

Fase 2 | Data Science: Programación y estadística con R

Equipo 9:

Tomás Hernández

Isidro Amaro

Christian E. Millán Hernández

José Luis Herrera Gallardo

Gustavo Gutierrez

Francisco Gómez

Repositorio:

<https://github.com/BeduDSEquipo9/G2DSF2-R/tree/main/S8>

ANÁLISIS DE LA INSEGURIDAD ALIMENTARIA EN MÉXICO



Contenido:

1. Planteamiento del problema del caso
2. Análisis descriptivo de la información
3. Cálculo de probabilidades para entender el problema en México.
4. Planteamiento de hipótesis estadísticas y conclusión sobre ellas para entender el problema en México
5. Estimación de un modelo de regresión logístico, para identificar los determinantes de la inseguridad alimentaria en México

1. PLANTEAMIENTO DEL PROBLEMA DEL CASO



Seguridad Alimentaria: De acuerdo a (Shamah-Levy, 2014) la inseguridad alimentaria se define como la disponibilidad limitada o incierta de alimentos nutricionalmente adecuados e inocuos; o la capacidad limitada e incierta de adquirir alimentos adecuados en formas socialmente aceptables. En México, casi una cuarta parte de la población enfrenta algún nivel de inseguridad alimentaria.



Un centro de salud nutricional está interesado en analizar estadísticamente y probabilísticamente el fenómeno de la inseguridad alimentaria para poder desarrollar un modelo que permita identificar los determinantes socioeconómicos detrás de ella.



Adicionalmente, se plantea el supuesto de que los hogares con menor nivel socioeconómico tienden a gastar más en productos no saludables que las personas con mayores niveles socioeconómicos y se quiere validar la veracidad de tal dicho.

2. ANÁLISIS DESCRIPTIVO DE LA INFORMACIÓN

Extracto de la base de datos: Encuesta Nacional de Salud y Nutrición 2012 del Instituto Nacional de Salud Pública en México.

Nombre de la variable	Tipo	Descripción
Nivel socioeconómico del hogar(nse5f)	Entero(int)	1 "Bajo", 2 "Medio bajo", 3 "Medio", 4 "Medio alto", 5 "Alto"
Zona geográfica (área):	Entero(int)	0 "Zona urbana", 1 "Zona rural"
Número de persona en el hogar (numpeho)	Entero(int)	
Recursos financieros distintos al ingreso laboral(refin):	Entero(int)	0 "no", 1 "sí"
Edad del jefe/a de familia (edadjef)	Entero(int)	
Sexo del jefe/a de familia(sexoje)	Entero(int)	0 "Hombre", 1 "Mujer"
Años de educación del jefe de familia (añosedu)	Entero(int)	
Logaritmo natural del gasto en alimentos saludables (ln_als)	Número(num)	
Logaritmo natural del gasto en alimentos no saludables (ln_alns)	Número(num)	
(Inseguridad alimentaria en el hogar (IA)	Entero(int)	0 "No presenta IA", 1 "Presenta IA"

La base de datos original:

- 40,809 observaciones
- 10 variables

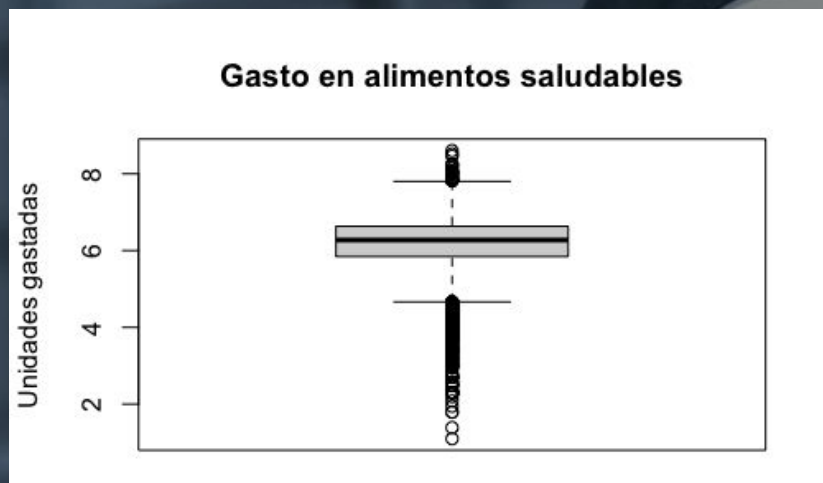
La base de datos después de limpieza:

- 20,529 observaciones
- 12 variables
 - als
 - alns

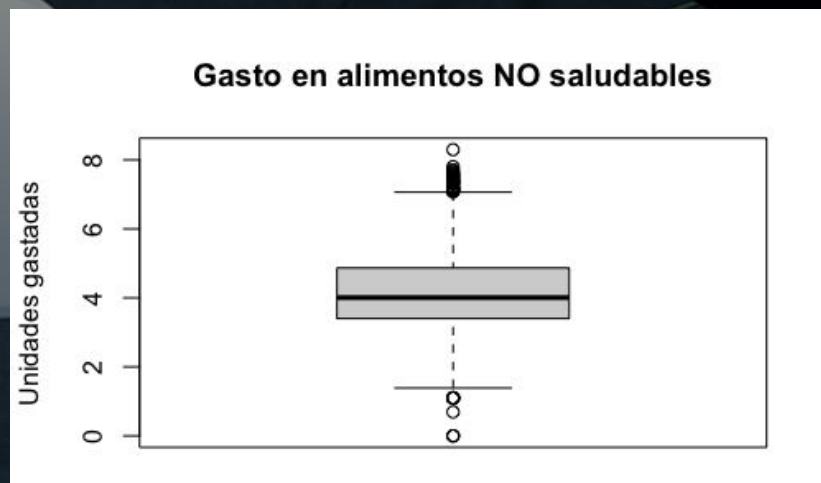
Tabla 1

2. ANÁLISIS DESCRIPTIVO DE LA INFORMACIÓN

Gráficas de los gastos en alimentos saludables (ln_als) y no saludables (ln_alns) en los hogares mexicanos



