

UNIVERSIDAD NACIONAL DE INGENIERÍA

**FACULTAD DE INGENIERÍA ECONÓMICA, ESTADÍSTICA Y CIENCIAS
SOCIALES**



TESIS

ESTIMACIÓN DE LA PREVALENCIA DE HIDATIDOSIS HUMANA CONSIDERANDO LA DISTRIBUCIÓN ESPACIAL EN EL CENTRO POBLADO DE CORPACANCHA, JUNÍN - PERÚ

PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERO ESTADÍSTICO

Elaborado por:

BACH. OSWALDO GABRIEL ERNESTO ESPINOZA HURTADO

Asesor:

MAG. CHRISTIAN AMAO SUXO

Lima, Junio, 2022

PERÚ

DEDICATORIA

A cada una de las personas que me acompañaron en este trayecto; tanto familiares, como amistades, profesores y mentores. A quienes siguen y a la memoria de quienes ya no están; en especial a mi abuelo, el maestro mecánico Luis Ernesto Hurtado Chiscul (Q.E.P.D.).

AGRADECIMIENTOS

Al profesor y asesor externo en esta investigación PhD. Erick Chacón, quien me asesoró y guió en la formulación y elaboración de la presente investigación.

Al profesor y asesor MSc. Christian Amao, quien me dio las herramientas básicas para introducirme en la redacción científica y aportó de forma crítica en la redacción de esta investigación.

Al equipo de investigación en Hidatidosis conformado por el PhD. MD. Saul Santivañez, la MSc. Karina Bardales, el BVetMed. Cesar Sedano y el BVetMed. Raul Enriquez, quienes me dieron asesoría temática y proporcionaron la base de datos en coordinación con área del Centro de Datos del Centro de Estudios Clínicos del Centro de Salud Global de la Universidad Peruana Cayetano Heredia.

Al fondo EULAC, el cual por medio de CONCYTEC (PRO CIENCIA) financió la presente investigación.

RESUMEN

ABSTRACT

PRÓLOGO

ÍNDICE

INTRODUCCIÓN	10
I. PLANTEAMIENTO DEL PROBLEMA	11
1.1. Descripción de la situación problemática	11
1.2. Formulación del problema	13
1.2.1. Problema General	13
1.2.2. Problemas Específicos	13
1.3. Objetivos de la investigación	14
1.3.1. Objetivos General	14
1.3.2. Objetivos Específicos	14
1.4. Justificación, alcances y limitaciones de la investigación	15
II. MARCO TEÓRICO CONCEPTUAL	16
2.1. Antecedentes	16
2.2. Bases Teóricas	17
2.2.1. Bases Teóricas Generales	17
2.2.2. Bases Teóricas Estadísticas	18
2.3. Enfoque teórico conceptual asumido por el investigador	26
2.3.1. Determinación del vector de pesos	27
2.3.2. Estimación de la prevalencia	30

2.3.3.	Factor de corrección en estudios de casos y controles	30
2.3.4.	Consideraciones para la simulación	32
2.4.	Marco legal	34
2.5.	Hipótesis	34
2.5.1.	Hipótesis General	34
2.5.2.	Hipótesis Específicas	34
III.	MARCO METODOLÓGICO	35
3.1.	Tipo, nivel y diseño de la investigación	35
3.2.	Población, muestra y tamaño de muestra	36
3.3.	Técnicas de análisis e instrumentos	37
3.3.1.	Análisis de datos	37
3.3.2.	Instrumentos de medición	38
3.4.	Cuadro de operacionalización de variables	40
3.5.	Matriz de consistencia	40
IV.	ANÁLISIS Y RESULTADOS	42
4.1.	Análisis Exploratorio	42
4.1.1.	Análisis de los factores de riesgo	47
4.1.2.	Análisis de la intensidad	51
4.2.	Simulaciones	55
4.3.	Modelamiento	58
4.3.1.	Intensidad	58
4.3.2.	Riesgo	60
4.4.	Inferencia	61
4.5.	Implementación	63

V. CONCLUSIONES Y RECOMENDACIONES	67
5.1. Conclusiones	67
5.2. Recomendaciones	69
REFERENCIA BIBLIOGRÁFICA	70
Anexos	73
A. Aspectos legales	74
B. Códigos empleados para el desarrollo de la tesis	75

Índice de tablas

2.1. Funciones para estimar una proporción considerando densidad poblacional, riesgo espacial y el tipo de muestreo	23
2.2. Estudio Caso-control	24
2.3. Caso-control con <i>ser seleccionado en la muestra</i> como exposición	30
3.1. Cuadro de operacionalización de variables	40
3.2. Matriz de consistencia	41
4.1. Factores de riesgo y prevalencia (n=141).	47
4.2. Factores de riesgo por resultado de la campaña (n=141).	48
4.3. Odd Ratio de la variable sexo	49
4.4. Odd Ratio de la variable edad	49
4.5. Odd Ratio de la variable perro	50
4.6. Resultados de la simulación	55
4.7. Métricas de ajuste para los modelos	60

Índice de figuras

3.1. Mapa urbano del centro poblado de Corpacancha (Junín, Perú)	39
4.1. Distribución en el espacio de las casas en el centro poblado de Corpacancha. . .	43
4.2. Distribución en el espacio de las casas habitadas en el centro poblado de Corpacancha.	44
4.3. Distribución de las casas con al menos un habitante que haya participado en la campaña realizada en el centro poblado de Corpacancha.	45
4.4. Prevalencia de Hidatidosis Humana en Corpacancha.	46
4.5. Prevalencia e intervalo de confianza por sexo.	48
4.6. Prevalencia e intervalo de confianza por tenencia de perros.	50
4.7. Intensidad poblacional	51
4.8. Intensidad muestral	52
4.9. Riesgo de participar en la muestra	53
4.10. Riesgo de tener la enfermedad	54
4.11. Densidad del parámetro de los escenarios 1 al 9	56
4.12. Densidad del parámetro de los escenarios 10 al 18	57
4.13. Residuales en la población	59
4.14. Residuales en la muestra	59
4.15. Prevalencia de Hidatidosis Humana en Corpacancha, añadiendo la prevalencia corregida.	61

4.16. Riesgo corregido de tener la enfermedad	62
4.17. Riesgo de participar en la muestra en Canchayllo	64
4.18. Riesgo de la enfermedad en la muestra en Canchayllo	65
4.19. Riesgo de la enfermedad en la muestra en Canchayllo	66
1.1. Carta de Aprobación Ética del Estudio Madre	74
2.1. QR enlazado al repositorio en GITHUB que contiene los códigos trabajados . .	75

INTRODUCCIÓN

El manejo de la información es de carácter fundamental en la toma de decisiones. Por lo que es de vital importancia que dicha información sea confiable. Pues no basta con solo poder obtener el valor de un indicador, sino que este mismo permita reflejar el valor real en la población en la que fue medido. Más aún si el objetivo de estos indicadores es el de determinar el impacto generado por una enfermedad en una población; ya que este no se limita solo al dinero utilizado en su tratamiento, sino también engloba otro tipo de gastos, como la pérdida en la esperanza y calidad de vida que conlleva padecerla. Por lo tanto, es necesario estimar indicadores como la prevalencia de forma insesgada. Por lo que de haber un sesgo de selección, es necesario establecer los pasos para corregirlo.

Partiendo de lo mencionado con anterioridad, la presente investigación se centró en el estudio de un método para la corrección de sesgo espacial que pueda ser aplicado en un caso real y que sirva como punto de partida en su implementación. Para esto, se empezó planteando la problemática de la situación, así como los objetivos de la investigación. Luego, se detalló en el Marco Teórico los elementos teóricos-conceptuales empleados en el desarrollo de la investigación. Después, se presentó el Marco Metodológico con el que se buscó responder a los objetivos de la investigación. Finalmente, se realizó el análisis correspondiente para las 2 poblaciones especificadas y se obtuvieron los resultados necesarios para la formulación de conclusiones acorde a las hipótesis presentadas.

Capítulo I

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la situación problemática

El impacto generado por una zoonosis (enfermedad transmitida de animales a humanos) en la economía de una sociedad consiste no solo en el dinero invertido para su tratamiento y prevención, o de las pérdidas que esta causa en las actividades ganaderas y agrícolas. El efecto que tiene sobre las personas también incluye la discapacidad y el cambio de estilo de vida que esta y su tratamiento conlleva (Shaw, Rushton, Roth, y Torgerson, 2017). Un caso de estas enfermedades es la hidatidosis; la cual se da como consecuencia de una infección parasitaria que desemboca en la formación de quiste focalizados en determinados órganos como el hígado, pulmón, riñón, entre otros. En el plano internacional, se han encontrado niveles de prevalencia (porcentaje de individuos positivos a una enfermedad) del 8.1 % y 13 % de hidatidosis en hígado y pulmón, respectivamente, en zonas ganaderas del norte de Etiopía (Kebede, Hagos, Girma, y Lobago, 2009). De manera análoga en la región de Sudamérica, durante los últimos 40 años se han encontrado niveles que varían entre 1.6 % y 14.2 % por región (Larrieu, Belloto, ARAMBULO III, y Tamayo, 2004). Estudios realizados con anterioridad en centros poblados ubicados en áreas rurales de Junín han determinado niveles de prevalencia entre 5 % y 18 % (Santivañez

y cols., 2010), convirtiendo a este departamento en una zona altamente endémica para esta zoonosis. Pese a esto, el efecto causado por esta enfermedad se encuentra infravalorado por la misma población (P. L. Moro y cols., 2011), lo cual complica al desarrollo de las investigaciones relacionadas a esta durante la recolección de datos.

La Organización Panamericana de la Salud (OPS), respecto a la recolección de datos de la Hidatidosis en humanos, menciona dentro de las fuentes oficiales de información a las encuestas poblacionales que utilizan ultrasonido (PAHO/WHO, 2017). De acuerdo a diferentes autores, estas suelen darse a modo de un muestreo por conveniencia mediante una campaña de despistaje gratuita en el centro de salud del centro poblado que se está estudiando. Esto implica un sesgo de selección debido a que la información recopilada suele encontrarse acumulada en quienes viven en las zonas aledañas a los centros de salud o en ciertas zonas específicas de la comunidad. Asimismo, esto permite la presencia de un potencial error en la estimación de la prevalencia de hidatidosis que podría estar subestimando o sobreestimando su valor real.

La presente investigación se centró en plantear una metodología que considere la distribución espacial de los individuos en el centro poblado respecto al centro de salud (o lugar en donde se haya realizado la campaña) en la estimación de la prevalencia buscando así corregir el problema del sesgo por medio de un factor de corrección. Para ello, se partió del supuesto de que la probabilidad de que un individuo de la población pertenezca a la muestra (esto es, haya sido enrolado en el estudio) siendo esta heterogénea y dependa de la ubicación de su vivienda.

1.2. Formulación del problema

Al estar planteando una metodología, surgió la necesidad de saber si esta es eficaz en comparación a la tradicional. Por esto, es necesario que la investigación pueda solucionar los siguientes problemas.

1.2.1. Problema General

¿Se puede emplear un método de estimación de prevalencia considerando la distribución espacial para reducir el sesgo del muestreo?

1.2.2. Problemas Específicos

- ¿Cómo se diferencian los resultados obtenidos de las estimaciones de la prevalencia mediante el método clásico y el error de estimación del método propuesto?
- ¿Cuál es el método adecuado de acuerdo a las condiciones de la investigación?
- ¿De cuánto es la prevalencia de la hidatidosis humana en los centros poblados de Corpacanca y Canchayllo?

1.3. Objetivos de la investigación

1.3.1. Objetivos General

Estudiar un método de estimación de prevalencia considerando la distribución espacial para reducir el sesgo del muestreo.

1.3.2. Objetivos Específicos

- Comparar los resultados obtenidos de las estimaciones de la prevalencia por el método tradicional y el método propuesto.
- Determinar el método adecuado de acuerdo a las condiciones de la investigación.
- Estimar la prevalencia de la hidatidosis humana en los centros poblados de Corpacancha y Canchayllo.

1.4. Justificación, alcances y limitaciones de la investigación

La justificación de la presente investigación radica en poder estimar correctamente desde un marco estadístico la prevalencia de la hidatidosis con el fin de poder contar con un indicador más preciso y que puede ser empleado en la toma de decisiones. De igual forma, el alcance se encuentra en presentar una metodología que nos permita determinar eficientemente la prevalencia de hidatidosis y así contribuir en el trabajo que tiene la Organización Mundial de la Salud (OMS) respecto a validar estrategias eficaces frente al control de la hidatidosis (WHO, 2019). Por otro lado, es necesario hacer mención que la presente investigación se centra netamente en estimar la prevalencia, mas no en los factores asociados a esta. Además, la metodología propuesta está limitada de acuerdo a los supuestos que esta plantea conforme a las condiciones en las que la información fue recolectada. Finalmente, al contar con datos reales, la investigación se encuentran sujeta a los principios éticos para estudios médicos que incluyen sujetos humanos (general assembly of the world medical association y cols., 2014).

Capítulo II

MARCO TEÓRICO CONCEPTUAL

2.1. Antecedentes

Estudios realizados en China han considerado la heterogeneidad de la probabilidad que tiene cada individuo en ser seleccionado para una muestra al momento de la estimación del número de pacientes que padecen cierta enfermedad, tratando la información obtenida bajo un enfoque bayesiano (Bailly, Daurès, Dunais, y Pradier, 2015). De igual forma, el considerar la distribución espacial de los individuos capturados en una determinada muestra se ha visto empleado en el campo de la ecología (Royle, Chandler, Sollmann, y Gardner, 2014); lo cual puede llegar a ser replicado en el campo de las ciencias epidemiológicas (Braeye y cols., 2016). Finalmente, es necesario mencionar que estudios anteriores realizados en la zona de Junin solo han empleado métodos tradicionales para la estimación de la prevalencia y factores asociados (Santivañez y cols., 2010).

