



Lección de ciencias de cinco minutos:

Uso de la prueba estadística de proporciones para revelar disparidades ocultas

Nelson Nelson

La ciencia es un proceso sistemático que se utiliza para resolver problemas complejos en el mundo natural. La ciencia también se utiliza para comprender el comportamiento humano. La estadística y las probabilidades son el lenguaje matemático de la ciencia, y cualquiera que desee participar, hacer uso o comprender los resultados de experimentos o pruebas científicas enfrentará la obligación de familiarizarse y dominar la discusión probabilística. procedimientos y conceptos probabilísticos.

Las decisiones científicas surgen cuantitativamente de una lógica matemática rigurosa y de la preponderancia o consistencia de la información. Este enfoque reconoce la incertidumbre inherente a muchas conclusiones científicas y se basa en una evaluación rigurosa de la solidez, variabilidad y reproducibilidad de la evidencia que puede respaldar diferentes conclusiones posibles.

De esta manera, la información probabilística proporciona una base para tomar decisiones informadas en un mundo donde la certeza es rara.

Mientras que un enfoque intuitivo o intelectual de la ciencia implica la formulación de nuevos conocimientos o conclusiones basadas en el pensamiento abstracto y conocimientos previos o suposiciones mantenidas, un enfoque empírico de la ciencia se basa en la observación, la experimentación y la observación y análisis de los datos. La ciencia empírica comienza con una hipótesis (a menudo en forma de pregunta o conjetura basada en pensamiento abstracto o inconsistencias en los datos observados o esperados) seguida de la recopilación y el análisis de datos.

En el paradigma de la prueba de significancia de hipótesis nula (NHST), un experimentador obtendrá sistemáticamente evidencia de una condición experimental y luego evaluará la solidez de la evidencia para las diferentes conclusiones posibles. Las diferentes conclusiones posibles a menudo se describen como la hipótesis nula (H_0) que establece esencialmente que no existen diferencias reales entre la evidencia observada.



como la frecuencia de algún evento, y la hipótesis alternativa (H_1 o H_A) que afirma que existe una diferencia real en las observaciones. La H_A es una declaración explícita de expectativa de que los datos observados, cuando se transformen en cantidades numéricas, serán estadísticamente significativos: que los datos observados serán inconsistentes con la información generada por el azar. En otras palabras, en el paradigma NHST, cuando la probabilidad de obtener los datos observados debido únicamente a la variación aleatoria es suficientemente pequeña, se puede suponer que los datos observados se deben a una diferencia real en las condiciones experimentales. En este caso se puede rechazar la H_0 y se puede decir que la H_A está respaldada por evidencia¹.

evaluar los datos observados o la evidencia sobre si las diferencias en la frecuencia de los eventos observados son consistentes con la variación aleatoria. Independientemente de si hay dos grupos o más, siempre hay dos conclusiones posibles: que las diferencias observadas entre los grupos están dentro del rango esperado que puede atribuirse a la variación aleatoria, o que son estadísticamente significativas. Los TOP pueden ser herramientas esenciales en el análisis de datos siempre que sea necesario investigar diferentes conclusiones posibles sobre las diferencias en las proporciones observadas entre grupos o categorías.

Aproximación de la prueba Z para la prueba de Proporciones de dos muestras

Introducción a la prueba de proporciones

Un método muy simple y útil para evaluar la fuerza probabilística de la evidencia para diferentes conclusiones posibles es una Prueba estadística de proporciones (TOP). Los TOP se pueden calcular de diferentes maneras, incluido el uso de la aproximación de la prueba z para dos grupos de muestras y mediante la prueba de chi-cuadrado (χ^2) para tres grupos de muestras. Los TOP están acostumbrados a

Una prueba z es una prueba estadística común que utiliza la distribución normal estándar (distribución gaussiana o curva de campana²) que tiene una media de 0 y una desviación estándar de 1. Muchos fenómenos naturales se ajustan a esta distribución. Debido a que las propiedades matemáticas de esta distribución son bien conocidas, para cualquier muestra de datos distribuidos normalmente es bastante fácil calcular la proporción de valores de datos que son mayores o menores que

¹ Sin embargo, no es lo mismo afirmar que una H_A está respaldada por evidencia que afirmar que la H_A ha sido probada o está respaldada por pruebas matemáticas y lógicas. Gran parte del conocimiento científico disponible existe en forma de teorías respaldadas por evidencia y por las cuales se reconoce que nuestro conocimiento actual sigue siendo incompleto. Casi siempre hay más que aprender sobre la realidad, el universo y el comportamiento humano.

