

Centro Universitario de Estudios Medioambientales

Seminarios de la reunión semanal del CUEM

Seminario: 2026-05-04

Expositor: Alfredo Rigalli

Tema: Cuando lo significativo puede ser insignificante

No se existe investigación si no existen datos. Un dato es un valor numérico, cualidad, resultado de análisis y prácticamente cualquier cosa surgida de la aplicación del proceso de medición y observación de un sistema. Recordemos que los pilares de un dato son tres. El contexto: un dato es acompañado de metadatos. La objetividad: un dato debe ser reproducible. El registro: el dato debe perdurar en algún soporte. Un dato es dato si permite u obliga a tomar una decisión. Ejemplo: concentración de arsénico de una muestra de agua = 5 ppb, agua de potable. Arsénico = 250 ppb, no consumirla.

Un dato nunca está solo, está acompañado de otros datos y de metadatos que forman su contexto. Los datos y los metadatos forman bases de datos. Estas bases de datos están a su vez asociadas a un dado suceso, a un procedimiento, a una o varias instituciones, son bases de datos globales, etc. La base de datos es definida por el individuo que lo maneja. Las bases de datos pueden pertenecer a small data, si ocupan poco espacio de memoria en una PC. Medium data si ocupan más lugar y cuando el espacio ocupado llega a los Terabytes podemos decir que se trata de bigdata.

Para llegar a conclusiones o tomar decisiones a partir de datos debemos aplicar herramientas de la estadística. Llamamos inferencia estadística a decisiones tomadas en base a los datos y la aplicación de estas herramientas. Habitualmente decimos que los datos nos indican algo significativo o no significativo, si los datos muestran algo diferente o no de lo conocido. Este significado estadístico se define por el valor del p-value. En general si p-value es menor de 0,05, decimos que es significativo, es decir estamos viendo algo nuevo o diferente a lo conocida. Si p-value es mayor a 0.05, es no significativo, es decir que nuestros datos nos están mostrando algo similar a lo conocido. El valor de p-value es fijado por el investigador según el objeto de estudio pero lo más común al menos en el ámbito que nos movemos es que tome el valor 0,05 o 0,01.

Cuando aumenta el número de datos que se someten a algún proceso de inferencia estadística, la probabilidad de hallar diferencias estadísticamente significativas aumenta, aun cuando esta diferencia es muy pequeña. Si el número de datos es relativamente grande, podemos hallar que la inferencia estadística nos diga que algo es diferente de otra cosa, sin que esa diferencia sea relevante o importante para la toma de decisión. La toma de una u otra decisión se debe hacer si hay diferencias estadísticas entre los datos y esa diferencia es relevante o importante.

Para evaluar la relevancia de la diferencia existen diferentes procedimientos. En este caso tomaremos y analizaremos la d de Cohen. El valor d puede tomar diferentes valores. Si es menor de 0.2 podemos decir que es irrelevante, si es mayor que 0.2 y menor que 0.5 es un efecto menor, si es mayor que 0,5 pero menor que 0,8 es un efecto medio y si es mayor que 0,8 es un efecto importante o muy relevante.

Ejemplos.

1- Comparamos dos grupos de datos utilizando t de Student y d de Cohen. Obtuvimos p-value=0,021 y d=0,98. Podemos decir que la diferencia es estadísticamente significativa y además es relevante o tiene importancia.

2- Comparamos dos grupos y obtuvimos p-value=0,00001 y d=0,11. Podemos decir que hay una diferencia altamente significativa, pero esta diferencia es irrelevante o carece de importancia.