



UNIVERSIDAD
NACIONAL
DE TUCUMÁN



FACULTAD DE
CIENCIAS
ECONOMICAS

"2020 - Año del General Manuel Belgrano"

“MEDICION DE LA POBREZA. ENFOQUE MULTIDIMENSIONAL”

Tesis para optar por el grado de Doctor en Estadística

Autora:

ADRIANA FÁTIMA PANICO

Tucumán, 2020



UNIVERSIDAD
NACIONAL
DE TUCUMÁN



FACULTAD DE
CIENCIAS
ECONÓMICAS

"2020 - Año del General Manuel Belgrano"

COMISIÓN DE SUPERVISIÓN:

Dr. SHRIKANT BANGDIWALA. Director

Dra. VIVIANA BEATRIZ LENCINA. Directora Asociada

Dr. MANUEL LUÍS CORDOMÍ

Dr. ESTEBAN ALBERTO NICOLINI

ÍNDICE GENERAL

RESUMEN

INTRODUCCIÓN	I
1. CAPÍTULO. POBREZA: CONCEPTOS Y METODOLOGÍAS	
1.1. Pobreza	1
1.1.1. Introducción	1
1.1.2. ¿Qué es la pobreza? Acepciones	2
1.1.3. Autopercepción de la pobreza	5
1.2. Marco conceptual	9
1.2.1. Marginalidad y dependencia asistencial	9
1.2.2. Pobreza y política social	11
1.3. Medición de la pobreza	14
1.3.1. Métodos directo e indirecto	14
1.3.2. Carencia y carencia severa	17
1.3.3. Indicadores de bienestar	19
1.3.4. Otro indicador de bienestar: Índice de bienes	24
1.4. Medición de la pobreza multidimensional	28
1.4.1. Cálculo del Índice de Privación Material de los Hogares	32
1.4.1.1. Privación patrimonial: condiciones habitacionales	32
1.4.1.2. Privación de recursos corrientes: capacidad	33
2. CAPÍTULO. REVISIÓN DE LA LITERATURA SOBRE ANÁLISIS DE LA POBREZA	
2.1. Introducción	35
2.2. Las mediciones de pobreza en Argentina. Principales estudios	37
2.3. Mediciones de pobreza en el mundo	51
2.4. Comentarios	57
3. CAPÍTULO. MÉTODOS MULTIVARIANTES. MODELOS DE ECUACIONES ESTRUCTURALES	
3.1. Introducción	59
3.2. Estimación no experimental y causalidad	60
3.2.1. Análisis de trayectorias: tipos de relaciones causales	61
3.2.2. Modelos de Ecuaciones Estructurales	67
3.2.3. Estrategias de modelización	70
3.2.4. Identificación del modelo	71
3.2.5. Métodos de estimación	72
3.3. Propuesta de un Modelo Estructural para medir pobreza multidimensional	74

3.4. Comentarios	76
4. CAPÍTULO. MÉTODOS DE ESTIMACIÓN Y PROCEDIMIENTOS DE AJUSTE	
4.1. Introducción	77
4.2. Características y clasificación de los Modelos de Variables Latentes	79
4.3. El Análisis Factorial desde el contexto de la Teoría de Respuesta al Ítem	83
4.3.1. Estimación de los parámetros del Modelo de Rasgos Latentes	88
4.4. El Modelo de Clases Latentes para datos binarios	89
4.4.1. Estimación de los parámetros del Modelo de Clases Latentes	91
4.4.2. Asignación de clases	92
4.5. Modelos para ítems no dicotómicos	93
4.5.1. Modelo de categorías nominales	96
4.6. Bondad del ajuste	96
4.6.1. Ajuste del Modelo de Clases Latentes y asignación de clases	102
4.7. Comentarios	105
5. CAPÍTULO. APLICACIÓN DEL CONCEPTO LATENTE PARA LA ESTIMACIÓN DEL BIENESTAR	
5.1. Introducción	107
5.2. Descripción de los datos	108
5.2.1. Las variables latentes	110
5.2.2. Análisis descriptivo	112
5.2.3. Correlaciones	116
5.3. Aproximaciones a la Teoría de Respuesta al Ítem	118
5.3.1. Estimación de parámetros de los ítems para la variable latente "funcionamiento del jefe de hogar"	122
5.3.2. Estimación de parámetros de los ítems para la variable latente "estar protegido"	127
5.3.3. Estimación de parámetros de los ítems para la variable latente "riesgo de entorno"	132
5.3.4. Estimación de parámetros de los ítems para la variable latente "índice de bienes"	134
5.3.5. Comentarios	137
5.4. Aproximaciones del Modelo de Clases Latentes	138
5.4.1. Ajuste de un Modelo de Clases Latentes para la dimensión "estar protegido"	139
5.4.2. Ajuste de un Modelo de Clases Latentes para las dimensiones "estar protegido" y "riesgo de entorno"	144

5.4.3. Ajuste de un Modelo de Clases Latentes para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"	148
5.4.4. Análisis comparativo de resultados	151
5.4.5. Ajuste de un Modelo de Clases Latentes para la dimensión "funcionamiento del jefe de hogar"	152
5.5. Resultados de los modelos de Variables Latentes aplicados. Comparaciones con los resultados de métodos tradicionales	155
5.6. Análisis de diferencias entre modelos	160
5.7. Comentarios	163
6. CAPÍTULO. CONCLUSIONES	
6.1. Conclusiones	164
6.2. Principales líneas de investigación	166
BIBLIOGRAFÍA	168

ÍNDICE DE TABLAS

Tabla 3.1: Medidas de correlación entre variables con distintas escalas	73
Tabla 4.1: Clasificación de técnicas estadísticas a usar según el tipo de variables Latentes y Manifiestas	79
Tabla 4.2: Caso general de más de dos categorías de respuesta reconfigurada para tratarse como respuesta binaria	94
Tabla 5.1: Pruebas de asociación entre los ítems de las dimensiones "funcionamiento del jefe de hogar" y "estar protegido"	117
Tabla 5.2: Pruebas de asociación entre los ítems de las dimensiones "índice de bienes" y "estar protegido"	117
Tabla 5.3: Pruebas de asociación entre los ítems de las dimensiones "riesgo de entorno" y "estar protegido"	118
Tabla 5.4: Resultados de las pruebas de bondad de ajuste por la IRT para la variable latente "funcionamiento del jefe de hogar"	124
Tabla 5.5: Resultados de las pruebas de bondad del ajuste del modelo IRT parsimonioso para la variable latente "funcionamiento del jefe de hogar"	126
Tabla 5.6: Parámetros estimados por IRT para los ítems "nivel de educación", "estado" y "sexo" correlacionados con "funcionamiento del jefe de hogar"	126
Tabla 5.7: Valores estimados de los parámetros de dificultad y discriminación por la IRT de los ítems que modelan la variable latente "estar protegido"	128
Tabla 5.8: Resultados de las pruebas de bondad del ajuste del modelo IRT parsimonioso para la variable latente "estar protegido"	128
Tabla 5.9: Resultados de las pruebas de bondad del ajuste del modelo IRT para la variable latente "estar protegido" sin el ítem "tenencia"	130
Tabla 5.10: Parámetros y correlación estimados de los ítems "vivienda", "agua", "baño", "cocina" y "hacinamiento" que componen la variable latente "estar protegido"	131
Tabla 5.11: Valores estimados de los parámetros de dificultad y discriminación por la IRT de los ítems que modelan la variable latente "riesgo de entorno"	133
Tabla 5.12: Resultados de las pruebas de bondad del ajuste del modelo IRT para la variable latente "índice de bienes"	135
Tabla 5.13: Parámetros estimados de los ítems que indican posesión de "radio", "televisores", "teléfono fijo", "celulares", "computadoras" y "servicio de internet"	136
Tabla 5.14: Parámetros y correlación estimados de los ítems que modelan la dimensión "índice de bienes"	136
Tabla 5.15: Índices de ajuste para comparar modelos con 2, 3 y 4 clases latentes para la dimensión "estar protegido"	140

Tabla 5.16: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando la dimensión "estar protegido"	141
Tabla 5.17: Proporción de hogares que presentan situación con bienestar y sin bienestar en la dimensión "estar protegido"	142
Tabla 5.18: Índices de ajuste para comparar modelos con 2, 3 y 4 clases latentes para las dimensiones "estar protegido" y "riesgo de entorno"	145
Tabla 5.19: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando la dimensión "estar protegido" y "riesgo de entorno"	146
Tabla 5.20: Clasificación y proporción de los hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para las dimensiones "estar protegido" y "riesgo de entorno"	147
Tabla 5.21: Índices de ajuste para comparar modelos con 2, 3 y 4 clases latentes para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"	149
Tabla 5.22: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"	150
Tabla 5.23: Clasificación y proporción de los hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"	150
Tabla 5.24: Porcentaje de hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes" de tres diferentes modelos	152
Tabla 5.25: Índices de ajuste para comparar modelos con 2 y 3 clases latentes para la dimensión "funcionamiento del jefe de hogar"	153
Tabla 5.26: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando la dimensión "funcionamiento del jefe de hogar"	154
Tabla 5.27: Clasificación y proporción de los hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para la dimensión "funcionamiento del jefe de hogar"	154
Tabla 5.28: Hogares bajo las líneas de pobreza e indigencia en el tercer trimestre de 2011	155
Tabla 5.29: Número de hogares con bienestar y sin bienestar según el método multidimensional versus el número de hogares bajo la línea de pobreza y sobre la línea de pobreza	156
Tabla 5.30: Número de hogares con bienestar y sin bienestar según el método multidimensional versus el número de hogares con y sin carencias	157
Tabla 5.31: Porcentajes de hogares que sufren privaciones calculados con las metodologías unidimensionales y multidimensionales	159

Tabla 5.32: Relación entre una prueba diagnóstica y la presencia o ausencia de "bienestar"	160
Tabla 5.33: Resultados de la clasificación de los hogares aplicando los Métodos Multidimensional y Unidimensional	161

ÍNDICE DE FIGURAS

Figura 1.1: Semáforo de la eliminación de la pobreza	8
Figura 1.2 Propuestas de Carencias y Carencias Severas. Metodología utilizada como una aproximación a la medición de la pobreza	18
Figura 1.3: Construcción del Índice de Privación Material de los Hogares	32
Figura 3.1: Esquema de modelo de trayectorias	62
Figura 3.2: Relaciones de causalidad	63
Figura 3.3: Modelos de Trayectoria recursivo y no recursivo	64
Figura 3.4: Las relaciones de causalidad entre las metodologías experimental y correlacional	65
Figura 3.5: El Modelo de Ecuaciones Estructurales como resultado de la integración de modelos y técnicas de diferentes disciplinas	66
Figura 3.6: Modelo general de Ecuaciones Estructurales completo	68
Figura 3.7: Modelo de Ecuaciones Estructurales para construir el índice de bienestar multidimensional	75
Figura 5.1: Proporción de hogares con situación favorable en los ítems relacionados al "funcionamiento de jefes de hogar"	113
Figura 5.2: Distribución porcentual de los rangos de edades de los jefes de hogar	113
Figura 5.3: Distribución porcentual de los niveles de educación de los jefes de hogar	114
Figura 5.4: Proporción de hogares con situación favorable en los ítems relacionados con el "estar protegidos" y "riesgo de entorno"	115
Figura 5.5 Proporción de hogares con situación desfavorable en los ítems relacionados al "índice de bienes"	116
Figura 5.6: Modelo Estructural para la variable latente "funcionamiento del jefe de hogar"	123
Figura 5.7: Curvas características de los ítems dicotómicos para la variable latente "funcionamiento del jefe de hogar"	125
Figura 5.8: Curvas características de los ítems dicotómicos "sexo", "nivel de educación" y "estado de ocupación" altamente correlacionados con la variable latente "funcionamiento del jefe de hogar"	127
Figura 5.9: Modelo Estructural para la variable latente "estar protegido" considerando los ítems "tipo de vivienda", "tiene agua", "tiene baño", "tiene cocina", "tenencia" y "hacinamiento"	127

Figura 5.10: Curvas características de los ítems "tipo de vivienda", "tiene agua", "tiene baño", "tiene cocina", "hacinamiento" y "tenencia" de la variable latente "estar protegido"	129
Figura 5.11: Modelo estructural para la variable latente "estar protegido" considerando los ítems "tipo de vivienda", "tiene agua", "tiene baño", "tiene cocina" y "hacinamiento"	130
Figura 5.12: Curvas características para los ítems que conforman la dimensión "estar protegido" con el modelo parsimonioso	131
Figura 5.13: Modelo estructural para la dimensión "riesgo de entorno" considerando los ítems "vivir cerca de basural", "vivir en zonas inundables" y "vivir en una villa"	132
Figura 5.14: Curvas características de los ítems que conforman la dimensión "riesgo de entorno"	133
Figura 5.15: Modelo Estructural para la variable latente "índice de bienes" considerando los ítems que indican posesión de "radio", "televisores", "teléfono fijo", "celulares", "computadoras" y "servicio de internet"	134
Figura 5.16: Curvas características de los ítems que indican posesión de "radio", "televisores", "teléfono fijo", "celulares", "computadoras" y "servicio de internet"	135
Figura 5.17: Curvas características de los ítems que conforman la dimensión "índice de bienes" para el modelo parsimonioso	137
Figura 5.18: Probabilidades condicionales estimadas para cada clase latente en la dimensión "estar protegido"	143
Figura 5.19: Modelo Estructural para las dimensiones "estar protegido" y "riesgo de entorno" considerando los ítems tipo de vivienda, tiene agua, tiene baño, tiene cocina, hacinamiento, vivir cerca de basural, vivir en zona inundable y vivir en una villa	144
Figura 5.20: Probabilidades condicionales de las clases latentes "con bienestar" y "sin bienestar" teniendo en cuenta las dimensiones "estar protegido" y "riesgo de entorno" simultáneamente	147
Figura 5.21: Modelo Estructural para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"	148
Figura 5.22: Probabilidades condicionales de clase latente para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"	151
Figura 5.23: Modelo Estructural parsimonioso para la dimensión "funcionamiento del jefe de hogar"	153
Figura 5.24: Gráfico de la Curva ROC. Evaluación del Método Multidimensional para la determinación de hogares "con bienestar" y "sin bienestar"	162
Figura 6.1: Modelo de clases latentes con variable agrupadora y dependencias	167

Resumen

El contexto general de esta investigación se enmarca en el estudio de la pobreza, concretamente en el análisis de la acepción multidimensional de la misma para el caso específico de Tucumán. El enfoque moderno de la pobreza se acerca más al concepto de bienestar en el que influyen numerosos factores, no sólo a nivel del ingreso monetario. La delimitación de la investigación se centra en el progreso de la visión tradicional de la pobreza hacia el campo multidimensional, acotando el análisis empírico a la situación de los aglomerados urbanos de Tucumán.

El objetivo principal de esta tesis es mostrar una metodología multidimensional, alternativa a los métodos tradicionales de Línea de Pobreza y Necesidades Básicas Insatisfechas para medir pobreza. Para ello se adoptan y desarrollan diferentes perspectivas de tratamiento de la información. Los Modelos de Ecuaciones Estructurales proporcionan una estrategia empírica de estimación del bienestar multidimensional y su aplicación supone un avance importante en la búsqueda de las relaciones causales entre constructos no observables. También, se desarrollan los conceptos de los Modelos de Variables Latentes y su conexión con los datos categóricos; las relaciones entre los Modelos Lineales y el Modelo de Rasgo Latente, particularmente, el Modelo de Respuesta al Ítem para datos categóricos y los Modelos de Clases Latentes como herramienta para la distribución óptima de la muestra en los segmentos identificados y el tamaño de cada segmento.

El índice de bienestar multidimensional propuesto, pretende ser una herramienta alternativa para detectar si los individuos logran alcanzar umbrales mínimos de bienestar (o estándar de vida) en cada una de las dimensiones propuestas.

Se destaca también, la continuidad en las metodologías de medición y estimación de indicadores para poder comparar resultados a través del tiempo y el espacio y analizar la evolución de Índices de Pobreza Multidimensional, lo que llevará a una aplicación adecuada y eficiente de las políticas públicas.

Agradecimientos:

Esta Tesis, realizada en la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán, representa un gran esfuerzo en el cual, directa o indirectamente, participaron distintas personas, entre colegas, familiares y amigos, brindando su opinión, marcando mis errores, dándome ánimo en los momentos de crisis y acompañando en los momentos de satisfacción. Este trabajo me ha permitido aprovechar la experiencia de muchas personas que deseo agradecer.

En primer lugar, a la Comisión de Seguimiento, integrada por los Doctores Shrikant Bangdiwala, Viviana Lencina, Manuel Luís Cordomí y Esteban Nicolini, por su valiosa dirección y apoyo para llegar al final del camino trazado y concluir la Tesis. Me gustaría agradecer muy especialmente al Dr. Manuel Luís Cordomí por su dedicación, orientación y aporte de conocimientos durante el proceso de redacción de la Tesis.

A la facultad de Ciencias Económicas por brindarme los conocimientos, el espacio y las herramientas para cursar y completar la carrera del Doctorado en Estadística. A mis compañeros, colegas y amigos.

Por supuesto, mi profundo agradecimiento a mi esposo e hijos por la paciencia, comprensión y el apoyo incondicional durante estos largos años de trabajo. También a mi madre, hermanos y sobrinos por el tiempo privado a la familia para llevar a cabo este proyecto.

A todos, muchas gracias.

ABREVIATURAS

AF	Análisis Factorial
AIC	Criterio de Información de Akaike
AUH	Asignación Universal por Hijo para la Protección Social
BM	Banco Mundial
BIC	Criterio de Información Bayesiano
BIC _{aj}	Criterio de información Bayesiano ajustado por el tamaño de muestra
CAPECO	Capacidad Económica
CBA	Canasta Básica Alimentaria
CBT	Canasta Básica Total
CCI	Curva Característica del Ítem
CEPAL	Comisión Económica para América Latina y el Caribe
CFA	Análisis Factorial Confirmatorio
CFI	Índice de Ajuste Comparativo
CMIN/DF	Razón de chi cuadrado sobre los grados de libertad
CODUSU	Código Usuario
CONDHAB	Condiciones Habitacionales
CONICET	Consejo Nacional de Investigaciones Científicas y Técnicas
CMIN/DF	Razón Chi Cuadrado sobre Grados de Libertad
CNPV	Censo Nacional de Población y Vivienda
DNI	Documento Nacional de Identidad
EAHU	Encuesta Anual de Hogares Urbanos
ECPF	Encuesta Continua de Presupuesto Familiar
EDSA	Encuesta de la Deuda Social Argentina
EM	Esperanza-Maximización
ENGHo	Encuesta Nacional de Gastos por Hogares
ENTIC	Encuesta Nacional de Tecnologías de la Información y la Comunicación
EPH	Encuesta Permanente de Hogares
EUROSTAT	Oficina Europea de Estadística
FA	Análisis Factorial
FFP	Fracción de Falsos Positivos
FGT	Índice de Pobreza Foster-Greer-Thorbecke
FVP	Fracción de Verdaderos Positivos
GBA	Gran Buenos Aires
GFI	Índice de Bondad de Ajuste
GL	Grados de Libertad
GLS	Mínimos Cuadrados Generalizados
IELDE	Instituto de Estudios Laborales y del Desarrollo Económico
IICS	Índice Integrado de Capacidades de Subsistencia
INDEC	Instituto Nacional de Estadísticas y Censos
INE	Instituto Nacional de Estadística
IPA	Investigación de la Pobreza en Argentina
IPC	Índice de Precios al Consumidor
IPH	Índice de Pobreza Humana
IPM	Índice de Pobreza Multidimensional

IPMH	Índice de Privación Material de los Hogares
IRT	Teoría de Respuesta al Ítem
LCM	Modelos de Clases Latentes
LGDS	Ley General de Desarrollo Social
LMR	Lo-Mendell-Rubin (Prueban de razón de verosimilitud)
LP	Línea de Pobreza
LPM	Modelos de Perfiles Latentes
LTM	Modelos de Rasgos Latentes
LVM	Modelos de Variables Latentes
MCL	Modelo de Clase Latente
MCLN	Modelo de Clase Latente Multinivel
MD	Privación Material (siglas en inglés)
ML	Máxima Verosimilitud (siglas en inglés)
MMNV	Marco Muestral Nacional de Viviendas
MRL	Modelo de Rasgo Latente
MV	Máxima Verosimilitud
NBI	Necesidades Básicas Insatisfechas
OCDE	Organización para la Cooperación y el Desarrollo Económicos
ONG	Organizaciones No Gubernamentales
OPHI	La Iniciativa de Pobreza y Desarrollo Humano de Oxford
PHOGUE	Panel de Hogares de la Unión Europea
PMD	Pobreza Multidimensional
PNUD	Programa de las Naciones Unidas para el Desarrollo
POM	Proportional Odds Model
PPA	Paridad de Poder Adquisitivo
RMSEA	Error Cuadrático Medio de Aproximación
ROC	Receiver Operating Characteristic
RVB	Razón de Verosimilitud paramétrica Bootstrapped
SEM	Modelo de Ecuaciones Estructurales
SEN	Sistema Estadístico Nacional
SEP	Sistema Estadístico Provincial
TIC	Tecnología de la Información y la Comunicación
TRI	Teoría de Respuesta al Ítem
UBACyT	Universidad Nacional de Buenos Aires Ciencia y Tecnología
UCA	Universidad Católica Argentina
UE	Unión Europea
UEM	Unidades Espaciales Menores
ULS	Mínimos Cuadrados no Ponderados
UNLP	Universidad Nacional de La Plata
UNTREF	Universidad Nacional de Tres de Febrero
UV	Variables Subyacentes (<i>Underlying Variables</i>)
VAE	Valor de los Años de Escolaridad Invertidos en el Mercado Laboral
VLMR	Razón de Verosimilitud de Vuong-Lo-Mendell-Rubin
WLS	Mínimos Cuadrados Ponderados

INTRODUCCION

El contexto general de esta investigación se enmarca en el estudio de la pobreza, concretamente en el análisis de la acepción multidimensional de la misma para el caso específico de Tucumán. El enfoque moderno de la pobreza se acerca más al concepto de bienestar en el que influyen numerosos factores, no sólo a nivel del ingreso monetario.

Las razones que justifican la elaboración de esta tesis doctoral se resumen en tres tipos diferentes pero complementarias: por un lado, existe una motivación personal tendiente a aportar ideas para detectar las carencias en los hogares de Tucumán y que sirvan para la implementación de políticas públicas tendientes a mejorar la calidad de vida de esos hogares. Es fundamental investigar en aquello que afecta al bienestar de las personas, máxime si se generan situaciones de desarraigo social. Por otro lado, aparece una justificación estrictamente económica, ya que, hoy por hoy la pobreza sigue siendo un asunto sin resolver y por ello existe el interés de aportar un "granito de arena" a la Comunidad Científica y Política. Esto ha estimulado la inquietud por encontrar una forma sencilla de medir la pobreza y de explicar sus causas, con el fin de actuar sobre ella. Por último, es un tema con muchas posibilidades de mejora, sobre todo en los medios necesarios para ello (información de alcance, bases de datos, técnicas para estimar pobreza multidimensional multivariante, software adecuado y temas de actualidad).

La delimitación de la investigación se centrará en el progreso de la visión tradicional de la pobreza hacia otro espacio multidimensional, y acotando el análisis empírico a la situación de los aglomerados urbanos de Tucumán.

Definir la "pobreza" es complejo, y aún más complejo es medirla, depende del punto de vista teórico con que se la plantee. Entonces, para estimar qué porción de una población no alcanza niveles de bienestar adecuados, se utilizan diversos métodos, desde los cuantitativos esenciales, hasta diseñar modelos complejos que incluyan variables de difícil medición y otras que no son medibles directamente.

Considerando su carácter multidimensional y arduo, en general se podría aseverar que la pobreza es una condición en la cual una persona, grupos de personas y/o familias tienen un nivel de bienestar inferior al mínimo necesario para la sobrevivencia. Se exponen las diferentes metodologías de medición de pobreza utilizadas habitualmente y sus resultados.

Por otro lado, el análisis multivariante es un conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente muchas variables medidas para cada individuo o unidad estudiada. En sentido estricto, muchas técnicas multivariantes son extensiones del análisis univariante (análisis de distribuciones de una sola variable) y del análisis bivariante (clasificaciones cruzadas, correlación, análisis de la varianza y regresiones simples utilizadas para analizar dos variables). Su razón de ser se basa en la mejor comprensión del fenómeno que se desea estudiar obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de lograr.

Existen diferentes técnicas y modelos, cada uno con su tipo de análisis. Se clasifican en métodos de: dependencia, interdependencia y estructurales. En los métodos de dependencia, las variables analizadas están divididas en dependientes e independientes y el propósito es determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes. En los métodos de interdependencia no hay distinción entre las variables dependientes y las independientes, y su objetivo es identificar qué variables están relacionadas, en qué forma y por qué. En cuanto a los estructurales, las variables dependientes e independientes están divididas en sendos grupos y su objetivo es analizar, no sólo cómo las variables independientes afectan a las dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí.

El establecimiento de la causalidad es uno de los problemas fundamentales que aborda la Estadística. Es un tópico importante en diversas áreas, desde las ciencias sociales -como la economía- hasta ciencias ambientales y de salud -como la psicología-. Es sabido que, antes de establecer una relación causal entre dos variables, debe identificarse una correlación entre ellas. Las metodologías clásicas de la Estadística permiten identificar, en forma relativamente sencilla, las correlaciones que no sean producto de la causalidad. Pero, una vez identificadas las relaciones estadísticamente significativas, estudiar una relación causal no es tarea fácil.

Las investigaciones que se realizan en el área de las ciencias sociales normalmente se enfrentan a dificultades en las cuales los constructos¹ de interés, desde el punto de vista de la teoría (por ejemplo, bienestar, pobreza, satisfacción) y de mayor capacidad explicativa de los fenómenos de interés, son multidimensionales y no se pueden medir directamente, ya que constituyen las denominadas Variables Latentes (VL). La medición de estas variables latentes se realiza indirectamente a través de múltiples variables observadas, llamadas variables manifiestas o indicadores (de tipo métricos) o ítems (de tipo categórico ordinal o dicotómico) (NAVARRO & ASUN, 2016).

Los métodos estadísticos basados en modelos de variables latentes juegan un rol importante en el análisis de datos multivariantes ya que se han desarrollado para reducir la dimensionalidad del conjunto de datos, manteniendo la mayor parte de la información posible. Esto se puede hacer debido a que muchas de las cuestiones planteadas en las variables medidas son, en realidad, aspectos de la misma característica básica.

El objetivo del análisis estadístico estándar es inferir parámetros de una distribución, a partir de muestras de dicha población; estudiar asociaciones estadísticas, para lo cual utiliza la probabilidad y técnicas de estimación, pero no hace una interpretación causal de los resultados. Si el estudio es experimental, se puede controlar el factor causal manteniendo fijo los otros factores, en este caso sí es factible hablar de relaciones causales. Sin embargo, muchas de las ciencias que utilizan la estadística vieron la necesidad de responderse preguntas de índole causal, a través de ensayos clínicos, como: ¿Cuál es la eficacia de una droga en una población determinada? ¿Qué porcentaje de crímenes del pasado podrían haberse evitado con una política determinada? ¿El tener un título universitario influye para que los ingresos de un individuo aumenten en el mercado laboral? Este es el tipo de preguntas causales a las que procura dar respuesta la inferencia causal. La causalidad trasciende la matemática, resultando de suma importancia en diversas áreas del conocimiento.

¹ Variable no observable directamente, de carácter teórico o abstracto medido por variables observables.

En la problemática de la pobreza no se puede, éticamente, hacer estudios experimentales, y por ende el tema de esta tesis es valioso. Identificar los efectos causales individuales excede las posibilidades en la medida en que no es posible saber qué es lo que hubiera ocurrido con un individuo si su situación laboral, educacional o de cualquier otra índole en general, hubiera sido contraria a la observada. Esta limitación demanda un enfoque diferente, dejando de lado la pregunta individual sobre el efecto causal de cierta condición, para introducir el efecto medio del mismo en toda la población.

En cierto tipo de situaciones, la observación se reemplaza por el "control estadístico", que requiere explicitar todas las variables implicadas en la investigación. En estos estudios no experimentales, las relaciones causales se infieren a partir de las relaciones observadas entre variables (muchas variables tienden a moverse conjuntamente). Desde el punto de vista estadístico, la variación conjunta entre variables se mide a través de ciertos estadísticos como la covarianza o la correlación.

En esta área de investigación, se requiere de una perspectiva multivariada de análisis de datos. Los Modelos de Ecuaciones Estructurales (SEM, por sus siglas en inglés) se consideran una extensión de los Modelos de Regresión Múltiple y de Análisis Factorial. Un SEM examina simultáneamente una serie de relaciones de dependencia. En este caso es particularmente útil, porque alguna variable dependiente se podría convertir en variable independiente en ulteriores relaciones de dependencia. Este conjunto de relaciones, cada una con variables dependientes e independientes, es la base del SEM (HAIR *et al.*, 1999).

El análisis multivariado comprende un grupo de técnicas de análisis de datos en continua expansión, el Análisis Factorial (FA, por sus siglas en inglés) es una de ellas. Se aplica para determinar interrelaciones entre un número elevado de variables métricas, explicando dichas interrelaciones en términos de un número menor de variables denominadas factores (si son inobservables) o componentes principales (si son observables). El objetivo es encontrar la forma de condensar la información contenida en las variables originales en un conjunto más pequeño de variables con la menor pérdida de información posible. Así, por ejemplo, si un analista financiero quiere determinar cuál es el estado financiero de una empresa a partir de un número de indicadores financieros, construyendo varios índices numéricos que definan su

situación, el problema se resolvería mediante un Análisis de Componentes Principales. Si un psicólogo quiere determinar los factores que caracterizan la inteligencia de un individuo a partir de sus respuestas a un test de inteligencia, utilizaría para resolver este problema un Análisis Factorial.

El FA de Componentes Principales es uno de los modelos estadísticos de mayor aplicación para estudiar las propiedades de variables escalares. Cuando los datos son categóricos no cumplen con las propiedades de normalidad de las distribuciones o linealidad de las relaciones entre las variables, que son restricciones del FA.

Actualmente, la aplicación de Modelos de Variables Latentes (LVM, por sus siglas en inglés) tiene particular interés para abordar el análisis con datos categóricos. Bajo este enfoque, se dispone de una variedad de métodos, técnicas y programas informáticos adecuados para aproximar los datos a los objetivos de la investigación.

En este conjunto de opciones hay aproximaciones teóricas y metodológicas especialmente relevantes. Habitualmente se destacan los enfoques clásicos que consisten en tratar a este tipo de datos como variables subyacentes (*underlying variables*, UV), en los que se establece una conexión entre variables categóricas y continuas. Otra alternativa conveniente es el Análisis Factorial desde la Teoría de Respuesta al Ítem (IRT, por sus siglas en inglés) que proporciona un conjunto adicional de modelos, métodos de estimación y de evaluación del ajuste, que conllevan el avance de diferentes supuestos (CHACON, 2011).

Esencialmente, el método de estimación de los parámetros del Modelo de Variables Latente es el Método de Máxima Verosimilitud. En cuanto a la bondad del ajuste, puede verificarse usando una prueba estándar para comparar las frecuencias observadas y esperadas a través del patrón de respuestas. Convencionalmente se refiere a estos estimadores como "frecuencias esperadas" cuando se usa la razón de verosimilitud frente a las Pruebas de Bondad de Ajuste de chi cuadrado de Pearson (BARTHOLOMEW *et al.*, 2008).

La pobreza es un fenómeno que abarca aspectos conceptuales y metodológicos muy variados de modo que, al elegir una técnica de medición, los investigadores deben tener en cuenta dichos aspectos. Generalmente, el proceso de medición involucra dos elementos: la "identificación" de las personas que se consideran pobres

y la "agregación" del bienestar de esos individuos en una medida de pobreza (FERES & MANCERO, 2001a).

En este trabajo de investigación se plantea considerar las distintas interpretaciones conceptuales del término "pobreza", y una revisión de las metodologías más utilizadas en los procesos de identificación y agregación. Tanto en los ámbitos conceptual como metodológico, se revisan las alternativas entre las nociones de pobreza "absoluta" y "relativa", entre los enfoques "directo" e "indirecto" y entre las perspectivas "objetiva" y "subjética". Finalmente, se concluye que ningún método de identificación y agregación es completo por sí solo, por lo que el uso combinado de los mismos, parece ser la opción más acertada.

Por otro lado, se considera una medición integral de la pobreza, desde un punto de vista multivariante, con el objeto de capturar no sólo el aspecto material de la situación de los individuos, sino también sus condiciones de vida y bienestar. Esto significa que el nivel de ingresos no es factor determinante para medir el bienestar de un hogar o un individuo, también necesita ciertas condiciones ambientales y sociales para poder desarrollarse y ejercer sus derechos a la educación, la salud y la libertad de pensamiento y conciencia.

Históricamente, el estudio científico de la pobreza se remonta a comienzos del siglo XX. Anteriormente a esa fecha se habían realizado algunas estimaciones sobre pobreza, pero el primero en "combinar la observación con un intento sistemático de medición de la extensión del problema", fue Booth entre 1892 y 1897, elaborando un mapa de pobreza de Londres (ATKINSON, 1987). Posteriormente, se realizó un estudio para medir la pobreza en York, y se utilizó un estándar de pobreza basado en requerimientos nutricionales (ROWNTREE, 1901). A partir de entonces se han desarrollado nuevos conceptos y nuevas metodologías para medir la pobreza, algunas de las cuales se reseñan en esta tesis.

Con el propósito de contribuir al problema de la medición de la pobreza, se presenta una metodología para modelar algunas dimensiones del bienestar, fundamentadas en el enfoque de las capacidades y los funcionamientos de Amartya Sen (SEN, 1987). En ese orden de conceptos, se propone una alternativa metodológica que utiliza los modelos de ecuaciones estructurales con múltiples

indicadores y múltiples causas, los cuales se enfocan a través del marco teórico de los modelos de variables latentes generalizados, rasgos latentes y clases latentes.

Desde el planteo inicial de la temática de esta tesis, se detectan problemas importantes: uno es relativo a la forma de medir la pobreza, donde tradicionalmente se la ha cuantificado única y exclusivamente a través del ingreso personal o del hogar, según sea la unidad de análisis. De esta forma, la información extraída es incompleta ya que no existe una combinación de elementos objetivos y subjetivos, ni la inclusión de factores que también influyen como el trabajo, la vivienda, la salud, los bienes durables, el ocio, etc. Además, en los últimos años se han desarrollado otros indicadores más potentes pero que no terminan de resolver la elección adecuada de variables, que cuantifiquen directamente la pobreza, y que ofrezcan las mediciones principales de los indicadores.

Ante estos inconvenientes, se procura contribuir con un "granito de arena", construyendo un indicador de pobreza multidimensional, que sea fácil de interpretar y que aporte información complementaria a otros estudios.

La propuesta es elaborar un modelo explicativo de la pobreza, considerando relaciones de causalidad.

Centrado el tema y las líneas de investigación que seguir, se plantea una serie de objetivos generales y específicos. Como consecuencia, la tesis se estructura de la siguiente manera: la primera parte refleja el marco conceptual y de referencia, necesario para respaldar y dar soporte al análisis empírico posterior. Y la segunda, eminentemente aplicada (marco empírico y aplicado), incluye las contribuciones esenciales de la investigación.

El primer capítulo, llamado "**Conceptos de pobreza**", contiene los conceptos relacionado con la pobreza en sí y los enfoques tradicionales de medición. Se resumen las principales acepciones existentes del término "pobreza", comenzando por las comúnmente conocidas, como la objetiva y la subjetiva, como la absoluta y la relativa y, finalizando por conceptos más innovadores como el de la nueva pobreza y la pobreza multidimensional. Con los diferentes criterios, aparecen con cierto detalle las principales líneas metodológicas que se deben tener en cuenta en investigaciones de este tipo: definición y tipos de umbral de la pobreza, dimensiones para tener en

cuenta, variables objeto de estudio, unidad de análisis, escalas de equivalencia y medidas de los indicadores; todo ello muy importantes para mantener un rigor adecuado y realizar comparaciones.

El capítulo dos, llamado "**Revisión de la literatura sobre análisis de pobreza**", se refiere a investigaciones empíricas relevantes realizadas en Argentina, en otros países de América Latina y en Europa. Algunas referencias argentinas se remontan a finales de la década del setenta. En los últimos años, ciertas instituciones y centros de investigación han hecho hincapié en el concepto de la pobreza multidimensional, entre ellos se destaca el Observatorio de la Deuda Social de la Universidad Católica Argentina.

En el capítulo tres, denominado "**Métodos Multivariantes**" se desarrolla el marco teórico de la metodología multivariante de los "Modelos de Ecuaciones Estructurales", que se propone como alternativa, a la Línea de Pobreza (LP) y/o Necesidades Básicas Insatisfechas (NBI), para medir carencias. El Modelo de Ecuaciones Estructurales (SEM) abarca una familia entera de modelos conocidos con diversos nombres, entre ellos, Análisis de la Estructura de la Covarianza, Análisis de Variable Latente, Análisis de Factor Confirmatorio y/o simplemente, Análisis LISREL². Todas estas técnicas tienen la particularidad de estimar una serie de regresiones múltiples distintas pero interrelacionadas, mediante la especificación del modelo estructural utilizado por el programa estadístico.

En la segunda etapa de la investigación se desarrolla el marco empírico y aplicado, donde se enmarcan tres capítulos bien diferenciados. En el primero de ellos (Capítulo 4), titulado "**Métodos de estimación y procedimientos de ajuste**", donde se describe la técnica del Análisis Factorial Confirmatorio (CFA, por sus siglas en inglés) y los métodos de aproximación cuando los datos no son métricos sino categóricos, Teoría de Respuesta al Ítem (IRT).

El CFA es un análisis encuadrado en los Modelos de Ecuaciones Estructurales, cuyo objetivo es el estudio de los modelos de medida, es decir, analizar las relaciones entre un grupo de indicadores (variables observables) y una o más variables latentes, factores o constructos. Sirve para determinar de antemano aspectos relevantes del

² Linear Structural Relationship, Software de computación, creado por Jöreskog – Sörbon en 1993.

modelo, fundamentados en la teoría previa y en la evidencia conocida. De este modo, se debe especificar, con anterioridad al análisis, qué factores y qué indicadores forman el modelo, si existe o no relación entre los factores y qué indicadores presentan saturaciones³ en cada factor.

En el Capítulo 5, llamado "**Aplicación del concepto latente para la estimación del bienestar**", se implementa la metodología explicada con detalle en los capítulos 3 y 4, desarrollando un modelo explicativo de la pobreza multidimensional. Se identifican las interrelaciones más importantes entre los factores con ciertas hipótesis de la investigación. Inicialmente se realiza un análisis descriptivo de los datos utilizados para realizar las estimaciones y comparaciones.

La justificación de aplicar este tipo de análisis radica en el aporte que ofrece para reducir información redundante y agruparla en variables latentes o constructos. Permite contrastar las relaciones de causalidad entre dimensiones, las cuales pueden comportarse como variables independientes en un primer momento y en endógenas en procesos ulteriores, por lo que ofrece la posibilidad de construir un diagrama de secuencias complejo, aunque fácil de interpretar.

Por otro lado, el Modelo de Clases Latentes es un marco metodológico muy adecuado para el análisis de la pobreza multidimensional porque resuelve el problema de la agregación y la definición del umbral. A partir de la definición de pobreza multidimensional utilizada por algunos organismos internacionales, es lógico inferir que la noción de pobreza es una definición no observada directamente. Puesto que el Modelo de Variables Latentes es un método estadístico multivariante que mide una variable no observada a partir de la información recogida en un conjunto de variables observas, se convierte en el más adecuado para realizar esta tarea. Además, dado que los indicadores disponibles son variables categóricas (en su mayoría dicotómicas), se hace necesaria la utilización del modelo de clases latentes, propuesto por (LAZARFELD, 1954) y (LAZARFELD & HENRY, 1968). Finalmente, al asignar a cada hogar a una clase distinta según el nivel y el tipo de pobreza latente sufrida, se solventa el problema de la arbitrariedad al establecer el umbral de pobreza.

³ Las saturaciones representan la varianza de los indicadores explicada por el factor. Indicadores con saturaciones superiores a 0,40 se consideran bien explicados.

Posteriormente, se elabora el modelo de pobreza multidimensional mediante un diagrama de secuencias, que representa todas las relaciones de causalidad señaladas. Dicho diagrama es inherente a las ecuaciones estructurales subyacentes, en el que se relacionan las variables latentes marcadas. Además, se usa el Método de los Modelos Rivales, donde a partir de un modelo base, se agregan nuevas relaciones para mejorar la calidad del ajuste, aunque, siempre manteniendo el rigor teórico base.

También, se aborda el problema de comparar el poder de clasificación de los modelos propuesto con el método de línea de pobreza a partir de las curvas ROC (*Receiver Operating Characteristic curves*).

Por último, el Capítulo 6 contiene las conclusiones y propuestas para futuras investigaciones sobre el tema.

CAPÍTULO 1

POBREZA: CONCEPTOS Y METODOLOGÍAS

1.1. Pobreza

En este capítulo se presentan las variadas definiciones de pobreza dadas por diferentes actores involucrados en la temática. Asimismo, se indican las metodologías que se utilizan habitualmente para medirla y la acción de algunas políticas públicas para enfrentar el fenómeno.

“La pobreza tiene muchas caras, sus múltiples definiciones explican el porqué de los diferentes métodos para medirlas” ... IPH de las Naciones Unidas

1.1.1. Introducción

La pobreza es tan antigua como la humanidad. Históricamente, la disparidad de intereses entre pobres y ricos ha provocado sucesos de diversas características que, a pesar de los avances en asistencia a los menos privilegiados, no han logrado modificar, sustancialmente, la situación de desigualdad en que vive la mayor parte de la población mundial. Sus diferentes significados y presentaciones han sido materia de estudio por historiadores, sociólogos y economistas. Sus causas se han identificado en una amplia variedad de fuentes, desde deficiencias en la administración del sostén del ingreso, a la injusticia del sistema económico y social.

En términos generales, la pobreza depende del desarrollo relativo alcanzado por un determinado país o región. De este modo, tanto el carácter como la magnitud de la

pobreza en los denominados países emergentes¹, difiere cualitativa y cuantitativamente de la misma condición en los países más avanzados. De igual manera, entre los países en desarrollo la pobreza no se presenta con la misma intensidad; en las naciones africanas y algunas asiáticas la pobreza reviste características cercanas a la miseria, incluso a la miseria absoluta. En otras regiones, como América Latina la pobreza es "menos pobre". Esto significa que, a pesar de no lograr satisfacer plenamente las necesidades básicas de un alto porcentaje de la población, la pobreza en esta región no alcanza el dramatismo de otras latitudes (ROMERO, 2000).

La mayoría de los estudios e investigaciones enfocan su atención en los estratos más desfavorecidos de la población; no obstante, entre ricos y pobres se encuentra la clase media, que permanentemente se esfuerza para ascender en la escala social o, al menos, no caer en la pobreza. Esto es esencialmente notorio en épocas de crisis, cuando el Estado descarga sobre la clase media, el mayor peso del reajuste fiscal y le restringe el ingreso cercenando sus posibilidades de progreso.

En definitiva, la preocupación por la pobreza se ha expresado por siglos, aun cuando no es prioridad en la agenda de la acción política.

1.1.2 ¿Qué es la pobreza? Acepciones.

En términos generales, la pobreza se refiere a la incapacidad de las personas de vivir una vida tolerable (PNUD, 1997). Entre los aspectos que la componen se menciona llevar una vida larga y saludable, tener educación y disfrutar de un nivel de vida decente, además de otros elementos como la libertad política, el respeto de los derechos humanos, la seguridad personal, el acceso al trabajo productivo y bien remunerado y la participación en la vida comunitaria. No obstante, dada la natural dificultad de medir algunos elementos constituyentes de la "calidad de vida", el estudio de la pobreza se ha restringido a los aspectos cuantificables –y generalmente materiales– de la misma, relacionados con el concepto de "nivel de vida".

¹ Es una clasificación que permite agrupar a aquellos mercados que ya no son "en vías de desarrollo" pero que a su vez no han alcanzado el estatus de "desarrollado".

El criterio habitual con el que se identifica la pobreza es el de una situación o forma de vida que surge como resultado de la imposibilidad de acceso y/o carencia de los recursos, para satisfacer las necesidades básicas humanas. Esta carencia incide en el desgaste del nivel y calidad de vida de las personas. (PAZ NÚÑEZ, 1996).

La pobreza puede concebirse como una situación de degradación generalizada (nutrición, educación, sanidad) que afecta a los individuos, tanto física, como fisiológica y psicológicamente, privándolos de sus capacidades básicas para ejercer sus derechos y mejorar su calidad de vida. Una de las principales dificultades a la hora de estudiar la pobreza es precisamente poner medida a este concepto de degradación para determinar el número de personas afectadas, en vista de los diversos factores a los que puede deberse.

Encontrar una definición única para "la pobreza" no es nada sencillo. Sin embargo, todo indica que, en esencia, es multidimensional y multicausal, por lo tanto, no se la puede abordar desde un solo ángulo, sino que debe ser planteada como una situación compleja que involucra factores de índole económico, social, educacional y político.

Hay componentes distributivos de la inequidad que derivan del sistema de retribución de los elementos que se encuentran implícitos en el estilo de desarrollo, que van más allá de los niveles de desigualdad aceptada socialmente. Estos elementos abarcan la distribución del ingreso y del patrimonio, incluyendo en éste, no solo sus aspectos físicos sino el patrimonio de conocimiento y habilidades y el acceso a la información. Entonces, más allá de la distribución del ingreso -que no se busca desestimar-, en el análisis, se deben incluir aspectos referidos al hogar entendido como una experiencia cultural de socialización cognitiva, social y de valores; también otros aspectos como el patrimonio de conocimientos, competencias y habilidades, como así mismo de acceso a la información. La distinción de estos componentes permite subrayar el carácter multidimensional del fenómeno de la inequidad social, por tanto, se reclama una comprensión integral del fenómeno (MARTINO, 2013).

Para interpretar mejor la naturaleza de la pobreza es necesario conocer los distintos puntos de vista que existen sobre la misma y que reflejan, de alguna manera, determinados intereses que son respaldados por los respectivos planteamientos teóricos o técnicos. Es por esta razón que no siempre son coincidentes los enfoques

de pobreza manejados por los organismos internacionales de crédito como el Banco Mundial y el Fondo Monetario Internacional, y los expresados por instituciones como la CEPAL, ONGs, organizaciones políticas y sindicales e investigadores independientes.

En documentos y publicaciones, el Banco Mundial establece que la pobreza: *"es hambre; es la carencia de protección; es estar enfermo y no tener con qué ir al médico; es no poder asistir a la escuela; no saber leer; no poder hablar correctamente; no tener un trabajo; es tener miedo al futuro; es vivir al día; la pobreza es perder un hijo debido a enfermedades provocadas por el uso de agua contaminada; es impotencia; es carecer de representación y libertad"* (MUNDIAL BANK, 2000/2001). En otra publicación se define a la pobreza como *"un fenómeno multidimensional, que incluye incapacidades para satisfacer las necesidades básicas, falta de control sobre los recursos, falta de educación y desarrollo de destrezas, deficiente salud, desnutrición, falta de vivienda, acceso limitado al agua y a los servicios sanitarios, vulnerabilidad a los cambios bruscos, violencia y crimen, falta de libertad política y de expresión"* (MUNDIAL BANK, 1999).

No existe un significado único del término "pobreza". En 1999, Paul Spicker identifica once posibles formas de interpretar esta palabra: *necesidad, estándar de vida, insuficiencia de recursos, carencia de seguridad básica, falta de titularidades, privación múltiple, exclusión, desigualdad, clase, dependencia y padecimiento inaceptable*.

Por su parte, en 1988 la CEPAL ha definido a la pobreza como "la situación de aquellos hogares que no logran reunir, en forma relativamente estable, los recursos necesarios para satisfacer las necesidades básicas de sus miembros". A ello puede agregarse que "la pobreza es un síndrome situacional en el que se asocian el infra consumo, la desnutrición, las precarias condiciones de vivienda, los bajos niveles educacionales, las malas condiciones sanitarias, una inserción inestable en el aparato productivo o dentro de los estratos primitivos del mismo, actitudes de desaliento y anomia², poca participación en los mecanismos de integración social, y quizás la adscripción a una escala particular de valores, diferenciada en alguna manera de la del resto de la sociedad" (ALTIMIR, 1979).

