





Programa de especialización

DATA DRIVEN DECISIONS 2026

Nivel Intermedio

SESIÓN 9

TRANSFORMAR AHORA,
trascender siempre

Sesión 9

Variabilidad, patrones y señales en los datos

Logro de aprendizaje

Al finalizar la sesión, lograrás:

- Interpretar la variabilidad utilizando medidas de dispersión básicas.
- Comparar grupos que pueden tener promedios similares pero comportamientos diferentes.
- Identificar tendencias, estabilidad, anomalías y otras señales relevantes en los datos.
- Utilizar Gemini para explicar resultados y formular hipótesis iniciales sin confundirlas con conclusiones.
- Realizar una exploración de tendencias y una proyección simple a partir de datos históricos.

Contenido:

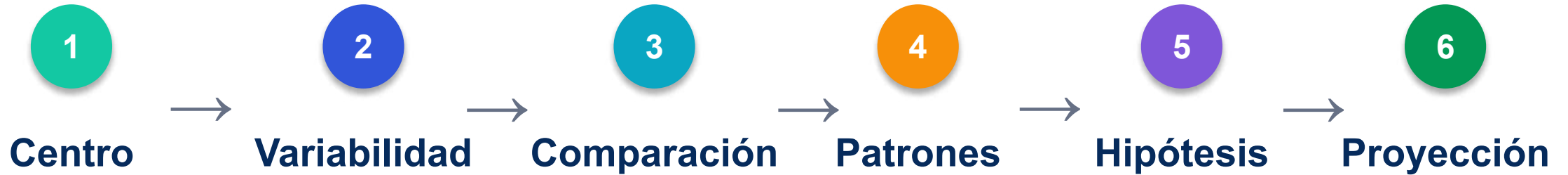
- | 1. Medidas de dispersión: introducción
- | 2. Medidas de dispersión: rango
- | 3. Medidas de dispersión/posición: cuartiles
- | 4. Medidas de dispersión: varianza y desviación estándar
- | 5. Anomalías
- | 6. Patrones
- | 7. Power BI
- | 8. Actividad en Power BI
- | 9. Gemini

¿Dónde estamos dentro del proceso?

Hasta ahora hemos aprendido a preparar y estructurar datos. La siguiente pregunta es qué debe ocurrir para que ese trabajo pueda ser utilizado por otras personas.



El flujo de hoy



Hoy pasamos de “¿cuál es el promedio?” a “¿qué tan consistente es ese promedio y qué señales aparecen detrás?”

Activación: dos programas, casi el mismo promedio

Ingeniería Informática

14,54

Promedio académico

Marketing

14,57

Promedio académico

¿Podemos afirmar que ambos programas se comportan de manera similar?

Todavía no tenemos suficiente información

Agregamos la mediana

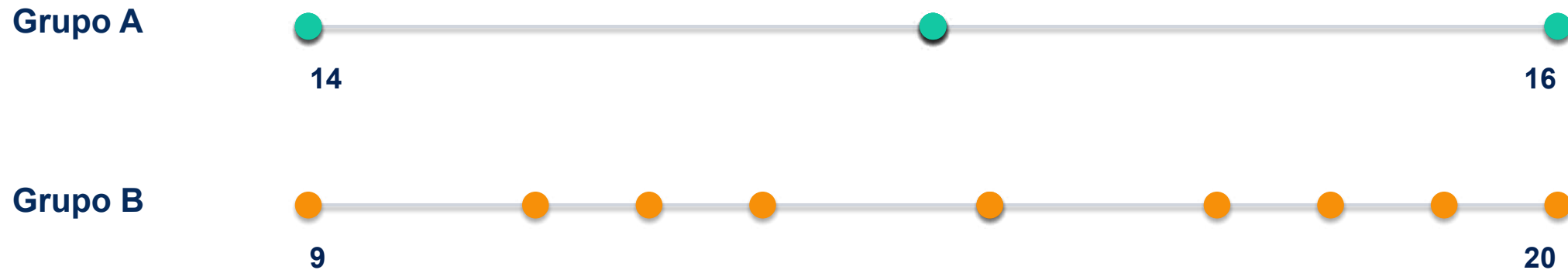
Programa	Promedio	Mediana
Ingeniería Informática	14,54	14,60
Marketing	14,57	13,90

Marketing ya muestra una señal: su centro no se comporta igual que el de Informática.

Pero aún falta una pregunta: ¿qué tan dispersos están los datos alrededor de ese centro?

El centro no basta

Dos grupos pueden tener el mismo promedio y niveles de estabilidad completamente diferentes.



El promedio no revela la amplitud de los datos

1. Medidas de dispersión: introducción

Variabilidad

¿Qué significa que los datos sean más o menos variables?

Variabilidad: cuánto se alejan los datos entre sí

La variabilidad describe cuánto cambian o se dispersan las observaciones dentro de un conjunto de datos.

Conocer solo el centro responde:

“¿Dónde están los datos?”

La dispersión agrega:

“¿Qué tan separados están?”

Centro + dispersión = una descripción mucho más completa del comportamiento de los datos.

¿Por qué importa la variabilidad?

1

Consistencia

¿Los resultados son estables o cambian mucho?

2

Riesgo

¿Hay valores que podrían generar resultados inesperados?

