

Empleo de métodos estadísticos para mejorar la exactitud en la determinación de los componentes de la caña de azúcar

Mauricio Ribas-García*, Ramón Consuegra-del-Rey, Marlen Alfonso-Alfonso, Susana Ravelo-Bravo

Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar (Icidca).

Via Banca 804 y Carretera Central. San Miguel del Padrón, La Habana, Cuba.

* rmauricio.ribas@icidca.azcuba.cu

RESUMEN

En el presente trabajo se evaluaron las tablas para la determinación de fibra y pol en caña, obtenida por la aplicación del método de la prensa hidráulica, descrito en el MACU (Manual Métodos Analíticos para Azúcar Crudo). Se analizaron muestras de caña correspondientes al 45 % de las variedades actualmente comercializadas. La evaluación consistió en comparar los resultados obtenidos por el método de la prensa hidráulica con el método de desintegración húmeda que es considerado un método de referencia.

Como resultado se obtuvo que lo contenido en el MACU aporta una exactitud aceptable, pero que es mejorable. El análisis estadístico de los experimentos dio lugar a nuevos modelos matemáticos que permiten determinar las variables con mayor exactitud.

Palabras clave: desintegrador húmedo, método prensa hidráulica, caña de azúcar, análisis de regresión.

ABSTRACT

In the present work the tables for the determination of fiber and pol in cane, obtained by the application of the hydraulic press method, described in the MACU (Manual Analytical Methods for Crude Sugar) were evaluated.

Cane samples corresponding to 45 % of commercial Cuban cane varieties were analyzed. The evaluation consisted of comparing the results obtained by the hydraulic press method with the wet disintegrator method which is considered a reference method.

As a result, it was obtained that MACU tables' accuracy is acceptable, however it could be better. New mathematical models derived from experiments statistical analysis permit a more accurate determination of fiber and pol.

Key words: wet disintegrator, hydraulic press method, sugar cane, regression analysis.

INTRODUCCIÓN

En las fábricas de azúcar cubanas se usa el método de la prensa hidráulica para la determinación de fibra y pol en caña. El manual "Métodos analíticos de azúcar crudo" (1), conocido por las siglas MACU, contiene tablas derivadas de ecuaciones en las que la fibra es función del peso de la torta derivada del prensado, mientras que la pol en caña lo es de la fibra y de la pol del jugo extraído por la prensa. Las tablas del MACU datan de los años 80 en los que la composición varietal de la caña es diferente a la actual, por lo que especialistas del Icidca (2) realizaron experimentos para valorarlas y mejorarlas (en caso

necesario) en el actual contexto de variedades de caña tomando como técnica de referencia el método de desintegración húmeda (DH).

En el presente trabajo se realiza un tratamiento estadístico de los resultados, dividido en tres etapas, que permitió evaluar la exactitud del método prensa actual implementado en el MACU (1) comparado con el método de referencia de Desintegración Húmeda (DH) (3-5), se correlacionaron los componentes de la caña determinados por DH con el peso de la torta húmeda del método prensa, teniendo en cuenta diferentes variedades y edades de la caña sembradas actualmente en Cuba. Por último, las nuevas expresiones obtenidas fueron sometidas a

un proceso de validación que demostró una mejor exactitud del método prensa, con las nuevas expresiones incorporadas, con respecto al método prensa que aparece descrito actualmente en el MACU.

MATERIALES Y MÉTODOS

Se tomaron 89 muestras de seis variedades de caña planta, que constituyen cerca del 45 % del total de caña sembrada en el país en la actualidad. Además, se evaluaron muestras de caña industrial del central “Jesús Rabí”.

Cada muestra fue sometida a un proceso de desfibrado, homogeneizado y se dividieron en dos porciones de ensayo. Una porción de ensayo de la muestra homogénea se sometió al proceso de prensado y a partir de la medición del peso de la torta húmeda se determinó el contenido de fibra por el método de la prensa hidráulica descrito en el MACU. La otra porción de ensayo fue utilizada para determinar el contenido de fibra mediante el método de desintegración húmeda (DH).

