

ACERCAMIENTO INFORMAL A LA INFERENCIA ESTADÍSTICA

Gudelia Figueroa Preciado, Irma Nancy Larios Rodríguez, María Elena Parra Ramos

Universidad de Sonora (México)

gfiguero@mat.uson.mx, nancy@mat.uson.mx, meparra@mat.uson.mx

Palabras clave: p-valor, nivel de significancia, nivel de confianza, inferencia estadística.

Key words: p-value, significance level, confidence level, statistical inference.

RESUMEN: La estadística es cada vez más utilizada en diversas áreas del conocimiento ya que permite, no sólo la toma de decisiones, sino el poder cuantificar la probabilidad de tomar éstas equivocadamente; cuantificación que es posible, por lo general, a través de técnicas estadísticas como son estimación de parámetros y pruebas de hipótesis. Estas técnicas requieren el manejo de ciertos conceptos estadísticos tales como nivel de significancia, nivel de confianza, p-valor, etcétera. En este trabajo se presentan diversas actividades que con el uso de simulación computacional, análisis descriptivo de datos y un cálculo sencillo de probabilidades, permiten realizar un acercamiento informal a la estadística inferencial y con ello tomar decisiones sin la necesidad de conocer los conceptos antes mencionados.

ABSTRACT: Statistics is widely used in many fields of knowledge because it not only allows us to take decisions, but quantify the probability of taking these incorrectly; the quantification of these decisions is usually possible throughout some statistical techniques, like parameter estimation and hypothesis tests. These techniques require the correct use of some statistical concepts like level of significance, confidence level, p-value, etcetera. In this work we present some activities involving the use of computer simulation, descriptive data analysis and some very simple probability computations; these activities allow students to have an informal approach to statistical inference and to take some decisions with no prior knowledge of the statistical concepts previously mentioned.

■ INTRODUCCIÓN

La estadística, cada vez más utilizada en diversos campos del conocimiento, nos permite tomar decisiones y cuantificar el grado de incertidumbre de éstas, razón por la cual se ha vuelto de uso común ya que es posible justificar de manera cuantitativa las decisiones que se toman, tanto en problemas científicos como aquellos que surgen en el quehacer diario. Esta toma de decisiones generalmente involucra herramientas como son la estimación de parámetros, pruebas de hipótesis y pruebas de significancia, que son técnicas muy útiles, ampliamente utilizadas hoy en día, pero cuyo uso y comprensión sigue siendo difícil para muchos estudiantes o usuarios de ellas, tal como lo señalan Garfield y Ahlgren (1998), Vallecillos y Batanero (1997), así como Figueroa, Larios y Parra (2014), quienes abordan las dificultades que presentan la mayoría de los estudiantes cuando se introducen los conceptos necesarios en la aplicación de la inferencia estadística, así como en el razonamiento inductivo que ésta conlleva.

Por otra parte, al analizar la currícula de muchas de las licenciaturas que se imparten en la Universidad de Sonora, se puede constatar que en varias de ellas sólo se incluye un curso de estadística descriptiva, por lo que, los estudiantes finalizan su carrera sin un conocimiento que les permita tomar decisiones con base en el análisis sencillo de una muestra. Aunado a lo anterior, en las carreras donde se incluye un curso de estadística inferencial, éste no considera muchas de las pruebas estadísticas que posiblemente el estudiante requerirá usar en su futuro desempeño profesional. Sin embargo, el objetivo no es aumentar la currícula con más cursos de estadística, sino desarrollar en los estudiantes un razonamiento intuitivo que les permita tomar una decisión, con base en la muestra observada y sin necesidad del manejo de los conceptos formales utilizados en estadística inferencial. Este razonamiento se lo proporciona el trabajar actividades que involucren un acercamiento informal a la estadística inferencial, el cual puede lograrse fácilmente con el uso de simulaciones computacionales, que hoy en día pueden efectuarse en software diverso y que permiten cuantificar la probabilidad de observar la muestra obtenida, bajo un supuesto establecido.

En el presente trabajo se muestran dos actividades en las que se resuelven problemas que requerirían el uso de pruebas de hipótesis o bien pruebas de significancia, pero que, con el manejo de simulaciones computacionales, que se realizaron en el software estadístico R, fue posible cuantificar la probabilidad de observar dichos datos, bajo un supuesto establecido y tomar una decisión.

