



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Estudios de Postgrado  
Maestría en Artes en Estadística Aplicada

**REGRESIÓN LINEAL APLICADA A LA ESTIMACIÓN DEL PESO DE  
LECHONES (*Sus scrofa domesticus*) EN UNA GRANJA PORCINA**

**LIC. ERICK ARTURO DE LA CRUZ NARCISO**

Asesorado por el MBA. Lic. José Rodrigo Trejo Roldán

Guatemala, junio de 2022



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**REGRESIÓN LINEAL APLICADA A LA ESTIMACIÓN DEL PESO DE  
LECHONES (*Sus scrofa domesticus*) EN UNA GRANJA PORCINA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**LIC. ERICK ARTURO DE LA CRUZ NARCISO**

ASESORADO POR EL MBA. LIC. JOSÉ RODRIGO TREJO ROLDÁN

AL CONFERÍRSELE EL TÍTULO DE

**MAESTRO EN ESTADÍSTICA APLICADA**

GUATEMALA, JUNIO DE 2022



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Mtro. Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL DE DEFENSA DE TESIS**

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Mtro. Ing. Edgar Darío Alvarez Coti
EXAMINADOR	Mtro. Ing. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtro. Ing. William Eduardo Fagiani Cruz
SECRETARIO	Mtro. Hugo Humberto Rivera Pérez



## HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

### **REGRESIÓN LINEAL APLICADA A LA ESTIMACIÓN DEL PESO DE LECHONES (*Sus scrofa domesticus*) EN UNA GRANJA PORCINA**

Tema que me fuera asignado por la Dirección de Escuela de Estudios de Postgrado con fecha 12 de agosto de 2021.

A handwritten signature in black ink, appearing to read 'Erick', with a large, stylized flourish extending to the right.

**Lic. Erick Arturo de la Cruz Narciso**



LNG.DECANATO.OI.468.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **REGRESIÓN LINEAL APLICADA A LA ESTIMACIÓN DEL PESO DE LECHONES (*Sus scrofa domesticus*) EN UNA GRANJA PORCINA**, presentado por: **Erick Arturo De La Cruz Narciso**, que pertenece al programa de Maestría en artes en Estadística aplicada después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Inga. Aurelia Anabela Cordova Estrada ★

Decana

Guatemala, junio de 2022

AACE/gaoc





**Guatemala, junio de 2022**

LNG.EEP.OI.468.2022

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

**"REGRESIÓN LINEAL APLICADA A LA ESTIMACIÓN DEL PESO DE LECHONES  
(Sus scrofa domesticus) EN UNA GRANJA PORCINA"**

presentado por **Erick Arturo De La Cruz Narciso** correspondiente al programa de **Maestría en artes en Estadística aplicada**; apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"

**Mtro. Ing. Edgar Darío Álvarez Cotí**  
Director

**Escuela de Estudios de Postgrado**  
**Facultad de Ingeniería**







Guatemala 29 de marzo 2022.

**M.A. Edgar Darío Álvarez Cotí**  
Director  
Escuela de Estudios de Postgrado  
Presente

**M.A. Ingeniero Álvarez Cotí:**

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado **"REGRESIÓN LINEAL APLICADA A LA ESTIMACIÓN DEL PESO DE LECHONES (*Sus scrofa domesticus*) EN UNA GRANJA PORCINA"** del estudiante **Erick Arturo De La Cruz Narciso** quien se identifica con número de carné 200512415 del programa de Maestría en Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,

**MSc. Ing. Edwin Adalberto Bracamonte Orozco**  
Coordinador  
Maestría en Estadística Aplicada  
Escuela de Estudios de Postgrado



Guatemala, noviembre de 2021

M.A. Ing. Edgar Darío Álvarez Cotí

Director

Escuela de Estudios de Postgrado

Presente

Estimado M.A. Ing. Álvarez Cotí

Por este medio informo a usted, que he revisado y aprobado el Trabajo de Graduación y el Artículo Científico: "REGRESIÓN LINEAL APLICADA A LA ESTIMACIÓN DEL PESO DE LECHONES (*Sus scrofa domesticus*) EN UNA GRANJA PORCINA" del estudiante Erick Arturo de la Cruz Narciso del programa de Maestría en Estadística Aplicada, identificada con número de carné: 999003139.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.



Lic. José Trejo Roldán

Colegiado 1763

---

Magister José Rodrigo Trejo Roldán

Colegiado No. 1763

Asesor de Tesis



## **ACTO QUE DEDICO A:**

**Dios**

Por permitirme la vida.

**Mi madre**

Erica Narciso. Por su amor y apoyo incondicional.

**Mi padre**

Pablo de la Cruz (q. d. e. p.). Por el orgullo que hubiese sentido de verme alcanzar esta meta.



## **AGRADECIMIENTOS A:**

- Facultad de Ingeniería** Por permitirme formarme profesionalmente.
- Mi asesor** Lic. Rodrigo Trejo. Por tomarse el tiempo para asesorarme y compartirme muchos de sus conocimientos.
- Mi novia** Andrea Paz. Por motivarme constantemente y estar presente en momentos buenos y malos.
- Carlos Ríos y Claudio Melini** Por su motivación, compañerismo y amistad.



## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES .....	III
LISTA DE SÍMBOLOS .....	V
GLOSARIO .....	VII
RESUMEN .....	IX
PLANTEAMIENTO DEL PROBLEMA.....	XI
OBJETIVOS.....	XV
RESUMEN DEL MARCO METODOLÓGICO .....	XVII
INTRODUCCIÓN .....	XXIII
1. MARCO REFERENCIAL.....	1
2. MARCO TEÓRICO.....	7
2.1. Fundamentos de la regresión lineal general.....	7
2.1.1. Análisis de correlación .....	7
2.1.2. Coeficiente de determinación $R^2$ .....	8
2.1.3. Coeficiente de determinación $R^2$ ajustado .....	9
2.1.4. Gráfico de cajas y bigotes.....	10
2.1.5. Verificación de supuestos .....	11
2.1.5.1. Normalidad .....	11
2.1.5.2. Homocedasticidad .....	14
2.1.5.3. Independencia .....	14
2.1.5.4. Colinealidad .....	16
2.2. Modelo lineal general .....	17
2.3. Análisis de varianza.....	19
2.4. Desarrollo fisiológico de los lechones.....	20

3.	PRESENTACIÓN DE RESULTADOS.....	21
3.1.	Modelo de regresión lineal .....	21
3.1.1.	Objetivo 1: Hallar el nivel de correlación entre el peso al nacimiento, el peso al destete y el peso de traslado a sitio 3 mediante modelos de regresión lineal.....	21
3.2.	Estimación del peso de lechones ( <i>Sus scrofa domesticus</i> ) al momento de su traslado a sitio 3 mediante un modelo de regresión lineal.....	22
3.2.1.	Objetivo 2: Evaluar el mejor modelo por medio del criterio de información de Akaike, el criterio de información Bayesiano y el R <sup>2</sup> ajustado. ....	25
3.2.2.	Objetivo 3: Comprobar la normalidad, homocedasticidad, independencia y falta de multicolinealidad del modelo. ....	28
3.2.2.1.	Normalidad .....	28
3.2.2.2.	Homocedasticidad .....	29
3.2.2.3.	Independencia.....	30
4.	DISCUSIÓN DE RESULTADOS.....	33
4.1.	Análisis interno.....	33
4.2.	Análisis externo.....	34
	CONCLUSIONES.....	37
	RECOMENDACIONES .....	39
	REFERENCIAS .....	41
	APÉNDICES.....	47

## ÍNDICE DE ILUSTRACIONES

### FIGURAS

1.	Diagrama de cajas y bigotes ( <i>Box-Plot</i> ) .....	11
2.	Gráfico de QQ-Plot para el modelo elegido.....	29
3.	Gráfico de residuos vs predichos .....	30
4.	Correlograma para el modelo de regresión lineal .....	31

### TABLAS

I.	Variables del estudio .....	XVIII
II.	Criterio de modelización de variables regresoras .....	XX
III.	Correlación de Pearson para el modelo .....	22
IV.	Resumen de resultados de coeficientes de regresión.....	23
V.	Resultados del ANOVA.....	24
VI.	Tabla de resultados de los criterios de información utilizados .....	25
VII.	Presentación de resultados para el modelo de regresión lineal M11 ....	27
VIII.	Presentación de resultados para el modelo de regresión lineal M12 ....	27
IX.	Presentación de resultados para el modelo de regresión lineal M15 ....	27



## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
%	Porcentaje



## GLOSARIO

<b>Aleatoriedad</b>	Es una característica que se asocia a una variable cuando su resultado no es previsible, únicamente al azar.
<b>Coefficiente de correlación</b>	Es la medida estadística de dependencia lineal que tienen al menos dos variables aleatorias.
<b>Coefficiente de determinación</b>	Es la medida estadística utilizada para la medición de futuros resultados en un modelo de predicción.
<b>Criterio Bayesiano</b>	Desarrollado por el ingeniero israelí Gideon Schwarz es un criterio para selección de modelos, mediante la función de probabilidad relacionando ajustes por adición de parámetros.
<b>Criterio de Akaike</b>	Desarrollado por el matemático japonés Hirotugu Akaike es un criterio de selección de modelos, mediante la medición de la calidad relativa en un conjunto de datos.
<b>Modelo de regresión lineal</b>	Modelo matemático utilizado para aproximar la relación de dependencia entre una variable de respuesta y sus regresoras.

<b>Normalidad</b>	Criterio estadístico para asumir que una población se comporta de forma de una distribución caracterizada por la simetría alrededor de la media, la cual coincide con la mediana formando una campana de Gauss.
<b>R</b>	Software de análisis estadístico utilizado con la interfaz <i>RStudio</i> .
<b>Variable dependiente</b>	Es la variable cuantitativa que es influida por los valores y fluctuaciones de las variables independientes que la afectan.
<b>Variable independiente</b>	Es la variable cuantitativa o estadísticamente como regresora, y representan las razones potenciales de variación sobre la variable dependiente.

## RESUMEN

El propósito de la presente investigación fue aportar una herramienta estadística para ayudar a comprender la relación entre edad y peso en lechones.

El objetivo general fue estimar el peso de lechones al momento de su traslado a sitio 3 mediante un modelo de regresión lineal, las variables independientes contempladas fueron el peso al nacimiento, el peso al destete y la edad al destete, seleccionando aquellas con las que se logró un mejor ajuste de los datos.

