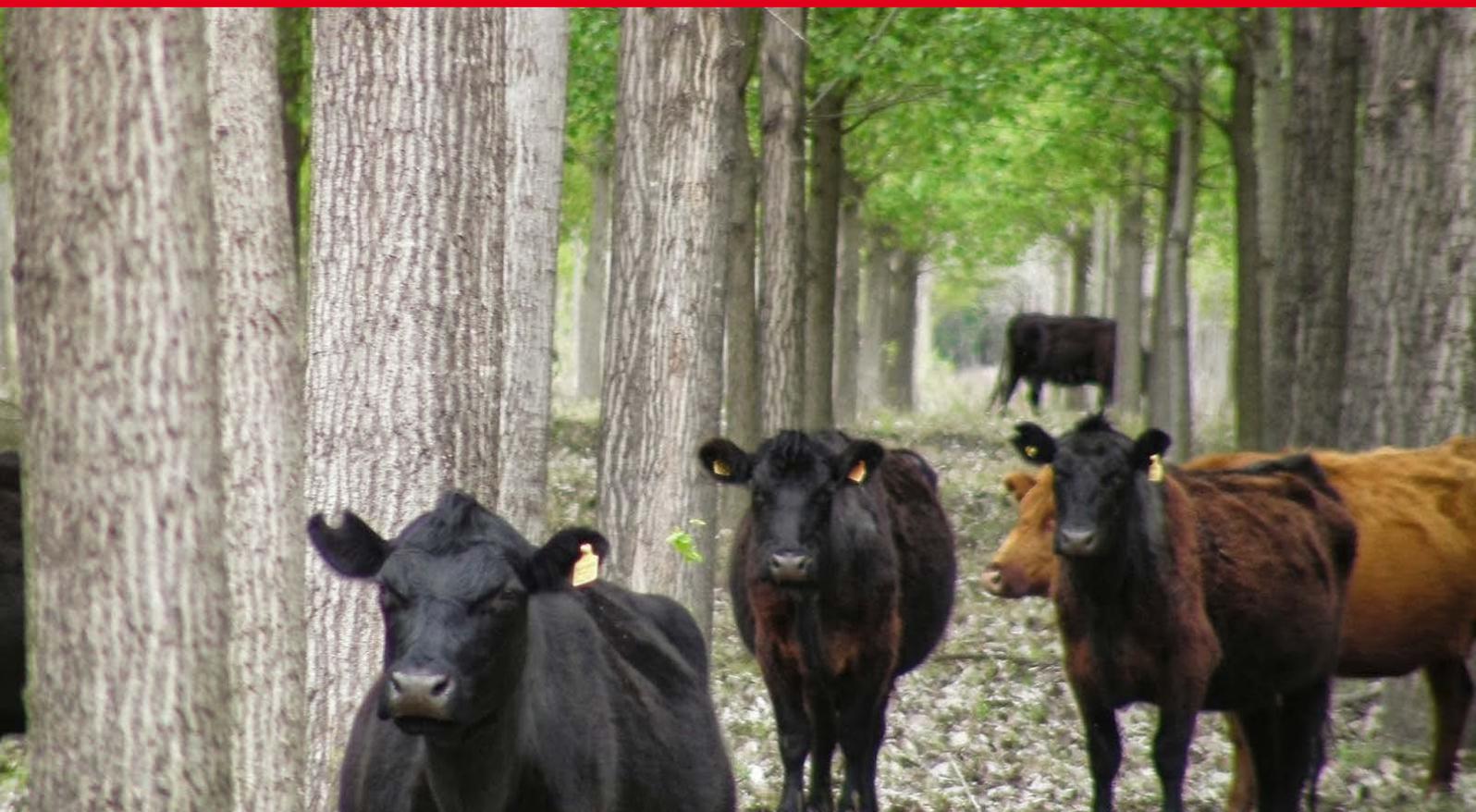




ENERO 2014

Características de la explotación ganadera bovina de Santa Fe

PROVINCIA DE SANTA FE



+ INFO: **Ministro de Economía**
Secretaría de Planificación y Política Económica
www.santafe.gov.ar/ipec



Gobierno de Santa Fe



**SANTA FE
AVANZA**

www.santafe.gov.ar

Señor Gobernador de la Provincia de Santa Fe
DR. Antonio Bonfatti

Ministro de Economía
CPN Ángel José Sciara

Secretario de Planificación y Política Económica
CPN Pablo Andrés Olivares

**Director Provincial del Instituto Provincial
de Estadística y Censos**
LIC. Jorge Alejandro Moore

Equipo Responsable
LIC. Alejandro Anderson
LIC. María Graciela Alvarez

Datos
LIC. Ester Forzani

Corrección
Manuel López de Tejada

Diseño y diagramación
Angelina Araiz



ÍNDICE

P.05	Resumen
P.06	1. Introducción
P.09	2. Resultados
P.11	3. Encuesta ganadera
P.11	3.1 Las variables
P.14	4. Modelo de regresión lineal
P.14	4.1. Linealidad de las variables predictoras y reducción del modelo
P.17	4.2. Análisis de la respuesta; normalidad, varianza en los errores y curvatura
P.22	4.3. Modelo final
P.23	5. Distritos atípicos
P.23	5.1 Detección de los distritos atípicos
P.25	5.2 Visualización de los distritos atípicos
P.27	6. Clasificación de distritos en grupos homogéneos
P.28	6.1 Determinación de la cantidad óptima de grupos
P.30	6.2 Determinación del mejor método de agrupamiento
P.31	6.3 Interpretación de los grupos
P.39	6.4 Etiquetado de los grupos
P.39	6.5 Mapa de ganado bovino por distrito según grupo
P.41	6.6 Separación de los grupos
P.43	6.7 Descripción de los grupos
P.51	7. Anexo
P.56	8. Referencia

ENERO 2014

CARACTERÍSTICAS DE LA EXPLOTACIÓN GANADERA BOVINA DE SANTA FE

PROVINCIA DE SANTA FE

Características de la Explotación Ganadera Bovina. Provincia de Santa Fe 2012

Resumen

Este informe está basado en los datos de la encuesta ganadera de junio del año 2012, desarrollada y realizada por el Instituto Provincial de Estadística y Censos (IPEC), la cual proporciona, entre otras, un total de 41 variables de carácter ganadero bovino. Los objetivos realizados plantean dos temas de interés, desmembrándose a lo largo de todo el trabajo varias conclusiones. Primero se estima la superficie agropecuaria en función de las variables más representativas y después, se agrupan y caracterizan los distritos de la provincia de acuerdo al comportamiento de estas variables. El primer procedimiento es desarrollado mediante el uso del método de regresión lineal múltiple, el cual reduce la base original a un total de 5 variables como las de mayor significancia; con éstas variables se sigue el resto del trabajo. Con el modelo obtenido en la regresión lineal múltiple se puede estimar la cantidad de superficie agropecuaria de cualquier distrito en cualquier año, solamente conociendo estas 5 variables en cuestión. La importancia de esto radica en que se puede estimar la superficie de explotaciones agropecuarias de cualquier distrito o año en el cual el dato sea incierto o falte; sólo basta conocer las 5 variables en cuestión, las cuales se puede conseguir desde otras fuentes. Segundo, considerando los 329 distritos con los que se trabaja, se agrupan de forma de reunir las características ganaderas bovinas similares. De esta manera, se obtienen 9 grupos finales, cada uno con sus características particulares e identificando las variables más influyentes. Por último, se visualiza cada grupo en el mapa con diferentes colores, así se puede reconocer agrupamientos geográficos.



1. Introducción

La cantidad de "Superficie de explotaciones agropecuarias" por distrito de la provincia de Santa Fe es una variable generada por medio de la Encuesta Ganadera de junio del año 2012, que según la superficie de catastro registra un 44,3% de cobertura sobre el total.

El trabajo está fundado sobre los 329 distritos, pertenecientes a la provincia de Santa Fe, que respondieron a la encuesta ganadera del año en cuestión. Aquí se excluyeron los 33 distritos que no contestaron a la encuesta (Anexo- tabla 15). El análisis, inicialmente, se hace respecto a un total de 41 variables, 8 de estas están vinculadas a la cantidad de explotaciones, superficies dedicadas a diferentes sectores y condiciones del suelo. Existen 8 variables que representan la existencia o mortandad del ganado bovino, 12 referentes al tambo, 6 sobre animales dedicados a la invernada y 7 para la cría (Sección 3.1). En cada uno de estos grupos hay variables que se deducen de otras; tales como "total de vacas" que depende de la suma de vacas para cría, invernada y tambo. Con el propósito de utilizar el mínimo de variables que sean a la vez las más representativas, la eliminación de redundancia constituirá en valorar la dependencia directa entre variables para luego ignorar aquellas que deriven en otras. Así, en consideración, la manera de trabajar sólo con totales deja como resultado las variables más representativas de la encuesta; ocho variables representando cada tema influyente en la ganadería bovina de la provincia de Santa Fe.

El modelo de regresión lineal múltiple se desarrolla con estas 8 variables, (si se puede reducir el modelo serán menos) una de estas es la variable "Respuesta" (la "Superficie de explotaciones agropecuarias") y las otras son las predictoras (las variables independientes). Con el fin de verificar las hipótesis de un modelo lineal, primero se dibujan los gráficos de dispersión entre las variables (Gráfico 2), observándose algunas que no cumplen con la relación lineal imprescindible para el modelo. La solución se establece transformando los predictores con el método Box-Cox. Luego se reduce el modelo eliminando variables redundantes con el método de selección de variables basado en AIC, alcanzando un total de 5 variables predictoras finales. Este modelo reducido se compara con el original para comprobar que no haya pérdida de información significativa; esto se hace mediante un f-test, el cual da como resultado que el modelo reducido es tan bueno como el original. Por otra parte, el análisis de las hipótesis del modelo; normalidad de los residuos y varianza constante de los mismos, indican que la primera no se cumple y la segunda no es constante. Para corregir el alejamiento de las hipótesis del modelo se transforma la respuesta finalmente con el método Box-Cox; lo que resuelve las condiciones inadecuadas.

La idea de tener pocas variables que sean representativas con las características ganadera bovina de la provincia de Santa Fe permitirá, al final del trabajo, representar estas características en un mapa de modo sencillo a interpretar. Estas están figuradas por las 5 variables predictoras que quedaron para el modelo, que son:

- 1- Cantidad Explotaciones Agropecuarias.
- 2- Cantidad Animales para Invernada.
- 3- Cantidad Animales para Cría.
- 4- Cantidad de Tambos.
- 5- Mortandad de Vacunos.

Con estas 5 variables se puede estimar la superficie de explotaciones agropecuaria de cualquier distrito de la provincia de Santa Fe de años cercanos al 2012 por medio del modelo lineal al cual llegamos (página 18):

$$Y = 11 + 0.93X_1 + 0.24X_2 + 0.46X_3 - 1.41X_4 + 0.37X_5$$

La variable respuesta, Y, es la variable "Superficie de explotaciones agropecuarias" con su debida transformación.

Las 5 variables predictoras, también transformadas, son X_1, X_2, X_3, X_4, X_5 .

La eficiencia de éste modelo también se prueba para el año 2010 y 2011 con buenos resultados. Por ejemplo; en el año 2010 el mejor de los casos es el distrito de San Lorenzo. El cual declaró 444,6 ha de superficie de explotaciones agropecuarias y nuestro modelo estima que en ese año ese distrito debería haber declarado 441,9 ha de superficie de explotaciones agropecuarias.

El segundo objetivo es agrupar distritos con características ganaderas similares. Caracterizar los grupos obtenidos, buscando las variables que predominan en cada uno, y visualizarlos en un mapa. Para esto, se detectan los datos atípicos multivariados; si no son tratados por separado podrían generar un número reducido de los grupos observados o conducir a resultados incorrectos. Una vez encontrados estos datos atípicos, 88 distritos (Anexo- tabla 16), se forman dos conjuntos, el que está formado por estos distritos y el que no. Luego se aplica el método de agrupamiento en cada uno de estos conjuntos por separado.

El método de agrupamiento "k-media" es utilizado aquí entre los métodos de agrupamiento que logran formar grupos con un alto grado de homogeneidad interna (dentro del grupo) y un alto grado de heterogeneidad externa (entre grupo). Para utilizarlo se necesita previamente conocer la cantidad de grupos para la partición, para esto se utiliza el método de agrupación jerárquica que no requiere hacer inferencias previas sobre el número de grupos y permite representar las sucesivas agrupaciones en forma de árbol (dendograma), con el cual se puede inferir el número de grupos para aplicar K-Medias. Los resultados que derivan de éste análisis son particiones de un tamaño de 5 grupos de distritos para los datos considerados atípicos, y 4 grupos de distritos para los otros datos, o sea 9 grupos en total.

Se caracterizan los grupos con las variables más significativas, y se los clasifican según las ponderaciones que se les atribuyen a las magnitudes de las medias de las variables.

Los grupos se clasifican de la siguiente manera:

Cría	Invernada	Mortandad	Tambo	Expl. Agrop.	Ponderación
grupo 4	grupo 4	grupo 3	grupo 3	grupo 1	Máximo
grupo 5	grupo 5	grupo 4 y 2	grupo 2	grupo 4	Muy bueno
	grupo 3	grupo 5 y 8	grupo 8	grupo 6	Bueno
grupo 9 y 1	grupo 9, 2 y 1			grupo 3, 2 y 5	Moderado
grupo 3 y 6	grupo 8 y 6	grupo 9 y 1	grupo 4 y 5	grupo 8	Poco
grupo 8, 2 y 7	grupo 7	grupo 6 y 7	grupo 9, 6,1 y 7	grupo 9 y 7	Muy poco

Se resume esta información en un nombre o etiqueta que defina aproximadamente la naturaleza del grupo.

Grupo	Etiqueta
1	Máxima cantidad de Explotaciones
2	Muy buena cantidad de Tambos
3	Máxima cantidad de Tambos
4	Máxima cantidad de Cría e Invernada
5	Muy Buena cantidad de Cría e Invernada
6	Buena cantidad de Explotaciones
7	Muy Poca Actividad Bovina
8	Buena cantidad de Tambos
9	Moderada cantidad de Cría e Invernada

Luego, cada grupo es visualizado en un mapa (Mapa 2), de modo que cada distrito sea identificado según el grupo al que pertenezca, con el fin de identificar patrones geográficos.

Finalmente se describe cada grupo con las variables más significativas y con otras incluidas en la base original.



2. Resultados

La obtención de las variables ganaderas bovinas más representativas proviene de las variables más influyentes para ajustar un modelo lineal (página 17) que explica la cantidad de superficie de explotaciones agropecuarias para cada distrito. Las variables que se obtienen son las siguientes:

Cantidad Explotaciones Agropecuarias.

Cantidad Animales para Invernada.

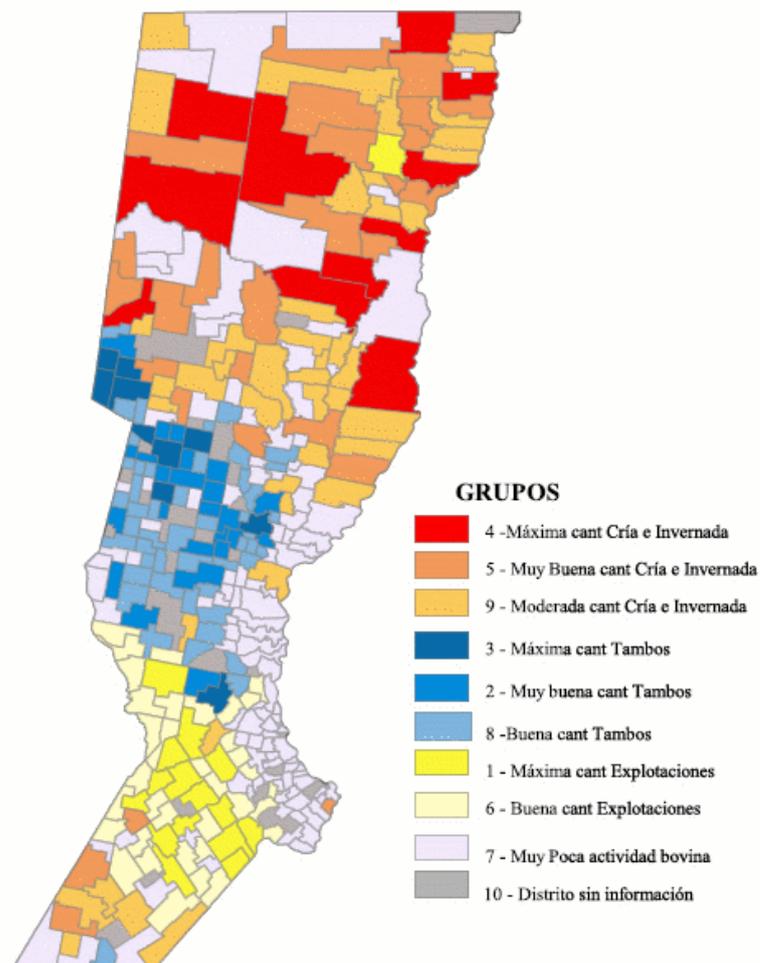
Cantidad Animales para Cría.

Cantidad de Tambos.

Mortandad de Vacunos.

