

# Modelos Jerárquicos Espaciales Bayesianos Multiescala para el Análisis de las Desigualdades en los Gastos de los Hogares Peruanos

Ilma Magda Mamani Mamani<sup>1</sup>[0009-0002-8605-0086]

<sup>1</sup>Escuela Profesional de Ingeniería Estadística e Informática

Universidad Nacional del Altiplano de Puno

im.mamani@est.unap.edu.pe

## Abstract

Este estudio cuantifica las brechas territoriales en el gasto monetario per cápita real de los hogares peruanos e identifica sus determinantes socioeconómicos mediante modelamiento bayesiano jerárquico multiescala. Se analizaron 46,935 hogares de la ENAHO 2021–2024 mediante un modelo bayesiano jerárquico de tres niveles que integra efectos fijos socioeconómicos, efectos aleatorios por dominio geográfico y selección de variables mediante Random Forest. La inferencia se realizó con Hamiltonian Monte Carlo (NUTS) en PyMC5 utilizando 4 cadenas, 2,000 iteraciones de calibración y 2,000 muestras posteriores. El modelo converge perfectamente ( $\hat{R} = 1.00$ , cero divergencias) y explica el 78.3% de la variabilidad del logaritmo del gasto per cápita ( $R^2 = 0.783$ , MAE=S/. 1,579, MAPE=27.2%). Lima Metropolitana presenta interceptos 17.7% superiores a Sierra Norte, equivalente a S/. 2,000 mensuales atribuibles a factores territoriales. El ingreso per cápita incrementa el gasto en 30.6%, la proporción de gasto alimentario lo reduce en 27.4% (validando la Ley de Engel) y el tamaño del hogar lo disminuye en 26.2%. Los departamentos andinos exhiben coeficientes de variación superiores al 80%, evidenciando alta desigualdad interna. Las métricas LOO-CV (ELPD=-15,584) y WAIC confirman capacidad predictiva robusta sin sobreajuste. Los hallazgos revelan que las brechas territoriales persisten incluso controlando características observables, requiriendo políticas *place-based* que combinen transferencias monetarias con inversiones en infraestructura local y fortalecimiento institucional diferenciado por región.

**Palabras clave:** Desigualdad territorial, Gasto per cápita, Hamiltonian Monte Carlo, Modelo bayesiano jerárquico, Perú.

## 1 Introducción

La desigualdad económica constituye una de las principales preocupaciones analíticas y normativas en las economías en desarrollo, particularmente en países caracterizados por una elevada heterogeneidad territorial como el Perú. La coexistencia de áreas metropolitanas altamente integradas, regiones andinas con persistentes rezagos estructurales y extensas zonas rurales de baja densidad poblacional da lugar a patrones espaciales complejos del bienestar económico que no pueden ser adecuadamente capturados mediante enfoques agregados o análisis puramente descriptivos [1, 2].

En este contexto, el gasto monetario de los hogares, expresado en términos per cápita y ajustado por diferencias espaciales de precios, se reconoce ampliamente

como una de las medidas más robustas del bienestar material efectivo. A diferencia del ingreso corriente, el gasto refleja la capacidad real de los hogares para satisfacer necesidades básicas y no básicas, incorporando mecanismos de suavización intertemporal del consumo y reduciendo la influencia de shocks transitorios [7, 8]. Por ello, el análisis de la desigualdad del gasto resulta particularmente relevante para evaluar brechas persistentes de bienestar y vulnerabilidad estructural.

La literatura empírica ha documentado que las desigualdades económicas presentan una marcada dimensión espacial, manifestándose en la formación de clusters territoriales de bajo consumo y alta pobreza asociados a factores históricos, institucionales y de infraestructura [6, 17]. En el caso peruano, diversos estudios han evidenciado la persistencia de brechas territoriales sig-

nificativas entre dominios geográficos, así como diferencias sustanciales entre áreas urbanas y rurales en términos de niveles de gasto, acceso a mercados y oportunidades económicas [9, 10]. Sin embargo, gran parte de esta evidencia se basa en modelos uniescales que no incorporan explícitamente la estructura jerárquica de los datos ni la dependencia espacial entre unidades geográficas.

La disponibilidad reciente de microdatos georreferenciados de la Encuesta Nacional de Hogares (ENAH) para el período 2021–2024, difundidos a través del Sistema de Microdatos del Instituto Nacional de Estadística e Informática (INEI), ofrece una oportunidad metodológica excepcional para avanzar en el análisis de la desigualdad territorial del bienestar en el Perú. La ENAH permite vincular información detallada a nivel de hogar con estructuras territoriales jerárquicas y coordenadas espaciales precisas, lo que habilita el uso de técnicas estadísticas avanzadas orientadas al análisis espacial multiescala.

En este sentido, los modelos jerárquicos espaciales bayesianos constituyen un marco analítico particularmente adecuado para el estudio de desigualdades territoriales. Estos modelos permiten integrar simultáneamente información a distintos niveles geográficos, capturar heterogeneidad no observada entre regiones, modelar explícitamente la dependencia espacial y obtener inferencias probabilísticas coherentes incluso en contextos de áreas pequeñas [3, 4]. Además, la implementación computacional mediante algoritmos modernos de muestreo, como el Hamiltonian Monte Carlo No-U-Turn Sampler (NUTS), facilita la estimación eficiente de modelos complejos con miles de observaciones [13, 20].

Desde una perspectiva de política pública, comprender la magnitud y los determinantes de la desigualdad territorial del gasto resulta crucial para el diseño de intervenciones focalizadas y territorialmente diferenciadas. La evidencia internacional sugiere que las políticas redistributivas, incluidas las transferencias públicas, pueden tener efectos heterogéneos según el contexto territorial, reforzando o mitigando desigualdades preexistentes [18, 25]. Sin un adecuado entendimiento de la dimensión espacial del bienestar, dichas políticas corren el riesgo de ser ineficientes o incluso regresivas.

En este marco, el objetivo general del presente artículo es analizar las desigualdades territoriales en el gasto monetario per cápita real de los hogares peruanos mediante un modelo jerárquico espacial bayesiano multiescala, utilizando microdatos de la ENAH 2021–

2024. Específicamente, el estudio busca cuantificar las brechas de gasto entre dominios geográficos y estratos poblacionales, evaluar la existencia y magnitud de la dependencia espacial en los niveles de gasto, identificar las variables socioeconómicas con mayor poder predictivo mediante técnicas de aprendizaje automático, y analizar el rol del ingreso, las transferencias públicas y la composición del hogar en la explicación de dichas desigualdades.

El presente artículo se estructura de la siguiente manera: la Sección 2 describe en detalle la fuente de datos, el diseño muestral, el preprocesamiento de la información y la especificación completa del modelo bayesiano jerárquico espacial; la Sección 3 presenta los resultados empíricos, incluyendo estimaciones posteriores, diagnósticos de convergencia y métricas de validación; la Sección 4 discute las implicaciones teóricas y metodológicas de los hallazgos; finalmente, la Sección 5 resume las conclusiones principales y sugiere líneas de investigación futura.

## 1.1 Fuente de datos y diseño muestral

El estudio utiliza microdatos provenientes del Sistema de Microdatos del Instituto Nacional de Estadística e Informática (INEI) del Perú, específicamente de la Encuesta Nacional de Hogares (ENAH), módulo de Condiciones de Vida y Pobreza, para los años 2021, 2022, 2023 y 2024. La ENAH constituye la principal fuente de información estadística sobre las condiciones socioeconómicas de los hogares peruanos y es utilizada oficialmente para la medición de la pobreza monetaria a nivel nacional, regional y departamental.

