CONTRASTES DE HIPÓTESIS PARAMÉTRICOS

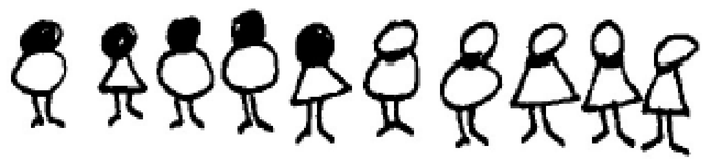
La estadística en cómic
(L. Gonick y W. Smith)

EJEMPLO: UN JUICIO INJUSTO

APELAREMOS EL VERIDICTO

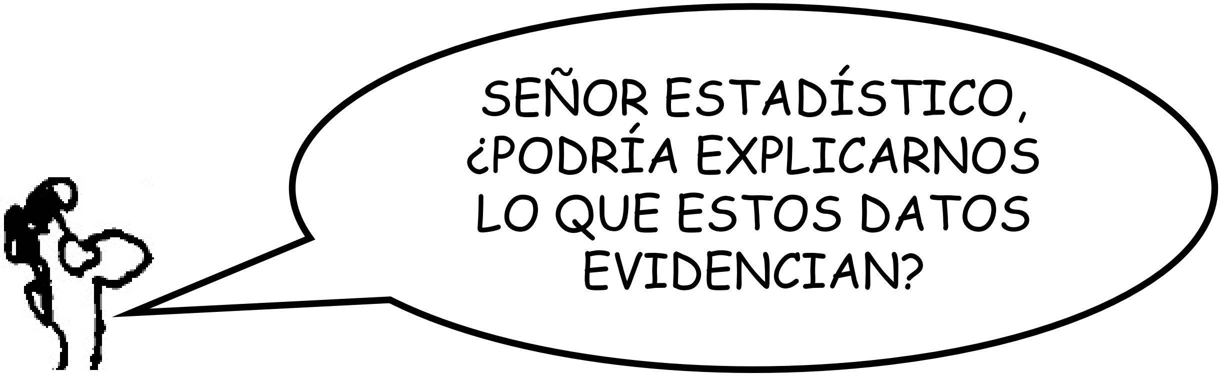
EN TEORÍA, LAS LISTAS DE JURADOS SE ELABORAN DE FORMA ALEATORIA A PARTIR DE UNA LISTA DE CIUDADANOS SUSCEPTIBLES DE SER ELEGIDOS. SIN EMBARGO, DURANTE LAS DÉCADAS DE LOS 50 Y 60, EN LOS ESTADOS DEL SUR, HABÍA MUY POCOS AFROAMERICANOS EN LAS LISTAS DE JURADOS. ASÍ QUE ALGUNOS ABOGADOS DE LA DEFENSA PUSIERON EN TELA DE JUICIO LOS VEREDICTOS. EN LA APELACIÓN, UN TESTIGO EXPERTO EN ESTADÍSTICA PRESENTÓ ESTA PRUEBA:

1) EL **50%** DE LOS CIUDADANOS SUSCEPTIBLES DE SER ELEGIDOS ERAN AFROAMERICANOS.



2) EN UNA LISTA DE 80 MIEMBROS POTENCIALES DEL JURADO, SÓLO **CUATRO** ERAN AFROAMERICANOS.

¿PUEDE SER ESTO FRUTO DE LA **PURA CASUALIDAD**?



ABOGADO DE LA DEFENSA

PARA COMPLETAR EL ARGUMENTO, SUPONGAMOS QUE LA ELECCIÓN DEL JURADO POTENCIAL FUE **ALEATORIA**, ENTONCES, EL NÚMERO DE AFROAMERICANOS DE LA LISTA DE 80 PERSONAS SERÍA LA VARIABLE ALEATORIA **BINOMIAL** X CON $n = 80$ PRUEBAS Y $p = 0{,}5$.

ENTONCES, LA POSIBILIDAD DE FORMAR UN JURADO CON HASTA 4 MIEMBROS AFROAMERICANOS ES $Pr(x \leq 4)$ O SEA, ALREDEDOR DE 0,0000000000000000014 (!).

COMO LA PROBABILIDAD ES TAN PEQUEÑA, LA LISTA EN CUESTIÓN, CON SÓLO CUATRO AFRO-AMERICANOS, RESULTA UNA **PRUEBA DE PESO** CONTRA LA **HIPÓTESIS** DE LA SELECCIÓN ALEATORIA.

¡¡EL JUICIO QUEDA ANULADO!!

¿QUÉ ES UN CONTRASTE DE HIPÓTESIS PARAMÉTRICO?

Es la técnica estadística que se usa para estudiar si una determinada afirmación acerca de cierto parámetro poblacional es confirmada o invalidada por los datos de una muestra extraída de dicha población.

<u>**EJEMPLO**</u>: ¿La selección del jurado es aleatoria? $\Updownarrow$ ¿p=0.5?

HIPÓTESIS A CONTRASTAR: LA NULA Y LA ALTERNATIVA

HIPÓTESIS NULA H_0

- Es la que se supone cierta, y debe aceptarse salvo prueba en contra
- Los datos muestrales pueden refutarla
- No debe ser rechazada sin una gran evidencia en contra

HIPÓTESIS ALTERNATIVA H_a

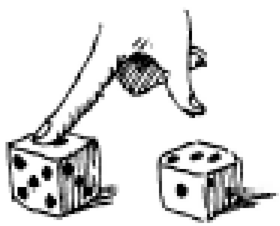
- Es la que niega la hipótesis nula
- Los datos muestrales pueden mostrar evidencias a favor.
- No debe ser aceptada sin una gran evidencia a favor

EJEMPLO:

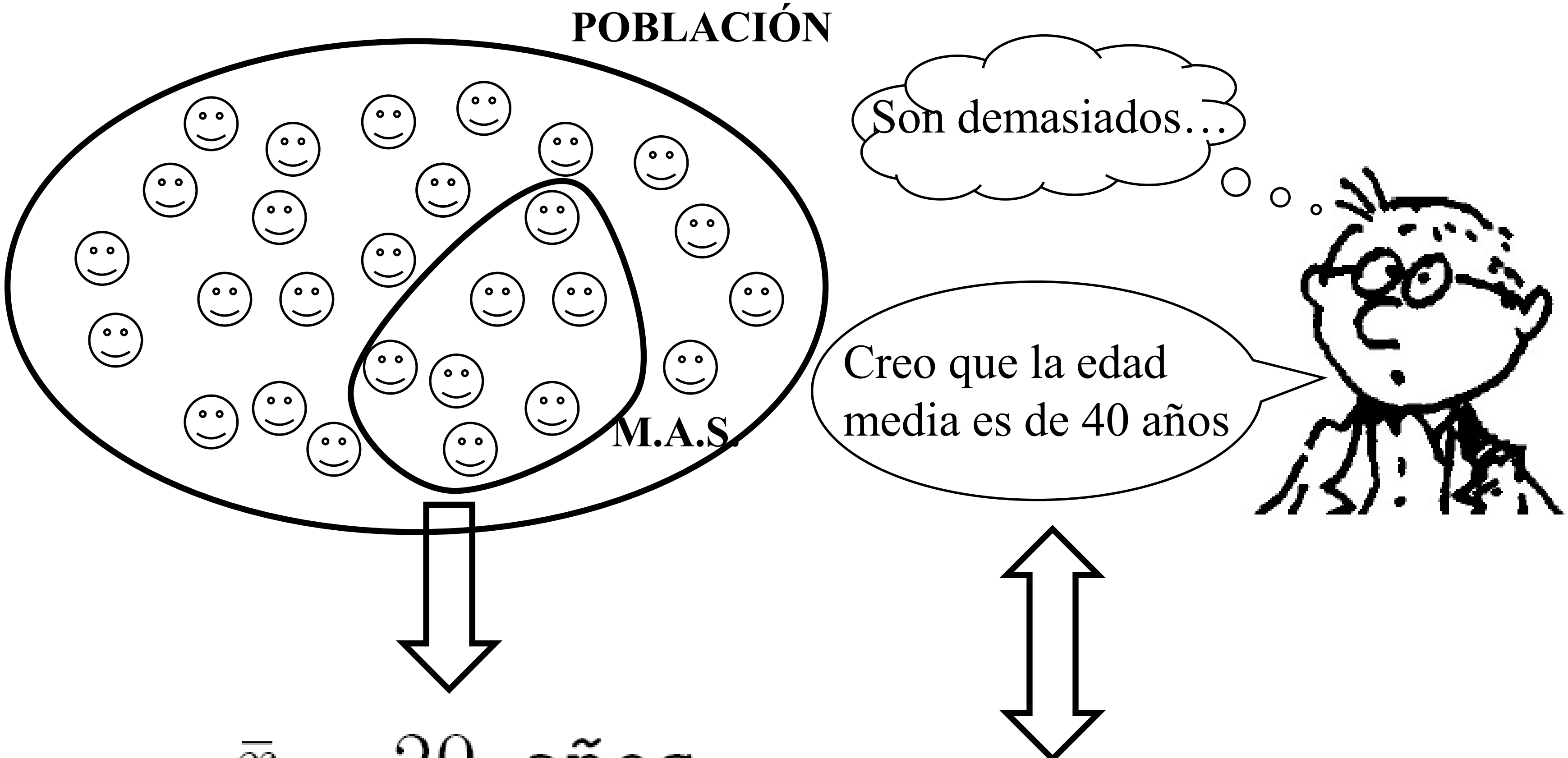
$$\begin{cases} H_0 : p = 0.5 \\ H_a : p \neq 0.5 \end{cases}$$

RAZONAMIENTO BÁSICO DEL CONTRASTE DE HIPÓTESIS

Localizar un suceso que sea muy improbable cuando la hipótesis nula se supone cierta; si, una vez extraída una muestra aleatoria acontece dicho suceso, o bien es que el azar nos ha jugado una mala pasada al elegir una muestra muy rara, o bien, como parece más razonable, la hipótesis nula es falsa.