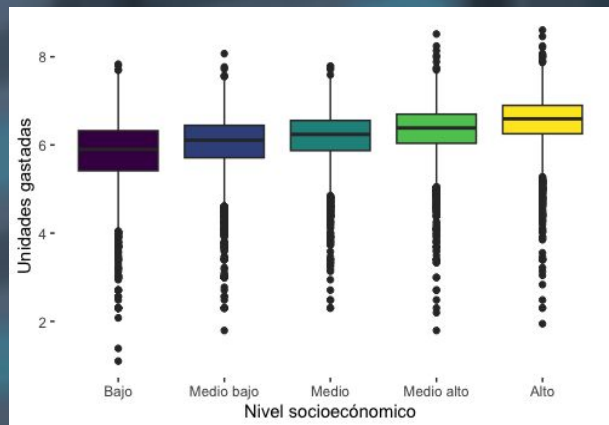
Gráfica 1. Los gasto en alimentos saludables (ln_als) muestran un mínimo: 1.09, primer cuartil: 5.84, media: 6.19 , mediana: 6.27, tercer cuartil: 6.63 y máximo: 8.6



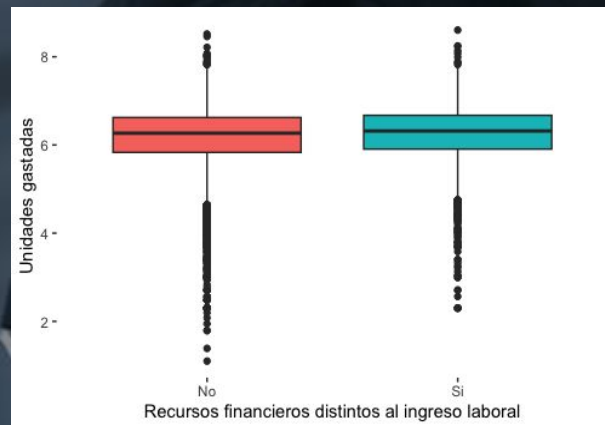
Gráfica 2. Los gastos en alimentos saludables (alns) muestran un mínimo: 0.0, primer cuartil: 3.4, media: 4.11 , mediana: 4.0, tercer cuartil: 4.86 y máximo: 8.29

