🐄 **Análisis de la producción lechera en granjas españolas** 🐄

**Analizando los factores clave que determinan su rendimiento.**

**Por: Miguel Diaz Perez de Juan y Dante Mathieu Schrantz Castro**

[**Acceso GitHub**](https://github.com/DanteSc03/stata)

En un sector tan competitivo como la producción de leche, ¿qué hace que algunas granjas sean más rentables que otras? ¿Es el número de vacas, la cantidad de tierra, la mano de obra, o la alimentación lo que marca la diferencia? En este análisis, intentaremos descifrar cómo estos factores afectan la producción de leche y cuál es la influencia real de cada uno en el rendimiento de las granjas. Usaremos datos reales de unas 250 granjas del norte de España, con registros de varios años, para entender mejor el comportamiento de las variables que afectan a esta industria.

# Objetivo

Nuestro objetivo es explorar la relación entre las variables explicativas (como vacas, tierra, mano de obra y alimentación) y la producción de leche. Usaremos nuestro conjunto de datos de producción lechera para obtener un modelo econométrico que nos permita entender cómo estos factores afectan el rendimiento de las granjas, y qué variables tienen mayor influencia.

**¿Por qué?**

La idea de este trabajo surgió una tarde entre Dante y yo, tomando un Colacao en medio de una conversación sobre la realidad de las granjas lecheras en el norte de España, especialmente en Galicia, de donde soy. Mientras charlábamos sobre cómo muchas de estas granjas luchan por mantenerse a flote, nos dimos cuenta de que era un tema que merecía ser estudiado. ¿Por qué no explorar cómo las variables económicas y operacionales de estas granjas afectan la producción de leche? Así nació nuestra motivación para analizar estos datos, con la intención de aportar alguna luz sobre cómo mejorar la eficiencia de las granjas lecheras y, de paso, comprender mejor la rentabilidad y la sostenibilidad del sector.

# Variables a considerar

Las variables presentes en el conjunto de datos son las siguientes:

**Entradas:**

**COWS:** Número de vacas en la granja.

**LAND:** Medida de la cantidad de terreno disponible.

**LABOR:** Número de trabajadores en la granja.

**FEED:** Cantidad de alimentación proporcionada.

**Variables transformadas (logaritmos):**

**X1 (log de COWS):** Logaritmo del número de vacas en la granja.

**X2 (log de LAND):** Logaritmo de la cantidad de terreno disponible.

**X3 (log de LABOR):** Logaritmo del número de trabajadores.

**X4 (log de FEED):** Logaritmo de la cantidad de alimentación proporcionada.

**Términos Translog (variables cuadráticas y de productos cruzados):**

**X11, X22, X33, X44:** Cuadráticos de las variables logarítmicas X1, X2, X3 y X4, respectivamente.

**X12, X13, X14, X23, X24, X34:** Productos cruzados de las variables logarítmicas, que capturan las interacciones no lineales entre ellas.

**Variables dummy de año:**

**YEAR93, YEAR94, ..., YEAR98:** Variables que indican el año de observación.

**Salida:**

**MILK:** La producción total de leche de cada granja.

**YIT (log de la producción de leche):** En nuestro caso no la usaremos, usaremos directamene el output de leche -> MILK para nuestros modelos y para el Time Series.

# Metodología - Datos de Panel

. **Datos de Panel**: Dado que tenemos datos de varias granjas a lo largo de varios años (N = 247, T = 6), el modelo adecuado para este análisis es un modelo de **panel**. Esto nos permitirá capturar tanto las variaciones a nivel de granja como las variaciones temporales.

. **Regresión de Panel**: Usaremos una regresión de panel para analizar cómo las variables de entrada y los efectos del año influyen en la producción de leche. Además, probaremos con efectos fijos y aleatorios para ver qué tipo de especificación se ajusta mejor a los datos.

. **Multicolinealidad**: Antes de realizar la regresión, verificaremos si existe multicolinealidad entre las variables de entrada, usando el índice de **VIF** (Factor de Inflación de la Varianza).

# Metodología - Time Series

. **Series Temporales**: Dado que nuestro objetivo es predecir la producción de leche (**MILK**) a partir de sus valores pasados y de los lags de la variable **COWS**, aplicamos un análisis de **series temporales**. En este análisis, consideramos que la producción de leche depende de los valores previos de la variable de vacas.

. **Autocorrelación y Lags**: Después de obtener la serie temporal, hemos realizado un análisis de **autocorrelación** para identificar si los **lags de la variable COWS** son significativos para predecir el valor futuro de la producción de leche. Esto nos ayudó a seleccionar los lags adecuados para el modelo.

. **Modelos AR (Autoregresivos)**: Para predecir la producción futura de leche, aplicamos un **modelo AR** usando los **lags de COWS** seleccionados. Esto nos permitió capturar la relación entre la producción de leche y los valores pasados de las vacas.

. **Evaluación de Resultados**: Finalmente, analizamos los resultados del modelo AR, evaluando la significancia de los coeficientes.

|  |
| --- |
| **import** pandas **as** pd **import** numpy **as** np **import** os os**.**chdir('/Applications/Stata/utilities') **from** pystata **import** config config**.**init('se')  **%stata** import delimited "/Users/danteschrantz/Desktop/UNAV/2024-2025/Modeling Trabajo Final/dairy.csv" |

In [ ]:

\_\_\_ \_\_\_\_ \_\_\_\_ \_\_\_\_ \_\_\_\_ ®

/\_\_ / \_\_\_\_/ / \_\_\_\_/ StataNow 18.5

\_\_\_/ / /\_\_\_/ / /\_\_\_/ SE—Standard Edition

Statistics and Data Science Copyright 1985-2023 StataCorp LLC

StataCorp

4905 Lakeway Drive

College Station, Texas 77845 USA

800-782-8272 https://www.stata.com

979-696-4600 service@stata.com

Stata license: Single-user , expiring 25 Nov 2024

Serial number: 401809408832 Licensed to: Dante Schrantz

Universidad de Navarra

Notes:

1. Unicode is supported; see help unicode\_advice.
2. Maximum number of variables is set to 5,000 but can be increased; see help set\_maxvar.

(encoding automatically selected: ISO-8859-1)

(28 vars, 1,482 obs)

## **Fuentes de Datos**

Los datos utilizados provienen de un conjunto de datos sobre la producción lechera en granjas españolas, disponible en la web de NYU Stern, que proporciona acceso a diversos conjuntos de datos de panel utilizados en econometría. Esta fuente es una recopilación de datasets que pueden ser utilizados para análisis económicos y agrícolas.