² Situaciones que derivan de la carencia de normas sociales o de su degradación.

En este orden de opiniones, Amartya Sen define a la pobreza como "la privación de capacidades básicas para funcionar dentro de la sociedad: una persona que carece de la oportunidad para conseguir ciertos niveles mínimos aceptables en dichas realizaciones o funcionamientos. Las realizaciones relevantes pueden comprender desde las físicas elementales, como estar bien alimentado, adecuadamente vestido, contar con un lugar digno donde vivir, evitar la morbilidad prevenible —entre otros muchos aspectos— hasta logros sociales más complejos como el nivel de participación que permita la sociedad" (SEN, 1984).

Dentro de la lógica planteada por Amartya Sen, Julio Boltvinik define a la pobreza como un proceso multidimensional en el que el bienestar de los hogares y las personas depende de seis fuentes: (i) el ingreso corriente; (ii) los activos no básicos y la capacidad de endeudamiento del hogar; (iii) el patrimonio familiar³; (iv) el acceso a bienes y servicios gratuitos; (v) el tiempo libre y el disponible para trabajo doméstico, educación y reposo y, (vi) los conocimientos de las personas⁴. Para este autor, los tres primeros representan lo que suele llamarse recursos económicos privados (expresables en términos monetarios); la cuarta fuente representa los recursos económicos públicos, llamados también "salario social". Finalmente, las dos últimas fuentes tienen sus propias unidades de medida: el recurso humano y el recurso tiempo (BOLTVINIK, 2003).

1.1.3 Autopercepción de la pobreza

En la mayoría de los estudios relacionados con la "cuantificación" de la pobreza, es frecuente encontrar un detallado análisis sobre las condiciones materiales de vida de las personas; se describe la situación de vivienda, educación, ingresos, ocupación y otros. Por otra parte, existe una amplia gama de estudios orientados a dimensionar el fenómeno de la pobreza. De manera similar, algunas investigaciones más complejas buscan vincular estadísticamente la relación entre pobreza, desocupación y

³ Entendido como el conjunto de activos y bienes durables que proporcionan servicios básicos a los hogares (vivienda y equipamiento doméstico básico).

⁴ Boltvinik señala que el conocimiento, no como medio para la obtención de ingresos, sino como satisfactores directos de la necesidad humana de entendimiento y como indicadores directos del grado de desarrollo cognitivo del ser humano.

crecimiento económico, con el fin de poder pronosticar los resultados que diversas líneas de acción pueden tener sobre los niveles de pobreza.

Sin pretender quitar importancia de tales estudios, se considera que, en general, éstos omiten un aspecto importante como es el diagnóstico de las causas y condiciones para salir de la condición de pobreza, desde el punto de vista de las actitudes y los esfuerzos que las propias personas pobres hacen por superar esas condiciones de vida.

Si bien existen importantes investigaciones que han registrado de manera cualitativa las condiciones comunitarias e individuales de vida de los más pobres, pocos trabajos han considerado las actitudes y los esfuerzos que realizan los propios pobres para mejorar su situación. Es importante considerar la potencialidad que ellos tienen, como sujetos de su propio desarrollo, para superar las condiciones socioeconómicas en las que viven (IRARRÁZVAL, 1995).

Por lo general, la literatura se refiere a "causales individuales" de la pobreza. Esto implica rescatar al individuo pobre y considerar que su opinión y sus perspectivas son elementos relevantes para la definición de la política social. Todo esto no significa, por cierto, un desconocimiento de las "causales estructurales" sino un enfoque diferente.

Ignacio Irarrázaval⁵ propuso crear diversos indicadores que registren el éxito económico relativo de los pobres de acuerdo con una combinación de variables. Siguiendo esta metodología, se creó un índice combinado que permite posicionar a cada familia de escasos recursos en un orden continuo según su nivel de surgimiento socioeconómico. El análisis de esta información permitió descubrir que existen ciertos patrones individuales, familiares y sociales asociados a este proceso. Así, se llegó a caracterizar cada uno de los extremos del orden continuo, mencionado previamente. Cerca de un extremo están las familias que registran condiciones objetivas de haber alcanzado un éxito socioeconómico relativo, dentro del segmento de pobreza en el que se encuentran. De esta manera, el grupo se caracteriza por tener actitudes, conductas

⁵ IGNACIO IRARRÁZVAL LLONA. Ph. D. en Planificación Social, London School of Economics. Profesor del Departamento de Economía de la Universidad de Chile. Investigador del Centro de Estudios Públicos.

y opiniones que reflejan una evidente aspiración a lograr una mejoría en su situación socioeconómica; muestran algunos avances concretos en esa dirección, a pesar de las restricciones propias del estado en el cual se encuentran. Por el contrario, en el otro extremo se ubican las familias que no muestran características de mejoramiento en sus condiciones socioeconómicas.

La visualización de la pobreza como un contrapunto entre las familias que están haciendo un esfuerzo por surgir económicamente y aquellas que no lo están, plantea un gran desafío para definir, a futuro, políticas sociales. Desde este punto de vista, se debería considerar al hogar "pobre" como centro del proceso de superación de la pobreza y se deberían crear incentivos para maximizar el esfuerzo individual.

Por otro lado, la psicología social se orienta hacia la acción comunitaria, concientizando y orientando a las comunidades a participar en la solución de sus propios problemas. Este proceso requiere que el grupo tenga acceso a recursos materiales y psicológicos y que, a través de éstos, adquiera mayor control sobre sus condiciones ambientales. En este contexto se presenta el desafío de no caer en el "asistencialismo-clientelismo" y desarrollar el compromiso con el cambio social (CASTILLO *et al.* 2006).

Asimismo, la Fundación Paraguaya⁶ implementó un programa denominado "Semáforo de Eliminación de Pobreza" con el fin de que cada familia determine su propio nivel de pobreza y desarrolle una estrategia personalizada para salir de esa situación en forma permanente.

El Semáforo de Eliminación de Pobreza es un instrumento que permite a las familias medir su nivel de pobreza e identificar estrategias personalizadas para solucionar sus carencias específicas. Define qué significa "no ser pobre" por medio de seis dimensiones: ingresos y empleo; salud y medio ambiente; vivienda e infraestructura; educación y cultura; organización y participación; interioridad y motivación. La Figura 1.1 ilustra las dimensiones de la pobreza, según la Fundación Paraguaya.

⁶ La **Fundación Paraguaya** es una empresa social sin fines de lucro, autosostenible, fundada en 1985 y desde entonces es pionera en microfinanzas y emprendedurismo en Paraguay.

Figura 1.1: Semáforo de la eliminación de la pobreza



Fuente: <http://fundacionirradia.org/2016/08/09/el-semaforo/> Fundación Irradia, agosto 2016

La fundación desarrolló esta innovación en base a una subdivisión de las dimensiones mencionadas en 50 indicadores, cada uno con tres niveles e imágenes simples que representan lo que es ser **pobre extremo (rojo)**, **pobre (amarillo)** y **no pobre (verde)**. Los hogares se autoevalúan con la asistencia de un software que refleja los tres niveles posibles de cada indicador, utilizando los colores de un semáforo (conocidos universalmente) e ilustraciones. De este modo, cada familia puede ver en qué consiste su pobreza y cuál es el siguiente nivel al cual pueden aspirar.

1.2 Marco conceptual

1.2.1 Marginalidad y dependencia asistencial

La década de los ochenta se ha caracterizado por una amplia polémica con relación a la efectividad de las políticas "antipobreza" implementadas en el mundo desarrollado. Dentro de este ámbito, uno de los fenómenos que ha adquirido mayor relevancia en la discusión académica sobre políticas sociales es el tema de la marginalidad.

El concepto de marginalidad aparece como un intento de equilibrar la investigación sobre la pobreza más allá de la cuestión puramente económica. Generalmente, la pobreza se asocia a un problema primordialmente monetario. Sin embargo, hay varios investigadores que analizan el fenómeno de la pobreza desde otro enfoque como el cultural, estilos de vida, aptitudes, valores y actitudes (FERES y MANCERO, 2001a, 2001b).

En 1989, la autora Van Haitsma establece un concepto de marginalidad muy utilizado y que determina que una persona se encuentra en condiciones de marginalidad o exclusión, cuando posee una vinculación endeble con el trabajo formal y cuyo contexto social tiende a debilitar aún más dicha vinculación. La relación con la capacidad de trabajo se refiere a la legitimidad de los ingresos y a la estabilidad o variabilidad de este. En este sentido, la legitimidad de los ingresos del trabajo dependerá de su origen; por ejemplo, los ingresos provenientes del crimen no tendrán legitimidad. Del mismo modo, las transferencias asistenciales del gobierno u otros organismos o instituciones tendrán en general, una legitimidad y estabilidad bajas. El contexto social se refiere a las estructuras sociales en las cuales el individuo está inserto: hogar, vecindario, redes sociales. Estos vínculos pueden o no reforzar este bajo nivel de relación con la fuerza de trabajo.

En cuanto a las causas del fenómeno de la marginalidad o exclusión, diversos autores han señalado una amplia gama de factores entre los que se destacan la dependencia asistencial, estructura familiar, bajo nivel educacional, crimen y descenso de los empleos urbanos. Algunos aspectos de estos elementos se mencionan brevemente en CORDERA CAMPOS *et al.*, (2008) y ZICCARDI, (2010).

La dependencia asistencial se refiere al hecho que las personas beneficiarias de los planes sociales monetarios del Estado realizan pocos esfuerzos por salir de la situación de pobreza en la cual están inmersas. De algún modo, los propios programas sociales contribuyen a generar situaciones de letargo y poca motivación de cambio en la situación sociocultural y económica; disminuyen los incentivos en la oferta laboral, lo que se traduce en la reducción de las horas de trabajo. *“Las políticas sociales asistencialistas se enfocan a mejorar las condiciones de vida, no las capacidades de las personas. Y este es el camino equivocado si se busca la erradicación de la pobreza y una mayor equidad social”* (ZAMAGNI, 2010)⁷.

En el aspecto familiar, algunas investigaciones han demostrado que, por lo general, existe una estrecha asociación entre las familias uniparentales, sobre todo si es una madre sola, y la marginalidad (MUJER, 2011). La importancia de la familia radica en la solvencia para transmitir valores que redundarán en la formación de las futuras generaciones.

El comportamiento “delictivo” es un tema que ha sido tratado considerando, principalmente, la evolución de la criminalidad. Se ha observado y demostrado que, existe una fuerte asociación entre mayor criminalidad y concentración de pobreza y exclusión o marginalidad en asentamientos urbanos ubicados, generalmente en los suburbios de las grandes ciudades con alta densidad de población (UNTREF, 2014).

El escaso o nulo acceso a la educación impacta en el problema de criminalidad, lo que lleva al surgimiento o profundizar aún más el problema de la marginalidad. Dentro de este contexto, sería interesante analizar cuál es la percepción que tienen los pobres y las personas en situación de exclusión respecto de la educación como vehículo de movilidad social.

En una investigación del año 1990, Woodrow Wilson postula que los cambios en la composición del empleo urbano han provocado la disminución de la demanda de trabajadores de baja calificación en las ciudades. En la actualidad, los empleos de baja calificación se encuentran en las periferias de las ciudades, generándose un desajuste espacial entre la oferta y la demanda por trabajo. De este modo, es posible encontrar

⁷ Stefano Zamagni, economista, profesor de la Universidad de Bologna y asesor papal en las cuestiones sociales.

barrios completos en los centros de las grandes ciudades en los cuales una muy alta proporción de personas adultas se encuentra desempleada. En estos lugares hay una marcada concentración de la marginalidad.

En resumen, existiría un consenso inicial entre los investigadores del tema en países desarrollados que, a partir de una prolongada permanencia como beneficiarios de los sistemas de asistencia social, podrían estarse generando incentivos perversos hacia la perpetuación y agravamiento de las condiciones de vida de ciertos grupos calificados de pobres. Por un lado, existe un amplio debate sobre los efectos de los programas sociales en términos de su capacidad para sacar a sus beneficiarios de la marginalidad y que puedan surgir, o bien para hacer que éstos tiendan a permanecer en la pobreza. En general, se pone en duda su efectividad debido a las relaciones de dependencia que generan. De la misma manera, se plantean diversos comportamientos y conductas que reforzarían el estancamiento de las personas en la pobreza. Es decir que, sería factible y relevante revisar hasta qué punto la población pobre presenta características de esfuerzo o estancamiento, acerca de superar su situación de pobreza.

Las investigaciones llevadas a cabo muestran componentes o factores que inciden en el afianzamiento del segmento de pobreza denominado marginalidad. En este sentido se consideran elementos como bajo nivel de educación, carencia de estructura familiar y trabajo urbano precario.

1.2.2 Pobreza y política social

La evolución de la política social en los países desarrollados suscita una variedad de interrogantes para el diseño de la política social de los países emergentes. De este modo, las estrategias de desarrollo social aplicadas en Argentina en la última década plantean incertidumbre para el futuro.

En la década de 1990 empieza a desarrollarse en Argentina, la globalización económica, caracterizándose por la apertura económica y su reorientación hacia el mercado externo, una modernización tecnológica importante, privatización de empresas de servicio y políticas de ajuste que terminaron produciendo el aumento del desempleo y la reducción del gasto público en coberturas sociales (SATRIANO, 2006).

Las políticas sociales concebidas posteriormente, como una estrategia del gobierno para intervenir en los vínculos sociales y paliar la pobreza, han demostrado una escasa capacidad para resolverlos porque no se localizan, precisamente, en el ámbito de la producción. Estas políticas son interpretadas sólo a la acción de los sectores sociales, parcializando la capacidad del Estado de responder a los efectos ocasionados por las políticas económicas y reduciendo su actuación a situaciones de emergencias.

En este sentido, los sectores más vulnerables se convierten en ámbitos de manipulación y clientelismo y el plano asistencial se limita a una provisión directa de algunos bienes de consumo (ayuda alimentaria, tickets familiares y/o cualquier otro subsidio "asistencialista") y puestos de trabajo precarizados. Por otra parte, es sabido que los costos sociales de este modelo se manifiestan de diversas maneras en las poblaciones pobres, porque los nuevos patrones de acumulación incrementaron la distribución desigual de los ingresos. La metodología de intervención opera atendiendo a un grupo de población que queda fuera del sistema y se lo incorpora a programas residuales y de bajo costo en el gasto público. Este sistema invierte el derecho ciudadano al constituir un clientelismo que intercambia favores basados en la idea de caridad.

En un trabajo del año 2000, Bustelo plantea que *"el pobre no es tan sólo pobre porque tiene carencia de bienes materiales, sino porque además es hecho pobre para constituirlo como dependiente de quien le da la dádiva y administra favores"*. En consecuencia, la salud pública y la educación para los sectores pobres quedan reducidas a una asistencia básica lejos del concepto de justicia distributiva.

Actualmente, el clientelismo político es uno de los problemas del asistencialismo, pero no el único. Sea clientelista o universal, el asistencialismo perpetúa la dependencia mental y entorpece la capacidad de los asistidos para convertirse en ciudadanos responsables de su futuro. La educación de calidad y la inversión que crea empleo son el único camino para disminuir la pobreza (HALPERIN *et al.*, 2008).

Por otro lado, también como un antecedente válido, es importante destacar que, en un contexto recesivo e inflacionario, el Estado argentino amplió el régimen de protección social dirigido a la población infantil en situación de vulnerabilidad a través de la "Asignación Universal por Hijo para la Protección Social" (AUH) (Decreto

1602/2009 –en vigencia desde noviembre de 2009). Este programa de transferencia de ingresos se sumó a otras líneas de asistencia económica directa, y/o indirecta preexistente, tales como créditos fiscales, pensiones no contributivas y/o seguridad social contributiva. Para acceder a dicho beneficio, el adulto responsable (padre, madre o tutor) y el niño/a, deben cumplir con ciertos requisitos. El beneficio está dirigido a hijos menores de 18 años o discapacitados de padres desocupados, monotributistas sociales, trabajadores no registrados o empleadas domésticas cuyos ingresos no superen el salario mínimo vital y móvil. Según un informe del Observatorio de la Deuda Social de la Universidad Católica Argentina, la AUH tuvo un impacto positivo importante, ya que se observó un aumento promedio del 25–30% en los ingresos medios de los hogares en situación de indigencia y del 10-15% en situación de pobreza. También, redujo en casi 20%, el riesgo de inseguridad alimentaria, en más de 60% el riesgo de no asistencia escolar y en 14% el riesgo de trabajo infantil, en el período 2010-2013.

Asimismo, el concepto de pobreza está relacionado con los derechos humanos y éstos son inherentes a las personas sin tener en cuenta su condición social. Las políticas públicas con perspectivas de derechos humanos deben considerar al individuo como “sujeto de derecho”, en tal sentido facultan a los pobres para que reivindiquen sus derechos económicos, sociales y culturales: alimentos, vivienda, educación, atención de salud, un trabajo digno y bien remunerado, seguridad social y participación en la toma de decisiones. Estos derechos les permiten exigir que se les rinda cuenta por la prestación de buenos servicios públicos, por políticas públicas a favor de los pobres y por un proceso participativo transparente y abierto, donde sus opiniones sean escuchadas, es decir que puedan ejercer libremente sus derechos civiles y políticos.

1.3 Medición de la pobreza

Como se observa y plantea en los párrafos anteriores, no es sencillo ni fácil definir la pobreza. Sin embargo, todo indica que es de esencia multidimensional y por lo tanto no se la puede abordar desde un solo ángulo, sino que debe ser presentada como una situación compleja que involucra factores de índole económica, social,

1602/2009 –en vigencia desde noviembre de 2009). Este programa de transferencia de ingresos se sumó a otras líneas de asistencia económica directa, y/o indirecta preexistente, tales como créditos fiscales, pensiones no contributivas y/o seguridad social contributiva. Para acceder a dicho beneficio, el adulto responsable (padre, madre o tutor) y el niño/a, deben cumplir con ciertos requisitos. El beneficio está dirigido a hijos menores de 18 años o discapacitados de padres desocupados, monotributistas sociales, trabajadores no registrados o empleadas domésticas cuyos ingresos no superen el salario mínimo vital y móvil. Según un informe del Observatorio de la Deuda Social de la Universidad Católica Argentina, la AUH tuvo un impacto positivo importante, ya que se observó un aumento promedio del 25–30% en los ingresos medios de los hogares en situación de indigencia y del 10-15% en situación de pobreza. También, redujo en casi 20%, el riesgo de inseguridad alimentaria, en más de 60% el riesgo de no asistencia escolar y en 14% el riesgo de trabajo infantil, en el período 2010-2013.

Asimismo, el concepto de pobreza está relacionado con los derechos humanos y éstos son inherentes a las personas sin tener en cuenta su condición social. Las políticas públicas con perspectivas de derechos humanos deben considerar al individuo como "sujeto de derecho", en tal sentido facultan a los pobres para que reivindiquen sus derechos económicos, sociales y culturales: alimentos, vivienda, educación, atención de salud, un trabajo digno y bien remunerado, seguridad social y participación en la toma de decisiones. Estos derechos les permiten exigir que se les rinda cuenta por la prestación de buenos servicios públicos, por políticas públicas a favor de los pobres y por un proceso participativo transparente y abierto, donde sus opiniones sean escuchadas, es decir que puedan ejercer libremente sus derechos civiles y políticos.

1.3 Medición de la pobreza

Como se observa y plantea en los párrafos anteriores, no es sencillo ni fácil definir la pobreza. Sin embargo, todo indica que es de esencia multidimensional y por lo tanto no se la puede abordar desde un solo ángulo, sino que debe ser presentada como una situación compleja que involucra factores de índole económica, social,

educacional, moral, política y hasta natural, como es el caso de los desastres naturales, migración forzosa de la población debido a guerras civiles y externas, que llevan a la miseria a millones de personas en el mundo. En tal sentido, las políticas propuestas para disminuir la pobreza deben ser de carácter integral concentrando su atención tanto en el mejoramiento material de las personas, como en el verdadero desarrollo de sus capacidades y fortalecimiento de sus derechos.

Así como la definición de la pobreza, su medición también es compleja y depende del punto de vista teórico con que se la plantee. Entonces, para su estimación se utilizan diversos métodos, desde los cuantitativos esenciales, hasta diseñar modelos complejos que incluyan variables de difícil medición. Para distinguir a los "pobres" de los "no pobres" es necesario comparar el bienestar de diferentes personas para evaluar si alguna de ellas tiene un nivel menor al "mínimo razonable", establecido por la sociedad. Cada forma de medir la pobreza tiene un indicador de bienestar y los resultados que se obtengan, probablemente serán muy sensibles al indicador elegido (FERES & MANCERO, 2001a).

Una vez elegido el o los indicadores de bienestar, es conveniente determinar el método que permita responder la pregunta: "¿desde qué nivel de bienestar se considera que una persona no es pobre?". Existen diferentes enfoques según se considere a la pobreza como un fenómeno absoluto o relativo, se mida "capacidad de consumo" versus "consumo efectivo", o se considere que no hay nadie mejor que los propios pobres para definir su situación a través de encuestas (INE, 2005).

1.3.1 Métodos directo e indirecto

Se puede adoptar un enfoque "directo", o uno "indirecto", para determinar si una familia es pobre. En el enfoque directo, un hogar es pobre si no satisface alguna o todas las necesidades básicas (hacinamiento, vivienda inconveniente, condiciones sanitarias, asistencia escolar, capacidad de subsistencia). Mientras que el enfoque indirecto considera pobres a aquellas personas que no cuentan con los recursos suficientes para satisfacer sus necesidades básicas.

En resumen, el método directo relaciona el bienestar con el consumo efectivamente realizado y el indirecto calcula el bienestar a través de la capacidad para realizar consumo.

El método indirecto se caracteriza por utilizar "líneas de pobreza", que establecen el ingreso o gasto mínimo que permite tener un nivel de vida conforme a ciertos estándares establecidos. Según este concepto, se considera que una familia es pobre si sus ingresos son menores a la "línea de la pobreza". Existen varias metodologías para construir la "línea de la pobreza", como se describe a continuación.

- a) *Consumo calórico:* Bajo este método la línea de pobreza corresponde al nivel de ingresos (o de gastos) con el cual una persona alcanza un consumo de calorías preestablecido. Las necesidades calóricas derivan de estudios nutricionales, suponiendo cierto nivel de actividad física. Es importante destacar que este procedimiento no equivale a medir la desnutrición, la cual se efectúa por observar si el consumo efectivo de calorías es insuficiente sin que se requiera conocer el ingreso monetario correspondiente.

En este contexto, el cálculo de la línea de la pobreza se realiza de dos maneras: una es seleccionando una submuestra de hogares con un consumo calórico aproximado al requerido y usar un ingreso promedio como línea de la pobreza. La otra forma es correr una regresión entre consumo calórico e ingreso y con la relación encontrada evaluar el ingreso necesario para consumir las calorías predeterminadas.

La ventaja de este método es que requiere de poca información y que no es necesario fijar expresamente un componente "no alimentario" en la línea de la pobreza. Por otro lado, la relación entre gasto y consumo energético puede variar según niveles de actividad física, precios relativos, preferencias, etc. y estas variaciones no necesariamente pueden producir cambios en el bienestar. Entonces no es un indicador de bienestar adecuado porque carece de cierta consistencia con otros indicadores como el gasto en consumo.

- b) *Método del costo de las necesidades básicas:* Este método utiliza una canasta básica de bienes (alimentarios y otros) y servicios; la línea de

pobreza es el gasto necesario para adquirir esa canasta básica. Se considera que la canasta está compuesta por los "bienes alimentarios" y "otros bienes". La canasta de bienes alimentarios es convenida de modo que satisfaga las necesidades básicas de alimentación y que minimice el costo de los nutrientes, según los precios vigentes. Sin embargo, esa combinación puede generar una canasta que no es compatible con las preferencias de la población.

Por otro lado, se presentan dos alternativas para la construcción de la canasta de "otros bienes". Una es proceder de manera similar que en el caso de la de alimentos e identificar requerimientos mínimos de cada necesidad como vivienda, vestimenta, educación, transporte, etc. Pero este procedimiento puede resultar poco apropiado porque la construcción de la línea de la pobreza estaría influenciada por las preferencias particulares de quienes participan.

La otra alternativa es no especificar los requerimientos mínimos sino utilizar la proporción observada de gastos en estos bienes, del gasto total de los hogares, en un grupo específico de la población. De este modo, la línea de pobreza se obtiene de dividir el valor de la canasta básica alimentaria por la proporción del gasto en alimentos. Esto se denomina "Método del Multiplicador" o "Método de Orshansky" (es el inverso del coeficiente de Engel).

Estas formas de cálculo presentan ciertas inconsistencias porque, por un lado, supone que los hogares logran cubrir sus necesidades de alimentación y estándares mínimos de otras necesidades "básicas" y por otro, no toma en cuenta la satisfacción de necesidades básicas provenientes de bienes y servicios, "gratuitos", suministrados por el Estado (por ejemplo, educación y salud). Tampoco se tiene en cuenta que el gasto en "otros bienes" depende del tamaño, la composición y la etapa del ciclo de vida de cada familia, entonces la elección de un coeficiente único para la construcción de la línea de la pobreza es "discutible".

A pesar de que este método intenta brindar un fundamento teórico objetivo de la construcción de la línea de la pobreza, no cabe duda de

que existen juicios de valor subjetivos, que se manifiestan en la preferencia del tipo de alimentos que satisfacen las necesidades nutricionales, en sus calidades y precios y en la construcción del multiplicador de Orshansky (AGUILAR GARZA & PARADA RIVERA, 2007).

1.3.2 Carencia y carencia severa

Los Métodos de Línea de Pobreza (LP) y de Necesidades Básicas Insatisfechas (NBI) descritos en los párrafos anteriores, son los más difundidos para la medición de la pobreza en la Argentina, pero cuestiones metodológicas propias (relacionadas, por ejemplo, con la periodicidad del relevamiento de información censal o mediciones del Índice del Costo de Vida y el Índice de Precios al Consumidor (IPC) para calcular el costo de una canasta básica de alimentos y bienes y servicios), generaron la necesidad de buscar mediciones alternativas⁸.

A diferencia de la línea de pobreza e indigencia que mide el Instituto Nacional de Estadística y Censos (INDEC) a partir del valor de una canasta (Básica total o Alimentaria según el caso), el indicador de "carencia" considera un porcentaje de la mediana de los ingresos de los hogares según la Encuesta Permanente de Hogares (EPH).

El cálculo parte de tomar la "mediana" del ingreso total familiar per cápita, con la cual se computan los niveles de carencia y carencia severa para cada hogar, de acuerdo con su número de integrantes, y se los compara con el ingreso total familiar. Para el caso de carencia, se toma el setenta y cinco por ciento (75%) de esa mediana y, para carencia severa, el veinticinco por ciento (25%), de esta forma es posible clasificar a los hogares en: "sin carencias", "con carencias" o "con carencias severas". Esta metodología tiene su antecedente en la esgrimida por el organismo estadístico europeo EUROSTAT, que utiliza estos índices para determinar la línea de pobreza o umbral -fija el umbral de pobreza en el sesenta por ciento (60%) de la mediana de la

⁸ La provincia de Córdoba avanzó en una nueva medición dirigida a calcular los conceptos de carencia y carencia severa con base en ingresos y no en una canasta de consumo, estableciendo la utilización de una metodología indirecta y relativa centrada en el ingreso total familiar del hogar.

distribución de los ingresos por unidad de consumo (adulto equivalente)-. También se utilizó dicha metodología para identificar (incluir/excluir) hogares con carencias como universo objetivo para planes asistenciales, tal como lo hizo la Dirección General de Estadística y Censos de la Provincia de Córdoba a pedido de su Ministerio de Desarrollo Social. La Figura 1.2 muestra el esquema de la metodología de cálculo.

Figura 1.2: Propuestas de Carencias y Carencias Severas. Metodología utilizada como una aproximación a la medición de la pobreza



Fuente: Dirección de Estadística de la Provincia de Córdoba. Argentina

Las magnitudes para el cálculo de carencia y carencia severa se definen como sigue:

- Me_{ipcf} = Mediana del Ingreso total familiar per cápita
- n = cantidad de personas en el hogar
- a = aporte marginal de cada persona extra del hogar = 0,74

(a = 0,74 identifica al coeficiente de "adulto equivalente" de una mujer entre 18 y 59 años y al promedio de los "adultos equivalentes" correspondientes a las personas

según sexo y edad, parámetros utilizados en la metodología de línea de Pobreza e Indigencia que calcula el INDEC – método oficial de cálculo de la pobreza en Argentina – para representar las necesidades alimentarias marginales de cada persona).

Carencia severa:

Un hogar con carencia severa es aquel en el que ingreso total familiar menor o igual al monto resultante de:

$$\underbrace{0,25 \times Me_{ipcf}}_{\text{Primera persona}} + \underbrace{(a \times 0,25 \times Me_{ipcf} \times (n-1))}_{\text{Resto de las personas}} \quad (1.1)$$

Carencia:

Un hogar con carencia es aquel en el que el ingreso total familiar menor o igual al monto resultante de:

$$\underbrace{\underbrace{0,25 \times Me_{ipcf}}_{\text{Primera persona}} + \underbrace{(a \times 0,25 \times Me_{ipcf} \times (n-1))}_{\text{Resto de las personas}}}_{\text{Primera persona}} + \underbrace{\underbrace{0,50 \times Me_{ipcf}}_{\text{Primera persona}} + \underbrace{(a \times 0,25 \times Me_{ipcf} \times (n-1))}_{\text{Resto de las personas}}}_{\text{Resto de las personas}} \quad (1.2)$$

Esta medición unidimensional basada en el ingreso familiar se usa habitualmente en Europa. En Argentina nunca se utilizó, sólo se la adoptó en la Provincia de Córdoba durante un período de tiempo en que no se realizaban mediciones oficiales de pobreza, durante los años 2014 y 2015. Esta metodología no fue aceptada por el INDEC.

1.3.3 Indicadores de bienestar

Comparando distintos niveles de bienestar se podría realizar la "identificación" de los pobres y para definir esos niveles, se debe determinar una variable cuantificable que represente al indicador del nivel de bienestar de las personas u hogares. Esta elección dependerá del concepto adoptado, pero también de la información disponible, que por lo general es insuficiente y/o deficiente. Algunos de estos indicadores se describen a continuación:

- a) *Ingreso vs consumo*: Estos indicadores de bienestar son los más utilizados en el análisis de la pobreza. Esto se debe a que, en general, el análisis económico estándar define a la pobreza como "nivel de vida", restringiendo el concepto al ámbito material.

Si la utilidad de las personas sólo dependiera del consumo presente, entonces, desde el punto de vista teórico, independientemente de la forma de financiamiento de tal consumo, el indicador más apropiado sería el consumo corriente y el ingreso sólo sería una aproximación al nivel de vida (GLEWWE & VAN DEL GAAG, 1990); (RAVALLION, 1992). En este contexto, se admitiría que el ingreso corriente podría subestimar o sobrestimar el nivel de vida. (ATKINSON, 1991). Sobrestima si la familia ahorra, por lo tanto, no todo el ingreso se utiliza para el gasto o bien hay racionamiento. La subestimación se produce cuando la familia recibe un crédito, entonces el consumo corriente no se ve restringido por el ingreso.

Este argumento no es, en absoluto, concluyente. La elección entre consumo e ingreso depende del objetivo propuesto para realizar la medición de la pobreza.

- b) *Ingreso total familiar vs ingreso per cápita*: Es razonable pensar que las necesidades de un hogar aumentan en la medida que aumenta el número de sus miembros, por lo que un indicador que tenga en cuenta el tamaño del hogar sería más apropiado que el ingreso total. De este modo, la medición de la pobreza se inclina a cuantificar el bienestar individual.

Sin embargo, es conveniente tener en cuenta que las necesidades individuales no son iguales para todos los miembros del hogar, dependen de ciertas características como, edad, género y actividad que realiza. En este sentido, sería preferible usar el indicador que considere las características determinantes de las necesidades individuales como el ingreso total ajustado por una "escala de equivalencias"⁹.

⁹ En algunos casos es difícil confeccionar esta escala de equivalencias porque no se dispone de información suficiente.

Este indicador, como los otros, tiene sus ventajas y desventajas, entre las primeras se puede mencionar que capta las diferencias en necesidades según el sexo, la edad, actividad y/o características demográficas. Mientras que la desventaja que posee es que la estimación de la demanda se realiza a partir de la información de encuestas de gastos y esta metodología carece de un problema teórico fundamental, centrado en que la demanda observada no tiene suficiente información para determinar adecuadamente niveles de bienestar.

- c) *Proporción del gasto en alimentos*: El uso de la proporción del gasto en alimentos de un hogar como indicador de bienestar, se basa en la ley de Engel, según la cual la proporción del gasto en comida tiende a disminuir cuando el ingreso aumenta, entonces el indicador de bienestar es la proporción de gasto en bienes "no alimentarios" (INDEC, 2003a).
- Como en el caso de los indicadores analizados anteriormente, éste también tiene ventajas y desventajas. A modo de ventaja, no requiere de ajustes según el tamaño del hogar, ni información sobre precios o ajuste por inflación. La desventaja es que no siempre se cumple para hogares muy pobres. La proporción gastada en alimentos difiere entre hogares, no sólo por las diferencias en el nivel de ingresos, sino también por los precios relativos, por las preferencias de la familia o por características demográficas distintas.
- d) *Indicadores nutricionales*: El consumo calórico que tiene una persona en el hogar, puede ser un parámetro de la calidad de vida dentro de ese hogar. Este concepto tiene sentido si se considera que la desnutrición es elemental en el fenómeno de la pobreza, sin embargo, sólo es una parte del "nivel de vida" y no un sinónimo de pobreza. Por esta razón y porque es complicado determinar los requerimientos nutricionales apropiados, es que este indicador es cuestionable.
- e) *Método antropométrico*: Se podrían conseguir aproximaciones de indicadores nutricionales (o indicadores de salud), tales como "talla según edad" y "peso según talla". Este indicador puede resultar útil

cuando se analiza el efecto adverso de la pobreza en la niñez y juventud, pero no permite realizar comparaciones de pobreza a nivel de toda la población. Se lo cuestiona en el sentido que puede ser incompatible con conceptos más amplios de bienestar.

- f) *Necesidades Básicas Insatisfechas (NBI)*: El método directo más utilizado para medir la pobreza en América Latina, es el denominado "Necesidades Básicas Insatisfechas" o NBI. Se basa en verificar si los hogares han satisfecho una serie de necesidades, previamente determinadas y considera pobres a los que no lo lograron. Este método sólo utiliza información del momento, no tiene en cuenta la capacidad del hogar para satisfacer las necesidades en el futuro.

La medición de la pobreza por este método requiere de ciertos ítems, como elección de las características de los hogares, que presentan algún marco significativo de privación y que están suficientemente asociados a estados de pobreza como para representar a las demás carencias que conforman tales situaciones (FERES & MANCERO, 2001b). Este documento señala que en Argentina los ingresos per cápita bajo la línea de la pobreza, están significativamente relacionados con el hacinamiento, que es el indicador más fuerte de NBI, pero no con la tenencia de la vivienda ni con la edad del jefe de hogar.

Si bien, las características propias de un país determinan cuál es el indicador más apropiado para medir la pobreza allí, hay ciertas carencias que se consideran el denominador común de la aplicación de esta metodología; las cuales son: (i) hacinamiento, (ii) vivienda inadecuada, (iii) inadecuado abastecimiento de agua potable, (iv) carencia o deficiencia del sistema sanitario, (v) no concurrencia a las escuelas de los menores en edad escolar, (vi) un indicador indirecto de capacidad económica.

Las privaciones (i y ii) caracterizan el acceso a una vivienda digna y se vinculan con la necesidad de privacidad, higiene y de protegerse del medio ambiente. La ausencia de estos elementos deteriora sustancialmente la calidad de vida (CEPAL & PNUD, 1989).

Las carencias (iii y iv) tienen que ver con la disponibilidad de agua y el acceso a servicios sanitarios básicos. La disponibilidad se refiere al abastecimiento de agua de buena calidad y cantidad suficiente para satisfacer las necesidades de consumo e higiene, lo cual se mide por sus características de potabilidad y la forma en que se suministra. También se distinguen dos características en el acceso al servicio sanitario, la disponibilidad de servicio higiénico y el sistema de eliminación de aguas servidas. En cuanto a la educación básica (v), se considera un requerimiento mínimo para que las personas puedan incorporarse a la vida productiva y social. Esta necesidad primordial consiste, no sólo en la asistencia a la escuela, sino también en recibir educación de calidad; lamentablemente, sólo se dispone información del primer requisito. La capacidad económica (vi), particularmente no mide una necesidad básica, sino que trata de reflejar la probabilidad que tiene la familia de generar recursos suficientes y capacidad de consumo. Este indicador toma en cuenta el nivel educacional del jefe de hogar para conocer, aproximadamente, los recursos que podría obtener y el número de personas en el hogar que dependen de los que aportan. Con estos parámetros se obtendría una idea de las necesidades que se cubrirían con dichos ingresos.

En conclusión, el método de las NBI es particularmente, adecuado para caracterizar la situación en la que viven los hogares carenciados, lo cual es muy útil a la hora de diseñar políticas enfocadas en aliviar determinadas necesidades básicas. A través de información obtenida de los censos de población y vivienda es factible registrar detalladamente la evolución de las necesidades básicas insatisfechas y construir "Mapas de Pobreza" que permitan identificar geográficamente esas privaciones y optimizar el gasto social destinado a brindar algún paliativo. Sin embargo, este método presenta algunas limitaciones como alternativa para la medición de la pobreza. El criterio que utiliza esta metodología para determinar si un hogar es pobre o no, es el no

satisfacer al menos una necesidad considerada básica pero dentro de esta cifra habrá hogares con diferentes grados de pobreza¹⁰.

1.3.4 Otro indicador de bienestar: índice de bienes

Como se describió en la sección anterior, la alternativa de medir la pobreza, a través de los ingresos personales, ingresos per cápita y/o cálculo de la línea de la pobreza, presenta ciertos problemas, tanto en su relevamiento (altas tasas de no respuesta en las encuestas), como en su análisis. Por otro lado, las disparidades en los bienes durables disponibles en los hogares suelen implicar el resultado de la "riqueza" acumulada mostrando otra dimensión de las desproporciones (MINUJIN & HEE BANG, 2002).

En diversos trabajos de investigación sobre alternativas para cuantificar la pobreza, se utiliza un "índice de bienes", basado en las características de la vivienda y en bienes durables del hogar y luego para determinar la distribución de los hogares se aplica el Método de Componentes Principales.

Estos trabajos se asemejan a estudios basados en el análisis de los denominados "activos sociales" o "índice de bienes", cuyo concepto se refiere al conjunto de recursos sociales (bienes durables y simbólicos-culturales) que inciden sobre las condiciones de bienestar de los hogares y sus perspectivas en el mediano y largo plazo. La construcción del índice de bienes se fundamenta en un conjunto de

¹⁰ Durante la década de los '90, se produjeron avances teóricos y metodológicos sobre la temática de la pobreza y las técnicas utilizadas para medir el fenómeno que pusieron de manifiesto algunas limitaciones importantes del método de NBI para la identificación de hogares pobres. El INDEC no ha sido ajeno a estos debates, mostró la necesidad de plantear dos modalidades respecto del estudio de la pobreza en la Argentina a partir de los datos del Censo 2001 orientando los esfuerzos en la elaboración de una nueva metodología que supere las limitaciones mencionadas. En ese contexto se elaboró el Índice de Privación Material de los Hogares (IPMH), metodología de identificación y agregación de las diferentes situaciones de pobreza, según el tipo y la intensidad de las privaciones que afectan a los hogares. La metodología del IPMH permite la identificación y agregación de las diferentes situaciones de pobreza. En la construcción del IPMH surgen cuatro categorías, ellas son: (a) hogares sin privación, (b) hogares con privación sólo de recursos corrientes, (c) hogares sólo con privación patrimonial (o de origen estructural); y (d) hogares con privación convergente (de recursos corrientes y patrimonial). Se consideran hogares con privación patrimonial aquellos que habitan en una vivienda con piso de tierra, techos sin cielorraso: de chapa, fibrocemento, plástico, cartón, caña, tabla, paja con barro, paja sola y que carecen de inodoro con descarga de agua. Para aproximarse a la definición de hogares con privación de recursos corrientes, se construye un indicador que tiene en cuenta la relación de los años de educación formal aprobados por los ocupados, jubilados y pensionados del hogar y la cantidad total de miembros del hogar, que de manera indirecta determina la capacidad económica de los hogares.

indicadores relacionados con las características de la vivienda y la propiedad de bienes durables específicos.

El supuesto de la selección de indicadores es que la posesión de estos bienes - en forma individual o combinada- constituyen elementos diferenciales y diferenciadores en términos sociales. Se considera, en general, que el acceso a ellos revela la situación patrimonial de los hogares, acorde a niveles y calidad de ingresos permanente (efecto de la riqueza acumulada), los cuales no están directamente relacionados con los niveles de ingresos corrientes. La ventaja principal de elaborar un índice numérico deriva de transformar un conjunto de variables categóricas en un indicador que sintetice el efecto conjunto de las propiedades de cada indicador como así también de las relaciones existentes entre ellos. Por otro lado, el hecho de ser un indicador numérico posibilita su utilización en procedimientos de análisis estadísticos que requieran de variables de este tipo, o puede ser "cerrado" en categorías cualitativas a partir de su ordenamiento en distintas clases de intervalos. Estas propiedades de características técnicas junto al concepto de obtener una medición de inequidad social confieren sentido a la construcción y evaluación de este índice de bienes.

Luego de la construcción del índice es importante determinar el peso correspondiente para cada ítem dentro de este índice numérico, para lo cual se utiliza el Método de Componentes Principales. Esta metodología es conveniente porque permite ponderar los indicadores en base a la forma y grado en que se encuentran asociados. Con el resultado se asigna a cada hogar un valor numérico que define su posición dentro del índice. De este modo los hogares son ubicados y agrupados a partir de los distintos niveles de asociación entre los indicadores seleccionados y de las formas de combinación de bienes durables que poseen. La ponderación de los distintos ítems se basa en los niveles de asociación a partir de combinaciones diferenciales existentes entre ellos.

El primer paso en el proceso de construcción del índice es la selección de los bienes que se incluyen a partir de los disponibles en las fuentes de datos (censos y/o encuestas específicas). El principal criterio es incluir bienes considerados *a priori* como socialmente diferenciadores.

Los ítems generalmente seleccionados por dimensión para la construcción del índice de bienes son los siguientes:

a) Características de vivienda:

- i) *Vivienda de materiales de alta calidad: paredes de ladrillo, piedra, bloque u hormigón; pisos de mosaico, madera, cerámica, alfombra; techo de teja, losa, baldosa, membrana, chapas metálicas, de fibrocemento o plástica.*
- ii) *Vivienda de materiales de baja calidad: paredes de adobe, chapa de metal, fibrocemento o plástico, cartón, desechos u otros; pisos y techos de cualquier tipo.*
- iii) *Vivienda con inodoro con descarga de agua*
- iv) *Vivienda con inodoro sin descarga de agua*
- v) *Vivienda sin inodoro*
- vi) *Vivienda con energía eléctrica*
- vii) *Obtención de agua por bomba (manual o motor)*
- viii) *Obtención de agua por fuente no entubada*
- ix) *Agua por cañería dentro de la casa*
- x) *Cocina con leña, carbón u otro combustible (exceptuando gas, electricidad y kerosene)*

b) Bienes durables:

- i) *Televisor color*
- ii) *Televisión por cable*
- iii) *Video grabador/reproductor*
- iv) *Teléfono fijo*
- v) *Teléfono celular*
- vi) *Computadora y conexión a internet*

- vii) *Heladera con freezer o freezer independiente*
- viii) *Heladera sin freezer*
- ix) *Lavarropas automático*
- x) *Horno microondas*
- xi) *Moto o motoneta*
- xii) *Automóvil hasta 5 años de antigüedad*

De la bibliografía consultada se destaca la conveniencia de aplicar el índice de bienes en la identificación de grupos característicos en la distribución diferencial de acceso de bienes y servicios sociales. En este contexto resulta un instrumento apropiado para el estudio de la inequidad social, especialmente cuando no se disponen de datos certeros sobre ingreso de los hogares, o hay una alta tasa de no respuesta en las encuestas de hogares. Sin embargo, existen algunas limitaciones, tanto en su construcción como en su aplicación.

Respecto a la construcción, el principal problema conceptual se centra en la selección final de los bienes que lo constituirán. Esto significa que por la forma que fue construido, el procedimiento anula los efectos "sobredimensionadores" de ítems altamente correlacionados y asigna pesos diferenciales, pero resulta un problema la determinación del conjunto de ítems que se deberían incluir, ya que la metodología puede ser aplicada en forma extendida incluyendo otros bienes que no están relacionados con los bienes durables. Pero queda el interrogante de: ¿en cuánto aportaría en la precisión del índice de bienes, la incorporación de otros indicadores no disponibles como la propiedad y valor de la vivienda además de otros bienes de capital?

En cuanto a su aplicabilidad también presenta algunas limitaciones. La condición fundamental para la composición del índice es la inclusión de bienes con capacidad de marcar diferencias en términos sociales. Sin embargo, su carácter diferencial puede variar sensiblemente en el tiempo. Por ejemplo, el acceso al servicio de telefonía celular o poseer computadora personal con conexión a Internet en la actualidad resulta menor que hace diez años (período entre censos). Asimismo, resulta difícil la

comparación en el tiempo ya que no se podría mantener el mismo conjunto de bienes. Las comparaciones entre censos deberían ser adecuadas a partir de un conjunto de bienes análogos para uno y otro, debido a las eventuales alteraciones de la capacidad diferencial de los mismos. Si bien, los resultados podrían ser comparables, habría una pérdida de valor explicativo, diferenciador de cada uno de los bienes que incluyan uno y otro relevamiento porque no se podría hacer una asociación uno a uno estricta, entre los estratos que se definan en cada caso.

Las perspectivas de aplicación y análisis considerando el índice de bienes pueden ser numerosas, debido, esencialmente, a la facilidad de relevamiento de los indicadores involucrados y a su condición de "proxy" del ingreso permanente. En este sentido se podría profundizar en el análisis de los grupos de hogares formados por el entrecruzamiento de los quintiles basados en el índice de bienes y aquellos construidos a partir del ingreso corriente de los hogares. La identificación de los distintos grupos conformados por este cruce resulta significativa en tanto vincula dos indicadores relacionados uno al corto plazo y el otro al mediano-largo, lo cual permite una aproximación a los efectos de los fenómenos dinámicos como los procesos de desigualdad social.

Otra línea de estudio consiste en el análisis de disparidades entre áreas geográficas definidas por divisiones políticas, lo que resultaría útil a la hora de diseñar programas y políticas sociales.

Por otro lado, la aplicación de esta metodología constituye un criterio más congruente de clasificación de favorecidos comparado con los índices clásicos de pobreza unidimensional basados en los ingresos corrientes.

1.4 Medición de pobreza multidimensional

Existe un consenso creciente respecto de la necesidad de que los países dispongan de mediciones multidimensionales de la pobreza, a fin de contar con información complementaria a la obtenida mediante los métodos convencionales, basados en los ingresos monetarios. En ello han influido distintos factores, como el surgimiento y predominio de nuevos enfoques sobre el desarrollo social y el bienestar, la constatación de las limitaciones de los indicadores monetarios como

aproximaciones al estándar de vida, la necesidad de una mayor alineación entre los indicadores de pobreza y las políticas tendientes a superarla

Este proceso se ha visto impulsado, en América Latina y otras regiones, por el predominio de nuevos marcos conceptuales sobre el desarrollo social y el bienestar, como el enfoque de derechos y las perspectivas de capacidades, que van más allá de la noción de pobreza limitada a la insatisfacción de necesidades básicas.

La medición multidimensional de la pobreza consiste en evaluar si las personas logran alcanzar umbrales mínimos de bienestar (o estándar de vida) en cada una de las dimensiones consideradas. Por tanto, de manera similar a la medición monetaria, se requiere información desagregada al nivel de las personas o de los hogares, como la que se obtiene de las encuestas de hogares o los censos de población.