² Esta distribución también se conoce a veces como distribución de Quetelet, en honor a Adolphe Quetelet, un estadístico belga que la utilizó para describir las características físicas y sociales humanas durante la primera mitad del siglo XIX. Se le conoce comúnmente como gaussiano porque Carl Friedrich Gauss realizó un extenso trabajo con la distribución en el contexto de observaciones astronómicas a principios del siglo XIX. También se le llama curva de campana debido a la característica forma de campana de la distribución. Aproximadamente el 68 % de los valores están dentro de ± 1 desviaciones estándar de la media, con aproximadamente el 95 % de los valores dentro de ± 2 desviaciones estándar y aproximadamente el 99 % de los valores dentro de ± 3 desviaciones estándar de la media. Y aunque otros tipos de distribución también son muy útiles, la distribución normal se ha utilizado ampliamente en prácticamente todas las áreas de la ciencia y el análisis de datos.

La prueba z es una prueba estadística común que utiliza la distribución normal estándar (distribución gaussiana o curva ll2) que tiene una media de 0 y una desviación estándar de 1. Muchos fenómenos que ocurren naturalmente se ajustan a esta distribución. Porque las propiedades matemáticas de esta distribución son más conocidas que cualquier valor individual. Un aspecto muy convencional de los datos TOP es binomial involucrando cualquier muestra de datos distribuidos normalmente, es bastante fácil calcular la proporción de valores de datos de esta distribución es que el número de observaciones y posibilidades son mayores o menores que cualquier valor individual. Un aspecto muy conveniente de esta distribución es que se ha publicado en tablas de valores z. Con un conjunto de datos de tamaño suficiente, las tablas de esta distribución pueden ser ampliamente utilizadas en todos los microcomputadores actuales (incluidos los libros móviles más recientes de esta distribución es que prácticamente con o formuladas con funciones integradas de esta distribución que prácticamente todas las funciones para calcular los valores z. mate la distribución normal estándar. 3 microcomputadora a dispositivos informáticos, tabletas, computadoras portátiles y computadoras de escritorio) incluye herramientas de cálculo bien distribuidas convenientes en lugar de una distribución binomial. En este caso, los datos TOP son binomiales: involucran el número de observaciones y posibilidades. sustituido por el cálculo de herramientas de cálculo reales con funciones integradas, con un conjunto de datos de tamaño suficiente, los valores discretos de la distribución binomial son reemplazados por las probabilidades reales asimptóticas.

con o aproximarse a la distribución normal estándar. Por esta razón, los cálculos estadísticos de la distribución normal por el cálculo de las probabilidades binomiales reales. Comparación, calórica de la diferencia entre Aproximación, en comparación matemática y estadística de la diferencia este uso, se refiere normalmente, el TOP implica la entre dos proporciones de muestra. Las formulas para el uso de una distribución conveniente en proporciones simples. Las TOP se muestra en las Figuras 1, 2 y 3.

Figura 1. Aproximación de la prueba Z para la prueba de proporciones.

Figura 1. Aproximación de la prueba Z para la prueba de proporciones.

Figura 2. Cálculo del error estándar agrupado.

$$Z = \frac{(p_1 - P_2)}{SE}$$

Figura 2. Cálculo del error estándar agrupado.

Figura 2. Cálculo del error estándar agrupado.

$$SE = \sqrt{(p(1-p) * (1/n_1 + 1/N_2))}$$

Figura 3. Cálculo de la proporción agrupada.

Esta distribución también se conoce a veces como distribución de Quetelet, por Adolphe Quetelet, un estadístico belga. Figura 3.

quien lo utilizó para describir características físicas y sociales humanas durante la primera mitad del siglo XIX. Se le conoce comúnmente como gaussiano porque Carl Friedrich Gauss realizó un extenso trabajo con la distribución en el contexto de las observaciones astronómicas a principios del siglo XIX. También se le llama curva de campana debido a la característica forma de campana de la distribución. Aproximadamente el 68% de los valores están dentro de +/- 1 desviaciones estándar de la media, con aproximadamente el 95% de los valores dentro de +/- 2 desviaciones estándar y aproximadamente el 99% de los valores desviaciones de la media. Aproximadamente el 99% de los valores dentro de +/- 3 estándar. Figura 3. Cálculo de la proporción agrupada.

