



UNIVERSIDAD DE GRANADA

Facultad de Ciencias

Doble Grado Ingeniería Informática y Matemáticas

ESTADÍSTICA MULTIVARIANTE

TRABAJO FINAL

Autores:

Moisés Jiménez Fernández

Sara Martín Rodríguez

Isabel Morro Tabares

Antonio Javier Rodríguez Romero

Adrián Jaén Fuentes

Curso 2025/26

Resumen

Este informe presenta un análisis estadístico sobre la seguridad alimentaria y el estado nutricional de menores de 5 años en Guatemala, usando datos de la ONG Acción Contra el Hambre. Se integra información de dos bases de datos y se realiza un análisis exploratorio univariante, para identificar patrones de datos faltantes y valores atípicos. Tras esto, se realiza un análisis multivariante, en el que se aplican técnicas de reducción de dimensionalidad como Análisis de Componentes Principales y Análisis Factorial, y técnicas de clasificación como Análisis Clúster y Análisis Discriminante. Como conclusión, los resultados permiten mejorar la toma de decisiones para la creación de intervenciones nutricionales en los entornos más desfavorecidos.

1. Introducción

La desnutrición crónica y la inseguridad alimentaria representan retos estructurales persistentes en Guatemala, afectando de manera crítica el desarrollo de la infancia en las regiones rurales. Para organizaciones como Acción Contra el Hambre, es fundamental contar con diagnósticos precisos que permitan optimizar sus recursos. En este informe, analizamos la relación entre las condiciones del entorno familiar y el estado nutricional de los niños menores de cinco años, utilizando un enfoque estadístico. Nuestro punto de partida fue la integración de dos fuentes de datos complementarias: la base de Hogares y la de Menores de 59 meses. Mediante la creación de una clave común basada en el identificador del hogar, unificamos un dataset de 705 registros de menores vinculados a sus contextos socioeconómicos. En la fase inicial, llevamos a cabo un análisis exploratorio univariante para asegurar la calidad de la información. Este proceso incluyó la detección de valores atípicos mediante el rango intercuartílico y un estudio detallado de los datos faltantes. Dada la complejidad del problema, se han aplicado técnicas multivariantes que nos permiten ver en más profundidad los detalles de este problema. Por un lado, empleamos métodos de reducción de dimensionalidad como el Análisis de Componentes Principales (PCA) y el Análisis Factorial, que nos permiten sintetizar múltiples indicadores (como las escalas ELCSA y CSI) en factores latentes de vulnerabilidad. Por otro lado, aplicamos técnicas de clasificación como el Análisis Clúster y el Análisis Discriminante para segmentar a la población e identificar los perfiles de riesgo más críticos. El objetivo de este trabajo es realizar un análisis estadístico multivariante integral que nos permita caracterizar los niveles de vulnerabilidad nutricional e identificar perfiles de riesgo específicos para facilitar la toma de decisiones estratégicas en los programas de ayuda humanitaria.

2. Materiales y Métodos

2.1. Descripción de la base de datos

Nuestra base de datos final es el resultado de unir dos bases de datos. El dataset resultante contiene una muestra de 705 menores de 5 años, vinculando a cada niño o niña las características socioeconómicas y de seguridad alimentaria de su hogar. A continuación mostramos estadísticas descriptivas básicas de las variables de la base de datos.

Cuadro 1: Estadísticos Descriptivos (Todas las variables)

