MD004 Test de Hipótesis

Máster universitario en Ciencia de los Datos / Data Science (MUDS)

Curso académico 2025/26
11 de noviembre de 2025

Xavier Vilasís
Gloria Aguilera
Ricard Sierra
Noelia Lage

INDICE

- Introducción al pensamiento inferencial
- Definiciones clave: H0, H1, errores I y II, p-valor
- Selección del estadístico de prueba
- Ejemplo práctico
- Diseño de un test de hipótesis
- Matriz de confusión y curva ROC
- Conceptos avanzados

Introducción al pensamiento inferencial

¿Dormir más horas mejora mi productividad?

¿El ejercicio matutino mejora mi nivel de energía durante el día?

¿Los niveles hormonales podrían afectar la eficacia de un fármaco?

¿Existe una diferencia significativa en la eficacia del tratamiento entre hombres y mujeres?

¿Las reseñas positivas incrementan las ventas de productos en un e-commerce?

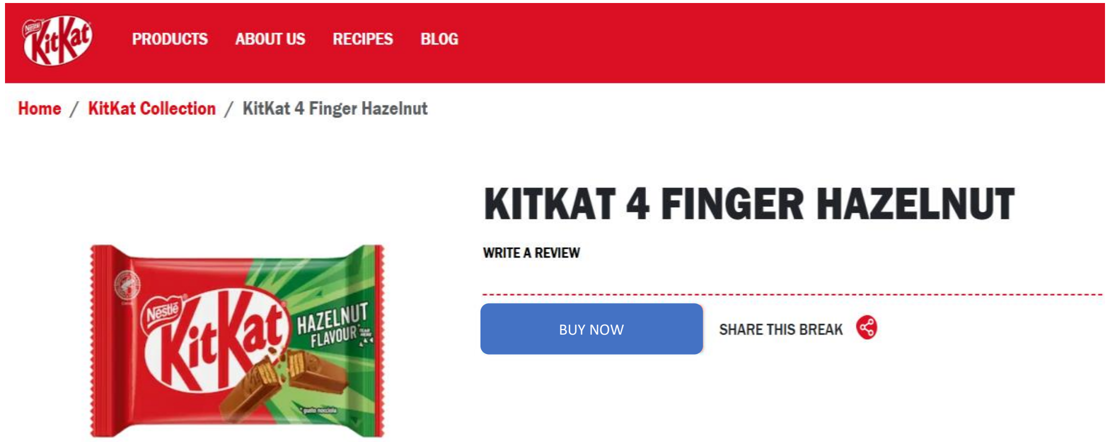
¿Cambiar el color de un botón CTA aumenta la tasa de conversión de ventas?

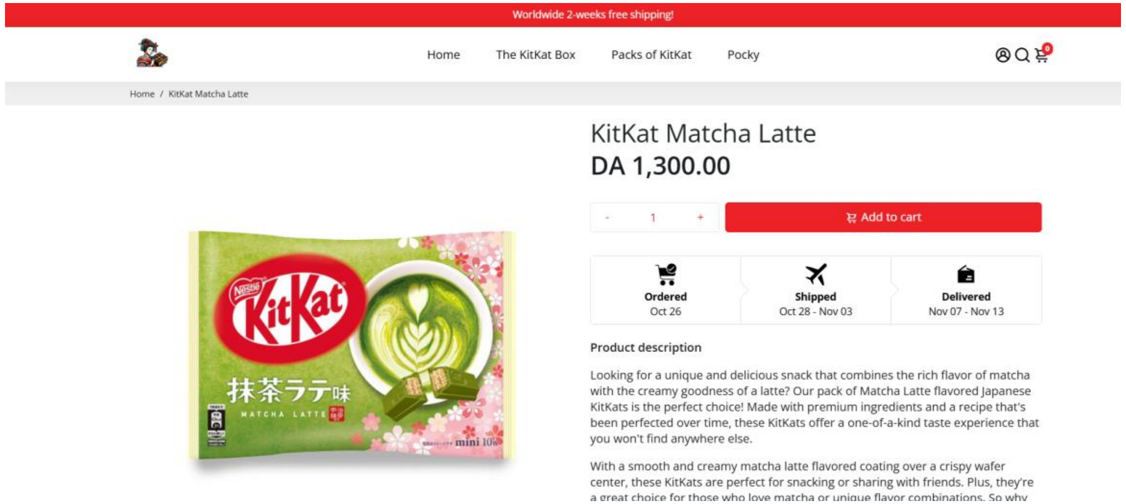
Fuente: link

Introducción al pensamiento inferencial

laSalle
Ramon Llull University

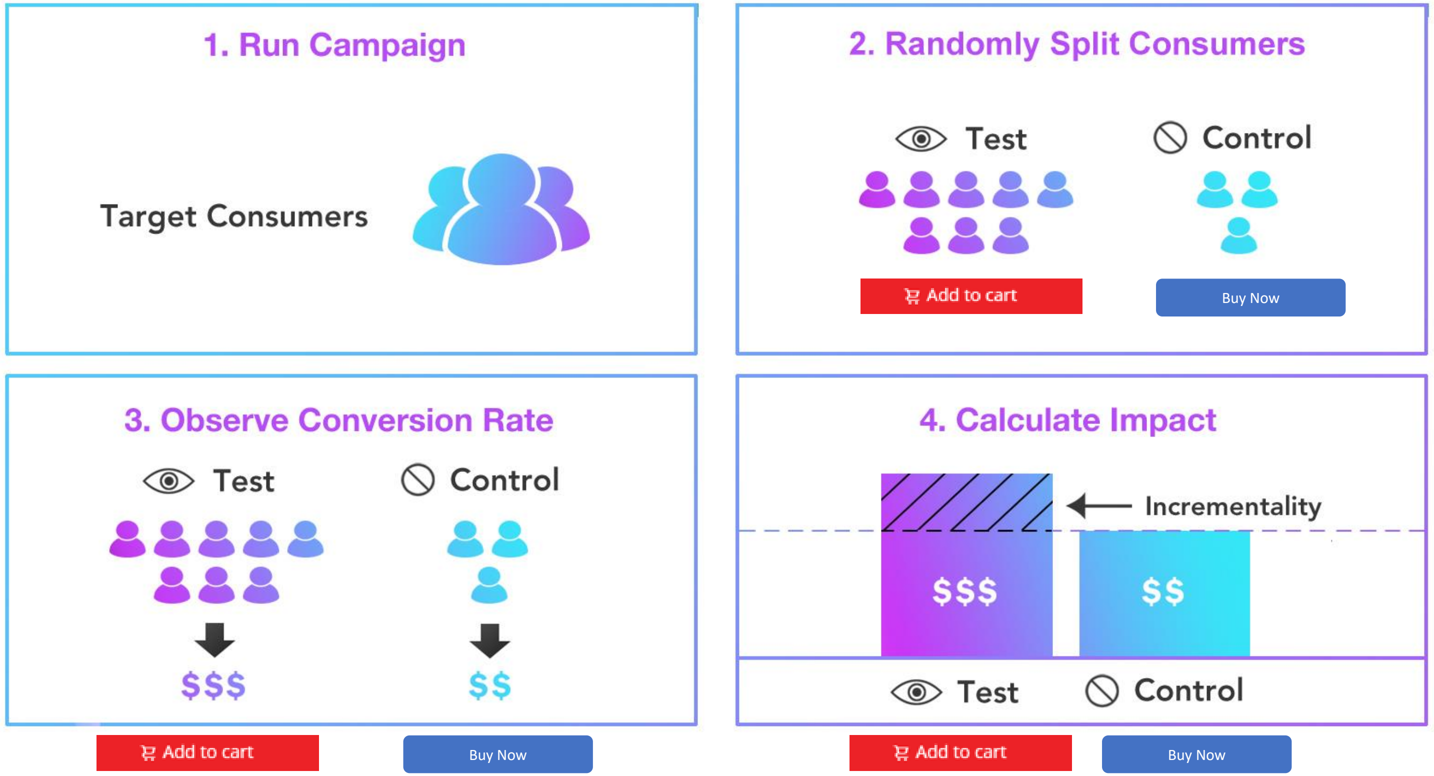
¿Cambiar el color de un botón CTA aumenta la tasa de conversión de ventas?