Se puede encontrar más información y acceder al conjunto de datos en el siguiente enlace: [NYU Stern Panel Data Sets](https://pages.stern.nyu.edu/~wgreene/Econometrics/PanelDataSets.htm).

|  |
| --- |
| **%stata** summarize |

In [ ]:

Variable | Obs Mean Std. dev. Min Max

-------------+-------------------------------------------------------- farm | 1,482 124 71.32624 1 247 year | 1,482 95.5 1.708402 93 98 cows | 1,482 22.11619 11.27415 4.5 82.3 land | 1,482 12.9857 6.172051 2 45.1 milk | 1,482 131106.6 92583.98 14410 727281

-------------+-------------------------------------------------------- labor | 1,482 1.671997 .5527831 1 4 feed | 1,482 57941.35 47981.24 3924.141 376731.6 yit | 1,482 11.57749 .6434377 9.575678 13.49707 x1 | 1,482 -1.40e-09 .4885577 -1.474623 1.431671 x2 | 1,482 5.16e-09 .4559323 -1.767564 1.348171

-------------+-------------------------------------------------------- x3 | 1,482 5.43e-09 .3428585 -.4570751 .9292222 x4 | 1,482 -1.22e-09 .7599885 -2.410496 2.153889 x11 | 1,482 .1192638 .1723436 1.80e-06 1.087256 x22 | 1,482 .103867 .1539671 .0000273 1.562141 x33 | 1,482 .0587363 .0534247 .0013315 .4317269

-------------+-------------------------------------------------------- x44 | 1,482 .2885964 .3847499 0 2.905246 x12 | 1,482 .1531676 .2855238 -.3743768 2.054143 x13 | 1,482 .0949598 .1775786 -.4079377 1.025777 x14 | 1,482 .3305641 .4785109 -.1321613 3.20606 x23 | 1,482 .052876 .1626291 -.4664559 1.252751

-------------+-------------------------------------------------------- x24 | 1,482 .1923081 .4297219 -.9513205 3.722819 x34 | 1,482 .1426242 .2619795 -.5546867 1.495623 year93 | 1,482 .1666667 .3728038 0 1 year94 | 1,482 .1666667 .3728038 0 1 year95 | 1,482 .1666667 .3728038 0 1

|  |
| --- |
| **%stata** label variable farm "Identificador único de la granja"  **%stata** label variable year "Año de la observación (93-98)"  **%stata** label variable cows "Número de vacas en la granja"  **%stata** label variable land "Superficie de tierra utilizada en hectáreas"  **%stata** label variable labor "Horas de trabajo de los empleados"  **%stata** label variable feed "Cantidad de alimentación proporcionada a las vacas"  **%stata** label variable yit "Logaritmo de la producción de leche"  **%stata** label variable x1 "Deviation from the mean (log)"  **%stata** label variable x2 "Deviation from the mean (log)"  **%stata** label variable x3 "Deviation from the mean (log)" **%stata** label variable x4 "Deviation from the mean (log)"  **%stata** label variable x11 "Square of X1"  **%stata** label variable x22 "Square of X2"  **%stata** label variable x33 "Square of X3"  **%stata** label variable x44 "Square of X4"  **%stata** label variable x12 "Cross-product of X1 and X2"  **%stata** label variable x13 "Cross-product of X1 and X3"  **%stata** label variable x14 "Cross-product of X1 and X4"  **%stata** label variable x23 "Cross-product of X2 and X3"  **%stata** label variable x24 "Cross-product of X2 and X4"  **%stata** label variable x34 "Cross-product of X3 and X4"  **%stata** label variable year93 "Dummy variable for year 1993"  **%stata** label variable year94 "Dummy variable for year 1994"  **%stata** label variable year95 "Dummy variable for year 1995"  **%stata** label variable year96 "Dummy variable for year 1996"  **%stata** label variable year97 "Dummy variable for year 1997"  **%stata** label variable year98 "Dummy variable for year 1998" |

-------------+-------------------------------------------------------- year96 | 1,482 .1666667 .3728038 0 1 year97 | 1,482 .1666667 .3728038 0 1 year98 | 1,482 .1666667 .3728038 0 1

In [ ]:

|  |
| --- |
| **%stata** describe |

In [ ]:

Contains data

Observations: 1,482

Variables: 28

-------------------------------------------------------------------------------

Variable Storage Display Value name type format label Variable label

------------------------------------------------------------------------------farm int %8.0g Identificador único de la granja year byte %8.0g Año de la observación (93-98) cows float %9.0g Número de vacas en la granja land float %9.0g Superficie de tierra utilizada en

hectáreas milk long %12.0g MILK labor float %9.0g Horas de trabajo de los empleados feed float %9.0g Cantidad de alimentación

proporcionada a las vacas

yit float %9.0g Logaritmo de la producción de leche x1 float %9.0g Deviation from the mean (log) x2 float %9.0g Deviation from the mean (log) x3 float %9.0g Deviation from the mean (log) x4 float %9.0g Deviation from the mean (log) x11 float %9.0g Square of X1 x22 float %9.0g Square of X2 x33 float %9.0g Square of X3 x44 float %9.0g Square of X4 x12 float %9.0g Cross-product of X1 and X2 x13 float %9.0g Cross-product of X1 and X3 x14 float %9.0g Cross-product of X1 and X4 x23 float %9.0g Cross-product of X2 and X3 x24 float %9.0g Cross-product of X2 and X4 x34 float %9.0g Cross-product of X3 and X4 year93 byte %8.0g Dummy variable for year 1993 year94 byte %8.0g Dummy variable for year 1994 year95 byte %8.0g Dummy variable for year 1995 year96 byte %8.0g Dummy variable for year 1996 year97 byte %8.0g Dummy variable for year 1997 year98 byte %8.0g Dummy variable for year 1998 ------------------------------------------------------------------------------Sorted by:

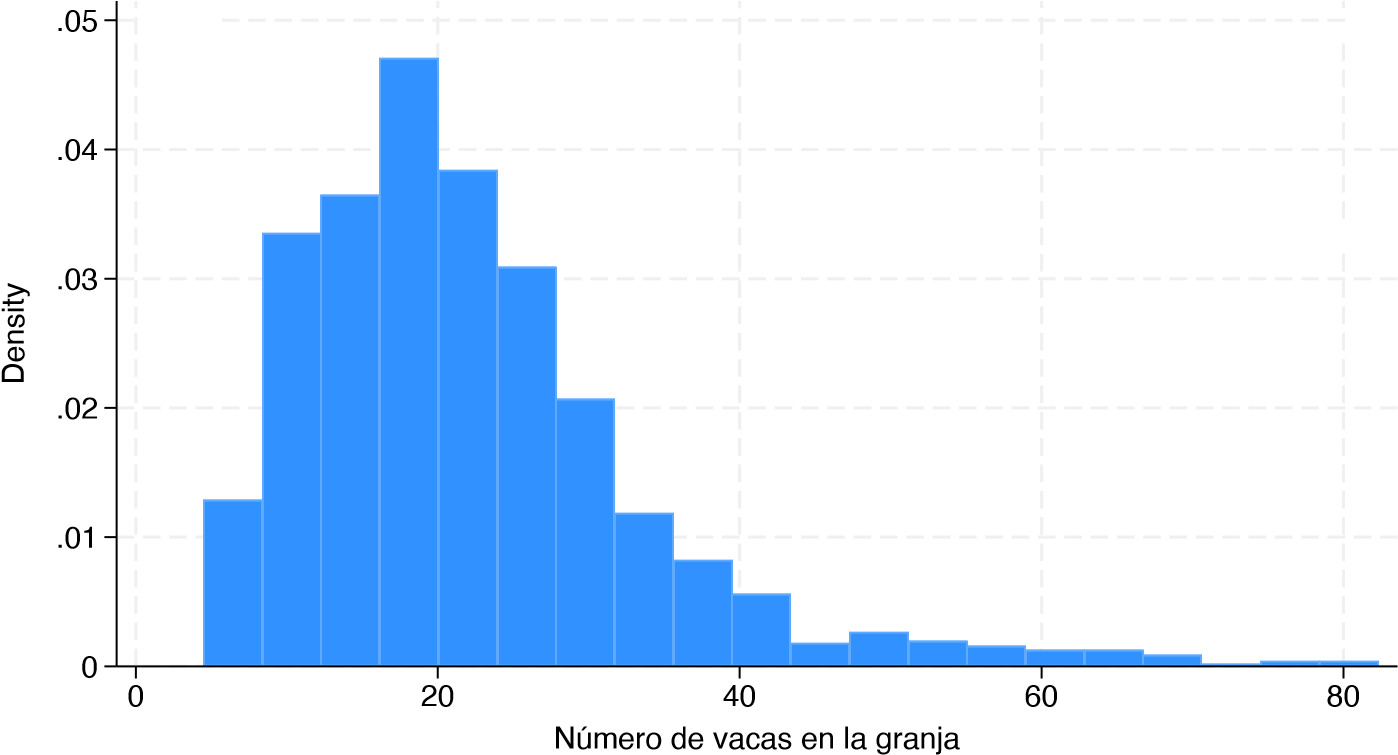
Note: Dataset has changed since last saved.

