

Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice

VOLUME 49

2020

NUMBER 2

Contents

Addendum to the 2011 Meta-analytic Survey – the Utah Four-Question Test (“Raskin Technique”) / ESS Editorial Staff	73
A Brief Comment on the Inhalation/Exhalation Ratios in Polygraph Scoring Donald J. Krapohl	79
Deception Detection through Text Analysis Aiman Shariff and Ritika Motwani	82
Electrodermal Responses: When is Bigger Really Better? Donald J. Krapohl	104
Bigger is Better for Automated Scoring: Analysis of Minimum Constraints for RQ/CQ Ratios Raymond Nelson	110
Using Virtual Reality to Improve Memory Recall and Detection of Deception in Forensic Interviews Joyce Yan Ting Sam, Lin Qiu, and Ky Phong Mai	121
Accuracy Effects for ESS and Three-Position Scores of Federal ZCT Exams Using the Grand Total Rule with Traditional/Federal and Multinomial Cutscores Raymond Nelson	157
Modification of the AFMGQT to Accommodate Single-Issue Screening: The British One-issue Screening Test Donald J. Krapohl, Donald Grubin, Tim Benson and Bernard Morris	176
Timely Non-deceptive Sexual History Polygraph Examinations are Correlated with Completion of Treatment but Not Correlated with Sexual Recidivism James E. Konopasek and Johneen Manno	184

Puntuación Automatizada para Más Grande es Mejor: Análisis de las Restricciones Mínimas de las Proporciones RQ/CQ

Raymond Nelson

Abstract

Se utilizó una muestra de archivo de $n=300$ exámenes poligráficos confirmados de campo, para estudiar los efectos de las proporciones mínimas de restricción, desde 1:1 y hasta 2:1 en incrementos de .05, para la extracción automática de características y para la asignación automática de puntuaciones. Para los datos respiratorios, el 95% de las puntuaciones fueron cero (0) con una proporción de restricción mínima de 1.6:1. Por el contrario, aproximadamente el 55% de las puntuaciones del EDA fueron distintas a cero y el 39% de las puntuaciones del cardio fueron distintas a cero con la misma proporción (1.6:1). Las proporciones LogRC fueron óptimas sin una restricción mínima, lo que indica que los métodos de puntuación automatizados son razonables cuando intentan hacer uso de cualquier diferencia medible que pueda extraerse entre las preguntas relevantes y de comparación. Para las puntuaciones enteras con signo, el coeficiente de correlación (similar al DEC) de las 2700 puntuaciones numéricas, no se vio fuertemente afectado por alguna restricción de los datos de cardio. El coeficiente de correlación para las puntuaciones numéricas de los datos del EDA se vio mínimamente afectado por la serie de restricciones, comenzando en .425 con 1.05:1, luego aumenta ligeramente a .450 con la proporción de 1.2:1 y termina en .385 con la proporción de restricción máxima de 2:1. Las correlaciones de puntuación de la respiración, junto con la correlación de la puntuación agregada (mostradas en naranja), sugieren que restringir la extracción de las puntuaciones de la respiración al rango entre 1.2:1 y 1.6:1 puede optimizar de manera útil la contribución de las puntuaciones de la respiración para lograr conclusiones correctas vs conclusiones incorrectas. Los datos de este análisis indican que no se necesita una proporción de restricción mínima para los métodos de análisis automatizados del EDA o cardio y proporcionan un soporte general a la validez de la regla "más grande es mejor".

Introducción

Todas las conclusiones científicas, tanto en la investigación como en las pruebas científicas, se realizan tomando en cuenta otras conclusiones posibles. El proceso científico tiene como objetivo evaluar la fuerza de la evidencia disponible, que apoya cada una de las diferentes conclusiones posibles que intentan responder preguntas básicas sobre el universo y la realidad. ¿Qué es? ¿Cómo funciona? ¿Por qué? Independientemente de si es ciencia al nivel de la física teórica o al nivel de decisiones prácticas forenses y de decisiones sobre el manejo de riesgos acerca de la mejor manera de proceder con un solo individuo, la reproducibilidad de



los resultados analíticos se ha convertido en un estándar o expectativa de facto en todos los ámbitos de la investigación y de las pruebas científicas (Peng, 2011). El requerimiento de la reproducibilidad puede observarse fácilmente en las pruebas para la evaluación de la credibilidad - comenzando por el hecho de que los datos de las pruebas se registran de forma permanente. La habilidad para registrar datos es fundamental para poder estudiar las señales de diferentes formas, y para que se puedan optimizar los métodos de análisis.

Los avances tecnológicos durante la historia temprana de la profesión poligráfica requirieron tanto del desarrollo de sensores que pudieran brindar acceso a las señales fisiológicas que están correlacionadas con el engaño y la veracidad, como de métodos para registrar los cambios fisiológicos para que pudieran ser estudiados más cuidadosa y repetidamente. Hoy sabemos que, aunque el engaño en sí mismo no puede medirse físicamente, es posible que nunca encontremos alguna actividad fisiológica que esté asociada exclusivamente con el engaño. Por el contrario, lo más probable es que todas las señales poligráficas sigan involucrando al sistema nervioso autónomo y a múltiples aspectos de la corteza cerebral. Y probablemente, toda la actividad fisiológica se mantendrá asociada con múltiples tipos de comportamiento humano. En resumen: no existe cosa tal como la "nariz de Pinocho".

La fuerza de la correlación de las diferentes señales fisiológicas, y el grado de covariación de las diferentes señales, seguirá siendo una preocupación subyacente para cualquier persona involucrada en la investigación de la validez del polígrafo (o en discusiones sobre la validez del polígrafo). Para que sean útiles, las señales fisiológicas idealmente deberán correlacionarse con el criterio a un nivel estadísticamente significativo, pero no covariarán tan fuertemente como para que se vuelvan redundantes. Las señales útiles contribuirán con información única, no redundante, con respecto a un modelo estructural (es decir, la representación matemática/estadística del fenómeno de interés). La redundancia de la señal fisiológica puede observarse cuando al agregar datos al modelo no se incrementa su eficacia, aunque se sepa que la información añadida está correlacionada con los fenómenos de interés. En otras palabras, no es cierto el adagio simplista de "más información es siempre mejor": más información es mejor cuando aumenta el tamaño del efecto de interés. Si la información añadida no aumenta el tamaño del efecto de interés, el efecto real será un aumento en el riesgo de confusión y falta de confiabilidad. El resultado de todo esto es que las pruebas científicas suelen construirse con señales cuya intensidad de correlación es moderada, porque las señales para las que la correlación de criterio es fuerte tenderán a covariar tanto que pueden llegar a ser redundantes.

