Sage Sphere International Journal

ISSN: 3091-1567

Vol. 4 Núm. 1 (2026)

SSIJ-Vol.4. N1. 084

Estructura multivariante del abastecimiento de caña de azúcar: componentes principales, asociaciones longitudinales y segmentación exploratoria

Multivariate structure of sugarcane supply: principal components, longitudinal associations, and exploratory clustering

Autores:
Harold Carlos Pinckert Ortiz
Investigador independiente, Santa Cruz de la Sierra, Bolivia
hpinckert@gmail.com
https://orcid.org/0009-0004-0450-7092

Autor de correspondencia: *Harold Carlos Pinckert Ortiz*, hpinckert@gmail.com

Enviado: 29-10-2026 **Revisado:** 10-11-2026
Aceptado: 19-11-2026 **Publicado:** 20-11-2026

Cómo citar este artículo:
Pinckert Ortiz, H. C. (2026). Estructura multivariante del abastecimiento de caña de azúcar: componentes principales, asociaciones longitudinales y segmentación exploratoria. *Sage Sphere International Journal*, *3*(1), 1-17. https://doi.org/10.63688/rgnj0m52

© 2026; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (https://creativecommons.org/licenses/by/4.0) que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea correctamente citada.

ISSN: 3091-1567

Vol. 4 Núm. 1 (2026)

RESUMEN

El objetivo fue caracterizar la estructura del abastecimiento de caña de azúcar y determinar qué variables se asociaron con la espera en un ingenio de Santa Cruz, Bolivia. Se analizaron cinco zafras (2020–2024) mediante un diseño observacional longitudinal. Los datos comprendieron 153.160 registros proveedor–día y 6.598 observaciones proveedor–zafra. El análisis de componentes principales (ACP) y la segmentación K-means utilizaron 6.135 casos completos. La regresión de la espera incluyó 6.170 observaciones de 1.356 proveedores, con errores robustos y un modelo de efectos fijos de proveedor y zafra. Cuatro componentes superaron los umbrales del análisis paralelo y explicaron el 63,65 % de la varianza. El índice KMO fue 0,547. En la regresión por mínimos cuadrados ordinarios, el R^2 fue 0,201. Dentro de un mismo proveedor, una mayor distancia y una mayor puntualidad se asociaron con menor espera. La silueta máxima fue 0,140 para dos conglomerados de 3.183 y 2.952 observaciones. Ambos difirieron en escala de entrega y tiempos de servicio, con amplio solapamiento. La cobertura de los registros y la variación de la espera que queda sin explicar delimitan el alcance de estos resultados al caso estudiado.

Palabras clave: caña de azúcar; análisis multivariante; componentes principales; regresión longitudinal; conglomerados.

ABSTRACT

This study examined the structure of sugarcane supply and the variables associated with waiting time at a mill in Santa Cruz, Bolivia. A longitudinal observational design covered five harvest seasons (2020–2024), with 153,160 supplier-day records and 6,598 supplier-season observations. Principal component analysis and K-means clustering used 6,135 complete observations. The waiting-time regression included 6,170 observations from 1,356 suppliers, with robust standard errors and a model with supplier and season fixed effects. Four components exceeded the parallel analysis thresholds and accounted for 63.65% of the variance. The KMO index was 0.547. The ordinary least squares regression yielded an R^2 of 0.201. Within suppliers, greater distance and higher punctuality were associated with shorter waits. The highest silhouette score was 0.140 for two clusters of 3,183 and 2,952 observations. The clusters differed in delivery scale and service times but overlapped substantially. Record coverage and the unexplained variation in waiting time limit the scope of these findings to the case studied.

Keywords: sugarcane; multivariate analysis; principal components; longitudinal regression; clustering.

Doi: https://doi.org/10.63688/rgnj0m52

ISSN: 3091-1567
Vol. 4 Núm. 1 (2026)

1. INTRODUCCIÓN

Un ingenio debe recibir en una misma jornada caña de proveedores que difieren en superficie, distancia y frecuencia de entrega. El tiempo que transcurre desde el corte también importa para la conservación de la sacarosa (Solomon, 2009). Las revisiones de Ahumada y Villalobos (2009), Taşkıner y Bilgen (2021) y Teixeira et al. (2023) sitúan la coordinación entre cosecha, transporte y procesamiento entre las decisiones centrales de estas cadenas.