2.2. Bases Teóricas

2.2.1. Bases Teóricas Generales

Hidatidosis Humana

La echinococcosis quística o hidatidosis humana (también conocida solo como *Hidatidosis*) es una zoonosis (enfermedad transmitida de animales a humanos) parasitaria incluida en la lista de enfermedades tropicales desatendidas por la OMS (Sarkar y cols., 2017). Es causada por el parásito *Echinococcus granulosus* en su estadio larvareo (Giri y Parija, 2012). Este parásito requiere dos hospederos mamíferos para completar su ciclo de vida: El ganado ovino (ovejas), quien es el hospedero intermediario, y el perro, como el hospedero definitivo. Este último adquiere el parásito alimentándose de vísceras de oveja infectadas con el parásito. El hombre podría considerarse como un hospedero intermediario *accidental* causado por el consumo de alimentos o agua contaminada con huevos presentes en las heces de los perros. Una vez adquirido el parásito, estos huevos eclosionan en el intestino delgado del hombre, en donde pasan a la circulación venosa hasta alojarse en el hígado (principal localización), pulmón (segunda localización de importancia) u otra víscera o tejido, donde se formará el quiste hidatídico (P. Moro y Schantz, 2009). Diversos autores concluyen que tanto la edad, el sexo y la tenencia de perros en el hogar, en otras variables, son factores de riesgo para la enfermedad (Santivañez y cols., 2010). Una de las formas en las que se diagnostica esta enfermedad es por medio de la prueba serológica de Western blot. Dicha prueba cuenta con una sensibilidad y especificidad para Hidatidosis del 94 % y 100 %, respectivamente (Davelois, Escalante, y Jara, 2016).

2.2.2. Bases Teóricas Estadísticas

Estimación de una proporción poblacional

Si seleccionamos una muestra aleatoria de tamaño n , la proporción muestral \hat{p} es la fracción de elementos de la muestra que poseen la característica de interés; para nuestro caso, esta será que el paciente sea positivo frente a tener la enfermedad. De esta forma, podemos decir que \hat{p} se puede estimar de la siguiente manera (Scheaffer, Mendenhall, Ott, y Gerow, 2011):

$$\hat{p} = \frac{n_+}{n} \quad (2.1)$$

Donde:

n_+ : Número de casos positivos en la muestra

n : Tamaño de la muestra

Si a los valores de la muestra se les asigna un valor numérico (Positivo: $X = 1$, Negativo: $X = 0$), entonces la estimación de \hat{p} puede ser expresada de la siguiente manera:

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.2)$$

Cuando n es lo suficientemente grande, el intervalo de confianza (a un determinado α) para la estimación se puede determinar de la siguiente manera:

$$\hat{p} \pm Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1} \left(\frac{N-n}{N} \right)}$$

Donde:

N : Tamaño poblacional

n : Tamaño de la muestra

$Z_{1-\alpha/2}$: Valor de $1 - \alpha/2$ en la distribución normal

Potencia estadística

Se entiende como potencia estadística a la probabilidad que la hipótesis alterna H_1 sea no rechazada cuando es verdadera. El valor para esta probabilidad es matemáticamente el complemento del error tipo II (β). Además, se encuentra comunmente relacionado a factores vinculados con la muestra y su recolección, como su tamaño (Ellis, 2010).

$$\text{Potencia estadística} = P(\text{Rechazar la hipótesis nula} | \text{La hipótesis es falsa}) = 1 - \beta$$

Sesgo

Al ser desconocido un parámetro θ , no es posible conocer hasta qué punto un estimador $\hat{\theta}$ se encuentra lejos o cerca de su verdadero valor. El error que aparece al tomar como verdadero al valor del estimador $\hat{\theta}$ es la diferencia entre $\theta - \hat{\theta}$. Si se pudiera obtener todas las posibles muestras y para cada una de ellas su correspondiente estimación, entonces una métrica para evaluar el error vendría a ser el error cuadrático medio (ECM):

$$ECM(\hat{\theta}) = E(\theta - \hat{\theta})^2 \quad (2.3)$$

Efectuando y ordenando:

$$ECM(\hat{\theta}) = V(\hat{\theta}) + [\theta - E(\hat{\theta})]^2$$

En donde la expresión $\theta - E(\hat{\theta})$ será denominada como sesgo (Ruiz-Maya y Pliego, 2004). Adicionalmente, es necesario hacer mención que desde un punto de vista epidemiológico, el sesgo puede ser entendido como un error sistemático en el diseño de un estudio que resulta en un error en la estimación. Cuando el error es producido durante el proceso de selección de los individuos de la muestra, recibe el nombre de sesgo de selección (Celentano, Mhs, y Szklo, 2019).

Procesos Puntuales Espaciales

Un proceso puntual espacial es un proceso estocástico cuyas realizaciones consisten en un conjunto finito o numerablemente infinito de puntos en el plano. En caso de tratarse de un proceso puntual espacial finito, se trata de una colección $\mathbf{X} = \{X_1, X_2, \dots\}$ de puntos en el espacio R^2 con un número finito de puntos en una región A . Los puntos en este tipo de procesos son medidos mediante la función de conteo $N(A)$ (número de puntos por unidad cuadrada) o la función de intensidad $\lambda(X)$ (densidad media de puntos) (Baddeley, Rubak, y Turner, 2015).

Procesos de Poisson

Entre los procesos puntuales espaciales, los más empleados son los procesos de Poisson. Estos son agrupados de acuerdo a la naturaleza de su función de intensidad $\lambda(X)$ (Baddeley y cols., 2015).

1. **Proceso de Poisson Homogéneo** Considerando una región arbitraria A y regiones disjuntas A_i para $i = 1, \dots, k$, entonces:

- i) $N(A)$ tiene una distribución de Poisson con media $\lambda|A|$, donde $|A|$ representa el área de la región A ($\lambda|A|$, función de intensidad, constante)
- ii) $N(A_1), \dots, N(A_k)$ son variables aleatorias independientes para todo k y A_i , lo que es conocido como Aleatoriedad Espacial Completa (CSR)

2. **Proceso de Poisson No-Homogéneo** Similar al caso anterior (Proceso de Poisson Homogéneo) con la diferencia de que la función de intensidad $\lambda(X)$ varía acorde a la ubicación de x .

- i) $N(A)$ tiene una distribución de Poisson con media $\int_A \lambda(x)dx$
- ii) $N(A_1), \dots, N(A_k)$ son CSR

3. **Proceso de Poisson Doblemente Estocástico** Proceso de Poisson No-Homogéneo con función de intensidad $\wedge(X)$ aleatoria.

i) $\{\wedge(X) = \lambda(X) : X \in R^2\}$ es un Proceso No-negativo Estocástico

ii) La condicional en $\{\wedge(X) = \lambda(X) : X \in R^2\}$ es un Proceso de Poisson No-Homogéneo con intensidad $\lambda(X)$

Procesos de Poisson Marcados

A la secuencia de pares $(X_1, Y_1), (X_2, Y_2), \dots$ se le conoce como un proceso de Poisson marcado cuando cada variable aleatoria Y_i está asociadas a cada X_i realización de un proceso de Poisson con función de intensidad $\lambda(X)$ (Pinsky y Karlin, 2010). En el caso que puntos marcados tengan como posibles valores de Y_i 1 y 0; la probabilidad para cada uno será:

$$P(Y_i = 1) = \theta$$

$$P(Y_i = 0) = 1 - \theta$$

Número esperado de puntos

En un proceso de Poisson el número esperado de puntos en el espacio $X = (x, y)$ está definido como el valor obtenido por medio de la integral de la función de intensidad $\lambda(X)$ en todo su dominio.

$$\int \int_X \lambda(x, y) dx dy$$

Suponiendo que la intensidad no depende de su ubicación en el espacio, entonces el valor de $\lambda(x, y)$ será una constante (λ). Considerando en base a este un proceso de Poisson marcado en el que θ varia espacialmente, entonces el número esperado de puntos *marcados* con el valor 1 ($Y_i = 1$) en el espacio $X = (x, y)$ está definido como el valor obtenido por medio de la integral

de la función de intensidad de los puntos marcados $\lambda_\theta(X)$ en todo su dominio. Esta función de intensidad $\lambda_\theta(X)$ es el producto de $\lambda(X)$ con la función espacial de θ ($f_\theta(x, y)$).

$$\int \int_X \lambda_\theta(x, y) dx dy = \int \int_X \lambda(x, y) f_\theta(x, y) dx dy$$

En caso que θ no dependa de su ubicación en el espacio, entonces este mantendrá su valor constante. Suponiendo que del proceso de Poisson marcado previamente planteado se obtiene una muestra en la que cada individuo i tiene una propabilidad π_i de ser seleccionado y esta depende de su ubicación espacial, entonces el número esperado de puntos *marcados* con el valor 1 ($Y_i = 1$) en la muestra en el espacio $X = (x, y)$ está definido como el valor obtenido por medio de la integral de la función de intensidad de los puntos marcados en la muestra $\lambda_{\theta_\pi}(X)$ en todo su dominio. Esta función de intensidad $\lambda_{\theta_\pi}(X)$ es el producto de $\lambda(X)$ con la función espacial de θ ($f_\theta(x, y)$) y la función espacial de π ($f_\pi(x, y)$).

$$\int \int_X \lambda_{\theta_\pi}(x, y) dx dy = \int \int_X \lambda(x, y) f_\theta(x, y) f_\pi(x, y) dx dy$$

En caso que π no dependa de su ubicación en el espacio, entonces este mantendrá su valor constante. De forma análoga, se puede determinar también cuántos puntos fueron seleccionados en total en la muestra con el valor obtenido por medio de la integral de la función de intensidad de los puntos en la muestra $\lambda_\pi(X)$ en todo su dominio. Esta función de intensidad es el producto de $\lambda(X)$ con la función espacial de π ($f_\pi(x, y)$).

$$\int \int_X \lambda_\pi(x, y) dx dy = \int \int_X \lambda(x, y) f_\pi(x, y) dx dy$$

Consecuentemente, en base a la Ecuación 2.1, para estimar una proporción considerando densidad poblacional, riesgo espacial y el tipo de muestreo se emplearía el siguiente cálculo:

$$\hat{p} = \frac{\int \int_X \lambda_{\theta\pi}(x, y) dx dy}{\int \int_X \lambda_{\pi}(x, y) dx dy} = \frac{\int \int_X \lambda(x, y) f_{\theta}(x, y) f_{\pi}(x, y) dx dy}{\int \int_X \lambda(x, y) f_{\pi}(x, y) dx dy} \quad (2.4)$$

Si bien la Ecuación 2.4 vendría a ser considerada como una generalización, en la Tabla 2.1 se puede ver el comportamiento de esta ecuación para cada uno de los escenarios.

Tabla 2.1: Funciones para estimar una proporción considerando densidad poblacional, riesgo espacial y el tipo de muestreo

	Densidad poblacional homogénea	
	Muestreo aleatorio simple	Muestreo con sesgo espacial
Riesgo espacial constante	$\hat{p} = \theta$	$\hat{p} = \theta$
Riesgo espacial variable	$\hat{p} = \int \int_X f_{\theta}(x, y) dx dy$	$\hat{p} = \frac{\int \int_X f_{\theta}(x, y) f_{\pi}(x, y) dx dy}{\int \int_X f_{\pi}(x, y) dx dy}$
	Densidad poblacional heterogénea	
	Muestreo aleatorio simple	Muestreo con sesgo espacial
Riesgo espacial constante	$\hat{p} = \theta$	$\hat{p} = \theta$
Riesgo espacial variable	$\hat{p} = \frac{\int \int_X \lambda(x, y) f_{\theta}(x, y) dx dy}{\int \int_X \lambda(x, y) dx dy}$	$\hat{p} = \frac{\int \int_X \lambda(x, y) f_{\theta}(x, y) f_{\pi}(x, y) dx dy}{\int \int_X \lambda(x, y) f_{\pi}(x, y) dx dy}$

Estudios de casos y controles

Un estudio de casos y controles se basa en identificar a los individuos positivos a la enfermedad (casos) y determinar su exposición. Luego, se compara esta información con un grupo negativo a la enfermedad (controles) (Celentano y cols., 2019). Una medida de asociación común dentro de estos estudios es el Odd Ratio (OR).

Tabla 2.2: Estudio Caso-control

		Enfermedad		Total
		Positivo	Negativo	
Exposición	Presente	a	b	$a + b$
	Ausente	c	d	$c + d$
Población		$a + c$	$b + d$	$a + b + c + d$

El valor del OR entre un grupo con otro se determina por medio del cociente entre los odds en el primer grupo y los odds en el segundo grupo. Por ejemplo, en la Tabla 2.3, el cálculo para el OR entre el grupo con la exposición presente y el grupo con la exposición ausente sería:

$$OR = \frac{a/c}{b/d} = \frac{ad}{cb}$$

Modelos Aditivos Generalizados (GAM)

Un modelo aditivo generalizado (GAM) es un modelo lineal generalizado (GLM) en el que los predictores lineales están dados por una suma especificada de funciones suavizadas de las p covariables más una constante (Wood, 2020).

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \epsilon \quad (2.5)$$

$$E(Y|X_1, X_2, \dots, X_p) = \alpha + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) \quad (2.6)$$

Un ejemplo de estos vendría a ser el modelo de regresión aditivo logístico (Friedman, Hastie, y Tibshirani, 2001):

$$\log \left(\frac{P(Y = 1|X)}{P(Y = 0|X)} \right) = \alpha + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) \quad (2.7)$$

Si definimos $\mu(X) = P(Y = 1|X)$ y la función de enlace $g(\mu) = \text{logit}(\mu)$, entonces la Ecuación 2.7 estaría expresada de la siguiente manera:

$$g(\mu(X)) = \alpha + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) \quad (2.8)$$

Función de Verosimilitud (L)

La función de verosimilitud mide la probabilidad de que los valores sean observados bajo cierto valor del parámetro, el cual puede variar. Para un conjunto de observaciones independientes x_1, x_2, \dots, x_n , la función de verosimilitud es matemáticamente igual a su función de probabilidad conjunta.

$$L(\theta; \mathbf{x}) = f_\theta(x_1)f_\theta(x_2) \dots f_\theta(x_n) \quad (2.9)$$

De forma complementaria, para un vector aleatorio $\mathbf{X} = (x_1, x_2, \dots, x_n)^t$, la función de verosimilitud ponderada WL (Wang, 2001) ha sido definida como:

$$WL(\theta; \mathbf{X}, \mathbf{w}) = \prod_{i=1}^m \prod_{j=1}^{n_i} f_\theta(x_{ij}; \theta)^{w_i} \quad (2.10)$$

$$\log WL(\theta; \mathbf{X}, \mathbf{w}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w_i f_\theta(y_{ij}; \theta) \quad (2.11)$$

Donde $\mathbf{w} = (w_1, w_2, \dots, w_n)^t$ es el vector de pesos.