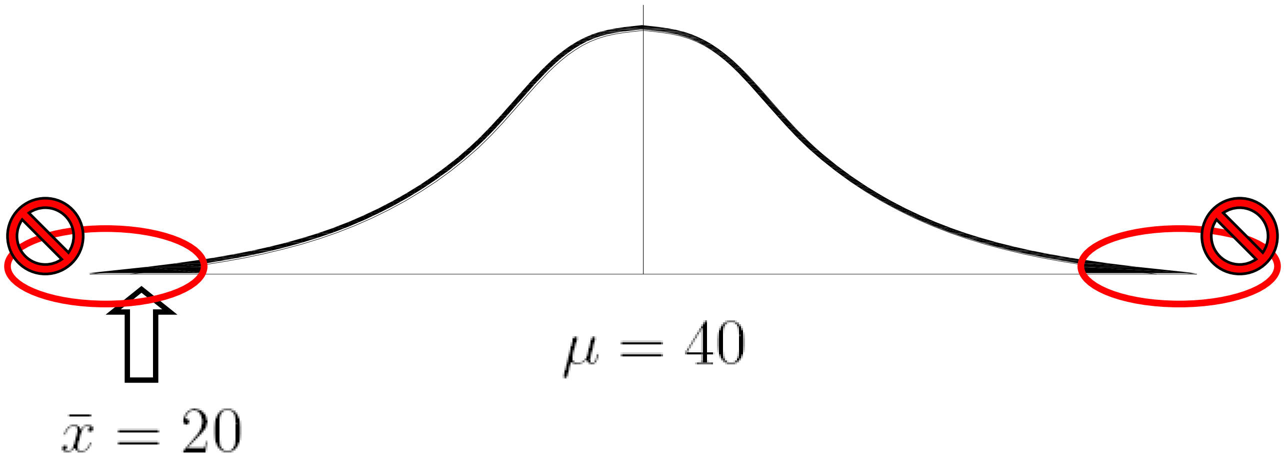
EJEMPLO: ¿Cuál es la edad media de la población?



$$\bar{x} = 20 \text{ años}$$

$$H_0 : \mu = 40 \text{ años}$$

Si suponemos que la hipótesis nula es cierta y normalidad en los datos, X = edad de un individuo de la población sigue una distribución normal de media 40…



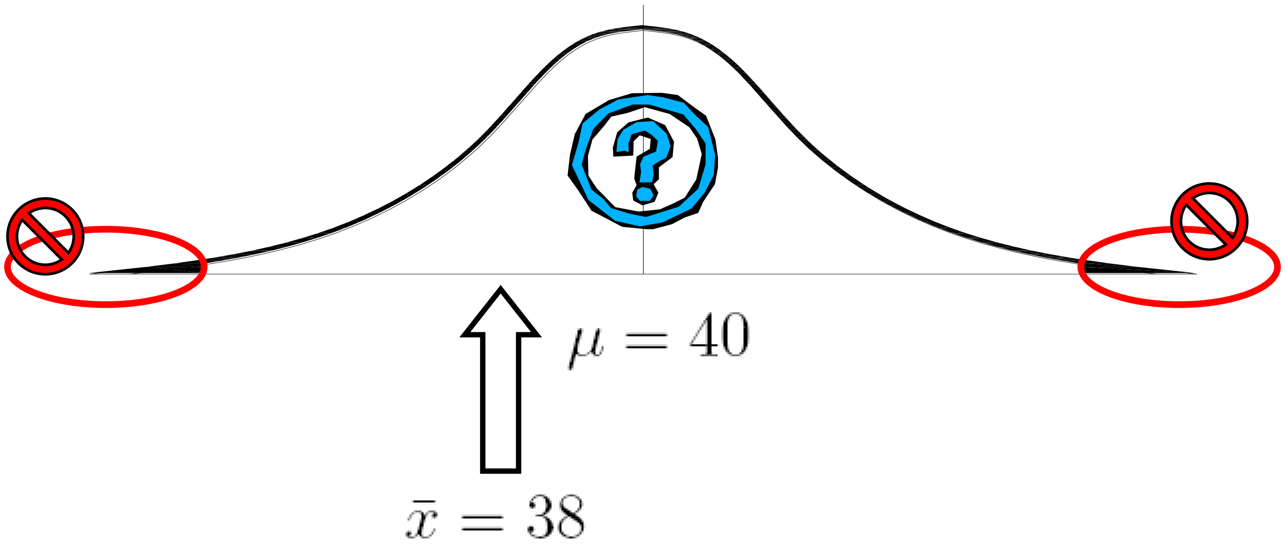
y por lo tanto, es muy improbable que la media muestral se encuentre en las zonas marcadas en rojo.

¡¡¡Sin embargo, ha ocurrido!!!

$\Downarrow$

Rechazamos que H_0 sea cierta

Pero, ¿qué hubiéramos podido afirmar si hubiésemos obtenido una media muestral de 38 años?



- No hay evidencias contra la hipótesis nula

-No podemos rechazar la hipótesis nula (NI ACEPTARLA)

- El contraste no es significativo.

Definición: Sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria de una caracteristicaX de una población con función de masa $P_\theta(x)$ (caso discreto), o con función de densidad $f_\theta(x)$(caso continuo) donde $\theta \in \Theta$ es desconocido. Sea $\Theta = \Theta_0 \cup \Theta_1$, $\mathrm{H}_0 : \theta \in \Theta_0$ y $\mathrm{H}_1 : \theta \in \Theta_1$

Un test consiste en decidir para cada muestra si aceptar o rechazar H_0

¿QUÉ ERRORES PODEMOS COMETER AL TOMAR UNA DECISIÓN?

ANALOGÍA: Consideremos el contraste de hipótesis como un detector de humos

ESTO ES LO QUE SE LLAMA UN **ERROR DE TIPO I**: UNA ALARMA SIN FUEGO. POR EL CONTRARIO, UN **ERROR DE TIPO II** ES UN FUEGO SIN ALARMA. TODOS LOS COCINEROS SABEN CÓMO EVITAR UN ERROR DE TIPO I: **QUITANDO LAS PILAS**. POR DESGRACIA ESTO AUMENTA LA INCIDENCIA DE ERRORES DE TIPO II.

DE IGUAL MODO, LA DISMINUCIÓN DE POSIBILIDADES DE ERRORES DE TIPO II, POR EJEMPLO, HACIENDO QUE LA ALARMA SEA HIPERSENSIBLE, PUEDE AUMENTAR EL NÚMERO DE FALSAS ALARMAS.

PODEMOS RESUMIR TODO ESTO EN LA SIGUIENTE TABLA:

	SIN FUEGO	CON FUEGO
SIN ALARMA	NO HAY ERROR	ERROR TIPO II
CON ALARMA	ERROR TIPO I	NO HAY ERROR

AHORA IDENTIFICAMOS: $\begin{cases} H_0 = \text{SIN FUEGO} \quad H_a = \text{CON FUEGO} \\ \text{CON ALARMA} = \text{RECHAZAR } H_0 \end{cases}$

	H_0 CIERTA	H_0 FALSA
ACEPTAR H_0	NO HAY ERROR	ERROR TIPO II
RECHAZAR H_0	ERROR TIPO I	NO HAY ERROR

POTENCIA DEL CONTRASTE

POTENCIA DEL CONTRASTE = probabilidad de rechazar la hipótesis nula.

$$\boxed{\textbf{POTENCIA=P(RECHAZAR } \boldsymbol{H_0}\textbf{)}}$$

<u>INTERESA QUE:</u>

1. La función de potencia tome valores próximos a cero cuando la hipótesis nula es cierta, es decir, que la probabilidad de error de tipo I sea pequeña

$$\text{P(RECHAZAR } H_0 \mid H_0 \text{ CIERTA)=P(ERROR TIPO I)}$$

2. La función de potencia tome valores próximos a uno cuando la hipótesis nula es falsa, es decir, que la probabilidad de error de tipo II sea pequeña

$$\text{P(RECHAZAR } H_0 \mid H_0 \text{ FALSA)=1-P(ACEPTAR } H_0 \mid H_0 \text{ FALSA)}$$
$$\text{=1-P(ERROR TIPO II)}$$

OBSERVACIÓN: Para un tamaño muestral fijo, no se pueden minimizar al mismo tiempo los riesgos de cometer estos dos tipos de errores.
Puede conseguirse disminuir simultáneamente el riesgo de ambos errores aumentando el tamaño muestral, pero esto encarece el estudio estadístico.

PROCEDIMIENTO A SEGUIR:

1. Se exige que la función de potencia no supere cierto valor cuando la hipótesis nula es cierta, es decir, se fija una cota superior para la probabilidad de cometer un error de tipo I.

2. Se procura que la función de potencia sea lo mayor posible cuando la hipótesis nula es falsa, es decir, se hace mínima la probabilidad de cometer un error de tipo II.

CONSECUENCIA: Se introduce una asimetría entre la hipótesis nula y la alternativa que hace que sea importante el papel que juega cada una.

