimiento promedio mes y por año osciló entre 193 (2011) a 236 (2016) crías de cerdo en el período 2010-2017, con desviaciones estándar entre 24.5 (año 2015) y 64.4 (año 2012). Se destaca que existe variabilidad relativamente alta, según expresa el coeficiente de variación, que fluctúa entre el 11.5 % (año 2015) y de 31.2 % (año 2012) esto pudiera estar asociado a que son datos en condiciones comerciales donde existe una gran variación en el manejo y alimentación de los animales. Existen meses con bajos nacimientos con valor mínimo a 90 nacimientos (año 2011) y otros que el máximo fue de 323 (años 2012 y 2013).

Tabla II.1. Resumen estadístico del indicador nacimiento en crías de cerdo mensual en el período 2010-2017.

	2010	2011	2012	2013	2014	2015	2016	2017
Frecuencia	12	12	12	12	12	12	12	12
Media	206	193	228	230	236	213	206	227
Mediana	208	206	222	237	240	213	208	235
Desviación estándar	26.1	60.3	64.4	47.3	35.5	24.5	26.1	37.0
Mínimo	172	90	103	130	161	173	172	161
Máximo	260	274	323	323	292	263	260	280

Coeficiente variación expresado en (%)	12.6	31.2	28.2	20.5	15	11.5	12.6	16.3
--	------	------	------	------	----	------	------	------

De igual forma en la Figura II.3 se corrobora la alta variabilidad del indicador productivo (nacimientos) en cada uno de los meses durante los ocho años en estudio, donde se observan momentos bruscos de crecimiento y decrecimiento de número de nacidos por meses por año, lo que pueden estar asociado a factores de manejo o de salud en la producción entre otros aspectos

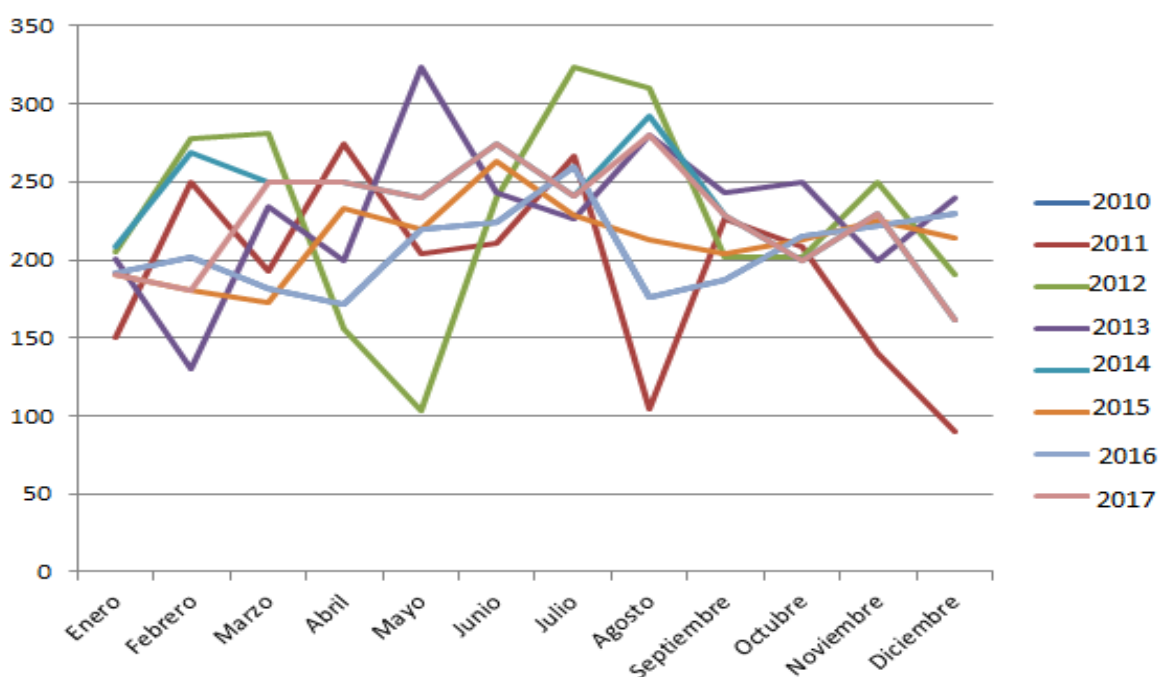


Figura II.3 Números de nacimiento mensual durante los 8 años en estudio.

En tal sentido Martínez (1998) expresó que el número de nacimiento depende de una serie de factores tanto fisiológicos como factores relacionados con el manejo al que son sometidas las cerdas y alega que el número y tamaño de la camada está influenciado por factores como: la tasa de ovulación, el índice de concepción,

la sobrevivencia embrionaria, la tasa de parición, la edad a la pubertad y el intervalo de destete entre otros.

De igual forma se considera que el efecto del verraco es importante sobre el indicador antes mencionado, por lo que el número de lechones nacidos por año en una operación comercial depende de una gran variedad de factores que se deben controlar. Otro aspecto a tener en consideración es la edad y peso al momento del servicio, detección del celo, momento óptimo del servicio y condiciones que afectan al mismo, así como el control de enfermedades infecciosas, las cuales son múltiples. Sobre este tema Martin (1996) plantea que las condiciones que afectan a la cerda durante la gestación pueden ser: la temperatura ambiente y las condiciones de manejo y alojamiento.

La tabla II.2 recoge los resultados alcanzados en el análisis de varianza, lo que indica que no existen diferencias estadísticamente significativas entre el número de nacimientos promedios alcanzado en los diferentes años con p-valor de 0.18

Tabla II.2 Análisis de Varianza (ANOVA) de los nacimientos entre los 8 años de estudio.

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	19236.0	7	2748.0	1.50	0.1776
Within groups	161170.0	88	1831.48		
Total (Corr.)	180407.0	95			

Serie del número de nacimientos de crías de cerdo por meses/año

La Figura II. 4 muestra el número de nacimientos en crías de cerdo mensual durante el período en estudio (2010-2017) en la misma puede observarse que existe variabilidad, outlier y sigue una tendencia no tan marcada que alterna entre

valores altos y bajos, no aparece periodicidad. Muestra el comportamiento de los nacimientos en los 96 meses observados que inicia en enero del 2010 hasta diciembre del 2018, donde existe una ligera tendencia creciente en los primeros meses de la serie y luego hay decrecimientos. De igual forma se aprecia una alta variabilidad, donde los picos más altos de nacimientos aparecen en los años 2012 (julio y agosto) y 2013 (mayo) y el más bajo en el 2011 (diciembre). Lo que indica la necesidad de realizar transformaciones a los datos para estacionar la serie de tiempo en media y varianza.

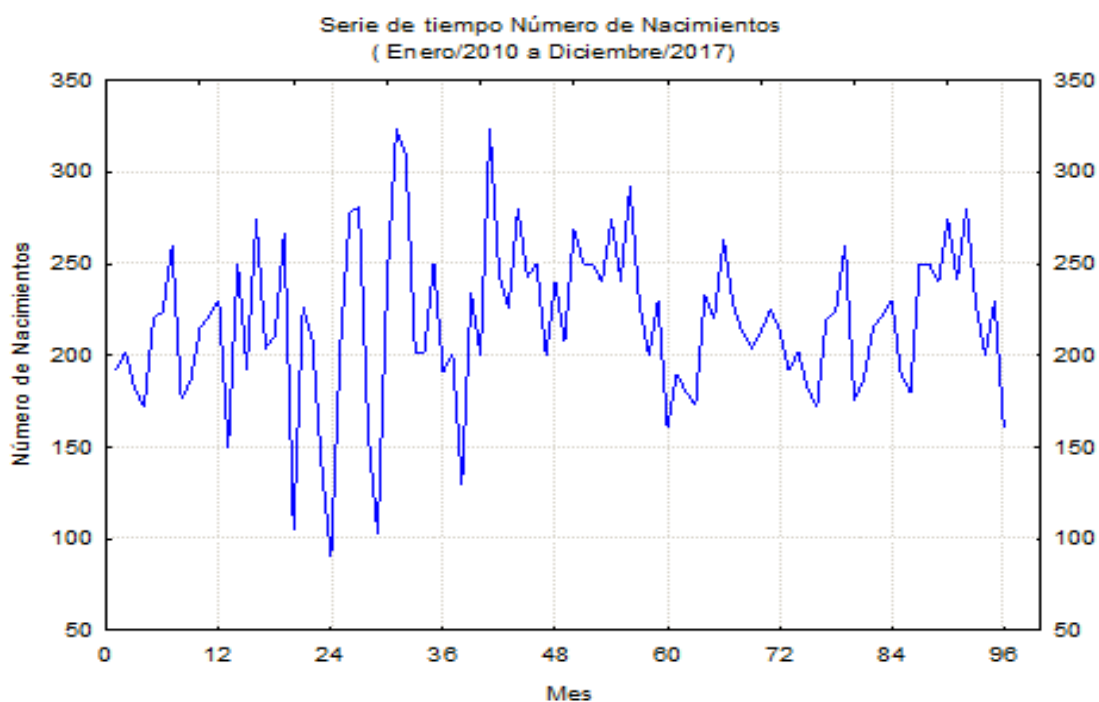


Figura II. 4. Serie sucesión de observaciones del número de nacimientos de cerdo, con registros mensuales (2010- 2017).

Transformaciones

En la figura II.5 se aprecia la transformación de logaritmo natural $\ln(x)$ y Diferenciación $D(-1)$ de la serie en nacimientos de crías de cerdos en el período

(2010-2017). Luego de aplicado los filtros la serie se modificó, eliminando la tendencia que se apreciaba y se logró menor variación. Como se muestra en la siguiente tabla.

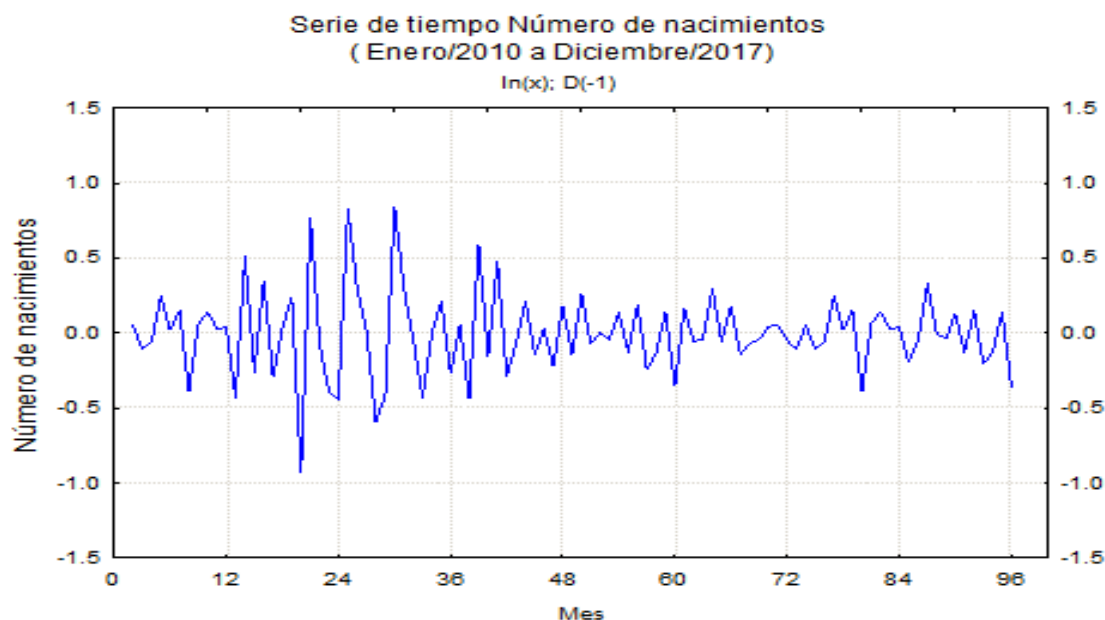


Figura II.5 Serie de tiempo al aplicar las transformaciones de logaritmo natural $\ln(x)$ y Diferenciación $D(-1)$ de la serie de nacimientos de crías de cerdos en el período (2010-2017).

En la tabla II.3 se muestra la estadística descriptiva de la serie original y luego de aplicado los filtros, donde se puede observar que los valores disminuyeron, existiendo menos variabilidad en los datos.

Tabla II. 3. Estadísticas descriptivas de la serie original y transformada.

