

Estudio de los Algoritmos para la Puntuación Poligráfica Computarizada Utilizando Características de Kircher

Raymond Nelson

Abstract

Se hizo un análisis de la literatura publicada acerca de los clasificadores estadísticos automatizados que utilizan un conjunto común de características de respuesta fisiológica. Estas características fueron presentadas y descritas por primera vez por investigadores de la Universidad de Utah durante la década de los 80's, y posteriormente fueron descritas por la profesión poligráfica como *características de Kircher*. Éstas incluyen el aumento en la amplitud de la actividad electrodérmica y cardiovascular, junto con una reducción de la actividad respiratoria. También puede incluirse la constricción o reducción de la amplitud del pulso vasomotor. Un elemento importante de estas características, además de su correlación estadística con el engaño y la veracidad, es que se pueden extraer de los datos de serie-tiempo registrados tanto de forma visual como por métodos computarizados automatizados. Dentro de los clasificadores estadísticos que se basan en estas características se incluyen a los siguientes: Análisis de Probabilidad, Sistema de Puntuación por Orden de Rango, Método de Permutación Objetiva, Método de Análisis Bootstrap, Sistema de Puntuación Empírica/Multinomial, Sistema de Puntuación Objetiva (versiones 1 y 2), y el Sistema de Puntuación Objetiva – versión 3. En los apéndices se resumen las características en el diseño de estos métodos de análisis.

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

Introducción

Al igual que todos los algoritmos de análisis de datos, los algoritmos de puntuación poligráfica están compuestos por una serie de operaciones fundamentales o funciones comunes. Estas incluyen: extracción de características, transformación numérica y reducción de datos, uso de alguna forma de función de verosimilitud, y reglas o métodos estructurados para la interpretación y clasificación de su producto o de sus resultados. Ya que todo análisis de datos comienza con la extracción de características - la identificación de variaciones útiles e informativas dentro de los datos disponibles - el perfeccionamiento del conocimiento relacionado con las características útiles de respuesta es un área fundamental para el desarrollo de diferentes enfoques metodológicos cuando se buscan soluciones para otras funciones posteriores dentro del mismo método de análisis de datos.

Un ejemplo de esto es un conjunto de características que son referidas como *características de Kircher* (Kircher, 1981, 1983; Kircher y Raskin, 1988), descritas por primera vez bajo este nombre por Krapohl y McManus (1999). De manera simple, estas características constan de una señal primaria para cada uno de los sensores de registro tradicionales: amplitud fásica de la respuesta electrodérmica, aumento fásico de la presión sanguínea relativa y reducción de la actividad respiratoria. La actividad vasomotora también puede incluirse dentro de este conjunto de características. Otros investigadores (Harris, Horner y McQuarrie, 2000; Kircher, Kristjansson, Gardner y Webb, 2004; Podlesny y Truslow, 1993; Rovner, 1986) también demostraron la eficacia de estas características de respuesta.

El desarrollo de un método para el análisis de datos es un proceso en el que primero se especifica la información de salida deseada - como por ejemplo una clasificación estadística del engaño o veracidad - y luego se deconstruye el proceso para alcanzar ese objetivo mediante un conjunto coherente de supuestos y una serie de funciones reproducibles. Todos los algoritmos están compuestos por cuatro funciones esenciales: extracción de características, transformación numérica y reducción de datos, algún tipo de función de verosimilitud y un conjunto de reglas o de procedimientos que se utilizan para interpretar el resultado de la prueba. Por supuesto que es posible realizar algunas otras descripciones dentro del proceso. Sin embargo, estas cuatro funciones básicas pueden generalizarse a casi todos los métodos de análisis de datos, ya sean manuales o automatizados, y sin importar que se basen en métodos tradicionales de clasificación y predicción estadística o en el

aprendizaje automatizado o de inteligencia artificial.

Al trabajar en sentido inverso a partir de estas funciones, la interpretación se refiere al proceso de traducir un resultado numérico y estadístico en información conceptual que pueda ser útil o informativa para personas ajenas o no involucradas en el proceso de evaluación o del análisis de los datos. En otras palabras, la interpretación sirve para responder a la pregunta: ¿qué significa realmente el resultado de la prueba? Los resultados categóricos de la prueba (es decir, positivo o negativo, y otros términos alegóricos) son la forma más simplista de interpretación y proporcionan la cantidad más pequeña de detalles acerca del significado científico de un resultado de prueba - pero a menudo proporcionan una interpretación más práctica o procesable. En la práctica poligráfica de campo, los resultados categóricos son el resultado de una regla de decisión¹ procesal. Sin embargo, los resultados categóricos pueden ser fuente de malentendidos y de confusión cuando se espera ingenuamente que sean infalibles. Por regla general, no se espera que las pruebas científicas sean infalibles. Aunque estén sujetos a una interpretación dicotómica, todos los resultados de las pruebas científicas son fundamentalmente probabilísticos y, por tanto, están inherentemente sujetos a la incertidumbre. Uno de los principales objetivos de cualquier prueba científica es cuantificar y/o reducir el grado de incertidumbre asociado a una conclusión.

