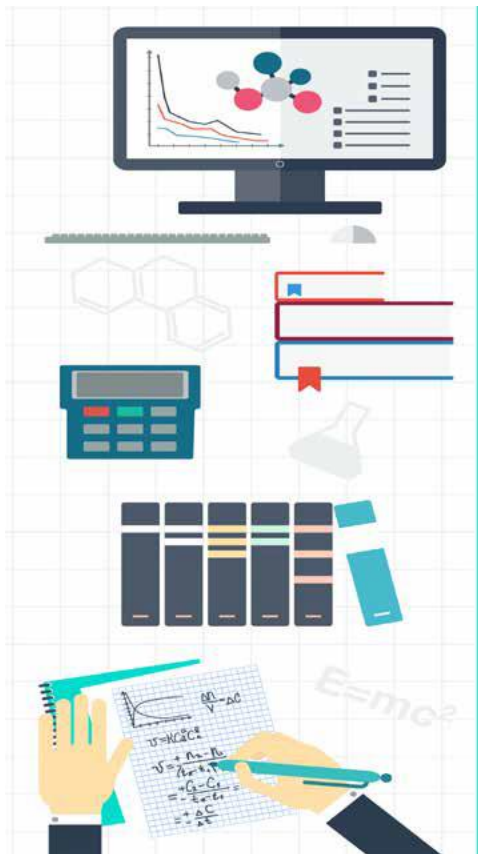


La Prueba Poligráfica como Experimento Científico en Un Solo Sujeto

Por Raymond Nelson



Las pruebas científicas son un proceso tanto de clasificación como de inferencia. Las pruebas científicas se pueden considerar como una forma de investigación o experimento en un solo sujeto, donde los datos, los puntajes y el resultado de la prueba probabilística son la evidencia que apoya una opinión o conclusión profesional para clasificar la prueba como perteneciente a una de las múltiples categorías dentro de los posibles resultados. Las opiniones profesionales y las

conclusiones científicas son por definición, declaraciones que tienen soporte en la evidencia.

En las pruebas e investigaciones científicas, la evidencia se expresa comúnmente a manera de información estadística o probabilística basada en un análisis replicable, numérico y cuantitativo de los datos de la prueba. Las conclusiones sin evidencia de soporte objetiva y cuantificable son solamente opiniones personales, incluso cuando las ofrece un profesional. A menudo, no existen soluciones deterministas perfectas, inmunes al comportamiento humano y a los efectos del azar aleatorio, para muchos fenómenos interesantes e importantes del mundo real (por ejemplo, la personalidad, el funcionamiento intelectual, rapport interpersonal y la discriminación entre el engaño y la veracidad). Un resultado de prueba estadísticamente significativo respalda una conclusión categórica o profesional, y es análogo a un resultado estadísticamente significativo de un experimento científico.

Discusión

Antes de continuar, sería útil recordar que las conclusiones científicas siempre ocurren con respecto a alguna alternativa. Las conclusiones científicas comienzan con sugerencias o hipótesis. Una hipótesis es

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor. rodolfo@poligrafia.com.mx

APA Magazine 2015, 48(5)

una forma de conjetura, especulación o explicación de la observación de algunos fenómenos. En la práctica, las hipótesis científicas son preguntas que deben ser apoyadas o refutadas por la evidencia. Una opinión personal, incluyendo las opiniones profesionales que se ofrecen sin un análisis cuantitativo replicable, puede considerarse una forma de hipótesis no investigada.

Prueba de significancia de hipótesis nulas.

La tradición de la prueba de hipótesis - formalmente llamada prueba de significancia de hipótesis nula - involucra la comparación de la fuerza de la evidencia de que una conclusión o hipótesis es correcta contra la fuerza de la evidencia de que la conclusión o hipótesis es errónea, o que los datos observados podrían haber ocurrido debido a causas aleatorias o no controladas. El nombre formal de la hipótesis es hipótesis alterna. El propósito de una hipótesis alterna es intentar proporcionar una explicación sistemática de algunos datos o fenómenos observables (es decir, cómo funciona el universo). En contraste con esto se encuentra la hipótesis nula que dice que los datos o fenómenos observados pudieron haber ocurrido como resultado de factores aleatorios o no controlados. Como corolario de la noción de que los datos o fenómenos observados se deben a causas aleatorias o no sistemáticas, la hipótesis alterna no puede sustentarse como una explicación sistemática de los datos o fenómenos observados.