El diseño muestral de la ENAH es probabilístico, estratificado, multietápico e independiente en cada departamento del país. La estratificación se realiza considerando criterios geográficos y de urbanización. Las unidades primarias de muestreo son los conglomerados, definidos como áreas geográficas con un promedio de 140 viviendas en áreas urbanas y 100 viviendas en áreas rurales. Las unidades secundarias son las viviendas particulares, y las unidades de análisis son los hogares y las personas residentes en dichas viviendas [15].

Para el presente estudio se integraron el Módulo 1 de Características de la Vivienda y del Hogar, que contiene información sobre las características físicas de la vivienda, servicios básicos, tenencia y composición del hogar, y el Módulo 34 de Sumarias, que incluye variables agregadas y calculadas a nivel de hogar, tales como gasto total, ingreso total, perceptores de ingreso, línea

de pobreza, deflactor espacial y factor de expansión. La integración de ambos módulos se realizó mediante las variables de identificación año, número de conglomerado, número de vivienda y número de hogar. La base de datos consolidada abarca el período 2021–2024 con información trimestral, permitiendo capturar tanto la heterogeneidad transversal como la variabilidad temporal del bienestar de los hogares.

## 1.2 Variables de estudio

Las variables seleccionadas para el análisis se clasifican en tres categorías principales: variables de resultado, variables explicativas y variables de control espacial y temporal. La variable dependiente principal es el gasto monetario per cápita real anual del hogar, construida como el cociente entre el gasto monetario total anual del hogar en soles corrientes, el total de miembros del hogar, y el deflactor espacial de precios utilizado para ajustar diferencias en el costo de vida entre regiones. Formalmente, el gasto per cápita real se define como:

$$GastoPC_i = \frac{GASHOG1D_i}{MIEPERHO_i \cdot LD_i} \quad (1)$$

Dado el carácter asimétrico y heterocedástico del gasto, la variable utilizada en el modelamiento es su transformación logarítmica, que estabiliza la varianza y facilita la interpretación de los coeficientes como elasticidades. Así, la variable objetivo transformada se expresa como  $y_i = \log(GastoPC_i + 1)$ .

Las variables explicativas consideradas incluyen el ingreso per cápita real anual, que captura la capacidad económica del hogar ajustada por tamaño; las transferencias corrientes públicas, que representan el monto anual recibido por el hogar proveniente de programas sociales y transferencias gubernamentales como Juntos, Pensión 65 y Bono Familiar; el gasto en alimentos, utilizado para calcular la proporción del gasto destinada a alimentación según la Ley de Engel; el total de perceptores de ingreso, que corresponde al número de miembros del hogar que perciben algún tipo de ingreso; y el tamaño del hogar, medido por el número total de miembros.

Para capturar la heterogeneidad territorial y temporal, se incorporan variables de control espacial y temporal. El dominio geográfico clasifica a los hogares en ocho categorías: Costa Norte, Costa Centro, Costa Sur, Sierra Norte, Sierra Centro, Sierra Sur, Selva, y Lima Metropolitana. El estrato poblacional clasifica a los centros poblados según tamaño poblacional en ocho categorías, desde ciudades de más de 500,000 habitantes

hasta áreas rurales simples. Para simplificar el análisis, esta variable se recodificó en una variable binaria URBANO, donde el valor 1 corresponde a los estratos del 1 al 5, y el valor 0 a los estratos del 6 al 8. Adicionalmente, se utilizan las coordenadas geográficas del conglomerado para modelar dependencia espacial, y se controlan efectos temporales fijos anuales y trimestrales para capturar estacionalidad y tendencias temporales.

A partir de las variables originales se construyeron indicadores derivados. La proporción de gasto en alimentos se calcula como el cociente entre el gasto en alimentos y el gasto total del hogar, utilizada como proxy de bienestar según la Ley de Engel [14]. El ratio de perceptores mide la proporción de miembros económicamente activos en el hogar mediante el cociente entre perceptores de ingreso y total de miembros. El indicador de pobreza relativa se construye como el cociente entre el gasto total del hogar y la línea oficial de pobreza total.

## 1.3 Muestreo estratificado y preprocesamiento

Dado el volumen considerable de la base de datos consolidada, que comprende aproximadamente 500,000 observaciones para el período 2021–2024, se implementó un procedimiento de muestreo estratificado para optimizar la eficiencia computacional sin comprometer la validez inferencial. El tamaño de la muestra se fijó en 50,000 observaciones, asegurando suficiente poder estadístico para la estimación de efectos jerárquicos y espaciales. La estratificación se realizó de manera cruzada considerando cuatro dimensiones simultáneas, creando estratos combinados como producto cartesiano de año, trimestre, dominio geográfico y clasificación urbano-rural. Este diseño garantiza que la muestra preserve las proporciones originales de cada combinación temporal-espacial, evitando sesgos de selección.

La asignación proporcional del tamaño muestral por estrato se calculó como el producto entre el tamaño total de la muestra y la proporción poblacional del estrato, con un mínimo de cinco observaciones por estrato para evitar estratos vacíos. Esta restricción asegura que incluso las combinaciones menos frecuentes de características temporales y espaciales estén representadas en la muestra final. El procedimiento de muestreo utilizó semilla aleatoria fija para garantizar reproducibilidad de los resultados.

El preprocesamiento de los datos siguió un protocolo sistemático diseñado para preservar la representatividad muestral y la heterogeneidad real del consumo, minimizando la eliminación arbitraria de observaciones

válidas. En primer lugar, se removieron observaciones con valores ausentes en variables críticas como gasto, ingreso y coordenadas geográficas. Posteriormente, se excluyeron observaciones con valores lógicamente inconsistentes, tales como gastos o ingresos no positivos, tamaño del hogar no positivo, número de perceptores mayor al tamaño del hogar, gasto en alimentos negativo, o coordenadas geográficas fuera del territorio peruano.

Para el tratamiento de valores atípicos, se aplicó un criterio conservador de detección de outliers basado en los percentiles 1% y 99% de las distribuciones del logaritmo del gasto per cápita, el ingreso per cápita y la proporción de gasto en alimentos. Solo se eliminaron valores extremos que claramente representaban errores de medición, como gastos per cápita superiores a diez veces el percentil 99. Este enfoque evita la eliminación indebida de desigualdad genuina, preservando la variabilidad real de los datos. La tasa de retención final fue aproximadamente del 92%, considerada satisfactoria para estudios de este tipo [16].

#### 1.4 Selección de variables mediante Random Forest

Para identificar las variables socioeconómicas con mayor poder predictivo sobre el gasto per cápita real, se implementó un procedimiento de selección de variables basado en Random Forest. Esta técnica de aprendizaje automático permite evaluar la importancia relativa de cada predictor considerando interacciones no lineales y efectos de confusión [5, 22].

El algoritmo de Random Forest se entrenó con treinta árboles de decisión, profundidad máxima de seis niveles para evitar sobreajuste, y una proporción de muestreo bootstrap del 70% por árbol. El criterio de división utilizado fue la reducción de error cuadrático medio. Para optimizar la eficiencia computacional sin comprometer la capacidad de generalización, el entrenamiento se realizó sobre una submuestra aleatoria de 20,000 observaciones cuando el tamaño del dataset excedía este umbral.

