

Respecto a la enseñanza del contraste de diferencia de medias para dos poblaciones normales y varianzas desconocidas

Alvaro Cortínez Pontoni¹

acortinezp@uta.cl

¹Departamento de Matemática, Universidad de Tarapacá, Arica, Chile

Resumen

En las asignaturas de pregrado, una componente muy importante es el tema de la inferencia y, en particular, las pruebas de hipótesis o, simplemente contrastes. Sin embargo, éstos son por lo general sólo vistos desde la perspectiva frecuentista. Los contrastes de hipótesis para diferencia de medias, por ejemplo, en que no se tiene información acerca de las varianzas, suelen ser tratados como una simple aproximación, siendo que existen ya en literatura soluciones prácticas muy fáciles de implementar. El presente artículo muestra cómo se puede tratar el problema basándose en la distribución t-student.

Palabras clave: Contrastos de Hipótesis, Bayes, Behrens y Fisher.

About teaching Hypothesis Testing of the Difference Between Two Mean of normal populations with unknown variances

Abstract

In undergraduate courses, a very important component are the subject of inference and, in particular, hypothesis testing or simply contrasts. However, these are usually only shown from the frequentist perspective. Hypothesis tests for difference of means, for example, where is no information about variances are usually treated as a simple approach, although there are practical solutions already in literature very easy to implement. This article shows how is possible to treat the problem based on the t-student distribution.

Keywords: Hypothesis Testing, Bayes, Behrens and Fisher.

À propos de l'enseignement de test d'hypothèse de la différence entre deux moyen des populations normales de variances inconnues

Résumé

Aux courses de pre-graduation, un élément très important est le sujet de l'inférence et, en particulier, la vérification de hypothèse ou, simplement de contrastes. Toutefois, ceux-ci sont généralement vues seulement de la perspective fréquentiel. Test d'hypothèses pour la différence de moyens, par exemple, où il n'y a pas d'information de variances, ce sont affichés comme un simple estimation, en étant qu'il déjà existe des solutions pratiques très facile d'implémenter dans la littérature. Cet article montre comment on peut gérer le problem basé en la distribution t-student.

Mots clés: Contrastes, Bayes, Behrens et Fisher.

1. INTRODUCCIÓN

En el proceso de aprendizaje de los estudiantes de pregrado, suele estar una asignatura de estadística, que comúnmente contiene elementos de estadística descriptiva, probabilidad y comienzos de inferencia. Se trata de una asignatura cuya dificultad no radica tanto en sus elementos

matemáticos, sino más bien conceptuales y, especialmente, de pensar en forma diferente las matemáticas. Mientras en cursos anteriores, los estudiantes se enfrentaban a las matemáticas de forma totalmente objetiva, se encuentran ahora con una asignatura que relativiza sus interpretaciones. Este cambio en el pensar es el que consigue que si no somos capaces de entregar correctamente los conceptos teóricos y de aplicaciones, los

estudiantes rápidamente desconectarán. Ya no les servirán las recetas o trucos para resolver los problemas, pues los problemas en sí no son complicados, más bien deberán ser capaces de utilizar el método adecuado y, sobretodo, extraer la información para interpretarla en el contexto del problema que se esté tratando.

Por otra parte, es importante destacar que la estadística está siendo enseñada en prácticamente todos sus niveles sólo desde ciertas perspectivas, sin considerar una visión más amplia que la haga más cercana al alumno, permitiéndole entender los fenómenos cotidianos a partir de la misma. De hecho, por ejemplo, en la mayoría de los programas de estadística inferencial tanto elemental como avanzada, que actualmente se imparten, ésta se trata desde el punto de vista frecuentista (Bernardo, 2002). De este modo, las interpretaciones difícilmente permiten al estudiante entenderlas como parte de la vida diaria. El hecho de que un intervalo de confianza o un contraste de hipótesis implique en su interpretación una reiteración indefinida del experimento bajo las mismas condiciones ya es un punto de bloqueo, pues los experimentos no siempre son juegos de azar o de laboratorio y situaciones de la vida cotidiana por lo general no se podrán replicar en condiciones similares. Esta noción de muestreo repetido de una población hipotética es la noción central de la inferencia estadística, lo cual, para los estudiantes crea una confusión a la hora de interpretar los resultados. Muchos profesores coinciden en la dificultad que significa la noción de distribución muestral (Albert, 2002). Entre los problemas que se plantean en los contrastes de hipótesis, por ejemplo, es que de las comparaciones de medias para poblaciones normales e independientes, sólo se consideran los casos en que las varianzas son conocidas o bien son desconocidas pero iguales. El caso más general, en que las varianzas también se consideran desconocidas, pero no necesariamente iguales, suele tratarse para muestras grandes. Sin embargo, el hecho de que las medias sean desconocidas implica que no hay motivos para suponer que en las aplicaciones habituales se puedan conocer las varianzas. Ni tampoco es de esperar que siempre se disponga de muchas observaciones.