Considerando las 5 variables por cada distrito, se agrupan los mismos en 9 grupos. Al caracterizar, etiquetar y describirlos se obtiene el siguiente mapa:

Características de la Explotación Ganadera Bovina por distrito. Provincia Santa Fe. Año 2012

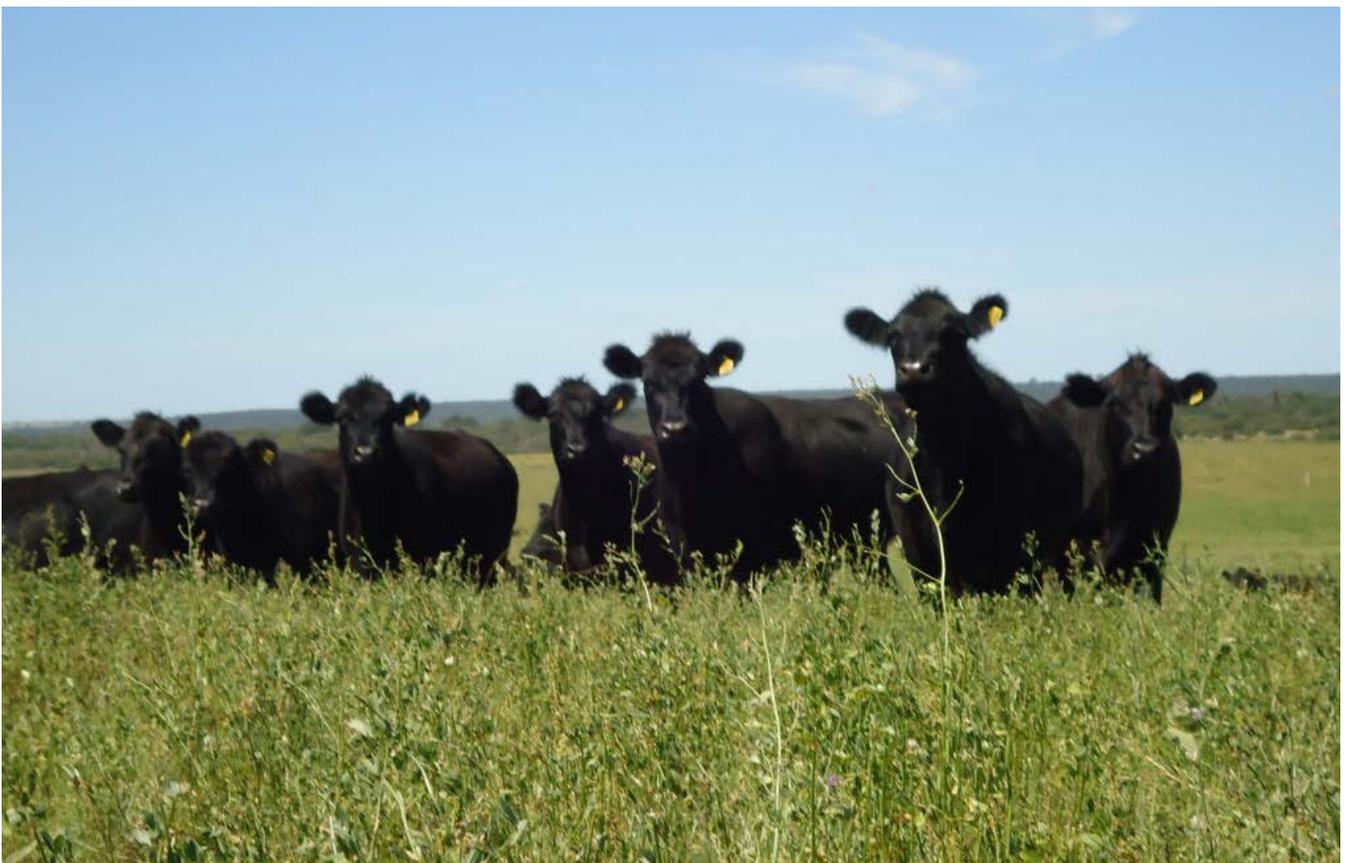


Fuente Datos: Encuesta Ganadera 2012

Fuente Mapa: IPEC

Se observa en el mapa que los grupos de distritos se pueden reagrupar en 4 grandes zonas, dejando de lado unos pocos distritos que geográficamente están fuera de los límites establecidos. La primer zona formada por 3 grupos de cría e invernada localizados en el norte de la provincia, donde se puede observar que los grupos que la forman se ubican de forma concéntrica de acuerdo a la cantidad de animales, estando en el centro la máxima existencia de cría e invernada. Una segunda zona formada por 3 grupos identificados como tambos que se asocian a la mortandad en el centro de la provincia, ordenándose los grupos que la componen de norte a sur de acuerdo a la cantidad de tambos. Una tercera zona integrada por 2 grupos caracterizados por la cantidad de explotaciones dedicadas a la agricultura y ubicada en el sur de la provincia con la misma estructura que la primer zona, el grupo con mayor cantidad de explotaciones en el centro de la zona. Por último la cuarta zona formada por un solo grupo de distritos sin actividad agropecuaria, ubicados en el cordón industrial.

Las características de cada zona y grupo se resumen en la tabla 14 del Anexo.



3. Encuesta ganadera

En el mes de junio de cada año, los productores agropecuarios de la provincia de Santa Fe deben responder a la encuesta ganadera de carácter provincial en las oficinas municipales o comunales a la que corresponda. Esta recaba información sobre explotación de la tierra que esté inscripta en el registro provincial de productores agropecuarios, a saber: destino de la tierra, existencia ganadera, avicultura, apicultura, mortandad, actividad tambera.

En este trabajo se cuenta con la información de la Encuesta Ganadera del año 2012, y solamente se usará las variables pertenecientes a los sectores que se puedan relacionar con el ganado bovino.

3.1 Las Variables

La lista de las variables con las que se cuentan para el trabajo son las siguientes:

- C0: Cantidad Explotaciones Agropecuarias.
- C1: Superficie Explotaciones Agropecuarias.
- C2: Superficie Dedicada a Ganadería.
- C3: Superficie Dedicada a Agricultura.
- C4: Superficie Montes Forestales y Frutales.
- C5: Superficie Granja, Floricultura Horticultura y Otras.
- C6: Superficie de Desperdicio.
- C7: Superficie en Has de Pastoreo de Animales.
- V1: Existencia ganado vacuno.
- CD: Vacas en total.
- CE: Vaquillonas en total.
- CF: Terneros y Terneras en total.
- CG: Toritos en total.
- CH: Toros en total.
- C43: Mortandad de Vacunos Adultos en total.
- C44: Mortandad de Terneros y Terneras en total.
- CA: Cantidad Animales para Tambo.
- C8: Vacas de Ordeñe para tambo.
- C9: Vacas Secas para tambo.
- C10: Terneras y Terneros para tambo.
- C11: Vaquillonas sin Servicio para tambo.
- C12: Vaquillonas con Servicios para tambo.
- C13: Toritos para tambo.
- C14: Toros para tambo.
- C48: Litros Obtenidos el 30/06/2012.
- CI: Cantidad de Tambos.
- C47: Litros Obtenidos del 01/07/11 al 30/06/12.
- C49: Litros Vendidos del 01/07/11 al 30/06/12.
- CB: Cantidad Animales para Invernada.

- C15: Novillos < 300Kgs para invernada.
 C16: Novillos de >300Kgs para invernada.
 C17: Vacas para invernada.
 C18: Vaquillonas (1 - 2) años para invernada.
 C19: Vaquillonas > de 2 años para invernada.
 CC: Cantidad Animales para Cría.
 C20: Vacas para cría.
 C21: Terneros y Terneras para cría.
 C22: Vaquillonas para cría.
 C23: Toritos para cría.
 C24: Toros para cría.
 C25: Bueyes y Torunos para cría.

Para tener en cuenta una previa eliminación de variables hay que considerar las relaciones directas que hay entre estas. Las siguientes sumas deben son consistentes:

$$C1 = C2 + \dots + C7$$

$$V1 = CA + CB + CC$$

$$CA = C8 + \dots + C14$$

$$CB = C15 + \dots + C19$$

$$CC = C20 + \dots + C25$$

$$CD = C8 + C20 + C9 + C17$$

$$CE = C11 + C12 + C18 + C19$$

$$CF = C10 + C21$$

$$CG = C13 + C23$$

$$CH = C14 + C24$$

En estas variables se encuentra la cantidad total de animales para cada sector. El total para tambo (CA), es la suma de la cantidad de vacas, terneros, toros, etcétera, en esa categoría. Sí consideramos los animales en cada sector como representativo para ese tipo, se puede prescindir de las variables C8,..., C14. Con el mismo criterio para animales destinados a invernada y cría, se tiene en cuenta, únicamente, las variables CB y CC, eliminando esta vez C15,..., C19, también C20,..., C25.

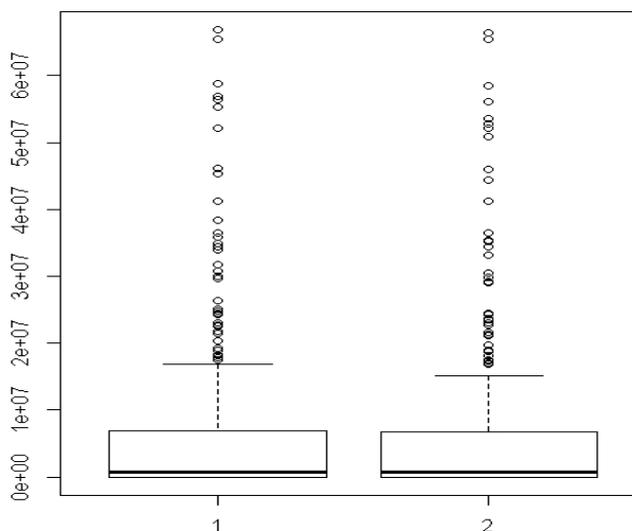
Como la cantidad de animales en total (V1) es la suma de animales dedicados al tambo, cría e internada (CA, CB, CC respectivamente), la información de esta variable es a la suma de la información las tres variables con las que se cuentan para el modelo, y se elimina también. De la misma forma pasa con el total de vacas (CD), vaquillonas (CE), terneros (CF), toritos (CG) y toros (CH).

En el caso de variables de superficie, C1 es el total de superficie de explotaciones agropecuarias, es la suma de las superficies dedicadas a diferentes sectores, como se muestra en las relaciones anteriores. Con la idea de trabajar con el total, se descarta C2,..., C7.

Con el mismo razonamiento anterior, es natural pensar en las variables C43 y C44 (mortalidad de animales adultos y mortalidad de terneros respectivamente) como una suma. Es decir, la mortalidad total; la suma de C43 y C44. A esta nueva variable se la llama CM; $CM = C43 + C44$.

Las variables C47 y C49 (cantidad de litros obtenidos y la cantidad de litros vendidos en el año) son datos similares, como se puede ver en el siguiente gráfico de cajas:

Gráfico 1. Box- plot de la cantidad de litros obtenidos y la cantidad de litros vendidos en el año



Además la correlación entre ambas variables es 0.9997. Están muy ligadas una con otra, por esto se decide trabajar sólo con la cantidad de litros obtenidos C47, que a diferencia de C49, es una variable conceptualmente más vinculada con la producción.

Las variables que quedan son las siguientes:

C1, C0, CA, CB, CC, CI, C47 y CM.

4. Modelo de Regresión Lineal

Se busca un ajuste lineal que modele la relación entre la variable dependiente "Superficie de explotaciones agropecuarias", las siete variables independientes que quedaron en la selección anterior y un término aleatorio ϵ . El modelo expresado es:

$$C1 = a + bC0 + cCA + dCB + eCC + fCI + gC47 + hCM + \epsilon$$

C1: La variable explicada o respuesta

C0, CA, CB, CC, CI, C47, CM: Variables explicativas o predictoras.

Las constantes b, c, d, e, f, g, h: Son los parámetros que miden la influencia que las variables predictoras tienen sobre la respuesta, a es la intersección o término "constante". Y ϵ es una perturbación aleatoria.

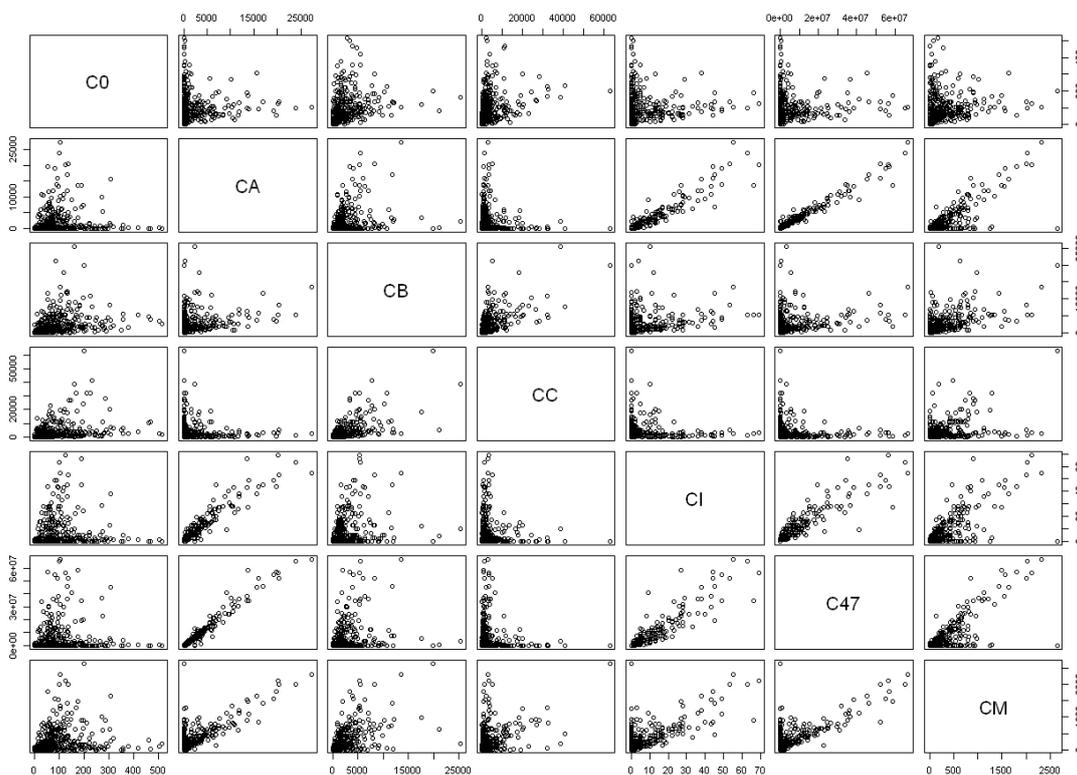
El modelo se representará del siguiente modo:

$$Modelo = modelo\ lineal\ (C1 \sim C0 + CA + CB + CC + CI + C47 + CM)$$

4.1 Linealidad de las variables predictoras y reducción del modelo.

Primero se debe verificar las hipótesis del modelo. Comprobar que la relación entre las variables predictoras sea lineal. La observación del siguiente gráfico, entre todas estas variables, muestra si hay linealidad o no.

Grafico 2. Dispersión entre variables



Se puede ver que hay variables que no guardan relación lineal, por lo que una transformación de las variables predictoras es necesaria. Para esto se utiliza el método Box-Cox. Es una transformación de tipo potencial, que en este caso su función es corregir la no linealidad en la relación (mejorar la correlación entre las variables).

Este método puede usarse solamente con variables positivas. En nuestro caso hay variables que tienen datos iguales a cero, para ellas se suma una unidad para que todos sus datos sean positivos.

Una vez transformadas las variables, se tiene las nuevas predictoras X_1, \dots, X_7 , las cuales se identifican en el siguiente test. El primer modelo a considerar, se muestra a continuación.

Modelo 1:

$$m1 = \text{modelo lineal}(C1 \sim X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7)$$

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-17152.40	2203.38	-7.785	9.71e-14 ***
$X_1=C_0^{0.33}$	884.85	202.31	4.374	1.65e-05 ***
$X_2=CA^{0.14}$	-216.15	379.63	-0.569	0.5695
$X_3=CB^{0.26}$	296.97	131.29	2.262	0.0244 *
$X_4=CC^{0.24}$	746.81	117.34	6.365	6.75e-10 ***
$X_5=CI^{-0.5}$	-8465.54	4937.45	-1.715	0.0874
$X_6=C_47^{0.14}$	66.47	124.90	0.532	0.5950
$X_7=CM^{0.23}$	974.12	231.46	4.209	3.34e-05 ***

Estos tests dicen que las variables X_2, X_5 y X_6 no son importantes para el modelo, si las demás variables. Aunque no asegura que quitarlas sea lo correcto, para verificar se ajusta un segundo modelo únicamente con las variables X_1, X_3, X_4 y X_7 y hacemos un F-test comparando el completo con el reducido.

Modelo 2:

$$m2 = \text{modelo lineal}(C1 \sim X_1 + X_3 + X_4 + X_7)$$

Analysis of Variance Table:

Model 1: $C1 \sim X1+X2+X3+X4+X5+X6+X7$

Model 2: $C1 \sim X1 + X3 + X4 + X7$

Res.Df	RSS		Df	Sum of Sq	F	Pr(>F)
1	321	5.1794e+10				
2	324	5.4847e+10	-3	-3052727866	6.3065	0.0003617 ***

El F-test resuelve un p-value de 0.0003617, por lo tanto rechazamos la hipótesis nula de que los modelos son similares y, el modelo reducido no es tan bueno como el original.