3

Comparación

¿Dos grupos con medias parecidas realmente se comportan igual?

4

Decisión

¿Podemos planificar usando un valor promedio con suficiente confianza?

Ejemplo: mismo promedio, distinta previsibilidad

Unidad A

Tiempos de atención: 4, 5, 5, 5, 6

Promedio = 5 días

Comportamiento relativamente estable

Unidad B

Tiempos de atención: 1, 2, 5, 8, 9

Promedio = 5 días

Comportamiento mucho más variable

El mismo promedio no implica el mismo nivel de predictibilidad.

2. Medidas de dispersión: rango

Rango: una primera mirada a la amplitud

¿Qué responde?

¿Qué tan separados están el valor mínimo y el máximo?

$$\text{Rango} = \text{Máximo} - \text{Mínimo}$$

Ejemplo

Mínimo = 10 | Máximo = 18

$$\text{Rango} = 8$$

Ventaja: fácil de interpretar. Limitación: solo utiliza los dos extremos.

Rango en nuestro dataset

Ingeniería Informática

5,2

Máx. 16,7 – Mín. 11,5

Marketing

8,5

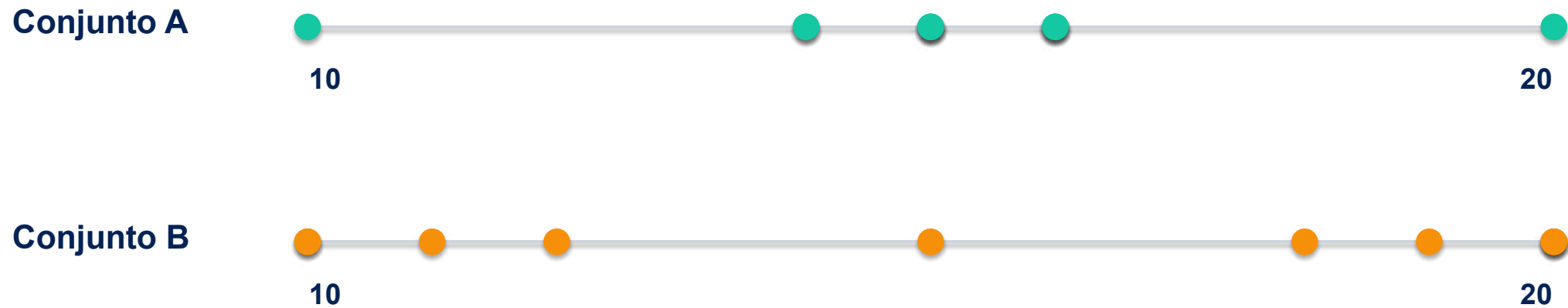
Máx. 19,7 – Mín. 11,2

Marketing cubre una amplitud mucho mayor, aunque su promedio sea prácticamente igual.

Primera evidencia de mayor variabilidad

El rango puede ser insuficiente

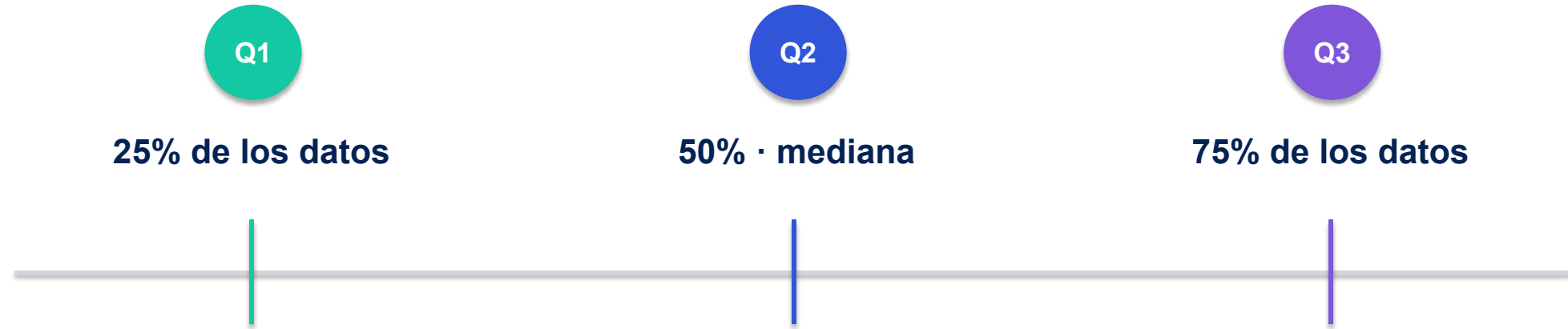
Dos conjuntos pueden compartir el mismo mínimo y máximo y distribuirse de forma muy diferente.



Necesitamos medidas que utilicen más información del conjunto.

3. Medidas de dispersión/posición: cuartiles

Cuartiles: dividir los datos ordenados en cuatro partes



Los cuartiles permiten mirar el centro de la distribución sin depender únicamente de los valores extremos.

Rango intercuartílico (IQR)

Mide la amplitud del 50% central de los datos.

$$\text{IQR} = Q3 - Q1$$



Es especialmente útil cuando existen valores extremos o distribuciones sesgadas.

4. Medidas de dispersión: varianza y desviación estándar

Varianza: la idea detrás de la dispersión

La varianza resume cuánto se separan los valores del promedio, utilizando las distancias al cuadrado.

