

Bootstrap y visualización de la variable P504B (Asistencia técnica) - PUNO 2024

ILMA MAGDA MAMANI MAMANI

16 de septiembre de 2025
Curso: Estadística Espacial

1. Introducción

Este documento presenta el análisis bootstrap y la visualización de la variable P504B, que representa el tipo de asistencia técnica recibida en el contexto de PUNO 2024. El análisis se realiza utilizando el lenguaje de programación R. Se cargan datos de un archivo .sav, se filtran para el departamento de Puno, y se aplica el método bootstrap para estimar las proporciones de cada categoría de asistencia técnica junto con sus intervalos de confianza al 95 %. Además, se generan gráficos para visualizar los resultados.

El objetivo es estimar de manera robusta las proporciones de las diferentes categorías de asistencia técnica, considerando la variabilidad muestral mediante remuestreo bootstrap. Esto es útil en estadística espacial para entender distribuciones en regiones específicas como Puno.

2. Código en R

A continuación, se presenta el código completo utilizado para el análisis.

```
# =====  
# Bootstrap y visualización de la variable P504B (Asistencia técnica) - PUNO  
# 2024  
# =====  
  
# Paquetes necesarios  
library(haven) # para leer .sav  
library(dplyr)  
library(ggplot2)  
library(tidyr)  
  
# Función para cargar datos filtrados  
cargar_datos_p504b_2024 <- function(ruta_archivo = "D:/decimo 10/Estadística  
Espacial/2024/2024 - DESCOMPRIMIDO/973-Modulo1905_vacunas/13_CAP500AB.sav") {  
  data <- read_sav(ruta_archivo)  
  puno_data_clean <- data %>%  
    filter(NOMBREDD == "PUNO") %>%  
    select(ANI0, CCDD, NOMBREDD, CCPP, NOMBREPV, CCDI, NOMBREDI, P504B) %>%  
    filter(!is.na(P504B)) %>%  
    mutate(  
      codigo_distrito = paste0(CCDD, CCPP, CCDI),  
      asistencia_codigo = as.numeric(P504B),  
      distrito_limpio = trimws(toupper(NOMBREDI)),  
      provincia_limpio = trimws(toupper(NOMBREPV))  
    )  
  return(puno_data_clean)  
}
```

```

# Bootstrap sobre variable P504B
puno_data <- cargar_datos_p504b_2024()
set.seed(123) # reproducibilidad
B <- 1000 # n mero de replicaciones bootstrap
bootstrap_props <- matrix(NA, nrow = B, ncol = 7)

for (i in 1:B) {
  sample_indices <- sample(1:nrow(puno_data), size = nrow(puno_data), replace =
    TRUE)
  bootstrap_sample <- puno_data[sample_indices, ]
  props <- table(factor(bootstrap_sample$asistencia_codigo, levels = 1:7)) / nrow(
    bootstrap_sample)
  bootstrap_props[i, ] <- props
}

# Resumen en data frame con etiquetas
etiquetas <- c(
  "1. Asesor autorizado",
  "2. M dico veterinario",
  "3. Ingeniero zootecnista",
  "4. Personal SENASA",
  "5. El mismo productor/a",
  "6. T cnico agropecuario",
  "7. Otro"
)

resumen_bootstrap <- data.frame(
  Categoria = factor(1:7, labels = etiquetas),
  Proporcion_media = colMeans(bootstrap_props),
  IC_2.5 = apply(bootstrap_props, 2, quantile, probs = 0.025),
  IC_97.5 = apply(bootstrap_props, 2, quantile, probs = 0.975)
)

print(resumen_bootstrap)

# Gr fico 1: Barras con intervalos de confianza
ggplot(resumen_bootstrap, aes(x = Categoria, y = Proporcion_media)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  geom_errorbar(aes(ymin = IC_2.5, ymax = IC_97.5), width = 0.2, color = "darkblue") +
  labs(
    title = "Proporciones estimadas por Bootstrap",
    subtitle = "Variable P504B - Tipo de asistencia t cnica (Puno 2024)",
    x = "Categor a de asistencia",
    y = "Proporci n"
  ) +
  theme_minimal() +
  coord_flip()

# Gr fico 2: Distribuci n bootstrap de TODAS las categor as (facetado)
df_long <- data.frame(bootstrap_props) %>%
  pivot_longer(cols = everything(), names_to = "Categoria", values_to = "
    Proporcion")

df_long$Categoria <- factor(df_long$Categoria,
  levels = paste0("X",1:7),
  labels = etiquetas)