Un caso particular se encuentra en un modelo de proceso de Poisson no homogéneo regido por un parámetro θ (Baddeley y cols., 2015).

$$L(\theta; \mathbf{X}) \propto \lambda_{\theta}(x_1)\lambda_{\theta}(x_2) \dots \lambda_{\theta}(x_n) \exp \left(- \int_W \lambda_{\theta}(u) \, du \right) \quad (2.12)$$

En donde $\lambda_{\theta}(x_1)\lambda_{\theta}(x_2) \dots \lambda_{\theta}(x_n)$ explica la contribución por los puntos en las ubicaciones; mientras que $\exp \left(- \int_W \lambda_{\theta}(u) \, du \right)$ explica la contribución por el número de puntos observados.

2.3. Enfoque teórico conceptual asumido por el investigador

Sea:

Y_i : Resultado de la prueba en el individuo i

$$Y_i \sim \text{Bernoulli}(\theta_i), Y_i = \begin{cases} 1 : \text{Positivo} \\ 0 : \text{Negativo} \end{cases}$$

$$f(Y_i = y_i | \theta_i) = \theta_i^{y_i} (1 - \theta_i)^{1-y_i}$$

Z_i : Condición del individuo i en la muestra

$$Z_i \sim \text{Bernoulli}(\pi_i), Z_i = \begin{cases} 1 : \text{Seleccionado} \\ 0 : \text{No seleccionado} \end{cases}$$

$$f(Z_i = z_i | \pi_i) = \pi_i^{z_i} (1 - \pi_i)^{1-z_i}$$

Además, cada individuo fue considerado con una realización de un proceso no homogéneo de Poisson; con su respectiva función de intensidad, tanto para la población en general, como aquellos que pertenecen a la muestra.

2.3.1. Determinación del vector de pesos

Sea una muestra que se comporta como un proceso de Poisson no homogéneo con función de densidad $\lambda(x_i)$, donde $x_i = (x_{i1}, x_{i2})$, la cual proviene de una población cuyo comportamiento es el de un proceso de Poisson no homogéneo con función de densidad $\lambda_p(x_i)$,

$$\begin{aligned}\lambda(x_i) &= \lambda_p(x_i)h(x_i) \\ &= \lambda_p(x_i)\exp(\beta_0 + h'(x_i)) \\ \log \lambda(x_i) &= \log \lambda_p(x_i) + \beta_0 + h'(x_i)\end{aligned}$$

donde $\log \lambda_p(x_i)$ es un offset, β_0 se encuentra asociado a la proporción de muestreo y $\beta_0 + h'(x_i)$ es el efecto del muestreo en cada ubicación i . En caso de que no hubiese un sesgo a nivel espacial, $h'(x_i)$ sería 0 y $\log \beta_0$ sería la proporción del muestreo.

El cociente obtenido entre la función de intensidad muestral y la función de intensidad poblacional para una realización i ($\lambda(x_i)/\lambda_p(x_i)$) es entendido como el *riesgo* de participar en la muestra para un individuo i . Partiendo de que el riesgo de participar en la muestra no es constante a nivel espacial, el peso asignado para corregir el sesgo sería el valor inverso del riesgo que tuvo el individuo i de participar en la muestra.

$$w_i = \frac{\lambda_p(x_i)}{\lambda(x_i)} \quad (2.13)$$

Validación del modelo

Para validar los modelos empleados al momento de determinar el vector de pesos w , se distribuyó los datos en dos grupos: entrenamiento (train) y prueba (test). En ello, utilizando los datos del grupo de entrenamiento, se ajustaron 99 modelos con un número diferente de funciones base para cada uno y se determinó un residual para cada uno. El residual para un proceso puntual (\mathcal{R}) para una determinada región B se entiende como la diferencia entre el

número de puntos observados y esperados en la región B:

$$\mathcal{R}(B) = n(\mathbf{x} \cap B) - \int_B \hat{\lambda}(u) du \quad (2.14)$$

donde \mathbf{x} es el proceso puntual observado, $n(\mathbf{x} \cap B)$ es el número de puntos de \mathbf{x} en la región B y $\hat{\lambda}(u)$ es la intensidad ajustada (Baddeley y cols., 2015). Posterior a esto, utilizando el grupo de prueba, se determinó el $\mathcal{R}(B)$ utilizando la función de intensidad ajustada con la información del grupo de entrenamiento. Con esto se seleccionó el número de funciones base que permite ajustar un modelo con menor $\mathcal{R}(B)$ en el grupo de prueba. Esto se ha realizado tanto al estimar la intensidad poblacional, como muestral.

Prueba de Monte-Carlo para la homogeneidad del riesgo

Cuando el riesgo de ser seleccionado en la muestra es constante, y que por ende no existe sesgo de selección, entonces la intensidad muestral es proporcional a la intensidad poblacional.

- $H_0 : \lambda(x)/\lambda_p(x) = k$
- $H_1 : \lambda(x)/\lambda_p(x) \neq k$

La prueba para la validación de la hipótesis nula se detalla en el Algoritmo 1, el cual está basado en simulaciones de Monte-Carlo.

Algoritmo 1: Validación de Hipótesis

- 1 Se determinan los n puntos de la muestra original ;
 - 2 Se ajusta la intensidad poblacional $\lambda_p(x)$;
 - 3 Se evalúa $\lambda_p(x)$ en cada punto de la muestra; **para j en $1:1000$ hacer**
 - 4 Obtener una muestra aleatoria de tamaño n de la población;
 - 5 Se evalúa $\lambda_{p_j}(x_i)$ en cada punto i de la muestra y se obtiene su logaritmo;
 - 6 Se ajusta la intensidad muestral $\lambda_j(x)$ con $\log \lambda_{p_j}(x_i)$ como *offset* ;
 - 7 Se evalúa $\lambda_j(x_i)$ en cada punto i de la muestra original;
 - 8 Se obtiene el cociente $k_{ij} = \lambda_j(x_i)/\lambda_p(x_i)$ en cada punto i de la muestra original
 - 9 Obtener los percentiles $P_{2,5}^i$ y $P_{97,5}^i$ para cada k_i ;
 - 10 **si** $\forall i \in 1 : n, P_{2,5}^i < \lambda(x_i)/\lambda_p(x_i) < P_{97,5}^i$ **entonces**
 - 11 No se rechaza H_0
 - 12 **en otro caso**
 - 13 Se rechaza H_0
-

2.3.2. Estimación de la prevalencia

Dado que el fin es estimar la prevalencia de la enfermedad en el centro poblado, fue necesario ajustar un GAM ponderado, con $\mathbf{w} = \{w_1, w_2, \dots, w_n\}$ como vector de pesos, en cada individuo participante de la muestra. Posterior a ello, el modelo se emplearía para estimar la prevalencia de la enfermedad en cada individuo de la población que no pertenece a la muestra (n_1^* y n_0^* para el número estimado de casos y controles, respectivamente). Con estos valores, se estimó una prevalencia corregida.

$$\tilde{p} = \frac{n_1 + n_1^*}{n_1 + n_1^* + n_0 + n_0^*} \quad (2.15)$$

Validación del modelo

A fin de evitar un posible sobreajuste o subajuste en la estimación, es necesario particionar la información proveniente de la muestra en dos grupos: entrenamiento y prueba. El modelo se ajustó usando el grupo de entrenamiento y se obtuvieron las métricas de ajuste (especificidad, sensibilidad y AUC) tanto en el grupo de entrenamiento, como en el grupo de prueba (Friedman y cols., 2001).

2.3.3. Factor de corrección en estudios de casos y controles

Supongase la Tabla 2.3. En esta, la correcta estimación de la prevalencia ($\hat{\theta}$) se da cuando su valor es el mismo tanto en la muestra como en la población empleando la Ecuación 2.1.

Tabla 2.3: Caso-control con *ser seleccionado en la muestra* como exposición

		Enfermedad		
		Positivo	Negativo	Total
Muestra	Seleccionado	n_1	n_0	n
	No seleccionado	a	b	$N - n$
Población		N_1	N_0	N

$$\theta = \frac{n_1}{n_1 + n_0} = \frac{N_1}{N_1 + N_0}$$

No obstante, la proporción entre casos y controles no suele mantenerse constante entre la población y la muestra por las características propias del estudio durante su etapa de recolección de información. Esto causa que el valor de θ sea obtenido mediante un estimador de máxima verosimilitud ponderado $\tilde{\theta}$ con $\mathbf{w} = (w_1, w_2)^t$ como vector de pesos para los casos y los controles. Partiendo de $Y_i \sim \text{Bernoulli}(\theta)$,

$$\begin{aligned} \text{WL}(\theta; \mathbf{Y}, \mathbf{w}) &= \prod_{i=1}^2 \prod_{j=1}^{n_i} [\theta^{y_j} (1 - \theta)^{1-y_j}]^{w_i} \\ \log \text{WL}(\theta; \mathbf{Y}, \mathbf{w}) &= \sum_{i=1}^2 \sum_{j=1}^{n_i} w_i \log [\theta^{y_j} (1 - \theta)^{1-y_j}] \\ \log \text{WL}(\theta; \mathbf{Y}, \mathbf{w}) &= \sum_{i=1}^2 \sum_{j=1}^{n_i} w_i [y_j \log(\theta) + (1 - y_j) \log(1 - \theta)] \\ \log \text{WL}(\theta; \mathbf{Y}, \mathbf{w}) &= w_1 n_1 \log(\theta) + w_2 n_0 \log(1 - \theta) \end{aligned}$$

Al maximizar esta función,

$$\begin{aligned} \frac{\partial}{\partial \theta} \log \text{WL}(\theta; \mathbf{Y}, \mathbf{w}) &= 0 \\ \frac{\partial}{\partial \theta} \log \text{WL}(\theta; \mathbf{Y}, \mathbf{w}) &= \frac{n_1}{\theta} - \frac{\rho n_0}{1 - \theta} = 0 \end{aligned}$$

Lo cual da como resultado que:

$$\tilde{\theta} = \frac{w_1 n_1}{w_1 n_1 + w_2 n_0}$$

Si $\mathbf{w} = (w_1, w_2)^t = (1, \rho)^t$,

$$\tilde{\theta} = \frac{n_1}{n_1 + \rho n_0} \tag{2.16}$$

Entonces una estimación correcta de $\tilde{\theta}$ se da cuando matemáticamente el valor de ρ es equivalente al Odds Ratio (OR) entre la muestra y la población.

$$\rho = \frac{n_1 N_0}{n_0 N_1} \quad (2.17)$$

Esto se aplica directamente en la estimación de la probabilidad de que un determinado individuo ubicado en un determinado punto en el espacio (x,y) sea un caso:

$$\hat{p}(x, y) = \frac{\hat{\lambda}_1(x, y)}{\hat{\lambda}_1(x, y) + \rho \hat{\lambda}_0(x, y)} \quad (2.18)$$

Donde

$\hat{\lambda}_1$: Función de intensidad de los casos en la muestra

$\hat{\lambda}_0$: Función de intensidad de los controles en la muestra

2.3.4. Consideraciones para la simulación

Dado que π y θ son probabilidades, entonces sus valores deben estar entre 0 y 1. Para esto, la formula espacial para cada una de estas ha de ser de la siguiente manera:

$$f_{\theta}(x, y) = \frac{1}{1 + e^{-\mu_{\theta}(x, y)}}$$

$$f_{\pi}(x, y) = \frac{1}{1 + e^{-\mu_{\pi}(x, y)}}$$

Donde:

$$\mu(x, y) = \alpha + g_1(x) + g_2(y)$$

Además, se definió a la función de intensidad λ como:

$$\lambda(x, y) = c_\lambda \left(1 - \sqrt{\frac{a_\lambda(x - h_\lambda)^2 + b_\lambda(y - k_\lambda)^2}{a_\lambda h_\lambda^2 + b_\lambda k_\lambda^2}} \right)$$

Las funciones de intensidad λ_π y λ_{θ_π} se comportan como:

$$\lambda_\pi(x, y) = \frac{c_\lambda}{(1 + \exp[-\mu_\pi(x, y)])} \left(1 - \sqrt{\frac{a_\lambda(x - h_\lambda)^2 + b_\lambda(y - k_\lambda)^2}{a_\lambda h_\lambda^2 + b_\lambda k_\lambda^2}} \right)$$

$$\lambda_{\theta_\pi}(x, y) = \frac{c_\lambda}{(1 + \exp[-\mu_\theta(x, y)])(1 + \exp[-\mu_\pi(x, y)])} \left(1 - \sqrt{\frac{a_\lambda(x - h_\lambda)^2 + b_\lambda(y - k_\lambda)^2}{a_\lambda h_\lambda^2 + b_\lambda k_\lambda^2}} \right)$$

Delimitando que:

$$\mu_\theta(x, y) = c_\theta + a_\theta(x - h_\theta)^2 + b_\theta(y - k_\theta)^2$$

$$\mu_\pi(x, y) = c_\pi + a_\pi(x - h_\pi)^2 + b_\pi(y - k_\pi)^2$$

2.4. Marco legal

La presente investigación se realizó gracias al apoyo del fondo EULAC por medio de FONDECYT. En relación a la parte ética de esto, el estudio madre cuenta con la aprobación ética de la Universidad Peruana Cayetano Heredia, tanto en la recolección de datos como en el manejo de la información (ver Anexo 1.1). Al ser un estudio secundario, la integridad de los participantes no se ha visto comprometida más allá de su privacidad. Por lo cual, con el fin de proteger la dicha privacidad se está trabajando con los código autogenerados de cada uno.

2.5. Hipótesis

2.5.1. Hipótesis General

- El método planteado en la presente investigación permite estimar la prevalencia de forma insesgada.