INTRODUCCIÓN

laSalle
Ramon Llull University

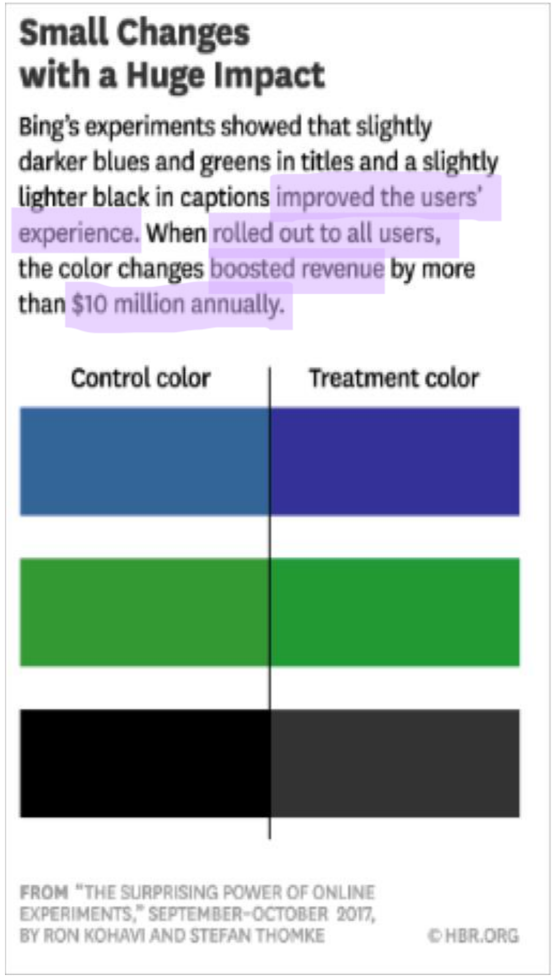


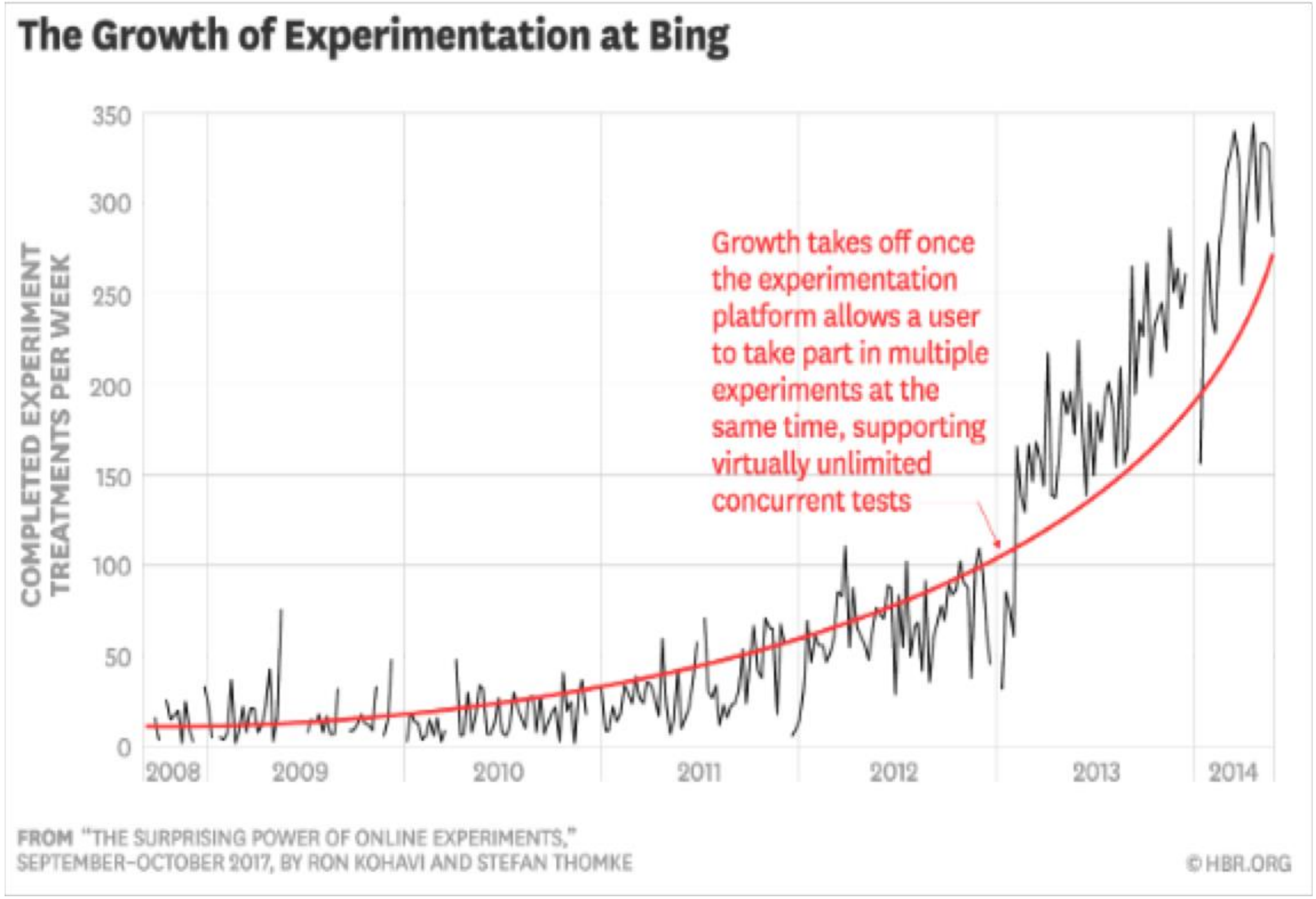
A partir de un estudio de una muestra de la población generalizamos acerca de una población.

Fuente: [link]

INTRODUCCIÓN

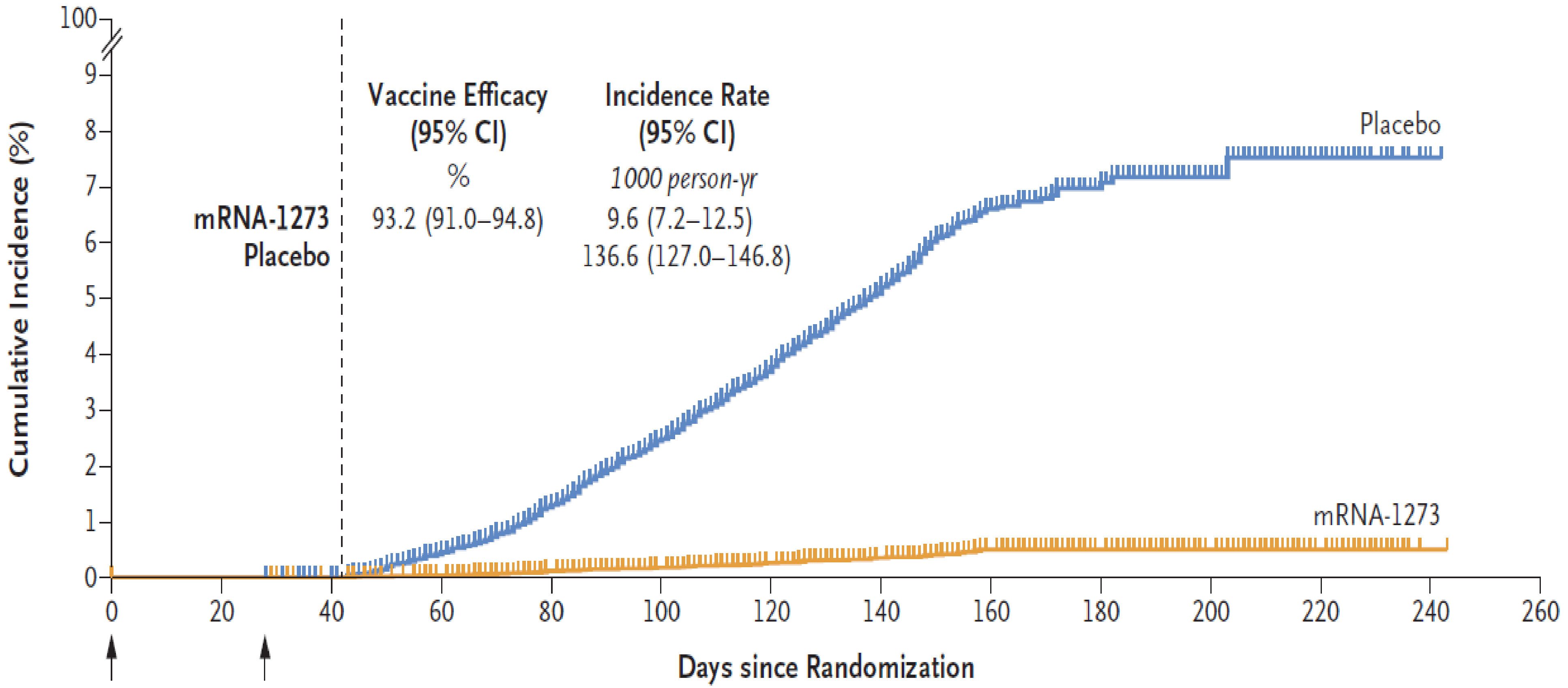
laSalle
Ramon Llull University





Fuente: link

INTRODUCCIÓN



No. at Risk

Placebo	14,164	14,164	14,134	13,030	13,733	12,970	11,199	7783	3323	953	336	64	5	0
mRNA-1273	14,287	14,287	14,281	14,246	14,096	13,584	12,196	9031	4252	1375	473	49	2	0

Fuente: link

INTRODUCCIÓN

Contexto	Grupo A (control)	Grupo B (Test)	Hipótesis
Marketing	Botón azul	Botón rojo	Ho: El cambio de color no tiene efecto en tasa de conversión H1: El cambio de color tiene efecto en tasa de conversión
Farmacéutica	Placebo	Vacuna	Ho: El tratamiento no reduce los contagios H1: El tratamiento reduce los contagios

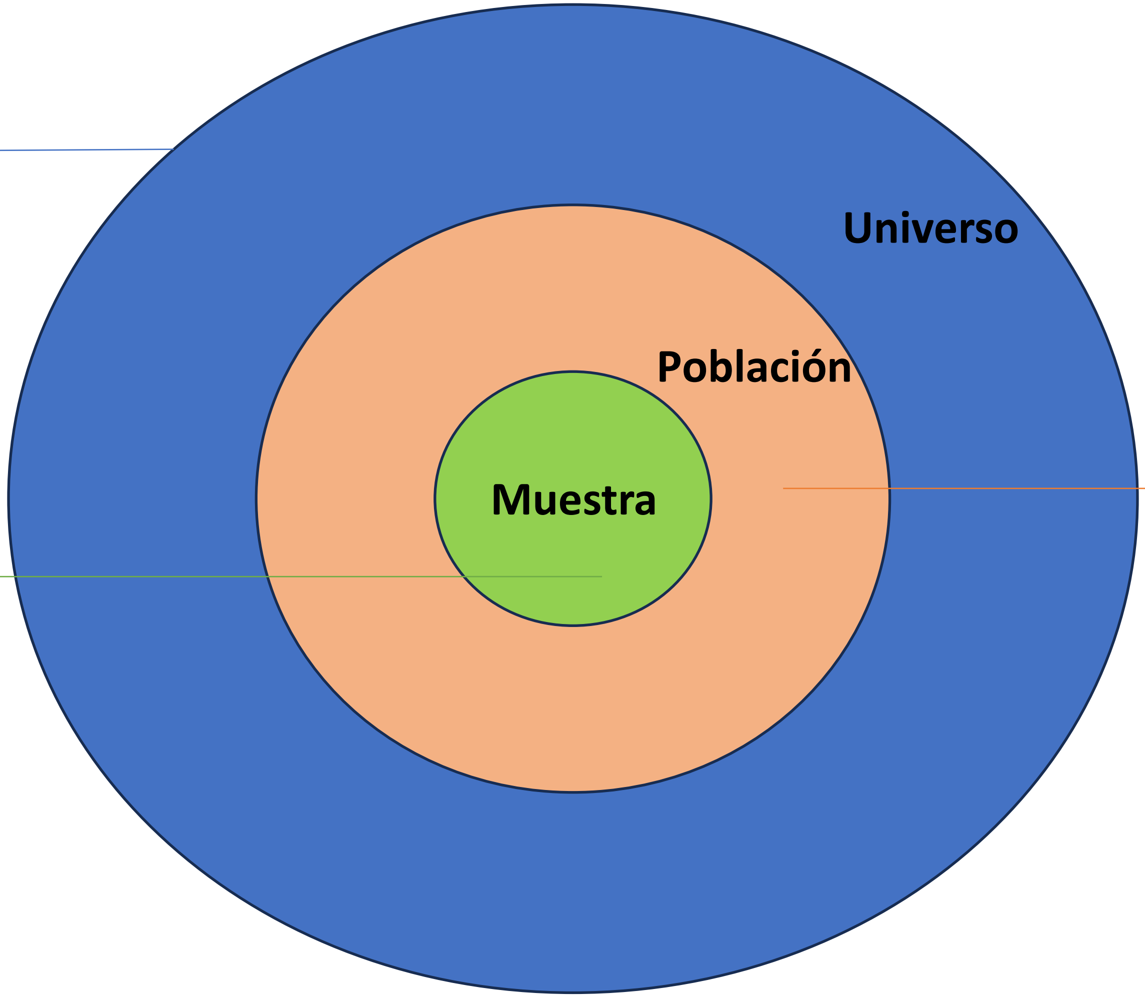
Test de contraste de hipótesis : es un método de inferencia estadística que, a partir de los datos muestrales, permite evaluar si las diferencias observadas proporcionan evidencia estadísticamente significativa para generalizar conclusiones a la población.