Asintótico, en este uso, significa "en el límite" o "cuando n llega al infinito". Entonces, la distribución binomial es asintóticamente normal cuando n tiende a infinito. Pero para cualquier n fijo, sigue siendo una distribución discreta (sin fracciones significativas o útiles entre los valores discretos) mientras que la distribución normal es continua (con un número infinito de 3). Asintótica, en este uso, significa "en el límite" o "cuando n va al infinito". Por lo tanto, la distribución binomial es asintóticamente y potencialmente valores fraccionarios más pequeños entre cada elemento de la escala). normal cuando n tiende a infinito. Pero para cualquier n fijo, sigue siendo una distribución discreta (sin fracciones útiles o significativas) entre los valores discretos) mientras que la distribución normal es continua (con un número infinito de valores fraccionarios potencialmente continuamente más pequeños entre cada ítem de la escala).

• p1 y p2 son las proporciones de la muestra (el número de eventos observados comparado con el número de eventos

Dónde:

- p1 y p2 son las proporciones muestrales (el número de eventos observados comparado con el número de eventos posibles)
- n1 y n2 son el número de eventos posibles (los tamaños de grupo de muestra)

El valor z (de la fórmula de la Figura 1) se puede utilizar para obtener un valor p de la distribución normal estándar. Y el valor p resultante se puede interpretar como la probabilidad de obtener los datos observados bajo la H0 (si la H0 fuera correcta).

Si el valor p es suficientemente pequeño (menos que un nivel alfa o tolerancia para el tipo 1)

error, que se indicó antes de obtener los datos y antes de completar los cálculos, entonces se puede rechazar la H0 y se puede decir que los datos respaldan el HA. Por el contrario, la H0 no puede rechazarse si el valor p excede el nivel alfa, en cuyo caso los datos no respaldan la HA. Sin embargo, esto sólo indica que la H0 no puede rechazarse y no quiere decir que los datos hayan demostrado que la H0 sea correcta.

Prueba X2 de proporciones para tres muestras

El TOP se puede ampliar a tres o más muestras mediante la prueba de chi-cuadrado (X2). Formalmente, la prueba X2 se utiliza para evaluar la independencia de las categorías muestrales. La Figura 4. muestra la fórmula del X2 TOP.

Figura 4. Prueba de proporciones X2 .

Figura 4. Prueba de proporciones X2 .

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Dónde:

- O es la frecuencia observada de ocurrencia, y E es la frecuencia esperada bajo H0

- E es la frecuencia esperada bajo H0

Aplicación histórica del TOP: abordar el sesgo de género en las admisiones universitarias

Aplicación Histórica del TOP:

Abordar el prejuicio de género en la universidad
Admisiones

Un ejemplo histórico de la aplicación del TOP implicó abordar el género

Un ejemplo histórico de la aplicación del TOP implicó abordar el sesgo de género en la universidad

Sesgo en las admisiones universitarias durante la década de 1970. Durante este período, surgieron preocupaciones sobre posibles sesgos de género en las prácticas de admisión de colegios y universidades en los Estados Unidos, específicamente en la Universidad de California en Berkeley. Se sospechaba que las mujeres se enfrentaban a discriminación, lo que daba lugar a una representación desigual en la educación superior. Posteriormente, estadísticos e investigadores aplicaron el TOP, específicamente la prueba z para dos muestras, para evaluar si había diferencias estadísticamente significativas en las proporciones de solicitantes masculinos y femeninos admitidos en instituciones educativas.

Los resultados iniciales que utilizaron datos de todos los departamentos académicos mostraron que las diferencias en las tasas de admisión a la universidad eran estadísticamente significativas y que los hombres eran admitidos con mayor frecuencia. Sin embargo, los resultados de cada departamento muestran una inversión de esta tendencia debido a las diferencias en la frecuencia (número) de solicitantes masculinos y femeninos en diferentes departamentos. Las mujeres postularon con mayor frecuencia a departamentos más competitivos, donde el rechazo era más probable, mientras que los hombres postularon con mayor frecuencia a departamentos con admisiones menos competitivas⁴. Aunque en general se admitió al 44% de los solicitantes masculinos, en comparación con el 35% de las solicitantes femeninas, cuando se analizaron los departamentos individuales, ningún departamento discriminó a las solicitantes femeninas, con un sesgo pequeño pero estadísticamente significativo a favor de las solicitantes femeninas. La tendencia de las mujeres a superar a los hombres en

Las admisiones y graduaciones universitarias se han observado de manera constante desde finales de la década de 1970, aunque las mujeres están subrepresentadas en algunas áreas de estudio.