|  |
| --- |
| **%stata** histogram cows, bin(20) title("Distribución del número de vacas en las granjas") xlabel(, grid) ylabel(, grid)  **%stata** histogram land, bin(20) title("Distribución del área de tierra utilizada en las granjas") xlabel(, grid) ylabel(, grid)  **%stata** histogram labor, bin(20) title("Distribución del trabajo (horas) en las granjas") xlabel(, grid) ylabel(, grid)  **%stata** histogram feed, bin(20) title("Distribución de la cantidad de alimento en las granjas") xlabel(, grid) ylabel(, grid) |

In [ ]:

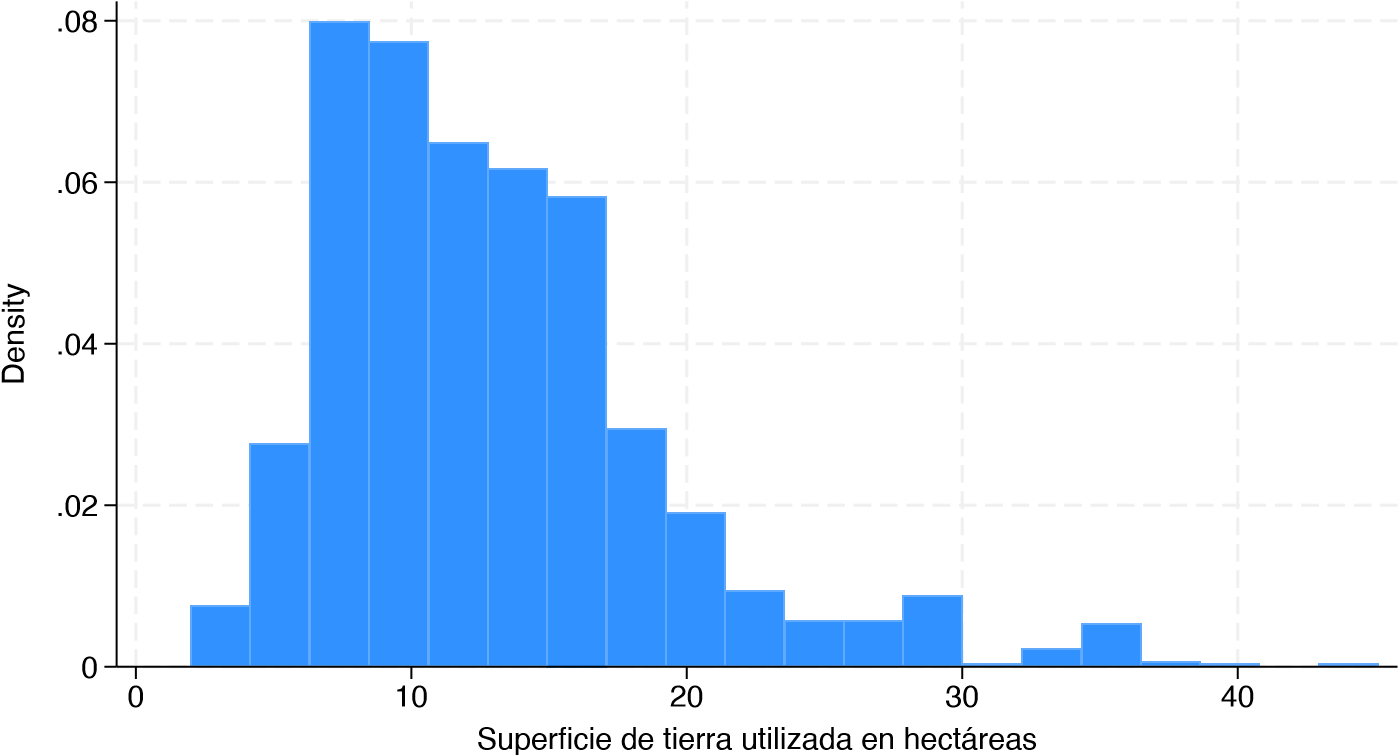
(bin=20, start=4.5, width=3.8900002)

## Distribución del número de vacas en las granjas



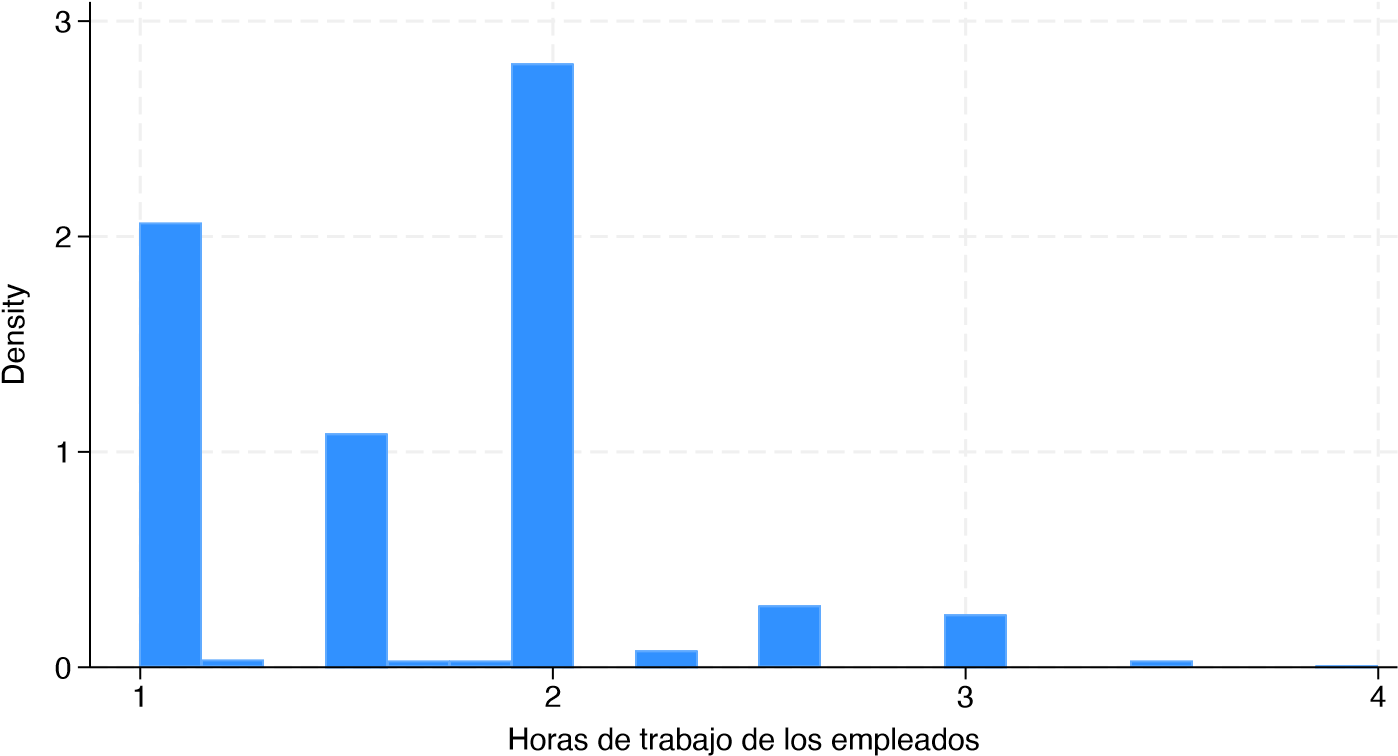
(bin=20, start=2, width=2.1549999)

