|  |  |
| --- | --- |
| Título del guion | La estadística y la probabilidad |
| Código del guion | MA\_10\_06\_CO |
| Descripción | Los problemas, herramientas y métodos de la estadística permiten tomar decisiones ante situaciones de incertidumbre, de manera que sirve de ayuda a cada individuo para producir y comunicar una mejor comprensión de la situación en un contexto particular  El azar es imprevisible, pero también produce regularidades que la probabilidad cuantifica y mide. |

[SECCIÓN 1] **1 La estadística descripctiva**

La estadística descriptiva es una rama de la estadística que se dedica a recolectar, sistematizar, ordenar, describir y analizar algunas características de una población, basándose en un conjunto de datos. Algunos ejemplos de esta rama son cuando se diseñan, aplican y analizan encuestas de preferencia de votos o cuando una empresa hace un estudio de los gustos y las necesidades para el lanzamiento de un nuevo producto. Por otro lado la estadística inferencial busca construir conclusiones generales para toda la población a partir del estudio de una muestra.

En ambos casos la estadística permite la construcción de parámetros que proporcionen una idea global cuantitativa de la población observada y, a partir de estos, se puede establecer conclusiones y tomar decisiones ante situaciones de incertidumbre.

La agrupación de datos es una herramienta de gran utilidad cuando se tiene una gran cantidad de datos. El proceso se denomina *distribución de frecuencias* dado que no solo se distribuyen los datos, sino a que tal distribución se realiza en subgrupos o clases o intervalos. Así, el total de datos aparece distribuido según sus repeticiones o *frecuencias* en tales subgrupos.

[SECCIÓN 2] **1.1 Las variables cualitativas**

Una vez se tiene un conjunto de datos éstos pueden ser sistematizados de diferentes formas, de acuerdo a sus características. Las *variables* son las diferentes características o aspectos que podemos estudiar en los individuos de una población y son necesarias para analizarla y tener alternativas de decisión.

Existen diferentes tipos de variables, las cuales pueden ser cualitativas o cuantitativas. Como su nombre lo indica, las variables cualitativas indican características no medibles o numerables de los individuos en una población, así que toman valores no numéricos. Ejemplo de tales variables son los colores del iris de los pobladores de una región, el género de las películas preferidas, el sector de residencia o el tema principal en los libros preferidos.

[SECCIÓN 2] **1.2 Las variables cuantitativas discretas**

Por su parte las variables cuantitativas recopilan atributos medibles de los individuos en una población, como pueden ser la edad, la estatura, la cantidad de hermanos, el tiempo de estudio o participación en redes sociales semanal, etc. En general, las variables cuantitativas toman valores numéricos y, debido a eso, se subclasifican en dos: discretas y continuas.

Las variables cuantitativas discretas son aquellas que toman valores del conjunto de los números naturales, como por ejemplo la cantidad de celulares que tenemos, la cantidad de ejemplares de nuestra colección favorita o la cantidad de hermanos.

[SECCIÓN 2] **1.3 Las variables cuantitativas continuas**

Entre las variables cuantitativas, se denominan variables cuantitativas continuas a aquellas que pueden tomar valores del conjunto de los números reales, es decir que se admiten todos los valores en un intervalo. Ejemplo de ellas son el peso de una persona, la estatura, el tiempo que dedica a actividades deportivas, etc.

Si los elementos objeto del estudio estadístico toman muchos valores diferentes, se agrupan formando intervalos de clase. Cada clase está delimitada por un valor inferior y otro superior, los cuales son llamados límites de la clase. Por ejemplo, si tenemos valores de un grupo de 237 datos que están entre 1 y 100, se pueden definir grupos para sintetizar la información, por medio de los intervalos: 1-25, 26-50, 51-75, 76-100. De esta forma, resumimos los datos en cuatro grupos, entre los que se distribuirán los 237 datos. En el ejemplo, los límites inferiores de las clases fueron 1, 26, 51 y 76, mientras que los límites superiores fueron 25, 50, 75 y 100.

Por su parte, la *marca de clase* es el punto medio de cada intervalo y es el valor que se utiliza para el cálculo de parámetros en contextos de variables cuantitativas, por lo cual será el valor que representará todo el intervalo de clase. Las marcas de clase de las 4 clases del ejemplo son 13, 38, 63 y 88.

Entre las formas de condensación de la información estadística, una de las más potentes y útiles para presentar información clara y concisa es la distribución de frecuencias. La tabla de frecuencias está formada por columnas en las cuales se pueden incluir intervalos de clase, frecuencias absolutas, frecuencias relativas, frecuencias absolutas acumuladas, frecuencias relativas acumuladas, marcas de clase y porcentajes. La aparición de tales columnas depende principalmente del tipo de variable estudiada.

Específicamente, los elementos de la tabla de frecuencias son:

* Los *intervalos de clase* muestran las agrupaciones o clases en que se van a distribuir los datos. El número *i* indica la cantidad total de clases, mientras que el número *n* indica la cantidad total de datos.
* La *frecuencia absoluta* (*fi*) es el número total de veces que se repiten los datos en cada clase.
* La *frecuencia relativa* (*hi*) es el cociente entre la frecuencia absoluta (*f*) en cada intervalo y el total de datos (*n*).
* La *frecuencia absoluta acumulada* (*F*) en cada clase es la suma de las frecuencias absolutas (*f*) de los valores menores o iguales a él.
* La *frecuencia relativa* acumulada (*H*) es el cociente entre la frecuencia relativa (*hi*) en cada intervalo y el total de datos (*n*).
* La *marca de clase* (*ci*) es el punto medio de cada intervalo y .solo tiene sentido para variables estadísticas cuantitativas

Por ejemplo, para obtener información acerca de los hábitos alimenticios de los estudiantes, se preguntó a cada uno de ellos por el tipo de alimento que consumía principalmente cada día. Por supuesto, se trata de una variable de tipo cualitativo.

Tomando las opciones de grupos nutricionales definidos por las sociedades nutricionales, las opciones para elegir el principal alimento consumido eran:

* Carne, pollo, huevos y frutos secos
* Pan, cereales, arroz y pasta
* Frutas y verduras
* Aceites, grasas y dulces
* Suplementos alimentarios
* Leche, yogurt y queso

Los resultados para los 36 estudiantes consultados fueron los siguientes:

|  |
| --- |
| Carne, pollo, huevos y frutos secos |
| Aceites, grasas y dulces |
| Pan, cereales, arroz y pasta |
| Pan, cereales, arroz y pasta |
| Frutas y verduras |
| Pan, cereales, arroz y pasta |
| Frutas y verduras |
| Carne, pollo, huevos y frutos secos |
| Leche, yogurt y queso |
| Carne, pollo, huevos y frutos secos |
| Frutas y verduras |
| Carne, pollo, huevos y frutos secos |
| Carne, pollo, huevos y frutos secos |
| Leche, yogurt y queso |
| Pan, cereales, arroz y pasta |
| Pan, cereales, arroz y pasta |
| Suplementos |
| Pan, cereales, arroz y pasta |
| Pan, cereales, arroz y pasta |
| Frutas y verduras |
| Aceites, grasas y dulces |
| Leche, yogurt y queso |
| Carne, pollo, huevos y frutos secos |
| Carne, pollo, huevos y frutos secos |
| Frutas y verduras |
| Suplementos |
| Pan, cereales, arroz y pasta |
| Aceites, grasas y dulces |
| Leche, yogurt y queso |
| Pan, cereales, arroz y pasta |
| Carne, pollo, huevos y frutos secos |
| Leche, yogurt y queso |
| Suplementos |
| Leche, yogurt y queso |
| Leche, yogurt y queso |
| Frutas y verduras |

Como se observa, el registro de las respuestas no permite de manera rápida ni precisa identificar el comportamiento nutricional de la población estudiada. Sin embargo, al organizar los datos en su tabla de frecuencias se obtiene:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Principal alimento consumido**  ***xi*** | **Frecuencia absoluta *(fi)*** | **Frecuencia relativa**  ***(hi= fi/n)*** | **Frecuencia absoluta acumulada *(Fi)*** | **Frecuencia relativa acumulada *(Hi= Fi/n )*** | **Porcentaje (%)**  **(*hi* escrita como porcentaje)** |
| Carne, pollo, huevos y frutos secos | 8 | 2/9 | 8 | 2/9 | 22.22% |
| Pan, cereales, arroz y pasta | 9 | 1/4 | 17 | 17/36 | 25.00% |
| Frutas y verduras | 6 | 1/6 | 23 | 23/36 | 16.67% |
| Aceites, grasas y dulces | 3 | 1/12 | 26 | 13/18 | 8.33% |
| Suplementos alimentarios | 3 | 1/12 | 29 | 29/36 | 8.33% |
| Leche, yogurt y queso | 7 | 7/36 | 36 | 1 | 19.44% |
| **TOTAL** | **36** | **1** | **1** |  | **100.00%** |

De esta manera se identifica por ejemplo que la cuarta parte de los estudiantes consume principalmente carbohidratos, y que la misma cantidad de estudiantes consume principalmente grasas o suplementos.

Por su parte, como segundo ejemplo, en una clase de 25 alumnos se ha realizado una encuesta sobre la máxima distancia a la que cada estudiante ha viajado, medida entre su actual lugar de residencia y la capital de un país extranjero adonde han viajado. En este caso se trata de una variable de tipo cuantitativo, por lo que, en este caso, la tabla contiene marcas de clase. Los resultados fueron agrupados en cinco intervalos de clase según la cantidad de kilómetros recorrida.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cantidad de kilómetros recorrida**  ***xi*** | **Marca de clase**  ***(ci)*** | **Frecuencia absoluta *(fi)*** | **Frecuencia relativa**  ***(hi= fi/n)*** | **Frecuencia absoluta acumulada *(Fi)*** | **Frecuencia relativa acumulada *(Hi= Fi/n )*** | **Porcentaje (%)**  **(*hi* escrita como porcentaje)** |
| [0, 2.000) | 1.000 | 14 | 14 / 25 = 0,56 | 14 | 14 / 25 | 56 % |
| [2.000, 4.000) | 3.000 | 5 | 5 / 25 = 0,2 | 14 + 5 = 19 | 19 / 25 | 20 % |
| [4.000, 6.000) | 5.000 | 1 | 1 / 25 = 0,04 | 19 + 1 = 20 | 20 / 25 | 4 % |
| [6.000, 8.000) | 7.000 | 2 | 2 / 25 = 0,08 | 20 + 2 = 22 | 22/ 25 | 8 % |
| [8.000, 10.000) | 9.000 | 3 | 3 / 25 = 0,12 | 22 + 3 = 25 | 25 / 25 | 12 % |
| TOTAL |  | 25 |  | 25 | 25 / 25 = 1 | 100.00% |

Programas como Microsoft Excel permiten crear las tablas de distribución, de manera que con los datos se puedan generar a su vez diagramas que presenten de manera sucinta la información. Por ejemplo en [VER](https://youtu.be/bKK0kXzwpgs) se ejemplifica una manera de generar una tabla de distribución de frecuencias.

[SECCIÓN 2] **1.4 Consolidación**

[SECCIÓN 1] **2 Los muestreos**

Un estudio estadístico siempre está basado en el análisis de un conjunto de elementos seleccionados y no de sucesos aislados, éste conjunto es llamado *población*. Algunos ejemplos son la totalidad de votantes para unas elecciones, la totalidad de bombillos producidos por una compañía o la totalidad de animales de un zoológico.

Si la población tiene un número excesivo de elementos, es necesario seleccionar un subconjunto o *muestra*, seleccionada con cuidado para que sea representativa del total de la población. El proceso de selección se denomina *muestreo****.***

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Incluir una imagen de la población y otra de la muestra para un estudio estadístico |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | 4º ESO/Matemáticas/La estadística/La terminología estadística  [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14643/InfoGuion/cuadernoestudio/images_xml/MT_10_10_img1_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14643/InfoGuion/cuadernoestudio/images_xml/MT_10_10_img1_zoom.jpg) |
| **Pie de imagen** | A la izquierda, la población del estudio para saber cuántas bolsas defectuosas de frutos secos se fabrican al día y, a la derecha, la muestra con la que se trabajará. |
| **Ubicación del pie de imagen** | Inferior |

Este proceso lo hacen en Colombia las industrias del café, para conocer cuántas bolsas defectuosas se fabrican cada día. Como no se pueden abrir todas las bolsas, el departamento de control de calidad elige unas muestras de cada lote de producción.

En este ejemplo la población sería el conjunto de bolsas café fabricadas y la muestra correspondería al número de bolsas que se toman para analizar y poder encontrar los resultados del estudio de unidades defectuosos en la muestra, e inferir a partir de tal estudio la cantidad de unidades defectuosas en el conjunto de la población. El muestreo es el método que se sigue para hacer la selección de las bolsas que se examinarán, teniendo en cuenta que el subconjunto elegido (muestra) debe representar al total de los datos (población).

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Población, muestra y muestreo** |
| **Contenido** | * La población: es el conjunto de todos los elementos que son objeto de un estudio. * La muestra: es un subconjunto representativo de la población. * El muestreo es el método que se sigue para seleccionar las muestras. |

[SECCIÓN 2] **2.1 Los tipos de muestreo**

Si bien existen varias formas de seleccionar una muestra en el conjunto de una población, no cualquier selección garantiza la condición de ser representativa de la población total. Por ejemplo, en una fábrica de bombillos, hacer la selección de 100 unidades como muestra de entre 10000 unidades producidas tomando al azar cualesquiera 100 unidades, puede representar a la población total de bombillos. Sin embargo, si se hace lo mismo preguntando por la intención de voto a la Alcaldía de una ciudad, quizá sea importante seleccionar personas mayores de edad, de diferentes localidades o sectores de la ciudad, de diferentes estratos y niveles educativos, de diferentes géneros, etc., es decir que elegir cualesquiera 100 de entre cada 10.000 puede no ser buena idea.

Un elemento fundamental para que un muestreo sea representativo es que esté formado por un número razonable de elementos. Si se comete el error de hacer conclusiones muy generales a partir de la observación de sólo una parte de una determinada población, ese error se denomina *error de muestreo*. Por otra parte, si se comete el error de generalizar conclusiones hacia una población más grande a aquella en la que originalmente se hizo el muestreo tal error se denomina *error de inferencia*.

Un elemento fundamental para que un muestreo sea aleatorio es que la muestra sea escogida al azar, garantizando con ello la neutralidad del analista respecto a los individuos analizados. Todo muestreo aleatorio descansa sobre el principio de equiprobabilidad, según el cual todos los individuos que pertenezcan a la población tienen la misma probabilidad de ser elegidos para formar parte de una muestra, con lo que se garantiza el rigor del análisis posterior.

Si bien existen ejemplos de muestreos no aleatorios, como en el caso en que un maestro elige de entre todos los estudiantes de la institución solo aquellos que son sus estudiantes, los muestreos representativos en estadística descriptiva deben ser además muestreos aleatorios.

Los muestreos aleatorios más utilizados son simples, sistemáticos, estratificados, por conglomerados o polietápicos. En general, se trata de elegir de entre una población de tamaño *N*, una muestra de tamaño *n*.

**Muestreo simple**

El muestreo aleatorio simple es el más sencillo de los métodos de muestreo. A partir de un listado completo de toda la población, se realiza una asignación de un número a cada individuo de la población. Posteriormente se acude al algún medio de selección como sorteo, asignación de números aleatorios predeterminados en una lista o generados por computador para elegir la cantidad de individuos que indique el tamaño de muestra requerido. De este modo se completa el muestreo simple.

En resumen, el proceso de muestreo simple selecciona, en una población de tamaño *N* una muestra de tamaño *n*, de manera que la probabilidad de que un elemento de la población se incluya en la muestra es de *n/N*. Aunque simple, el método requiere que se conozca de antemano a toda la población para poder listarla y la selección de muestras pequeñas puede no ser representativa de la población. Puede ser útil en muestreos para estudios de calidad en objetos producto de cadenas de producción y en general en poblaciones homogéneas y tiene la ventaja de que permite hacer cálculos de medidas de variación y dispersión muy fácilmente.

Veamos un ejemplo: En el marco del proyecto institucional de ahorro y protección financiera de un colegio, 20 estudiantes de bachillerato de diferentes grados viajarán a Cartagena para representar a su colegio en el Foro Nacional de Educación. Uno de los cursos, de 30 estudiantes, deberá elegir aleatoriamente 2 representantes. Se tiene la lista de estudiantes y la respuesta ante la pregunta ¿Ahorra?

|  |  |
| --- | --- |
| Estudiante | ¿Ahorra? |
| Camilo | No |
| Tomás | No |
| Fausto | No |
| Valeria | No |
| Martín | No |
| Jimena | Sí |
| Julián | No |
| Andrés | Sí |
| Santiago | Sí |
| Luisa | No |
| Elizabeth | No |
| Juana | Sí |
| Liliana | No |
| Ana | Sí |
| Laura | No |
| Lucas | Sí |
| Alejandro | No |
| Sara | Sí |
| Mateo | No |
| Daniel | Sí |
| Laura | No |
| Natalia | Sí |
| Juan | Sí |
| Tatiana | Sí |
| Mauricio | No |
| Joaquín | Sí |
| Claudia | Sí |
| Marcela | No |
| Sebastián | No |
| Daniela | No |

Elige, mediante un muestreo aleatorio simple a los dos representantes de este curso. Para lograrlo, puedes hacer un sorteo poniendo los números del 1 al 30 en una bolsa y sacando dos de ellos, elegir aquellos dos estudiantes cuyos códigos hayan sido los números centrales en el sorteo del baloto de un día específico, generando números aleatorios y eligiendo a los dos con el número mayor, etc.

Lo primero que debe hacerse es enumerar a la población.

|  |  |  |
| --- | --- | --- |
| No. | Estudiante | ¿Ahorra? |
| 1 | Camilo | No |
| 2 | Tomás | No |
| 3 | Fausto | No |
| 4 | Valeria | No |
| 5 | Martín | No |
| 6 | Jimena | Sí |
| 7 | Julián | No |
| 8 | Andrés | Sí |
| 9 | Santiago | Sí |
| 10 | Luisa | No |
| 11 | Elizabeth | No |
| 12 | Juana | Sí |
| 13 | Liliana | No |
| 14 | Ana | Sí |
| 15 | Laura | No |
| 16 | Lucas | Sí |
| 17 | Alejandro | No |
| 18 | Sara | Sí |
| 19 | Mateo | No |
| 20 | Daniel | Sí |
| 21 | Laura | No |
| 22 | Natalia | Sí |
| 23 | Juan | Sí |
| 24 | Tatiana | Sí |
| 25 | Mauricio | No |
| 26 | Joaquín | Sí |
| 27 | Claudia | Sí |
| 28 | Marcela | No |
| 29 | Sebastián | No |
| 30 | Daniela | No |

Una primera forma de hacer la selección es generando números aleatorios del 1 al 30, usando la función de Excel “=aleatorio.entre(1;30)”. Los dos primeros números que aparezcan serán las posiciones en el listado de los estudiantes elegidos. El resultado es 7 y 23, así que Julián y Juan serán los representantes.

Para otra forma de hacer la selección, generaremos números aleatorios del 100 al 200, usando la función de Excel “=aleatorio.entre(100;200)”. Los dos estudiantes que tengan los números mayores serán los elegidos. Para lograrlo, frente al primero escribimos ““=aleatorio.entre(100;200)” y arrastramos para que frente a cada uno haya un número aleatorio.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Generación de números aleatorios usando Excel |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Asignación de números aleatorios a una población. |
| **Ubicación del pie de imagen** | Inferior |

Concluimos el muestreo eligiendo a los seleccionados. En este caso el listado final quedó:

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Estudiante | ¿Ahorra? | Aleatorio |
| 1 | Camilo | No | 162 |
| 2 | Tomás | No | 182 |
| 3 | Fausto | No | 140 |
| 4 | Valeria | No | 117 |
| 5 | Martín | No | 108 |
| 6 | Jimena | Sí | 189 |
| 7 | Julián | No | 102 |
| 8 | Andrés | Sí | 116 |
| 9 | Santiago | Sí | 186 |
| 10 | Luisa | No | 161 |
| 11 | Elizabeth | No | 158 |
| 12 | Juana | Sí | 150 |
| 13 | Liliana | No | 175 |
| 14 | Ana | Sí | 134 |
| 15 | Laura | No | 198 |
| 16 | Lucas | Sí | 178 |
| 17 | Alejandro | No | 200 |
| 18 | Sara | Sí | 160 |
| 19 | Mateo | No | 121 |
| 20 | Daniel | Sí | 130 |
| 21 | Laura | No | 173 |
| 22 | Natalia | Sí | 110 |
| 23 | Juan | Sí | 114 |
| 24 | Tatiana | Sí | 194 |
| 25 | Mauricio | No | 157 |
| 26 | Joaquín | Sí | 199 |
| 27 | Claudia | Sí | 163 |
| 28 | Marcela | No | 172 |
| 29 | Sebastián | No | 117 |
| 30 | Daniela | No | 136 |

Por lo que Alejandro y Joaquín serán los representantes, pues el criterio de selección fue elegir a los dos estudiantes que tengan los números mayores aleatorios. Aunque se obtiene un muestreo simple, un problema de este método de generación de números aleatorios y posterior ordenación es que en ocasiones se repiten los números, si bien se subsana fácilmente generándolos hasta que no suceda que hay más elementos en la muestra que los requeridos.

Otra forma, que suele ser la más empleada y correcta al generar números aleatorios para hacer el muestreo es generando números aleatorios de una cantidad grande de cifras y seleccionando las primeras que respondan a la numeración elegida. Por ejemplo se generan cinco números aleatorios entre 0 y 1’000.000 y se eligen siempre números de dos cifras hasta que tengan sentido en el listado propuesto. Para generarlos se usa la función “=aleatorio.entre(0;1000000)”, con lo que se obtiene:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 336565 | 611783 | 391602 | 654499 | 618010 |

Los números de dos cifras son 33, 65, 65, 61, 17, 83, 39, 16, 02, etc., pero como necesitamos elegir dos con sentido en nuestro listado, 17 y 16 son los primeros. Así, en este muestreo los representantes serían Alejandro y Lucas, que tienen numeración 17 y 16.

**Muestreo sistemático**

El muestreo aleatorio sistemático inicia del mismo modo que el muestreo aleatorio simple, enumerando a todos los miembros de la población. Luego se escoge un único elemento al azar y se completa la muestra eligiendo de manera periódica el resto. La diferencia radica en que, en lugar de extraer *n* números que corresponderán a la muestra, se extrae solo uno, que se llamará *i*. Se parte de ese número aleatorio *i*, que fue elegido al azar, y para completar la muestra se eligen los números *i*, *i+k*, *i+2k*, *i+3k*,..., *i+(n-1)k*, es decir se toman los individuos de *k* en *k*, donde *k* el resultado de dividir el tamaño de la población *N* entre el tamaño de la muestra *n*, con lo cual *k = N/n*. El número *k* se llama *constante de muestreo*.

El riesgo de este tipo de muestreo es que la constante de muestreo homogeniza de alguna manera a la población, y si la homogeneidad está asociada con el fenómeno de interés, las estimaciones obtenidas a partir de la muestra pueden contener sesgo de selección. Supongamos que estamos seleccionando una muestra sobre listas de 10 individuos en los que los 5 primeros son adultos y los 5 últimos son menores de edad. Al aplicar un muestreo aleatorio sistemático con constante de muestreo *k = 10* siempre se seleccionan, o sólo niños o sólo adultos, perdiendo con ello la característica de representatividad requerida en el muestreo. Por el contrario, si la población está ordenada siguiendo una tendencia conocida, este tipo de muestreo asegura una cobertura de unidades de todos los tipos.

En nuestro problema de elegir a los dos representantes del curso, el muestreo sistemático se haría de la siguiente manera: Primero, se elige al azar un número entre 1 y 30, por ejemplo usando Excel, que arroja el número 20, que llamaremos *i*. Así, el primer representante es Daniel, que tiene el número 20.

