

ANÁLISIS BÁSICO DE DATOS

APLICACIONES A LAS CIENCIAS BIOMÉDICAS Y SOCIALES

Vivimos inmersos en la sociedad de la información y nos vemos abrumados por grandes cantidades de datos procedentes de muy diversas fuentes en las que hemos de apoyarnos para mejorar nuestro conocimiento y tomar decisiones en ambiente de incertidumbre. Para ello, será necesario disponer de ciertos conocimientos básicos que nos permitan ser capaces de diseñar la manera correcta de obtener esta información, de organizar y disponerla de forma que pueda utilizarse de una manera sencilla, de resumirla adecuadamente, de analizarla objetivamente y, en definitiva, de hacer un uso de ella, de tal manera que sea una herramienta efectiva de soporte a la decisión. Todas estas capacidades serán realizadas a través de un apropiado uso de la metodología estadística.

La Estadística ha experimentado en los últimos 150 años una rápida y continua evolución: desde su viejo propósito de elaborar la contabilidad de los estados ha pasado a ser una herramienta fundamental y necesaria en las ciencias empíricas, permitiendo dar respuesta científica a gran cantidad de problemas. Hoy día, un país estadísticamente avanzado se caracteriza: porque la mayoría de sus decisiones sociales y económicas están basadas en estadísticas relevantes, sus medios de comunicación proporcionan información veraz sobre la esencia de la información contenida en informes estadísticos serios y existe un buen sistema de enseñanza de la Estadística en sus diferentes niveles. Cualquier ciencia que precise de un análisis de datos será un potencial campo de aplicación para la Estadística, siendo su intervención fundamental en actividades relacionadas con las administraciones públicas, la economía y las finanzas o la industria y los servicios. En todos estos importantes campos de aplicación, la metodología estadística debe intervenir en la etapa inicial de identificación y formulación de los problemas, en el posterior diseño de la investigación, en la recogida y codificación de la información, en su análisis, en el ajuste y validación de modelos, en la elaboración de resultados y conclusiones, en su presentación y en la propuesta de acciones futuras. En particular, ocupan un lugar destacado los métodos estadísticos descriptivos que nos permiten organizar, representar gráficamente, resumir convenientemente y analizar la información recogida.

El objetivo fundamental de este libro es precisamente proporcionar unos conocimientos básicos pero suficientes sobre esta importante parte de la metodología estadística. Estructurado en diez capítulos, se presentan de forma clara, directa y bien ilustrada, las principales técnicas estadísticas para el análisis descriptivo de los datos. Tras un primer capítulo introductorio en el que el autor nos proporciona información sobre los orígenes y evolución de la Estadística, sus principales partes, su metodología y algunas notas históricas sobre la evolución en nuestro país, se describen, tanto para el caso de una sola variable (capítulos II al IV) como para el caso de dos variables (capítulos V y VI), los métodos estadísticos necesarios para el análisis descriptivo. Se proporciona también información relativa a técnicas estadísticas para la descripción de poblaciones dinámicas (capítulos VII y VIII), el análisis exploratorio de datos (capítulo IX) y, finalmente, un interesantísimo capítulo dedicado al desarrollo de algunas aplicaciones en el campo de la economía, la geografía, el medio ambiente, las ciencias sociales, la educación física, la industria y la medicina.

En definitiva, se trata de un libro esencialmente útil para estudiantes de bachillerato y de titulaciones universitarias que requieren apoyo estadístico y en el que hay que agradecer a su autor el esfuerzo realizado para que resulte ser un libro ameno y atractivo.

A continuación desarrollamos una de las aplicaciones, los Diagramas de Pareto, que figuran en el Capítulo X, como método para el estudio del Control de Calidad.

EL CONTROL DE CALIDAD

El concepto tradicional de *calidad* era que los productos y servicios tienen que cumplir las especificaciones de los fabricantes o que deben satisfacer las necesidades del usuario.

El concepto moderno de calidad se refiere a la variabilidad, es decir, mientras menos variabilidad haya en la fabricación de un producto o en la prestación de un servicio, más calidad habrá.

