



Capability Building

Data DRIVEN

El modelo que responde la pregunta

MÓDULO 3

Módulo 3:

El modelo que responde la pregunta

Como vimos anteriormente, partimos de una pregunta que producirá una hipótesis. Para testear nuestra hipótesis primero debemos diseñar la ecuación que nos permitirá confirmarla o refutarla, luego definir los datos que necesitamos obtener para completar la ecuación, y si los datos no se encuentran disponibles, diseñar los estudios necesarios para obtenerlos.

Por lo tanto, en este módulo nos vamos a concentrar en dos grandes temas: la 'ecuación' y la 'data' de la ecuación. La ecuación es el análisis que debemos hacer con nuestros datos. ¿Y por qué nos referimos al análisis como ecuación? Porque en lugar de ir primero a la data disponible, analizarla e intentar generar hallazgos y conclusiones, lo hacemos de modo inverso. Primero pensamos el tipo de relación o ecuación (pasos algebraicos) que resolverán nuestra incógnita (o pregunta inicial) y luego buscamos los datos necesarios para la ecuación o los generamos. La ecuación, en definitiva, es un modelo matemático.

ATENCIÓN: aunque tengamos una única incógnita, no hay una única ecuación para obtener la respuesta. Lo que nos ayudará a elegir entre las respuestas es el costo, la conveniencia y la confiabilidad de la respuesta.

Por ejemplo, si mi hipótesis es que el candidato oficialista X ganará las elecciones, una ecuación podría ser mostrar la relación de la tasa de desempleo, el crecimiento económico, el índice de aproba-

ción y el índice de confianza del consumidor que tomaré como indicadores para demostrar el hecho de que en determinado porcentaje de casos, cuando dichos indicadores son altos, gana el candidato oficialista.

Otra ecuación podría ser la cantidad de gente que declara escogerlo entre otros candidatos. En este caso la ecuación sería una encuesta, o lo que llamamos un experimento que nos permitió generar datos para resolver nuestra ecuación.

El punto es, puedo generar distintos tipos de ecuaciones para responder lo mismo, y no en todas las ecuaciones tener los datos necesarios para calcularla.

A continuación, profundizaremos en herramientas comunes y útiles en el planteo de ecuaciones, y posteriormente, en experimentos comunes para generar datos.

Herramientas comunes de análisis

La ecuación es nuestro modelo matemático. Para que nuestro modelo no falle, existen una serie de conceptos fundamentales que cualquiera que quiera dirigir sus acciones con datos debería manejar. Estos conceptos son:

- ★ Valor esperado.
- ★ Retorno sobre inversión.
- ★ Correlación.
- ★ Causalidad.

El valor esperado

(esperanza matemática o expected value)

El valor esperado, también llamado esperanza matemática o promedio ponderado, es igual a la sumatoria de las probabilidades de que exista un suceso, multiplicado por el valor de dicho suceso. Este valor se produce cuando ordenamos las informaciones en base a la ponderación entre el resultado esperado y la probabilidad de ocurrencia del mismo.

Veamos un caso como ejemplo:

Imaginemos una moneda con dos caras: cara y cruz. ¿Cuál sería la esperanza matemática (valor esperado) de que salga cara? Dado que la moneda solo puede caer en una de esas dos posiciones y ambas tienen la misma probabilidad de ocurrencia,

diremos que la esperanza matemática de que salga cara es una de entre dos posibles sucesos, o lo que es lo mismo, el 50% de las veces.

Ahora, te invito a resolver el siguiente caso:

Imagina por un momento que Estados Unidos se está preparando para el brote de una enfermedad asiática inusual que se espera que mate a 600.000 personas. Se han propuesto dos programas alternativos para combatir la enfermedad. Supongamos que las estimaciones científicas exactas de las consecuencias de los programas son las siguientes.

Programa A: Si se adopta el Programa A, se salvarán 200.000 personas.

Programa B: Si se adopta el Programa B, hay una probabilidad de un tercio de que se salven 600.000 personas y una probabilidad de dos tercios de que no se salve ninguna.

Si tuvieras que elegir el Programa A o el Programa B basándote únicamente en la información proporcionada en el problema, ¿qué programa elegirías? una elección por intuición o realizaste algún cálculo para seleccionar tu respuesta?

La respuesta lógica es elegir el programa con el mayor valor de resultado posible de vidas salvadas. Por lo tanto, no debería sorprendernos que 9 de cada 10 personas elija el Programa A.