Dicho formalmente, una hipótesis tiene soporte cuando la hipótesis nula puede rechazarse porque no es probable que

hubiera causado los datos observados. Las hipótesis que están respaldadas por la evidencia se pueden retener para ser consideradas como conclusiones o explicaciones sistemáticas potencialmente útiles, de los datos o fenómenos observados (es decir, de cómo funciona el universo). Una hipótesis no tiene soporte si no hay suficiente evidencia para rechazar la hipótesis nula. Los conceptos básicos de la prueba de hipótesis se pueden aplicar a una sola observación de prueba, de manera similar a la forma en que se aplican en una muestra de investigación. El estándar de inferencia clásico, frecuentista, es que una conclusión o hipótesis tiene soporte cuando la probabilidad de que los datos observados ocurrieron debido a causas distintas a la conclusión sistemática es suficientemente baja, habiendo decidido antes de la prueba una tolerancia máxima al error.

Los principios de la ciencia y de las pruebas nos obligan a reconocer que no existe una prueba determinista o perfecta para la que no exista la posibilidad de error. El objetivo de una prueba o experimento científico es cuantificar el margen de error con respecto a una posible conclusión. Las pruebas de hipótesis, independientemente de si se trata de un caso único o de un grupo de casos, comienzan con una declaración explícita de nuestra tolerancia al error. Descuidar esto alentaría expectativas equivocadas de perfección.

Alfa, valor-p y significancia estadística.

El límite alfa (α) es el término estadístico común que se utiliza para describir el nivel de tolerancia al error. Alpha a menudo se establece en $\alpha = .05$, que se refiere a una

tolerancia de error del 5%. Algunas circunstancias podrían justificar un límite alfa más restrictivo; en esos casos se puede utilizar $\alpha = .01$ para limitar la tasa de error observado a un nivel del 1%. Por supuesto, habrá una disminución en la tasa de clasificación para los casos que van a ser estadísticamente significativos con este nivel. Bajo diferentes circunstancias podría obtenerse un beneficio por la reducción de los resultados que no sean estadísticamente significativos o inconclusos, en cuyo caso el alfa se puede establecer a $\alpha = .10$ para lo que se puede esperar que la tasa de error observado se limite a menos del 10%, al tiempo que se incrementa el número de casos que se clasifican como estadísticamente significativos.

Los límites alfa siempre se establecen a priori (es decir, antes de iniciar una prueba o experimento). De lo contrario, existe un mayor riesgo de tener errores y resultados de prueba manipulados. En el contexto poligráfico, el alfa es análogo al puntaje de corte, para el que existe un puntaje de corte para clasificaciones de engaño y un puntaje de corte diferente para clasificaciones de veracidad. Las puntuaciones de corte numéricas para las pruebas científicas poligráficas se calculan en función del límite alfa deseado y del modelo y datos de referencia que describen las distribuciones esperadas de las puntuaciones de personas con engaño y veraces. En la práctica, los límites alfa y los puntajes de corte del polígrafo se establecerán con mayor frecuencia como una cuestión de política de la agencia.

El valor-p (valor de probabilidad) es el término estadístico común utilizado para describir el cálculo estimado de la

probabilidad de que los datos observados sean el resultado de causas aleatorias o no controladas. Formalmente, un valor-p es la proporción de casos que se espera que produzcan un resultado bajo la hipótesis nula, similar o más extremo que el resultado observado. En la práctica, los resultados de las pruebas e investigaciones científicas también se describen categóricamente en cuanto a si los resultados son significativos (es decir, estadísticamente significativos) o no significativos (es decir, no estadísticamente significativos). Se dice que un resultado es significativo en el nivel alfa (por ejemplo, estadísticamente significativo en el nivel .05) cuando el valor p (probabilidad de error) es menor que alfa (tolerancia al error).

Clasificación: resultados categóricos.

Cuando los resultados probabilísticos son estadísticamente significativos, se nos permite clasificar categóricamente a los resultados de la prueba como positivos o indicativos de la presencia del asunto problema que se está evaluando. Cuando el resultado no es estadísticamente significativo, se nos permite clasificar el resultado como negativo o no indicativo de la presencia del asunto problema que se está evaluando. Estos términos se consideran abstracciones objetivas y neutrales en comparación con términos emocionalmente cargados como aprobar y reprobar.

