

INFERENCIA INFORMAL EN CLASES DE ESTADÍSTICA USANDO CÓDIGO R EN DISPOSITIVOS MÓVILES

Calandra, María Valeria^a; D'Urzo, Paula G.^b, Di Paolantonio, Anyelen^a, De Cortazar, Cecilia^a

^aUIDET GAMEFI, Dto.de Ciencias Básicas, Facultad de Ingeniería, UNLP

^b Dto. de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata
maria.calandra@ing.unlp.edu.ar

Introducción

El objetivo de este trabajo es desarrollar dispositivos que promuevan la enseñanza funcional de temas específicos de test de hipótesis para los alumnos de carreras científico-tecnológicas. Dentro de los temas más importante de inferencia estadística para la aplicación en el campo experimental, se encuentra el tema de test de hipótesis, pero también son abundantes los reportes sobre las dificultades sobre su enseñanza y aprendizaje. Los cálculos e interpretación de los errores de Tipo I y de Tipo II son fundamentales para entender su lógica. Por eso es que se propone una enseñanza contextualizada, para estas temáticas con el uso de una aplicación que se puede implementar en código de programación R en teléfonos móviles en el que los alumnos pueden simular el proceso de extracción de una muestra aleatoria de una fábrica bajo ciertas condiciones para realizar la certificación de un artículo mediante un test de hipótesis y calcular empíricamente las probabilidades de errores de Tipo I y de Tipo II. El estudiante tiene a su disposición el código del programa y lo puede modificar de modo de simular distintos escenarios de trabajo. Esta propuesta permite emular el trabajo científico de investigación en inferencia estadística. Además, permite que los estudiantes apliquen el lenguaje de programación R de fundamental importancia en el trabajo científico en el campo de la estadística. Esta propuesta permite que los alumnos tomen un rol activo en la construcción del conocimiento de estas temáticas ya que pueden intervenir desde la simulación de la muestra con la que se realizará el test como en la construcción de común acuerdo en el transcurso de la clase de los conceptos de error de Tipo I y de Tipo II y la re significación de los mismos para la situación planteada.

La motivación para hacer este trabajo surgió de la problemática de la enseñanza de estas temáticas. Existen numerosos errores e interpretaciones incorrectas del contraste de hipótesis, que se han encontrado incluso en los trabajos de investigación (Batanero, 2000; Falk y Greenbaum, 1995; Harradine y otros, 2011). El establecimiento de hipótesis adecuadas a la situación es el primer paso en la resolución de un problema de contraste de hipótesis pero presenta grandes dificultades de comprensión para los estudiantes que no logran identificar cuáles serían las hipótesis adecuadas en cada caso, no comprenden el papel que juegan en el proceso o confunden las hipótesis nula y alternativa (Vallecillos Jiménez, 1997). Liu y Thompson (2009) encontraron una falta de comprensión de la lógica de los test de hipótesis ligada a la comprensión de los resultados. Además, Vallecillos Jiménez (1994) realizó una amplia investigación sobre el aprendizaje del tema del contraste de hipótesis estadísticas en alumnos universitarios y detectó que algunos alumnos suponen que la suma de las probabilidades de cometer un error de Tipo I y un error de Tipo II es uno y esto no es así, los dos tipos de errores son eventos incompatibles, pero no son sucesos complementarios. Se han encontrado estudiantes que creen que el cambio del nivel de significación no afecta al riesgo de error de Tipo I en la decisión. Algunos confunden la significación estadística y con relevancia práctica, otros asocian un resultado significativo como uno que corrobora la hipótesis nula. Batanero y otros (2017) observó que los alumnos manifiestan una dificultad en cuanto a la comprensión, cuando calculan la probabilidad de cometer un error de Tipo I perdiendo de vista la condición (la verdad de la hipótesis nula). Muchos estudiantes asocian de forma inmediata el rechazo de la hipótesis al error de Tipo I. Vallecillos Jiménez (1997) encontró conflictos en la lógica de la interpretación de los test de hipótesis ligadas a la suposición de que los test de hipótesis proporcionan una demostración matemática deductiva de la verdad de la hipótesis y que el nivel de significación de un test