## Distribución del área de tierra utilizada en las granjas



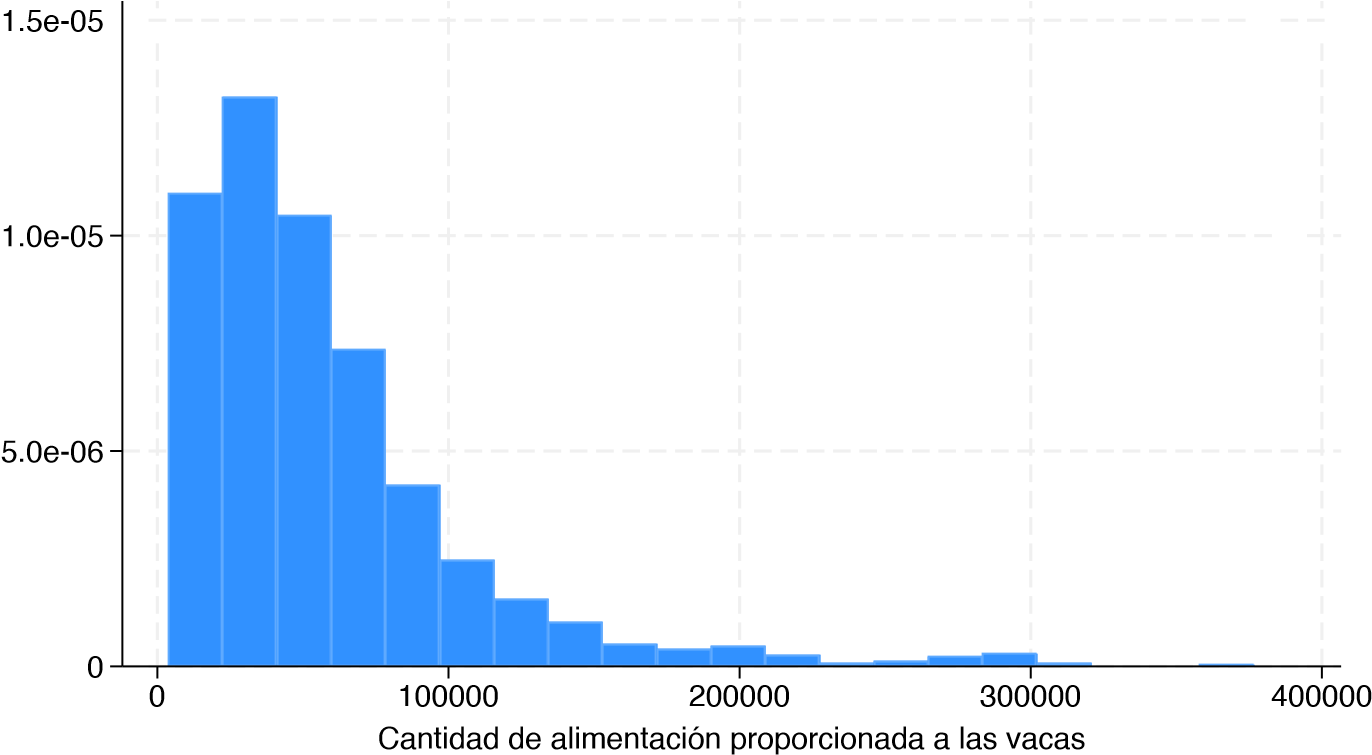
(bin=20, start=1, width=.15)

## Distribución del trabajo (horas) en las granjas



(bin=20, start=3924.1411, width=18640.371)

## Distribución de la cantidad de alimento en las granjas



Density

### 1. Distribución del Número de Vacas en las Granjas🐮

**Forma de la Distribución**: La distribución es asimétrica positiva (con sesgo a la derecha), lo cual significa que la mayoría de las granjas tienen un número relativamente bajo de vacas, mientras que pocas granjas tienen un número mas alto de vacas.

**Moda**: El máximo de la distribución se encuentra entre 15 y 25 vacas, lo que sugiere que la mayoría de las granjas tienen un número de vacas en este rango.

**Cola Derecha**: La cola derecha es larga, indicando que hay algunas granjas con un número significativamente mayor de vacas, aunque son pocas en comparación con la mayoría.

### 2. Distribución del Área de Tierra Utilizada en las Granjas⛰

**Forma de la Distribución**: La distribución del área de tierra también muestra un sesgo a la derecha, similar a la distribución del número de vacas. Esto indica que la mayoría de las granjas tienen una superficie relativamente pequeña, mientras que hay pocas con grandes extensiones de tierra.

**Moda**: La moda se encuentra entre 8 y 12 hectáreas, lo cual sugiere que la mayoría de las granjas utilizan una superficie en este rango. **Cola Derecha**: La cola es más larga, lo que implica que hay algunas granjas que usan una gran cantidad de tierra, aunque son pocas.

### 3. Distribución del Trabajo en las Granjas

**Forma de la Distribución**: La distribución es claramente asimétrica y discontinua, mostrando varios picos. Hay dos valores de densidad altos, correspondientes a las horas de trabajo en la granja, posiblemente indicando categorías comunes.

**Picos**:

El primer pico, correspondiente a un valor de alrededor de 1, podría indicar granjas con trabajo muy reducido o granjas pequeñas con pocos empleados.

El segundo pico, más alto, se encuentra alrededor de 2. Esto sugiere que muchas granjas tienen una asignación más regular de trabajo, que puede corresponder a la mayoría de las actividades requeridas.

**Anomalías**: La presencia de otros picos pequeños sugiere variabilidad en la cantidad de trabajo entre granjas. Esto podría deberse a diferencias en el tamaño de las granjas o a la estacionalidad del trabajo.

### 4. Distribución de la Cantidad de Alimento Proporcionada a las Vacas🥗

**Forma de la Distribución**: Esta distribución tiene un sesgo positivo muy marcado, lo que indica que la mayoría de las granjas proporcionan cantidades relativamente bajas de alimento, mientras que algunas pocas granjas proporcionan una cantidad considerablemente mayor.