Es fundamental estadísticamente, pero su unidad queda al cuadrado y es menos intuitiva para comunicar.

- 1 Calcula la distancia de cada valor respecto a la media.
- 2 Eleva esas distancias al cuadrado.
- 3 Promedia las distancias cuadradas.

Desviación estándar: dispersión en la misma unidad

La desviación estándar es la raíz cuadrada de la varianza.

$$\text{Desv. estándar} = \sqrt{\text{Varianza}}$$

Una desviación estándar pequeña



Datos más concentrados alrededor de la media.

Una desviación estándar grande



Datos más dispersos.

Se interpreta en la misma unidad que la variable

Informática vs. Marketing: ahora aparece la diferencia

Programa	Promedio	Mediana	Rango	Desv. estándar
Ingeniería Informática	14,54	14,60	5,2	0,88
Marketing	14,57	13,90	8,5	1,85

Marketing tiene una desviación estándar más de dos veces mayor.

El promedio era casi idéntico. La variabilidad revela una realidad distinta.

¿Una desviación estándar alta siempre es mala?

No. La dispersión es una característica del comportamiento de los datos, no un juicio de valor.

Puede ser problemática

si buscamos consistencia en tiempos de atención.

Puede ser esperable

si analizamos diversidad de perfiles o ingresos.

Puede ser una señal

si cambia abruptamente entre periodos.

La interpretación depende del contexto de la decisión

5. Anomalías

Anomalías y señales

No todo valor diferente es un error

Valor atípico, error y anomalía no son lo mismo

1

Error

Dato incorrecto por digitación, formato, duplicación o captura.

2

Valor atípico

Observación válida pero alejada del comportamiento más común.

3

Anomalía

Comportamiento inesperado respecto a un patrón o referencia.

Primero valida el dato. Después interpreta por qué es diferente.

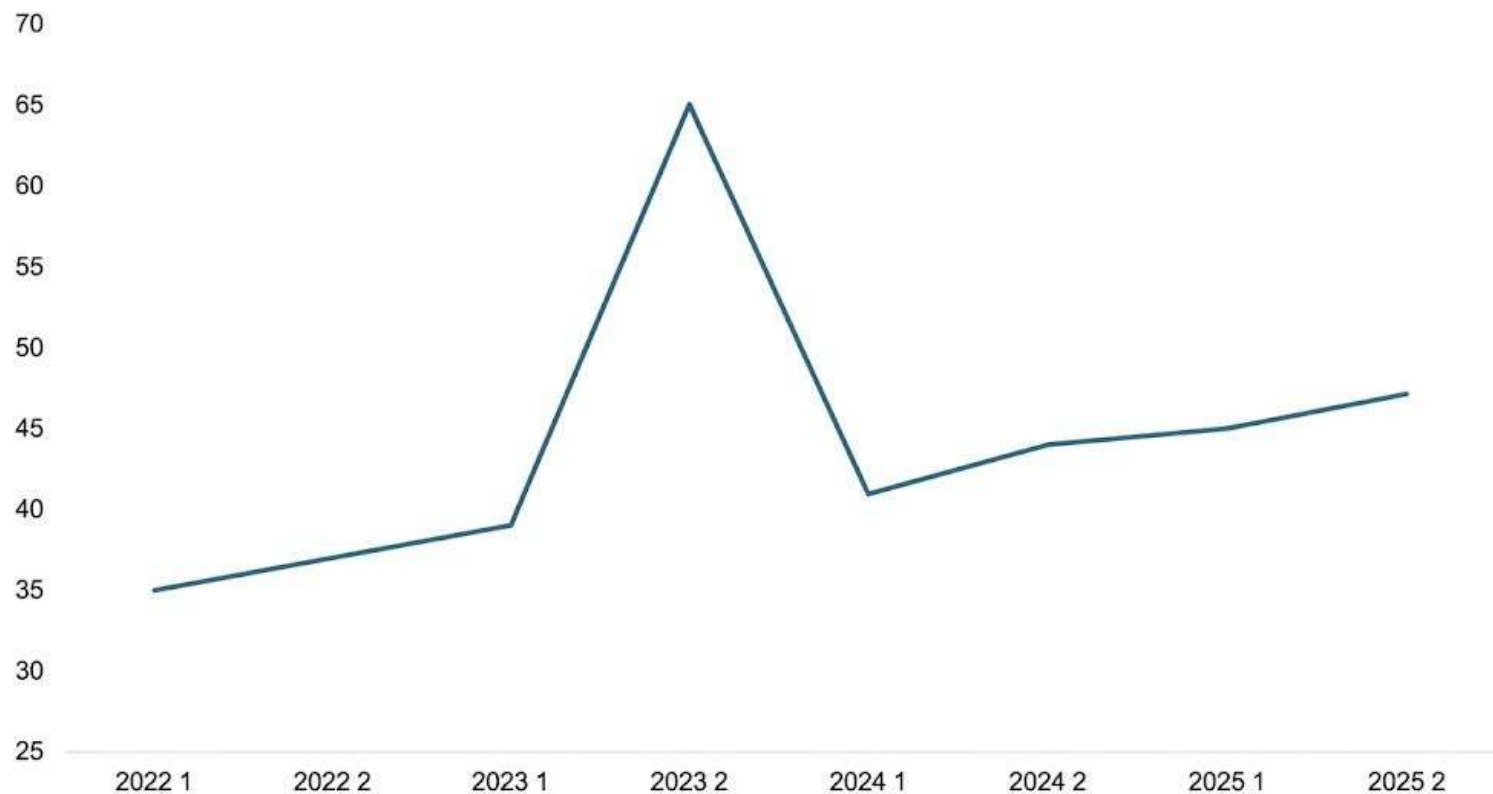
Una anomalía puede ser la información más importante

Eliminar automáticamente lo “raro” puede borrar precisamente la señal que necesitamos investigar.

