Elementos de Estadística

Pedro Vicente Esteban Duarte

Presentación

La estadística es una rama de las matemáticas que se ocupa de la recolección, procesamiento y análisis de la información para hacer inferencias que ayudan en la toma de decisiones en diversos campos del conocimiento, como la Sociología, la Psicología, la Política, la Física, la Química, entre otras.

Su estudio y comprensión se hace cada vez más necesario desde los primeros años de escolaridad, pues fortalece el pensamiento aleatorio y sistemas de datos, que se ocupa de los problemas en los que para su solución es necesario pensar en términos de probabilidades.

El módulo tiene los siguientes objetivos:

Objetivo general

Estudiar los conceptos básicos sobre la recolección, procesamiento e inferencia de datos cualitativos y cuantitativos.

Objetivos específicos

- Clasificar conjuntos de datos.
- Encontrar los estadísticos (moda, media, mediana, rango, varianza, desviación estándar) para conjuntos de datos.
- Agrupar datos de acuerdo con sus características.
- Realizar tablas de frecuencias y calcular la frecuencia relativa y acumulada.
- Hacer inferencias a partir de las tablas de frecuencias.
- Utilizar las reglas de conteo en el cálculo de probabilidades.

Los conceptos expuestos y los ejercicios planteados permiten comprender conceptos fundamentales de la estadística básica.

El tiempo estimado para la solución del taller es de tres (3) horas.

En su estudio y solución le deseamos muchos éxitos.

1. Recolección de la información

Desde la antigüedad, se tienen registros de recolección de información con diversos fines: censar una población, militares, rendimiento de las cosechas, cría del ganado, entre otros. De ello se tienen registros en arcilla cocida y papiros, en civilizaciones como la de Babilonia, de Egipto, de Grecia, entre otras[1].

En el Siglo $XVIII$, los datos relacionados con el censo y otras actividades propias del comercio se denominaron “datos del estado”. El término *estadística* fue introducido, en el siglo XIX, por el inglés John Sinclair (1754-1835). A través de la historia, son muchos los matemáticos que han contribuido en la construcción de de la teoría y en el análisis de los resultados de las aplicaciones que se hacen de esta rama del conocimiento a diversas ciencias[2].

La recolección de la información parte de los intereses que se tengan para realizar un estudio, en relación con una población, el tipo de variables y de los datos que se quieren observar, la cantidad de datos que se quieren recolectar, la capacidad de procesamiento que se tenga, los recursos humanos y financieros con que se cuente para llevar a buen término el análisis de los datos.

En periódicos, revistas, libros de texto, radio y televisión se encuentran muchos ejemplos de la forma de recolectar la información, del procesamiento de la misma y los fines con la que se usa.

1.1. Tipos de datos

Los estudios, en los que se requiere del análisis estadístico, comienzan por la recolección de la información, sobre una o varias características de interés de un conjunto de datos. La unidad básica de información es el *dato* y el conjunto de datos conforman la *muestra* que va a ser analizada. Con la muestra se construyen los *estadísticos*, que son los valores, que permiten hacer inferencias sobre la población de la que se extraen los datos.

Población: cualquier conjunto de referencia sobre el que se hace una observación. Los estudiantes de un colegio, los compañeros de clase, las notas obtenidas en un examen de matemáticas, los hinchas de un

[1] Best, Joel (2001). Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists. University of California Press. ISBN 0-520-21978-3.

[2] Hacking, Ian (1990). The Taming of Chance. Cambridge University Press. ISBN 0-521-38884-8.

equipo de fútbol que van a ver un partido al estadio, las personas que utilizan una determinada marca de zapatos, entre otros. El tamaño de la población, se denota generalmente por la letra **N** mayúscula.

Variable de interés: es la característica que interesa observar en una población. En una misma población un investigador puede estar interesado en observar varias características, cada una de esas características es una variable. En un colegio, puede ser de interés las notas de los estudiantes en cada una de las áreas de estudio: Matemáticas, Sociales, Ciencias, entre otras.

Tipos de variables: las variables pueden ser *cualitativas* o *cuantitativas*. Las *cualitativas* son aquellas que describen cualidades, alto, bajo, rojo, verde, bueno, malo, regular, primero, segundo, entre otras. Las *cuantitativas* son las que como resultado se obtienen valores numéricos con los que tiene sentido realizar operaciones aritméticas: kilogramos, metros, el resultado de la nota en un examen, la estatura de una persona, el dinero que tiene alguien en el bolsillo, entre otras.

Dato: cualquier resultado de la observación de la variable de interés. Por ejemplo, para las notas de matemáticas en un determinado colegio, si al encuestar un alumnos dice que obtuvo un 7, ese es un dato para la variable observada.

Muestra: cualquier subconjunto de la población. Si la población son los estudiantes de un colegio, los siguientes subconjuntos, dependiendo el estudio que se quiera realizar, pueden ser una muestra: los estudiantes que llegaron tarde el primer lunes del último mes, los estudiantes del grado octavo, los que llegan a pie al colegio, los que viven en barrios distintos en los que esta ubicado el colegio, entre otros. El tamaño de la muestra se denota por la letra **n** minúscula.

Estadístico: es el resultado de cualquier procedimiento que se haga con los datos.

Ejemplo

Un colegio, en el grado octavo, cuenta con 60 estudiantes, se quieren observar algunas características de este grupo. Para ello, se entrevista a 10 estudiantes, seleccionados al azar y, se les pregunta por la nota en español y el deporte que practicaron en el último período académico. Los resultados obtenidos fueron los siguientes:

Nota de español: 7, 8, 9, 6, 5, 7, 7, 9, 5, 6.

Deporte que práctica: voleibol, fútbol, patinaje, atletismo, fútbol, fútbol, voleibol, atletismo, karate, fútbol.

La **población** objeto de estudio son los alumnos del grado octavo de un colegio en particular. Note que el tamaño de la población es de 60, por lo tanto $N = 60$.

Variables de interés:

1. De acuerdo con el enunciado del ejercicio, la nota de español del grupo objeto de estudio, es una variable *cuantitativa*. Las notas se pueden sumar, sacar promedios, entre otros.

2. De acuerdo con el enunciado del ejercicios, el deporte que practican los alumnos del grupo objeto de estudio, es una variable *cualitativa*. No tiene sentido sumar, sacar promedios, entre otras operaciones, con los deportes reportados.

La *muestra* son los diez estudiantes seleccionados al azar, su tamaño es: $n = 10$.

Los **datos** son cada una de las respuesta dadas por los estudiantes del curso.

Para la variable *cuantitativa*, cada una de las notas reportadas por los alumnos entrevistados corresponden a los datos.

Para la variable *cualitativa*, cada uno de los deportes que los alumnos reportaron que habían practicado durante el período académico corresponden a los datos.

Estadísticos para la variable cuantitativa:

Si se organizan de menor a mayor las notas obtenidas en español se tiene lo siguiente:

Notas de español: 5, 5, 6, 6, 7, 7, 7, 8, 9, 9.

A partir de esta actividad se puede decir que la diferencia en la nota de español es de 4 unidades. Este valor se obtiene de realizar una operación aritmética con los datos recolectados, por lo tanto es un **estadístico**. Otro puede ser el 7, que es la nota que más se repite en la muestra. Como se puede observar, a partir de los datos de una muestra se pueden obtener varios estadísticos.

Estadísticos para la variable cualitativa:

Si se organizan en orden alfabético los deportes practicados, por el grupo de estudiantes observados, se tiene:

atletismo, atletismo, fútbol, fútbol, fútbol, fútbol, karate, patinaje, voleibol, voleibol

El deporte más practicado por los estudiantes observados es el fútbol, que sirve como un estadístico de la muestra. Esto se obtuvo a partir de observar características de la muestra recolectada.

Inferencias para la variable cuantitativa:

Al observar los resultados y ver que la nota que más reportaron los estudiantes es 7 y que hay más estudiantes por debajo de esa nota que por encima, se infiere que los estudiantes necesitan refuerzo en español para los temas de ese período.

Inferencias para la variable cualitativa:

Es posible que al ver los resultados del área de deportes decida promover más la práctica de los deportes menos populares.

Como se puede observar, recolectar y organizar la información sobre poblaciones y variables de interés, le ayuda a las personas que toman decisiones a estructurar y asignar, de una mejor manera, los recursos que tienen a su disposición, para el bien de una comunidad o para el mejoramiento de los procesos que se siguen en una empresa.

2. Algunos estadísticos básicos

Como se mencionó anteriormente, un *estadístico* es el resultado de una operación (datos cualitativos) u observación (datos cuantitativos) que se hace con los datos de una muestra aleatoria. Algunos de ellos son: la media o promedio, la mediana, la moda, el rango, la varianza, la desviación estándar, entre otros.

2.1. La media o promedio $\overline{x}$ de una muestra

Si se tiene una muestra cuantitativa $x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$ de tamaño n, la media o promedio de la muestra está dada por:

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \ldots + x_i + \ldots + x_n}{n}$$

La media $(\overline{x})$ permite hacer inferencias acerca de los datos de la población de la que proviene la muestra, puesto que no siempre es posible tener todos los datos de la población.

Por ejemplo, en un grupo de octavo hay 30 estudiantes que presentaron el examen del período de matemáticas, se seleccionaron aleatoriamente 6 estudiantes y se les preguntó por su resultados, obteniendo lo siguiente:

$$8, 6, 7, 8, 8, 5$$

El promedio de la muestra es:

$$\overline{x} = \sum_{i=1}^{6} \frac{x_i}{6} = \frac{5+6+7+8+8+8}{6} = 7$$

Se puede interpretar que $\overline{x} = 7$, es un resultado cercano al promedio de las notas de toda la población observada.

Algunas propiedades de la media

Algunas de las propiedades de la media son las siguientes:

a. La media $(\overline{x})$ es un valor que se encuentra entre el dato menor y el dato mayor.

b. La media es muy sensible a datos pequeños. Si en nuestro ejemplo, se hubieran seleccionado solamente los estudiantes que perdieron el examen la media no sería un valor representativo de las notas del grupo de estudiantes.

c. La media es muy sensible a datos grandes. Si en nuestro ejemplo, se hubieran seleccionado solamente los estudiantes que ganaron el examen la media no sería un valor representativo de las notas del grupo de estudiantes.

Ejercicio

En una encuesta en un barrio, se seleccionaron 7 tiendas al azar y se les preguntó por el precio del jabón en barra. Los resultados fueron los siguientes: 600, 615, 630, 620, 625, 610, 630 pesos por barra. El promedio de los precios de la barra de jabón es

a. $ 635

b. $ 618.57

c. $ 617.20

d. $ 611

2.2. La mediana M_e de una muestra

La mediana es el dato que se encuentra en el centro de una muestra ordenada, note que para hallarla, los datos deben ser cuantitativos. Para encontrarlo hay que organizar los datos de menor a mayor o de menor a mayor. Se tienen dos casos:

Si el tamaño de la muestra es impar, la M_e es el dato que se encuentra en todo el centro de los datos después de ordenarlos. Por ejemplo x_1, x_2, x_3, x_4, x_5, la mediana es x_3.

Si el tamaño de la muestra es par, la M_e es el promedio de los dos tatos que están en el centro. Por ejemplo, si la muestra está dada por $x_1, x_2, x_3, x_4, x_5, x_6$ la mediana es $\frac{x_3+x_4}{2}$. Note que en este caso, la mediana no es un dato de la muestra.

La mediana indica que dato se encuentra en el centro de la muestra y por lo tanto da información sobre la cantidad de datos que están a la izquierda o a la derecha de ese valor en la muestra.

Ejercicio

En una encuesta en un barrio, se seleccionaron 7 tiendas al azar y se les preguntó por el precio del jabón en barra. Los resultados fueron los siguientes: 600, 615, 630, 620, 625, 610, 630, pesos por barra. La mediana (recuerde que hay que que ordenar los datos) de los precios de la barra de jabón es

a. $ 630

b. $ 625

c. $ 620

d. $ 610

Ejercicio

En una encuesta en un barrio, se seleccionaron 8 tiendas al azar y se les preguntó por el precio del jabón en barra. Los resultados fueron los siguientes: 600, 615, 630, 620, 625, 610, 630, 630, pesos por barra. La mediana (recuerde que hay que que ordenar los datos) de los precios de la barra de jabón es

a. $ 630.5

b. $ 625

c. $ 622

d. $ 622.5

2.3. La moda M_o de una muestra

La moda (M_o) es el dato que más se repite (que tiene mayor la frecuencia) en la muestra. Si dos datos diferentes tienen la mayor frecuencia en la muestra, se dice que la muestra es bimodal.

La moda es un estadístico que se puede encontrar para muestras cuantitativas y cualitativas.

Por ejemplo, para la muestra de las ocho tiendas que venden jabón de barra, se recolectaron los siguientes datos:

$$600,\ 615,\ 630,\ 620,\ 625,\ 610,\ 630,\ 630$$

Por lo tanto la moda es $M_o = 630$, que es el dato que más se repite en la muestra tomada.

A la entrada al estadio se observan los tipos de comidas rápidas que consumen las personas que ingresan a ver un partido de fútbol. Se tomó una muestra de tamaño 8 y se observó lo siguiente: hamburguesa, hamburguesa, perro, perro, hamburguesa, perro, hamburguesa, salchipapa. A partir de esta muestra se puede afirmar que la moda es la "hamburguesa", que es la comida rápida que más consumen los hinchas del equipo al que van a ver jugar.

Ejercicio

A la entrada de un colegio, se anotó los minutos de retardo con que llegaron los últimos 10 estudiantes en un día cualquiera de la semana. Los resultados fueron los siguientes: 15, 5, 6, 6, 7, 8, 10, 9, 2, 7 minutos. La moda de la muestra es

a. 6, 7 minutos, la muestra es bimodal.

b. 7 minutos.

c. 5 minutos.

d. 10, 15 minutos, son las mayores tardanzas.

Los valores media $(\overline{x})$, mediana, (M_e), moda (M_o), son estadísticos que se clasifican como de tendencia central. Otros estadísticos como el rango, la varianza y la desviación estándar indican medidas de dispersión de la muestra.

2.4. El rango R de una muestra

Rango R de una muestra cuantitativa es la diferencia entre el mayor y menor dato de la muestra. Es una medida de la dispersión de los datos de la muestra.

$$R = x_{max} - x_{min}$$

Para la muestra de la tardanza de los estudiantes al Colegio: 15, 5, 6, 6, 7, 8, 10, 9, 2, 7, se tiene que

$x_{max} = 15$, $x_{min} = 2$. Por lo tanto, el rango está dado por:

$$R = x_{max} - x_{min} = 15 - 2 = 13$$

minutos.