es la probabilidad de que H_0 sea cierta dado que se ha rechazado. Es decir, se interpreta incorrectamente el resultado del test por creer que el contraste demuestra la hipótesis o porque calcula su probabilidad. También en la misma publicación se identifican errores en el aprendizaje de los alumnos asociados al cálculo e interpretación del p-valor. El valor p se calcula mediante el cálculo de la probabilidad de observar el valor empírico del estadístico o un valor más extremo, dado que la hipótesis nula es verdadera y varía de una muestra a otra y muchas veces es confundido con el nivel de significación que fija el experimentador antes de realizar el test. Un concepto que se suele comprender erróneamente es el nivel de significación. Respecto a este concepto se pueden encontrar numerosos conflictos, el más frecuente consiste en intercambiar los dos términos de la probabilidad condicional, interpretándolo como la probabilidad de que la hipótesis nula sea cierta, una vez que la decisión de rechazarla se ha tomado (Krauss y Wassner, 2002). También se han reportado conflictos para distinguir lo que es un parámetro y lo que es un estadístico o no reconocen que un estadístico es una variable aleatoria, lo que lleva a plantear mal los test de hipótesis, por ejemplo, los alumnos confunden la media muestral con la media poblacional (Korin, 2021; Harradine y otros, 2011).

Marco Teórico

Las nuevas teorías didácticas conciben al alumno como un verdadero protagonista de su aprendizaje, en donde el profesor proponga actividades contextualizadas que permitan el surgimiento de un modo funcional las técnicas, elementos tecnológicos y organizaciones matemáticas que se quieren enseñar. La Teoría Antropológica de lo Didáctico (TAD), propone la aplicación de un dispositivo didáctico que denomina Actividad de Estudio e Investigación (AEI), para enfrentar lo que ha denominado la monumentalización del saber y la pérdida del sentido de lo que se estudia en algunos ámbitos educativos (Chevallard, 2013). Dicho dispositivo permite introducir en los sistemas de enseñanza procesos de estudio funcionales, tanto a nivel cognitivo como procedimental, concibiendo al alumno con un rol activo en el proceso de enseñanza/aprendizaje. Para ello, las AEI se organizan en torno a una pregunta generatriz Q_0 , seleccionada por el profesor, que tenga el potencial de generar el estudio por parte de los alumnos de ciertos contenidos matemáticos. Se propone una investigación y debate, en este caso, centrado en torno a la pregunta generatriz: Q_0 : ¿Cómo certificar una característica de un producto o proceso? La búsqueda de respuesta a la pregunta también generará más preguntas derivadas, que permitirán contextualizar la propuesta, como puede ser ¿Cómo certificar la longitud del alabe de la turbina de un motor? Las respuestas a las preguntas junto con otras derivadas de una actividad específica llevarán al estudio de técnicas y de elementos tecnológicos-teóricos específicos relativos a test de hipótesis, como la probabilidad de cometer un error de Tipo I: ¿Cuál es la probabilidad de que se concluya que la pieza no certifica los estándares cuando en realidad si lo hace? O la probabilidad de cometer un error de Tipo II: ¿Cuál es la probabilidad de que la auditoría concluya que la pieza certifica los estándares cuando en realidad no lo hace?

Propuesta

La propuesta que se presenta a continuación está dirigida, en particular, a alumnos de carreras científico-tecnológicas que estén cursando probabilidad y estadística. La AEI servirá para complementar la enseñanza formal de estos conceptos y se resolverá mediante la aplicación de inferencia informal (Zieffler y otros, 2008) con el uso de la aplicación para dispositivos móviles y web: Run R Script - Online Statistical Data Analysis - Version 1.1 (Liila Tech, 2018) que permite programar usando el lenguaje R versión 3.5.2 (Team R.C., 2021), y Python (Python Software Foundation, 2017), es un entorno de desarrollo integrado (IDE) en línea, incluye una consola, un editor de código que admite la ejecución directa de código, así como herramientas, archivos de datos definidos por el usuario, almacenamiento de código en la nube. Este software genera una cuenta en la nube para cada usuario, donde se almacenan los códigos que desee el mismo, con el fin de ser

compartidos con los usuarios mediante diferentes métodos (redes sociales, email, etc.), permitiendo ser ejecutados sin poseer una instalación del programa. Se propone, en este caso, compartir con los alumnos en el aula las rutinas mediante un código QR que “encripte” el link a la dirección web de la rutina en forma directa.