En la selección de indicadores y umbrales se debe tomar muy en cuenta la disponibilidad de información relevante en una fuente de referencia única. Para ilustrar mejor esta necesidad, se puede recurrir a la experiencia de Latinoamérica. Una de las iniciativas pioneras en la medición multidimensional de la pobreza fue el método de las Necesidades Básicas Insatisfechas (NBI), promovido por la CEPAL en la década de 1980 como una alternativa para aprovechar la información de los censos de población en la región. La selección de dimensiones del método NBI estuvo fuertemente condicionada por las variables disponibles en los censos. Si bien se incluyeron dimensiones relevantes para la satisfacción de necesidades, como vivienda, agua y saneamiento, educación y capacidad de consumo, se excluyeron otras, como nutrición, salud y empleo.

Las metodologías que se vienen desarrollando, en general, presentan ciertas similitudes en la consideración de las dimensiones y la elección de los indicadores usados para su estimación.

Es así como a partir de 2008 en México se promulga la Ley General de Desarrollo Social (LGDS) que hace obligatoria la medición multidimensional de la pobreza y establece que deben utilizarse por lo menos ocho indicadores: I. Ingreso corriente per cápita; II. Rezago educativo promedio en el hogar; III. Acceso a los servicios de salud; IV. Acceso a la seguridad social; V. Calidad y espacios de la vivienda; VI. Acceso a los servicios básicos en la vivienda; VII. Acceso a la

alimentación; y VIII. Grado de cohesión social. De estos indicadores, los siete primeros se refieren a atributos del individuo / hogar, pero el octavo es un atributo de la sociedad, lo cual ha sido un problema para la práctica oficial de medición, como en el caso de Colombia a seguir.

En Colombia se sigue la metodología de ALKIRE & FOSTER (2007) para la identificación y agregación de la población en condición de pobreza. Las dimensiones e indicadores son: I. EDUCACIÓN: i) años de escolaridad; ii) asistencia escolar para niños entre 1 y 8 años; II. SALUD: i) mortalidad infantil; ii) nutrición; III. ESTANDAR DE VIDA: i) electricidad; ii) agua para consumo; iii) saneamiento; iv) material de pisos; v) combustible para cocinar; vi) posesión de activos.

Brasil definió las siguientes dimensiones e indicadores: I. VULNERABILIDAD: recursos para satisfacer necesidades básicas; II. ACCESO AL CONOCIMIENTO: i) alfabetismo; ii) escolaridad formal; iii) calificación profesional; III. ACCESO AL TRABAJO: i) disponibilidad de trabajo; ii) calidad; iii) productividad de los puestos de trabajo disponibles; IV. ESCASEZ DE RECURSOS: i) extrema pobreza; ii) pobreza; iii) capacidad de generación de recursos; V. DESARROLLO INFANTIL: i) trabajo infantil; ii) acceso a la escuela; iii) progreso escolar; V. CARENCIAS HABITACIONALES: i) propiedad del inmueble; ii) déficit habitacional; iii) capacidad de abrigar; iv) acceso inadecuado al agua; v) desagote sanitario inadecuado; vi) sin recolección de basura domiciliaria; vii) sin acceso a la electricidad; viii) sin acceso a bienes durables (PAES DE BARROS *et al.*, 2006).

Por otro lado, la Oficina de Estadística Europea (EUROSTAT), más concretamente el Subgrupo de Indicadores del Comité de Protección Social, está desarrollando una metodología para la construcción de indicadores no monetarios de privación. El objetivo es que estos indicadores aporten información complementaria a la que ya proporcionan el resto de los indicadores de inclusión social, aunque sin la ambición de cubrir todas las áreas existentes de la exclusión social. Toman una serie de elementos, bienes, servicios o condiciones materiales de vida como indicadores básicos que se utilizan para detectar la privación material, los cuales se agrupan de la siguiente manera: I. DIFICULTADES ECONÓMICAS: i) poder pagar vacaciones fuera de casa, al menos una semana al año; ii) retrasos en los pagos de la hipoteca, el alquiler, recibos de agua, electricidad, etc., compras a crédito; iii) poder pagar una

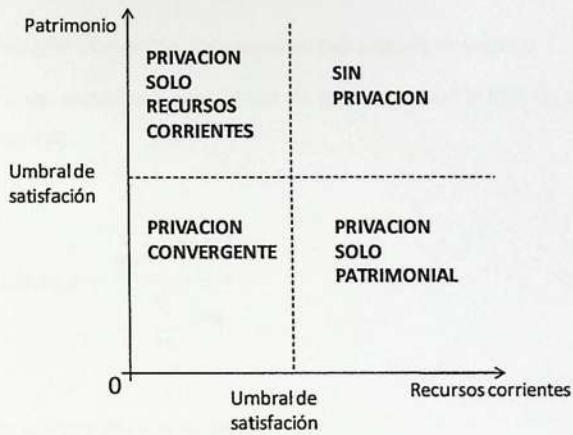
comida de carne, pollo o pescado (o equivalente vegetariano), al menos cada dos días; iv) poder permitirse mantener la vivienda a una temperatura adecuada durante los meses fríos; II. BIENES DURADEROS: i) se considerará que un hogar está privado del bien si la carencia se debe a falta de recursos económicos (TV en color, teléfono, automóvil para uso personal); ii) vivienda (existencia de goteras, humedades en paredes, techos o cimientos, o podredumbre en pisos, marcos de ventanas o puertas, escasez de luz natural en alguna habitación, baño o ducha en la vivienda, inodoro con agua corriente en el interior de la vivienda para uso exclusivo del hogar).

La preocupación por la medición de la pobreza en Argentina, así como en otros de América Latina, tuvo un fuerte impulso desde mediados de la década de 1980. En la década de 1990 se desarrollaron nuevas perspectivas y metodologías, lo que implicó revisar las ventajas y limitaciones de los métodos utilizados hasta el momento, sobre todo el de las NBI. Muchos investigadores señalaron que resultaba importante contar con métodos que utilizaran los datos censales debido a que consideraban al total de la población relevada en los distintos lugares del país. Por este motivo, al realizarse el censo del año 2001, además de continuar con la información de NBI, el Instituto Nacional de Estadística y Censos (INDEC) elaboró una nueva metodología para brindar información sobre los niveles de privación de los hogares. Este método, denominado Índice de Privación Material de los Hogares (IPMH), considera dos dimensiones: I. CONDICIONES PATRIMONIALES, caracterizadas por el acceso a una vivienda confortable (materiales adecuados, existencia de inodoro con descarga de agua); II. LA DISPONIBILIDAD DE RECURSOS EN DINERO, llamados recursos corrientes. Sin embargo, como los censos de población no registran información sobre los niveles de ingreso de las personas, se definió un indicador que permite aproximarse a esos valores de manera indirecta, es decir, utilizando datos relevados por el censo. El indicador, denominado capacidad económica del hogar, permite distinguir niveles entre los hogares a partir del análisis de los años de educación que poseen todos los integrantes del hogar que tienen ingresos y de la cantidad de personas que dependen de ellos. La combinación de esas dimensiones permite distinguir: -hogares que tienen privaciones de tipo estructural (patrimonial, vinculada a la vivienda); -hogares con dificultades o privaciones para obtener ingresos monetarios; - hogares con ambos tipos de privación; - hogares sin privaciones.

1.4.1 Cálculo del Índice de Privación Material de los Hogares

El Índice de Privación Material de los Hogares (IPMH) se construye a partir de la combinación de dos indicadores: el de condiciones habitacionales (CONDHAB) y el de capacidad económica de los hogares (CAPECO). En función a la valoración que se le asigne a cada uno de estos indicadores en los hogares, se determinará si el hogar tiene o no privación material, y en caso afirmativo, el tipo de privación que presenta.

Figura 1.3: Construcción del Índice de Privación Material de los Hogares



Fuente: INDEC

Cada una de sus dimensiones se plantea empíricamente a partir de indicadores contruidos con datos censales que cumplen los siguientes requisitos: son observables en todo el universo de hogares, permiten reconocer diferentes aspectos de la privación y son susceptibles de elaborarse con otras fuentes.

1.4.1.1 Privación Patrimonial: condiciones habitacionales

Para el cálculo de la dimensión "privación patrimonial" se seleccionan un grupo de variables referidas a los materiales constructivos de la vivienda: material de los pisos, techos y presencia de cielorraso. Se adopta el criterio para la aproximación del nivel de calidad de éstos, la durabilidad y aislamiento que proporcionan, entendiendo

que algunos son más aptos que otros y que ofrecen distintas condiciones de habitabilidad.

La segunda dimensión se estima a través del indicador “**Tenencia de instalación de baño con arrastre de agua en el inodoro**”, es la que resume de manera óptima las condiciones sanitarias de los hogares.

El indicador Condiciones Habitacionales (CONDHAB) se construye a través de un criterio de combinación de condiciones suficientes; se considera con privación en la dimensión habitacional a los hogares que carecen de inodoro con descarga de agua o habitan en una vivienda con materiales constructivos insuficientes.

1.4.1.2 Privación Recursos Corrientes: capacidad económica

El indicador de capacidad económica de los hogares (CAPECO) es el resultado de la siguiente fórmula:

$$CAPECO = \frac{\sum_{i=1}^n (CP_i * VAE_i)}{\sum_{i=1}^n Aeq_i} \quad (1.3)$$

donde

n : Total de integrantes del hogar

CP : Condición de percepción (asume distintos valores según la condición de actividad, la edad, el sexo y el lugar de residencia)

VAE : Valor de los años de escolaridad invertidos en el mercado laboral

Aeq : Valor de la unidad de adulto equivalente de cada integrante del hogar (varía de acuerdo con sexo y edad, siguiendo una tabla de necesidades calóricas y nutricionales) (GÓMEZ, *et al.*, 2004).

En la actualidad, la información disponible en las encuestas de hogares para implementar un método multidimensional es superior a la de los censos. No obstante, en muchos casos continúa siendo insuficiente, sobre todo si se tienen en cuenta los cambios en los requerimientos para poder satisfacer necesidades básicas y participar en la sociedad. Por ejemplo, en educación, se cuenta con indicadores de acceso de la población en edad escolar, pero no sobre las competencias de alfabetización de los

adultos; en el ámbito de la vivienda, se dispone prácticamente de las mismas variables y categorías contenidas en los censos de los años ochenta, y en varias encuestas de América Latina no se releva el ámbito de la salud, entre ellas las de Argentina. En los casos en que se dispone de variables útiles, por ejemplo, sobre mortalidad infantil y desnutrición en las encuestas sobre la base de indicadores múltiples y en las de demografía y salud, estas no suelen ir acompañadas de otras variables relevantes (tales como, la situación laboral o los ingresos del hogar), lo que dificulta la aplicación de indicadores desarrollados en un país aplicarlo en otro.

El Índice de Bienestar Multidimensional que se propone estimar en esta tesis está compuesto por algunas dimensiones que no son tenidas en cuenta en el IPMH, como ser la funcionalidad del jefe del hogar y del medio en que se desenvuelve. También la dimensión índice de bienes, compuesto por el estado patrimonial del hogar (tenencia de la vivienda) y posesión de ciertos artículos y servicios, indicativos de confort. Además, se consideran algunas características de la vivienda, que no están incluidas en el IPMH, como su ubicación (cerca de basural, zona inundable, en villa de emergencia) y el hacinamiento.

Capítulo 2

Revisión de la literatura sobre el análisis de la pobreza

2.1 Introducción

Tomando como referencia el primer capítulo, donde se describen las diferentes acepciones de la pobreza, sus enfoques, metodologías y las causas que explican este fenómeno, es oportuno realizar una revisión de los diversos estudios más importantes de pobreza unidimensional y multidimensional, publicados hasta el momento en Europa, Argentina y otros países de América Latina.

Este será el punto de partida para conocer la evolución de las investigaciones llevadas a cabo desde la década de los ochenta¹, época en que se realizaron las primeras estimaciones oficiales sobre pobreza en Argentina, con la finalidad de examinar cuáles han sido las tendencias fundamentales y sus resultados.

Recién en ese momento, el Organismo de Estadísticas Oficiales (INDEC) manifestó interés en medir la pobreza, a pesar de que el incremento de esta realidad se venía manifestando desde aproximadamente quince años antes², aparte de que, las primeras propuestas metodológicas para cuantificarla fueron realizadas a fines del siglo XIX y principios del siglo XX (ATKINSON, 1987). La bibliografía especializada coincide con el diagnóstico realizado por el Instituto Nacional de Estadísticas y Censos (INDEC, 1984) en el sentido que, hasta comienzos de la década del setenta, la

¹ Vale aclarar que existen ejercicios de estimación previos - como por ejemplo los cálculos realizados para la CEPAL por Oscar Altimir a fines de los setenta, o aquellos llevados a cabo por Beccaria y Minujín a mediados de la década siguiente- que no son considerados al realizar esta afirmación, debido a que no formaban parte del sistema estadístico nacional. Sin embargo, estos ensayos no pueden dejar de ser mencionados, ya que los mismos fueron referencias ineludibles para el diseño de la metodología oficial de estimación de la pobreza en nuestro país.

² Se considera que la pobreza ya había alcanzado niveles preocupantes antes de la existencia de información oficial, cualquiera sea la metodología aplicada. A su vez, aún en ausencia de datos contruidos con el propósito de informar respecto a este fenómeno, es posible identificar situaciones de pobreza en nuestro país a fines de la década de los sesenta y comienzos de los setenta (BECCARIA & GROISMAN, 2007).

pobreza existía, pero como un fenómeno marginal o de magnitud relativa moderada. Sin embargo, luego de las transformaciones económicas introducidas a mediados de esa misma década, la problemática comienza a expandirse y profundizarse, obligando a los diversos organismos oficiales a abordar la construcción sistemática de información oficial que evidencien el alcance, la localización y la magnitud de las diversas situaciones de pobreza.

A nivel global, existen tres metodologías tradiciones para medir pobreza: los estudios de **pobreza objetiva, relativa y subjetiva**. En la primera, se suelen registrar las estadísticas oficiales del mundo en desarrollo. Cuenta con un gran abanico de técnicas. Entre las más populares se encuentran la Canasta de Satisfacción de Necesidades Básicas, el Índice de Necesidades Básicas Insatisfechas, el Método Integrado, y el dólar Paridades de Poder Adquisitivo (PPA) por persona al día.

La medición de pobreza relativa se usa frecuentemente en los países europeos y en la Organización para la Cooperación y el Desarrollo Económicos (OCDE). También existen varias alternativas, aunque muy similares en su construcción: la Línea de Pobreza (LP) del 60% de la mediana de ingresos (o del 50%, 30%, etc.) y la LP del 40% del promedio de ingresos. Por último, están los estudios de pobreza subjetivos, que también construyen LP, pero basadas en las opiniones y percepciones de las personas sobre el ingreso mínimo indispensable para la supervivencia. Estos últimos no han sido considerados aún como parte de los métodos oficiales de medición.

Algo común en los países de América Latina y en los organismos internacionales es que utilizan parámetros muy bajos. Eso lleva a considerar como "no pobres" a personas que en realidad están en una situación de vulnerabilidad. La mayoría de los organismos internacionales y de los países latinoamericanos se inclinan por los métodos objetivos. El Banco Mundial (BM) usa una medición con una LP estimada como el promedio de las LP de los países más pobres del mundo; así obtiene la LP de 1,25 dólares PPA.

La Cepal usa una LP que es igual al costo de una canasta de alimentos crudos multiplicada por 2 en el contexto urbano y por 1.75 en el rural. A diferencia de los anteriores, considera los requerimientos nutricionales de las personas según sexo y edad y una relación empírica observada entre el gasto en alimentos y el gasto en otros

rubros. Tiene un fundamento en alimentación, pero en los otros rubros puede generarse cierta confusión.

La importancia de las mediciones multidimensionales se basa en el cruce de los criterios objetivos con los relativos, y, sobre todo, que incluyan las variables subjetivas. Si bien éstas son las más difíciles de calcular, son imprescindibles para comprender el fenómeno en toda su expresión. Sólo así se podría actuar efectivamente desde el Estado.

Existe un consenso creciente respecto de la necesidad de que los países dispongan de mediciones multidimensionales de la pobreza, a fin de contar con información complementaria a la obtenida mediante los métodos convencionales, basados en los ingresos monetarios. En ello han influido distintos factores, como el surgimiento y predominio de nuevos enfoques sobre el desarrollo social y el bienestar, la constatación de las limitaciones de los indicadores monetarios como aproximaciones al estándar de vida, la necesidad de una mayor alineación entre los indicadores de pobreza y las políticas tendientes a superarla.

Este proceso se ha visto impulsado, en Europa, América Latina y otras latitudes, por el predominio de nuevos marcos conceptuales sobre el desarrollo social y el bienestar, como el enfoque de derechos y las perspectivas de capacidades, que van más allá de la noción de pobreza limitada a la insatisfacción de necesidades básicas.

2.2 Las mediciones de pobreza en Argentina.

Principales estudios

Las primeras estimaciones oficiales de la pobreza se realizaron para todo el país a mediados de los ochenta, a través de la aplicación del Método de las Necesidades Básicas Insatisfechas (NBI), con los datos del Censo Nacional de Población y Vivienda (CNPV) de 1980 (INDEC, 1984); y hacia fines de esa misma década, en el marco de la Investigación de la Pobreza en la Argentina (IPA), se establecieron las bases para el cálculo oficial mediante el Método de la Línea de Pobreza (LP). No obstante, la metodología definitiva para la estimación de la pobreza por ingresos comenzó a emplearse en forma metódica con información de la Encuesta Permanente de Hogares

(EPH), y sólo para el Gran Buenos Aires (GBA)³, recién a principios de los años noventa (ARAKAKI, 2011).

A partir de ese momento, la incidencia de la pobreza fue aumentando al punto de alcanzar niveles alarmantes. En este contexto, no se consideró la existencia de cambios en la composición del universo de hogares en situación de pobreza. Por tal motivo, las instituciones como las universidades se propusieron llevar a cabo proyectos de investigación enmarcados en el tema.

El Departamento de Investigación Institucional, Programa Observatorio de la Deuda Social Argentina de la Universidad Católica Argentina, es un centro de investigación, extensión y formación de recursos humanos, creado en el año 2002. Sus funciones se enmarcan en un programa que reúne proyectos de investigación patrocinados por entidades que promueven el desarrollo de la ciencia y la tecnología, de instituciones internacionales y de organizaciones no gubernamentales. Posee un equipo multidisciplinario integrado por investigadores y becarios (tanto de la propia Universidad Católica Argentina-UCA como del Consejo Nacional de Investigaciones Científicas y Técnicas-CONICET) abocados al estudio integral de la realidad nacional argentina dimensionada en términos de logros y déficit de desarrollo.

Desde 2004, el Programa realiza anualmente una encuesta nacional con indicadores de desarrollo humano e integración social (Encuesta de la Deuda Social Argentina, EDSA) a hogares, poblaciones y niños residentes de áreas urbanas del país. Actualmente, en el marco del proyecto "La Argentina del Bicentenario 2010-2016", la encuesta releva 5700 hogares. A partir de estos relevamientos, el Observatorio de la Deuda Social dispone de una base de datos primarios con indicadores sociales auténticos, metodológicamente confiables y con extensa aplicación y reconocimiento académico y político-institucional.

De estas investigaciones surgen permanentemente, diversos documentos que conforman una serie de estudios, destinados a monitorear las vulneradas capacidades de desarrollo humano que permanecen vigentes en la actual etapa político-institucional y económica argentina. Los mismos se concentran en el análisis de las privaciones en las capacidades elementales y de subsistencia para la vida humana desde un enfoque

³ Los datos referidos al resto de los aglomerados urbanos comenzaron a producirse a partir del 2002, a través de la aplicación de una "metodología de transición" (INDEC, 2002).

dinámico y multidimensional. Para ello, se parte del reconocimiento conceptual y empírico de cuatro dimensiones de estudio principales: (a) estar bien alimentado y no padecer hambre; (b) gozar de buena salud y estar protegido de enfermedades; (c) disponer de un hábitat doméstico adecuado y (d) disponer de medios de vida suficientes. Entre los documentos relacionados a esta temática se pueden mencionar los siguientes:

1. (LÉPORE & SALVIA, 2005): *"La naturaleza multidimensional de la pobreza. Algunas aproximaciones empíricas al déficit de capacidades de subsistencia"*

La fuente de datos utilizada para elaborar este informe es una encuesta multipropósito con diseño en panel focalizada territorialmente en espacios residenciales que aglomeran sectores con mayor vulnerabilidad social. La misma se aplica con frecuencia semestral sobre una muestra aleatoria de 1.100 casos representativa de la población mayor de 18 años residente en importantes centros urbanos del país (Área Metropolitana de Buenos Aires, el Gran Córdoba, el Gran Salta, el Gran Resistencia, el Gran Mendoza, Bahía Blanca, en la Provincia de Buenos Aires y Neuquén-Plotier, en provincia de Neuquén). El diseño muestral de la EDSA contiene un sistema de rotación que permite el seguimiento de los casos en el tiempo, lo que hace posible la confección de paneles para análisis de datos longitudinales. Para la elaboración de la muestra se aplicó un procedimiento de estratificación y conglomeración de las unidades muestrales con similar perfil socioeducativo, mediante el promedio del nivel de educación de los jefes de hogar residentes en los radios censales de los aglomerados relevados. La distribución de la muestra en cada una de las áreas urbanas representadas permitió clasificar a las unidades censales en cuatro tipos diferenciados de "espacios residenciales socioeducativos": muy bajo, bajo, medio bajo y medio alto. Con base en la hipótesis de la segregación espacial, los tres primeros fueron reagrupados en la categoría "espacios residenciales de vulnerabilidad", en tanto que el cuarto se empleó con fines de control y comparación⁴.

⁴ Para mayor especificación sobre la metodología y el diseño muestral empleados por la EDSA, puede consultarse el Informe Técnico en TAMI & SALVIA, (2004) o en el primer documento de la Serie Monitoreo de la Deuda Social /Documento 1 / 2005 (SALVIA, 2005).

La estructura de análisis consistió en presentar, para cada una de las dimensiones identificadas, un conjunto de indicadores asociados a umbrales mínimos de satisfacción, los que fueron evaluados en términos de la desigual incidencia mostrada según la localización residencial socioeducativa. El resultado verificado en los espacios de control, característicos de clases medias integradas, actuó como parámetro de comparación socioculturalmente instituido y económicamente viable.

También elaboraron un Índice Integrado de Capacidades de Subsistencia (IICS), a partir de una selección de indicadores claves correspondientes a cada una de las dimensiones de realización identificadas.

La conclusión de este informe fue que la segregación residencial tenía un efecto neto en la determinación de las probabilidades de experimentar privaciones de capacidades de subsistencia. La probabilidad de los hogares de sufrir déficit de subsistencia es mayor en aquellos establecidos en un espacio residencial de vulnerabilidad, independientemente de otros factores.

2. (SALVIA, COMAS, & STEFANI, 2010)⁵: *"Heterogeneidad estructural y acceso diferencial a empleos de calidad en dos momentos de crecimiento económico 1998 – 2006"*⁶

En esta publicación se abordó el impacto de la heterogeneidad estructural sobre la segmentación del mercado de trabajo en el ámbito urbano de la Argentina, entre el último momento de crecimiento del período de reformas estructurales y apertura externa de los años noventa y el período expansivo iniciado en 2003. Los investigadores partieron de la hipótesis de que la heterogeneidad estructural y la formación de excedentes relativos de fuerza de trabajo, generada por los procesos de reforma y apertura económica, persistían en las condiciones de funcionamiento económico-social del momento en que se realizó el trabajo.

⁵ Agustín Salvia: Doctor en Ciencias Sociales, Investigador del CONICET. Director del Programa CEyDS. Instituto de Investigaciones Gino Germani; Guillermina Comas: Becaria UBACYT del programa CEyDS. Federico Stefani: Becario CONICET del programa CEyDS.

⁶ Este trabajo pertenece al Equipo Cambio Estructural y Desigualdad Social, dirigido por Agustín Salvia, con sede en el Instituto de Investigaciones Gino Germani de la Facultad de Ciencias Sociales- UBA/ Buenos Aires.

La metodología utilizada, se basó en análisis estadísticos de estática comparada a partir de los microdatos de la Encuesta Permanente de Hogares EPH-INDEC, correspondientes a los relevamientos de octubre de los años 1998 y 2001, bajo la modalidad EPH-Puntual, y a los relevamientos de los segundos semestres de los años 2003-2006, bajo la metodología de la EPH continua. Dadas las diferencias metodológicas que presentaban ambos tipos de encuestas, se emplearon procedimientos de ajuste sobre los datos estimados.

Se contrastaron los cambios en la estructura económico-ocupacional, centrando el análisis tanto en la calidad de las relaciones laborales, la composición sectorial de la fuerza de trabajo, así como, en la relación entre ambas dimensiones. También se elaboraron dos Modelos de Regresión Logística con el fin de determinar la incidencia del sector de inserción y de la categoría ocupacional sobre las probabilidades diferenciales de acceso a empleos de calidad. La base del estudio fue el carácter estructural de la segmentación laboral y de la existencia de fuerza de trabajo no integrada al sistema.

Los investigadores llegaron a la conclusión que los períodos de crecimiento, bajo una estructura heterogénea y dividida, aparentemente no pudo generar procesos integradores. En este contexto, se destacó una vez más la vigencia de un comportamiento relativamente "procíclico" del sector informal, pero sobre todo del segmento marginal del empleo. Este último no pareció haber sufrido cambio alguno frente a las reglas macroeconómicas e institucionales del momento.

Por otro lado, sin dejar de reconocer las mejoras que experimentaron los indicadores económicos y ocupacionales hasta el 2006, las mismas tuvieron un impacto diferencial, resultando más favorecidas las inserciones en el sector formal. Si bien es importante reconocer que el período económico evaluado resulta breve para esgrimir argumentos concluyentes, los autores consideraron que las tendencias mostraron una escasa capacidad de reversión de los problemas de empleo, pobreza y desigualdad. En este sentido, la profundización de las desigualdades en la estructura social del trabajo sugirió la necesidad de considerar más la parte estructural y menos vinculada al contexto económico e institucional, como clave explicativa de los procesos que reproducen la heterogeneidad y segmentación laboral.

3. (TUÑÓN *et al.*, 2012)⁷: “Los indocumentados en Argentina. La cara invisible de la pobreza”

Cuando se habla sobre población indocumentada, habitualmente se relaciona las migraciones internacionales informales y la condición de aquellos que se encuentran viviendo en el territorio nacional sin una identidad certificada en Argentina. Sin embargo, se trata de personas que, en su lugar de origen, cuentan con una documentación que lo acredita, lo que disminuye la gravedad del problema. Sin embargo, existe otro sector que nació en territorio argentino y no obtuvieron su Documento Nacional de Identidad por parte del Registro Nacional de las Personas (Ministerio del Interior) ya que nunca fueron registradas o inscriptas al nacer en los Registros Civiles Provinciales correspondientes, por lo que carecen de todo tipo de identidad legal.

La ausencia de un Documento Nacional de Identidad (DNI) vulnera el derecho a la identidad y en consecuencia otros derechos también se ven afectados. Los niños y niñas que no fueron registrados no existen para el Estado y, por lo tanto, no pueden ejercer sus derechos de una manera plena. La ausencia de un documento de identidad implica ser “invisible” para los registros estatales, situación que genera exclusión, profundizando y perpetuando las condiciones de pobreza y precariedad.

Las estimaciones presentadas en el trabajo que se cita muestran que en 2011 en las zonas urbanas de Argentina el 1,6 % de los niños y niñas entre 0 y 17 años nacidos en el país no tenía DNI (lo que equivale a aproximadamente 168.000 chicos)⁸. Como primera aproximación a la problemática se analizaron distintos factores sociodemográficos de los niños/as y adolescentes argentinos sin documento como son la edad y el sexo. Al respecto, si bien no se observaron

⁷ El presente informe fue realizado en el marco de los estudios Programa del Observatorio de la Deuda Social Argentina (ODSA) de la Universidad Católica Argentina (UCA), en alianza con el Instituto Abierto para el Desarrollo y Estudio de Políticas Públicas (IADEPP). La coordinación institucional estuvo a cargo de Alicia Casermeiro Pereson y Agustín Salvia por el ODSA, y Jorge Álvarez por el IADEPP. El estudio fue desarrollado bajo la dirección de Ianiña Tuñón y fue elaborado por Helga Fourcade, María Sol González y Natalia Reggini; y la colaboración de Jorge Álvarez, Luis Freitas, Axel Cantlon, y Carlos Celaya.

⁸ Las estimaciones fueron realizadas en base a los microdatos de la Encuesta de la Deuda Social Argentina. Encuesta realizada en el marco del Programa del Observatorio de la Deuda Social Argentina de la UCA, y que cuenta con la colaboración de las Fundaciones Arcor y Telefónica. Ver detalles metodológicos de la EDSA en www.uca.edu.ar/observatorio.

diferencias estadísticamente significativas por sexo, sí se advirtieron desigualdades en el interior de cada grupo de edad. En este sentido, la situación más crítica la registraban los niños/as argentinos entre 0 y 4 años, en tanto 2,3% no tenía documento; mientras que dicha situación afectaba al 1% promedio de los niños/as entre 5 y 17.

Se observó, además, una estrecha relación en términos del espacio geográfico. Al respecto se advirtió una fuerte correlación con las características del espacio de residencia, en tanto los niños/as en villas o asentamientos registraban mayor probabilidad de no tener DNI que sus pares en áreas de urbanización formal media. A medida que empeoraban las condiciones del espacio socio-residencial se incrementaba la propensión a la informalidad del derecho a la identidad.

Por otro lado, se analizó la situación de los niños/as indocumentados argentinos con relación a los derechos esenciales de tener una vida saludable, de participar de procesos de formación en el campo educativo y no trabajar.

Con respecto al derecho a la salud, tuvieron en cuenta dos indicadores objetivos, la proporción de niños/as y adolescentes menores de 18 años que no tenían cobertura de salud y que, además, no realizaron consultas médicas desde más de un año antes e incluso nunca hicieron un control. En cuanto al derecho a la educación, se evaluó la propensión a no asistir a la escuela en las edades correspondientes al ciclo inicial obligatorio (sala de 5 años) y hasta la finalización de la escuela secundaria, es decir la escolarización en el ciclo educativo obligatorio. Por último, se analizó la propensión al trabajo en actividades domésticas o en actividades no domésticas en el mercado.

El trabajo infantil, en cualquiera de sus formas, se opone al ejercicio de otros derechos y expone al niño/a a múltiples riesgos, tales como accidentes, lesiones, maltrato, abuso, entre tantos otros, lo cual incrementa aún más su situación de vulnerabilidad.

Las causas de ser indocumentado son diversas, las trabas burocráticas de los procedimientos de registro tardío y extemporáneo, la distancia geográfica y sobre todo las condiciones de pobreza. Estas dificultades originan la primera situación de vulnerabilidad que es acompañada por una serie de privaciones

concatenadas, vinculadas al no cumplimiento del derecho a la identidad con el que debería contar todo niño y niña que nace en Argentina.

El impacto negativo se advirtió en el acceso de las personas a una educación completa, al sistema de salud y a las diferentes líneas de seguridad social que brinda el Estado para la población en situación de vulnerabilidad o pobreza. Asimismo, afecta la capacidad de representación de las personas ante las autoridades en general y el sistema de justicia, interfiere en su posibilidad de adquirir créditos o acceder a una cuenta bancaria; lo imposibilita a insertarse en el mercado laboral como empleado formal y dificulta el libre movimiento de las personas, tanto dentro como fuera del país.

4. (ARAKAKI, 2011): *"La pobreza en Argentina 1974-2006. Construcción y análisis de la información"*

Algunas universidades de gestión pública también realizaron proyectos de investigación abocados a la problemática de la pobreza. Es así como en el marco de la beca UBACyT - Categoría Estímulo⁹, se inició una importante investigación, cuyo objetivo principal era "contribuir a alcanzar una visión más acabada del fenómeno de la pobreza en la Argentina"¹⁰. Con este fin, se propuso elaborar series estadísticas que permitieran estudiar con mayor profundidad la evolución de esta pobreza, para lo cual se tomaron dos vías distintas. Por un lado, sobre la base de los criterios de clasificación empleados en nuestro país, Necesidades Básicas Insatisfechas (NBI) y Línea de Pobreza (LP) se calculó una serie de indicadores complementarios a los utilizados en forma oficial. Por otro lado, se empleó un criterio de clasificación denominado método combinado o bidimensional, el cual surge de utilizar en forma conjunta los criterios de NBI y LP.

⁹ Beca UBACyT - Categoría Estímulo otorgada por la Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires para la realización del proyecto "Medida Alternativa de la Pobreza. Aplicación al Caso Argentino. 1991 - 2006", como parte del Proyecto UBACyT E-003, "Crisis socioeconómica y perspectivas del empleo en la Argentina actual".

¹⁰ Pese a que en el plan de trabajo original proponía abordar el estudio de la pobreza en todo el país, el análisis se limitó al caso del GBA debido a las dificultades vinculadas a la fuente de información.

En ese trabajo se presentaron ciertos inconvenientes como las restricciones metodológicas, ya que, los procedimientos empleados en forma oficial en Argentina utilizaban datos provenientes de diferentes fuentes de información -lo que llevaba a diferencias en los instrumentos de captación, en la frecuencia de publicación de los datos, en el período de vigencia de cada metodología, etc.- no sólo obstaculizaba la comparación de los resultados obtenidos de los mismos, sino también la complementariedad entre ellos, posibilidad que contribuiría en forma significativa al estudio de la pobreza.

La Encuesta Permanente de Hogares comenzó a realizarse, en su versión Puntual, en 1973 y mantuvo esta dinámica de funcionamiento, realizando dos relevamientos anuales, durante treinta años. Sin embargo, al momento de esta investigación, sólo algunas de las bases correspondientes al período 1973-1987 se encontraban disponibles, lo cual limitó los alcances temporales de la misma.

Posteriormente, en el año 2003, el INDEC realizó una reformulación integral de la EPH, que incluyó modificaciones referidas a los cuestionarios, la forma de relevamiento, etc. (INDEC, 2003a). Con el propósito de analizar el impacto sobre las estimaciones y la comparabilidad entre las mismas, se clasificaron los cambios en dos tipos. Por un lado, se encuentran aquellos que afectan en forma más directa a los indicadores, entre los que se destaca la eliminación de algunas variables necesarias, pero esto impide la construcción de series de largo plazo en algunos casos. Por otro lado, se hallan las modificaciones aplicadas en el modo de captación de la información -más específicamente, en los cuestionarios y la frecuencia de los relevamientos que afectan a la homogeneidad de las series construidas en base a datos captados bajo la modalidad Puntual y Continua. En consecuencia, para obtener series de largo plazo (desde 1974 hasta el 2006) comparables, era necesario empalmar los resultados que surgen de las dos versiones de la EPH. Sin embargo, resultó imposible debido a que la base correspondiente al primer semestre de 2003 (el único punto de superposición entre ambos relevamientos) no contaba con las variables necesarias para empalmar ambas series¹¹. Por lo tanto, estos dos

¹¹ Luego de la intervención del INDEC iniciada en 2007 se publicaron bases modificadas, las cuales incluían una nueva versión de aquellas difundidas bajo la modalidad Continua (esto es, 2003-2007). Este cambio permitió realizar estimaciones de pobreza por NBI que contemplaron

inconvenientes no dieron otra posibilidad que presentar (en cada caso) dos series (1974-2003 y 2003-2006) que, si bien no eran estrictamente comparables, permitieron obtener algunas conclusiones respecto a los cambios de tendencias entre uno y otro período.

Se analizó la evolución de la pobreza en el GBA con diferentes propuestas metodológicas: se calculó el índice de recuento tanto para la condición de NBI como para cada componente. También, en el marco del enfoque indirecto, se construyeron cinco indicadores distintos (índice de recuento, índice de intensidad estandarizada, brecha de pobreza, índice de Sen y el índice FGT2). Finalmente, se realizó la estimación de los distintos universos definidos por el método combinado o bidimensional.

De esta forma, si bien el número total de hogares que presentaba algún tipo de inconveniente para satisfacer sus necesidades se encontraba en niveles similares a los del comienzo del período analizado, las características de la pobreza parecían haber cambiado considerablemente. Así, mientras al inicio la forma predominante de la pobreza era la falta de acceso a bienes y servicios básicos, al momento de la investigación, se caracterizaba por la insuficiencia de ingresos.

5. (CONCONI, 2011): *"Pobreza multidimensional en Argentina: Ampliando las medidas tradicionales de pobreza por ingreso y NBI"*¹²

Este trabajo se divide en una parte introductoria que consiste en la caracterización de la población argentina y un análisis de la pobreza,

todas las variables empleadas en el período 1974-2003; pero resultó imposible replicar los cálculos de pobreza por ingresos realizados por el organismo, ya que la información referida a las canastas no se encontró disponible en las bases y que estas últimas no se presentan con la frecuencia requerida. A su vez, tampoco se han publicado bases que puedan ser utilizadas como punto de empalme con la serie puntual. Todo esto, sumado al hecho de que las bases corregidas se encuentran "sospechadas" de manipulación, no se las consideró en el trabajo en cuestión.

¹² Tesis de Maestría en Economía de la Universidad Nacional de La Plata de la autora en 2009. Publicado como Documento de Trabajo N° 90 de la Serie Documentos de Trabajo del Departamento de Economía de la Facultad de Ciencias Económicas de la UNLP en 2001 (ISBN 1853-3930).

considerando estimaciones por ingresos y por NBI¹³. En la segunda parte, se analizaron medidas de pobreza multidimensional en base a las metodologías de BOURGUIGNON y CHAKRAVARTY, (2003) y ALKIRE y FOSTER, (2008).

En el enfoque por NBI se consideró la posibilidad de construir indicadores que tenían en cuenta otro tipo de carencias. En consecuencia, el indicador $NBI = 1$ cuando en el hogar se verifica al menos una de las siguientes condiciones: hacinamiento, vivienda precaria, condiciones sanitarias deficientes, falta de acceso al agua potable, escolaridad deficiente de los niños del hogar, escolaridad deficiente del jefe de hogar y capacidad de subsistencia limitada.

Se observó que esta medida de pobreza generaba una proporción de pobres superior que la monetaria para cualquier línea de pobreza. De acuerdo con este enfoque, cerca del 37% de la población argentina tenía NBI de algún tipo. Sin embargo, y a diferencia de la pobreza monetaria, esta medida era más estable, dado que entre 1992 y 2003 no se observaban diferencias significativas, con un porcentaje de pobres cercano al 40%.

En cuanto a considerar una medición multidimensional de la pobreza, la autora aplicó una metodología reductora de datos (Análisis Factorial por Componentes Principales) para evaluar cuáles debían ser las dimensiones por incluir en el análisis. Esta técnica consideraba que el ingreso, empleo, educación y demás variables seleccionadas explicaban el comportamiento de una variable latente (en este caso, se consideraba que era una noción de "bienestar" o "pobreza ampliada").

De acuerdo con la fuente de datos disponible, los elementos que pudieron estimarse en este trabajo fueron los siguientes: **Dimensión vivienda** ("*habitar en una vivienda adecuada*"); **Dimensión ingresos** ("*poseer un ingreso adecuado*"); **Dimensión laboral** ("*ser capaz de obtener un trabajo*"); **Dimensión educación** ("*obtener educación básica*"); **Dimensión ambiente** ("*vivir en un ambiente limpio y seguro*").

La conclusión preliminar que se extrajo de los resultados de las estimaciones fue que el ingreso por sí mismo resultaba insuficiente para explicar una noción de

¹³ La fuente de datos fue la EPH puntual de 1992 a 2002 y la continua de 2003 a 2006.

bienestar (o de pobreza ampliada). De allí que por lo menos dos dimensiones eran relevantes y debían incluirse en el análisis: *características de la vivienda y situación laboral*. Por otro lado, tampoco era adecuado considerar pobreza sólo por NBI, ya que el segundo factor (relacionado con las características que suelen incluirse las mediciones de NBI) explicaba sólo cerca del 20% de la variabilidad total de la variable latente.

Como una segunda conclusión, la autora puso en evidencia que las dimensiones de ingreso, vivienda y empleo eran necesarias para analizar adecuadamente la pobreza, independientemente de la etapa del ciclo económico en que se encontraba el país. Los resultados se mantuvieron tanto en los años de estabilidad, como en los de recesión, crisis y recuperación. Por lo tanto, una política social que busca atacar correctamente el problema de la pobreza debería considerar conjuntamente estas tres dimensiones, y no en forma aislada.

La pobreza multidimensional por este enfoque presentaba la misma tendencia que la explicada para las estimaciones con la medida anterior: se redujo en 1998, creció en 2002 ante la fuerte crisis económica y volvió a caer en 2006. Sin embargo, se encontró que la tasa de pobreza en este caso era inferior a la anterior. Lo cual se explica porque la metodología de ALKIRE y FOSTER (2008) realiza un ajuste al recuento de pobreza, mediante el cual multiplican este indicador por la proporción de privaciones que sufren los pobres en promedio.

6. (PAZ, 2014): "*Pobreza multidimensional en la Argentina: Asimetrías regionales (Parte I)*"¹⁴

En este documento se presentaron los resultados de una primera Medición de la Pobreza Multidimensional (PMD) en la Argentina para Unidades Espaciales Menores (UEM), tales como ciudades grandes y centros urbanos de menor magnitud demográfica. Para ello se usó la Encuesta Anual de Hogares Urbanos que releva información de áreas urbanas de 2000 habitantes y más.

¹⁴ Documento de trabajo N° 11. INSTITUTO DE ESTUDIOS LABORALES Y DEL DESARROLLO ECONÓMICO (ielde). Facultad de Ciencias Económicas, Jurídicas y Sociales. Universidad Nacional de Salta (UNSa). Salta. Argentina. ISSN 1852-1118 (impreso), ISSN 1852-1223 (en línea).

Se entiende por PMD las privaciones que experimentan las personas en dimensiones diversas: monetaria, vivienda adecuada, saneamiento básico e inclusión social. En este contexto se definieron las UEM como "jurisdicciones", según la denominación que usa la Encuesta Anual de Hogares Urbanos (EAHU), para referirse a las ciudades grandes, intermedias y más chicas, desde el punto de vista demográfico. Las diferencias entre jurisdicciones derivadas de ese cálculo se denominaron "asimetrías regionales".

A los tradicionales indicadores usados en la medición de la pobreza en América Latina y en la Argentina en particular (necesidades básicas insatisfechas e ingresos insuficientes), se agregaron otros relacionados con la exclusión social o lo que se ha denominado "frontera de exclusión/inclusión social".

Con el fin de plantear este último tema se elaboró un concepto que permite identificar instituciones claves, partiendo del nexo entre la persona y las sociedades que la relacionan con otras personas a lo largo de su vida. Se consideró que un individuo transita por cuatro grandes etapas durante su vida, por lo tanto, se pensó en cuáles son las instituciones que tienden a integrar o incluir (como opuesto a mantenerse al margen, excluido) a esta persona a la vida social. La situación individual frente a estas instituciones definió "la frontera de la exclusión/inclusión social".

A partir de esta idea, se determinaron tres grandes instituciones que se corresponden con las etapas de la vida identificadas: escuela, mercado laboral y seguridad social. Así, una persona se considerará no integrada o excluida si es un niño o adolescente –en sentido genérico– que no asiste a la escuela, o joven que no estudia ni trabaja, o adulto en edad central que no trabaja, o que trabaja en condiciones poco adecuadas; o adulto mayor que no cuenta con cobertura de la seguridad social. Para evaluar el bienestar de la Argentina se consideraron las privaciones en las dimensiones siguientes: ingresos, necesidades básicas, integración social.

Para la evaluación de la pobreza multidimensional (PMD) se tuvieron en cuenta 4 dimensiones de la pobreza, de acuerdo con privaciones en 4 capacidades consideradas básicas: a) capacidad económica; b) vivienda adecuada; c) saneamiento básico; d) inclusión/integración social.

En la dimensión "capacidad económica" se incluyeron tres privaciones: (i) ingresos insuficientes; (ii) capacidad del jefe de generar ingreso; (iii) capacidad del grupo familiar de mantener un ingreso estable y más o menos seguro. Cada una de estas privaciones contó a la vez con uno o más indicadores: ingresos familiares por debajo del mínimo indispensable para cubrir los gastos que demanda la Canasta Básica Total (CBT); nivel educativo del jefe por debajo de los 7 años; y estrategias hogareñas precarias¹⁵.

Entre los resultados más relevantes se destaca, el crecimiento económico con impacto asimétrico, no sólo a nivel de grupos, sino también de Unidades Espaciales Menores (UEM) dentro del país. También se constató que si bien la exclusión laboral es muy elevada (el indicador que más pesa en el total de privaciones y el que arroja tasas de incidencia más altas), resultó muy pareja a nivel de UEM. En contrapartida, se vio que otros indicadores, como por ejemplo la asistencia de niñas y niños a la escuela o el saneamiento, dieron discrepancias que podrían servir como base para reflexionar la política pública a nivel de gobiernos provinciales o municipales.

Un aspecto muy importante que se marcó en este documento es que no siempre las dimensiones más importantes en términos de incidencia, brecha o severidad de la pobreza son las que causan las mayores asimetrías entre UEM. Se advirtió que la dimensión "exclusión/inclusión social" era la de mayor peso en todas las jurisdicciones del país, pero también fue el indicador de menor variabilidad entre ellas.

2.3 Mediciones de pobreza en el mundo

Desde la década de 1980, la CEPAL ha realizado sistemáticamente mediciones de la pobreza por ingresos en América Latina. Las primeras estimaciones se remontan a fines de los años setenta (ALTIMIR, 1979) y a principios de los noventa se calcularon nuevas líneas de pobreza, con base en las encuestas de ingresos y gastos realizadas

¹⁵ Las "estrategias precarias" son aquellas que denotan una situación de debilidad o carencia. Se supone así que si los ingresos corrientes y de fuente genuina no son suficientes y los miembros del hogar deben recurrir a ayudas externas, endeudamiento, o a fuentes que no son sustentables en el tiempo, ese hogar está en una condición desventajosa en términos de bienestar, con lo cual se lo podría considerar "pobre" en esta dimensión.

en los años ochenta (CEPAL, 1991). El contexto actual es diferente al de tres décadas atrás, la mayoría de los países cuentan con mediciones nacionales, generalmente con carácter oficial. Estas mediciones utilizan metodologías y criterios de aplicación que responden a las necesidades y restricciones de cada contexto nacional, pero no permiten generar una visión regional comparable de la pobreza y la desigualdad.

Las mediciones de la pobreza de la CEPAL tienen por objetivo la comparabilidad regional y proveen una mirada complementaria a la que proviene de las mediciones oficiales nacionales. Realiza estimaciones que buscan ser lo más comparables posible entre países y que permiten caracterizar adecuadamente el panorama regional. La búsqueda de la comparabilidad se basa en aplicar un esquema metodológico estandarizado y homogeneizar, en la mayor medida posible, los criterios aplicados. Uno de esos criterios es adecuar las encuestas de hogares, ya que la información con la que se cuenta es insuficiente y poco comparable.

La medición multidimensional para América Latina incorpora elementos provenientes de perspectivas de derechos, capacidades y necesidades básicas insatisfechas. Las metodologías que se vienen desarrollando, en general, presentan ciertas similitudes en la consideración de las dimensiones y la elección de los indicadores usados para su estimación.

En 2008 se promulgó en México la Ley General de Desarrollo Social (LGDS) que hace obligatoria la medición multidimensional de la pobreza (CONEVAL, 2008). En Colombia se sigue la metodología de ALKIRE y FOSTER (2008) para la identificación y agregación de la población en condición de pobreza. Asimismo, Brasil y otros países de Sudamérica fueron incorporando mediciones multidimensionales basados en el Índice de Desarrollo Humano, pero adoptando criterios propios para determinar las dimensiones.

La Oficina de Estadística Europea (EUROSTAT), más concretamente el Subgrupo de Indicadores del Comité de Protección Social, está desarrollando una metodología para la construcción de indicadores no monetarios de privación. El objetivo es que estos indicadores aporten información complementaria a la que ya proporcionan el resto de los indicadores de inclusión social, aunque sin la ambición de cubrir todas las áreas existentes de la exclusión social.

La Iniciativa de Pobreza y Desarrollo Humano de Oxford (OPHI) es un centro de investigación económica dentro del Departamento de Desarrollo Internacional de Oxford en la Universidad de Oxford, Inglaterra. Fue creado en 2007 y su propósito fundamental es construir un marco económico multidimensional para reducir la pobreza basado en las experiencias y valores de las personas. Su directora es Sabina Alkire, autora también de diversos trabajos de investigación sobre la temática de la pobreza.