```

```
ggplot(df_long, aes(x = Proporción)) +
  geom_histogram(bins = 30, fill = "lightgreen", color = "black") +
  facet_wrap(~Categoría, scales = "free") +
  labs(
    title = "Distribuciones bootstrap de proporciones por categoría",
    x = "Proporción",
    y = "Frecuencia"
  ) +
  theme_minimal()
```

3. Resultados

3.1. Tabla de Resumen Bootstrap

La siguiente tabla muestra las proporciones medias estimadas mediante bootstrap para cada categoría, junto con los intervalos de confianza al 95 %.

Categoría	Proporción media	IC 2.5 %	IC 97.5 %
1. Asesor autorizado	0.003162901	0.001050972	0.005793484
2. Médico veterinario	0.090007357	0.076720967	0.101944298
3. Ingeniero zootecnista	0.004211771	0.001576458	0.007356805
4. Personal SENASA	0.074166579	0.062532843	0.085667367
5. El mismo productor/a	0.105678403	0.091946926	0.119298476
6. Técnico agropecuario	0.704306884	0.683657383	0.725170783
7. Otro	0.018466106	0.012611666	0.024697846

Cuadro 1: Resumen de proporciones bootstrap para P504B.

3.2. Gráficos

3.2.1. Gráfico 1: Barras con Intervalos de Confianza

Este gráfico muestra las proporciones estimadas para cada categoría con barras y los intervalos de confianza al 95 % representados por líneas de error. La categoría dominante es "Técnico agropecuario" con una proporción cercana al 70 %.

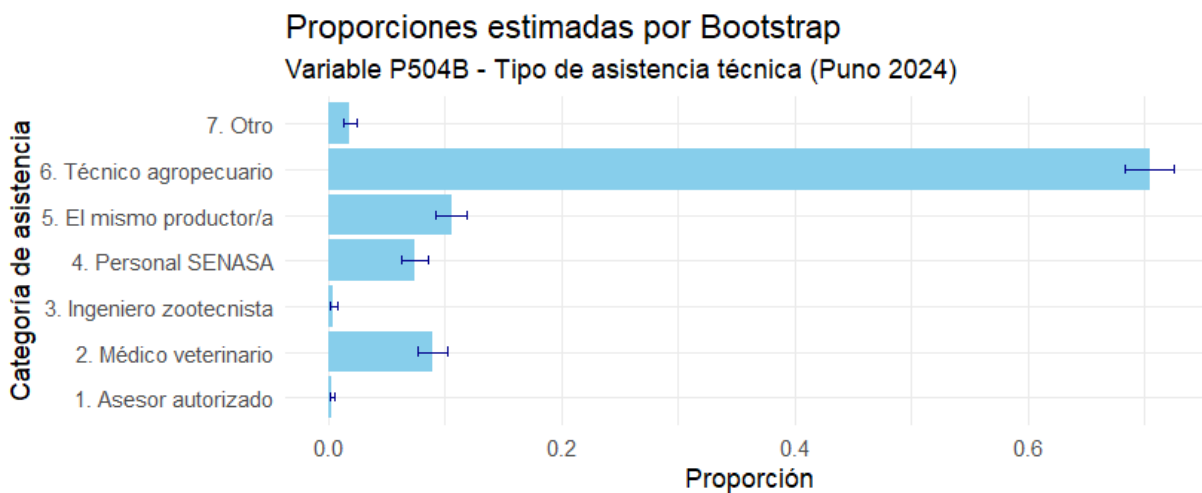


Figura 1: Barras con Intervalos de Confianza

3.2.2. Gráfico 2: Distribuciones Bootstrap Facetadas

Este gráfico facetado presenta histogramas de las distribuciones bootstrap para cada categoría, mostrando la variabilidad de las proporciones en las 1000 replicaciones. Las distribuciones son aproximadamente normales, lo que valida el uso de quantiles para los intervalos de confianza.