Este valor indica que entre la menor y la mayor tardanza hay 13 minutos de diferencia.

Note que para datos cualitativos no tiene sentido hablar del rango de una muestra.

Ejercicio

En una encuesta en un barrio, se seleccionaron 8 tiendas al azar y se les preguntó por el precio del jabón en barra. Los resultados fueron los siguientes: 600, 615, 630, 620, 625, 610, 630, 630, pesos por barra. El rango de la muestra es

a. \$ 10

b. \$ 15

c. \$ 20

d. \$ 30

2.5. La varianza s^2 de una muestra

La varianza de los datos de una muestra, es una medida de dispersión y se puede calcular en relación a diferentes valores: a la media, la mediana, la moda, o cualquier otro valor de interés de la muestra. Generalmente se hace con relación a la media de la muestra. Si la muestra es de tamaño n, la varianza está dada por:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}$$

o por

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$$

Se puede definir como el "casi promedio" de los cuadrados de las desviaciones $\left((x_i - \overline{x})^2\right)$ de los datos con respecto a la media $(\overline{x})$ muestral.

Si se utiliza la segunda fórmula, lo que se está indicando es que se está teniendo en cuenta los "grados de libertad" de la muestra. Esto indica que para calcular la varianza ya se ha utilizado un vez los datos (para calcular $\overline{x}$) por lo tanto se ha perdido un grado de libertad. En la calculadora o en el Excel ® para calcular este estadístico, se utiliza la segunda fórmula. En los ejemplos que siguen y en los problemas planteados, se utilizará, también, la segunda fórmula.

Si se tienen datos de toda la población, se habla de la varianza poblacional. Si la población es de tamaño N, y su media es $\overline{X}$, la varianza poblacional se calcula de la siguiente manera:

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{X})^2}{N-1}$$

Calcular estadísticos para toda la población resulta "casi" imposible en la mayoría de los casos. Es por ello que se hacen inferencias para toda la población a partir de los estadísticos muestrales.

Note que la varianza muetral se denota por s^2 y la poblacional por σ^2, la media muestral por $\overline{x}$ y la poblacional por $\overline{X}$. Siempre hay que estar atentos a lo que se quiere calcular.

Por ejemplo, para un grupo de octavo de 30 estudiantes que presentaron el examen del período de Matemáticas, se seleccionaron aleatoriamente 6 estudiantes y se les preguntó por su resultados obteniendo lo siguiente:

$$8, 6, 7, 8, 8, 5$$

El promedio de la muestra es:

$$\overline{x} = \sum_{i=1}^{6} \frac{x_i}{6} = \frac{5+6+7+8+8+8}{6} = 7$$

La varianza de la muestra esta dada por

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1} = \frac{(8-7)^2 + (6-7)^2 + (7-7)^2 + (8-7)^2 + (8-7)^2 + (5-7)^2}{6-1} = \frac{8}{5} = 1.6$$

Note que la varianza tiene unidades cuadradas, que en la mayoría de las aplicaciones no tiene sentido práctico. Por ello, se utiliza la desviación estándar como una unidad de la medida de la dispersión de los datos de una muestra.

Ejercicio

En una encuesta en un barrio, se seleccionaron 8 tiendas al azar y se les preguntó por el precio del jabón en barra. Los resultados fueron los siguientes: 600, 615, 630, 620, 625, 610, 630, 630, peso por barra. La varianza muestral es

a. \$ 121.42

b. \$ 600

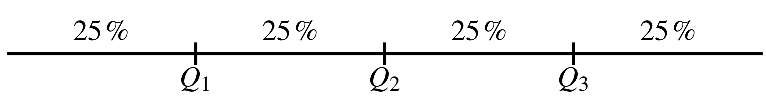
c. \$ 615

d. \$ 120.15

2.6. Cuartiles

Son medidas de posición que dividen en cuatro partes porcentuales iguales a una distribución ordenada de datos. Esto quiere decir, que si se van a encontrar los cuartiles para una muestra específica hay que ordenar los datos.

Los cuartiles se representan por Q_1, Q_2 y Q_3, como se muestra en la siguiente figura:



Como puede observarse en la gráfica, los tres cuartiles, dividen los datos en cuatro partes iguales, en las que entre cada uno de ellos está el 25 % de los datos de la muestra.

Esto quiere decir que el 25 % de los datos son menores o iguales que Q_1. El 50 % de los datos son menores o iguales que Q_2. El 75 % de los datos son menores o iguales que Q_3.

Para calcular los cuartiles se utiliza la siguiente fórmula:

$$Q_k = k\left(\frac{n}{4}\right)$$

En donde k es el número del cuartil y n el tamaño de la muestra.

Ejemplo

Se tomó una muestra aleatoria de tamaño 20 de la temperatura en grados centígrados a las dos de la tarden en el Centro de Medellín. Los datos obtenidos fueron los siguientes:

18, 25, 27, 19, 30, 22, 23, 24, 27, 29, 21, 30, 31, 17, 28, 27, 22, 23, 25, 26, 19

Encuentre los tres cuartiles de la muestra.

Como la muestra es de tamaño $n = 20$, los cuartiles se calculan de la siguiente manera:

$Q_1 = 1\left(\frac{20}{4}\right) = 5$, quiere decir que el quinto dato, en la muestra ordenada, corresponde con el primer cuartil.

$Q_2 = 2\left(\frac{20}{4}\right) = 10$, quiere decir que el décimo dato, en la muestra ordenada, corresponde con el segundo cuartil.

$Q_3 = 3\left(\frac{20}{4}\right) = 15$, quiere decir que el dato en la posición quince, en la muestra ordenada, corresponde con el tercer cuartil.

Al ordenar la muestra se tiene que

17, 18, 19, 19, **21**, 22, 22, 23, 23, **24**, 25, 25, 26, 27, **27**, 27, 28, 29, 30, 30, 31

$Q_1 = 21$, que es el dato que está en la posición 5 en la muestra ordenada. Se puede interpretar que el 25 % de los días observados, la temperatura en Medellín, a las 2 de la tarde era menor a 21 grados centígrados.

$Q_2 = 24$, que es el dato que está en la posición 10 en la muestra ordenada. Se puede interpretar que el 50 % de los días observados, la temperatura en Medellín, a las 2 de la tarde era menor a 24 grados centígrados.

$Q_3 = 27$, que es el dato que está en la posición quince en la muestra ordenada. Se puede interpretar que el 75 % de los días observados, la temperatura en Medellín a las 2 de la tarde era menor a 27 grados centígrados.

En la muestra ordenada, los cuartiles se muestran resaltados.

Ejercicio

Se toma una muestra aleatoria de las notas de matemáticas de 16 estudiantes en un Colegio. Los datos obtenidos fueron los siguientes: 3, 8, 9, 10, 4, 5, 9, 3, 4, 7, 2, 8, 3, 7, 5, 7
Los cuartiles Q_1, Q_2 y Q_3, respectivamente, son

a. 3, 5, 9

b. 2, 4, 8

c. 4, 5, 9

d. 3, 5, 8

2.7. La desviación estándar s de una muestra

La desviación estándar de una muestra es la raíz cuadrada de la varianza muestral, es decir:

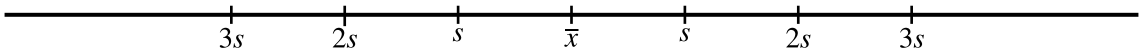
$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}}$$

o por

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}}$$

La desviación estándar tiene las mismas unidades de la variable observada. Se utiliza para encontrar los datos que se encuentra a una, dos, tres o más desviaciones estándar de la media. En lo que sigue, se utilizará la segunda fórmula para encontrar la desviación estándar de una muestra.

Observe el siguiente dibujo en el que se ilustra esta situación:



Por ejemplo, para un grupo de octavo de 30 estudiantes, que presentaron el examen del período de matemáticas, se seleccionaron aleatoriamente 6 estudiantes y se les preguntó por su resultados obteniendo lo siguiente:

$$8, 6, 7, 8, 8, 5$$

El promedio de la muestra es:

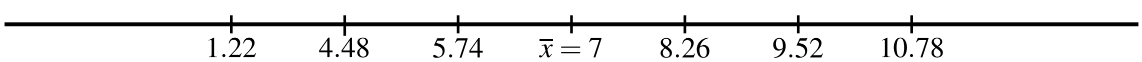
$$\overline{x} = \sum_{i=1}^{6} \frac{x_i}{6} = \frac{5+6+7+8+8+8}{6} = 7$$

La varianza de la muestra está dada por

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1} = \frac{(8-7)^2 + (6-7)^2 + (7-7)^2 + (8-7)^2 + (8-7)^2 + (5-7)^2}{6-1} = \frac{8}{5} = 1.6$$

La desviación estándar es $s = \sqrt{s^2} = \sqrt{1.6} = 1.26$

Al graficar los datos obtenidos sobre la recta, se tiene:



Los datos que están a una desviación estándar de la media son:

La media de la muestra es $\overline{x} = 7$. Como la varianza se calculó con respecto a la media muestral, este valor se coloca en el centro de la línea.

Por lo tanto, los valores de la muestra que están a una distancia de una desviación estándar de la media muestral son lo que están en el intervalo:

$$\overline{x}-s<x_i<\overline{x}+s \Rightarrow 7-1.26<x_i<7+1.26 \Rightarrow 5.74<x_i<8.26$$

De acuerdo con lo observado en la muestra dada, los valores encontrados son: 6, 7, 8, 8 y 8. Note que la mayoría de los valores están a una desviación estándar de la media. Esto indica que la mayoría de los datos están agrupados alrededor de la media, a una desviación estándar.

Los que están a dos desviaciones estándar de la media, están en los intervalos, dados por $4.48<x_i<5.74$ o $8.26<x_i<9.52$. En este ejemplo, el único dato que cumple estas condiciones es $x=5$.

Los intervalos que están a tres desviaciones estándar de la media son $1.22<x_i<4.48$ o $9.52<x_i<10.78$. Para los datos analizados, ninguno se encuentra a tres desviaciones de la media.

Recordar que el rango, la varianza muestral (s^2) y la desviación estándar (s) son medidas de la dispersión de los datos.

Por ejemplo, si para el mismo grupo de estudiantes, los resultados de la encuesta hubieran sido los siguientes: 5, 10, 8, 6, 9, 4. Al encontrar la media, el rango, la varianza y la desviación estándar se obtienen los siguientes resultados:

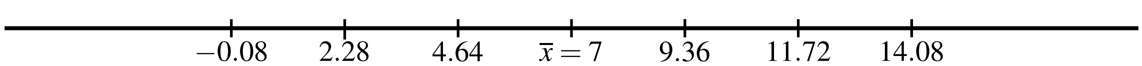
La media, $\overline{x}=7$, que es igual a la de la muestra anterior.

El rango es $R=10-4=6$ que es mayor que el anterior que fue de 3. Indicando una mayor dispersión de los datos.

La varianza muestral, $s^2=5.6$, que comparada con la anterior $(s^2=1.6)$, lo que confirma una mayor dispersión de los datos.

La desviación estándar muestral, $s=2.36$ es mayor que la anterior que era $s=1.26$, indicando una mayor dispersión de los datos de la muestra.

Al colocar la media, la primera, segunda y tercera desviaciones estándar muestrales sobre una recta se obtiene:



Los datos de la muestra, que están a una desviación estándar de la media son: 5, 6, 8 y 9.

Los datos de la muestra, que están a dos desviación estándar de la media son: 4 y 10

No hay datos a tres desviaciones estándar de la media.

Ejemplo

En una institución educativa se tomó una muestra al azar, de tamaño $n = 30$, y se les preguntó a los alumnos sobre las notas obtenidas en el área del lenguaje, para el segundo período académico. Los datos recolectados fueron los siguientes:

2, 10, 4, 5, 6, 7, 9, 3, 7, 4, 7, 9, 10, 2, 1, 5, 6, 7, 8, 8, 8, 4, 9, 2, 4, 7, 6, 5, 7, 9

Resolver cada uno de los siguientes aspectos relacionados con la muestra dada:

a. Tipo de variable a analizar.

b. La media

c. La moda

d. La mediana

e. El rango

f. La varianza

g. La desviación estándar

h. Los datos de la muestra que están a una, dos y tres desviaciones estándar de la media.

i. De una interpretación de los resultados obtenidos.

Solución

a. La variable dada es *cuantitativa*, pues tiene sentido realizar operaciones aritméticas con los datos suministrados. La variable observada es: la calificación obtenida en el área de lenguaje de una muestra de 30 estudiantes en una institución educativa.

b. La media muestral está dada por:

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 6.03$$

lo cual indica que en promedio los estudiantes, en el segundo período académico, ganaron el área de lenguaje, pero con una nota baja. Esto debería alertar a los profesores del área para tomar medidas correctivas y mejorar los procesos evaluativos, de enseñanza, entre otros.

c. La moda es el dato que más se repite en la muestra. Este es $M_o = 7$, indica que la mayoría de los estudiantes en la muestra, tienen esta nota como calificación.

d. La mediana de la muestra, es el dato que se encuentra en el centro. Como la muestra es de tamaño $n = 30$, que es par, se deben de tomar los datos que están en el centro y encontrar su promedio. Al ordenar los datos se obtiene:

   1, 2, 2, 2, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6, **6**, **7**, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10

   Los dos datos que están resaltados en negro son los que se encuentran en el centro de la muestra después de ordenarla. Por lo tanto $M_e = \frac{6+7}{2} = 6.5$. Este valor no se encuentra en los datos analizados. Indica que a la derecha de 6.5 hay la misma cantidad de datos que a la izquierda del mismo valor.

e. El rango es la diferencia entre el dato mayor y el dato menor. En este caso, entre 10 y 1. Por lo tanto $R = 10 - 1 = 9$, que indica que la muestra es dispersa.

f. La varianza esta dada por:

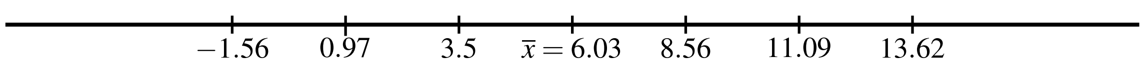
$$s^2 = \sum_{i=1}^{30} \frac{(x_i - \overline{x})^2}{30 - 1} = 6.44$$

   que en el contexto de las notas de una muestra de estudiantes no tiene sentido.

g. La desviación estándar de la muestra está dada por:

$$s = \sqrt{\sum_{i=1}^{30} \frac{(x_i - \overline{x})^2}{30 - 1}} = 2.53$$

   esto indica que los datos recolectados son dispersos. Al encontrar los intervalos para la primera, segunda y tercera desviación estándar con respecto a la media y graficarlos se obtiene:



h. Los datos de la muestra que están a una, dos y tres desviaciones estándar de la media.

