

APLICACIÓN DE MODELOS LINEALES MIXTOS PARA ESTUDIAR LA PRODUCCIÓN DE LECHE EN VACAS DE LA RAZA LUCERNA^a

APPLICATION OF LINEAR MIXED MODELS TO STUDY MILK YIELD PER DAY IN COWS OF THE LUCERNA BREED

FREDDY HERNÁNDEZ BARAJAS ^b, LUIS GABRIEL GONZÁLEZ HERRERA ^{c *}

Recibido 22-01-2020, aceptado 02-05-2020, versión final 18-05-2020.

Artículo Investigación

RESUMEN: A Colombia se le considera como un país muy diverso y con múltiples opciones productivas a nivel ganadero. La raza colombiana Lucerna se caracteriza por presentar buenos desempeños productivos y reproductivos en el trópico bajo de Colombia, debido a su buena rusticidad. Existen pocos estudios que permitan identificar las ventajas productivas que se pueden obtener, al utilizar este tipo de animales. El estudio de la producción de leche por día es una herramienta importante para evaluar la eficiencia de un hato lechero; esta variable es afectada por factores genéticos y no genéticos. Entre tanto, la evaluación de Condición Corporal (CC) por ejemplo, ha sido propuesta como herramienta en el manejo nutricional, reproductivo y de salud en hatos bovinos, por ser un buen indicador del estado nutricional de la vaca y por ende, del nivel de desempeño que pueden tener las hembras para producir leche durante la lactancia. En la presente investigación se utilizaron los registros productivos de 197 bovinos colombianos sintéticos de la raza Lucerna y sus cruces, procedentes de una empresa agropecuaria ubicada en el Valle del Cauca, Colombia. La variable producción de leche por día fue modelada en función del componente racial, la edad de la vaca al momento del parto y la condición corporal. Para la construcción de los modelos fue utilizada una estrategia incremental, iniciando con un modelo lineal clásico con efectos fijos, agregando posteriormente efectos aleatorios y terminando con la modelación de la varianza y las correlaciones entre las observaciones para cada animal. Para identificar el modelo con el mejor ajuste, se utilizaron la prueba de razón de verosimilitud y el criterio de información de Akaike. El modelo que mejor se ajustó para explicar la producción de leche por día fue el que incluyó la variable condición corporal, la edad de la vaca al momento del parto y un término cuadrático de la edad. El modelo estimado acompañó bien la evolución de la producción de leche por día a lo largo de las lactancias, adicionalmente, el coeficiente de correlación de Pearson entre el valor real y el valor estimado fue de 0,792. Los supuestos del modelo se cumplieron satisfactoriamente. Las mayores producciones de leche fueron observadas para vacas con condición corporal alta y edades al parto entre los 100 y 150 meses.

PALABRAS CLAVE: Condición corporal; edad de la vaca al parto; modelación estadística; producción de leche por

^aHernández, F. & González-Herrera, L. G. (2020). Aplicación de modelos lineales mixtos para estudiar la producción de leche en vacas de la raza Lucerna. *Rev. Fac. Cienc.*, 9 (2), 23–42. DOI: <https://doi.org/10.15446/rev.fac.cienc.v9n2.84681>

^bPh.D. en Estadística. Profesor Asistente, Facultad de Ciencias, Universidad Nacional de Colombia sede Medellín.

^cPh.D. en Genética y Mejoramiento Animal. Profesor Asociado, Facultad de Ciencias Agrarias, Departamento de Producción Animal, Grupo de Investigación en Biodiversidad y Genética Molecular (BIOGEM), Universidad Nacional de Colombia sede Medellín.

* Autor para correspondencia: lugonzalezhe@unal.edu.co

día.

ABSTRACT: Colombia is considered a very diverse country with multiple productive options at the livestock level. The Colombian race Lucerna is characterized by good productive and reproductive performance in the low tropic of Colombia, due to its good hardiness. There are few studies that allow to identify the productive advantages that can be obtained, when using this type of animals. The study of dairy milk yield is an important tool to evaluate the efficiency of a milk production system; this variable is affected by genetic and non-genetic factors. On the other hand, the evaluation of Body Condition, has been proposed as a tool in the nutritional, reproductive and health management of bovine herds, as it is a good indicator of the nutritional status of the cow and therefore of the performance level that females can have to produce milk during lactation. The productive records of 197 synthetic Colombian cattle of the Lucerna breed and their crosses were used, coming from a farm located in Valle del Cauca, Colombia. The variable dairy milk yield per parturition was modeled according to the racial component, age of the cow at the time of delivery, body condition and the interaction between these last two variables. For the construction of the models used in the description of the evaluated characteristic, an incremental strategy was used, starting with a classical linear model with fixed effects, adding random effects later and ending with the correlation estimate between the correlations for each animal. To identify the model with the best fit, the likelihood ratio test and the Akaike Information Criterion were used. The final model to explain the dairy milk yield variation was the one that included the body condition, the age of the cow at the time of delivery, the interaction between these two variables and a random intercept. The estimated model accompanied well the evolution of milk production per day throughout the eight lactations. The Pearson correlation coefficient between the real value and the estimated value was 0.792. The assumptions of the model were satisfactorily fulfilled. Animals with low body condition had lower daily milk production with increasing age at the time of delivery. The highest milk yields were observed for cows with high body condition and calving ages between 100 and 150 months.

KEYWORDS: Body condition; cow's age at calving; statistical modeling; dairy milk yield.

1. INTRODUCCIÓN

A Colombia se le considera como un país muy diverso y con múltiples opciones productivas, principalmente por la gran variedad de ecosistemas que posee. El país cuenta con siete razas criollas y dos colombianas, las cuales pueden ser usadas en todo tipo de explotación bovina en condiciones de trópico, ya sea carne y/o producción de leche.

Una de las dos razas colombianas es la Lucerna, la cual se formó en la Hacienda Lucerna de Bugalagrande, producto del cruce entre las razas Hartón del Valle (Criolla colombiana), Holstein y Shorthorn lechero, buscando animales adaptados al trópico, con buena rusticidad, aptitud para la producción de leche y mejor conformación para carne que otros bovinos locales (Restrepo, 2014; Giraldo & Viveros, 2015). Las vacas Lucerna presentan parámetros productivos sobresalientes como por ejemplo, producción media anual de leche de 2925 kg y tiempo promedio entre partos de 382,1 días (Restrepo, 2014; Giraldo & Viveros, 2015). Entre tanto, existen pocos estudios que permitan identificar las ventajas productivas que se pueden obtener, al utilizar este tipo de animales en sistemas productivos propios de nuestro país.

Para el caso de sistemas de producción de leche, la producción de leche por día, se convierte en una variable importante a la hora de evaluar la eficiencia del sistema. Entre tanto, múltiples factores genéticos y no genéticos, pueden influenciar el desempeño de las vacas en relación a la producción de leche.

La evaluación de Condición Corporal (CC) por ejemplo, ha sido propuesta como herramienta en el manejo nutricional, reproductivo y de salud en Hatos bovinos, por ser un buen indicador del estado nutricional de la vaca (Hady *et al.*, 1994) y por ende, del nivel de desempeño que pueden tener las hembras para producir leche durante la lactancia. La calificación de la condición corporal en ganado lechero es una medición subjetiva visual y táctil de la cantidad de grasa subcutánea que posee una vaca (Wildman *et al.*, 1982; Edmonson *et al.*, 1989; Ferguson *et al.*, 1994; Saborío-Montero & Sánchez, 2014). La metodología descrita por Wildman *et al.* (1982) para calificar la condición corporal, utiliza una escala de 1 a 5 puntos, con incrementos de 0,25 puntos, en donde una vaca con puntaje de 1 es famélica (emaciada) y una vaca con 5 puntos es obesa.

