|  |  |
| --- | --- |
| Título del guion | La estadística y la combinatoria |
| Código del guion | MA\_10\_06\_CO |
| Descripción | La estadística permite estudiar las características de una población utilizando variables, como las de dispersión y las de posición. El azar es imprevisible, pero también produce regularidades que la probabilidad cuantifica y mide. |

[SECCIÓN 1] **1 La estadística**

La estadística se puede dividir, a grandes rasgos, en dos áreas: estadística descriptiva y estadística inductiva.

La **estadística descriptiva** se dedica a recolectar, sistematizar, ordenar, describir y analizar algunas características de una población, con base en un conjunto de datos.

Por otro lado, la **estadística inductiva** o 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 pueden establecer conclusiones y tomar decisiones ante situaciones de incertidumbre.

La **agrupación de datos** es una herramienta de gran utilidad cuando se tiene un gran número de datos. El proceso se denomina **distribución de frecuencias**dado que no solo se distribuyen los datos, sino que tal distribución se realiza en subgrupos, 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**

Las **variables cualitativas** indican **características no medibles o numerables** de los individuos en una población, por lo tanto, 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.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Una variable cualitativa estadística** |
| **Contenido** | Una variable cualitativa es estadística si es posible clasificar los datos obtenidos en clases bien definidas. |

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

Las **variables cuantitativas** toman **valores numéricos,** pues recopilan atributos medibles de los individuos en una población, como pueden ser la edad, la estatura, el número de hermanos, el tiempo de estudio, participación semanal en redes sociales, etc.

Las variables cuantitativas se clasifican en discretas y continuas.

Las **variables cuantitativas discretas** solo pueden tomar **valores enteros**, es decir que son el resultado de contar valores con **números naturales**, por ejemplo el número de personas en el hogar.

Las **variables cuantitativas continuas** son aquellas que pueden tomar valores del conjunto de los **números reales**, es decir que se admiten todos los valores en un intervalo. Son el resultado de medir, por ejemplo la estatura, el peso de una persona, el tiempo que dedica una persona en actividades deportivas, etc.

[SECCIÓN 2] **1.3 Los intervalos de clase**

Si las **variables** toman un **número grande de valores** o si la variable es **continua,** se agrupan formando **intervalos de clase**. Estos intervalos tienen la misma **amplitud** y se conocen como **clases**.

Cada clase está delimitada por un valor inferior y otro superior, los cuales son llamados **límites de la clase**. La **marca de clase** es el punto medio de cada intervalo; este valor 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.

* Por ejemplo, si de un grupo de 237 datos se tienen valores que están entre 1 y 100, se pueden definir grupos para sintetizar la información; los datos se resumen en cuatro clases con la misma amplitud, entre las que se distribuyen los 237 datos:

1 - 25, 26 - 50, 51 - 75, 76 - 100

Los límites inferiores de las clases son: 1, 26, 51 y 76, los límites superiores son: 25, 50, 75 y 100, mientras que las marcas de clase son: 13, 38, 63 y 88.

[SECCIÓN 2] **1.4 Las tablas de frecuencias**

En un estudio estadístico se utilizan las **tablas de frecuencias** para ordenar y resumir los datos estadísticos. La tabla 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 esas columnas depende principalmente del tipo de variable estudiada.

* 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* señala el total de datos.
* La **frecuencia absoluta** (*fi*) es el número total de **veces que se repite** un valor en cada clase.
* La **frecuencia relativa** (*hi*) es el cociente entre la *fi* en cada intervalo y el total de datos (*N*).
* La **frecuencia absoluta acumulada** (*Fi*) en cada clase es la suma de las frecuencias absolutas (*fi*) de los valores menores o iguales a él; solo tiene sentido para variables estadísticas cuantitativas.
* La **frecuencia relativa acumulada** (*Hi*) es el cociente entre la frecuencia relativa (*hi*) en cada intervalo y el total de datos (*N*).

Las tablas de frecuencias pueden recoger información de distribuciones **continuas** o **discretas**.

* Por ejemplo, en una clase de 25 alumnos se ha realizado una encuesta sobre las distancias entre su lugar de residencia y la capital de un país extranjero adonde han viajado. En esta ocasión se trata de una variable de tipo cuantitativo, por lo que, en este caso, la tabla contiene marcas de clase.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tabla de frecuencias que agrupa en cinco intervalos de clase los km recorridos** | | | | | | |
| **Número de kilómetros recorrido** | **Marca de clase**  ***(ci)*** | ***fi*** | ***hi*** | ***Fi*** | ***Hi*** | ***hi* %** |
| [0, 2000) | 1000 | 14 | 14/25 = 0,56 | 14 | 14/25 | 56 % |
| [2000, 4000) | 3000 | 5 | 5/25 = 0,2 | 14 + 5 = 19 | 19/25 | 20 % |
| [4000, 6000) | 5000 | 1 | 1/25 = 0,04 | 19 + 1 = 20 | 20/25 | 4 % |
| [6000, 8000) | 7000 | 2 | 2/25 = 0,08 | 20 + 2 = 22 | 22/25 | 8 % |
| [8000, 10 000) | 9000 | 3 | 3/25 = 0,12 | 22 + 3 = 25 | 25/25 | 12 % |
| Total |  | 25 |  | 25 | 25/25 = 1 | 100 % |

* Para obtener información acerca de los hábitos alimentarios de los estudiantes, se preguntó a cada uno de ellos por el tipo de alimento que consumía principalmente cada día.

En este ejemplo se trata de una variable de tipo cualitativo. Tomando las opciones de grupos nutricionales definidos por las sociedades nutricionales se tiene:

|  |  |  |
| --- | --- | --- |
| **Opciones para elegir el principal alimento consumido** | | |
| Carne, pollo, huevos y frutos secos | Frutas y verduras | Suplementos alimentarios |
| Pan, cereales, arroz y pasta | Aceites, grasas y dulces | Leche, yogur 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, yogur y queso; Carne, pollo, huevos y frutos secos; Frutas y verduras; Carne, pollo, huevos y frutos secos; Carne, pollo, huevos y frutos secos; Leche, yogur 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, yogur 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, yogur y queso; Pan, cereales, arroz y pasta; Carne, pollo, huevos y frutos secos; Leche, yogur y queso; Suplementos; Leche, yogur y queso; Leche, yogur y queso; Frutas y verduras.