Singh y Pathak (1994) estudiaron conjuntamente la cosecha mecanizada y el número de camiones en Tailandia. En ese mismo país, Piewthongngam et al. (2009) relacionaron el crecimiento del cultivo con el calendario de abastecimiento. En Brasil se han examinado la elección de proveedores y el volumen de molienda (Paiva & Morabito, 2009), la incertidumbre de la cosecha (da Silva et al., 2015) y las rutas de las cosechadoras (Santoro et al., 2017). En Sudáfrica, Stray et al. (2012) analizaron la programación estacional de los productores.

Bocanegra-Herrera y Vidal-Holguín (2016) simularon el abastecimiento para estudiar cómo se coordinan sus operaciones. Los trabajos sobre cosecha y transporte de Arjona et al. (2001), Higgins et al. (2004), Higgins y Davies (2005) y López-Milán et al. (2006) muestran por qué conviene observar juntos el volumen movilizado y el tiempo de servicio. En los registros de un ingenio, parte de esa relación también puede proceder de cómo se contabilizan las entregas.

El objetivo fue caracterizar la estructura multivariante del abastecimiento de Empresa A durante las zafras 2020–2024. Se examinaron tres preguntas. ¿Qué dimensiones resumen la variación de los indicadores? ¿Qué características se asocian con la espera cuando se comparan proveedores y cuando se sigue al mismo proveedor? ¿Hasta qué punto se distinguen los perfiles obtenidos mediante agrupamiento?

2. METODOLOGÍA

Diseño, fuentes y unidades de análisis

Se realizó un estudio observacional de caso único en Santa Cruz, Bolivia. Empresa A facilitó boletas de recepción y pesaje, análisis de calidad de la caña y maestros de proveedores y propiedades. Se incluyeron los proveedores con entregas identificables en al menos una de las cinco zafras. La clave diaria combinó el código de proveedor con la fecha industrial. La

Doi: https://doi.org/10.63688/rgnj0m52

clave anual combinó el mismo código con la zafra. Las matrices consolidadas contenían una observación por clave.

Se utilizaron dos bases con las mismas 6.598 claves anuales y 1.565 proveedores distintos. La primera conserva el detalle proveedor–día y sus agregados anuales para las correlaciones, el ACP y K-means. La segunda contiene el panel anual utilizado para la regresión. En 2020, la primera inicia el 6 de junio y la segunda el 4 de junio. Para las otras cuatro zafras coinciden las fechas de inicio y cierre. Los métodos y las muestras se cotejaron con el capítulo V y el anexo 3 de Pinckert Ortiz (2026).

Selección de observaciones y medición

Se sumaron las toneladas por proveedor y fecha. En la primera base, el campo denominado «viajes» cuenta los ítems de boleta. Un ticket físico puede incluir varios ítems. El indicador consignado como capacidad del camión equivale, en las 6.598 claves, a las toneladas divididas por esos ítems. En el ACP y K-means se interpreta como carga por ítem. En la regresión, la carga efectiva divide las toneladas por el número de tickets distintos del proveedor en la zafra. En 2024 se contaron 97.077 ítems y 66.931 tickets.

Los días activos son las fechas distintas con entregas. La espera se expresa en minutos desde el ingreso registrado a playa hasta el inicio de la atención, y la descarga abarca desde el inicio hasta el fin de la atención. La puntualidad se expresa como proporción y su definición depende de la base utilizada. En la primera base, las medias anuales de espera, descarga y puntualidad ponderan el dato diario por el número de ítems con información. Sacarosa aparente (Pol), fibra y materia extraña se expresan en porcentaje. El área de referencia procede del maestro de propiedades.

Tabla 1

Cobertura histórica y selección de observaciones por zafra

Zafra	Proveedor–día	Proveedor–zafra	ACP yK-means	ExcluidasACP / K-means	Regresión	Excluidasregresión
2020	30.883	1.299	1.102	197	1.117	182
2021	31.281	1.306	1.164	142	1.182	124
2022	30.920	1.333	1.252	81	1.252	81
2023	32.976	1.332	1.304	28	1.306	26

Doi: https://doi.org/10.63688/rgnj0m52

Zafra	Proveedor–día	Proveedor–zafra	ACP yK-means	ExcluidasACP / K-means	Regresión	Excluidasregresión
2024	27.100	1.328	1.313	15	1.313	15
Total	153.160	6.598	6.135	463	6.170	428

Nota. Los recuentos diarios corresponden a la primera base. Las demás columnas cuentan observaciones proveedor–zafra. Los totales de cada método se calcularon por separado. Cálculos a partir de los registros administrativos de Empresa A.