Según Roche *et al.* (2007), citado por Emam *et al.* (2017), existen evidencias que la condición corporal al momento del parto y la pérdida de la misma a lo largo de la lactancia, afectan la producción de leche. Emam *et al.* (2017) observaron que la producción de leche fue afectada por la condición corporal antes del parto, siendo que vacas que presentaban mejor condición corporal, presentaron un nivel de producción de leche mayor.

Jílek *et al.* (2008) y Mishra *et al.* (2016), mencionan que la pérdida de condición corporal durante el primer mes de lactancia incrementa el número de días abiertos y días para la presentación del primer estro y para el primer servicio después del parto. Además, una baja condición corporal presentada en cualquier momento de la lactancia temprana, está asociada con la prolongación de la actividad ovárica, baja frecuencia de la pulsación de la hormona LH, baja respuesta a la estimulación con gonadotropinas y una consecuente disminución de la competencia funcional de los oocitos (Diskin *et al.*, 2003; Gearhart *et al.*, 1990).

De otro lado, es bien conocido que la edad de la vaca al momento del parto presenta influencia sobre la producción de leche, siendo ese efecto lineal y cuadrático (Cerón *et al.*, 2003) o únicamente lineal (Pérez & Gómez, 2009). Teke & Murat (2013) observaron una menor media de producción de leche total en vacas con edades tempranas al primer parto (menos de 23 meses), comparadas con aquellas que parían con edades más avanzadas. Vacas con edades al primer parto entre 23 y 24,5 meses, presentaban el máximo de producción de leche en el primer parto, siendo que a partir de los 24 meses de edad, la producción de leche total no era afectada por la edad. Un mayor desarrollo del tejido de la glándula mamaria en vacas más tardías al momento del primer parto, sería la explicación de este hallazgo. Los mismos autores observaron además, que vacas con edades al primer parto entre los 23 y 31 meses, eran aquellas que conseguían permanecer por mayor tiempo en el rebaño, ya que este tipo de animales son los que presentan producción de leche superior a lo largo de la vida productiva, lo que hace que el productor decida mantenerlos en el rebaño por más tiempo. Mostert *et al.* (2001) observaron menor producción de leche en vacas jóvenes y en vacas viejas, pero una

mayor producción de leche en vacas con edades intermedias. Los autores verificaron una gran influencia de la edad al parto sobre la producción de leche diaria al inicio de la lactancia. En ese estudio, las vacas con partos superiores al primero produjeron más cantidad de leche que las novillas.

Vacek *et al.* (2015) observaron menor producción de leche en el primer parto, en novillas que presentaron mejor nivel de condición corporal antes del parto, en comparación con novillas de condición corporal baja y media. Los autores también observaron un incremento en el número de servicios por concepción en vacas con condición corporal alta, lo que generalmente se traduce en un largo intervalo entre partos, con consecuente aumento de la edad al parto.

Con base en todo lo mencionado anteriormente, es necesario analizar el efecto que tienen algunos factores propios de los animales, como por ejemplo la edad de la vaca al parto y la condición corporal, sobre la producción de leche diaria, utilizando herramientas estadísticas, de manera que los productores puedan hacer uso de los resultados para mejorar el estatus productivo de sus animales. El objetivo del presente estudio fue el de contribuir con un modelo estadístico apropiado para describir la producción promedio de leche por día por parto en vacas de la raza Lucerna y sus cruces, en función de las características de cada una de las vacas y de información medioambiental. En la sección (2) del presente artículo se caracteriza la población de estudio, se presenta un análisis descriptivo de la base de datos y se hace una introducción a los modelos estadísticos a usar. En la sección (3) se muestra el proceso de la construcción de los modelos y los resultados relevantes. Finalmente en la sección (4) se presentan las conclusiones del estudio.

2. MATERIALES Y MÉTODOS

En esta sección se describe la población de animales en estudio, las variables observadas para estos animales, se muestra un análisis descriptivo de las variables y conceptos básicos sobre los modelos lineales usados para estudiar la relación entre las covariables y la producción de leche por día.

2.1. Población de estudio y variables consideradas

Para llevar a cabo el estudio se utilizaron los registros productivos de bovinos colombianos sintéticos de la raza Lucerna y sus cruces procedentes de la empresa agropecuaria Reserva Natural el Hatico, localizada en el departamento del Valle del Cauca, en el municipio de El Cerrito. La Reserva Natural el Hatico se encuentra ubicada a 3°27' N y 76°32' W, a 960 m.s.n., con temperatura promedio de 24°, precipitación promedio anual de 1300 mm, humedad relativa promedio anual de 80 % y zona de vida correspondiente a bosque muy seco tropical (*bms – T*). El manejo alimenticio de los bovinos se basa en sistemas agroforestales silvopastoriles intensivos, manejados en asociación con especies de gramíneas (*Cynodon plectostachyus*, *Panicum maximun*) y leguminosas arbóreas (Mahecha *et al.*, 2002).

La información de la base de datos fue recolectada desde el año 1970 hasta el año 2016, está formada por siete variables y contiene el seguimiento a 920 vacas que fueron monitoreadas durante varios partos. Para poder observar el comportamiento de la producción de leche por día a lo largo de los partos, se consideraron sólo 197 vacas que cumplieron con el criterio de haber tenido cinco o más partos. El ordeño de las vacas se realiza en forma mecánica una vez al día y las vacas son suplementadas al momento del ordeño con concentrado. Las variables que componen la base de datos son las siguientes:

- Id: número de cuatro cifras para identificar el animal.
- ppd: producción promedio de leche por día en litros por cada animal, para cada lactancia.
- edad: edad del animal en meses.
- parto: número de parto del animal.
- cr: indica el componente racial del animal y puede ser *cr1*, *cr2*, *cr3* o *cr4*.
- cc: variable continua que indica la condición corporal del animal.
- año: año en el cual parió el animal.
- epoca: época en la cual parió el animal.

La variable respuesta ppd corresponde a la producción de leche por día. Cuatro estaciones de parto fueron creadas de acuerdo a la intensidad de lluvias y de sequía que se presenta en la región, las cuatro estaciones son:

- Estación 1, de enero a marzo,
- Estación 2, de abril a junio,
- Estación 3, de julio a septiembre,
- Estación 4, de octubre a diciembre.

Los componentes raciales, fueron agrupados de la siguiente manera: *cr1*: < 50 % Lucerna, *cr2*: 50 – 74 % Lucerna, *cr3*: 75 – 93 % Lucerna y *cr4*: 94 – 100 % Lucerna.

La variable condición corporal es evaluada al momento del parto por el médico veterinario con base en la cantidad de grasa subcutánea presente al nivel de la raíz de la cola, espacios intercostales y en el dorso del animal, con calificaciones variando entre 1 (vaca emaciada) y 5 (vaca obesa). En la Figura 1 se muestra una ilustración presentada por Meléndez (2015) de la escala usada para medir la condición corporal.

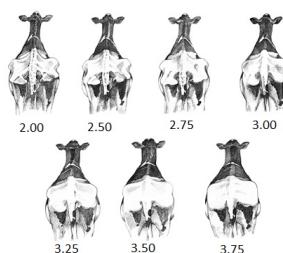


Figura 1: Ilustración de la medición corporal. Fuente Meléndez (2015).

2.2. Análisis descriptivo de la base de datos

En la Figura 2 se muestra el diagrama de barras para el porcentaje de animales que hasta el momento habían presentado cierta cantidad de partos. De las 920 vacas, 31 % tienen un solo parto, 21,6 % de las vacas tienen dos partos y así sucesivamente.

En la Figura 3 se muestra la densidad (usando kernel gaussiano) para la producción de leche por día por animal, diferenciando por el número de parto en el cual se encuentra. Se observa que para los partos 1 y 2 la producción de leche es menor ya que las densidades en color negro y azul están más a la izquierda; a medida que aumenta el parto del animal, la producción de leche se centra alrededor de 11 litros aproximadamente.

