Revista de Educación Matemática
Volumen 34, N° 1 (2019), páginas 11 – 27
Unión Matemática Argentina - FAMAF (UNC)

CONTRASTES DE HIPÓTESIS MEDIANTE TÉCNICAS DE SIMULACIÓN

Aldana M. González Montoro

Resumen. El avance de la tecnología y la disponibilidad de grandes cantidades de datos hacen a la estadística cada vez más indispensable en todas la diciplinas. Luego, es importante contar con estrategias para enseñar la estadística de forma que resulte atractiva a estudiantes de todos los perfiles.

Los contrastes de hipótesis son una herramienta de la inferencia estadística que nos permiten responder preguntas acerca de una población basándonos en los datos de una muestra.

El objetivo del presente artículo es el de construir la intuición de los contrastes de hipótesis utilizando técnicas de simulación. Estas técnicas nos permiten abordar el desarrollo de los métodos estadísticos desde el comienzo de un curso, sin necesidad de presentar la teoría matemática.

Introducción

El continuo avance en la tecnología nos provée cada día de enormes cantidades de datos en tópicos de lo más variados: economía, educación, películas, deportes, comida, medicina, de opinión pública, comportamiento social, por mencionar algunos. Además, la potencia computacional actual nos permite también analizar esos datos en formas muy eficientes. En este escenario, la estadística se vuelve indispensable en casi todas las disciplinas. Por esto, la inclusión de la enseñanza de la estadística en todos los ámbitos (niveles y disciplinas) se vuelve cada día más necesaria. Sin embargo, es muy normal que tras tomar unas clases o un curso introductorio de estadística, las y los estudiantes terminen con una colección de fórmulas y métodos que creen no relacionados entre sí. Esto es porque para llegar a tratar los temas centrales de inferencia estadística, como *intervalos de confianza* o los *contrastes de hipótesis* primero hay que recorrer la teoría de la probabilidad, los modelos probabilísticos y las distribuciones estadísticas, lo que implica introducir hipótesis, fórmulas y teoremas. Tanta formalidad matemática hace muy fácil que se pierda de vista el objetivo principal, que es el de analizar un conjunto de datos. En este artículo se busca mostrar que es posible introducir los temas centrales de la estadística de forma intuitiva, antes de presentar la teoría.

Muchas veces queremos responder preguntas acerca de una población basándonos en la información de una muestra. Por ejemplo, supongamos que le preguntamos a 100 estudiantes de un colegio si quieren ir a la Universidad y 61 de ellos/as nos responden que sí. ¿Podemos afirmar que más del $60\,\%$ de las y los estudiantes de ese colegio quieren ir a la Universidad? O, consideremos el caso en que un equipo médico en un ensayo clínico, le da a 10 pacientes con cierta enfermedad un nuevo medicamento y a otros 10 un placebo. Pasado un tiempo ven que 8 del primer grupo y 6 del segundo mejoraron. ¿Son estos resultados concluyentes para decir que el nuevo medicamento sirve? O, supongamos que basándonos en la evidencia queremos determinar si Lionel Messi es mejor que la media de los jugadores de futbol pateando penales o no. Los procedimientos estadísticos que permiten contestar este tipo de preguntas se llaman contrastes de hipótesis. Estos, usan la información provista por una muestra para inferir acerca de la veracidad de una conjetura.

Aunque el camino para llegar a la teoría de los contrastes de hipótesis es largo y algo abstracto, es posible introducirlos conceptualemente sin utilizar modelos formales ni fórmulas matemáticas. En este artículo, mostraremos una forma de hacerlo, haciendo uso de métodos de simulación. Así, una vez expuesta la idea detrás de los contrastes y comprendido su uso, puede presentarse la teoría, que resultará más fácil de comprender.

Al mismo tiempo, mostraremos los métodos de simulación como herramienta para la enseñanza de la estadística mediante ejemplos y la aplicación a datos reales. En general, los métodos enseñanza basados en la simulación permiten un abordaje más cercano a la intuición de las y los estudiantes introduciendo el proceso estadístico completo desde el comienzo del curso. Además, el uso de tecnología desde las primeras clases, vuelve la materia más llamativa para el estudiantado y hace más fácil mostrar la potencia de la estadística como herramienta para encontrar respuestas a preguntas interesantes. Existen libros de texto con este enfoque, por ejemplo Lock et al. y Tintle et al. (Lock, Lock, Morgan, Lock, y Lock, 2013; Tintle y cols., 2015), aunque están en inglés.

Todas las simulaciones que se muestran en este artículo son explicadas en términos de experimentos físicos para facilitar el entendimiento. Luego fueron realizadas por computadora utilizando el software estadístico libre R (Team, 2018), pero no es parte de los objetivos de este artículo mostrar cómo se realizan. Existen muchas opciones de software libre y fácil de usar para llevarlas a cabo, algunas se listan al final del artículo.

Conceptos previos

En su famoso libro "The Design of Experiments” (Fisher, 1971), el estadístico inglés Ronald Fisher expone la siguiente situación. Una señora afirma que ella puede determinar, con solo probarla, si al preparar una taza de té con leche, la leche fue agregada antes o después del té. Supongamos que le ofrecemos una taza a esta señora y ella señala acertadamente que la leche fue incorporada luego del té. ¿Es esto suficiente para decir que la señora es realmente capaz de distinguir el orden? Claramente no, podría haber adivinado por casualidad. Supongamos ahora que le ofrecemos 8 tazas en vez de una. En este caso, aunque sea cierto que puede distinguir, no esperamos que acierte todas las veces, podemos esperar que se equivoque algunas. Pero, ¿cómo podemos saber si miente o no? Nuestro objetivo en este ejemplo es determinar cuál es el verdadero caso.

Los contrastes de hipótesis nos permiten responder preguntas como estas tomando en cuenta la evidencia que hay en los datos. Pero, para hablar de contrastes de hipótesis es necesario manejar algunos conceptos estadísticos previos, que no serán expuestos en detalle aquí. Sin embargo, haremos un recorrido rápido por algunos de ellos. Sí es importante destacar que todos los conceptos previos necesarios para llegar a los contrastes de hipótesis pueden ser abordados en un curso de la misma forma que se expone en este artículo, utilizando ejemplos, datos reales y simulaciones por computadora.

Recorramos brevemente el proceso estadístico:

1. El proceso comienza con el planteo de una pregunta que puede ser respondida mediante el análisis de un conjunto de datos.
2. Luego, viene la recolección de los datos en una *muestra*. Esto involucra elegir la *población de estudio*, armar un plan de muestreo, y realizar la recolección de forma sistemática y cuidada.
3. Al comenzar el análisis, los datos deben ser explorados, esta etapa se llama *análisis exploratorio de datos* e involucra realizar tablas y gráficos y buscar patrones que tengan que ver con nuestra pregunta.
4. **Inferencia estadística**. Se usa la información de la muestra para buscar tendencias y estudiar la fuerza de esas tendencias . Por último, se usan estos resultados para entender y sacar conclusiones acerca de la población de estudio entera. Esto permite tomar decisiones informadas acerca del problema en cuestión.

En todo este proceso debemos entender la diferencia entre *población*, que incluye todos los individuos u objetos de interés, y *muestra*, que es un subconjunto de esa población. También es importante conocer la diferencia entre *parámetro* y *estadístico muestral*. El primero, es un número que describe una característica de la población entera, es un número fijo, generalmente desconocido. El segundo es un número que se calcula con los datos de la muestra y por lo tanto es aleatorio.