El filtro del ACP y K-means excluyó 463 filas. Faltaban solo área y distancia en 418, solo calidad en 32, área, distancia y calidad en 6, solo descarga en 5 y área, distancia y descarga en 2. En la regresión, después de recuperar distancias, se excluyeron 428 filas. De ellas, 147 carecían solo de área, 279 de área y distancia y 2 solo de descarga. Las categorías son mutuamente excluyentes. Los 6.135 casos del ACP también integraron la regresión, que incorporó 35 casos adicionales.

Correlaciones y componentes principales

Las correlaciones diarias se calcularon mediante el coeficiente de Pearson sobre los pares con datos. Para el ACP se aplicó ln(1 + x) a las siete variables señaladas en la Tabla 2. Se estandarizaron los doce indicadores y se descompuso la matriz de correlaciones sin rotación. Las cargas son los autovectores multiplicados por la raíz de sus autovalores (Jolliffe & Cadima, 2016). La adecuación se examinó con KMO y la prueba de esfericidad (Bartlett, 1950; Kaiser, 1970; Kaiser & Rice, 1974). El criterio de Horn (1965) comparó cada autovalor con el percentil 95 de 1.000 matrices gaussianas independientes del mismo tamaño.

Regresión longitudinal de la espera

Para el proveedor i y la zafra t, la respuesta fue el logaritmo natural de uno más su espera media W, medida en minutos. El modelo por mínimos cuadrados ordinarios (OLS) incluyó nueve covariables y cuatro indicadores de zafra, con 2020 como referencia (Montgomery et al., 2021). Su especificación fue

$$\ln(1 + W_{it}) = \alpha + X_{it}\beta + \delta_t + \varepsilon_{it}$$

El vector X reúne ln(1 + área en ha), distancia en km, carga en t/ticket, ln(1 + descarga en min), puntualidad entre 0 y 1, sacarosa y fibra en puntos porcentuales, ln(1 + materia extraña en %) y ln(1 + toneladas por día activo). Las toneladas por día activo se obtuvieron al dividir

la caña del proveedor por el número de fechas en las que entregó. Se usaron tickets únicos por zafra y proveedor para calcular la carga, conservando las toneladas de los ítems que integran cada ticket.

La puntualidad del ACP y K-means es la proporción de ítems cuya fecha industrial coincide con la fecha de cupo. En la regresión, para cada proveedor y día industrial se calculó la fracción de ítems cuya fecha de documento coincide con la de cupo. Se promediaron esas fracciones diarias ponderándolas por las toneladas del día. Ambas definiciones se verificaron en las cinco zafras.

Se mantuvo la distancia observada cuando estaba disponible. Para las distancias faltantes se enlazó el código de proveedor con sus propiedades y se usó la mediana de las distancias de las propiedades distintas. Así se recuperaron 1.345 de 1.624 valores faltantes. De los recuperados, 1.344 ingresaron en la regresión. Para 1.354 proveedores con ambas fuentes se comparó la mediana de sus distancias observadas con la mediana del maestro. La diferencia absoluta tuvo mediana de 1,37 km y media de 3,57 km.

Se examinaron la heterocedasticidad con la prueba de Breusch y Pagan (1979) y la colinealidad con factores de inflación de varianza. Se estimaron errores estándar HC3 y errores agrupados por proveedor para los mismos coeficientes OLS (Cameron & Miller, 2015; MacKinnon & White, 1985; White, 1980). El modelo de efectos fijos sustituyó el intercepto común por uno específico de cada proveedor y retuvo los indicadores de zafra. Se centraron la respuesta, las covariables y dichos indicadores dentro del proveedor. El área quedó absorbida por ser constante en el tiempo.