   A una desviación estándar de la media muestral, están los datos que se encuentran en el intervalo: $3.5 < x_i < 8.56$, estos son: 4, 4, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8

   Los que están a dos desviaciones estándar de la media muestral son los que están en el intervalo: $0.97 < x_i < 3.5$ unido $8.56 < x_i < 11.09$, estos son: 1, 2, 2, 2, 3, 9, 9, 9, 9, 10, 10

   No hay datos que se encuentren a tres desviaciones estándar de la media muestral.

i. Los estadísticos asociados con una muestra aleatoria son herramientas valiosas en la toma de decisiones, para las personas del área de la que proviene la muestra, en este caso, del área de lenguaje de una institución educativa. Los profesores pueden tomar acciones para que las notas no estén tan dispersas, subir más el promedio, entre otras.

Ejemplo

En una encuesta realizada a 30 estudiantes de una escuela femenina, acerca del color de la blusa que prefieren usar el fin de semana. Se obtuvieron los siguientes resultados:

Azul, Verde, Amarillo, Azul, Azul, Verde, Verde, Amarillo, Rojo, Rojo, Rojo, Verde, Verde, Azul, Azul, Amarillo, Negro, Negro, Verde, Verde, Rojo, Azul, Azul, Verde, Azul, Verde, Negro, Anaranjado, Negro, Verde

Resuelva las siguientes aspectos relacionados con la muestra dada:

a. Describa la variable a observar y su tipo.

b. De acuerdo con el tipo de variable, encuentre los estadísticos que son pertinentes.

Solución

a. La variable observada es: El color de la blusa que prefieren usar las estudiantes de una escuela los fines de semana. La variable es *cualitativa*, pues no es posible realizar, con los datos y operaciones aritméticas que tengan sentido matemático.

b. De acuerdo con los datos suministrados, el estadístico pertinente es la moda. Para poder mirar, con facilidad los datos que se repiten, los organizamos de acuerdo con el orden alfabético, de donde se obtiene:

   Amarillo, Amarillo, Amarillo, Anaranjado, Azul, Azul, Azul, Azul, Azul, Azul, Azul, Azul, Negro, Negro, Negro, Negro, Rojo, Rojo, Rojo, Rojo, Verde, Verde, Verde, Verde, Verde, Verde, Verde, Verde, Verde,Verde

   Al realizar el recuento se observa que el color que más se repite es el “Verde”, por lo tanto, ese color es la moda.

Cuando se tienen muestras grandes, es preferible utilizar técnicas de agrupamiento de los datos. Estas técnicas permiten realizar distintos gráficos que pueden ser útiles para presentar e interpretar la información, bien sea de tipo cualitativo o cuantitativo.

Observación

Hay datos que pueden presentarse como números, pero en realidad representan datos cualitativos.

Por ejemplo: 1^0, 2^0, 3^0 4^0, ..., que pueden ser los lugares en los que llegaron los corredores en una competencia atlética. Hay que tener cuidado con el tipo de dato que se presenta para saber los estadísticos que se pueden calcular a partir de ellos.

3. Agrupamiento de datos de una muestra aleatoria

Para agrupar los datos de una muestra aleatoria, hay que tener en cuenta si estos son *cualitativos* o *cuantitativos*.

Si los datos son *cualitativos*, se puede formar una categoría o clase por cada una de las cualidades de los datos de la muestra, por ejemplo, si corresponden a:

a. Colores: cada uno de los colores de la muestra puede ser una categoría o clase.

b. Si corresponden a "bueno", "regular" o "malo", los datos que correspondan a una de estas características puede formar una categoría o clase.

c. Entre otros, ...

Si los datos son cuantitativos, se pueden formar intervalos para agruparlos. Cada uno de los intervalos definidos forman una categoría o clase.

Observación

Hay que tener en cuenta que tanto para variables cualitativas o cuantitativas, un mismo dato o valor no puede pertenecer a dos categorías o clases distintas. Es decir, al definir las clases, estas deben ser disyuntas.

3.1. Número de clases N_c para datos cuantitativos

Existen varias fórmulas para calcular el número de clases (N_c) para datos cuantitativos. Si la muestra es de tamaño n, se tiene las siguientes formas:

a. A partir de la regla de Sturges, que propone que el número de clases para agrupar los datos de una muestra de tamaño n está dado por $N_c = 1 + 3.332\log_{10}(n)$.

   Así, por ejemplo, para una muestra cuantitativa de tamaño $n = 30$, el número de clases a utilizar es: $N_c = 1 + 3.332\log_{10}(30) = 1 + (3.332)(1.477) = 5.9$ clases. La persona encargada de procesar la información debe tomar la decisión de considerar 5 ó 6 clases.

b. Otro criterio muy utilizado es que el número de clase a utilizar es la raíz cuadrada del número de datos, es decir $N_c = \sqrt{n}$, en dónde n es el tamaño de la muestra.

   Por ejemplo, si la muestra es de tamaño 30, $N_c = \sqrt{30} = 5.47$, que como en el caso anterior, el investigador debe tomar la decisión de tomar 5 ó 6 clases.

c. Otro criterio para calcular el número de clases, para agrupar los datos de una muestra de tamaño n, es que si $20 \leq n \leq 100$, el número máximo de clases es 5.

d. Al encontrar el número de clases y comenzar a ubicar los datos en cada una de las clases, se sugiere que no quede ninguna clase sin datos.

Observación

Como se puede apreciar, el cálculo de clases es un método empírico y una buena distribución de los datos está asociada con la experiencia del investigador, del tipo de datos y de los fines para los que se va a utilizar la información.

3.2. Ancho de las clases A_c para datos cuantitativos

A partir de la determinación del número de clases hay que calcular el ancho (A_c) de cada una de ellas. Para esto se divide el rango (R) de la muestra, entre el número de clases definidas, esto es:

$$A_c = \frac{\text{Rango}}{\text{Número de clases}} = \frac{R}{N_c}$$

Luego de determinar el número de clases y del ancho de las clases se procede a calcular cada una de las clases. Para ello se parte del dato menor.

3.3. Cálculo de las clases para datos cuantitativos

Para calcular las clases se construyen intervalos que tengan el ancho (A_c) de la clase determinado, partiendo del dato menor encontrado en la muestra.

En el primer intervalo a construir se tiene como extremo inferior el dato menor de la muestra y como extremo superior, el dato menor más el ancho de la clase. Hay que determinar cuál de los extremos es abierto y cuál cerrado. Cabe recordar que los intervalos no deben de tener valores en común, puesto que un dato no puede ser ubicado en dos clases distintas. Para calcular las otras clases se procede de igual forma.

Ejemplo

Retomemos los 30 datos analizados anteriormente.

En una institución educativa se tomo una muestra al azar, de tamaño $n = 30$, y se les pregunto a los alumnos sobre las notas obtenidas en el área del lenguaje para el segundo período académico. Los datos recolectados fueron los siguientes:

2, 10, 4, 5, 6, 7, 9, 3, 7, 4, 7, 9, 10, 2, 1, 5, 6, 7, 8, 8, 8, 4, 9, 2, 4, 7, 6, 5, 7, 9

Determine:

a. El rango R de la muestra.

b. El número de clases N_c.

c. El ancho de las clases A_c.

d. Cada una de las clases.

Solución

a. El rango R de la muestra está dado por la diferencia entre el dato mayor, menos el dato menor. El dato mayor es 10 y el menor 1.

$$R = 10 - 1$$

b. Para calcular el número de clases N_c, se parte del tamaño de la muestra, en este caso $n = 30$. Si se utiliza la fórmula de Sturges, se tiene que $N_c = 1 + 3.332\log_{10}(30) = 1 + (3.332)(1.477) = 5.9$

Ahora, si se determina a partir de la raíz cuadrada del tamaño de la muestra se tiene que:

$$N_c = \sqrt{30} = 5.47$$

Si se tiene en cuenta que el tamaño de la muestra es $n = 30$, que es un número entre 20 y 100, se sugiere no tener más de 5 clases.

Por lo tanto, se toma $N_c = 6$

c. El ancho de las clases está dado por el cociente entre el rango y el número de clases.

$$A_C = \frac{R}{N_c} = \frac{9}{5} = 1.8$$

d. Determine las clases.

Para la construcción, teniendo en cuenta que las clases deben ser disyuntas, se tomará la convención de que el extremo izquierdo es cerrado y el derecho abierto.

Cálculo de la primera clase

Se parte del dato menor, en este caso es 1. Este valor se toma como el extremo inferior de la primera clase. El extremo superior es el dato menor más el ancho de clase, esto es $1 + A_c = 1 + 1.8 = 2.8$

Por lo tanto, la primera clase está determinada por el intervalo $[1, 2.8)$

Cálculo de la segunda clase

Para la segunda clase, se toma como extremo inferior, el extremo superior de la primera clase, esto es 2.8. Para el extremo superior se le suma el ancho de la clase, esto es $1.8 + A_c = 1.8 + 1.8 = 3.6$.

Para determinar las otras clases se procede de igual forma, de dónde se tiene que:

Primera clase: $[1, 1.8)$

Segunda clase: $[1.8, 3.6)$

Tercera clase: $[3.6, 5.4)$

Cuarta clase: $[5.4, 7.2)$

Quinta clase: $[7.2, 9)$

Como se puede observar en la muestra, los datos con valores 9 o 10 no están dentro de ninguna de las clases construidas. En un estudio estadístico, hay que tener en cuenta todos los datos suministrados. Para incluirlos, se tiene dos posibilidades: (a) ampliar la última clase, que en este caso sería $[7.2, 10]$, pero quedaría con muchos datos, (b) con los datos excluidos crear una nueva clase, en este caso sería $[9, 10]$. En este ejemplo se decidió utilizar la segunda opción expuesta. Note que en todas las clases hay datos.

Para continuar estudiando la muestra, se colocan las clases en una tabla como la siguiente:

Clase
$[1, 1.8)$
$[1.8, 3.6)$
$[3.6, 5.4)$
$[5.4, 7.2)$
$[7.2, 9)$
$[9, 10]$
Totales

Note que al tabular las clases, estas se ubican de arriba hacia abajo, en el orden en que se fueron construyendo. Esto es muy importante, puesto el buen análisis de los resultados obtenidos se basa en hacer los proceso ordenadamente.

3.3.1. Recuento de los datos

Después de determinar las clases hay que contar los datos que hay en cada una de ellas. Para ello se hace un recuento, que se puede hacer más fácilmente a partir del ordenamiento de los datos suministrados. Esto es:

1, 2, 2, 2, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10

De esta forma, para la primera clase se cuentan los datos que están en el intervalo $[1, 1.8)$, que en este caso es únicamente el 1.

Para la segunda clase, se cuentan todos los datos que están en el intervalo $[1.8, 3.6)$, que corresponden a los datos iguales a 2 y a 3.

Este procedimiento se realiza con todas las clases obtenidas.

El conteo se puede tabular en una tabla como la siguiente.

Clases	Recuento
$[1, 1.8)$	I
$[1.8, 3.6)$	IIII
$[3.6, 5.4)$	IIIIIII
$[5.4, 7.2)$	IIIIIIIII
$[7.2, 9)$	III
$[9, 10]$	IIIIII
Totales	30

3.3.2. Frecuencia simple f_i

La frecuencia simple f_i se encuentra para cada una de las clases, que se han determinado para la muestra dada, es el número de datos que hay en cada una de las clases. En el ejemplo, anterior para la clase $[5.4, 7.2)$ la frecuencia relativa es $f = 9$, que es el resultado del recuento de los datos que están en ese intervalo. La tabla se puede seguir ampliando de la siguiente manera:

Clases	Recuento	f_i
$[1, 1.8)$	I	1
$[1.8, 3.6)$	IIII	4
$[3.6, 5.4)$	IIIIIII	7
$[5.4, 7.2)$	IIIIIIIII	9
$[7.2, 9)$	III	3
$[9, 10]$	IIIIII	6
Totales	30	30

Para la frecuencia simple se pueden realizar gráficas de barras horizontales o verticales, denominadas histogramas de frecuencias. Hacer uno de los dos tipos de estos dos gráficos está en función de la información que se tenga y de la forma como el investigador prefiera presentarla. Si se realiza un histograma vertical, en el eje horizontal se escriben las clases y en el vertical las frecuencias de cada una de ellas, como se muestra en la Figura 2. Note la facilidad con la que se puede leer la información al graficarla.

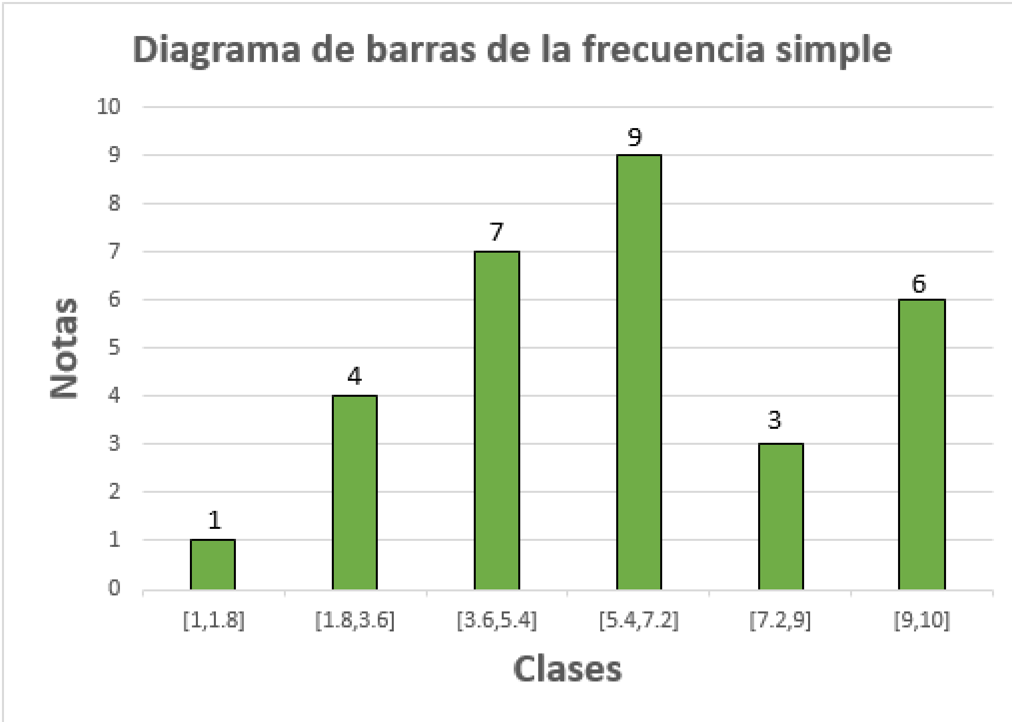


Figura 1: Histograma de frecuencias simples.

