

# Estructuración de Proyecto para Data Analytics y Estadística

## ¿Cómo están distribuidos los rendimientos?

Medida	Valor
Count	314.000000
Mean	93.105096
Std	31.975025
Min	4.500000
25%	72.650000
50%	93.000000
75%	114.425000
Max	188.300000

### Interpretación

**Count:** 314 — Hay entonces 314 registros válidos (no nulos) en esta columna.

**Mean:** 93.10 — El rendimiento promedio por hectárea es de 93.10 (qq/ha).

**Std:** 31.98 — Los datos varían en promedio +/- 31.98 respecto al promedio. Esto indica una dispersión considerable.

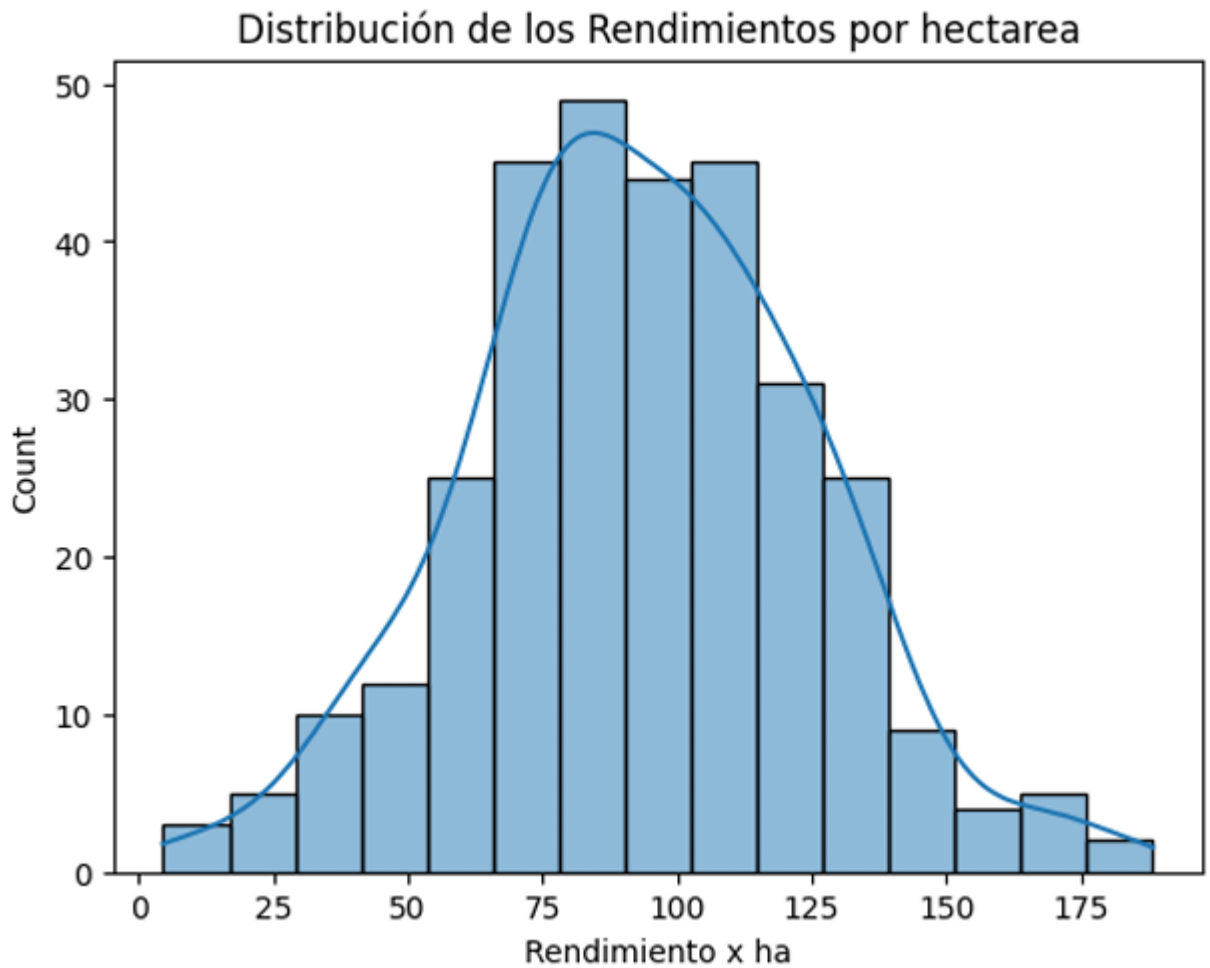
**Min:** 4.50 — El valor más bajo es muy inferior al promedio, lo cual sugiere que puede haber algunos outliers negativos.

**25%:** 72.65 — El 25% de los datos está por debajo de 72.65.

**50%:** 93.00 — La mediana, es decir el valor central, es casi igual a la media, lo que sugiere que la distribución podría ser simétrica.

**75%:** 114.43 — El 75% de los datos están por debajo de este valor.

**Max:** 188.30 — El valor más alto es el doble del promedio, indicando posibles outliers positivos o rendimientos excepcionalmente buenos.



**Análisis de forma:**

1. **¿Es simétrica o asimétrica?** Casi simétrica, con ligera asimetría hacia la derecha.
2. **¿Tiene forma de campana (normal)?** Sí, se aproxima a una forma normal.
3. **¿Está sesgada a la izquierda o derecha?** Ligeramente sesgada a la derecha (media mayor a mediana).
4. **¿Hay muchos ceros o valores extremos?** No ceros, pero sí outliers altos y bajos.
5. **¿Se concentra en ciertos rangos?** Sí, la mayoría está entre 73 y 114.

### Conclusiones finales:

- Los rendimientos observados muestran una distribución más natural, con cierta dispersión y presencia de outliers.
- Los rendimientos objetivo están fuertemente concentrados y parecen definidos por un estándar.

