

Modelado Avanzado con PLS-SEM:

Dominando la No Linealidad, la Heterogeneidad y el Poder Predictivo para la Investigación de Negocios

Fredy Quispe Gómez

Rossana Esmila Cortez Lopez

Yanet Mamani Vargas

Sandra Ines Ponce Umiña

Oscar René Barrientos Huamán

Alberto Magno Cutipa Limache

Manuel Anchapuri Quispe

Fredy Quispe Gómez
Rossana Esmila Cortez Lopez
Yanet Mamani Vargas
Sandra Ines Ponce Umiña
Oscar René Barrientos Huamán
Alberto Magno Cutipa Limache
Manuel Anchapuri Quispe

**Modelado Avanzado con PLS-SEM: Dominando
la no linealidad, la heterogeneidad y el poder predictivo
para la investigación de negocios.**

Modelado Avanzado con PLS-SEM: Dominando la no linealidad, la heterogeneidad y el poder predictivo para la investigación de negocios

Autores:

Fredy Quispe Gómez

Rossana Esmila Cortez Lopez

Yanet Mamani Vargas

Sandra Ines Ponce Umiña

Oscar René Barrientos Huamán

Alberto Magno Cutipa Limache

Manuel Anchapuri Quispe

Editor

Manuel Anchapuri Quispe

Av. La torre Nro 773

manchapuri@unap.edu.pe

Puno – Perú

Primera edición, febrero del 2026

Versión digital

DEPÓSITO LEGAL DEL LIBRO ELECTRÓNICO N° 2026-00291

ISBN N° 978-612-03-1884-3

Disponible en: <https://demokno.info/publicaciones/>

Diseño y diagramación

Micol Zayetsy Anchapuri Ramos

Diseño de cubierta:

Manuel Anchapuri Q.

Toma de decisiones estratégicas: La integración de NCA, HOC e IPMA en el Análisis PLS-SEM Avanzado

Fredy Quispe Gómez

Rossana Esmila Cortez Lopez

Yanet Mamani Vargas

Sandra Ines Ponce Umiña

Oscar René Barrientos Huamán

Alberto Magno Cutipa Limache

Manuel Anchapuri Quispe

Como referenciar:

Quispe Gómez, F., Cortez Lopez, R. E., Mamani Vargas, Y., Ponce Umiña, S. I., Barrientos Huamán, O. R., Cutipa Limache, A. M., y Anchapuri Quispe, M. (2026). *Toma de decisiones estratégicas: La integración de NCA, HOC e IPMA en el Análisis PLS-SEM Avanzado*. Demokno.

<https://demokno.info/publicaciones/>



Toma de decisiones estratégicas: La integración de NCA, HOC e IPMA en el Análisis PLS-SEM Avanzado © 2026 by Quispe Gómez, F., Cortez Lopez, R. E., Mamani Vargas, Y., Ponce Umiña, S. I., Barrientos Huamán, O. R., Cutipa Limache, A. M., y Anchapuri Quispe, M., is licensed under Creative Commons Attribution-ShareAlike 4.0 International Para ver una copia de esta licencia, visite <https://creativecommons.org/licenses/by-sa/4.0/>

“La mejor venganza es ser diferente a quien causó el daño.”

“Nada tiene tanto poder para ampliar la mente que la capacidad de investigar de manera sistemática y real todo lo que es susceptible de observación en la vida”

Marco Aurelio,

"Elecciones fáciles, vida difícil. Elecciones difíciles, vida fácil"

Jerzy Gregorek

Dedicatoria:

A los curiosos, a los persistentes y a los que no temen adentrarse en la complejidad de los Modelos de Ecuaciones Estructurales, les dedicamos este trabajo. Que encuentren en estas páginas no solo fórmulas y métodos, sino también la inspiración para seguir explorando, cuestionando y mejorando. Porque cada investigador que avanza, eleva el conocimiento colectivo y abre nuevas puertas al futuro.

Pensamiento:

*"Nunca te despojes de toda
defensa, porque tarde o temprano
serás atacado por los que antes te
respetaban".*

*“Solo viviré una vez; por
lo tanto, Cuánto bien haga y
cuanta bondad pueda mostrar a un
Ser Humano, he de hacerlo ahora.
No debe aplazarlo ni olvidarlo,
Pues no volveré a pasar por aquí”.*

PRESENTACIÓN DEL LIBRO

Por los autores:

La complejidad como ventaja competitiva: una nueva era para el modelado estructural

La investigación en ciencias empresariales vive una paradoja. Mientras que la disponibilidad de datos y la capacidad computacional han crecido exponencialmente, muchos investigadores y analistas siguen atados a modelos estadísticos diseñados para un mundo que ya no existe: un mundo de relaciones lineales, poblaciones homogéneas y causalidades simples.

Este libro, "Modelado Avanzado con PLS-SEM: Dominando la No Linealidad, la Heterogeneidad y el Poder Predictivo", surge como respuesta a esa brecha. No es un manual introductorio; es un manifiesto técnico para el investigador que ha comprendido que la realidad empresarial es intrínsecamente "ruidosa" y compleja, y que busca herramientas capaces de modelar esa complejidad en lugar de ignorarla.

¿Por qué este libro ahora?

La transición de los métodos basados en covarianza (CB-SEM) a los métodos basados en varianza (PLS-SEM) ya no es una novedad; es el estándar de facto para la predicción. Sin embargo, la mayoría de la literatura se detiene en los fundamentos. Esta obra comienza donde las otras terminan.

Bajo la filosofía de la ciencia abierta y utilizando la potencia del entorno R, este texto guía al lector a través de los desafíos más sofisticados del análisis moderno:

1. **Del Promedio a la Estructura Latente:** Desmantelamos la falacia de la homogeneidad. A través de técnicas de segmentación avanzada como FIMIX-PLS (Finite Mixture) y PLS-POS (Prediction-Oriented Segmentation), enseñamos a descubrir subgrupos de datos ocultos que cambian radicalmente la interpretación de los resultados.
2. **La Realidad no es una Línea Recta:** Los rendimientos decrecientes y los puntos de saturación son comunes en los negocios. Aquí aprenderá a modelar y visualizar relaciones no lineales (cuadráticas y cúbicas), superando la miopía de los coeficientes lineales tradicionales.
3. **Rigor en la Medición:** Introducimos el Análisis Tetrádico Confirmatorio (CTA-PLS), permitiendo que sea la evidencia empírica, y no solo la intuición, la que decida si un constructo debe ser modelado como formativo o reflexivo.
4. **Convergencia con Data Science:** Finalmente, conectamos el modelado estructural con la revolución del Big Data y la Inteligencia Artificial, abordando la ética algorítmica y el uso de SEM en la gobernanza de datos.

Al lector exigente:

Este texto está dirigido a académicos de postgrado, investigadores doctorales y científicos de datos aplicados que necesitan ir más allá de la "significancia estadística". Si su objetivo es simplemente ajustar un modelo, existen otros manuales. Pero si su objetivo es *predecir* con precisión, entender la heterogeneidad de sus mercados y producir evidencia robusta que soporte la toma de decisiones estratégicas, entonces ha encontrado su guía.

La complejidad ya no es un obstáculo; con las herramientas adecuadas, es su mayor ventaja

Índice

PRESENTACIÓN DEL LIBRO	9
Índice	12
INTRODUCCIÓN	19
PRÓLOGO	23
FUNDAMENTOS DEL SEM: FILOSOFÍA Y ALCANCE	25
1.1 Definición y alcance del modelado de ecuaciones estructurales	26
1.2 De la regresión al SEM (Técnicas de 1ra vs. 2da Generación)	27
1.3 Anatomía del Modelo	28
1.3.1 Tipos de variables SEM	28
1.3.2 Arquitectura y componentes de un modelo SEM	30
1.3.3 Especificación del Modelo de Medida: Reflectivo vs. Formativo	32
1.3.4 Términos de Error	35
1.4 Fundamentos teóricos del SEM	37
1.5 Paradigmas del SEM: Covarianza (CB-SEM) vs. Varianza (PLS-SEM)	38
1.5.1 El Enfoque Confirmatorio: CB-SEM (Covariance-Based)	39
1.5.2 El Enfoque Predictivo: PLS-SEM (Partial Least Squares)	40
1.5.3 Marco de Decisión	41
1.6 Justificación del Enfoque PLS-SEM en ciencias empresariales	42

1.6.1	PLS-SEM en Ciencias Empresariales	42
1.6.2	La importancia de predecir constructos	42
1.7	Aplicabilidad gerencial y estratégica	43
1.8	Entorno informático SEM	44
1.8.1	PLS-SEM (Enfoque basado en la varianza)	44
1.8.2	CB-SEM (Enfoque basado en la covarianza)	45
DISEÑO DE INVESTIGACIÓN Y ESTRATEGIA MUESTRAL		48
1.1	Decisión analítica: el marco estratégico	50
1.2	Teoría del muestreo en SEM: calidad y robustez	51
1.3	Potencia estadística (1-beta) y decisiones metodológicas	53
2.1.1	El Estándar de Oro: El umbral del 80%	53
2.1.2	Implicaciones Críticas para PLS-SEM	55
1.4	Métodos de estimación del tamaño muestral	56
1.4.1	Nivel 1: Aproximación Heurística	56
1.4.2	Nivel 2: Análisis de Potencia a Priori	57
1.4.3	Nivel 3: Simulaciones y criterios de complejidad	59
1.5	Diseño para la investigación exploratoria y predictiva	61
PROTOCOLO DE EVALUACIÓN Y REPORTE DE RESULTADOS		66
3.1	Fundamentos metodológicos de la validación	67
3.1.1	Los dos niveles de teoría en SEM	67
3.1.2	El Principio de la "Secuencia de Dos Pasos"	69
3.2	Paso 1: Validación del modelo de medida (Outer Model)	70
3.2.1	Confiabilidad del Indicador	71
3.2.2	Consistencia Interna	72
3.2.3	Validez Convergente	73

3.2.4	Validez discriminante	74
3.3	Paso 2: Validación del Modelo Estructural (Inner Model)	74
3.3.1	Evaluación de la Colinealidad (VIF)	75
3.3.2	Evaluación de las Relaciones Causales	76
3.3.3	Coeficiente de Determinación (R^2)	77
3.3.4	Tamaño del Efecto (f^2)	77
3.3.5	Relevancia Predictiva (Q^2)	78
3.4	Consideraciones sobre el ajuste global y la predicción.	79
3.4.1	Fundamento algorítmico (Incongruencia Matemática)	80
3.4.2	Estado experimental de las métricas alternativas	80
3.4.3	Objetivo Epistemológico (Explicación vs. Predicción)	81
3.5	Síntesis y estándares de reporte académico	82
3.6	Guía de lecturas avanzadas y modelos de referencia	84
	MODELADO AVANZADO Y EVALUACIÓN DE MODELOS	90
4.1	Modelado de relaciones no lineales: Efectos cuadráticos	91
4.2	Fundamentos del análisis confirmatorio de tétradas (CTA-PLS)	98
4.3	PLS-Predict y Evaluación Predictiva Out-of-Sample	102
4.4	Estrategias de validación cruzada avanzada y comparación de modelos	104
4.5	Síntesis: Valor gerencial de la no linealidad, CTA y PLS-Predict	107
4.6	Investigaciones sobre: Modelado avanzado y evaluación de modelos, en bases de datos indexadas:	109
	MODELADO DE LA HETEROGENEIDAD OBSERVADA	114

5.1	Pre-requisito de la Invariancia de la Medición (MICOM) en PLS-SEM	116
5.2	Introducción al MGA y el Método Paramétrico Tradicional	119
5.3	Rigor Estadístico: MGA basado en Bootstrap	121
5.4	Superando Limitaciones: La Prueba de Permutación (Permutation MGA)	124
5.5	Estrategias Avanzadas: Manejo de la comparación entre múltiples grupos ($k > 2$)	126
5.6	Aplicación Práctica y Workflow: Un Estudio de Caso	128
5.7	Implicaciones gerenciales: Traducción de la heterogeneidad a estrategias contingentes	129
5.8	Investigaciones sobre: Modelado de la heterogeneidad observada, en bases de datos indexadas:	131
	MODELADO DE LA HETEROGENEIDAD NO OBSERVADA	135
6.1	Fundamentos Teóricos y Racional del FIMIX-PLS	137
6.2	Criterios de selección y evaluación del número de segmentos en FIMIX-PLS	139
6.3	Segmentación Orientada a Predicciones para Mínimos Cuadrados Parciales (PLS-POS)	141
6.4	Comparación y selección estratégica de la técnica de segmentación	142
6.5	Interpretación Detallada y Caracterización de los Segmentos Latentes	144
6.6	Guía de Aplicación Práctica y Workflow de Segmentación en SmartPLS	147

6.7	Implicaciones gerenciales: El valor estratégico de la segmentación no observada	149
	DE LOS DATOS A LA ESTRATEGIA	154
7.1	Fundamentos de la ciencia de datos y el rol del PLS-SEM en el analytics empresarial	156
7.2	Internet del comportamiento y su relevancia para el modelado	160
7.3	Preprocesamiento de Big data para el PLS-SEM	161
7.4	Flujograma de preprocesamiento de Big Data	165
7.5	El Modelo PLS-SEM como Herramienta de Data Storytelling	169
7.6	Rigor y replicabilidad: Integración de Output de PLS-SEM	171
7.7	Traducción de hallazgos: De los coeficientes estructurales a las decisiones estratégicas	173
7.8	Comunicación gerencial: Diseño de dashboards y visualizaciones estratégicas	175
7.9	Ética en el modelado causal: Implicaciones de la IA y el uso responsable de PLS-SEM	177
7.10	Artículos sobre sobre las implicaciones de la IA, big data, internet de las cosas (IoT) e internet del comportamiento en PLS-SEM en revistas indexadas.	178
	EPÍLOGO	188
	ACRÓNIMOS	190
	GLOSARIO	191

INTRODUCCIÓN

La frontera de la investigación: de la confirmación a la predicción estratégica

"La ciencia no se trata de confirmar lo que ya creemos saber, sino de tener el coraje de explorar las anomalías que desafían nuestros modelos."

La investigación en ciencias empresariales ha llegado a un punto de inflexión irreversible. Durante décadas, nos hemos conformado con modelos que asumen un mundo ordenado: relaciones perfectamente lineales, muestras homogéneas y poblaciones que se comportan como un promedio monolítico. Sin embargo, la realidad de los mercados, el comportamiento del consumidor y la dinámica organizacional es intrínsecamente desordenada, no lineal y profundamente heterogénea.

Este libro, "Modelado Avanzado con PLS-SEM", nace de una necesidad crítica: la caja de herramientas convencional ya no es suficiente.

Si en el pasado el objetivo era obtener un índice de ajuste aceptable para publicar un paper, hoy el imperativo es la relevancia predictiva. De nada sirve un modelo teóricamente perfecto si falla estrepitosamente al intentar anticipar el comportamiento de un cliente fuera de la muestra. Estamos transitando de un paradigma puramente explicativo (centrado en

el R^2 y la covarianza) a un paradigma predictivo (centrado en la varianza y la generalización), donde la técnica PLS-SEM se erige no como una alternativa "suave", sino como el estándar robusto para la complejidad.

A lo largo de estas páginas, desafiaremos las simplificaciones habituales. Nos adentraremos en la "caja negra" de la heterogeneidad no observada, utilizando algoritmos de segmentación latente como FIMIX-PLS y PLS-POS para descubrir subgrupos de datos que permanecen invisibles a los ojos de la estadística tradicional. Cuestionaremos la linealidad de las hipótesis, aprendiendo a modelar curvas y puntos de inflexión que revelan la verdadera naturaleza de los rendimientos decrecientes en la gestión.

Lejos de las restricciones del software comercial, aquí abrazamos la reproducibilidad y la integración con las fronteras de la Ciencia de Datos, la Inteligencia Artificial y la ética en el manejo de información.

Este texto no es para quien busca el camino fácil. Es para el investigador que entiende que la verdadera ventaja competitiva tanto académica como gerencial reside en dominar los matices que otros ignoran.

Bienvenidos a la arquitectura avanzada de la investigación moderna.

Manuel Anchapuri

Objetivos del Texto

Proveer al investigador avanzado de un marco metodológico y técnico para dominar la aplicación de Modelos de Ecuaciones Estructurales (PLS-SEM) en escenarios de alta complejidad, capacitándolo para gestionar la no linealidad, descubrir la heterogeneidad no observada y validar el poder predictivo de sus teorías bajo los estándares de la ciencia de datos contemporánea.

Objetivos Específicos

1. **Fundamentar el Paradigma Predictivo:** Establecer las bases epistemológicas que justifican la transición desde los modelos basados en covarianza (CB-SEM) hacia los enfoques basados en varianza (PLS-SEM), priorizando la capacidad de predicción sobre la confirmación teórica pura en entornos de incertidumbre empresarial.
2. **Validar Empíricamente la Medición (CTA-PLS):** Superar la distinción teórica subjetiva entre constructos reflexivos y formativos mediante la aplicación del Análisis Tetrádico Confirmatorio (CTA-PLS), permitiendo que la naturaleza de los datos dicte la especificación correcta del modelo de medida.
3. **Desentrañar la Heterogeneidad No Observada:** Diagnosticar y tratar la existencia de subgrupos ocultos en los datos (segmentación latente) que pueden sesgar los resultados globales, utilizando

técnicas avanzadas de mixtura finita (FIMIX-PLS) y segmentación orientada a la predicción (PLS-POS).

4. **Modelar la No Linealidad de los Mercados:** Habilitar al analista para detectar y estimar relaciones no lineales (efectos cuadráticos o cúbicos) entre variables, reconociendo que los fenómenos de negocio (como la satisfacción y el rendimiento) raramente siguen trayectorias perfectamente rectas.
5. **Integrar SEM con la Era del Big Data e IA:** Conectar la metodología PLS-SEM con las fronteras de la inteligencia artificial y el análisis de datos masivos (Big Data/IoT), discutiendo las implicaciones éticas y prácticas de aplicar estos modelos en la gobernanza de datos y la toma de decisiones automatizada.

Fredy Quispe Gómez.

PRÓLOGO

Más Allá del Promedio: La Búsqueda de la Estructura Oculta

Durante mucho tiempo, la investigación en ciencias sociales y empresariales ha operado bajo una tregua silenciosa con la realidad. Para hacer nuestros modelos manejables, aceptamos simplificaciones drásticas: asumimos que las relaciones entre variables son líneas rectas, que nuestras muestras son grupos homogéneos y que el objetivo final es confirmar una teoría preexistente, no predecir un comportamiento futuro.

Este libro, "**Modelado Avanzado con PLS-SEM**", rompe esa tregua.

Nos encontramos en una era donde la generación de datos ha superado nuestra capacidad teórica para explicarlos. Los fenómenos que estudiamos la lealtad del cliente, la adopción tecnológica, la cultura organizacional no son lineales ni uniformes. Son ecosistemas complejos, llenos de umbrales, rendimientos decrecientes y subgrupos latentes que escapan a la detección de las técnicas convencionales. Seguir aplicando los mismos modelos de hace veinte años en el entorno actual no es solo anacrónico; es una oportunidad perdida.

Esta obra es una invitación a mirar debajo de la superficie del "promedio". A través de la metodología PLS-SEM ejecutada en R, exploramos territorios que a menudo se evitan por su dificultad técnica: la heterogeneidad no observada y la no linealidad.

¿Por qué un producto tiene éxito en un segmento y fracasa en otro, aunque el modelo global sugiera un efecto positivo? La respuesta suele estar oculta en la heterogeneidad, y aquí aprenderemos a desenterrarla mediante algoritmos de mixtura finita (FIMIX-PLS) y segmentación orientada a la predicción (PLS-POS).

¿Hasta qué punto es cierto que "más es mejor"? Cuestionaremos la linealidad asumiendo que las relaciones empresariales a menudo siguen curvas cuadráticas, donde el exceso de un atributo puede volverse contraproducente.

Pero más allá de la técnica, este texto propone un cambio de mentalidad. Abogamos por una convergencia necesaria entre la econometría tradicional y la Ciencia de Datos moderna. Al priorizar el poder predictivo sobre el ajuste del modelo, alineamos la investigación académica con las demandas de la práctica gerencial, donde la anticipación precisa vale más que la explicación elegante.

El lector no encontrará aquí atajos. En su lugar, encontrará un rigor metodológico sin concesiones y la libertad del código abierto. Es hora de que nuestros modelos dejen de ser caricaturas simplificadas del mundo y empiecen a reflejar su verdadera, caótica y fascinante complejidad.

Fredy Quispe Gómez
Perú, febrero de 2026

CAPITULO I

FUNDAMENTOS DEL SEM: FILOSOFÍA Y ALCANCE

Este capítulo establece los cimientos conceptuales del modelado de ecuaciones estructurales (SEM), resaltando su importancia estratégica para la investigación en ciencias empresariales. El enfoque presenta la evolución de las técnicas de primera y segunda generación, así como la transición metodológica desde los modelos basados en covarianza hacia los enfoques predictivos basados en la varianza. Estas aproximaciones han permitido a los investigadores analizar constructos latentes con una precisión cada vez mayor, articulando teoría y evidencia empírica en entornos caracterizados por la complejidad y la incertidumbre del mercado.

La discusión incorpora los fundamentos teóricos que sustentan el SEM y examina las implicaciones gerenciales de su aplicación. Se profundiza en la pertinencia de estas técnicas para la toma de decisiones basada en datos, resaltando su utilidad en la comprensión de fenómenos organizacionales multidimensionales, así como en el diseño de modelos explicativos y predictivos robustos.

1.1 Definición y alcance del modelado de ecuaciones estructurales

El SEM es una familia de técnicas estadísticas multivariantes de segunda generación, extremadamente versátil, que permite a los investigadores especificar, estimar y evaluar modelos de relaciones causales entre variables latentes (constructos) y variables observadas (indicadores).

A diferencia de la regresión múltiple tradicional, que solo considera relaciones entre variables observadas, el SEM integra dos componentes clave en un solo marco analítico:

1. **Modelo de medición (*Outer Model*):** Define la relación entre las variables latentes (no observables directamente, como la "Satisfacción del cliente" o el "Liderazgo transformacional") y sus indicadores observados (ítems del cuestionario). Su objetivo es evaluar la validez y la fiabilidad de las medidas.
2. **Modelo estructural (*Inner Model*):** Especifica las relaciones causales hipotetizadas entre los constructos latentes, permitiendo evaluar la dirección, magnitud y significancia estadística de los vínculos teóricos

La capacidad del SEM para modelar y corregir el error de medición “un problema endémico en la investigación de encuestas”, al tratar las variables latentes como libres de error, le otorga una ventaja considerable en la prueba de hipótesis complejas (Hair et al., 2022).

1.2 De la regresión al SEM (Técnicas de 1ra vs. 2da Generación)

El análisis estadístico es una herramienta indispensable para los investigadores en el ámbito empresarial y de ciencias sociales. Con el avance del *hardware* y *software* informático, las aplicaciones de los métodos estadísticos se han expandido significativamente, permitiendo el acceso a metodologías más sofisticadas (Hair et al., 2019).

Inicialmente, la investigación se apoyaba en las técnicas de primera generación, como la regresión múltiple, la regresión logística y el análisis de varianza (ANOVA). Si bien estos métodos han sido fundamentales, comparten limitaciones críticas cuando el objetivo es estimar las relaciones entre medidas de conceptos teóricos (Haenlein y Kaplan, 2004). Estas deficiencias son cruciales, ya que las ciencias sociales abordan rutinariamente conceptos no observables como *percepciones*, *actitudes* e *intenciones*.

Los métodos de primera generación presentan tres limitaciones fundamentales (Haenlein y Kaplan, 2004):

1. **Estructura simple:** La presunción de una estructura de modelo simple que no permite el análisis simultáneo de múltiples relaciones de dependencia.
2. **Observabilidad total:** La suposición de que todas las variables son directamente observables, lo cual es incompatible con el trabajo de variable latentes teóricos o latentes.

3. **Ausencia de error de medición:** La asunción de que todas las variables se miden sin error, ignorando el error de escala inherente a la investigación de encuestas.

Para superar estas barreras, los investigadores han recurrido a las técnicas de segunda generación, colectivamente conocidas como modelado de ecuaciones estructurales (SEM). SEM permite modelar y estimar simultáneamente relaciones complejas entre múltiples variables, al tiempo que tiene en cuenta el error de medición en las variables observadas. En consecuencia, este procedimiento proporciona una medida más precisa de los conceptos teóricos relevantes (Cole y Preacher, 2014).

1.3 Anatomía del Modelo

1.3.1 Tipos de variables SEM

En el contexto de un PLS-SEM, las variables se distinguen por su naturaleza de medición (observabilidad) y por su rol causal dentro de la estructura teórica.

Tabla 1

Variables según su nivel de observación

Tipo de Variable	Concepto	Rol y Ejemplo
Variable Observada o Indicador	Variables que se miden directamente en los sujetos de estudio (datos primarios). Son los ítems o preguntas de un cuestionario.	Actúan como representaciones empíricas de los constructos. Ejemplo: Una pregunta específica en una escala Likert.

Tipo de Variable	Concepto	Rol y Ejemplo
Variable Latente (Constructo)	Concepto teórico no observable directamente. Es la característica subyacente que se desea medir, la cual se estima a partir de la varianza común de sus indicadores.	Representa el constructo teórico y se asume que está libre de error aleatorio (ya que el error se modela en el indicador). Ejemplo: La Satisfacción del Cliente, El clima laboral.
Error de Medición ϵ (epsilon)	Componente de error asociado a cada variable observada (indicador).	Representa la varianza única de cada indicador que no es explicada por la variable latente a la que pertenece.

Tabla 2

Variables según su rol causal (Estructura del Modelo)

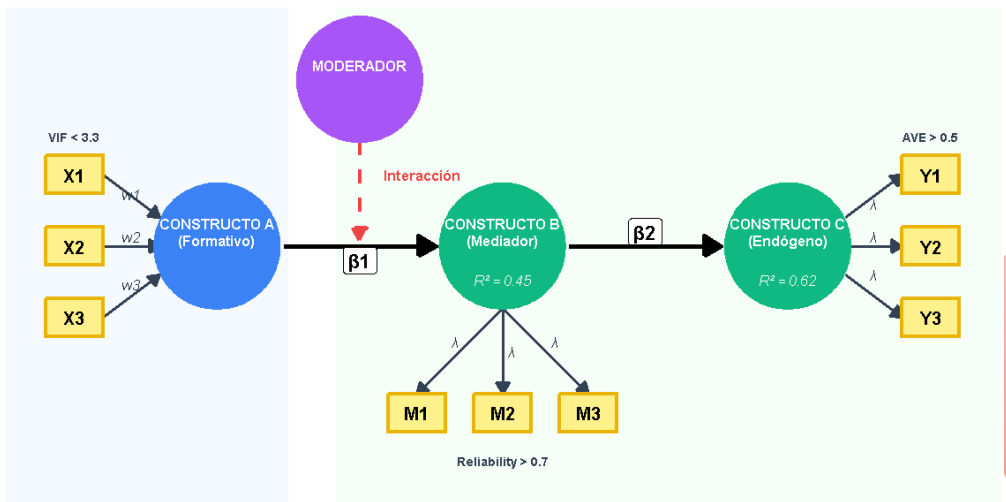
Tipo de Variable	Concepto	Rol y Características
Variable Exógena	Variable que inicia una cadena causal. Afecta a otras variables dentro del modelo, pero no recibe ningún efecto de ellas.	Actúan como variables independientes. Son los puntos de partida del modelo estructural.
Variable Endógena	Variable que recibe efectos causales de una o más variables (exógenas o endógenas) dentro del modelo.	Actúan como variables dependientes o mediadoras. 1. Toda variable endógena debe estar asociada a un término de error residual.
Error Residual ζ (zeta)	Componente de error asociado a cada variable endógena.	Representa el error de predicción; es decir, la porción de la varianza de la variable endógena que no es explicada por sus variables predictoras en el modelo.