NIVEL DE SIGNIFICACIÓN

NIVEL DE SIGNIFICACIÓN = α = máxima probabilidad de cometer un error de tipo I = máxima probabilidad de rechazar la hipótesis nula cuando es cierta

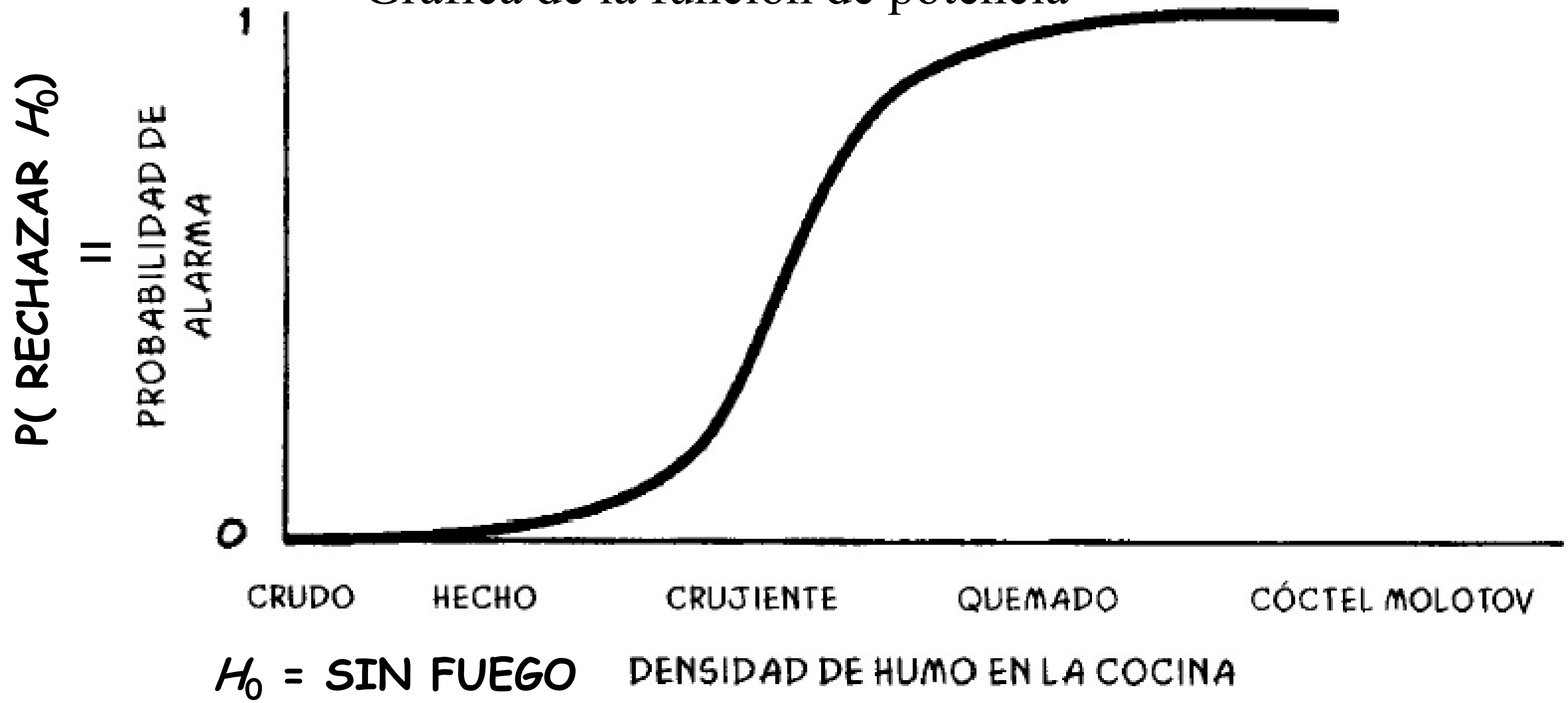
El nivel de significación es un número que se fija antes de realizar el contraste. Evidentemente, queremos que sea pequeño. Por tradición suele tomarse 0.1, 0.05 ó 0.01.

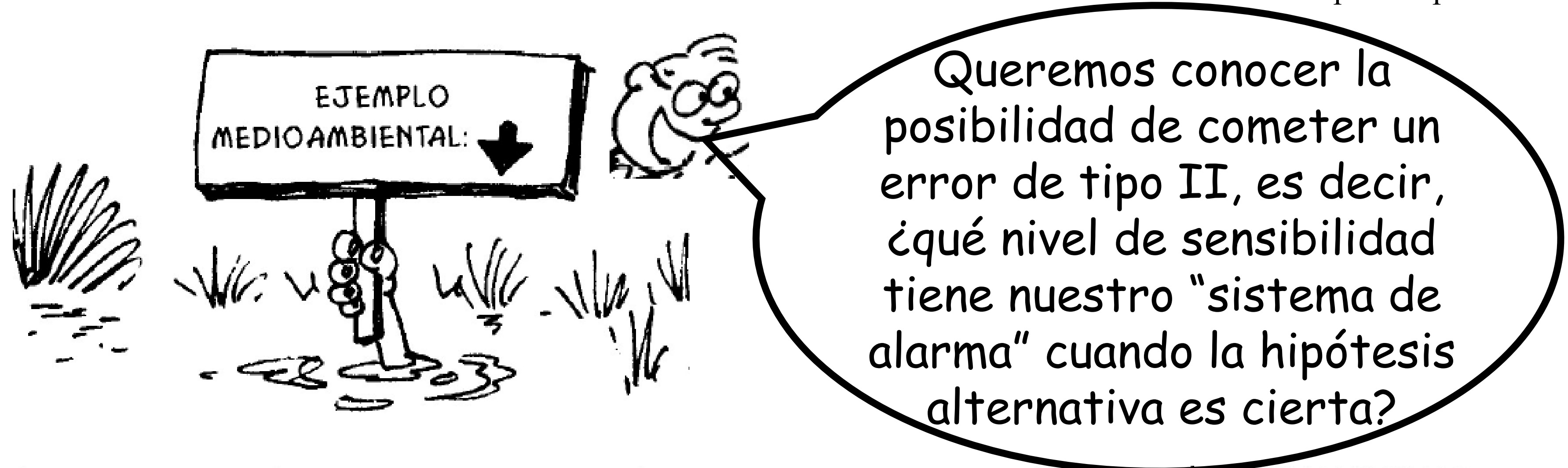
CONSECUENCIA: Los contrastes de hipótesis son muy conservadores con la hipótesis nula, es decir, para que la hipótesis nula sea rechazada tiene que haber una gran evidencia muestral en su contra.

CURVA DE POTENCIA

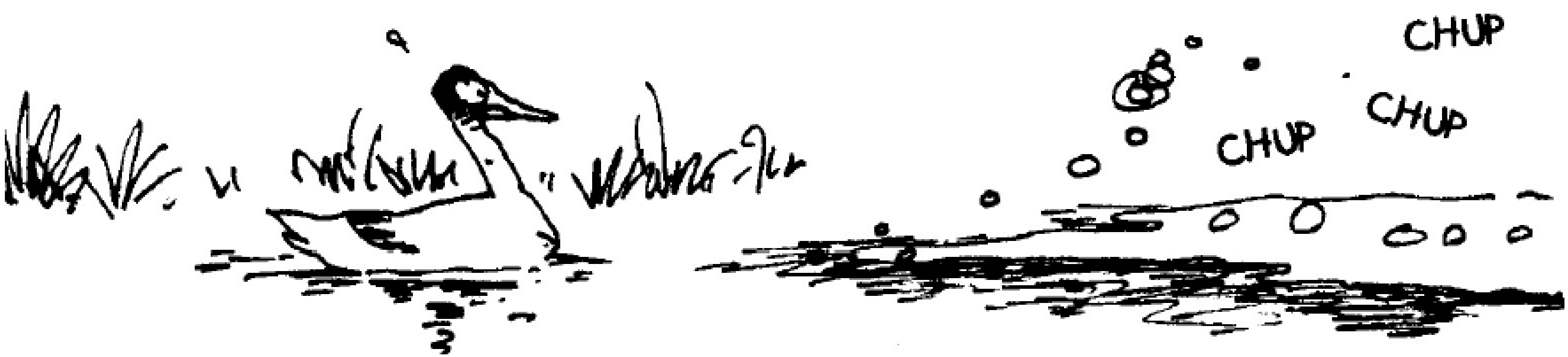
UNA FORMA DE VISUALIZAR EL EFECTO DE LOS ANÁLISIS DE POTENCIA ES DIBUJAR LA GRÁFICA DE LA PROBABILIDAD DEL RECHAZO DE H_0 Y EL ESTADO REAL DEL SISTEMA DE ALARMA. EN EL CASO DE LA ALARMA PARA HUMOS, LA PROBABILIDAD ASCIENDE HASTA 1 A MEDIDA QUE EL HUMO SE HACE MÁS DENSO.

Gráfica de la función de potencia





ANTES, LAS FÁBRICAS QUE VERTÍAN SUS DESECHOS EN LAS VÍAS FLUVIALES DEBÍAN DEMOSTRAR QUE EL VERTIDO NO TENÍA EFECTOS NOCIVOS PARA LA VIDA ANIMAL. ESTO ES LA H_0. EL QUE CONTAMINABA PODÍA SEGUIR HACIÉNDOLO HASTA QUE LA HIPÓTESIS NULA FUERA RECHAZADA POR ALCANZAR EL NIVEL DE SIGNIFICACIÓN 0,05.

