



Analítica de datos

Recolectando datos: Experimentos (Pruebas A/B)

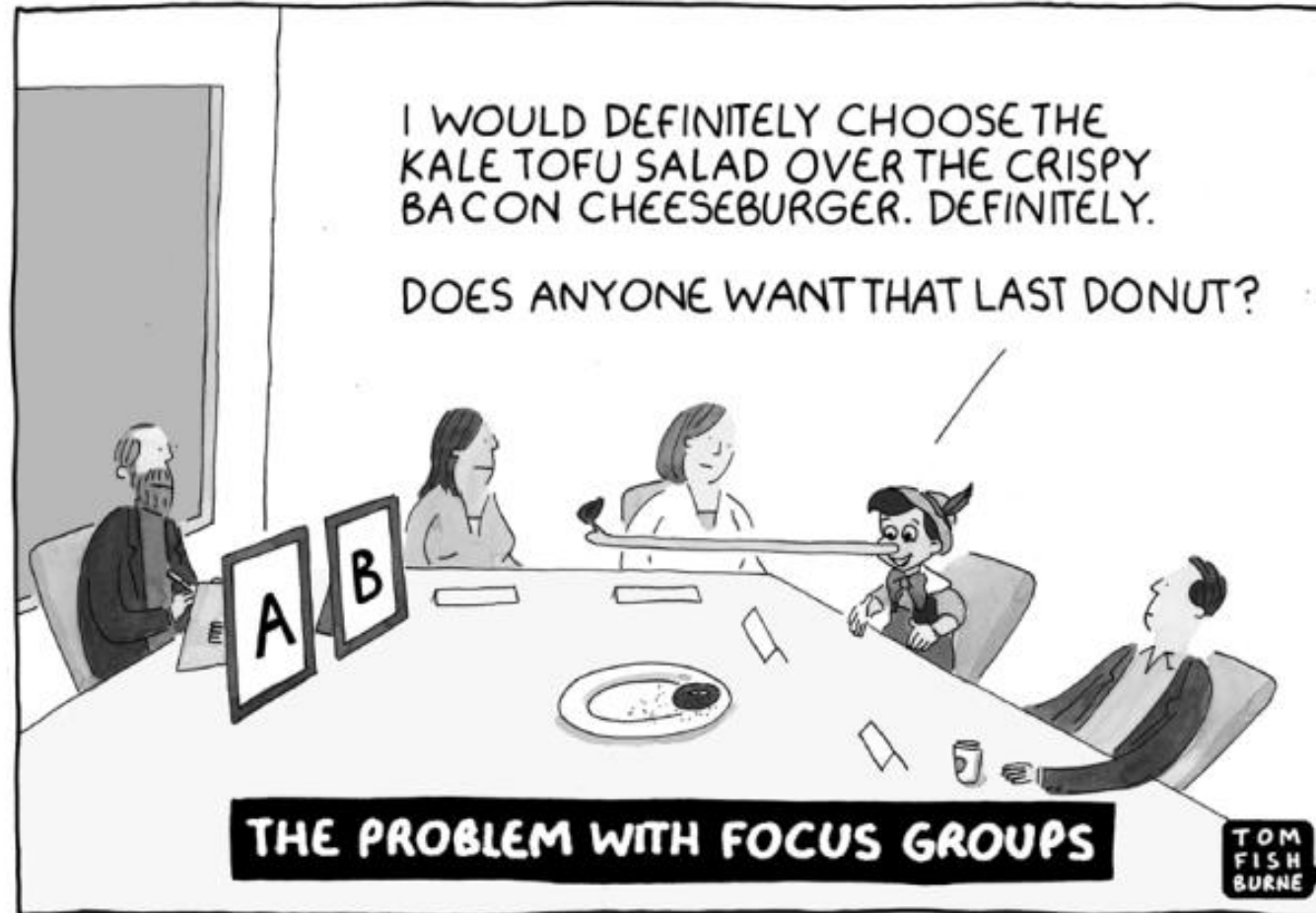


Pontificia Universidad
JAVERIANA
Bogotá

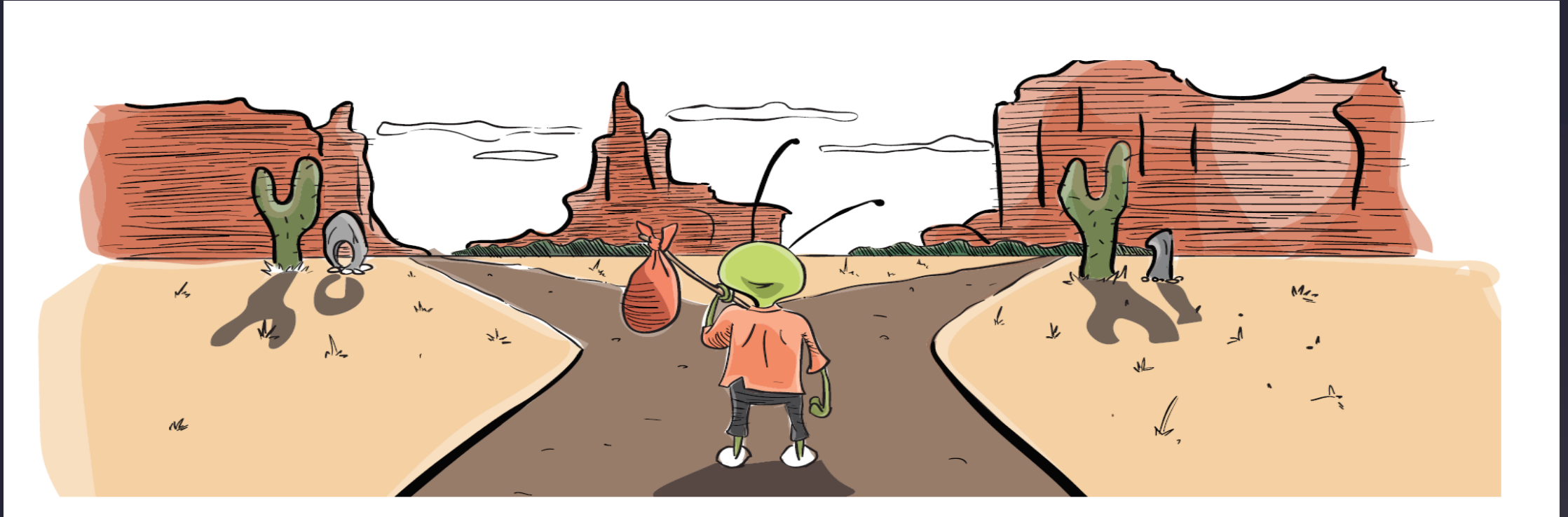
Profesor: Nicolás Velásquez

El problema de las encuestas y grupos focales

The saying-doing gap in focus groups



¿Si elegimos un camino, cómo sabemos que fue la mejor opción?



www.menti.com code: 2468 8322

Los experimentos como herramienta para evaluar las relaciones causales.