Lo que se busca con este artículo es comparar los contrastes de hipótesis paramétricos sobre diferencia de medias de dos poblaciones normales. Se parte desde el punto de vista clásico, que es el que más difundido está. A continuación, se presentan las formulaciones bayesianas, mostrando que se ajustan de forma muy precisa al método científico de razonamiento.

Al no plantear restricciones sobre las varianzas, se verá que el enfoque bayesiano puede ser muy adecuado a la solución del contraste de hipótesis para diferencia de medias con varianzas desconocidas, que es el llamado Problema de Behrens y Fisher. Se propone, de esta forma, que ese caso sea tratado de forma adecuada y no como una solución sólo para cuando se tienen tamaños muestrales suficientemente grandes. Se espera que el estudiante al enfrentarse con un problema de comparación de medias, no tenga que restringirse o abundar en supuestos para encontrar una solución e interpretación.

Desde el punto de vista clásico, las pruebas de significación tienen su origen en los trabajos desarrollados en los años 20 por Fisher y en los años 30 por Neyman y Pearson. Ellos brindan un criterio objetivo para calificar las diferencias que se presentan al comparar los resultados de dos muestras, con la intención de explicar si dichas diferencias se mantienen dentro de los límites previstos por el diseño estadístico (un error y una confianza esperados) o si, por el contrario, la diferencia entre ellas resulta lo suficientemente grande como para inferir que ha ocurrido un cambio real en el indicador. La mayoría de los libros de inferencia estadística adoptan este procedimiento y es posible encontrarlo de una manera muy completa en Lehmann (2006). Las inferencias realizadas de esta forma, en muchas ocasiones pasan por alto elementos bastante más relevantes del análisis de los datos que el resultado del valor p ("p value") y sólo centran la discusión en el resultado de esta prueba.

La Inferencia Bayesiana, por su parte, aplica un marco teórico similar a la inferencia clásica u objetiva: hay un parámetro poblacional, digamos θ , que toma valores a partir de un espacio paramétrico Θ y respecto al cual se desea realizar inferencias; y se tiene un modelo que determina la probabilidad de observar diferentes valores de la variable aleatoria X en un espacio muestral Ω , bajo diferentes valores de los parámetros. Sin embargo, la diferencia fundamental es que la inferencia bayesiana considera al parámetro como una variable aleatoria, como una modelización de su incertidumbre, mientras que la clásica lo considera como un valor fijo pero desconocido (Bolstad, 2007; Lee, 2012). Este aspecto es de gran importancia, pues conduce a una aproximación diferente para realizar el modelamiento del problema y la inferencia propiamente dicha.

El pensamiento bayesiano y el área de aplicación de la metodología bayesiana es la misma que la de la estadística frecuentista, pero ofrece mayores ventajas en situaciones como los estudios de equivalencia (inferencia estadística) o con el tipo de situaciones que se ve enfrentado el científico habitualmente (teoría de decisión).

2. LA ENSEÑANZA DE LA ESTADÍSTICA DESDE EL PUNTO DE VISTA BAYESIANO

Diversos son los autores que en los últimos años han propuesto la introducción de elementos subjetivistas en la enseñanza de la estadística. Algunos de ellos, incluso proponen que la estadística se enseñe directamente utilizando el paradigma bayesiano. Berry (1997) afirma que la aproximación bayesiana es fácil de enseñar a los estudiantes de cursos de estadística elemental. Los estudiantes aprenden acerca del proceso de la ciencia. Sin embargo, ¿por qué no se ha incorporado la estadística bayesiana en la enseñanza? Berry presenta algunas razones:

1. La estadística bayesiana es inherentemente muy difícil para ser enseñada en cursos de nivel elemental;
2. La aproximación bayesiana es subjetiva y no provee métodos que lleven a los estándares de la objetividad requerida en ciencias;

3. Los bayesianos no se ponen de acuerdo en la elección de distribuciones a priori apropiadas;

4. Los métodos frecuentistas dominan en las disciplinas sustanciales.

En la aproximación bayesiana existen una serie de ideas básicas, de las cuales fluyen los cálculos y las inferencias. Por ejemplo, el Teorema de Bayes. El hecho de que la incertidumbre sea interpretada mediante probabilidades, los desarrollos subsecuentes se basan, por lo tanto, en seguir procesos de pensamiento lógico.