Tipo de Variable	Concepto	Rol y Características
Variable de Agrupación	Variable categórica que define la pertenencia de los sujetos a diferentes subpoblaciones o grupos discretos.	Se utiliza para comparar las relaciones estructurales entre distintos segmentos. (ej., en Análisis Multi-Grupo o MGA). Ejemplo: Género, Nivel Educativo, Ubicación.

1.3.2 Arquitectura y componentes de un modelo SEM

El modelo Path representa el "plano maestro" de la investigación en SEM. Esta visualización permite articular la compleja interacción entre lo abstracto (teoría) y lo empírico (datos). La solidez de este diagrama depende de la correcta alineación de cuatro pilares constitutivos: los constructos, los indicadores observables, las rutas de causalidad y los términos de error. A continuación, la Figura 1 desglosa la anatomía técnica de estos componentes.

Figura 1
Elementos básicos de un modelo SEM



Como se muestra en la Figura 1, el modelo SEM-PLS integra de forma simultánea dos niveles analíticos claramente diferenciados: el modelo de medida y el modelo estructural.

El **modelo de medida (outer model)** permite especificar y evaluar la relación entre los constructos latentes y sus indicadores observables, distinguiendo explícitamente la naturaleza causal de cada tipo de constructo. En primer lugar, los constructos formativos, representados en color azul y ubicados en el lado izquierdo del diagrama, se caracterizan porque los indicadores actúan como causas del constructo latente. En este contexto, la evaluación se centra en la significancia y relevancia de los pesos externos (w), así como en el diagnóstico de multicolinealidad entre indicadores, habitualmente mediante el Factor de Inflación de la Varianza (VIF).

En contraste, los constructos reflectivos, identificados en color verde y situados en la zona central y derecha del modelo, asumen una lógica causal inversa, en la que el constructo latente explica la variación de sus indicadores. La evaluación de estos bloques requiere el análisis de la magnitud de las cargas externas (λ lambda), junto con los criterios de fiabilidad interna y validez convergente, particularmente la fiabilidad compuesta y la varianza media extraída (AVE).

El **modelo estructural (inner model)** representa las relaciones causales hipotetizadas entre los constructos latentes, estimadas a través de

los coeficientes de trayectoria (β beta). Este nivel del modelo permite evaluar la estructura causal subyacente y la coherencia teórica del sistema propuesto. La figura ilustra, además, la capacidad del SEM para incorporar variables mediadoras, que reciben y transmiten efectos causales, así como efectos de moderación, representados mediante términos de interacción, en los cuales una tercera variable condiciona la intensidad o dirección de una relación causal.

Finalmente, los valores de R^2 mostrados en los constructos endógenos expresan la proporción de varianza explicada por el modelo, constituyéndose en un criterio fundamental para evaluar su capacidad explicativa y predictiva, en concordancia con los objetivos analíticos del enfoque SEM-PLS

1.3.3 Especificación del Modelo de Medida: Reflectivo vs. Formativo

Es fundamental para la validez del modelo distinguir la dirección de la causalidad entre el constructo latente y sus indicadores, ya que esto determina las pruebas estadísticas de evaluación. La literatura metodológica identifica dos enfoques opuestos (Hair et al., 2022):

- **Modelos Reflectivos (Modo A):** La dirección de la causalidad va desde el constructo hacia los indicadores (Constructo \rightarrow Indicadores). Se asume que los indicadores

son una "manifestación" o reflejo del constructo subyacente. Por ejemplo, la "Satisfacción" causa que un cliente puntúe alto en varias preguntas. En este enfoque, se espera que los indicadores estén altamente correlacionados entre sí y sean intercambiables; eliminar uno no altera el significado del constructo.

- **Modelos Formativos (Modo B):** La dirección de la causalidad va desde los indicadores hacia el constructo (Indicadores → Constructo). Aquí, los indicadores "causan" o forman la variable latente. Por ejemplo, el "Nivel Socioeconómico" está formado por la educación, el ingreso y la ocupación; estos no necesariamente correlacionan (se puede tener mucha educación y bajo ingreso) y no son intercambiables. Omitir un indicador en este modelo altera la naturaleza del constructo, perdiendo una parte de su significado (Sarstedt et al., 2019).

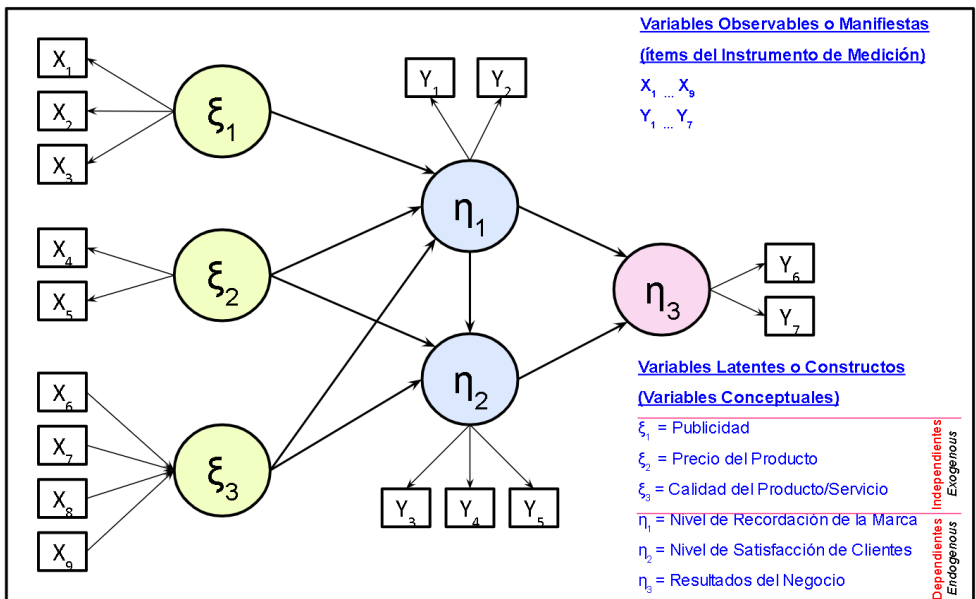
Tabla 3

Elementos gráficos del modelo

Elemento	Representación Gráfica	Definición Teórica
Constructos (Variables Latentes)	Círculos u óvalos	Conceptos teóricos no directamente observables (ej., Satisfacción, Innovación). Pueden ser exógenos (independientes, sin flechas de predicción que les lleguen) o

Elemento	Representación Gráfica	Definición Teórica
		endógenos (dependientes, con flechas que les llegan).
Variables Medidas (Indicadores)	Rectángulos	Observaciones directamente medidas (datos brutos), también conocidas como variables manifiestas.
Relaciones (Hipótesis)	Flechas de una sola punta	Indican una relación predictiva o causal entre constructos o entre un constructo y sus indicadores. Estas conexiones se fundamentan en la lógica y la teoría estructural.
Términos de Error	Flechas de una sola punta que apuntan al elemento	Representan la varianza que el modelo no puede explicar.

Figura 2
Modelo conceptual SEM



Fuente: Saravia (2024) Universidad del Pacifico PIE

En la Figura 2. Modelo conceptual SEM presenta la estructura teórica propuesta para la investigación, operativizando la revisión de literatura en una red nomológica contrastable. Se visualizan las relaciones de dependencia entre los constructos exógenos (predictores) y endógenos (resultado), estableciendo visualmente las hipótesis (H_1, H_2...) que serán sometidas a prueba estadística mediante el algoritmo PLS

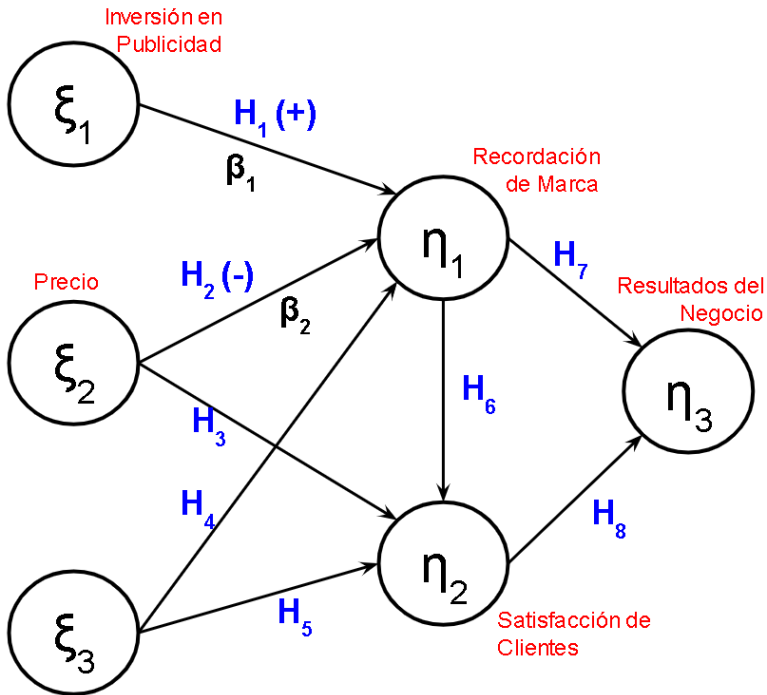
1.3.4 Términos de Error

Los términos de error son esenciales para el rigor estadístico, pero su ubicación depende del tipo de variable:

- Existen para constructos endógenos, ya que su varianza es parcialmente explicada por otros constructos.
- Existen para indicadores medidos de forma reflectiva, pues el constructo no explica el 100% de la varianza del indicador.
- No existen para constructos exógenos o indicadores formativos, ya que estos últimos son las causas o definidores del constructo, y su error se gestiona mediante la colinealidad.

Figura 3

SEM: Modelo Teórico e Hipótesis



Nota. Modelo teórico simplificado Saravia (2024) Universidad del Pacifico PIE

En la Figura 3. SEM: Modelo Teórico e Hipótesis se aprecia el diagrama formaliza la red de causalidad propuesta, identificando explícitamente las hipótesis de investigación (H_1, H_2, \dots, H_n) que vinculan a los constructos. Esta simplificación del modelo de Saravia (2024) delimita el alcance de la inferencia estadística, especificando la dirección esperada (signo positivo o negativo) de los coeficientes path (beta) que serán estimados en el análisis estructural.

Ejemplos de hipótesis del modelo

- H₁: “La inversión en Publicidad influye de manera positiva y significativa en la Recordación de Marca”
- H₂: “El Precio influye de manera negativa y significativa en la Recordación de Marca”
-
- H₆: “La Recordación de Marca influye significativamente de manera directa, e indirecta con la Satisfacción como variable mediadora, en los Resultados”
- H₇: “La Recordación de Marca influye de manera positiva y significativa en los Resultados”

1.4 Fundamentos teóricos del SEM

La motivación para la adopción generalizada de los modelos SEM en la investigación de negocios es doble: radica en su rigor teórico para la prueba de hipótesis y en su aplicabilidad práctica para la toma de decisiones.

Rigor en la prueba de teorías.

Históricamente, la regresión y el análisis de varianza han limitado la capacidad de los investigadores para testear redes complejas de causalidad. El SEM, en cambio, permite:

- **Modelar múltiples relaciones simultáneamente:** Es posible evaluar cómo un conjunto de variables exógenas influye en múltiples variables endógenas al mismo tiempo. Por ejemplo, cómo la cultura organizacional impacta la innovación y esta, a

su vez, afecta el desempeño financiero, mientras que el liderazgo modera una de estas relaciones (Sarstedt et al., 2020).

- **Análisis de mecanismos (mediación y moderación):** El SEM es la técnica predilecta para desentrañar los procesos causales (*cómo y cuándo* ocurren los efectos). La mediación explica los mecanismos intervinientes, mientras que la moderación explica las condiciones contextuales que amplifican o reducen los efectos (Hayes, 2018; Preacher & Hayes, 2008).
- **Gestión del error de medición:** Al especificar las variables latentes, el SEM permite al investigador aislar el error estocástico de las relaciones estructurales, mejorando la fiabilidad de las inferencias causales y la validez de las variables latentes, una necesidad crítica en las ciencias sociales (Bagozzi y Yi, 2012).

1.5 Paradigmas del SEM: Covarianza (CB-SEM) vs. Varianza (PLS-SEM)

En la literatura académica sobre Modelos de Ecuaciones Estructurales (SEM), existe una distinción fundamental que a menudo confunde al investigador novel. El término "SEM" no se refiere a una única técnica estadística, sino que engloba dos escuelas de pensamiento o paradigmas distintos que, aunque comparten el objetivo de analizar relaciones estructurales, difieren radicalmente en sus algoritmos matemáticos, supuestos estadísticos y objetivos de investigación.

Esta dicotomía divide al SEM en dos enfoques principales, cada uno impulsado por una motivación causal distinta: el SEM basado en

Covarianza (CB-SEM) y el SEM basado en Varianza, comúnmente conocido como Mínimos Cuadrados Parciales (PLS-SEM).

1.5.1 El Enfoque Confirmatorio: CB-SEM (Covariance-Based)

El modelado basado en covarianza es el enfoque clásico, popularizado por Jöreskog (1973). Su filosofía se centra en la confirmación y verificación teórica.

- **Lógica Estadística:** Su algoritmo, generalmente de Máxima Verosimilitud (Maximum Likelihood), intenta minimizar la discrepancia entre la matriz de covarianza observada en los datos y la matriz de covarianza teórica implícita en el modelo hipotético. Conceptualmente, considera las variables latentes como factores comunes que explican la covariación entre sus indicadores.
- **Objetivo:** Se utiliza para determinar qué tan bien un modelo teórico *ajusta* a los datos. Es el enfoque ideal cuando el objetivo es confirmar teorías ya establecidas, comparar modelos competitivos o rechazar hipótesis estrictas (Hair et al., 2022).
- **Requisitos:** Exige normalidad multivariante estricta, tamaños de muestra grandes y se adhiere preferentemente a modelos de medición reflexivos.

1.5.2 El Enfoque Predictivo: PLS-SEM (Partial Least Squares)

Por otro lado, el PLS-SEM es un enfoque "causal-predictivo" desarrollado para situaciones donde la teoría es menos madura o el objetivo es la aplicación práctica.

- **Lógica Estadística:** Es una técnica basada en compuestos, lo que implica que las variables latentes se estiman como combinaciones lineales de sus indicadores. A diferencia del CB-SEM, no busca reproducir la matriz de covarianza, sino maximizar la varianza explicada (R^2) de las variables endógenas clave.
- **Objetivo:** Su foco se coloca en el poder predictivo. La validación del modelo se desplaza del ajuste global hacia la capacidad para predecir nuevos datos (out-of-sample predictive power). Es especialmente popular en ciencias empresariales por su capacidad para identificar "drivers" clave de resultados como la lealtad o la rentabilidad.
- **Flexibilidad:** Destaca por su robustez con datos no normales, su capacidad para manejar constructos formativos y su eficacia con tamaños de muestra pequeños o medianos (Sarstedt et al., 2019; Hair et al., 2022).

1.5.3 Marco de Decisión

La elección entre CB-SEM y PLS-SEM no debe ser arbitraria, sino basarse en la naturaleza de la investigación y los datos disponibles. Hair et al. (2019) proponen los siguientes criterios para guiar esta decisión:

Tabla 4

Criterios de Selección Metodológica

Criterio	CB-SEM (Confirmatorio)	PLS-SEM (Predictivo)
Objetivo	Contrastar teorías fuertes, confirmar modelos o evaluar ajuste global.	Predecir constructos clave, desarrollar teoría exploratoria o identificar factores determinantes.
Normalidad de Datos	Requiere normalidad multivariante (curva de campana).	No requiere normalidad (robusto ante datos asimétricos).
Tipo de Medidas	Preferentemente constructos reflexivos.	Maneja constructos reflexivos y formativos indistintamente.
Tamaño de Muestra	Requiere muestras grandes para potencia estadística ($N > 200-300$).	Funciona eficazmente con muestras pequeñas, aunque requiere potencia adecuada.
Complejidad	Limitado a modelos pequeños o medianos (< 50 ítems).	Capaz de gestionar modelos muy complejos (100+ ítems).
Prevalencia	Psicología, Sociología (Ciencias del Comportamiento).	Marketing, Estrategia, Sistemas de Información (Ciencias Empresariales).

Nota. Adaptado de Hair et al. (2019) y Rigdon et al. (2017).

1.6 Justificación del Enfoque PLS-SEM en ciencias empresariales

1.6.1 PLS-SEM en Ciencias Empresariales

Dado que esta obra se centra en las ciencias empresariales y la toma de decisiones gerenciales, el enfoque metodológico predominante en los capítulos siguientes será el PLS-SEM. Esta decisión se fundamenta en tres realidades de la investigación aplicada a negocios:

1. Los datos empresariales (ventas, satisfacción, percepción) rara vez siguen una distribución normal perfecta.
2. Los modelos de gestión suelen utilizar constructos formativos (ej. el "Éxito Empresarial" está *formado* por la rentabilidad y el crecimiento, no *reflejado* por ellos), los cuales son difíciles de modelar en CB-SEM.
3. El interés del gerente o consultor reside en la predicción (R^2) y la priorización de recursos (IPMA), capacidades nativas del PLS-SEM.

1.6.2 La importancia de predecir constructos

El foco de PLS-SEM se coloca en el poder predictivo del modelo. En lugar de replicar la estructura de covarianza, el método busca maximizar la varianza explicada (R^2) en las variables endógenas y evaluar la capacidad del modelo para predecir nuevos

datos. En este sentido, la validación del modelo se desplaza del ajuste a la capacidad predictiva.

Para reflejar este enfoque, Shmueli et al. (2016) han propuesto procedimientos de evaluación más recientes y apropiados, diseñados específicamente para la naturaleza orientada a la predicción de PLS-SEM, como el uso de la capacidad predictiva fuera de la muestra (*out-of-sample predictive power*). Estos nuevos métodos son cruciales para validar los resultados de PLS-SEM.

1.7 Aplicabilidad gerencial y estratégica

La capacidad del SEM para modelar procesos complejos se traduce directamente en inteligencia estratégica para la toma de decisiones:

- **Identificación de palancas estratégicas:** Al cuantificar los coeficientes de sendero, el SEM revela cuáles son las variables exógenas o mediadoras con mayor impacto en un resultado empresarial clave (ej., la lealtad del cliente, la retención de empleados o el desempeño innovador).
- **Validación de modelos de negocio:** Permite validar empíricamente los modelos lógicos de las empresas. Por ejemplo, un modelo que postula que la formación lleva a la competencia técnica y está a la productividad puede ser evaluado por completo en un solo modelo SEM (Shmueli et al., 2019).
- **Análisis de inversión:** Los resultados de SEM guían la asignación de recursos. Si se demuestra que la "calidad del servicio" tiene un

efecto causal significativo y fuerte sobre la "intención de recompra", la gerencia tiene justificación empírica para priorizar la inversión en la mejora de la calidad.

El SEM, en cualquiera de sus variantes (CB-SEM o PLS-SEM), ofrece el marco más completo para testear la plausibilidad de las teorías, garantizando que las conclusiones extraídas del análisis no solo sean estadísticamente sólidas, sino también estratégicamente accionables (Hair et al., 2022). Su dominio es fundamental para la investigación empírica de vanguardia en la disciplina empresarial.

1.8 Entorno informático SEM

Si bien históricamente el análisis de ecuaciones estructurales ha estado dominado por software comercial propietario (como SmartPLS, IBM Amos o LISREL), la tendencia actual en la investigación de alto impacto se inclina hacia la reproducibilidad computacional y el uso de herramientas de código abierto.

Para la implementación efectiva de los conceptos abordados en este texto, se establecen los siguientes estándares en el entorno R:

1.8.1 PLS-SEM (Enfoque basado en la varianza)

- **Paquete Recomendado:** `seminr` (SeminR).
- **Justificación:** Desarrollado en colaboración con los principales teóricos del método (Hair, Sarstedt, Ringle), `seminr` ofrece la sintaxis más intuitiva y robusta para

especificar modelos formativos y reflexivos. A diferencia de las interfaces gráficas rígidas, permite una personalización avanzada de los algoritmos de estimación, bootstrapping y evaluación predictiva (PLSpredict), integrando las métricas más recientes de la literatura metodológica.

1.8.2 CB-SEM (Enfoque basado en la covarianza)

- **Paquete Recomendado:** lavaan (**Latent Variable Analysis**).
- **Justificación:** Es el estándar en R para el modelado confirmatorio. Su capacidad para replicar resultados de software como Mplus o Amos, sumado a su gratuidad y flexibilidad para manejar estimadores robustos (MLR) ante la falta de normalidad, lo convierte en la herramienta indispensable para análisis de invarianza y modelos de medida estrictos.

Tabla 5

Matriz de selección de software en R

Objetivo del Análisis	Enfoque Metodológico	Paquete de R Recomendado	Función Principal
Predicción / Exploración	PLS-SEM (Partial Least Squares)	Library (sempr)	estimate_pls()
Confirmación Teórica	CB-SEM (Covariance-Based)	Library (lavaan)	cfa() o sem()

Objetivo del Análisis	Enfoque Metodológico	Paquete de R Recomendado	Función Principal
Diagramas de Senderos	Visualización Gráfica	Library (semPlot)	semPaths()

En el capítulo se ha establecido los cimientos conceptuales necesarios para comprender el Modelado de Ecuaciones Estructurales (SEM) no solo como una herramienta estadística avanzada, sino como un enfoque metodológico indispensable para la investigación en ciencias empresariales. Hemos transitado desde las limitaciones de las técnicas de primera generación hacia la flexibilidad del SEM, destacando su capacidad única para modelar variables latentes y errores de medición simultáneamente (Hair et al., 2022).

La discusión central ha evidenciado la dicotomía entre el enfoque basado en la covarianza (CB-SEM), orientado a la confirmación teórica, y el enfoque basado en la varianza (PLS-SEM), centrado en la predicción. Para el ámbito de la gestión y los negocios, donde la incertidumbre del mercado exige herramientas capaces de anticipar resultados, el enfoque predictivo del PLS-SEM se revela como la opción más robusta. Al priorizar la maximización de la varianza explicada (R^2) de los constructos dependientes, el investigador puede ofrecer "implicaciones gerenciales" que van más allá de la significancia estadística, aportando valor real a la toma de decisiones estratégicas (Sarstedt et al., 2019).

Finalmente, la integración de estos fundamentos teóricos con un entorno informático moderno, basado en el lenguaje R, prepara el terreno

para la aplicación práctica. El dominio de la sintaxis y la lógica del modelo estructural no es un fin en sí mismo, sino el medio para desenmarañar la complejidad de los fenómenos organizacionales. En los capítulos subsiguientes, trasladaremos estos conceptos abstractos como la distinción entre modelos reflexivos y formativos o la gestión de variables exógenas y endógenas hacia la ejecución técnica directa, permitiendo al lector transformar hipótesis teóricas en modelos empíricos validados..

CAPITULO II

DISEÑO DE INVESTIGACIÓN Y ESTRATEGIA MUESTRAL

Una vez definido el paradigma metodológico y justificada la elección entre el enfoque basado en la covarianza (CB-SEM) o en la varianza (PLS-SEM), tal como se discutió en el capítulo precedente, el investigador se enfrenta a una fase crítica que precede a cualquier recolección de datos: la arquitectura del diseño de investigación. La eficacia del Modelado de Ecuaciones Estructurales (SEM) no reside únicamente en la potencia de sus algoritmos matemáticos, sino en la solidez de las decisiones estratégicas tomadas durante la planificación del estudio (Hair et al., 2022).

En el ámbito de las ciencias empresariales y del comportamiento, los fenómenos rara vez ocurren en condiciones de laboratorio controladas. Por el contrario, la heterogeneidad inherente a los entornos competitivos y organizacionales introduce niveles de "ruido" y variabilidad que exigen diseños muestrales rigurosos (Sarstedt et al., 2022). En este contexto, la selección de la muestra deja de ser un mero trámite logístico para convertirse en un imperativo metodológico. Un diseño muestral deficiente no solo limita la capacidad de generalización estadística, sino que

compromete la estabilidad de las estimaciones de los parámetros y, fundamentalmente, la validez predictiva del modelo (Shmueli et al., 2019).

En consecuencia, este capítulo sostiene que la insuficiencia muestral ya sea por tamaño o por composición no constituye simplemente una limitación técnica subsanable, sino un error metodológico de origen que amenaza la reproducibilidad de la ciencia (Kock y Hadaya, 2018). A lo largo de las siguientes secciones, se desmitificará la premisa de que PLS-SEM es una "bala de plata" para muestras extremadamente pequeñas y se establecerán los protocolos de potencia estadística necesarios para garantizar que los hallazgos empíricos posean la robustez requerida para la toma de decisiones gerenciales.

2.1 Decisión analítica: el marco estratégico

La elección entre el modelado basado en la covarianza (CB-SEM) y el modelado basado en la varianza (PLS-SEM) no constituye una decisión arbitraria ni meramente técnica; es un imperativo estratégico que condiciona todo el diseño muestral. Antes de calcular el tamaño de la muestra, el investigador debe definir la naturaleza epistemológica de su estudio, ya que cada enfoque persigue objetivos diametralmente opuestos (Hair et al., 2019).

Si el objetivo primordial es la confirmación teórica, el diseño debe orientarse hacia el CB-SEM. Este enfoque exige una teoría preexistente sólida y establecida. Su meta analítica es evaluar el ajuste global del modelo; es decir, confirmar si la matriz de covarianza teórica se ajusta perfectamente a la observada en los datos. En consecuencia, el diseño muestral bajo esta ruta debe ser riguroso, satisfaciendo supuestos estrictos de normalidad multivariante y requiriendo tamaños de muestra grandes para garantizar la estabilidad de los parámetros.

Por el contrario, si el objetivo es la predicción y el desarrollo teórico, el diseño debe alinearse con el PLS-SEM. Aquí, la prioridad se desplaza del ajuste global hacia la maximización de la varianza explicada (R^2) de los constructos endógenos y la identificación de sus impulsores clave. En este contexto, el PLS-SEM opera como una "*fuerza de ataque rápida y flexible*": su algoritmo no depende de la normalidad estricta de los datos y mantiene una robustez notable incluso con muestras más limitadas (Hair et al., 2022).

Esta distinción es crítica para investigadores que operan en mercados emergentes, evalúan tecnologías disruptivas o analizan modelos de negocio innovadores donde la teoría es incipiente. En tales escenarios, donde la incertidumbre es alta y la teoría escasa, la capacidad del PLS-SEM para predecir el rendimiento del fenómeno se convierte en la ventaja metodológica decisiva frente a la rigidez confirmatoria del CB-SEM

Tabla 6

CB-SEM vs, PLS-SEM

Criterio de Decisión	CB-SEM (Basado en Covarianzas)	PLS-SEM (Mínimos Cuadrados Parciales)
Normalidad de los datos	Exigente: Requiere la suposición de distribución normal multivariante.	No Paramétrico: No requiere supuestos de distribución sobre los datos (útil con datos no normales o Likert).
Tamaño de la muestra	Muestras grandes: Se requiere un tamaño de muestra mayor para obtener parámetros estables y un buen ajuste global.	Muestras pequeñas: Robusto con muestras pequeñas (aunque siempre se recomiendan $N > 100$). Sigue la regla del $10 \times$ (10 veces el máximo de flechas apuntando a un constructo).
Complejidad del modelo	Preferido para modelos con pocas variables latentes y muchos indicadores por constructo.	Preferido para modelos complejos con muchas variables latentes y múltiples relaciones estructurales.