Por ejemplo, el censo de 2010 de la República Argentina dice que el $56,05\,\%$ de la población de tres años y más de la provincia de Córdoba utiliza computadora (GEOCENSO). El número $0{,}5605$ es un parámetro poblacional, la *proporción* (real) de la población de tres años y más de la provincia de Córdoba utiliza computadora, a esta proporción la denotamos por p. Supongamos ahora que tomamos una muestra aleatoria tamaño $n = 200$ de esa población, de las cuales 123 usan computadora. En este caso, la proporción de personas que usan computadora en la muestra, que la denotamos con $\hat{p}$, es $\hat{p} = 123/200 = 0,615$. Este número es el valor que asume el estadístico muestral en este caso particular y depende de la muestra elegida.

Otros parámetros poblacionales y sus compañeros estadísticos muestrales usuales son, por ejemplo, la media poblacional y la media muestral (promedio), la varianza poblacional y muestral y la mediana poblacional y muestral.

Los estadísticos muestrales se usan como *estimadores puntuales* de los parámetros poblacionales. Esto es, aproximaciones a los parámetros desconocidos, basados en los datos de una muestra. Si tomamos dos muestras distintas, los valores que asuma el estadístico muestral serán distintos. Es decir, los estadísticos muestrales tienen *variabilidad* y por lo tanto tienen una distribución, que se llama *distribución muestral*.

La distribución muestral es la distribución del estadístico muestral calculado para muchas muestras del mismo tamaño de la misma población. Este es el concepto más importante para el tema de este artículo, los contrastes de hipótesis.

Veamos un ejemplo. Consideremos las Universidades Nacionales Argentinas. Del INDEC podemos obtener los datos de cantidad de egresados por Universidad en 2015. La tabla se puede bajar de la página del INDEC. Hay 47 Universidades que tuvieron egresado/as ese año. Según los datos, la media de egresado/as por universidad en 2015 es de $2097{,}06$. Como estos datos corresponden al total de la población de Universidades Nacionales, tenemos que el parámetro poblacional μ es conocido y $\mu = 2097{,}06$. Consideremos varias muestras tamaño 10 de estas universidades. Para cada muestra obtendremos una media (muestral) distinta y distinta de μ. En la Figura 1 a) podemos observar la distribución muestral del estadístico *media muestral*, al que podemos llamar $\overline{x}$, en este caso. En la Figura 1 b) y c) observamos la distribución muestral de $\overline{x}$ pero para otros tamaños de muestra, $n = 20$ y $n = 3$. Se puede observar el efecto del tamaño de la muestra (a mayor tamaño menor variabilidad) y que las distribuciones están centradas alrededor del parámetro poblacional.

En general, para la mayoría de los estimadores que utilicemos, si las muestras están elegidas aleatoriamente y n es suficientemente grande, la distribución muestral estará centrada en el valor del parámetro poblacional y será simétrica con forma de campana. La distribución muestral expone cuánto puede variar el estadístico de muestra a muestra, es decir, cuál es su variabilidad.

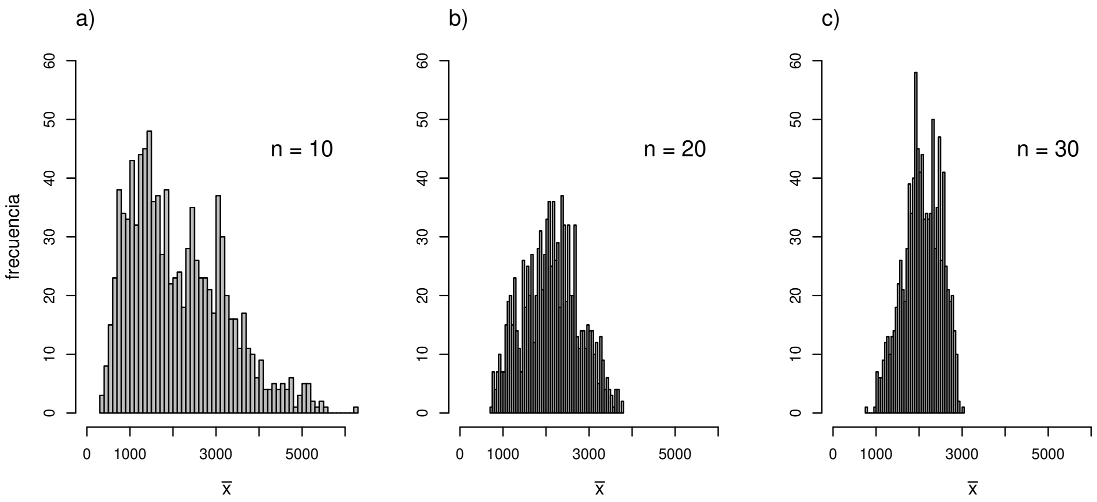

Figura 1. Distribución muestral del estadístico media de egresados de Universidades Nacionales en 2015 para muestras tamaño 10 (a), 20 (b) y 30 (c). Cada distribución corresponde a 1000 muestras aleatorias de la población real de 47 Universidades.

Contrastes de Hipótesis

La señora que toma té con leche. Volvamos al ejemplo de la señora que toma té. Llamemos π la proporción de veces que la señora acertaría si fuésemos capaces de darle infinitas tazas de té con leche, esto es, la probabilidad de acierto. Esa es la cantidad o parámetro poblacional que nos interesa estimar. Nuestra conjetura es que ese número es grande y que, por lo tanto, la señora no nos miente. Si la señora no sabe distinguir, y nos está engañando, entonces suponemos que ella elige en cada caso una de las opciones al azar, con probabilidad $0{,}5$. Es decir, si nuestra conjetura es falsa, entonces $\pi = 0{,}5$. Como no podemos darle infintas tazas, estimamos π a partir de una muestra que, en este caso, es de tamaño 8. Supongamos que, de las 8 tazas que le ofrecemos, la señora distingue correctamente en 7 de ellas el orden de los ingredientes. Luego, podemos estimar π con el estadístico muestral $\hat{p} = 7/8 = 0{,}875$.

En este contexto pueden haber pasado dos cosas: la señora realmente sabe distinguir o está adivinando pero tuvo mucha suerte. La explicación más razonable parece ser que la señora sí sabe distinguir. Vamos a proponer una prueba para justificar esta explicación. Pero, antes de seguir, pongamos un poco en contexto estadístico al ejemplo. Tenemos una conjetura acerca de un parámetro poblacional (π) y tenemos una cantidad muestral ($\hat{p}$) que utilizaremos para determinar si nuestra conjetura es verdadera o no. Esta es la escencia de los *contrastes de hipótesis*: usar la información de una muestra para evaluar una afirmación sobre una población.

Un contraste de hipótesis consta de dos afirmaciones contradictorias, las *hipótesis*, a las que llamamos *hipótesis nula*, denotada por H_0, e *hipótesis alternativa*, denotada por H_a. La afirmación que representa nuestra conjetura se asigna a la hipótesis alternativa (en el ejemplo: la señora sabe realmente distinguir). Generalmente, la hipótesis nula es la afirmación de que no hay efecto (la señora elige al azar). Para que las hipótesis no sean ambiguas, se escriben en función de un parámetro poblacional. En nuestro ejemplo podemos escribir:

- $H_0: \ \pi = 0{,}5$ (la señora está eligiendo al azar).
- $H_a: \ \pi > 0{,}5$ (la señora sabe distinguir).

La pregunta aquí es si $\hat{p} = 0{,}875$ es evidencia suficiente para poder descartar la hipótesis nula. Esto es, utilizaremos $\hat{p}$ para desarrollar nuestra prueba. El estadístico muestral que se utiliza en el contraste de hipótesis se llama *estadístico de prueba*.