DEFINICIONES > TEST DE HIPÓTESIS

laSalle
Ramon Llull University

El **universo** es el conjunto **más amplio y teórico** de individuos o elementos sobre los que se podría aplicar el estudio o el medicamento.

Todas las personas del mundo *que podrían llegar a usar o necesitar este medicamento.*



La población es el conjunto **específico, real y actual** de personas o elementos sobre los que **queremos sacar conclusiones**, es decir, sobre los que deseamos inferir resultados.

Es un grupo más concreto y actual, formado por las personas que cumplen con las condiciones para tomar ese medicamento ahora mismo. Son las que realmente podrían beneficiarse del tratamiento hoy..

Subconjunto representativo de esa población que realmente participa en el estudio o experimento.

Las personas seleccionadas para el ensayo clínico *(por ejemplo, 500 pacientes que cumplen ciertos criterios).*

laSalle
Ramon Llull University

HIPÓTESIS	Supuesto sobre los parámetros de una o más **poblaciones**. (Supuestos sobre una población y no sobre una muestra)

H_0: Hipótesis Nula

Hipótesis que se desea contrastar. Representa la **situación base o sin efecto**. Se asume como **cierta** hasta que los **datos demuestren lo contrario.**

Ej: Ho: La tasa de conversión (o compra) es la misma entre los dos grupos (control: azul, test: rojo)

La tasa de incidencia de la enfermedad es la misma en los dos grupos (placebo, test)

H_1: Hipótesis Alternativa

Suposición **alternativa a la hipótesis nula** que queremos **contrastar** y **comprobar**. Sugiere que existe un **cambio o diferencia**. Es la conclusión que se busca confirmar tras el análisis estadístico.

H1: La tasa de conversión es mayor en el grupo test (rojo) que en el grupo control (azul).

La tasa de incidencia de la enfermedad es menor en el grupo test que en el grupo placebo (unilaterial)

DEFINICIONES > TEST DE HIPÓTESIS

laSalle
Ramon Llull University

Testeo de CTA

H_0: Hipótesis Nula: *La **media de ventas** es la misma en ambos grupos (control: azul, test: rojo).*

H_1: Hipótesis Alternativa: La **media de ventas** es **mayor en el grupo test (rojo)** que en el grupo control (azul).

Muestra:
N=100
- **Color azul (control):** versión actual o tradicional → 1 500 euros de media en ventas
- **Color rojo (experimental):** nueva versión →2 000 euros de media en ventas

- **H_0:** Cambio color CTA no aumenta ventas (μrojo=μazul)
- **H_1:** Cambio color CTA incrementa las ventas (μrojo > μazul)

Pregunta: ¿La diferencia de 500 € en las ventas medias entre los grupos test y control es **estadísticamente significativa**, o podría ser debida al azar?

DEFINICIONES > TEST DE HIPÓTESIS

laSalle
Ramon Llull University

Testeo de Fármacos

$\boldsymbol{H_0}$**: Hipótesis Nula:** El fármaco no tiene ningún efecto sobre tasa de contagio de los pacientes

$\boldsymbol{H_1}$**: Hipótesis Alternativa:** El fármaco mejora la tasa de contagio de los pacientes

Muestra:
N=2000
- Grupo sin fármaco: 1500 contagiados / 2000 → **75 %**
- Grupo con fármaco: 100 contagiados / 2000 → **5 %**

Queremos saber si esta diferencia del 75 % al 5 % es **estadísticamente significativa**.

H_0: El fármaco no reduce la tasa de contagio ($p_1 = p_2$)
H_1: El fármaco reduce la tasa de contagio ($p_1 > p_2$)

Pregunta: ¿Podemos afirmar, con evidencia estadística, que el fármaco **reduce realmente la tasa de contagio**, o esa diferencia podría ser solo **una casualidad** en la muestra?

DEFINICIONES > TEST DE HIPÓTESIS

laSalle
Ramon Llull University

TEST ESTADÍSTICO

Procedimiento que permite evaluar si la **diferencia observada en una muestra** es lo suficientemente **significativa** como para **inferirla a la población**. Nos ayuda a decidir si las diferencias observadas se deben al azar del muestreo o si reflejan un efecto real.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

df = n - 1

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

df=n1+n2-2

- Nos permite comparar lo que observamos en la muestra con lo que esperaríamos si la hipótesis nula (H_0) fuera cierta.
- Nos indica qué tan lejos está lo que observamos en la muestra (esa diferencia de 500 € entre grupos o tasa de contagio) de lo que esperaríamos si no existiera ningún efecto real (Ho cierta).
- Cuanto mayor sea esa distancia, más evidencia tenemos de que el cambio sí tiene un efecto real.

DEFINICIONES > TEST DE HIPÓTESIS

TEST ESTADÍSTICO	Procedimiento que permite evaluar si la **diferencia observada en una muestra** es lo suficientemente **significativa** como para **inferirla a la población**. Nos ayuda a decidir si las diferencias observadas se deben al azar del muestreo o si reflejan un efecto real.

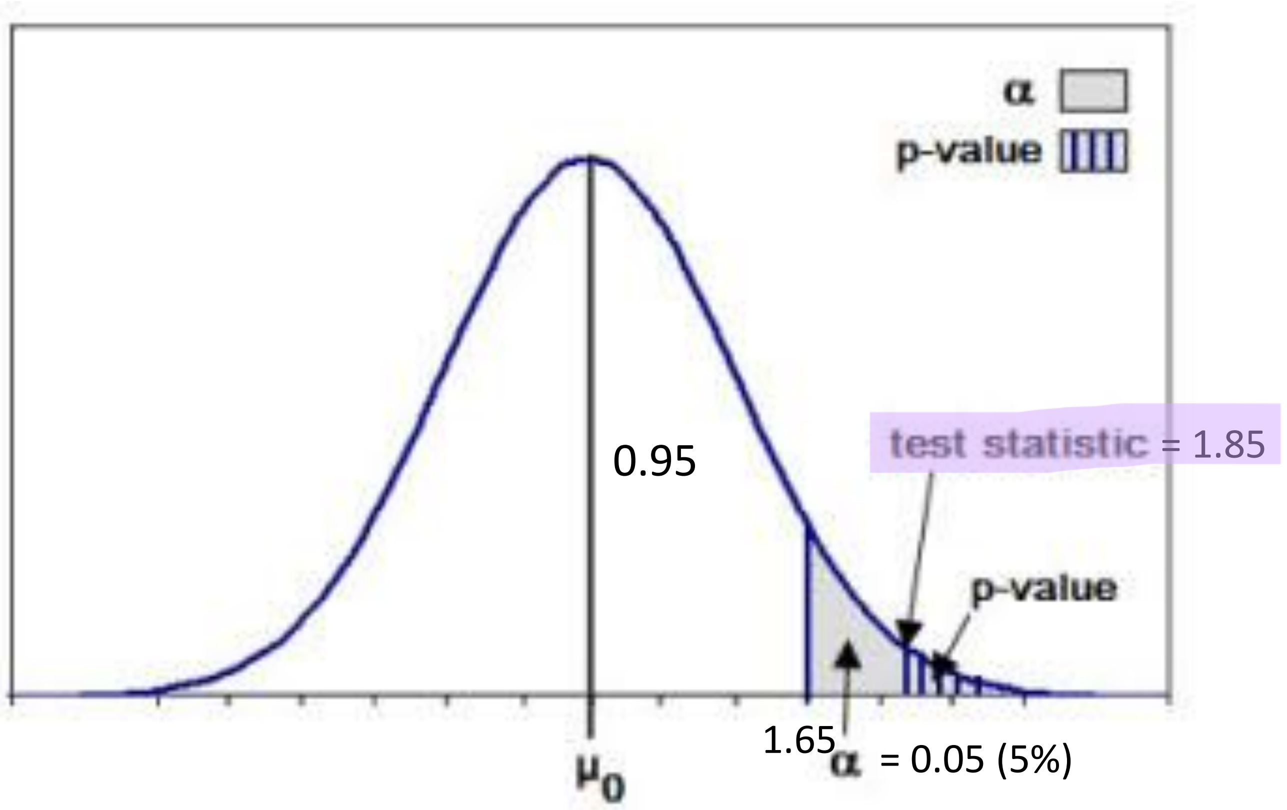
α = 0.05 (5%)