Una *función de verosimilitud* es un mecanismo que se utiliza para obtener un valor estadístico reproducible para los datos observados en la prueba. Las funciones de verosimilitud pueden adoptar muchas formas, incluyendo las distribuciones empíricas y matemáticas. Una función de verosimilitud también puede considerarse como los parámetros y fórmulas utilizados para calcular una distribución de referencia. Un ejemplo sencillo y práctico de una función de verosimilitud es una tabla publicada de valores para una distribución de referencia. La forma más básica y sencilla de función de verosimilitud es una puntuación de corte numérico a partir del que se da un resultado categórico de prueba. Las puntuaciones de corte numéricas pueden pensarse como la asociación a una probabilidad estadística de que el resultado de una prueba sea correcto o incorrecto.

¹Las reglas de decisión poligráficas utilizan información numérica agregada, que pueden ser puntuaciones totales y subtotales. [Véase Nelson (2018) para una descripción de las diferentes reglas de decisión, incluyendo la regla de gran total (GTR), la regla de dos etapas (TSR), la regla de puntuación subtotal (SSR) entre otras].

Antes de poder calcular cualquier valor estadístico, si los datos registrados no son numéricos, deben someterse a una *transformación numérica y a una reducción de datos*. Los procedimientos para obtener estos valores numéricos pueden variar ampliamente, y pueden incluir el uso de una medición física cuando es procedente, escalas de Likert, ordenamiento por rangos, uso de proporciones, puntuaciones z u otros valores matemáticos o estadísticos. Los métodos de reducción de datos también pueden variar, y pueden incluir la suma, el promedio, el promedio ponderado, las funciones discriminantes, las funciones logarítmicas, y el remuestreo entre otros métodos. El objetivo funcional de la transformación de datos y de la reducción numérica es transformar los datos de prueba registrados en un conjunto de valores numéricos útiles, y reducir esos valores a un pequeño conjunto de números con los que se pueda obtenerse un valor estadístico.

Antes de poder calcular cualquier resultado numérico o estadístico, deben extraerse características de respuesta útiles de los datos registrados. *La extracción de características* es el principio de cualquier método de análisis de datos algorítmico o procedimental. Todos los datos son una combinación de señal y ruido. La extracción de características es el proceso de identificación y aislamiento de la información de respuesta de interés, lo que incluye la identificación del inicio y fin de la respuesta. En condiciones ideales, la relación señal/ruido es muy alta y es muy fácil aislar la información la señal útil del ruido inútil. La investigación acerca de la extracción de características es fundamental para el desarrollo de soluciones para todas las funciones analíticas posteriores².

Método

Se examinó la bibliografía publicada acerca de las descripciones de los clasificadores estadísticos basados en características de respuesta fisiológica descritos por investigadores de la Universidad de Utah (Kircher y Raskin, 1988). Se enumeraron las características de diseño y desarrollo de estos métodos.

²Por simplicidad, en este documento se ignoran una serie de funciones importantes previas al análisis de datos que incluyen el desarrollo de sensores, el desarrollo de los estímulos, la administración de la prueba, el procesamiento de las señales y el registro de datos.

Se buscó información sobre el tipo de método de decisión y el tipo de clasificador estadístico, al igual que de los métodos para la transformación numérica y la reducción de datos - incluyendo la selección de pares de preguntas de comparación-relevantes, cuando esto aplicaba. Se describen los métodos generales para el desarrollo de una función estadística de verosimilitud para cada uno de los algoritmos de análisis. Por último, se describen reglas de procedimiento para la interpretación o clasificación del engaño y la veracidad para cada uno de los métodos de análisis, dependiendo de cómo se aplican en los exámenes poligráficos de asunto único y de asuntos múltiples.

Resultados

Se encontraron siete clasificadores estadísticos diferentes en la literatura publicada. Estos incluyen: Análisis de Probabilidad (PA; Kircher & Raskin, 1988), Sistema de Puntuación por Orden de Rango (ROSS; Honts & Driscoll, 1987; 1988), Método de Análisis Bootstrap (BAM; Honts & Devitt, 1992), Sistema de Puntuación Objetiva - versiones 1-2 (OSS 1-2: Krapohl, 2002; Krapohl & Mc- Manus, 1999), Método de Puntuación Objetiva por Permutación (OPS; MacLaren & Krapohl 2003), Sistema de Puntuación Objetiva versión 3 (OSS- 3; Nelson, Krapohl & Handler, 2008), y el Sistema de Puntuación Empírica - Multinomial (ESS/ESS-M; Nelson, 2017a; Nelson, Krapohl & Handler, 2008).