- Los experimentos son una buena alternativa para crear **contrafactuales** adecuados.
- Los experimentos nos permiten buscar datos que no están disponibles sobre el terreno.
- Los experimentos (bien diseñados) proporcionan efectos causales.

El problema del análisis de datos: la causalidad

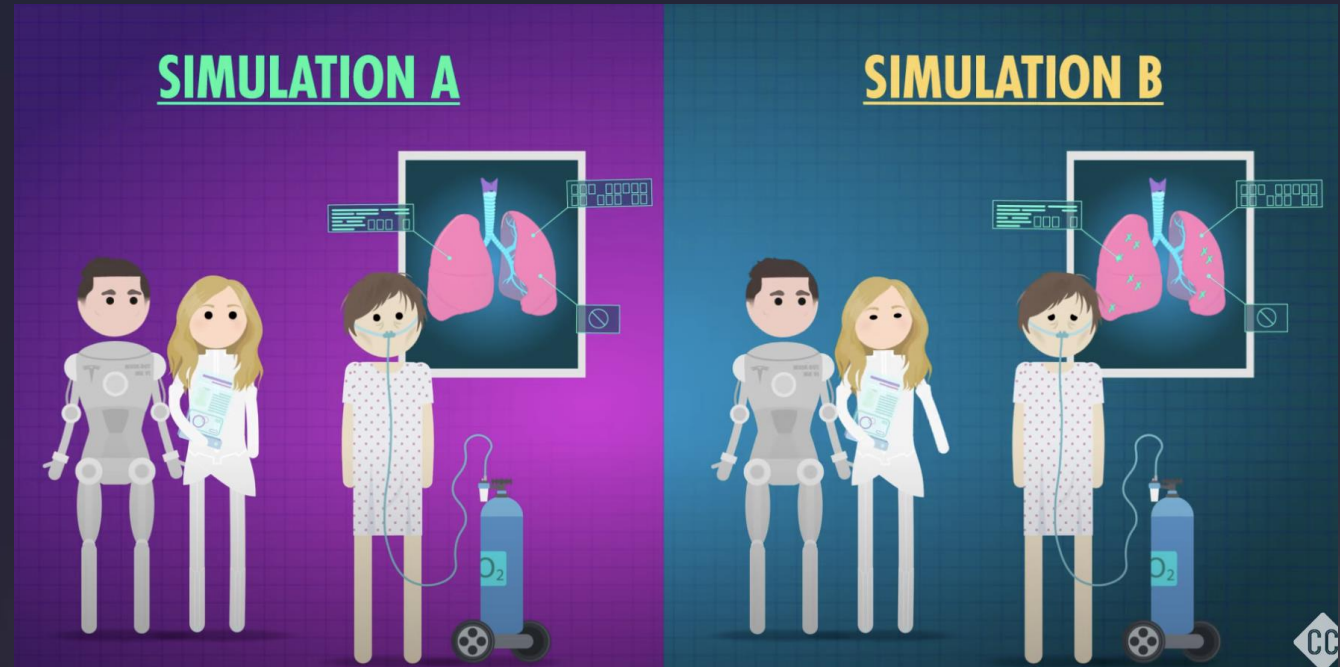
The problem with correlation



El problema del análisis de datos: la causalidad

https://www.youtube.com/watch?v=DwzqSp2rMI8&ab_channel=ProductivityGuy

https://www.youtube.com/watch?v=kkBDa-ICvyY&ab_channel=CrashCourse



¿Qué es un experimento?

- Un proceso mediante el cual se testea una hipótesis.
- Ejemplos de hipótesis que se pueden testear con experimentos:
 - La hidroxicloroquina es un medicamento efectivo contra el coronavirus.
 - Abordar aviones de pasajeros comenzando por los asientos de atrás y hacia adelante es más eficiente que hacerlo sin orden.
 - Organizar reuniones de brainstorming para innovar en procesos de la empresa, hace sus procesos más eficientes.
 - Pagar por anunciar algún producto en YouTube es beneficioso para las ventas.
 - Un nuevo elemento en la página web hace que las personas sean más propensas a comprar un producto.

Plantear la hipótesis

- El primero paso es **plantear la hipótesis** que queremos testear.
- Debemos **especificar la variable que analizaremos para saber si la hipótesis se cumple** o no, y que dicha variable es la adecuada.

En ese sentido, los ejemplos anteriores necesitan todavía un poco de trabajo...

Determinar variable(s) a medir

“La hidroxiclороquina es un medicamento efectivo contra el nuevo coronavirus”

Algunas alternativas:

- Duración de la enfermedad desde que se presentaron síntomas hasta que desaparecen.
- % que progresa a estado grave (hospital).
- % que necesita oxígeno.

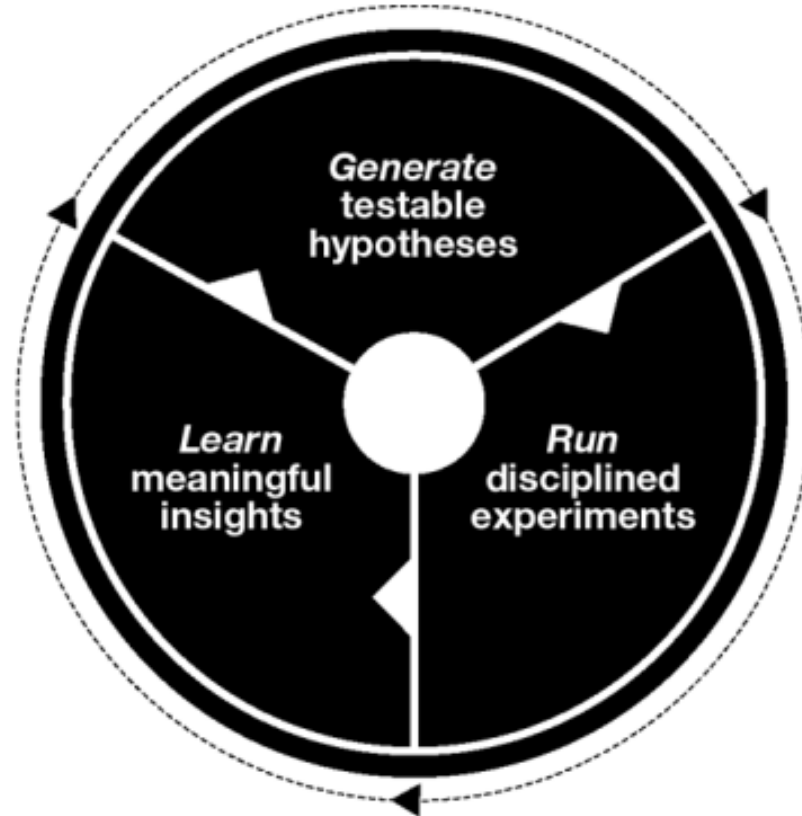
Determinar variable(s) a medir

“Organizar reuniones semanales de 1 hora de brainstorming para innovar en procesos de la empresa, hace sus procesos más eficientes.