La investigación empleó un enfoque cuantitativo, con un diseño no experimental y alcance correlacional. Para la identificación de cada cerdo se utilizó un código impreso en un arete plástico colocado en su oreja izquierda, se registró para cada uno de ellos la fecha y el peso de nacimiento, destete y traslado a sitio 3, posteriormente se tabularon los datos, luego se hizo la modelización, y el cálculo del  $R^2$  ajustado, y de los criterios de información de Akaike y Bayesiano, luego se seleccionó el mejor modelo y se comprobaron los supuestos de normalidad, homocedasticidad e independencia.

El principal resultado mostró que los modelos de regresión lineal son una buena opción para generar estimaciones de pesos en lechones.

Con lo evaluado se logró concluir que el mejor modelo fue el de regresión lineal múltiple con el peso al nacimiento, peso al destete y edad al destete como variables independientes dado que fue el que presentó la mejor bondad de ajuste,

con el  $R^2$  ajustado más alto (0.43) y los criterios de información de Akaike y Bayesiano más bajos.

Se sugirió darles seguimiento a los cerdos del estudio para generar un modelo de regresión para estimar el peso a las 22 semanas de vida, esto puede ayudar de manera importante a la planificación de la fase final del desarrollo de los cerdos, así como a planificar mejor la venta.

## PLANTEAMIENTO DEL PROBLEMA

El cerdo doméstico tiene un papel importante en la nutrición mundial debido a su alto valor proteico y forma parte de la cultura gastronómica de Guatemala; esto ha permitido que la industria porcina en el país sea fuerte, sin embargo el creciente costo de las materias primas tiende a reducir las utilidades, ante ello muchas empresas buscan captar gran cantidad de información de variables zootécnicas a nivel de campo y de hacer estimaciones precisas sobre el desempeño futuro de los animales a fin de mejorar la eficiencia.

Respecto a las estimaciones, toma especial importancia la concerniente al peso de los cerdos en sus distintas fases productivas, entre más tempranas se puedan realizar mejor para la toma de decisiones, entre las etapas más críticas del cerdo debido a múltiples factores tanto fisiológicos como ambientales está la llamada fase de transición, que culmina al momento de su traslado del sitio 2 al sitio 3 y uno de los indicadores más importantes sin lugar a dudas es el peso que cada cerdo tiene al momento del traslado.

Hay empresas que hacen estimaciones predictivas sobre el peso que tendrán los lechones al momento de trasladarse al sitio 3, muchas de esas estimaciones se basan en apreciaciones empíricas como la visualización de la condición corporal de los lechones en determinado momento de su desarrollo.

A nivel empírico parece haber cierta relación entre diferentes variables como el peso y la edad de los lechones, sin embargo, puede resultar confuso hacer estimaciones sin un modelo estadístico, por tanto, en esta investigación se tomaron dichas variables y se modelizaron mediante el método de regresión

lineal, así pues, se pretende hallar el modelo con mejor ajuste para la estimación del peso de lechones al momento del traslado a sitio 3.

- Contexto general

La granja donde se realizó el estudio es de tipo comercial multisitio, las granjas de este tipo son explotaciones productivas que tienen instalaciones bien diferenciadas para alojar cerdos en distintas fases, posee un área denominada sitio 1 que aloja cerdas gestantes, cerdas lactantes y a lechones aún no destetados; también tiene un área denominada sitio 2 que aloja lechones desde su destete hasta las 11 semanas de vida, y por último tiene un área denominada sitio 3 que aloja cerdos desde las 11 semanas aproximadamente hasta su venta.

El presente estudio se inició en el sitio 1 y finalizó en el sitio 2 y la población utilizada fueron los lechones nacidos entre el 18 y el 23 de julio del 2021.

Actualmente en la granja se hacen estimaciones empíricas sobre el peso que tendrán los lechones previos a ser trasladados al sitio 3, dichas estimaciones resultan altamente imprecisas y por tanto tienen poca utilidad.

- Descripción del problema

Dada la aparente relación observada mediante un análisis descriptivo entre diferentes indicadores productivos como el peso al nacer, el peso al destete, la edad al destete y el peso al momento del traslado a sitio 3, se dedujo que probablemente existía correlación entre ellos, y que de ser así dichas variables serían susceptibles de ser modeladas para realizar estimaciones.

Uno de los indicadores de mayor interés en la producción porcina es el peso de traslado a sitio 3 y su correcta estimación representaría una oportunidad de mejora en manejos, en eficiencia, y en proyección de costos de producción, sin embargo, el ritmo de trabajo acelerado y enfocado en la parte operativa conlleva que dichos análisis no se realicen.

Ante esto fue necesario desarrollar un modelo que permitiera hacer estimaciones confiables, y así tomar decisiones con fundamento estadístico.

- Formulación del problema

- Pregunta central

¿Cuál es el modelo de regresión lineal con mejor ajuste para la estimación del peso de lechones (*Sus scrofa domesticus*) al momento del traslado a sitio 3?

- Preguntas auxiliares

- ¿Cuál es el nivel de correlación entre las variables regresoras y la variable respuesta?
    - ¿Cuál es el modelo que presenta mejores criterios de información (de Akaike y Bayesiano) y mejor  $R^2$  ajustado?
    - ¿Los datos cumplen los supuestos de normalidad, homocedasticidad e independencia?

- Delimitación del problema

El problema se delimitó a lechones desde su nacimiento hasta el momento en el que son trasladados al sitio 3 en una granja comercial multisitio ubicada en el municipio de Río Bravo del departamento de Suchitepéquez, se utilizó información generada entre el 11 de julio y el 4 de agosto del 2021.

## OBJETIVOS

### General

Estimar el peso de lechones (*Sus scrofa domesticus*) al momento de su traslado a sitio 3 mediante un modelo de regresión lineal, con el peso al nacimiento, el peso al destete y la edad al destete como variables independientes.

### Específicos

- Hallar el nivel de correlación entre el peso al nacimiento, el peso al destete y el peso de traslado a sitio 3 mediante modelos de regresión lineal.
- Evaluar el mejor modelo por medio del criterio de información de Akaike, el criterio de información Bayesiano y el  $R^2$  ajustado.
- Comprobar la normalidad, homocedasticidad, independencia y falta de multicolinealidad del modelo.



## RESUMEN DEL MARCO METODOLÓGICO

La presente investigación tiene un enfoque cuantitativo, con un diseño no experimental y un alcance correlacional, las variables tienen una escala de tipo continua, cuantitativa y de razón.

- Características del estudio

El enfoque del estudio propuesto fue cuantitativo ya que se evaluó la relación numérica entre las variables mediante el diseño de un modelo de regresión lineal.

El alcance fue correlacional dado que se pretendió que los modelos generados describiesen el peso de traslado a sitio 3 en función del peso al nacimiento, peso al destete y edad al destete.

Las balanzas utilizadas siempre son calibradas anualmente, mediante el departamento de metrología.

El diseño adoptado fue no experimental, pues para diseñar el modelo de regresión lineal con mejor ajuste se utilizaron los datos tal cual se generaron en su estado original sin ninguna manipulación, se registró cada peso de los lechones nacidos entre el 18 y 23 de julio del 2021 y posteriormente se registró el peso de cada lechón en sus subsiguientes fases críticas de desarrollo, las cuáles son el destete y el traslado a sitio 3.

- Unidades de análisis

La población del estudio fueron los lechones nacidos entre el 18 y el 23 de julio del 2021, correspondientes al lote identificado con el código 21.29, dicha población se encontraba dividida en dos subpoblaciones dadas por el sexo, quedando entonces una subpoblación de machos y otra de hembras, de la cual se extrajeron las muestras de forma aleatoria simple sin reemplazo, que fueron analizadas en su totalidad.

- Variables

Tabla I. **Variabes del estudio**

<b>Variable</b>	<b>Definición teórica</b>	<b>Definición operativa</b>
Peso al nacimiento	Peso del lechón nacido vivo (LNV) tomado como máximo 12 horas después del nacimiento. Es una variable cuantitativa, continua, de razón.	Esta variable se medirá en una báscula digital, está dada en libras y su escala es de tipo continua., cuantitativa, de razón.
Peso al destete	Peso del lechón justo al momento de ser separado permanentemente de su madre y previo a ser trasladado al sitio 2. Es una variable cuantitativa, continua, de razón.	Esta variable se medirá en una báscula digital, está dada en libras y su escala es de tipo continua, cuantitativa, de razón.
Edad al destete	Edad en días al momento del destete, resulta de sumar los días transcurridos desde el día del nacimiento hasta el día en que es destetado. Es una variable cuantitativa, continua, de razón.	Esta variable está dada en días y su escala es de tipo continuo, cuantitativa, de razón.

Continuación tabla I.

Edad al traslado	Edad en días al momento del traslado al sitio 3, resulta de la suma de los días desde el nacimiento hasta el traslado al sitio 3. Es una variable cuantitativa, continua, de razón.	Esta variable está dada en días y su escala es de tipo continuo, cuantitativa, de razón.
Peso de traslado al sitio 3.	Peso del lechón inmediatamente antes de ser trasladado al sitio 3. Esta es la variable de respuesta.	Esta variable se medirá en una báscula digital, está dada en libras y su escala es de tipo continua, cuantitativa, de razón.

Fuente: elaboración propia, empleando Microsoft Excel.

- Fases del estudio

Se realizó una etapa investigativa y descriptiva, donde se estudiaron comportamientos previos. Luego se ejecutó la parte de análisis de acuerdo con las fases siguientes.

- Fase 1: Recolección de datos

Se procedió a identificar de manera individual a cada lechón nacido entre el 18 y el 23 de julio, esto se realizó mediante un arete plástico con un número correlativo, luego se anotó el peso en cada una de las fases que son el nacimiento, el destete y el traslado a sitio 3.

- Fase 2: Análisis descriptivo de datos

Se realizó una gráfica de dispersión para observar de manera preliminar la existencia de correlación entre las variables. Se realizaron gráficas de cajas y bigotes para apreciar la posible presencia de datos atípicos y se generaron gráficas de densidad para observar simetría, curtosis y normalidad.

- Fase 3: Modelización

Se procedió a diseñar los modelos de regresión lineal, se calculó el número de modelos basado en las posibles combinaciones de las variables regresoras, dando como resultado cuatro modelos de regresión lineal simple y once modelos de regresión lineal múltiple.

Se modelaron las ecuaciones con las variables regresoras, utilizando el criterio a continuación.

Tabla II. **Criterio de modelización de variables regresoras**

<b>Variable / Modelo</b>	<b>X1: Peso al nacimiento</b>	<b>X2: Peso al destete</b>	<b>X3: Edad al destete</b>	<b>X4: Edad al traslado sitio 3</b>
M1	X			
M2		X		
M3			X	
M4				X
M5	X	X		
M6	X		X	
M7	X			X
M8		X	X	
M9		X		X

Continuación tabla II.

