Trabajo Práctico 1: ¿La gloria es el dinero?

Introducción

La idea de convertir unos pocos pesos en una fortuna con un solo clic tiene algo de magia, de esa que seduce con la promesa de éxito instantáneo. Masivo. ¿Qué puede salir mal? Total... *"Está todo bajo control, yo decido cuándo frenar"*. El problema es que, a veces de manera silenciosa, el juego compulsivo puede desencadenar problemas de largo plazo en la salud mental y el funcionamiento social de los individuos.

Las apuestas en línea han ganado una creciente popularidad entre los adolescentes, impulsadas por la accesibilidad de plataformas digitales y la constante exposición a la publicidad en redes sociales y eventos deportivos. A esto se suma que se trata de una actividad escasamente regulada por el Estado y a la que cualquier persona puede acceder con relativa facilidad.

Estas apuestas pueden dividirse en dos grandes categorías: los casinos virtuales, que incluyen juegos como ruleta o blackjack, y las apuestas deportivas, donde los jóvenes apuestan sobre el resultado de eventos deportivos, a menudo influenciados por pronósticos y estrategias compartidas en comunidades en línea. Ambas modalidades presentan riesgos relevantes, ya que la facilidad para depositar dinero y la sensación de control pueden llevar a un espiral de pérdidas y nuevas apuestas en busca de recuperar lo perdido.

La sensación de ganar y perder plata genera adrenalina. La adolescencia, una etapa crucial de desarrollo de los seres humanos, es especialmente susceptible a la promesa del dinero fácil y a la gratificación instantánea que pueden otorgar los juegos de apuestas, debido a la inmadurez de las áreas cerebrales encargadas del control de impulsos, lo que incrementa la vulnerabilidad a comportamientos adictivos. A esto se suman factores sociales como la presión de grupo y la búsqueda de aceptación, donde ganar una apuesta se convierte en una forma de demostrar conocimiento y obtener reputación entre pares.

La ilusión de ganar dinero rápido, junto con la normalización de las apuestas a través de la publicidad, crea un cóctel peligroso que atrapa a los jóvenes en un ciclo de riesgo y compulsión. A esto se suma la aparición de gurúes financieros y esquemas piramidales que prometen riqueza rápida, donde quienes no participan son vistos como ingenuos o fracasados. Esta presión social y la promesa de ganancias fáciles refuerzan la participación en un entorno cada vez más riesgoso y manipulador.

Como si todo esto fuera poco, indagar en un grupo sobre temas sensibles puede resultar complejo. Muchas personas, por vergüenza, miedo al juicio social o simplemente por desconfianza, optan por no responder con sinceridad a ciertas preguntas, por ejemplo en el marco de una encuesta. En el caso de las apuestas, admitir la frecuencia con la que se participa, las cantidades apostadas o las pérdidas sufridas puede generar resistencia a contestar con la verdad. Algunas personas evitan responder, mientras que otras pueden proporcionar respuestas erróneas de manera deliberada. Este fenómeno, conocido como sesgo de respuesta, representa un desafío para quienes buscan obtener datos fiables en encuestas y estudios sobre este tema.

Partiendo de la premisa de que una forma de incrementar la cooperación de los encuestados es garantizar la protección de información sensible, una posible forma de mitigar este sesgo es la técnica de respuesta aleatorizada, que permite que los encuestados respondan de manera más sincera sin temor a ser identificados. En la técnica de respuesta aleatorizada, se introduce una cuota de azar con el objetivo de preservar la privacidad de la persona que responde.

Nos centraremos en el problema de querer realizar inferencias sobre π_A, la proporción de estudiantes de una escuela que participan de apuestas deportivas en línea.

Estudiaremos dos técnicas de respuesta aleatorizada. En lugar de de responder directamente a la pregunta *"¿Participaste alguna vez en apuestas deportivas en línea?"*, se le pide al encuestado que utilice un mecanismo aleatorio, como arrojar una moneda. Dependiendo del resultado de este mecanismo, el encuestado sigue ciertas instrucciones para responder. En las variantes que estudiaremos, por ejemplo, si sale cara, el estudiante contesta a la pregunta *"¿Participaste alguna vez en apuestas deportivas en línea?"* mientras que si sale cruz, responde a otra pregunta. Así, como el investigador no conoce el resultado del mecanismo aleatorio no puede saber a qué pregunta está respondiendo el encuestado. No obstante, a partir de los resultados obtenidos para toda la muestra, pueden obtenerse conclusiones válidas.

En el caso más general, diremos que p es la probabilidad (conocida), debida al mecanismo aleatorio, de que el encuestado responda a la pregunta *"¿Participaste alguna vez en apuestas deportivas en línea?"* mientras que $(1-p)$ es la probabilidad de que responda a la otra pregunta.

