Introducción

En las áreas relacionadas con los negocios y las ciencias sociales es usual trabajar con información organizada y resumida para así tener una mejor apreciación, comprensión e interpretación del fenómeno que deseamos estudiar. Para esto acudimos al auxilio de la estadística descriptiva la cual nos proporciona técnicas para organizar y simplificar la información contenida en un conjunto de datos.

Cuando se lleva a cabo una investigación, la primera tarea consiste en recolectar toda la información relevante. En esta etapa inicial, generalmente observamos que los datos se encuentran de manera desorganizada, lo cual dificulta su apreciación. Para solucionar lo anterior utilizamos métodos tabulares o métodos gráficos, los cuales permiten organizar los datos para hacer su entendimiento más fácil, y resaltar así los rasgos más sobresalientes de la información.

Por ejemplo, cuando un analista de empresas trata de reconocer la configuración que tienen las ventas de una compañía, poco podría averiguar mirando cada una las facturas expedidas de un periodo determinado. En este caso, el analista únicamente tendría una idea muy vaga sobre el nivel de ventas que esta compañía ha alcanzado durante el periodo bajo estudio. Para tener una mejor interpretación, el analista puede acudir al auxilio de las técnicas proporcionadas por la estadística descriptiva, ya sea utilizando métodos tabulares o métodos gráficos, con la finalidad de extraer toda la información contenida en las facturas y darle un sentido lógico a las ventas. De esta manera, el analista puede apreciar con mayor claridad el comportamiento de la demanda por artículos, estimar los costos, las ganancias, así como el número de empleados que serían requeridos en cada periodo del año por esta compañía.

Otro ejemplo que podemos señalar es el relacionado con el flujo migratorio internacional en México. Si deseamos estudiar el fenómeno migratorio, por ejemplo en el Aeropuerto Internacional de la Ciudad de México, muy poca información podríamos obtener observando cada una de las formas migratorias que los extranjeros llenan cuando ingresan a nuestro país. En lugar de esto se pueden utilizar métodos tabulares o métodos gráficos para conocer, de manera resumida, las tendencias del flujo migratorio durante los últimos años. Incluso, podemos observar los meses, los días o las horas en que el flujo migratorio se intensifica en el Aeropuerto Internacional de la Ciudad de México y el propósito que tienen los extranjeros al visitar nuestro país, ya sea que tengan fines turísticos, de negocios o de residencia.

En esta unidad conocerás algunos métodos tabulares y gráficos proporcionados por la estadística descriptiva. En primer lugar se expondrán las tablas de distribución de frecuencias,

las tablas de frecuencias relativas y las tablas de frecuencias relativas acumuladas; posteriormente conocerás algunos métodos gráficos que se utilizan para organizar e interpretar datos como: el diagrama circular, el diagrama de barras, los gráficos de líneas, de dispersión e histograma. Por último, en las secciones 2.3 y 2.4 se estudiarán los polígonos de frecuencia relativa y los polígonos de frecuencia relativa acumulada, los cuales son métodos gráficos especiales que se encuentran íntimamente relacionados con la información proporcionada por las tablas de frecuencia relativa y las tablas de frecuencia relativa acumulada.

El hecho de que estas técnicas para organizar datos sean estudiadas en una unidad especial y al inicio de este curso tiene diversas ventajas. En primer lugar, resulta de gran importancia que el alumno tenga su primer contacto con la estadística utilizando los métodos más elementales y fáciles de comprender. De esta manera, el alumno conocerá la inmediata aplicación y gran utilidad que la estadística descriptiva tiene para la toma de decisiones.

En muchas ocasiones utilizarás los métodos tabulares y los métodos gráficos en tu desempeño profesional, ya sea para realizar un estudio de mercado, para analizar la situación financiera de una empresa o de un sector, para mostrar los niveles de audiencia de un programa de televisión o para exponer con mayor claridad tus ideas en una conferencia o en una reunión de trabajo.