M10			X	X
M11	X	X	X	
M12	X	X		X
M13	X		X	X
M14		X	X	X
M15	X	X	X	X

Fuente: elaboración propia, empleando Microsoft Excel.

De los modelos evaluados, se comprobaron los supuestos para cada uno de ellos (normalidad, homocedasticidad e independencia), para posteriormente elegir el mejor modelo basado en el criterio de información de Akaike (AIC), el criterio de información Bayesiano (BIC) y el  $R^2$  ajustado.

- Fase 4: Interpretación de resultados

Luego de haber generado los modelos y haber utilizado los criterios de información AIC, BIC y  $R^2$  ajustado se interpretaron los resultados obtenidos para sustentar la elección del modelo con base en el principio de parsimonia.

- Fase 5: Elaboración de informe final

Se redactó un informe final para comunicar de manera clara, precisa y efectiva los hallazgos del estudio, y con esto lograr que dicha investigación sea una fuente de consulta útil para otros investigadores.



## INTRODUCCIÓN

La presente investigación fue realizada con el propósito de aportar una herramienta estadística para ayudar a comprender la relación entre edad y peso en lechones, y dada esta comprensión generar un modelo de regresión lineal que tuviera una óptima bondad de ajuste y siguiera el principio de parsimonia, esta estimación basada en un modelo estadístico es relevante dado que representa una mejora respecto a las estimaciones empíricas que son en general poco precisas y de difícil validación, por otro lado tener una estimación fiable ayuda a tomar decisiones operativas oportunamente.

La regresión lineal es un modelo matemático que permite un análisis tanto descriptivo como predictivo, de acuerdo con Montgomery (2011) el análisis de regresión consiste en modelar e investigar la relación entre dos o más variables, esto le confiere una aplicabilidad amplia en diferentes campos del conocimiento. Entre estos campos del conocimiento se encuentra la industria en general y la porcicultura en particular, la cuál es la segunda industria más importante del país solo por detrás de la avicultura.

En la granja tecnificada estudiada, se hacen solamente estimaciones empíricas sobre el peso que tendrán los lechones. Estas estimaciones son altamente imprecisas y de poca utilidad, debido a esto se hizo necesario desarrollar una herramienta estadística que proporciona información más precisa y cuantificable. La investigación contribuyó en general a la mejora de la eficiencia, y en particular con el aporte de un criterio estadístico para la toma de decisiones tempranas como la eutanasia en lechones de bajo peso o la aplicación de

suplementos de manera oportuna, adicionalmente a esto ayudó a estimar de manera más fiable los costos de producción de cada lote de lechones.

La metodología consistió en la descripción y definición del alcance y enfoque del estudio, luego se definió la variable dependiente y las variables independientes, entonces se procedió a la recolección de datos de manera individual para cada unidad de estudio y luego se procedió a realizar la modelización y a la selección del mejor modelo basado en los criterios de información de Akaike, Bayesiano y al  $R^2$  ajustado.

El estudio contribuyó a brindar información acerca de la relación existente entre el peso de traslado a sitio 3 y las variables independientes. Esto sirvió para desarrollar una herramienta estadística que permitió estimaciones más precisas y objetivas que las realizadas antes del estudio.

Este estudio está dividido en cuatro capítulos. El capítulo I muestra el marco referencial que le da sustento estadístico a la investigación. En el capítulo II se presenta el marco teórico donde se describen los fundamentos de la regresión lineal, el análisis de correlación, el coeficiente de determinación y también los supuestos de normalidad, homocedasticidad e independencia, además también se presentan las características generales de la porcicultura, haciendo énfasis en el desarrollo de los lechones.

En el capítulo III se exponen los resultados de la investigación donde se describió el comportamiento de los datos tanto en forma tabular como gráfica, se muestran gráficas de dispersión, gráfica de caja y bigotes y gráficas de densidad; también se muestran los modelos diseñados y sus características más relevantes, así como los residuos y su análisis. El capítulo IV consiste en la discusión de resultados, dando lugar al análisis externo e interno.

## 1. MARCO REFERENCIAL

Los modelos de regresión lineal son ampliamente utilizados en diferentes ramas del conocimiento, y su aplicabilidad no se limita al análisis puramente descriptivo, esto ya lo hace ver Sánchez (2016) al indicar que los modelos de regresión lineal son fundamentales para proveer capacidad predictiva a los modelos calculados, de acuerdo a esta capacidad predictiva se puede entender su amplio uso en la industria y la porcicultura no es la excepción.

La regresión lineal como herramienta estadística para hacer estimaciones en la porcicultura es de gran utilidad, una muestra de ello es el estudio de Didivich Charneca y Françoise (2017) en el cuál logró estimar que un aumento de 100 g en el peso al nacimiento de los lechones se relacionaba de manera lineal y directa con un aumento de 26 gramos en la ingesta de calostro, y dado que a mayor consumo de calostro, mayor ganancia diaria de peso, se puede inferir que el peso al nacimiento tiene una relación con el peso al destete.

Coherentes con lo anterior son los resultados de Gondret *et. al.* (2005) quién encontró que lechones de bajo peso al nacimiento tenían un peso más bajo al destete y también un peso más bajo a los 68 días.

Por su parte, Hanke (2010) indica que es importante que cuando se desarrollan los pronósticos de datos, se comprenda el pasado de los mismos y usen estos datos históricos, en su buen juicio para hacer planes inteligentes que satisfagan las demandas del futuro.

El análisis de datos y el uso de la estadística inferencial para la resolución de problemas, siempre ha sido una herramienta ventajosa y asertiva, y ha permitido desarrollar conocimiento para la toma de decisiones. Como afirma Sánchez (2016), los modelos de regresión se consideran el objetivo fundamental en una investigación, sobre todo cuando se evalúa o se necesita conocer la relación existente entre una variable aleatoria regresora, y por consiguiente la respuesta que se origina.

El postula que los estudios de regresión lineal pueden llevar a cabo múltiples análisis, en consecuencia, un proceso de modelado donde se busca que los datos pertenezcan a una distribución normal. Estos estudios de regresión lineal permiten construir modelos predictivos o de otra índole estadística.

Los modelos de regresión lineal cobran fuerza como una herramienta valiosa de predicción. Sarmiento (2008), menciona que es de gran importancia para las empresas el poder contar con buenos modelos de pronósticos de las variables más críticas y sensibles.

La regresión lineal como técnica de predicción puede también considerar el hecho de utilizarse en métodos multivariados. Como lo menciona Valle (2018), el análisis de regresión lineal multivariado permite encontrar las relaciones existentes de variables químicas y físicas propias en una concentración de cemento. Este estudio permitió determinar la resistencia a la compresión bajo ciertas características experimentales, resaltando la importancia de esta técnica estadística al hacer estudios experimentales correspondientes, sin embargo, la regresión lineal multivariante escapa de los alcances de este estudio.

La construcción de modelos de regresión lineal también requiere que el investigador tenga un buen precedente en cuanto al orden teórico y metodológico

para sustentar la investigación de una manera correcta. En este orden de ideas, Bermejo (2020), menciona cómo la aplicación de la regresión lineal, para evaluar el nivel de la competencia laboral en un ámbito empresarial, dispuso de una herramienta de análisis estadístico que le brindó un sustento científico al análisis de los datos de forma cuantitativa.

Si al considerar que el análisis estadístico, le da un sustento científico y cuantitativo a este tipo de estudios categóricos, es de considerar que las variables deben estar bien definidas, para que el estudio de regresión lineal sea una herramienta verídica. Para este estudio se evaluó el nivel de competencias laborales y se realizó una propuesta de acciones que permitan mejorar la efectividad de los resultados económicos e incentivos.

El modelo de regresión lineal permitió evidenciar ciertos aspectos cuantitativos y su relación con aspectos que se definieron en la evaluación. Es importante lograr ajustar y entender las variables de forma en que, al ajustar las proyecciones al futuro de estos datos, permiten explicar de una forma segura la información esperada.

Los estudios realizados por Ceiro-Catasu *et. al.* (2021), afirmaron que al realizar un experimento de un análisis de presencia de una agente bacteriano hongo hematófago *Pochonia chlamydosporia* en distintos tipos de suelos mediante análisis estadístico ANOVA, permitió evidenciar que existían diferencias de la presencia de este agente según los suelos donde se había realizado el experimento.

Se aplicó un análisis de regresión lineal luego de comprobar estas diferencias, mediante pruebas *post-hoc*, para determinar la relación existente entre las variables de estudio y las poblaciones nativas de microorganismos.

Este estudio de regresión lineal permitió sustentar las relaciones que pudieran existir entre las variables definidas para poder inferir en cuanto al comportamiento o la explicación que la variable de respuesta tiene sobre las distintas variables regresoras, dando una mejor explicación al experimento. El análisis de regresión lineal tiene también la aplicación, en cuanto a los estudios de correlación entre variables.

Una buena manera de evaluar los métodos actuales y las técnicas futuras mediante análisis de datos se mediante un estudio de regresión lineal. Al utilizar la técnica metodológica propuesta y comparar con los métodos actuales, se obtiene un buen nivel de comparación entre los métodos. Esto se sostiene mediante lo propuesto por Puchades *et. al.* (2019), mediante la aplicación de los modelos de regresión lineal en el estudio de la resistencia de la caña de azúcar a la enfermedad Mosaico.

El estudio realizó un comparativo entre la metodología vigente utilizada y un método propuesto. Para lo cual se hizo un análisis de regresión lineal, que permitiera evaluar la variable de respuesta y la relación que tenían los tratamientos nuevos ofrecidos. Este análisis permitió concluir que, aunque en términos generales ambos métodos tuvieron respuestas similares, el nuevo tratamiento puede funcionar como una alternativa al actual, permitiendo obtener una buena respuesta.

Esta técnica estadística permite medir la relación que tienen las variables regresoras y su explicación a la variable de respuesta, y comparar con otros métodos o técnicas vigentes para descartar, validar o simplemente comparar comportamientos.

Estas referencias sirvieron de base teórica a este estudio, debido a que demuestran la flexibilidad de la aplicación de los modelos de regresión lineal para predecir la variable dependiente, en función a las variables regresoras. Por lo que se desarrollará el estudio con el uso de esta técnica estadística matemática.



## **2. MARCO TEÓRICO**

La estadística es una ciencia que tiene aplicación en cualquier área que sea susceptible de generar datos, sus dos grandes ramas son la estadística descriptiva y la estadística inferencial, esta última es particularmente útil ya que permite a través de una muestra realizar inferencias de una población, lográndose generar modelos con estimaciones predictivas, uno de los modelos más utilizados es el de la regresión lineal, Hanke (2010) indica que una vez establecida la relación lineal se puede emplear el conocimiento de la variable independiente para pronosticar la variable dependiente, esto da lugar a una regresión lineal simple, mientras que si hay más de una variable independiente el modelo corresponde a una regresión lineal múltiple.