La información de interés para esta revisión, además del uso de las características de Kircher y el sensor de registro asociado, incluyó lo siguiente: modelo de decisión, clasificador estadístico, transformación numérica, método o procedimiento para la selección de los pares de análisis RQ y CQ, método de reducción de datos, tipo de función de verosimilitud y reglas de decisión del procedimiento. El tipo de modelo de decisión se refiere al método general por el que se logra una clasificación; puede incluir el uso de una prueba z, la discriminación de señales gaussiana-gaussiana y Bayes simple entre otros métodos. Se observaron diferentes tipos de clasificadores estadísticos, entre ellos los valores p, posibilidades posteriores, comparaciones de punto y de puntuaciones de corte que se asignan a tasas de TN y TP, entre otros métodos. Los examinadores poligráficos de campo que deseen una mejor comprensión de las diferencias entre los métodos de puntuación manual y automatizada podrían estar interesados en toda la variedad de métodos empleados para la selección y comparación de los valores entre pares de RQ y CQ. Los métodos para la transformación numérica incluyeron

transformaciones de rango, diferencia en las puntuaciones y puntuaciones z , entre otros métodos. Las estrategias para reducir los subtotales de los sensores y de las preguntas hacia puntuaciones de gran total incluyeron la suma, el promedio y el promedio ponderado. Se observaron diversos tipos de funciones de verosimilitud, como fueron las distribuciones empíricas, distribuciones multinomiales, distribuciones Bootstrap y permutaciones, entre otros métodos. Las reglas de decisión incluyeron el uso de puntuaciones subtotales para exámenes de asuntos múltiples y puntuaciones de gran total para exámenes de asunto único, entre otras soluciones de procedimiento. Las características de diseño y desarrollo de cada uno de estos métodos de análisis se muestran en los Apéndices A a G.

Conclusión

Desde la década de los 80's, las características de Kircher han sido una solución útil y eficiente ante el reto de la extracción de características poligráficas. La existencia de un conjunto de características fácil de identificar y de utilizar ha facilitado el estudio y el desarrollo de diversos tipos de clasificadores estadísticos. Estas características se describieron por primera vez en el desarrollo del algoritmo del Análisis de Probabilidades (Kircher y Raskin, 1988). En la literatura publicada actual, se pueden encontrar siete diferentes algoritmos computarizados que hacen uso de este conjunto de características poligráficas comunes. Muchos de estos métodos están disponibles en productos comerciales y profesionales utilizados por los poligrafistas de campo.

Ya que todos los métodos de análisis de esta revisión se basan en las características de Kircher, todos incluyen sensores de registro similares. Pueden existir algunas diferencias en el procesamiento de señal con los distintos instrumentos poligráficos. Sin embargo, las diferencias en el procesamiento de señal están fuera del alcance de este proyecto. En los métodos de análisis estudiados se incluyen los distintos tipos de clasificadores estadísticos. Entre ellos se incluyen la estimación de verosimilitud máxima, el análisis de discriminación lineal, la detección/discriminación de señales gaussianas-gaussianas, los métodos de permutación y bootstrap, y las transformaciones de rango. Algunos métodos utilizan clasificadores bayesianos simples, en los que el resultado posterior puede considerarse pragmáticamente como una probabilidad de engaño o de veracidad.

Aunque se considera como una de las características de Kircher, los datos de la respuesta

vasomotora no se incluyen en la mayoría de estos métodos de análisis. Aunque se puede encontrar información acerca de la actividad vasomotora en la literatura publicada - incluyendo publicaciones de algunos desarrolladores de los algoritmos de puntuación disponibles - no se ha incluido en el modelo estructural de los métodos de puntuación publicados y disponibles. Las razones de esto no se han discutido por completo. Sin embargo, cabe suponer que es muy probable que los datos vasomotores se habrían incluido si hubieran mejorado el tamaño del efecto de los modelos estructurales publicados. Por lo tanto, se puede plantear la hipótesis de que los datos vasomotores, aunque quizás correlacionados con el criterio de interés, pueden no haber mejorado los modelos estructurales descritos en la literatura publicada. La adición de datos vasomotores requeriría una base suficiente de datos para volver a desarrollar las diversas funciones de verosimilitud y estudiar los tamaños de los efectos resultantes. Un método de análisis, el ESS-M (Nelson, 2017a) incluye una función de verosimilitud que puede incluir datos vasomotores. Sin embargo, la información publicada no muestra ninguna diferencia en el tamaño del efecto cuando se incluye la información vasomotora (Nelson, 2017b). Una mejor comprensión de los datos potenciales de respuesta vasomotora y de los tamaños del efecto para los métodos de puntuación automatizado, requerirá la réplica y ampliación de estos métodos de algoritmos, además del recálculo o nuevo desarrollo de las funciones de verosimilitud asociadas.