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

df = n - 1

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

df=n1+n2-2



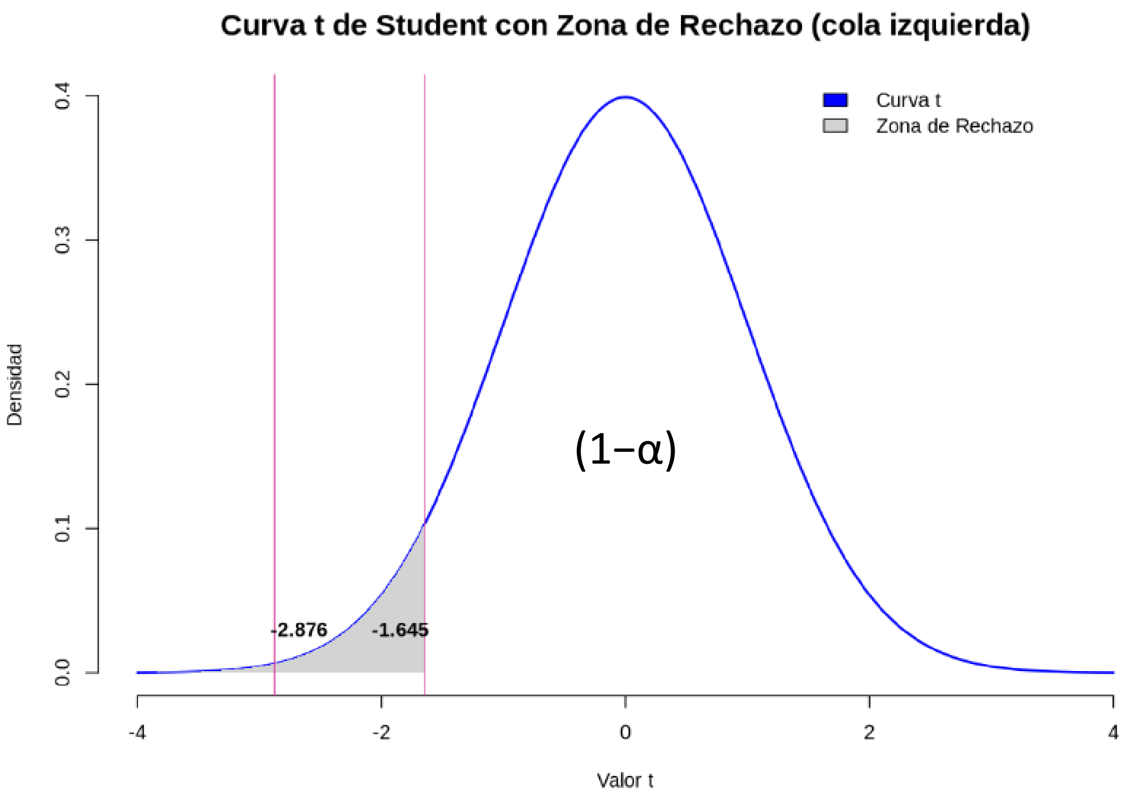
DEFINICIONES > TEST DE HIPÓTESIS

laSalle
Ramon Llull University

Distribuciones estadísticas	cómo se distribuyen o dispersan los valores de un conjunto de datos en un espacio muestral
Región Crítica	Región de los valores muestrales t a partir de los cuáles se rechaza la hipótesis nula
Nivel de significancia (α) – Error Tipo I	Define qué tan estrictos somos para cometer un error de tipo uno. Probabilidad de rechazar la hipótesis nula cuando en realidad es verdadera. Probabilidad máxima de cometer error tipo I.
Error tipo II (β)	La probabilidad de NO rechazar la hipótesis nula cuando en realidad es falsa
Región de aceptación	Región de los valores muestrales t a partir de los cuáles se acepta la hipótesis nula

DEFINICIONES > TEST DE HIPÓTESIS

Distribuciones estadísticas	Región Crítica	Nivel de significancia (α) – Error Tipo I	Error tipo II (β)



Dado un valor muestral t podemos calcular la probabilidad de pertenecer al subconjunto H_0 o H_1.

El objetivo por lo tanto es encontrar un valor límite *tcut* para el cuál, los datos obtenidos por debajo del mismo pertenezcan a H_0 y por encima del mismo pertenezcan a H_1.

DEFINICIONES > TEST DE HIPÓTESIS

laSalle
Ramon Llull University

P-valor	Es la probabilidad de obtener un resultado igual o más extremo que el observado, si la hipótesis nula (H_0) fuera cierta.

Nos dice **qué tan probables son los datos observados (valor del test estadístico)** si la **hipótesis nula (H_0)** fuera cierta.

P-valor bajo (<0.05): es poco probable obtener los resultados observados (ese valor de test estadístico) si la Ho fuera cierta. Por lo tanto, **hay evidencia suficiente para rechazar la hipótesis nula**. Hay evidencia a favor de H1. Indica **efecto o diferencia significativa**.

P-valor alto (>0.05): **No hay** evidencia suficiente para rechazar H_0. Los datos son **compatibles con H_0** (no se detecta diferencia significativa).

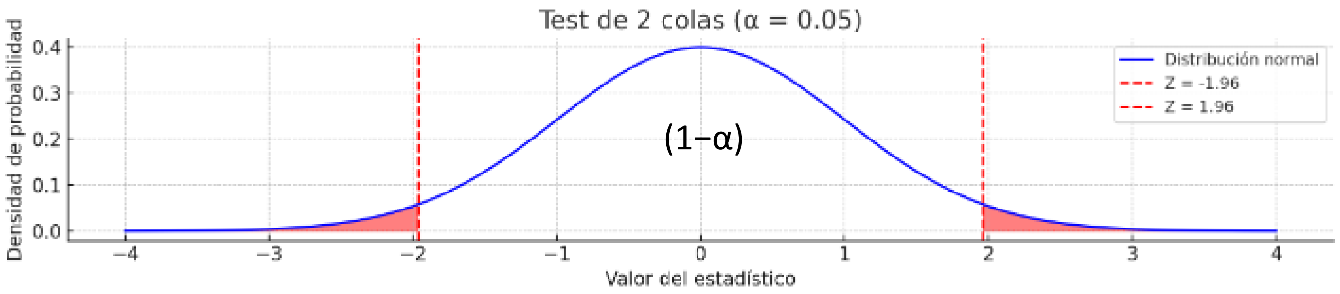
Nivel de Confianza	Probabilidad de que un intervalo de confianza contenga el valor verdadero del parámetro poblacional que se está estimando ($1-\alpha$)

Con un **nivel de confianza del 95%** indicamos que estamos dispuestos a correr un **5% de riesgo** de que el intervalo de confianza no contenga el valor del parámetro.

DEFINICIONES > TEST DE HIPÓTESIS

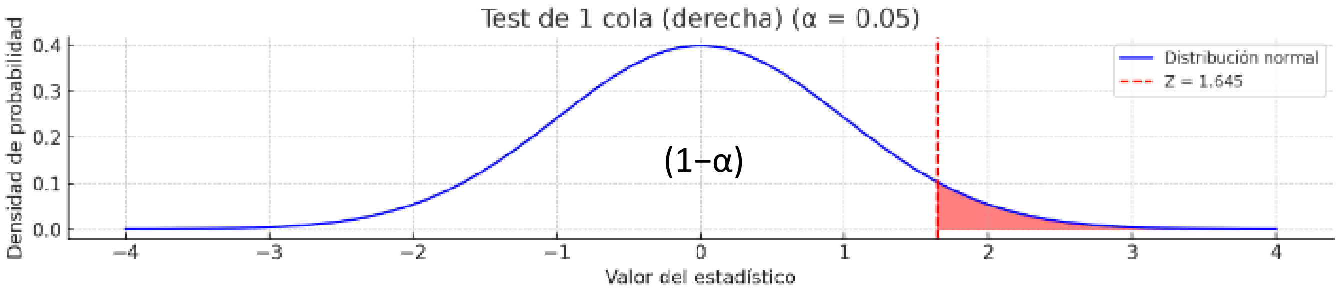
Test de hipótesis 1 o 2 colas

Test de 2 colas: Evalúa si hay una diferencia, sin importar el sentido
Test de 1 cola (derecha): Evalúa diferencias en una sola dirección positiva
Test de 1 cola (izquierda): Evalúa diferencias en una sola dirección negativa



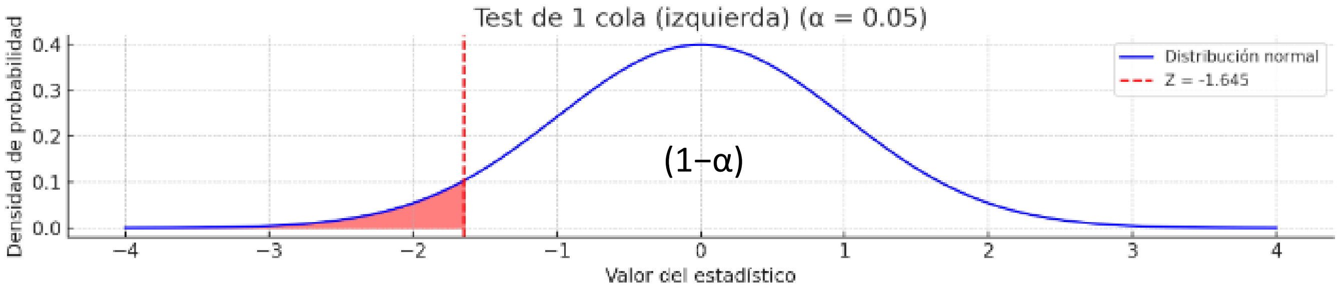
Hipótesis nula (H_0): $\mu = \mu_0$

Hipótesis alternativa (H_1): $\mu \neq \mu_0$



Hipótesis nula (H_0): $\mu \leq \mu_0$

Hipótesis alternativa (H_1): $\mu > \mu_0$



Hipótesis nula (H_0): $\mu \geq \mu_0$

Hipótesis alternativa (H_1): $\mu < \mu_0$

SELECCIÓN ESTADÍTICO DE PRUEBA

laSalle
Ramon Llull University

1

Distribución Z

Distribución Gaussiana.

Base de la inferencia estadística.Gracias al Teorema Central del Límite, las medias muestrales tienden a seguir una distribución normal cuando el tamaño de la muestra es grande.

- El valor de **la muestra es > 30**

Y

- Valor de la **varianza/desviación estándar** de la población **conocida**

2

Distribución T

Distribución T de Student.

Curva t más ancha y con colas más largas. El test se vuelve más exigente: los valores críticos aumentan, es más difícil rechazar H_0 y la diferencia debe ser mayor para considerarse significativa.

El valor de la **muestra es < 30**

Y/O

- Valor de la **varianza/desviación estándar** de la población **es desconocida**

SELECCIÓN ESTADÍTICO DE PRUEBA

laSalle
Ramon Llull University

1 Distribución Z

$\alpha = 0.05$

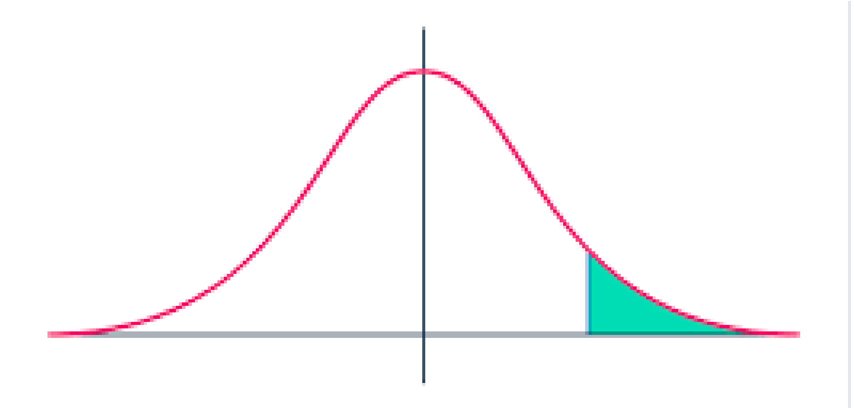
Estadístico observado (Zo) es **2.012**, y queremos comparar ese valor con el **valor crítico (Zcrítico)** de la distribución normal estándar.