### **2.1. Fundamentos de la regresión lineal general**

El modelo de regresión lineal se compone básicamente de elementos como lo son las variables regresoras, las cuales tienen propiedades cuantitativas que permiten desarrollar análisis de las variables explicativas y su relación con la variable de respuesta, dependiente o endógena.

#### **2.1.1. Análisis de correlación**

El coeficiente de correlación es una asociación entre variables, Newbold (2008) lo describe como una medida estandarizada de la relación lineal entre dos variables, y agrega que dicho coeficiente indica tanto el sentido como el grado de la relación.

Con respecto al sentido y al grado del coeficiente de correlación se puede decir que: el coeficiente de correlación va de -1 a +1 y cuanto más cerca se encuentra r de +1, más cerca se encuentran los datos de puntos de una línea recta ascendente que indica una relación lineal positiva. Cuanto más cerca se encuentra r de -1, más cerca se encuentran los datos de puntos de una línea recta descendente que indica una relación lineal negativa. Cuando  $r = 0$  no existe ninguna relación lineal entre x e y (Newbold, 2008).

La ecuación para el cálculo del coeficiente de correlación es la siguiente:

$$r = \frac{\text{cov}(X,Y)}{s_x s_y} \quad (\text{Ec. 1})$$

Donde:

Cov (X,Y) = Covarianza de los datos pareados X, Y

$S_x$  = Desviación estándar muestral de X

$S_y$  = Desviación estándar muestral de Y

### **2.1.2. Coeficiente de determinación $R^2$**

El coeficiente de determinación  $R^2$  es una medida del ajuste del modelo, Newbold (2018) indica que se considera que el coeficiente de determinación es el porcentaje de la variabilidad de Y que es explicado por la ecuación de regresión.

La ecuación del coeficiente de determinación resulta de restar 1 menos la suma de cuadrados del error (SCE) dentro de la suma de cuadrados total (SCT):

$$R^2 = 1 - \frac{SCE}{SCT} \quad (\text{Ec. 2})$$

El coeficiente de determinación es importante para evaluar el ajuste de un modelo, si el  $R^2$  se utiliza para comparar dos modelos hay que considerar lo mencionado por Newbold (2008) cuando expresa que cuando dos modelos de regresión tengan el mismo conjunto de variables dependientes observadas siempre pueden compararse utilizando el coeficiente de determinación  $R^2$ , sin embargo, advierte que las interpretaciones generales de  $R^2$  que se aplican a todas las ecuaciones de regresión son peligrosas.

### 2.1.3. Coeficiente de determinación $R^2$ ajustado

El  $R^2$  normalmente aumenta entre más número de variables independientes agreguemos al modelo, por tanto, se suele realizar un ajuste, este ajuste da como resultado precisamente el llamado  $R^2$  ajustado.

El  $R^2$  ajustado es el coeficiente modificado para justificar el número de variables y el tamaño de la muestra (Triola, 2018).

El  $R^2$  ajustado permite comparar mejor los modelos de regresión múltiple que tienen diferentes números de variables independientes (Newbold, 2008).

La ecuación del  $R^2$  ajustado es la siguiente:

$$R^2 \text{ ajustado} = 1 - \frac{SCE/(n-K-1)}{SCT/(n-1)} \quad (\text{Ec. 3})$$

Donde:

SCE = Suma de cuadrados del error  
SCT = Suma de cuadrados total  
n = Tamaño de la muestra  
K = Número de variables independientes

#### **2.1.4. Gráfico de cajas y bigotes**

La gráfica de caja y bigote es ampliamente utilizada ya que es muy útil para el análisis descriptivo de las medidas de posición de un conjunto de datos.

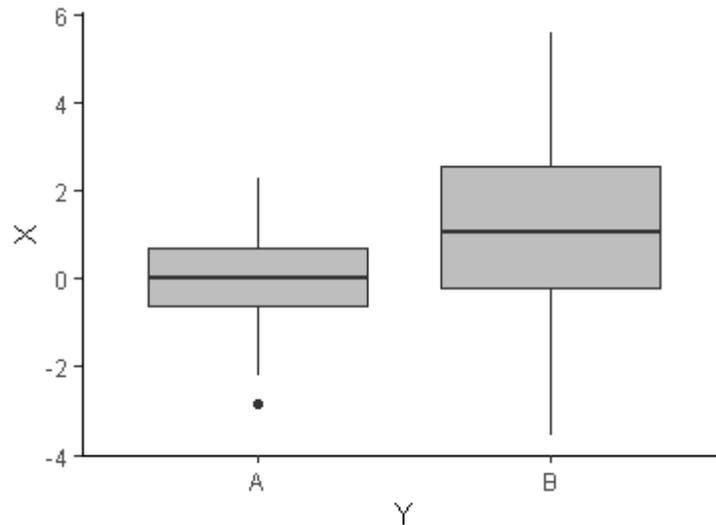
Es una gráfica de un conjunto de datos que consiste en una línea, que se extiende desde el valor mínimo hasta el valor máximo, y una caja con líneas trazadas en el primer cuartil, Q1, la mediana y el tercer cuartil, Q3 (Triola, 2018).

Existe una variante de la gráfica de caja, Triola (2018) le llama gráfica de cuadro modificada y señala que en ella los valores extremos son representados como puntos especiales, Walpole, Myers, Myers y Yeng (2012) menciona que pueden considerarse valores extremos aquellos cuya distancia desde la caja exceda 1.5 veces el rango intercuartil (en cualquier dirección).

Las gráficas de cuadro son especialmente útiles para comparar conjuntos de datos, en particular cuando pertenecen a la misma escala (Triola 2018).

Para muestras razonablemente grandes la presentación indica el centro de localización, la variabilidad y el grado de asimetría (Walpole, Myers, Myers y Yeng, 2012).

Figura 1. **Diagrama de cajas y bigotes (*Box-Plot*)**



Fuente: elaboración propia, empleando ggplot2.

### 2.1.5. Verificación de supuestos

Los modelos de regresión lineal simple y múltiple se generan basados en ciertos supuestos que hay que verificar para que el modelo tenga validez, estos supuestos se evalúan mediante el análisis de los residuos del modelo y son la normalidad, la homocedasticidad, la independencia.

#### 2.1.5.1. Normalidad

La distribución normal es una importante distribución continua, ya que, por medio de ella, se pueden aproximar muchas poblaciones útiles. Para identificar la distribución normal se requiere conocer la media y la desviación estándar. Una curva normal es simétrica y en forma de campana (Hanke, 2010).

Para realizar la inferencia estadística, se supone que el error ( $\varepsilon$ ) sigue una distribución normal de media 0 y varianza  $\sigma^2$  (Newbold, 2008).

La violación del supuesto de normalidad hace que las interpretaciones de los resultados no sean las que, a priori, se podrían deducir del uso de las pruebas en sí mismas (Pedrosa, Juarros-Basterretxea, Robles, Basteiro y Garica, 2014).

Para la evaluación de la normalidad existen varios métodos, tanto gráficos como numéricos.

- Gráfico cuantil-cuantil

Este gráfico es uno de los más utilizados para evaluar el supuesto de normalidad.

Para la generación de este gráfico, García y Sánchez (2006) indica que se grafican las posiciones de los cuantiles de la distribución de datos con respecto a los de otra distribución, la cual en este caso corresponde a una distribución teórica (normal), la interpretación consiste en asumir normalidad si los datos tienden en su mayoría a estar sobre la línea recta de la distribución teórica trazada.

Este método es muy práctico, sin embargo, tal como indica Ogunleye, Oyejola y Obisesan (2018), pese a que los métodos gráficos son útiles, no proporcionan suficiente evidencia concluyente de que el supuesto de normalidad efectivamente se cumple.

- Prueba de bondad de ajuste normalidad de Shapiro - Wilk

Dado que las pruebas gráficas no suelen ser confirmatorias, se debe utilizar una prueba estadística para confirmar o rechazar la normalidad de los datos, la prueba de Shapiro – Wilk es una de las más utilizadas para este fin.

Esta prueba calcula un estadístico llamado W el cuál es el estadístico de contraste para confirmar o rechazar la normalidad, los detalles de esta prueba fueron publicados en una revista científica en el año 1965, en dicha publicación Shapiro y Wilk (1965) resalta la ventaja de que es una prueba fácil de calcular una vez que se tenga la tabla de coeficientes disponible y menciona también su sensibilidad incluso para muestras menores a 20, en la misma publicación Shapiro y Wilk (1965) indica como desventaja que para muestras grandes los valores pueden ser difíciles de tabular; sin embargo, actualmente con el uso de software estadístico con alto poder de cómputo esto no resulta ningún inconveniente.

En un artículo comparativo entre pruebas de normalidad, Pedrosa, Juarros-Basterretxea, Robles, Basteiro y Garica (2014), indica que la prueba de Shapiro–Wilk es una de las más consistentes a la variación muestral, también señala que tiene un poder de detección razonable para muestras menores de 50 y que su mejor poder de detección lo presentó en muestras con n = 50.

$$\frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{Ec. 4})$$

### **2.1.5.2. Homocedasticidad**

Para entender este concepto, es importante lo que cita Gujarati y Porter (2009), al indicar que la homocedasticidad se postula como una igualdad de varianzas.

En cuanto al modelo de regresión, los residuos se deben distribuir con varianza constante, para ello se puede realizar un diagnóstico gráfico a través del gráfico de Residuos contra Predichos, y se puede evaluar de acuerdo con Ludwig, Thomas, Stefan y Marx (2013), mediante la prueba de Breusch-Pagan cuya base es un modelo multiplicativo para las varianzas del error, si el valor-p < 0.05 se rechaza la hipótesis de que las varianzas no son iguales.

### **2.1.5.3. Independencia**

Los residuos del modelo deben presentar independencia, es decir que no deben tener un patrón definido ya que esto supondría un sesgo importante.

La independencia se puede evaluar de manera gráfica, y a través de pruebas estadísticas.

- Gráfico de residuos estandarizados ordenados

Una herramienta sencilla para diagnosticar errores correlacionados son los gráficos de dispersión de residuos estandarizados a lo largo del tiempo (o en el orden en el que fueron tomados), si un residuo positivo (negativo) es seguido en tendencia por un residuo positivo (negativo) los errores están correlacionados positivamente, así mismo si en hay tendencia de signos alterna en los residuos hay correlación negativa (Ludwig, Thomas, Stefan y Marx, 2013).