Se puede observar que 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 una tabla de frecuencias, se obtiene:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tabla de frecuencias de los hábitos alimentarios de los estudiantes** | | | | | |
| **Principal alimento consumido**  ***xi*** | ***fi*** | ***hi*** | ***Fi*** | ***Hi*** | ***hi* %** |
| 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 % |
| 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, yogur y queso | 7 | 7/36 | 36 | 1 | 19,44 % |
| **Total** | **36** | **1** | **1** |  | **100 %** |

Con la tabla de frecuencias se identifica que la cuarta parte de los estudiantes consume principalmente carbohidratos, y que el mismo número de estudiantes consume principalmente grasas o suplementos.

[SECCIÓN 2] **1.5 Consolidación**

Actividades para consolidar lo que has aprendido en esta sección.

[SECCIÓN 1] **2 El muestreo**

Un estudio estadístico siempre está basado en el análisis de un conjunto de elementos seleccionados y no de sucesos aislados; a este conjunto se llama **población**. Si no es posible o conveniente examinar a toda la población que se estudia, es necesario seleccionar un subconjunto o **muestra**, escogida con cuidado para que sea representativa del total de la población. El proceso de selección se denomina **muestreo*.***

|  |  |
| --- | --- |
| **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. |

Por ejemplo, en Colombia, para conocer cuántas bolsas defectuosas se fabrican cada día, el departamento de control de calidad de las industrias del café elige unas muestras de cada lote de producción. En este caso la población es el conjunto de bolsas de café fabricadas y la muestra corresponde al número de bolsas que se toman para analizar y poder encontrar los resultados del estudio de unidades defectuosas en la muestra, e inferir, a partir de ese estudio, el número 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).

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 sacar conclusiones muy generales a partir de la observación de una sola parte de determinada población, ese error se denomina **error de muestreo**. Por otra parte, si se incurre en el error de extraer conclusiones hacia una población más grande a aquella en la que originalmente se hizo el muestreo, tal error se denomina **error de inferencia**.

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 bombillas, seleccionar 100 unidades como muestra de entre 10 000 unidades producidas tomando al azar cualesquiera 100 unidades, puede representar a la población total de bombillas. 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 diversos estratos y niveles educativos, de distintos géneros, etc., es decir que elegir cualesquiera 100 de entre cada 10 000 habitantes puede no ser un error de muestreo.

Un elemento fundamental para que un muestreo sea aleatorio es **que la muestra sea escogida al azar**, lo cual garantiza 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 docente 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 o probabilísticos.

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

Los métodos de muestreo probabilísticos son aquellos en los que **todos los individuos tienen la misma probabilidad de ser seleccionados para formar parte de una muestra**; solo los métodos de muestreo probabilístico nos aseguran la representatividad de la muestra extraída y, por ello, son los más recomendados. Dentro de los métodos de muestreo probabilísticos se encuentran los siguientes tipos:

[SECCIÓN 3] **2.1.1 El muestreo aleatorio simple**

El **muestreo aleatorio simple** utiliza el siguiente procedimiento: primero, a partir de un listado completo de toda la población, se realiza una asignación de un número a cada individuo de ella; posteriormente, se acude a algún medio de selección como sorteo, asignación de números aleatorios predeterminados en una lista o generados por computador, para seleccionar el número de individuos que indique el tamaño de muestra requerido, de este modo se completa el muestreo simple.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El muestreo aleatorio simple** |
| **Contenido** | 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, en este método es necesario 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; tiene la ventaja de que permite hacer cálculos de medidas de variación y dispersión muy fácilmente.

[SECCIÓN 3] **2.1.2 El muestreo aleatorio sistemático**

El **muestreo aleatorio sistemático** inicia del mismo modo que el muestreo aleatorio simple, es decir que se enumeran todos los miembros de la población y, posteriormente, se selecciona **un único elemento al azar**; la muestra se completa eligiendo de manera **periódica** el resto.

En el muestreo aleatorio simple se extraen *n* números que corresponden a la muestra, mientras que en el muestreo aleatorio sistemático 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 +* 2*k*), (*i +* 3*k*),..., es decir que de forma **periódica** se toman los individuos de *k* en *k*, donde *k* es el cociente entre el tamaño de la población *N* y 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 solo niños o solo 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.

[SECCIÓN 3] **2.1.3 El muestreo aleatorio estratificado**

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**.

Consiste en considerar categorías típicas diferentes entre sí (**estratos**) que poseen gran homogeneidad respecto a alguna característica. En cada uno de los estratos puede aplicarse un muestreo simple o sistemático para elegir los elementos de la muestra, pues cada estrato funciona independientemente.

El muestreo aleatorio estratificado inicia reconociendo **las características homogéneas** que garanticen representatividad en la muestra según el tipo de estudio poblacional que se va a realizar. Aunque en Colombia la palabra “estrato” se usa con frecuencia para notar el nivel de calidad de la zona en que se ubica una vivienda, también se puede estratificar según la profesión, el municipio de residencia, el género, el estado civil, etc. Sin embargo, como no todos los estratos tienen el mismo número 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 número de individuos y perjudica a los que tienen mayor número 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.

[SECCIÓN 3] **2.1.4 El muestreo aleatorio por conglomerados**

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, conocida como **conglomerado.**

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 dichas 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.

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.

Este tipo de muestreo es eficaz cuando la población es muy grande y dispersa; presenta la ventaja de que no se requiere tener un listado de todas las unidades poblacionales, sino solo el listado de conglomerados como unidades primarias de muestreo.

[SECCIÓN 2] **2.2 Consolidación**

SECCIÓN 1] **3 Las medidas estadísticas**

Luego de tener claridad respecto al tipo de variable al que se 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–, el siguiente paso cuando se trabaja con **variables cuantitativas** para estudiar los datos recogidos, es el cálculo de diferentes magnitudes características. Se definen diversas **medidas** capaces de resumir toda la información recogida en un pequeño grupo de valores. Estas medidas resumen, mejor conocidas como **medidas estadísticas,** permiten comparar una muestra con otras y dar una idea rápida de cómo se distribuyen los datos.