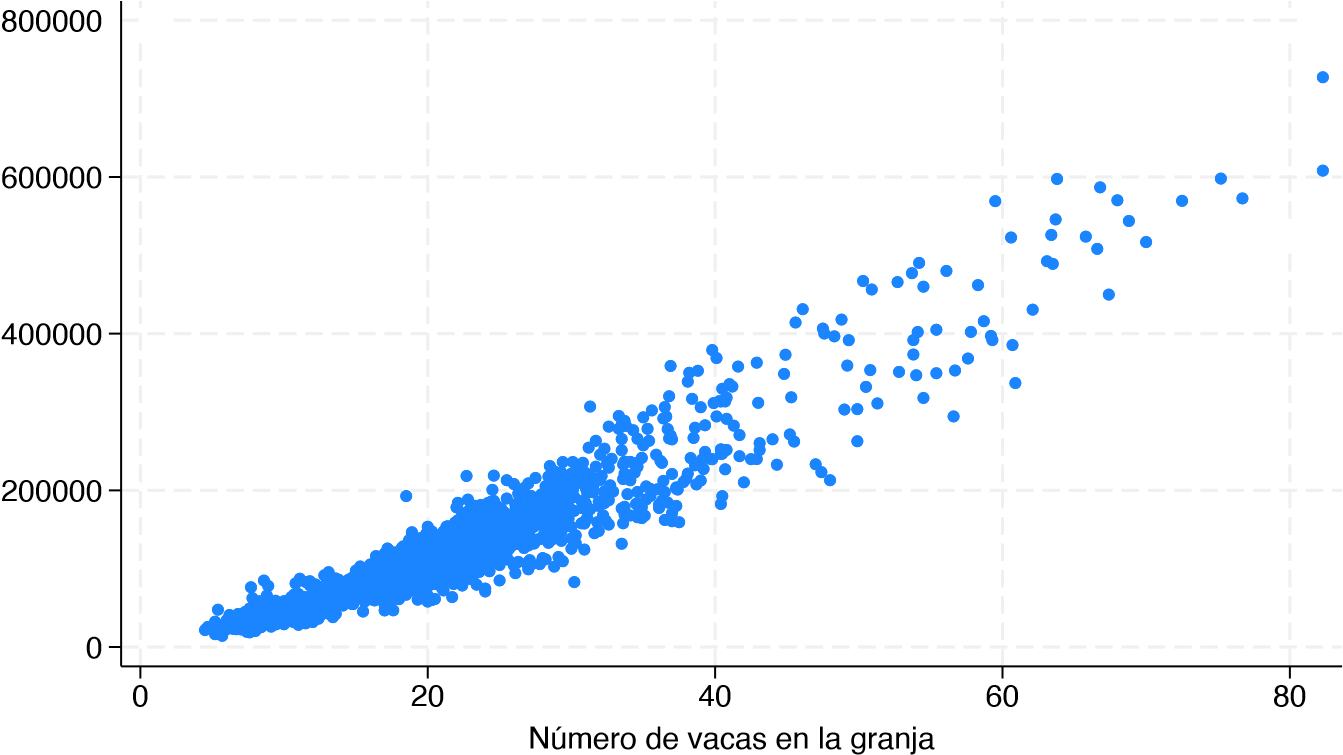
**Moda**: La mayor densidad de alimentación se encuentra en el rango más bajo, probablemente por debajo de 50,000 unidades. Esto sugiere que la mayoría de las granjas tienen una política de alimentación moderada.

**Cola Derecha Larga**: La cola derecha larga muestra que hay algunas granjas que utilizan mucha más alimentación. Esto podría deberse a que estas granjas tienen más vacas, son más intensivas, o manejan ganado con dietas más especializadas.

|  |
| --- |
| **%stata** scatter milk cows, title("Relación entre la producción de leche y el número de vacas")  **%stata** scatter milk land, title("Relación entre la producción de leche y el área de tierra")  **%stata** scatter milk feed, title("Relación entre la producción de leche y la cantidad de alimento") |

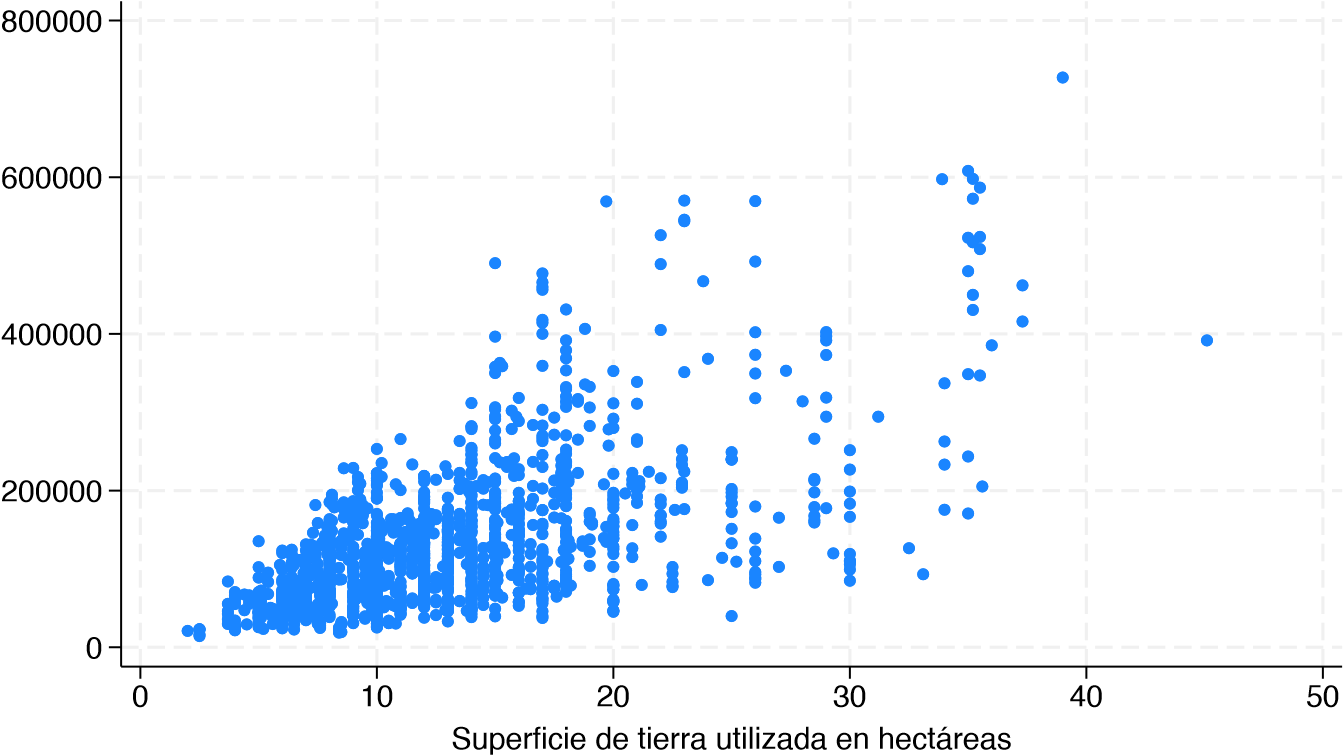
In [ ]:

## Relación entre la producción de leche y el número de vacas



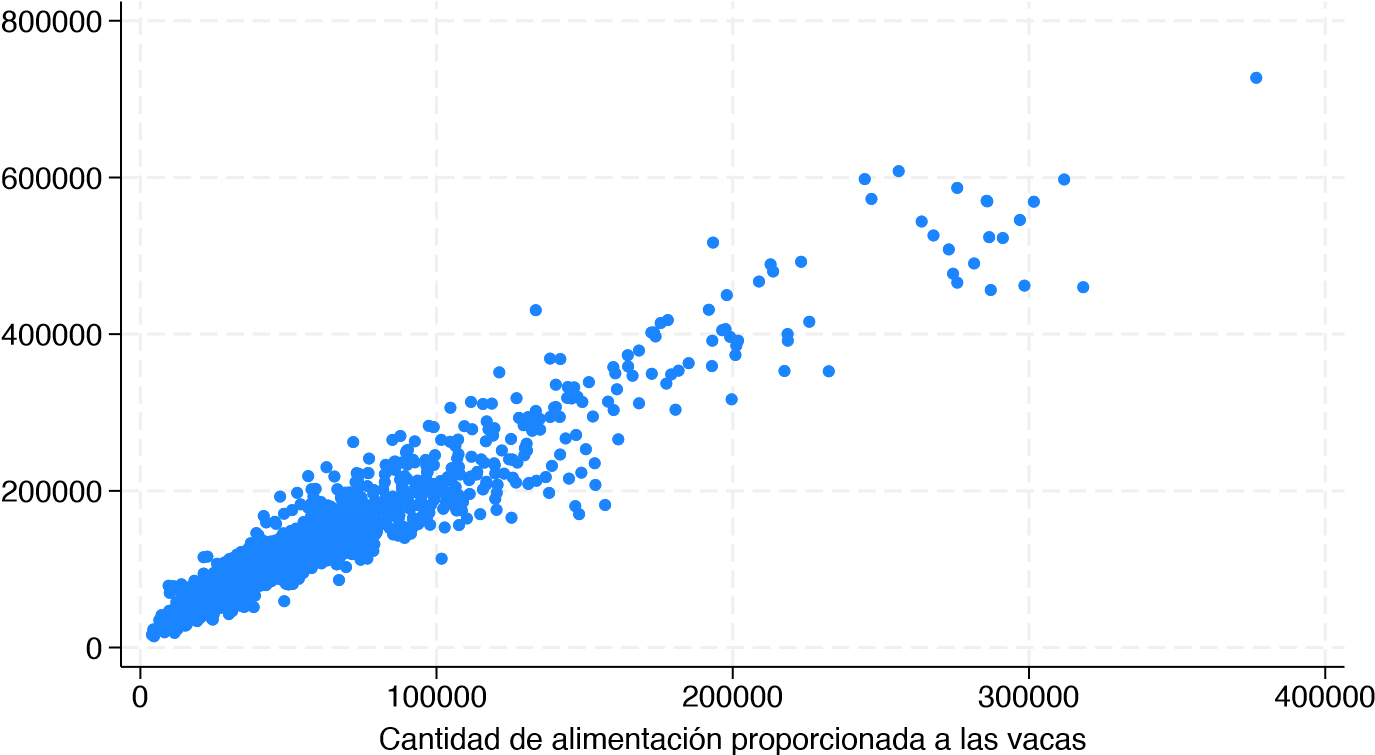
MILK

## Relación entre la producción de leche y el área de tierra



MILK

## Relación entre la producción de leche y la cantidad de alimento



MILK

### 1. Relación entre Producción de Leche y Número de Vacas 🥛🐮

**Correlación Positiva Fuerte**: El gráfico muestra una clara relación positiva entre el número de vacas y la producción de leche. A medida que aumenta el número de vacas, también lo hace la producción de leche. Esta correlación tiene una tendencia lineal, lo que indica que el número de vacas es un factor determinante y directo en la producción de leche.

**Dispersión**: Aunque la mayoría de los puntos siguen una tendencia bastante clara, hay cierta variabilidad en la producción de leche entre granjas con un número similar de vacas, lo cual podría deberse a factores adicionales, como la eficiencia de manejo, la calidad del alimento, o la salud del ganado.

### 2. Relación entre Producción de Leche y Área de Tierra Utilizada 🥛⛰

**Relación Positiva con Variabilidad**: También se observa una tendencia positiva en la relación entre la superficie de tierra utilizada y la producción de leche, aunque la correlación parece ser más dispersa comparada con el número de vacas. Esto sugiere que, aunque más tierra puede estar asociada con una mayor producción de leche, no necesariamente es el único factor, y hay más variabilidad en la producción con respecto a la cantidad de tierra.

**Tendencia No Lineal**: La producción de leche tiende a aumentar con más área de tierra, pero no de manera tan predecible como con el número de vacas. Podría ser que el tamaño de la tierra también dependa de cómo está siendo utilizada: por ejemplo, para pastos, almacenamiento, o incluso áreas que no contribuyen directamente a la producción.

### 3. Relación entre Producción de Leche y Cantidad de Alimento Proporcionado 🥛🥗

**Correlación Positiva Fuerte**: Este gráfico muestra una relación positiva muy fuerte entre la cantidad de alimento proporcionado y la producción de leche. A medida que aumenta la cantidad de alimento, también lo hace la producción de leche de manera casi proporcional. Esto sugiere que el alimento es uno de los principales factores que impulsa la producción de leche.

**Menos Variabilidad**: La dispersión de los puntos alrededor de la línea de tendencia es menor que en los otros gráficos, lo cual indica que la alimentación tiene un impacto muy predecible y consistente sobre la producción de leche. Esto hace que sea un factor clave para incrementar la producción de manera más controlada.

|  |
| --- |
| **%stata** cor milk cows land labor feed  **%stata** sum milk cows land labor feed |

In [ ]:

(obs=1,482)

| milk cows land labor feed

-------------+-------------------------------------------- milk | 1.0000 cows | 0.9466 1.0000 land | 0.6228 0.7022 1.0000 labor | 0.5403 0.5618 0.3561 1.0000 feed | 0.9575 0.9003 0.5609 0.5059 1.0000

Variable | Obs Mean Std. dev. Min Max

-------------+-------------------------------------------------------- milk | 1,482 131106.6 92583.98 14410 727281 cows | 1,482 22.11619 11.27415 4.5 82.3 land | 1,482 12.9857 6.172051 2 45.1 labor | 1,482 1.671997 .5527831 1 4 feed | 1,482 57941.35 47981.24 3924.141 376731.6

Variable | Obs Mean Std. dev. Min Max

-------------+-------------------------------------------------------- milk | 1,482 131106.6 92583.98 14410 727281 cows | 1,482 22.11619 11.27415 4.5 82.3 land | 1,482 12.9857 6.172051 2 45.1 labor | 1,482 1.671997 .5527831 1 4 feed | 1,482 57941.35 47981.24 3924.141 376731.6

### Análisis de Correlación y Estadísticas Descriptivas

La producción de leche se correlaciona fuertemente con el alimento ( 0.9575 ) y el número de vacas ( 0.9466 ). Tierra ( 0.6228 ) y trabajo ( 0.5403 ) tienen una correlación más baja con la producción. La cantidad de alimento y el número de vacas son los principales factores para aumentar la producción.