■ REFERENTE TEÓRICO

La mayoría de los cursos de estadística, a nivel universitario, tienen como propósito llegar a desarrollar y utilizar algunos métodos de inferencia estadística. A pesar de la valiosa utilidad de la estadística inferencial, diversos estudios de investigación como los publicados por Hubbard y Bayarri (2003), Garfield y Ahlgren (1998), concuerdan con respecto a la dificultad que muestran los estudiantes tanto para comprender, como para utilizar técnicas inferenciales. Dicha problemática ha generado preocupación en el seno de la comunidad académica y ha propiciado el buscar alternativas menos formales para comprender el razonamiento que conlleva el tomar una decisión, con base en la probabilidad de ocurrencia de la muestra observada.

Investigadores como Makar, Baker y Ben-Zvi (2011), así como Garfield y Ben-Zvi (2008), proponen desarrollar un razonamiento informal de la estadística inferencial, incluso desde los niveles básicos de educación. El término inferencia informal o razonamiento informal de la inferencia estadística, ha sido tema de interés en recientes investigaciones en educación estadística, de forma tal que en los años recientes se ha publicado un gran número de artículos al respecto; en ellos se explica que un razonamiento inferencial informal se puede desarrollar con el uso de actividades que permitan obtener conclusiones de manera informal, o bien realizar predicciones acerca de la población partiendo de la observación de patrones, representaciones, medidas estadísticas y modelos estadísticos de muestras aleatorias, mientras que al mismo tiempo se presta atención a las limitaciones que el tamaño de la muestra puede tener al realizar estas inferencias. De la misma manera, Zieffler, Garfield, Delmas y Reading (2008), opinan que este término se refiere a la forma en la cual los estudiantes utilizan un razonamiento informal para elaborar argumentos que soporten inferencias acerca de poblaciones no conocidas, a partir de muestras observadas, e indican también que se debe ir más allá de los datos al realizar generalizaciones, esto es, se deben utilizar los datos como evidencia y hacer uso de un lenguaje probabilístico para expresar la incertidumbre que conllevan estas generalizaciones.

■ METODOLOGÍA

A continuación se presentan dos actividades en las que se debe tomar una decisión acerca del parámetro o parámetros en estudio. Estas actividades se han utilizado exitosamente durante varios semestres y sólo requieren el uso de algún software que permita realizar las simulaciones que en las mismas se detallan.

Primera actividad

Esta primera actividad resulta muy familiar a los estudiantes ya que consiste en lanzar varias veces una moneda que supondremos es legal, es decir, una moneda donde la probabilidad de cara es igual a la probabilidad de sello, ambas fijas en 0.5. La variable en estudio puede ser el contabilizar el número de caras que resultan en cierto número de lanzamientos. Para abordar este tipo de problema los estudiantes deben haber visto previamente distribuciones discretas de probabilidad, de esta manera logran identificar fácilmente a esta variable, como una variable aleatoria que sigue una distribución binomial con parámetros n y p , que denotan el número de lanzamientos independientes realizados y la probabilidad de éxito, en este caso la probabilidad de obtener cara.

Se recomienda que el número de ensayos a realizar sea tal que permita calcular fácilmente probabilidades de ocurrencia de la variable aleatoria. Aprovechando la familiaridad que los estudiantes tienen con este experimento, se les pregunta cuáles son los resultados más probables de ocurrir, si suponemos que la moneda a utilizar es legal y se realizan veinte lanzamientos. Las respuestas que externan comúnmente son el considerar muy probable que el resultado sea 8, 9, 10, 11 o bien 12 caras; otros eventos les resultan improbables o raros, si la moneda es legal, esto es, si $p = 0.5$.