[SECCIÓN 2] **3.1 Las medidas de centralización**

Las **medidas de tendencia central** o valores centrales indican un valor central alrededor del cual se distribuyen los datos, es decir, el valor promedio de los datos. Los valores de posición central más utilizados son: de tamaño, la **media aritmética;**de posición, la**mediana,**y de frecuencia, la**moda***.*

[SECCIÓN 3] **3.1.1 La media aritmética**

Si se tiene una muestra de tamaño *N*, donde la variable estadística *x* toma los valores *x*1, *x*2,…, *xN*, la **media** **aritmética** se obtiene al adicionar todos los valores y dividir el resultado entre el número total de datos.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La media aritmética** |
| **Contenido** | La media aritmética de una variable estadística es el cociente entre la suma de todos los valores de la misma y el número de estos.  <<MA\_10\_06\_001>> |

La media puede ser un número que no tenga sentido en el contexto propuesto; tiene sentido en contextos de estudio de **variables continuas,** como el estudio de la variabilidad del peso o la estatura 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 arroja como resultado que las personas leen 1,5 libros o compran 2,5 pares de zapatos.

Por ejemplo, un estudiante ha obtenido los datos del peso de cuatro de sus compañeros, los cuales son 45 kg, 40 kg, 51 kg y 45 kg. Para establecer un valor medio del peso de sus cuatro compañeros, calcula la media aritméticaadicionando las cantidades que representan los pesos de sus compañeros (45 + 40 + 51 + 45). Luego, divide la suma entre el número total de individuos (181/4 = 45,25 kg), lo cual corresponde a la media aritmética o promedio aritmético del peso de los cuatro estudiantes.

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

Para encontrar la media aritmética de un conjunto de datos proveniente de una tabla de frecuencias, el cálculo consiste en adicionar los productos de las marcas de clase (*ci*) por las frecuencias relativas (*fi*) de cada clase y, luego, el resultado se divide por la cantidad de datos. Así, si se tiene *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:

<<MA\_10\_06\_002>>

Por ejemplo, la siguiente tabla de frecuencias muestra los resultados de una encuesta aplicada en una clase de 25 estudiantes 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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tabla de frecuencias sobre la mayor distancia que ha viajado cada estudiante** | | | | | | |
| **Número de kilómetros recorrido**  ***xi*** | **Marca de clase**  **(*ci*)** | **Frecuencia absoluta (*fi*)** | **Frecuencia relativa**  **(*hi = fi/n*)** | **Frecuencia absoluta acumulada (*Fi*)** | **Frecuencia relativa acumulada (*Hi = Fi/n*)** | ***hi* %** |
| [0, 2000) | 1000 | 14 | 14/25 = 0,56 | 14 | 14/25 | 56 % |
| [2000, 4000) | 3000 | 5 | 5/25 = 0,2 | 14 + 5 = 19 | 19/25 | 20 % |
| [4000, 6000) | 5000 | 1 | 1/25 = 0,04 | 19 + 1 = 20 | 20/25 | 4 % |
| [6000, 8000) | 7000 | 2 | 2/25 = 0,08 | 20 + 2 = 22 | 22/25 | 8 % |
| [8000, 10 000) | 9000 | 3 | 3/25 = 0,12 | 22 + 3 = 25 | 25/25 | 12 % |
| **Total** |  | 25 |  | 25 | 25/25 = 1 | 100 % |

Para el cálculo de la distancia media a la que ha viajado el grupo de estudiantes, se tiene que *N =* 25, mientras que *k =* 5, que indica el número de clases de la tabla. Ya con ello, se establece que la distancia media es:

<<MA\_10\_06\_003>>

Por lo tanto, el grupo de estudiantes ha viajado, en promedio, 3000 km.

[SECCIÓN 3] **3.1.2 La mediana**

Una medida importante de tendencia central es la **mediana** (*Med*) de un conjunto de datos; se define como una medida central que corresponde al valor que se encuentra **en el centro de la serie ordenada de datos**. Si el número de datos es impar, la mediana es el valor central del conjunto de *N* datos, que deja el mismo número de datos a la derecha que a la izquierda de la lista ordenada. Si el número *N* de datos es par, 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 de la lista ordenada.

Por ejemplo, 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”. El número 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 el número de datos es par o impar; en este caso como *N =* 20, la cantidad de datos es par.
* Debido a que el número de datos es par, se seleccionan los dos valores centrales de la lista ordenada y se halla 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 el 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:

<<MA\_10\_06\_004>>

Esta es la **mediana** del número de goles marcados por el equipo en la fase “Todos contra todos”.

La mediana también se puede calcular 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, mediante la siguiente ecuación:

<<MA\_10\_06\_005>>

Por ejemplo, en un estudio para identificar la talla para las chaquetas de los estudiantes de la promoción 2017, las estaturas de un grupo 100 de estudiantes se agrupan en cinco intervalos:

|  |  |  |
| --- | --- | --- |
| **Datos agrupados de las estaturas de los estudiantes** | | |
| **Estatura (en cm*)*** | ***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 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 *l*3= 166; la amplitud de la tercera clase en que se encuentra la mediana es *a*3 *=* 3, la frecuencia acumulada anterior a la clase tres en que está la mediana es *F*2= 30, la frecuencia absoluta de la clase 3 es *f*3 = 55 y la mitad de los datos es *N/*2 = 50. Entonces, el cálculo de la mediana queda:

<<MA\_10\_06\_006>>

Luego la mediana para el conjunto de datos agrupados es 167,09 cm de estatura; eso significa que 50 estudiantes tienen una estatura menor que 167,09 cm y los otros 50 tienen mayor estatura.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La mediana** |
| **Contenido** | La mediana es el valor que divide una serie de datos en partes iguales. El número de datos que queda por debajo y por arriba de la mediana son iguales. |

[SECCIÓN 3] **3.1.3 La moda**

La **moda** (*Mo*)de un conjunto de datos indica la tendencia de dicho conjunto, es decir, el dato que presenta mayor frecuencia. 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 del número 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 el mismo número de veces, se dirá que la población tiene una distribución bimodal, trimodal, etc.