Valor Z:

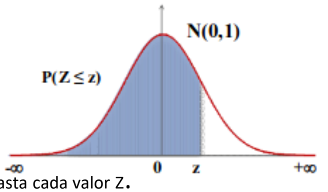
$$\mathbf{z} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \; ; \; \mathbf{z} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Buscamos en tabla el valor de **Zcrítico** tal que: $P\left(Z \leq Z_{crítico}\right) = 0.95$

Es decir, el valor de Z que deja **el 95% del área** bajo la curva a su izquierda.



FUNCIÓN DE DISTRIBUCIÓN NORMAL N(0,1)



Probabilidad acumulada hasta cada valor Z.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Nota: En el interior de la tabla se da la probabilidad de que la variable aleatoria Z, con distribución N(0,1), esté por debajo del valor z.

SELECCIÓN ESTADÍTICO DE PRUEBA

laSalle
Ramon Llull University

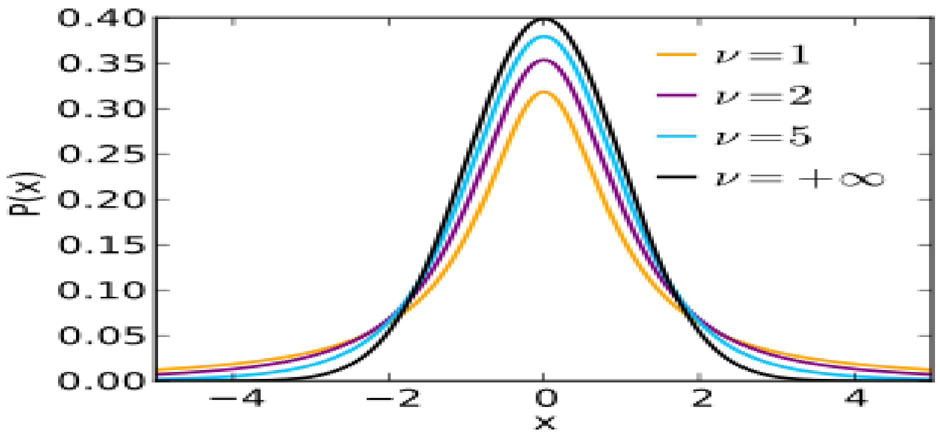
2 Distribución T

En muestras reducidas o σ Curva t más ancha y con colas más largas.

El test se vuelve más exigente: los valores críticos aumentan, es más difícil rechazar H_0 y la diferencia debe ser mayor para considerarse significativa.

Valor t:

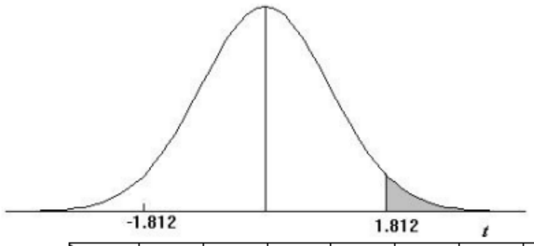
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$



v=gl

TABLA 2: DISTRIBUCIÓN t DE STUDENT

Puntos de porcentaje de la distribución t



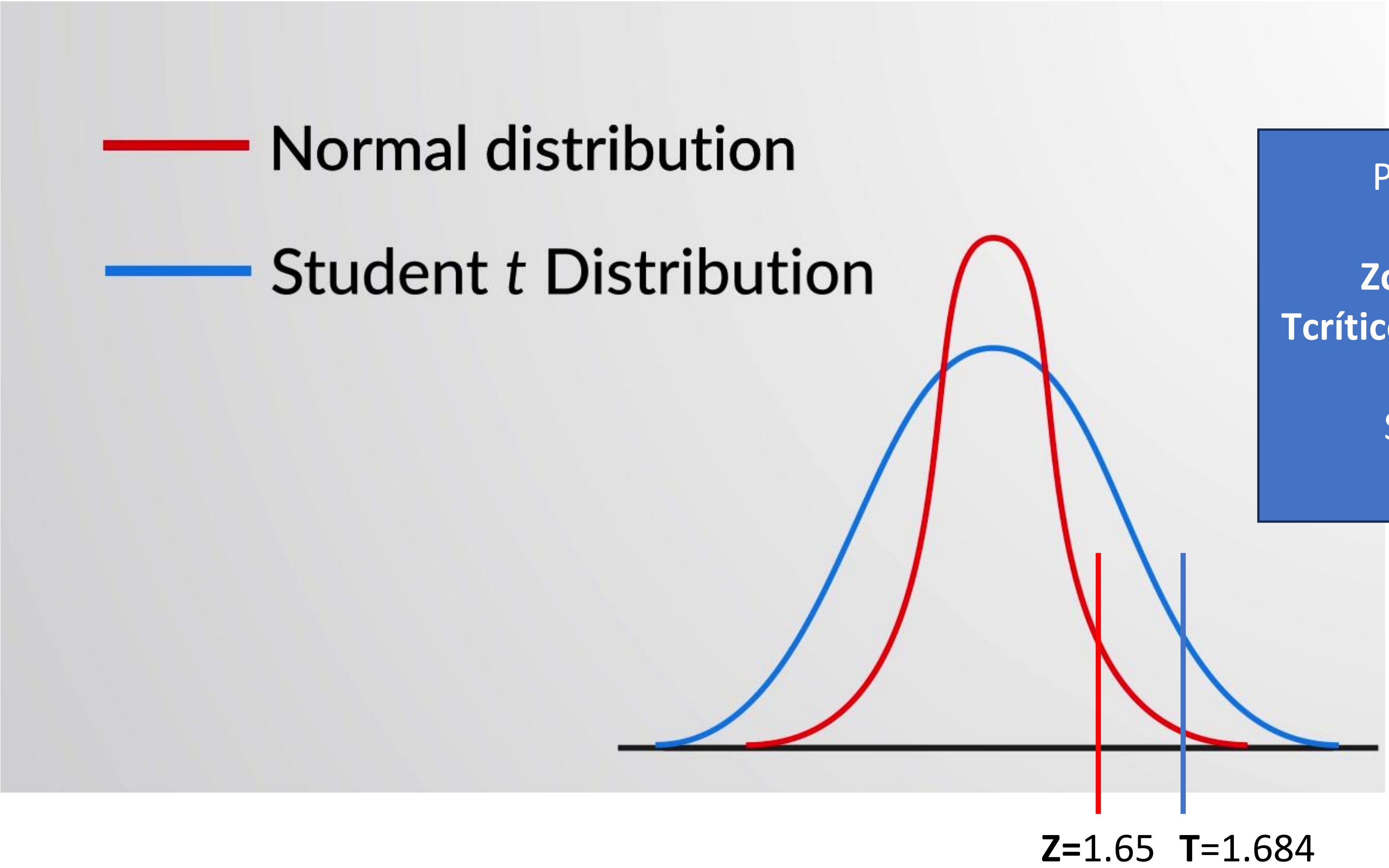
Ejemplo

Para ϕ = 10 grados de libertad:

$P[t > 1.812] = 0.05$
$P[t < -1.812] = 0.05$

r \ α	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

Curva t más ancha y con colas más largas. El test se vuelve más exigente: los valores críticos aumentan, es más difícil rechazar H_0 y la diferencia debe ser mayor para considerarse significativa.



Para x= 0.05

Zcrítico= 1.65
Tcrítico (gl=40) = 1.684

Si Zo =1.67
To = 1.67

SELECCIÓN ESTADÍTICO DE PRUEBA

laSalle
Ramon Llull University

t
- **n < 30** y/o
- ***σ*2/ *σ* <u>desconocida</u>**

z
- **n > 30** y/o
- ***σ*2/ *σ* <u>conocida</u>**

Tipos de tests

One sample

- Comparar la media de una muestra con una media poblacional conocida o con un valor de referencia (benchmark).
- *Ej: Estudiar si la media de las notas de una clase difiere de la media nacional (μ=7.2)*

Si n**< 30** y/o ***σ*2/ *σ* <u>desconocida</u>** -- > $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ con gl=n-1

Si n **> 30** y/o ***σ*2/ *σ* <u>conocida</u>** -- > $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Two samples

- Comparar las medias de dos grupos/ poblaciones mediante muestras independientes y determinar si existe una diferencia significativa entre ellas.
- *Ej: Comparar si la media de notas de los alumnos del grupo A difiere de la media de notas del grupo B.*

Si n**< 30** y/o ***σ*2/ *σ* <u>desconocida</u>** -- > $t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ con gl=n1+n2-2

Si n **> 30** y/o ***σ*2/ *σ* <u>conocida</u>** -- > $Z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

SELECCIÓN ESTADÍTICO DE PRUEBA

t

z

- **Muestra grande**

Tipos de tests

Paired t-test

Comparar **dos medias obtenidas de muestras dependientes,** datos provienen de los mismos individuos u observaciones antes y después de un cambio.

Ej: Profesor evalúa si su nueva estrategia mejora las notas comparando, en n alumnos, las calificaciones antes y después del cambio.

One-sample proportion test

Comparar una proporción observada en una muestra con una proporción poblacional conocida o valor de referencia (benchmark).

Ej: Comprobar si la proporción de personas que aprueban un examen en un instituto ($\hat{p}$) es diferente de la proporción nacional conocida ($p_0 = 0.6$).

Two-sample proportion test

Comparar dos proporciones y ver si existe una diferencia significativa entre ellas

Ej: Comparar la proporción de aprobados en dos ciudades diferentes ($\hat{p}_1$ y $\hat{p}_2$).

EJEMPLO: COMPARATIVA MUESTRAL

laSalle
Ramon Llull University

Una empresa de fabricación de bombillas produce bombillas con una **vida media de 1000 horas**. Recientemente han recibido quejas de clientes que la **vida útil es inferior a las 1000 horas promocionadas** y se quiere validar esta hipótesis:

1 **Definir hipótesis**

H_0: Hipótesis Nula: Situación Inicial:
H_1: Hipótesis Alternativa: Suponemos que la media de la población es inferior a 1000h.

