

INFORME DEL ANÁLISIS DE LOS DETERMINANTES DE LA PRODUCTIVIDAD DE SOYA 2023

PIDARA

Proyecto Integral de Diversificación
Agroproductiva y Reconversión Agrícola



Ministerio de
Agricultura y Ganadería

Informe del análisis de los determinantes de la productividad de soya 2023

Proyecto Integral de Diversificación Agroproductiva y Reconversión Agrícola-PIDARA

Quito – Ecuador

2024

RESUMEN

El rendimiento nacional de soya al 12 % de humedad y 1 % de impurezas, para 2023, fue de 2.25 t/ha. La producción de la oleaginosa se concentró en las provincias de Guayas y Los Ríos, las cuales mostraron rendimientos de 2.38 t/ha y 2.16 t/ha, respectivamente. Asimismo, se estableció mediante la interpretación de imágenes satelitales, una superficie sembrada nacional de 7,696 hectáreas entre la época lluviosa y la época seca, esta última comprendida entre junio y diciembre. En el presente informe, a partir de las variables obtenidas se plantea una modelización econométrica de carácter explicativo del comportamiento de las principales variables en función del rendimiento. Los factores limitantes en el rendimiento obtenidos fueron el peso de la muestra, el porcentaje de humedad y las variedades sembradas. Dado que el peso de la muestra y el porcentaje de humedad son parte de la fórmula de rendimiento neto, se aplicaron tratamientos diferenciados, los cuales se plasman en dos modelos bifactoriales con interacción, los cuales explican el comportamiento de los rendimientos de Soya para el año 2023.

Palabras clave: rendimiento, humedad, superficie, modelo econométrico, variedades.

ABSTRACT

The national soybean yield at 12 % moisture and 1 % impurities for 2023 was 2.25 t/ha. Oilseed production was concentrated in the provinces of Guayas and Los Ríos, which showed yields of 2.38 t/ha and 2.16 t/ha, respectively. Likewise, the interpretation of satellite images established a national planted area of 7,696 hectares between the rainy and dry seasons, where the latter, between June and December. In this report, based on the variables obtained, an econometric modeling is proposed to explain the behavior of the main variables based on yield performance. The limiting factors in the yield were: the weight of the sample, the percentage of humidity and the soybean varieties planted. Given that the weight of the sample and the percentage of humidity are part of the net yield formula, both were differentiated in two treatments, whose were reflected in two bifactor models with interaction, which explain the behavior of Soybean yields for the year 2023.

Keywords: yield, humidity, planted area, econometric modeling, varieties.

Contenido

1. ANTECEDENTES	5
1.1 Operativos de rendimientos objetivos	5
1.2 Justificación o el objetivo de determinantes de la productividad	5
2. METODOLOGÍA	5
2.1 Determinantes de la productividad	5
2.2 Selección de Características.....	5
2.2.1 Variable de Respuesta.....	5
2.2.2 Factores influyentes en el rendimiento.....	6
2.3 Remoción de atípicos.....	6
2.4 Transformación de variables	6
2.5 Planteamiento de la modelización	6
3. RESULTADOS	7
3.1 Análisis de varianza del primer modelo	7
3.1.1 Cuadro de Análisis de varianza del primer modelo	7
3.1.2 Supuestos del primer modelo.....	9
3.2 Análisis de varianza del segundo modelo	10
3.2.1 Cuadro de análisis de varianza del segundo modelo	10
3.2.2 Supuestos del segundo modelo	11
4. BIBLIOGRAFÍA	13

1. ANTECEDENTES

1.1 Operativos de rendimientos objetivos

La producción de soya predomina en la época seca, por lo que el levantamiento de información en territorio se realiza en el mes de septiembre, en las provincias de Guayas y Los Ríos, en las que se concentra casi la totalidad de la superficie sembrada, a nivel nacional. El proceso inicia con el monitoreo permanente de las siembras y cosechas del cultivo a través de la supervisión del personal técnico de territorio. Para determinar el tamaño de la muestra se implementa el método de Muestreo Aleatorio Simple Estratificado, en el que se incluyen parámetros de nivel de confianza y términos de error, con la finalidad de obtener información representativa de la productividad a nivel nacional y desagregado a nivel provincial.