2.1. Métodos tabulares

Existen diversas formas para organizar datos mediante métodos tabulares, destacando aquellos que toman en cuenta la frecuencia o el número de veces en que se presentan los valores en la serie de datos. En esta sección se expondrán los siguientes métodos tabulares:

- Tabla de distribución de frecuencias.
- Tabla de distribución de frecuencias relativas.
- Tabla de distribución de frecuencias acumuladas.

Estos métodos son muy utilizados para analizar el comportamiento de un fenómeno o el desempeño de algún mercado. Su objetivo radica en facilitar la interpretación del conjunto de datos, pues sirven para organizar la información que comúnmente los analistas o en la alta gerencia usan para la toma de decisiones.

2.1.1. Tabla de distribución de frecuencias

También conocida como tabla de frecuencias, es un método tabular que resume la información contenida en un conjunto de datos, organizándolos según su clase y su frecuencia. Es una técnica propia de la estadística descriptiva pues simplifica la información con el propósito de facilitar su interpretación.

Para la construcción de una tabla de distribución de frecuencias generalmente se destinan dos columnas; a la primera columna se le asignan las clases en que se clasifican los datos de la serie; a la segunda columna se le asignan las frecuencias. Una **clase** expresa un valor o un conjunto de valores o categorías en que son agrupados los datos de una serie. La **frecuencia** señala la cantidad de datos que están incluidos dentro de una clase.

El número de renglones de la tabla de distribución de frecuencias depende del número de clases en que se divide la serie de datos. La tabla de distribución de frecuencias puede ser utilizada para organizar y resumir datos cualitativos y datos cuantitativos.

Ejemplo 1

En la siguiente tabla de distribución de frecuencias se exponen los trámites migratorios recibidos en la Delegación Regional del Instituto Nacional de Migración en el Distrito Federal, de enero a junio de 2001, para los extranjeros cuyo propósito fue la residencia temporal o residencia permanente en nuestro país, clasificados según su nacionalidad:

Clase	Frecuencia
Colombia	5 120
Cuba	4 563
Estados Unidos	4 042
Argentina	2 740
España	2 594
Corea del Sur	1 857
Francia	1 697
Perú	1 499
Brasil	1 403
Venezuela	1 381
Alemania	1 316
Chile	1 298
Japón	1 100
China Popular	1 019
Otras nacionalidades	14 339
Total	45 968

Fuente: Programa Bianual de Mejora Regulatoria 2001-2003 de la Secretaría de Gobernación, www.cofemer.gob.mx

Tabla 2.1. Trámites migratorios según la nacionalidad del extranjero, entre enero y junio de 2001.

Este es un ejemplo en que una tabla de distribución de frecuencias organiza *datos cualitativos*, pues la variable que se maneja, nacionalidad del extranjero que solicita residencia en México, expresa un nivel de calidad, no de cantidad. En la primera columna se inscriben las clases, que en este caso son los diferentes países de donde son nacionales los extranjeros, mientras que en la segunda columna se asignan las frecuencias respectivas para cada clase, que en este caso son los trámites solicitados ante el Instituto Nacional de Migración entre los meses de enero a junio del año 2001.

Esta tabla de distribución de frecuencias resume la información contenida en las 45 968 formas migratorias, facilitando su análisis e interpretación. De no haberse utilizado un método tabular, probablemente se hubiera entorpecido el análisis migratorio al observar la nacionalidad de cada una de las solicitudes migratorias {colombiano, español, chino, colombiano, argentino, colombiano, alemán, brasileño, brasileño, estadounidense, chino, estadounidense, etcétera.

Cuando se emplean datos cualitativos organizados a través de tablas de distribución de frecuencias, generalmente las clases se ordenan de mayor a menor frecuencia. En el ejemplo 1, Colombia tiene 5 120 solicitudes de residencia y es la primera clase en aparecer en la tabla, seguida por Cuba que tiene 4 563 y así sucesivamente hasta llegar a China Popular con 1 019 trámites migratorios. Al final se expone una clase especial en la que se establecen las nacionalidades restantes, con una frecuencia de 14 339.