En un momento posterior y después de analizar los resultados que consideraron más probables cuando $p = 0.5$, se les solicita proponer posibles resultados, de los veinte lanzamientos, si la moneda lanzada tuviera una probabilidad de $p = 0.6$, $p = 0.7$ y $P = 0.8$, de que el lanzamiento resulte en cara. Una vez discutidos los posibles resultados propuestos por los estudiantes y con el

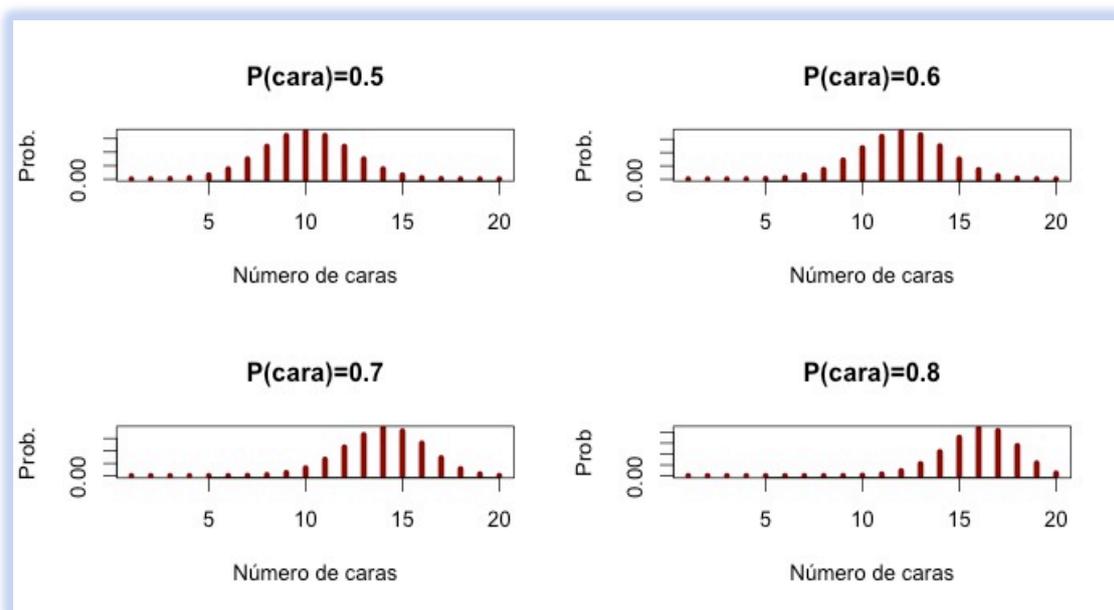
apoyo de algún software estadístico, aquí se utilizó el software libre R, se calcula la probabilidad de obtener desde cero hasta veinte caras, en veinte lanzamientos. Este cálculo se realiza suponiendo primero que la moneda es legal, esto es, que $p = 0.5$, y después se repite este procedimiento suponiendo que la probabilidad de que la moneda caiga en cara, es de $p = 0.6$, $p = 0.7$ y $p = 0.8$. Las probabilidades asociadas a los eventos $P(X = 0)$, $P(X = 1)$, ..., $P(X = 20)$, se muestran en la Figura 1.

Después del análisis anterior se plantea a los estudiantes el suponer que se obtuvieron 16 caras, en los veinte lanzamientos de la moneda y se les solicita que, analizando las distribuciones mostradas en la Figura 1, respondan: ¿cuál de ellas concuerda más con la muestra observada? Luego se les solicita considerar como cierto el supuesto de que la moneda es legal y seleccionar el escenario que mejor respalda dicho supuesto.

Cuando los estudiantes observan los diagramas presentados en la Figura 1 deciden, casi de manera unánime, que el evento de obtener diez y seis caras, en veinte lanzamientos, está respaldado mayormente cuando la probabilidad de cara al lanzar la moneda, es $P = 0.8$, seguida de $P = 0.7$, pero ese evento lo consideran raro si $P = 0.6$, y más raro aún cuando $P = 0.5$; así pues, de esta manera tan sencilla, los estudiantes han decidido rechazar el supuesto de que la moneda es legal cuando observen la ocurrencia de diez y seis caras en veinte lanzamientos. Esto es, han realizado de manera informal una prueba de significancia sin necesidad de hablar o calcular un p-valor.

Es muy útil retomar esta actividad posteriormente, cuando se introduce el tema de pruebas de significancia o bien, cuando se presenta el enfoque de pruebas de hipótesis, pues los diagramas presentados en la Figura 1 permiten visualizar de una manera muy sencilla, el rechazo de la hipótesis nula para diferentes niveles de significancia.

Figura 1. Distribuciones binomiales para $n=20$ y diferentes valores de p .