	Media	Desv. Est.	Min.	Max.	N
Var 1	217.8750	43.5777	90.000	323.000	96
$\ln(x); D(-1)$	-0.0018	0.2839	-0.933	0.845910	95

Función de Autocorrelación y Autocorrelación parcial

La Figura II.6 muestra la gráfica el correlograma de la función de autocorrelación (FAC), la misma señala el valor de (q) que alcanzó el número 1. En el caso la gráfica II.7 se aprecia el Correlograma de la función de autocorrelación parcial (FACP) en la cual se expresa el valor de (p) el cual fue 4.

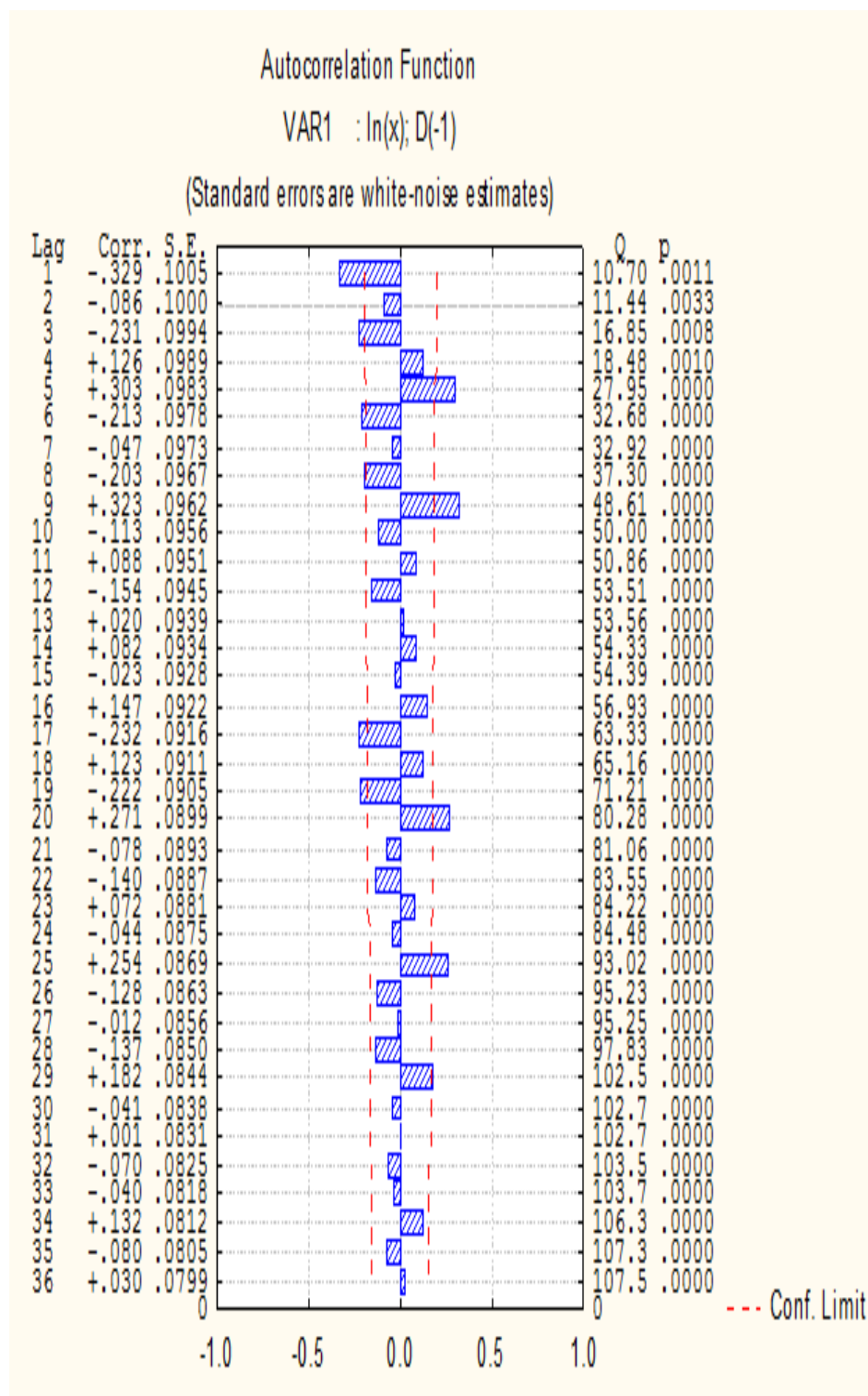


Figura II.6. Correlograma de la función de Autocorrelación (FAC)

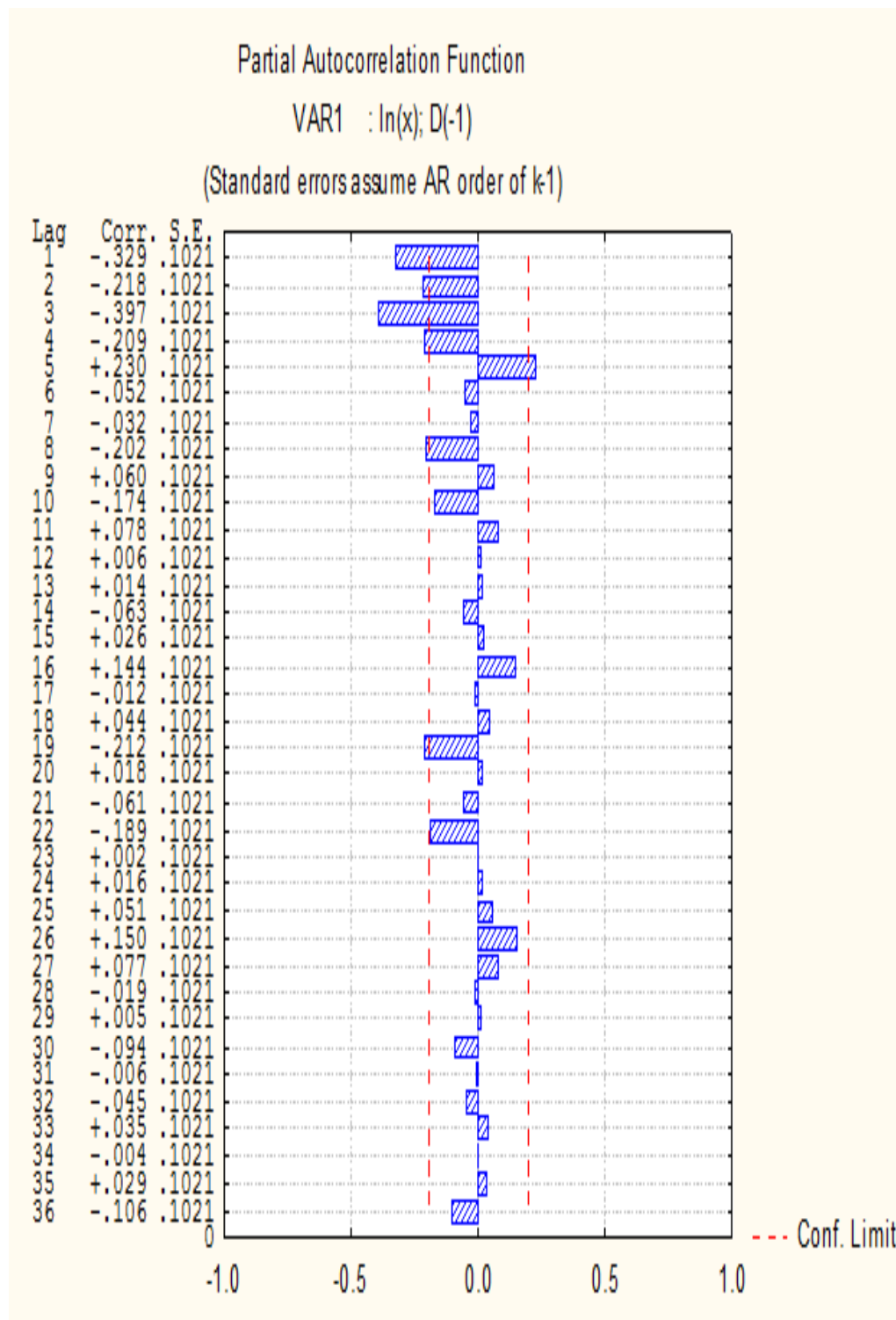


Figura II.7. Correlograma de la función de Autocorrelación parcial. (FACP)

El modelo ARIMA que sugieren los correlogramas observados es el siguiente:

$$ARIMA(p, d, q)_s = ARIMA(4, 1, 1)_1$$

Modelo

$$Z_t = -0.001 + Ut - 0.551Z_{t-1} + 0.9483 + 0.22074 + 0.35843 + 0.99950$$

Obtenido el modelo se verificaron los supuestos:

1.- Admisible: En este modelo el valor $p = 0,9483$ es muy cercano a 1 por lo tanto no es admisible.

2.- Parsimonioso: El modelo no estima intervalos

3.- Estable: No arroja información sobre las correlaciones, por lo que no pueden estimarse.

El modelo que se obtuvo mediante la observación de los Correlogramas no fue bueno dado que no cumplió los supuestos, de ser Admisible, Parsimonioso y Estable. Por lo anterior se propuso el siguiente modelo ARIMA, resultado de verificar el ajuste de varios modelos siendo este el único que se ajustó. Es importante destacar que como se transformó la serie original, el modelo encontrado es el de la transformación, no el de la serie original.

Modelo Autorregresivo propuesto:

$$ARIMA(p, d, q)_s = ARIMA(1, 1, 0)_1$$

Modelo:

$$Z_t = -0.0010 + U_t - 0.3566Z_{t-1}$$

El modelo resultante fue elegido ya que cumplía en ser:

Admisible: Es admisible dado que p es menor a 1

Parsimonioso: El intervalo va de -0.5518 a -.01612 por tanto no incluye el 0

Estable: Los parámetros no se encuentran correlacionados dado que solo existe 1 parámetro en el modelo.

El Valor de MS=0,07128 al ser el único modelo de buen ajuste no fue necesario compararlos con otros.

Residuos:

Como el modelo cumple con los primeros tres requisitos, se procede a validar los supuestos de los residuos del modelo:

La Figura II.8 muestra la gráfica la distribución de los residuales con respecto al modelo, en el cual se puede apreciar la normalidad del mismo.

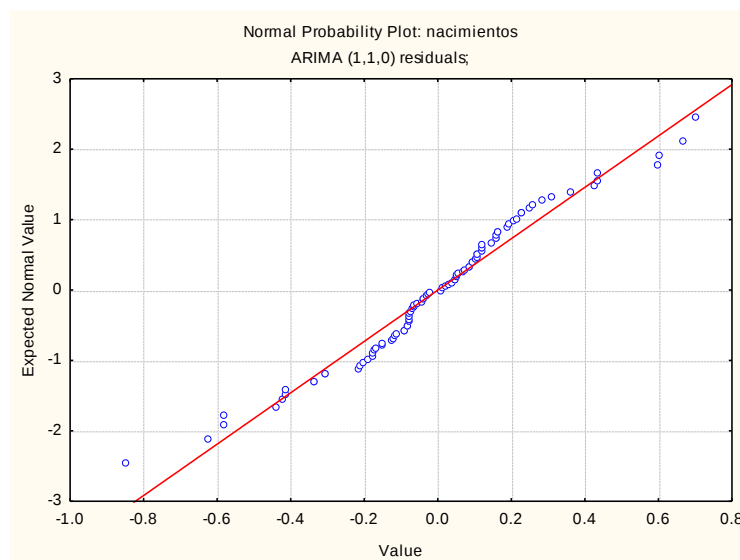


Figura II.8. Distribución normal de residuos.

La Figura IV.9 muestra la gráfica de auto correlación de los residuos para ver la independencia, en este caso se graficaron 50 barras de las cuales se salen de los límites siete y la regla permite que se salgan de los límites a lo más el 5% de las barras graficadas, por lo tanto en este caso no se cumple el supuesto.

