

47 Pruebas de significación

¿Cómo saber si el problema realmente está declinando? La mayoría de los problemas varía en intensidad, incluso cuando nadie está respondiendo a ellos. Por ejemplo, en promedio hay nueve robos de vehículos en el centro de alguna ciudad, pero raramente hay semanas en las que ocurre esa cantidad exacta de robos. En cambio, 95% de las semanas tienen entre 7 y 11 robos, y en 5% de las semanas menos de 7 o mayores a 11. Tal variación aleatoria es normal. Claramente, una reducción del robo de vehículos de un promedio de nueve por semana a un promedio de seis por semana está dentro de la norma, y podría ser debido exclusivamente a la aleatoriedad. En otras palabras, la respuesta al problema puede no haber tenido impacto alguno. Si el cambio en el problema es reducido, la variación aleatoria es una explicación razonable. Si el número de robos cayó a aproximadamente cero a la semana, con algún robo ocasional a la semana, es bastante probable que no haya sido un resultado de la implementación; ya que es anormal. ¿Cómo podemos saber si el cambio en el problema está dentro de su variación normal?

La respuesta es usar una **prueba de significación**. Una prueba de este tipo compara la variación en el problema antes y después de la implementación. Dice la probabilidad aleatoria de cambio en el problema. Su implementación puede haber causado el cambio si resulta una baja probabilidad aleatoria de cambio. Si hay menos de 5% de oportunidad de que el problema haya cambiado debido a fluctuaciones azarosas, rechazamos la explicación de aleatoriedad como una causa del cambio. Aquí, 5% es el nivel de importancia, y es una selección convencional. En suma, “apostamos” que otra cosa distinta a la aleatoriedad causó el cambio, si hay una probabilidad lo “suficientemente baja”, la aleatoriedad podría ser la causa. Esta probabilidad se define de acuerdo al nivel de importancia. A menudo nuestra opción es 5%, pero es posible escoger otros niveles.

Se puede escoger un nivel más severo, como 1%. Mientras más rígido sea el nivel de importancia seleccionado, es mayor la probabilidad de concluir que la implementación fue ineficaz aun así haya funcionado. Pero esta opción puede usarse si el costo de la implementación es tan alto que necesita tener plena certeza de su funcionamiento. De vez en cuando, los analistas usan niveles menos rígidos, como 10%. Mientras más flexible sea el nivel elegido, hay más posibilidades de considerar exitosa una implementación que de hecho pudo haber fracasado. Se podría querer escoger un nivel menos rígido si el problema es serio, las medidas de solución no son particularmente buenas, y usted está muy interesado en la cuestión de que surja accidentalmente una buena implementación. Hay dos maneras de usar los niveles de importancia. Los científicos sociales típicamente los usan como umbrales de rechazo: debajo del

nivel usted rechaza la oportunidad y sobre el nivel usted la acepta como causa. En resumen, las pruebas de importancia hacen la decisión. Si usted sigue esta estrategia, siempre escoja el nivel de importancia antes de calcularlo. De hecho, debe seleccionarse al iniciar el proceso de evaluación para evitar “entrampar” los cálculos del resultado deseado.

Un mejor acercamiento es usar las pruebas de significación como auxiliares en la toma de decisiones. Los tomadores de decisiones usan la información de las pruebas de significación junto con otros hechos (la gravedad del problema, el programa, sus costos, la reducción absoluta de los problemas y varias más) para tomar una medida informada. Muchas ciencias, como la medicina, siguen este acercamiento. Si usted lo hace también, use un **valor p** en lugar de la prueba de importancia. El valor p es una probabilidad exacta de que el cambio en el problema se debe a la oportunidad. Así, un valor p de 0.062 le dice que la oportunidad de cometer un error si acepta la implementación es una porción de más de 6%. Esto puede interpretarse bruscamente diciendo que de 100 decisiones tales, la decisión que rechaza la aleatoriedad en favor de la implementación estará equivocada aproximadamente 6 veces. ¿Usted o sus colegas tomarían tal riesgo? depende de muchas cosas, ¿no lo cree?

Hay muchos programas estadísticos que pueden hacer los cálculos requeridos. La dificultad que usted enfrentará es cómo escoger entre tantos métodos para determinar el nivel de importancia o el valor p. Usted puede dejar al programa definir los valores por defecto, pero no hay ninguna garantía de que estos correspondan a su situación. La dificultad radica en las presunciones ocultas detrás de los cálculos. Las siguientes son algunas consideraciones importantes.

La variación en un problema es llamada **distribución**. En el ejemplo sobre la distribución de semanas (durante un periodo largo) con cero robos, un robo, dos robos, y así. Muchos programas observan una campana de distribución **normal**. Esto a menudo no es apropiado porque una distribución normal asume que los números son continuos (como el ingreso de las personas, o los índices delictivos) en lugar de los eventos discretos (como las estadísticas delictivas). Si usted está usando datos brutos, verifique que en el programa que se piensa utilizar hay una opción para usar algo conocido como **distribución binomial**.

También, las distribuciones normales son simétricas, con el promedio en el centro. La mayoría de los datos delictivos están muy distorsionados – algunas personas, lugares u horarios concentran la mayoría de los eventos delictivos y muchas personas, lugares o tiempos tienen poco o ningún evento (vea Paso 19). Esto es porque el delito relativamente raro, particularmente en intervalos de tiempo reducidos y lugares pequeños. Esto también implica la oportunidad de representar más fácilmente las fluctuaciones. Para encontrar un cambio significativo en un problema cuando los eventos son raros, tendrá que indagar

más profunda y largamente antes y después de la implementación. Esto típicamente significa más periodos de tiempo. En este caso, la distribución de **Poisson** es un mejor método cuando se trata de eventos raros.

La investigación de oportunidad puede volverse muy compleja. Si resulta muy importante el resultado de una prueba de importancia o un valor p y usted no es experto en teoría de probabilidad o estadísticas, debe buscar ayuda especializada de las universidades locales u otras organizaciones que usan estadísticas regularmente.