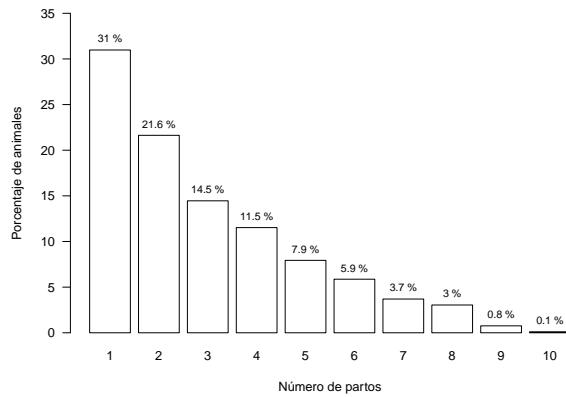


Figura 2: Diagrama de barras para el porcentaje de animales que presentan cierta cantidad de partos. Fuente: Elaboración propia.

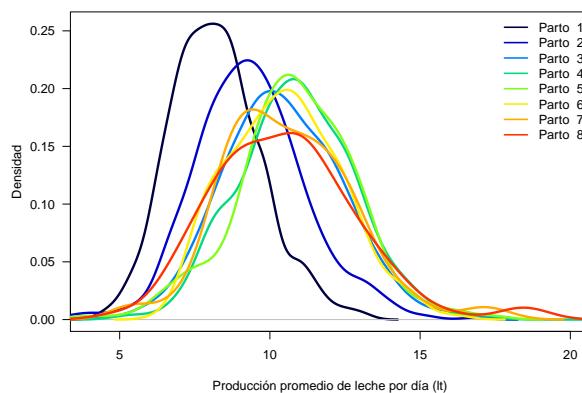


Figura 3: Densidad para la producción de leche promedio por día según el número de parto en el cual se encontraba el animal.

Fuente: Elaboración propia.

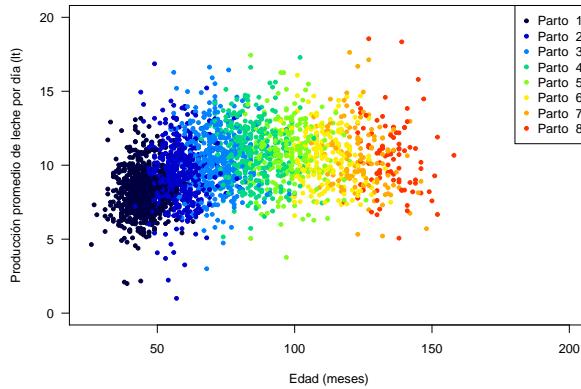


Figura 4: Relación entre la producción promedio de leche por día y la edad en meses diferenciando por parto. Fuente: Elaboración propia.

En la Figura 4 se muestra la relación entre la producción promedio de leche por día en función de la edad del animal y del parto en el que se encuentra. De la figura se observa que las vacas jóvenes (de partos 1 y 2) presentan la menor producción de leche, sin embargo, la producción de leche mejora a medida que la vaca aumenta en edad (aumenta el parto). En la nube de puntos se observa una curvatura con un máximo de producción de leche alrededor de los 100 meses de edad (parto 5).

En la Figura 5 se muestra la relación entre la producción promedio de leche por día en función de la condición corporal del animal y del parto en el que se encuentra. De la figura parece que a medida que aumenta la condición corporal la producción tiende a aumentar.

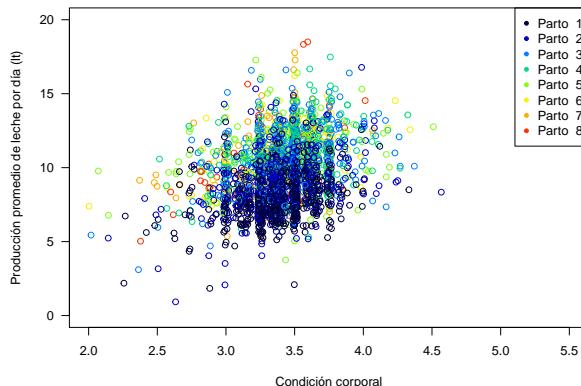


Figura 5: Relación entre la producción promedio de leche por día y la condición corporal del animal diferenciando por parto. Fuente: Elaboración propia.

En la Figura 6 se presenta un diagrama de dispersión que muestra la evolución de la producción de leche promedio por día por parto en función de la edad en meses del animal para 36 de los 197 animales considerados. De esta figura se observan varios patrones en la evolución de la producción de leche. Por ejemplo, para el animal con identificación 3091 se observa un crecimiento en la producción a medida que aumenta la edad, mientras que para los animales 2419 y 2534 se observa una disminución en la producción a medida que aumenta la edad.

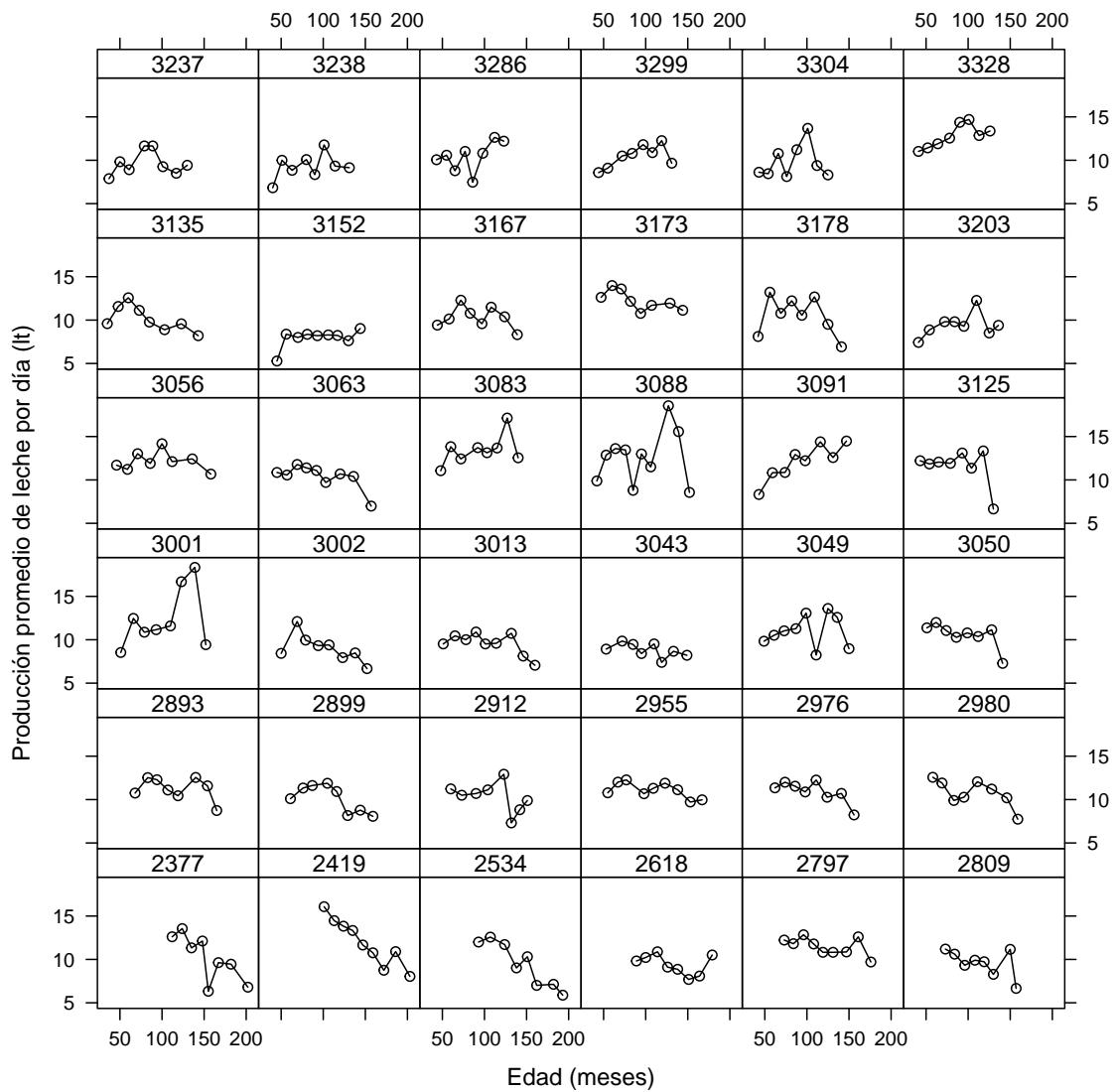


Figura 6: Relación entre la producción promedio de leche por día por parto y la edad en meses diferenciando por animal para 36 de los 197 animales considerados. Fuente: Elaboración propia.