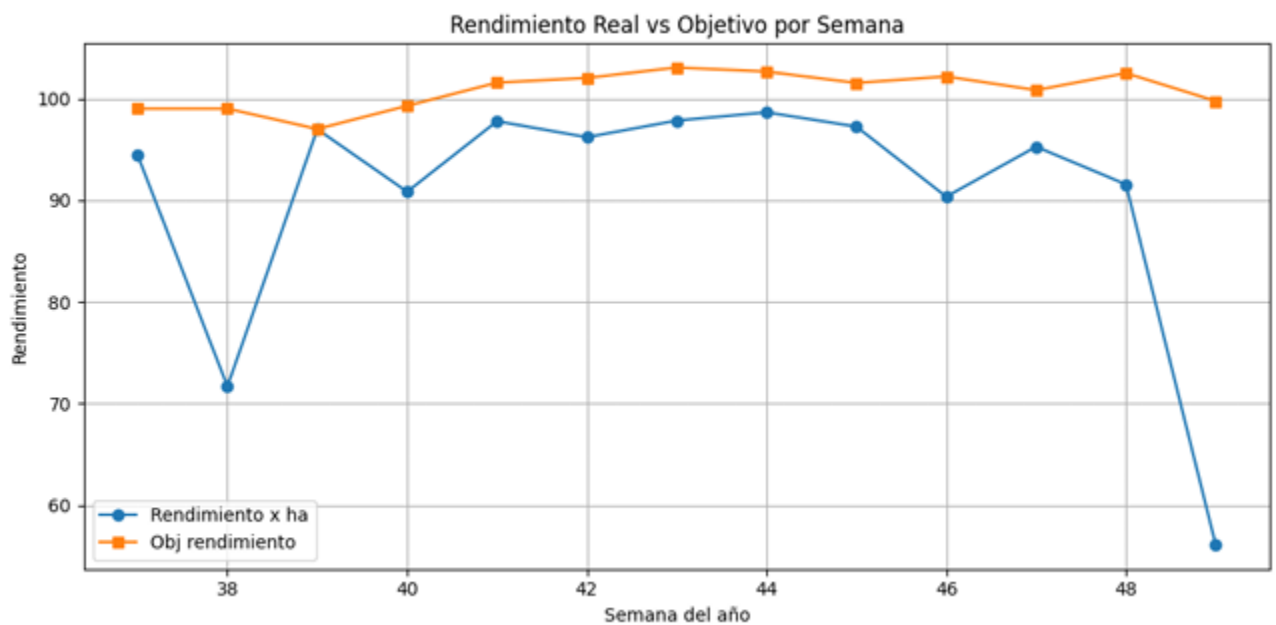
## 2. ¿Cuál es el rendimiento promedio?

- **Rendimiento promedio real por hectárea:** 93.11 aproximadamente.
- **Rendimiento objetivo promedio:** 101.25 aproximadamente.

### En comparación:

1. En promedio, los rendimientos reales están 8.14 unidades por debajo del objetivo.
2. Esto indica una brecha entre el rendimiento esperado y lo alcanzado.

## 3. ¿Cuánto varían los rendimientos entre ellos?



En general, los rendimientos reales por hectárea se mantuvieron por debajo del rendimiento objetivo a lo largo de las semanas 37 a 49. Las semanas 37 y 49 presentan las mayores desviaciones negativas, con una diferencia de más de 25 puntos respecto al objetivo.

Las semanas más cercanas al cumplimiento del objetivo fueron de la 41 a la 45, donde la diferencia se redujo a menos de 5 puntos. Esto sugiere que hubo factores que afectaron

negativamente la producción en semanas específicas, mientras que en otras semanas se mantuvo un buen nivel de eficiencia. Sería útil investigar qué ocurrió en las semanas con mayor caída para corregir posibles problemas en el futuro.

## ¿Cuál es la desviación entre el rendimiento esperado y el rendimiento real?

- **Desviación promedio:** 25.68 unidades.
- **Error porcentual medio (MAPE):** 52.30%.
- **Desviación estándar de las diferencias:** 31.08 unidades.

### Ejemplos de diferencias absolutas (algunas observaciones):

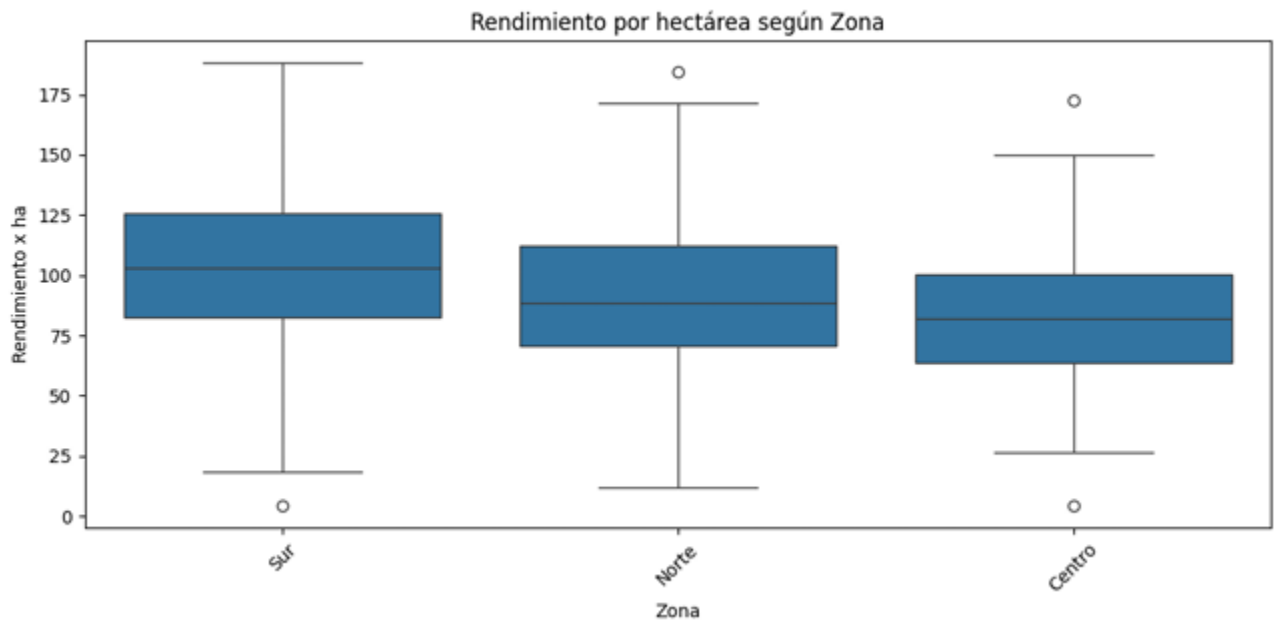
Índice	Diferencia absoluta
0	82.3
1	78.4
2	69.5
3	70.4
4	65.7
...	
311	67.0
312	87.3
313	93.8
314	97.3
315	94.5