El hecho de que la aproximación bayesiana parta de creencias a priori subjetivas, precisamente se acerca más al razonamiento dentro de las ciencias, en que el investigador inicia su estudio en base a las hipótesis (subjetivas) que él formula.

Bolstad (2002) ha realizado muchas experiencias enseñando en cursos de estadística elemental el punto de vista bayesiano. Según él, ha obtenido buenos resultados. Sus estudiantes entienden las ideas claves de la inferencia bayesiana y aprecian que se pueden obtener mejores logros que con la visión frecuentista. También pudieron entender la importancia en la aleatorización durante la recolección de datos.

¿Cuáles son las ventajas de enseñar la inferencia estadística desde el punto de vista bayesiano? Según Albert (2002), pensar las probabilidades de forma bayesiana es más intuitiva que de forma frecuentista. Refleja mejor el sentido común acerca de la incertidumbre que los estudiantes tienen antes de tomar el curso de estadística. Habitualmente las personas usan expresiones como "más posible", "raro", "siempre", "nunca" para reflejar los diferentes grados de incertidumbre. Por otra parte, en los problemas de estimación, queremos tener la confianza de que un intervalo estimado contiene el parámetro de interés. Desde el punto de vista frecuentista, uno sólo puede confiar en el resultado del intervalo de confianza o el test de hipótesis bajo el supuesto de un gran número de repeticiones., pero esto no es útil cuando se realizan inferencias sobre un único conjunto de datos. La interpretación bayesiana es directa, pues tiene sentido asignarle probabilidades a los parámetros.

Batanero, Garfield, Ottaviani, y Truran, J. (2000) piensan que la escuela bayesiana no ha alcanzado en educación la misma relevancia que la escuela clásica, debido a razones filosóficas y tecnológicas. Ella afirma que en cada curso introductorio de inferencia los estudiantes debieran ser introducidos a los dos enfoques, sean los profesores o no bayesianos. Y exclama: "¡También pienso que es más fácil introducir estos fundamentos a los estudiantes que tratar de hacerles comprender un test clásico de hipótesis para diferencia de dos medias con varianzas desconocidas pero diferentes!". (p.11). Esta es la primera alusión a que debe incorporarse en la didáctica de la inferencia estadística el Problema de Behrens y Fisher. Sin embargo, este problema no ha sido abordado didácticamente hasta el momento.

Los Contrastes de Hipótesis de por sí generan problemas en los estudiantes. Vallencillos y Batanero (1997) enumeran algunos de ellos: interpretaciones incorrectas del nivel de significación incompreensión del tipo de validación de las hipótesis, dificultades en establecer las hipótesis. Hoekstra, Morey y Rouder (2014) se centra en las interpretaciones incorrectas de los intervalos de confianzas y los contrastes de hipótesis. Además, pone en un mismo nivel dichos errores de interpretación, tanto del punto de vista clásico como bayesiano. Tratar los intervalos frecuentistas y bayesianos de forma intercambiable lleva a pensar incorrectamente. Por ejemplo, la lógica de rechazar el valor de un parámetro si está fuera del intervalo de confianza no es válida para el intervalo creíble bayesiano, pues de hecho no es una técnica bayesiana válida.

Finalmente, es interesante destacar algunas de las concepciones erróneas que los estudiantes (y las personas en general) suelen tener en torno a la inferencia (Shaughnessy, 1992):

- confianza injustificada en muestras pequeñas;
- poco respeto por las pequeñas diferencias en las muestras grandes;
- idea errónea de que cualquier diferencia entre las medias de dos grupos es significativa.

Según Shaughnessy, la mayoría de las personas no reconocerían una diferencia significativa desde el punto de vista estadístico entre muestras grandes y adecuadas cuando se las mostraran, ni se darían cuenta de que una muestra seleccionada en forma cuidadosa de unos cuantos cientos de instancias puede decir mucho de una población muy grande.

3. EL CASO DE DOS POBLACIONES NORMALES

Uno de los casos que más habitualmente se estudian en inferencia estadística, es el de los contrastes sobre los parámetros de poblaciones normales. Considerando que la distribución normal se encuentra en la mayoría de los ejemplos prácticos, es muy simple de usar, y además surge como un caso límite para otras distribuciones cuando se tienen tamaños muestrales grandes, es de gran utilidad estudiar sus parámetros, realizando los contrastes adecuados en torno a ellos.