Luego, como *N = 30* y *n* = 2, entonces *k = N/n= 15*. Como *i+k=20+15=35*, pero nuestra lista llega a 30, iniciamos de nuevo hasta llegar a Martín, que tiene el número 5. Nótese que entre Daniel y Martín hay la misma cantidad de estudiantes:

|  |  |  |
| --- | --- | --- |
| No. | Estudiante | ¿Ahorra? |
| 1 | Camilo | No |
| 2 | Tomás | No |
| 3 | Fausto | No |
| 4 | Valeria | No |
| 5 | Martín | No |
| 6 | Jimena | Sí |
| 7 | Julián | No |
| 8 | Andrés | Sí |
| 9 | Santiago | Sí |
| 10 | Luisa | No |
| 11 | Elizabeth | No |
| 12 | Juana | Sí |
| 13 | Liliana | No |
| 14 | Ana | Sí |
| 15 | Laura | No |
| 16 | Lucas | Sí |
| 17 | Alejandro | No |
| 18 | Sara | Sí |
| 19 | Mateo | No |
| 20 | Daniel | Sí |
| 21 | Laura | No |
| 22 | Natalia | Sí |
| 23 | Juan | Sí |
| 24 | Tatiana | Sí |
| 25 | Mauricio | No |
| 26 | Joaquín | Sí |
| 27 | Claudia | Sí |
| 28 | Marcela | No |
| 29 | Sebastián | No |
| 30 | Daniela | No |

En general, en el muestreo aleatorio sistemático la muestra queda homogéneamente distribuida en el conjunto de la población.

**Muestreo estratificado**

¿Notaste que hasta ahora ninguna niña ha sido elegida para representar al colegio y que no se ha usado el dato que parece relevante respecto a una práctica del ahorro? Por este tipo de circunstancia, el muestreo aleatorio estratificado inicia considerando categorías diferentes entre sí, que sean homogéneas respecto a alguna característica particular con el fin de garantizar la diversidad de la población en la muestra.

En el muestreo aleatorio estratificado cada categoría o *estrato* de representación funciona de manera independiente. En cada uno de los estratos puede aplicarse un muestreo simple o sistemático para elegir los elementos de la muestra.

El muestreo aleatorio estratificado inicia reconociendo las características homogéneas que garanticen representatividad en la muestra según el tipo de estudio poblacional a realizar. Aunque en Colombia la palabra “estrato” se usa frecuentemente para notar el nivel de calidad de la zona en que se ubica una vivienda, se puede estratificar también según la profesión, el municipio de residencia, el género, el estado civil, etc. Sin embargo, como no todos los estratos tienen la misma cantidad de individuos, debe detallarse la distribución de la muestra en los diferentes estratos; tal distribución se conoce como *afijación*.

Existen varios tipos de afijación, entre los que se encuentran la afijación simple, la afijación proporcional, la afijación de mínima varianza y la afijación óptima. En la afijación simple, a cada estrato le corresponde igual número de elementos muestrales; es óptima si los estratos están homogéneamente distribuidos en la población, pero si no es así tal afijación favorece a los estratos con menor cantidad de individuos y perjudica a los que tienen mayor cantidad de individuos. En la afijación proporcional, la distribución de elementos en la muestra es directamente proporcional al tamaño de la población en cada estrato. La afijación de mínima varianza y la afijación óptima requieren cálculos previos de desviación y varianza, por lo que no las especificaremos aquí.

En el ejemplo de seleccionar los representantes del curso que viajarán a Cartagena a representar al colegio, pueden considerarse varias estratificaciones. Una puede ser por género y, la otra, por iniciativa de ahorro.

Para realizar el muestreo estratificado categorizando por género, iniciamos diferenciando la lista por las categorías

|  |  |
| --- | --- |
| **ESTRATO 1** | **ESTRATO 2** |
| |  |  |  |  | | --- | --- | --- | --- | | **No.** | **Estudiante** | **¿Ahorra?** | **Género** | | 1 | Camilo | No | H | | 2 | Tomás | No | H | | 3 | Fausto | No | H | | 4 | Martín | No | H | | 5 | Julián | No | H | | 6 | Andrés | Sí | H | | 7 | Santiago | Sí | H | | 8 | Lucas | Sí | H | | 9 | Alejandro | No | H | | 10 | Mateo | No | H | | 11 | Daniel | Sí | H | | 12 | Juan | Sí | H | | 13 | Mauricio | No | H | | 14 | Joaquín | Sí | H | | 15 | Sebastián | No | H | | |  |  |  |  | | --- | --- | --- | --- | | **No.** | **Estudiante** | **¿Ahorra?** | **Género** | | 1 | Valeria | No | M | | 2 | Jimena | Sí | M | | 3 | Luisa | No | M | | 4 | Elizabeth | No | M | | 5 | Juana | Sí | M | | 6 | Liliana | No | M | | 7 | Ana | Sí | M | | 8 | Laura | No | M | | 9 | Sara | Sí | M | | 10 | Laura | No | M | | 11 | Natalia | Sí | M | | 12 | Tatiana | Sí | M | | 13 | Claudia | Sí | M | | 14 | Marcela | No | M | | 15 | Daniela | No | M | |

Luego de ello, aplicamos un muestreo simple o sistemático a cada uno de los estratos.

Haremos un muestreo simple en cada estrato. Para el primer estrato, usando un listado de números aleatorios para cada estrato y seleccionando números de dos cifras, se obtiene:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 713578 | 97227 | 53412 | 313227 | 86868 |

Los números son 71, 35, 78, 97, 22, 75, 34, 12, 31… Así, el muestreo arroja como ganador a Juan, que tiene el número 12 de entre los niños.

Repitiendo el proceso para el segundo estrato, se generan los números aleatorios y se procede igual. Los números aleatorios son:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 730992 | 112529 | 572350 | 774040 | 69027 |

Los números son 73, 09, 92,… Así, el muestreo arroja como ganadora a Sara, que tiene el número 9, de entre las niñas. Con este muestreo estratificado, los representantes del curso son Juan y Sara, y se garantizó que haya un niño y una niña.

En este caso no se calcula afijación, porque los estratos tienen la misma cantidad de individuos, y la probabilidad de que un niño sea escogido es 1/15, que es la misma de que una niña sea escogida.

En un muestreo que categorice la iniciativa de ahorro, también, iniciamos diferenciando la lista por las categorías

|  |  |
| --- | --- |
| **ESTRATO 1** | **ESTRATO 2** |
| |  |  |  | | --- | --- | --- | | **No.** | **Estudiante** | **¿Ahorra?** | | 1 | Jimena | Sí | | 2 | Andrés | Sí | | 3 | Santiago | Sí | | 4 | Juana | Sí | | 5 | Ana | Sí | | 6 | Lucas | Sí | | 7 | Sara | Sí | | 8 | Daniel | Sí | | 9 | Natalia | Sí | | 10 | Juan | Sí | | 11 | Tatiana | Sí | | 12 | Joaquín | Sí | | 13 | Claudia | Sí | | |  |  |  | | --- | --- | --- | | **No.** | **Estudiante** | **¿Ahorra?** | | 1 | Camilo | No | | 2 | Tomás | No | | 3 | Fausto | No | | 4 | Valeria | No | | 5 | Martín | No | | 6 | Julián | No | | 7 | Luisa | No | | 8 | Elizabeth | No | | 9 | Liliana | No | | 10 | Laura | No | | 11 | Alejandro | No | | 12 | Mateo | No | | 13 | Laura | No | | 14 | Mauricio | No | | 15 | Marcela | No | | 16 | Sebastián | No | | 17 | Daniela | No | |

Luego de ello, se aplica un muestreo simple o sistemático a cada uno de los estratos.

Haremos un muestreo simple en cada estrato. Para el primer estrato, usando un listado de números aleatorios para cada estrato y seleccionando números de dos cifras, se obtiene:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 649907 | 278501 | 929878 | 42982 | 666179 |

Los números son 64, 99, 07, 27,… Así, el muestreo arroja como ganadora a Sara, que tiene el número 7 de entre los que sí ahorran.

Repitiendo el proceso para el segundo estrato, se generan los números aleatorios y se procede igual. Los números aleatorios son:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 755076 | 831991 | 719148 | 496980 | 47466 |

Los números son 75, 50, 76, 83, 19, 91, 71, 91, 48 49, 69, 80, 47, 46 y 6,… Así, el muestreo arroja como ganador a Julián, que tiene el número 6 de entre los que no ahorran. Con este muestreo estratificado, los representantes del curso son Sara y Julián, y se garantizó que haya un estudiante que ahorra y uno que no ahorra.

En este caso se aplicó una afijación simple, porque a cada estrato le correspondió igual número de elementos muestrales, que en este caso fue uno. Sin embargo, como los estratos no tienen la misma cantidad de individuos, la probabilidad de que un estudiante que ahorra sea escogido es de 1/13, que no es igual a la probabilidad de que sea escogido un estudiante que no ahorra, que es de 1/17. En este caso, los estudiantes que no ahorran tienen mayor probabilidad de ser elegidos en la muestra.

**Muestreo por conglomerados**

A diferencia de los métodos de muestreo previos, en el muestreo por conglomerados la unidad muestral no son los elementos de la población, sino subgrupos de elementos de la población que conforman una nueva unidad. Tal unidad se conoce como *conglomerado*. Así, el muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados y en investigar después todos los elementos pertenecientes a los conglomerados elegidos, de manera que se alcance el tamaño muestral seleccionado.

Para efectuar el muestreo por conglomerados, lo primero que se hace es fraccionar la población en subgrupos que sean convenientes para el muestreo. Tales subgrupos serán los conglomerados y serán las unidades de estudio. Luego, se seleccionan tales unidades por algún método de muestreo aleatorio simple o sistemático para finalmente constituir la muestra con todos los elementos de los conglomerados seleccionados al azar. Este tipo de muestreo es eficaz cuando la población es muy grande y dispersa y presenta la ventaja de que no se requiere tener un listado de todas las unidades poblacionales, sino tan solo el listado de conglomerados como unidades primarias de muestreo.

En el ejemplo de la selección de los representantes del curso en el Foro de Educación, podrían elegirse como conglomerados las profesiones de los padres de los estudiantes. Luego se selecciona al azar la cantidad de profesiones que completen la muestra. Suponga que en la selección al azar se eligieron como conglomerados “Militar” y otro, “Abogado”. Si hay dos estudiantes que sean hijos de militar, ellos serán los representantes del curso. Si hay más de dos, deberán seleccionarse dos de entre el total de estudiantes que cumplan la condición. Si hay uno o ninguno, habrá que escoger entre los que son hijos de abogados el segundo representante del curso.

En general los estudios poblacionales son complejos, por lo es que es de gran utilidad usar el muestreo polietápico que opera en etapas sucesivas, sobre las que se aplica el método de muestreo aleatorio que se considere más adecuado en cada etapa. En [VER](http://recursostic.educacion.es/descartes/web/materiales_didacticos/muestreo_poblaciones_ccg/tipos_muestreo.htm) hay información adicional acerca de los tipos de muestreo

[SECCIÓN 3] **2.2 Consolidación**

SECCIÓN 1] **3 Las medidas estadísticas**

Luego de tener claridad respecto al tipo de variable al que refiere un estudio estadístico, de hacer el muestro que garantice que el estudio será manejable y representativo de la población y de tener el conjunto de datos –bien sea listado en detalle o representado en una tabla de frecuencias–, se caracterizan de manera cuantitativa datos más específicos respecto a la población y a la manera en que se distribuyen los datos.

Las medidas de tendencia central como la media, la moda y la mediana son utilizadas en este proceso para establecer el comportamiento, la dispersión o la agrupación del conjunto de datos.

Tal cuantificación puede medir, bien la tendencia de la población, bien su grado de dispersión, o bien los comportamientos que relacionan conjuntamente tanto la tendencia como la dispersión de los datos. Las medidas de posición central y no central y las medidas de dispersión permiten hacer la caracterización.

[SECCIÓN 2] **3.1 Las medidas de tendencia central**

Las medidas de tendencia central o valores centrales indican un valor central alrededor del cual se distribuyen el resto de valores de la distribución. Los valores de posición centrales más utilizados son la *media aritmética,* la *mediana* y la *moda.*

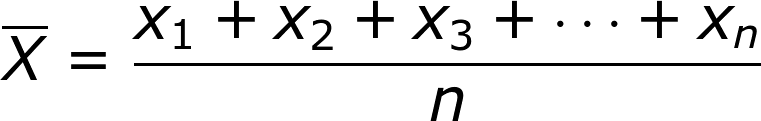
**La media aritmética**

La media aritmética representa en un único valor la cantidad total de la variable, distribuida a partes iguales entre cada observación; se obtiene al sumar todos los datos y dividir el resultado entre el número total de datos. Para representarla se utiliza la letra mayúscula *X* con una barra horizontal sobre ella: *X̄*. No debe confundirse con el uso de la letra minúscula *x* que representa cualquiera de los datos de la muestra.

La media puede ser un número que no tenga sentido en el contexto propuesto; tiene sentido en contextos continuos como el estudio de la variabilidad del peso o la altura de una muestra en una población, pero no necesariamente lo tiene en contextos de estudio de variables discretas, como por ejemplo en el caso en que nos arroja como resultado que las personas leen 1,5 libros o compran 2,5 pares de zapatos.

Los estudiantes suelen tener un acercamiento al cálculo de la media aritmética, pues es la medida de tendencia central que suele aplicarse para el cálculo de las notas en las instituciones escolares que basan las calificaciones en datos numéricos.

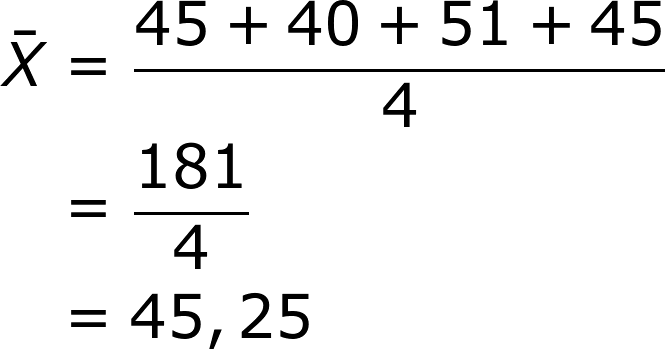
Específicamente, si *x1*, *x2*, *x3*,…, *xn*, representa nuestro conjunto de *n* datos, la media aritmética corresponde a:



donde los *xi* corresponden a cada uno de entre los *n* datos. Si se cuenta con un listado de datos en Excel sobre los cuales se desea calcular la media aritmética, se aplica la función =PROMEDIO (), variando en el conjunto de datos.

Para entender en qué consiste la media aritmética *X̄*, veamos un ejemplo:

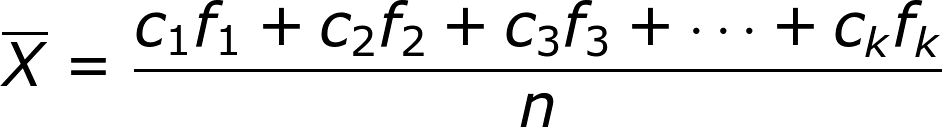
Un estudiante ha obtenido los datos del peso en kilogramos de cuatro de sus compañeros, los cuales son 45, 40, 51 y 45 Kilos. Para establecer un valor medio del peso de sus cuatro compañeros, calcula la media aritméticasumando las cantidades que representan los pesos sus compañeros. Luego, divide el entre el número total de individuos, que en este caso es 4. Entonces, el cálculo realizado en el primer momento es: *(45 + 40 + 51 + 45) = 181*, que es la suma de los pesos de los cuatro. En el segundo momento, lo divide y obtiene *181 / 4 = 45,25* kilos, lo cual corresponde a la media aritmética del conjunto de datos y se escribe: *X̄ = 45,25* kilos. En la escritura estándar ello significa:



Lo que significa que la media o promedio aritmético del peso de los cuatro estudiantes es de 45,25 kilos.

En el caso en que el conjunto de datos se presente en una tabla de frecuencias, el cálculo de media requiere el uso de las frecuencias relativas.

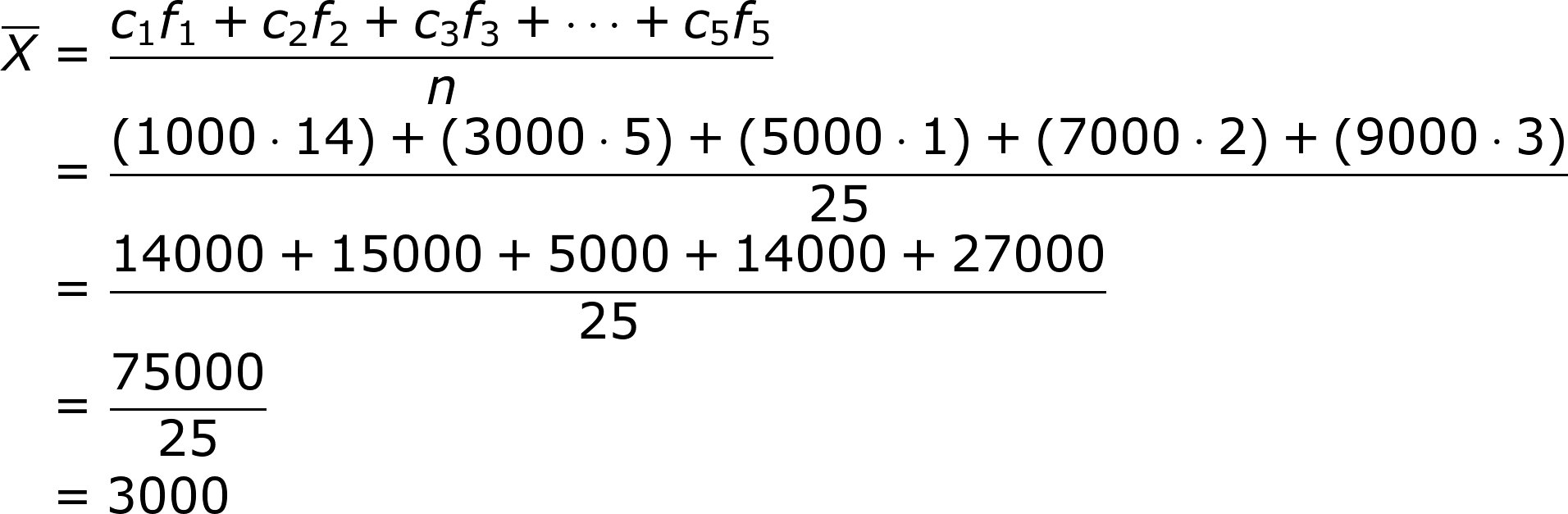
Para encontrar la media aritmética ( *X̄* ) de un conjunto de datos proveniente de una tabla de frecuencias, el cálculo consiste en sumar los productos de las marcas de clase (*ci*) por las frecuencias relativas (*fi*) de cada clase y luego dividir todo por la cantidad de datos. Así, si tenemos *k* clases con marcas de clase *c1*, *c2*, *c3*, …, *ck*  y sus frecuencias relativas respectivas *f1*, *f2*, *f3*, …, *fk*, de un conjunto de *n* datos, el resultado de la media aritmética se obtiene calculando:



Por ejemplo la tabla de frecuencias siguiente muestra los resultados de una encuesta aplicada en una clase de 25 alumnos en la que se les preguntó la máxima distancia a la que cada uno había viajado, medida entre su actual lugar de residencia y la capital del país extranjero visitado.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cantidad de kilómetros recorrida**  ***xi*** | **Marca de clase**  ***(ci)*** | **Frecuencia absoluta *(fi)*** | **Frecuencia relativa**  ***(hi= fi/n)*** | **Frecuencia absoluta acumulada *(Fi)*** | **Frecuencia relativa acumulada *(Hi= Fi/n )*** | **Porcentaje (%)**  **(*hi* escrita como porcentaje)** |
| [0, 2.000) | 1.000 | 14 | 14 / 25 = 0,56 | 14 | 14 / 25 | 56 % |
| [2.000, 4.000) | 3.000 | 5 | 5 / 25 = 0,2 | 14 + 5 = 19 | 19 / 25 | 20 % |
| [4.000, 6.000) | 5.000 | 1 | 1 / 25 = 0,04 | 19 + 1 = 20 | 20 / 25 | 4 % |
| [6.000, 8.000) | 7.000 | 2 | 2 / 25 = 0,08 | 20 + 2 = 22 | 22/ 25 | 8 % |
| [8.000, 10.000) | 9.000 | 3 | 3 / 25 = 0,12 | 22 + 3 = 25 | 25 / 25 | 12 % |
| TOTAL |  | 25 |  | 25 | 25 / 25 = 1 | 100.00% |

Para el cálculo de la distancia media a la que ha viajado el grupo de estudiantes, nótese que *n=25* y corresponde a la cantidad de datos, mientras que *k=5* indica la cantidad de clases de la tabla. Ya con ello, se establece que la distancia media es:

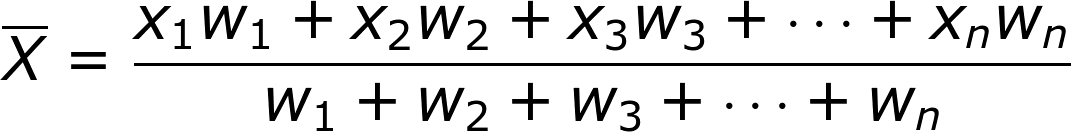


Así que el grupo de estudiantes ha viajado, en promedio, 3000 kilómetros.

**La media aritmética ponderada**

No en todas las ocasiones los datos obtenidos tienen la misma relevancia o peso para mostrar el comportamiento de los datos en la población. Cuando ello ocurre, se asignan pesos a los datos, dependiendo de su relevancia en el marco del estudio a realizar y de esa manera se obtiene la media ponderada del conjunto de datos.

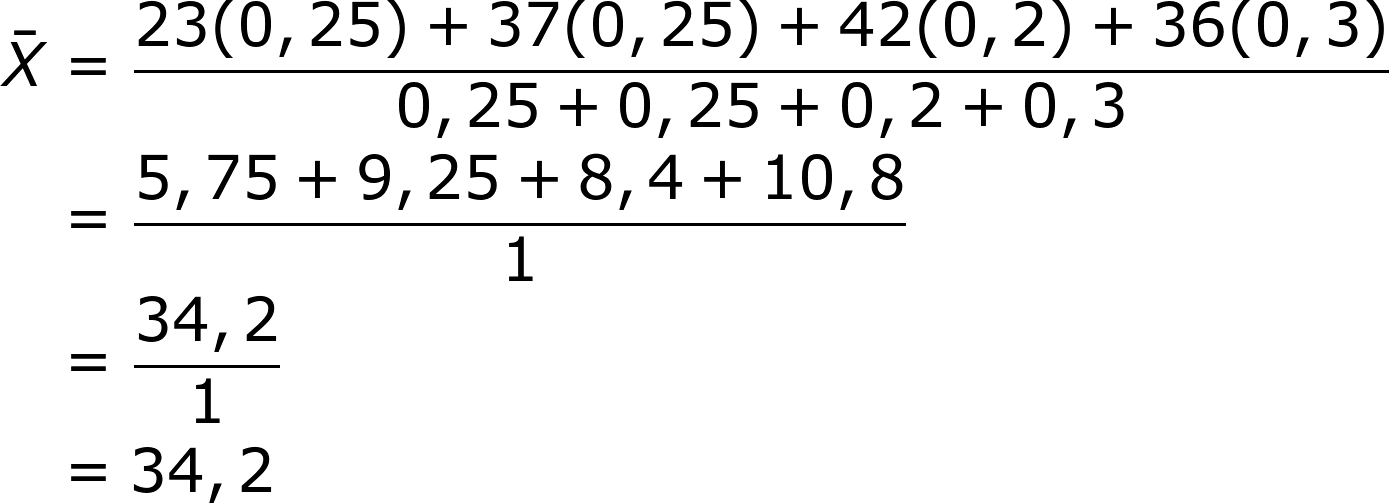
La media ponderada se calcula haciendo los productos entre los datos y su nivel de relevancia, que se expresa como un porcentaje. Más específicamente, si *x1*, *x2*, *x3*, …, *xn*, representa nuestro conjunto de *n* datos y si *w1*, *w2*, *w3*, …, *wn*, representa el conjunto de *n* pesos respectivos, la media ponderada se define de la siguiente forma:

****

Veamos el cálculo de la media ponderada en un ejemplo concreto:

Para valorar los desempeños de los estudiantes en primer semestre, una profesora hace dos pruebas escritas parciales y una prueba general final; valora además la actitud en clase. Cada una de las pruebas parciales corresponde al 25 % sobre la nota final, la actitud en clase corresponde al 20 % y -por reglamento- la prueba final corresponde al 30 % de la nota de la materia. ¿Cuál será la puntuación de un alumno que en las dos pruebas parciales sacó 23 y 37, respectivamente, y cuyas notas en el examen final y en actitud han sido 36 y 42 respectivamente?

