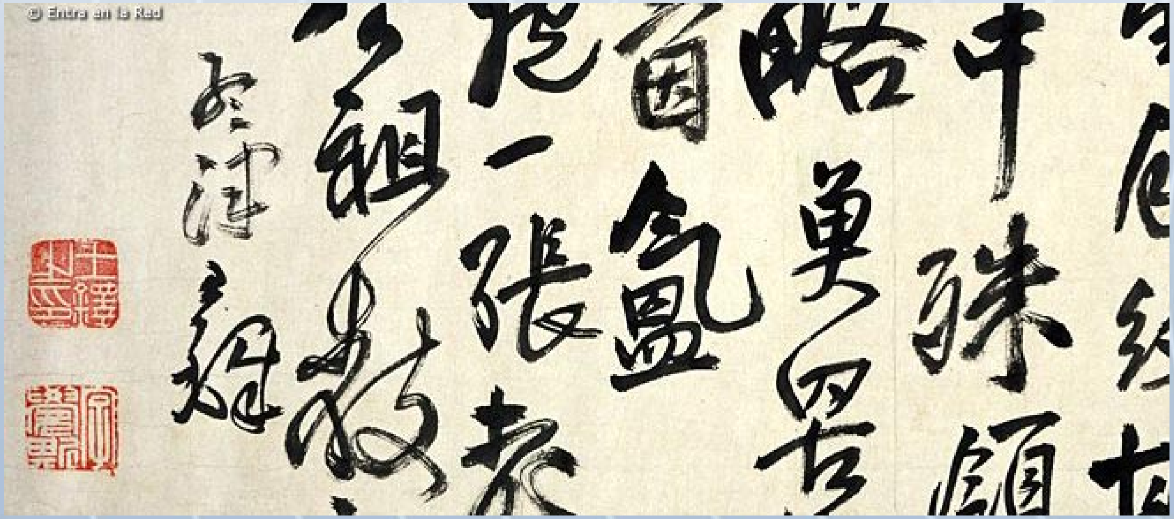
UNIVERSIDAD POLITÉCNICA DE MADRID

conento

Aprendiendo idiomas VI

María Jesús Vázquez Gallo



Formación Conento: febrero 2009

1. Introducción. Contraste de hipótesis paramétricas.
2. Contraste de hipótesis sobre la media.
3. Contraste de hipótesis sobre una proporción.
4. Contraste de hipótesis sobre varias poblaciones.
5. Continuará…

Seguimos aprendiendo sobre el contraste de hipótesis, una técnica estadística que nos permite juzgar, a partir de una muestra de una población, si es razonable aceptar una supuesta propiedad suya.

Ejemplo: Queremos saber si es razonable afirmar, a partir de un muestreo, que la proporción de mujeres casadas de entre 20 y 30 años en Madrid es del 30%, controlando la probabilidad de juzgarlo no razonable aun siendo cierto.

Esto sería un ejemplo de contraste de hipótesis paramétrico.

La hipótesis nula (la que queremos contrastar) se hace sobre el parámetro del modelo: p la proporción de mujeres casadas.

$$\begin{cases} H_0: & p = 0.3 \\ H_1: & p \neq 0.3 \end{cases}$$

En la próxima entrega, hablaremos de contrastes no paramétricos

En ellos se hacen hipótesis relacionadas con el tipo de distribución de probabilidad que ha generado los datos.

Ejemplo: Contrastes de normalidad.

Dos tipos de hipótesis paramétricas:

- Hipótesis simple: sobre un único valor del parámetro.
  Ejemplo: En el caso de las mujeres casadas, H_0: $p = 0.3$.

- Hipótesis compuesta: sobre más de un valor del parámetro.
  Ejemplo: En el mismo caso, H_0: p menor o igual que 0.3.

Pueden realizarse hipótesis sobre:

- Una población. Ejemplo: en un lote de productos, la proporción de defectuosos no supera el 5%. H_0: $p < 0.05$.

- Varias poblaciones. Ejemplo: el salario medio de hombres y mujeres coincide. H_0: $\mu_x - \mu_y = 0$.

Ejemplo: un fabricante de cartuchos para impresora afirma que su duración media es de 1000 páginas.
Se supone que la variable aleatoria x "duración de un cartucho" sigue una distribución normal de media μ.

Vamos a contrastar la hipótesis paramétrica simple: H_0: $\mu = 1000$, a partir de una muestra de 25 cartuchos, cuya media es 1020.

