

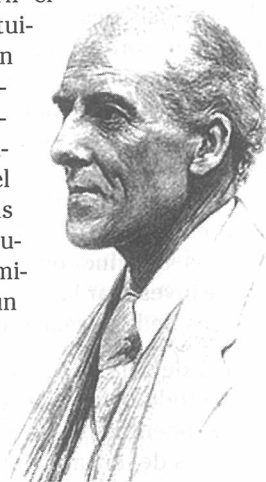
UNA PANORÁMICA DEL ANÁLISIS DE DATOS

Miguel Ángel González Sierra

Por costumbre, se ha considerado la estadística descriptiva y la teórica como dos ramas de la misma con distinta metodología. Años atrás, se trataba de resumir los resultados en términos de “estadística descriptiva” calculando medidas de situación y dispersión, momentos de órdenes más elevados o diversos índices, y también exponer determinadas características de los datos a través de gráficos tales como histogramas, diagramas de barras y gráficas bidimensionales. No se hace referencia al mecanismo estocástico (o distribución de probabilidad) que originó los datos presentados. Las estadísticas descriptivas así efectuadas se utilizan en la comparación de diferentes grupos de datos. Incluso se prescriben diversas reglas para escoger entre estadísticos alternativos, tales como media, mediana y moda, dependiendo de la naturaleza de los datos y de las preguntas que deban ser contestadas. Tales análisis estadísticos son conocidos como análisis descriptivos de datos. En estadística teórica, el objetivo es de nuevo el resultado de los datos, pero con referencia a una determinada familia de distribución de probabilidad (o modelo) subyacente. El resumen estadístico descriptivo, en este caso, depende básicamente del modelo estocástico específico, y las distribuciones de probabilidad correspondientes se utilizan para especificar márgenes de incertidumbre referidos a parámetros desconocidos. Esta metodología se conoce con el nombre de análisis inferencial.

Karl Pearson fue el primero en tratar de cubrir el hueco entre ambos tipos de análisis. Utilizaba la intuición facilitada por el análisis descriptivo basado en los momentos e histogramas para obtener referencias acerca de la familia de distribuciones subyacentes. Para este menester, inventó el primero y quizás el más importante test, consistente en utilizar el estadístico ji-cuadrado, para contrastar la hipótesis de que uno o más datos provenían de una distribución de probabilidad perteneciente a una determinada familia. Dicho test marcó el comienzo de un nuevo modo de tomar decisiones.

Karl Pearson creó un sistema de distribuciones de probabilidad, que se podían generar a partir de sus cuatro primeros momentos. Un hermoso ejemplo de trabajo de investigación fue llevado a cabo por Karl Pearson a través del uso de histogramas y el test ji-cuadrado, consistente en el descubrimiento de que la distribución del tamaño de los tripanosomas hallados en determinados animales, es una mezcla de dos distribuciones normales.



Karl Pearson