El objetivo fundamental de la inferencia de dos poblaciones es el poder tomar las decisiones respecto a la comparación de los parámetros en dos poblaciones (por ejemplo normales). Es así como se busca saber, por ejemplo si se pueden o no considerar iguales las medias y/o varianzas de ambas poblaciones.

Al comparar las medias de dos poblaciones, realizamos el "Contraste para Diferencia de Medias". Para eso es útil saber si las varianzas de dichas poblaciones son o no conocidas. Además, en caso de ser desconocidas, se puede

comparar si son o no iguales. Todo esto con el fin de buscar el estadístico del contraste más adecuado.

La forma más común de empezar, en los cursos de inferencia, es comparando medias de dos poblaciones normales de varianzas conocidas. En este caso, tenemos dos muestras aleatorias, x_1, x_2, \dots, x_{n_1} e y_1, y_2, \dots, y_{n_2} , de tamaños muestrales n_1 y n_2 , y de poblaciones normales $N(\mu_1; \sigma_1^2)$ y $N(\mu_2; \sigma_2^2)$, respectivamente, donde las varianzas σ_1^2 y σ_2^2 son conocidas.

Partimos de la distribución en el muestreo de los promedios:

$$\bar{x}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ y } \bar{x}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

y se usa el estadístico:

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Basta ahora con definir las hipótesis, el nivel de significancia del test y el estadístico bajo la hipótesis nula para construir la Región Crítica o de rechazo y tomar una decisión.

El funcionamiento es bastante mecánico. Es necesario, eso sí, dejar claro cómo formular las hipótesis, el criterio para definir el nivel de significancia y, finalmente, llevar adecuadamente el resultado al contexto del problema que se está estudiando.

3.1. Comparación de Medias de dos Poblaciones Normales Independientes y de Varianzas Desconocidas.

Cuando se busca realizar contrastes sobre las medias de poblaciones normales, es muy habitual no tener información de las varianzas. Si las muestras son lo suficientemente grandes, se puede utilizar el Teorema Central del Límite y por consiguiente la aproximación normal.

Considérense las muestras aleatorias x_1, x_2, \dots, x_{n_1} e y_1, y_2, \dots, y_{n_2} , de poblaciones con distribución normal $N(\mu_1; \sigma_1^2)$ y $N(\mu_2; \sigma_2^2)$, respectivamente, donde las varianzas σ_1^2 y σ_2^2 son ahora desconocidas. Si las muestras son grandes, trabajamos con el estadístico del contraste bajo la hipótesis nula:

$$Z_c = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0; 1)$$

Sin embargo, si las muestras son pequeñas, toma mayor importancia la información que tengamos de las varianzas. El caso en que las varianzas se puedan considerar iguales se tiene una solución basada en una varianza común (mancomunada), a partir de la cual se construye el estadístico del contraste:

$$T_c = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{S_m^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

$$\text{con } S_m^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

El hecho de que no se conozcan las medias y, que por lo tanto se estén haciendo inferencias sobre ellas, lleva a pensar que también es muy probable que las varianzas no se conozcan. Estamos así frente al problema de Behrens y Fisher.

3.2. El Problema de Behrens y Fisher

El llamado Problema de Behrens y Fisher fue planteado primero por Behrens (1929) y Fisher (1935). Se trata de un problema abierto aún. Aparentemente es un problema muy simple, básicamente una extensión de los casos anteriores (varianzas conocidas y varianzas desconocidas e iguales), pero no lo es. No se ha podido encontrar que exista una solución clásica exacta que cumpla los criterios clásicos de pruebas de hipótesis confiables. Esto genera un problema en el momento de enseñar este tema en los cursos de pregrado, pues, sobre todo a estudiantes que no son de áreas afines a la estadística, es difícil explicarles que deben solucionar un problema solamente para situaciones específicas o en base a aproximaciones.

En el caso en que las muestras sean razonablemente grandes, las diferencias entre las distintas soluciones que se hayan planteado en la literatura son muy pequeñas y de hecho basta con usar el Teorema del Límite Central. Cuando los tamaños de las muestras son pequeños, sin embargo, surgen diferentes teorías para dar lugar a soluciones diferentes. Diversos autores han tratado el Problema de Behrens y Fisher desde que fuera planteado: Fisher (1939), Welch (1938), Aspin (1948), Fisher & Healy (1956), Patil (1965), Lee & Gurland (1975), Weerahandi (1987), Ruben (2002) entre muchos otros.