2.5.2. Hipótesis Específicas

- El método planteado en la presente investigación por lo general estima de una mejor forma la prevalencia que el método tradicional.
- El método planteado en la presente investigación es el adecuado para la situación que se está observando.
- el método planteado en la presente investigación puede ser utilizado en otros centros poblados.

Capítulo III

MARCO METODOLÓGICO

3.1. Tipo, nivel y diseño de la investigación

Estudio madre

El estudio madre de donde provienen los datos es de carácter observacional y consiste en una cohorte prospectiva iniciada en el 2017 en comunidades ganaderas del departamento de Junin bajo el nombre de VIRSEL. Posterior a ello, bajo el nombre de HYCOM, se hizo una segunda recolección de datos en el 2018. Finalmente, este estudio fue modificado en el 2019 ampliando su alcance al eliminar la limitante del departamento; lo cual permitió la recolección de información en la región de Huancavelica en el 2020.

Estudio de tesis

El estudio de tesis se trata de una investigación de carácter observacional y de corte transversal con la información recolectada en el estudio madre en los centros poblados de Corpacancha y Canchayllo.

3.2. Población, muestra y tamaño de muestra

La población del presente estudio comprende a los habitantes del centro poblado de Corpacancha, Junín-Perú. La información recolectada (muestra) respecto a la tenencia de la enfermedad proviene de dos intervenciones realizadas con anterioridad en la comunidad:

1. VIRSEL: Primera intervención realizado en Octubre del 2017. La información fue recolectada por medio de una campaña de despistaje gratuita en el centro de salud del lugar.
2. HYCOM: Segunda intervención realizada en Junio del 2018. La información fue recolectada por medio de una campaña de despistaje gratuita en la que se visitaron las casas de los habitantes.

Por otro lado, la información referente al otro centro poblado sobre el que se implementará la metodología (Canchayllo, Junín-Perú) proviene de tres intervenciones realizadas en los años 2019 (Septiembre) y 2020 (Diciembre). Esta recolección de información se llevó a cabo gracias al fondo EULAC, a través de CONCYTEC (PRO CIENCIA), quienes también apoyaron esta investigación.

El poder trabajar con información en relación a la tenencia de hidatidosis humana proveniente de dos intervenciones diferentes realizadas con varios meses de diferencias es posible dado al periodo de incubación de la enfermedad (Ver Sección 2.2.1). Por otro lado, en lo que respecta a la información sobre la distribución espacial de los habitantes, esta proviene de un censo realizado en la comunidad en simultáneo a los estudios VIRSEL y HYCOM. De este se tienen las coordenadas geográficas de cada casa de la comunidad.

3.3. Técnicas de análisis e instrumentos

3.3.1. Análisis de datos

Se inició con una limpieza de la base de datos a fin de contar solo con la información con la que se trabajó, creando las variables necesarias para su procesamiento. Después, se consolidó en una sola base de datos con las coordenadas de cada casa y sus distancias al centro de salud. Posterior a esto, se procedió a realizar un análisis exploratorio de las variables y los resultados fueron reportados por medio de tablas.

Previo al procesamiento de la información real por medio del método planteado, fue necesario estudiarlo. Por ello, su eficacia se evaluó dentro de los siguientes escenarios:

1. Densidad poblacional homogénea, riesgo espacial constante
2. Densidad poblacional heterogénea, riesgo espacial variable

Para cada uno de estos 2 escenarios, las simulaciones se realizaron tomando en consideración diferentes niveles de prevalencia de hidatidosis humana. De igual forma, en cada uno se simularon dos tipos de muestreos: uno totalmente aleatorio y el otro con un sesgo de selección hacia los más próximos al punto en que se recolectó la información por medio de la campaña. Esto fue considerando diferentes niveles de cobertura (porcentaje de individuos de la población que participaron en la campaña de despistaje). Estos escenarios han sido simulados como procesos puntuales. Estos han sido procesos de Poisson homogéneos y no-homogéneos, según sea el caso constante o variable. Con cada uno de los escenarios simulados, se procedió a estimar la prevalencia por medio del método planteado (ver Sección 2.3). Se comparó el valor de esta estimación con el resultado obtenido por el método tradicional (ver Ecuación 2.2) y el nivel de prevalencia establecido en cada simulación. Con esto se determinó bajo qué condiciones el método propuesto resulta eficiente en comparación al método tradicional tomando como criterio al ECM de las estimaciones (ver Ecuación 2.3).

Al pasar a la data real, además de utilizar la información obtenida en la intervención VIRSEL por medio de la campaña de salud realizada en el Centro Poblado de Corpacancha (Junín, Perú), se usó la información proveniente del censo realizado a la comunidad para la distribución espacial de los pobladores. A continuación, para determinar la probabilidad de que cada individuo de la comunidad haya participado en la intervención VIRSEL se ajustó un GAM (ver Ecuación 2.7) con las coordenadas del habitante, considerando también como covariables al sexo y al rango de edad. Después, a fin de tener un control en las estimaciones por ambos métodos, cada una de estas fue comparada con el valor de la prevalencia observada en la intervención HYCOM. Finalmente se implementó el método sobre la población de un centro poblado diferente (Canchayllo) y se obtuvo una prevalencia corregida.

El código de lo trabajado en esta sección en R, tanto el análisis como las simulaciones, se encuentra en el QR presentado en el Anexo B.

3.3.2. Instrumentos de medición

Prueba diagnóstica

Para determinar la tenencia de Hidatidosis Humana en el estudio VIRSEL se empleó la prueba diagnóstica de Wester Blot. Esta prueba consiste en vertir sangre sobre un papel especial y según el número de bandas que esta presenta, se determina si el individuo presenta o no Hidatidosis. Por otro lado, en el estudio HYCOM, la prevalencia de la enfermedad se determinó por medio de una ecografía y una tomografía, además del Wester Blot.

Geolocalización

Durante el censo realizado en paralelo se determinó la casa correspondiente para cada individuo de la comunidad. De igual forma, por medio de un equipo de GPS se determinaron las coordenadas para cada casa de la comunidad. Esta información fue almacenada digitalmente en formato KML.

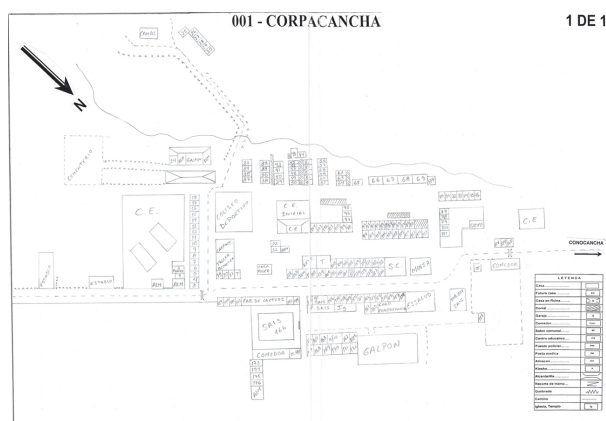


Figura 3.1: Mapa urbano del centro poblado de Corpacancha (Junín, Perú)

3.4. Cuadro de operacionalización de variables

Tabla 3.1: Cuadro de operacionalización de variables

Variable	Definición Conceptual	Tipo	Valores
Distancia	Distancia en metros desde la casa de un individuo al centro de EsSalud	Continua	Número real mayor a 0
Prevalencia	Proporción de individuos que presentan la enfermedad	Continua	Número real comprendido entre 0 y 1
Sexo	Sexo del individuo	Nominal	0: Femenino 1: Masculino
Edad	Tiempo de vida en años del individuo	Discreta	Número entero mayor o igual a 0
Rango de edad	Tiempo de vida en años del individuo agrupado en rangos	Ordinal	1: Menores de 15 años 2: De 16 a 30 años 3: De 31 a 40 años 4: De 41 a 50 años 5: De 51 a 65 años 6: De 66 años a más

3.5. Matriz de consistencia

En la siguiente página, en la Tabla 3.2, encontramos la matriz de consistencia del estudio

Tabla 3.2: Matriz de consistencia

Problema	Objetivos	Hipótesis	Variables e Indicadores	Metodología
Problema General ¿Se puede emplear un método de estimación de prevalencia considerando la distribución espacial para reducir el sesgo del muestreo?	Objetivo General Estudiar un método de estimación de prevalencia considerando la distribución espacial para reducir el sesgo del muestreo	Hipótesis General El método planteado estima la prevalencia de forma insesgada	Variable dependiente Prevalencia: Porcentaje de individuos que presentan la enfermedad	Tipo de Investigación Básica.
Problemas Específicos ¿Cómo se diferencian los resultados obtenidos de las estimaciones de la prevalencia mediante el método clásico y el error de estimación del método propuesto? ¿Cuál es el método adecuado de acuerdo a las condiciones de la investigación?	Objetivos Específicos Comparar los resultados obtenidos de las estimaciones de la prevalencia por el método tradicional y el método propuesto Determinar el método adecuado de acuerdo a las condiciones de la investigación.	Hipótesis Específica El método planteado presenta un menor Error Cuadrático Medio en la estimación de la prevalencia que el método tradicional. El método planteado es el adecuado para la situación que se está observando	Variable independiente Ubicación de la vivienda del individuo Presencia de la enfermedad en el individuo Covariables epidemiológicas	Tipo de Investigación Observacional. A fin de comparar los métodos antes de aplicarlos en la data real, se simularán escenarios tomando en consideración diferentes condiciones. La información del caso a tratar proviene de dos campañas médicas realizadas en la comunidad.
¿De cuánta es la prevalencia de la hidatidosis humana en el centro poblado de Corpacancha mediante el método propuesto y de cuánta sería en un centro poblado diferente?	Estimar la prevalencia de la hidatidosis humana en el centro poblado de Corpacancha y en un centro poblado diferente mediante el método propuesto.			

Capítulo IV

ANÁLISIS Y RESULTADOS

4.1. Análisis Exploratorio

El centro poblado de Corpacancha cuenta con 177 casas, de las cuales 105 (59.3 %) están habitadas. Del censo realizado en 2018, se conoce que la población es de un total de 309 habitantes; de los cuales 141 (45.6 %) participaron en la campaña de salud. Tras haber realizado la limpieza correspondiente a los datos de la muestra proveniente del estudio VIRSEL, se conoce que la Hidatidosis Humana en el Centro poblado de Corpacancha tiene una prevalencia del $0.241 (\pm 0.054 \text{ IC}_{95} \%)$. Este estudio contó con una potencia estadística del 22.1 por ciento. La segunda campaña realizada junto al censo permitió aumentar la cobertura a un total de 196 participantes (63.4 %). Como resultado, la prevalencia disminuyó a $0.235 (\pm 0.036 \text{ IC}_{95} \%)$; indicando que la prevalencia del primer estudio se encontraba sobreestimada. Este estudio contó con una potencia estadística del 28.8 por ciento.

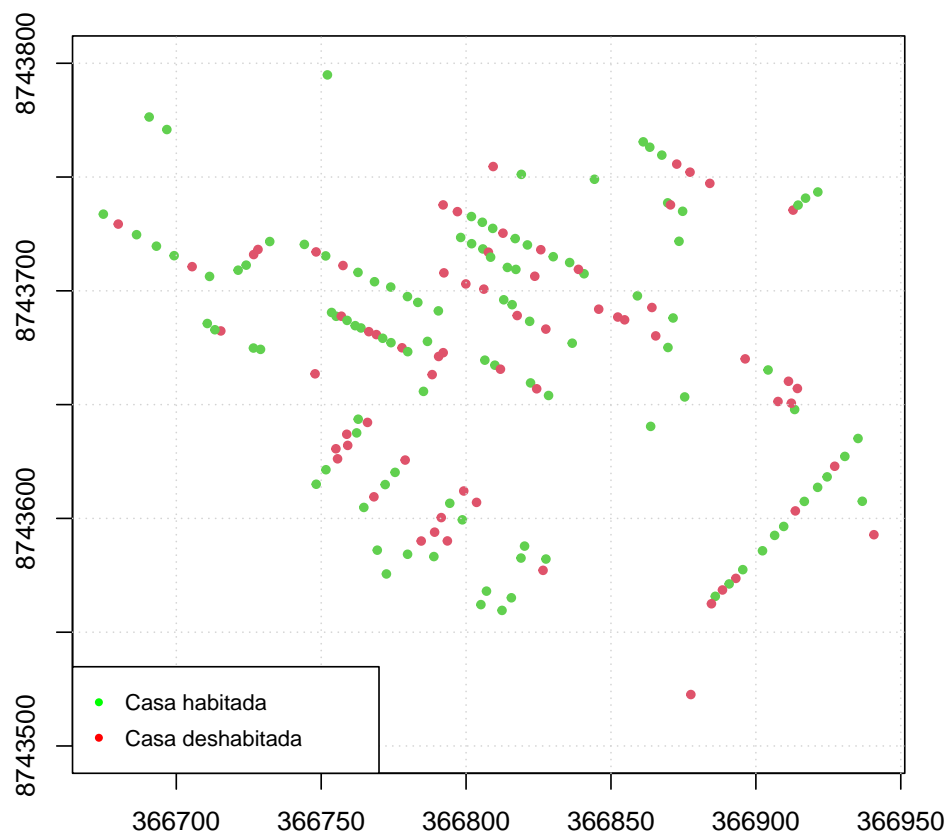


Figura 4.1: Distribución en el espacio de las casas en el centro poblado de Corpacancha.