En los contrastes de hipótesis la decisión de descartar la hipótesis nula se toma en función de la distribución del estadístico de prueba bajo la suposición de que la hipótesis nula es verdadera. Razonemos esto en el contexto del ejemplo. Si la señora estuviera eligiendo al azar con probabilidad $0{,}5$, el proceso de adivinar en cada caso es comparable al de tirar una moneda al aire y elegir leche-té si sale cara y té-leche si sale seca. Podemos usar este hecho para *simular* el ecenario en el que la señora simplemente adivida al azar (con probabilidad $0{,}5$) y así ver si el resultado del experimento real es un resultado esperable en ese ecenario. Al tirar una moneda estamos imitando la elección (al azar) de la señora entre leche-té y té-leche. Luego, para imitar el experimento original, debemos tirar 8 veces la moneda y anotar la proporción de veces que salió cara (o seca). Supongamos que tiramos la moneda 8 veces y obtenemos 5 caras. Luego, el estadístico simulado en este caso es $5/8 = 0{,}625$. Si repetimos la simulación varias veces obtendremos distintos resultados, pues este es un proceso aleatorio.

De esta manera, podemos obtener la distribución muestral del estadístico $\hat{p}$ bajo la suposición de que $\pi = 0{,}5$, esto es, la distribución muestral de $\hat{p}$ asumiendo la hipótesis de que la señora elige al azar (¡y por lo tanto nos miente!). Esa distribución está centrada en el parámetro poblacional, que en este caso es $0{,}5$ y nos permite contestar preguntas como ¿qué proporciones de cara son las más probables de obtener? o ¿cuánta variabilidad tiene el estadístico muestral simulado?

En la Figura 2 se muestra el resultado de repetir el experimento simulado 1000 veces. Podemos ver que los resultados que más ocurren son 3, 4 y 5. El 2 y el 6 también ocurren bastantes veces, el 0 y el 8 ocurrieron 4 veces y el 7 treinta y cuatro veces. Podríamos considerar cualquier resultado entre 2 y 6 como típico, pero sacar menos de 2 o más de 6 lo podemos considerar inusual.

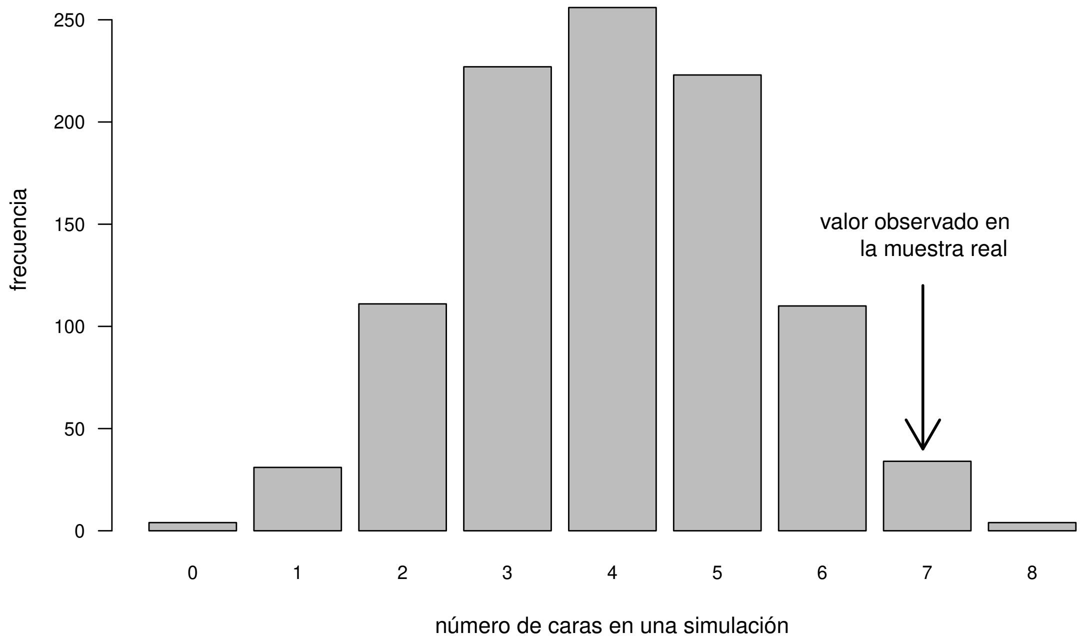

FIGURA 2. Simulaciones del experimento de arrojar una moneda 8 veces y anotar el número de caras. 1000 repeticiones.

En nuestro ejemplo, la señora adivinó 7 de las 8 veces, ese es un resultado inusual bajo la hipótesis nula. Entonces, el modelo de la moneda nos dice que hay una fuerte evidencia de que la señora no nos está mintiendo y efectivamente, puede distinguir el orden de los ingredientes. En estos casos, cuando un resultado como el estadístico calculado de la muestra real es muy poco probable de obtener por mero azar cuando se asume que la hipótesis nula es verdadera, decimos que el resultado es *estadísticamente significativo*. Si el resultado es estadísticamente significativo, tenemos evidencia suficiente en contra de H_0 y a favor de H_a.

Un concepto muy importante en estadística es el del *p-valor*. El p-valor es la probabilidad de obtener un valor tan o más extremo que el observado (en la dirección de la hipótesis alternativa), cuando la hipótesis nula es verdadera. Este valor es una herramienta que nos permite medir la fuerza de la evidencia en contra de la hipótesis nula. Mientras más chico es el p-valor, más improbable es observar un valor como el del estadístico muestral, bajo la hipótesis nula, por mero chance. Es decir, mientras más chico el p-valor, más estadísticamente significativa es la evidencia en contra de la hipótesis nula. Una forma de estimar el p-valor es calculando la proporción de valores de la distribución simulada que son tan o más inusuales que el estadístico muestral observado.

En nuestro ejemplo salieron 38 valores tan o más extremos que el muestral (7's y 8's). Luego, el p-valor es $38/1000 = 0{,}038$. Una pequeña discusión sobre cuán pequeño debe ser un p-valor para considerar la evidencia significativa se puede encontrar un poco más adelante.

Messi y los penales. Hasta el día en que escribí este artículo Lionel Messi había pateado 111 penales de los cuales falló 26. Asumiendo que un jugador promedio acierta el $75\,\%$ de sus penales, ¿es la evidencia suficiente para decir que Messi es un pateador de penales por encima de la norma? Si consideramos π la probabilidad de que Messi falle un penal, podemos plantear las hipótesis de la siguiente manera:

- $H_0: \ \pi = 0{,}25.$
- $H_a: \ \pi < 0{,}25.$

La hipótesis nula es que Messi tiene probabilidad de $0{,}25$ de fallar el penal y la alternativa es que él tiene probabilidad menor a $0{,}25$ de fallarlo. Ahora el experimento de la moneda ya no nos sirve como simulación para imitar la distribución muestral del estadístico pues debemos imitar el hecho de que la probabilidad de fallar es $0{,}25$, no $0{,}5$. Sin embargo, es fácil diseñar una simulación para este caso. Por ejemplo, podemos tirar un dado de 4 caras honesto donde un resultado (digamos el número 1) simule el fallo del penal y los otros 3 simulen el acierto. Así, el fallo tiene probabilidad $1/4 = 0{,}25$ y el acierto $3/4 = 0{,}75$, (otra posibilidad es poner 4 bolillas en un bolillero, una de un color, representando el fallo y las otras de otro color, representando el acierto). Los dados de 4 caras existen y son muy utilizados, entre otros, en los juegos de rol, Figura 3.