La siguiente es una solución aproximada, frecuentista, que se suele presentar en cursos de pregrado:

Sean x_1, x_2, \dots, x_{n_1} e y_1, y_2, \dots, y_{n_2} , dos muestras aleatorias de n_1 y n_2 observaciones tomadas respectivamente de dos poblaciones con distribución normal $N(\mu_1; \sigma_1^2)$ y $N(\mu_2; \sigma_2^2)$, donde σ_1^2 y σ_2^2 son desconocidas. Dadas las hipótesis:

$$H_0: \mu_1 - \mu_2 = d_0 \\ H_1: \mu_1 - \mu_2 \neq d_0$$

Usamos el estadístico del contraste (suponiendo la hipótesis nula cierta):

$$T_c = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_g$$

donde

$$g = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right]^2}{\frac{\left(\frac{s_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{n_2 - 1}} - 2$$

(g es redondeado al entero más próximo).

Entonces la región crítica de la prueba, una vez fijado en nivel de significancia (o tamaño de la prueba) α es:

$$RC = \{T_c : |T_c| > t_{g, 1-\alpha/2}\}$$

Esta es la llamada "Aproximación de Welch" al problema de Behrens y Fischer y g la aproximación de Welch a los grados de libertad.

Existen otras aproximaciones clásicas, que si bien dan una solución práctica, no son muy confiables, por lo que no serán tratados en este documento.

4. SOLUCIÓN BAYESIANA AL PROBLEMA DE BEHRENS Y FISHER

Desde el punto de vista bayesiano, los problemas suelen tener un camino bien definido: se requiere una distribución de probabilidades para la muestra, una distribución de probabilidades a priori para los parámetros y luego se actúa en función del Teorema de Bayes, identificando en lo posible la distribución a posteriori del parámetro. Este procedimiento es especialmente útil cuando estamos haciendo contrastes de hipótesis sobre los parámetros, pues en este caso nos basamos en la distribución de los mismos y en cómo ésta varía, a través del Teorema de Bayes, desde la *a priori* hasta la *a posteriori*.

En un problema relacionado con los parámetros de una distribución normal, puede ser más realista hacer el supuesto de que ambos parámetros son desconocidos en lugar de que sólo la varianza lo sea.

Para comenzar, observemos el caso de una población Normal:

$$X \sim N(\mu; \sigma^2) \Leftrightarrow f(x|\mu; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(x - \mu)^2}{\sigma^2}\right\}$$

La función de densidad conjunta para una muestra de tamaño n se obtiene como un producto de las densidades univariantes:

$$\begin{aligned} f(x_1, x_2, \dots, x_n | \mu; \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(x_1 - \mu)^2}{\sigma^2}\right\}^{-\frac{1}{2}} \cdot \dots \\ &\cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(x_n - \mu)^2}{\sigma^2}\right\}^{-\frac{1}{2}} \end{aligned}$$

Desarrollando este producto se puede comprobar que:

$$f(x_1, x_2, \dots, x_n | \mu; \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \cdot \exp\left[-\frac{1}{2}\{S + n(\bar{x} - \mu)^2\}/\sigma^2\right]$$

en el que se define

$$S = \sum (x_i - \bar{x})^2$$

Se requiere de una distribución a priori para los parámetros. Al no tener estudios anteriores, la forma más simple de enfrentar este problema es considerando una distribución a priori de referencia:

$$\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$$

que es el producto de las distribuciones a priori de referencia $\pi(\mu) \propto 1$ para μ y $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ para σ^2 . En Lee (2012) y Berger (1985) se demuestra y argumenta que a priori μ y σ^2 se pueden considerar independientes. Por lo tanto, mediante un desarrollo algebraico no muy complicado se llega a:

$$\begin{aligned} \pi(\mu, \sigma^2 | x) &\propto (\sigma^2)^{-(v+1)/2-1} \\ &\cdot \exp\left[-\frac{1}{2}\{S + n(\bar{x} - \mu)^2\}/\sigma^2\right] \end{aligned}$$

donde $v = n - 1$.

Observando esta última expresión (la distribución a posteriori), se puede ver que hay un producto de dos distribuciones, una normal y una gamma inversa, lo cual es razonable con los supuestos sobre las distribuciones de μ y σ^2 .

