


# Fundamentos de Datos para la Toma de Decisiones Estratégicas con IA

Clase 01 - El valor estratégico de los datos y fundamentos para entender la IA

Sebastián Egaña Santibáñez 

## Enlaces del profesor

-  [sebastian.egana@udd.cl](mailto:sebastian.egana@udd.cl)
-  <https://segana.netlify.app>
-  <https://www.linkedin.com/in/sebastian-egana-santibanez/>
-  <https://github.com/sebaegana>

## Clase 01

### Contexto actual

- Volumen de datos global crece >40% anual (IDC, 2024).
- Solo ~30% de las empresas se considera realmente *data-driven* (McKinsey, 2023).
- El valor no está en acumular datos, sino en **estructurarlos, analizarlos e interpretarlos**.

## Rol de la IA

- IA = capacidad de **extraer patrones y predicciones**
- Integra datos estructurados y no estructurados (texto, imágenes, audio, video, documentos) y los convierte en información utilizable; arquitecturas empresariales actuales permiten extraer y mapear contenido no estructurado y combinarlo con recuperación + generación (RAG)
- Permite **reducir incertidumbre y mejorar la velocidad y calidad** de la decisión

## Estrategia basada en datos

### Diferencia entre decisiones intuitivas vs. basadas en datos.

Por ejemplo, en el ámbito personal, ¿cómo decidir entre una oferta de trabajo que implica un mayor salario pero requiere re localizarse o una oferta con menor salario pero un mejor equilibrio entre la vida laboral y personal?

La dualidad acá está entre tomar una decisión en base a nuestro instinto o en base a un sistematización de puntos a favor o puntos en contra. Esto independiente del **método** o **metodología**.

Para esto, debemos desarrollar un framework orientado a la toma de decisiones; en la actualidad se habla de **decision science**, como el campo multidisciplinario orientado a desarrollar un enfoque sistemático y orientado en datos (data-driven) para resolver problemas y optimizar desempeño y resultados.

Como áreas relacionadas se encuentran las matemáticas, estadística, psicología, economía, ciencias de la computación, etc.

El punto principal se basa en tomar decisiones informadas; una herramienta común a las ciencias es el método científico.

### Método científico

Corresponde a la metodología aplicada en ciencias orientada a entender la realidad y generar nuevos conocimientos en base a la observación sistemática, medición y experimentación en base al análisis y formulación de hipótesis. Se puede descomponer en las siguiente etapas:

1. Observación
2. Planteamiento de hipótesis
3. Experimentación / Medición
4. Análisis de resultados

## 5. Conclusión

En este sentido, **decision science** busca tomar los aspectos positivos del conocimiento científico en el ámbito de la toma de decisiones para organizaciones.

Podemos afirmar lo siguiente: “Cada decisión en salud, finanzas o negocios se basa en una predicción — explícita o implícita. El valor/desafío está en hacerla con evidencia”.

- Decisiones: independiente del tipo de institución, es necesario tomar decisiones de gestión
- Predicción: visión del futuro en base a la información del pasado
- Evidencia: relación con el método científico y **decision science**

### Framework de trabajo

El enfoque de **decision science** plantea el siguiente framework:

1. Definir el problema
2. Reunir y analizar datos
3. Desarrollar y evaluar alternativas
4. Seleccionar e implementar soluciones

Se debe tener claridad en la toma de decisiones (no desarrollar/invertir en nada que no tenga sustento en datos), identificar tendencias y generar predicciones para reducir/interiorizar la incertidumbre.

- Por ejemplo: generar una gestión correcta de camas dentro de un recinto hospitalario para aumentar la rentabilidad/bienestar de maximizar la ocupación del recinto.

### Sesgos más allá de los datos

⚠ Un warning relevante es la necesidad de estar siempre atentos a los posibles sesgos inconscientes dentro de la toma de decisiones:

- Confirmation bias: busca información que confirma mis creencias (contratación: cv versus entrevista)
- Anchoring bias: sobre confianza en la información inicial (precios del cyber)
- Loss aversion: preferencia hacia evadir pérdidas sobre (perder un 10% de descuento versus ganar un 10% de descuento)
- Available heuristic: sobre estimar la frecuencia de un evento en base a lo fácil que son de recordar (caída de aviones)

## **Ejemplo**

¿Qué se prefiere?