FIGURA 3. Dados de distinto número de caras, usualmente usados en juegos de rol. El azul es de cuatro caras, el rojo de 10. Fotos sacadas de (Wikipedia, 2019)

Cada simulación consistirá en 111 lanzamientos del dado y la cantidad de resultados iguales a 1 será la cantidad de fallos asumiendo que la hipótesis nula es verdadera. El estadístico muestral, en este caso, es $\hat{p} = \frac{26}{111} = 0{,}234$. En la Figura 4 observamos la distribución muestral del estadístico $\hat{p}$, proporción penales fallados por Messi, bajo la suposición de que tiene el triple de probabilidad de acertar que de fallar.

Se puede ver que el valor del estadístico observado de la muestra es un valor bastante frecuente en esta simulación, lo que indice que no hay evidencia de que Messi sea mejor pateador de penales que la media de los jugadores. Para calcular el p-valor, tenemos que contar cuantos valores tan o más extremos que el obtenido con la muestra real se observan en la distribución muestral, en la dirección de la hipótesis alternativa. De hecho hay 399 valores *menores o iguales* al 26. Luego, el p-valor estimado es $399/1000 = 0{,}399$, que es un número grande.

Elige un dígito. Consideremos como tercer ejemplo, un juego que se puede implementar en clase. Supongamos que pedimos a cada persona de un grupo de 20 que escriba un número del 0 al 9 en un papel. Si cada persona elige al azar el número, esperaríamos observar una cantidad muy similar (alrededor de dos) de cada uno de los dígitos. Supongamos que al hacer el recuento observamos que hay seis 3's. ¿Es esto evidencia de que las personas suelen elegir el número 3 más de lo esperado? Si las personas eligen realmente al azar esperaríamos que hayan una proporción de $0{,}1$ de 3's en la muestra (ya que hay 10 dígitos). Si queremos realizar un contraste de hipótesis en este caso podemos plantear las hipótesis en términos de la probabilidad π de que una persona cualquiera elija el 3:

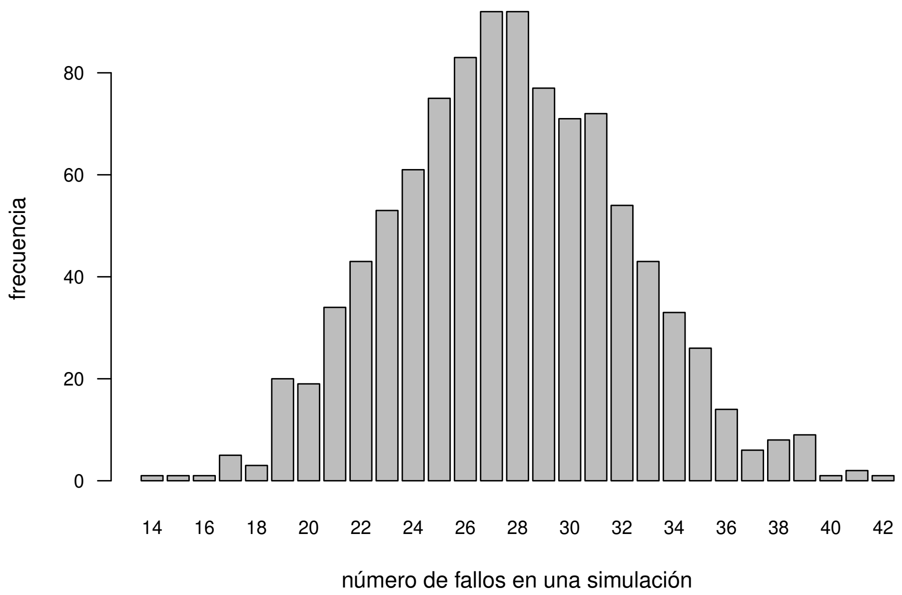

FIGURA 4. Simulaciones del experimento de arrojar un dado de cuatro caras y anotar el número de 1's. 1000 repeticiones.

- $H_0: \pi = 0{,}1$
- $H_a: \pi \neq 0{,}1$.

El estadístico que consideramos en este caso para realizar el contraste es la proporción de 3's observados en la muestra, $\hat{p} = 6/20 = 0{,}3$. Para imitar la distribución muestral del estadístico debemos imitar el hecho de que la probabilidad de obtener un 3 sea $0{,}1$. Podemos realizar la simulación con una técnica parecida a la del ejercicio anterior: tirar un dado de 10 caras 20 veces donde cada resultado representa la elección de una persona y contar la cantidad de 3's. Este experimento se debe repetir muchas veces para obtener la distribución del estadístico muestral, bajo H_0. La Figura 5 muestra el resultado (realizado por computadora) de repetir 1000 veces el experimento de elegir 20 dígitos al azar, cada uno con probabilidad $0{,}1$. En este caso mostramos la proporción (en vez del conteo) de 3's en cada muestra. Podemos observar que en este caso, la distribución muestral del estadístico no es simétrica. Esto ocurre porque la variable que consideramos es una proporción y, por lo tanto, no puede tomar valores negativos.

En esta simulación, el $0{,}3$ (esto es, que el 3 fuera elegido 6 veces) ocurrió ocho veces, el $0{,}35$ (que corresponde a 7 veces) tres veces y el $0{,}4$ (8 veces) una sola vez. Esto parece indicar que hay evidencia de que las personas eligen el 3 con más

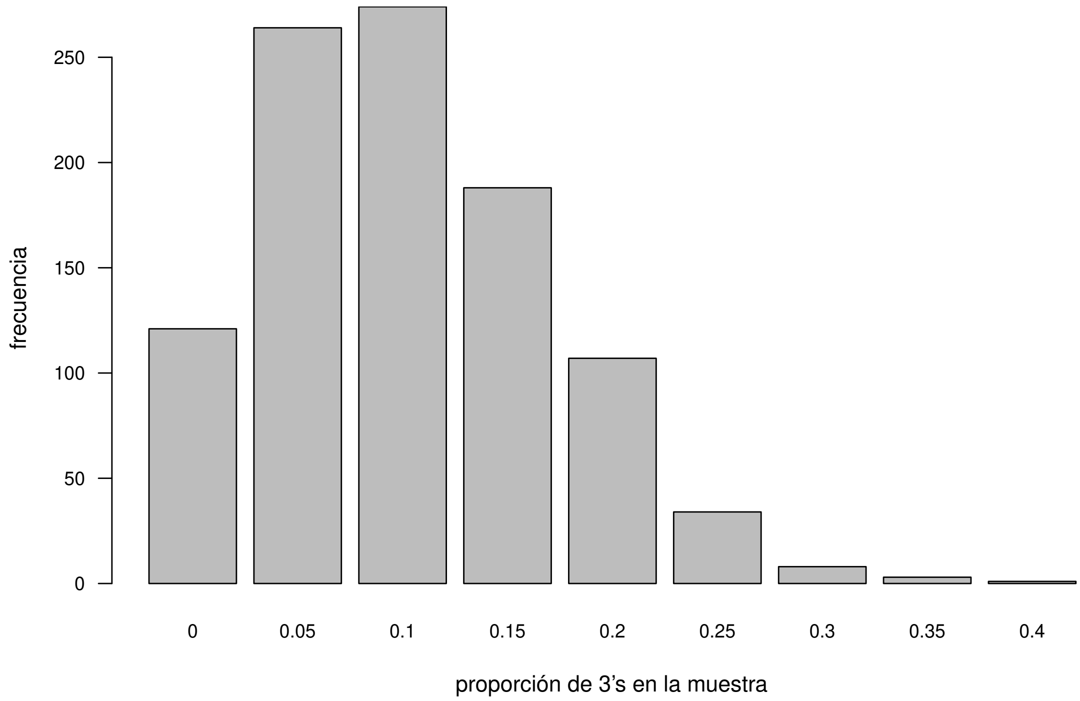

FIGURA 5. Simulaciones del experimento de arrojar un dado de 10 caras 20 veces y anotar la proporción de 3's. 1000 repeticiones.

frecuencia que la esperada por mero azar. Para calcular el p-valor en este caso tenemos que tener en cuenta que la hipótesis nula es $\pi \neq 0{,}1$, por lo tanto debemos considerar valores tan o más extremos que el observado en la muestra para las dos colas de la distribución. En este caso, al no poder observar valores tan extremos como el $0{,}3$ en la cola izquierda, el p-valor lo calculamos como 2 veces la cantidad de valores tan o más extremos que el 6: p-valor $= 2 \times 12/1000 = 0{,}024$. Este número es pequeño, lo que confirma nuestra conjetura, la evidencia es estadísticamente significativa.