Las limitaciones de este proyecto son varias, e incluyen el hecho de que este proyecto sólo pretende proporcionar un resumen descriptivo de las características en el diseño de estos diferentes métodos. No se incluye en este informe ninguna descripción matemática o de procedimiento de los métodos de puntuación identificados. Otra limitación de este estudio es que no incluye otros métodos de análisis computarizados que hagan uso de otras características de puntuación. Pueden existir otros métodos de análisis en las publicaciones, incluyendo métodos que se basan en métodos de extracción de características patentados y de boutique, y características de respuesta que son menos comunes o intuitivas para los profesionales poligráficos de campo - por ejemplo, características de respuesta espectral. Este proyecto se limita a métodos que existen en publicaciones abiertas, y no incluye algoritmos que están sujetos a restricciones del propietario o de propiedad intelectual. Otras investigaciones deberían abordar la necesidad de obtener información sobre esos métodos. Por último, los algoritmos incluidos en este informe no incluyen información sobre el tamaño de los efectos. Otras investigaciones deberían abordar la necesidad de obtener información sobre los tamaños de los efectos, incluyendo las tasas de sensibilidad, especificidad, falsos

positivos y falsos negativos, y los errores de medición asociados. En el futuro habrá que seguir investigando las posibles ventajas de las distintas soluciones a la serie de retos inherentes al análisis y la clasificación automatizados de datos estadísticos.

Este proyecto es una breve descripción de la información conceptual, aunque no técnica, que puede ser útil a los lectores que deseen una introducción al tema de los algoritmos de puntuación computarizada del polígrafo, junto con una introducción a la amplitud de la actividad en esta área a lo largo de los últimos 35 años o más. Además del hecho de que la disponibilidad de un conjunto útil de características de respuesta conocidas ha permitido a una variedad de investigadores estudiar la aplicación de diferentes métodos estadísticos a la clasificación del engaño y veracidad, las características de Kircher tienen la ventaja de que pueden proporcionar una comprensión intuitiva a los examinadores poligráficos de campo que desean entender qué detalles de la fisiología registrada se incluyen en el análisis. De hecho, las características de Kircher pueden extraerse manualmente o de forma automatizada. Se espera que esta información sea útil o informativa para aquellos interesados en el desarrollo de algoritmos de análisis de datos poligráficos, y para los profesionales de la poligrafía de campo que deseen comprender mejor las diferencias entre los diversos métodos de análisis y los procedimientos tradicionales de puntuación manual.

Referencias

- Harris, J. Horner A. & McQuarrie, A. D. (2000). *An Evaluation of the Criteria Taught by the Department of Defense Polygraph Institute for Interpreting Polygraph Examinations*. Department of Defense Polygraph Institute report No DoDPI00-R-0007
- Honts, C. R. & Devitt, M. K. (1992). *Bootstrap decision making for polygraph examinations*. Report number DoDPI92-R-0002. Department of Defense Polygraph Institute report No DoDPI92-R-0002. Reprinted in *Polygraph*, 3, (1), 1-47.
- Honts, C. R. & Driscoll, L. N. (1987). An evaluation of the reliability and validity of rank order and standard numerical scoring of polygraph charts. *Polygraph*, 16, 241-257.
- Honts, C. R. & Driscoll, L. N. (1988). A field validity study of rank order scoring system (ROSS) in multiple issue control question tests. *Polygraph*, 17, 1-15.
- Kircher, J. C. (1981). *Computerized chart evaluation in the detection of deception*. University of Utah.
- Kircher, J. C. (1983). *Computerized decision making and patterns of activation in the detection of deception*. Doctoral dissertation, University of Utah, Salt Lake City. Dissertation Abstracts International, 44, 345.
- Kircher, J. C., Kristjansson, S. D., Gardner, M. D. & Webb, A. (2004). *Human and computer decision-making in the psychophysiological detection of deception*. University of Utah.
- Kircher, J. C. & Raskin, D. C. (1988). Human versus computerized

evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291-302.

Krapohl, D. J. (2002). Short report: Update for the objective scoring system. *Polygraph*, 31, 298-302. Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209-222.

MacLaren, V. V. & Krapohl, D. J. (2003). Objective Assessment of Comparison Question Polygraphy. *Polygraph*, 32, 107-126.

Nelson, R., Krapohl, D. & 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, 185-215.

Nelson, R. (2017a). Multinomial reference distributions for the Empirical Scoring System. *Polygraph & Forensic Credibility Assessment*, 46 (2), 81-115.

Nelson, R. (2017b). Updated numerical distributions for the Empirical Scoring System. An accuracy demonstration with archival datasets with and without the Vasomotor Sensor. *Polygraph & Forensic Credibility Assessment*, 46 (2), 116-131.

Nelson, R. (2018). Practical polygraph: a survey and description of decision rules. *APA Magazine*, 51(2), 127-133.

Nelson, R. (2019). Literature survey of structural weighting of polygraph signals: why double the EDA? *Polygraph & Forensic Credibility Assessment*, 48 (2), 105-112.

Podlesny, J. A. & Truslow, C. M. (1993). Validity of an expanded-

issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78, 788-797.