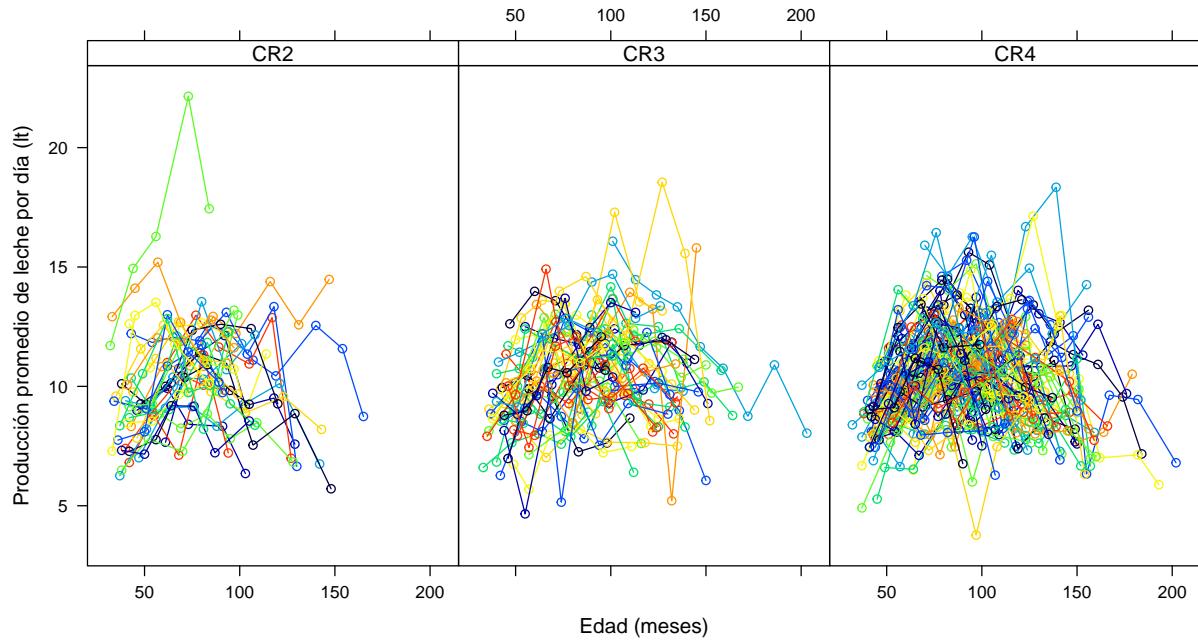


Figura 7: Relación entre la producción promedio de leche por día por parto y la edad en meses para los 197 animales. Fuente: Elaboración propia.

En la Figura 7 se presenta un diagrama de dispersión entre la producción promedio de leche por día por parto y la edad en meses del animal diferenciando por el componente racial de los 197 animales. La figura está compuesta de tres páneles, en el de la izquierda hay 32 líneas que corresponden a los animales con componente racial *cr2*, en los páneles central y derecho hay 55 y 110 animales de componente racial *cr3* y *cr4* respectivamente. En esta figura no se observan evidencias claras de que el componente racial tenga una influencia en la variación de la producción de leche.

En la Figura 8 se muestra el diagrama de dispersión para la producción de leche promedio por día por parto en función de la condición corporal del animal para 36 animales de los 197 que presentaron al menos cinco partos. De este diagrama se observa que para cuatro animales (3002, 3050, 2809 y 3083), las nubes de puntos tienen una orientación con pendiente negativa, mientras que para los 32 animales restantes las nubes de puntos tienen orientación con pendiente positiva, siendo el animal 2980 el que presenta la máxima correlación entre las variables con un valor de 0,90.

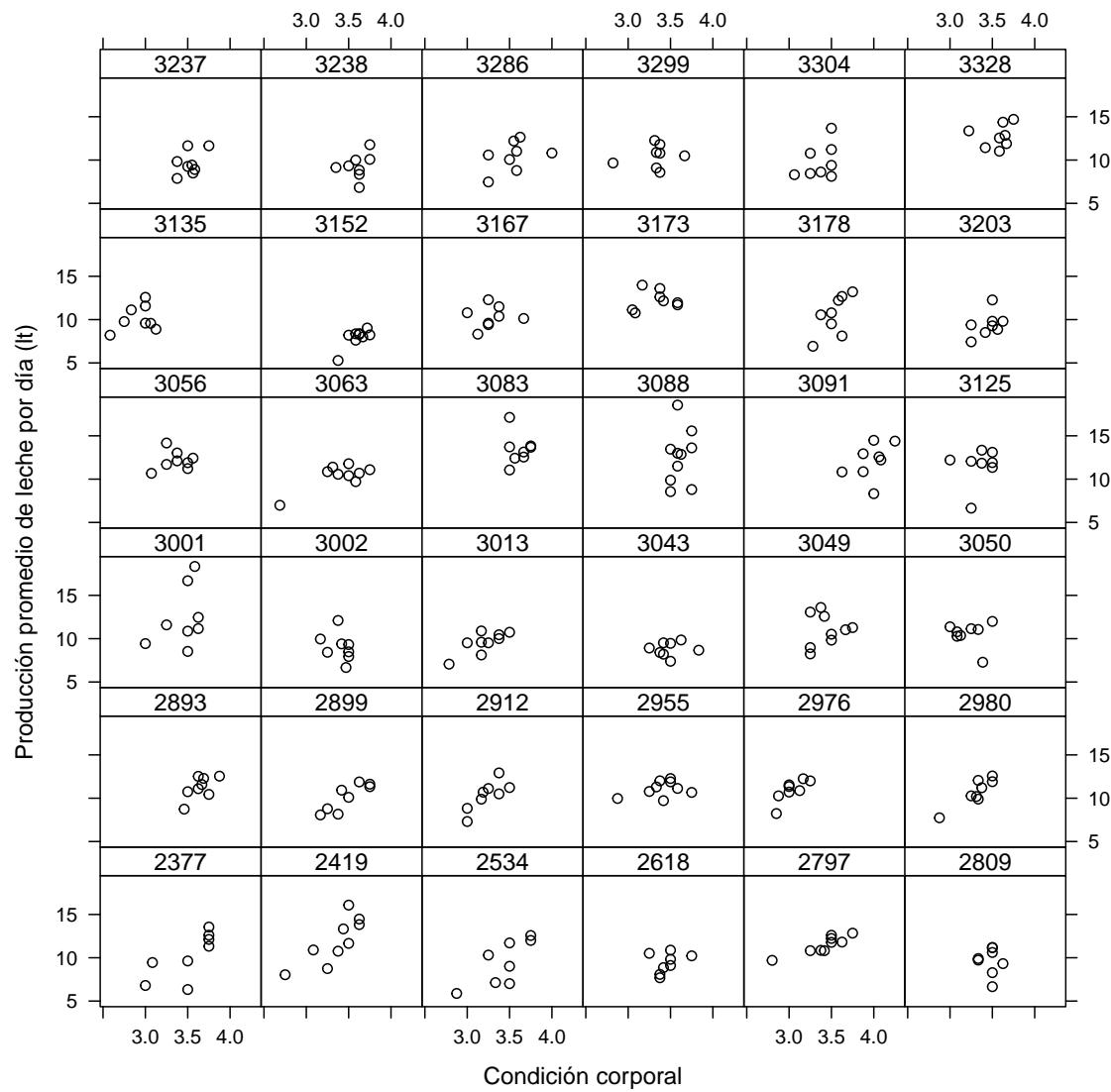


Figura 8: Relación entre la producción promedio de leche por día y la condición corporal para 36 de los 197 animales considerados. Fuente: Elaboración propia.

