

Si te gusta la estadística, bancate los metámeros

Anonymous

No Institute Given

Resumen Keywords: Hasta · tres · palabras clave

1. Introducción

En 1973 Frank Anscombe creó cuatro sets de datos que compartían la media y el desvío de cada variable y su coeficiente de correlación pero que lucían muy distintos cuando se los grafica (ref). Desde entonces, el cuarteto de Anscombe se usa para ilustrar la importancia de visualizar los datos crudos en vez de confiar en los estadísticos. A pesar de eso, no existe mucha investigación sobre el fenómeno general de “datasets disintos con los mismos estadísticos” del cual el cuarteto de Anscombe es sólo un ejemplo. Además usar un conjunto de datos creados hace 50 años para enseñar esto da la impresión de que es un caso único o extraordinario.

En este artículo propongo el nombre de “metámeros estadísticos” en analogía al concepto de colorimetría y presento el paquete metamer, que implementa el algoritmo de blabla (ref) para la creación automática de metámeros.

2. Fuente de metamerismo estadístico

El Demonio de Laplace no sabe ni necesita saber estadística. Él tiene pue- de conocer la posición y velocidad de cada partícula del universo y usar ese conocimiento para predecir su evolución. Pero los seres humanos no podemos analizar más de unos pocos números por vez. Si queremos entender el universo tenemos que resumir grand cantidad de observaciones en unos pocos números. Necesitamos saber estadística.

Ese es el objetivo de la gran mayoría de los métodos estadísticos: conseguir algunos pocos números representativos y comprensibles a partir de sets de datos demasiado grandes para entenderlos. El proceso necesariamente elimina información (distintos métodos priorizan diferentes parte de la información) lo cual implica necesariamente que los mismos pocos números pueden ser producidos por una infinidad de sets de datos distintos. Por ejemplo, se necesitan N momentos estadísticos para caracterizar unívocamente un set de datos de N observaciones¹. Como colorario, existen infinitos sets de datos de N observaciones que comparten los mismos $n < N$ momentos.

¹ Técnicamente unívocamente a menos a menos de una permutación.

Sea una transformación estadística $E : A \rightarrow B$. Voy a llamar metámeros de E a todo elemento de $a \in A$ tal que $E(a) = b_0$ con $b_0 \in B$.

El Cuarteto de Anscombe debe ser entendido bajo esta idea. Va más allá de Si pensamos en una función que