**DESCRIPCIÓN DE LOS CONJUNTOS DE DATOS**

1. **INTRODUCCIÓN**

Las investigaciones científicas frecuentemente arrojan una cantidad de datos sueltos que son difíciles de entender para el lector, estos datos sueltos se pueden representar mediante gráficos para su mejor comprensión y comparación, ahí nace la importancia de saber cómo representar gráficamente los conjuntos de datos.

Existen múltiples formas en que se puede organizar y manejar un conjunto grande de datos para proporcionar una interpretación visual rápida de la información que este contiene. Dichas herramientas estadísticas permiten describir un conjunto de datos originales de manera concisa y fácil de leer de acuerdo a la siguiente secuencia:

DESCRIPCIÓN DE LOS CONJUNTOS DE DATOS

DISTRIBUCIÓN DE FRECUENCIAS

GRÁFICOS

DISTRIBUCIÓN DE FRECUENCIAS ABSOLUTAS

HISTOGRMA

DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS

DIAGRAMA DE BARRAS

DISTRIBUCIÓN DE FRECUENCIAS ACUMULADAS

DIAGRAMA CIRCULAR

TABLAS DE CONTINGENCIA

DIAGRAMA DE LINEAS

**2.- DISTRIBUCIÓN DE FRECUENCIAS**

**2.1.- DISTRIBUCIÓN DE FRECUENCIAS ABSOLUTAS**

Para iniciar a tratar la temática es importante basarnos en un ejemplo, por ejemplo el tomar como ejemplo los datos recolectados sobre los pasajeros que viajaron en los últimos 50 días en una aerolínea determinada.

En la siguiente tabla tenemos datos sobre el número de pasajeros que han decidido viajar con la aerolínea P&P. Tales datos correspondientes a los últimos 50 días se muestran en la Tabla N° 1, sin embargo, con estos datos en bruto es improbable que el director pueda obtener información útil y significativa respecto a las operaciones de vuelo. Los datos no están organizados y es difícil llegar a una conclusión significativa simplemente revisando una serie de números anotados en un papel. Es preciso agrupar y presentar los datos de manera concisa y reveladora para facilitar el acceso a la información que contienen. Primero se analizará cómo puede utilizarse una distribución de frecuencia para organizar el conjunto de datos.

**Tabla 1. DATOS DEL NÚMERO DE PASAJEROS DE LA AEROLINEA P&P**

**EN LOS ÚLTIMOS 50 DÍAS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 68 | 71 | 77 | 83 | 79 |
| 72 | 74 | 57 | 67 | 69 |
| 50 | 60 | 70 | 66 | 76 |
| 70 | 84 | 59 | 75 | 94 |
| 65 | 72 | 85 | 79 | 71 |
| 83 | 84 | 74 | 82 | 97 |
| 77 | 73 | 78 | 93 | 95 |
| 78 | 81 | 79 | 90 | 83 |
| 80 | 84 | 91 | 101 | 86 |
| 93 | 92 | 102 | 80 | 69 |

Clase. Es el número de subconjuntos en que se han agrupado los datos. Cada clase se puede denominar mediante una letra, un número o alguna característica del subconjunto. Intervalo de clase. Es un conjunto de elementos que forman a una clase, conteniendo un límite inferior y un límite superior. Tamaño de clase. Es la diferencia entre dos límites inferiores o superiores de clases sucesivas. Frecuencia. Es el número de datos que pertenecen a cada clase.

Una distribución de frecuencias (o tabla de frecuencias) ordenará los datos en clases y se registrará el número de observaciones en cada clase, tal y como se muestra en la tabla N° 2.

