CONTENIDO DEL CURSO DE PROBABILIDAD Y ESTADÍSTICA

1. LA PROBABILIDAD Y ESTADÍSTICA Y SU RELACIÓN CON LOS EVENTOS

La Probabilidad y la Estadística se ven involucradas implícitamente en nuestra vida cotidiana; Son herramientas que utilizamos para interpretar una cantidad diversa de información y de datos; nos ayudan a valorar riesgos y dar mayor certeza al momento de tomar una decisión o interpretar un resultado.

Con este curso se pretende que los estudiantes aprendan los contenidos de la Estadística Descriptiva y la Probabilidad con el propósito de desarrollar el pensamiento matemático, además de que será un tema introductorio para los estudiantes que estén interesados en entender cómo funcionan las matemáticas aplicadas al análisis de datos.

Los temas a tratar, incluyen: Estadística descriptiva, técnica de conteo, permutaciones, combinaciones, distribuciones, y muchos temas más.

Objetivo 1.1 Enseñar a los estudiantes conceptos básicos referentes a la probabilidad y logren la comprensión de los mismos.

Todos en nuestra vida hemos escuchado frases como: "Es probable que mañana llueva", "Cuando tiramos un dado, es más probable sacar un número menor a tres que sacar un 6", "Es poco probable que me saque la lotería" y muchas más frases como estas; como ves, todos tenemos una noción intuitiva de probabilidad, pero, te has preguntado ¿qué es exactamente? Bueno, la **probabilidad** es simplemente qué tan posible es que ocurra una cosa o suceda al azar.; y al análisis de los eventos gobernados por la probabilidad se le llama **estadística.**

Estos son algunos enunciados que te ayudarán a tener una mejor noción acerca de qué es lo que verás en los ejercicios cuando estudiemos a la probabilidad:

- La probabilidad de un evento solo puede ser un número entre 0 y 1.
- La probabilidad de que ocurra un suceso puede escribirse como una fracción, un número decimal o como un porcentaje.
- La probabilidad del evento A suele escribirse como $P(A)$.
- Si $P(A) > P(B)$, el evento A tiene una mayor probabilidad de ocurrir que el evento B.
- Si $P(A) = P(B)$, los eventos A y B tienen la misma probabilidad de ocurrir.

Un **experimento** es cualquier proceso que proporciona datos, numéricos o no numéricos. Un ejemplo de un *experimento numérico* sería meter monedas de diferentes denominaciones en una caja e ir sacándolas ya que el resultado de este proceso, sería un número (valor de la moneda). Un *experimento no numérico* podría ser el clásico ejemplo donde metes varias canicas de colores en una bolsa y debido al azar, el resultado será un color, no un número.

El **espacio muestral** (Se representa por las letras S o Ω) es el conjunto de elementos que representan a todos los posibles resultados de un experimento; cada uno de estos resultados, se denominan **sucesos elementales** que serían subconjuntos del espacio muestral. Podemos plantear también que los sucesos son los resultados que queremos considerar de entre todos los resultados posibles, es decir, de entre todo el espacio muestral.

Un **suceso seguro** es aquel que contiene todo el espacio muestral.

Un **suceso imposible** es el caso contrario, cuando el suceso no contiene ningún elemento del espacio muestral.

Normalmente, estos sucesos se representan con A=B=0, que es el conjunto vacío, o lo que es lo mismo, estamos diciendo que no hay ningún resultado posible que cumpla el suceso.

Ahora bien, vamos a representar estos términos con un ejemplo para que nos quede mucho más claro:

Imaginemos que tenemos un dado de 6 caras, nuestro experimento sería, lanzar el dado y obtener como resultado el número de puntos de la cara que queda hacia arriba.

Nuestro espacio muestral estaría dado por Ω = {1, 2, 3, 4, 5, 6} ya que son todos los resultados posibles que podríamos obtener al lanzar el dado.

Cada uno de los elementos que conforman a Ω, es decir, el 1, el 2, el 3, el 4, el 5 y el 6, son cada uno, un suceso elemental del espacio muestral.

Un suceso seguro, para este caso, podría ser “sacar un número mayor a 0 y menor a 10”, ya que todos los elementos del espacio muestral cumplen con esa condición; por el contrario, un suceso imposible sería “obtener un 7 al lanzar el dado” ya que no hay ningún resultado posible que cumpla este suceso.

ACTIVIDAD PROPUESTA:

Tenemos una caja cerrada, con un ejemplar de cada moneda mexicana como las de la imagen, abrimos la caja y sacamos una moneda al azar para saber su valor en pesos mexicanos. Marca la respuesta correcta:

1. Se trata de:
a) Un experimento numérico
b) Un experimento no numérico

2. El espacio muestral es:
a) Ω = {1, 2, 3, 4, 5, 6}
b) Ω = {0.5, 1, 2, 5, 10, 20}
c) Ω = {50, 1, 2, 5, 10, 20}

3. No es un suceso elemental:
a) 2
b) 3
c) 1

4. Es un suceso seguro:
a) Sacar una moneda cuyo valor sea múltiplo de 2
b) sacar una moneda de tres pesos
c) Sacar una moneda cuyo valor sea mayor a 0 y menor a 100

RESPUESTAS:

1. **a**
2. **b (El problema dice “valor en pesos” y 50 centavos es 50/100=0.5 pesos)**
3. **b**
4. **c**

Objetivo 1.2 Determinar la probabilidad de que ocurra un evento al contar resultados empleando la fórmula de Laplace.

Cuando hablamos de experimentos probabilísticos sencillos, como lanzar una moneda al aire, tirar un dado, sacar una carta de un mazo, etc, donde sabemos que los resultados son **equiprobables**, es decir, que cada suceso elemental de nuestro espacio muestral tiene la misma probabilidad de salir como resultado, utilizamos una fórmula súper sencilla e importante para determinar la probabilidad de uno de estos sucesos, estamos hablando de la **Regla de Laplace** que nos dice que:

La probabilidad de que ocurra un suceso A se calcula de la siguiente manera:

$$P(A) = \frac{\text{Casos favorables de A}}{\text{Casos posibles}}$$

Debemos de tener sumo cuidado en ordenar los resultados y comprobar que sean equiprobables.

EJEMPLO:

¿Cuál es la probabilidad de que al lanzar una moneda, salga cruz?
Los casos favorables, en este caso sólo es 1, que el resultado sea cruz.
Los casos posibles son 2, que el resultado sea cara o sea cruz.

$$P(A) = \frac{1}{2}$$

La probabilidad de que al lanzar una moneda y que el resultado sea cruz, es de ½ o bien, 0.5 que resulta de dividir 1 entre dos. El resultado también lo podemos convertir en porcentaje, tomando en cuenta que la unidad es el 100%, 0.5 sería equivalente al 50%.

EJERCICIOS RESUELTOS:

a) ¿Cuál es la probabilidad de que al lanzar un dado salga como resultado 5?

Al lanzar un dado, el espacio muestral está dado por Ω= {1, 2, 3, 4, 5, 6}; por lo tanto hay 6 casos posibles. Si el 5 sólo es 1 caso favorable, la probabilidad de que al lanzar un dado el resultado será:

$$P(A) = \frac{1}{6}$$

b) En una bolsa, hay 10 bolas numeradas del 11 al 20, algunas rojas y otras verdes. Sacamos sin mirar una bola, ¿Cuál es la probabilidad de sacar un número primo? 11, 13, 17, 19

Los números primos ubicados entre el 11 y el 20 son el 11, 13, 17 y el 19 → 4 Resultados favorables

Ω= {11, 12, 13, 14, 15, 16, 17, 18, 19, 20} → 10 Resultados posibles

$$P(A) = \frac{4}{10} = \frac{2}{5}$$

c) En una comida hay 28 hombres y 32 mujeres. Han comido carne 16 hombres y 20 mujeres, comiendo pescado el resto. Si se elige una de estas personas al azar, ¿Cuál es la probabilidad de que la persona elegida sea hombre?

La información de lo que han comido es irrelevante. Lo importante aquí es que los casos favorables son 28, que es el número de hombres que hay en la cena; y el número de casos posibles sería el número total de personas en la cena, es decir, la suma de hombres y mujeres (28 hombres + 32 mujeres), 60 personas en total.

$$P(A) = \frac{28}{60} = \frac{14}{30} = \frac{7}{15}$$

d) ¿Cuál es la probabilidad de al sacar una carta al azar de un naipe inglés (52) cartas, esta sea un as? Expresar la probabilidad en número decimal.

Si los resultados favorables son 4, ya que hay un as por figura, y los resultados posibles son 52, la probabilidad de obtener un as del mazo es:

$$P(A) = \frac{4}{52} = 0.769 = \frac{1}{13}$$

e) Violeta lanza un dado, ¿Cuál es la probabilidad de que ella obtenga un número menor que 3? Expresar el valor de la probabilidad en porcentaje.

Ω= {1, 2, 3, 4, 5, 6} → 6 casos posibles

1 y 2 son los sucesos menores a 3 → 2 Casos favorables

$$P(A) = \frac{2}{6} = \frac{1}{3} = 33.\overline{33}\%$$

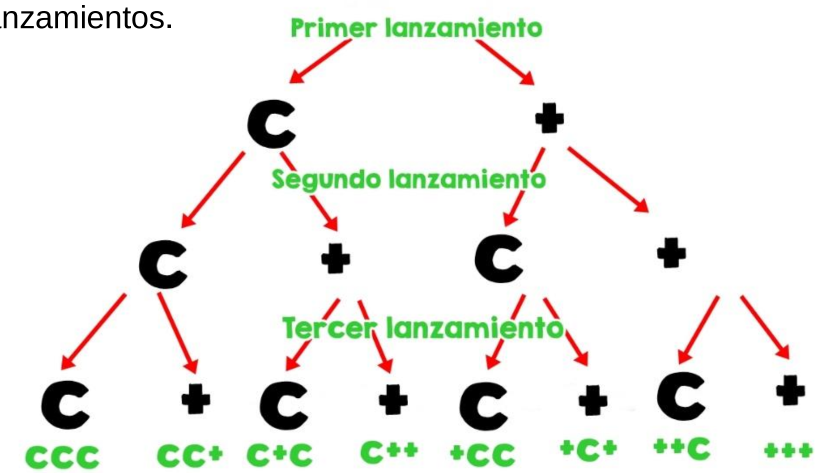
Objetivo 1.3 Explicar en qué consiste la probabilidad compuesta, así como analizar el uso de diagramas de árbol para resolver experimentos compuestos.