Ejemplo 2

A continuación se muestra el número de accidentes en la avenida Insurgentes que una compañía de seguros contabilizó durante 30 diferentes días. Los datos se presentan en desorden:

3 4 5 0 1 6
2 3 1 2 0 7
5 5 4 3 6 4
4 3 7 0 6 2
2 3 1 1 1 2

Cuando tenemos que manejar demasiada información, la tabla de distribución de frecuencias puede ser modificada agregando una nueva columna, que generalmente se encuentra en la parte central de la tabla. Esta nueva columna lleva el nombre de tabulación, asignándole marcas (|) para identificar cada una de las observaciones de la clase.

X	Tabulación	
0	| | |	3
1	| | | | |	5
2	| | | | |	5
3	| | | | |	5
4	| | | |	4
5	| | |	3
6	| | |	3
7	| |	2
Total		30

Tabla 2.2. Tabla de frecuencias del número de accidentes contabilizados.

En la información recabada originalmente por la compañía de seguros no se podía apreciar el comportamiento de los accidentes registrados en la avenida Insurgentes, por lo que esta información tuvo que ser organizada mediante una tabla de distribución de frecuencias.

En este segundo ejemplo, la tabla de distribución de frecuencias organiza *datos cuantitativos*, pues la variable "número de accidentes ocurridos en un día", es cuantificable. En la primera columna se asigna el número de accidentes ocurridos por día (por ejemplo, en un día pueden ocurrir 0, 1, 2, 3, 4, 5, 6, o 7 accidentes). En la segunda columna se asigna la tabulación (|) para cada una de las observaciones que se señala en cada clase. En la última columna se asignan las frecuencias o número de días en que ocurrieron los accidentes señalados en cada clase (no confundir que la frecuencia es número de días y no número de accidentes). Cabe señalar que la suma de todas las frecuencias es igual al número de datos.

De la tabla 2.2 se puede apreciar que durante tres días no ocurrió ningún accidente, en cinco diferentes días se presentó un accidente diario, en cinco diferentes días se observaron dos accidentes diarios, en otros cinco diferentes días ocurrieron tres accidentes diarios, y así sucesivamente hasta observar que en dos diferentes días llegaron a presentarse siete accidente diarios.

En la primera columna de la tabla de frecuencias 2.2 se ha asignado un sólo valor en cada clase. Sin embargo, cuando existen muchos valores es recomendable efectuar una simplificación adicional mediante el uso de tablas de distribuciones de frecuencias con intervalos que agrupen varios valores en una clase.

Una tabla de distribución de frecuencias con intervalos de clase es un método tabular que organiza y resume información únicamente para datos cuantitativos, agrupando varios valores en cada una de las clases. Los valores de cada clase se expresan a través de intervalos, por ejemplo 100 – 199, lo que quiere decir que en esa clase se establecerán los datos cuyos valores se encuentren entre 100 y 199.

Ejemplo 3

Del XII Censo de Población y Vivienda 2000, se sabe que la población en México se distribuye, según su edad, de la siguiente manera:

Edades	Frecuencia	Hombres	Mujeres
0 – 9 años	21 850 480	11 079 017	10 771 463
10 – 19 años	20 728 628	10 345 385	10 383 243
20 – 29 años	17 228 877	8 165 082	9 063 795
30 – 39 años	13 489 061	6 406 684	7 082 377
40 – 49 años	9 266 924	4 451 948	4 814 976
50 – 59 años	5 917 184	2 858 105	3 059 079
60 – 69 años	3 858 931	1 825 070	2 033 861
70 – 79 años	2 110 944	1 000 303	1 110 641
80 – 89 años	773 927	342 371	431 556
90 – más años	184 598	76 284	108 314
No especificado	2 073 858	1 053 732	1 020 126
Total	97 483 412	47 603 981	49 879 431

Fuente: XII Censo General de Población y Vivienda 2000, www.inegi.gob.mx

Tabla 2.3. Distribución de la población en México de acuerdo con la edad.

En este caso se presenta una tabla de distribución de frecuencias con intervalos de clase, pues en la primera columna se puede observar que las edades se han agrupado mediante intervalos. Por ejemplo, en el año 2000 había 21 850 480 mexicanos que tenían entre 0 y 9 años de edad, de los cuales, 11 079 017 eran hombres, mientras que 10 771 463 eran mujeres. Observa que esta clase y la que señala las personas que no especificaron su edad son las únicas clases en las que existen más hombres que mujeres.

