



Departamento de Automática y Sistemas Computacionales

TRABAJO DE DIPLOMA

**Título: Relaciones entre datos de cosecha de combinadas
cañeras Case IH A8000**

Autor: Mathieu Badhel Ekoya Bouya

Tutores: MsC. Robby Gustabello Cogle

Ing. Carlos Alejandro Pérez García

Este documento es Propiedad Patrimonial de la Universidad Central “Marta Abreu” de Las Villas, y se encuentra depositado en los fondos de la Biblioteca Universitaria “Chiqui Gómez Lubian” subordinada a la Dirección de Información Científico Técnica de la mencionada casa de altos estudios.

Se autoriza su utilización bajo la licencia siguiente:

Atribución- No Comercial- Compartir Igual



Para cualquier información contacte con:

Dirección de Información Científico Técnica. Universidad Central “Marta Abreu” de Las Villas. Carretera a Camajuaní. Km 5½. Santa Clara. Villa Clara. Cuba. CP. 54 830

Teléfonos.: +53 01 42281503-1419

PENSAMIENTO

“A los hombres les encanta maravillarse, esto es la semilla de la ciencia”

Ralph Waldo Emerson

DEDICATORIA

A mi papá Léon Alexandre Bouya por ser mi ejemplo y por sostenerme tanto moralmente
como financieramente durante toda mi estancia en Cuba.

A mi querida Cendra Kidjo por ayudarme tanto moralmente como académicamente para
cumplir mis metas.

A mis padres por preocuparse siempre de mí y por poner mi vida en la mano de DIOS.

Al estado congoleño por darme la oportunidad de estudiar en Cuba.

AGRADECIMIENTOS

A mi familia por sus llamadas y sus mensajes cariñosos que fortalecieron mi motivación.

A mis tutores Robby Gustabello Cogle y Carlos Pérez García por sus consejos y su
paciencia.

A mis profesores, mis compañeros de clase y la relación internacional por apoyarme desde
el principio de la carrera hasta al final. Esto ha sido una buena experiencia.

A mis amigos por ayudarme a encontrar un equilibrio entre la redacción de la tesis y la vida
social.

RESUMEN

El proceso de recolección de la caña de azúcar ha evolucionado gracias a la cosecha mecanizada moderna. Las cosechadoras altamente automatizadas como CASE-IH A8000 vienen equipadas con dispositivos tecnológicos capaces de recopilar datos en el proceso agrícola. En el presente trabajo se determinan las relaciones entre los datos de cosecha de las combinadas cañeras CASE-IH A8000. Para ello se emplean los métodos: regresión lineal, correlación y análisis de componentes principales a través de los softwares *SMS Advanced*, *WEKA* y *SPSS*. Mediante la experimentación de 3 conjuntos de datos, se tributa a relaciones entre variables continuas de los datos citados.

Palabras claves: Variables, relación, agricultura, datos de cosecha.

ABSTRACT

The process of harvesting sugarcane has greatly evolved through mechanized agriculture. Highly automated harvesting machines, such as the CASE IH A8000, are equipped with technological devices capable of acquiring data in an agricultural process. The purpose of this work is to determine the relationship between harvesting data to improve agricultural practices. Therefore, we use statistical methods like correlation, regression and the principal component analysis (PCA) of three different software: SMS Advanced, WEKA and SPSS. Based on the experimentation from three datasets, we discover an association between most continuous quantitative variables.

Keywords: Variables, relations, agriculture, harvesting data.

TABLA DE CONTENIDOS

PENSAMIENTO	iii
DEDICATORIA	iv
AGRADECIMIENTOS	v
RESUMEN	vi
ABSTRACT.....	vii
INTRODUCCIÓN	1
CAPÍTULO 1. Herramientas Empleadas En El Análisis de Conjuntos de Datos.....	5
1.1 Desempeño de una cosechadora Case IH A8000.....	5
1.1.1 Cosecha mecanizada	5
1.1.2 Cosechadora CASE IH A8000	6
1.2 Variables de máquinas agrícolas y sus posibles relaciones	8
1.3 Estudio de relaciones en un conjunto de datos	9
1.3.1 Análisis exploratorio de datos.....	9
1.3.3 Correlación.....	10
1.3.4 Regresión	11
1.3.2 Análisis de componente principal.....	12
1.4. Software para el análisis de datos	17
1.4.1 <i>SMS Advanced</i>	17

1.4.2 <i>Paquete de estadística para las ciencias sociales</i>	18
1.4.3 <i>WEKA</i>	19
1.5 Consideraciones finales del capítulo.....	20
CAPÍTULO 2. Estudio de relaciones en los datos de cosecha.....	22
2.1. Datos de cosecha.....	22
2.2 Tratamiento de datos.....	26
2.3 Análisis de correlación.....	29
2.4 Análisis de regresión.....	31
2.5 Análisis de componentes principales	32
2.6 Consideraciones del capítulo 2	34
CAPÍTULO 3. ANALISIS DE RESULTADOS.....	35
3.1 Resultados de la correlación	35
3.2 Análisis de regresión.....	38
3.3 Análisis de componentes principales	40
3.4 Análisis económico y medioambiental	43
3.5 Consideraciones del capítulo 3	44
CONCLUSIONES Y RECOMENDACIONES	45
Conclusiones	45
Recomendaciones	46
REFERENCIAS BIBLIOGRÁFICAS	47
ANEXOS	51
Anexo I Fragmento de diagramas de dispersión de variables del conjunto 3.....	51
Anexo II Histogramas de variables del conjunto 3	52
Anexo III Comparación de datos transformadas y no transformadas con correlación	53

Anexo IV Estadística descriptiva de variables no transformadas	54
Anexo V Estadística descriptiva de variables transformadas	55
Anexo VI Análisis de correlación mediante	56
Anexo VII Análisis de correlación mediante <i>SPSS</i>	59
Anexo VIII Análisis de regresión mediante <i>WEKA</i>	62
Anexo IX Análisis de regresión mediante <i>SPSS</i>	64
Anexo XI Varianza Total Explicada del conjunto 1	68
Anexo XVI Matriz De Componentes Correlacionadas Con Variables Iniciales Conjunto3	73

INTRODUCCIÓN

El desarrollo sostenible de la agricultura en el mundo se ha vuelto realidad gracias a la modernización de la agricultura mecanizada. En efecto cada año se trata de mejorar la calidad de los productos y el rendimiento de la producción basándose en la observación, el análisis y la interpretación de grandes cantidades de informaciones generadas por los vehículos agrícolas modernos. Estos datos producen cambios enormes en las prácticas agronómicas tradicionales. Antes, la toma de decisión se elaboraba únicamente con el personal en el campo basándose en sus tradiciones. Actualmente, con el desarrollo de la ciencia y las tecnologías de la información, se requieren estrategias más elaboradas con el objetivo de satisfacer necesidades macroeconómicas y aumentar el índice de productividad agrícola. A esta modernización se le conoce como Agricultura de Precisión que según (Schimmelpfennig 2016), es una tecnología de la información en el área agrícola. Gracias a elementos como la cartografía de los suelos y la intervención de sistemas de geoposicionamiento, ella permite a los agricultores modificar sus prácticas de producción. Las informaciones brindadas por los elementos de la agricultura de precisión son: mapas de rendimiento, características del suelo, entre otras. Esta información se obtiene gracias a los sistemas automatizados implementados en los equipos agrícolas, las fotos aéreas o de forma manual por muestreo del suelo (Jallas and Crétenet 2003). De acuerdo con (Valero Ubierna, Navas Gracia et al. 2010), para entenderlo mejor, se puede definir como el hecho de “efectuar la intervención agrícola correcta, en el momento adecuado y en el lugar preciso”. Según la misma fuente, la agricultura de precisión se resume a un ciclo realimentado (Figura 1) donde las correcciones y las mejoras en los procesos agrícolas se hacen a partir del análisis de la información producida por las máquinas agrícolas y otros dispositivos

tecnológicos afiliados en este proceso. Esta información sirve de base para modular la aplicación de insumos teniendo en cuenta la variabilidad del terreno.



Figura 1 Ciclo del agricultura de precisión

Sobre la base de esto y la mejora de otros elementos como la irrigación inteligente y la genómica de precisión (importante para la producción de plantas más saludables y resistentes a enfermedades) se puede considerar que la cantidad de alimentos agrícolas está creciendo, aunque las condiciones climáticas están deteriorándose. Sin embargo, según (McGuire 2015), el 10% de la población mundial sufre de hambre. Este número parece muy insignificante, pero representa más o menos 800 millones de personas. Otra fuente afirma que los datos de la FAO indican que hay alimentos suficientes para erradicar el hambre en nuestro planeta, pero el problema reside en el poder de compra de alimentos, es decir, hay gente que no tiene suficiente dinero para alimentarse (Zamilpa, Rindermann et al. 2016).

En Cuba, las mayores aproximaciones a la agricultura de precisión han tenido lugares en la explotación de la caña de azúcar. A partir de 2014 el país importó diversas gamas de vehículos altamente automatizados de la compañía brasileña Case IH. Con esta iniciativa

se pretende elevar la producción y disminuir costos relacionados con esta última. Las cosechadoras cañeras han sido uno de los elementos que han favorecido el fortalecimiento de alianzas entre el Grupo Empresarial AZCUBA, la Estación Territorial de Investigaciones de la Caña de Azúcar (ETICA) y la Universidad Central “Marta Abreu” de Las Villas (UCLV).

Bajo su amparo, se han realizado varias acciones satisfactorias fundamentalmente desde el punto de vista investigativo. Entre ellas, se pueden mencionar:

- La caracterización y la descripción del sistema de automatización existente en la cosechadora cañera (Molina Hernández 2015).
- Una metodología para la extracción de los datos de cosecha del monitor de rendimiento ubicado en la combinada y el manejo de datos en cuestión a partir de softwares especializados (Pérez García 2016).
- Una metodología para la generación de informes de cosecha (Sánchez Ibáñez 2017).

Con tales antecedentes y el incremento del número de cosechadoras en la provincia y el país, así como la necesidad de profundizar en el estudio del comportamiento de las máquinas se enfrenta **el problema científico** que da lugar a la presente investigación. Este consiste en el desconocimiento de las relaciones existente entre los datos de cosecha extraídos de la combinada cañera CASE-IH A8000. Para aportar una solución a este problema se traza el siguiente objetivo general:

Objetivo general

Determinar la relación existente entre las variables de cosecha de la combinada cañera CASE-IH A8000 para usarla como base en la obtención de un posible modelo de operación.

Objetivos específicos

1. Revisar la bibliografía científica para tener un conocimiento sobre las variables importantes de las máquinas agrícolas modernas.
2. Analizar métodos y herramientas empleados para la determinación de relaciones entre conjuntos de datos.
3. Caracterizar los datos de cosecha adquiridos por la CASE-IH A8000.

4. Aplicar los métodos y herramientas disponibles para la determinación de las relaciones de dependencia.
5. Analizar los resultados obtenidos.

Organización del informe

El informe de la investigación se estructura en introducción, 3 capítulos, conclusiones, referencias bibliográficas y anexos. Los diferentes temas que se abordan en cada capítulo se encuentran estructurados de la forma siguiente:

- En el primer capítulo se procede a una revisión bibliográfica sobre la combinada cañera y sobre las variables y sus relaciones. Luego se habla de las técnicas estadísticas utilizadas en la literatura para relacionar variables existentes en un conjunto de datos. Al final se presentan diferentes programas para facilitar el manejo de dichos datos.
- En el segundo capítulo se caracterizan los datos de cosecha. Luego, se propone una metodología para relacionar las variables que cumplan con los requisitos del análisis estadísticas gracias a la ayuda del software de la agricultura de precisión *SMS Advanced* y del software estadístico *IBM SPSS* y *WEKA*.
- En el último capítulo se presentan los resultados obtenidos del análisis estadístico y se analizan dichos resultados.

Por fin, se concluye el trabajo dando los aspectos relevantes de la investigación y se presenta un grupo de recomendaciones para futuras investigaciones.

CAPÍTULO 1. Herramientas Empleadas En El Análisis de Conjuntos de Datos

Este capítulo está dividido en 4 partes principales. En la primera parte se realiza una búsqueda bibliográfica sobre el proceso de formación y de adquisición de datos producidos por la cosechadora para que el lector pueda entender la fuente de los datos. Luego en la segunda y tercera parte se hace una revisión bibliográfica sobre las variables de máquinas agrícolas modernas y se presentan métodos utilizados en la literatura para el análisis de conjuntos de datos. Al final se proponen programas utilizados para el análisis de datos.

1.1 Desempeño de una cosechadora Case IH A8000

1.1.1 Cosecha mecanizada

El proceso de corte de las cosechadoras cañeras CASE IH A8000 comienza por sacar las puntas con el despuntador, luego interviene el divisor de líneas y el rodillo tumbador para direccionar las cañas hacia el cortador de base. Cuando se finaliza el corte basal, las cañas se mueven dentro de la máquina a través de los rodillos alimentadores y el picador pica estas cañas en trozos de 30 a 40 cm y los deposita en el cesto del elevador, donde se va a realizar una primera limpieza soplando tierra e impurezas con un ventilador montado en el extractor primario. Al final, la caña se dirige hacia el segundo extractor para una segunda limpieza y para la descarga del producto en un sistema de transbordo (Marcasso 2017). Este sistema de transbordo es conectado a un tractor y llevará la materia prima a lo largo del proceso de cosecha (Figura 1.1).

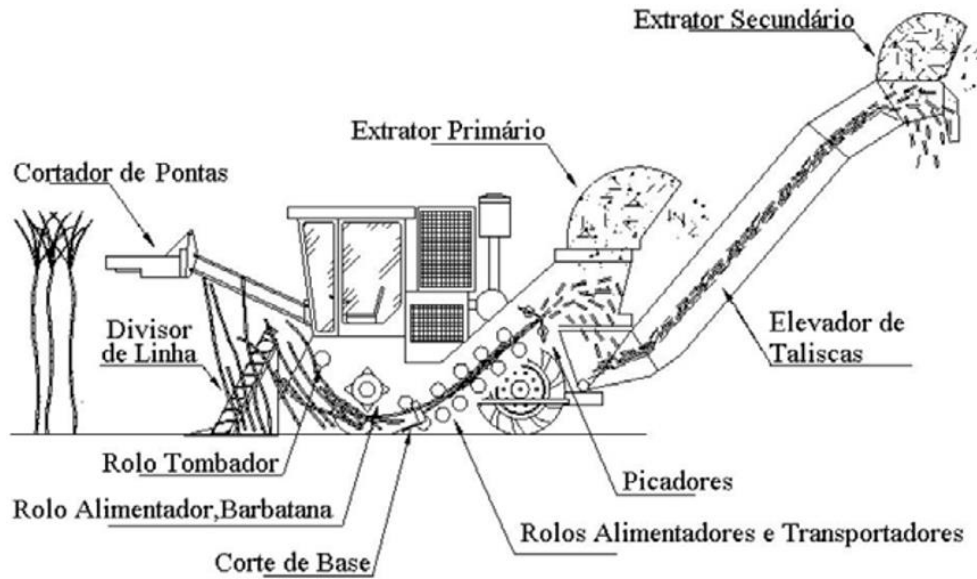


Figura 1.1 Parte del proceso

1.1.2 Cosechadora CASE IH A8000



Figura 1.2 Cosechadora cañera Case IH

La cosechadora cañera A8000 dispone de dos modelos (Figura 1.2), el primero tiene ruedas similares a los tractores, mientras que el segundo A8800 cuenta con estera *Heavy duty* D5M sellada y lubricada. La elección entre las dos depende fuertemente del tipo de terreno que se va a labrar. Sin embargo, el sistema de automatización es el mismo. Esta máquina está diseñada para garantizar un manejo sencillo y una alta comodidad al operario.

En su trabajo de investigación, (Molina Hernández 2015) hace un análisis completo del sistema de automatización de la cosechadora cañera Case IH A8000, enfatizando los dispositivos e instrumentaciones relacionados con el área del control automático. En base a estas potencialidades descubiertas, él lo considera como un sistema digital de control distribuido y reafirma que la cosechadora es apta y tiene prestaciones para la realización de la agricultura de precisión.

Según el mismo autor, esta cosechadora cuenta con un alto grado de automatización. Los principales componentes que la conforman son:

- Sensores
- Actuadores
- Control de máquina
- Mano derecha
- Red CAN
- Monitor de rendimiento
- Receptor GPS
- Unidad de control de motores

Los dispositivos tales como el GPS, los sensores, la red Can y el monitor de rendimiento forman parte del proceso de formación y de adquisición de los datos de cosecha. Los sensores se encuentran al nivel más bajo del sistema de automatización y están en contacto con los procesos, sensando y transformando magnitudes en otra señal proporcional, capaz de ser utilizada para la descripción de un proceso determinado. De otro lado, la red Can es la encargada de transmitir la información de los sensores hacia el monitor de rendimiento para la visualización de diversas variables y su posible extracción para un posterior análisis. De hecho (Pérez García 2016), propone una estrategia para sacar los datos de cosecha en la cosechadora por la vía del monitor de rendimiento, para luego importarlos en dos programas de agricultura de precisión (*AFS Desktop* y *SMS Advanced*). El autor describió paso a paso el proceso de extracción e importación de datos mediante un dispositivo de almacenamiento.