Para comenzar la actividad el profesor propone un debate acerca de la certificación de un producto y de los métodos estadísticos para certificar la calidad del mismo, que iniciará el estudio del saber a construir. La pregunta generatriz del debate será: ***Q₀: ¿Cómo certificar una característica de un producto?***, esta pregunta actuaría como eje articulador para la reconstrucción de los temas relativos a test de hipótesis. A su vez esta pregunta podría derivar en otras *¿Qué significa certificar un producto? ¿Qué productos se pueden certificar? ¿Qué característica de un producto se podría certificar?* Supongamos el caso en que se desea certificar una dimensión de un alabe de turbina de un motor a reacción de un avión (ver Figura 1).

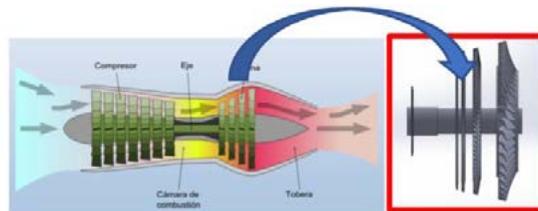


Figura1. Esquema Típico de un motor a reacción de avión, a la izquierda el rotor de la turbina de varios discos.

Como se observa en la sección de la turbina existen varios discos conformados por una cierta cantidad de alabes, insertados en los mismos, estos deben poseer una longitud media específica en función del diseño del motor (Ver Figura 2)

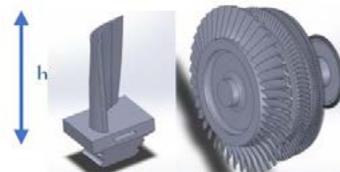


Figura 2. Imagen de un alabe y de los discos donde se insertan los mismos.

En un caso particular, de un motor de una marca determinada, la longitud de un alabe para el disco de mayor diámetro se puede modelar como una variable aleatoria con distribución aproximadamente Normal con media $\mu = 269$ mm, y una desviación estándar $\sigma = 0,15$ mm.

Actividad:

Una fábrica de motores a reacción está trabajando de acuerdo a los estándares de calidad. De todas formas, anualmente debe someterse a una auditoría para verificar dicha situación. La auditoría se realiza mediante un test de hipótesis aplicado a la longitud media de los álabes producidos mediante una muestra aleatoria extraída de la producción.

Propuesta 1: El objetivo de esta propuesta es resignificar desde un punto de vista empírico el concepto de error de Tipo I y de probabilidad de error de Tipo I para un test de hipótesis.

Para realizar la actividad se le propone a cada alumno generar con la aplicación Run R Script una muestra aleatoria X_1, X_2, \dots, X_{10} donde cada X_i tenga distribución Normal con media 269 mm y desviación estándar $\sigma = 0,15$ mm, que serviría para simular la extracción

de una muestra aleatoria de la fábrica que está trabajando de acuerdo a los estándares de calidad.

Para poder realizar la certificación de la longitud media de la producción de los álabes de esa marca producidos por una fábrica, se propone a cada alumno, realizar con la muestra obtenida el siguiente test de hipótesis bilateral:

$$H_0: \mu=269\text{mm} \quad \text{contra la hipótesis alternativa } H_1: \mu \neq 269\text{mm}$$

Donde rechazar H_0 significaría que la fábrica no cumple los estándares. La regla de decisión del test podría ser rechazar H_0 si $\left| \frac{\bar{X}-269}{\frac{0,15}{\sqrt{10}}} \right| > z_{0,05}$ siendo $\bar{X} = \frac{X_1+X_2+\dots+X_{10}}{10}$ Donde $z_{0,05}$ corresponde al percentil 95 de la distribución Normal estándar.