Para conocer la respuesta se calcula la media aritmética ponderada:



Así, el desempeño del estudiante en esa materia es de 34,2. Si las notas son calculadas sobre 50 y la nota aprobatoria es 30, pasó la materia. Si las notas son calculadas sobre 100 y la nota aprobatoria es 50, perdió la materia. Si está en posgrado, las notas son calculadas sobre 50 y la nota aprobatoria es 35, así que el estudiante reprobó la materia.

En el caso en el que los datos para el cálculo de la media ponderada provengan de una tabla de distribución de frecuencias, habría que garantizar que los pesos de la ponderación varían justamente en las clases establecidas. De lo contrario, habría que expandir los datos y hacer las ponderaciones como en el caso para datos no agrupados.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La media aritmética y ponderada** |
| **Contenido** | Llamamos media aritmética ( *X̄* ) de un conjunto *x1*, *x2*, *x3*, …, *xn* de *n* datos al cociente de la suma de todos ellos por el número *n*.    Si el cálculo de la media aritmética se hace sobre una tabla de distribución de frecuencias, con *k* clases, marcas de clase *c1*, *c2*, *c3*, …, *ck*  y frecuencias relativas correspondientes *f1*, *f2*, *f3*, …, *fk*, de un conjunto de *n* datos, la media aritmética es:    Para encontrar la media aritmética ponderada ( *X̄* ) de un conjunto *x1*, *x2*, *x3*, …, *xn*  de *n* datos, se suman los productos de estos números por sus pesos respectivos *w1*, *w2*, *w3*, …, *wn* (en %) y el resultado se divide por la suma de los pesos (que corresponde al 100 %, osea que debe sumar 1). |

**La mediana**

Otra importante medida de tendencia central es la llamada ***mediana*** de un conjunto de datos, que corresponde al valor que se encuentra **en el centro de la serie ordenada de datos**. Para representarla se utiliza la el apócope *Med*. Si la cantidad de datos es impar, la mediana es el valor central entre el conjunto de *n* datos, que deja la misma cantidad de datos a derecha e izquierda de la lista ordenada. Si la cantidad *n* de datos es impar, no habrá un único valor que está en el centro de la lista ordenada; en ese caso la mediana corresponde a la media aritmética entre los dos valores centrales en la lista ordenada. La función en Excel que permite hacer el cálculo de la mediana es = MEDIANA(), con variación en el conjunto de datos.

|  |  |
| --- | --- |
| Si *n* es par | Si *n* es impar |
| donde los *xi* están en la lista ordenada de *n* datos | donde los *xi* están en la lista ordenada de *n* datos |

Para ejemplificar el cálculo de la mediana veamos los siguientes ejemplos:

En el campeonato local de la liga colombiana de fútbol se juegan, para la misma temporada, dos fases: la fase “Todos contra todos” compuesta por 20 fechas y la fase de “finales”. La cantidad de goles que marcó un equipo en las 20 fechas de la fase “Todos contra todos” se presenta a continuación:

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

Para calcular la mediana de los datos, se procede de la siguiente manera:

* Se ordena la lista de datos, de menor a mayor

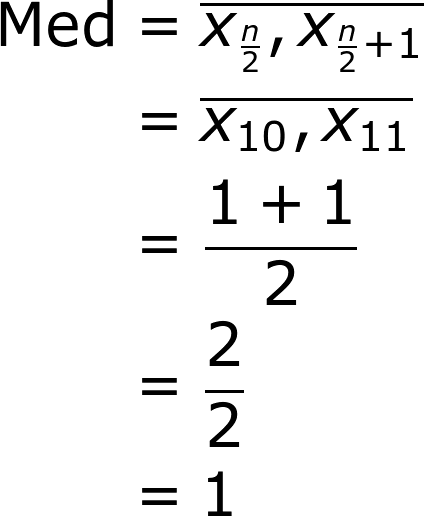
0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3

* Se identifica si la cantidad de datos es par o impar

en este caso, como *n=20*, la cantidad de datos es par

* Si la cantidad de datos es impar, se escoge el valor central de la lista ordenada. Este no es el caso, pues tenemos cantidad de datos par.
* Si la cantidad de datos es par, se escogen los dos valores centrales de la lista ordenada y se hace la media aritmética entre ellos.
* 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3

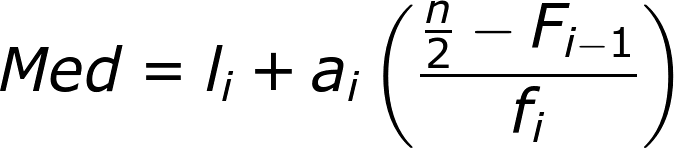
En nuestro listado, los dos valores centrales son los de las posiciones 10 y 11 de la lista ordenada, que dejan 9 datos a cada lado de la lista ordenada. La mediana es la media aritmética entre ellos dos, es decir que se obtiene:



Así, la mediana es la media aritmética o promedio de la pareja de **valores centrales 1**; esta es la **mediana** de la cantidad de goles marcados por el equipo en la fase “Todos contra todos”.

La mediana también puede ser encontrada cuando el conjunto de datos ha sido agrupado por intervalos. En este caso se debe conocer el límite inferior de la clase donde se encuentra la mediana (*li*), la frecuencia acumulada anterior a la clase donde está la mediana (*Fi-1*), la frecuencia absoluta de la clase donde está la mediana (*fi*) y la amplitud de la clase en que se encuentra la mediana (*ai*).

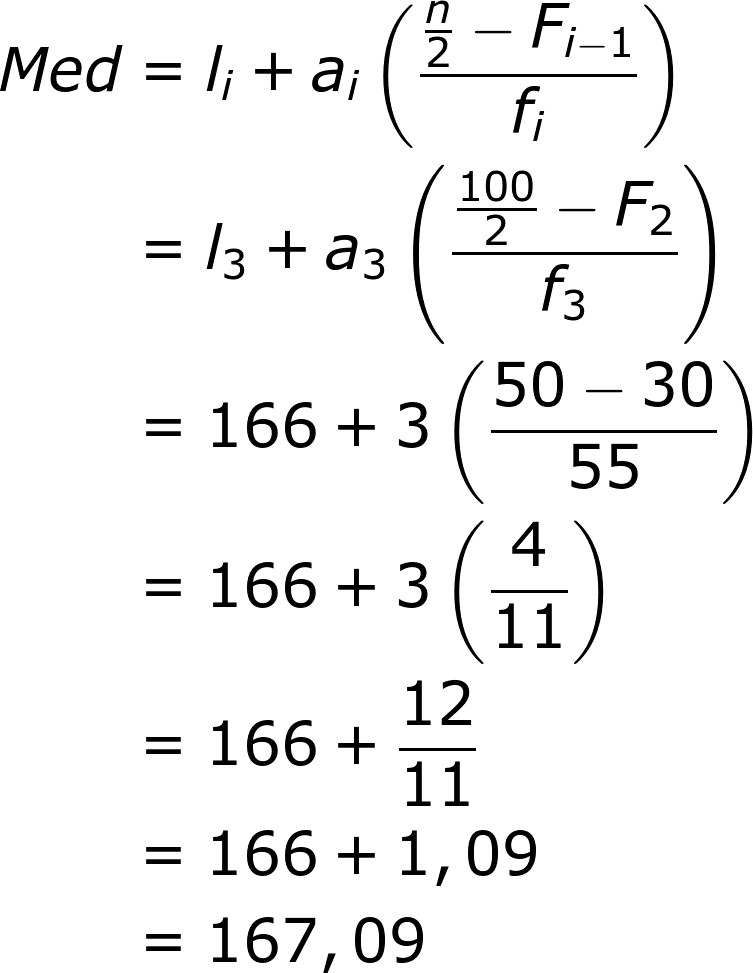
Para el cálculo se procede buscando el valor que corresponde a la posición *n / 2*, y se hace el cálculo de la mediana en la clase *i* a la que pertenezca la mediana, por medio de la siguiente ecuación:



Por ejemplo, en un estudio para identificar tallajes para las chaquetas de la promoción, se agrupan las estaturas de un grupo 100 de estudiantes en cinco intervalos:

|  |  |  |
| --- | --- | --- |
| ***Estatura (en cms)*** | ***fi*** | ***Fi*** |
| [160, 163) | 7 | 7 |
| [163, 166) | 23 | 30 |
| [166, 169) | 55 | 85 |
| [169, 172) | 10 | 95 |
| [172, 175) | 5 | 100 |

En este caso, la clase en la que se encuentra la mediana es la clase *i =* 3, pues es en esa clase en la que se sobrepasa la mitad de los datos ordenados. Así, para este caso, con *i =* 3, se tiene que el límite inferior de la tercera clase es 166, es decir *l3 = 166*; la amplitud de la tercera clase en que se encuentra la mediana es *a3 =* 3, la frecuencia acumulada anterior a la clase tres en que está la mediana es *F2 = 30*, la frecuencia absoluta de la clase 3 es *f3 = 55* y la mitad de los datos es *n/2 = 50*. Así, el cálculo de la mediana queda:



Por lo que la mediana para el conjunto de datos agrupados es 167,09 cms de estatura, lo que significa que 50 estudiantes miden menos de 167, 09 y los otros 50 miden más que 167,09.

**La moda**

La **moda** de un conjunto de datos indica la tendencia de dicho conjunto, es decir el dato más frecuente que corresponde al valor que más veces se repite. La moda se representa como *Mod* y es siempre un valor que tuvo ocurrencia en el conjunto de datos. Se trata de una medida de tendencia central que se puede calcular tanto para variables cualitativas como cuantitativas. El cálculo de la moda para datos sin agrupar se realiza haciendo el conteo de la cantidad de veces que se repite cada dato y eligiendo como moda aquel que tenga la mayor repetición. En caso de que varios datos se repitan la misma cantidad de veces, se dirá que la población tiene una distribución bimodal, trimodal y, en general, multimodal.

Por ejemplo en el siguiente par de series se presenta el resultado de una encuesta en la que se preguntó a los 29 estudiantes de un salón la edad en la que experimentaron su primer beso. Los resultados para los 11 estudiantes de género masculino fueron los siguientes:

Edades para los hombres: 13,14, 14, 13, 15, 15, 14, 15, 13, 15, 15

En este conjunto de datos se puede evidenciar que 15 es la moda o el valor más repetido.

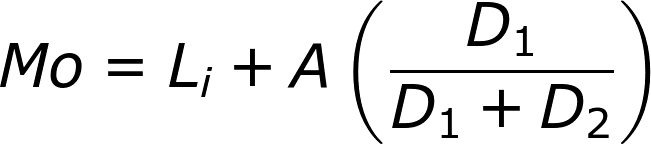
Por su parte, en el grupo de género femenino, las restantes 18 estudiantes respondieron lo siguiente:

Edades para las mujeres: 14, 13,14, 14, 13, 15, 12, 15, 14, 15, 13, 15, 15, 14, 15, 15, 14, 14.

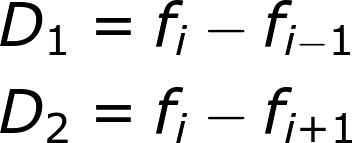
En este conjunto de datos, las edades 14 y 15 tienen la misma frecuencia, pues se repiten 7 veces cada una. Cuando existen dos modas o dos valores que tiene valores de mayor frecuencia se dice que el conjunto de datos tiene una distribución bimodal.

En el caso en que se estudie la población total del salón sin estratificarla por género, en los datos de los 29 estudiantes la distribución es unimodal, con *Mod = 15*.

Para encontrar la moda en un conjunto de datos agrupados es necesario tener en cuenta el límite inferior de la clase modal o clase en la que se presenta la máxima frecuencia (*Li*), la frecuencia absoluta de la clase modal (*fi*), la frecuencia absoluta de la clase inmediatamente inferior a la clase modal, (*fi-1*), que llamaremos *clase premodal*, la frecuencia absoluta de la clase inmediatamente posterior a la clase modal (*fi+1*), que llamaremos *clase postmodal* y la amplitud de las clases (*A*). El cálculo de la moda en este caso responde a la ecuación:



donde *D1* y *D2* representan, la diferencia entre la clase modal y la premodal, y la diferencia entre la clase modal y la postmodal respectivamente. Es decir que:

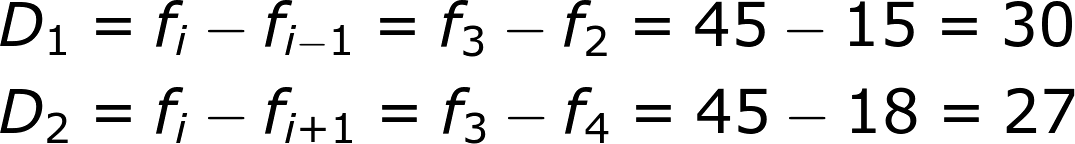


Nótese que los valores *D1* y *D2* son siempre positivos, dado que *fi* es el máximo entre las frecuencias.

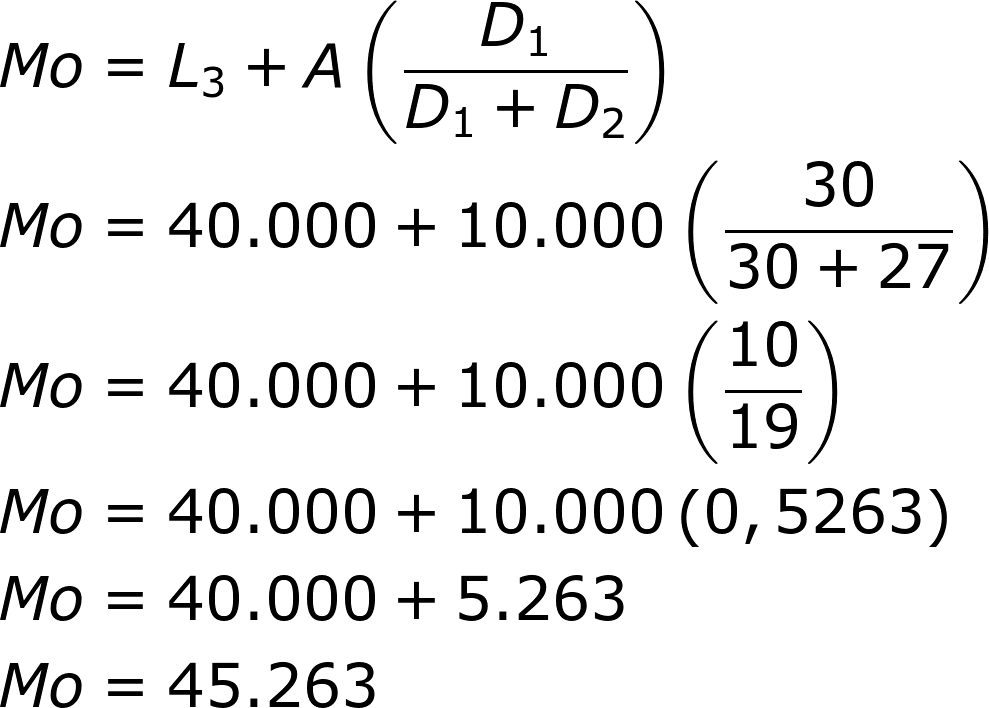
Para comprender el funcionamiento de la ecuación para el cálculo de la moda para datos agrupados, veamos el siguiente ejemplo: la tabla representa los resultados respecto a la cantidad de dinero o salario, en pesos, que gana un grupo de 100 padres de familia.

|  |  |
| --- | --- |
| **Dinero ganado en un día** | ***fi*** |
| [20.000, 30.000) | 10 |
| [30.000, 40.000) | 15 |
| [40.000, 50.000) | 45 |
| [50.000, 60.000) | 18 |
| [60.000, 70.000) | 12 |

Entonces, inicialmente observamos cuál es la clase en la que aparece la frecuencia absoluta más alta. En este caso la clase con frecuencia máxima es la tercera, es decir que *i = 3*. Ahora bien, calculando los valores de las diferencias *D1* y *D2* se obtiene:



Ya con los valores de las dos diferencias, se procede al cálculo de la moda. Así:



De esta manera, la moda o la mayor frecuencia en los datos recolectados es de $45.263, lo cual significa que la mayoría de los padres del grupo obtiene ese salario cada día.

[SECCIÓN 2] **3.2 Las medidas de dispersión**

Si bien el conocimiento de las medidas de tendencia central para una población arroja información importante para lograr caracterizar dicha población, es cierto también que dos distribuciones pueden tener las mismas medidas de centralización y, aun así, ser muy diferentes.

La estadística reconoce por ejemplo que cuando la desviación de los valores observados respecto a un valor central es muy grande, las medidas de tendencia central no son representativas del total; reconoce así que es necesario caracterizar mejor la población. Para lograrlo, trata de estudiar y cuantificar la dispersión de los datos en la muestra o en la población, lo cual se conoce como ***medidas de dispersión***.

En el objetivo de completar el conocimiento de la distribución se proponen las siguientes medidas de dispersión: el rango, la desviación media, la varianza, la desviación estándar y el coeficiente de variación.

**El rango**

Una primera forma de **medir la dispersión se logra estudiando qué tan grande es el intervalo en el que varían los datos.** Para obtener esa medida se calcula el recorrido o rango, es decir, la **diferencia** entre la observación **máxima** y la **mínima en la población**, el cual se simboliza mediante la letra R.

Cuando los datos se distribuyen de manera homogénea, el rango caracteriza bien la dispersión de la población. Sin embargo, ya que el rango depende exclusivamente de las observaciones máxima y mínima, la dispersión puede estar desajustada en el caso de que dichos extremos sean valores atípicos. Como identificación de dispersión, se considera que cuanto mayor sea el rango o recorrido, más grande será la dispersión de los datos.

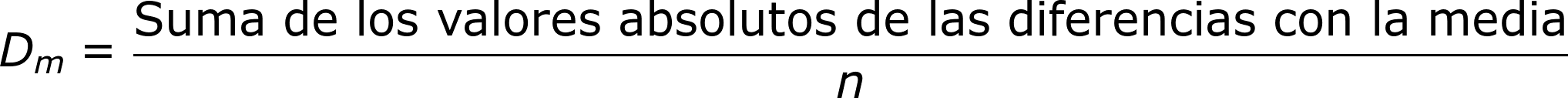
Por ejemplo, en el estudio de una población se recogieron los datos acerca de la cantidad de lotes de un artículo, producidos en un turno de la fábrica. El número mínimo de lotes fue 23 y el máximo 89. Por tanto, el rango o recorrido sería:

Rango o recorrido = 89 – 23 = 66 lotes.

Entonces, se sabe que la cantidad de lotes por turno puede variar entre 66 datos, pero no se tienen más datos acerca de la población.

**La desviación media**

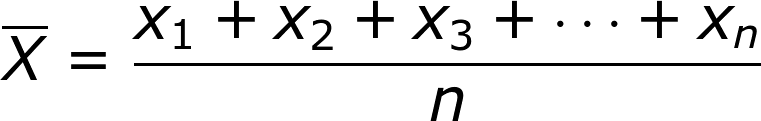
Otra forma de caracterizar la dispersión de la población se realiza midiendo qué tan lejos están las observaciones respecto a la media, las cuales se conocen como *desviaciones respecto a la media*. La media aritmética de los valores absolutos de estas desviaciones es lo que se conoce como *desviación media* de la población con *n* datos, que se representa por *Dm.* Así:



Para ejemplificar el cálculo de la desviación media veamos el siguiente ejemplo: Si la cantidad de goles que marcó un equipo en las 20 fechas de la fase “Todos contra todos” es

2, 1, 0, 1, 3, 0, 1, 1, 0, 2, 1, 0, 1, 0, 1, 1, 2, 0, 2, 1, ¿cuál es la desviación media?

Para calcularla, inicialmente se calcula la media aritmética del conjunto de datos.



Ya que la suma de los datos es 20 y que la cantidad de datos es 20, entonces la media aritmética es *X̄* = 20/20 = 1.

Las desviaciones respecto a la media son las siguientes:

|  |  |  |
| --- | --- | --- |
| **Cantidad de goles**  ***xi*** | **Desviaciones respecto a la media**  ***xi - X̄*** | **Valores absolutos de las desviaciones respecto a la media**  **| *xi - X̄|*** |
| 2 | 2 – 1 = 1 | 1 |
| 1 | 1 – 1 = 0 | 0 |
| 0 | 0 – 1 = **-1** | 1 |
| 1 | 1 – 1 = 0 | 0 |
| 3 | 3 – 1= 2 | 2 |
| 0 | 0 – 1= **-1** | 1 |
| 1 | 1 – 1= 0 | 0 |
| 1 | 1 – 1= 0 | 0 |
| 0 | 0 – 1 = **-1** | 1 |
| 2 | 2 – 1 = 1 | 1 |
| 1 | 1 – 1= 0 | 0 |
| 0 | 0 – 1 = **-1** | 1 |
| 1 | 1 – 1= 0 | 0 |
| 0 | 0 – 1= **-1** | 1 |
| 1 | 1 – 1= 0 | 0 |
| 1 | 1 – 1= 0 | 0 |
| 2 | 2 – 1 = 1 | 1 |
| 0 | 0 – 1 = **-1** | 1 |
| 2 | 2 – 1 = 1 | 1 |
| 1 | 1 – 1= 0 | 0 |
| **Desviación media = Media aritmética de los valores absolutos de las desviaciones** | | *Dm* = 12/20 = 0,6 |

Así, la desviación media del conjunto de datos es *Dm* = 0,6, que indica el grado de dispersión (o alejamiento) de los datos respecto a la media.

Nota que algunas de las diferencias resultan negativas. Por ello se requiere calcular los valores absolutos de las desviaciones, de manera que la desviación media es siempre una media aritmética de valores positivos. Cuando los datos están agrupados, se toma la marca de clase como la media aritmética de los extremos de un intervalo, lo que simplifica el cálculo de la desviación media pues de ese modo se sustituye cada intervalo por un solo número.

**La varianza**

**La varianza** (o ***σ2***) de un conjunto de valores **es la media aritmética de los cuadrados de las desviaciones** de esos valores respecto a la media aritmética. Se expresa en unidades cuadradas.

