ISSN: 2951-6641 (papel) 2951-8156 (online)
https://doi.org/10.54790/rccs.69

NOTA DE INVESTIGACIÓN/RESEARCH NOTE

El estudiante como actor racional ante exámenes tipo test y cómo puntuar sus errores

Students in Tests as Rational Actors and How to Mark Their Mistakes

Daniel Guinea-Martín
Universidad de Málaga, España
daniel.guinea@uma.es

Recibido/Received: 10/6/2023
Aceptado/Accepted: 25/7/2023

RESUMEN

La experiencia de los sociólogos que enseñamos la disciplina en la universidad muestra que el uso de los exámenes tipo test con respuestas cerradas se ha generalizado en nuestra disciplina, entre otras. Si no se penalizan las respuestas erróneas, estos test se transforman en un contexto que favorece el surgimiento de un actor racional en la forma del estudiante-jugador. Este minimiza su esfuerzo aprovechando la probabilidad de acertar con que le obsequia la suerte. Aplicando la definición de variable aleatoria de tipo Bernoulli, en esta nota metodológica presentamos la puntuación que deben tener las respuestas erróneas en exámenes de cualquier extensión en su número de preguntas y opciones de respuesta. Desviarse de esta puntuación supone, o bien favorecer el surgimiento del estudiante-jugador, o bien penalizar en exceso la asunción del riesgo a equivocarse que todo estudiante debe afrontar en un test.

PALABRAS CLAVE: acción racional; estudiante; evaluación; examen; nota; penalización; pregunta cerrada; puntuación; test.

CÓMO CITAR: Guinea Martín, D. (2023). El estudiante como actor racional ante exámenes tipo test y cómo puntuar sus errores. *Revista Centra de Ciencias Sociales*, 2(2), 97-114. https://doi.org/10.54790/rccs.69

English version can be read on https://doi.org/10.54790/rccs.69

ABSTRACT

Sociologists teaching at universities have seen the rise and expansion of tests with closed-ended questions. If wrong answers are not penalized, such tests become a context that favors the student qua gambler or rational actor that minimizes effort. The latter chooses answers randomly. Taking the definition of a Bernoulli random variable, I show the grade that wrong answers should have in tests with any number of closed-ended questions and options to answer. Grading otherwise would either encourage the student qua gambler, or penalize too much the risk-taking that answering a test requires.

KEYWORDS: closed-ended question; exam; grading; penalty; rational action; student; test.

1. Introducción

Los exámenes tipo test (a los que me referiré de aquí en adelante como exámenes o test simplemente) han sido objeto de reflexión sociológica desde hace más de cincuenta años (véase, por ejemplo, Goslin y Glass, 1967). Sin embargo, su expansión más allá de los controvertidos test de inteligencia y las pruebas de acceso a universidades estadounidenses es más reciente. Como señala Edwards (2006), su proliferación puede vincularse al argumento parsoniano sobre la escuela: en las sociedades meritocráticas, esta institución cumple la doble función de (a) selección del talento, y (b) la consiguiente asignación de individuos a ocupaciones (véase, sin embargo, Bennett de Marrais y LeCompte, 1998, para una crítica de este argumento).

En los últimos años los test se realizan en numerosos grados universitarios españoles, incluyendo los de ciencias sociales, donde las materias discursivas abundan y, por ello, no parece la mejor manera de evaluar los conocimientos adquiridos. Como sociólogos deberíamos estudiar las causas de esta tendencia. Sin ser este el objeto de esta breve nota metodológica, sí deseo mencionar que la razón más citada en mi experiencia es que alumnos y docentes prefieren los test porque estos requieren menos esfuerzo que el examen con preguntas abiertas: los docentes lo corrigen más fácil y rápidamente; además, se reduce el conflicto en las revisiones o reclamaciones de notas. Por su parte, el alumno se ahorra escribir un razonamiento propio, así como la posible comisión de faltas gramaticales y ortográficas en su desarrollo. Otra ventaja de los test es que, aun partiendo de que las decisiones sobre su diseño y contenidos son subjetivas, la estandarización de sus respuestas cerradas hace más objetiva su evaluación (Andreasen, Rasmussen e Ydesen, 2013).

Sea como fuere, el objetivo de esta nota metodológica es advertir de que el test en el que las respuestas erróneas no se penalizan es un contexto social que favorece el cálculo racional del famoso *homo economicus* (Levitt y List, 2008). En este caso, su modelo ideal-típico, en el sentido weberiano (Weber, [1922] 1982), es el estudiante exclusivamente motivado por el fin de aprobar con el menor esfuerzo (es decir, sin estudiar). Tal estereotipo del estudiante ocioso responde las preguntas del test al azar. Ante esta situación, numerosos docentes desconocen cómo se deberían valorar las respuestas erróneas para reducir a cero la esperanza que tiene el estudiante-jugador de aprobar. En algunos casos no penalizan, otorgando con ello una amplia ventaja al estudiante que nada ha estudiado. Como se verá más adelante, esta equivale a la probabilidad de acertar al azar la respuesta correcta entre las opciones de respuesta que tiene cada pregunta. (Varios estudios, véanse, por ejemplo, Dehnad, Nasser y Hosseini, 2014, y Schneid *et al.*, 2014, concluyen que tres opciones de respuesta es el número óptimo para maximizar la validez y fiabilidad de los test.) Traducido a puntos, el estudiante-jugador obtendría en promedio la nota máxima del examen, por ejemplo 10, multiplicada por la probabilidad anterior; por ejemplo, si hay 3 posibles respuestas, en promedio su nota final sería $10 \cdot \frac{1}{3} = 3{,}3$ puntos.

En otros casos penalizan demasiado poco, con lo que continúa otorgándose una ventaja al estudiante ocioso que se situará en el intervalo $\left(0, \frac{1}{número de posibles respuestas}\right)$, dependiendo de la penalización empleada.

Y, finalmente, en algunos casos penalizan en exceso las respuestas incorrectas en relación con aquella que eliminaría el papel del azar. Esta situación desplaza la penalización a la izquierda en la línea de los reales: la del estudiante-jugador pasa a ser negativa, y la del estudiante que se ha preparado pasa a ser inferior al valor numérico que mejor refleja sus conocimientos.

