

LOS TEST DE HIPÓTESIS PARA ALUMNOS DE BACHILLERATO

**Díaz Leno, M.S.
Martín Rodríguez, J.
Tardáguila García, P.**

1. INTRODUCCIÓN:

La Inferencia Estadística consta de dos partes: Estimación y Contrastes de Hipótesis. La técnica de Contraste de Hipótesis es preciso para establecer procedimientos para aceptar o rechazar hipótesis estadísticas emitidas acerca de un parámetro, u otra característica de la población.

La única forma de saber con certeza absoluta que una hipótesis estadística es verdadera, es examinar toda la población. Pero esto, en la mayoría de los casos, resulta imposible (por falta de medios económicos, imposibilidades técnicas, etc.). Por lo tanto, la decisión debe adoptarse a partir de los resultados de una muestra de la población (supuesta representativa), que nos inducirá a tomar la decisión sobre la verdad o falsedad de la hipótesis. Pero es difícil ésta decisión, porque aunque sepamos exactamente el valor del parámetro de la población, en las muestras es muy difícil que se verifique ese valor exacto, por lo que debemos decidir unos límites de valores del parámetro en la muestra, que nos puedan llevar a la decisión de aceptar el valor del parámetro poblacional.

2. CONCEPTOS BÁSICOS

A la hipótesis que se desea contrastar la denominaremos **Hipótesis nula** (H_0).

*Esta **hipótesis nula** es la que se somete a comprobación, y es la que se acepta o rechaza, como la conclusión final de un contraste.*

Esta hipótesis nula lleva consigo una **hipótesis alternativa**, denotada por H_a o H_1 .

*La **hipótesis alternativa** será la que se acepta si se rechaza H_0 y viceversa*

El contraste de hipótesis, es pues, un mecanismo mediante el cual se rechaza la hipótesis nula cuando existan diferencias significativas entre los valores muestrales y los valores teóricos, y se acepta en caso contrario. Estas variables se medirán mediante una variable denominada **estadígrafo de contraste**, o **estadístico de contraste**, que sigue una distribución determinada conocida, y que para cada muestra tomará un valor particular.

*La **región crítica** es el conjunto de valores del estadístico de contraste que nos induce a rechazar la hipótesis nula*

*La **región de aceptación** es el conjunto de los valores del estadístico que nos induce a aceptar la hipótesis nula.*

Obviamente la conclusión tras un contraste de hipótesis puede ser cierta o no, ya que no sabemos con certeza cuál es la situación verdadera. Esto nos puede llevar a las situaciones reflejadas en el siguiente cuadro:

		Decisión	
		Aceptar H_0	Rechazar H_0
Hipótesis cierta	H_0	Correcta	Error tipo I
	H_1	Error tipo II	Correcta

Nivel de significación del contraste es la probabilidad de cometer un error del tipo I, es decir, de rechazar la hipótesis nula siendo cierta, y se acostumbra a denotar por α

La probabilidad de cometer error del tipo II se denota por β

$$\beta = P(\text{cometer error tipo II}) = P(\text{aceptar } H_0 \text{ siendo falsa})$$

La **potencia del contraste**, es la probabilidad de rechazar la hipótesis nula siendo ésta falsa

Según que la región crítica contenga una o dos regiones, diremos que el contraste es **unilateral**, o **bilateral**.

3. CONTRASTE PARA LA MEDIA DE UNA POBLACIÓN NORMAL

* **Contraste para la media de una población normal, con varianza poblacional conocida**

Supongamos que queremos contrastar la hipótesis de que la media μ de una población normal, toma un valor específico μ_0 , cuando la varianza s^2 de la población es conocida.

– En éste caso, la **hipótesis nula** será, en general: $H_0 (\mu = \mu_0)$

– Mientras que la **hipótesis alternativa** puede tener diversas expresiones:

$H_a (\mu = \mu_1)$, o bien, $H_a (\mu < \mu_0)$, $H_a (\mu > \mu_0)$, o bien $H_a (\mu \in \mu_0)$.

– **Estadígrafo de contraste:** El contraste se efectuará tomando muestras aleatorias de tamaño n . Conocemos que la distribución de la media muestral sigue una distribución normal $N(\mu, s/\sqrt{n})$. Entonces, si \bar{x} es la media de una muestra de tamaño n , la variable

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

seguirá, una distribución normal estándar $N(0,1)$.

– **Nivel de significación:** El nivel de significación será α , que generalmente tomará los valores 0.1, 0.05 ó 0.01.

– **Región crítica:** Si la hipótesis nula fuese cierta, $\mu = \mu_0$, cabe esperar que la media muestral se distribuya en torno al valor μ_0 , es decir, $\bar{x} - \mu_0$ tendrá un valor elevado para que existan evidencias de que la hipótesis

nula sea falsa, es decir, la variable Z tomará un valor absoluto grande; así pues, la región crítica estará formada por los valores de Z elevados, tanto positivos como negativos. Para especificar cuando se consideran elevados, teniendo en cuenta la distribución de Z , serán aquellos que sean mayores, en valor absoluto, que $z_{\alpha/2}$, en el contraste bilateral, o que z_α en el contraste unilateral. (Ver figura 1).