Luego se plantea a los estudiantes las siguientes preguntas *¿Alguno rechazó H_0 con la muestra seleccionada? ¿Es correcto haber rechazado H_0 ? ¿Si se realiza el test 10.000 veces con distintas muestras, cuantas veces rechazaríamos H_0 ? ¿Se podría representar gráficamente usando las 10.000 muestras de tamaño 10 la distribución empírica de \bar{X} ? ¿Cómo se podría representar la región de rechazo de H_0 en dicha gráfica? ¿Cómo se podría calcular la frecuencia relativa de veces que la fábrica no satisface los estándares de calidad? ¿Cómo se podría interpretar esa proporción en términos de nuestro problema? ¿Cómo se podría calcular la frecuencia relativa de veces que la fábrica podría certificar la pieza?* Para contestar estas preguntas los estudiantes deben correr una rutina utilizando Run R Script con el código que se muestra a continuación en el enlace que se brinda a los mismos mediante la vinculación con un código QR <https://r.varisk.xyz/r/220611/demo-Rout-1558816771-dark.html>

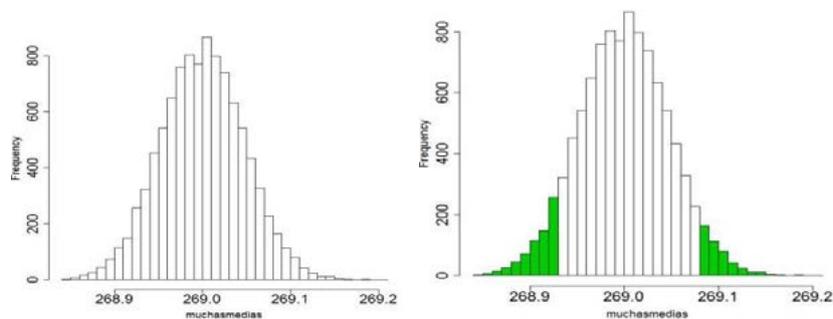


Figura 3. Histogramas de la distribución de frecuencias de las medias muestrales Propuesta 1.

El saber hacer estas tareas se justifica por el hecho de cada alumno puede generar sus propias muestras e identificar gráficamente la distribución de las frecuencias absolutas de \bar{X} cuando la distribución subyacente de la muestra es Normal con media 269 y desviación estándar 0.15 (Figura 3, izquierda). Esto serviría para contestar la pregunta: *¿Se podría representar gráficamente usando las 10.000 muestras de tamaño 10 la distribución empírica de \bar{X} ?* Luego pueden identificar gráficamente en color verde en el histograma (Figura 3, derecha), la frecuencia de veces que se rechazará H_0 cuando en realidad es verdadera, con lo que se puede estimar la probabilidad de cometer un error de Tipo I si se la divide por 10.000:

$$P(\text{error de Tipo I}) = P(\text{Rechazar } H_0 / \mu = 269) \cong \frac{103}{10.000}$$

Esta, difiere un poco de la verdadera probabilidad que es 0,1 por tratarse de una simulación. Esto serviría para contestar estas preguntas: *¿Qué proporción de \bar{X} caen en la zona de rechazo? ¿Cómo se podría interpretar esa proporción en términos de nuestro problema?*

En el mismo gráfico de la Figura 3 de la derecha, se puede apreciar la región complementaria, que no está pintada de verde, que serviría para estimar la probabilidad del test de no rechazar H_0 cuando la misma es verdadera que es lo que se esperaría que sucediera en este caso y no corresponde al error de Tipo II.

Esta forma de trabajo permite que los estudiantes intervengan en el diseño del programa realizando distintas tareas: generar las muestras aleatorias bajo las condiciones requeridas, calcular el promedio de cada una de ellas, construir un histograma con los promedios, identificar en el histograma los promedios que caen en la zona de rechazo del test, cambiar si quisieran el tamaño de las muestras o el número de simulaciones. Por otro lado, esta propuesta permitiría que los alumnos se familiaricen con el uso del lenguaje de programación R, que es fundamental para cálculos en estadística. Esta forma de trabajo muestra cómo se llevan a cabo las investigaciones estadísticas y las principales ideas que subyacen en dichas investigaciones.