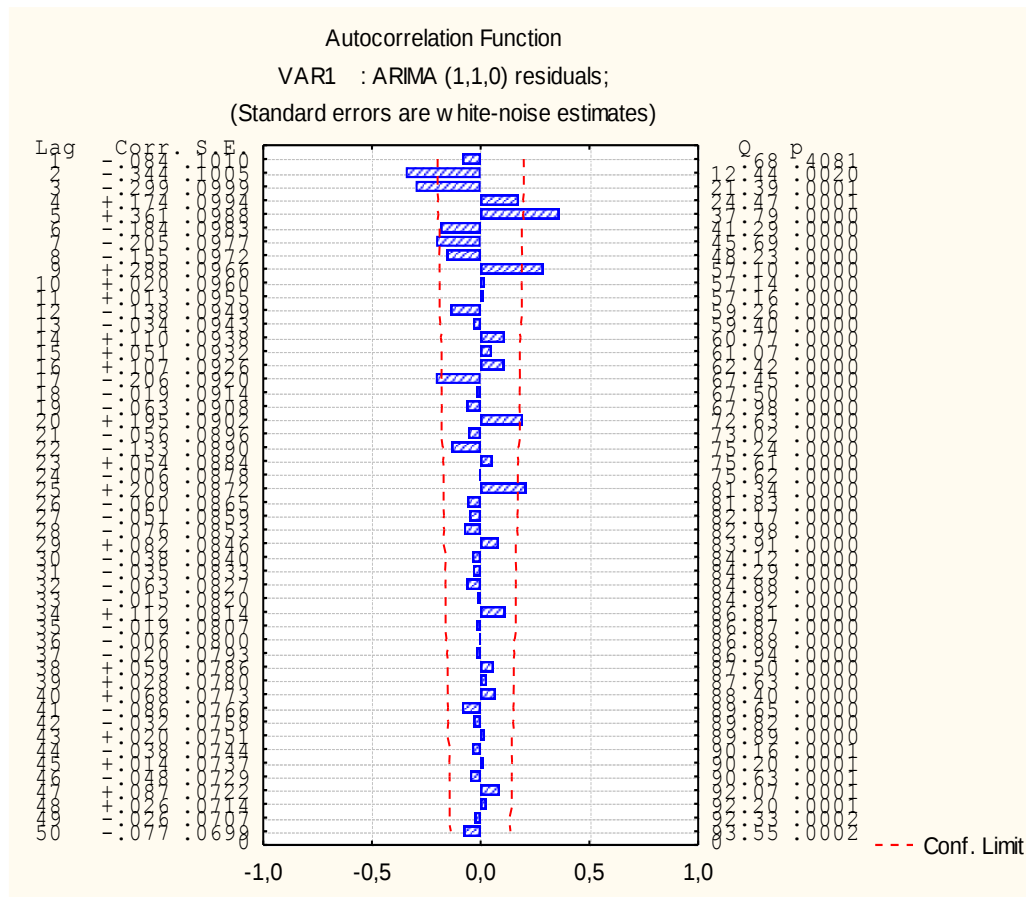


Figura II. 9. Independencia de función de Autocorrelación de residuos.

A pesar de no cumplir con uno de los residuos analizados el modelo es bueno ya que cumple con la mayoría de los parámetros analizados, por ello se procedió a predecir los valores correspondiente a los meses enero y febrero del 2018 y ver los límites superior e inferior que predicho por el modelo.

La tabla II.4 muestra la predicción de dos meses después del valor observado de los nacimientos en crías de cerdo con el modelo $ARIMA(1,1,0)_1$. Se aprecia que los límites inferiores predichos marcan valores de 102 y 117 nacimiento de crías de cerdo en el mes, alcanzando un promedio de 109 nacimientos.

Tabla II.4. Predicción de dos meses después del valor observado de los nacimientos con el modelo $ARIMA(1,1,0)_1$ con un 90 % de confianza.

	Predicciones	Nivel inferior (90%)	Nivel superior (90%)
97(enero 2018)	182	117	284
98(enero 2018)	174	102	295
Promedio de límite inferior		109	

En el análisis para un total de 96 observaciones mensuales del número de nacimiento se pudo estimar que la media general fue **109**, este valor es un indicador de detección temprana de alerta, es decir en este primer estudio se estima que bajo condiciones de explotación comercial cuando el número de nacimiento mensual desciendan a 109 nacimientos en un mes, debe constituir una señal de alarma y activar acciones de investigación en la granja involucrada.

El presente resultado, a pesar de basarse solo en el estudio de una granja, tiene un valor práctico al permitir implementar un sistema de alarmas en esa misma unidad sin necesidad de realizar diariamente procesamiento estadístico complejo de la serie de datos, en particular, a nivel de granja donde por lo general no existen equipos de cómputo.

VARIABLE MORTALIDAD

La tabla II.5 muestra la estadística descriptiva de la mortalidad durante los ocho

años en estudio. Donde los valores medios se comportaron entre 0.090 (año 2010) y 0.109 (año 2011) mortalidad/año, alcanzando los mayores coeficiente de variación en los primeros tres años con los valores que oscilaron entre 12.54 % y 29.93 %, en los últimos cinco años estos valores están por debajo de 4,62 % lo que indica que los datos se comportaron similares en cuanto a la variabilidad relativa alcanzada en cada grupo.

Tabla II.5 Resumen estadístico del indicador mortalidad en crías de cerdo mensual en el período 2010-2017.

	2010	2011	2012	2013	2014	2015	2016	2017
Frecuencia	12	12	12	12	12	12	12	12
Media	0.090	0.109	0.091	0.097	0.096	0.093	0.095	0.095
Mediana	0.097	0.097	0.097	0.097	0.095	0.095	0.096	0.095
Desviación Estándar	0.011	0.032	0.012	0.001	0.004	0.004	0.002	0.002
Mínimo	0.071	0.089	0.061	0.095	0.091	0.085	0.091	0.090
Máximo	0.100	0.180	0.100	0.100	0.100	0.100	0.100	0.090
Coef variación (%)	12.54	29.93	13.77	1.63	4.62	4.69	2.49	2.86

De igual forma se muestra en el Figura II.10 que al realizar el análisis de la mortalidad con gráfica de puntos para cada uno de los meses durante los ocho años, se aprecia durante todo el período en estudio frecuentes pérdidas de cerditos, donde hubo momentos aberrantes, los cuales fueron más sobresalientes en el año 2011.

Mortalidad en Crías

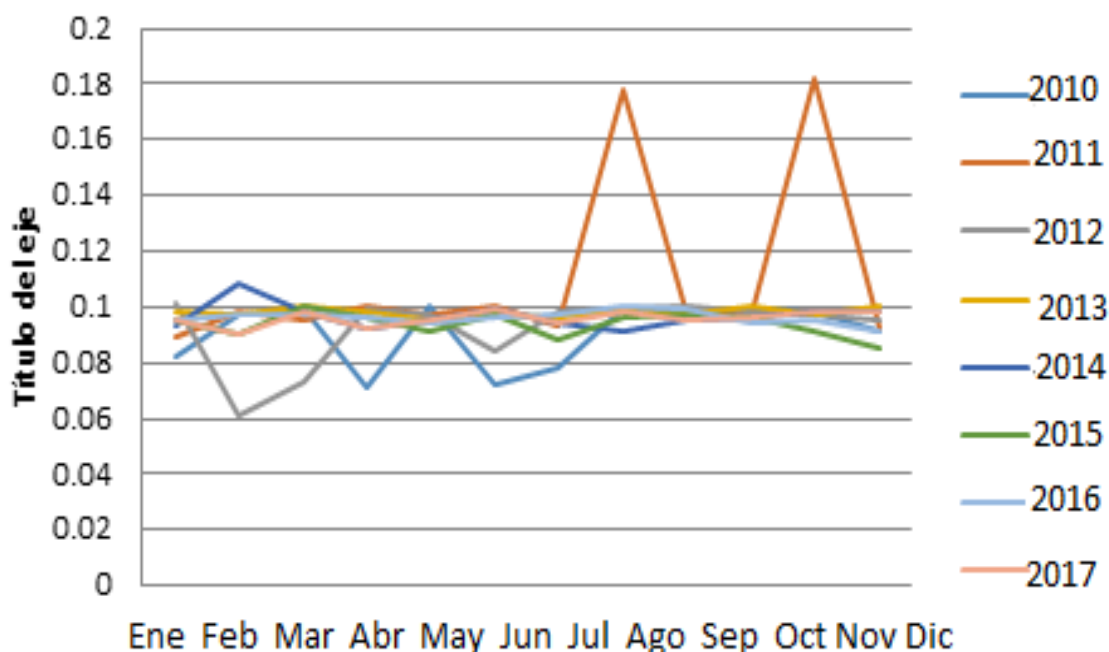


Figura II. 10. Gráfica de mortalidad mensual durante los ocho años en estudio

Las principales causas de mortalidad en esta primera etapa de la producción porcina pueden variar tanto por factores ambientales, de manejo y por enfermedades. Las principales causa de muerte en crías de cerdo en Cuba son, desnutrición, trastornos gastrointestinales, aplastamiento, colibacilosis, sacrificio sanitario y trastornos respiratorios.

El análisis de varianza de la tabla II.11, mostró con un 95% que existen diferencias estadísticamente significativas entre la mortalidad promedios alcanzado en los diferentes años con p-valor de 0.0262.

Tabla II.11 Análisis de varianza (ANOVA) de mortalidad durante los ocho años en estudio

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	0.0030128	7	0.000430409	2.41	0.0262
Within groups	0.0156903	88	0.000178299		
Total (Corr.)	0.0187032	95			

La tabla II. 6 hace referencia a la prueba de rangos múltiples de Duncan (1955) la cual muestra que existe diferencia estadísticamente significativa entre la mortalidad en crías en el año 2011 con respecto al resto de los años en estudio.

Tabla II. 6. Prueba de rangos múltiples de Duncan

Método: 95.0 de Duncan			
	cantidad	media	homogeneidad de grupos
2010	12	0.0903333	X
2012	12	0.0918333	X
2015	12	0.09355	X
2017	12	0.0956667	X
2016	12	0.0959167	X
2014	12	0.0964167	X
2013	12	0.09775	X
2011	12	0.109917	X

Análisis de la serie original de mortalidad

A continuación se presenta la Figura II.12 la cual muestra la serie original de mortalidad en crías de cerdo mensual en el período del 2010-2017. Es de destacar que desde el 2010 hasta 2013 se presentó una alta variabilidad, con momentos

bruscos de crecimiento, en el resto de la serie los valores estuvieron variando pero en general se comportaron entre los valores medios.

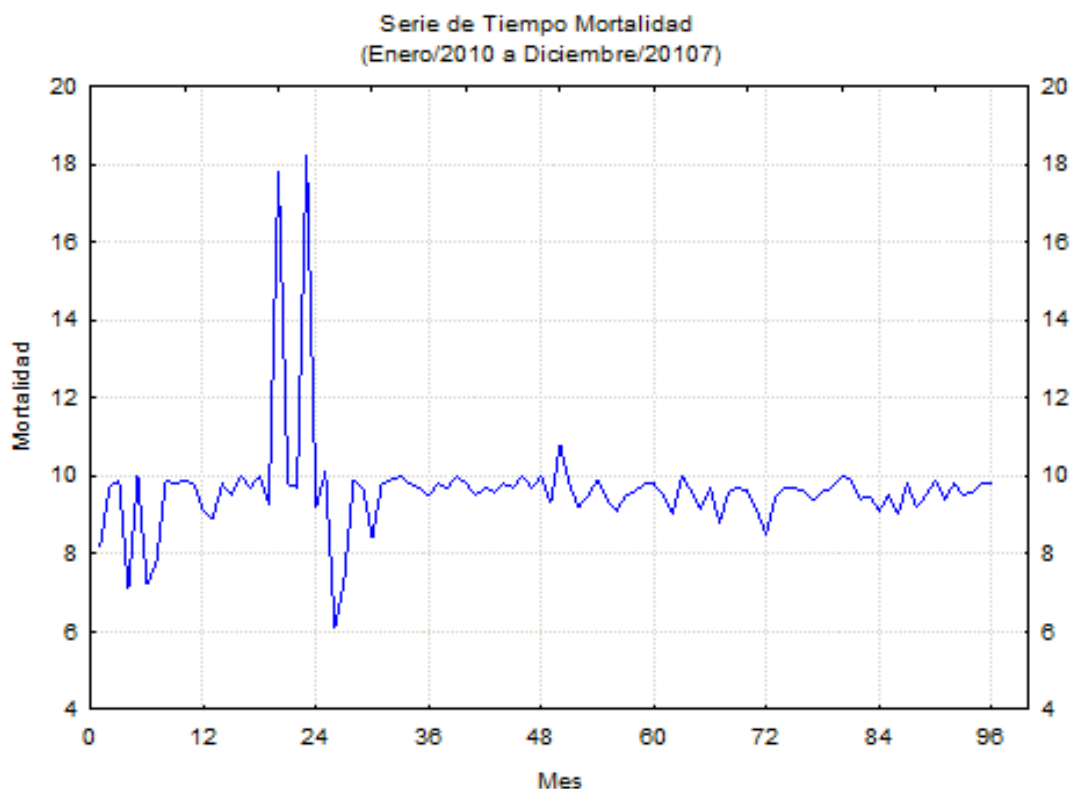


Figura II. 12 Gráfico de la sucesión de observaciones de la mortalidad en crías de cerdo por mes en el período 2010-2017

La misma figura al mostrar variabilidad, consistencia, existencia de valores aberrantes y lo más importante no es estacionaria, fue necesario aplicar transformaciones a la serie, de acuerdo con la metodología de Box-Jenkins.