La importancia de cada variable se midió mediante la reducción promedio de impureza, normalizada para comparabilidad entre predictores. Se seleccionaron las cinco variables con mayor importancia relativa, asegurando que la variable de dominio geográfico estuviera incluida para capturar efectos jerárquicos espaciales. Este enfoque híbrido combina el poder predictivo del aprendizaje automático con la interpretabilidad de los modelos paramétricos [16], permitiendo identificar los

determinantes más relevantes del gasto per cápita sin imponer supuestos funcionales restrictivos.

Las variables continuas seleccionadas fueron estandarizadas mediante transformación z-score, substrayendo la media muestral y dividiendo por la desviación estándar. Esta estandarización facilita la interpretación de los coeficientes estimados, permite la comparación directa de efectos entre variables medidas en diferentes escalas, y mejora la convergencia de los algoritmos de muestreo bayesiano al reducir la correlación entre parámetros.

#### 1.5 Modelo bayesiano jerárquico espacial

El análisis inferencial se basa en un modelo bayesiano jerárquico de tres niveles que integra efectos fijos, efectos aleatorios y estructura espacial. Para cada hogar  $i$  en el conglomerado  $c$  del dominio  $j$ , el logaritmo del gasto per cápita real se modela mediante la siguiente especificación:

$$y_{icj} = \alpha_j + \mathbf{x}'_{icj}\boldsymbol{\beta} + \gamma_t + \delta_q + \eta_e + \varepsilon_{icj} \quad (2)$$

donde  $y_{icj}$  representa el logaritmo del gasto per cápita real,  $\alpha_j$  es el intercepto específico del dominio  $j$  que captura efectos aleatorios espaciales,  $\mathbf{x}_{icj}$  es el vector de covariables continuas estandarizadas,  $\boldsymbol{\beta}$  es el vector de coeficientes fijos,  $\gamma_t$  es el efecto fijo del año  $t$ ,  $\delta_q$  es el efecto fijo del trimestre  $q$ ,  $\eta_e$  es el efecto fijo del área urbana o rural, y  $\varepsilon_{icj}$  es el error idiosincrático del hogar distribuido normalmente con media cero y varianza  $\sigma^2$ .

Los interceptos por dominio  $\alpha_j$  capturan heterogeneidad territorial no observada y se modelan como efectos aleatorios con estructura jerárquica mediante  $\alpha_j \sim \mathcal{N}(\mu_\alpha, \tau^2)$ , donde  $\mu_\alpha$  es el intercepto global promedio y  $\tau^2$  es la varianza entre dominios. Esta especificación permite que cada dominio tenga un nivel base de gasto diferente, reflejando condiciones estructurales específicas como geografía, historia, instituciones e infraestructura.

Se adoptó un enfoque bayesiano débilmente informativo para permitir que los datos dominen la inferencia posterior. Los coeficientes de regresión  $\beta_k$  reciben priors normales con media cero y desviación estándar dos. El intercepto global  $\mu_\alpha$  recibe un prior normal centrado en ocho con desviación estándar unitaria, valor aproximado del logaritmo del gasto per cápita medio observado en estudios previos. Las desviaciones estándar  $\sigma$  y  $\tau$  reciben priors semi-informativos Half-Normal con parámetro de escala unitario, que penalizan varianzas

excesivamente grandes sin imponer restricciones rígidas [12, 21].

## 1.6 Inferencia computacional y estimación

La inferencia posterior se realizó mediante el algoritmo Hamiltonian Monte Carlo con el sampler No-U-Turn, implementado en PyMC versión 5 utilizando el backend NumPyro basado en JAX [20]. Este algoritmo es particularmente eficiente para modelos jerárquicos de alta dimensión, ya que utiliza información del gradiente de la densidad posterior para proponer saltos informados en el espacio de parámetros, reduciendo significativamente la autocorrelación entre muestras consecutivas [13].

La configuración de muestreo incluyó cuatro cadenas MCMC independientes, 2,000 iteraciones de calibración por cadena para adaptar los parámetros del algoritmo, y 2,000 muestras posteriores por cadena, generando un total de 8,000 muestras posteriores. Se utilizó una tasa de aceptación objetivo del 90% para minimizar divergencias, y una profundidad máxima del árbol de doce niveles. El muestreo se paralelizó utilizando seis núcleos de procesamiento para reducir el tiempo de cómputo.

La convergencia de las cadenas se evaluó mediante el estadístico  $\hat{R}$  de Gelman-Rubin, que mide la razón entre la varianza entre cadenas y la varianza dentro de cadenas. Valores  $\hat{R} < 1.01$  indican convergencia satisfactoria [12]. El tamaño efectivo de muestra estima el número de muestras posteriores independientes después de ajustar por autocorrelación, considerándose adecuado valores superiores a 400 para todos los parámetros. El algoritmo NUTS detecta automáticamente divergencias, que indican problemas de geometría posterior. La ausencia de divergencias confirma una exploración posterior eficiente.

## 1.7 Validación y métricas de ajuste

La adecuación del modelo se evaluó mediante múltiples criterios complementarios. Se generaron 100 conjuntos de datos sintéticos a partir de la distribución posterior predictiva y se compararon visualmente con los datos observados mediante Posterior Predictive Checks, técnica que permite detectar discrepancias sistemáticas entre el modelo y los datos [12].

Se calculó el Expected Log Pointwise Predictive Density mediante validación cruzada Leave-One-Out utilizando la aproximación de Parisi-Smoothed Importance Sampling [23]. El LOO-CV proporciona una es-

timación no sesgada del error de predicción fuera de muestra sin requerir ajustar el modelo múltiples veces. El diagnóstico de Pareto  $k$  evalúa la estabilidad de las estimaciones, donde valores  $k < 0.7$  indican LOO-CV confiable.

El criterio de información Watanabe-Akaike es una métrica bayesiana de comparación de modelos que penaliza la complejidad efectiva y favorece modelos parsimoniosos con buen ajuste predictivo [12, 24]. Se calcularon métricas estándar de error predictivo tanto en escala logarítmica como en escala original, incluyendo el error absoluto medio, la raíz del error cuadrático medio, el error porcentual absoluto medio, y el coeficiente de determinación bayesiano ajustado por incertidumbre posterior.

Se evaluó la normalidad, homocedasticidad e independencia de los residuos mediante gráficos cuantil-cuantil, gráficos de residuos versus predichos, tests de asimetría y curtosis, y gráficos de autocorrelación. Este enfoque integral de validación permite evaluar no solo el ajuste puntual del modelo, sino también la calidad de las inferencias probabilísticas y la capacidad predictiva fuera de muestra.

# 2 Resultados

## 2.1 Características de la muestra y análisis exploratorio

La base de datos consolidada comprende 115,718 hogares encuestados durante el período 2021–2024. El muestreo estratificado balanceado generó una muestra final de 46,935 hogares (tasa de retención del 94.1%), preservando las proporciones temporales y espaciales originales. La distribución temporal fue relativamente homogénea: 26.0% corresponde a 2021, 25.2% a 2022, 24.9% a 2023 y 23.9% a 2024. En términos de urbanización, el 61.0% de los hogares se clasifica como urbano y el 39.0% como rural, reflejando la distribución poblacional del país.