Segmentación y entorno de cálculo

Se ajustó K-means sobre once de los doce indicadores del ACP. Se excluyeron los días activos y se mantuvieron las transformaciones y la estandarización. Para cada k de 2 a 6 se empleó el algoritmo de Lloyd, inicialización k-means++, diez inicializaciones y semilla 42 (Arthur & Vassilvitskii, 2007; MacQueen, 1967). Se prefieren valores altos de silueta y Caliński–Harabasz, y bajos de Davies–Bouldin (Caliński & Harabasz, 1974; Davies & Bouldin, 1979; Rousseeuw, 1987). Cada unidad agrupada fue un proveedor en una zafra.

Los resultados de las Tablas 1–4 se calcularon para este artículo a partir de las bases de Empresa A. Se utilizaron Python 3.12.14, pandas 2.2.3, NumPy 2.3.5, SciPy 1.17.0 y scikit-

learn 1.8.0. La regresión y sus covarianzas se calcularon con álgebra matricial. El apéndice detalla las operaciones necesarias para reproducir la inferencia.

3. RESULTADOS

Correlaciones diarias y componentes principales

En 153.160 observaciones proveedor–día, la cantidad de ítems se asoció con el tonelaje entregado ($r = 0{,}817$), y una mayor espera se asoció con menor puntualidad ($r = -0{,}538$). En la matriz anual de 6.135 casos, el KMO global fue 0,547. La prueba de Bartlett rechazó la matriz identidad, $\chi^2(66) = 51.097{,}65$, $p < 0{,}001$. Los cuatro primeros componentes superaron el umbral de Horn y explicaron en conjunto el 63,65 % de la varianza (Tabla 2).

Tabla 2

Cargas, autovalores y varianza del análisis de componentes principales

Indicador	CP1	CP2	CP3	CP4	CP5[2]
Área de referencia (ha)[1]	0,494	−0,332	0,380	−0,149	−0,087
Distancia al ingenio (km)	0,144	0,018	0,289	−0,380	0,688
Días activos (días)[1]	0,878	0,288	−0,068	0,242	0,111
Ítems de boleta (n)[1]	0,925	0,236	−0,047	0,243	0,072
Carga por ítem (t/ítem)	0,269	−0,637	0,204	−0,416	−0,166
Caña entregada (t)[1]	0,972	0,022	0,028	0,090	0,014
Espera media (min)[1]	0,235	0,698	−0,075	−0,442	−0,189
Descarga media (min)[1]	−0,277	0,246	−0,018	−0,075	0,594
Puntualidad (proporción)	−0,034	−0,662	−0,064	0,552	0,241
Sacarosa aparente (%)	−0,214	0,421	0,156	0,205	−0,072
Fibra (%)	−0,398	0,474	0,443	0,352	−0,037

Doi: https://doi.org/10.63688/rgnj0m52

Indicador	CP1	CP2	CP3	CP4	CP5[2]
Materia extraña (%)[1]	0,015	0,024	0,833	0,108	−0,098
Autovalor observado	3,244	2,045	1,201	1,149	0,989
Umbral de Horn (P95)	1,088	1,065	1,050	1,037	1,025
Varianza explicada (%)	27,031	17,038	10,006	9,572	8,242
Varianza acumulada (%)	27,031	44,069	54,075	63,647	71,889

Nota. CP = componente principal. [1] Se aplicó ln(1 + x) antes de estandarizar. [2] Eje descriptivo. El signo de cada eje se orientó hacia cargas positivas de toneladas, espera, materia extraña, puntualidad y distancia, respectivamente. Cálculos a partir del panel anual de Empresa A.

CP1 representa la escala de entrega. Sus mayores cargas corresponden al tonelaje (0,972), al número de ítems (0,925) y a los días activos (0,878). En CP2, una mayor espera (0,698) coincide con menor puntualidad (−0,662) y menor carga por ítem (−0,637). CP3 recoge sobre todo la presencia de materia extraña (0,833), acompañada de fibra (0,443). En CP4, la puntualidad (0,552) se opone a la espera (−0,442), a la carga por ítem (−0,416) y a la distancia (−0,380). CP5 relaciona distancia y descarga. Su autovalor de 0,989 queda por debajo de su umbral de 1,025.