Para evaluar la proximidad entre los valores de fibra en caña reportados por el método prensa descrito en el MACU y por el método de desintegración húmeda se calculó el error cuadrático medio (ECM), la desviación estándar residual (RSD) y la desv. estándar residual relativa según las ecuaciones 1, 2 y 3, respectivamente.

$$ECM = \frac{\sum_{i=1}^n (F_i^{MACU} - F_i^{DH})^2}{n-1} \quad (1)$$

$$RSD = \sqrt{ECM} \quad (2)$$

$$RSD = \frac{RSD}{\text{Valor medio DH}} \cdot 100 \quad (3)$$

DESARROLLO

La tabla 1 contiene una comparación de la fibra en caña determinada mediante el método prensa descrito en el MACU con respecto al método de referencia que es el de desintegración húmeda (DH).

El resultado de la comparación dio lugar a una desviación estándar residual $RSD=0,28$ lo que indica que los valores de fibra en caña obtenidos por el método prensa descrito en el MACU se diferencian como promedio en 0,28 unidades respecto a la fibra en caña determinada por el DH. No obstante, hay variedades de caña donde esta dispersión puede llegar a ser de hasta 0,44 unidades.

La desviación estándar residual relativa $RSD \% = 1,87 \%$, se considera aceptable. Sin embargo, los valores de fibra en caña determinados por DH siempre son mayores que los obtenidos por el método prensa descrito en el MACU independientemente de la variedad de caña (tabla 1).

Teniendo en cuenta lo anterior se puede concluir que las determinaciones de la fibra en caña obtenidas mediante el método prensa no se alejan mucho de la realidad, pero es posible obtener valores más exactos de este componente si se correlaciona la fibra en caña obtenida por DH con el peso de la torta húmeda medida por el método prensa descrito en el MACU.

Modelo de regresión para determinar fibra en caña

A continuación, se decidió aplicar la técnica de análisis de regresión con el objetivo de obtener un modelo con una buena capacidad de predicción que permita estimar los valores de fibra en caña determinados por el DH, a partir del peso de la torta húmeda obtenida por el método prensa.

Se realizó un análisis de residuos para identificar las observaciones más alejadas del modelo. Esto se logra con el cálculo de los Residuos Estudentizados para cada valor predicho por el modelo y considerando como observaciones más alejadas del modelo aquellas que tienen valores de Residuos Estudentizados mayores que 2,5. También fue necesario identificar aquellas observaciones que tienen una alta influencia sobre los parámetros del modelo. Son observaciones que si se eliminan cambian los valores de los parámetros del modelo porque tienen una influencia mucho mayor que la influencia de un

Tabla 1. Comparación de los valores de fibra en caña obtenidos por el MACU y por DH

Variedad y edad	No. de muestras	Fibra DH (%)	Fibra MACU (%)	RSD	RSD%
C86-12 (15 meses)	26	14,23	14,00	0,27	1,90
Co 997 (12 meses)	10	17,00	16,75	0,44	2,59
C323-68 (17 meses)	10	14,76	14,64	0,26	1,76
C86-56 (17 meses)	5	14,55	14,24	0,35	2,41
Pruebas Rabí	14	14,96	14,76	0,28	1,87
C92-325 (12 meses)	10	15,35	15,18	0,24	1,56
C1051-73 (12 meses)	10	15,50	15,29	0,23	1,48
Contenido disperso	5	13,77	13,59	0,24	1,74
Total	89	14,97	14,76	0,28	1,87

punto promedio. Las tablas 2 y 3 muestran los resultados de estos análisis. Se encontraron 5 Residuos Estudentizados mayores que 2,5, lo que es considerado suficientemente grande, y se decidió remover estas 5 observaciones del modelo (tabla 2). Hay 2 observaciones con más de 5 veces el valor de influencia promedio (tabla 3).