Es decir:

$$\begin{aligned} \pi(\mu | x) &\propto \{S + n(\bar{x} - \mu)^2\}^{-(v+1)/2} \\ \pi(\sigma | x) &\propto \sigma^{-v/2-1} \exp\left(-\frac{1}{2} S / \sigma^2\right) \end{aligned}$$

Por lo tanto, desde el punto de vista bayesiano es posible realizar inferencias sobre los parámetros del modelo normal, pues toda la información actualizada de ellos está contenida en las distribuciones a posteriori y, de hecho son distribuciones conocidas. Por lo tanto, ya sea con el uso de tablas o de computador, se puede efectuar cualquier cálculo a partir de ellas.

4.1. La Distribución de Behrens y Fisher

Extenderemos, ahora, el desarrollo anterior para dos poblaciones normales. Tanto las medias como las varianzas son desconocidas. Volvemos a los vectores independientes $x = (x_1, x_2, \dots, x_{n1})$ e $y = (y_1, y_2, \dots, y_{n2})$ de modo que $x_i \sim N(\mu_1, \sigma_1^2)$ e $y_i \sim N(\mu_2, \sigma_2^2)$

Dadas las hipótesis:

$$\begin{aligned} H_0: \mu_1 - \mu_2 &= \text{do} \\ H_1: \mu_1 - \mu_2 &\neq \text{do} \end{aligned}$$

Definamos:

$$\begin{aligned} s_x^2 &= \frac{S_x}{n_1 - 1} = \frac{S_y}{v_1} \\ s_y^2 &= \frac{S_y}{n_2 - 1} = \frac{S_y}{v_2} \end{aligned}$$

Es necesario encontrar distribuciones a priori para los parámetros μ_1, μ_2, σ_1^2 y σ_2^2 . En principio supondremos distribuciones de referencia. De este modo, según Lee se

deduce que las distribuciones a posteriori de μ_1 y μ_2 son independientes tales que:

$$T_x = \frac{\mu_1 - \bar{x}}{s_x/\sqrt{n_1}} \sim t_{v_1} \quad \text{y} \quad T_y = \frac{\mu_2 - \bar{y}}{s_y/\sqrt{n_2}} \sim t_{v_2}$$

Se definen los estadísticos T y θ mediante:

$$T = \frac{\bar{x} - \bar{y} - d_0}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}}$$

$$\tan(\theta) = \sqrt{\frac{s_x^2/n_1}{s_y^2/n_2}}$$

(consideremos a θ dentro del primer cuadrante). Se puede comprobar que:

$$T = T_x \cos \theta - T_y \sin \theta$$

Considerando que a partir de los datos se conoce θ y que las distribuciones de T_x y T_y también son conocidas, podemos obtener la distribución de T . Dicha distribución está tabulada y se conoce con el nombre de Behrens (Behrens, 1929). La denotaremos como

$$T \sim \text{BF}(f_1, f_2, \theta).$$

De esta forma, tenemos una solución exacta al problema de Behrens y Fisher. Las complicaciones que puedan aparecer van más bien por el lado de los cálculos, pues el problema teórico quedaría así resuelto. Sin embargo, con programas computacionales ya no tenemos restricción alguna y puede ser utilizada sin mayor problema en cualquier curso de inferencia.

Desde el punto de vista pedagógico, al disponer de las herramientas adecuadas, es más razonable enseñar la solución partiendo de la distribución exacta más que con formas aproximadas. El Modelos de Behrens no suele estar incluido en la lista de modelos que se les enseña a los estudiantes y, de hecho, no requiere estarlo, pues es obtenido a partir del desarrollo del Teorema de Bayes utilizando modelos conocidos. Por supuesto que si a priori se utilizara alguna distribución distinta a la no informativa, quizás basados en estudios previos, el problema de Behrens y Fisher obtendrá mejores resultados.

Considerando que los Contrastes de Hipótesis también son útiles para estudiantes no relacionados con áreas de matemática o ingeniería, puede ser razonable trabajar una aproximación de la solución exacta, de modo que se puedan utilizar herramientas más “cercanas” a ellos. A continuación se presenta una aproximación al Problema de Behrens y Fisher.

4.2. Aproximación de Patil

Para cursos en que el cálculo numérico no es un requisito o los estudiantes no tienen una buena base matemática, se puede acudir a la aproximación de Patil (1965).