2. ANÁLISIS DESCRIPTIVO DE LA INFORMACIÓN

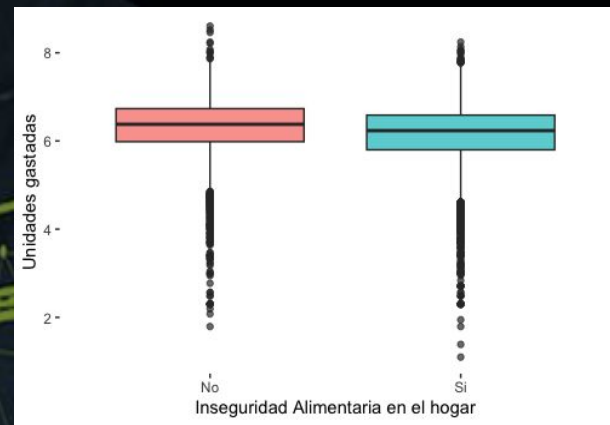
3EDU



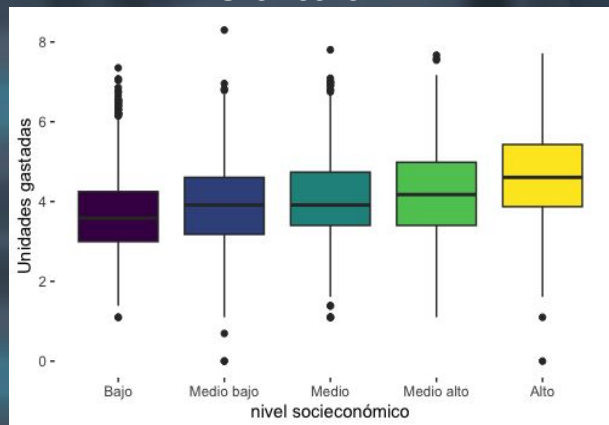
Gráfica 3



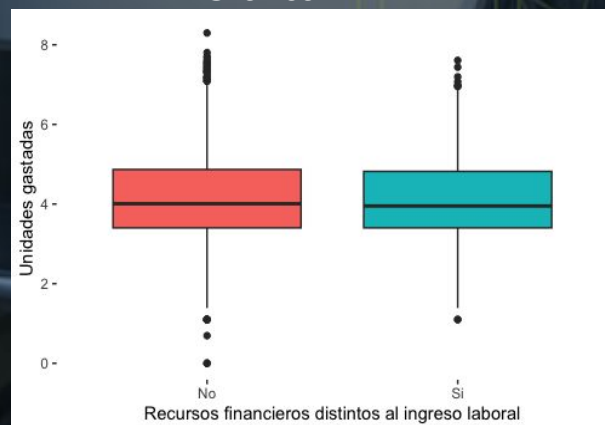
Gráfica 4



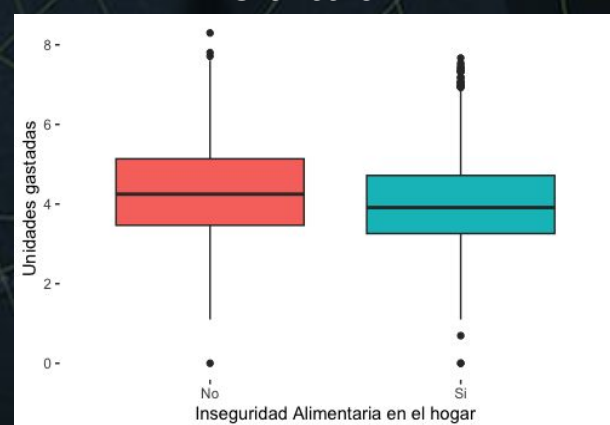
Gráfica 5



Gráfica 6

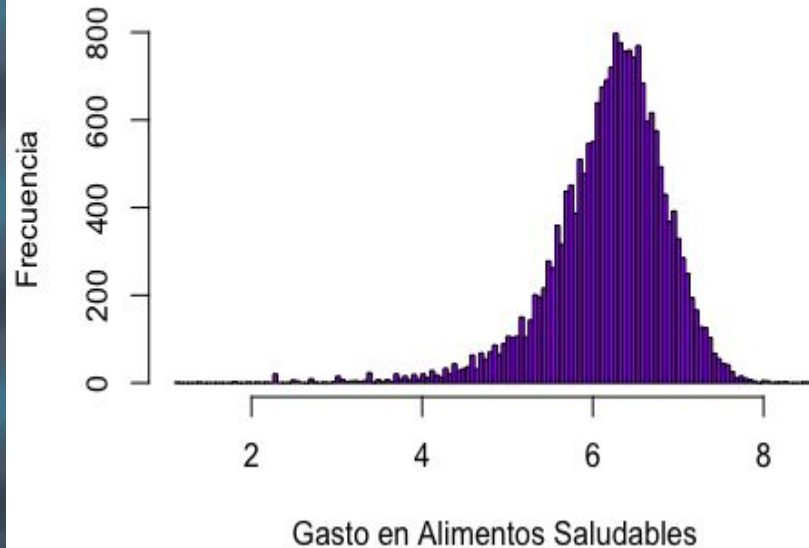


Gráfica 7



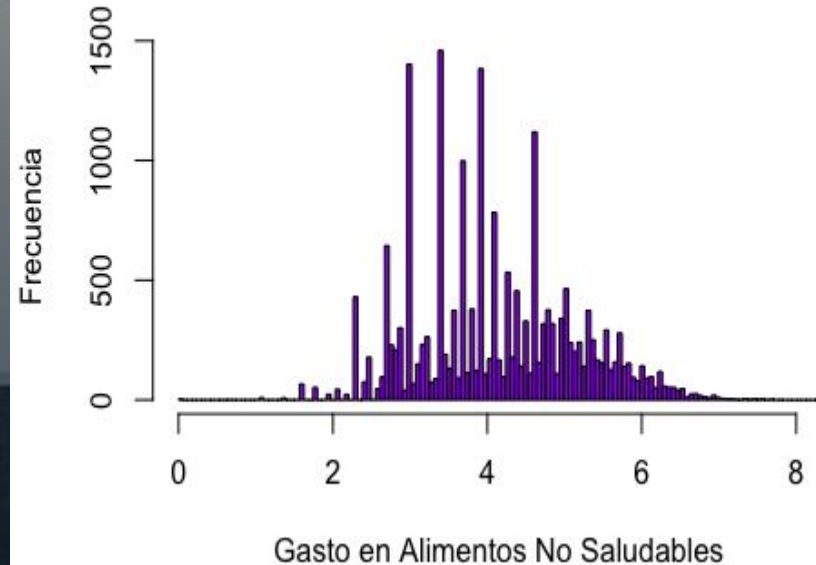
Gráfica 8

Histograma de Gasto en Alimentos Saludables



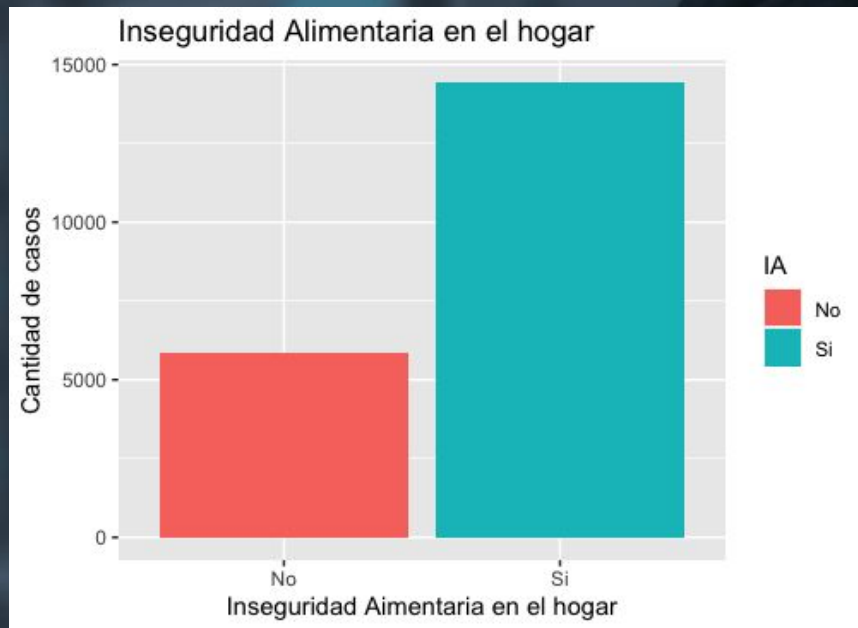
Gráfica 9 Histograma de frecuencia de gasto de alimentos saludables (als). Las medidas de forma indican que tienen un sesgo a la izquierda ($\text{skewness} = -1.19$) y una forma leptocúrtica ($\text{curtosis} = 6.6$).