Segunda actividad

A continuación se propone una actividad que permite al estudiante realizar, de una manera informal y sencilla, la comparación de las medias de dos poblaciones.

Consideremos la situación de comparar dos diferentes tipos de empaques de carnes. Uno de ellos es un empaque al vacío y el otro utiliza CO₂. Para llevar a cabo esta comparación se dispone de siete cortes de carnes, similares entre sí. Se seleccionan aleatoriamente tres de ellos y se empacan al vacío, mientras que los cuatro restantes se empacan utilizando una mezcla de CO₂. Después de cinco días de permanecer en una temperatura controlada, similar al ambiente que tendrían en un exhibidor de supermercado, se cuenta el número de bacterias presentes en cada uno de los cortes de carne. Para analizar este tipo de resultados, es muy común trabajar con el logaritmo del número de bacterias, obteniendo:

Empaque al vacío: 4.48, 5.43, 4.84.

Empaque con CO₂: 3.53, 3.38, 3.63, 3.74.

Analicemos de una manera descriptiva estas muestras, que sólo tienen tres y cuatro datos respectivamente, y de las cuales se obtienen las siguientes medidas:

Tabla 1. Medidas descriptivas para los dos tipos de empaque

Tipo de empaque	Mínimo	Máximo	Media	Desviación estándar
Al vacío	4.48	5.43	4.917	0.4796
Con CO ₂	3.38	3.74	3.570	0.1530

La diferencia entre los conteos promedios de bacteria observados con el empaque CO₂ y el empaque al vacío resulta:

$$3.570 - 4.917 = -1.347$$

Las preguntas que surgen, de manera natural, son: ¿Cómo interpretar esa diferencia? ¿Cómo utilizar esta información para decidir si los dos tipos de empaques de carne producen, en promedio, la misma cantidad de bacterias?

Para responder las preguntas anteriores debemos analizar si el valor -1.347 representa un resultado que podría esperarse, o presentarse, cuando los dos tipos de empaque producen, en promedio, el mismo número de bacterias. Para realizar este análisis, volvamos a los resultados observados en la muestra de siete cortes de carne, de acuerdo al tipo de empaque:

Tabla 2. Resultados obtenidos con los dos tipos de empaque

4.48	5.43	4.84	3.53	3.38	3.63	3.74
VACIO	VACIO	VACIO	CO ₂	CO ₂	CO ₂	CO ₂

¿Podrían estos resultados haber ocurrido por azar o en realidad el conteo observado de bacterias sí depende del tipo de empaque utilizado? Si suponemos que dichos resultados ocurren por azar, entonces el conteo de bacterias observado en la muestra, ocurriría independientemente del tipo de empaque que se utilizara; esto podemos simularlo fijando el renglón de datos y permutando los tipos de empaque; con ello estamos suponiendo que los resultados ocurren por azar y no dependen del tipo de empaque.

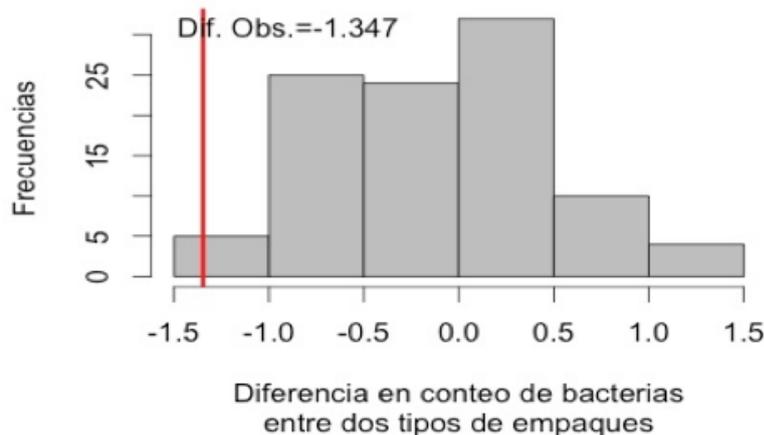
La permutación de las etiquetas “Vacío”, “Vacío”, “Vacío”, “CO2”, “CO2”, “CO2” y “CO2”, pueden realizarse fácilmente en un software como R, y el repetirla digamos cien veces, nos permite analizar el comportamiento que el azar arrojaría. A continuación se muestra una parte de las cien permutaciones realizadas en el software R.