- Situación A: Ganar 500 USD con seguridad
- Situación B: 50% de probabilidad de ganar 1.000 USD y 50% de no ganar nada

Resultado:

- La mayoría elige A, pero ¿es correcto eso?

## **Sobre la racionalidad**

Se asume que siempre estamos en un contexto de racionalidad perfecta, lo que implica la existencia de información completa y la no existencia de sesgos ni emociones. A pesar de esto, en la realidad, se toman decisiones en racionalidad incompleta o limitada:

1. Los seres humanos tenemos límites cognitivos
2. Tiempo e información limitados
3. Estamos sujetos a emociones y sesgos



## Utilización de datos

Jack Maple y COMPSTAT Center



Figure 1: Recuperada en: <https://www.nydailynews.com/2013/09/16/hamill-bill-de-blasio-says-bill-bratton-on-short-list-to-be-police-commissioner>

## Tiempo promedio de hospitalización:

En un caso simple, podríamos pensar en el promedio de días que pasa cada paciente por alguna patología en particular.

Tenamos los siguientes datos de altas pacientes con neumonía y queremos predecir el tiempo promedio de hospitalización en días:

```
tiempo <- c(2,3,1,5,7,8,9,3,4,6,2,10,11,7,5,4,12,8,9,6)
alta    <- c(1,1,1,1,1,0,0,1,1,1,1,0,0,1,1,1,0,1,1,1)
```

Acá la censura corresponde a los pacientes que no han sido dados de alta (alta = 0).

Table 1: Conteos: total, eventos (DELTA=1), censuras (DELTA=0)

Total	Eventos_DELTA1	Censuras_DELTA0	Proporcion_evento
20	15	5	0.75

Table 2: Resumen global de TCC (días)

n	min	mean	mean_obs	sd	median	max
20	1	6.1	4.8	3.161	6	12

Veamos el análisis de Weibull para estos datos:

Table 3: Parámetros Weibull 2P (ajuste con censura)

Caso	shape (k)	scale ( )
Con censura	1.5848	7.8569

Table 4: Mediana, media y percentiles t\_p (p=0.1, 0.5, 0.9) con censura

Caso	Mediana	Media	p=0.1	p=0.5	p=0.9
Con censura	6.2347	7.0506	1.8993	6.2347	13.2985

## Concepto de Data-Driven Organization.

Es algo que **al parecer** todos tenemos claros, pero cuáles son las características de estas empresas (ESE, 2021 y Castro, 2023):

- Estrategia y gobierno de datos: en el centro debe estar la mejora en la toma de decisiones
- Democratizar el acceso a los datos: tanto como el acceso, como las herramientas
- Promover la colaboración entre equipos multidisciplinarios
- Crear programas sólidos de calidad de datos
- Impulsar la innovación continua para integrar el cambio
- Fomentar la cultura Data-Driven en todos los niveles de la organización

## Experimentación

Dentro del contexto de **decision science** hablamos de la conexión con el método científico; elemento clave es la experimentación.

- Debemos pasar de intuir a medir: [The Surprising Power of Online Experiments](#)
- Si no experimentas, adivinas: [Great Strategy Starts with Experimentation](#)
- Cultura de experimentación: [Democratizing online controlled experiments at Booking.com](#)

## Algunos ejemplos

Epigenética y la utilización de ratones 🐭

- Experimentación en ratones, ¿por qué?
- No mutaciones del ADN, pero que pueden ser heredables.

NASA versus Space X 🚀

- “Disminuir infinitamente el riesgo aumenta infinitamente el costo”
- Space X con un costo de 67 million por lanzamiento para el Falcon 9; Nasa SLS sobre los \$4 billion USD ([más info](#))

## Actividad

### “Decisiones con o sin datos”

Duración: 15–20 minutos Objetivo: Reconocer cómo se suelen tomar decisiones (intuición vs. evidencia).

Instrucciones:



- Grupos aleatorios entre 3 - 4 personas.
- Cada equipo comparte un ejemplo de una decisión importante tomada en su trabajo.

