

Tipos de datos y su recogida

Se incluyen en este documento algunas ideas sobre los conjuntos de datos que se pueden utilizar en el trabajo final del curso. Se clasifican, en primer lugar, por el tipo de datos que se pueden usar, y después por el tipo de recogida de los mismos.

Tipo de datos

Cualitativos

Son aquellos que están constituidos por palabras o símbolos no numéricos: Películas preferidas, barrio al que perteneces, nombre y apellidos, SI/NO/No sabe, etc., como pueden ser:

- Respuestas a una encuesta: *SI/NO, BIEN/MAL/REGULAR*
- Sexo, Nacionalidad, Comunidad Autónoma, Nombre y apellidos.
- Marcas de motos, títulos de películas o actores preferidos.

A estos datos se les llama también **modalidades o atributos**. Con ellos las únicas operaciones que podemos efectuar son las de contar las veces en las que se presenta cada modalidad, por ejemplo, en una votación contar los votos afirmativos, los negativos y las abstenciones.

El número de veces que aparece cada modalidad recibe el nombre de **Frecuencia** o frecuencia absoluta. Observa esta tabla, sacada de una encuesta sobre el consumo de bebidas:

TABLA DE FRECUENCIAS SOBRE PREFERENCIAS DE BEBIDAS

Tipos de bebidas	Frecuencias
Vino	13
Zumos	8
Coca-cola	10
Tónica	2
Whisky	1
Cerveza	23
Café	12
Otras	11

Es una variable cualitativa, porque no se representa con números.

Lo único que podemos hacer es contar cada tipo de bebida: Hay 23 que prefieren cerveza, 2 que beben tónica, etc. y esos números son las **frecuencias** y se representan con la letra **n**.

Estos datos cualitativos motivan mucho a los alumnos, pues permiten recoger datos que obtienen respuestas de sus compañeros a encuestas, del tipo SI/NO o elección de personajes. Su inconveniente es que se quedan muy cortos en el aprendizaje del Análisis de Datos. La tarea más útil en este tipo de trabajos es la de creación de gráficos (especialmente de barras y de sectores), pero apenas permiten manipular las tablas y enriquecerlas con cálculos. Sólo la *moda* tiene sentido con ellos.

Ejemplos de encuestas de tipo cualitativo:

- Opiniones sobre temas con preguntas del tipo SI/NO
- Sondeos en los que hay que elegir entre varias posibilidades

- Elegir el personaje más popular de la televisión, o actores favoritos
- El alumno o alumna más popular (o entre profesores)
- ¿A quién vas a votar?

- Aportación de ideas, soluciones o comentarios (¿qué hacemos el día del patrón?)

- ¿Cómo mejorarías el tráfico?
- ¿Qué te ha interesado más de este capítulo?
- ¿Qué cualidad aprecias más en tus amigos?

El inconveniente de esta modalidad es que alguien ha de hacer de filtro para unificar respuestas poco claras o ambiguas, así como la creación de apartados de *Otras respuestas*.

- Valoraciones de tipo ordinal sobre varias cuestiones (valora de 0 a 5, elige BIEN, MAL, REGULAR...)

- Estado de la disciplina en el Centro
- Características de tu forma de estudiar
- Dinámica interna del grupo de alumnos
- Personajes favoritos, o marcas de coche, programas de televisión, etc.

Otros estudios

Vocal más frecuente en los distintos idiomas

Marcas de los coches que están aparcados en ciertas calles del barrio a lo largo de varios días

Color, fruta o marca preferidos

Operaciones estadísticas

En los datos cualitativos lo único que podemos estudiar son las frecuencias, los gráficos y el estudio de la moda. No obstante, son suficientes para que los alumnos se interesen, pues su recogida es muy dinámica.

Cuantitativos

Son aquellos datos que se expresan mediante un número. Son los más frecuentes y los que permiten análisis de datos más profundos. Para realizar un trabajo escolar de cierta envergadura son preferibles los datos cuantitativos. Su problema es que requieren precisión en las medidas, salvo que provengan de la operación de contar.

Si entre cada dos datos puede haber una infinidad de ellos, se llaman continuos, y si entre un dato y otro siempre hay un hueco o salto, se llaman **discretos**.

Son datos discretos: Número de hijos, nota de un examen, goles a favor o en contra, etc.

Son continuos: El peso, la estatura, la densidad de un líquido, la fuerza de un muelle, etc.

En muchos estudios los datos aparecen aislados, sin frecuencias. Por ejemplo, las notas de un examen, como se puede ver en la siguiente lista:

Notas de una prueba de Inglés:

0 , 0 , 1 , 2 , 3 , 3 , 4 , 4 , 4 , 5 , 5 , 5 , 6 , 6 , 7 , 8 , 8 , 9 , 4 , 6 , 2 , 2 , 3 , 4

En la Hoja de Cálculo puedes escribir los datos en columna, para que ocupen menos espacio.

Lo normal es que vengan los datos acompañados de frecuencias, en una tabla, con lo que estamos en el terreno natural de la Hoja de Cálculo.

Variable	Frecuencia
4	4
6	8
8	13
10	5
12	4
14	1

Ejemplos de trabajos de tipo cuantitativo:

- Conoce a tus compañeros:
 - Tabla de edades
 - Reparto en el año de las fechas de cumpleaños
 - Horas de estudio diarias, o de televisión, u hora de recogida los viernes y sábados.

- Número de hermanos por familia
 - Frecuencia de latidos del corazón
-
- **Conoce tu Centro:**
 - Dimensiones de cada dependencia
 - Edades de los profesores
 - Estudio de calificaciones numéricas
 - Puntualidad: Número de compañeros que entran desde diez minutos antes a diez minutos después de la hora.
 - Datos de Jefatura de Estudios o Tutorías (no confidenciales)
 - **Conoce tu entorno:**
 - Número de pisos de las casas del barrio
 - Coches que paran en un semáforo según las horas del día
 - Personas que entran en el mercado, o en el Ayuntamiento, etc.

Operaciones estadísticas

Este tipo de datos admite la práctica totalidad de las operaciones estadísticas elementales.:

Tablas de frecuencias: absolutas, relativas, acumuladas, etc.

Gráficos: De barras, de sectores, lineales, etc.

Medidas: Media, mediana, moda, desviación típica, varianza, asimetría y curtosis.

Estadísticos de posición: Cuartiles, percentiles,...

Estadísticos de comparación: Números índices.

Continuos o agrupados

Cuando los datos son muy numerosos, o son continuos (con posibilidad de muchos decimales), se suelen agrupar por intervalos. Son continuos casi todos los datos antropométricos: estatura, peso, glucosa en sangre, colesterol, etc., así como la mayoría de las medidas de tipo físico.

Peso	Frecuencia
30 a 40	4
40 a 50	12
50 a 60	24
60 a 70	18
70 a 80	6
80 a 90	2

Con este tipo de datos podemos emprender muchos experimentos y recogidas de datos distintas:

- Datos antropométricos: estatura, peso, longitudes de brazos, capacidad pulmonar, etc.
- Datos geográficos: Poblaciones, extensiones, longitud de los ríos, etc.
- Procedentes de los medios:
 - Tablas y gráficos económicos
 - Reparto de la riqueza, o de los sueldos
 - Estadísticas sobre problemas actuales: maltrato a la mujer, opiniones políticas, etc.
- Medidas obtenidas en los Laboratorios, en experimentos al aire libre o mediante simulación en ordenador.
- Distribuciones de notas de los alumnos.

Operaciones estadísticas

A las propias de los datos cuantitativos se puede añadir la de ajuste entre distribuciones teóricas y empíricas.

Bidimensionales

En algunos experimentos las medidas que se obtienen son dobles, pertenecientes a dos variables distintas, a las que llamaremos X e Y respectivamente. Este tipo de estudios es muy frecuente. Daremos algunos ejemplos:

- Comparación entre mortalidad y natalidad de distintos países.
- Ídem entre extensión y población.
- Diferencias de renta entre la población en general y los titulados universitarios.
- Pruebas *pretest* y *postest*.
- Influencia de la latitud en la temperatura media.
- Ídem de las horas de estudio en la calificación en una asignatura.

Tipos de variables

Las dos variables que se comparan pueden ser de igual naturaleza, ambas nominales u ordinales o de intervalo, o bien de distinta, lo que da lugar a muchos casos posibles. Nos limitaremos al caso en el que ambas variables son cuantitativas discretas.

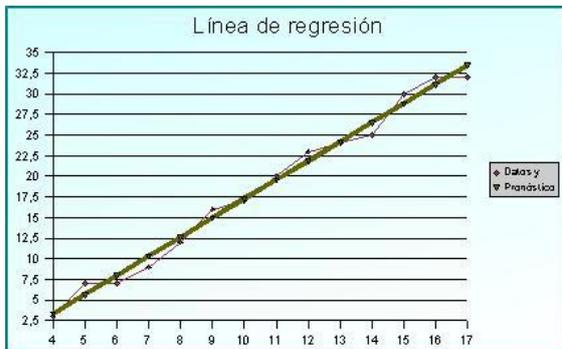
Tablas simples de comparación de dos datos cuantitativos

El caso bidimensional más simple es el formado por datos XY emparejados.

Alumnos	X: Examen de Geografía	Y: Examen de Matemáticas
Julia	4	7
Pedro	6	5
Miguel	5	4

Marta	2	3
.....

En estos casos cada par de valores representa a un sujeto o medición. Se representan mediante gráficos de dispersión XY



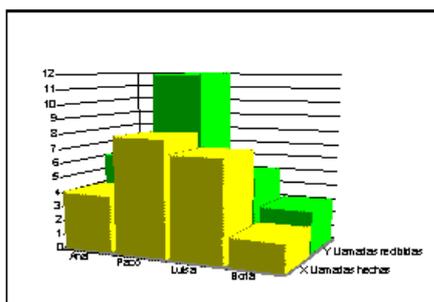
Tablas de doble entrada

En ellas la X y la Y pueden ser de distinta naturaleza, por lo que se disponen en tabla de doble entrada. Cuando existen frecuencias, es el mejor método, pues permite tratar una variable por columnas y otra por filas.

La siguiente tabla muestra la distribución de las llamadas telefónicas con origen o destino en los cuatro hijos de una pareja.

	X Llamadas hechas	Y Llamadas recibidas
Ana	4	6
Paco	8	12
Luisa	7	5
Borja	2	3

Estas tablas de doble entrada con frecuencias admiten una representación gráfica muy intuitiva mediante barras (columnas) ordenadas en varios conjuntos mediante tres ejes.



Operaciones estadísticas

Los datos de tipo bidimensional permiten varias operaciones distintas, aunque las más importantes son la correlación y regresión. Podemos destacar:

- Obtención de distintas frecuencias: marginales, condicionales, acumuladas, etc.
 - Comparación de medias y varianzas.
 - Estudio de la correlación. Coeficiente de Pearson.
 - Estudio de la regresión lineal y no lineal.
 - Pruebas de independencia y ajuste.
-

Recogida de datos

En nuestra época existen muchas fuentes distintas de datos, desde el muestreo de campo, como pueden ser las encuestas y las medidas realizadas por los alumnos hasta las clásicas Enciclopedias y Anuarios y las propias de nuestra época: los Medios de Comunicación e Internet.

Incluimos algunas ideas para la recogida de datos:

Trabajo de campo

Se pueden emprender las medidas y encuestas contenidas en el apartado anterior:

Medidas en:

- Los compañeros
- El Centro
- El Barrio
- Nuestra casa
- Experimentos de Laboratorio

En Enciclopedias, Anuarios y Libros de Texto

Sólo incluimos ejemplos diversos. No se pueden abarcar todos, pero los que figuran pueden ser fuente de inspiración para elegir otros similares:

Tu región

Evolución histórica de las poblaciones, mortalidad, natalidad, etc.
Incidencia de la gripe en los últimos años y su reparto por meses.
Climogramas de los últimos años.
Evolución de la tasa de paro.

España

Estadísticas actuales e históricas sobre las CCAA.
Reparto de los minifundios y los latifundios.

Reparto de las fuentes de energía: Térmica, hidráulica, nuclear, eólica.
Reparto de producción de materias primas por CCAA. Localización de industrias.
Índices varios por CCAA o provincias: habitantes por km², producción agrícola por habitante, etc.
Evolución de la esperanza de vida y distinción por sexos.
Porcentajes de inmigrantes por CCAA o regiones.
Evolución del índice de precios al consumo.
Pirámides de población en las distintas épocas.

Europa y el resto del mundo

Consumo energético por países, CCAA, regiones o continentes.
Evolución histórica del consumo de petróleo.
Estudio de la relación entre las poblaciones de los países y la de sus metrópolis.
Reparto mundial de la riqueza.
Porcentaje de uso de electrodomésticos en las distintas poblaciones: frigoríficos, ordenadores, lavadoras...

Datos procedentes de la Prensa e Internet

En este apartado incluiremos todos los temas de actualidad o de curiosidades que no suelen figurar en los anuarios. Por otra parte Internet permite acceder a los Institutos de Estadística de las distintas CCAA, que son más difíciles de encontrar en los textos.

Intención de voto a lo largo de una campaña electoral.
Estadísticas en problemas que preocupan a la población: maltrato, paro, terrorismo, coste de la vivienda, etc.
Datos de tipo popular: Tablas de clasificación deportivas, datos de cantantes, edades de actores y actrices...
Aficiones: Internet está lleno de páginas sobre aficiones que pueden constituir fuentes de datos.