Pregunta correcta: “¿este valor es incorrecto o representa un evento real?”

Calidad de datos antes que interpretación

Caso: el pico de Marketing



2023-2 registra 65 matrículas frente a 39 en el periodo anterior.

¿Qué podemos afirmar y qué debemos investigar?

Lo que los datos sí muestran

- 1 Hubo un aumento abrupto en 2023-2.
- 2 El valor no sigue el comportamiento de los periodos vecinos.
- 3 El periodo 2023-2 fue identificado estadísticamente como dato atípico al superar el límite del Rango Inter cuartílico (IQR).

Lo que requiere una hipótesis

- 1 ¿Hubo una campaña de admisión?
- 2 ¿Cambió la oferta o la modalidad?
- 3 ¿Existe un cambio en la fuente de datos?

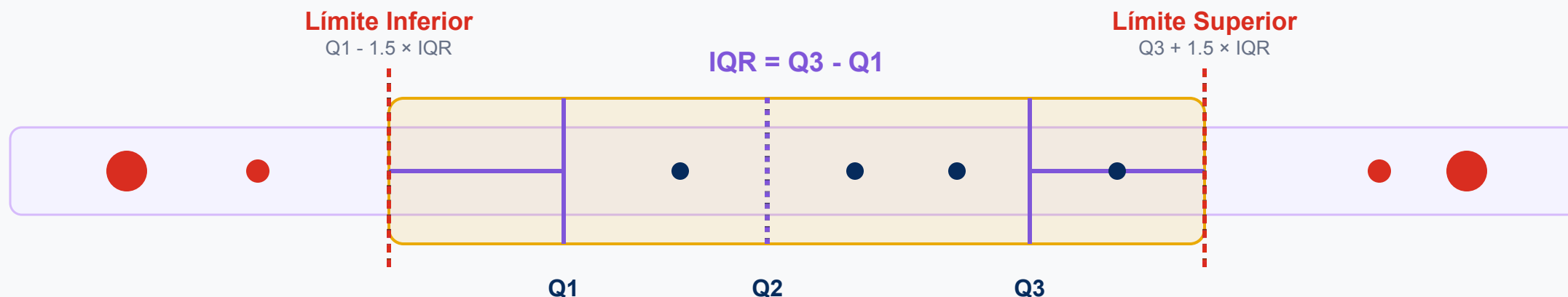
Una hipótesis orienta la investigación. No reemplaza la evidencia.

¿Cómo se identifican los datos atípicos con el IQR?

1. Calcular Cuartiles y IQR

2. Definir Límites

3. Detección



Cualquier valor ubicado por **debajo del Límite Inferior** o por **encima del Límite Superior** se clasifica estadísticamente como un **dato atípico (outlier)**.

El Rango Intercuartílico evalúa el 50% central de los datos para aislar valores extremos.

6. Patrones

Patrones

Pasar de observar puntos aislados a reconocer comportamientos

¿Qué señales podemos buscar en una serie de datos?

1

Crecimiento

Aumento sostenido a través del tiempo.

2

Descenso

Reducción progresiva.

3

Estabilidad

Variaciones pequeñas alrededor de un nivel.

4

Meseta

Crecimiento inicial seguido de estabilización.

5

Pico / caída

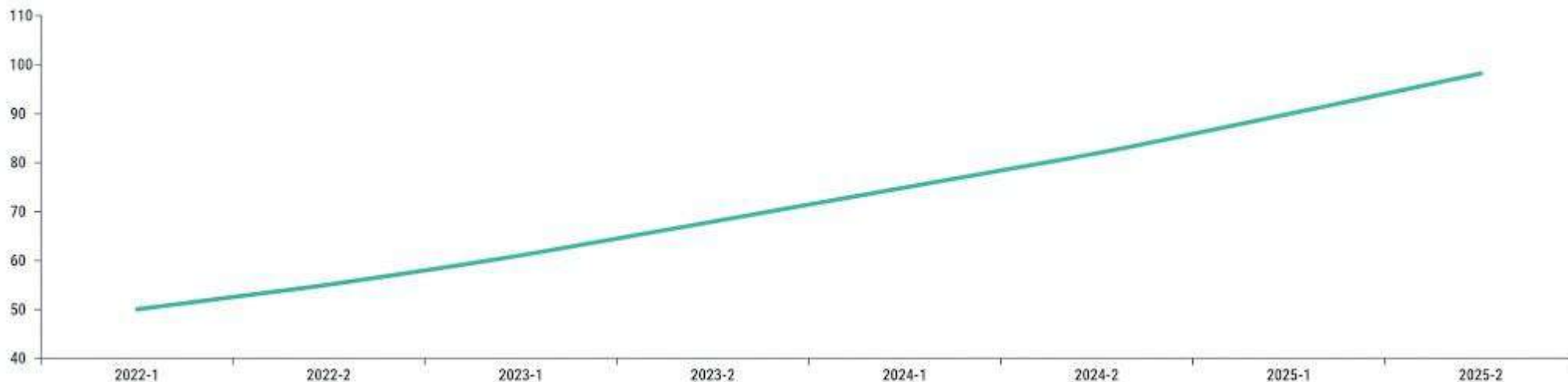
Cambio puntual que rompe el patrón.

