



Ecosistema de investigación de CIEM IA+

Compendio de Fundamentos: Análisis Exploratorio de Datos (AED) y Técnicas Emergentes (Fase 2)

Autor: Estadístico León Darío Bello Parías

Dirección científica: CIEM IA+

Enfoque: Aprender Haciendo con Datos e Inteligencia

Contacto: <https://ciemonline.com/ecosistema-ciem-ia/>

1. Introducción: La Inteligencia de Datos en el Mercadeo

El Análisis Exploratorio de Datos (AED) no es un simple trámite técnico; es el cimiento estratégico sobre el cual se edifica cualquier campaña de marketing de alto impacto. En mi experiencia como consultor senior, he visto cómo la inversión en pauta se convierte en una apuesta ciega cuando carece de un diagnóstico inicial riguroso. Sin una estructura de datos sólida, cualquier intento de optimización es, en el mejor de los casos, una conjetura afortunada.

En el ecosistema de CRM y bases de datos de clientes, la "**Higiene del Dato**" es el determinante real del Retorno de Inversión (ROI). Una base de datos depurada no solo mejora la precisión de la segmentación, sino que protege la reputación de la marca al evitar comunicaciones erróneas. En el ecosistema **CIEM IA+**, operamos bajo el principio de "Seguridad Honesta": garantizamos que los hallazgos sean estadísticamente significativos y éticamente fiables. Esta transparencia metodológica es lo que otorga credibilidad final a los resultados presentados ante una junta directiva. Para alcanzar este nivel de maestría, es imperativo transitar por protocolos técnicos que aseguren la calidad desde la raíz.

Ecosistema Estadístico CIEM IA+

Director Estadístico León Darío Bello Parías – WhatsApp + 57 316 575 9247

2. Núcleo Técnico: El Protocolo AED PRO (Fase 2)

Antes de proceder a cualquier modelado de datos, debemos exigir una validación técnica absoluta. Como siempre reitero en mis sesiones de formación: **"Detectar un faltante a tiempo es evitar una conclusión equivocada después"**. Si el insumo está contaminado, la estrategia resultante será defectuosa.

Validación de Supuestos y el Mito del "Promedio"

En el análisis de mercado, ignorar supuestos como la normalidad o la independencia invalida cualquier prueba posterior. Tomemos como referencia un caso real de nuestra academia: una variable "Edad" con un rango de 0 a 150 años. Al procesar los datos, observamos una **Media de 30.67** y una **Mediana de 29**. Sin embargo, el indicador crítico es la **Curtosis de 31.25**.

Desde una visión pragmática de negocios, esta Curtosis tan elevada nos revela una distribución extremadamente apuntada (leptocúrtica) con valores atípicos severos. Esto nos dice que el "consumidor promedio" de 30 años es un mito estadístico en este dataset; la realidad del mercado está concentrada en picos específicos o distorsionada por errores de registro que deben ser saneados para no sesgar la inversión.

Técnicas Emergentes y Agentes de IA

Hoy, los Agentes de IA permiten una validación que trasciende las limitaciones de las hojas de cálculo simples. Mientras Excel nos limita a una visión bidimensional, la IA puede realizar un reconocimiento de patrones multidimensionales, detectando anomalías en el comportamiento de compra que el ojo humano omitiría. Esta potencia analítica es la que nos permite transitar del ruido a la señal.

3. Fiabilidad Inter-Jueces y Validación de Raters

Cuando la investigación de mercados depende de criterios humanos —como clasificar la temperatura de un *lead* o el análisis de sentimiento en redes sociales— la consistencia es obligatoria. No podemos permitir que el éxito de una estrategia dependa de la subjetividad de un evaluador.