Tabla 2. Residuos atípicos

Fila	Y	Y predicha	Residuo	Residuo Estuden.
33	17,30	16,52	0,548	3,07
36	17,33	16,68	0,654	3,75
44	14,80	15,18	-0,381	-2,68
45	14,90	15,29	-0,392	-2,54
52	13,28	13,71	-0,431	-2,85

Tabla 3. Puntos influyentes

Fila	Influencia	Distancia de Mahalanobis	DFITS
54	0,116	10,419	-0,487
87	0,1168	10,419	-0,4876

Influencia media de un solo punto = 0,022

El modelo ajustado con los 4 puntos anteriores excluidos muestra que el término correspondiente al peso de la torta (PT) es estadísticamente significativo con un nivel de confianza del 95,0% (tabla 4).

El modelo obtenido responde a la siguiente ecuación:

$$Fibra_{caña} = 2,22502 + 0,0850224 \cdot Peso_{torta} \quad (4)$$

Puesto que el valor-P en la tabla ANOVA es menor que 0,05 (tabla 5), existe una relación estadísticamente significativa entre las variables con un

Tabla 4. Resultados del modelo para la fibra

Parámetro	Estimación	Error estándar	Estadístico T	Valor-P
Constante	2,225	0,216	10,284	0,000
PT	0,0850	0,0014	59,295	0,000

Tabla 5. Análisis de Varianza

Fuente	Suma de cuadrados	GI	Cuadrado medio	Razón-F	Valor-P
Modelo	67,438	1	67,438	3515,90	0,000
Residuo	1,515	79	0,0192		
Total (Corr.)	68,953	80			

R-cuadrada = 97,80 por ciento

R-cuadrado (ajustado para g.l.) = 97,77 por ciento

Error estándar del est. = 0,138

Estadístico Durbin-Watson = 2,20903 (P=0,807)

Autocorrelación de residuos en retraso 1 = -0,107

nivel de confianza del 95,0 %. El modelo es capaz de explicar más de 97,8 % de la variabilidad de los datos de fibra en caña con un error estándar de las estimaciones igual a 0,14. Esto significa que en promedio los valores de fibra en caña determinados por el modelo se diferencian en 0,14 unidades de los valores observados utilizando el método de referencia de DH.

La prueba de Durbin-Watson (DW) para determinar si hay alguna correlación significativa basada en el orden en el que se presentan los datos dio no significativa (valor-P > 0,05). Por lo tanto, no hay indicación de una auto correlación serial en los residuos con un nivel de confianza del 95,0 %.

Los gráficos de los residuos del modelo muestran que todos sus valores se distribuyen aleatoriamente sobre la línea horizontal por lo que no hay necesidad de incorporarle otros términos al modelo. Los residuos no revelan presencia de heteroscedasticidad en los datos, o sea, la variabilidad de los datos cambia conforme la media cambia. En este caso la variabilidad de los datos es la misma para todos los valores predichos (figura 1).

Se puede concluir que, mediante el uso de la ecuación (4) se mejora la exactitud de la determinación de fibra en caña por el método prensa, ya que disminuye la desviación estándar residual (RSD) de 0,28 (tabla 1) a 0,14.

Determinación de la pol en caña

La determinación de la pol en caña responde a expresiones matemáticas en las que la variable es función de la pol del jugo derivado del prensado y de la fibra en caña, la última dependiente del peso de la torta.

Se realizó un estudio con 40 muestras de diferentes variedades de caña que se comercializan actualmente en Cuba para relacionar la razón:

$$N = \frac{Pol_{caña}}{Pol_{jp}}$$

con la fibra haciendo uso del método de DH, también se usa Pol_{jp} para jugo de prensa.

El gráfico de la figura 3 muestra una relación lineal inversa entre ambas variables.