La probabilidad compuesta o condicionada mide la probabilidad de un determinado suceso conociendo información previa sobre otro suceso; es decir, aquella donde intervienen más de un experimento aleatorio. Por ejemplo, tirar dos monedas al aire y mirar si salen dos caras o sacar dos canicas de una bolsa y que una sea blanca.

Para tener una noción un poco más clara de cómo esto ocurre, nos ayudamos de los **diagramas de árbol**.

EJEMPLO: Se lanza tres veces una moneda al aire. Queremos saber la probabilidad de que ocurra el suceso A= Sacar al menos dos cruces en los tres tiros.

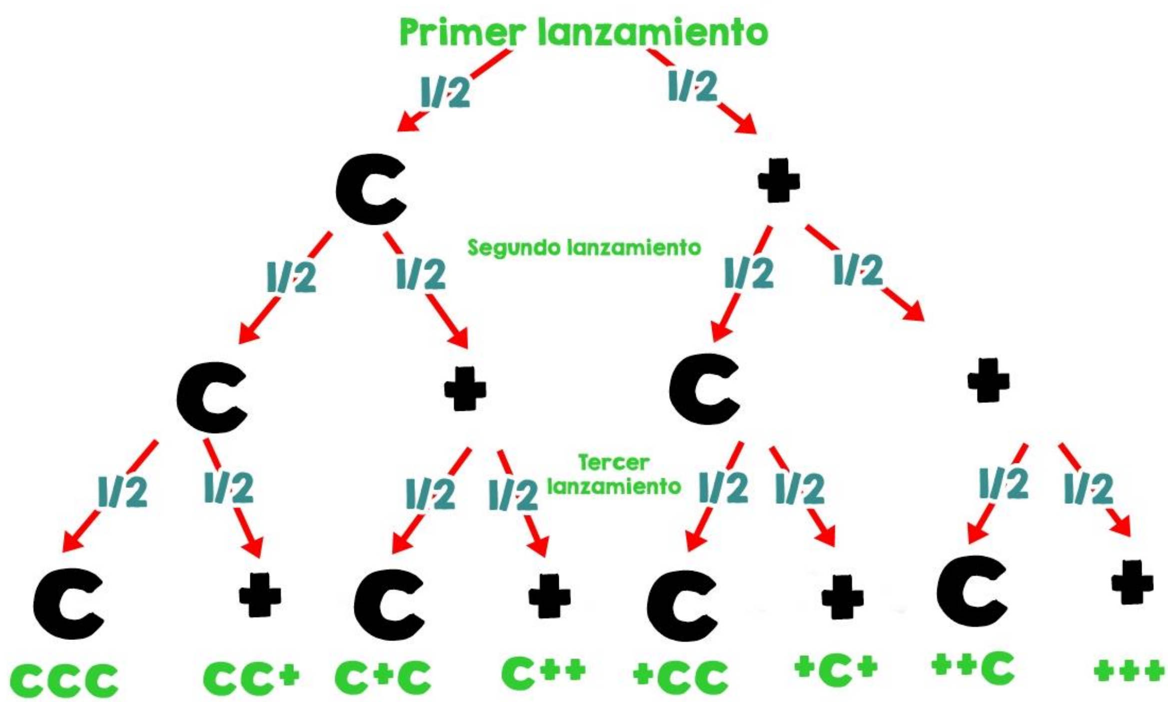
SOLUCIÓN: Para tener un panorama más claro de cuáles son los sucesos posibles en este experimento, nos ayudamos de un diagrama de árbol y observamos cómo cada **rama** los posibles caminos o sucesos que pueden resultar de los tres lanzamientos.



Si primero sale C, luego + y en el último lanzamiento vuelve a salir +, esta ramificación corresponde al suceso elemental C++. De esta manera, interpretamos que el espacio muestral está definido por Ω = {CCC, CC+, C+C, C++, +CC, +C+, ++C, +++}.

Para encontrar la probabilidad de cada rama, o cada suceso elemental, es necesario multiplicar las probabilidades de todas las ramas que hemos seguido hasta llegar al final del árbol.

Si cada que se hace un lanzamiento, la probabilidad de que salga C o + es de ½, el diagrama quedaría de esta manera:



Así, sabiendo que cada pequeña rama vale ½ y que para obtener el valor de cada suceso elemental sólo basta con multiplicar las ramas hasta llegar al final, quiere decir que cada suceso elemental tiene una probabilidad de salir de:

$$P(Cada\ rama\ completa) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

Ahora, si se desea saber la probabilidad de que ocurra el suceso A= Sacar al menos dos cruces en los tres tiros, los resultados posibles según nuestro espacio muestral son: C++, +C+, ++C y +++.

$$P(C++) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

$$P(+C+) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

$$P(++C) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

$$P(+++) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

Por lo tanto, si para cada rama completa, la probabilidad es de $\frac{1}{8}$, el valor de estos 4 sucesos elementales viene dado por la suma de estas cuatro probabilidades.

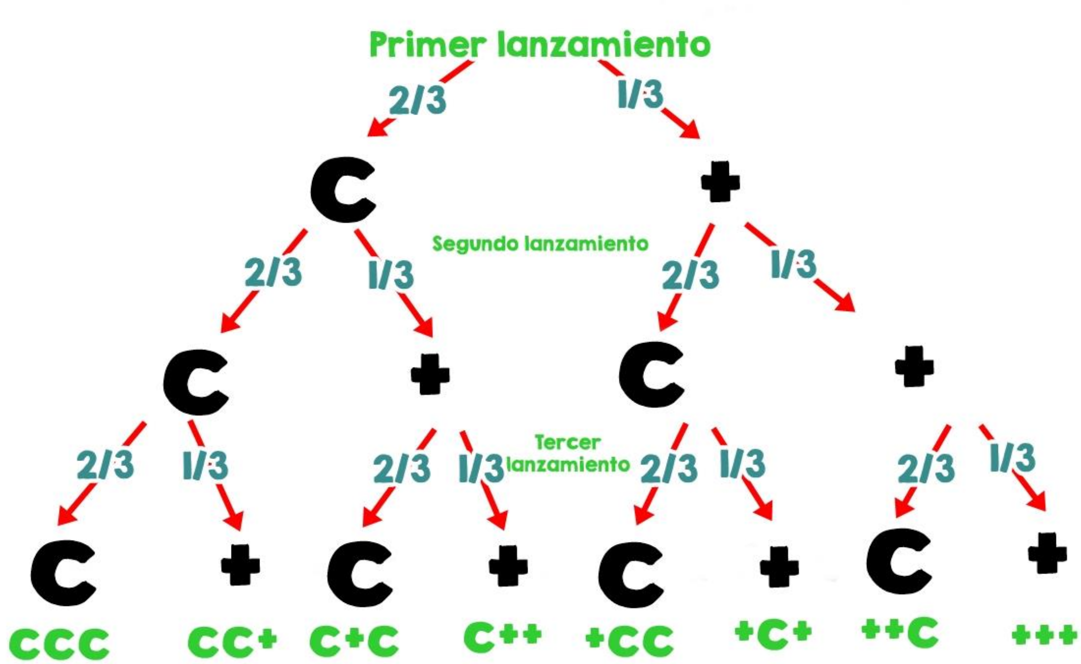
$$P(A) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

EJERCICIO RESUELTO:

Calcular la probabilidad de que al lanzar una moneda descompensada tres veces, nos salgan por lo menos dos caras, no importando el orden de estos resultados. Nota: La cara tiene una probabilidad de salir de 2/3.

SOLUCIÓN: Al estar descompensada la moneda, nos da a entender que es más probable que salga más un lado que el otro; y si en el problema nos dice que la cara tiene una probabilidad de salir de 2/3, la cruz tendría sólo 1/3 de probabilidad.

Nuestro diagrama de árbol nos quedaría de la siguiente manera:



Los casos favorables de A son, CCC, CC+, C+C y +CC; y la probabilidad de cada suceso elemental que contiene dos caras quedaría de la siguiente manera:

$$P(CCC) = \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{8}{27}$$

$$P(CC+) = \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{4}{27}$$

$$P(C+C) = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{4}{27}$$

$$P(+CC) = \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{4}{27}$$

Por lo tanto, la probabilidad de que al lanzar esta moneda descompensada, el resultado sea de por lo menos dos caras, es de:

$$P(A) = \frac{8}{27} + \frac{4}{27} + \frac{4}{27} + \frac{4}{27} = \frac{20}{27} = 0.740 = 74\%$$

Junto a la probabilidad compuesta, le acompañan términos importantes que debemos tomar en cuenta; como si los sucesos son compatibles o no.

Consideremos los siguientes sucesos al lanzar un dado de 6 caras: A= {6,1} B= {1,3} y C= 4. Observamos que si sacamos un 1, se cumple el suceso A como B; decimos entonces, que son **sucesos compatibles** ($A \cap B=0$), por el contrario, A y C son **sucesos incompatibles** ya que no se pueden dar al mismo tiempo ($A \cap B \neq 0$).

Un **suceso complementario** es el subconjunto que unido a los sucesos de determinada condición, da como resultado un suceso seguro.

La **Unión** de A y B se simboliza como **AUB** es el suceso formado por todos los resultados que cumplen A o cumplen B.

Si A= {2, 4, 6} y B= {4, 5, 6}, la unión de A y B sería igual a:

AUB= {2, 4, 6} U {4, 5, 6} = {2, 4, 5, 6}

Para calcular la **PROBABILIDAD DE UNA UNIÓN**, seguimos la siguiente regla:

$$P(AUB) = P(A) + P(B)$$

$$P(AUB) = P(A) + P(B) – P(A \cap B)$$

La **intersección** de A y B se simboliza como **A∩B** es el suceso formado por todos los resultados que cumplen A y cumplen B.

