**Ejecutar el experimento y recolectar datos**

* Correrlo hasta que **cada grupo** llegue al tamaño de muestra calculado.
* No mirar los resultados intermedios “con intención de parar si parece ganador” (evita el “peeking” que sesga los resultados).

 **Asignar aleatoriamente** cada lead a control (A) o tratamiento (B) y guardar su grupo (sticky, sin cambiar después).

 **Programar la llamada** según el grupo: A usa el horario actual, B usa el horario óptimo predicho por el modelo.

 **Llamar y registrar**: guardar hora programada, hora real, resultado (contacta o no), intentos y si se cumplió la ventana (AM, MD, PM).

 **Monitorear diariamente**: número de leads por grupo, % de contactos, desviaciones en la asignación (SRM), llamadas fuera de ventana, registros faltantes.

 **Mantener consistencia**: si se detectan errores de asignación, problemas de registro o sesgos, pausar y corregir.

 **Cerrar** cuando todos los leads hayan tenido oportunidad de responder dentro del periodo definido (ej. 72h).

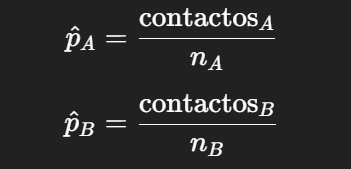
 **Analizar** con test estadístico (z-test o t-test) sobre la tasa de contacto de A vs. B y decidir si adoptar el horario predicho.

**7. es donde transformas los datos crudos de llamadas en evidencia estadística para decidir si el horario recomendado por tu modelo (Tratamiento B) es realmente mejor que el horario tradicional (Control A).**

**Preparar datos**

* **Dataset mínimo**:
  + lead\_id
  + grupo (A o B) — asignación inicial
  + resultado (1 si contestó, 0 si no)
  + hora\_llamada y ventana (AM, MD, PM) para QA
* **Filtros previos**:
  + Excluir leads que no recibieron llamada dentro de la ventana definida.
  + Excluir duplicados.
  + Confirmar que no hubo **SRM** (Sample Ratio Mismatch): verificar si el % asignado a cada grupo está cercano a lo planeado (ej. 50%-50%).

Calcular métricas primarias

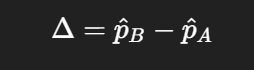


donde nAn\_AnA​ y nBn\_BnB​ son el número de leads únicos por grupo.

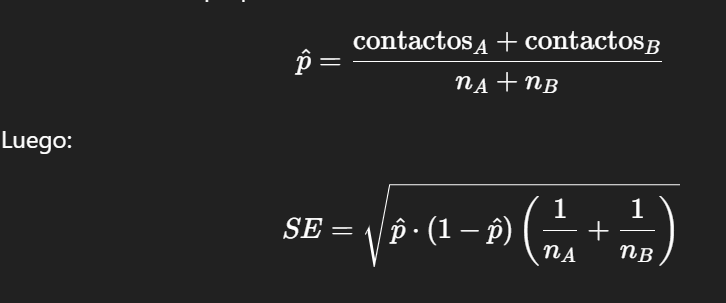
**alcular diferencia de proporciones**

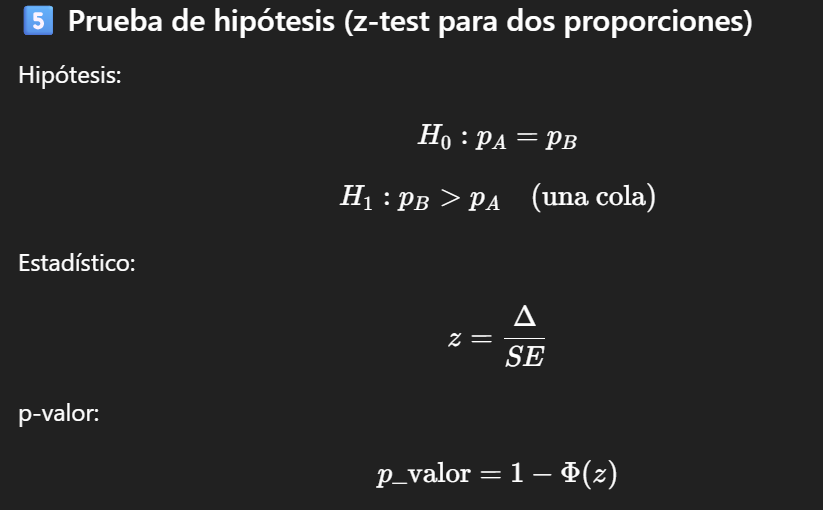
Δ=p^B−p^A\Delta = \hat{p}\_B - \hat{p}\_AΔ=p^​B​−p^​A​

Esto es la magnitud del efecto observado.