El análisis de residuos permitió identificar a 5 valores atípicos y no se observaron puntos con alta influencia en las estimaciones de los parámetros del modelo (tabla 6).

El modelo ajustado muestra que el término correspondiente a la fibra en caña determinada por DH es estadísticamente significativo con un nivel de confianza del 95,0 % (tabla 7).

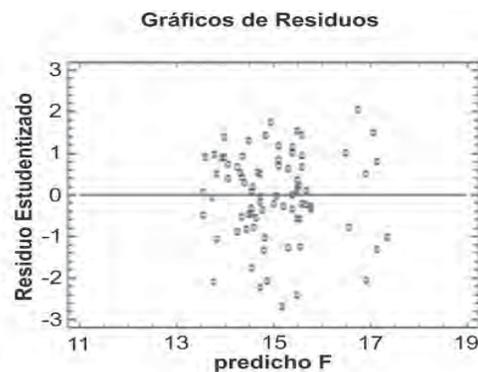
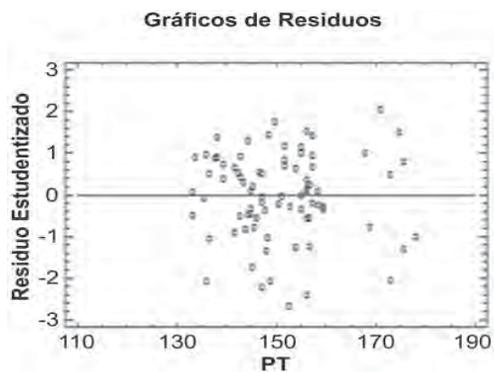


Figura 1. Gráficos de residuos.

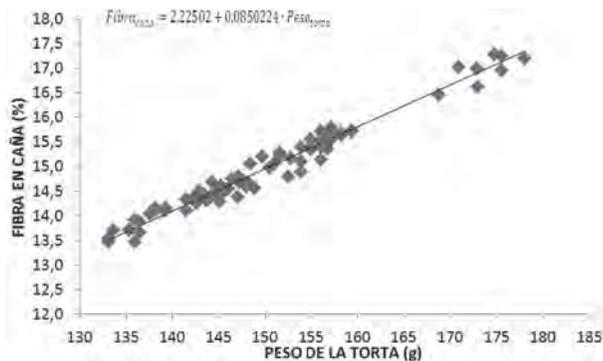


Figura 2. Modelo de regresión ajustado.

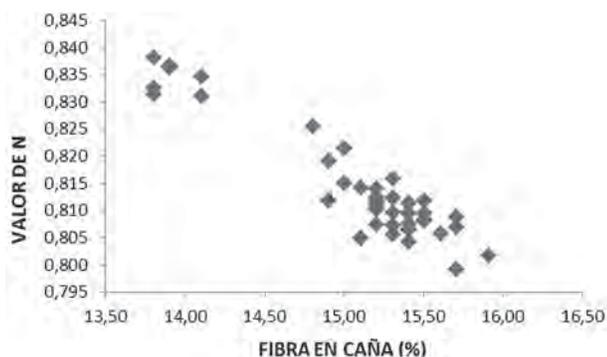


Figura 3. Relación lineal inversa.

Tabla 6. Residuos atípicos

Fila	Y	Y predicha	Residuo	Residuo Estuden.
41	0,800	0,812	-0,0122	-3,04
16	0,815	0,806	0,0095	2,53
37	0,825	0,816	0,0090	2,53
22	0,805	0,814	-0,0088	-2,63
34	0,798	0,805	-0,0072	-2,29

Tabla 7. Resultados del modelo para N

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
Constante	1,067	0,0130	81,524	0,000
Fibra	-0,0167	0,0009	-19,314	0,000

Puesto que el valor-P en la tabla ANOVA es menor que 0,05, existe una relación estadísticamente significativa entre las variables con un nivel de confianza del 95,0 %. El modelo es capaz de explicar el 90,5 % de la variabilidad de los datos de fibra en caña con un error estándar de las estimaciones igual a 0,003. Esto significa que en promedio los valores de la razón N predichos por el modelo se diferencian en 0,003 unidades de los valores observados.

