

Detección de copia en pruebas del estado

Diego Jara Alvaro Riascos Mauricio Romero

- 1 Introducción
- 2 Revisión de la Literatura
- 3 Descripción del Modelo
- 4 Resultados del Índice para Pruebas Independientes
- 5 Resultados y observaciones

- Este trabajo implementa la metodología de detección de copia en exámenes de múltiples opciones expuesta en (Sotaridona, van der Linden Meijer 2006) basada en el índice Kappa de Cohen.
- Presentamos los resultados de aplicar esta metodología a las pruebas del Estado ICFES 2009 II.
- Con este apoyo se pueden generar alertas y priorizar investigaciones detalladas caso por caso.

- Desde el punto de vista de la literatura académica especializada, el aporte del trabajo es evaluar, con datos reales, las bondades del índice Kappa propuesto por los autores y que ellos validan con datos artificiales generados mediante la simulación de un modelo de respuesta nominal.
- En particular, los resultados ponen en evidencia que, si bien el índice es muy valioso en la práctica, éste adolece de problemas de especificación difíciles de superar utilizando datos reales.

Revisión de la Literatura

- La mayoría de los modelos de detección de copia existentes (van der Linden & Sotaridona 2006) se basan en calcular diferentes estadísticos que cuentan el número de respuestas comunes entre el individuo sospechoso de hacer copia y aquél de quien copia (la fuente).
- Las principales diferencias de estos estadísticos son
 - 1 El conjunto de respuestas sobre las que se calcula el índice (todas las respuestas, correctas e incorrectas, solo las respuestas incorrectas, solo las respuestas incorrectas condicional a la habilidad de los dos individuos, etc.)
 - 2 La distribución que se supone tiene este estadístico.

Índice Kappa

- Una característica importante del índice Kappa es que no se basa en un modelo de respuesta nominal. Es decir, para su aplicación no es necesario primero estimar la probabilidad de que los individuos respondan la opción k en una pregunta. Las hipótesis básicas son:
 - 1 La función de respuesta a una pregunta es probabilística.
 - 2 Es posible seleccionar una muestra en la cual las respuestas son independientes (en particular, no hay copia).

Índice Kappa

- En la primera versión que presentamos del índice Kappa se hace una hipótesis muy fuerte: que la probabilidad conjunta de dos individuos a responder la opción k es igual para todas las preguntas.
- Esta es una hipótesis muy fuerte y por esta razón se utiliza una técnica de recodificación de las respuestas para mitigar el problema.

Índice Kappa

- Para comenzar, se supone que tenemos un cuestionario con N preguntas, n posibles respuestas a cada pregunta y dos estudiantes c (sospechoso de copia) y s (sospechoso de ser la fuente) que responden de forma independiente cada pregunta (no se copian). Los resultados de las respuestas de cada estudiante se pueden representar por una matriz $n \times n$.
- En la celda (i, i) (la diagonal) van a aparecer el número de respuestas coincidentes entre c y s con respuesta común i y en una celda (i, j) (por fuera de la diagonal), va a aparecer el número de veces que s respondió la opción j en las preguntas en que c respondió la opción i .

Índice Kappa

- Sea π_{vv} la probabilidad de que una pregunta sea clasificada en la celda (v, v) y π_{v+} , π_{+v} la probabilidad de que la pregunta sea clasificada en la fila o columna v respectivamente. La probabilidad observada de que los individuos coincidan en una pregunta dada, π_0 es:

$$\pi_0 = \sum_v \pi_{vv} \quad (1)$$

Índice Kappa

- Ahora, bajo la hipótesis de que la distribución conjunta de c y s para las diferentes opciones es igual para todas las preguntas, se puede demostrar que la probabilidad de que coincidan en por lo menos una pregunta π_e es:

$$\pi_e = \sum_v \pi_{v+} \pi_{+v} \quad (2)$$

Índice Kappa

- El estadístico κ (Kappa) se define como:

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e} \quad (3)$$

- La prueba estadística para detectar sospechosos de copia se basa en la siguiente hipótesis:

$$H_0 : \kappa = 0$$

$$H_1 : \kappa \neq 0$$

Índice Kappa

- Con las hipótesis mencionadas hasta el momento se puede obtener la distribución asintótica de κ , en el número de preguntas N : sea $\hat{\kappa}$ el estadístico que se obtiene de sustituir las probabilidades por las frecuencias observadas (proporciones en la diagonal, fila y columna). Entonces asintóticamente $\hat{\kappa}$ se distribuye $N(\mu, \sigma^2)$ donde:

$$\mu = \kappa$$
$$\sigma^2 = \frac{1}{N} \left(\frac{\pi_0(1 - \pi_0)}{(1 - \pi_e)^2} + a + b \right)$$

y a y b son constantes.