**Tabla 2. DISTRIBUCIÓN DE FRECUENCIAS PARA PASAJEROS DE AEROLÍNEAS P&P**

|  |  |  |  |
| --- | --- | --- | --- |
| **Clase (pasajeros)** | **Cuenta** | **Frecuencia (días)** | **Marca de clase****Punto medio** |
| 50 a 59 | III | 3 | 54.5 |
| 60 a 69 | IIIII II | 7 | 64.5 |
| 70 a 79 | IIIII IIIII IIIII III | 18 | 74.5 |
| 80 a 89 | IIIII IIIII II | 12 | 84.5 |
| 90 a 99 | IIIII III | 8 | 94.5 |
| 100 a 109 | II | 2 | 104.5 |
|  |  | 50 |  |

Por ejemplo, ahora se puede ver fácilmente que, que en 18 de los 50 días, entre 70 y 79 pasajeros volaron en P&P. En ningún momento la lista de pasajeros diarios excedió de 109. La aerolínea rara vez transportó menos de 60 pasajeros. El director puede ahora detectar características que no eran evidentes en el análisis de datos en bruto de la tabla N° 1 Estas características son útiles para la toma de decisiones inteligentes y bien informadas respecto a las operaciones de vuelo.

**Calculo de clases**

El número de clases en una tabla de frecuencias es algo arbitrario. En general, la tabla debería tener entre 5 y 20 clases. Muy pocas clases no revelarían ningún detalle sobre los datos y demasiadas clases serían tan confusas como la misma lista de datos originales. Se puede seguir una regla simple para aproximar el número de clases a utilizar, c, es:

$$Número de clases=2^{c}\geq n$$

En donde n es el número de observaciones. El número de clases es la menor potencia a la cual se eleva 2, de manera que el resultado sea igual a o mayor que el número de observaciones. En el ejemplo de P&P, se tiene que n=50 observaciones.

$$Número de clases=2^{c}\geq 50$$

Despejando c, lo cual puede hacerse fácilmente con una calculadora manual, se encuentra que $2^{6}\geq 64$

Esta regla sugiere que debería haber seis clases en la tabla de frecuencias. Por razones de conveniencia, puede utilizarse un número mayor o menor de clases.

El punto medio de la clase, M se calcula como promedio de los limites superior e inferior de dicha clase. El punto medio para la primera clase en la tabla es (50+59) /2=54.5

El intervalo de clase es el rango de valores encontrados dentro de una clase. Se determina restando el límite superior (o inferior) de una clase del límite inferior (o superior) de la clase siguiente. El intervalo para la primera clase en la Tabla 2 es (60-50) =9. Es deseable que todos los intervalos de clase sean de igual tamaño, ya que facilita las interpretaciones estadísticas en usos subsiguientes. Sin embargo, puede ser conveniente utilizar intervalos abiertos que no mencionan un límite inferior para la primera clase o un límite superior para la última clase. La ultima clase puede leerse como “100 o más”

**Cálculo del intervalo de clase**

En la construcción original de una tabla de frecuencias el intervalo de clase puede determinarse como:

Intervalo de clase para una tabla de frecuencia IC= $\frac{valor mas grande-valor más pequeño}{Número deseado de clases}$

Ya que se deciden seis clases para la tabla de frecuencia, el intervalo de clase se convierte en:

IC=$\frac{102-50}{5}=8.7$

Debido a que 8.7 es un número poco práctico, el intervalo puede ajustarse levemente hacia arriba o hacia abajo. Por razones de conveniencia, se seleccionó el intervalo 9 para formar la tabla.

**FRECUENCIA ACUMULADA**

Es la suma (o total acumulado) de todas las frecuencias hasta el punto actual del conjunto de datos.

Una distribución de frecuencia acumulada “menor que” se construye sumando las frecuencias de cada clase, la tabla muestra que en ningún momento volaron menos de 50 pasajeros en P&. La frecuencia acumulada de la primera clase por lo tanto es cero. En tres días, menos de 60 pasajeros abordaron vuelos. La frecuencia acumulada de la segunda clase es tres. Debido a que hubo 10 días en los cuales menos de 70 pasajeros volaron. La tabla muestra que la frecuencia acumulada de la tercera clase es 3+7=10. De nuevo las frecuencias acumuladas de las clases restantes se determinan de manera similar

**Tabla 3. DISTRIBUCIÓN DE FRECUENCIAS ACUMULADAS PARA PASAJEROS DE AEROLÍNEAS P&*P***

|  |  |  |
| --- | --- | --- |
| **Clase (pasajeros)** | **Frecuencia (días) n** | **Frecuencia acumulada (días) N** |
| Menos de 50  | 0 | 0 |
| Menos de 60 | 3 | 3 |
| Menos de 70 | 7 | 10 |
| Menos de 80 | 18 | 28 |
| Menos de 90 | 12 | 40 |
| Menos de 100 | 8 | 48 |
| Menos de 110 | 2 | 50 |