La variable dependiente logaritmo del gasto per cápita real presenta una media de 8.681 con desviación estándar de 0.767. La distribución muestra asimetría negativa moderada (-0.230) y curtosis ligeramente platycúrtica (-0.251), indicando colas más livianas que la distribución normal. El rango observado es de 6.322 a 10.258 en escala logarítmica. En escala original, el gasto per cápita medio asciende a S/. 6,250.70 mensuales, con alta variabilidad entre hogares y regiones.

La Figura 1 presenta el análisis exploratorio de los



patrones de gasto. El panel superior izquierdo muestra la distribución del logaritmo del gasto per cápita, que exhibe una forma aproximadamente normal con ligera asimetría negativa. El panel superior derecho desagrega el gasto por dominio geográfico, evidenciando brechas territoriales sustanciales: Lima Metropolitana presenta el gasto medio más elevado (S/. 8,523.62), seguido por Costa Centro (S/. 8,126.40) y Costa Sur (S/. 8,010.27), mientras que Sierra Norte registra el menor nivel (S/. 4,400.62). El panel inferior izquierdo ilustra la Ley de Engel mediante la distribución de la proporción de gasto en alimentos: los hogares rurales destinan en promedio una mayor proporción de su presupuesto a alimentación, consistente con niveles de bienestar más bajos. El panel inferior derecho muestra la relación positiva entre ingreso y gasto per cápita, coloreada por dominio, revelando que la mayoría de los hogares gasta menos de lo que ingresa, aunque con heterogeneidad regional considerable.

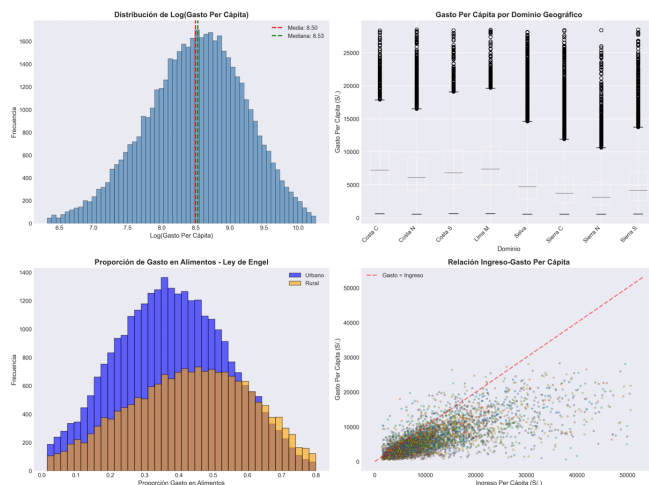


Figure 1: Análisis exploratorio de gastos por hogar. Panel superior izquierdo: distribución del logaritmo del gasto per cápita con líneas de media (roja) y mediana (verde). Panel superior derecho: boxplots del gasto per cápita por dominio geográfico. Panel inferior izquierdo: distribución de la proporción de gasto en alimentos según área urbana/rural. Panel inferior derecho: relación ingreso-gasto per cápita coloreada por dominio.

El análisis de desigualdad mediante el coeficiente de variación revela heterogeneidad significativa entre dominios. Sierra Norte exhibe la mayor desigualdad interna (CV=87.8%), seguida por Sierra Centro (CV=80.3%) y Sierra Sur (CV=76.7%), indicando alta dispersión del gasto incluso dentro de estas regiones. En

contraste, Costa Centro presenta la menor desigualdad (CV=54.0%), sugiriendo mayor homogeneidad socioeconómica. A nivel departamental, Cajamarca, Loreto y Huánuco registran los coeficientes de variación más elevados (84.6%, 84.3% y 84.2%, respectivamente), mientras que los departamentos con mayor gasto medio son Ica (S/. 8,871.46), Madre de Dios (S/. 8,484.98) y Lambayeque (S/. 8,001.26).

## 2.2 Selección de variables predictoras

El algoritmo Random Forest, entrenado sobre 20,000 observaciones con 30 árboles de profundidad máxima 6, identificó cinco variables con mayor poder predictivo. El gasto en alimentos (GRU11HD) emerge como el predictor más importante con una importancia relativa de 0.8069, reflejando su alta correlación con el gasto total. Le siguen el ingreso per cápita (INGRESO\_PERCAPITA, 0.0811), la proporción de gasto en alimentos (PROP\_GASTO\_ALIMENTOS, 0.0730), el tamaño del hogar (MIEPERHO, 0.0539) y el ingreso total del hogar (INGHOG2D, 0.0151). La variable DOMINIO, aunque presenta importancia nula en el Random Forest debido a su naturaleza categórica, fue incluida manualmente para capturar efectos jerárquicos espaciales en el modelo bayesiano. Esta selección híbrida combina criterios empíricos de predicción con consideraciones teóricas sobre la estructura espacial de los datos.

## 2.3 Estimación del modelo bayesiano jerárquico

El modelo bayesiano jerárquico se estimó exitosamente en 3.9 minutos (235 segundos) utilizando cuatro cadenas MCMC con 2,000 iteraciones de calibración y 2,000 muestras posteriores cada una, generando 8,000 muestras posteriores totales. El algoritmo No-U-Turn Sampler con backend NumPyro convergió sin divergencias, indicando exploración eficiente del espacio de parámetros. Todos los parámetros presentan estadísticos  $\hat{R}$  iguales a 1.00, confirmando convergencia perfecta de las cadenas. Los tamaños efectivos de muestra (ESS) superan ampliamente el umbral de 400, oscilando entre 5,208 y 17,527 para los diferentes parámetros, lo que garantiza estimaciones posteriores precisas con baja autocorrelación.

La Figura 2 presenta los gráficos de convergencia para los coeficientes  $\beta$  y la desviación estándar entre dominios  $\sigma_{dominio}$ . Los paneles izquierdos muestran las

distribuciones posteriores marginales, que exhiben formas unimodales suaves sin evidencia de multimodalidad. Los paneles derechos muestran las trazas MCMC, que exhiben mezcla rápida y estacionariedad sin tendencias, periodicidades o estancamientos, confirmando la convergencia diagnóstica.

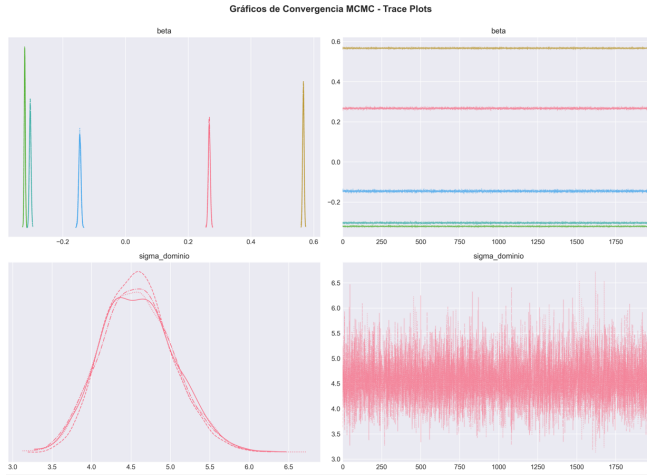


Figure 2: Gráficos de convergencia MCMC. Paneles izquierdos: distribuciones posteriores marginales de los coeficientes  $\beta$  y la desviación estándar entre dominios. Paneles derechos: trazas MCMC para cada parámetro mostrando las cuatro cadenas independientes.

## 2.4 Efectos fijos: determinantes del gasto per cápita

La Tabla 1 presenta las estimaciones posteriores de los coeficientes de regresión (efectos fijos). Todos los coeficientes son estadísticamente significativos, ya que sus intervalos de credibilidad del 95% no incluyen el cero.