Asociaciones longitudinales con la espera

La regresión incluyó 6.170 observaciones y 1.356 proveedores, de los cuales 1.313 participaron en dos o más zafras. El ajuste OLS fue $R^2 = 0,201$ y R^2 ajustado = 0,200. La prueba de Breusch–Pagan indicó heterocedasticidad (LM = 146,81, $p < 0,001$). Los factores de inflación de varianza oscilaron entre 1,06 y 2,58.

Tabla 3

Coeficientes de la regresión de ln(1 + espera), errores estándar e intervalos de confianza

Mínimos cuadrados ordinarios con errores HC3

Variable	β	EE	IC 95 %inferior	IC 95 %superior	p
Constante	7,35229	0,32249	6,72011	7,98448	< 0,0001
ln(1 + área)	−0,03898	0,00741	−0,05351	−0,02445	< 0,0001
Distancia (km)	0,00039	0,00064	−0,00086	0,00164	0,5414
Carga (t/ticket)	0,00657	0,00135	0,00392	0,00922	< 0,0001
ln(1 + descarga)	−0,03905	0,03119	−0,10020	0,02210	0,2107
Puntualidad (0–1)	−0,03636	0,06730	−0,16828	0,09556	0,5890
Sacarosa (%)	0,04143	0,01271	0,01651	0,06635	0,0011
Fibra (%)	−0,08121	0,01944	−0,11932	−0,04310	< 0,0001
ln(1 + materia extraña)	−0,07143	0,03620	−0,14239	−0,00047	0,0485
ln(1 + t/día activo)	−0,08550	0,01741	−0,11963	−0,05136	< 0,0001
Zafra 2021	0,03732	0,02048	−0,00283	0,07747	0,0685
Zafra 2022	0,14342	0,02136	0,10156	0,18529	< 0,0001
Zafra 2023	−0,14829	0,02542	−0,19813	−0,09845	< 0,0001
Zafra 2024	−0,59982	0,02615	−0,65108	−0,54855	< 0,0001

- Mínimos cuadrados ordinarios con errores agrupados por proveedor

Variable	β	EE	IC 95 %inferior	IC 95 %superior	p
Constante	7,35229	0,33256	6,69990	8,00469	< 0,0001
ln(1 + área)	−0,03898	0,00896	−0,05656	−0,02140	< 0,0001
Distancia (km)	0,00039	0,00076	−0,00109	0,00187	0,6060
Carga (t/ticket)	0,00657	0,00147	0,00370	0,00944	< 0,0001
ln(1 + descarga)	−0,03905	0,03439	−0,10652	0,02842	0,2564
Puntualidad (0–1)	−0,03636	0,06320	−0,16034	0,08762	0,5652
Sacarosa (%)	0,04143	0,01240	0,01709	0,06576	0,00086
Fibra (%)	−0,08121	0,01868	−0,11785	−0,04457	< 0,0001
ln(1 + materia extraña)	−0,07143	0,03546	−0,14099	−0,00187	0,0442

Variable	β	EE	IC 95 %inferior	IC 95 %superior	p
ln(1 + t/día activo)	−0,08550	0,01919	−0,12313	−0,04786	< 0,0001
Zafra 2021	0,03732	0,01905	−0,00006	0,07470	0,0504
Zafra 2022	0,14342	0,02096	0,10231	0,18454	< 0,0001
Zafra 2023	−0,14829	0,02352	−0,19443	−0,10216	< 0,0001
Zafra 2024	−0,59982	0,02587	−0,65057	−0,54907	< 0,0001

Nota. β = coeficiente en la escala ln(1 + minutos de espera). EE = error estándar. IC = intervalo de confianza bilateral. La zafra de referencia es 2020. Los paneles A y B comparten coeficientes. Cálculos a partir del panel anual y las boletas de Empresa A.