Aunque los profesores a veces no calculan correctamente cómo penalizar las respuestas incorrectas, algunos estudiantes emplean estrategias diseñadas para obtener la máxima nota posible en los tests con el mínimo estudio posible (véanse, por ejemplo, Psiconociendo, 2022, y Sentipia, 2022). Existe incluso una campaña para que el Tribunal Constitucional declare ilegal la penalización por respuesta incorrecta en los test (véase Icaro100, 2010) en base a la justificación jurídica propuesta por Muñoz Clares y Caballero Salinas (2019).

Ante tal situación, con esta nota metodológica se pretende que tanto profesores como alumnos aborden la cuestión (a) conociendo sus detalles técnicos, y (b) siendo conscientes del tipo ideal de alumno que uno u otro tipo de puntuación, sin penalización vs. con penalización justa, favorece: el jugador vs. el responsable, respectivamente. El resultado final de una correcta calibración de la puntuación, corrigiendo por respuestas equivocadas, es que la nota final del test refleja numéricamente el estado del conocimiento del alumno.

Además, y al contrario que algunos artículos que versan sobre la misma materia y que me han inspirado a escribir esta nota (véase, por ejemplo, Morales, 2017; entre los muchos recursos sobre estos temas que se pueden encontrar en la web, destaco los de la American Statistical Association, 2013; Bickis, 2017 y Stanbrough, 2009), presento la fórmula general para calcular la nota de las respuestas erróneas, independientemente del número de preguntas que tenga el examen, y del número de opciones de respuesta que tenga cada una. La única asunción constante es que el valor total del examen es de diez puntos, como es usual en la escuela y universidad española. (En cualquier caso, si se usara otra puntuación, simplemente bastaría con sustituir el número 10 por el que corresponda en las fórmulas que siguen.)

El resultado es un cálculo sencillo, pero no trivial. El cómputo requiere el dominio de algunos conceptos estadísticos básicos: espacio muestral, variable aleatoria discreta de tipo Bernoulli y binomial, así como su esperanza y función de probabilidad o cuantía. En esta nota explicaré brevemente, y ejemplificaré, estos conceptos que se encuentran tratados más extensamente en cualquier manual de estadística (véase, por ejemplo, Martín-Pliego y Ruiz-Maya, 2006).

2. El actor racional y el papel del azar en los test

Llevando al límite el modelo del estudiante exclusivamente motivado por el fin de aprobar con el menor esfuerzo, este equivale a no estudiar nada. Pero aun no estudiando nada, si no existe penalización por respuesta incorrecta, tal estudiante *qua* agente racional responde a las preguntas cerradas porque puede adivinar por puro azar la respuesta correcta de entre las opciones de respuesta.

Elegir aleatoriamente una respuesta tiene una probabilidad $p = \frac{1}{m}$ de acertar, y una probabilidad $q = 1 - p = 1 - \frac{1}{m} = \frac{m-1}{m}$ de equivocarse. Ciertamente, la probabilidad de acertar al azar todas las preguntas de un examen es menor, puesto que, si asumimos que cada pregunta es una prueba independiente de las otras preguntas, lo cual es razonable para el caso del estudiante que no ha estudiado nada, entonces la probabilidad de acertar las *n* preguntas de un examen es $p^n < p$.

Esta situación replica los juegos de azar que en el siglo XVIII modeló el matemático suizo Jacob Bernoulli ([1713] 1993), conociéndose por ello como experimentos o pruebas de Bernoulli. La cuestión fundamental es la siguiente: los estudiantes que no estudian nada y que eligen sus respuestas al azar son premiados si las normas de puntuación del examen otorgan puntos positivos a las respuestas correctas, por ejemplo, un punto, y cero puntos tanto a las no respuestas como a las respuestas incorrectas. En este caso, el estudiante-jugador puede esperar ganar $1 \cdot p + 0 \cdot q = 1 \cdot \frac{1}{m} + 0 \cdot \frac{m-1}{m} = \frac{1}{m} = p$ puntos en cada pregunta del examen, una cantidad nada desdeñable.

El objetivo de este artículo es explicar el cálculo anterior, así como la norma de puntuación que anularía completamente tal ventaja *a priori*, de tal modo que, en promedio, el estudiante-jugador obtenga un cero en el examen.

3. Conceptos fundamentales

Responder a una pregunta con opciones de respuesta cerradas es un ejemplo de prueba o experimento llamado «de Bernoulli» porque solo hay dos posibles resultados para el estudiante: acierto o error, que podemos denotar como ω_1 y ω_2, respectivamente, y que conforman el llamado «espacio muestral» Ω de la prueba. Pero con eventos no podemos operar matemáticamente. Para hacerlo introducimos la función «variable aleatoria» (v.a. de aquí en adelante); en este caso la denominamos $Y \equiv$ «acertar al azar la respuesta a una pregunta», que traduce, por así decir, los dos eventos $\Omega = \{\omega_1, \omega_2\}$ del espacio muestral Ω a los dos valores numéricos de la v.a. Y: $Y = \{y_1, y_2\}$. Por ello es usual ver en los manuales de estadística la siguiente notación:

$$(1)\ Y : \Omega \rightarrow R.$$

La expresión (1) significa que la v.a. Y es una función definida sobre el espacio muestral (conjunto de los resultados de un experimento aleatorio) que toma valores en el cuerpo de los números reales R. Sin embargo, en experimentos o pruebas de Bernoulli, la función v.a. traduce los resultados de la prueba al campo de los enteros Z:

$$Y{:}\Omega \to Z$$

En concreto, como he hecho anteriormente, se suele reservar (por mor de la claridad expositiva y la comodidad operativa) el entero 1 al «éxito» de la prueba (acertar en nuestro ejemplo corresponde con el valor 1 de la v.a.: $Y = y_1 = 1$) y el entero cero a su contrario, el «fracaso» o error ($Y = y_2 = 0$) (Baclawski, 2008, p. 48).