Índice de masa corporal y sobrepeso. En este ejemplo, consideramos una variable continua. El índice de masa corporal (IMC) es un indicador que se utiliza frecuentemente para identificar el sobrepeso y la obesidad en los adultos. Este índice es una relación entre el peso y la talla (altura) que se calcula dividiendo el peso de una persona en kilos por el cuadrado de su talla en metros (kg/m^2). Se considera que una persona tiene sobrepeso si el IMC es igual o mayor a 25 y un IMC de más de 30 indica obesidad. Para este ejemplo, consideremos los datos de la Encuesta Nacional de Factores de Riesgo (ENFR). La ENFR se lleva a cabo en Argentina como estrategia nacional de prevención y control de enfermedades no transmisibles. La encuesta se realiza a una muestra de personas de 18 años o más que viven en conglomerados de 5000 o más habitantes. Consideremos para este ejemplo la población de adultos hombres de más de 70 años. En la ENFR hay 1109 hombres de más de 70 años y el IMC medio de esa muestra es $\overline{x} = 26{,}96$. ¿Significa esto que la población de hombres mayores de 70 años en Argentina (que viven en conglomerados de 5000 o más habitantes) sufre de sobrepeso o será un efecto de la aleatoriedad de la muestra? La hipótesis que representa la conjetura de que hay sobrepeso, en términos de la media poblacional μ, es que ésta es mayor a 25. Y queremos ver si hay evidencia de este hecho en contra de que la media es igual a 25. Es decir:

- $H_0 : \mu = 25$
- $H_a : \mu > 25$

Para evaluar la significación de la evidencia en esta muestra, necesitamos aproximar la distribución muestral del estadístico $\overline{x}$ bajo la hipótesis nula. La estrategia de la moneda o el dado ya no nos sirve en este contexto dado que lo que necesitamos son muestras que reflejen la estructura de los datos originales, que son observaciones de una variable aleatoria continua (IMC). Para este tipo de problemas se utilizan métodos de aleatorización o remuestreo, en los que se utilizan los datos originales para generar muestras artificiales. De esta manera, nos evitamos tener que conocer (o aproximar) el proceso que generó los datos para poder imirtalos. Por otro lado, para poder sacar conclusiones válidas, los datos artificiales deben ser consistentes con la hipótesis nula. En este ejemplo, esto significa que debemos generar muestras que sean consistentes con la hipótesis de que la media es $25 kg/m^2$. Para lograr esto, sumaremos a cada dato de la muestra original una constante, de manera tal que los datos trasladados tengan media 25. Luego, tomaremos muestras aleatorias *con* reemplazo tamaño $n = 1109$ de la muestra original trasladada.

En la Figura 6 a) podemos observar los datos correspondientes a los 1109 hombres mayores de 70 años de todo el territorio Nacional, que participaron en la ENFR en el diagrama de cajas y bigotes de la izquierda y una muestra aleatorizadas en el diagrmaa de la derecha. En el panel b) observamos un histograma de las medias de los IMC calculados para 1000 muestras aleatorizadas, esto es la distribución del IMC medio bajo la hipótesis nula. Ya en este histograma podemos observar que el valor de IMC medio de la muestra original está lejos de los valores que uno esperaría tener si la media de la población fuera 25 y esto implica que hay evidencias suficientes para decir que la población de hombres mayores de 70 años argentina sufre de sobrepeso. De todas maneras, si queremos calcular el p-valor en este caso, lo hacemos de la misma forma que en los ejemplos anteriores, contamos cuantos resultados son igual o mayores a $26{,}96$. Como no hay valores tan extremos entre los resultados de la simulaciones, p-valor$= 0$

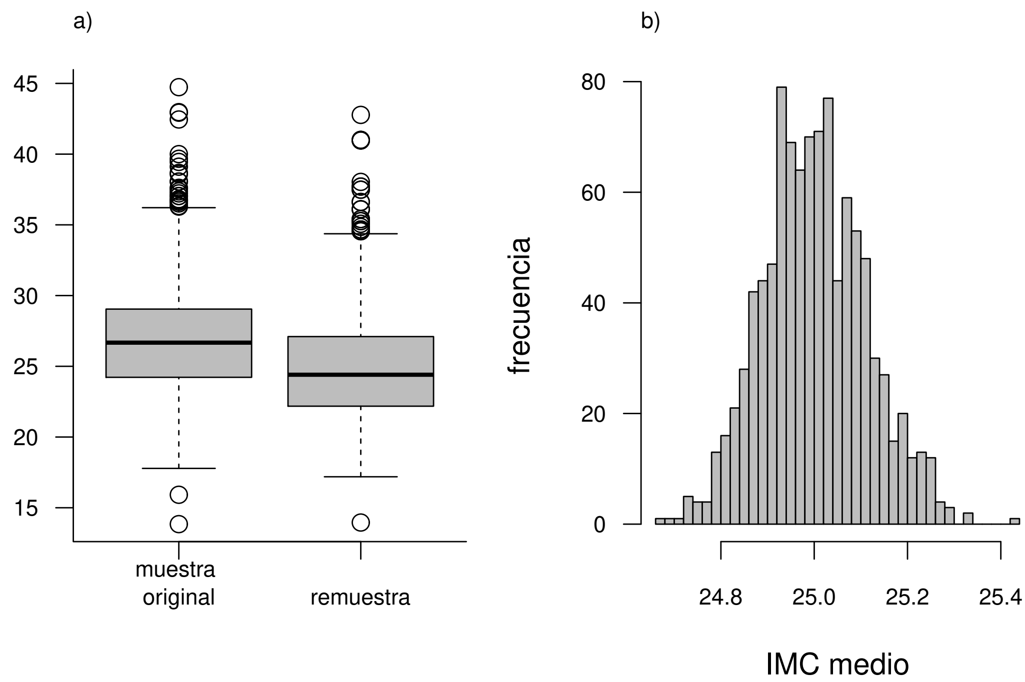

FIGURA 6. a) Diagramas de caja y bigotes de la muestra original de IMC de adultos hombres mayores de 70 años (izquierda) y de una muestra aleatorizada del mismo tamaño de la original (derecha). b) Histograma de las medias del IMC de 1000 muestras aleatorizdas.

Los pollos. Veamos un último ejemplo con otra variable cuantitativa continua. Vamos a considerar un experimento en el que unos pollos recién nacidos fueron asignados al azar en seis grupos y cada grupo recibió un suplemento alimenticio diferente. Después de seis semanas se registró el peso (en gramos) de cada pollo (Anónimo, 1948). Estos datos (como muchos otros) están en internet libres para usar.