Tabla 3 (continuación)

Asociaciones dentro del mismo proveedor

Efectos fijos de proveedor y zafra con errores agrupados por proveedor

Variable	β	EE	IC 95 %inferior	IC 95 %superior	p
ln(1 + área)	Absorbida	—	—	—	—
Distancia (km)	−0,00485	0,00171	−0,00820	−0,00150	0,0046
Carga (t/ticket)	0,00758	0,00257	0,00253	0,01262	0,0033
ln(1 + descarga)	0,04461	0,03366	−0,02142	0,11063	0,1853
Puntualidad (0–1)	−0,22506	0,06724	−0,35696	−0,09316	0,00084
Sacarosa (%)	0,04975	0,01562	0,01910	0,08040	0,0015
Fibra (%)	−0,05788	0,02030	−0,09771	−0,01806	0,0044
ln(1 + materia extraña)	−0,03986	0,03721	−0,11286	0,03313	0,2842
ln(1 + t/día activo)	−0,12213	0,03604	−0,19284	−0,05142	0,00072
Zafra 2021	0,04541	0,01991	0,00635	0,08447	0,0227
Zafra 2022	0,16380	0,02229	0,12006	0,20753	< 0,0001
Zafra 2023	−0,12951	0,02559	−0,17971	−0,07931	< 0,0001
Zafra 2024	−0,55688	0,02872	−0,61323	−0,50053	< 0,0001

Doi: https://doi.org/10.63688/rgnj0m52

Nota. El modelo estima los contrastes de zafra frente a 2020 después de retirar la media de cada proveedor. El área no varía dentro de proveedor. La transformación absorbe los interceptos individuales. Cálculos a partir del panel anual y las boletas de Empresa A.

Área, carga por ticket, sacarosa, fibra, materia extraña y toneladas por día activo se asociaron con la espera al nivel de 0,05 con ambas covarianzas OLS. La descarga presentó intervalos que incluyeron cero en las tres especificaciones. Frente a 2020, el coeficiente OLS fue positivo en 2022 y negativo en 2023 y 2024. El contraste 2021–2020 quedó cerca del umbral convencional con errores agrupados ($p = 0,0504$) y fue significativo dentro de proveedor ($p = 0,0227$).

Distancia y puntualidad se asociaron con menor espera en el modelo de efectos fijos. Para el mismo proveedor, un aumento de 10 puntos porcentuales en puntualidad se asoció con una variación de −2,23 % en 1 + espera (IC 95 % de −3,51 % a −0,93 %), manteniendo constantes las demás covariables. Esta asociación corresponde al indicador de puntualidad basado en la fecha de documento. La materia extraña perdió significancia al incorporar los efectos fijos. Carga por ticket, sacarosa, fibra y toneladas por día activo conservaron el signo de la asociación.

Separación de los conglomerados

La silueta y el índice de Caliński–Harabasz alcanzaron su máximo con dos grupos. Davies–Bouldin favoreció seis. Se eligió $k = 2$ para describir el contraste de escala más sencillo. La silueta de 0,140 indica que numerosos registros tienen perfiles próximos a los de ambos grupos (Tabla 4).

Tabla 4

Métricas del agrupamiento y perfil de los dos conglomerados

Comparación de las cinco soluciones

Número de grupos (k)	Silueta	Caliński–Harabasz	Davies–Bouldin
2	0,140	1.072,4	2,220
3	0,136	934,2	2,091
4	0,105	814,9	2,090
5	0,101	722,6	2,070
6	0,104	665,9	1,984

Medianas de los indicadores en unidades originales

Indicador	Conglomerado 1	Conglomerado 2
Observaciones proveedor–zafra	3.183	2.952
Proporción de la muestra (%)	51,88	48,12
Área de referencia (ha)	47,2	184,1
Distancia al ingenio (km)	25,52	29,83
Ítems de boleta (n)	26	99
Carga por ítem (t/ítem)	14,46	27,09
Caña entregada (t)	413,73	2.325,51
Espera media (min)	494,28	425,79
Descarga media (min)	102,68	91,26
Puntualidad (%)	66,67	72,34
Sacarosa aparente (%)	16,68	16,11
Fibra (%)	16,64	16,00
Materia extraña (%)	1,20	1,20

Nota. n = 6.135 en todas las soluciones. El panel B presenta el recuento y el porcentaje en las dos primeras filas y la mediana en las demás. Se nombró grupo 1 al de menor tonelaje mediano. Cálculos a partir del panel anual de Empresa A.

El primer grupo reunió 3.183 observaciones proveedor–zafra, con una mediana de 414 t entregadas y 494 minutos de espera. El segundo reunió 2.952, con medianas de 2.326 t y 426 minutos. La materia extraña tuvo medianas prácticamente iguales. La variación en escala de entrega y en tiempos de servicio fue más visible que la diferencia en ese indicador de calidad.

4. DISCUSIÓN

El primer componente concentra tres medidas que crecen con la actividad del proveedor. La relación de los ítems con el tonelaje concuerda con la importancia del flujo de transporte en los estudios de Higgins et al. (2004) y López-Milán et al. (2006). Parte de esa relación es contable, porque el tonelaje de un ticket puede distribuirse entre varios ítems. Distinguir ambos denominadores permite conservar el significado de las cargas del ACP y el de la carga física por ticket en la regresión.

El desempeño temporal y la calidad física aportan variación que el tonelaje por sí solo deja fuera. La adecuación KMO limitada aconseja interpretar las cargas como un resumen de las observaciones de Empresa A (Jolliffe & Cadima, 2016). El análisis paralelo aporta un criterio empírico para retener cuatro componentes. El quinto añade varianza descriptiva, aunque permanece por debajo de lo esperable en la comparación aleatoria.

Doi: https://doi.org/10.63688/rgnj0m52

Los efectos fijos permiten distinguir cambios dentro de una misma relación de suministro de las diferencias entre proveedores. El signo negativo de la distancia dentro del proveedor puede coexistir con cambios de propiedad de origen, rutas o condiciones de recepción. Además, una parte de las distancias se recuperó del maestro. Estos registros tampoco describen toda la congestión de cada jornada. El diseño admite una lectura asociativa y deja sin identificar el efecto causal de acortar una ruta o modificar un horario.

La separación débil de los conglomerados concuerda con diferencias de escala que se distribuyen gradualmente. La pertenencia puede cambiar entre zafras. Por ello los dos perfiles sirven para describir la población cubierta y carecen de fundamento como regla automática de prioridad. Las exclusiones por faltantes fueron mayores al inicio de la serie, lo que restringe la comparación de perfiles con otras poblaciones. Una futura aplicación predictiva tendría que reservar zafras posteriores para evaluación temporal y examinar por separado los errores fuera de muestra (Bergmeir & Benítez, 2012; Tashman, 2000). Si se comparan dos pronósticos, la diferencia de pérdidas puede evaluarse con el procedimiento de Diebold y Mariano (1995).

5. CONCLUSIÓN

Cuatro componentes resumieron el 63,65 % de la variación histórica y distinguieron escala de entrega, condiciones de servicio, calidad física y puntualidad vinculada a la accesibilidad. El panel mostró que la distancia y la puntualidad se asociaron con la espera al seguir al mismo proveedor entre zafras. El ajuste de la regresión fue limitado. Dos conglomerados describieron diferencias de volumen y de tiempos de servicio, con amplio solapamiento. La selección de casos completos, la distancia recuperada y la agregación anual delimitan la interpretación y su generalización a otros ingenios.

REFERENCIAS BIBLIOGRÁFICAS

Ahumada, O., & Villalobos, J. R. (2009). Application of planning models in the agri-food supply chain: A review. *European Journal of Operational Research, 196*(1), 1–20. https://doi.org/10.1016/j.ejor.2008.02.014

Arjona, E., Bueno, G., & Salazar, L. (2001). An activity simulation model for the analysis of harvesting and transportation systems of a sugarcane plantation. *Computers and*

Electronics in Agriculture, 32(3), 247–264. https://doi.org/10.1016/S0168-1699(01)00168-5

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In N. Bansal, K. Pruhs, & C. Stein (Eds.), *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027–1035). Society for Industrial and Applied Mathematics. https://research.google/pubs/k-means-the-advantages-of-careful-seeding/

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology, 3*(2), 77–85. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bocanegra-Herrera, C. C., & Vidal-Holguín, C. J. (2016). Development of a simulation model as a decision support system for sugarcane supply. *DYNA, 83*(198), 180–186. https://doi.org/10.15446/dyna.v83n198.52719

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294. https://doi.org/10.2307/1911963

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

da Silva, A. F., Marins, F. A. S., & Dias, E. X. (2015). Addressing uncertainty in sugarcane harvest planning through a revised multi-choice goal programming model. *Applied Mathematical Modelling, 39*(18), 5540–5558. https://doi.org/10.1016/j.apm.2015.01.007