Si A= {2, 4, 6} y B= {4, 5, 6}, la intersección de A y B sería igual a:

AUB= {2, 4, 6} ∩ {4, 5, 6} = {4, 6}

Para calcular la **PROBABILIDAD DE UNA INTERSECCIÓN**, seguimos la siguiente regla:

$$P(A \cap B)= P(A) \cdot P(B)$$

$$P(A \cap B)= P(A/B) \cdot P(B)$$

Cuando se producen procesos estocásticos consecutivamente de un espacio muestral, puede darse cualquiera de estas dos situaciones:

- Los sucesos son independientes uno del otro.
- Cada suceso es condicionado por el anterior.

Cuando un suceso A influye en el resultado de un segundo suceso B, se dice que tiene una **PROBABILIDAD CONDICIONADA**, expresada como P(B/A), cuyo valor es:

$$\boxed{P(B/A) = \frac{P(A \cap B)}{P(A)} \; si \; P(A) \neq 0}$$

De la fórmula de la probabilidad condicionada, podemos encontrar una expresión que nos resultará muy útil:

$$\boxed{P(A \cap B) = P(B/A) \cdot P(A)}$$

Esta fórmula define al **PRINCIPIO DE LA PROBABILIDAD COMPUESTA**.

EJEMPLO 1: Un 15% de los pacientes atendidos en un hospital son hipertensos, un 10% son obesos y un 3% son hipertensos y obesos. ¿Qué probabilidad hay de que elegido un paciente al azar sea obeso o hipertenso?

$\cap$ = Y

SOLUCIÓN:

$\cup$ = ó

A= {Hipertenso} P(A)= 15% P(A)= 0.15

B= {Obeso} P(B)= 10% P(B)= 0.1

A∩B= {Hipertenso y obeso} P(A∩B)= 3% P(A∩B)= 0.03

AUB= {Hipertenso u obeso} P(AUB)= 0.15+0.01-0.03= 0.22

La probabilidad de que al elegir un paciente al azar y sea obeso o hipertenso es del 22%, de 0.22 o de 22/100.

EJEMPLO 2: Determine el valor de la probabilidad de obtener un 2, sabiendo que ha salido un número par al lanzar un dado al aire.

SOLUCIÓN:

Primero se definen los sucesos:

P(A) = P(al lanzar el dado, el resultado es par) A={2, 4, 6}

P(B) = P(Que salga el número 2) B={2}

P(A∩B) = P(Que al lanzar un dado, el resultado sea par y sea el 2) A∩B={2}

Se calculan las probabilidades clásicas para cada caso usando la fórmula de Laplace:

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

$$P(\mathrm{A} \cap \mathrm{B}) = \frac{1}{6}$$

Finalmente, se calcula la probabilidad condicionada:

$$P(B/A) = \frac{\mathrm{P}(\mathrm{A} \cap \mathrm{B})}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{2}{6} = \frac{1}{3}$$

La probabilidad de que obtengamos un 2, sabiendo que ha salido un número par al lanzar un dado al aire es de 1/3, 0.33 o del 33.33%

EJEMPLO 3: Se tiene una caja con tres bolas blancas y dos rojas, si se extraen al azar dos bolas consecutivamente y sin reemplazamiento de la caja, determine la probabilidad de que:

a. Ambas sean blancas.

b. Ambas sean rojas.

c. Sean de diferente color.

SOLUCIÓN:

Se definen los sucesos:

B= {extraer una bola blanca}

R= {extraer una bola roja}

a) La probabilidad de extraer ambas bolas blancas será

$$P(B_1 \cap B_2) = P(B_1) \cdot P(B_2/B_1) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{6}{20} = \frac{3}{10} = 0.3 = 30\%$$

Los valores de $P(B_1)\ y\ P(B_2/B_1)$ se determinaron en base a la probabilidad clásica, según la fórmula de Laplace.

b) La probabilidad de extraer ambas bolas rojas será

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2/R_1) = \left(\frac{2}{5}\right)\left(\frac{1}{4}\right) = \frac{2}{20} = \frac{1}{10} = 0.1 = 10\%$$

c) La probabilidad de extraer bolas de diferente color implica que la primera extracción sea una bola blanca y la segunda roja o viceversa, lo cual implica una unión de probabilidades, por lo tanto, este valor de probabilidad será:

P(una roja y una blanca) =

$$P(B_1 \cap R_2) + P(R_1 \cap B_2) = \left(\frac{3}{5} \cdot \frac{2}{4}\right) + \left(\frac{2}{5} \cdot \frac{3}{4}\right) = \frac{6}{20} + \frac{6}{20} = \frac{12}{20} = \frac{3}{5} = 0.6 = 60\%$$

EJERCICIOS RESUELTOS:

1. Se sabe que el 50% de la población fuma, un 65% es hipertensa y que el 25% fuma y es hipertensa. ¿Cuál es la probabilidad de que una persona elegida al azar sea una persona fumadora o hipertensa?

SOLUCIÓN:

A= {ser hipertenso}   P(A)= 0.65

B= {ser fumador}   P(B)= 0.5

A ∩ B = {ser hipertenso y fumador}   P(A ∩ B)= 0.25

AUB= {ser hipertenso o fumador}   P(AUB)= 0.65+0.5-0.25= 0.9

La probabilidad de que al elegir un paciente al azar y sea fumador o hipertenso es del 90%, de 0.9 o de 9/10.

2. En una ciudad hay dos camiones de bomberos. La probabilidad de que un cada camión esté disponible si se necesitara es de 90%. Hay un incendio, se desea saber el valor de la probabilidad de que:

   a. Ninguno,

   b. Solo uno,

   c. Ambos,

   Estén disponibles para acudir al llamado.

SOLUCIÓN:

a. La probabilidad según el ejercicio de que no estén disponibles sería del 10%, ya que 100% - 90% = 10%. Ahora bien, la probabilidad de que ninguno esté disponible (ND) es de:

$$P(ninguno) = P(ND_1 \cap ND_2) = P(ND_1) \cdot P(ND_2) = 0.1 * 0.1 = 0.01 = 1\%$$

b. La probabilidad de solo un camión esté disponible, implica que el primero lo esté y el segundo no o viceversa, lo cual es una unión de probabilidades, por lo tanto, tal valor de probabilidad será:

$$P(sólo\ uno) = P(D_1 \cap ND_2) + P(ND_1 \cap D_2) = (0.9 * 0.1) + (0.1 * 0.9) = 0.18 = 18\%$$

c. La probabilidad de que ambos estén disponibles es de:

$$P(ambos) = P(D_1 \cap D_2) = P(D_1) \cdot P(D_2) = 0.9 * 0.9 = 0.81 = 81\%$$

3. En una habitación de cierto hospital, la probabilidad de que un paciente sea ingresado con problemas de presión arterial es del 70%, la probabilidad de que ingrese con problemas renales es 0,45 y la de que ingrese con ambos problemas es de 15/50. Hallar:
   a. La probabilidad de que un paciente que ha sido ingresado con problemas de presión arterial, padezca también de problemas renales.
   b. La probabilidad de que un paciente que ha sido ingresado con problemas renales, presente problemas de presión arterial.

SOLUCIÓN:

P= {Paciente que ingresa con problemas de presión arterial}
R= {Paciente que ingresa con problemas renales}
Las probabilidades las convertimos a decimales.
a. La probabilidad de que un paciente con problemas de presión arterial también tenga problemas renales será:

$$P(R/P) = \frac{P(P \cap R)}{P(P)} = \frac{0.3}{0.7} = 0.428 \cong 0.43\ o\ 43\%$$

b. La probabilidad de que un paciente que ha sido ingresado con problemas renales, presente problemas de presión arterial será:

$$P(P/R) = \frac{P(P \cap R)}{P(R)} = \frac{0.3}{0.45} = 0.666 \cong 0.67\ o\ 67\%$$

Objetivo 1.4 Explicar la diferencia entre variaciones, permutaciones y combinaciones, así como las fórmulas para la obtención de resultados.

La **combinatoria** es la parte de la matemática que estudia la formación de subconjuntos o agrupaciones de elementos partiendo de un conjunto dado, teniendo en cuenta la ordenación y el número de esos elementos.

Ahora me toca explicarte cuáles son las características de estos subgrupos; veremos las tres formas posibles de agruparlos y seleccionar a los elementos; estas son las variaciones, permutaciones y combinaciones.

Pero antes, quiero hacer una explicación breve de lo que es el famoso “factorial”. El factorial de un número natural n es el resultado del producto de este número por todos los números naturales que lo anteceden, excluyendo al cero y se representa por el símbolo !.

a) Variaciones

Estos subgrupos ocurren cuando tenemos que agrupar cierto número de elementos en una cantidad específica dentro del grupo.

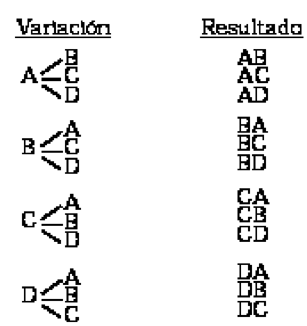
Para determinar las variaciones de un grupo de n elementos tomados de r en r, seguimos la siguiente fórmula:

$$V_{n,r} = \frac{n!}{(n-r)!}$$

Ejemplo: ¿Cuántas variaciones existen entre los elementos A, B, C y D tomados de 2 en 2?

$$V_{4,2} = \frac{4!}{(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 4 \cdot 3 = 12$$

R. 12 variaciones.



b) Permutaciones

Son variaciones de n elementos tomados en grupos de r, en el que n = r. Cada agrupación difiere de las restantes sólo en el orden de colocación de los elementos, y en cada grupo intervienen todos los elementos del conjunto. El número de permutaciones ordinarias formadas con n elementos viene dado por:

$$P_n = n(n-1)(n-2)\ldots(1) = n!$$

Permutación			Resultado
A	B	C	ABC
	C	B	ACB
B	A	C	BAC
	C	A	BCA
C	A	B	CAB
	B	A	CBA

Ejemplo: ¿Cuántas permutaciones ordinarias (sin repetición) se pueden formar con los elementos A, B, C?

$P_3 = 3! = 3 \cdot 2 \cdot 1 = 6$ R. 6 permutaciones.

c) Combinaciones

Se obtienen al seleccionar de un conjunto de n elementos grupos de r, de tal forma que sin importar el orden de sus elementos, estos no se repiten.

El número de combinaciones ordinarias (sin repetición) que se pueden formar con n elementos tomados de r en r se calcula a partir de la siguiente fórmula:

$$C_{n,r} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r}$$

La expresión $\binom{n}{r}$ se lee «n sobre r» y se denomina **número combinatorio**.