### Ejemplos de error porcentual (algunas observaciones):

Índice	Error porcentual (%)
0	43.71
1	42.52
2	39.60
3	40.84
4	38.26
...	
311	291.30
312	466.84
313	768.85
314	2070.21
315	2100.00

**Desviación promedio:** 25.68  
**Error porcentual medio (MAPE):** 52.30%  
**Desviación estándar de las diferencias:** 31.08

#### 4. ¿Hay alguna relación entre los rendimientos y los parámetros en amarillo?



#### ANÁLISIS POR ZONA:

##### Zona Sur:

- Mediana: Cerca de 100.
- Rango intercuartílico (IQR) (la caja): Aproximadamente de 75 a 130.
- Valores extremos (outliers): Un valor muy bajo cerca de 5.
- Máximo y mínimo (sin contar outliers): Aproximadamente de 20 a 185.

##### Zona Norte:

- Mediana: Cerca de 95.
- IQR: De unos 70 a 120.
- Outliers: Uno bajo (cerca de 10) y uno alto (alrededor de 180).
- Máximos y mínimos (sin outliers): Entre 20 y 160 aprox.

##### Zona Centro:

- Mediana: Alrededor de 90.
- IQR: De 65 a 110.

- Outliers: Uno bajo (cerca de 10) y uno alto (cerca de 170).
- Rango general: De 25 a 150 aprox.

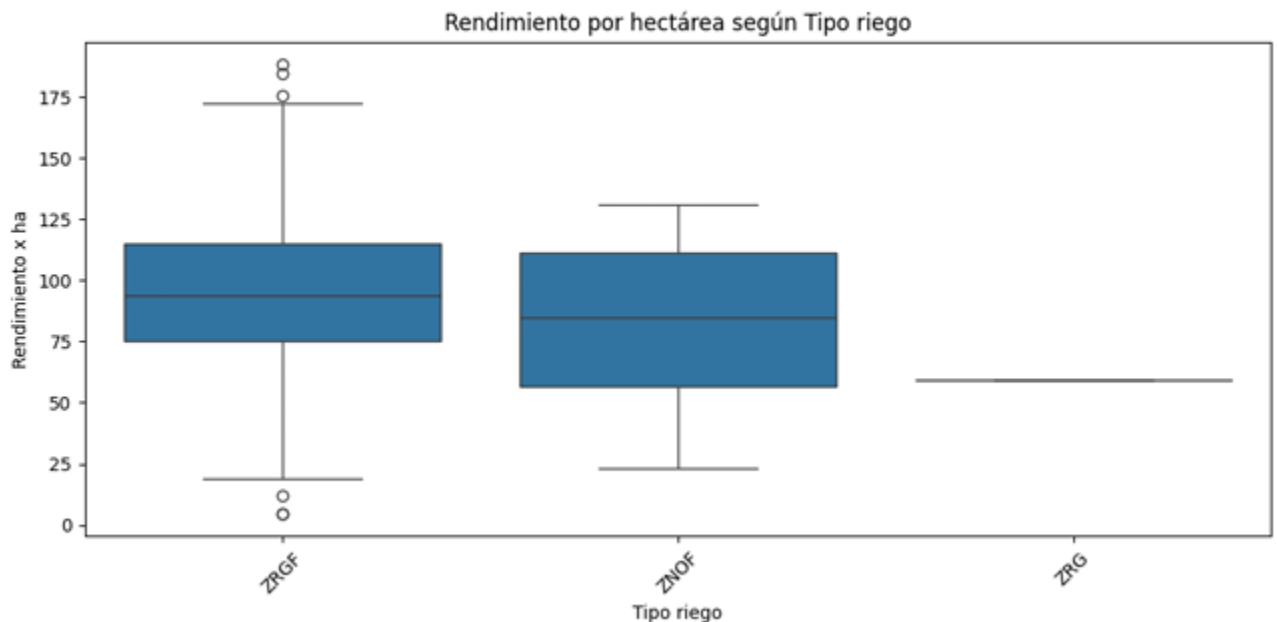
]Centro: 19.16% Norte: 6.90% Sur: 1.65%

La zona Centro presenta el mejor rendimiento relativo. El Sur tiene el desempeño más bajo, lo que podría indicar condiciones menos favorables o prácticas menos eficientes.

### Comparación general:

- **Mayor rendimiento:** En promedio, la zona Sur tiene la mediana más alta.
- **Dispersión:** La zona Sur también parece tener una mayor dispersión (rango más amplio).
- **Outliers:** Todas las zonas presentan valores atípicos bajos, y Centro y Norte también tienen altos.
- **Rendimiento más consistente:** La zona Centro tiene la caja (IQR) más compacta.

## ANÁLISIS DE RENDIMIENTO POR HECTÁREA SEGÚN SU TIPO DE RIEGO:



## 1. ZRF

- **Mediana:** 100
- **IQR:** Entre 75 y 120
- **Rango general:** De 10 a 185
- **Outliers:** Varios valores bajos ( 5-10) y algunos muy altos ( 180)

## 2. ZNOF

- **Mediana:** También alrededor de 100
- **IQR:** Más amplio, de 50 a 125
- **Rango general:** Más estrecho que ZRF, de 20 a 130
- **Outliers:** Menos visibles; parece más homogéneo, pero con mayor dispersión hacia el mínimo