Entonces, para reducir, se decide seleccionar la fórmula basada en el modelo de selección usando el criterio AIC (Selection Using the Akaike Information Criterion):

Después de la sucesión de pasos en el método se elimina del modelo principal las variables $X2$ y $X6$, resultando en un tercero de la siguiente forma:

Modelo 3:

$$m3 = \text{moedlo lineal}(C1 \sim X1 + X3 + X4 + X5 + X7)$$

La comparación de los modelos:

Analysis of Variance Table:

Model 1: $C1 \sim Z$

Model 3: $C1 \sim X1 + X3 + X4 + X5 + X7$

Res.Df	RSS		Df	Sum of Sq	F	Pr(>F)
1	321	5.1794e+10				
3	323	5.1860e+10	-2	-66167126	0.205	0.8147

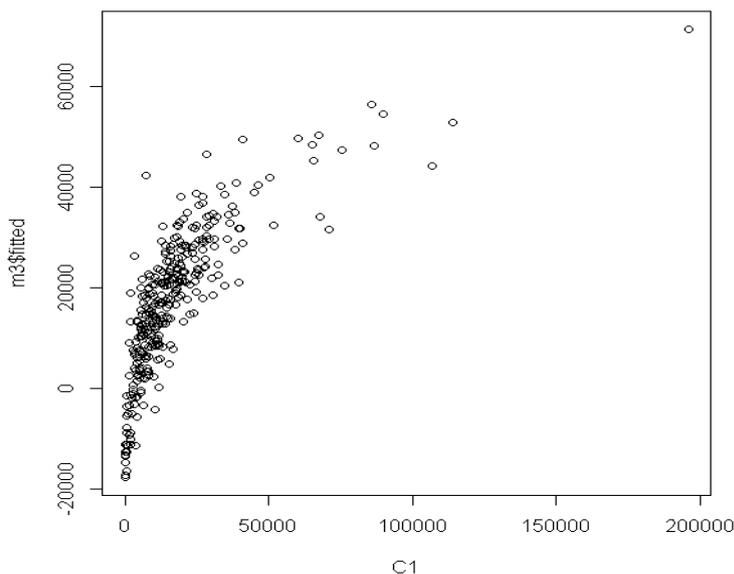
Esta vez el p-value es 0.8147, por lo tanto, no rechaza la suposición que el modelo es tan bueno como el original.

4.2 Análisis de la respuesta; normalidad, varianza en los errores y curvatura.

La normalidad de los errores también es una hipótesis que debe ser comprobada, es decir, asegurarse que los errores sigan una distribución normal, y que la varianza de estos sea constante.

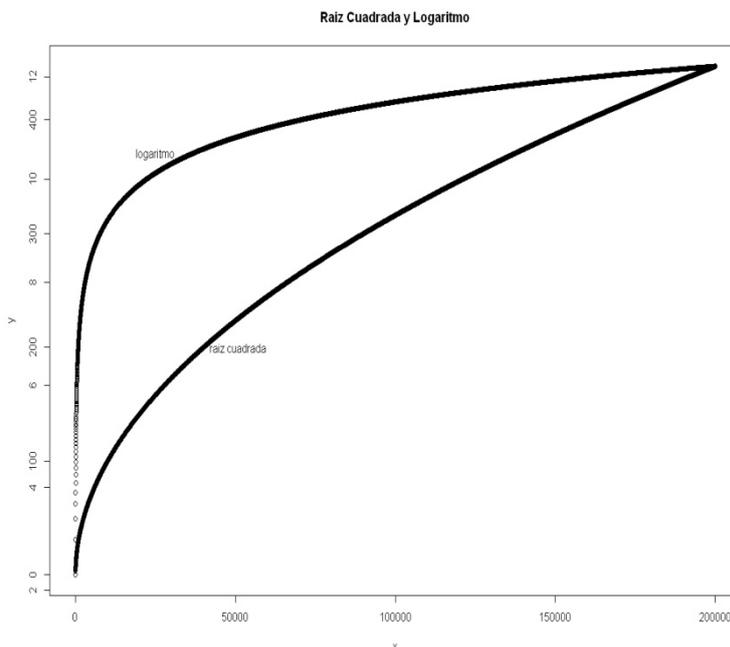
Primero se analiza la respuesta graficando los valores de la variable "superficie de explotaciones agropecuarias" contra los valores ajustados en el modelo 3, para observar alguna tendencia.

Gráfico 3. Variable dependiente versus valores ajustados del modelo 3



De la gráfica anterior, se puede decir que la respuesta necesita transformación. La idea inicial es probar transformar con una función que tenga como gráfico una curva similar a la que se ve en la tendencia. Las siguientes curvas, de las funciones "logaritmo" y "raíz cuadrada" son las candidatas:

Gráfico 4. Logaritmo y raíz cuadrada de la variable superficie de explotaciones

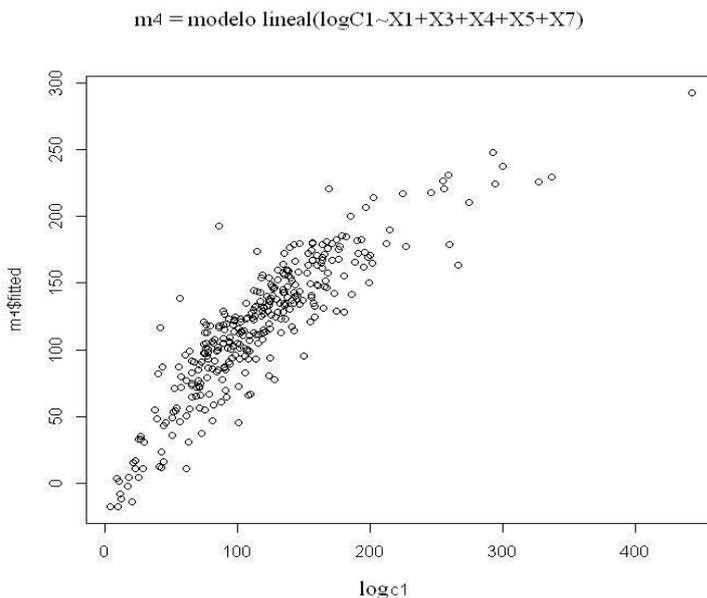


Primero se prueba transformando con la función "logaritmo". El cuarto modelo es el siguiente:

Modelo 4:

$$m4 = \text{modelo lineal}(\log C1 \sim X1 + X3 + X4 + X5 + X7)$$

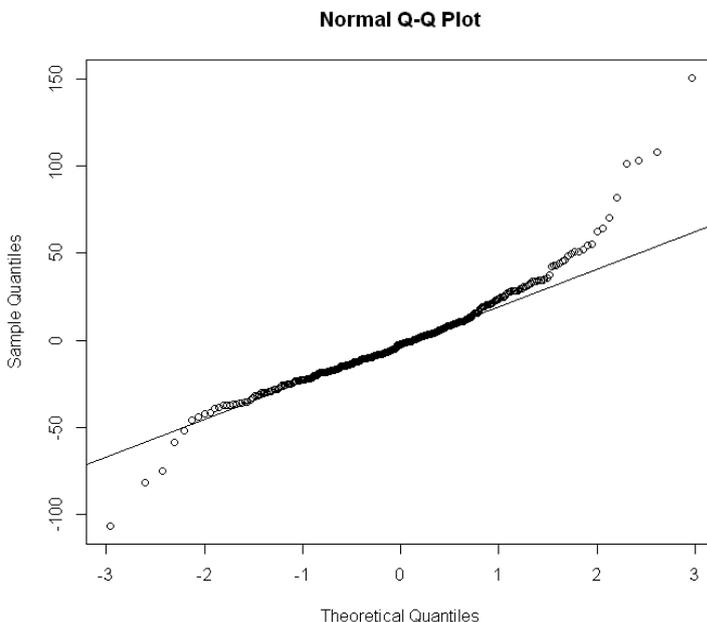
Gráfico 5. Variable dependiente versus valores ajustados del modelo 4



Esta transformación de la respuesta tiene un mejor ajuste lineal.

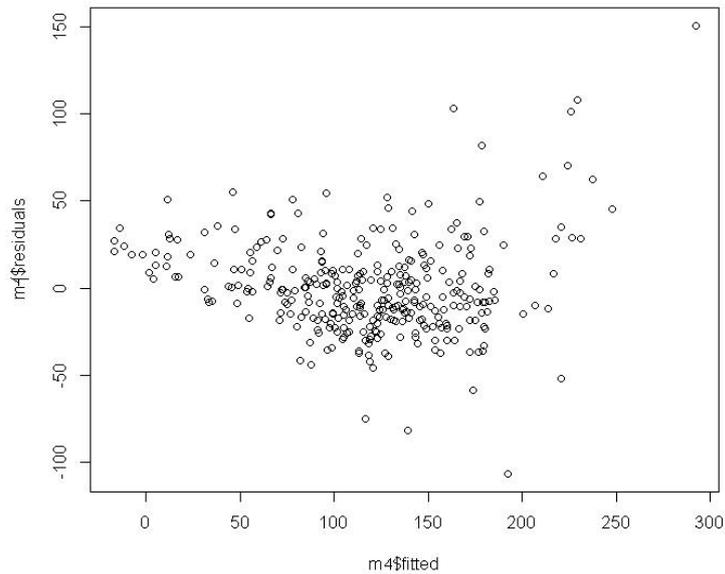
También normaliza los errores según el siguiente gráfico probabilística.

Gráfico 6. Q-Q plot del modelo 4



La varianza de los errores no es constante. Se puede ver del siguiente gráfico que la varianza aparenta no ser constante.

Gráfico 7. Valores ajustados versus residuos del modelo 4



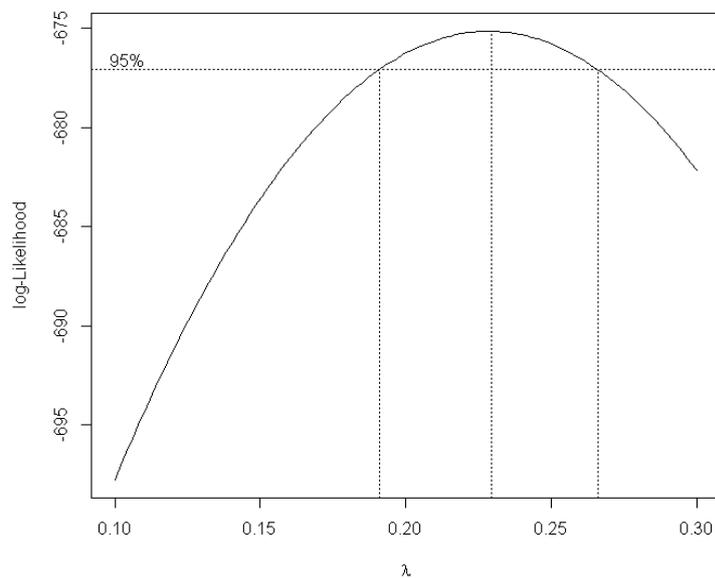
Sucede lo mismo si la respuesta se transforma con la raíz cuadrada.

Se usa el método Box-Cox para transformar la respuesta. La transformación según este método es la siguiente:

$$C_{1tr} = \frac{(C_1^\lambda - 1)}{\lambda}$$

El valor λ puede obtenerse del siguiente gráfico log-likelihood.

Gráfico 8. Log-likelihood de λ



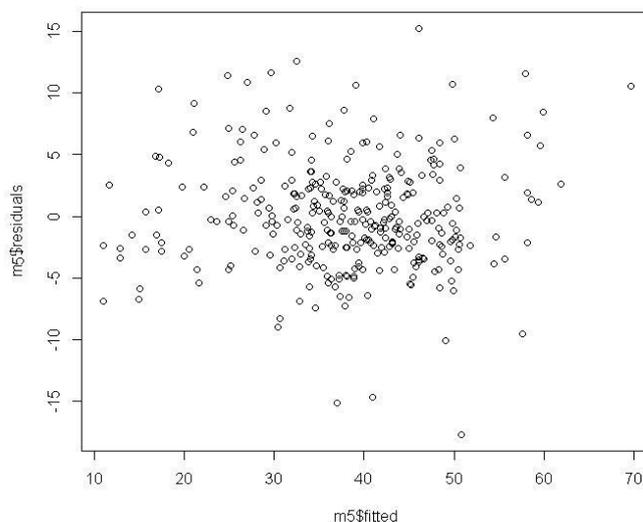
Se elige el valor $\lambda = 0.25$.

Modelo 5:

$$m5 = \text{modelo lineal} \left(\frac{C1^{0.25} - 1}{0.25} \sim X1 + X3 + X4 + X5 + X7 \right)$$

Esta transformación deja constante la varianza de los errores, como aparenta ser en el siguiente gráfico.

Gráfico 9. Valores ajustados versus residuos del modelo 5

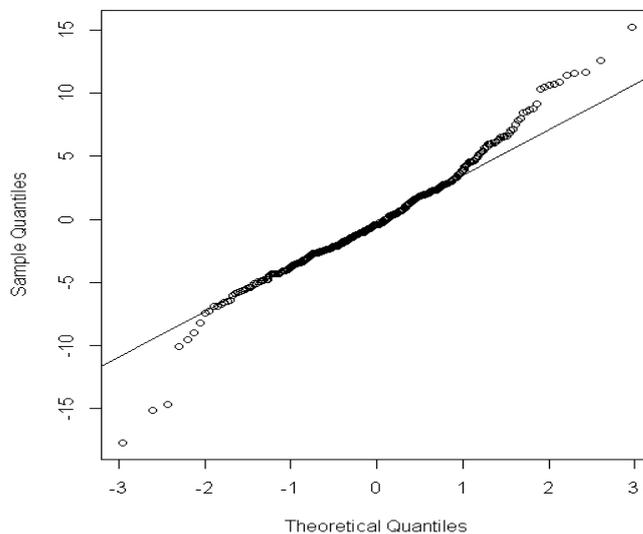


Y como se verifica haciendo el Breusch-Pagan test, el cual da un p-value igual a 0.28.

Además la normalidad de los errores mejora.

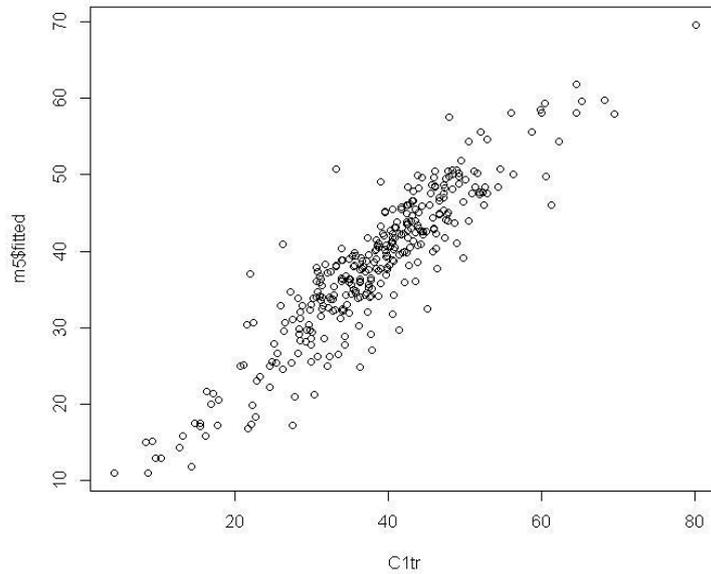
Gráfico 10. Q-Q plot del modelo 5

Normal Q-Q Plot



En gráfico respuesta versus los valores ajustados se observa un mejor comportamiento lineal.

Gráfico 11. Variable dependiente versus valores ajustados del modelo 5



Se puede ver de los dos siguientes gráficos que también mejora la curvatura:

Gráfico 12. Curvatura del modelo 3 por variable

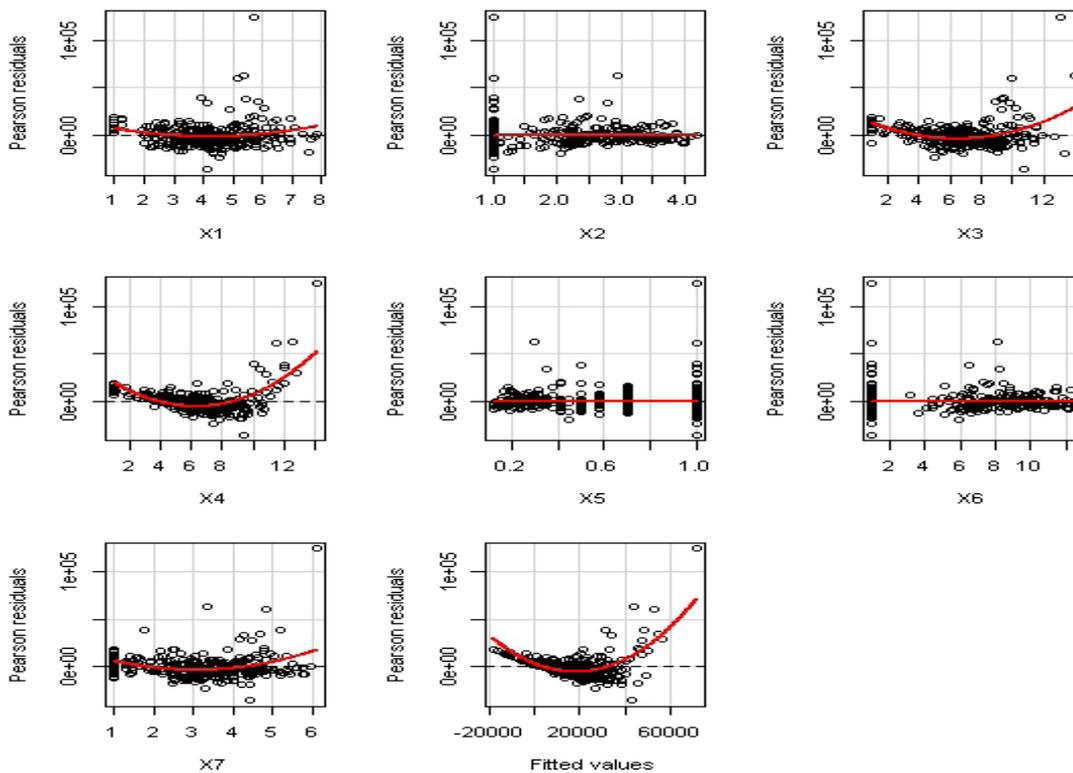
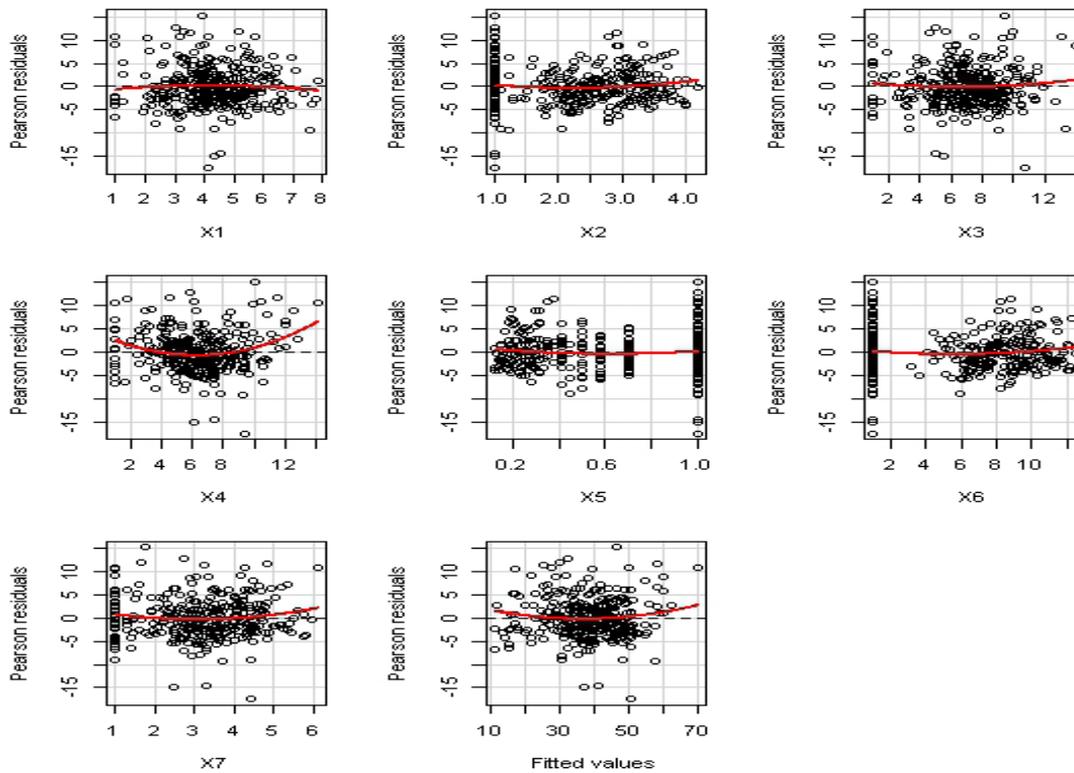