El número de combinaciones de n elementos en grupos de r es igual al de variaciones de n en grupos de r dividido por el número de permutaciones de los r elementos de cada grupo. Es decir:

$$C_{n,r} = \frac{V_{n,r}}{P_r}$$

Combinación		Resultado
A	B	AB
	C	AC
	D	AD
B	C	BC
	D	BD
C	D	CD

¿Cuál es el número de permutaciones que pueden formarse con los elementos A, B, C y D tomados de dos en dos?

$$C_{4,2} = \frac{4!}{2!\,(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 2 \cdot 1} = \frac{24}{4} = 6$$

R. 6 Permutaciones

EJEMPLOS:

1. ¿De cuántas maneras pueden sentarse 10 personas en un banco si hay sólo 4 sitios disponibles?

$$V_{10,4} = \frac{10!}{(10-4)!} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$$

R. De 5040 maneras.

2. En una clase de 10 alumnos van a distribuirse 3 premios, suponiendo que los alumnos no pueden recibir más de un premio, cuántas maneras posibles hay de repartir los premios si:
   a. Los premios son diferentes.
   b. Los premios son iguales.

   a. Los premios son diferentes, no es lo mismo ganar el primer premio que el segundo.

$$V_{10,3} = \frac{10!}{(10-3)!} = 10 \cdot 9 \cdot 8 = 720\ maneras$$

   b. Los premios son iguales, no importa el orden, pueden distribuirse de

$$C_{10,3} = \frac{10!}{3!\,(10-3)!} = \frac{10!}{3!\,7!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120\ maneras$$

3. Hay que colocar a 5 hombres y 4 mujeres en una fila de modo que las mujeres ocupen los lugares pares. ¿De cuántas maneras puede hacerse?
   Como la fila es de 9 personas HMHMHMHMH en total, hay 4 posiciones pares (que deben ser ocupadas por las 4 mujeres) y 5 posiciones impares (para los 5 hombres).

$$P_4 = 4! = 24\ mujeres$$

$$P_5 = 5! = 120\ hombres$$

$$24 \cdot 120 = 2880\ maneras\ posibles$$

EJERCICIOS RESUELTOS:

1. Cuántos números de 4 dígitos pueden formarse con las cifras de 1 al 9…
    a. Permitiendo repeticiones
    b. Sin repeticiones
    c. Si el último dígito es 1 y no se permiten repeticiones

SOLUCIÓN:

a. Permitiendo repeticiones e importa el orden (no es lo mismo el 1224 que el 2214):

$$V_{9,4} = 9^4 = 6561\ números$$

b. No se permiten repeticiones e importa el orden:

$$V_{9,4} = \frac{9!}{(9-4)!} = 9 \cdot 8 \cdot 7 \cdot 6 = 3024\ números$$

c. Fijamos el último dígito (el número 1 está en la última posición) y como no puede haber repeticiones de números, nos quedan 8 números para 3 posiciones) se obtiene un total de:

$$V_{8,3} = \frac{8!}{(8-3)!} = 8 \cdot 7 \cdot 6 = 336\ números$$

2. Tres atletas toman parte en una competencia, ¿De cuántas maneras podrán llegar a la meta? (Pueden llegar juntos)

SOLUCIÓN:

Existen varias posibilidades:

- Si llegan los tres juntos, entonces sólo hay **1 posibilidad**.
- Si llegan dos juntos, existen
  $C_{3,2} = \frac{3!}{2!(3-2)!} = \frac{3}{1} = \mathbf{3\ grupos}$ de dos que llegan juntos y $P_2 = 2! = 2 \cdot$

$1 = \mathbf{2\ posibilidades}$ distintas del grupo de dos y el otro atleta, por lo que existen en total 3*2= **6 posibilidades**

- Si llegan los tres por separado, existen $P_3 = 3! = 3 \cdot 2 \cdot 1 = \mathbf{6\ posibilidades}$

Por lo tanto, pueden llegar a la meta de 1+6+6= **13 maneras distintas.** **R.**

3. En un hospital se usan 5 símbolos para clasificar las historias clínicas de sus pacientes, de manera que los dos primeros son letras y los tres últimos son dígitos. Suponiendo que hay 25 letras, cuántas historias clínicas podrían hacerse si:
   a. No hay restricciones sobre letras y números.
   b. Las dos letras no pueden ser iguales.

SOLUCIÓN:

a. Es necesario tener en cuenta el orden de las dos letras elegidas y que pueden repetirse, hay entonces $VR_{25,2} = 25^2 = 625$ posibilidades para las letras. Se hace lo mismo con los dígitos, serían entonces $VR_{10,3} = 10^3 = 1000$ posibilidades. El total de las historias clínicas posibles es entonces de Total= 625*1000=**625000**

b. Se hace de forma similar que el inciso a; con la diferencia de que ahora las letras no pueden repetirse; entonces hay $V_{25,2} = \frac{25!}{(25-2)!} = 25 \cdot 24 = 600$ posibilidades para las letras y $VR_{10,3} = 10^3 = 1000$ posibilidades para los números, resultando que hay en Total= 600*1000= **600000 historias clínicas**.

Objetivo 1.4 Interpretar brevemente el papel que juega el muestreo en la ciencia de datos.

2. DISTRIBUCIÓN DE DATOS

El conocimiento de la distribución de datos es esencial para elegir el método estadístico correcto.

En estadística descriptiva hemos comentado que los datos analizados pueden presentarse comúnmente mediante tres formas básicas: en forma textual, mediante cuadros estadísticos y mediante gráficas.

Objetivo 2.1 Explicar detalladamente cómo fabricar una tabla para que nuestros datos estén organizados.

La tabla de frecuencias (o distribución de frecuencias) es una tabla que muestra la distribución de los datos mediante sus frecuencias. Se utiliza para variables cuantitativas o cualitativas ordinales.

La tabla de frecuencias es una herramienta que permite ordenar los datos de manera que se presentan numéricamente las características de la distribución de un conjunto de datos o muestra.

X_i	Frecuencia absoluta (n_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa (fi = ni/N)	Frecuencia relativa acumulada (Fi=Ni/N)
1	7	7	0,06	0,06
2	19	26	0,15	0,21
3	25	51	0,20	0,41
4	12	63	0,10	0,50
5	23	86	0,18	0,69
6	15	101	0,12	0,81
7	8	109	0,06	0,87
8	16	125	0,13	1,00
Total	**125**	**125**	**1**	**1**

Construcción de la tabla de frecuencias

1. En la primera columna se ordenan de menor a mayor los diferentes valores que tiene la variable en el conjunto de datos.
2. En las siguientes columnas (segunda y tercera) se ponen las frecuencias absolutas**,** que es el conteo de cada una de estas variables y las frecuencias absolutas acumuladas que resulta de ir sumando la frecuencia absoluta con la frecuencia absoluta acumulada.
3. Las columnas cuarta y quinta contienen la las frecuencias relativas y las frecuencias relativas acumuladas.
4. Adicionalmente (opcional) se pueden incluir dos columnas (sexta y séptima), representando la frecuencia y la frecuencia relativa acumulada como tanto por cien. Estos porcentajes se obtienen multiplicando las dos frecuencias por cien.

Tipos de frecuencias

a. Frecuencia absoluta

La frecuencia absoluta (n_i) de un valor X_i es el número de veces que el valor está en el conjunto (X_1, X_2,…, X_N).

La suma de las frecuencias absolutas de todos los elementos diferentes del conjunto debe ser el número total de sujetos *N*. Si el conjunto tiene *k* números (o categorías) diferentes, entonces:

$$\sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_k = N$$

b. Frecuencia absoluta acumulada.

La frecuencia absoluta acumulada (N_i) de un valor X_i del conjunto (X_1, X_2,…, X_N) es la suma de las frecuencias absolutas de los valores menores o iguales a X_i, es decir:

$$N_i = n_1 + n_2 + \ldots + n_i$$

c. Frecuencia relativa.

La frecuencia relativa (f_i) de un valor X_i es la proporción de valores iguales a X_i en el conjunto de datos (X_1, X_2,…, X_N). Es decir, la frecuencia relativa es la frecuencia absoluta dividida por el número total de elementos *N*:

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

siendo $(X_1, X_2, ..., X_N)$ el conjunto de datos y n_i el total de valores igual a X_i

Las frecuencias relativas son valores entre 0 y 1, $0 \leq f_i \leq 1$. La suma de las frecuencias relativas de todos los sujetos da 1. Supongamos que en el conjunto tenemos *k* números (o categorías) diferentes, entonces:

$$\sum_{i=1}^{k} f_i = f_1 + f_2 + \ldots + f_k = 1$$

Si se multiplica la frecuencia relativa por cien se obtiene el **porcentaje**.

d. Frecuencia relativa acumulada.

Definimos la frecuencia relativa acumulada (Fi) de un valor X_i como la proporción de valores iguales o menores a X_i en el conjunto de datos (X_1, X_2,…, X_N). Es decir, la frecuencia relativa acumulada es la frecuencia absoluta acumulada dividida por el número total de sujetos *N*:

$$F_i = \frac{N_i}{N}$$

siendo $(X_1, X_2, ..., X_N)$ el conjunto de datos y N_i el total de valores igual o menor a X_i

La frecuencia relativa acumulada de cada valor siempre es mayor que la frecuencia relativa. De hecho, la frecuencia relativa acumulada de un elemento es la suma de las frecuencias relativas de los elementos menores o iguales a él, es decir:

$$F_i = f_1 + f_2 + \ldots + f_i$$

Ejemplo

Una profesora tiene la lista de las calificaciones en matemáticas de 30 alumnos de su clase. Las notas son las siguientes:

NOTAS EN MATEMÁTICAS DE 30 ALUMNOS									
6	10	5	5	4	4	6	6	5	4
6	7	7	5	6	3	6	7	9	5
6	5	7	3	8	8	4	7	8	9

1) Frecuencia absoluta

X_i	Frecuencia absoluta (n_i)
3	2
4	4
5	6
6	7
7	5
8	3
9	2
10	1
Total	30

Se realiza el **recuento** de la variable que se estudia (notas) para ver el número de veces que aparece cada nota.