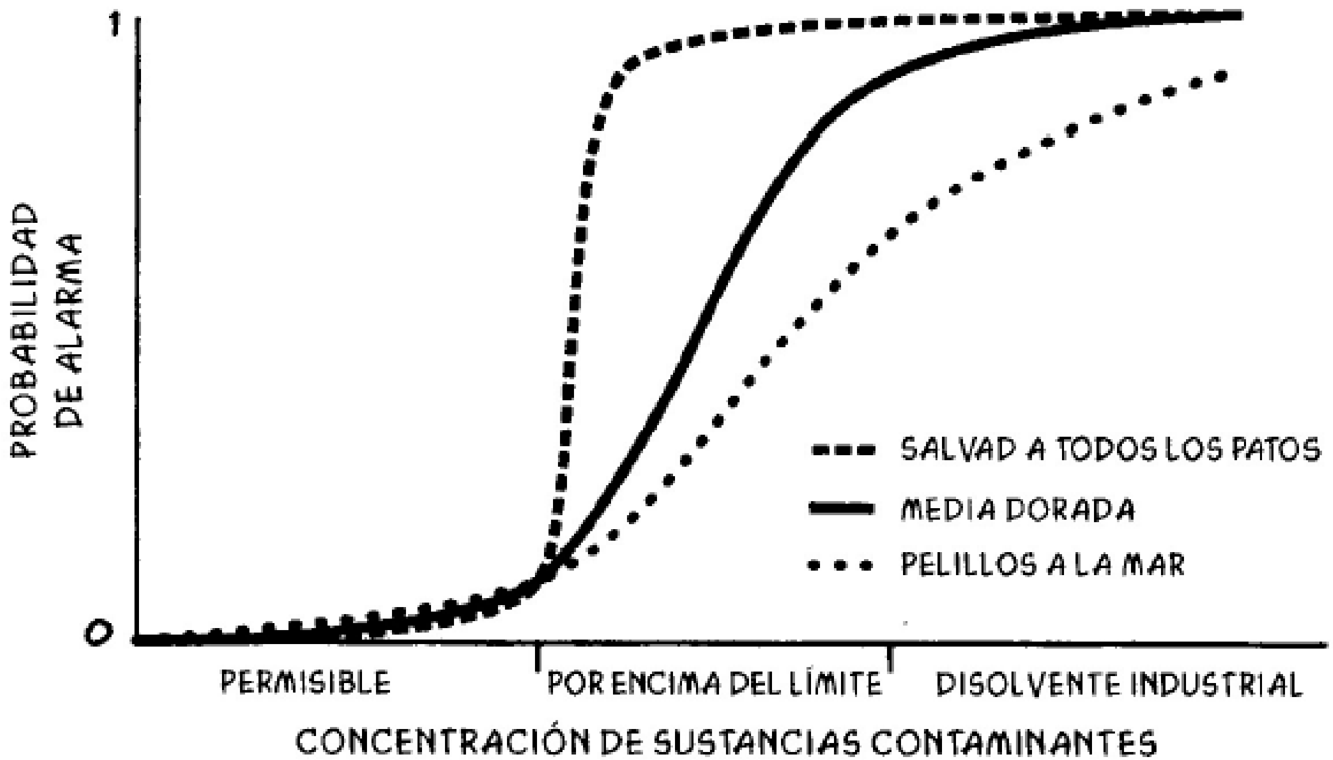


ASÍ QUE, CUANDO EL RESPONSABLE DE LA CONTAMINACIÓN CREÍA ESTAR VIOLANDO LOS LÍMITES ESTABLECIDOS POR LA AGENCIA PARA LA PROTECCIÓN DEL MEDIO AMBIENTE, LLEVABA A CABO UN PLAN DE SEGUIMIENTO DE CONTAMINACIÓN NADA EFECTIVO.

EL CULPABLE DE LA CONTAMINACIÓN SE SIENTE MUY SATISFECHO, PORQUE, COMO EN NUESTRA ALARMA DE HUMO SIN PILAS, SU ANÁLISIS TIENE POCAS POSIBILIDADES O NINGUNA DE HACER SALTAR LA ALARMA.


MM...
SE ME HA DESINTEGRADO LA BOTA.
TE GUSTARÁ SABER QUE LOS LEGISLADORES MEDIOAMBIENTALES CADA VEZ EXIGEN MÁS PROGRAMAS DE SEGUIMIENTO PARA DEMOSTRAR QUE TIENEN UNA PROBABILIDAD MUY ALTA DE DETECTAR GRAVES CASOS DE CONTAMINACIÓN. EL **ANÁLISIS DE POTENCIA** REVELA A MENUDO DEFECTOS OCULTOS EN LOS PROGRAMAS DE SEGUIMIENTO.

AQUÍ, ESTÁN REPRESENTADAS LAS CURVAS DE EFECTIVIDAD DE LOS TRES PROGRAMAS DE SEGUIMIENTO. LA DE **SALVAD A TODOS LOS PATOS** (CON UN COSTE DE 5 MILLONES DE DÓLARES), LA **MEDIA DORADA** (CON UN COSTE DE 500.000 DÓLARES) Y **PELILLOS A LA MAR** (TAMBIÉN, CON UN COSTE DE 500.000 DÓLARES). CUANTO MAYOR SEA LA POTENCIA DE LA PRUEBA MAYOR SERÁ LA PRONUNCIACIÓN DE LA CURVA.



H_0 = EL VERTIDO NO TIENE EFECTOS NOCIVOS PARA LA VIDA ANIMAL

ELECCIÓN DE LA HIPÓTESIS NULA

SITUACIÓN SENCILLA: Queremos contrastar

$\theta = \theta_0$ frente a $\theta \neq \theta_0$.

Se toma H_0: $\theta = \theta_0$.

SITUACIÓN DELICADA: Queremos contrastar

$\theta \leq \theta_0$ frente a $\theta > \theta_0$.

Puesto que los contrastes de hipótesis son conservadores con la hipótesis nula, ya que ésta no debe ser rechazada sin una gran evidencia en contra, debemos tomar como hipótesis alternativa aquella que deseemos "probar" estadísticamente, es decir, la hipótesis nula debe de ser la contraria de la que queremos "probar" estadísticamente.

EJEMPLO

Unos laboratorios farmacéuticos están estudiando un nuevo medicamento contra cierta enfermedad, que lanzarán al mercado únicamente si cura la enfermedad en más del 80% de los casos. ¿Qué deberíamos tomar como hipótesis nula?

Si llamamos p a la proporción de casos en los que el medicamento cura la enfermedad, los laboratorios sólo lanzarán el producto al mercado si tienen una fuerte evidencia muestral de que el producto es efectivo, es decir, si consiguen "probar" estadísticamente que $p>0.8$. En consecuencia, deberíamos tomar como hipótesis nula

$$H_0\text{: p} \leq 0.8$$

ESTADÍSTICO DEL CONTRASTE

Para contrastar una hipótesis nula frente a una alternativa se utiliza un contraste de hipótesis, el cual basa su decisión en la evidencia aportada por una muestra de tamaño n, o más en concreto, en una función suya T, llamada **ESTADÍSTICO DEL CONTRASTE**.

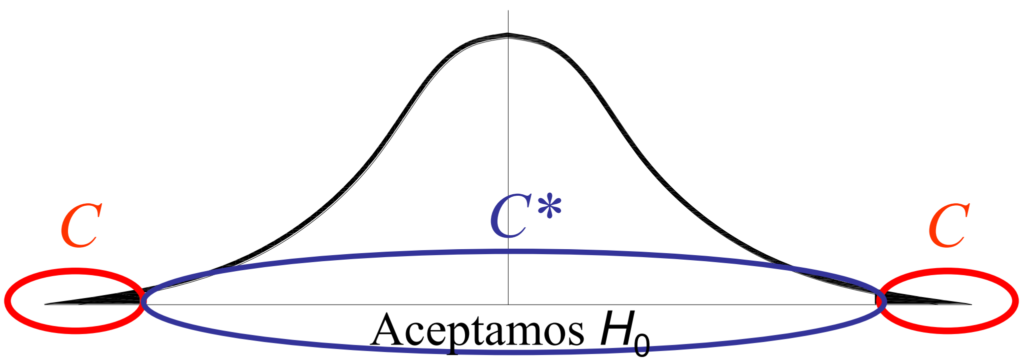
Habitualmente, el estadístico del contraste es una función del estimador natural, que vimos en el tema “estimación puntual”, asociado al parámetro poblacional al que hacen referencia las hipótesis del contraste.

El uso del estimador natural suele dar resultados óptimos, en el sentido de que se minimiza el riesgo de cometer errores al tomar una decisión.

REGIÓN CRÍTICA Y REGIÓN DE ACEPTACIÓN

REGIÓN CRÍTICA = C = conjunto de valores de T donde se rechaza la hipótesis nula

REGIÓN DE ACEPTACIÓN = C^* = conjunto de valores de T donde se acepta la hipótesis nula



OBSERVACIÓN: $C \cup C^*$= conjunto de posibles valores de T

OBSERVACIÓN 1: C es el conjunto de valores improbables de T, cuando se supone cierta la hipótesis nula.

OBSERVACIÓN 2: C^* es el conjunto complementario de C.

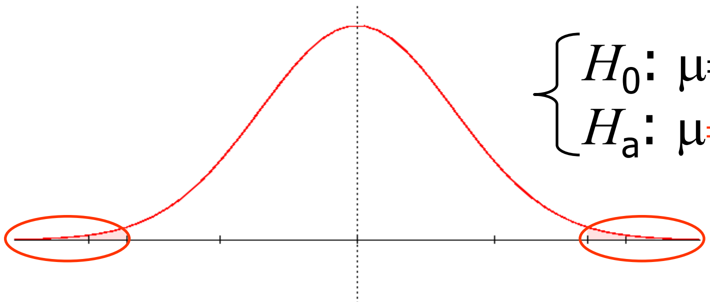
OBSERVACIÓN 3: C es conocida antes de realizar el contraste, es decir, antes de extraer una m.a.s. de la población que refutarían o confirmarían la hipótesis nula.

OBSERVACIÓN 4: C depende del nivel de significación y de la hipótesis alternativa.

CONTRASTES BILATERALES Y CONTRASTES UNILATERALES

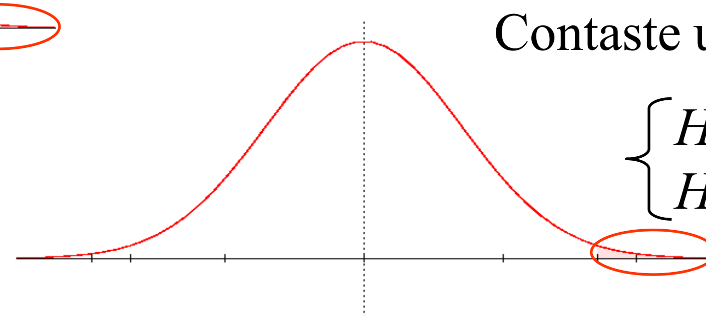
Contaste bilateral

$$\begin{cases} H_0: \mu = \mu_0 \\ H_a: \mu \neq \mu_0 \end{cases}$$



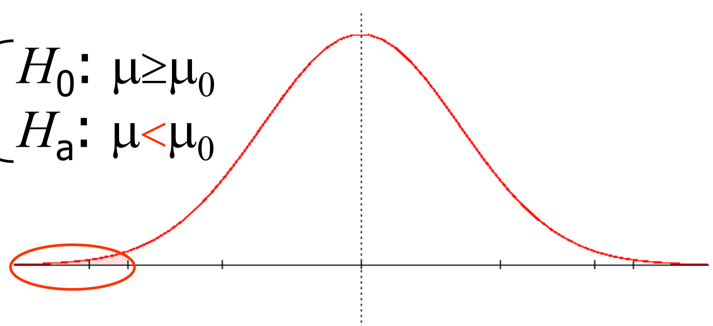
Contaste unilateral

$$\begin{cases} H_0: \mu \leq \mu_0 \\ H_a: \mu > \mu_0 \end{cases}$$



Contaste unilateral

$$\begin{cases} H_0: \mu \geq \mu_0 \\ H_a: \mu < \mu_0 \end{cases}$$



EJEMPLO: Se quiere averiguar si el consumo habitual de cierto producto modifica el nivel de colesterol. Si denotamos por μ al nivel medio de colesterol en sangre de los consumidores del producto, el problema es decidir si coincide con el nivel medio en el conjunto de personas sanas que es de 200 mg/dl.

$$\begin{cases} H_0 : \mu = 200 \\ H_a : \mu \neq 200 \end{cases}$$

Para ello se eligen al azar n personas a las que, tras un consumo prolongado del producto en cuestión, se mide su nivel de colesterol.