Durante la primera parte del siglo XX el kimógrafo era la mejor tecnología disponible para poder registrar datos poligráficos para su posterior análisis y reanálisis. Los métodos de análisis hasta la mitad del siglo se basaban casi uniformemente en la intuición no cuantificable y en la experiencia del observador experto. Con el tiempo, hacia la segunda mitad del siglo XX, la necesidad de mejorar la consistencia y de desarrollar habilidades entre grupos de expertos condujo a poner énfasis en los sistemas de puntuación numérica, como lo es el sistema de siete posiciones y el método de puntuación de tres posiciones. Hacia la segunda mitad del siglo XX observamos un aumento exponencial en la disponibilidad de tecnología



informática. Una tecnología computacional potente y (relativamente) barata ha influido en casi todos los aspectos de la vida social y profesional - incluyendo al ocio, el entretenimiento, la comunicación, el transporte, la educación, la administración, el empleo, las noticias y la información, publicaciones, e incluso la ciencia y las pruebas científicas.

Hoy en día - bien entrado el siglo XXI - no hay ningún entorno social o de la ciencia que no haga uso de la tecnología informática para registrar y analizar datos. El Kimógrafo de principios del siglo XX ha sido virtualmente suplantado por completo en la práctica de campo del polígrafo y en la investigación poligráfica, por convertidores analógicos a digitales y sistemas de codificación computarizados que registran los datos del polígrafo no como un trazo de un cilindro o rollo de papel, sino como series de tiempo de números almacenados en un medio electrónico. Los datos se procesan para su visualización en una forma tradicional de trazos de series de tiempo en la pantalla de una computadora. Un aspecto conveniente de todo esto es que los examinadores de polígrafo más antiguos pueden imprimir sus "gráficos" en papel e inspeccionarlos visualmente de manera similar a lo que hacían décadas pasadas.

Mientras que los factores que influyeron en los trazos lineales en la época de los primeros instrumentos poligráficos eran totalmente mecánicos - ya que involucraba la masa en movimiento de un hardware cuidadosamente diseñado, incluyendo los coeficientes de fricción de los pivotes y los cojinetes, así como los innumerables ajustes y calibraciones necesarias para garantizar que los datos, codificados en forma de tinta sobre papel, fueran útiles - el procesamiento de la señal del polígrafo hoy en día es más cuidadoso y preciso gracias a los esfuerzos de los ingenieros eléctricos que entienden nuestros requisitos de hardware y a través de ingenieros de software y científicos de datos que pueden permitirnos hacer uso de los métodos de procesamiento de señales digitales y estadísticas más poderoso que los que utilizábamos durante la época en que todos los cálculos se hacían manualmente.

La ingeniería electrónica y el procesamiento digital de señales pueden proporcionar una mayor precisión y confiabilidad y mucha mayor comodidad y economía en comparación con las soluciones mecánicas del pasado. A diferencia de los sistemas poligráficos mecánicos, en los que el filtrado y regulación eran a veces un subproducto no intencional o imprevisto de la fricción del peso de la pluma de tinta sobre el papel, los sistemas poligráficos computarizados de hoy - con altas tasas de muestreo y alta resolución de conversión analógico-digital - pueden proporcionar datos de mayor fidelidad, en términos de registro y representación de la actividad fisiológica, como nunca en el pasado.

De manera simultánea a los avances en los métodos de evaluación poligráfica, de procesamiento de señales y de datos, los métodos de análisis de datos también han avanzado gracias la disponibilidad de tecnologías informáticas. Hoy en día, los profesionales de la poligrafía tienen acceso tanto a las distribuciones de referencia empíricas (Krapohl & McManus, 1999; Krapohl, 2002; Nelson, Krapohl & Handler, 2008; Nelson y Handler, 2015) como a distribuciones de referencia multinomiales (Nelson, 2017; 2018). La disponibilidad de modelos estadísticos de referencia basados en computadoras ha llevado a una aplicación potencialmente conveniente, tanto de los métodos estadísticos frecuentistas como de los



métodos estadísticos bayesianos en la práctica de campo.

En pleno siglo XXI, tenemos la capacidad de registrar las señales poligráficas de manera digital y de un uso potencial y conveniente de poderosos métodos matemáticos y estadísticos que pueden ir mucho más allá de lo que los profesionales de la poligrafía podrían estar dispuestos a intentar con lápiz y papel. También tenemos la capacidad de automatizar la extracción de características - que sería inherentemente más confiable que una extracción de características mediante métodos de reconocimiento de patrones visuales que puede haber sido la mejor solución disponible con los instrumentos poligráficos análogos. El engaño y la veracidad son problemas complejos - comenzando por la compleja asimetría de lograr una definición epistemológica/filosófica de lo que es el engaño y la veracidad. Tampoco es de extrañar que el análisis de los datos de las pruebas para la evaluación de la credibilidad sea intrínsecamente complejo - y que por tanto, esté sujeto a diferentes tipos de sesgo, subjetividad e inconsistencia.

El nivel de complejidad con respecto a la extracción de características poligráficas se hace muy aparente cuando se toma en cuenta la combinación o interacción de factores que pueden influir en una puntuación numérica, que incluye la extracción de características tanto de la pregunta relevante (RQ) como de la pregunta de comparación (CQ) al tiempo que se realiza una comparación entre estos dos valores. La solución más realista para el futuro de la profesión poligráfica será aprovechar el poder de los ordenadores digitales para el registro del proceso y para el análisis todos los tipos de complejidades e interacciones, que incluyen la extracción automática de las características y de la asignación de puntajes. Hacer lo contrario - limitar la metodología poligráfica a los métodos de mediados del siglo pasado, previa a la informática- eventualmente provocará un perjuicio. Afortunadamente existe una gran cantidad de conocimientos y metodologías para el logro de esta meta. Hoy en día, las poderosas herramientas informáticas y de análisis son abundantes y baratas - a menudo son gratuitas y de código abierto. Este proyecto es un estudio de optimización de la asignación automatizada de la puntuación numérica en función de la proporción de los pares de RQ y CQ. El tema de interés es si existe un conjunto de restricciones de proporción mínima entre RQ/CQ que maximice la información de diagnóstico obtenida en las puntuaciones numéricas para cada uno de los sensores de registro poligráfico.