Cada valor y_1, y_2 de la v.a. Y tiene una probabilidad asociada de verificarse en una prueba o experimento dado o, en nuestro contexto, en cada pregunta de examen: el estudiante acierta ($\omega_1 = \text{acierto}, y_1 = 1$) con una probabilidad $p = \frac{1}{m}$ y se equivoca ($\omega_1 = error, y_2 = 0$) con una probabilidad $\frac{m-1}{m}$. En resumen, la v.a. tiene una distribución de probabilidad que, en nuestro caso, también se llama distribución de Bernoulli y que se expresa sucintamente como $Y \sim \text{Bernoulli}\ (p)$.

Se llama «esperanza» (un operador matemático denotado por $E[\cdot]$) de una v.a. Y a su «valor esperado»: $E[Y]$. Este depende de su distribución de probabilidad, tendiendo hacia los valores de Y con mayor probabilidad asociada. Cuando contamos con datos empíricos, el equivalente es la media: esta se aproxima a la esperanza a medida que la muestra aumenta o se repite[1].

En el caso discreto la esperanza se calcula como el sumatorio del producto de la probabilidad (p_i) que tiene cada valor i de la variable $Y(y_i)$ por ese mismo valor y_i (en nuestro caso $i = \{1,2\}$): $E[Y] = \sum_i y_i\, p_i = y_1 p_1 + y_2 p_2 = y_1 \frac{1}{m} + y_2 \left(1 - \frac{1}{m}\right) = y_1 \frac{1}{m} + y_2 \frac{m-1}{m}$.

Cuando la respuesta correcta se puntúa con el entero 1 ($y_1 = 1$) y la incorrecta con el entero cero ($y_2 = 0$), el valor esperado es p: $E[Y] = 1 \cdot \frac{1}{m} + 0 \cdot \frac{m-1}{m} = \frac{1}{m} = p$.

4. Extensión a cualquier tipo de puntuación por pregunta

Puntuar con un 1 la respuesta correcta tiene sentido si: (a) la nota máxima que se puede obtener son 10 puntos que representan la perfección absoluta en la respuesta y (b) el examen consta de 10 preguntas de igual valor; o si el examen tiene preguntas y la nota final no está restringida a 10 puntos, sino que es $1 \cdot n = n$ puntos.

En la siguiente discusión mantendré fijo el supuesto (a) puesto que la escala de notas de cero a 10 puntos para la evaluación de los exámenes es la más común en nuestro sistema educativo. Sin embargo, el supuesto (b) es cuestionable ya que los test varían el número de preguntas que plantean; de hecho, rara vez se limitan a solo 10 preguntas, puesto que compensan no tener que razonar explícitamente con exigir la respuesta a muchas preguntas.

La nota máxima del examen, 10, debe repartirse entre las *n* preguntas, con lo que el valor de la respuesta correcta, llamémosle *u*, debe ser $u = \frac{10}{n}$ puntos. Por ejemplo, con n=20 preguntas, $u = \frac{10}{n} = \frac{10}{20} = \frac{1}{2} = 0.5$ puntos. Con ello vemos que la imagen o rango de nuestra v.a. Bernoulli (es decir, el conjunto de valores que puede tomar) se corresponde con el conjunto de los números racionales (Q) y no en el de los enteros (Z). Es decir, en estos casos se verifica que:

$$Y : \Omega \rightarrow \mathrm{Q}$$

Como he mencionado anteriormente, esta particularidad no se tiene usualmente en cuenta en las exposiciones sobre cómo puntuar exámenes de tipo test (véase, por ejemplo, Morales, 2017), pero es de gran importancia para resolver el *quid* de la cuestión: cómo puntuar cada respuesta incorrecta: ¿qué número racional debemos asignar a y_2, el segundo posible valor de nuestra v.a. *Y* que representa a los errores en las respuestas? Esta ignorancia la podemos expresar formalmente en la llamada función de probabilidad o de cuantía de nuestra v.a. *Y*, es decir, la función que informa de la probabilidad con que la v.a. *Y* adopta cada uno de sus y_i valores, $P(Y=y_1)$ o $P(Y=y_2)$:

$$Y = \begin{cases} Y(\omega_1 = \text{acierto}) = y_1 = u = \dfrac{10}{n}, P\left(Y = y_1 = \dfrac{10}{n}\right) = \dfrac{1}{m} = p \\ Y(\omega_2 = \text{error}) = y_2 = ?, P(Y = y_2 = ?) = \dfrac{m-1}{m} = q = 1 - p \end{cases}$$

Un docente que comparte el juicio de valor según el cual la nota de un estudiante que no ha estudiado nada debe ser el equivalente numérico a la nada, es decir, un cero, estará de acuerdo con que la cantidad y_2, por ahora desconocida, debe ser tal que centre la esperanza de la variable *Y* en el cero, $E[Y]=0$, y no en $p = \frac{1}{m}$ como era el caso de la v.a. *Y* de tipo Bernoulli con la que empezamos la exposición, es decir, aquella que asocia el valor cero a las equivocaciones. Por lo tanto, y dado que la definición de esperanza matemática es $E[Y]=\sum_i y_i p_i$, nuestra situación se representa como sigue:

$$E[Y] = y_1 \cdot p + y_2 \cdot q = \frac{10}{n} \cdot \frac{1}{m} + y_2 \frac{(m-1)}{m}$$

Para lograr que esta esperanza valga cero, $E[Y] = \frac{10}{n} \cdot \frac{1}{m} + y_2 \cdot \frac{m-1}{m} = 0$ la nota por respuesta errónea, y_2, debe adoptar un valor preciso y, para conocerlo, escribimos la expresión de la esperanza igualada a cero en términos de y_2:

$$\frac{10}{n}\cdot\frac{1}{m}+y_2\cdot\frac{m-1}{m}=0$$
$$y_2\cdot\frac{m-1}{m}=\frac{-10}{nm}$$
$$y_2=\frac{-10m}{nm(m-1)}=\frac{-10}{n(m-1)}=\frac{10}{n}\cdot\left(\frac{-1}{m-1}\right)=u\cdot\left(\frac{-1}{m-1}\right)=\frac{-u}{m-1}$$

(Para la discusión posterior en ocasiones conviene simplificar la notación denotando con v la penalización $\frac{-1}{m-1}$ que se aplica a cada respuesta incorrecta, de modo que $y_2=u\cdot\left(\frac{-1}{m-1}\right)=uv$.) Es decir, en un examen de tipo test cuya nota máxima son 10 puntos, con n preguntas, donde cada pregunta tiene m posibles respuestas, puntuando con $y_2=\frac{-10}{n(m-1)}=u\left(\frac{-1}{m-1}\right)=uv$ cada respuesta incorrecta, el alumno que juega a adivinar las respuestas azarosamente tiene una nota esperada de cero.