APLICACIÓN DE FILTROS

La Figura II. 13 muestra las transformaciones de logaritmo natural $\ln(x)$ y Diferenciación $D(-1)$ de la serie de mortalidad original. En la cual se observa luego de aplicado los filtros se aprecian sus modificaciones.

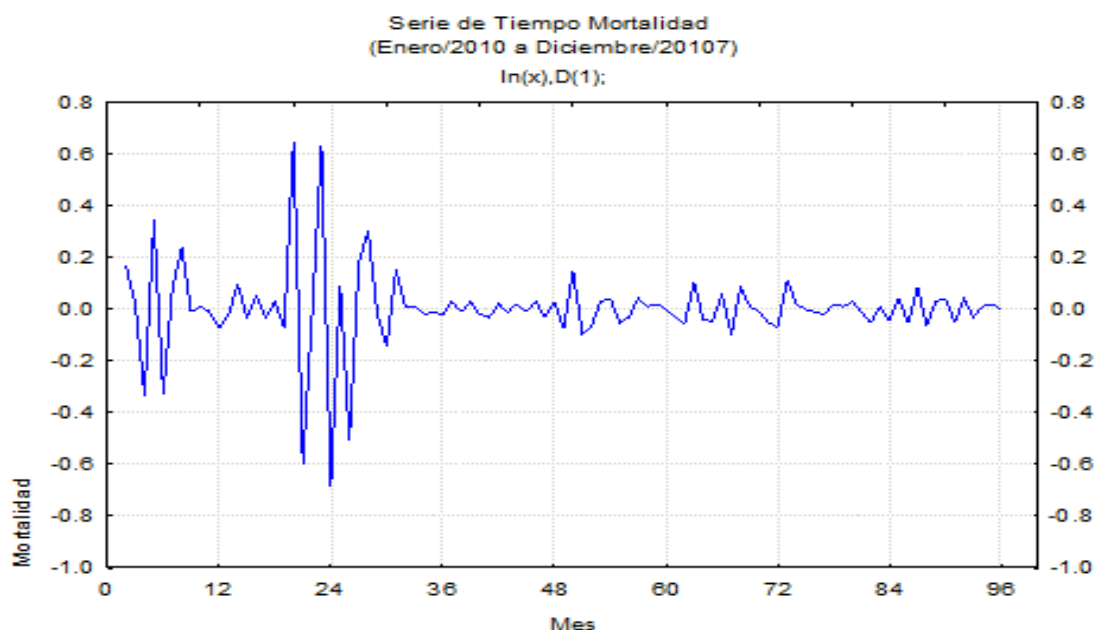


Figura II.13. Transformación de logaritmo natural $\ln(x)$ y Diferenciación $D(-1)$ de la serie de mortalidad original.

La tabla II.7 indica en la serie de tiempo original valores medios de mortalidad en cerditos de 9.6 y una desviación estándar de 1.4. Se puede observar las transformaciones de logaritmo natural $\ln(x)$ y Diferenciación $D(-1)$

Tabla II.7 Estadísticas descriptivas serie original (2010-2017)

	Media	Desviación Estándar	Mínimo	Máximo	N
Variable	9.64	1.4	6.1	18.2	96
$\ln(x);D(-1)$	0.000130	0.169000	-0.682218	0.64918	95

Autocorrelogramas

La Figura II.14 muestra el Correlograma de la función de auto correlación (FAC), la misma muestra el valor de (q) que el alcanzó el valor de 1. En caso de la Figura

II.15 el Correlograma de la función Autorregresiva parcial (FACP) que en la cual se aprecia el valor de (p) el valor obtenido es 2.

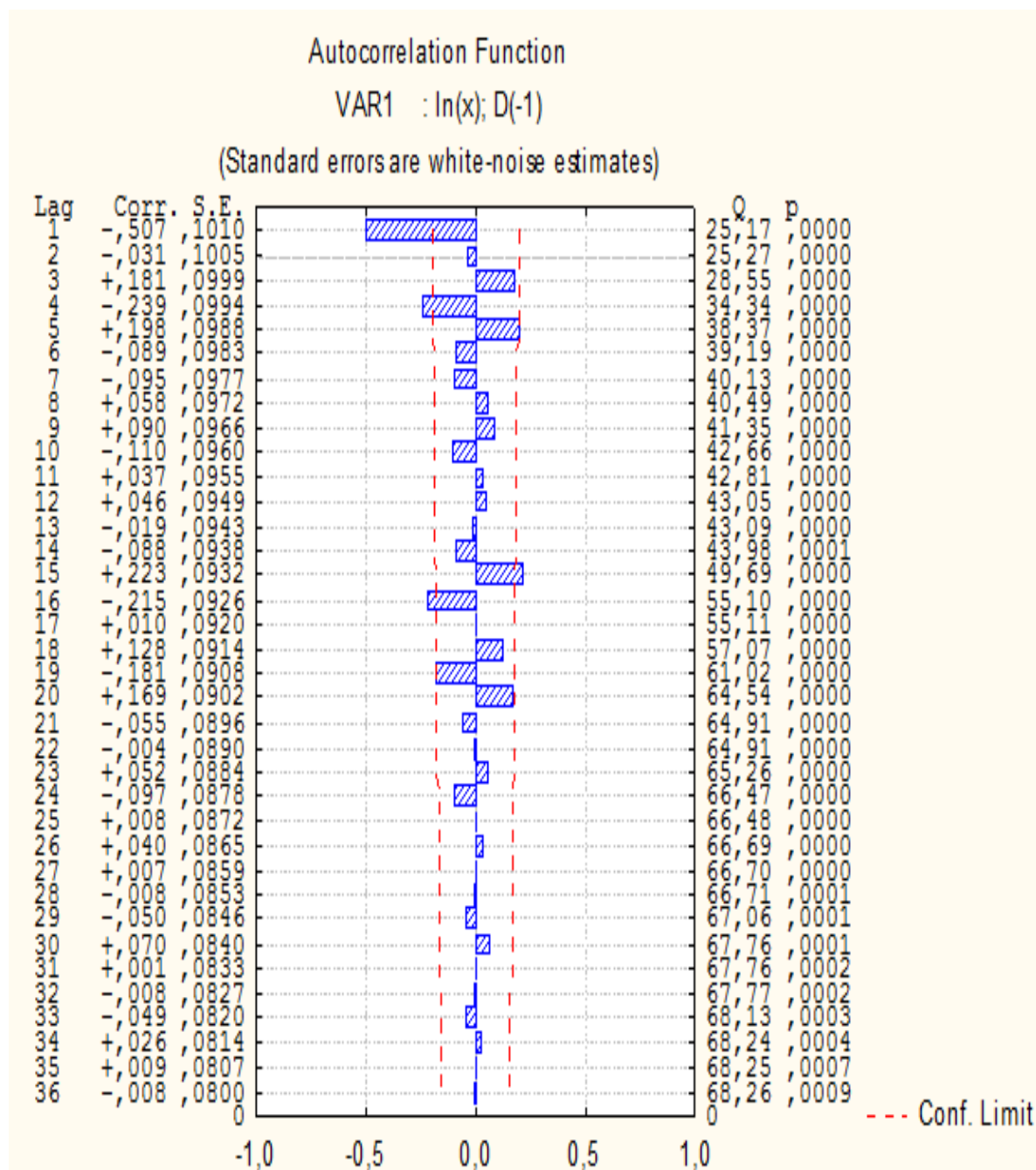


Figura II.14. Correlograma de la Función de Autocorrelación(FAC) de la serie original de mortalidad.

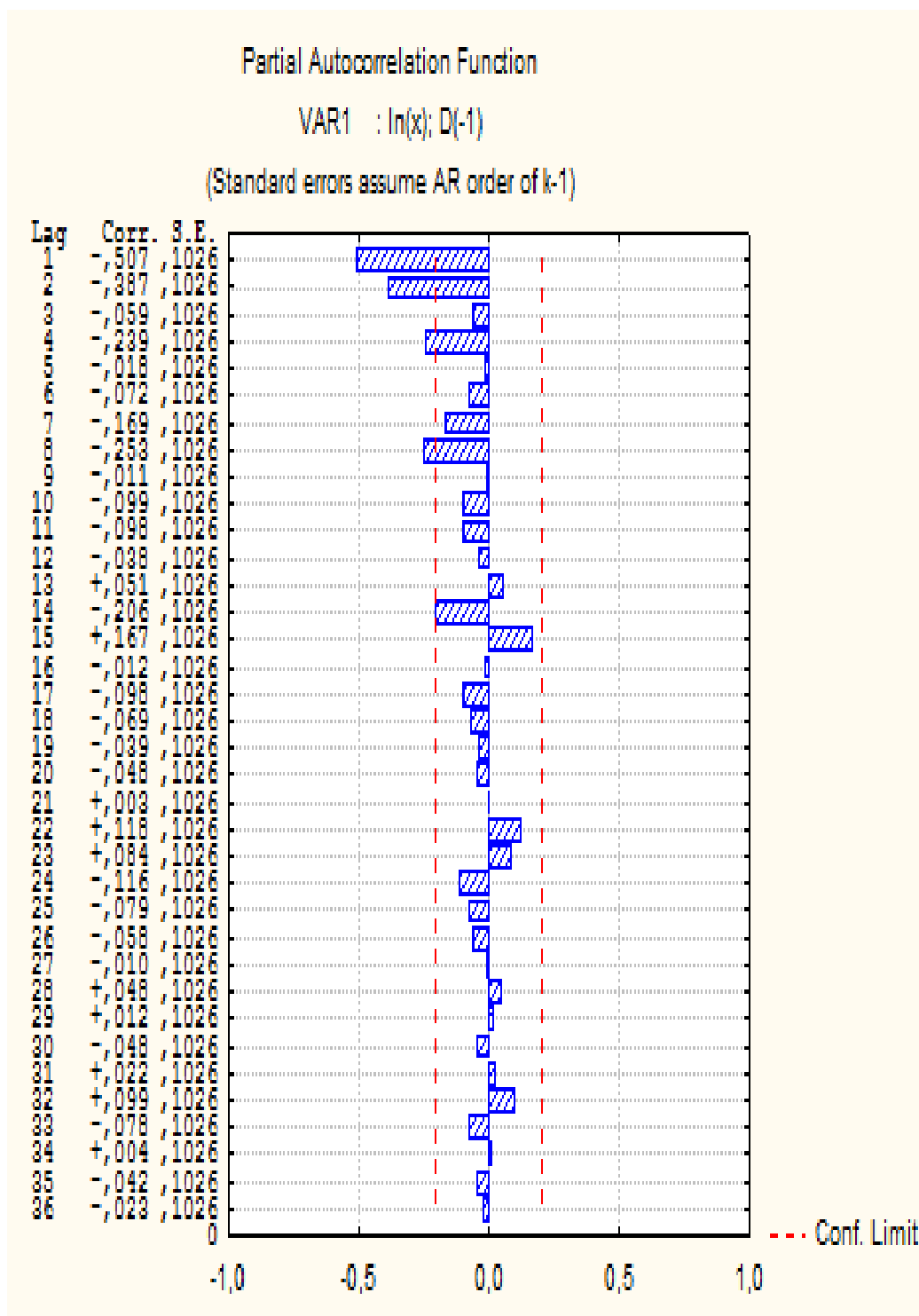


Figura II.15 .Correlograma de la Función de Autocorrelación Parcial (FACP) de la serie original de mortalidad.