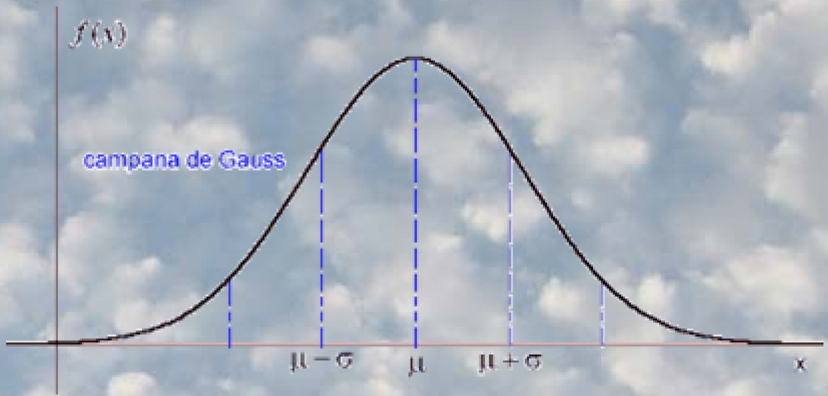
Etapas del contraste

1. **Definición de hipótesis.**
2. **Definición de estadístico de contraste.**
3. **Indicar el valor del nivel de significación.**
4. **Decisión sobre las hipótesis según el valor crítico del contraste.**

2. La elección del estadístico de contraste depende de si la varianza de la distribución normal en cuestión es conocida o no.

¿Recuerdas? La varianza es una medida de dispersión. Es igual la media de los cuadrados de las diferencias entre cada valor de la variable y la media. Su raíz cuadrada es la desviación típica σ, que se interpreta como la desviación en promedio de los datos con respecto a la media.

Aprendiendo idiomas I



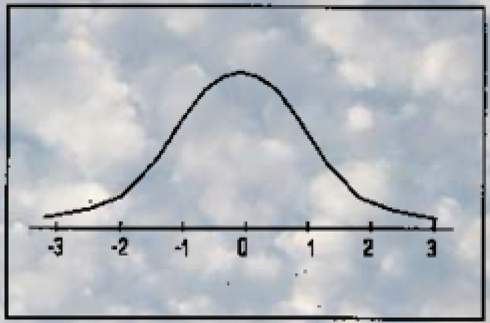
La gráfica de la función de densidad de la distribución normal es simétrica respecto a la media μ, con un máximo en μ, y dos puntos de inflexión, situados a ambos lados de la media y a distancia σ de ella.

Si la varianza σ² de la distribución normal es conocida, se toma como estadístico de contraste la media de los datos de la muestra $\bar{x}$ (puesto que nuestra hipótesis versaba sobre la media) o, mejor, se toma:

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

tamaño de la muestra.

De esta forma, cuando la variable x sigue una normal N(μ, σ²), de media μ y varianza σ², resulta que la variable z sigue una normal N(0,1) como la de la figura.

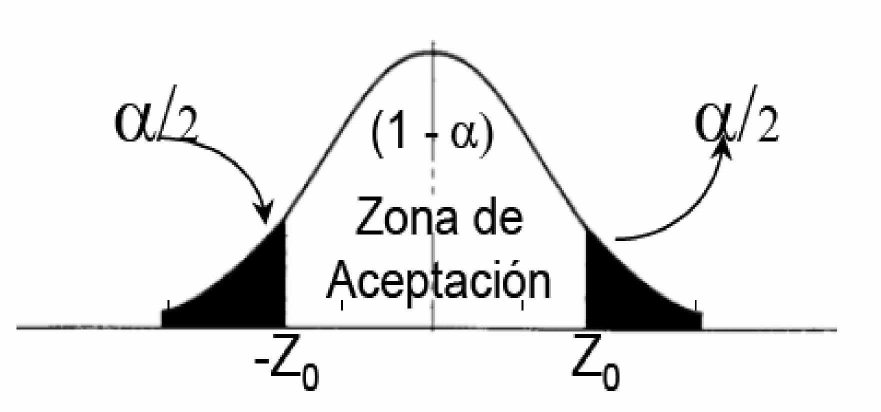


3. Indicar el valor del nivel de significación.

¿Recuerdas lo visto sobre cómo escoger el nivel de significación en Aprendiendo idiomas V ? Tomemos aquí $\alpha = 0.05$.

Se rechazará la hipótesis cuando esta discrepancia se considere demasiado grande:

$$|z| > z_0$$



¿Cómo se obtiene z_0 ?

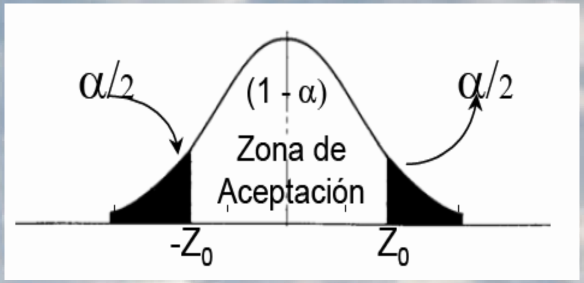
El valor z_0 se calcula una vez elegido el nivel de significación α, es decir, la mayor probabilidad de rechazar la hipótesis nula siendo cierta.

Idea: El estadístico de contraste medirá la distancia entre lo que dice la muestra y la hipótesis nula.

En este caso, debe cumplirse que la probabilidad de que z supere z_0 sea $\alpha/2$ (o de manera equivalente, que la probabilidad de que z sea menor o igual que z_0 sea $\alpha/2$).

Mirando las tablas de la distribución normal (la que sigue z), obtenemos el valor z_0. En nuestro caso: $z_0 = 1.96$. Este valor divide en dos regiones el espacio de posibles valores del estadístico z.