Aplicación del TOP al Polígrafo Análisis de contramedidas

La técnica de preguntas de comparación (CQT) es una forma de prueba de evaluación de credibilidad científica (detección de mentiras) que se basa en la comparación de respuestas fisiológicas a preguntas relevantes (RQ) sobre el evento en cuestión con respuestas a preguntas de comparación (CQ). Todas las formas de pruebas científicas se ocupan de la validez de constructo de los datos y resultados de las pruebas. Las pruebas de polígrafo no detectan ni miden el engaño per se, simplemente porque el engaño es una construcción amorfa). En cambio, las pruebas de polígrafo se basan en señales autónomas que se correlacionan con el engaño en el CQT y para las cuales se pueden desarrollar modelos estadísticos utilizando combinaciones de diferentes señales. Porque se utilizan pruebas científicas de todo tipo para cuantificar fenómenos que no pueden estar sujetos a una observación determinista perfecta (inmune al comportamiento humano y no afectada por variaciones aleatorias) o a una medición física (que requiere un fenómeno físico y estaría sujeta a pruebas científicas de todo tipo). sólo al error de medición aleatorio), los resultados de las pruebas científicas son inherentemente probabilísticos y se basan en la relación estadística entre un fenómeno de interés que no se puede medir físicamente y señales proxy que están disponibles para su registro y medición. hasta el grado

⁴ Esta inversión de las tendencias estadísticas para los datos de todo el grupo y de los subgrupos se conoce como paradoja de Simpson.



Como toda actividad fisiológica está asociada con múltiples formas de actividad humana, las pruebas poligráficas, como muchas pruebas científicas, pueden incluir algunas vulnerabilidades a la manipulación o falsificación estratégica, y se pueden utilizar métodos estadísticos para monitorear y reducir tales vulnerabilidades.

Una dificultad para aquellos que intentan manipular o falsificar voluntariamente los datos de sus pruebas de polígrafo será que existe cierta probabilidad de que la actividad de falsificación voluntaria pueda producir datos que sean cualitativa o cuantitativamente distintos de la actividad autónoma normal. Los examinadores de polígrafo se refieren a las señales de datos inusuales o atípicas como "artefactos" o, más generalmente, como "actividad fisiológica atípica". Una dificultad relacionada es que los artefactos de datos a veces pueden ser el resultado de causas involuntarias o inocentes, y la mera presencia de artefactos de datos puede generar sospechas o acusaciones de falsificación en algunas circunstancias. Debido a que el CQT es un procedimiento estandarizado y sistemático, los esfuerzos de falsificación también deben ser sistemáticos para que sean efectivos. Es decir, una simulación exitosa en el CQT debe revertir sistemáticamente la carga de cambios en la actividad fisiológica que ocurren en respuesta a diferentes tipos de estímulos de prueba según la teoría analítica de la prueba del polígrafo. Así, si bien una actividad sistemática constante puede tener el mayor potencial para alterar los valores numéricos resultantes, también tendrá la mayor vulnerabilidad para ser observada fácilmente. Esfuerzos de falsificación más sofisticados podrían implicar el uso de una variedad de actividades estratégicas que se ejecutan de manera inconsistente o pseudoaleatoria y que pueden ser menos fáciles de observar, pero que, si no son lo suficientemente sistemáticas, también

lograr la inversión deseada de la carga de los datos fisiológicos.

Una dificultad para los practicantes de campo del polígrafo es que, si bien la mera observación de una actividad inusual puede no ser difícil, los intentos de hacer atribuciones precisas sobre la motivación o la intención de dicha actividad requieren una visión precisa de la motivación o el estado del examinado. mental: si los artefactos observados son sistemáticos y deliberados o aleatorios e involuntarios. Muchos artefactos del polígrafo pueden producir datos igualmente atípicos, independientemente de si son voluntarios y sistemáticos o involuntarios. Si fuera posible obtener una visión precisa del estado mental del examinado, podría excluirse la necesidad de realizar una prueba de polígrafo. Además, al considerar las complicaciones actualmente insuperables que rodean la posibilidad de leer la mente en la actualidad, se vuelve problemático, y por lo tanto incorrecto, intentar hacer atribuciones de que los artefactos observados son voluntarios o sistemáticos cuando la atribución a un No se ha descartado de manera científica aceptable una causa inocente y plausible.