Gráfico 13. Curvatura del modelo 5 por variable



4.3 Modelo final

Se elige el modelo 5, calculado en la sección anterior, como el modelo que mejor ajusta los datos. Este tiene la siguiente forma:

$$E(C1^{tr} | X1, X3, X4, X5, X7) = 11 + 0.93X1 + 0.24X3 + 0.46X4 - 1.41X5 + 0.37X7$$

Reemplazando por las variables originales y después de despejar $C1^{0.25}$ el resultado es:

$$C1^{0.25} = 3.75 + 0.23C0^{0.33} + 0.06CB^{0.26} + 0.12(CC + 1)^{0.24} - 0.35(CI + 1)^{-0.5} + 0.09(CM + 1)^{0.23} + \epsilon$$

5. Distritos atípicos

Los datos atípicos multivariados son observaciones que se consideran extrañas no por el valor que toman en una determinada variable, sino en el conjunto de aquellas. Son mucho más difíciles de identificar que los datos atípicos unidimensionales, dado que no pueden considerarse "valores extremos", como sucede cuando se tiene una única variable bajo estudio. Su presencia tiene efectos todavía más perjudiciales que en el caso unidimensional, porque distorsionan no sólo los valores de la medida de posición (media) o de dispersión (varianza), sino muy especialmente las correlaciones entre las variables. Además pueden distorsionar la formación de los grupos, cuando un elemento se encuentra suficientemente lejos del resto, los demás datos parecen un grupo compacto por comparación. Por otra parte, los datos atípicos pueden ser importantes por sí mismos, porque en algunos casos revelarán aquellos datos que posean un comportamiento excepcional desde un punto de vista multidimensional, por lo que apuntarán a grupos importantes.

5.1 Detección de los distritos atípicos

Una vez determinadas las 5 variables de estudio, se detectan los datos atípicos multivariados. Para esto se utiliza la sentencia `sign1` del paquete `mvoutliers` de "R", donde el cuantil utilizado como valor crítico para la detección de los datos atípicos es 0.99. Este da como resultado 88 distritos identificados como atípicos (Anexo- tabla16).

Estos se pueden ver en los siguientes gráficos de diagnóstico, donde se representa la distancia utilizada por el método atribuida a cada distrito indicando el corte que divide los distritos atípicos (arriba del corte) y los distritos no atípicos (debajo del corte) basados en el cuantil elegido.

Gráfico 14. Distancia por distrito

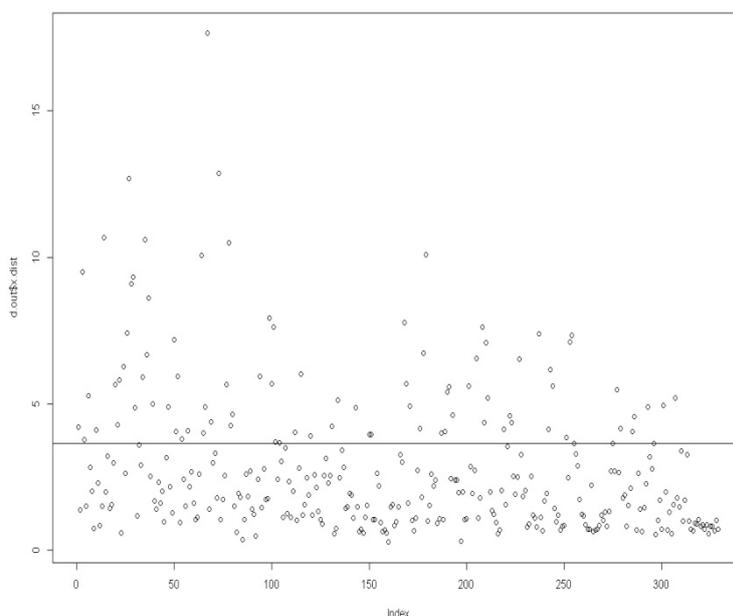
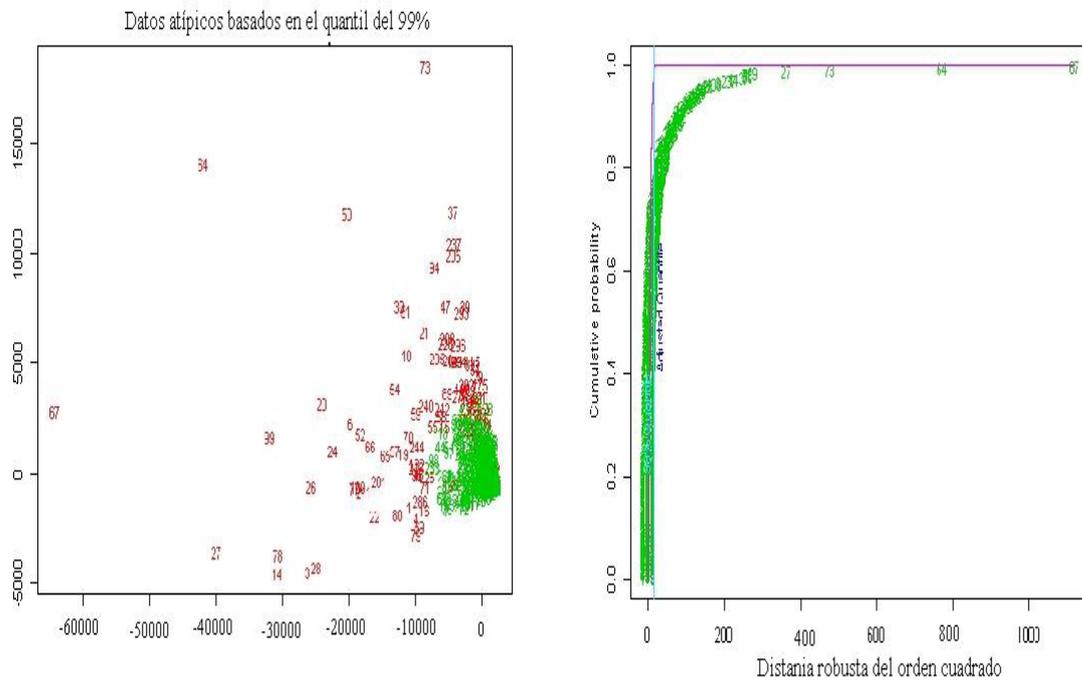


Gráfico 15



Una comparación para los distritos atípicos y los distritos no atípicos muy útil puede brindar la siguiente tabla, donde se usa las medidas descriptivas como comparación, con el corte que divide ambos datos.

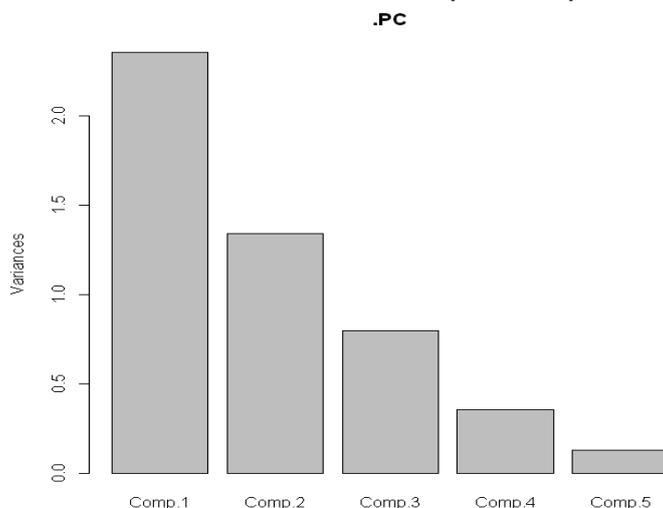
Tabla 1

	Distritos no atípicos				Distritos atípicos			
	Min	Media	Mediana	Max	Min	Media	Mediana	Max
CO	1	80.02	67	281	9	162.7	122	515
CB	0	1709.12	1234	8248	477	5710.75	4591	25397
CC	0	2437.59	1395	12747	0	9659.9	4549	63035
CI	0	4.25	1	26	0	16.71	6.5	69
CM	0	171.93	113	1120	0	714.28	649	2644

5.2 Visualización de los distritos atípicos

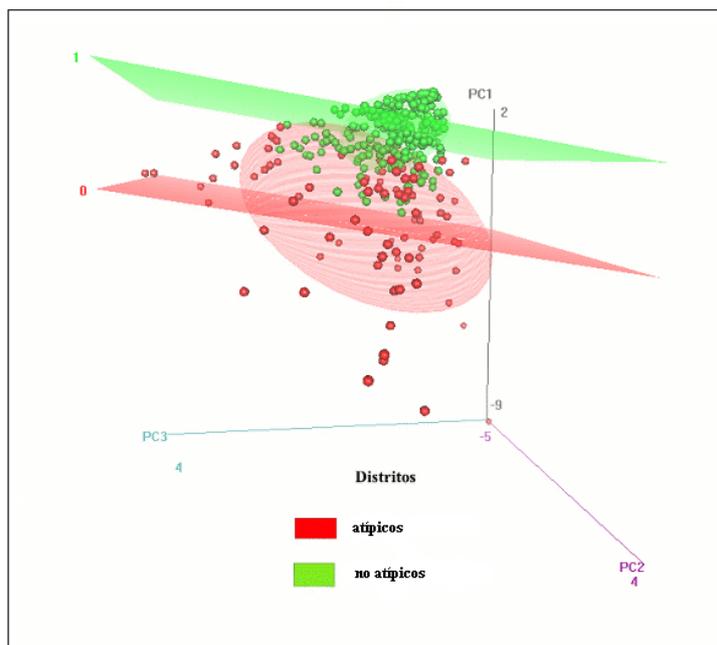
La visualización de los datos en gráficos 3-dimensional logra una síntesis del conocimiento; por lo que es conveniente reducir a tres la cantidad de variables con las que trabajamos mediante el método de Componentes Principales. Este método construye nuevas variables que resultan ser combinaciones lineales de las variables originales. Las tres componentes principales de las variables en estudio, representan el 90,4 % de la variabilidad de la información, sin tener mayor pérdida de la misma.

Gráfico 16. Contribución a la variabilidad por componente principal.



Por lo tanto se representan los distritos como puntos de distintos colores, rojo para distritos atípicos y verde para distritos no atípicos en el espacio tridimensional.

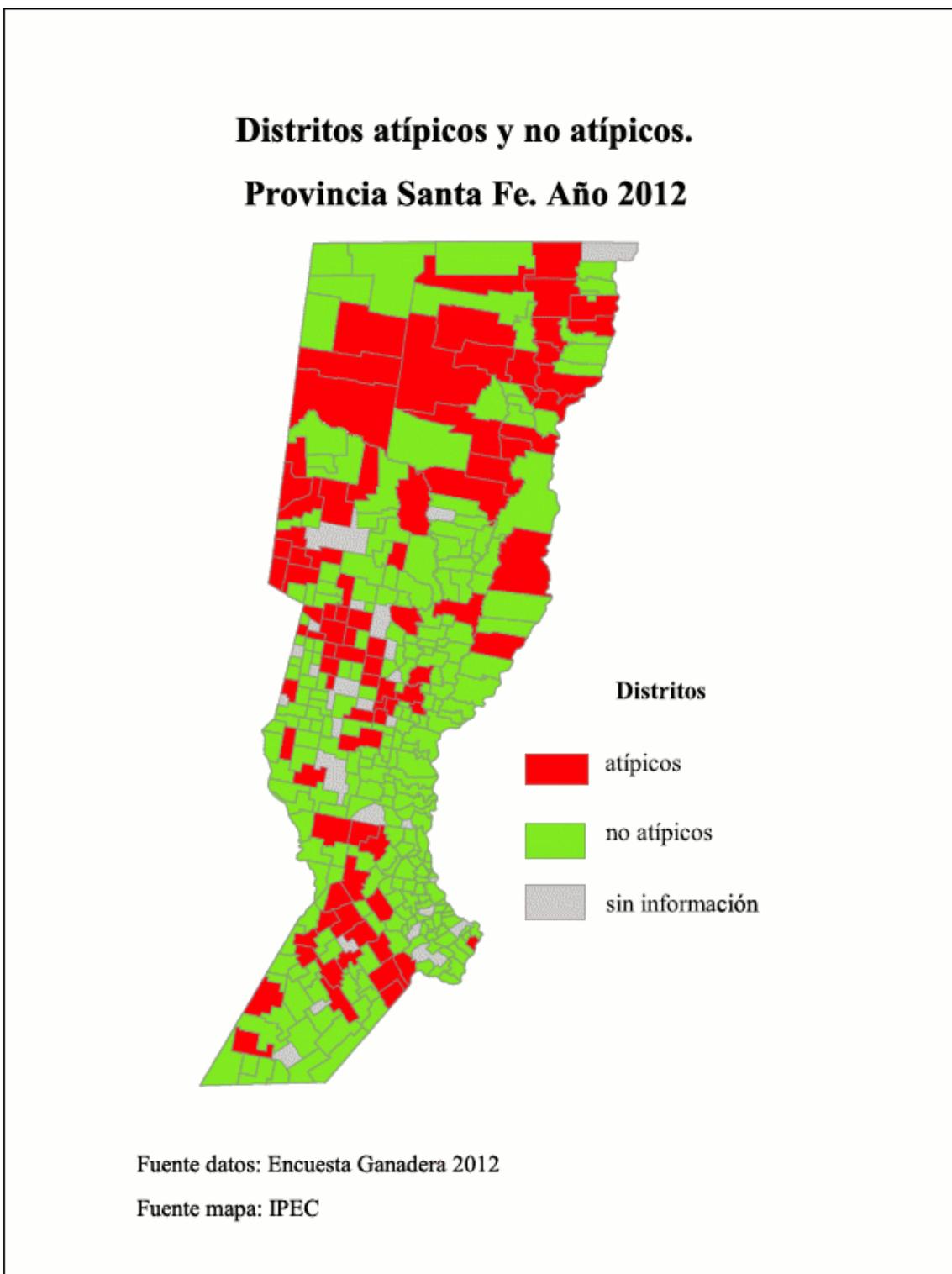
Gráfico 17. Distritos atípicos y no atípicos según las 3 componentes principales.



Se comprueba visualmente una separación entre los distritos no atípicos y los atípicos.

En el siguiente mapa de la provincia se pintan los distritos atípicos en rojo y los distritos no atípicos en verde. Los distritos que no contestaron la encuesta se dejan en color gris. El listado de los distritos atípicos se encuentra en Anexo 3.

Mapa 1



6. Clasificación de distritos en grupos homogéneos

En cada uno de los dos conjuntos formados, uno con los datos atípicos y el otro con los no atípicos, se agrupan los distritos de cada conjunto basándose en las características que poseen. Este agrupamiento clasifica los distritos de forma que cada uno de ellos sea similar, respecto a las variables observadas, a los que pertenecen a su mismo grupo. Por lo tanto los grupos que se forman muestran un alto grado de homogeneidad interna (dentro del grupo) y un alto grado de heterogeneidad externa (entre grupo).

Esto se conoce como análisis de agrupamiento. Los puntos esenciales al desarrollarlo, una vez elegidas las variables de estudio son: elección de una medida de distancia a utilizar en la ejecución del método de agrupamiento, determinación de la cantidad de grupos, elección del método de agrupación y la validación de los resultados.

Como el objetivo es agrupar objetos similares, se necesita alguna medida para evaluar las diferencias y similitudes entre objetos. El concepto de similitud es fundamental en este tipo de análisis, ésta es una medida de correspondencia o semejanza entre objetos que van a ser agrupados. La estrategia más común consiste en medir la equivalencia en términos de la distancia entre los pares de objetos. Los objetos con distancias reducidas entre ellos son más parecidos entre sí que aquellos que tienen distancias mayores y se agruparán dentro del mismo grupo. De esta manera cualquier objeto se puede comparar con cualquier otro objeto a través de la medida de similitud.