- Correlograma

Una forma de evaluar de manera visual la autocorrelación o independencia es el llamado correlograma, esta gráfica se basa de acuerdo a Ludwig, Thomas, Stefan y Marx (2013), en la función empírica de autocorrelación cuya ecuación es la siguiente:

$$ACF_{(j)} = \frac{\text{cov}(\varepsilon_{j3} \varepsilon_{i-j})}{\text{var}(\varepsilon_i)} \quad (\text{Ec. 5})$$

Donde:

Cov = Covarianza

Var = Varianza

- Prueba de Durbin-Watson

La independencia también se puede evaluar mediante el cálculo de un estadístico de contraste mediante la prueba de Durbin-Watson.

La prueba de Durbin Watson fue publicada por primera vez allá por el año de 1951, en la introducción del artículo menciona Durbin y Watson (1971), que el desarrollo de la prueba surgió para evaluar la independencia de los errores, además hace énfasis en lo común de utilizar el método de mínimos cuadrados en datos de series temporales lo cual es incorrecto dado que las observaciones sucesivas están correlacionadas.

Ludwig, Thomas, Stefan y Marx (2013), menciona que la prueba de Durbin-Watson provee evidencia de autocorrelación (es decir de falta de independencia) aunque no determina el tipo de la misma.

La prueba de Durbin-Watson considera dos hipótesis  $H_0 = \rho = 0$  y  $H_1 = \rho \neq 0$ .

El estadístico de contraste  $d$  es el utilizado en esta prueba, si  $d$  es muy cercano a 2, entonces el valor  $p$  es próximo a 0, por tanto, se rechazaría la hipótesis nula (Ludwig, Thomas, Stefan y Marx, 2013).

El rechazo de la hipótesis nula conlleva la confirmación de autocorrelación, así mismo el no rechazo de la hipótesis nula conlleva la confirmación de falta de autocorrelación.

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} \quad (\text{Ec. 6})$$

#### **2.1.5.4. Colinealidad**

La colinealidad es la situación en la que las variables independientes de una ecuación de regresión múltiple están altamente intercorrelacionadas (Hanke, 2010).

Una colinealidad perfecta es causada por un error en la formulación del modelo lineal, como no emplear una categoría de referencia para las variables ficticias (*dummy*).

Para la detección de colinealidad se puede utilizar la prueba estadística llamada VIF (factor de inflación de la varianza).

Debido a su simplicidad e interpretación directa, el VIF (o su raíz cuadrada) es el diagnóstico básico de colinealidad.

La ecuación para el cálculo del VIF es la siguiente:

$$VIF_j = 1 / (1 - R_j^2) \quad (\text{Ec. 7})$$

Cuando la colinealidad no se limita a pares de variables explicativas se le denomina multicolinealidad.

## 2.2. Modelo lineal general

La regresión lineal tiene como objetivo diseñar estadísticamente, un modelo en forma matemática del comportamiento de las variables regresoras y en función de la respuesta de una variable dependiente.

Pretende explicar la estimación de parámetros necesarios para que los datos experimentales u observacionales, puedan estimar el comportamiento de la variable de respuesta. Gutiérrez y Vara (2008), explican que, si se tienen dos variables cuantitativas X y Y, se supone que se explica el comportamiento de Y basándose en los valores que toma X en determinado tiempo y condiciones. A esto se le conoce como el modelo lineal general simple y esta descrito con la ecuación siguiente.

$$Y = \beta_0 + \beta_1 x \quad (\text{Ec. 8})$$

De esta ecuación se busca estimar y contrastar la hipótesis nula que dice que la pendiente es estadísticamente significativa de cero. Ya que de ser así se anularía el efecto de  $\beta_i$ .

$$H_0: \beta_i = 0 \quad (\text{Ec. 9})$$

$$H_a: \beta \neq 0 \quad (\text{Ec. 10})$$

Si en el sistema lineal se tiene más de una variable regresora, y se tiene siempre un sistema de variables que explican de forma interrelacionada el comportamiento de la variabilidad de la independiente, debe definirse bajo otra característica estadística.

Gutiérrez y Vara (2008), también postulan que, para estos casos, se considera que en muchas situaciones prácticas existen varias variables independientes. Afirmación que es muy acertada en el campo experimental, debido a que la cantidad de fuentes de variación siempre son muchas.

De ser así será necesario considerar que se quiere predecir o entender el comportamiento de Y, la cual se encontraría en función de estas regresoras. Para ello se debe considerar la siguiente ecuación.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_p X_{p_i} + \varepsilon_i \quad (\text{Ec. 11})$$

Para afirmar o descartar este postulado, es necesario evaluar las variables regresoras, las cuales se considera que influyen o tienen una relación con la variable de respuesta Y.

$$H_o : \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (\text{Ec. 12})$$

$$H_a : \beta_j \neq 0 \quad (\text{Ec. 13})$$

Para evaluar el efecto que tiene cualquier pendiente b sobre una variable X se debe contrastar la hipótesis nula planteada, para que se pueda evaluar el nivel de contribución que tiene esta variable sobre la explicación de la variable de

respuesta, en el modelo lineal múltiple. Rechazar  $H_0$ , se traduciría en que al menos, se tiene un término explicativo en las variables regresoras que, permitiría al modelo ser capaz de explicar la variabilidad de la variable independiente.

$$t_j = \frac{\hat{\beta}_j}{s.e.(\hat{\beta}_j)} \quad (\text{Ec. 14})$$

### 2.3. Análisis de varianza

El análisis de varianza consiste en separar la variación total observada en cada una de las fuentes que contribuye a la misma (Gutierrez, 2008).

Así pues, sus aplicaciones son diversas, en una regresión lineal es útil para evaluar la significancia de las variables predictoras en el modelo.

El resultado de un ANOVA es el estadístico de prueba F, que debe ser comparado con un valor crítico correspondiente a la significancia seleccionada, su ecuación es la siguiente:

$$F_0 = \frac{CM_R}{CM_E} \quad (\text{Ec. 15})$$

Donde  $CM_R$  se refiere a los cuadrados medios de la regresión y  $CM_E$  a los cuadrados medios del error, quedando formados los cuadrados medios de la regresión de la siguiente forma:

$$CM_R = \frac{SC_R}{1} \quad (\text{Ec. 16})$$

Y los cuadrados medios del error de la siguiente forma:

$$CM_E = \frac{SC_E}{n-2} \quad (\text{Ec. 17})$$

Donde  $SC_R$  se refiere a la suma de cuadrados de la regresión y se forma de la siguiente manera:

$$SC_R = \widehat{\beta}_1 S_{xy} \quad (\text{Ec. 18})$$

Mientras que  $sc_E$  se refiere a la suma de cuadrados del error y se forma así:

$$sc_E = S_{yy} - \widehat{\beta}_1 S_{yy} \quad (\text{Ec. 19})$$

Donde:

$$S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (\text{Ec. 20})$$

#### **2.4. Desarrollo fisiológico de los lechones**

El desarrollo fisiológico del lechón desde su nacimiento hasta el momento de su traslado al sitio 3 está relacionado con múltiples factores, el peso al nacimiento, demostrando que los lechones más pesados al nacimiento tienen una mayor ganancia de peso durante la lactación, también demostró que la posición del pezón de la ubre de la cual se alimenta el lechón está relacionada con su ganancia de peso, esto tiene sentido ya que los pezones más craneales tienen mayor producción de leche que los más caudales, así mismo demuestra que el tipo de alojamiento y el tipo de alimentación influyen en la ganancia diaria de peso.

### 3. PRESENTACIÓN DE RESULTADOS

De acuerdo con los objetivos propuestos, se presentan los siguientes resultados.

#### 3.1. Modelo de regresión lineal

Con el fin de llevar a cabo la estimación del peso de lechones al momento de su traslado a sitio 3 se desarrollaron diferentes modelos de regresión lineal, que van desde modelos de regresión lineal simple, pasando por modelos de regresión lineal múltiple con 2, 3 y hasta 4 variables dependientes.

La tabulación de los datos se llevó a cabo en una hoja electrónica de Excel y el análisis estadístico se hizo mediante el software R con la ayuda del entorno de desarrollo integrado *RStudio*, presentándose para los diferentes objetivos los siguientes resultados.

##### 3.1.1. **Objetivo 1: Hallar el nivel de correlación entre el peso al nacimiento, el peso al destete y el peso de traslado a sitio 3 mediante modelos de regresión lineal**

Se calculó el índice de correlación a través del método de Pearson para la variable dependiente en función de cada variable regresora.

Las variables que presentaron una correlación más fuerte fueron el peso al destete y el peso al nacimiento, con una correlación de 0.5408 y 0.4794

respectivamente; mientras que las variables con correlación más débil fueron la edad al destete y la edad al traslado.

Ambas variables presentaron una correlación de 0.3324, como podemos notar todas las variables presentaron una correlación positiva con un grado intermedio de fuerza, a continuación, se muestra en forma de tabla la matriz de correlación.

Tabla III. **Correlación de Pearson para el modelo**

Tabla de correlación mediante el método de Pearson				
	x1	x2	x3	x4
	Peso al nacimiento	Peso al destete	Edad al destete	Edad al traslado
y				
Peso al traslado	0.4794	0.5408	0.3324	0.3324

Fuente: elaboración propia, empleando Microsoft Excel.

### 3.2. Estimación del peso de lechones (*Sus scrofa domesticus*) al momento de su traslado a sitio 3 mediante un modelo de regresión lineal

El mejor modelo encontrado para estimar el peso de los lechones al momento de su traslado a sitio 3 mediante un modelo de regresión lineal fue el modelo 11, cuya fórmula la podemos describir de la siguiente forma:

$$lm \text{ (fórmula = } y \sim x1 + x2 + x3) \quad (\text{Ec. 21})$$

Donde:

lm = es un modelo de regresión lineal, cuya fórmula está dada por una variable dependiente y con 3 variables independientes ( $x_1 + x_2 + x_3$ )

Y = es la variable independiente peso al traslado

X1 = es la variable independiente peso al nacimiento

X2 = es la variable independiente peso al destete

X3 = es la variable independiente edad al destete

Con este modelo, definido gracias a los criterios de información de Akaike, Bayesiano y  $R^2$  ajustado se logró realizar el ajuste lineal que minimiza los errores cuadráticos, y que dados los valores obtenidos para el intercepto y la pendiente logran estimar el peso de los lechones al momento del traslado a sitio 3.