Histograma de Gasto en Alimentos No Saludables



Gráfica 10. Histograma de frecuencia de gasto de alimentos no saludables (alns). Las medidas de forma indican que es simétrica ($\text{skewness} = 0.24$) y una forma platicúrtica ($\text{curtosis} = 2.57$).

2. ANÁLISIS DESCRIPTIVO DE LA INFORMACIÓN



Gráfica 11. Se muestra que el 71% de las familias que participaron en la muestra presentan inseguridad alimentaria en el hogar.

71% de las familias que participaron en la muestra presentan inseguridad alimentaria en el hogar.

29% de las familias que participaron en la muestra no presenta inseguridad alimentaria.

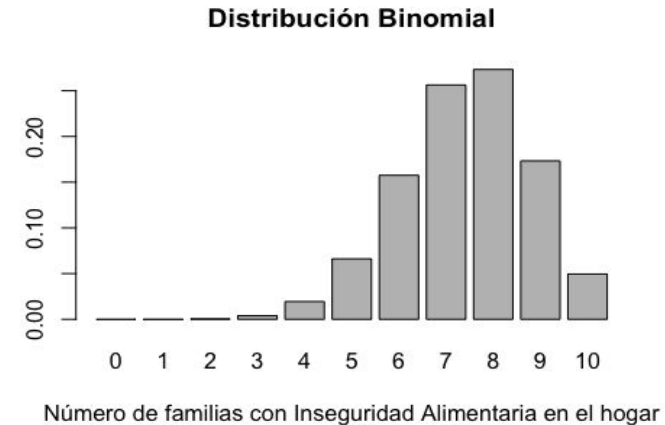
3. CÁLCULO DE PROBABILIDADES PARA ENTENDER EL PROBLEMA EN MÉXICO



Cálculo de probabilidad de que una familia presente o no inseguridad alimentaria mediante una distribución Binomial.

En el supuesto de que el centro de salud nutricional realiza por día 10 estudios a familias para determinar si presentan Inseguridad Alimentaria (IA), ¿Cuál es la probabilidad de que 7 familias presenten IA?

La probabilidad de que 7 familias presenten Inseguridad Alimentaria en el hogar es de 25%



Gráfica 12. Distribución Binomial resultante.
Media = 0.74091, desviación estándar = 0.43815.
Media teórica esperada = 0.74, desviación estándar es = 0.43863

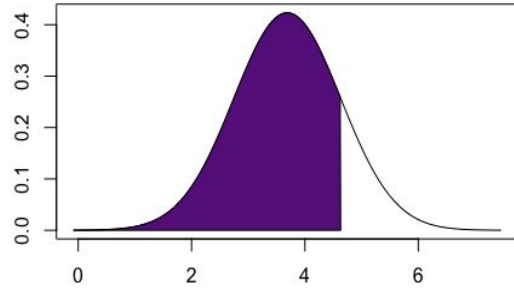
3. CÁLCULO DE PROBABILIDADES PARA ENTENDER EL PROBLEMA EN MÉXICO

Cálculo de probabilidad de los gastos en alimentos no saludables de una familia de acuerdo a su nivel socioeconómico mediante una distribución normal.

A partir de los gastos de alimentos no saludables se realiza un análisis probabilístico, considerando que los gastos promedio de familia en alimentos no saludables se ajustan a una distribución normal.

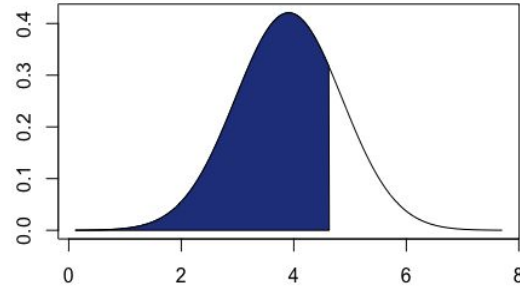
Nivel socioeconómico	Media	Mediana	Moda	Varianza
Bajo	3.68	3.58	2.99	0.94
Medio bajo	3.9	3.91	3.40	0.94
Medio	4.05	3.91	3.40	0.98
Medio alto	4.23	4.17	3.91	1.02
Alto	4.61	4.6	4.60	1.05

Probabilidad $P(X \leq 4.63)$ NSE Bajo



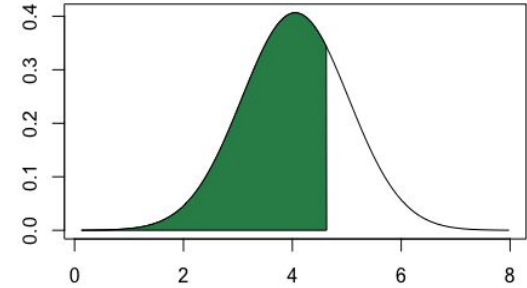
$\mu = 3.68$, $\sigma = 0.94$ y $p = 0.8410548$