Cinco programas, cinco comportamientos

Programa	2022-1	2023-2	2025-2	Señal principal
Ing. Informática	50	68	98	Crecimiento sostenido
Ing. Civil	72	60	47	Descenso gradual
Administración	45	61	69	Crecimiento → meseta
Marketing	35	65	47	Pico atípico
Psicología	58	61	63	Estabilidad

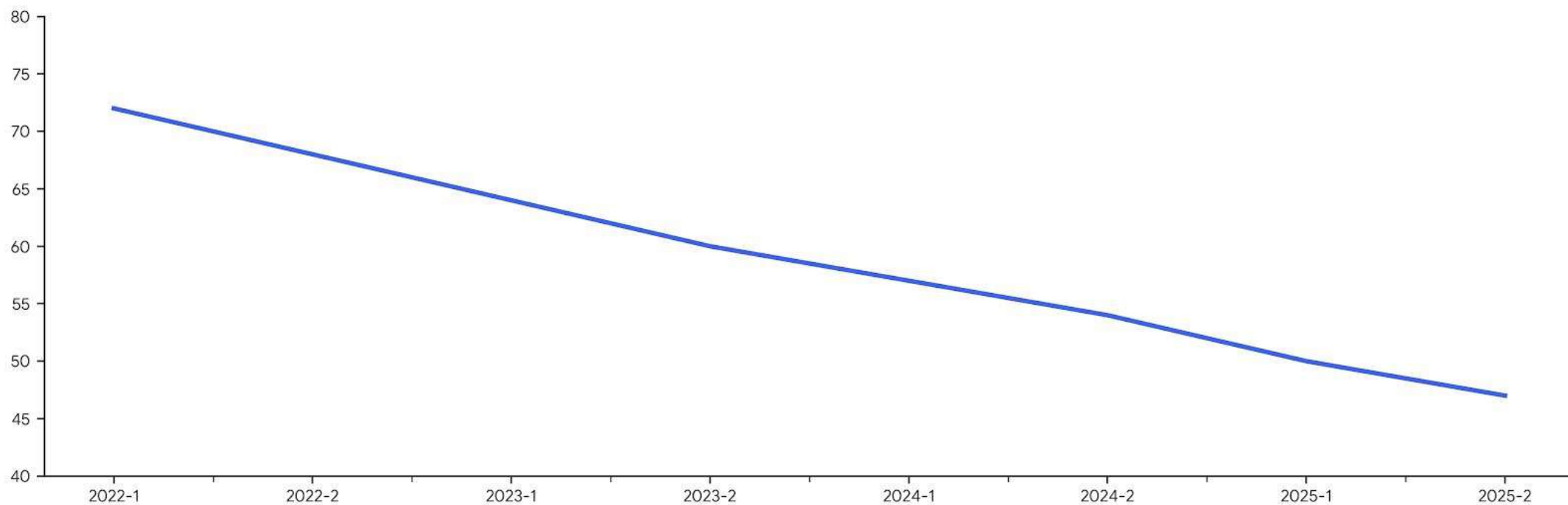
La misma institución puede contener patrones muy distintos