N1=n1 =0

N2= N1+ n2=0+3=3

N3= N2+ n3=3+7=10

**2.2.- FRECUENCIA RELATIVA**

Una distribución de frecuencia relativa expresa la frecuencia dentro de una clase como un personaje del número total de observaciones. En nuestro caso actual, la frecuencia relativa de una clase se determina como la frecuencia de dicha clase dividida por 50. Por ejemplo, la tabla muestra que la frecuencia relativa de la tercera clase es 18/50=36%. Esto permite sacar conclusiones sobre el número de observaciones en una clase respecto a toda la muestra.

**Tabla 4. DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS PARA PASAJEROS DE AEROLÍNEAS P&P**

|  |  |  |
| --- | --- | --- |
| **Clase (pasajeros)** | **Frecuencia (días) n** | **Frecuencia relativa (%) h** |
| 50 a 59 | 3 | 6% |
| 60 a 69 | 7 | 14% |
| 70 a 79 | 18 | 36% |
| 80 a 89 | 12 | 24% |
| 90 a 99  | 8 | 16% |
| 100 a 109 | 2 | 4% |
|   | 50 | 100% |

h1=n1/N=3/50=6%

h2=n2/N=7/50=14%

h3=n3/N=18/50=36%

**3.- TABLAS DE CONTINGENCIA O TABLAS CRUZADAS**

Las tablas de frecuencia pueden organizar datos de solo una variable a la vez. Si se desea examinar o comparar dos variables, una tabla de contingencia resulta de mucha utilidad. Se supone que además de recolectar información sobre el número de pasajeros de P&P, también se obtuvieron datos sobre las edades de los pasajeros y el número de vuelos en los que se registraron cada año. Ambas variables pueden verse en detalle mediante una tabla de contingencia que enumera el número de vuelos en la parte superior y las edades debajo. Tal como se muestra en la tabla. Al dividir la edad en cuatro categorías y los vuelos en tres categorías, se han creado 12 celdas en la tabla. Ahora se puede determinar si estas dos características están relacionadas. Por ejemplo, en la última columna se puede ver que el número más grande de pasajeros, 15 o 30%, está en la categoría de edad comprendida entre los 40 y los 65 años y vuelan más de cinco veces al año. El número más pequeño de pasajeros, tan solo 4 personas son menores de 25 años o tienen 65 años o más. Una mayoría, 28 personas, toma de forma típica más de 5 vuelos cada año.

**Tabla 5. TABLA DE CONTINGENCIA DEL NÚMERO DE VUELOS POR RANGO DE EDAD PARA PASAJEROS DE AEROLÍNEAS P&P**



**4.- GRÁFICOS**

**4.1.- GRÁFICOS DE BARRAS**

Los **diagramas de barras**sirven para comparar dos o más valores, en este se representan variables cualitativas ordinales. En el eje horizontal se representan las diferentes categorí­as y sobre él se levantan unas columnas o barras cuya altura es proporcional a la frecuencia de cada categorí­a, la **frecuencia**es la cantidad de veces que la variable se repite durante un experimento. También podrí­amos utilizar este tipo de gráfico para representar variables cuantitativas discretas, pero lo que no es correcto hacer es usarlo para las variables cualitativas nominales.

Colocando la variable en el eje horizontal y la frecuencia en el eje vertical obtendremos un **diagrama de barras vertical.**

**Figura 1. GRÁFICO DE BARRAS PARA LOS PASAJEROS DE P&P**

Si en cambio lo invertimos, colocando la variable en el eje de ordenadas y la frecuencia en el eje de abscisas, tendremos un **diagrama de barras horizontal.**

**Figura 2. GRÁFICO DE BARRAS PARA LOS PASAJEROS DE P&P**

* Todas las barras, rectángulos o paralelepípedos deben tener el mismo grosor
* Los espacios entre las barras deben ser de la misma magnitud. No debe ser inferior que la mitad de una barra, ni mayor que el ancho de la misma
* La escala de la frecuencia debe empezar de cero
* Deben dibujarse a buen criterio, líneas de fondo en la gráfica ya que ellas facilitan la lectura de los valores
* No se deben recargar las barras tratando de expresar muchos productos en cada una de ellas
* Si el gráfico tiene muchas barras es preferible reemplazarlo por un diagrama lineal