2.3. Modelos lineales

Uno de los modelos estadísticos más utilizados en la práctica para analizar datos es el modelo lineal clásico (o modelo de regresión múltiple); este modelo tiene sus orígenes a principios del siglo *XIX* gracias a los trabajos realizados por Legendre y Gauss, quienes usaron el modelo lineal clásico para analizar datos sobre astronomía (Stigler, 1990; James *et al.*, 2017).

En el modelo lineal clásico se asume que la variable respuesta Y tiene distribución normal con varianza

constante y media μ_i que depende de las k variables explicativas (covariables) x_1, x_2, \dots, x_k . El modelo lineal clásico se puede resumir por medio de la siguiente expresión:

$$\begin{aligned} Y_i &\stackrel{ind}{\sim} \mathcal{N}(\mu_i, \sigma^2), \\ \mu_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}, \end{aligned} \quad (1)$$

donde el subíndice $i = 1, 2, \dots, n$ representa cada una las observaciones de la muestra, el término β_0 corresponde al valor esperado de Y cuando todas las covariables son nulas, mientras que los restantes coeficientes β asociados a las k covariables representan el efecto fijo marginal de cada una de las covariables sobre el valor promedio de Y .

El modelo lineal clásico (1) se puede escribir en forma vectorial así:

$$\begin{aligned} Y &\sim \mathcal{N}(\mu, \sigma^2 I_n), \\ \mu &= X\beta, \end{aligned} \quad (2)$$

donde $Y = (Y_1, Y_2, \dots, Y_n)^\top$ es el vector de respuestas; X representa la matriz de diseño conocida de dimensión $n \times p$ (siendo $p = k + 1$), la cual tiene en la primera columna unos mientras que las restantes columnas corresponden a las k covariables; el elemento $\beta = (\beta_0, \beta_1, \dots, \beta_k)^\top$ es el vector de coeficientes y I_n es la matriz identidad de $n \times n$. Este modelo lineal clásico también se conoce como modelo de efectos fijos debido a que el vector de efectos β es el mismo para cada una de las observaciones.

El vector de parámetros θ de este modelo está formado por el vector de efectos fijos β y por la varianza σ^2 , es decir, $\theta = (\beta, \sigma^2)^\top$. La estimación del vector β se puede realizar por medio del método de mínimos cuadrados o por el método de máxima verosimilitud (el cual requiere la normalidad de Y) y con ambos métodos se obtiene la misma solución. El estimador de máxima verosimilitud para el vector β está dado por la siguiente expresión:

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y. \quad (3)$$

Usando el vector estimado $\hat{\beta}$ se puede calcular el vector de medias estimado $\hat{\mu} = X\hat{\beta}$ con el cual se obtienen los residuales del modelo $\hat{\epsilon} = Y - \hat{\mu}$. El estimador insesgado de σ^2 está dado por

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}^\top \hat{\epsilon}}{n - (k + 1)} \quad (4)$$

Una suposición del modelo lineal clásico mostrado en la expresión (1), es que las observaciones Y_i son independientes entre sí y en algunas aplicaciones es usual encontrarse con varias mediciones de la variable respuesta Y para un mismo sujeto o grupo i , es decir, que en lugar de tener un solo Y_i , en la base de datos se tendrán n_i observaciones en el vector $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})^\top$ para el sujeto o grupo i . Para este tipo de situaciones ya no es apropiado usar el modelo lineal clásico debido a que hay una dependencia entre las respuestas observadas para un mismo sujeto o grupo. Una de las formas de abordar este tipo de situaciones es por medio de los modelos longitudinales (modelos mixtos).

Los modelos mixtos fueron propuestos por Laird & Ware (1982) y en ellos se asume que existe una relación entre el vector de observaciones Y_i del sujeto o grupo i y las covariables por medio de la siguiente expresión:

$$\begin{aligned} Y_i | b_i &\sim \mathcal{N}(X_i\beta + Z_i b_i, \Sigma_i), \\ b_i &\sim \mathcal{N}(0, D), \end{aligned} \quad (5)$$

donde X_i y Z_i son matrices de diseño conocidas con la información de las covariables, siendo X_i de dimensión $n_i \times p$ y Z_i de dimensión $n_i \times q$. El elemento β representa el vector de efectos fijos general, b_i el vector de efectos aleatorios exclusivos para el grupo i y las matrices.

El vector b_i en la expresión (5) es llamado efecto aleatorio porque éste cambia la media de sujeto a sujeto y su función es la de mejorar el ajuste general dado por el elemento $X_i\beta$ al agregar la cantidad $Z_i b_i$. El modelo dado en la expresión (5) es llamado también modelo mixto porque involucra tanto efectos fijos (β) como efectos aleatorios (b_i).

La distribución marginal de Y_i está dada por:

$$f_i(Y_i) = \int f(Y_i|b_i) f(b_i) db_i \quad (6)$$

donde $f(Y_i|b_i)$ y $f(b_i)$ corresponden a las densidades normales mostradas en la expresión (5). Esta distribución marginal tiene forma cerrada y se puede mostrar fácilmente que la distribución de Y_i es una normal multivariada con vector de medias y matriz de covarianzas como se muestra a continuación:

$$Y_i \sim \mathcal{N}(X_i\beta, V_i), \quad (7)$$

donde $V_i = Z_i D Z_i^\top + \Sigma_i$. El vector de parámetros en este caso es $\theta = (\beta, \alpha)^\top$ donde α consiste de los $q(q+1)/2$ elementos diferentes de la matriz D y todos los elementos de la matriz Σ_i .

3. RESULTADOS

En esta sección se presentan los modelos que se consideraron para explicar la producción de leche promedio por día por parto (*ppd*) de cada animal, en función de las variables condición corporal (*cc*), componente racial (*cr*) y edad en meses del animal (*edad*).

Para la construcción de los modelos se usó una estrategia incremental, se inició con un modelo lineal clásico con efectos fijos, luego se fueron agregando efectos aleatorios para mejorar el modelo, luego se modeló la varianza de la variable respuesta y por último se modeló la correlación entre las observaciones para cada animal. Para la construcción de los modelos se usaron los datos de los animales que presentaron al menos cinco partos, eso significa que se usó información de 197 animales del total de 920 animales.