Algunas alternativas:

- Duración de ciertos/todos los procesos.
- Pasos en el proceso.
- Procesos que utilizan tecnologías de la información.

The experimentation wheel



Generate Testable Hypotheses

Thomke, S. H. (2020). *Experimentation works: The surprising power of business experiments*. Harvard Business Press.

Control y tratamiento: ¿Qué son?

- Todo experimento debe tener **un control y al menos, un tratamiento.**

Control: grupo en el que no se interviene.

Tratamiento(s): grupo(s) en los que se interviene/se hace algún cambio.

¿Por qué es necesario tener un grupo de control?

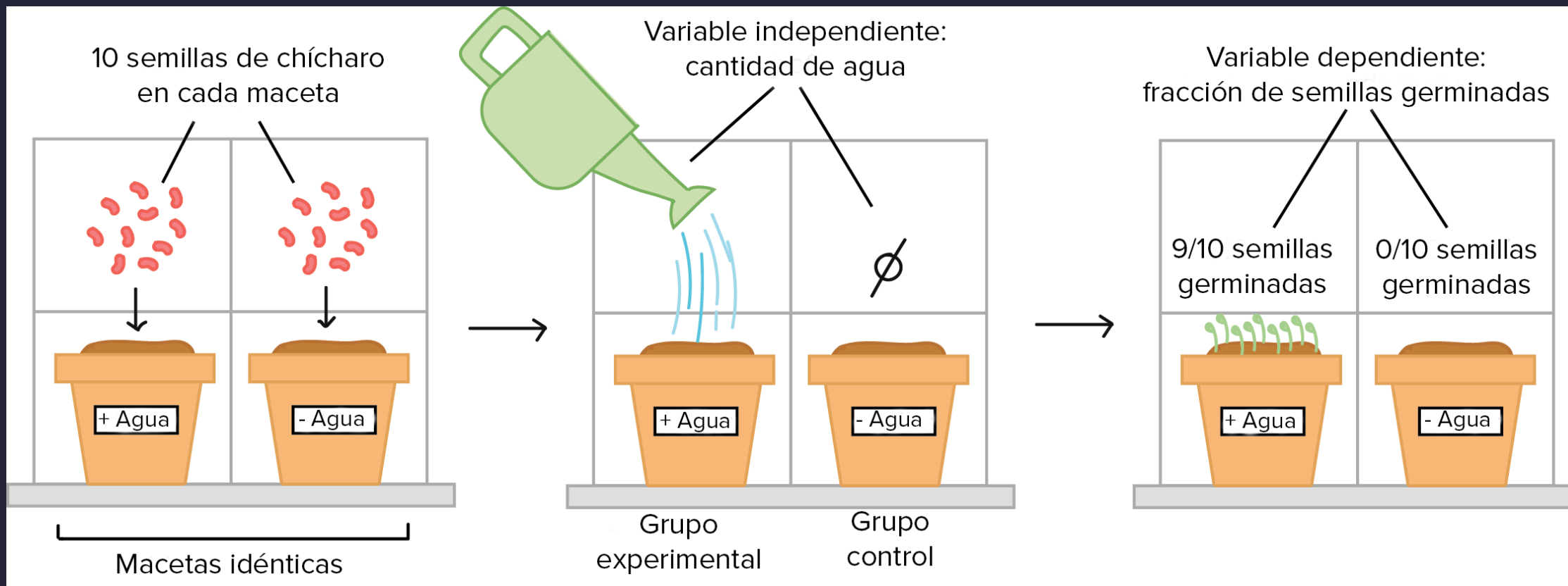
CONTROL GROUP



EXPERIMENTAL GROUP



Experiment: Will student test scores be affected by distracting sounds in the testing environment?



- Al momento de comprar control y tratamiento, **eliminamos** (“controlamos”) **cualquier otro factor que pudiese haber influido en nuestra variable de interés** y que no se deba a nuestra intervención / cambio.
- Estos otros factores deberían influenciar a grupos control y tratamiento por igual.
- **Es decir, el control nos permite determinar si la intervención verdaderamente CAUSÓ un cambio/efecto en la variable de interés.**

¿Quién queda en el control y quién queda en el/los tratamientos?

- Supongamos que tenemos 1 control y n tratamientos (donde n puede ser 1,2,...)
- Control y tratamientos serían (normalmente) de tamaño $1/(n+1)$ por las personas a estudiar.

Ejemplo:

1000 personas en mi organización que participan en el experimento (ya sea como control, tratamiento 1 o tratamiento 2).

Entonces $1/3$ de las 1000 personas (333) entrarían al control, $1/3$ (333) al tratamiento 1 y $1/3$ (334) al tratamiento 2.

¿Quién queda en el control y quién queda en el/los tratamientos?

Sigamos con nuestro ejemplo de 1 control y 2 tratamientos.

Dos alternativas para realizar la asignación:

Asignación aleatoria simple: pido a Excel que de la lista de personas me seleccione a $1/3$ (333). Esos 333 quedan en el control. Lo mismo para los dos tratamientos (eliminando las personas que ya han sido asignadas).

Asignación aleatoria estratificada: si yo se, por ejemplo, que el 80% de los 1000 son hombres y el 20% son mujeres, entonces me tengo que asegurar que en cada grupo quedan más o menos 80% de hombres y 20% de mujeres.

¿Quién queda en el control y quién queda en el/los tratamientos?

Asignación aleatoria estratificada: si yo se, por ejemplo, que el 80% de los 1000 son hombres y el 20% son mujeres, entonces me tengo que asegurar que en cada grupo quedan más o menos 80% de hombres y 20% de mujeres.

- Control y tratamientos tienen que seguir siendo del mismo tamaño (333), pero con 80% hombres y 20% mujeres.
- Es decir, en cada grupo tendré que tener $333 \times 0.8 = 266$ hombres y el resto, 67, mujeres.
- Pido a Excel que de la lista de 1000 personas seleccione a 266 hombres y esos los asigno al control....

¿Quién queda en el control y quién queda en el/los tratamientos?

- Es fundamental, para asegurarnos que cualquier diferencia entre control y tratamiento (en la variable de interés) se deba a nuestra intervención que la asignación sea aleatoria.
- De lo contrario, si observo una diferencia entre control y tratamiento, esto se podría deber a...

¿Cómo determino que el control es distinto al (a los) tratamiento(s)?