Sin embargo, si observamos detenidamente, efectivamente se trata de un análisis de esperanza matemática, en donde en ambos casos se espera que se salven, en promedio, 200.000 personas. Si llegaste a esta conclusión, te felicito: formas parte de menos del 5% de la población que sabe cómo resolverlo.

Ahora analicemos el cálculo rápidamente para que quede claro para todos:

- $600.000 \times \frac{1}{3}$ de que se salven todos, equivale a 200.000 personas salvadas; o
- $600.000 \times \frac{2}{3}$ de que no se salve nadie, equivale a 400.000 personas muertas.

Por lo tanto, el valor esperado, o esperanza matemática, es ordenar las informaciones en base a la ponderación entre el resultado esperado y la probabilidad de ocurrencia del mismo.

El Retorno de la inversión

El retorno de la inversión o *Return of Investment* (ROI) es una ecuación que nos permite analizar los beneficios generados a partir de una inversión determinada. Es un cálculo sumamente útil para comparar diferentes escenarios de inversiones (iniciativas o propuestas de negocio) y determinar cuál de ellas produciría la mayor ganancia para la compañía.

La forma más simple de la fórmula para el ROI involucra sólo dos valores: **el costo de la inversión y la ganancia de la inversión**. La fórmula a utilizar es la siguiente:

$$\text{ROI} (\%) = \frac{\text{Ganancia} - \text{Costo de la inversión}}{\text{Costo de la inversión}} \times 100$$

Vamos con un ejemplo:

Una empresa está considerando adquirir una nueva forma de tecnología de sistemas de información por \$500.000 dólares. En base a esta nueva tecnología, se espera procesar la información de los diferentes departamentos de manera más eficiente, generando un ahorro de \$1 millón de dólares en 20 años.

¿Cuál es el ROI de la inversión?

$$\text{ROI}_{20} = \frac{1.000.000 - 500.000}{500.000} \times 100 = 100\%$$

Significa que el resultado del proyecto otorga un retorno del 100% de la inversión. Sin embargo, si indagamos en el ROI del primer año, vemos que el cálculo sería:

$$\text{ROI}_1 = \frac{50.000 - 500.000}{500.000} \times 100 = -90\%$$

¿Qué significa un ROI negativo?

Independientemente del denominador, significa que la ganancia obtenida es menor a la inversión realizada, por lo que ni siquiera recuperamos el monto de la inversión. En contraposición, si el ROI es positivo

significa que hemos ganado más que el costo que gastamos en la inversión.

En el ejemplo, es evidente que deben esperar al menos 10 años para recuperar la inversión (donde el ROI equivaldrá a 0), para recién en el año 11 generar un ROI de 10%.

$$\text{ROI}_{11} = \frac{550.000 - 500.000}{500.000} \times 100 = 10\%$$

¿Por qué es importante calcular el ROI?

Nos permite justificar o defender la inversión propuesta (ya sea de tiempo, recursos, dinero, etc.) para una determinada iniciativa.

Nos permite establecer un criterio objetivo para seleccionar entre distintas alternativas de proyectos, iniciativas, u oportunidades.

De todos modos, parece sencillo calcular el ROI cuando el costo y las ganancias son cifras conocidas y constantes, pero esto sucede pocas veces. La realidad es que a lo largo del tiempo, el costo y las ganancias pueden verse afectados por diferentes factores. Por ejemplo, ¿qué pasaría si hubiera que sacar un préstamo para pagar la inversión?, ¿O qué pasaría si los beneficios no evolucionan en forma lineal? También pueden presentarse otras complicaciones con respecto a la forma de calcular la inversión y definir qué gastos incluir. Pensemos en una campaña de marketing: la inversión debería considerar sólo el costo de los medios, ¿o también debería incluir la inversión de tiempo del personal para crear el anuncio?

El ROI de la actividad de las redes sociales por lo general parece muy alto si solo consideramos los costos económicos, pero si consideramos los recursos humanos necesarios para desarrollar contenidos y dar respuesta a las publicaciones de los

consumidores las 24 horas del día, los 7 días de la semana, el ROI naturalmente disminuye.

Otro desafío radica en estimar los beneficios incrementales. ¿Cuáles habrían sido nuestras ventas y utilidades si no hubiésemos gastado en este programa de marketing? Además, el beneficio de una acción no siempre es fácilmente medible en términos financieros, ¿cómo podemos considerar los beneficios de mejorar la lealtad de clientes y disminuir su *churn*?