En julio de 1999 se inauguró en la Universidad de Bristol el Centro Townsend para la Investigación Internacional sobre la Pobreza. Bajo la dirección de David Gordon se dedica a la investigación de la pobreza multidimensional, tanto en los países desarrollados como en desarrollo. Proporciona investigación y datos esenciales a gobiernos, ONG y organismos del sector privado que luchan por reducir la pobreza en el Reino Unido y a nivel internacional.

A continuación, se mencionan algunos trabajos relevantes de estudios de pobreza a nivel internacional.

1. (ALKIRE y FOSTER, 2008) *"Recuento y medición multidimensional de la pobreza"*

Este trabajo propone una nueva metodología para la medición multidimensional de la pobreza que consiste en un método de identificación, que amplía los enfoques tradicionales de intersección y unión, y una clase de mediciones de pobreza que satisface una variedad de propiedades esperadas, incluyendo la descomposición. En la etapa de identificación se usan dos tipos de línea de corte: en primer lugar, una línea de corte dentro de cada dimensión para determinar si una persona u hogar sufre privaciones en esa dimensión; en segundo lugar, una línea de corte entre las dimensiones que identifica a los pobres contando la cantidad de dimensiones en las cuales sufre privaciones. El paso de agregación emplea las mediciones de FGT (Índice de Pobreza Foster-Greer-Thorbecke)¹⁶, ajustadas adecuadamente para dar cuenta de la

¹⁶ El FGT es un índice que mide las carencias en el consumo privado y toma como referencia una determinada línea de pobreza individual, obtenida a partir de un salario mínimo diario, de la población total y de la población económicamente activa. Fue desarrollado por los profesores Erik Thorbecke, Joel Greer y James Foster de la Universidad de Cornell.

multidimensionalidad. El método de identificación es particularmente adecuado para ser utilizado con datos ordinales, al igual que la tasa ajustada de recuento.

Si bien, los autores ponen énfasis en las ventajas del enfoque, también observan que hay otros aspectos que ameritan ser estudiados con mayor profundidad. Inicialmente, como el método de identificación se basa en las líneas de corte, éste es sensible a ciertos cambios, pero insensible a otros. En segundo término, no se destacan las potenciales interrelaciones que pueden existir entre dimensiones cuando las variables son cuantitativas. Evidentemente, el método de identificación toma en cuenta un tipo de vínculo bastante burdo entre las dimensiones, ya que una persona debe sufrir privaciones en un determinado número de dimensiones para ser considerada pobre. Los autores suponen que es necesario enfrentar problemas adicionales cuando se consideran las interrelaciones entre las dimensiones. Tampoco se estableció que las interrelaciones potenciales deban verse reflejadas en una metodología general para evaluar la pobreza multidimensional.

En conclusión, la metodología propuesta en esta investigación ofrece una base neutral sobre la cual se pueden desarrollar versiones con consideraciones más refinadas de la interconexión entre las dimensiones. Como ejemplos ilustrativos de aplicación de la metodología, se utilizaron datos de Indonesia y Estados Unidos.

2. (ALKIRE y ALIM, 2019) *"El estado de la pobreza infantil multidimensional en Asia Meridional: una visión contextual y de género"*

Este artículo propone métodos para evidenciar las desigualdades entre los géneros y dentro de los hogares mediante el análisis a nivel individual y hogar utilizando el mismo conjunto de datos. En particular, se examina: i) la proporción de niños desfavorecidos que viven en hogares multidimensionalmente pobres; ii) de los niños desfavorecidos cuántos son niñas; y iii) la proporción de niños desfavorecidos que viven en hogares en los que no se priva a otros niños en ese mismo indicador. También se observa iv) qué otras privaciones y carencias experimentan los niños, además de las privaciones focales. Por último, se estudia la proporción de personas que viven en hogares donde niños de edades

diferentes sufren simultáneamente dos privaciones diferentes. Además, se realizan análisis más complejos que combinan información sobre la situación de privación de más de un miembro elegible, y niños que completaron seis años de escolaridad, aunque los adultos en su hogar no lo hayan hecho. El principal aporte de este estudio es la aplicación de la metodología de análisis multinivel para calcular el IPM a fin de explicar la pobreza infantil en múltiples dimensiones. Se obtuvo el IPM global para siete países del sur de Asia.

3. (GUIO *et al.*, 2017)¹⁷ "Revisión de las variables de privación material de la UE"

En esta investigación, los autores plantean el concepto de Privación Material (MD, siglas en inglés), basado en una selección de artículos (bienes o servicios) que se consideran necesarios para que las personas tengan un nivel de vida "aceptable" en el país donde viven. El análisis distingue entre los hogares que no pueden permitirse un determinado artículo (los que carecen de medios materiales) y los que no lo poseen por otra razón. La UE utiliza actualmente nueve elementos para medir la MD. Los hogares con graves carencias materiales se definen como los que carecen de al menos cuatro de estos nueve elementos. Estos análisis incluyen una evaluación detallada de la estructura dimensional de todo el conjunto de elementos seleccionados, así como su idoneidad, validez, fiabilidad y adicionalidad. También se analiza la agregación de elementos de MD y se proponen indicadores sólidos de MD para toda la población y para los niños. El uso de los Modelos de Teoría de Respuesta al Ítem (TRI) permite obtener información adicional útil sobre las propiedades de cada ítem en los indicadores de MD propuestos.

¹⁷ Los autores de este trabajo son: **Anne-Catherine Guio** (*Luxembourg Institute of Socio-Economic Research*); **Davis Gordon** (*Director of the Townsend Centre for International Poverty Research and the Bristol Poverty Institute at the University of Bristol, UK*); **Héctor Najera Catalán** (*School for Policy Studies, University of Bristol, Bristol, UK*) y **Marco Pomati** (*School of Social Sciences, Cardiff University, UK*).

Cabe destacar que este trabajo forma parte de la colección de metodologías y documentos de trabajo de EUROSTAT, que son publicaciones técnicas para expertos estadísticos que trabajan en un ámbito determinado.

4. (CEPAL, 2018) *"Medición de la pobreza por ingresos. Actualización metodológica y resultados"*

En este documento se presenta una actualización de las estimaciones de la pobreza realizadas por la CEPAL, con el fin de brindar un panorama de la situación a nivel regional que tenga mayor comparabilidad con respecto a la que se obtiene directamente de las mediciones efectuadas en los dieciocho países¹⁸ analizados.

La actualización aborda dos aspectos, el uso de datos provenientes de fuentes de información más recientes y la revisión de algunos aspectos puntuales de la metodología. Por ejemplo, para comparar las líneas de pobreza extrema y de pobreza de los distintos países se requiere expresarlas en una unidad común, como el dólar. Este procedimiento no está exento de limitaciones, ya que el tipo de cambio oficial puede no reflejar la paridad real entre las monedas y los valores en dólares corrientes no evidencian las discrepancias entre los niveles de precios de los países. Aun así, esta práctica permite tener una visión aproximada del monto que se requiere en cada país para satisfacer algunas necesidades básicas.

Los resultados obtenidos corroboran que la pobreza absoluta en América Latina tiene una incidencia elevada, aunque con una marcada heterogeneidad entre los países. Sin embargo, en distintos informes publicados en el Panorama Social de América Latina, la CEPAL estimó para la región una tendencia decreciente en los niveles de pobreza a principios de la década del 2000 y hasta comienzos de la década de 2010.

En 2016, las líneas de pobreza extrema para las áreas urbanas del conjunto de dieciocho países alcanzaron en promedio simple un valor de 65 dólares por persona al mes y las líneas de pobreza un valor de 132 dólares por persona al

¹⁸ (Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, México, Honduras, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, Venezuela).

mes. Aplicando el tipo de cambio mencionado anteriormente, se observan valores particularmente altos para Argentina y Venezuela.

Las estimaciones presentadas en este documento tienen dos puntos de referencia con los cuales resulta natural efectuar una comparación: las estimaciones de pobreza elaboradas por la CEPAL hasta 2015 y las estimaciones oficiales nacionales.

5. (CONEVAL, 2008) *"Metodología para la medición de la pobreza multidimensional en México"*

Este documento contiene los criterios metodológicos que el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) ha adoptado para la medición multidimensional de la pobreza en México.

La Ley General de Desarrollo Social (LGDS), promulgada en 2004, establece como uno de sus objetivos centrales, "...garantizar el pleno ejercicio de los derechos sociales consagrados en la Constitución Política y asegurando el acceso de toda la población al desarrollo social".

Los ordenamientos de la LGDS identifican dos grandes enfoques para analizar el carácter multidimensional de la pobreza: el enfoque de bienestar y el enfoque de derechos. El primero circunscribe las aproximaciones de necesidades básicas insatisfechas, de activos, de capacidades, entre otras; el segundo está asociado a la existencia de garantías fundamentales, inalienables, insustituibles e interdependientes.

La identificación de la población en situación de pobreza se realiza en dos etapas: en la primera, se determina si los ingresos de una persona son insuficientes para la satisfacción de sus necesidades y si presenta carencias en cada uno de los indicadores; en la segunda, se combinan los indicadores generados en la etapa previa, a fin de identificar a la población en situación de pobreza multidimensional.

En cuanto a la metodología, dado que cada uno de los dos espacios que definen la pobreza brinda un diagnóstico de las limitaciones y restricciones que enfrentan las personas, se establecen criterios diferenciados para definir la presencia o ausencia de carencias en cada uno de ellos. En el espacio del bienestar se establece una cantidad mínima de recursos monetarios (definida por la línea de bienestar) requeridos para satisfacer las necesidades básicas de las personas.

En el espacio de los derechos sociales, se considera que una persona está imposibilitada para ejercer uno o más derechos cuando presenta carencia en al menos uno de los seis indicadores señalados en la LGDS: rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, servicios básicos en la vivienda y acceso a la alimentación.

A partir de ahí, se definen tres tipos de medidas de pobreza multidimensional: de incidencia (porcentaje de personas que padece algún tipo de carencia económica o social), de profundidad (dos tipos, una asociada al espacio del bienestar y otra que concierne directamente al índice de privación social) y de intensidad (se construye a partir de la multiplicación de una medida de incidencia y una medida de profundidad).

México es el primer país cuyo marco normativo crea las disposiciones legales para adoptar una medición de la pobreza que reconoce el carácter multidimensional de esta problemática social.

2.4 Comentarios

Existe amplio consenso en la literatura sobre pobreza respecto a la necesidad de considerar medidas que superen a los cálculos tradicionales basados únicamente en ingresos y NBI. En los últimos años se han realizado muchos estudios orientados a presentar nuevas metodologías superadoras, ampliando el número de dimensiones incorporadas al análisis y estudiando la forma en que éstas influyen en el bienestar. No existe un patrón único que determine qué metodología debe utilizarse, ni qué dimensiones son necesarias incluir, todo dependerá del contexto temporal y espacial.

Otro punto destacable que se rescata de la bibliografía es la continuidad de uso de los métodos de medición y estimación de indicadores para poder comparar resultados a través del tiempo y el espacio y analizar la evolución de índices de pobreza multidimensional, lo que llevará a la aplicación de políticas públicas.

En definitiva, a través de las mediciones sistemáticas de la pobreza en el tiempo, los países pueden determinar la orientación y magnitud de las acciones que deben llevar adelante para superarla. Les permiten evaluar su efectividad y, eventualmente, les faculta para cambiar o justificar determinadas decisiones de

política pública. Por lo tanto, la alternativa de medición que se escoja no es una decisión trivial. Tiene consecuencias, ya que obliga a enfocarse en algunos aspectos y no en otros, que suele sesgar la acción del Estado en alguna dirección particular.

Capítulo 3

Métodos Multivariantes: Modelos de Ecuaciones Estructurales

3.1 Introducción

Este capítulo tiene por objeto revisar las características principales del Análisis Factorial Confirmatorio (CFA, *Confirmatory Factorial Analysis*), procedimiento encuadrado en los Modelos de Ecuaciones Estructurales (SEM, *Structural Equation Models*). Los SEM se centralizan en el estudio de los modelos de medida, es decir, en analizar las relaciones entre un conjunto de indicadores o variables manifiestas y una o más variables latentes o factores, que no pueden ser observados directamente. Los indicadores pueden ser los ítems de un test, las puntuaciones obtenidas por los individuos en distintas escalas o las preguntas en las encuestas de hogares.

En el área de la psicometría se hace hincapié en que, al tomar las mediciones, se cometen errores (en econometría, generalmente, no se tienen en cuenta los errores de medición); en consecuencia, se fueron desarrollando modelos para estudiar conceptos abstractos, no físicos, que se miden en forma indirecta. Por ejemplo, no se puede obtener una medición directa de la felicidad, la inteligencia, la calidad de un docente, la satisfacción de un cliente con respecto a un producto o servicio, el bienestar de una familia o su capital cultural. Estas variables que no pueden ser medidas directamente se denominan "variables latentes" y su valor depende de las "variables observadas o manifiestas". Estas últimas pueden ser medidas a través de instrumentos como cuestionarios, resultados de exámenes o por observación directa del comportamiento de los individuos (SANCHEZ RIVERO, 2004).

La forma de modelar un fenómeno que requiere representar relaciones entre variables latentes y manifiestas es a través de los modelos de ecuaciones estructurales, cuyo desarrollo se ha convertido en los últimos años en uno de los

procedimientos de análisis más utilizados en investigación en ciencias sociales. En este aspecto los SEM resultan ser un método valioso para medir la pobreza multidimensional ya que a través de ellos se pueden establecer relaciones de causalidad entre las dimensiones consideradas para cuantificar el fenómeno y los indicadores que componen tales dimensiones.

El modelo de ecuaciones estructurales abarca una familia entera de modelos conocidos con diversos nombres, entre ellos, Análisis de la Estructura de la Covarianza, Análisis de Variable Latente, Análisis de Factor Confirmatorio y/o simplemente, Análisis LISREL¹ (JÖRESKOG & SÖRBOM, 1993). Tienen la particularidad de estimar una serie de regresiones múltiples distintas pero interrelacionadas mediante la especificación del modelo estructural utilizado por el programa estadístico (BATISTA FOGUET & COENDERS GALLART, 2000).

El Análisis Factorial Confirmatorio postula ciertas relaciones entre las variables observadas y las latentes considerando un patrón, previamente especificado para los parámetros del modelo (factores de carga, parámetros estructurales, varianzas residuales). Se utiliza principalmente para hacer pruebas de hipótesis que surjan de la teoría, de esta forma, el número de variables latentes y de indicadores que se usarán para medir cada variable latente se determinan de antemano. Esto implica que habría suficiente conocimiento para definir la hipótesis del modelo, que son las relaciones entre las variables latentes o constructos y las variables observadas o indicadores que ellos explican.

3.2 Estimación no experimental y causalidad

El concepto de causalidad está afianzado en muchos aspectos de la investigación científica y parte de la idea de que “nada sucede sin una causa”, todo efecto es consecuencia de una causa inmediata, pudiéndose, de esa manera, ordenar una sucesión de efectos y causas.

¹ **Linear Structural Relationship**, Software de computación, creado por Jöreskog – Sörbon en 1993. (Kari Jöreskog y Dag Sorbom, profesores de la Universidad de Uppsala, Suecia, desarrollaron el programa LISREL en los años '70. Su versión más reciente es la 8.8 de agosto de 2009).

El experimento, como herramienta para verificar la causalidad de las relaciones entre variables, se desarrolló en las ciencias físicas debido, especialmente, a su amplio fundamento teórico que permite conocer las fuentes de variación afines con el fenómeno bajo estudio, además, susceptible de implementar procedimientos de control.

Por otro lado, las ciencias sociales y del comportamiento suelen enfrentarse a procesos cuya teoría es relativamente escasa y carecen de medios para controlar experimentalmente el proceso de recolección de información y, en consecuencia, observar las variables. En estos estudios no experimentales la causalidad de las relaciones se infieren a partir de las relaciones observadas entre las variables (BATISTA FOGUET & COENDERS GALLART, 2000).

Los Modelos de Ecuaciones Estructurales con variables latentes, constituyen una herramienta útil para el estudio de relaciones causales de tipo lineal sobre los constructos. Estos modelos no prueban la causalidad, pero ayudan en la toma de decisiones, rechazando las hipótesis causales cuando se contradicen con los datos, es decir, con la estructura de covarianzas o correlaciones subyacente entre las variables.

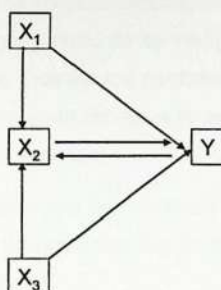
Existen variables que tienen comportamiento conjunto pero la mera asociación estadística entre variables no es una condición suficiente para que exista causalidad. Rex Kline afirma ... "La condición necesaria y suficiente del principio de causalidad podría expresarse en estos términos: una variable A es causa de B si siempre que se da A acontece B, y nunca acontece B si previamente no se ha dado A" (KLINE, 2005). Sólo existe relación causal en el sentido $A \rightarrow B$, puesto que la causalidad es asimétrica. Sin embargo, existe causalidad cuando se encuentra una relación entre dos variables y es posible determinar que no es espuria (o no causal).

3.2.1 Análisis de trayectorias: tipos de relaciones causales

Por lo general, para describir la relación entre una respuesta cuantitativa (Y) de interés y una variable predictora (X) que, se supone, es influyente en esa respuesta (Y), se recurre a un Modelo de Regresión Lineal Simple. En base a información adecuada sobre estas características se estiman los parámetros de ese modelo, se verifica su validez y precisión y se acepta o descarta según corresponda. Cuando hay

más de un factor influyente sobre Y , como X_1 , X_2 y X_3 , se podría plantear un modelo de Regresión Lineal Múltiple para Y con X_1 , X_2 y X_3 , como variables predictoras. Pero, si la situación es más compleja, por ejemplo, si Y influye en los valores que toma X_2 y, a su vez, X_2 está influenciada por X_1 y X_3 . En este caso se necesitaría más de una ecuación para describir toda esta estructura, como muestra la Figura 3.1.

Figura 3.1: Esquema de modelo de trayectorias



Fuente: Esquema extraído de Kline, 2005

Estas ideas comenzaron a ser planteadas sistemáticamente por Sewall Wright (WRIGHT, 1934) en su análisis de trayectorias (*Path Analysis*) y se fortalecieron con (JÖRESKOG, 1971), (JÖRESKOG & GOLDBERGER, 1975). Las asociaciones detectadas entre variables observables o, entre variables manifiestas y latentes, llevan a plantear relaciones de causalidad.

El concepto de análisis causal en las ciencias sociales hace referencia al conjunto de estrategias y técnicas de elaboración de modelos causales que expliquen los fenómenos, con objeto de contrastarlos empíricamente. Sus orígenes se encuentran en el *Path-Analysis*, literalmente traducido como Análisis de Trayectorias, cuyo objetivo es el estudio de los efectos de unas variables consideradas como causas sobre otras tomadas como efectos.

Los elementos que componen el análisis de trayectorias son: el diagrama de trayectorias, el modelo de trayectorias y los coeficientes de Wright. El diagrama de

trayectorias es un gráfico en donde se encuentran representadas las relaciones de causalidad que se supone existen entre un conjunto de variables.

El Análisis de Trayectorias es una técnica similar a la regresión, pero con capacidad explicativa, que estudia los efectos directos e indirectos en el conjunto de las variables observables, asumiendo la existencia de relaciones lineales, errores de regresión no correlacionados y ausencia de errores de medición de las variables. Los coeficientes de Wright (p_{ij} : donde i se refiere a la variable efecto y j a la variable causa) explican el impacto de una variable en otra mediante su descomposición en tres bloques: sendero de la variable independiente a la intermedia, trayectorias de la intermedia a la dependiente y resto de senderos que llevan a la variable final, que no incluyen a la interviniente. Usando los coeficientes de Wright se pueden obtener las diferentes correlaciones entre las variables analizando el conjunto de los efectos, sean éstos directos, indirectos o espurios.

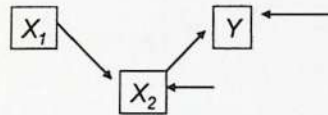
Los efectos causales entre las variables se agrupan en directos, indirectos y espurios, y se pueden representar empleando diagramas de rutas, como muestran las figuras (3.2a), (3.2b) y (3.2c):

Figura 3.2: Relaciones de Causalidad

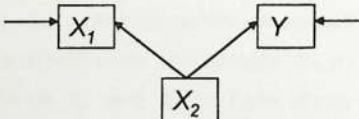
Relación causal directa X_1 causa Y



Relación causal indirecta
 X_1 causa X_2 a través del efecto de Y



Relación espuria o no causal X_2 causa efecto sobre X_1 e Y



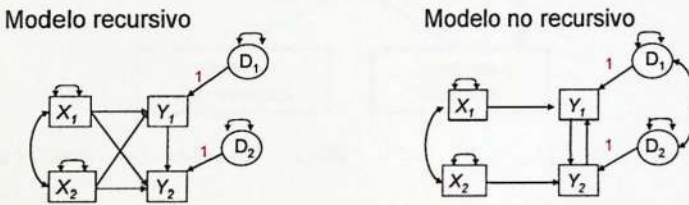
Fuente: Esquema extraído de Kline, 2005

El objetivo principal de un análisis de trayectorias es estimar los aspectos causales frente a los no causales de las correlaciones observadas. Parte de ese objetivo es evaluar qué tan bien explica los datos (correlaciones o covarianzas observadas). Si el modelo no es rechazado no se puede concluir automáticamente que la hipótesis de causalidad es verdadera (VALDIVIESO TABORGA, 2012).

La investigación no experimental puede aproximarse a la experimental partiendo de modelos causales en donde se introduzcan las variables necesarias para refutar aquellas relaciones entre variables no derivadas de la teoría. El modelo recursivo causal incluye todas las posibles relaciones entre las variables del modelo, de modo que la relación entre cada par de variable sea unidireccional, es decir, causal.

Básicamente hay dos tipos de modelos de senderos: los "modelos recursivos" son los más sencillos y tienen dos características básicas, sus residuos (o perturbaciones) no están correlacionados y todos los efectos causales son unidireccionales. Los "modelos no recursivos" tienen lazos de realimentación o pueden tener perturbaciones correlacionadas.

Figura 3.3: Modelos de Trayectoria recursivo y no recursivo



Fuente: Esquema extraído de Kline, 2005

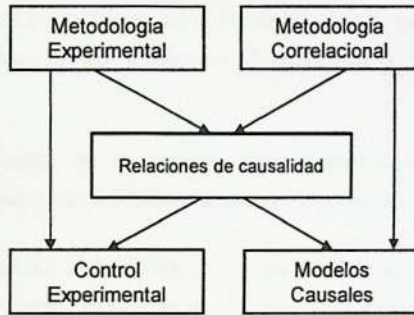
El modelo de la izquierda de la Figura 3.3 es recursivo porque sus errores son independientes y ninguna variable es simultáneamente causa y efecto de otra variable, directa o indirectamente. Por ejemplo, X_1 , X_2 e Y_1 se especifican como causas directa o indirecta de Y_2 , pero Y_2 no tiene efecto de retroalimentación sobre una de sus causas presuntas. En contraste, el modelo de la derecha es no recursivo porque tiene

un lazo de realimentación directo, en el cual Y_1 e Y_2 se especifican como causas y efectos, una de la otra. Además, se observa la presencia de una correlación entre las perturbaciones.

El desarrollo de Modelos de Trayectorias dio lugar a un importante avance en los estudios correlacionales al descubrir más que meras asociaciones entre variables, sino que se aproxima cada vez más al descubrimiento de relaciones causales. En este sentido, constituyen un nexo entre la metodología experimental y la correlacional.

La Figura 3.4 muestra un esquema de la vinculación entre las metodologías mencionadas (KLINE, 2005).

Figura 3.4: Las relaciones de causalidad entre las metodologías experimental y correlacional



Fuente: VI Seminario Científico (SAID, 2008)²

En la metodología experimental, los cambios en la variable dependiente se observan cuando se introducen diferentes niveles en la variable independiente. Además, debido a su propia estructura de control, garantiza una relación causal, ya que elimina la presencia de otras variables conexas, que podrían constituir posibles causas explicativas. En el caso de la metodología correlacional, la existencia o no de asociación se determina por técnicas estadísticas como el Análisis de Correlación y/o Análisis de Regresión. Los procedimientos de control aseguran que se elimine la

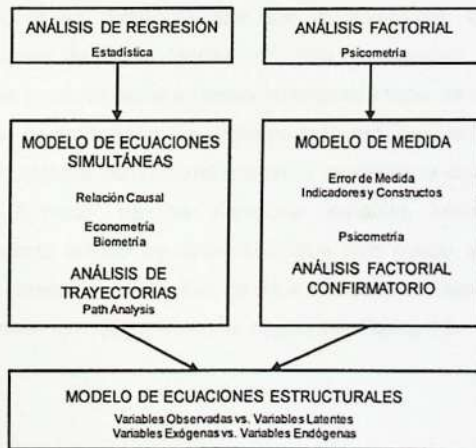
² Metodología de Investigación sobre Discapacidad. Introducción al uso de Ecuaciones Estructurales. VI Simposio Científico SAID, 2008.

posibilidad de un "factor causal común". Por otra parte, las variables conexas pueden ser eliminadas como posibles causas sobre la base de un modelo teórico que establezca que son irrelevantes al proceso causal, o bien mediante el control estadístico.

Para enfrentar estas dificultades, se realiza una adecuada revisión de hipótesis alternativas y el ajuste de modelos con varias ecuaciones, que se basan en la idea de causalidad. Es decir que su aplicación, como técnica para realizar inferencias causales a partir de simples correlaciones, debe basarse prioritariamente en la teoría.

Los modelos SEM constituyen la integración de modelos y técnicas surgidos de diferentes campos como la Estadística, la Biometría, la Econometría y la Psicometría. La estadística hizo su aporte con el Análisis de Regresión. El concepto de variables latentes y errores de medida, así como la relación entre los indicadores o variables observables y las variables latentes o constructos proviene de la teoría Psicométrica a través del Análisis Factorial y la Teoría de la Fiabilidad. Por su parte, la Econometría interviene con el análisis de las influencias direccionales y simultáneas de las variables.

Figura 3.5: El Modelo de Ecuaciones Estructurales como resultado de la integración de modelos y técnicas de diferentes disciplinas



Fuente: VI Seminario Científico (SAID, 2008)

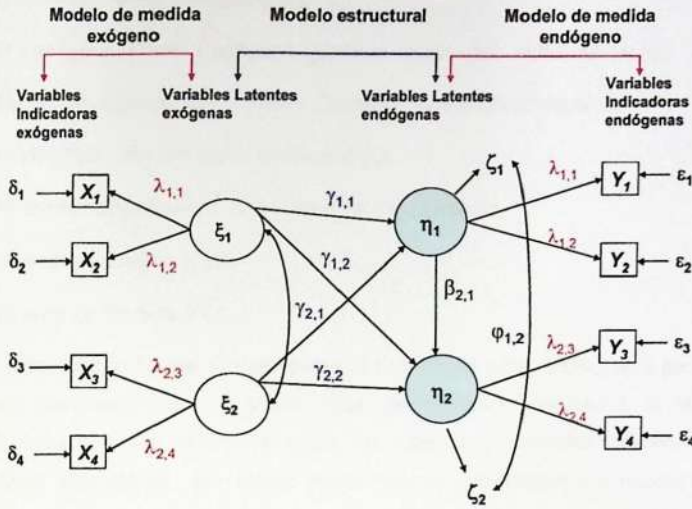
El Modelo de Regresión Lineal Simple trata de explicar una variable dependiente en función de una única variable independiente susceptible de observación y se considera que son medidas sin error. En este modelo no se puede hablar de una relación causa-efecto. Este modelo presenta limitaciones porque, raramente se analiza la relación entre un único predictor y un criterio. En la mayoría de los casos se utiliza un conjunto de variables predictoras, todas ellas observables, como en el caso anterior, entonces se plantea un Modelo de Regresión Lineal Múltiple (MALDONADO *et al.*, 2016).

En las ciencias sociales y comportamentales es bastante frecuente encontrar efectos mediadores o indirectos de ciertas variables, es decir, que se manifiestan dentro de la secuencia causal de otras variables. En estos casos, el modelo de regresión resulta insuficiente. Para poder considerar simultáneamente las relaciones entre las diferentes variables implicadas en un fenómeno, se recurre a los SEM o Modelos Causales, que permiten analizar las interrelaciones entre varias variables al mismo tiempo. Con esta metodología una misma variable puede ser endógena en ciertas ecuaciones y exógena en otras, como en el caso de los efectos mediadores (VERDUGO, *et al.*, 2008).

3.2.2 Modelos de Ecuaciones Estructurales

Las técnicas SEM se distinguen por dos características fundamentales: i) estimación de relaciones de dependencias múltiples y cruzadas; ii) capacidad de representar conceptos no observados en estas relaciones y tener en cuenta el error de medida en el proceso de estimación. En términos más sencillos, estima una serie de regresiones múltiples distintas pero interrelacionadas mediante la especificación de un modelo estructural. Además, permite incorporar variables latentes, que hacen referencia a un concepto teórico no observable que sólo puede ser aproximado y estimado mediante variables manifiestas (o indicadores). Un Modelo General de Ecuaciones Estructurales se representa como muestra la Figura 3.6.

Figura 3.6: Modelo general de Ecuaciones Estructurales completo



Fuente: (KLINE, 2005)

Partes del modelo:

- Círculos o Elipses:** contienen las variables latentes: (ξ) exógenas; (η) endógenas.
- Rectángulos:** contienen las variables indicadoras u observables: (X) exógenas; (Y) endógenas
- Flechas unidireccionales:** desde las variables latentes a las observables indica una relación causal (los factores son causas posibles de las variables indicadoras)
- λ_{ij} : parámetro que cuantifica las relaciones (coeficiente de regresión de cada variable observada sobre su latente)
- ξ : factores exógenos
- η : factores endógenos

g) **Relaciones que definen el modelo: causales** -tipo unidireccional-

i) Exógeno \longrightarrow Endógeno (γ),

ii) Endógeno \longrightarrow Endógeno (β); *covariación* -tipo bidireccional (φ)

h) **Especificación de los errores: Causales** -tipo unidireccional-

i) de predicción del factor endógeno (ζ),

ii) inter-endógenos (φ); Entre latentes e indicadores,

iii) error de medida X (δ),

iv) error de medida Y (ε).

La especificación del Modelo General Estructural para el diagrama de la Figura 3.4, está compuesto por dos submodelos, de medida y estructural. El Modelo de Medida especifica las relaciones entre los factores o variables latentes con sus respectivos indicadores, del mismo modo que se especifican las relaciones entre variables observables y latentes en un Análisis Factorial Confirmatorio (CFA).

$$\text{Modelo de Medida Exógeno: } X = \Lambda_X \xi + \delta \quad (3.1)$$

$$\text{donde: } \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ 0 & \lambda_{23} \\ 0 & \lambda_{24} \end{pmatrix} \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix}$$

$$\text{Modelo de Medida Endógeno: } Y = \Lambda_Y \eta + \varepsilon \quad (3.2)$$

$$\text{donde: } \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ 0 & \lambda_{23} \\ 0 & \lambda_{24} \end{pmatrix} \cdot \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

El Modelo Estructural indica las relaciones direccionales de las variables latentes entre sí, es decir, que son ecuaciones que expresan relaciones entre factores, *i. e.:*

Modelo estructural:
$$\eta = B\eta + \Gamma\xi + \zeta \quad (3.3)$$

donde:
$$\begin{cases} \eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1 \\ \eta_2 = \beta_{12}\eta_1 + \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \zeta_2 \end{cases}$$

Esta configuración indica que el SEM es una combinación de un modelo psicométrico de Análisis Factorial y de otro Econométrico. Otra característica del SEM es que su estadística básica es la covarianza.

La Covarianza entre dos variables X e Y se define según la siguiente expresión:

$$Cov_{XY} = r_{XY} \cdot S_X \cdot S_Y \quad (3.4)$$

donde r_{XY} es el Coeficiente de Correlación Lineal de Pearson entre las variables X e Y , S_X y S_Y son sus respectivas desviaciones estándar. Siendo la covarianza la estadística básica de SEM, en el análisis se consideran dos objetivos esenciales, inferir los esquemas de correlación entre un grupo de variables y, en lo posible, explicar su varianza con el modelo especificado (KLINE, 2005).

3.2.3 Estrategias de modelización

Desde el punto de vista teórico, en la modelización de ecuaciones estructurales se estiman coeficientes desconocidos en un conjunto de ecuaciones estructurales lineales. Las variables que intervienen en el sistema de ecuaciones son: variables manifiestas y variables latentes. La modelización de ecuaciones estructurales considera que existe una combinación causal entre un conjunto de variables latentes, las cuales pueden surgir como combinaciones lineales de las indicadoras.

La flexibilidad de SEM permite utilizar múltiples técnicas en función de los objetivos planteados. (JÖRESKOG, 1993) distingue tres estrategias que se pueden adoptar en relación con los modelos de ecuaciones estructurales:

- a) **Modelización confirmatoria:** se especifica un modelo aislado y SEM se utiliza para evaluar su significación estadística. Aunque esta aplicación parezca rigurosa, no es la prueba más rígida del modelo propuesto. En base a diversas investigaciones se ha mostrado que las técnicas desarrolladas para evaluar los modelos de ecuaciones estructurales tienen un "sesgo confirmatorio" que se inclina a demostrar que el modelo se ajusta a los datos. Por lo tanto, si el modelo propuesto tiene un ajuste aceptable no significa que esté "probado", otros potenciales modelos pueden tener ajustes igualmente aceptables. Esto lleva a afirmar que la prueba más rigurosa se consigue comparando modelos rivales.
- b) **Contraste de modelos:** el análisis más fuerte de un modelo propuesto es identificar y contrastar modelos que representan las verdaderas y distintas relaciones hipotéticas estructurales. Los modelos equivalentes proporcionan otra perspectiva sobre el desarrollo de un conjunto de modelos rivales.
- c) **Desarrollo del modelo:** En general, la teoría sólo puede dar un punto de partida para el desarrollo de un modelo con justificación teórica que pueda ser evaluado empíricamente. Entonces se debe emplear SEM, no sólo para contrastar el modelo empíricamente sino también para obtener perspectivas sobre su re-especificación. Es importante tener en cuenta que la nueva especificación debe realizarse siempre en base a la teoría más que encontrar una justificación empírica (KLINE, 2005).

3.2.4 Identificación del modelo

En el problema de la identificación en los modelos de ecuaciones simultáneas, es necesario tener en cuenta siempre que la estimación correcta contempla, en cada ecuación, el valor de estimación de la otra u otras ecuaciones. Si el modelo no es identificable es imposible obtener un único valor para los parámetros libres.

Cuando los modelos contienen variables latentes es fundamental fijar la escala de cada uno de ellos para que sea identificable. Otra situación que podría generar el

problema de la no identificabilidad es cuando hay más parámetros libres que ecuaciones y por lo tanto uno o varios parámetros quedan expresados en términos de otros, o bien cuando se obtienen valores inadmisibles como covarianzas negativas.

Para determinar si un modelo es identificable, es necesario conocer el número de parámetros libres y de variables observadas y bastará con que satisfaga la siguiente desigualdad:

$$t \leq \frac{(p+q)(p+q+1)}{2} \quad (3.5)$$

donde t es el número de parámetros libres y $(p+q)$ es el número de variables observadas. Esta regla es necesaria pero no suficiente para garantizar la identificabilidad del modelo.

3.2.5 Métodos de estimación

La hipótesis básica de SEM se limita a probar que la matriz de varianzas y covarianzas poblacional es igual a una matriz de varianzas y covarianzas asociada al modelo teórico, es decir que:

$$\Sigma = \Sigma(\theta) \quad (3.6)$$

donde Σ es la matriz poblacional y $\Sigma(\theta)$ es la matriz asociada al modelo propuesto. En realidad, en la práctica no se da la igualdad como tal, entonces, el objetivo es estimar $\hat{\theta}$ de modo que Σ sea lo más cercana posible a $\Sigma(\hat{\theta})$.

Como estimador de la matriz de varianzas y covarianzas poblacional se utiliza la matriz de varianzas-covarianzas muestral, S . La diferencia $(S - \Sigma(\hat{\theta}))$ indica la discrepancia entre lo observado por medio de los datos y las estimaciones obtenidas del modelo.

Los métodos de estimación más utilizados son Máxima Verosimilitud (ML, siglas en inglés), Mínimos Cuadrados no Ponderados (ULS, siglas en inglés), Mínimos

Cuadrados Ponderados (WLS, siglas en inglés) y Mínimos Cuadrados Generalizados (GLS, siglas en inglés). El método ML es, quizás, el más usado y funciona bajo el supuesto de normalidad de los datos, al igual que el método de GLS. Este último y el método ULS son análogos al método de Mínimos Cuadrados Ordinarios empleados en Regresión Lineal, aunque GLS se pondera con una matriz de pesos.

El método WLS, conocido también como de distribución asintótica libre, se utiliza cuando no se cumplen los supuestos de normalidad de los datos. De hecho, es necesario si el modelo contiene una o más variables categóricas y por lo tanto se trabaja con matrices policóricas, poliseriales y tetracóricas. Particularmente, este método requiere que la muestra sea considerablemente grande ($n > 250$).

La Tabla 3.1 presenta el tipo de correlación que conviene calcular de acuerdo con el orden de medición de las variables involucradas.

Tabla 3.1: Medidas de correlación entre variables con distintas escalas

Escala de medición	Continua	Ordinal	Dicotómica
Continua	Pearson	Poliserial	Punto biserial
Ordinal		Policórica	Policórica
Dicotómica			Tetracórica

Fuente: (GENEVAL, 2009)

Cada una de las variables categóricas proviene de un proceso de categorización de una variable continua. Es decir, que se supone que subyace una variable latente continua con distribución normal. En las investigaciones realizadas en el área de las Ciencias Sociales es habitual trabajar con variables categóricas (nivel socioeconómico, bienestar, nivel de satisfacción de un servicio, evaluación de la calidad de un producto, y tantas otras), entonces, cuando se construye un Modelo de Ecuaciones Estructurales se debe trabajar con una matriz de correlación mixta, que contenga varias de las correlaciones indicadas en la tabla anterior (MANZANO PATIÑO y ZAMORA MUÑOZ, 2009).

3.3 Propuesta de un Modelo Estructural para medir pobreza multidimensional

La cuantificación del fenómeno de la pobreza llevó a adoptar y desarrollar diferentes perspectivas de tratamiento de la información. La estrecha relación entre la medición de un fenómeno complejo como la pobreza y el planteo de políticas sociales adecuadas muestra la relevancia y la necesidad de profundizar el análisis de las metodologías de elaboración de los datos. En las últimas décadas, el uso de la información censal y de encuestas de hogares para el estudio de la pobreza se ha convertido en un recurso fundamental para orientar la formulación de políticas sociales.

Actualmente, en varios países de América Latina y en la Unión Europea, se vienen estimando índices de pobreza multidimensional, cada uno adecuado a la idiosincrasia de cada región y país. Generalmente, el índice se construye a través de la combinación de una serie de variables o indicadores que manifiestan una dimensión diferente del concepto. Los indicadores se expresan de acuerdo a un umbral de privación, los cuales son definidos como la incapacidad que tiene un hogar i de alcanzar un nivel z_j en el indicador x_j , de este modo, habrá privación cuando $x_j < z_j$. Consecuentemente, se debe definir con precisión cuáles serán los criterios para definir los umbrales de privación en cada variable. Una vez que los indicadores han sido seleccionados y los umbrales definidos se debe establecer las ponderaciones que estarán asociadas a cada indicador. De este modo, un valor de privación para cada hogar se calculará como una suma ponderada del número de privaciones (CAPPELLARI & JENKINS, 2006).

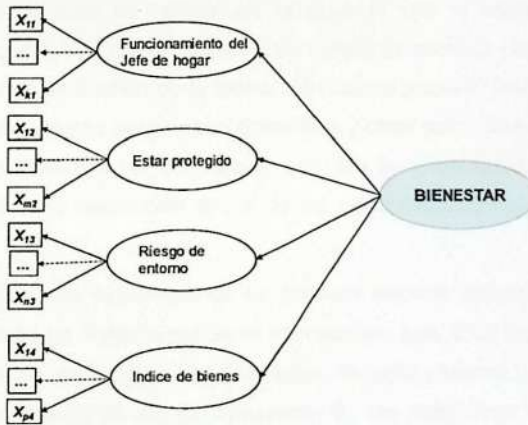
En este trabajo se propone construir un índice de bienestar multidimensional considerando cuatro dimensiones: "funcionamiento del jefe de hogar", referida a la capacidad de éste para generar recursos; "estar protegido", que tiene que ver con los materiales de la vivienda; "riesgo de entorno", relacionada a la seguridad ambiental; "índice de bienes", referida a la capacidad de acumular "riqueza" a través de los bienes durables.

Como se mencionó al inicio de este capítulo, el Modelo del Análisis Factorial estándar implica dos niveles diferentes de variables: factores inobservables e

indicadores observables de estos factores (ítem). Cada una de las dimensiones propuestas es una variable latente que estará compuesta por una combinación adecuada de indicadores (variables manifiestas).

El Método Generalizado de los Mínimos Cuadrados para el Análisis Factorial de Datos Categóricos requiere un nivel intermedio entre los datos observados y la variable latente. Este nivel corresponde a las variables latentes de respuesta, una para cada categoría observada, que, por lo general, se cortan en puntos de umbral correspondientes a juicios observados como Si-NO (respuesta afirmativa o negativa). La correlación tetracórica supone la relación entre dos variables latentes continuas que han sido categorizadas en determinados puntos umbral. El Análisis Factorial puede entonces aplicarse directamente porque se está trabajando con indicadores continuos, aunque inobservables. Sin embargo, existen otros procedimientos de tratamiento de variables categóricas, relacionados con SEM, con mejores resultados que el análisis factorial. Estos son los Modelos de Rasgos Latentes y clases latentes que se desarrollarán en el capítulo 4. La Figura 3.7 muestra el esquema del Modelo de Ecuaciones Estructurales propuesto para construir el índice de bienestar multidimensional.

Figura 3.7: Modelo de Ecuaciones Estructurales para construir el índice de bienestar multidimensional



Fuente: Elaboración propia

En la Figura 3.7 se observa que el índice de bienestar multidimensional es una variable latente, así como las dimensiones que se proponen para su construcción, por lo tanto, están representadas por elipses. Los indicadores utilizados para medir cada dimensión son variables manifiestas, entonces se representan con rectángulos.

Además, las dimensiones son exógenas respecto a los indicadores y endógenas con respecto a la variable latente "bienestar", lo cual está claramente indicado por el sentido de las flechas.

Combinando los conceptos de Análisis Factorial con el Análisis de Trayectorias y de Ecuaciones Lineales Simultáneas, se pueden describir gráfica y analíticamente las relaciones que, se suponen que existen, entre las variables manifiestas y las latentes, considerando la dirección de cada una de esas relaciones. Además, a partir de la información muestral se puede estimar y juzgar su importancia, simplificando, probablemente, los diagramas iniciales hasta obtener un modelo parsimonioso.

3.4 Comentarios

Un aspecto central en la estimación de cualquier tipo de modelo estadístico es que debe estar sujeto al planteamiento de una teoría debidamente asentada en el área de conocimiento en que se esté trabajando. Esto es particularmente exigible en el caso de los SEM, lo cual se refleja en el cuidado por la especificación y la identificación del modelo. El primer aspecto se refiere al correcto planteamiento del sistema de ecuaciones en función de la teoría subyacente (cumplimiento de supuestos básicos, definición de algunos parámetros como fijos y otros como libres o estimables); el segundo aspecto tiene que ver con que la cantidad de información disponible sea suficiente para tener una estimación única de los parámetros libres, más de una o ninguna.

La cuantificación del fenómeno de la pobreza permite adoptar y desarrollar diferentes perspectivas de tratamiento de la información. Los SEM proporcionan una estrategia empírica de estimación del bienestar multidimensional y su aplicación supone un avance importante en la búsqueda de las relaciones causales entre constructos no observables.

Capítulo 4

Métodos de estimación y procedimientos de ajuste

4.1 Introducción

En el capítulo anterior se mencionó que en el ámbito de las ciencias sociales existe una gran cantidad de situaciones o fenómenos que no pueden ser observados o cuantificados directamente, ya sea porque se trata de un concepto abstracto o una característica subyacente (la inteligencia, la calidad de vida, el bienestar, los resultados de una gestión gubernamental, entre muchos otros). Los conceptos de esta naturaleza pueden agruparse como variables latentes, es decir, subyacen en el fenómeno bajo estudio, pero no son directamente observables. Su análisis se realiza mediante variables manifiestas, que se consideran indicadoras de las latentes.

Las investigaciones sobre las llamadas variables latentes han dado lugar al surgimiento de los modelos de variables latentes. Existe una serie de técnicas y modelos estadísticos que, aunque han demostrado su utilidad en el análisis, son poco conocidos por los investigadores sociales. El objeto principal del presente capítulo es plantear los aspectos metodológicos más destacados de estas técnicas y modelos, así como promover su uso mostrando una aplicación en el fenómeno latente del bienestar multidimensional.

Las investigaciones en las que se incluyen variables latentes tienen como finalidad detectar si las relaciones entre las variables manifiestas pueden ser explicadas por una o más variables latentes. De esta forma, los análisis de este tipo reducen el número de variables de un estudio y definen las relaciones entre las variables observadas.

Para encontrar las relaciones causa-efecto entre las variables se utiliza el Análisis Multivariado. Esto es un grupo de técnicas de análisis de datos en continua expansión. El Análisis Factorial (FA, por sus siglas en inglés) es una de esas técnicas y se aplica para determinar interrelaciones entre un número elevado de variables

métricas, explicando dichas interrelaciones en términos de un número menor de variables denominadas "factores" (si son inobservables) o "componentes principales" (si son observables) (SALVADOR FIGUERAS, 2000). El objetivo es encontrar la forma de condensar la información contenida en las variables originales en un conjunto más pequeño de variables con la menor pérdida de información posible. Así, por ejemplo, si un analista financiero quiere determinar cuál es el estado financiero de una empresa a partir de un número de indicadores financieros, construyendo varios índices numéricos que definan su situación, el problema se resolvería mediante un Análisis de Componentes Principales. Si un psicólogo quiere determinar los factores que caracterizan la inteligencia de un individuo a partir de sus respuestas a un test de inteligencia, utilizaría para resolver este problema un Análisis Factorial.

El Análisis de Componentes Principales es una de las técnicas estadísticas de mayor aplicación para estudiar las propiedades de una escala. Por ejemplo, cuando los datos son categóricos, éstos no cumplen con las propiedades de normalidad de las distribuciones o linealidad de las relaciones entre las variables, que son restricciones del FA, y, por lo tanto, limitan su aplicación.

Actualmente, la aplicación de Modelos de Variables Latentes (LVM siglas en inglés) tiene particular interés para abordar el Análisis con Datos Categóricos. Bajo el enfoque de los LVM, se dispone de una variedad de métodos, técnicas y programas informáticos adecuados para aproximar los datos a los objetivos de la investigación.

En este conjunto de opciones hay aproximaciones teóricas y metodológicas especialmente relevantes. Habitualmente se destacan los enfoques clásicos que consisten en tratar a este tipo de datos como Variables Subyacentes (*underlying variables*, UV), en los que se establece una conexión entre variables categóricas y continuas. Otra alternativa conveniente es el Análisis Factorial desde la Teoría de Respuesta al Ítem (IRT por sus siglas en inglés) que proporciona un conjunto adicional de modelos, métodos de estimación y de evaluación del ajuste, que conllevan el avance de diferentes supuestos (CHACON, 2008).

4.2 Características y clasificación de los Modelos de Variables Latentes

El tipo de escala en que se miden las variables manifiestas y latentes (cuantitativas, nominales, ordinales, por intervalos), lleva a la clasificación de los Modelos de Variables Latentes distinguiéndose cuatro tipos (ver Tabla 4.1) (BARTHOLOMEW, *et al.* 2011). Las variables cuantitativas toman valores en un conjunto de números y pueden ser discretas o continuas, por ejemplo, edad, ingresos monetarios, o magnitudes físicas como temperatura, presión atmosférica, peso, talla, entre otras. Las variables categóricas asignan un nivel entre un conjunto de categorías, que pueden estar ordenadas o no, ejemplo de este tipo son: sexo, estado de ocupación, etc. Si se da cierto orden en las categorías, a veces son tratadas con el mismo criterio que las variables cuantitativas, por ejemplo, el nivel de educación asignando un puntaje o score de menor a mayor para indicar los niveles alcanzados.