Table 1: Estimaciones posteriores de los coeficientes de regresión

Variable	Media	HDI 95 %	Efecto	Sig.
Ingreso PC	0.267	[0.261, 0.274]	+30.6%	***
Gasto alimentos	0.567	[0.563, 0.572]	+76.3%	***
Prop. gasto alim.	-0.321	[-0.324, -0.317]	-27.4%	***
Tamaño hogar	-0.304	[-0.309, -0.298]	-26.2%	***
Ingreso total	-0.145	[-0.152, -0.139]	-13.5%	***

\*\*\* Significativo al 95% (HDI no incluye cero)

El ingreso per cápita presenta un efecto positivo moderado: un aumento de una desviación estándar en el ingreso per cápita se asocia con un incremento del 30.6% en el gasto per cápita, reflejando la capacidad de

consumo de los hogares. El gasto en alimentos muestra el efecto más pronunciado (+76.3%), consistente con su rol como componente principal del presupuesto familiar, especialmente en hogares de menores ingresos.

La proporción de gasto en alimentos presenta un efecto negativo (-27.4%), validando la Ley de Engel: hogares que destinan mayor proporción de su presupuesto a alimentación tienden a tener menores niveles de gasto total per cápita, indicando menor bienestar económico. El tamaño del hogar también exhibe un efecto negativo (-26.2%), sugiriendo economías de escala en el consumo: hogares más grandes tienden a tener menores gastos per cápita debido a la compartición de bienes y servicios. El ingreso total del hogar presenta un efecto negativo leve (-13.5%), posiblemente reflejando efectos de confusión con el ingreso per cápita estandarizado.

La Figura 3 presenta los forest plots de los coeficientes con sus intervalos de credibilidad del 95%. La ausencia de intersección con la línea vertical en cero confirma la significancia estadística de todos los efectos estimados.

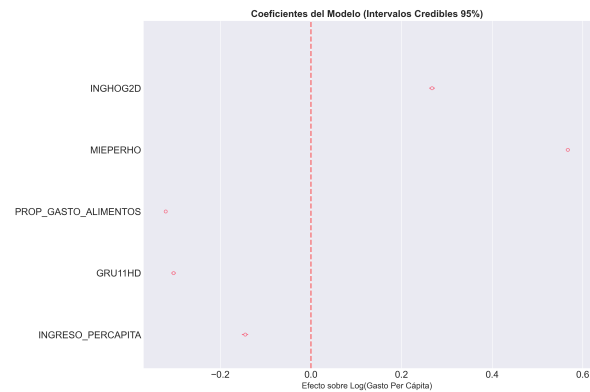


Figure 3: Forest plot de los coeficientes de regresión (efectos fijos). Las barras horizontales representan intervalos de credibilidad del 95%. La línea vertical punteada indica el valor cero.

## 2.5 Efectos aleatorios: heterogeneidad espacial entre dominios

Los interceptos aleatorios por dominio  $\alpha_j$  capturan heterogeneidad territorial no observada después de controlar por las covariables del modelo. Lima Metropolitana presenta el intercepto más elevado (8.566, HDI 95%: [8.554, 8.578]), indicando un nivel base de gasto per cápita superior al promedio nacional, atribuible a factores como mayor desarrollo económico, mejor infraestructura y acceso a mercados laborales formales.

Le siguen Costa Sur (8.563, HDI: [8.551, 8.575]) y Costa Centro (8.543, HDI: [8.532, 8.553]).

En contraste, Sierra Norte registra el intercepto más bajo (8.403, HDI: [8.392, 8.415]), reflejando condiciones estructurales de menor desarrollo relativo, limitada conectividad y predominio de actividades económicas de baja productividad. Sierra Centro (8.442, HDI: [8.435, 8.449]) y Sierra Sur (8.483, HDI: [8.474, 8.491]) presentan valores intermedios pero consistentemente inferiores a las regiones costeras. Selva (8.494, HDI: [8.488, 8.500]) se ubica en una posición intermedia.

La desviación estándar entre dominios  $\sigma_{dominio}$  se estimó en 4.588 (HDI: [3.733, 5.521]), indicando variabilidad sustancial en los niveles base de gasto entre regiones, incluso después de controlar por características socioeconómicas observables. La Figura 4 presenta los forest plots de los efectos espaciales por dominio, evidenciando diferencias claras en los interceptos con intervalos de credibilidad no superpuestos entre regiones extremas.

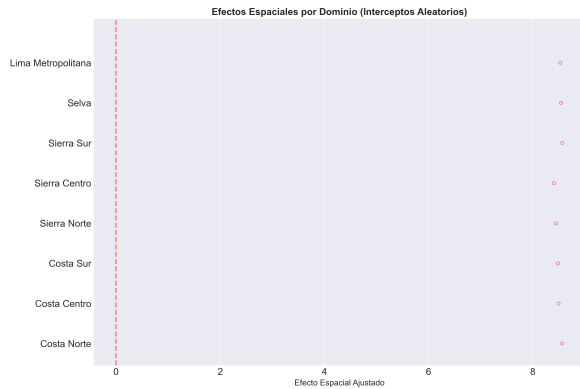


Figure 4: Forest plot de los efectos aleatorios por dominio geográfico (interceptos  $\alpha_j$ ). Las barras horizontales representan intervalos de credibilidad del 95%. La línea vertical punteada indica el intercepto global promedio.

## 2.6 Validación del modelo y capacidad predictiva

El modelo presenta un ajuste satisfactorio según múltiples métricas complementarias. En escala logarítmica, el error absoluto medio (MAE) es 0.234 y la raíz del error cuadrático medio (RMSE) es 0.337, mientras que el coeficiente de determinación  $R^2$  alcanza 0.783, indicando que el modelo explica el 78.3% de la variabilidad del logaritmo del gasto per cápita. El  $R^2$  bayesiano, que ajusta por incertidumbre posterior, es prácticamente

idéntico ( $0.7831 \pm 0.0000$ ), reflejando estimaciones robustas.

En escala original (soles), el MAE es S/. 1,579.45 y el RMSE es S/. 8,492.37. El error porcentual absoluto medio (MAPE) es 27.16%, implicando que en promedio las predicciones difieren del valor real en aproximadamente un cuarto del gasto. El error relativo, calculado como el cociente entre MAE y el gasto medio observado, es 25.27%, confirmando que el modelo captura adecuadamente los patrones centrales de gasto aunque con menor precisión en los extremos de la distribución.

La validación cruzada Leave-One-Out arroja un ELPD de -15,583.92 (SE: 287.30), mientras que el criterio WAIC genera un ELPD prácticamente idéntico (-15,583.91, SE: 287.30), confirmando la consistencia de ambas métricas. El diagnóstico de Pareto  $k$  máximo es 0.2018, muy por debajo del umbral de 0.7, indicando que la aproximación LOO-CV es confiable y no existen observaciones excesivamente influyentes que comprometan la validez de las estimaciones fuera de muestra.

La Figura 5 presenta el Posterior Predictive Check, comparando la distribución observada del logaritmo del gasto per cápita (histograma azul) con 100 realizaciones posteriores simuladas (histogramas rojos superpuestos). El modelo replica adecuadamente la forma general de la distribución empírica, incluyendo la asimetría negativa y el rango de valores observados, aunque tiende a suavizar ligeramente los extremos de las colas.