Rovner L. I. (1986). The accuracy of physiological detection of deception for subjects with prior knowledge. *Polygraph*, 15(1), 1-39

Apéndice A. Análisis de Probabilidad

Análisis de Probabilidad (Kircher y Raskin, 1988)	
Sensores	Respiración (torácica y abdominal), Electrodermico, Cardiovascular.
Características de respuesta	Reducción de la actividad respiratoria, amplitud fásica de la respuesta electrodermica, incremento cardiovascular fásico de la presión sanguínea relativa.
Modelo de decisión	El algoritmo calcula una función discriminatoria que sirve como clasificador estadístico - un valor estadístico para el que los puntajes de corte de decisión pueden hacer clasificaciones categóricas de engaño o veracidad.
Clasificador estadístico	Análisis Bayesiano que utiliza una función de verosimilitud obtenida mediante un análisis discriminante. Los resultados pueden considerarse como una probabilidad posterior de engaño o veracidad.
Selección de CQ	Para cada sensor, la media de todas las RQ entre gráficos se compara con la media entre todas las CQ entre los gráficos
Transformación numérica	Valores numéricos transformados en puntuaciones Z para cada sensor utilizando las RQ y CQ combinadas.
Reducción de datos	Se promedian las puntuaciones Z entre gráficos para los sensores individuales de todas las RQ y para todas las CQ. Se combinan las puntuaciones Z de los sensores mediante una función de ponderación estructural que se obtuvo mediante un análisis de discriminación lineal.
Función de verosimilitud	Se utilizan dos fórmulas de verosimilitud para calcular los valores de verosimilitud complementarios del engaño y veracidad.
Reglas de decisión: asunto único	GTR
Reglas de decisión: asunto múltiple	Ninguna
Comentarios	Existen coeficientes estructurales de parte de los desarrolladores y también por estudios de replicación. La publicación describe la aplicación del análisis probabilístico (PA) a los formatos de examen de asunto único. La aplicación del algoritmo de PA a exámenes de asuntos múltiples puede requerir un cambio en la agregación de las RQ y CQ dentro de los gráficos y también entre ellos. Las ventajas de los esquemas de transformación por separado, dentro del gráfico y entre gráficos, no se han descrito completamente en la publicación, sin embargo, el algoritmo subsecuente ha demostrado la aplicación de estos a exámenes de asuntos múltiples. Por esta razón, sería posible adaptar el análisis probabilístico con múltiples formatos de prueba.

Apéndice B. Sistema de Puntuación de Orden de Rangos

ROSS (Honts y Driscoll, 1987, 1988)	
Sensores	Respiración (torácica y abdominal), Electrodermico, Cardiovascular, Vasomotor
Modelo de decisión	El modelo de decisión ROSS es similar a un método de discriminación de señales gaussiano-gaussiano que utiliza distribuciones de suma de rangos derivados empíricamente para los casos de culpables e inocentes.
Clasificador estadístico	Los clasificadores estadísticos son tasas TP, TN, FP y FN derivadas empíricamente.
Características de respuesta	Reducción de la actividad respiratoria, amplitud física de la respuesta electrodermica, incremento cardiovascular físico de la presión sanguínea relativa.
Selección de CQ	Para el análisis, las CQ y RQ no se aparean como en la puntuación poligráfica tradicional. En su lugar, el análisis de rangos de clasificación comienza con la asignación de puntajes de rango enteros para todos los estímulos de prueba, incluyendo a todas las CQ y a las RQ juntas dentro de cada gráfico registrado.
Transformación numérica	Las puntuaciones enteras se le asignan a la variación del orden de rango de las RQ y las CQ para cada sensor dentro de cada gráfico. Las puntuaciones de clasificación se asignan en orden inverso, en el que a la respuesta con el mayor cambio fisiológico se le asigna un valor de rango igual al número total de RQ y CQ, y la respuesta más pequeña recibe un valor de clasificación de 1.
Reducción de datos	Los valores de rango se suman para todas las RQ de todos los gráficos y también para todas las CQ de todos los gráficos. A continuación, se calcula una puntuación CQtotal: $RQ_{total} = \text{DiferenciaDeRango}$.
Función de verosimilitud	Se pueden calcular las distribuciones empíricas para las puntuaciones de DiferenciaDeRango (<i>RankDifference</i>) y se pueden seleccionar las puntuaciones de corte numéricas para lograr los tamaños de efecto deseados.
Reglas de decisión: asunto único	GTR
Reglas de decisión: asunto múltiple	SSR
Comentarios	Las transformaciones de orden de rango son una solución no paramétrica común, y a veces pueden optimizar la robustez de los datos confusos y difíciles, con cierto costo potencial debido a la granularidad de los valores de clasificación. Se obtiene cada valor de clasificación al comparar cada respuesta con todas las demás respuestas (todas las demás RQ y CQ), lo que puede complicar las suposiciones de varianza RQ independiente para exámenes de asuntos múltiples.