Nota. Basado en Hair et al. (2019).

2.2 Teoría del muestreo en SEM: calidad y robustez

Uno de los mitos más persistentes en la investigación con Modelos de Ecuaciones Estructurales es la creencia de que el enfoque PLS-SEM funciona como una "bala de plata" capaz de validar modelos complejos con muestras extremadamente pequeñas. Si bien el algoritmo es robusto ante distribuciones no normales, la literatura metodológica reciente advierte

sobre la "falacia del muestreo inadecuado". Este error metodológico ocurre cuando el investigador interpreta la flexibilidad del PLS-SEM como una licencia para utilizar tamaños muestrales insuficientes, ignorando la complejidad estructural del modelo y la potencia estadística necesaria para detectar efectos significativos (Kline, 2016; Ringle et al., 2022).

Las consecuencias de sucumbir a esta falacia son severas para la integridad del estudio. Una muestra que no cumple con los requisitos mínimos de potencia afecta directamente la estabilidad de los coeficientes path, incrementa la varianza del error y, lo más crítico, compromete la relevancia predictiva del modelo. Como señalan Hair et al. (2021) y Henseler (2023), los resultados derivados de un diseño muestral pobre suelen ser inestables e irreproducibles, anulando la capacidad del modelo para ofrecer inferencias válidas sobre la población objetivo.

Finalmente, es imperativo comprender que la robustez del muestreo trasciende la mera cantidad de observaciones (N). La calidad del dato exige una evaluación integral que incluya la ausencia de sesgos de no respuesta, la consistencia conceptual del marco muestral y la representatividad genuina del grupo de análisis. Incluso con un N elevado, si la recolección de datos carece de rigor en estos factores complementarios, la capacidad de generalización y predicción del modelo se verá fracturada (Sarstedt et al., 2022; Sarstedt y Henseler, 2023)

2.3 Potencia estadística (1-beta) y decisiones metodológicas

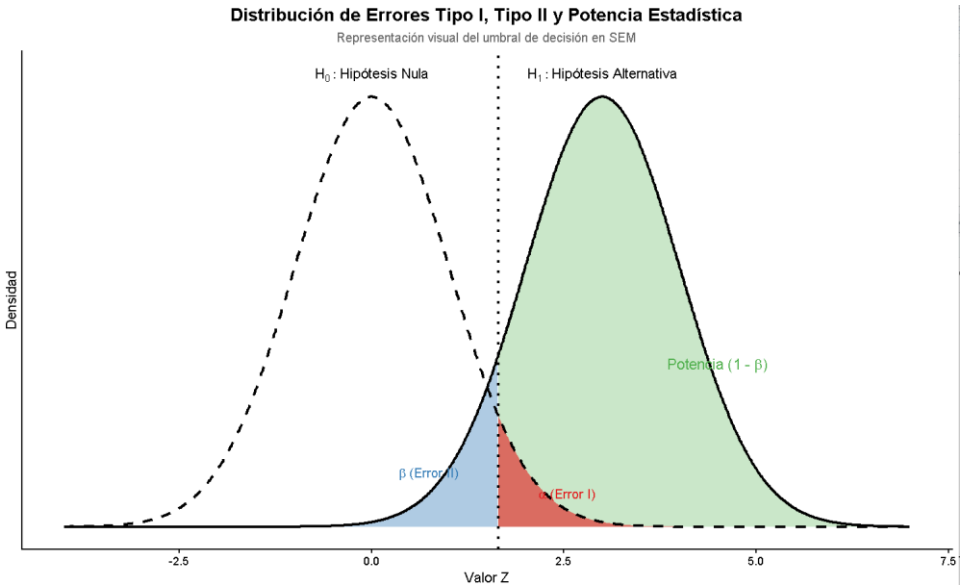
Más allá de cumplir con un número arbitrario de encuestas, el diseño muestral debe obedecer a la lógica de la potencia estadística (1-beta). En términos formales, la potencia se define como la probabilidad condicional de rechazar la hipótesis nula (H_0) cuando esta es realmente falsa; es decir, la capacidad del test para detectar un efecto significativo cuando dicho efecto existe en la realidad (Cohen, 1992).

1.8.3 El Estándar de Oro: El umbral del 80%

La comunidad científica ha adoptado universalmente el criterio propuesto por Jacob Cohen, estableciendo una potencia mínima del 80% (0.80) como el estándar de rigor. Esto implica que el investigador acepta un riesgo máximo del 20% ($\beta = 0.20$) de cometer un **Error Tipo II** (falsos negativos), o lo que es lo mismo, fallar en detectar una relación que verdaderamente existe. Operar por debajo de este umbral convierte la investigación en un ejercicio de azar, donde la ausencia de significancia estadística puede deberse simplemente a la falta de sensibilidad de la muestra y no a la ausencia del fenómeno.

Figura 4

Interacción entre Errores Estadísticos y Potencia (1-beta)



La gráfica ilustra el dilema fundamental de la inferencia estadística en SEM: la capacidad de distinguir entre el "ruido" (Hipótesis Nula, H_0) y la "señal" verdadera (Hipótesis Alternativa, H_1).

1. **El Umbral de Decisión (Línea Punteada Vertical):** Representa el valor crítico (generalmente asociado a un $\alpha = 0.05$). Cualquier estadístico t o z que caiga a la derecha de esta línea se considera "significativo".
2. **Zona Roja (alpha - Error Tipo I):** Es la probabilidad de "Falso Positivo". Ocurre cuando el investigador rechaza la hipótesis nula (afirma que hay un efecto) cuando en realidad no existe. Este error se controla fijando el nivel de significancia (usualmente al 5%).

3. **Zona Azul (beta - Error Tipo II):** Es la probabilidad de "Falso Negativo". Representa el riesgo de que el efecto exista realmente (estamos bajo la curva sólida H_1), pero nuestros datos no sean lo suficientemente contundentes para cruzar el umbral crítico. Este es el riesgo principal de usar muestras pequeñas en PLS-SEM: la zona azul crece y oscurece la verdad.

4. **Zona Verde (Potencia Estadística, 1-beta):** Es la probabilidad de éxito. Representa el área bajo la curva alternativa que supera el umbral crítico. Un diseño muestral robusto busca maximizar esta área verde hasta alcanzar, al menos, el 80%.

Para aumentar el área verde (Potencia) y reducir la azul (Error II), el investigador tiene dos caminos matemáticos: aumentar la magnitud del efecto (separar las curvas) o, lo que es más controlable, aumentar el tamaño de la muestra para "adelgazar" las curvas (reducir el error estándar), minimizando así el solapamiento entre las hipótesis.

1.8.4 Implicaciones Críticas para PLS-SEM

La potencia cobra una relevancia crítica en el entorno PLS-SEM, donde frecuentemente se modelan fenómenos de ciencias del comportamiento con tamaños del efecto (f^2) moderados o pequeños. Existe una relación inversamente proporcional que el investigador debe gestionar: a medida que aumenta la complejidad del modelo (número de predictores) y disminuye la magnitud del efecto esperado, la potencia estadística cae drásticamente si el tamaño muestral (n) se mantiene constante.

Si el diseño muestral es débil, el modelo PLS-SEM perderá su "agudeza visual", volviéndose incapaz de identificar relaciones estructurales sutiles, pero teóricamente relevantes. Por ello, la recomendación práctica antes de iniciar el trabajo de campo es realizar un análisis de sensibilidad ($G^*Power > Sensitivity Analysis$). Este procedimiento permite calcular cuál es el tamaño del efecto mínimo que la muestra disponible será capaz de detectar con una potencia del 80%, ajustando así las expectativas teóricas a la realidad empírica del estudio (Hair et al., 2021)

2.4 Métodos de estimación del tamaño muestral

La determinación del tamaño de la muestra en PLS-SEM no es un procedimiento monolítico, sino una decisión jerárquica que depende del nivel de rigor exigido por la investigación. La literatura actual permite clasificar las estrategias de estimación en tres niveles de complejidad ascendente:

2.4.1 Nivel 1: Aproximación Heurística

(La regla de las "10 veces")

Históricamente, este ha sido el método más citado debido a su simplicidad operativa. Propuesto originalmente por Barclay et al. (1995), este criterio establece que el tamaño mínimo de la muestra debe ser 10 veces el número máximo de flechas estructurales

dirigidas a un constructo latente en el modelo (ya sean indicadores formativos o rutas estructurales).

- **Advertencia Metodológica:** Aunque útil para estimaciones preliminares rápidas, la literatura contemporánea (Hair et al., 2022) advierte que este método carece de precisión estadística. No se recomienda como justificación única en investigaciones de alto impacto o tesis doctorales, ya que tiende a subestimar los requerimientos muestrales en modelos complejos, ignorando la potencia estadística y el tamaño del efecto.

2.4.2 Nivel 2: Análisis de Potencia a Priori

(El Estándar Áureo)

Para cumplir con los estándares de las revistas indexadas (Q1/Q2), la estimación debe basarse en el análisis de potencia estadística. Este enfoque utiliza algoritmos, disponibles en software como G*Power, para calcular el N exacto en función de tres parámetros críticos (Cohen, 1992; Faul et al., 2009):

1. **Potencia Estadística (1-beta):** Generalmente fijada en 0.80 (80%), asegurando una alta probabilidad de detectar efectos existentes.

2. **Nivel de Significación (α):** Típicamente 0.05, definiendo el umbral de error tipo I tolerado.
3. **Tamaño del Efecto (f^2):** La magnitud mínima de la relación que el investigador espera detectar (pequeño, mediano o grande).

Determinación del Tamaño del Efecto Esperado (f^2)

A diferencia del nivel de significancia ($\alpha = 0.05$) y la potencia ($1 - \beta = 0.80$), que son convenciones estadísticas fijas, el tamaño del efecto (f^2) es un parámetro que depende de la naturaleza del fenómeno estudiado. Este valor representa la magnitud de la relación que el investigador espera encontrar entre las variables.

Para la estimación *a priori* del tamaño muestral, Cohen (1992) establece tres umbrales de referencia ampliamente aceptados en las ciencias del comportamiento:

- **Efecto Pequeño ($f^2 = 0.02$):** Requiere muestras grandes para ser detectado. Es común en estudios donde múltiples factores influyen levemente en la variable dependiente.
- **Efecto Mediano ($f^2 = 0.15$):** Estándar recomendado cuando no existen antecedentes previos. Representa una influencia visible a "simple vista" estadística.
- **Efecto Grande ($f^2 = 0.35$):** Implica una relación muy fuerte y evidente, detectable con muestras reducidas.

Nota: En ausencia de estudios previos o meta-análisis que sugieran un valor específico, se recomienda utilizar el criterio de **Efecto Mediano ($f^2 = 0.15$)** como punto de partida conservador para el cálculo en G*Power. Asumir un efecto grande sin justificación puede derivar en una subestimación peligrosa del tamaño muestral necesario.

2.4.3 Nivel 3: Simulaciones y criterios de complejidad

En el nivel más alto de sofisticación, especialmente necesario cuando los supuestos distribucionales son severamente violados o la arquitectura del modelo es inusual, se recurre a métodos avanzados:

- **Simulaciones Monte Carlo:** Permiten evaluar la estabilidad de los parámetros bajo miles de escenarios hipotéticos de remuestreo, ofreciendo la justificación más robusta posible (Hair et al., 2021).
- **Criterios de Estructura del Modelo:** Enfoques como el de Marcoulides y Saunders (2006) proponen tablas de referencia que ajustan el tamaño muestral necesario basándose en la complejidad conceptual (número de constructos y relaciones) y la comunalidad de las variables, siendo vitales para modelos de gran escala.

Tabla 7*Comparativa de métodos en PLS-SEM*

Método de Estimación	Ventaja Principal	Desventaja Principal
Regla de las 10 veces (Heurística)	Simplicidad operativa: Permite un cálculo inmediato sin necesidad de software externo. Ideal para verificaciones rápidas.	Falta de precisión: No considera la potencia estadística ni el tamaño del efecto. Es frecuentemente rechazada en revistas de alto impacto.
Análisis de Potencia (GPower / Cohen)*	Rigor Científico: Es el estándar actual. Garantiza el control sobre los errores Tipo I y II, asegurando la validez estadística.	Dependencia de parámetros: Requiere que el investigador estime a priori el tamaño del efecto esperado (f^2), lo cual puede ser difícil sin estudios previos.
Simulaciones Monte Carlo (Complejidad del modelo)	Máxima Robustez: Evalúa la estabilidad de los parámetros bajo miles de escenarios, ideal para datos no normales o modelos muy complejos.	Alta Exigencia Técnica: Requiere mayor capacidad computacional y un conocimiento avanzado de programación o configuración de software.

Nota. Elaboración basada en Hair et al. (2022) y Cohen (1992).

Escenarios de Muestra Fija: El Análisis de Sensibilidad (Post-hoc)

Frecuentemente, el investigador opera bajo restricciones logísticas o presupuestarias que hacen que el tamaño de la muestra sea inelástico (por ejemplo, al trabajar con una base de datos histórica cerrada o una población finita muy reducida). En estos casos, donde no es factible aumentar el N para alcanzar la potencia deseada, el cálculo *a priori* pierde su utilidad práctica.

La alternativa metodológica rigurosa es ejecutar un **Análisis de Sensibilidad** (*Sensitivity Analysis*). Esta función, disponible en software como G*Power (*Compute required effect size given alpha, power, and N*), invierte la lógica de la ecuación: en lugar de preguntar cuántos sujetos se necesitan, el investigador introduce el N fijo disponible para determinar cuál es el tamaño del efecto mínimo (f^2) que el modelo será capaz de detectar con fiabilidad.

- **Criterio de Decisión:** Si el análisis revela que la muestra disponible solo tiene potencia para detectar efectos "grandes" ($f^2 > 0.35$), el investigador debe reconocer explícitamente esta limitación. Esto implica advertir que las relaciones causales sutiles o moderadas podrían existir en la realidad, pero pasarán desapercibidas en el estudio (riesgo de Error Tipo II) debido a la falta de sensibilidad de los datos (Faul et al., 2009).

2.5 Diseño para la investigación exploratoria y predictiva

La elección del diseño muestral y analítico culmina con la alineación entre la técnica y el estado de madurez de la teoría. Históricamente, Herman Wold (1985), el creador del método PLS, enfatizó que la utilidad primordial de este enfoque reside en su capacidad para el "modelado suave" (*soft modeling*), permitiendo al investigador evaluar mejoras tentativas en la estructura del modelo como la adición o supresión

de variables latentes de manera ágil y orientada a la maximización de la capacidad predictiva.

En el contexto contemporáneo de las ciencias empresariales, el diseño basado en PLS-SEM se establece como el estándar preferente para el desarrollo teórico por tres razones metodológicas fundamentales:

1. **Robustez ante la No-Normalidad de los Datos:** La realidad empírica de los negocios rara vez se ajusta a la curva de campana ideal. El uso extensivo de escalas Likert y la asimetría natural de datos financieros o comportamentales violan frecuentemente el supuesto de normalidad multivariante exigido por el CB-SEM. El diseño PLS-SEM gestiona estas distribuciones con una robustez superior, evitando sesgos en la estimación de parámetros sin requerir transformaciones de datos artificiales (Hair et al., 2022).
2. **Focalización en la Relevancia Gerencial (R^2):** Mientras que otros enfoques priorizan el ajuste global (*model fit*), el diseño PLS-SEM responde a la pregunta pragmática de la gestión estratégica: "*¿Qué constructos poseen el mayor poder para predecir el rendimiento, la satisfacción o la lealtad?*". Al centrarse en maximizar el coeficiente de determinación (R^2) de las variables endógenas, el método alinea el rigor estadístico con la utilidad práctica para la toma de decisiones.

- 3. Flexibilidad en el Continuo Exploratorio-Confirmatorio:** Es crucial superar la dicotomía rígida que encasilla al PLS-SEM únicamente como una herramienta exploratoria. Si bien es ideal para teorías en desarrollo, Hair et al. (2017) clarifican que este enfoque es igualmente válido para diseños confirmatorios cuando el objetivo final de la investigación es la **predicción** de constructos clave. Esta dualidad permite al investigador transitar desde la generación de hipótesis hasta su validación predictiva dentro de un mismo marco metodológico unificado.

Con un diseño muestral fundamentado en la potencia estadística y una clara justificación del enfoque predictivo, el investigador ha establecido los cimientos sólidos para la fase de ejecución. En el siguiente capítulo, abordaremos el Protocolo de Evaluación, detallando los criterios de calidad que transforman estos datos en evidencia científica válida

"La integridad del análisis que sigue depende de este cimiento: solo una muestra con la potencia adecuada faculta a las métricas del Capítulo III para discriminar, sin ambigüedad, entre la causalidad verdadera y el ruido estadístico."

Referencias

- Barclay, D., Higgins, C., y Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies*, 2(2), 285–309.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Faul, F., Erdfelder, E., Buchner, A., y Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., y Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3^a ed.). Sage.
- Hair, J. F., Matthews, L. M., Matthews, R. L., y Sarstedt, M. (2017). PLS-SEM or CB-SEM: Updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, 1(2), 107–123. <https://doi.org/10.1504/IJMDA.2017.087624>
- Hair, J. F., Risher, J. J., Sarstedt, M., y Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Henseler, J. (2023). *Composite-based structural equation modeling: Analyzing latent and emergent variables*. Guilford Press.

- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4ª ed.). Guilford Press.
- Kock, N., y Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227–261. <https://doi.org/10.1111/isj.12131>
- Marcoulides, G. A., y Saunders, C. (2006). PLS: A silver bullet? *MIS Quarterly*, 30(2), III–IX. <https://doi.org/10.2307/25148727>
- Ringle, C. M., Sarstedt, M., Sinkovics, N., y Sinkovics, R. R. (2022). A perspective on using partial least squares structural equation modeling in data articles. *Data in Brief*, 48, 109074.
- Sarstedt, M., Hair, J. F., Pick, M., Lienggaard, B. D., Radomir, L., y Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing*, 39(5), 1035–1064. <https://doi.org/10.1002/mar.21640>
- Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Thiele, K. O., y Ringle, C. M. (2019). Prediction-oriented model evaluation in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, 53(11), 2322–2347. <https://doi.org/10.1108/EJM-02-2019-0189>
- Wold, H. (1985). Partial least squares. En S. Kotz y N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 6, pp. 581–591). Wiley..

CAPITULO III

PROTOCOLO DE EVALUACIÓN Y REPORTE DE RESULTADOS

Tras haber definido en el capítulo anterior la estrategia de muestreo y garantizado la potencia estadística necesaria para la detección de efectos, la investigación entra en su fase analítica decisiva: la validación empírica. Sin embargo, tener datos recolectados no es sinónimo de tener evidencia científica. Los datos son, en esta etapa, materia prima que debe someterse a un riguroso control de calidad antes de poder extraer conclusiones gerenciales o teóricas.

Este capítulo detalla el protocolo sistemático de evaluación en PLS-SEM. A diferencia de las técnicas de primera generación (como la regresión múltiple), donde a menudo se asume que las mediciones son perfectas, el SEM exige un escrutinio dual: primero debemos demostrar que nuestros instrumentos de medición funcionan (fiabilidad y validez) y, solo entonces, podremos evaluar si nuestras hipótesis teóricas son ciertas (predicción y causalidad). A lo largo de las siguientes páginas, desglosaremos este proceso en una hoja de ruta estandarizada que transformará los datos crudos en hallazgos robustos y publicables.

3.1 Fundamentos metodológicos de la validación

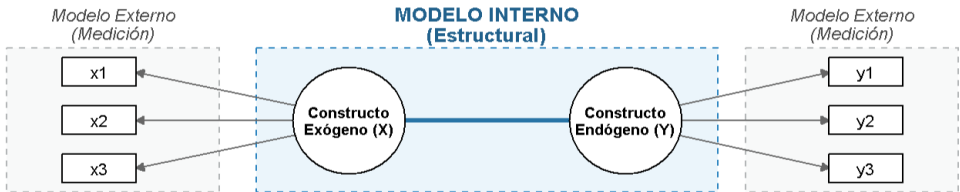
La validación en el Modelado de Ecuaciones Estructurales (SEM) trasciende la mera formalidad de "cumplir con los índices". Se trata de un proceso sistemático de filtrado diseñado para proteger al investigador de cometer errores de inferencia. Su función es garantizar que los constructos abstractos (como "Lealtad" o "Innovación") y sus relaciones causales sean teóricamente coherentes y empíricamente robustos.

Desde una perspectiva epistemológica, este proceso articula los principios de la teoría de la medición con la evaluación predictiva. El objetivo final es asegurar que las conclusiones del estudio no sean artefactos estadísticos, sino insumos confiables para la toma de decisiones estratégicas en entornos de incertidumbre (Hair et al., 2021).

3.1.1 Los dos niveles de teoría en SEM

Para operacionalizar este análisis sin perderse en la complejidad matemática, es imperativo distinguir que un modelo SEM no es una entidad monolítica, sino la fusión de dos niveles de teoría que operan simultáneamente. Como se detalla en la siguiente Tabla, cada nivel cumple una función metodológica distinta y ataca una parte diferente del modelo gráfico.

Figura 5
Arquitectura dual del modelo SEM:
Integración de la teoría de la medición y la teoría estructural)



La Figura 5 desglosa visualmente la complejidad del modelado SEM en tres componentes analíticos diferenciados, facilitando la comprensión de la "Secuencia de Dos Pasos" descrita en la sección anterior:

1. El "Cinturón" de Medición (Zonas Grises Laterales):

Los paneles laterales, etiquetados como Modelo Externo, representan la Teoría de la Medición¹. Aquí se observa la dirección de las flechas finas partiendo desde los constructos (círculos) hacia los indicadores (rectángulos). Esta dirección certifica que estamos ante un modelo reflexivo: se asume teóricamente que la variable latente (X o Y) es la causa que genera los valores en los ítems observados ($x_1 \dots x_3$). La validación de estas zonas grises (Paso 1) es el prerequisite de calidad de datos.

2. El "Núcleo" Estructural (Zona Azul Central):

El panel central, destacado en azul, contiene la Teoría Estructural². Representa la hipótesis del investigador sobre cómo funciona la realidad. Nótese que la flecha que conecta X con Y es visualmente más gruesa y destacada. Esto simboliza que, aunque el modelo requiere de todas las partes para funcionar, el objetivo epistemológico final es evaluar la fuerza y significancia de esta relación causal central (Paso 2).

3. Jerarquía del Análisis:

La disposición gráfica refuerza que no es posible acceder al "núcleo" (evaluar si X influye en Y) sin atravesar primero el "cinturón" de medición. Si los indicadores en las zonas grises no capturan adecuadamente los constructos (bajas cargas o falta de fiabilidad), la flecha central azul carecerá de validez empírica, independientemente de lo que indiquen los cálculos matemáticos.

Tabla 8

Dimensiones teóricas y componentes del modelo SEM

Tipo de Teoría	Componente del Modelo	Función Metodológica (El "Qué" hace)
Teoría de Medición	Modelo Externo (Outer Model)	Especifica cómo los indicadores observados (preguntas de la encuesta) materializan un concepto abstracto. Define las reglas de correspondencia entre la realidad empírica y la teoría latente (reflexiva o formativa).
Teoría Estructural	Modelo Interno (Inner Model)	Especifica la red de relaciones causales entre los constructos latentes. Determina las hipótesis predictivas (las flechas entre círculos) y la arquitectura del fenómeno estudiado.

Nota. Adaptado de Hair et al. (2019) y Gregory (1980).

3.1.2 El Principio de la "Secuencia de Dos Pasos"

Esta distinción entre medición y estructura fundamenta la regla de oro de la validación en SEM: la "Secuencia de Dos Pasos". Este principio, inquebrantable en la literatura contemporánea, establece una jerarquía lógica en el análisis:

"Ningún modelo estructural (relaciones entre variables) puede ser interpretado si previamente no se ha demostrado que sus instrumentos de medida son válidos y confiables" (Hair et al., 2021).

La lógica didáctica detrás de este principio es análoga a la calibración de instrumentos en ingeniería: no tiene sentido medir la velocidad de un vehículo si el velocímetro está roto. En SEM, intentar analizar hipótesis causales (Modelo Interno) con mediciones deficientes (Modelo Externo) carece de sentido epistemológico. Si los datos que alimentan el modelo son "ruido", cualquier relación causal encontrada será espuria, conduciendo a lo que se conoce como falacias estadísticas.

Por tanto, el protocolo de validación que seguiremos en este capítulo respeta estrictamente este orden:

1. **Paso 1:** Validar el instrumento de medición (¿Medimos bien lo que decimos medir?).
2. **Paso 2:** Examinar las relaciones estructurales (¿Influye la variable A en la variable B?).

3.2 Paso 1: Validación del modelo de medida (Outer Model)

El primer filtro analítico del protocolo consiste en evaluar la correspondencia métrica entre el constructo latente y sus indicadores observables. Dado que la inmensa mayoría de las escalas en ciencias del comportamiento son de naturaleza reflexiva (donde el constructo "causa"

las respuestas en los ítems), esta sección detalla los criterios específicos para este tipo de medición.

Para los modelos reflexivos, la validación exige el cumplimiento estricto y secuencial de cuatro criterios estadísticos jerárquicos:

⚠ Nota de Advertencia Metodológica: Alcance del Protocolo

Es crucial distinguir la naturaleza de su modelo antes de aplicar los siguientes criterios. Las métricas que se detallan a continuación (Cargas, Alpha de Cronbach, AVE) son válidas exclusivamente para modelos de medida reflexivos (Modo A).

Si su investigación incorpora constructos formativos (donde los indicadores causan o forman la variable latente, ej. Índice de Nivel Socioeconómico), aplicar estas métricas constituye un error metodológico grave. Los modelos formativos requieren un protocolo de validación totalmente distinto, centrado en la evaluación de la colinealidad (**VIF**) de los indicadores y la significancia estadística de los pesos externos (*outer weights*), en lugar de las cargas factoriales (Hair et al., 2021).

3.2.1 Confiabilidad del Indicador

Este análisis examina la carga factorial externa (*outer loading*) de cada ítem individual para determinar si el indicador realmente pertenece al constructo.

- **Criterio de Magnitud:** El estándar metodológico establece que las cargas deben ser superiores a 0.7081. Este umbral garantiza que el constructo explica más del 50% de la varianza del indicador ($0.708^2 > 0.50$), asegurando que la varianza compartida supera a la varianza del error. Cargas inferiores sugieren que el ítem introduce más "ruido" que información sustantiva.

Nota de Rigor: Significancia Estadística

No basta con observar la magnitud de la carga; es imperativo confirmar que esta no sea producto del azar. Además de superar el umbral de 0.708, cada carga debe ser estadísticamente significativa.

Esto se verifica mediante el procedimiento de Bootstrapping. El investigador debe comprobar que el estadístico t asociado a cada carga sea mayor a 1.96 (para un nivel de confianza del 95%). Si una carga es alta en valor absoluto pero su $t < 1.96$, el indicador no es confiable y debe ser eliminado.

3.2.2 Consistencia Interna

Este criterio evalúa si el conjunto de indicadores que componen un constructo mide la misma magnitud teórica de manera

homogénea. Para su verificación, se utilizan dos coeficientes complementarios:

- **Alpha de Cronbach:** Actúa como el límite inferior conservador de la fiabilidad, asumiendo (a menudo erróneamente) que todos los indicadores tienen cargas iguales.
- **Fiabilidad Compuesta (*Composite Reliability*):** Representa el límite superior y es una medida más precisa en PLS-SEM al considerar las cargas individuales de cada ítem.

