

Centro Universitario de Estudios Medioambientales

Seminarios de la reunión semanal del CUEM

Seminario: 2025-06-30

Expositor: Silvina P. Vaquero

Tema: Introducción a la estadística

La estadística es la rama de la matemática que recopila, organiza, analiza, interpreta y presenta datos, con el objetivo de posibilitar la toma de decisiones. Para ello, se vale de gráficas, tablas, medidas de tendencia central (tales como media, mediana, modo), medidas de dispersión (por ejemplo, rango, varianza, desvío estándar), pruebas de hipótesis, intervalos de confianza, regresiones, etc.

Podemos dividirla en dos grandes grupos, según sus principales características: Estadística descriptiva, que recoge, organiza, presenta y analiza un conjunto de datos para describir sus características principales, sin sacar conclusiones; y Estadística inferencial, que sí posibilita generalizar, estimar y predecir sobre una población, a partir de una muestra.

Si bien es una herramienta extremadamente efectiva, su uso no está exento de equivocaciones al interpretar resultados. ¿Por qué puede pasar esto? Algunos de los motivos incluyen muestras no representativas, errores de muestreo, interpretación incorrecta de los resultados, errores humanos o técnicos, mala elección de los métodos estadísticos, probabilidad de error en las pruebas de hipótesis. Existen dos tipos de probabilidad de error: Tipo I, que representa rechazar una hipótesis verdadera, y Tipo II, que se asocia a no rechazar una hipótesis falsa.

Uno de los motivos asociados a los errores en la interpretación de la estadística es el empleo de muestras no representativas, pero, ¿cuáles son las condiciones para que una muestra sea representativa? Veamos: definir claramente la población a analizar, elegir el tipo de muestreo adecuado, tener un tamaño de muestra adecuado, evitar sesgos y recolectar los datos de forma consistente. Y, ahora, ¿qué tipos de muestreos existen? No probabilístico y probabilístico. El no probabilístico no suele recomendarse cuando se busca representatividad, pero es útil en estudios exploratorios. El probabilístico, por su parte, es más confiable, ya que cada individuo tiene una probabilidad conocida y diferente de cero de ser elegido. A su vez, este último, consta de diferentes métodos, que incluyen: aleatorio simple (selección al azar), estratificado (división por subgrupos y toma de muestra proporcional de cada uno), sistemático (punto de inicio aleatorio y, luego, cada n -ésimo individuo), o por conglomerados (grupos enteros al azar).

¿A qué puede obedecer la incorrecta interpretación de los resultados? A múltiples factores, entre los que se hallan confundir correlación con causalidad, sesgos de la muestra, manipulación de gráficos, mal uso de las medidas estadísticas, omisión del contexto, sesgos del analista, o desconocimiento del significado del *p-value*.

Para concluir, veamos qué significa este valor, el *p-value*, y cómo se interpreta. Sabemos que es un valor empleado en estadística inferencial, que ayuda a diferenciar los resultados que obedecen al azar del muestreo de aquellos que son estadísticamente significativos. En otras palabras, un *p-value* menor a 0,05 (en algunos estudios son más exigentes y aplican un valor inferior a 0,01) significa que es poco probable que el resultado se deba al azar, mientras que un valor superior a 0,05 se asocia a que existe una probabilidad de que el resultado se deba al azar.