El modelo (mod0) del cual se inició es un modelo de referencia el cual desconoce la correlación entre las observaciones repetidas de cada animal. Este modelo contiene sólo efectos fijos debido a la condición corporal, a la edad del parto, a la interacción entre condición corporal y edad, términos cuadráticos de la edad (debido a la curvatura observada en la Figura 4) y condición corporal, y por último el efecto del componente racial ($cr2$, $cr3$ y $cr4$). El modelo inicial se muestra a continuación:

$$\begin{aligned} ppd_{ij} &\sim \mathcal{N}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \beta_0 + \beta_1 cc_{ij} + \beta_2 edad_{ij} + \beta_3 cc_{ij}edad_{ij} + \beta_4 cc_{ij} + \beta_5 edad_{ij}^2 + \beta_6 cr3_i + \beta_7 cr4_i. \end{aligned} \quad (8)$$

El subíndice $i = 1, 2, \dots, 197$ representa cada uno de los animales considerados mientras que el subíndice $j = 1, 2, \dots, n_i$ indica las observaciones repetidas para cada animal. Para este modelo se encontró que los efectos del componente racial, el cuadrado de la condición corporal y la interacción entre condición corporal y edad no eran significativos; por esta razón se depuró el modelo inicial eliminando uno a uno los términos y se creó un nuevo modelo reducido (mod1) el cual se muestra a continuación:

$$\begin{aligned} ppd_{ij} &\sim \mathcal{N}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \beta_0 + \beta_1 cc_{ij} + \beta_2 edad_{ij} + \beta_3 edad_{ij}^2. \end{aligned} \quad (9)$$

A partir del modelo anterior se construyó un modelo (mod2) que se caracteriza por tener un intercepto aleatorio denotado por b_{0i} ; el modelo se puede resumir a continuación:

$$\begin{aligned} ppd_{ij} &\sim \mathcal{N}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \beta_0 + \beta_1 cc_{ij} + \beta_2 edad_{ij} + \beta_3 edad_{ij}^2 + b_{0i}, \\ b_{0i} &\sim N(0, \sigma_{b0}^2). \end{aligned} \quad (10)$$

La prueba de razón de verosimilitud fue utilizada para comparar el modelo de efectos fijos (9) con el modelo de intercepto aleatorio (10). La hipótesis nula de la prueba fue $H_0 : \sigma_{b0}^2 = 0$ frente a la hipótesis alterna $H_A : \sigma_{b0}^2 > 0$. El valor P de la prueba fue menor que 0,0001, lo que indica que el modelo con intercepto aleatorio explica mejor la producción promedio de leche diaria.

El siguiente modelo (mod3) generaliza el modelo anterior porque tiene un intercepto aleatorio b_{0i} y una pendiente aleatoria b_{1i} para la edad; el modelo considerado se resume a continuación:

$$\begin{aligned} ppd_{ij} &\sim \mathcal{N}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \beta_0 + \beta_1 cc_{ij} + \beta_2 edad_{ij} + \beta_3 edad_{ij}^2 + b_{0i} + b_{1i} edad_{ij}, \\ \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{b0}^2 & \sigma_{b01} \\ \sigma_{b01} & \sigma_{b1}^2 \end{pmatrix}\right). \end{aligned} \quad (11)$$

La prueba de razón de verosimilitud fue nuevamente utilizada para comparar los modelos dados en las expresiones (10) y (11), en este caso la hipótesis nula de la prueba fue:

$$H_0 : \begin{pmatrix} \sigma_{b0}^2 & 0 \\ 0 & 0 \end{pmatrix},$$

frente a la hipótesis alterna

$$H_A : \begin{pmatrix} \sigma_{b0}^2 & \sigma_{b01} \\ \sigma_{b01} & \sigma_{b1}^2 \end{pmatrix}.$$

El valor P de la prueba fue menor de 0,0001 y, a un nivel de significancia del 5%, se puede decir que el modelo con intercepto y pendiente aleatoria logra explicar mejor la media de la variable respuesta.

El siguiente modelo considerado (mod4) generaliza el modelo anterior al incluir un segundo efecto aleatorio asociado a la condición corporal, este modelo se puede resumir por la siguiente expresión:

$$\begin{aligned} ppd_{ij} &\sim \mathcal{N}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \beta_0 + \beta_1 cc_{ij} + \beta_2 edad_{ij} + \beta_3 edad_{ij}^2 + b_{0i} + b_{1i} edad_{ij} + b_{2i} cc_{ij}, \\ \begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{b0}^2 & \sigma_{b01} & \sigma_{b02} \\ \cdot & \sigma_{b1}^2 & \sigma_{b12} \\ \cdot & \cdot & \sigma_{b2}^2 \end{pmatrix} \right). \end{aligned} \quad (12)$$

Al usar la prueba de razón de verosimilitud para comparar los modelos de las expresiones (11) y (12) se obtuvo un valor P cercano a uno, esto indica que no es necesario incluir la segunda pendiente aleatoria en el modelo.

Siguiendo la estrategia incremental para la creación de modelos se construyeron cuatro submodelos del modelo mod3 dado en la expresión (11), estos modelos se caracterizan porque la desviación σ fue modelada en función de la condición corporal, de la edad y de la media estimada $\hat{\mu}_{ij}$. Los cuatro submodelos del modelo mod3 se pueden resumir como se muestra a continuación:

$$\begin{aligned} \text{mod5} \quad \sigma_{ij} &= \sigma |cc_{ij}|^\delta \\ \text{mod6} \quad \sigma_{ij} &= \sigma |edad_{ij}|^\delta \\ \text{mod7} \quad \sigma_{ij} &= \sigma \hat{\mu}_{ij}^\delta \\ \text{mod8} \quad \sigma_{ij} &= \sigma \mu_{ij} \end{aligned}$$

Al usar la prueba razón de verosimilitud para comparar los modelos anteriores con mod3 hubo un empate entre mod5 y mod8, pero se decidió elegir el modelo mod5 debido a que mod5 presentó el mayor valor de correlación (0,792) entre y_{ij} y \hat{y}_{ij} .

Finalmente, se modeló la correlación entre las observaciones para cada animal, para esto se construyeron otros tres modelos en los cuales se consideró una estructura de correlación simétrica (mod9), una estructura $AR(1)$ (mod10) y una estructura $MA(1)$ (mod11). Al comparar estos tres últimos modelos con el modelo mod5 se decidió tomar como modelo final el modelo mod5 porque este último presentó un menor *AIC* (*Akaike Information Criterion* (Akaike, 1974)) y mayor valor de correlación entre y_{ij} y \hat{y}_{ij} .

A continuación en (13) se presentan las expresiones para el modelo estimado mod5. En la primera línea

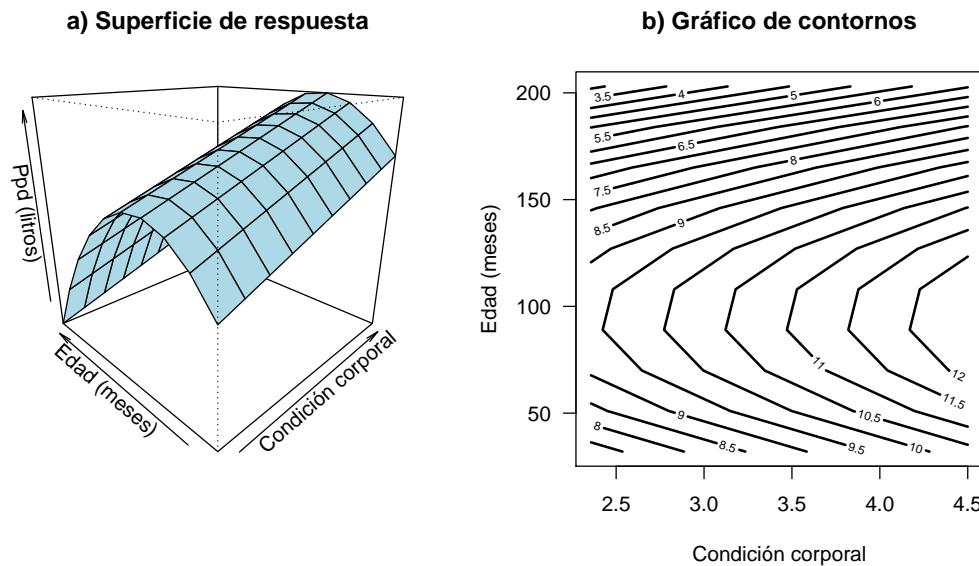


Figura 9: Superficie de respuesta (izquierda) y gráficos de contornos (derecha) para la producción de leche promedio por día estimada (\widehat{ppd}). Fuente: Elaboración propia.

se encuentra la media estimada para la variable ppd del animal i en el momento j ; en la segunda línea se observa la desviación estimada para la variable ppd y en la última línea se tienen las estimaciones para los elementos de la matriz de covarianzas de los efectos aleatorios b_{0i} y b_{1i} mostrados en la expresión (11). Los elementos \tilde{b}_{0i} y \tilde{b}_{1i} son los valores predichos para el intercepto aleatorio y la pendiente aleatoria del i -ésimo animal.