Patrón 1: Crecimiento sostenido



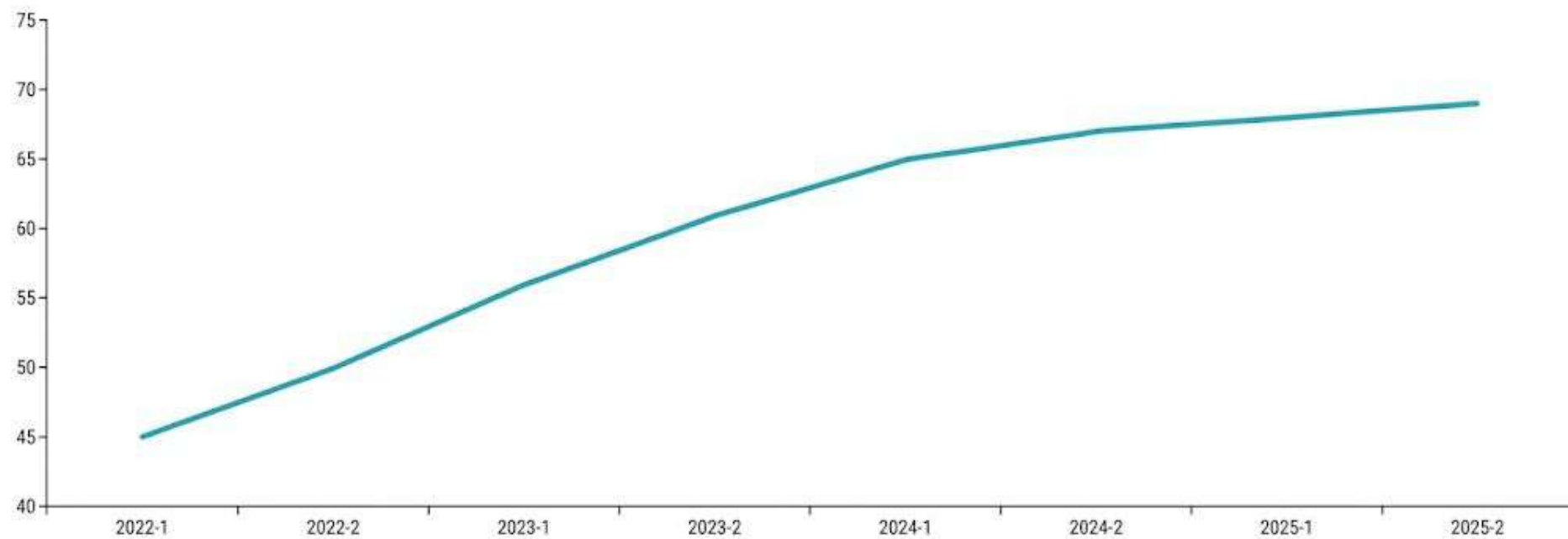
50 → 98 matrículas en ocho periodos

Patrón 2: descenso gradual



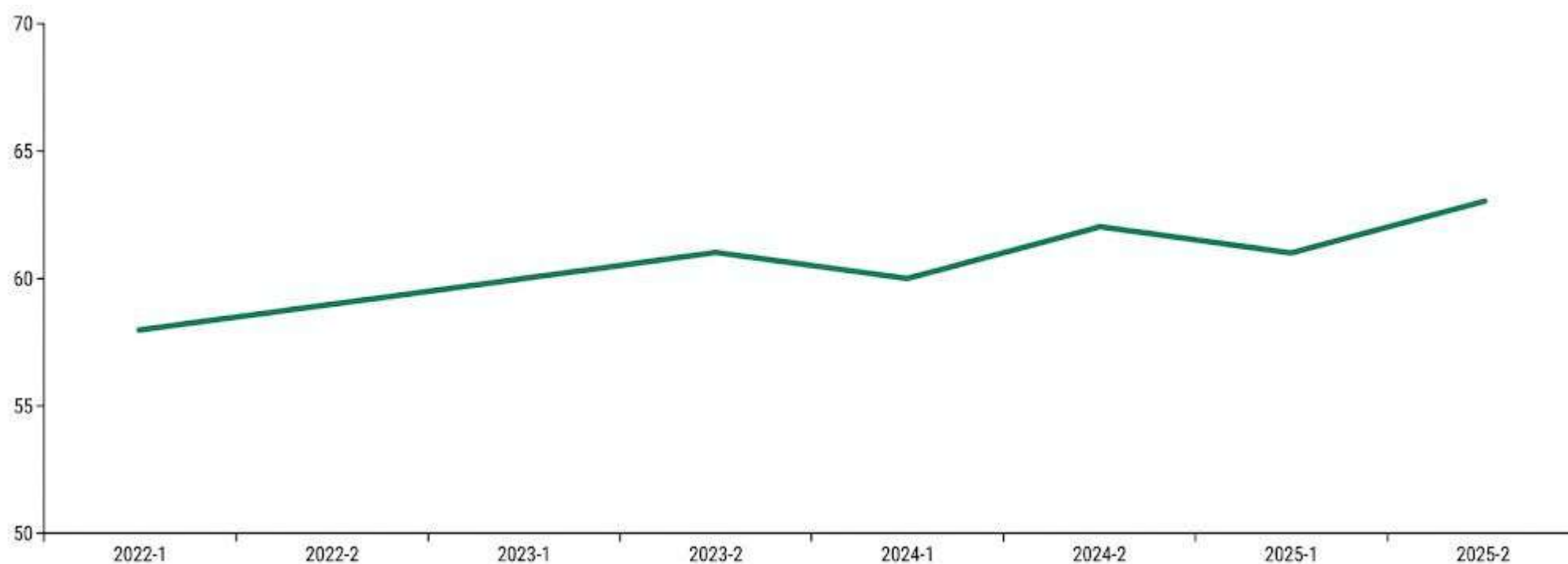
72 → 47 matrículas en ocho periodos

Patrón 3: crecimiento y estabilización



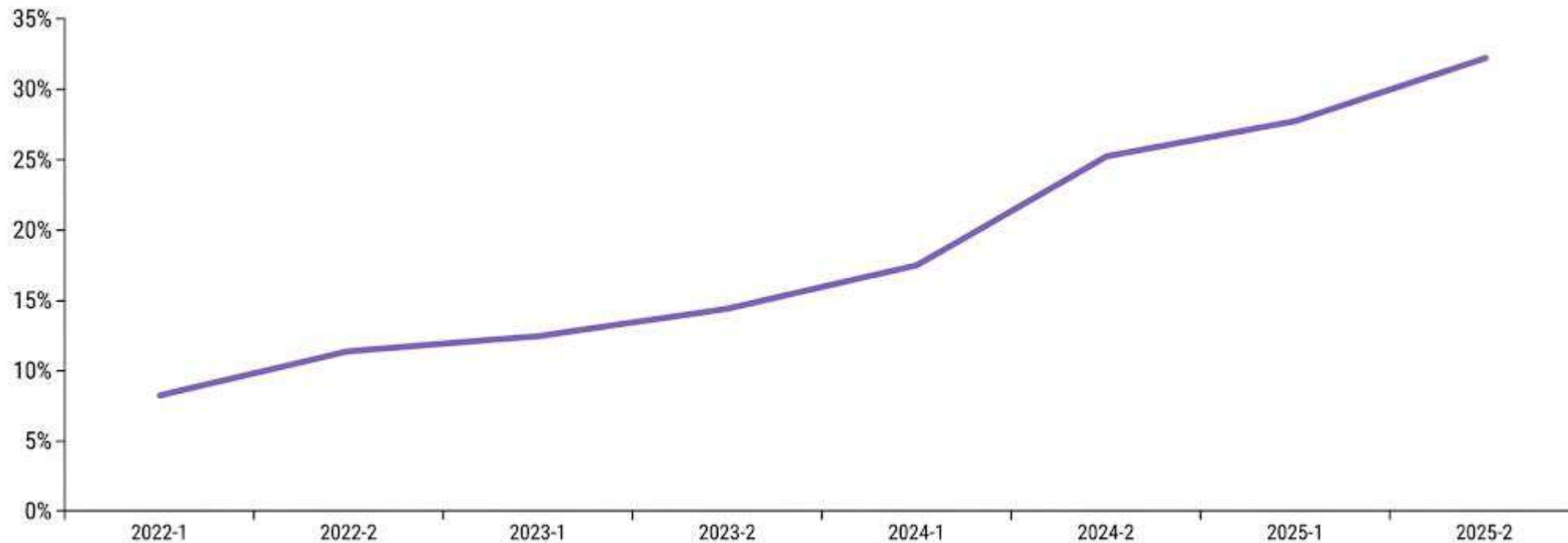
El crecimiento se desacelera: 65 → 67 → 68 → 69

Patrón 4: estabilidad



El volumen se mantiene cerca de 60 matrículas por periodo

Otra señal institucional: crecimiento de la modalidad virtual



8,3% → 32,2% de las matrículas

Un patrón observado no es una causa

Patrón observado

“La participación de la modalidad virtual aumenta a través de los periodos.”

Explicación causal

“Aumentó porque los estudiantes prefieren la virtualidad.”

La segunda afirmación requiere evidencia adicional

Lectura comparativa: ¿con qué estoy comparando?

1

Tiempo

Actual vs. anterior /
histórico

2

Grupo

Programa vs. programa

3

Referencia

Unidad vs. institución

4

Distribución

Centro vs. dispersión

Una cifra aislada informa. Una comparación bien elegida permite interpretar.

Antes de interpretar una diferencia, pregunta

- 1 ¿La diferencia es grande en términos absolutos o relativos?
- 2 ¿El volumen de observaciones es comparable entre los grupos?
- 3 ¿La diferencia se mantiene en varios periodos o aparece una sola vez?
- 4 ¿Existe un cambio en la composición del grupo o en la fuente de datos?
- 5 ¿La variabilidad interna podría explicar parte de la diferencia?

7. Power BI

Power BI

Construir medidas de dispersión para comparar con contexto

Medidas que agregaremos al PBIX

Medida	Pregunta que responde	Función DAX