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Higgins, A., Antony, G., Sandell, G., Davies, I., Prestwidge, D., & Andrew, B. (2004). A framework for integrating a complex harvesting and transport system for sugar production. *Agricultural Systems, 82*(2), 99–115. https://doi.org/10.1016/j.agsy.2003.12.004

Higgins, A., & Davies, I. (2005). A simulation model for capacity planning in sugarcane transport. *Computers and Electronics in Agriculture, 47*(2), 85–102. https://doi.org/10.1016/j.compag.2004.10.006

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), Article 20150202. https://doi.org/10.1098/rsta.2015.0202

Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika, 35*(4), 401–415. https://doi.org/10.1007/BF02291817

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement, 34*(1), 111–117. https://doi.org/10.1177/001316447403400115

López-Milán, E., Miquel-Fernández, S., & Plà-Aragonés, L. M. (2006). Sugar cane transportation in Cuba, a case study. *European Journal of Operational Research, 174*(1), 374–386. https://doi.org/10.1016/j.ejor.2005.01.028

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics, 29*(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability: Volume 1. Statistics* (pp. 281–297). University of California Press. https://projecteuclid.org/euclid.bsmsp/1200512992

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). Wiley. https://www.wiley.com/en-gb/shop/general-introductory-statistics/introduction-to-linear-regression-analysis-6th-edition-p-9781119578727

Paiva, R. P. O., & Morabito, R. (2009). An optimization model for the aggregate production planning of a Brazilian sugar and ethanol milling company. *Annals of Operations Research, 169*(1), 117–130. https://doi.org/10.1007/s10479-008-0428-9

Piewthongngam, K., Pathumnakul, S., & Setthanan, K. (2009). Application of crop growth simulation and mathematical modeling to supply chain management in the Thai sugar industry. *Agricultural Systems, 102*(1–3), 58–66. https://doi.org/10.1016/j.agsy.2009.07.002

Pinckert Ortiz, H. C. (2026). *Diseño y evaluación de un sistema dinámico basado en análisis multivariante, pronóstico y optimización matemática para la asignación eficiente de cupos diarios de proveedores en la agroindustria cañera de Santa Cruz, Bolivia* [Manuscrito de tesis doctoral]. Universidad Autónoma Gabriel René Moreno.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Santoro, E., Soler, E. M., & Cherri, A. C. (2017). Route optimization in mechanized sugarcane harvesting. *Computers and Electronics in Agriculture, 141*, 140–146. https://doi.org/10.1016/j.compag.2017.07.013

Singh, G., & Pathak, B. K. (1994). A decision support system for mechanical harvesting and transportation of sugarcane in Thailand. *Computers and Electronics in Agriculture, 11*(2–3), 173–182. https://doi.org/10.1016/0168-1699(94)90006-X

Solomon, S. (2009). Post-harvest deterioration of sugarcane. *Sugar Tech, 11*(2), 109–123. https://doi.org/10.1007/s12355-009-0018-4

Stray, B. J., van Vuuren, J. H., & Bezuidenhout, C. N. (2012). An optimisation-based seasonal sugarcane harvest scheduling decision support system for commercial growers in South Africa. *Computers and Electronics in Agriculture, 83*, 21–31. https://doi.org/10.1016/j.compag.2012.01.009

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting, 16*(4), 437–450. https://doi.org/10.1016/S0169-2070(00)00065-0

Taşkıner, T., & Bilgen, B. (2021). Optimization models for harvest and production planning in agri-food supply chain: A systematic review. *Logistics, 5*(3), Article 52. https://doi.org/10.3390/logistics5030052

Teixeira, E. dos S., Rangel, S., Florentino, H. de O., & de Araujo, S. A. (2023). A review of mathematical optimization models applied to the sugarcane supply chain. *International Transactions in Operational Research, 30*(4), 1755–1788. https://doi.org/10.1111/itor.13056

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817–838. https://doi.org/10.2307/1912934

CONFLICTO DE INTERESES

No existe.

FINANCIAMIENTO

No aplica.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Harold Carlos Pinckert Ortiz.
Metodología: Harold Carlos Pinckert Ortiz.
Investigación: Harold Carlos Pinckert Ortiz.
Redacción – borrador original: Harold Carlos Pinckert Ortiz.
Redacción – revisión y edición: Harold Carlos Pinckert Ortiz.