

# GAUSSIAN RANDOM FIELDS

Aplicación a Datos Agrícolas: Superficie, Producción y Rendimiento

Estadística Espacial

Luna Turpo  
Rosmery

17 de septiembre de 2025

# Contenido

- 1 Introducción y Objetivos
- 2 Datos y Metodología
- 3 Análisis de Normalidad
- 4 Parámetros del Campo Gaussiano
- 5 Fundamentos Matemáticos
- 6 Simulaciones del Campo Gaussiano
- 7 Visualizaciones del Análisis
- 8 Validación del Modelo
- 9 Resultados y Análisis
- 10 Conclusiones

## Análisis Gaussiano Funcional

El análisis gaussiano funcional es una técnica estadística que permite modelar datos multivariados usando distribuciones normales multivariadas, particularmente útil para:

- Modelar relaciones complejas entre variables
- Generar simulaciones realistas
- Predecir comportamientos futuros

## Aplicación Agrícola

En este estudio analizamos datos agrícolas peruanos para entender las relaciones entre superficie cultivada, producción y rendimiento.

# Objetivos del Análisis

- 1 **Caracterizar** las distribuciones de variables agrícolas clave
- 2 **Modelar** las relaciones entre superficie, producción y rendimiento
- 3 **Simular** escenarios agrícolas usando campos gaussianos
- 4 **Validar** la calidad del modelo estadístico
- 5 **Generar** predicciones para diferentes escenarios

# Descripción de los Datos

## Fuente de Datos

Base de datos agrícola con **165,711** observaciones iniciales y **127** variables.

## Variables Principales Analizadas

- P217\_SUP\_ha: Superficie cultivada (hectáreas)
- P219\_CANT\_1: Producción total (kilogramos)
- P220\_1\_PRE\_KG: Precio por kilogramo
- P220\_1\_VAL: Valor total de producción
- P204\_NOM: Tipo de producto agrícola

# Preparación de Datos

## Proceso de Limpieza

- 1 Filtrado de valores faltantes y valores extremos
- 2 Restricción:  $0 < \text{superficie} < 1000$  hectáreas
- 3 Cálculo de rendimiento = producción / superficie
- 4 Transformación logarítmica para normalización

## Resultado Final

**71,854** observaciones válidas para el análisis gaussiano.

## Variables Transformadas

$$\log\_superficie = \ln(superficie\_ha) \quad (1)$$

$$\log\_produccion = \ln(produccion\_kg) \quad (2)$$

$$\log\_rendimiento = \ln\left(\frac{produccion\_kg}{superficie\_ha}\right) \quad (3)$$

## Justificación

La transformación logarítmica es común en datos agrícolas porque:

- Reduce la asimetría de las distribuciones
- Estabiliza la varianza
- Mejora la normalidad de los datos

# Tests de Normalidad

Cuadro: Resultados de Tests de Normalidad

Variable	Anderson-Darling	Lilliefors	Asimetría
log_superficie	0.000000	0.000000	-0.0428
log_produccion	0.000000	0.000000	0.7253
log_rendimiento	0.000000	0.000000	0.1271

## Interpretación

- P-valores = 0 indican desviación de normalidad estricta
- log\_rendimiento tiene menor asimetría (0.1271)
- log\_superficie es casi simétrica (-0.0428)
- El gran tamaño muestral hace los tests muy sensibles



# Selección de Variables

## Variables Seleccionadas para el Campo Gaussiano

- **Variable 1:** log\_superficie
- **Variable 2:** log\_rendimiento

## Criterios de Selección

- 1 Menor asimetría en las distribuciones
- 2 Relevancia agronómica: superficie y rendimiento son fundamentales
- 3 Interpretabilidad de resultados
- 4 Calidad de la aproximación normal

# Parámetros Estimados

## Vector de Medias

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{\log\_superficie} \\ \mu_{\log\_rendimiento} \end{pmatrix} = \begin{pmatrix} -1,1828 \\ 5,8980 \end{pmatrix} \quad (4)$$

## Desviaciones Estándar

$$\sigma_{\log\_superficie} = 1,9079 \quad (5)$$

$$\sigma_{\log\_rendimiento} = 2,4148 \quad (6)$$

## Correlación Bivariada

$$\rho = -0,2657$$

**Correlación negativa moderada** entre  $\log(\text{superficie})$  y

$\log(\text{rendimiento})$

# Interpretación de Parámetros

## Significado Agronómico

- Media  $\log\_superficie = -1.18$ : La superficie típica es  $e^{-1,18} \approx 0,31$  hectáreas
- Media  $\log\_rendimiento = 5.90$ : El rendimiento típico es  $e^{5,90} \approx 365$  kg/ha
- Correlación negativa: A mayor superficie, menor rendimiento por hectárea

## Implicaciones

La correlación negativa sugiere posibles diseconomías de escala o diferencias en intensidad de cultivo entre parcelas pequeñas y grandes.