Calculemos la varianza de la población usada en el ejemplo anterior. En este caso:

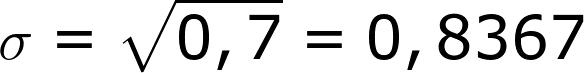
|  |  |  |
| --- | --- | --- |
| **Cantidad de goles**  ***xi*** | **Desviaciones respecto a la media**  ***xi - X̄*** | **Cuadrados de las desviaciones respecto a la media**  **( *xi - X̄ )2*** |
| 2 | 2 – 1 = 1 | 1 |
| 1 | 1 – 1 = 0 | 0 |
| 0 | 0 – 1 = **-1** | 1 |
| 1 | 1 – 1 = 0 | 0 |
| 3 | 3 – 1= 2 | 4 |
| 0 | 0 – 1= **-1** | 1 |
| 1 | 1 – 1= 0 | 0 |
| 1 | 1 – 1= 0 | 0 |
| 0 | 0 – 1 = **-1** | 1 |
| 2 | 2 – 1 = 1 | 1 |
| 1 | 1 – 1= 0 | 0 |
| 0 | 0 – 1 = **-1** | 1 |
| 1 | 1 – 1= 0 | 0 |
| 0 | 0 – 1= **-1** | 1 |
| 1 | 1 – 1= 0 | 0 |
| 1 | 1 – 1= 0 | 0 |
| 2 | 2 – 1 = 1 | 1 |
| 0 | 0 – 1 = **-1** | 1 |
| 2 | 2 – 1 = 1 | 1 |
| 1 | 1 – 1= 0 | 0 |
| **Varianza = Media aritmética de los cuadrados de las desviaciones** | | ***σ2*** = 14/20 = 0,7 |

Así que la varianza en este caso será ***σ2*** = 0,7 goles***2***

**La desviación estándar**

**Para expresar la desviación en las mismas unidades que las usadas al hacer las observaciones utilizamos como medida de dispersión la raíz cuadrada positiva de la varianza. Dicha raíz cuadrada se llama *desviación estándar*, que se representa por *σ*. La desviación estándar sirve para medir el grado de dispersión, pues cuando dos distribuciones tienen la misma media, la desviación estándar indica lo alejados que están los valores respecto a esa media.**

**La desviación estándar** para la cantidad de goles que marcó el equipo en las 20 fechas de la fase “Todos contra todos” tuvo como varianza ***σ2*** = 0,7 goles***2*, por lo que la desviación estándar para el conjunto de datos será:**



El cálculo de las medidas de dispersión para el caso en que los datos se encuentran agrupados, se sigue el procedimiento para hacer el cálculo de la media aritmética y, posteriormente, se calculan las dispersiones y sus cuadrados, a fin de calcular tanto la varianza como la desviación estándar. El procedimiento se ejemplifica a continuación:

Se ha obtenido la medida de la altura de 100 personas de una institución. Los resultados obtenidos, agrupados en intervalos han sido:

|  |  |
| --- | --- |
| **Intervalo**  **(Talla en cm)** | **Frecuencia absoluta**  ***Fi*** |
| [155, 160) | 6 |
| [160, 165) | 16 |
| [165, 170) | 24 |
| [170, 175) | 27 |
| [175, 180) | 19 |
| [180, 185) | 8 |
| ***n*** | 100 |

Inicialmente, se incluye una columna, para sustituir cada intervalo, por su respectiva marca de clase, que corresponde el punto medio del intervalo:

|  |  |  |
| --- | --- | --- |
| **Intervalo**  **(talla en cm)** | **Marca de clase**  ***xi*** | **Frecuencia absoluta**  ***fi*** |
| [155, 160) | (155, 160) / 2 = 157,5 | 6 |
| [160, 165) | (160, 165) / 2 = 162,5 | 16 |
| [165, 170) | (165, 170) / 2 = 167,5 | 24 |
| [170, 175) | (170, 175) / 2 = 172,5 | 27 |
| [175, 180) | (175, 180) / 2 = 177,5 | 19 |
| [180, 185) | (180, 185) / 2 = 182,5 | 8 |
| ***n*** | | **100** |

A continuación se utilizan las marcas de clase para encontrar la media de las estaturas medidas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervalo**  **(talla en cm)** | **Marca de clase**  ***xi*** | **Frecuencia absoluta**  ***fi*** | ***fi* *xi*** |
| [155, 160) | (155, 160) / 2 = 157,5 | 6 | 945 |
| [160, 165) | (160, 165) / 2 = 162,5 | 16 | 2.600 |
| [165, 170) | (165, 170) / 2 = 167,5 | 24 | 4.020 |
| [170, 175) | (170, 175) / 2 = 172,5 | 27 | 4.658 |
| [175, 180) | (175, 180) / 2 = 177,5 | 19 | 3.373 |
| [180, 185) | (180, 185) / 2 = 182,5 | 8 | 1.460 |
|  |  | **100** | **17.055** |

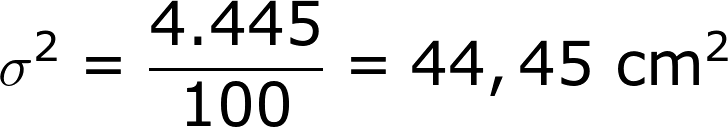
Con los resultados de la última columna y sabiendo que la cantidad *n* total de datos es *n = 100*, se obtiene la media aritmética del conjunto de datos:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14643/InfoGuion/cuadernoestudio/images_xml/MT_10_10_formula6_resized.gif

Para realizar el cálculo de la varianza y de la desviación estándar se calculan las desviaciones y sus cuadrados. En la última columna de la derecha de la siguiente tablar se recogen los datos para poder calcular la varianza y la desviación estándar.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Intervalo**  **(talla en cm)** | **Marca de clase**  ***xi*** | **Frecuencia absoluta**  ***fi*** | ***(xi* – *X̄)2*** | ***fi* *(xi* – *X̄)2*** |
| [155, 160) | (155, 160) / 2 = 157,5 | 6 | 170 | 1.022 |
| [160, 165) | (160, 165) / 2 = 162,5 | 16 | 65 | 1.037 |
| [165, 170) | (165, 170) / 2 = 167,5 | 24 | 9 | 223 |
| [170, 175) | (170, 175) / 2 = 172,5 | 27 | 4 | 103 |
| [175, 180) | (175, 180) / 2 = 177,5 | 19 | 48 | 918 |
| [180, 185) | (180, 185) / 2 = 182,5 | 8 | 143 | 1.142 |
| ***n*** | | **100** |  | **4.445** |

En base a lo anterior, se deduce que la varianza es:



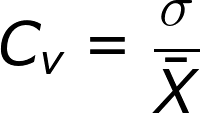
Sacando la raíz cuadrada positiva a la varianza se obtiene la desviación estándar que será entonces:



**El coeficiente de variación**

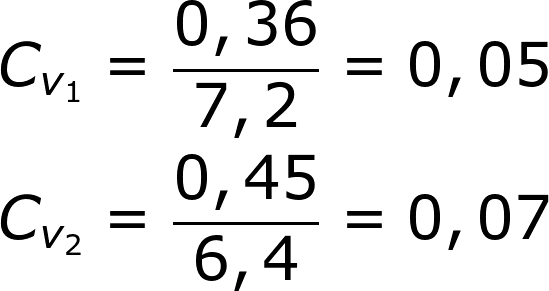
Para **comparar** dos distribuciones heterogéneas con distinta media y saber la dispersión de sus valores, podemos utilizar el **coeficiente de variación**, que normalmente se expresa en porcentaje (%). El coeficiente de variación es el cociente de la desviación estándar entre la media y se simboliza mediante Cv. Cuando se comparan dos distribuciones, la que tiene mayor coeficiente de variación es también la que tiene mayor dispersión.

El coeficiente de variación se calcula dividiendo la desviación estándar entre la media:



Por ejemplo, la nota media en matemáticas de la clase de 10º A ha sido de 7,2 y su desviación estándar de 0,36. En cambio, en la clase de 10º B ha sido de 6,4 y desviación estándar de 0,45. ¿Cuál es el coeficiente de variación de la clase de 10º A? ¿Y el de 10º B?

Aplicando la ecuación se obtiene:



Así, como el coeficiente de variación para los estudiantes del primer curso fue de 0,05, eso significa que la variación es del 5 % en 10º A. Para el segundo curso el coeficiente de variación de 0,07 indica que la variación es del 7 % en 10º B. Por lo tanto se concluye que las notas de matemáticas han estado más dispersas en la clase de 10º B.

[SECCIÓN 2] **3.3 Las medidas de posición no central**

Las medidas estadísticas de posición pueden ser **valores centrales** o valores **no centrales**. En estadística tanto los primeros como los segundos se calculan para hacer la caracterización de la población. Los más comunes valores de posición central, llamados también medidas de tendencia central son la **media aritmética**, la **mediana** y la **moda.**

**Por su parte, las medidas de posición no central sirven para dividir** un conjunto de valores de una variable estadística en subconjuntos con la misma cantidad de valores, **y para establecer el comportamiento de los datos en diferentes puntos. Como estas medidas son complementarias a los valores centrales, pueden ayudar a dar información más completa del conjunto de datos,** generar diferentes representaciones de la muestra que ayuden a entender el comportamiento de la misma y a hacer inferencias acerca del comportamiento de la población de estudio. Las medidas de tendencia no central más utilizadas en estadística son los cuartiles, los quintiles, los deciles y los percentiles.

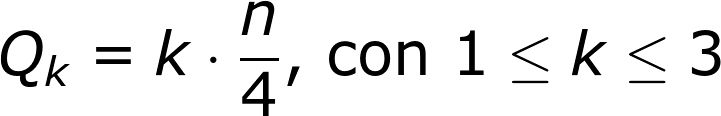
**Los cuartiles**

Los **cuartiles, como medida de posición no central,** dividen el conjunto de datos en cuatro grupos iguales. Es decir, se si tuviera la lista de 100 datos ordenados de menor a mayor, los cuartiles los dividirían en subgrupos de 25 datos; en una población de 60 datos, los cuartiles los dividirían en subconjuntos de 15 datos.

Los cuartiles son tres, Q1, Q2 y Q3. El primer cuartil es Q1, hasta el que se encuentran el primer 25% de los datos. El segundo cuartil es Q2 hasta el que se encuentra el segundo 25% de los datos, es decir el 50% de los datos y por ello coincide exactamente con la mediana. Finalmente, el tercer cuartil es Q3, hasta el que se encuentra el tercer 25% de los datos, es decir el 75 % de los datos. Así, entre el primer y el tercer cuartil, es decir entre Q1 y Q3, se encuentra ubicada la mitad de la población.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Representación de los cuartiles sobre un segmento de recta |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | http://dieumsnh.qfb.umich.mx/estadistica/medidasd%20de%20posicion_archivos/image002.jpg |
| **Pie de imagen** | Representación de los cuartiles |
| **Ubicación del pie de imagen** | Inferior o lateral |

La ecuación que permite calcular la posición de cada cuartil en un conjunto de *n* datos es:



La variación del parámetro *k* permite obtener la posición sobre la lista ordenada –de menor a mayor– de datos para los tres cuartiles. Así, para *k = 1*, se obtiene la posición del primer cuartil; para *k = 2*, se obtiene la posición del segundo cuartil o mediana; y para *k = 3*, se obtiene la posición del tercer cuartil.

Otra forma de decirlo es que el primer cuartil *Q1* corresponde a la mediana de la primera mitad de valores, el segundo cuartil *Q2* es la mediana general de la serie y que el tercer cuartil *Q3* corresponde a la mediana de la segunda mitad de valores.

Veamos el proceso de cálculo de los cuartiles en un ejemplo concreto: En una fábrica, la cantidad de lotes producidos por cada una de las 15 máquinas de producción ha sido anotado y se presenta en el siguiente listado:

10, 32, 19, 25, 15, 7, 17, 16, 24, 30, 12, 29, 27, 18, 27

Para hacer el cálculo de los cuartiles, se ordenan las observaciones en orden creciente, teniendo en cuenta que la cantidad de datos es *n = 15*:

7, 10, 12, **15**, 16, 17, 18, **19**, 24, 25, 27, **27**, 29, 30, 32

**A continuación se puede, bien usar la ecuación para calcular la posición de cada cuartil, o se puede iniciar** localizando la mediana *Q2* y luego calcular la posición de *Q1* y de *Q3* como medianas de la primera y segunda mitad de los datos.

Si se toma esta segunda opción, nótese que como la cantidad de datos es impar, la mediana corresponde al dato que está en el centro de la lista ordenada y que deja a cada lado la misma cantidad de datos, en este caso 7 a cada lado:

7, 10, 12, **15**, 16, 17, 18, **19**, 24, 25, 27, **27**, 29, 30, 32

entonces



**que corresponde a la mediana o segundo cuartil.**

**Para encontrar los cuartiles** *Q1* y de *Q3* se realiza el mismo procedimiento anterior, para los datos que quedan a cada lado de la mediana. Entonces, la primera mitad de los datos ordenados contiene 7 datos y la mediana de ese conjunto es el primer cuartil o *Q1*:

7, 10, 12, **15**, 16, 17, 18

Entonces



Por su parte la segunda mitad de los datos ordenados también contiene 7 datos y la mediana de ese conjunto es el tercer cuartil o *Q3*:

24, 25, 27, **27**, 29, 30, 32

Así que



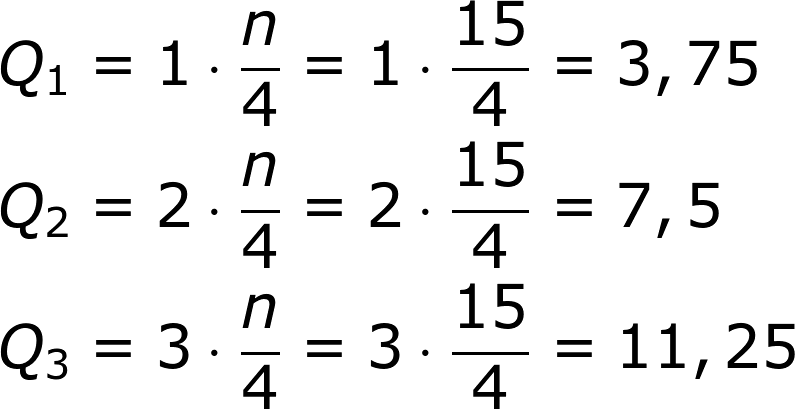
Entonces la distribución de los datos según las medidas de posición cuartílica para los 15 datos queda:

7, 10, 12, **15**, 16, 17, 18, **19**, 24, 25, 27, **27**, 29, 30, 32

*Q1 Q2 Q3*

Con lo que se concluye que la mitad de los datos está entre 15 y 27

En caso de aplicar la fórmula de los cuartiles en el mismo ejemplo, se obtiene:



Entonces el **primer cuartil** ***Q1* se ubica en la** posición 3,75, es decir, es el cuarto valor, que en este caso es 15. Ese valor corresponde a la mediana de las observaciones situadas a la **izquierda de la mediana de la totalidad**, que en este caso, es la mediana de las 7 primeras observaciones.

El segundo cuartil ***Q2*** ocupa la posición 7,5 de entre las 15 observaciones ordenadas; en este caso, es el número 19, que coincide con la que ya se había encontrado tomando el valor central del conjunto ordenado de datos.

Finalmente, el **tercer cuartil** ***Q3* se ubica en la** posición 11,25, es decir, es el doceavo valor de la lista ordenada, que en este caso es 27. Ese valor corresponde a la mediana de las observaciones situadas a la **derecha de la mediana de la totalidad**, que en este caso, es la mediana de las 7 últimas observaciones.

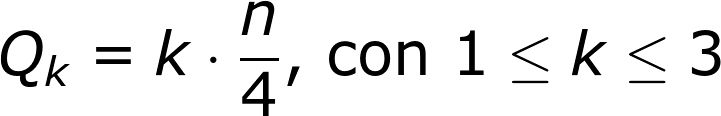
Con cualquiera de los dos métodos se obtienen los cuartiles:

7, 10, 12, **15**, 16, 17, 18, **19**, 24, 25, 27, **27**, 29, 30, 32

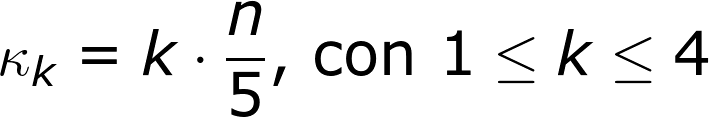
*Q1 Q2 Q3*

**Los quintiles, los deciles y los percentiles**

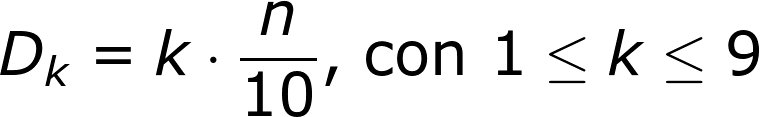
Los **cuartiles** son **3 valores** de la variable estadística que dividen los *n* datos en **4 partes iguales, que se representan con la letra Q y el subíndice correspondiente, que varía entre 1 y 3**. Los cuartiles dividen a la distribución en 4 partes, donde cada cuartil representa un 25 % de la muestra, es decir, una cuarta parte. Para calcular la posición en la lista ordenada de los datos que corresponden a cada cuartil se aplica la ecuación:



De manera análoga, los **quintiles** son los **4 valores** que dividen los *n* datos en **5 partes iguales y se representan con la letra Kappa y el subíndice correspondiente**. Los quintiles dividen a la distribución en 5 partes iguales, donde cada quintil representa un 20 % de la muestra, es decir, una quinta parte. El cálculo de la posición de los quintiles sobre la lista ordenada de datos es:



Por su parte, los **deciles** son los **9 valores** que dividen los *n* datos en **10 partes iguales. Análogamente a las medidas de posición previas se representan con una letra, en este caso la D, y el subíndice correspondiente, que varía entre 1 y 4**. Los deciles dividen a la distribución en 10 partes iguales, donde cada decil representa un 10 % de la muestra, es decir, una décima parte. El cálculo de la posición de los deciles sobre la lista ordenada de datos es:



Finalmente, se conocen como **percentiles** los **99 valores** que dividen los datos en 100 partes iguales. La representación en este caso usa la letra P y el subíndice correspondiente, que varía entre 1 y 99. Los percentiles dividen a la distribución en 100 partes iguales, donde cada percentil representa un 1 % de la muestra, es decir, una centésima parte. Para calcular la posición en la lista ordenada de los datos que corresponden a cada percentil se aplica la ecuación:



[SECCIÓN 2] **3.4 Aplicaciones de la estadística unidimensional**

En estadística la utilidad que tienen las medidas de tendencia central y no central y las medidas de posición se relacionan tanto con la necesidad de organizar y representar los datos, como con la de hacer inferencias y tomar decisiones apoyadas en los resultados obtenidos.

En la organización de los datos y en la presentación de su distribución conviene el apoyo de software, sobre todo si se trata de una gran cantidad de datos. Se usan medidas de variabilidad (rango, desviación estándar y varianza), medidas de centralidad (media aritmética, mediana, moda), medidas de no centralidad (cuartiles, quintiles, percentiles), diagramas (de cajas y bigotes), reconocimiento de valores atípicos, etc. Todo ello se hace para reconocer la distribución de los datos, bajo la comprensión de que lo importante de la representación elegida es la posibilidad de visualización de la variabilidad.

Una de las más importantes formas de presentación de la distribución de un conjunto de datos de una población es el diagrama de caja. Además, varios diagramas de caja provenientes de muestras aleatorias de una misma población permiten estudiar la variabilidad muestral y la variabilidad del azar.

Así, en el estudio estadístico no solo es importante hacer uso de lenguaje formal para expresar población, muestra, frecuencias, representatividad, variación y centralidad, sino para comprender los cambios que se dan al comparar diferentes datos provenientes de las muestras aleatorias

Entonces, las aplicaciones de la estadística unidimensional permite establecer el comportamiento de los datos de una población o de una muestra, comparar datos provenientes de diferentes muestras de una misma población o reconocer qué tanta variabilidad introduce el azar, para de ese modo poder realizar inferencias más formales generalizando el comportamiento de la muestra a la población, según las relaciones encontradas en el análisis estadístico.

**Los diagramas de cajas**

Uno de los principales usos de los cuartiles está en que sirven como marcas para hacer la representación de la distribución de los datos en diagramas de cajas que describen de manera condensada el centro y la dispersión del conjunto de datos.

Para crear un diagrama de caja de un conjunto de datos, se requiere de cinco datos esenciales:

* La **mediana o** *Q2*, para ubicar el valor central de las observaciones.
* Los **cuartiles** *Q1* y *Q3* para ubicar al 50% central del conjunto de datos. La diferencia entre la diferencia entre Q3 y *Q1* se llama rango intercuartílico y de denota como IQR.
* Las observaciones individuales **mínima (Mín) y máxima (Máx)** para indicar la dispersión.

Estos cinco números, resumen de una distribución, se expresan en un nuevo diagrama, conocido como **diagrama de caja**. Si uno de los datos se encuentra **a más de 1,5 IQR** de los extremos de la caja, se considera que es una **observación atípica** y se representa en el diagrama de caja individualmente con un punto. De ese modo se controla que los datos atípicos no “trasladen” la distribución y por ellos es de gran importancia hacer el cálculo del IQR.

Entonces, el parámetro de identificación de datos atípicos o IQR es:



y el cálculo de los valores atípicos se realiza calculando marcas atípicas inferiores y superiores. Para valores atípicos inferiores, la marca se calcula haciendo la diferencia entre *Q1* y 1,5 IQR es decir calculando:



Para valores atípicos superiores, la marca se calcula haciendo la suma entre *Q3* y 1,5 IQR es decir calculando:



Si un dato está por debajo de la marca de datos atípicos inferiores, o por encima de la marca de datos atípicos superiores, se considera un dato atípico.

El diagrama de caja se puede dibujar en posición horizontal o vertical, a condición de que el diagrama siempre incluya una escala. El diagrama de caja queda constituido al identificar si hay datos atípicos, y ubicar sobre un segmento escalado Mín, *Q1*, *Q2*, *Q3* y Máx. Luego se hace una caja alrededor de *Q1* y *Q3*, segmentos en los intervalos (Min, *Q1*) y (*Q3*, Máx) y se marcan como puntos aislados los datos atípicos. De ese modo se completa el diagrama.

Algunos ejemplos de diagramas de caja sin datos atípicos se observan en la imagen:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Representación de algunos diagramas de caja |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | http://www.estadisticaparatodos.es/taller/graficas/i/bigote3.gifTomada de http://www.estadisticaparatodos.es/taller/graficas/i/bigote3.gif  http://e-ducativa.catedu.es/44700165/aula/archivos/repositorio/2000/2006/html/caja_big1.png Tomada de http://e-ducativa.catedu.es/44700165/aula/archivos/repositorio//2000/2006/html/caja\_big1.png |
| **Pie de imagen** | Representación de algunos diagramas de caja |
| **Ubicación del pie de imagen** | Inferior o lateral |

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Los diagrama de caja** |
| **Contenido** | Los **diagramas de caja** son gráficos basados en **cuartiles** que son muy útiles **para la comparación de más de una distribución en un mismo gráfico**. Están compuestos por un rectángulo, la caja y dos brazos. Es un gráfico que suministra información sobre los valores **mínimo** y **máximo**, los **cuartiles** *Q1*, *Q2* o **mediana** y *Q3* y sobre la existencia de **valores atípicos** y **simetría de la distribución**. |

**Comparación de datos de la misma población**

Por ejemplo, una fábrica está realizando un estudio para identificar la productividad de su empresa en relación con los turnos de trabajo. Para ello, han tomado los datos de la producción de lotes en cada turno, durante los días hábiles que la empresa y cada turno funcionaron durante el mes anterior. Los datos han sido ordenados en orden creciente y son los siguientes:

Lotes que se han realizado en el turno de mañana de una fábrica:

23, 25, 26, 27, **28**, 3**2**, 34, 35, 37, **40**, **42**, 44, 50, 60, **75**, **85**, 87, 88, 88, 89

Lotes que se han realizado en el turno de noche:

7, 10, 12, **15**, 16, 17, 18, **19**, 24, 25, 27, **27**, 29, 30, 32

El cálculo de cuartiles para el turno de la mañana es el siguiente:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **VALOR**  **MÍNIMO** | **Q1** | **MEDIANA**  **Q2** | **Q3** | **VALOR**  **MÁXIMO** | **IQR**  **(Q3 - *Q1*)** |
| 23 | 30 | 41 | 80 | 89 | 50 |

que puede corroborarse en la lista ordenada de datos:

23, 25, 26, 27, **28**, **32**, 34, 35, 37, **40**, **42**, 44, 50, 60, **75**, **85**, 87, 88, 88, 89

*Q1 Q2 Q3*

De la misma manera, el cálculo de cuartiles para el turno de la mañana es el siguiente:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **VALOR MÍNIMO** | **Q1** | **MEDIANA**  **Q2** | **Q3** | **VALOR MÁXIMO** | **IQR**  **(Q3 - *Q1*)** |
| 7 | 15 | 19 | 27 | 32 | 12 |

que también puede corroborarse en la lista ordenada de datos:

7, 10, 12, **15**, 16, 17, 18, **19**, 24, 25, 27, **27**, 29, 30, 32

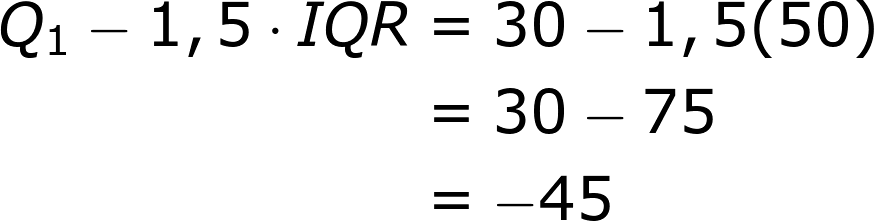
*Q1 Q2 Q3*

El IQR del turno de mañana es de 80 – 30 = 50 lotes producidos, mientras que el IQR del turno de noche: 27 – 15 = 12 lotes producidos.