En ambos casos, valores situados entre 0.70 y 0.95 se consideran satisfactorios. Valores superiores a 0.95 pueden indicar redundancia de ítems (semánticamente idénticos), lo cual no es deseable.

3.2.3 Validez Convergente

Se determina mediante la Varianza Extraída Media (AVE). Este indicador global del constructo debe ser superior a 0.50. Un valor por encima de este umbral garantiza que, en promedio, el constructo latente es capaz de explicar más de la mitad de la varianza de sus propios indicadores, convergiendo así en una medición sólida y representativa del fenómeno estudiado.

3.2.4 Validez discriminante

Es la prueba crítica para demostrar que cada constructo es empíricamente distinto de los demás y no está capturando fenómenos solapados (validez divergente). Aunque históricamente se utilizó el criterio de Fornell-Larcker, la literatura contemporánea (Henseler et al., 2015; Hair et al., 2021) ha demostrado su baja sensibilidad. Por ello, se recomienda priorizar la ratio **HTMT** (*Heterotrait-Monotrait Ratio*).

- **Criterio de decisión:** Valores de HTMT inferiores a 0.85 (o 0.90 en conceptos teóricamente similares) confirman que existe una distinción real entre las variables latentes. Si el HTMT supera el umbral de 1.0, existe una falta total de validez discriminante.

3.3 Paso 2: Validación del Modelo Estructural (Inner Model)

Una vez certificada la fiabilidad y validez de los instrumentos de medida (Paso 1), el investigador tiene "luz verde" metodológica para proceder a la evaluación del modelo estructural. En esta fase, el foco del análisis cambia radicalmente: el objetivo ya no es verificar la calidad de los datos, sino determinar la **significancia causal** de las relaciones teóricas y la **capacidad predictiva** del modelo propuesto.

La evaluación del modelo interno se estructura en tres dimensiones analíticas fundamentales:

3.3.1 Evaluación de la Colinealidad (VIF)

El primer paso crítico en el modelo estructural es examinar la colinealidad lateral. Este fenómeno ocurre cuando dos o más variables latentes independientes (predictores) están altamente correlacionadas entre sí, lo que dificulta que el algoritmo estime con precisión la influencia única de cada una sobre la variable dependiente. La presencia de alta colinealidad puede distorsionar los coeficientes de ruta, volviéndolos inestables o incluso invirtiendo su signo contraintuitivamente.

Para diagnosticar este problema, se utiliza el **Factor de Inflación de la Varianza (VIF)**. Se debe examinar el VIF de cada conjunto de constructos predictores para cada variable endógena.

- **Criterio de Decisión:**

- **VIF < 3.0 (Ideal):** Indica que no hay problemas de colinealidad. Es el estándar de rigor actual recomendado por Hair et al. (2021).
- **VIF < 5.0 (Tolerable):** Se considera aceptable en estudios exploratorios o con teorías menos maduras, aunque implica cierto grado de solapamiento informativo.
- **VIF \geq 5.0 (Crítico):** Indica problemas severos de colinealidad. El investigador debe tomar medidas correctivas, como eliminar uno de los constructos redundantes o fusionarlos en un constructo de orden superior.

Una vez descartada la colinealidad ($VIF < 3$ o < 5), se puede proceder con confianza a la evaluación de las relaciones causales.

3.3.2 Evaluación de las Relaciones Causales

(Path Coefficients)

Los coeficientes de ruta o *path coefficients* (beta) representan las hipótesis centrales del estudio; indican la fuerza y la dirección de la relación entre las variables latentes. Dado que PLS-SEM es una técnica no paramétrica que no asume normalidad en la distribución de los datos, la significancia estadística de estos coeficientes no puede determinarse mediante las pruebas t tradicionales.

En su lugar, se utiliza el procedimiento de Bootstrapping (remuestreo). Esta técnica re-muestra los datos originales para estimar el error estándar de los coeficientes.

- **Configuración Técnica:** Para garantizar la estabilidad de los resultados, se recomienda configurar el software analítico (ej. SmartPLS, R-semPLS, ADANCO o WarpPLS) con un mínimo de 5,000 submuestras (idealmente 10,000). Una configuración con menos submuestras podría generar valores p imprecisos.

- **Criterio de decisión:** A partir del remuestreo, se analizan los Intervalos de Confianza (*Bias-Corrected Confidence Intervals*). Si el intervalo para un coeficiente beta no incluye el cero (ej. [0.15; 0.45]), se concluye que el efecto es estadísticamente significativo con un 95% de confianza (Preacher y Hayes, 2008)

3.3.3 Coeficiente de Determinación (R²)

Mientras que la significancia nos dice si un efecto existe, estos dos indicadores nos dicen "cuánto importa" ese efecto.

- **Coeficiente de Determinación (R²):** Evalúa el poder explicativo del modelo (*in-sample prediction*). Representa la cantidad de varianza de la variable endógena (dependiente) que es explicada por sus predictores. En investigaciones de marketing y gestión, los valores de referencia suelen interpretarse así (Hair et al., 2021):
 - **0.75:** Poder explicativo **Sustancial**.
 - **0.50:** Poder explicativo **Moderado**.
 - **0.25:** Poder explicativo **Débil**.

1.1.1 Tamaño del Efecto (f²)

Es una medida de sensibilidad que aísla la contribución específica de cada predictor individual. Responde a la pregunta gerencial: "Si elimino esta variable del modelo, ¿cuánto pierdo en

capacidad de predicción?". Permite identificar qué constructos son los impulsores clave del fenómeno y cuáles tienen un impacto meramente marginal.

3.3.4 Relevancia Predictiva (Q²)

Finalmente, se evalúa la capacidad del modelo para predecir datos que no fueron utilizados en la estimación inicial (*out-of-sample prediction*). Históricamente, este análisis se realizaba mediante el procedimiento de Blindfolding, que calcula el valor Q² de Stone-Geisser omitiendo datos sistemáticamente.

- **Criterio tradicional:** Un valor de $Q^2 > 0$ indica que el modelo tiene relevancia predictiva.

Sin embargo, la literatura metodológica más reciente (Shmueli et al., 2019) señala que el *Blindfolding* no es una medida de predicción pura, ya que reutiliza los datos de la muestra. Por ello, se recomienda encarecidamente utilizar el algoritmo **PLSpredict**.

- **La ventaja de PLSpredict:** Esta técnica divide la muestra en subconjuntos de entrenamiento y prueba (*k-fold cross-validation*), permitiendo evaluar si el modelo SEM predice mejor el comportamiento de los nuevos casos que un modelo lineal ingenuo (LM).

- **Criterio de decisión actualizado:** Si los errores de predicción del modelo PLS (RMSE o MAE) son menores que los del modelo lineal (LM), el modelo posee un poder predictivo alto y robusto.

1.2 Consideraciones sobre el ajuste global y la predicción.

Una de las áreas de mayor controversia y evolución en la literatura metodológica reciente reside en el concepto de "Ajuste del Modelo" (*Model Fit*). Históricamente, los investigadores formados en la tradición del CB-SEM (basado en covarianzas, como AMOS o LISREL) han buscado validar sus modelos mediante índices globales como el CFI, RMSEA o χ^2 . Sin embargo, trasladar esta lógica de forma automática al entorno de PLS-SEM constituye un error conceptual que puede derivar en el rechazo injustificado de modelos predictivos válidos.

A diferencia del CB-SEM, cuyo algoritmo busca minimizar la discrepancia entre la matriz de covarianza teórica y la observada, el PLS-SEM es un enfoque causal-predictivo que maximiza la varianza explicada de los constructos endógenos. Esta diferencia fundamental ha llevado a algunos críticos a sostener erróneamente que PLS-SEM es menos riguroso para la confirmación de teorías (Westland, 2015). No obstante, la realidad metodológica es que PLS-SEM prioriza la predicción sobre el ajuste global (Shmueli et al., 2019).

Siguiendo las directrices actualizadas de Hair et al. (2021) y Rigdon (2012), evaluar un modelo PLS exclusivamente bajo la óptica de las métricas de ajuste tradicionales es inadecuado por tres razones estructurales:

3.3.5 Fundamento algorítmico (Incongruencia Matemática)

El algoritmo PLS se basa en regresiones de mínimos cuadrados ordinarios (OLS) diseñadas para la predicción, no para la optimización del ajuste global. Las métricas basadas en χ^2 asumen que el objetivo es reproducir perfectamente la matriz de covarianza de la muestra. Dado que PLS no tiene función de bondad de ajuste global en su función objetivo, aplicar estas métricas equivale a juzgar una herramienta (el martillo) por su capacidad para realizar una tarea para la que no fue diseñada (atornillar).

3.3.6 Estado experimental de las métricas alternativas

Aunque metodólogos como Henseler et al. (2016) han propuesto métricas de ajuste específicas para PLS, como el SRMR (*Standardized Root Mean Square Residual*) o las pruebas de ajuste exacto (d_{ULS} , d_G), su uso debe ser extremadamente cauteloso.

- **Limitación actual:** Los umbrales de corte (por ejemplo, $SRMR < 0.08$) son tentativos y no han sido validados robustamente en todas las condiciones de heterogeneidad de datos. Por ello, Hair et al. (2021) advierten que el rechazo de

un modelo PLS basado únicamente en un "mal ajuste" según estas métricas provisionales es una práctica desaconsejada.

3.3.7 Objetivo Epistemológico (Explicación vs. Predicción)

Existe una distinción filosófica crucial: un modelo puede tener un "ajuste" excelente (reproducir bien los datos pasados) y, sin embargo, tener una capacidad nula para predecir eventos futuros. En el ámbito de la gestión estratégica y el marketing, el interés suele centrarse en la capacidad del modelo para anticipar comportamientos (Hair et al., 2017).

- **El cambio de paradigma:** El rigor en PLS-SEM no se demuestra mediante la replicación de covarianzas (ajuste), sino mediante la **precisión predictiva**.

Conclusión del debate:

La calidad científica en PLS-SEM debe juzgarse desplazando el foco del "Fit" a la "Predicción". El investigador debe priorizar la evaluación de la potencia predictiva fuera de muestra (out-of-sample prediction), utilizando herramientas como PLSpredict, y la robustez de los coeficientes estructurales, en lugar de obsesionarse con índices de ajuste global que poseen poco valor diagnóstico en este contexto (Shmueli et al., 2019).

3.4 Síntesis y estándares de reporte académico

La validación estadística carece de impacto si no se comunica con transparencia y rigor. Un error frecuente en investigadores noveles es presentar capturas de pantalla indiscriminadas del software o saturar el texto con índices irrelevantes, omitiendo métricas esenciales para la replicabilidad científica.

Para subsanar esta brecha, la **Tabla 9** se presenta con una doble función:

1. **Resumen Ejecutivo:** Compendia los umbrales de decisión aceptados en la literatura contemporánea para aceptar o rechazar hipótesis.
2. **Plantilla de Reporte:** Estructura la jerarquía de información que debe aparecer obligatoriamente en la sección de "Resultados" de una tesis o artículo científico.

Se recomienda al investigador utilizar esta tabla como una "lista de verificación" (*checklist*) antes de enviar su manuscrito. Si su reporte de resultados cubre explícitamente cada una de las fases aquí descritas (Medición, Estructura y Ajuste), estará cumpliendo con los estándares de publicación exigidos por revistas de cuartil superior (Q1/Q2).

Tabla 9

Resumen del protocolo de validación en PLS-SEM (Modelos Reflexivos)

Fase de Validación	Criterio / Métrica	Umbral de Aceptación (Regla General)	Referencia Clave
1. MODELO DE MEDIDA	Cargas Factoriales (Outer Loadings)	$\lambda \geq 0.708$ (Aceptable >0.40 si su eliminación no mejora la fiabilidad, pero <0.70 requiere justificación)	Hair et al. (2021)
	Consistencia Interna (Reliability)	Alpha de Cronbach: 0.70 - 0.95 Fiabilidad Compuesta (ρ_c): 0.70 - 0.95	Nunnally & Bernstein (1994)
	Validez Convergente	$AVE \geq 0.50$ (El constructo explica $>50\%$ de la varianza de sus ítems)	Bagozzi & Yi (1988)
	Validez Discriminante	HTMT ratio < 0.85 (Estricto) o < 0.90 (Conceptos similares) IC del HTMT no debe incluir el 1.	Henseler et al. (2015)
2. MODELO ESTRUCTURAL	Colinealidad	$VIF < 3.0$ (Ideal) o $VIF < 5.0$ (Tolerable)	Hair et al. (2021)
	Significancia de Rutas (Path Coefficients)	$t\text{-value} > 1.96$ ($p < 0.05$) IC del Bootstrap (BCa) no debe incluir el 0.	Preacher & Hayes (2008)
	Coefficiente de Determinación	$R^2 \approx 0.75$ (Sustancial), 0.50 (Moderado), 0.25 (Débil)	Hair et al. (2021)
	Tamaño del Efecto	$f^2 \geq 0.02$ (Pequeño), $f^2 \geq 0.15$ (Medio), $f^2 \geq 0.35$ (Grande)	Cohen (1988)

Fase de Validación	Criterio / Métrica	Umbral de Aceptación (Regla General)	Referencia Clave
	Relevancia Predictiva	$Q^2 > 0$ (Tiene relevancia predictiva) Evaluación preferente mediante PLSpredict.	Shmueli et al. (2019)
3. AJUSTE GLOBAL	Ajuste del Modelo (<i>Model Fit</i>)	$SRMR < 0.08$ (<i>Uso diagnóstico; no usar como criterio único de rechazo.</i>)	Henseler et al. (2016)

Nota. Los umbrales presentados son referenciales para investigaciones en ciencias sociales y empresariales; pueden variar según la madurez de la teoría.

3.5 Guía de lecturas avanzadas y modelos de referencia

El protocolo de validación expuesto en las secciones precedentes constituye el estándar metodológico actual. Sin embargo, la investigación científica es una disciplina viva que se adapta a las particularidades de cada objeto de estudio. Para facilitar la transición de la teoría a la práctica investigativa, ofrecemos aquí un compendio de lecturas complementarias y modelos de referencia.

Esta sección se distingue de la bibliografía general por su carácter instrumental. Los siguientes trabajos han sido seleccionados por su rigor en la ejecución del análisis post-hoc y la claridad en la presentación de la evidencia empírica. Se recomienda al investigador novelesco utilizar estos textos como "plantillas de calidad" (*benchmark*) para evaluar la estructura y profundidad de sus propios reportes de resultados en tesis o artículos científicos

1. Soriano, J.L. & Mejía-Trejo, J. (2022). Modelado de Ecuaciones Estructurales en el campo de las Ciencias de la Administración. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 33, 242-263.
<https://www.econstor.eu/bitstream/10419/286270/1/1814673792.pdf>
 - Analiza la pertinencia y aplicación práctica del SEM en la gestión y administración, y revisa comparativamente los software disponibles en el área empresarial.

2. Ikram, A.A.D.W., et al. (2025). Empleo de modelos de ecuaciones estructurales para analizar factores de políticas públicas en gobiernos locales. *Sustainability*, 17(15), 6855.
<https://www.mdpi.com/2071-1050/17/15/6855>
 - Presenta aplicaciones de SEM para medir el impacto de variables latentes en la gestión pública y políticas públicas utilizando LISREL.

3. Maynez Guaderrama, A.I. & López Torres, V.G. (2025). Modelos de ecuaciones estructurales como técnica de análisis en ciencias administrativas. Astra Editorial. <https://astraeditorialshop.com/wp-content/uploads/2025/04/Modelos-de-ecuaciones-estructurales-como-tecnica.pdf>
 - Libro colectivo con ejemplos detallados de SEM tanto en gestión pública como en empresas, incluyendo aplicaciones con software gratuito (R, PLS-SEM, multigrupo).

4. Natesan, S.D. (2023). Un enfoque de ecuaciones estructurales para la implementación de políticas públicas. *Journal of Policy Modeling*, 45(2), 401-417.
<https://www.sciencedirect.com/science/article/pii/S0970389623000307>
 - Utiliza SEM para evaluar cómo variables de redes sociales y demografía influyen en la implementación de políticas públicas.

5. Vidal-Portilla, E., et al. (2023). Modelamiento mediante ecuaciones estructurales (PLS-SEM) en la investigación en

ciencias empresariales. Contabilidad y Negocios, 18(36).

<https://revistas.pucp.edu.pe/index.php/contabilidadyNegocios/article/view/26930>

· Ejemplo de uso de PLS-SEM aplicado a factores de competitividad en empresas públicas y privadas.

6. Ruiz, J. & Guerrero, J. (2025). Tutorial de modelado de ecuaciones estructurales con software libre para gestión pública y empresarial. Astra Editorial. https://astraeditorialshop.com/wp-content/uploads/2025/04/Modelos-de-ecuaciones-estructurales-como-tecnica_.pdf
· Incluido dentro del libro anterior, muestra paso a paso cómo usar R y paquetes libres para SEM en análisis de gestión pública.
7. “Ecuaciones Estructurales Aplicadas a la Toma de Decisiones”. Revista Internacional de Ciencias Sociales y Humanidades, 13(1), 2024.
<https://www.reincisol.com/ojs/index.php/reincisol/article/view/1007>
· Analiza la utilidad del modelado SEM en la toma de decisiones estratégicas en el sector público y en empresas.

Referencias

- Bagozzi, R. P., y Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94. <https://doi.org/10.1007/BF02723327>
- Bagozzi, R. P., y Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, 40(1), 8–34. <https://doi.org/10.1007/s11747-011-0278-x>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2^a ed.). Lawrence Erlbaum Associates.
- Gregory, R. J. (1980). *Psychological testing: History, principles, and applications*. Allyn & Bacon.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., y Sarstedt, M. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science*, 45(5), 616–632. <https://doi.org/10.1007/s11747-017-0517-x>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., y Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3^a ed.). Sage Publications.
- Hair, J. F., Risher, J. J., Sarstedt, M., y Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>

- Henseler, J., Hubona, G., y Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, 116(1), 2–20. <https://doi.org/10.1108/IMDS-09-2015-0382>
- Henseler, J., Ringle, C. M., y Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Henseler, J., y Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, 28(2), 565–580. <https://doi.org/10.1007/s00180-012-0317-1>
- Nunnally, J. C., y Bernstein, I. H. (1994). *Psychometric theory* (3ª ed.). McGraw-Hill.
- Preacher, K. J., y Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/BRM.40.3.879>
- Rigdon, E. E. (2012). Rethinking partial least squares modeling: In praise of simple methods. *Long Range Planning*, 45(5–6), 341–358. <https://doi.org/10.1016/j.lrp.2012.09.010>
- Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., y Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, 53(11), 2322–2347. <https://doi.org/10.1108/EJM-02-2019-0189>

Westland, J. C. (2015). *Structural Equation Models: From Paths to Networks*. Springer. <https://doi.org/10.1007/978-3-319-16507-3>

CAPITULO IV

MODELADO AVANZADO Y EVALUACIÓN DE MODELOS

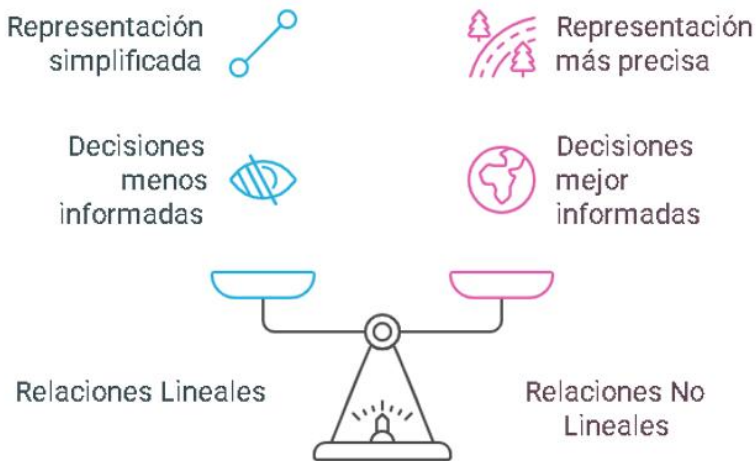
Imagina que quieres entender cómo se relacionan diferentes factores en el mundo de los negocios, como la satisfacción de un cliente con su lealtad. tradicionalmente, los investigadores suponen que estas relaciones son lineales, es decir, como una línea recta: si uno aumenta, el otro también lo hace de forma constante. Pero, ¿y si la realidad es más compleja? A menudo, estas conexiones forman curvas, donde el efecto cambia según el nivel en que se encuentre cada factor.

Pensemos en un ejemplo concreto. La relación entre la satisfacción y la lealtad no siempre es una línea recta. Al principio, un aumento en la satisfacción puede hacer que la lealtad crezca mucho; pero después de cierto punto, ese mismo aumento tiene un impacto cada vez menor. Esta curva, similar a una parábola, es lo que se conoce como un efecto no lineal. Para capturar estas sutilezas, los investigadores utilizan herramientas estadísticas especiales que permiten añadir estos "giros" a sus análisis. De este modo, no se limitan a ver líneas rectas, sino que pueden mapear las curvas reales que describen cómo se comportan las cosas.

Entender que las relaciones en los negocios a menudo siguen curvas en lugar de líneas rectas nos da una imagen mucho más rica y precisa de la

realidad. Este enfoque permite a las empresas tomar decisiones mejor informadas, ajustar sus estrategias con mayor precisión y, en definitiva, comprender mejor los complejos mecanismos que rigen el comportamiento de sus clientes y del mercado.

Relaciones lineales vs. Relaciones no lineales



4.1 Modelado de relaciones no lineales: Efectos cuadráticos

La asunción de relaciones lineales simples entre variables es generalizada en los primeros modelos de gestión, economía y ciencias del comportamiento. Sin embargo, en contextos organizacionales complejos, tal linealidad suele ser insostenible y puede llevar a conclusiones erróneas. Conceptos como la Ley de Rendimientos Decrecientes propuesta por la economía clásica ilustran claramente que incrementos sucesivos en una variable (por ejemplo, inversión en recursos o trabajo) tienden a generar

beneficios marginales cada vez menores más allá de un umbral (Samuelson & Nordhaus, 2010).

En términos prácticos, muchos efectos estratégicos, de innovación, o de gestión presentan formas curvilíneas. Por ejemplo, una inversión creciente en Investigación y Desarrollo (I+D) puede, hasta cierto punto, mejorar la rentabilidad de la empresa, pero tras un nivel óptimo, un exceso de inversión puede generar retornos decrecientes o incluso negativos (Benson y Ziedonis, 2009). Este fenómeno es reflejo de relaciones cuadráticas, donde la variable predictora afecta al resultado con un patrón en forma de U o de \cap (curva convexa o cóncava).

Metodología: centrado y creación del término cuadrático

El modelado de relaciones cuadráticas requiere transformar la variable de interés X para evitar problemas de colinealidad y facilitar la interpretación. Se recomienda primero centrar o estandarizar X , restando su media o dividiendo por su desviación estándar (Aiken & West, 1991; Hair et al., 2019):

$$X_{\text{centrado}} = X - \bar{X}$$

$$X_{\text{estandarizado}} = (X - \bar{X}) / SD_X$$

Luego, se crea el término cuadrático:

$$X^2_{\text{centrado}} = (X_{\text{centrado}})^2$$

Este término se incorpora al modelo de regresión estructural, permitiendo identificar efectos no lineales.

Ejemplo: Inversión en I+D y Rentabilidad

Suponga que se analiza el impacto de la inversión en I+D sobre la rentabilidad (Y) de una empresa. El modelo incluiría tanto I+D centrada como $(I+D)^2$ centrada:

$$Y = \beta_0 + \beta_1(I + D) + \beta_2(I + D)^2_{\text{centrado}} + \varepsilon$$

Y: Variable Dependiente o variable de respuesta.

β_0 : Es el Intercepto (el valor de Y cuando todos los demás términos son cero).

β_1 y β_2 : Son los Coeficientes de Regresión (pesos que indican el impacto de los términos lineales y cuadráticos en Y).

$(I + D)$: Es el término Lineal que representa la combinación de una o dos variables independientes (p. ej., I + D).

$(I + D)^2_{\text{centrado}}$: Es el término Cuadrático (elevado al cuadrado) de la variable combinada (I + D). El subíndice "centrado" indica que esta variable se ha restado de su media antes de elevarla al cuadrado, una práctica común para reducir la multicolinealidad en modelos polinómicos.

ε : Es el Término de Error (el residuo no explicado por el modelo).

Si $\beta_2 < 0$, existe una relación cuadrática invertida: más allá de cierto punto, mayores inversiones en I+D reducen la rentabilidad, mostrando

rendimientos decrecientes (Benson & Ziedonis, 2009; Hair et al., 2019). Esta metodología refina la comprensión gerencial, permite identificar puntos óptimos y evita recomendaciones excesivamente simplistas o riesgosas.

Guía SmartPLS: Implementación y prueba de significancia de efectos no lineales (Término Cuadrático)

PASO 1: Prepara tus datos e ingresa tu modelo básico

- Carga el archivo de datos en SmartPLS.
- Asegúrate de tener incluidas las variables latentes y sus respectivos indicadores en el modelo gráfico.

PASO 2: Centrado o estandarización de la variable predictora

- Antes de agregar el término cuadrático, es recomendable centrar (restar la media) o estandarizar (transformar a z-score) la variable exógena para minimizar la colinealidad y facilitar la interpretación.
- Puedes crear el nuevo indicador (X_{centrado}) fuera de SmartPLS y cargar la base de datos, o emplear herramientas de transformación en el propio software.

PASO 3: Crear el término cuadrático dentro del modelo gráfico

- Haz clic derecho sobre la variable latente predictora en el modelo gráfico.
- Selecciona “Crear término de interacción”, elige “Efecto cuadrático”.
- Especifica la variable para la que deseas el cuadrado (X). SmartPLS generará automáticamente X^2 como una variable latente de segundo orden.
- Una flecha se agregará entre este nuevo término cuadrático y la variable dependiente.

PASO 4: Configurar y ejecutar el algoritmo PLS

- Haz clic en el botón de algoritmos y selecciona “PLS” para cálculos estándar.
- Selecciona la opción bootstrap para obtener intervalos de confianza y el p-value de los parámetros.
- Ejecuta el modelo. Espera a que finalice el cálculo.

PASO 5: Prueba de significancia y análisis del p-value

- En la pestaña de resultados, localiza el coeficiente estimado para X^2 (término cuadrático).

- Observa el p-value asociado.
 - Si el p-value es menor a 0.05, el efecto cuadrático es estadísticamente significativo.
 - Un valor negativo indica una relación en forma de \cap ; positivo, una forma de U .

PASO 6: Cálculo e interpretación del punto de inflexión (Turning Point)

- Anota los coeficientes β_1 (de X) y β_2 (de X^2) que SmartPLS reporta para la variable dependiente.
- Utiliza la fórmula del punto de inflexión:

$$X^* = -\beta_1 / (2 * \beta_2)$$

Esta fórmula se utiliza típicamente en **Regresión Polinómica Cuadrática** (modelos que incluyen un término al cuadrado, como la fórmula que convertimos anteriormente) para encontrar el **valor óptimo** o el **vértice** de la parábola.

- X^* : Representa el valor de la variable independiente X (o el predictor combinado) en el que la variable dependiente Y alcanza su valor **máximo o mínimo** (el vértice de la curva).
- β_1 : Es el negativo del coeficiente del término **lineal** en el modelo de regresión.

- β_2 : Es dos veces el coeficiente del término **cuadrático** en el modelo de regresión.
- El resultado X^* representa el nivel óptimo de X donde el efecto sobre la variable dependiente se revierte o maximiza.
- Interpreta gerencialmente:
 - Si X^* cae dentro del rango de los datos, identifica el nivel óptimo real de la variable predictora.
 - Si cae fuera, advierte cautela o posibles límites en la interpretación práctica.

