# CONTRASTE DE HIPÓTESIS. UNA PROPUESTA PARA EL LABORATORIO.

**César Gutiérrez**[1]

[1]*Escuela Universitaria Politécnica. Universidad de Valladolid.*

**Resumen:** El contraste de hipótesis es un elemento de importancia capital en cualquier procedimiento de inferencia estadística. La comprensión de esta técnica es, por tanto, uno de los objetivos fundamentales que el alumno de Informática debe alcanzar al cursar la asignatura de Estadística. Esta ponencia propone algunos problemas que pueden ayudar en dicho cometido.

## 1.- INTRODUCCIÓN.

El contraste de hipótesis es un concepto estadístico difícil de enseñar. La experiencia muestra como el alumno encuentra grandes dificultades para conjugar todos los elementos que en él intervienen y comprender, a partir de ellos, el funcionamiento y las ideas que hay detrás de esta técnica estadística de manera satisfactoria.

Por ello es importante aprovechar cualquier oportunidad que se nos presente para incidir sobre este concepto. El laboratorio es una de ellas. La resolución de un problema con ordenador libera al alumno de la realización de operaciones de forma que su atención se concentra fundamentalmente en el planteamiento del mismo y en la elección de un procedimiento adecuado para resolverlo.

Podemos conseguir, mediante determinados problemas y cuestiones, que el alumno razone sobre los contrastes estadísticos, de forma que no se limite únicamente a resolverlos, como si de un algoritmo se tratara, sino que comprenda su funcionamiento y por qué se hacen así las cosas y no de otra manera. El profesor juega un papel esencial en este aprendizaje, guiando a los alumnos en sus razonamientos, planteando alternativas que clarifiquen los conceptos, fomentando el debate y el interés sobre las cuestiones que se están tratando, y en definitiva, actuando en todos aquellos aspectos que favorezcan la compresión completa de este concepto estadístico.

La finalidad de esta comunicación es mostrar algunos ejemplos que pueden ser útiles en esta línea. Se trata de enunciados que han sido propuestos en las prácticas correspondientes a la asignaturas Estadística y Estadística II de los estudios de Ingeniería Informática en la Universidad de Valladolid.

## 2.- PLANTEAMIENTO DE HIPÓTESIS.

Un ejercicio interesante que obliga al alumno a razonar sobre la elección de hipótesis en un contraste es cualquier enunciado concreto que se ajuste a la siguiente situación general:

Sea X una variable aleatoria con distribución $N(\mu,\sigma)$ que representa una característica de importancia en cierto contexto. Esa importancia se traduce en que un valor medio $\mu$ inferior a $\mu_0$ tiene consecuencias fatales en dicho contexto.

1. Resolver los siguientes contrastes a nivel $\alpha$:

$H_0$: $\mu \leq \mu_0$ contra $H_1$: $\mu > \mu_0$ y

$H_0$: $\mu \geq \mu_0$ contra $H_1$: $\mu < \mu_0$.

(la muestra aleatoria obliga a aceptar la hipótesis nula en ambos planteamientos).

1□¿Cuál es la decisión correcta?

En este tipo de problemas, el alumno debe plantearse las consecuencias que conlleva una decisión equivocada y razonar sobre el diseño del test y las probabilidades de error $\alpha$ y $\beta$ que de dicho diseño se derivan.

## 3.- REGIÓN CRÍTICA, PROBABILIDADES DE ERROR Y POTENCIA.

El siguiente enunciado tiene como finalidad que el alumno comprenda el criterio de la máxima potencia en el diseño de un test:

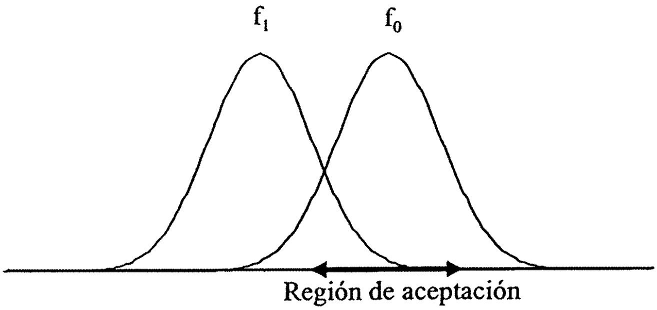
Sea X una variable aleatoria continua con valores sobre el intervalo [0,1]. Plantee el mejor test posible para contrastar a nivel $\alpha$:
$H_0$: $X \sim U(0,1)$ frente $H_1$: $X \sim f_1(x)$

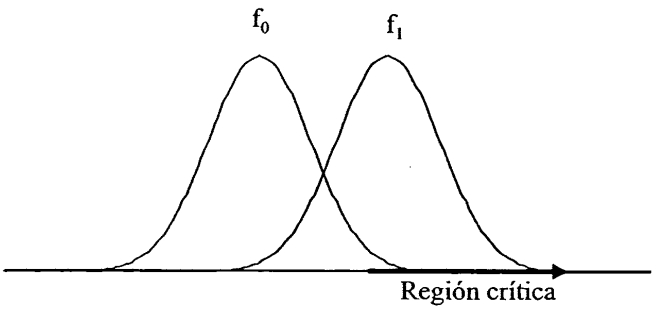
$f_1(x)$ podría ser, por ejemplo, una función de densidad creciente (o decreciente) que tenga por rango el intervalo [0,1] y cuya gráfica sea sencilla de dibujar. El alumno debe comprender que cualquier subintervalo de longitud $\alpha$ define una región crítica de ese nivel y razonar sobre la gráfica para elegir el test más potente.