La ecuación de regresión lineal que responde al modelo propuesto quedó definida de la siguiente manera:

$$Y = 26.0744 + 5.4974x_1 + 1.1914x_2 + 0.8755x_3 \quad (\text{Ec. 22})$$

Tabla IV. **Resumen de resultados de coeficientes de regresión**

Resultados de la regresión usando Peso al momento de traslado a sitio 3 como criterio							
Predictor	<i>B</i>	<i>B</i> 95 % IC [LI, LS]	<i>beta</i>	<i>beta</i> 95 % CI [LI, LS]	<i>sr</i> <sup>2</sup>	<i>sr</i> <sup>2</sup> 95 % CI [LI, LS]	<i>r</i>
(Intercepto)	26.07* *	[9.34, 42.81]					
x1	5.50**	[2.20, 8.79]	0.37	[0.15, 0.59]	.11	[-.01, .24]	.48* *
x2	1.19**	[0.41, 1.97]	0.34	[0.12, 0.57]	.09	[-.02, .21]	.54* *

Continuación tabla IV.

x3	0.88**	[0.26, 1.49]	0.30	[0.09, 0.50]	.08	[-.02, .19]	.33*
							$R^2 = .460^{**}$
							95 %
							IC[.24, .58]
<p>Nota. b representa regresiones ponderadas no estandarizadas. beta indica las ponderaciones de regresión estandarizadas.  <math>sr^2</math> representa la correlación semiparcial al cuadrado. r representa la correlación de orden cero.          LI y LS indican el límite inferior y el límite superior de un intervalo de confianza respectivamente.          * indica <math>p &lt; .05</math>. ** indica <math>p &lt; .01</math>.</p>							

Fuente: elaboración propia, empleando Microsoft Excel.

El  $R^2$  del modelo fue de 0.460, por tanto, el modelo lineal generado tiene un buen ajuste a los datos. Los coeficientes de correlación de las tres variables independientes resultaron significativos, lo que demuestra que cada una de ellas tiene influencia en el valor esperado de Y.

Tabla V. Resultados del ANOVA

Resultados del ANOVA usando el peso de traslado a sitio 3 como criterio							
Predictor	Suma de Cuadrados	gl	Cuadrados Medios	F	p	$\eta^2$ parcial	$\eta^2$ parcial 95 % CI [LI, LS]
(Intercepto)	489.35	1	489.35	9.76	.003		
Peso nacimiento	560.94	1	560.94	11.19	.002	.17	[.03, .34]
Peso destete	469.34	1	469.34	9.36	.003	.15	[.02, .32]
Edad destete	414.03	1	414.03	8.26	.006	.13	[.01, .30]
Error	2707.89	54	50.15				
<p>Nota. LI y LS representan el límite inferior y el límite superior respectivamente de los intervalos de confianza.</p>							

Fuente: elaboración propia, empleando Microsoft Excel.

Con base a lo anterior, se logró estimar de manera confiable el peso de los lechones al momento del traslado a sitio 3, sin embargo, el  $R^2$  muy probablemente se puede mejorar incluyendo otras variables independientes adicionales a las evaluadas en el presente estudio.

Se puede observar también que el cuadrado medio del error es bajo comparado con los cuadrados medios de las variables independientes, esto es un indicador de que dichas variables explican gran parte de la variabilidad del modelo, observando el valor p también se comprueba este hecho ya cada variable independiente tuvo un valor p menor a 0.05 por tanto todas son estadísticamente significativas.

### 3.2.1. **Objetivo 2: Evaluar el mejor modelo por medio del criterio de información de Akaike, el criterio de información Bayesiano y el $R^2$ ajustado**

A continuación, se muestra una tabla con los resultados del  $R^2$  ajustado, criterio de información de Akaike (AIC) y criterio de información Bayesiano (BIC) para cada uno de los modelos.

Tabla VI. **Tabla de resultados de los criterios de información utilizados**

<b>Modelo</b>	<b><math>R^2</math> ajustado</b>	<b>AIC</b>	<b>BIC</b>
m1	0.2161	414.1069	420.2882
m2	0.2798	409.1865	415.3678
m3	0.0946	422.4634	428.6447
m4	0.0946	422.4634	428.6647
m5	0.3547	403.7709	412.0127
m6	0.3433	404.7896	413.0314
m7	0.3433	404.7896	413.0314

Continuación tabla VI.

m8	0.3244	406.4381	414.6798
m9	0.3244	406.4381	414.6798
m10	0.0946	422.4634	428.6447
m11	0.4210	397.5189	470.8211
m12	0.4210	397.5189	407.8211
m13	0.3433	404.7896	413.0314
m14	0.3244	406.4381	414.6798
m15	0.4210	397.5189	407.8211

Fuente: elaboración propia, empleando Microsoft Excel.

De acuerdo al  $R^2$  los mejores modelos son: m11, m12 y m15, ya que son los que poseen los valores más altos y por tanto son los que explican en mayor medida el comportamiento de la variable dependiente.

Los mejores modelos según el criterio de información de Akaike (AIC) son los modelos m11, m12 y m15 dado que son los que poseen valores más pequeños y por tanto tienen una menor pérdida de información siendo entonces los modelos mejor ajustados; sin embargo, utilizando el criterio de información Bayesiano (BIC) no queda claro cuál modelo es mejor ya que todos los modelos tienen un índice similar, esto resulta interesante dado que el BIC penaliza más la adición de variables al modelo, es decir que el BIC valora más la parsimonia que el AIC.

Ya establecidos los mejores 3 modelos se procedió a analizar el índice de la inflación de la varianza (VIF) para cada uno de ellos, este índice nos permitió establecer si hay o no multicolinealidad, los resultados se muestran a continuación:

Tabla VII. **Presentación de resultados para el modelo de regresión lineal M11**

<b>Modelo 11</b>			
<b>Variable</b>	<b>X1</b>	<b>X2</b>	<b>X3</b>
<b>VIF</b>	1.2122	1.2486	1.06343

Fuente: elaboración propia, empleando Microsoft Excel.

Tabla VIII. **Presentación de resultados para el modelo de regresión lineal M12**

<b>Modelo 12</b>			
<b>Variable</b>	<b>X1</b>	<b>X2</b>	<b>X4</b>
<b>VIF</b>	1.2122	1.2486	1.06343

Fuente: elaboración propia, empleando Microsoft Excel.

Tabla IX. **Presentación de resultados para el modelo de regresión lineal M15**

<b>Modelo 15</b>				
<b>Variable</b>	<b>X1</b>	<b>X2</b>	<b>X3</b>	<b>X4</b>
<b>VIF</b>	1.2266	1.2616	5953195	5953047

Fuente: elaboración propia, empleando Microsoft Excel.

Valores VIF bajos se relacionan con poca multicolinealidad, y valores VIF altos se relacionan con mucha multicolinealidad, en general se acepta que valores VIF arriba de 10 muestran alta multicolinealidad, por esto se determinó que hay multicolinealidad entre la variable X3 y X4 y por ende el modelo 15 no resulta apropiado.

Se concluyó que el mejor modelo es el m11 ya que es el que posee junto con el m12 los mejores valores tanto de  $R^2$  ajustado como de AIC y BIC, sin embargo, el modelo 11 tiene una ventaja práctica respecto al modelo 12, esta ventaja se debe a que posee entre sus variables independientes la variable X3 (edad al destete), mientras que el modelo 12 posee en su lugar la variable X4 (edad al traslado), esto hace que con el modelo 11 se puedan hacer estimaciones más tempranas que con el modelo 12.

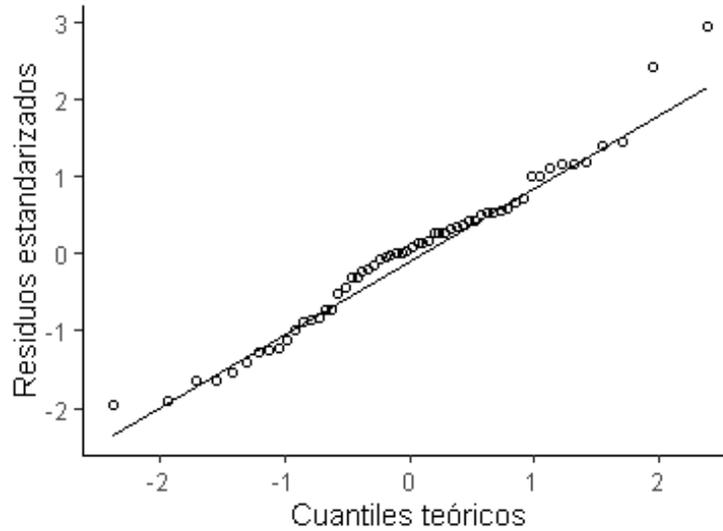
### **3.2.2. Objetivo 3: Comprobar la normalidad, homocedasticidad, independencia y falta de multicolinealidad del modelo.**

Se procedió a la realización de los cálculos de los supuestos necesarios para la comprobación del uso de métodos paramétricos en el modelo.

#### **3.2.2.1. Normalidad**

Gráficamente mediante el QQplot se observó que los residuos tienen normalidad, para corroborar dicha normalidad se realizó la prueba de Shapiro-Wilk que presentó un valor para el estadístico W de 0.0973 que corresponde a un valor P de 0.242 y dado que el valor P es mayor a 0.05 se confirma que los residuos siguen una distribución normal.

Figura 2. **Gráfico de QQ-Plot para el modelo elegido**

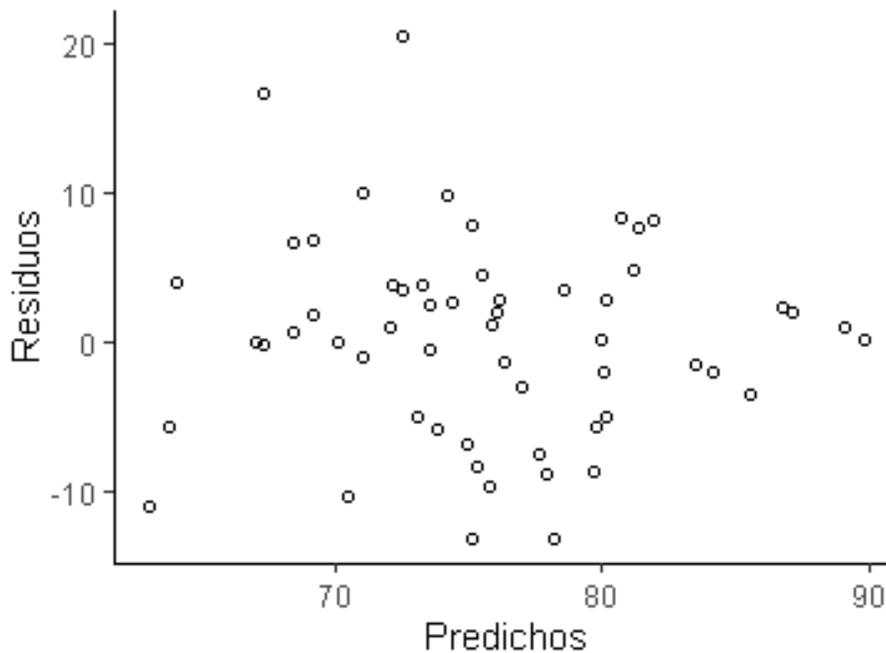


Fuente: elaboración propia, empleando QQ-Plot.