Por lo tanto, se selecciona una distancia que aproxime los distritos con semejanza en sus características. Esta es la distancia euclídea, ya que todas las variables consideradas son continuas y están medidas en una misma escala, pero como es sensible a las diferencias de magnitud de los valores de las variables se procede a estandarizar las variables elegidas antes de comenzar con el análisis de agrupamiento.

Tanto para la determinación del número de grupos como del método de agrupación apropiado se procesan diferentes métodos de agrupamiento. En general se distinguen los métodos jerárquicos a los métodos no jerárquicos.

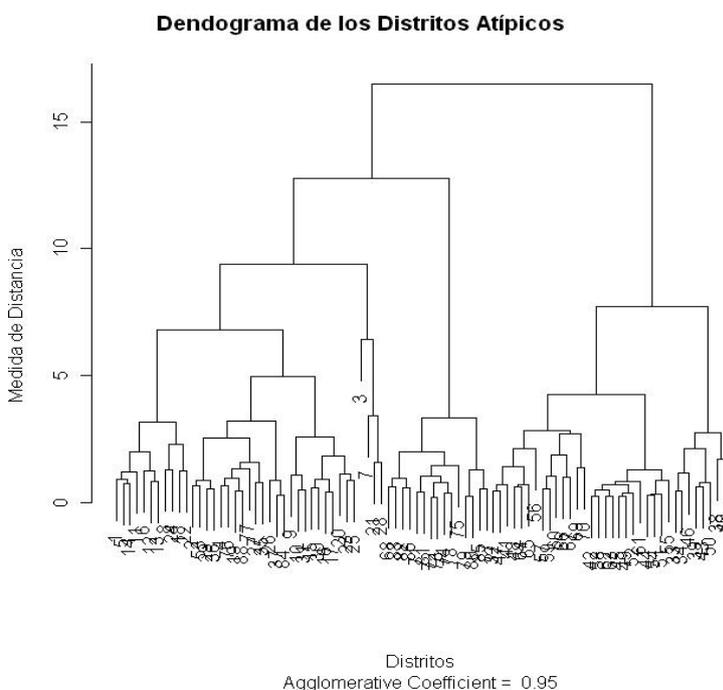
Los métodos jerárquicos se caracterizan por desarrollar una estructura de árbol. Pueden ser por aglomeración o por división. Los métodos por aglomeración empiezan con cada objeto en grupos separados y se van agrupando sucesivamente en grupos hasta que se forma un único grupo. Los divisorios por el contrario parten de un grupo general que se va separando en grupos hasta llegar a grupos con un solo objeto. Estos métodos son útiles para determinar la cantidad de grupos, pero sobrevaloran los casos atípicos y pueden combinar inicialmente grupos no deseables que persisten en el proceso hasta llegar a resultados incorrectos.

Los métodos no jerárquicos, menos susceptibles a la existencia de casos atípicos y a la medida de distancia utilizada, logran optimizar los resultados. Siempre y cuando se utilicen métodos no jerárquicos con los centros de los agrupamientos iniciales de los centros surgidos de los resultados jerárquicos. Por lo tanto se decide combinar ambos tipos de algoritmos en dos etapas: primero determinar la cantidad de grupos mediante un método jerárquico y, una vez definida la mejor solución en cuanto al número de grupos, utilizar un algoritmo no jerárquico.

6.1 Determinación de la cantidad óptima de grupos

Para la determinación de la cantidad apropiada de grupos, se observa el dendograma que surge al aplicar el método de cluster jerárquico ascendente con el método de Ward. El dendograma, representación gráfica del proceso de agrupación, muestra líneas horizontales dónde se unen los grupos. Las posiciones de estas líneas en la escala indican las distancias en que se unen los grupos. Las distancias en las primeras etapas tienen por lo general magnitudes pequeñas similares, mientras que en las últimas etapas las distancias en las que se combinan los grupos son grandes. Al considerar las diferencias entre las distancias de los pasos sucesivos se forman saltos (líneas verticales), los cuales son más pronunciados en las últimas etapas. Entonces se toma como los posibles números de grupos a las cantidades de los mayores saltos. Cabe mencionar que estos mayores saltos son los que se corresponden con los grupos más heterogéneos.

Gráfico 18. Dendograma de los distritos atípicos



El dendograma de distritos atípicos muestra los posibles números de grupos a considerar que están en el rango de 2 a 5.

Tabla 3. Criterio de validación para determinar la cantidad de grupos para los distritos no atípicos

Método	Medida	K	Cantidad de Distritos			Criterio de validación		
						Hubert Gamma	Dunn	WB Ratio
Ward	Euclidea	2	110	131		0.303	0.071	0.746
		3	110	66	65	0.508	0.072	0.568
		4	110	34	32	65	0.578	0.072

En este caso, la mejor opción es la elección de 4 grupos.

6.2 Determinación del mejor método de agrupamiento

A continuación se aplican distintos métodos no jerárquicos con los datos con valores atípicos y con los datos con valores no atípicos, teniendo en cuenta el número óptimo de grupos ya establecido en cada caso. Entre los métodos existentes se utiliza el método k-media, Pam y Clara.

Método k-media: Dado un conjunto de datos y el número de grupos k , el algoritmo genera una partición de los datos con k grupos. El objetivo de k-media es minimizar el error cuadrático medio de todos los grupos. El error cuadrático de un grupo se define como la media de los cuadrados de las diferencias de los objetos de un grupo al centroide del grupo. El centroide de un cluster es el punto medio de todos los objetos del grupo. El procedimiento general es definir los centroides iniciales (dados por el resultado del algoritmo jerárquico seleccionado al determinar el número de grupos) y luego tomar cada punto de la base de datos y situarlo en la clase cuyo centroide sea el más cercano. El próximo paso es recalcular el centroide de cada grupo y volver a distribuir todos los objetos según el centroide más cercano. El proceso se repite hasta cubrir el objetivo donde ya no hay cambios en los grupos de un paso al siguiente.

Método Pam: encuentra k grupos, con un objeto representativo (menoide) en cada uno de ellos. El menoide es el objeto con menor distancia media al resto de los casos del grupo y resulta ser el punto localizado más centradamente en el mismo. A diferencia de los centroides del método k-media los menooides son objetos de la base de datos. Este algoritmo consiste en dos fases. En la primera fase, se obtiene un agrupamiento inicial, seleccionando sucesivamente k objetos representativos. En la segunda fase se intenta mejorar el conjunto de objetos representativos y por lo tanto también mejorar el agrupamiento generado por este conjunto, reemplazando iterativamente un menoide por uno de los no menooides mientras la distancia total del agrupamiento resultante sea mejorada.

Método Clara: Este algoritmo, en lugar de considerar todo el conjunto de datos, elige una muestra como representativa de los mismos y elige los menooides de esta muestra usando PAM. Luego de sacar varias muestras y aplicar PAM en cada una de ellas, da como resultado el mejor de los agrupamientos.

Se debe elegir cual es el método de agrupación más conveniente entre los métodos presentados, donde para la elección se usan los mismos índices que en el caso de la cantidad de grupos.

Tabla 4. Criterio de validación del método a considerar para los distritos atípicos.

Método	Medida	K	Cantidad de Distritos					Criterio de validación		
								Hubert Gamma	Dunn	WB Ratio
Ward	Euclidea	5	15	24	19	15	15	0.466	0.070	0.515
kmeans	Euclidea	5	16	24	10	11	27	0.535	0.121	0.476
pam	Euclidea	5	25	14	19	15	15	0.460	0.072	0.523
clara	Euclidea	5	24	15	25	9	15	0.490	0.063	0.509

Se elige el método k-media.

Tabla 5. Criterio de validación del método a considerar para los distritos no atípicos.

Método	Medida	K	Cantidad de Distritos					Criterio de validación		
								Hubert Gamma	Dunn	WB Ratio
Ward	Euclidea	4	110	34	32	65	0.578	0.072	0.498	
kmeans	Euclidea	4	35	114	49	43	0.630	0.047	0.460	
pam	Euclidea	4	44	101	59	37	0.586	0.049	0.474	
clara	Euclidea	4	46	99	45	51	0.578	0.044	0.474	

Se elige el método k-media.

6.3 Interpretación de los grupos formados

Cada grupo debe ser interpretado y etiquetado de acuerdo a las características y descripción de los mismos. Esta etiqueta debe resumir de la mejor forma la información de cada grupo, describiendo su naturaleza.

Se llama caracterización de cada grupo a la presentación de las variables más significativas de los mismos. Para detectarlas se realiza un t-test que mide la desviación entre los valores relativos pertenecientes al grupo y los valores globales con respecto a todos los distritos. Luego se ordenan de mayor a menor los valores del test; donde los valores más altos son para las variables que más contribuyen al grupo y los valores más bajos para las variables que menos contribuyen. Las variables que no aparecen en la caracterización son las que no tienen una diferencia significativa entre la media del grupo a la media global, estas se dice que no caracterizan al grupo.

La caracterización de cada grupo se puede ver en las siguientes tablas.

Se recuerda que los grupos 1 a 5 están formados por los distritos atípicos, es decir, los distritos en los que las variables tienen valores muy distintos a la media global, en general los valores más altos. Mientras que los grupos 6 a 9 están formados por de los distritos no atípicos, por lo tanto el valor de las variables en este caso son cercanos a la media global.

Tabla 6. Valor test, media global y media grupal por variable significativa de cada grupo**Grupo 1.**

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CO	12,0	360,1	16	102,1
CM	-1,7	149,9	16	317,0
CI	-2,0	1,6	16	7,6

Grupo 2.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CI	10,0	32,0	24	7,6
CM	6,1	809,6	24	317,0
CC	-2,1	1.406,0	24	4.369,4

Grupo 3.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CI	11,6	52,5	10	7,6
CM	11,1	1.734,8	10	317,0
CB	3,8	6.693,0	10	2.779,5
CO	1,1	132,1	10	102,1

Grupo 4.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CC	13,4	32.442,6	11	4.369,4
CB	8,3	10.917,5	11	2.779,5
CM	4,7	896,3	11	317,0
CO	3,9	203,8	11	102,1

Grupo 5.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CB	7,4	7.294,7	27	2.779,5
CC	6,4	12.647,7	27	4.369,4
CM	2,6	511,9	27	317,0
CI	-1,6	4,0	27	7,6

Grupo 6.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CO	6,7	196,7	35	102,1
CB	-1,3	2.103,1	35	2.779,5
CC	-1,7	2.504,1	35	4.369,4
CI	-3,0	1,7	35	7,6
CM	-3,4	96,0	35	317,0

Grupo 7.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CC	-6,1	1.123,0	114	4.369,4
CI	-6,6	1,4	114	7,6
CO	-7,9	49,6	114	102,1
CM	-8,2	62,7	114	317,0
CB	-8,6	634,2	114	2.779,5

Grupo 8.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CI	4,3	14,6	49	7,6
CM	1,8	415,1	49	317,0
CB	-1,2	2.266,4	49	2.779,5
CO	-2,0	79,1	49	102,1
CC	-3,0	1.564,4	49	4.369,4

Grupo 9.

Variable	Valor test	Media del Grupo	Nº de distritos	Media Global
CC	2,5	6.863,6	43	4.369,4
CB	1,8	3.603,2	43	2.779,5
CM	-1,2	246,1	43	317,0
CO	-2,8	66,7	43	102,1
CI	-3,1	2,1	43	7,6

Para el conocimiento de la información de cada grupo, también se calculan las medidas descriptivas de las variables en los grupos y se grafican los box-plot correspondientes.

Tabla 7. Medidas descriptivas de la variable "Cantidad de explotaciones agropecuarias" por grupo.

Variable: C0- Cantidad de Explotaciones Agropecuarias			
Grupos	Media	Desvío estándar	Nº de distritos
1	360,1	91,5	16
2	101,6	57,2	24
3	132,1	70,8	10
4	203,8	56,2	11
5	94,6	40,2	27
6	196,7	41,7	35
7	49,6	35,8	114
8	79,1	36,0	49
9	66,7	43,2	43

Gráfico 20. Box-Plot de la variable "Cantidad de explotaciones agropecuarias" por grupo.

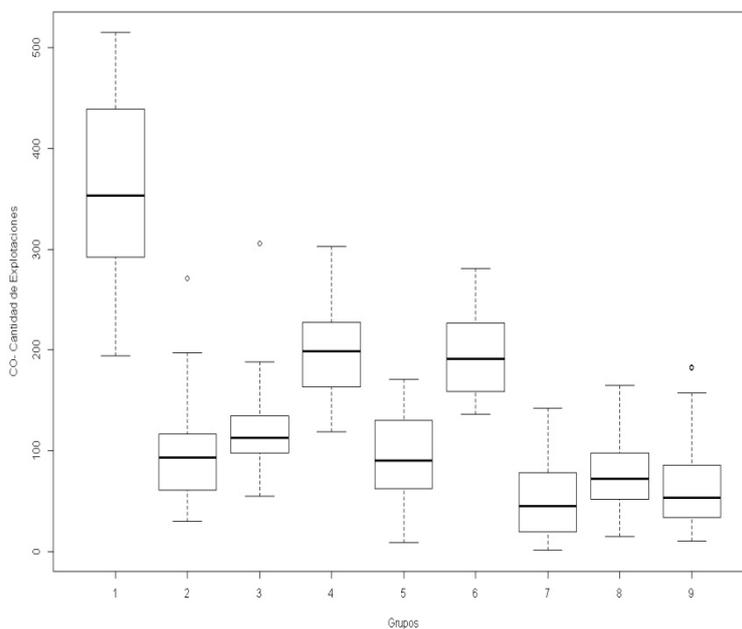


Tabla 8. Medidas descriptivas de la variable "Cantidad de animales para invernada" por grupo.

Variable: CB- Cantidad de animales para invernada			
Grupos	Media	Desvío estándar	Nº de distritos
1	2.807,6	1.678,9	16
2	3.068,6	1.628,4	24
3	6.693,0	3.436,3	10
4	10.917,5	7.081,1	11
5	7.294,7	4.274,6	27
6	2.103,1	1.438,2	35
7	634,2	657,6	114
8	2.266,4	1.378,0	49
9	3.603,2	1.846,4	43

Gráfico 21. Box-Plot de la variable "Cantidad de animales para invernada" por grupo.

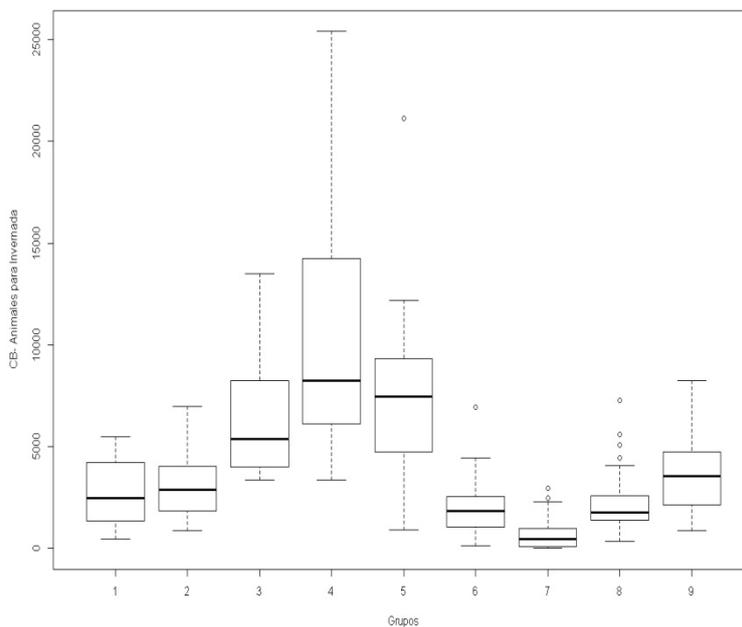


Tabla 9. Medidas descriptivas de la variable "Cantidad de animales para cría" por grupo.

Variable: CC- Cantidad de animales para cría			
Grupos	Media	Desvío estándar	Nº de distritos
1	5.774,1	3.529,0	16
2	1.406,0	1.243,5	24
3	2.558,6	1.293,4	10
4	32.442,6	12.343,0	11
5	12.647,7	6.630,0	27
6	2.504,1	1.527,6	35
7	1.123,0	1.253,9	114
8	1.564,4	1.584,7	49
9	6.863,6	2.616,2	43

Gráfico 22. Box-Plot de la variable "Cantidad de animales para cría" por grupo.

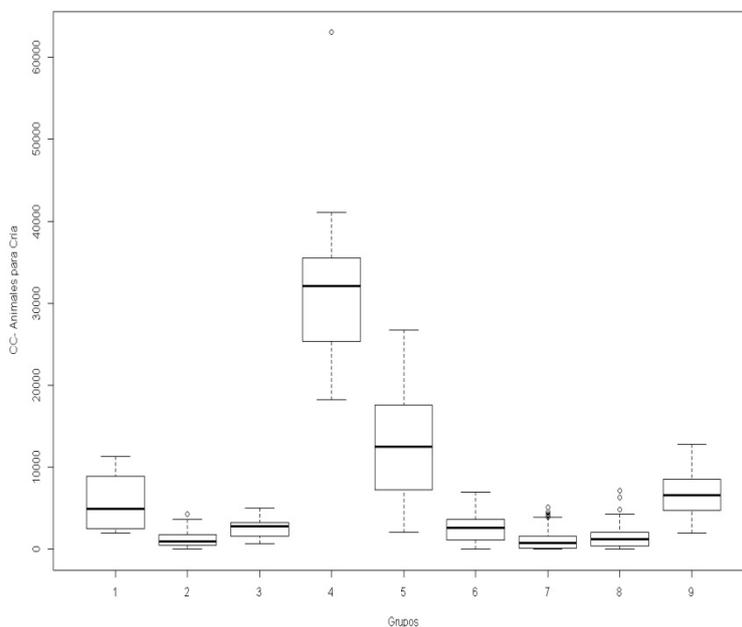


Tabla 10. Medidas descriptivas de la variable "Cantidad de Tambos" por grupo.