Si la hipótesis nula es cierta, el estimador natural de μ, la media muestral, debe tomar un valor cercano a $200=\mu_0$, por lo tanto, rechazaremos H_0 si $\bar{x}$ queda "lejos" de 200, es decir,

$C = \{|\bar{x} - \mu_0| > c\}$, para cierto valor de c a determinar.

El grado de "alejamiento" permitido dependerá del tamaño muestral *n*, y del nivel de siginifcación α, fijado, ya que

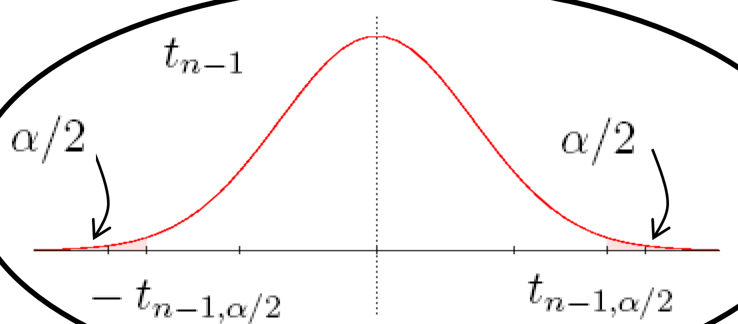
$$\alpha = P(\texttt{RECHAZAR } H_0 | H_0 \texttt{ CIERTA}) = P(\overline{X} \in C)$$

$$\Updownarrow$$

Si suponemos normalidad en los datos, $\dfrac{\overline{X}-\mu}{S\sqrt{n}} \rightsquigarrow t_{n-1}$

$$\alpha = P\left(|t_{n-1}| > c\,\frac{\sqrt{n}}{s}\right)$$

$$\Downarrow$$



$$c = t_{n-1,\alpha/2}\,\frac{s}{\sqrt{n}}$$

$$\Downarrow$$

$$\boxed{C = \left\{ |\bar{x} - \mu_0| > t_{n-1,\alpha/2}\,\frac{s}{\sqrt{n}} \right\}}$$

En resumen, bajo las condiciones del ejemplo, el contraste de hipótesis al nivel de significación α para contrastar

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \neq \mu_0 \end{cases} \text{ consiste en } \begin{cases} \texttt{rechazar } H_0 \texttt{ si } \dfrac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > t_{n-1,\alpha/2} \\ \texttt{aceptar } H_0 \texttt{ si } \dfrac{|\bar{x} - \mu_0|}{s/\sqrt{n}} \leq t_{n-1,\alpha/2} \end{cases}$$

En concreto, si en el ejemplo del colesterol fijamos α=0.05, y extraemos una m.a.s. de tamaño 10 con media muestral 202 y cuasivarianza muestral 289, tenemos que

$$\frac{|202 - 200|}{\sqrt{289/10}} = 0.372 < t_{9,0.025} = 2.262$$

$$\Downarrow$$

$$\boxed{\text{Aceptamos } H_0 \text{ al nivel de significación } 0.05}$$

<u>EJEMPLO:</u> Si el producto considerado en el ejemplo anterior es un aperitivo elaborado con cierto aceite, nadie se preocupará porque su consumo pueda disminuir el nivel de colesterol, por lo que las hipótesis a contrastar son:

$$\begin{cases} H_0 : \mu \leq 200 \\ H_a : \mu > 200 \end{cases}$$

Únicamente se rechazará la hipótesis nula si la media muestral es "mucho mayor" de $200 = \mu_0$, es decir,

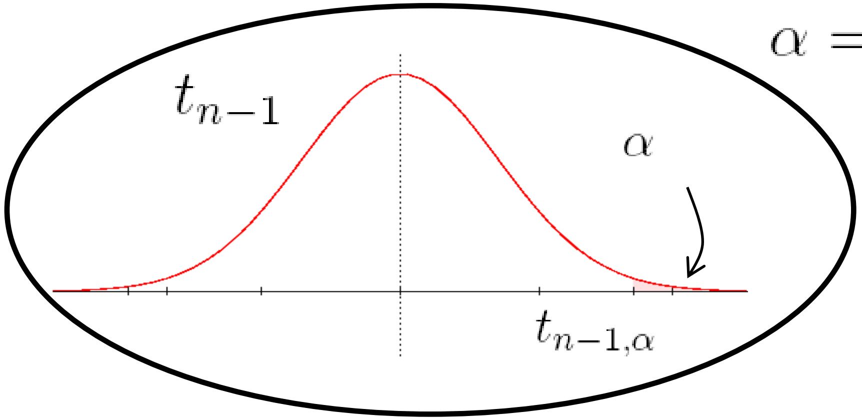
$C = \{\bar{x} > \mu_0 + c\}$, para cierto valor de c a determinar.

Para un nivel de siginifcación α, debe de verificarse:

$$\alpha = P(\overline{X} \in C) = P(\overline{X} > \mu_0 + c)$$

$$\Updownarrow$$

$$\alpha = P\left(t_{n-1} > c\,\frac{\sqrt{n}}{s}\right)$$



$$\Downarrow$$

$$c = t_{n-1,\alpha}\,\frac{s}{\sqrt{n}}$$

$$\Downarrow$$

$$\boxed{C = \left\{\bar{x} > \mu_0 + t_{n-1,\alpha}\,\frac{s}{\sqrt{n}}\right\}}$$

En resumen, bajo las condiciones del ejemplo, el contraste de hipótesis al nivel de significación α para contrastar

$$\left\{\begin{array}{l} H_0 : \mu \leq \mu_0 \\ H_a : \mu > \mu_0 \end{array}\right. \text{ consiste en } \left\{\begin{array}{l} \texttt{rechazar } H_0 \texttt{ si } \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{n-1,\alpha} \\[2ex] \texttt{aceptar } H_0 \texttt{ si } \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}} \leq t_{n-1,\alpha} \end{array}\right.$$

En concreto, si fijamos α=0.05, y extraemos una m.a.s. de tamaño 10 con media muestral 202 y cuasivarianza muestral 289, tenemos que

$$\frac{202-200}{\sqrt{289/10}} = 0.372 < t_{9,0.05} = 1.833$$

$$\Downarrow$$

$\boxed{\text{Aceptamos } H_0 \text{ al nivel de significación } 0.05}$

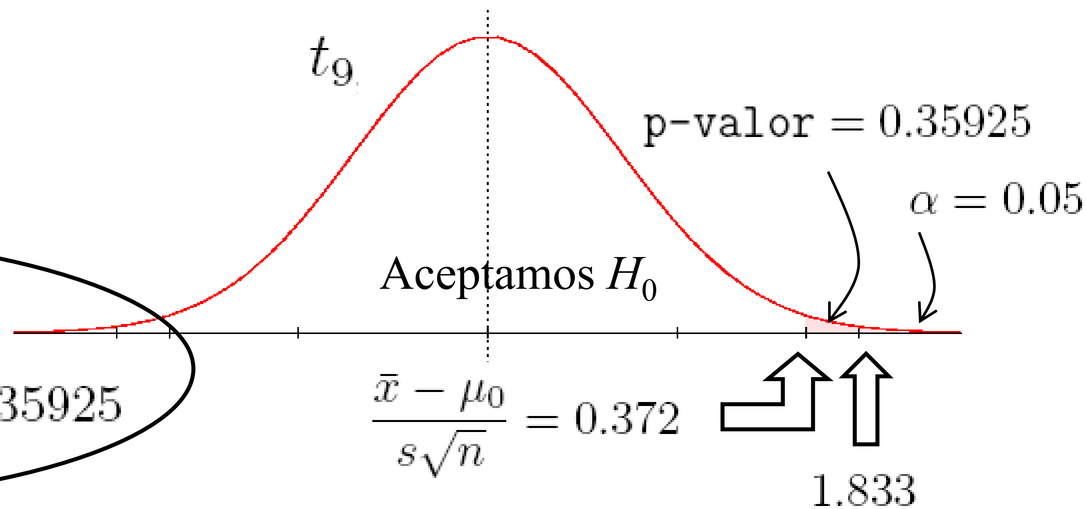
EL P-VALOR

El P-VALOR de una muestra es la probabilidad de obtener un resultado menos compatible con la hipótesis nula que el proporcionado por la muestra.

<u>EJEMPLO ANTERIOR:</u>

H_0 puede rechazarse con un

$\texttt{p-valor} = P(t_9 > 0.372) = 0.35925$



P-VALOR = mínimo nivel de significación que hace que rechacemos H_0

Cuando no se tiene claro qué valor de α fijar, es decir, qué riesgo se está dispuesto a correr al rechazar H_0 cuando es cierta, se suele adjuntar a la decisión de rechazar H_0 el riesgo que ello conlleva, es decir, el p-valor.