## 3. ZRG

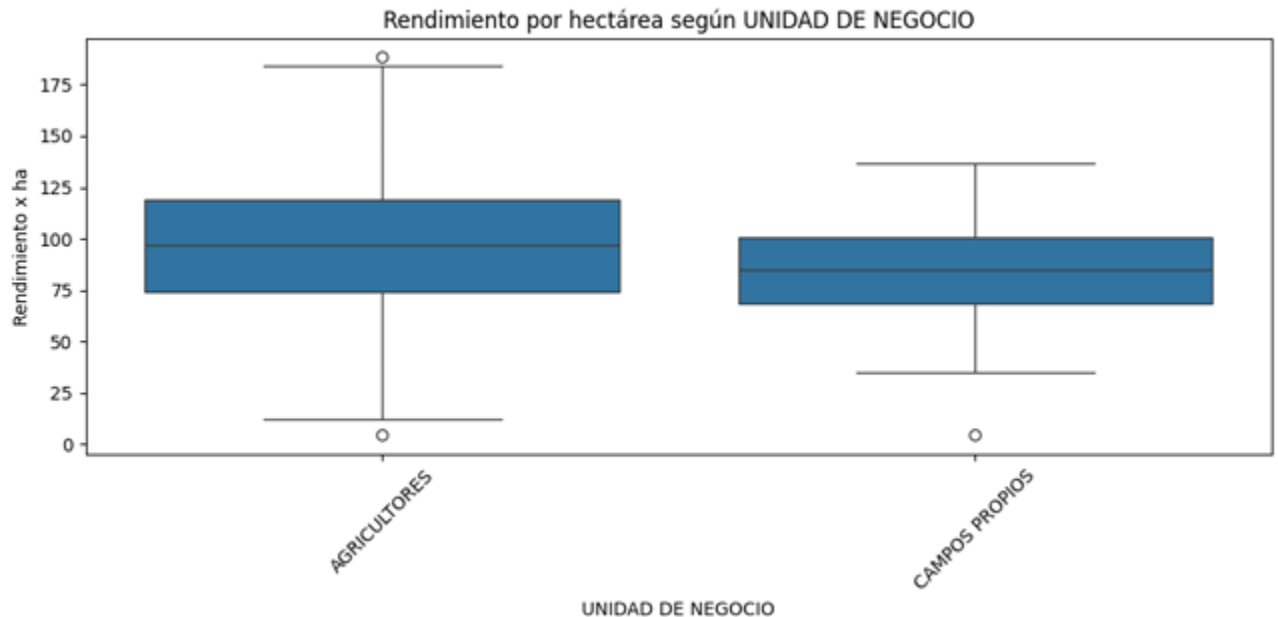
- **Observación clave:** Solo hay un valor (línea horizontal sin caja), lo que sugiere que solo hay un dato disponible — no se puede comparar estadísticamente con los otros tipos.

## Comparación:

- ZRF y ZNOF tienen rendimientos similares en la mediana.
- ZRF muestra más variabilidad (y más outliers).
- ZRG tiene datos insuficientes para análisis.
- **ZNOF:** 7.02%
- **ZRG:** 44.25%
- **ZRGF:** 8.42%

El riego tipo ZRG muestra el mayor sobrecumplimiento, incluso sin fertilización explícita, lo que sugiere alta eficiencia o ventajas particulares. ZRGF, con fertilización, no logra igual desempeño.

# ANÁLISIS DE RENDIMIENTO POR EL TIPO DE UNIDAD DE NEGOCIO



## 1. AGRICULTORES

- **Mediana:** Cerca de 105
- **IQR:** De 80 a 130
- **Rango total:** De 10 a 185
- **Outliers:** Varios bajos ( 5-10) y uno muy alto ( 180)

## 2. CAMPOS PROPIOS

- **Mediana:** Más baja, 90
- **IQR:** 60 a 110
- **Rango:** Más estrecho que Agricultores ( 20 a 150)
- **Outliers:** Uno muy bajo ( 5), menos extremos

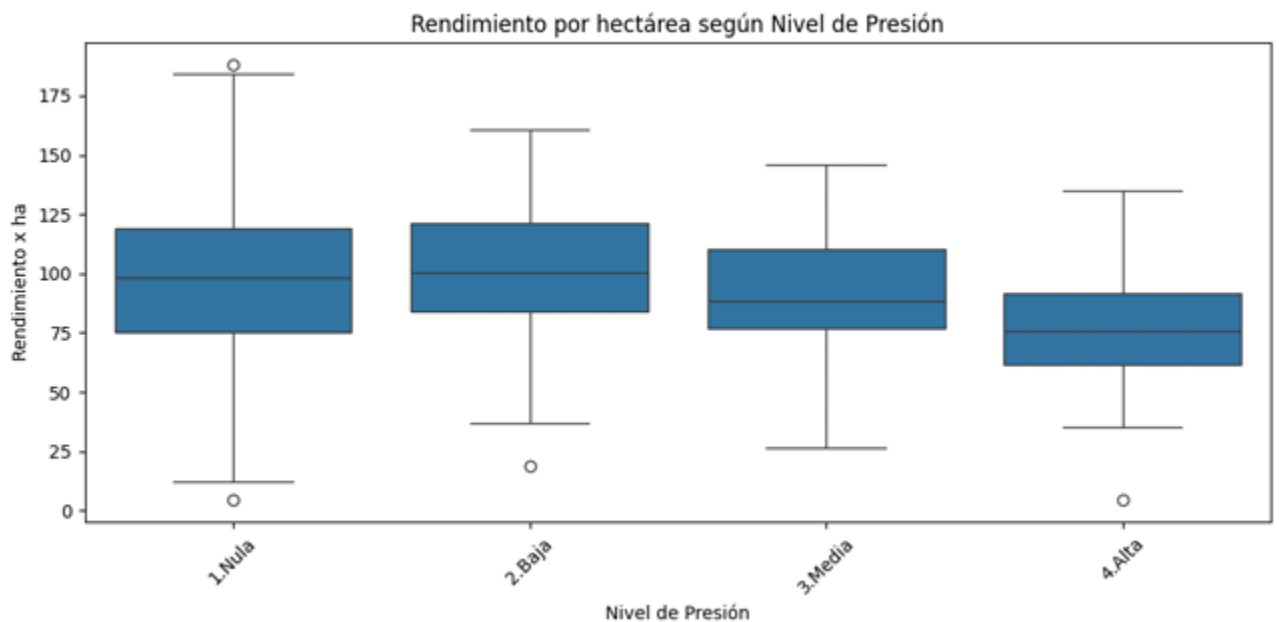
### Comparación:

- **AGRICULTORES** tienen mayor rendimiento mediano y mayor variabilidad.

- **CAMPOS PROPIOS** presentan menor rendimiento promedio y menos dispersión, aunque también tienen outliers.
- **Tipo de riego:** No hay grandes diferencias entre ZRF y ZNOF, pero ZRF parece tener rendimientos más altos y dispersos.
- **Unidad de negocio: AGRICULTORES** superan a **CAMPOS PROPIOS** tanto en mediana como en rango superior de rendimientos, aunque también presentan más variabilidad y outliers.
- **Agricultores:** 6.37%
- **Campos propios:** 14.26%

Los campos propios lograron un mejor rendimiento relativo, posiblemente debido a mayor inversión, supervisión o manejo técnico más riguroso.