Para este ejemplo, consideraremos solo los pollos alimentados a base de harinas cárnicas (11 pollos) y con linaza (12 pollos). En la Figura 7 a) se muestran los pesos de los pollos de estos dos grupos junto con sus medias muestrales: $\overline{x}_C$ para el grupo alimentado con harinas cárnicas y $\overline{x}_L$ para el grupo alimentado con linaza. Se puede observar que $\bar{x}_C > \bar{x}_L$ y que casi todos (9 de 11) de los pollos alimentados con harinas cárnicas son más pesados que la media de los alimentados con linaza. ¿Será esto evidencia de que pollos alimentados con la dieta a base de harinas cárnicas aumentaron significativamente más de peso que aquellos alimentados con linaza? ¿O será este resultado mera consecuencia del azar?

Si llamamos μ_L a la media poblacional de los pesos de los pollos alimentados con linaza y μ_C a la de los alimentados con harinas cárnicas, podemos plantear las hipótesis de interés como

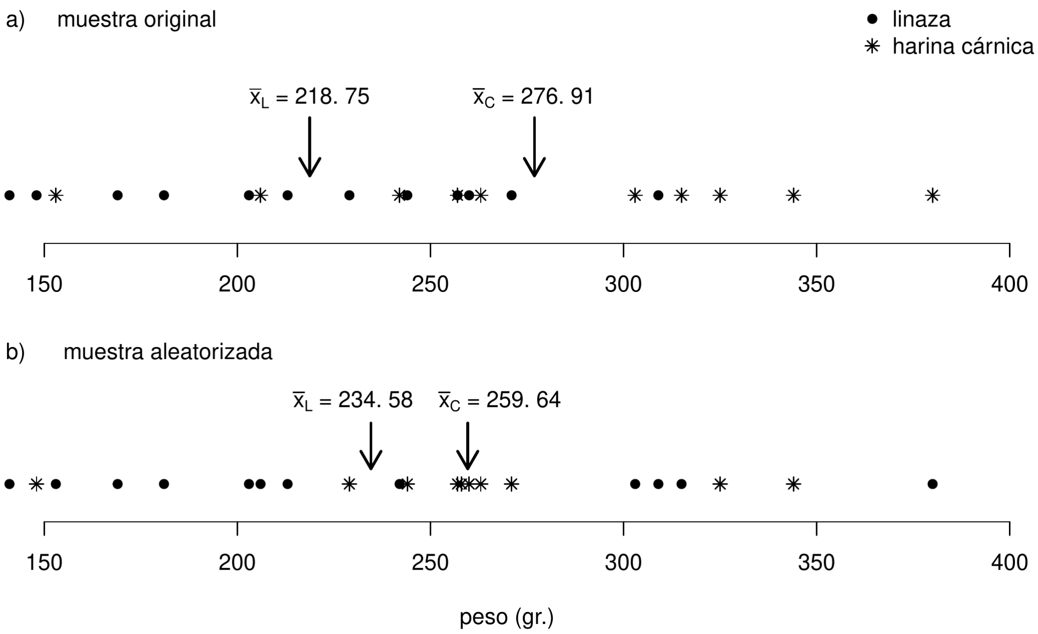

FIGURA 7. Peso de los pollos alimentados con linaza y harinas cárnicas. a) Muestra original. b) Muestra aleatorizada.

- $H_0 : \mu_C = \mu_L$
- $H_a : \mu_C > \mu_L$

La hipótesis nula es equivalente a decir que la diferencia de las medias poblacionales es cero. El estadístico muestral que consideraremos en este caso es la diferencia en medias muestrales. Para estos datos, la diferencia observada es

$$D = 276{,}91 - 218{,}75 = 58{,}16$$

Luego, para determinar si la diferencia observada es significativamente diferente de cero, debemos ver si $58{,}16$ es un valor frecuente en la distribución de la diferencia de medias asumiendo que la hipótesis nula es verdadera. La estrategia utilizada en el ejemplo anterior no nos sirve aquí dado que lo que tenemos que reflejar en este caso es que los valores observados bajo los grupos *harinas cárnicas* y *linaza* en realidad vienen de distribuciones con igual media.

Si H_0 es verdadera el peso de un pollo sería el mismo habiendo recibido cualquiera de los dos alimentos. Dicho de otro modo, cualquiera de los valores observados en el grupo linaza podría haber sido observado en el grupo harinas cárnicas y viceversa, si se hubieran asignado los pollos de otra forma. Esto nos da una idea de cómo generar las muestras aleatorizadas en el escenario en que las asignaciones a los grupos no influyen en el peso de los pollos. Podemos separar aleatoriamente los 23 valores observados en dos grupos: harinas cárnicas y linaza (respetando los tamaños muestrales) y calcular D con la nueva asignación. Para simular este proceso podríamos, por ejemplo, tomar 23 pedazos de papel, en cada uno escribir un valor de la muestra, mezclarlos y luego armar dos pilas, una de once y otra de doce, donde cada pila representa cada grupo. Por ejemplo, en la Figura 7 b), se muestra una de estas asignaciones aleatorias. En esta muestra artificial, $D = \bar{x}_C - \bar{x}_L = 259{,}64 - 234{,}58 = 25{,}06$.

Repitiendo muchas veces el proceso de mezclar los papelitos y separarlos en dos pilas, simulamos el proceso de separar aleatoriamente los pollos en dos grupos, asumiendo que la alimentación no influye en su crecimiento. Así, aproximamos la distribución de la diferencia de medias bajo H_0 y podemos ver dónde cae el estadístico observado $D = 58{,}16$.

La Figura 8 muestra la diferencia de medias en 1000 de estas muestras aleatorizadas. Observar que la distribución está centrada en cero, como es de esperar dado que la hipótesis nula es $\mu_C - \mu_L = 0$. En rojo se muestran las medias mayores a $58{,}16$, que son 16. Luego, el p-valor es $16/1000 = 0{,}016$ y podemos decir que hay evidencia estadísticamente significativa de que la alimentación a base de harinas cárnicas hace crecer más a los pollos que la alimentación a base de linaza.

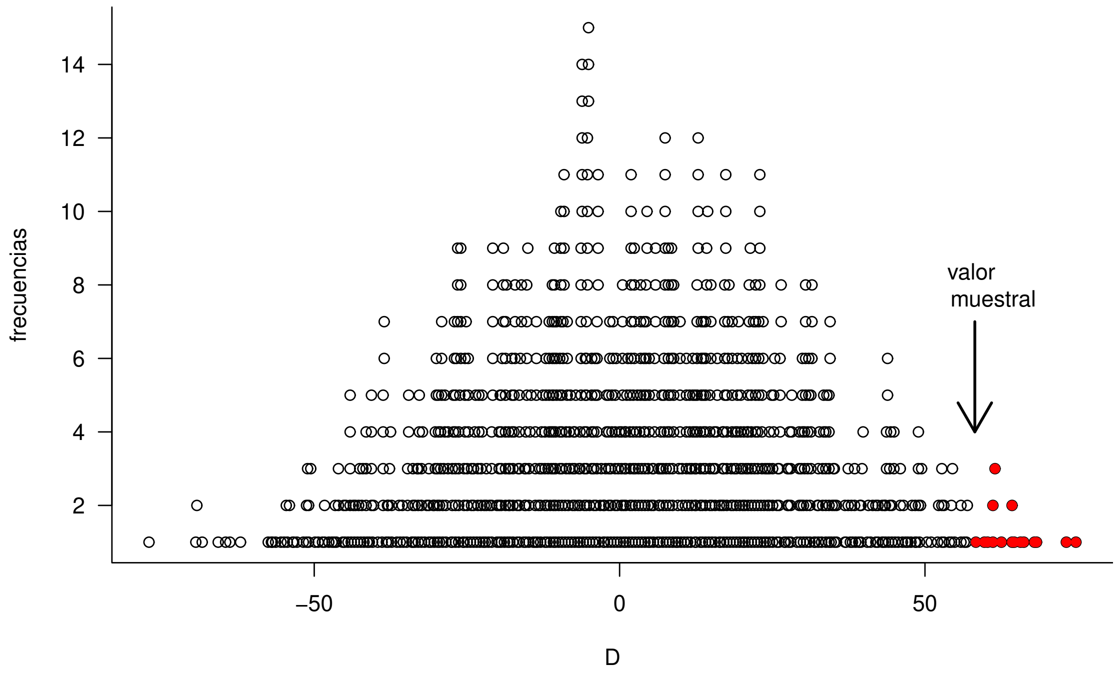

FIGURA 8. Diferencias en media para 1000 muestras aleatorizadas.