Calcular error estándar (SE)





Intervalo de confianza (IC) 95%

IC95%​=Δ±1.96⋅SE

**Decisión**

* Si p\_valor<αp\\_\text{valor} < \alphap\_valor<α (ej. 0.05) y Δ>0\Delta > 0Δ>0 → **B gana**.
* Si no, no se rechaza H0H\_0H0​ (no evidencia suficiente de mejora).

**Actualización del experimento**

En un test frecuentista clásico **no se actualiza durante la ejecución** para evitar error Tipo I inflado.

* Puedes monitorear métricas descriptivas, pero **sin hacer pruebas de significancia antes del tamaño final**.
* Si necesitas detener antes, debes aplicar correcciones como **Pocock** o **O'Brien-Fleming** para ajustar el α.

Paso 8

El **Paso 8: Intervalo / Credibilidad** es donde pones “barreras de confianza” alrededor de tu estimación, para no quedarte solo con el valor puntual (la diferencia observada entre A y B), sino también medir la **incertidumbre**.

Te lo explico en tus dos enfoques:

**1️⃣ Frecuentista – Intervalo de Confianza (IC)**

* **Objetivo:**  
  Estimar un rango donde probablemente se encuentra la verdadera diferencia pB−pAp\_B - p\_ApB​−pA​ en la población, con un cierto nivel de confianza (ej. 95%).
* **Fórmula para dos proporciones:**

IC95%=Δ±z1−α/2⋅SEIC\_{95\%} = \Delta \pm z\_{1-\alpha/2} \cdot SEIC95%​=Δ±z1−α/2​⋅SE

**Decisión**

Es el paso donde decides si tu modelo de horario (Tratamiento B) es **mejor que el horario tradicional (Control A)**.

**Frecuentista**

* Basado en p-valor e IC:
  1. Revisa si p\_valor<αp\\_\text{valor} < \alphap\_valor<α (ej. 0.05).
  2. Revisa si la diferencia observada Δ=p^B−p^A>0\Delta = \hat{p}\_B - \hat{p}\_A > 0Δ=p^​B​−p^​A​>0.
* **Reglas:**
  1. Si ambos se cumplen → **B gana** (modelo recomienda mejor hora).
  2. Si no → no hay evidencia de mejora.
* IC también ayuda a ver magnitud: si el IC incluye 0 → efecto no significativo.

10

**🔟 Duración / cuándo parar**

Determina cuánto tiempo o cuántos leads debes observar antes de tomar la decisión.

**Frecuentista**

* Basado en **tamaño de muestra calculado**:
  + Fórmula de dos proporciones con α y β deseados.
  + Parar cuando se alcanzan nAn\_AnA​ y nBn\_BnB​.

**Frecuentista**

* statsmodels → para z-test de proporciones, calcular IC y tamaño de muestra.

**1️⃣ Fase de planificación (antes de ejecutar el experimento)**

* Determinas **α** (nivel de significancia, ej. 0.05) y **β** (poder, ej. 0.8).
* Estimas la **proporción esperada de éxito** en Control (pAp\_ApA​) y la mejora mínima que consideras relevante (δ=pB−pA\delta = p\_B - p\_Aδ=pB​−pA​).
* Con esto, calculas **número de clientes por grupo**:

n=2(z1−α/2+z1−β)2 p(1−p)δ2n = 2 \frac{(z\_{1-\alpha/2}+z\_{1-\beta})^2 \, p(1-p)}{\delta^2}n=2δ2(z1−α/2​+z1−β​)2p(1−p)​

* Luego, **duración estimada** = número de clientes / leads disponibles por día.

**2️⃣ Durante el experimento**

* Monitorizas el progreso, pero **no decides antes de tiempo** si se cumple el tamaño de muestra (para evitar inflar el error tipo I).