Una vez analizado los supuestos se presenta de manera resumen la validación del modelo obtenido por el correlograma de la serie mortalidad en crías de cerdo:

$$ARIMA(p, d, q)s = ARIMA(2,1,1)_1$$

Modelo:

$$Z_t = -0.00121 + Ut - 0.42215Z_{t-1} + 0.89083U_{t-1}$$

Admisible: En este modelo el valor de $p = 0,042149$ no es cercano a 1 por lo tanto es admisible.

Estable: El modelo no es estable ya que los parámetros están altamente correlacionados ($\rho = 0.861582$).

Parsimonioso: No es parsimonioso dado que la estimación puntual y el intervalo con un 95% de confianza del parámetro incluye al cero: en caso de p_1 (-0,208836 a 0,293135) y p_2 (-0,17909 a 0,305534)

Por lo anterior se propuso el siguiente modelo ARIMA, resultado de verificar el ajuste de varios modelos siendo este con el cual se ha obtenido mejor resultado.

Modelo autorregresivo propuesto:

$$ARIMA(p, d, q)s = ARIMA(1,1,0)_1$$

Modelo:

$$Z_t = -0.0126 + Ut - 0.86437 Z_{t-1}$$

El modelo resultante fue elegido ya que cumple con ser:

Admisible: Es admisible dado que p es menor a 1

Parsimonioso: El intervalo va de -0.6852 a -0.03282 por tanto no incluye el 0.

Estable: Los parámetros no se encuentran correlacionados dado que solo existe 1 parámetro en el modelo.

El valor de **MS= 0,016**. Al ser el único modelo que se ajustó no fue necesario compararlo.

Residuos:

La Figura II.16 muestra la distribución de los residuales con respecto al modelo, en este gráfico de probabilidad se muestra que no cumplen con el supuesto de normalidad, esto puede estar dado por los puntos aberrantes que presenta la serie.

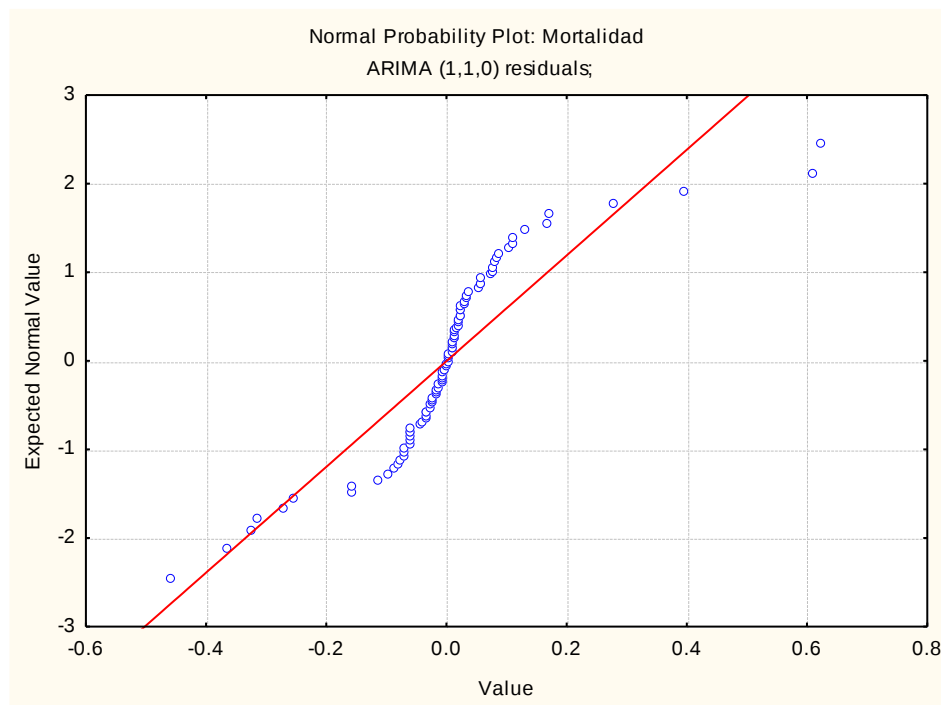


Figura II.16. Distribución de los residuos.

La Figura II.17 muestra la auto correlación de los residuos para ver la independencia, en este caso se graficaron 50 barras de las cuales se salen tres y la regla permite que se salgan de los límites a lo más el 5% de las barras graficadas, por lo tanto en este caso si se cumple el supuesto.

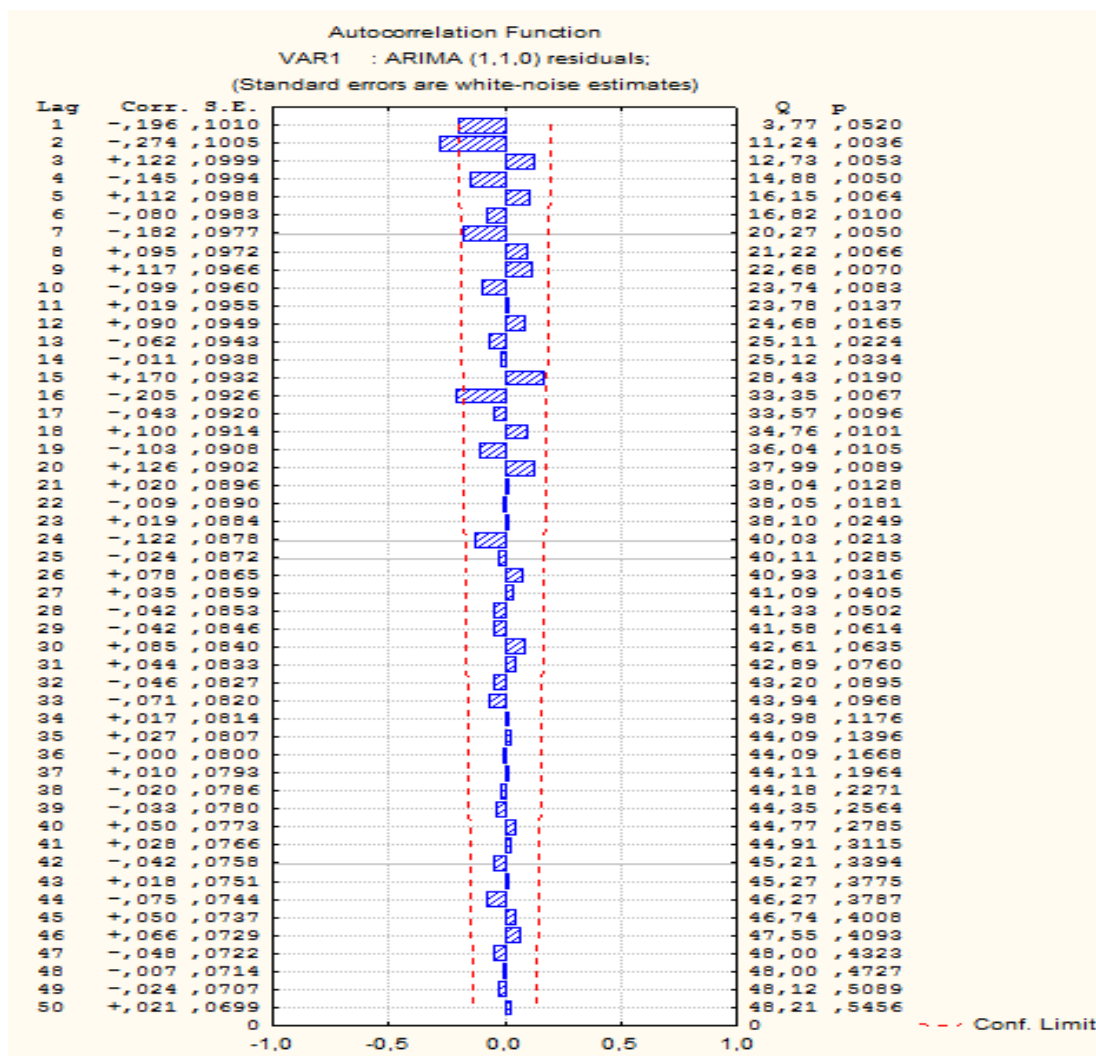


Figura II.17. Independencia de función de Autocorrelación de residuos.

Como muestra la tabla II.8 la mortalidad predicha para los próximo dos meses es de 9.82% y 9.83 %. Es importante destacar que si la mortalidad alcanza los

niveles de 12.52 o 12.82 se pueden considerar como una alarma en la cual hay que tomar medidas para evitar mayores pérdidas.

Tabla II.8 Predicción de mortalidad en crías de cerdo en los próximos dos meses

	Predicción	Límite inferior (90%)	Límite superior (90%)
97	9.82%	7.70%	12.52%
98	9.83%	7.49%	12.82%

Análisis de la serie recortada

Como la serie al inicio del período tuvo gran variabilidad con puntos aberrantes en el indicador analizado se obtuvo una nueva serie, minimizando así la variabilidad de los datos, la cual no considerara los meses que pertenecen a los tres primeros años, por lo que la serie se formó desde el 2013 – 2017.

La Figura II.18 muestra dicha serie en la cual la misma presenta tendencia negativa en los primeros tres años, a partir del año 2015 se hace positiva, es decir se observa una disminución de la mortalidad de crías de cerdo por meses y luego fueron aumentando, pero se mantuvieron en los valores promedios.

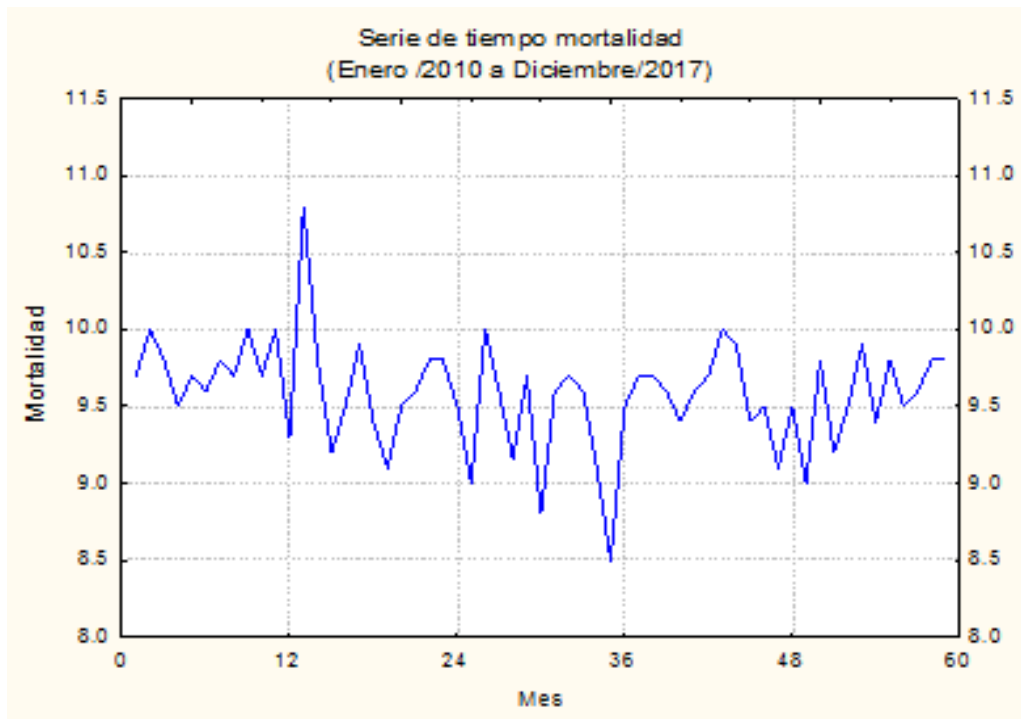


Figura II.18 Mortalidad de la serie recortada en crías de cerdo mensual en el período 2013-2017.

APLICACIÓN DE FILTROS

La Figura II.19 muestra la aplicación de las transformación de logaritmo natural $\ln(x)$ y Diferenciación $D(-1)$ de la serie de mortalidad recortada . En la cual se observa luego de aplicado los filtros que la serie se modificaron y se aprecia que ya no hay tendencia ni variación.