3.3.3. Frecuencia simple acumulada F_i

La frecuencia simple acumulada F_i, para una clase cualquiera, es la suma de todas las frecuencias simples de las clases anteriores más la frecuencia simple de la clase que se está observando.

Por ejemplo, F_i par la clase $[3.6, 5.4)$ es las suma de $1+4+7=12$ que es la suma de las frecuencias simples de las clases que son anteriores a ella, más la frecuencia simple de la misma clase. Los resultados obtenidos se muestran en la siguiente tabla:

Clases	Recuento	f_i	F_i
$[1, 1.8)$	I	1	1
$[1.8, 3.6)$	IIII	4	5
$[3.6, 5.4)$	IIIIIII	7	12
$[5.4, 7.2)$	IIIIIIIII	9	21
$[7.2, 9)$	III	3	24
$[9, 10]$	IIIIII	6	30
Totales	30	30	

3.3.4. Frecuencia relativa (f_r) y frecuencia relativa acumulada (F_r)

La frecuencia relativa (f_r), se obtiene al dividir cada una de las frecuencias simples (f_i) por el número de datos de la muestra.

La frecuencia relativa acumulada (F_r) se obtiene dividiendo cada una de las frecuencias acumuladas por el número de datos de la muestra. Note que la mayor frecuencia relativa acumulada es 1, que corresponde a la última clase.

Al agregarle estas dos columnas a la tabla se obtiene lo siguiente:

Clases	Recuento	f_i	F_i	f_r	F_r
$[1,1.8)$	I	1	1	1/30	1/30
$[1.8,3.6)$	IIII	4	5	4/30	5/30
$[3.6,5.4)$	IIIIIII	7	12	7/30	12/30
$[5.4,7.2)$	IIIIIIIII	9	21	9/30	21/30
$[7.2,9)$	III	3	24	3/30	24/30
$[9,10]$	IIIIII	6	30	6/30	30/30
Totales	30	30			

3.3.5. Frecuencia relativa porcentual $(f_r\%)$ y frecuencia relativa acumulada porcentual $(F_r\%)$

Tanto la frecuencia relativa porcentual $(f_r\%)$, como la frecuencia relativa acumulada porcentual $(F_r\%)$, se obtienen multiplicando las respectivas frecuencias por 100.

Por ejemplo para la clase $[3.6,5.4)$, la frecuencia relativa $f_r = \frac{7}{30}$, por lo tanto la frecuencia relativa porcentual es $f_r\% = f_r(100) = 23.333$, lo que significa que el porcentaje de los datos de la muestra que están en esta clase es el 23.333 por ciento.

Para la misma clase, por ejemplo para la clase $[3.6,5.4)$, la frecuencia relativa acumulada porcentual es $F_r\% = F_r(100)$ esto es $F_r\% = F_r(100) = \frac{12}{30}(100) = 0.4(100) = 40$, lo que significa que el 40 por ciento de los datos de la muestra están en esta clase unida a los datos de las anteriores.

Al completar la tabla con estos nuevos valores se obtiene (ver Tabla 1):

Clases	Recuento	f_i	F_i	f_r	F_r	$f_r\%$	$F_r\%$
$[1,1.8)$	I	1	1	1/30	1/30	3.333	3.333
$[1.8,3.6)$	IIII	4	5	4/30	5/30	13.333	16.666
$[3.6,5.4)$	IIIIIII	7	12	7/30	12/30	23.333	40.000
$[5.4,7.2)$	IIIIIIIII	9	21	9/30	21/30	30.000	70.000
$[7.2,9)$	III	3	24	3/30	24/30	10.000	80.000
$[9,10]$	IIIIII	6	30	6/30	30/30	20.000	100.000
Totales	30	30					

Tabla 1: Frecuencias para datos cuantitativos.

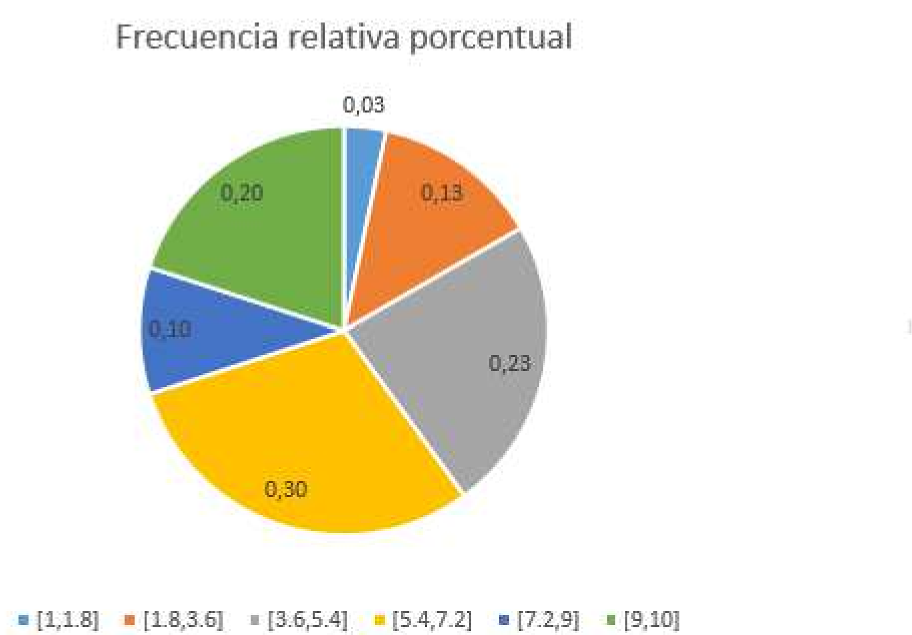


Figura 2: Diagrama de torta para la $f_r\%$.

Para la frecuencia relativa porcentual, un tipo de gráfica que se utilizan son las llamadas "tortas" en las que es más fácil observar el porcentaje de tiene cada una de las clases.

A partir de completar una tabla como está se pueden hacer muchas preguntas relativas a la muestra y sus respuestas se pueden extrapolar a toda la población.

Así, por ejemplo, ¿cuál es el porcentaje de alumnos que obtuvieron notas de 6 ó 7? Para encontrar la respuesta a esta pregunta, se observa que estas notas pertenecen a la clase $[5.4, 7.2)$ que su frecuencia relativa (f_r) es $f_r = \frac{9}{30}$. Es preferible entregarla en términos de porcentajes, por lo que se tiene $f_r\% = \frac{9}{30}(100) = 30$. Lo que significa que el 30 % de los alumnos de la muestra tienen notas de 6 ó 7. Si la muestra está bien tomada y es representativa de la población, este resultado se puede generalizar a toda la población diciendo: *que el 30 % de los alumnos de la institución educativa, en el área de español, aproximadamente, obtuvieron una nota de 6 ó 7.*

Otra pregunta que se podría formular es: ¿qué porcentaje de estudiantes perdió la prueba de español en el segundo período? La respuesta se puede obtener a partir de la frecuencia relativa acumulada porcentual. Para ello se observa que que las notas $x_i \leq 5$ están en la unión de las clases $[1, 1.8)$, $[1.8, 3.6)$ y $[3.6, 5.4)$. Toda la frecuencia relativa acumulada porcentual $(F_r\%)$, se encuentra en la clase $[3.6, 5.4)$. Es decir, $(F_r\%) = 40$, lo que indica que el 40 por ciento de los estudiantes en la muestra, perdieron la prueba de español en el segundo período. Como en le caso anterior, si la muestra esta bien tomada y es representativa del Área de Lenguaje, se puede afirmar que aproximadamente el 40 % de los estudiantes de esa institución, en el segundo período perdieron el Área de Lenguaje.

Ejercicio

Con base en la Tabla 1 encuentre la frecuencia acumulada porcentual $(F_r\%)$ de los estudiantes que obtuvieron una nota menor o igual que 7 en la prueba de lenguaje.

a. 3.333 por ciento.

b. 30 por ciento.

c. 40 por ciento.

d. 70 por ciento.

Ejercicio

Con base en la Tabla 1 encuentre la frecuencia (f_r) para los estudiantes que obtuvieron una nota 4 ó 5, en la prueba de lenguaje.

a. $\frac{7}{30}$

b. $\frac{9}{30}$

c. $\frac{5}{30}$

d. $\frac{1}{3}$

3.4. Agrupamiento de datos de una muestra aleatoria cualitativa

Para agrupar datos de una muestra aleatoria cualitativa, fundamentalmente, se siguen los mismos pasos que para una muestra cuantitativa. Las clases se definen a partir de las cualidades de la muestra. Si son colores, cada uno de los colores puede definirse como una clase, si son preferencias por equipos de fútbol, cada uno de los equipos puede definirse como una clase, el investigador debe de tener claro las cualidades que quiere observar en un conjunto de datos y de esta forma tener criterios para agrupar los datos.

Ejemplo

En una encuesta realizada a 30 estudiantes de una escuela femenina, acerca del color de la blusa que prefieren usar el fin de semana. Se obtuvieron los siguientes resultados:

Azul, Verde, Amarillo, Azul, Azul, Verde, Verde, Amarillo, Rojo, Rojo, Rojo, Verde, Verde, Azul, Azul, Amarillo, Negro, Negro, Verde, Verde, Rojo, Azul, Azul, Verde, Azul, Verde, Negro, Anaranjado, Negro, Verde

Solucione cada uno de los siguientes aspectos:

a. Defina las clases y utilice el orden alfabético para organizarlas.

b. Llene la tabla de frecuencias.

c. Encuentre la frecuencia de que las niñas usen blusa azul el fin de semana.

d. Encuentre la frecuencia relativa de las niñas que usan blusa verde o roja el fin de semana.

Solución

a. Al ordenar los datos alfabéticamente se tiene:

   Amarillo, Amarillo, Amarillo, Anaranjado, Azul, Azul, Azul, Azul, Azul, Azul, Azul, Azul, Negro, Negro, Negro, Negro, Rojo, Rojo, Rojo, Rojo, Verde, Verde, Verde, Verde, Verde, Verde, Verde, Verde, Verde, Verde

   Por lo tanto se pueden definir las clases como cada uno de los diferentes colores que se encuentran en la muestra de tamaño 30. Es decir: Amarillo, Anaranjado, Azul, Negro, Rojo, Verde.

   De esta forma, las clases ya están organizadas alfabéticamente.

b. Al llenar la tabla de frecuencias, se obtiene la Tabla 2, en la página 29.

c. La frecuencia relativa (f_r) de que las niñas usan blusa azul el fin de semana es $f_r = \frac{12}{30}$ y la frecuencia relativa porcentual es $(f_r\%)$ es $f_r\% = \frac{12}{30}(100) = 26.666$ por ciento. Se puede interpretar diciendo que aproximadamente el 26.666 por ciento de las niñas del Colegio utilizan blusa azul los fines de semana.

Clases	Recuento	f_i	F_i	f_r	F_r	$f_r\,\%$	$F_r\,\%$
Amarillo	III	3	3	3/30	3/30	10.000	10.000
Anaranjado	I	1	4	1/30	4/30	3.333	13.333
Azul	IIII IIII	8	12	8/30	12/30	26.666	40.000
Negro	IIII	4	16	4/30	16/30	13.333	53.333
Rojo	IIII	4	20	4/30	20/30	13.333	66.666
Verde	IIII IIII II	10	30	10/30	30/30	33.333	100.000
Totales	30	30					

Tabla 2: Tabla de frecuencias para datos *cualitativos*.

d. La frecuencia relativa de las niñas que usan blusa verde o roja el fin de semana, es la frecuencia relativa de las niñas que usan blusa verde más las que usan blusa roja. Esto es:

$$f_{r_v} + f_{r_r} = \frac{10}{30}(100) + \frac{4}{30}(100) = 33.333 + 13.333 = 46.666$$

Una interpretación es que los fines de semana, las niñas del colegio, utilizan blusa azul o roja, aproximadamente, en un 46.666 por ciento.

Ejercicio

Con base en la Tabla 2 encuentre la frecuencia acumulada porcentual $(F_r\,\%)$para las niñas que usan blusa amarillas o anaranjada los fines de semana.

a. 66.666 por ciento.

b. 10 por ciento.

c. 13.333 por ciento.

d. 40 por ciento.

Ejercicio

Con base en la Tabla 2 encuentre la frecuencia (f_r) para las niñas que usan blusa verde los fines de semana.

a. $\frac{2}{15}$

b. $\frac{8}{30}$

c. $\frac{5}{30}$

d. $\frac{1}{3}$

4. Actividades de ejercitación

En los siguientes ejercicios desarrollar cada uno de los aspectos pedidos.

4.1. Apreciación de la calidad de un colegio

En un colegio se tomó una muestra aleatoria de tamaño 50. Se le preguntó a los alumnos por su apreciación acerca de la calidad del mismo. Los resultados obtenidos fueron los siguientes:

Bueno, Regular, Bueno, Excelente, Excelente, Regular, Regular, Bueno, Bueno, Bueno, Bueno, Malo, Bueno, Bueno, Bueno, Excelente, Excelente, Malo, Bueno, Bueno, Excelente, Excelente, Excelente, Bueno, Bueno, Bueno, Regular, Excelente, Malo, Regular, Regular, Bueno, Malo, Regular, Excelente, Bueno, Bueno, Malo, Bueno, Regular, Excelente, Excelente, Malo, Bueno, Regular, Excelente, Excelente, Malo, Malo, Bueno

Resuelva cada uno de los siguientes ejercicios.

a. Defina la variable observada y clasifiquela como *cualitativa* o *cuantitativa*.

b. Encuentre la moda de la muestra y de una interpretación de la misma.

c. Complete la Tabla 3.

d. La frecuencia relativa de los estudiantes que tienen una apreciación “Excelente” del colegio es

e. La frecuencia relativa de los estudiantes que tienen una apreciación “Regular” del colegio es

f. La frecuencia acumulada de los estudiantes que tienen una apreciación “Regular” del colegio es

Clases	Recuento	f_i	F_i	f_r	F_r	$f_r\%$	$F_r\%$
Excelente							
Bueno							
Regular							
Malo							
Totales							

Tabla 3: Frecuencias de la apreciación de los estudiantes del colegio

g. La frecuencia relativa porcentual de los estudiantes que tienen una apreciación “Mala” del colegio es

h. La frecuencia acumulada porcentual de los estudiantes que tienen una apreciación “Buena” del colegio es

4.2. Apreciación de los jugadores de la selección Colombia

En un grado noveno de una institución educativa, hay diversas opiniones sobre cuál es el mejor jugador de la Selección Colombia. A Jaime se le ocurrió hacer una encuesta sobre las preferencias de sus compañeros, obteniendo los resultados que se muestran en la Tabla 4.