Por lo tanto, la función de cuantía de nuestra v.a. Y de tipo Bernoulli que toma valores en el conjunto de números racionales Q y con esperanza cero es la siguiente:

$$Y=\begin{cases} Y(\text{acierto})=y_1=u=\dfrac{10}{n}, P(Y=y_1)=\dfrac{1}{m}\\ Y(\text{error})=y_2=uv=u\cdot\left(\dfrac{-1}{m-1}\right)=uv, P(Y=y_2)=\dfrac{m-1}{m}\end{cases}$$

donde tenemos: (a) los valores $Y(\text{acierto})=y_1=u$ e $Y(\text{error})=y_2=uv$; (b) n preguntas en el examen; (c) m posibles respuestas por pregunta.

En resumen, $\frac{10}{n}\cdot\left(\frac{-1}{m-1}\right)$ es la puntuación que debe tener una respuesta errónea en un test para que, quien juega al azar con su respuesta, obtenga un cero en promedio. Nótese que esta expresión es una función de dos parámetros: las n preguntas de las que conste el examen y las m opciones de respuesta que tiene cada pregunta. Como he mencionado anteriormente, en esta formulación se mantiene la nota máxima del examen fija en 10 puntos. Si no fuera así, este sería un tercer parámetro k del que dependería la puntuación por respuesta correcta e incorrecta. En la conclusión se reproduce esta fórmula, acercándola al lenguaje natural para que trascienda como mensaje principal de esta nota metodológica y los docentes interesados puedan aplicarla fácilmente a su situación.

Comprobemos que, efectivamente, haciendo que $y_2=uv$, el estudiante-jugador obtiene en promedio un cero en cada pregunta.

$$E[Y]=y_1\cdot p+y_2\cdot q=\frac{10}{n}\cdot\frac{1}{m}-\frac{10}{n(m-1)}\cdot\frac{(m-1)}{m}=\frac{10}{nm}-\frac{10}{nm}=0$$

Dado este esquema de puntuación, se derivan los siguientes hechos:

1. La nota máxima del examen con preguntas es efectivamente $n \cdot \frac{10}{n} = 10$;
2. El mínimo es $n \cdot \frac{10}{n} \cdot \left(\frac{-1}{m-1}\right) = \frac{-10}{m-1}$;
3. Si alguien quisiera adoptar normas de puntuación que asignan el valor $w \neq uv$ a $Y(\text{error}) = y_2$, caben dos posibilidades:

    - si $w < uv$ las normas de puntuación penalizan demasiado severamente las respuestas incorrectas, cohibiendo la asunción del riesgo que supone contestar cuando se tiene la mínima duda sobre cuál es la opción correcta;
    - si $w > uv$ las normas de puntuación penalizan demasiado levemente las respuestas incorrectas, aproximando el estudiante real bajo examen al modelo ideal-típico del actor racional que he denominado el «estudiante-jugador».

4.1. Nota esperada en el conjunto del examen

Por la linealidad del operador esperanza, $E[\cdot]$, y asumiendo que las n preguntas dan lugar a n $Y_1, Y_2, \ldots, Y_n$ v.a. independientes e idénticamente distribuidas (i.i.d.) con la función de cuantía descrita anteriormente, la nota global del examen esperada será igual a cero: $E[n \cdot Y] = n \cdot E[Y] = n \cdot 0 = 0$.

4.2. Ejemplos

Partamos de un ejemplo donde la función de cuantía de la v.a. $Y \sim \text{Bernoulli}\left(p = \frac{1}{5}\right)$ refleja un test con $n = 30$ preguntas y $m = 5$ opciones de respuesta:

$$Y = \begin{cases} Y(\text{acierto}) = y_1 = \frac{10}{30} = \frac{1}{3}, P\left(Y = y_1 = \frac{1}{3}\right) = \frac{1}{5} \\ Y(\text{error}) = y_2 = \frac{1}{3}\left(\frac{-1}{5-1}\right) = \frac{-1}{12}, P\left(Y = y_2 = \frac{-1}{12}\right) = \frac{4}{5} \end{cases}$$

Comprobemos que es así:

$$E[Y] = \frac{1}{3} \cdot \frac{1}{5} - \frac{1}{12} \cdot \frac{4}{5} = \frac{1}{15} - \frac{4}{60} = \frac{1}{15} - \frac{1}{15} = 0$$

Otro ejemplo es un test con $n = 20$ preguntas y $m = 3$ opciones de respuesta por pregunta. Entonces cada respuesta correcta vale $u = \frac{10}{20} = \frac{1}{2}$ puntos, y cada respuesta incorrecta debería evaluarse con

$$\frac{1}{2} \cdot \left(\frac{-1}{3-1}\right) = \frac{-1}{2} \cdot \frac{1}{2} = \frac{-1}{4} = -0.25 \text{ puntos}$$

Con ello la esperanza de la v.a. Y es igual a cero:

$$E[Y] = y_1 \cdot p + y_2 \cdot q = \frac{1}{2} \cdot \frac{1}{3} - \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6} - \frac{1}{6} = 0$$

Si el mismo examen tuviera preguntas con $m = 4$ opciones de respuesta, cada error debería puntuarse con $\frac{1}{2} \cdot \left(\frac{-1}{4-1}\right) = \frac{-1}{2} \cdot \frac{1}{3} = \frac{-1}{6} \approx -0.17$ puntos. Con ello, en efecto, su esperanza es:

$$E[Y] = \frac{1}{2} \cdot \frac{1}{4} - \frac{1}{6} \cdot \frac{3}{4} = \frac{1}{8} - \frac{1}{8} = 0$$