Métodos

Datos

Los datos para este proyecto fueron $n=300$ exámenes confirmados de campo, que se llevaron a cabo utilizando el formato Federal Zone Comparison Test (FZCT, Department of Defense, 2006). Los casos de la muestra fueron llevados a cabo por diversas agencias policiales federales, estatales y municipales y fueron posteriormente incluidos en el archivo de casos confirmados del Department of Defense Polygraph Institute (llamado ahora National Center for Credibility Assessment) (*en español, Instituto Poligráfico del Departamento de Defensa, ahora el Centro Nacional para la Evaluación de la Credibilidad*). Todos los casos constaron de tres iteraciones de una secuencia de preguntas que contenía tres preguntas relevantes



(RQs) y tres preguntas de comparación (CQs), además de otras preguntas de procedimiento que no están sujetas a un análisis numérico o estadístico. Todos los exámenes tuvieron tres gráficas completas de prueba. [Consulte Nelson (2015) y Department of Defense (2006) para obtener información general sobre esta prueba con preguntas de comparación y de cómo se realizaron los casos de la muestra].

Todos los casos de la muestra se aplicaron con el conjunto estándar de sensores, que incluyen sensores de respiración superior e inferior, un sensor de actividad electrodérmica y un sensor de actividad cardiovascular de los cuales se obtuvieron las respuestas y se asignaron las puntuaciones numéricas. Esta muestra se utilizó previamente para el desarrollo del método de puntuación OSS-2 (Krapohl, 2002), y en su momento se extrajeron las características de respuesta de los datos registrados utilizando un programa de software (Extract.exe, Harris, en Krapohl & McManus, 1999) que se desarrolló para extraer objetivamente las mediciones de las características de Kircher en los datos respiratorios, EDA y cardio de los datos poligráficos computarizados. Se contó con un total de 21600 mediciones para los $n=300$ casos de campo con tres iteraciones de una secuencia de preguntas que incluía tres RQs y tres CQs. Los datos se importaron a Language R para su análisis estadístico informático (R Core Team, 2019).

Análisis

Se evaluaron todas las iteraciones de todas las preguntas relevantes (RQs) utilizando la pregunta de comparación seleccionada por el procedimiento estandarizado en el formato FZCT. Las RQ están etiquetadas como R5, R7 y R10, mientras que las preguntas de comparación (CQ) como C4, C6 y C9. Para cada sensor, la primera RQ, R5, se evaluó contra las CQs ubicadas inmediatamente antes y después de esta RQ - C4 o C6 del primer gráfico registrado, aunque las preguntas pueden ser rotadas para los gráficos posteriores - dependiendo de qué CQ produjo el mayor cambio en la actividad fisiológica. La segunda y tercer RQ, R7, y R10 fueron evaluadas contra la CQ precedente - C6 para R7 y C9 para R10, aunque el orden puede ser rotado en algunas de las gráficas de prueba registradas. Para cada par de preguntas se calculó una proporción RQ/CQ, denominada Proporción RC. Para los sensores EDA y cardio, cuando los valores extraídos son más grandes, indican mayores cambios fisiológicos. Por el contrario, para el sensor respiratorio, los valores extraídos más pequeños representan mayores cambios en la actividad fisiológica.

Las Proporciones RC se ajustaron a una distribución asimétrica, delimitada entre 0 y ∞ (infinito) con una media de 1 y un rango potencialmente infinito de valores entre 0 y 1, junto con un rango potencialmente infinito de valores entre 1 e infinito.



Cuando el valor de la RQ fue mayor que el valor de la CQ, la Proporción RC fue un valor entre 1 e infinito. Cuando el valor de la CQ fue mayor que el valor RQ, la Proporción RC fue un valor decimal entre 0 y 1. Para evitar esta asimetría, se tomó el logaritmo natural para cada Proporción RC, denominada Proporción logRC. La distribución resultante de las Proporciones logRC fue una distribución simétrica con una media de 0 y un número infinito de valores potenciales entre 0 y ∞ (infinito) junto con un número infinito de valores potenciales entre 0 y $-\infty$ (infinito negativo). Las Proporciones RC entre 0 y 1 produjeron valores negativos de Proporciones logRC entre 1 e infinito negativo, mientras que las Proporciones RC entre 1 e infinito produjeron Proporciones logRC positivos entre 0 e infinito. El cuadro 1 muestra un ejemplo del uso del logaritmo natural para producir proporciones que son simétricas en torno a 0.

Tabla 1. Ejemplos que muestran el uso del logaritmo natural para lograr una distribución simétrica de las proporciones logRC				
	Valor RQ	Valor CQ	Proporción RC	Proporción logRC
Ex 1	300	200	1.5	0.4054651
Ex 2	200	300	.67	-0.4054651

Obsérvese en la Tabla 1, cómo las Proporciones RC no son simétricas en torno a 1, mientras que las Proporciones logRC son simétricas en torno a 0. Esta simetría permite utilizar cálculos estadísticos lineales como el coeficiente de correlación. Antes de seguir adelante, fue necesario ajustar los valores del signo de las Proporciones logRC para los datos de EDA y cardio, de modo que las Proporciones logRC negativas correspondieran a puntuaciones de engaño, al tiempo que las Proporciones logRC positivas correspondieran a puntuaciones de veracidad para todos los sensores, incluyendo los de la respiración, EDA y cardio.