Tabla 4.1: Clasificación de técnicas estadísticas a usar según el tipo de variables Latentes y Manifiestas

<div style="display: flex; justify-content: space-between;"> VARIABLES MANIFIESTAS VARIABLES LATENTES </div>	Variables Métricas	Variables Categóricas
Variables Métricas	Análisis Factorial (FA)	Modelos de Rasgos Latentes (LTM)
Variables Categóricas	Modelos de Perfiles Latentes (LPM)	Modelos de Clase Latente (LCM)

Fuente: (BARTHOLOMEW *et al.*, 2008)

Esta clasificación no es en ningún caso exhaustiva, sino que recoge los nombres más usados para ciertos modelos de variables latentes que son conocidos dentro de algunas disciplinas.

El objetivo principal de estos modelos es reducir la complejidad de un conjunto de datos explicando la asociación entre variables manifiestas (u observadas) en términos de un conjunto más pequeño de factores (FA, por sus siglas en inglés), clases latentes

(LCM, por sus siglas en inglés), rasgos latentes (LTM, por sus siglas en inglés) o perfiles latentes (LPM, por sus siglas en inglés).

En el FA y LTM, se consideran las variables latentes como variables continuas normalmente distribuidas, las cuales se modelan a través de la variabilidad de las variables manifiestas. Por otra parte, en LPM y LCM, las variables latentes son categóricas, entonces, se considera que provienen de una distribución binomial o multinomial, ejemplo: "sufre" o "no sufre" carencias. Las variables manifiestas (u observadas) en FA y LPM son continuas, se supone que sus distribuciones condicionales son normales. En LTM y LCM, los indicadores o variables manifiestas son categóricos (binarios u ordinales) y se infiere que sus distribuciones condicionales son binomiales o multinomiales.

En la formulación general de FA se considera un conjunto p de variables observadas o manifiestas, \mathbf{X} , que se resumen en q factores o variables latentes no observadas \mathbf{L} , con ($q < p$). Siguiendo el Modelo de Regresión Múltiple, cada variable observada se caracteriza como una combinación lineal de los factores, según la siguiente expresión:

$$x_{ik} = \alpha_{ik0} + \alpha_{ik1}L_1 + \alpha_{ik2}L_2 + \dots + \alpha_{ikq}L_q + e_{ik} \quad (i = 1, \dots, p); (k = 1, \dots, n) \quad (4.1)$$

donde L_1, \dots, L_q son los factores o variables latentes; e_1, \dots, e_p son los términos de error; $\alpha_{ik1}, \dots, \alpha_{ikq}$ son las cargas factoriales y α_{ik0} es el término constante o intercepto y el subíndice i designa la i -ésima variable y k designa el k -ésimo individuo. La ecuación (4.1) representa el Modelo General de Análisis Factorial Lineal. (BARTHOLOMEW *et al.*, 2008)

Aunque el Análisis Factorial Lineal se basa en la idea de la Regresión Múltiple, en lugar de tener una variable dependiente se tienen p -variables dependientes (observadas), esto hace más compleja la comprobación de supuestos. Se debe verificar, entonces, que las variables manifiestas no estén correlacionadas, y que, además, estén explicadas a partir de las variables no observadas o factores. Bajo estos conceptos, se consideran los siguientes supuestos:

- i. Las variables latentes L_1, \dots, L_q , no están correlacionadas y se distribuyen normalmente con media cero y varianza uno.

- ii. Los términos de error e_1, \dots, e_p , no están correlacionados y se distribuyen con media cero, pero pueden tener diferentes varianzas $Var(e_i) = \sigma_i^2 \quad (i = 1, \dots, p)$.
- iii. Las variables latentes y los errores no están correlacionados.
- iv. Tanto las variables latentes como los errores siguen una distribución normal multivariante.

El supuesto (iv) implica que las variables manifiestas X_1, \dots, X_p tienen una distribución normal multivariante, lo cual lleva a un Modelo Normal Lineal de Análisis Factorial. El supuesto (ii) indica que las correlaciones entre las variables manifiestas están suficientemente explicadas por las latentes (FERNÁNDEZ ARAUZ, 2015).

Cuando las variables observadas son categóricas el Modelo Lineal General resulta inadecuado. El problema es que no se cumplen los supuestos básicos ya que si los errores se distribuyen normalmente con media cero y varianza σ^2 y las variables latentes siguen una distribución normal estándar, el vector de variables latentes \mathbf{L} y los términos de error e_i pueden tomar valores independientes y, por lo tanto, las variables manifiestas también podrían tomar cualquier valor. Esto es lo que invalida el Modelo Factorial Lineal para variables categóricas (binarias y ordinales). En consecuencia, es necesario un modelo diferente que permita relacionar el vector de variables latentes \mathbf{L} con las variables observadas.

El Modelo de Rasgo Latente (LTM) resuelve satisfactoriamente los problemas asociados a las matrices de correlaciones entre ítems categóricos. Este modelo presenta dos enfoques de aproximación, la aproximación de variables subyacentes (UV), la cual considera que, para cada variable manifiesta, X_i existe una variable ficticia continua, X_i^* que se distribuye normalmente con media μ_i y varianza σ_i^2 . La esencia de este método es tratar las X_i^* bajo el modelo lineal general del análisis factorial. La segunda aproximación se da desde la Teoría de Respuesta al Ítem (IRT), que es un modelo estadístico que relaciona la probabilidad de respuesta a un ítem en función de un parámetro específico del ítem y el nivel de rasgo latente que presenta un individuo.

La probabilidad de la respuesta positivamente a un determinado ítem es una función de dos componentes (los parámetros del ítem y los parámetros para el individuo que responde) y refleja dónde está localizada esta función en el factor subyacente. Cada ítem de la escala se puede representar gráficamente con una función no lineal (por lo general logística) que muestra las propiedades del ítem.

Una forma alternativa de escribir el Modelo Factorial Lineal es como la esperanza condicional (o valor medio) de X_i dado el vector de variables latentes \mathbf{L} . Cuando X_i es binaria, el valor esperado de X_i dadas las L_j ($j=1, \dots, q$) es igual a $P(X_i=1|\mathbf{L}) = \pi_i(\mathbf{L})$, donde $\pi_i(\mathbf{L})$ es la probabilidad condicional que la variable binaria X_i sea igual a uno, dados los valores de las q variables latentes. Entonces se especifica la forma de la probabilidad $\pi_i(\mathbf{L})$ como una función lineal de L_1, L_2, \dots, L_q . La función elegida se denomina "función de enlace".

En cuanto al Modelo de Clase Latente (LCM), tanto las variables latentes como las manifiestas son categóricas, con dos o más categorías. Éste permite detectar la heterogeneidad de una población identificando el menor número posible de grupos presentes en el universo que se estudia.

En el LCM se define una clase latente por el criterio de la independencia condicional, un supuesto fuerte que debe cumplirse, puesto que, dentro de cada clase latente cada variable manifiesta es estadísticamente independiente de las otras variables. Esto indica que las variables latentes son las causantes de la relación observada en las variables manifiestas, por lo tanto, no existe una relación directa entre las variables manifiestas; toda la asociación observada entre las variables manifiestas está medida o explicada por las variables latentes (LAZARSELD & HENRY, 1968), (GOODMAN, 1974).

Otro de los supuestos básicos es el de homogeneidad interna de las variables latentes: cada uno de los miembros de una clase latente tiene una distribución de probabilidad igual respecto a la de la variable latente, y ésta será diferente a la de los individuos pertenecientes a otras clases. La parametrización del modelo se puede realizar de dos maneras diferentes, mediante probabilidades condicionadas o por un modelo log-lineal. El LCM supone un modelo paramétrico para obtener la clase de pertenencia de un hogar a una determinada clase latente (por ejemplo: "con bienestar"

o "sin bienestar"), y lo hace tomando como base los datos observados. Los parámetros del modelo son las probabilidades de las clases latentes y las probabilidades condicionales, es decir; (1) la existencia de C clases latentes o subpoblaciones o submuestras (no observables directamente) y (2) las probabilidades condicionales, también denominadas probabilidades a posteriori, las cuales se obtienen cuando las categorías de las variables (ítems) manifiestas pertenecen a determinada categoría definida en la clase latente.

4.3 El Análisis Factorial desde el contexto de la Teoría de Respuesta al Ítem

Como se mencionó anteriormente, dentro de los métodos de Análisis factorial confirmatorio, existe un caso especial que son los modelos de "Teoría del Respuesta del Ítem" (IRT, por sus siglas en inglés). Estos modelos asumen que solo hay una variable latente y por ello son solo un caso especial de FA. Existen tres tipos de IRT de uno, dos y tres parámetros. Para el cálculo de pobreza (o bienestar) la literatura recomienda usar el de dos parámetros porque informa sobre dos elementos que son clave.

El modelo IRT supone que la "privación" es un rasgo latente no observable que no se puede medir directamente, como decir talla o peso, ya que es un concepto en lugar de una magnitud física. Sin embargo, este concepto de "privación" se puede medir indirectamente utilizando preguntas de encuestas sociales sobre la capacidad del encuestado para pagar ciertos bienes de consumo duradero, morar en una vivienda digna, sin riesgos de contaminación ambiental y participar en actividades sociales. Describe la relación entre las respuestas de un individuo (u hogar) a los ítems del cuestionario y un rasgo latente no observado como el nivel de felicidad, de bienestar o privación material (FUSCO & DICKES, 2008).

El modelo IRT es una de las aproximaciones más utilizadas cuando los ítems (variables manifiestas) son dicotómicos y consiste en construir un modelo que exprese la probabilidad de una respuesta positiva dada por un individuo, como función de las variables latentes. Si se considera una variable X_{ij} para representar la respuesta del hogar j al ítem i , esta variable tomará el valor 1 si la respuesta es "positiva" a

carecer del ítem, y 0 si la respuesta es "negativa". La probabilidad de responder "positivamente" al ítem i dependerá del parámetro θ , que es el valor que toma la privación (latente) de cada hogar. Así, las familias que tengan privación latente tendrán probabilidades más altas de responder positivamente a los ítems que aquellos con menor privación latente. Por lo tanto, la probabilidad $P_i(\theta)$ es una función monótona creciente del parámetro θ y se denomina Función de Respuesta al Ítem o Curva Característica de Ítem (CCI).

La función $P_i(\theta)$, según el modelo que se proponga para su estimación, dependerá también de otros dos parámetros: dificultad, $\lambda_{i0} \in (-\infty, +\infty)$ y discriminación, $\lambda_{i1} \in (-\infty, +\infty)$.

La "dificultad" de un ítem se define por la probabilidad de que la persona (u hogar) carezca de ese ítem, por lo tanto, se espera que, un indicador de privación incluya ítems con rango de diferentes "puntajes" de severidad, es decir, algunos ítems de baja severidad, algunos de severidad media y otros de alta severidad. Por otro lado, la "discriminación" mide qué tan bien ese ítem hace la diferencia entre los "privados" y los "no privados" del ítem en cuestión, por lo tanto, se esperan puntuaciones altas de "discriminación".

La probabilidad $\pi_i(\mathbf{L})$ denota la probabilidad de "respuesta positiva", entonces, el Modelo General de Rasgo Latente de dos parámetros se define como:

$$\pi_i(\mathbf{L}) = \frac{\exp\left(\lambda_{i0} + \sum_{j=1}^q \lambda_{ij} L_j\right)}{1 + \exp\left(\lambda_{i0} + \sum_{j=1}^q \lambda_{ij} L_j\right)} \quad (4.2)$$

Un caso especial importante es cuando se toma $q = 1$, de este modo, se obtiene el Modelo de Rasgo Latente Unidimensional.

$$\pi_i(L) = \frac{\exp(\lambda_{i0} + \lambda_{i1} L)}{1 + \exp(\lambda_{i0} + \lambda_{i1} L)} \quad (4.3)$$

Para una mejor interpretación, el modelo para $\pi_i(\mathbf{L})$, lineal en las variables latentes, se denota, mediante la transformación logit de la siguiente manera:

$$\text{logit}\pi_i(\mathbf{L}) = \ln \frac{\pi_i(\mathbf{L})}{1 - \pi_i(\mathbf{L})} = \lambda_{i0} + \sum_{j=1}^q \lambda_{ij} L_j \quad (4.4)$$

A cada variable observada X_i le corresponden $q+1$ parámetros a estimar (λ_{i0} y los parámetros discriminantes $\lambda_{i1}, \dots, \lambda_{iq}$). El parámetro λ_{i0} es la probabilidad de que una persona (u hogar) responda positivamente a la carencia de un ítem, mientras que, cuanto mayor sea el valor de λ_{ij} para la variable observada X_i , mayor será la diferencia en la probabilidad de obtener una respuesta positiva entre dos hogares que están situados a cierta distancia en la dimensión latente. Cuanto más elevado sea el parámetro, más fácil resultará discriminar entre los hogares en función de sus respuestas a cada variable observada (BARTHOLOMEW *et al.*, 2008).

La estimación de los parámetros permite asignar los valores correspondientes de las variables latentes a cada individuo u hogar en función de la presencia o no de carencias en las condiciones observadas. Toda la información sobre las variables latentes se encuentra en su distribución posterior dado un conjunto de respuestas observadas, que se denomina "patrón de respuestas".

Para obtener la puntuación del factor en el modelo de rasgo latente es necesario buscar un predictor adecuado para cada L_j dadas X_1, \dots, X_p . Usando la idea de regresión se sugiere recurrir al valor de la media condicional (esperanza condicional) $E(L_j | X_1, \dots, X_p)$, ($j = 1, \dots, q$). Pero estas medias no son combinaciones lineales de X_1, \dots, X_p , no obstante, se pueden calcular fácilmente: son funciones monótonicas (para la función de enlace logit) de las puntuaciones de las componentes, dadas por:

$$Y_j = \sum_{i=1}^p \lambda_{ij} x_i \quad (j = 1, \dots, q) \quad (4.5)$$

Las componentes, son una suma ponderada de las respuestas observadas usando como ponderaciones los coeficientes de discriminación (λ_{ij}), que, además, determinan la posición de los hogares en una escala dentro de la dimensión latente.

La IRT se diferencia del FA por la transformación del modelo matemático sobre el cual se fundamenta, pasando de un modelo lineal a un modelo logístico, para adecuarse al tipo de datos. Otra diferencia es que se deja de lado el concepto de puntuación verdadera y puntuación observada, y sólo se tiene en cuenta el nivel de rasgo expresado con " θ ". Este parámetro se estima para cada hogar, en cada ítem, reemplazando la escala de medida, por una escala que conceptualmente está en el intervalo de $(-\infty, +\infty)$, y que la media se encuentra en 0 (cero). Cuando los valores de capacidad se estandarizan, el parámetro de dificultad, λ_{i0} , suele variar en el intervalo $(-2, +2)$, de modo que cuando está próximo a dos indica una elevada dificultad en el ítem correspondiente. El criterio de dificultad se mide en unidades de desviación estándar de la media, se rechazan los ítems con dificultad mayor al valor absoluto de tres desviaciones estándar de la media (EUROSTAT, 2012).

Por su parte, la discriminación λ_{ij} está relacionada con la pendiente de la curva Característica del Ítem (CCI) en el punto λ_{i0} de forma que cuanto mayor sea la pendiente de la curva, mayores serán las diferencias en las probabilidades $P_i(\theta)$ para pequeñas variaciones en el mismo valor latente, lo que indica que el ítem en cuestión tiene una gran capacidad de discriminación entre los hogares con escasa privación y los que tienen elevada privación latente. A pesar de que, este parámetro también está definido en el intervalo $(-\infty, +\infty)$, cuando es negativo habitualmente el ítem es eliminado puesto que su definición es errónea y/o se produjo una pérdida de información de los hogares con mayor privación latente (significa que la probabilidad $P_i(\theta)$ es menor para los individuos de privación baja que para los de alta). Por otro lado, no es habitual que exceda el valor dos, por lo que en la práctica el rango usual es $(0, 2)$. Los valores próximos a dos denotan ítems con gran poder de discriminación, de la misma forma que valores próximos a cero se asocian a ítems escasamente discriminatorios. El criterio adoptado es descartar todos los ítems cuya correlación con la variable latente es inferior a 0,4 (EUROSTAT, 2012).

Los diferentes modelos que formalizan la probabilidad de responder positivamente a un ítem (considerando la privación como latente) en función de los

parámetros θ , λ_j y λ_{j_0} parten de una serie de hipótesis relativas a los datos en los que se van a aplicar dichos modelos. Entre estas hipótesis hay dos que son fundamentales: las hipótesis de "independencia local" y de "unidimensionalidad".

La hipótesis de unidimensionalidad establece que el conjunto de ítems utilizados pretende cuantificar una única dimensión o variable latente. En base a este supuesto, se definen los ítems que miden cada una de las dimensiones latentes: "estar protegido", "riesgo de entorno", "índice de bienes" y "funcionamiento del jefe de hogar", individualmente, y por lo general, se descartan los que no discriminan (EUROSTAT, 2012).

Por su parte, la hipótesis de independencia local significa que, dado un valor determinado de privación latente (θ), la probabilidad de que un hogar responda positivamente a un ítem determinado es independiente (localmente) de los demás ítems utilizados para la medición. Por este supuesto, la probabilidad conjunta de observar un determinado patrón de respuesta es el producto de las probabilidades de respuesta de cada ítem.

En realidad, estas dos hipótesis están relacionadas. Así, si después de fijar un valor latente concreto (θ), los ítems no son localmente independientes, es decir, están correlacionados, se debe a que estos ítems tienen varias dimensiones latentes en común, de forma que si se considera una única dimensión y las demás dimensiones se mantienen constantes, entonces los ítems serán independientes. Por consiguiente, cuando se verifica la hipótesis de unidimensionalidad, se cumplirá también la independencia local, de modo que estas dos hipótesis son equivalentes. Sin embargo, la más apropiada para medir el bienestar es la hipótesis de unidimensionalidad.

Los Modelos de Respuesta al Ítem surgen para paliar una serie de deficiencias que resultan de la teoría clásica con respecto a las pruebas de ajuste de los modelos. Pueden resumirse en el hecho de que estos modelos suponen que todas las diferencias en respuestas de los individuos a un ítem se deben exclusivamente a la capacidad de dichos individuos, considerándose constantes todas las demás fuentes potenciales de variación en las respuestas. Además, los modelos clásicos presentan problemas en cuanto a la confiabilidad de las pruebas y giran en torno a éstas en conjunto, en lugar de considerar separadamente cada uno de los ítems que lo forman.

En el mundo real más de un factor puede ser causante de las respuestas, por lo que se plantean modelos multidimensionales.

4.3.1 Estimación de los parámetros del Modelo de Rasgos Latentes

Esencialmente, el método de estimación de los parámetros del modelo de rasgo latente es el Método de Máxima Verosimilitud.

El criterio de máxima verosimilitud conjunta consiste en estimar simultáneamente los parámetros de los ítems y de los individuos. La muestra de las observaciones es una matriz $\mathbf{X} = (x_{ki})$, con $k = 1, \dots, n$ el subíndice sobre los individuos (hogares) y $i = 1, \dots, p$ el subíndice sobre los ítems. Cada respuesta x_{ki} de \mathbf{X} es la respuesta del individuo k al ítem i . El objetivo es obtener los estimadores $\hat{\theta}$ y los estimadores de los parámetros del ítem i , a partir de la matriz \mathbf{X} . El logaritmo de la función de verosimilitud conjunta se expresa según:

$$L(\theta, \lambda_{ki}, \lambda_{i0}, \mathbf{x}) = \prod_{k=1}^n \prod_{i=1}^p P_i(\theta, \lambda_{ki}, \lambda_{i0})^{x_{ki}} [1 - P_i(\theta, \lambda_{ki}, \lambda_{i0})]^{1-x_{ki}} \quad (4.6)$$

Sin embargo, algunos problemas de convergencia y propiedades no deseables de los estimadores de máxima verosimilitud conjunta han motivado que el método más adecuado de estimación para los modelos de rasgo latente sea el de máxima verosimilitud marginal, en el cual, considerando que la distribución de la variable latente está dada por la función de densidad $f(\theta)$, la probabilidad marginal asociada al vector de respuestas $\mathbf{x} = (x_1, \dots, x_n)$, se calcula de la siguiente forma:

$$P(\mathbf{x} | \lambda_j, \lambda_0) = \int_{-\infty}^{\infty} \prod_{i=1}^n P_i(\theta, \lambda_{ij}, \lambda_{i0})^{x_i} [1 - P_i(\theta, \lambda_{ij}, \lambda_{i0})]^{1-x_i} f(\theta) d\theta \quad (4.7)$$

A partir de la expresión anterior, la función de verosimilitud marginal viene dada por:

$$L(\lambda_j, \lambda_0, \mathbf{x}) = \prod_{x=1}^{2^p} \pi_x^{r_x} \quad (4.8)$$

siendo π_x la probabilidad asociada al vector de respuestas \mathbf{x} y r_x la frecuencia observada de dicho vector. De la maximización del logaritmo de la expresión (4.8) se obtienen los estimadores por máxima verosimilitud marginal de los parámetros λ_j y λ_0 . El inconveniente de este método es que el elevado número de posibles vectores de respuesta hace que el proceso de estimación sea muy lento cuando el número de ítems es grande. Para salvar este inconveniente se introdujo el algoritmo Esperanza-Maximización (algoritmo EM) de modo que es plenamente operativo cualquiera sea el número de ítems. Los estimadores de los parámetros de los ítems obtenidos por máxima verosimilitud marginal son consistentes y asintóticamente eficientes.

4.4 El Modelo de Clases Latentes para datos binarios

Las relaciones de dependencia entre las variables categóricas de una tabla de contingencia se deben, en la mayoría de los casos, a una asociación entre cada una de ellas y la variable latente.

El modelo de clase latente surge debido a la necesidad de explicar la relación existente entre un conjunto de p variables observadas directamente $\mathbf{X}' = (x_1, \dots, x_p)$, medidas sobre una muestra de n individuos (u hogares). Dicha relación podría estar definida mediante las variables latentes, denotadas por L , las cuales se expresan a través del vector $\mathbf{L}' = (L_1, \dots, L_q)$, con $q < p$. Cada categoría de L se denomina "clase latente".

Si se considera una variable latente L con C categorías o clases latentes y p variables manifiestas (x_1, \dots, x_p) de la variable latente, estas variables conforman el modelo de clases latentes, el cual está definido por:

$$P(\mathbf{X} = \mathbf{x}) = \sum_{j=1}^C P(L = j, \mathbf{X} = \mathbf{x}) \quad (4.9)$$

Donde (x_1, \dots, x_p) es un patrón de respuestas cualquiera, $P(\mathbf{X} = \mathbf{x})$ es la probabilidad conjunta de que las variables manifiestas sean iguales a un cierto patrón de respuestas y $P(L = j, \mathbf{X} = \mathbf{x})$ es la probabilidad conjunta de tener un patrón de respuestas \mathbf{x} , que pertenece a la clase latente j . El modelo (4.9) puede expresarse como:

$$P(\mathbf{X} = \mathbf{x}) = \sum_{j=1}^C P(L = j)P(\mathbf{X} = \mathbf{x} | L = j) \quad (4.10)$$

Siendo $P(L = j)$ la probabilidad de pertenecer a la clase j (probabilidad a priori) y $P(\mathbf{X} = \mathbf{x} | L = j)$ la probabilidad condicional de obtener un determinado patrón de respuestas para un hogar perteneciente a la clase j . De este modo, la expresión (4.10) puede escribirse:

$$\begin{aligned} P(\mathbf{X} = \mathbf{x}) &= \sum_{j=1}^C P(L = j, \mathbf{X} = \mathbf{x}) \\ &= \sum_{j=1}^C P(L = j)P(\mathbf{X} = \mathbf{x} | L = j) \\ &= \sum_{j=1}^C P(L = j) \prod_{i=1}^p P(x_i = x_i | L = j) \end{aligned} \quad (4.11)$$

donde $P(\mathbf{X} = \mathbf{x}_i | L = j)$ es la probabilidad de obtener un determinado valor en la variable \mathbf{X}_p para un hogar de la clase j . Los hogares con un patrón de respuesta definido se clasifican dentro de la clase latente j aplicando una asignación modal, es decir, los hogares se asignan a la clase latente para la cual su probabilidad a posteriori, $P(L = j | \mathbf{X} = \mathbf{x})$, es mayor (CASTRO LÓPEZ & TENORIO AGUIRRE, 2010). Para estimar esta probabilidad se emplea el Teorema de Bayes:

$$P(L = j | \mathbf{X} = \mathbf{x}) = \frac{P(L = j, \mathbf{X} = \mathbf{x})}{P(\mathbf{X} = \mathbf{x})} \quad (4.12)$$

La probabilidad de que un individuo responda positivamente a un ítem (o variable manifiesta) se define en forma condicional en función de las M clases latentes.

$$P(\mathbf{x} = 1 | L = j), \quad j = 1, \dots, C \quad (4.13)$$

El modelo no impone ninguna forma funcional sobre las probabilidades condicionales. En este caso, las probabilidades condicionales son los parámetros de interés junto con las probabilidades previas.

Tanto el MCL como el de MRL permiten sintetizar un conjunto de ítems en un único índice, a partir de la correlación entre ellos y su mutua dependencia con la variable latente.

4.4.1 Estimación de los parámetros del Modelo de Clases Latentes

Las estimaciones máximo-verosímiles de las probabilidades condicionales y a posteriori en Modelos de Clases Latentes resultan un poco más complicadas que para los modelos log-lineales.

Se utilizan varios métodos de estimación, entre los cuales se encuentran: el algoritmo de Newton-Raphson y el algoritmo EM. La mayoría de los paquetes computacionales usa el algoritmo EM, el cual es un procedimiento iterativo de estimación que consta de dos pasos: en el paso **E** se calculan todos los valores esperados dados los valores observados y los "actuales" parámetros del modelo; en el paso **M** se maximiza la función de verosimilitud de todos los datos a partir de los valores esperados calculados en el paso anterior. Esto implica el cálculo de estimaciones actualizadas de los parámetros del modelo como si no faltaran datos. Las iteraciones continúan hasta que se alcanza la convergencia.

Teóricamente, el algoritmo de estimación converge en la mejor solución global (único conjunto de valores de parámetros) de todos los valores posibles. Sin embargo, en algunas ocasiones, el algoritmo de estimación puede converger en una solución máxima local (es la mejor solución en un entorno del espacio de parámetros, pero no es el máximo global).

Los máximos locales están relacionados con la complejidad del modelo, es decir que, a medida que aumenta el número de clases es más frecuente que el algoritmo converja a un máximo local. Para no incurrir en el error de considerar a éste como el

máximo global, se debe ejecutar el algoritmo de estimación varias veces con diferentes valores de inicio de parámetros y verificar que se alcanza la misma solución cada vez, o si hay diferencias, elegir la mejor solución (MONROY CASORLA *et al.*, 2009).

Las probabilidades de las clases latentes describen la distribución de los niveles detectados en una variable latente, a través de los cuales las variables manifiestas son independientes. A estos subgrupos generalmente se los identifica como clases, o grupos. En las probabilidades de las clases latentes hay dos aspectos importantes: 1) el número de clases C de la variable latente L que se modelan para explicar las relaciones entre las variables manifiestas, y 2) el tamaño relativo de cada una de las clases latentes indica cómo se distribuye la población entre el número total de clases C ; así, se pueden identificar grupos normativos, minoritarios o equivalentes en la muestra. Generalmente, la evaluación se inicia con el modelo de una clase y posteriormente, se va incrementado una clase $C + 1$ hasta encontrar el modelo que se ajuste mejor a los datos.

En el proceso de ajuste del LCM puede surgir otro problema conocido como la "subidentificación". En términos generales, significa que hay demasiados parámetros para estimar a partir de los datos disponibles. Una diferencia importante entre los LCM y los modelos de la IRT es que estos últimos imponen relaciones funcionales (tal como la log-lineal) entre la probabilidad de dar una respuesta positiva y la variable latente. Los LCM no imponen ninguna restricción de este tipo de las probabilidades. Entonces, en LCM hay muchos más parámetros para estimar, lo cual tiene implicancias prácticas para la complejidad del ajuste de los modelos (BARTHOLOMEW, *et al.*, 2008).

4.4.2 Asignación de clases

Una vez decidido cuál es el mejor modelo que ajusta los datos, se deben asignar los individuos a las clases identificadas de acuerdo con sus respuestas (este proceso es equivalente al score en el patrón de respuestas en IRT). La asignación de clase se realiza al estimar la probabilidad de que un individuo con un patrón de respuesta particular pertenezca a una determinada clase. Esta probabilidad se designa como probabilidad posterior y se denota según la siguiente expresión:

$$\Pr(\text{que un individuo esté en la clase } j | x_1, \dots, x_p) \quad j = 1, \dots, C \quad (4.14)$$

Dada la probabilidad, la regla de asignación se realiza mediante la probabilidad modal, es decir que los individuos situados en una celda determinada de la tabla serán asignados a la clase latente cuya probabilidad sea mayor. Sin embargo, no existe una correspondencia absoluta entre la clase latente que realmente ocupa un individuo y la clase a la cual se le asigna.

Es posible que algunas probabilidades estén cerca de cero o de uno. En ciertos casos, la posición es más ambigua, por ejemplo, cuando la probabilidad es 0,55, lo que genera dudas respecto a cuál clase debe ser asignado el individuo. Ante esta situación, los individuos pueden asignarse a una u otra clase latente con escasa incertidumbre. Las probabilidades condicionales reflejan las características de los miembros de una clase latente, por lo que de acuerdo con ellas se asigna el nombre a la clase (BARTHOLOMEW *et al.*, 2008).

4.5 Modelos para ítems no dicotómicos

Las variables cualitativas constituyen un conjunto de cualidades no numéricas llamadas categorías, modalidades o niveles – ejemplos: sexo (mujer, hombre), ideología política (liberal, moderada, conservadora), estado civil (soltero, casado, divorciado, viudo), nivel de estudios (ninguno, primario, medio, universitario), etc. - Para cualquier variable cualitativa es obligatorio que las clases sean exhaustivas (proporcionen suficientes valores para clasificar a toda una muestra o población) y mutuamente excluyentes (cada individuo se clasifica en una y solo una categoría).

Se clasifican en dicotómicas y politómicas (según el número de categorías). Las "dicotómicas" sólo tienen dos modalidades, por ejemplo, sexo (Varón y Mujer), padece una enfermedad (Si, No), el fenómeno es de respuesta binaria; las "politómicas" tienen más de dos categorías, ejemplo, fenómenos de respuesta múltiple, lugar de nacimiento, clase social, etc.

También se pueden clasificar en nominales, ordinales y por intervalos. Las "nominal", son aquellas en las que no se define un orden natural entre sus categorías, por ejemplo, la religión (católica, judía, protestante, otra). Las "ordinal" cuando existen

relaciones de orden entre las categorías, entonces se pueden distinguir niveles, mayor, menor, igual o preferencia entre los individuos. Por ejemplo, el estatus social (alta, media, baja). Aquellas por "intervalos", proceden de variables cuantitativas agrupadas en intervalos, ejemplos de este tipo son niveles de ingreso, grupos etarios, o niveles de presión sanguínea.

En el caso de ítems binarios, el objetivo consiste en construir un modelo que exprese la probabilidad de una respuesta positiva $\{\pi_i(\mathbf{L})\}$. En el caso de variables categóricas, donde hay más de dos categorías, es necesario especificar las probabilidades para cada categoría. Las variables observadas se denotan X_1, \dots, X_p . Se considera que para cada variable X_i hay m_i categorías (c_1, \dots, c_{m_i}) y es necesario definir una probabilidad de respuesta para cada categoría. Sea $\{\pi_{i(s)}(\mathbf{L})\}$ la probabilidad que, dada L , una respuesta esté en la categoría s para la variable X_i . El caso general de más de dos categorías de respuestas se puede reconfigurar para tratarse como respuesta binaria, como muestra el siguiente esquema:

Tabla 4.2: Caso general de más de dos categorías de respuesta reconfigurada para tratarse como respuesta binaria

Categorías	0	1
Probabilidad de respuesta	$1 - \pi_i(\mathbf{L})$	$\pi_i(\mathbf{L})$

Categorías	1	...	s	...	m_i
Probabilidad de respuesta	$\pi_{i(1)}(\mathbf{L})$...	$\pi_{i(s)}(\mathbf{L})$...	$\pi_{i(m_i)}(\mathbf{L})$

Fuente: (Bartholomew *et al.*, 2008)

En ambos casos la suma de las probabilidades es igual a uno. En el caso binario, se deriva el modelo logístico que expresa el log de la probabilidad de una respuesta en la categoría 1 como una función lineal de L . La cuestión es cómo

generalizar el argumento cuando hay más de dos categorías. Las categorías se pueden dividir en dos grupos: un grupo contiene las categorías $(1, 2, \dots, s)$ y en el otro están las categorías $(s + 1, s + 2, \dots, m_i)$ y se indica a qué grupo pertenece la respuesta. De este modo, la variable categórica se reduce a una variable binaria. Por lo tanto, parece razonable pensar que cualquier modelo elegido para el caso de tener más de dos categorías, debe ser coherente con el usado para el caso binario. De este modo, se puede expresar:

$$\begin{aligned} \gamma_{i(s)}(\mathbf{L}) &= \Pr(x_i \leq s) = \pi_{i(1)}(\mathbf{L}) + \pi_{i(2)}(\mathbf{L}) + \dots + \pi_{i(s)}(\mathbf{L}) \quad y \\ 1 - \gamma_{i(s)}(\mathbf{L}) &= \Pr(x_i > s) = \pi_{i(s+1)}(\mathbf{L}) + \pi_{i(s+2)}(\mathbf{L}) + \dots + \pi_{i(m_i)}(\mathbf{L}) \end{aligned} \quad (4.15)$$

Donde x_i indica la categoría, donde cae la i -ésima variable manifiesta. Las probabilidades $\gamma_{i(s)}(\mathbf{L})$ son las "probabilidades de respuesta acumulada".

Para este caso, el modelo se puede formular de acuerdo con el grupo que se considere como "respuesta positiva", generalmente, en términos de $(1 - \gamma)$ para relacionarlo directamente con el modelo binario, donde $(s = 1, \dots, m_i; i = 1, \dots, p)$, esto es:

$$\log \left[\frac{\gamma_{i(s)}(\mathbf{L})}{1 - \gamma_{i(s)}(\mathbf{L})} \right] = \lambda_{i0} + \sum_{j=1}^q \lambda_{ij} L_j \quad (4.16)$$

En lugar de usar la función de enlace logit, se puede usar la función normal (probit). Si el modelo usa la función logit como función de enlace, se llama *proportional odds model* (POM). El nombre de POM proviene del hecho, de que en el caso de una variable latente, la diferencia entre dos logit acumulativos para dos individuos con puntuaciones $L_{1(a)}$ y $L_{1(b)}$ es proporcional a $L_{1(a)} - L_{1(b)}$, por lo tanto, hay un parámetro dificultad $\lambda_{i(s)}$ para cada categoría. Esto refleja el hecho de que al aumentar los niveles ($\lambda_{i(s)}$) para una respuesta dada, la dificultad del ítem también aumenta. Un

orden en las categorías implica que los parámetros de dificultad también están ordenados: $\lambda_{i(1)}, \dots, \lambda_{i(m_i)}$.

Sin embargo, los parámetros de discriminación λ_{ij} permanecen iguales entre las categorías de la misma variable, es decir que el poder discriminatorio de los ítems no depende de la distancia entre dos categorías adyacentes. Las probabilidades π_s se obtienen desde las γ_s por:

$$\pi_{i(s)} = \gamma_{i(s)}(L) - \gamma_{i(s-1)}(L) \quad (4.17)$$

donde $\pi_{i(1)}(L) = \gamma_{i(1)}(L)$ y $\gamma_{i(m_i)}(L) = 1$.

Por lo general, $\gamma_{i(s)}(L)$ se denomina "función de respuesta acumulada" y $\pi_{i(s)}$ es la "función de respuesta de categoría" (BARTHOLOMEW *et al.*, 2008).

4.5.1 Modelo de categorías nominales

El modelo de respuesta nominal es un modelo logístico de dos parámetros ("c", el intercepto, y "a", el parámetro de discriminación). La probabilidad de responder a la categoría k del ítem i dado el nivel j de bienestar del hogar se expresa formalmente como:

$$P_{ik}(\theta_j) = \frac{\exp(c_{ik} + a_{ik}\theta_j)}{\sum_{k=1}^m \exp(c_{ik} + a_{ik}\theta_j)} \quad (4.18)$$

donde a_{ik} es el parámetro discriminación para la categoría k del ítem i y c_{ik} es el intercepto de la categoría k del ítem i . Ambos parámetros se calculan para la categoría en la que se segmenta el ítem politómico (DE ANDRADE *et al.*, 2000).

Para que el modelo esté identificado se consideran restricciones en ambos parámetros: $\sum_{k=1}^m a_{ik} = 0$ y $\sum_{k=1}^m c_{ik} = 0$ (PÉREZ GIL, 2004), (BOCK, 1972).

4.6 Bondad del ajuste

En cuanto al modelado estadístico, es importante elegir dentro de un conjunto de modelos alternativos el más apropiado para los datos disponibles. La etapa de diagnóstico de la bondad del ajuste se refiere a la validez de los supuestos del modelo especificado para determinar si éste es correcto y si sirve como aproximación al fenómeno real, precisando así, su poder de predicción.

Las medidas de calidad del ajuste pueden ser de tres tipos, a saber:

- (1) "Medidas absolutas del ajuste", que evalúan el ajuste global del modelo;
- (2) "Medidas del ajuste incremental", que compara el modelo propuesto con otros modelos especificados por el investigador; o
- (3) "Medidas de ajuste de la parsimonia", utilizadas para comparar modelos con diferentes números de parámetros estimados, siendo su propósito determinar la calidad del ajuste obtenido por cada coeficiente estimado (HAIR *et al.*, 1999).

En la literatura se recomienda utilizar varias medidas para evaluar el ajuste del modelo. Las más utilizadas son el estadístico χ^2 (chi-cuadrado), la razón chi cuadrado sobre los grados de libertad (CMIN/DF), el cambio en χ^2 entre los modelos alternativos, el índice de ajuste comparativo (CFI, por sus siglas en inglés), el índice de bondad de ajuste (GFI, por sus siglas en inglés) y el error cuadrático medio de aproximación (RMSEA, por sus siglas en inglés).

Los valores de CFI y GFI varían entre cero y uno (1 indica ajuste perfecto). Los valores superiores a 0,9 sugieren un ajuste bastante satisfactorio entre las estructuras teóricas y los datos empíricos. El chi-cuadrado debe ser no significativo para un buen ajuste, lo contrario implica que la estructura del modelo teórico propuesto es significativamente diferente a la indicada por la matriz de covarianzas observada. No obstante, este estadístico es sensible al tamaño muestral, entonces se lo debe interpretar con precaución.

Si la razón de chi cuadrado sobre los grados de libertad varía entre 2 y 5 se considera un buen ajuste. Cuando se comparan diferentes modelos teóricos, la

reducción significativa en χ^2 de un modelo a otro, también indica un ajuste más adecuado a los datos.

Por último, el índice RMSEA se considera óptimo cuando sus valores son inferiores a 0,06 (CUPANI, 2012).

Los índices AIC y BIC (Criterio de Información de Akaike y Criterio de Información Bayesiano, respectivamente) son dos criterios de uso frecuente para la selección de modelos. El AIC fue propuesto por Akaike en 1974 como un estimador insesgado asintótico de la información de Kullback-Leibler esperada, entre un modelo seleccionado ajustado y el verdadero modelo. El BIC fue propuesto por Schwarz en 1978 como una aproximación a una transformación de la probabilidad posterior de un modelo candidato.

Los indicadores con criterio de información abordan el compromiso entre la complejidad y la capacidad predictiva de un modelo. Cuanto más complejo sea un modelo peor será su capacidad para predecir en un amplio espectro de situaciones. Es decir, cuantas más variables e interrelaciones entre los componentes incorpore un modelo, más concretas serán sus predicciones, pero también menos generalizables. Si un modelo es sencillo podrá predecir en un mayor número de situaciones aceptando cierto error de precisión. Ambos utilizan máxima verosimilitud como criterio de bondad de ajuste y el número de parámetros como medida de complejidad (ARAYA & SEPÚLVEDA, 2010).

La bondad del ajuste puede verificarse de varias maneras. Una forma es usando una prueba estándar de bondad de ajuste para comparar las frecuencias observadas y esperadas a través del patrón de respuestas. Convencionalmente se refiere a estos estimadores como "frecuencias esperadas" cuando se usa la razón de verosimilitud frente a las pruebas de bondad de ajuste de chi cuadrado de Pearson. De hecho, dado que se ajusta el modelo al elegir el valor del parámetro para que estas distribuciones estén lo más cerca posible, la cercanía sería mínima. Una prueba basada en esa medida es la prueba de razón de verosimilitud. Su estadístico de prueba se define como indica la expresión (4.18) (BARTHOLOMEW et al., 2008):

$$G^2 = 2 \sum_{r=1}^{2^r} O(r) \ln \frac{O(r)}{E(r)} \quad (4.18)$$

donde r representa el patrón de respuestas, $O(r)$ y $E(r)$ representan las frecuencias observadas y esperadas respectivamente, del patrón de respuestas r . Alternativamente se puede utilizar el estadístico de prueba chi-cuadrado de Pearson, dado por la expresión (4.19):

$$\chi^2 = 2 \sum_{r=1}^{2^p} \frac{(O(r) - E(r))^2}{E(r)} \quad (4.19)$$

Ambos estadísticos se distribuyen aproximadamente como χ^2 con grados de libertad igual al número de diferentes patrones de respuesta menos el número de parámetros independientes menos 1, esto es $[2^p - p(q+1) - 1]$. Si el tamaño de la muestra n es mucho más grande que el número total de respuestas diferentes, dado por 2^p , entonces las frecuencias observadas y esperadas serán razonablemente grandes y la aproximación, sobre la cual se basa la prueba, será válida. Sin embargo, cuando el número de variables binarias es grande, varios patrones de respuestas tendrán frecuencias esperadas muy pequeñas. La recomendación es que todas las frecuencias esperadas no sean menores a cinco, para que la aproximación chi cuadrado del estadístico sea válida. Si muchas frecuencias esperadas son excesivamente pequeñas las pruebas chi-cuadrado y razón de verosimilitud no alcanzarán la distribución chi-cuadrado y desde el punto de vista práctico, no podrán usarse como pruebas de bondad de ajuste. El problema se puede superar en cierta medida agrupando los patrones de respuestas con frecuencias esperadas menores a 5, pero que podría llevar rápidamente a una situación en la que no se dejan grados de libertad para realizar la prueba. En tal caso, será necesario realizar el análisis con otro enfoque, tales como los criterios de información de Akaike y Bayesiano, los cuales se basan en el uso de principios de selección de modelos (BARTHOLOMEW *et al.*, 2008).

Por otro lado, la prueba exacta de Fisher permite analizar la asociación de dos variables categóricas cuando la muestra es muy pequeña y no se cumplen las condiciones necesarias para que la aplicación de la prueba chi cuadrado sea adecuada. La prueba exacta de Fisher se basa en evaluar la probabilidad asociada a cada una de las tablas de dos por dos que se pueden formar manteniendo los mismos totales de filas y columnas que los de la tabla observada. Cada una de estas probabilidades se obtiene bajo la hipótesis nula de independencia de las dos variables

que se están considerando. La probabilidad exacta de observar un conjunto concreto de frecuencias cuando se asume independencia y los totales de filas y columnas son fijos, está dada por la distribución hipergeométrica. Luego estas probabilidades se usan para calcular el valor " p " asociado a la prueba exacta de Fisher, el cual indica la probabilidad de obtener una diferencia entre los grupos mayor o igual a la observada bajo la hipótesis nula de independencia. Si esta probabilidad es pequeña ($p < 0,05$) se debe rechazar la hipótesis de partida y asumir que las dos variables no son independientes (KOTZ & JOHNSON, 1982).

La Teoría de la Información Estadística (medidas de información) constituye un método alternativo de estimación y contraste. Las medidas de entropía tratan de cuantificar la incertidumbre asociada a un experimento aleatorio. Las medidas de entropía asignan un valor numérico a cada distribución de probabilidad, materializando la idea intuitiva de mayor o menor incertidumbre. Por otra parte, también pueden usarse para cuantificar el grado de homogeneidad con que la probabilidad se distribuye entre los distintos sucesos y por tanto como medidas de la "aleatoriedad" de una variable X .

Teniendo en cuenta estos conceptos, la entropía se utiliza para determinar si el número de clases latentes estimadas es el correcto. Indica el nivel de separación de las clases, mientras más alto sea mejor será la separación de las clases latentes.

Como norma general, la entropía de Shannon (SHANNON, 1948) cuantifica la incertidumbre en una variable aleatoria X a través de su función de densidad de probabilidad, $\Lambda_n = \left[P = (p_1, p_2, \dots, p_n) : p_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n p_i = 1 \right]$ y se define por:

$$H(X) = -\sum p(x) \cdot \ln[p(x)] \quad (4.20)$$

El tamaño relativo de cada una de las clases latentes muestra cómo se distribuye la población entre el número total de clases y se pueden identificar los grupos normativos, minoritarios o equivalentes. El tamaño de las clases es un parámetro elemental para comparar muestras o poblaciones.

Cuando se trata de modelos no lineales existen criterios de bondad del ajuste alternativos, pero deben interpretarse con cautela porque tienen validez limitada. Uno de ellos es la "prueba de Hosmer-Lemeshow", utilizada sólo en variables de respuesta binaria. Consiste en ordenar la muestra de acuerdo con sus probabilidades predichas de menor a mayor y luego dividir en grupos (g) del mismo tamaño, preferentemente en deciles, $\left(n'_k = \frac{n}{10}\right)$ (n es el tamaño de la muestra; $k = 1, \dots, 10$; $g = 10$). Para cada grupo se calculan las frecuencias observadas y esperadas de respuestas. La frecuencia esperada de respuestas positivas es la suma de las probabilidades predichas sobre los individuos en el grupo, mientras que la frecuencia esperada de respuestas negativas es el tamaño del grupo menos la frecuencia esperada de eventos. Cualquiera sea la estrategia de agrupamiento, el estadístico de bondad de ajuste de Hosmer-Lemeshow, \hat{C} , se obtiene calculando el estadístico chi-cuadrado de Pearson de una tabla $g \times 2$ de frecuencias observadas y estimadas. La fórmula para el cálculo es la siguiente:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \hat{\pi}_k)^2}{n'_k \hat{\pi}_k (1 - \hat{\pi}_k)} \quad (4.21)$$

La distribución de \hat{C} es muy próxima a una distribución chi-cuadrado de Pearson, con el número de grupos menos 2 (dos) grados de libertad. Al igual que las pruebas clásicas, los p -valores bajos sugieren el rechazo del modelo (HOSMER & LEMESHOW, 2000).

Por otra parte, las curvas del tipo *Receiver Operating Characteristic (ROC)* son un gráfico en el que se representa la sensibilidad en función de $(1 - \text{especificidad})$. Se utilizan para comprobar la efectividad del modelo en la clasificación de observaciones, para lo cual se construye una tabla de clasificación donde se cruza el verdadero valor de la observación (1 ó 0), con su predicho según el modelo que se considere. La predicción suele hacerse con respecto a un valor de referencia arbitrario π_0 de manera que $x_1 = 1$ si $\hat{\pi}_i > \pi_0$ mientras que $x_1 = 0$ si $\hat{\pi}_i \leq \pi_0$. Habitualmente, el valor que se toma es $\pi_0 = 0,5$. La capacidad predictiva del modelo se resume mediante el

concepto de "sensibilidad". Esta prueba proporciona una fácil comprensión de la precisión, es decir, de la capacidad que posee para discriminar a través de todo el rango de valores. Sin embargo, no tiene aplicación cuando los resultados de la prueba son medidos en una escala dicotómica (RODRÍGUEZ ALONSO & GONZÁLEZ SUÁREZ, 2002).