Desde el apartado 2.5.2, en la Gestión de Rendimientos Objetivos, dentro de sus atribuciones y responsabilidades, se encuentra el cumplimiento de documentos de planificación, gestión y calendario para la generación de información de rendimientos agropecuarios, así como un informe consolidado de información de rendimientos levantados, a nivel territorial.

1.2 Justificación o el objetivo de determinantes de la productividad

La consideración de variables que se encuentran asociadas a la productividad en el cultivo de soya, se los puede delimitar tanto en los que se miden a nivel de campo, que fundamentalmente son: variedad, tipo de semilla y todos los componentes de condiciones medioambientales, así como una textura franco-arenosa. A nivel de laboratorio, para la estimación de los rendimientos netos, existen dos variables limitantes como son: el porcentaje de humedad y el peso total de la muestra, dado que pueden inflar o depreciar el rendimiento.

2. METODOLOGÍA

2.1 Determinantes de la productividad

Las presentes variables se denotan como los principales influyentes en la productividad, dado el levantamiento de información y sus análisis correlacionales entre las variables.

Variedad

Tipo de semilla

Peso de la muestra

Porcentaje (%) de humedad

2.2 Selección de Características

2.2.1 Variable de Respuesta

Rendimiento Neto: es el rendimiento en toneladas métricas por hectárea, considerando el porcentaje de humedad e impurezas.

2.2.2 Factores influyentes en el rendimiento

Peso de la muestra: comprende la cantidad cosechada de una unidad experimental que, en este caso comprende a una planta, por lo que se define como una submuestra, dado el número de plantas que se seleccionan en la fase de campo del operativo de rendimientos objetivos.

Variación: dada la naturaleza de la Soya, en el operativo de rendimientos objetivos, se encontró que las variedades mayoritarias son las desarrolladas por el INIAP (303, 304 y 310), SO ITAV (Híbrido), P34 (Panorama). Por otro lado, se encuentra un grupo mayoritario al que pertenece la otra variedad o la persona productora no conoce de la misma.

2.3 Remoción de atípicos

Para la variable de respuesta se han corrido los test del rango intercuantil y el test de grubbs para delimitar los atípicos a nivel provincial; dónde para la provincia de Los Ríos se definieron cinco atípicos, mientras que, en Guayas, pese a que el Histograma delimita una tendencia de rendimiento hacia el límite superior, no se definen atípicos, lo cual permite definir al peso total de la muestra y a la humedad como limitantes del rendimiento.

2.4 Transformación de variables

Dado que el porcentaje de humedad y el peso de la muestra son variables numéricas, se procede a una transformación a variables categóricas nominales, considerando que en el modelo deben ser parametrizables en función al rendimiento y criterios de delimitación para que corresponda con la ordenada, de tal forma las dos variables transformadas permitirán el entrenamiento de los modelos, definición de hipótesis de respuesta.

El campo provincia es una variable cualitativa, pero el entrenamiento del modelo aditivo; debe ser una variable categórica ordinal, ya que cumple con la función de bloque, en cuanto a los factores de: variedad, peso de la muestra y porcentaje de humedad.

2.5 Planteamiento de la modelización

La variación del rendimiento con respecto a las variedades (Conocidas; INIAP 303, 304, 310, P34-Panorama y SO ITAV (Híbrida) versus las variedades que la persona productora no conoce o seleccionó Otra, diferente. Los porcentajes de humedad y peso de la muestra se tratarán en dos modelos diferentes con la siguiente estructura:

Modelo 1: Se usa como factor limitante al peso de la muestra:

$$Y_{ijk} = u + B_i + V_j + e(V)_{ij} + P_k + VP_{jk} + e(P)_{jk} \text{ Dónde;}$$

Y_{ijk} = Rendimiento neto

u = Media general

B_i = Efecto del i -ésimo bloque

V_j = Efecto del j -ésima variedad

$e(V)_{ij}$ = Error para la variedad

$P_k =$ Efecto del k -ésimo peso de la muestra

$VP_{jk} =$ Efecto de la interacción Variedad \times Peso de la muestra

$e(P)_{jk} =$ Error para el peso de la muestra

Modelo 2: Se usa como factor limitante al porcentaje de humedad de la muestra:

$$Y_{ijk} = u + B_i + V_j + e(V)_{ij} + H_k + VH_{jk} + e(H)_{jk} \text{ Dónde;}$$