Una vez realizado el recuento, se representan las frecuencias absolutas de cada una de las notas (n_i). Las frecuencias son: $n_1(3)=2$, $n_2(4)=4$, $n_3(5)=6$, $n_4(6)=7$, $n_5(7)=5$, $n_6(8)=3$, $n_7(9)=2$ y $n_8(10)=1$.

2) Frecuencia absoluta acumulada

Se calculan las frecuencias absolutas acumuladas (N_i) como la suma de las frecuencias absolutas de los valores menores o iguales a X_i:

X_i	Frecuencia absoluta (n_i)	Frecuencia absoluta acumulada (N_i)
3	2	2
4	4	6
5	6	12
6	7	19
7	5	24
8	3	27
9	2	29
10	1	30
Total	30	30

$\mathbf{N_1(3)}=n_1(3)=\mathbf{2}$
$\mathbf{N_2(4)}=n_1(3)+n_2(4)=2+4=\mathbf{6}$
$\mathbf{N_3(5)}=n_1(3)+n_2(4)+n_3(5)=2+4+6=\mathbf{12}$
$\mathbf{N_4(6)}=n_1(3)+n_2(4)+n_3(5)+n_4(6)=2+4+6+7=\mathbf{19}$
$\mathbf{N_5(7)}=n_1(3)+n_2(4)+n_3(5)+n_4(6)+n_5(7)=2+4+6+7+5=\mathbf{24}$
$\mathbf{N_6(8)}=n_1(3)+n_2(4)+n_3(5)+n_4(6)+n_5(7)+n_6(8)=2+4+6+7+5+3=\mathbf{27}$
$\mathbf{N_7(9)}=n_1(3)+n_2(4)+n_3(5)+n_4(6)+n_5(7)+n_6(8)+n_7(9)=2+4+6+7+5+3+2=\mathbf{29}$
$\mathbf{N_8(10)}=n_1(3)+n_2(4)+n_3(5)+n_4(6)+n_5(7)+n_6(8)+n_7(9)+n_8(10)=2+4+6+7+5+3+2+1=\mathbf{30}$

3) Frecuencia relativa

Se calcula la frecuencia relativa de cada elemento como la división de la frecuencia absoluta entre el total de elementos *N*=30.

X_i	Frecuencia absoluta (n_i)	Frecuencia relativa (fi = ni/N)	Frecuencia relativa (fi = ni/N) en %
3	2	0,07	7%
4	4	0,13	13%
5	6	0,20	20%
6	7	0,23	23%
7	5	0,17	17%
8	3	0,10	10%
9	2	0,07	7%
10	1	0,03	3%
Total	**30**	**1**	**100%**

$\mathbf{f_1(3)} = n_1(3)/N = 2/30 = \mathbf{0,07}$
$\mathbf{f_2(4)} = n_2(4)/N = 4/30 = \mathbf{0,13}$
$\mathbf{f_3(5)} = n_3(5)/N = 6/30 = \mathbf{0,20}$
$\mathbf{f_4(6)} = n_4(6)/N = 7/30 = \mathbf{0,23}$
$\mathbf{f_5(7)} = n_5(7)/N = 5/30 = \mathbf{0,17}$
$\mathbf{f_6(8)} = n_6(8)/N = 3/30 = \mathbf{0,10}$
$\mathbf{f_7(9)} = n_7(9)/N = 2/30 = \mathbf{0,07}$
$\mathbf{f_8(10)} = n_8(10)/N = 1/30 = \mathbf{0,03}$

Se pueden calcular las frecuencias relativas en porcentaje (%) multiplicándolas por 100.

4) Frecuencia relativa acumulada

Para obtener la frecuencia relativa acumulada se divide la frecuencia absoluta acumulada entre el número total de elementos (*N*=30). Esto da el tanto por uno de elementos iguales o menores a los elementos que se estudia. Las frecuencias relativas acumuladas son las siguientes:

X_i	Frecuencia absoluta (n_i)	Frecuencia relativa ($f_i = n_i/N$)	Frecuencia relativa acumulada ($F_i=N_i/N$)	Frecuencia relativa acumulada (Fi=Ni/N) en %
3	2	0,07	0,07	7%
4	4	0,13	0,20	20%
5	6	0,20	0,40	40%
6	7	0,23	0,63	63%
7	5	0,17	0,80	80%
8	3	0,10	0,90	90%
9	2	0,07	0,97	97%
10	1	0,03	1,00	100%
Total	**30**	**1**	**1**	**100%**

$\mathbf{F_1(3)}=f_1(3)=\mathbf{0,07}$
$\mathbf{F_2(4)}=f_1(3)+f_2(4)=0,07+0,13=\mathbf{0,20}$
$\mathbf{F_3(5)}=f_1(3)+f_2(4)+f_3(5)=0,07+0,13+0,20=\mathbf{0,40}$
$\mathbf{F_4(6)}=f_1(3)+f_2(4)+f_3(5)+f_4(6)=0,07+0,13+0,20+0,23=\mathbf{0,63}$
$\mathbf{F_5(7)}=f_1(3)+f_2(4)+f_3(5)+f_4(6)+f_5(7)=0,07+0,13+0,20+0,23+0,17=\mathbf{0,80}$
$\mathbf{F_6(8)}=f_1(3)+f_2(4)+f_3(5)+f_4(6)+f_5(7)+f_6(8)=0,07+0,13+0,20+0,23+0,17+0,10=\mathbf{0,90}$
$\mathbf{F_7(9)}=f_1(3)+f_2(4)+f_3(5)+f_4(6)+f_5(7)+f_6(8)+f_7(9) =0,07+0,13+0,20+0,23+0,17+0,10+0,07=\mathbf{0,97}$
$\mathbf{F_8(10)}=f_1(3)+f_2(4)+f_3(5)+f_4(6)+f_5(7)+f_6(8)+f_7(9)+f_8(10)=0,07+0,13+0,20+0,23+0,17+0,10+0,07+0,03= \mathbf{1,00}$

Se pueden calcular las frecuencias relativas acumuladas en porcentaje (%) multiplicándolas por 100.

5) Tabla de frecuencias

Una vez se han calculado todas las frecuencias, se construye la tabla de frecuencias. La tabla es la siguiente:

X_i	Frecuencia absoluta (n_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa (fi = ni/N)	Frecuencia relativa acumulada (Fi=Ni/N)
3	2	2	0,07	0,07
4	4	6	0,13	0,20
5	6	12	0,20	0,40
6	7	19	0,23	0,63
7	5	24	0,17	0,80
8	3	27	0,10	0,90
9	2	29	0,07	0,97
10	1	30	0,03	1,00
Total	30	30	1	1

Adicionalmente, se pueden incluir dos columnas con los **porcentajes** de las frecuencias relativas y frecuencias relativas acumuladas. Se obtiene la siguiente tabla:

X_i	Frecuencia absoluta (n_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa (fi = ni/N)	Frecuencia relativa acumulada (Fi=Ni/N)	Frecuencia relativa (fi = ni/N) en %	Frecuencia relativa acumulada (Fi=Ni/N) en %
3	2	2	0,07	0,07	7%	7%
4	4	6	0,13	0,20	13%	20%
5	6	12	0,20	0,40	20%	40%
6	7	19	0,23	0,63	23%	63%
7	5	24	0,17	0,80	17%	80%
8	3	27	0,10	0,90	10%	90%
9	2	29	0,07	0,97	7%	97%
10	1	30	0,03	1,00	3%	100%
Total	30	30	1	1	100%	100%

Objetivo 2.2 Evaluar nuestra distribución de datos y aprender a crear diversos tipos de gráficas a partir de la misma.

Los diversos tipos de gráficas, pueden proveer instantáneamente información sobre la distribución de un conjunto de datos. Los histogramas pueden ayudarle a observar:

- Si los datos se agrupan en torno a un valor individual o si los datos tienen múltiples picos o modas.
- Si los datos están diseminados con poca densidad en un rango amplio o si los datos se encuentran dentro de un rango pequeño.
- Si los datos son asimétricos o simétricos.

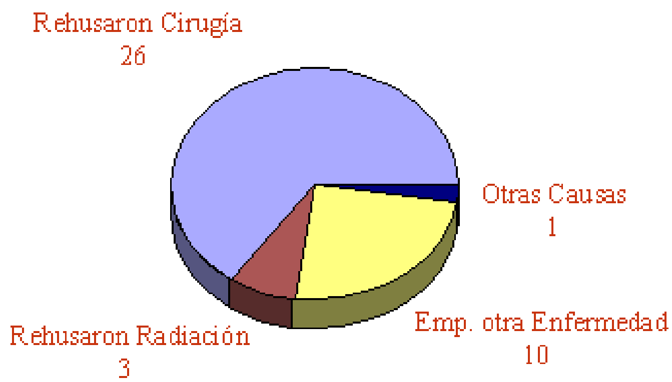
1. Diagrama de sectores:

Esta representación gráfica consiste en dividir un círculo en tantos sectores circulares como modalidades presente el carácter cualitativo, asignando un ángulo central a cada sector circular proporcional a la frecuencia absoluta n_i, consiguiendo de esta manera un sector con área proporcional también a n_i.

EJEMPLO: Los ángulos que corresponden a las cuatro modalidades de la tabla adjunta serán:

	Número de casos	**Ángulo(grados)**
Rehusaron cirugía	26	234°
Rehusaron radiación	3	27°
Empeoraron por una enfermedad ajena al cáncer	10	90°
Otras causas	1	9°

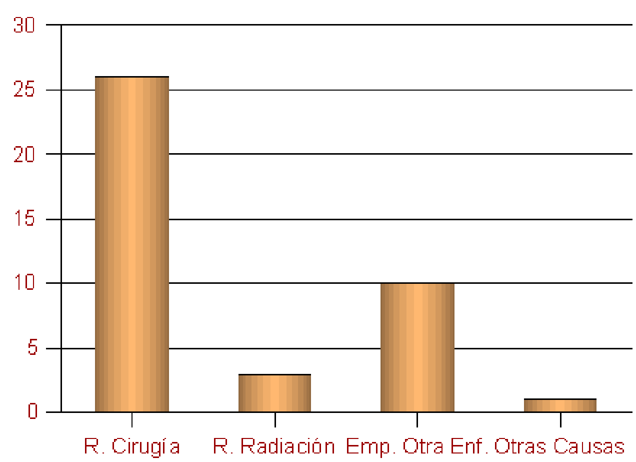
Y su representación en un diagrama de sectores será:



2. Diagrama de rectángulos:

Esta representación gráfica consiste en construir tantos rectángulos como modalidades presente el carácter cualitativo en estudio, todo ellos con base de igual amplitud. La altura se toma igual a la frecuencia absolua o relativa (según la distribución de frecuencias que estemos representando), consiguiendo de esta manera rectángulos con áreas proporcionales a las frecuencias que se quieren representar.