## ANÁLISIS DE RENDIMIENTO POR HECTÁREA SEGÚN EL NIVEL DE PRESIÓN



### Niveles de presión:

1. Nula
2. Baja
3. Media
4. Alta

## 1. Nula y 2. Baja

- **Medianas:** Altas ( 100-110)
- **Rangos amplios:** Muchos valores altos, incluso outliers por encima de 175
- **IQR:** Similares ( 80 a 130)

## 3. Media

- **Mediana:** Ligeramente menor ( 95)
- **Observación:** Menor dispersión en general

## 4. Alta

- **Mediana:** Más baja, entre 75 y 80
- **IQR:** Más bajo y estrecho ( 60 a 90)
- **Observación:** Menos valores extremos altos

## 1 Nivel de Presión: variabilidad respecto de la comparativa

- **Nula:** 4.41%
- **Baja:** 2.11%
- **Media:** 10.15%
- **Alta:** 22.96%

Contra lo esperado, a mayor presión se obtuvieron mejores rendimientos. Esto podría indicar que los lotes con mayor presión recibieron más atención o inversión. También podría haber confusión respecto a qué representa exactamente "presión".

## Conclusión:

A mayor presión, menor el rendimiento por hectárea. La tendencia es descendente del rendimiento conforme aumenta la presión. La diferencia entre "Nula/Baja" y "Alta" es significativa visualmente. La presencia de \*Orobanche\* y el aumento en el nivel de presión están asociados con una disminución clara en el rendimiento por hectárea. Es probable que ambas variables estén relacionadas entre sí (la presión puede deberse en parte a \*Orobanche\* u otras plagas).

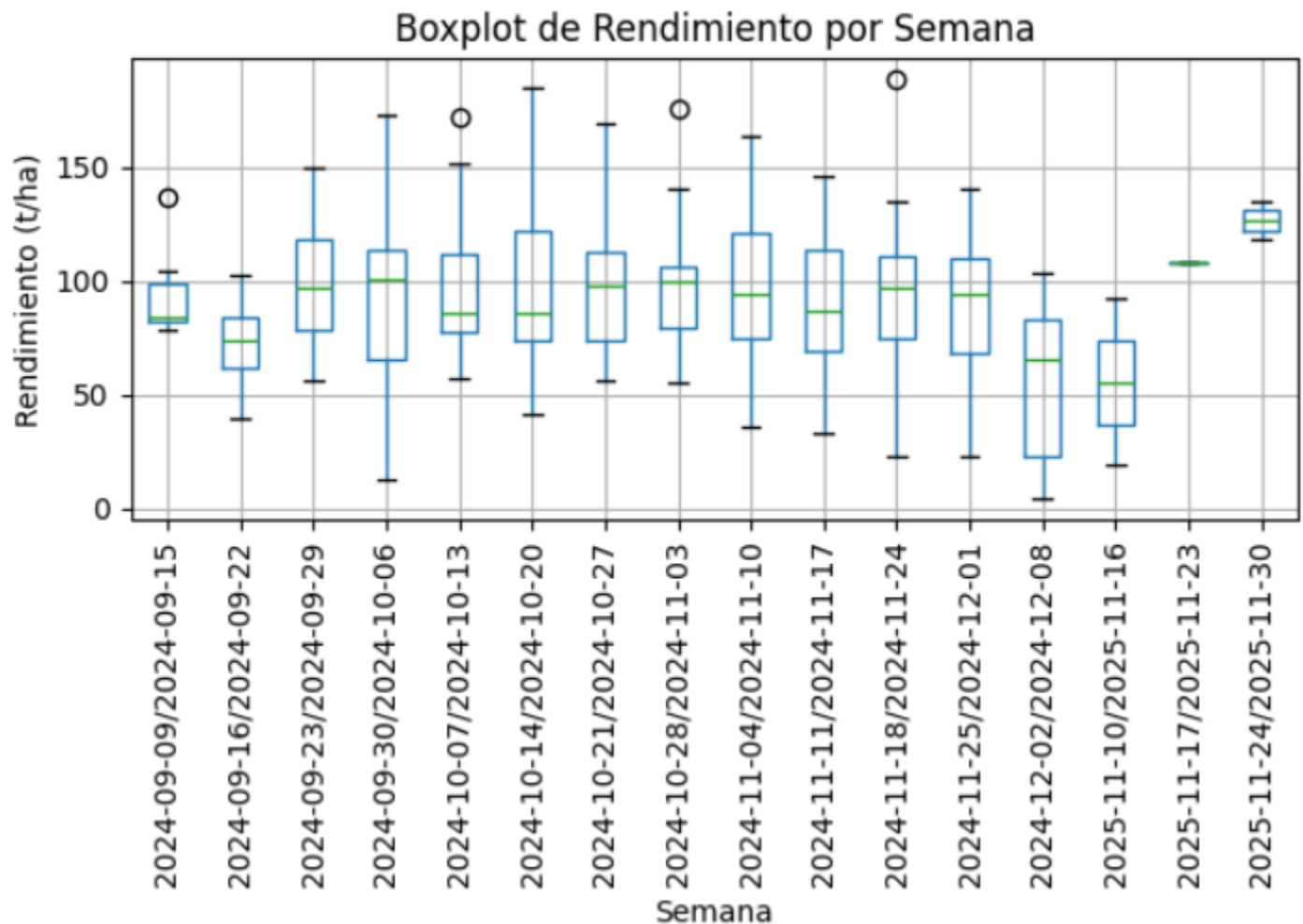
## ¿Hay alguna relación entre el nivel de presión y el rendimiento?

Sí, hay una relación inversa. Del gráfico, se observa que:

- A medida que el nivel de presión aumenta (de 1: Nula a 4: Alta), el rendimiento tiende a disminuir.
- La mediana del rendimiento baja progresivamente con cada aumento en el nivel de presión.
- También parece haber una mayor variabilidad en niveles bajos de presión y una menor variabilidad (rendimientos más consistentemente bajos) en niveles de presión alta.

**Conclusión:** A mayor nivel de presión, menor rendimiento por hectárea. Esto sugiere que el nivel de presión (posiblemente de plaga, enfermedad o maleza) tiene un efecto negativo sobre el rendimiento del cultivo.

