

Semana 3

Prueba de hipótesis *Valentín Vergara Hidd*

Usualmente utilizamos herramientas estadísticas para describir una situación (crear modelos); o para tomar decisiones. En ambos casos, lo que buscamos es que el resultado sea **preciso**. Es decir, aceptando que al trabajar con muestras no lograremos una precisión del 100 %, lograr resultados los suficientemente precisos como para *confiar* en ellos.

En este documento les presentaré la idea de **prueba de hipótesis**, que constituye el “esqueleto” de la creación de modelos o de la toma de decisiones. Hay ciertos procedimientos que cambian dependiendo de la prueba estadística, pero la lógica se mantiene igual. Es importante tener esto muy claro, porque de ahora en adelante, es la principal herramienta que vamos a utilizar.

Precisando conceptos

Es probable que ustedes tengan algún grado de familiaridad con el concepto de hipótesis, dado que lo utilizamos constantemente en ciencia, o incluso en el lenguaje cotidiano. Precisamente esto es un problema, ya que nos lleva a imprecisiones.

Por tanto, lo primero que definiremos es el uso particular de hipótesis que trabajaremos en este documento (y en los siguientes)

Definición 1: Hipótesis

En estadística, definimos **hipótesis** como una proposición lógica sobre una propiedad de la población. Como tal, podemos comprobar el valor de verdad de dicha proposición. Es decir, es verdadera o es falsa, de forma excluyente.

Un corolario de la definición anterior (Triola, 2018), es que utilizaremos el procedimiento de la **prueba de hipótesis** para probar la veracidad de la proposición contenida en ella. Quizás esto sea más evidente con un ejemplo.

Ejemplo 1: Puntajes PSU

En un video anterior, les comenté que la PSU está diseñada para ser una prueba cuya distribución de puntajes tiene $\mu = 550$ y $\sigma = 110$.

Sólo sabiendo eso, supongamos que contamos con los puntajes de toda una generación que egresa de un colegio, con 63 estudiantes. Para esta generación, si X es la variable definida como el puntaje PSU, medimos $\bar{x} = 588,7$ y $s_x = 83,9$.

Una hipótesis (proposición) válida sería: *la muestra no es representativa de la población*. Los valores no son exactamente los esperados, de acuerdo a los parámetros poblacionales.

Sin embargo, sabemos que es posible cierta *desviación* respecto de ellos. Para saber si la proposición escrita en cursiva es verdadera o falsa, necesitamos hacer una prueba de hipótesis.

Pasos de la prueba

El procedimiento de la prueba de hipótesis se puede resumir en cuatro pasos. Usualmente, utilizamos un computador para hacer cálculos, que en su mayoría se encuentran únicamente en el tercer paso. Todo lo demás involucra más lógica que aritmética, por lo que, si bien es muy útil utilizar un computador para hacer cálculos que de otra forma serían extremadamente tediosos, lo fundamental es tener clara la lógica subyacente a esta prueba de hipótesis.

A continuación describiré con detalle los cuatro pasos, al mismo tiempo que mantendremos la descripción lo más genérica posible. Esto porque mi intención en este documento es presentar la lógica, dejando los detalles particulares para cada *familia* de pruebas.

Paso 1: Establecer hipótesis nula e hipótesis alternativa

En este paso se busca crear definiciones operacionales que nos permitirán aceptar o rechazar proposiciones que no necesariamente son exactamente iguales a la pregunta de investigación. Estas proposiciones corresponden a la hipótesis nula e hipótesis alternativa.

No necesariamente corresponden a la pregunta de investigación, pero sí tienen relación con ella. Esto se debe a que generalmente formulamos la hipótesis nula en términos de una igualdad y la o las hipótesis alternativas en términos de una desigualdad. La justificación de esta forma de proceder es que en términos lógicos, podemos aceptar o rechazar una igualdad más fácilmente que una desigualdad.

Definición 2: Hipótesis nula y alternativa

Una **hipótesis nula** H_0 es una proposición lógica en la que se define el valor de un parámetro poblacional como *igual* a un valor propuesto por el investigador.

La **hipótesis alternativa** H_1, H_2, \dots es una proposición que se formula de tal manera en que el parámetro poblacional *difiere* del valor propuesto. Al resolver un problema, puede haber más de una hipótesis nula.

Para transformar la pregunta de investigación en hipótesis, se sugieren los siguientes pasos.