Variable: CI- Cantidad de Tambos			
Grupos	Media	Desvío estándar	Nº de distritos
1	1,6	2,2	16
2	32,0	7,5	24
3	52,5	10,6	10
4	4,0	5,2	11
5	4,0	5,8	27
6	1,7	2,0	35
7	1,4	2,4	114
8	14,6	4,5	49
9	2,1	2,8	43

Gráfico 23. Box-Plot de la variable "Cantidad de Tambos" por grupo.

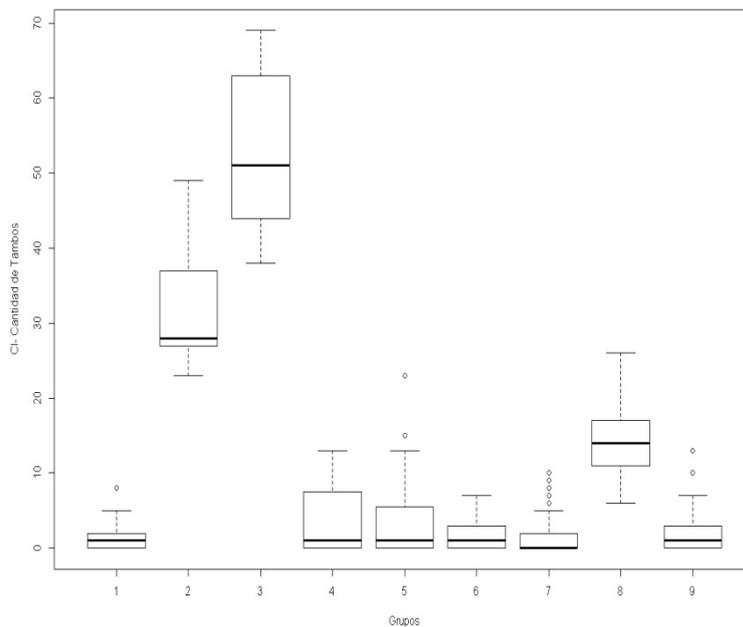
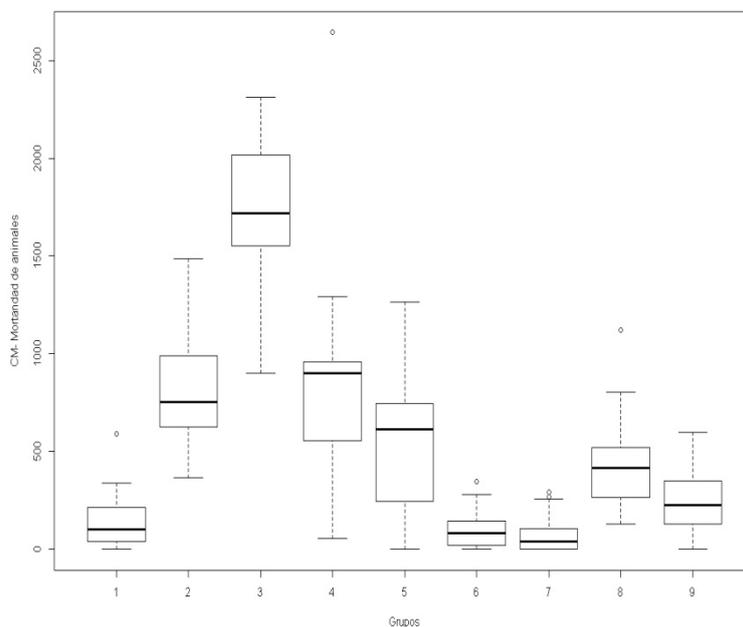


Tabla 11. Medidas descriptivas de la variable "Mortandad de animales" por grupo.

Variable: CM- Mortandad de animales			
Grupos	Media	Desvío estándar	Nº de distritos
1	149,9	156,4	16
2	809,6	284,5	24
3	1.734,8	407,2	10
4	896,3	683,6	11
5	511,9	333,4	27
6	96,0	88,7	35
7	62,7	72,0	114
8	415,1	190,0	49
9	246,1	148,6	43

Gráfico 24. Box-Plot de la variable "Mortandad de animales" por grupo.



6.4 Etiquetado de los grupos formados

A fin de asignar una etiqueta precisa que describa la naturaleza de los grupos se construye una tabla de doble entrada, donde cada grupo se clasifica según la ponderación de la variable que se le atribuye al mismo.

Tabla 12. Clasificación de los grupos según las variables y las ponderaciones.

Cría	Invernada	Mortandad	Tambo	Expl. Agro.	Ponderación
grupo 4	grupo 4	grupo 3	grupo 3	grupo 1	Máximo
grupo 5	grupo 5	grupo 4 y 2	grupo 2	grupo 4	Muy bueno
	grupo 3	grupo 5 y 8	grupo 8	grupo 6	Bueno
grupo 9 y 1	grupo 9, 2 y 1			grupo 3, 2 y 5	Moderado
grupo 3 y 6	grupo 8 y 6	grupo 9 y 1	grupo 4 y 5	grupo 8	Poco
grupo 8, 2 y 7	grupo 7	grupo 6 y 7	grupo 9, 6,1 y 7	grupo 9 y 7	Muy poco

Se determinan las siguientes etiquetas de acuerdo a las variables más significativas de cada grupo según su ponderación.

Tabla 13. Grupos por etiquetas

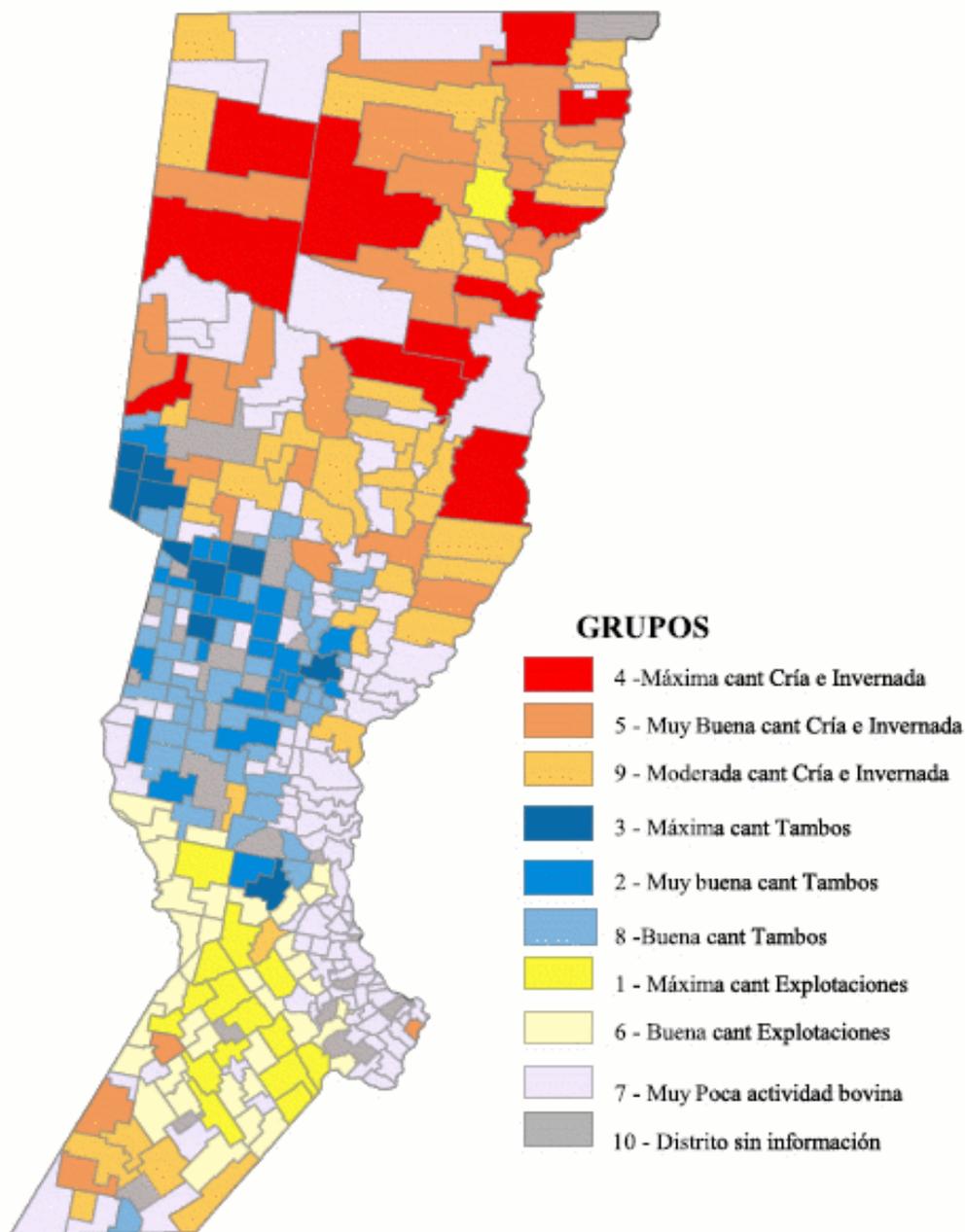
Grupo	Etiqueta
1	Máxima cantidad de Explotaciones
2	Muy buena cantidad de Tambos
3	Máxima cantidad de Tambos
4	Máxima cantidad de Cría e Invernada
5	Muy Buena cantidad de Cría e Invernada
6	Buena cantidad de Explotaciones
7	Muy Poca Actividad Bovina
8	Buena cantidad de Tambos
9	Moderada cantidad de Cría e Invernada

6.5 Mapa de ganado bovino por distrito según grupo

Una vez etiquetado cada uno de los nueve grupos con su característica particular, se grafica el mapa de la provincia de Santa Fe con cada distrito coloreado de un color específico según el grupo al cual corresponda. De esta manera puede verse zonas con colores predominantes, que pueden describirse con las características conforme a la etiqueta del grupo correspondiente.

Mapa 2

**Características de la Explotación Ganadera Bovina por distrito.
Provincia Santa Fe. Año 2012**



Fuente Datos: Encuesta Ganadera 2012

Fuente Mapa: IPEC

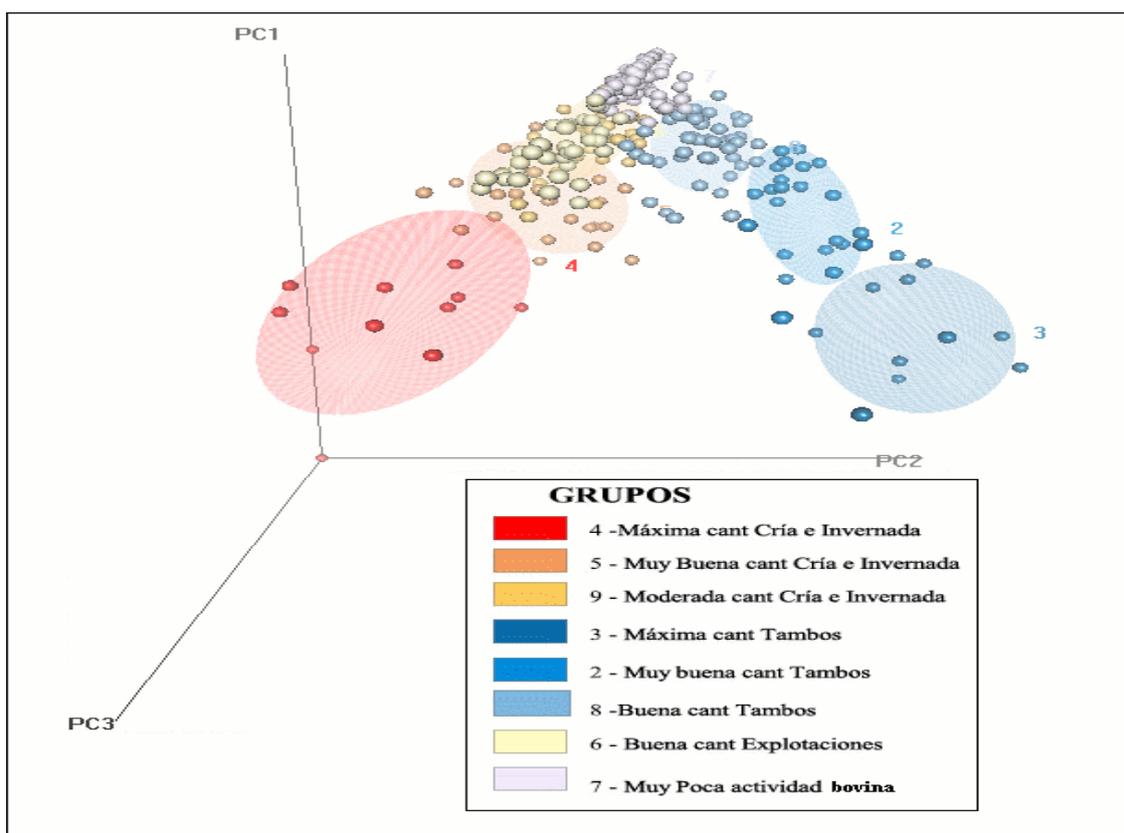
6.6 Separación de los grupos

Por medio de una inspección visual se puede verificar si la clasificación en grupos logra su objetivo. Ya que en la presentación geométrica los distritos dentro de un grupo estarán muy cercanos unos de otros y entre grupos diferentes estarán muy apartados. Grupos ideales pueden definirse como conjuntos de distritos compactos y aislados.

Desde un punto de vista geométrico, los distritos representados como puntos de distintos colores, según al grupo que pertenezcan, forman nubes de puntos por grupo en el espacio de 3 dimensiones.

Se presenta el gráfico resultante desde 2 puntos de vista distintos, para facilitar la visualización de los grupos según las variables que los etiquetan. Por limitaciones del software R que no puede colorear más de 8 grupos, no se visualizan los distritos del grupo 1 (color amarillo oscuro en el mapa), que se caracteriza por la máxima cantidad de explotaciones.

Gráfico 25. Distritos por grupo según las 3 componentes principales- 1º punto de vista.

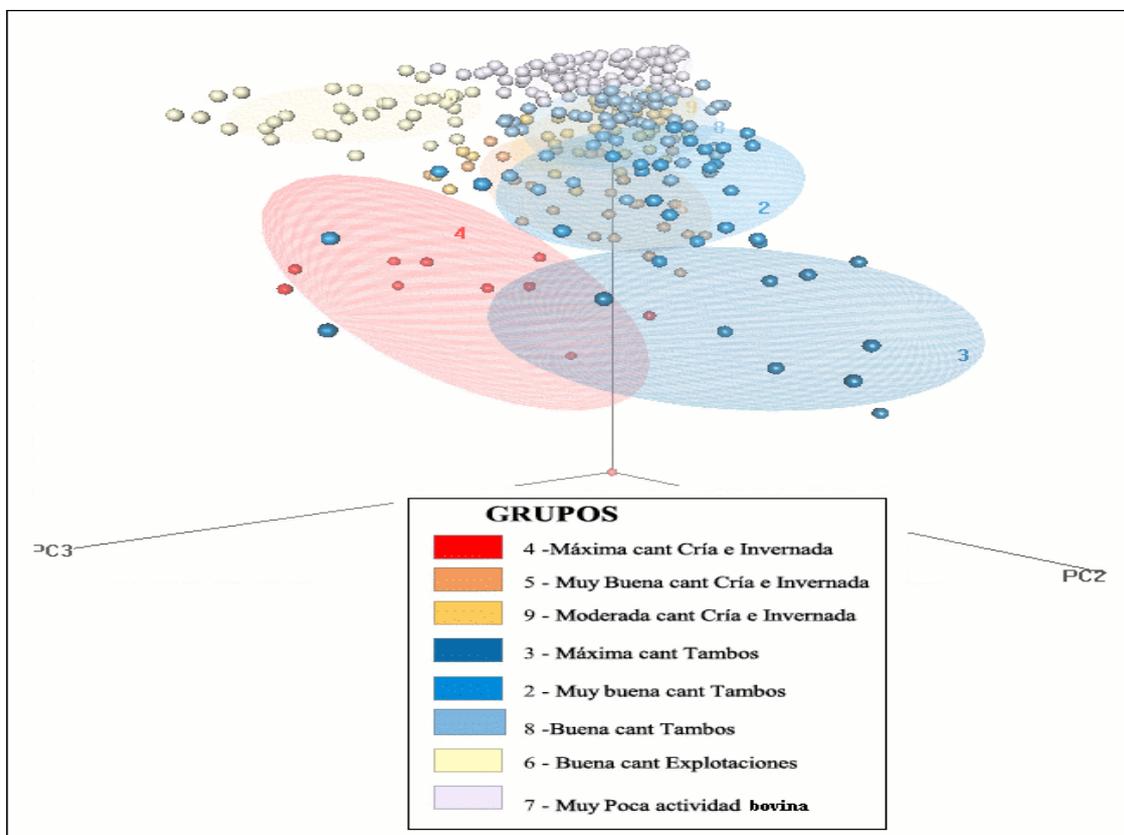


La componente 1, PC1, caracterizada por la contribución en forma negativa por todas las variables, sobretudo por las variables "Cantidad de mortandad", "Cantidad de animales para invernada" y "Cantidad de animales para cría" provoca que los grupos con mayor cantidad de animales para Cría e Invernada y mayor Mortandad se ubiquen más cerca del origen. Además, la disminución de los valores de estas variables hace que los grupos se vayan alineando, alejándose del origen. Es el caso de los grupos 4, 5 y 9 etiquetados como Máxima, Muy Buena y Moderada cantidad de animales para Cría e Invernada y los grupos 3, 2 y 8 que se caracterizan por la cantidad

de Tambos, etiquetados como Máxima, Muy Buena y Buena cantidad de tambos. Ubicándose los grupos 7 y 6 por arriba de todos, que son precisamente los grupos etiquetados con Poca actividad ganadera y Buena Cantidad de Explotaciones.