Por lo tanto:

$$H_0: \mu = 1000h$$
$$H_1: \mu < 1000h$$

2 ***Datos de la muestra***

Tomamos una muestra de **40 bombillas** para realizar el estudio y obtenemos una **vida media de 975** horas y una **desviación estándar de 50 horas.**

Valores comunes:
- α = 5%
- α = 1%

EJEMPLO: COMPARATIVA MUESTRAL

3 *Selección de estadístico de prueba*

- Tamaño de la muestra (n): 40 bombillas
- Media muestral (x): 975 horas
- Desviación estándar muestral (s): 50 horas
- Media hipotética de la población (µ0): 1000 horas

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \qquad t = \frac{975 - 1000}{\frac{50}{\sqrt{40}}} \qquad t \approx -3.16$$

4 *Determinamos área rechazo si nivel significancia = 0.05*

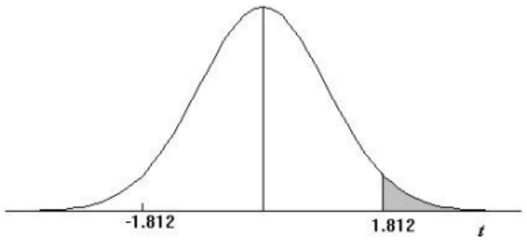
Valor crítico

- con n−1=39 grados de libertad
- nivel de significancia de α=0.05

$$t_{crítico} \approx -1.685$$

TABLA 2: DISTRIBUCIÓN t DE STUDENT

Puntos de porcentaje de la distribución t



Ejemplo

Para ϕ = 10 grados de libertad:

$P[\, t > 1.812] = 0.05$
$P[\, t < -1.812] = 0.05$

r \ α	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

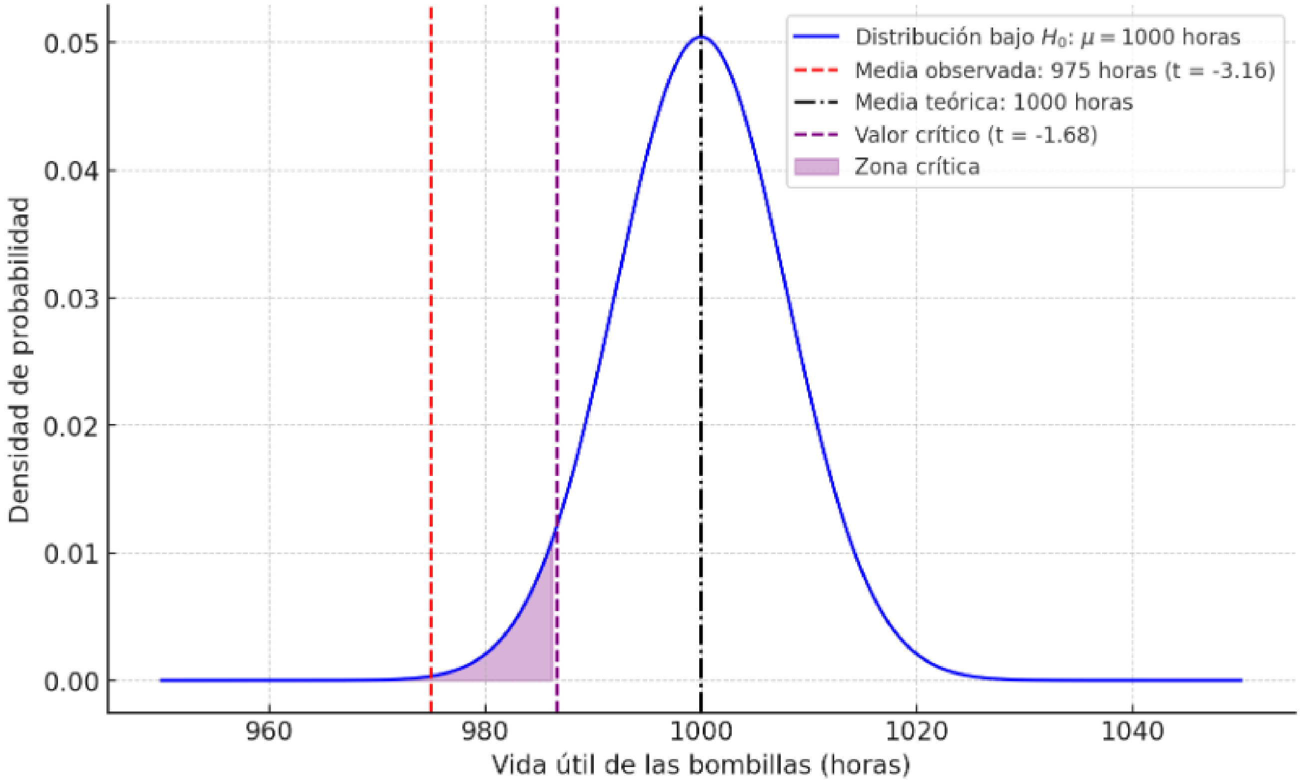
EJEMPLO: COMPARATIVA MUESTRAL

RESOLUCIÓN

$H_0 : \mu = 1000$ horas

$H_1 : \mu < 1000$ horas

distribución de la vida útil de las bombillas



To=-3.16 **Tcrítico = -1.685**

Existe evidencia suficiente para afirmar que la media poblacional es **inferior a 1000h**.

DISEÑO DE UN TEST DE HIPÓTESIS

1. Definir universo, población y muestra
2. Definición de Hipótesis Nula y alternativa
3. Selección de estadístico de prueba
4. Seleccionar el nivel de significancia
5. Calcular el estadístico de prueba y determinar la región de rechazo
6. Toma de decisión

MATRIZ DE CONFUSIÓN BINARIA

Matriz de confusión Binaria:

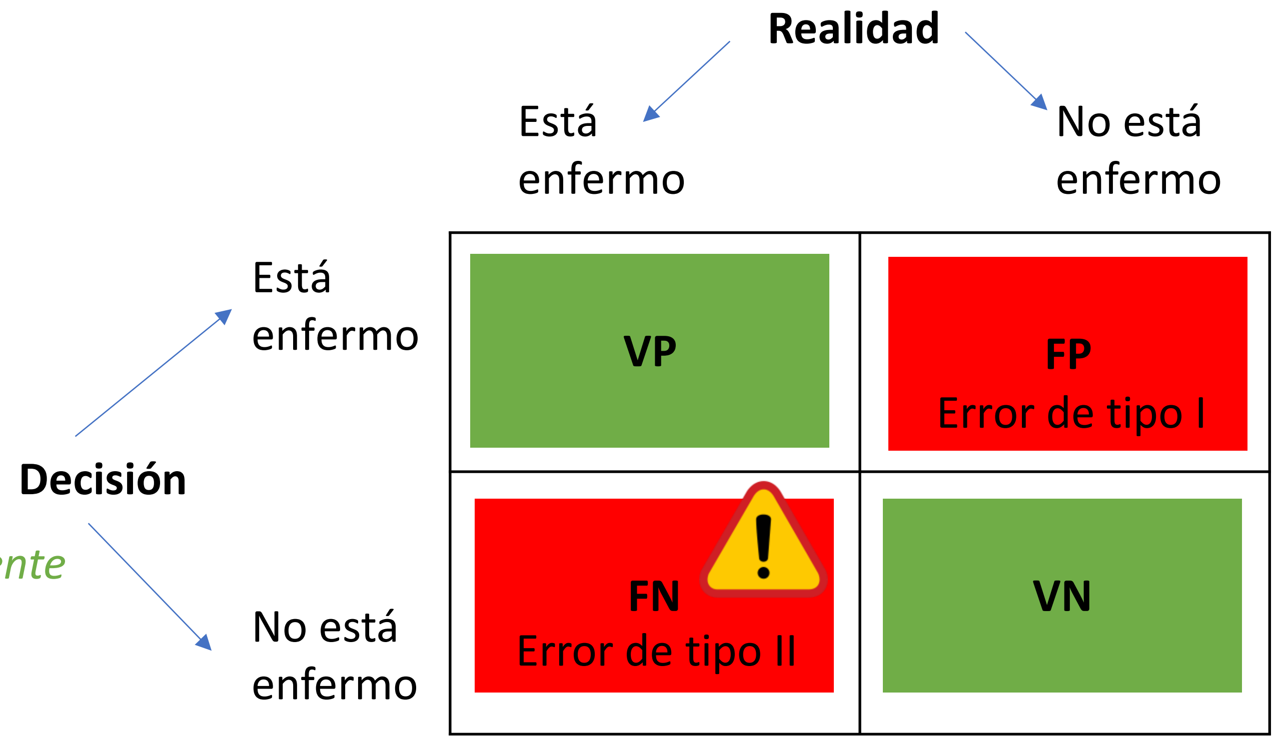
Al tomar una decisión binaria (Rechazar Ho o no) podemos acertar/ equivocarnos de diferentes maneras.
La matriz de decisión nos ayuda a visualizar todas las posibles combinaciones de esos posibles aciertos y errores.

Ho: no está enfermo

H1: está enfermo

- **Verdadero positivo:** Rechazar Ho
- **Verdadero negativo:** No rechazar Ho

- **Falso Positivo:** Error de tipo 1
  - *Ejemplo: Falsa alarma – Diagnosticar un paciente como enfermo cuando en realidad está sano.*
- **Falso Negativo:** Error de tipo 2
  - *Ejemplo: Diagnosticar un paciente como sano cuando en realidad está enfermo*



MATRIZ DE CONFUSIÓN BINARIA

Métricas asociadas Curva a la matriz de confusion binaria:

		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$	F_1 score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$		

Capacidad de detectar VP

Capacidad de detectar TN

Fuente - Wikipedia: https://en.wikipedia.org/wiki/Receiver_operating_characteristic

MATRIZ DE CONFUSIÓN BINARIA

Sensitivity o True Positive Rate (TPR): Equivalente a la Tasa de acierto

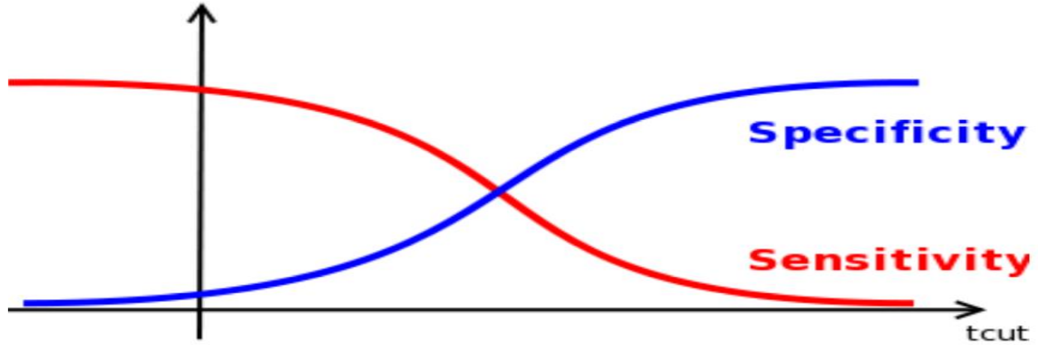
$$TPR = TP/P = TP/(TP + FN)$$

Speficificy o False Positive Rate (SPC): Equivalente a la Tasa de error

$$SPC = TN/N = TN/(FP + TN)$$

Las curvas de sensitividad y especificidad evolucionan complemetariamente dependiendo del valor de *t-cut,*

El t-cut cambia en función de lo exigentes que seamos con nuestro error tipo I.