**4.2.- GRÁFICO DE BARRAS DOBLES, MÚLTIPLES**

Es un conjunto de barras simples identificados por distintos colores o sombreado, u**na** gráfica de barras múltiple **es similar a una gráfica de barras, pero representa a dos o tres piezas de información de cada elemento en el eje horizontal**(x)(x)**, en lugar de uno.** Una gráfica de barra doble se puede usar para comparar las respuestas de una encuesta a un grupo contra otro grupo. Una gráfica de triple barra puede ser utilizada para comparar los datos registrados durante un período de tres días o durante varios días.

Se pidió a treinta y cinco niñas y treinta y cinco niños a manifestar su deporte favorito. Los resultados de la encuesta se muestran en la siguiente gráfica.

**Figura 3. GRÁFICO DE BARRAS DOBLES PARA LOS DEPORTES FAVORITOS DE LOS NIÑOS DEL BARRIO DE C&Y**



**4.3.- GRÁFICO DE BARRAS COMPUESTAS O PROPORCIONALES**

Se caracterizan por presentar en una sola figura geométrica, datos cuyo conjunto forman un todo, mostrando la proporción de cada una de las partes con relación al total. Se dibuja el diagrama de barras de una serie y se continua las barras dibujadas con otras de distinto sombreado que representarán la segunda serie estadística. De esta forma, cada barra tendrá una altura, que será la suma de la frecuencia absoluta del suceso que representa la barra en la primera serie mas la frecuencia absoluta del suceso en la segunda serie.

La diferencia con los gráficos de barras simples, dobles o múltiples está en que todos los datos están representados en una sola barra, la cual está dividida en porciones que pueden representar también el peso porcentual de cada una de las partes con relación al 100% que forma la totalidad de ellas.

Es útil cuando se tienen varias variables a representar, con el fin de establecer comparaciones entre ellas (siempre que su unidad de medida sea la misma).

**Figura 4. GRÁFICO DE BARRAS COMPUESTAS DE LAS VENTAS POR MES DE DOS PRODUCTOS DE LA EMPRESA G&J**



**5.- HISTOGRAMA**

Un histograma representa una distribución de frecuencias que se utiliza (o debe) para representar la frecuencia de las variables cuantitativas continuas. Aquí­ no es la altura, sino el área de la barra lo que es proporcional a la frecuencia de ese intervalo, y está en relación con la probabilidad con la que cada intervalo puede presentarse. Como veis en la figura 2, las columnas, a diferencia del diagrama de barras, están todas juntas y el punto medio es el que da el nombre al intervalo. Los intervalos no tienen por qué ser todos iguales (aunque es lo más habitual), pero siempre tendrán un área mayor aquellos intervalos con mayor frecuencia.

**Figura 5. HISTOGRAMA PARA LOS PASAJEROS DE P&P**

En el eje vertical se representan las frecuencias, y en el eje horizontal los valores de las variables, normalmente señalando las marcas de clase, es decir, la mitad del intervalo en el que están agrupados los datos.

**Figura 6. HISTOGRAMA PARA EL PESO DE LOS NIÑOS DEL BARRIO C&Y**





Para realizar el polígono de frecuencias, se deben unir los puntos medios o las marcas de clase de los intervalos. Se emplean los polígonos de frecuencias cuando es necesario **graficar o resaltar distintas distribuciones conjuntas o bien una clasificación cruzada** de una variable cuantitativa continua, junto con otra variable cualitativa o cuantitativa discreta, todo dentro de un mismo gráfico.

Este tipo de grafico es útil porque ponen en relieve y aclaran las tendencias que no se captan fácilmente en las tablas.

**6.- DIAGRAMA CIRCULAR (GRÁFICOS DE SECTORES O DE PASTEL)**

El **diagrama circular** (también llamado **gráfica circular**, **gráfica de pastel** o **diagrama de sectores**) sirve para representar [variables cualitativas](https://www.universoformulas.com/estadistica/descriptiva/variables-estadisticas/) o discretas. Se utiliza para representar la proporción de elementos de cada uno de los valores de la variable.