La componente 2, PC2, que lleva principalmente la información de las variables "Cantidad de Tambos" en forma positiva y de la variable "Cantidad de animales para cría" en forma negativa, provoca que el conjunto de grupos de Cría e Invernada con el conjunto de grupos de Tambos se separen. Arrastrando esta dimensión al grupo de Tambos hacia la derecha del gráfico.

Gráfico 26. Distritos por grupo según las 3 componentes principales- 2º punto de vista.



La tercer componente PC3, influenciada por la variable "Cantidad de Explotaciones agropecuarias", provoca que los grupos se alejen del origen y sean arrastrados por esta dimensión en la medida que los valores de esta variable sean mayores. Se observa en el gráfico que el grupo 6; "Buena cantidad de Explotaciones", es el más afectado por esta característica. Recordamos que el grupo 1, "Máxima cantidad de Explotaciones" no fue graficado.

6.7 Descripción de los grupos

Con las variables que caracterizan cada grupo, la descripción de las medidas estadísticas, cada box-plot, la tabla de clasificación, la tabla de etiquetado y los gráficos de distritos por grupo según las 3 componentes principales se detallan los grupos.

El grupo 1: (16 distritos) se caracteriza por la máxima cantidad de explotaciones agropecuarias, con poca mortandad y muy pocos tambos. Además tiene una moderada cantidad de animales para cría e internada. La mayoría de esos distritos se agrupan en la zona sur de la provincia.

El grupo 2: (24 distritos) queda definido por una muy buena cantidad de tambos y muy buena cantidad de mortandad de animales, además se caracteriza porque tiene muy pocos animales para cría. Los animales para internada como la cantidad de explotaciones se mantienen similares a los respectivos valores de las medias globales. Este grupo se distribuye en el centro de la provincia.

El grupo 3: (10 distritos) está formado por los distritos con la máxima cantidad de tambos y la máxima mortandad, buena cantidad de animales para internada y moderada cantidad de explotaciones. Los animales para cría son pocos.

Este grupo mantiene todos los valores de sus variables superiores a los de grupo 2. Se localiza en el centro de la provincia, más al norte del grupo 2.

El grupo 4: (11 distritos) se caracteriza por presentar la máxima cantidad de animales para cría e internada. Además la media de la mortandad se triplica con respecto a la mortandad de su media global y la media de la cantidad de explotaciones duplica esta medida, ponderándose estas 2 variables como Muy Buenas. Por otro lado dispone de pocos tambos.

Estos distritos se ubican en el norte de la provincia.

El grupo 5: (27 distritos) se identifica positivamente con las variables de internada, cría y mortandad, casi de la misma manera que el grupo 4, con la salvedad que las medias de todas las variables tienen un valor menor a las medias de las variables del grupo 4. También se caracteriza por la variable cantidad de tambos en forma negativa, ya que la media en este grupo es menor a la media global.

Los distritos se distribuyen alrededor de los distritos del grupo 4, en el norte de la provincia.

El grupo 6: (35 distritos) la única variable que lo define positivamente es la cantidad de explotaciones, siendo su valor medio menor al correspondiente del grupo 1, ponderándose como bueno. El resto de las variables mantienen sus medias por debajo de las medias globales, atribuyéndoles una ponderación entre poco y muy poco. Además estas medias son menores a las del grupo 1, salvo la variable cantidad de tambos que tiene un valor similar a la de dicho grupo.

Los distritos rodean a los distritos del grupo 1 en el sur de la provincia.

El grupo 7: (114 distritos) se destaca con las 5 variables caracterizadas en forma negativa. Todas tienen los valores de sus medias mínimas en cada grupo.

Es evidente que los distritos que conforman este grupo tienen muy poca actividad bovina. Gran parte de estos

distritos se alinean desde el sur de la provincia hasta el departamento Garay a lo largo de la costa del río Paraná, el resto se distribuye sin ningún orden en la provincia. La mayoría de estos últimos distritos tienen baja cobertura de respuesta en la encuesta, lo que puede explicar la pertenencia a este grupo.

El grupo 8: (49 distritos) se caracteriza positivamente por las variables de cantidad de tambos y mortandad, siendo sus valores medios casi la mitad de los correspondientes del grupo 2, resultando sus ponderaciones buenas. Además se caracteriza negativamente en las variables invernada, cantidad de explotaciones y cría. Los valores de las medias de las 2 primeras variables mencionadas anteriormente son menores que las del grupo 2, mientras que el correspondiente a la variable cría es levemente mayor.

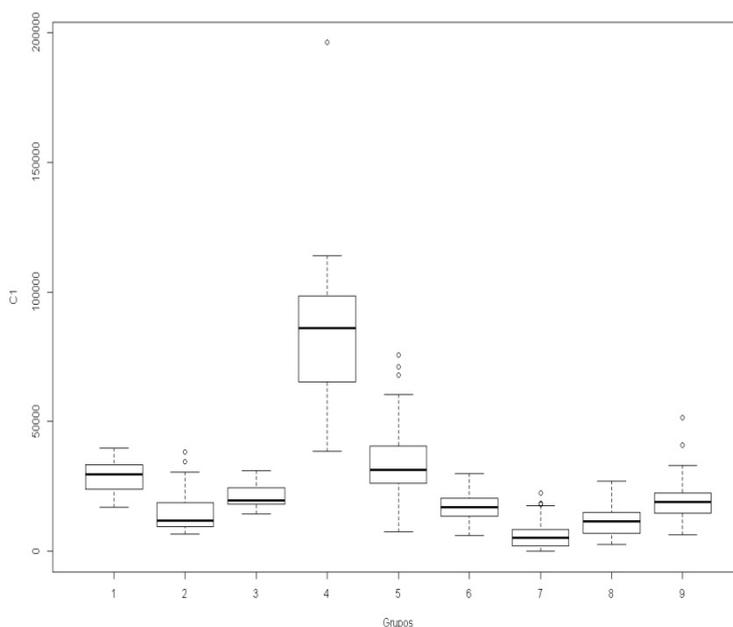
Los distritos se ubican en general más al sur que los del grupo 3 y 2, estableciendo una zona caracterizada por la cantidad de tambos y mortandad en el centro de la provincia.

El grupo 9: (43 distritos) se destacan las variables de cría e invernada, diferenciándose los valores medios a la mitad con los valores del grupo 5. Las variables de mortandad, cantidad de explotaciones y cantidad de tambos caracterizan al grupo en forma negativa y mantienen sus valores medios por debajo del grupo 5.

Estos distritos envuelven a los del grupo 4 y 5, conformando una zona de cría e invernada en el norte de la provincia.

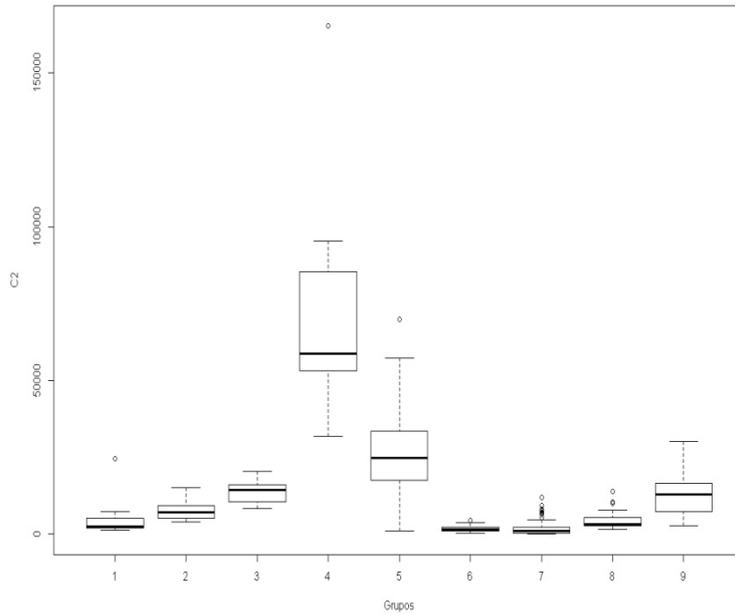
Para obtener mayor información de los grupos se observa cada box-plot de otras variables incluidas en la base original en función de los grupos.

Gráfico 27. Superficie de las explotaciones agropecuarias por grupo.



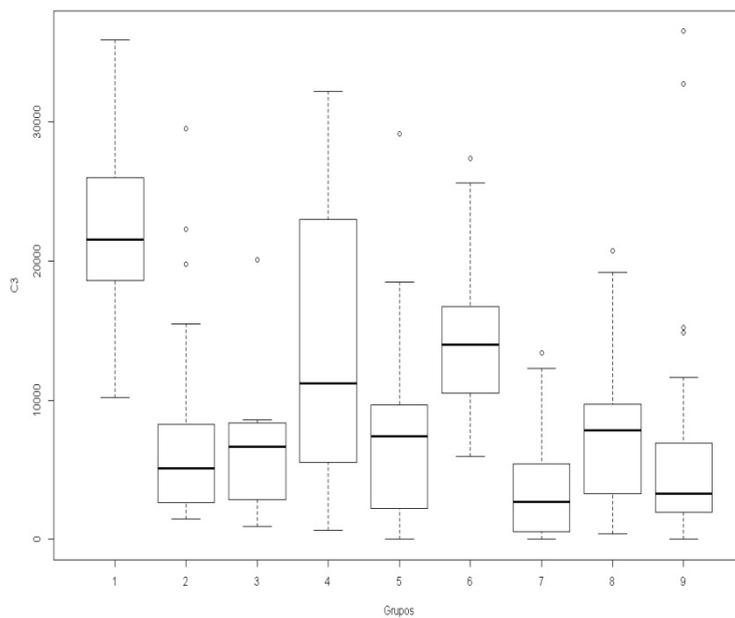
Se observa que los grupos 4 y 5 tienen los distritos con mayores superficies para las explotaciones. Le sigue el grupo 1 que lidera al resto de los grupos en menor medida.

Gráfico 28. Superficie de las explotaciones dedicadas a la ganadería por grupo.



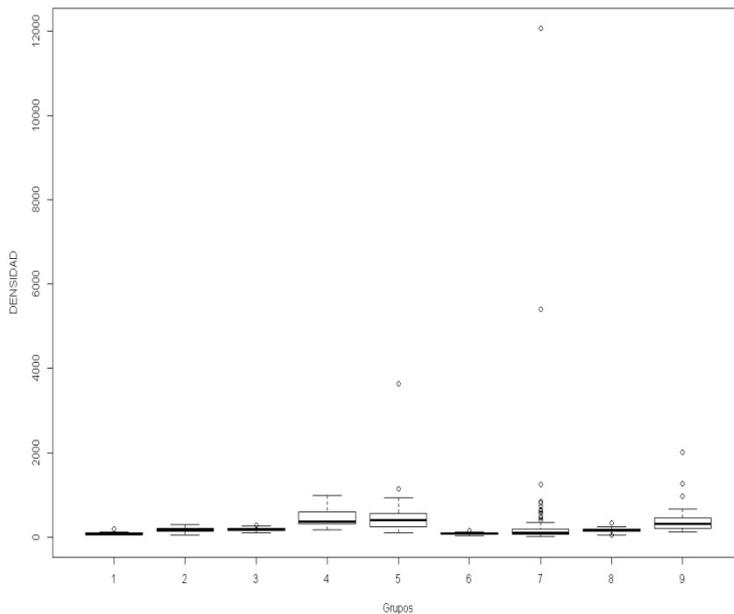
Los grupos 4 y 5 son los que mayores extensiones de tierra tienen dedicadas a la ganadería, liderando el resto de los grupos el 9. El grupo 1 es el que menor superficie de las explotaciones tiene dedicada a la ganadería.

Gráfico 29. Superficie de las explotaciones dedicadas a la agricultura por grupo.



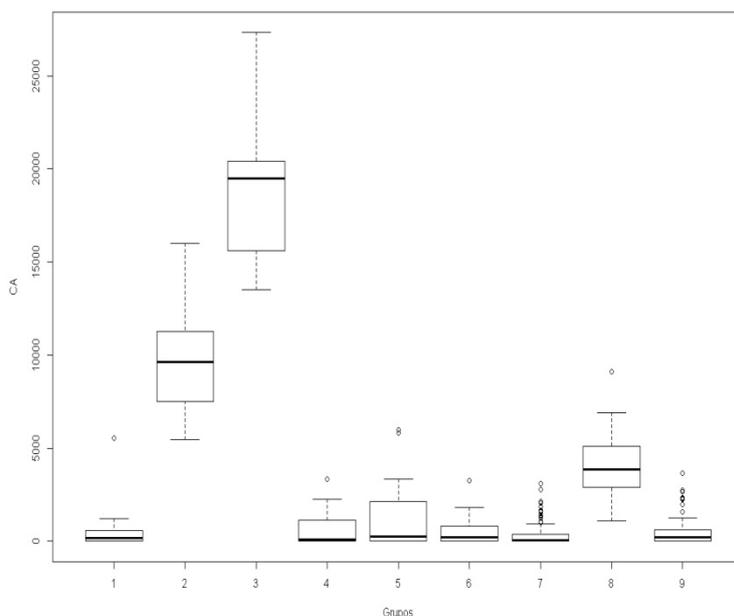
Los grupos 1 y 6 en la variable "Superficie dedicada a la agricultura" presentan sus medianas más altas, estando casi todos sus datos por arriba de los datos de los otros grupos salvo del grupo 4 que tiene valores muy altos y muy bajos.

Gráfico 30. Densidad (superficie / cantidad de explotaciones) de las explotaciones por grupo.



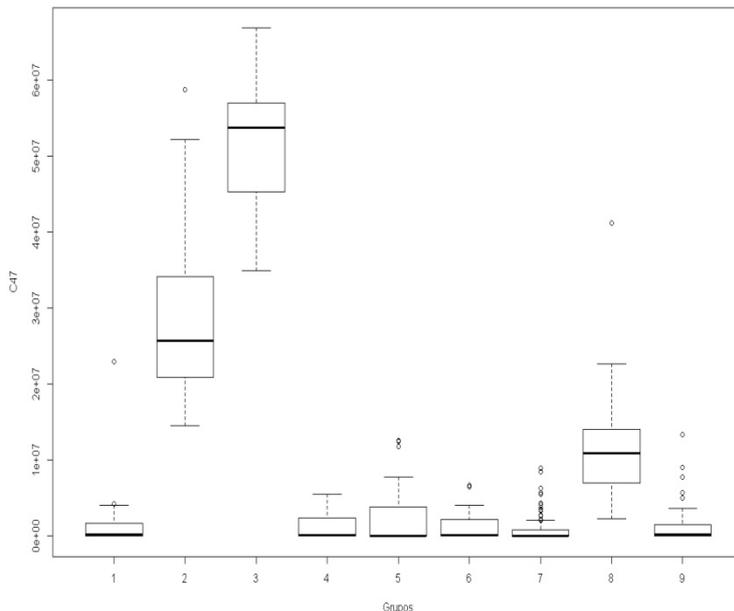
Se observa que los grupos 4, 5 y 9 tienen sus explotaciones más grandes comparadas con los demás grupos.

Gráfico 31. Cantidad de animales para tambos por grupo.



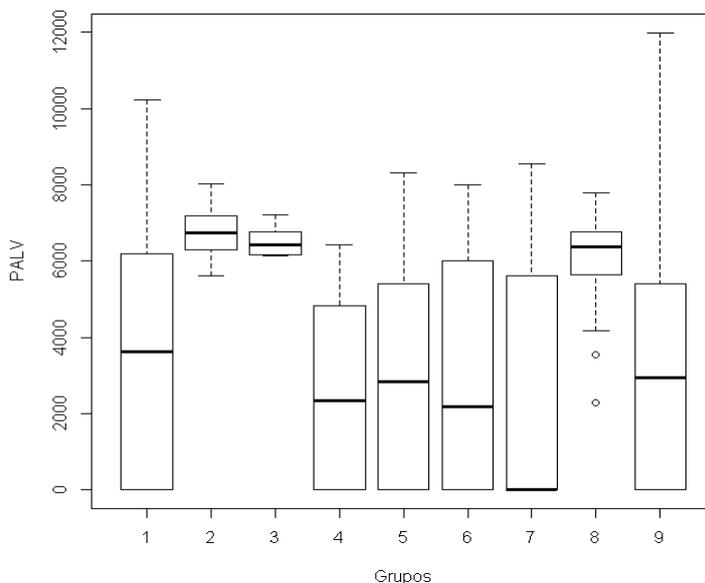
Los animales para tambos se agrupan en los grupos 3, 2 y 8, habiendo muy pocos animales de esta categoría en los demás grupos.

Gráfico 32. Litros obtenidos del 01/07/11 al 30/06/12 para tambos por grupo.



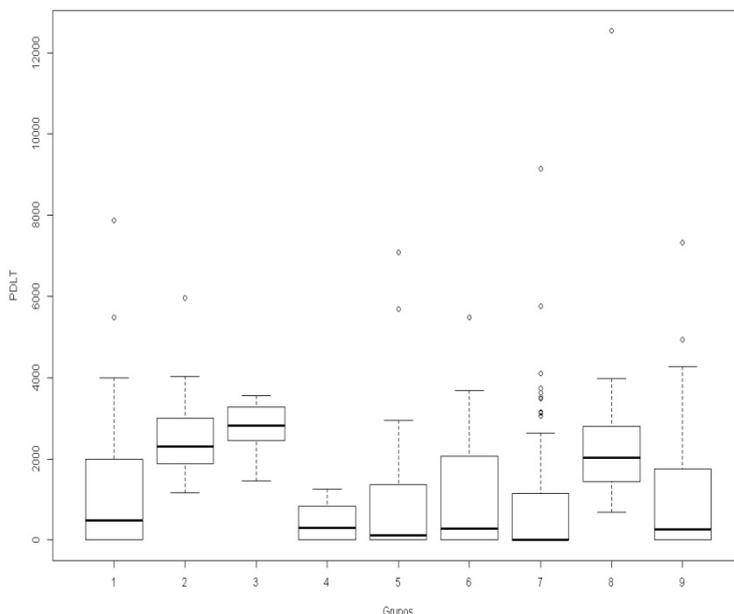
La cantidad de litros obtenidos durante el año de referencia está de acuerdo a la cantidad de animales para tambos.