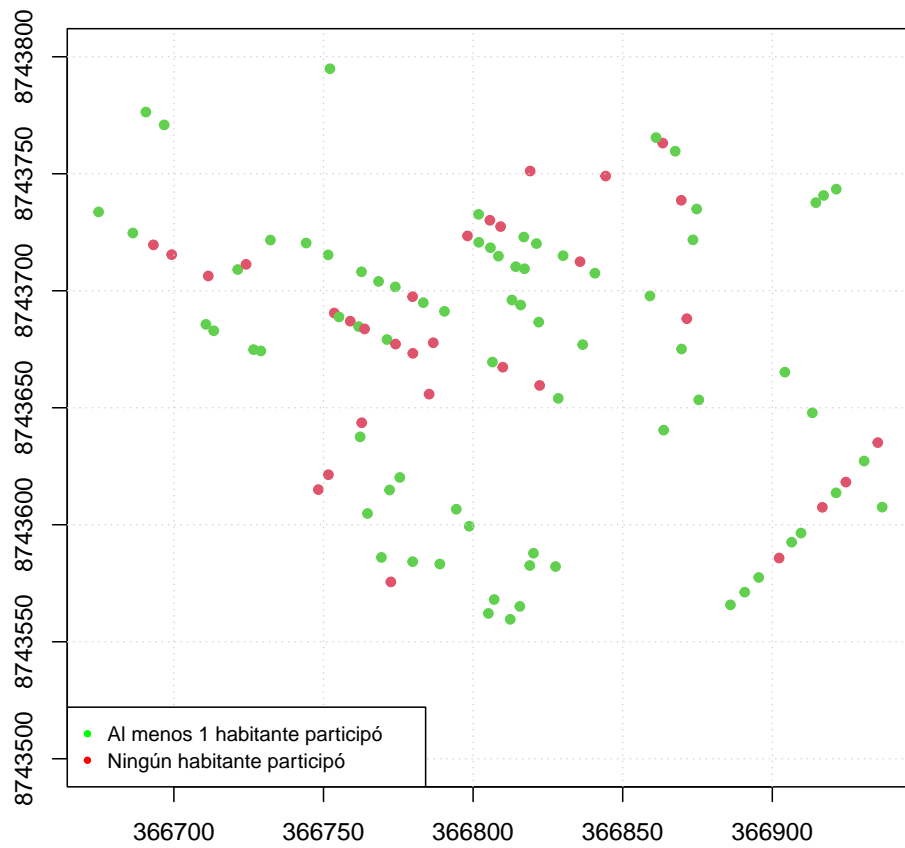


Figura 4.2: Distribución en el espacio de las casas habitadas en el centro poblado de Corpacancha.

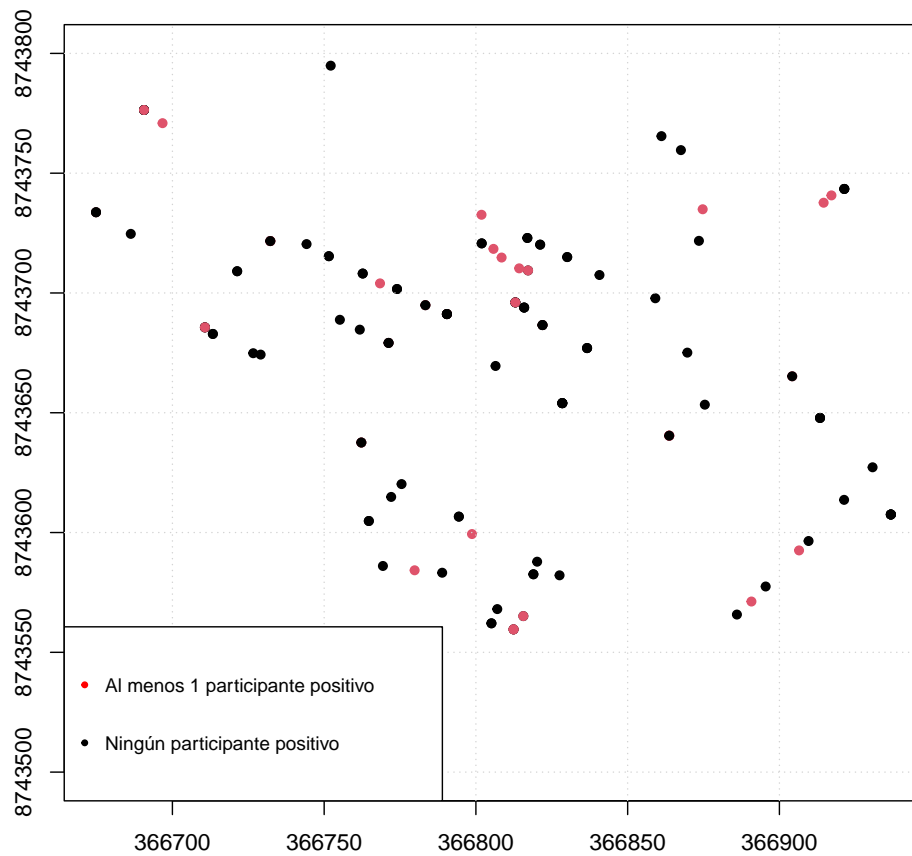


Figura 4.3: Distribución de las casas con al menos un habitante que haya participado en la campaña realizada en el centro poblado de Corpacancha.

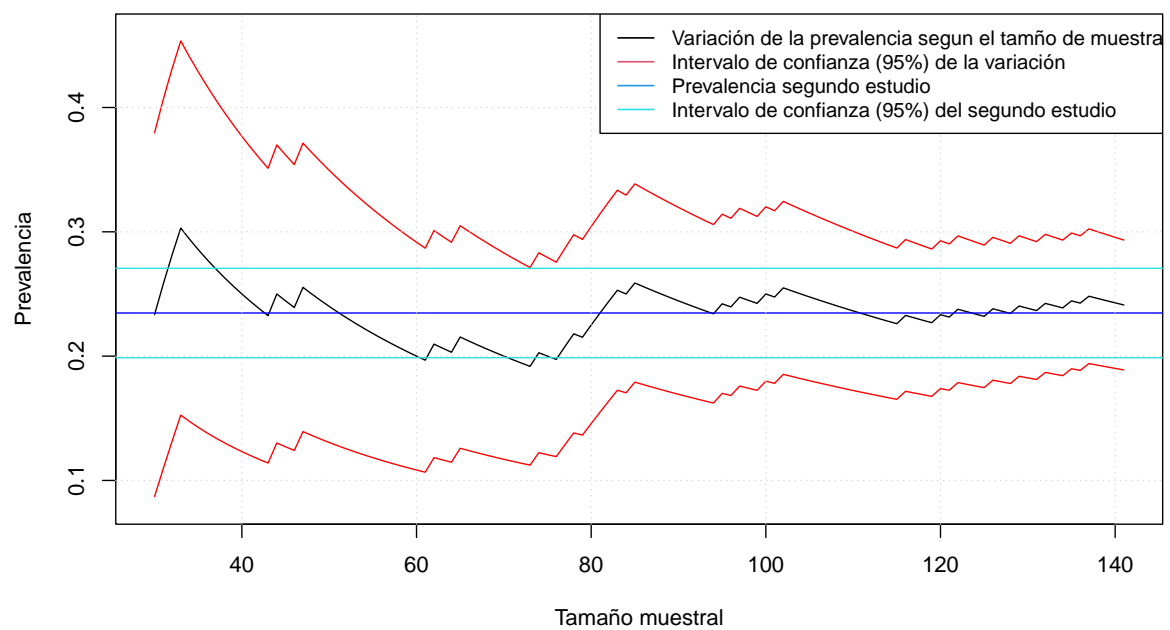


Figura 4.4: Prevalencia de Hidatidosis Humana en Corpacancha.

4.1.1. Análisis de los factores de riesgo

De la base de datos, se seleccionaron como covariables a los factores de riesgo determinados en estudios anteriores en la región (Santivañez y cols., 2010). El resumen de esta información, tanto univariado, como bivariado respecto a la prevalencia, se puede encontrar en las tablas 4.1 y 4.2. En esta segunda tabla se observaron diferencias que fueron necesarias estudiar.

Tabla 4.1: Factores de riesgo y prevalencia (n=141).

Características		n	%
Sexo			
	Femenino	83	58.9
	Masculino	58	41.1
Edad *		33.27 (\pm 3.19)	
Número de perros			
	0	95	67.4
	1	15	10.6
	2	17	12.1
	3	11	7.8
	4	3	2.1
Prevalencia**		0.241 (\pm 0.054)	

* Promedio (IC 95 %)

** Proporción (IC 95 %)

Tabla 4.2: Factores de riesgo por resultado de la campaña (n=141).

Factor de riesgo	Positivo	Negativo
Sexo		
Male	16 (27.6 %)	42 (72.4 %)
Female	18 (21.7 %)	65 (78.3 %)
Edad*		
	36.1 \pm 6.7	32.4 \pm 3.6
Número de perros		
0	21 (22.1 %)	74 (77.9 %)
1	2 (13.3 %)	13 (86.7 %)
2	3 (17.6 %)	14 (82.4 %)
3	7 (63.6 %)	4 (36.4 %)
4	1 (33.3 %)	2 (66.7 %)

* Media por resultado

Sexo

En la tabla 4.2, se observó que hay una diferencia en la prevalencia por sexo de la persona. Pese a esto, la diferencia encontrada no era significativa debido a la superposición de los intervalos de confianza de cada una (Fig. 4.5).

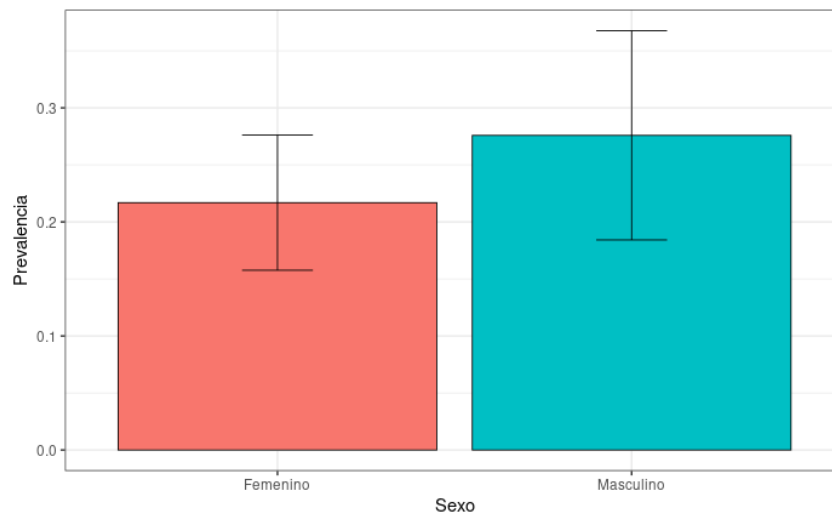


Figura 4.5: Prevalencia e intervalo de confianza por sexo.

Ajustando un modelo lineal generalizado entre el sexo y la prevalencia de hidatidosis, se observó que la posibilidad de tener hidatidosis siendo hombre es 1.38 veces la posibilidad siendo mujer (Tabla 4.3). Pese a que los resultados no son significativos, esto no descarta que se considere como factor de riesgo debido a que el p-valor se encuentra relacionado con la potencia estadística que tuvo la muestra (Ellis, 2010).

Tabla 4.3: Odd Ratio de la variable sexo

	OR*	IC 95 %	p-valor
Sexo = Masculino	1.38	[0.63; 3.00]	0.421

*obtenido al ajustar un modelo lineal generalizado

Edad

En la tabla 4.2, se observó que hay una diferencia en la edad promedio entre el grupo positivo a la enfermedad y el grupo negativo a la enfermedad. No obstante, esta diferencia resultó ser no significativa pues sus intervalos de confianza se superponen. Ajustando un modelo lineal generalizado entre el edad y la prevalencia de hidatidosis, se observó que la posibilidad de tener hidatidosis se incrementaba 1.01 veces con cada año de edad (Tabla 4.4). Pese a que los resultados no son significativos, esto no descarta que se considere como factor de riesgo. Esto debido a que el p-valor se encuentra relacionado con la potencia estadística que tuvo la muestra (Ellis, 2010).

Tabla 4.4: Odd Ratio de la variable edad

	OR*	IC 95 %	p-valor
Edad	1.01	[0.99; 1.03]	0.334

*obtenido al ajustar un modelo lineal generalizado

Número de perros

En la tabla 4.2, se observó que hay una diferencia en la prevalencia por el número de perros que se tienen. La diferencia encontrada era significativa cuando se dicotomizaba la variable del número de perros. Esto debido a la no superposición de los intervalos de confianza de cada una (Fig. 4.6). Ajustando un modelo lineal generalizado entre el número de perros y la prevalencia de hidatidosis, se observó que la posibilidad de tener hidatidosis teniendo menos de tres perros es 0.19 veces la posibilidad teniendo al menos tres perros (Tabla 4.3). Esto determinó que en adelante, la variable número de perros sea tomada como cualitativa con dos categorías: menos de tres perros y al menos tres perros.

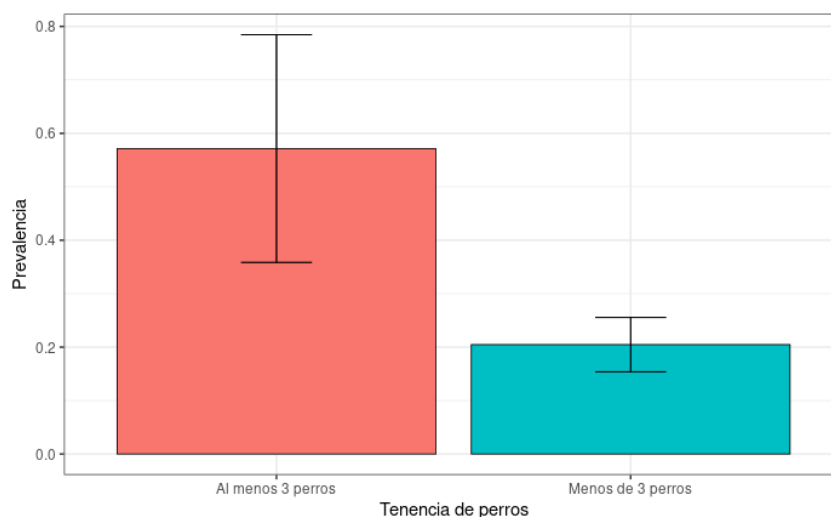


Figura 4.6: Prevalencia e intervalo de confianza por tenencia de perros.