1.2 Variables de máquinas agrícolas y sus posibles relaciones

Sobre las variables y sus posibles relaciones, pocos autores se han expresado. Entre los trabajos vinculados con esta temática se pueden mencionar:

(Lyra 2012) analizó el comportamiento del consumo del combustible con respecto a la velocidad y la rotación del motor (RPM) en dos cosechadoras de caña diferente (Case IH A7700 y Case IH A8800), en diferente intervalo de velocidad (5-6,6-7,7-8 km/h) y con variación de la rotación del motor (1900,2000 y 2100 RPM) utilizando dos flujómetros volumétricos instalados entre el filtro y la bomba y en el retorno del combustible al tanque respectivamente. Se pudo determinar que la característica del consumo del combustible tenía una cierta relación con la velocidad y el régimen (RPM).

(González Valdés, Rodríguez López et al. 2010) hicieron un análisis de consumo de combustible sobre vehículos agrícolas (tractores) usando datos muestrales. Se aplicaron técnicas estadísticas de correlación y regresión lineal simple sobre el consumo del combustible contra la carga del motor, el consumo del combustible contra la fuerza del motor y la fuerza del motor contra la carga utilizando el programa estadístico *STATGRAPHICS plus 5.0*. Sus resultados fueron satisfactorios ya que encontraron coeficientes de correlación mayor que 0.9 en las tres combinaciones posibles indicando una fuerte relación entre dichas variables. Se pudo generar también graficas de regresión y ecuaciones de los modelos en las tres combinaciones

(Voltarelli 2015) considera que el alto consumo horario ha sido siempre un problema en tiempo de cosecha de caña de azúcar, ya que cuando la cosechadora está operando, el consumo del combustible se avecina a los 50 o 60 L/h.

La velocidad de cosecha de caña de azúcar durante la zafra es una variable decisiva para el proceso de cosecha. Ella se ajusta en dependencia del terreno que se va a cosechar y puede alcanzar un valor máximo de 9 km/h. En efecto, la característica del suelo y el operador, pueden influenciar a la hora de aplicarla, siendo esta velocidad directamente proporcional a la productividad y la calidad de caña cortada (Belardo 2010; Testa 2014).

(Pereira Filho and Rodrigues 2015) afirman que la productividad se puede calcular como:

$$Productividad(ha/h) = \frac{velocidad(m/h) * Ancho(m) * Factor\ del\ campo}{10000m^2}$$

Donde el valor del factor del campo en el tiempo de cosecha varía entre 50% y 70%.

1.3 Estudio de relaciones en un conjunto de datos

Según (Stengel 2017), la manipulación de datos para los resultados favorables no se limita solo a la observación. En efecto, para comprender mejor la utilidad de dichos datos, hay que analizarlos. El estudio de conjuntos de datos implica el uso de técnicas del dominio de la ciencia como la estadística para analizar e interpretar datos para luego caracterizar mejor un fenómeno.

1.3.1 Análisis exploratorio de datos

Generalmente, en la estadística se realiza un análisis exploratorio de datos (EDA por sus siglas en inglés) antes de hacer cualquier análisis cuantitativo en un conjunto de datos multivariado. El EDA facilita la comprensión de datos y determina si los datos respetan las hipótesis para la estadística inferencial (correlación, regresión, etc.) (Rojo 2006).

El EDA tiene como meta ilustrar las semejanzas, divergencias y relaciones existentes en un conjunto voluminoso de datos a partir de las características que tienen sus atributos. Por tanto, esta técnica es de gran utilidad para la visualización de patrón de distribución en un conjunto de datos, no obstante, no llega concretar una explicación clara sobre la existencia de dichos patrones (Leech, Barrett et al. 2013).

El EDA es fácil de abordar ya que el análisis se centra sobre la representación gráfica de datos y la estadística descriptiva dígame mediana, varianza, desviación estándar entre otros. Ilustrando datos de manera visual, se puede desarrollar capacidad de reconocimiento de curvas para sacar conclusiones sobre las tendencias del conjunto bajo estudio.

De una forma general, se puede considerar el EDA como un análisis de datos que usa diversas técnicas para:

- Tener un conocimiento previo sobre las variables
- Extraer variables más significativas

- Detectar valores extremos y posibles anomalías
- Presentar hipótesis sobre el comportamiento de las variables
- Seleccionar mejor las herramientas y las técnicas estadísticas

1.3.3 Correlación

La correlación es un método estadístico que permite determinar el grado de asociación existente entre dos variables. Se expresa por el coeficiente de correlación que permite calificar esta relación y también de un diagrama de dispersión que da una representación gráfica. El diagrama mapea los puntos de intersección entre las variables relacionadas. Este conjunto de puntos forma una nube, cuya forma permite juzgar la intensidad y el sentido de la relación. El coeficiente de correlación varía entre -1 y 1 y la intensidad de la relación se clasifica en fuerte, débil y nula, por mencionar algunos. Este coeficiente puede ser de Spearman o de Pearson. Ambos tienen prácticamente las mismas características. La única diferencia es que el de Spearman se recomienda cuando se quiere analizar una relación monótona donde las dos variables tienden a moverse en una dirección, sin necesidad de tener una relación lineal (Figura 1.3)(Aguilar 2019).

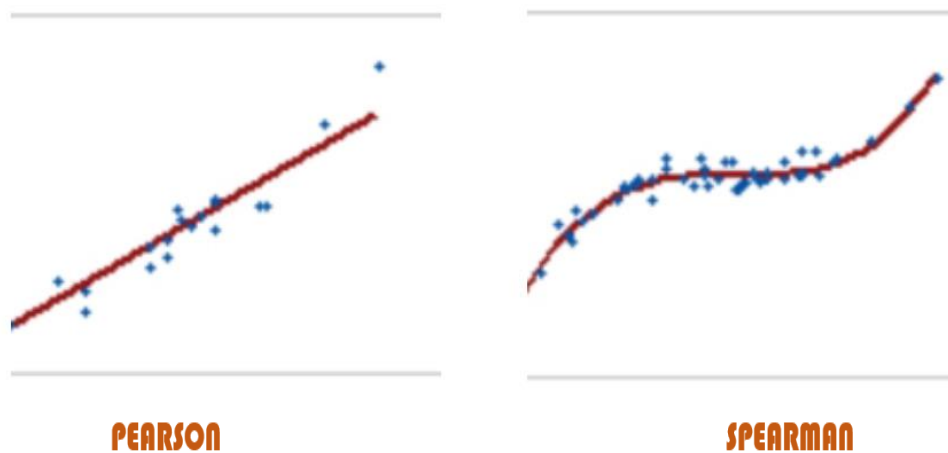


Figura 1.3 Tendencia de la correlación de Pearson y Spearman

El coeficiente de Pearson se determina por: $\rho = \frac{cov(x,y)}{desv(x)*desv(y)}$

Donde $cov(x, y)$ es la covarianza entre la variable y y x mientras que $desv(x)$ y $desv(y)$ son las desviaciones estándar de las variables y y x .

Según (Mukaka 2012), la regla de oro que siempre se ha utilizado para la calificación del coeficiente de correlación es el siguiente (Tabla 1):

Tabla 1

Clasificación del coeficiente de correlación

Coeficiente de correlación	Interpretación
0.90 a 1.00 (-0.9 a -1.00)	Muy fuerte
0.70 a 0.90 (-0.70 a -0.90)	Fuerte
0.50 a 0.70 (-0.50 a -0.70)	Moderado
0.30 a 0.50 (-0.30 a -0.50)	Débil
0.00 a 0.30 (-0.00 a -0.30)	Despreciable

1.3.4 Regresión

La regresión lineal se expresa por la ecuación: $y = a + bx$; donde los coeficientes a y b son la intercepción de la recta y la pendiente respectivamente. Con esta ecuación, se pueden predecir los valores de y a partir de los valores de x . El coeficiente de regresión “ R^2 ” explica la variancia de la variable dependiente. Este coeficiente varía entre 0 y 1 o en porcentaje (0 a 100). Cuanto más tiende el coeficiente a 1 mejor es la predicción. Esta regresión puede ser lineal simple o múltiple (multivariable).

El parámetro a representa la media de la variable y (\bar{Y}) y el coeficiente b se calcula como:

$$b = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Generalmente, para la realización del análisis de regresión, los pares analizados deben cumplir con las pruebas de linealidad y de homocedasticidad. La linealidad es una prueba

que afirma que se puede trazar una línea recta para describir la relación entre un par de variables, mientras que la homocedasticidad comprueba la variación de una variable respecto a otra. Esto se logra saber observando los diagramas de dispersión. Generalmente la nube de punto debe ser oval (Figura 1.5)(Rocheftort 2009).

1.3.2 Análisis de componente principal

A veces, las variables encontradas en un conjunto de datos son elevadas. Ello complejiza el análisis y aumenta las imprecisiones sobre los resultados esperados. Por tanto, resulta útil reducir la dimensionalidad del proceso, condensando datos redundantes que se pueden encontrar en los conjuntos de datos sin perder muchas informaciones. El análisis de componentes principales (PCA, por sus siglas en inglés) es una de las técnicas más antiguas sobre el análisis de datos multivariantes. Fue creada en 1901 por Karl Pearson y luego fue integrada en la disciplina de estadística en 1933 por Harold Hotelling. Debido a su complejidad, su uso se hizo popular cuando apareció el medio de cómputo (Bencherki and Moustafa 2018) .

El PCA es una técnica estadística de exploración de datos avanzado que se ha usado frecuentemente en muchas ramas de la ciencia. Se revela de gran utilidad cuando se quiere conocer la estructura de un conjunto de datos, sintetizarlo y resumirlo, ya que se estima que la eficiencia de un modelo de inteligencia artificial depende de la calidad de los datos (Delgado-Trejos, Piedrahita et al. 2006; Jaouad, Mahjoubi et al. 2015).

El PCA puede aplicarse sobre conjuntos de datos cuantitativos de tipo matricial, donde las filas son observaciones y las columnas son variables. Según (Brahimi 2011) este método lleva:

- Un análisis rápido de relación mediante matrices de correlación entre variables
- Un análisis de datos en toda su forma (tabla, diagrama de dispersión en dos o tres dimensiones)
- Una construcción de un conjunto de factores no correlacionados donde las nuevas variables representan unas combinaciones lineales de variables originales.

La calidad de estas matrices de correlación puede cuantificarse mediante el índice de Kaiser-Meyer-Olkin y las pruebas de Bartlett. El índice de KMO varía entre 0 y 1. Cuanto más tiende a 1 mejor es la calidad de las relaciones entre variables. Por otro lado, el índice de Bartlett debe respetar una probabilidad inferior a 0.10 (Cerny and Kaiser 1977; Le Moal 2002). Estos índices se ven afectados por la cantidad de variables, de observaciones y la cantidad de los factores.

La esencia de esta técnica tiene un vínculo con el análisis geométrico y el álgebra lineal. Supóngase una matriz de datos rectangular X donde se observan las medidas de “p” variables con respecto a “n” individuos:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Considerando el análisis con dos variables iniciales X_1 y X_2 , lo primero que se hace es representar el diagrama de dispersión de dichas variables. Se observa una nube de puntos (Figura 1.4) que describe una cierta tendencia de relación (Villardón 2002).

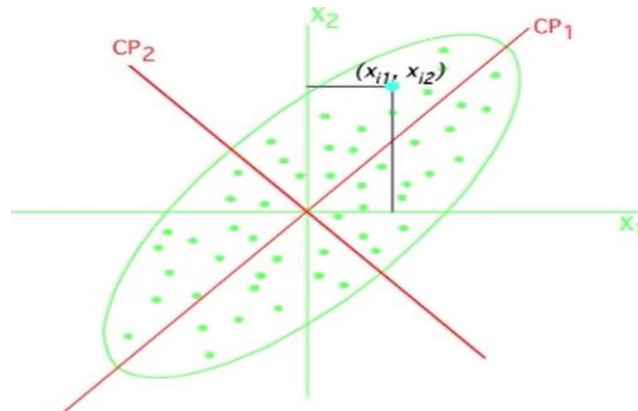


Figura 1.4 Tendencia de la Nube de puntos

Lo que se quiere obtener es un diagrama más simple y representativo que contenga la mayor cantidad de la información del conjunto de datos inicial y que tenga menor dimensionalidad. Esto se logra haciendo un cambio de coordenada donde las nuevas variables llamadas componentes principales, recogerán la mayor parte de la variabilidad del sistema. En la (Figura 1.5) se ve claramente que el sentido del primer componente debe recoger la mayor parte de la variabilidad. El segundo componente siempre será ortogonal al

primero y recogerá la mayor parte de la variabilidad restante. El número de componentes seguirá creciendo ortogonalmente en función de la dimensionalidad y de la cantidad de variables iniciales (Villardón 2002).

Sin embargo, esta tendencia no siempre se aprecia. Por lo que es evidente realizar una formulación matemática para detectar el sentido de los componentes mediante los vectores (Figura 1.5).

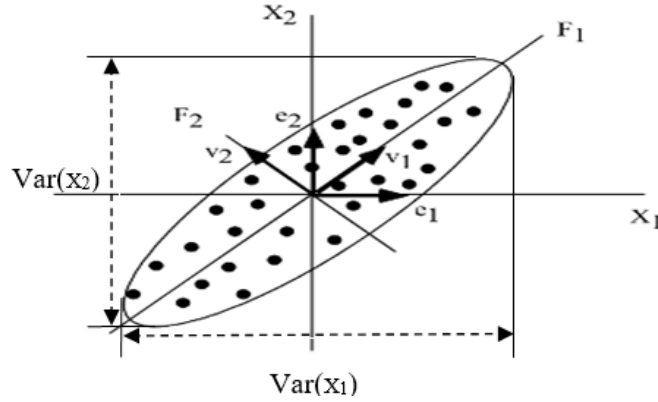


Figura 1.5 Descripción del diagrama

Estos vectores (V) se calculan en función de la matriz de covarianza y de los valores propios (Λ) asociados a cada componente según:

$$\begin{pmatrix} Var(x_1) & Cov(x_1, x_2) \\ Cov(x_1, x_2) & Var(x_2) \end{pmatrix} V = \Lambda V$$

Donde: $var(x)$ es la varianza de la nube de puntos en el eje x e y respectivamente y

$cov(x_1, x_2)$ es la covarianza de la nube de puntos.

Teniendo la matriz de covarianza, se puede calcular fácilmente los valores propios, para luego determinar los vectores propios.

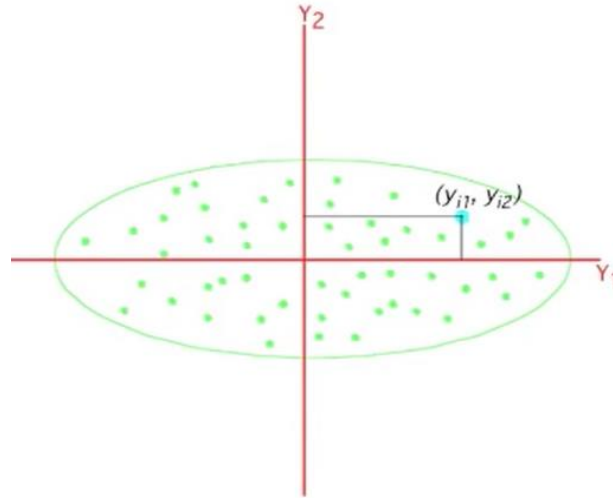


Figura 1.6 Cambio de coordenadas

Puesto que se han cambiado las coordenadas (Figura 1.6), se debe referir los puntos con respecto a los componentes principales tomando dichos componentes como combinaciones lineales de las variables iniciales (Villardón 2002).

$$y_{i1} = x_{i1}v_{11} + x_{i2}v_{12}$$

$$y_{i2} = x_{i1}v_{21} + x_{i2}v_{22}$$

Ahora, si se quiere conocer las nuevas observaciones de cada componente, se debe tener en cuenta las observaciones de todas las variables iniciales x_{ij} y de los coeficientes v_{ij} que son vectores en la dirección de los componentes.

En cuanto a la reducción de la dimensionalidad, tres criterios se toman en cuenta (Le Moal 2002):

- El criterio de Kaiser que consiste en que los valores propios inferiores a 1 son despreciables y no se consideran los componentes asociados a ellos.
- Si hay una cantidad de componentes que recogen al menos 80% de la información inicial en un porcentaje de varianza acumulada, se puede considerar como buena aproximación del conjunto inicial y se puede ignorar el resto de las componentes.

- La prueba del codo, donde se observa la gráfica de los valores propios y se consideran solo los valores propios que se encuentran del lado izquierdo del punto de inflexión. De forma gráfica, se traza una recta alrededor de los puntos que están alineados y se consideran solo los valores propios que se encuentran por encima de la recta (Figura 1.7).