OBSERVACIÓN 1: En general, el cálculo del p-valor no es fácil sin un ordenador.

OBSERVACIÓN 2:

-Si $\alpha \geq$p-valor, se deber rechazar H_0 al nivel de significación α

-Si $\alpha <$p-valor, se deber aceptar H_0 al nivel de significación α

OBSERVACIÓN 3: El p-valor no se conoce hasta que se ha extraído una m.a.s. de la población, mientras que el nivel de significación se fija antes de extraer ninguna muestra.

OBSERVACIÓN 4:

-Si p-valor<0.01, se deber rechazar H_0

-Si p-valor>0.1, se deber aceptar H_0

-Si $0.01 \leq$ p-valor≤ 0.1, el resultado es ambiguo. Es recomendable extraer otra muestra y volver a realizar el contraste.

CONTRASTES ÓPTIMOS

La determinación del contraste de hipótesis óptimo dependerá de las suposiciones que se hagan sobre la población: normalidad, varianza conocida o desconocida, muestra grande o pequeña; y también de las hipótesis que se deseen contrastar.

A continuación veremos contrastes que la estadística propone como óptimos en cada una de las situaciones que se consideran, para una y para dos poblaciones.

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA μ DE UNA POBLACIÓN NORMAL CON σ CONOCIDA

$$
\begin{cases}
\texttt{- Si } H_0 : \mu = \mu_0, \qquad C = \left\{ \dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2} \right\} \\
\texttt{- Si } H_0 : \mu \leq \mu_0, \qquad C = \left\{ \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha} \right\} \\
\texttt{- Si } H_0 : \mu \geq \mu_0, \qquad C = \left\{ \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_{1-\alpha} = -z_{\alpha} \right\}
\end{cases}
$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA μ DE UNA POBLACIÓN NORMAL CON σ DESCONOCIDA

$$\begin{cases} \texttt{- Si}\ \ H_0 : \mu = \mu_0, & C = \left\{ \dfrac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > t_{n-1,\alpha/2} \right\} \\ \texttt{- Si}\ \ H_0 : \mu \leq \mu_0, & C = \left\{ \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha} \right\} \\ \texttt{- Si}\ \ H_0 : \mu \geq \mu_0, & C = \left\{ \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} < t_{n-1,1-\alpha} = -t_{n-1,\alpha} \right\} \end{cases}$$

EJEMPLO 1: Hace 10 años se realizó un estudio sobre la estatura de los individuos de una población, cuyo histograma sugirió para dicha variable una distribución normal de media 1.68 m y desviación típica de 0.064 m.
Ahora se quiere analizar si la estatura media de dicha población ha variado con el tiempo, para lo cual se tomó una muestra de tamaño 15, de la que se obtuvo una media muestral de 1.73 m. y una cuasivarianza muestral de 0.415 .

(A) Si admitimos que la dispersión de dicha población no ha variado en estos 10 años, ¿podemos afirmar que la estatura no ha variado?

(B) ¿Y si no tenemos la certeza de que la dispersión de dicha población no haya variado en los 10 años?

$$\begin{cases} H_0 : \mu = 1.68 \\ H_a : \mu \neq 1.68 \end{cases}$$

(A) Elegimos como nivel de significación $\alpha=0.05 \Rightarrow$ $\alpha/2=0.025$

$$\frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} = \frac{|1.73-1.68|}{0.064/\sqrt{15}} = 3.026 > z_{0.025} = 1.96$$

$\Downarrow$

Rechazamos H_0 al nivel de significación 0.05

$$\texttt{p-valor} = P(|Z| > 3.026) = 2P(Z > 3.026) = 0.0025$$

Confirma la decisión de rechazar H_0

(B) $$\frac{|\bar{x}-\mu_0|}{s/\sqrt{n}} = \frac{|1.73-1.68|}{\sqrt{41.5/15}} = 0.03 < t_{14,0.025} = 2.145$$

$\Downarrow$

Aceptamos H_0 al nivel de significación 0.05

$$\texttt{p-valor} = P(|t_{14}| > 0.03) = 2P(t_{14} > 0.03) = 0.97649$$

Confirma la decisión de aceptar H_0

<u>**EJEMPLO 2:**</u> Un laboratorio piensa que un nuevo procedimiento de fabricación alarga la duración de ciertos componentes electrónicos, que era hasta la actualidad de dos años de media.
Con objeto de comprobarlo, se extrajo una m.a.s. de 18 unidades, que proporcionó una media muestral de 2.8 años y una cuasidesviación típica de 1.2 años.
Bajo la hipótesis de que la duración de los componentes electrónicos sigue una distribución normal, ¿podemos afirmar que los laboratorios tienen razón?

$$\begin{cases} H_0 : \mu \leq 2 \\ H_a : \mu > 2 \end{cases} \qquad \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2.8 - 2}{1.2/\sqrt{18}} = 2.8284$$

$$\Downarrow$$

$$\texttt{p-valor} = P(t_{17} > 2.8284) = 0.00579$$

Confirma la decisión de rechazar H_0

Podemos afirmar que los laboratorios tienen razón.

EJEMPLO 3: Cierto disco duro tiene un tiempo de acceso de 30 milisegundos. La empresa cree que un nuevo sistema de producción reduce significativamente dicho tiempo.
Para contrastar esta creencia, se extrae una m.a.s. de 10 unidades producidas por el nuevo sistema, con una media y una cuasivarianza muestral de 28 milisegundos y 16 milisegundos al cuadrados respectivamente.
Si suponemos normalidad en los datos, ¿podemos afirmar con un nivel de significación del 0.05 que la creencia de la empresa es correcta?

$$\left\{\begin{array}{l} H_0 : \mu \geq 30 \\ H_a : \mu < 30 \end{array}\right. \quad \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{28-30}{4/\sqrt{10}} = -1.58 > t_{9,0.95} = -1.833$$

$$\Downarrow$$

No podemos afirmar que la creencia de la empresa sea correcta.

$\boxed{\text{Con un nivel del 0.05, no podemos rechazar } H_0}$

EJEMPLO 4: La longitud craneal es un factor determinante en la clasificación de restos humanos del paleolítico. En concreto, una longitud craneal de 63 cm lleva a clasificar un esqueleto como de raza Neandertal.
Después del hallazgo de 30 esqueletos en una necrópolis, los arqueólogos determinaron una longitud craneal media de 59 cm y una desviación típica muestral de 5cm.
¿Puede concluirse que los restos son de raza Neandertal?

$$\begin{cases} H_0 : \mu = 63 \\ H_a : \mu \neq 63 \end{cases} \qquad \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} = \frac{|59 - 63|}{5.085/\sqrt{30}} = 4.309$$

$$s = 5\sqrt{30/29} = 5.085$$

$$\Downarrow$$

$$\texttt{p-valor} = P(|Z| > 4.309) = 2P(Z > 4.309) = 0.00002$$

$$\Downarrow$$

Rechazamos H_0, es decir, rechazamos que los restos sean de raza Neandertal.

EJEMPLO 5: En una muestra de 49 estudiantes de un centro, sometidos a un cierto test, la media y la desviación típica muestrales de las puntuaciones obtenidas fueron 39 y 11 respectivamente.
Sabiendo que la puntuación media habitual es 30, nos preguntamos si los estudiantes del centro tienen una habilidad superior a la usual en la resolución del test.

$$\begin{cases} H_0 : \mu \leq 30 \\ H_a : \mu > 30 \end{cases} \qquad \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{39 - 30}{11.11/\sqrt{49}} = 5.6685$$

$$s = 11\sqrt{49/48} = 11.11$$

$$\Downarrow$$

$$\texttt{p-valor} = P(Z > 5.6685) = 7.2(10)^{-9}$$

$$\Downarrow$$

Rechazamos H_0, y por lo tanto, aceptamos que los estudiantes del centro tienen una habilidad superior.

Contrastes para la diferencia de las medias

Sentido Común

Datos Independientes

Test Igualdad De varianzas

$\sigma_1 = \sigma_2$

Test diferencia De medias para varianzas iguales

$\sigma_1 \neq \sigma_2$

Test diferencia De medias para varianzas diferentes

Datos Emparejados

Se investiga el comportamiento de la Variable D=X-Y

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA EL COCIENTE DE LAS VARIANZAS DE DOS POBLACIONES NORMALES INDEPENDIENTES CON MEDIAS DESCONOCIDAS

$$\begin{cases} \text{- Si } H_0 : \sigma_1^2 = \sigma_2^2, \qquad C = \left\{ \dfrac{s_1^2}{s_2^2} \notin \left[F_{n_1-1, n_2-1, 1-\alpha/2}, F_{n_1-1, n_2-1, \alpha/2} \right] \right\} \\ \text{- Si } H_0 : \sigma_1^2 \leq \sigma_2^2, \qquad C = \left\{ \dfrac{s_1^2}{s_2^2} > F_{n_1-1, n_2-1, \alpha} \right\} \\ \text{- Si } H_0 : \sigma_1^2 \geq \sigma_2^2, \qquad C = \left\{ \dfrac{s_1^2}{s_2^2} < F_{n_1-1, n_2-1, 1-\alpha} \right\} \end{cases}$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA EL COCIENTE DE LAS VARIANZAS DE DOS POBLACIONES NORMALES INDEPENDIENTES CON MEDIAS CONOCIDAS

$$\begin{cases} \texttt{- Si } H_0: \sigma_1^2 = \sigma_2^2, & C = \left\{ \dfrac{\sum_{i=1}^{n_1} (x_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (y_i - \mu_2)^2 / n_2} \notin \left[F_{n_1, n_2, 1-\alpha/2}, F_{n_1 - 1, n_2 - 1, \alpha/2} \right] \right\} \\ \texttt{- Si } H_0: \sigma_1^2 \leq \sigma_2^2, & C = \left\{ \dfrac{\sum_{i=1}^{n_1} (x_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (y_i - \mu_2)^2 / n_2} > F_{n_1, n_2, \alpha} \right\} \\ \texttt{- Si } H_0: \sigma_1^2 \geq \sigma_2^2, & C = \left\{ \dfrac{\sum_{i=1}^{n_1} (x_i - \mu_1)^2 / n_1}{\sum_{i=1}^{n_2} (y_i - \mu_2)^2 / n_2} < F_{n_1, n_2, 1-\alpha} \right\} \end{cases}$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA DIFERENCIA DE LAS MEDIAS DE DOS POBLACIONES NORMALES INDEPENDIENTES CON VARIANZAS CONOCIDAS