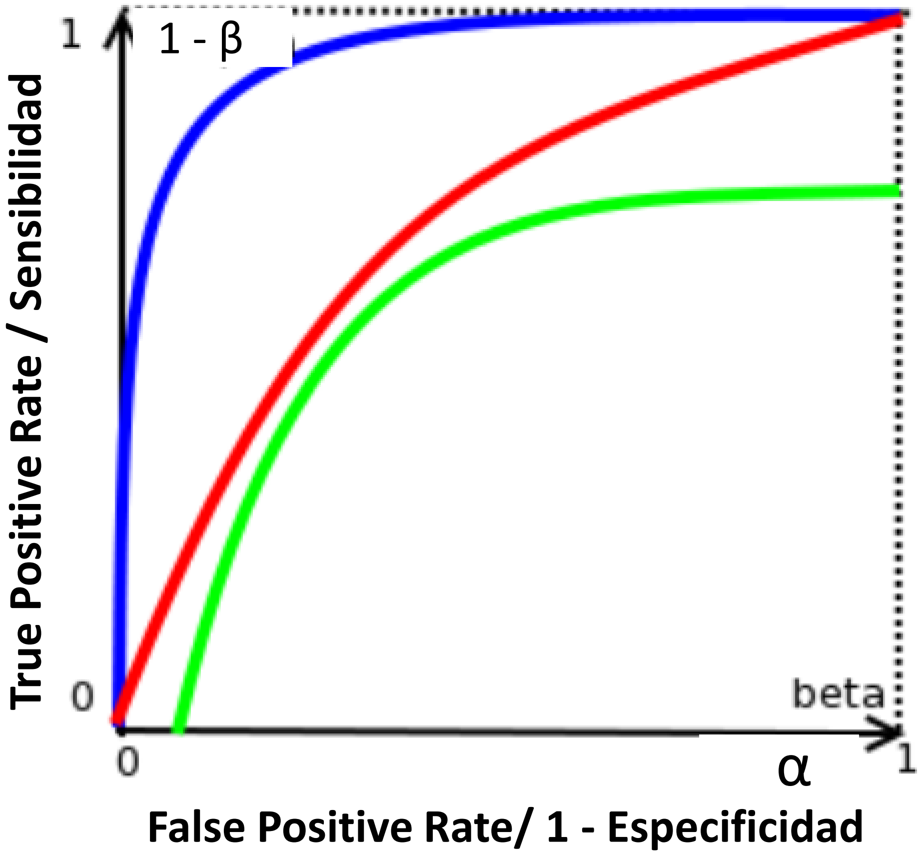


MATRIZ DE CONFUSIÓN BINARIA

laSalle
Ramon Llull University

Curva ROC:

- *La **curva ROC** (Receiver Operating Characteristic) es una representación gráfica que muestra la relación entre la **Sensibilidad (True Positive Rate)** y el **1 - Especificidad (False Positive Rate)** de un modelo de clasificación.*
- *La **Sensibilidad** está relacionada con el **Error de Tipo II (β),** mientras que el False Positive Rate **(1 - Especificidad)** está vinculado al **Error de Tipo I (α).***
- Cada punto en la curva representa **un posible tcut**, un posible equilibrio entre sensibilidad y especificidad
- Las curvas ayudan a **comparar** distintos tests o modelos de clasificación, mostrando su **capacidad** para **distinguir entre clases positivas y negativas.**
- El área bajo la curva ROC **(AUC)** representa la probabilidad de que el modelo asigne una **puntuación mayor** a una **observación positiva** que a una negativa seleccionadas al azar.



Si el **AUC = 0.5** → el test no sirve
Si el **AUC = 1** → el test es perfecto (detecta todos los VP sin FP).

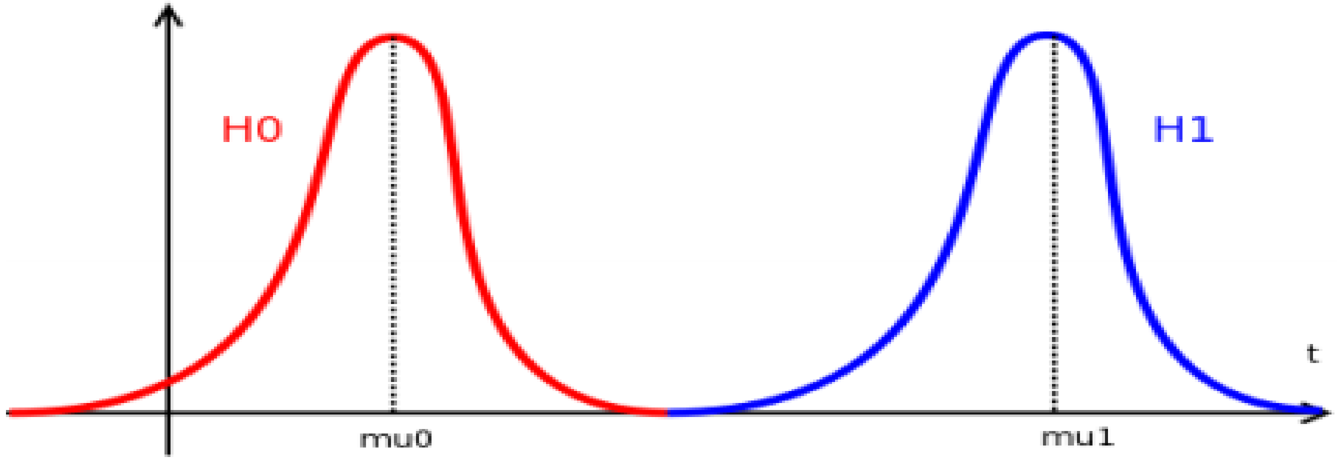
DISCRIMINACIÓN DE HIPÓTESIS

laSalle
Ramon Llull University

Discriminante de Fisher

El objetivo de buen estudio estadístico: distinguir claramente cuándo una hipótesis es cierta y cuándo no

Ejemplo: determinar si una vacuna no es efectiva (Ho) o si lo es (H1)



$$
\begin{array}{llll}
H_0 & x_i & \longrightarrow & \mu_{0i},\ \sigma^{(2)}_{0ij} \\
H_1 & x_i & \longrightarrow & \mu_{1i},\ \sigma^{(2)}_{1ij}
\end{array}
$$

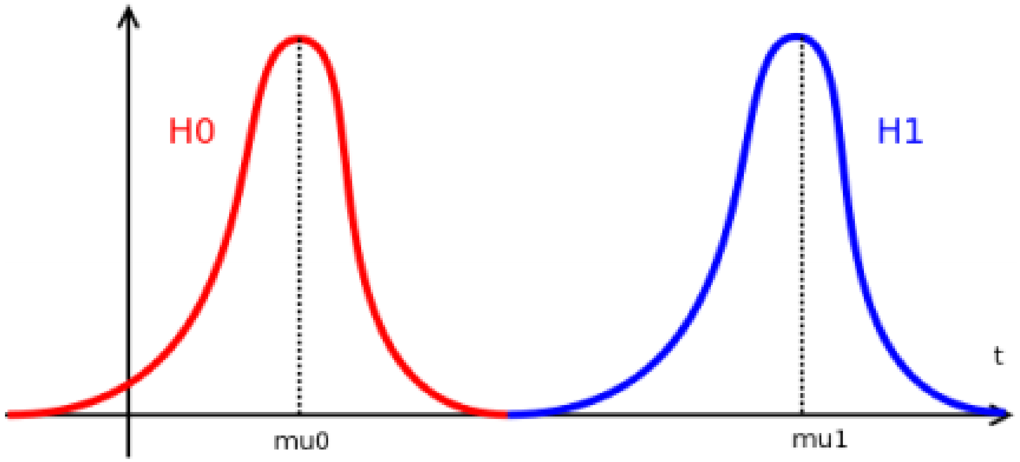
- Cada curva muestra cómo se distribuyen los resultados si cada hipótesis fuera cierta. Cuanto **más separadas** estén, **mejor discriminación** del efecto. Si se **solapan**, el efecto no es estadísticamente distinguible del azar.

- Fisher formalizó este concepto con el **Discriminante Lineal de Fisher** ,una función que **maximiza la separación entre las medias (μ_0 y μ_1)** y minimiza el solapamiento.

DISCRIMINACIÓN DE HIPÓTESIS

Objetivos del discriminante lineal de Fisher:

- **Maximizar la separación de las medias** entre las distribuciones: las medias de los dos grupos deben estar lo más separadas posible.
- **Minimizar la varianza dentro de cada grupo:** ambos conjuntos deberían tener una varianza pequeña, reduciendo el solapamiento y asegurando consistencia dentro de cada hipótesis.
- *Visualmente: dos curvas bien separadas con poca dispersión*



- Mide cuántas desviaciones estándar separan las dos medias.
- Nos dice cuán lejos está la media de la muestra respecto a lo que esperaríamos si H_0 fuera cierto (no hubiera efecto).
- Cuanto mayor sea ese valor, mayor evidencia de que existe una diferencia significativa.