- En otras palabras, ¿cómo determinamos que hay un efecto de mi intervención en la variable de interés?
- Cuando queremos saber si dos grupos son distintos en cuanto a alguna variable, hacemos...

¿Cómo determino que el control es distinto al (a los) tratamiento(s)?

- Cuando queremos saber si dos grupos son distintos en cuanto a alguna variable, hacemos...

PRUEBA DE HIPOTESIS DE MEDIAS POBLACIONALES

- Si p-value es menor al límite de riesgo (1%, 5% o 10%), entonces puedo decir que ambos grupos son distintos en cuanto a esa variable.
- Puedo hacer varias comparaciones, si hay mas de 1 tratamiento, pero siempre entre 2 grupos: control vs tratamiento 1, control vs tratamiento 2, tratamiento 1 vs tratamiento 2, etc...

Algunos puntos importantes...

- Los experimentos nos permiten establecer **causalidad**.
 - Implemento acción, estrategia, intervención y observo que las personas/divisiones/municipios... se comportan de manera distinta a las del control cuando se les aplica la intervención (importancia del control para establecer causalidad).
 - Entonces la intervención está causando el cambio en comportamiento. No es una simple correlación.

Algunos puntos importantes...

- Mínimo 30-40 personas por grupo (prueba de hipótesis).
- Tener cuidado con:
 - Efecto contagio
 - Efecto demanda del experimentador
 - Cuestiones éticas

Ética en los experimentos

La regla es simple: hay que cuidar la integridad física y psicológica de las personas, no hacerlas correr riesgos más allá de los normales por participar en el experimento.

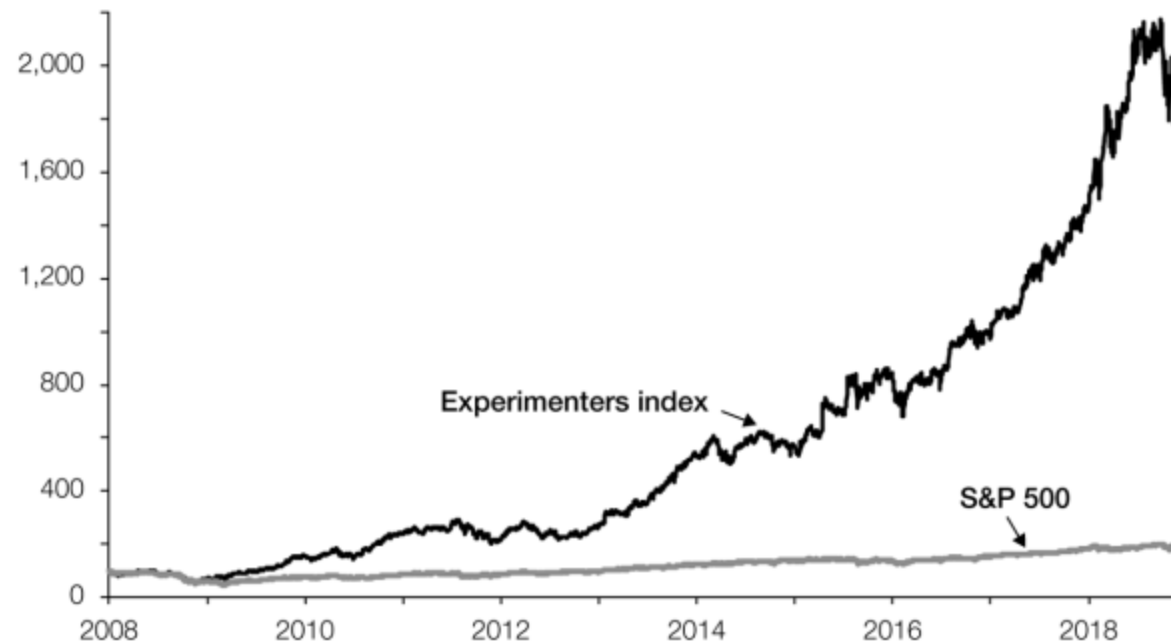
¿Por qué usar experimentos en una empresa?

- Las ideas verdaderamente innovadoras van en contra de nuestra experiencia y nuestras suposiciones, o de la sabiduría convencional.
- Ya se trate de mejorar la experiencia del cliente, probar nuevos modelos o el desarrollo de nuevos productos y servicios, incluso los gerentes más experimentados a menudo se equivocan, les guste o no.
- *"Vemos pruebas todos los días que la gente es terrible adivinando. Nuestras predicciones sobre el comportamiento de los clientes se equivocan nueve de cada diez veces".*
 - CEO Booking

¿Por qué usar experimentos en una empresa?

Stock performance of leading experimenters (January 2, 2008 = 100)*

An equally weighted index consisting of Amazon, ETSY, Facebook, Google, Microsoft, Netflix, and Booking Holdings. These companies have spent years building infrastructures and cultures for large-scale experimentation.



Source: Bloomberg 2019.

***Of course, correlation isn't causation—share prices are the result of many factors, and the sample is limited to public companies. But we should still**

Thomke, S. H.
(2020). *Experimentation works: The surprising power of business experiments*. Harvard Business Press.

Predicciones célebres sobre el comportamiento de los clientes

"[El iPhone es] el teléfono más caro del mundo, y no atrae a los clientes empresariales porque no tiene teclado, lo que hace que no sea una buena máquina de correo electrónico".

-Steve Ballmer, consejero delegado de Microsoft (2007)

"La gente nos ha dicho una y otra vez que no quiere alquilar su música... no quiere suscripciones".

-Steve Jobs, consejero delegado de Apple (2003)

"La televisión no podrá retener ningún mercado que capte después de los seis primeros meses. La gente pronto se cansará de mirar una caja de contrachapado todas las noches".

-Atribuido al jefe de los estudios 20th Century Fox, Darryl F. Zanuck
(1946)

HiPPOs at work



El reto de la innovación

- La idea es que, al igual que en una fábrica, la variabilidad y la incertidumbre son indeseables y deben eliminarse.
- **Aquí radica el dilema:** por definición, la novedad crea incertidumbre porque no sabemos qué funcionará y qué no. Dicho de otro modo la incertidumbre es necesaria porque crea oportunidades.
- La solución es complementar el análisis de datos con los business experiments.

Preguntas para realizar buenos experimentos empresariales

Hipótesis

- ¿La hipótesis se basa en observaciones, ideas o datos?
- ¿Se centra el experimento en una medida de gestión comprobable?
- ¿Tiene variables medibles y puede demostrarse su falsedad?
- ¿Qué se espera aprender del experimento?

Adhesión

- ¿Qué cambios concretos se introducirían en función de los resultados?
- ¿Cómo garantizará la organización que no se ignoren los resultados?
- ¿Cómo encaja el experimento en el programa general de aprendizaje y las prioridades estratégicas de la organización?