Es a través de este proceso de inferencia estadística (es decir, que no es probable que los datos observados se deban al azar) que podemos realizar conclusiones profesionales basadas en evidencia

probabilística, sin dejar de ser responsables del hecho de que los resultados de las pruebas y los experimentos científicos son conclusiones acerca de fenómenos amorfos para los que no se puede lograr una observación determinista simple y perfecta ni una medición física/lineal directa. Conceptualmente, el valor-p es un término estadístico que es análogo al puntaje de la prueba de polígrafo y se calcula a partir del puntaje numérico de la prueba contra el modelo de referencia estadística que se puede derivar empíricamente (es decir, datos de muestra o de población) o teóricamente (es decir, mediante hechos basados en pruebas matemáticas y lógica formal).

A menudo, los resultados categóricos de las pruebas poligráficas han involucrado términos descriptivos muy específicos dentro del contexto poligráfico. Una conclusión categórica para los resultados poligráficos positivos podría ser que se indica engaño cuando los datos difieren en un nivel estadísticamente significativo de la distribución de referencia estadística de las personas veraces. En el contexto de polígrafos exploratorios, a menudo afirmamos de manera más general que hay reacciones significativas. De manera similar, una conclusión categórica podría ser que no se indica engaño cuando los datos poligráficos difieren en un nivel estadísticamente significativo del modelo o de los datos de referencia de las personas engañosas. En el contexto de polígrafos exploratorios, a menudo afirmamos que no hay reacciones significativas con el fin de simplificar y aclarar la comunicación entre poligrafistas. Independientemente del esquema de etiquetado que se seleccione, las implicaciones científicas son las mismas

para describir el resultado categórico de la prueba: no existe una prueba perfecta y todos los resultados de la prueba son fundamentalmente probabilísticos. Utilizamos el lenguaje de significancia estadística y de la investigación científica para comunicar el resultado de la prueba de polígrafo probabilístico y categórico. Los resultados categóricos de los exámenes de polígrafo están respaldados por cálculos probabilísticos basados en la puntuación de la prueba y el valor p.

Un examinador que utiliza un formato de prueba poligráfica con preguntas de comparación evaluará una de dos hipótesis, engaño o veracidad, dependiendo de si los mayores cambios en la actividad fisiológica se cargan hacia los estímulos relevantes o los de comparación. Los puntajes numéricos de los resultados positivos, indicativos de engaño en las pruebas de polígrafo con preguntas de comparación, rechazarán la hipótesis nula de que los datos no son significativamente diferentes del modelo de referencia de los veraces. Los puntajes de las pruebas que se dicen negativos, indicativos de veracidad, rechazarán la hipótesis nula de que los datos no son significativamente diferentes del modelo de referencia para el engaño. Formalmente, tanto las conclusiones de engaño como de veracidad son estadísticamente significativas (es decir, positivas), aunque en la práctica las clasificaciones de veracidad se describen pragmáticamente como negativas o indicativas de ausencia de engaño con respecto al problema o asunto que se está evaluando.

En el contexto de la prueba de polígrafo con preguntas de comparación, la hipótesis

alterna de que un examinado ha sido engañoso se evalúa en contra la hipótesis nula de que los datos del examinado no difieren significativamente del modelo o de los datos de referencia de las personas veraces. Esto es especialmente obvio cuando se entiende el contexto de la prueba poligráfica exploratoria en la cual no se conocen incidentes o alegatos con respecto a los problemas comportamentales que se están investigando. De manera similar, la hipótesis alterna de que un examinado ha sido veraz se evaluará contra la hipótesis nula de que los datos de prueba observados no difieren significativamente de los datos o del modelo de referencia para personas engañosas. Una aplicación clásica de los principios de la prueba de hipótesis de la prueba poligráfica de preguntas de comparación sería evaluar la hipótesis de dos colas utilizando una distribución de referencia teórica calculada a partir de datos matemáticos y lógicos. Con esta aplicación, la hipótesis nula podría ser que el patrón de respuestas observado no difiere de la distribución de referencia teórica. La hipótesis alternativa podría afirmar que el patrón observado de cambios fisiológicos se carga de manera sistemática, no aleatoria, diferencialmente ante los estímulos relevantes o de comparación de una manera que se puede demostrar que discrimina entre el engaño y la verdad.