Probabilidad $P(X \leq 4.63)$ NSE Medio bajo



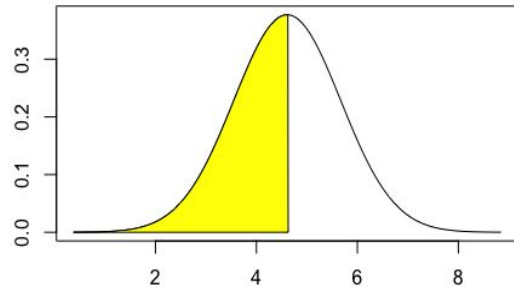
$\mu = 3.09$, $\sigma = 0.94$ y $p = 0.7761948$

Probabilidad $P(X \leq 4.63)$ NSE Medio



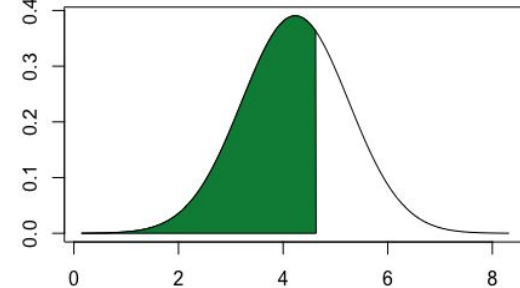
$\mu = 4.05$, $\sigma = 0.98$ y $p = 0.7214145$

Probabilidad $P(X \leq 4.63)$ NSE Alto



$\mu = 4.61$, $\sigma = 1.05$, $p = 0.5071862$

Probabilidad $P(X \leq 4.63)$ NSE Medio alto



$\mu = 4.23$, $\sigma = 1.02$ y $p = 0.6515184$

3. CÁLCULO DE PROBABILIDADES PARA ENTENDER EL PROBLEMA EN MÉXICO

¿Cuál es la probabilidad de que el gasto en alimentos no saludables sea menor o igual a medio salario mínimo en cada nivel socioeconómico?

Nivel socioeconómico	Probabilidad
Bajo	0.84105
Medio bajo	0.77619
Medio	0.72141
Medio alto	0.65151
Alto	0.50718

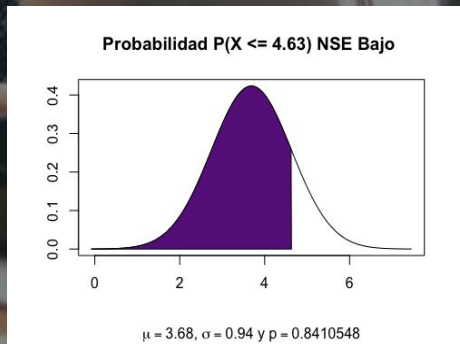
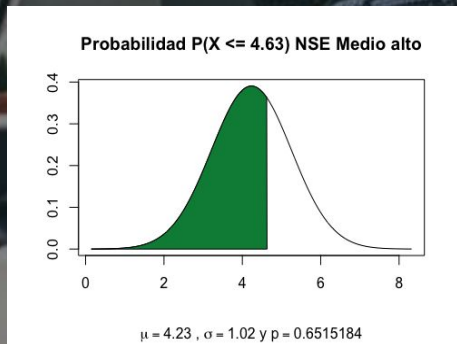


Tabla 3. Resumen de las probabilidades obtenidas de gastos de alimentos no saludables por niveles socioeconómicos

Ejemplos de gráficas de probabilidad de gastos de alimentos no saludables por nivel socioeconómico.

4. HIPÓTESIS ESTADÍSTICAS Y CONCLUSIONES SOBRE EL PROBLEMA EN MÉXICO



*“La mayoría de las personas afirman que los hogares con menor nivel socioeconómico tienden a gastar más en productos **no saludables** que las personas con mayores niveles socioeconómicos y que esto, entre otros determinantes, lleva a que un hogar presente cierta inseguridad alimentaria.”*

Se busca saber si, en promedio, quienes pertenecen al nivel socioeconómico más bajo gastan más en alimentos no saludables que los del nivel más alto.

Ho: $\text{prom_alns_nse5f_bajo} \geq \text{prom_alns_nse5f_alto}$

Ha: $\text{prom_alns_nse5f_bajo} < \text{prom_alns_nse5f_alto}$

A continuación se analiza si existe evidencia estadística para determinar si, en promedio, el nivel socioeconómico tiene efectos sobre el gasto en alimentos no saludables.

Ho: $\text{prom_alns_bajo} = \text{prom_alns_medio_bajo} = \text{prom_alns_medio} = \text{prom_alns_medio_alto} = \text{prom_alns_alto}$