4.6.1 Ajuste del Modelo de Clases Latentes y asignación de clases

Sea $\pi_{ij} = \Pr(x_i = 1|j)$ la probabilidad de que un individuo de la clase j elegido aleatoriamente responda en forma positiva al ítem i , para $(i = 1, \dots, p; j = 1, \dots, C)$. De este modo, π_{ij} es la probabilidad condicional de una respuesta positiva al ítem i , dada la pertenencia a la clase j .

Sea η_j la proporción de la población en la clase latente j . A veces se refiere a η_j como la probabilidad a priori de pertenecer a la clase j . En los modelos de análisis factorial, se condicionan los valores de q factores o rasgos latentes, (y_1, \dots, y_q) y se modela la distribución condicional de (x_1, \dots, x_p) dado (y_1, \dots, y_q) . En este caso hay una única variable latente categórica que toma valores $(1, \dots, C)$. Por convención se escribe π_{ij} o $\Pr(x_i = 1|j)$ en vez de $\Pr(x_i = 1|y = j)$.

Los modelos se ajustan iterativamente para obtener los estimadores de máxima verosimilitud $\hat{\pi}_{ij}$ de π_{ij} y $\hat{\eta}_j$ de η_j . Durante ese proceso, se pueden obtener máximos locales en lugar de máximos globales, llevando a soluciones que no son las verdaderas. Para salvar esta dificultad, la bibliografía recomienda repetir el proceso varias veces con diferentes puntos de partida y si se repite la solución, entonces es altamente probable que sea el resultado verdadero.

En base a experiencias de diferentes investigadores, los máximos locales no presentan mayores problemas cuando se ajustan con dos o tres clases, el riesgo es mayor cuando aumenta el número de clases.

El objetivo esencial del LCM es asignar los individuos a las clases identificadas usando sus respuestas. La pregunta es: ¿qué se puede decir sobre la pertenencia del individuo después de haber respondido a los ítems? El problema se resuelve

estimando la probabilidad que un individuo, con un determinado patrón de respuestas cae en una clase particular.

$$\Pr(\text{el hogar está en la clase } k | x_1, \dots, x_p) \quad (k = 1, \dots, C) \quad (4.22)$$

Por otro lado, es importante valorar cuál de los modelos propuestos se ajusta mejor a los datos. En el LCM los estadísticos más usados para evaluar el ajuste de los modelos son la razón de verosimilitud (L^2) y los criterios de información AIC, BIC y bayesiano ajustado por el tamaño muestral. Estos estadísticos ponderan, según el número de parámetros, la bondad del ajuste de un modelo medido por el valor de máxima verosimilitud obtenido. Como regla para la selección del modelo que mejor se ajusta a los datos, se debe identificar el modelo que obtenga el menor valor de AIC y BIC (GILL, 2001). Las siguientes expresiones definen los criterios de información:

$$\begin{cases} AIC = 2p - 2 \log(L) \\ BIC = -2 \log(L) + p \log(n) \\ BIC_{aj} = BIC \text{ ajustado por el tamaño de la muestra } (n^* = \frac{(n+2)}{24}) \end{cases} \quad (4.23)$$

donde:

- p : Número de variables explicativas del modelo
- n : Tamaño de la muestra

El criterio de información Bayesiano ajustado por el tamaño de la muestra se obtiene reemplazando n por n^* en la fórmula de BIC.

Existen otros procedimientos formales que ayudan a decidir cuál de los modelos planteados es preferible, como la prueba de razón de verosimilitud de Vuong, Lo, Mendell y Rubin y su versión ajustada de Lo, Mendell y Rubin (NYLUND *et al.*, 2007) (ASPAROUHOV & MUTHEN, 2012). Lo que hacen estos procedimientos estadísticos es testar que un modelo con una clase menos es superior al de referencia, que cuenta con una clase más. Es decir que la hipótesis nula es considerar que un modelo con $k-1$ clases es suficiente para describir los datos y es preferible al de k clases. Si la probabilidad de la prueba es menor a 0,05 se rechaza la hipótesis nula.

El modelo puede parametrizarse a través de las probabilidades condicionadas o un modelo log-lineal. En esta tesis se realiza a través de las probabilidades. Sea un

conjunto de variables categóricas o ítems x_1, \dots, x_p con un número de categorías I_1, \dots, I_p respectivamente. Por otro lado, sea L_q una variable latente con K número de clases. Las ecuaciones básicas del modelo de clase latente son:

$$\pi_{i_1, \dots, i_p} = \sum_{j=1}^J \pi_{i_1, \dots, i_p | k} \quad (4.24)$$

$$\pi_{i_1, \dots, i_p | k} = \pi_k \pi_{i_1 | k} \dots \pi_{i_p | k} \quad (4.25)$$

La expresión, $\pi_{i_1, \dots, i_p, k}$ representa la probabilidad de estar en la celda (i_1, \dots, i_p, k) de la distribución conjunta x_1, \dots, x_p, L_q . Asimismo, π_k es la probabilidad de pertenecer a la clase latente k y $\pi_{i_1, \dots, i_p | k}$ es la probabilidad de tener un patrón de respuesta concreta dado $L_q = k$. El resto de los parámetros π son probabilidades condicionadas. De este modo, los parámetros del Modelo de Clases Latentes son las probabilidades condicionadas y las probabilidades de las clases latentes, las cuales tendrán las siguientes restricciones:

$$\sum_{i_1=1}^{I_1} \pi_{i_1 | k} = \dots = \sum_{i_p=1}^{I_p} \pi_{i_p | k} = 1 \quad \text{y} \quad \sum_{j=1}^J \pi_k = 1 \quad (4.26)$$

Finalmente, una vez que se han identificado las clases, como cada individuo se asigna a la clase latente más probable dado su patrón de respuesta, es importante considerar la probabilidad de una mala clasificación. El error de clasificación se basa en la asignación modal para cada vector de respuestas y en las frecuencias de dichos vectores, considerando el tamaño de la muestra. Un análisis complementario es calcular la "proporción esperada de los errores de clasificación" al asignar cada individuo a la clase cuya probabilidad sea mayor, por la expresión:

$$E = \sum \left[1 - \text{máx} \left(\pi_{k|h_1, \dots, h_p} \right) \right] \hat{\pi}_{h_1, \dots, h_p} \quad (4.27)$$

Valores de error de clasificación más cercanos a cero son indicadores de una mejor clasificación.

Existe otro indicador de la calidad de la asignación a una clase concreta. Es una modificación del contraste λ de Goodman y Kruskal realizada por Clogg en 1981 (CLOGG, 1981) e indica la reducción del error o mejora de la predicción al clasificar las observaciones en lugar de asignar todas a la clase modal. Se define como:

$$\lambda = \frac{[1 - \text{máx}(\hat{\pi}_k)] - E}{1 - \text{máx}(\hat{\pi}_k)} \quad (4.28)$$

Se considera el valor crítico para la clasificación, sólo los casos que presentan una probabilidad de pertenencia a una clase determinada mayor a 0,75.

4.7 Comentarios

En este capítulo se desarrollaron los conceptos de los Modelos de Variables Latentes y su conexión con los datos categóricos; las relaciones entre los Modelos Lineales y el Modelo de Respuesta al Ítem (IRT) para datos categóricos; los métodos de estimación de parámetros y su relación con la evaluación del ajuste del modelo. Asimismo, se justifica que la Teoría de Respuesta al Ítem facilita al investigador una metodología que construye una variable latente continua a partir de múltiples variables observadas discretas (dicotómicas o policotómicas).

En cuanto a las particularidades de los Modelos de Clases Latentes, la principal ventaja es su carácter confirmatorio. Al igual que otras técnicas estadísticas empleadas para clasificar, como el Análisis Factorial o el Análisis Cluster, el Análisis de Clases Latentes es un método exploratorio de poblaciones o muestras, sin

embargo, por encima de su naturaleza exploratoria, el Análisis de Clases Latentes permite realizar todo tipo de investigaciones confirmatorias sobre la naturaleza del concepto latente (como la propia existencia del concepto, la adecuación de los indicadores empleados para su estudio, la óptima distribución de la población en los segmentos identificados, el tamaño de cada segmento, el comportamiento de los individuos ubicados en cada grupo, etc.). No obstante, también presenta ciertas limitaciones, tales como que requiere muestras muy grandes para que las estimaciones de los parámetros sean estables.

En resumen, todas estas ventajas, en su conjunto, suponen una clara mejora respecto a los planteamientos del modelo clásico y son, sin duda, las mejores razones para justificar el carácter hegemónico que actualmente tienen estos modelos. Se presentan como un marco ilimitado de referencia para resolver problemas de medición.

Para ajustar los diferentes modelos IRT y LCM, se usa el programa Mplus (MUTHEN & MUTHEN, 2010). Este paquete presenta la versatilidad necesaria para el análisis de modelos LTM y LCM con múltiples variables categóricas. Otras características especiales incluyen interacciones de variables latentes y análisis de factores no lineales utilizando máxima verosimilitud y un módulo de gráficos posterior al procesamiento, entre otras múltiples funciones.

Estos modelos bien utilizados se convierten en una herramienta fundamental para analizar fuentes de datos como las de las encuestas de hogares. En el caso de Argentina, la Encuesta Permanente de Hogares (EPH) y la Encuesta Nacional de Gastos por Hogares (ENGHo) son las principales fuentes para obtener información sobre la situación social, laboral, educacional, de necesidades básicas, de consumo y preferencias de hogares en los aglomerados urbanos; también son útiles para analizar el impacto de programas y acciones de políticas públicas.

En este sentido, las encuestas se constituyen en insumos importantes para el análisis y la toma de decisiones. Por supuesto que esto se logrará, en la medida que los datos se procesen y analicen apropiadamente, que se obtengan resultados que después se conviertan en información valiosa, y que, finalmente, la información se convierta en conocimiento.

Capítulo 5

Aplicación del concepto latente para la estimación del bienestar

5.1 Introducción

En este Capítulo se propone la construcción de un indicador alternativo para la medición de la pobreza. Concretamente se pone el enfoque en la pobreza desde una perspectiva multidimensional, tanto desde la óptica de la medición como desde la explicación del estado en sí mismo.

Un rasgo muy importante para tener en cuenta es que el indicador tiene una doble lectura. Si bien trata de cuantificar la pobreza y diferenciar por tipos, dado que, la acepción asumida es multidimensional, se acerca en gran medida, al concepto de bienestar de una persona u hogar. De este modo, el indicador podría ser considerado como un medidor del bienestar o del "malestar" (si se concentra en los niveles desfavorecidos). Por lo tanto, la pobreza multidimensional se asocia con el grado de "malestar" en tanto y en cuanto es una pérdida de bienestar.

Se propone un indicador de bienestar multidimensional que proporcione una alternativa a los índices de pobreza revisados en el Capítulo 3, y que aporte información para tomar decisiones de políticas públicas en aspectos socioeconómicos.

En los capítulos anteriores se presenta una justificación teórica para la especificación de las relaciones de dependencia, modificaciones de las relaciones propuestas y demás aspectos de la estimación de un modelo. En el caso de Modelos de Ecuaciones Estructurales (SEM) se puede recurrir a tres estrategias: (1) la modelización confirmatoria, (2) confrontación de modelos y (3) desarrollo del modelo.

La teoría estadística establece que un coeficiente de regresión está compuesto de dos elementos: el "verdadero" o coeficiente estructural entre las variables

dependiente e independiente y la **fiabilidad** de la variable predictor. La fiabilidad es el grado en el que la variable independiente está libre de "error de medida".

Dado que todas las relaciones de dependencia se basan en la correlación observada entre las variables, se espera "fortalecer" las correlaciones utilizando un modelo de dependencia para hacer estimaciones más precisas de los coeficientes estructurales teniendo en cuenta la correlación atribuible, sobre todo, a cualquiera de los problemas de medida (POSA LARA, 2008).

Para elaborar dicho indicador de pobreza (bienestar) se emplea la base de microdatos de la Encuesta Anual de Hogares Urbanos (EAHU) del año 2011¹. A partir de allí, se aplica una metodología basada en técnicas estadísticas multivariantes de análisis de datos –tratadas en el Capítulo 3– para agregar variables, ponderarlas y realizar cuantificaciones con el índice.

Concretamente, se utiliza la metodología de Teoría de Respuesta al Ítem como un Análisis Factorial Confirmatorio para datos categóricos. Una vez seleccionados los ítems que presentan mayor correlación en la dimensión considerada, se aplica el Análisis de Clases Latentes en Modelos de Ecuaciones Estructurales para evaluar cuántas clases latentes presenta la muestra de hogares analizada, en cada dimensión.

5.2 Descripción de los datos

La fuente de datos utilizada para aplicar las metodologías descriptas es la Encuesta Anual de Hogares Urbanos (EAHU) del año 2011 para la Provincia de Tucumán, Argentina.

La EAHU, deriva de la extensión del operativo continuo "*Encuesta Permanente de Hogares - 31 Aglomerados Urbanos*", a través de la incorporación a la muestra de viviendas particulares pertenecientes a localidades de 2.000 y más habitantes, no comprendidas en los dominios de estimación del operativo continuo, para todas las

¹ Encuesta Anual de Hogares Urbanos (EAHU). Diseño de registro y estructura para las bases de microdatos individual y hogar. Vigente desde el tercer trimestre de 2010. Primera edición en formato digital, Buenos Aires, diciembre de 2011.

<https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos>

provincias con excepción de la de Tierra del Fuego, Antártida e Islas del Atlántico Sur y la Ciudad Autónoma de Buenos Aires. Se realizó todos los terceros trimestres desde los años 2010 hasta 2015, y sus dominios de estimación fueron: "Total Nacional de población urbana residente en hogares particulares" y "totales provinciales de población urbana residente en hogares particulares". Se basa en una muestra probabilística, estratificada y polietápica, obtenida por la Dirección de Metodología Estadística del INDEC del Marco de Muestreo Nacional de Viviendas (MMNV), que incluye las viviendas seleccionadas de treinta y un (31) aglomerados urbanos para el operativo EPH. El tamaño de la muestra era de aproximadamente 46.000 viviendas (INDEC, 2011).

Los SEM requieren de muestras grandes para evitar estimaciones imprecisas; además, otros factores afectan los requerimientos del tamaño de la muestra, como la complejidad del modelo, la ausencia de normalidad, la cantidad de valores perdidos, y el tipo de estimación (KLINE, 2005). Sin embargo, la literatura no hace mención alguna respecto a que el tipo de muestreo tenga alguna influencia en la estimación de estos modelos.

En las encuestas de hogares cada registro tiene un número de identificación (CODUSU)², que permite relacionar una vivienda con los hogares y personas que la componen a lo largo de los trimestres en que participa. En la base Hogar todos los hogares que pertenecen a una misma vivienda poseen el mismo CODUSU. Para identificar los hogares se debe utilizar CODUSU y NRO_HOGAR. En la base Individual todos los miembros del hogar tienen el mismo CODUSU y NRO_HOGAR, pero se diferencian por el número de COMPONENTE (EAHU Diseño de registros y estructura de las bases de microdatos, INDEC 2011).

Por otro lado, el Sistema Estadístico Nacional (SEN), encabezado por el INDEC, ha realizado diversos operativos tendientes a la obtención de estadísticas sobre Tecnologías de la Información y la Comunicación (TIC). Entre otras acciones, durante el Censo del Bicentenario realizado el 27 de octubre de 2010, se indagó en la totalidad de los hogares del país respecto a que, si posee computadoras, telefonía

² CODUSU (Código Usuario): en las bases cada registro tiene un número de identificación (CODUSU) que permite relacionar una vivienda con los hogares y personas que lo componen, a lo largo de los cuatro trimestres en los que participa. En la base HOGAR todos los hogares que pertenecen a la misma vivienda tienen el mismo CODUSU, se distinguen por el registro de NRO_HOGAR

fija y telefonía móvil, al tiempo que se preguntó a toda la población del país sobre el uso de computadoras de tres o más años.

Así se desarrolló la primera Encuesta Nacional sobre Acceso y Uso de Tecnologías de la Información y la Comunicación (ENTIC) en Hogares y Personas, que permite contar con información sobre la perspectiva de los usos y accesos de los hogares y de las personas a dichas tecnologías en Argentina.

La ENTIC se administró a todos los hogares y a las personas de 10 años y más, entrevistados para la Encuesta Anual de Hogares Urbanos (EAHU). Se relevaron datos de los accesos a Radio, Televisión, Telefonía (fija y móvil), Computadora e Internet en los Hogares, según el tipo de vivienda y su zona de ubicación, la existencia de hacinamiento crítico, así como la distribución del acceso a estos bienes TIC según quintiles de ingreso total familiar.

La fuente de los datos incluidos corresponde a la ENTIC, relevada en el país durante el tercer trimestre de 2011 en los mismos hogares donde se llevó a cabo la EAHU, haciendo posible la identificación de los correspondientes datos a través del mismo CODUSU y NRO_HOGAR.

Como se mencionó anteriormente, el trabajo se centra en analizar el bienestar de la muestra de la población urbana de la provincia de Tucumán, por lo tanto, de la base completa EAHU se filtraron los datos correspondientes a Tucumán. La muestra quedó con 1.058 hogares y 4.324 personas que, expandiendo a toda la población, representan 338.401 hogares y 1.257.329 personas.

Para el análisis multidimensional del bienestar, se usa el hogar como unidad de análisis en lugar de las personas que lo componen. Las variables elegidas para caracterizar la medición de la pobreza son de carácter variado, e incluyen privaciones referidas a los bienes (televisor, teléfonos fijos y móviles, computadoras y servicio de internet), condiciones de vivienda, su tenencia y ubicación, como así también variables referidas al jefe de hogar que determinan su capacidad económica.

5.2.1 Las variables latentes

Para el análisis multidimensional de la pobreza, se consideran las siguientes variables latentes:

L1, "funcionamiento del jefe de hogar";

L2, "estar bien protegido";

L3, "libre de riesgo de entorno";

L4, "índice de bienes";

Los ítems para L1 corresponden a las características del jefe de hogar: sexo, edad, tiene pareja (es decir si convive con su cónyuge), posee obra social, nivel de educación alcanzado (variable categórica, considerando los ciclos escolares completos) y estado de ocupación. En la base original la variable estado de ocupación tiene tres categorías (ocupado, desocupado, inactivo). Dicha variable fue transformada a dos categorías: 0: ocupado y 1: no ocupado; en esta última se consideran los jubilados, pensionados y todos aquellos individuos que no viven de un ingreso laboral. Las edades de los jefes de hogar, está comprendida entre los 18 y 94 años, al tener un rango tan extenso se agrupó en cinco categorías: categoría 1: de 18 a 29 años; categoría 2: de 30 a 44 años; categoría 3: de 45 a 59 años; categoría 4: de 60 a 69 años; categoría 5: de 70 y más años. La clasificación es en general dicotómica salvo en los ítems edad y el nivel de educación.

Los ítems de L2 se refieren a las características de la vivienda: tipo de vivienda (0 si es de buena condición, 1 si es precaria), si tiene agua de red por cañerías, si el baño está dentro de la vivienda, tipo de materiales de la cubierta externa del techo y si posee cuarto de cocina. La tenencia³ (la variable tenencia, se refiere a si el jefe de hogar es propietario del terreno y/o de la vivienda que está sobre dicho terreno) y si hay hacinamiento; esta última variable calculada como la cantidad de personas por habitación para dormir (tres o más personas se considera que hay hacinamiento)

Los indicadores elegidos para la variable latente L3 son: si la vivienda está ubicada en villa de emergencia, cerca de basural o en zona inundable. Todos los indicadores de estas dimensiones son dicotómicos.

La variable latente L4 "índice de bienes", representa el efecto de que el hogar cuenta con ciertos bienes durables: tiene radio, teléfono fijo, teléfono celular, computadora y servicio de Internet. Todas son dicotómicas: 0 tiene, 1 no tiene.

³ La EPH asigna esta denominación a la variable.

5.2.2 Análisis descriptivo

El propósito es analizar el bienestar de los hogares pertenecientes a los aglomerados urbanos de la provincia de Tucumán. Estos son: el aglomerado Gran Tucumán⁴, Tafi Viejo y otros del interior de la provincia de dos mil (2.000) y más habitantes⁵.

Los datos fueron ponderados sólo para realizar el análisis descriptivo. Los resultados obtenidos del análisis formal son simplemente para evaluar la metodología y no para hacer inferencia sobre la situación de la población en su totalidad. Como se trata de una muestra compleja, existen limitaciones en la ponderación de los datos⁶.

La Figura 5.1 muestra la distribución en porcentajes de la situación favorable de cada uno de los ítems dicotómicos de la dimensión "funcionamiento del jefe de hogar"⁷. Entre los jefes de hogar, el 69% son varones, el 63% tiene pareja, el 77% posee obra social, el 53 % tienen alguna ocupación remunerada y el 42% tiene nivel de educación aceptable.

Se consideran ocupados los jefes de hogar que viven de su trabajo, ya sea empleado o por cuenta propia; inactivos son los jubilados, amas de casa, estudiantes que tienen ingresos por becas y personas que viven de alguna renta o pensión; los desocupados no tienen trabajo fijo, viven de subsidios y/o changas. Esta variable se re-categorizó en 0: ocupado, 1: no ocupado. El 42% de los jefes de hogar alcanzó un nivel de educación aceptable, como un indicador de bienestar, aunque este ítem no se haya dicotomizado. Se considera "nivel de educación aceptable" a los siguientes niveles: secundario completo, terciario o universitario incompleto y completo.

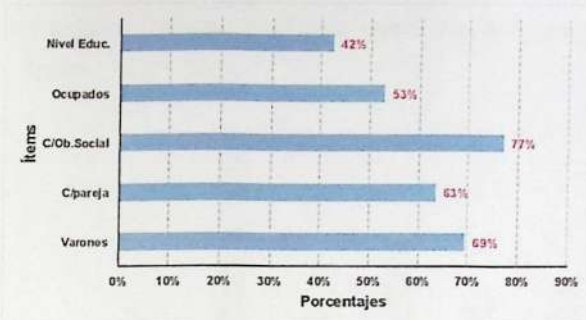
⁴ Los municipios que conforman el aglomerado Gran Tucumán son: San Miguel de Tucumán, Yerba Buena, Las Talitas, Banda del Río Sali y Alderetes.

⁵ En la EAHU estos aglomerados se designan como "resto de Tucumán".

⁶ La introducción de los pesos en una muestra compleja no es el tema de investigación de esta tesis.

⁷ Los datos de la encuesta vienen codificados desde el origen en: 1 para la situación favorable y 2 para la situación desfavorable. Para el análisis realizado en este trabajo y siguiendo con el mismo criterio de utilizar el menor valor para la situación favorable, se re-categorizó en 0 cuando no hay carencia y 1 para el caso contrario.

Figura 5.1: Proporción de hogares con situación favorable en los ítems relacionados al “funcionamiento” de jefes de hogar



Fuente: Elaboración propia en base a los datos ponderados de la EAHU para Tucumán

El ítem **edad**, agrupado en cinco categorías se distribuye como muestra la Figura 5.2, la mayor concentración está en los rangos de 30 a 44 años (31,1%) y de 45 a 59 años (27,8%).

Figura 5.2: Distribución porcentual de los rangos de edades de los jefes de hogar

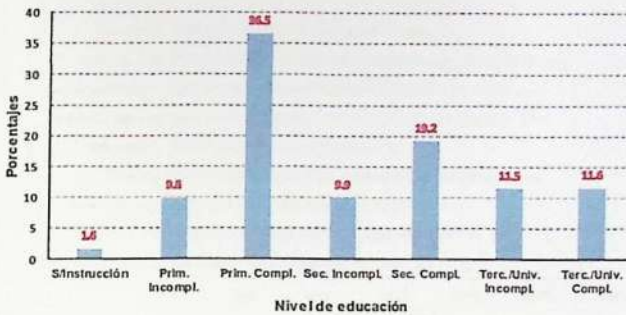


Fuente: Elaboración propia en base a los datos ponderados de la EAHU para Tucumán

Otro ítem importante que hace al “funcionamiento” del sostén del hogar es su **nivel de educación**, determinado en las siguientes categorías: 1: sin instrucción (1,6%), 2: primario incompleto (9,8%), 3: primario completo (36,5%), 4: secundario

incompleto (9,9%), 5: secundario completo (19,2%), 6: universitario incompleto (11,5%) y 7: terciario y/o universitario completo (11,6%) (Figura 5.3).

Figura 5.3: Distribución porcentual de los niveles de educación de los jefes de hogar



Fuente: Elaboración propia en base a los datos ponderados de la EAHU para Tucumán

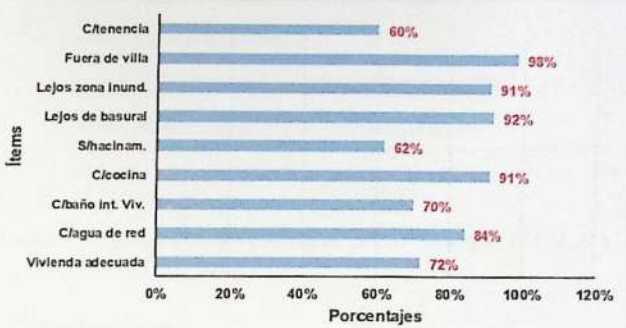
Considerando las variables características de la vivienda, se observa que casi el 60% de las familias poseen una vivienda adecuada. En cuanto a si poseen agua y baño, se indica que el 84% tiene agua de red y el 70% tiene baño instalado con descarga en el interior de la vivienda. El 91% tiene cuarto de cocina exclusivamente para esa función. En general, se destaca que la mayoría de las viviendas goza de buenas condiciones sanitarias, sin embargo, sufren de hacinamiento; el 38% de las familias.

Otro dato considerado para estimar la variable latente "estar protegido" es la propiedad de la vivienda y el terreno donde está emplazada. Se observa que el 60% de los jefes de hogar son dueños del lugar que habita con su familia.

Se supone, además, un potencial riesgo ambiental debido a la ubicación de la vivienda, para lo cual se consideran los ítems "vivir fuera de una villa", "lejos de basural" y de "zona inundable". Se observa que casi un 10% de las viviendas se encuentra en zonas de riesgo ambiental. Sin embargo, estos resultados son relativos, ya que esta encuesta se realizó en zonas urbanas de la provincia de Tucumán, y aunque se hayan ponderado, no se tienen en cuenta los lugares donde están emplazadas las industrias de la caña de azúcar y las citricolas (principales

contaminantes del ambiente). La Figura 5.4 muestra los porcentajes de los ítems descritos en situación favorable.

Figura 5.4: Proporción de hogares con situación favorable en los ítems relacionados con las variables latentes "estar protegidos" y "riesgo de entorno"



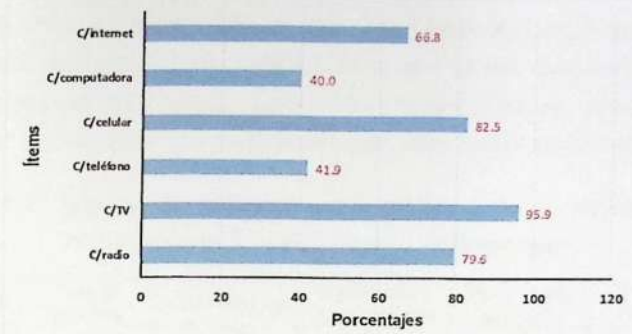
Fuente: Elaboración propia en base a los datos ponderados de la EAHU para Tucumán

Según el enfoque de Amartya Sen (SEN, 1999), para medir el bienestar se deben considerar, además de los ítems mencionados en los párrafos anteriores, las características de los bienes, tomando en cuenta las formas cómo los miembros de un hogar los utilizan. Esa es la base del concepto de funcionamientos, entendidos como los logros o éxitos de una familia (o persona) para hacer y ser lo que desea.

Se usa un conjunto de variables observadas en las viviendas que dependen de los bienes y servicios empleados, tales como: si tiene radio, televisores, líneas de telefonía fija, líneas activas de teléfono celular, computadora y servicio de internet. Todas las variables son dicotómicas, codificadas con: 0 (tiene) y 1 (no tiene).

Nuevamente se destaca la situación favorable en cada uno de los ítems del índice de bienes (Figura 5.5). El 80% de los hogares tienen radio, el 96% tiene TV, el 42% tiene línea activa de teléfono fijo, casi el 83% tiene una o más líneas activas de celular, el 40% tiene computadora, ya sea PC o portátil y el 67% posee servicio de internet. Se distingue que el porcentaje con telefonía móvil es mucho mayor que el de telefonía fija y los hogares que tienen servicio de internet es de casi un 20% menos de los que tienen computadoras.

Figura 5.5: Proporción de hogares con situación favorable en los ítems relacionados al "índice de bienes"



Fuente: Elaboración propia en base a los datos ponderados de la EAHU para Tucumán

5.2.3 Correlaciones

Como parte del análisis descriptivo, es importante analizar si existe correlación entre las variables observadas, ya que de ello dependerá el diseño de Modelos de Ecuaciones Estructurales y su convergencia. Para determinar si existe asociación entre los ítems se usan las tablas de contingencia, las cuales admiten diferentes procedimientos estadísticos para su análisis como las pruebas chi-cuadrado, exacta de Fisher, la Q de Cochran, entre otras. En este caso, se expone el cálculo e interpretación de la prueba exacta de Fisher como método estándar de análisis.

Bajo el supuesto de independencia el valor de p indica la probabilidad de obtener una diferencia entre las variables manifiestas mayor o igual a la observada. Si ($p < 0,05$) se rechaza la hipótesis nula de partida y se asume que los ítems de la prueba están asociados. En caso contrario, se dirá que no existe evidencia estadística de asociación entre ambas variables.

En la Tabla 5.1 se indican los ítems de las dimensiones "funcionamiento del jefe" y "estar protegido" que tienen asociación. Sólo estado de ocupación tiene correlación con tipo de vivienda, hacinamiento, tener baño dentro de la vivienda y tiene cuarto exclusivo para cocina; tiene pareja y sexo del jefe de hogar están

asociadas a hacinamiento. En cuanto a los otros ítems, no presentan evidencia alguna de correlación.

Intuitivamente, se podría suponer que existe cierta asociación entre las variables "estar protegido" (tipo de vivienda, tiene agua de red, tiene baño en el interior de la vivienda, tiene cocina, hacinamiento y tenencia) con los "bienes" que pueda poseer la misma (radio, TV, teléfono fijo, celular/es, computadora e internet).

Tabla 5.1: Pruebas de asociación entre los ítems de las dimensiones "funcionamiento del jefe de hogar" y "estar protegido"

ITEMS	Tipo de vivienda	Hacinamiento	Baño dentro de la vivienda	Tiene cuarto de cocina	Tenencia
Estado	$p=0,002$	$p=0,001$	$p=0,024$	$p=0,01$	S/asociación
Obra social	$p=0,002$	S/asociación	S/asociación	S/asociación	S/asociación
Pareja	S/asociación	$p=0,008$	S/asociación	S/asociación	S/asociación
Sexo	S/asociación	$p=0,001$	S/asociación	S/asociación	S/asociación

Fuente: Elaboración propia en base a los resultados de las pruebas

Tabla 5.2: Pruebas de asociación entre los ítems de las dimensiones "índice de bienes" y "estar protegido"

ITEMS	Vivienda adecuada	Hacinam.	Baño dentro de la vivienda	Tiene cuarto de cocina	Tenencia	Agua de red
Radio	$p=0,001$	S/asociac.	$p=0,004$	$p=0,01$	S/asociac.	$p=0,001$
TV	$p=0,001$	S/asociac.	$p=0,016$	$p=0,001$	S/asociac.	$p=0,003$
Teléfono	$p=0,003$	$p=0,001$	$p=0,001$	$p=0,001$	S/asociac.	$p=0,001$
Celular	$p=0,001$	S/asociac.	$p=0,009$	$p=0,001$	S/asociac.	$p=0,001$
Comput.	$p=0,001$	$p=0,001$	$p=0,001$	$p=0,001$	S/asociac.	$p=0,001$
Internet	$p=0,007$	$p=0,001$	$p=0,008$	S/asociac.	$p=0,001$	$p=0,001$

Fuente: Elaboración propia en base a los resultados de las pruebas

Entre casi todos los ítems de las dimensiones "estar protegido" e "índice de bienes" se observa una fuerte asociación, con excepción del ítem tenencia que sólo está asociado con tener servicio de internet y el ítem hacinamiento que no está asociado con tener radio, TV y teléfono celular. Luego, tener cuarto de cocina no está asociado con tener internet.

Asimismo, se analizan las asociaciones entre los ítems de las dimensiones "estar protegido" y "riesgo de entorno". La Tabla 5.3 muestra los resultados.

Tabla 5.3: Pruebas de asociación entre los ítems de las dimensiones "riesgo de entorno" y "estar protegido"

ITEMS	Vivienda adecuada	Hacinam.	Baño dentro de la vivienda	Tiene cuarto de cocina	Tenencia	Agua de red
Cerca de basural	$p=0,001$	S/asociac.	$p=0,008$	$p=0,03$	S/asociac.	$p=0,001$
Zona inundable	$p=0,001$	S/asociac.	$p=0,002$	$p=0,001$	S/asociac.	$p=0,001$
Villa	$p=0,001$	$p=0,006$	$p=0,005$	$p=0,001$	$p=0,002$	$p=0,001$

Fuente: Elaboración propia en base a los resultados de las pruebas

En general, se observa una fuerte asociación, salvo hacinamiento y tenencia que no están asociados a estar cerca de basural y en zona inundable.

5.3 Aproximaciones a la Teoría de Respuesta al Ítem

La Teoría de Respuesta al Ítem (IRT) está basada en un modelo probabilístico formal y toma en cuenta el nivel de facilidad y capacidad discriminativa de cada ítem y la correlación entre ellos. Su principal función es determinar el número y la naturaleza de las variables latentes o factores que explican la variación y covariación entre un conjunto de medidas observadas, comúnmente conocidas como indicadores o ítems. Un factor (también llamado variable latente, factor latente subyacente o constructo) es una variable no observable que influye en más de una medida observada y determina las correlaciones entre estas medidas observadas.

En general, para la modelización con variables latentes la bibliografía considera seis pasos: especificación, identificación, estimación de parámetros, evaluación del ajuste, re-especificación del modelo y si éste ajusta adecuadamente a los datos, se interpretan de resultados derivados del mismo (KAPLAN,2003; KLINE, 2005).

La "especificación" del modelo está basada principalmente en los conocimientos teóricos del fenómeno estudiado y su aplicación al planteamiento de las ecuaciones matemáticas relativas a los efectos causales de las variables latentes y a las expresiones que las relacionan con los indicadores o variables observables.

Si el modelo teórico es correcto, se procede a la "identificación", en donde se debe asegurar que los parámetros del modelo puedan ser estimados. El modelo está identificado si existe una solución única para cada uno de los parámetros, es decir que pueden estimarse a partir de los elementos de la matriz de covarianzas de las variables observables. Existen algunas reglas generales para identificar un modelo, una de ellas es la regla de los grados de libertad. Los grados de libertad (gl) se definen como la diferencia entre el número de varianzas y covarianzas, y el de parámetros a estimar, entonces, el número de grados de libertad no puede ser negativo. Denotando el número total de variables con $s = p + q$, siendo p las variables endógenas y q las variables exógenas, se tiene que el número de elementos no redundantes es igual a: $\frac{s(s+1)}{2}$, y al número total de parámetros que han de ser estimados en el modelo como t , se definen los grados de libertad: $gl = \frac{s(s+1)}{2} - t$. Según el valor de gl se pueden clasificar los modelos en: **nunca identificado ($gl < 0$)**, modelos en los que los parámetros toman infinitos valores, y por ellos están indeterminados; **posiblemente identificado ($gl = 0$)**, modelos en los que puede existir una única solución para los parámetros que iguale la matriz de covarianzas observada e implicada; **posiblemente sobre identificados ($gl > 0$)**, modelos que incluyen menos parámetros que varianzas y covarianzas. El modelo sobre identificado tiene más información en la matriz de datos que la cantidad de parámetros a estimar, por lo tanto, los grados de libertad son números positivos ($gl > 0$). En cambio, el modelo infra estimado tiene grados de libertad negativos ($gl < 0$), lo que significa que la información es insuficiente para estimar un número de parámetros superior. Como recomendación usual, para evitar problemas, la bibliografía sugiere un mínimo de tres ítems por cada variable latente (CUPANI, 2012).

La "estimación" implica determinar los valores de los parámetros desconocidos y su respectivo error de medición. Se ajusta el modelo por la elección de los valores de los parámetros para que la matriz de covarianza predicha por el modelo sea lo más próxima posible a la matriz observada. La distribución conjunta queda completamente determinada por las covarianzas, de modo que las distribuciones observada y predicha son cercanas. Una de las técnicas ampliamente empleada en la mayoría de los programas informáticos para la estimación de modelos

estructurales, es el de Máxima Verosimilitud (MV), que es eficiente y asintóticamente insesgada cuando se cumplen los supuestos de normalidad multivariada y facilita la convergencia de las estimaciones aún en ausencia de normalidad⁵. Sin embargo, las variables tratadas en esta tesis son categóricas, por lo tanto, los supuestos no se cumplen.

La etapa de "evaluación del ajuste", o diagnóstico de la bondad del ajuste se refiere a la exactitud de los supuestos del modelo especificado para determinar si éste es correcto y sirve como aproximación al fenómeno real, precisando, de este modo, su poder de predicción.

Generalmente, el primer modelo propuesto no es el que mejor ajusta, en consecuencia, se buscan métodos para mejorar el ajuste del modelo y/o su correspondencia con la teoría subyacente. En tal caso, se re-especifica el modelo, añadiendo o eliminando los parámetros estimados del modelo original. Antes de tratar algunos enfoques para identificar las modificaciones del modelo, es aconsejable hacer tales modificaciones con cuidado y considerando las justificaciones teóricas antes que las empíricamente deseables. Para realizar una re-especificación se deben examinar los índices de modificación. El valor del índice de modificación corresponde aproximadamente a la reducción en el chi-cuadrado que se produciría si el coeficiente fuera estimado.

Finalmente, además de considerar el ajuste del modelo, debe prestarse atención a la significación de los parámetros estimados para realizar una interpretación correcta de éstos.

Cuando se aplica el Modelo de Rasgos Latentes (LTM, siglas en inglés) en el campo de la psicometría, se pretende medir la capacidad de los individuos a través de diferentes preguntas. Esta capacidad representa un fenómeno "positivo", sea inteligencia o cualquier otra habilidad de las personas, entonces, en este caso el parámetro referido al individuo se denomina "habilidad". Cuando lo que se mide es la

⁵ La función log-verosimilitud es: $\log L = -\frac{1}{2}(N-1)\left\{\log|\Sigma(\theta)| + tr\left[S\Sigma(\theta^{-1})\right]\right\} + c$;

donde L es la función de verosimilitud, N el tamaño de la muestra, S la matriz de covarianzas de la muestra, $\Sigma(\theta)$ la matriz de covarianzas del modelo y θ es el vector de los parámetros.

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta} P_i(\theta) \right]^2}{P_i(\theta) Q_i(\theta)} \quad (5.1)$$

donde,

$I_i(\theta)$ es la información proporcionada por el ítem i en el nivel de bienestar θ ;

$P_i(\theta) = P(X_{ij} = 1 | \theta)$ y $Q_i(\theta) = 1 - P_i(\theta)$

La información proporcionada por la prueba es simplemente la suma de las informaciones de cada ítem.

El Método de Máxima Verosimilitud, mencionado en párrafos anteriores, proporciona una estimación puntual de los parámetros. En muchas ocasiones es conveniente tener también una indicación de la precisión de la estimación en términos de un intervalo de confianza para el valor verdadero de los parámetros. Así, el valor estimado por Máxima Verosimilitud de un parámetro se puede considerar aproximadamente con distribución normal cuya media es su valor verdadero y su varianza es el inverso de la función de información de la prueba. Es decir que, la amplitud del intervalo de confianza para θ dependerá del valor de la función de información (DE ANDRADE *et al.*, 2000).

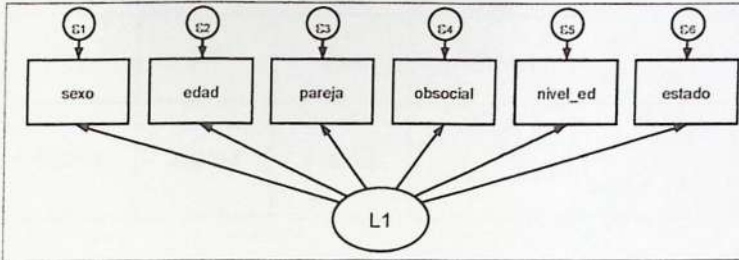
En las siguientes subsecciones, se estiman los modelos usando la aproximación de la IRT para determinar cuáles indicadores son significativos a las variables latentes "funcionamiento del jefe de hogar" (L1), "estar protegido" (L2), "riesgo de entorno" (L3) e "índice de bienes" (L4).

5.3.1 Estimación de parámetros de los ítems para la variable latente "funcionamiento del jefe de hogar"

En el primer modelo, cuyo esquema se muestra en la Figura 5.6, se considera la dimensión "funcionamiento del jefe de hogar" como la variable latente L1 y los ítems observados que caracterizan a la persona a cargo del hogar (sexo, edad, situación conyugal, cobertura médico asistencial, nivel de educación y estado de ocupación), los cuales fueron descriptos en secciones previas. Las variables nivel de educación y edad tienen más de dos categorías, mientras que las demás son

dicotómicas. La Figura 5.6 representa el modelo con seis ítems y un factor (o variable latente, L1) para respuestas categóricas.

Figura 5.6: Modelo Estructural para la variable latente “funcionamiento del jefe de hogar”



Una vez determinado el conjunto de ítems y estimados sus parámetros, así como el nivel de bienestar en la dimensión de cada hogar de la muestra, la IRT permite estimar la precisión y el error típico de medición para diferentes niveles del rasgo. La función de información depende esencialmente de la pendiente de la curva (capacidad de discriminación del ítem) y del error estándar de medición. Es decir, cuanto mayor sea la pendiente y menor el error de medición de un ítem, mayor será la información (CORTADA & CUPANI, 2013).

Para analizar la bondad de ajuste se consideran las pruebas chi cuadrado de razón de verosimilitud (G^2) y de Pearson (χ^2) (Capítulo 4, Sección 4.6), ambos estadísticos se distribuyen aproximadamente como χ^2 con grados de libertad igual al número de diferentes patrones de respuesta menos el número de parámetros independientes menos uno ($[2^p - p(q+1) - 1]$). Cuando las pruebas chi cuadrado de Pearson y de razón de verosimilitud no son significativas ($p\text{-val} > 0,05$), el modelo está bien ajustado.

Otra forma de analizar el ajuste es a través de los Criterios de Información de Akaike (AIC), Bayesiano (BIC) y Bayesiano ajustado por tamaño muestral (BIC_n),

mientras más parsimonioso⁹ es el modelo mejor es el ajuste. Los resultados de las pruebas mencionadas se muestran en la Tabla 5.4.

Tabla 5.4: Resultados de las pruebas de bondad de ajuste por la IRT para la variable latente "funcionamiento del jefe de hogar"

AIC	BIC	BIC _{aj}	χ^2 Pearson	G^2 razón de verosimilitud
11571,1	11670,4	11606,9	924,78 gl = 539 (*) p-val = 0,01	825,89 gl = 539 (*) ¹⁰ p-val = 0,01

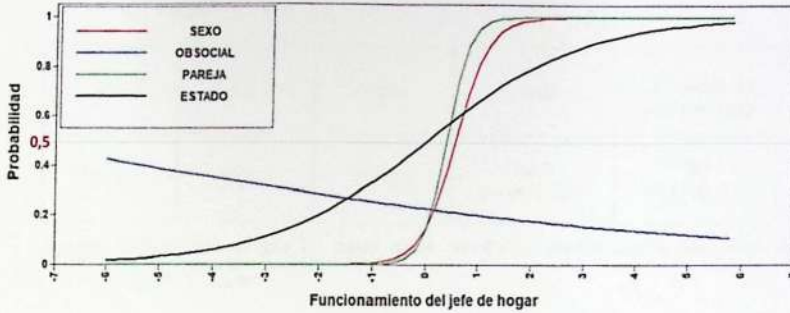
Como se mencionó en párrafos anteriores, el modelo está compuesto por seis ítems, por lo tanto, para cada uno se deben estimar sus respectivos parámetros de dificultad y de discriminación. Para los ítems dicotómicos: sexo, pareja, obra social y estado se estiman los cuatro parámetros de dificultad: λ_{10} , λ_{20} , λ_{30} , λ_{40} y los cuatro de discriminación: λ_{11} , λ_{21} , λ_{31} , λ_{41} (un total de 8-parámetros). Para el ítem nivel de educación se estima un parámetro de dificultad, λ_{50} y seis parámetros de discriminación (número de categorías menos uno) y para el ítem edad se estima un parámetro de dificultad, λ_{60} y cuatro parámetros de discriminación. En total hay 20 parámetros para estimar (parámetros libres).

La Figura 5.7 muestra las Curvas Características de los Ítems dicotómicos y claramente se observa que la curva del ítem "obra social" al no tener curvatura y ser de pendiente negativa, no discrimina.

⁹ "Parsimonia" es simplicidad. Los criterios de información abordan el compromiso entre la complejidad y la capacidad predictiva de un modelo. Cuando más complejo sea un modelo peor será su capacidad para predecir en un espectro amplio de situaciones. Si un modelo es sencillo e incluye sólo los componentes de mayor importancia para el sistema, podrá predecir mejor.

¹⁰ (*) Los grados de libertad se obtienen de la siguiente expresión: $2^4 \cdot 5 \cdot 7 - 20 - 1 = 539$, donde el 2^4 corresponde a los 4 ítems dicotómicos, un ítem de 5 categorías y un ítem de 7 categorías y 20 es el número de parámetros libres.

Figura 5.7: Curvas características de los ítems dicotómicos para la variable latente "funcionamiento del jefe de hogar"



Fuente: Gráfico extraído del programa Mplus

La variable nivel de educación tiene 7 categorías que dificultan el ajuste adecuado del modelo y la interpretación de los resultados. Teniendo en cuenta que en Argentina la escolaridad obligatoria incluye el secundario completo, se recategoriza en dos jerarquías; desde el nivel "sin instrucción" hasta secundario incompleto se considera que no completó la escolarización obligatoria entonces se le asigna 1 y el resto, desde secundario completo hasta terciario o universitario completo (si completó la escolarización obligatoria), se le asigna 0.

Se prueba un segundo modelo eliminando los ítems obra social, edad y pareja y se incorpora el nivel de educación con dos categorías. Los resultados del análisis de bondad del ajuste se muestran en la Tabla 5.5. Se observa que las pruebas chi cuadrado de Pearson y de razón de verosimilitud siguen siendo significativas, pero sus respectivos valores disminuyeron. Por otro lado, los resultados de los criterios de información también disminuyen notablemente, es decir que, al eliminar los ítems antes mencionados, mejora el ajuste.

Tabla 5.5: Resultados de las pruebas de bondad del ajuste del modelo IRT parsimonioso para la variable latente “funcionamiento del jefe de hogar”

AIC	BIC	BIC _{aj}	χ^2 Pearson	G^2 razón de verosimilitud
7243,2	7307,7	7266,5	83,6 p-val=0,001	84 p-val=0,001

Fuente: Elaboración propia en base a los resultados dados por la ejecución del programa Mplus

En el proceso de estimación, el programa utiliza un estimador de máxima verosimilitud con errores estándares robustos a través de un algoritmo de integración numérico. La Tabla 5.6 muestra los parámetros de discriminación y dificultad estimados por el método descrito en el párrafo anterior y la correlación de cada ítem con la variable latente “funcionamiento del jefe de hogar”.