Variable	Media	D.Típica
ID_UNICO	1752069,407	526480,899
MONTHS	29,6	17,366
EDAD_304	12,0139	8,5033
AMAMANT_305	0,9722	0,1646
TIEMPO_1RA_VEZ_306	58,2607	76,5349
AMAMANT_HOY_307	0,8964	0,3052
AMAMANT_MESES_308	13,6923	6,2211
SUPL_VIT_A_309	0,5233	0,5004
AGUA_310_A	0,6446	0,4795
FORM_310_B1	0,101	0,3019
FORM_VEZ_310_B2	0,2509	1,0902
LECH_310_C1	0,0976	0,2972
LECH_310_C2	0,2624	1,0944
ATOL_310_D	0,5331	0,4998
JUGO_310_E	0,2125	0,4098
CAFE_310_F	0,4669	0,4998
LIQU_310_G	0,223	0,417
PACH_311	0,3693	0,4835
PURE_312_A	0,6434	0,4798
FRIJO_312_B	0,5559	0,4977
LACT_312_C	0,1439	0,3516
CARN_312_D	0,1088	0,3119
HUEV_312_E	0,3042	0,4609
VERD_312_F	0,3776	0,4856
FRUT_312_G	0,2727	0,4461
GRAS_312_H	0,3182	0,4666
AZUC_312_I	0,5874	0,4932
COMIDAS_VEZ_313	2,2657	1,6435
GA1	0,6434	0,4798
GA2	0,5559	0,4977
GA3	0,1439	0,3516
GA4	0,1088	0,3119
GA5	0,3042	0,4609
GA6	0,3776	0,4856
GA7	0,2727	0,4461
GA8	0,5874	0,4932
GA9	0,3182	0,4666
DA_MIN_SC	3,3112	2,4303
ID_HOGAR	17520,6312	5264,8157
ESTRATO	1,5994	0,4904
HOGAR	9,2798	5,2149
TAMANHO_HOG	7,125	3,4017
HOGAR_A_y	0,8618	0,3453

Continúa en la siguiente página...

Cuadro 1 – continuación de la página anterior

Variable	Media	D.Típica
HOGAR_B	0,3533	0,4783
HOGAR_C	0,5472	0,4981
ADULTOS_D	0,5623	0,4965
ADULTOS_E	0,2379	0,4261
ADULTOS_F	0,4243	0,4946
ADULTOS_G	0,2643	0,4413
ADULTOS_H	0,1681	0,3742
MENORES_I	0,5333	0,4992
MENORES_J	0,5457	0,4983
MENORES_K	0,2026	0,4022
MENORES_L	0,3476	0,4766
MENORES_M	0,3052	0,4608
MENORES_N	0,1891	0,3919
MENORES_O	0,1243	0,3301
MAIZ_A	6,9858	11,3413
FRIJOL_B	3,604	8,1533
IES_A	2,2751	2,6349
IES_B	0,6933	1,3358
IES_C	1,2371	1,8755
IES_D	0,7692	1,4834
IES_E	0,2246	0,8018
IES_F	0,2165	0,8978
IES_G	0,057	0,4496
IES_H	1,3252	2,171
IES_I	1,1492	2,0344
IES_J	0,2892	0,913
IES_K	0,2	0,9761
IES_L	0,7379	1,6754
IES_M	0,0798	0,3953
EMPLEO_JORNAL_A1	0,698	0,4595
EMPLEO_OTRO_B1	0,2692	0,4439
COMERCIO_C1	0,0855	0,2798
VENTA_PROD_D1	0,0954	0,294
CLUSTER_A_y	15,1136	9,0404
HOGAR_Z	9,2798	5,2149
NUM_FUENT_INGRES	1,1481	0,4043
MAIZ_CLASF_2	1,1239	1,1921
FRIJOL_CLASF_2	0,6268	1,0164
stress_coping_OCC	1,5378	0,8437
crisiscoping_OCC	1,5339	1,5007
emergencycoping_OCC	1,1314	1,8028
stress_coping_ORI	1,5378	0,8437
crisiscoping_ORI	1,5339	1,5007
emergencycoping_ORI	1,1086	1,7916