Si, en cambio, el examen tuviera $n = 40$ preguntas con $m = 3$ opciones de respuesta, acertar valdría $u = \frac{10}{40} = \frac{1}{4}$ puntos y equivocarse $uv = \frac{1}{4} \cdot \left(\frac{-1}{3-1}\right) = \frac{-1}{4} \cdot \frac{1}{2} = \frac{-1}{8} = -0.125$ puntos. Con lo cual el valor esperado es:

$$E[Y] = \frac{1}{4} \cdot \frac{1}{3} - \frac{1}{4 \cdot 2} \cdot \frac{2}{3} = \frac{1}{4} \cdot \frac{1}{3} - \frac{1}{4} \cdot \frac{1}{3} = 0$$

La misma situación con $m = 4$ opciones de respuesta daría lugar a evaluar con $\frac{1}{4} \cdot \left(\frac{-1}{4-1}\right) = \frac{-1}{4} \cdot \frac{1}{3} = \frac{-1}{12} \approx -0.083$ puntos la respuesta incorrecta. En efecto, en tal caso su esperanza se mantiene en cero:

$$E[Y] = \frac{1}{4} \cdot \frac{1}{3} - \frac{1}{8} \cdot \frac{2}{3} = \frac{1}{12} - \frac{1}{12} = 0$$

Para resumir, cuantas más opciones de respuesta y más preguntas haya en un examen, menor será la penalización por respuesta incorrecta.

5. Número de respuestas correctas en n preguntas: paso de la distribución de Bernoulli a la distribución binomial

En una ocasión me equivoqué en un paso del código que uso para alternar aleatoriamente la posición de la respuesta correcta en un test con $n = 20$ preguntas y $m = 3$ opciones de respuesta. El resultado fue que en la primera convocatoria ordinaria de la asignatura «Sociología de la diversidad», en el grado en Sociología de la UNED, la respuesta correcta se encontró siempre en la tercera posición. Un alumno que no había estudiado nada, y que se presentó a la segunda convocatoria ordinaria (de recuperación), decidió seguir apostando por la respuesta (c) en cada pregunta del test, cuando en esa convocatoria la respuesta correcta ya variaba aleatoriamente de pregunta en pregunta. Con ello acertó al azar la respuesta a 8 preguntas, y se equivocó en 12, obteniendo una nota de $8 \cdot \frac{1}{2} - 12 \cdot \frac{1}{4} = 1$ punto «gratis», i.e., el 10% de la nota total del examen. Pero si no hubiera penalizado las respuestas incorrectas según he explicado en la sección anterior, el alumno habría conseguido $8 \cdot \frac{1}{2} = 4$ puntos por pura suerte, el 40% de la nota total. Con este resultado final el alumno podría considerar reclamar la nota final del curso argumentando que estaba muy cerca del aprobado en el examen, como es frecuente que ocurra entre los alumnos cuya nota se aproxima al valor cinco por la izquierda.

Una pregunta interesante es: ¿cuál es la probabilidad de obtener 8 aciertos de 20 intentos? Obtener x aciertos en n intentos o pruebas describe el proceso generador de datos de una variable aleatoria con una distribución binomial. Llamemos a esta variable aleatoria $X \equiv$ «**número de aciertos en n preguntas de un test**». Además, este modelo estadístico asume que cada una de las n preguntas del test es independiente de las otras y que, además, la v.a. asociada a cada pregunta está idénticamente distribuida, i.e., sigue una distribución Bernoulli con probabilidad p[2].

Bajo estas condiciones la probabilidad de obtener exactamente x aciertos en n preguntas se expresa con $P(X = x)$, y la función de cuantía de la variable $X \sim B(n, p)$ es:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, y = 0,1, \ldots ,20$$

Siendo X una v.a. binomial, sabemos que su esperanza es $E[X] = np$ (Martín-Pliego y Ruiz-Maya, 2006), por lo que, en el caso $n = 20$, $p = \frac{1}{3}$, $E[X] = 20 \cdot \frac{1}{3} = 6.\overline{6}$. En el ejemplo citado el resultado ha sido algo mayor que la esperanza, puesto que han ocurrido $x = 8$ aciertos. Podemos cuantificar la probabilidad de que esto ocurra como sigue:

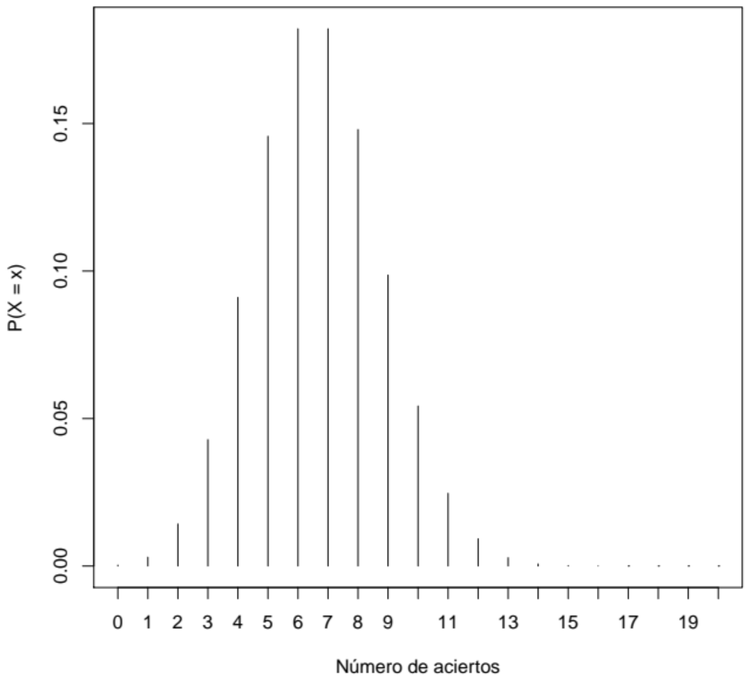
$$P(X = x = 8) = \binom{20}{8} p^8 q^{12} = \frac{20!}{12!\,8!}\, p^8 q^{12} = 125970 \cdot \left(\frac{1}{3}\right)^8 \cdot \left(\frac{2}{3}\right)^{12}$$
$$= 125970 \cdot 0{,}0001524158 \cdot 0{,}007707347 = 0{,}1479797.$$

Es decir, alrededor de un 15% de los estudiantes-jugadores obtendrán 8 aciertos por azar o, alternativamente, si uno de estos estudiantes repitiera el test *ad infinitum*, obtendría 8 aciertos por azar en casi el 15% de las ocasiones.