Para este tipo de tablas, el número preciso de clases que ha de adoptarse para una variable determinada depende muchas veces del juicio personal, de la experiencia y de otras consideraciones relativas a la construcción de distribuciones de frecuencia; sin embargo, a continuación se señalan algunos criterios que deben ser tomados en cuenta en el agrupamiento de valores en clases:

1. Para determinar el número de clases no deben utilizarse menos de 5 o más de 20 clases. El número apropiado de clases depende de la cantidad de datos y de valores con los que se trabaje.

2. Las clases deben formarse de tal modo que todos los datos se puedan incluir.

3. Debe asegurarse que cada valor pueda ubicarse en una única clase, estableciendo de manera clara sus límites inferiores y superiores de manera que las clases sean ajenas.

4. Es aconsejable que las clases tomen igual amplitud, donde la amplitud se obtiene dividiendo el número de datos entre el número de clases.

No obstante, existe un método más formal para determinar el número de clases que deben existir para clasificar un conjunto de datos, el cual viene dado por la regla de Sturges. Esta regla es sugerida para aquel lector que desee adentrarse en el tema de la organización de datos, misma que se expresa de la siguiente manera:

Regla de Sturges $C = 3.3 (\log n) + 1$

Donde
C = Número de clases.
n = Tamaño del conjunto de datos que se examinan.

Cabe señalar que esta regla es una aproximación para determinar el número de clases en las que podrían clasificarse los datos de una población o muestra. Las clases establecidas por esta regla en muchas ocasiones deben ser redondeadas de acuerdo con el juicio personal del investigador.

2.1.2. Tabla de distribución de frecuencias relativas

Es un método tabular que resume la información contenida en un conjunto de datos, organizándolos según su clase y su frecuencia relativa. Puede ser utilizada para organizar datos cualitativos o datos cuantitativos. Es muy parecida a la tabla de frecuencia, con la salvedad de que se agrega una columna que señala las frecuencias relativas que corresponden a cada una de las clases. La *frecuencia relativa* de una clase (Fr) se obtiene dividiendo la frecuencia de la clase entre el total de datos.

$$Frecuencia\ relativa\ (Fr) = \frac{Frecuencia}{Total}$$

Ejemplo 4

De la información contenida en la tabla 2.3, se puede encontrar la siguiente tabla de distribución de frecuencias relativas para la población en México, según la edad:

Edades	Frecuencia	Frecuencia relativa	Porcentaje
0 – 9 años	21 850 480	0.2241	22.41%
10 – 19 años	20 728 628	0.2126	21.26%
20 – 29 años	17 228 877	0.1768	17.68%
30 – 39 años	13 489 061	0.1384	13.84%
40 – 49 años	9 266 924	0.0950	9.50%
50 – 59 años	5 917 184	0.0606	6.06%
60 – 69 años	3 858 931	0.0396	3.96%
70 – 79 años	2 110 944	0.0217	2.17%
80 – 89 años	773 927	0.0080	0.80%
90 – más años	184 598	0.0019	0.19%
No especificado	2 073 858	0.0213	2.13%
Total	97 483 412	1.00	100%

Fuente: XII Censo General de Población y Vivienda 2000, www.inegi.gob.mx

Tabla 2.4. Frecuencia relativa de la población en México.

Una tabla de distribución de frecuencias relativas tiene varias ventajas sobre la tabla de distribución de frecuencias; una de ellas es que las frecuencias relativas se pueden convertir fácilmente a porcentajes, multiplicando cada una de ellas por 100.

Por ejemplo, en la tabla 2.4 se puede observar que la frecuencia relativa de los mexicanos cuya edad estaba entre los 0 y 9 años es de 0.2241. Si esta frecuencia relativa es multiplicada por 100, se puede decir que las personas entre 0 y 9 años representaban 22.41% del total de la población en México, mientras que los adolescentes representaron 21.26% de la población.