1. Identificar el elemento de la pregunta de investigación que se puede medir, y por tanto, que se puede contrastar con un valor teórico; para luego expresarlo en términos simbólicos. Retomando el ejemplo 1:

$$H_0 : \mu = \bar{x} = 588,7 \quad (1)$$

Noten que la proposición que queremos probar **no** es esa, sino que el número que obtuvimos a partir de nuestra muestra no es representativo de la población. Lo formulamos de esa forma, porque es aquella que más fácilmente se puede escribir (simbólicamente) en términos de una igualdad.

2. Una vez formulada la hipótesis nula, se formula la hipótesis alternativa. Dependiendo de la informa con la que se cuenta, puede ser cualquiera de las siguientes:

$$H_1 : \mu \neq \bar{x} \quad (2)$$

$$H_2 : \mu < \bar{x} \quad (3)$$

$$H_3 : \mu > \bar{x} \quad (4)$$

Dada la información que tenemos, podríamos usar H_1 o H_3 . La elección depende del grado de precisión que queremos lograr con la respuesta. En este caso, H_3 es más precisa que H_1

Paso 2: Seleccionar el nivel de confianza de las conclusiones

Como ya he mencionado, podemos utilizar todas las herramientas que hemos revisado hasta ahora para crear modelos o para tomar decisiones. En ambos casos, los resultados nunca son exactamente precisos. Idealmente, son suficientemente confiables para poder hacer algo con ellos, pero todo esto depende de este paso y los valores que ajustemos en él.

Es conveniente distinguir entre los tipos de errores que tratamos de *evitar* al realizar cualquier prueba de hipótesis. Estos tienen que ver con la conclusión que obtenemos respecto de la hipótesis nula y el *estado del mundo*. Es conveniente mostrar las distintas posibilidades.

	H_0 Verdadera	H_0 Falsa
Investigación rechaza H_0	Error tipo I	No hay problema
Investigación NO rechaza H_0	No Hay Problema	Error tipo II

Usualmente se denomina al error tipo I como α y al error tipo II como β . Generalmente trabajamos con α en mente, tratando de evitar cometer este error; para lo cual operacionalizamos esta cantidad como la **probabilidad de cometer un error tipo I**. Es decir, lo ideal es que $\alpha \rightarrow 0$, pero siendo realistas, debemos fijar un valor de α lo suficientemente pequeño como para confiar en el resultado, pero no tan pequeño que haga imposible trabajar con una muestra.

Esto varía de disciplina en disciplina, pero cuando tenemos muestras pequeñas y conocemos poco de la población, generalmente $\alpha \leq 0,05$ es suficiente. Esto se interpreta de la siguiente forma: si repito la medición en 100 muestras similares a la que realmente observé, voy a cometer un error tipo I en aproximadamente 5 de esas 100 veces.

Definición 3: Nivel de confianza de las conclusiones

Formalmente, definiremos *alpha* como la probabilidad de cometer un error de tipo I. Es decir, la probabilidad de rechazar erróneamente H_0 cuando es verdadera.

$$\alpha = \Pr(\text{Rechazar } H_0 | H_0 \text{ es verdadera}) \quad (5)$$

Paso 3: Determinar la prueba estadística apropiada

Dependiendo de lo que se haya formulado en el paso 1, se debe considerar cuál es la prueba estadística apropiada y cuál es la distribución de probabilidad al tomar muchos resultados provenientes de muestras. En el caso del ejemplo 1, estamos trabajando con una media aritmética. Sabemos (por el teorema del límite central), que las medias de diferentes muestras tienen una distribución normal. No siempre es el caso, si lo que nos interesa no es una media aritmética sino que otra cosa. Veremos algunos casos diferentes en lo que queda de este curso.

Paso 4: Decisión

Una vez que se tienen los resultados de la prueba estadística, ajustada a la distribución apropiada, se pueden utilizar dos criterios para tomar la decisión respecto de aceptar o no H_0

Valor p : Lo que generalmente aprendemos en clases de estadística en pregrado. Su uso tiene que ver con la distribución de las muestras y las regiones bajo la curva de la función de densidad de dicha distribución.

Intervalo de confianza: En lugar de entregar un valor, nos entrega un *intervalo* al que el verdadero parámetro (poblacional) pertenece.

En ciertos casos extremos, obtendremos resultados diferentes utilizando valores p o intervalos de confianza. Lo que usualmente ocurre es que ambos criterios concuerdan. Últimamente se han puesto bajo escrutinio los valores p , debido a su uso indiscriminado, muchas veces en situaciones en que no se justifica, o cuando sus resultados pueden ser engañosos (muestras muy grandes, por ejemplo). Por tanto, es pertinente conocer ambos criterios, fácilmente implementables.

Nos referiremos con más detalle a estos dos criterios en el siguiente documento.

Referencias

Triola, Mario. 2018. *Elementary statistics*. United States: Pearson.