Tabla 4.5: Odd Ratio de la variable perro

	OR*	IC 95 %	p-valor
Perros = 1	0.54	[0.08; 2.17]	0.443
Perros = 2	0.76	[0.16; 2.59]	0.681
Perros = 3	6.17	[1.70; 25.49]	0.007
Perros = 4	1.76	[0.08; 19.28]	0.650
Perros = Menos de 3 perros	0.19	[0.06; 0.60]	0.005

*obtenido al ajustar un modelo lineal generalizado

4.1.2. Análisis de la intensidad

Se observó que la población se concentra al sureste del centro de salud en el que se hizo la campaña (Fig. 4.7). Se encontró un comportamiento similar en los habitantes que participaron en la campaña (Fig. 4.8). No obstante, era necesario observar si la razón de las intensidades se mantiene constante en el espacio. En la figura 4.9 se observó que el riesgo de participar en la muestra no es constante en el área; especialmente en la zona noroeste. Esto planteó la existencia de un posible sesgo espacial.

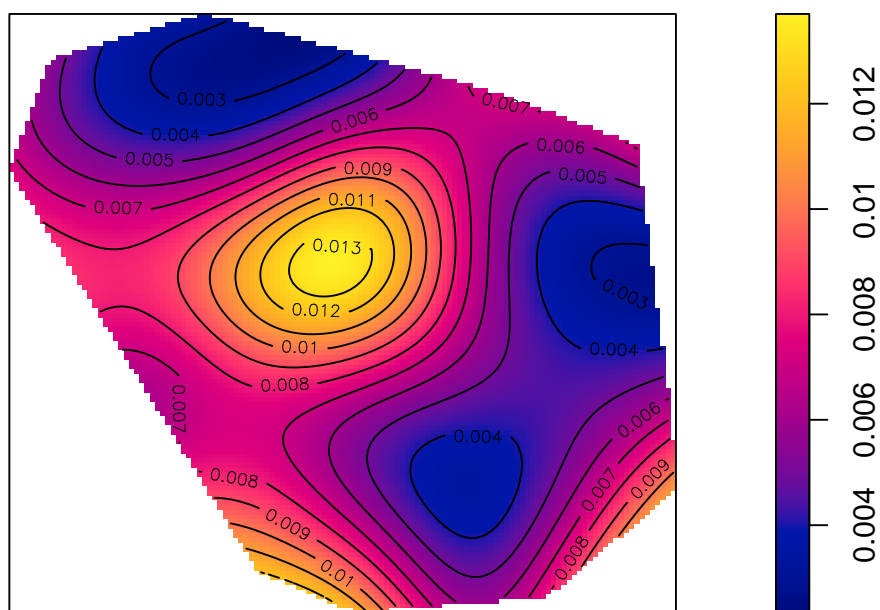


Figura 4.7: Intensidad poblacional

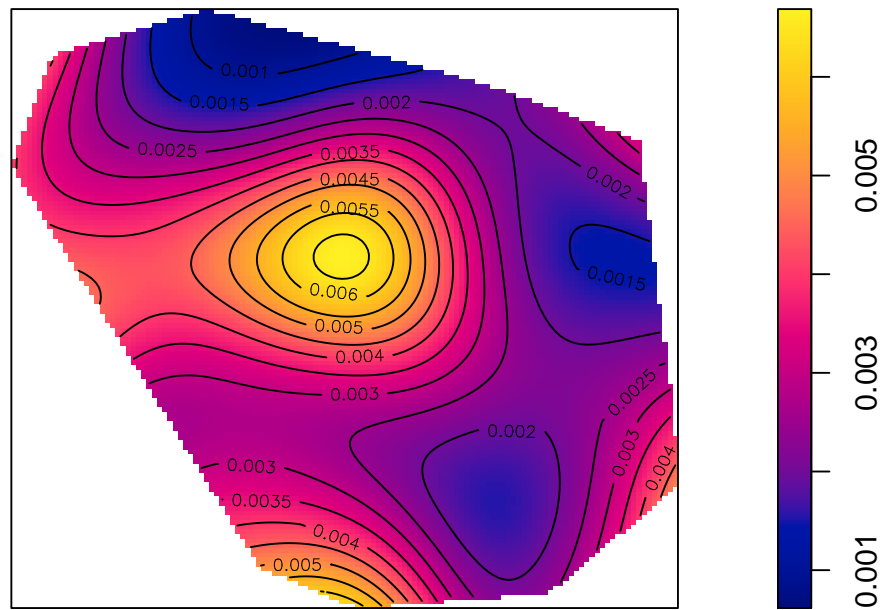


Figura 4.8: Intensidad muestral

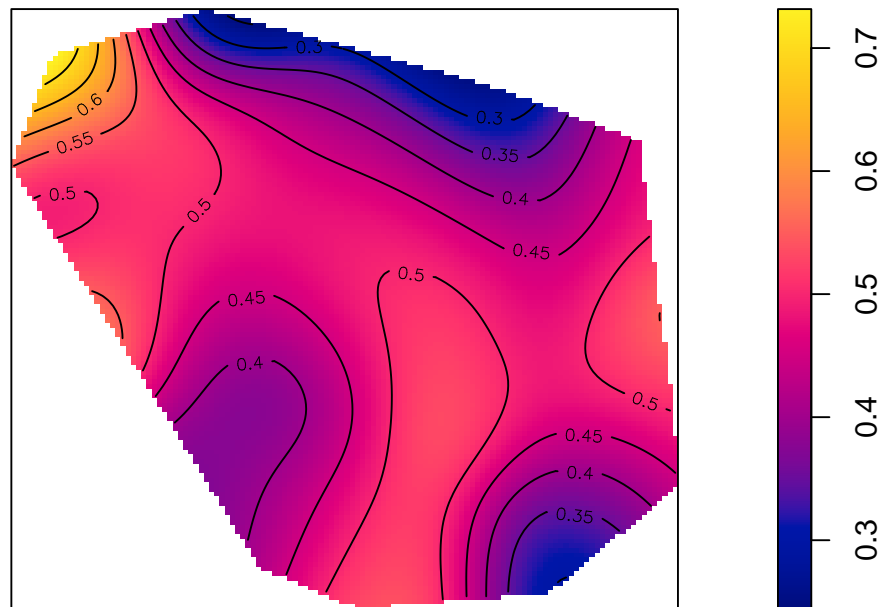


Figura 4.9: Riesgo de participar en la muestra

Lo visto en la Figura 4.9 indicó la necesidad de corregir el riesgo de tener la enfermedad a nivel espacial (Fig. 4.10).

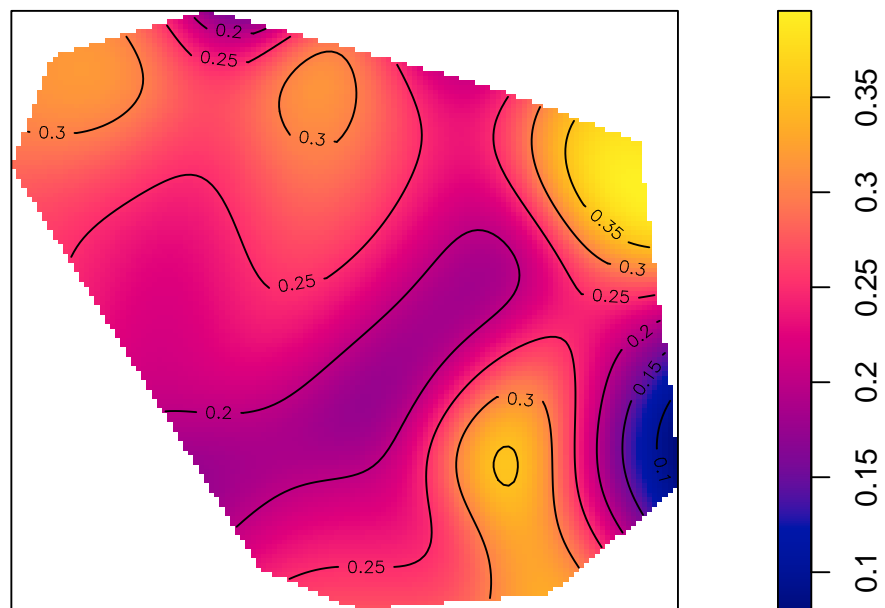


Figura 4.10: Riesgo de tener la enfermedad

4.2. Simulaciones

Cada una de las simulaciones y sus resultados fueron evaluados por escenario (Tabla 4.6, Figura 4.11 y Figura 4.12). Si bien el $ECM(\tilde{\theta})$ resultó ser mayor al $ECM(\hat{\theta})$ en cada escenario, también se observó que en su mayoría estos superaron el 15 % de éxito; o sea, que en 13 de 18 escenarios, al menos el 15 % de las simulaciones, la estimación de la prevalencia por medio del método presentado tenía un valor más cercano al real que la estimación por el método convencional. Este comportamiento también se observó en los escenarios con las características más similares a la data real, a nivel de cobertura y tamaño poblacional. Pese a esto, las curvas de densidad en cada escenario presentaron notorias diferencias entre si.

Tabla 4.6: Resultados de la simulación

Escenario	m	Éxitos (%)	\hat{n}	Cobertura	Prevalencia*	$ECM(\hat{\theta})$	$ECM(\tilde{\theta})$
1	506	33.20	249	34.67	0.252	0.0014	0.0037
2	510	23.92	249	44.95	0.250	0.0009	0.0049
3	509	21.61	250	54.86	0.251	0.0006	0.0053
4	510	23.53	349	35.09	0.251	0.0010	0.0066
5	510	15.69	350	44.94	0.249	0.0007	0.0080
6	510	10.78	350	54.86	0.252	0.0004	0.0087
7	510	14.12	451	35.09	0.249	0.0007	0.0118
8	510	8.43	450	45.08	0.251	0.0005	0.0108
9	510	6.47	450	55.21	0.251	0.0003	0.0098
10	509	15.52	247	34.28	0.241	0.0040	0.0271
11	499	15.83	246	44.38	0.243	0.0009	0.0153
12	510	16.86	247	55.41	0.242	0.0007	0.0088
13	503	20.68	344	34.50	0.242	0.0033	0.0259
14	506	16.60	342	44.20	0.243	0.0007	0.0127
15	509	15.52	343	55.47	0.243	0.0005	0.0080
16	510	21.96	440	34.40	0.241	0.0034	0.0236
17	510	15.69	440	44.16	0.242	0.0005	0.0124
18	510	14.51	440	55.36	0.242	0.0004	0.0088