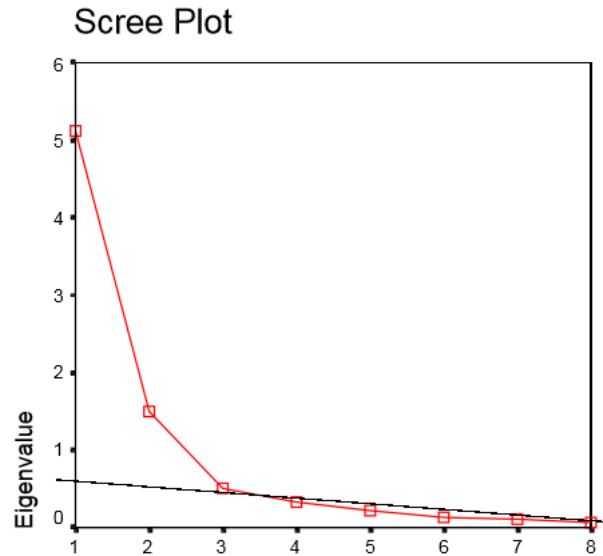


Figura 1.7 Criterio del codo

De forma general, si se quiere realizar el PCA en un conjunto de datos de dimensionalidad superior, se puede tener una matriz $Y=XV$.

$$\begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1p} \\ \vdots & \ddots & \vdots \\ v_{p1} & \cdots & v_{pp} \end{pmatrix}$$

Los aspectos a tener en cuenta tienen como expectativa optimizar los resultados (de la Maza 2017):

1. Si las variables son de diferentes unidades se debe proceder a la normalización o transformación.
2. Si las variables tienen las mismas unidades se puede:
 - Elegir solo variables que tienen mayor variabilidad o

- Transformar todas las variables y considerarlas con el mismo peso.

Generalmente, estas técnicas de normalización se dedican a transformar con fórmulas muy simples de translación, de cambio de escala y rotaciones, con el objetivo de homogeneizar las varianzas del conjunto de datos. Varios métodos se proponen para transformar variables (la raíz cuadrada, el logaritmo y el inverso). Uno de los problemas es que estas técnicas de transformación comprimen mucho la parte derecha de la distribución, pero hay varias técnicas para paliar este asunto (Díaz Blanco 2000; Rodríguez Ayán and Ruiz Díaz 2008).

Dado que las técnicas estadísticas de correlación, regresión y de exploración de datos implican un trabajo muy fuerte de matemática, sobre todo cuando se trata de datos con variables múltiples, se recomienda fuertemente el uso de herramientas computacionales apropiadas para reducir el tiempo de realización y aumentar la flexibilidad, cuando se debe realizar varias pruebas en diferentes conjuntos de datos.

1.4. Software para el análisis de datos

Los programas como *SPSS*, *Statgraphics*, *Rstudio* y *WEKA* se emplean con frecuencia en la resolución de diversas tareas estadísticas. Del otro lado, los datos agrícolas se manejan con programas específicos uno de estos programas es el *SMS Advanced*.

1.4.1 SMS Advanced

Diseñado por el grupo Ag Leader, *SMS* es una referencia en el manejo de datos de cosecha. En efecto, su compatibilidad con varias firmas productoras de máquinas le hace imprescindible a la hora de analizar datos de cosecha. Este programa tiene tres versiones: *SMS Mobile*, *SMS Basic* y *SMS Advanced*. Este último posee diversas herramientas tanto para la presentación de la información como para el análisis y el manejo de datos agrícolas, pero su accesibilidad se limita por licencias. *SMS Advanced* dispone de herramientas capaces de generar ecuaciones y analizar los terrenos. También tiene herramientas que comprueban la relación entre el rendimiento y la topografía (Sánchez Ibáñez, 2017).

Este software se destaca por tener en su versión avanzada un conjunto de herramienta para el análisis estadístico de datos. En ella se puede conocer características descriptivas como el promedio de las variables, sus valores mínimos y máximos. Además, se puede aplicar

análisis de correlación para conocer las relaciones entre variables del conjunto (Pérez García 2016).

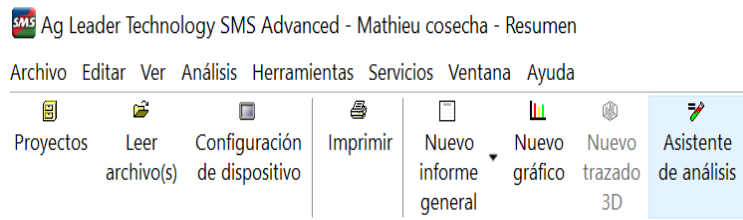


Figura 1.8 Barra de herramientas SMS Advanced

El análisis de correlación se encuentra en la barra de herramienta (Figura 1.8) “asistente de análisis” (Figura 1.9).

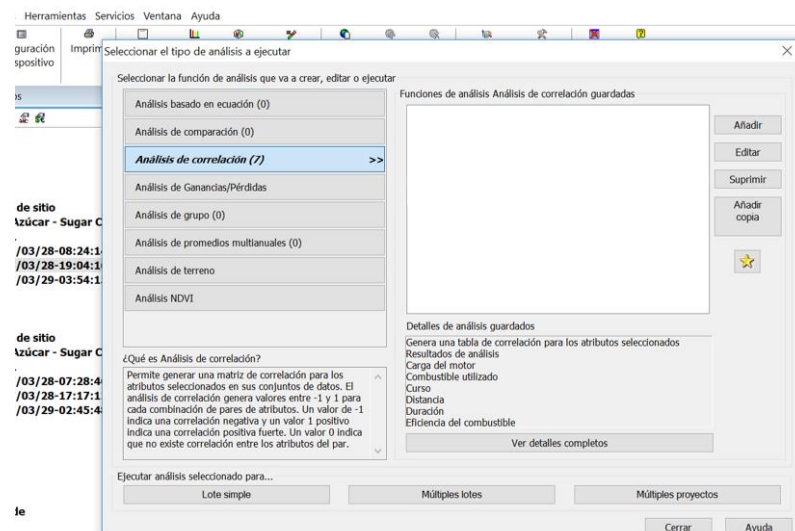


Figura 1.9 Análisis de correlación con SMS Advanced

1.4.2 Paquete de estadística para las ciencias sociales

El paquete de estadística para las Ciencias Sociales (*SPSS* por sus siglas en inglés) es un programa fácil de usar y muy fiable para el análisis de datos (George and Mallery 2016). Desarrollado en 1965 por tres estudiantes de la universidad de Stanford en Estados Unidos y luego comprado por el grupo IBM en 2009, *SPSS* responde a la necesidad de ahorrar tiempo, ya que antes, para resolver tareas estadísticas en computadoras había que escribir códigos de programación muy largos (Nie, Bent et al. 1975). Desde su creación, *SPSS* es usado en muchas entidades como la NASA y las universidades especialmente de humanidades y en el ámbito comercial. Con un solo clic en un ícono se pueden realizar

correlaciones, regresiones, diagramas, ANOVA y la implementación de redes neuronales. *SPSS* es muy demandada por su lenguaje de comando simple y su manual de uso muy integral. Los datos pueden exportarse de diversos programas hacia *SPSS*. *SPSS* define e importa los nombres de las variables, los tipos, los títulos y los valores designados ahorrando tiempo al usuario (Figura 1.12). Gracias a *SPSS* se pueden realizar muchas tareas (George and Mallery 2016):

- Tarea estadística (genera frecuencias, tablas y estadísticas).
- Tarea de modelización (permite crear y validar modelos predictivos gracias a la estadística avanzada).
- Tarea de análisis de texto (ayuda a los administradores a poner información importante de encuestas con participantes).
- Tarea de concepto visual (permite generar gráficos y otras ilustraciones).

Otra ventaja que tiene este software es la capacidad de definir la significatividad de la mayoría de las operaciones. Este índice de significatividad “P” debe ser inferior a 0.05 para confirmar que el análisis realizado no es producto de una casualidad.

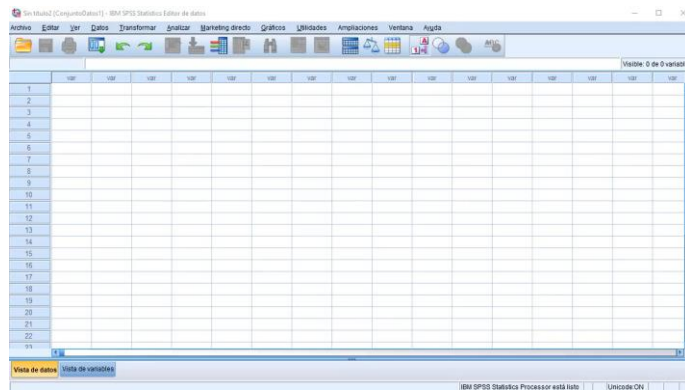


Figura 1.12 Vista de variables

1.4.3 WEKA

WEKA es un programa de inteligencia artificial escrito en Java. Su nombre se deriva de las siglas en inglés de Ambiente *Waikato* para el Análisis de Conocimiento. Este programa tiene como ventaja de ser libre y completamente gratuito. En un inicio, estaba diseñado para el análisis de datos en el campo de la agricultura con otro nombre, pero con el tiempo

se han ido agregando prestaciones y algoritmos hasta redefinirle como herramienta para la inteligencia artificial (Jagtap 2013).

Posee herramientas para la visualización de gráficos, la programación de algoritmos para el análisis de datos y la realización de modelo predictivo e inteligencia artificial. Gracias a su interfaz gráfico de usuario, su uso es bastante fácil ya que permite analizar diferentes técnicas tal como los algoritmos de clasificación, el tratamiento previo de datos, la asociación, el reagrupamiento, el *clustering*, la regresión y la exploración de datos mediante visualización de gráficos (ver Anexos I y II). Además, permite realizar tareas de transformación gracias a sus diferentes filtros implementados (Wimukthi Yatagama Lokuge 2015).



Figura 1.13 Página principal de WEKA

Otras ventajas brindadas por este software son(Rodríguez Suárez and Díaz Amador 2009):

- Su portabilidad dado que esta implementado en el lenguaje de programación Java y es compatible con varias plataformas informáticas modernas.
- Posee una gama de técnicas de modelado y de tratamiento previo de datos.
- Su posibilidad de acceder a bases de datos de tipo SQL gracias a Java *Database Connectivity*.

1.5 Consideraciones finales del capítulo

- De las variables agrícolas, la velocidad, el consumo del combustible y la productividad han sido siempre una preocupación.

- Una de las formas para detectar relaciones entre conjunto es usando métodos del análisis estadístico. Los métodos de la estadística inferencial tal como la correlación permiten cuantificar el grado de asociación entre pares de variables. Por otro la regresión es una medida de estimación de una variable en dependencia de la otra.
- El PCA es un método estadístico de exploración de datos avanzada que permite conocer las estructuras subyacentes. Esta técnica puede ser útil cuando se quiere condensar conjuntos de datos reagrupando variables similares en el punto de vista estadístico.
- *SPSS* y *WEKA* poseen varias herramientas para facilitar el abordaje del análisis estadístico. Además, *SMS Advanced* se anexa al análisis y al manejo de datos en el mundo agrícola.

CAPÍTULO 2. Estudio de relaciones en los datos de cosecha

En este capítulo se presentan las etapas implementadas para desarrollar un estudio de relación entre variables de los datos de cosecha. En un primer momento, se realiza una exploración de datos. En un segundo momento, se proceden a la caracterización de los datos de cosecha y una filtración de datos. En un tercero, se hace un análisis de correlación entre variables que constituye los conjuntos de datos. En un cuarto momento, se realizan un análisis de regresión. Por último, en un quinto, se procede a una síntesis de los conjuntos de datos respectivamente.

2.1. Datos de cosecha

Para esta investigación, se han tomado 8 conjuntos de datos de cosecha, correspondientes a una máquina que operaba en fechas diferentes en la misma granja. El proceso de extracción se hizo respetando la metodología propuesta por (Pérez García 2016).

Después de haber creado un proyecto en el programa en *SMS Advanced*, se importaron todos los conjuntos de datos en el “Árbol de administración” para buscar conjuntos significativos. Los elegidos dependieron de:

- La cantidad de observaciones
- La cantidad de horas de labor
- Las hectáreas cosechadas por horas
- Los mapas de rendimiento
- La cantidad de variables activas

Esto se logra saber explorando diferentes grupos de conjuntos situado en el “Árbol de administración” y eligiendo “Verificación del sitio” para observar las características de los

conjuntos de datos. Las características se observan en “General” y “Valores del conjunto de datos” (Figura 2.1).

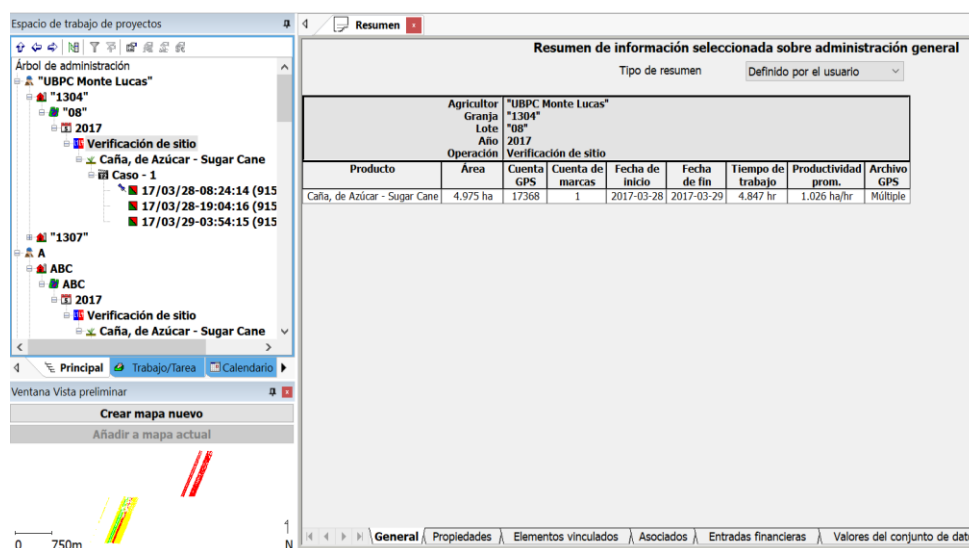


Figura 2.1 Detalles del conjunto

Tres conjuntos de datos fueron retenidos para participar en este estudio. El primer conjunto tiene una duración de 4.7 horas y unas observaciones que llegan hasta 16884. El segundo, contiene 1.143 horas de labor y 4130 observaciones y el tercer conjunto posee 13.66 horas laborales con 48975 observaciones.

En la tabla 2 se pueden apreciar las variables extraídas de la cosechadora con el uso del *SMS Advanced*.

Tabla 2

Variables encontradas en los datos de cosecha usando SMS Advanced

1.Tiempo	2. Elevación	3.Carga motor	4.Velocidad	5. Duración	6.Nivel Fuerza
7. ID objeto	8.Voltaje batería	9. Configuración cóncava	10.Configuración criba 2	11.Distance	12. RPM motor
13. Lote	14.Fecha	15.Desviación Y	16.Longitud	17.Producto	18.Productividad
19.Curso	20.Estado dif.	21. RPM transportador	22.Latitud	23.Configuración criba1	24. Filtro aire ok
25. Modo cortador	26.Cuenta de área	27.Consumo combustible(L/hr)	28.Conjunto de datos	29.Ancho de franja(m)	30.Temperatura aceite hidráulico
31.Fuerza motor	32.Estado de trabajo	33.Consumo combustible(ha/l)	34.Aceite hidráulico bajo	35.Torn. desc. Detall.	36.Temperatura refrigerante (°C)
37.Número de paso	38.Nivel aceite bajo	39.Eficiencia combustible(ha/l)	40.Estado elev. Detall	41.Nivel.bajo refrigerante	42.Combustible utilizado(L)
43.Filtro aceite ok	44.Presión cortador base	45.Velocidad ventilador rpm	46.Codigo Parada E	47.OK Filtro combustible	48.Eficiencia combustible(km/l)
49.Temperatura refrigerante fuera de rango		50. Presión aceite m Fuera de rango	51.Temperatura aceite hidráulico fuera de rango		52.Presión aceite motor (kPa)

El software *SMS Advanced* reveló la existencia de 52 variables. Estos datos fueron exportados hacia *WEKA* para un análisis exploratorio con el fin de conocer mejor la estructura de los datos y luego caracterizarlos. En el *WEKA* se pudo observar tendencias descriptivas de todas las variables mediante la visualización de histogramas, diagramas de dispersión (ver Anexos I y II). Además, se observó características descriptivas como las medias, las desviaciones estándar, los valores mínimos y máximos.

Teniendo en cuenta la exploración de datos, los nombres que llevan las variables y otras características que fueron observadas por *SMS Advanced*, como la preselección de variables importantes para el análisis estadístico, se pudo deducir que estos datos de cosecha se constituyen por variables cualitativas, binarias y cuantitativas continuas y discretas.

Los conjuntos de datos de cosecha se conforman por:

- 3 variables indican fechas y hora: 1,14 y 28.
- La variable (17. Producto) indica el tipo de cultivo (caña de azúcar).
- La variable (9. Lote) corresponde al número del lote.
- 17 variables cualitativas y responden a 2 o 3 estados (si o no y alto, medio o bajo): 6,7,9,10,13,17,20,23,24,25,26,32,34,37,40,41,43,46,47,49,50 y 51.
- 5 son variables cuantitativas discretas:1,7,13,23 y 37 ya que además de no poseer parte decimal, estas variables no tienen unidades.
- 25 son cuantitativas continuas: 2,3,4,5,8,11,12,15,16,18,19,21,22,27,29,30,31,33,36, 39,42,44,45,48 y 52.