Ejemplo 5

Durante 40 días una distribuidora automotriz tuvo las siguientes ventas:

51 62 73 83 92 61 58 72
63 64 71 77 69 78 57 65
76 52 54 41 67 85 73 82
78 72 56 68 84 88 53 81
79 100 75 66 55 65 74 48

A partir de los datos anteriores construimos la tabla de distribución de frecuencias relativa.

Clases	Tabulación	Frecuencia (f)	Frecuencia relativa (fr)	Porcentaje
41 – 50	||	2	0.05	5%
51 – 60	||||| |||	8	0.20	20%
61 – 70	||||| |||||	10	0.25	25%
71 – 80	||||| ||||| ||	12	0.30	30%
81 – 90	||||| |	6	0.15	15%
91 – 100	||	2	0.05	5%
Σ		40	1	100%

Tabla 2.5. Distribución de frecuencias de la venta de automóviles

La frecuencia relativa muestra la proporción o porcentaje de la frecuencia de clase con respecto al total de frecuencias. Para la primera clase, la frecuencia relativa se obtiene dividiendo 2/40 = 0.05, indicando que la frecuencia de esa clase es 5% del total de frecuencias, denotando que de los cuarenta días sólo 5% (dos días) registraron ventas que oscilaron entre 41 y 50 (que es la clase). Para la segunda clase la frecuencia relativa es 8/40 = 0.2, por lo que en 20% de los cuarenta días (ocho) se registraron ventas entre 51 y 60, y así sucesivamente.

2.1.3. Tabla de distribución de frecuencias relativas acumuladas

Es un método tabular que resume la información contenida en un conjunto de datos, organizándolos según su clase y su frecuencia relativa acumulada (Fra). Ésta sólo es utilizada para datos cuantitativos. Es muy parecida a la tabla de frecuencia acumulada, con la salvedad de que se le agrega una columna en la que se asignan las frecuencias relativas acumuladas que corresponden a cada uno de las clases.

La **frecuencia relativa acumulada** de una clase (Fra) es la suma de la frecuencia relativa de esa misma clase y la frecuencia relativa acumulada de la clase inmediata anterior de menor valor.

Ejemplo 6

De la información contenida en la tabla 2.3 correspondiente al XII Censo de Población y Vivienda 2000, se obtiene la siguiente tabla de frecuencias relativas acumuladas para la población en México, distribuida según su edad:

Edades	Frecuencia	Frecuencia relativa	Frecuencia relativa acumulada
0 – 9 años	21 850 480	0.2241	0.2241
10 – 19 años	20 728 628	0.2126	0.4367
20 – 29 años	17 228 877	0.1768	0.6135
30 – 39 años	13 489 061	0.1384	0.7519
40 – 49 años	9 266 924	0.0950	0.8469
50 – 59 años	5 917 184	0.0606	0.9075
60 – 69 años	3 858 931	0.0396	0.9471
70 – 79 años	2 110 944	0.0217	0.9688
80 – 89 años	773 927	0.0080	0.9768
90 – más años	184 598	0.0019	0.9787
No especificado	2 073 858	0.0213	1.00
Total	97 483 412	1.00	

Fuente: XII Censo General de Población y Vivienda 2000, www.inegi.gob.mx

Tabla 2.6. Frecuencia relativa acumulada de la población en México.

La frecuencia relativa acumulada para la primera clase se obtiene repitiendo el valor de la frecuencia relativa, caso excepcional; en el caso de la segunda clase se obtiene sumando la primera frecuencia relativa acumulada y la frecuencia relativa de la segunda clase, de lo cual tenemos 0.2241 + 0.2126 = 0.4367; en el caso de la tercera clase se obtiene sumando la segunda frecuencia relativa acumulada con la tercera frecuencia relativa, de lo cual resulta 0.1768 + 0.4367 = 0.6135 y así sucesivamente.