Cuadro 2: SEX

Valor	Frec.	% Válido
f	353	50.07 %
m	349	49.50 %
(Perdidos)	3	0.43 %

Cuadro 3: DIARREA

Valor	Frec.	% Válido
0.0	500	71.33 %
1.0	201	28.67 %

Cuadro 4: FIEBRE

Valor	Frec.	% Válido
0.0	387	55.13 %
1.0	315	44.87 %

Cuadro 5: INFEC.RE

Valor	Frec.	% Válido
1.0	412	58.69 %
0.0	290	41.31 %

Cuadro 6: DESPARAS

Valor	Frec.	% Válido
0.0	361	59.28 %
1.0	248	40.72 %

Cuadro 7: VACC_SAR

Valor	Frec.	% Válido
1.0	458	72.24 %
0.0	151	23.82 %
2.0	25	3.94 %

Cuadro 8: CLUSTER_A_X

Cluster	Frec.	%
09	39	5.53 %
14	30	4.26 %
07	29	4.11 %
24	29	4.11 %
05	28	3.97 %
20	27	3.83 %
02	27	3.83 %
11	26	3.69 %
01	25	3.55 %
12	25	3.55 %
19	25	3.55 %
13	24	3.40 %
08	24	3.40 %
06	23	3.26 %
04	23	3.26 %
17	22	3.12 %
03	22	3.12 %
30	20	2.84 %
15	20	2.84 %
10	20	2.84 %
18	20	2.84 %
32	19	2.70 %
23	19	2.70 %
25	19	2.70 %
29	19	2.70 %
27	17	2.41 %
28	15	2.13 %
22	14	1.99 %
16	13	1.84 %
26	13	1.84 %
21	13	1.84 %
31	13	1.84 %
(V)	3	0.43 %

Cuadro 9: HOGAR_A_X

Hogar	Frec.	%
02	47	6.67 %
14	45	6.38 %
09	44	6.24 %
16	44	6.24 %
05	43	6.10 %
06	42	5.96 %
04	41	5.82 %
03	40	5.67 %
15	40	5.67 %
08	39	5.53 %
01	39	5.53 %
12	37	5.25 %
10	37	5.25 %
07	35	4.96 %
17	34	4.82 %
18	34	4.82 %
13	32	4.54 %
11	29	4.11 %
(V)	3	0.43 %

Cuadro 10: VACC_SAR_AGRUP

Valor	Frec.	%	Válido
1.0	483	76.18 %	
0.0	151	23.82 %	

Cuadro 11: GAM

Valor	Frec.	%	Válido
0.0	681	98.70 %	
1.0	9	1.30 %	

<u>Cuadro 12: SUNTING</u>		
Valor	Frec.	% Válido
1.0	380	55.07 %
0.0	310	44.93 %

<u>Cuadro 13: UNDERWEIGHT</u>		
Valor	Frec.	% Válido
0.0	561	81.07 %
1.0	131	18.93 %

<u>Cuadro 14: SEXO_303</u>		
Sexo	Frec.	% Válido
(Vacío)	417	59.15 %
F	150	21.28 %
M	138	19.57 %

<u>Cuadro 15: LACT_TEMP</u>		
Valor	Frec.	% Válido
1.0	211	73.26 %
0.0	77	26.74 %

<u>Cuadro 16: LACT_EXCL</u>		
Valor	Frec.	% Válido
0.0	246	85.71 %
1.0	41	14.29 %

<u>Cuadro 17: INTRO_ALIM</u>		
Valor	Frec.	% Válido
1.0	216	75.52 %
0.0	70	24.48 %

<u>Cuadro 18: DA_MIN_CLASF</u>		
Valor	Frec.	% Válido
0.0	143	50.00 %
1.0	143	50.00 %

<u>Cuadro 19: FRQ_MIN_COM</u>		
Valor	Frec.	% Válido
0.0	231	80.77 %
1.0	55	19.23 %

<u>Cuadro 20: DIET_MIN_ACEP</u>		
Valor	Frec.	% Válido
0.0	243	84.97 %
1.0	43	15.03 %

<u>Cuadro 21: LACT_Y_SSOLD</u>		
Valor	Frec.	% Válido
0.0	282	98.26 %
1.0	5	1.74 %