Preguntas para realizar buenos experimentos empresariales

Viabilidad

- ¿Tiene el experimento una predicción comprobable?
- ¿Cuál es el tamaño de muestra necesario? Nota: El tamaño de la muestra dependerá del efecto esperado (por ejemplo, un aumento del 5% en las ventas).
- ¿Puede la organización llevar a cabo el experimento en los lugares de prueba durante el tiempo necesario?

Fiabilidad

- ¿Coinciden las características del grupo de control con las del grupo de tratamiento?
- ¿Puede realizarse el experimento a "ciegas" o a "doble ciegas"?
- ¿Se han eliminado los sesgos restantes mediante análisis estadísticos u otras técnicas?
- ¿Otros que realicen la misma prueba obtendrían resultados similares?

Preguntas para realizar buenos experimentos empresariales

Causalidad

- ¿Hemos captado todas las variables que podrían influir en nuestros parámetros?
- ¿Podemos relacionar intervenciones específicas con el efecto observado?
- ¿Cuál es la solidez de las pruebas?
- ¿Nos sentimos cómodos actuando sin pruebas de causalidad?

Valor

- ¿Se ha planteado la organización un despliegue selectivo, es decir, uno que tenga en cuenta el efecto de una iniciativa propuesta en diferentes clientes, mercados y segmentos, para concentrar las inversiones en áreas en las que la rentabilidad potencial es mayor?
- ¿Ha implantado la organización sólo los componentes de una iniciativa con el mayor retorno de la inversión?
- ¿Conoce mejor la organización qué variables causan qué efectos?

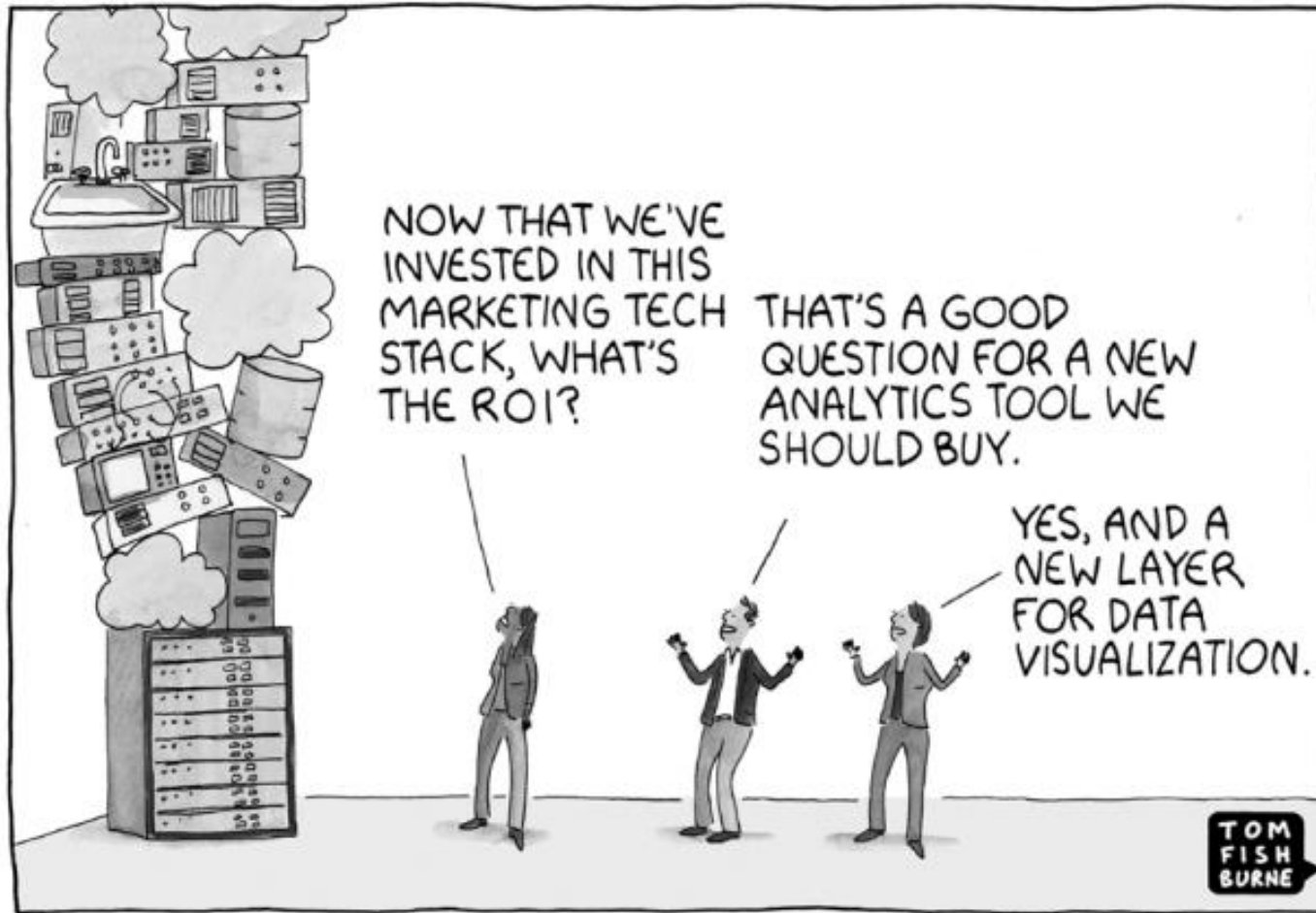
Preguntas para realizar buenos experimentos empresariales

Decisiones

¿Reconocemos que no todas las decisiones empresariales pueden o deben ser resueltas mediante experimentos, sino que debe probarse todo lo que pueda probarse?

¿Estamos utilizando pruebas experimentales para añadir transparencia a nuestro proceso de toma de decisiones?

Tools don't automatically improve performance



Algunos Ejemplos sencillos Pruebas A/B

Longer-headline experiment

A. Control (existing display)

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

bing
MS Beta

flowers

358,000,000 RESULTS

Flowers at 1-800-FLOWERS® Ads
1800Flowers.com
Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

FTD® - Flowers
www.FTD.com
Get Same Day Flowers in Hours! Buy Now for 25% Off Best Sellers.

Send Flowers from \$19.99
www.ProFlowers.com
Send Roses, Tulips & Other Flowers. "Best Value" -Wall Street Journal.
proflowers.com is rated ★★★★★ on Bizrate (1307 reviews)

50% Off All Flowers
www.BloomsToday.com
All Flowers on the Site are 50% Off. Take Advantage and Buy Today!

B. Treatment (new idea called "Long Ad Titles")

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

bing
MS Beta

flowers

358,000,000 RESULTS

FTD® - Flowers - Get Same Day Flowers in Hours! Ads
www.FTD.com
Buy Now for 25% Off Best Sellers.