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, por la edad que tenían cuando viajaron al extranjero por primera vez. Los resultados fueron los siguientes:

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

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

En el conjunto de edades de los hombres se puede evidenciar que 15 es la moda o el valor más repetido. En el conjunto de datos de las mujeres, por su parte, las edades 14 y 15 tienen la misma frecuencia, pues cada una se repite 7 veces. Cuando existen dos modas o dos datos que tiene valores de mayor frecuencia se dice que el conjunto de datos tiene una distribución bimodal.

Si se llega a estudiar la población total del salón sin estratificarla por género, la distribución de los datos de los 29 estudiantes 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 posmodal**– y la amplitud de las clases (*A*). El cálculo de la moda en este caso responde a la ecuación:

<<MA\_10\_06\_007>>

Por ejemplo, la siguiente tabla presenta los resultados respecto a la cantidad de dinero o salario, en pesos, que gana un grupo de 100 padres de familia.

|  |  |
| --- | --- |
| **Tabla de frecuencias del salario por día que ganan los padres de familia de los estudiantes** | |
| **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 |

En este caso la clase con frecuencia máxima es la tercera, es decir *i* = 3. Entonces, se obtiene:

<<MA\_10\_06\_008>>

De esta manera, la moda o la mayor frecuencia en los datos recolectados es $ 45 263, lo cual significa que la mayoría de los padres del grupo obtiene ese salario cada día.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Moda** |
| **Contenido** | La moda de un conjunto de datos es el valor que presenta mayor frecuencia. Se representa por *Mo*. |

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

Las **medidas de dispersión** son valores numéricos que miden la dispersión o variabilidad entre los datos.

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, razón por la cual en estos casos se hace necesario caracterizar mejor la población. Para lograrlo, es imprescindible estudiar y cuantificar la dispersión de los datos en la muestra o en la población, lo cual se conoce como **medidas de dispersión**.

Con el fin 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.

[SECCIÓN 3] **3.2.1 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 los valores máximo y mínimo que toma la variable estadística.

Cuando los datos se distribuyen de manera homogénea, el rango caracteriza bien la dispersión de la población. Sin embargo, como el rango depende exclusivamente de dos observaciones, 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 del número de lotes de un artículo, producidos en un turno de la fábrica. El número mínimo de lotes fue 23 y el máximo 89. Por lo tanto, el rango o recorrido será:

*R* = 89 – 23 = 66

Entonces, se sabe que el número de lotes por turno puede variar entre 66 datos, pero no se tiene más información acerca de la población.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El rango** |
| **Contenido** | El rango de un conjunto de datos es la diferencia numérica entre el dato mayor del conjunto y el dato menor. |

[SECCIÓN 3] **3.2.2 La desviación media**

Otra forma de caracterizar la dispersión de la muestra es comparar cada uno de los datos con el valor de la media, los valores que resultan de la media 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** respecto a la media aritmética de la población con *N* datos, que se representa por ***Dm****.*

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La desviación media** |
| **Contenido** | La **desviación media** es la media aritmética de las diferencias absolutas entre los valores de la variable y la media aritmética de la muestra.  <<MA\_10\_06\_009>> |

Por ejemplo, cuál es la desviación media de los goles que marcó un equipo, si el número de tantos que convirtió 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

Para calcularla, inicialmente se halla la media aritmética del conjunto de datos. Como la suma de estos es 20 y el número de datos es 20, entonces la media aritmética es 20/20 = 1.

|  |  |  |
| --- | --- | --- |
| **Desviaciones respecto a la media** | | |
| **Número 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.

Se observa 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.

[SECCIÓN 3] **3.2.3 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.

Por ejemplo, los 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

|  |  |  |
| --- | --- | --- |
| **Varianza de los goles marcados por el equipo** | | |
| **Número 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.**

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La varianza** |
| **Contenido** | La varianza es una medida que pretende establecer la cercanía de cada uno de los datos respecto a la media. |

[SECCIÓN 3] **3.2.4 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 conoce como 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.**

**Por ejemplo, si la desviación estándar** para el número de goles que marcó el equipo en las 20 fechas de la fase “Todos contra todos” tuvo como varianza ***σ2*** = 0,7 goles**2, la desviación estándar para el conjunto de datos será la raíz cuadra de 0,7, es decir, σ = 0,8367.**

Para el cálculo de la dispersión cuando los datos se encuentran agrupados, se sigue el procedimiento para hacer el cómputo de la media aritmética y, posteriormente, se calculan las dispersiones y sus cuadrados, con el fin de hallar tanto la varianza como la desviación estándar.

Por ejemplo, se ha obtenido la medida de la estatura de cien personas de una institución. Los resultados logrados, agrupados en intervalos, han sido:

|  |  |
| --- | --- |
| **Intervalos de los resultados obtenidos al medir la estatura** | |
| **Intervalo**  **(estatura en cm)** | **Frecuencia absoluta** |
| [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; a continuación, se utilizan las marcas de clase para encontrar la media de las estaturas medidas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Media de las estaturas medidas** | | | |
| **Intervalo**  **(estatura 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 | 2600 |
| [165, 170) | (165, 170) / 2 = 167,5 | 24 | 4020 |
| [170, 175) | (170, 175) / 2 = 172,5 | 27 | 4658 |
| [175, 180) | (175, 180) / 2 = 177,5 | 19 | 3373 |
| [180, 185) | (180, 185) / 2 = 182,5 | 8 | 1460 |
|  |  | **100** | **17 056** |

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

<<MA\_10\_06\_010>>

Para determinar la varianza y la desviación estándar se calculan las desviaciones y sus cuadrados.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Datos para calcular la varianza y la desviación estándar** | | | | |
| **Intervalo** | **Marca de clase**  ***xi*** | **Frecuencia absoluta**  ***fi*** | **(*xi* – *x̄*)2** | ***fi* (*xi* – *x̄*)2** |
| [155, 160) | (155, 160) / 2 = 157,5 | 6 | 170 | 1020 |
| [160, 165) | (160, 165) / 2 = 162,5 | 16 | 65 | 1040 |
| [165, 170) | (165, 170) / 2 = 167,5 | 24 | 9 | 216 |
| [170, 175) | (170, 175) / 2 = 172,5 | 27 | 4 | 108 |
| [175, 180) | (175, 180) / 2 = 177,5 | 19 | 48 | 912 |
| [180, 185) | (180, 185) / 2 = 182,5 | 8 | 143 | 1144 |
| ***N*** | | **100** |  | **4440** |