<u>Cuadro 22: LACT_Y_LIQ</u>		
Valor	Frec.	% Válido
0.0	282	98.26 %
1.0	5	1.74 %

<u>Cuadro 23: LACT_Y_LECH</u>		
Valor	Frec.	% Válido
0.0	283	98.61 %
1.0	4	1.39 %

<u>Cuadro 24: LACT_Y_AGUA</u>		
Valor	Frec.	% Válido
0.0	278	96.86 %
1.0	9	3.14 %

<u>Cuadro 25: NO_LACT</u>		
Valor	Frec.	% Válido
0.0	280	97.22 %
1.0	8	2.78 %

Cuadro 26: INT_EDAD_2MES

Valor	Frec.	%
7.0	33	11.50 %
6.0	32	11.15 %
8.0	28	9.76 %
1.0	27	9.41 %
5.0	24	8.36 %
11.0	24	8.36 %
12.0	23	8.01 %
9.0	21	7.32 %
10.0	21	7.32 %
2.0	20	6.97 %
3.0	19	6.62 %
4.0	15	5.23 %

Cuadro 27: DPTO

DPTO	Frec.	% Válido
QUI	182	25.85 %
CHQ	103	14.63 %
JUT	87	12.36 %
TOT	72	10.23 %
ZAC	71	10.09 %
HUE	59	8.38 %
SMA	52	7.39 %
BVE	42	5.97 %
JAL	18	2.56 %
ELP	18	2.56 %

Cuadro 28: HH_MENORES

Valor	Frec.	% Válido
1.0	703	99.86 %
0.0	1	0.14 %

Cuadro 29: HH_MENORES_5A

Valor	Frec.	% Válido
1.0	703	99.86 %
0.0	1	0.14 %

Cuadro 30: HH_MENORES_2A

Valor	Frec.	% Válido
1.0	438	62.22 %
0.0	266	37.78 %

Cuadro 31: ELCSA_ADULT_CLASF

Valor	Frec.	%
2.0	293	42.65 %
3.0	228	33.19 %
4.0	97	14.12 %
1.0	69	10.04 %

Cuadro 32: ELCSA_ADULT_SC

Valor	Frec.	%
1.0	157	22.85 %
4.0	109	15.87 %
3.0	86	12.52 %
0.0	69	10.04 %
8.0	67	9.75 %
5.0	66	9.61 %
6.0	53	7.71 %
2.0	50	7.28 %
7.0	30	4.37 %

Cuadro 33: ELCSA_MEN18_CLASF

Valor	Frec.	%
2.0	293	42.96 %
3.0	203	29.77 %
4.0	119	17.45 %
1.0	67	9.82 %

Cuadro 34: ELCSA_MEN18_SC

Valor	Frec.	%
1.0	148	21.70 %
5.0	68	9.97 %
0.0	67	9.82 %
6.0	63	9.24 %
8.0	43	6.30 %
15.0	42	6.16 %
9.0	34	4.99 %
7.0	33	4.84 %
10.0	30	4.40 %
4.0	29	4.25 %
11.0	26	3.81 %
3.0	25	3.67 %
2.0	23	3.37 %
13.0	21	3.08 %
12.0	17	2.49 %
14.0	13	1.91 %

Cuadro 35: ELCSA_MEN5_SC

Valor	Frec.	%
1.0	148	21.70 %
5.0	68	9.97 %
0.0	67	9.82 %
6.0	63	9.24 %
8.0	43	6.30 %
15.0	42	6.16 %
9.0	34	4.99 %
7.0	33	4.84 %
10.0	30	4.40 %
4.0	29	4.25 %
11.0	26	3.81 %
3.0	25	3.67 %
2.0	23	3.37 %
13.0	21	3.08 %
12.0	17	2.49 %
14.0	13	1.91 %

