CONTRASTES DE HIPÓTESIS
Resumen

1) Contrastes de Hipótesis Estadísticas

Un contraste de hipótesis es un procedimiento para saber si los datos de una muestra avalan el valor que suponemos que tiene un parámetro poblacional.

2) Procedimiento

El procedimiento, una vez extraída la muestra, consiste en:

1) Se realiza una afirmación sobre el valor del parámetro poblacional objeto de estudio. A esta afirmación se le llama **Hipótesis nula**. Se la designa por H_0.
La H_0 se formula mediante los signos $=, \leq$ ó $\geq$, pero nunca $\neq$ ni $<$ ni $>$.
Por ejemplo: $H_0: \ \mu = 10$.
Al mismo tiempo, se formula la **Hipótesis alternativa**, que recogerá una afirmación contraria a la nula. Se la designa por H_1. Lo que afirma H_1 puede ser exactamente lo contrario de H_0 o, según las condiciones del problema, otra situación incompatible con H_0.
Por ejemplo: H_1: $\mu \neq 10$.

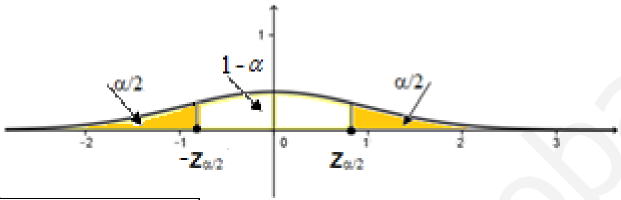
2) Se escoge el **estadístico de contraste**, a partir de un *estimador* del parámetro objeto de estudio cuya distribución es conocida. Hay un *estadístico de contraste* específico para cada tipo de parámetro poblacional objeto de estudio.
Por ejemplo, si estudiamos la media poblacional μ, el *estadístico de contraste* estará basado en $\bar{x}$, cuyo valor es un *estimador* del valor de μ (su valor se aproxima al de μ).
El *estadístico de contraste* es una variable aleatoria, pues su valor depende de la muestra, que es aleatoria. Siempre trabajaremos con *estadísticos* que tendrán una distribución Normal si sabemos que la población que estudiamos también es Normal. En caso contrario, aplicaremos el *Teorema Central del Límite* que, para tamaños muestrales mayores o iguales que 30, permite suponer que la distribución es Normal con resultados aceptables.
La distribución de dicho *estadístico* la tendremos si suponemos cierta la afirmación realizada en H_0.
Por ejemplo, si $H_0 : \ \mu = 10$ y trabajamos con tamaño muestral $n = 50$, si conocemos que $\sigma = 3 \ \Rightarrow \ \bar{x} \in N\left(10; 3/\sqrt{100}\right) = N(10;\ 0.3)$. Nótese que precisamos conocer σ. Si no es así y las muestras son de tamaño mayores o iguales que 30, podría *estimarse* por la *cuasidesviación típica muestral*.
<u>El estadístico de contraste se toma tipificado</u>: su distribución será una $N(0;\ 1)$.
Por ejemplo, el *estadístico de contraste* cuando $H_0 : \mu = \mu_0$ es: $Z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \in N(0;1)$.

3) Se elige el **nivel de significación** α. Éste valor será la probabilidad de que, con los datos de la muestra, se acepte que lo válido es H_1. O sea, que supuesta cierta la H_0, la muestra aporte datos tan discrepantes como para rechazar que sea realmente cierta: $\alpha = P(H_1 / H_0)$
Se llama **nivel de confianza** a $1 - \alpha$ (probabilidad de aceptar H_0, supuesta cierta H_0: el contraste se construye con esta suposición): $1 - \alpha = P(H_0 / H_0)$.

4) A partir α se construye la **región de aceptación**: RA_α, y / o su complementaria, la **región crítica**: RC_α. Como son complementarias, conocida una, tendremos la otra. Si el valor del *estadístico de contraste* está en la *región de aceptación*, daremos como válida la hipótesis nula H_0 con un *nivel de confianza* $1-\alpha$. Y si está en la *región crítica*, aceptaremos H_1 con un *nivel de significación* α.

La región de aceptación depende de que el contraste sea **bilateral** o **unilateral**.

El contraste es **bilateral** si la H_0 se formula con un =. Por ejemplo: $H_0: \ \mu = 10$. Si lo que afirma H_0 es válido, el *estadístico de contraste*, que está tipificado, tomará valores cercanos a su media, que es 0. Sólo si los valores están muy alejados de 0 hay que suponer que H_0 no puede darse por válida, porque las diferencias no se deben al azar. Por tanto:



$\boxed{RA_\alpha = (-z_{\alpha/2}, z_{\alpha/2})}$, siendo $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$

Como:

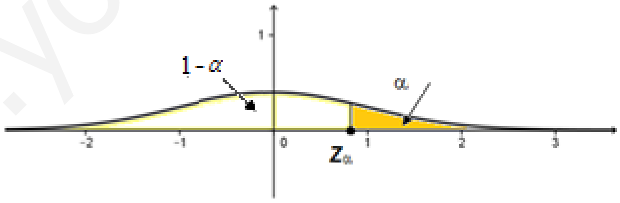
$$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = P(Z \le z_{\alpha/2}) - P(Z \le -z_{\alpha/2}) = \text{(por simetría de la Normal)}$$
$$= P(Z \le z_{\alpha/2}) - P(Z \ge z_{\alpha/2}) = \text{(por suceso contrario)} = P(Z \le z_{\alpha/2}) - [1 - P(Z \le z_{\alpha/2})] =$$
$$= 2\,P(Z \le z_{\alpha/2}) - 1$$

se tiene que:

$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha \iff 2\,P(Z \le z_{\alpha/2}) - 1 = 1-\alpha \iff \boxed{P(Z \le z_{\alpha/2}) = 1-\alpha/2}$

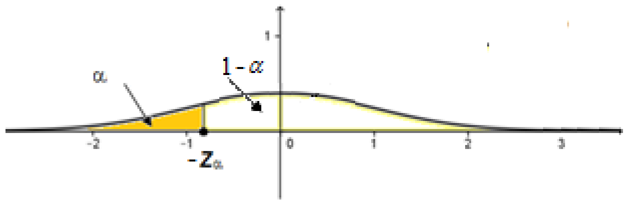
El contraste es **unilateral** si H_0 se formula con un $\le$ ó $\ge$.

- Si contiene un $\le$ (ej: $H_0: \ \mu \le 10$), habrá que dar como válida H_0 si los valores del estadístico de contraste están por debajo de 0 o si rebasan un poco, pero de forma que dicho exceso deba a factores aleatorios. En este caso:



$\boxed{RA_\alpha = (-\infty, z_\alpha)}$, siendo $\boxed{P(Z \le z_\alpha) = 1-\alpha}$

- Si contiene un $\ge$ (ej: $H_0: \ \mu \ge 10$), habrá que dar como válida H_0 si los valores del estadístico de contraste están por encima de 0 o si rebasan un poco por debajo, pero de forma se deba a factores aleatorios. En este caso:



$\boxed{RA_\alpha = (-z_\alpha, +\infty)}$, siendo $P(Z \ge -z_\alpha) = 1-\alpha \iff \boxed{P(Z \le z_\alpha) = 1-\alpha}$

3) Estadísticos de contraste

Contraste para la media poblacional

Como siempre, suponemos que la muestra procede de una población Normal o que $n \geq 30$.

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

siendo:

- μ_0 el valor contenido en la hipótesis nula
- $\bar{x}$ la media muestral
- n el tamaño de la muestral
- σ la desviación típica de la población, que debemos conocer.

Si no se conoce σ pero $n \geq 30$, puede estimarse por la *cuasidesviación típica muestral*.

Contraste para la proporción de una distribución Binomial

Suponemos $n \geq 30$:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$$

siendo:

- p_0 el valor contenido en la hipótesis nula, y $q_0 = 1 - p_0$.
- $\hat{p}$ la proporción muestral
- n el tamaño de la muestral

Contraste para la diferencia de medias poblacionales

Como siempre, suponemos que las dos muestras proceden de poblaciones Normales o que sus tamaños son mayores o iguales a 30.

$H_0: \ \mu_1 - \mu_2 = 0$

$H_1: \ \mu_1 - \mu_2 \neq 0$

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

siendo:

- $\bar{x}_1$ y $\bar{x}_2$ las medias muestrales
- n_1 y n_2 los tamaños de las muestras
- σ_1 y σ_2 las desviaciones típicas de las poblaciones, que debemos conocer.

Si no se conocen las desviaciones típicas poblacionales pero los tamaños muestrales son mayores o iguales que 30, para ambas muestras, pueden estimarse por las *cuasidesviaciones típicas muestrales*.

4) Errores

- **Error de Tipo I**: Se comete cuando el contraste decide rechazar H_0 cuando en realidad es cierta. Es decir: $\alpha = P(H_1 \ / \ H_0)$. Coincide con el *nivel de significación*.
- **Error de Tipo II**: Se comete cuando no rechazamos H_0 cuando en realidad es falsa. Es decir: $\beta = P(H_0 \ / \ H_1)$.

- **Potencia** del contraste es $1 - \beta = P(H_1 / H_1)$.

Podemos disminuir el error de tipo I tanto como queramos, pero esto hace que aumente el error de tipo II. Pero al agrandar el tamaño de la muestra n, disminuye el error de tipo II.
No siempre puede calcularse la potencia del contraste: sólo cuando la hipótesis alternativa no se limita a negar la hipótesis nula, sino que realiza una cuantificación del valor del parámetro que se estudia.

Cuando se intenta disminuir el error de tipo I (disminuyendo el valor de α), aumenta el error de tipo II. La forma de disminuir ambos a la vez es aumentar el tamaño muestral n.