El enfoque científico común ante la necesidad de tomar decisiones bajo circunstancias que están sujetas a una incertidumbre inherente es utilizar métodos estadísticos para cuantificar el margen de incertidumbre o nivel de confianza que los datos pueden proporcionar en apoyo de las diferentes opciones posibles. conclusión. El TOP proporciona un medio viable para investigar la posibilidad de falsificación sistemática mediante la evaluación de la frecuencia y las proporciones de los artefactos de datos que ocurren en respuesta a diferentes tipos de estímulos de prueba. Si TOP indica una diferencia significativa en las proporciones de artefactos de datos durante los CQ en comparación

sugiere que pueden estar en uso contramedidas. Esta información puede ser valiosa para evaluar la confiabilidad de los resultados de las pruebas del polígrafo y ajustar las estrategias de interrogatorio en consecuencia.

Conclusión

La ciencia es un proceso estructurado que ayuda a resolver problemas complejos en el mundo natural y comprender el comportamiento humano.

Utiliza estadísticas y probabilidades como lenguaje matemático. Las conclusiones científicas, con incertidumbres inherentes, se derivan de una lógica matemática rigurosa y de información consistente. Las probabilidades forman la base de las decisiones informadas en

un mundo donde la certeza es difícil de alcanzar. Mientras que la ciencia intuitiva se basa en suposiciones y pensamientos abstractos, la ciencia empírica se basa en la observación y el análisis de datos. La prueba de significancia de hipótesis nulas (NHST) es un método en el que se utiliza evidencia de experimentos para evaluar la validez de hipótesis nulas y alternativas. La Prueba de Proporciones (TOP) es un método estadístico para analizar la solidez de la evidencia. Puede ejecutarse mediante pruebas z para dos muestras o la prueba chi-cuadrado (χ^2) para tres o más muestras. Históricamente, el TOP se ha utilizado para abordar los prejuicios de género en las admisiones universitarias y tiene aplicaciones en el análisis de contramedidas poligráficas.



Referencias

- Agresti, A. (2002). Análisis de datos categóricos (2ª ed.). John Wiley & Sons, Inc.
- Agresti, A. y Coull, BA (1998). Aproximado es mejor que "exacto" para la estimación de intervalos de proporciones binomiales. *El estadístico americano*, 52(2), 119-126.
- Bickel, PJ, Hammel, EA y O'Connell, JW (1975). Sesgo sexual en las admisiones de posgrado: datos de Berkeley. *Ciencia*, 187(4175), 398-404.
- Bishop, YM, Fienberg, SE y Holland, PW (1975). Análisis multivariado discreto: el-teoría y práctica. Prensa del MIT.
- Bruning, J. L y Kintz, BL (1997) Manual computacional de estadística. (4ª edición). Perahijo.
- Cochran, WG (1954). Algunos métodos para fortalecer las pruebas χ^2 comunes. *Biometrics*, 10(4), 417-451.
- Cohen, J. (1994). La Tierra es redonda ($p < .05$). *Psicólogo estadounidense*, 49(12), 997.
- Conover, WJ (1999). Estadística práctica no paramétrica (3ª ed.). John Wiley e hijos.
- Cumming, G. (2014). Las nuevas estadísticas: por qué y cómo. *Ciencia psicológica*, 25(1), 7-29.
- Everitt, BS (1977). El Análisis de Tablas de Contingencia. Chapman y Hall.
- Pescador, RA (1935). El diseño de experimentos. Edimburgo: Oliver y Boyd.
- Gentil, JE (2009). Estadística Computacional. Nueva York: Springer.
- Honts, CR (2004). La prueba de preguntas de comparación. En PA Granhag y LA Strömwall (Eds.), *La detección del engaño en contextos forenses* (págs. 14-40). Prensa de la Universidad de Cambridge.
- Neyman, J. y Pearson, ES (1933). Sobre el problema de las pruebas más eficientes de hipótesis estadísticas. *Transacciones filosóficas de la Royal Society de Londres. Serie A*, que contiene artículos de carácter físico o matemático, 231 (694-706), 289-337.