# Matriz de Covarianza

## Matriz de Covarianza Estimada

$$\Sigma = \begin{pmatrix} \sigma_{\log\_superficie}^2 & \sigma_{\log\_superficie, \log\_rendimiento} \\ \sigma_{\log\_superficie, \log\_rendimiento} & \sigma_{\log\_rendimiento}^2 \end{pmatrix} \quad (7)$$

$$\Sigma = \begin{pmatrix} 3,640 & -1,225 \\ -1,225 & 5,831 \end{pmatrix} \quad (8)$$

## Estructura de Dependencia

La covarianza negativa (-1.225) confirma la relación inversa entre superficie y rendimiento logarítmicos.

# Distribución Normal Bivariada

## Función de Densidad

La función de densidad de probabilidad del campo gaussiano bivariado es:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (9)$$

donde  $x = (x_1, x_2)^T$  representa  $\log(\text{superficie})$  y  $\log(\text{rendimiento})$ .

## Parámetros Clave

- $\mu$ : Vector de medias
- $\Sigma$ : Matriz de covarianza
- $|\Sigma|$ : Determinante de la matriz de covarianza

## Elipse de Concentración

La distribución normal bivariada forma elipses de concentración donde:

- Centro: Vector de medias  $\mu$
- Orientación: Determinada por los vectores propios de  $\Sigma$
- Forma: Determinada por los valores propios de  $\Sigma$

## En Nuestro Caso

- Centro en  $(-1.18, 5.90)$
- Correlación negativa genera elipse inclinada
- Mayor variabilidad en  $\log(\text{rendimiento})$

# Estrategia de Simulación

## Distribución Gaussiana Bivariada

$$X \sim \mathcal{N}(\mu, \Sigma)$$

## Escenarios Simulados

- 1 Escenario Base: 2,000 simulaciones con parámetros estimados
- 2 Escenario Conservador: 500 simulaciones con  $\mu \times 0,9$  y  $\Sigma \times 0,8$
- 3 Escenario Optimista: 500 simulaciones con  $\mu \times 1,1$  y  $\Sigma \times 1,2$

## Método

Utilizamos la función `mvrnorm` del paquete MASS en R para generar muestras de la distribución normal multivariada.

# Interpretación de Escenarios

## Escenario Conservador

- Medias reducidas en 10 %: menor productividad esperada
- Varianza reducida en 20 %: menor incertidumbre
- Representa condiciones adversas (sequía, plagas, etc.)

## Escenario Optimista

- Medias aumentadas en 10 %: mayor productividad esperada
- Varianza aumentada en 20 %: mayor incertidumbre
- Representa condiciones favorables (tecnología, clima óptimo)

## Aplicación Práctica

Estos escenarios permiten evaluación de riesgos y planificación agrícola bajo diferentes condiciones.



# Distribuciones Marginales

## Gráficos de Densidad

- Histogramas: Distribución observada de  $\log(\text{superficie})$  y  $\log(\text{rendimiento})$
- Curva roja: Densidad empírica (kernel density)
- Curva azul punteada: Distribución normal teórica
- Interpretación: Evaluación visual de la aproximación normal

## Hallazgo Visual

La transformación logarítmica mejora considerablemente la aproximación a la normalidad, especialmente para  $\log(\text{superficie})$ .

# Q-Q Plots (Gráficos Cuantil-Cuantil)

## Metodología Q-Q

- Compara cuantiles observados vs. cuantiles teóricos normales
- Línea recta: Perfecta normalidad
- Desviaciones: Indican apartamiento de normalidad

## Interpretación

- log\_superficie: Sigue aproximadamente la línea recta (buena normalidad)
- log\_rendimiento: Ligeras desviaciones en las colas
- Ambas variables son adecuadas para el modelo gaussiano

# Campo Gaussiano Bivariado

## Visualización del Campo

- Puntos rojos: Datos observados (71,854 observaciones)
- Puntos azules: Simulaciones del campo gaussiano (2,000)
- Patrón: Evaluación visual del ajuste del modelo

## Resultado

Las simulaciones reproducen adecuadamente el patrón de dispersión y la correlación negativa observada en los datos reales.

# Contornos de Densidad

## Análisis de Contornos

- Contornos rojos: Densidad de datos observados
- Contornos azules: Densidad de simulaciones gaussianas
- Comparación: Evaluación de la similitud de patrones

## Interpretación

La superposición de contornos confirma que el modelo gaussiano captura correctamente la estructura de dependencia bivariada.

