Asistencia para la inteligencia empresarial

Introducción a la estadística
Ejemplos

Servicio Nacional de Aprendizaje SENA

Introducción a la estadística - Ejemplos

A continuación, se presenta una serie de casos de análisis y ejemplos para complementar la información presentada para las técnicas y herramientas de la estadística, para el procesamiento de datos como son: las distribuciones de frecuencia, las medidas de tendencia central y las medidas de dispersión.

Distribuciones de frecuencia para datos simples: se denomina también distribuciones de frecuencia relativa. Como ejemplo, se presenta en la tabla 1 de distribución de frecuencia, para las edades de un grupo de personas que se consideran clientes potenciales para la compra de un nuevo producto. Se realiza un estudio de mercado con 55 clientes, y se obtienen los datos mostrados en esta tabla. Se obtiene, para cada edad, la cantidad de sujetos que la poseen, mostrado en la columna “Frecuencia (f)” y, además, la proporción en % de cada edad, sobre el total de sujetos. Se obtiene, en líneas generales, que la mayor proporción de posibles clientes para el producto, son personas de 47 años con una proporción de 16,36 %. Los clientes menos potenciales, son los de 55 años, con una proporción de 5,45 %.

Tabla 1

Ejemplo de distribución de frecuencia para datos simples

Edad	Frecuencia (f)	Proporción (%)
18	4	7,27
22	8	14,55
46	6	10,91
47	9	16,36
55	3	5,45
63	4	7,27
34	6	10,91
36	3	5,45
28	8	14,55
65	4	7,27
Total	55	100,00

Distribuciones de frecuencia para datos agrupados: para el mismo caso, se presentan ahora los intervalos de clase, conocidos como agrupamientos de datos por rangos. La tabla 2, está conformada por 4 columnas, que se detallan con sus respectivos cálculos.

Tabla 2

Ejemplo de tabla de distribución de frecuencia de datos agrupados

Columna 1	Columna 2	Columna 3	Columna 4
Intervalo	Frecuencia absoluta (fi)	Frecuencia relativa (fr)	%
18 - 24	12	0,22	21,82
25 - 31	8	0,15	14,55
32 - 38	9	0,16	16,36
39 - 45	0	0,00	0,00
46 - 52	15	0,27	27,27
53 - 59	3	0,05	5,45
60 - 66	8	0,15	14,55
Total	55	1,00	100,00

● La primera columna, determina el intervalo en el que se agrupan los datos. El intervalo se calcula a través de diferentes pasos:

Paso 1. *Cálculo del rango (R)*: el rango es igual al valor mayor de los datos, menos el valor menor. Para el caso en estudio, sería 65-18 = 47

Paso 2. *Cálculo del intervalo (I):* es igual a la raíz cuadrada de la cantidad de casos. Para el caso de estudio, es la √55 = 7,42

Paso 3. Cálculo de la amplitud del intervalo: es igual al Rango (R) dividido entre el Intervalo (I). Para el caso de estudio, sería = 47 / 7,42 = 6,33 ≈ 6

Toda vez que se tiene un intervalo, a partir del menor número se empieza a sumar el valor de la amplitud del intervalo. Para el caso en estudio, el primer intervalo estaría entre los 18 y 24 años, ya que 18+6 = 24. Surge así, un total de 7 intervalos.

● La segunda columna posee la Frecuencia absoluta (fi), es decir, la cantidad de casos dentro de cada intervalo.

● La tercera columna muestra la Frecuencia relativa (fr), la cual resulta de la división de la Frecuencia absoluta (fi), entre la sumatoria de las frecuencias absolutas, es decir, el número de casos en estudio.

$$fr = \frac{fi}{\sum fi \; (casos\ de\ estudio)} \qquad (1)$$

Para la primera fila, se tendría una fr = 12/55 = 0,22, como se muestra en la columna 4.
Al finalizar, la fr debe sumar 1. Esta frecuencia relativa, como en el caso de la frecuencia simple, indica el porcentaje total de casos en cada intervalo.

- La cuarta columna muestra la proporción (%), correspondiente al porcentaje calculado al dividir la frecuencia absoluta de cada intervalo (fi) entre el total de casos.

Media: para un conjunto de 20 alumnos de un aula, se requiere saber la media del índice académico en una escala del 1 al 5. Al sumar todos los promedios, se obtiene una sumatoria total de promedios de 83,7 puntos, dividido entre 20 estudiantes, resulta una media de 4,18 puntos, lo cual da indicios de un aula con un buen nivel de desempeño académico general.