La variable (7.ID objeto) es el incremento de cada instante de captura. Los datos se capturan cada un segundo aproximadamente. Con la variable 7, se puede conocer el número de observaciones y se pueden estimar las horas de trabajo de la cosechadora.

El tiempo que se demora para la captura de dos instantes consecutivos se almacena en la variable (5. Duración).

Las 17 variables que responden a los estados como son la señalización para ilustrar al operario los estados de la máquina, y los rangos de variables ya que las variables como presión aceite motor, temperatura refrigerante, temperatura aceite hidráulico, nivel de refrigerante y el nivel de aceite deben estar en un rango determinado para el desempeño de la cosechadora. Por tanto, se puede deducir que las variables citadas poseen una relación discreta con sus respectivas variables continuas.

Según (Voltarelli 2015; Marcasso 2017) la temperatura del refrigerante debe estar entre 85° y 95°, la temperatura del aceite hidráulico no debe sobrepasar los 82° y la presión del aceite se ubica entre 15 y 40 psi. En caso de que no se cumplen estos preceptos, el desempeño de las cosechadoras puede verse afectado.

Como las variables no continuas contribuyen muy poco a la caracterización de un proceso, se consideró profundizar en las 25 variables cuantitativas continuas.

De las variables continuas, 4 variables se consideraron como problemáticas. La (16.longitud y 22.latitud) fueron descartadas ya que las cosechadoras son diseñadas para operar en terrenos planos, por tanto, además de ser externas a la máquina, sus variaciones son ínfimas. Estas hipótesis se corroboran con *SMS Advanced*, ya que, por ser especializado en la agricultura, también considera que las variables (16.longitud y 22.latitud) no son características para el análisis estadístico.

Por otro lado, el análisis exploratorio reveló una variable nula (21.RPM transportador), la otra variable (29. Ancho de franja) era constante para dos conjuntos y presentaba una ligera variación para el primer conjunto. Para tener resultados similares de acuerdo con los tres conjuntos, se decidió no tomarla en cuenta. Al final, 21 variables fueron retenidas para los análisis de correlación, regresión y componentes principales (Tabla 3).

Tabla 3

Variables continuas

2.Elevación	3.Carga motor	4.Velocidad	5. Duración	8.Voltaje batería
11.Distance	12.RPMmotor	15.DesviaciónY	18.Productividad	19.Curso
21. RPM transportador	27.Consumo combustible(L/hr)	30.Temperatura aceite hidráulico	31.Fuerza motor	
33.Consumo combustible (L/ha)	36.Temperatura refrigerante	39.Eficiencia combustible ha/l	42.Combustible utilizado(L)	44.Presión cortador base
48.Eficiencia combustible(km/l)	45.Velocidad ventilador rpm	52. Presión aceite (kPa)		

2.2 Tratamiento de datos

Previamente, se hizo un tratamiento de datos para acondicionar las variables. Este tratamiento se hizo teniendo en cuenta las consideraciones del análisis exploratorio de datos realizado por el programa *WEKA*. Esta exploración evidenció las diferencias de escalas entre variables y la necesidad de homogeneizar las varianzas realizando unas transformaciones.

En este sentido, las características de cada una de las variables han sido verificadas para asegurar que cumplían con los requisitos del PCA y de la regresión lineal. De acuerdo con la teoría, esta transformación o reducción se tenía que hacer dado que las variables presentes en el conjunto de datos no poseían las mismas unidades. La transformación usada fue una función logarítmica de base 10. En este caso, la reducción se hizo sobre las variables que presentaban una escala grande respecto a las otras para poder tener un rango homogéneo de varianza entre todas las variables. Esta reducción se hizo con el medio de la calculadora incluido en *SPSS* (Figura 2.2 y 2.3).

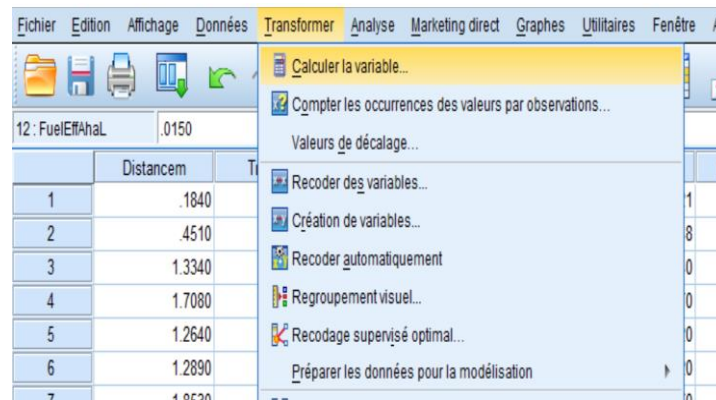


Figura 2.2 Transformación de variables

Dado las características de la función logarítmica:

- Indefinida para valor inferior o igual a 0.
- Negativa en el intervalo de 0 a 1.

Varias estrategias se implementaron para paliar este problema (Rodríguez Ayán and Ruiz Díaz 2008):

- Las observaciones de las variables a normalizar deben ser mayor que cero.
- Si existen observaciones menores que uno, adicionarles un valor a todas las observaciones de la variable de tal forma que el valor mínimo de la variable sea al menos uno.
- En el caso que el histograma tienda a la derecha alrededor de la media, usar " $LG10(Max + 1 - Variable)$ " donde *Max* es el máximo de la variable.

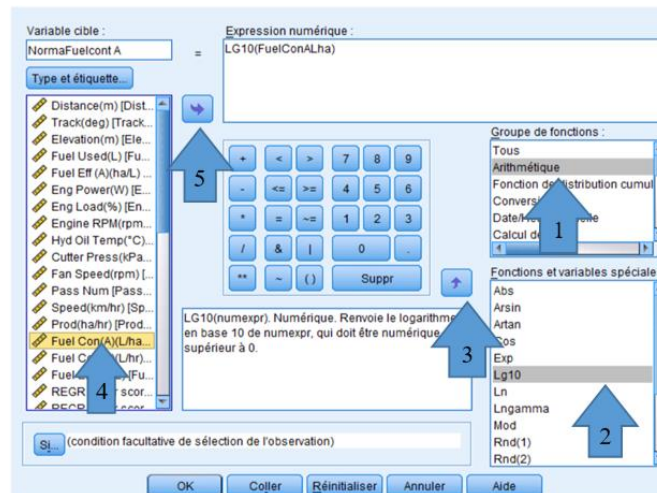


Figura 2.3 Pasos para la función logarítmica

También se hizo un análisis de correlación bivariado entre pares de variables transformadas y no transformadas para chequear la calidad de las variables transformadas respecto a las variables no transformadas, ya que el objetivo era sobre todo reducir la escala para que todas las variables puedan ser comparadas con el mismo peso. Esta prueba demostró una ligera diferencia entre las variables transformadas y no transformadas en cuanto a su correlación (ver Anexo III).

Muchas variables han sido reducidas en dependencia del conjunto analizado y de las características descriptivas. Por lo general, se redujeron todas las variables que presentaban una media superior a 10 de tal modo que las tendencias centrales de los datos sean menores o iguales a la unidad (ver Anexo IV y V).

En los Anexos IV y V se puede observar las características descriptivas de los 3 conjuntos analizados tales como las medias y las desviaciones estándar. Después de haber transformado las variables, se puede constatar que los valores están bastante homogeneizados respecto a las características descriptivas sin transformar. Todas las medias son por lo general inferiores a 6. De igual forma las desviaciones estándar son en su mayoría por debajo de 1.

2.3 Análisis de correlación

El análisis de correlación se hizo con tres programas. En un primer momento se realizó con el software *SMS Advanced*. Luego, se hizo con el software *WEKA*. Por último, se realizó una correlación de Spearman para pares no lineales.

En el *SMS Advanced*, se hizo tres procedimientos de correlación. Los procedimientos realizados, corresponden a cada uno de los conjuntos elegidos. Esto se logra filtrando los conjuntos (Figura 2.4). Esta filtración es simplemente una selección del conjunto de datos que se desea analizar y se debe repetir al final del procedimiento para confirmar el conjunto elegido.

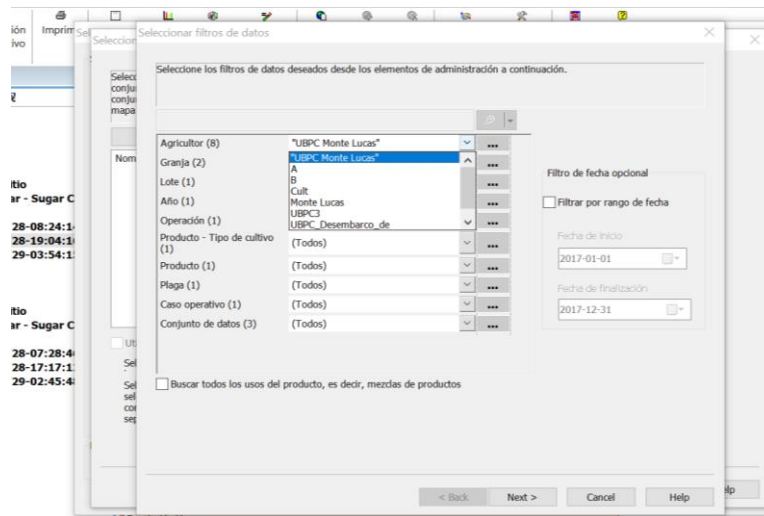


Figura 2.4 Selección del conjunto de datos

Además, se hizo una correlación punto a punto entre cada par de variables, ya que todas las variables tenían las mismas cantidades de observaciones. *SMS Advanced* no toma en cuenta las variables (16.longitud y 22.latitud) para los análisis de correlación, por tanto preseleccionó 23 variables continuas de forma automática para el análisis de correlación. Enseguida se procedió a la selección de 21 variables como se había mencionado anteriormente que las variables (21.RPM transportador y 29. Ancho de franja) eran nulas y constantes. En esta misma etapa, se aprovecha la posibilidad que dispone el programa para normalizar fácilmente variables (Figura 2.5). Estas etapas se hicieron para todas las variables analizables en los distintos conjuntos. Finalmente, se generaron matrices de correlación para los tres conjuntos de forma separada.

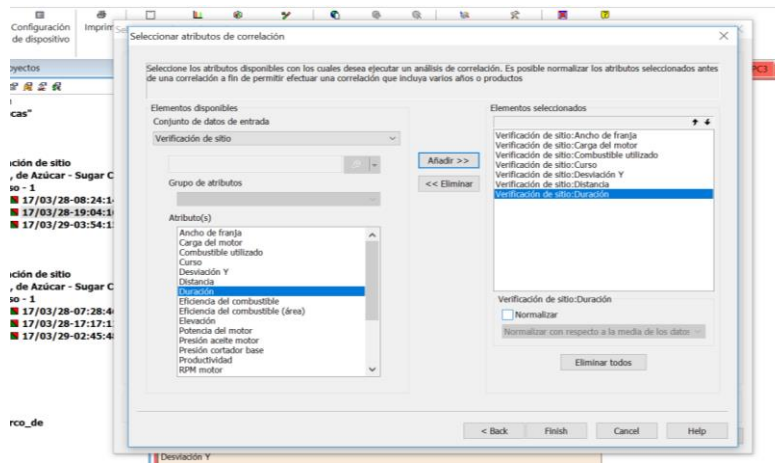


Figura 2.5 Preparación de datos

El mismo análisis de correlación ha sido hecho por cada variable en WEKA. Este análisis de correlación se encuentra en las herramientas “*Select attributes*” (Figura 2.6).

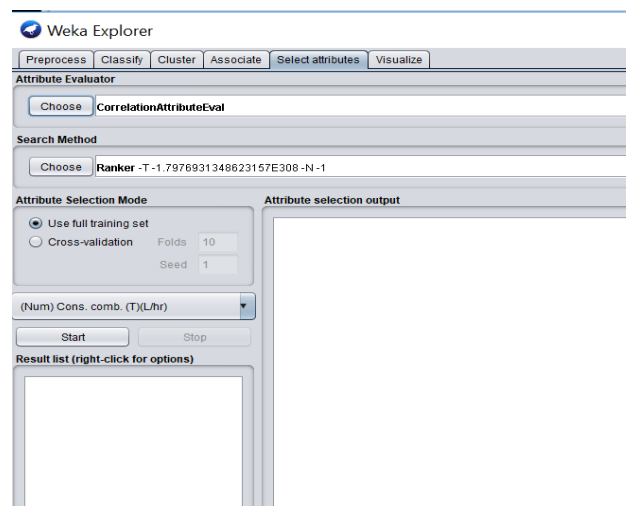


Figura 2.6 Análisis de correlación

La exploración de datos reveló también relaciones con características no lineales (Anexo I). Por tanto, se hizo un análisis de correlación de Spearman para estos pares de variables con la herramienta *SPSS*.

Dado la multitud de pares que pueden existir debido a la combinación de 21 variables y los coeficientes de correlación producidos por las distintas matrices de correlación generadas por los programas, se clasificaron las relaciones en un orden mostrando solo coeficientes de correlación que presentaban buena repetibilidad en los tres conjuntos de datos de cada programa como para descartar relaciones dudosas.

2.4 Análisis de regresión

Se hizo un análisis de regresión lineal simple sobre los pares de variables que presentaban coeficientes de correlación con una buena repetibilidad. De estos pares de variables, se analizaron únicamente relaciones que presentaban una buena linealidad y homocedasticidad.

De igual forma que se realizó el análisis de correlación, este proceso se hizo para los tres conjuntos de datos. Este análisis se hizo en dos programas: *WEKA* y *SPSS*.

En *WEKA*, la regresión lineal simple se hizo eligiendo este método en el grupo de herramientas de clasificación brindadas por el programa. El software generó unas ecuaciones de dependencia para cada par (Figura 2.7).

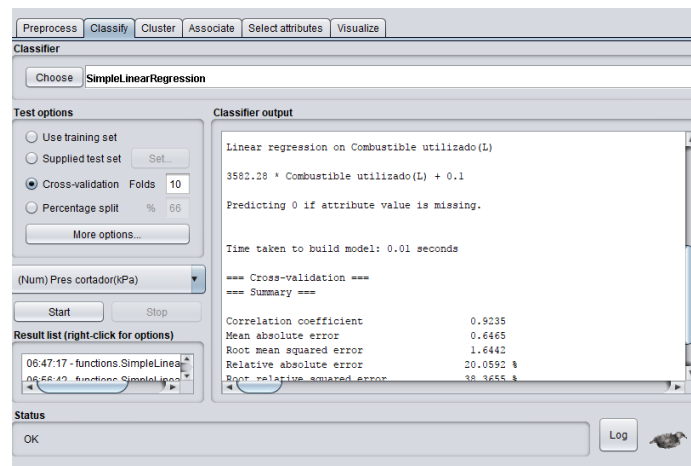


Figura 2.7 Análisis de regresión lineal simple

En *SPSS*, esta técnica se realiza mediante la herramienta de análisis (Figura 2.8). Además de generar ecuaciones de dependencia, esta misma regresión lineal simple ha permitido la obtención de coeficientes de regresión y también un índice de significación para indicar la calidad de las regresiones.

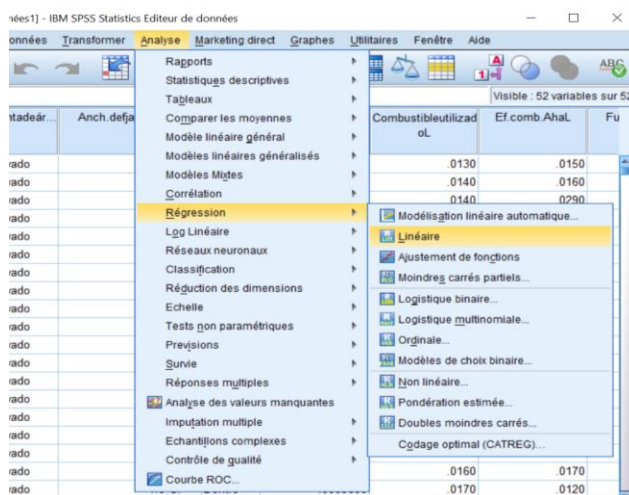


Figura 2.8 Análisis de regresión

2.5 Análisis de componentes principales

Esta técnica de reducción de dimensionalidad puede llevarse a cabo con la herramienta estadística *SPSS* mediante la función del análisis factorial (Figura 2.9).

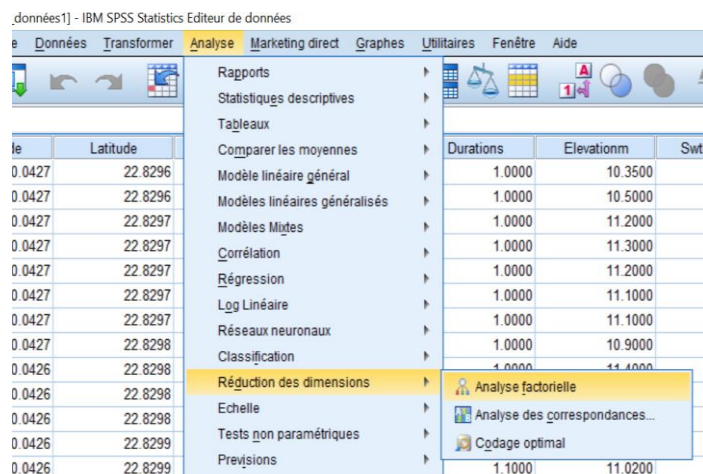


Figura 2.9 Análisis factorial

El análisis de componentes principales ha sido usado en los tres conjuntos de datos. Los conjuntos de datos fueron sometidos a una serie de pruebas tales como el Índice de KMO y la prueba de esfericidad de Bartlett en *SPSS*. Estas pruebas son una medida de la calidad de las relaciones y de la matriz de correlación. Se pudo generar también características descriptivas como las medias y las desviaciones estándar (ver Anexo V), para asegurarse que dichos conjuntos transformados seguían cumpliendo con los criterios del PCA (Figura 2.10).