Apéndice C. Sistema de Puntuación de Remuestreo (*Bootstrap*)

Sistema de Puntuación de Remuestreo (<i>Bootstrap</i>) (Honts y Devit, 1992)	
Sensores	Respiración (torácica y abdominal), Electrodérmico, Cardiovascular, Vasomotor
Modelo de decisión	Prueba Z de una diferencia observada entre CQ-RQ utilizando una distribución nula de remuestreo (<i>Bootstrap</i>).
Clasificador estadístico	Valor P, que indica la probabilidad de obtener una puntuación igual o más extrema que la puntuación observada bajo la hipótesis nula de que no hay diferencia entre el valor de CQ y RQ.
Características de respuesta	Reducción de la actividad respiratoria, amplitud fásica de la respuesta electrodérmica, incremento cardiovascular fásico de la presión sanguínea relativa.
Selección de CQ	Cada RQ se aparea con las CQ anteriores para calcular los valores de diferencia CQ-RQ después de transformar todos los valores RQ y CQ en puntuaciones Z.
Transformación numérica	Las presentaciones de todas las RQ y CQ, para cada sensor, de todos los gráficos combinados, se transforman en puntuaciones Z. De esta forma, todas las puntuaciones de todos los sensores y de todos los gráficos tienen un valor de escala común que puede fácilmente sujetarse a un remuestreo (<i>bootstrapping</i>).
Reducción de datos	Los puntajes CQ – RQ = diferencia-Z se calcularon para obtener las puntuaciones Z. Existe la hipótesis de que las puntuaciones de diferencia-Z se cargan en valores mayores a cero para los sujetos inocentes, y menores y cero para sujetos culpables. Las puntuaciones de diferencia-Z se agregan mediante una sumatoria para obtener una puntuación única de diferencia-Z para cada examen.
Función de verosimilitud	Se calcula una distribución nula para cada examen combinando todas las puntuaciones Z de RQ y CQ, para todos los gráficos y todos los sensores en un solo vector, y luego se hace un remuestreo de una distribución nula (muestreo aleatorio con reemplazo) al tiempo que se asignan valores arbitrarios como CQ o RQ.
Reglas de decisión: asunto único	GTR
Reglas de decisión: asunto múltiple	NA
Comentarios	Este método se describió utilizando exámenes de asunto único con el mismo número de RQ y CQ, pero podría adaptarse a formatos de prueba de asuntos múltiples y formatos de prueba con números desiguales de RQ y CQ.

Apéndice D. Sistema de Puntuación Pobjetiva (OSS/OSS-2)

OSS/OSS-2 (Krapohl y McManus, 1999; Krapohl, 2002)	
Sensores	Respiración (torácica y abdominal), Electrodermico, Cardiovascular,
Modelo de decisión	Modelo de discriminación de señales Gaussiano-Gaussiano (clasificador de detección de señales aplicado a una tarea de discriminación de señales).
Clasificador estadístico	Valor P, indicativo de la verosimilitud del estadístico de prueba observada bajo la distribución representada por los datos de entrenamiento, confirmados como opuestos a la clasificación seleccionada (engaño o veraz).
Características de respuesta	Reducción de la actividad respiratoria, amplitud fásica de la respuesta electrodermica, incremento cardiovascular fásico de la presión sanguínea relativa.
Selección de CQ	Cada RQ se compara con la CQ previa.
Transformación numérica	Las proporciones R/C se transforman en puntuaciones enteras utilizando una distribución de contenedores séptiles uniformes. Las puntuaciones de 7 posiciones del OSS-2 [-3, -2, -2, 0, +1, +2, +3] difieren de las puntuaciones poligráficas de 7 posiciones tradicionales en que el rango de puntuaciones del OSS-2 puede ocurrir con la misma probabilidad, mientras que las puntuaciones tradicionales de 7 posiciones se cargan cerca del 0, y las puntuaciones más alejadas de 0 ocurren con menos frecuencia.
Reducción de datos	Las puntuaciones enteras se agregan mediante la suma de cada RQ, para todos los sensores y todas las gráficas. Luego, los subtotales de las RQ se suman para obtener una puntuación total general. Debido a que las puntuaciones de números enteros se agregan mediante sumatoria, no hay diferencia si las puntuaciones de los sensores se suman primero entre gráficas o si se suman dentro de las gráficas.
Función de Verosimilitud	Las tablas de referencia OSS-2 se derivan empíricamente.
Reglas de decisión: asunto único	GTR
Reglas de decisión: asunto múltiple	ninguno
Comentarios	Las funciones de verosimilitud del OSS-2 (tablas de referencia) están disponibles para exámenes de polígrafo de un solo asunto con 3 RQ y 3 gráficas. El diseño de sumatoria implica que la función de probabilidad puede ser menos robusta cuando existen datos faltantes y con artefactos, y puede sobrecargarse cuando se usan más de tres gráficas, y puede verse sesgada con formatos de prueba con números desiguales de RQ y CQ. No se han publicado las funciones de verosimilitud del OSS para exámenes de asuntos múltiples. OSS-1 y OSS-2 comenzaron como protocolos de puntuación manual para los cuales la estructura y los procedimientos estaban lo suficientemente estructurados y sin ambigüedades, lo que conducía fácilmente a la automatización. En consecuencia, OSS y OSS-2 ahora son métodos de análisis automatizados de facto.