$$\begin{aligned} \widehat{ppd}_{ij} &= 1,073 + 1,431 cc_{ij} + 0,105 edad_{ij} - 0,001 edad_{ij}^2 + \tilde{b}_{0i} + \tilde{b}_{1i} edad_{ij}, \\ \hat{\sigma} &= 0,520 |cc_{ij}|^{0,804}, \\ \hat{\sigma}_{b0} &= 1,670, \quad \hat{\sigma}_{b01} = -0,018, \quad \hat{\sigma}_{b1} = 0,015 \end{aligned} \quad (13)$$

En la Figura 9 se muestra la representación gráfica de la expresión \widehat{ppd} anterior. En el panel izquierdo se tiene una superficie de respuesta para \widehat{ppd} y en el panel derecho el gráfico de contornos para \widehat{ppd} . De esta figura se observa que hay un incremento en \widehat{ppd} a medida que aumenta la edad y la condición corporal. Sin embargo, el efecto cuadrático de la edad conlleva a una disminución en la condición corporal y por ende en \widehat{ppd} .

En la Figura 10 se muestra un diagrama de dispersión entre la producción de leche promedio por día y la edad para los 197 animales considerados en el estudio. La línea de color negro representa la producción de leche estimada \widehat{ppd} dada en la expresión (13). De esta figura se observa que el modelo estimado \widehat{ppd} logra acompañar la evolución de la producción de leche promedio por día. Este resultado se confirma por valor del coeficiente de correlación de Pearson calculado entre ppd y \widehat{ppd} que fue de 0,792.

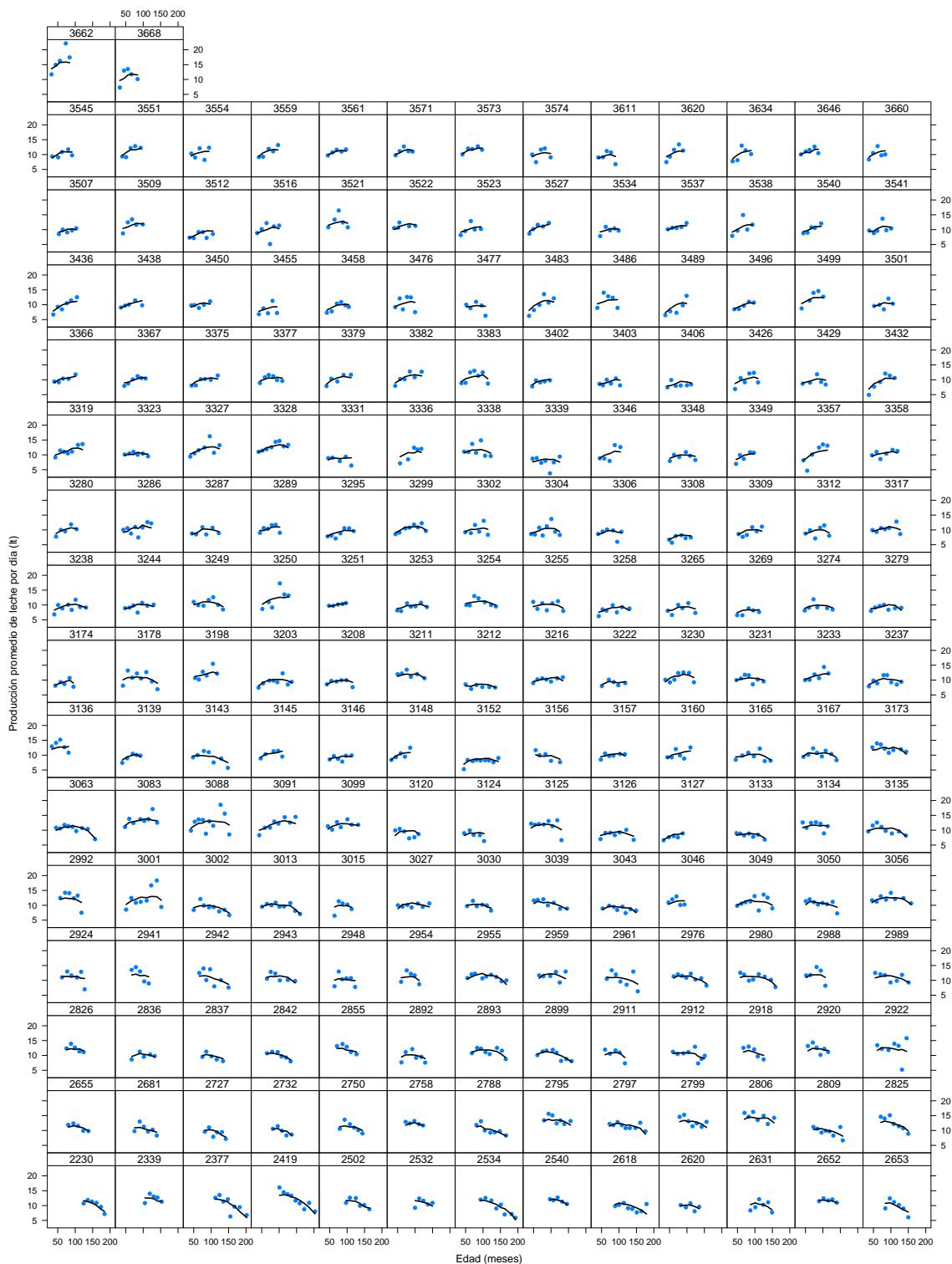


Figura 10: Relación entre la producción promedio de leche por día (ppd) y la edad en meses para los 197 animales. Los puntos azules (●●●) representan los datos observados y la curva (—) representa la estimación de \widehat{ppd} . Fuente: Elaboración propia.

En la Figura 11 se muestran tres gráficos útiles para chequear el cumplimiento de las suposiciones del modelo ajustado. En la parte de arriba a la izquierda se tiene un gráfico de residuales estandarizados versus los valores ajustados \widehat{ppd} ; de este gráfico se observa que la mayoría de los residuales están en el intervalo $(-3, 3)$ y que no hay evidencias de un patrón claro en la nube de puntos. El modelo mod5 fue ajustado asumiendo que $\sigma_{ij} = \sigma|ccij|^{\delta}$ y por lo tanto la desviación de los errores ya tiene incluída la relación con uno de los predictores. En la parte de arriba a la derecha de la Figura 11 se tiene un gráfico de normalidad para los residuales estandarizados; de este gráfico se ve que la mayoría de los puntos están alineados según la línea de referencia y que una pequeña cantidad de los residuales se alejan al inicio y al final. En la parte de abajo están los gráficos de normalidad para los valores predichos \tilde{b}_0 y \tilde{b}_1 , se observa que ambos se ajustan bien a un patrón normal.