Deben responder:

1. ¿Qué datos se usaron (si los hubo)?
2. ¿Qué sesgos pudieron influir?
3. ¿Cómo habría cambiado la decisión si se hubiesen considerado más datos?
4. Cada grupo elige un caso para comentar brevemente al curso

## **Fundamentos estadísticos esenciales**

Como conceptos iniciales, en relación a los modelos debemos recordar lo siguiente:

- Buenos modelos no son suficientes sin impacto
- La indefición de lo que consideramos éxito, conlleva a la pérdida de tiempo
- Lo que consideramos éxito debe ser acordado de antemano

## **Medidas de éxito**

- Desempeño: exactitud (accuracy), precisión u otros indicadores técnicos
- Tiempo: entrega en plazo
- Costo: relación con presupuesto
- Calidad: Claridad en el código, documentación y reproducibilidad
- Impacto en stakeholders: satisfacción, comunicación, etc.

## **Conceptos clave: promedio, varianza, correlación, causalidad.**

Podemos partir por lo más conocidos, como las medidas de tendencia central y de dispersión y correlación

- Media, mediana, moda
- Varianza y desviación estándar
- Correlación

Precio de acción de LATAM:

## LATAM Airlines Group

Último cierre: \$21,65 CLP (2025-10-15)

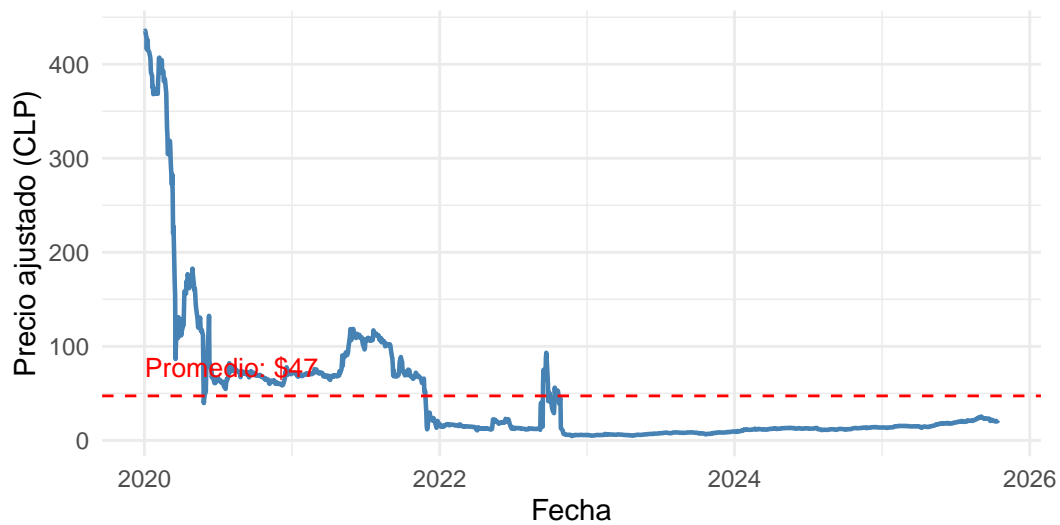


Fuente: Yahoo Finance via tidyquant

Veamos una ventana de tiempo distinta:

## LATAM Airlines Group

Último cierre: \$21,65 CLP (2025-10-15)



Fuente: Yahoo Finance via tidyquant

Por último:

## LATAM Airlines Group

Último cierre: \$21,65 CLP (2025-10-15)



Fuente: Yahoo Finance via tidyquant

### Tipos de modelos

Por lo general, se habla de dos tipos de modelos:

- Modelos de regresión: variable objetivo es numérica o continua
- Modelos de clasificación: variable objetivo es categórica o discreta

Otros modelos relacionados con la clasificación:

- Clustering: buscan descubrir patrones dentro de los datos

Dependiendo el tipo de aprendizaje se puede hablar de supervisados y no supervisados.