EJEMPLO : La representación gráfica de la distribución de frecuencias absolutas del ejemplo anterior será de la forma:



3. Histograma:

Al ser esta [redacted] una representación por áreas, hay que distinguir si los intervalos en los que aparecen agrupados los datos son de igual amplitud o no. Si la amplitud de los intervalos es constante, dicha amplitud puede tomarse como unidad y al ser

Frecuencia (área) = amplitud del intervalo · altura

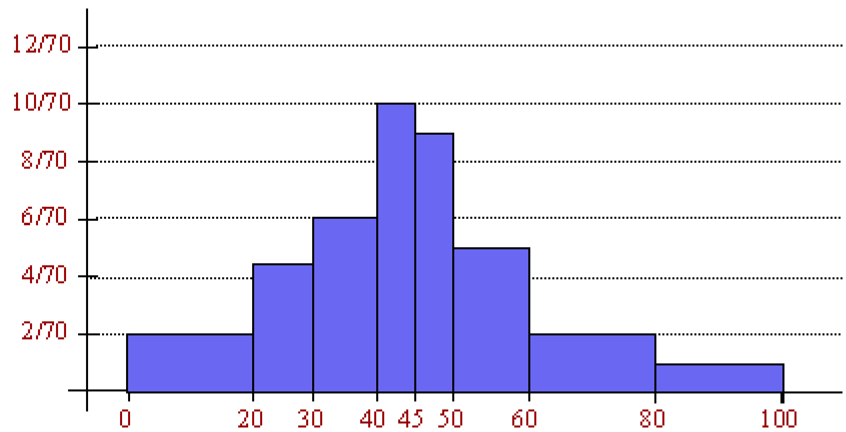
La altura correspondiente a cada intervalo puede tomarse igual a la frecuencia.

Si los intervalos tienen diferente amplitud, se toma alguna de ellas como unidad (generalmente la menor) y se levantan alturas para cada intervalo de forma que la ecuación anterior se cumpla.

EJEMPLO: Si tuviéramos una distribución de frecuencias como la siguiente, correspondiente a puntuaciones obtenidas en un test psicológico y en la que los intervalos son de diferente amplitud

l_i	n_i	f_i
0-20	8	8/70
20-30	9	9/70
30-40	12	12/70
40-45	10	10/70
45-50	9	9/70
50-60	10	10/70
60-80	8	8/70
80-100	4	4/70
	$\sum n_i = 70$	$\sum f_i = 1$

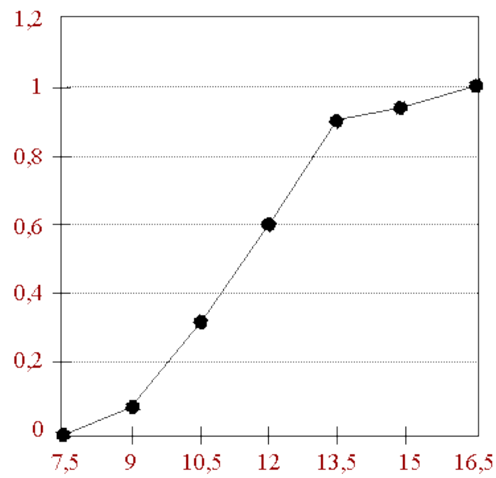
Tomando la amplitud 5 como unidad, deberemos levantar para el primer intervalo una altura de 2/70 para que el área sea la frecuencia relativa 8/70. Procediendo de la misma manera con el resto de los intervalos obtendríamos como representación gráfica la figura siguiente:



Obsérvese que la suma de todas las áreas debe ser 1, tanto si los intervalos de la distribución de frecuencias relativas son o no de igual amplitud.

4. Polígono de frecuencias acumuladas:

Se utiliza para representar distribuciones de frecuencias (relativas o absolutas) acumuladas. Consiste en representar la gráfica de una función que una por segmentos las alturas correspondientes a los extremos superiores de cada intervalo, tengan o no todos igual amplitud, siendo dicha altura igual a la frecuencia acumulada, dando una altura cero al extremo inferior del primer intervalo y siendo constante a partir del extremo superior del último.



EJERCICIOS:

Sea la variable “Religión que se profesa”

0: Católica
1: Protestante
2: Otra
3: Ninguna

Se han obtenido datos para una muestra de 50 personas, los resultados fueron los siguientes:

2,0,0,3,0,2,0,1,0,3,0,1,0,3,0,3,0,1,0,3,0,3,0,1,1,3,2,1,2,2,2,1,2,2,1,2,3, 3,1,3,3,3,2,3,3,3,3,1,3,3.

Realizar la tabla de frecuencia de datos y una gráfica representativa a los resultados.

SOLUCIÓN:

X_i	Frecuencia absoluta n_i	Frecuencia absoluta acumulada N_i	Frecuencia relativa $(f_i = n_i/N)$	Frecuencia relativa acumulada $(F_i = N_i/N)$	Frecuencia relativa en %	Frecuencia relativa acumulada en %
0	12	12	0.24	0.24	24%	24%
1	10	22	0.20	0.44	20%	44%
2	10	32	0.20	0.64	20%	64%
3	18	50	0.36	1.00	36%	100%
	50		1.00		100%	

Objetivo 2.3 Explicar qué son las medidas de tendencia central y el uso de cada una de ellas.

Las medidas de tendencia central ayudan a conocer de forma aproximada, el comportamiento de una distribución estadística.

- **Media aritmética.**

La media x (también llamada promedio o media aritmética) de un conjunto de datos ($X_1, X_2, \ldots, X_N$) es una medida de posición central. La definimos como el valor característico de la serie de datos resultado de la suma de todas las observaciones dividido por el número total de datos.

$$Media(X) = \overline{x} = \frac{\sum_{i=1}^{N} X_i}{N}$$

siendo $(X_1, X_2, ..., X_N)$ el conjunto de observaciones

Es decir:

$$Media(X) = \overline{x} = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

EJEMPLO:

Tenemos las edades de los once jugadores de un equipo de fútbol y queremos calcular su media.

29	31	20	19	26	25	26	30	18	23	26

Para ello, sumamos todas las edades y las dividimos por el número total de elementos, o sea once jugadores. **La media es de 24.82**.

$$Media = \frac{29 + 31 + 20 + \ldots + 26}{11} = \frac{273}{11} = 24,82$$

- **Mediana**

La mediana (Me(X)), o mediana estadística, es el elemento de un conjunto de datos ordenados (X1,X2,…,XN) que deja a izquierda y derecha la mitad de valores. Sea (X1,X2,…,XN) un conjunto de datos ordenado. El cálculo de la mediana depende de si el número de elementos N es par o impar.

Si N es impar, la mediana es el valor que está al medio, es decir:

$$\boxed{Mediana(X) = X_{\frac{N+1}{2}}}$$

Si N es par, la mediana es la media de los dos valores del centro, N/2 y N/2+1:

$$\boxed{Mediana(X) = Media(X_{\frac{N}{2}}, X_{\frac{N}{2}+1}) = \frac{X_{\frac{N}{2}} + X_{\frac{N}{2}+1}}{2}}$$

EJEMPLO:

Suponemos que tenemos una muestra con las edades de los once jugadores de un equipo de fútbol.

29	31	20	19	26	25	26	30	18	23	26

Para calcular la mediana necesitaríamos ordenar los elementos de menor a mayor y ver cuál es el elemento que deja a izquierda y derecha el mismo número de elementos.

18	19	20	23	25	**26**	26	26	29	30	31

Como el número de elementos del conjunto es impar, la mediana es el sujeto número 6, que se encuentra en el medio del conjunto. Por lo tanto la **Mediana(X)=26.**

- **Moda**

La moda (Mo(X)), o moda estadística, es el valor más repetido del conjunto de datos, es decir, el valor cuya frecuencia relativa es mayor. En un conjunto puede haber más de una moda.

EJEMPLO:

Tenemos una muestra de las once edades de los jugadores de un equipo de fútbol.

29	31	20	19	26	25	26	30	18	23	26

Hacemos recuento del elemento que más se repite en el conjunto.

29	31	20	19	**26**	25	**26**	30	18	23	**26**

La edad que más se repite es 26, por lo que la **moda del conjunto es 26**.

3. GRÁFICAS DE DISPERSIÓN

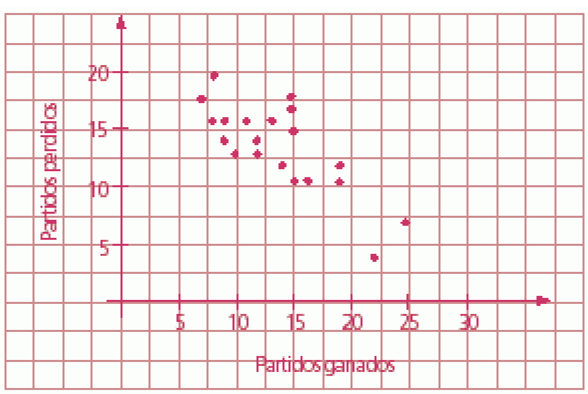
Las gráficas de dispersión nos muestran ciertas relaciones y patrones en los datos de estudio. Estas gráficas las representamos en un diagrama matemático mediante coordenadas cartesianas que generan puntos dispersos alrededor del plano y es así como con ayuda de la estadística, podemos valernos de esta información y obtener datos futuros. En los diversos tipos de investigación, frecuentemente se miden numerosas variables, muchas veces interesa determinar si existe relación entre algunas de esas variables, o predecir el valor de una de ellas conociendo el valor de otras; y es ahí donde nos valemos de las asociaciones lineales.

Objetivo 3.1 Aprender a crear gráficas de dispersión a partir del estudio de los datos.

Un diagrama de dispersión consiste en la representación gráfica de dos variables para un conjunto de datos. Analiza la relación entre dos variables, conociendo qué tanto se afectan entre sí o qué tan independientes son una de la otra.