|  |
| --- |
| **%stata** regress milk cows land labor feed |

In [ ]:

Source | SS df MS Number of obs = 1,482

-------------+---------------------------------- F(4, 1477) = 7779.62 Model | 1.2120e+13 4 3.0299e+12 Prob > F = 0.0000

Residual | 5.7524e+11 1,477 389465851 R-squared = 0.9547

-------------+---------------------------------- Adj R-squared = 0.9546

Total | 1.2695e+13 1,481 8.5718e+09 Root MSE = 19735

----------------------------------------------------------------------------- milk | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+--------------------------------------------------------------- cows | 3644.012 129.804 28.07 0.000 3389.392 3898.632 land | -58.63673 120.1844 -0.49 0.626 -294.387 177.1136 labor | 2031.415 1123.863 1.81 0.071 -173.1226 4235.952 feed | 1.069114 .0252323 42.37 0.000 1.019619 1.118609 \_cons | -14066.09 1828.365 -7.69 0.000 -17652.56 -10479.63

------------------------------------------------------------------------------

|  |
| --- |
| **%stata** vif |

In [ ]:

Variable | VIF 1/VIF

-------------+--------------------- cows | 8.14 0.122792 feed | 5.57 0.179415 land | 2.09 0.477924 labor | 1.47 0.681362

-------------+----------------------

Mean VIF | 4.32

Medimos la colinealidad entre las variables por el VIF. Un VIF que supera los 10 nos señala colinealidad, sin embargo, en esta situación, todos los VIFs se encuentran por debajo de 10, lo que indica que no existe una multicolinealidad relevante entre nuestras variables.

Las vacas y el alimento son factores esenciales para comprender la producción de leche. La región no ejerce un efecto relevante en este modelo, mientras que la variable labor ejerce un efecto residual. El modelo presenta un adecuado ajuste (R-cuadrado = 0.9547), lo que indica que la regresión recoge la mayor parte de la variabilidad en la producción láctea, aunque existen posibles mejoras.

# Datos de Panel

|  |
| --- |
| **%stata** xtset farm year |

In [ ]:

Panel variable: farm (strongly balanced)

Time variable: year, 93 to 98 Delta: 1 unit

|  |
| --- |
| **%stata** xtreg milk cows land labor feed, fe  **%stata** estimates store fe\_model |

In [ ]:

Fixed-effects (within) regression Number of obs = 1,482

Group variable: farm Number of groups = 247

R-squared: Obs per group:

Within = 0.8759 min = 6

Between = 0.9642 avg = 6.0

Overall = 0.9536 max = 6

F(4, 1231) = 2172.03 corr(u\_i, Xb) = -0.0685 Prob > F = 0.0000

----------------------------------------------------------------------------- milk | Coefficient Std. err. t P>|t| [95% conf. interval] -------------+--------------------------------------------------------------- cows | 4110.793 174.6928 23.53 0.000 3768.065 4453.522 land | 162.0506 167.2217 0.97 0.333 -166.0204 490.1216 labor | 2815.646 2221.462 1.27 0.205 -1542.626 7173.917 feed | .9590173 .0267543 35.85 0.000 .9065282 1.011506 \_cons | -22187.36 4521.29 -4.91 0.000 -31057.65 -13317.07 -------------+--------------------------------------------------------------- sigma\_u | 16519.186 sigma\_e | 12346.393 rho | .64160039 (fraction of variance due to u\_i)

------------------------------------------------------------------------------

|  |
| --- |
| **%stata** xtreg milk cows land labor feed, re  **%stata** estimates store re\_model |

F test that all u\_i=0: F(246, 1231) = 10.34 Prob > F = 0.0000 In [ ]:

Random-effects GLS regression Number of obs = 1,482

Group variable: farm Number of groups = 247

R-squared: Obs per group:

Within = 0.8757 min = 6

Between = 0.9649 avg = 6.0

Overall = 0.9543 max = 6

Wald chi2(4) = 15458.17 corr(u\_i, X) = 0 (assumed) Prob > chi2 = 0.0000

----------------------------------------------------------------------------- milk | Coefficient Std. err. z P>|z| [95% conf. interval] -------------+--------------------------------------------------------------- cows | 3936.384 140.5181 28.01 0.000 3660.973 4211.794 land | 42.54122 140.092 0.30 0.761 -232.0341 317.1165 labor | 2262.435 1578.092 1.43 0.152 -830.5691 5355.439 feed | .9900496 .0241388 41.01 0.000 .9427385 1.037361 \_cons | -17651.26 2922.842 -6.04 0.000 -23379.93 -11922.6 -------------+--------------------------------------------------------------- sigma\_u | 15445.67 sigma\_e | 12346.393 rho | .61014682 (fraction of variance due to u\_i)

------------------------------------------------------------------------------

|  |
| --- |
| **%stata** xttest0 |

In [ ]:

Breusch and Pagan Lagrangian multiplier test for random effects milk[farm,t] = Xb + u[farm] + e[farm,t]

Estimated results:

| Var SD = sqrt(Var) ---------+---------------------------- milk | 8.57e+09 92583.98 e | 1.52e+08 12346.39 u | 2.39e+08 15445.67

Test: Var(u) = 0

chibar2(01) = 1345.08

Prob > chibar2 = 0.0000

|  |
| --- |
| **%stata** hausman fe\_model re\_model |

In [ ]:

Note: the rank of the differenced variance matrix (3) does not equal the number of coefficients being tested (4); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

---- Coefficients ----

| (b) (B) (b-B) sqrt(diag(V\_b-V\_B)) | fe\_model re\_model Difference Std. err.

-------------+--------------------------------------------------------------- cows | 4110.793 3936.384 174.4096 103.7895 land | 162.0506 42.54122 119.5094 91.30892 labor | 2815.646 2262.435 553.2107 1563.496 feed | .9590173 .9900496 -.0310323 .0115374

----------------------------------------------------------------------------- b = Consistent under H0 and Ha; obtained from xtreg. B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

chi2(3) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)

= 4.13 Prob > chi2 = 0.2479

En este estudio, hemos tenido en cuenta dos métodos para modelar los datos del panel: 1.los efectos fijos (FE) y 2.los efectos aleatorios (RE). Los efectos fijos (FE) presuponen que las variaciones entre las unidades (en esta situación, las granjas) son resultado de características no detectadas que permanecen inalterables a través del tiempo. Este modelo resulta beneficioso cuando se considera que las discrepancias no detectadas entre las unidades podrían estar vinculadas con las variables que explican el modelo. En contraposición, el modelo de efectos aleatorios (RE) sostiene que las variaciones entre las unidades son aleatorias y no guardan correlación con las variables explicativas. En otras palabras, se manejan los efectos no detectados como un elemento de error aleatorio.

La selección entre FE y RE se fundamenta en la relevancia de los efectos aleatorios, que se evaluó a través del test de Breusch-Pagan, logrando un valor p bastante bajo (0.0000), lo que señala que los efectos aleatorios son relevantes. Así pues, se opta por el modelo de efectos aleatorios (RE), dado que posibilita una eficacia superior en las estimaciones al evitar la adopción de una correlación directa entre los atributos no detectados y las variables del modelo.