Tabla 3. Cien permutaciones de las etiquetas de los tipos de empaque.

4.48	5.43	4.84	3.53	3.38	3.63	3.74
VACIO	CO2	CO2	VACIO	VACIO	CO2	CO2
CO2	VACIO	VACIO	VACIO	CO2	CO2	CO2
VACIO	VACIO	VACIO	CO2	CO2	CO2	CO2
CO2	VACIO	CO2	CO2	VACIO	VACIO	CO2
.						
.						
.						
VACIO	CO2	VACIO	CO2	CO2	VACIO	CO2

El siguiente paso es calcular, para cada una de las cien muestras simuladas, la diferencia entre los conteos promedios de bacterias obtenidos con el empaque que contiene CO2 y el empaque al vacío. Con estas cien diferencias de medias muestrales se construyó el histograma que se muestra en la Figura 2. Ubiquemos ahora, en este histograma, la diferencia media obtenida, en el experimento realizado con los siete cortes de carne, o sea, el valor de -1.347, pues esto permite visualizar qué tan probable o improbable resulta el observar una diferencia de -1.347, cuando el histograma se construye suponiendo que los conteos de bacterias observados pueden ocurrir bajo cualquier tipo de empaque; esto es, que los dos tipos de empaque pueden originar cualesquiera de los siete datos observados.

Figura 2. Histograma de las diferencias de medias obtenidas en la simulación



Cuando el estudiante nota que lo observado en la muestra no es un evento que ocurra comúnmente por azar, pues es un evento que se localiza en la cola inferior de resultados contruidos bajo el supuesto de que las bacterias crecen por igual en los dos tipos de empaque, le resulta fácil concluir que los empaques deben producir, en promedio, un conteo de bacterias diferente. Inclusive, observando las medias de la Tabla 1, puede decir que es más conveniente utilizar el empaque donde se utiliza CO₂.

Es importante señalar que para resolver el problema planteado en esta segunda actividad, no fue necesario el verificar supuestos distribucionales, el plantear hipótesis, el definir nivel de significancia o bien el calcular un p-valor. La conclusión se obtuvo de un análisis muy intuitivo que puede ser fácilmente comprendido desde un nivel descriptivo de la estadística y sin la necesidad de introducir conceptos de inferencia estadística.

■ CONCLUSIONES

La experiencia obtenida al utilizar actividades que involucran un tipo de razonamiento informal de la inferencia estadística, como las presentadas en los dos ejemplos incluidos, ha mostrado que éstas permiten al estudiante tomar una decisión, con base en la muestra observada, que es equiparable a la decisión que obtendría realizando pruebas de hipótesis o bien pruebas de significancia.

Es innegable la gran ayuda que el uso de software brinda para la realización de estas actividades, por lo que se sugiere enormemente su uso ya que éste permite al estudiante familiarizarse con el aspecto de cuantificar un evento ocurrido, en términos de una probabilidad calculada bajo un supuesto establecido.

Finalmente, es importante mencionar que todos los aspectos que se analizan de manera muy intuitiva en estas actividades, facilitan la presentación de conceptos como p-valor, hipótesis nula, nivel de significancia, etcétera, que se presentarán posteriormente en los temas de inferencia estadística.

■ REFERENCIAS BIBLIOGRÁFICAS

- Figueroa, G., Larios, N., & Parra, M. (2014). ¿Prueba de hipótesis, intervalo de confianza o prueba de significancia? *Acta Latinoamericana de Matemática Educativa*, 27, págs. 491-498. P. Leston (Ed).
- Garfield, J. B., & Ben-Zvi, D. (2008). Connecting Research and Teaching Practice. En *Developing Students' Statistical Reasoning*. The Netherlands: Springer.
- Garfield, J., & Ahlgren, A. (1998). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, 19 (1), 44-63.
- Hubbard, R., & Bayarri, M. (2003). Confusion Over Measures of Evidence (p 's) Versus Errors (α 's) in Classical Statistical Testing. *The American Statistician*, 57 (3), 171-182.
- Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13 (1-2), 152-173.
- Vallecillos, A., & Batanero, C. (1997). Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitarios. *Recherches en Didactique des Mathematiques* (17), 29-48.
- 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.