Diagramas de Pareto

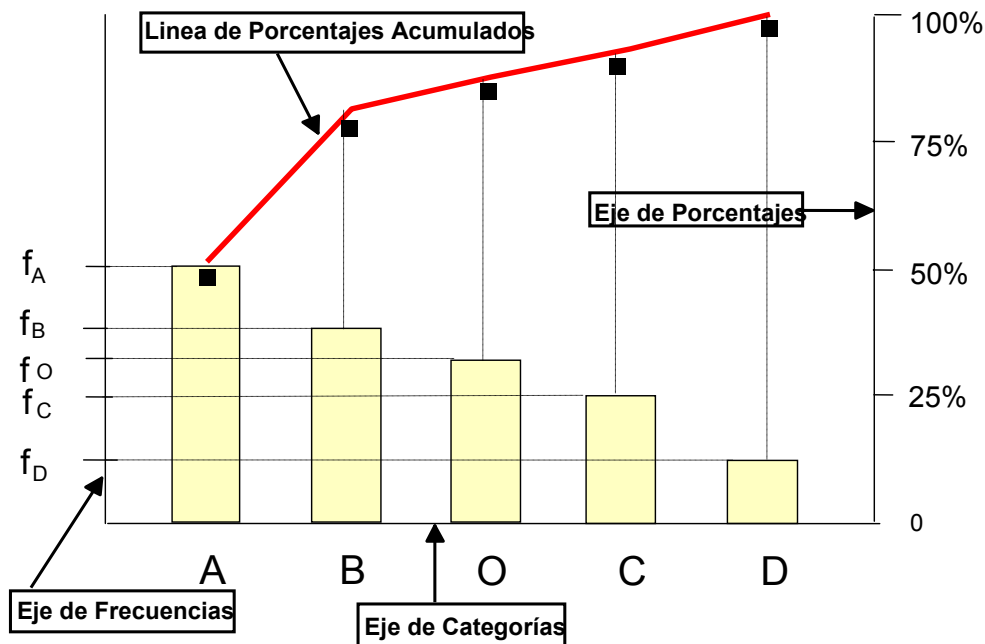
En 1906, el economista italiano Vilfredo Pareto descubrió que el 20% de las personas controlaban el 80% de las riquezas en Italia. Mas adelante, Joseph Juran descubrió que la regla 80/20 se daba en multitud de situaciones y en concreto en los estudios de la calidad.

Un diagrama de Pareto sirve para ver gráficamente cuales son las principales causas por las que un producto o servicio es defectuosos y consiguientemente, poder separar los aspectos fundamentales del problema de los aspectos triviales. Para su construcción se disponen previamente los datos en una tabla donde:

- En la primera columna se colocan las causas o motivos de los defectos o reclamaciones.
- En una segunda columna se colocan las frecuencias absolutas de esas causas, ordenadas de mayor a menor ($f_A \geq f_B \geq \dots \geq f_o$). Si una de las causas es la suma de muchas otras que aparecen en pocas ocasiones, se suele nominar con el epígrafe *otras*, y su frecuencia puede ser mayor que las anteriores.
- La tercera columna es la de las frecuencias absolutas acumuladas.
- La cuarta columna es la de las los porcentajes de cada causa.
- La quinta y última columna es la de los porcentajes acumulados.

CAUSAS	f_i	F_i	p_i	P_i
<i>A</i>	f_A	F_A	p_A	P_A
<i>B</i>	f_B	F_B	p_B	P_B
<i>C</i>	f_C	F_C	p_C	P_C
.....
<i>D</i>	f_D	F_D	p_D	P_D
<i>O (otras)</i>	f_o	F_o	p_o	$P_o = 100%$
	$f_i = N$			

El diagrama consta de un eje horizontal, donde se colocan las causas y dos eje verticales: en el de la izquierda se gradúan las f_i , es decir, el número de casos de cada causa y en el de la derecha los P_i (porcentajes acumulados). A continuación, se levantan barras cuyas alturas son las frecuencias de cada causa y una línea que representa los porcentajes acumulados. En realidad, es una doble representación: la primera consiste en un diagrama de barras con las f_i y la segunda un diagrama lineal con los P_i . Un Diagrama de Pareto tendría una forma:



El Diagrama de Pareto se configura con la regla **80/20**, es decir, el 80% de los defectos en el servicio o de los fallos en la producción son debidos al 20% de las causas, o dicho de otra forma, una cantidad pequeña de categorías producen una cantidad importante de defectos.