### **3.2.2.2. Homocedasticidad**

Visualmente mediante la gráfica de residuos contra predichos se pudo notar constancia en la varianza, por tanto, se cumple con el supuesto de homocedasticidad. La prueba de Breusch-Pagan indicó un valor BP = 3.132 el cual es equivalente a un p-valor de 0.371, el cual es mayor a la significancia concluyendo que el modelo es homocedástico.

Figura 3. **Gráfico de residuos vs predichos**

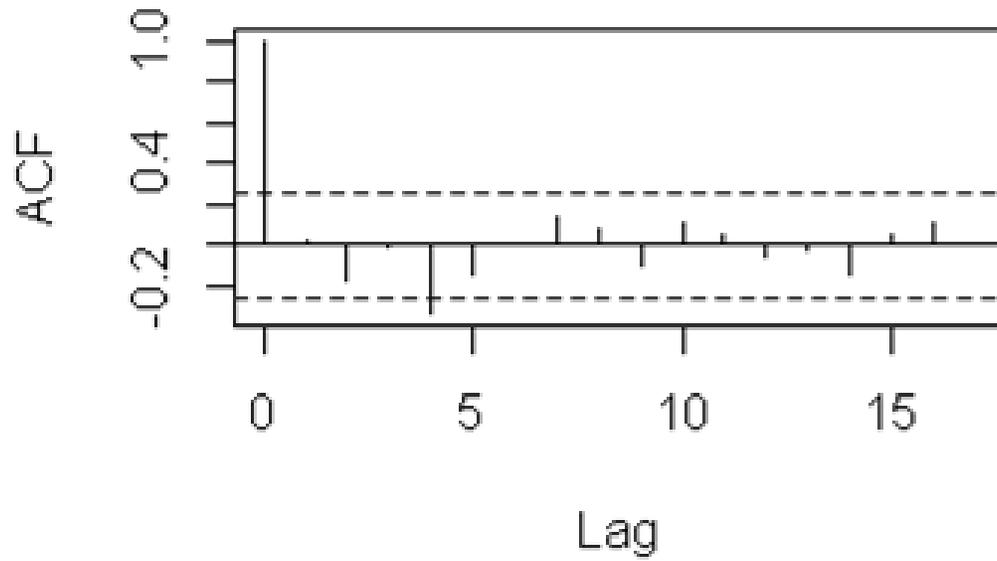


Fuente: elaboración propia, empleando QQ-Plot.

### 3.2.2.3. Independencia

La independencia se evaluó mediante un correlograma, la regla indica que si más del 10 % de las barras están fuera de las líneas punteadas que representan intervalos de confianza al 95 % significa que los datos son independientes, tal es el caso del modelo evaluado. La prueba de Durbin-Watson dio un estadístico  $DW = 1.918$ , equivalente a un valor P de 0.326 confirmando la independencia del modelo.

Figura 4. **Correlograma para el modelo de regresión lineal**



Fuente: elaboración propia, empleando QQ-Plot.



## **4. DISCUSIÓN DE RESULTADOS**

La presente investigación se realizó con el fin de aportar una herramienta estadística para ayudar a comprender la relación entre edad y peso en lechones y de esta manera poder realizar estimaciones que permitan tomar decisiones que impacten positivamente a la producción porcina.

### **4.1. Análisis interno**

Para que un modelo de regresión lineal sea válido, los residuos deben distribuirse con normalidad, este análisis se realizó de manera gráfica a través de un gráfico cuantil-cuantil y mediante una prueba estadística formal que fue la prueba de Shapiro-Wilk, en ambas pruebas se demostró la normalidad, cabe señalar que la prueba de Shapiro-Wilk fue la más óptima para este estudio ya que indica Ogunleye, Oyejola y Obisesan (2018), al realizar comparativas entre pruebas de normalidad encontró que la prueba de Shapiro-Wilk fue la más consistente.

Otro supuesto necesario que deben cumplir los residuos es el de homocedasticidad, el mismo se comprobó gráficamente mediante la gráfica de residuos contra predichos gracias a que no se observó ningún patrón definido.

El supuesto de independencia también se cumplió con la ayuda del correlograma.

Así pues, gracias a que los residuos cumplieron los requisitos, el modelo tiene validez, aunque no deja de tener limitaciones ya que deben evaluarse los supuestos de normalidad en cada población de la que se quieran hacer estimaciones basadas en este modelo.

#### **4.2. Análisis externo**

Para el análisis de los resultados obtenidos, en su investigación Iwasawa *et. al.* (2004), desarrolló un modelo de regresión para estimar el peso de cerdas basado en la medida de flanco a flanco, encontrando una correlación positiva fuerte entre la variable independiente y la variable dependiente.

Murcia (2015) relacionó el peso al nacer con el peso al destete y el peso a los 18 meses de crías bovinas de raza Brahman hallando una correlación baja entre el peso al nacimiento y el peso al destete de entre 0.13 y 0.17 mientras que en la presente investigación se encontró una correlación entre el peso al nacimiento y el peso al traslado al sitio 3 de 0.4794 lo cual permite que el peso al nacimiento sea más fiable para hacer estimaciones en la especie porcina que en la especie bovina.

Las variable regresora con mayor correlación respecto a la variable independiente fue el peso al nacimiento, el impacto del peso al nacimiento en el desempeño de los lechones lo evidencia en su estudio Didivich Charneca y Françoise (2017), en el cuál halló que los lechones con menor peso al nacimiento tuvieron mayor mortalidad, además encontró que el consumo de calostro aumentaba  $26 \pm 1.6$  gramos por cada 100 gramos más de peso al nacimiento, dado que el calostro es una fuente importante no solo de inmunoglobulinas sino también de energía y nutrientes en general, un alto consumo contribuye a una mayor ganancia diaria de peso.

La relación lineal encontrada en este estudio tiene implicaciones tanto en la variable dependiente que fue el peso de traslado a sitio 3 y también en otras variables no contempladas en este estudio, pero relevantes en la fase de finalización de los cerdos, esto queda evidenciado en el estudio de Gondret *et. al.* (2005) en el que demuestra que el peso al nacimiento impacta en la velocidad de crecimiento dado que los cerdos con bajo peso al nacimiento tardaron 12 días más en alcanzar el mismo peso al sacrificio que los cerdos de alto peso al nacimiento.

La relevancia de estimar el peso al traslado a sitio 3 basado en indicadores como el peso al nacimiento, el peso al destete y la edad al destete radica en que se pueden tomar decisiones tempranas de manejo para mejorar el desempeño de los cerdos en sus siguientes etapas de desarrollo, en este estudio se evidenció que buenos parámetros en las etapas tempranas se correlacionan con pesos altos al momento del traslado a sitio 3 permitiendo hacer estimaciones mediante un modelo de regresión lineal, esta relación entre el desempeño en las etapas tempranas y el desempeño en las etapas subsiguientes están acordes con el estudio de Fix *et. al.* (2009) quién encontró que un mayor peso al nacimiento se asocia a un mayor peso en el futuro y una ganancia diaria de peso más rápida, este conocimiento ayuda a tomar mejores decisiones enfocadas en mejorar la eficiencia productiva, haciendo que los modelos de regresión lineal se perfilen como una herramienta muy útil en la estimación de parámetros zootécnicos y su aplicación permite que las explotaciones productivas sean más eficientes.



## CONCLUSIONES

1. Se definió el nivel de correlación de la variable dependiente (peso de traslado a sitio 3) respecto a cada una de las variables independientes, y se concluye que la correlación es aceptable. El coeficiente de determinación general del modelo fue del 0.46, esto significa que las variables utilizadas explican el 46 % de la variable dependiente.
2. Utilizando los criterios de información de Akaike, Bayesiano y  $R^2$  ajustado, fueron 3 los mejores modelos: el modelo m11, m12 y m15, sin embargo dado que el modelo m15 incluye a todas las variables no resulta ser el más parsimonioso, y aplicando la prueba del factor de inflación de la varianza (VIF) resultó con colinealidad, lo cual lo invalidó y quedó descartado, entre el m11 y el m12 el mejor modelo se decidió que es el m11 debido a que la variable que no comparte con el m12 es la edad al destete, mientras que el m12 la variable que no comparte con el m11 es la edad al traslado a sitio 3, esto implica que el modelo m11 sea capaz de estimar el peso de los lechones con anticipación, es decir que resulta ser predictivo, mientras que el modelo m12 resulta ser inferencial pero no predictivo.
3. El modelo evaluado cumplió con los supuestos de normalidad, homocedasticidad e independencia, lo que significa que la generación de un modelo de regresión lineal basado en las variables utilizadas es válida.
4. Se estimó el peso de traslado a sitio 3 de lechones (*Sus scrofa domesticus*) mediante un modelo de regresión lineal, utilizando las variables regresoras que tuvieron un coeficiente de determinación más

alto y que demostraron explicar mejor los datos, estas variables fueron el peso al nacimiento, el peso al destete y la edad al destete.

## RECOMENDACIONES

1. Evaluar la correlación entre el peso de traslado a sitio 3 y otras variables no contempladas en este estudio como la cantidad de ingesta de calostro.
2. Someter a los métodos de criterio de información, al modelo de regresión lineal, con el fin de seleccionar el mejor ajuste, para el pronóstico de los pesos de los lechones al momento del traslado a sitio 3, y así tomar decisiones en busca de mejorar la eficiencia productiva.
3. Realizar el análisis de los residuos del modelo para la verificación del cumplimiento de los supuestos de normalidad, homocedasticidad e independencia.
4. Evaluar un modelo de proyección del peso de lechones (*Sus scrofa domestica*) mediante la utilización de otras variables adicionales que se consideren significativas a la variable de respuesta, durante las ocho semanas posteriores al destete de los lechones, con el fin de desarrollar un modelo con mejor ajuste.



## REFERENCIAS

1. Aceituno, V. (2018). *Importancia de la aplicación de modelos de regresión lineal en proyecciones de la cobertura boscosa y de captura de CO<sub>2</sub>, para la venta de certificados de carbono en Guatemala, período 2017-2030* (Tesis de maestría). Universidad de San Carlos de Guatemala, Guatemala.
2. Asencio, A. (2018). *Modelo de regresión lineal aplicado a la cantidad de contenedores* (Tesis de maestría). Universidad de San Carlos de Guatemala, Guatemala.
3. Bermejo, M. (enero, 2020). Tratamiento del nivel de competencias laborales desde la regresión lineal simple. *Retos de la dirección*, 14(2), 264-277. Recuperado de <http://scielo.sld.cu/pdf/rdir/v14n1/2306-9155-rdir-14-01-264.pdf>.
4. Ceiro-Catasu, W. Hidalgo, M. Hidalgo, L. Arévalo, J. García, M. y Mazon, J. (enero, 2021). Establecimiento in-vitro del hongo hematófago *Pochonia chlamydosporia* var. *catenulata* en diferentes suelos. *Terra Latinoamericana*, 39, 1-7.
5. Didivich, J. Charneca, R. y Françoise, T. (2017). *Relationship between birth order, birth weight, colostrum intake, acquisition of passive immunity and pre-weaning mortality of piglets*. Évora, Portugal: Universidad de Évora.