$$\left\{\begin{array}{ll}
\texttt{- Si}\ \ H_0:\mu_1=\mu_2, & C=\left\{\dfrac{|\overline{x}_1-\overline{x}_2|}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}>z_{\alpha/2}\right\} \\
\texttt{- Si}\ \ H_0:\mu_1\leq\mu_2, & C=\left\{\dfrac{\overline{x}_1-\overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}>z_{\alpha}\right\} \\
\texttt{- Si}\ \ H_0:\mu_1\geq\mu_2, & C=\left\{\dfrac{\overline{x}_1-\overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}<z_{1-\alpha}\right\}
\end{array}\right.$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA DIFERENCIA DE LAS MEDIAS DE DOS POBLACIONES NORMALES INDEPENDIENTES CON VARIANZAS DESCONOCIDAS E IGUALES

$$\left\{\begin{array}{ll}
\text{- Si } H_0: \mu_1 = \mu_2, & C = \left\{ \dfrac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} > t_{n_1+n_2-2,\alpha/2} \right\} \\
\text{- Si } H_0: \mu_1 \leq \mu_2, & C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} > t_{n_1+n_2-2,\alpha} \right\} \\
\text{- Si } H_0: \mu_1 \geq \mu_2, & C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} < t_{n_1+n_2-2,1-\alpha} \right\}
\end{array}\right.$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA DIFERENCIA DE LAS MEDIAS DE DOS POBLACIONES NORMALES INDEPENDIENTES CON VARIANZAS DESCONOCIDAS Y DISTINTAS

$$\begin{cases} \text{- Si } H_0 : \mu_1 = \mu_2, \qquad C = \left\{ \dfrac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} > t_{f,\alpha/2} \right\} \\ \text{- Si } H_0 : \mu_1 \leq \mu_2, \qquad C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} > t_{f,\alpha} \right\} \\ \text{- Si } H_0 : \mu_1 \geq \mu_2, \qquad C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} < t_{f,1-\alpha} \right\} \end{cases}$$

$f =$ entero más próximo a

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} - 2$$

EJEMPLO 8: Con el objeto de averiguar si difieren significativamente los pesos de los individuos de dos poblaciones independientes, se eligieron al azar 13 individuos de la primera y 10 de la segunda, obteniéndose los siguientes resultados:

$$\overline{x}_1 = 63.9 \qquad \overline{x}_2 = 67.8 \qquad s_1 = 9.16 \qquad s_2 = 8.37$$

Con una significación del 0.1, ¿qué conclusión podemos extraer de estos datos?

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

¿Debemos suponer varianzas iguales?

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_a : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

$$F_{12,9,0.95} = \frac{1}{F_{9,12,0.05}} = 0.3576$$

$$\frac{s_1^2}{s_2^2} = \frac{(9.16)^2}{(8.37)^2} = 1.198 \in [F_{12,9,0.95}, F_{12,9,0.05}] = [0.3576, 3.0729]$$

Aceptamos que las varianzas son iguales

$$\frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$= \frac{|63.9 - 67.8|}{\sqrt{\frac{12(9.16)^2 + 9(8.37)^2}{21}}\sqrt{\frac{1}{13}+\frac{1}{10}}} = 1.05 < t_{21,0.05} = 1.72$$

Aceptamos H_0, es decir, no hay diferencias significativas.

$$\texttt{p-valor} = p\left(|t_{21}| > 1.05\right) = 2p\left(t_{21} > 1.05\right) = 0.3$$

<u>EJEMPLO 9:</u> Un experimento realizado por la universidad de Harvard, pretendía averiguar si el consumo habitual de aspirinas reducía el riesgo de infarto.

AL GRUPO 1 SE LE ADMINISTRA UN PLACEBO. UNA PASTILLA IDÉNTICA A LA ASPIRINA PERO QUE NO CONTIENE ASPIRINA.

AL GRUPO 2 SE LE ADMINISTRA UNA ASPIRINA DIARIA.

A la vista de los datos de la siguiente tabla, ¿qué podemos afirmar?

	INFARTO	NO INFARTO	n	INCIDENCIA DE INFARTO
PLACEBO	**239**	**10.795**	**11.034**	$\hat{p}_1 = \frac{239}{11.034} = \mathbf{0{,}0217}$
ASPIRINA	**139**	**10.898**	**11.037**	$\hat{p}_2 = \frac{139}{11.037} = \mathbf{0{,}0126}$

$$\Rightarrow \bar{p} = \frac{378}{22.071}$$

H_0. LA HIPÓTESIS NULA, ES QUE LA ASPIRINA NO TIENE EFECTO: $p_1 \leq p_2$.

H_a. LA HIPÓTESIS ALTERNATIVA, ES QUE LA ASPIRINA REDUCE LA INCIDENCIA DE INFARTO: $p_1 > p_2$.

$$T = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

ENTONCES

$$Z_{OBS} = \frac{0{,}0091}{0{,}00175} = 5{,}20$$

VALOR P = PR($Z \geq Z_{OBS}$) = PR($Z \geq 5{,}2$) = **0,0000001**

SI LA HIPÓTESIS NULA FUERA VERDADERA, LA PROBABILIDAD DE OBSERVAR UN EFECTO ASÍ DE GRANDE SERÍA DE **UNA ENTRE DIEZ MILLONES**. ¡ES UNA PRUEBA DE MUCHO PESO CONTRA H_0!

CONTRASTES PARA DATOS APAREADOS

EJEMPLO 10: El propietario de una gran flota de taxis quiere comparar la cantidad de gasolina de dos tipos que es consumida por sus taxis.
A la vista de los datos de la siguiente tabla, ¿qué podemos afirmar?

TAXI	GASOLINA A	GASOLINA B
1	27,01	26,95
2	20,00	20,44
3	23,41	25,05
4	25,22	26,32
5	30,11	29,56
6	25,55	26,60
7	22,23	22,93
8	19,78	20,23
9	33,45	33,95
10	25,22	26,01

TAXI	GASOLINA A	GASOLINA B	DIFERENCIA
1	27,01	26,95	0,06
2	20,00	20,44	−0,44
3	23,41	25,05	−1,64
4	25,22	26,32	−1,10
5	30,11	29,56	0,55
6	25,55	26,60	−1,05
7	22,23	22,93	−0,70
8	19,78	20,23	−0,45
9	33,45	33,95	−0,50
10	25,22	26,01	−0,79
MEDIA	25,20	25,80	−0,60
DESVIACIÓN TÍPICA	4,27	4,10	0,61

H_0: $\mu_d = 0$

H_a: $\mu_d \neq 0$

Si suponemos normalidad en los datos...

$$T = \frac{|\bar{d} - (\mu_d)_0|}{s/\sqrt{n}}$$

$$\text{VALOR P} = Pr(|t| \geq |t_{OBS}|) = Pr\left(|t| \geq \frac{0,6}{0,19}\right) = Pr(|t| \geq 3,15) = 0,012 < 0,05$$

Rechazamos H_0, es decir, los datos evidencian una diferencia en el consumo de los dos tipos de gasolina

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA μ DE UNA POBLACIÓN NO NECESARIAMENTE NORMAL CON σ CONOCIDA Y TAMAÑO MUESTRAL GRANDE

$$\begin{cases} \text{- Si } H_0 : \mu = \mu_0, \quad \mathrm{C} = \left\{ \dfrac{|\,\bar{x} - \mu_0\,|}{\sigma / \sqrt{n}} > z_{\alpha/2} \right\} \\ \text{- Si } H_0 : \mu \leq \mu_0, \quad \mathrm{C} = \left\{ \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} > z_{\alpha} \right\} \\ \text{- Si } H_0 : \mu \geq \mu_0, \quad \mathrm{C} = \left\{ \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} < z_{1-\alpha} = -z_{\alpha} \right\} \end{cases}$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA μ DE UNA POBLACIÓN NO NECESARIAMENTE NORMAL CON σ DESCONOCIDA Y TAMAÑO MUESTRAL GRANDE