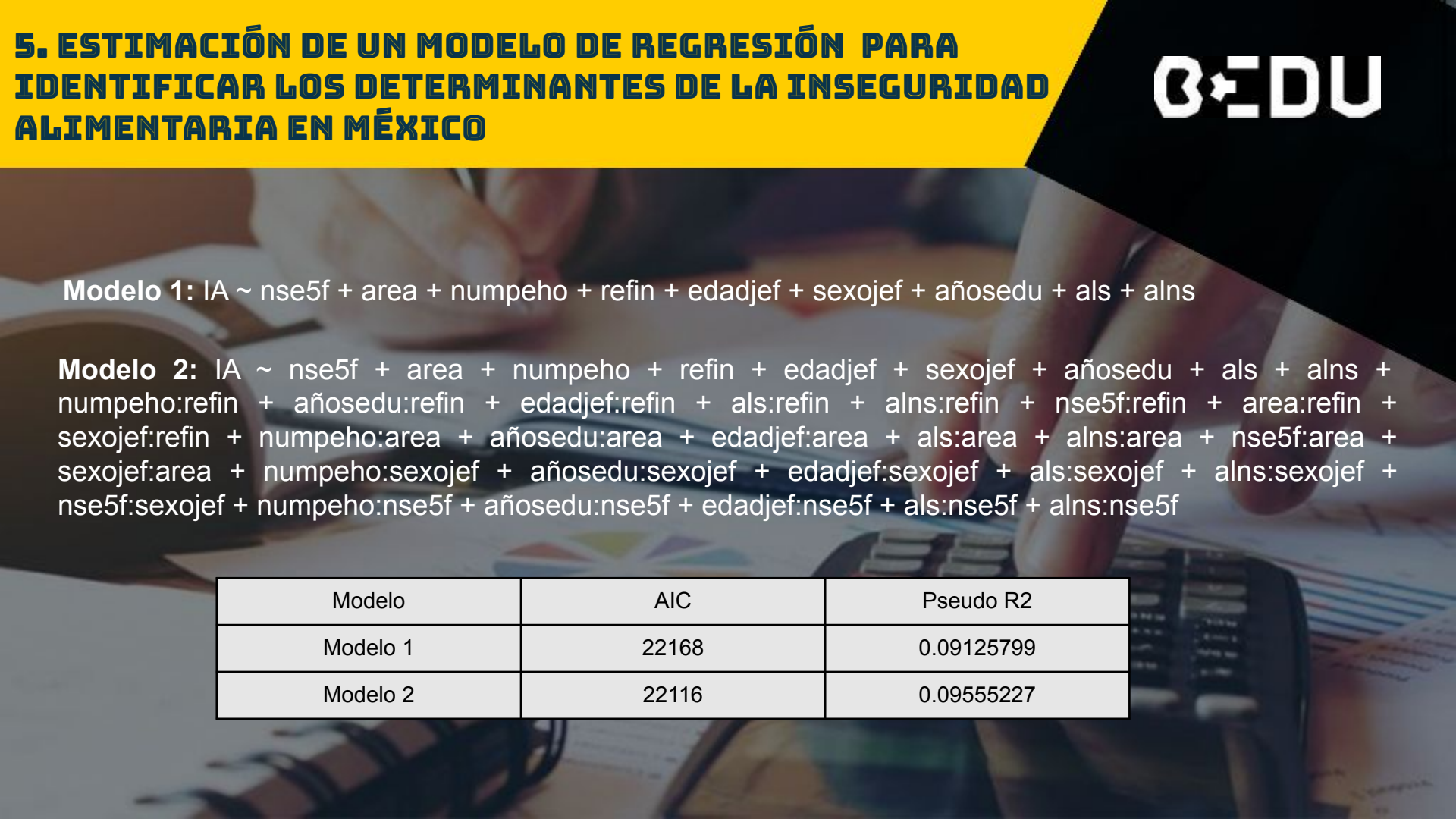
Ha: Al menos uno es diferente

Se analizó también el comportamiento del gasto en alimentos no saludables según la **zona geográfica** y si existen o no **ingresos extra** en el hogar.

4. HIPÓTESIS ESTADÍSTICAS Y CONCLUSIONES SOBRE EL PROBLEMA EN MÉXICO

Los **RESULTADOS**

para alimentos **saludables**
fueron muy similares, excepto
para el ingreso extra.

