

Análisis de Sistemas de Riego en Puno usando Bootstrap No Paramétrico

Estudiante Yhack Bryan Aycaya Paco
Docente Dr. Fred Torres Cruz
Curso Estadística Espacial
Institución Universidad Nacional del Altiplano (UNA PUNO)

Posicionamiento del trabajo

Este trabajo se enmarca en el curso de Estadística Espacial de la Universidad Nacional del Altiplano (UNA PUNO). Se analiza la variable P_{212} del dataset de la Encuesta Nacional Agraria (ENA) 2023, que describe los sistemas de riego para cultivos en el departamento de Puno. La variable es categórica, con 7 categorías: 1 = Lluvia (Secano), 2 = Río, 3 = Manantial o puquio, 4 = Pozo/agua subterránea, 5 = Reservorio, 6 = Pequeño reservorio, 7 = Otro. Se procesaron 8801 observaciones válidas de 94 distritos. Dado que la distribución de P_{212} es altamente sesgada (94.47 % en Lluvia/Secano) y no normal (prueba Kolmogorov-Smirnov: p-valor $\hat{2}.2e-16$), se empleó el bootstrap no paramétrico (método percentil) para estimar proporciones e intervalos de confianza (IC) al 95 %, evitando suposiciones paramétricas inadecuadas. Este análisis es relevante para entender la dependencia agrícola de sistemas no irrigados en Puno, con implicaciones para políticas frente al cambio climático.

Objetivos

El objetivo principal es estimar las proporciones de los sistemas de riego utilizados en Puno y calcular sus IC al 95 % mediante bootstrap no paramétrico (método percentil). Específicamente, se busca:

- Determinar la prevalencia de cada sistema de riego en las unidades agrícolas de Puno.
- Proporcionar estimaciones robustas de las proporciones, sin asumir normalidad.
- Analizar la dependencia de la agricultura en sistemas no irrigados (e.g., Lluvia/Secano) para informar políticas agrícolas.

Plan de trabajo

El análisis se desarrolló en los siguientes pasos:

1. Carga y filtrado del dataset ENA 2023 para Puno, seleccionando P_{212} y eliminando valores faltantes (8801 observaciones válidas).
2. Conversión de P_{212} a variable categórica con etiquetas descriptivas.
3. Cálculo de frecuencias y proporciones observadas para cada categoría.
4. Aplicación de bootstrap no paramétrico con 1000 réplicas para estimar IC al 95 % (método percentil).
5. Cálculo de sesgo y error estándar bootstrap para evaluar la robustez de las estimaciones.

Resultados

Frecuencias y proporciones

Se calcularon las frecuencias y proporciones de cada categoría de P_{212} :

| Categoría | Frecuencia | Proporción |
|-----------------|------------|------------|
| Lluvia (Secano) | 8314 | 0.9447 |

| | | |
|-----------------------|-----|--------|
| Río | 253 | 0.0287 |
| Manantial o puquio | 179 | 0.0203 |
| Pozo/agua subterránea | 15 | 0.0017 |
| Reservorio | 13 | 0.0015 |
| Pequeño reservorio | 10 | 0.0011 |
| Otro | 17 | 0.0019 |

Cuadro 1: Frecuencias y proporciones de sistemas de riego en Puno.

Intervalos de confianza (Bootstrap Percentil)

Se generaron 1000 réplicas bootstrap para estimar los IC al 95 %:

| Categoría | Proporción Observada | IC 95 % [Límite Inferior, Límite Superior] |
|-----------------------|----------------------|--|
| Lluvia (Secano) | 0.9447 | [0.9401, 0.9493] |
| Río | 0.0287 | [0.0255, 0.0325] |
| Manantial o puquio | 0.0203 | [0.0173, 0.0234] |
| Pozo/agua subterránea | 0.0017 | [0.0009, 0.0026] |
| Reservorio | 0.0015 | [0.0007, 0.0023] |
| Pequeño reservorio | 0.0011 | [0.0005, 0.0019] |
| Otro | 0.0019 | [0.0010, 0.0028] |

Cuadro 2: Intervalos de confianza bootstrap para proporciones.

Sesgo y error estándar

El sesgo y error estándar bootstrap son:

| Categoría | Sesgo | Error Estándar |
|-----------------------|-----------|----------------|
| Lluvia (Secano) | -0.000037 | 0.002363 |
| Río | 0.000021 | 0.001735 |
| Manantial o puquio | -0.000011 | 0.001538 |
| Pozo/agua subterránea | 0.000013 | 0.000435 |
| Reservorio | 0.000008 | 0.000410 |
| Pequeño reservorio | 0.000004 | 0.000355 |
| Otro | 0.000002 | 0.000470 |

Cuadro 3: Sesgo y error estándar bootstrap para proporciones.

Justificación del uso del bootstrap no paramétrico (método percentil)

El método bootstrap no paramétrico (percentil) se utilizó por las siguientes razones:

- **Distribución no normal:** La variable P_{212} es categórica y altamente sesgada (94.47 % en Lluvia/Secano), como confirmó la prueba Kolmogorov-Smirnov (p-valor ¡2.2e-16). Esto descarta métodos paramétricos que asumen normalidad.
- **Robustez para proporciones:** El método percentil calcula los IC directamente de los percentiles 2.5 % y 97.5 % de las réplicas bootstrap, manejando distribuciones sesgadas y categorías raras (¡0.2 %).
- **Tamaño muestral grande:** Con 8801 observaciones, el bootstrap converge rápidamente, proporcionando estimaciones precisas (errores estándar bajos, e.g., 0.002363 para Lluvia/Secano).
- **Sin suposiciones paramétricas:** El bootstrap captura la variabilidad empírica sin asumir una distribución específica, ideal para datos categóricos.

Conclusiones

El 94.47 % de las unidades agrícolas en Puno dependen de Lluvia/Secano (IC: [94.01 %, 94.93 %]), evidenciando una alta vulnerabilidad climática. Los sistemas de riego como Río (2.87 %) y Manantial o puquio (2.03 %) son minoritarios, mientras que los sistemas artificiales son raros (¡0.2 %). El bootstrap no paramétrico asegura estimaciones robustas sin suposiciones de normalidad. Estos resultados sugieren la necesidad de políticas para promover sistemas de riego en Puno frente al cambio climático.