Responda cada una de los siguientes aspectos:

a. Defina la variable y el tipo de variable.

b. Determine el tamaño de la muestra.

c. Determine la moda de la muestra.

d. Complete la Tabla 4.

e. Encuentre la frecuencia relativa para “James” y “Cuadrado”. De una intepretación de cada uno de los resultados.

f. Realice un histograma de frecuencias.

g. Realice un diagrama de “Torta” para las frecuencias relativas porcentuales.

Clases	Recuento	f_i	F_i	f_r	F_r	$f_r\%$	$F_r\%$
Falcao	IIII IIII						
James	IIII IIII III						
Ospina	IIII III						
Cuadrado	IIII II						
Bonilla	IIII						
Totales							

Tabla 4: Recuento de apreciación de jugadores de la Selección Colombia.

4.2.1. Histograma

En el siguiente histograma se presenta el el control de peso de 50 recién nacidos.

A partir de del histograma, resuelva los siguientes aspectos.

a. Defina la variable de interés.

b. Determine si la variable es cualitativa o cuantitativa.

c. Completa la Tabla de frecuencias 5.

4.2.2. Diagrama de torta

Ejercicio Se le preguntó a 200 estudiantes sobre su materia preferida. El gráfico representa los resultados obtenidos:

a. Defina el tipo de variable observada.

b. Determine si es cualitativa o cuantitativa.

c. Según el diagrama circular indicar la cantidad de estudiantes que prefieren cada materia.

d. Realiza un diagrama de barras horizontales sobre la preferencia de materias utilizando los datos del ejercicio anterior.

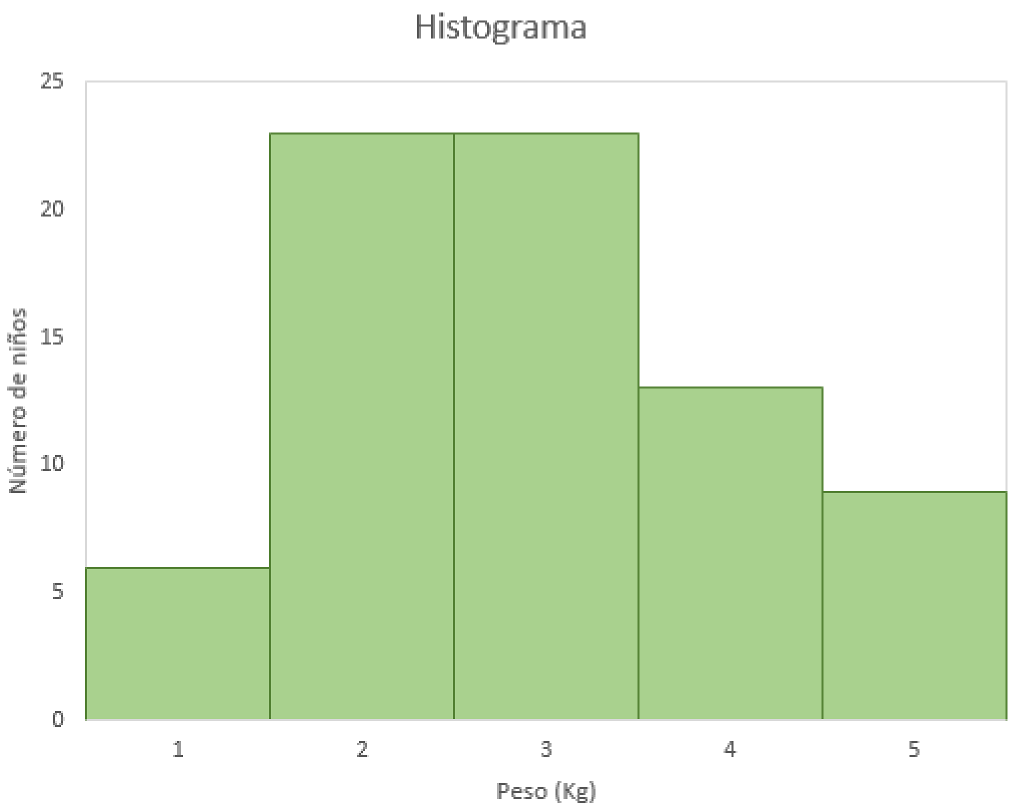


Figura 3: Histograma de frecuencias simples.

Clases	Recuento	f_i	F_i	f_r	F_r	$f_r\%$	$F_r\%$
$[2.5,3)$							
$[3,3.5)$							
Ospina							
$[3.5,4)$							
$[4,4.])$							
Totales							

Tabla 5: Peso de recién nacidos.

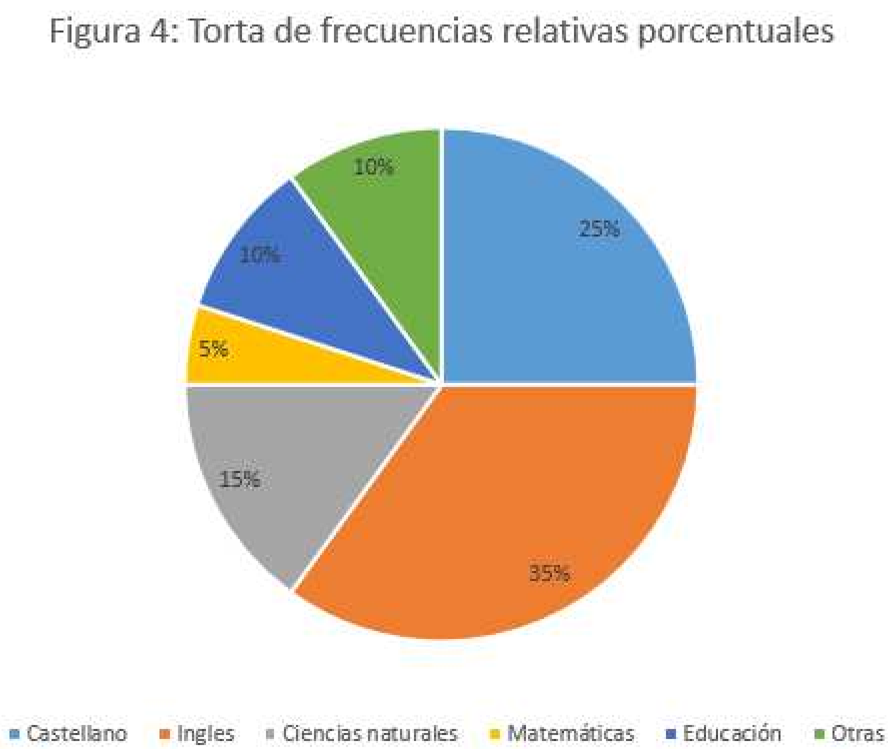


Figura 4: Torta de frecuencias relativas porcentuales.

e. ¿Cuál es la materia preferida y por qué?

f. Organiza las materias en orden de preferencia, iniciando por la materia que menos gusta, hasta la que más agrado representa para los estudiantes. ¿Cuál materia queda en la mitad?

Ejercicio

La siguiente gráfica muestra los resultados obtenidos en un estudio sobre lo que ahorra semanalmente cada uno de los jóvenes de un curso.

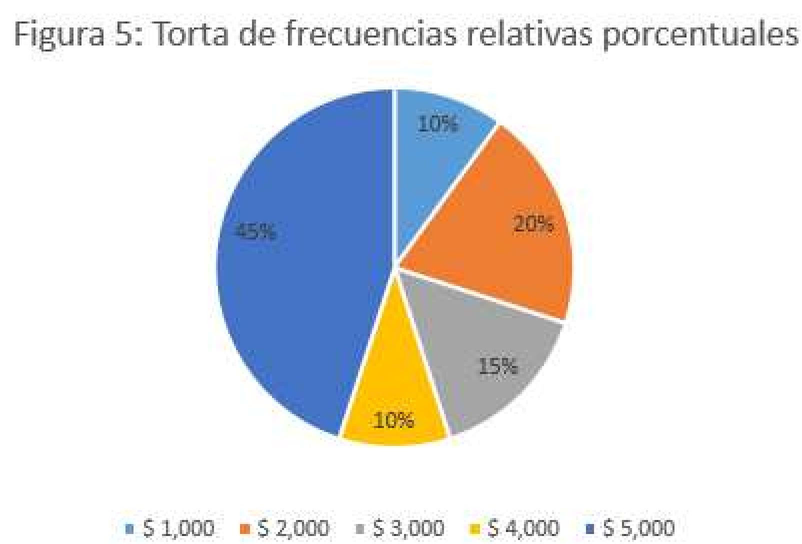


Figura 5: Torta de frecuencias relativas porcentuales.

a. Defina el tipo de variable observada.

b. Determine si es *cualitativa* o *cuantitativa*.

c. ¿Cuál es la variable que se considera en este estudio?

d. ¿El dato más frecuente en este estudio es?

e. A partir de la información dada en la gráfica, ¿Cuál es el número total de datos?

f. ¿Cuáles son las medidas de tendencia central, de localización y de dispersión?

5. Reglas de conteo

En diferentes actividades de la vida diaria es necesario saber de cuántas maneras o formas se puede realizar una actividad o proceso. Algunas preguntas que normalmente se hacen las personas son: (a) De cuántas maneras puedo seleccionar números en el Baloto, (b) si tengo cinco camisas, tres pantalones y dos pares de zapatos, ¿de cuántas formas diferentes me puedo vestir?, (c) en el campeonato de fútbol, ¿de cuántas formas posibles se pueden seleccionar los ocho equipos que llegan a las semifinales?, entre otras.

5.1. Regla de la multiplicación

La regla de multiplicación o principio multiplicativos expresa que si el proceso p_1 de puede realizar de n_1 maneras distintas, el proceso p_2 de n_2 maneras distintas, el proceso p_3 de n_3 maneras distintas. Los tres procesos a la vez, se pueden realizar de:

$$n_1 \times n_2 \times n_3$$

maneras diferentes. Note que la regla se puede aplicar para cualquier número de procesos en los que sepamos el número de formas diferentes en las que se puede realizar cada uno.

Ejemplo

Luis tiene 5 pantalones diferentes, 6 correas, 4 camisas y 3 pares de tenis. Encuentre el número de formas diferentes en las que se puede vestir.

Solución

Número de pantalones (p_1) es cinco por lo tanto $n_1 = 5$.

Número de correas (p_2) es seis por lo tanto $n_2 = 6$.

Número de camisas (p_3) es cuatro por lo tanto $n_3 = 4$.

Número de tenis (p_4) es 3 por lo tanto $n_4 = 3$.

Por lo tanto, el número de formas diferentes en las que se puede vestir es:

$$n_1 \times n_2 \times n_3 \times n_4 = 5 \times 6 \times 4 \times 3 = 360$$

Note que con un número reducido de prendas de vestir, puede estar casi todo un año sin repetir la forma de vestir.

Ejercicio

En la etapa final de fútbol profesional de primera división, cuatro equipos: Nacional (N), Medellín (M), Junior (J), Santa Fé (S), disputan el primer y segundo lugar (campeón y subcampeón). El número de maneras diferentes que estos equipos pueden ubicarse en dichos lugares es

a. 12

b. 16

c. 9

d. 4

5.2. Principio aditivo

Si se desea realizar una actividad que tiene formas alternativas para ejecutarse, en la que cada una de ellas puede hacer de maneras diferentes, se aplica el principio aditivo. Esto es, si la forma alternativa A_1 se puede realizar de n_1 maneras diferentes, la A_2 se puede realizar de n_2 formas diferentes, la A_3 se puede realizar de n_3 formas diferentes. Entonces la actividad se puede realizar de $n_1+n_2+n_3$. formas diferentes.

Ejemplo

Juan va a comprar un coche, le gustan el Mazda (M), el Renault (R) o el Kia (K). Para el Mazda puede seleccionar el 2, el 3 ó el 6, con los colores rojo, azul o negro. El Renault puede seleccionar entre Clio, Logan o Sandero con colores verde o azul. Para el Kia puede seleccionar entre el Sportage o Sorento con colores azulo negro. Encuentre las posibilidades de comprar uno de estos vehículos.

Solución

La actividad a realizar es comprar un vehículo.

Las formas alternativas son cada una de las marcas: Mazda, Renault o Kia.

Un Mazda se puede seleccionar de 3 por 3 maneras (tres modelos y tres colores). Es decir que $m = 3\times 3 = 9$.

Un Renault se puede seleccionar de 3 por 2 maneras (tres modelos y dos colores). Es decir que $r = 3\times 2 = 6$.

Un Kia se puede seleccionar de 2 por 2 maneras (dos modelos y dos colores). Es decir que $k = 2\times 2 = 4$.

Por lo tanto, teniendo en cuenta las condiciones dadas Juan tiene $m+r+k = 9+6+4 = 19$ formas diferentes de seleccionar un vehículo.

Note que para encontrar el número de posibilidades de realizar las formas alternativas se aplicó el principio multiplicativo.

Ejercicio

La familia Ospina quiere comprar una lavadora de ropa. A la señora le gusta una de las siguientes marcas: Haceb, Whirpool o General Electric, cuando acude a hacer la compra se encuentra que la lavadora H se presenta en dos tipos (8 u 11 kg), en cuatro colores diferentes y puede ser automática o semi automática, mientras que la lavadora W, se presenta en tres tipos de carga (8, 11 o 15 kg), en dos colores diferentes y puede ser automática o semiautomática y la lavadora GE, se presenta en un solo tipo de carga, de 11 kg, dos colores diferentes y solo hay semiautomática. El número de formas en qué la familia Ospina puede seleccionar el electrodoméstico es

a. 3

b. 6

c. 15

d. 30

5.3. Factorial de un número

Aunque el factorial de un número no es una regla de conteo, este concepto se utiliza para encontrar permutaciones o combinaciones de conjuntos de datos. Si n es un entero mayor que cero el factorial de n $(n!)$, es el producto de todos los enteros positivos menores o iguales que n. Esto es :

$$n! = n(n-1)(n-2(n-3)...(3)(2)(1)$$

Para el entero cero (0), se define que $0! = 1$.