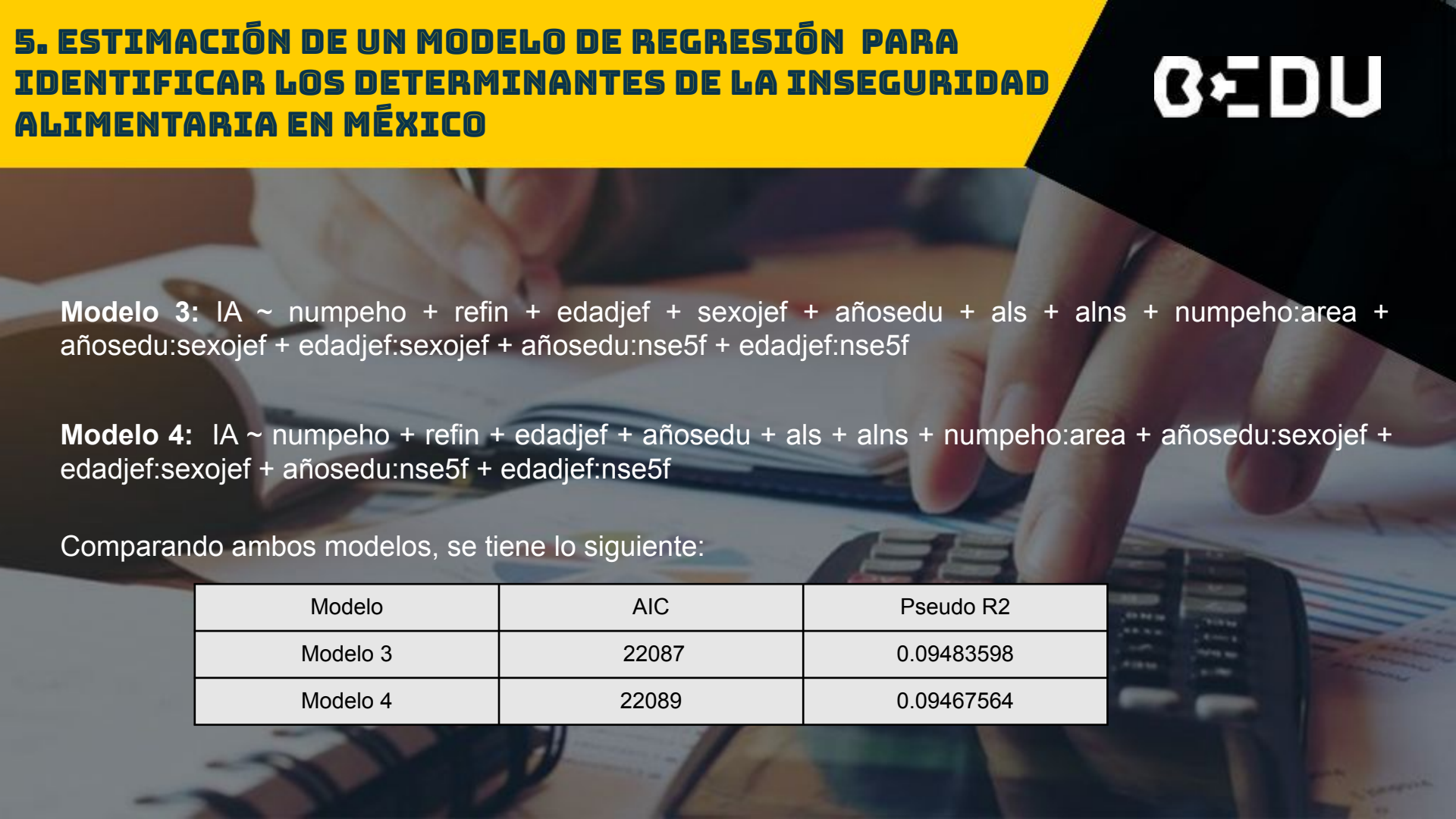


5. ESTIMACIÓN DE UN MODELO DE REGRESIÓN PARA IDENTIFICAR LOS DETERMINANTES DE LA INSEGURIDAD ALIMENTARIA EN MÉXICO

Modelo 1: $IA \sim nse5f + area + numpeho + refin + edadjef + sexojef + añosedu + als + alns$

Modelo 2: $IA \sim nse5f + area + numpeho + refin + edadjef + sexojef + añosedu + als + alns + numpeho:refin + añosedu:refin + edadjef:refin + als:refin + alns:refin + nse5f:refin + area:refin + sexojef:refin + numpeho:area + añosedu:area + edadjef:area + als:area + alns:area + nse5f:area + sexojef:area + numpeho:sexojef + añosedu:sexojef + edadjef:sexojef + als:sexojef + alns:sexojef + nse5f:sexojef + numpeho:nse5f + añosedu:nse5f + edadjef:nse5f + als:nse5f + alns:nse5f$

Modelo	AIC	Pseudo R2
Modelo 1	22168	0.09125799
Modelo 2	22116	0.09555227



5. ESTIMACIÓN DE UN MODELO DE REGRESIÓN PARA IDENTIFICAR LOS DETERMINANTES DE LA INSEGURIDAD ALIMENTARIA EN MÉXICO

Modelo 3: $IA \sim \text{numpeho} + \text{refin} + \text{edadjef} + \text{sexojef} + \text{añosedu} + \text{als} + \text{alns} + \text{numpeho:area} + \text{añosedu:sexojef} + \text{edadjef:sexojef} + \text{añosedu:nse5f} + \text{edadjef:nse5f}$

Modelo 4: $IA \sim \text{numpeho} + \text{refin} + \text{edadjef} + \text{añosedu} + \text{als} + \text{alns} + \text{numpeho:area} + \text{añosedu:sexojef} + \text{edadjef:sexojef} + \text{añosedu:nse5f} + \text{edadjef:nse5f}$

Comparando ambos modelos, se tiene lo siguiente:

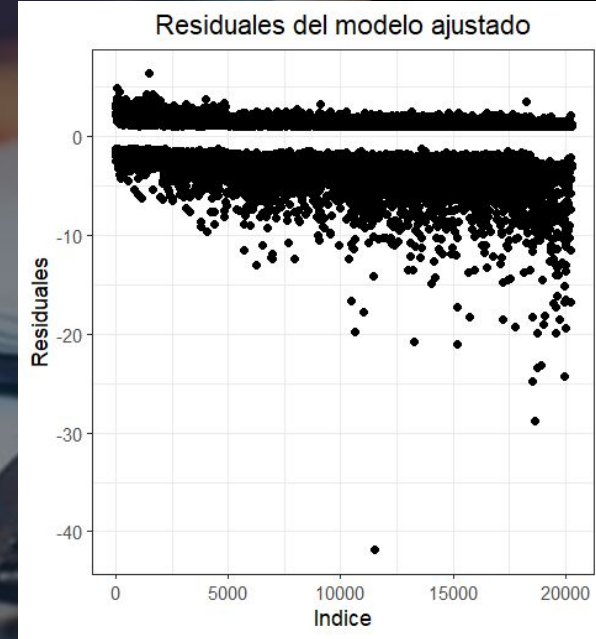
Modelo	AIC	Pseudo R2
Modelo 3	22087	0.09483598
Modelo 4	22089	0.09467564