Gráfico 33. Producción anual lechera por vaca en ordeño por grupo.



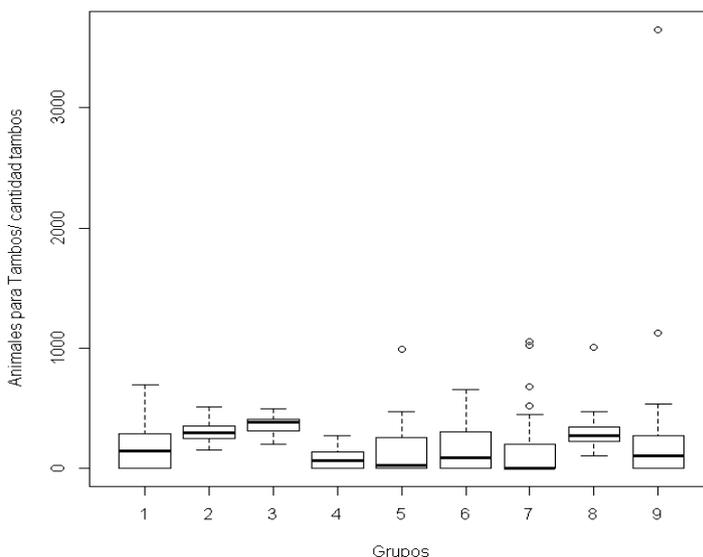
Se observa que las medianas de la producción de litros anuales por vaca en ordeño son superiores en los grupos 2, 3 y 8. Las medianas en estos grupos están alrededor de los 6500 litros anuales por vaca, mientras que en el resto de los grupos no superan los 3500 litros. Es decir que la producción de leche anual por vaca es muy superior en los grupos etiquetados con tambos con respecto a los otros grupos.

Gráfico 34. Producción diaria de litros por Tambo por grupo.



La producción diaria de litros por Tambo es mayor en los grupos 3, 2 y 8, oscilando las medianas en los 2500 litros diarios por tambo, mientras que en los tambos de los otros grupos las medianas no superan los 500 litros.

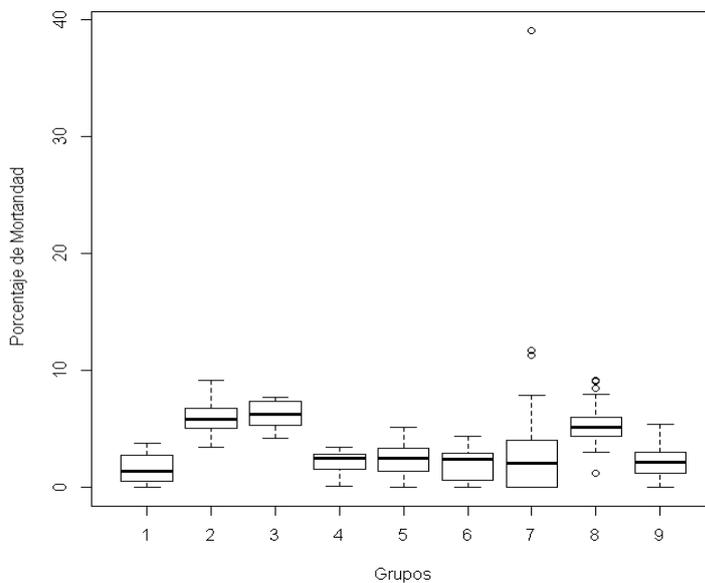
Gráfico 35. Tamaño de los Tambos (animales para Tambos/ cantidad de Tambos) por grupo.



Los tambos más grandes de la provincia, medidos por la cantidad de animales se localizan en los grupos 3, 2 y 8, es decir en la zona tambera. Son tambos donde los valores de sus medianas figuran entre 380 y 270 de animales, mientras que en los otros sus medianas no superan los 140 animales.

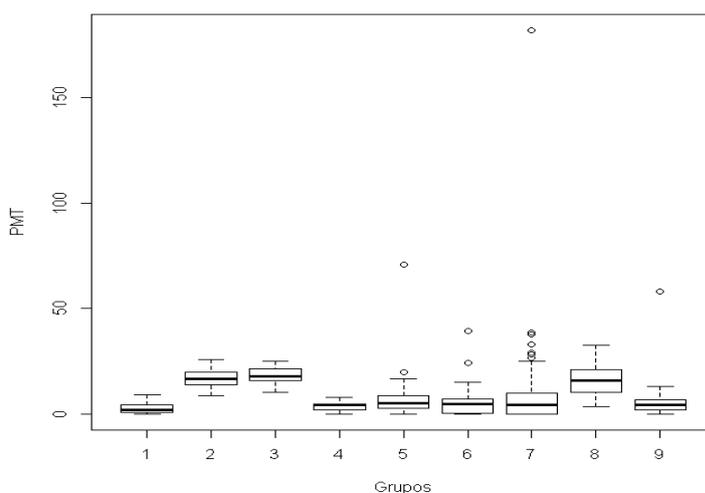
Cabe mencionar que el grupo 7 mantiene los valores de las medianas en cero en casi todas las variables relacionadas con los tambos.

Gráfico 36. Porcentaje de Mortandad por grupo.



El porcentaje de la mortandad de los grupos 3, 2 y 8 es superior con respecto a los demás grupos. Los valores de las medianas son respectivamente 6,3 %, 5,8 % y 5,2 %, mientras que las medianas de los otros grupos no superan el 2,9 %.

Gráfico 37. Porcentaje de Mortandad de Terneros y Terneras por grupo.



Las medianas de los porcentajes de mortandad de terneros y de terneras en los grupos 3, 2 y 9 son del 18 %, 16,8 % y del 15,9 % respectivamente. Las medianas en los otros grupos no superan el 5 %.

En base a las descripciones de los grupos se puede deducir la existencia de 4 grandes zonas. Una zona identificada con grupos por animales para la cría, que se relacionan con los grupos por animales para invernada, ubicándose en el norte de la provincia y siendo representada por 21 % de los distritos que contestaron a la encuesta.

Otra zona valorada por la cantidad de tambos asociada principalmente con la mortandad y limitándose al centro de la provincia, constituida por el 25 % de los distritos.

Una 3º zona formada por el 15 % de los distritos localizados en el sur de la provincia, los cuales se caracterizan por la cantidad de explotaciones con moderada o poca actividad bovina, con actividad agrícola.

La 4º y última zona identificada solamente por el grupo que más distritos poseen, con el 22 % del total de los mismos, se caracteriza por la ausencia de la actividad bovina y de las explotaciones agropecuarias. Esta zona está ubicada en el cordón industrial. Los distritos que pertenecen al grupo 7 y están fuera de la zona se deben en su mayoría a la baja cobertura de respuesta.

Por último, se resumen todas las características obtenidas anteriormente por grupo y por zona en la tabla 14 del anexo.



7. Anexo

Tabla 14. Características por grupo y zona

Zonas	Grupo	Etiquetado	Cantidad de distritos del grupo	Localización de los distritos dentro de la provincia	Características Principales	Otras Características
1 Cría e Invernada	4	Máxima cantidad de Cría e Invernada	11	Norte	<ul style="list-style-type: none"> <input type="checkbox"/> Máxima cant. de animales para cría (promedio = 32443 vacunos) <input type="checkbox"/> Máxima cant. de animales para Invernada (promedio = 10917 vacunos) <input type="checkbox"/> Muy Buena Mortandad (promedio = 896 vacunos) <input type="checkbox"/> Muy Buena cant. explotaciones (promedio = 203 exp.) <input type="checkbox"/> Pocos tambos (promedio = 4 tambos) 	<ul style="list-style-type: none"> <input type="checkbox"/> Explotaciones más extensas <input type="checkbox"/> La mayoría de explotaciones dedicadas a la ganadería
	5	Muy Buena cant. de Cría e Invernada	27	Alrededor de los del grupo 4	<ul style="list-style-type: none"> <input type="checkbox"/> Muy buena cant. de animales para cría (promedio = 12648 vacunos) <input type="checkbox"/> Muy buena cant. de animales para invernada (promedio = 7295 vacunos) <input type="checkbox"/> Buena Mortandad (promedio = 511 vacunos) <input type="checkbox"/> Moderada cant. explotaciones (promedio = 40 exp.) <input type="checkbox"/> Pocos tambos (promedio = 4 tambos) 	
	9	Moderada cant. de Cría e Invernada	43	Envuelven a los distritos del grupo 4 y 5	<ul style="list-style-type: none"> <input type="checkbox"/> Moderada cant. animales para cría (promedio = 6864 vacas) <input type="checkbox"/> Moderada cant. animales para invernada (promedio = 3603 vacas) <input type="checkbox"/> Poca Mortandad (promedio = 246 vacunos) <input type="checkbox"/> Muy pocas explotaciones (promedio = 67 exp.) <input type="checkbox"/> Muy pocos tambos (promedio = 2 tambos) 	

2 Tambo	3	Máxima cant. de tambos	10	Centro	<ul style="list-style-type: none"> <input type="checkbox"/> Máxima cant. de tambos (promedio = 52 tambos) <input type="checkbox"/> Máxima cant. de mortandad (promedio = 1735 vacunos) <input type="checkbox"/> Animales para invernada (promedio = 6693 vacunos) <input type="checkbox"/> Moderada cant. explotaciones (promedio = 132 exp.) <input type="checkbox"/> Pocos animales para cría (promedio = 2559 vacunos) 	<ul style="list-style-type: none"> <input type="checkbox"/> El promedio en los grupos tamberos es de 19051, 9573 y de 4133 animales para tambos, no superando el promedio de 765 animales en los otros grupos tamberos. <input type="checkbox"/> El promedio de la producción anual de litros obtenidos está entre los 51773727 l y de 28478748 l en los grupos tamberos, mientras que los demás grupos no superan un promedio de 2696303 l. <input type="checkbox"/> El promedio de la producción anual de leche por vaca está entre 6746 l y 6170 l en los grupos tamberos. Además el promedio diario de l de leche por tambo, está entre 2771 l y 2290 l en estos grupos. Los demás grupos no superan el promedio de 3643 l anual de leche por vaca y de 1533 l diario de leche por tambo. <input type="checkbox"/> Los tambos más grandes se dan en estos 3 grupos donde el promedio está entre 370 y 292 animales por tambo, mientras que en el resto de los tambos los promedios están por debajo de las 250 cabezas por tambo. <input type="checkbox"/> El promedio del porcentaje de la mortandad de vacunos en estos tres grupos oscila entre el 6,2 % y el 5,3 %, mientras en el resto de los grupos el porcentaje no supera el 2,9 <input type="checkbox"/> El promedio de los porcentajes de mortandad en terneros está entre los 18,3 % y el 16,1 % en estos 3 grupos, mientras en el resto de los grupos el porcentaje no supera el 8,5 %
	2	Muy Buena cant. de tambos	24	Centro, por debajo de los distritos del grupo 3	<ul style="list-style-type: none"> <input type="checkbox"/> Muy Buena cant. de tambos (promedio = 32 tambos) <input type="checkbox"/> Muy Buena cant. de mortandad (promedio = 809 vacunos) <input type="checkbox"/> Moderados animales para invernada (promedio = 3067 vacunos) <input type="checkbox"/> Moderada cant. de explotaciones (promedio = 101 exp.) <input type="checkbox"/> Muy pocos animales para cría (promedio = 1406 vacunos) 	
	8	Buena cant. de tambos	49	Centro, por debajo de los distritos del grupo 2	<ul style="list-style-type: none"> <input type="checkbox"/> Buena cant. de tambos (promedio = 15 tambos) <input type="checkbox"/> Buena cant. de mortandad (promedio = 415 vacunos) <input type="checkbox"/> Pocos animales para invernada (promedio = 2266 vacunos) <input type="checkbox"/> Pocas explotaciones (promedio = 79 exp.) <input type="checkbox"/> Muy poca cría (promedio = 1564 vacunos) 	
3 Agrícola	1	Máxima cant. explotaciones	16	Sur	<ul style="list-style-type: none"> <input type="checkbox"/> Máxima cant. de explotaciones (promedio = 360 exp) <input type="checkbox"/> Moderada cant. animales para cría (promedio = 5774 vacunos) <input type="checkbox"/> Moderada cant animales para invernada (promedio = 2808 vacunos) <input type="checkbox"/> Poca mortandad (promedio = 150 vacunos) <input type="checkbox"/> Muy pocos tambos (promedio = 2 tambos) 	<ul style="list-style-type: none"> <input type="checkbox"/> Explotaciones dedicadas a la agricultura
	6	Buena cant. explotaciones	35	Sur, alrededor de los distritos del grupo 1	<ul style="list-style-type: none"> <input type="checkbox"/> Buena cant. de explotaciones (promedio = 197 exp.) <input type="checkbox"/> Pocos animales para cría (promedio = 2504 vacunos) <input type="checkbox"/> Pocos animales para invernada (promedio = 2103 vacunos) <input type="checkbox"/> Muy poca mortandad (promedio = 96 vacunos) <input type="checkbox"/> Muy pocos tambos (promedio = 2 tambos) 	

<p>4 Industrial</p>	<p>7</p>	<p>Muy poca actividad bovina</p>	<p>114</p>	<p>Cordón industrial y dispersos por la provincia</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Muy pocos animales para cría (promedio = 1123 vacunos) <input type="checkbox"/> Muy pocos tambos (promedio = 2 tambos) <input type="checkbox"/> Muy poca cant. de explotaciones (promedio = 50 exp.) <input type="checkbox"/> Muy poca mortandad (promedio = 63 vacunos) <input type="checkbox"/> Muy pocos animales para invernada (promedio = 634 vacunos) 	<ul style="list-style-type: none"> <input type="checkbox"/> Pocas explotaciones y las explotaciones que hay tienen poca actividad agropecuaria.
---------------------------------------	----------	----------------------------------	------------	---	---	--

Tabla 15. Listado de distritos que no respondieron a la encuesta ganadera 2012

DEPARTAMENTO	DISTRITO
CASEROS	Villada
CASTELLANOS	Aurelia
	Colonia Iturraspe
	Frontera
	Hugentobler
	Ramona
	San Antonio
	San Vicente
CONSTITUCIÓN	Susana
	Virginia
GENERAL LÓPEZ	Santa Teresa
	Sargento Cabral
GENERAL OBLIGADO	Chapuy
	Christophersen
LAS COLONIAS	Florencia
	Colonia Ituzangó
	Hipatía
	Rivadavia
	San Carlos Norte
ROSARIO	Santa María Centro
	Soutomayor
	Acebal
	Figuera
SAN CRISTÓBAL	Granadero Baigorria
	Piñero
	Arrufó
SAN JERÓNIMO	Portugalete
	Maciel
SAN JUSTO	San Genaro
	La Camila
SAN MARTÍN	Cañada de Gómez
	Casas
	Trall

Tabla 16. Listado de distritos atípicos

DEPARTAMENTO	DISTRITO
BELGRANO	Las Rosas
CASEROS	Arequito Berabevú Bigand Casilda Chabás Godekén San José de la Esquina
CASTELLANOS	Aldao Ataliva Colonia Raquel Egusquiza Eusebia Humberto I Josefina Lehmann Presidente Roca Sunchales Tacural Tacurales Villa San José
CONSTITUCIÓN	Alcorta Juncal Máximo Paz Theobald
GARAY	Helvecia
GENERAL LÓPEZ	Amenabar Chovet Elortondo Firmat Maggiolo
GENERAL OBLIGADO	Avellaneda El Sombrerito Ingeniero Chanourdi La Sarita Lanteri Nicanor Molinas Reconquista Villa Ana Villa Guillermina Villa Ocampo
IRIONDO	Cañada de Gómez Clason Totoras Villa Eloísa
LACAPITAL	Nelson

LAS COLONIAS	Empalme San Carlos Esperanza Felicia Humboldt La Pelada Nuevo Torino Sa Pereyra San Jerónimo del Sauce San Jerónimo Norte Santa Clara de Buena Vista Santa María Norte Sarmiento
NUEVE DE JULIO	Montefiore Pozo Borrado San Bernardo Tostado
SAN CRISTÓBAL	Aguará Ambrosetti Ceres Colonia Rosa Curupaytí Hersilia La Lucila Las Avispas Moisés Ville Monte Oscuridad San Guillermo Suardi Villa Trinidad
SAN JAVIER	Colonia Durán Romang San Javier
SAN JUSTO	San Justo
SAN MARTÍN	Carlos Pellegrini Crispi
VERA	Calchaqui Cañada Ombú Fortín Olmos Garabato Intiyaco Margarita Vera

8. Referencia

Mixed – Effets Models in S and S – PLUS, José C. Pinheiro, Douglas M. Bates.

Applied Linear Regression , Sanford Weisberg.

Outliers Identification in High Fimensions.P Filzmoser, R. Maronna, and M. Werner

Reconocimiento de Patronos. L. F. Escudero

El problema de los ouliers multivariantes en el análisis de sectores clave y cluster industrial. Antonio Morillas, Bárbara Diaz.

Algoritmos de agrupamiento, D. Pascual 1, F. Pla 2, S. Sánchez.

Identification of Multivariate Outliers: A Performance Study, Peter Filzmoser, Vienna University of Technology, Austria.

Análisis multivariante, Hair, Anderson, Tatham, Black.

Aproximaciones al diagnóstico de la situación social de la Provincia de Santa Fe. Liborio M., Alvarez G., Stieffel M., Merello J, Balbi S, Aguilar V, et al (2008). “.

Software utilizado:

El Sistema de Información Geográfica “GvSIG”.

El lenguaje y entorno de programación para análisis estadístico “R”.