Propuesta 2: El objetivo de esta propuesta es resignificar desde un punto de vista empírico el concepto de error de Tipo II y de probabilidad de error de Tipo II para el test de hipótesis aplicado a la misma actividad.

Para realizar la actividad se le propone a cada alumno generar con la aplicación Run R Script una muestra aleatoria X_1, X_2, \dots, X_{10} donde cada X_i tenga distribución Normal con media 269,04 mm y desviación estándar $\sigma = 0,15$ mm, que serviría para simular la extracción de una muestra aleatoria de la fábrica que no está trabajando de acuerdo a los estándares de calidad.

Luego se plantean las preguntas *¿Algún alumno no pudo rechazar H_0 con la muestra seleccionada? ¿Si se realiza el test 10000 veces con distintas muestras, cuantas veces rechazaríamos H_0 ? ¿Se podría representar gráficamente usando las 10000 muestras de tamaño 10 la distribución empírica de \bar{X} ? ¿Para qué proporción de valores de \bar{X} se rechaza H_0 ? ¿Qué proporción de valores de \bar{X} no caen en la zona de rechazo? ¿Cómo se podría identificar esa región en el histograma? ¿Cómo se podría interpretar esta proporción en términos de nuestro problema? ¿Qué proporción de valores de \bar{X} caen en la zona de rechazo de H_0 ? ¿Cómo se podría interpretar esta proporción en términos de nuestro problema?* Para contestar estas preguntas los estudiantes deben correr una rutina utilizando Run R Script con el código que se muestra a continuación en enlace que se brinda a los mismos mediante la vinculación con un código QR <https://r.varisk.xyz/r/220611/demo-Rout-1530181579-dark.html>

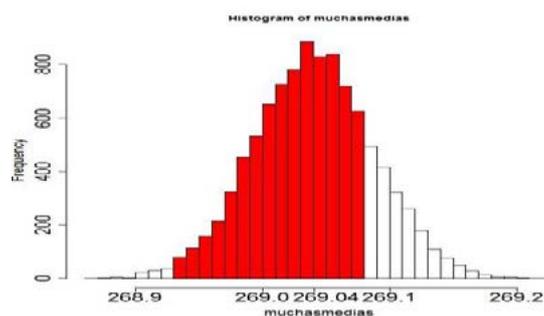


Figura 4. Histograma de la distribución de frecuencias de las medias muestrales Propuesta 2.

El alumno puede generar sus propias muestras e identificar la región de aceptación de H_0 cuando es falsa o error de Tipo II (ver zona de color rojo, en Figura 4), mediante la distribución empírica de \bar{X} cuando la distribución subyacente de la muestra es Normal con media 269,04 y desviación estándar 0,15. Lo que permitiría estimar empíricamente la probabilidad de cometer un error de Tipo II, es decir:

$$P(\text{error de Tipo II}) = (No\ Rechazar\ H_0/\mu = 269,04) \cong \frac{8.431}{10.000}$$

Se puede realizar nuevamente la experiencia con una muestra de tamaño $n = 20$ o más para ver que a medida que se toman muestras más grandes de la producción dicha probabilidad disminuye. También se podría generar muestras con una media más lejos de 269, por ejemplo 269,06 en lugar de 269,04 mm y observar también que la probabilidad de cometer error de Tipo II disminuye.