Índice Kappa

- Ahora, en el caso específico de la prueba de hipótesis mencionada arriba, $\hat{\kappa}$ se distribuye asintóticamente $N(0, \sigma^2)$.
- En el análisis realizado sobre todas las pruebas del Estado se utilizó un nivel de confianza del 99.9%.

Índice Kappa

- Un supuesto clave en la derivación de esta prueba es que se supone que las probabilidades con las que c y s escogen una respuesta no dependen de la pregunta en consideración. Éste se usa en dos instancias.
 - 1 Para aplicar el teorema Central del Límite (algo que no constituye una verdadera limitación ya que existen otras versiones del teorema Central del Límite que son aplicables a esta situación).
 - 2 Para calcular las probabilidades π_{v+} y π_{+v} que son sólo válidas con la hipótesis de que la probabilidad de las respuestas no depende de la pregunta.
- Esta última hipótesis es muy fuerte y por esta razón se propone hacer una recodificación de las variables.

Recodificación de Variables

- La idea de la recodificación se basa en la observación de que una permutación de las opciones de las preguntas no afectaría la suma de la diagonal de la matriz mencionada anteriormente, que es donde se registran las respuestas coincidentes.
- Utilizando esta idea, el objetivo es recodificar las respuestas de tal forma que la probabilidad conjunta de respuesta sea igual para todas las preguntas.
- Primero se calcula la distribución empírica de las respuestas a cada pregunta y se ordenan las respuestas a cada pregunta según su frecuencia.

Recodificación de Variables

- Además se fuerza que la respuesta 1 corresponda a la correcta. Las demás respuestas siguen el ordenamiento empírico entre las que no son correctas.
- Obsérvese que la recodificación no iguala la distribución de las respuestas entre preguntas pero reduce sus diferencias manteniendo la consistencia con los datos observados.
- Ahora, la diferencia de la distribución entre preguntas puede depender de la habilidad de los individuos. Por esta razón en la tercera versión de índice Kappa se condiciona la distribución empírica de las respuestas a la habilidad medida como el número de respuestas correctas.

Recodificación de Variables Condicional

- Esta es la versión del índice Kappa que se ha implementado como un algoritmo de detección de copia.
- De esta forma para cada número posible de respuestas correctas y para cada pregunta se calcula una distribución empírica de las respuestas.

Pruebas Independientes

- Si se toman parejas compuestas por individuos que presentan pruebas en salones distintos, es de esperarse que el índice sospeche de copia en uno de cada mil casos, si las suposiciones de construcción del modelo se cumplen en la práctica.
- Para verificar esto, se construyen 50,000 parejas de forma aleatoria e independiente (sobre distintos salones) para cada forma.

Pruebas Independientes

Cuadro: Casos de copia por mil parejas independientes, para las 36 formas analizadas.

Forma	Tasa de copia	Forma	Tasa de copia	Forma	Tasa de copia
2	3.208333	106	1.041667	130	4.775
3	3.283333	108	2.141667	131	5.266667
4	5.9	120	5.233333	132	3.583333
8	3.258333	121	5.333333	133	3.625
10	3.075	122	5.483333	134	3.383333
41	4.15	123	3.033333	135	3.125
42	5.791667	124	2.691667	136	3.316667
43	2.341667	125	2.883333	137	3.241667
44	1.016667	126	3.225	138	3.266667
51	16.533333	127	2.9	139	3.541667
100	0.833333	128	2.883333	140	3.141667
101	1.216667	129	6.166667	149	19.575

Pruebas Independientes

- Los resultados sugieren una fuerte diferencia entre las probabilidades de respuesta en las preguntas, que se suponían idénticas para la definición del índice.
- En efecto, los patrones de respuesta varían bastante con las preguntas, como se puede observar en las siguientes figuras, que muestran para tres formas distintas las probabilidades de respuesta usadas para cuatro preguntas. Para cada prueba se escogió un nivel de habilidad arbitrariamente.