Flowers at 1-800-FLOWERS® | 1800flowers.com
1800Flowers.com
Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

Send Flowers from \$19.99 - Send Roses, Tulips & Other Flowers
www.ProFlowers.com
"Best Value" -Wall Street Journal.
proflowers.com is rated ★★★★★ on Bizrate (1307 reviews)

\$19.99 - Cheap Flowers - Delivery Today By A Local Florist!
www.FromYouFlowers.com

The diagram illustrates the experimental design. In the Control section (A), four ads are shown. In the Treatment section (B), the same four ads are shown but with longer, underlined titles. Arrows indicate the mapping: from the underlined title 'Get Same Day Flowers in Hours!' in the Control FTD ad to the new title 'Get Same Day Flowers in Hours!' in the Treatment FTD ad; from the underlined title 'Send Roses, Tulips & Other Flowers' in the Control ProFlowers ad to the new title 'Send Roses, Tulips & Other Flowers' in the Treatment ProFlowers ad; from the underlined title 'Send Flowers from \$19.99' in the Control ProFlowers ad to the new title 'Send Flowers from \$19.99' in the Treatment ProFlowers ad; and from the underlined title '50% Off All Flowers' in the Control BloomsToday ad to the new title '\$19.99 - Cheap Flowers - Delivery Today By A Local Florist!' in the Treatment FromYouFlowers ad.

Source: R. Kohavi, D. Tang, Y. Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge, UK: Cambridge University Press, in press).

Example of A/B test (Booking)

A. Control (shows current practice)			B. Treatment (adds checkout date above children's ages)		
Rooms 1 ▾	Adults 2 ▾	Children 2 ▾	Rooms 1 ▾	Adults 2 ▾	Children 2 ▾
Ages of children at checkout 4 ▾ 7 ▾			Children's ages on Jul 23, 2016 4 ▾ 7 ▾		

Source: S. Thomke and D. Beyersdorfer, “Booking.com,” Harvard Business School Case 619-015 (Boston: Harvard Business School Publishing, 2018).

The ad link experiment

A. Control

Esurance® Auto Insurance - You Could Save 28% with Esurance. www.esurance.com/California Get Your Free Online Quote Today!	Ads
--	-----

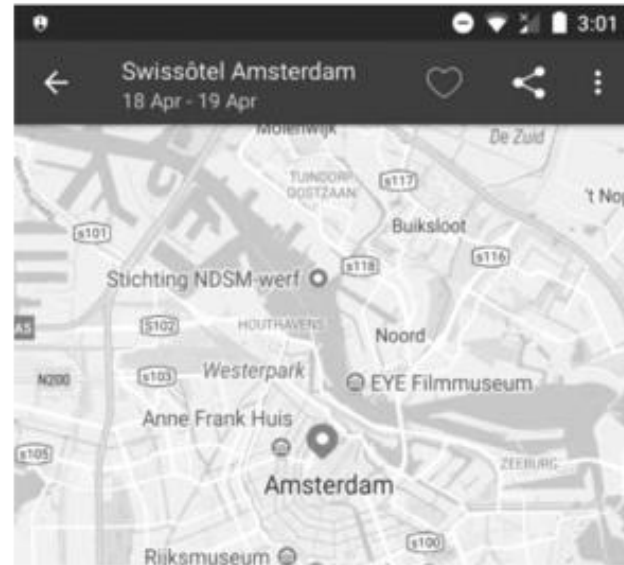
B. Treatment with site links at the bottom

Esurance® Auto Insurance - You Could Save 28% with Esurance. www.esurance.com/California Get Your Free Online Quote Today! Get a Quote · Find Discounts · An Allstate Company · Compare Rates	Ads
--	-----

Source: R. Kohavi et al., “Online Controlled Experiments at Large Scale,” Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’13), Chicago, August 11–14, 2013, New York: ACM, 2013.

Testing a walkability assessment

A. Control
(shows current practice)



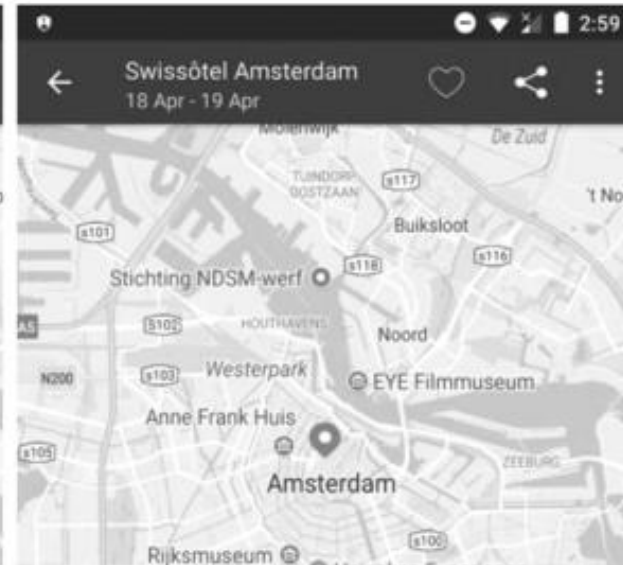
Swissôtel Amsterdam



9.5 - Exceptional location!

Based on 2966 reviews

B. Treatment
(adds walkability assessment)



Swissôtel Amsterdam



9.5 - Exceptional location!

Guests **loved walking around** in this neighborhood

Source: S. Thomke and D. Beyersdorfer, "Booking.com," Harvard Business School Case No. 619-015 (Boston: Harvard Business School Publishing, 2018).

¿Cómo crearían un experimento para resolver estas preguntas?

1. ¿Cuál es el impacto de cambiar el diseño de nuestra página de inicio en las tasas de conversión de ventas?
2. ¿Es más efectivo enviar correos electrónicos promocionales a nuestros clientes por la mañana o por la tarde?
3. ¿Cuál es la mejor ubicación para colocar un botón de llamado a la acción (CTA) en nuestro sitio web para aumentar las tasas de clics?
4. ¿Qué tipo de mensaje es más efectivo para motivar a los clientes a realizar una compra: ¿mensajes de urgencia, de escasez o de beneficios?
5. ¿Cuál es la longitud óptima de una descripción de producto que maximiza las tasas de conversión de ventas?

Does Management Matter? Evidence from India.

The Quarterly Journal of Economics

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2012).

Pregunta: ¿Pueden las diferencias en las prácticas de gestión (management) explicar las diferencias en la productividad de las empresas?

Puede parecer obvio, pero a pesar de la enorme literatura que se remonta al menos a Walker (1887), no hay consenso.

Identificación: Aleatorizar una intervención de gestión sobre un grupo de grandes empresas textiles indias.