Se tomaron veintiún mil seiscientas (2100) mediciones de los $n=300$ casos, a partir de las cuales se calcularon un total de 10800 Proporciones logRC para las tres iteraciones de las tres RQs, para los sensores de respiración torácica y abdominal, el sensor EDA y el sensor cardio para cada uno de los $n=300$ casos de la muestra. Tras combinar los datos de los sensores de respiración torácica y abdominal, se obtuvieron 8100 Proporciones RQs, incluyendo 2700 valores para cada sensor de registro: respiración, EDA y cardio. Para mantener la coherencia con la intuición tradicional de las puntuaciones enteras utilizadas en la práctica poligráfica de campo, las Proporciones logRC de valor con signo + correspondieron a la veracidad, mientras que las puntuaciones enteras de valor de signo - correspondieron al engaño. Durante el análisis, se asignaron puntuaciones enteras con signo automatizadas a las 10800 proporciones logRC utilizando la regla bigger-is-better (*más grande es mejor*) (BIBR: National Center for Credibility Assessment, 2017). Las puntuaciones numéricas de este tipo son similares a las que asignarían los expertos humanos utilizando métodos de puntuación manual como el método de puntuación federal de tres posiciones (National Center for Credibility Assessment, 2017) o el Sistema de Puntuación Empírica (ESS: Nelson, Krapohl & Handler, 2008).



Datos respiratorios

Las Proporciones logRC torácicas y abdominales se combinaron en un único vector de 5400 valores y la correlación de punto-biserial para las puntuaciones de la respiración fue de $r_{pb}=.184$. Solo para el sensor de respiración torácica el valor fue $r_{pb}=.209$, y para el sensor abdominal fue $r_{pb}=.161$. Para los datos combinados de los sensores respiratorios, la Proporción logRC máxima fue de 3.8 (una proporción de 45:1). Se aplicó iterativamente un valor máximo de restricción desde +/-2 hasta +/0 3.5 y se encontró que se optimizaban las correlaciones de punto-biserial en un +/-2.7 (una proporción de 14.9). Es decir, las Proporciones logRC se forzaron a cero (0) si se superaban los valores de 2.7 o -2.7. Las Proporciones LogRC para los datos respiratorios contribuían más con puntuaciones incorrectas que correctas cuando superaban este nivel. Veinticuatro (24) de las 5400 puntuaciones respiratorias (<0.5%) superaron la restricción máxima. Con el valor máximo de restricción, las correlaciones de punto-biserial fueron $r_{pb}=.216$ para el sensor torácico y $r_{pb}=.182$ para el sensor abdominal. Una restricción externa o máxima puede mejorar los coeficientes de correlación para las puntuaciones agregadas de la respiración. Sin embargo, como el objetivo de este proyecto era estudiar las restricciones óptimas mínimas, no se realizó ninguna optimización adicional para la restricción externa de las puntuaciones respiratorias. La proporción de restricción máxima se mantuvo para el resto del análisis.

Entonces, las puntuaciones respiratorias para los sensores torácicos y abdominales se combinaron en un único conjunto de 2700 puntuaciones utilizando el procedimiento descrito por Nelson y Krapohl (2017). Utilizando ese procedimiento la Proporción logRC combinada se presionó a cero (0) cuando los valores de signo eran opuestos, y se fijó en el valor con el mayor valor absoluto si no eran opuestos. Después de combinar las dos Proporciones logRC a una sola puntuación respiratoria para cada iteración de cada RQ para cada caso, la correlación punto-biserial fue $r_{pb}=.211$. Los datos respiratorios también se combinaron promediando las proporciones logRC de los dos sensores de respiración. La correlación punto-biserial para el método donde se promedió la combinación de los datos en los sensores fue de $r_{pb}=.215$, y superó la del método de procedimiento. Para el vector de respiración combinada, cinco (5) de las 2700 proporciones logRC de respiración fueron cero (0). Los vectores separados de proporción logRC fueron retenidos para su análisis, incluyendo 2700 valores para el sensor torácico y 2700 valores para los sensores abdominales. La información torácica y abdominal se podría combinar en un paso posterior para cada iteración de una serie de restricciones de proporción mínima.

Las proporciones logRC se estandarizaron a una media de cero (0) y una desviación estándar de uno (1). No hubo diferencias en la correlación ($r_{pb}=.215$) para las Proporciones logRC estandarizadas para para los sensores de respiración torácica y abdominal promediados. Dado que la estandarización no ofrecía ninguna ventaja, el resto del análisis se completó con los datos respiratorios no estandarizados. Los valores estandarizados tienen una métrica común, con media=0 y ds=1, y pueden calcularse posteriormente cuando se combinan datos de



diferentes sensores.

Las Proporciones LogRC para los datos de respiración combinados se agregaron promediando las tres iteraciones de las tres RQs para cada caso. La correlación de punto-biserial para la media de la Proporción logRC para los datos respiratorios fue $r_{pb}=.408$. Para los sensores individuales, la correlación punto-biserial después de agregar las puntuaciones de cada caso fue $r_{pb}=.420$ para el sensor torácico y $r_{pb}=.359$ para el sensor abdominal con la restricción más alta. Sin la restricción máxima las correlaciones fueron $r_{pb}=.401$ para el sensor torácico y $r_{pb}=.317$ para el abdominal. No fue sorpresa que las Proporciones logRC agregadas para cada caso produjeran un coeficiente de correlación más fuerte que las Proporciones logRC de la presentación de cada RQ. La agregación de datos tiene el efecto de mejorar la relación señal-ruido dentro de la información extraída en los datos registrados.

Se evaluó una serie de restricciones mínimas desde 1.05:1 hasta 2:1 en incrementos de 0.05. Las proporciones de 1.05 a 2 se transformaron a sus logaritmos naturales para que pudieran aplicarse simétricamente a las Proporciones logRC que tienen valores similares con signo + y - a la intuición tradicional que los examinadores poligráficos de campo utilizan para los resultados numéricos de veracidad y engaño. La serie de restricciones posibles se aplicó iterativamente a la muestra de $n=300$ casos. Para cada iteración de la serie de restricciones mínimas, las Proporciones logRC se presionaron a cero si no superaban la restricción. Las puntuaciones enteras utilizando el BIBR también se presionaron a cero para los valores para en los que el logRC se presionó a cero. El valor de la restricción se aplicó por separado a los datos de los sensores torácicos y abdominales antes de combinar los datos de los sensores utilizando el método descrito por Nelson y Krapohl (2017). Las Proporciones LogRC se agregaron promediando todas las iteraciones de todas las RC para cada caso, y la correlación de punto-biserial se calculó entonces para los estados de criterio de caso codificados como [+1, -1].