Conclusión

Los principios de la toma de decisiones científicas se han aplicado dentro del contexto poligráfico durante muchos años, desde la primera vez que un examinador ofreció una conclusión u opinión de que un resultado era significativo. Hoy, sin

embargo, somos conscientes del hecho de que el significado no se puede inferir visual o intuitivamente. No se pueden ofrecer declaraciones o conclusiones sobre la significancia sin el cálculo matemático del nivel de significancia estadística. La significancia no se puede determinar mediante una sola presentación de una pregunta de estímulo de prueba, y no se puede determinar sin obtener primero todos los datos de prueba necesarios para su cálculo. Las discusiones acerca de la significancia dentro del contexto científico pertenecen al ámbito del análisis cuantitativo y probabilístico y la conclusión categórica que está íntimamente conectada con la tradición de la prueba de hipótesis.

Se pueden realizar conclusiones categóricas y probabilísticas tanto para polígrafos diagnóstico de eventos específicos, utilizando puntajes de grandes totales y subtotales, como para polígrafos exploratorios de asuntos múltiples, para los cuales generalmente no se utilizan puntajes de grandes totales. En general, las declaraciones y conclusiones de la significancia serán más precisas cuando se basen en una mayor cantidad de datos (es decir, más estímulos de prueba o más iteraciones de los estímulos de prueba), pero serán menos precisos en la medida que se realicen más conclusiones probabilísticas al mismo tiempo (es decir, cuando se utilizan puntuaciones subtotales). Todos los cálculos y conclusiones categóricas de significancia estarán sujetos a las leyes de la probabilidad. Ayudará a la profesión poligráfica a avanzar el que los profesionales de campo pueden estar más familiarizados con los principios básicos de la probabilidad, pruebas y experimentos

científicos. Los objetivos prácticos de la investigación y las pruebas científicas continuarán involucrando tanto la clasificación como la inferencia estadística mientras las pruebas sean necesarias y útiles. No se espera que las pruebas científicas sean perfectas. Se espera que cuantifiquen el margen de error y de incertidumbre de manera que las tasas de error observadas estén dentro de nuestra tolerancia al error establecida.

Aunque la práctica de la prueba de significancia de hipótesis nulas ha estado rodeada de algunas complejidades y controversias, los principios básicos han permanecido incrustados en la tradición científica de manera tal que todos los estudiantes de estadística desde la escuela secundaria hasta la escuela de posgrado deben estudiar la aplicación práctica de estos principios básicos. Prácticamente todos los estudiantes de todos los cursos de introducción a la estadística también aprenden sobre las pruebas generales (es decir, pruebas F y otros métodos) que evalúan un grupo de hipótesis estadísticas como una sola hipótesis sin la necesidad de correcciones estadísticas para manejar y reportar los efectos de la multiplicidad, y el uso de correcciones estadísticas para efectos de multiplicidad como los esperados cuando se utilizan puntajes subtotales en polígrafos con preguntas de comparación.

Aunque nunca se va a requerir que los practicantes de campo realicen los cálculos matemáticos y estadísticos reales, cualquier profesional que desee convertirse en experto en pruebas y en los aspectos analíticos del polígrafo, debe estar familiarizado con el lenguaje y los conceptos de las pruebas científicas. Una comprensión bien

desarrollada de la prueba del polígrafo como una aplicación contextual de los principios de la toma de decisiones estadísticas, puede ayudar a transmitir de manera más eficaz los resultados de la prueba de manera que satisfaga y eduque a otros sobre la base científica y el uso potencial de los resultados del polígrafo en procedimientos legales que de alguna manera se construyen en torno a conceptos de peso similar como el concepto de la evidencia.

Además de una comprensión básica de los conceptos de la inferencia frecuentista (p. ej., valor p, alfa, significancia estadística, etc.), también será útil e importante para el profesional del polígrafo que se familiarice con los conceptos de la inferencia Bayesiana. Los métodos de inferencia Bayesiana han demostrado su utilidad en una gran variedad de contextos, sin embargo, diferentes supuestos a priori pueden llevar a diferentes resultados a posteriori. En la práctica, la inferencia Bayesiana puede proporcionar resultados de probabilidad que para algunas personas pueden resultar más intuitivos que los valores-p y los límites alfa. A diferencia de la inferencia frecuentista, la inferencia Bayesiana no involucra el uso de un alfa arbitrario con un límite de .05, o .01, o .10, pero sí utiliza una declaración explícita a priori de nuestra suposición con respecto a la probabilidad previa o tasa base (es decir, antes de la prueba) de cada conclusión alternativa. Los exámenes de polígrafo también se pueden considerar como un experimento Bayesiano de un solo sujeto, pero eso tendrá que ser objeto de otra publicación.