- La región de Rechazo, con probabilidad α, si H_0 es cierta.
- La región de Aceptación, con probabilidad $1-\alpha$, si H_0 es cierta.



4. Decisión sobre las hipótesis según el valor crítico del contraste.

Finalmente, se calcula el llamado valor crítico del contraste, es decir, el valor del estadístico de contraste para la muestra:

$$z = \frac{1020 - 1000}{50 / \sqrt{25}} = 2$$

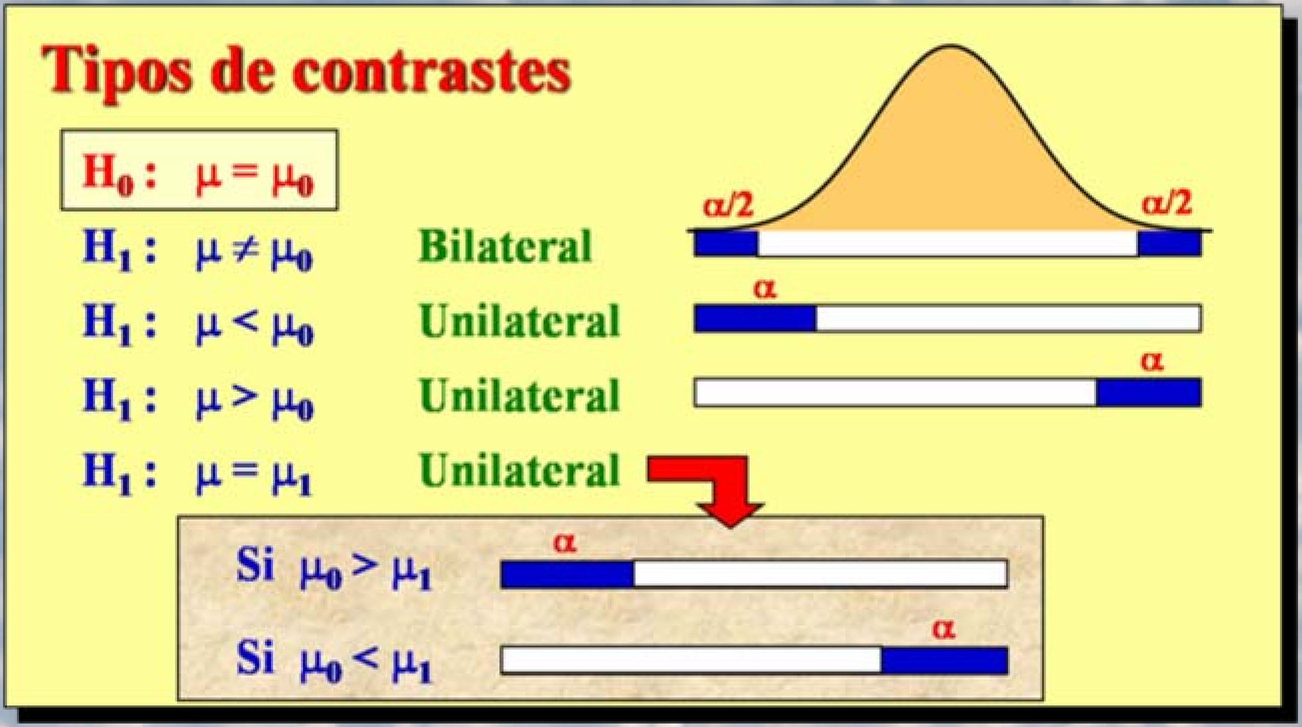
Recordando que 1020 es la media muestral, 25 el tamaño de la muestra y suponiendo que la desviación típica es 50.

Como el valor crítico del contraste pertenece a la región de rechazo (2>1.96) → se rechaza la hipótesis nula H_0 con nivel de significación $\alpha=0.05$

¿Podría ocurrir que la H0 fuese cierta?
Si puede ocurrir, ¿con qué probabilidad sucede?

Tipos de contraste según la hipótesis nula:

Dependen de cómo se establezca la hipótesis nula H_0 (y en consecuencia, la hipótesis alternativa H_1).



Si la varianza es desconocida, se toma como estadístico de contraste:

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

tamaño de la muestra.

¿quién es s?

$$s = \sqrt{\frac{\sum \left(x_i - \bar{x}\right)^2}{n - 1}}$$

Este estimador se obtiene a partir de la muestra y sustituye a la varianza.

Si la hipótesis versa sobre una proporción. Por ejemplo:

$H_0: \quad p = p_0$

y dicha proporción en la muestra toma el valor p*, se toma como estadístico de contraste:

$$z = \frac{p^* - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

que sigue una distribución normal N(0,1)

Si se consideran, por ejemplo, dos poblaciones x e y, y la hipótesis nula versa sobre sus medias:

H_0: $\mu_x - \mu_y = \delta_0$.

se toma como estadístico de contraste:

$$z = \frac{\bar{x} - \bar{y} - \delta_0}{\sqrt{\frac{\sigma^2{}_x}{n} + \frac{\sigma^2{}_y}{m}}}$$

donde n y m son los tamaños de las muestras, que sigue una distribución normal N(0,1)


Continuará...