m: Número de simulaciones por escenario

n: Número promedio de personas por escenario

Cobertura*: Cobertura de muestreo promedio por escenario

Prevalencia*: Prevalencia promedio por escenario

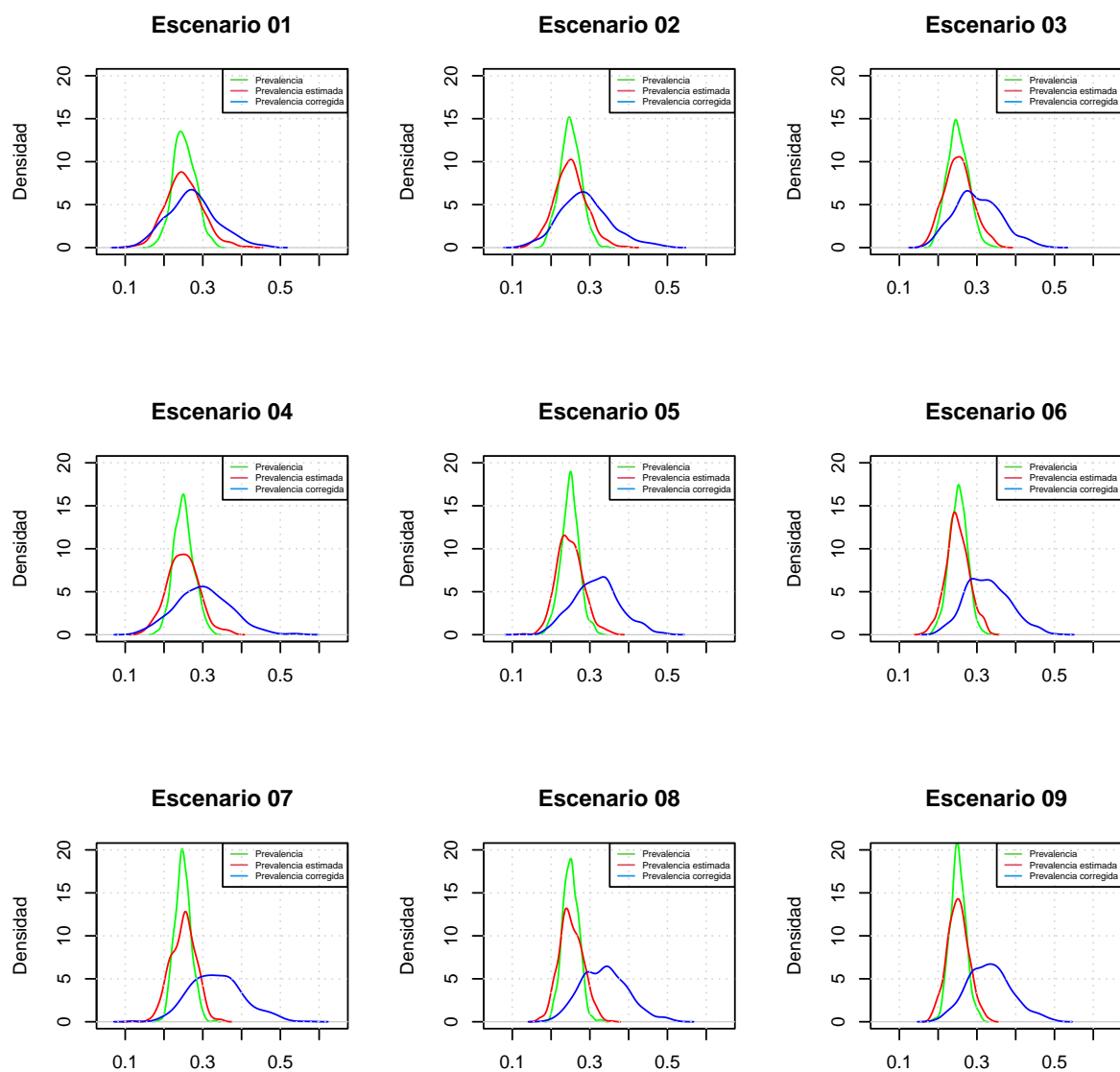


Figura 4.11: Densidad del parámetro de los escenarios 1 al 9

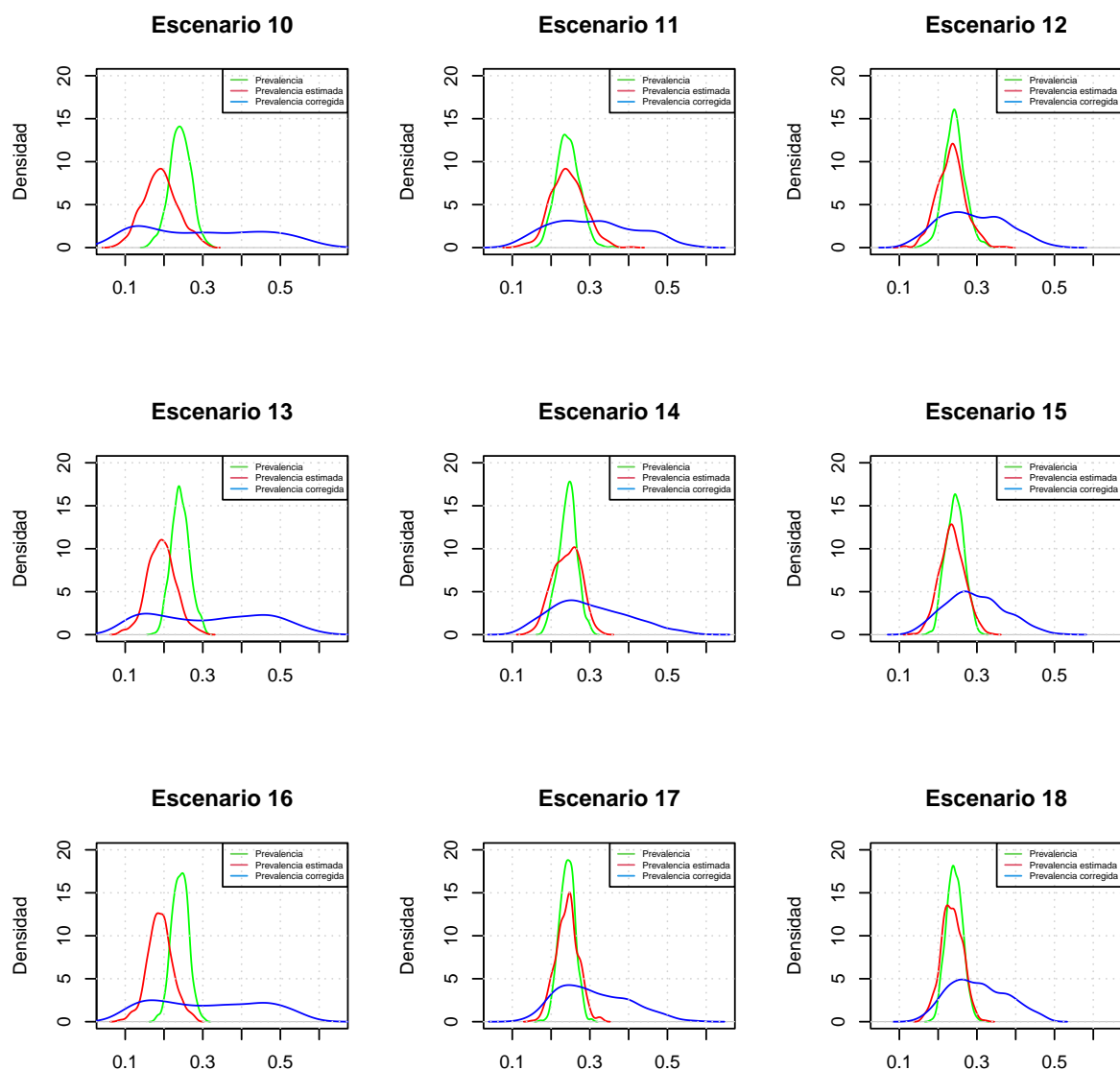


Figura 4.12: Densidad del parámetro de los escenarios 10 al 18

4.3. Modelamiento

4.3.1. Intensidad

En la estimación de la función de intensidad para la población y la muestra, los modelos se ajustaron con 14 y 16 funciones de base, respectivamente, en base a sus residuales (Véase Fig. 4.13 y 4.14)

Homogeneidad de riesgo

En base al algoritmo 1, se determinó que el riesgo al ser seleccionado es no homogéneo, debido a que en más de una observación, el cociente entre las dos intensidades no se ha encontrado en el intervalo de confianza correspondiente. Este resultado confirmó lo visto en la Figura 4.9.

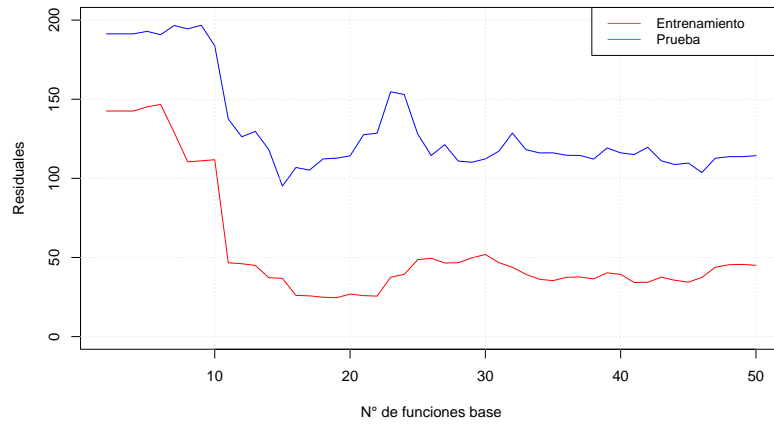


Figura 4.13: Residuales en la población

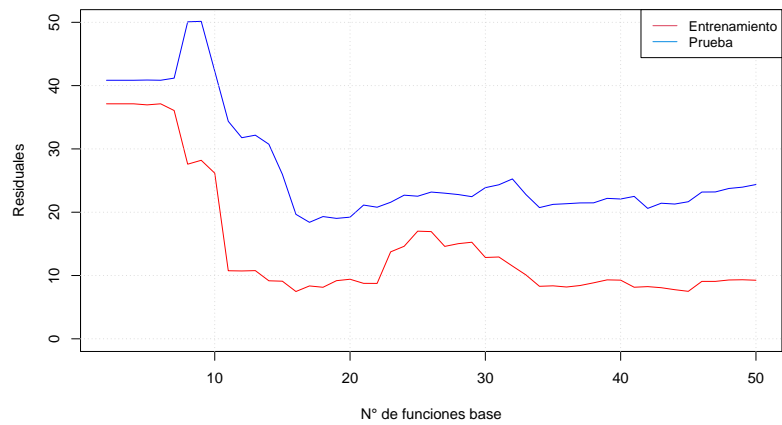


Figura 4.14: Residuales en la muestra

4.3.2. Riesgo

Comparando las métricas de ajuste del grupo de entrenamiento y del grupo de prueba de cada modelo ajustado (Tabla 4.7), se encontró una mejora en los modelos al introducir el factor de corrección w , salvo aquel modelo que solo consideró el efecto espacial de las observaciones (Modelo espacial). Este mismo comportamiento se observó cuando a las covariables comprendidas como factores de riesgo (edad, sexo, tener al menos 3 perros) se les añadió la variable del efecto espacial. En base a estos resultados, se seleccionó al *Modelo completo (w)*, compuesto por los factores de riesgo, el efecto espacial y el factor de corrección w , como el modelo más adecuado para la inferencia y estimación del valor de la prevalencia corregida.

Tabla 4.7: Métricas de ajuste para los modelos

Modelo	Entrenamiento			Prueba		
	Esp.	Sens.	AUC	Esp.	Sens.	AUC
Modelo espacial	1	0	0.5381	-	-	0.6766
Modelo espacial (w)	0.96	0.125	0.5864	-	-	0.5453
Modelo edad	1	0.0833	0.5578	0.03125	1	0.5672
Modelo edad (w)	0.9333	0.2917	0.6122	0.59375	0.6	0.6453
Modelo sexo	1	0	0.5358	1	0	0.5469
Modelo sexo (w)	1	0	0.4642	1	0	0.5469
Modelo perros	0.9333	0.3333	0.6333	0.9688	0	0.149
Modelo perros (w)	0.9333	0.3333	0.6333	0.9688	0	0.149
Modelo sexo edad	1	0.0833	0.6369	0.03125	1	0.5406
Modelo sexo edad (w)	0.9467	0.4583	0.8708	0.5625	0.8	0.7562
Modelo s. e. p.	0.96	0.375	0.74	0.0938	0.9	0.5141
Modelo s. e. p. (w)	0.9467	0.4583	0.7144	0.625	0.5	0.5328
Modelo completo	0.96	0.333	0.7403	0.53125	0.5	0.6
Modelo completo (w)	0.9733	0.375	0.7503	0.5625	0.8	0.7438

4.4. Inferencia

El modelo final, *modelo completo* (w), fue utilizado para inferir la presencia de la enfermedad en los datos no muestreados. Con esto se pudo estimar una prevalencia corregida del 0.22. Este valor, si bien se encontraba dentro del intervalo de confianza para el primer estudio, estaba por debajo de su valor puntual (0.241), comportamiento que se ha visto repetido en el segundo estudio (0.235). Esto indicó la presencia de una sobrestimación en la prevalencia del primer estudio.

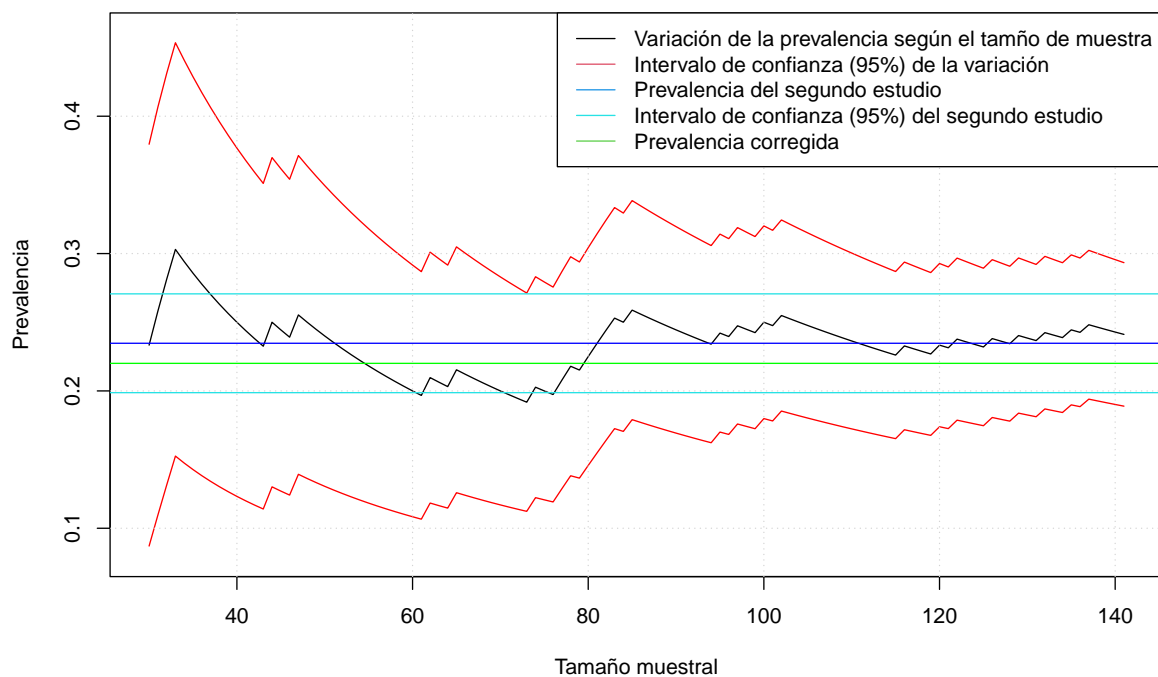


Figura 4.15: Prevalencia de Hidatidosis Humana en Corpacancha, añadiendo la prevalencia corregida.

Esto permitió estimar un factor de corrección para la intensidad de $\rho = 1.126$ (Véase 2.17). Empleando esto en 2.18, se observó el cambio en el riesgo corregido de tener la enfermedad a nivel espacial en la población (Figura 4.16).

$$\hat{p}(x, y) = \frac{\hat{\lambda}_1(x, y)}{\hat{\lambda}_1(x, y) + \rho \hat{\lambda}_0(x, y)} = \frac{\hat{\lambda}_1(x, y)}{\hat{\lambda}_1(x, y) + 1,126 * \hat{\lambda}_0(x, y)}$$

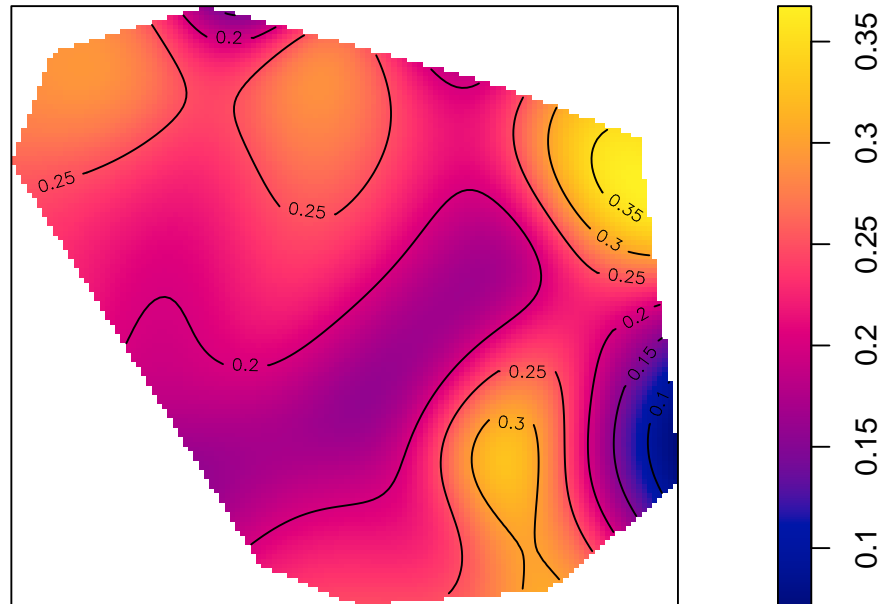


Figura 4.16: Riesgo corregido de tener la enfermedad

4.5. Implementación

Con el fin de analizar la utilidad del método propuesto pero en otros contextos, se aplicó en el centro poblado de Canchayllo y se tuvo como resultado una prevalencia del 0,150 ($\pm 0,049$ IC_{95 %}).