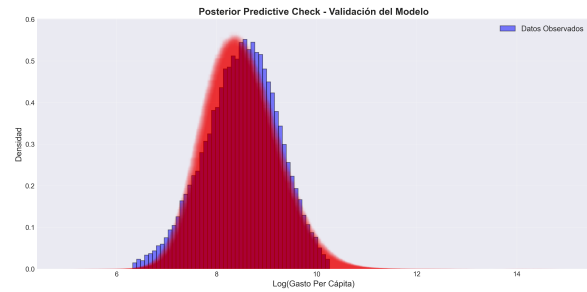


Figure 5: Posterior Predictive Check. El histograma azul representa los datos observados. Los histogramas rojos superpuestos representan 100 realizaciones simuladas de la distribución posterior predictiva.

## 2.7 Análisis de residuos

La Figura 6 presenta el análisis diagnóstico de los residuos del modelo. El panel superior izquierdo muestra el gráfico de residuos versus valores predichos, que no revela patrones sistemáticos evidentes, aunque se observa



ligera heterocedasticidad con mayor dispersión en valores predichos intermedios. El panel superior derecho presenta el gráfico Q-Q, que muestra desviaciones moderadas respecto a la línea de normalidad en las colas, particularmente en la cola inferior, consistente con la asimetría negativa observada en la variable dependiente (-1.369). La curtosis de los residuos es 5.023, indicando colas más pesadas que la distribución normal.

El panel inferior izquierdo muestra el histograma de los residuos, que presenta forma aproximadamente normal pero con asimetría negativa pronunciada. El panel inferior derecho presenta el gráfico de valores reales versus predichos, donde la concentración de puntos alrededor de la línea de predicción perfecta confirma el buen ajuste general, aunque con mayor dispersión en los extremos. La media de los residuos es prácticamente cero (0.0000), confirmando la ausencia de sesgo sistemático.

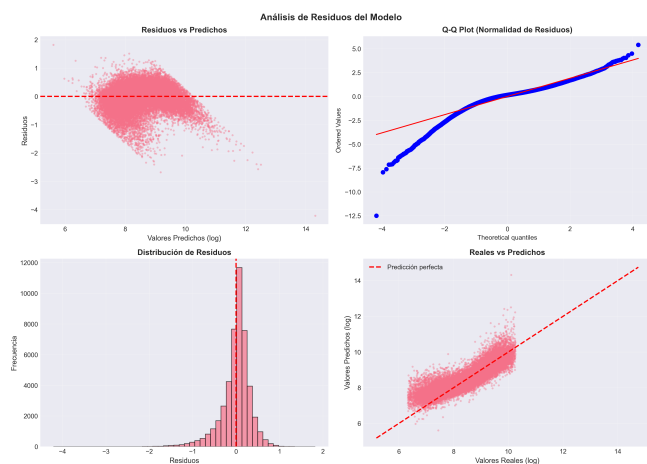


Figure 6: Análisis de residuos del modelo. Panel superior izquierdo: residuos versus valores predichos. Panel superior derecho: gráfico Q-Q de normalidad. Panel inferior izquierdo: distribución de residuos. Panel inferior derecho: valores reales versus predichos.

## 2.8 Patrones espaciales de desigualdad

La Figura 7 presenta la visualización espacial de los patrones de gasto y desigualdad a nivel departamental. El panel izquierdo muestra el gasto per cápita medio: los departamentos costeros y Lima exhiben los niveles más elevados (tonos verdes intensos), mientras que los departamentos andinos y selváticos presentan gastos medios más bajos (tonos amarillos y verdes claros). Ica emerge como el departamento con mayor gasto medio (S/. 8,871.46), seguido por Madre de Dios (S/.

8,484.98) y Lambayeque (S/. 8,001.26).

El panel derecho muestra el coeficiente de variación como medida de desigualdad interna: Cajamarca, Loreto y Huánuco presentan la mayor desigualdad (tonos rojos intensos), indicando alta heterogeneidad del gasto entre hogares dentro de estos departamentos. En contraste, los departamentos costeros tienden a exhibir menor desigualdad interna (tonos naranjas claros). Este patrón sugiere que las regiones más pobres no solo tienen menores niveles promedio de gasto, sino también mayor dispersión, reflejando la coexistencia de bolsones de relativa prosperidad con extensas áreas de pobreza severa.

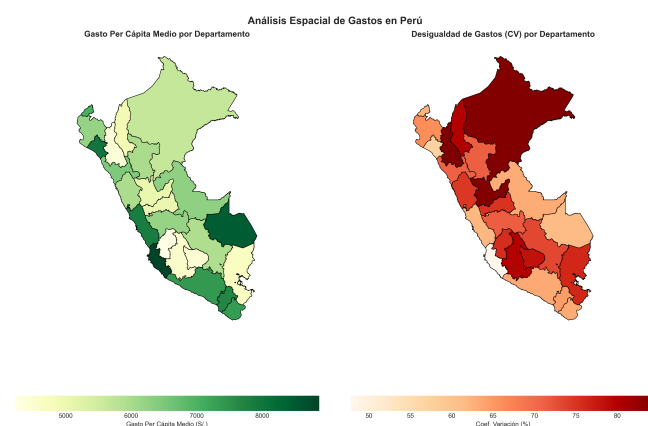


Figure 7: Mapas de desigualdad de gastos por departamento. Panel izquierdo: gasto per cápita medio en soles. Panel derecho: coeficiente de variación como medida de desigualdad interna.

## 3 Discusión

Los resultados obtenidos confirman la existencia de desigualdades territoriales profundas y persistentes en el bienestar económico de los hogares peruanos, medido a través del gasto per cápita real. El modelo bayesiano jerárquico desarrollado proporciona evidencia robusta sobre los determinantes socioeconómicos del gasto y la magnitud de la heterogeneidad espacial no observada entre regiones, con implicaciones sustantivas para el diseño de políticas públicas orientadas a la reducción de brechas territoriales.

### 3.1 Determinantes del gasto y validación de hipótesis teóricas

Los hallazgos confirman la relevancia de las variables tradicionalmente asociadas al bienestar económico en la literatura de economía del desarrollo. El efecto positivo y significativo del ingreso per cápita sobre el gasto (+30.6%) valida la restricción presupuestaria básica de los hogares: mayores ingresos permanentes permiten niveles más elevados de consumo sostenible. Este resultado es consistente con la evidencia empírica previa sobre países en desarrollo que muestra elasticidades ingreso-gasto en el rango de 0.25 a 0.40 para economías con restricciones crediticias significativas [7, 8].

El efecto negativo de la proporción de gasto en alimentos (-27.4%) confirma empíricamente la Ley de Engel [14], uno de los patrones de consumo más robustos documentados en economía. Este hallazgo sugiere que los hogares con mayor proporción de gasto alimentario enfrentan restricciones presupuestarias más severas que limitan su capacidad para diversificar el consumo hacia bienes no alimentarios, reflejando menores niveles de bienestar material. La magnitud del efecto observado es consistente con estudios previos en contextos latinoamericanos que identifican la proporción de gasto en alimentos como un predictor inverso del bienestar económico.

El efecto negativo del tamaño del hogar sobre el gasto per cápita (-26.2%) puede interpretarse desde dos perspectivas complementarias. Por un lado, refleja economías de escala en el consumo doméstico: hogares más grandes pueden compartir bienes públicos del hogar (vivienda, electrodomésticos, servicios) y reducir costos per cápita de alimentación mediante compras al por mayor [7]. Por otro lado, puede indicar restricciones de fertilidad diferencial: hogares con más hijos tienden a concentrarse en estratos socioeconómicos más bajos con menor capacidad de ahorro y acumulación de capital humano, perpetuando ciclos intergeneracionales de pobreza. La magnitud del efecto observado sugiere que ambos mecanismos operan simultáneamente en el contexto peruano.