¿Hay alguna relación entre la fecha de plantación y el rendimiento?



## 2 Análisis del Mapa de Rendimiento Promedio por Zona Geográfica

<sup>1</sup>.

### 2.1 Descripción General

El mapa muestra puntos georreferenciados (latitud vs. longitud) que representan distintas zonas agrícolas, coloreados según su rendimiento promedio medido en toneladas por hectárea (t/ha). La barra de color a la derecha indica la escala de rendimiento, que va desde los tonos púrpura (rendimiento bajo) hasta los tonos amarillo brillante (rendimiento alto).

<sup>1</sup>Consulta el mapa interactivo.

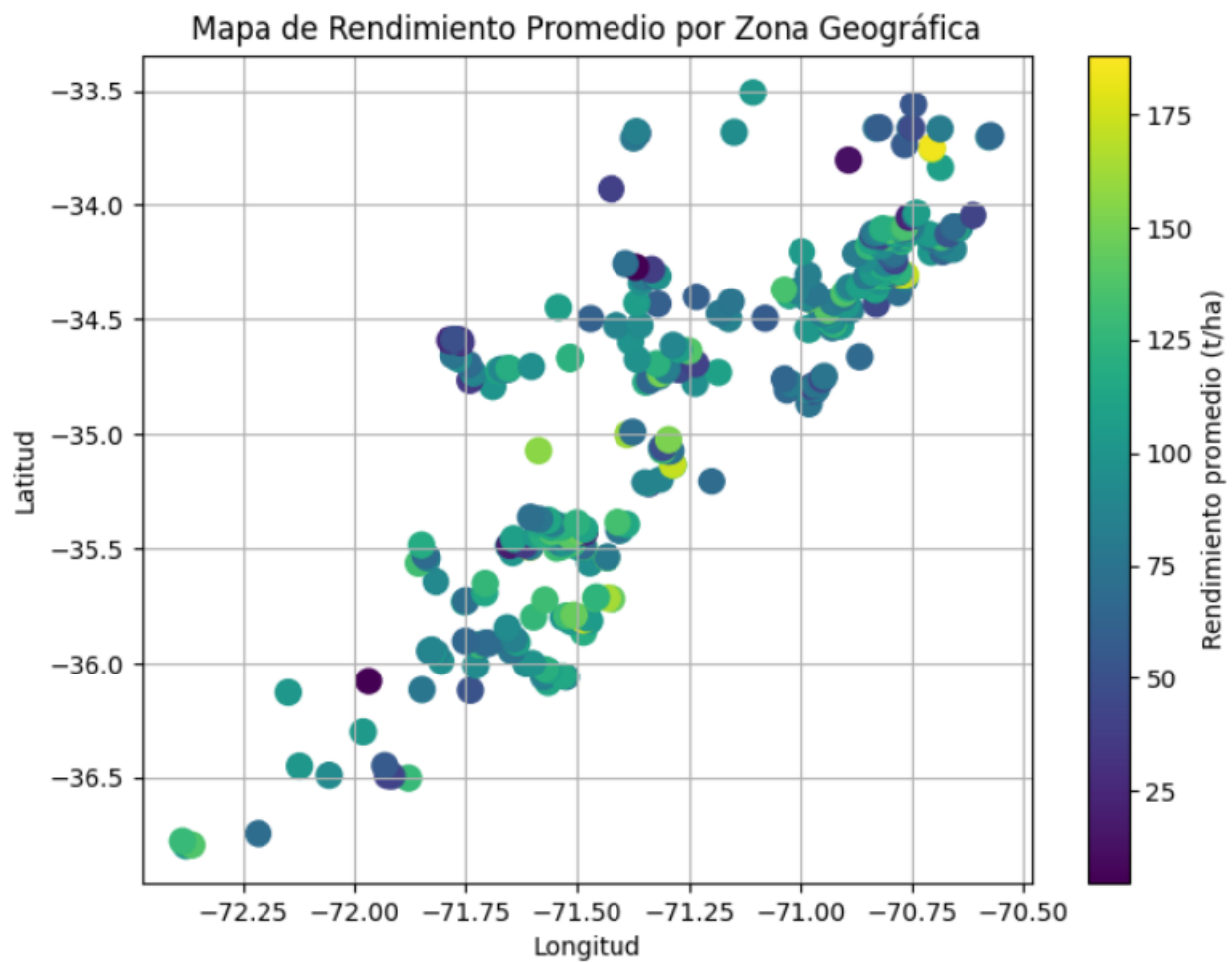


Figure 1: Mapa de rendimiento promedio por hectárea según ubicación geográfica.

## 2.2 Distribución Geográfica

- La mayoría de los puntos se agrupan en una franja que va aproximadamente de latitud  $-36.5$  a  $-33.5$ , y longitud de  $-72.25$  a  $-70.5$ , lo que sugiere una región agrícola continua.
- Se observa una tendencia en la dispersión geográfica: los puntos siguen una trayectoria diagonal de suroeste a noreste, lo cual podría estar asociado a un valle o corredor agrícola (como el Valle Central de Chile).

## 2.3 Rendimiento Promedio

- **Altos rendimientos (amarillo):** Se encuentran en zonas dispersas, sin una concentración específica. Esto podría indicar presencia de buenas prácticas agrícolas, disponibilidad de agua o suelos fértiles.
- **Bajos rendimientos (púrpura):** También distribuidos, pero con cierta concentración hacia los bordes o zonas menos densas, lo que podría reflejar condiciones subóptimas o menor acceso a recursos.