Ejercicios

a. $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

b. $\frac{7!}{5!} = \frac{7 \times 6 \times 5!}{5!} = 42$

Note que para simplificar expresiones que contengan factorial, este se puede expresar como:

$n! = n(n-1)!$

$n! = n(n-1)(n-2)!$

$n! = n(n-1)(n-2)(n-3)!$, entre otras maneras.

Ejercicio

Al expandir 10! y simplificar $\frac{10!}{7!}$, se obtiene, respectivamente

a. 1.500.700 y 720

b. 3.628.800 y 720

c. 10.000 y 72

d. 3.628.800 y 1000

5.4. Permutaciones y combinaciones

Los conceptos de permutaciones y combinaciones ayudan a establecer el número de maneras en las que se puede realizar una tarea, dependiendo si importa o no el orden. Para ello es necesario establecer en que casos se aplican las permutación las combinaciones.

Permutación: aquí importa el orden y la posición de un elemento en un arreglo. Por ejemplo, seleccionar el primer, segundo, tercero y cuarto puestos en un campeonato de voleibol de playa jugado entre 10 equipos.

Combinación: se utiliza en arreglos de elementos en donde no interesa el lugar o posición que ocupa cada uno de los elementos que lo conforman. Por ejemplo, seleccionar 5 niños de la clase de matemáticas para representar al grupo en un concurso.

5.4.1. Permutaciones

Se tienen 7 alumnos que están concursando para obtener los tres primeros lugares en un concurso de lenguaje. ¿De cuántas maneras se pueden seleccionar el primero, segundo y terceros lugares?

Para ocupar el primer lugar hay 7 posibilidades.

Para ocupar el segundo lugar, después de que se selecciona el primero, solo quedan 6 posibilidades.

Para el tercer lugar, después de seleccionar el primero y el segundo, solo quedan 5 posibilidades.

Al aplicar el Principio multiplicativo se tiene que el número de posibilidades de seleccionar los tres primeros lugares es $7 \times 6 \times 5 = 210$ maneras posibles.

En general, si de un conjunto de n elementos se desean seleccionar r, con $r \leq n$ elementos en los que importa el orden se tiene que el número posibilidades es $n(n-1)(n-2)\ldots(n-r+1)$. A esto se le denomina permutaciones. La fórmula general para encontrar el número de permutaciones de un conjunto de tamaño n tomando r elementos $(r \leq n)$ está dado por:

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

Ejemplo

En una carrera automovilística compiten 6 autos. El orden de largada se hace totalmente al azar. Encontrar el número de posibilidades para seleccionar el orden de la largada.

En este caso importan el orden, por lo tanto se trata de una permutación. Note que $n = 6$ y $r = 6$. Todos los seis autos van a competir. Por lo tanto se tiene que

$$P(6,6) = 6P6 = \frac{6!}{(6-6)!} = \frac{6!}{0!} = \frac{6!}{1} = 6(5)(4)(3)(2) = 720$$

Luego, hay 720 maneras de seleccionar el orden de partida.

Ejercicio

En la Nascar, en una carrera especial compiten 8 autos. El número total de dar el orden de partida y de ocupar los tres primeros lugares, respectivamente, es

a. 200, 6

b. 8, 3

c. 40.320, 336

d. 336, 40320

5.4.2. Combinaciones

En un arreglo de tamaño n en el que se quieren extraer r elementos $(r \leq n)$ en los que no importa el orden, un elemento pude ocupar cualquiera de las r posibles posiciones. La primera la puede ocupar de r maneras, la segunda de $r-1$ maneras, continuando con este razonamiento, el elemento puede ocupar cualquiera de las r posiciones de $r!$ maneras. Por lo tanto la fórmula para calcular combinaciones de un arreglo de tamaño n tomando r elementos es

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

Observando la fórmula de las permutaciones y de las combinaciones se tiene que:

$$C(n,r) = \frac{nPr}{r!}$$

A partir de la igualdad anterior se puede observar como las combinaciones de r elementos tomados de n, pueden ser obtenidas a partir de las permutaciones de r elementos tomados de entre n, esto se debe a que como en las combinaciones no importa el orden de los elementos. Al dividir las permutaciones de los n elementos entre $r!$, se está eliminando el orden y por lo tanto transformando las permutaciones en combinaciones.

Ejemplo

Un grupo de 10 estudiantes de una institución educativa, quiere enviar una representación de 4 de ellos para hablar con el rector acerca de la realización de un campeonato interno de fútbol. Encuentre el número de formas en las que se pueden seleccionar a los 4 representantes.

Solución

En este caso no importa el orden. No se están definiendo cargos o diferentes responsabilidades dentro del grupo que va a representar, por lo tanto la solución se encuentra a partir del uso de la fórmula de las combinaciones. Aquí, se tiene que $n = 10$ y $r = 4$, por lo tanto:

$$\binom{10}{4} = C(10,4) = 10C4 = \frac{10!}{4!(10-4)!} = \frac{10!}{4!6!} = 210$$

Formas posibles de seleccionar a los cuatro representantes que hablarán con el rector.

Ejercicio

En vacaciones, se reúnen 20 amigos para realizar un campeonato de fútbol entre ellos. Quieren conformar equipos de a cinco jugadores. El número de formas posibles de seleccionar los cuatro equipos es

a. 4

b. 15.504

c. 5

d. 16.823

5.4.3. Otros ejemplos

Las reglas de conteo son muy importantes en la estadística y la probabilidad. Se debe practicar los suficiente para poder distinguir entre los diferentes casos que se puedan presentar. A continuación se exponen algunos ejemplos adicionales:

a. Combinaciones

   a. Si se cuenta con 10 alumnos que voluntariamente quieren formar parte del comité de convivencia de un colegio, ¿de cuántas maneras se pueden seleccionar a seis de ellos?

b. Permutaciones

   En un colegio hay 10 equipos de fútbol. Si entre todos ellos juegan un torneo, y tienen las mismas posibilidades de ganar el campeonato, cuántas maneras existen para estar entre los cuatro primeros lugares.

   Solución

   En un campeonato de fútbol, ningún equipo participante puede ocupar dos lugares a la vez, por lo tanto, obtener un lugar u otro en la tabla de posiciones son eventos disjuntos. Una forma de solucionar el ejercicio es aplicando el principio multiplicativo, que dice que si una tarea se puede hacer de n maneras diferentes y otra tarea distinta se puede hacer de m formas diferentes, las dos tareas se pueden hacer de n por m maneras diferentes, es decir, nm maneras.

   Primer puesto: Se puede obtener de diez (10) maneras diferentes. Hay diez equipos en el campeonato.

   Segundo puesto: Al haber un primer puesto ocupado, quedan nueve (9) posibilidades para el segundo.

   Tercer puesto: Al continuar con el mismo razonamiento, hay ocho (8) posibilidades de ser ocupado.

   Cuarto puesto: Hay siete (7) manera de ser ocupado.

   Por lo tanto, seleccionar los cuatro primeros puesto de un campeonato de fútbol, en el que participan 10 equipos se puede hacer de

$$10 \times 9 \times 8 \times 7 = 5040$$

maneras diferentes.

En este caso, el orden que ocupen los equipos si importa, por lo tanto, otra forma de solucionar el problema es aplicando la fórmula de las permutaciones directamente, es decir:

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

en dónde n es el número de elementos del conjunto del que se quieren seleccionar r sin repetición.

En nuestro caso, $n = 10$ el número de equipos que participan en el campeonato. $r = 4$, el número de equipos que se seleccionan para los primeros 4 lugares. Note que ningún equipo puede ocupar dos lugares. Por lo tanto:

$$P(10,4) = 10P4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = \frac{10\times 9\times 8\times 7\times 6\times 5\times 4\times 3\times 2\times 1}{6\times 5\times 4\times 3\times 2\times 1} = 10\times 9\times 8\times 7 = 5040$$

maneras diferentes.

Solución

En este caso, no importa el orden. Cada uno de los estudiantes seleccionados tiene el mismo grado de importancia: pertenecer a comité. Por lo tanto, el número de formas posibles de hacer la selección corresponde a una combinación. En este, caso $n = 10$, el número de todos los estudiantes que quieren participar. $r = 6$, el número de estudiantes que se pueden escoger, luego se aplica la fórmula:

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

Al reemplazar los valores dados, se obtiene:

$$C(10,6) = 10C6 = \frac{10!}{6!(10-6)!} = \frac{10!}{6!\times 4!} = \frac{10\times 9\times 8\times 7\times 6!}{6!\times 4\times 3\times 2\times 1} = \frac{10\times 9\times 8\times 7}{4\times 3\times 2\times 1} = 5\times 3\times 2\times 7 = 210$$

Por lo tanto hay 210 formas diferentes de seleccionar los miembros de dicho comité.

b. Si entre los 10 alumnos hay 4 mujeres, ¿cuántos de los grupos del comité tendrán a 3 mujeres?

Solución

De los 10 alumnos hay seis hombres. Si el comité, debe estar conformado por 3 mujeres, entonces se deben seleccionar tres hombres. En este caso, no importa el orden. Por lo tanto el problema se soluciona aplicando combinaciones.

Maneras de seleccionar 3 hombres de un total de 6 es $C(6,3) = 6C3 = 20$.

Maneras de seleccionar 3 mujeres de un total de 4 es $C(4,3) = 4C3 = 4$.

Como elegir a hombres o a mujeres son eventos independientes, para encontrar el total de formas de seleccionar el comité, con las condiciones dadas, se obtiene aplicando el principio de multiplicación, es decir:

$$C(6,3) \times C(4,3) = \frac{6!}{3!(6-3)!} \times \frac{4!}{3!(3-3)!} = 20 \times 4 = 80$$

Formas diferentes de seleccionar el comité con las condiciones pedidas, es decir, que en el comité hayan tres mujeres.

c. ¿Cuántos de los grupos del comité contarán con 3 mujeres por lo menos?

Para que en el comité hayan tres mujeres por lo menos, la posibilidad son que hayan comités con 3 ó con 4 mujeres. Por lo tanto se deben de seleccionar todas las posibilidades de seleccionar comités con 3 mujeres y comités con 4 mujeres y luego sumarlos.

Note que seleccionar comités con 3 mujeres es la solución del problema anterior, luego la respuesta es $C(6,3) \times C(4,3) = 80$.

Para seleccionar comités con exactamente cuatro mujeres se tiene que tener en cuenta que, en este caso, los comités contarían solamente con dos hombres, por lo tanto, el número de posibilidades de seleccionar este tipo de comités es:

$$C(6,2) \times C(4,4) = 15 \times 1 = 15$$

formas diferentes de seleccionar un comité en el que hayan 4 mujeres.

Por lo tanto la solución del ejercicio está dada por:

$$C(6,3) \times C(4,3) + C(6,2) \times C(4,4) = 80 + 15 = 95$$

Que es el número de comités que se pueden seleccionar en el que hayan como mínimo 3 mujeres.

Conclusión

Las reglas de conteo tienen muchas posibilidades de uso. Depende, fundamentalmente, del tipo de ejercicio que se plantea y la forma de solucionarlo.

En cada caso, hay que leer detenidamente para poder seleccionar adecuadamente la regla a aplicar.

6. Breve introducción a la probabilidad

La probabilidad es una rama de la matemática que cuantifica numéricamente la posibilidad de que un evento o suceso ocurra bajo determinadas condiciones. Hay que tener en cuenta que el hecho que se le asigne un valor a la posibilidad de ocurrencia de un evento eso no quiere decir que siempre que se den esas mismas condiciones el suceso tiene que ocurrir.

En la matemática se estudian fundamentalmente tres tipos de eventos:

Determinísticos: Aquellos que bajo las mismas condiciones el resultado esperado siempre es el mismo. Por ejemplo, lanzar una piedra hacia arriba. En este caso se espera que la piedra siempre retorne a la tierra.

Probabilísticos: Son los eventos en los que a pesar de tener las mismas condiciones el resultado del experimento no se sabe con anticipación. Por ejemplo, si una persona lanza una piedra varias veces, la distancia a la que cae no es la misma en cada acción.

Determinísticos y probabilísticos: Son aquellos en que una parte del experimento tiene un resultado determinado y otra es probabilístico. Por ejemplo el costo del recibo de la energía que paga un hogar mes a mes. La parte determinística está dada por los costos fijos que cobran las empresas públicas por el mantenimiento de la red eléctrica y la parte probabilística está dada por el costo de los kilovatios hora que consume el hogar mes a mes. Note que no siempre es el mismo consumo.

La Teoría de la Probabilidad estudia estos dos últimos tipos de eventos.

Ejercicios

Encuentre en su entorno diferentes fenómenos que cumplan con las características de ser determinísticos o probabilísticos o que sean de las dos formas a la vez.

Para estudiar el tema de probabilidades hay que comprender dos conceptos previos, estos son: espacio muestral y evento o suceso.

Espacio muestral

Es el conjunto formado por todos los resultados posibles de un experimento aleatorio.

Ejemplos

Lanzar una moneda al aire es un experimento aleatorio, los resultados posibles son cara (c) o sello (s). Por lo tanto el espacio muestral se puede definir como el conjunto $S = \{c, s\}$.

Otro ejemplo clásico es lanzamiento de un dado y observar el número que queda en la cara superior. El espacio muestral está dado por: $D = \{1, 2, 3, 4, 5, 6\}$

Ejercicio

Se lanzan al aire dos monedas a la vez, de denominación diferente y se anota el resultado que se obtiene en cada uno de los lanzamientos. El espacio muestral M está dado por

a. $M = \{(c, s), (s, c)\}$

b. $M = \{(c, c), (s, s)\}$

c. $M = \{(c, c), (c, s), (s, s)\}$

d. $M = \{(c, c), (c, s), (s, c), (s, s)\}$

Evento

Un evento es un subconjunto de elementos de un espacio muestral.

Ejemplos

Para el caso del lanzamiento de una moneda los eventos posibles son $\emptyset$, {c}, {s}, {c,s}. Recordar que el conjunto vacío $\emptyset$ es subcojunto de culquier cojunto. Un conjunto es su subconjunto del mismo conjunto. Cuando el evento es el conjunto vacío ($\emptyset$) se dice que es el evento imposible. Todo el espacio muestral es un evento seguro.

Para espacios muestrales M finitos de tamaño n, el número posible de eventos es 2^n que es el cardinal (número de elementos), del conjunto de partes de M que es $\mathcal{P}(M)$.