Mínimo	¿Cuál es el menor valor?	MIN
Máximo	¿Cuál es el mayor valor?	MAX
Rango académico	¿Cuál es la amplitud?	MAX – MIN
Desv. estándar	¿Qué tan dispersos están los valores?	STDEV.P
Variación vs. periodo anterior	¿Cuánto cambió?	DIVIDE + contexto temporal

El objetivo sigue siendo comprender qué calcula cada medida, no memorizar sintaxis.

DAX: valor mínimo y máximo

Mínimo =MIN(Matriculas[Promedio_Semestral])

Maximo = MAX(Matriculas[Promedio_Semestral])

Mínimo global

9,5

Máximo global

19,7

Ambas medidas cambian cuando aplicamos filtros de programa, periodo o modalidad.

DAX: rango académico

Rango académico = [Promedio máximo] - [Promedio mínimo]

Buena práctica: reutilizar medidas existentes mejora la lectura y el mantenimiento del modelo.

La medida debe responder correctamente al contexto de filtros.

DAX: desviación estándar

Desviación estándar = STDEV.P(Matriculas[Promedio_Semestral])

Informática: 0,88

Marketing: 1,85

La misma fórmula revela resultados distintos al cambiar el programa.

¿STDEV.P o STDEV.S?

STDEV.P

Cuando los datos que estás analizando representan la población completa que quieres describir.

STDEV.S

Cuando los datos son una muestra y quieres estimar la variabilidad de una población mayor.

Para el laboratorio utilizaremos STDEV.P sobre el dataset completo.

Resultado esperado: compara centro + dispersión

Programa	Promedio	Mediana	Rango	Desv. estándar
Ing. Informática	14,54	14,60	5,2	0,88
Ing. Civil	13,33	13,35	7,6	1,32
Marketing	14,57	13,90	8,5	1,85
Psicología	15,26	15,30	4,4	0,75

No busques “el mejor” número. Busca una lectura coherente con la pregunta.