Para utilizar esta aproximación, es necesario encontrar:

$$a = \left(\frac{n_1 - 1}{n_1 - 3}\right) \sin^2 \theta + \left(\frac{n_2 - 1}{n_2 - 3}\right) \cos^2 \theta$$

$$b = \frac{(n_1 - 1)^2}{(n_1 - 3)^2(n_1 - 5)} \sin^4 \theta + \frac{(n_2 - 1)^2}{(n_2 - 3)^2(n_2 - 5)} \cos^4 \theta$$

$$g_1 = 4 + \frac{a^2}{b}$$

$$g_2 = \sqrt{\frac{a(g_1 - 2)}{b}}$$

Entonces, aproximadamente

$$T/g_2 \sim t_{g_1}.$$

Debido a que g_1 no es necesariamente un número entero, puede ser necesaria una interpolación en la tabla de la distribución t - student para utilizar esta aproximación, aunque también es útil usar un programa computacional, por ejemplo Excel. Obsérvese que al trabajar con una distribución t -student, no requerimos de tablas adicionales.

El siguiente es un ejemplo simple, de elaboración propia, que puede ser visto en cualquier clase elemental de contraste de hipótesis.

4.3. Un ejemplo

Se está estudiando el aumento de peso (durante 53 días) de cierto tipo de ratas bajo diferentes dietas. El primer grupo, formado por $n_1=13$ ratas recibió una dieta de alto valor proteico, obteniéndose los siguientes datos:

Rata número	Aumento de peso (x_i)
1	134
2	146
3	118
4	125
5	124
6	161
7	113

Rata número	Aumento de peso (x_i)
8	83
9	106
10	129
11	90
12	124
13	107

El segundo grupo, de $n_2 = 7$ ratas recibió una dieta baja en proteínas:

Rata número	Aumento de peso (y_i)
1	70
2	118

3	101
4	85
5	107
6	132
7	94

Queremos contrastar si hay una diferencia significativa en el aumento de peso entre los dos grupos alimentados por dietas diferentes:

$$H_0: \mu_1 = \mu_2 \Leftrightarrow H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 \neq \mu_2 \Leftrightarrow H_1: \mu_1 - \mu_2 \neq 0$$

Para realizar el contraste, extraemos la siguiente información:

$$n_1 = 13, n_2 = 7$$

$$f_1 = 12, f_2 = 6$$

$$\bar{x} = 120, \bar{y} = 101$$

$$S_x = 5378.045,$$

$$S_y = 2551.98$$

$$s_x^2 = 448.17,$$

$$s_y^2 = 425.33$$

Por lo tanto:

$$\tan \theta = \frac{s_x/\sqrt{n_1}}{s_y/\sqrt{n_2}} = \left(\frac{448.17/13}{425.33/7} \right)^{\frac{1}{2}} = 0.8$$

de modo que $\theta = 37$ (aproximadamente), y:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{(s_x^2/n_1 + s_y^2/n_2)}} = 1.91$$

Así $T \sim BF(12,6,37)$. En una tabla de BF o mediante un programa computacional a un nivel del 5% de significación (bilateral) se tiene $BF(12,6,37) = 2.36$.

El mismo resultado se puede obtener directamente de la aproximación de Patil. Se tienen los siguientes valores:

$$a = 1.38, b = 0.41, g_1 = 8.66, g_2 = 1.03.$$

Así $T/1.03 \sim t_{8.66}$ y tenemos que a un nivel del 5%, $t_{8.66}$ es 2.3060. Se concluye, entonces, que a un nivel del 5% de significación se puede afirmar que los datos no arrojan suficiente evidencia para indicar que hay diferencia significativa en el aumento de peso entre los dos grupos de ratas alimentados por dietas diferentes.

Para efectos de la enseñanza de la inferencia, aun cuando pueda surgir algún problema computacional, el nivel de comprensión al que se puede llegar desde el punto de vista bayesiano es mayor que simplemente planteando el problema y dando soluciones que únicamente son aproximaciones a lo que no se conoce como solución misma.

5. CONCLUSIONES

La inferencia estadística en general y, los contrastes de hipótesis en particular, suelen ser enseñados en cursos de pregrado sólo desde el punto de vista frecuentista.

El contraste de hipótesis para diferencia de medias en que las varianzas son desconocidas y no necesariamente iguales suele enseñarse en los cursos de pregrado en base a una aproximación o para casos en que los tamaños muestrales son suficientemente grandes para utilizar el modelo normal.

El Problema de Behrens y Fisher tiene una solución bayesiana exacta, basada en la distribución de Behrens y Fisher y utilizando una distribución a priori no informativa.