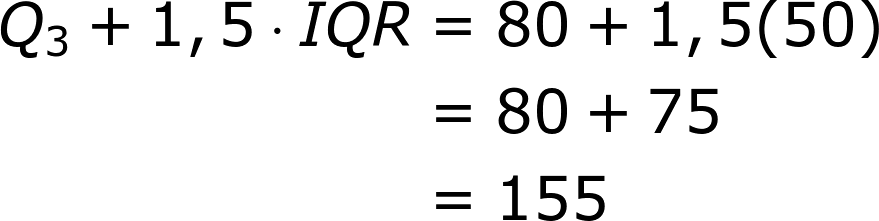
Para la construcción del diagrama de caja se dibuja un eje con la escala que se quiera utilizar. A continuación se construye un rectángulo o caja, cuyo lado izquierdo y derecho (o inferior y superior si la caja es vertical) van del primer al tercer cuartil. Por tanto, el ancho (o altura, si la caja es vertical) de la caja es la amplitud del 50 % de los datos centrales.

Debido a que si un punto se encuentra **a más de 1,5 IQR** de los extremos de la caja, se considera que es una **observación atípica**, se debe identificar si en el conjunto de datos algún dato es atípico.

Para identificar si hay valores atípicos en el turno de la mañana se calculan la marca inferior y la marca superior. La marca inferior corresponde a:

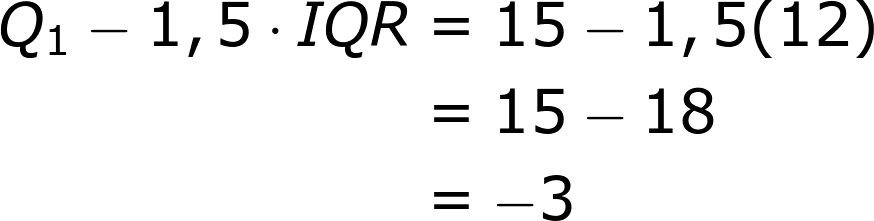


mientras que la marca inferior es:

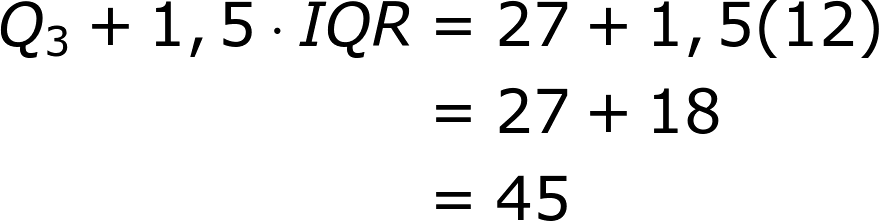


Así que como en el conjunto de datos para el turno de la mañana no hay valores inferiores a la marca atípica inferior, es decir a -45, ni datos superiores a la marca atípica superior, es decir a 155, no hay datos atípicos en el turno de la mañana.

Tampoco hay datos atípicos en el turno de la noche, donde las marcas atípicas inferior y superior son:



y



ya que en el conjunto de datos para el turno de la noche ningún dato es menor que -3 ni superior que 45.

Entonces, por último, se extienden segmentos perpendiculares a los lados derecho (superior) e izquierdo (inferior) hasta los valores máximo y mínimo respectivamente, que no sean observaciones atípicas. Los datos atípicos, de haberlos, se marcan como puntos.

Finalmente el diagrama queda:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/La estadística/Las medidas de posición/Los valores no centrales |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | 4º ESO/Matemáticas/La estadística/Las medidas de posición/Los valores no centrales    **OJO!!! Ordenaron mal los datos en el cuaderno de estudio presente en Aula Planeta y por eso algunas medidas cuartílicas son erróneas** |
| **Pie de imagen** | Diagramas de caja en un mismo gráfico para comparar el número de lotes realizados por el turno de mañana y el turno de noche. |
| **Ubicación del pie de imagen** | Inferior |

El análisis de la distribución permite concluir que en la noche hay menor producción de lotes, pero menor variación en los resultados de dicha producción.

Puedes encontrar más información teórica acerca de la estadística unidimensional en [VER](http://www.fuenterrebollo.com/Economicas2013/unidimensional-teoria.pdf) y algunos ejercicios resueltos en [VER](http://www.fuenterrebollo.com/Economicas2013/unidimensional-ejercicios.pdf).

[SECCIÓN 2] **3.5 Consolidación**

**[SECCIÓN 1] 4. La probabilidad**

La probabilidad es la disciplina matemática que estudia los experimentos o fenómenos aleatorios donde interviene el azar. Muchos procesos de la vida cotidiana, como procesos naturales biológicos, físicos, químicos y sociales (género, altura, peso, color del pelo, estado del clima, inundaciones, nacimientos en una ciudad, etc.), se comportan como juegos de azar, en el sentido de que aunque no se sabe cuál será el resultado, sí hay unos resultados posibles y delimitados entre los cuáles puede estar el resultado. El sorteo en estos casos ya está realizado a través de mecanismos complejos, muchas veces desconocidos. La probabilidad se encarga de recoger los resultados y analizarlos de forma que permitan predecir y tomar decisiones futuras.

**[SECCIÓN 2] 4.1 Los espacios muestrales**

La probabilidad permite cuantificar los resultados de un **experimento aleatorio**. Se denomina experimento a cualquier método de recogida de datos que puede, o no, producir un valor numérico.

Un experimento, si se repite bajo idénticas circunstancias, puede ser de dos tipos:

* **Determinista**: cuando siempre se obtiene el mismo resultado, es decir cuando el resultado es completamente predecible. Por ejemplo, si se quiere ejecutar el experimento de “extraer una papeleta de una urna llena de papeletas rectangulares y observar su forma”, es completamente obvio que siempre se sacará una papeleta de forma rectangular, lo que significa que el resultado es predecible y siempre será el mismo: una papeleta rectangular.
* **Aleatorio**: cuando no se obtiene siempre el mismo resultado, es decir que el resultado es impredecible. Por ejemplo, si se quiere ejecutar el experimento de “extraer una papeleta de una urna llena de papeletas circulares, triangulares, cuadradas y rectangulares y observar su forma”, es claro que no se puede asegurar la forma de una papeleta elegida al azar. Ello significa que el resultado es impredecible y no siempre será el mismo.

Los fenómenos o experimentos aleatorios tienen en común las siguientes características:

* **Se conoce el conjunto de todos los resultados posibles, pero no el que se obtendrá al realizar el experimento**. Por ejemplo, del experimento de “extraer una bola de una bolsa que contiene 7 bolas numeradas del 1 al 7”, sabemos que saldrá una bola con el número 1, 2, 3, 4, 5, 6 ó 7, pero no cuál de ellas. Se denomina **espacio muestral** al conjunto de posibles resultados de un experimento aleatorio y se representa con la letra Ω**.** En este caso Ω = {1, 2, 3, 4, 5, 6, 7}.
* **Se puede repetir tantas veces como se quiera en condiciones casi idénticas**. Para el experimento de selección de las bolas enumeradas, si cada vez que se saca una bola, se anota el resultado y se devuelve la bola a la bolsa, la repetición sucesiva del experimento no asegura que vuelva a salir la misma bola aunque se proceda de forma muy parecida en todas las repeticiones.
* **Cualquier modificación de las condiciones iniciales de la repetición puede alterar el resultado**. Si en el experimento de extracción de las bolas enumeradas no se devuelve la bola extraída, o se añaden dos bolas con el mismo número al sacar una o se lanza un dado para definir si el resultado se anota o no o cualquier otra modificación de las condiciones iniciales del experimento genera un nuevo experimento en el que la probabilidad de un resultado difiere de la probabilidad del experimento inicial.
* Si el experimento se repite un gran número de veces, entonces aparece algún modelo de regularidad estadística en los resultados obtenidos. En el caso en que se repita la extracción de una de las 7 bolas una gran cantidad de veces, se equilibrarán los resultados y las repeticiones de los resultados será equivalente.

Los resultados posibles de un experimento aleatorio son todos aquellos que se pueden observar en la realización del experimento. Se denomina **espacio muestral** al conjunto de posibles resultados de un experimento aleatorio y se representa con la letra Ω**.** Por ejemplo en el experimento aleatorio de “sacar una bola de una urna con bolas negras y blancas, y observar su color”, el espacio muestral tiene dos elementos: Ω = {negra, blanca}.

Por su parte cualquier hecho o resultado que se pretenda estudiar en un experimento aleatorio se denomina **suceso** y se representa con una letra mayúscula. Así, un suceso es cualquier subconjunto del espacio muestral de un experimento aleatorio. Por ejemplo el suceso “A: La bola extraída es blanca” es uno de los sucesos para el experimento de “sacar una bola de una urna con bolas negras y blancas, y observar su color”

Decimos que un determinado suceso ha tenido lugar (es decir, se verifica) si el resultado de la experiencia aleatoria ha sido alguno de los elementos de ese suceso. Por ejemplo, si se hace el experimento de “hacer girar una ruleta dividida en 6 sectores idénticos numerados del 1 al 6”, y ha salido un 3 al hacer girar la flecha, podemos asegurar que ha tenido lugar el suceso “obtener un número mayor que 2”.

Los tipos de sucesos que pueden darse para un experimento aleatorio son los siguientes:

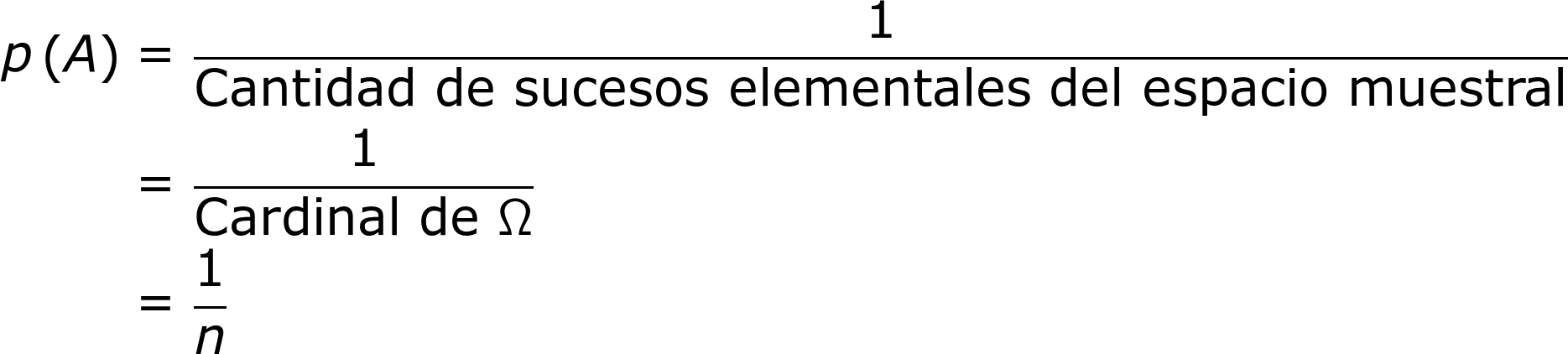
* **Los sucesos elementales o simples**, que son los formados por un único resultado del espacio muestral. Así, en el ejemplo de la ruleta, los sucesos simples son: A: que caiga en 1, B: que caiga en 2, C: que caiga en 3, D: que caiga en 4, E: que caiga en 5 y F: que caiga en 6. Tales sucesos también se pueden denotar como {1}, {2}, {3}, {4}, {5}, {6} si se interpreta que corresponden con los números de cada uno de los sectores en que está dividida la ruleta. Podría definirse el suceso P: que caiga en un número par o I: que caiga en un número impar, pero tales sucesos no son simples, sino compuestos. En el experimento “lanzar una moneda al aire dos veces consecutivas”, los sucesos simples son {(C, C)}, {(C, X)}, {(X, C)}, {(X, X)}, donde C = cara y X = cruz. En el experimento de “lanzar 3 monedas al aire”, los sucesos simples son:{(C, C, C)}, {(C, C, X)}, {(C, X, X)}, {(X, X, C)}, {(X, C, C)}, {(C, X, C)}, {(X, C, X)}, {(X, X, X)}
* **Los sucesos compuestos**, son aquellos formados por dos o más sucesos elementales. Por ejemplo, un suceso compuesto del experimento aleatorio de la ruleta descrita sería “A: obtener un número impar: {1, 3, 5}”.Un suceso del experimento de “lanzar una moneda al aire dos veces consecutivas” sería el suceso “A: obtener al menos una cruz: {(C, X), (X, C), (X, X)}”.
* **Los sucesos imposibles**, son aquellos que nunca ocurrirán. Se denotan con el símbolo de conjunto vacío (∅). Por ejemplo para el experimento de lanzar dos veces consecutivas una moneda un suceso imposible es que caiga el número 1, así como es imposible que en el giro de la ruleta descrita se obtenga X. Son imposibles porque el resultado no hace parte del espacio muestral
* **Los sucesos seguros** son aquellos sucesos que están formados por todos los resultados posibles del experimento. Ocurren siempre y coinciden con el espacio muestral.

Además, dos sucesos cualesquiera pueden ser compatibles o incompatibles:

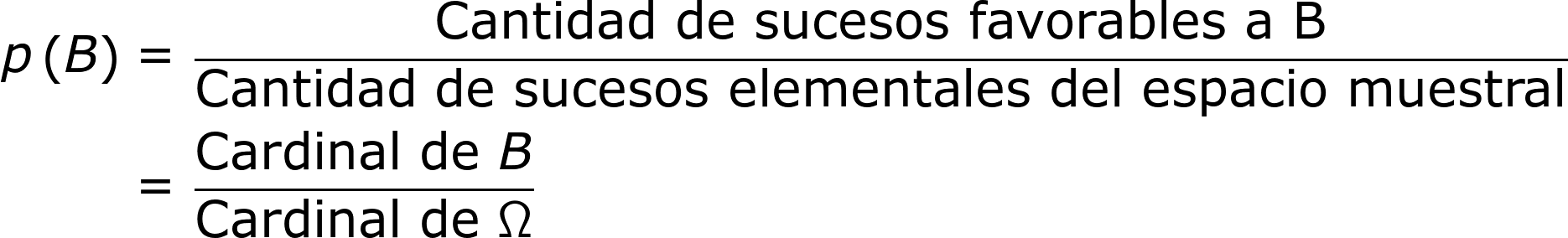
* **Los sucesos compatibles**, son aquellos que se pueden dar simultáneamente. Por ejemplo en el experimento aleatorio de seleccionar dos veces consecutivas una bola de una bolsa que contiene bolas blancas, negras, amarillas , azules y rojas, tenemos el suceso A: {blanco, negro} y el suceso B: {blanco, rojo}. Observamos que “blanco” es un caso favorable a los dos sucesos.
* **Los sucesos incompatibles**, son aquellos que no se pueden dar simultáneamente. Su intersección es igual al conjunto vacío (∅), es decir, no tienen elementos comunes. Por ejemplo, para el experimento descrito el suceso {blanco, negro} es incompatible con el suceso {azul, rojo}; por este motivo, los sucesos elementales son siempre incompatibles.

**[SECCIÓN 3] 4.1.1 Probabilidad y conteo**

La probabilidad de un suceso es el número que representa la proporción de veces que podemos esperar que un suceso se verifique cuando el experimento es repetido muchas veces en idénticas condiciones. Uno de los resultados más importantes de la probabilidad es el Teorema de Laplace, según el cual la probabilidad de ocurrencia de un suceso corresponde a la razón entre la cantidad de opciones favorables al suceso, sobre la cantidad de opciones posibles. Por ejemplo, si un suceso A es elemental o simple, la probabilidad de que ocurra es:



Por su parte, si B es un suceso de un experimento aleatorio en que todos los casos posibles tienen la misma probabilidad, se cumple que la probabilidad de dicho suceso, sea simple o compuesto es:



Fíjate en la siguiente ruleta:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Probabilidad/La probabilidad de un suceso |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) |  |
| Pie de imagen | Ruleta de un juego de azar |
| Ubicación del pie de imagen | Inferior |

Al jugar a la ruleta podemos ejemplificar la probabilidad de diferentes sucesos. Por ejemplo

¿cuál es la probabilidad de cada suceso elemental?, ¿cuál es la probabilidad de obtener el suceso A = {rojo, amarillo} al hacer girar la ruleta?

Las probabilidades para los sucesos elementales serán las siguientes:

p({rojo}) = p({azul}) = p({rosa oscuro}) = p({rosa claro}) = p({morado})= p({mostaza}) = 1/8.

p({gris}) = 1/4.

p({rojo, amarillo}) = 1/8 + 1/8 = 2/8 = 1/4.

Sin embargo, no en todas las ocasiones ocurre que el espacio muestral tenga tan pocos elementos, o que las probabilidades de los sucesos simples sean iguales.

**CONTEO O COMBINATORIA**

Para realizar al cálculo de probabilidades se requiere conocer el tamaño del espacio muestral del experimento aleatorio, lo cual se logra mediante un proceso que en matemáticas se denomina conteo o combinatoria. Antes de contar el número de posibilidades, deberemos tener en cuenta lo siguiente:

* Si importa o no el **orden** en que se pondrán los elementos que hacen parte del espacio muestral, en cuyo caso tendremos:

Muestras ordenadas.

Muestras no ordenadas.

* Si se pueden o no **repetir** los elementos que hacen parte del espacio muestral, lo cual generará conteos:

Con repetición.

Sin repetición.

En la mayoría de problemas de combinatoria, solo hará falta esta observación para conocer el camino a seguir. Veamos algunos ejemplos:

En una clase se quiere elegir dos representantes para un concurso. Como en ocasiones un estudiante puede faltar, se elegirá el que concursará y un suplente. En una primera votación, salen elegidos tres candidatos: María, Laura y Pedro. En una segunda votación, se elige a uno de entre los tres candidatos, escribiendo en la papeleta un nombre para el concursante y otro para el suplente. ¿Cuántas papeletas diferentes pueden salir en esta segunda votación?

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Combinatoria/¿Qué es la combinatoria? |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img1_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img1_zoom.jpg) |
| Pie de imagen | Representación que permite averiguar cuántas papeletas diferentes pueden salir en la elección de un par de representantes de un curso. |
| Ubicación del pie de imagen | Inferior |

Para representar todas las opciones posibles, se utiliza un **diagrama de árbol**, que es una representación gráfica muy útil para contar todas las posibles maneras de combinar una cantidad finita de elementos de uno o de varios conjuntos.

Para el primer nombre, como concursante, existen tres posibilidades:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Combinatoria/¿Qué es la combinatoria? |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img2_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img2_zoom.jpg) |
| Pie de imagen | Representación mediante **diagrama de árbol** de los nombres de los **posibles concursantes**. |
| Ubicación del pie de imagen | Inferior |

Una vez escrito el primer nombre, solo nos quedan dos posibilidades para el segundo nombre, que será el suplente:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Combinatoria/ ¿Qué es la combinatoria? |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img3_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img3_zoom.jpg) |
| Pie de imagen | Representación mediante **diagrama de árbol** que muestra **todas las posibilidades** que tienen los alumnos de salir elegidos como suplentes. |
| Ubicación del pie de imagen | Inferior |

Así, en total hay 3 × 2 = 6 posibilidades

|  |  |
| --- | --- |
| **Recuerda** | |
| Contenido | El **factorial** de un número **entero positivo** es el producto de todos los números naturales anteriores o iguales a él. Se escribe **n!**, se lee “**n** f**actorial**” y para n > 0 se define como:  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula1_resized.gif  El factorial de 0 se define como:  0! = 1 |

**Muestras ordenadas**

Las **muestras ordenadas** se **clasifican** dependiendo de si intervienen o no todos los elementos del conjunto inicial en la muestra. Según si intervienen o no todos los elementos del conjunto, las muestras ordenadas se clasifican así:

* Muestra ordenada en la que no intervienen todos los elementos. Se conoce como **variación**.
* Muestra ordenada en la que intervienen todos los elementos. Se conoce como **permutación.**

Tanto para las variaciones como las permutaciones las apariciones de los elementos del conjunto pueden ser **con repetición** o **sin repetición de elementos**.

**Las variaciones**

Las **variaciones** son formas de **agrupar algunos de los elementos** de un conjunto en las cuales **aunque no todos** los elementos del conjunto intervienen en cada muestra, sí es importante el orden en que se ubican tales elementos. El cálculo de la cantidad de variaciones posibles en un conjunto depende de si las variaciones se hacen sin repetición o con repetición.

**Variaciones sin repetición**

Las **variaciones sin repetición** son las distintas **agrupaciones ordenadas** que se pueden formar tomando ***n*** elementos distintos de un conjunto de ***m*** elementos, cuando el orden es significativo y **no se pueden repetir los elementos**. Su fórmula es:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula2_resized.gif

Fíjate en lo siguiente:

* **El primer factor es igual al número de elementos del conjunto inicial**.
* Los factores van disminuyendo de uno en uno.
* **Hay tantos factores como elementos tiene una muestra**.

Veamos un ejemplo: En una votación salen elegidos cuatro candidatos para representar al instituto en un concurso: María, Laura, Carlos y Pedro. En una segunda votación, se elige entre ellos quién será el concursante, quién el primer suplente y quién el segundo suplente, escribiendo los nombres según este orden en la papeleta. ¿Cuántas papeletas diferentes pueden salir?

En este caso se requiere elegir 3 personas de entre un conjunto de 4 personas, en que el orden es importante. Los tres nombres que aparecen en las papeletas se llaman **muestras ordenadas** de un conjunto con cuatro elementos, que son los candidatos. O lo que se conoce también como **variaciones sin repetición** de tres nombres (o ***n***) escogidos de entre un total de cuatro (o ***m***).

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Combinatoria/ Muestras ordenadas |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img4_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img4_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_11\_img4\_zoom.jpg |
| Pie de imagen | Representación que permite averiguar cuántas papeletas diferentes pueden salir en la elección de tres representantes de un curso, en un grupo de cuatro. |
| Ubicación del pie de imagen | Inferior |

Así, de un conjunto de cuatro nombres, ¿cuántas variaciones o **agrupaciones ordenadas** de tres nombres, **sin repetición**, pueden salir?

Para el primer nombre, correspondiente al concursante, tenemos cuatro posibilidades. Una vez elegido el concursante, solo nos quedarán tres nombres para elegir al primer sustituto. Una vez elegido el concursante y el primer sustituto, solo nos quedarán dos nombres para elegir al segundo sustituto.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Combinatoria/ Muestras ordenadas |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img5_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img5_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_11\_img5\_zoom.jpg |
| Pie de imagen | Representación mediante diagrama **de árbol** que muestra **todas las variaciones posibles.** |
| Ubicación del pie de imagen | Inferior |

Fíjate que en este caso **el orden de los elementos es importante**. No es lo mismo una papeleta que ponga María, Laura, Carlos que otra que ponga Laura, María, Carlos. Aunque son las mismas personas, en la primera papeleta la concursante sería María y en la otra papeleta, Laura. Entonces, Como muestra el diagrama de árbol, el número total de agrupaciones posibles es 4 × 3 × 2 = 24

Veamos un segundo ejemplo: si tenemos tiras de tela de cinco colores diferentes, ¿cuántas banderas distintas de tres franjas podemos construir si es importante el orden de los colores y no se pueden repetir colores?

En este caso se trata de una variación sin repetición de tres colores (***n*)** escogidos de entre un total de cinco colores (***m*)***.*

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Combinatoria/ Muestras ordenadas |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img6_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img6_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_11\_img6\_zoom.jpg |
| Pie de imagen | Representación que permite averiguar cuántas banderas de tres colores **sin repetición** se pueden hacer si los colores son escogidos de entre un conjunto de cinco colores. |
| Ubicación del pie de imagen | Inferior |

Realizando el mismo análisis anterior, deducimos que para el primer color tenemos cinco posibilidades. Una vez elegido ese primer color, solo nos quedarán cuatro colores más para elegir el segundo color. Una vez elegidos el primer y el segundo color, nos quedarán tres colores para elegir el tercer color de la bandera. Por lo tanto el número total de variaciones es 5 × 4 × 3 = 60 banderas

Si se aplica la fórmula de **variaciones sin repetición se obtiene**:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula3_resized.gif

Fíjate en lo siguiente:

* **El primer factor es igual al número de elementos del conjunto inicial (cinco)**.
* Los otros van disminuyendo de uno en uno.
* **Hay tantos factores como elementos tiene una muestra (tres)**.