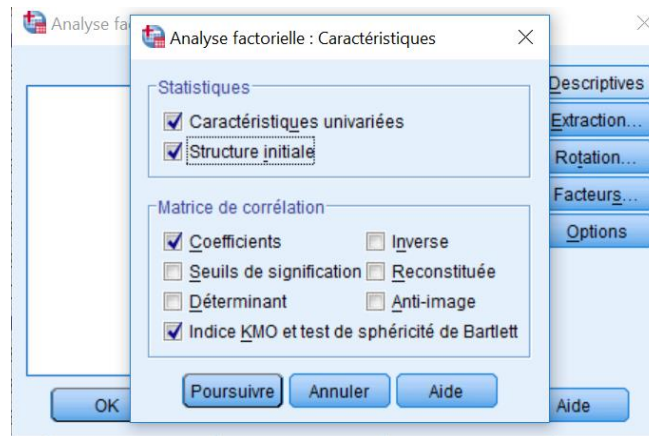


Figura 2.10 Definición de características del análisis

Todos los conjuntos pasaron estas pruebas con éxito. Los coeficientes de Kaiser-Meyer-Olkin para cada conjunto respectivamente tenían un valor de: 0.641, 0.809 y 0.683. De manera similar, la prueba de esfericidad de Bartlett era significativa para todos los conjuntos con una “P” igual a 0.00. Por tanto, el análisis de componentes principales era realizable.

Para la extracción de los componentes principales, se generaron matrices de correlación ya que se quiere conocer también posibles relaciones entre variables. Se generó un diagrama de valores propios y se eligió los componentes principales con el criterio de Kaiser excluyendo valores propios inferiores a 1 (Figura 2.11).

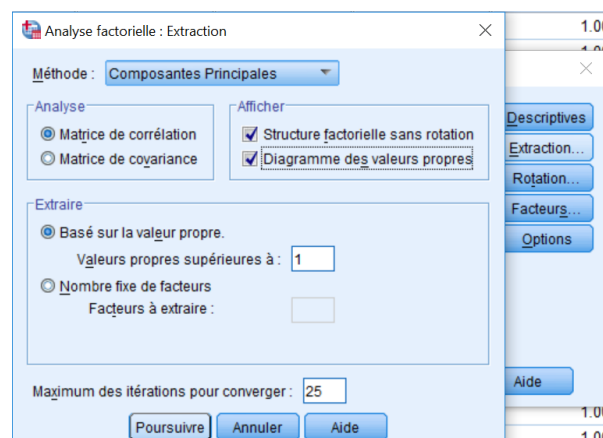


Figura 2.11 Extracción de componentes principales

El análisis de componentes principales realizado con SPSS generó una gran cantidad de resultados. Se puede observar estadística descriptiva, matriz de correlación, calidad de la

información representada en los componentes principales, diagramas valores propios, varianza representada y correlación de cada componente respecto a cada variable inicial.

2.6 Consideraciones del capítulo 2

- Los datos de cosecha están constituidos por 52 variables. Estos conjuntos de datos poseen una gran cantidad de variables que pueden ser cuantitativas continuas y discretas, cualitativas y dicotómicas. Las variables más características para los análisis estadísticos son continuas. Estos datos se capturan cada segundo.
- Para llevar a cabo este estudio, se utilizaron tres programas. Estos programas agilizaron el proceso del análisis de datos proporcionando una interfaz amigable que simplificó considerablemente el trabajo.
- El programa de agricultura de precisión *SMS Advanced* ha sido de mucha ayuda puesto que sostuvo nuestras hipótesis preseleccionando de forma automática 23 variables continuas para el análisis de correlación. En este programa se generó matrices de correlación fácilmente.
- El *WEKA* permitió una exploración visual de datos, una generación de matrices de correlación y un análisis de regresión.
- En el *SPSS* se hizo una reducción del conjunto de datos para resaltar nuevas variables características que servirán de base para un futuro modelado de este proceso. En este mismo programa se realizaron un análisis de regresión lineal sobre las variables.

CAPÍTULO 3. ANALISIS DE RESULTADOS

En este capítulo, se analizan los resultados obtenidos de la metodología implementada para los análisis de correlación, regresión lineal y componentes principales. Luego se dan a conocer los posibles impactos económicos y medioambientales.

3.1 Resultados de la correlación

En el Anexo VI, se tiene el análisis de correlación realizado con los dos programas (*SMS Advanced*, *WEKA* y *SPSS* respectivamente) y en el anexo VII se muestran los resultados de las matrices de correlación generadas por el análisis de componentes principales en *SPSS*. Las columnas representan diferentes conjuntos analizados mientras que las filas representan las combinaciones de pares de variables analizadas. En realidad, se puede extender la tabla ya que hay muchas combinaciones posibles, pero se seleccionaron solo pares que tenían buena repetibilidad del coeficiente de correlación de acuerdo con todos los conjuntos analizados. Todos los pares analizados tenían coeficientes significativos con valores de “P” iguales a 0.00.

En todos los conjuntos y mediante todos programas, los pares analizados han presentado una cierta similitud. Los pares como: Eficiencia del combustible (Km/L)-Eficiencia del combustible (ha/L), Velocidad-Productividad, Velocidad-Distancia, Productividad-Distancia, Combustible Utilizado-Consumo combustible(L/hr), Fuerza motor-Combustible Utilizado, Distancia-Eficiencia de combustible (Km/L), Productividad-Eficiencia combustible (Km/L), Distancia-Eficiencia de combustible (ha/L), Productividad-Eficiencia combustible (ha/L), Velocidad-Eficiencia combustible (Km/L), Velocidad-Eficiencia combustible (ha/L), presentaron una relación fuerte con coeficientes de correlación que muchas veces sobrepasan los 0.9 (Figura 3.1).

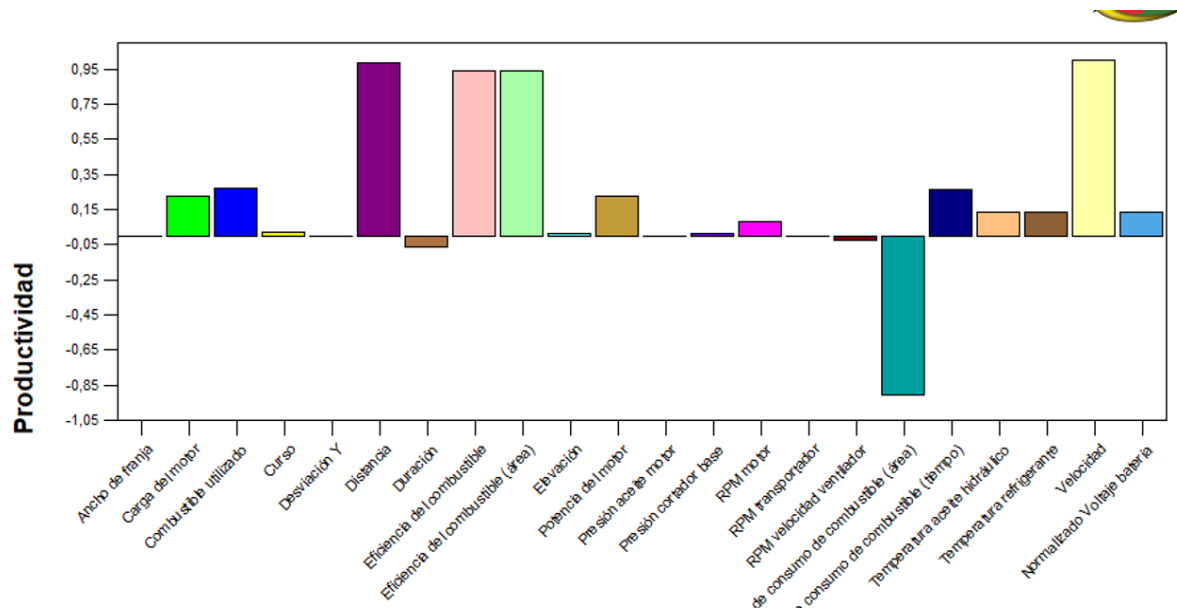


Figura 3.1 Correlación productividad contra todas las variables

De igual manera, los pares como: Carga motor-Fuerza motor, Carga motor-Combustible utilizado, Fuerza motor-Consumo combustible(L/hr) y Carga motor-Consumo combustible(L/hr) presentaron unas correlaciones fuertes con coeficientes de correlación entre 0.76 y 0.89 (Figura 3.2).

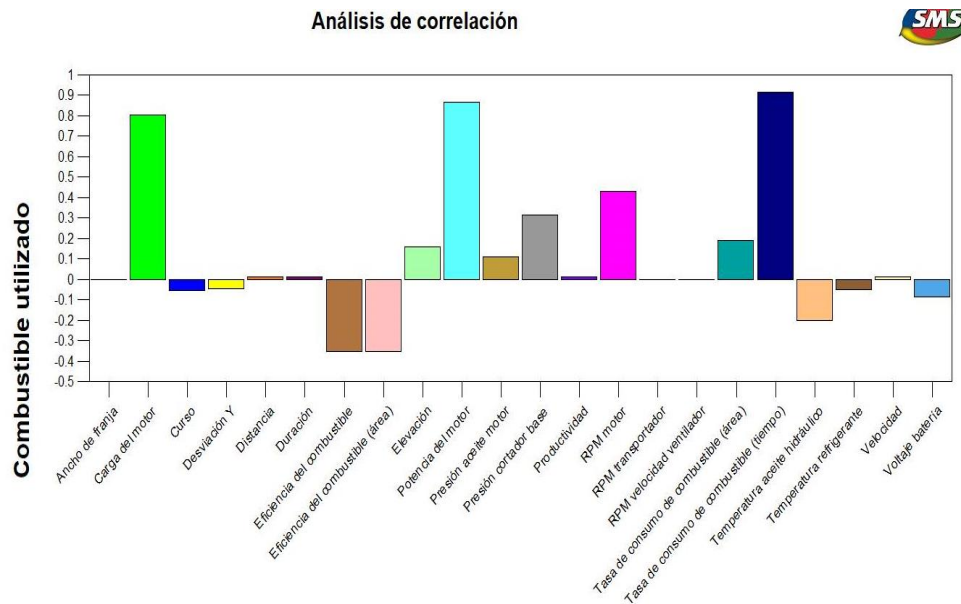


Figura 3.2 Correlación combustible utilizado contra todas las variables

Los pares como: Consumo combustible (L/hr)-Duración, Fuerza motor-Presión cortador base y Temperatura Refrigerante-Temperatura aceite hidraulico presentaron unas correlaciones débiles con coeficientes de correlaciones entre 0.25 y 4.40.

El par RPM motor-Presión aceite motor tiene un coeficiente de correlación moderada entre 0.48 y 0.59 aunque en algún caso presentó una correlación muy fuerte.

Los dos pares: Consumo combustible (L/hr)-Presión cortador base y Combustible utilizado- Presión cortador base indicaron unas correlaciones casi despreciables con valores de sus coeficientes de correlación ubicados entre 0.26 y 0.30.

Por fin, los pares como: Eficiencia combustible (ha/L)-Consumo combustible (ha/h), Eficiencia combustible (Km/L)-Consumo combustible (L/ha), Productividad- Consumo combustible (ha/h), Velocidad-Consumo combustible (L/ha) y Distancia-Consumo combustible (L/ha) presentaron una relación no lineal e inversamente proporcional. Los análisis de correlación de Spearman indicaron relaciones muy fuertes con coeficientes que sobrepasaban los 0.90 en la mayoría de los casos (Tabla 4 y figura 3.3).

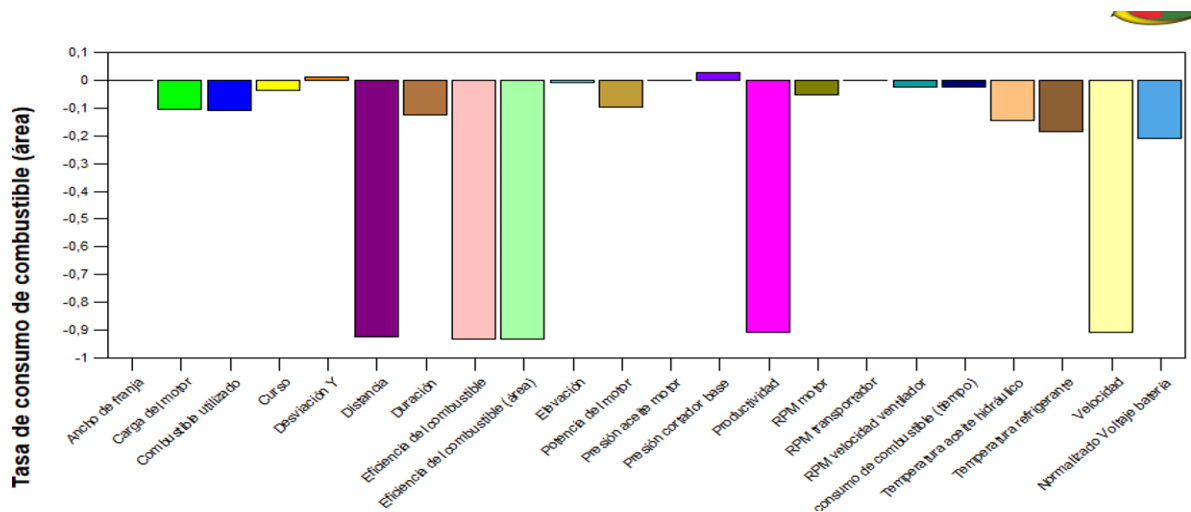


Figura 3.3 Correlación consumo en unidad de área contra todas las variables

Tabla 4

Análisis de correlación no lineal de Spearman con SPSS

Conjuntos Pares analizados	Conjunto 1	Conjunto 2	Conjunto 3
$y = \text{Velocidad}$ $x = \text{Consumo combustible(L/ha)}$	-0.853 (P=0.00)	-0.934 (P=0.00)	-0.937 (P=0.00)
$y = \text{Consumo combustible(L/ha)}$ $x = \text{Productividad}$	-0.920 (P=0.00)	-0.934 (P=0.00)	-0.937 (P=0.00)
$y = \text{Consumo combustible(L/ha)}$ $x = \text{Eficiencia combustible(Km/L)}$	- 0.904 (P=0.00)	-1.000 (P=0.00)	-1.000 (P=0.00)
$y = \text{Distancia}$ $x = \text{Consumo combustible(L/ha)}$	-0.912 (P=0.00)	-0.949 (P=0.00)	-0.946 (P=0.00)
$y = \text{Consumo combustible(L/ha)}$ $x = \text{Eficiencia combustible(ha/L)}$	-0.998 (P=0.00)	-0.996 (P=0.00)	-0.997 (P=0.00)

3.2 Análisis de regresión

El análisis de regresión mediante los programas *WEKA* y *SPSS* dieron resultados casi similares con respecto a las ecuaciones de predicción. Las únicas diferencias son los hechos de que *WEKA* redondea los coeficientes a y b de la ecuación de la recta (Anexo XV y XVI) y no devuelve el coeficiente de regresión, mientras que el *SPSS* presenta una mayor cantidad de información como el coeficiente de regresión y el índice de significatividad estadística “p”. Todos los pares analizados eran significativos con los “P” nulos indicando que los resultados son significativos.

El coeficiente de regresión del par Eficiencia del combustible (Km/L)-Eficiencia del combustible (ha/L) se ubica entre 0.846 y 0.996. Esto indica que la eficiencia del combustible en km/L puede explicar entre 84% y 99.6% de la varianza de la eficiencia en unidad de área.

La velocidad puede explicar entre 85% y 100% de la varianza de la productividad. La distancia puede explicar entre 83% y 98% de la varianza de la productividad. La velocidad de desplazamiento puede explicar 98% de la varianza la distancia y esta distancia puede explicar entre 84% y 91% de la varianza de la eficiencia del combustible en unidad métrica y entre 76% y 90% de la eficiencia en unidad de área. La velocidad de desplazamiento puede explicar entre 82% y 90% de la varianza de la eficiencia en unidad métrica y esta eficiencia puede explicar entre 66% y 90% de la varianza de la productividad. Por fin, la velocidad puede explicar entre 75% y 90% de la varianza de la eficiencia en unidad de área y esta eficiencia puede explicar entre 85% y 90% de la varianza de la productividad.