Mediana: para el caso de 11 datos ordenados de forma consecutiva:
1, 2, 2, 4, 7, 9, 10, 11, 12, 15, 17
La Mediana es = 9
Cuando se trata de una cantidad par, se suman los dos números centrales y se dividen entre 2.
1, 2, 2, 4, 7, 9, 10, 11, 12, 15, 17, 20
La Mediana es = (9 + 10) /2 = 19/2 = 9

Moda: para el caso de los siguientes datos del ejemplo anterior, la moda es 20.
1, 2, 2, 4, 7, 9, 10, 11, 12, 15, 17, 20, 20, 20

Mediana ponderada: en el caso de la empresa, se pretende obtener un valor promedio de aceptación de un producto en función de la votación que realizaron 27 clientes, respecto a la importancia del atributo de valor más importante. En la tabla 3 se muestran los datos.

Tabla 3

Ejemplo de media ponderada

Columna 1		Columna 2	Columna 3	Columna 4	
Atributo de calidad		Valoración del atributo	Cantidad de clientes (n)	Ponderación	
1	Precio	0,30	6	1,8	
2	Funcionalidad	0,05	4	0,2	
3	Disponibilidad	0,05	2	0,1	
4	Diversidad	0,20	3	0,6	
5	Servicio postventa	0,10	5	0,5	
6	Garantías	0,20	4	0,8	
7	Medios de pago	0,10	3	0,3	
	Total	1	27	4,3	
		Media aritmética	3,86	4,30	Media ponderada

- La Columna 1, contiene el atributo de valor que se está evaluando.
- La Columna 2, contiene la valoración, que sobre la base de 100 % o 1, se le da a cada atributo, por parte de quien realiza el estudio.
- La Columna 3, contiene el conteo de opiniones para cada atributo, con respecto a los 27 clientes.
- La Columna 4, es la ponderación, calculada a partir de la multiplicación de la valoración del atributo (conocido como peso), por el valor de la frecuencia de las opiniones.
- En la última fila, del lado izquierdo, se observa el valor de la media aritmética, calculada con la suma de las opiniones, entre el total de atributos valorados, es decir, 27/7 = 3,86.
- En la última fila, del lado derecho, se observa el valor de la media ponderada, resultante de dividir el total de la ponderación, entre la sumatoria del peso, es decir, 4,3/1 = 4,30.

En este caso, el analista pudiera concluir que la media de valoración de cada atributo por cliente es de 4,30 clientes. También pudiera concluir que las valoraciones por encima de 4,30, tal es el caso del precio y el servicio postventa, son en definitiva, los atributos que más deberían tomarse en cuenta para la introducción del nuevo producto.

Media geométrica: como ejemplo para el caso de las estaturas de 4 bailarinas de ballet clásico, se tiene la recolección de las siguientes: la fórmula se representaría de la siguiente manera, obteniéndose una estatura promedio de 1,23.

MG= 4√(1,20*1,15*1,33*1,25) = 1,23

Media armónica: como ejemplo, para el mismo caso de las edades de 10 alumnos en un aula, se tienen los resultados de media armónica de la Tabla 4.

Tabla 4

Ejemplo de media armónica

	Columna 1	Columna 2
	Edad	**1/n**
1	**18**	**0,06**
2	**17**	**0,06**
3	**16**	**0,06**
4	**17**	**0,06**
5	**18**	**0,06**
6	**16**	**0,06**
7	**15**	**0,07**
8	**17**	**0,06**

9	**16**	**0,06**	
10	**16**	**0,06**	
		0,60	**Total**
		16,15	**Media armónica**

- La Columna 1, representa el sujeto en estudio (alumno)
- la Columna 2, contiene la edad de cada sujeto
- la Columna 3, contiene el inverso (1/n), de la edad de cada sujeto, es decir, 1/edad.
- En la última fila se calcula la media armónica, siendo el resultado de dividir 10/0,60 = 16,55. Es decir, el promedio de edad de los estudiantes de la clase con media armónica es de 16,55 años.

Cuartiles, deciles y percentiles: estos tres elementos son considerados en estadística, medidas de posición. Los cuartiles (Q), son medidas que dividen la serie estadística de datos en cuatro grupos de números iguales, conformando cada uno, sobre la base de un 100 %, el 25 % de la población (25*4 = 100). Existe el cuartil Q4 va de 0 – 25, el cuartil Q3 va de 26 – 50, el cuartil Q2 va de 51 a 75, y el Q1 va de 75 a 100. Esto quiere decir que, en esos cuartiles, se ubica cierta cantidad de datos. Por ejemplo, en el caso de las revistas científicas, las mismas cotizan su impacto científico de calidad, al pertenecer a un determinado cuartil en función del tipo de artículo que se publica, de la trayectoria de la revista, de la calidad de los investigadores, entre otros elementos de medición. En el contexto mundial, los investigadores más cotizados publican sus artículos en revistas Q1 y Q2.