Qué hicieron: realizar un experimento para evaluar el impacto de prácticas de gestión más modernas.

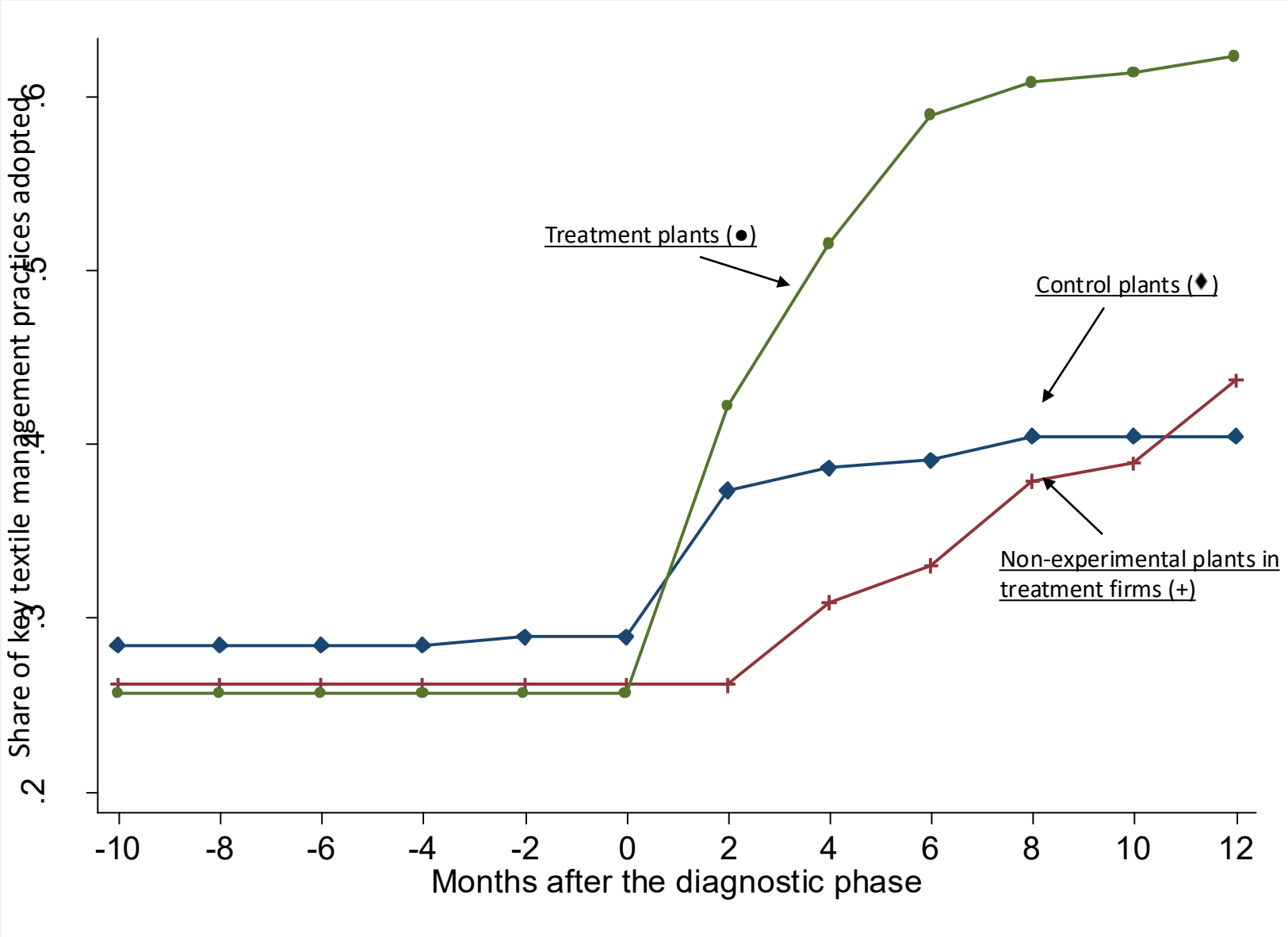
Experimento en 20 plantas propiedad de 17 grandes empresas (multiplanta con \approx 300 empleados cada una) cerca de Bombay que fabrican tejido de algodón.

Las plantas de tratamiento reciben cinco meses de una intervención estándar de consultoría de gestión, los controles reciben 1 mes.

Consultoría sobre 38 prácticas específicas relacionadas con las operaciones de la fábrica, el control de calidad e inventario y la gestión de recursos humanos.

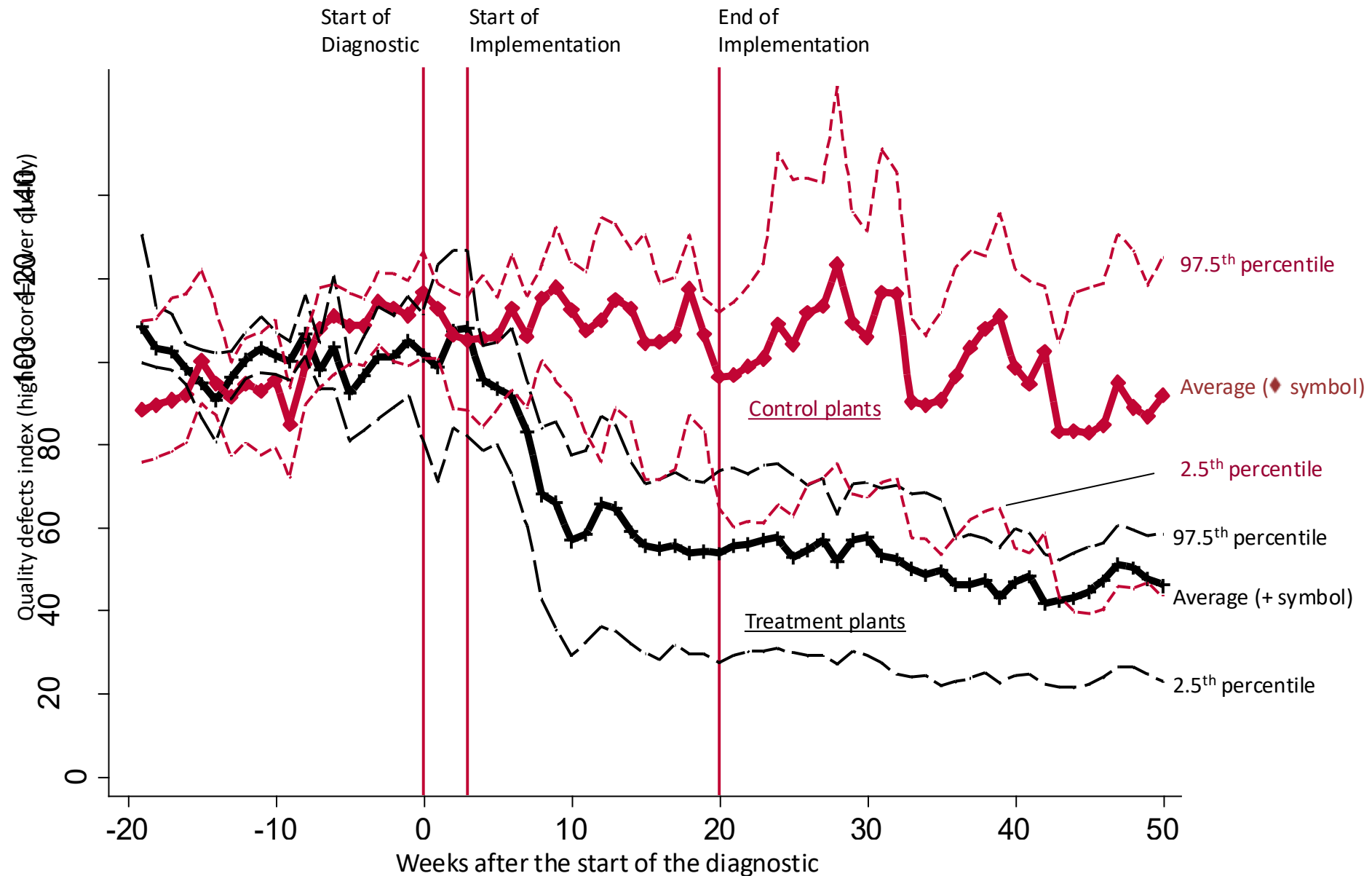
Recopilación de datos semanales en todas las plantas de 2008 a 2010.