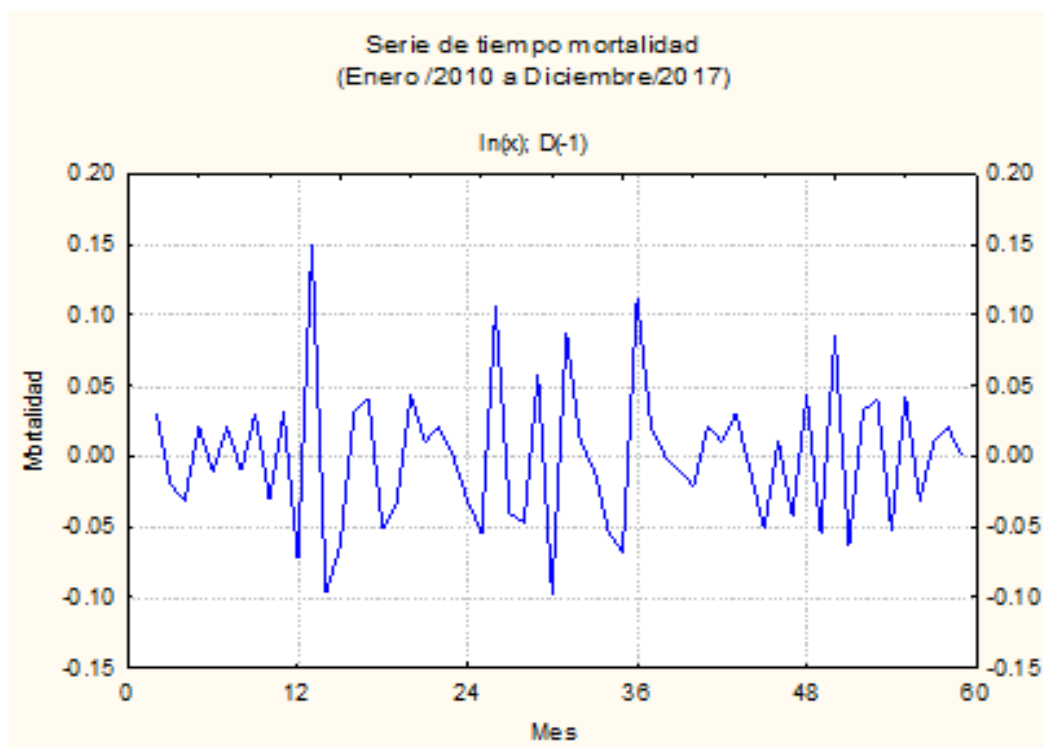


Figura II.19. Transformación de logaritmo natural $\ln(x)$ y Diferenciación $D(-1)$ de la serie de mortalidad recortada.

En la tabla II.9 se observan las estadísticas descriptivas de la serie original y de los respectivos filtros que se utilizaron para que la serie quedara estacionarizada en media y en varianza.

Tabla II.9. Estadísticas descriptivas de la serie y los filtros

	Media	Desviación. Estándar	Mínimo	Máximo	N
Var	9.586	0.348	8.5	10.8	60
Var: $\ln(x)$: $D(-1)$	0.00000	0.5404	-0.0973	0.1495	59

En la figura II. 20 se presenta la Función de Autocorrelación (FAC), para determinar el posible valor de la parte autorregresiva del modelo (p), en este caso obtuvo el valor 1 y para la figura II. 21 la Función de Autocorrelación (FACP) para conocer el valor de (q), que el mismo fue de 4. La serie al no presentar un patrón se determinó que no muestra periodicidad.

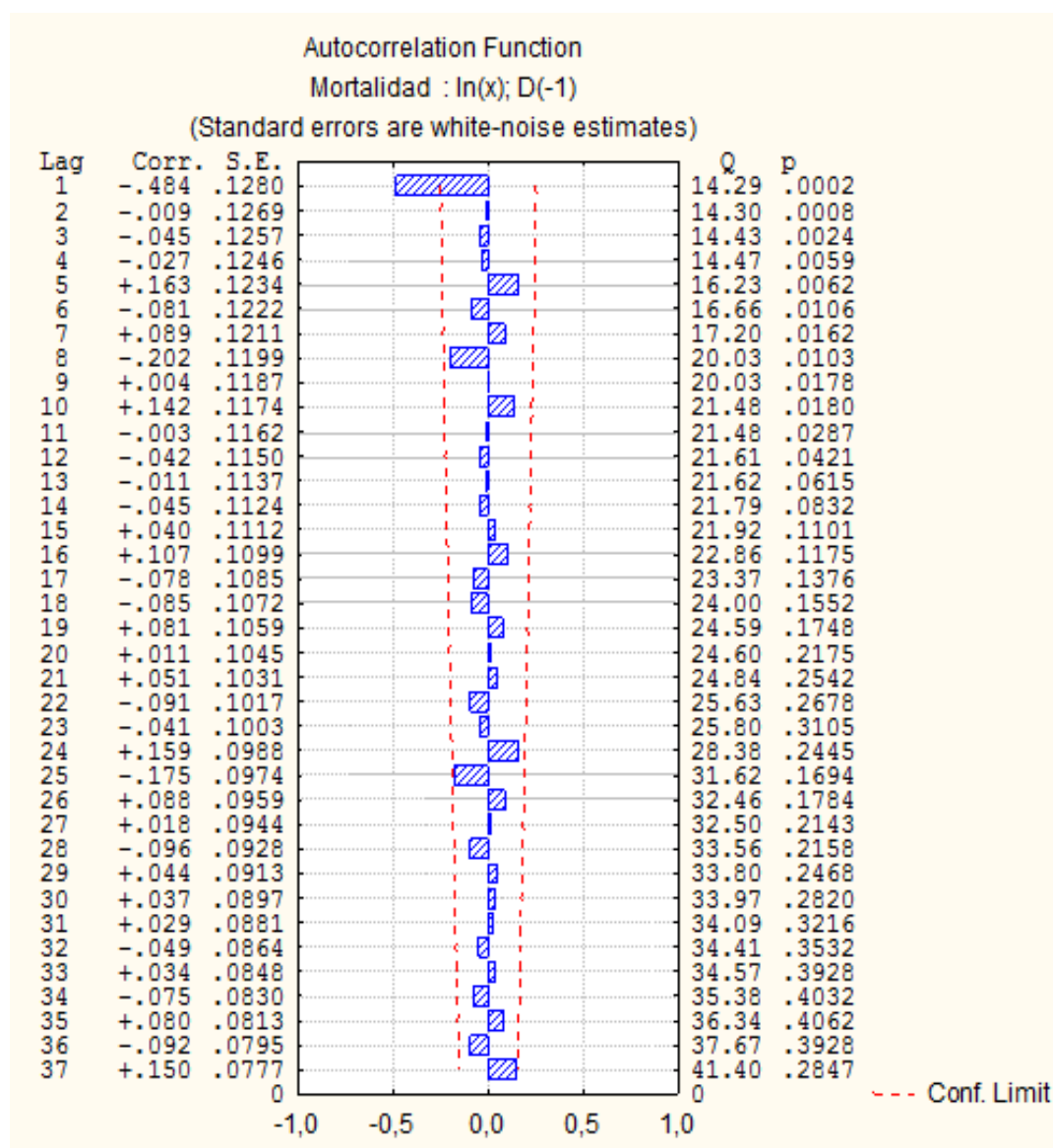


Figura IV.20. Autocorrelograma de la Función de Autocorrelación (FAC) de la serie mortalidad recortada.

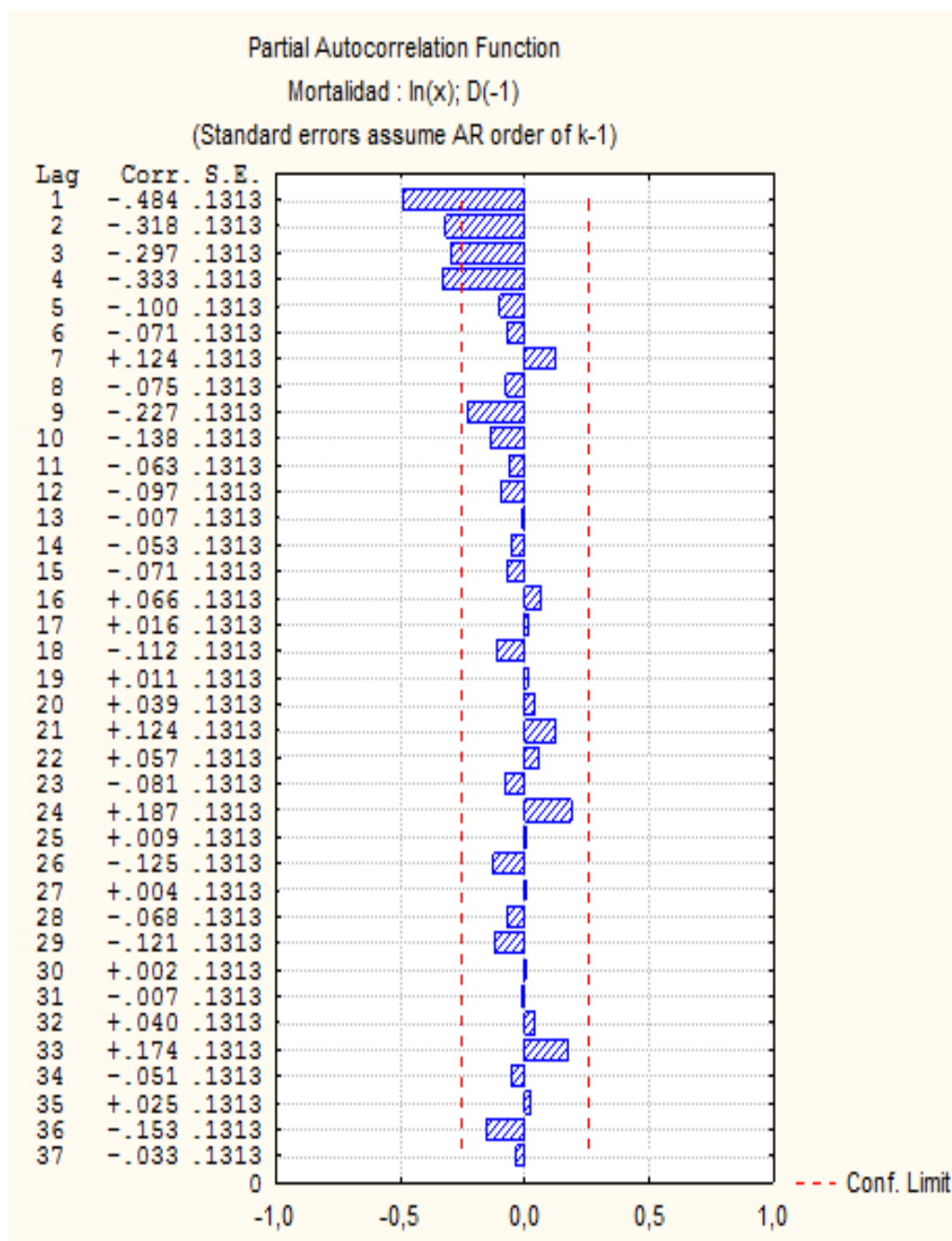


Figura IV.21. Autocorrelograma de la Función de Autocorrelación (FACP) de la serie mortalidad recortada.

Una vez analizado los supuestos se presenta de manera resumen la validación del modelo obtenido por el correlograma de la serie mortalidad en crías de cerdo:

$$ARIMA(p, d, q)s = ARIMA(4, 1, 1)_1$$

Modelo:

$$Z_t = -0.0001 + Ut + 0.5551Z_{t-12} + 0.99950U_{t-12}$$

Al correr el modelo mediante la observación de los correlogramas no fue bueno dado que no cumple los supuestos de ser:

Admisible: Es admisible dado que $p=0,5551$ es menor a 1.

Parsimonioso: No es parsimonioso ya que el intervalo va desde -0,407892 a 0,248090, en el cual incluye al 0.

Estable: el modelo no es bueno, ya que los parámetros se encuentran altamente correlacionados ($\rho = 0.861582$).

Por lo anterior se propone el siguiente modelo ARIMA, resultado de verificar el ajuste de varios modelos siendo este con el que obtuvo mejor resultado.

Modelo autorregresivo propuesto:

$$ARIMA(p, d, q)s = ARIMA(1, 1, 0)_1$$

Modelo:

$$Z_t = -0.0004 + Ut + 0.4856Z_{t-1}$$

El modelo resultante fue elegido ya que cumplió ser:

Admisible: Es admisible dado que $p = 0,4856$ es menor a 1.

Parsimonioso: Es parsimonioso dado que la estimación puntual y el intervalo con un 95% de confianza del parámetro no incluye al cero: -0,17195542 a 0,251702.

Estable: Los parámetros no se encuentran correlacionados dado que solo existe 1 parámetro en el modelo.