Cuadro 36: ELCSA_MEN5_CLASF

Valor	Frec.	%
2.0	293	42.96 %
3.0	203	29.77 %
4.0	119	17.45 %
1.0	67	9.82 %

Cuadro 37: ELCSA_MEN2_SC

Valor	Frec.	%
1.0	148	21.70 %
5.0	68	9.97 %
0.0	67	9.82 %
6.0	63	9.24 %
8.0	43	6.30 %
15.0	42	6.16 %
9.0	34	4.99 %
7.0	33	4.84 %
10.0	30	4.40 %
4.0	29	4.25 %
11.0	26	3.81 %
3.0	25	3.67 %
2.0	23	3.37 %
13.0	21	3.08 %
12.0	17	2.49 %
14.0	13	1.91 %

Cuadro 38: ELCSA_MEN2_CLASF

Valor	Frec.	%
2.0	293	42.96 %
3.0	203	29.77 %
4.0	119	17.45 %
1.0	67	9.82 %

Cuadro 39: MAIZ_CLASF

Valor	Frec.	%	Válido
1.0	474	67.52 %	
2.0	103	14.67 %	
3.0	102	14.53 %	
4.0	23	3.28 %	

Cuadro 40: FRIJOL_CLASF

Valor	Frec.	%	Válido
1.0	571	81.34 %	
2.0	70	9.97 %	
3.0	52	7.41 %	
4.0	9	1.28 %	

Cuadro 41: IES_A1

Valor	Frec.	%	Válido
1.0	441	63.18 %	
0.0	257	36.82 %	

Cuadro 42: IES_B2

Valor	Frec.	%	Válido
0.0	483	68.90 %	
1.0	218	31.10 %	

Cuadro 43: IES_C3

Valor	Frec.	%	Válido
0.0	381	54.43 %	
1.0	319	45.57 %	

Cuadro 44: IES_D4

Valor	Frec.	%	Válido
0.0	479	68.23 %	
1.0	223	31.77 %	

Cuadro 45: IES_E5

Valor	Frec.	%	Válido
0.0	570	87.69 %	
1.0	80	12.31 %	

Cuadro 46: IES_F6

Valor	Frec.	%	Válido
0.0	646	92.02 %	
1.0	56	7.98 %	

Cuadro 47: IES_G7

Valor	Frec.	%	Válido
0.0	681	97.01 %	
1.0	21	2.99 %	

Cuadro 48: IES_H8

Valor	Frec.	%	Válido
0.0	423	60.60 %	
1.0	275	39.40 %	

Cuadro 49: IES_I9

Valor	Frec.	%	Válido
0.0	459	65.85 %	
1.0	238	34.15 %	

Cuadro 50: IES_J10		
Valor	Frec.	% Válido
0.0	614	87.46 %
1.0	88	12.54 %

Cuadro 51: IES_K11		
Valor	Frec.	% Válido
0.0	652	93.14 %
1.0	48	6.86 %

Cuadro 52: IES_L12		
Valor	Frec.	% Válido
0.0	529	75.36 %
1.0	173	24.64 %

Cuadro 53: IES_M13		
Valor	Frec.	% Válido
0.0	668	95.16 %
1.0	34	4.84 %

Cuadro 54: TAMÑO_HOG_CLASF

Valor	Frec.	%
5.0	102	14.49 %
7.0	102	14.49 %
4.0	101	14.35 %
8.0	86	12.22 %
6.0	83	11.79 %
11.0	71	10.09 %
10.0	56	7.95 %
9.0	56	7.95 %
3.0	43	6.11 %
2.0	4	0.57 %

Cuadro 55: CSI_final

Valor	Frec.	% Válido
2.0	204	29.06 %
4.0	198	28.21 %
3.0	151	21.51 %
1.0	149	21.23 %

