

Parte A

Estadística Descriptiva

Introducción a la estadística descriptiva

La primera parte de este libro está dedicada a la estadística descriptiva. Atendiendo a lo que tradicionalmente se ha entendido por descriptiva se estaría hablando de un conjunto de herramientas, formado por coeficientes y técnicas, que tratan de resumir la información contenida en un conjunto de datos. Sin embargo, la estadística descriptiva es mucho más que eso, en realidad es una parte fundamental de cualquier análisis estadístico complejo, en la que se empiezan a tomar decisiones que afectarán al conjunto de la investigación. Los coeficientes descriptivos darán información sobre la estructura de la población que se estudia, indicando, por ejemplo, si ésta es simétrica, si realmente se trata de una única población o hay una superposición de poblaciones, también pueden detectarse valores extraordinariamente raros, etc.

Desde otra óptica, la mayoría de los coeficientes descriptivos tendrán su homólogo inferencial o poblacional, que necesariamente deberán ser estudiados a la luz de aquellos. Haciendo una pequeña abstracción muchos de los coeficientes descriptivos, los más importantes, se convierten en poblacionales al sustituir frecuencias por probabilidades.

En resumen, el análisis descriptivo es una parte inseparable de cualquier análisis estadístico, que puede tener su continuidad con un análisis inferencial cuando los datos que se manejan se corresponden con una muestra probabilística extraída de una población.

Esta primera parte del libro está compuesta por tres capítulos, el

4 *Introducción a la estadística descriptiva*

primero de ellos aborda el problema unidimensional. Se trata de identificar la información que se va a analizar, bien sean variables cuantitativas o de clase, procediéndose a organizarla en distribuciones de frecuencias. Se indica que la primera toma de contacto con las peculiaridades de una distribución se obtiene a través de sus representaciones gráficas y se da al menos una representación para cada uno de los tipos de datos que se manejan. Se calculan todos los coeficientes tradicionales: medidas de centralización, de posición, de dispersión y de forma; se obtienen los momentos respecto al origen y respecto a la media, indicándose que generalizan la mayoría de las medidas anteriores. Se introduce la desigualdad de Tchebychev, poniéndose de manifiesto la relación existente entre la varianza y la media aritmética. Se estudian las transformaciones de variables, haciendo ver que el objetivo es conseguir distribuciones más regulares, que sean comparables, más simétricas; entre todas las transformaciones se dedica especial atención a la normalización o tipificación. Por último, se hace una breve incursión al análisis exploratorio de datos, recurriendo a representaciones, como los diagramas de cajas, que resaltan las regularidades y las especificidades del conjunto de datos, entre las que cabe destacar la presencia de observaciones candidatas a ser valores extraños o anómalos.

El capítulo segundo supone una generalización al caso de que conjuntamente se tenga más de una variable, viéndose con detenimiento el caso bivariable y destacando el hecho de la posible existencia de relaciones entre dichas variables. La existencia de dependencias merece una especial atención por las consecuencias que de ella se derivan en muchas técnicas estadísticas. Se introducen coeficientes que expresarán el grado de relación entre las variables, distinguiendo los casos en que éstas sean continuas, ordenadas o de clase; lo que conduce a definir medidas de correlación, concordancia y contingencia o asociación.

En el último capítulo de esta parte se aborda el problema del ajuste y la regresión en el plano, lo que supone un primer acercamiento a la modelización estadística. El desarrollo del tema se hace planteando un modelo lineal, empleándose para la estimación de los parámetros el método de los mínimos cuadrados. El análisis de la bondad del ajuste se realiza a través del coeficiente de determinación. El método de la

regresión a la media permite calibrar la calidad de los posibles ajustes a realizar. También se analizan algunas extensiones a los casos de modelos linealizables y polinomiales.