La prueba de Hausman nos da un valor de chi2 de 15.78 y una probabilidad de 0.0456, lo cual significa que rechazamos la hipótesis nula. Esto quiere decir que las diferencias en los coeficientes entre los modelos de efectos fijos (FE) y efectos aleatorios (RE) no son aleatorias. Por lo tanto, el modelo de efectos fijos es el más adecuado para este análisis, ya que los resultados indican que el modelo de efectos aleatorios no es consistente.

( Añadir que: El test de Breusch-Pagan es un diagnóstico adicional que ayuda a confirmar o rechazar la importancia de los efectos aleatorios en el modelo, pero, si la prueba de Hausman indica que los efectos fijos son preferibles, lo más recomendable es seguir con el modelo de efectos fijos, incluso si el test de Breusch-Pagan sugiere lo contrario; de ahí nuestra conslusión. )

|  |
| --- |
| **%stata** xtreg milk cows land labor i.year, fe |

In [ ]:

Fixed-effects (within) regression Number of obs = 1,482

Group variable: farm Number of groups = 247

R-squared: Obs per group:

Within = 0.7724 min = 6

Between = 0.9171 avg = 6.0

Overall = 0.8996 max = 6

F(8, 1227) = 520.54 corr(u\_i, Xb) = -0.1582 Prob > F = 0.0000

----------------------------------------------------------------------------- milk | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+--------------------------------------------------------------- cows | 7878.819 177.9992 44.26 0.000 7529.602 8228.035 land | 254.9118 227.8185 1.12 0.263 -192.0451 701.8687 labor | 2706.876 3020.834 0.90 0.370 -3219.697 8633.448

| year |

1. | 4085.824 1517.234 2.69 0.007 1109.164 7062.483
2. | 10740.8 1533.057 7.01 0.000 7733.094 13748.5
3. | 12846.19 1591.418 8.07 0.000 9723.987 15968.39
4. | 13848.95 1643.148 8.43 0.000 10625.26 17072.64 98 | 17789.39 1714.461 10.38 0.000 14425.79 21152.99 |

\_cons | -60864.18 6072.751 -10.02 0.000 -72778.31 -48950.06

-------------+--------------------------------------------------------------- sigma\_u | 25421.17 sigma\_e | 16746.626 rho | .69736278 (fraction of variance due to u\_i)

------------------------------------------------------------------------------

F test that all u\_i=0: F(246, 1227) = 12.68 Prob > F = 0.0000

El modelo de efectos fijos (FE) indica que tanto la cantidad de vacas como el gasto en comida son los factores clave que influyen de manera positiva en la producción de leche. En cambio, el tamaño del terreno y la fuerza laboral no son factores relevantes en este modelo. Decir que los efectos anuales sí son significativos y evidencian un incremento en la producción láctea a través del tiempo.

El **Rho** del 65.55% sugiere que una parte significativa de la variabilidad en la producción de leche entre las granjas se debe a factores específicos de cada granja que NO hemos considerado en nuestras variables explicativas. Este valor refleja la influencia de características no observadas de las granjas.

Por otro lado, el **R² dentro del modelo de efectos fijos** es alto (88.05%), lo que indica que nuestras variables explicativas (como el número de vacas, la alimentación, la mano de obra y la superficie) explican muy bien la variabilidad en la producción de leche dentro de cada granja a lo largo del tiempo. Esto significa que nuestro modelo es eficaz para capturar los efectos de las variables observadas sobre la producción de leche dentro de las granjas.

➡ Entonces, aunque ciertamente nuestras variables explicativas son muy efectivas para modelar la producción de leche, aún existen factores específicos de cada granja no observados que influyen en los resultados.

# Series Temporales ⏳

|  |
| --- |
| **%stata** tsset farm year |

In [ ]:

Panel variable: farm (strongly balanced)

Time variable: year, 93 to 98 Delta: 1 unit

|  |
| --- |
| **%stata** gen L1\_cows = L.cows  **%stata** gen L2\_cows = L2.cows  **%stata** gen L3\_cows = L3.cows  **%stata** gen L4\_cows = L4.cows |

In [ ]:

(247 missing values generated)

(494 missing values generated)

(741 missing values generated)

(988 missing values generated)

|  |
| --- |
| **%stata** correlate milk L1\_cows L2\_cows L3\_cows L4\_cows |

In [ ]:

(obs=494)

| milk L1\_cows L2\_cows L3\_cows L4\_cows

-------------+-------------------------------------------- milk | 1.0000

L1\_cows | 0.9267 1.0000

L2\_cows | 0.9092 0.9764 1.0000

L3\_cows | 0.8898 0.9607 0.9821 1.0000

L4\_cows | 0.8615 0.9382 0.9633 0.9770 1.0000

Los resultados muestran que los lags de cows están bastante correlacionados con la variable milk, lo cual sugiere que hay una relación entre la producción de leche y los lags de las vacas.

|  |
| --- |
| **%stata** reg milk L1\_cows L2\_cows L3\_cows L4\_cows |

In [ ]:

Source | SS df MS Number of obs = 494

-------------+---------------------------------- F(4, 489) = 759.05 Model | 5.1042e+12 4 1.2760e+12 Prob > F = 0.0000

Residual | 8.2206e+11 489 1.6811e+09 R-squared = 0.8613

-------------+---------------------------------- Adj R-squared = 0.8601

Total | 5.9262e+12 493 1.2021e+10 Root MSE = 41001

----------------------------------------------------------------------------- milk | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+--------------------------------------------------------------- L1\_cows | 7783.1 739.9378 10.52 0.000 6329.251 9236.95

L2\_cows | 2480.868 1158.843 2.14 0.033 203.9413 4757.795

L3\_cows | 375.027 1201.02 0.31 0.755 -1984.769 2734.823

L4\_cows | -2076.562 887.4636 -2.34 0.020 -3820.274 -332.8492

\_cons | -47768.03 4239.328 -11.27 0.000 -56097.58 -39438.48 ------------------------------------------------------------------------------

Coeficientes y significancia:

L1\_cows: Coeficiente de 7783.1, t = 10.52, p-value < 0.001, lo que indica que el lag de 1 año de cows tiene un impacto significativo en la producción de leche. L2\_cows: Coeficiente de 2480.86, t = 2.14, p-value = 0.033, indica que el lag de 2 años de cows también es significativo, aunque con una significancia algo menor. L3\_cows: Coeficiente de 375.027, t = 0.31, p-value = 0.755, el lag de 3 años no es significativo.

L4\_cows: Coeficiente de -2076.56, t = -2.34, p-value = 0.020, el lag de 4 años es significativo.

R-cuadrado: nuestro 0.8613 sugiere que aproximadamente el 86.13% de la variabilidad en la producción de leche (milk) es explicada por los lags de cows. Prob > F =

0.0000 indica que el modelo es significativo en general.

Los lags de cows de 1, 2 y 4 años son significativos para predecir la producción de leche. El lag 3 no parece tener un impacto significativo en la predicción de la producción de leche.

**Bibliografia** 📚

[Datos](https://pages.stern.nyu.edu/~wgreene/Econometrics/PanelDataSets.htm)

[Documentación PyStata](https://www.stata.com/python/pystata18/install.html)

[Documentación NumPy](https://numpy.org/devdocs/)

[Documenatación Pandas](https://pandas.pydata.org/docs/)