Pruebas Independientes

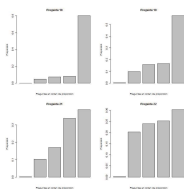


Figura: Probabilidades de respuesta, Forma 2, Habilidad 22, preguntas 19-22.

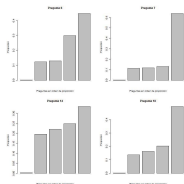


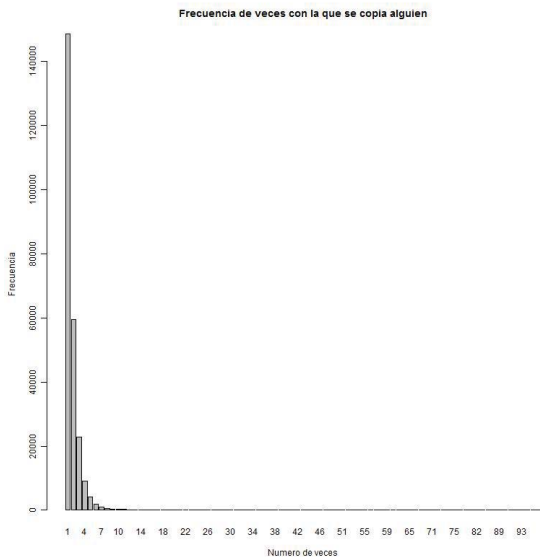
Figura: Probabilidades de respuesta, Forma 3, Habilidad 6, preguntas 6, 7, 14, 18.

Pruebas Independientes

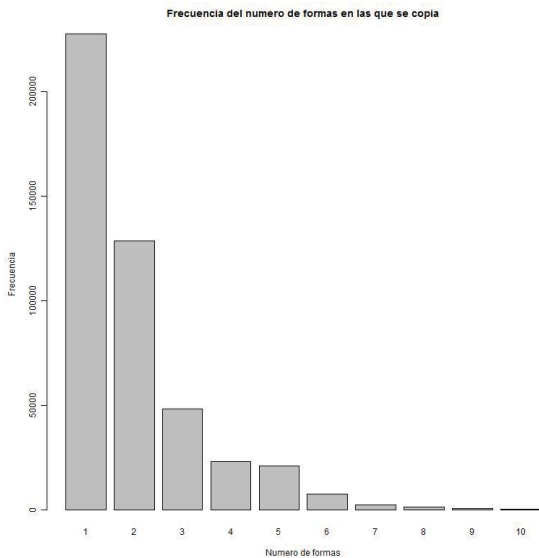
- Estas observaciones sugieren explorar direcciones para subsanar el impacto que las suposiciones de las distribuciones de las preguntas tienen sobre la detección de casos potenciales de copia.
- Una posibilidad que no requiere alterar el esquema de trabajo es corregir la media del estadístico del índice con la media encontrada para pruebas independientes.

- Se consideran 36 formas como exámenes separados. Es decir, si un área toma varias formas, cada una se considera como un examen independiente. Así, no se analizan casos de copia de dos personas en una misma área, si tenían formas diferentes.
- El porcentaje de parejas sospechosas de copia es de 2.94 %. Este valor corresponde a las parejas ordenadas que son detectadas como sospechosas.
- Considerablemente mayor, es el porcentaje de salones en donde se sospecha hubo copia. Este es 96.94 %.

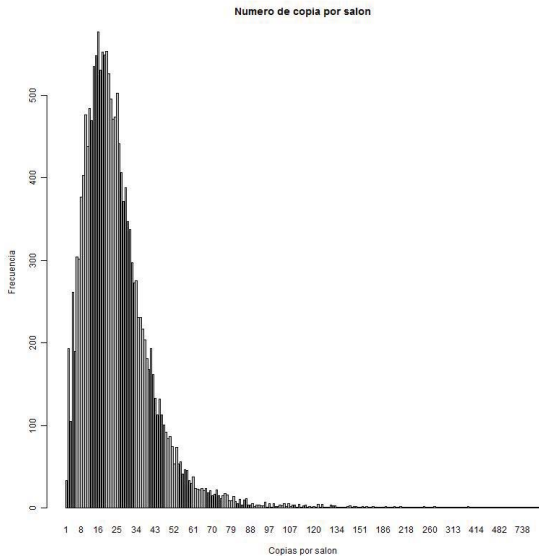
Número de veces que copia una persona (variando por salón, forma y fuente de copia)