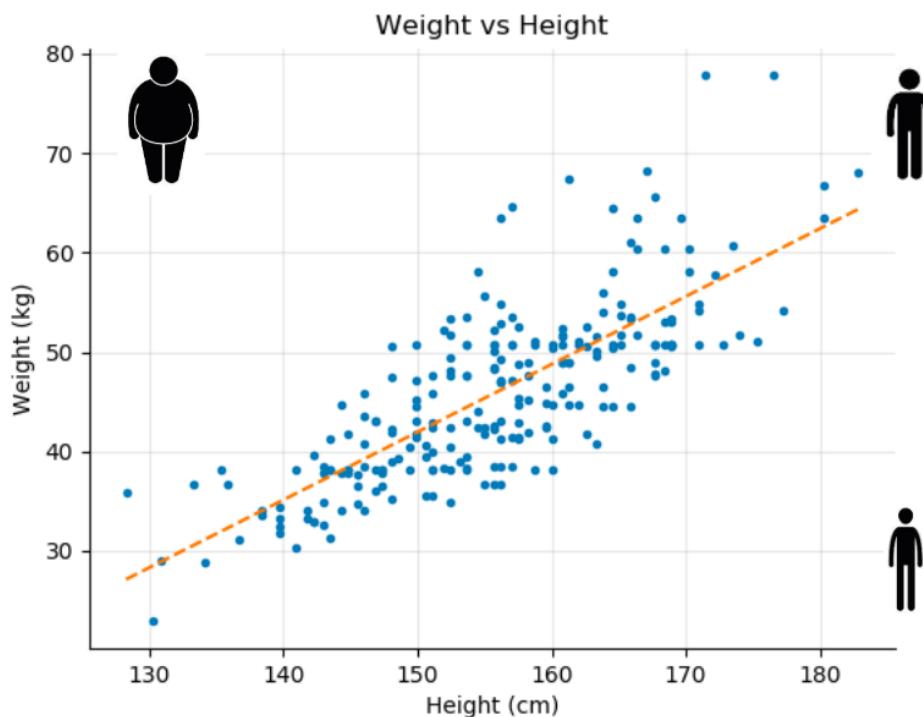
Como vemos, el concepto del ROI es simple pero su cálculo puede volverse complejo muy fácilmente. De todos modos, es de suma importancia implementarlo al momento de presentar cualquier iniciativa o acción, ya que nos brinda información valiosa sobre el potencial rendimiento de nuestra propuesta y su aplicación es necesaria para todas las decisiones de negocio.

Correlación

La correlación es un término que usamos a diario, pero si vamos a utilizarlo para manejar datos es importante definirlo y entenderlo con propiedad. Además, es un elemento fundamental en la construcción de algoritmos.

La correlación es una medida estadística que expresa hasta qué punto dos variables están relacionadas linealmente (lo que significa que cambian juntas a un ritmo constante). Cuando dos variables están correlacionadas, el valor de una variable podría, bajo ciertas condiciones, predecir el valor de la otra: esto es lo que nos interesa para tomar decisiones de manera asertiva.

Por ejemplo, una correlación positiva es aquella de altura y peso. Si grafico un grupo amplio de gente, seguramente veré un gráfico con los siguientes puntos.



Queda claro que a mayor altura, las personas pesan más aunque puede que haya algunos puntos que no siguen la lógica (aquellos muy delgados o excedidos de peso).

Esto es lo que se conoce como una correlación positiva: el aumento de una variable se relaciona con el aumento de la segunda variable.

La correlación también puede ser negativa. Por ejemplo, a mayores niveles de pobreza existen menores niveles de logros educativos.

¿De qué manera podemos utilizar la correlación para nuestras ecuaciones y análisis?

La correlación brinda cuatro usos útiles:

A. Predicción: Si existe una relación entre dos variables, podemos hacer predicciones sobre el comportamiento de una a partir de otra.

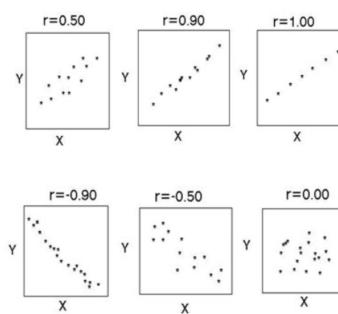
B. Validez: Validar el comportamiento de una nueva medida en base a una correlación que existió en el pasado con una medida semejante.

C. Fiabilidad: Confiabilidad en que las medidas observadas sean consistentes con medidas tomadas en el pasado.

D. Verificación: Validar una idea o teoría si puedo predecir un comportamiento.

¿Y qué tan fuerte es la correlación que descubrimos?