6. Durbin, J. y Watson, G. (marzo, 1971). Testing for serial correlation in least squares regression. *Biometrik*, 58(1), 1-19. Recuperado de <https://www.jstor.org/stable/2334313>.
7. Fix, J., Cassady, J., Herring, W., Holl, J., Culbertson, M. y See, M. (2009). *Effect of piglet birth weight on body weight, growth, backfat, and longissimus muscle area of commercial market swine*. Estados Unidos: Universidad Estatal de Carolina del Norte.
8. García, C. y Sánchez, M. (2006). *Herramientas de Análisis Estadístico y Espacial. Herramientas de Análisis Estadístico y Espacial aplicadas en la separación Línea Base – Anomalía Geoquímica. Implicaciones en Exploración de Recursos Naturales*. Bogotá, Colombia: Revisión metodológica.
9. Gondret, F., Lefaucheur, L., Juin, H., Louveau, I. y Lebret, B. (enero, 2005). Low birth weight is associated with enlarged muscle fiber area and impaired meat tenderness of the longissimus muscle in pigs. *Journal of Animal*, 84(1), 93-103. Recuperado de <https://academic.oup.com/jas/article-abstract/84/1/93/4804155>.
10. Gujarati, D. y Porter D. (2009). *Econometría*. México: Mc Graw Hill.
11. Gutiérrez, P. y Vara, R. (2008) *Análisis y diseño de experimentos*. México: McGraw Hill.
12. Hanke, J. (2010). *Pronósticos en los negocios*. México: Pearson educación.

13. Iwasawa, T., Young, M., Keegan, T., Tokach, M., Goodband, R., DeRouchey, J., Nelssen, J. y Dritz, S. (2004). *Comparison of heart girth or flank-to-flank measurements for predicting sow weight*. Estados Unidos: FAO.
14. Llinás, H. (2017). *Estadística inferencial*. Colombia: Editorial Universidad del Norte.
15. Ludwig, F., Thomas, K., Stefan, L. y Marx, B. (2013). *Regression: Models, Methods and Applications*. Berlín, Alemania: Springer-Verlag.
16. Montgomery, D. y Hines, W. (1996). *Probabilidad y estadística para ingeniería y administración*. México: Editorial Continental.
17. Murcia, J. (2015). *Relación del peso al nacer con el peso al destete y el peso a los 18 meses en un hato de cría Brahman en Tame (Arauca)*. (Tesis de licenciatura). Universidad de la Salle, Colombia.
18. Newbold, P. (2008). *Estadística para administración y economía*. Madrid, España: Pearson - Prentice Hall.
19. Ogunleye, L., Oyejola, B. y Obisesan, K. (marzo, 2018). Comparison of Some Common Tests for Normality. *International Journal of Probability and Statistics*, 7(5), 130-137. Recuperado de <http://article.sapub.org/10.5923.j.ijps.20180705.02.html>.
20. Pedrosa, I., Juarros-Basterretxea, J., Robles, A., Basteiro, J. y Garica, E. (enero, 2014). Pruebas de bondad de ajuste en distribuciones

simétricas, ¿Qué estadístico utilizar? *Revista Psicológica*, 14(1), 245-254.

21. Puchades, Y., La O, M., Rodríguez, E. y Rodríguez, M. (octubre, 2019). Modelos lineales para evaluar resistencia a la enfermedad mosaico de la caña de azúcar. *Centro Agrícola*, 46(4), 30-37. Recuperado de <http://cagricola.uclv.edu.cu/index.php/es/>.
22. Reyes, J., Flores, C., Vargas, J. y Ramírez, P. (abril, 2007). Una aplicación del modelo de regresión logística en la predicción del rendimiento estudiantil. *Estudios Pedagógicos*, XXXIII(2), 101-120. Recuperado de <https://scielo.conicyt.cl/pdf/estped/v33n2/art06.pdf>.
23. Sánchez, D. (2016). *Control estadístico y modelos de regresión lineal. Una forma práctica de control de puentes* (Tesis de doctorado). Universidad de Cantabria, España.
24. Sarmiento, E. (enero, 2008). Predicción con series de tiempo y regresión. *Panorama*, 2(4), 1-19.
25. Shapiro, S. y Wilk, B. (junio, 1965). An analysis of variance test for normality. General Electric Co. and Bell Telephone Laboratories. *Biometrika*, 52(3-4), 591-611. Recuperado de <https://www.jstor.org/stable/2333709>.
26. Triola, M. (2018). *Estadística*. México: Pearson Hispanoamérica.
27. Valle, D. (2018). *Modelo de regresión lineal multivariado aplicado a la relación entre variables químicas del clinker y variables físicas del*

*cemento* (Tesis de maestría). Universidad de San Carlos de Guatemala, Guatemala.

28. Walpole, R., Myers, R., Myers, S. y Yeng, K. (2012). *Probabilidad y estadística para ingeniería y ciencias*. México: Pearson Prentice-Hall.
29. Webster, A. (2000). *Estadística aplicada a los negocios y la economía*. Bogotá, Colombia: McGraw Hill.



# APÉNDICES

## Apéndice 1. Matriz de coherencia

Pregunta de investigación	Objetivo	Conclusión	Recomendación	Resultado
¿Cuál es el nivel de correlación entre las variables regresoras y la variable respuesta?	Hallar el nivel de correlación entre el peso al nacimiento, el peso al destete y el peso de traslado a sitio 3 mediante modelos de regresión lineal.	Se definió el nivel de correlación de la variable dependiente (peso de traslado a sitio 3) respecto a cada una de las variables independientes, y se concluye que la correlación es aceptable. El coeficiente de determinación general del modelo fue del 46 %, esto significa que las variables utilizadas explican el 46 % de la variable dependiente.	Se sugiere evaluar la correlación entre el peso de traslado a sitio 3 y otras variables no contempladas en este estudio como la cantidad de ingesta de calostro.	Las correlaciones para las variables dependientes respecto a la variable independiente son las siguientes: (Peso al nacimiento = .479), (Peso al destete = .540), (Edad al destete = .332) y (Edad al traslado = .332)
¿Los datos cumplen los supuestos de normalidad, homocedasticidad e independencia?	Comprobar normalidad, homocedasticidad e independencia del modelo por medio de los supuestos de los modelos de regresión.	El modelo evaluado cumplió con los supuestos de normalidad, homocedasticidad e independencia, lo que significa que la generación de un modelo de regresión lineal basado en las variables utilizadas es válida.	Se hace necesario realizar el análisis de los residuos del modelo para la verificación del cumplimiento de los supuestos de normalidad, homocedasticidad e independencia.	Para los supuestos evaluados se obtuvo los siguientes resultados: Normalidad Shapiro-Wilk ( $p = .242$ ). El modelo presenta homocedasticidad evaluada mediante el gráfico de residuos versus predichos y se usó la prueba de Breusch-Pagan dando como respuesta que el modelo es homocedástico. Es independiente, ya que se evaluaron los residuos mediante el correlograma y la prueba de Durbin-Watson.

Continuación apéndice 1.

<p>¿Cuál es el modelo que presenta mejores criterios de información (de Akaike y Bayesiano) y mejor R<sup>2</sup> ajustado?</p>	<p>Evaluar el mejor modelo por medio del criterio de información de Akaike, el criterio de información Bayesiano y el R<sup>2</sup> ajustado.</p>	<p>Utilizando los criterios de información de Akaike, Bayesiano y R<sup>2</sup> ajustado, fueron 3 los mejores modelos: El modelo m11, el m12 y el m15, sin embargo dado que el modelo m15 incluye a todas las variables no resulta ser el más parsimonioso, y aplicándole la prueba del factor de inflación de la varianza (VIF) resultó con colinealidad, lo cual lo invalidó y quedó descartado, entre el m11 y el m12 el mejor modelo se decidió que es el m11 debido a que la variable que no comparte con el m12 es la edad al destete mientras que el m12 la variable que no comparte con el m11 es la edad al traslado a sitio 3, esto implica que el modelo m11 sea capaz de estimar el peso de los lechones con anticipación, es decir que resulta ser predictivo, mientras que el modelo m12 resulta ser inferencial pero no predictivo.</p>	<p>Se sugiere someter a los métodos de criterio de información, al modelo de regresión lineal, con el fin de seleccionar el mejor ajuste, para el pronóstico de los pesos de los lechones al momento del traslado a sitio 3, y así tomar decisiones en busca de mejorar la eficiencia productiva.</p>	<p>El mejor modelo fue el modelo m11. Cuyas variables independientes fueron peso al nacimiento, peso al destete y edad al destete. Dicho modelo presento un R<sup>2</sup> ajustado de .421, un criterio AIC de 397.518 y un BIC de 407.821.</p>
<p>¿Cuál es el modelo de regresión lineal con mejor ajuste para la estimación del peso de lechones (<i>Sus scrofa domesticus</i>)?</p>	<p>Estimar el peso de lechones (<i>Sus scrofa domesticus</i>) al momento de su traslado a sitio 3 mediante un modelo de regresión lineal, con el peso al nacimiento y el peso al destete como variables independientes</p>	<p>Se estimó el peso de traslado a sitio 3 de lechones (<i>Sus scrofa domesticus</i>) mediante un modelo de regresión lineal, utilizando las variables regresoras que tuvieron un coeficiente de determinación más alto y que demostraron explicar mejor los datos, estas variables fueron el peso al nacimiento, el peso al destete y la edad al destete.</p>	<p>Evaluar un modelo de proyección del peso de lechones (<i>Sus scrofa domesticus</i>) mediante la utilización de otras variables adicionales que se consideren significativas a la variable de respuesta, durante las ocho semanas posteriores al destete de los lechones, con el fin de desarrollar un modelo con mejor ajuste.</p>	<p>Se obtuvo un modelo de regresión lineal múltiple con características de normalidad, homoscedasticidad e independencia. Con una correlación positiva moderada y bajos criterios de información. El modelo servirá para la estimación del peso de los lechones al momento de traslado al sitio 3.</p>

Fuente: elaboración propia, empleando Microsoft Excel.