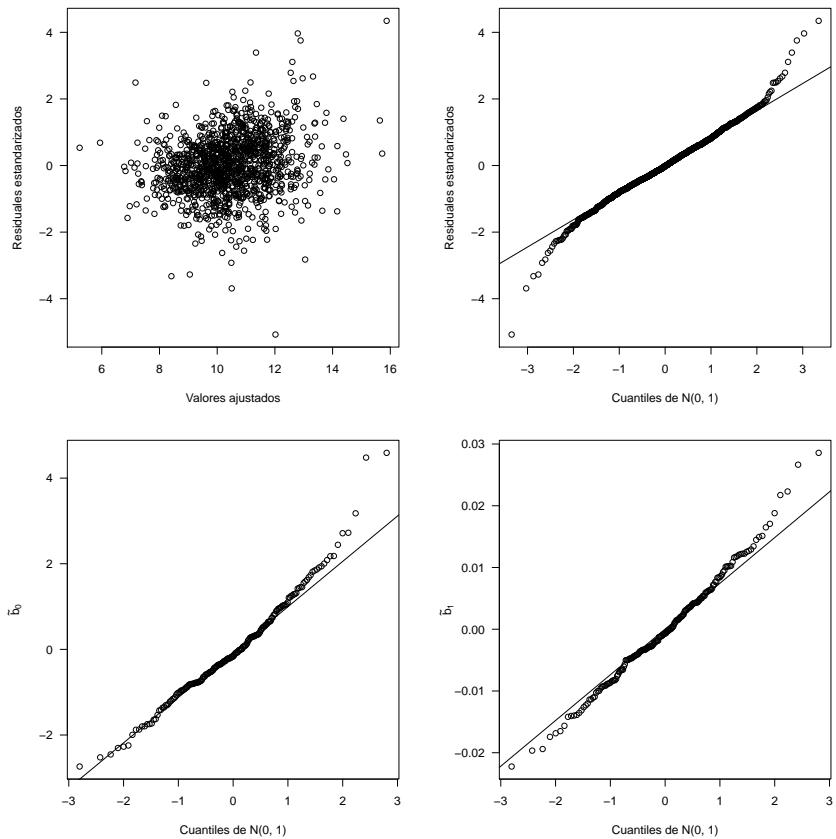


Figura 11: Análisis de residuales. Arriba a la izquierda los residuales estandarizados versus valores ajustados; arriba a la derecha el gráfico de normalidad para los residuales estandarizados; abajo a la izquierda el gráfico de normalidad para los interceptos aleatorios predichos \tilde{b}_0 ; abajo a la derecha el gráfico de normalidad para las pendientes aleatorias predichas \tilde{b}_1 . Fuente: Elaboración propia.

4. CONCLUSIONES

En este artículo se usó un modelo lineal mixto que incluyó efectos fijos y aleatorios para explicar la producción promedio de leche por día para un grupo de animales de la raza Lucerna; las variables iniciales consideradas en el modelo fueron componente racial, condición corporal y edad al momento del parto (en meses).

Del estudio se encontró que sólo las variables condición corporal y edad fueron suficientes (ver expresión (13) y Figura 9) para explicar la producción de leche promedio por día, en vacas Lucerna y sus cruces hasta el orden de parto número 8. Del modelo final se observa que el nivel máximo de producción de leche por día se dá para animales con edades alrededor de 100 meses y condición corporal de 4,5.

Estos resultados pueden ser una herramienta útil para el productor a la hora de evaluar el rendimiento productivo de las vacas e implementar estrategias encaminadas a mejorar la producción de leche diaria durante la lactancia.

Como trabajos futuros se pueden construir otros modelos incluyendo mayor número de animales, mayor número de variables y mayor número de órdenes de parto, con el fin de crear un modelo que tenga en cuenta mayor información. Como en este trabajo se asumió normalidad para la variable respuesta, es posible usar modelos lineales mixtos en los cuales la variable respuesta siga una distribución diferente a la normal.

Referencias

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), 716–723.
- Cerón, M., Tonhati, H., Costa, C., Solarte, C. & Benavides, O. (2003). Factores de ajuste para producción de leche en bovinos Holstein colombiano. *Revista Colombiana de Ciencias Pecuarias*, 16(1), 26–32.
- Diskin, M., Mackey, D., Roche, J. & Sreenan, J. (2003). Effects of nutrition and metabolic status on circulating hormones and ovarian follicle development in cattle. *Animal Reproduction Science*, 78, 345–370.
- Edmonson, A. J., Lean, I. J., Weaver, L. D., Farver, T. & Webster, G. (1989). A body condition scoring chart of Holstein dairy cows. *Journal of Dairy Science*, 72, 68–78.
- Emam, N., Riasi, A., Shahneh, A., Celi, P. & Ghoreishi, S. (2017). Effect of pre-calving body condition score and previous lactation on bcs change, blood metabolites, oxidative stress and milk production in holstein dairy cows. *Italian Journal of Animal Science*, 16 (3), 474–483.

- Ferguson, J. D., Galligan, D. T. & Thomsen, N. (1994). Principal descriptors of body condition score in Holstein cows, *Journal of Dairy Science*, 77, 2695–2703.
- Gearhart, M., Curtis, C., Erb, H., Smith, R., Sniffen, C., Chase, L. & Cooper, M. (1990). Relationship of changes in condition score to cow health in Holsteins. *Journal of Dairy Science*, 73, 3132–3140.
- Giraldo, A. N. & Viveros, A. (2015). Estimación de parámetros de la curva de lactancia en ganado con diferente proporción de Lucerna y su relación con algunos factores de relevancia económica. Tesis de Maestría, Universidad Tecnológica de Pereira, Risaralda, Colombia.
- Hady, W., Domecq, J. & Kaneene, J. (1994). Frequency and precision of body condition scoring in dairy cattle. *Journal of Dairy Science*, 77, 1543–1547.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics). New York, NY: Springer.
- Jílek, F., Pytloun, P., Kubešová, M., Štípková, M., Bouška, J. & Volek, J. (2008). Relationships among body condition score, milk yield and reproduction in Czech Fleckvieh cows. *Czech J. Anim. Sci.*, 9, 357–367.
- Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38 (4), 963–974.
- Mahecha, L., Angulo, J. & Manrique, L. (2002). Predicción del peso vivo a través del perímetro torácico en la raza bovina Lucerna. *Revista Colombiana de Ciencias Pecuarias*, 15 (1), 88-91.
- Meléndez, P. (2015). La condición corporal: Una herramienta práctica para evaluar la nutrición energética del ganado lechero. *Periódico el Mercurio, sección Campo*.
- Mishra, S., Kumari, K. & Dubey, A. (2016). Body Condition Scoring of Dairy Cattle: A Review. *Research & Reviews: Journal of Veterinary Sciences*, 2 (1), 58–65.
- Mostert, B. E., Theron, H. E. & Kanfer, F. H. J. (2001). The effect of calving season and age at calving on production traits of south african dairy cattle. *South African Journal of Animal Science*, 31 (3), 205–214.
- Pérez, G. & Gómez, M. (2009). Factores genéticos y ambientales que afectan el comportamiento productivo de un rebaño pardo suizo en el trópico. 2. Intervalo entre partos y su relación con la producción de leche. *Revista Científica*, 19 (1), 77–83.
- R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Restrepo, A. (2014). Estimación de parámetros genéticos para la stayability y su asociación con otras características de interés económico en la raza colombiana lucerna. Master's thesis, Universidad Tecnológica de Pereira, Risaralda, Colombia.

- Roche, J., MacDonald, K., Burke, C., Lee, J. & Berry, D. (2007). Associations among body condition score, body weight, and reproductive performance in Seasonal-Calving dairy cattle. *Journal of Dairy Science*, 90, 376–391.
- Saborío-Montero, A. & Sánchez, J. (2014). Evaluacióon de la condición corporal en un hato de vacas Jersey en pastoreo en la zona alta de Cartago. *Agronomía Costarricense*, 38, 55–65.
- Stigler, S. M. (1990). The History of Statistics: The Measurement of Uncertainty before 1900. Belknap Press.
- Teke, B. & Murat, H. (2013). Effect of age at first calving on first lactation milk yield, lifetime milk yield and lifetime in turkish holsteins of the mediterranean region in turkey. *Bulgarian Journal of Agricultural Science*, 19 (5), 1126 – 1129.
- Vacek, M., Krpálková, L., Syrůček, J., Štípková, M. & Janecká, M.(2015). Relationships between growth and body condition development during the rearing period and performance in the first three lactations in holstein cows. *Czech Journal of Animal Science*, 60 (9), 417–425.
- Wildman, E., Jones, G., Wagner, P., Boman, R., Trout, H. & Lesch, T. (1982). A dairy cow body condition scoring system and its relationships to standard production characteristics. *Journal of Dairy Science*, 65, 495–501.