El coeficiente de correlación (r) se mide en un rango de entre -1 y 1, e indica hasta qué punto las dos variables se mueven en la misma dirección.



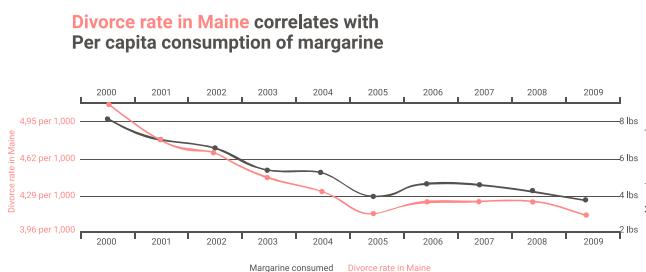
Una correlación de -1 indica una **correlación negativa perfecta**, lo que significa que cuando una variable sube, la otra siempre baja. Una correlación de +1 indica una **correlación positiva perfecta**, lo que significa que cuando una variable sube, la otra siempre sube.

Causalidad

Al hablar de correlación debemos detenernos obligatoriamente en el concepto de **Causalidad**. La causalidad implica que, si el valor de una variable cambia, entonces el valor de la segunda variable cambiará en consecuencia. Sin embargo, la **correlación no es suficiente para determinar un efecto causal, incluso cuando las dos variables están fuertemente correlacionadas entre sí**. Esto indica que la correlación no siempre significa causalidad.

Una correlación puede no implicar causalidad por varias razones:

- ★ Se pueden correlacionar dos variables entre sí por pura coincidencia. Por ejemplo, la tasa de divorcios en el estado de Maine, EEUU, se correlaciona el consumo per capita de margarina. En este caso, la correlación es casi perfecta. (correlación = 0,9925), pero no existe una conexión clara entre las dos variables¹.



- ★ Otra razón común por la cual la correlación no es lo mismo que la causalidad es que la causalidad

es una relación asimétrica (X causa Y, es diferente de Y causa X), mientras que la correlación es una relación simétrica. Por tanto, cuando X se correlaciona con Y, no nos dice si X está causando Y o Y está causando X.

- ★ Un pronóstico del tiempo a menudo se correlaciona con el tiempo real, pero el pronóstico del tiempo claramente no causa el pronóstico.
- ★ Las empresas que gastan más en informática (IT) tienden a tener un mejor desempeño (por ejemplo, ganancias). Pero, ¿podemos concluir que más gasto en IT conduce a mayores ganancias? ¿Deberíamos recomendar que todas las empresas inviertan más en IT?
- ★ Otra situación común en la que la correlación puede fallar en predecir la causalidad es cuando X e Y están correlacionados entre sí, pero la correlación se ve afectada por un tercer factor, Z. Por ejemplo, a medida que aumentan las ventas de helados, la tasa de muertes por ahogamiento aumenta bruscamente. Sin embargo, es poco probable que el consumo de helado provoque el ahogamiento. La razón de esta correlación tiene que ver con las altas temperaturas, que conducen a un mayor consumo de helado y al mismo tiempo a un mayor uso de piletas o espacios de agua.

Ahora supongamos que un gerente de negocio se enfrenta a la siguiente pregunta:

¿Debería invertir en capacitación universitaria para mis colaboradores?

Para responder esta pregunta amplia, podríamos comenzar por analizar la relación entre títulos universitarios y la productividad. Sin embargo, incluso si se observa una asociación positiva entre los títulos universitarios y la productividad, es difícil saber si esta relación es causal.

Es probable que haya otras diferencias subyacentes entre personas con y sin títulos. Es decir que sería necesario un estudio adicional para comprender si existe o no una causalidad real.

Realicemos una pequeña variación en la pregunta:

¿Debería contratar a más colaboradores con título universitario?

¹: Extraído de spurious-correlations. (<https://tylervigen.com/spurious-correlations>)

Ahora podríamos volver a observar la correlación entre los títulos universitarios y la productividad para considerar si contrataría trabajadores más productivos ajustando los criterios de contratación a trabajadores con un título universitario. En este caso sería útil la correlación, ya que está ayudando a predecir quién será (más) productivo.

Existe una diferencia clave entre estas dos preguntas:

“¿Debería contratar a más colaboradores con título universitario?” - es **un problema de predicción**, en el que buscamos observar si los títulos universitarios predicen que los colaboradores serán más productivos. *“¿Debería invertir en capacitación universitaria para mis colaboradores?”* - es **una inferencia causal**. Para determinar con éxito si los títulos universitarios ayudarán a mejorar el desempeño actual de los colaboradores, se necesita aplicar herramientas de inferencia causal, focalizados en comprender el impacto causal de realizar un cambio.