Con base en lo anterior, se deduce que la varianza y la desviación estándar son:

<<MA\_10\_06\_011>>

<<MA\_10\_06\_012>>

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La desviación estándar** |
| **Contenido** | La desviación estándar, notada como **σ**, es la raíz cuadrada positiva de la varianza. |

[SECCIÓN 3] **3.2.5 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 que resulta de dividir 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.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El coeficiente de variación** |
| **Contenido** | El coeficiente de variación se calcula dividiendo la desviación estándar entre la media.  <<MA\_10\_06\_013>> |

Por ejemplo, la nota media en matemáticas del curso 10A ha sido 7,2 y su desviación estándar 0,36. En cambio, en el curso 10B la nota media ha sido 6,4 y su desviación estándar 0,45. Calculando el coeficiente de variación de las dos clases se tiene:

<<MA\_10\_06\_014>>

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

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

**Las medidas de posición no central sirven para dividir** un conjunto de valores de una variable estadística en subconjuntos con el mismo número de valores, **y para establecer el comportamiento de los datos en diferentes puntos.**

**Estas medidas son complementarias de los valores centrales y ayudan 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.

[SECCIÓN 3] **3.3.1 Los cuartiles**

Los **cuartiles, como medida de posición no central,** dividen el conjunto de datos en cuatro grupos iguales.

Los cuartiles son tres, *Q*1, *Q*2 y *Q*3. El primer cuartil es *Q*1; este deja por debajo el 25 % de los datos. El segundo cuartil es *Q*2 y coincide con la mediana. El tercer cuartil es *Q*3 y deja por debajo el 75 % de los datos.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_06\_CO\_IMG01 |
| **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  En la gráfica, escribir el símbolo de % separado un espacio de la cantidad, así: 0 % 25 % … |
| **Pie de imagen** | Representación de los cuartiles; en la imagen se puede ver que entre el primer y el tercer cuartil, es decir, entre *Q*1 y *Q*3, se encuentra ubicada la mitad de la población. |
| **Ubicación del pie de imagen** | Inferior |

Por ejemplo, en una fábrica, el número de lotes producidos por cada una de las 15 máquinas de manufactura es:

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 el número de datos es *N* = 15:

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

**A continuación, se puede** localizar la mediana *Q*2 y luego calcular las posiciones de *Q*1y de *Q*3 como medianas de la primera y segunda mitad de los datos.

Como el número de datos es impar, la mediana corresponde al dato que está en el centro de la lista ordenada y que deja a cada lado igual número de datos, en este caso 7:

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

Entonces, *Q*2 = 19 que **corresponde a la mediana o segundo cuartil.**

La primera mitad de los datos ordenados contiene 7 datos y la mediana de ese conjunto es el primer cuartil o *Q*1:

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

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 *Q*3:

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

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

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

[SECCIÓN 2] **3.4 Los diagramas de caja**

El **diagrama de caja** es un resumen gráfico en el que se describen varias de las características más destacadas de un conjunto de datos, como los datos inusuales del estudio y la dispersión de los datos respecto a los cuartiles.

Para crear un diagrama de caja de un conjunto de datos, son necesarios cinco datos esenciales:

* La **mediana o** *Q*2, para ubicar el valor central de las observaciones.
* Los **cuartiles** *Q*1 y *Q*3 para ubicar el 50 % central del conjunto de datos.
* La diferencia entre *Q*3 y *Q*1 se llama rango intercuartílico y se denota como IQR.
* Las observaciones individuales **mínima (Mín) y máxima (Máx)** para indicar la dispersión.

Estos cinco valores, 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, por ello es de gran importancia hacer el cálculo del IQR.

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

1,5 ∙ IQR

El cálculo de los valores atípicos se lleva a cabo calculando marcas atípicas inferiores y superiores. Para valores atípicos inferiores, la marca se calcula realizando la sustracción entre *Q*1 y 1,5 ∙ IQR, es decir:

*Q*1 ‒ (1,5 ∙ IQR)

Para valores atípicos superiores, la marca se calcula haciendo la adición entre *Q*3 y **(1,5 ∙ IQR),** es decir, calculando:

*Q*3 + (1,5 ∙ IQR)

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, con la 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., *Q*1, *Q*2, *Q*3y Máx. Luego se hace una caja alrededor de *Q*1 y *Q*3, segmentos en los intervalos (Mín., *Q*1) y (*Q*3, Máx.) y se marcan como puntos aislados los datos atípicos. De ese modo se completa el diagrama.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Los diagramas de caja** |
| **Contenido** | Los **diagramas de caja** son gráficos basados en **cuartiles,** 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** *Q*1, *Q*2 (o **mediana**) y *Q*3 y sobre la existencia de **valores atípicos** y **simetría de la distribución**. |

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

Lotes que se han producido en el turno de la mañana 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 elaborado en el turno de noche:

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cálculo de cuartiles para el turno de la mañana siguiente** | | | | | |
| **Valor**  **mínimo** | ***Q*1** | ***Q*2** | ***Q*3** | **Valor**  **máximo** | **IQR**  **(*Q*3 ‒ *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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cálculo de cuartiles para el turno de la noche siguiente** | | | | | |
| **Valor mínimo** | ***Q*1** | ***Q*2** | ***Q*3** | **Valor máximo** | **IQR**  **(*Q*3 ‒ *Q*1)** |
| 7 | 15 | 19 | 27 | 32 | 12 |

que también puede confirmarse en la lista ordenada de datos:

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

El IQR del turno de la mañana es 80 – 30 = 50 lotes producidos, mientras que el IQR del turno de la noche es: 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 traza un rectángulo o caja, cuyos lados izquierdo y derecho (o inferior y superior si la caja es vertical) van del primero al tercer cuartil. Por lo tanto, el ancho (o altura, si la caja es vertical) de la caja es la amplitud de 50 % de los datos centrales.