RESUMEN

1. Confirma la significancia estadística del efecto cuadrático (p-value).
2. Calcula el punto óptimo a partir de los coeficientes:

$$X^* = -\beta_1 / (2 * \beta_2)$$

3. Compara el X^* con tu contexto y comunica recomendaciones. Por ejemplo: *“La rentabilidad máxima por inversión en I+D se alcanza con un nivel de inversión equivalente a X^* ; invertir más podría generar retornos decrecientes.”*

4.2 Fundamentos del análisis confirmatorio de tétradas (CTA-PLS)

El Análisis Confirmatorio de Tétradas (CTA-PLS) es una metodología psicométrica avanzada implementada en el contexto de modelos de ecuaciones estructurales, especialmente con estimación PLS. El propósito fundamental del CTA es la evaluación rigurosa de la **validez discriminante**, es decir, garantizar que cada constructo teórico sea distinto y represente una dimensión específica, en lugar de emerger como un simple artefacto estadístico o una sola medida común (Dijkstra & Henseler, 2011).

¿Qué es una tétrada?

Una tétrada es una combinación matemática que toma cuatro ítems observados (indicators) de un constructo y compara sus covarianzas de forma específica. El análisis confirmatorio de tétradas (CTA) aprovecha esta comparación para detectar si la estructura interna del constructo refleja las diferencias entre los ítems esto es, si la varianza entre combinaciones de pares de indicadores es suficientemente pequeña como para pensar que el constructo representa una entidad real y no una suma artificial.

Propósito y valor agregado del CTA

El análisis confirmatorio de tétradas (CTA) se utiliza para analizar si los constructos son verdaderamente multidimensionales y si la estructura de correlación interna entre los indicadores respalda su validez discriminante.

- Si las varianzas observadas en las tétradas son **significativamente más pequeñas** que lo que se observaría en caso de que todos los indicadores compartieran una única fuente común (por ejemplo, sesgo de un método o una sola medida), ello respalda la validez discriminante.
- Si, por el contrario, las diferencias son grandes, el constructo puede estar alterado por ruido, sesgo o artefactos que impiden considerar la validez discriminante.

CTA y HTMT: Complementariedad

Mientras el HTMT (Heterotrait-Monotrait Ratio) es una comparación entre la correlación de indicadores de diferentes constructos (heterotrait) y de igual constructo (monotrait) para evaluar validez discriminante (Henseler et al., 2015), el CTA va más allá al analizar la *estructura interna* mediante las tétradas. El HTMT identifica cuándo los constructos se solapan demasiado; el CTA detecta si la estructura interna es consistente con un verdadero constructo latente.

Metodología del CTA

El CTA se basa en la estimación y comparación de las varianzas de todas las posibles tétradas generadas a partir de los indicadores de un constructo. Si esas varianzas son consistentemente más bajas, hay evidencia de que las diferencias entre ítems son más pequeñas de lo que serían por

artefacto, es decir, el constructo es teóricamente sólido (Dijkstra & Henseler, 2011; Dijkstra, 2014).

Metodología de Evaluación con el CTA-PLS

El Análisis Confirmatorio de Tétradas (CTA-PLS) es un procedimiento estadístico especializado para evaluar la validez discriminante de los constructos en modelos de ecuaciones estructurales estimados por Mínimos Cuadrados Parciales (PLS-SEM). El CTA examina si la diferencia entre indicadores de un constructo latente es, en realidad, menor que la esperada si solo hubiera una medida común (artefacto), permitiendo identificar modelos de medición auténticos y sólidos (Dijkstra, 2014; Dijkstra & Henseler, 2011).

Ejecución del CTA

1. Preparación del modelo:

El investigador debe especificar el modelo reflejo en PLS-SEM con todos los indicadores de los constructos latentes correctamente definidos.

2. Cálculo matemático:

En el CTA, se generan todas las tétradas (combinaciones posibles de cuatro indicadores de un constructo). Se calcula la varianza de

cada tétlada, confrontándola con la hipótesis nula de una única fuente común.

3. Obtención del p-value:

El software reporta para cada constructo el p-value del test de tétladas, resultado de una prueba de hipótesis que evalúa si las diferencias internas son estadísticamente menores a las esperadas por artefacto.

Interpretación crítica del p-value y regla de decisión

- **Umbral estadístico:**

El criterio universal adoptado en la literatura (Dijkstra & Henseler, 2011; Hair et al., 2019) es el nivel de significancia alfa = 0.05.

- **Decisión:**

- Si el **p-value** ≥ 0.05 , no se rechaza la hipótesis nula, lo que significa que la estructura interna del constructo es distinta a una medida común y se confirma una adecuada **validez discriminante**. El modelo de medición es aceptable y puede ser utilizado para interpretación teórica y práctica.
- Si el **p-value** < 0.05 , se rechaza la hipótesis nula, indicando que la diferencia entre los ítems del constructo no es suficientemente pequeña como para ser considerada válida.

Esto sugiere problemas de validez discriminante y el investigador debe **re-evaluar o re-especificar** el modelo de medición. Posibles acciones incluyen eliminar o modificar indicadores, dividir el constructo, revisar la especificación teórica o aplicar el CTA a versiones ajustadas.

El resultado del CTA es una señal clara para tomar decisiones. Un p-value menor que 0.05 obliga a revisar profundamente la validez conceptual y estadística del modelo; un p-value igual o mayor a 0.05 otorga confianza en la diferenciación de los constructos latentes, fortaleciendo la interpretación del modelo. El CTA, junto al HTMT, proporciona un estándar avanzado y transparente para validar la estructura interna de medición en investigaciones cuantitativas.

4.3 PLS-Predict y Evaluación Predictiva Out-of-Sample

En la frontera de la investigación predictiva, la verdadera utilidad de un modelo estructural va más allá del ajuste interno (in-sample fit). El alto impacto y valor práctico de los modelos en gestión, marketing, economía o salud radica en su **capacidad de predecir comportamientos, actitudes o resultados futuros** en datos no utilizados durante la estimación inicial. Por ello, la **evaluación predictiva out-of-sample** es un criterio esencial para juzgar la robustez y utilidad real de los modelos PLS-SEM (Hair et al., 2019; Shmueli et al., 2016).

PLS-Predict es la herramienta estandarizada para evaluar la capacidad predictiva fuera de muestra de los modelos PLS-SEM. Su lógica se fundamenta en la **validación cruzada**:

- El conjunto de datos se divide en múltiples partes (folds).
- El modelo se estima sobre el conjunto de entrenamiento (in-sample),
- Se predicen los valores de los indicadores dependientes sobre el conjunto de prueba (out-of-sample).
- Este proceso se repite varias veces, y se recopilan los errores de predicción obtenidos.

El resultado: una estimación realista del poder predictivo del modelo en nuevos casos.

Criterio Clave: Q_{predict}^2 y Comparación de Errores

- **Q_{predict}^2** : Es una métrica que indica el poder de predicción del modelo. Se calcula como:

$$Q_{\text{predict}}^2 = 1 - (\text{error de predicción} / \text{varianza de los datos})$$

Un valor positivo (> 0) indica que el modelo tiene capacidad predictiva; valores cercanos a 1 implican alta precisión en la predicción out-of-sample (Shmueli et al., 2016).

- **Errores de predicción (RMSE/MAE):** Las métricas como el Error Cuadrático Medio de Predicción (RMSE) y el Error Absoluto Medio (MAE) permiten comparar el desempeño predictivo del modelo PLS-SEM con modelos tradicionales de regresión lineal.
 - Si el RMSE o MAE de PLS-SEM es menor que el de la regresión lineal, el modelo PLS demuestra superioridad predictiva.
 - Si los errores del PLS-SEM son mayores, puede ser necesario revisar las especificaciones del modelo o considerar alternativas.

Implicancia Gerencial

La evaluación con PLS-Predict orienta al investigador hacia modelos que no solo explican, sino que anticipan comportamientos y resultados en contextos reales. Únicamente mediante la validación fuera de muestra ($Q_{\text{predict}}^2_{\text{predic}}$, RMSE, MAE) se pueden ofrecer recomendaciones sólidas y resilientes para la toma de decisiones, la innovación y la gestión basada en datos futuros y escenarios desconocidos.

4.4 Estrategias de validación cruzada avanzada y comparación de modelos

La robustez y replicabilidad de un modelo estadístico dependen no solo del ajuste interno, sino de su estabilidad y capacidad de generalización

ante nuevos datos. Para lograr esto, se aplican estrategias avanzadas de **validación cruzada** y técnicas rigurosas de **comparación de modelos**, asegurando que la elección final tenga el mayor poder explicativo y predictivo posible.

Validación Cruzada: Proceso y valoración

La **validación cruzada (Cross-Validation)** consiste en dividir el conjunto de datos en varias partes o "folds". El modelo se ajusta en el conjunto de entrenamiento y se evalúa sobre el conjunto de prueba. Este ciclo se repite múltiples veces, rotando los datos entre entrenamiento y prueba hasta que todos los casos han sido validados fuera de muestra. Las variantes comunes incluyen la validación *k-fold*, *leave-one-out* y *repeated cross-validation* (James et al., 2013).

Este procedimiento permite:

- Medir la **estabilidad** del modelo, reduciendo el riesgo de sobreajuste.
- Obtener estimaciones más realistas de la capacidad predictiva mediante métricas como RMSE, MAE y Q^2 .
- Identificar debilidades estructurales que no son evidentes con una sola partición de datos.

En investigación aplicada es habitual contrastar diferentes modelos teóricos, por ejemplo, un modelo jerárquico versus uno anidado, o modelos con diferentes niveles de complejidad.

Las principales estrategias para comparar incluyen:

1. **Bondad de Ajuste Global (GOF):**

- Aplicar índices como el **SRMR** (Standardized Root Mean Square Residual), **NFI** (Normed Fit Index) y otros criterios globales para evaluar el ajuste total del modelo.
- Modelos con menor SRMR o mayor NFI se consideran mejor ajustados (Henseler et al., 2014).

2. **Evaluación Predictiva Out-of-Sample:**

- Utilizar validación cruzada para comparar el poder predictivo. El modelo con menor error (RMSE/MAE) y mayor Q^2 en los datos de prueba será el preferido.

3. **Pruebas Específicas de Comparación:**

- Pruebas estadísticas como el **diferencial de log-verosimilitud**, criterios de información (AIC, BIC) o el test de diferencia de baja dimensionalidad, según el tipo de modelo.

4. Modelos Jerárquicos vs. Anidados:

- Para comparar un modelo jerárquico (con constructos de segundo orden) frente a uno anidado (solo primer orden), se debe observar si el modelo complejo añade capacidad predictiva significativa, utilizando validación cruzada y criterios GOF.
- Si la mejora es marginal, se prefiere el modelo más parsimonioso.

Regla de Decisión Gerencial

El modelo óptimo será aquel que:

- Demuestre **estabilidad** en los resultados según validación cruzada.
- Posea un **ajuste global superior** (menor error, mejores índices).
- Sea más **parsimonioso** cuando las diferencias explicativas entre modelos sean pequeñas.
- Ofrezca **mayor potencia predictiva** en escenarios reales.

4.5 Valor gerencial de la no linealidad, CTA y PLS-Predict

En el entorno empresarial contemporáneo, la alta dirección necesita modelos analíticos que no solo expliquen lo que ocurre sino anticipen con

precisión sus consecuencias. La implementación de técnicas avanzadas como la **No Linealidad**, el **Análisis Confirmatorio de Tétradas (CTA)** y **PLS-Predict** marca la diferencia entre análisis convencionales y una arquitectura metodológica verdaderamente robusta y gerencialmente útil.

La **No Linealidad** permite descubrir que el impacto de las decisiones no siempre es proporcional por ejemplo, reconocer límites óptimos en inversión o gestión que evitan rendimientos decrecientes y riesgos innecesarios. Este entendimiento ayuda a la alta dirección a afinar la asignación de recursos, calibrar estrategias y anticipar efectos inesperados según los diferentes niveles de acción.

El **CTA** aporta rigor en la evaluación de la validez discriminante de los constructos del modelo, garantizando que cada variable analizada responde a dimensiones teóricas genuinas y no a artefactos estadísticos. Ello fortalece la confianza ejecutiva en que las relaciones y conclusiones derivadas del análisis reflejan realidades organizacionales y no distorsiones, permitiendo tomar decisiones con certidumbre conceptual.

Por su parte, **PLS-Predict** lleva el análisis más allá del ajuste interno, evaluando el poder predictivo real sobre datos no vistos. La alta dirección obtiene así no solo explicaciones de lo sucedido, sino estimaciones precisas de lo que podría ocurrir, posibilitando la planeación y el control estratégico en escenarios de alta incertidumbre.

En suma, un modelo validado con estas tres técnicas no solo es estadísticamente superior, sino también **más preciso, flexible y orientado a la acción ejecutiva**. Permite decidir qué hacer, hasta dónde hacerlo y con qué probabilidad de éxito, dotando a la alta dirección de ventajas competitivas y capacidad de respuesta fundamentada ante la complejidad empresarial contemporánea.

4.6 Investigaciones sobre: Modelado avanzado y evaluación de modelos, en bases de datos:

1. **Richter, N.F., et al. (2024).**

“Combination of PLS-SEM and machine learning algorithms for identifying nonlinear effects.”

<https://www.sciencedirect.com/science/article/pii/S0148296323008123>

Analiza cómo SmartPLS puede combinarse con algoritmos de machine learning para detectar efectos no lineales en variables latentes, aplicando la técnica en contextos de mercadeo y gestión empresarial avanzada.

2. **Gudergan, S.P., et al. (2025).**

“Advanced partial least squares SEM in business research.”

<https://www.sciencedirect.com/science/article/pii/S0148296324005915>

Estudio sobre aplicaciones avanzadas de PLS-SEM, incluyendo no linealidad, en el análisis de desempeño organizacional y toma de decisiones en empresas.

3. **Thien, L.M. (2024).**

“Relaciones lineales y no lineales entre liderazgo y aprendizaje profesional con PLS-SEM.”

<https://www.nature.com/articles/s41599-023-02500-5>

Investiga, usando SmartPLS, cómo el liderazgo organiza efectos

no lineales en el desarrollo profesional, relevante para la gestión de recursos humanos en administración pública y privada.

4. **SmartPLS Team (2019).**

“Procedimiento para estimar relaciones no lineales en SmartPLS.”

<https://www.smartpls.com/documentation/algorithms-and-techniques/nonlinear/>

Documentación oficial que explica cómo implementar análisis de efectos cuadráticos y no lineales en modelos empresariales y proyectos de gestión pública.

5. **Martínez Ávila, M. (2018).**

“Aplicación de la técnica PLS-SEM en la gestión del conocimiento empresarial.”

http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-74672018000100130

Estudio sobre cómo SmartPLS permite modelar relaciones no lineales entre factores de gestión del conocimiento y desempeño organizacional en empresas.

6. **Magno, F. (2024).**

“A brief review of PLS-SEM in quality management, including nonlinear effects.”

<https://www.emerald.com/tqm/article/36/5/1242/1217482/A-brief-review-of-partial-least-squares-structural>

Revisión sobre el uso de SmartPLS para identificar relaciones no lineales en calidad total y procesos de mejora continua en sectores empresariales.

7. **Henseler, J. (2016).**

“Using PLS path modeling in new technology business research.”

<https://www.emerald.com/imds/article/116/1/2/176733/Using-PLS-path-modeling-in-new-technology-research>

Describe como SmartPLS ayuda a modelar efectos no lineales en la adopción de nuevas tecnologías en empresas, relevante para el cambio tecnológico en la administración pública y privada.

8. **Gudergan, S.P. (2022).**

“Nonlinear relationships in family business research.”

<https://www.sciencedirect.com/science/article/abs/pii/S1877858521000383>

Analiza y compara la detección de relaciones no lineales en

empresas familiares mediante SmartPLS, aportando metodología para gestión estratégica.

9. **SmartPLS Team (2024).**

“Valoración de los resultados PLS-SEM 3ª parte.”

https://www.smartpls.com/primer-book-case-studies/manual_de_PLS-SEM_ejemplo_ilustrativo_capítulo_6.pdf

Manual de casos aplicados de SmartPLS, con secciones dedicadas al modelado de no linealidad en escenarios empresariales complejos.

10. **Basco, R. et al. (2019).**

“PLS-SEM modelling in family and public management research.”

<https://dialnet.unirioja.es/descarga/articulo/6308426.pdf>

Ejemplifica el uso de SmartPLS para analizar relaciones no lineales tanto en gestión pública como en investigación sobre empresas familiares.

Referencias

- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Sage.
- Basco, R., Hair, J. F., Ringle, C. M., & Sarstedt, M. (2021). *Nonlinear Effects in PLS-SEM*.
- Benson, D., & Ziedonis, R. H. (2009). Corporate Venture Capital as a Window on New Technologies: Implications for the Performance of Corporate Investors when Acquiring Start-ups. *Organization Science*, 20(2), 329-351. <https://doi.org/10.1287/orsc.1080.0385>
- Dijkstra, T. K. (2014). Confirmatory Tetrad Analysis in PLS Path Modeling. *Journal of Business Research*, 67(5), 884-890. <https://doi.org/10.1016/j.jbusres.2013.08.003>
- Dijkstra, T. K., & Henseler, J. (2011). Linear Indices in the Identification of Reflective Constructs in PLS Path Modeling. *MIS Quarterly*, 35(1), 111-117.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2019). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). Sage.
- Henseler, J., Hubona, G. S., & Ray, P. A. (2016). Using PLS Path Modeling in New Technology Research: Updated Guidelines. *Industrial Management & Data Systems*, 116(1), 2-20. <https://doi.org/10.1108/IMDS-09-2015-0382>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling. *Journal of the Academy of Marketing Science*, 43(1), 115-135. <https://doi.org/10.1007/s11747-014-0403-8>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>
- Samuelson, P. A., & Nordhaus, W. D. (2010). *Economía* (19th ed.). McGraw-Hill.

- Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. (2016). The Elephant in the Room: Predictive Performance of PLS Models. *Journal of Business Research*, 69(10), 4552-4564. <https://doi.org/10.1016/j.jbusres.2016.03.049>
- SmartPLS GmbH. (s. f.). *Algorithms and Techniques: Nonlinear Effects*. <https://www.smartpls.com/documentation/algorithms-and-techniques/nonlinear>

CAPITULO V

MODELADO DE LA HETEROGENEIDAD OBSERVADA

La Realidad Oculta tras el Promedio: El Desafío de la Heterogeneidad

En la investigación de negocios y ciencias sociales, uno de los supuestos más frecuentes y peligrosos es la homogeneidad de los datos. A menudo, los investigadores asumen que el modelo teórico propuesto funciona de manera universal para todos los individuos de la muestra. Sin embargo, la realidad empresarial es intrínsecamente diversa: lo que fideliza a un cliente en un mercado emergente puede no tener efecto en un mercado consolidado; los incentivos que motivan a una generación joven pueden ser irrelevantes para una generación mayor.

Al analizar los datos como un único conjunto monolítico, corremos el riesgo de obtener resultados que representan un "promedio" matemático, pero que no describen fielmente a ningún grupo en particular. Este fenómeno puede ocultar diferencias estratégicas críticas o, peor aún, llevar a la generalización de conclusiones erróneas.

Este capítulo se centra en el Modelado de la Heterogeneidad Observada, abordando situaciones donde las diferencias entre los sujetos

proviene de una característica categórica conocida a priori, como el género, el país de origen, el sector industrial o el tipo de empresa.

Para gestionar esta diversidad con rigor científico, exploraremos el Análisis Multigrupo (MGA) en el entorno PLS-SEM. No obstante, antes de comparar si los coeficientes estructurales difieren entre grupos, debemos superar un desafío metodológico fundamental: la Invariancia de la Medición. ¿Estamos midiendo el mismo constructo de la misma manera en todos los grupos? Sin garantizar esta equivalencia a través del protocolo MICOM (*Measurement Invariance of Composite Models*), cualquier comparación posterior carecería de validez empírica.

A lo largo de estas páginas, dejaremos atrás las pruebas paramétricas tradicionales (sujetas a estrictos supuestos de normalidad) para adoptar métodos no paramétricos avanzados, como el Bootstrap MGA y la Prueba de Permutación, reconocidos hoy como los estándares de oro para la comparación robusta de grupos.

El objetivo final de este capítulo no es solo estadístico, sino gerencial: dotar al investigador de la capacidad para detectar matices estructurales que permitan diseñar estrategias contingentes y segmentadas, transformando la complejidad de los datos en decisiones de alta precisión

5.1 Pre-requisito de la Invariancia de la Medición (MICOM) en PLS-SEM

Imaginemos un estudio en ciencias empresariales que busca comparar el impacto del liderazgo transformacional en el desempeño laboral entre empresas de dos países distintos, como México y Brasil. Para que la comparación de los coeficientes estructurales entre ambos contextos sea válida y útil para la alta dirección, es imprescindible antes garantizar que el constructo "liderazgo transformacional" realmente significa lo mismo y es capturado de igual manera en cada grupo. Si los indicadores usados para medir el liderazgo reflejan distintos comportamientos, significados o sesgos culturales, cualquier diferencia en los modelos podría deberse a errores de medición y no a verdaderas diferencias entre los países.

Asegurar este principio fundamental se llama **invariancia de la medición**, y se verifica en PLS-SEM mediante el protocolo MICOM (Measurement Invariance of Composite Models). Sin esta comprobación, las comparaciones pueden llevar a decisiones erróneas y estrategias que no son transferibles de un contexto a otro.

¿Por qué es esencial la invariancia?

La invariancia asegura que los instrumentos estadísticos miden los constructos de manera idéntica en todos los segmentos analizados (grupos de país, género, sector, etc.). Si no se cumple, los resultados comparativos pueden estar sesgados, atribuyendo variaciones a los modelos estructurales

cuando en realidad reflejan diferencias de percepción, interpretación o cultura (Henseler et al., 2016).

Proceso MICOM: Tres Pasos Clave

1. Invariancia Configuracional

- Se verifica que el modelo especificado tenga la misma estructura, los mismos indicadores y rutas en todos los grupos. En nuestro ejemplo, los ítems para liderazgo deben ser los mismos en México y Brasil y su relación con desempeño debe estar modelada de forma idéntica.
- Es requisito técnico y conceptual para cualquier comparación posterior.

2. Invariancia Composicional

- Evalúa si la manera en que los indicadores se combinan para formar el constructo es equivalente entre grupos. Esto se realiza calculando la correlación entre los scores de los constructos mediante permutaciones.
- **Regla de decisión:** La diferencia absoluta de correlación (c) debe ser menor o igual a 0.05 (o, en estudios exploratorios, hasta 0.10). Si se cumple, hay invariancia composicional.

- Determina si existen diferencias significativas en los promedios y/o las varianzas de los constructos entre los grupos analizados.
- Se utiliza la permutación, comparando los estadísticos observados con los esperados bajo invariancia.
- **Regla de decisión:** La diferencia debe estar dentro del umbral de tolerancia ($c \leq 0.05$ o $c \leq 0.10$), indicando invariancia total. Si solo los dos primeros pasos se cumplen, existe invariancia parcial, suficiente para comparar relaciones estructurales (coeficientes) pero no medias/varianzas.

Implicaciones Prácticas y Gerenciales

- **Invariancia total:** Permite comparar tanto las relaciones estructurales como los valores medios y las dispersiones de los constructos. Los resultados son robustos y aplicables gerencialmente.
- **Invariancia parcial:** Habilita la comparación solo de coeficientes estructurales, no de medias o niveles absolutos.
- **Fallo en invariancia:** Cualquier comparación entre grupos es potencialmente sesgada y debería evitarse, pues no puede atribuirse con seguridad a diferencias reales en la teoría empresarial.

La alta dirección debe exigir que todo análisis multigrupo en modelos estructurales se apoye en esta validación, garantizando que las decisiones y estrategias que se derivan sean igualmente válidas, útiles y accionables en los diferentes segmentos o mercados.

5.2 Introducción al MGA y el Método Paramétrico Tradicional

En la investigación multigrupo dentro de modelos de ecuaciones estructurales, un desafío esencial es determinar si los coeficientes estructurales por ejemplo, el impacto de un estilo de liderazgo sobre el desempeño laboral (β) son estadísticamente diferentes entre dos o más grupos, como distintos países, géneros o unidades organizacionales. La técnica central para este objetivo es el **Análisis Multigrupo (MGA, por sus siglas en inglés)**.

El **MGA** permite comparar directamente parámetros del modelo entre grupos, respondiendo preguntas clave como: “¿El efecto de la motivación en la productividad es igual en equipos de ventas frente a equipos de soporte?” o “¿La influencia de la satisfacción laboral varía entre sucursales nacionales e internacionales?”.

El **MGA Paramétrico Tradicional** fue la primera herramienta utilizada en análisis multigrupo. Se basa en la prueba t de Hotelling-Fisher, que compara los coeficientes estructurales (β) entre grupos bajo la siguiente lógica:

- Calcula la diferencia entre los coeficientes estimados para cada grupo.
- Estandariza esta diferencia usando una estadística t, dependiendo de la varianza y el tamaño muestral de cada grupo.
- Contrasta el valor obtenido contra una distribución t para juzgar si la diferencia es significativa (habitualmente a un nivel de significancia de 0.05).

Sin embargo, el uso del método paramétrico tradicional presenta limitaciones serias:

- **Asume normalidad** en la distribución de los errores del modelo.
- **Requiere homocedasticidad** (igualdad de varianzas) entre grupos.
- Estos supuestos suelen no cumplirse en la práctica, especialmente en ciencias sociales y datos empresariales reales, generando riesgo de falsas conclusiones.

Por este motivo, la literatura contemporánea recomienda métodos alternativos más robustos y flexibles, como el **bootstrap** y la **permutación**, que no requieren esas suposiciones estrictas y ofrecen estimaciones empíricas más fiables sobre la diferencia real entre grupos.

En suma, aunque el MGA paramétrico tradicional estableció las bases de la comparación multigrupo, hoy es considerado una herramienta

menos rigurosa, especialmente en contextos que demandan validez y replicabilidad alta. Los métodos basados en remuestreo ha pasado a ser el estándar metodológico moderno para comparar coeficientes estructurales de forma precisa y robusta.

Actualmente, la literatura recomienda métodos más robustos para comparación multigrupo, como el bootstrap y la permutación, ya que éstos no dependen de la normalidad ni de la homogeneidad de varianzas y proporcionan resultados más válidos y replicables (Hair et al., 2019; Sarstedt et al., 2011).

5.3 Rigor Estadístico: MGA basado en Bootstrap

El MGA basado en Bootstrap representa un avance significativo en la comparación de coeficientes estructurales (β) entre grupos en modelos de ecuaciones estructurales. A diferencia del MGA paramétrico tradicional, este método no depende de supuestos exigentes como la normalidad o la homocedasticidad de los errores, lo que lo hace especialmente adecuado para datos reales y estructuras complejas comúnmente encontradas en las ciencias empresariales (Hair et al., 2019; Sarstedt et al., 2011).

El **Bootstrap MGA** utiliza técnicas de remuestreo para estimar la distribución empírica de los parámetros y permite realizar inferencias robustas sobre diferencias entre grupos. Una variante popular es el enfoque de *Partial Confidence Interval Overlap* (PCIO), que evalúa si los intervalos

de confianza de los coeficientes de los grupos se solapan respecto a la diferencia deseada.