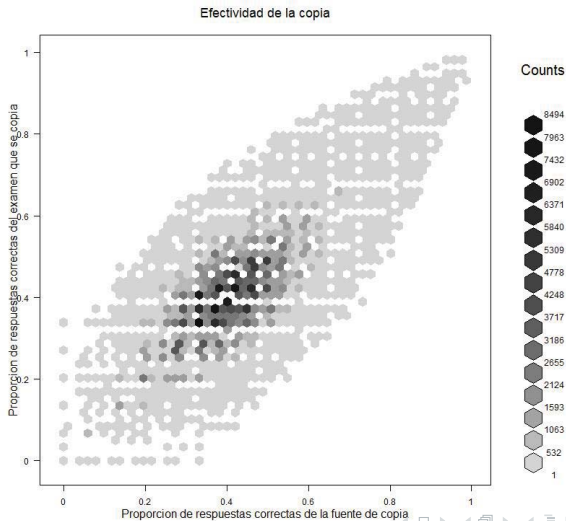
Número de formas en las que se detecta un sospechoso de copia.



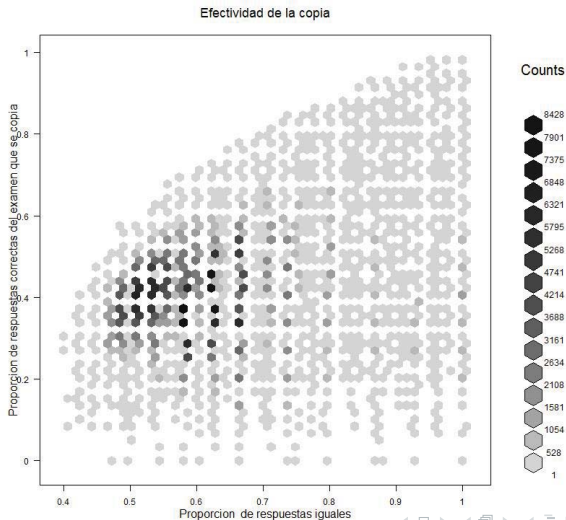
Número de parejas (ordenadas) detectadas como sospechosas en un salón.



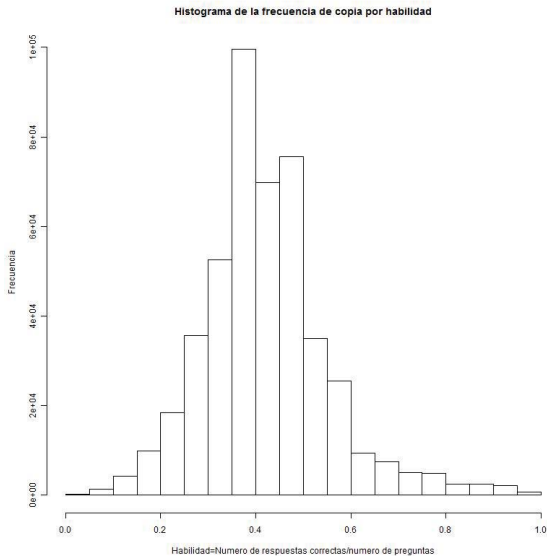
Relación del número de respuestas correctas entre individuo que copia y fuente, tomando las parejas sospechosas de copia



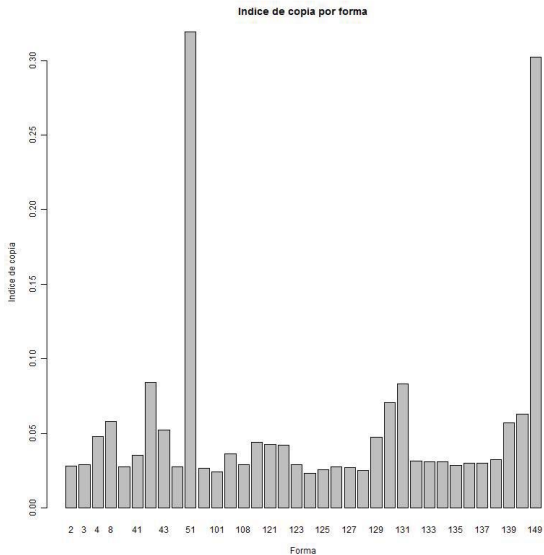
Relación entre número de respuestas correctas del sospechoso de copia y número de respuestas iguales con la fuente.