Una tabla de frecuencias relativas acumuladas tiene varias ventajas para el análisis de la información contenida en un conjunto de datos. Una de ellas es que la información que proporciona la frecuencia relativa acumulada puede ser utilizada para mostrar el porcentaje que representan varias clases agrupadas. Por ejemplo, de la tabla 2.6 se puede señalar que 43.67% de la población en México tiene menos de 20 años, 61.35% tiene menos de 30 años o 75.19% tiene menos de 40 años, lo que nos hace señalar que la población en México está integrada en su mayoría por gente joven. De la misma manera, podemos señalar que 23.34% de la población en México se encuentra entre los 30 y 49 años de edad; obteniendo esta cifra de la diferencia entre las frecuencias relativas acumuladas de la población menor a 50 años y la de menor a los 30 años (0.8469 – 0.6135 = 0.2334).

Ejemplo 7

A partir de los datos de la tabla 2.5, sobre las ventas de una distribuidora automotriz obtener la frecuencia relativa acumulada (Fra).

Clases	Tabulación	Frecuencia (f)	Frecuencia relativa (fr)	Frecuencia relativa acumulada
41 – 50	||	2	0.05	0.05
51 – 60	||||| |||	8	0.20	0.25
61 – 70	||||| |||||	10	0.25	0.50
71 – 80	||||| ||||| ||	12	0.30	0.80
81 – 90	||||| |	6	0.15	0.95
91 – 100	||	2	0.05	1
Σ		40	1	

Tabla 2.7. Frecuencias relativas y acumuladas de la venta de automóviles.

Para obtener la frecuencia relativa acumulada (Fra) se debe sumar a la frecuencia relativa de la clase la frecuencia relativa acumulada de la clase anterior, por esto, para la segunda clase la frecuencia relativa acumulada es 0.05+0.20=0.25 debido a que la frecuencia relativa acumulada de la primera clase es 0.05 y la frecuencia relativa de la segunda clase es 0.2.

EJERCICIOS

Ejercicio 1

1. De acuerdo con la información proporcionada por el XII Censo de Población y Vivienda, se sabe que durante el año 2000 había 6 590 348 profesionistas, distribuidos, según su área de estudio, de la siguiente manera: 215 402 en el área agropecuaria; 644 226 en el área de salud; 118 024 en el área de ciencias naturales y exactas; 2 470 708 en el área de ciencias sociales y administrativas, 1 156 098 en educación y humanidades; 1 550 486 en ingeniería y tecnología y, 435 404 que no especificó. Elabora una tabla de distribución de frecuencias. *(Fuente: www.inegi.gob.mx)*

2. A continuación se exponen las calificaciones finales de 42 alumnos inscritos en un curso de Estadística II, impartido en el campus Marina de una conocida universidad en la Ciudad de México:

Las calificaciones finales obtenidas, en el orden de la lista del profesor, fueron: 5, 10, 10, 10, 10, 5, 6, 5, 6, 10, 5, 9, 8, 5, 9, 9, 6, 8, 9, 5, 5, 9, 8, 10, 10, 7, 7, 9, 5, 10, 10, 5, 5, 8, 6, 5, 8, 5, 6, 5, 6, 5.

a) Construye la tabla de distribuciones de frecuencias e indica cuál fue la nota que se obtuvo con mayor frecuencia en este curso.
b) Encuentra la tabla de distribuciones de frecuencias relativas y señala qué porcentaje representó cada una de las notas.
c) Elabora una tabla de distribuciones de frecuencias relativas acumuladas.

3. Con la siguiente información de productos vendidos por hora en una tienda de autoservicio:

48	56	50	36	40
65	74	61	43	60
67	31	51	65	54
66	36	53	67	37
40	41	60	83	46
68	37	58	61	54
53	61	70	77	46
62	38	81	68	37
70	63	90	44	45
36	46	60	55	36

a) Construye la tabla de distribuciones de frecuencias e indica cuál fue la clase o monto de artículos vendidos que más frecuencia tuvo esta tienda de autoservicio, de acuerdo con las siguientes clases:

31 – 40
41 – 50
51 – 60
61 – 70
71 – 80
81 – 90

b) Encuentra la tabla de distribuciones de frecuencias relativas y señala qué porcentaje representó cada una de las notas.

c) Elabora una tabla de distribuciones de frecuencias relativas acumuladas.