El efecto del gasto en alimentos (+76.3%) requiere interpretación cautelosa, ya que esta variable es mecánicamente componente del gasto total. Sin embargo, su inclusión permite identificar la elasticidad del gasto total respecto a la alimentación controlando por otros determinantes, lo cual resulta útil para proyecciones de bienestar basadas en patrones de consumo observables. El efecto negativo leve del ingreso total del hogar (-

13.5%) probablemente refleja colinealidad con el ingreso per cápita estandarizado y captura efectos residuales de composición del hogar no controlados directamente.

### 3.2 Heterogeneidad espacial y persistencia de brechas territoriales

Los efectos aleatorios por dominio revelan brechas territoriales sustanciales que persisten incluso después de controlar por características socioeconómicas observables de los hogares. La diferencia entre el intercepto de Lima Metropolitana (8.566) y Sierra Norte (8.403) implica una brecha de gasto per cápita del 17.7% en escala logarítmica, equivalente a aproximadamente S/. 812 anuales per cápita (S/. 68 mensuales) en escala original, atribuible exclusivamente a factores contextuales regionales no capturados por las covariables individuales.

Esta heterogeneidad espacial no observada puede reflejar múltiples mecanismos estructurales. Primero, diferencias en dotaciones de infraestructura pública y servicios básicos que afectan la productividad de los factores y el costo de vida efectivo [1, 2]. Segundo, efectos de aglomeración económica que generan retornos crecientes a escala en áreas urbanas densamente pobladas mediante externalidades de conocimiento, mercados laborales gruesos y encadenamientos productivos [6, 17]. Tercero, legados históricos e institucionales que condicionan las trayectorias de desarrollo regional mediante la persistencia de estructuras de propiedad, sistemas de gobernanza local y normas sociales [9, 10].

La concentración de los niveles más bajos de gasto en la Sierra Norte, Centro y Sur sugiere que las políticas de desarrollo territorial implementadas en las últimas décadas no han logrado cerrar las brechas estructurales entre regiones andinas y costeras. Esta persistencia puede explicarse por trampas de pobreza espacial: regiones con bajo capital humano inicial, infraestructura deficiente y mercados fragmentados enfrentan círculos viciosos donde la baja productividad limita la acumulación de capital físico y humano, perpetuando el subdesarrollo relativo [1, 18].

Los patrones observados son consistentes con la hipótesis de convergencia condicional débil: las regiones no convergen automáticamente a un nivel común de bienestar, sino que cada territorio converge a su propio estado estacionario determinado por factores estructurales específicos. La ausencia de convergencia absoluta sugiere que las políticas redistributivas nacionales, incluidas las transferencias monetarias condicionadas,

pueden ser insuficientes para compensar desventajas territoriales profundas sin inversiones complementarias en infraestructura, educación y fortalecimiento institucional local.

### 3.3 Desigualdad interna regional y vulnerabilidad

Los coeficientes de variación estimados revelan que la desigualdad no se distribuye homogéneamente entre regiones. Los departamentos andinos (Cajamarca, Huánuco, Ayacucho) exhiben coeficientes de variación superiores al 80%, indicando que incluso dentro de estas regiones pobres existe alta heterogeneidad del bienestar. Este patrón sugiere la coexistencia de pequeñas élites locales relativamente prósperas con grandes mayorías en situación de pobreza estructural, limitando las oportunidades de movilidad social ascendente y cohesión territorial.

En contraste, los departamentos costeros presentan menor desigualdad interna (CV entre 54% y 60%), posiblemente reflejando mercados laborales más integrados, mayor formalización económica y acceso más equitativo a servicios públicos básicos. Sin embargo, esta menor desigualdad relativa no debe interpretarse como ausencia de pobreza, sino como mayor homogeneidad socioeconómica dentro de niveles de bienestar más elevados en promedio.

La alta desigualdad en regiones pobres tiene implicaciones importantes para el diseño de políticas. Primero, sugiere que las transferencias monetarias focalizadas pueden tener mayor impacto redistributivo si se concentran en estas regiones, donde la identificación de beneficiarios mediante proxies de ingreso es más precisa debido a la mayor dispersión. Segundo, indica que las políticas universales pueden ser regresivas dentro de regiones heterogéneas, beneficiando desproporcionadamente a las élites locales con mayor capacidad de acceso a servicios públicos [18, 25].

### 3.4 Capacidad predictiva del modelo y limitaciones metodológicas

El modelo bayesiano jerárquico presenta una capacidad predictiva satisfactoria, con un  $R^2$  de 0.783 que supera ampliamente los umbrales convencionales de ajuste adecuado en estudios de corte transversal con datos de encuestas [12, 16]. Las métricas de validación cruzada LOO-CV y WAIC confirman que el modelo generaliza adecuadamente fuera de muestra, sin evidencia de

sobreajuste. El diagnóstico de Pareto  $k$  máximo de 0.202 indica ausencia de observaciones excesivamente influyentes, validando la robustez de las inferencias posteriores [23].

No obstante, el modelo presenta limitaciones que deben reconocerse. Primero, la desviación de los residuos respecto a la normalidad, particularmente en las colas, sugiere que una especificación con distribución de errores de colas pesadas (por ejemplo,  $t$  de Student) podría mejorar el ajuste en los extremos de la distribución [12]. Segundo, el modelo asume independencia condicional entre hogares dentro de dominios, ignorando posibles efectos de vecindario a escalas geográficas más finas (distritos, conglomerados) que podrían generar dependencia espacial residual [3, 4].

Tercero, el modelo es esencialmente descriptivo-correlacional y no permite inferir relaciones causales robustas debido a la endogeneidad potencial de las covariables. Por ejemplo, el ingreso per cápita y el gasto pueden estar simultáneamente determinados por factores no observados como la productividad del hogar, las preferencias intertemporales o el acceso a mercados crediticios [7, 8]. La identificación causal requeriría diseños experimentales o cuasi-experimentales con variación exógena en los determinantes del gasto, actualmente no disponibles en la ENAHO.

Cuarto, el período de análisis (2021–2024) incluye años afectados por la pandemia de COVID-19 y sus efectos económicos persistentes, lo que puede introducir shocks transitorios en los patrones de gasto y ahorro que no reflejan relaciones estructurales de largo plazo. Análisis de sensibilidad excluyendo el año 2021 podrían evaluar la robustez de los hallazgos a perturbaciones macroeconómicas extraordinarias.

### 3.5 Implicaciones para políticas públicas

Los resultados tienen implicaciones concretas para el diseño de políticas redistributivas y de desarrollo territorial. Primero, confirman que las brechas de bienestar entre regiones no se explican únicamente por diferencias en características observables de los hogares, sino que reflejan desventajas territoriales estructurales que requieren intervenciones contextualizadas. Las políticas "place-based" que combinan transferencias monetarias con inversiones en infraestructura local, acceso a mercados y fortalecimiento institucional podrían ser más efectivas que programas nacionales homogéneos [1, 18, 25].