Apéndice E. Sistema de Calificación por Permutación

Sistema de Calificación por Permutación (MacLaren y Krapohl, 2003)	
Sensores	Respiración (torácica y abdominal), Electrodermico, Cardiovascular, Vasomotor
Modelo de decisión	Un clasificador bayesiano simple que utiliza el teorema de Bayes en forma de posibilidades (<i>odds</i>), donde los valores obtenidos $p/(1-p)$ son una permutación de las puntuaciones enteras uniformes de 7 posiciones.
Clasificador estadístico	El uso del teorema de Bayes significa que los resultados pueden considerarse, en términos prácticos, como una probabilidad posterior de engaño o veracidad.
Características de respuesta	Reducción de la actividad respiratoria, amplitud fásica de la respuesta electrodermica, incremento cardiovascular fásico de la presión sanguínea relativa.
Selección de CQ	Los RQ se aparean con las CQ previas.
Transformación numérica	Se calcula una proporción para cada par de RQ/CQ, después de lo cual se asignan dos conjuntos de puntuaciones enteras utilizando dos distribuciones de contenedores séptiles uniformes que se calcularon a partir de casos confirmados de culpabilidad e inocencia.
Reducción de datos	Se suman dos conjuntos de puntuaciones enteras para obtener dos puntuaciones de gran total (CulpableTotal e InocenteTotal) que luego se comparan con una función de probabilidad del PSS.
Función de Verosimilitud	La función PSS de verosimilitud es la permutación de todas las puntuaciones de 7 posiciones posibles si no están asociadas sistemáticamente a culpabilidad o inocencia. La distribución exacta incluye 6.57×10^{22} combinaciones posibles. Se puede calcular utilizando una fórmula combinatoria y se puede aproximar fácilmente mediante simulación.
Reglas de decisión: asunto único	GTR
Reglas de decisión: asunto múltiple	ninguna
Comentarios	El PSS se desarrolló con un examen que consta de tres presentaciones de una secuencia de preguntas que incluye tres RQ y tres CQ. La adaptación del método PSS a exámenes de asuntos múltiples y a formatos de prueba con dos a cuatro RQ y con cuatro o cinco presentaciones requiere datos disponibles de casos confirmados para calcular las distribuciones séptiles uniformes, además del recálculo de la función de verosimilitud de permutación.

Apéndice F. Sistema de Calificación Objetiva – Versión 3

OSS-3 (Nelson, Krapohl y Handler, 2008)	
Sensores	Respiración (torácica y abdominal), electrodérmico, cardiovascular
Modelo de decisión	Modelo de discriminación de señales Gaussiano-Gaussiano (clasificador de detección de señales aplicado a una tarea de discriminación de señales).
Clasificador estadístico	Valor P, indicativo de la verosimilitud de que el estadístico de prueba observada bajo la distribución representada por los datos de entrenamiento confirmados es opuesto a la clasificación seleccionada (engaño o veraz).
Características de respuesta	Reducción de la actividad respiratoria, amplitud fásica de la respuesta electrodérmica, incremento cardiovascular fásico de la presión sanguínea relativa.
Selección de CQ	La media de las CQ se compara contra cada RQ.
Transformación numérica	Las proporciones Log R/C estandarizadas van contra los datos de entrenamiento. Las puntuaciones de los sensores son puntuaciones estándar (media = 0, desviación estándar = 1) de -3 a +3. Las proporciones log R/C estandarizadas indican el número de desviaciones estándar que una respuesta observada tiene por encima o por debajo de la media de los datos de entrenamiento, cuando se combinan los casos de culpables e inocentes. Estas puntuaciones son intuitivamente similares a la noción de puntuaciones poligráficas de 7 posiciones, pero con decimales.
Reducción de datos	La gran media es la media de las puntuaciones RQ entre gráficas. Las puntuaciones RQ entre gráficas son las medias de las puntuaciones medias ponderadas de los sensores dentro de la gráfica. Las puntuaciones de los sensores son proporciones logarítmicas R/C estandarizadas. Los coeficientes de ponderación de sensores se obtienen a través del análisis de discriminación lineal y también pueden calcularse mediante regresión logística, remuestreo (<i>bootstrapping</i>) y otros métodos con poca diferencia en la función de ponderación resultante.
Función de verosimilitud	Las tablas de referencia del OSS-3 se derivan empíricamente de casos confirmados de culpabilidad e inocencia.
Reglas de decisión: asunto único	GTR, TSR
Reglas de decisión: asunto múltiple	La regla de decisión SSR del OSS-3 usa el método ANOVA K-W para diferencias y similitudes evaluadas entre las RQ para reducir la presencia de resultados inconclusos en exámenes de asuntos múltiples (bajo el supuesto de que ser veraz en todas las RQ no dará como resultado diferencias significativas en las puntuaciones RQ).