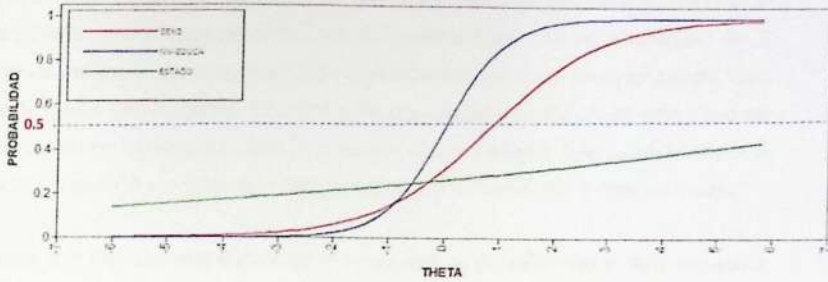
Tabla 5.6: Parámetros estimados por IRT para los ítems “nivel de educación”, “estado” y “sexo” correlacionados con “funcionamiento del jefe de hogar”

Item	$\hat{\lambda}_{i0}$ (dificultad)	$\hat{\lambda}_{i1}$ (discriminación)	$\hat{\pi}_i(I)$ (correlación)
Nivel de Educación	0,02	1,1	0,85
Estado	7,6	0,08	0,75
Sexo	0,8	0,6	0,97

Fuente: Elaboración propia en base a los resultados del programa Mplus

El ítem “sexo” es el de mayor correlación (0,97), le siguen “nivel de educación” (0,85) y “estado de ocupación” (0,75). Es decir que, esos ítems son los que más discriminan entre hogares con bienestar y sin bienestar en la dimensión “funcionamiento del jefe de hogar”. La Figura 5.8 muestra las Curvas Características de los Ítems dicotómicos.

Figura 5.8: Curvas características de los ítems dicotómicos “sexo”, “nivel de educación” y “estado de ocupación” altamente correlacionados con la variable latente “funcionamiento del jefe de hogar”

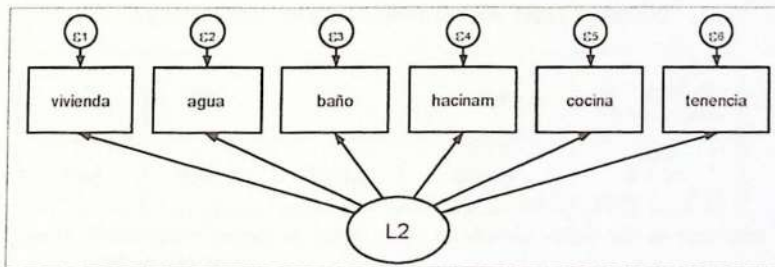


Fuente: Gráfico extraído del programa Mplus

5.3.2 Estimación de parámetros de los ítems para la variable latente “estar protegido”

La segunda dimensión que contribuye al bienestar es “estar protegido”, representada en la variable latente L2. Los ítems considerados para esta dimensión son: “tipo de vivienda”, “tiene agua”, “tiene baño”, “tiene cocina”, “tenencia” y “hacinamiento”. Se ajusta un Modelo de Rasgo Latente de dos parámetros y un factor con los ítems mencionados, los cuales son todos dicotómicos. El primer modelo propuesto contempla todos los indicadores como indica la Figura 5.9.

Figura 5.9: Modelo Estructural para la variable latente “estar protegido” considerando los ítems “tipo de vivienda”, “tiene agua”, “tiene baño”, “tiene cocina”, “tenencia” y “hacinamiento”



La Tabla 5.7 muestra los resultados de IRT de dos parámetros para cada uno de los seis ítems de "estar protegido". La columna marcada como "dificultad" se puede interpretar como que el hogar estaría privado de ese ítem. El criterio de severidad se establece en 3 desviaciones estándar de la media, es decir que se deben eliminar todos los ítems que tengan desviaciones estándar superior al valor absoluto de 3 desviaciones estándar. En la misma tabla se indica qué tan bien discrimina cada ítem entre los hogares encuestados. El criterio de discriminación usado en este caso es descartar todos los elementos cuya correlación con la variable latente sea inferior a 0,4. Los ítems que no cumplen con esta condición son "tenencia" y "hacinamiento".

Tabla 5.7: Valores estimados de los parámetros de dificultad y discriminación por la IRT de los ítems que modelan la variable latente "estar protegido"

Ítems	Dificultad	Discriminación
Tipo de vivienda	0,85	1,11
Tiene agua dentro o fuera de la vivienda	1,44	2,01
Tiene baño con/sin descarga de agua	1,94	4,54
Tiene cuarto exclusivo de cocina	1,94	1,00
Hacinamiento	1,79	0,05
Tenencia de la propiedad	2,69	0,23

Fuente: Elaboración propia en base a los resultados del programa Mplus

Tabla 5.8: Resultados de las pruebas de bondad del ajuste del modelo IRT parsimonioso para la variable latente "estar protegido"

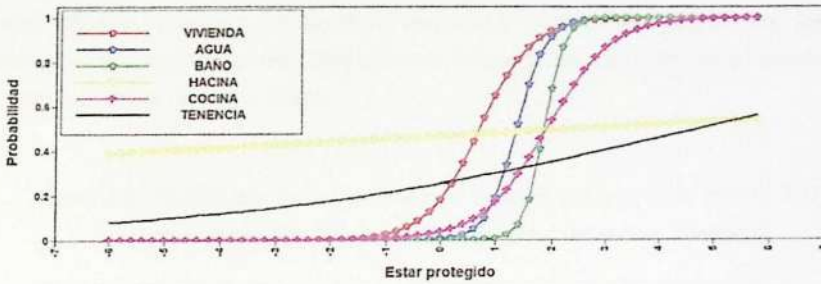
AIC	BIC	BIC _{aj}	χ^2 Pearson	G^2 razón de verosimilitud
2595	2646,3	2608,2	62,2 gl = 51 p-val = 0,09	63,5 gl = 51 p-val = 0,11

Fuente: Elaboración propia en base a los resultados dados por la ejecución del programa Mplus

La Tabla 5.8 muestra los resultados de las pruebas de bondad de ajuste. Los estadísticos chi cuadrado de Pearson y de razón de verosimilitud son no significativos, lo que indica un buen ajuste, sin embargo, los resultados de los parámetros antes mencionados y las curvas características (Figura 5.10), indican que se deberían descartar los ítems hacinamiento y tenencia.

La Figura 5.10 muestra que las curvas de los ítems hacinamiento y tenencia están muy suavizadas, lo cual indica que dichos ítems no discriminan entre los hogares con estar o no protegidos.

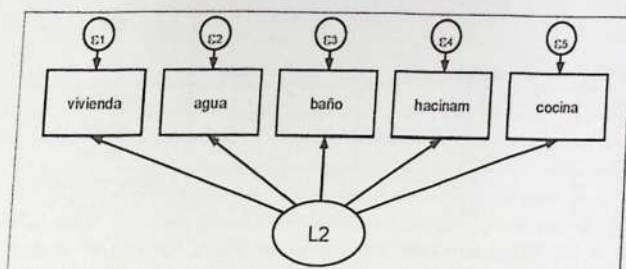
Figura 5.10: Curvas características de los ítems "tipo de vivienda", "tiene agua", "tiene baño", "tiene cocina", "hacinamiento" y "tenencia" de la variable latente "estar protegido"



Fuente: Gráfico extraído del programa Mplus

Se prueba otro modelo descartando sólo el ítem "tenencia", el ítem hacinamiento no se descarta porque es una necesidad básica y se analizan las pruebas de bondad de ajuste nuevamente, las mismas se muestran en la Tabla 5.9. La Figura 5.11 representa el nuevo modelo IRT sin el ítem "tenencia".

Figura 5.11: Modelo estructural para la variable latente "estar protegido" considerando los ítems "tipo de vivienda", "tiene agua", "tiene baño", "tiene cocina" y "hacinamiento"



Los valores de los estadísticos chi cuadrado de Pearson y de razón de verosimilitud disminuyeron notablemente respecto al modelo anterior. Además, los criterios de información también disminuyeron lo que indica que el segundo modelo (Figura 5.11) tiene un mejor ajuste.

Tabla 5.9: Resultados de las pruebas de bondad del ajuste del modelo IRT para la variable latente "estar protegido" sin el ítem "tenencia"

AIC	BIC	BIC _{aj}	χ^2 Pearson	G^2 razón de verosimilitud
1988,9	2031,6	1999,8	16,68 gl = 21 p-val = 0,73	16,78 gl = 21 p-val = 0,72

Fuente: Elaboración propia en base a los resultados dados por la ejecución del programa Mplus

La Tabla 5.10 muestra los parámetros de discriminación, de dificultad y la correlación de cada ítem con la variable latente "estar protegido". El ítem que más discrimina entre los hogares es el tener baño en el interior de la vivienda, le sigue en importancia el tener agua de red por cañerías internas, mientras que, el hacinamiento es la que menos discrimina.

Tabla 5.10: Parámetros y correlación estimados de los ítems "vivienda", "agua", "baño", "cocina" y "hacinamiento" que componen la variable latente "estar protegido"

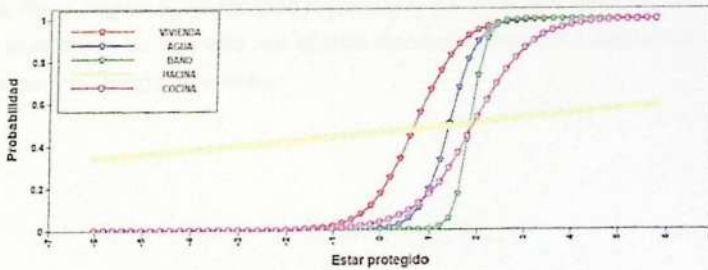
Ítem	$\hat{\lambda}_{i0}$ (dificultad)	$\hat{\lambda}_{i1}$ (discriminación)	$\hat{\pi}_i(L)$ (correlación)
Vivienda	0,82	1,22	0,77
Agua	1,47	1,84	0,88
Baño	1,93	5,37	0,98
Hacinamiento	0,90	0,09	0,48
Cocina	1,95	0,99	0,78

Fuente: Elaboración propia en base a los resultados dados por la ejecución del programa Mplus

El parámetro de discriminación en el hacinamiento es 0,09 (nuevamente menor a 0,4), por lo tanto es un ítem que no discrimina demasiado entre los hogares con bienestar y sin bienestar, sin embargo el parámetro de dificultad tiene un valor aceptable (menor al valor absoluto de tres).

La Figura 5.12 muestra las curvas características de respuestas de los ítems tipo de vivienda, tiene agua de red por cañerías internas, tiene baño en el interior de la vivienda, tiene cuarto exclusivo para cocina y hacinamiento.

Figura 5.12: Curvas características para los ítems que conforman la dimensión "estar protegido" con el modelo parsimonioso

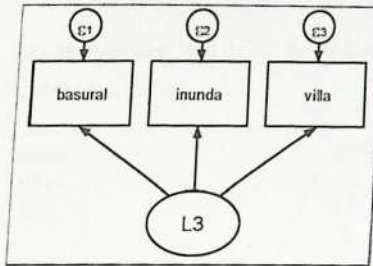


Fuente: Gráfico extraído del programa Mplus

5.3.3 Estimación de parámetros de los ítems para la variable latente "riesgo de entorno"

La tercera dimensión propuesta es el "riesgo de entorno" (L3) cuyos ítems son: "vivir cerca de basural", "vivir en zonas inundables" y "vivir en una villa"¹¹, tres causales de falta de bienestar en cuanto a seguridad ambiental y salubridad. Nuevamente se utiliza un Modelo de Rasgo Latente con dos parámetros y un factor, como indica la Figura 5.13.

Figura 5.13: Modelo estructural para la dimensión "riesgo de entorno" considerando los ítems "vivir cerca de basural", "vivir en zonas inundables" y "vivir en una villa"



El ajuste resultó muy satisfactorio, las pruebas chi cuadrado de Pearson y de razón de verosimilitud no son significativas. Por otro lado, las curvas características de los ítems (Figura 5.14) también indican que los ítems vivir cerca de basural, en zona inundable y en una villa son buenos discriminadores para determinar entre el grupo con y sin riesgo de entorno.

¹¹ Se considera "villa" a un asentamiento de viviendas en un terreno ajeno, que no reúnen las mínimas condiciones de comodidad, ni cuentan con los servicios básicos de agua, energía eléctrica y/o red de gas.

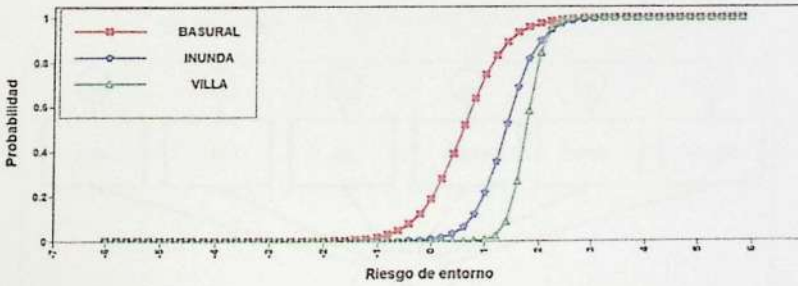
Tabla 5.11: Valores estimados de los parámetros de dificultad y discriminación por la IRT de los ítems que modelan la variable latente "riesgo de entorno"

Ítem	$\hat{\lambda}_{i0}$ (dificultad)	$\hat{\lambda}_{i1}$ (discriminación)	$\hat{\pi}_i(L)$ (correlación)
Basural	1,57	1,50	0,81
Inunda	1,37	2,03	0,89
Villa	1,75	3,93	0,97

Fuente: Elaboración propia según los resultados del programa Mplus

La Tabla 5.11 muestra los parámetros de dificultad y discriminación estimados por máxima verosimilitud. Se observa que los tres ítems están fuertemente correlacionados con la variable latente "riesgo de entorno".

Figura 5.14: Curvas características de los ítems que conforman la dimensión "riesgo de entorno"



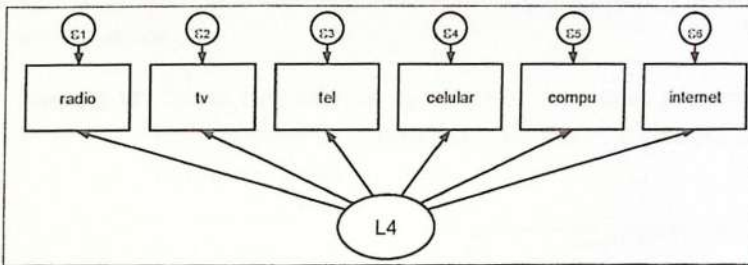
Fuente: Gráfico extraído del programa Mplus

Las curvas observadas en la Figura 5.14 se corroboran con los resultados mostrados en la Tabla 5.9.

5.3.4 Estimación de parámetros de los ítems para la variable latente "índice de bienes"

Las posesiones o bienes patrimoniales también son un buen indicio del bienestar de un hogar. En este sentido se considera un Modelo de Rasgo Latente cuyo factor es el "índice de bienes" con los ítems que indican posesión de: "radio", "televisores", "teléfono fijo", "celulares", "computadoras" (personales o notebooks) y "servicio de internet". El ítem *tenencia* (la familia es propietaria de la casa, del terreno donde ésta está asentada o de ambos) es un bien patrimonial, sin embargo, no tiene influencia en el índice de bienes porque éste se refiere a los bienes durables. Se estima un primer modelo IRT con todos ítems mencionados, el cual se muestra en la Figura 5.15.

Figura 5.15: Modelo Estructural para la variable latente "índice de bienes" considerando los ítems que indican posesión de "radio", "televisores", "teléfono fijo", "celulares", "computadoras" y "servicio de internet"



La Tabla 5.12 muestra los resultados de las pruebas de bondad de ajuste y los criterios de información. Se observa que, tanto las prueba chi cuadrado de Pearson como el chi cuadrado de razón de verosimilitud son significativas y sus valores muy altos, lo cual indica que el modelo no es bueno.

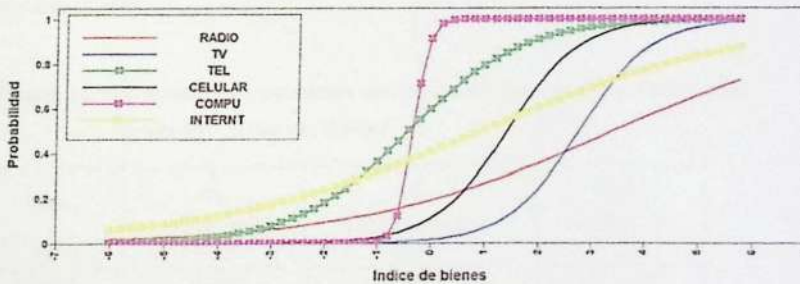
Tabla 5.12 Resultados de las pruebas de bondad del ajuste del modelo IRT para la variable latente "índice de bienes"

AIC	BIC	BIC _{aj}	χ^2 Pearson	G^2 razón de verosimilitud
6367,6	6427,1	6389	$\chi^2 = 152$ gl = 51 p-val = 0,01	$\chi^2 = 135,6$ gl = 51 p-val = 0,01

Fuente: Elaboración propia según los resultados del programa Mplus

La Figura 5.16 muestra las CCI para esta dimensión. Las curvas de los ítems "tiene radio" y "tiene internet" están muy suavizadas, por lo tanto, son los que menos discriminan entre los hogares con y sin bienestar en posesión de bienes, mientras que "tiene computadora" es el ítem que más discrimina. Los valores de los parámetros estimados, mostrados en la Tabla 5.13, indican que la dificultad del ítem "tiene radio" es superior al valor absoluto de 3, mientras que el parámetro discriminación de este mismo ítem y del ítem "tener internet" son menores a 0,4. Por otro lado, el parámetro discriminación del ítem "tener computadora" es bastante superior a los demás.

Figura 5.16: Curvas características de los ítems que indican posesión de "radio", "televisores", "teléfono fijo", "celulares", "computadoras" y "servicio de internet"



Fuente: Gráfico extraído del programa Mplus

Tabla 5.13: Parámetros estimados de los ítems que indican posesión de "radio", "televisores", "teléfono fijo", "celulares", "computadoras" y "servicio de internet"

Ítems	Dificultad	Discriminación
Tiene radio	3,45	0,25
Tiene TV	2,79	0,87
Tiene línea de Teléfono activa	-0,41	0,56
Tiene línea de celular activa	1,42	0,87
Tiene computadoras	-0,33	4,25
Tiene servicio de internet	0,97	0,23

Fuente: Elaboración propia según los resultados del programa Mplus

Se prueba otro modelo en el que se eliminan los ítems para los cuales el parámetro de dificultad es mayor al valor absoluto de tres desvíos estándar y el parámetro de discriminación es menor a 0,4. Con este criterio, los ítems que más discriminan entre las familias con y sin bienestar patrimonial son "tiene TV", "tiene una o más líneas activas de celular" y "tiene computadoras". Las pruebas chi cuadrado de bondad de ajuste resultaron no significativas y los criterios de información tuvieron una notable disminución, lo cual indica que el modelo ajustó bien al conjunto de acuerdo con el patrón de respuestas de la encuesta. La Tabla 5.14 muestra los resultados de los parámetros estimados y la correlación de cada ítem con la variable latente.

Tabla 5.14: Parámetros y correlación estimados de los ítems que modelan la dimensión "índice de bienes"

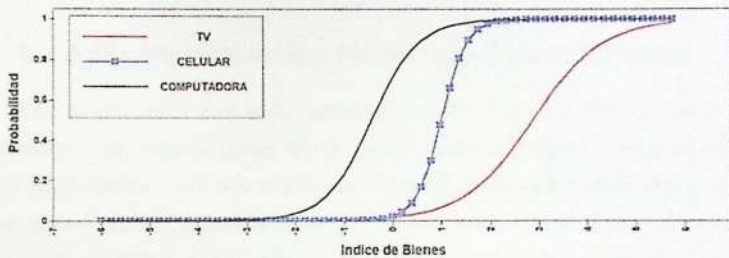
Item	$\hat{\lambda}_{i0}$ (dificultad)	$\hat{\lambda}_{i1}$ (discriminación)	$\hat{\pi}_i(L)$ (correlación)
TV	2.92	0.81	0.80
Celular	1.03	2.22	0.86
Computadora	-0.41	1.21	0.65

Fuente: Elaboración propia según los resultados del programa Mplus

De acuerdo con los parámetros estimados se advierte que el ítem que más discrimina entre hogares es tener líneas de celular activas (una o más), le sigue tener televisión y en menor medida poseer computadoras.

La Figura 5.17 muestra las curvas de respuesta de los ítems: tener televisores, celulares y computadoras en el hogar.

Figura 5.17: Curvas características de los ítems que conforman la dimensión "índice de bienes" para el modelo parsimonioso



Fuente: Gráfico extraído del programa Mplus

5.3.5 Comentarios

En base a la fuente de datos trabajada, los ítems que más discriminan para la dimensión "funcionamiento del jefe de hogar" son: sexo, el nivel de educación y el estado de ocupación. Para la dimensión "estar protegido" se consideran los ítems con mayor poder de discriminación a: tipo de vivienda, tener agua de red por cañerías internas, tener baño con descarga en el interior de la vivienda, que la vivienda tenga un cuarto exclusivo para la cocina y en menor medida el hacinamiento. También, vivir cerca de basural, en zona inundable y en una villa son indicadores determinantes de bienestar; estos ítems corresponden a la dimensión "riesgo de entorno". Por último, la dimensión "índice de bienes" es otro factor que brinda confort y es parte importante del bienestar de los hogares. Los ítems de esta dimensión con mayor poder de discriminación resultaron: tener TV, tener una o más líneas activas de celular y computadoras.

Asimismo, se realizó un análisis de estabilidad para determinar si los resultados obtenidos dependen o no de la muestra. Para ello se quitó aleatoriamente un 10%

de la muestra tres veces y se repitió el procedimiento, en todos los casos resultaron selectos (más correlacionados con la dimensión correspondiente), los mismos ítems.

Como IRT está basado en un modelo probabilístico formal y toma en cuenta el nivel de facilidad y capacidad discriminativa de cada ítem, la correlación entre ítems y la distribución original de los datos, es muy útil para determinar los ítems que más discriminan entre los hogares con bienestar y sin bienestar, además, esta propiedad permitiría diseñar mejor las preguntas de encuestas de calidad de vida.

5.4 Aproximaciones del Modelo de Clases Latentes

En las secciones anteriores se analizaron los modelos IRT ajustados a las dimensiones "funcionamiento del jefe de hogar", "estar protegido", "riesgo de entorno" e "índice de bienes" para determinar los ítems relevantes para cada una. A su vez, estas dimensiones conforman el índice multidimensional de bienestar propuesto (Figura 3.7 del Capítulo 3) y por el enfoque de derechos se considera que todas tienen la misma ponderación.

En esta sección se utilizan Modelos de Clases Latentes (LCM), para clasificar a los individuos (en este caso, hogares) resumiendo la información proporcionada por este conjunto de variables determinado por IRT. Para la dimensión "estar protegido" los indicadores son: tipo de vivienda, tiene agua, tiene baño, tiene cocina y hacinamiento; para la dimensión "riesgo de entorno" se consideran los indicadores: vivir cerca de basural, vivir en zona inundable y vivir en una villa. Por último, para la dimensión "índice de bienes" los indicadores seleccionados son: posesión de TV, teléfonos celulares y computadoras.

La dimensión "funcionamiento del jefe de hogar" no se considera en conjunto con las otras dimensiones porque los ítems se miden a nivel individual, mientras que el resto está referido al hogar.

La relación entre los indicadores (variables manifiestas) debe verificar el principio de independencia local. Dicho principio supone que, dentro de cada categoría de la variable latente, las variables observadas son estadísticamente independientes, es decir, las variables de la tabla de contingencia son condicionadamente independientes dada una clase determinada de la variable

latente. Toda la asociación observada entre las variables manifiestas está medida o explicada por las variables latentes. El no cumplimiento del principio trae como consecuencia que los estadísticos de ajuste del modelo (χ^2 y/o G^2)¹² sean demasiado grandes, los valores de las estimaciones de los parámetros del modelo se distorsionan, los errores estándar para los estimadores son grandes, y los estimadores de las diferencias de sus varianzas no son consistentes.

Otro de los supuestos del modelo es el que las clases latentes son internamente homogéneas, es decir, todos los miembros de una clase latente tienen la misma distribución de probabilidades con respecto a la variable latente: hogares de diferentes clases presentarán características diferentes.

Para los datos analizados, la matriz de información, teniendo en cuenta los ítems relevantes está formada por 14 variables categóricas (sexo, estado de ocupación, nivel de educación, tipo de vivienda, tiene agua de red, tiene baño dentro de la vivienda, hacinamiento, tiene cuarto de cocina, cerca de basural, en zona inundable, en villa, tiene TV, tiene computadora, tiene internet) sobre una muestra de 1058 hogares, donde cada fila de la matriz contiene el conjunto de respuestas dado para cada una de las variables (patrón de respuesta).

En cuanto a la asignación de los ítems, el análisis posterior está relacionado con su ubicación que pertenecen a una clase determinada. Los elementos se clasifican dentro de la clase latente más probable. Es decir que, para cada patrón de respuesta se analiza el conjunto de probabilidades, y se asigna el hogar a la clase latente en la cual la probabilidad es mayor.

5.4.1 Ajuste de un Modelo de Clases Latentes para la dimensión "estar protegido"

En una primera etapa consideramos los ítems referidos a las condiciones y comodidades de la vivienda, éstas son: el tipo de vivienda propiamente dicha, si tiene agua de red, baño con descarga de agua, cuarto de cocina y si presenta

¹² Estadísticos de las pruebas chi-cuadrado de Pearson y chi-cuadrado de razón de verosimilitud, (ver sección 4.6, del Capítulo 4).

hacinamiento (Figura 5.9). Las variables son todas dicotómicas, por lo tanto, sólo informan si las comodidades se satisfacen o no.

Estadísticamente, la hipótesis de homogeneidad de la población se corresponde con el modelo de una clase latente según el cual se plantea la independencia condicional de las variables. La aceptación de un modelo con más de una clase latente significa que la población puede dividirse en tantos grupos como clases tenga el modelo.

Para correcta elección del número de clases, el procedimiento es comparar modelos con distinto número de clases latentes. Para tal fin, se propuso una medida de *bootstrapping* condicional (prueba de Vuong-Lo-Mendell-Rubin, VLMR) en función de la diferencia en el valor del logaritmo de la verosimilitud (NYLUND *et al.*, 2007). A partir de considerar como verdadero un modelo con $k + 1$ clases latentes (menos restringido), se evalúa la diferencia en el ajuste entre dicho modelo y un modelo con k clases (más restringido), y si la diferencia resulta significativa se rechaza el modelo con menos clases latentes; en cambio, si la diferencia no es significativa se utiliza el modelo con menos clases latentes ya que es más parsimonioso.

En la Tabla 5.15 se pueden apreciar diferentes índices de ajuste para los tres modelos que se pusieron a prueba. Dado que actualmente no existe un único método ampliamente aceptado para comparar modelos con diferente número de clases latentes, es necesario combinar la teoría sustantiva que se utilizó como base para construir la prueba, así como los diferentes índices de ajuste diseñados para comparar modelos con un número creciente de clases latentes.

Tabla 5.15: Índices de ajuste para comparar modelos con 2, 3 y 4 clases latentes para la dimensión "estar protegido"

Modelo	L ²	p-val	Entropía	AIC	BIC	BIC _{AJ}	VLMR p-val	LMR p-val	RVB
2 clases	31.4	0.05	0.858	4063.2	4117.8	4082.9	0.01	0.01	0.01
3 clases	15.8	0.32	0.677	4059.7	4144.1	4090.1	0.14	0.15	0.04
4 clases	13.2	0.10	0.753	4069.5	4183.2	4110.2	0.16	0.17	1.00

Fuente: Elaboración propia en base a los resultados dados por la ejecución del programa Mplus

Observando la tabla anterior se determina la calidad de ajuste por el estadístico L^2 junto a su probabilidad asociada (p -*val*); también por los criterios de información de Akaike (AIC), Bayesiano (BIC) y el BIC ajustado. Por otro lado, las pruebas de razón de verosimilitud Vuong-Lo-Mendell-Rubin (VLMR) y Lo-Mendell-Rubin (LMR) resultan no significativas sugiriendo que dos clases son suficientes. En cuanto a la prueba de Razón de Verosimilitud paramétrica Bootstrapped (RVB), con la cual se compara si es preferible el modelo de dos o de tres clases, se determina que el modelo de tres clases no es necesario (p -*val* = 0,04, es significativa). Yang (2006), a través de ciertas pruebas de simulación, demostró que el BIC ajustado y LMR (Lo-Mendell-Rubin) son las mejores pruebas para determinar el número de clases. Finalmente, el valor de la entropía es mayor para el modelo de dos clases.

Habiendo determinado que el modelo de 2-clases es el más adecuado, se observan en la Tabla 5.16 los resultados. Cada fila es un ítem y las dos columnas son las probabilidades de responder afirmativamente a la carencia dado que el hogar pertenece a la clase.

Tabla 5.16: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando la dimensión "estar protegido"

	Clase 1 (hogares sin bienestar)	Clase 2 (hogares con bienestar)	Indicador
Ítem 1	0.903	0.199	Tipo de vivienda
Ítem 2	0.705	0.017	Agua de red
Ítem 3	0.288	0.001	Baño con descarga
Ítem 4	0.507	0.459	Hacinamiento
Ítem 5	0.459	0.035	Hab. exclus. p/cocinar

Fuente: Elaboración propia en base a los resultados del programa Mplus

Los hogares pertenecientes a la clase 1 y a la clase 2 tienen un 90,3%, y 19,9% respectivamente, de responder que su vivienda es inadecuada. Los hogares de la clase 1 tienen 70,5% de probabilidad de responder que carecen de agua de red, mientras que las familias pertenecientes a las clases 2 tiene 1,7% de probabilidades de responder que no poseen agua de red. En cuanto al baño con descarga y dentro

de la vivienda sólo la clase 1 tiene una probabilidad de 28,8% de carecer de este indicador básico. El ítem hacinamiento tiene casi el mismo nivel de respuesta para las dos clases, 50,7% y 45,9% lo cual podría explicarse por el déficit habitacional existente para todas las clases, (no es un buen discriminador de clase). Luego, el no poseer una habitación exclusiva para cocinar, el 45,9% pertenece a la clase 1, y sólo el 3,5% a la clase 2. Con estos resultados, se puede etiquetar a la clase 1 como "hogares sin bienestar" y a la clase 2 como "hogares con bienestar" en la dimensión "estar protegido".

La Tabla 5.17(a) muestra las proporciones para los patrones de clase latente basados en probabilidades posteriores estimadas y la Tabla 5.17(b) muestra la clasificación de los hogares basados en su mayor probabilidad de pertenecer a la clase latente. Estas son dos maneras de estimar el tamaño de las clases¹³.

Tabla 5.17: Proporción de hogares que presentan situación con bienestar y sin bienestar en la dimensión "estar protegido"

(a)			(b)		
Clases	Cantidad	Proporción	Clases	Cantidad	Proporción
Sin bienestar	139	0,13	Sin bienestar	132	0,13
Con bienestar	919	0,87	Con bienestar	926	0,87

Fuente: Elaboración propia en base a los resultados del programa Mplus

Las clases latentes se etiquetan como "con bienestar" y "sin bienestar". El 87% de los hogares de la muestra tienen bienestar y el 13% carece del mismo, de acuerdo con los indicadores tipo de vivienda, tener agua de red, baño con descarga, no sufrir hacinamiento y tener una habitación exclusiva para cocinar.

Por otro lado, es posible cometer errores al asignar los hogares a una clase latente determinada, lo que se denomina "error de mala clasificación". Para estimar dicho error, se calcula su proporción esperada asignando cada hogar a la clase de mayor probabilidad. Esta clasificación se basa en la probabilidad posterior de

¹³ Todas las probabilidades son estimadas con el programa Mplus.

pertenencia a una clase latente K , dado un patrón de respuestas determinado, que recurriendo al Teorema de Bayes se expresa como:

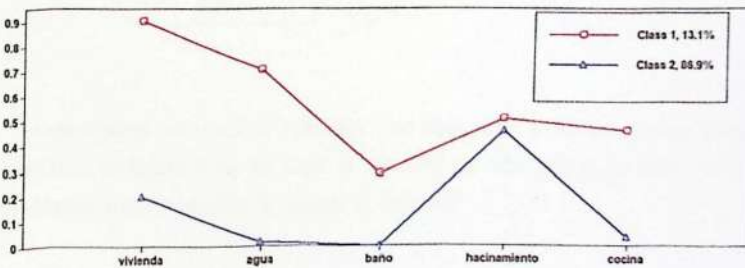
$$P(k/x_i) = \frac{P(x_i/k)\pi_k}{\sum_{k=1}^K P(x_i/k)\pi_k} \quad (5.2)$$

donde k es una determinada clase, x_i es la respuesta afirmativa ($X = 1$), respuesta negativa ($X = 0$); π_k es la probabilidad de la clase latente.

Luego se clasifican los casos que presentan un patrón de respuestas determinado en la clase latente con mayor probabilidad posterior, lo cual se conoce con el nombre de asignación modal. También se puede considerar un valor crítico para la clasificación, por ejemplo, clasificar sólo los casos que presentan una probabilidad de pertenencia a una clase determinada mayor a 0,75. En cuanto al error de clasificación, es decir, cómo se asignan apropiadamente los casos con cierto valor de respuesta a una clase latente determinada, los valores más cercanos a cero son indicadores de una mejor clasificación. En este caso la proporción es de 0,1025, lo cual indica que, se esperan errores de asignación pequeños.

La Figura 5.18 muestra las probabilidades condicionales de las clases latentes. El 13% de la muestra pertenece a la clase latente etiquetada como "sin bienestar".

Figura 5.18: Probabilidades condicionales estimadas para cada clase latente en la dimensión "estar protegido"

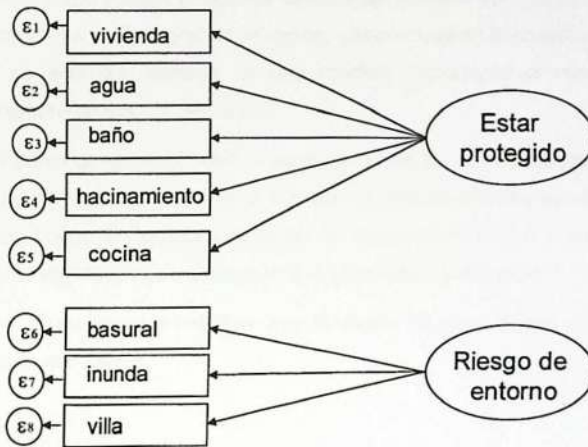


Fuente: Gráfico extraído del programa Mplus

5.4.2 Ajuste de un Modelo de Clases Latentes para las dimensiones "estar protegido" y "riesgo de entorno"

Se consideran en este caso dos dimensiones juntas, "estar protegido" y "riesgo de entorno". Se desea determinar cuántas categorías de bienestar presenta la muestra y finalmente, realizar un análisis integral de la situación de bienestar. Se agregan los ítems vivir cerca de basural, vivir en zona inundable y vivir en una villa. La Figura 5.19 muestra el modelo SEM propuesto.

Figura 5.19: Modelo Estructural para las dimensiones "estar protegido" y "riesgo de entorno" considerando los ítems tipo de vivienda, tiene agua, tiene baño, tiene cocina, hacinamiento, vivir cerca de basural, vivir en zona inundable y vivir en una villa



Nuevamente se prueban modelos con dos, tres y cuatro clases latentes. Observando la Tabla 5.18 se elige el modelo de dos clases latentes para las dimensiones "estar protegido" y "riesgo de entorno".

Tabla 5.18: Índices de ajuste para comparar modelos con 2, 3 y 4 clases latentes para las dimensiones "estar protegido" y "riesgo de entorno"

Modelo	L ²	p-val	Entropía	AIC	BIC	BIC _{AJ}	VLMLR p-val	LMR p-val	RVB
2 clases	367	0.43	0.898	5309.6	5403.9	5343.6	0.01	0.01	0.01
3 clases	250	0.99	0.849	5198.2	5342.1	5250.0	0.99	0.99	0.01
4 clases	186	0.10	0.856	5154.4	5348.0	5224.1	0.60	0.60	1.00

Fuente: Elaboración propia en base a los resultados del programa Mplus

Para determinar el número de clases latentes, generalmente, se consideran los criterios de información porque pequeñas diferencias pueden ser estadísticamente significativas dado una medida grande de casos, como ocurre en la muestra utilizada. El parámetro de entropía también es una medida importante al momento de establecer el número adecuado de clases.

Con los valores de ajuste obtenidos, se elige el modelo dos clases, teniendo en cuenta que tiene mayor valor de entropía, aunque en éste los criterios de información no son menores. Luego, las pruebas de razón de verosimilitud VMLR y verosimilitud ajustada LMR, no significativas indican que dos clases son suficientes.

La Tabla 5.19 muestra los resultados en la escala de probabilidad de los ítems para el modelo de dos clases.

Tabla 5.19: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando la dimensión "estar protegido" y "riesgo de entorno"

	Clase 1	Clase 2	Indicador
Item 1	0.962	0.188	Tipo de vivienda
Item 2	0.611	0.029	Agua de red
Item 3	0.283	0.000	Baño con descarga
Item 4	0.456	0.466	Hacinamiento
Item 5	0.393	0.044	Hab. exclus. p/cocinar
Item 6	0.233	0.034	Vivir cerca de basural
Item 7	0.363	0.031	Zona inundable
Item 8	0.371	0.002	Vivir en una villa

Fuente: Elaboración propia en base a los resultados del programa Mplus

Se observa que los hogares pertenecientes a la clase 2 tienen menores probabilidades de responder positivamente a sufrir carencias en todos los ítems, excepto en el ítem hacinamiento¹⁴.

La proporción esperada de los errores de clasificación al asignar los hogares a cada clase es muy pequeña, aproximadamente 0,099.

Al considerar indicadores referidos a las comodidades y condiciones de la vivienda juntamente con los referidos al riesgo de entorno, nuevamente quedan determinadas 2 clases. En la tabla anterior se observa que los hogares se pueden clasificar en: "sin bienestar" (clase 1) y "con bienestar" (clase 2). Esta última categoría concentra el mayor porcentaje de los hogares de la muestra (88%), mientras que el 12% de los hogares sufre toda clase de carencias, excepto hacinamiento que tiene similar probabilidad en ambas clases.

La Tabla 5.20(a) muestra las proporciones para los patrones de clase latente basados en probabilidades posteriores estimadas y la Tabla 5.20(b) muestra la

¹⁴ Se considera que hay hacinamiento cuando hay tres o más personas que duermen en una misma habitación

clasificación de los hogares basados en su mayor probabilidad de pertenecer a la clase latente.

Tabla 5.20: Clasificación y proporción de los hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para las dimensiones "estar protegido" y "riesgo de entorno"

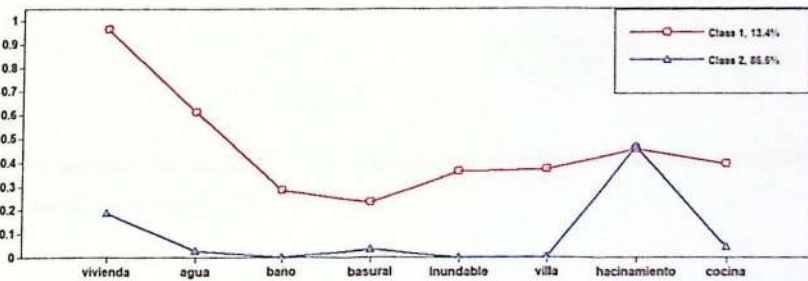
(a)			(b)		
Clase	Cantidad	Proporción	Clase	Cantidad	Proporción
Sin bienestar	141	0.13	Sin bienestar	130	0.12
Con Bienestar	916	0.87	Con Bienestar	928	0.88

Fuente: Elaboración propia en base a los resultados del programa Mplus

Comparando las Tablas 5.17 y 5.20 se observa que la proporción de hogares con bienestar y sin bienestar mantiene similar distribución al analizar simultáneamente las dimensiones "estar protegido" y "riesgo de entorno"

La Figura 5.20 muestra las probabilidades condicionales estimadas de cada clase. El porcentaje de la muestra que pertenece a la situación "sin bienestar", es aproximadamente 13,4%.

Figura 5.20: Probabilidades condicionales de las clases latentes "con bienestar" y "sin bienestar" teniendo en cuenta las dimensiones "estar protegido" y "riesgo de entorno" simultáneamente



Fuente: Gráfico extraído del programa Mplus

5.4.3 Ajuste de un Modelo de Clases Latentes para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"

En esta sección se describe el ajuste de un modelo considerando las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes", teniendo en cuenta los ítems de cada una que resultaron significativos en los modelos de IRT. Estos son: "tipo de vivienda", "tiene agua de red", "tiene baño", "sufre hacinamiento", "tiene cocina", "está ubicada cerca de "basural", "en zona inundable" en una "villa", "tiene TV", "tiene línea/s activa/s de celulares" y "tiene computadora (portátil o PC)". La Figura 5.21 muestra el modelo SEM que representa las tres dimensiones con sus respectivos ítems.

Figura 5.21: Modelo Estructural para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"



Se probaron modelos con 2, 3 y 4 clases, cuyas pruebas de bondad de ajuste se muestran en la Tabla 5.21.

Tabla 5.21: Índices de ajuste para comparar modelos con 2, 3 y 4 clases latentes para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"

Modelo	L ²	p-val	Entropía	AIC	BIC	BIC _{AJ}	VMLR p-val	LMR p-val	RVB
2 clases	716,7	0.99	0.886	7958,4	8082,5	8003,1	0.00	0.00	0.00
3 clases	567,7	0.99	0.939	7815,8	8004,5	7883,8	0.00	0.00	0.00
4 clases	440,4	0.99	0.899	7705,7	7958,9	7796,9	0.82	0.82	0.00

Fuente: Elaboración propia en base a los resultados del programa Mplus

Observando los índices de la Tabla 5.21 el modelo más conveniente es el de tres clases porque presenta el valor más alto de entropía. Las pruebas de razón de verosimilitud VLMR y verosimilitud ajustada LMR, son significativas y la prueba de Razón de Verosimilitud paramétrica Bootstrapped es muy significativa ($p\text{-value}=0.001$), quiere decir que el modelo de tres clases es preferible al de dos clases, además, también los criterios de información son menores. Los índices de las pruebas de verosimilitud y de los criterios de información para el modelo de cuatro clases son menores que en los otros modelos. Sin embargo, el máximo valor del logaritmo de verosimilitud no se replica, lo cual significa que el máximo alcanzado no es global, por lo tanto, este modelo se descarta.

La Tabla 5.22 muestra los resultados en la escala de probabilidad de los ítems de las dimensiones "estar protegidos", de "riesgo de entorno" y de "índice de bienes" para el modelo de tres clases latentes. Los hogares con mayores probabilidades de responder afirmativamente que carece del correspondiente ítem, en general se encuentran en la clase 1, por lo tanto, se etiqueta a esta clase como "sin bienestar". Los hogares de la clase 2 presentan, en general, probabilidades bajas de respuesta positiva a la carencia en todos los ítems, entonces se etiqueta como "con bienestar" y a la clase 3, que presenta algunas probabilidades más altas con respecto a las de la clase 2 y más bajas con respecto a los de la clase 1, se denomina "con bienestar medio".

Tabla 5.22: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"

	Clase 1	Clase 2	Clase 3	Ítem
Item 1	0.955	0.126	0.240	Tipo de vivienda
Item 2	0.624	0.024	0.034	Agua de red
Item 3	0.292	0.001	0.001	Baño con descarga
Item 7	0.447	0.338	0.526	Hacinamiento
Item 8	0.410	0.024	0.057	Hab. exclus. p/cocinar
Item 4	0.248	0.054	0.017	Vivir cerca de basural
Item 5	0.369	0.031	0.033	Zona inundable
Item 6	0.375	0.007	0.001	Vivir en una villa
Item9	0.115	0.010	0.040	Tener TV
Item10	0.328	0.021	0.252	Línea de celular
Item11	0.917	0.001	0.998	Computadoras

Fuente: Elaboración propia en base a los resultados del programa Mplus

Las probabilidades posteriores estimadas, es decir la probabilidad de que cada hogar por separado pertenezca a una clase latente y la clasificación de los hogares basados en su mayor probabilidad de pertenecer a la clase latente, se muestran en las tablas 5.23(a) y 5.23(b), respectivamente.

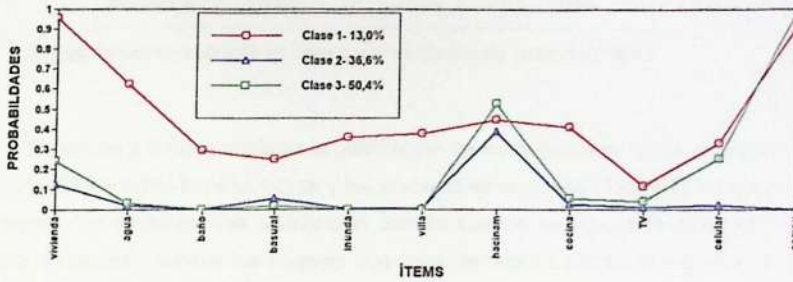
Tabla 5.23: Clasificación y proporción de los hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"

(a)			(b)		
Clase	Cantidad	Proporción	Clase	Cantidad	Proporción
1	137	0,13	1	128	0.13
2	533	0,51	2	538	0.51
3	387	0,37	3	392	0.37

Fuente: Elaboración propia en base a los resultados del programa Mplus

Considerando las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes", se observa que la mitad (51%) de los hogares de la muestra goza de bienestar, el 37% tiene bienestar medio y el 13% carece de bienestar. En este caso, la proporción esperada de los errores de clasificación es de 0,1474. La Figura 5.23 muestra las probabilidades condicionales estimadas de cada clase.

Figura 5.22: Probabilidades condicionales de clase latente para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes"



5.4.4 Análisis comparativo de resultados

Una vez expuestos los resultados obtenidos en el ajuste de los Modelos de Clases Latentes para una dimensión y luego los modelos ajustados al agregar las otras dimensiones, se presenta el análisis de los resultados en forma comparativa entre los modelos con la dimensión "estar protegido" solamente (Modelo 1), con dos dimensiones agregando "riesgo de entorno" (Modelo 2) y finalmente incorporando la dimensión "índice de bienes" (Modelo 3). En los dos primeros casos, el modelo que mejor ajustó fue el de 2 clases, las cuales se etiquetaron como "con bienestar" y "sin bienestar" y se obtuvo la distribución de los hogares de la muestra en la clase correspondiente de acuerdo con las probabilidades estimadas. En el tercer caso, aparece una tercera clase latente que divide a la muestra de manera diferente, es decir, "con bienestar", "con bienestar medio" y "sin bienestar"

En la Tabla 5.24 se muestra la distribución de los hogares en las clases latentes luego de ajustar los tres modelos mencionados en el párrafo anterior.

Tabla 5.24: Porcentaje de hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para las dimensiones "estar protegido", "riesgo de entorno" e "índice de bienes" de tres diferentes modelos

	Con bienestar	Con bienestar medio	Sin bienestar
Modelo 1	87%	-	13%
Modelo 2	88%	-	12%
Modelo 3	51%	37%	13%

Fuente: Elaboración propia en base a los resultados del programa Mplus

En los dos primeros modelos la distribución de los hogares no se modifica, la mayoría (87% y 88%) tiene bienestar y los porcentajes restantes (13% y 12%) sufre carencias. Sin embargo, esa distribución cambia cuando se agrega la dimensión "índice de bienes", aunque los hogares con bienestar siguen siendo la mayoría se reducen a 51%. Se manifiesta otra clase latente con bienestar medio que concentra el 37% de los hogares, pero el porcentaje de hogares sin bienestar no se reduce; permanece en el mismo nivel (13%).

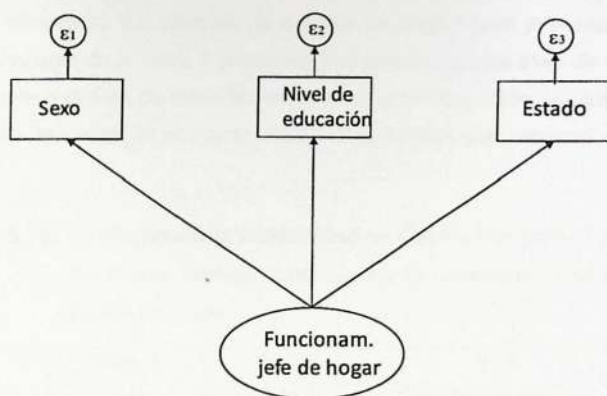
Con estos resultados se podría afirmar que la medición de bienes que posee el hogar ofrece otra evaluación del estado socioeconómico, aunque esta medida, en forma de activos, sea utilizada con menos frecuencia.