En este sentido, ambas variables se representan como un punto en el plano cartesiano y de acuerdo a la relación que exista entre ellas, definimos su tipo de correlación.

En estas distribuciones se manejan **variables estadísticas bidimensionales**, que constituyen pares de valores de cada una de las variables elementales que intervienen, denotados por (x_i, y_j).



Ejemplo de un gráfico de dispersión (nube de puntos), que relaciona dos variables de una distribución bidimensional

¿Cómo hacer un diagrama de dispersión paso a paso?

- **Paso 1:** Determinar cuál es la situación. Si no entendemos qué es lo que está ocurriendo, no podremos establecer las variables a estudiar.

- **Paso 2:** Determina las variables a estudiar. Si ya determinaste las variables a estudiar, es porque crees que puede existir una relación entre ellas que te permita caracterizar la situación.

- **Paso 3:** Recolecta los datos de las variables: Si ya los tienes, perfecto. Si no, definimos un período de tiempo para conseguir los datos de las variables antes definidas. Recuerda que los datos de las dos variables deben estar dados en el mismo período de tiempo.

- **Paso 4:** Ubica los valores en el eje respectivo. Por lo general, la variable independiente es aquella que no está influenciada por la otra y se ubica en el *eje x*. La variable dependiente que es la que se ve afectada por la otra variable se ubica en el eje y. Así pues, procedemos a ubicar los valores en el plano cartesiano de acuerdo a su variable (x, y)

- **Paso 5:** Determina el coeficiente de correlación: El coeficiente de correlación debe verse reflejado en la forma que toma el gráfico de dispersión. Es el cociente de la covarianza y la multiplicación de la desviación típica de las dos variables.

- **Paso 6:** Analizamos: Con base en el coeficiente y en el gráfico, definimos cuál es la relación de las dos variables y tomamos las decisiones pertinentes.

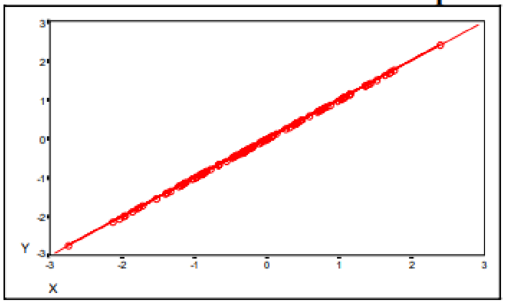
Objetivo 3.2 Generar asociaciones lineales a partir del coeficiente de relación y aprender a interpretar visualmente una recta de tendencia a partir de la imagen del diagrama de dispersión.

Con base en el comportamiento que toman las variables de estudio, podemos encontrar 3 tipos de correlación: Positiva, negativa y nula.

- ### Correlación directa (Covarianza +)

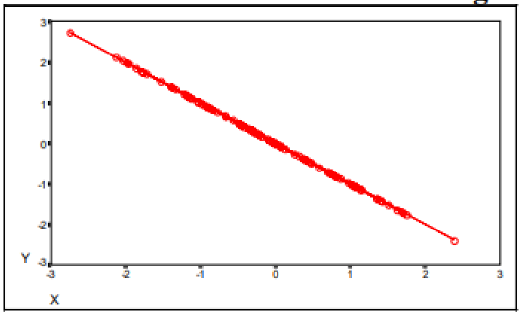
Se presenta cuando una variable aumenta o disminuye y la otra también, respectivamente. Hay una relación proporcional. Por ejemplo para un vendedor de carros, si él vende más carros (variable 1), va a ganar más dinero (variable 2).

Gráfica 1. Relación lineal exacta positiva.



- ### Correlación inversa (Covarianza -)

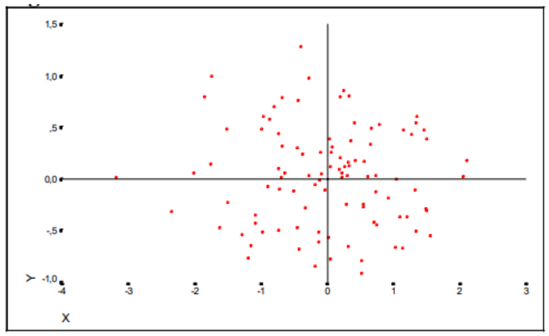
Gráfica 2. Relación lineal exacta negativa.



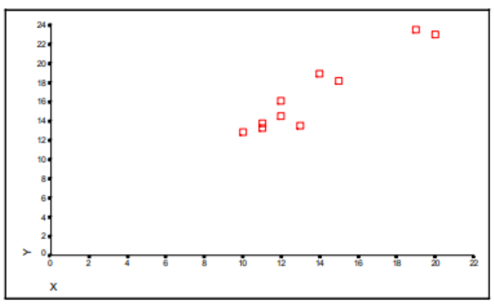
Se presenta cuando una variable se comporta de forma contraria o a la otra, es decir que si una variable aumenta, la otra disminuye. Hay una relación inversa proporcional. Por ejemplo para la construcción de un edificio, entre más trabajadores estén construyendo un edificio (variable 1), menos tiempo se necesitará para tenerlo listo (variable 2)

- ### Correlación nula

Si no encuentras un comportamiento entre las variables, existe una correlación nula.



EJEMPLO DE CÓMO DETERMINAR EL TIPO DE CORRELACIÓN USANDO LA FÓRMULA DE LA COVARIANZA



i	X_i	Y_i	X_iY_i
1	12	14,55	174,6
2	10	12,85	128,5
3	11	13,3	146,3
4	13	13,53	175,89
5	15	18,18	272,7
6	14	18,94	265,16
7	12	16,11	193,32
8	11	13,82	152,02
9	19	23,53	447,07
10	20	23,02	460,4
Suma	137	167,83	2415,96
Media	13,7	16,783	

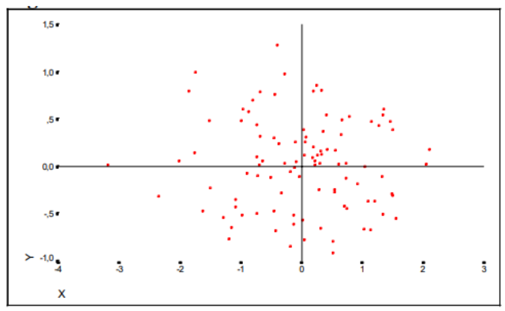
$$s_{XY}=\frac{\sum_{i=1}^{n}(X_i-\bar{x})(Y_i-\bar{y})}{n}=\frac{\sum_{i=1}^{n}X_iY_i}{n}-\overline{XY}=\frac{2415,96}{10}-13,7\cdot 16,783=11,67$$

Por tanto, existe ***asociación positiva*** entre ambas variables.

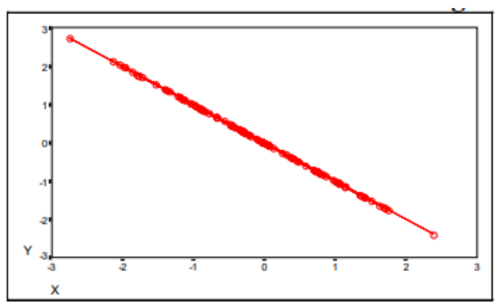
Otros tipos de clasificación de correlación

Otros tipos de clasificación están basados en qué tan **fuerte o débil** es el tipo de correlación, y se aprecian a simple vista en nuestro diagrama de dispersión.

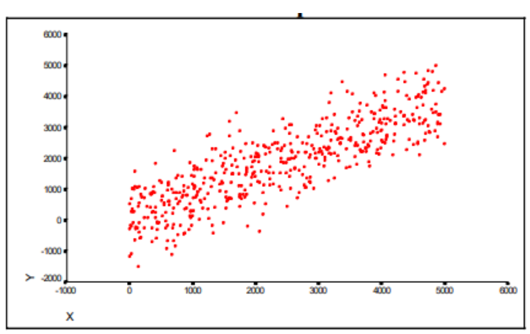
- **Sin correlación:** La misma correlación nula, los puntos están dispersos, formando una tipo circunferencia.



- **Correlación fuerte:** Los puntos de la nube están cercanos a la línea de regresión.



- **Correlación débil:** Los puntos de la nube no están muy cercanos a la línea de regresión.



El coeficiente de correlación en un diagrama de dispersión

El coeficiente de correlación nos describe cómo es la relación existente entre dos variables, en otras palabras, al conocer este número sabemos si la correlación es positiva o negativa y qué tan fuerte o débil es.

Este parámetro se define como el cociente entre la covarianza de la distribución y el producto de las desviaciones típicas de cada una de las variables.

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

Se usa la letra r para expresarla, veamos cómo:

- **r=1** La correlación es positiva perfecta. Si una variable crece, la otra también lo hace en una proporción constante. Es una relación directa, por eso si trazamos una línea de ajuste esta va pasar por todos y cada uno de los puntos.

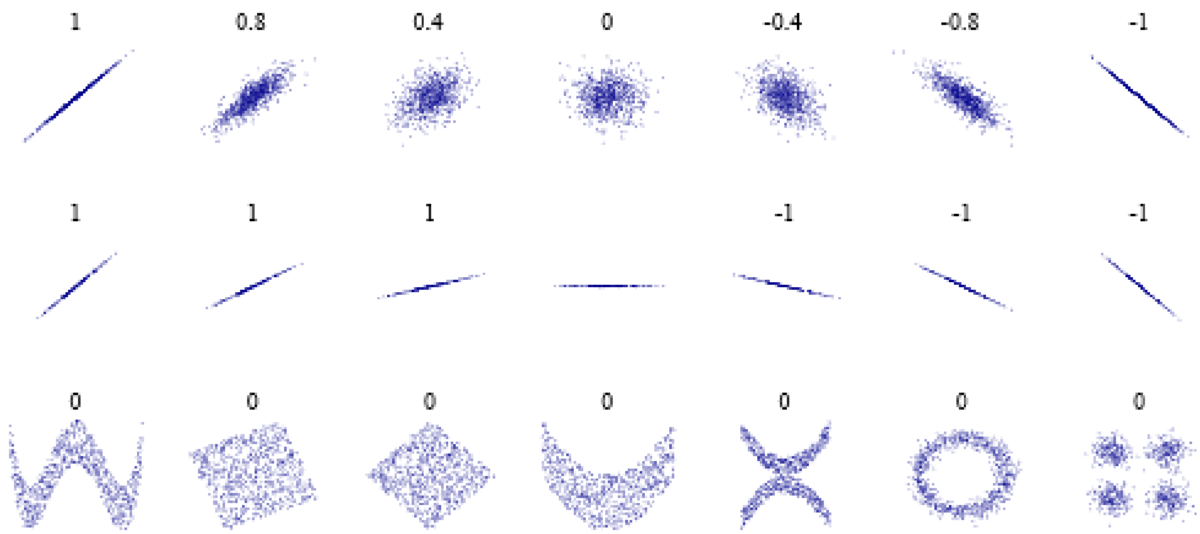
- **0<r<1** Es cuando r esta entre 0 y 1 sin llegar a ser 0 y 1. Es una correlación positiva. El grado de cercanía de 1 define qué tan directa y proporcional es la relación entre ambas variables, por ende entre más cerca esté de 0, más débil será su correlación negativa.