Comentarios	<p>El propósito del OSS-3 fue ampliar el conocimiento disponible del OSS-1 y OSS-2 hacia una gran variedad de formatos de examen, incluyendo los exámenes de asunto único y de asuntos múltiples, con 2, 3 y 4 RQ y 3, 4 o 5 gráficas. El OSS-3 fue diseñado para ser robusto ante ciertos datos faltantes y con artefactos, y para utilizar la 2a de cualquier pregunta repetida dentro de una gráfica. La reducción de datos utilizando el promedio significa que las funciones de verosimilitud del OSS-3 se aplican más fácilmente a exámenes con 3, 4 o 5 presentaciones de la secuencia de preguntas, y la transformación $\log(RQ/CQ_{media})$ puede aplicarse a formatos de prueba con números desiguales de RQ y CQ. El algoritmo también incluye capacidad de marcar segmentos con artefactos e inutilizables para excluirlos del análisis. Los segmentos con artefactos se pueden analizar utilizando una prueba de proporciones, para hacer inferencias sobre su causa, ya sea sistemática o aleatoria.</p>
-------------	--

Apéndice G. Sistema de Calificación Empírica (ESS/ESS-M)

ESS/ESS-M (Nelson, 2017a; Nelson, Handler y Krapohl, 2008)	
Sensores	Respiración (torácica y abdominal), Electrodermico, Cardiovascular, Vasomotor
Modelo de decisión	El ESS se basa en un modelo de discriminación de señales Gaussiano-Gaussiano. ESS es un clasificador bayesiano simple. ESS también se puede estudiar y utilizar con puntajes de corte tradicionales/federales para lograr los tamaños de efecto estudiados empíricamente.
Clasificador estadístico	El ESS utiliza un valor p, obtenido de una distribución empírica, que describe la verosimilitud de los datos observados bajo una hipótesis específica. Las puntuaciones de corte para el ESS se seleccionaron empíricamente, para restringir los errores de FN y FP a los niveles alfa deseados. Los resultados del ESS-M se pueden expresar como posibilidades posteriores de engaño o veracidad. En términos prácticos, el uso del teorema de Bayes significa que los resultados pueden considerarse como una probabilidad posterior de engaño o veracidad. El ESS-M también proporciona las posibilidades de límite inferior del intervalo creíble posterior de alfa-1 para engaño o veracidad, lo que indica la verosimilitud de obtener un resultado categórico similar al repetir el procedimiento de prueba.
Características de respuesta	Reducción de la actividad respiratoria, amplitud fásica de la respuesta electrodermica, incremento cardiovascular fásico de la presión sanguínea relativa.
Selección de CQ	Las RQ se aparean con las CQ de acuerdo con los procedimientos tradicionales utilizados por los examinadores de campo para cada formato diferente de prueba poligráfica. En general, las RQ se comparan con la CQ anterior o posterior que tenga el mayor cambio en la actividad fisiológica siempre que sea posible, y con la CQ anterior cuando no haya dos CQ disponibles.
Transformación numérica	Las puntuaciones enteras se asignan comparando las diferencias en la magnitud de la respuesta para los pares de RQ y CQ. Los pares de preguntas se pueden utilizar de manera ingenua o pueden estar sujetos a coeficientes de optimización para reducir las puntuaciones que pueden producirse a causa de ruidos espurios o aleatorios. Las puntuaciones enteras del EDA se duplican antes de la sumatoria, de modo que la contribución estructural de los datos EDA es mayor que la de los otros sensores. [Ver Nelson, 2019].
Reducción de datos	Las puntuaciones enteras se suman para cada RQ, para todos los sensores entre gráficas. Luego, los subtotales de las RQ se suman para obtener una puntuación de gran total. La sumatoria a través de este proceso significa que los valores están disponibles para formatos de prueba de asunto único y asuntos múltiples.

Función de verosimilitud	Las funciones de verosimilitud del ESS se calcularon empíricamente usando solo los sensores de respiración, EDA y cardiovascular. Las funciones de verosimilitud multinomial del ESS se calculan matemáticamente utilizando la teoría analítica del CQT y están disponibles para exámenes poligráficos de asunto único y asuntos múltiples con 2, 3 y 4 RQ con 3, 4 o 5 gráficas, incluidos los sensores de respiración, EDA y cardiovascular, además del sensor vasomotor como opcional. La verosimilitud de las puntuaciones de corte tradicionales son las tasas empíricas de TP, TN, FP y FN.
Reglas de decisión: asunto único	GTR, TSR, FZR
Reglas de decisión: asunto múltiple	SSR
Comentarios	El ESS y el ESS-M se introdujeron como métodos de calificación manual. Las publicaciones de investigación han hecho uso de modelos ESS y ESS-M totalmente automatizados.