Ajuste de intensidad

Se ajustaron las intensidades correspondientes para la población y la muestra. El cociente de estas intensidades se puede ver en la Figura 4.17, tal como en la sección 4.3.1. La no homogeneidad observada en esta figura se confirmó tras hacer la evaluación en base al algoritmo 1, debido a que en más de una observación, el cociente entre las dos intensidades no se ha encontrado en el intervalo de confianza correspondiente de acuerdo al algoritmo.

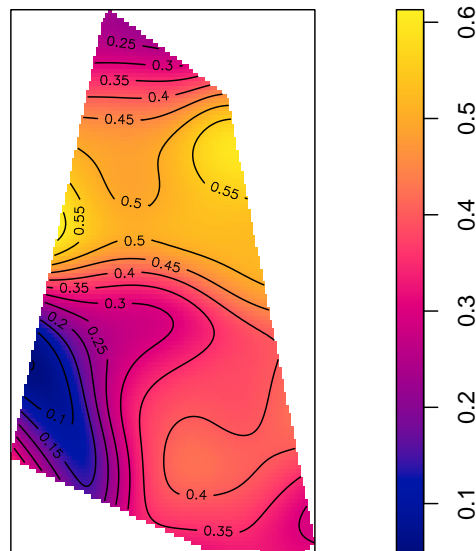


Figura 4.17: Riesgo de participar en la muestra en Canchayllo

Ajuste del modelo

Se seleccionó un modelo GAM usando los pesos obtenidos según la metodología propuesta. Se tomó como covariable a la tenencia de perros (Tener o no tener más de 3 perros). Este modelo tuvo un AUC de 0.664 y 0.763 para los datos de entrenamiento y prueba, respectivamente

Inferencia

Tras haber hecho el ajuste correspondiente, se obtuvo una prevalencia corregida, la cual ascendió a 0.245. Asimismo, el riesgo espacial fue corregido, lo cual muestra una observable diferencia con el riesgo antes de realizar la corrección. Además, se puede observar la zona en la que el riesgo se concentra.

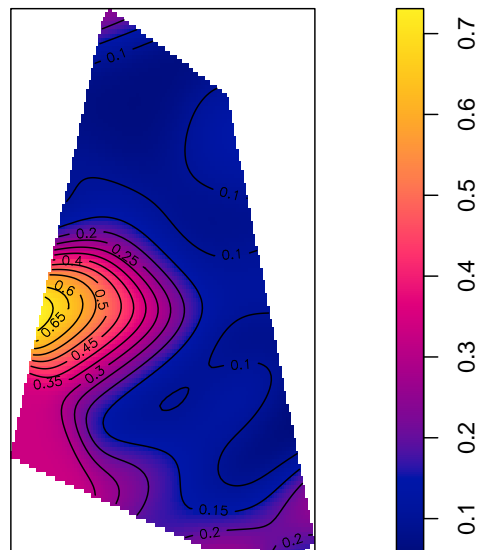


Figura 4.18: Riesgo de la enfermedad en la muestra en Canchayllo

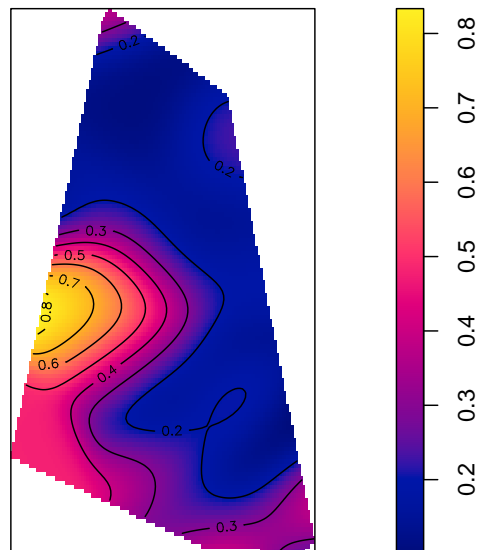


Figura 4.19: Riesgo de la enfermedad en la muestra en Canchayllo

Capítulo V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Durante el análisis de las simulaciones se encontraron diferencias observables entre la distribución de la prevalencia estimada y la corregida. No obstante, y tal como se hizo mención en la sección de resultados, en un porcentaje notorio (más del 15 % en más de un escenario) el método propuesto para corregir la prevalencia obtuvo un resultado más cercano al real. En ese sentido, y considerando que estas simulaciones solo consideraron la componente espacial en la variabilidad de los datos y no otras covariables que permitan explicar su comportamiento, se concluyó de esta forma en que por el momento no existía la suficiente evidencia estadística que permita rechazar la hipótesis de que el método propuesto por lo general estima de una mejor forma la prevalencia que el método tradicional.

Por otro lado, de acuerdo a los resultados obtenidos en la prueba de Monte Carlo para la muestra correspondiente a la primera intervención en Corpacancha (la muestra no proviene de un muestreo totalmente aleatorio a nivel especial), se conocía que a causa de la no homogeneidad del

riesgo espacial a ser seleccionados en la muestra era necesario utilizar un método que integre el factor espacial en el proceso de estimación de la prevalencia. Pese a los resultados obtenidos en el estudio del método planteado a través de las simulaciones, se aplicó con los datos reales. El método estudiado para corregir la prevalencia en el primer estudio generó resultados coherentes en relación a lo que el segundo estudio pudo obtener. En base a esto, se concluyó que la prevalencia del primer estudio (0.241) estuvo sobreestimada. El empleo del método presentado en esta investigación permitió poder estimar esta prevalencia de forma correcta al corregir el sesgo de selección. Al trabajar con la información recolectada en Canchayllo se determinó, en base a la prueba de Monte Carlo, un sesgo en la selección a nivel espacial, por lo que fue necesario emplear también el método propuesto para corregir con ello la prevalencia y el riesgo espacial a la enfermedad. Así, se concluyó que esta prevalencia se encontraba subestimada (0.245) debido al sesgo.

Durante la comparación de los modelos, se pudo observar que añadir el factor de ponderación w por sí solo no obtenía una mejora en las métricas de ajuste. Por lo que fue necesario que este sea empleado en conjunto con los factores de riesgo. Esto evidenciaría la necesidad de no depender solo del análisis espacial al momento de buscar explicar la variabilidad de la información; sino también de la información sociodemográfica epidemiológica de los participantes del estudio.

5.2. Recomendaciones

Considerando lo planteado en la Sección 2.3 y en base a las conclusiones presentadas, se recomienda que en estudios posteriores se plantee una prueba estadística que determine de forma objetiva el empleo de una metodología de corrección de sesgo espacial como la presentada. Para ello, la prueba debe estar vinculada al concepto de si el valor estimado de la prevalencia es insesgado y si el posible sesgo de selección se encuentra relacionado con el factor espacial u otra covariable. En este sentido, la prueba deberá evaluar si se rechaza o no la hipótesis nula de que la proporción de riesgo (PR) de la enfermedad entre el grupo seleccionado y el no seleccionado es 1.

De igual forma, se recomienda continuar en el estudio del método en sí bajo otros enfoques. En ello se tendría que analizar su comportamiento al añadir covariables como factores de riesgo en diversas condiciones y distribuciones de los datos que sean distintas a las presentadas. En ese sentido, se podrá determinar los criterios mediante los cuales el método pueda ser generalizado en estudios completamente diferentes. Con ello también poder desarrollar una regla de decisión que permita considerar inválida la muestra, basándose en la magnitud del sesgo de selección presente en ella.

Finalmente, también se recomienda hacer un estudio de impacto del método propuesto para determinar su viabilidad y su capacidad de implementabilidad en estudios epidemiológicos. Por ello se consideraría estudiar la rentabilidad del método, así como la importancia de su uso en el control, prevención y erradicación de enfermedades.

REFERENCIA BIBLIOGRÁFICA

- Baddeley, A., Rubak, E., y Turner, R. (2015). *Spatial point patterns: methodology and applications with r*. CRC press.
- Bailly, L., Daurès, J. P., Dunais, B., y Pradier, C. (2015). Bayesian estimation of a cancer population by capture-recapture with individual capture heterogeneity and small sample. *BMC medical research methodology*, 15, 39.
- Braeye, T., Verheagen, J., Mignon, A., Flipse, W., Pierard, D., Huygen, K., . . . Hens, N. (2016). Capture-recapture estimators in epidemiology with applications to pertussis and pneumococcal invasive disease surveillance. *PloS one*, 11(8), e0159832.
- Celentano, D. D., Mhs, S., y Szklo, M. (2019). *Gordis. epidemiología*. Elsevier.
- Davelois, K., Escalante, H., y Jara, C. (2016). Rendimiento diagnóstico del western blot para detectar simultáneamente anticuerpos en pacientes con cisticercosis, hidatidosis y fascioliasis humana. *Revista peruana de medicina experimental y salud publica*, 33, 616–624.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge university press.
- Friedman, J., Hastie, T., y Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1) (n.º 10). Springer series in statistics New York.
- general assembly of the world medical association, y cols. (2014). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists*, 81(3), 14.

- Giri, S., y Parija, S. C. (2012). A review on diagnostic and preventive aspects of cystic echinococcosis and human cysticercosis. *Tropical parasitology*, 2(2), 99.
- Kebede, W., Hagos, A., Girma, Z., y Lobago, F. (2009). Echinococcosis/hydatidosis: its prevalence, economic and public health significance in tigray region, north ethiopia. *Tropical Animal Health and Production*, 41(6), 865–871.
- Larrieu, E., Belloto, A., ARAMBULO III, P., y Tamayo, H. (2004). Echinococcosis quística: epidemiología y control en américa del sur. *Parasitología latinoamericana*, 59(1-2), 82–89.
- Moro, P., y Schantz, P. M. (2009). Echinococcosis: a review. *International journal of Infectious diseases*, 13(2), 125–133.
- Moro, P. L., Budke, C. M., Schantz, P. M., Vasquez, J., Santivañez, S. J., y Villavicencio, J. (2011, 05). Economic impact of cystic echinococcosis in peru. *PLOS Neglected Tropical Diseases*, 5, 1-6.
- PAHO/WHO. (2017). *Prevention and control of hydatidosis at local level: South american initiative for the control and surveillance of cystic echinococcosis/hydatidosis*. PANAF-TOSA.
- Pinsky, M., y Karlin, S. (2010). *An introduction to stochastic modeling*. Academic press.
- Royle, J., Chandler, R., Sollmann, R., y Gardner, B. (2014). *Spatial capture-recapture*. Academic Press.
- Ruiz-Maya, L., y Pliego, F. J. M. (2004). *Fundamentos de inferencia estadística*. AC.
- Santivañez, S. J., Naquira, C., Gavidia, C. M., Tello, L., Hernandez, E., Brunetti, E., ... Garcia, H. H. (2010). Factores domiciliarios asociados con la presencia de hidatidosis humana en tres comunidades rurales de junín, Perú. *Revista Peruana de Medicina Experimental y Salud Pública*, 27, 498–505.
- Sarkar, S., Roy, H., Saha, P., Sengupta, M., Sarder, K., y Sengupta, M. (2017). Cystic echinococcosis: a neglected disease at usual and unusual locations. *Tropical parasitology*, 7(1),

51.

Scheaffer, R., Mendenhall, W., Ott, L., y Gerow, K. G. (2011). *Elementary survey sampling*. Cengage Learning.

Shaw, A., Rushton, J., Roth, F., y Torgerson, P. R. (2017). Dalys, dollars and dogs: how best to analyse the economics of controlling zoonoses. *Revue scientifique et technique (International Office of Epizootics)*, 36(1), 147–161.

Wang, S. X. (2001). *Maximum weighted likelihood estimation* (Tesis Doctoral no publicada). University of British Columbia.

WHO. (2019). *Equinococosis*. <https://www.who.int/es/news-room/fact-sheets/detail/echinococcosis>.

Wood, S. (2020). Mixed gam computation vehicle with automatic smoothness estimation [Manual de software informático]. (R package version 1.8-33)

Anexos

Anexo A

Aspectos legales



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

Dirección Universitaria de
INVESTIGACIÓN, CIENCIA Y
TECNOLOGÍA (DUICT)

CONSTANCIA

El Presidente del Comité Institucional de Ética en Investigación (CIEI) de la Universidad Peruana Cayetano Heredia hace constar que el comité institucional de ética en investigación aprobó la ENMIENDA/MODIFICACIÓN del proyecto de investigación señalado a continuación.

Título del Proyecto : "Prevalencia de equinocosis quística humana en comunidades ganaderas de Junín, Perú".

Código de inscripción : 102270

Investigador principal : García Lescano, Héctor Hugo
Santivañez Salazar, Saul Javier

La enmienda/modificación corresponde a los siguientes documentos:

1. Protocolo de investigación, versión 3.0 de fecha 22 de agosto del 2019.
2. Consentimiento informado, versión 3.0 de fecha 22 de agosto del 2019.
3. Asentimiento informado (menores de 12 años), versión 3.0 de fecha 22 de agosto del 2019.
4. Asentimiento informado (12 a 17 años), versión 3.0 de fecha 22 de agosto del 2019.
5. Consentimiento informado (padres), versión 3.0 de fecha 22 de agosto del 2019.

Lima, 10 de setiembre del 2019.

F. Samalvides
Dra. Frine Samalvides Cuba
Presidenta
Comité Institucional de Ética en Investigación

Av. Honorio Delgado 430, SMP 15102
Apartado postal 4314
(511) 319-0000 anexo 201352
duict@oficinas-upch.pe
www.cayetano.edu.pe

Figura 1.1: Carta de Aprobación Ética del Estudio Madre

Anexo B

Códigos empleados para el desarrollo de la tesis



Figura 2.1: QR enlazado al repositorio en GITHUB que contiene los códigos trabajados