El valor de **MS= 0,0019**. Al ser el único modelo que se ajustó no fue necesario compararlo.

Residuos:

La Figura II.22 muestra la distribución de los residuales con respecto al modelo, el cual si cumplen con el supuesto de normalidad.

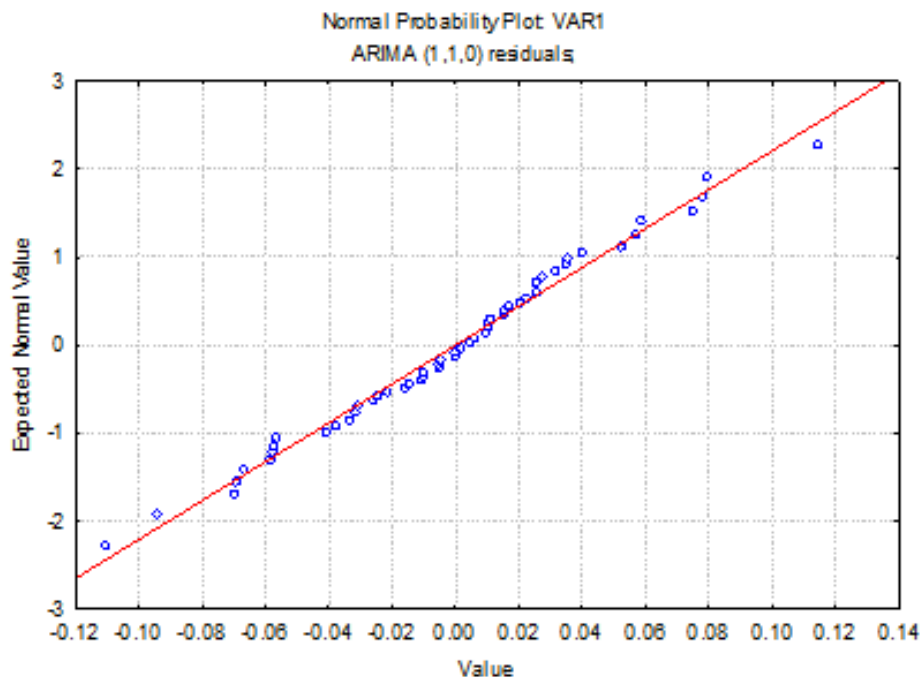


Figura II.22. Distribución normal de residuos.

La Figura II.23 muestra la autocorrelación de los residuos para ver la independencia, en este caso se graficaron 50 barras de las cuales se salen dos y la regla permite que se salgan de los límites a lo más el 5% de las barras graficadas, por lo tanto en este caso si se cumple el supuesto.

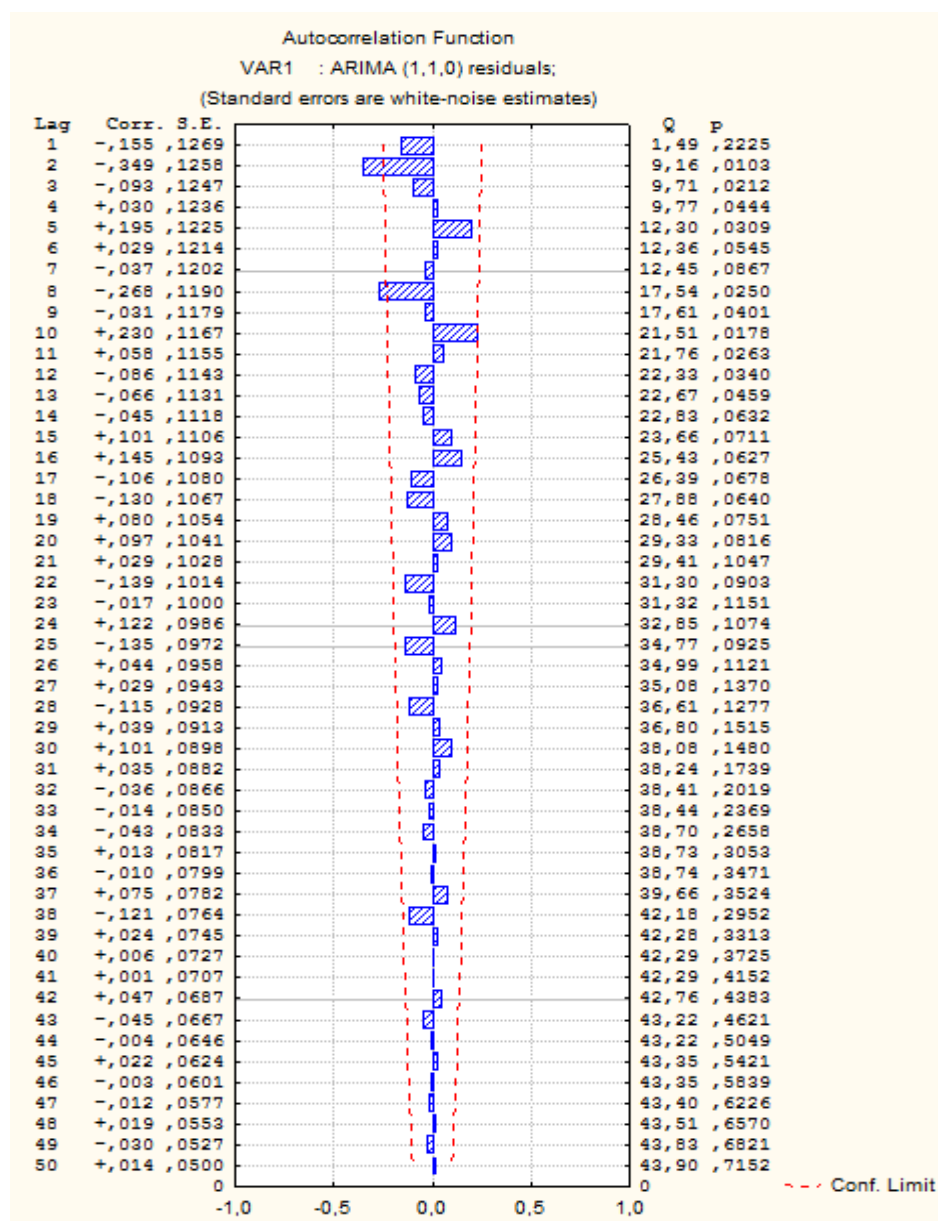


Figura II.23. Independencia de función de auto correlación de residuos.

Como muestra la (Tabla II.10) con un 90 % de confianza la mortalidad predicha para los meses enero y febrero del 2018 deben ser de 9.8, en este caso el límite superior alcanzó valores de 10.55 y 10.65, los mismos se tuvieron en consideración a la hora de declarar el valor de alarma.

En la tabla II.10 .Predicción de la serie de mortalidad modificada dos meses después de la última observación.

N de año	Predicción	Nivel inferior (90%)	Nivel superior (90%)
61	9.8	9.09	10.55
62	9.8	9.01	10.65

La tabla II.11 muestra los valores predichos del límite superior obtenido en la serie original (2010-2017) y la serie recortada (2013-2017), los cuales al ser promediados se obtuvo un valor único como indicador de alarma de 11.6 % de mortalidad en crías.

Tabla II. 11. Promedio de los límites superiores pronosticada en la serie original y la serie recortada de mortalidad en crías de cerdo.

	Límite superior de la serie original	Límite superior de la serie recortada	Promedio en ambas
1	12.52%	10.55 %	11.6
2	12.82%	10.65%	

Un análisis general del ciclo en estudio con un total de 96 observaciones del indicador mortalidad en crías de cerdo en la serie original y 60 en caso de la mortalidad en crías de la serie recortada se pudo estimar que la mediana general mortalidad del limite superior fue **11.6**, este valor es un indicador de detección temprana de alerta, es decir en este primer estudio se estima que bajo

condiciones de explotación comercial cuando la mortalidad alcanza 11.6 % en un mes , debe constituir una señal de alarma y activar acciones de investigación en la granja involucrada.

No obstante, en Cuba no se cuenta con estudios similares en mortalidad de crías de cerdo que permitan conocer cuán buenas son los resultados obtenidos en cuanto la alerta. Pero investigaciones asociadas a la vigilancia sindrómica en aves, utilizan procedimientos similares basados en series temporales, como muestran los resultados alcanzados por Beltran-Alcrudo (2009) quien propone sistemas de alertas tempranas para detectar circulación de cepas de influenza aviar baja patogenicidad en granjas ponedoras bajo condiciones comerciales de explotación, en la cual la aplicación de estos métodos fue eficiente.

De igual forma estos procedimientos han sido utilizados en la salud pública por *Guanche* y col, (2005) con resultados positivos en estudios de morbilidad y letalidad por asma bronquial en atención secundaria.

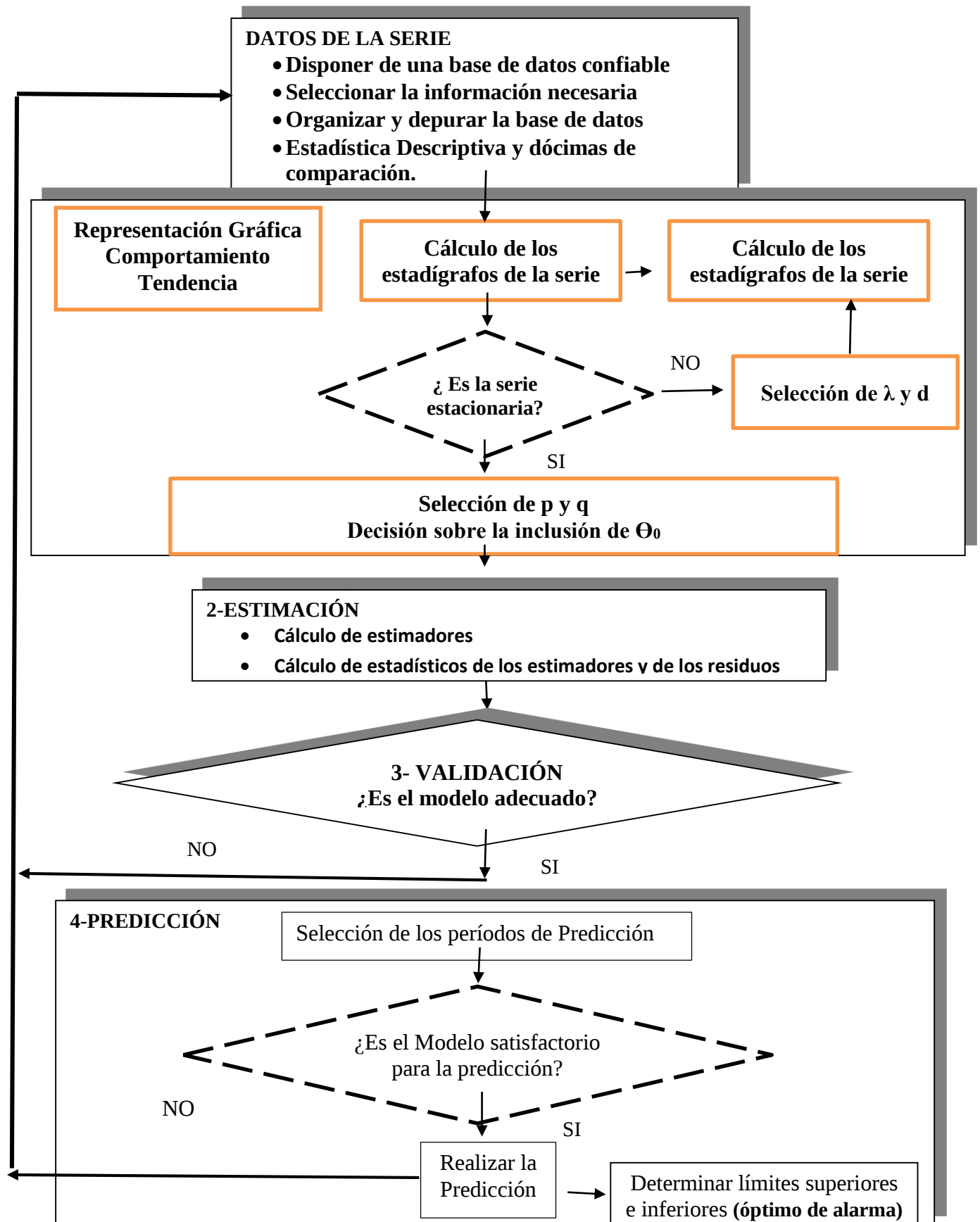
Por todo lo anterior expuesto se ultima, que el empleo de series temporales con modelos autorregresivos ARIMA constituye una vía eficiente para determinar el momento óptimo para activar una alarma en los sistemas de vigilancia sindrómica, estos resultados propician la proyección de una propuesta de acciones metodológicas para realizar sistemas de alerta que contribuyan a disminuir los riesgos para la salud animal y alcanzar los indicadores de crianza porcina con eficacia.