5. ESTIMACIÓN DE UN MODELO DE REGRESIÓN PARA IDENTIFICAR LOS DETERMINANTES DE LA INSEGURIDAD ALIMENTARIA EN MÉXICO

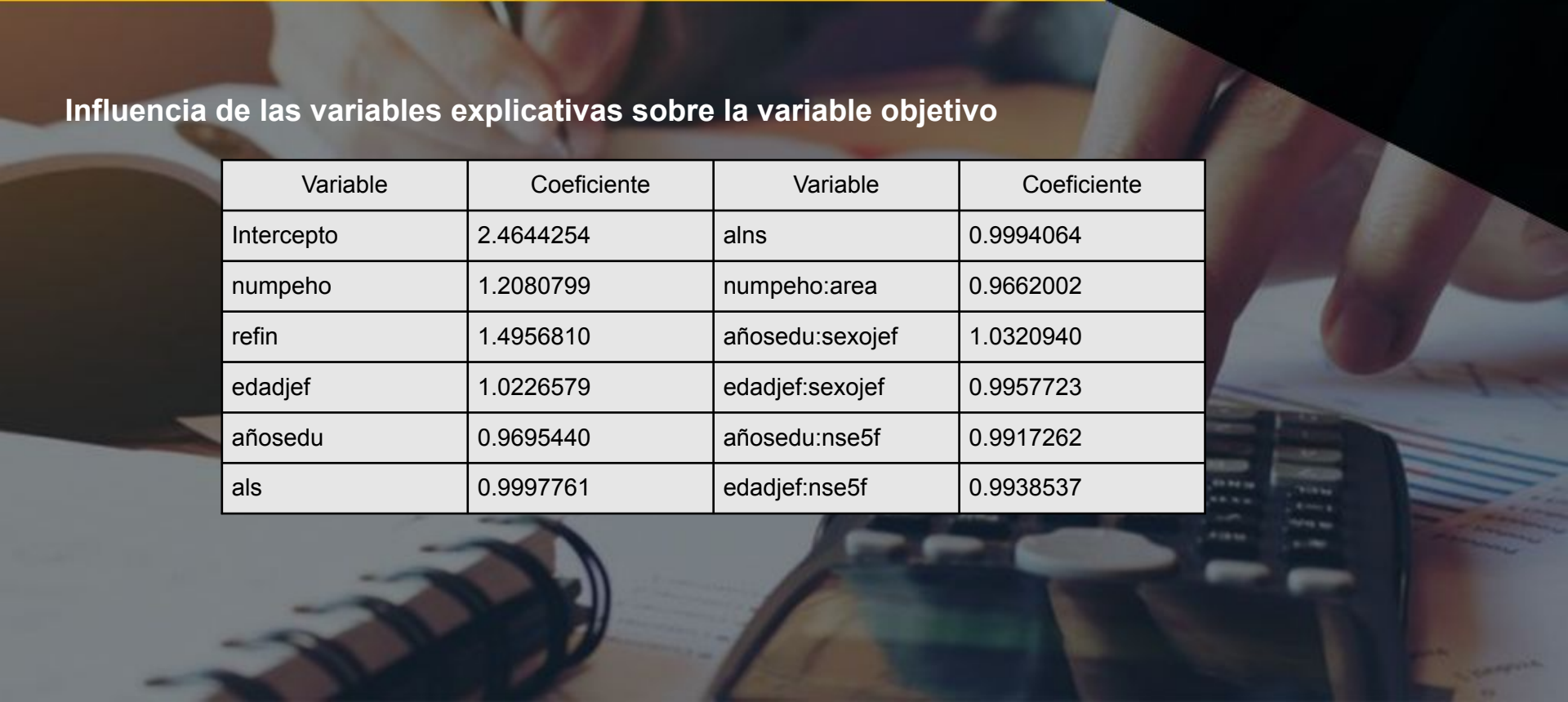


Validación de supuestos modelo 4:

- a) Multicolinealidad: **OK**
- b) Independencia: **OK**
- c) Linealidad: **OK**



En el gráfico de residuales no se ve ningún patrón definido que pudiera indicar un grado de dependencia entre las observaciones.



Influencia de las variables explicativas sobre la variable objetivo

Variable	Coeficiente	Variable	Coeficiente
Intercepto	2.4644254	alns	0.9994064
numpeho	1.2080799	numpeho:area	0.9662002
refin	1.4956810	añosedu:sexojef	1.0320940
edadjef	1.0226579	edadjef:sexojef	0.9957723
añosedu	0.9695440	añosedu:nse5f	0.9917262
als	0.9997761	edadjef:nse5f	0.9938537

5. ESTIMACIÓN DE UN MODELO DE REGRESIÓN PARA IDENTIFICAR LOS DETERMINANTES DE LA INSEGURIDAD ALIMENTARIA EN MÉXICO

Finalmente, se calculan los intervalos de confianza para los coeficientes de los predictores:

Variable	2.5%	97.5%	Variable	2.5%	97.5%
Intercepto	1.9933520	3.0480279	alns	0.9991845	0.9996264
numpeho	1.1820358	1.2349334	numpeho:area	0.9481619	0.9846686
refin	1.3706795	1.6334161	añosedu:sexojef	1.0199859	1.0443211
edadjef	1.0184790	1.0268991	edadjef:sexojef	0.9929893	0.9985863
añosedu	0.9526730	0.9865828	añosedu:nse5f	0.9879187	0.9955657
als	0.9996744	0.9998777	edadjef:nse5f	0.9929008	0.9947976

De este modo, se recomendó un modelo que explica el fenómeno de la inseguridad alimentaria en los hogares, identificando los factores que influyen más y en qué proporción, así como su relación con la variable estudiada.

¡GRACIAS!