El problema se puede completar solicitando la potencia o la probabilidad $\beta$ de cometer un error tipo II del test obtenido. De esta manera conseguiremos que el alumno piense sobre el significado de estas probabilidades, calculándolas explícitamente.

A continuación muestro una cuestión que puede ayudar a comprender los valores $\alpha$ y $\beta$ como lo que son, probabilidades:

Dibuja en cada caso las áreas correspondientes a los valores $\alpha$ y $\beta$ del contraste $H_0$: $f_0$ frente a $H_1$: $f_1$.




Al resolver esta cuestión, el alumno analiza el significado de las probabilidades de error y la forma de obtenerlas, tomando la densidad y región adecuada según corresponda.

## 4.- RELACIÓN ENTRE INTERVALOS DE CONFIANZA Y TEST DE HIPÓTESIS.

Consideremos cualquier situación concreta que se ajuste al siguiente planteamiento general:

Tenemos una variable aleatoria X con distribución $N(\mu,\sigma)$ y se ha realizado el contraste $H_0$: $\mu = \mu_0$ contra $H_1$: $\mu \neq \mu_0$ a nivel $\alpha$.

Podemos plantear la siguiente cuestión: ¿qué valores $\mu_0$ hacen significativo el contraste?. El alumno debe pensar intuitivamente que dichos valores no son "admisibles" para la media y relacionar ese pensamiento con la idea de intervalo de confianza para resolver la cuestión planteada.

Siguiendo la misma línea se puede proponer la situación recíproca: ¿sabiendo que el resultado del contraste ha sido aceptar $H_0$, se puede concluir que $\mu_0$ pertenece al intervalo de confianza para $\mu$ de nivel $-\alpha$?

## 5.- P-VALOR Y NIVEL DE SIGNIFICACIÓN.

El cálculo del p-valor en un problema donde la distribución subyacente es normal no supone ninguna complicación. Esto hace que, frecuentemente, el alumno pierda de vista el significado de este concepto. La siguiente cuestión puede ser útil en este sentido:

Escribir todos los niveles de significación $\alpha \in (0,1)$ para los que el test es no significativo.

Esta cuestión se puede formular en cualquier problema en el que se resuelva un test y obliga, necesariamente, a plantearse la idea de p-valor.

Otro enunciado interesante en este sentido es:

Supongamos una situación concreta basada en una variable aleatoria X con distribución $N(\mu,\sigma)$ y consideremos que se ha resulto un mismo test sobre $\mu$ dos veces a partir de sendas muestras aleatorias del mismo tamaño y en ambos casos la decisión ha sido aceptar $H_0$. En estas condiciones, ¿qué muestra aporta mayor evidencia sobre la veracidad de la hipótesis nula?

Considerar el p-valor como una medida del grado de evidencia mostrado por los datos sobre la hipótesis nula dará la pista para resolver el problema. Así conseguiremos que el alumno no piense en el p-valor, simplemente, como un procedimiento alternativo para tomar una decisión y que no pierda de vista que la decisión de un test es una afirmación basada en la evidencia empírica.

## 6.- SIGNIFICATIVIDAD ESTADÍSTICA Y SIGNIFICATIVIDAD PRÁCTICA.

La última consideración que realizo tiene que ver con la aplicación práctica del resultado de un test. Concretamente, es un inciso sobre la interpretación de la decisión adoptada en una prueba de hipótesis cuando la muestra aleatoria de datos sobre la que se ejecuta el procedimiento es grande. En esta situación, el alumno debe ser consciente de que la potencia del test puede ser muy elevada incluso en valores del parámetro próximos a la hipótesis nula. De esta manera, el test puede detectar cualquier alejamiento pequeño con respecto al valor o valores de $H_0$, aun cuando la diferencia no tiene ninguna importancia desde un punto de vista práctico.

Un ejercicio interesante que obliga al alumno a razonar sobre esta cuestión es cualquier enunciado concreto que se ajuste a la siguiente situación general:

Sea X una variable aleatoria con distribución $N(\mu,\sigma)$. X representa una característica sobre la que se quiere contrastar $H_0$: $\mu \leq \mu_0$ contra $H_1$: $\mu > \mu_0$ a nivel $\alpha$. Se sabe que un alejamiento del valor nulo $\mu_0$ en una unidad no es de importancia. A partir de la muestra aleatoria $x_1,\ldots,x_n$, ¿qué decisión adoptarías?. Obtén un intervalo de confianza de nivel $-\alpha$ y, a partir de él, razona si la decisión anterior es la adecuada. ¿Qué influencia tiene el tamaño muestral en el resultado?

Escogiendo adecuadamente la muestra (simulación) podemos conseguir que el resultado del test sea rechazar la hipótesis nula y que el intervalo de confianza contenga valores inferiores a $\mu_0 + 1$ (necesitaremos para ello un tamaño muestral grande).

Con este problema, el alumno no olvidará el sentido práctico de los contrastes de hipótesis, además de relacionar la potencia de un test con el tamaño muestral.

## 7.- CONCLUSIONES.

Las clases de laboratorio permiten que el alumno, no solo plantee y resuelva problemas prácticos aplicando uno u otro procedimiento estadístico, sino que proporcionan una excelente oportunidad para ahondar en las ideas que hay detrás de estas técnicas. La interacción profesor-alumno y una adecuada selección de preguntas y problemas puede conseguir que ciertos conceptos estadísticos complejos, como es el contraste de hipótesis, se comprendan de forma satisfactoria.