**Variaciones con repetición**

Las variaciones con repetición son las distintas agrupaciones que se pueden formar tomando *n* elementos iguales o distintos de un conjunto de *m* elementos. Los elementos que se obtienen reciben el nombre de **muestras ordenadas con repetición** o **variaciones con repetición**. Si **se repiten los elementos**, podemos hacerlo hasta tantas veces como elementos tenga la agrupación. Su ecuación es:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula4_resized.gif

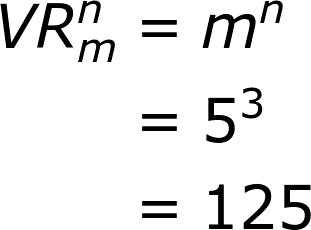
Fíjate en lo siguiente:

* El resultado es una potencia, donde la base es el número de elementos del conjunto inicial y el exponente es el número de elementos de la muestra.

Para entender en qué consisten las variaciones con repetición, veamos el siguiente ejemplo: Para saber cuántos boletos crear para una rifa, los organizadores se preguntan ¿cuántos números de tres cifras podemos escribir utilizando los números 1, 2, 3, 4, 5?

En primer lugar, algunos números posibles serían: 135, 341, 555, 111, etc. En este caso todas las cifras (primera, segunda y tercera) tendrán siempre cinco posibilidades, ya que escribir un 1, por ejemplo, en las centenas, no significa que no se pueda volver a escribir en las decenas e, incluso, en las unidades.

Así, como se pueden repetir los cinco números para formar los números de tres cifras, donde el orden importa (ya que no es el mismo número el 351 que el 153 o que el 513), por este motivo, el total de posibilidades será 5 × 5 × 5 = 125. Al aplicar la ecuación con *n = 3* y *m = 5* se obtiene:

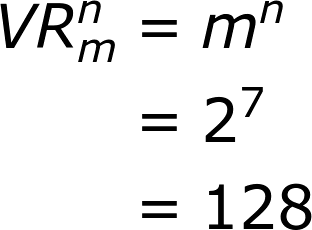


Para un segundo ejemplo, se lanza una moneda al aire siete veces seguidas y se escribe por orden los resultados, ¿cuántos resultados distintos puede haber?

En primer es claro que se trata de muestras ordenadas con repetición, pues en los siete lanzamientos se repetirán las opciones “cara” o “sello”. Además, no es lo mismo que caigan 4 caras seguidas y 3 sellos seguidos, que esas mismas 4 caras y 3 sellos intercalados cara-sello-cara....

El razonamiento de conteo sería del siguiente estilo: en el primer lanzamiento hay dos posibilidades: cara o sello. Sea cual sea el resultado de ese primer lanzamiento, también hay dos posibilidades para el segundo: cara o sello. Análogamente en el tercer, cuarto, quinto, sexto y séptimo lanzamiento, se mantiene el hecho de que aparecen dos posibilidades. Entonces en total hay 2 × 2 × 2 × 2 × 2 × 2 × 2 = 27 = 128 posibilidades. Así que en total se obtienen 128 resultados posibles después de haber lanzado siete veces seguidas una moneda.

El resultado se corrobora si se aplica la ecuación de **variaciones con repetición**:

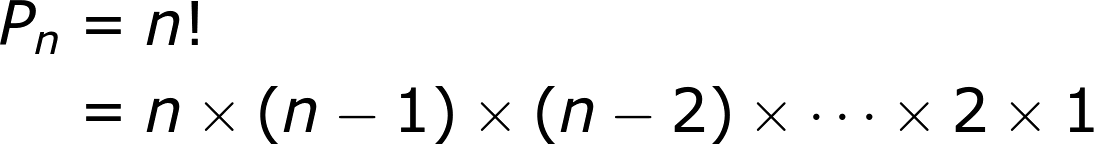


**4.3 Las permutaciones**

Las permutaciones son las distintas formas de agrupar los elementos de un conjunto tomando simultáneamente todos los elementos del conjunto. Para conformar tales agrupaciones influye el orden en que se presentan los datos. Las permutaciones pueden realizarse sin repetición y con repetición.

**Permutaciones sin repetición**

Las permutaciones sin repetición son las distintas formas de ordenar un grupo de *n* elementos distintos, por lo que la única diferencia entre ellas es el orden de ubicación de sus elementos. Pueden verse también como variaciones sin repetición en las que *n = m*. Para calcular la cantidad de posibles permutaciones sin repetición se aplica la ecuación:



Fíjate en que:

* + El número de permutaciones sin repetición es igual a un producto de números naturales consecutivos.
  + El primer factor es el número de elementos de la permutación.
  + Los otros factores van decreciendo de uno en uno hasta llegar al número 1.
  + Se trata de calcular el factorial de *n*, siendo *n* el número de elementos del conjunto.

Veamos el siguiente ejemplo: De cuántas formas se pueden ordenar en una estantería los siguientes cinco cómics diferentes:

Uno de Superman.

Uno de Spiderman.

Uno de El capitán Trueno.

Uno de Los 4 fantásticos.

Uno de El increíble Hulk.

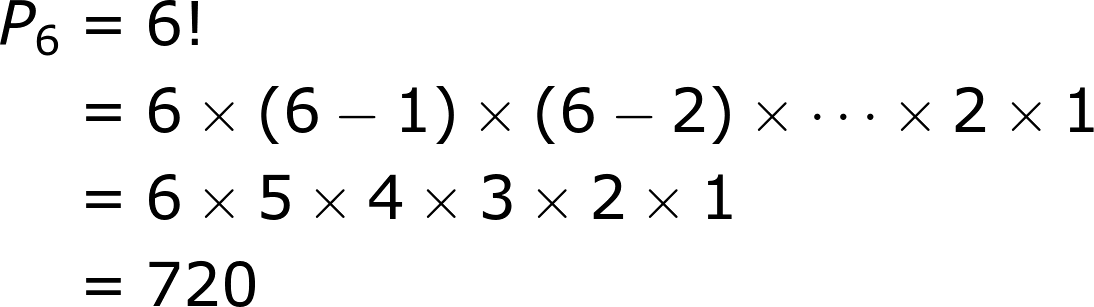
Los cinco cómics son el **conjunto inicial** y coinciden con la **muestra** de cinco elementos, en este caso cómics que queremos ordenar. En el ejemplo es claro que **en una misma muestra no se pueden repetir** dos elementos, pues eso equivaldría a poner dos veces el mismo cómic.

El razonamiento para calcular cuántas distintas permutaciones hay puede ser como el siguiente: para la **primera** posición hay **cinco posibilidades, es decir que se** puede escoger de entre todos los cómics cuál poner primero. Hecha esa elección, para la **segunda** posición quedan **cuatro posibilidades** y para la tercera posición quedan tan solo **tres posibilidades. Finalmente para cuarta** y quinta posición quedarán respectivamente 2 y 1 cómic. En este punto no quedan cómics sin ordenar y se terminan las opciones. Entonces, en total se tienen 5! = 5 × 4 × 3 × 2 × 1 = 5 × **(5 − 1)** × **(5 − 2)** × **(5 − 3)** × **(5 − 4)** = 120 posibilidades.

Como un segundo ejemplo veamos el siguiente: ¿cuántas permutaciones o anagramas se pueden hacer con las letras de la palabra “fútbol”?

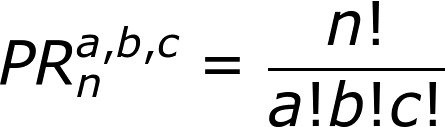
Para responder a la pregunta, se nota que se trata de ordenar un total de seis letras distintas, sin repetir ninguna, de todas las formas posibles. Así que el número de posibilidades para la **primera** posición es igual a **seis pues se puede** escoger entre todas las letras F, Ú, T, B, O y L cuál de ellas irá en la primera posición. Consecutivamente las posibilidades para la **segunda** posición son **cinco, pa**ra la **tercera** posición **cuatro, p**ara la **cuarta** posición **tres, p**ara la **quinta dos y para la sexta una** Entonces, en total se tienen 6! = 6 × 5 × 4 × 3 × 2 × 1 = 6 × **(6 − 1)** × **(6 − 2)** × **(6 − 3)** × **(6 − 4)** × **(6 − 5)** = 720 posibilidades.

Si se aplica la ecuación para el **cálculo del total**  de **permutaciones sin repetición para un conjunto de seis elementos se** llega al mismo resultado:



**Permutaciones con repetición**

Las permutaciones con repetición de *n* elementos son las agrupaciones en las que el primer elemento se repite *a* veces, el segundo se repite *b* veces, el tercero c veces, etc., de tal forma que en cada grupo formado por los *n* elementos cada elemento aparece el número de veces indicado. La ecuación que permite calcular la cantidad de resultados que se obtienen al permutar de esta manera es:



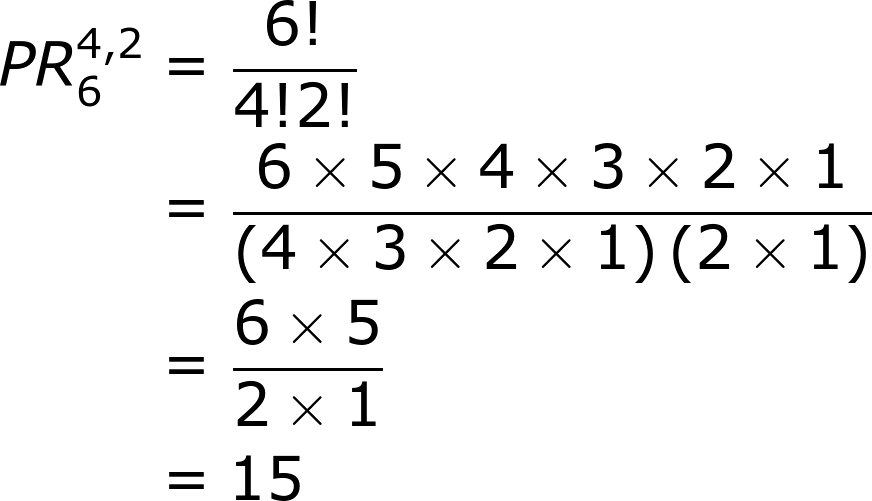
Para entender en qué consisten las permutaciones con repetición, veamos un ejemplo. El código secreto de seis cifras para abrir una caja fuerte, está formado por cuatro unos y dos cincos (1, 1, 1, 1, 5, 5), pero no recordamos el orden correcto. ¿Cuántas posibilidades tenemos para probar?

Se trata de construir todos los códigos de seis cifras posibles, utilizando cuatro veces el número 1 y dos veces el número 5. La lista extensa aparece en la siguiente imagen:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Combinatoria/ Muestras ordenadas |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img7_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img7_zoom.jpg)  <http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img7_zoom.jpg> |
| **Pie de imagen** | **Diagrama de** todos los **posibles códigos de seis cifras formados por cuatro unos y dos cincos.** |
| **Ubicación del pie de imagen** | Inferior o lateral |

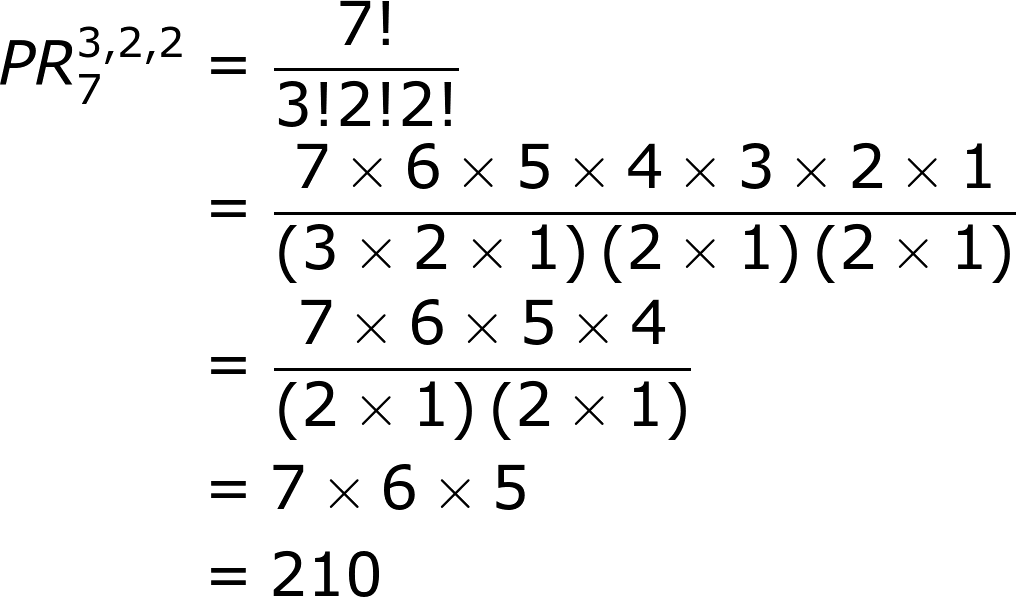
En el diagrama se presentan todas las posibles **permutaciones con repetición** de seis elementos en las que el primer elemento se repite cuatro veces y el segundo dos veces. En total resultan 15 posibles códigos de forma que, como máximo se tendrá que probar 15 veces para poder abrir la caja fuerte.

El mismo resultado se obtiene si se aplica la ecuación para calcular la cantidad **de permutaciones con repetición, donde** *n = 6*, *a = 4* y *b = 2*:



Como segundo ejemplo veamos el siguiente: si tenemos tres camisetas, dos pantalones y dos chaquetas, ¿de cuántas maneras distintas se puede ordenar esa ropa en el armario?

En este caso se trata de ordenar un total de 7 elementos de forma que aparezcan tres camisetas, dos pantalones y dos chaquetas, contando con que **la única diferencia** entre las distintas agrupaciones es **el orden** de ubicación. Entonces de debe calcular la cantidad de permutaciones con repetición de siete elementos en las que el primer elemento se repite tres veces, el segundo dos veces y el tercero dos veces. En este caso entonces *n = 7*, *a = 3*, *b = 2* y *c = 2*. Al aplicar la ecuación se obtiene:



Por lo tanto hay 210 maneras de disponer la ropa en el armario.

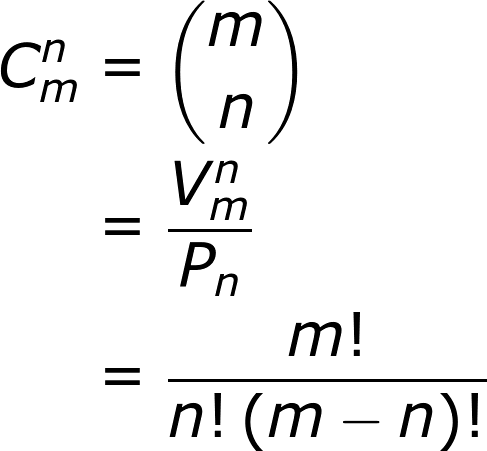
**Muestras no ordenadas**

En ocasiones **no interesa el orden** en que se van poniendo los elementos de una muestra de un conjunto. En el caso en que eso ocurre, el nombre que se da al conteo de las posibles formas de seleccionar segmentos o la totalidad de los elementos de un conjunto se denomina ***combinaciones***.

Como en el caso de las muestras ordenadas, las combinaciones se pueden construir **sin** o **con repetición de los elementos del conjunto**.

**Las combinaciones sin repetición**

Las **combinaciones sin repetición** son las **agrupaciones** que se pueden formar tomando ***n*** elementos de un conjunto total de ***m*** elementos distintos, en las que **el orden no es significativo**. La ecuación que permite calcular la cantidad de resultados que se obtienen al combinar sin repetición tales muestras es:



El número resultante se conoce también como **coeficiente binomial** o **número combinatorio**, porque expresa los coeficientes que aparecen cuando se expande un binomio, que son exactamente los números del triángulo de Pascal.

Veamos el funcionamiento de las combinaciones sin repetición mediante un ejemplo: Seis amigos (Luis, Ruth, Ana, Juan, Pepe y Mar) participaron en un concurso de baile y ganaron como premio dos entradas para ir a un concierto. Ahora, deben decidir cuáles de los seis irán al concierto. ¿Cuántas posibles parejas se pueden conformar para hacer uso del premio?

**Un razonamiento inicial es que el orden no importa** porque da igual conseguir la primera entrada que la segunda. Del conjunto de seis elementos, preparamos todas las muestras ordenadas posibles de dos elementos. Para ello, utilizamos las iniciales de los nombres para representar a cada uno de los amigos, o los enumeramos o asignamos letras del alfabeto a cada uno, para identificar claramente los elementos del conjunto. En este caso las iniciales de los nombres son todas distintas, entonces el diagrama de árbol que resume las opciones es:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Combinatoria/ Muestras no ordenadas |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img8_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img8_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_11\_img8\_zoom.jpg |
| **Pie de imagen** | Diagrama de árbol y tabla resumen de las combinaciones sin repetición |
| **Ubicación del pie de imagen** | Inferior |

De las 6 × 5 = 30 muestras ordenadas que ilustra el diagrama de árbol, vemos que hay muestras que contienen a las mismas personas, aunque en distinto orden. En la tabla se han indicado poniéndolas en la misma columna. Las muestras no ordenadas o subconjuntos de dos elementos escogidos de entre seis son entonces solo la mitad del total que nos han resultado, es decir, **15**.

En el ejemplo se observa que el resultado es el cociente entre:

* La cantidad de muestras no ordenadas o subconjuntos de ***n*** elementos escogidos de entre ***m*** de un conjunto, corresponde a Vmn o total de muestras ordenadas, es decir son las **variaciones sin repetición** de ***n*** elementos escogidos entre un total de ***m***.
* El total de permutaciones sin repetición de los *n* elementos, que es Pn.

En nuestro ejemplo, Vmn = V62 serían las variaciones sin repetición de dos elementos entre seis:

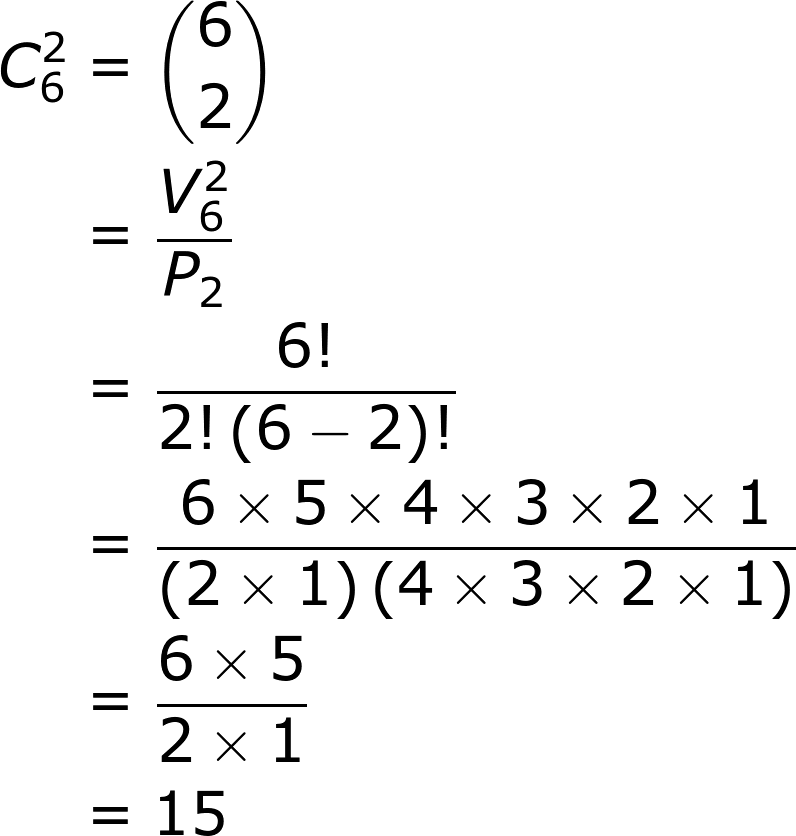
http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula11_resized.gif

y *Pn = P2 = 2!* es el total de **permutaciones sin repetición de los** **2** **elementos** de cada muestra.

Finalmente se realiza el cociente de estos dos resultados, con lo que se obtiene la cantidad de muestras no ordenadas o subconjuntos de 2 elementos del conjunto de 6 elementos, que son:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula12_resized.gif

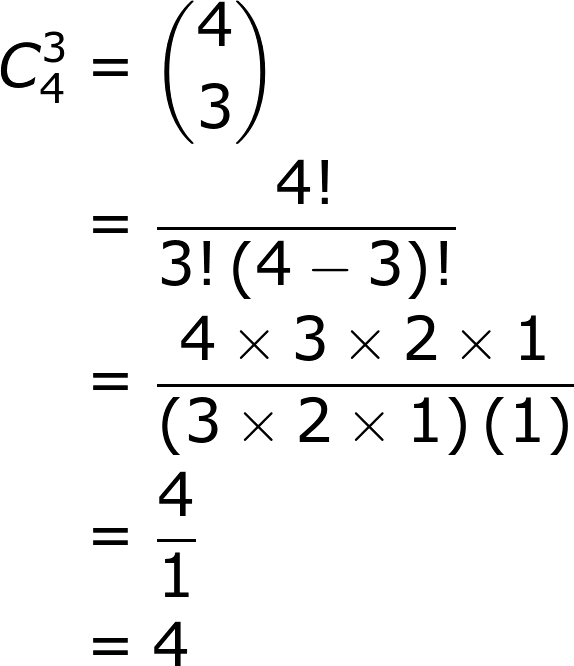
que es el mismo resultado que el obtenido para indicar el número de subconjuntos de dos elementos de un conjunto de seis en total, aplicando el ***coeficiente binomial*** o ***número combinatorio***, que también expresa el número de combinaciones sin repetición. Aplicando la ecuación que permite calcular la cantidad de resultados que se obtienen al combinar sin repetición se obtiene:



Otro ejemplo: ¿cuántos triángulos se pueden dibujar utilizando como vértices de los triángulos los del cuadrilátero dado?

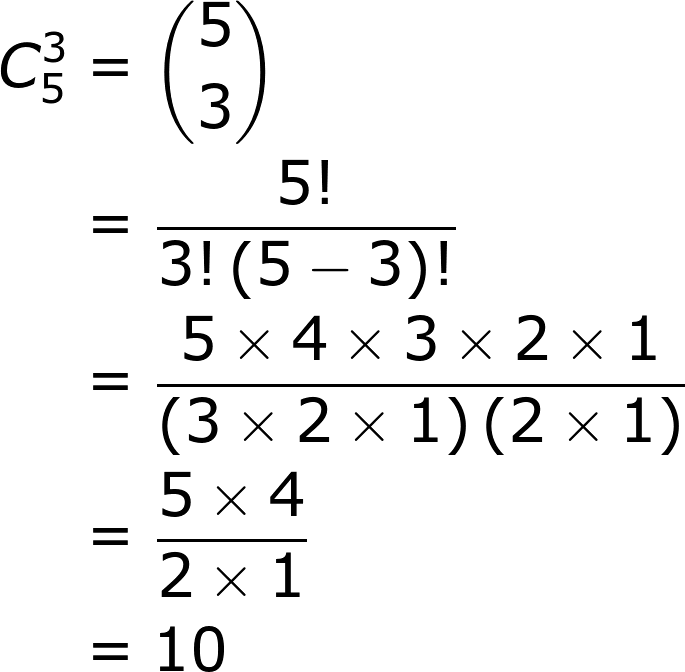
|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Combinatoria/ Muestras no ordenadas |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img9_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img9_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_11\_img9\_zoom.jpg |
| **Pie de imagen** | Cada triángulo es una muestra no ordenada o subconjunto de tres elementos (vértices del triángulo) escogidos entre un total de cuatro (vértices del cuadrilátero). |
| **Ubicación del pie de imagen** | Inferior |

Se trata de **muestras no ordenadas** de tres elementos escogidos entre un total de cuatro, por lo que para averiguar el número de triángulos posibles se aplica la ecuación con *n = 3* y *m = 4*:



El anterior resultado significa que es posible dibujar máximo cuatro triángulos utilizando los vértices de un cuadrilátero.