Las principales ventajas de utilizar un Diagrama de Pareto son:

- Observar, con una simple inspección, cuál es la causa o cuales son las causas que tienen mayor importancia en el mal funcionamiento de un servicio o en los fallos de un determinado producto.
- Proporcionar una visión simple y rápida de la importancia relativa de los problemas y cuales son las causas claves (también llamadas *focos vitales*) de un determinado problema.
- Comparar las mejoras cuando se han aplicado medidas correctoras para solucionar los fallos o problemas.

Para aclarar todos estos conceptos, vamos a ver con un caso concreto, donde estarían las principales causas de las reclamaciones que se han presentado en el Servicio de Atención al Paciente de un determinado hospital

Ejemplo

Una vez que disponemos de los datos sobre las reclamaciones de todo el año, elaboramos la **Hoja de Control**, que es una tabla simple donde en la primera columna ponemos los motivos de las reclamaciones y en la segunda, el número de estas.

HOJA DE CONTROL	
Tipo de Reclamación	Cantidad
Lista de Espera	37
Retraso en Cita Previa	12
Falta de Asistencia	8
Incumplimiento de Horario	15
Demora en la Asistencia	28
Trato Personal	19
Organización y Normas	79
Transporte	6
Disconformidad con la Asistencia	79
Suspensión de la Asistencia	4
Otros (Extravío Objetos, Insuficiencias del Proceso, Derivación a Otros Centros, así hasta 15 causas)	13
	300

A partir de la Hoja de Control, elaboramos la tabla que vimos anteriormente:

Tipo de Reclamación	f_i	F_i	p_i	P_i
Disconformidad con la Asistencia	79	79	26,3 %	26,3 %
Organización y Normas	79	158	26,3 %	52,7 %
Lista de Espera	37	195	12,3 %	65 %
Demora en la Asistencia	28	223	9,3 %	74,3 %
Trato Personal	19	242	6,3 %	80,7 %
Incumplimiento de Horario	15	257	5 %	85,7 %
Otros	13	270	4,3 %	90 %
Retraso en Cita Previa	12	282	4 %	94 %
Falta de Asistencia	8	290	2,7 %	96,7 %
Transporte	6	296	2 %	98,7 %
Suspensión de la Asistencia	4	300	1,3 %	100 %
<div style="display: flex; justify-content: space-between; width: 100%;"> ■ ■ </div> <p>LOS VALORES DE ESTAS DOS COLUMNAS SE CORRESPONDEN CON LOS DOS EJES DE ORDENADAS DEL DIAGRAMA DE PARETO</p>				

NOTAS

1. La segunda columna, la correspondiente al número de casos de cada causa, se ha ordenado en orden decreciente.
2. Si sumamos los datos de la tercera columna, la de porcentajes, obtenemos $p_i = 99,8 \%$, debido a los redondeos.
3. La quinta columna, la de los porcentajes acumulados, es la que realmente nos aporta las principales causas de las reclamaciones. Aquí, es conveniente observar que sabiendo que el hospital tiene registradas 25 causas de reclamación, el 20% de las causas (las cinco primeras de esta última tabla), generan el 80,7% de las reclamaciones.
4. Antes de dibujar el Diagrama de Pareto, es conveniente dibujar el diagrama de barras de las columnas primera y segunda.

El Diagrama de Pareto construido con SPSS, donde observamos que las cinco primeras causa son las que hay que mejorar y fundamentalmente las dos primeras: Disconformidad con la Asistencia Organización y Normas, que suponen un 52,7% de las reclamaciones. es:

DIAGRAMA DE PARETO DE RECLAMACIONES