Para ciertos datos, pertenecer a un cuartil, posee mayor importancia, por ejemplo, la valoración de los volúmenes de venta, cuyos datos mensuales se aspira pertenezcan como mínimo, a un cuartil 75, siendo lo mejor el movimiento del dato entre el cuartil 75 a 100. Por el contrario, el Índice de Masa Corporal de un grupo de individuos, se aspira, en contra de la obesidad, que se encuentre por debajo del cuartil 25.

Así como los cuartiles, los deciles (D), dividen a los datos en 10 partes iguales, mientras que los percentiles (P), en cien partes o grupos iguales. Todo dependerá de la magnitud de los datos y de los intereses del análisis.

Desviación media: como ejemplo, para el estudio de los promedios de 20 estudiantes de una clase, se obtuvo una media aritmética de 4,18; con una desviación media de 2,23, resultante de dividir 9,33/4,185. En la Tabla 5 se observan las desviaciones medias parciales, y la desviación media general.

Tabla 5

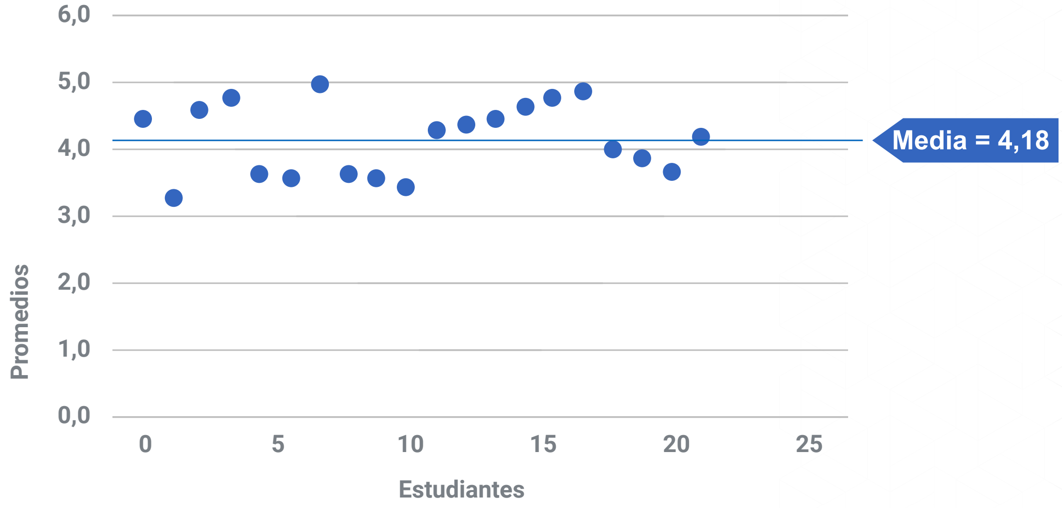
Ejemplo de cálculo de desviación media

Estudiante	Promedio	DM	Valor absoluto	
1.	4,5	0,32	0,32	
2.	3,3	-0,89	0,885	
3.	4,6	0,42	0,42	
4.	4,8	0,62	0,62	
5.	3,7	-0,48	0,485	
6.	3,6	-0,59	0,585	
7.	5,0	0,82	0,82	
8.	3,7	-0,48	0,485	
9.	3,6	-0,59	0,585	
10	3,5	-0,69	0,685	
11.	4,3	0,12	0,12	
12.	4,4	0,22	0,22	
13.	4,5	0,32	0,32	
14.	4,7	0,52	0,52	
15.	4,8	0,62	0,62	
16.	4,9	0,72	0,72	
17.	4,0	-0,19	0,19	
18.	3,9	-0,29	0,29	
19.	3,7	-0,48	0,485	
20.	4,2	0,02	0,02	
Media general	4,185		9,33	
			2,23	**Desviación media del conjunto**

Desviación estándar: como ejemplo, para el mismo conjunto de datos de los 20 estudiantes y su promedio. Con el uso del Excel, a través de la fórmula estadística, se obtiene una SD de 0,53. Se considera una SD que muestra dispersión de los datos, al estar por encima de 0,50. En la Figura 1, se tiene la gráfica de SD para el caso en estudio, en la cual se observa dicha dispersión.

Figura 1

Gráfica de SD para el caso de estudio



Varianza como ejemplo, para el mismo conjunto de datos de los 20 estudiantes y su promedio. Con el uso del Excel, a través de la fórmula estadística, se obtiene una SD2 de 0,27. Se considera una SD2 aunque por debajo de 0,50, con datos dispersos, debiendo tender al máximo a "0", aunque la combinación de la desviación estándar con la varianza, refleja una menor dispersión.