Últimas consideraciones. Hemos hablado de significación estadística y p-valor pero sin embargo, no hemos dado una regla para determinar cuán chico debe ser un p-valor para poder considerar la evidencia estadísticamente significativa. El punto de corte entre rechazar y no rechazar una hipótesis nula se llama *nivel de* *significación* y generalmente se denota con la letra griega α. Por ejemplo, si $\alpha = 0{,}05$, decimos que estamos llevando a cabo un contraste de hipótesis de nivel $0{,}05$, rechazaremos la hipótesis nula si el p-valor resulta menor a $0{,}05$ y diremos que se rechaza la hipótesis nula a favor de la alternativa a nivel $0{,}05$ o $5\,\%$. Aunque el nivel de significación hay que elegirlo para cada problema particular, hay bastante consenso en las siguientes conclusiones. Si p-valor $> 0{,}1$ entonces no hay mucha evidencia en contra de H_0. Si $0{,}05 < p$-valor $\leq 0{,}1$ entonces hay alguna evidencia en contra de H_0. Si $0{,}001 < p$-valor $\leq 0{,}05$ entonces hay fuerte evidencia en contra de H_0 y si p-valor $\leq 0{,}01$ entonces hay evidencia muy fuerte en contra de H_0.

El nivel de significación tiene que ver con los errores que se pueden cometer al tomar una decisión. Esto es, podemos rechazar la hipótesis nula cuando esta es en realidad verdadera (error tipo I) o, podemos no rechazarla cuando es falsa (error tipo II). El nivel de significación, α, representa la máxima probabilidad tolerable de cometer el error tipo I. Por esto, la elección del nivel de significación debe hacerse para cada problema en particular. Pero esto se escapa al objetivo de este artículo.

Las simulaciones

Si bien las simulaciones pueden muchas veces pensarse como procesos físicos (como tirar una moneda) para que sean más fácilmente interpretables, luego, éstas se realizan utilizando una computadora. Aunque no entra en los objetivos de este artículo el mostrar cómo realizar estas simulaciones, podemos nombrar algunas opciones. Se pueden usar softwar estadísticos libres y gratuitos como

- R
- GeoGebra
- Infostat (versión estudiantil)

Otras aplicaciones online, por nombrar algunas (éstas en inglés):

- StatKey - del libro “Unlocking the power of data"(Lock y cols., 2013)
- Apps del libro "Introduction to Statistical Investigations" (Tintle y cols., 2015)

Conclusiones

La capacidad de desarrollar herramientas para obtener información de un conjunto de datos nos habilita a tomar decisiones en nuestra vida personal y profesional, cualquiera sea el área en la que nos desarrollemos. Por esto, la incorporación de la estadística en los planes de estudio de todos los niveles y disciplinas se vuelve cada día más necesaria. Para enseñar estadística en áreas alejadas de la matemática es necesario implementar métodos que muestren la potencia y utilidad de los métodos estadísticos tempranamente y de forma atractiva. Los métodos basados en simulaciones permiten incorporar los conceptos de estadística, que muchas veces son difíciles y abstractos, de manera intuitiva, utilizando ejemplos y datos reales desde el primer día, sin atascarnos en las formalidades matemáticas. En este artículo se mostró este hecho con uno de los temas centrales de la inferencia estadística, los contrastes de hipótesis.

Referencias

Anónimo. (1948). –. *Biometrics*, 214. doi: 10.2307/3001566

Fisher, R. A. (1971). *The design of experiments (9th ed.)*. Macmillan. ISBN 0-02-844690-9.

Lock, R., Lock, P., Morgan, K., Lock, E., y Lock, D. (2013). *Statistics: Unlocking the power of data*. John Wiley & Sons, Inc.

Team, R. C. (2018). *A language and environment for statistical computing. r foundation for statistical computing*. Vienna, Austria. Descargado de www.R-project.org

Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., y VanderStoep, J. (2015). *Introduction to statistical investigations: High school binding*. John Wiley & Sons, Inc.

Wikipedia. (2019). *Dados de rol. (2019, 3 de marzo)*. Descargado de https://es.wikipedia.org/w/index.php?title=Dados_de_rol&oldid=114332241. ([Fecha de consulta: 19:00, marzo 25, 2019])

Aldana M. González Montoro
Universidad Nacional de Córdoba .
FAMAF, Av. Medina Allende s/n, Ciudad Universitaria, CP:X5000HUA Córdoba, Argentina.
(✉) aldana.gonzalez.montor@unc.edu.ar

Recibido: *28 de febrero de 2019*.
Aceptado: *24 de abril de 2019*.
Publicado en línea: *7 de mayo de 2019*.

33 es suma de 3 cubos perfectos?

Existen muchos resultados sobre suma de cuadrados. Por ejemplo, todos estamos familiarizados desde el colegio con la ecuación

$$x^2 + y^2 = z^2.$$

Sabemos por el Teorema de Pitágoras, que si x, y, z son las longitudes de los lados de un triángulo rectángulo, entonces x, y, z la satisfacen. Hay soluciones enteras como $(3, 4, 5)$ ó $(5, 12, 13)$. Mas aún, existen infinitas soluciones enteras, que son las conocidas ternas pitagóricas.

Un famoso teorema de Fermat afirma que todo número primo impar p es suma de dos enteros al cuadrado si y sólo si éste tiene resto 1 al dividirlo por 4. En símbolos,

$$p = x^2 + y^2 \quad \text{tiene solución} \qquad \Leftrightarrow \qquad p \equiv 1 \pmod{4}.$$

Más generalmente, un número n (no necesariamente primo) es suma de dos enteros al cuadrado $n = x^2 + y^2$ si y sólo si la descomposición prima de n no contiene primos congruentes a 3 mod 4 elevados a potencias impares.

El teorema de los 3 cuadrados de Legendre afirma que

$$n = x^2 + y^2 + z^2$$

tiene solución en enteros si y sólo si n no es de la forma $n = 4^a(8b+7)$ para enteros a, b. Por otra parte, tenemos el teorema de los 4 cuadrados de Lagrange que afirma que todo número natural puede ser escrito como la suma de 4 enteros al cuadrado. Es decir, dado $n \in \mathbb{N}$, la ecuación

$$n = x^2 + y^2 + z^2 + w^2$$

siempre tiene solución entera.

Para potencias mayores que dos, tenemos el conocido como el último Teorema de Fermat, conjeturado por éste y probado por Andrew Wiles en 1995, que dice que la ecuación

$$x^n + y^n = z^n$$

no tiene soluciones enteras no triviales (o sea sin usar 0 y ± 1) para $n \geq 3$.