Con las tecnologías actuales, el Problema de Behrens y Fisher puede ser tratado sin mayor dificultad, incluso en carreras no relacionadas con las matemáticas.

REFERENCIAS

- Albert, J. (2002). College students' perception of probability. (Tech. Report 2002-12). Bowling Green State University, Department of Mathematics and Statistics.
- Aspin, A., (1948). An examination and further development of a formula arising in the problem of comparing two mean values, *Biometrika*, 35(1), 88-96.
- Batanero, C., Garfield, J.B., Ottaviani, M.G. & Truran, J. (2000). Research in statistical education: some priority questions. *Statistical Education Research Newsletter*, 1(2), 2-6.
- Behrens, W. A. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen, *Landwirtschaftliche Jahrbücher*, 68, 807-837.
- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, (2nd Ed). Berlin: Springer-Verlag.
- Bernardo, J. (2002). Un programa de síntesis para la enseñanza universitaria de la estadística matemática contemporánea. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*, 95(1-2), 81-99.
- Berry, D. (1997). Teaching elementary bayesian statistics with real applications in science. *The American Statistician* 51(3), 241-246.
- Bolstad, W.M. (2007). *Introduction to bayesian Statistics* (2nd. Ed.). Hoboken, New Jersey, EEUU: John Wiley & Sons, Inc.
- Bolstad, W.M. (2002). Teaching bayesian statistics to undergraduates: Who, What, Where, When, Why and How. *Proceedings of ICOTS6 International Conference on Teaching Statistics*, Capetown, South Africa.
- Fisher, R. A., (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, 6(4), 391-398.
- Fisher, R. A., (1939). The comparison of samples with possibly unequal variances, *Annals of Eugenics*, 9(2), 174-180.

- Fisher, R. A., & Healy, M. J.R (1956). New tables of Behrens test of significance. *Journal of the Royal Statistical Society, Series B*, 18 (2), 212-216.
- Hoekstra, R, Morey, R. & Rouder, J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21(5), 1157-1164.
- Lee, A., Gurland, J., (1975). Size and power of tests for equality of means of two normal populations with unequal variances, *Journal of the American Statistical Association*, 70(3), 933-944.
- Lee, P. (2012). *Bayesian Statistics An Introduction* (4th Ed.). London. UK: John Wiley and Sons.
- Lehmann, E.L. & Romano, J.P. (2006). *Testing statistical hypotheses* (3rd. Ed.) New York, EEUU: Springer Texts in Statistics.
- Patil, V. H., (1965). Approximations to the Behrens-Fisher distributions, *Biometrika*, 52(1), 267-271.
- Ruben, H., (2002). A simple conservative and robust solution of the Behrens-Fisher Problem. *Sankhya: The Indian Journal of Statistics*, 64(A-1), 139-155.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. En D. A. Grows (Eds.), *Handbook of research on mathematics teaching and learning* (pp. 465 – 494). New York: MacMillan.
- Vallecillos, A. y Batanero, C. (1997). Aprendizaje y enseñanza del contraste de hipótesis: concepciones y errores. *Enseñanza de las Ciencias*, 15(2), 189-197.
- Weerahandi, S., (1987). Testing regression equality with unequal variances. *Econometrica*, 55(5), 1211-1215.
- Welch, B.L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350–362.

Alvaro Cortínez Pontoni

Estadístico y Licenciado en Matemáticas en la Pontificia Universidad Católica de Chile (1991), Doctor en Ciencias Matemáticas en el departamento de Estadística de la Universidad Complutense de Madrid (1998). Actualmente es Profesor Asociado del Departamento de Matemática de la Universidad de Tarapacá, Arica, Chile. Es miembro del Grupo de Investigación en Enseñanza de la Estadística Bayesiana y pertenece a la Red Latinoamericana de Investigación en Educación Estadística Su investigación se ha centrado en la Estadística Bayesiana, en particular en el Análisis de Confiabilidad. Sus líneas de investigación actual son los modelos lineales desde el punto de vista bayesiano y la didáctica de la estadística. En estadística está estudiando el problema R & R sin réplicas, buscando una solución bayesiana y su correspondiente implementación. En educación estadística, analiza la factibilidad de incorporar ideas bayesianas en los contenidos de probabilidad e inferencia estadística para cursos de pregrado. En particular, se estudia la concepción del parámetro y los obstáculos de aprendizajes que asociados a él puedan presentarse en la comprensión de la inferencia estadística.