Estadístico de comparación z:

$$z = \frac{\bar{x}_{H1} - \bar{x}_{H0}}{\sqrt{\frac{S_{H1}^2}{n_{H1}} + \frac{S_{H1}^2}{n_{H0}}}}$$

Estadístico de comparación t:

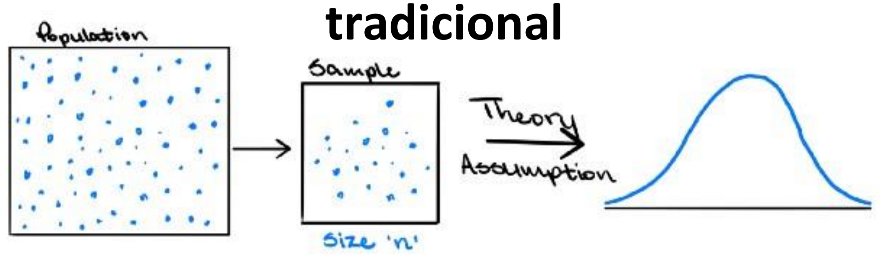
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

con gl=n1+n2-2

Bootstrapping

Hasta ahora…
Hemos asumido que los datos siguen una **distribución teórica (normal)**
Y que contamos con **muestras grandes**

Test de Hipótesis tradicional



Pero… ¿qué ocurre si:
- La muestra es **pequeña**
- Presencia de **outliers**
- Los datos **no son normales**

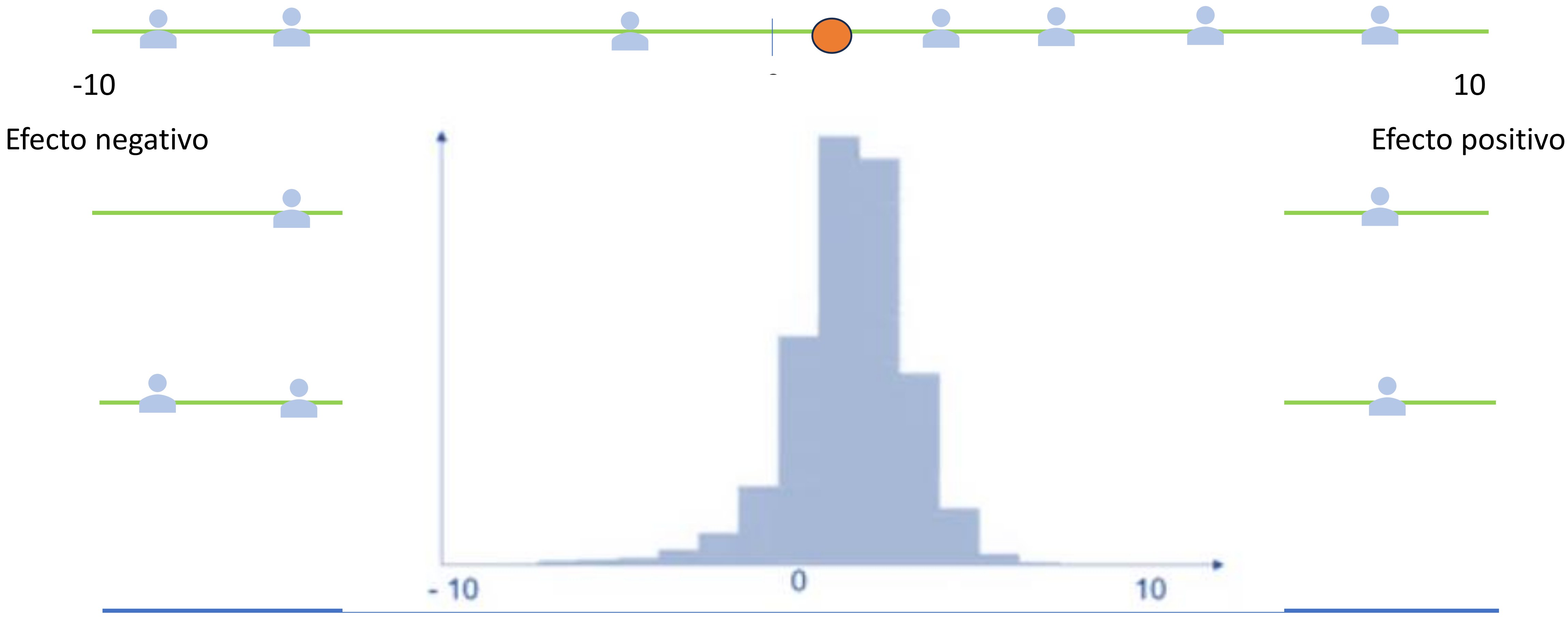
Fuente: Ejemplo de bootstrapping en kaggle

CONCEPTOS AVANZADOS SOBRE A/B TESTING

Bootstrapping

$H_0: \mu = 0$

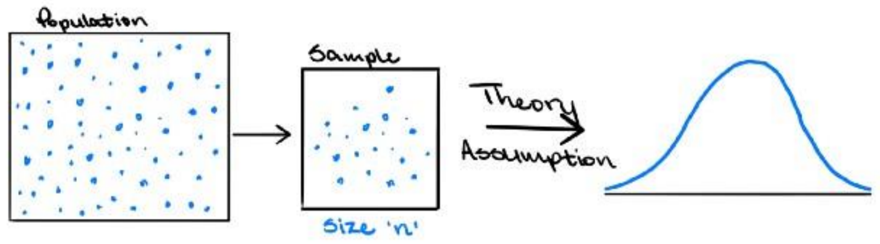
$H_1: \mu \neq 0$



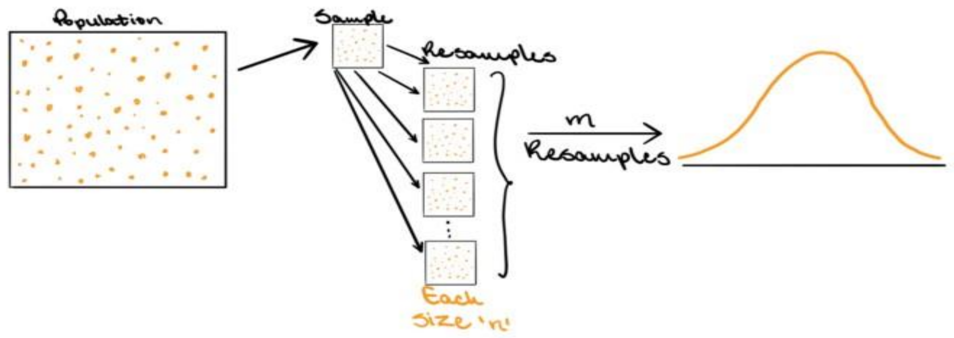
Bootstrapping

Técnica estadística por la cual, partiendo de un conjunto de datos *n,* generamos *m* subconjuntos de datos de tamaño *n'*. Generamos muchas réplicas de nuestra muestra, calculamos estadístico de interés y construímos su distribución empírica.

Test de Hipótesis tradicional



Test de Hipótesis con Bootstrap



Bootstrapping nos permite realizar test de hipótesis de forma más robusta, incluso cuando tenemos pocos datos o valores atípicos.

Fuente: Ejemplo de bootstrapping en kaggle

Test de Hipótesis Bayesianos

En estadística hay dos enfoques:

1

Frecuentista

2

Bayesiano

Test de Hipótesis Bayesianos

En estadística hay dos enfoques: Queremos estimar la **altura media real** de todos los alumnos de La Salle.

1 Frecuentista

Procedimiento:
Tomamos una muestra aleatoria de **50 estudiantes** y calculamos una media de **1.74 m**, con un **intervalo de confianza del 90% [1.70 m – 1.78 m]**.

Interpretación frecuentista:
Si repitiéramos este experimento muchas veces, tomando distintas muestras de 50 alumnos cada vez, el 90% de los intervalos de confianza que construyamos contendrán la verdadera media de altura de todos los alumnos de La Salle.

La media real es fija pero desconocida, lo que cambia son las muestras.

2 Bayesiano

Procedimiento:
Partimos de una creencia previa: antes de medir, creemos (por experiencia previa) que la altura media ronda 1.70 m.

Después de medir a 50 estudiantes, actualizamos esa creencia y obtenemos una distribución posterior según la cual la altura media está entre 1.72 m y 1.77 m con un 90% de probabilidad.

Interpretación bayesiana:
Dado lo que creíamos antes y los datos que hemos observado, hay un 90% de probabilidad de que la altura media real de los alumnos de La Salle esté entre 1.72 m y 1.77 m

Los datos observados se consideran fijos, lo aleatorio es el parámetro, que refleja nuestra incertidumbre.

Aspecto	**Frecuentista**	**Bayesiana**
Objetivo	**Estimar parámetros fijos y evaluar métodos de inferencia.**	Actualizar creencias sobre los parámetros a medida que se observan nuevos datos.
Qué considera fijo	El parámetro (por ejemplo, la media real).	Los datos observados (una vez recogidos).
Qué considera aleatorio	Los datos de la muestra (varían en cada experimento).	El parámetro desconocido (representa nuestra incertidumbre).
Interpretación de la probabilidad	Frecuencia relativa de un evento al repetir el experimento muchas veces.	Grado de creencia o incertidumbre sobre un valor o hipótesis.
Intervalo del 90%	El 90% de los intervalos obtenidos en repeticiones del experimento contendrán la media real.	Hay un 90% de probabilidad de que la media real esté dentro del intervalo.
Uso del conocimiento previo	No se considera (solo los datos observados).	Se incorpora mediante una distribución a priori, que se actualiza con los datos (a posteriori).
Tipo de inferencia	Evaluación de métodos y contrastes de hipótesis.	Actualización de conocimiento y estimación probabilística.
Ejemplo intuitivo	“Si repito el experimento muchas veces, 9 de cada 10 intervalos contendrán la media real.”	“Dado lo que sé y los datos que tengo, hay un 90% de probabilidad de que la media esté entre estos valores.”

Test de Hipótesis Bayesianos

Hasta ahora hemos trabajado con el enfoque:

1

Frecuentista

Quiero saber altura promedio de adultos de mi pueblo:

"El intervalo de confianza del 90% para la altura media está entre 1,68 m y 1,75 m"

Pero…
Qué significa?

Hay un 90% de probabilidad de que la media real poblacional esté en ese rango (interpretación incorrecta).

Si repitiéramos el muestreo muchas veces, el 90% de los intervalos de confianza incluiría la media real poblacional (interpretación correcta).

Razones para [confiar]() y [desconfirar]() en los Tests de Hipótesis Bayesianos

Test de Hipótesis Bayesianos

Basados en estadística bayesiana, en donde la evidencia de las hipótesis se expresa en probabilidades

2

Bayesiano

Quiero saber altura promedio de adultos de mi pueblo”

Hay un 90% de probabilidad de que la media real poblacional esté en ese rango.

Razones para confiar y desconfirar en los Tests de Hipótesis Bayesianos

MD004 Test de Hipótesis

Máster universitario en Ciencia de los Datos / Data Science (MUDS)

Curso académico 2024/25
13 de noviembre de 2024

Xavier Vilasís Cardona
Ángel Berián
Gloria Aguilera