$Y_{ijk} =$ Rendimiento neto

$u =$ Media general

$B_i =$ Efecto del i -ésimo bloque

$V_j =$ Efecto del j -ésima variedad

$e(V)_{ij} =$ Error para la variedad

$P_k =$ Efecto del k -ésimo porcentaje de humedad

$VP_{jk} =$ Efecto de la interacción Variedad \times Porcentaje de humedad

$e(P)_{jk} =$ Error para el porcentaje de humedad

El planteamiento de ambos modelos se da, debido a que el peso de la muestra en el primer modelo permitirá analizar la tendencia de incremento del rendimiento con respecto a los intervalos superiores del peso de la muestra. Mientras tanto, el segundo modelo permitirá contrarrestar los porcentajes de humedad mayoritarios que se encuentran en los intervalos inferiores de la tabla de frecuencia, para verificar si esta variable presenta diferencias significativas entre los grupos de datos formados a nivel provincial. El factor que se adhiere junto con el peso de la muestra y el porcentaje de humedad; es la variedad.

3. RESULTADOS

Puesto que, los modelos comprenden un análisis bifactorial bloqueado en relación con el rendimiento; se establecieron criterios de definición, tanto para el primer modelo, donde se pretende evaluar si los pesos de la muestra de los tres intervalos superiores obtenidos a través de una tabla de frecuencia, inflan el rendimiento de soya. Mientras tanto, en el segundo modelo se quiere evaluar los porcentajes de humedad de los dos intervalos inferiores, dado el comportamiento de los porcentajes de humedad. En los modelos descritos se incluyó a las variedades, dado que para los siguientes operativos se ampliaron las variedades a muestrearse.

3.1 Análisis de varianza del primer modelo

3.1.1 Cuadro de Análisis de varianza del primer modelo

A partir del análisis de varianza, se verifica que los pesos de los tres intervalos superiores de la tabla de frecuencias, presentan diferencias significativas, dado que se

utilizaron para parametrizar los diferenciales de rendimiento al alza con la inclusión del factor peso de la muestra y en efecto, según una prueba de comparación de medias de Tukey, son superiores a los rendimientos provinciales (Guayas 2.38 t/ha y Los Ríos 2.16 t/ha) y al rendimiento ponderado nacional 2.25 t/ha), lo que se verifica en la Tabla 1.

Tabla 1. Análisis de varianza de los componentes del primer modelo, en función del rendimiento

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
Rendimiento	36	0,92	0,89	5,45

Cuadro de Análisis de la Varianza (SC tipo III)

	F.V.	SC	gl	CM	F	p-valor	(Error)
Modelo		6,31	9	0,70	32,73	<0,0001	
Variedad		4,3E-04	1	4,3E-04	0,20	0,7330	(Variedad*Bloque)
Peso de la muestra		4,93	2	2,47	115,23	<0,0001	
Bloque		1,27	1	1,27	59,18	<0,0001	
Variedad*Peso de la muestr..		0,01	2	3,4E-03	0,16	0,8550	
Variedad*Bloque		2,1E-03	1	2,1E-03	0,10	0,7544	
Peso de la muestra*Bloque		0,10	2	0,05	2,24	0,1265	
Error		0,56	26	0,02			
Total		6,86	35				

Tabla 2. Prueba de comparación de medias de Tukey para los factores de peso de muestra y variedad en función del rendimiento

```

Test:Tukey Alfa=0,05 DMS=0,00000
Error: 0,0021 gl: 1
Variedad Medias n E.E.
Conocidos 2,69 18 0,01 A
OTRO 2,68 18 0,01 B
Medias con una letra común no son significativamente diferentes (p > 0,05)
  
```

```

Test:Tukey Alfa=0,05 DMS=0,14842
Error: 0,0214 gl: 26
Peso de la muestra Medias n E.E.
3 3,15 12 0,04 A
2 2,66 12 0,04 B
1 2,25 12 0,04 C
Medias con una letra común no son significativamente diferentes (p > 0,05)
  
```

3.1.2 Supuestos del primer modelo

3.1.2.1 Normalidad

Para comprobar el supuesto se corrió, a nivel gráfico, el test de Q-Q plot de los cuantiles observados versus los cuantiles de la normal, así como el test de Shapiro – Wilks, que evalúa los residuos de la variable rendimiento ($p = 0.5657$), los cuales fueron superiores a $p = 0.05$, lo que permite aceptar la hipótesis nula, que los datos presentan una distribución normal.