PROPUESTAS DE ACCIONES METODOLÓGICAS PARA INVESTIGACIONES RELACIONADAS CON VIGILANCIA SINDRÓMICA Y SISTEMAS DE ALERTAS CRÍAS DE CERDOS.

La propuesta de acciones metodológicas relacionada con la aplicación de herramientas estadísticas y modelos ARIMAS se resume en el esquema II.1, entre los aspectos que se proponen sobresalen con mayor interés los siguientes:

1. Estadística descriptiva de las variables en estudio (Número de nacimiento y mortalidad)
2. Representación gráfica de las variables en el período analizado y análisis de comportamiento y tendencia
3. Pruebas de comparación teniendo en cuenta factor mes y años. .
4. Aplicación de la Metodología de (Box y Jenkins, 1978) que incluye desde Identificación, Estimación de Parámetros, Verificación del modelo hasta las predicciones.
5. Selección de subciclos en caso necesario
6. Determinar el límite superior en caso de mortalidad y límite inferior en la variable número de nacimiento como momentos de alarma en los sistemas de producción.

Se resume en el esquema el procedimiento general utilizado en esta tesis, el cual es aplicable para investigaciones relacionadas con vigilancia sindrómica y sistemas de alertas en crías de cerdos.



CONCLUSIONES

- Los modelos $ARIMA(1,1,0)_1$ para la series de nacimientos y mortalidad en crias porcinas, resultaron efectivos para detectar tendencias de valores aberrantes, consistentes con señales de alarma ante descensos de los nacimientos por debajo de **106 crias** o el aumento de la mortalidad sobre **11.6%** en un mes, bajo las condiciones en estudio.
- Los modelos autoregresivos (ARIMA), en datos asociados a los indicadores número de nacimientos y mortalidad en la primera etapa de la crianza porcina especializada son efectivos para detectar alarma en los sistemas de vigilancia sindrómica.

RECOMENDACIONES

1. Comunicar a la empresa a la cual pertenecen la granja, el potencial de mitigación de la magnitud de las pérdidas mediante la implementación de sistemas de vigilancia sindrómica.
2. Extender el estudio a otras granjas y regiones del país, con vistas a evaluar regularidad y factibilidad de generalización como sistema de vigilancia sindrómica.

REFERENCIAS BIBLIOGRÁFICAS

1. ACPA. (Asociación Cubana de Producción Animal. (2008). Enfermedades entéricas contagiosas del cerdo en Cuba. Extracto de la presentación realizada en el Encuentro Porcicultura. Instituto de Medicina Veterinaria.
2. Alba, A., Dorea, F. C., Arinero, I., Sánchez, J., Cordon, R., Puig, P. y Revie, C. W. (2015). Exploring the surveillance potential of mortality data: nine years of bovine fallen stock data collected in Catalonia PLoS, One, 10. (Spain).
3. Beltrán-Alcrudo (2009). A flock-tailored early warning system for low pathogenic avian influenza(LPAI) in commercial egg laying flocks.[En línea] Disponible en: <www.elsevier.com/locate/prevetmed>(Consultado: diciembre 2017).
4. Box, G. y Jenkins, G. (1976). Times Séries Analysis, forecasting and control. Holden-Day Series in Time Series Analysis. G. M. Jenkins and E. Parzen, Editors. San Francisco. 553p.
5. Coutin, M. G. (2007). Las Series temporales. Unidad de Análisis y Tendencias en Salud. MINSAP. La Habana. 33 págs.
6. Cuba MINAG. (2016). Manual de Procedimientos Técnicos Para la crianza porcina. Ministerio de la Agricultura. Instituto de Investigaciones Porcinas (IIP). La Habana. 118 págs.
7. Del Valle, J. Guerra, C. W. (2012). La multicolinealidad en modelos de Regresión Lineal Múltiple. Revista Ciencias Técnica Agropecuarias, Cuba. ISSN – 2071 – 0054, RNPS – 2177.
8. Delgado, A. (2003). Comportamiento reproductivo en cerdas Yorkshire x Landrace; influencias del número de partos y la época del año sobre algunos índices. Anuario Universidad de Camagüey, Cuba.
9. Dórea, F.C; Muckle, C.A.;Kelton, D.;McClure, J.;McEwen, B.J.;McNab, W.B. (2013). Exploratory analysis of methods for automated classification of Laboratory Test Orders into Syndromic Groups in Veterinary Medicine. PLoS ONE 8(3): e57334. [En línea] Disponible en:<www.elsevier.com/locate/prevetmed> (Consultado: enero 2018).
10. Duncan, D. B. 1955. Multiple range and multiple F tests. Biometrics:11:1

11. Espallargas, D.S. y Solis, M. (2012). *Econometría y Series temporales. Aplicaciones*. La Habana .Editorial Félix Varela, 354 págs.
12. FAO, (2002). *Agricultura mundial: hacia los años 2015/2030*. Food and Agriculture Organization (FAO). Roma, versión electrónica disponible en el sitio: <http://www.fao.org/docrep/fao/004/y3557S>.
13. Fernández L, Tonhati, H; Albuquerque, L.G; Aspilcueta-Borquis, R.R y Menéndez Buxadera, A. (2011). Modelos de regresiones aleatorias para la estimación de parámetros genéticos y estudios de curvas de lactancia del Holstein en Cuba. *Revista Cubana de Ciencia Agrícola*, Tomo 45, Número 1.
14. Fernández-, L; Lara A.M; Pereyra, A.M; Guerra, CW; De Calzadilla, J. (2013). Estadística Aplicada a la Ingeniería Agrícola y a las Ciencias Agropecuarias. Su contribución en la docencia, investigación y transferencia de conocimiento. *Revista Ciencias Técnicas Agropecuarias*, 22(2).
15. Fernández, Lucia; Guerra, C.W; De Calzadilla, J; Lim. N.U. (2017). Desarrollo de la Modelación Estadístico–Matemática en las Ciencias Agrarias. Retos y perspectivas. *Revista Investigación Operacional*. VOL 38. , No.5, 462-467.
16. German R.R., Lee L.M, Horan J.M, Milstein R.L, Pertowski C.A. y Waller M.N. (2001). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep*, 50:1-35.
17. Guanche, H; García E; Gutiérrez F. (2005). Morbilidad y letalidad hospitalaria por asma bronquial. *Rev Cubana Med [serie en Internet]*. [citado 2 Oct 2007];44(3-4). Disponible en: http://bvs.sld.cu/revistas/med/vol44_3-4_05/med033-405.htm
18. Guerra, C.W; Menéndez, E; Barrera, R. y Egaña, E. (2010). *Estadística*. La Habana. Editorial Félix Varela 2da Ed. ISBN 959-258-421-4.
19. Hasler, B., Howe, K. S. y Stark, K. D. (2011). Conceptualising the technical relationship of animal disease surveillance to intervention and mitigation as a basis for economic analysis. *BMC Health Serv Res*, 11, 225 págs.

20. Herrera, M. (2013) .Métodos Estadísticos alternativos de análisis con variables discretas y categóricas en investigaciones agropecuarias. Tesis presentada en opción al grado de Doctor en Ciencias. Instituto de Ciencia Animal.
21. Hoinville, L. J.; Alban, L; Drewe, J. A; Gibbens, J. C; Gustafson, L; Hasler, B; Saegerman, C; Salman, M. y Stark, K. D. (2013). Proposed terms and concepts for describing and evaluating animal-health surveillance systems. *Prev Vet Med*, 112, 1-12 págs.
22. Jackson, M.L, Baer, A., Painter, I. y Duchin, J. A. (2007) Simulation study comparing aberration detection algorithms for syndromic surveillance. *BMC Medical Informatics and Decision Making*.7:6:10.1186/1472-6947-7-6.
23. Katz, R., May, L., Baker, J. y Test, E. (2011). Redefining syndromic surveillance. *J Epidemiol Glob Health*, 1, 21-31.
24. Lawson, A. B. (2013). Bayesian disease mapping: hierarchical modeling in spatial epidemiology, CRC press.p 25-35
25. Lazo, P. L. y Gutiérrez, C. Y. (2011). Estudio de factores de riesgo vinculados a la mortalidad neonatal en una granja porcina. Disponible en: <http://www.engormix.com/mailreddirect/notif.asp?nkid=85>.
26. Lombardo, J. S., Burkom, H. y Pavlin, J. (2004). ESSENCE II and the framework for evaluating syndromic surveillance systems. *MMWR Morb Mortal Wkly Rep*, 53: 159-65.
27. Macnab, Y. C. (2011). On Gaussian Markov random fields and Bayesian disease mapping. *Stat Methods Med Res*, 20, 49-68.
28. Martin, S. (1996). Bases fisiológicas en el manejo de las hembras reproductoras. En *Reproducción del Cerdo*. División del Sistema de Universidad Abierta. FMVZ-UNAM. México
29. Martínez, R. (1998) Principales factores que afectan la reproducción en el cerdo. CIENCIA VETERINARIA. Departamento de Producción Animal: Cerdos Facultad de Medicina Veterinaria y Zootecnia. Universidad Nacional Autónoma de México Ciudad Universitaria, 04510, México, D.F.
30. Microsoft® Office, (2010). Proofing Tools© 2010 Microsoft Corporation. All rights reserved.

31. Morante, F. (2018). La estadística en las investigaciones científicas. Disponible en: <https://www.masscience.com/2016/03/27/la-estadistica-en-las-investigaciones-cientificas>.
32. Peña, D. (2005). Análisis de series temporales. Madrid. Editorial Alianza. pp.245.
33. Pires, S.R., Mederos, C.M., Diéguez, F.J. y Sosa, R.(2007). Producción porcina a pequeña y mediana escala. Grupo de Investigación de Producción Porcina. La Habana, versión electrónica disponible en disco compacto.
34. Quiles, S.A; Méndez H.L. (2004). Producción porcina intensiva. Editorial: S.A. Agrícola española,158 págs.
35. Ramírez F., Ordóñez, R. and Iriarte, J.M. (2005). Vigilancia en salud pública: más allá de las enfermedades transmisibles. *Gac. Sanit.* 5:19:1813 págs.
36. Ramírez Juan Ruiz, Hernández-Rodríguez Gabriela Eréndira y Zulueta Rodríguez Ramón, Concepción. Modelos Box-Jenkins: Time Series Analysis Forecasting in SugarCane Production (2011), Facultad de Ciencias Agrícolas. Universidad Veracruzana. pp103-109
37. Rodríguez, L. (2015) Modelación y simulación de la producción de biomasa de *Pennisetum Purpureum* Schum vc. King grass y su aplicación en la alimentación. Tesis presentada en opción al grado de Doctor en Ciencias. Instituto de Ciencia Animal. 100 pág.
38. StatgraphicsPlus Centurion versión. XV.II (2015).
39. StatSoft. INC (2004). STATISTICA. Data analysis software system. Version 7. www.statsoft.com.
40. Vallejo, R. (2012). Introducción a las Series Cronológicas. Modelos ingenuos. Universidad Técnica Federico Santa María. España.
41. Van Metre, D. C., Barkey, D. Q., Salman, M. D. y Morley, P. S. (2009). Development of a syndromic surveillance system for detection of disease among livestock entering an auction market. *J Am Vet Med Assoc*, 234, 658-64.

42. Vázquez, Y; Guerra, C.W; Sánchez, O.E. (2014) .Modelación estadístico-computacional para el estudio de la sostenibilidad socioeconómica de la empresa pecuaria valle del Perú. Revista Investigación Operacional. 35(2): 121-129.
43. Vourc'h, G; Bridges, V. E; Gibbens, J; De Groot, B. D., McIntyre, L; Poland, R. y Barnouin, J. (2006). Detecting Emerging Diseases in Farm Animals through Clinical Observations. Emerging Infectious Diseases, 12: 204-210.
44. Ward, M. P. y Carpenter, T. E. (2000). Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. Prev Vet Med, 45, 257-84.