Las Proporciones LogRC para cada presentación de cada RQ se transformaron a puntuaciones enteras con signo en las que las puntuaciones torácicas y abdominales se combinaron con el método descrito anteriormente. Los valores de los signos tanto de los LogRC como de las puntuaciones enteras con signo se ajustaron a la intuición tradicional para las puntuaciones de signo usadas por los examinadores poligráficos de campo. Las puntuaciones de valor positivo (+) corresponden a la veracidad, y las de valor con signo negativo (-) corresponden al engaño. Para cada iteración de la serie de proporciones de restricción mínima se calculó la Proporción logRC diferente a cero - también la proporción de los puntajes enteros con signo diferentes a cero - para las 2700 puntuaciones después de combinar los datos de los sensores torácicos y abdominales. La proporción de puntuaciones correctas enteras con signo, diferentes a cero, se calcularon comparando las puntuaciones enteras contra el estado del criterio de caso codificado como [+ , -]. Por último, las puntuaciones enteras con signo se sumaron para cada caso y se calculó la correlación de punto-biserial para las puntuaciones numéricas contra el estado del criterio, codificado como [+1, -1] para cada iteración de la proporción de restricción mínima.



Para cada iteración de la proporción de restricción mínima, la correlación de las 2700 puntuaciones enteras con signo contra el estado del criterio del caso, se calculó mediante un procedimiento similar al del coeficiente de eficiencia de detección (DEC; Kircher, Horowitz, & Raskin, 1988). El DEC se calcula mediante la correlación de Pearson entre los códigos de puntuación [+1, 0, -1] y el estado del criterio [+1, -1], y son informativos porque representan la fuerza de la información con respecto a los resultados correctos, incorrectos e inconclusos en una sola estadística de correlación. Esta aplicación del DEC difiere de su uso normal en que el DEC se utilizó para clasificaciones realizadas con cortes numéricos o estadísticos, agregando puntuaciones para cada caso utilizando el conjunto completo de sensores de registro; aquí se utilizó con las puntuaciones individuales de un solo sensor y sin puntuaciones numéricas. En este uso, la correlación DEC puede pensarse como una correlación de puntuación numérica; proporciona una medición de la fuerza de la información de las puntuaciones numéricas en cada proporción de restricción mínima.

Datos EDA

Para las puntuaciones del EDA, la correlación de punto-biserial para las 2700 Proporciones logRC fue $r_{pb}=.433$. La Proporción logRC máxima fue de 3.4, que corresponde a una relación de 30:1 en la que el valor de la CQ superó el de la RQ. La Proporción logRC mínima fue de -4,7, que corresponde a una proporción de 110:1, en la que el valor de la RQ superaba al de la CQ.

Se aplicó iterativamente una relación de restricción máxima de +/-2 a +/-20 y se encontró que maximizaba la correlación de punto-biserial con una proporción de restricción máxima de 7:1 con $r_{pb}=.439$. Las proporciones LogRC de EDA que excedieron este nivel contribuyeron con más puntuaciones incorrectas que correctas. El resto del análisis se completó con esta relación de restricción máxima. Noventa (90) de las 2700 Proporciones LogRC (3,3%) superaron este valor de restricción.

Entonces se estandarizaron las proporciones LogRC para evaluar su efecto sobre los coeficientes de correlación. La estandarización de las Proporciones LogRC no cambió las correlaciones de punto-biserial. Debido a que este proyecto no incluyó la sumatoria de datos entre sensores, el resto del análisis se completó con las Proporciones LogRC no estandarizadas. La estandarización proporciona ventajas y realizarse posteriormente.

Después de agregar las Proporciones logRC para todas las iteraciones de todas las RQs para cada uno de los $n=300$ casos de la muestra, la correlación de punto-biserial fue $r_{pb}=.751$. Agregar los datos tiene el efecto de mejorar la relación señal/ruido, y por lo tanto, no es sorprendente que la correlación de las Proporciones logRC agregadas supere la correlación de las Proporciones logRC



individuales.

Se evaluó una serie de proporciones de restricción mínimas desde 1.05:1 y hasta 2:1 en incrementos de 0.05. Los resultados de los datos del EDA se volvieron a calcular para cada iteración de la restricción, incluyendo la correlación de punto-biserial de las Proporciones logRC agregadas, contra el estado del criterio de cada uno de los $n=300$ casos de muestra. Para cada iteración de la serie de restricciones mínimas, las Proporciones logRC se presionaron a cero si no superaban la restricción. Las 2700 Proporciones logRC también se transformaron en puntuaciones enteras con signo [+1, 0, -1] y se calculó la correlación del criterio de las puntuaciones con signo para cada iteración de la serie de restricciones mínimas. También se calcularon los resultados para la proporción de puntuaciones de signos correctas y de la proporción de puntuaciones distintas de cero. Finalmente se sumaron las puntuaciones enteras con signo para cada uno de los $n=300$ casos y se calculó la correlación de las puntuaciones enteras agregadas contra el estado del criterio del caso.

Datos cardiovasculares

Para las puntuaciones del cardio, la correlación de punto-biserial para las 2700 Proporciones logRC fue de $r_{pb}=.179$. La Proporción logRC máxima fue de 1.9, lo que corresponde a una Proporción de 6.7:1. La Proporción logRC mínima fue de -3.9, lo que corresponde a una relación de 52:1.

Se aplicó iterativamente una relación de restricción máxima a los datos de cardio de +/-2 a +/-20 y se encontró que se optimizaba la correlación de punto-biserial con una relación de restricción máxima de 12:1 con un $r_{pb}=.180$. Las Proporciones LogRC que superaban este nivel de 12:1 eran más propensas a generar puntuaciones de cardio incorrectas que correctas. El resto del análisis se completó con esta relación de restricción máxima. Dos (2) de las 2700 Proporciones LogRC superaron el valor de restricción de 12:1.

Entonces se estandarizaron las Proporciones LogRC para evaluar el efecto sobre los coeficientes de correlación. La estandarización de las proporciones LogRC no cambió las correlaciones de punto-biserial. Debido a que este proyecto no incluyó la sumatoria de datos entre sensores, el análisis posterior de los datos de cardio se completó con las Proporciones LogRC no estandarizadas. La estandarización de los datos de cardio se puede realizar más adelante, cuando se combinen los datos de todo el conjunto de sensores de registro.