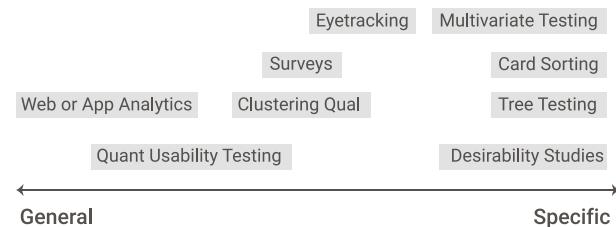
En el mundo corporativo, nos enfrentamos en forma casi constante a decisiones que implican pensar en el **impacto causal** de las diferentes opciones. *¿Mejorará la adopción de un producto si implementamos un programa de fidelidad? ¿Los salarios más altos reducirán la rotación? ¿Podremos incrementar el market share implementando la campaña de marketing? ¿Atraeremos nuevos clientes realizando publicidad en redes sociales?*

Todas estas preguntas requieren ser respondidas utilizando la inferencia causal y la aplicación de estudios observacionales frente al cambio de las variables en cuestión, por ejemplo, aplicando un ensayo aleatorio controlado.

Tipos comunes de experimentos

Una vez definida la ecuación que nos permitirá confirmar o refutar la hipótesis, debemos definir cómo conseguir la data que nos permitirá resolver la ecuación. En muchos casos, la data no existe y debemos producirla. Esto podemos lograrlo por medio de estudios o ensayos que diseñemos específicamente para conseguir los datos que necesitamos.

El siguiente cuadro ilustra distintos tipos de ensayos típicamente utilizados para la recopilación de data en productos digitales.



A continuación vamos a repasar algunos de los mismos. La siguiente no comprende una lista exhaustiva y priorizo aquellos más comúnmente utilizados:

Web or App Analysis:

- son métodos que miden el comportamiento del usuario en un sitio web o en una aplicación para realizar un seguimiento del rendimiento y detectar el potencial de mejora. Las métricas típicas que se analizarán y de las que aprender incluyen: usuarios activos diarios y mensuales, la cantidad de inicios de sesión, vistas de página o vistas de pantalla, tiempo de sesión y comportamiento de clic y desplazamiento.

Quant usability testing:

- método que mide la usabilidad de un producto digital en números. Mientras que los estudios de usabilidad cualitativos tienen como objetivo comprender si los usuarios encuentran útiles las funciones y dónde esperarían que se ubicara una función en particular, las pruebas de usabilidad cuantitativas miden las tasas de éxito de terminar una tarea o el tiempo dedicado a una tarea. Estos

estudios se repiten con el tiempo para comparar la efectividad de los cambios en el producto.

Surveys: herramienta común que permite recopilar todo tipo de datos cuantificables. Una encuesta puede pedir a los participantes calificaciones, proporciones y respuestas a preguntas de opción múltiple. Los tipos de preguntas pueden ser cerradas o abiertas. Un ejemplo de una herramienta de encuesta ampliamente utilizada es el NPS, que representa la lealtad de los usuarios.

A/B testing: método utilizado para probar diferentes versiones de diseño de una función entre sí. Cuando se cambia un elemento, como el color de un botón o una llamada a la acción, se mide y esto identifica qué versión funciona mejor. La combinación de pruebas A/B con análisis web o de aplicaciones facilita la identificación de qué versión tiene qué impacto en el comportamiento del usuario y las métricas de conversión.

Multivariate testing: tienen un objetivo similar al de las pruebas A/B. Sin embargo, miden el impacto en la conversión cuando hay múltiples cambios de elementos de diseño o de todo el diseño de la página al mismo tiempo.

Tree testing: método basado en tareas que identifica las jerarquías de información que tienen más sentido para los usuarios. Se le pide al participante que cree un árbol lógico de categorías y subcategorías con varios términos abstractos y específicos.

Desirability studies: ayudan a comprender cómo las audiencias evalúan el atractivo estético, la fuerza de la marca o el tono de voz de una oferta. Por tanto, se muestran imágenes o secuencias de vídeo de un producto o característica o bien los participantes utilizan el prototipo. Luego se les pide que describan su respuesta al producto eligiendo de una lista de adjetivos.



Capability Building

✉ hello@capabilitybuild.com

🌐 www.capabilitybuild.com

📍 175 SW 7th Street,
Suite 1501 Miami, FL 33130

SEGUINOS EN
NUESTRAS REDES!