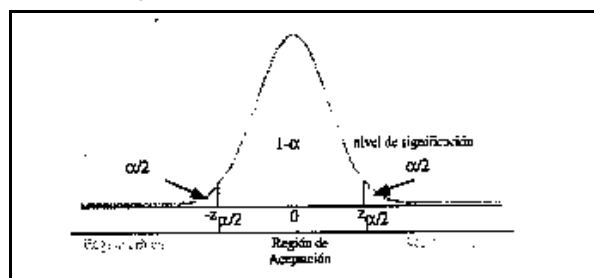


Figura 1: Regiones crítica y de aceptación en un contraste bilateral cuando la distribución es normal estándar.

* **Contraste para la media de una población normal, con varianza poblacional desconocida**

MUESTRAS GRANDES

– **Estadígrafo de contraste:**

En el mismo caso que en el párrafo anterior, y con las mismas hipótesis, si el tamaño de la muestra es suficientemente grande ($n > 30$), aunque sea desconocida la varianza poblacional, se consiguen buenos resultados utilizando como estimador de la varianza poblacional la cuasi-varianza muestral y, por lo tanto, se puede tomar como estadístico de contraste el mismo que se tomó cuando la varianza poblacional era conocida; es decir

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

es una variable que sigue una distribución normal estándar $N(0,1)$, por lo que el razonamiento es idéntico al caso anterior.

MUESTRAS PEQUEÑAS

– **Estadígrafo de contraste:**

Si el tamaño de la muestra es pequeño ($n < 30$), sabemos que la variable

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n-1}}$$

seguirá, una t con $n-1$ grados de libertad. Se trabaja con los mismos valores de α que en el caso anterior.

$$R.C. = \{t/t < -t_{\alpha, n-1} \text{ o } t > t_{\alpha, n-1}\}$$

4. CONTRASTE DE HIPÓTESIS PARA LA IGUALDAD DE MEDIAS DE DOS POBLACIONES NORMALES:

En este apartado consideraremos dos poblaciones con distribuciones normales con medias μ_1 y μ_2 y varianzas s_1^2 y s_2^2 respectivamente, de las cuales extraemos muestras aleatorias independientes de tamaños n_1 y n_2 respectivamente. El objetivo de éste apartado será determinar si las dos poblaciones pueden considerarse con la

misma media poblacional, es decir, la hipótesis nula será H_0 ($\mu_1 = \mu_2$), mientras que la hipótesis alternativa puede tener diversas expresiones: H_a ($\mu_1 < \mu_2$), o bien, H_a ($\mu_1 > \mu_2$), o bien H_a ($\mu_1 \neq \mu_2$).

De la misma manera que en los casos anteriores, el procedimiento consiste en elegir el estadígrafo más adecuado, y, teniendo en cuenta el nivel de significación, determinar la región crítica.

La figura 2 muestra, de forma esquemática, el proceso para decidir el tipo de contraste más adecuado para poblaciones normales.