### Métricas de modelos

Algunas de las métricas más utilizadas son:

- Accuracy: medición general de éxito
- Precision: cuántos predichos positivos fueron efectivamente correctos \_\_ Recall: cuán bien el modelo predice a los que efectivamente eran positivos
- F1 Score: es la media armónica entre ambas
- Mean absolute error (MAE): tamaño promedio de la predicción de errores
- Mean percentage error (MAPE): cuán lejos están las predicciones en porcentaje

## Matriz de confusión

		Ground truth		
		+	-	
Predicted	+	True positive (TP)	False positive (FP)	Precision = TP / (TP + FP)
	-	False negative (FN)	True negative (TN)	
		Recall = TP / (TP + FN)		Accuracy = (TP + TN) / (TP + FP + TN + FN)

Figure 2: Recuperada en: <https://www.innovatiana.com/es/post/understand-confusion-matrix-in-ai>

## Precision

Quieres evitar falsos positivos: es decir, no marcar algo como positivo si no lo es

Ejemplo	Por qué importa más la <i>precision</i>
<b>Filtros de spam en correo electrónico</b>	Es mejor dejar pasar un correo spam que bloquear uno legítimo. Un falso positivo (bloquear un correo válido) es más dañino.
<b>Sistemas de detección de fraude en tarjetas</b>	Acusar erróneamente una transacción legítima como fraude genera molestia y pérdida de confianza del cliente.
<b>Diagnóstico automático de enfermedades raras</b> (como soporte, no reemplazo médico)	Un falso positivo puede causar ansiedad y exámenes innecesarios costosos.

## Recall

Quieres evitar falsos negativos: es decir, no dejar pasar algo que sí era positivo.

Ejemplo	Por qué importa más el <i>recall</i>
<b>Detección de cáncer o enfermedades graves</b>	Es más grave no detectar un caso real (falso negativo) que generar una alarma falsa que luego se descarte.
<b>Detección de fraude bancario</b>	Prefieres marcar más casos sospechosos (aunque algunos sean falsos) para revisar todos los posibles fraudes.
<b>Sistemas de seguridad o vigilancia</b>	Mejor analizar más alertas que dejar pasar una amenaza real.

## Incertidumbre y confianza en la interpretación de resultados.

Se debe tener en cuenta que la incertidumbre siempre va a existir y se debe convivir con eso. Esto no implica que no la interioricemos dentro de nuestra toma de decisiones o que no la intentemos medir.

Una buena técnica es generar ejercicios que añaden incertidumbre a los modelos calculados.

- En el análisis de datos, la incertidumbre no se elimina, se cuantifica. Por eso usamos intervalos de confianza, desviaciones estándar o bandas de predicción. Estas herramientas nos permiten comunicar no solo qué esperamos que ocurra, sino cuán seguros estamos de ello.
- La clave no es eliminar la incertidumbre, sino tomar decisiones que sean robustas frente a ella. En un entorno data-driven, la confianza no proviene de la certeza, sino de la capacidad de medir y comunicar el riesgo asociado a cada decisión.

## Simulaciones de Monte Carlo

Una simulación corresponde a imitar en un entorno controlado para experimentar posibles comportamientos futuros de un sistema particular. En este caso lo que se simula es el valor futuro de la serie de tiempo que estamos intentando predecir.

El concepto de **Monte Carlo** corresponde al muestreo aleatorio repetido de múltiples valores futuro para poder incorporar incertidumbre dentro de nuestras predicciones.

## Aplicación en evaluación de proyectos

En el caso de proyectos, podemos realizar un análisis de carácter estocástico, lo que implica asumir que alguna de las variables no es conocida pero si que puede estar situada dentro de algunos parámetros.

Veamos el siguiente ejemplo:

Una empresa está evaluando la introducción de un nuevo producto y desea conocer la probabilidad que obtenga pérdidas. Para cumplir dicho propósito lo contrata a usted para construir un modelo financiero del negocio.

Después de realizar la debida investigación, usted decide construir una simulación de Monte Carlo que le permita generar la distribución de probabilidad de las utilidades del negocio, modelando separadamente los ingresos y los costos totales, considerando los siguientes elementos:

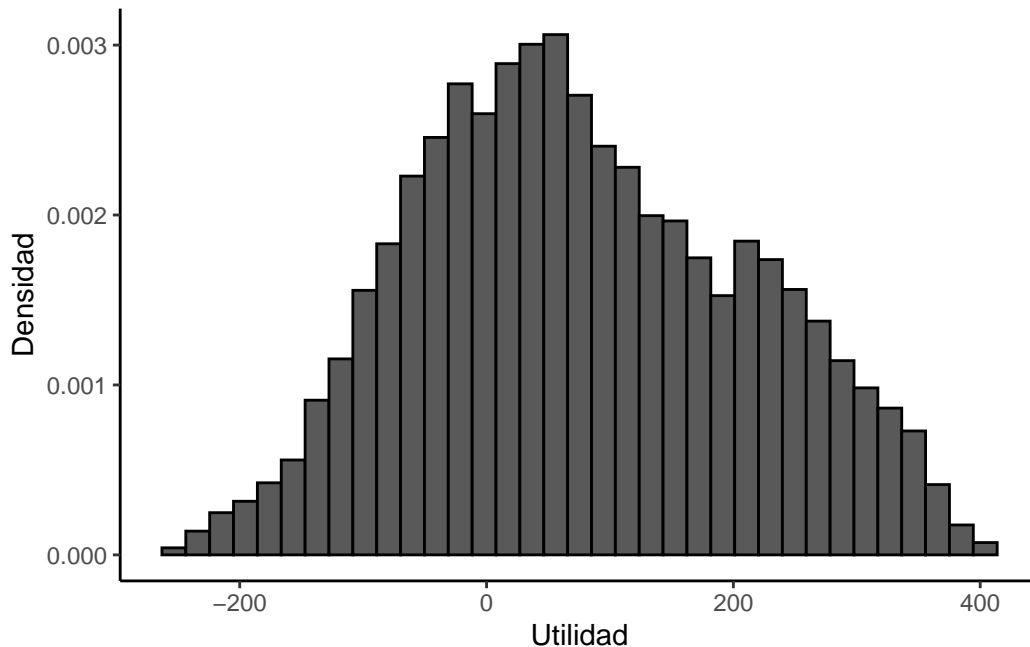
Por el lado de los ingresos totales, se considerarán tres escenarios (A, B y C) para el precio y la cantidad, los cuales aparecen descritos en la siguiente tabla:

Variable	A	B	C
<b>Precio</b>	12	13	16
<b>Cantidad</b>	110	100	80
<b>Probabilidad</b>	1/4	1/2	1/4

En el caso de los costos totales, los costos fijos son iguales a \$150 y los costos variables unitarios son constantes y son modelados usando una distribución triangular con los siguientes parámetros:

Parámetro	Valores
<b>Valor mínimo</b>	9
<b>Valor máximo</b>	13
<b>Moda</b>	11

A. ¿Cuál es la probabilidad que el negocio tenga pérdidas?



```
[1] "Prob(utilidad<0) = 0.31"
```

B. Calcule el valor esperado de las utilidades del negocio

```
[1] "Valor esperado utilidad = 76.65"
```

## Referencias

- ESE. (2021). Data-Driven Organization: ¿Qué es una organización dirigida por datos? Resumen Transformación Digital N° 23. ESE. Recuperado de [https://www.es.cl/es/site/artic/20210518/asocfile/20210518180050/data\\_driven\\_organization\\_que\\_es\\_una\\_organizacion\\_dirigida\\_por\\_datos.pdf](https://www.es.cl/es/site/artic/20210518/asocfile/20210518180050/data_driven_organization_que_es_una_organizacion_dirigida_por_datos.pdf)

Castro Reyes, D. A. (2023). ¿Qué pueden hacer las organizaciones para convertirse en Data-Driven? EY Colombia. Recuperado de [https://www.ey.com/es\\_co/insights/consulting/analytics-consulting-services/organizaciones-convertirse-data-driven](https://www.ey.com/es_co/insights/consulting/analytics-consulting-services/organizaciones-convertirse-data-driven)

- Vecillas Martin, D., Berruezo Fernández, C., & Gento Municio, A. M. (2025). Systematic Review of Discrete Event Simulation in Healthcare and Statistics Distributions. Applied Sciences, 15(4), 1861. [Enlace aquí](#)

- Heins, J., Schoenfelder, J., Heider, S., Heller, A. R., & Brunner, J. O. (2022). A scalable forecasting framework to predict COVID-19 hospital bed occupancy. *INFORMS Journal on Applied Analytics*, 52(6), 508-523. [Enlace aquí](#)