Metodología del MGA basado en Bootstrap:

1. Ejecución del bootstrap en el modelo multigrupo:

Cada grupo se somete a un procedimiento de remuestreo (por ejemplo, 5,000 réplicas), donde, en cada réplica, se estima el modelo y se obtienen los coeficientes estructurales.

2. Cálculo de la diferencia entre los coeficientes:

Para cada bootstrap, se calcula la diferencia entre el coeficiente estructural del Grupo 1 y el del Grupo 2:

$$\Delta\beta = \beta_{\text{Grupo 1}} - \beta_{\text{Grupo 2}}$$

En los modelos de ecuaciones estructurales para calcular la diferencia en el peso o la fuerza de un predictor (β) entre dos grupos distintos (Grupo 1 y Grupo 2).

- $\Delta\beta$ (**Delta Beta**): Es la **diferencia** en los coeficientes de regresión (o pesos estandarizados) entre los dos grupos.
- $\beta_{\text{Grupo 1}}$: Es el coeficiente de regresión (β) para una variable predictora específica en el **primer grupo**.
- $\beta_{\text{Grupo 2}}$: Es el coeficiente de regresión (β) para la misma variable predictora específica en el **segundo grupo**.

Esta diferencia se suele calcular para realizar una prueba estadística (como la prueba de invarianza de la pendiente de regresión) para determinar si la relación entre la variable predictora y la variable dependiente **es significativamente diferente** entre los dos grupos.

Esto genera una distribución empírica de diferencias.

3. **Determinación de la significancia usando intervalos de confianza:**

Se obtienen intervalos de confianza (usualmente al 95%) para $\Delta\beta$.

- Si el intervalo excluye el valor cero, hay evidencia estadística de que los grupos difieren de manera significativa en la relación modelada.
- Si el intervalo incluye cero, no se puede afirmar que exista una diferencia estadísticamente significativa entre los coeficientes de los grupos para ese parámetro.

Esta metodología confiere un rigor superior y resultados más confiables y replicables para la comparación multigrupo, haciendo que el bootstrap MGA sea el estándar recomendado en la práctica moderna del modelado estructural con PLS-SEM (Hair et al., 2019).

5.4 Superando Limitaciones: La Prueba de Permutación (Permutation MGA)

En la evaluación de heterogeneidad entre grupos en modelos PLS-SEM, la **prueba de permutación** se ha establecido como el estándar de oro metodológico, superando ampliamente tanto los métodos paramétricos tradicionales como el enfoque bootstrap. Su principal fortaleza recae en la generación de una **distribución nula empírica** para la diferencia de coeficientes, fundamentada en la aleatorización real de los datos (Henseler et al., 2009; Sarstedt et al., 2011).

La lógica de la prueba de permutación se basa en mezclar las observaciones de los grupos originales (por ejemplo, dos segmentos de clientes o países) y reasignar etiquetas de grupo de forma aleatoria en cada iteración. Así, se construyen submuestras que simulan el escenario en que **no existe heterogeneidad real** entre los grupos. Luego, se calcula la diferencia de coeficientes estructurales ($\Delta\beta$) para cada permutación, generando una distribución nula precisa de la diferencia esperada bajo la hipótesis de igualdad.

Metodología de la Permutation MGA:

1. Se combinan todos los casos de ambos grupos en una sola muestra.

2. Mediante permutaciones repetidas (por ejemplo, 5,000), se asignan etiquetas de grupo aleatoriamente, manteniendo los tamaños originales.
3. Se estima el modelo PLS-SEM para cada permutación y se calcula la diferencia de coeficientes ($\Delta\beta$) en cada iteración.
4. Se compara la diferencia observada (real) de coeficientes contra la distribución nula obtenida mediante permutación.

Regla de Decisión

- El método proporciona un valor p (p-value) bilateral para la diferencia observada:
 - Si el p-value es **menor o igual a 0.05** (bilateral), existe evidencia estadística de **heterogeneidad significativa** en el coeficiente estructural entre los grupos.
 - Para pruebas unilaterales, el umbral recomendado es **$p \leq 0.025$** .
- Si el p-value es mayor que estos umbrales, no se puede concluir que exista heterogeneidad relevante.

Ventajas y Robustez

La prueba de permutación **no requiere ninguna asunción distribucional** como normalidad o igualdad de varianzas, lo que la hace robusta ante violaciones de supuestos y la heterogeneidad real de los datos empresariales y sociales. Por eso, se considera el método preferido y más preciso para contrastar diferencias entre grupos y respaldar decisiones gerenciales o científicas con alta confianza (Sarstedt et al., 2011; Hair et al., 2019).

5.5 Estrategias Avanzadas: Manejo de la comparación entre múltiples grupos ($k > 2$)

En investigaciones con múltiples grupos, como diferentes generaciones o países, es fundamental comparar coeficientes estructurales de forma rigurosa. Cuando el número de grupos excede dos ($k > 2$), el Análisis Multigrupo (MGA) debe aplicarse mediante comparaciones pareadas: Grupo 1 vs. Grupo 2, Grupo 1 vs. Grupo 3, Grupo 2 vs. Grupo 3, y así sucesivamente (Hair et al., 2019; Sarstedt et al., 2011).

Inflación del Error de Tipo I

La realización de múltiples comparaciones incrementa el riesgo de encontrar diferencias significativas por azar (error de Tipo I). Este fenómeno, conocido como el problema de las comparaciones múltiples, puede comprometer la validez estadística de los hallazgos si no se controla adecuadamente (Hair et al., 2019).

Métodos de Corrección y Pruebas Omnibus

Para mitigar este riesgo, se recomienda:

- **Corrección de Bonferroni:**

Ajusta el nivel de significancia dividiéndolo por el número total de comparaciones ($\alpha_{\text{ajustado}} = \alpha/m$). Así, con seis comparaciones y un α original de 0.05, el nuevo umbral sería 0.0083 para cada prueba individual. Esta corrección reduce la probabilidad de errores Tipo I y aumenta la robustez de los resultados (Hair et al., 2019; Sarstedt et al., 2011).

- **Prueba Omnibus:**

Cuando el software lo permite, se recomienda realizar una prueba ómnibus antes de las comparaciones pareadas. La prueba omnibus evalúa la hipótesis global de que todos los grupos son iguales, y solo si se rechaza esta hipótesis procede la comparación por pares (Sarstedt et al., 2011).

Es fundamental que el investigador justifique explícitamente la estrategia de corrección empleada y reporte los ajustes y umbrales de manera transparente.

5.6 Aplicación Práctica y Workflow: Un Estudio de Caso

Para consolidar la teoría y la práctica en modelos de ecuaciones estructurales multigrupo, presentemos un flujo de trabajo aplicado en el estudio “Diferencias de Género en la Lealtad a la Marca”, donde el objetivo es evaluar si el efecto de la “Confianza” sobre la “Lealtad” es diferente entre hombres y mujeres.

Paso 1: Evaluación del MICOM (Measurement Invariance of Composite Models)

El primer paso consiste en comprobar la invariancia de la medición entre los grupos (género). Una vez preparado el modelo, se ejecuta el protocolo MICOM que incluye: invariancia configuracional, composicional y de media/varianza (Hair et al., 2019).

- Si la invariancia no se cumple (fallo en MICOM), la interpretación de cualquier diferencia estructural queda comprometida y no es válido continuar con el análisis comparativo.
- Si se confirma la invariancia parcial o total, puede procederse con seguridad.

Paso 2: Procedimiento del MGA (Multigroup Analysis)

Una vez garantizada la invariancia, se procede a comparar el coeficiente estructural clave, es decir, el efecto de “Confianza” sobre

“Lealtad” en ambos segmentos (hombres y mujeres). Se recomienda utilizar la prueba de permutación como método robusto para evitar supuestos distribucionales y proporcionar resultados fiables (Sarstedt et al., 2011).

Paso 3: Presentación y Decisión con MGA Permutación

En el análisis, se obtienen los valores $\beta_{\text{Confianza, Hombres}}$ y $\beta_{\text{Confianza, Mujeres}}$. El MGA por permutación genera un p-value bilateral para la diferencia de coeficientes.

- Si el p-value es menor o igual a 0.05, la diferencia es significativa: el impacto de la confianza en la lealtad varía estructuralmente entre hombres y mujeres (Hair et al., 2019; Sarstedt et al., 2011).
- Si el p-value supera el umbral, se concluye que no existen diferencias sustanciales entre grupos.

Este flujo lógico asegura que la comparación multigrupo sea rigurosa y que cualquier recomendación gerencial resulte de hallazgos estadísticamente sólidos y comparaciones válidas.

5.7 Implicaciones gerenciales: Traducción de la heterogeneidad a estrategias contingentes

La incorporación del Análisis Multigrupo (MGA) en la gestión

competitiva decisiva. El valor del MGA radica en su capacidad para revelar **heterogeneidad en el funcionamiento de los mercados, segmentos y grupos de clientes**, permitiendo a la alta dirección fundamentar el diseño de **estrategias contingentes** y diferenciadas en función de evidencia empírica robusta (Hair et al., 2019; Sarstedt et al., 2011).

Cuando el MGA confirma que el impacto de una variable clave (X, como “Confianza”) sobre un resultado estratégico (Y, como “Lealtad a la Marca”) es significativamente más fuerte en el Grupo A que en el Grupo B, la empresa está ante una poderosa evidencia para **asignar recursos de manera diferenciada**. Por ejemplo, si “Confianza” predice “Lealtad” mucho más en jóvenes que en adultos, las acciones de marketing, comunicación y experiencia de producto deben concentrarse proporcionalmente en el segmento joven, maximizando así el Retorno de la Inversión (ROI).

Esto permite a la administración:

- Priorizar la inversión en el grupo más receptivo al estímulo o intervención.
- Optimizar el uso del presupuesto de marketing y recursos operativos.
- Justificar estrategias segmentadas ante el directorio y stakeholders, mostrando un **enfoque basado en datos** para alcanzar mejores resultados.

- Monitorear y ajustar contingentemente las políticas comerciales ante evidencia nueva, sustentando decisiones en la dinámica real de los mercados.

Así, el MGA no sólo identifica diferencias estructurales entre grupos, sino que empodera a la alta dirección para construir estrategias y asignaciones óptimas, personalizadas y ajustables, asegurando que cada inversión rinda su máximo potencial contextual.

5.8 Investigaciones sobre: Modelado de la heterogeneidad observada, en bases de datos:

1. **Gudergan, S.P., et al. (2025). Advanced partial least squares SEM in business research.**
<https://www.sciencedirect.com/science/article/pii/S0148296324005915>
 Revisión sobre técnicas avanzadas de PLS-SEM, abordando el análisis de moderadores y el modelado de la heterogeneidad observada en relaciones estructurales empresariales.
2. **CC Astuti et al. (2025). Segmentation of partial least squares structural equation models using kernel-based clustering.**
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12408404/>
 Propone un método de segmentación para solventar la heterogeneidad no observada en modelos PLS-SEM mediante clustering, útil en gestión organizacional y pública.
3. **SmartPLS Team (2019). Finite Mixture Partial Least Squares (FIMIX-PLS).**
<https://www.smartpls.com/documentation/algorithms-and-techniques/fimix-pls/>
 Tutorial sobre FIMIX-PLS, técnica para segmentar y modelar heterogeneidad no observada en datos empresariales, relevante para estudios de gestión y marketing.

4. **Soriano, J.L. (2022). Modelado de ecuaciones estructurales en el campo de las ciencias de la administración.**
<https://www.econstor.eu/bitstream/10419/286270/1/1814673792.pdf>
 Analiza aplicaciones de SEM y SmartPLS en la administración, abordando la detección y ajuste por heterogeneidad observada en modelos empresariales.
5. **Martínez Ávila, M. (2018). Aplicación de la técnica PLS-SEM en la gestión del conocimiento empresarial.**
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-74672018000100130
 Explica el procedimiento de modelado SEM en SmartPLS incluyendo cómo manejar diferentes segmentos y heterogeneidad de datos organizacionales.
6. **Ratzmann, M. (2016). Capturing heterogeneity and PLS-SEM prediction ability.**
<https://www.sciencedirect.com/science/article/abs/pii/S0148296316301333>
 Evalúa cómo la heterogeneidad observada puede condicionar la capacidad predictiva en PLS-SEM para innovación y confianza empresarial.
7. **Shela, V. (2023). Una revisión sistemática de modelos PLS-SEM y heterogeneidad.**
<https://www.sciencedirect.com/science/article/pii/S2405844023096846>
 Revisión sobre la aplicación de PLS-SEM, abordando el manejo de la heterogeneidad observada en mediciones formativas y reflexivas en negocios.
8. **Ismael, K. (2021). Investigación de la calidad percibida con PLS-SEM y heterogeneidad observada.**
<https://www.mdpi.com/2071-1050/13/23/13018>
 Estudio de la calidad percibida y su relación con diferentes segmentos de consumidores, modelando la heterogeneidad en contexto público.
9. **Hair, J.F. et al. (2016). Identifying and treating unobserved heterogeneity with FIMIX-PLS.**
<https://epub.ub.uni-muenchen.de/96135/>

Aborda segmentación de clases latentes y estrategias para tratar la heterogeneidad no observada en SmartPLS, relevante para administración.

10. **Magno, F. (2024). Una breve revisión del uso de PLS-SEM en gestión de calidad y heterogeneidad.**
<https://www.emerald.com/tqm/article/36/5/1242/1217482/A-brief-review-of-partial-least-squares-structural>

Explora cómo PLS-SEM ayuda a modelar y segmentar la heterogeneidad en estudios de gestión de calidad empresarial.

Referencias

- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2019). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). Sage.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing Measurement Invariance of Composites Using Partial Least Squares. *International Marketing Review*, 33(3), 405-431. <https://doi.org/10.1108/IMR-09-2014-0304>
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The Use of Partial Least Squares Path Modeling in International Marketing. *Advances in International Marketing*, 20, 277-319.
- Keil, M., Tan, B. C. Y., Wei, K.-K., Saarinen, T., Tuunainen, V., & Wassenaar, A. (2000). A Cross-Cultural Study on Escalation of Commitment Behavior in Software Projects. *MIS Quarterly*, 24(2), 299-325. <https://doi.org/10.2307/3250940>
- Sarstedt, M., Henseler, J., & Ringle, C. M. (2011). Multigroup Analysis in Partial Least Squares Path Modeling: Alternative Methods and Empirical Results. *Advances in International Marketing*, 22, 195-218.

CAPITULO VI

MODELADO DE LA HETEROGENEIDAD NO OBSERVADA

Descubriendo la Estructura Oculta del Mercado

En la investigación empresarial convencional, a menudo caemos en la tentación de segmentar a nuestros sujetos de estudio basándonos en lo visible: género, edad, tamaño de empresa o ubicación geográfica. Sin embargo, la realidad estratégica es mucho más esquivada. ¿Qué sucede cuando dos clientes de la misma edad y ciudad responden de manera opuesta a una misma campaña de marketing? ¿Por qué dos empleados con el mismo cargo y antigüedad muestran niveles de compromiso radicalmente distintos?

La respuesta reside en la Heterogeneidad No Observada (o latente). Este concepto reconoce que dentro de cualquier conjunto de datos existen subgrupos "invisibles" cuyos modelos mentales y causales difieren significativamente, pero que no pueden ser identificados mediante variables demográficas a priori. Estos segmentos se agrupan en función de factores psicológicos, culturales o motivacionales profundos que escapan a la segmentación tradicional.

Este capítulo le guiará a través de las metodologías más avanzadas para detectar y modelar esta complejidad oculta dentro del marco PLS-SEM:

1. **FIMIX-PLS (Finite Mixture Partial Least Squares):** Aprenderá a utilizar esta técnica de clases latentes para identificar teóricamente cuántos segmentos existen realmente en sus datos y cómo difieren sus estructuras causales. Exploraremos los criterios de información (AIC, BIC) y las pruebas de ratio de verosimilitud (LMR, BLRT) necesarios para determinar el número óptimo de grupos.
2. **PLS-POS (Prediction-Oriented Segmentation):** Introduciremos una técnica alternativa diseñada no solo para ajustar el modelo, sino para maximizar la capacidad predictiva (R^2) de un constructo objetivo. Veremos cómo esta herramienta es vital para la toma de decisiones gerenciales cuando el objetivo es la optimización de resultados (ventas, lealtad, eficiencia).
3. **Del Dato a la Estrategia:** Más allá de la matemática, nos enfocaremos en la interpretación gerencial. Descubrirá cómo caracterizar estos segmentos latentes (ej. "Clientes Racionalistas" vs. "Emocionales") analizando sus diferencias en medias, cargas y coeficientes, transformando la estadística abstracta en estrategias de personalización profunda y asignación eficiente de recursos.

Prepárese para dejar de mirar a su mercado como un bloque monolítico y comenzar a ver la rica diversidad estratégica que se esconde bajo la superficie de sus datos

6.1 Fundamentos Teóricos y Racional del FIMIX-PLS

En la investigación organizacional y administrativa, la noción de **heterogeneidad no observada** (o latente) es fundamental para entender la complejidad de los comportamientos empresariales y de mercado. La heterogeneidad no observada hace referencia a la existencia de subgrupos dentro de una muestra cuyos modelos causales difieren significativamente, pero que **no se pueden definir objetivamente** por variables a priori como género, edad, sector o ubicación. Estos segmentos pueden estar influenciados por factores psicológicos, culturales, o dinámicas organizacionales que no son capturados de manera directa en el conjunto de datos (Becker et al., 2013).

Por ejemplo, en una empresa internacional, los empleados podrían agruparse en segmentos latentes según su orientación hacia la innovación— independientemente de su departamento, antigüedad o país y estos segmentos presentar diferentes patrones en la relación entre formación, ambiente laboral y desempeño.

Para abordar esta complejidad, se ha desarrollado el método Finite Mixture Partial Least Squares (FIMIX-PLS). FIMIX-PLS es una técnica de análisis de clases latentes que integra el marco PLS-SEM y permite explorar

y estimar estructuras subyacentes no observadas en los datos. En vez de asumir que la totalidad de los casos sigue un único modelo causal, FIMIX-PLS identifica el **número óptimo de segmentos latentes** y estima los parámetros específicos del modelo PLS-SEM para cada grupo. Esto facilita una interpretación más precisa y alineada con la realidad compleja y segmentada de las organizaciones (Becker et al., 2013; Sarstedt et al., 2011).

El procedimiento incluye:

1. Ejecutar FIMIX-PLS, que utiliza algoritmos de mezcla finita para asignar probabilísticamente los casos a diferentes clases latentes.
2. Determinar el número óptimo de segmentos utilizando criterios como el Bayesian Information Criterion (BIC).
3. Estimar los coeficientes estructurales y de medición propios para cada segmento, permitiendo comparar cómo operan los modelos causales entre grupos no definidos a priori.

La aplicación de FIMIX-PLS en ciencias administrativas permite revelar tipologías organizacionales ocultas, segmentar mercados o identificar patrones de respuesta en clientes, empleados o unidades estratégicas, optimizando la toma de decisiones y las estrategias de personalización empresarial.

6.2 Criterios de selección y evaluación del número de segmentos en FIMIX-PLS

Uno de los pasos clave en el análisis de clases latentes con FIMIX-PLS es la selección del número óptimo de segmentos, cada uno con patrones causales distintos, pero no observables directamente. La decisión debe balancear la precisión del ajuste con la parsimonia del modelo, utilizando métricas estadísticas robustas.

1. Indicadores de Información: AIC, CAIC y BIC

Los criterios de información como el Akaike Information Criterion (AIC), el Bayesian Information Criterion (BIC) y el Consistent Akaike Information Criterion (CAIC) permiten comparar modelos con diferente número de clases latentes:

- **AIC:** Penaliza la complejidad del modelo y busca minimizar la pérdida de información.
- **BIC y CAIC:** Penalizan más fuertemente los modelos complejos, favoreciendo modelos más parsimoniosos.

Regla operativa:

Se calculan estos indicadores para el modelo con 1 segmento, 2 segmentos, 3 segmentos, etc. El número óptimo de segmentos será aquel que tenga el **valor más bajo** en estas métricas.

Por ejemplo, si al pasar de dos a tres segmentos el BIC baja de 420 a 415, el ajuste mejora; si agregar un cuarto segmento lo sube a 418, el mejor modelo es el de tres segmentos.

2. Test de la Ratio de Verosimilitud: LMR y BLRT

El criterio más importante en la práctica es la Ratio de Verosimilitud de Lo-Mendell-Rubin (LMR) y el Bootstrapped Likelihood Ratio Test (BLRT):

- Estos tests comparan de forma formal el modelo con k segmentos frente a uno con $k-1$ segmentos.
- Se calcula la diferencia entre la verosimilitud de ambos modelos y se determina si es estadísticamente significativa.
- Un **p-value bajo (por ejemplo, < 0.05)** indica que el modelo con k segmentos mejora significativamente el ajuste, permitiendo rechazar la solución con menos segmentos.

Ejemplo:

Supongamos que se analizan clientes de una empresa y se prueba modelos de hasta cuatro segmentos:

- El modelo de dos segmentos tiene $BIC = 430$; el de tres segmentos, $BIC = 420$; el de cuatro segmentos, $BIC = 425$.

- El BLRT arroja $p = 0.02$ al comparar dos vs. tres segmentos (significativo) y $p = 0.20$ al comparar tres vs. cuatro segmentos (no significativo).
- Se selecciona tres segmentos como óptimo, pues tiene el BIC más bajo y el BLRT confirma que mejora significativamente respecto a dos segmentos, pero agregar un cuarto no ofrece ventaja.

La convergencia de estos criterios proporciona una base estadística robusta para determinar el número de segmentos y garantiza que cada grupo identificado tenga relevancia práctica y validez empírica en la segmentación organizacional.

6.3 Segmentación Orientada a Predicciones para Mínimos Cuadrados Parciales (PLS-POS)

En el ámbito de la segmentación avanzada, la técnica PLS-POS (Partial Least Squares Prediction-Oriented Segmentation) surge como una alternativa metodológica robusta frente a técnicas tradicionales como FIMIX-PLS. Mientras FIMIX prioriza el ajuste estadístico global del modelo, PLS-POS enfatiza la capacidad predictiva de cada segmento en relación al constructo objetivo que interesa al investigador (Becker et al., 2018).

La lógica de PLS-POS se fundamenta en maximizar el coeficiente de determinación (R^2) de la variable dependiente clave, asegurando que la

segmentación identificada contribuya directamente a mejorar la precisión en la predicción de resultados relevantes. Esta característica resulta especialmente útil cuando el propósito principal del análisis es optimizar las predicciones y no solo mejorar el ajuste del modelo, siendo idónea para aplicaciones en marketing, tipología de clientes e incluso gestión pública.

Por ejemplo, en una entidad gubernamental que busca segmentar ciudadanos para anticipar el nivel de satisfacción con un nuevo servicio digital, PLS-POS permite formar grupos latentes cuya configuración esté orientada a maximizar la capacidad de predecir la satisfacción. Así, la administración puede diseñar estrategias adaptadas a los segmentos con mayor potencial de respuesta, focalizando esfuerzos en áreas donde la R^2 es óptima, incrementando la efectividad de campañas públicas o inversiones en atención ciudadana.

En contraste con el ajuste tradicional, PLS-POS da prioridad a la utilidad predictiva de la segmentación, una ventaja decisiva para responsables de marketing, gestores de servicios y analistas públicos que buscan optimizar recursos y resultados a partir de datos reales y enfoques predictivos.

6.4 Comparación y selección estratégica de la técnica de segmentación

La elección adecuada de la técnica de segmentación en modelos PLS-SEM es esencial para alinear los resultados metodológicos con los

objetivos estratégicos del estudio. Existen dos enfoques avanzados, FIMIX-PLS y PLS-POS, cada uno con fortalezas diferenciadas y una aplicabilidad ideal que debe ser cuidadosamente considerada y justificada al reportar resultados.

Marco de decisión

- **FIMIX-PLS** es recomendable cuando el propósito principal es la identificación teórica de subgrupos latentes que presentan diferencias en las relaciones causales del modelo. Es la opción preferida para la comprensión y exploración de patrones estructurales diversos dentro de la muestra, facilitando el descubrimiento de segmentos con trayectorias causales particulares (Becker et al., 2013; Sarstedt et al., 2011). Se emplea cuando se busca profundizar en las bases teóricas del comportamiento organizacional, consumidor o ciudadano.
- **PLS-POS** se orienta a contextos donde la maximización de la capacidad predictiva es prioritaria. Esta técnica segmenta los datos para optimizar la explicación (mayor R^2) del constructo objetivo, resultando ideal para escenarios de toma de decisiones, marketing analítico o gestión pública basada en la predicción y la optimización de recursos (Becker et al., 2018).

Instrucciones para reportar la justificación metodológica

En el apartado de Análisis de Datos del reporte de investigación, el investigador debe:

- Especificar claramente el objetivo principal del análisis de segmentación (teórico-causal o predictivo-decisional).
- Justificar la elección de FIMIX-PLS cuando el interés radica en la comprensión de heterogeneidad latente y las diferencias estructurales, citando la literatura relevante y el marco conceptual del estudio.
- Optar por PLS-POS y explicitarlo cuando la meta es maximizar la capacidad explicativa para soportar acciones concretas o decisiones gerenciales.
- Describir de forma transparente los criterios comparados, las métricas utilizadas para la selección y los beneficios de la técnica elegida para los objetivos del trabajo.

6.5 Interpretación Detallada y Caracterización de los Segmentos Latentes

Una vez identificados los segmentos latentes mediante técnicas como FIMIX-PLS o PLS-POS, la verdadera aportación gerencial depende de la caracterización precisa de cada grupo. Este análisis es fundamental para transformar resultados estadísticos en perfiles útiles y accionables para la estrategia organizacional y la gestión del cambio.

La caracterización de los segmentos debe abordarse a través de tres conjuntos de métricas clave:

1. **Diferencias en las puntuaciones medias latentes (Desempeño)**

Analizar las medias de los constructos latentes para cada segmento permite identificar grupos con mejor o peor desempeño en dimensiones críticas.

Ejemplo: En el contexto de satisfacción y lealtad a una marca, un segmento puede presentar puntuaciones medias significativamente más altas en “Satisfacción del Cliente” que otros, posicionándolo como prioritario para acciones de fidelización o retención.

2. **Diferencias en las cargas factoriales (Patrón de Medición del Segmento)**

Examinar las cargas factoriales por segmento revela cómo cada grupo interpreta y responde a los indicadores de un constructo. Puede haber segmentos que atribuyen mayor importancia a ciertos elementos del servicio (ej., “Rapidez de Respuesta” vs. “Amabilidad” en atención al cliente).

Ejemplo: En recursos humanos, un segmento de empleados podría valorar mucho el equilibrio vida-trabajo en su medición del “Clima Laboral”, mientras otro prioriza la “Remuneración”.

3. Diferencias en los Coeficientes Estructurales (β) El Hallazgo Clave

Comparar los coeficientes estructurales (relaciones causales) entre segmentos es el mayor aporte del análisis: permite descubrir modelos causales únicos por grupo.

Ejemplo: En marketing, el efecto de la “Calidad Percibida” sobre la “Lealtad” podría ser fuerte en un segmento de clientes tecnológicos, pero débil en un segmento tradicional; o en gestión pública, la influencia de la “Confianza en la Institución” sobre la “Participación Ciudadana” puede variar drásticamente según el perfil latente.

Esta triple caracterización convierte los segmentos en tipologías concretas, informando el diseño de políticas, productos y estrategias personalizadas. Gerencialmente, permite responder: ¿Qué grupo ofrece mayor retorno potencial? ¿Dónde invertir recursos para obtener máximo impacto? ¿Qué necesidades, motivaciones y patrones son únicos en cada segmento?