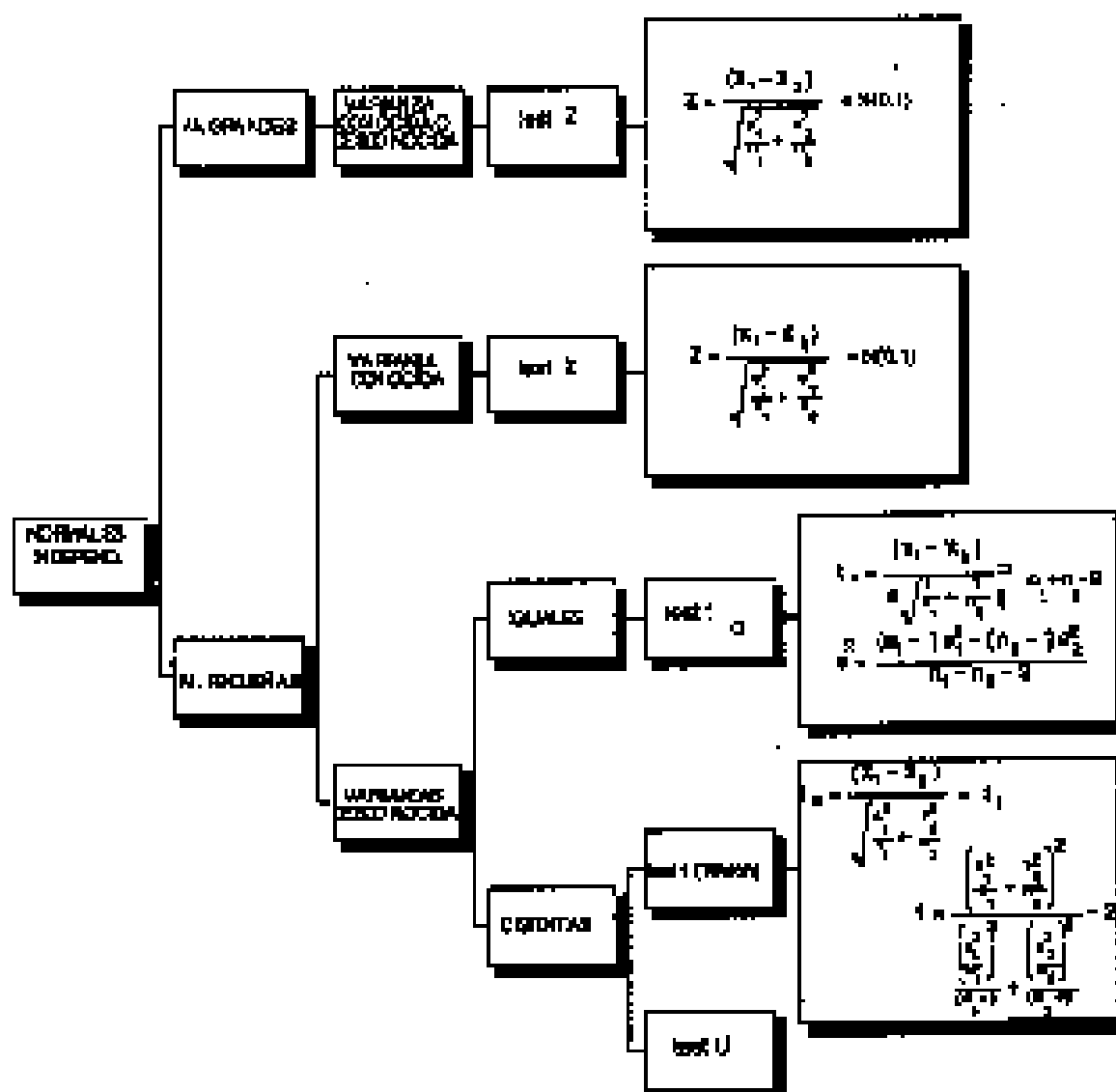


Figura 2: Árbol de decisiones para el contraste de comparación de las medias de dos poblaciones normales.

5. CONTRASTE PARA DISTRIBUCIONES BINOMIALES:

Estudiaremos sólo contrastes en los que sea posible aproximaciones de la binomial mediante la normal, por lo que estudiaremos sólo los casos de muestras grandes, de tamaño > 30 .

* Contraste para el parámetro p de una distribución Binomial

Partimos de una población que se ajuste al modelo binomial $B(n, p)$, siendo p la probabilidad de «éxito»; denotaremos por \hat{p} a la proporción muestral de casos favorables y por p_0 el valor hipotético con el que queremos contrastar el valor del parámetro p .

– Hipótesis de partida

En éste caso:

$H_0 : p = p_0$ $H_a : p \in p_0$ en el contraste bilateral,
o bien $H_a : p > p_0$, en el contraste unilateral (también $H_a : p < p_0$).

– Estadígrafo de contraste:

Como conocemos que la distribución binomial $B(n, p)$ se aproxima mediante una variable normal $N(np, \sqrt{np(1-p)})$, entonces, se verifica que la variable

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

se distribuye como una distribución normal estándar $N(0,1)$.

– Región crítica:

$$R.C. = \{Z/Z < -Z_{\alpha/2} \text{ y } Z < Z_{\alpha/2}\}$$

* Contraste para la igualdad de los parámetros de dos distribuciones binomiales

Partimos, en éste caso, de dos distribuciones binomiales $B(n_1, p_1)$ y $B(n_2, p_2)$ respectivamente. En las muestras, los parámetros muestrales serán \hat{p}_1 y \hat{p}_2 respectivamente.

– Hipótesis de partida:

$$H_0 : p_1 = p_2 \quad H_a : p_1 \neq p_2$$

– Estadígrafo de contraste:

Ahora, teniendo en cuenta las propiedades de las distribuciones normales, por las que se aproximan las binomiales, se verifica que la variable

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

se distribuye, como una distribución normal estándar $N(0,1)$

– Región crítica:

La región crítica será análoga a todas aquéllas en el que el estadístico de contraste sigue una distribución nor-

BIBLIOGRAFÍA

- FREEDMAN, PISANI, PURVES and ADHIKARI (1993): «*Estadística*». Antoni Bosch Editor.
 GALINDO, M. P. (1984): «*Exposición intuitiva de métodos estadísticos. (Fundamentos y aplicaciones a Biología, Medicina y otras ciencias)*». Universidad de Salamanca.
 G. M. CLARKE and D. COOKE (1992): «*A basic course in Statistics*». Arnold.
 MARTÍN ANDRÉS, A.; LUNA DEL CASTILLO, J. DE D. (1990): «*Bioestadística para las Ciencias de la Salud*». Ediciones Norma.