Figure 2: The adoption of key textile management practices over time



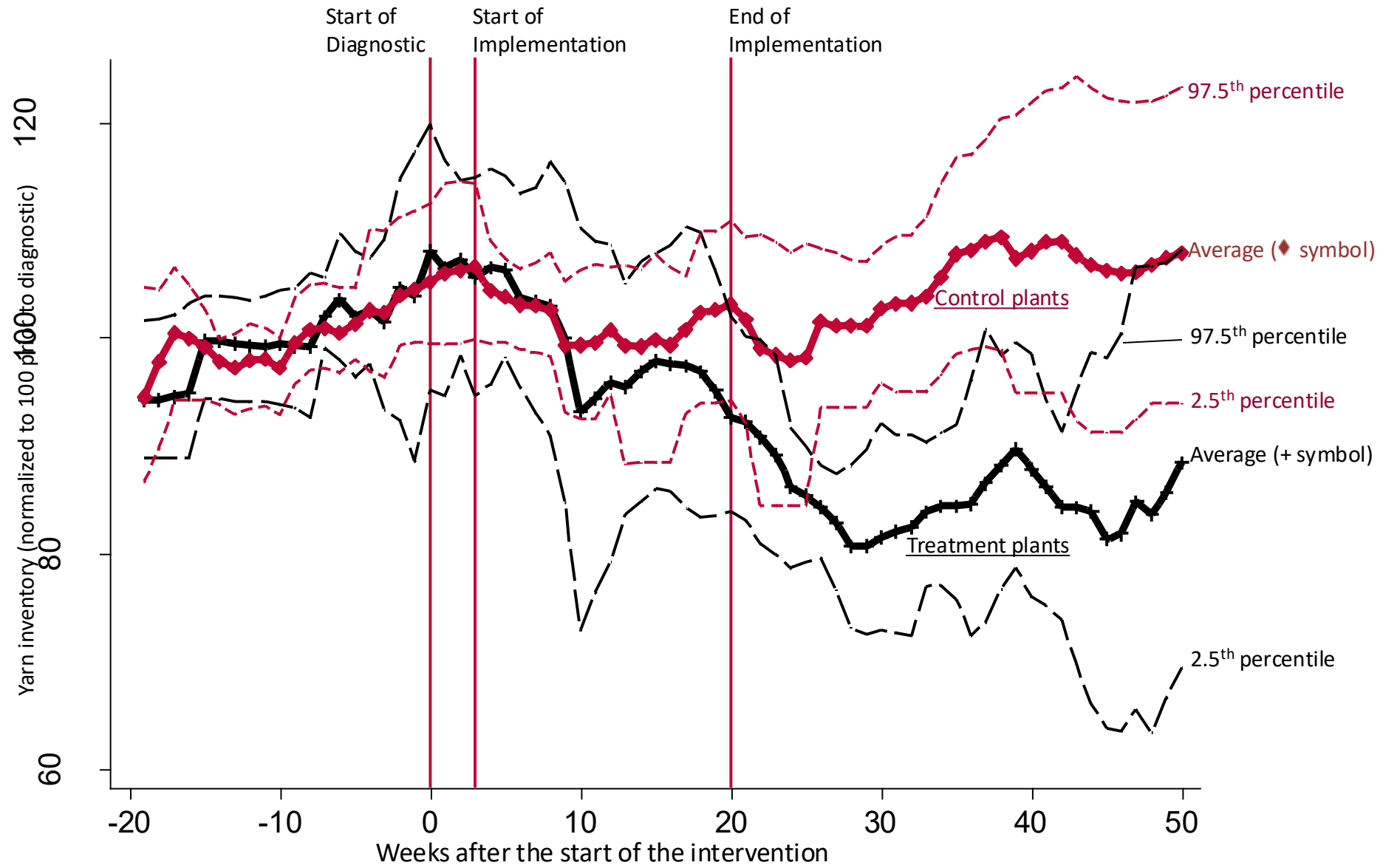
Notes: Average adoption rates of the 38 key textile manufacturing management practices listed in Table 2. Shown separately for the 14 treatment plants (round symbol), 6 control plants (diamond symbol) and the 5 non-experimental plants in the treatment firms which the consultants did not provide any direct consulting assistance to (+ symbol). Scores range from 0 (if none of the group of plants have adopted any of the 38 management practices) to 1 (if all of the group of plants have adopted all of the 38 management practices). Initial differences across all the groups are not statistically significant.

Figure 3: Quality defects index for the treatment and control plants



Notes: Displays the average weekly quality defects index, which is a weighted index of quality defects, so a higher score means lower quality. This is plotted for the 14 treatment plants (+ symbols) and the 6 control plants (♦ symbols). Values normalized so both series have an average of 100 prior to the start of the intervention. To obtain confidence intervals we bootstrapped the firms with replacement 250 times.

Figure 4: Yarn inventory for the treatment and control plants



Conclusiones

La gestión importa en las empresas indias: las prácticas más modernas tienen un gran impacto en la productividad y la rentabilidad.

Una de las principales razones de la mala gestión parece ser la falta de información, que la competencia limitada permite que persista.