Esta metodología propulsa una gestión basada en evidencia y diferenciación, recomendando no solo “a quién” dirigir la acción, sino también “cómo” y “con qué foco” hacerlo, rediseñando la ventaja competitiva y la atención segmentada en ciencias empresariales.

6.6 Guía de Aplicación Práctica y Workflow de Segmentación en SmartPLS

I. Configurar el Modelo Base PLS

- Importa el archivo de datos y define tu modelo PLS-SEM base en la interfaz gráfica: identifica los constructos latentes, relaciones de caminos y los indicadores asociados.

II. Ejecutar FIMIX-PLS con un Rango de 2 a 5 Segmentos

- Abre la sección de segmentación y selecciona FIMIX-PLS como algoritmo de mezcla finita.
- Especifica el rango de segmentos a probar (por ejemplo, de 2 hasta 5).
- Ejecuta el análisis, asegurándote de que la opción para guardar la variable de pertenencia a segmentos esté activada.

III. Analizar los Criterios de Selección (AIC, BIC, LMR)

- Examina los valores de AIC, BIC y CAIC para cada número de segmentos. Recuerda, valores más bajos indican mejor ajuste.
- Consulta el resultado del test de significado Lo-Mendell-Rubin (LMR) o el Bootstrapped Likelihood Ratio Test (BLRT). Un p-

value significativo indica que el número de segmentos actual mejora la opción anterior.

- Selecciona el número óptimo de segmentos según los mejores criterios y la significancia.

IV. Extraer la Variable de Pertenencia al Segmento Latente del Output de FIMIX

- Una vez identificado el modelo óptimo, exporta la variable de pertenencia al segmento latente (por ejemplo, “Segmento FIMIX”).
- Esta variable asigna a cada caso de tu base de datos un segmento determinado.

V. Utilizar Esta Variable como Moderador Categórico en un MGA Tradicional (o MGA Permutación)

- Vuelve a la sección de Análisis Multigrupo (MGA) en SmartPLS.
- Emplea la variable “Segmento FIMIX” como moderador categórico, creando los subgrupos según la pertenencia latente definida.
- Realiza el MGA tradicional (paramétrico) o, preferiblemente, el MGA por permutación para comparar los coeficientes estructurales (β), cargas factoriales y medias entre los segmentos.

- Verifica la significancia de las diferencias entre segmentos para validar la relevancia práctica de la segmentación inicial.

Resultado: Este workflow integra la segmentación latente, la optimización con criterios estadísticos y la validación de diferencias entre grupos, permitiendo al usuario de SmartPLS 4 transformar la heterogeneidad estadística en insights gerenciales y estrategias diferenciadas basadas en evidencia rigurosa.

6.7 Implicaciones gerenciales: El valor estratégico de la segmentación no observada

La segmentación no observada representa una evolución crítica en la gestión empresarial moderna, al permitir la identificación de grupos latentes con patrones comportamentales y motivacionales únicos, que no surgen de variables sociodemográficas predefinidas como edad, género o ingresos. Mediante técnicas como FIMIX-PLS y PLS-POS, se pueden descubrir tipologías tales como “Clientes Racionalistas” y “Clientes Emocionales”, que reflejan diferencias profundas en el proceso de decisión y respuesta ante productos y comunicaciones de la empresa (Becker et al., 2013; Becker et al., 2018).

La utilidad fundamental de esta aproximación radica en el diseño de estrategias de producto y comunicación altamente personalizadas. Por ejemplo, los “Clientes Racionalistas” pueden responder mejor a mensajes que enfatizan la funcionalidad, eficiencia y economía, mientras que los

“Clientes Emocionales” accionan frente a estímulos basados en storytelling, identidad de marca y experiencias memorables. Estas estrategias segmentadas superan ampliamente el valor de la personalización basada solo en perfiles sociodemográficos, permitiendo a la empresa asignar recursos publicitarios, innovar productos y definir canales de venta allí donde el impacto es máximo.

El resultado es una asignación de recursos significativamente más eficiente. La inversión y el esfuerzo se concentran en los segmentos de mayor potencial de respuesta y fidelización, optimizando el Retorno de la Inversión (ROI) y reduciendo el desperdicio en campañas o desarrollos no alineados con las verdaderas motivaciones de cada grupo (Sarstedt et al., 2011; Becker et al., 2018). Además, esta aproximación capacita a la alta dirección para anticipar cambios en el mercado, monitorear la evolución de tipologías y ajustar la estrategia de forma dinámica ante nuevas segmentaciones emergentes.

En conclusión, la segmentación no observada transforma la investigación de mercados y la gestión estratégica permitiendo a las organizaciones tomar decisiones basadas en la heterogeneidad real de sus clientes y públicos, potenciando la personalización profunda y la maximización del valor en cada acción gerencial.

6.8 Investigaciones sobre: Modelado de la heterogeneidad no observada, en la literatura científica:

1. **Heterogeneidad no observada en la satisfacción laboral del sector público peruano**
http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2219-71682024000300224
 Analiza la heterogeneidad no observada en la satisfacción laboral—usando técnicas latentes—en trabajadores públicos de Perú, mostrando cómo ignorar esta fuente puede afectar los resultados.
2. **La heterogeneidad del consumidor en los modelos de elección**
<https://dialnet.unirioja.es/descarga/articulo/2504994.pdf>
 Revisión de mecanismos de modelado de preferencias y respuestas personales, con especial énfasis en el impacto de la heterogeneidad no observada en las funciones de elección en marketing.
3. **Solving unobserved heterogeneity with latent class inflated Poisson regression model**
https://pmc.ncbi.nlm.nih.gov/articles/PMC9336471/pdf/CJAS_49_1929875.pdf
 Propuesta metodológica y aplicación del modelo Poisson latente para abordar la heterogeneidad no observada en datos de conteo, con ejemplos en encuestas de salud.
4. **Finite Mixture Partial Least Squares (FIMIX-PLS)**
<https://www.smartpls.com/documentation/algorithms-and-techniques/fimix-pls/>
 Revisión y fundamentación del método FIMIX-PLS para detectar segmentos latentes y tratar heterogeneidad no observada en modelos PLS-SEM.
5. **A New Formulation for Latent Class Models**
<https://www.sheffield.ac.uk/media/3920/download>
 Artículo técnico que revisa la aplicación de modelos de clases latentes para capturar heterogeneidad no observada en ciencias sociales y económicas.

6. **A general latent class approach to unobserved heterogeneity**
<https://repository.tilburguniversity.edu/server/api/core/bitstreams/77bd40b3-ead2-4fba-9a38-fa866630c57c/content>
 Metodología extendida de clases latentes para corregir heterogeneidad no observada, con aplicaciones empíricas en distintas disciplinas.
7. **Endogeneidad y heterogeneidad no observada: Menú de hoy**
<https://ccp.ucr.ac.cr/documentos/portal/conversatorios/2010/gpinto.pdf>
 Ensayo sobre modelos econométricos y el desafío que representa la heterogeneidad no observada para la validez de las inferencias.
8. **Breve reseña del uso del modelado de ecuaciones estructurales PLS en calidad total**
https://www.researchgate.net/publication/363890517_A_Brief_Review_of_Partial_Least_Squares_Structural_Equation_Modeling_PLS-SEM_Use_in_Quality_Management_Studies
 Incluye una revisión de FIMIX-PLS y otras técnicas de clases latentes aplicadas a la gestión de la calidad.
9. **Modelación de heterogeneidad en la percepción cualitativa de paisajes**
<https://repositorio.uc.cl/handle/11534/38940>
 Ejemplo aplicado sobre cómo la heterogeneidad no observada puede alterar significativamente la interpretación de percepciones cualitativas en estudios medioambientales.
10. **A General Latent Class Approach to Unobserved Heterogeneity in the Analysis of Event History Data**
<https://jeroenvermunt.nl/hagenaars2002a.pdf>
 Propuesta metodológica avanzada para incorporar clases latentes en el modelado de heterogeneidad no observada en estudios longitudinales.

Referencias

- Becker, J.-M., Ringle, C. M., Sarstedt, M., & Völckner, F. (2013). How Collinearity Affects Mixture Regression Results. *Marketing Letters*, 24(4), 775-791. <https://doi.org/10.1007/s11002-012-9219-9>
- Becker, J.-M., Ringle, C. M., Sarstedt, M., & Völckner, F. (2018). Prediction-Oriented Segmentation in Partial Least Squares Path Modeling. *European Journal of Marketing*, 52(5/6), 1044-1066.
- Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). A Critical Look at the Use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, 36(1), iii-xiv. <https://doi.org/10.2307/41410402>
- Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Unobserved Heterogeneity in Structural Equation Models: A Comparison of FIMIX-PLS with Standard Segmentation Approaches. *International Journal of Research in Marketing*, 28(3), 231-241. <https://doi.org/10.1016/j.ijresmar.2011.04.001>

CAPITULO VII

DE LOS DATOS A LA ESTRATEGIA

La revolución digital ha impulsado una transformación fundamental en la forma en que las organizaciones convierten los datos en estrategia. En esta nueva era, la ciencia de datos se ha consolidado como una disciplina esencial, integrando grandes volúmenes de información con pensamiento crítico, tecnologías abiertas y criterios de transparencia para mejorar la toma de decisiones en todos los niveles del management empresarial (Adler, 2023). En este marco, el analytic empresarial deja atrás el enfoque descriptivo tradicional para adoptar herramientas predictivas-explicativas que permitan navegar entornos complejos y competir en mercados dinámicos.

El modelado de ecuaciones estructurales con mínimos cuadrados parciales (PLS-SEM) representa una de las más poderosas y versátiles metodologías en la analítica avanzada actual. Permite abordar relaciones lineales y no lineales, trabajar con constructos latentes clave como satisfacción o lealtad, y se adapta tanto a grandes volúmenes de datos como a fuentes no tradicionales, superando las limitaciones de la regresión clásica (Demir, 2025; Guenther et al., 2023; Sarstedt et al., 2022). A la vez, cierra la brecha entre el machine learning orientado sólo a la predicción y el

análisis regresivo, aportando valor tanto explicativo como predictivo, y elevando los estándares de reproducibilidad, colaboración y ciencia abierta (Richter et al., 2024; Alshurideh et al., 2023).

En el ámbito actual, el creciente flujo de datos digitales generado por el Internet del Comportamiento (IoB), así como Big Data proveniente de redes sociales, dispositivos IoT y fuentes diversas, plantea nuevos desafíos y oportunidades al modelado causal avanzado. El capítulo explora cómo integrar dichos datos en modelos PLS-SEM robustos mediante buenas prácticas de preprocesamiento desde la limpieza, reducción de dimensionalidad y gestión de datos faltantes, hasta la ponderación para asegurar representatividad muestral (Luengo Martín, 2020; Hair et al., 2017). Ejemplos aplicados y flujogramas ilustran cómo transformar información bruta en resultados accionables.

En esta nueva frontera, el modelo PLS-SEM emerge no sólo como herramienta estadística, sino como un auténtico mapa visual de relaciones causa-efecto, eslabonando hipótesis estratégicas con resultados tangibles y sirviendo como base para el data storytelling de alto impacto ejecutivo (Ringle et al., 2023; Hair et al., 2016). El texto enfatiza también los estándares de rigor y ciencia abierta en la presentación de outputs, la traducción directa de métricas en decisiones estratégicas (diagnóstico, priorización y acción), y la necesidad de dashboards y visualizaciones adaptadas para la gestión avanzada del conocimiento.

Se aborda la responsabilidad ética en el uso del PLS-SEM y la convergencia con Inteligencia Artificial, subrayando el rol del modelado causal en la explicabilidad y justicia de los sistemas automatizados y la relevancia de mitigar sesgos desde la recolección y modelado. El capítulo cierra con una selección curada de artículos científicos de acceso abierto que documentan casos, retos y oportunidades prácticas en la intersección de IA, Big Data, IoT, IoB y modelado PLS-SEM en ciencias empresariales y gestión pública.

Este enfoque integral ofrece un marco actualizado para que investigadores y tomadores de decisión traduzcan la complejidad de los datos en estrategia efectiva, sostenible y ética dentro de la economía digital.

7.1 **Fundamentos de la ciencia de datos y el rol del PLS-SEM en el analytics empresarial**

La Ciencia de datos ha transformado la analítica empresarial al permitir la articulación de datos masivos con pensamiento crítico, tecnología OCD (Open, Collaborative, Data-driven) y criterios de transparencia. El Business Analytics integra la recopilación, procesamiento y análisis de datos para informar y optimizar la toma de decisiones empresariales en todos los niveles (Adler, 2023).

En el contexto actual, el PLS-SEM (Partial Least Squares Structural Equation Modeling) sobresale como herramienta de *Análisis Predictivo-Explicativo*, combinando el poder de la

predicción (común en machine learning) con la capacidad explicativa causal tradicional de modelos de regresión. El PLS-SEM está orientado a predecir constructos latentes clave como satisfacción, lealtad, o valor percibido y permite simultanear complejas relaciones multivariadas en entornos de Big Data empresarial (Guenther et al., 2023; Sarstedt et al., 2022).

La ventaja crítica del PLS-SEM reside en su capacidad para:

- Modelar tanto relaciones lineales como no lineales y latentes, superando el alcance de la regresión clásica que se limita a variables observadas (Demir, 2025).
- Cerrar la brecha entre algoritmos de Machine Learning priorizan la predicción sobre la explicación causal y modelos econométricos tradicionales que explican, pero carecen de potencia predictiva y flexibilidad con datos complejos (Richter et al., 2024; Alshurideh et al., 2023).
- Fortalecer la reproducibilidad y transparencia mediante buenas prácticas de ciencia abierta, permitiendo el desarrollo teórico robusto y replicable para gestión y marketing (Adler, 2023; Sarstedt et al., 2022).
- Articular insights aplicables con tecnologías OCD y pensamiento crítico, donde la interpretación y el análisis colaborativo potencian el valor estratégico de los resultados.

El diseño y el uso del PLS-SEM en la analítica empresarial habilita la formulación de teorías basadas en evidencia y permite que los equipos de dirección operativa, marketing y estrategia fundamenten sus decisiones en datos reales y replicables. De este modo, el PLS-SEM logra que la ciencia de datos sirva tanto para predecir como para entender —optimizando el impacto empresarial de la analítica avanzada.

Ejemplos y Aplicaciones

El Business Analytics se entiende como el uso estratégico y sistemático de los datos empresariales para generar valor, optimizando procesos y fundamentando la toma de decisiones en evidencia (Adler, 2023; Sarstedt et al., 2022). En este marco, el PLS-SEM destaca como herramienta de análisis predictivo-explicativo, cerrando la brecha entre la sofisticación predictiva de machine learning y la explicación causal de la regresión clásica.

- **PLS-SEM en Evaluación de la Satisfacción del Cliente con Big Data Empresarial**

Sarstedt et al. (2022) muestran que, aplicando PLS-SEM a bases de datos masivas de clientes, es posible predecir la “Satisfacción del Cliente” como constructo latente usando variables observables de experiencia, interacción y percepción. El modelo revela tanto los factores

más influyentes (explicación causal) como permite predecir la satisfacción futura tras cambios en estrategias operativas, proporcionando insights accionables para el diseño de servicios.

- **Integración de PLS-SEM con Algoritmos de Machine Learning**

Richter et al. (2024) y Guenther et al. (2023) comparan el uso de PLS-SEM con métodos de machine learning para predecir la intención de compra en marketing digital. Resulta que PLS-SEM no solo iguala la capacidad predictiva de algoritmos avanzados, sino que, además, proporciona explicaciones claras de cómo variables como “confianza” y “valor percibido” impactan realmente la decisión de compra. Esto posibilita estrategias más fundamentadas y transparentes.

- **Open Science y pensamiento crítico en Analytics empresarial**

Adler (2023) documenta cómo las prácticas OCD (Open, Collaborative, Data-driven) y la ciencia abierta en el desarrollo de modelos PLS-SEM facilitan la colaboración entre departamentos y la validación crítica de teorías empresariales, asegurando que los resultados sean

replicables y transparentes. Estos principios permiten auditar los modelos, compartir resultados y mejorar continuamente la calidad de las decisiones.

Relación y cierre metodológico

Los ejemplos expuestos demuestran que el PLS-SEM transforma datos empresariales en conocimiento explicativo y predictivo, permitiendo a empresas y consultores tomar mejores decisiones, diseñar productos orientados al cliente y maximizar el valor de proyectos de análisis con pensamiento crítico.

7.2 Internet del comportamiento y su relevancia para el modelado

El concepto de Internet del Comportamiento (Internet of Behavior, IoB) refiere al proceso de capturar, analizar y utilizar datos generados por las interacciones digitales de las personas con el fin de entender, predecir e influir en su comportamiento a través de plataformas y dispositivos conectados. Se basa en la integración de tecnologías de internet, analítica avanzada, inteligencia artificial y ciencias del comportamiento para identificar patrones de toma de decisiones y preferencias.

En el ámbito del modelado empresarial avanzado, la IoB proporciona flujos constantes de datos no tradicionales como rastros

de navegación, registros de compras, opiniones en redes sociales y comportamientos en plataformas digitales que pueden ser incorporados en modelos PLS-SEM como nuevos indicadores o variables latentes, enriqueciendo la comprensión causal de los factores que impulsa la conducta de clientes y empleados.

La inclusión de datos provenientes del IoB permite refinar la segmentación, diseñar productos y estrategias de comunicación más personalizadas, y anticipar tendencias emergentes en los mercados. Además, facilita la evaluación del impacto de las iniciativas empresariales en tiempo real, al alimentar modelos predictivos-explicativos con fuentes de información masiva y contextualizadas. Este enfoque contribuye a maximizar la capacidad del PLS-SEM para combinar el valor explicativo (relaciones causales) con el predictivo (previsión de eventos o resultados), posicionando la analítica sobre comportamiento digital como un activo estratégico para la toma de decisiones basada en ciencia de datos.

7.3 **Preprocesamiento de Big data para el PLS-SEM**

El auge del Big Data y la proliferación de fuentes no tradicionales, como datos provenientes de redes sociales, sensores IoT o web scraping, plantean importantes desafíos al incorporar estos datos en modelos avanzados como el PLS-SEM (Partial Least

Squares Structural Equation Modeling). La correcta integración y transformación de estos datos es fundamental para producir modelos robustos y válidos (Luengo Martín, 2020; Hair et al., 2017).

Desafíos al integrar datos no tradicionales en PLS-SEM

- **Heterogeneidad y alta dimensionalidad:** Los datos de redes sociales o plataformas web pueden contener miles de variables y formatos diversos, complicando la estructura del modelo.
- **Datos faltantes y ruido:** Las grandes bases de datos, especialmente las recolectadas automáticamente, tienden a ser incompletas, contener errores o estar desbalanceadas respecto a los segmentos de interés.
- **Representatividad y ponderación:** La sobre/infrarepresentación de ciertos grupos puede sesgar los resultados si no se aplican métodos de ponderación adecuados (Luengo Martín, 2020).

Estrategias de preprocesamiento críticas

1. Reducción de dimensionalidad

- Técnicas como el análisis factorial, PCA o el propio uso de PLS-SEM con variables latentes de orden superior ayudan a sintetizar la información y eliminar redundancias, facilitando modelos manejables y evitando la maldición de la dimensionalidad (Ringle et al., 2020).
- Ejemplo: Extraer variables latentes de “sentimiento del cliente” a partir de miles de comentarios en redes sociales usando modelos de medición con PLS-SEM.

2. Gestión de datos faltantes

- Estudios recientes recomiendan superar la simple eliminación de casos (listwise deletion) y la imputación por medias, aplicando métodos más avanzados como la imputación por regresión, EM (Expectation Maximization) o FIML (Full Information Maximum Likelihood), que minimizan sesgos y pérdida de información (Amusa & Hossana, 2019; PMC9336471, 2024).
- Ejemplo: Amusa y Hossana (2019) demuestran que la imputación por EM mejora la precisión de los

coeficientes en PLS-SEM en comparación con métodos tradicionales de imputación.

3. Ponderación de casos para representatividad

- El uso de pesos muestrales permite corregir desbalances de representación, ajustando la contribución de cada caso según su presencia o importancia real en la población objetivo. Esto es especialmente relevante cuando los datos provienen de scraping web o plataformas autoselectivas (Luengo Martín, 2020).
- Ejemplo: Aplicar ponderaciones en una base de datos de reseñas de clientes donde ciertos segmentos (por ejemplo, jóvenes urbanos) están sobrerrepresentados.

Integrar Big Data en modelos PLS-SEM requiere una etapa de preprocesamiento robusta, con estrategias de reducción de dimensionalidad, manejo avanzado de datos faltantes y ponderación, apoyadas en la transparencia metodológica recomendada por la ciencia abierta. Estos pasos aseguran resultados replicables y verdaderamente representativos, facilitando el uso crítico y responsable del Big Data en la investigación y en la práctica empresarial.

7.4 **Flujograma de preprocesamiento de Big Data**

1. Recolección y Consolidación de Datos

- Reúne datos de diversas fuentes (redes sociales, formularios web, sensores, CRM).
- Unifica el formato y estructura los datos en una base común.

2. Limpieza de Datos

- Elimina duplicados y registros irrelevantes.
- Detecta y corrige errores de escritura o codificación.
- Normaliza las variables categóricas y numéricas.

3. Gestión de Datos Faltantes

- Analiza la matriz de valores faltantes por variable y caso.
- Decide la estrategia adecuada (imputación por media, EM, FIML, o eliminación de casos según el patrón de falta y su impacto).

4. Reducción de Dimensionalidad

- Identifica variables redundantes o altamente correlacionadas.
- Aplica métodos como PCA o genera constructos latentes de orden superior mediante PLS-SEM para sintetizar información.

5. Ponderación y Representatividad

- Analiza la representatividad de los segmentos clave.
- Calcula y aplica pesos muestrales cuando sea necesario para ajustar la distribución del dataset.

6. Transformación y Preparación Final

- Escala y normaliza las variables según los requisitos del modelo.
- Codifica las variables categóricas (dummy/one-hot encoding si es necesario para PLS-SEM).
- Divide el dataset en muestras de entrenamiento y validación (validación cruzada si aplica).

7. Ejecución del Modelo PLS-SEM

- Especifica el modelo estructural y de medida.
- Ejecuta el análisis, verifica supuestos y documenta el flujo de preprocesamiento.

Ejemplo aplicado de preprocesamiento de Big Data

A continuación, se presenta un caso tipo paso a paso utilizando un conjunto de datos de comentarios recopilados vía web scraping de reseñas de clientes en una tienda electrónica. Los pasos

incluyen ejemplos de variables y la referencia a la herramienta SmartPLS.

Ejemplo de Variables:

- Texto del comentario (ComentarioCliente)
- Puntaje de satisfacción (SatisfaccionEscala1_5)
- Edad y Género (Edad, Genero)
- Ciudad (CiudadCliente)

Flujograma paso a paso

1. Recolección y Consolidación

- Descarga comentarios y metadatos de la web.
- Consolida todos los comentarios y variables sociodemográficas en una sola base de datos (ejemplo: archivo CSV).

2. Limpieza de Datos

- Elimina duplicados: Si un usuario dejó el mismo comentario más de una vez, se conserva un registro.
- Corrige errores en Edad y Ciudad (valores imposibles o mal escritos).
- Estandariza géneros ("Hombre"/"Mujer" → 1/0).

3. Gestión de Datos Faltantes

- Por ejemplo, si el 3% de los comentarios carece de puntaje de satisfacción, aplica imputación por media o FIML según la MAR (Missing at Random).
- En SmartPLS: cargar la base y escoger la opción de tratamiento de datos faltantes más adecuada.

4. Reducción de Dimensionalidad

- Del texto del comentario, extrae variables latentes usando análisis de sentimientos o agrupación de palabras clave ("CalidadServicio", "EntregaRapida", etc.).
- Usa estos factores para crear constructos latentes de orden superior en SmartPLS (ejemplo: "Percepción Global").

5. Ponderación y Representatividad

- Si la mayoría de los comentarios provienen de una sola ciudad, crea un peso correctivo proporcional para ajustar la representatividad geográfica.
- En SmartPLS, el peso puede cargarse como variable extra e indicarse para el análisis.

6. Transformación y Preparación Final

- Escala el puntaje de satisfacción (por ejemplo: normaliza a rango 0–1).
- Convierte las variables de ciudad en dummies si las vas a usar como moderadoras/segregación en MGA.
- Divide los datos si planeas validación cruzada (entrenamiento/test).

7. Ejecución del Modelo en SmartPLS

- Especifica el modelo de medida con constructos latentes de "Percepción del Cliente" y modelo estructural con "Satisfacción del Cliente".
- Ejecuta el PLS-SEM, valida el modelo y realiza análisis multigrupo (MGA) según variables dummies o peso poblacional.

Este workflow es adaptable y puede implementarse en SmartPLS (para modelado visual), en R (paquete `plspm`) o Python (`pyPLS`) con módulos adicionales para NLP y manejo avanzado de datos faltantes.

7.5 El Modelo PLS-SEM como Herramienta de Data Storytelling

En la era de la analítica empresarial, el modelo PLS-SEM (Partial Least Squares Structural Equation Modeling) no debe comunicarse a la alta dirección únicamente como un diagrama

estadístico, sino como el Mapa de Relaciones Causa-Efecto de la Empresa: una narrativa visual que conecta las hipótesis estratégicas con los resultados clave de negocio (Ringle et al., 2023; Hair et al., 2016). Presentar un diagrama de sendero PLS-SEM convierte la estadística en una historia ejecutiva, donde cada camino (flecha) explica no solo *qué* sucede, sino *por qué* y *cómo* las variables empresariales clave interactúan para producir el éxito o el fracaso.

Consideremos el caso desarrollado por Ringle, Wende y Becker (2022), donde la reputación corporativa se configura como el antecedente de la satisfacción y esta, a su vez, impacta la lealtad del cliente. El diagrama PLS-SEM no busca ser una maraña de coeficientes abstractos, sino una narrativa visual comprensible para ejecutivos, permitiendo observar, por ejemplo, cómo mejorar la reputación corporativa refuerza la satisfacción y, finalmente, la lealtad dando sentido de dirección, prioridades y posibles intervenciones (SmartPLS Primer Book, 2022).

Este enfoque de **Data Storytelling causal** facilita la comprensión transversal en equipos multidisciplinarios y acelera la aceptación y aplicación de los insights. Numerosos estudios destacan que, al traducir el modelo estadístico en historias visuales de causa y efecto, se logra conectar la interpretación ejecutiva con la evidencia cuantitativa, generando confianza en la toma de decisiones basada en datos (Ayanwale et al., 2021; SSBFNET,

2024). De hecho, la narrativa visual es clave para que los tomadores de decisión no técnicos se apropien de la lógica estadística y la integren en los procesos estratégicos, como la priorización de inversiones, la gestión del riesgo o la alineación cultural.

La visualización del PLS-SEM como mapa causal no solo aumenta la transparencia y la credibilidad del análisis, sino que habilita a la organización para contar historias robustas sobre su propia realidad, anticipar escenarios futuros y cerrar el ciclo entre los datos y las decisiones estratégicas.

7.6 **Rigor y replicabilidad: Integración de Output de PLS-SEM**

En el contexto actual de la investigación empírica y analítica, los estándares de transparencia y reproducibilidad son fundamentales para el avance científico y la confianza en los resultados. Para estudios que emplean el modelado avanzado PLS-SEM (Partial Least Squares Structural Equation Modeling), la integración con las buenas prácticas de la ciencia es obligatoria y debe ser verificada en todos los procesos editoriales (Adler, 2023; Guenther et al., 2023).