¿Cuál es la probabilidad para los 20 resultados posibles, desde los cero hasta los 20 aciertos obtenidos por azar? El gráfico 1 de la función de cuantía de $X \sim B\left(n = 20, p = \frac{1}{3}\right)$ lo muestra:

Gráfico 1

Distribución de probabilidad de X~B(n=20, p=1/3)



En el gráfico 1 observamos que la probabilidad más alta corresponde a los resultados muestrales que coinciden con los enteros alrededor de la esperanza matemática de $E[X]$: 6 aciertos y 14 errores, por un lado, y 7 aciertos y 13 errores por otro. En efecto, esta distribución binomial tiene dos valores con probabilidad máxima, es decir, dos modas M_0, verificándose la siguiente desigualdad (la prueba se encuentra en Arnáiz,1986, citado en Martín-Pliego y Ruiz-Maya, 2006, p. 188):

$$\begin{aligned} np - q &\leq M_0 \leq np + p \\ 20 \cdot \frac{1}{3} - \frac{2}{3} &\leq M_o \leq 20 \cdot \frac{1}{3} + \frac{1}{3} \\ 6 &\leq M_o \leq 7. \end{aligned}$$

Y el valor exacto de esta, la probabilidad más alta de esta distribución de cuantía, es:

$$P(X = 6) = \binom{20}{6} p^6 q^{14} = 0{,}1821288;$$
$$P(X = 7) = \binom{20}{7} p^7 q^{13} = 0{,}1821288.$$

Es decir, observaremos que alrededor de un 18% de los estudiantes jugadores obtendrán 6 aciertos y 14 errores, y el mismo porcentaje obtendrá 7 aciertos y 13 errores. Con estos, los resultados más comunes, y no teniendo dudas a estas alturas de que en un test de este tipo la respuesta correcta vale $\frac{1}{2}$ puntos y la incorrecta $\frac{-1}{4}$ puntos, se obtendrán las siguientes notas finales, respectivamente:

$$\begin{aligned} \text{Nota con 6 aciertos} &= 6 \cdot \frac{1}{2} \quad -14 \cdot \frac{1}{4} = -0{,}5 \\ \text{Nota con 7 aciertos} &= 7 \cdot \frac{1}{2} \quad -13 \cdot \frac{1}{4} = 0{,}25 \end{aligned}$$

Por el contrario, tener tan mala suerte como para no acertar al azar ninguna de las 20 preguntas es una circunstancia que solo aflige al

$$P(X = 0) \cdot 100 = \binom{20}{0} p^0 q^{20} \cdot 100 = q^{20} \cdot 100 = 0.0003007287 \cdot 100 \approx 0.03\%$$

de los estudiantes-jugadores o, alternativamente, de las ocasiones en que un estudiante juega a adivinar las respuestas. En el otro extremo, acertar todas las preguntas por pura suerte es un evento que tiene una probabilidad asociada aún más pequeña, de solo (expresada en tantos por ciento)

$$P(X = 20) \cdot 100 = \binom{20}{20} p^{20} q^0 \cdot 100 = p^{20} \cdot 100 = 2.867972 \cdot 10^{-10} \cdot 100 = 2.867972 \cdot 10^{-8} = 0.0000000286792\%$$

es decir, es un evento que esperamos que ocurra con una frecuencia cercana a las tres ocasiones cada cien millones de intentos.

Otra pregunta relevante concierne al número mínimo de aciertos necesarios para aprobar el test con las características dadas ($n = 20$, $m = 3$), i.e., para obtener 5 o más puntos, que es la nota convencionalmente interpretada como un «aprobado» en exámenes cuya nota máxima es 10. Entonces:

$$\begin{aligned} x \cdot \frac{1}{2} &\quad -(20 - x) \cdot \frac{1}{4} \geq 5 \\ \frac{x}{2} &\quad \frac{-20 - x}{4} \geq 5 \\ \frac{2x - 20 + x}{4} &\quad \geq 5 \\ 3x - 20 &\quad \geq 20 \\ 3x &\quad \geq 40 \\ x \geq \frac{40}{3} &\quad \geq 13{,}\overline{3} \end{aligned}$$

Por lo tanto, con un test de estas características el alumno necesita acertar 14 respuestas, ya que con 13 todavía estará por debajo del 5:

$$\text{Nota con 13 aciertos} = 13 \cdot \frac{1}{2} \quad -7 \cdot \frac{1}{4} = 4{,}75$$
$$\text{Nota con 14 aciertos} = 14 \cdot \frac{1}{2} \quad -6 \cdot \frac{1}{4} = 5{,}5$$

Y la probabilidad de obtener los 14 aciertos necesarios para aprobar el test es de

$$P(X = 14) = \binom{20}{14} p^{14} q^{6} = 0.0007114406$$

Es decir, de un 0,07%: lo lograrán siete de cada diez mil estudiantes o intentos.

En comparación, si no existiera penalización por respuesta errónea bastaría con acertar la mitad de las respuestas para aprobar, con la comodidad de poder probar suerte en todas las preguntas. Es decir, se podría aprobar aun errando en la mitad de las respuestas. La probabilidad de que esto ocurra en un examen con 20 preguntas con tres opciones de respuesta cada una es de

$$P(X = 10) = \binom{20}{10} p^{10} q^{10} = 0.0542592$$

Alrededor de cinco estudiantes-jugadores de cada cien aprobarían el test. Nótese que el orden de magnitud de esta cantidad de aprobados es cien veces superior al que existe cuando se penalizan las respuestas incorrectas y el estudiante responde a todas las preguntas.