Distribución de la habilidad entre los sospechosos de copia.



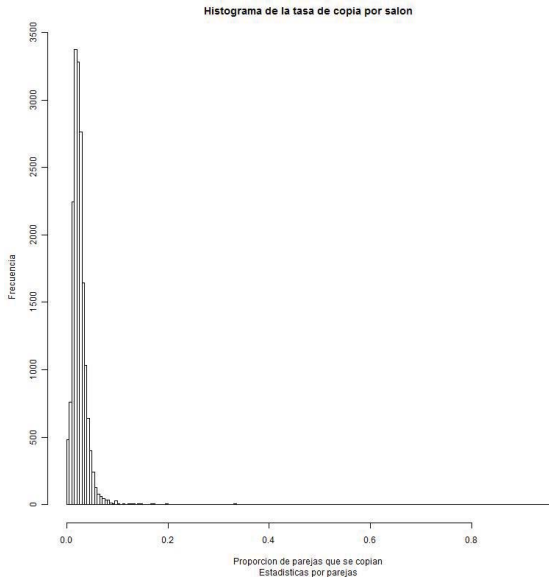
Distribución de la habilidad entre los sospechosos de copia.



Cuadro: Proporción de estudiantes sospechosos de copia en cada forma.

Forma	Tasa de copia	Forma	Tasa de copia	Forma	Tasa de copia
2	0.028456	106	0.036627	130	0.070667
3	0.029356	108	0.029161	131	0.083363
4	0.048014	120	0.044204	132	0.031487
8	0.058311	121	0.042585	133	0.030958
10	0.027542	122	0.042462	134	0.030931
41	0.035594	123	0.029392	135	0.028689
42	0.084407	124	0.023428	136	0.030233
43	0.052603	125	0.025745	137	0.030009
44	0.027599	126	0.02779	138	0.032762
51	0.319311	127	0.027268	139	0.057442
100	0.026786	128	0.025485	140	0.063148
101	0.024447	129	0.047734	149	0.302057

Proporción de parejas que se copian en cada forma.



Resultados

- La mayor proporción de estudiantes sospechosos de copia se encuentra en la habilidad media.
- La copia no es muy efectiva. La copia no siempre lleva a respuestas correctas.
- Las formas con menos seguridad (149 y 51) son las que tienen una mayor proporción de copia.

Datos Curiosos

- Hay casos de aparente copia masiva, en las que muchos estudiantes coinciden en todas las respuestas de una forma, en un mismo salón. Por ejemplo, hay casos de 17 estudiantes de un mismo salón con respuestas idénticas para la forma 100 (6 bien de 15); 15 en la forma 43 (6 bien de 15); 13 en las formas 42 (15 bien de 24), 43 (6 bien de 15) y 100 (6 bien de 15).
- Hay un caso de cuatro personas con las mismas respuestas en tres formas. Asimismo, se observan dos casos de tres personas con respuestas idénticas en cuatro formas.

Datos Curiosos

- Un grupo de 13 personas (del salón 437 788) tuvieron respuestas idénticas en dos formas (la 43 y la 100). Adicionalmente, hubo otras cuatro personas con idénticas respuestas en la forma 100, y otras dos en la forma 43. El número de respuestas correctas en las dos formas fue de 6 entre 15 preguntas.
- Hay 14 personas involucradas en 30 casos de respuestas idénticas. 13 de éstas forman parte del caso masivo mencionado en el punto anterior. 5
- Hay más de 35 parejas que tuvieron respuestas idénticas en 4 formas, nueve con respuestas idénticas en 5 formas, y hubo incluso dos con respuestas idénticas en 6 formas.

- Angoff, W. (1974), 'The development of statistical indices for detecting cheaters', *Journal of the American Statistical Association* .
- Baird, L.S., J. (1980), 'Current trends in college cheating', *Psychology in the schools* .
- Belleza, F. & Belleza, S. (1989), 'Detection of cheating on multiple-choice test by using error-similarity analysis', *Teaching of Psychology* .
- Cohen, J. (1960), 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement* pp. 37–46.
- Frary, B., Tideman, N. & Watts, T. (2009), 'Indices of cheating on multiple choice tests'.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*.
- Holland, P. (1996), 'Assesing unusual agreement between the incorrect answers of two examinees using the k index: statistical theory and empirical support', *ETS technical report* .