Después de agregar Las Proporciones logRC para todas las iteraciones de todas las RQs para cada uno de los $n=300$ casos de muestra, la correlación de punto-biserial fue de $r_{pb}=.460$. La agregación de los datos cardiovasculares mejora la



relación señal/ruido y, por lo tanto, no es sorprendente que la correlación de las Proporciones logRC agregadas supere la correlación de las Proporciones logRC individuales.

A los datos de cardio se les aplicó la misma serie de proporciones de restricción mínima de 1.05:1 a 2:1 con incrementos de 0.05. Los resultados se volvieron a calcular para cada iteración de la restricción, incluyendo la correlación de punto-biserial de las Proporciones logRC agregadas contra el estado del criterio de cada uno de los n=300 casos de la muestra. Para cada iteración de la serie de restricciones mínimas, las Proporciones logRC se presionaron a cero si no superaban la restricción. Las 2700 Proporciones logRC también se transformaron en puntuaciones enteras con signo [+1, 0, -1] y se calculó la correlación del criterio con las puntuaciones con signo. También se calcularon los resultados de la proporción de puntuaciones con signo correctas y la proporción de puntuaciones distintas a cero para los datos de cardio. Por último, se sumaron las puntuaciones enteras con signo para cada uno de los n=300 casos y se calculó la correlación de las puntuaciones enteras agregadas de cardio contra el estado de criterio de caso.

Resultados

La tabla 2 muestra los coeficientes de correlación media de punto-biserial para las 2700 proporciones logRC y del estado del criterio para cada sensor de registro, junto con el coeficiente de correlación media de punto-biserial para las Proporciones logRC agregadas y de las puntuaciones enteras sumadas. Se puede observar en la Tabla 2 que las correlaciones para las puntuaciones agregadas superan a las individuales. Este es un ejemplo del valor de usar formatos de pruebas poligráficas con múltiples RQs y múltiples iteraciones de la secuencia de preguntas.

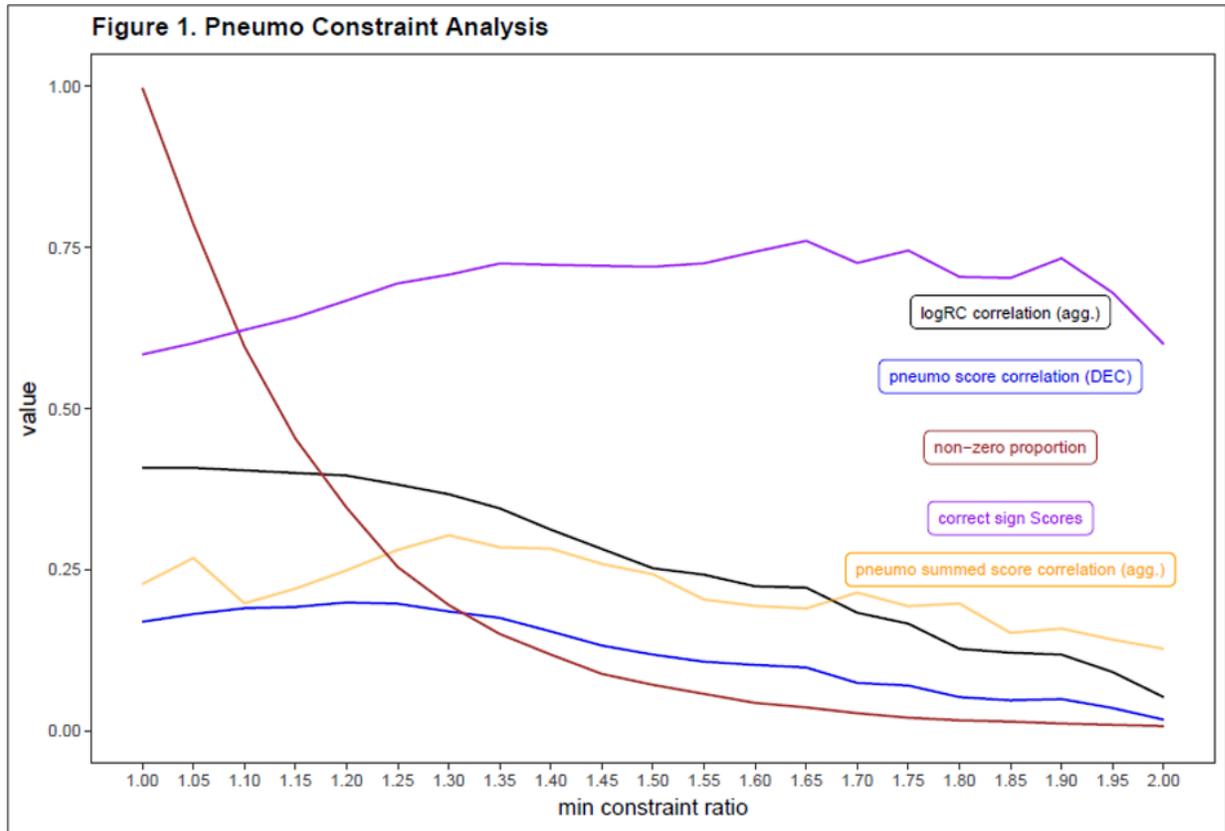
Tabla 2. Correlaciones punto-biserial para las proporciones logRC, logRC agregado y de las puntuaciones enteras sumadas.