Del otro lado, la carga del motor puede explicar entre 96% y 99% de la varianza de la fuerza o la potencia generada por el motor. El consumo horario del combustible puede explicar entre 81% y 89% de la varianza del combustible utilizado. La carga del motor puede explicar entre 72% y 77% de la varianza del combustible utilizado. La fuerza o la potencia del motor puede explicar entre 62% y 73% de la varianza del consumo horario del combustible. Esta fuerza también puede explicar entre 76% y 83% de la varianza del combustible. La carga puede explicarse entre 63% y 68% de la varianza de consumo horario del combustible.

Aunque el consumo en unidad de área posee una relación no lineal con sus pares respectivos (Tabla 4), sus ecuaciones y sus coeficientes de regresión eran considerable.

Los resultados indicaron que, respecto al consumo en unidad de área, la velocidad y la productividad pueden explicarse entre 47% y 50% de su varianza. La distancia puede explicar entre 48% y 60% de su varianza. La eficiencia en unidad de área y en unidad métrica puede explicar entre 48% y 62% de su varianza.

Las variables relacionadas con presión del cortador de base (combustible utilizado, carga, fuerza y consumo) presentaron una mala regresión con varianzas explicadas por lo general menor que 10%. Aunque sus coeficientes de correlación alcanzaban 0.3 o 0.4.

Los pares de variables tales como carga-Fuerza, carga-consumo combustible(L/hr) y consumo combustible(L/hr)-Fuerza presentaron coeficientes de correlación y regresión altos (Anexos VI, VII y XVI). Este hallazgo apoya los resultados obtenidos por (González Valdés, Rodríguez López et al. 2010) durante su análisis de relación entre carga del motor, consumo de combustible y Fuerza del motor en máquinas agrícolas y ligeras usando la herramienta computacional *Statgraphics*.

Los pares de variables como Eficiencia combustible (A)(ha/L)-Eficiencia combustible (km/L) presentaron una correlación muy fuerte y un coeficiente de regresión del mismo orden en la mayoría de los casos. Esto indica una posibilidad de que la información proviene de fuentes muy cercanas. Además, puede ser que una de estas variables ha sido derivada mediante una ecuación lineal. Este supuesto también se extiende para otros pares como Velocidad-Productividad, Velocidad-Distancia, Productividad-Distancia, Carga motor-Fuerza motor, Combustible Utilizado-Consumo combustible(L/hr), Fuerza motor-Combustible Utilizado, Distancia-Eficiencia de combustible (Km/L), Productividad-Eficiencia combustible (Km/L), Distancia-Eficiencia de combustible (ha/L), Productividad-Eficiencia combustible (ha/L), Velocidad-Eficiencia combustible (Km/L), Velocidad-Eficiencia combustible (ha/L). Por tanto, esto ilustra que muchas variables continuas de los datos de cosecha derivan de ecuaciones lineales.

La ecuación del autor (Pereira Filho and Rodrigues 2015) para el cálculo de la productividad tiene un sentido ya que la pequeña variación en la variable (29. ancho de franja) producida en el conjunto 1 produjo una ligera variación del coeficiente de correlación. Esto indica que la productividad no depende solo de la velocidad.

La alta correlación entre la velocidad y la productividad sostiene los resultados de los trabajos de (Testa 2014) y de (Belardo 2010) sobre la relación de proporcionalidad que tenían ambas variables, siendo cuanto mayor es la velocidad, mayor será la productividad en el campo.

3.3 Análisis de componentes principales

Las variables candidatas al análisis de componentes principales fueron bien representadas en los nuevos conjuntos respectivamente. De forma general, no se perdió mucha

información. Las variables que fueron menos representativas son: Elevación, temperatura refrigerante, temperatura aceite hidráulico, presión cortadora de base y desviación Y. Estos problemas ocurrieron en mayor medida en los conjuntos 1 y 3 (ver Anexo X).

Los tres conjuntos de datos iniciales fueron reducidos en componentes principales. El primer conjunto tuvo 7 componentes principales con un volumen de información retenida de 84.367 % (ver Anexo XI). En el segundo se obtuvo 6 componentes principales con 89.773 % de la información inicial (ver Anexo XII). Por fin, el tercer conjunto tuvo 7 componentes principales 79.526 (ver Anexo XIII). La extracción de los componentes se hizo siguiendo el criterio Kaiser. Este criterio retiene solo los componentes que tiene valores propios mayores o iguales a 1.

El análisis de componentes principales reveló que cada componente tenía una cierta relación con las variables iniciales.

Para el primer conjunto (ver Anexo XIV):

1. El primer componente tiene una relación fuerte con: Eficiencia del combustible (km/L), eficiencia combustible(ha/L), velocidad, productividad, consumo combustible (L/ha) y distancia.
2. El segundo componente está fuertemente correlacionado con las variables: Carga motor, fuerza motora, consumo combustible (L/hr) y combustible utilizado.
3. El tercer componente está más correlacionado con: Desviación Y, velocidad ventilador y voltaje batería.
4. El cuarto componente está más correlacionado con: RPM motor y presión aceite motor.
5. El quinto componente está más correlacionado con: Temperatura aceite hidráulico, temperatura refrigerante y elevación.
6. El sexto componente está más correlacionado con: Duración, temperatura refrigerante y curso.
7. El último componente está más correlacionado con duración.

Para el conjunto 2 (ver Anexo XV):

1. El primer componente tiene una relación fuerte con: Eficiencia del combustible (km/L), eficiencia combustible(ha/L), velocidad, productividad, consumo combustible

- (L/ha) y distancia, temperatura aceite hidráulico, temperatura refrigerante, presión cortador de base, velocidad ventilador, RPM motor y presión aceite motor.
2. El segundo componente está fuertemente correlacionado con las variables: Eficiencia del combustible (km/L), eficiencia combustible(ha/L), velocidad, productividad, consumo combustible (L/ha) y distancia, temperatura refrigerante, presión cortador de base, velocidad ventilador, voltaje batería, RPM motor y presión aceite motor.
 3. El tercer componente está más correlacionado con: Carga motor, fuerza motora, consumo combustible (L/hr) y combustible utilizado.
 4. El cuarto componente está más correlacionado con: Duración.
 5. El quinto componente está más correlacionado con: Duración, temperatura refrigerante y curso.
 6. El sexto componente está más correlacionado con: Desviación Y.

Para el tercer conjunto (ver Anexo XVI):

1. El primer componente tiene una relación fuerte con: Eficiencia del combustible (km/L), eficiencia combustible(ha/L), velocidad, productividad, consumo combustible (L/ha) y distancia.
2. El segundo componente está fuertemente correlacionado con las variables: Carga motor, fuerza motora, consumo combustible (L/hr) y combustible utilizado.
3. El tercer componente está más correlacionado con: RPM motor y presión aceite motor.
4. El cuarto componente está fuertemente correlacionado con: Temperatura aceite hidráulico y temperatura refrigerante.
5. El quinto componente está más correlacionado con duración.
6. El sexto componente está más correlacionado con: Voltaje batería y curso.
7. El último componente está más correlacionado con: Voltaje batería, elevación y desviación Y.

El análisis de componentes principales y el análisis de correlación han permitido detectar variables que se reagrupan o se conglomeran.

Primer grupo:

- Eficiencia combustible (Km/L)
- Eficiencia combustible (A) (ha/L)

- Velocidad
- Productividad
- Distancia
- Consumo combustible (L/ha)

Segundo grupo:

- Fuerza
- Carga
- Consumo horario del combustible
- Combustible utilizado

Se puede observar que las variables que conforman cada grupo están fuertemente correlacionadas entre sí y también siempre se reagrupan independientemente del conjunto de datos analizado en el análisis de componentes principales (Anexos XIV, XV y XVI). Es cierto que el análisis de componentes principales es un estudio de varianza, pero este estudio demostró que se puede clasificar o identificar variables similares si se hace varias pruebas con conjuntos diferentes y a condición de que los individuos que se desean clasificar tengan una correlación considerable entre ellos.

3.4 Análisis económico y medioambiental

El principal impacto económico del presente trabajo está relacionado con el ahorro de tiempo que puede significar la aplicación de esta investigación. Con la determinación de las relaciones entre los datos de cosecha es posible conocer, por ejemplo, qué variables se necesitan para obtener un modelo de funcionamiento de las cosechadoras CASE-IH A8000. Mientras menos variables se precisen, más sencillo será usarlo y menor costo computacional tendrá, por lo que los requerimientos para su uso no requerirán una gran inversión. Por otro lado, puede conocerse qué variables influyen sobre el combustible consumido y actuar sobre éstas para disminuir su uso. Evidentemente tal cuestión también tiene un efecto medioambiental ya que con la disminución del consumo también se disminuye la emisión de gases de efecto invernadero a la atmosfera.

La demostración del empleo de *WEKA* contribuye a la estrategia de migración hacia el software libre que se lleva a cabo en Cuba.

3.5 Consideraciones del capítulo 3

- 22 pares de variables presentan relaciones considerables con coeficientes que sobrepasa los 0.5. De estos 22 pares, todos presentaron coeficiente de regresión bastante considerable a excepto del par RPM motor-Presión aceite motor.
- Los conjuntos por lo general se pueden reducir en 6 componentes principales como mínimo.
- Los dos grupos de variables clasificadas presentan individuos que son estadísticamente casi similares brindando una misma información.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

1. Generalmente, los estudios asociados a las máquinas agrícolas hoy en día se enfocan en el conocimiento de las variables operacionales para elevar la productividad y para la reducción del consumo de combustible.
2. Varios métodos estadísticos como el análisis de correlación, análisis de componentes principales y el análisis de regresión permitieron evidenciar las relaciones, detectar similitudes entre variables y conocer las influencias de un gran número variable para el desempeño operacional de la cosechadora cañera.
3. Los datos de cosecha están compuestos de 52 variables. Ellos están constituidos por variables cualitativas, binarias, cuantitativas continuas y discretas. Entre ellas, las que se pueden utilizar para caracterizar el proceso son mayoritariamente cuantitativas continuas.
4. Las herramientas computacionales agilizaron el proceso de los análisis ahorrando tiempo y devolviendo informaciones detalladas de tal forma que se ha logrado interpretar un volumen de información con más facilidad.
5. 28 pares de variables tienen relación. De esas 28, 22 poseen una relación considerable con el coeficiente de correlación mayor de 0.5. Con estas 22, se pueden estimar 21 pares con ecuaciones de regresión.
6. Los conjuntos de datos pueden sintetizarse en 6 componentes principales como mínimo. De estos conjuntos, existen dos grupos de variables cuyos individuos provienen de fuentes muy cercanas en el punto de vista estadística.

Recomendaciones

1. Utilizar este procedimiento para el tratamiento y la síntesis de datos de cosecha para el diseño de un modelo predictivo.
2. Analizar relaciones de forma específica para tener conclusiones más precisas.
3. Generalizar este estudio aplicando estos análisis en otros datos producidas en otras máquinas agrícolas.
4. Proceder a una mejor caracterización de datos de cosecha.

REFERENCIAS BIBLIOGRÁFICAS

Aguilar, M. P. P. (2019). "Indicadores de desempeño académico como predictores de captación de recursos financieros." Convergencia: Revista de ciencias sociales(79): 8.

Belardo, G. (2010). Avaliação de desempenho efetivo de três colhedoras em cana-de-açúcar (*Saccharum spp*) sem queima. 2010. 136 p, Dissertação (Mestrado em Ciências)-Escola Superior de Agricultura "Luiz de

Bencherki, G. and B. Moustafa (2018). "Implémentation d'un système de reconnaissance de visages à base de PCA."

Brahimi, N. (2011). "Construction et exploitation d'une base de données sur les semiconducteurs Magnétiques Dilués."

Cerny, B. A. and H. F. Kaiser (1977). "A study of a measure of sampling adequacy for factor-analytic correlation matrices." Multivariate behavioral research **12**(1): 43-47.

de la Maza, L. E. L. (2017). "Tesis de diploma para optar por el título de ingeniero químico."

Delgado-Trejos, E., J. H. R. Piedrahita, et al. (2006). "Reducción de dimensiones para clasificación de datos multidimensionales usando medidas de información." Scientia et Technica **3**(32): 181-186.

Díaz Blanco, I. (2000). "Detección e identificación de fallos en procesos industriales mediante técnicas de procesamiento digital de señal y redes neuronales. Aplicación y mantenimiento predictivo de accionamientos eléctricos."

George, D. and P. Mallery (2016). IBM SPSS statistics 23 step by step: A simple guide and reference, Routledge.

González Valdés, R. P., Y. Rodríguez López, et al. (2010). "Consumo de combustible de los motores de combustión interna." Revista Ciencias Técnicas Agropecuarias **19**(1): 01-08.

Jagtap, S. B. (2013). "Census data mining and data analysis using WEKA." arXiv preprint arXiv:1310.4647.

Jallas, E. and M. Crétenet (2003). Aide à la décision en agriculture et gestion des ressources naturelles: note introductive. Savanes africaines: des espaces en mutation, des acteurs face à de nouveaux défis. Actes du colloque, Garoua, Cameroun, Cirad-Prasac.

Jaouad, M., A. Mahjoubi, et al. (2015). "Identification des structures spatiales des pratiques agricoles dans les régions arides tunisiennes: Cartographie statistique de données d'enquêtes socioéconomiques."

Le Moal, L. (2002). "L'analyse en composantes principales." LACP sous SPSS.

Leech, N., K. Barrett, et al. (2013). SPSS for intermediate statistics: Use and interpretation, Routledge.

Lyra, G. A. d. (2012). "Consumo de combustível de duas colhedoras de cana-de-açúcar em função da velocidade e rotação de motor."

Marcasso, G. (2017). "Estudo de caso da falha de um motor diesel 9.0 l John Deere de uma colhedora de cana de açúcar John Deere 3522 2l."

McGuire, S. (2015). FAO, IFAD, and WFP. The state of food insecurity in the world 2015: meeting the 2015 international hunger targets: taking stock of uneven progress. Rome: FAO, 2015, Oxford University Press.

Molina Hernández, J. (2015). Cosechadoras cañeras CASE IH A8000 para la Agricultura de Precisión en Cuba, Universidad Central" Marta Abreu" de Las Villas.

Mukaka, M. M. (2012). "A guide to appropriate use of correlation coefficient in medical research." Malawi Medical Journal **24**(3): 69-71.

Nie, N. H., D. H. Bent, et al. (1975). "SPSS: Statistical package for the social sciences."

Pereira Filho, I. and J. Rodrigues (2015). "Sorgo: o produtor pergunta, a Embrapa responde." Embrapa Milho e Sorgo-Livro técnico (INFOTECA-E).

Pérez García, C. A. (2016). Metodología para la extracción de datos de cosecha en cosechadoras cañeras Case IH A8000, Universidad Central" Marta Abreu" de Las Villas. Facultad de Ingeniería

Rocheffort, C. (2009). "Structure et fonctionnement des services québécois de néonatalogie et leurs relations avec la qualité des soins et des indicateurs de santé au travail chez les infirmières."

Rodríguez Ayán, M. N. and M. Á. Ruiz Díaz (2008). "Atenuación de la asimetría y de la curtosis de las puntuaciones observadas mediante transformaciones de variables: Incidencia sobre la estructura factorial." Psicológica **29**(2).

Rodríguez Suárez, Y. and A. Díaz Amador (2009). "Herramientas de minería de datos." Revista Cubana de Ciencias Informáticas **3**(3-4).

Rojo, J. M. (2006). "Análisis descriptivo y exploratorio de datos." Laboratorio de Estadística del Instituto de Economía y Geografía Consejo Superior de Investigaciones Científicas, Madrid.

Sánchez Ibáñez, N. D. (2017). Metodología para la creación de informes de cosecha de Case-IH A8000 con SMS advanced, Universidad Central" Marta Abreu" de Las Villas. Facultad de Ingeniería

Schimmelpfennig, D. (2016). Farm profits and adoption of precision agriculture, United States Department of Agriculture, Economic Research Service.

Stengel, R. (2017). "Probability and Statistics!".

Testa, J. V. P. (2014). "Desempenho operacional e energético de colhedoras de cana-de-açúcar (*Saccharum* spp.) para uma e duas linhas da cultura."

Valero Ubierna, C., L. M. Navas Gracia, et al. (2010). Ahorro y eficiencia energética en la Agricultura de Precisión, Instituto para la Diversificación y Ahorro de la Energía.

Villardón, J. L. V. (2002). "Análisis de componentes principales." Cataluña: UOC, Departamento de Estadística **32**.

Voltarelli, M. A. (2015). "Ferramentas da qualidade na colheita mecanizada de cana-de-açúcar."

Wimukthi Yatagama Lokuge, S. (2015). Extensión del módulo de visualización científica de Weka, Universidad Central "Marta Abreu" de Las Villas.

Zamilpa, J., R. S. Rindermann, et al. (2016). "State of the art about the criticism of organic agriculture." Acta Universitaria **26**(2): 20-29.