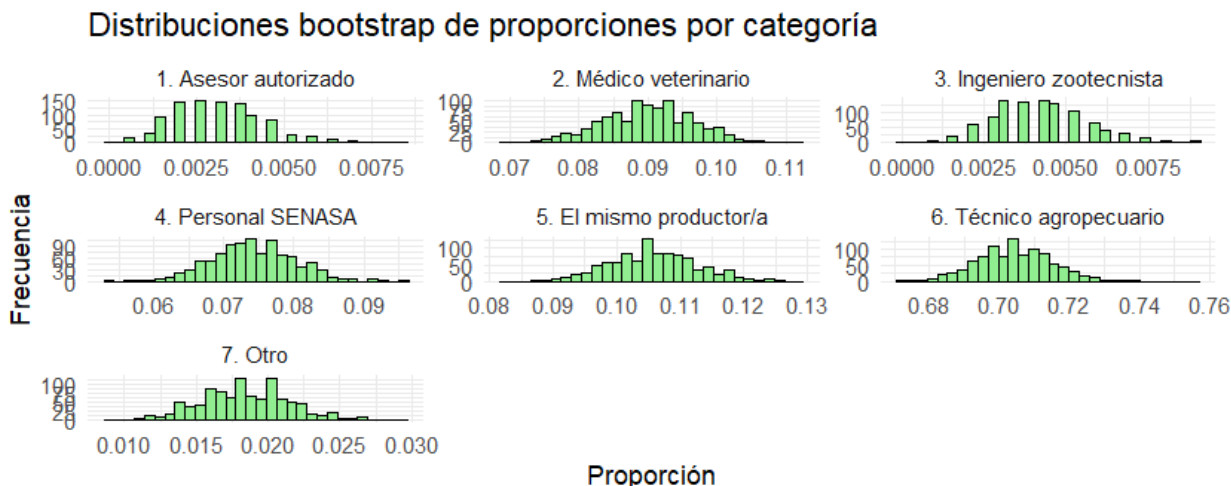


Figura 2: Distribuciones Bootstrap Facetadas

4. Explicación del Código

El código comienza cargando paquetes necesarios como `haven` para leer archivos `.sav` (de SPSS), `dplyr` para manipulación de datos, `ggplot2` para visualización y `tidyr` para reformato.

Se define una función `cargar_datos_p504b.2024` que lee el archivo, filtra por el departamento "PUNO", selecciona variables relevantes, elimina NA en P504B y crea nuevas variables limpias.

Luego, se aplica bootstrap: se fijan una semilla para reproducibilidad, se define $B=1000$ replicaciones. En un bucle, se remuestrea con reemplazo, calcula proporciones para categorías 1 a 7 y las almacena.

Se crea un data frame de resumen con medias y quantiles (IC 95%).

Finalmente, se generan dos gráficos: uno de barras con error bars y otro de histogramas facetados.

5. Interpretación de los Resultados

El análisis bootstrap revela que la categoría más frecuente de asistencia técnica en Puno 2024 es "6. Técnico agropecuario" con una proporción media de aproximadamente 0.704 (70.4%), y un intervalo de confianza entre 0.684 y 0.725, indicando una alta prevalencia y baja variabilidad relativa.

Le siguen "5. El mismo productor/a" (10.6%) y "2. Médico veterinario" (9.0%), mientras que categorías como "1. Asesor autorizado" (0.3%) y "3. Ingeniero zootecnista" (0.4%) son muy raras, con intervalos que incluyen valores cercanos a cero.

Esto sugiere que en Puno, la asistencia técnica está dominada por técnicos agropecuarios, posiblemente debido a la accesibilidad y costo en contextos rurales. Las categorías menos comunes podrían indicar falta de especialistas autorizados o preferencia por auto-asistencia.

Los intervalos de confianza, derivados de quantiles bootstrap, proporcionan una medida no paramétrica de incertidumbre, útil en datos posiblemente no normales. Las distribuciones bootstrap confirman que las estimaciones son robustas.

En el contexto de estadística espacial, estos resultados podrían extenderse a mapeo geográfico de asistencias por distrito, usando las variables de código de distrito incluidas.