6. Conclusión

Esta nota metodológica contiene una conclusión principal para todo docente que use test para evaluar a sus alumnos. Sin hacer uso de la poca intuitiva notación matemática y sus variables mudas (m, n, etc.), la conclusión se puede expresar como sigue: la respuesta correcta debe valer...

$$\frac{\text{Nota máxima del text}}{\text{Número de preguntas del test}} \text{ puntos}$$

Y la respuesta incorrecta debe valer...

$$\frac{\text{Nota máxima del text}}{\text{Número de preguntas del test}} \cdot \frac{-1}{\text{Número de opciones por pregunta} - 1} \text{ puntos}$$

Explicar claramente este esquema de puntuación a los estudiantes, incidiendo en la baja probabilidad de aprobar el test por puro azar, modificará la «definición de la situación» (Thomas y Thomas, 1928) y esta, a su vez, influirá en la preparación del estudiante: el test ya no es un contexto social fecundo para la aparición del jugador racional que saca el máximo partido de su probabilidad de acertar al azar.

Sin embargo, conviene concluir este artículo llamando la atención de que, como explica Pes (2009), la estrategia del estudiante-jugador puede extenderse al cálculo del número de preguntas que es óptimo responder cuando las respuestas erróneas se penalizan. Siguiendo sus consejos, hay estudiantes que acceden a repositorios con todos los test realizados hasta la fecha en su asignatura (los de la UNED, por ejemplo, están a disposición de los miembros de su comunidad; véase UNED, 2020). Con ellos, estudian memorizando las preguntas que históricamente han aparecido más frecuentemente en las distintas convocatorias. Si el día de su examen aparecen un número suficiente de estas preguntas memorizadas para sumar un aprobado, su mejor estrategia es no responder a ninguna otra pregunta.

Para eludir este subterfugio propio ahora no del tipo ideal del estudiante-jugador, pero sí del estudiante-estratega que busca aprobar por la mínima, memorizando mecánicamente y sin dominar los contenidos de la asignatura, el docente podría incentivar la respuesta a todas las preguntas penalizando por igual tanto la no respuesta como la respuesta errónea. Al fin y al cabo, si asumimos que cada pregunta tiene una respuesta correcta, la no respuesta es en sí misma una respuesta errónea. En resumen, el docente podría contrarrestar al estudiante-estratega puntuando positivamente el conocimiento del alumno, y negativamente su desconocimiento, representado este último tanto por la no respuesta como por la respuesta equivocada. De este modo, además, se lidiaría con la preocupación de Muñoz Clares y Caballero Salinas (2019) relativa al tratamiento diferencial de las no respuestas y las respuestas equivocadas (base jurídica de la campaña para prohibir la penalización en los test promovida por Icaro100, 2010).

Con todo ello, se establecería una nueva definición de la situación donde el estudiante es consciente de la necesidad de estudiar seriamente la asignatura. Si lo hace aumentará su probabilidad de conocer la respuesta correcta hasta el máximo $p = 1$ (certeza absoluta), o bien hasta una cantidad p superior a la que se corresponde al puro azar. Ello es así puesto que el estudiante que hace honor a su nombre y estudia será capaz de restringir las posibles respuestas correctas a un subconjunto de las opciones originales. Como recuerdan Dehnad, Nasser y Hosseini (2014), esta situación, favorecida por los test con tres opciones por pregunta donde la conjetura se restringe a dos opciones, ya no se corresponde con un acto de adivinanza azarosa. Ahora se trata de una «conjetura informada» con un grado de acierto final en relación directa con la intensidad del estudio y la comprensión de la asignatura. Es decir, el examen tipo test se acerca a su ideal como instrumento para valorar numéricamente el grado de conocimiento alcanzado por el estudiante.

Una cuestión relacionada con el tema de este artículo que queda abierta para futuras aportaciones la plantean los distintos formatos de preguntas tipo test que existen además de las de única opción de respuesta correcta. Case y Swanson (2001) ofrecen un tratamiento sistemático de estas alternativas, entre las que sobresalen las preguntas con múltiples opciones correctas de respuesta (Palés-Argullós, 2010). Siguiendo la aproximación a la cuestión de Gaviria (2020), podrían considerarse cómo puntuar estos otros formatos alternativos de test para que la esperanza del estudiante-jugador se mantenga en cero.

7. Agradecimientos

Agradezco a dos revisores anónimos de la *Revista Centra*, y a mis colegas Verónica de Miguel y Ricardo Mora, por los comentarios recibidos a un borrador preliminar de este artículo. Los errores y erratas que seguramente, y por desgracia, perduran son de mi exclusiva responsabilidad.

8. Bibliografía

American Statistical Association (2013). *Should you guess? Interview with Rajeeva Laxman Karandikar*. Acceso 8 julio 2023. Disponible en: http://www.worldofstatistics.org/2013/03/04/should-you-guess/.

Andreasen, K. E., Rasmussen, A. e Ydesen, C. (2013). Standardized testing. En J. Ainsworth (Ed.), *Sociology of Education: An A-to-Z Guide* (pp. 737-739). Thousand Oaks, CA: Sage.

Arnáiz, G. (1986). *Introducción a la estadística teórica*. Madrid: Lex Nova.

Baclawski, K. (2008). *Introduction to probability with R*. Boca Raton, FL: Chapman & Hall/CRC. https://doi.org/10.1201/9781420065220

Bennett de Marrais, K. y LeCompte, M. (1998). *The way schools work: A sociological analysis of education*. London: Longman/Addison Wesley.

Bernoulli, J. ([1713] 1993). *Teoría de probabilidades (Ars conjectandi).* Madrid: Sociedad Española de Historia de las Ciencias y de las Técnicas, Llull.

Bickis, Mikelis G. (2017). *STATS 244.3. Elementary Statistical Concepts. Scoring Multiple Choice Exams.* Acceso 8 julio 2023. Disponible en: http://math.usask.ca/~bickis/courses/stats244/marking.html.