Si un punto se encuentra **a más de 1,5 IQR** de los extremos de la caja, se considera que es una **observación atípica**, por lo tanto, se debe identificar si en el conjunto de datos alguno es atípico.

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

*Q*1 ‒ 1,5 ∙ IQR = 30 ‒ 1,5(50) = ‒45

Mientras que la marca inferior es:

*Q*3 + 1,5 ∙ IQR = 80 + 1,5(50) = 155

En consecuencia, 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:

*Q*1 ‒ 1,5 ∙ IQR = 15 ‒ 1,5(12) = ‒3

*Q*3 + 1,5 ∙ IQR = 27 + 1,5(12) = 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.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_06\_CO\_IMG02 |
| **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!!! Se cambia un dato en la tabla, el 30.** |
| **Pie de imagen** | En la imagen se observan dos **diagramas de caja** en un mismo gráfico para comparar el número de lotes realizados por el turno de la mañana y el turno de la noche. |
| **Ubicación del pie de imagen** | Inferior |

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

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El 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 provee información sobre los valores **mínimo** y **máximo**, los **cuartiles** *Q*1, *Q*2 (o **mediana**) y *Q*3 y sobre la existencia de **valores atípicos** y **simetría de la distribución**. |

[SECCIÓN 2] **3.5 Consolidación**

**[SECCIÓN 1] 4. La combinatoria**

Para realizar el cálculo de probabilidades se necesita 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 el **orden** o no en que se pondrán los elementos que hacen parte del espacio muestral, en cuyo caso se tienen:

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.

En una clase se quiere elegir dos representantes para un concurso. Como en ocasiones un estudiante puede faltar, se escogerá el que concursará y un suplente.

En una primera votación, salen elegidos tres candidatos: María, Laura y Pedro.

En una segunda votación, se selecciona a uno de 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 | MA\_10\_02\_CO\_IMG05 |
| Descripción | 4º ESO/Matemáticas/La 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**.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **El diagrama de árbol** |
| **Contenido** | Un **diagrama de árbol** es una **representación gráfica** útil para contar todas las posibles maneras de **combinar** un número de elementos finitos **de uno o más conjuntos**. |

* 1. Para el primer nombre, como concursante, existen tres posibilidades:

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código | MA\_10\_02\_CO\_IMG06 |
| Descripción | 4º ESO/Matemáticas/La 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 un **diagrama de árbol** de los nombres de los **posibles concursantes**. |
| Ubicación del pie de imagen | Inferior |

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

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| Código | MA\_10\_02\_CO\_IMG07 |
| 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 un **diagrama de árbol** que muestra **todas las posibilidades** que tienen los estudiantes de ser elegidos como suplentes. |
| Ubicación del pie de imagen | Inferior |

Así, en total hay 3 ∙ 2 = 6 posibilidades.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Factorial de un número** |
| **Contenido** | Se llama factorial de un número natural *n* al producto de los *n* factores consecutivos desde *n* hasta 1. Se denota por *n*! |

El **factorial** de un número **entero positivo** es el producto de todos los números naturales anteriores o iguales a él. Se escribe ***n*!** y se lee “***n* factorial**”.

*n*! = *n* ∙ (*n* ‒ 1) ∙ (*n* ‒ 2) ∙… ∙ 1

El factorial de 0 se define como:

0! = 1

Por ejemplo, si se calcula el factorial de 8:

8! = 8 · 7 · 6 · 5 · 4 · 3 · 2 · 1 = 40 320

Si se calcula el factorial de 6:

6! = 6 · 5 · 4 · 3 · 2 · 1 = 720

**[SECCIÓN 2] 4.1 Las permutaciones**

Las permutaciones son las distintas formas de agrupar los elementos de un conjunto teniendo en cuenta lo siguiente:

* Influye el **orden** en que se presentan los datos.
* **Intervienen todos los elementos** del conjunto inicial.

Las permutaciones pueden realizarse sin repetición y con repetición.

**[SECCIÓN 3] 4.1.1 Las permutaciones sin repetición**

De cuántas formas se pueden ordenar en una estantería los siguientes cinco cómics:

* 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 (cómics) que se quieren ordenar. En el ejemplo es claro que **en una misma muestra no se pueden repetir** dos elementos, pues eso equivaldría a ubicar dos veces el mismo cómic.

El número de posibilidades será:

* Para la **primera** posición: **cinco posibilidades**. Se puede escoger de entre todos el que se quiere colocar primero.
* Para la **segunda** posición: **cuatro posibilidades (5 − 1)**. Ahora quedan solo cuatro cómics para escoger el que se ubicará en segundo lugar.
* Para la **tercera** posición: **tres posibilidades (5 − 2)**. Ahora solo quedan tres cómics para escoger.
* Para la **cuarta** posición: **dos posibilidades (5 − 3)**. Solo se tienen dos cómics sin ordenar. Para la **quinta** posición: **una posibilidad (5 − 4)**. Solo queda un cómic por ordenar. Ahora ya están todos ordenados.

Total: 5 · 4 · 3 · 2 · 1 = 120 posibilidades.

La novedad de este ejemplo está en que **todos los elementos** del conjunto **intervienen en cada muestra**. Es una **muestra ordenada sin repetición de tantos elementos como el conjunto inicial**. Estos casos reciben el nombre de **permutaciones sin repetición**.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Las permutaciones sin repetición** |
| **Contenido** | 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 el número de posibles permutaciones sin repetición se aplica la ecuación:  *Pn* = *n*! = *n* × (*n* ‒ 1) × (*n* ‒ 2) × … × 1  Se debe resaltar 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. |

Por ejemplo, ¿cuántas permutaciones o anagramas se pueden hacer con las letras de la palabra “fútbol”?

Se trata de ordenar un total de seis letras distintas, sin repetir ninguna, de todas las formas posibles.