5.4.5 Ajuste de un Modelo de Clases Latentes para la dimensión "funcionamiento del jefe de hogar"

En esta sección se presentan los resultados de ajustar un LCM para el "funcionamiento del jefe de hogar". Como se mencionó al comienzo de la Sección 5.4, esta dimensión no se considera en conjunto con las otras dimensiones porque los ítems nivel de educación, sexo y estado de ocupación, se miden a nivel individual, mientras que el resto está referido al hogar. El modelo adecuado para esta situación es el Modelo de Clases Latentes Multiniveles (MLCM), cuyos conceptos teóricos no

serán tratados en esta tesis. La Figura 5.23 muestra el modelo SEM que representa el "funcionamiento del jefe de hogar".

Figura 5.23: Modelo Estructural parsimonioso para la dimensión "funcionamiento del jefe de hogar"



La Tabla 5.25 muestra los resultados de las pruebas de bondad de ajuste para los modelos de 2 y 3 clases. También se probó un modelo de cuatro clases, pero no se pudieron obtener resultados porque no hubo convergencia.

Tabla 5.25: Índices de ajuste para comparar modelos con 2 y 3 clases latentes para la dimensión "funcionamiento del jefe de hogar"

Modelo	Entropía	p-val	AIC	BIC	BIC _{AJ}	VMLR p-val	LMR p-val	RVB
2 clases	0,85	0,01	3375,3	3410	3387,8	0,01	0,01	0,01
3 clases	0,47	0,01	3382,5	3437,1	3402,2	0,38	0,39	0,99

Fuente: Elaboración propia en base a los resultados del programa Mplus

De acuerdo con los resultados mostrados en la Tabla 5.25, el modelo más adecuado es de dos clases. El ajuste del modelo de tres clases presenta el inconveniente que los máximos obtenidos en el proceso son locales, lo cual lleva a

que las estimaciones no sean correctas. Por otro lado, las pruebas de razón de verosimilitud VLMR y verosimilitud ajustada LMR son no significativas para el modelo de tres clases, lo cual indica que el modelo de dos clases es preferible al de tres clases.

La Tabla 5.26 muestra los resultados en la escala de probabilidades de los ítems de la dimensión "funcionamiento del jefe de hogar" para el modelo de dos clases. Los hogares de la clase 1 presentan probabilidades más altas de responder afirmativamente a que es de sexo femenino y no tiene ocupación. En cuanto a las categorías del ítem nivel de ocupación, las probabilidades son similares en ambas clases.

Tabla 5.26: Distribuciones de probabilidad de Clase e Ítem para el modelo de dos clases latentes considerando la dimensión "funcionamiento del jefe de hogar"

	Clase 1	Clase 2	Ítem
Ítem 1	0,90	0,11	Sexo del jefe de hogar
Ítem 2	0,87	0,01	Nivel de educación
Ítem 3	0,21	0,81	Estado de ocupación

Fuente: Elaboración propia en base a los resultados del programa Mplus

La probabilidad (probabilidades posteriores estimadas) de que cada hogar por separado pertenezca a una clase latente y la clasificación de los hogares basados en su mayor probabilidad de pertenecer a la clase latente, se muestran en las Tablas 5.27(a) y 5.27(b), respectivamente.

Tabla 5.27: Clasificación y proporción de los hogares pertenecientes a cada clase latente, de acuerdo con las probabilidades posteriores estimadas para la dimensión "funcionamiento del jefe de hogar"

(a)			(b)		
Clase	Cantidad	Proporción	Clase	Cantidad	Proporción
1	773	0,73	1	795	0,75
2	285	0,27	2	263	0,25

Fuente: Elaboración propia en base a los resultados del programa Mplus

En base al enfoque de derechos¹⁵ se considera que cada dimensión se pondera de igual manera, entonces, para que la dimensión "funcionamiento del jefe de hogar" no quede excluida de la composición del indicador de bienestar final, se adopta un criterio similar al de las NBI. De acuerdo con este criterio, basta que en una sola dimensión el hogar carezca de bienestar para que se considere "sin bienestar".

5.5. Resultados de los Modelos de Variables Latentes aplicados. Comparaciones con los resultados de métodos tradicionales

En esta sección se comparan los resultados de las diferentes metodologías utilizadas para medir pobreza utilizando los datos de la Encuesta Anual de Hogares Urbanos (EAHU) del año 2011, descrita en la sección 5.2.

La medición de la pobreza por ingresos (línea de pobreza) es el método más utilizado a nivel mundial, a pesar de sus limitaciones. Es una metodología unidimensional y las fuentes de información para realizar los cálculos son las encuestas de hogares (en Argentina, particularmente, se usa la EPH). La Tabla 5.28 muestra los valores absolutos y porcentajes obtenidos por el método indirecto (datos inferidos para la provincia de Tucumán)¹⁶ y con los datos muestrales.

Tabla 5.28: Hogares bajo las líneas de pobreza e indigencia en el tercer trimestre de 2011

Datos inferidos			Datos muestrales		
Nº hogares indigentes	Nº hogares pobres	Total hogares	Nº hogares indigentes	Nº hogares pobres	Total hogares
15.152	82.477	338.621	49	281	1058
4.5%	24.4%		4.6%	26.6%	

Fuente: Elaboración propia en base a la EAHU 2011, publicada por INDEC (INDEC, 2011)

¹⁵ Desarrollo Humano Sostenible y Desarrollo de Capacidades en un contexto institucional democrático (PNUD) (PNUD, 2008).

¹⁶ Los resultados se obtuvieron con los datos ponderados.

La Tabla 5.29 muestra la distribución de los hogares pobres y no pobres (por el Método de la Línea de Pobreza), en las columnas y los hogares sin bienestar y con bienestar (por el Método Multidimensional) en las filas. Se comparan los resultados sólo a nivel muestral.

Tabla 5.29: Número de hogares con bienestar y sin bienestar según el método multidimensional versus el número de hogares bajo la línea de pobreza y sobre la línea de pobreza

Unidimensional Multidimensional	Hogares pobres (Línea de Pobreza)	Hogares no pobres (Línea de Pobreza)
Hogares s/bienestar	168	400
Hogares c/ bienestar	162	328

Fuente: Elaboración propia en base a los resultados obtenidos por los Métodos Multidimensional y Línea de Pobreza (Método unidimensional)

Se observa que ambos Métodos coinciden en que son "pobres" o "sin bienestar", sólo en 168 hogares de la muestra y 328 hogares coinciden en que "no son pobres" y "tienen bienestar". Por otro lado, el número de hogares considerados "no pobres" por el criterio de la Línea de Pobreza (LP) y "sin bienestar" multidimensionalmente, es de 400, el más alto de las cuatro combinaciones y la menor cantidad de hogares se encuentran en la combinación inversa a ésta. Cabe destacar, que se está comparando un método multidimensional con otro unidimensional. Los resultados de la Tabla 5.29, indican que el primero es más eficiente para clasificar los hogares, según sus condiciones, ya que hay 232 hogares que no son pobres desde el punto de vista de ingresos, pero sí lo son en otras dimensiones.

El indicador, "carencia" también es de tipo unidimensional y se basa en ingresos, como el de LP. Se calcula para los aglomerados urbanos de la provincia de Tucumán como alternativa a las líneas de pobreza e indigencia durante algunos períodos. Sin embargo, los resultados no fueron aceptados por el INDEC y sólo se

publicaron en el SEP¹⁷. Con el propósito de realizar algunas comparaciones, se calcula el indicador de carencia con los datos de la EAHU 2011 para los aglomerados urbanos de Tucumán. La Tabla 5.30 muestra los resultados.

Tabla 5.30: Número de hogares con bienestar y sin bienestar según el método multidimensional versus el número de hogares con y sin carencias

Unidimensional	Hogares con carencias	Hogares sin carencias
Multidimensional		
Hogares s/bienestar	102	466
Hogares c/ bienestar	93	397

Fuente: Elaboración propia en base a los resultados obtenidos por los Métodos Multidimensional y de Carencia (Método unidimensional)

La clasificación de los hogares con el Método Multidimensional y con el Método de Carencia difiere muy poco de la obtenida en el caso anterior.

El Índice de Privación Material de los Hogares (IPMH, descrito en la sección 1.3.1), adoptado en Argentina para computar pobreza multidimensional, comprende dos dimensiones y se calcula en base a datos censales. Las dos únicas mediciones realizadas, hasta el momento, por el INDEC, se obtuvieron luego de efectuarse los censos de Población y Vivienda de los años 2001 y 2010. Empero, el censo 2010 presentó diferencias en la forma de medir ciertas variables, involucradas en el cálculo del IPMH, por lo tanto, resultó muy difícil comparar con el índice calculado con los datos del censo 2001. Por ejemplo, la variable "condición de actividad" es una componente importante en el cálculo de CAPECO. En el censo 2010 la misma fue categorizada como: "ocupado", "desocupado" e "inactivo", mientras que en el Censo 2001 estaba desagregada en más categorías: "sólo trabaja"; "trabaja/estudia"; "trabaja/es jubilado", "trabaja/estudia/es jubilado", "sólo busca trabajo", "busca trabajo/estudia", "busca trabajo/es jubilado"; "busca trabajo/

¹⁷ SEP (Sistema Estadístico Provincial) del Gobierno de la Provincia de Tucumán.
<http://sep.tucuman.gob.ar/>

estudia/es jubilado"; "solo es jubilado"; "es jubilado/estudia"; "sólo estudia"; "otra situación".

Para poder comparar entre un censo y otro, en la Dirección de Estadística de Tucumán se propuso agrupar las categorías del censo 2001 de la siguiente manera: i) se sumaron las categorías que "tiene trabajo" (4 categorías) para equiparar a ocupado; ii) se sumaron las que tienen "busca trabajo" (4 categorías) para que sea equivalente a desocupado; por último iii) se agrupan las categorías que contiene a "sólo es jubilado", "sólo estudia" y combinaciones de estas últimas para asemejar a inactivo. En cuanto al Valor de los Años de Escolaridad Invertidos en el Mercado Laboral (VAE), también debió ser transformado a través de una función cuadrática para hacer comparables los resultados¹⁸. Por otro lado, se realizó el cálculo con la asistencia del INDEC, pero los resultados obtenidos no fueron "rigurosos".

Asimismo, la metodología multidimensional propuesta no es comparable con el IPMH, ya que las fuentes de información para elaborar cada uno son diferentes. Por esta misma razón, tampoco es comparable con el método directo o NBI.

Las diferentes metodologías de medición de pobreza involucran algunos ítems o dimensiones que se pueden comparar porque son similares, como vivienda, condiciones sanitarias, hacinamiento y capacidad económica. La Tabla 5.31 compara los resultados obtenidos usando las metodologías de las Necesidades Básicas Insatisfechas (NBI), el Índice de Privación Material de los Hogares (IPMH) y Modelos de Clases Latentes (MCL) en las dimensiones señaladas.

En cuanto a los métodos unidimensionales, como "Línea de Pobreza" y "Carencia" sólo se podría comparar con la dimensión relacionada a "subsistencia" o "capacidad económica de los hogares".

La Tabla 5.31 muestra el porcentaje de hogares que sufren diferentes tipos de privaciones. El Método de Clases Latentes, que tiene en cuenta cuatro dimensiones bien definidas ("funcionamiento del jefe de hogar", "estar protegido", "riesgo de entorno" e "índice de bienes"), indica que el 13,4% de los hogares carece de bienestar en una o más de estas dimensiones. El Método de la NBI desagregado indica que el

¹⁸ Sistema Estadístico Provincial (SEP). <http://sep.tucuman.gob.ar/group/pobreza-e-indigencia>

3,9% de los hogares no posee vivienda digna, el 4,5% carece de condiciones sanitarias adecuadas, el 5,6% sufre hacinamiento y el 1,4% no tiene capacidad económica. Considerando todos estos aspectos juntos, el 15,4% de los hogares sufre de al menos uno. La metodología de las NBI tiene características multidimensionales, pero no contempla la dimensión referida a los bienes durables.

Tabla 5.31: Porcentajes de hogares que sufren privaciones calculados con las metodologías unidimensionales y multidimensionales

	Condiciones de vivienda	Condiciones sanitarias	Hacinamiento	Capacidad económica	Bienes durables	Total
NBI ^{19(*)}	3,9%	4,5%	5,6%	1,4%		15,4%
IPMH (*)		14,4%				
CL	13,4%					
LP				24,5%		
Carencia				16,4%		

Fuente: Elaboración propia en base a datos extraídos del SEP

En cuanto a comparar el Método de Clases Latentes con la otra metodología multidimensional por la que se obtiene el Índice de Privación Material de los Hogares (IPMH), la cual también se mide con datos censales, se observa una diferencia de un punto porcentual. Cabe aclarar que el IPMH sólo contempla dos dimensiones y cada dimensión se construye con información muy agregada. Este indicador tiene una desventaja importante, si se cambian las definiciones de las variables de un censo a otro los resultados no pueden ser comparables.

Por otro lado, se observa que, por los métodos unidimensionales, las mediciones por ingresos (línea de pobreza) y carencia dan resultados muy dispares, 24,5% y 16,4% respectivamente. Cabe destacar que, los datos utilizados para obtener estas mediciones provienen de las Encuestas de Hogares, donde existe una pregunta sobre los "ingresos individuales y totales por hogar". Sin embargo, no siempre es respondida con veracidad, también presenta muchos datos faltantes y/o

^{19 (*)} Fuente: Sistema Estadístico Provincial (SEP), <http://sep.tucuman.gob.ar/group/pobreza-e-indigencia>

puede acarrear los errores propios de imputación. Por tales motivos la medición es incompleta o con más dudas que certezas.

5.6 Análisis de diferencias entre modelos

Las curvas ROC son herramientas utilizadas en la evaluación del rendimiento de índices binarios. Indican de manera visual la relación entre la precisión y la sensibilidad de un modelo, a la vez que sirven para comparar el rendimiento de distintos modelos de clasificación.

Generalmente, la evaluación se expresa como sensibilidad y especificidad diagnósticas. Cuando se utiliza una prueba dicotómica (una cuyos resultados se puedan interpretar directamente como positivos o negativos), la sensibilidad es la probabilidad de clasificar correctamente a un individuo cuyo estado real sea el definido como positivo respecto a la condición que estudia la prueba, razón por la que también es denominada Fracción de Verdaderos Positivos (FVP). La especificidad es la probabilidad de clasificar correctamente a un individuo cuyo estado real sea el definido como negativo. Es igual al resultado de restar a uno la Fracción de Falsos Positivos (FFP).

En el caso particular de clasificar la muestra de hogares en una tabla de contingencia por el resultado de aplicar Modelos de Clases Latentes y su situación respecto a que tienen o no bienestar, se puede estimar la sensibilidad y la especificidad del resultado. La Tabla 5.32 muestra la relación entre una prueba diagnóstica y la presencia o ausencia de "bienestar".

Tabla 5.32: Relación entre una prueba diagnóstica y la presencia o ausencia de "bienestar"

		Clasificación verdadera	
		Pobre	No pobre
Resultados de la metodología	S/bienestar	Verdadero positivo (VP)	Falso positivo (FP)
	C/bienestar	Falso negativo (FN)	Verdadero negativo (VN)

Fuente: (LÓPEZ & PITA FERNÁNDEZ, 2003. Actualizado en 2010).

Se define, entonces:

$$\text{Sensibilidad} = \frac{VP}{VP + FN} = FVP \text{ (fracción de verdaderos positivos)} \quad (5.3)$$

$$\begin{aligned} \text{Especificidad} &= \frac{VN}{VN + FP} = FVN \text{ (fracción de verdaderos negativos)} \\ &= 1 - FFP \text{ (fracción de falsos positivos)} \end{aligned} \quad (5.4)$$

Los conceptos de sensibilidad y especificidad permiten validar una prueba diagnóstica. Proporcionan información respecto a la probabilidad de obtener un resultado concreto ("con bienestar" o "sin bienestar") en función de la verdadera condición de un hogar en relación con la pobreza. La Tabla 5.33 muestra la clasificación de los hogares aplicando el Método Multidimensional de Clases Latentes y el Unidimensional de Línea de Pobreza.

Tabla 5.33: Resultados de la clasificación de los hogares aplicando los Métodos Multidimensional y Unidimensional

		Clasificación verdadera	
		Pobre	No pobre
Resultados de la metodología	S/bienestar	478	3
	C/bienestar	88	489

Fuente: Elaboración propia

$$\text{Sensibilidad} = \frac{478}{566} = 0,84 \text{ (fracción de verdaderos positivos)}$$

$$\text{Especificidad} = \frac{489}{492} = 0,99 \text{ (fracción de verdaderos negativos)}$$

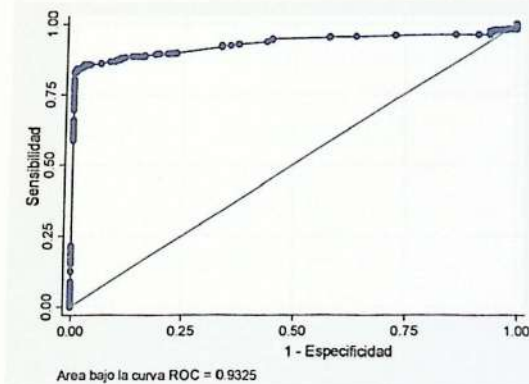
La curva ROC es un gráfico en el que se observan todos los pares "sensibilidad/especificidad" (designados S y E respectivamente), resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados. En el eje y de coordenadas se sitúa la sensibilidad o fracción de verdaderos positivos, definida

como se expuso anteriormente y calculada en el grupo de pobres (Figura 5.25). En el eje x se sitúa la fracción de falsos positivos o 1-especificidad calculada en el subgrupo no pobre. Cada punto de la curva representa un par S/1-E correspondiente a un nivel de decisión determinado. Una prueba con discriminación perfecta, sin solapamiento de resultados en las dos poblaciones, tiene una curva ROC que pasa por la esquina superior izquierda, donde S y E toman valores máximos (S y E = 1). Una prueba sin discriminación, con igual distribución de resultados en los dos subgrupos, da una línea diagonal de 45°, desde la esquina inferior izquierda hasta la superior derecha. La mayoría de las curvas ROC caen entre estos dos extremos. Cualitativamente, cuanto más próxima es una curva ROC a la esquina superior izquierda, más alta es la exactitud global de la prueba (GIBBONS, 2007).

El área bajo la curva ROC se emplea como un índice de la exactitud global de la prueba: la exactitud máxima correspondería a un valor de ABC de 1 y la mínima a uno de 0.5 (si fuera menor de 0.5 se debería invertir el criterio de positividad de la prueba).

En la Figura 5.25 se observa la curva ROC obtenida para la muestra de hogares estudiada con el fin de establecer si dichos hogares tienen o no bienestar. El área bajo la curva es 0,93, lo que significa que hay 93% de probabilidad de que el Modelo de Clases Latentes pueda distinguir entre los hogares con y sin bienestar.

Figura 5.24: Gráfico de la Curva ROC. Evaluación del Método Multidimensional para la determinación de hogares "con bienestar" y "sin bienestar"



5.7 Comentarios

En este capítulo se ha desarrollado un proceso para la composición de un indicador de bienestar multidimensional aplicando metodologías multivariantes innovadoras como la Teoría de Respuesta al Ítem (Modelo de Rasgo Latente) y el Modelo de Clase Latente. Se expusieron los fundamentos y la teoría de estimación clásica para los Modelos de Clases Latentes, considerando variables manifiestas (Ítems) de naturaleza categórica que son las que determinan las clases.

Según las dimensiones consideradas y los ítems asociados a cada dimensión, los modelos que mejor ajustaron a los datos fueron de dos clases latentes bien definidas.

Los datos utilizados para aplicación de los modelos mencionados corresponden a la Encuesta Anual de Hogares Urbanos, junto con la Encuesta Nacional sobre Acceso y Uso de Tecnologías de la Información y la Comunicación (ENTIC), por lo tanto, el análisis se limitó a la información contenida en ambas encuestas.

Por otro lado, no se realiza inferencia alguna puesto que el objetivo principal es mostrar una metodología multidimensional, alternativa a los métodos tradicionales de Líneas de Pobreza y Necesidades Básicas Insatisfechas (NBI) para medir pobreza.

Capítulo 6

Conclusiones

El objetivo propuesto en esta tesis es construir un indicador de bienestar multidimensional que sea fácil de interpretar, de utilizar en ulteriores estudios y que aporte información distinta y/o complementaria a las tradicionales medidas unidimensionales (por ingresos -línea de pobreza- y por NBI). Examinar su utilidad en predicción de condiciones y situaciones y que tenga propiedades psicométricas, es decir que sea "confiable" (mide en forma consistente) y que tenga "validez" (mide realmente el atributo). El principal aporte es la recapitulación de los conceptos y de las principales metodologías junto con los estudios más importantes realizados en Argentina y en algunos países de Europa y América Latina.

El contexto general se enmarca en el estudio de la pobreza, concretamente en el análisis de la acepción multidimensional de la misma para el caso específico de Tucumán. El enfoque moderno de la pobreza se acerca más al concepto de bienestar en el que influyen numerosos factores, no sólo a nivel del ingreso monetario.

Las razones que justifican la elaboración de esta tesis doctoral se resumen en tres tipos diferentes pero complementarias: por un lado, existe una motivación personal tendiente a aportar ideas para detectar las carencias en los hogares de Tucumán y que sirvan para la implementación de políticas públicas tendientes a mejorar la calidad de vida de esos hogares. Es fundamental investigar en aquello que afecta al bienestar de las personas, máxime si se generan situaciones de desarraigo social. Por otro lado, aparece una justificación estrictamente económica, ya que, hoy por hoy la pobreza sigue siendo un asunto sin resolver y por ello existe el interés de aportar un "granito de arena" a la Comunidad Científica y Política. Esto incentiva la creatividad por encontrar una forma sencilla de medir la pobreza y de explicar sus causas, con el fin de actuar sobre ella.

Durante el proceso de investigación, se realiza una extensa búsqueda de bibliografía sobre pobreza y su diversidad de formas y dimensiones para establecer el marco teórico. Además, la participación en numerosos seminarios, cursos, talleres,

foros y otros encuentros científicos, también contribuyeron con importantes conocimientos y criterios para la formación del marco teórico y la metodología estadística utilizada.

La función del índice multidimensional propuesto permite contar con información complementaria a la obtenida mediante los métodos convencionales. También sirve para dar respuesta a distintos factores sobre el desarrollo social y el bienestar y evalúa si las personas u hogares logran alcanzar umbrales mínimos de bienestar (o estándar de vida) en cada una de las dimensiones consideradas.

Por otro lado, si bien se incluyen dimensiones relevantes para la satisfacción de necesidades, como vivienda, agua y saneamiento, educación y capacidad de consumo y bienes durables, se excluyeron otras, como nutrición, mortalidad infantil y salud tanto o más importantes que las mencionadas anteriormente para medir el bienestar. Existen encuestas específicas de salud donde se dispone de estos datos, pero dichas encuestas no suelen ir acompañadas de otras variables relevantes tales como, la situación laboral, los ingresos del hogar o las características y materiales de la vivienda, entonces no se puede incluir como otra medición. Por otro lado, estas condiciones llevan a que índices desarrollados en un país tengan dificultades de aplicación en otros países.

En cuanto a las metodologías empleadas, los Modelos de Variables Latentes resultan bastante convenientes para el análisis multidimensional de ítems categóricos. Por un lado, los modelos de Teoría de Respuesta al Ítem (IRT) se aplican como un Análisis Factorial Confirmatorio para determinar los ítems de mayor correlación con las dimensiones (rasgos latentes) y, por otro lado, con los Modelos de Clases Latentes (LCM) se logra diferenciar en cuántas clases latentes se divide la muestra en cada dimensión. De esa manera, es posible determinar los ítems que más discriminan entre los hogares con bienestar y sin bienestar. Además, la posibilidad de seleccionar los ítems de mayor correlación con las variables latentes permite diseñar mejor las preguntas en las encuestas.

Otro punto importante para tener en cuenta es la continuidad en las metodologías de medición y estimación de indicadores para poder comparar resultados a través del tiempo y el espacio y analizar la evolución de índices de

pobreza multidimensional, lo que lleva a una aplicación adecuada y eficiente de las políticas públicas.

Por último, es un tema con muchas posibilidades de mejora, sobre todo en los medios necesarios para ello de: "información de alcance, bases de datos, técnicas de pobreza multidimensional multivariantes, software adecuado y temas de actualidad".

6.1 Principales líneas de investigación

Para concluir este trabajo, se mencionan algunas líneas de investigación que se pretende desarrollar en adelante. En primer lugar, actualizar los aspectos meramente empíricos con una nueva encuesta que sustituye a las EPH, una encuesta de condiciones de vida. En segundo lugar, mejorar las variables que incluye el indicador en la medida de lo posible y realizar comparaciones con otros países de América Latina. También, se tiene interés en encontrar soluciones metodológicas para dar carácter dinámico al modelo explicativo de la pobreza multidimensional.

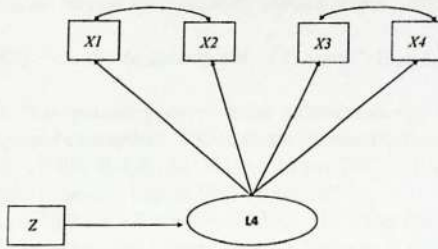
Por otro lado, se desea investigar el Modelo de Clases Latentes Multinivel (MLCM). Este modelo es una extensión muy importante del Modelo de Clases Latentes, cuyo objetivo principal es el de construir perfiles a partir de un conjunto de variables categóricas observadas; bajo el supuesto de independencia entre observaciones, lo que frecuentemente es vulnerado si se analizan datos que presentan cierta jerarquía, con el fin de obtener una solución óptima del número de subclases que pertenecen a las variables latentes.

Las encuestas de hogares tienen una estructura jerárquica natural (personas dentro de los hogares), entonces en el mismo Modelo de Ecuaciones Estructurales se podrían incorporar variables latentes independientes (bienestar o alguna dimensión del bienestar) pertenecientes a diferentes niveles de agregación. Por ejemplo, como se explicó en la sección 5.4.5, la variable latente "funcionamiento del jefe de hogar" está asociada a las características propias del jefe de hogar y algunas de esas características son compartidas por otros miembros del mismo hogar; entonces el análisis del bienestar multidimensional en los hogares se puede ampliar considerando no sólo al jefe de hogar, sino también a los otros miembros.

Otra innovación al Modelo Básico de Clases Latentes es la posibilidad de relajar el supuesto de independencia local. Una forma de mejorar los modelos que presentan un ajuste deficiente a los datos es permitir asociaciones entre las variables manifiestas, así como dependencias entre las covariables y los indicadores (ver Figura 6.1).

En diversos modelos, al incorporar dependencias entre las variables manifiestas, los errores en la clasificación pueden disminuir notablemente. Por ejemplo, al tiempo que se incorpora el efecto de la covariable Z sobre la variable latente L , se pueden modelar dependencias entre las variables $X_1 - X_2$, así como entre $X_3 - X_4$. La Figura 6.1 muestra un modelo SEM de clases latentes con una variable agrupadora y dependencias.

Figura 6.1: Modelo de clases latentes con variable agrupadora y dependencias



Fuente: (CENEVAL, 2009).

Por último, como reflexión personal, en base a toda la lectura y estudios realizados, se plantea el análisis de la pobreza y su difícil solución desde un punto de vista multidisciplinar ya que no sólo es un problema económico; y, por otro lado, a reflexionar sobre ¿qué es la pobreza?, ¿qué factores son importantes en el proceso?, ¿qué debe hacer la sociedad en general y los gobiernos en particular? ...

Estos planteos incentivan la motivación de actualización continua y de aportar nuevas ideas en futuras investigaciones.

BIBLIOGRAFÍA

- AGUILAR GARZA, L.; PARADA RIVERA, M. (2007). "La pobreza al interior del capitalismo: ¿Fenómeno marginal o resultado estructural del sistema económico?" Universidad Centroamericana "José Simeón Cañas". Antiguo Cuscatlán, El Salvador.
- ALKIRE, S.; FOSTER, J. (2008). "Recuento y medición multidimensional de la pobreza". OPHI, Serie Documentos de trabajo. N° 7. Oxford, Inglaterra.
- ALKIRE, s.; ALIM, A. (2019). "El estado de la pobreza infantil multidimensional en Asia meridional: una visión contextual y de género". OPHI Working Paper N° 127 - ISSN 2040-8188.
- ALTIMIR, O. (1979). "La dimensión de la pobreza en América Latina". Cuadernos de la CEPAL, 1-101.
- ARAKAKI, A. (2011). "La pobreza en Argentina 1974-2006. Construcción y análisis de la información". CEPED - Documento de Trabajo N° 15, 1-112. Buenos Aires. Argentina.
- ARAYA, C.; SEPÚLVEDA, R. (2010). "Análisis de clases latentes en tablas poco ocupadas. Revista de Matemática. Teoría y Aplicaciones". ISSN: 1409-2433, 25 - 40.
- ATKINSON, A. B. (1987). "On the Measurement of Poverty". *Econometrica*, vol 55, N° 4, 749-764.
- ATKINSON, A. (1991). "Comparing poverty rates internationally: lesson from recent studies in development countries". *World Bank Economic Review*, 5(1), pp 3-21.
- ASPAROUHOV, T. ; MUTHEN, B. (2012). "Using Mplus TECH11 and TECH14 to test the number of latent classes". Mplus Web Notes: N° 14.
- BARTHOLOMEW, D.; STEELE, F.; MOUSTAKI, I.; GALBRAITH, J. (2008). "Analysis of multivariate social science data". Chapman & Hall.
- BARTHOLOMEW, D. ; KNOTT, M.; MOUSTAKI, I. (2011). "Latent Variable Models and Factor Analysis: A Unified Approach". 3rd Edition. Copyright © 2011 John Wiley & Sons, Ltd.
- BATISTA FOGUET, J. M.; COENDERS GALLART, G. (2000). "Modelo de ecuaciones estructurales". Madrid: La Muralla, S. A.
- BECCARIA, L.; GROISMAN, F. (2007). "Informalidad y pobreza en la Argentina". Publicación UNGS - Sociedad Argentina de Economía Política, pp 1-19.
- BOCK. (1972). "Estimación de los parámetros del ítem y capacidad latente cuando las respuestas son dos o más categorías nominales". *Psychometrika*, 37, pp 29-51
- BOLTVINIK, J. (2003). "Concepto y medición de la pobreza. La necesidad de ampliar la mirada". *Papeles de Población*, vol 9, N° 38, ISSN:1405-7425, 9-25.
- BOURGUIGNON, F.; CHAKRAVARTY, S. (2003). "La medición de la pobreza multidimensional". *Journal of Economic Inequality*, pp 25-49.
- CAPPELLARI, L.; JENKINS, S. (2006). "Summarising Multiple Deprivation Indicators". Research Gate.
- CASTILLO, O.; MORALES, U.; GRANADOS, E.; LÓPEZ, A.; VILLAR, Y.; BERNABEU, N. (2006). "Reflexiones: Desarrollo local con equidad de género".

- San Salvador, El Salvador: Programa de las Naciones Unidas para el Desarrollo (PNUD). ISBN 99923-874-2-4.
- CASTRO LÓPEZ, C.; TENORIO AGUIRRE, A. (2010). "*Modelos de clase latente aplicados a encuestas de percepción ciudadana*". Revista Legislativa de Estudios Sociales y de Opinión Pública, Vol 3. N° 6 PP 187-200.
- CEPAL y PNUD. (1989). "*Ecuador. Mapa de Necesidades Básicas Insatisfechas*". Naciones Unidas - División Estadística y Proyecciones. PNUD-RLA/86/004.
- CEPAL. (1991). "*Magnitud de la pobreza en América Latina en los años ochenta*". Publicaciones de las Naciones Unidas - ISSN 0256-9795 ISBN 92-1-321356-5 .
- CEPAL. (2018). "*Medición de la pobreza por ingresos. Actualización metodológica y resultados*". Santiago: Metodologías de la CEPAL, N° 2 (LC/PUB.2018/22-P).
- CENEVAL (2009). "*Escalamiento de variables de contexto*". México, D.F.: Centro Nacional de Evaluación para la Educación Superior. Cuaderno N°7.
- CHACON, E. (2008). "*Análisis de variables ordinales en modelos de variable latente: evaluación de métodos de estimación y procedimientos de ajuste*". Madrid: Universidad Autónoma de Madrid.
- CHACON, E. (2011). "*Análisis de variables ordinales en modelos de variables latentes*". LAP Lambert Acad. Publ. Editorial - EAE - 978-3-8465-6549-0.
- CLOGG. (1981). "*New development in latent structure analysis*". En Jackson D.J. y Bogotta E.F. (eds) "*Factor Analysis and Measurement in Sociological Research*", Beverly Hills : Sage, pp 215-246.
- CONCONI, A. (2011). "*Pobreza multidimensional en Argentina: ampliando las medidas tradicionales de pobreza por ingresos y NBI*". Facultad de Ciencias Económicas - Universidad Nacional de la Plata - ISSN 1853-3930. Documento de trabajo N° 90.
- CONEVAL (2008). "*Metodología para la medición multidimensional de la pobreza en México*". Ciudad de México: Consejo Nacional de Evaluación de la Política de Desarrollo Social.
- CORDERA CAMPOS, R.; RAMÍREZ KURI, P.; ZICCARDI, A.; VANEGAS, L. (2008). "*Pobreza, desigualdad y exclusión social en la ciudad del siglo XXI*". Dialnet, ISBN 978-607-3-000-43-7.
- CORTADA, N.; CUPANI, M. (2013). "*Teoría de Respuesta al Ítem*". En S. Tornimbeni, E. Pérez, & F. Olaz, "Introducción a la Psicometría". Paidós Mexicana, ISBN-13: 978-9501260854.
- CUPANI, M. (2012). "*Análisis de Ecuaciones Estructurales.: conceptos, etapas de desarrollo y un ejemplo de aplicación*". Revista Tesis N° 1, 186-199.
- DE ANDRADE, D.; RIBEIRO, H.; DA CUNHA, R. (2000). "*Teoria da Resposta ao Item: Conceitos e Aplicações*". San Pablo - Brasil: Associação Brasileira de Estatística.
- EUROSTAT. (2012). "*Measuring material deprivation in the EU. Indicators for the whole population and child-specific indicators*". Eurostat. Methodologies & Working Papers.
- EVERITT, B. S. (1984). "*Introduction to latent variable models*". Biometrical Journal, 107 pp.

- FERES, J. C.; MANCERO, X. (2001a). "Enfoques para la medición de la pobreza". Breve revisión de la literatura. CEPAL. Serie Estudios Estadísticos y Prospectivos N° 4, 1-45.
- FERES, J.; MANCERO, X. (2001b). "El Método de las Necesidades Básicas Insatisfechas y sus aplicaciones en América Latina". CEPAL - SERIE - Estudios Estadísticos y Prospectivos N° 7.
- FERNÁNDEZ ARAUZ, A. (2015). "Aplicación del Análisis Factorial Confirmatorio a un modelo de medición del rendimiento académico en lectura". Ciencias Económicas 33-N°2: 2015/ 39-66 / ISSN: 0252-9521.
- FUSCO, A.; DICKES, P. (2008). "The Rasch Model and Multidimensional Poverty Measurement". En Quantitative Approaches To Multidimensional Poverty Measurement. Palgrave MacMillan.
- GIBBONS, J. A. (2007). "Expression of human protein phosphatase-1 in *saccharomyces cerevisiae* highlights the role of phosphatase isoforms in regulating eukaryotic functions". Journal Article | Research Support, N.I.H., Extramural.
- GILL, P. (2001). « Modelos Lineales Generalizados: un enfoque unificado ». Sage University Papers Series. Thousand Oaks. CA: Sage
- GLEWWE, P.; VAN DEL GAAG, J. (1990). "Identificación de los pobres en los países en desarrollo. ¿Importan las definiciones diferentes?" World Development, 803-814.
- GÓMEZ, A.; ALVAREZ, G.; MARIO, S.; OLMOS, F.; INDEC. (2004). "Metodología de elaboración del Índice de Privación Material de los Hogares (IPMH)". Buenos Aires: DNESyP/DEP/P5/PID Serie Pobreza - INDEC.
- GOODMAN, L. (1974). "Exploratory latent structure analysis using both identifiable and unidentifiable models". Biometrika, Vol. 61, Issue 2, 215-231.
- GUIO, A.; GORDON, D.; NAJERA, H.; POMATI, M. (2017). "Revisión de las variables de privación material de la UE. Luxemburg". Publications Office of the European Union
- HAIR, J.; ANDERSON, R.; TATHAN, R.; BLACK, W. (1999). "Análisis multivariante". Madrid: Prentice Hall Iberia.
- HALPERIN, L.; LABIAGUERRE, J.; DELPECH, C.; GONZÁLEZ, M.; HOREN, B.; VILLADEAMIGO, J.; MÜLLER, G. (2008). "Políticas sociales en la Argentina: entre la ciudadanía plena y el asistencialismo focalizado en la contención del pauperismo". Buenos Aires: CEPED - Cuaderno N°10 - ISBN 978-950-29-1070-3
- HOSMER, D., & LEMESHOW, S. (2000). "Applied logistic regression". 2° Ed. JOHN WILEY & SONS, INC.
- INDEC (1984). "La Pobreza En Argentina". Buenos Aires: Informes de Prensa - INDEC.
- INDEC (2002). Incidencia de la pobreza y de la indigencia en el aglomerado Gran Buenos Aires. Buenos Aires: Información de Prensa - INDEC ISSN 0327-7968.
- INDEC (2003a). "A cerca del método para la medición de la pobreza en Argentina". Buenos Aires: Informe INDEC.
- INDEC (2003b). "Incidencia de la pobreza y de la indigencia en los aglomerados urbanos". Buenos Aires: Información de Prensa del INDEC, ISSN 0327-7968.

- INDEC (2011). "Encuesta Anual de Hogares Urbanos. Diseño de registro y estructura para las bases de microdatos individual y hogar". Buenos Aires: INDEC - ISBN 978-950-896-413-7.
- INE (2005). "La Pobreza y su Dimensión. Presentación de diversos métodos de obtención de medidas de pobreza".
<https://www.ine.es/daco/daco42/sociales/pobreza.pdf>
- IRARRÁZAVAL, I. (1995). "Habilitación, pobreza y política social". Estudios Públicos, N° 59, 100-165.
- JÖRESKOG, K. (1971). "Statistical Analysis of Sets of Congeneric Tests". Psychometrika, 109-133.
- JÖRESKOG, K.; GOLDBERGER, A. (1975) "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable". Journal of American Statistical Association, Vol 70 N° 351, 1975: 631-639.
- JÖRESKOG, K. (1993) "Testing Structural Equation Models". Advances Journalism and Communication. Vol 2 N° 3 : pp. 294-316.
- KAPLAN, D. (2003). "Structural equation modeling". Foundations and extensions. ResearchGate - RG.
- KLINE, R. (2005). "Principles and practice of structural equation modeling". 2° Ed. N. York: Guilford Press.
- KOTZ, S.; JOHNSON, N. (1982). "Encyclopedia of Statistical Sciences". Vol. 3. USA: Wiley-Interscience publications. ISBN 0-471-05549-2.
- LAZARSFELD, P.; HENRY, N. (1968). *Latent Structure Analysis*". New York: Houghton, Mifflin.
- LAZARSFELD, P. (1954). "El Enfoque Matemático de las Ciencias Sociales". Editorial Centro de Investigaciones Sociológicas. Colección: Clásicos del Pensamiento Social. ISBN 9788474767292.
- LEENEN, I. (2014). "Virtudes y limitaciones de la Teoría de Respuesta al Ítem para la evaluación educativa en las ciencias médicas". Investigación en Educación Médica, 40-55.
- LÉPORE, E.; SALVIA, A. (2005). "La Naturaleza Multidimensional de la Pobreza. Algunas aproximaciones empíricas al déficit de capacidades de subsistencia". Tandil: Universidad Nacional de Tandil - Asociación de Estudios de la Población Argentina (AEPA).
- LÓPEZ, G.; PITA FERNÁNTEZ, S. (2003, actualizado en 2010). "Pruebas Diagnósticas". Unidad de Epidemiología Clínica y Bioestadística. Cad Aten Primaria, pp 120-124.
- MALDONADO, L.; KRONMÜLLER, E.; GUTIÉRREZ CROCCO, I. (2016). "Estrategia para la Inferencia Causal y Planificación de Estudios Observacionales en las Ciencias Sociales: el caso de Chaitén post erupción del 2008". Revista de Ciencia Política, vol 36, N° 3.
- MANZANO PATIÑO, A.; ZAMORA MUÑOZ, S. (2009) "Sistema de Ecuaciones Estructurales: una herramienta de investigación". Cuaderno Técnico N°4, México: CENEVAL.
- MARTINO, A. (2013). "Pobreza y Política". ICIMISS.
http://www.institutocirculomss.com.ar/Opiniones/archivo/martino_pobresa.htm

- MINUJIN, A.; HEE BANG, J. (2002). "Indicadores de Inequidad Social. Acerca del uso del "Índice de Bienes" para la Distribución de los Hogares". Desarrollo Económico N° 165, vol. 42. ISSN 0046-001X, 129-146.
- MONROY CASORLA, L., VIDAL URIBE, R., & SAADE HAZIN, A. (2009). "Análisis de Clases Latentes. Una técnica para detectar heterogeneidad en poblaciones". MÉXICO: CENEVAL.
- MUJER, I. (2011). "Las madres solas sufren más riesgo de pobreza y de exclusión social y laboral". Diario El Mundo, edición del 9 de junio de 2011.
- MUNDIAL BANK, (1999). "Informe sobre el desarrollo mundial". Banco Mundial - Informe N° 19761.
- MUNDIAL BANK, (2000/2001). "Informe sobre el desarrollo mundial. Lucha contra la pobreza". Madrid - Barcelona - México: Ediciones Mundi-Prensa.
- MUTHEN, L. ; MUTHEN, B. (2010). "Mplus: Statistical Analysis with Latent Variables. User's guide". Los Angeles: Los Angeles, Muthen & Muthen.
- NAJERA CATALÁN, H. (2016). "Multiple Deprivation, Severity and Latent Sub-Groups: Advantages of Factor Mixture Modelling for Analysing Material Deprivation". Social Indicators Research **131**, 681-700
<https://doi.org/10.1007/s11205-016-1272-y>
- NAVARRO, K.; ASÚN, R. (2016). "Desarrollos resientes en Estadística. Aportes teórico-metodológicos a la investigación sociológica". Sociología y Tecnociencia. E-ISSN: 1989-8487, pp 1-13.
- NYLUND, K.; ASPAROUHOV, T. ; MUTHEN, B. (2007) "Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modelling: A Monte Carlo Simulation Study". Journal Structural Equation Modeling: A Multidisciplinary Journal, vol 14 – Issue 4
<https://doi.org/10.1080/10705510701575396>
- PAES DE BARROS, R.; DE CARVALHO, M.; FRANCO, S. (2006). "Pobreza Multidimensional no Brasil". Texto para Discussao N°1227.
- PAZ NÚÑEZ, E. (1996). "La Pobreza". Estudios Sociales:
<https://www.monografias.com/trabajos89/pobreza-situacion/pobreza-situacion.shtml>
- PAZ, J. (2014). "Pobreza multidimensional en Argentina. Asimetrías regionales (Parte I)". Instituto de Estudios Laborales y del Desarrollo Económico (IELDE) - Documento de trabajo N° 11, 1-49.
- PÉREZ GIL, J. A. (2004). "Modelos para ítems politómicos de respuestas discretas". Universidad de Sevilla: Dpto de Psicología Experimental.
- PNUD (1997). "Informe sobre Desarrollo Humano". Ediciones Mundi-prensa, Madrid
- PNUD (2008). "Desarrollo de Capacidades". Notas de Práctica del PNUD, Nueva York.
- POSA LARA, C. (2008). "Pobreza multidimensional: el caso específico español a través del panel de hogares de la Unión Europea". Tesis doctoral. Universidad Complutense de Madrid. ISBN: 978-84-692-0035-3.
- RAVALLION, M. (1992). "Comparaciones de pobreza. Una guía de métodos y conceptos". LSMS Working Paper N°88. ISSN: 0253-4517, 1-138.

- RODRÍGUEZ ALONSO, E.; GONZÁLEZ SUÁREZ, R. (2002). "Análisis de las curvas ROC: un método útil para evaluar procedimientos diagnósticos". Revista Cubana de Endocrinología v.13 n.2 ISSN 1561-2953.
- ROMERO, A. (2000). "El mundo de la pobreza (Primera parte)". Tendencias, Vol 1, Nº2, 35-59.
- ROWNTREE, B. (1901). "Poverty: A Study of Town Life". Macmillan and Co. York, Inglaterra. ISBN 1-86134-202-0.
- SAID (2008). "Metodología en la Investigación sobre Discapacidad. Introducción al uso de Ecuaciones Estructurales". Publicaciones de INICO. Colección Actas.
- SALVADOR FIGUERAS, M. (2000). "Introducción al Análisis Multivariante". Zaragoza. <http://www.5campus.com/leccion/anamul/>
- SALVIA, A.; COMAS, G.; STEFANI, F. (2010). "Heterogeneidad estructural y acceso diferencial a empleos de calidad en dos momentos de crecimiento económico". VI Congreso de la Asociación Latinoamericana de Sociología del Trabajo. México: ALAST.
- SALVIA, A. (2005). "El Desarrollo Humano y la Segmentación Social de los Espacios Urbanos". Serie Monitoreo de la Deuda Social Argentina – Documento 1. Buenos Aires, Argentina.
- SANCHEZ RIVERO, M. (2004). "Introducción a la Teoría de Respuesta al Ítem. Una Herramienta para el Análisis de Variables Latentes". XVIII Reunión Anual Asepelt-España. León, España.
- SATRIANO, C. (2006). "Pobreza, políticas públicas y políticas sociales". Mad. Nº 15, 1-15.
- SEN, A. (1984). "Resource Values and Development". Basil Blackwell Publisher. Oxford.
- SEN, A. (1987). "Commodities and Capabilities". Oxford University Press. Oxford.
- SEN, A. (1999). "Development as Freedom". Oxford University Press. Oxford.
- SHANNON, C. (1948). "A mathematical theory of communication". Bell System Technical Journal, 27: 379-423.
- TAMI, F.; SALVIA, A. (2004). "Desarrollo Humano y Deuda Social. Aproximaciones conceptuales y metodológicas". Barómetro de la Deuda Social Argentina/1: Las Grandes Desigualdades. EDUCA. Buenos Aires, Argentina.
- TUÑÓN, I.; FOURCADE, H.; GONZÁLEZ, M.; REGGINI, N. (2012). "Los indocumentados en Argentina. La cara invisible de la pobreza". Biblioteca digital de la UCA - ODSA - Documento de trabajo .
- UNTREF. (2014). "Delito, Marginalidad y Desempeño Institucional en la Argentina: Resultado de una encuesta de presos condenados". Centro Estudios Latinoamericanos sobre Inseguridad y Violencia. ISBN 978-987-1889-41-9. Sáenz Peña.
- VALDIVIESO TABORGA, C. (2012). "Estudio de la Especificación de Medición Formativa y Reflexiva de las Dimensiones de la Calidad de Servicio de Clases dentro de un Modelo Comprehensivo de Relación con el Valor Percibido, la Satisfacción y las Intenciones de Comportamiento de Postclase". Tesis de Doctorado en Economía y Administración de Empresas. Universidad Privada Boliviana. Cochabamba, Bolivia.

- VERDUGO, M.; CRESPO, M.; BADÍA, M.; ARIA, B. (2008). "*Metodología en la Investigación sobre la Discapacidad. Introducción al uso de Ecuaciones Estructurales*". VI Simposio Científico SAID. Publicaciones del INICO. Colección Actas. ISBN: 978-84-691-5852-4, 2008. Salamanca, España.
- WRIGHT, S. (1934). "*The Method of Path Coefficient*". *Mathematical Statistics*, pp 161-215.
- YANG, C. (2006). "*Evaluación de modelos de análisis de clases latentes en la identificación de fenotipos cualitativos*". *Estadísticas Computacionales y Análisis de Datos*; 50 (4): 1090-1104.
- ZAMAGNI, S. (2010). "*El asistencialismo social produce una gran pobreza*". *Diario La Nación*, del 18 de agosto de 2010.
- ZICCARDI, A. (2010). "*Pobreza Urbana, Marginalidad y Exclusión Social*". *Revista Ciencia* N° 26, pp 26-35.
www.revistaciencia.amc.edu.mx/images/revista/61_4/pdf/05_Pobreza_Urbana.pdf