Case, S. M. y Swanson, D. B. (2001). *Constructing Written Test Questions for the Basic and Clinical Sciences*. 3.ª edición. Philadelphia, PA: National Board of Medical Examiners.

Dehnad, A., Nasser, H. y Hosseini, A. F. (2014). A Comparison Between Three-and Four-Option Multiple Choice Questions. *Procedia. Social and Behavioral Sciences*, *98*, 398-403. https://doi.org/10.1016/j.sbspro.2014.03.432

Edwards, N. T. (2006). The historical and social foundations of standardized testing: In search of a balance between learning and evaluation. *Shiken and Shiken Research Journal: JALT Testing and Evaluations SIG Newsletter*, *10*, 8-16.

Gaviria, J. L. (2020). *Recomendaciones para la Elaboración de las Preguntas*. Vicerrectorado de Tecnología y Sostenibilidad, Universidad Complutense de Madrid. Acceso 8 julio 2023. Disponible en: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiBoYOztv___AhWEXcAKHf-nAIcQFnoECA4QAQ&url=https%3A%2F%2Fderecho.ucm.es%2Ffile%2Frecomendaciones-para-elaborar-preguntas-a-traves-de-herramienta-cuestionario-moodle&usg=AOvVaw0cZadEz-i2iJ-rQwHGdZsp&opi=89978449.

Goslin, D. A. y Glass, D. C. (1967). The social effects of standardized testing in American elementary and secondary Schools. *Sociology of Education*, *40* (2), 115-131. https://doi.org/10.2307/2112040

Icaro100 (2010). Campaña de impugnación de exámenes tipo test alegando la ilegalidad de la penalización. *Tribuna del Jurista.* Acceso 8 julio 2023. Disponible en: https://

tribunadeljurista.foroes.org/t1486-campaa-de-impugnacin-de-exmenes-tipo-test-alegando-la-ilegalidad-de-la-penalizacin-adjunto-formulario-de-impugnacin.

Levitt, S. D. y List, J. A. (2008). Homo economicus evolves. *Science, 319* (5865), 909-910. https://doi.org/10.1126/science.1153640

Martín-Pliego, F. J. y Ruiz-Maya, L. (2006). *Fundamentos de Probabilidad.* Madrid: Ediciones Paraninfo.

Morales, M. A. (2017). Las matemáticas de la fórmula de puntuación de exámenes test. *El País,* 22 de febrero. Acceso 8 julio 2023. Disponible en: https://elpais.com/elpais/2017/02/22/el_aleph/1487760836225306.html.

Muñoz Clares, J. y Caballero Salinas, J. M. (2019). Problemática de los exámenes test con merma por respuesta errónea. *Revista Aranzadi Doctrinal*, 7.

Palés-Argullós, J. (2010). ¿Cómo elaborar correctamente preguntas de elección múltiple? *Educación Médica*, 3. https://doi.org/10.4321/S1575-18132010000300005

Pes, C. (2009). *¿Cómo Aprobar Un Examen Tipo Test?* Acceso 8 julio 2023. Disponible en: https://www.carlospes.com/articulos/como_aprobar_un_examen_tipo_test.php.

Psiconociendo (2022). *12 Trucos Para Aprobar Un Examen Tipo Test Sin Estudiar*. Youtube. Acceso 8 julio 2023. Disponible en: https://www.youtube.com/watch?v=gR9yHCOW05c&list=RDLVuv0hBxbx3Tk&index=6.

Schneid, S. D., Armour, C., Park, Y. S., Yudkowsky, R. y Bordage, G. (2014). Reducing the number of options on multiple-choice questions: Response time, psychometrics and standard Setting. *Medical Education, 48* (10): 1020-1027. https://doi.org/10.1111/medu.12525

Sentipia (2022). *Cómo Responder Exámenes Tipo Test para Aprobar.* Youtube. Acceso 8 julio 2023. Disponible en: https://www.youtube.com/watch?v=uv0hBxbx3Tk&list=RDLVuv0hBxbx3Tk&index=1.

Stanbrough, J. L. (2009). *The guessing penalty. Let's do the math, shall we?* Batesville, IN: Batesville Community School Corporation. Acceso 8 julio 2023. Disponible en: http://www.batesville.k12.in.us/Physics/CalcNet/Intro/guessing_penalty.htm

Thomas, W. I. y Thomas, D. S. (1928). *The Child in America: Behavior Problems and Programs*. New York, NY: Knopf.

UNED (2020). *Acceso al Depósito de Exámenes*. Acceso 8 julio 2023. Disponible en: http://www.calatayud.uned.es/examenes/examenes_step_0.asp.

Weber, Max ([1922] 1982). *Ensayos sobre metodología sociológica.* Buenos Aires: Amorrortu.

Notas

1 Con nuestro caso de estudio, el resultado empírico que ofrece la media de un conjunto de datos se aproxima al valor de la esperanza de la distribución de probabilidad cuando (a) el mismo estudiante-jugador juega al azar en el examen n veces, con $n \to \infty$; o (b) hay un número k de estudiantes que responden al azar la pregunta del examen, con $k \to \infty$.

2 La distribución Bernoulli es el caso particular de la distribución binomial para $n = 1$ pruebas. Luego $X \sim \text{Bernoulli}\ (p) = B\ (1, p)$ y, por ello, la variable aleatoria se enuncia como acierto/éxito (vs. error/fracaso) e, implícitamente, se quiere decir, «en una única prueba o experimento».

El autor

Daniel Guinea Martín es profesor titular de Sociología en la Universidad de Málaga, donde enseña métodos, sociología de la educación y del turismo. Doctor por el Instituto Universitario Europeo de Florencia, máster por la London School of Economics, diplomado por la Universidad de California en San Diego, así como licenciado en Sociología y graduado en Estadística por la Universidad Complutense de Madrid. Se especializa en métodos de investigación social, desigualdad y segregación. Ha publicado en revistas como *American Sociological Review*, *Social Science Research* y *Ethnography*, entre otras.