Técnicas de respuesta aleatorizada

Warner

Un mecanismo aleatorio selecciona la pregunta que se le presenta a la persona que responde. Con probabilidad p se le pregunta si es cierto que alguna vez participó en apuestas deportivas en línea (Q_1), mientras que con probabilidad $(1-p)$ se le pregunta si es cierto que nunca participó en apuestas deportivas en línea (Q_2). Es decir, en general, con probabilidad p se le pregunta si pertenece a una categoría A, mientras que con probabilidad $(1-p)$ se le pregunta si pertenece a la categoría A^c (el complemento de A).

Greenberg y otros

Un mecanismo aleatorio selecciona la pregunta que se le presenta a la persona que responde. Con probabilidad p se le pregunta si es cierto que alguna vez participó en apuestas deportivas en línea (Q_1), mientras que con probabilidad $(1-p)$ se le pregunta si nació en un mes de 31 días (Q_2). Es decir, en general, con probabilidad p se le pregunta si pertenece a una categoría A, mientras que con probabilidad $(1-p)$ se le pregunta si pertenece a una categoría B (no controversial o no polémica).

Existen dos versiones de esta modalidad, en relación al conocimiento o desconocimiento de π_B, la proporción de elementos en la población que pertenece a la categoría B. En el primer caso, se conoce de antemano π_B (como en el caso presentado, en que $\pi_B = \frac{7}{12}$). Mientras que en el segundo caso, no se conoce π_B. Nos limitaremos al estudio de la primera situación.

Actividades

Para todo lo que sigue, tenga en consideración lo siguiente:

- Se toma una muestra al azar con reemplazo de tamaño $n = 100$ de una población de tamaño $N = 1000$.
- Para nuestros estudios comparativos, supondremos que el porcentaje de alumnos que apuesta (lo que se quisiera estimar) es 40%.
- Cuando se les pregunta directamente si han hecho alguna vez apuestas, los estudiantes que sí han apostado alguna vez mienten con probabilidad μ.
- Si se utilizan técnicas de respuesta aleatorizada, los estudiantes no mienten.

Comenzaremos realizando un estudio de simulación para estudiar el efecto de la mentira en las estimaciones. En esta primera aproximación, consideraremos que se realiza la pregunta directa.

1. Proponga un modelo bayesiano que, a partir de encuestar a n estudiantes, permita estimar π_a. Explique la elección de la función de verosimilitud y el *prior* e indique cómse obtiene el *posterior*.
2. Utilizando R, simule la obtención de una muestra para el caso en que los estudiantes no mienten y para el caso en que los estudiantes mienten con tres niveles de mentira μ bajo, medio y alto. Compare los resultados de la inferencia.
3. Realice ahora 1000 simulaciones y compare los resultados de las inferencias.

Considere ahora el caso del método propuesto por Warner:

4. Según este método, ¿cuál es la probabilidad (llamémosla λ_W) de que un estudiante responda que apuesta? ¿cuál es la probabilidad de que un estudiante responda queno apuesta?
5. A partir de lo anterior, proponga un modelo razonable sobre cómo se generan los datos.
6. Considere un *prior* uniforme y halle el *posterior* exacto.

Ayuda

Escriba el *posterior* dejando el denominador de la Regla de Bayes expresado como Z. Luego, muestre que:

$$Z = \frac{B(1-p; y+1, n-y+1) - B(p; y+1, n-y+1)}{1-2p}$$

donde $B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ es la función beta incompleta, p es la probabilidad con la que se hace la pregunta alternativa e y es la cantidad depersonas que responden que apuestan.
Para resolver la integral, utilice el método de sustitución y hágalo en términos de λ_W. No es tan terrible como parece.

7. Grafique el *posterior* para diferentes valores de p y concluya.
8. ¿Qué pasaría si el porcentaje de alumnos que apuesta fuera diferente al 40%? Analice los resultados en función de diferentes niveles de p y π_a.
9. Escriba una función de R que le permita realizar la inferencia (en forma aproximada) con un *prior* beta no necesariamente uniforme.

Para el método propuesto por Greenberg:

10. ¿Cuál es la probabilidad λ_G de que un estudiante responda que apuesta? ¿y de que responda que no apuesta?
11. Escriba una función de R que le permita realizar la inferencia (en forma aproximada) con un *prior* beta no necesariamente uniforme.

Para terminar, es hora de comparar todos los escenarios.

12. Utilizando R, simule la obtención de una muestra para el caso en que los estudiantes no mienten, mienten con tres niveles de mentira μ (bajo, medio y alto), se utiliza el método de Werner y se utiliza el método de Greenberg. Compare los resultados de las inferencias en cada caso.
13. Realice ahora 1000 simulaciones y analice los resultados.