ANEXOS

Anexo I Fragmento de diagramas de dispersión de variables del conjunto 3



Anexo II Histogramas de variables del conjunto 3



Anexo III Comparación de datos transformadas y no transformadas con correlación

	Distanci atrans	Carga trans	Tempaceit ehidtrans	Prescortad basetrans	Velociventit rans	Consumo areatrans	Consumohor ariotrans	Curso trans	Temprefri gertran	Voltba ttrans	Elevacio ntrans	Fuerzatr ans	RPMmo ttrans	Veloci trans
Distancia(m)	.980					-.915								
Carga mot.(%)		.998					.800					.987		
Cargatrans							.802					.989		
Temp. aceite hid.(°C)			-.946											
Tempaceitehidtrans														
Pres cortador(kPa)				.975										
Prescortadbasetrans														
Velocidad ventilador(rpm)					-.914									
Velociventitrans														
Cons. comb. (A)(L/ha)	-.775					.869								-.822
Consumoareatrans	-.954													-.958
Cons. comb. (T)(L/hr)		.798					.996					.821		
Consumohorariotrans		.802										.827		
Curso(deg)								.942						
Cursotrans														
Temp refrigerante(°C)									.999					
Temprefrigertran														
Volts bat(V)										1.000				
Voltbattrans														
Elevación(m)											.955			
Elevaciontrans														
Fuerza mot.(W)		.988					.825							
Fuerzatrans		.989					.827							
RPM motor(rpm)													.996	
RPMmottrans														
Velocitrans	.988					-.958								
Velocidad(km/hr)	.974	.057		-.054	-.074	-.910	.074	.043	.130	.060	-.065	.060	.029	.961

Anexo IV Estadística descriptiva de variables no transformadas

Estadística descriptiva de variables no transformadas

	Conjunto 1		Conjunto 2		Conjunto 3	
	Media	Desv. Esta.	Media	Desv. Esta.	Media	Desv. Esta.
Eficiencia combustible(ha/L)	.020249	.0047767	.017146	.0039590	.016445	.0044216
Eficiencia combustible (Km/L)	.121397	.0266494	.114254	.0263089	.109662	.0293702
Productividad	1.00144	.241931	.907453	.2119085	.816017	.2089288
Velocidad	5.99406	1.311814	6.049732	1.4127652	5.440140	1.3928615
Carga motor	69.576	5.5601	74.320630	4.2927370	70.084539	5.3361364
Consumo combustible(L/ha)	52.9906	20.20936	62.135383	20.7523804	67.059325	33.4576633
Consumo combustible(L/hr)	49.5468	4.52363	53.024007	3.7692384	49.894343	4.2855160
Fuerza motor	193591.53	15703.541	206248.87	11914.020	194827.18	14392.744
Distancia(m)	1.67096	.368053	1.683019	.3947533	1.516886	.3908435
Presión cortador base	5679.580	1591.7705	5253.902978	1045.6697464	5759.983106	1463.9437576
Temperatura aceite hidráulico	57.8853	2.74410	53.484465	5.5299181	56.839253	4.8862906
Velocidad ventilador	566.053	500.6781	1023.520291	35.8621278	1002.605211	32.3339042
RPM motor	1798.720	24.6665	1796.653	40.3103	1797.620196	13.2615928
Temperatura refrigerante	85.4636	.71556	85.120886	2.1540251	85.485591	.8364646
Curso	119.223	93.1036	116.592930	94.5168762	121.606130	95.2776693
Voltaje batería	13.2308	.11131	13.0675	.36578	13.2223	.16060
Elevación	3.05043	3.410343	9.251746	3.3384042	10.781099	3.7248672
Desviación Y(m)	-.000349070	.0006798515	.000742746	.0005514178	.000144649	.0007497566
Combustible utilizado(L)	.013802	.0011918	.014740	.0009440	.013901	.0011048
Duración(s)	1.004	.0308	1.001695	.0310787	1.004012	.0326357

Anexo V Estadística descriptiva de variables transformadas

Estadística descriptiva de variables transformadas

	Conjunto 1		Conjunto 2		Conjunto 3	
	Media	Desv. Esta.	Media	Desv. Esta.	Media	Desv. Esta.
Eficiencia combustible(ha/L)	.020249	.0047767	.017146	.0039590	.016445	.0044216
Eficiencia combustible (Km/L)	.121397	.0266494	.114254	.0263089	.109662	.0293702
Productividad	1.001440	.2419305	.907453	.2119085	.816017	.2089288
Velocidad	.8143	.09487	6.049732	1.4127652	5.440140	1.3928615
Carga motor	1.8410	.03522	1.8704	.02546	1.8443	.03344
Consumo combustible(L/ha)	1.7070	.11305	1.7782	.10697	1.8015	.13193
Consumo combustible(L/hr)	1.6932	.04036	1.7234	.03129	1.6964	.03766
Fuerza motor	5.2854	.03576	5.3137	.02534	5.2884	.03249
Distancia(m)	1.670956	.3680529	1.683019	.3947533	1.516886	.3908435
Presión cortador base	3.7433	.08948	3.7126	.10769	3.7500	.08946
Temperatura aceite hidráulico	1.7620	.02233	1.7332	.06069	.9541	.22315
Velocidad ventilador	1.6904	1.48909	3.0091	.06729	3.0012	.02408
RPM motor	3.2549	.00616	3.2533	.07164	3.2547	.00341
Temperatura refrigerante	1.9318	.00364	1.9343	.04297	1.9319	.00449
Curso	1.8546	.50780	1.8380	.51129	1.8583	.52186
Voltaje batería	1.1216	.00366	1.1478	.02620	1.1213	.00525
Elevación	.4386	.39336	.9776	.19383	1.0465	.15533
Desviación Y(m)	-.000349070	.000679851	.000742746	.0005514178	.000144649	.0007497566
		5				
Combustible utilizado(L)	.013802	.0011918	.014740	.0009440	.013901	.0011048
Duración(s)	1.003749	.0308161	1.001695	.0310787	1.004012	.0326357
Presión aceite motor	2.7745	.00059	2.7739	.06106	2.7745	.00032

Anexo VI Análisis de correlación mediante

Análisis de correlación mediante

Programa	SMS Advanced			WEKA		
Conjuntos	Conjunto 1	Conjun to 2	Conjun to 3	Conjunto 1	Conjun to 2	Conjunt o 3
Pares analizados						
$y = \text{Eficiencia combustible(ha/L)}$	1.000	1.000	1.000	0.920	0.997	0.998
$x = \text{Eficiencia combustible(Km/L)}$						
$y = \text{Productividad}$ $x = \text{Velocidad}$	0.930	1.000	1.000	0.927	0.926	1.000
$y = \text{Fuerza motor}$ $x = \text{Carga motor}$	0.980	0.990	0.990	0.980	0.990	0.989
$y = \text{Distancia}$ $x = \text{Velocidad}$	0.990	0.991	0.992	0.990	0.99	0.991
$y = \text{Distancia}$ $x = \text{productividad}$	0.916	0.991	0.992	0.915	0.91	0.991
$y = \text{Consumo combustible (L/hr)}$ $x = \text{Combustible utilizado}$	0.943	0.900	0.923	0.943	0.900	0.923
$y = \text{Eficiencia combustible(Km/L)}$ $x = \text{Distancia}$	0.920	0.963	0.950	0.919	0.962	0.953
$y = \text{Fuerza motor}$ $x = \text{Combustible utilizado}$	0.911	0.874	0.892	0.910	0.874	0.891
$y = \text{Velocidad}$ $x = \text{Eficiencia combustible(Km/L)}$	0.910	0.953	0.940	0.910	0.953	0.945

$y = \text{productividad}$	0.815	0.953	0.940	0.921	0.950	0.945
$x = \text{Eficiencia combustible(Km/L)}$						
$y = \text{Distancia}$	0.920	0.963	0.950	0.873	0.960	0.952
$x = \text{Eficiencia combustible(ha/L)}$						
$y = \text{Velocidad}$	0.910	0.953	0.940	0.865	0.950	0.945
$x = \text{Eficiencia combustible(ha/L)}$						
$y = \text{productividad}$	0.815	0.953	0.940	0.815	0.953	0.945
$x = \text{Eficiencia combustible(ha/L)}$						
$y = \text{Carga motor}$	0.880	0.849	0.866	0.880	0.848	0.865
$x = \text{Combustible utilizado}$						
$y = \text{Fuerza motor}$	0.858	0.788	0.823	0.86	0.788	0.822
$x = \text{Consumo combustible(L/hr)}$						
$y = \text{Carga motor}$	0.829	0.764	0.799	0.830	0.763	0.798
$x = \text{Consumo combustible(L/hr)}$						
$y = \text{Velocidad}$	-0.718	-0.765	-0.690	-0.708	-0.708	-0.690
$x = \text{Consumo combustible(L/ha)}$						
$y = \text{Consumo combustible(L/ha)}$	-0.720	-0.765	-0.690	-0.720	-0.720	-0.690
$x = \text{Productividad}$						
$y = \text{Consumo combustible(L/ha)}$	-0.727	-0.790	-0.699	-0.750	-0.790	-0.699
$x = \text{Eficiencia combustible(Km/L)}$						
$y = \text{Consumo combustible(L/ha)}$	-0.727	-0.790	-0.699	-0.756	-0.789	-0.699
$x = \text{Eficiencia combustible(ha/L)}$						
$y = \text{Distancia}$	-0.711	-0.770	-0.692	-0.710	-0.770	-0.692
$x = \text{Consumo combustible(L/ha)}$						

$y = \text{Presión cortador base}$	0.352	0.4	0.354	0.352	0.401	0.353
$x = \text{Carga motor}$						
$y = \text{Consumo combustible(L/hr)}$	-0.325	-0.434	-0.370	-0.320	-0.434	-0.369
$x = \text{Duración}$						
$y = \text{Presión cortador base}$	0.333	0.404	0.351	0.333	0.404	0.351
$x = \text{Fuerza motor}$						
$y = \text{Temperatura Aceite Hidraulico}$	0.256	0.372	0.411	0.256	0.372	0.410
$x = \text{Temp.refrigerante}$						
$y = \text{Presión cortador base}$	0.296	0.309	0.297	0.296	0.309	0.296
$x = \text{Consumo combustible(L/hr)}$						
$y = \text{Presión cortador base}$	0.271	0.262	0.266	0.271	0.262	0.265
$x = \text{Combustible utilizado}$						
$y = \text{Presión aceite motor } x = \text{RPM motor}$	0.589	0.98	0.485	0.588	0.98	0.485

Anexo VII Análisis de correlación mediante SPSS

Análisis de correlación mediante SPSS

Conjuntos Pares analizados	Conjunto 1	Conjunto 2	Conjunto 3
$y = \text{Eficiencia combustible(ha/L)}$	0.919(p=0.00)	0.997(p=0.00)	0.998(p=0.00)
$x = \text{Eficiencia combustible(Km/L)}$			
$y = \text{Productividad}$ $x = \text{Velocidad}$	0.927(p=0.00)	1.000(p=0.00)	1.000(p=0.00)
$y = \text{Fuerza motor}$ $x = \text{Carga motor}$	0.980(p=0.00)	0.990(p=0.00)	0.990(p=0.00)
$y = \text{Distancia}$ $x = \text{Velocidad}$	0.990(p=0.00)	0.990(p=0.00)	0.992(p=0.00)
$y = \text{Distancia}$ $x = \text{productividad}$	0.916(p=0.00)	0.990(p=0.00)	0.992(p=0.00)
$y = \text{Eficiencia comb. (Km/L)}$ $x = \text{Distancia}$	0.873(p=0.00)	0.960(p=0.00)	0.952(p=0.00)
$y = \text{Distancia}$ $x = \text{Eficiencia comb. (ha/L)}$	0.873(p=0.00)	0.960(p=0.00)	0.952(p=0.00)
$y = \text{productividad}$ $x = \text{Eficiencia comb. (ha/L)}$	0.922(p=0.00)	0.950(p=0.00)	0.944(p=0.00)
$y = \text{Consumo combust(L/ha)}$ $x = \text{Eficiencia comb. (ha/L)}$	-0.756(p=0.00)	-0.790(p=0.00)	-0.697(p=0.00)

$y = \text{Velocidad}$	0.866(p=0.00)	0.950(p=0.00)	0.944(p=0.00)
$x = \text{Eficiencia comb. (ha/L)}$			
$y = \text{Consumo comb (L/hr)}$	0.943(p=0.00)	0.900(p=0.00)	0.923(p=0.00)
$x = \text{Combustible utilizado}$			
$y = \text{Distancia}$	-0.711(p=0.00)	-0.770(p=0.00)	-0.692(p=0.00)
$x = \text{Consumo comb(L/ha)}$			
$y = \text{Velocidad}$	0.866(p=0.00)	0.950(p=0.00)	0.944(p=0.00)
$x = \text{Eficiencia comb. (Km/L)}$			
$y = \text{productividad}$	0.922(p=0.00)	0.950(p=0.00)	0.944(p=0.00)
$x = \text{Eficiencia combustible(Km/L)}$			
$y = \text{Velocidad}$	-0.708(p=0.00)	-0.760(p=0.00)	-0.691(p=0.00)
$x = \text{Consumo combustible(L/ha)}$			
$y = \text{Consumo combustible(L/ha)}$	-0.720(p=0.00)	-0.760(p=0.00)	-0.691(p=0.00)
$x = \text{Productividad}$			
$y = \text{Consumo combustible(L/ha)}$	-0.756(p=0.00)	-0.790(p=0.00)	-0.697(p=0.00)
$x = \text{Eficiencia combustible(Km/L)}$			
$y = \text{Carga motor}$	0.880(p=0.00)	0.850(p=0.00)	0.866(p=0.00)
$x = \text{Combustible utilizado}$			
$y = \text{Fuerza motor}$	0.911(p=0.00)	0.874(p=0.00)	0.892(p=0.00)
$x = \text{Combustible utilizado}$			
$y = \text{Carga motor}$	0.829(p=0.00)	0.764(p=0.00)	0.799(p=0.00)
$x = \text{Consumo combustible(L/hr)}$			
$y = \text{Fuerza motor}$	0.858(p=0.00)	0.788(p=0.00)	0.823(p=0.00)

$x = \text{Consumo combustible(L/hr)}$

$y = \text{Presión cortador base}$ 0.352(p=0.00) 0.400(p=0.00) 0.354(p=0.00)

$x = \text{Carga motor}$

$y = \text{Consumo combustible(L/hr)}$ -0.323(p=0.00) -0.430(p=0.00) -0.369(p=0.00)

$x = \text{Duración}$

$y = \text{Presión cortador base}$ 0.333(p=0.00) 0.400(p=0.00) 0.351(p=0.00)

$x = \text{Fuerza motor}$

$y = \text{Temperatura aceite hidraulico}$ 0.256(p=0.00) 0.372(p=0.00) 0.411(p=0.00)

$x = \text{Temperatura refrigerante}$

$y = \text{Presión cortador base}$ 0.296(p=0.00) 0.310(p=0.00) 0.297(p=0.00)

$x = \text{Combustible utilizado}$

$y = \text{Presión aceite motor } x = \text{RPM motor}$ 0.589(p=0.000) 0.98(p=0.000) 0.485(p=0.000)

Anexo VIII Análisis de regresión mediante WEKA

Análisis de regresión mediante WEKA

Conjuntos Pares analizados	Conjunto 1	Conjunto 2	Conjunto 3
$y = \text{Eficiencia combustible(ha/L)}$ $x = \text{Eficiencia comb. (Km/L)}$	$5.13x + 0.02$	$6.63x$	$6.63x$
$y = \text{Productividad}$ $x = \text{Velocidad}$	$0.17x - 0.02$	$0.15x$	$0.15x$
$y = \text{Fuerza motor}$ $x = \text{Carga motor}$	$2767.43x$	$2748.52x + 1977.28$	$2669.8x + 7715.39$
$y = \text{Distancia}$ $x = \text{productividad}$	$1.39x + 0.28$	$1.85x + 0.01$	$1.85x$
$y = \text{Distancia}$ $x = \text{Velocidad}$	$0.28x + 0.01$	$0.28x + 0.01$	$0.28x$
$y = \text{Eficiencia comb. (Km/L)}$ $x = \text{Distancia}$	$0.07x + 0.01$	$0.06x + 0.01$	$0.01x$
$y = \text{Consumo combustible (L/hr)}$ $x = \text{Combustible utilizado}$	$3580.16x + 0.13$	$3595.47x + 0.03$	$3582.28x + 0.1$
$y = \text{Distancia}$ $x = \text{Consumo combustible(L/ha)}$	$-0.01x + 2.36$	$-0.01x + 2.59$	$-0.01x + 2.06$
$y = \text{Velocidad}$ $x = \text{Eficiencia combustible(Km/L)}$	$44.74x + 0.56$	$51.19x + 0.2$	$44.85x + 0.52$
$y = \text{productividad}$ $x = \text{Eficiencia combustible(Km/L)}$	$7.4x + 0.1$	$7.68x + 0.03$	$6.73x + 0.08$

$y = \text{Velocidad}$			
$x = \text{Consumo combustible(L/ha)}$	$-0.05x + 8.43$	$-0.05x + 9.28$	$-0.03x + 7.37$
$y = \text{Consumo combustible(L/ha)}$			
$x = \text{Productividad}$	$-60.16x + 113.23$	$-74.87x + 130.08$	$-110.6x + 157.31$
$y = \text{Consumo combustible(L/ha)}$	$-550.99x + 119.88$	$-624.34x + 133.47$	$-796.06x + 154.36$
$x = \text{Eficiencia combustible(Km/L)}$			
$y = \text{Carga motor}$	$4106.02x + 12.9$	$3859x + 17.44$	$4181x + 11.97$
$x = \text{Consumo combustible(L/hr)}$			
$y = \text{Fuerza motor}$	$12001445.82x + 27945103$	$1036107.88x + 43579.84$	$11615174.22x + 33367.36$
$x = \text{Combustible utilizado}$			
$y = \text{Carga motor}$	$1.02x + 19.08$	$0.87x + 28.19$	$0.99x + 20.46$
$x = \text{Consumo comb(L/hr)}$			
$y = \text{Fuerza motor}$	$2979.34x + 45974.83$	$2491.26x + 74152.28$	$2763.57x + 56940.75$
$x = \text{Consumo combustible(L/hr)}$			