# Mapa de Calor de Densidad

## Función de Densidad

- Colores cálidos: Mayor densidad de probabilidad
- Colores fríos: Menor densidad de probabilidad
- Puntos blancos: Observaciones reales superpuestas

## Insights

- La máxima densidad se concentra cerca de las medias estimadas
- La forma elíptica confirma la correlación bivariada
- Las observaciones se concentran en las regiones de alta densidad teórica

# Tests de Bondad de Ajuste

## Test de Kolmogorov-Smirnov

Comparación entre datos observados y simulaciones:

- log\_superficie: p-valor = 0.000016
- log\_rendimiento: p-valor  $\approx 0$

## Interpretación

- P-valores muy pequeños indican diferencias detectables
- Resultado esperado dado el gran tamaño muestral (71,854 obs.)
- Tests muy sensibles con muestras grandes
- La aproximación gaussiana sigue siendo útil para simulación

# Evaluación Visual del Modelo

## Métodos de Validación Visual

- 1 Q-Q Plots: Comparación cuantil-cuantil con distribución normal
- 2 Histogramas: Densidad observada vs. densidad teórica normal
- 3 Gráficos bivariados: Dispersión de datos vs. simulaciones
- 4 Contornos de densidad: Comparación de patrones bivariados
- 5 Mapas de calor: Visualización de la función de densidad

## Resultado General

Las visualizaciones muestran una aproximación razonable del modelo gaussiano a los datos agrícolas, especialmente en las regiones centrales de la distribución.

# Matriz de Correlación - Variables Originales

## Correlaciones Importantes

Análisis de correlaciones entre variables originales (superficie, producción, rendimiento, precio):

- Permite entender relaciones económicas
- Identifica variables redundantes o complementarias
- Orienta decisiones de política agrícola

## Insight Clave

La correlación negativa entre superficie y rendimiento sugiere que las explotaciones más grandes tienden a tener menor intensidad productiva por hectárea.



## Usos Prácticos del Campo Gaussiano

- 1 Planificación Agrícola: Simulación de resultados bajo diferentes estrategias
- 2 Análisis de Riesgo: Evaluación de probabilidades de eventos extremos
- 3 Políticas Públicas: Modelado de impactos de intervenciones
- 4 Seguro Agrícola: Cálculo de primas basado en distribuciones de riesgo
- 5 Investigación: Generación de datos sintéticos para estudios

# Limitaciones del Modelo

## Limitaciones Identificadas

- Normalidad: Los datos no siguen perfectamente una distribución normal
- Linealidad: El modelo asume relaciones lineales en escala logarítmica
- Estacionaridad: No considera cambios temporales en parámetros
- Independencia: No modela dependencias espaciales entre parcelas

## Recomendaciones

- Considerar modelos de mezcla de gaussianas
- Incorporar efectos espaciales y temporales
- Explorar transformaciones alternativas
- Validar en submuestras por región o tipo de cultivo

## Logros Principales

- Modelo gaussiano bivariado exitosamente ajustado a 71,854 observaciones
- Parámetros estimados:  $\mu = (-1,18, 5,90)$ ,  $\rho = -0,27$
- 3 escenarios de simulación implementados (base, conservador, optimista)
- 2,000 simulaciones principales más 1,000 simulaciones de escenarios
- Validación completa con múltiples métodos estadísticos y visuales

# Conclusiones Principales

## Hallazgos Clave

- 1 La transformación logarítmica mejora significativamente la normalidad
- 2 Existe una relación inversa moderada entre superficie y rendimiento
- 3 El modelo gaussiano proporciona una aproximación útil para simulación
- 4 Los escenarios permiten análisis de sensibilidad efectivo

## Valor Agregado

Este análisis proporciona una base sólida para la toma de decisiones en el sector agrícola peruano mediante simulación estadística robusta.

## Extensiones Propuestas

- 1 Análisis Temporal: Incorporar series de tiempo para capturar tendencias
- 2 Efectos Espaciales: Modelar autocorrelación geográfica
- 3 Variables Adicionales: Incluir clima, tipo de suelo, tecnología
- 4 Modelos No-Paramétricos: Explorar alternativas más flexibles
- 5 Aplicación por Cultivos: Análisis específico por tipo de producto

## Meta Final

Desarrollar un sistema de apoyo a decisiones basado en simulación gaussiana para el sector agrícola.

# ¿Preguntas?

Análisis Gaussiano Funcional  
Datos Agrícolas

Gracias por su atención