¿Y si se utilizaran los vértices de un pentágono? En ese caso se trataría de una muestra no ordenada de tres elementos, escogidos entre un total de cinco, para los que la cantidad de combinaciones posibles sería:

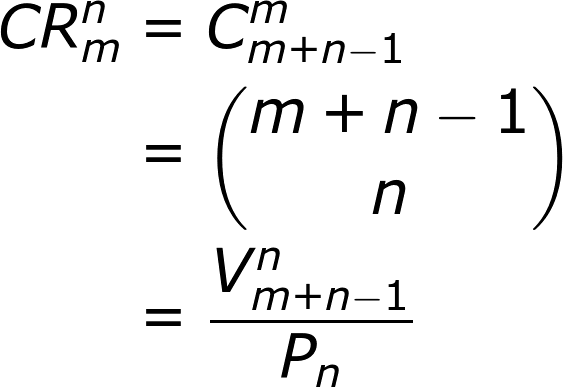


Así que la máxima cantidad posible de triángulos que se pueden dibujar utilizando los vértices de un pentágono son 10 triángulos.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El triángulo de Pascal** |
| **Contenido** | Con los números combinatorios podemos formar el llamado triángulo de Pascal, que se usa para escribir los valores de los números combinatorios o coeficientes binomiales, encontrar los números triangulares, fragmentar las potencias de dos y otros cálculos matemáticos sin necesidad de hacer ningún cálculo.   |  |  | | --- | --- | | **Imagen (fotografía, gráfica o ilustración)** | | | **Código** |  | | **Descripción** | 4º ESO/Matemáticas/Combinatoria/ Muestras no ordenadas | | **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img11_small.jpg  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_11\_img11\_zoom.jpg | | **Pie de imagen** | Representación del triángulo de Pascal como combinatoria (izquierda) y como sumas (derecha) | | **Ubicación del pie de imagen** | Inferior o lateral |   Se trata de un triángulo de números enteros, infinito y simétrico. El número superior es un 1, la segunda fila corresponde a los números combinatorios de 1, la tercera fila a los combinatorios de 2, la cuarta a los de 3 y así sucesivamente. Todas las filas acaban y empiezan en 1. Todas las filas son simétricas. Un número que no esté en un extremo de la fila es la suma de los dos que están situados en la fila superior encima de él, a izquierda y derecha.  Como combinatorias, los elementos del triángulo corresponden a las muestras no ordenadas de los elementos de un conjunto y son de mucha utilidad para encontrar los coeficientes de un binomio de grado alto, es decir encontrar los números de una determinada fila, sin tener que generar el triángulo completo. |

**Las combinaciones con repetición**

Las **combinaciones con repetición** son las **agrupaciones** que se pueden formar tomando ***n*** elementos **que pueden repetirse** de un conjunto total de ***m*** elementos, cuando **el orden no es significativo**. La ecuación que permite calcular la cantidad de resultados que se obtienen al combinar de esta manera es:



Por ejemplo, en una tienda venden cuatro tipos diferentes de papas fritas (con sal, con sabor a jamón, con sabor a salsa de barbacoa y con sabor a queso). Si se quieren comprar tres paquetes de papas, ¿de cuántas formas podemos elegirlas?

Un primer razonamiento es que se trata de una muestra de tres elementos a elegir de un conjunto de cuatro donde el **orden no importa**, pues da igual que se elijan dos paquetes de papas con sal y uno de papas con sabor a jamón, que un paquete de papas sabor jamón y dos paquetes de papas con sal. Un segundo razonamiento es que en la selección no **entran todos los elementos del conjunto**, sino que tan solo se eligen tres de entre los cuatro posibles. Finalmente, **como se repiten** los elementos, podemos elegir más de un paquete del mismo tipo.

El razonamiento previo señala que el conteo a realizar debe establecer la cantidad máxima de **combinaciones con repetición de 3 elementos, dentro de un conjunto de 4 elementos.**

**Si se asignan las denominaciones PS a un paquete de papas con sal, PJ a un paquete de papas sabor a jamón, PB a un paquete de papas sabor a barbacoa y PQ a un paquete de papas sabor a queso, entonces el diagrama de árbol que corresponde a las opciones posibles para elegir tres paquetes de papas es:**

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Combinatoria/ Muestras no ordenadas |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img10_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img10_zoom.jpg)**OJO!!!! Cambiar el nombre de “bolsas de patatas” por “paquetes de papas”**  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_11\_img10\_zoom.jpg |
| **Pie de imagen** | Diagrama de árbol con las combinaciones con repetición de tres elementos de un conjunto de cuatro, para el caso de la selección de los tres sabores de tres paquetes de papas elegidos entre cuatro sabores posibles. |
| **Ubicación del pie de imagen** | Inferior |

En el diagrama de árbol se aprecian las 20 combinaciones posibles, que resultan de las sumas de los distintos tipo de “rama”: 1 rama de 4 brazos, 2 ramas de 3 brazos, 3 ramas de 2 brazos y 4 ramas de un brazo, que resulta respectivamente en (1 × 4) + (2 × 3) + (3 × 2) + (4 × 1) = 4 + 6 + 6 + 4 = 20 opciones:

Para especificar el sentido del cálculo de las **combinaciones con repetición, nótese que por ejemplo las posibles** combinaciones con repetición de orden 1 son aquellas en las que se elige **un elemento de un conjunto de cuatro**. De esas hay 4, o equivalentemente:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula17_resized.gif

Por su parte **las posibles** combinaciones con repetición de orden 2 resultan ser la misma cantidad que aparece al construir las combinaciones sin repetición con un elemento más: combinaciones sin repetición de **elegir dos elementos de un conjunto de 4 + 1**. De esas hay 4, o equivalentemente:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula18_resized.gif

Finalmente, **las posibles** combinaciones con repetición de orden 3 resultan ser la misma cantidad que aparece al construir las combinaciones sin repetición con un elemento más. De la misma manera que en los casos precedentes, es igual que construir las **combinaciones sin repetición con un elemento más**, es decir:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_formula19_resized.gif

El resultado anterior corrobora lo encontrado usando el diagrama de árbol, es decir que las muestras no ordenadas de tres elementos de un conjunto de cuatro con repetición son 20. Además, como en los demás órdenes ocurrirá algo similar, se concluye que:



Un elemento fundamental para realizar el cálculo de probabilidades es definir la cantidad de casos favorables a un suceso, dentro de los casos posibles. Los casos posibles se calculan mediante los conteos de muestras ordenadas o no, en las que influye el orden o no, con posibilidad de repetición o no, presentados hasta aquí.

**[SECCIÓN 2] 1.2 Probabilidad y conjuntos**

Dada la definición de espacio muestral como conjunto de posibles resultados de un experimento aleatorio, y a la definición de suceso como subconjunto del espacio muestral, vemos que el contexto de tratamiento de los conceptos probabilísticos es de corte conjuntista. Sin embargo, como la probabilidad debe ser cuantificable, falta definir de qué manera se realizan los cálculos de probabilidades, que son numéricos, en relación con las operaciones entre conjuntos.

Para lograr hacer esa correspondencia, se definen las operaciones entre sucesos. Para ejemplificar las operaciones entre sucesos para un experimento aleatorio, podemos utilizar una representación conjuntista llamada Diagrama de Venn:

* **UNIÓN:** Dados dos sucesos A y B, del mismo espacio muestral, la unión de sucesos se verifica cuando ocurre A o cuando ocurre B. El suceso unión de A y B es un nuevo suceso, formado por todos los elementos de A y todos los elementos de B y se representa como A ∪ B.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/ Experimentos aleatorios |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img2_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img2_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_12\_img2\_zoom.jpg |
| **Pie de imagen** | Diagrama de Venn de la unión de dos sucesos simples. |
| **Ubicación del pie de imagen** | Inferior |

Por ejemplo, si se hace el experimento de “hacer girar una ruleta dividida en 6 sectores idénticos numerados del 1 al 6”, el suceso “obtener un número mayor o igual que 5” puede verse como la unión de los sucesos simples A: “obtener 5” y B: “obtener 6”. Entonces el nuevo evento A ∪ B será “obtener 5 ó obtener 6”, que equivale a “obtener un número mayor o igual que 5”

* **INTERSECCIÓN:** Dados dos sucesos A y B, del mismo espacio muestral, la **intersección de sucesos** solo se verifica cuando A y B ocurren al mismo tiempo, por lo que A y B deben ser eventos compatibles. El suceso intersección de A y B se denota por A ∩ B y es un nuevo suceso, formado por todos los elementos que son simultáneamente de A y de B.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/ Experimentos aleatorios |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img3_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img3_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_12\_img3\_zoom.jpg |
| **Pie de imagen** | Diagrama de Venn de la interseccción de dos sucesos simples. |
| **Ubicación del pie de imagen** | Inferior o lateral |

Por ejemplo, en el experimento de “hacer girar una ruleta dividida en 6 sectores idénticos numerados del 1 al 6”, el suceso “obtener un número primo” puede verse como la intersección de los sucesos simples A: “obtener un número impar” y B: “obtener un número solo divisible por sí mismo”. Entonces el nuevo evento A **∩** B será “obtener un número impar y obtener un número solo divisible por sí mismo”, que equivale al evento intersección A **∩** B: “obtener un número primo”.

* **DIFERENCIA:** Dados dos sucesos A y B, del mismo espacio muestral, el suceso diferencia se verifica cuando ocurre A, pero no ocurre B. El suceso diferencia de A y B se denota por A - B y es un nuevo suceso, formado por todos los elementos que pertenecen a A pero que no pertenecen a B. Nótese que no es lo mismo A – B que B – A.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Representación mediante diagrama de Venn de la diferencia entre dos conjuntos |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | https://upload.wikimedia.org/wikipedia/commons/thumb/e/ec/SetDifferenceA.svg/200px-SetDifferenceA.svg.png https://upload.wikimedia.org/wikipedia/commons/thumb/6/6c/SetDifferenceB.svg/200px-SetDifferenceB.svg.png  A – B B – A  Tomados de <https://upload.wikimedia.org/wikipedia/commons/thumb/e/ec/SetDifferenceA.svg/200px-> y SetDifferenceA.svg.pnghttps://upload.wikimedia.org/wikipedia/commons/thumb/6/6c/SetDifferenceB.svg/200px-SetDifferenceB.svg.png |
| **Pie de imagen** | Diagrama de Venn de la diferencia de dos sucesos simples. |
| **Ubicación del pie de imagen** | Inferior o lateral |

Por ejemplo, en el experimento aleatorio de la ruleta, de los sucesos A = {1, 5, 6} y B = {2, 3, 5, 6}, un suceso diferencia sería A - B = {1}, mientras que B – A = {2, 3}.

* **DIFERENCIA SIMÉTRICA:** Dados dos sucesos A y B, del mismo espacio muestral, el suceso diferencia simétrica se verifica cuando ocurre A, pero no ocurre B o cuando ocurre B pero no ocurre A. El suceso diferencia simétrica de A y B se denota por A ∆ B y es un nuevo suceso, formado por la unión de los sucesos A - B y B - A. Así A ∆ B = (A – B) ∪ (B – A) = (A ∪ B) – (A **∩** B)

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Representación mediante diagrama de Venn de la diferencia simétrica entre dos conjuntos |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | https://upload.wikimedia.org/wikipedia/commons/thumb/f/f2/SetSymmetricDifference.svg/220px-SetSymmetricDifference.svg.pngTomada de https://upload.wikimedia.org/wikipedia/commons/thumb/f/f2/SetSymmetricDifference.svg/220px-SetSymmetricDifference.svg.png |
| **Pie de imagen** | Diagrama de Venn de la diferencia simétrica de dos sucesos simples. |
| **Ubicación del pie de imagen** | Inferior o lateral |

Por ejemplo, en el experimento aleatorio del giro de la ruleta, de los sucesos A = {1, 5, 6} y B = {2, 3, 5, 6}, la diferencia simétrica sería A ∆ B = {1, 2, 3}. Es decir que es la unión de los sucesos, quitándoles su intersección.

* **COMPLEMENTO:** Dado un suceso A de un espacio muestral, el suceso complemento o suceso contrario es el suceso que tiene lugar siempre que no se verifica A. El suceso complemento de A se denota AC. Dos **sucesos complementarios A y** AC son evidentemente **incompatibles, y su unión es el espacio muestral**.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Representación mediante diagrama de Venn del complemento de un conjunto, en el contexto de espacios muestrales |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | https://upload.wikimedia.org/wikipedia/commons/thumb/b/b9/SetComplement.svg/220px-SetComplement.svg.png**OJO: Escribir en la parte sombreada azul AC en lugar de la U y escribir por fuera del conjunto el símbolo Ω (Omega) para indicar que el conjunto universal es el espacio muestral**  Tomada de https://upload.wikimedia.org/wikipedia/commons/thumb/b/b9/SetComplement.svg/220px-SetComplement.svg.png |
| **Pie de imagen** | Diagrama de Venn del complemento AC de un suceso A en un espacio muestral Ω. |
| **Ubicación del pie de imagen** | Inferior o lateral |

Por ejemplo, en el experimento aleatorio del giro de la ruleta el contrario de A = {2, 4, 6} es AC = {1, 3, 5}.

Veamos algunas operaciones entre sucesos, para el experimento de lanzar un dado y los sucesos *A*: “obtener un número par, que en este caso corresponde a A = {2, 4, 6} y *B*: “obtener un número mayor que 2”, es decir B = {3, 4, 5, 6}. En el Diagrama de Venn el rectángulo representa el espacio muestral y las elipses los sucesos A y B, subconjuntos del espacio muestral.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Probabilidad/Los fenómenos aleatorios/Las operaciones con sucesos. |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14873/InfoGuion/cuadernoestudio/images_xml/MT_09_12_img3_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14873/InfoGuion/cuadernoestudio/images_xml/MT_09_12_img3_zoom.jpg) |
| Pie de imagen | Diagrama de Venn en el que se representan los sucesos A: “obtener un número par” y B: “obtener un número mayor que 2” para el experimento aleatorio de lanzar un dado. |
| Ubicación del pie de imagen | Inferior |

Entonces:

A ∪ B = {2, 3, 4, 5, 6}

A ∩ B = {4, 6}

Como un segundo ejemplo se realiza el experimento aleatorio de “tirar un dado dos veces consecutivas”. Se halla a continuación su espacio muestral:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Probabilidad/Los fenómenos aleatorios/Las operaciones con sucesos. |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Resultados posibles de tirar dos dados de seis caras** | | | | | | | (1, 1) | (1, 2) | (1, 3) | (1, 4) | (1, 5) | (1, 6) | | (2, 1) | (2, 2) | (2, 3) | (2, 4) | (2, 5) | (2, 6) | | (3, 1) | (3, 2) | (3, 3) | (3, 4) | (3, 5) | (3, 6) | | (4, 1) | (4, 2) | (4, 3) | (4, 4) | (4, 5) | (4, 6) | | (5, 1) | (5, 2) | (5, 3) | (5, 4) | (5, 5) | (5, 6) | | (6, 1) | (6, 2) | (6, 3) | (6, 4) | (6, 5) | (6, 6) | |
| Pie de imagen | Resultados posibles de tirar dos dados de seis caras. |
| Ubicación del pie de imagen | Inferior |

Seleccionamos entre todos los resultados aquellos que correspondan a los siguientes sucesos:

* Suceso *A* = “obtener pares”.

*A* = {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}.

* Suceso *B* = “obtener una suma par”.

*B* = {(1, 1), (1, 3), (1, 5), (2, 2), (2, 4), (2, 6), (3, 1), (3, 3), (3, 5), (4, 2), (4, 4), (4, 6), (5, 1), (5, 3),(5, 5),(6, 2),(6, 4),(6, 6)}.

* Suceso *C* = “obtener una suma igual a 9”.

*C* = {(3, 6), (4, 5), (5, 4), (6, 3)}.

* Suceso *D* = “obtener una suma impar”.

*D* = {(1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5), (3, 2), (3, 4), (3, 6), (4, 1), (4, 3), (4, 5), (5, 2), (5, 4), (5, 6), (6, 1), (6, 3), (6, 5)}

* Suceso *A* ∪ *C* = “obtener pares” u “obtener una suma de 9”.

*A* ∪ *C* = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (3, 6), (4, 5), (5, 4), (6, 3)}.

* Suceso *A* ∪ *D* = “obtener pares” u “obtener una suma impar”.

*A* ∪ *D* = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5), (3, 2), (3, 4), (3, 6), (4, 1), (4, 3), (4, 5), (5, 2), (5, 4), (5, 6), (6, 1), (6, 3), (6, 5)}

* Suceso A ∩ C = “obtener pares” y “obtener una suma de 9”.

A ∩ C = {∅}.

* Suceso complementario o contrario de A ∪ D (es decir, aquellos resultados que faltan al conjunto A ∪ D para completar el espacio muestral).

(A ∪ D )C = {(1, 3), (1, 5), (2, 4), (2, 6), (3, 1), (3, 5), (4, 2), (4, 6), (5, 1), (5, 3), (6, 2), (6, 4)}

* ¿Qué sucesos son incompatibles?

Los sucesos A y C, los sucesos A y D, los sucesos C y B, y los sucesos B y D.

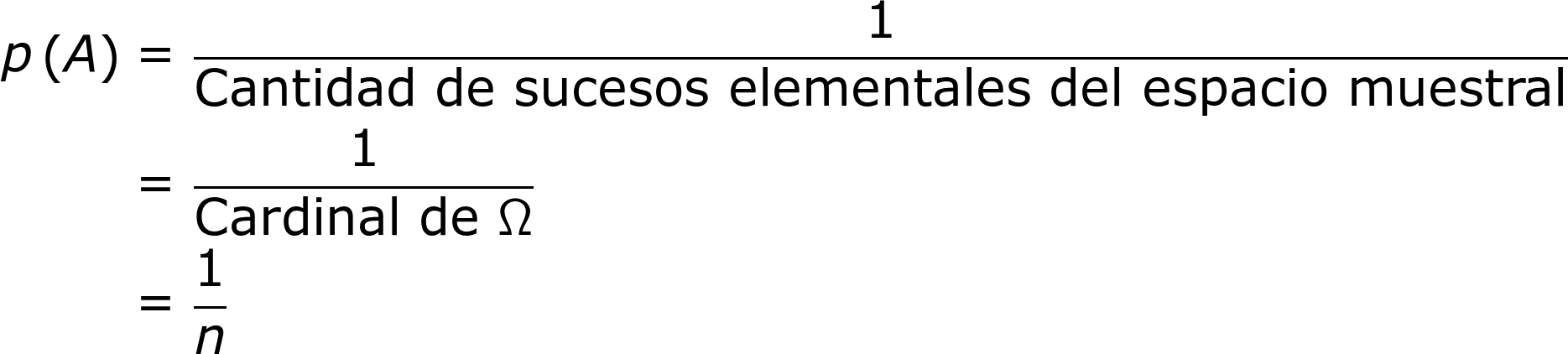
* De todos estos sucesos incompatibles, ¿cuáles son también sucesos contrarios?

Los sucesos *B* y *D*, porque o saldrá una suma par o saldrá una suma impar. Además, la unión de ambos corresponde al espacio muestral.

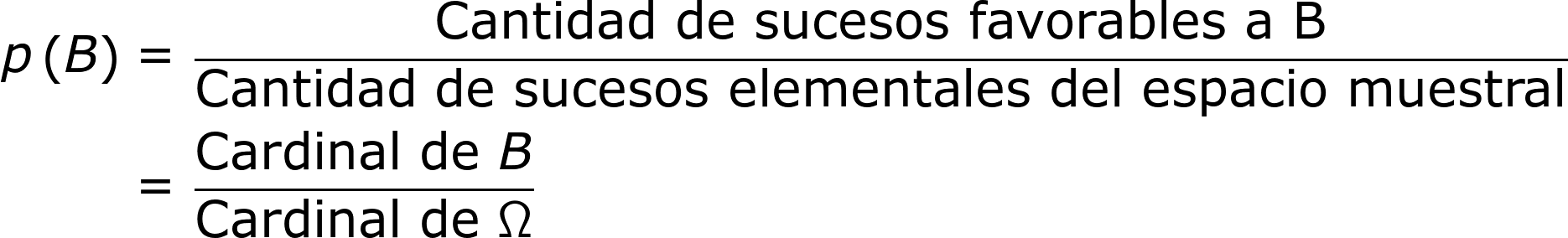
[**SECCIÓN 2] 1.3 Consolidación**

**[SECCIÓN 1] 2 Concepto de probabilidad**

La probabilidad de un suceso es el número que representa la proporción de veces que podemos esperar que un suceso se verifique cuando el experimento es repetido muchas veces en idénticas condiciones. Uno de los resultados más importantes de la probabilidad es el Teorema de Laplace, según el cual la probabilidad de ocurrencia de un suceso corresponde a la razón entre la cantidad de opciones favorables al suceso, sobre la cantidad de opciones posibles. Por ejemplo, si un suceso A es elemental o simple, la probabilidad de que ocurra es:



Por su parte, si B es un suceso de un experimento aleatorio en que todos los casos posibles tienen la misma probabilidad, se cumple que la probabilidad de dicho suceso, sea simple o compuesto es:



Por ejemplo, si se tiene una urna con 40 bolas: 20 rojas, 10 verdes, 5 azules y 5 amarillas, y se realiza el experimento aleatorio de extraer de la bolsa una bola al azar, la probabilidad de cada color será:

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14873/InfoGuion/cuadernoestudio/images_xml/MT_09_12_formula15_resized.gif

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14873/InfoGuion/cuadernoestudio/images_xml/MT_09_12_formula16_resized.gif

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14873/InfoGuion/cuadernoestudio/images_xml/MT_09_12_formula17_resized.gif

http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14873/InfoGuion/cuadernoestudio/images_xml/MT_09_12_formula18_resized.gif

Hay experiencias que, en principio, no corresponden con resultados equiprobables, pero que con una leve reformulación se pueden transformar, lo permite aplicar el Teorema de Laplace. Por ejemplo, si para el experimento aleatorio de “lanzar dos dados y estudiar su suma” se define el espacio muestral como Ω = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}, los casos posibles no son equiprobables. En cambio, si se hace una distinción entre los dos dados, se llega a detallar un espacio muestral con 36 casos posibles equiprobables:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | 4º ESO/Matemáticas/Probabilidad/Los fenómenos aleatorios/Las operaciones con sucesos. |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Resultados posibles de tirar dos dados de seis caras** | | | | | | | (1, 1) | (1, 2) | (1, 3) | (1, 4) | (1, 5) | (1, 6) | | (2, 1) | (2, 2) | (2, 3) | (2, 4) | (2, 5) | (2, 6) | | (3, 1) | (3, 2) | (3, 3) | (3, 4) | (3, 5) | (3, 6) | | (4, 1) | (4, 2) | (4, 3) | (4, 4) | (4, 5) | (4, 6) | | (5, 1) | (5, 2) | (5, 3) | (5, 4) | (5, 5) | (5, 6) | | (6, 1) | (6, 2) | (6, 3) | (6, 4) | (6, 5) | (6, 6) | |
| Pie de imagen | Resultados posibles de tirar dos dados de seis caras. |
| Ubicación del pie de imagen | Inferior |

En este caso el espacio muestral Ω cuenta con 36 opciones, todas ellas igualmente probables. ¿Cuál es la probabilidad de los sucesos enunciados?

* Suceso *A* = “lograr dos números iguales”.

*A* = {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}.

*p(A) = 6/36 = 1/6*

* Suceso *B* = “obtener una suma par”.

*B* = {(1, 1), (1, 3), (1, 5), (2, 2), (2, 4), (2, 6), (3, 1), (3, 3), (3, 5), (4, 2), (4, 4), (4, 6), (5, 1), (5, 3),(5, 5),(6, 2),(6, 4),(6, 6)}.

*p (B) = 18/36 = 1/2*

* Suceso *C* = “obtener una suma igual a 9”.

*C* = {(3, 6), (4, 5), (5, 4), (6, 3)}.

*p (C) = 4/36 = 1/9*

* Suceso *D* = “obtener una suma impar”.

*D* = {(1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5), (3, 2), (3, 4), (3, 6), (4, 1), (4, 3), (4, 5), (5, 2), (5, 4), (5, 6), (6, 1), (6, 3), (6, 5)}.

*p (D) = 18/36 = 1/2*

* Suceso *A* ∪ *C* = “lograr dos números iguales” u “obtener una suma de 9”.

*A* ∪ *C* = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (3, 6), (4, 5), (5, 4), (6, 3)}.

*p (A* ∪ *C) = 10/36 = p(A) + p(C) = 1/6* + *1/9 = 10/36*

* Suceso *A* ∪ *D* = “lograr dos números iguales” u “obtener una suma impar”.