Gráfico 1: Test de Q-Q plot de normalidad

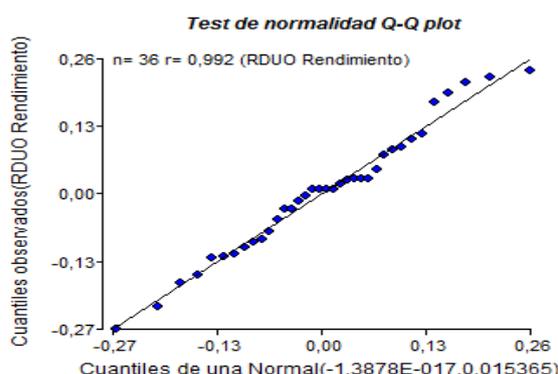


Tabla 3. Prueba de Shapiro-Wilks

Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
RDUO Rendimiento	36	0,00	0,12	0,96	0,5657

3.1.2.2 Homocedasticidad

En el Gráfico 2, las varianzas homogéneas se denotan, dada la nube de puntos distribuida de forma aleatoria. Además, para comprobar la validez del supuesto, se corrió una prueba de Levene para los dos factores relacionados con el rendimiento, donde el peso de la muestra y la variedad presentan probabilidades de 0.1026 y 0.6755, mayores a 0.05, lo que establece que las varianzas son homogéneas, aceptando la hipótesis nula.

Gráfico 2: Prueba de Homocedasticidad de residuos vs los predichos

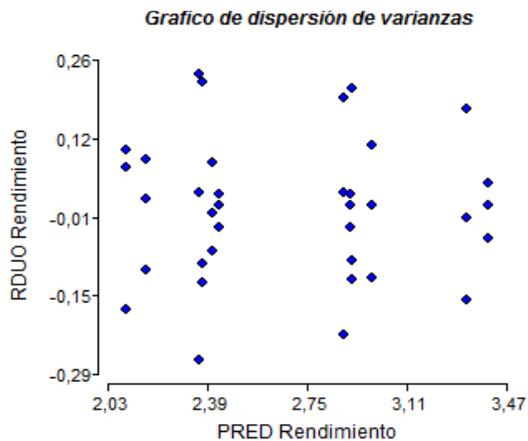


Tabla 4: Prueba de Levene para los factores de variedad y peso de la muestra

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
RABS Rendimiento	36	0,01	0,00	82,84

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	1,1E-03	1	1,1E-03	0,18	0,6755
Variedad	1,1E-03	1	1,1E-03	0,18	0,6755
Error	0,21	34	0,01		
Total	0,21	35			

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
RABS Rendimiento	36	0,13	0,08	78,69

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	0,03	2	0,01	2,44	0,1026
Peso de la muestra	0,03	2	0,01	2,44	0,1026
Error	0,18	33	0,01		
Total	0,21	35			

3.2 Análisis de varianza del segundo modelo

3.2.1 Cuadro de análisis de varianza del segundo modelo

A partir del análisis de varianza, se determina que los porcentajes de humedad de los dos intervalos inferiores de la tabla de frecuencias, dónde se halla la mayor representatividad de la data, presentan diferencias significativas, tanto el porcentaje de humedad, como la interacción entre el porcentaje de humedad y la provincia.

Tabla 5: Análisis de varianza de los componentes del segundo modelo, en función del rendimiento.

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
Rendimiento	24	0,64	0,48	27,84

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	10,01	7	1,43	4,07	0,0096
Variedad	0,02	1	0,02	0,06	0,8033
Humedad	3,07	1	3,07	8,74	0,0093
Bloque	1,72	1	1,72	4,88	0,0421
Variedad*Humedad	0,08	1	0,08	0,23	0,6367
Variedad*Bloque	1,39	1	1,39	3,94	0,0646
Humedad*Bloque	3,68	1	3,68	10,47	0,0052
Variedad*Humedad*Bloque	0,05	1	0,05	0,14	0,7141
Error	5,63	16	0,35		
Total	15,64	23			

Test:Tukey Alfa=0,05 DMS=0,51334

Error: 0,3518 gl: 16

Variedad Medias n E.E.

OTRO 2,16 12 0,17 A

Conocidos 2,10 12 0,17 A

Medias con una letra común no son significativamente diferentes (p > 0,05)

Test:Tukey Alfa=0,05 DMS=0,51334

Error: 0,3518 gl: 16

Humedad Medias n E.E.