Estándares de publicación para PLS-SEM:

1. **Compartir archivos de datos:** El manuscrito debe ser acompañado por los archivos de datos brutos y/o transformados, alojados en

repositorios de acceso abierto reconocidos (ejemplo: Zenodo, Figshare, OSF), garantizando el anonimato y la protección de la privacidad, conforme a lo dispuesto en las normativas éticas internacionales (Adler, 2023).

2. **Compartir archivos de código y scripts:** Los archivos del modelo (SmartPLS `.splsm`, R scripts para `lavaan` o `plspm`) deben ser depositados en repositorios públicos, con instrucciones detalladas para su ejecución y personalización, facilitando la replicación exacta del análisis (Guenther et al., 2023; Sarstedt et al., 2022).
3. **Anexo completo de resultados:** El artículo debe incluir (como documentos suplementarios) un reporte integral: matrices de loads y weights, pruebas de validez y fiabilidad (AVE, CR, Alfa de Cronbach), y tablas completas de p-values del bootstrapping y MGA, permitiendo la validación independiente por revisores y futuros investigadores.

El cumplimiento riguroso de estos estándares asegura el potencial de replicabilidad total del estudio, incrementa la credibilidad científica y habilita la reutilización de los resultados en meta-análisis y desarrollos futuros (Adler, 2023; Sarstedt et al., 2022).

7.7 Traducción de hallazgos: De los coeficientes estructurales a las decisiones estratégicas

La fortaleza del modelo PLS-SEM reside en transformar métricas complejas en recomendaciones ejecutivas claras. La siguiente guía de traducción ayuda a convertir resultados estadísticos en acciones estratégicas concretas, avanzando del diagnóstico al despliegue operativo.

Nivel I: Diagnóstico—Interpretación de R^2 y Q^2

- **R^2 (Coeficiente de determinación):** Indica cuánta varianza de la variable dependiente explica el modelo. Cuanto mayor el R^2 , mayor capacidad diagnóstica y explicativa tiene la estrategia; por ejemplo, un $R^2 = 0.45$ sugiere que el 45% de la variabilidad de la lealtad del cliente se explica por los factores modelados.
- **Q^2 (Relevancia predictiva):** Señala si el modelo predice bien fuera de muestra y permite priorizar los constructos con mejor aporte futuro a la organización.

Nivel II: Priorización—Uso de beta como “elasticidad causal”

- El **Coeficiente de Sendero Estandarizado (beta)** se interpreta como la elasticidad causal: por cada punto que

mejora X, se espera una mejora de beta unidades en Y, manteniendo las demás variables constantes.

- Por ejemplo, si $\beta = 0.45$ para el camino "Confianza \rightarrow Lealtad", un aumento de una desviación estándar en confianza genera un incremento de 0.45 desviaciones estándar en lealtad, guiando así la prioridad de intervención.

Nivel III: Acción—Resultados del IPMA y NCA para asignación óptima de recursos

- El **IPMA** (Importance-Performance Map Analysis) identifica áreas donde la importancia (efecto directo en el resultado clave) y el desempeño (nivel actual) están desalineados, señalando oportunidades de mejora o debilidades críticas para direccionar inversiones.
- La **NCA** (Necessary Condition Analysis) determina si un factor es condición necesaria para alcanzar un resultado estratégico. Si se revela que cierto umbral en innovación es indispensable para obtener altos niveles de venta, esa dimensión se convierte en un requisito mínimo no negociable en la asignación de recursos.

Este enfoque por niveles asegura que las métricas PLS-SEM y sus extensiones no solo describan lo que sucede, sino que guíen

acciones concretas: diagnosticar el estado actual, priorizar iniciativas según impacto esperado y definir asignaciones basadas en criterios de importancia y necesidad contingente.

7.8 **Comunicación gerencial: Diseño de dashboards y visualizaciones estratégicas**

El avance en la analítica empresarial requiere que los hallazgos de modelos PLS-SEM trasciendan las tablas estáticas para integrarse en entornos dinámicos de toma de decisiones. El rol del investigador es traducir la complejidad estadística en narrativas visuales que catalicen la acción gerencial. Para esto, la creación de dashboards (tableros de control) interactivos es esencial, permitiendo monitoreo en tiempo real y enfoque selectivo en los aspectos más estratégicos (Guenther et al., 2023; Hair et al., 2017).

Dos visualizaciones destacan por su valor directivo:

1) El Mapa IPMA (Importance-Performance Map Analysis):

- Este gráfico avanzado posiciona constructos o variables según su importancia relativa en el resultado clave (basado en coeficientes de sendero) y su nivel actual de desempeño (promedio o puntuación latente).

- Permite identificar áreas críticas: zonas de alta importancia y bajo desempeño señalan prioridades de inversión y mejora inmediata, mientras que áreas de alta importancia y alto desempeño se consolidan como fortalezas.
- Un dashboard efectivo permite filtrar segmentos, modelar escenarios y mostrar evoluciones temporales, ofreciendo una guía visual directa para la asignación de recursos (Hair et al., 2017; Sarstedt et al., 2022).

2) Diagrama de Sankey o de flujo para cadena de valor causal (X → M → Y):

- El diagrama de Sankey convierte el diagrama de sendero clásico en un mapa de flujos, donde el ancho de cada conexión representa la magnitud del coeficiente causal entre los constructos (Ringle et al., 2023).
- Permite visualizar la cadena de valor desde el input (ej. inversión en capacitación), pasando por mediadores estratégicos (ej. innovación interna), hasta el resultado deseado (ej. rentabilidad).
- Esta visualización ayuda a la gerencia a comprender y comunicar, de forma intuitiva, las rutas más eficaces para alcanzar objetivos corporativos, y a planificar intervenciones secuenciales y contingentes.

La comunicación gerencial exige que los dashboards combinen visualizaciones integradas y adaptables que permitan una lectura proactiva y estratégica de los resultados PLS-SEM, empoderando la toma de decisiones basadas en evidencia visual y contexto de negocio (Guenther et al., 2023).

7.9 **Ética en el modelado causal: Implicaciones de la IA y el uso responsable de PLS-SEM**

El análisis causal con PLS-SEM exige una consideración ética rigurosa tanto en la recolección y manejo de datos como en la interpretación e implicancias de los hallazgos. Los investigadores tienen la obligación de mitigar sesgos en el modelado y la medición, especialmente en constructos sociales sensibles, ya que estos sesgos pueden trasladarse a decisiones empresariales automatizadas con consecuencias significativas para individuos y grupos (Nallalathan et al., 2024; Kock, 2015). La transparencia en los instrumentos y procedimientos analíticos, la documentación detallada de los supuestos y la justificación crítica de los modelos son requisitos fundamentales para preservar la integridad científica y la confianza de los stakeholders (Nallalathan et al., 2024).

Respecto a la integración con la Inteligencia Artificial (IA), el PLS-SEM cumple un rol clave como mecanismo de validación y explicación causal de predicciones generadas por algoritmos de IA.

Mientras muchos sistemas basados en IA logran elevada capacidad predictiva, suelen carecer de explicabilidad y pueden amplificar sesgos presentes en los datos de entrenamiento. El uso de PLS-SEM permite a los analistas desenmarañar y verificar la lógica causal subyacente en las predicciones, aportando evidencia sobre la justicia y equidad de las decisiones automatizadas (Guenther et al., 2023; Nallaluthan et al., 2024).

Ético es también reconocer el impacto social de las recomendaciones derivadas de modelos causales, evitando generalizaciones irresponsables y siendo cautelosos con conclusiones que puedan marginar a minorías o resultar en discriminación indirecta. La adopción de principios de la investigación, la publicación de código y datos, y el reporte completo de supuestos y limitaciones son prácticas imprescindibles para avanzar hacia una investigación realmente responsable y justa (Adler, 2023; Nallaluthan et al., 2024).

7.10 **Artículos sobre sobre las implicaciones de la IA, big data, internet de las cosas (IoT) e internet del comportamiento en PLS-SEM en revistas indexadas.**

1. Jonny, Kriswanto, H., & Matsumura, T. (2021). Building an Implementation Model of IoT and Big Data and Its Improvement. *International Journal of Technology*, 12(4), 700–710. <https://ijtech.eng.ui.ac.id/article/view/5178>

- Desarrolla un modelo de implementación de IoT y Big Data utilizando PLS-SEM para optimizar el rendimiento empresarial, ilustrando relaciones causales y mejoras continuas en el entorno industrial.
2. Zulkarnain, A., Furqani, R., Nugroho, R., & Abdillah, A. (2025). Advancing Management Strategies with AI and IoT for Operational Excellence and Competitive Edge. *Automated Technology Management*, 6(2), 45–59.
<https://ijc.ilearning.co/index.php/ATM/article/download/2396/871/8132>
 - Aplica PLS-SEM para analizar el impacto de IA y IoT en la eficiencia operativa y la ventaja competitiva en empresas tecnológicas, identificando caminos causales para la transformación digital.
 3. Chinnaraju, K. (2025). Partial least squares structural equation modeling (PLS-SEM) in the AI Era: Innovative methodological guide and framework for business research. *Magnascientia Advanced Research & Reviews*, 6(1), 220-232.
<https://magnascientiapub.com/journals/msarr/sites/default/files/MSARR-2025-0048.pdf>
 - Guía metodológica avanzada sobre la integración de PLS-SEM con IA y machine learning para investigación empresarial, abordando implicaciones causales y predictivas.
 4. Ningsih, K., Sari, S. R., & Rahman, A. (2023). Structural equation modeling for IoT and big data implementation in business performance. *AIP Conference Proceedings*, 2485(1), 020007.
<https://pubs.aip.org/aip/acp/article-abstract/2485/1/020007/2905926/Structural-equation-modelling-for-IoT-and-big-data?redirectedFrom=fulltext>

- Propone el uso combinado de SEM/PLS para examinar los efectos de IoT y big data en el desempeño empresarial, subrayando el rol de la innovación gerencial.
5. Musdholifah, F., & Suhendra, W. (2024). Structural equation modeling of internet of things (IoT) adoption for Indonesian Village-Owned Enterprises (BUMDes). *Informasi*, 51(1), 87–98. <https://scholarhub.uny.ac.id/informasi/vol51/iss1/8/>
 - Analiza factores tecnológicos y psicológicos en la adopción de IoT usando PLS-SEM en empresas públicas rurales, ayudando a diseñar intervenciones para la gestión pública.
 6. De la Garza Montemayor, D., & Gómez Díaz de León, M. A. (2024). Inteligencia artificial y big data: nuevos paradigmas de la Comunicación Política y la Gobernanza Digital. *Revista Latina de Comunicación Social*, 82, 456–472. <https://dialnet.unirioja.es/descarga/articulo/9482600.pdf>
 - Explora cómo la IA y big data transforman la gobernanza pública y la comunicación política, integrando hallazgos para mejorar la gestión e interacción ciudadana.
 7. Universitat Oberta de Catalunya. (2018). Datos masivos y datos abiertos para una gobernanza inteligente. *UOC Reports*, 34, 1–34. <https://recerca.uoc.edu/documentos/6560f507d280522600938ee7>
 - Estudio sobre el uso de big data y datos abiertos en la gestión pública, con recomendaciones analíticas para la toma de decisiones participativa y transparente.
 8. Banco Interamericano de Desarrollo (BID). (2024). Uso responsable de la IA para las políticas públicas. *Manuales técnicos del BID*, 100, 1–64. <https://publications.iadb.org/publications/spanish/document/IA-Responsable-Manual-tecnico-Ciclo-de-vida-de-la-inteligencia-artificial.pdf>

- Manual de buenas prácticas éticas en el ciclo de vida de IA y big data para la gestión pública, incluyendo lineamientos de transparencia y explicabilidad.
9. Subiyakto, B., & Ahlan, A. (2018). A PLS-SEM Analysis. *Asian Journal of Information Science*, 7(3), 82–90.
<http://archive.sciendo.com/AJIS/ajis.2018.7.issue-3/ajis-2018-0064/ajis-2018-0064.pdf>
 - Ejemplo aplicado de PLS-SEM para modelar la aceptación tecnológica en organizaciones, útil para la gestión empresarial y pública.
10. SCIRP Open Access. (2025). PLS-SEM - Articles (Colección). *Scientific Research Publishing*.
[https://www.scirp.org/\(S\(vtj3fa45qm1ean45%20vffcz55\)\)/journal/articles?searchCode=PLS-SEM&searchField=keyword&page=1](https://www.scirp.org/(S(vtj3fa45qm1ean45%20vffcz55))/journal/articles?searchCode=PLS-SEM&searchField=keyword&page=1)
 - Colección de artículos revisados sobre la aplicación de PLS-SEM, IA, big data y tecnología en gestión y administración pública.

Referencias

- Abdillah, A., Furqani, R., Nugroho, R., & Zulkarnain, A. (2025). Advancing Management Strategies with AI and IoT for Operational Excellence and Competitive Edge. *Automated Technology Management*, 6(2), 45-59.
- Adler, S. J. (2023). Toward open science in PLS-SEM: Assessing the state of transparency and reproducibility. *Journal of Business Research*, 155, 14-26.
- Alshurideh, M., Salloum, S. A., & Al Kurdi, B. (2023). Predicting actual use of m-learning systems: A comparative approach with PLS-SEM and ML algorithms. *Education and Information Technologies*, 28, 1105-1127.
- Amusa, L. B., & Hossana, T. (2019). An Empirical Comparison of Some Missing Data Treatments in PLS-SEM. *PLOS ONE*, 14(11), e0238995. <https://doi.org/10.1371/journal.pone.0238995>
- Ayanwale, M. A., Shittu, O., & Sanni, O. (2021). A PLS-SEM approach for analyzing learners' satisfaction in online education during COVID-19. *International Journal of Innovation, Creativity and Change*, 15(10), 373-388.
- Banco Interamericano de Desarrollo. (2024). *Uso responsable de la IA para las políticas públicas* (Manuales técnicos del BID No. 100).

Banco Interamericano de Desarrollo.

<https://publications.iadb.org/publications/spanish/document/IA-Responsable-Manual-tecnico-Ciclo-de-vida-de-la-inteligencia-artificial.pdf>

Chinnaraju, K. (2025). Partial least squares structural equation modeling (PLS-SEM) in the AI Era: Innovative methodological guide and framework for business research. *Magnascientia Advanced Research & Reviews*, 6(1), 220-232.

De la Garza Montemayor, D., & Gómez Díaz de León, M. A. (2024). Inteligencia artificial y big data: Nuevos paradigmas de la Comunicación Política y la Gobernanza Digital. *Revista Latina de Comunicación Social*, 82, 456-472.

Demir, S. (2025). Analyzing the Implementation of PLS-SEM in Educational Technology: A Systematic Review. *Journal of Educational Computing*, 54(2), 232-245.

Dolgui, A. (2025). Internet of behaviors: Conceptual model, practical and research applications. *International Journal of Production Research*, 62(2), 123-139. <https://doi.org/10.1080/00207543.2024.2372008>

Guenther, P., Guenther, M., Ringle, C. M., & Sarstedt, M. (2023). Improving PLS-SEM use for business marketing research. *Industrial*

Marketing Management, *111*, 127-142.
<https://doi.org/10.1016/j.indmarman.2023.03.010>

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2.^a ed.). SAGE Publications.

Hair, J. F., Sarstedt, M., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Partial least squares structural equation modeling in business research. *Journal of Business Research*, *69*(10), 3593-3599.
<https://doi.org/10.1016/j.jbusres.2016.08.002>

Jonny, Kriswanto, H., & Matsumura, T. (2021). Building an Implementation Model of IoT and Big Data and Its Improvement. *International Journal of Technology*, *12*(4), 700-710.

Judijanto, L. (2024). Uncovering Global Trends in Internet of Behaviors (IoB): Bibliometric Analysis and Future Directions. *EastSouth Institute Conference Series*, *3*(1), 100-115.

Kock, N. (2015). Common Method Bias in PLS-SEM: A Full Collinearity Assessment Approach. *International Journal of e-Collaboration*, *11*(4), 1-10. <https://doi.org/10.4018/ijec.2015100101>

Luengo Martín, J. (2020). *Big data preprocessing: Enabling smart data*. Springer. <https://digibug.ugr.es/handle/10481/65588>

- Musdholifah, F., & Suhendra, W. (2024). Structural equation modeling of internet of things (IoT) adoption for Indonesian Village-Owned Enterprises (BUMDes). *Informasi*, 51(1), 87-98.
- Nallaluthan, K., Kamaruddin, S., Thurasamy, R., Ghouri, A. M., & Kanapathy, K. (2024). Quantitative Data Analysis using PLS-SEM (SmartPLS): Issues and Challenges in Ethical Consideration. *International Business Education Journal*, 17(2), 41-54.
- Ningsih, K., Sari, S. R., & Rahman, A. (2023). Structural equation modeling for IoT and big data implementation in business performance. *AIP Conference Proceedings*, 2485(1), 020007.
- Richter, N. F., Sarstedt, M., Schlaegel, C., & Gudergan, S. P. (2024). Combining PLS-SEM and selected machine learning algorithms. *Journal of Marketing Analytics*, 12, 201-220.
- Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2020). Partial least squares structural equation modeling in HRM research. *The International Journal of Human Resource Management*, 31(12), 1617-1643. <https://doi.org/10.1080/09585192.2017.1416653>
- Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2023). A perspective on using partial least squares structural equation modelling (PLS-SEM). *Data in Brief*, 49, 109192. <https://doi.org/10.1016/j.dib.2023.109192>

- Ringle, C. M., Wende, S., & Becker, J.-M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)*. SmartPLS Academy. https://www.smartpls.com/primer-book-case-studies/primer_3e_chap2_case_new.pdf
- Sadeqi-Arani, Z. (2025). Benefits and Application of IoB in Educational Businesses. *Journal of Education and Health Promotion*, 14, 25.
- Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2022). Progress in partial least squares structural equation modeling in marketing analytics. *Journal of Marketing Analytics*, 10(4), 341-357.
- Subiyakto, B., & Ahlan, A. (2018). A PLS-SEM Analysis. *Asian Journal of Information Science*, 7(3), 82-90.
- Surachman, S., Wijaya, E., & Rahayu, S. (2024). Examining the efficacy of the internet of behavior (IoB) in the identification of client needs on electronic commerce platforms. *Journal of Theoretical and Applied Information Technology*, 102(8), 3459-3470.
- Universitat Oberta de Catalunya. (2018). *Datos masivos y datos abiertos para una gobernanza inteligente* (UOC Reports No. 34). Universitat Oberta de Catalunya. <https://recerca.uoc.edu/documentos/6560f507d280522600938ee7>
- Yang, C. L., Chen, X., & Wang, J. (2021). The use of social media data mining and PLS-SEM to analyze psychological adaptation to air

pollution. *International Journal of Environmental Research and Public Health*, 18(10), 5270. <https://doi.org/10.3390/ijerph18105270>


EPÍLOGO

El fin de la inocencia estadística: hacia una ciencia de datos responsable

Al cerrar este volumen, es probable que la forma en que usted concibe la investigación empresarial haya cambiado para siempre. Hemos recorrido un largo camino desde las certezas cómodas de los modelos lineales y las muestras homogéneas, adentrándonos en el terreno mucho más exigente pero infinitamente más rico de la no linealidad y la heterogeneidad no observada.

Este libro no se escribió para ofrecer recetas fáciles, sino para otorgar poder. El poder de detectar lo que otros ignoran. Ahora, cuando se enfrenta a un conjunto de datos que parece no tener sentido, sabrá que quizás no es "ruido", sino una estructura latente esperando ser descubierta mediante FIMIX-PLS o PLS-POS. Cuando una teoría establecida no se cumpla, tendrá la audacia técnica para preguntar si la relación es cuadrática o si el constructo fue mal especificado, aplicando el Análisis Tetrádico Confirmatorio (CTA-PLS) para dejar que los datos hablen.

Sin embargo, con este poder analítico viene una responsabilidad ineludible. Como discutimos en los capítulos finales, la convergencia entre PLS-SEM, Big Data e Inteligencia Artificial nos sitúa en una nueva frontera. Ya no somos simples estadísticos; somos arquitectos de algoritmos que pueden influir en decisiones de política pública, estrategias



corporativas y vidas humanas. La capacidad de predecir el comportamiento con alta precisión no nos exime de la obligación ética de comprender el "porqué" detrás de la predicción.

La investigación de negocios del futuro no pertenecerá a quien acumule más datos, sino a quien sepa modelar mejor su complejidad. Usted lleva ahora en su caja de herramientas el rigor para no conformarse con el "promedio" y la visión para buscar la verdad en los matices.

La estadística tradicional ha muerto. ¡Larga vida a la ciencia de la predicción!

Manuel Anchapuri

ACRÓNIMOS

Acrónimo	Significado
SEM	Modelado de Ecuaciones Estructurales (Structural Equation Modeling).
PLS-SEM	Modelado de Ecuaciones Estructurales por Mínimos Cuadrados Parciales (Partial Least Squares Structural Equation Modeling).
AVE	Varianza Extraída Promedio.
CFA	Análisis Factorial Confirmatorio (Confirmatory Factor Analysis).
CB-SEM	Modelado de Ecuaciones Estructurales Basado en Covarianzas (Covariance-Based Structural Equation Modeling).
R ²	Coefficiente de Determinación.
f ²	Tamaño del efecto (Effect Size).
VIF	Factor de Inflación de la Varianza (Variance Inflation Factor).
CMB	Sesgo del Método Común (Common Method Bias).
IPMA	Análisis de Importancia–Desempeño (Importance–Performance Map Analysis).
NCA	Análisis de Condición Necesaria (Necessary Condition Analysis).
PLSpredict	Evaluación del poder predictivo fuera de la muestra en PLS-SEM.

GLOSARIO

Término	Definición
Modelado de Ecuaciones Estructurales (SEM)	Marco metodológico multivariante que permite analizar de manera simultánea relaciones entre variables observadas y constructos latentes, integrando modelos de medida y modelos estructurales.
PLS-SEM (Partial Least Squares Structural Equation Modeling)	Enfoque de ecuaciones estructurales basado en la varianza, orientado a la explicación y predicción de fenómenos complejos, especialmente adecuado para contextos con alta complejidad teórica y heterogeneidad.
Constructo latente	Variable teórica no observable directamente, inferida a partir de indicadores empíricos que representan dimensiones conceptuales del fenómeno estudiado.
Modelo de medida	Componente del SEM que especifica la relación entre los constructos latentes y sus indicadores, determinando cómo se operacionalizan los conceptos teóricos.
Modelo estructural	Componente del SEM que representa las relaciones causales hipotetizadas entre los constructos latentes, permitiendo evaluar efectos directos, indirectos y totales.
No linealidad	Característica de las relaciones teóricas en las que los cambios en una variable no generan efectos proporcionales en otra, reflejando comportamientos asimétricos o dependientes de umbrales.
Heterogeneidad no observada	Presencia de subpoblaciones latentes dentro de la muestra que difieren en sus patrones estructurales, y que, de no ser identificadas, pueden distorsionar las inferencias teóricas.

Poder predictivo	Capacidad de un modelo teórico para anticipar valores futuros o no observados de los constructos dependientes, más allá de la significancia estadística de los coeficientes.
Relaciones asimétricas	Vínculos teóricos en los que el efecto de una variable sobre otra difiere según el nivel o la dirección del cambio, desafiando los supuestos de simetría clásica.
Umbrales teóricos	Puntos críticos a partir de los cuales una variable comienza a generar efectos sustantivos sobre otra, relevantes para la interpretación estratégica de los modelos.
Varianza explicada	Proporción de la variabilidad de un constructo endógeno que es atribuible a sus predictores dentro del modelo estructural.
Inferencia teórica	Proceso mediante el cual se derivan conclusiones conceptuales a partir de los resultados del modelo, evaluando la coherencia entre teoría, hipótesis y evidencia empírica.
Complejidad estructural	Nivel de interconexión y multiplicidad de relaciones entre constructos dentro de un modelo, asociado a fenómenos organizacionales dinámicos y multifactoriales.
Validez sustantiva	Grado en que los resultados del modelo aportan significado teórico y utilidad interpretativa para la comprensión del fenómeno investigado.
Investigación predictiva	Enfoque de investigación orientado a evaluar la capacidad de los modelos teóricos para anticipar comportamientos y resultados relevantes en contextos reales.

Autores



Fredy Quispe Gómez.

Doctor en Administración: Magíster Scientiae en Contabilidad y Administración con mención en Marketing y Negocios Internacionales ;Estudios de Maestría en Investigación y Docencia Universitaria ,Universidad Nacional del Altiplano y Licenciado en Administración de Empresas Docente Universitario Universidad Nacional del Altiplano Facultad de Ciencias Administrativas y Humanas;Ex Delegado ante el consejo Directivo nacional del colegio de licenciados en administración -CLAD Fundador y presidente del Directorio Cámara internacional de empresarios y emprendedores de Latinoamérica en capacitación ,asesoría y consultoría empresarial.

<https://orcid.org/0000-0001-6988-4666>

fquispeg@unap.edu.pe



Rossana Esmila Cortez López

Licenciada en Administración, experiencia en el sector Público y Privado; estudios de especialización en Gestión Pública, Marketing, Negocios Internacionales.

<https://orcid.org/0009-0002-5191-3486>

rosxsanna@gmail.com



Sandra Ines Ponce Umiña

Candidata a Magíster en Contabilidad y Administración con mención en Administración y Finanzas. Docente de la Universidad Nacional del Altiplano - Puno. Cuenta con experiencia en gestión pública, desarrollo económico local, desarrollo humano y participación ciudadana. Asimismo, posee trayectoria especializada en la administración de registros de transporte y gerencia de terminales terrestres, desempeñándose tanto en instituciones públicas como privadas

<https://orcid.org/0009-0007-6296-4509>

sainpoum@gmail.com



Alberto Magno Cutipa Limache

Doctor en Administración (UNA). Magíster en Administración (UJCM) y Licenciado en Administración (UNA). Docente de pre y posgrado en la Facultad de Ciencias Contables y Administrativas de la UNA. Exbecario del Programa de Intercambio Educativo de la Universidad del Pacífico. Ganador del Premio a la Publicación Científica UNA 2020. Investigador principal de proyectos, como una aplicación móvil para artesanas. Futuras líneas: Marketing e innovación, utilizando SmartPLS y Amos

<https://orcid.org/0000-0001-8584-6424>

acutipa@unap.edu.pe



Manuel Anchapuri Q.

Licenciado en Administración magíster en Contabilidad y Administración, con especialización en Administración y Finanzas, y Doctor en Administración y Contabilidad con Especialización en Big Data Analytics Aplicada a los Negocios, y Gobernabilidad, Gerencia Política y Gestión Pública, así como en Gobernabilidad e Innovación Pública.

<https://orcid.org/0000-0002-0339-2631>

manchapuri@unap.edu.pe

¿CONFIRMACIÓN O PREDICCIÓN?

Asumir linealidad y homogeneidad en mercados complejos ya no es una abstracción aceptable; es un error metodológico.

"Modelado Avanzado con PLS-SEM" redefine el estándar de la investigación cuantitativa, guiando el tránsito desde el paradigma explicativo (covarianza) hacia el predictivo (varianza).

Una obra esencial para dominar:

- **Heterogeneidad Latente:** Descubra segmentos ocultos teóricamente relevantes.
- **Dinámicas No Lineales:** Modele la realidad de los rendimientos decrecientes.
- **Rigor Epistémico:** Valide la naturaleza de sus constructos mediante el Análisis Tetrádico (CTA).

Deje de ajustar modelos al pasado. Convierta la complejidad teórica en su mayor ventaja científica.

La complejidad es su nueva ventaja competitiva



ISBN: 978-612-03-1884-3



9 786120 318843