	**Temas de Estadística Práctica** **Antonio Roldán Martínez** **Proyecto** http://www.hojamat.es/
	Tema 8: Tests de hipótesis **Resumen teórico**

Tests de hipótesis

Concepto de test de hipótesis

Un test de hipótesis (o contraste) es un proceso, compuesto de varios pasos muy concretos, que nos permite aceptar o rechazar una hipótesis en términos estadísticos. Desarrollamos esto con más extensión:

Proceso de contraste de hipótesis

1) Planteamiento de las hipótesis

Un contraste se comienza con una afirmación. En este curso consistirá en afirmar un valor concreto de un parámetro o de la diferencia o cociente de parámetros. Llamaremos ***Hipótesis nula H_0***. a la afirmación que hacemos sobre los parámetros de una población y cuya validez deseamos contrastar: La estatura media en Andalucía es de 1,74, este grupo tiene dos puntos más de media que este otro, la varianza de esta población siempre es 54, en todos los experimentos que se efectúen, etc.

Hipótesis nula H_0 es la afirmación cuya validez se desea contrastar.

Frente a esa afirmación podemos oponer otra, a la que llamamos ***hipótesis alternativa H_1***. Suele ser una desigualdad que se opone a la igualdad que afirmamos. Así, si llamamos **m** a la media de una población, podíamos plantear como hipótesis nula

H_0 : m=234

y su hipótesis alternativa podría ser **H_1 : m<>234**, es decir, que la media puede ser mayor o menor que 234. En este caso hablaremos de un *contraste bilateral* o *de dos colas*.

Si tomamos como hipótesis alternativa **H_1 : m>234** o bien **H_1 : m<234** expresaremos que estamos usando un contraste *unilateral* o *de una cola*.

Decidir un contraste u otro depende del contexto. Si experimentamos con un analgésico no nos planteamos que aumente el dolor, ni tampoco que una cantidad de agua disminuya la humedad.

2) Supuestos

Debemos tener en cuenta siempre qué supuestos estamos aceptando sobre la población, si es es simétrica, normal, continua… y sobre la muestra, si es aleatoria simple, es de tamaño mayor que 30…

En este curso, por su especial orientación, no daremos mucha importancia a los supuestos, aunque los nombraremos cuando sea oportuno.

3) Estadístico de contraste

Es la expresión matemática, calculada a partir de la muestra, que nos servirá para tomar la decisión. Debemos conocer la función de distribución que posee. Las tres más usadas son: la Normal, la T de Student y la Chi-cuadrado.

4) Nivel de significación

En todo contraste hay que poner una barrera entre aceptar la hipótesis nula **H_0** o rechazarla, y en ese caso aceptar la alternativa **H_1** . Construiremos dos zonas, la primera de las cuales, ***De rechazo***, compuesta por valores, que por estar alejados de la hipótesis nula, se suponen que no han aparecido al azar, y que por tanto **H_0** es la que ha cambiado. La probabilidad de que unos valores caigan en la región de rechazo, **a pesar de que H_0 sea verdadera**, se conoce con el nombre de ***nivel de significación a***, y se suele fijar con un valor pequeño: 5%, 1%, … Así los experimentos son conservadores, pues asignan muy poca probabilidad a la hipótesis alternativa **H_1**.

La zona complementaria, ***zona de aceptación***, tendrá como probabilidad **1 - a**, generalmente muy grande, del orden del 95%, 96%, o 99%.

5) Toma de decisión

Una vez realizado el experimento en una muestra y calculado el *estadístico de contraste*, según dónde caiga, ***aceptaremos o rechazaremos*** la hipótesis nula. A veces, cuando se rechaza, mediante las técnicas de estimación se propone un valor alternativo.

También se suele añadir a la decisión, no sólo el nivel de significación a, sino también la probabilidad de que en un experimento realizado en las mismas condiciones y con la hipótesis nula considerada verdadera, produzca valores del estadístico **menores** que el obtenido en nuestro caso. A este valor lo llamaremos ***p-valor***. Si es mayor que 1-a en contraste unilaterales, o bien mayor que 1-a/2 en bilaterales, rechazaremos la hipótesis nula.

A continuación resumimos en dos tablas los contrastes más importantes existentes sobre una muestra y sobre dos. Para más detalles o ejemplos es preferible acudir a un texto de Estadística Inferencial.

Significado de los símbolos: **m** media de la población , **s** desviación típica de la población, **p** proporción en la muestra, **P** proporción en la población , **S** desviación típica en la muestra, **D** diferencia entre medias, $\mathbf{m_D}$ media de la diferencia en la población.

Contrastes sobre una muestra

Estimador	Supuestos	Distribución	Estadístico de contraste
$\overline{X}$	s^2 conocida Población normal y/o n>30	N(0,1)	$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$

$\overline{X}$	s^2 desconocida Población normal, al menos aprox.	T_{n-1}	$T = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n-1}}$
p	Población binomial np>5	N(0,1)	$Z = \dfrac{p - P}{\sqrt{PQ / n}} \approx \dfrac{p - P}{\sqrt{1/(4n)}}$
S^2	Población normal	χ^2_{n-1}	$T = \dfrac{nS^2}{\sigma_0^2}$

Contrastes sobre dos muestras

Estimador	Supuestos	Distribución	Estadístico de contraste
$\overline{X}_2 - \overline{X}_1$	s^2 conocida e igual en las dos muestras Independencia Población normal y/o n>30	N(0,1)	$\dfrac{X_2 - X_1 - (\mu_2 - \mu_1)}{\sqrt{\left(\dfrac{\sigma_2^2}{n_2} + \dfrac{\sigma_1^2}{n_1}\right)}}$
$\overline{X}_2 - \overline{X}_1$	s^2 desconocida e igual en las dos muestras Independencia Población normal	$T_{n1+n2-2}$	$\dfrac{X_2 - X_1 - (\mu_2 - \mu_1)}{\sqrt{\left(\dfrac{1}{n_2} + \dfrac{1}{n_1}\right)\left(\dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\right)}}$

$\overline{X}_2 - \overline{X}_1$	s^2 desconocidas y distintas en las dos muestras Independencia Población normal	N(0,1)	$$\frac{X_2 - X_1 - (\mu_2 - \mu_1)}{\sqrt{\left(\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}\right)}}$$ Se supone que las dos desviaciones típicas de la población se sustituyen por sus estimadores insesgados, las cuasidesviaciones típicas
$\overline{X}_2 - \overline{X}_1$	s^2 desconocida e igual en las dos muestras Datos apareados Población normal	T_{n-1}	$$\frac{\overline{D} - \mu_d}{\sqrt{\left(\frac{\sigma_d^2}{n}\right)}}$$
$p_2 = p_1$	Población binomial Igualdad de proporciones en muestras grandes e independientes Se suponen proporciones iguales en la población	N(0,1)	$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ siendo $$p = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2}$$
$p_2 - p_1 = d$	Población binomial Diferencia de proporciones en muestras grandes e independientes, sin el supuesto de igualda de proporciones en la población	N(0,1)	$$Z = \frac{p_2 - p_1 - (P_2 - P_1)}{\sqrt{P_1Q_1 / n_1 + P_2Q_2 / n_2}}$$

S_2^2 / S_1^2	Poblaciones normales	Fn1-1,n2-1	$F = \frac{S_2^2}{S_1^2}$