1 2,49 12 0,17 A

2 1,77 12 0,17 B

Medias con una letra común no son significativamente diferentes (p > 0,05)

Test:Tukey Alfa=0,05 DMS=0,97976

Error: 0,3518 gl: 16

Variedad Humedad Medias n E.E.

OTRO 1 2,58 6 0,24 A

Conocidos 1 2,40 6 0,24 A

Conocidos 2 1,80 6 0,24 A

OTRO 2 1,75 6 0,24 A

Medias con una letra común no son significativamente diferentes (p > 0,05)

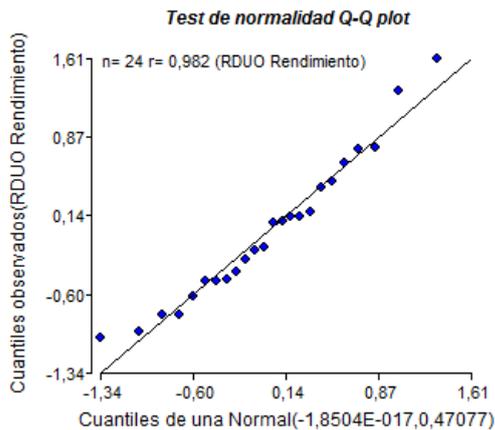
3.2.2 Supuestos del segundo modelo

3.2.2.1 Normalidad

Tal como se procedió con el primer modelo; se corrió el test de Q-Q plot de los cuantiles observados versus los cuantiles de la normal, para el análisis gráfico; así como el test de Shapiro – Wilks, que evalúa los residuos de la variable rendimiento ($p = 0.4039$), los cuales fueron superiores a $p = 0.05$, lo que permite aceptar la hipótesis nula, que los datos presentan una distribución normal.

Gráfico 3: Test de Q-Q plot de normalidad

Tabla 6: Prueba de Shapiro-Wilks



Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
RDUO Rendimiento	24	0,00	0,69	0,94	0,4039

3.2.1.2 Homocedasticidad

Las varianzas homogéneas se validan mediante la prueba de Levene para los dos factores relacionados con el rendimiento, donde el porcentaje de humedad y la variedad presentan probabilidades de 0.7266 y 0.5605, respectivamente, los cuales son mayores a 0.05, lo que establece que las varianzas son homogéneas, aceptando la hipótesis nula.

Gráfico 4: Prueba de Homocedasticidad de residuos vs los predichos

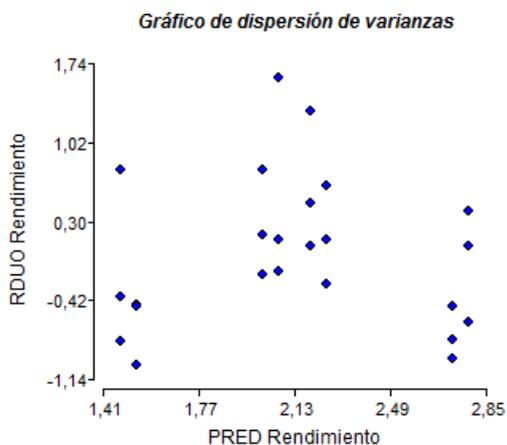


Tabla 7: Prueba de Levene para los factores de variedad y porcentaje de humedad

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
RABS Rendimiento	24	0,02	0,00	74,16

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	0,06	1	0,06	0,35	0,5605
Variedad	0,06	1	0,06	0,35	0,5605
Error	3,61	22	0,16		
Total	3,67	23			

Análisis de la varianza

Variable	N	R ²	R ² Aj	CV
RABS Rendimiento	24	0,01	0,00	74,53

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	0,02	1	0,02	0,13	0,7266
Humedad	0,02	1	0,02	0,13	0,7266
Error	3,65	22	0,17		
Total	3,67	23			

4. BIBLIOGRAFÍA

Instituto Nacional de Estadística y Censos, (2022). *Tabulados ESPAC de Soya*. Tomado de <https://www.ecuadorencifras.gob.ec/informacion-agroambiental-y-tecnificacion-agropecuaria-2022/>

Ministerio de Agricultura y Ganadería, (2023). *Modelo de Producción Estadística de los Operativos de Rendimientos Objetivos*. Primera Edición; pp 7-26.