## 2.4 Posibles Interpretaciones

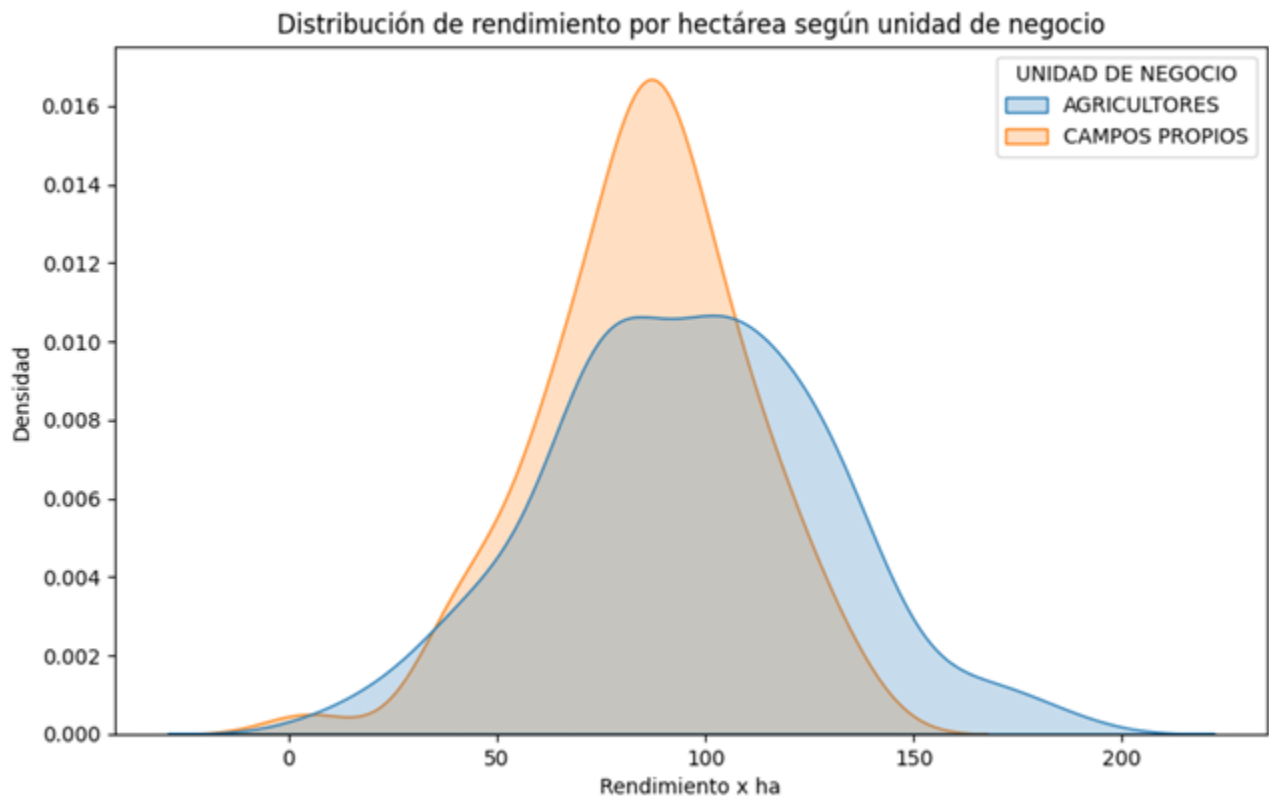
- **Factores ambientales:** El patrón observado podría estar influenciado por la altitud, el tipo de suelo, la disponibilidad hídrica o el clima regional.
- **Factores socioeconómicos:** Las diferencias de rendimiento pueden relacionarse con inversión en tecnología, tamaño del predio o acceso a infraestructura.

Se observa gran variabilidad. Algunos ejemplos:

- **2024-09-11:** 16.87%
- **2024-09-14:**  $-38.28\%$
- **2025-11-14:** 82.36%
- **2025-11-29:**  $-31.11\%$

Esto indica que la fecha de plantación es crítica para el éxito del cultivo. Plantaciones en fechas subóptimas reducen notablemente el rendimiento.

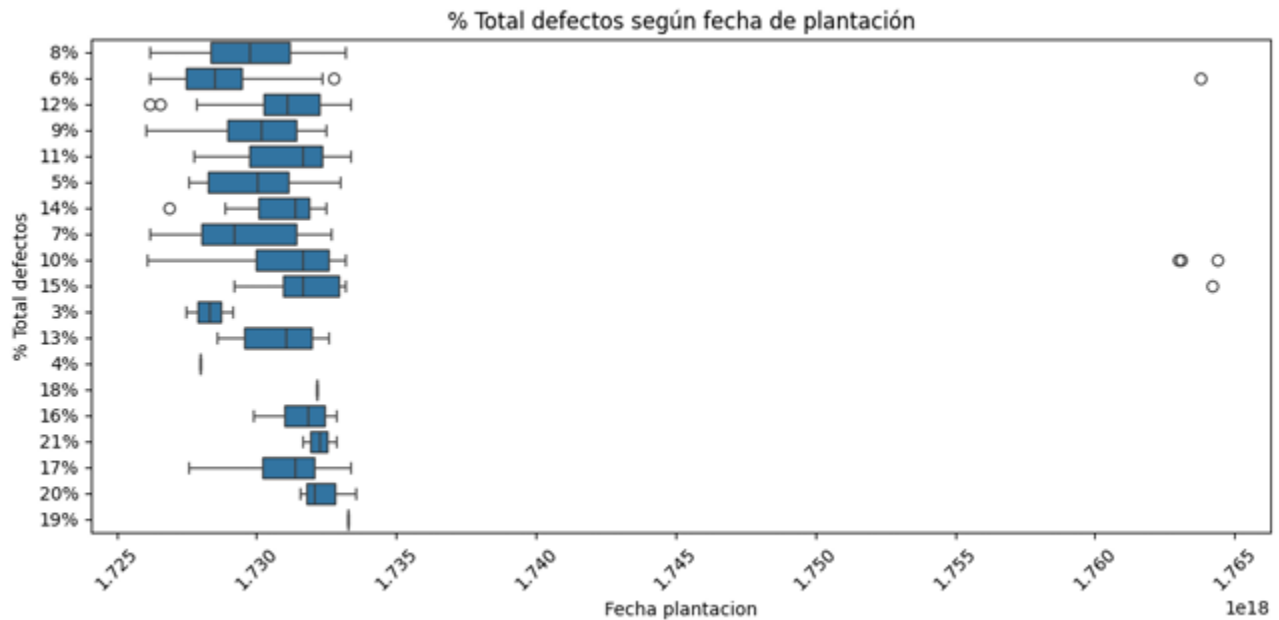
## ¿Cómo es la distribución del rendimiento entre agricultores y campos propios?