El número de posibilidades será:

* Para la **primera** posición: **seis posibilidades**. Se puede escoger entre todas las letras (F, Ú, T, B, O, L) cuál se coloca en la primer posición.
* Para la **segunda** posición: **cinco posibilidades (6 − 1)**.
* Para la **tercera** posición: **cuatro posibilidades (6 − 2)**.
* Para la **cuarta** posición: **tres posibilidades (6 − 3)**.
* Para la **quinta** posición: **dos posibilidades (6 − 4).**
* Para la **sexta** posición: **una posibilidad (6 − 5)**.

Total: 6! = 6 × 5 × 4 × 3 × 2 × 1 = 720 posibilidades.

Si se aplica la **fórmula** de **permutaciones sin repetición**, se llega al mismo resultado para ordenar una muestra de seis elementos:

*P*6 = 6! = 6 ∙ (6 ‒ 1) ∙ (6 ‒ 2) ∙ (6 ‒ 3) ∙ (6 ‒ 4) ∙ (6 ‒ 5) = 6 × 5 × 4 × 3 × 2 × 1 = 720

**[SECCIÓN 3] 4.1.2 Las permutaciones con repetición**

El código secreto de seis cifras para abrir una caja fuerte está formado por las siguientes cifras: 1, 1, 1, 1, 5, 5, pero no se conoce el orden correcto. ¿Cuántas posibilidades se tienen para probar cuál es el código?

Se trata de construir todos los códigos de seis cifras posibles, utilizando cuatro veces el número 1 y dos veces el número 5.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_02\_CO\_IMG08 |
| **Descripción** | 4º ESO/Matemáticas/La combinatoria/ Las 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** | **Para construir todos los códigos posibles de seis cifras de una forma ordenada se utiliza un diagrama de árbol.** |
| **Ubicación del pie de imagen** | Inferior |

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

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **Las permutaciones con repetición** |
| **Contenido** | 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 modo que en cada grupo formado por los ***n*** elementos, **cada elemento aparece el número de veces indicado** y que **dos grupos se diferencien solo en el orden de colocación.**  La ecuación que permite calcular el número de resultados que se obtienen al permutar de esta manera es:  <<MA\_10\_06\_015>> |

En el ejemplo del código secreto de seis cifras para abrir una caja fuerte, se obtiene el mismo resultado si se aplica la ecuación para calcular el número **de permutaciones con repetición, donde** *n =* 6, *a* = 4 y *b* = 2:

<<MA\_10\_06\_016>>

Si por ejemplo se tienen 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.

Se debe calcular el número 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, por lo tanto:

<<MA\_10\_06\_017>>

Por consiguiente, se tienen 210 maneras de disponer la ropa en el armario.

**[SECCIÓN 2] 4.2 Las combinaciones**

En ocasiones **no interesa el orden** en que se van ubicando los elementos de la muestra de un conjunto. Cuando 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** **repetición** o **con repetición de los elementos del conjunto**.

**[SECCIÓN 3] 4.2.1 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**; por consiguiente, para **que dos grupos sean diferentes** tienen que tener **al menos un elemento distinto**.

La **fórmula** que permite calcular el número de resultados que se obtienen al combinar sin repetición tales muestras es:

<<MA\_10\_06\_018>>

El número resultante se conoce como **coeficiente binomial** o **número combinatorio**, porque expresa el número de combinaciones sin repetición de ***n*** elementos de una muestra total de ***m***.

Por ejemplo, seis amigos (Luis, Ruth, Ana, Juan, Pepe y Mary) tienen dos entradas para ir a un concierto y quieren repartirlas entre sí. Ahora, deben decidir quiénes de los seis irán al concierto; ¿cuántas posibles parejas se pueden conformar para hacer uso de las entradas?

**El orden no importa porque da igual conseguir la primera entrada que la segunda.**

Del conjunto de seis elementos se preparan todas las muestras ordenadas posibles de dos elementos. Para ello, se utilizan las iniciales de los nombres para representar a cada uno de los amigos, para así identificar claramente los elementos del conjunto.

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_02\_CO\_IMG09 |
| **Descripción** | 4º ESO/Matemáticas/La combinatoria/Las 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) |
| **Pie de imagen** | El diagrama de árbol y la tabla resumen las **combinaciones sin repetición.** |
| **Ubicación del pie de imagen** | Inferior |

De las 30 muestras ordenadas que ilustra el diagrama de árbol, se observa que hay muestras que contienen a las mismas personas, aunque en distinto orden.

En la tabla se han indicado ubicándolas 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 resumen, para obtener el número total de muestras no ordenadas o subconjuntos de ***n*** elementos escogidos entre los ***m*** de un conjunto, se debe:

1. Se calcula el total de muestras ordenadas, es decir, las **variaciones sin repetición** de ***n*** elementos escogidos entre un total de ***m***:

En el ejemplo, serían las variaciones sin repetición de dos elementos entre seis:

<<MA\_10\_06\_019>>

1. Se busca el total de **permutaciones sin repetición de los *n* elementos** de cada muestra:

En el ejemplo: *P*2 = 2!

1. Se realiza **el cociente** de estos dos números:

<<MA\_10\_06\_020>>

Las muestras no ordenadas o subconjuntos de 2 elementos del conjunto de 6 elementos son 15.

Para indicar el número de subconjuntos de dos elementos de un conjunto de seis en total, se utiliza el símbolo de **coeficiente binomial** o **número combinatorio**, ya que expresa el número de combinaciones sin repetición. En consecuencia:

<<MA\_10\_06\_021>>

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

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_02\_CO\_IMG10 |
| **Descripción** | 4º ESO/Matemáticas/La combinatoria/Las 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) |
| **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 tanto, para averiguar el número de triángulos posibles se aplica la ecuación con *n* = 3 y *m =* 4:

<<MA\_10\_06\_022>>

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 trata de una muestra no ordenada de tres elementos, escogidos entre un total de cinco, para los que el número de combinaciones posibles será:

<<MA\_10\_06\_023>>

Así que el máximo posible de triángulos que se pueden dibujar utilizando los vértices de un pentágono son 10 triángulos.

**[SECCIÓN 3] 4.2.2 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**. Por lo tanto, para que **dos grupos sean diferentes** tienen que tener **al menos un elemento diferente**.