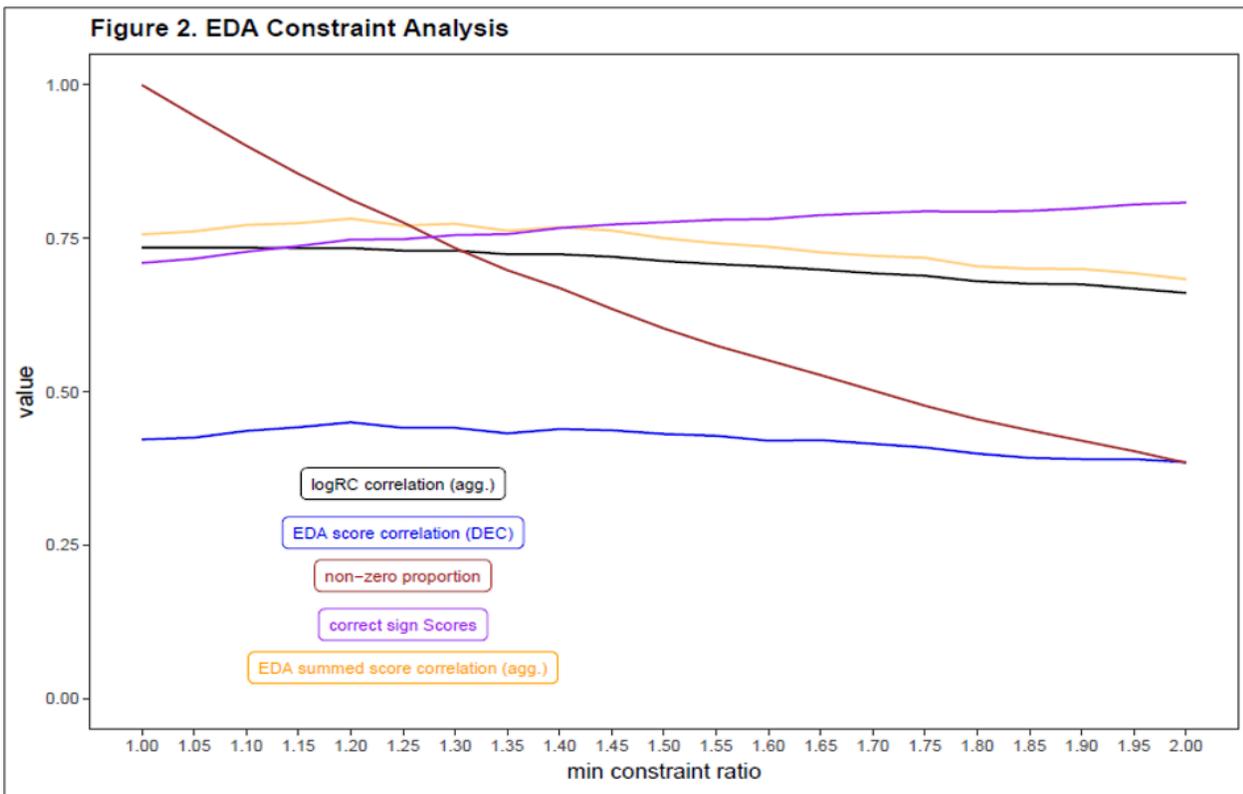
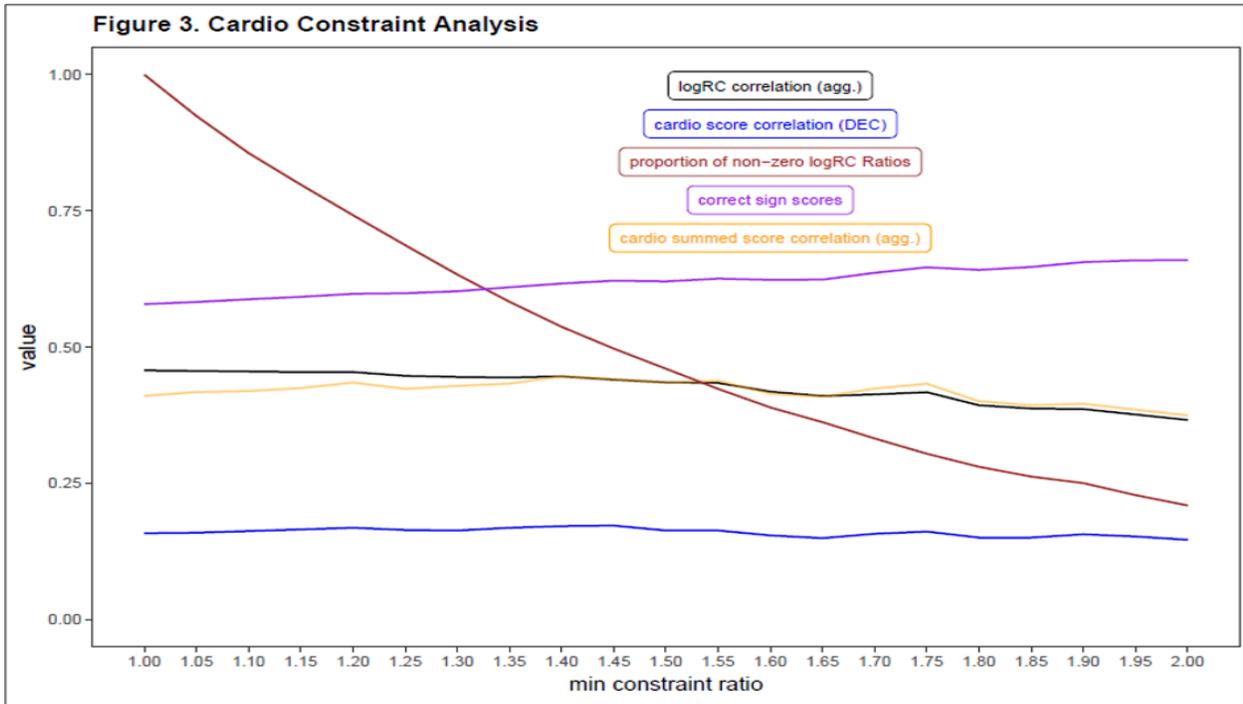
	Proporciones logRC (2700 puntajes)	Proporciones Agregadas logRC (n=300)	Puntajes Enteros Sumados (n=300)
Respiración	UP=.216,LP=.182,comb=.211	UP=.420, LP=.359, comb=.401	.228
EDA	.439	.751	.749
Cardio	.183	.460	.410

La figura 1 muestra un gráfico con la serie de restricciones de relación mínima, de 1.05:1 a 2:1, con los datos respiratorios, incluyendo la correlación de punto-biserial para las Proporciones logRC y las puntuaciones numéricas después de agregar los datos para cada caso. También se muestra la correlación de punto-biserial para las 2700 puntuaciones numéricas, junto con la proporción de puntuaciones diferentes a



cero para cada nivel de restricción mínima y la proporción de puntuaciones correctas diferentes a cero. La figura 2 muestra un gráfico con la misma información para los datos del EDA. La figura 3 muestra la misma información para los datos de cardio.