Análisis Comparativo de Coeficientes (Contexto JASP)

La elección del coeficiente debe ser quirúrgica, dependiendo de la escala y el número de jueces:

Coeficiente	Evaluable	Escala	Observación de Valor Estratégico
Cohen's Kappa	2	Nominal / Ordinal	El estándar para parejas de jueces.
Fleiss' Kappa	2 o más	Nominal / Ordinal	Crucial para equipos de ventas o soporte amplios.
Krippendorff's Alpha	2 o más	Nominal / Ordinal	El más versátil; maneja datos faltantes con robustez.

Nota Técnica sobre el Desacuerdo: En escalas ordinales (ej. calificación de servicio), es vital utilizar el **Weighted Kappa** (Kappa Ponderado). En marketing, un desacuerdo entre "Interesado" y "Muy Interesado" es menos grave que entre "Interesado" y "No Interesado". El coeficiente ponderado reconoce estos grados de discrepancia, proporcionando una métrica de fiabilidad mucho más realista para el negocio.

4. Metodología de "Higiene del Dato": Excel vs. SPSS

Mi lema es claro: **"No se trata de usar un software, sino de saber cuándo usar cada uno"**. Excel es nuestra mesa de disección inicial; SPSS es nuestro laboratorio de alta precisión.

Protocolo de Limpieza en Excel

La metodología CIEM establece dos rutas de acción para la normalización:

1. **Opción 1 (Ordenamiento Manual):** Organizar la base por la variable objetivo para identificar visualmente agrupaciones de errores.
2. **Opción 2 (Normalización Técnica):** Utilizar **Ctrl+L (Buscar/Reemplazar)** para unificar etiquetas (ej. convertir "Fem", "femenino" y "f" en el estándar "F").
3. **Diagnóstico Cuantitativo:** Activar el complemento "Análisis de datos" para obtener el resumen descriptivo que revele la salud de la variable.

Superioridad de SPSS y la Bitácora de Cambios

SPSS supera a Excel en la gestión de grandes volúmenes y en la **imputación de datos**. Un mandato innegociable en este proceso es **documentar cada cambio en la bitácora**. Esta trazabilidad es lo que permite auditar la limpieza y garantiza que la transformación de los datos sea reversible y justificada.

El Dilema de la Imputación Estratégica

Ante valores faltantes (NA), el consultor debe decidir:

- **Imputar:** Si el faltante es marginal y el sesgo es mínimo.
- **Eliminar:** Si la integridad de la variable está comprometida.
- **Dejar explícito el NA:** Estrategia clave en segmentación. Por ejemplo, un cliente que deja el campo "Ingresos" en blanco puede pertenecer a un segmento con alta conciencia de privacidad, convirtiendo el "dato faltante" en un indicador conductual por sí mismo.

5. Aplicación Práctica: Caso de Estudio en Comportamiento del Consumidor

La verdadera maestría consiste en traducir una tabla de frecuencias en una decisión de presupuesto. Supongamos que analizamos el "Gasto Mensual" de una base de clientes.

Si el análisis descriptivo nos arroja una **Curtosis de 31.25**, como analista senior, mi recomendación a la junta no sería "seguir el promedio". Esa curtosis me indica que hay un grupo pequeño de clientes con un gasto masivo (outliers) que están inflando la media. Si invertimos basándonos solo en el promedio de 30.67, estaríamos desperdiciando pauta en un segmento inexistente.

La decisión estratégica correcta sería pivotar: aislar a esos "Heavy Users" identificados por la curtosis y diseñar una campaña de lealtad exclusiva, mientras que para la gran masa (representada por la mediana de 29), se mantiene una estrategia de volumen. Este es el poder de convertir un histograma en un plan de acción rentable.

6. Referencias Bibliográficas (Estándar CIEM IA+)

- Bello Parias, L. D. (2025). *Módulo 2: Limpieza de datos y estandarización para garantizar calidad*. Academia CIEM.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*.
- Hair, J. F., & Anderson, R. E. (2025). *Multivariate Data Analysis: A Global Perspective* (9th ed.).
- JASP Team (2024). *JASP Help Guide: Validation and Rater Agreement* (Version 0.19).
- Krippendorff, K. (1970). Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement*