PARTE I

1. ¿Qué es la Estadística?

La Estadística es una rama de la matemática que se refiere a la recolección, análisis e interpretación de los datos obtenidos en un estudio. Es aplicable a una amplia variedad de disciplinas, desde la física hasta las ciencias sociales, ciencias de la salud como la Psicología y la Medicina, y usada en la toma de decisiones en áreas de negocios e instituciones gubernamentales.

Se puede decir que la Estadística se divide en dos ramas:

La **estadística descriptiva**, que se dedica a los métodos de recolección, descripción, visualización y resumen de originados a partir de los fenómenos en estudio. Los datos pueden ser resumidos numérica o gráficamente. Ejemplos básicos de descriptores numéricos son: la media y la desviación estándar. Resúmenes gráficos incluyen varios tipos de figuras y gráficos.

La **inferencia estadística**, que se dedica a la generación de los modelos, inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta lo aleatorio e incertidumbre en las observaciones. Se usa para modelar patrones en los datos y extraer inferencias acerca de la población de estudio. Estas inferencias pueden tomar la forma de respuestas a preguntas si/no (prueba de hipótesis), estimaciones de características numéricas (estimación), pronósticos de futuras observaciones, descripciones de asociación (correlación) o modelamiento de relaciones entre variables (análisis de regresión). Otras técnicas de modelamiento incluyen ANOVA, series de tiempo y minería de datos.

Ambas ramas (descriptiva e inferencial) comprenden la **estadística aplicada**. Hay también una disciplina llamada **estadística matemática**, la cual se refiere a las bases teóricas de la materia.

La palabra estadísticas también se refiere al resultado de aplicar un algoritmo estadístico a un conjunto de datos, como en estadísticas económicas, estadísticas criminales, etc.

2. Conceptos previos: variables estadísticas.

Llamamos **población** a todo conjunto de objetos, individuos,… que presentan características en común. Ya que en multitud de ocasiones no siempre es posible estudiar a todos los elementos de la población, se emplea una **muestra** que es un subconjunto de la población el cual si se puede estudiar al completo.

Al número de los elementos de la población se le llama tamaño de la población. Análogo para la muestra.

Se denomina **carácter** a una cualidad observable de los elementos de la población o muestra, que puede tomar distintos valores.

> Ejemplo: *Un claro ejemplo, es cuando consideramos un estudio sobre la altura de los eucaliptos en Marín. La población sería todos los eucaliptos de Marín, la muestra serían aquellos que vamos a estudiar (ya que estudiarlos todos es muy laborioso). Claramente el carácter va a ser la altura, pues es la cualidad observable que tienen en común todos los eucaliptos y es objeto de estudio.*

Los caracteres de los elementos de la población se dividen en:

- **Cuantitativos o variables:** son aquellos que son medibles numéricamente (altura, peso, talla, edad).

- **Cualitativos o atributos:** son aquellos que no son medibles (sexo, estado civil, intención de voto).

Distinguimos dos tipos de variables:

- **Variables Discretas:** Una variable estadística es discreta cuando los posibles valores que toma son finitos o numerables (número de hijos, peso, número de coches que fabrica diariamente Opel,…)

- **Variables continuas:** Una variable estadística es continua si sus valores son un número infinito, o bien entre dos valores que pueda tomar la variable existen multitud de otros intermedios (la talla, salario mensual, velocidad viaja un avión …)

Cuando una población es de pequeño tamaño, se puede estudiar el comportamiento de todos y cada uno de sus individuos, recogiendo toda la información en gráficas, tablas y parámetros estadísticos.

A esto se le llama Estadística descriptiva.

3. Tablas estadísticas.

Supongamos que tenemos los datos procedentes de una variable estadística, (por ejemplo, las edades en años de los alumnos de esta clase), veamos de qué forma se puede estructurar dicha información de modo que sea posible manejarla posteriormente.

Frecuencia absoluta: se llama frecuencia absoluta de un valor o modalidad xi de una variable estadística o V.E. al número de individuos de la población que presentan dicho valor. Se representa por fi.

La suma de frecuencias absolutas de todos los valores de una población es igual al tamaño N de la población.

$$\sum f_i = N$$

Frecuencia relativa: se llama frecuencia relativa de un valor xi de una V.E. a la proporción de individuos de la población que presentan dicho valor, es decir, el cociente entre la frecuencia absoluta y el tamaño de la población, representándose por hi.

$$h_i = \frac{f_i}{N}$$

Se cumple que la suma de frecuencias relativas de todos los valores de una población es igual a 1.

$$\sum h_i = 1$$

Frecuencia acumulada absoluta: se llama frecuencia acumulada absoluta de un valor xi de una V.E. a la suma de las frecuencias absolutas de todos los valores anteriores a dicho valor, incluido él. Se representa por Fi

Frecuencia acumulada relativa: se llama frecuencia acumulada relativa de un valor xi de una V.E. a la suma de las frecuencias relativas de todos los valores anteriores a dicho valor, incluido él. Se representa por Hi

Una tabla estadística es un conjunto de columnas en las que se representan los valores de una variable estadística y sus frecuencias. En una primera columna se representan los valores de la variable, y en las sucesivas las frecuencias absolutas, relativas, absolutas acumuladas y relativas acumuladas.

OBSERVACIÓN: En caso de V.E. continuas, en una primera columna se colocan los intervalos de valores donde se engloban estos, a continuación, una columna con los puntos medios de cada intervalo, llamados marcas de clase.

La tabla sería de la forma:

Intervalos $[L_i, L_{i+1}$		***Frecuencia absoluta*** (f_i)	***Frecuencia Absoluta acumulada*** (F_i)	***Frecuencia relativa*** (h_i)	***Frecuencia relativa acumulada*** (H_i)

4. Gráficos estadísticos

El siguiente paso, tras haber tabulado los datos de un estudio estadístico, será proceder a su síntesis mediante unas graficas estadísticas, que recogerán visualmente toda la información posible, y que se basan en la proporcionalidad entre sus áreas y las frecuencias correspondientes. Existen varios tipos de gráficas, dependiendo del carácter que se estudia:

para caracteres cualitativos; diagrama de barras, diagrama de sectores y pictograma.

para variables estadísticas discretas; diagrama de barras, diagrama de sectores, pictograma, diagrama de frecuencias y diagrama de frecuencias acumuladas.

para variables estadísticas continuas; histograma, diagrama de sectores, pictograma, diagrama de frecuencias y diagrama de frecuencias acumuladas.

5. Parámetros estadísticos

Las tablas y gráficos estadísticos nos aportan información cualitativa de la variable estudiada en la población. Con el fin de realizar un estudio cuantitativo, se van a utilizar los parámetros estadísticos. Los parámetros estadísticos se clasifican en 2 grupos:

- Medidas de posición o centralización.
- Medidas de dispersión.

5.1. Medidas de posición o centralización

Las medidas de posición nos indican como se agrupan los datos observados. Dentro de ellas, se dividen en **medidas de posición no central** (cuartiles, deciles, percentiles) y **medidas de posición central** (media, mediana y moda).

MEDIA ARITMÉTICA O MEDIA

Es una medida de posición central y como tal da información de cuál es el valor central de todos los datos recogidos de la muestra o población.

Según la naturaleza de la los datos el cálculo de la media aritmética sigue distintas expresiones:

- Datos simples (todos los valores distintos): Considerando una variable estadística X cuyos valores son $x_1, x_2, \ldots . x_N$ y sea N el número total de datos (tamaño de la muestra o población), se denota la **media aritmética para datos simples** como:

$$\bar{x} = \frac{x_1 + x_2, \ldots . + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Ejemplo:

Si queremos saber cuánto viven de media en cautividad las cebras en el zoológico de Vigo y disponemos de los datos de todas las cebras que hubo.

Se tiene que en total el zoológico de Vigo albergo a 8 cebras (tamaño de la muestra es N=8) y los registros de cada una de ellas son: 8.7, 4.2, 3, 6.5, 7.3, 9.4 ,6 y 7.7 años.

De esta forma se tiene que una idea de cuánto vivieron las cebras sería calcular la media, que nos proporcionaría el valor medio de esperanza de vida de todas ellas.

$$\bar{x} = \frac{8.7 + 4.2 + 3 + 6.5 + 7.3 + 9.4 + 6 + 7.7}{8} = \frac{1}{8}\sum_{i=1}^{8} x_i = \ 6.6\ a\tilde{n}os$$

- Datos agrupados (algunos valores se repiten): Considerando una variable X y un tamaño de la muestra N. Si tenemos un total de $x_1, x_2, \ldots . x_N$ valores y sabiendo que muchos de ellos se repiten lo cual nos permite agruparlos y se tienen ahora $x_1, x_2, \ldots . x_k$ valores distintos con K<N. Siendo $h_1, h_2, \ldots ., h_k$ las frecuencias relativas de los valores distintos y $n_1, n_2, \ldots, n_k$ las frecuencias absolutas de los valores distintos, se define la **media aritmética para valores agrupados** como:

$$\bar{x} = x_1.h_1 + \ x_2.h_2 + .. + \ x_k.h_k = \sum_{i=1}^{k} x_i h_i = \frac{1}{N}\sum_{i=1}^{k} x_i f_i$$

Ejemplo:

Si queremos saber la edad media de los alumnos de 2º Bach de un instituto. Tenemos que hay 28 alumnos de los cuales 11 tienen 16 años, 13 tienen 17 años y 4 tienen 18 años.

Como podemos observar el tamaño de la muestra es N=28, además hay datos que se repiten (hay muchos alumnos que tienen la misma edad) se trata entonces de datos agrupados. Luego solo consideramos 3 datos distintos:

$x_1 = 16\ a\tilde{n}os,\ x_2 = 17\ a\tilde{n}os \quad y\ x_3 = 18\ a\tilde{n}os$ *(K=3)*

Veamos los valores de la frecuencia absoluta y relativa para cada uno de estos datos mediante la siguiente tabla:

x_1	*17*
x_2	*16*
.....	
x_{28}	*17*

Realizar la tabla con 28 valores es muy laborioso, por ello vamos solo a emplear los que sean distintos y utlizamos las frecuencias

x_i	f_i	h_i
16	*11*	*0.39*
17	*13*	*0.47*
18	*4*	*0.14*
total	*28*	*1*

De esta forma la media aritmética va a ser:

$$\bar{x} = \frac{16.11 + 17.13 + 18.4}{28} = \frac{1}{28}\sum_{i=1}^{3} x_i = 16.0.39 + 17.0.47 + 18.0.14$$
$$= 16.65\ a\tilde{n}os$$

- Datos continuos (no son finitos): Considerando como N el tamaño total de la muestra y X la variable continua se actúa de forma similar a los datos agrupados salvo que ahora los valores $x_1, x_2, \ldots . x_k$ con K<N tiene que ser calculados y se llaman **marcas de clase.**

Primeramente si tenemos una variable con datos continuos, se acota por intervalos de forma que se tienen intervalos del tipo:

$$[L_1, L_2], [L_2, L_3], \ldots . , [L_{n-1}, L_n]$$

Ya que no conocemos el valor de cada x_i pero si el intervalo donde está contenido, calculamos las **marcas de clase** de la siguiente forma. Sea **a** la **amplitud o longitud del intervalo**, $a_i = [L_i - L_{i-1}]$. De tener la misma amplitud todos los intervalos ($a_1 = a_2 = .. = a_n$) se tiene que cada marca de clase es:

$$x_i = \frac{L_i + L_{i-1}}{2}$$

De esta forma ya podemos operar como si fuese una variable simple o agrupada y calcular cualquier parámetro de forma normal (media, mediana, moda,…).

Así pues, si los datos son continuos, calculando las marcas de clase $x_1, x_2, \ldots . x_k$ (las denotamos igual que los valores para datos simples o agrupados) y sabiendo las frecuencias absolutas $n_1, n_2, \ldots, n_k$ y las frecuencias relativas $f_1, f_2, \ldots . , f_k$ se tiene que la **media aritmética para datos continuos** toma la forma:

$$\bar{x} = x_1.h_1 + x_2.h_2 + .. + x_k.h_k = \sum_{i=1}^{k} x_i h_i = \frac{1}{N}\sum_{i=1}^{k} x_i f_i$$

Ejemplo:

Si queremos saber la altura media de los alumnos de 2º Bach del instituto. Sabemos que nunca dos alumnos medirán lo mismo pues según la precisión de la medición siempre habrá variaciones (diferencia de centímetros, milímetros, …), luego se tata de una variable con datos continuos, es decir es necesario acotar en intervalos donde se contenga a la altura del alumno.

Tenemos que hay 28 alumnos de los cuales 3 miden entre 1,50 y 1,60, 13 miden entre 1,60 y 1,70, 10 miden entre 1,70 y 1,80 y 2 mide entre 1,80 y 1,90.

Como podemos observar el tamaño de la muestra es N=28, además hay datos que se repiten (hay muchos alumnos que están en el mismo intervalo de altura) se trata entonces de datos continuos agrupados. Luego solo consideramos 4 intervalos distintos:

$L_1 = [1.50, 1.60],\ L_2 = [1.60, 1.70],\ L_3 = [1.70, 1.80]\ \ y\ \ L_4 = [1.80, 1.90]$

Observamos que todos los intervalos tienen la misma amplitud. Calculamos ahora las marcas de clase:

$$x_1 = \frac{1.60 + 1.50}{2} = 1.55, \qquad x_2 = \frac{1.70 + 1.60}{2} = 1.65,$$
$$x_3 = \frac{180 + 1.70}{2} = 1.75 \qquad x_4 = \frac{1.90 + 1.80}{2} = 1.85$$

Ahora es momento de realizar la tabla indicando las frecuencias absolutas y relativas de cada marca de clase:

x_i	f_i	h_i
1.55	*3*	*0.10*
1.65	*13*	*0.47*
1.75	*10*	*0.36*
1.85	*2*	*0.07*
total	*28*	*1*

De esta forma la media aritmética va a ser:

$$\bar{x} = \frac{1{,}55.3 + 1{,}65.13 + 1.75.10 + 1{,}85.2}{28} = \frac{1}{28}\sum_{i=1}^{4} x_i = 1.69\,m$$

Las ventajas que presenta la media aritmética son las siguientes:

- Está perfecta determinada y es única.
- Tiene un significado interpretativo muy claro.
- Es muy sencilla de calcular.
- Para su cálculo se emplean todos los valores de la muestra.

Por el contrario, la media aritmética presenta el siguiente inconveniente:

- Si los valores de los extremos son muy dispares, estos influyen demasiado en el cálculo de la media haciendo que pierda valor significativo. Un claro ejemplo es si queremos saber la nota media de 5 alumnos y sabemos que fueron: un 5.12, 5.1, 5.23, 5.19 y 9. Se tiene que al calcular la media el valor es $\bar{x} = 5{,}94$ que no parece muy válido sabiendo que 4 de los 5 alumnos tienen una nota próxima a 5.1

MEDIANA

Se define **la mediana (M)** como el valor de la variable que ocupa el lugar que divide a los datos en dos partes iguales, habiendo tantos valores por encima como por debajo.

Al igual que sucede con la media, en función de la naturaleza de los datos se tienen las siguientes fórmulas para calcular la mediana:

- Datos simples (todos los valores son distintos):Sea X la variable estadística con valores son $x_1, x_2, \ldots . x_N$ y sea N el número total de datos (tamaño de la muestra o población), se tiene que la mediana vale:
  a) Si **N es impar: la mediana** es el valor que ocupa la posición (N+1)/2, es decir, el valor central

Ejemplo:

Las calificaciones en la asignatura de Matemáticas de 39 alumnos de una clase viene dada por la siguiente tabla:

Calificaciones	1	2	3	4	5	6	7	8	9
Número de alumnos	2	2	4	5	8	9	3	4	2

Primero se hallan las frecuencias absolutas acumuladas N_i. Así, aplicando la fórmula asociada a la mediana para n impar, se obtiene $X(39+1)/2 = X20$.

- Ni-1< n/2 < Ni = N19 < 19,5 < N20

Por tanto la mediana será el valor de la variable que ocupe el vigésimo lugar.En este ejemplo, 21 (frecuencia absoluta acumulada para Xi = 5) > 19,5 con lo que Me = 5 puntos, la mitad de la clase ha obtenido un 5 o menos, y la otra mitad un 5 o más.

xi	fi	Ni
1	2	2
2	2	4
3	4	8
4	5	13
5	8	21 > 19,5
6	9	30
7	3	33
8	4	37
9	2	39

  b) Si n es par, la mediana es la media aritmética de los dos valores centrales. Es decir la media de los valores que ocupan las posiciones N/2 y (N+1)/2.

Ejemplo:

Las calificaciones en la asignatura de Matemáticas de 38 alumnos de una clase viene dada por la siguiente tabla (debajo):

Calificaciones	1	2	3	4	5	6	7	8	9
Número de alumnos	2	2	4	5	6	9	4	4	2

Primero se hallan las frecuencias absolutas acumuladas N_i. Así, aplicando la fórmula asociada a la mediana para n par, se obtiene la siguiente fórmula: $X = n/2 ==> X = (38/2) => X = 19$ (Donde n= 38 alumnos divididos entre dos).

- Ni-1< n/2 < Ni = N18 < 19 < N19

Con lo cual la mediana será la media aritmética de los valores de la variable que ocupen el decimonoveno y el vigésimo lugar. En el ejemplo el lugar decimonoveno lo ocupa el 5 y el vigésimo el 6 con lo que Me = (5+6)/2 = 5,5 puntos, la mitad de la clase ha obtenido un 5,5 o menos y la otra mitad un 5,5 o más.

xi	fi	Ni+w
1	2	2
2	2	4
3	4	8
4	5	13
5	6	19 = 19
6	9	28
7	4	32
8	4	36
9	2	38

La mediana, como medida de posición central, presenta las siguientes ventajas:

- Interpretación muy sencilla.
- Solo influyen los datos centrales (aunque los valores extremos estén muy separados, no afectan al cálculo de la mediana).
- Su cálculo es sencillo.

Por el contrario, los principales inconvenientes son:

- No tiene una fórmula matemática única, ya que depende de si los datos son pares o impares.

MODA

Se llama **moda (Mo)** de un estudia estadístico descriptivo al valor (o valores) que se corresponden con el de mayor frecuencia. Cabe destacar que un estudio puede tener una o varias modas, en caso de tener una se llama **unimodal**, si tiene dos es **bimodal**.

PERCENTIL

Se define como **percentil de orden r** y se denota como $\boldsymbol{P_r}$ al valor de la variable que representa el r% del total de la frecuencia absoluta acumulada.

Los percentiles son medidas de posición no central y los más empleados son $\boldsymbol{P_{25}, P_{50}, P_{75}}$ se denominan **cuartiles**. Asimismo, se tiene que $\boldsymbol{P_{50}}$ es el percentil que ocupa la posición central, es decir, coincide con la mediana.

Otros percentiles muy empleados son los **deciles** y son $\boldsymbol{P_{10}, P_{20}, \dots, P_{90}}$

5.2 Medidas de dispersión.

Las medidas de dispersión, muestran la variabilidad de una distribución, indicándonos por medio de un número si los diferentes valores que toma una variable en un estudio estadístico están muy alejados de la media de los valores. Cuanto mayor sea este número mayor dispersión hay respecto de la media en los valores

Las medidas de dispersión más empleadas son la varianza y la desviación típica. A continuación, se muestra la definición de cada una y su expresión.

VARIANZA

Se define **la varianza** de una distribución de frecuencias como la media aritmética de los cuadrados de las desviaciones respecto de la media. Es el índice de dispersión más empleado y se denota por $\boldsymbol{\sigma^2}$ **o V(x).**

En función de los datos, se tienen las siguientes expresiones para calcular la varianza:

- Datos simples (todos los valores son distintos):Sea X la variable estadística con valores son $x_1, x_2, \ldots . x_N$ y sea N el número total de datos (tamaño de la muestra o población), se tiene que la **varianza para datos simples** toma la forma:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots . + (x_N - \bar{x})^2}{N} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

- Datos agrupados (algunos valores se repiten): Considerando una variable X y un tamaño de la muestra N. Si tenemos un total de $x_1, x_2, \ldots . x_N$ valores y sabiendo que muchos de ellos se repiten lo cual nos permite agruparlos y se tienen ahora $x_1, x_2, \ldots . x_k$ valores distintos con K<N. Siendo $f_1, f_2, \ldots , f_k$ las frecuencias relativas de los valores distintos y $n_1, n_2, \ldots , n_k$ las frecuencias absolutas de los valores distintos, se define la **varianza para datos agrupados** como:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots . + (x_N - \bar{x})^2 f_N}{N} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 f_i$$

O también para el cálculo de la varianza con datos agrupados se emplea:

$$\sigma^2 = \frac{\sum x_i^2 \cdot f_i}{N} - \bar{x}^2$$

DESVIACIÓN TÍPICA.

Se define la desviación típica como la raíz cuadrada positiva de la varianza y se denota por **σ,** sirviendo para medir el grado de separación de los datos; a mayor σ, mayor dispersión de los datos respecto a la media.

Realmente, no hay mucha diferencia en la información que da la varianza o la desviación típica, salvo que la varianza, al ser una media de diferencias al cuadrado6 respecto los datos iniciales no está en el mismo rango que los datos iniciales, está al cuadrado. Pero esa diferencia de escala se corrige al tomar la raíz cuadrada, ya que la desviación típica vuelve a estar a la misma escala que los datos iniciales.

Cuanto más cercano a cero sea $\boldsymbol{\sigma^2}$ (o σ), la media será más representativa, al estar próximos los datos. Cuanto m´as grande sea $\boldsymbol{\sigma^2}$ (o σ), los datos serán más dispersos.

PARTE II

6. Distribuciones de probabilidad y variables aleatorias

Las distribuciones de probabilidad son idealizaciones de las distribuciones de frecuencias relativas de un fenómeno estadístico y se usan para realizar estudios sobre experimentos ya que en la práctica no pueden repetirse indefinidamente.

Si lanzamos una moneda 1000 veces, la distribución de probabilidad (o la situación ideal) es que la cara salga un 50 % de las veces, 500, con probabilidad 0’5, al igual que la cruz. La distribución de frecuencias será algo muy parecido, 51 % a 49 %, 52 % a 48 %... que por la Ley de los Grandes números deberá converger a la distribución de frecuencias a medida que aumenten los lanzamientos.

La distribución de probabilidad anterior es la teoría, lo ideal, el modelo. En la realidad lo que saldrá es algo parecido, una distribución de frecuencias que a medida que aumenten los lanzamientos se parecerá más a la distribución de probabilidad.

Variable aleatoria

Definición: Una función que transforma cada resultado posible de un experimento aleatorio (el espacio muestral) en un número real. Su propósito es cuantificar los resultados de un evento incierto para poder tratarlos matemáticamente y predecir su comportamiento probabilístico.

Una **variable estadística** es una característica de una población que se puede medir o contar, mientras que una **variable aleatoria** es una función que asocia un valor numérico a cada resultado de un experimento aleatorio. La principal diferencia es que la variable aleatoria se usa para modelar fenómenos de azar y obtener distribuciones de probabilidad, mientras que la variable estadística es la característica observable en sí misma, como la altura de los estudiantes en una clase.

Ejemplos de variables aleatorias.:

- Al lanzar dos monedas, la variable aleatoria puede ser el número de caras (puede tomar los valores 0, 1, o 2).
- La altura de una persona seleccionada al azar de la población.
- El número de llamadas telefónicas recibidas en una central en un minuto.

- **Ejemplos de variables estadísticas:**
- La altura de un grupo de personas, tomada directamente de la población.
- El peso de los objetos de una fábrica.
- El color de ojos de un grupo de estudiantes.

Relación entre ambas

- Una variable estadística puede ser representada matemáticamente por una variable aleatoria. Por ejemplo, si medimos la altura de estudiantes universitarios, la altura es la variable estadística y la función que asigna a cada estudiante su altura es una variable aleatoria.
- La variable aleatoria es una herramienta teórica que se utiliza para analizar las variables estadísticas en contextos de incertidumbre, permitiendo calcular probabilidades y hacer inferencias sobre la población a partir de una muestra.

7. Distribución de probabilidad para una variable discreta

Recordar que una **variable aleatoria** X asocia a cada suceso del espacio muestral E un número real. Por ejemplo, tomando el experimento de contar el número de caras al lanzar 2 monedas, la variable X "contar el número de caras" puede tomar los valores: 0,1 y 2.

$$E = \{CC, C+, +C, ++\}$$
$$Variable\ aleatoria\ X = contar\ el\ número\ de\ caras$$
$$X(CC) = 2$$
$$X(C+) = 1$$
$$X(+C) = 1$$
$$X(++) = 0$$

Asimismo, una variable aleatoria X se dice **discreta** cuando toma valores aislados, como en este ejemplo.

Función de probabilidad (de densidad o de masa de probabilidad)

La **función de probabilidad** asigna a cada valor $x_1, x_2, \dots, x_n$ de la variable aleatoria discreta X una probabilidad. Es decir, toma la forma:

$$f(x_i) = P(X = x_i) = P_i$$

Por tratarse de una probabilidad se va a cumplir que:

- $0 \leq P(X = x_i) \leq 1$
- $\sum_{i=1}^{n} P(X = x_i) = 1$

Función de distribución

La **función de distribución** se construye a partir de la distribución de la variable X y se define como la siguiente probabilidad:

$$F(x_i) = P(X \leq x_i)$$

Es decir, es la probabilidad de que ocurra ese suceso y todos los anteriores.

Para entender esto veamos el ejemplo del número de caras:

$$E = \{CC, C+, +C, ++\}$$

$$Variable\ aleatoria\ X = contar\ el\ número\ de\ caras$$

$$X = 0 \Rightarrow f(0) = P(CC) = \frac{1}{4}$$

$$X = 1 \Rightarrow f(1) = P(C+, +C) = \frac{2}{4}$$

$$X = 2 \Rightarrow f(2) = P(++) = \frac{1}{4}$$

Ejemplo:

Tomemos el experimento de lanzar dos dados y hallar la suma de los resultados en cada lanzamiento. Se tiene que la variable aleatoria discreta X va ser "la suma de los dos dados" y los valores que puede tomar están entre 2 y 16. Por tanto de enunciado se deduce:

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X (suma)	2	3	4	5	6	7	8	9	10	11	12
$P(X=i)$ $2 \leq i \leq 12$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Con este ejemplo si nos piden la probabilidad de obtener un 2 o bien un 7 se pide calcular la función de probabilidad:

$$f(2) = P(X = 2) = \frac{1}{36}$$

$$f(7) = P(X = 7) = \frac{6}{36}$$

Siguiendo con el mismo ejemplo, si se pide calcular la probabilidad de que al lanzar dos dados la suma sea igual o inferior a 3 nos están pidiendo calcular una función de distribución:

$$F(3) = P(X \leq 3) = P(X = 2) + P(X = 3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}$$

Media y Varianza de una distribución de probabilidad discreta

En las distribuciones de probabilidad podemos calcular la media y la varianza de manera análoga a lo que se hace con las distribuciones estadísticas.

Si tenemos una distribución de probabilidad discreta siendo la variable X, si toma los valores $x_1, x_2, \dots, x_n$ con probabilidades $p_1, p_2, \dots, P_n$ se define **la media o esperanza matemática** como:

$$E(X) = \mu = x_1p_1 + x_2p_2 + \cdots . + x_np_n = \sum_{i=1}^{n} x_ip_i = \sum_{i=1}^{n} x_i \cdot f(x_i)$$

Por otro lado, se define **la varianza** y **la desviación típica** de una distribución de probabilidad discreta como:

$$\text{Varianza -> } \sigma^2 = \sum_{i=1}^{n} {x_i}^2 p_i - \mu^2$$

$$\text{Desviación típica -> } \sigma = \sqrt{\sum_{i=1}^{n} {x_i}^2 p_i - \mu^2}$$

OBSERVACIÓN: en las distribuciones de probabilidad para designar a la media se utiliza la letra μ en vez de $\bar{x}$.

Media, varianza y desviación típica de la variable X número de caras:

$$\mu = E(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = 1$$

$$\sigma^2 = Var(X) = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{2}{4} + 2^2 \cdot \frac{1}{4} - 1^2 = \frac{1}{2}$$

$$\sigma = \sqrt{1/2}$$

8. Distribución binomial

Es la distribución de probabilidad discreta más empleada en la práctica. La **distribución binomial** se emplea cuando el fenómeno a estudiar queda determinado por dos situaciones: éxito o fracaso, si o no, hombre o mujer, es decir, cuando en lo que estudiamos solo nos interesan dos opciones a favor o en contra.

Las características fundamentales de una distribución binomial son:

- Cada prueba del experimento aleatorio presenta dos únicas opciones: éxito (E) o fracaso (F).

- Se realizan n pruebas o experimentos aleatorios e independientes unos de otros.
- La probabilidad de éxito es constante en las n pruebas: P(E)=p
- La probabilidad de fracaso es constante en las n pruebas: P(F)=1-p=q

En una distribución binomial, si consideramos n como el número de experimentos o pruebas y p como la probabilidad de éxito se tiene que la distribución binomial queda completamente definida por estos dos valores y se representa como: B(n,p).

La **función de probabilidad** de una distribución binomial B(n,p) donde se pide calcular la probabilidad de r éxitos toma la forman:

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

Media, varianza y desviación típica de una binomial B(n,p) como:

$$\mu = np.$$
$$\sigma^2 = np(1-p).$$
$$\sigma = \sqrt{np(1-p)}$$

Ejemplo:

Si en una determinada región, la tasa de paro entre su población activa es del 12%, si se pregunta a 10 personas de esa población, elegidos al azar, por su situación laboral. ¿Cuál es la probabilidad de que haya dos parados dentro de los encuestados? ¿Y menos de tres parados?

Primeramente tras leer el problema vemos que estudiamos la tasa de paro, es decir, estar en paro es éxito y no estar en paro es fracaso, luego se puede asociar el problema a una binomial. Ya que se encuestan a 10 personas, se tiene que n=10 y la probabilidad de éxito (de estar en paro) es de 12%, de otra forma, p=0.12

Por tanto se tiene B(10, 0.12), sabiendo esto la probabilidad de que haya dos parados entre los encuestados es:

$$P(X=2) = \binom{10}{2} 0{,}12^2 \cdot 0{,}88^8 = \frac{10 \cdot 9}{2} 0{,}12^2 \cdot 0{,}88^8 = 45 \cdot 0{,}0051787... = 0{,}233$$

La probabilidad de menos de tres parados, toma la forma:

$$P(X<3) = P(X=0) + P(X=1) + P(X=2) = \binom{10}{0} 0{,}12^0 \cdot 0{,}88^1 + \binom{10}{1} 0{,}12^1 \cdot 0{,}88^9 + \binom{10}{2} 0{,}12^2 \cdot 0{,}88^8 = 0{,}89$$

Asimismo si queremos saber ahora cual es la probabilidad de al menos tres parados, se tiene:

$$P(X \geq 3) = 1 - P(X < 3) = 1 - 0.89 = 0.11$$

9. Distribución de probabilidad para una variable continua

Una variable estadística X es continua cuando puede tomar todos los valores de un intervalo. Ejemplos de variables continuas son las estaturas, pesos, tiempos de espera de un autobús, tamaño de manzanas,…En las distribuciones continuas la probabilidad de un valor concreto es 0, pues cualquier medida nunca es exacta 1,722347262….m. De esta forma:

P(X=1,722347262..)=0

En lugar de calcular probabilidades en puntos concretos se van a hacer por intervalos: P(1,72< X<1,73).

Función de probabilidad o de densidad

Para variables continuas, X, la función de probabilidad también se llama función de densidad, f(x), permite calcular la probabilidad para distribuciones continuas. Ya que no hay probabilidad de valores exactos, la probabilidad de que una variable tome el valor en un intervalo se va a corresponder con el área del plano determinado por los extremos del intervalo y la función de densidad, es decir, por la integral siguiente:

$$P(a < X \leq b) = \int_a^b f(x)dx$$

Esta función de densidad verifica:

$i)\ f(x) \geq 0$

$ii)\ \int_{-\infty}^{+\infty} f(x)dx = 1$

Función de distribución

Llamaremos **función de distribución**, a la función definida por:

$$F(x_i) = P(X \leq x_i) = \int_{-\infty}^{x_i} f(x)dx$$

Gráficamente, la función de distribución es el área comprendida entre la función de densidad $f(x)$ y el eje X, desde $-\infty$ hasta x_i.

Además $F'(x) = f(x)$

Y $P(a < x \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_a^b f(x)dx$

Ejemplo:

Una variable aleatoria tiene por función de densidad: $f(x) = \begin{cases} 0 \ \ si\ \ x < 0 \\ 2x\ si\ 0 \leq x \leq 1 \\ 0 \ \ si\ x > 1 \end{cases}$

Comprobar que efectivamente es una función de densidad y calcula su función de distribución y $P(0{,}25 < X < 0{,}75)$

Evidentemente $f(x) \geq 0$

$\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 2x = 1$. Por tanto es una función de densidad.

Además, la función de distribución es:

$$F(x) = \begin{cases} 0 \; si \; x < 0 \\ x^2 \; si \; 0 \leq x \leq 1 \\ 1 \; si \; x > 1 \end{cases}$$

$P(0{,}25 < X < 0{,}75) = F(0{,}75) - F(0{,}25) = \frac{9}{16} - \frac{1}{16} = 0{,}5$

Media y Varianza de una distribución de probabilidad continua

Dada una variable aleatoria X continua, se llama media o esperanza matemática de la variable X al valor:

$$\mu = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

Y se llama varianza de la variable X al valor:

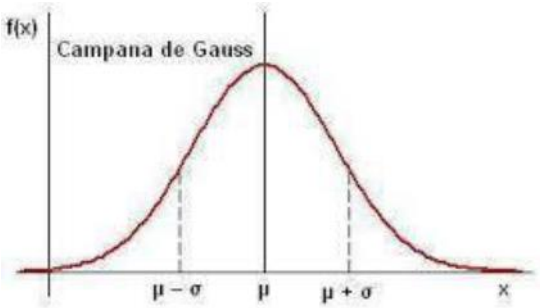
$$\sigma^2 = Var(X) = E((X - E(x)^2) = \int_{-\infty}^{+\infty} (x - E(X))^2 \cdot f(x)dx$$
$$= \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - E(X)^2$$

10. Distribución normal

La distribución normal fue usada por primera vez en 1753 por De Moivre, pero no tuvo continuidad hasta el siglo XIX, en que Gauss y Laplace la volvieron a utilizar. Por eso a veces se la conoce como la distribución de Gauss-Laplace.

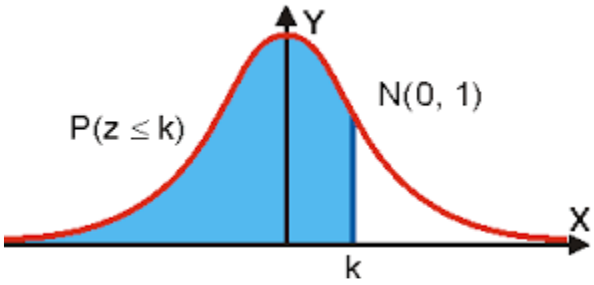
Se creía que, en la práctica, la mayor parte de las distribuciones eran de este tipo, y por eso se le llamaron NORMALES, y a las demás, distribuciones ANORMALES.

Una variable aleatoria continua X sigue una distribución normal $N(\mu, \sigma)$ si su función de densidad viene dada por : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, siendo μ la media y σ la desviación típica. Gráficamente:



Dentro de las distribuciones normales, la más importante es la que tiene media 0 y desviación típica 1. Se denota por $Z \in N(0,1)$ y su función de densidad es $f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}}$. Esta función es simétrica respecto al eje OY ya que $\mu = 0$.

Función de distribución : $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$ que geométricamente representa el área limitada por la función de densidad y el eje X, desde $-\infty$ hasta x:



Las probabilidades de una variable Z con distribución N(0,1) están tabuladas por lo que no es necesario calcular esas integrales (la tabla está al final).

La distribución N(0,1) es la más importante, ya que cualquier otra variable con una distribución normal $X \in N(\mu, \sigma)$, se puede transformar en una $Z \in N(0,1)$, esto se llama tipificar y se hace mediante el cambio de variable: $z = \frac{x-\mu}{\sigma}$

MANEJO DE LA TABLA:

- $P(X \leq a). Ejemplo: P(X \leq 1{,}85) = 0{,}9687\ directamente\ en\ la\ tabla$
- $P(X \leq -a) = P(X > a) = 1 - P(X \leq a)$
- $P(X > a) = 1 - P(X \leq a)$
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$
- $P(-a \leq X \leq b) = P(X \leq b) - P(X \leq -a) = P(X \leq b) - [1 - P(X \leq a)]$
- $P(-a \leq X \leq -b) = P(b \leq X \leq a) = P(X \leq a) - P(X \leq b)$
- $P(-a \leq X \leq a) = P(X \leq a) - P(X - a) = P(X \leq a) - [1 - P(X \leq a)] = 2P(X \leq a) - 1$

ÁREAS NOTABLES

- ✓ Área comprendida entre $\mu - \sigma \ \ y \ \ \mu + \sigma$
  $P(\mu - \sigma \leq x \leq \mu + \sigma) = P(-1 \leq Z \leq 1) = 2P(Z \leq 1) - 1 = 0{,}6826.$
  O sea, aproximadamente el 68% de las observaciones en una distribución $N(\mu, \sigma)$ se encuentran en ese intervalo.
- ✓ Área comprendida entre $\mu - 2\sigma \ \ y \ \ \mu + 2\sigma$
  $P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) = 2P(Z \leq 2) - 1 = 0{,}9544.$
  O sea, aproximadamente el 95% de las observaciones están dentro de ese intervalo
- ✓ Área comprendida entre $\mu - 3\sigma \ \ y \ \ \mu + 3\sigma$
  $P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) = 2P(Z \leq 3) - 1 = 0{,}9974.$
  O sea, casi el 100%.

Ejemplo:

Los resultados de un examen siguen una distribución normal de media 78 y desviación típica 36. Se pide:

a) ¿Cuál es la probabilidad de que una persona que se presenta al examen obtenga una calificación superior a 70?
b) ¿Cuál es la probabilidad de que tenga una puntuación entre 60 y 80?

Solución:

a) $P(X \geq 70) = P\left(\frac{X-\mu}{\sigma} \geq \frac{70-78}{36}\right) = P(Z \geq -0{,}2222) = P(Z \leq 0{,}2222) = 0{,}5871$

b) $P(60 \leq X \leq 98) = P\left(\frac{60-78}{36} \leq X \leq \frac{98-78}{36}\right) = P(-0{,}5 \leq Z \leq 0{,}5555) = P(Z \leq 0{,}5555) - P(Z \leq -0{,}5) = P(Z \leq 0{,}5555) - (1 - P(Z \leq 0{,}5) = P(Z \leq 0{,}5555) + P(Z \leq 0{,}5) - 1 = 0{,}7088 + 0{,}6915 - 1 = 0{,}4003$

11. Aproximación de la distribución por la distribución normal

Se puede probar que una distribución binomial $X \in B(n,p)$ suficientemente grande, bajo las condiciones $np > 5\ y\ nq > 5$, converge a una normal de media $\mu = np$ y desviación típica $\sigma = \sqrt{npq}$

Esto se conoce como teorema de Moivre: $X \in B(n,p) \equiv N(np, \sqrt{npq})$

Corrección por continuidad:

La distribución binomial es de variable discreta y, por tanto, tiene sentido calcular probabilidades en puntos fijados. Sin embargo, la distribución normal es de variable continua, y carece de sentido calcular probabilidades puntuales ya que todas son nulas. La aproximación de una variable discreta, X por una variable continua X', genera un error que se corrige modificando el intervalo donde se quiere calcular la probabilidad.

$P(X = a) = P(a - 0{,}5 < X' \leq a + 0{,}5)$

$P(a \leq X \leq b) = P(a - 0{,}5 \leq X' \leq b + 0{,}5)$

$P(X < a) = P(X' \leq a - 0{,}5)$

$P(X > a) = P(X' \geq a + 0{,}5)$