2.2. Métodos estadísticos

Para abordar los objetivos planteados, hemos estructurado nuestro flujo de trabajo en tres fases, aplicando diversas herramientas de análisis univariante y multivariante. En la primera fase, tras la integración y limpieza de las bases de datos (eliminación de columnas no relevantes), hemos llevado a cabo un análisis exploratorio univariante. En este punto, empleamos estadísticos de tendencia central, dispersión y forma (asimetría y curtosis). Para la gestión de la calidad de los datos, identificamos el porcentaje de valores faltantes y determinamos si su patrón es aleatorio mediante pruebas de homogeneidad, específicamente el test t-Student para variables cuantitativas y el test Chi-cuadrado para categóricas. Además, para la detección de outliers, aplicamos el método del rango intercuartílico. Por otro lado, previo a la aplicación de técnicas de reducción de dimensionalidad o multivariantes en general, se comprueba la verificación de hipótesis necesarias para la aplicación de estos métodos. Principalmente, estas pruebas van dirigidas al estudio de la normalidad de cada componente de estudio, la normalidad multivariante como conjunto y a la correlación que mantienen estas variables entre ellas. En este caso, los métodos aplicados son: el análisis de gráficas Q-Q y el test de Sapiro-Wilk para la normalidad de las variables; el test de Bartlett para la correlación; el test de Henze-Zirkler para la normalidad multivariante. En la segunda fase, nos centramos en la reducción de

dimensionalidad para simplificar la estructura de los datos. Posteriormente, realizamos el Análisis de Componentes Principales (PCA) y el Análisis Factorial, utilizando el gráfico de sedimentación y la regla de Kaiser para la selección de los factores y componentes más significativos. Como paso previo, evaluamos la distribución de la muestra mediante el test de normalidad multivariante de Henze-Zirkler. En la tercera fase, procedimos a la clasificación y segmentación de la muestra. Aplicamos el Análisis Cluster mediante el algoritmo K-means, determinando el número óptimo de grupos a través del método del codo. Finalmente, para validar la consistencia de los grupos identificados y entender qué variables contribuyen más a su diferenciación, empleamos Análisis Discriminante. Todo el análisis ha sido desarrollado usando el lenguaje de programación Python, apoyándonos en librerías especializadas como scikit-learn, factor_analyzer y pingouin.

3. Resultados

En cuanto a la verificación de las hipótesis necesarias para los distintos métodos que se aplicarán, se llega a la conclusión de manera bastante fehaciente de que, por un lado, las variables de estudio no siguen en su amplia mayoría una distribución normal univariante y, coherentemente, en su conjunto no se distribuyen según una normal multivariante. Sin embargo, debido a la robustez de los métodos que se aplicarán posteriormente y al gran tamaño muestral con el que contamos ($N > 700$), podremos aplicar las técnicas descritas previamente suponiendo cierta la hipótesis de normalidad.

Por otro lado, en cuanto a la correlación de las variables, una vez obtenemos los campos puramente cuantitativos, obtenemos que entre ellos existe una correlación.

Por lo tanto, podemos concluir que será válida la aplicación de los métodos y técnicas descritas en el apartado anterior y, consecuentemente, sus resultados.

En cuanto a la reducción de la dimensionalidad, tras validar la correlación (test de Bartlett significativo, con un p-valor $p < 0,001$), se procedió a la extracción de las componentes principales. Siguiendo la Regla de Kaiser se seleccionaron 4 componentes principales que explican el 52.49 % de la varianza total. La primera componente explica el 15.42 % de la varianza y presenta cargas elevadas en EDAD304 y MONTHS. La segunda (13.45 %) está dominada por MAIZA y FRIJOLB. La tercera (11.9 %) tiene su mayor peso en TIEMPO1RAVEZ, y la cuarta (11.72 %) relaciona de forma inversa TAMANHOHOG (-0.51) con CLUSTERAY (0.69)

En cuanto a la clasificación, el algoritmo K-Means ($k=2$) segmentó la muestra en dos grupos. Al realizar la validación cruzada con la variable respuesta (SUNTING), la tabla de contingencia mostró una distribución donde los clústeres no representan una separación unívoca respecto a la variable respuesta.

Respecto al análisis discriminante, los modelos LDA Y QDA se evaluaron para predecir la variable SUNTING. La regresión lineal obtuvo un Accuracy del 55.67 %, con un Recall del 77.68 %, pero una baja especificidad (28.57 %). La regresión cuadrática mostró un rendimiento similar, con un Accuracy del 57.14 %. Además, el área bajo de la curva para ambos modelos estuvo entre 0.55 y 0.57, indicando una capacidad predictiva limitada para esta variable respuesta.