Discusión

Para cada uno de los sensores de registros respiratorio, EDA y cardio, la correlación de punto-biserial entre las Proporciones logRC agregadas fue mayor cuando no existió restricción mínima. La aplicación de la serie de restricciones mínimas crecientes dio lugar a un debilitamiento continuo del coeficiente de punto-biserial. No es sorprendente que, para cada uno de los sensores de registro, la proporción de valores distintos de cero - tanto las Proporciones logRC como las puntuaciones numéricas - fuera mayor sin restricción mínima y se hiciera progresivamente más pequeña en la medida que aumentaba el coeficiente de restricción mínima. Este efecto fue más pronunciado para los datos respiratorios en los que el 95% de las puntuaciones fue de cero (0) con una proporción de restricción mínima de 1.6:1. Por el contrario, aproximadamente el 55% de las puntuaciones de EDA y el 39% de las puntuaciones de cardio fueron distintas a cero con la misma proporción (1.6:1). El coeficiente de correlación (similar al DEC) para las 2700 puntuaciones numéricas no se vio realmente afectado por los datos de cardio. El coeficiente de correlación para las puntuaciones del EDA se vio mínimamente afectado por la serie de restricciones, comenzando en 0.411 y aumentando ligeramente a 0.450 con la proporción de 1.2:1, y terminando en 0.385 con la proporción de restricción máxima de 2:1.

La proporción de puntuaciones con signo correcto - tanto de las Proporciones logRC como de las puntuaciones numéricas - aumentó ligeramente a lo largo del rango de proporciones de restricción mínimas crecientes para los datos del EDA y cardio. Sin embargo, la magnitud de este aumento fue sustancialmente menor que el aumento en el número de puntuaciones de cero (0) para estos sensores. En el caso de los datos respiratorios, la proporción de puntuaciones con signo correcto aumentó hasta alcanzar un pico en una proporción de 1.35:1 y se volvió inestable a partir de una relación superior a 1.6:1. Esta inestabilidad puede atribuirse al pequeño número de puntuaciones diferentes a cero que se mantuvieron con las proporciones de restricción de 1.6:1 y superiores.

La correlación de punto-biserial para las Proporciones logRC (mostradas en negro en las Figuras 1, 2 y 3), junto con la correlación de puntuaciones (mostrada en azul) - similar a una correlación DEC del resultado, pero calculada en este análisis sin puntuaciones numéricas - proporciona una síntesis conveniente de la información compleja en este análisis. Esta correlación agrupa información acerca de las puntuaciones correctas, incorrectas y nulas (0) en un único índice numérico para el que se puede aplicar la intuición tradicional de los coeficientes de correlación. En el caso de las puntuaciones cardiovasculares, no puede identificarse restricción mínima alguna que aumente la eficacia de las puntuaciones extraídas de los datos registrados. En el caso de las puntuaciones del EDA, el efecto de una proporción de restricción mínima para mejorar el coeficiente de correlación de las puntuaciones numéricas fue mínimo. Lo que



queda por saber es si existe alguna ventaja estadísticamente significativa para el uso de una proporción de restricción mínima para las puntuaciones numéricas. Sin embargo, estos datos sugieren que no hay ninguna ventaja en el uso de una restricción mínima con las Proporciones logRC utilizadas en los métodos de puntuación automatizados - y esta misma conclusión se observaría utilizando proporciones sin una transformación logarítmica. Las correlaciones de puntuación para los puntajes respiratorios y la correlación de puntuación agregada (mostrada en naranja), sugieren que restringir la extracción de puntajes al rango de 1.2:1 y 1.6:1 puede ser útil para optimizar la contribución de las puntuaciones de la respiración a las conclusiones correctas frente a las incorrectas.

Una limitación obvia de este estudio es la falta de una prueba de significancia estadística. Es posible incluir dicha prueba, pero requeriría de una metodología muy compleja que aumentaría considerablemente su sobrecarga para los lectores y podría reducir el nivel de interés en este tema tan importante. La optimización estadística de la extracción de características y de las puntuaciones numéricas es un reto analítico no trivial que merece una mayor atención en esta publicación. Se pensó que limitar este proyecto a un estudio de correlación, y de la presentación de los resultados analíticos de alta dimensión tres gráficos, podría servir para mantener la legibilidad y la claridad entre los lectores interesados en el tema. Los datos de este análisis apoyan la validez de la BIBR como una solución razonable, y sugieren que no es necesaria una proporción de restricción mínima para los métodos de análisis automatizados. (Se presume que los algoritmos de puntuación computarizada ya hacen uso de cualquier diferencia medible que pueda extraerse entre las preguntas relevantes y de comparación). Otra limitación de este análisis es la falta de una segunda muestra contra la cual comparar estos resultados. Se recomienda un interés e investigación continua en la optimización de la extracción de características y en la transformación numérica para los métodos de análisis automatizados y manuales de los datos de prueba.



Referencias

- Kircher, J.C., Horowitz, S.W, and Raskin, D.C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*, 12(1), 79 – 90.
- Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209-222.
- Krapohl, D.J. (2002). Short report: Update for the Objective Scoring System. *Polygraph*, 31(4), 298 – 302.
- National Center for Credibility Assessment (2017). *Test Data Analysis: Numerical Evaluation Scoring System Pamphlet*. Available from the author. (Retrieved from <http://www.antipolygraph.org> 4-13-2019).
- Nelson, R. & Krapohl, D. K. (2017) Practical polygraph: a recommendation for combining the upper and lower respiration data for a single respiration score. *APA Magazine*. 50(6), 31-41.
- Nelson, R., Krapohl, D.J., and Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph*, 37(3), 185-215.
- Nelson, R. & Handler, M. (2015). Statistical reference distributions for comparison question polygraphs. *Polygraph*, 44(1), 91-114.
- Nelson, R. & Handler, M. (2018). Practical polygraph: seven things to know about feature extraction with electrodermal and cardio data. *APA Magazine*, 51(1), 116-132.
- Nelson, R. (2017). Multinomial reference distributions for the Empirical Scoring System. *Polygraph and Forensic Credibility Assessment*, 46 (2), 81-115.
- Nelson, R. (2018). Multinomial reference distributions for three-position scores of comparison question polygraph examinations. *Polygraph and Forensic Credibility Assessment*, 47(2), 158- 175.
- Peng, R. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227.
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.