8. Actividad en Power BI

Actividad

Construye la segunda capa analítica del dashboard

Actividad: variabilidad y señales

Tiempo sugerido: 35–40 minutos

- 1 Abre el PBIX común utilizado en la sesión 8.
- 2 Crea las medidas de mínimo, máximo, rango y desviación estándar.
- 3 Compara al menos cuatro programas utilizando promedio + mediana + dispersión.
- 4 Analiza la evolución de matrículas por periodo e identifica al menos dos patrones y una señal atípica.
- 5 Redacta una interpretación breve diferenciando evidencia e hipótesis.

Parte 1: medidas obligatorias

Medida	Pista DAX
Promedio mínimo	MIN
Promedio máximo	MAX
Rango académico	Máx. – Mín.
Desviación estándar	STDEV.P
Una medida adicional	Elección del participante

Gemini puede ayudar con la fórmula, pero debes explicar qué calcula.

Parte 2: encuentra señales en los datos

- 1 ¿Qué programa muestra mayor variabilidad académica?
- 2 ¿Qué programa parece más estable?
- 3 ¿Dónde encuentras una tendencia sostenida?
- 4 ¿Qué serie contiene un pico o caída que rompe el patrón?
- 5 ¿Qué hipótesis investigarías antes de explicar esa señal?

Tu mini-dashboard debe crecer, no empezar de cero



Checklist de validación

- ✓ ¿Las medidas cambian al aplicar filtros?
- ✓ ¿El rango corresponde a máximo – mínimo?
- ✓ ¿La desviación estándar está en la misma escala del promedio?
- ✓ ¿Puedes explicar por qué Marketing es más variable?
- ✓ ¿Distingues un patrón observado de una hipótesis causal?

9. Gemini

Gemini

De calculadora de fórmulas a copiloto de interpretación

Gemini: primero describe, después interpreta

1. Describir

“Compara promedio, mediana, rango y desviación estándar de los programas. Señala únicamente diferencias observables en los valores.”

2. Interpretar

“Ahora propón posibles explicaciones como hipótesis. Separa claramente evidencia, inferencia y preguntas que debemos investigar.”

Separar evidencia de hipótesis reduce alucinaciones analíticas

Prompt 1: explicar variabilidad

Tengo estas medidas por programa: promedio, mediana, rango y desviación estándar.

1. Explica qué aporta cada medida.
2. Compara los programas sin inventar causas.
3. Señala dónde el promedio podría ser insuficiente.
4. Indica qué validaciones adicionales harías.

Primero evidencia; después interpretación

Prompt 2: identificar patrones

Analiza la evolución de matrículas por programa entre 2022-1 y 2025-2.

Clasifica cada serie como crecimiento, descenso, estabilidad, meseta o comportamiento atípico.

Para cada clasificación, cita los valores que sustentan tu conclusión.

No propongas causas todavía.

Pedir evidencia numérica obliga a Gemini a justificar la clasificación.

Prompt 3: formular hipótesis sin convertirlas en hechos

Marketing presenta un pico de matrículas en 2023-2.

Formula 4 hipótesis plausibles que podrían explicar el cambio.

Para cada hipótesis, indica qué dato o evidencia necesitaríamos para validarla o descartarla.

No presentes ninguna hipótesis como causa confirmada.

Una buena hipótesis debe venir acompañada de una prueba posible

Evidencia, inferencia e hipótesis

Nivel	Ejemplo	¿Qué tan seguros es?
Evidencia	Marketing pasó de 39 a 65 matrículas.	Observado
Inferencia	El cambio rompe el comportamiento de periodos vecinos.	Sustentado por comparación
Hipótesis	Podría haber existido una campaña de admisión.	Debe validarse

Gemini debe ayudarte a mantener separados estos tres niveles.

10. Resumen de la sesión

3 prácticas a consolidar

Medir

**Centro + dispersión
para describir mejor
los datos.**

Comparar

**Grupos, periodos y
referencias con
contexto.**

Investigar

**Patrones y anomalías
sin convertir hipótesis
en hechos.**

La IA acelera el análisis; el criterio determina qué conclusiones son defendibles.

Cierre: un resumen de las 2 últimas sesiones

Pregunta	Medida / enfoque
¿Dónde está el centro?	Media · mediana · moda
¿Qué tan dispersos están?	Rango · IQR · desviación estándar
¿Qué cambia entre grupos?	Lectura comparativa
¿Qué comportamiento aparece en el tiempo?	Tendencias · estabilidad · anomalías
¿Qué podría explicar una señal?	Hipótesis que requieren evidencia

Una buena interpretación combina medidas, contexto y preguntas de validación.

Próxima sesión

Sesión 10 · Cómo elegir la visualización adecuada

Barras

Líneas

Dispersión

Histogramas

Cajas

Ahora que sabemos qué medir e interpretar, aprenderemos cómo representarlo.

Tarea: Próxima sesión

- Leer guía básica de visualizaciones en power BI
- Trae un gráfico o visualización



Evaluación de la Sesión



<https://docs.google.com/forms/d/e/1FAIpQLSefDWVWtRfqMuFD1mJMU3xsePBLH3zG-sga1nLQpqBigGBwyA/viewform?usp=dialog>

Fin Sesión 9



La *transformación digital*
es parte de nuestro

Plan
Estratégico
Institucional

23
27



Conoce más ↗