La prueba de Durbin-Watson (DW) para determinar si hay alguna correlación significativa basada en el orden en el que se presentan los datos dio no significativa (valor-P > 0,05), no hay indicación de una auto correlación serial en los residuos con un nivel de confianza del 95,0 %.

Los gráficos de los residuos del modelo muestran que todos sus valores se distribuyen aleatoriamente sobre la línea horizontal por lo que no hay necesidad de incorporarle otros términos al modelo. Los residuos no revelan falta de homogeneidad en la varianza de los datos (figura 4). Por lo tanto, el modelo propuesto es igualmente preciso en todo el rango de variación de las variables.

El modelo queda definido mediante la siguiente ecuación:

$$N = 1,06708 - 0,0167386 \cdot Fibra_{caña} \quad (5)$$

Luego la $Pol_{caña}$ puede ser determinada a través de:

$$Pol_{caña} = N \cdot Pol_{jp} \quad (6)$$

Las ecuaciones 4, 5 y 6 ofrecen un nuevo modelo que permite determinar la $Pol_{caña}$ a partir de la medición del peso de la torta húmeda obtenida por el método prensa.

Validación de los resultados del modelo

El modelo dado por las ecuaciones 4, 5 y 6 fue sometido a una prueba de validación en el contexto de 3 variedades de caña y se ratificó su notable exactitud y superioridad

Tabla 9. Valores promedios de Pol en caña obtenidos por diferentes métodos

Variedad y edad	Muestras	$Pol_{Caña}^{DH}$ (%)	$Pol_{Caña}^{MACU}$ (%)	$Pol_{Caña}^{NF}$ (%)	RSD_{MACU}^{DH} (%)	RSD_{NF}^{DH} (%)
CP-5243. 11 meses	20	13,39	13,62	13,43	0,24	0,05
C90-469. 10 meses	14	14,31	14,52	14,32	0,23	0,03
Caña sonda Central Rabí	12	11,89	12,05	11,88	0,17	0,03

respecto a las ecuaciones del MACU. Se cumple que para las 3 variedades analizadas los valores de desviación estándar residual (RSD_{NF}^{DH})

entre el DH y la nueva variante propuesta (tabla 5) es menor que cuando se comparan las expresiones actuales del MACU con el DH (RSD_{MACU}^{DH})

CONCLUSIONES

1. Se evaluó la exactitud de las determinaciones de

los componentes de la caña por el método prensa y se concluyó que podría ser mejorada.

2. La metodología estadística utilizada permitió obtener nuevas expresiones para la determinación de los componentes de la caña (fibra y pol) por el método prensa que resultan más exactas que las actualmente contenidas en el MACU, luego de compararse ambas variantes con los resultados de la aplicación del método de desintegración húmeda usado como referencia.

REFERENCIAS BIBLIOGRÁFICAS

1. Sanfiel, F., Fernández, F. (2006). Métodos Analíticos para Azúcar Crudo. Publicaciones Azucareras. Ciudad de La Habana. Cuba.
 2. Alfonso, M.; *et al.* (2016). Informe de etapa 2 del proyecto FONCI/ ICIDCA Evaluación de la recuperación de azúcar en el proceso y pago de la caña mediante estándares de alta eficiencia en el cálculo de "RPC meta. 10 páginas.
 3. Wright, P. G. (2005). Process Benchmarking in Cane Sugar Factories. Proc. Aust. Soc. Sugar Cane Technol., 27: 437-453.
 4. Sanallah, (2013). An Alternative Method for Determination of Fibre % Cane. Proc. Aust. Soc. Sugar Cane Technol., 28 1-5.
 5. Rein, P. (2007). Cane Sugar Engineering. Verlag Dr. Albert Bartens, Berlin, 768.
-