Para el caso del lanzamiento de una moneda se tiene que:

$$\mathcal{P}(M) = \{\emptyset, \{c\}, \{s\}, \{c, s\}\}$$

son todos los subconjuntos posibles del espacio muestral, por lo tanto, todos los eventos que se pueden definir.

Definir claramente el espacio muestral de un experimento aleatorio y sus eventos es un requisito fundamental para poder calcular correctamente las probabilidades asociadas.

Ejercicio

Se realiza un experimento aleatorio y se encontró que el espacio muestral está dado por $A = \{a,b,c\}$, el conjunto de todos los eventos posibles es

a. $\mathcal{P}(A) = \{\{a\},\{b\},\{c\},\{a,b\},\{a,c\},\{b,c\},\{a,b,c\}\}$

b. $\mathcal{P}(A) = \{\emptyset,\{a\},\{b\},\{c\},\{a,b\},\{a,c\},\{b,c\}\}$

c. $\mathcal{P}(A) = \{\emptyset,\{a\},\{b\},\{c\},\{a,b\},\{a,c\},\{b,c\},\{a,b,c\}\}$

d. $\mathcal{P}(A) = \{\emptyset,\{a\},\{b\},\{c\},\{a,b\},\{b,a\},\{b,c\},\{a,b,c\}\}$

Ejercicio

Lanzar un dado no cargado y anotar el número que queda en la cara superior es un experimento aleatorio. El número posible de eventos del espacio muestral $D = \{1,2,3,4,5,6\}$, está dado por

a. 64

b. 32

c. 20

d. 6

6.1. La probabilidad

En el estudio de la probabilidad, se distinguen tres tipos:

Subjetiva: ocurre cuando se le asigna una probabilidad de ocurrencia a un evento o suceso de acuerdo con la experiencia, la opinión personal o la intuición, que tiene una persona en relación con un suceso o evento. Por ejemplo, Juan puede afirmar que hay una probabilidad del 60 % de que a las tres de la tarde del sábado esté lloviendo. Observe que parte de la opinión o experiencia que tiene Juan de la observación del clima en una determinada época del año.

Clásica o a priori: si un suceso puede ocurrir de n maneras mutuamente excluyentes e igualmente probables y r de ellas poseen una característica A, entonces:

$$P(A) = \frac{r}{n} = \frac{\text{Número de casos favorables a A}}{\text{Número de casos posibles}}$$

Por ejemplo: $P(\text{Salga un cinco al tirar un dado legal}) = \dfrac{1}{6}$

Empírica o frecuencial: surge de la experimentación que se realiza en las diferentes ramas del conocimiento, como la Física, la Química, la Matemática, la Biología, las Ciencias Sociales, entre otras. Al realizar un experimento aleatorio, se obtienen muchos resultados que muestran frecuencia de ocurrencia. Así, si un experimento aleatorio, se ha realizado n veces y el resultado (suceso) A ha ocurrido r veces, la probabilidad de A, está dada por $P(A) = \frac{r}{n}$.

Por ejemplo, si se encuestan a 30 personas acerca de sus gustos por diferentes tipos de deportes y 5 manifestaron preferencias por el fútbol, se tiene que el número de veces que se realizó el experimento (hacer la encuesta) es 30, por lo tanto $n = 30$. Las personas que respondieron la encuesta y manifestaron preferencias (frecuencia) por el fútbol es de cinco, luego $r = 5$. Si se define el suceso A como el gustos por el fútbol, $P(A) = \frac{5}{30} = \frac{1}{6}$.

6.1.1. Propiedades de la probabilidad

Al calcular la probabilidad de un evento o suceso, se deben de tener en cuenta las siguientes propiedades:

a. La probabilidad de un evento imposible es 0 (cero). Anteriormente se denotó el evento imposible como $\emptyset$, luego $P(\emptyset) = 0$.

b. Para un experimento aleatorio, a cada uno de sus resultados se le asigna una probabilidad entre 0 y 1. Es decir, si A_i es un resultado de un experimento $0 \leq P(A_i) \leq 1$, para toda i.

c. Si A es un evento de un espacio muestral, $P(A^c) = 1 - P(A)$

d. La suma de las probabilidades de todos los resultados posibles A_i de un experimento es igual a 1. Es decir: $P(A_1) + P(A_2) + \ldots + P(A_i) + \ldots + P(A_n) = 1$ en donde $1 \leq i \leq n$, y n es el número de resultados posibles.

e. Si A_i y A_j, son eventos disjuntos (no tienen elementos en común) $P(A_i \cup A_j) = P(A_i) + P(A_j)$, $P(A_i \cap A_j) = 0$

f. Si A_i y A_j, tienen elementos en común, $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j)$.

g. La probabilidad del espacio muestral es 1. Si M es todo el espacio muestral, $P(M) = 1$, a M se le llama el evento seguro.

Ejemplo

Si se lanza un dado legal y se anota el número que aparece en la cara superior, el espacio muestral es $D = \{1, 2, 3, 4, 5, 6\}$. Si A_i es el suceso de que aparezca el número i, encuentre las siguientes probabilidades: $P(A_1)$, $P(A_3 \cup A_4)$ y $P(A_2 \cap A_5)$.

Solución

A_1 es el evento que indica que al lanzar un dado y observar la cara superior se obtuvo el número 1. Por lo tanto, los casos favorables es 1, luego $r = 1$. Los casos posibles son 6, luego $n = 6$ que es el tamaño del espacio muestral o casos favorables. Por lo tanto $P(A_1) = \frac{1}{6}$.

A_3 y A_4 son eventos disjuntos. En el experimento, no puede ocurrir que se obtengan los número 3 y 4 al mismo tiempo al lanzar el dado bajo las condiciones dadas. Por lo tanto $P(A_3 \cup A_4) = P(A_3) + P(A_4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

A_2 y A_5 son eventos disjuntos. En el experimento, no puede ocurrir que se obtengan los número 3 y 5 al mismo tiempo al lanzar el dado, bajo las condiciones dadas, de donde $A_2 \cap A_5 = \emptyset$. Por lo tanto $P(A_3 \cap A_4) = P(\emptyset) = 0$.

Ejemplo

En el grado octavo de una institución educativa, hay 75 alumnos que compiten en diferentes equipos. En fútbol, el equipo está conformado por 25 alumnos, de los cuales también juegan basquet, cuyo equipo esta conformado por 15 estudiantes. Los 40 restantes, compiten en otras disciplinas. Encuentre la probabilidad de que un estudiante:

a. Compite en fútbol y basquet.

b. Compite en fútbol o basquet.

c. Compite solamente en fútbol.

d. Compite solamente en basquet.

e. Compite en otro deporte diferente a fútbol o basquet.

Solución

El espacio muestral son los 75 estudiantes que compiten (C) en las disciplinas en las que participan en el grado octavo. Por lo tanto $n = 75$, que es el tamaño del espacio muestral. De acuerdo con el ejercicio se pueden definir los siguientes eventos:

F: estudiantes que compiten en fútbol.

B: estudiantes que compiten en basquet.

O: estudiantes que compiten en otras disciplinas.

SF: estudiantes que compiten solamente en fútbol.

SB: estudiantes que compiten solamente en basquet.

Para encontrar el tamaño de cada evento se puede hacer a partir de un diagrama de Venn, de la siguiente manera:

Figura:

De dónde se tiene que:

a. El evento "Compite en fútbol y basquet" está dado por $F \cap B$. De acuerdo con la gráfica **??**, el número de posibilidades de particiapar en los dos eventos es de 5. Por lo tanto, la probabilidad está dada por:

$$P(F \cap B) = \frac{5}{25} = \frac{1}{5} = 0.2$$

b. El evento "Compite en fútbol o basquet", se puede expresar como $F \cup B$, que de acuerdo con la figura tiene 35 posibilidades a favor. Por lo tanto la probabilidad de que un estudiante esté compitiendo en uno de estos dos deporte, está dada por:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{25}{75} + \frac{15}{75} - \frac{5}{75} = \frac{35}{75} = \frac{7}{15} \approx 0.466$$

c. El evento "Compite solamente en fútbol" está dado por $F - B$, que tiene 20 casos favorables. Por lo tanto, la probabilidad está dada por:

$$P(F - B) = P(SF) = \frac{20}{75} = \frac{4}{15} \approx 0.266$$

d. El evento "Compite solamente en basquet" esta dado por $B - F$, que tiene 10 casos favorables. Por lo tanto, la probabilidad esta dada por:

$$P(B - B) = P(SB) = \frac{10}{75} = \frac{2}{15} \approx 0.133$$

e. El evento "Compite en otro deporte diferente a fútbol o basquet" esta dado por $C - (F \cup B)$, este evento tiene 40 casos favorables. Otra forma es teniendo en cuenta que los estudiantes juegan otros deportes está dado por $(F \cup B)^c$. De dónde se tiene que la probabilidad está dada por:

$$P(C - (F \cup B)) = \frac{40}{75} = \frac{8}{15} \approx 0.533$$

$$P((F \cup B)^c) = 1 - P(F \cup B) \approx 1 - 0.466 \approx 0.533$$

Ejercicio

Suponga que se tiene el espacio muestral $M = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$, en donde $A_1, \ldots, A_7$ denotan puntos muestrales. Cada evento tiene las siguientes probabilidades: $P(A_1) = 0.05$, $P(A_2) = 0.20$, $P(A_3) = 0.20$, $P(A_4) = 0.25$, $P(A_5) = 0.15$, $P(A_6) = 0.10$, $P(A_7) = 0.05$.
Si $Z = \{A_1, A_4, A_6\}$, $X = \{A_2, A_4, A_7\}$, $Y = \{A_2, A_3, A_5, A_7\}$
La probabilidad de $P(Z)$, $P(X)$ y $P(Y)$, respectivamente, es

a. 0.4, 0.5 y 0.6

b. 0.5, 0.4 y 0.6

c. 0.5, 0.5 y 0.4

d. 0.5, 0.6 y 0.6

Ejercicio

Suponga que se tiene el espacio muestral $M = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$, en donde $A_1, \ldots, A_7$ denotan puntos muestrales. Cada evento tiene las siguientes probabilidades: $P(A_1) = 0.05$, $P(A_2) = 0.20$, $P(A_3) = 0.20$, $P(A_4) = 0.25$, $P(A_5) = 0.15$, $P(A_6) = 0.10$, $P(A_7) = 0.05$.
Si $Z = \{A_1, A_4, A_6\}$, $X = \{A_2, A_4, A_7\}$, $Y = \{A_2, A_3, A_5, A_7\}$
La probabilidad de $P(Z \cup X)$ es

a. 0.4

b. 0.65

c. 0.3

d. 0.8

Ejercicio

Suponga que se tiene el espacio muestral $M = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$, en donde $A_1, \ldots, A_7$ denotan puntos muestrales. Cada evento tiene las siguientes probabilidades: $P(A_1) = 0.05$, $P(A_2) = 0.20$, $P(A_3) = 0.20$, $P(A_4) = 0.25$, $P(A_5) = 0.15$, $P(A_6) = 0.10$, $P(A_7) = 0.05$.
Si $Z = \{A_1, A_4, A_6\}$, $X = \{A_2, A_4, A_7\}$, $Y = \{A_2, A_3, A_5, A_7\}$
La probabilidad de $P(Z \cap X)$ es

a. 0.45

b. 0.65

c. 0.25

d. 0.7

Ejercicio

Suponga que se tiene el espacio muestral $M = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$, en donde $A_1, \ldots, A_7$ denotan puntos muestrales. Cada evento tiene las siguientes probabilidades: $P(A_1) = 0.05$, $P(A_2) = 0.20$, $P(A_3) = 0.20$, $P(A_4) = 0.25$, $P(A_5) = 0.15$, $P(A_6) = 0.10$, $P(A_7) = 0.05$.
Si $Z = \{A_1, A_4, A_6\}$, $X = \{A_2, A_4, A_7\}$, $Y = \{A_2, A_3, A_5, A_7\}$
La probabilidad de $P(Y^c)$ es

a. 0.45

b. 0.65

c. 0.3

d. 0.6

Ejercicio

En un colegio le prestan computadores portatiles a sus estudiante. Una encuesta sobre el uso de estos equipos, en el último mes, reveló que 48.5 % lo prestaron por razones de estudio, 54 % por razones de estudio y de juego, 30 % por razones de estudio y de juego.
La probabilidad de que un estudiante haya pedido prestado el computador por razones de estudio o de juego en el último mes es

a. 1

b. 0.725

c. 0.185

d. 0.6

Ejercicio

En un colegio le prestan computadores portátiles a sus estudiante. Una encuesta sobre el uso de estos equipos, en el último mes, reveló que 48.5 % lo prestaron por razones de estudio, 54 % por razones de estudio y de juego, 30 % por razones de estudio y de juego.
La probabilidad de que un estudiante no haya pedido prestado un computador en el último mes es

a. 0.9

b. 0.725

c. 0.275

d. 0.625

Ejercicio

En un colegio le prestan computadores portátiles a sus estudiante. Una encuesta sobre el uso de estos equipos, en el último mes, reveló que 48.5 % lo prestaron por razones de estudio, 54 % por razones de estudio y de juego, 30 % por razones de estudio y de juego.
La probabilidad de que un estudiante haya pedido prestado un computador solo para jugar es

a. 0.24

b. 0.8

c. 0.275

d. 0.7

Ejercicio

En un colegio le prestan computadores portátiles a sus estudiante. Una encuesta sobre el uso de estos equipos, en el último mes, reveló que 48.5 % lo prestaron por razones de estudio, 54 % por razones de estudio y de juego, 30 % por razones de estudio y de juego.
La probabilidad de que un estudiante haya pedido prestado un computador solo para estudiar es

a. 0.24

b. 0.18

c. 0.275

d. 0.185

7. Ejercicios

1. La siguiente variable corresponde a una variable cuantitativa

   A. Comida favorita
   B. Color de los ojos de tus compañeros de clase
   C. Número de alumnos de un colegio
   D. Estado de ánimo de un grupo de empleados en una empresa

2. La siguiente variable corresponde a una variable cualitativa

a. Número de goles marcados en un campeonato por el equipo campeón en un campeonato
b. Estatura en centímetros de los estudiantes de un colegio
c. Número de alumnos de un colegio
d. La profesión de una persona

3. La siguiente variable corresponde a una variable discreta

a. Temperaturas registrada en un salón de clase
b. Número de hijos de 10 familias
c. Tiempo de vida útil de un computador
d. Estatura en centímetros de un grupo de 40 personas