Anexo IX Análisis de regresión mediante SPSS

Análisis de regresión mediante SPSS

Conjuntos Pares analizados	Conjunto 1	Conjunto 2	Conjunto 3
$y = \text{Eficiencia combustible(ha/L)}$	$5.13x + 0.018$	$6.628x + 0.001$	$6.66x$
$x = \text{Eficiencia comb. (Km/L)}$	$(R^2 = 0.846)$	$(R^2 = 0.995)$	$(R^2 = 0.996)$
$y = \text{Productividad}$	$0.171x - 0.023$	$0.15 x$	$0.15x$
$x = \text{Velocidad}$	$(R^2 = 0.859)$	$(R^2 = 1)$	$(R^2 = 1)$
$y = \text{Fuerza motor}$	$2767.4x + 1045.71$	$2748.518x + 1977$	$2669.8x + 7715.39$
$x = \text{Carga motor}$	$(R^2 = 0.960)$	$(R^2 = 0.981)$	$(R^2 = 0.99)$
$y = \text{Distancia}$	$1.393x + 0.276$	$1.846x + 0.008$	$1.85x + 0.003$
$x = \text{productividad}$	$(R^2 = 0.838)$	$(R^2 = 0.982)$	$(R^2 = 0.983)$
$y = \text{Distancia}$	$0.278x + 0.006$	$0.277x + 0.008$	$0.278x + 0.003$
$x = \text{Velocidad}$	$(R^2 = 0.980)$	$(R^2 = 0.982)$	$(R^2 = 0.983)$
$y = \text{Eficiencia comb. (Km/L)}$ $x = \text{Distancia}$	$0.067x + 0.01$	$0.064x + 0.006$	$0.072x + 0.001$
	$(R^2 = 0.845)$	$(R^2 = 0.927)$	$(R^2 = 0.909)$
$y = \text{Distancia}$	$67.301x + 0.308$	$95.74x + 0.041$	$84.181x + 0.133$
$x = \text{Eficiencia comb. (ha/L)}$	$(R^2 = 0.76)$	$(R^2 = 0.9)$	$(R^2 = 0.9)$
$y = \text{Consumo combustible (L/hr)}$	$3580.165x + 0.133$	$3595.467x + 0.028$	$3582.277x + 0.098$
$x = \text{Combustible utilizado}$	$(R^2 = 0.89)$	$(R^2 = 0.811)$	$(R^2 = 0.853)$

$y = \text{Velocidad}$	$44.741x + 0.563$	$51.191x + 0.201$	$44.853x + 0.521$
$x = \text{Eficiencia combustible(Km/L)}$	$(R2 = 0.826)$	$(R2 = 0.909)$	$(R2 = 0.895)$
$y = \text{Velocidad}$	$237.796x + 1.179$	$339.18x + 0.233$	$297.527x + 0.547$
$x = \text{Eficiencia comb. (ha/L)}$	$(R2 = 0.75)$	$(R2 = 0.9)$	$(R2 = 0.9)$
$y = \text{productividad}$	$7.395x + 0.104$	$7.679x + 0.03$	$6.728x + 0.078$
$x = \text{Eficiencia comb. (Km/L)}$	$(R2 = 0.664)$	$(R2 = 0.909)$	$(R2 = 0.895)$
$y = \text{productividad}$	$46.684x + 0.056$ $(R^2=0.85)$	$50.875x + 0.035$ $(R^2=0.9)$	$44.629x + 0.082$
$x = \text{Eficiencia comb. (ha/L)}$			$(R2 = 0.89)$
$y = \text{Velocidad}$			
$x = \text{Consumo combustible(L/ha)}$	$-0.046x + 8.43$ $(R2 = 0.501)$	$-0.052x + 9.284$ $(R2 = 0.584)$	$-0.029x + 7.368$ $(R2 = 0.477)$
$y = \text{Consumo combust(L/ha)}$			
$x = \text{Productividad}$	$-60.155x + 113.232$ $(R2 = 0.519)$	$-74.872x + 130.078$ $(R2 = 0.585)$	$-110.596x + 157.307$ $(R2 = 0.477)$
$y = \text{Consumo combustible(L/ha)}$	$-550.986x + 119.879$	$-624.337x + 133.468$	$-796.058x + 154.357$
$x = \text{Eficiencia combustible(Km/L)}$	$(R2 = 0.528)$	$(R2 = 0.626)$	$(R2 = 0.488)$
$y = \text{Consumo combust(L/ha)}$	$-3199.563x + 117.777$	$-4138.469x + 133.095$	$-5274.465x + 153.79$
$x = \text{Eficiencia combustible(ha/L)}$	$(R2 = 0.572)$	$(R2 = 0.623)$	$(R2 = 0.486)$
$y = \text{Carga motor}$	$4106.021x + 12.904$	$3858.998x + 17.44$	$4181.997x + 11.965$
$x = \text{Combustible utilizado}$	$(R2 = 0.775)$	$(R2 = 0.720)$	$(R2 = 0.749)$

$y = \text{Fuerza motor}$	$12001446x + 27946$	$11036108x + 43580$	$11615174x + 33367$
$x = \text{Combustible utilizado}$	$(R2 = 0.83)$	$(R2 = 0.765)$	$(R2 = 0.795)$
$y = \text{Carga motor}$	$1.010x + 19.081$	$0.87x + 28.189$	$0.995x + 20.46$
$x = \text{Consumo combustible(L/hr)}$	$(R2 = 0.687)$	$(R2 = 0.584)$	$(R2 = 0.638)$
$y = \text{Fuerza motor}$	$2979x + 45975$	$2491.26x + 74152$	$2763.568x + 56940.75$
$x = \text{Consumo combustible(L/hr)}$	$(R2 = 0.737)$	$(R2 = 0.621)$	$(R2 = 0.677)$
$y = \text{Distancia}$	$-0.013x + 2.357$	$-0.015x + 2.594$	$-0.008x + 2.059$
$x = \text{Consumo combustible(L/ha)}$	$(R2 = 0.505)$	$(R2 = 0.594)$	$(R2 = 0.479)$

Anexo X Calidad de Representación

Cualidad de Representación

	Conjunto 1		Conjunto 2		Conjunto 3	
	Inicial	Extracción	Inicial	Extracción	Inicial	Extracción
Eficiencia combustible(ha/L)	1.000	.982	1.000	.991	1.000	.989
Eficiencia combustible(Km/L)	1.000	.985	1.000	.994	1.000	.991
productividad	1.000	.982	1.000	.993	1.000	.994
Velocidad	1.000	.979	1.000	.993	1.000	.994
Carga motor	1.000	.915	1.000	.927	1.000	.930
Consumo combustible(L/ha)	1.000	.966	1.000	.929	1.000	.924
Consumo combustible(L/hr)	1.000	.964	1.000	.958	1.000	.965
Fuerza motor	1.000	.940	1.000	.950	1.000	.942
Distancia(m)	1.000	.983	1.000	.992	1.000	.993
Presión cortador base	1.000	.437	1.000	.731	1.000	.449
Temperatura aceite hidráulico	1.000	.698	1.000	.670	1.000	.625
Velocidad ventilador	1.000	.896	1.000	.969	1.000	.337
RPM motor	1.000	.804	1.000	.977	1.000	.784
Temperatura refrigerante	1.000	.593	1.000	.977	1.000	.553
Curso	1.000	.801	1.000	.736	1.000	.587
Voltaje batería	1.000	.834	1.000	.947	1.000	.567
Elevación	1.000	.494	1.000	.428	1.000	.454
Desviación Y(m)	1.000	.537	1.000	.915	1.000	.519
Combustible utilizado(L)	1.000	.931	1.000	.915	1.000	.923
Duración(s)	1.000	.928	1.000	.977	1.000	.936
Presión aceite motor	1.000	.779	1.000	.978	1.000	.728

Anexo XI Varianza Total Explicada del conjunto 1

Varianza Total Explicada del conjunto 1

Componente	Valores propios iniciales			Suma de factores retenidos		
	Total	% varianza	% acumulada	Total	% varianza	% acumulada
1	5.931	28.242	28.242	5.931	28.242	28.242
2	4.153	19.777	48.018	4.153	19.777	48.018
3	2.225	10.596	58.614	2.225	10.596	58.614
4	1.533	7.302	65.916	1.533	7.302	65.916
5	1.381	6.575	72.491	1.381	6.575	72.491
6	1.142	5.438	77.929	1.142	5.438	77.929
7	1.064	5.069	82.997	1.064	5.069	82.997
8	.916	4.361	87.359			
9	.731	3.481	90.840			
10	.645	3.070	93.910			
11	.503	2.396	96.305			
12	.390	1.856	98.161			
13	.157	.746	98.907			
14	.138	.658	99.565			
15	.065	.308	99.874			
16	.010	.049	99.922			
17	.007	.032	99.954			
18	.005	.022	99.976			
19	.003	.016	99.992			
20	.001	.006	99.998			
21	.000	.002	100.000			

Anexo XII Varianza Total Explicada del conjunto 2

Varianza Total Explicada del conjunto 2

Componente	Valores propios iniciales			Suma de factores retenidos		
	Total	% varianza	% acumulada	Total	% varianza	% acumulada
1	6.126	29.172	29.172	6.126	29.172	29.172
2	5.829	27.756	56.929	5.829	27.756	56.929
3	3.734	17.779	74.707	3.734	17.779	74.707
4	1.175	5.597	80.304	1.175	5.597	80.304
5	1.081	5.149	85.453	1.081	5.149	85.453
6	1.004	4.780	90.233	1.004	4.780	90.233
7	.913	4.350	94.583			
8	.392	1.865	96.447			
9	.334	1.590	98.038			
10	.233	1.111	99.148			
11	.081	.388	99.536			
12	.051	.244	99.780			
13	.018	.088	99.868			
14	.008	.039	99.907			
15	.008	.037	99.944			
16	.008	.037	99.980			
17	.002	.009	99.989			
18	.001	.007	99.997			
19	.001	.003	99.999			
20	.000	.001	100.000			
21	0.000	0.000	100.000			

Anexo XIII Varianza Total Explicada del conjunto 3

Varianza Total Explicada del conjunto 3

Componente	Valores propios iniciales			Suma de factores retenidos		
	Total	% varianza	% acumula	Total	% varianza	% acumula
1	5.978	28.466	28.466	5.978	28.466	28.466
2	3.906	18.601	47.067	3.906	18.601	47.067
3	1.588	7.561	54.628	1.588	7.561	54.628
4	1.430	6.809	61.436	1.430	6.809	61.436
5	1.144	5.450	66.886	1.144	5.450	66.886
6	1.095	5.214	72.100	1.095	5.214	72.100
7	1.041	4.958	77.058	1.041	4.958	77.058
8	.977	4.654	81.712			
9	.902	4.295	86.007			
10	.849	4.044	90.051			
11	.742	3.533	93.584			
12	.577	2.747	96.331			
13	.524	2.494	98.825			
14	.138	.656	99.481			
15	.087	.415	99.896			
16	.012	.056	99.952			
17	.005	.024	99.977			
18	.002	.011	99.988			
19	.002	.008	99.997			
20	.001	.003	100.000			
21	9.580E-7	4.562E-6	100.000			

Anexo XIV Matriz De Componentes Correlacionadas Con Variables Iniciales Conjunto1

Matriz De Componentes Correlacionadas Con Variables Iniciales Conjunto1

	Componentes						
	1	2	3	4	5	6	7
Eficiencia combustible(ha/L)	.883	-.434	-.110	-.005	-.030	.028	-.005
Eficiencia combustible(Km/L)	.795	-.540	.215	-.093	-.075	.008	-.009
productividad	.981	-.096	-.066	-.010	.009	-.059	-.057
Velocidad	.927	-.168	.271	-.074	-.040	-.081	-.072
Carga motor	.401	.828	.069	-.234	.055	.016	.079
Consumo combustible(L/ha)	-.893	.403	.060	-.020	.031	-.022	.011
Consumo combustible(L/hr)	.364	.862	.126	-.003	.070	-.214	-.148
Fuerza motor	.415	.855	.132	-.099	.035	.035	.085
Distancia(m)	.928	-.195	.260	-.117	-.055	.010	.019
Presión cortador base	.042	.408	-.030	-.441	.116	.239	.048
Temperatura aceite hidráulico	.088	-.269	.087	.148	.740	-.045	.199
Velocidad ventilador	.372	.177	-.816	.213	.107	.059	.017
RPM motor	.179	.335	.356	.713	-.108	.105	.046
Temperatura refrigerante	.036	.012	.239	-.074	.525	.434	-.253
Curso	.049	.013	.027	-.013	.040	-.531	.717
Voltaje batería	.403	.236	-.764	.153	-.041	.085	.004
Elevación	-.001	-.132	.153	.135	.647	-.118	.057
Desviación Y(m)	.238	.107	-.613	.220	.203	.053	.025
Combustible utilizado(L)	.390	.867	.134	-.069	.031	.014	.058
Duración(s)	.031	-.125	.047	-.099	-.135	.701	.624
Presión aceite motor	.136	.202	.344	.748	-.135	.152	.026

Anexo XV Matriz De Componentes Correlacionadas Con Variables Iniciales Conjunto2

Matriz De Componentes conjunto 2

	Componentes					
	1	2	3	4	5	6
Eficiencia combustible(ha/L)	.627	-.724	-.265	.033	-.038	.006
Eficiencia combustible(Km/L)	.628	-.726	-.265	.032	-.038	.006
productividad	.647	-.753	.003	-.064	-.070	-.007
Velocidad	.647	-.752	.003	-.064	-.070	-.007
Carga motor	.161	-.166	.916	.149	.114	.006
Consumo combustible(L/ha)	-.616	.704	.228	-.039	.028	-.001
Consumo combustible(L/hr)	.118	-.149	.905	-.303	-.097	-.041
Fuerza motor	.162	-.168	.929	.141	.114	.009
Distancia(m)	.649	-.753	-.020	.051	-.015	.010
Presión cortador base	.614	.521	.260	.121	.001	-.028
Temperatura aceite hidráulico	.689	.303	-.223	-.123	.181	-.080
Velocidad ventilador	.746	.642	.002	-.012	-.015	.013
RPM motor	.732	.663	.007	.005	-.047	.024
Temperatura refrigerante	.753	.639	.004	.004	-.030	.018
Curso	.055	-.032	-.116	-.301	.744	-.273
Voltaje batería	.742	.624	-.076	-.024	-.016	.014
Elevación	-.115	.132	.062	.335	-.507	-.158
Desviación Y(m)	-.009	.002	.003	-.177	.086	.936
Combustible utilizado(L)	.148	-.177	.920	.081	.088	.017
Duración(s)	.044	-.038	-.180	.870	.410	.129
Presión aceite motor	.731	.664	.005	.007	-.047	.023

Anexo XVI Matriz De Componentes Correlacionadas Con Variables Iniciales Conjunto3

Matriz De Componentes conjunto 3

	Componentes						
	1	2	3	4	5	6	7
Eficiencia combustible(ha/L)	.992	.024	-.005	-.058	.037	-.009	-.007
Eficiencia combustible(Km/L)	.993	.023	-.006	-.058	.037	-.009	-.008
productividad	.943	.310	.017	-.064	-.062	-.019	-.024
Velocidad	.943	.310	.017	-.064	-.062	-.019	-.024
Carga motor	-.224	.916	-.148	-.074	.091	.038	.063
Consumo combustible(L/ha)	-.957	-.067	-.001	.061	-.029	.014	.012
Consumo combustible(L/hr)	-.243	.899	.092	-.014	-.296	-.017	-.035
Fuerza motor	-.226	.933	-.059	-.045	.098	.043	.064
Distancia(m)	.948	.297	.006	-.058	.055	-.004	-.001
Presión cortador base	-.154	.367	-.362	.339	.137	.148	.057
Temperatura aceite hidráulico	-.334	.136	.160	-.648	.137	-.174	.021
Velocidad ventilador	.080	.005	-.038	.343	-.018	-.330	-.319
RPM motor	-.014	.137	.834	.251	.062	.052	.007
Temperatura refrigerante	.173	.044	-.159	.672	-.126	.117	.118
Curso	.050	.015	-.014	.015	-.019	.723	-.245
Voltaje batería	.057	.075	-.051	.226	.002	-.486	.519
Elevación	-.119	.115	-.008	-.122	.179	-.273	-.552
Desviación Y(m)	.052	-.176	.161	-.327	-.221	.252	.490
Combustible utilizado(L)	-.238	.927	.045	.001	.054	.027	.038
Duración(s)	.079	-.092	-.082	.044	.929	.121	.188
Presión aceite motor	.000	.105	.798	.243	.144	.007	-.015