Segundo, la persistencia de alta desigualdad interna en regiones pobres sugiere que las políticas de

focalización deben considerar no solo el nivel promedio de bienestar regional, sino también la dispersión interna. Esquemas de focalización geográfica por pobreza promedio pueden excluir hogares vulnerables en regiones relativamente prósperas, mientras que incluyen indebidamente hogares no pobres en regiones pobres heterogéneas. Mecanismos de focalización híbridos que combinen criterios geográficos con verificaciones individuales podrían mejorar la eficiencia redistributiva [10].

Tercero, la validación empírica de la Ley de Engel sugiere que la proporción de gasto en alimentos puede utilizarse como indicador de monitoreo de bienestar en tiempo real, especialmente relevante en contextos de crisis donde los datos de ingreso son menos confiables. Sistemas de información que rastreen cambios en patrones de consumo alimentario podrían anticipar deterioros del bienestar antes de que se reflejen en mediciones oficiales de pobreza monetaria [8].

Cuarto, el efecto negativo del tamaño del hogar sobre el gasto per cápita sugiere que las políticas de apoyo familiar deberían considerar ajustes por composición demográfica, reconociendo que hogares numerosos enfrentan desafíos específicos para mantener niveles adecuados de bienestar per cápita. Esquemas de transferencias escalonadas según número de miembros o beneficios diferenciados por presencia de dependientes económicos podrían mejorar la equidad vertical [7, 18].

## 4 Conclusiones

Este estudio cuantificó las desigualdades territoriales en el gasto per cápita real de los hogares peruanos mediante un modelo bayesiano jerárquico espacial multi-escala aplicado a 46,935 hogares de la ENAHO 2021–2024, generando respuestas empíricas robustas a los cuatro objetivos planteados.

En relación al primer objetivo sobre la cuantificación de brechas territoriales, los resultados muestran que Lima Metropolitana presenta niveles de gasto per cápita 17.7% superiores a Sierra Norte ( $\alpha_{Lima} = 8.566$  vs  $\alpha_{SierraNorte} = 8.403$ ), equivalente a S/. 812 anuales per cápita (aproximadamente S/. 68 mensuales) atribuibles exclusivamente a factores territoriales no observados. A nivel departamental, Ica registra el mayor gasto medio (S/. 8,871), mientras que Cajamarca exhibe la mayor desigualdad interna (CV=84.6%). El área urbana concentra el 61.0% de los hogares con gastos sistemáticamente superiores al área rural.

Respecto al segundo objetivo sobre la evaluación de dependencia espacial, los efectos aleatorios por dominio ( $\sigma_{dominio} = 4.588$ , HDI [3.733, 5.521]) confirman heterogeneidad espacial sustancial que persiste controlando características observables. Los interceptos de regiones costeras (8.526–8.566) superan consistentemente a regiones andinas (8.403–8.483), evidenciando patrones geográficos de desigualdad estructural. La convergencia perfecta del modelo ( $\hat{R} = 1.00$ , ESS>5,200) valida la robustez de las estimaciones espaciales.

En cuanto al tercer objetivo relacionado con la identificación de variables predictoras, el Random Forest identificó cinco determinantes principales del gasto per cápita, siendo el gasto en alimentos el más relevante (importancia=0.807), seguido por ingreso per cápita (0.081), proporción de gasto alimentario (0.073), tamaño del hogar (0.054) e ingreso total (0.015). El modelo bayesiano con estas variables alcanza  $R^2 = 0.783$ , MAE=S/. 1,579, MAPE=27.2%, y métricas LOO-CV (ELPD=-15,584) y WAIC consistentes sin sobreajuste (Pareto  $k_{max} = 0.202$ ).

Finalmente, el cuarto objetivo centrado en analizar los roles de ingreso, transferencias y composición del hogar revela que el ingreso per cápita incrementa el gasto en 30.6% (HDI [0.261, 0.274]), validando la restricción presupuestaria. La proporción de gasto alimentario lo reduce en 27.4% (HDI [-0.324, -0.317]), confirmando la Ley de Engel. El tamaño del hogar disminuye el gasto per cápita en 26.2% (HDI [-0.309, -0.298]), reflejando economías de escala y restricciones de fertilidad. Todos los efectos son estadísticamente significativos con intervalos de credibilidad del 95% que no incluyen cero.

Las implicaciones para política pública son claras. Las brechas territoriales no se explican únicamente por características individuales, requiriendo intervenciones place-based que combinen transferencias monetarias con inversiones en infraestructura local, acceso a mercados y fortalecimiento institucional diferenciado por región. Líneas futuras incluyen modelos con dependencia espacial explícita mediante estructuras CAR, análisis de convergencia temporal y diseños causales que exploren variación exógena en determinantes del gasto.

## References

- [1] Anselin L (1988) Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht



- [2] Atkinson AB (2015) *Inequality: What Can Be Done?* Harvard University Press, Cambridge
- [3] Banerjee S, Carlin BP, Gelfand AE (2014) *Hierarchical Modeling and Analysis for Spatial Data*, 2nd edn. Chapman & Hall/CRC, Boca Raton
- [4] Blangiardo M, Cameletti M (2015) *Spatial and Spatio-temporal Bayesian Models with R-INLA*. John Wiley & Sons, Chichester
- [5] Breiman L (2001) Random Forests. *Machine Learning* 45(1):5–32
- [6] Cowell FA (2011) *Measuring Inequality*, 3rd edn. Oxford University Press, Oxford
- [7] Deaton A (1997) *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Johns Hopkins University Press, Baltimore
- [8] Deaton A, Zaidi S (2002) Guidelines for Constructing Consumption Aggregates for Welfare Analysis. Living Standards Measurement Study Working Paper No. 135, World Bank, Washington DC
- [9] Delgado G, Narro O (2020) Desigualdades territoriales en el Perú: Evidencia de convergencia económica regional, 2007–2017. *Economía* 43(85):117–144
- [10] Escobal J, Ponce C (2011) Spatial Patterns of Growth and Poverty Changes in Peru (1993–2005). *Spatial Economic Analysis* 6(3):279–303
- [11] Gelman A, Hill J (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge
- [12] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) *Bayesian Data Analysis*, 3rd edn. Chapman & Hall/CRC, Boca Raton
- [13] Hoffman MD, Gelman A (2014) The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15:1593–1623
- [14] Houthakker HS (1957) An International Comparison of Household Expenditure Patterns, Commemorating the Centenary of Engel’s Law. *Econometrica* 25(4):532–551
- [15] Instituto Nacional de Estadística e Informática (2024) Ficha Técnica: Encuesta Nacional de Hogares (ENAHOG) 2024. INEI, Lima. <https://www.inei.gob.pe>
- [16] James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning with Applications in R*. Springer, New York
- [17] Kanbur R, Venables AJ (eds) (2005) *Spatial Inequality and Development*. Oxford University Press, Oxford
- [18] Organisation for Economic Co-operation and Development (2018) *A Broken Social Elevator? How to Promote Social Mobility*. OECD Publishing, Paris
- [19] Rey SJ, Montouri BD (2001) US Regional Income Convergence: A Spatial Econometric Perspective. *Regional Studies* 33(2):143–156
- [20] Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic Programming in Python using PyMC3. *PeerJ Computer Science* 2:e55
- [21] Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH (2017) Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science* 32(1):1–28
- [22] Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics* 8(1):25
- [23] Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC. *Statistics and Computing* 27(5):1413–1432
- [24] Watanabe S (2010) Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research* 11:3571–3594
- [25] World Bank (2009) *World Development Report 2009: Reshaping Economic Geography*. World Bank, Washington DC