La **fórmula** que permite calcular el número de resultados que se obtienen al combinar de esta manera es:

<<MA\_10\_06\_024>>

Por ejemplo, en una tienda venden cuatro tipos diferentes de papas fritas (con sabores a suero costeño, a jamón, a salsa BBQ y a queso). Si se quiere comprar tres paquetes de papas, ¿de cuántas formas se pueden elegir?

Se trata de una muestra de tres elementos a elegir de un conjunto de cuatro, donde se cumple lo siguiente:

* El **orden NO importa**. Da igual que se elijan dos paquetes de papas con sabor a suero costeño y un paquete con sabor a jamón, que un paquete con sabor a jamón y dos con sabor a suero costeño.
* **No entran todos los elementos**, pues de los cuatro posibles solo se eligen tres.
* **Sí se repiten** los elementos, ya que se puede elegir más de un paquete del mismo tipo.

Esto quiere decir que el conteo que se va a realizar debe establecer el número máxima de **combinaciones con repetición de 3 elementos, dentro de un conjunto de 4 elementos.**

**Para construir el diagrama de árbol que corresponde a las opciones posibles para elegir tres paquetes de papas, se asignan las denominaciones PS a un paquete de papas con sabor a suero costeño, PJ a un paquete de papas sabor a jamón, PB a un paquete de papas sabor a BBQ y PQ a un paquete de papas sabor a queso.**

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_02\_CO\_IMG11 |
| **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 adiciones 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 en forma equivalente:

<<MA\_10\_06\_025>>

Por otra parte, **las posibles** combinaciones con repetición de **orden 2** resultan ser el mismo número 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**.

<<MA\_10\_06\_026>>

Finalmente, **las posibles** combinaciones con repetición de **orden 3** resultan ser el mismo número que aparece al construir **las combinaciones sin repetición con un elemento más**, es decir:

<<MA\_10\_06\_027>>

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.

**[SECCIÓN 2] 4.3 La probabilidad condicionada**

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

Pero también puede ocurrir que los sucesos dependan entre sí. Por ejemplo, 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*”.

|  |  |
| --- | --- |
| **Destacado** | |
| **Título** | **La probabilidad condicionada** |
| **Contenido** | Si *A* y *B* son dos sucesos de un experimento aleatorio y *A* no es un suceso imposible, es decir, *p*(*A*) ≠ 0, la **probabilidad** de *B* **condicionada** a *A* se define como:  <<MA\_10\_06\_034>>  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: número de casos favorables de *B* entre los que también son favorables a *A*.  <<MA\_10\_06\_035>> |

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

* 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.   Si *A* y *B* son independientes: *p*(*A* ∩ *B*) = *p*(*A*) × *p*(*B*).   * 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*). |

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

* *A*: “obtener un 6”.
* *B*: “obtener cara”.

Para ello, se procede con los siguientes pasos:

* + 1. Para analizar la independencia de los sucesos inicialmente se define el **espacio muestral** del experimento, que se condensa en el diagrama de árbol, en el que se hace la convención C = cara y que X = sello

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_02\_CO\_IMG13 |
| **Descripción** | 4º ESO/Matemáticas/La probabilidad/La 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) |
| **Pie de imagen** | Espacio muestral para el experimento aleatorio de lanzar una moneda y después tirar un dado. |
| **Ubicación del pie de imagen** | Inferior |

El espacio muestral es *S* = {(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.

* + 1. 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)}

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

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

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

*p*(*A* ∩ *B*) = 1/12

* 1. Se comprueba que 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*) = *p*(*A) · p*(*B*)

1/12 = 1/12

Como los resultados de las probabilidades son iguales, queda comprobado que los sucesos *A* y *B* son sucesos independientes.

En ocasiones, no es tan evidente la independencia de los 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.**

**[SECCIÓN 3] 4.3.1 Las 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 en 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** | MA\_10\_02\_CO\_IMG14 |
| **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) |
| **Pie de imagen** | Tabla de contingencia y probabilidades presentes en ella; se puede observar que cada individuo debe quedar clasificado en una casilla y solo en una. |
| **Ubicación del pie de imagen** | Inferior |

Por ejemplo, en una clase de 30 estudiantes, 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, se van a indicar los datos en una **tabla de contingencia.**

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_02\_CO\_IMG15 |
| **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 estudiantes 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 |

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

|  |  |
| --- | --- |
| **Imagen (fotografía, gráfica o ilustración)** | |
| **Código** | MA\_10\_02\_CO\_IMG16 |
| **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** |
| **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 estudiantes, 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

El mismo resultado se obtiene al aplicar la ecuación de probabilidad de sucesos dependientes.

*p*(*M* ∩ *T*) = *p*(*M*) × *p*(*T/M*) = 19/30 × 7/19 = 7/30

Para 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 lo 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 idéntico resultado.

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

[SECCIÓN 1] **Competencias**

Pon a prueba tus capacidades y aplica lo aprendido con este recurso.

[SECCIÓN 1]**Fin de tema**

|  |  |
| --- | --- |
| **Mapa conceptual** | |
| **Código** | XX\_00\_00\_REC00 |
| **Título** | Mapa conceptual |
| **Descripción** | Mapa conceptual del tema La estadística y la probabilidad |

|  |  |
| --- | --- |
| **Evaluación: recurso nuevo** | |
| **Código** | XX\_00\_00\_REC00 |
| **Título** | Evaluación |
| **Descripción** | Actividad que permite evaluar los conocimientos sobre el tema La estadística y la probabilidad |

|  |  |  |
| --- | --- | --- |
| **Webs de referencia** | | |
| **Código** | XX\_00\_00\_REC00 | |
| **Web 01** | *Ejemplo de cómo construir una tabla de frecuencias en Excel* | [VER](https://youtu.be/bKK0kXzwpgs) |
| **Web 02** | *Estudia más sobre la estadística unidimensional al observar algunos ejercicios resueltos* | [VER](http://www.fuenterrebollo.com/Economicas2013/unidimensional-ejercicios.pdf) |
| **Web 03** | *Observa algunos ejercicios resueltos de probabilidad condicionada* | [*URL*](http://www3.uah.es/jmmartinezmediano/Segundo%20CS/MCCSS%20Tema%2009b%20Problemas%20de%20probabilidad%20condicionada.pdf) |