Una pregunta natural y aparentemente inocente, que surge en el mismo espíritu que los comentarios previos, es la siguiente: *¿Se puede escribir cualquier número natural como suma de 3 cubos?* Es decir, dado $a \in \mathbb{N}$, ¿tiene la ecuación

$$a = x^3 + y^3 + z^3 \tag{1}$$

solución en enteros x, y, z? Ejemplos sencillos con solución son los tres primeros naturales: $1 = 1^3 + 0^3 + 0^3$, $2 = 1^3 + 1^3 + 0^3$ y $3 = 1^3 + 1^3 + 1^3$. ¿Qué pasa con 4 y 5?

La pregunta general está lejos de ser respondida. Por un lado, se sabe que algunos números no pueden ser representados como suma de 3 cubos. Los números de la forma $9k \pm 4$ (o sea los que dan resto 4 ó 5 al dividir por 9) no pueden escribirse como suma de 3 cubos. En efecto, es fácil chequear que los cubos módulo 9 son 0, 1 ó -1. En símbolos

$$x^3 \equiv 0, 1, -1 \pmod 9$$

para todo $x \in \mathbb{Z}$. Basta chequear que

$$0^3, 3^3, 6^3 \equiv 0 \pmod 9, \quad 1^3, 4^3, 7^3 \equiv 1 \pmod 9, \quad 2^3, 5^3, 8^3 \equiv -1 \pmod 9.$$

Luego, como suma de 3 cubos podríamos obtener (o no) números de la forma 9, $9k \pm 1$, $9k \pm 2$ y $9k \pm 3$, pero nunca de la forma $9k \pm 4$.

Por otra parte, en algunos casos existen infinitas soluciones de (1) para el mismo natural a. Por ejemplo:

$$\begin{aligned} 1 &= (9t^3 + 1)^3 + (9t^4)^3 + (-9t^4 - 3t)^3, \\ 2 &= (6t^3 + 1)^3 + (1 - 6t^3)^3 + (-6t^2)^3, \end{aligned}$$

para cualquier $t \in \mathbb{N}$ (¡chequear!). Notar que para $t = 1$ en la primer ecuación tenemos $1 = 10^3 + 9^3 - 12^3$, de donde sale ¡el número de la muy conocida anécdota[1] del taxi entre Hardy y Ramanujan!

La lista con los primeros números naturales como suma de 3 cubos es la siguiente:

$$\begin{array}{ll}
6 = 2^3 + (-1)^3 + (-1)^3, & 7 = 2^3 + (-1)^3 + 0^3, \\
8 = 2^3 + 0^3 + 0^3, & 9 = 2^3 + 1^3 + 0^3, \\
10 = 2^3 + 1^3 + 1^3, & 11 = 3^3 + (-2)^3 + (-2)^3, \\
12 = 10^3 + 7^3 + (-11)^3, & 15 = 2^3 + 2^3 + (-1)^3, \\
16 = 2^3 + 2^3 + 0^3, & 17 = 2^3 + 2^3 + 1^3, \\
18 = 3^3 + (-2)^3 + (-1)^3, & 19 = 3^3 + (-2)^3 + 0^3, \\
20 = 3^3 + (-2)^3 + 1^3, & 21 = 16^3 + (-14)^3 + (-11)^3, \\
24 = 2^3 + 2^3 + 2^3, & 25 = 3^3 + (-1)^3 + (-1)^3, \\
26 = 3^3 + (-1)^3 + 0^3, & 27 = 3^3 + 0^3 + 0^3, \\
28 = 3^3 + 1^3 + 0^3, & 29 = 3^3 + 1^3 + 1^3,
\end{array}$$

[1] Se dice que estando Srinivasa Ramanujan (el genio indio de la teoría de números) en el hospital, el gran matemático inglés G. H. Hardy fue a visitarlo y le comentó que había tomado un taxi cuya patente era 1729 y que le parecía que este número no tenía ninguna propiedad interesante. Ramanujan meditó un segundo y le respondió: "1729 es el menor entero que puede escribirse como suma de cubos de dos formas distintas". Esto dió lugar a los números llamados 'taxicab'. El número $Ta(n)$ es el menor número que puede escribirse como suma de dos cubos en n formas distintas. De este modo $T(1) = 2 = 1^3 + 1^3$, $T(2) = 1729 = 9^3 + 10^3 = 1^3 + 12^3$, $T(3) = 87.539.139 = 167^3 + 436^3 = 228^3 + 423^3 = 255^3 + 414^3$ y sólo se conocen $Ta(4)$, $Ta(5)$ y $Ta(6)$.

donde hemos descartado los números $4, 5, 13, 14, 22, 23, 31, 32$, pues ya sabemos que estos no pueden expresarse cono suma de cubos, por ser de la forma $9k+4$ ó $9k+5$. Notar que en la lista damos la menor solución posible, ya que por ejemplo también tenemos $3 = 4^3 + 4^3 + (-5)^3$.

Si bien todos estos números pueden ser hallados a mano con mayor o menor esfuerzo, notemos que también tenemos $26 = (114.844.365)^3 + (110.902.301)^3 + (-142.254.840)^3$; y que para el siguiente se complicaría un poco mas, dado que

$$30 = (-283.059.965)^3 + (-2.218.888.517)^3 + 2.220.422.932^3$$

donde

$$\begin{aligned} 283.059.965^3 &= 22.679.597.663.705.862.245.457.125, \\ 2.218.888.517^3 &= 10.924.622.727.902.378.924.946.084.413, \\ 2.220.422.932^3 &= 10.947.302.325.566.084.787.191.541.568, \end{aligned}$$

una expresión que sorprende y hasta asusta un poco. Peor aun, para el 33 tenemos la asombrosa expresión hallada muy recientemente

$$33 = (8.866.128.975.287.528)^3 + (-8.778.405.442.862.239)^3 + (-2.736.111.468.807.040)^3$$

donde

$$\begin{aligned} 8.866.128.975.287.528^3 &= 696.950.821.015.779.435.648.178.972.565.490.929.714.876.221.952, \\ 8.778.405.442.862.239^3 &= 676.467.453.392.982.277.424.361.019.810.585.360.331.722.557.919, \\ 2.736.111.468.807.040^3 &= 20.483.367.622.797.158.223.817.952.754.905.569.383.153.664.000. \end{aligned}$$

¡Sin palabras!

La lista sigue: $34 = 3^3 + 2^3 + (-1)^3$, $35 = 3^3 + 2^3 + 0^3$, $36 = 3^3 + 2^3 + 1^3$, $37 = 4^3 + (-3)^3 + 0^3$, $38 = 4^3 + (-3)^3 + 1^3$,

$$39 = 134.476^3 + 117.367^3 + (-159.380)^3.$$

Sabemos que 40 y 41 hay que descartarlos. Sin embargo el 42 no se sabe. Este es el menor número para el que no se sabe si puede escribirse como suma de 3 cubos. ¿Te animás a encontrarlo?

Este es un ejemplo más que muestra que en aritmética, enunciados tan simples que cualquiera puede entender pueden ser tremendamente difíciles de resolver, incluso para los mejores matemáticos y las mejores computadoras.

La representación del 39 fue hallada en 1992 por Heath-Brown, Lioen and te Riele. La representación del 30 fue encontrada por Beck, Pine, Tarrant and Yarbrough Jensen en 1999. El problema para el número 33 estuvo abierto por 64 años y fue resuelto en.. ¡marzo de este año! por un matemático de la Universidad de Bristol llamado Andrew Booker.