- **r=0** La correlación es nula, es decir que no existe una relación lineal entre ambas variables. Qué tal si pruebas buscando otro tipo de relación.

- **-1<r<0** Es cuando r esta entre -1 y 0 sin llegar a ser –1 y 0. Es una correlación negativa. El grado de cercanía a -1 define que tan inversa y proporcional es la relación entre ambas variables, por ende entre más cerca esté de 0, más debíl será su correlación negativa.

- **r=-1** La correlación es negativa perfecta. Si una variable crece, la otra va a disminuir en proporción constante. Es una relación directa e inversa, por lo tanto una línea de ajuste va a tocar todos los puntos graficados.

Un ejemplo más claro de todo lo mencionado se representa en la siguiente imagen:



Ejemplo de diagrama de dispersión

Vamos a ver desde una problemática empresarial, un ejemplo resuelto de diagrama de dispersión para el área de calidad.

Imagina que una imprenta está abriendo una nueva área de producción para la impresión de periódicos, y en este momento se encuentra haciendo todos los ensayos y pruebas para determinar la cantidad de tinta de cada color que deberían tener las máquinas.

Como prueba inicial, han decidido establecer la relación de errores de impresión según el grado de llenado de los recipientes de tinta de la máquina.

Bien, definida la situación, iniciamos desde el **paso 2**:
Las variables a estudiar para este ejemplo de gráfico de dispersión en calidad son:

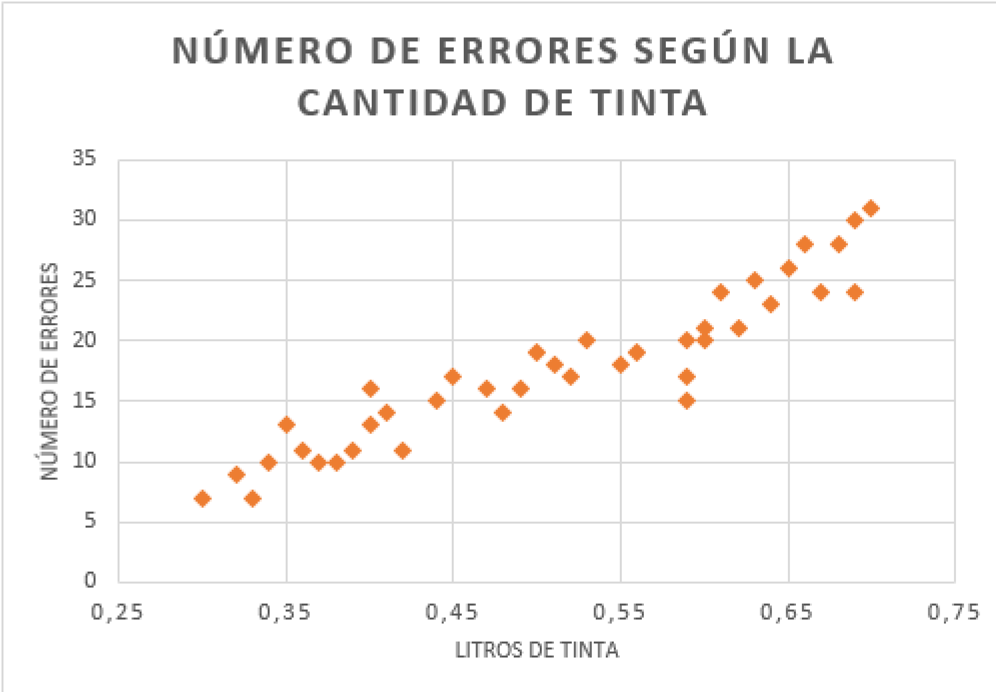
- Cantidad de tinta en litros
- Número de errores de impresión

Para el **paso 3**, comenzamos a recolectar las variables. En nuestro caso, el departamento de control de calidad hace 50 corridas o pruebas durante 5 días continuos.
Los resultados, en la tabla de la derecha.

Cantidad de tinta (Litros)	Número de errores
0,47	16
0,48	14
0,69	30
0,7	31
0,59	15
0,59	17
0,37	10
0,62	21
0,39	11
0,35	13
0,68	28
0,52	17
0,42	11
0,51	18
0,5	19
0,34	10
0,41	14
0,3	7
0,53	20
0,33	7
0,36	11
0,4	16
0,4	13
0,69	24
0,61	24
0,32	9
0,66	28
0,64	23
0,45	17
0,59	20
0,6	21
0,56	19
0,6	20
0,55	18
0,44	15
0,49	16
0,63	25
0,65	26
0,38	10
0,67	24

Para el **paso 4** ubicamos los ejes según las variables que tenemos. Al estar el número de errores influenciado por la cantidad de tinta, lo ubicamos como el eje y. Por consiguiente, el eje x es la cantidad de tinta. Ahora sí, hacemos el gráfico de dispersión.

Paso 5: Determinamos el coeficiente de correlación. Para nuestro ejemplo resuelto, obtenemos 0,94, ¿se ve esto reflejado en el gráfico? Por supuesto que sí, fíjate que los puntos están muy cerca unos de los otros, lo que indica que los valores se correlacionan fuertemente, es decir que la relación entre un aumento en los litros de tinta, impacta directamente en el número de errores en la impresión de posters. De hecho se hace evidente si miramos la tabla, no hay grandes saltos entre datos si miramos el número de errores.



Paso 6: Analizamos. Evidentemente hay una relación positiva fuerte entre la cantidad de tinta con la que se carga el tubo de la máquina y el número de errores generados en la impresión de los posters. Un paso siguiente para un problema de este tipo, sería buscar la forma de aprovechar la capacidad restante de la máquina, por ejemplo usar más tubos y más pequeños.

4. DISPERSIÓN DE DISTRIBUCIONES

Las medidas de dispersión o variación, nos indican si los datos distribuidos están próximos entre sí o sí están dispersos; es decir, nos muestran qué tan esparcidos se halla los datos. Además, estas medidas de dispersión nos permiten apreciar la distancia de nuestros datos a un valor central, que nos ayudará a saber qué tan dispersas están dos o más distribuciones.

4.1 Explicar qué son las medidas de dispersión, tales como el rango, la varianza y la desviación estándar, así como las fórmulas para su obtención.

Las medidas de tendencia central ofrecen una idea aproximada del comportamiento de una serie estadística. No obstante, no resultan suficientes para expresar sus características: una misma media puede provenir de valores cercanos a la misma o resultar de la confluencia de datos estadísticos enormemente dispares. Para conocer en qué grado las medidas de tendencia central son representativas de la serie, se han de complementar con medidas de dispersión como la varianza o la desviación típica.

Las medidas de centralización ayudan a determinar el «centro de gravedad» de una distribución estadística. Para describir el comportamiento general de la serie se necesita, sin embargo, una información complementaria para saber si los datos están dispersos o agrupados.

Así, las **medidas de dispersión** pueden definirse como los valores numéricos cuyo objeto es analizar el grado de separación de los valores de una serie estadística con respecto a las **medidas de tendencia central** consideradas.

Rango

La medida de dispersión más inmediata es el **recorrido** de la distribución estadística, también llamado **rango** o **amplitud**. Dada una serie de valores x_1, x_2, ..., x_n, su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores:

$$Re = x_i\ (\text{máx}) - x_i\ (\text{mín}), \text{ siendo } i = 1, 2, \ldots, n.$$

Desviación media.

Como medida de dispersión más frecuentemente utilizada, la **desviación media** se define como la **media aritmética** de los valores absolutos de la desviación de cada valor de la variable con respecto a la media. Su formulación matemática es la siguiente:

$$DM = \frac{f_1|x_1 - \overline{x}| + f_2|x_2 - \overline{x}| + \ldots + f_n|x_n - \overline{x}|}{f_1 + f_2 + \ldots + f_n} = $$
$$= \frac{\sum f_i |x_i - \overline{x}|}{\sum f_i} \text{ con } i = 1, 2, 3, \ldots, n.$$

Varianza y desviación típica

La desviación media no siempre suministra una idea clara del grado de separación entre los valores de una variable estadística. Para estudios científicos, se prefiere utilizar una pareja de parámetros relacionados que se conocen como **varianza** y **desviación típica**.

La varianza se define como el cociente entre la suma de los cuadrados de las desviaciones de los valores de la variable y el número de datos del estudio. Matemáticamente, se expresa como:

$$V = \frac{f_1(x_1 - \overline{x})^2 + f_2(x_2 - \overline{x})^2 + \ldots + f_n(x_n - \overline{x})^2}{f_1 + f_2 + \ldots + f_n} = $$
$$= \frac{\sum f_i (x_i - \overline{x})^2}{\sum f_i} \text{ con } i = 1, 2, 3, \ldots, n.$$

Por su parte, la desviación típica, simbolizada por , se define sencillamente como la raíz cuadrada de la varianza:

$$\sigma = \sqrt{V}$$

Por lo tanto, se tiene que:

$$\sigma = \sqrt{\frac{\sum f_i(x_i - \overline{x})^2}{\sum f_i}}$$

La varianza y la desviación típica, cada una con su respectivo valor, se usan indistintamente en los estudios estadísticos.

4.2 Explicar qué es una regresión lineal simple y la finalidad de la misma en problemas sencillos