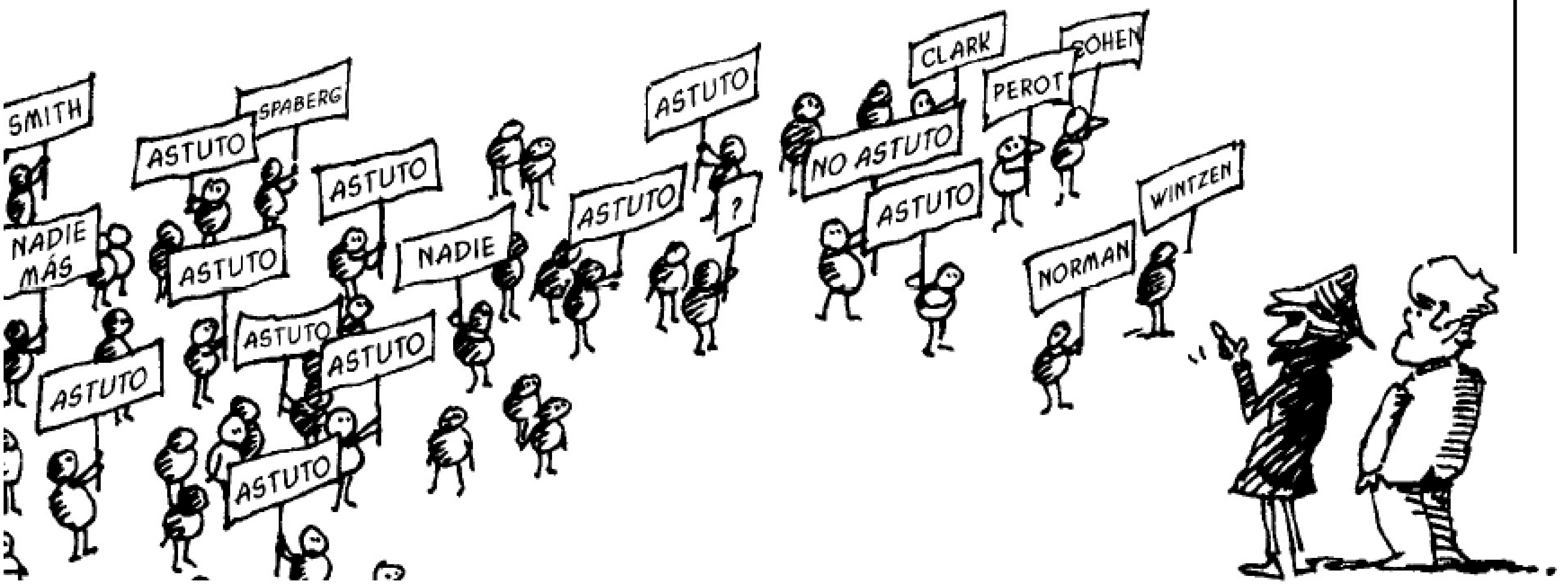
$$
\begin{cases}
\text{-Si } H_0 : \mu = \mu_0, \quad \mathrm{C} = \left\{ \dfrac{|\bar{x} - \mu_0|}{s / \sqrt{n}} > z_{\alpha/2} \right\} \\
\text{-Si } H_0 : \mu \leq \mu_0, \quad \mathrm{C} = \left\{ \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} > z_{\alpha} \right\} \\
\text{-Si } H_0 : \mu \geq \mu_0, \quad \mathrm{C} = \left\{ \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} < z_{1-\alpha} = -z_{\alpha} \right\}
\end{cases}
$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA EL PARÁMETRO p DE UNA POBLACIÓN BINOMIAL CON MUESTRAS GRANDES

$$\begin{cases} \text{- Si } H_0 : p = p_0, & C = \left\{ \dfrac{|\widehat{p} - p_0|}{\sqrt{p_0(1-p_0)/n}} > z_{\alpha/2} \right\} \\ \text{- Si } H_0 : p \leq p_0, & C = \left\{ \dfrac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} > z_{\alpha} \right\} \\ \text{- Si } H_0 : p \geq p_0, & C = \left\{ \dfrac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} < z_{1-\alpha} = -z_{\alpha} \right\} \end{cases}$$

EJEMPLO 6: El senador Astuto quiere averiguar si obtendrá mayoría absoluta en las próximas elecciones. Para ello seleccionó una m.a.s. de 1000 habitantes a los que les preguntó si tenían intención de votarle o no, a lo que el 55% respondió que sí.

A la vista de estos datos, ¿podemos afirmar que el senador Astuto obtendrá mayoría absoluta?



1) LAS HIPÓTESIS SON:

$$H_0 : p \leq 0{,}5$$

$$H_a : p > 0{,}5$$

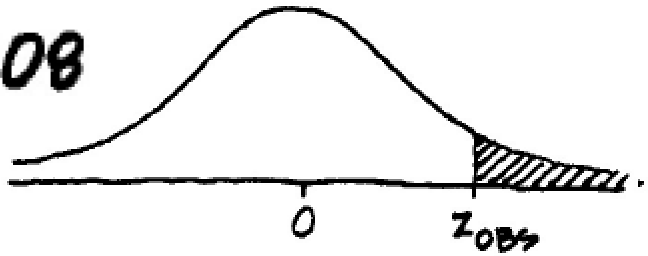
RECUERDA: La hipótesis alternativa debe de ser la que quieres probar estadísticamente.

2) LA PRUEBA ESTADÍSTICA ES:

$$T = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} \Rightarrow z_{OBS} = \frac{0{,}55 - 0{,}50}{\sqrt{(0{,}5)(0{,}5)}/\sqrt{1.000}} = 3{,}16$$

3) EL VALOR p ES:

$$Pr(Z > z_{OBS}) = Pr(z \geq 3{,}16) = 0{,}0008$$



4) ASTUTO, QUE ES BASTANTE CONSERVADOR, TOMA UN NIVEL DE SIGNIFICACIÓN α DE 0,01 Y OBSERVA QUE

$$Pr(Z > z_{OBS}) = 0{,}0008 < \alpha$$

ASÍ, EL SENADOR RECHAZA LA HIPÓTESIS NULA. ÉL (Y SUS PARTIDARIOS) PUEDEN ESTAR SEGUROS DE SU VENTAJA ELECTORAL.

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA EL PARÁMETRO λ DE UNA POBLACIÓN DE POISSON CON MUESTRAS GRANDE

$$\begin{cases} \texttt{- Si } H_0 : \lambda = \lambda_0, & C = \left\{ \dfrac{|\widehat{\lambda} - \lambda_0|}{\sqrt{\lambda_0/n}} > z_{\alpha/2} \right\} \\ \texttt{- Si } H_0 : \lambda \leq \lambda_0, & C = \left\{ \dfrac{\widehat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} > z_{\alpha} \right\} \\ \texttt{- Si } H_0 : \lambda \geq \lambda_0, & C = \left\{ \dfrac{\widehat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} < z_{1-\alpha} = -z_{\alpha} \right\} \end{cases}$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA σ DE UNA POBLACIÓN NORMAL CON μ DESCONOCIDA

$$\begin{cases} \texttt{- Si } H_0: \sigma^2 = \sigma_0^2 & C = \left\{ \dfrac{(n-1)s^2}{\sigma_0^2} \notin \left[\chi^2_{n-1,1-\alpha/2}, \chi^2_{n-1,\alpha/2} \right] \right\} \\ \texttt{- Si } H_0: \sigma^2 \leq \sigma_0^2 & C = \left\{ \dfrac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\alpha} \right\} \\ \texttt{- Si } H_0: \sigma^2 \geq \sigma_0^2 & C = \left\{ \dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha} \right\} \end{cases}$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA σ DE UNA POBLACIÓN NORMAL CON μ CONOCIDA

$$\begin{cases} \texttt{- Si} \;\; H_0 : \sigma^2 = \sigma_0^2 \quad C = \left\{ \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma_0^2} \notin \left[\chi^2_{n,1-\alpha/2}, \chi^2_{n,\alpha/2}\right] \right\} \\ \texttt{- Si} \;\; H_0 : \sigma^2 \leq \sigma_0^2 \quad C = \left\{ \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma_0^2} > \chi^2_{n,\alpha} \right\} \\ \texttt{- Si} \;\; H_0 : \sigma^2 \geq \sigma_0^2 \quad C = \left\{ \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma_0^2} < \chi^2_{n,1-\alpha} \right\} \end{cases}$$

EJEMPLO 7: Con el objeto de estudiar la variabilidad de la cotización de determinadas acciones, se seleccionaron al azar 16 días, y se obtuvo una cuasidesviación típica muestral de 0.7 enteros.
Admitiendo normalidad en las cotizaciones, se quiere contrastar, al nivel 0.05, si la desviación típica es mayor que 0.6 enteros.

$$\begin{cases} H_0 : \sigma \leq 0.6 \\ H_a : \sigma > 0.6 \end{cases} \qquad \frac{s\sqrt{n-1}}{\sigma_0} = \frac{0.7\sqrt{15}}{0.6} = 4.518 < \sqrt{\chi^2_{15,0.05}} = 5$$

Aceptamos H_0.

$$\texttt{p-valor} = p\left(\sqrt{\chi^2_{15}} > 4.518\right) = p\left(\chi^2_{15} > 20.41\right) = 0.1568$$

Confirma la decisión de aceptar H_0

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA DIFERENCIA DE LAS MEDIAS DE DOS POBLACIONES INDEPENDIENTES NO NORMALES CON VARIANZAS CONOCIDAS Y MUESTRAS GRANDES

$$\left\{\begin{array}{ll} \text{- Si } H_0 : \mu_1 = \mu_2, & C = \left\{ \dfrac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} > z_{\alpha/2} \right\} \\ \text{- Si } H_0 : \mu_1 \leq \mu_2, & C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} > z_{\alpha} \right\} \\ \text{- Si } H_0 : \mu_1 \geq \mu_2, & C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < z_{1-\alpha} \right\} \end{array}\right.$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA DIFERENCIA DE LAS MEDIAS DE DOS POBLACIONES INDEPENDIENTES NO NORMALES CON VARIANZAS DESCONOCIDAS Y MUESTRAS GRANDES

$$\left\{\begin{array}{ll} \texttt{- Si } H_0: \mu_1 = \mu_2, & C = \left\{ \dfrac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} > z_{\alpha/2} \right\} \\ \texttt{- Si } H_0: \mu_1 \leq \mu_2, & C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} > z_{\alpha} \right\} \\ \texttt{- Si } H_0: \mu_1 \geq \mu_2, & C = \left\{ \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} < z_{1-\alpha} \right\} \end{array}\right.$$

CONTRASTES AL NIVEL DE SIGNIFICACIÓN α PARA LA DIFERENCIA DE LAS PROPORCIONES DE DOS POBLACIONES BINOMIALES INDEPENDIENTES CON MUESTRAS GRANDES

$$\begin{cases} \text{- Si } H_0: p_1 = p_2, \quad C = \left\{ \dfrac{|\widehat{p}_1 - \widehat{p}_2|}{\sqrt{\bar{p}(1-\bar{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} > z_{\alpha/2} \right\} \\ \text{- Si } H_0: p_1 \leq p_2, \quad C = \left\{ \dfrac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} > z_{\alpha} \right\} \\ \text{- Si } H_0: p_1 \geq p_2, \quad C = \left\{ \dfrac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} < z_{1-\alpha} \right\} \end{cases}$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2},$$

x_1 y x_2 son el número de éxitos observados en cada población

FIN