Consiste en partir el [círculo](https://www.universoformulas.com/matematicas/geometria/circulo/) en porciones proporcionales a la [frecuencia relativa](https://www.universoformulas.com/estadistica/descriptiva/frecuencia-relativa/). Entiéndase como porción la parte del [círculo](https://www.universoformulas.com/matematicas/geometria/circulo/) que representa a cada valor que toma la [variable](https://www.universoformulas.com/estadistica/descriptiva/variables-estadisticas/).

Es importante asegurarse que todos los segmentos sumen el 100%. Utiliza un gráfico circular si quieres **ilustrar el dominio de una categoría sobre las demás, se debe utilizar** menos de 6 sectores. Si los sectores del círculo tienen aproximadamente el mismo valor, considera usar un gráfico de barras o columnas en su lugar. Para construir el gráfico de sector, se utiliza una circunferencia, cuyo circulo se divide en sectores proporcionales a las magnitudes de los valores de la variable que representan. Al total le corresponde el círculo completo, es decir los 360° de la circunferencia y por una regla de tres simple se encontrará el número de grados que le corresponde a cada parte:

Total------------------------360°

Frecuencia--------------------X

**Figura 7. GRÁFICO DE SECTORES PARA LOS PASAJEROS DE P&P**

**6.1 DIAGRAMA DE LINEAS**

Los gráficos de líneas muestran tendencias o cambios a lo largo del tiempo mostrando una serie de puntos de datos conectados por líneas rectas. Puede mostrar un solo campo de datos como varias líneas basadas en diferentes categorías a través del campo.

Se utilizan gráficos de líneas para hacer un seguimiento de los cambios a lo largo de períodos de tiempo breves o extensos y para ayudar en análisis de datos predictivos. Si existen cambios pequeños y frecuentes en la serie, los gráficos de líneas son más eficaces que los gráficos de barras para visualizar el cambio a lo largo del tiempo. Los gráficos de líneas también resultan útiles para comparar los cambios a lo largo del mismo período de tiempo en varios grupos o categorías.

Gráfico de líneas simple. Se utiliza para analizar la tendencia que ha tenido una variable en un determinado periodo de tiempo.

**Figura 8. DIAGRAMA DE LÍNEAS DE LAS HORAS DE ESTUDIO PROMEDIO EN LA CARRERA DE ECONOMÍA EN EL PERIODO 2000-2022**

Gráfico de líneas multiseries curvas. Se utiliza para comparar la tendencia de más de una variable en un determinado periodo de tiempo

**Figura 9. DIAGRAMA DE LÍNEAS DE LAS HORAS DE ESTUDIO PROMEDIO EN LA CARRERA DE CONTADURÍA PÚBLICA, ARQUITECTURA Y MEDICINA EN EL PERIODO 2000-2022**

# **Bibliografía**

*Concepto*. (s.f.). Obtenido de https://concepto.de/poligono-de-frecuencias/

*Excel total*. (s.f.). Obtenido de https://exceltotal.com/partes-de-un-grafico-de-excel/

*Grafico de barras*. (s.f.). Obtenido de https://graficodebarras.com/

*Infogram*. (s.f.). Obtenido de https://infogram.com/es/crear/grafico-circular

*KC-12*. (s.f.). Obtenido de https://flexbooks.ck12.org/cbook/ck-12-conceptos-de-matem%C3%A1ticas-de-la-escuela-secundaria-grado-7-en-espa%C3%B1ol/section/11.6/primary/lesson/gr%C3%A1ficas-de-barras-m%C3%BAltiples/

Molina, M. (29 de octubre de 2019). *Anesteciar*. Obtenido de https://anestesiar.org/2014/como-un-huevo-una-castana-diferencias-entre-un-diagrama-de-barras-y-un-histograma/

Moya Calderon , R. (2015). *Estadística Descrptiva.* Perú: San Marcos E.I.R.L.

Webster, A. (s.f.). *Estadística aplicada a los negocios y la economía.*