4. Discusión

Como se dijo en la introducción, el objetivo de este trabajo era realizar un análisis multivariante completo para caracterizar los niveles de vulnerabilidad nutricional e identificar perfiles de riesgo específicos. Tras analizar los resultados, podemos discutir en qué medidas se han alcanzado estos objetivos basándonos en la evidencia obtenida.

En primer lugar, el objetivo de caracterizar la vulnerabilidad se ha logrado a través de las técnicas de reducción de dimensionalidad. Los resultados muestran una realidad compleja: la vulnerabilidad depende de varios factores. El hecho de que la primera componente principal y el factor 2 estén dominados por la edad (EDAD304 y MONTHS) confirma que el crecimiento es una característica importante de la muestra. Sin embargo, la aparición de un factor centrado en las prácticas de salud iniciales (el factor 1 y la componente principal 3, dominadas por TIEMPO1RAVEZ306) demuestra que las acciones tomadas en las primeras horas de vida tienen un peso estadístico independiente de la situación económica o demográfica del hogar. Por otra parte el factor de subsistencia (maíz y frijol) confirma de dependencia estructural de estos hogares hacia los granos básicos.

En segundo lugar, respecto a identificar perfiles de riesgo mediante modelos de clasificación, los resultados de los modelos LDA y QDA ofrecen información valiosa. Aunque la precisión global es de en torno al 56 %, la alta sensibilidad (77.68 %) obtenida indica que el modelo es eficaz detectando niños que realmente sufren desnutrición crónica. Para una ONG, es preferible tener un modelo que “alerte” de más (baja especificidad) a uno que ignore casos críticos (falsos negativos). No obstante, la baja especificidad y el AUC de 0.57 sugieren que la desnutrición crónica es un fenómeno multicausal; las variables de seguridad alimentaria y reservas son necesarias, pero no suficientes por sí solas para predecir el estado clínico del menor.

Finalmente, la validación Clúster refuerza esta conclusión. Al observar la tabla cruzada, comprobamos que los grupos formados por el algoritmo no discriminan de forma perfecta entre niños sanos y desnutridos. Esto indica que los hogares con perfiles socioeconómicos similares pueden presentar resultados en cuanto a la nutrición distintos. Un hallazgo llamativo es la baja communalidad del tamaño del hogar (TAMANHOHOG), lo que sugiere que, en esta muestra específica, tener una familia numerosa no es predictor directo de peores prácticas de lactancia o menor reserva de granos.

En conclusión, se han alcanzado los objetivos de caracterización y segmentación, proporcionando a Acción Contra el Hambre una base técnica para entender que el riesgo nutricional en Guatemala debe abordarse desde varios flancos: priorizando prácticas de lactancia temprana y el acceso a granos básicos, pero sabiendo la complejidad que impide una clasificación binaria del estado de salud del menor.

5. Conclusión

Este trabajo nos ha permitido dar una visión sobre la situación nutricional y seguridad alimentaria en las regiones estudiadas de Guatemala. Hemos conseguido transformar datos complejos en información útil para las necesidades de la organización Acción Contra el Hambre. Como punto fuerte del estudio, destacamos la solidez del enfoque multivariante. La combinación de las técnicas de reducción de dimensionalidad y de clasificación ha permitido identificar patrones de vulnerabilidad que el análisis univariante no puede

detectar. Sin embargo, el trabajo presenta limitaciones, como la falta de calidad de los datos. Estos presentan patrones de datos faltantes no aleatorios y valores atípicos en las mediciones, lo que sugiere posibles errores en la recolección de los datos en campo. Como mejoras futuras, proponemos la codificación de las fuentes de ingreso del Módulo V para profundizar en el perfil socioeconómico, así como el desarrollo de un estudio que evalúe la estabilidad de los clústeres identificados.