Conclusiones

Esta propuesta de enseñanza se ha aplicado a alumnos de la facultad de ingeniería de la Universidad Nacional de La Plata, como trabajo futuro se evaluará su impacto en el aprendizaje. La misma pregunta generatriz es posible presentarla en otros contextos y orientar la propuesta para distintas carreras científico-tecnológicas. El uso de la aplicación Run R Script ofrece grandes ventajas para la enseñanza dado que se puede implementar en un aula de clases, pudiéndose ejecutar en un dispositivo móvil y sin tener la aplicación instalada, esto resulta superior a otras propuestas de este tipo (Batanero y Diaz, 2015), ya que esta aplicación no solo tiene esta capacidad, sino que además permite programar y variar los códigos, lo cual expande el universo de posibilidad de estudios de diferentes aspectos de la temática. Además, se les facilita a los estudiantes el código del programa propuesto para la actividad, de modo que ellos simplemente lo ejecuten como está o puedan cambiar algunas variables para explorar el comportamiento de las probabilidades de los errores de Tipo I y II en distintas situaciones, como, por ejemplo: ¿Qué pasa con dichas probabilidades si aumentamos el número de piezas extraída de la fábrica? El hecho de usar la simulación para los cálculos los ayudaría a comprender mejor cuál es la función del condicionante en el cálculo de las probabilidades de error de Tipo I y de Tipo II y de sus complementos. Además, permitiría emular el trabajo científico en inferencia estadística que parte de un hecho empírico para luego generalizarlo a la población.

Bibliografía

- Arteaga, G.R. Cañadas, M.M. Gea, B. Giacomone y M.M. López-Martín (Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*. Disponibles en, enfoqueontosemiotico.ugr.es/civeos.html
- Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning*, 2(1-2), 75-98.
- Batanero, C. & Diaz, C. (2015). Aproximación informal al contraste de hipótesis. M. Contreras (Ed.), *II Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatoria*. Granada: SEIEM.
- Batanero, C., Díaz, C. & López-Martín, M.M. (2017). Significados del contraste de hipótesis, configuraciones epistémicas asociadas y algunos conflictos semióticos. En J.M. Contreras, P. Arteaga, G.R. Cañadas, M.M. Gea, B. Giacomone y M.M. López-Martín (Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*. Disponible en: <http://enfoqueontosemiotico.ugr.es/civeos.html>.
- Chevallard, Y. (2013). Enseñar matemáticas en la sociedad de mañana: alegato a favor de un contraparádigma emergente. *Journal of Research in Mathematics Education*, 2(2), 161-182.
- Falk, R., & Greenbaum, C.W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology*, 5(1), 75-98. <https://doi.org/10.1177/0959354395051004>
- Harradine, A., Batanero, C. & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 235-246). Springer, Netherlands. https://doi.org/10.1007/978-94-007-1131-0_24
- Korin, C. (2021). La comprensión de los test de hipótesis estadísticos. Un estudio con alumnos universitarios. *Revista De Educación Matemática*. Recuperado a partir de <https://revistas.unc.edu.ar/index.php/REM/article/view/10263>

- Krauss, S. & Wassner, K. (2002). How significance tests should be presented to avoid the typical misinterpretations. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Statistical Institute and International Association for Statistical Education.
- Liila Tech (2018). Run R Script - Online Statistical Data Analysis - Version 1.1 Build 226-208 – Android
- Liu, Y. & Thompson, P.W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies: An International Journal*, 4(2), 126-138. <https://doi.org/10.1080/15544800902741564>
- Python Software Foundation (2017). Python Software Foundation. Retrieved from <https://www.python.org/>
- Team, R. C. (2021). R: A Language and Environment for Statistical Computing.
- Vallecillos Jiménez, A. (1994). Estudio teórico-experimental de errores y concepciones sobre el contraste estadístico de hipótesis en estudiantes universitarios. [Tesis Doctoral]. Universidad de Granada, Departamento de Didáctica de la Matemática, España.
- Vallecillos Jiménez, A. (1997). El papel de las hipótesis estadísticas en los contrastes: concepciones y dificultades de aprendizaje. *Educación Matemática*, 9(2), 5-20.
- Vanderman, S.B. (2002). Providing "real" context in statistical quality control courses for engineers. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Statistical Institute and International Association for Statistical Education.
- Viles, E. (2007). Análisis didáctico de la estadística y la calidad en los estudios de Ingeniería Industrial. *Tecnura*, 11(21), 54–62. <https://doi.org/10.14483/22487638.6272>
- Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.