1. **¿Tiene forma de campana?** La distribución del rendimiento por hectárea en la unidad de negocio “*AGRICULTORES*” se asemeja a una curva de campana, aunque presenta una forma ligeramente aplanada. Por otro lado, “*CAMPOS PROPIOS*” también muestra una distribución unimodal, pero con una forma más estrecha y puntiaguda, menos simétrica que la anterior.
2. **¿Está sesgada?** En “*AGRICULTORES*”, la distribución presenta un leve sesgo positivo (hacia la derecha), evidenciado por una cola más extendida en ese extremo. En contraste, la distribución de “*CAMPOS PROPIOS*” muestra un sesgo negativo (hacia la izquierda), con una cola más prolongada hacia valores más bajos de rendimiento.
3. **¿Dónde se concentran los datos?** Para “*AGRICULTORES*”, la mayor densidad de observaciones se encuentra entre 80 y 120 unidades de rendimiento por hectárea, con un pico central bien definido. En cambio, “*CAMPOS PROPIOS*” concentra sus datos en un rango más bajo, aproximadamente entre 60 y 100 unidades, con un pico más agudo alrededor de los 80.
4. **¿Se distribuyen diferente entre unidades?** Sí, existe una diferencia clara entre las distribuciones. “*CAMPOS PROPIOS*” tiende a concentrar sus rendimientos en valores ligeramente más bajos y en un rango más estrecho, mientras que “*AGRICULTORES*”

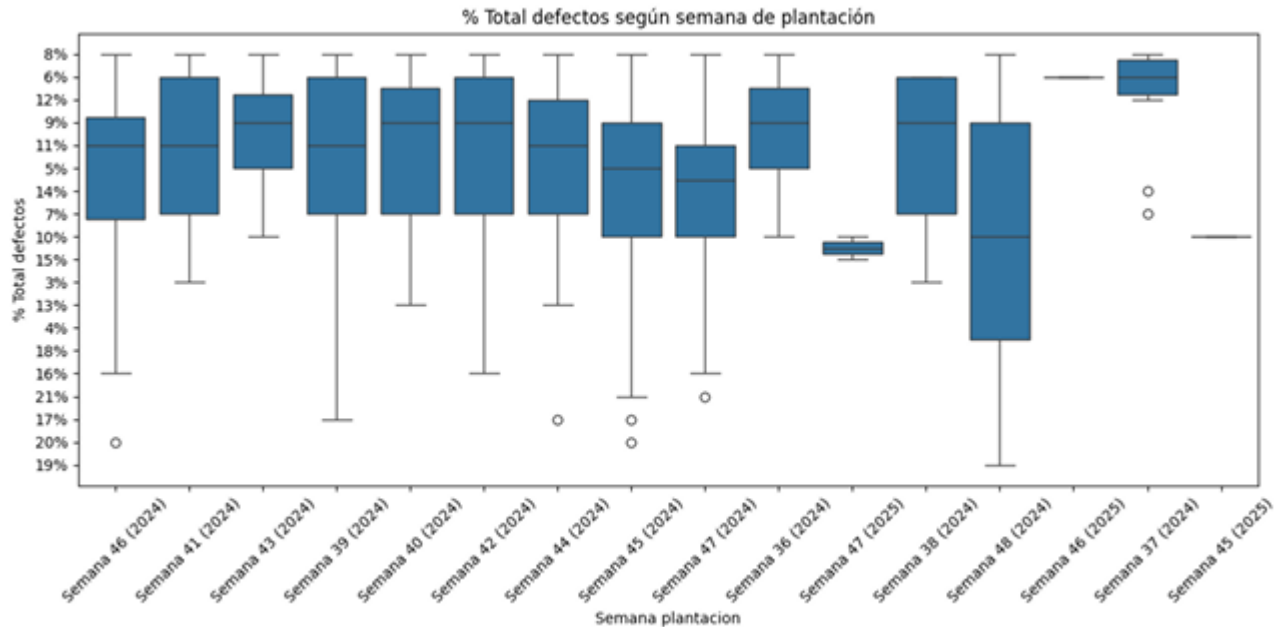
muestra una distribución más dispersa, con tendencia a rendimientos ligeramente superiores en promedio. Estas diferencias se reflejan tanto en la forma (más achatada vs. más concentrada) como en el sesgo de cada curva, lo que sugiere comportamientos distintos en el rendimiento según la unidad de negocio.

## ¿Hay alguna relación entre el porcentaje de defecto y la fecha de plantación?



### 1. Resumen del gráfico:

- **Eje X (Semana de plantación):** Agrupadas por semana y año. Algunas semanas se repiten en distintos años (por ejemplo, Semana 46 en 2024 y 2025).
- **Eje Y (% Total de defectos):** Varía aproximadamente entre 0% y 20%.
- **Boxplots:** Representan la variabilidad del porcentaje de defectos por semana; la línea dentro de cada caja indica la mediana.
- **Valores atípicos:** Se muestran como puntos fuera de los bigotes de las cajas.



## 2. Observaciones clave:

- **Variabilidad por semana:** La dispersión del porcentaje de defectos varía considerablemente entre semanas; algunas tienen mayor concentración (cajas pequeñas) y otras mayor dispersión (cajas amplias).
- **Comparación entre años para una misma semana:**
  - **Semana 46:** En 2024, la mediana está en torno al 9–10%. En 2025, sube considerablemente, cerca del 16%.
  - **Semana 47:** Ambas presentan medianas altas, pero en 2025 se mantiene el nivel elevado de la semana anterior.
  - **Semana 36:** En 2025, se observa un nivel más bajo de defectos en comparación con 2024.
  - **Semana 37:** Comportamiento inverso: mayor mediana en 2024 que en 2025.
- **Tendencia temporal dentro del mismo año:** No se aprecia una tendencia clara de aumento o disminución del porcentaje de defectos con el avance de las semanas. Por ejemplo, en 2024, semanas consecutivas pueden tener porcentajes de defectos tanto altos como bajos.

## 3. Posible patrón observado:

- El **año de plantación** parece tener mayor influencia en el porcentaje total de defectos que la semana específica.
- En 2025, las semanas 46 y 47 presentan consistentemente porcentajes más altos de defectos que en 2024.
- Para otras semanas (como la 36 y 37), el comportamiento es inconsistente, lo cual sugiere la influencia de otros factores.

- En resumen, no hay una tendencia clara por semana, pero sí diferencias notables entre años, lo que apunta a factores anuales (clima, prácticas agrícolas, calidad de insumos, etc.) como posibles responsables de la variabilidad en los defectos.