*A* ∪ *D* = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5), (3, 2), (3, 4), (3, 6), (4, 1), (4, 3), (4, 5), (5, 2), (5, 4), (5, 6), (6, 1), (6, 3), (6, 5)}

*p (A* ∪ *D) = 24/36 = 2/3 = p(A) + p(D) = 1/6* + *1/2 = 2/3*

* Suceso A ∩ C = “lograr dos números iguales” y “obtener una suma de 9”.

A ∩ C = {∅}.

p (A ∩ C) = 0

* Suceso complementario o contrario de A ∪ D (es decir, aquellos resultados que faltan al conjunto A ∪ D para completar el espacio muestral).

(A ∪ D )C = {(1, 3), (1, 5), (2, 4), (2, 6), (3, 1), (3, 5), (4, 2), (4, 6), (5, 1), (5, 3), (6, 2), (6, 4)}

p {(A ∪ D )C} = 12/36 = 1/3 = 1- p (A ∪ D ) = 1 – 1/3 = 2/3

**Propiedades del cálculo de probabilidades**

Para el cálculo de probabilidades deben tenerse en cuenta las siguientes propiedades:

* Para **cualquier suceso** **A, bien sea simple o compuesto**, se cumple que su probabilidad no puede ser menor que 0 ni mayor que 1. Es decir que

*0 ≤ p(A) ≤ 1*

La anterior expresión equivale a las probabilidades porcentuales, es decir que:

*0 ≤ p%(A) ≤ 100.*

* La probabilidad de un **suceso** seguro es 1:

*p(Ω) = 1*

que expresado en porcentaje es:

*p%(Ω) = 100.*

* La probabilidad de un suceso **imposible** es 0:

*p(∅) = 0*

Nótese que Ω como espacio muestral es el complemento del conjunto vacío (∅), es decir que ΩC = ∅ y ∅C = Ω.

* Si **dos sucesos** A y B son **incompatibles**: la probabilidad de su unión es la suma de sus probabilidades.

*p(A ∪ B) = p(A) + p(B)*

* Si **dos sucesos** A y B son **compatibles**: se verifica que la probabilidad de su unión es igual a la suma de sus probabilidades, menos la probabilidad de la intersección.

*p(A ∪ B) = p(A) + p(B) – p(A ∩ B)*

* Si **A y AC** son **sucesos complementarios**, entonces la suma de sus probabilidades es igual a 1. De lo anterior se deduce que la probabilidad de cualquier suceso es igual a 1 menos la probabilidad de su complementario.

*p(A) + p(****AC****) = 1*

*p(****AC****) = 1 - p(A)*

La anterior propiedad se deduce debido a que *A* y ***AC*** son mutuamente excluyentes, por lo que *A* ∪ ***AC*** = Ω y como es sabido P(Ω) = 1.

* **La suma de las probabilidades de todos los sucesos elementales tiene que ser 1**.
* Si un suceso está incluido en otro, su probabilidad es menor o igual a la de este, es decir que si A ⊂ B ⇒ p(A) ≤ p(B)

Fíjate en el siguiente dado:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código |  |
| Descripción | Imagen de un dado común de seis caras |
| Código Shutterstock (o URL o la ruta en AulaPlaneta) | Resultado de imagen para dado de 4 caras |
| Pie de imagen | Dado de seis caras común de un juego de azar |
| Ubicación del pie de imagen | Inferior |

Al jugar a lanzar un dado podemos ejemplificar la probabilidad de diferentes sucesos. Por ejemplo ¿cuál es la probabilidad de cada suceso elemental?, ¿cuál es la probabilidad de obtener el suceso A: “obtener un número par” o B: “obtener un número mayor o igual que 5”?

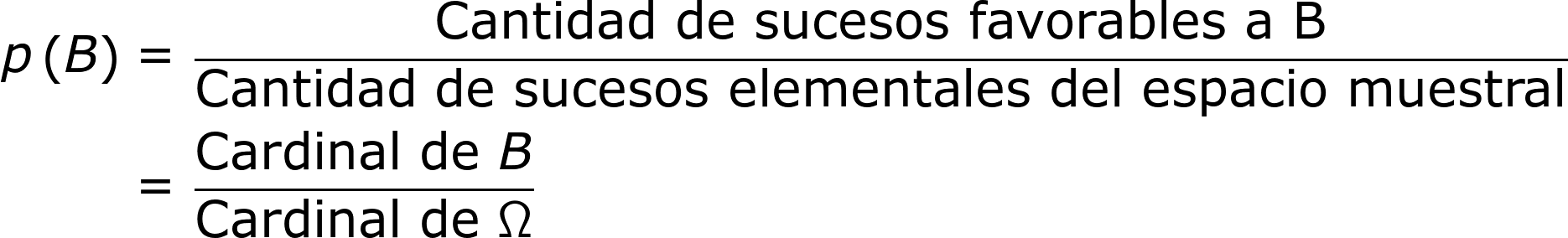
El espacio muestral para el experimento aleatorio es *Ω = {1, 2, 3, 4, 5, 6}*. Las probabilidades para los sucesos elementales serán *p({1}) = p({2}) = p({3}) = p({4}) = p({5})= p({6}) = 1/6.*

La probabilidad de obtener un número par o *p(A)* es *p(A) = p({2}) + p({4}) + p({6}) = 3/6 = 1/2*. Por su parte, la probabilidad de obtener un número mayor o igual que 5 o *p(B)* es *p(B)* = *p({5}) + p({6}) = 2/6 = 1/3*

Para un segundo ejemplo, si en una tienda venden cuatro tipos diferentes de papas fritas (con sal, con sabor a jamón, con sabor a salsa de barbacoa y con sabor a queso) y se quieren comprar tres paquetes de papas. ¿Cuál es la probabilidad de que ninguno de los tres paquetes sea sabor a pollo? ¿Cuál la probabilidad de que exactamente dos paquetes tengan sabor a queso?

Para la primera pregunta, la respuesta es inmediata: como entre los tipos de papas fritas no hay opción de sabor a pollo, el suceso de que ningún paquete tenga sabor a pollo es un suceso siempre cierto, es decir que la probabilidad de que ninguno sea de sabor a pollo es 1. Escrito en un lenguaje más formal: sea A el suceso A: “Ningún paquete tiene sabor a pollo”, entonces como el espacio muestral es Ω = {**PS: paquete de papas con sal, PJ: paquete de papas sabor a jamón, PB: paquete de papas sabor a barbacoa y PQ:** paquete de papas sabor a queso}, entonces el suceso PP: **paquete de papas sabor a pollo es un suceso imposible, entonces su probabilidad es p (PP) = 0. Como el evento A siempre ocurre es decir que siempre ocurre que ningún paquete tiene sabor a pollo, entonces p(A) = 1.**

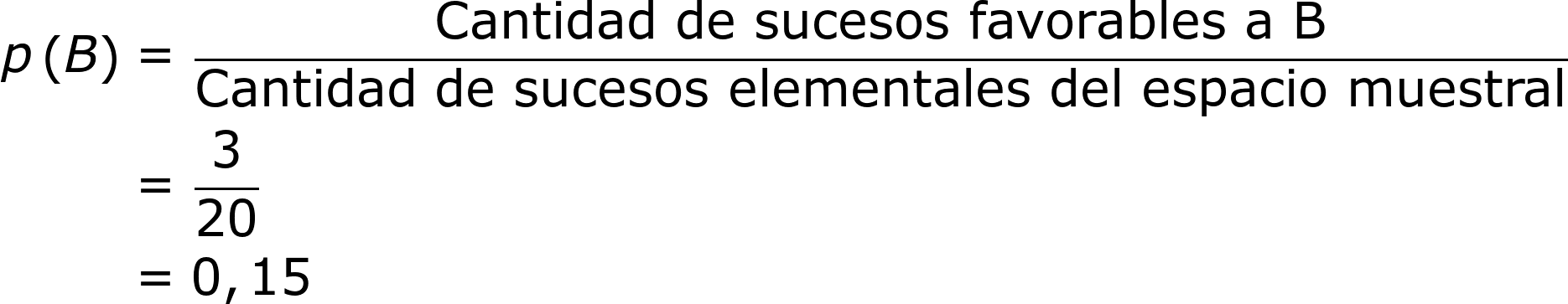
Para la segunda pregunta, teniendo el diagrama de todas las opciones posibles, entonces, sea B el evento B: “Exactamente dos de entre los tres paquetes tienen sabor a queso”. Como:



entonces realizando el conteo de los casos favorables sobre los casos posibles, se obtiene:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | Modificación de la imagen de 4º ESO/Matemáticas/Combinatoria/ Muestras no ordenadas disponible en <http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14644/InfoGuion/cuadernoestudio/images_xml/MT_10_11_img10_zoom.jpg>, seleccionando los casos en que aparecen exactamente dos paquetes de papas sabor a queso |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** |  |
| **Pie de imagen** | Diagrama de árbol con las combinaciones para la selección de sabores de tres paquetes de papas elegidos entre cuatro sabores posibles, resaltando los casos en que aparecen exactamente dos paquetes de papas sabor a queso. |
| **Ubicación del pie de imagen** | Inferior |

En el diagrama de árbol se aprecian las 20 las combinaciones para la selección de sabores. Están resaltados los 3 casos en los que aparecen exactamente dos paquetes de papas sabor a queso, así que:

**

Así que la probabilidad de que aparezcan exactamente dos paquetes de papa sabor a queso es del 15%.

**Experimentos aleatorios compuestos**

Un **experimento compuesto** es el que está formado por **varios sucesos elementales** o simples, realizados de forma consecutiva.

|  |  |
| --- | --- |
| **Recuerda** | |
| **Contenido** | Cuando dos **sucesos** pueden **ocurrir simultáneamente** se denominan **compatibles**; en caso contrario, se dice que son incompatibles. |

Para facilitar el análisis de un **experimento compuesto**, se suele utilizar un **diagrama de árbol**. En este diagrama se representan las diversas posibilidades de desarrollo que tiene un suceso compuesto, de forma que las diversas ramas o caminos sean mutuamente excluyentes.

|  |  |
| --- | --- |
| **Recuerda** | |
| **Contenido** | Un **diagrama de árbol** es una representación gráfica para construir y contar todas las posibles maneras de combinar elementos de uno o más conjuntos. Mediante un diagrama de árbol se **condensan de forma gráfica todas las agrupaciones posibles**.  Un diagrama de árbol debe considerar todas las posibilidades, en caso contrario proporciona probabilidades erróneas. |

En los diagramas de árbol al servicio del cálculo de probabilidades, en cada rama se señala la probabilidad de que pase aquello que se indica. En el diagrama de árbol el cálculo en cada desprendimiento de ramas se realiza como si los sucesos indicados por las ramas que llevan hasta allí se hubieran realizado. Ello se evidencia porque l**a suma de las probabilidades de cada abanico de ramas es igual a 1**.

Para calcular una probabilidad, solo hay que dibujar el camino correspondiente en un diagrama de árbol y realizar el **producto de las probabilidades de todas las ramas** que lo forman. Esta multiplicación recibe el nombre de **regla del producto**.

Para calcular la probabilidad de un suceso, es necesario:

* Examinar todos los **caminos** que le son **favorables**.
* **Calcular la probabilidad** de cada camino **multiplicando las probabilidades** de las ramas que lo forman.
* **Sumar las probabilidades de todos los caminos favorables**.

La **regla del producto resulta ser entonces** una forma de calcular la probabilidad del suceso intersección, A ∩ B, o del suceso “A y después B” o “B, dado que ya ocurrió A”. Consiste en **multiplicar las probabilidades de los sucesos simples** que forman el camino. Posteriormente, para calcular la probabilidad del suceso compuesto, **se suman las probabilidades** de todos los caminos que le son favorables.

Por ejemplo, en una lotería hay 100 números, con dos premios, uno de 500 millones de pesos (representado como I en el diagrama) y otro de 200 millones de pesos, (representado por el II, en el diagrama). Si se han comprado 2 billetes, ¿qué probabilidad hay de no ganar nada, de ganar el primer premio, de ganar el segundo premio y de ganar ambos premios?

Para poder dar respuesta a las pregunta, se construye el **espacio muestral con los resultados posibles y las probabilidades de los sucesos elementales**. Una forma de lograr hacerlo es a través de un **diagrama de árbol:**

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/ Experimentos aleatorios |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_%20img5_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_%20img5_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_12\_%20img5\_zoom.jpg |
| **Pie de imagen** | Diagrama de árbol de las probabilidades de obtener premio en una lotería comprando dos billetes. |
| **Ubicación del pie de imagen** | Inferior |

Así, las **probabilidades** de los posibles **sucesos** quedan recogidas en la siguiente **tabla**:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/Experimentos aleatorios compuestos |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img6_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img6_zoom.jpg) **OJO!!!! Quitar los valores en euros, o cambiarlos por 0 millones, 200 millones, 500 millones y 700 millones en cada fila respectivamente.** |
| **Pie de imagen** | Tabla resumen de las probabilidades de los sucesos posibles. La suma de todas las probabilidades debe ser igual a 1. |
| **Ubicación del pie de imagen** | Inferior |

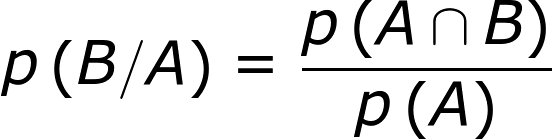
**[SECCIÓN 2] 2.1 Probabilidad condicional**

En muchas ocasiones, la ocurrencia de un suceso no interfiere sobre la ocurrencia de otro. Por ejemplo si se lanza primero un dado y luego una moneda, un resultado no interfiere sobre el segundo; tales sucesos se conocen como sucesos independientes. Cuando los sucesos son independientes, la probabilidad de un suceso compuesto está dada por el producto de las probabilidades de los sucesos simples que lo conforman.

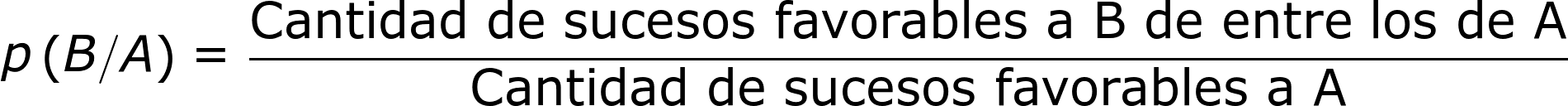
Por el contrario, cuando se realizan observaciones de **sucesos compuestos**, puede suceder que uno dependa del otro. Los sucesos A: “estudiar para trigonometría” y B: “aprobar trigonometría” son **sucesos dependientes**, ya que cuanto más se estudia, más probabilidad hay de aprobar la asignatura.

La probabilidad de que se verifique un suceso B (aprobar) cuando ha ocurrido otro A (estudiar) se llama **probabilidad condicionada** y se expresa por *p(B/A)*, que se lee “la probabilidad de B, dado A” o “la probabilidad de B condicionada a A”. En ese caso, la probabilidad no es la suma de los sucesos que lo componen.

Si A y B son dos sucesos de un experimento aleatorio y A no es un suceso imposible, es decir, p(A) ≠ 0, se define la **probabilidad** de B **condicionada** a A como:



En el caso de que se trate de un experimento para el que se pueda aplicar la fórmula de Laplace, el número anterior puede ser calculado así: cantidad de casos favorables de B entre los que también son favorables a A.



A partir de la definición del cálculo de la probabilidad condicionada y según la regla del producto entonces:

* Si A y B son dependientes: p (A ∩ B) =p (A) × p (B/A).
* Si A y B son independientes: p (A ∩ B)= p (A) × p (B) ya que en ese caso p (B/A) = p(B).

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Los sucesos dependientes y los independientes** |
| **Contenido** | Los experimentos simples que forman un experimento compuesto pueden ser independientes o dependientes.   * Dos sucesos son independientes cuando el resultado de cada uno de ellos no depende del resultado del otro, es decir que el resultado de cada suceso no depende de los demás; * Dos sucesos son dependientes, cuando el resultado de cada uno de ellos influye en las probabilidades de los demás. * Si A y B son dependientes: p (A ∩ B) =p (A) × p (B/A). * Si A y B son independientes: p (A ∩ B)= p (A) × p (B). |

Por ejemplo, se realiza el experimento de lanzar una moneda y después lanzar un dado y se quiere calcular la probabilidad de los siguientes dos sucesos:

A: “obtener un 6”.

B: “obtener cara”.

Para analizar la independencia de los siguientes dos sucesos, inicialmente se define el **espacio muestral** del experimento, que se condensa en el diagrama de árbol, en el que se hace la convención de que C = cara y de que X = cruz:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/ Probabilidad condicionada |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img7_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img7_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_12\_img7\_zoom.jpg |
| **Pie de imagen** | Espacio muestral para el experimento aleatorio de lanzar una moneda y después lanzar un dado |
| **Ubicación del pie de imagen** | Inferior |

El espacio muestral es E = {(C, 1), (C, 2), (C, 3), (C, 4), (C, 5), (C, 6), (X, 1), (X, 2), (X, 3), (X, 4), (X, 5), (X, 6)}, compuesto por 12 elementos.

Para analizar la probabilidad condicionada de los sucesos A: “obtener un 6” y B: “obtener cara”, se siguen los siguientes pasos

* Se escriben los **sucesos simples** que componen los dos sucesos a comparar y el suceso intersección:

A = {(C, 6), (X, 6)}; B = {(C, 1), (C, 2), (C, 3), (C, 4), (C, 5), (C, 6)}

A ∩ B = {(C, 6)}

* Se calculan las **probabilidades de los sucesos** y **la probabilidad de la intersección de sucesos**.

p (A) = 2/12 =1/6

p (B) = 6/12 = 1/2

p (A ∩ B) = 1/12

* Se comprueba si el **producto de las probabilidades de los sucesos** es **igual** a la **probabilidad de la intersección** de los mismos. Si son iguales, los sucesos son independientes, de lo contrario son dependientes. En este caso

*p (A) × p (B) = 2/12 × 6/12 = 12/144 = 1/12*

*p (A ∩ B) = 1/12 = p(A) × p (B)*

Así, como los resultados de las probabilidades son iguales, se ha comprobado que los sucesos *A* y *B* son sucesos independientes.

En ocasiones, no es tan evidente la independencia de sucesos. Por ello, es imprescindible **calcular**, por un lado, la **probabilidad de la intersección** y por otro, el producto de las probabilidades de cada suceso, a fin de comprobar si el resultado obtenido en cada cálculo es el mismo. Si **son iguales, se asegura la independencia de los sucesos. Si no, se debe recurrir a otras estrategias para realizar el cálculo de la probabilidad condicionada, usando por ejemplo diagramas de árbol o tablas de contingencia.**

**Tablas de contingencia**

**Cuando se clasifican los datos de un grupo que tienen más de una modalidad, referidos a dos características distintas, los datos se pueden expresar a través de una tabla de doble entrada, de manera que se garantice que cada individuo ha quedado clasificado en una y solo una casilla. Cuando es así, la tabla se conoce como tabla de contingencia.**

**Entonces, las entradas en la tabla discriminan los grupos según sus modalidades y características, de la siguiente manera:**

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/ Probabilidad condicionada |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img7b_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img7b_zoom.jpg)  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_12\_img7b\_zoom.jpg |
| **Pie de imagen** | Tabla de contingencia y probabilidades presentes en ella |
| **Ubicación del pie de imagen** | Inferior |

A partir del esquema anterior se observa que en una tabla de contingencia cada individuo debe quedar clasificado en una y solo en una casilla.

Por ejemplo, en una clase de 30 alumnos, 19 son mujeres y 11 hombres. Si 7 mujeres y 8 hombres llevan tenis y se escoge una persona al azar, cuál es la probabilidad de los siguientes sucesos:

A: “que sea una mujer y lleve tenis”.

B: “que sea una mujer, sabiendo que lleva tenis”.

Para facilitar los cálculos, vamos a indicar los datos en una **tabla de contingencia, donde H significa “Hombre”, M “mujer”, T “lleva tenis” y T’ “no lleva tenis”**:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/ Probabilidad condicionada |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img7c_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img7c_zoom.jpg) |
| **Pie de imagen** | Tabla de contingencia para los individuos clasificados según el género y el tipo de calzado. **H significa “Hombre”, M “mujer”, T “lleva tenis” y T’ “no lleva tenis”** |
| **Ubicación del pie de imagen** | Inferior o lateral |

Estos mismos datos quedarían distribuidos en un diagrama de árbol de la siguiente manera:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** |  |
| **Descripción** | 4º ESO/Matemáticas/Probabilidad/ Probabilidad condicionada |
| **Código Shutterstock (o URL o la ruta en AulaPlaneta)** | [http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img8_small.jpg](http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images_xml/MT_10_12_img8_zoom.jpg) **OJO!!!! CAMBIAR “tejanos” POR “tenis”, “chicas” POR “mujeres” Y “chicos” POR “hombres”. OJO**  http://profesores.aulaplaneta.com/DNNPlayerPackages/Package14645/InfoGuion/cuadernoestudio/images\_xml/MT\_10\_12\_img8\_zoom.jpg |
| **Pie de imagen** | Diagrama de árbol en el que se detalla el espacio muestral del experimento. |
| **Ubicación del pie de imagen** | Inferior |

En el diagrama de árbol se especifican las probabilidades de cada rama y, al final, las probabilidades de cada camino.

Para responder a la primera pregunta, se observa que de los 30 alumnos, 7 son mujeres con tenis; luego la probabilidad de que la persona elegida sea una mujer con tenis es:

*p (M ∩ T) = 7/30 = 0,23*

Se obtiene el mismo resultado al aplicar la ecuación de probabilidad de sucesos dependientes,

p (M ∩ T) = p (M) ×·p (T/M)

p (M ∩ T) = 19/30 ×·7/19

p (M ∩ T) = 7/30

En el caso del suceso B: “que sea una mujer sabiendo que lleva tenis”, está condicionada a “llevar tenis”. De los 15 casos que hay de personas con tenis, 7 son mujeres, por tanto, la probabilidad es:

p (M/T) = 7/15

Si se aplica la ecuación de la probabilidad de ser mujer condicionada a llevar tenis, se obtiene el mismo resultado:

p (M/T) = p (M ∩ T) / p (T) = (7/30) / (15/30) = 7/15

(No es posible tener una Sección 4)

[SECCIÓN 1]**Fin de unidad**

|  |  |
| --- | --- |
| **Mapa conceptual** | |
| **Código** | XX\_00\_00\_REC00 |
| **Título** | Mapa conceptual |
| **Descripción** |  |

|  |  |
| --- | --- |
| **Evaluación: recurso nuevo** | |
| **Código** | XX\_00\_00\_REC00 |
| **Título** |  |
| **Descripción** |  |

|  |  |  |
| --- | --- | --- |
| **Webs de referencia** | | |
| **Código** | XX\_00\_00\_REC00 | |
| **Web 01** | *Probabilidad y estadística* | [*URL*](http://lya.fciencias.unam.mx/lars/libros/pe-agosto-2006.pdf) |
| **Web 02** | *Probabilidad* | [*URL*](http://www.virtual.unal.edu.co/cursos/ciencias/2010620/) |
| **Web 03** | *Ejercicios resueltos de probabilidad condicionada* | [*URL*](http://www3.uah.es/jmmartinezmediano/Segundo%20CS/MCCSS%20Tema%2009b%20Problemas%20de%20probabilidad%20condicionada.pdf) |

SECCIÓN 1] **4 Estadística bidimensional**

[SECCIÓN 1]**Ejercitación y competencias**

|  |  |
| --- | --- |
| **Mapa conceptual** | |
| **Código** | XX\_00\_00\_REC00 |
| **Título** | Mapa conceptual |
| **Descripción** |  |

|  |  |
| --- | --- |
| **Evaluación: recurso nuevo** | |
| **Código** | XX\_00\_00\_REC00 |
| **Título** |  |
| **Descripción** |  |

|  |  |  |
| --- | --- | --- |
| **Webs de referencia** | | |
| **Código** | XX\_00\_00\_REC00 | |
| **Web 01** | *La estadística y la propuesta de un currículo por competencias* | [*URL*](https://repensarlasmatematicas.files.wordpress.com/2014/01/s68-material-de-referencia.pdf) |
| **Web 02** | *Estadística descriptiva unidimensional* | [*URL*](http://www.fuenterrebollo.com/Economicas2013/unidimensional-teoria.pdf) |
| **Web 03** | *Estadística bidimensional* | [*URL*](http://platea.pntic.mec.es/jfgarcia/editorialsm/es4_op_b_sm_esfera/unidad15.pdf) |