4. Las calificaciones en matemáticas de un grupo de estudiantes fueron: 8, 1, 10, 3, 7, 3, 5, 6, 10, 9. El rango y la media son respectivamente iguales a:

a. 9 6.5
b. 10 6.2
c. 10 6.5
d. 9 6.2

5. Las estaturas en metros de un grupo de 15 estudiantes son: 1.7, 1.5, 1.7, 1.0, 1.6, 1.2, 1.6, 1.4, 1.8, 1.8, 1.5, 1.4, 1.6, 1.6, 1.5. Entonces, la media y la varianza de la muestra son respectivamente iguales a

a. 1.5266 $0,0433$
b. 1.5266 $0,0464$
c. 1.6666 $0,0464$
d. 1.5266 $0,0550$

6. Las calificaciones en biología de un grupo de diez estudiantes fueron 8, 9, 8, 9, 4, 5, 4, 6, 8, 3. La moda y la mediana son respectivamente iguales a

a. 8 6
b. 4 7
c. 8 7
d. 8 8

7. En cierta región los registros que se tomaron a 10 familias, sobre el número de hijos fueron: 3, 8, 2, 5, 3, 1, 4, 2, 5, 2. La moda y la mediana son respectivamente iguales a

a. 3 3.0
b. 3 3.5
c. 2 3.0
d. 2 3.3

8. En una escuela rural, diez estudiantes obtuvieron en español los siguientes puntajes:
   4, 7, 8, 5, 2, 2, 3, 5, 3, 5. La media y la moda son respectivamente iguales a:

   a. 4.4 3
   b. 3.6 5
   c. 4.0 2
   d. 4.4 5

9. Las calificaciones que obtuvieron en sociales diez estudiantes, en una escuela para niños, fueron las siguientes: 9.0, 2.0, 3.0, 8.0, 3.0, 8.0, 3.0, 6.0, 4.0, 5.0. La mediana y la media son respectivamente iguales a

   a. 3.0 5.0
   b. 3.0 5.1
   c. 4.5 5.1
   d. 4.5 6.0

10. Suponga que las calificaciones que obtuvieron en matemáticas diez estudiantes, en una institución educativa, fueron las siguientes: 9, 2, 8, 2, 7, 4, 5, 4, 4, 5. La mediana y la moda son respectivamente iguales a:

    a. 4.8 4
    b. 4.5 4
    c. 4.5 2
    d. 5.5 4

11. El peso en kilogramos de quince productos cárnicos en un supermercado son: 2.5, 2.0, 1.0, 1.3, 1.8, 1.7, 1.6, 1.4, 1.5, 1.5, 2.4, 2.5, 1.7, 1.3, 1.5. Entonces la media y la desviación estándar de la muestra son respectivamente iguales a

    a. 1.7133 1.7022

    b. 1.7133 0.4395

    c. 1.6 6.4549

    d. 1.7133 0.1936

12. El peso en kilogramos de quince productos cárnicos en un supermercado son: 2.5, 2.0, 1.0, 1.3, 1.8, 1.7, 1.6, 1.4, 1.5, 1.5, 2.4, 2.5, 1.7, 1.3, 1.5. Entonces el número de datos que están a menos de una desviación estándar es tal que

    a. Es menor que 4

    b. Está entre 4 y 6

    c. Está entre 6 y 10

    d. Es mayor que 10

13. En una prueba a diez estudiantes, se obtuvieron las siguientes calificaciones en matemáticas: 5, 8, 6, 7, 4, 8, 10, 7, 10, 9. La media y las calificaciones que están a una desviación estándar de la media son respectivamente iguales a

    a. 7; 6, 7, 7, 8, 8, 9

    b. 7.4; 6, 7, 7, 8, 8, 9

    c. 7.4; 7, 7, 8, 8, 9

    d. 7.4; 6, 7, 7, 8, 8

14. En una prueba a diez estudiantes, se obtuvieron las siguientes calificaciones en matemáticas: 5, 8, 6, 7, 4, 8, 10, 7, 10, 9. El porcentaje de alumnos sacaron calificaciones que están a una desviación estándar de la media es

    a. 60 %
    b. 40 %
    c. 50 %
    d. 80 %

15. En una prueba a diez estudiantes, se obtuvieron las siguientes calificaciones en biología: 5, 8, 6, 7, 4, 8, 10, 7, 10, 9. La media y las calificaciones que están a dos desviación estándar de la media son respectivamente iguales a

    a. 7; 4, 5, 10, 10
    b. 7.4; 4, 5, 6, 7, 7, 8, 8, 9, 10, 10
    c. 7; 4, 5, 6, 7, 7, 8, 8, 9, 10, 10
    d. 7.4; 4, 5, 10, 10

16. En un campeonato participan seis equipos. El total de posibilidades en que puede terminar el campeonato es igual a

    a. 15
    b. 46656
    c. 720
    d. 32

17. En un campeonato participan siete equipos. El total de posibilidades en que puede terminar el campeonato es igual a

    a. 49
    b. 5040
    c. 720
    d. 490

18. En un campeonato participan ocho equipos, de los cuales únicamente clasifican cuatro. El total de posibles clasificaciones que se pueden presentar es igual a

    a. 70
    b. 1680
    c. 840
    d. 256

19. De un grupo de 8 personas conformado por 5 hombres y 3 mujeres, se requiere conformar un comité de 5 personas, conformado por 3 hombres y 2 mujeres. El número de posibilidades está dado por

    a. 56
    b. 36
    c. 40
    d. 30

20. De un grupo de 10 personas conformado por 6 hombres y 4 mujeres, se requiere conformar un comité de 5 personas, en donde el número de mujeres pueden ser 2 o 3. El número de posibilidades está dado por

    a. 112
    b. 60
    c. 180
    d. 120

21. Para realizar un procedimiento de control de calidad en una fábrica, se revisan al azar 4 de 10 productos para examinar y determinar si son defectuosos. El número de muestras posibles es igual a

    a. 720
    b. 210
    c. 120
    d. 40

22. En una rifa se utilizan de manera aleatoria 4 letras de un grupo de 9 letras diferentes, para determinar el ganador. El número de posibles selecciones es igual a

    a. 210
    b. 630
    c. 3024
    d. 126

23. En cierto cruce, un vehículo tiene tres opciones: a) dar vuelta a la izquierda, b) dar vuelta a la derecha, y c) seguir de largo. La probabilidad de que los vehículos den vuelta a la izquierda es de 0.25, la probabilidad que los vehículos den vuelta a la derecha es de 0.20. Entonces la probabilidad de que los vehículos sigan de largo es igual a

    a. 0.45
    b. 0.55
    c. 0.75
    d. 0.80

24. Se requiere conformar comités de dos físicos y de dos químicos, sabiendo que se disponen de cuatro químicos y tres físicos. El número de posibles comités es igual a

    a. 21
    b. 42
    c. 9
    d. 18

25. En la tabla aparece el registro acerca del número de caries de cada uno de los estudiantes de un grupo de 20.

Número de caries	Frecuencia absoluta	Frecuencia relativa
0	2	0.10
1	4	0.20
2	5	0.25
3	8	x
4	y	z

El número medio de caries es igual a

a. 5
b. 2.1
c. 2.2
d. 8.8

26. En la tabla aparece el registro acerca del número de caries de cada uno de los estudiantes de un grupo de 20.

Número de caries	Frecuencia absoluta	Frecuencia relativa
0	2	0.10
1	4	0.20
2	5	0.25
3	8	x
4	y	z

La desviación estándar de la muestra es igual a

a. 1.2526
b. 1.0241
c. 1.1192
d. 1.0909

27. En la tabla aparece el registro acerca del número de caries de cada uno de los estudiantes de un grupo de 20.

Número de caries	Frecuencia absoluta	Frecuencia relativa
0	2	0.10
1	x	y
2	5	0.25
3	8	0.40
4	1	0.05

El porcentaje de estudiantes que registraron una caries es igual a:

a. 10 %
b. 20 %
c. 30 %
d. 40 %

28. En la tabla aparece el registro acerca del número de caries de cada uno de los estudiantes de un grupo de 20.

Número de caries	Frecuencia absoluta	Frecuencia relativa
0	2	0.10
1	4	0.20
2	5	0.25
3	8	0.40
4	1	0.05

El porcentaje de estudiantes que registraron más de dos caries es igual a

a. 45 %

b. 70 %

c. 65 %

d. 30 %

29. Las calificaciones obtenidas por un grupo de 16 estudiantes sobre un examen de física aparecen en la siguiente tabla.

6	7	3	8
10	1	5	9
7	7	3	6
10	9	6	8

El porcentaje de estudiantes que registraron una nota superior a 8 es igual a

a. 20 %

b. 28 %

c. 25 %

d. 30 %

30. Las calificaciones obtenidas por un grupo de 16 estudiantes sobre un examen de física, aparecen en la siguiente tabla.

6	7	3	8
10	1	5	9
7	7	3	6
10	9	6	8

El porcentaje de estudiantes que registraron una nota a menos de una desviación estándar de la media es igual a

a. 71.24 %

b. 31.25 %

c. 65.42 %

d. 68.75 %

31. El número de maneras en que se pueden colocar siete bolas, donde hay cuatro bolas rojas y tres bolas verdes, es igual a

a. 35

b. 12

c. 144

d. 6

32. Se ha encargado la impresión de un documento a una imprenta, la cual imprime 18 documentos defectuosos de cada 150. La probabilidad de que al elegir un documento, éste salga defectuoso es igual a

a. 0.82

b. 0.18

c. 0.12

d. 0.88

33. En cierta fábrica se ha medido la longitud de 16 piezas de características similares. Los registros en cm. son los siguientes: 6.2, 11.5, 12.3, 12.1, 8.8, 7.5, 5.5, 5.9, 6.6, 8.3, 9.8, 10.4, 12.8, 12.4, 12.5, 7.2. En la siguiente tabla aparecen los registros en cuatro clases.

Longitud en cm.	Número de piezas
[5,7)	4
[7,9)	
[9,11)	
[11,13]	6

Al completar la tablas se tiene que la frecuencia relativa correspondiente a los registros que están en la clase $[7,9)$ es igual a

a. 0.20

b. 0.25

c. 0.30

d. 0.50

34. En cierta fábrica se ha medido la longitud de 16 piezas de características similares. Los registros en cm. son los siguientes: 12.3, 12.1, 8.8, 7.5, 5.5, 5.9, 6.2, 6.6, 8.3, 9.8, 10.4, 12.8, 12.4, 11.5, 12.5, 7.2. En la siguiente tabla aparecen los registros en cuatro clases.

Longitud en cm.	Número de piezas
[5,7)	4
[7,9)	
[9,11)	
[11,13]	

Al completar la tablas se tiene que la frecuencia relativa porcentual, correspondiente a los registros que están en la clase $[11,13]$ es igual a

a. 25%

b. 0.25%

c. 0.375%

d. 37.5%

35. En un registro sobre un grupo de personas alcohólicas se tiene que el 35 % tienen el padre alcohólico, mientra que el 25 % de los mismos tienen la madre alcohólica. Si se sabe que el 50 % tienen al menos uno de los padres alcohólico, entonces el porcentaje de personas que tienen ambos padres alcohólicos es igual a

    a. 25 %
    b. 20 %
    c. 15 %
    d. 10 %

36. En un registro sobre un grupo de personas alcohólicos se tiene que el 50 % tienen el padre alcohólico, mientras que el 35 % de los mismos tienen la madre alcohólica. Si se sabe que el 70 % tienen al menos uno de los padres alcohólico, entonces el porcentaje de personas que tienen el padre alcohólico y la madre no alcohólica es igual a

    a. 35 %
    b. 20 %
    c. 15 %
    d. 10 %

37. En un examen, el 25 % de los hombres lo perdieron, así mismo el 10 % de las mujeres lo perdieron. Sabiendo que 40 estudiantes presentaron el examen y 7 estudiantes lo perdieron, el número de mujeres que ganaron el examen es igual a

    a. 10
    b. 18
    c. 12
    d. 15

38. En un examen de español el 20% de los hombres perdieron el examen, mientras que el 25% de las mujeres perdieron el examen. Sabiendo que 40 estudiantes presentaron el examen y 9 estudiantes lo perdieron, el número de hombres que ganaron el examen is igual a

    a. 16
    b. 15
    c. 12
    d. 20

39. La ventas en cierto número de empresas son la que se indican a continuación

Ventas (millones)	5	3	4	6
Empresas	3	4	6	5

    Las ventas medias y la desviación estándar de la muestra son respectivamente iguales a

    a. 4.0 millones 1.004
    b. 4.5 millones 1.200
    c. 4.5 millones 1.1180
    d. 4.5 millones 1.1504

40. En un campeonato participan 6 equipos, de los cuales únicamente clasifican cuatro. El número de posibilidades de que clasifique un equipo entre los primeros cuatro es igual a

    a. 256
    b. 30
    c. 15
    d. 360

41. Dados los números 1, 2, 3, 4, y 5, la cantidad de los posibles números que se pueden formar de 3 dígitos, sin repetir dígito, es igual a

    a. 10
    b. 60
    c. 30
    d. 120

42. Dados 8 bombillos, de los cuales, 5 son rojos y 3 son verdes, el número posible de arreglos de 8 bombillos que se pueden realizar es igual a

    a. 28
    b. 42
    c. 56
    d. 112

43. De un grupo de 10 personas se eligen 4 para tomar una foto en forma lineal, respetando el orden, el número de disposiciones posibles es igual a

    a. 420
    b. 210
    c. 151200
    d. 5040

44. De un grupo de 10 personas se eligen 4 para tomar una foto en forma lineal, sin importar el orden, el número de disposiciones posibles es igual a

    a. 420
    b. 210
    c. 2520
    d. 5040

45. En una reunión de 10 socios se pretende elegir a tres de estos socios, para los cargos de un Presidente, un Secretario y un Tesorero. El número de posibles elecciones es igual a

    a. 360
    b. 1000
    c. 120
    d. 720

46. En una reunión de 10 socios se pretende elegir a una comisión conformada por cuatro de estos socios. El número de posibles comisiones es igual a

    a. 210
    b. 5040
    c. 256
    d. 10000

47. En una reunión de 10 socios se pretende elegir a una comisión conformada por cuatro de estos socios. El número de posibles comisiones es igual a

    a. 210
    b. 5040
    c. 256
    d. 10000

48. Se dispone de cinco bolas de colores: roja, verde, amarilla, blanca y negra. El número de maneras en que se pueden dividir las cinco bolas en dos grupos, uno con dos bolas y otro con tres bolas, es igual a

    a. 60
    b. 10
    c. 36
    d. 50

49. Se dispone de 9 jugadores para conformar equipos de Básquetbol de 5 jugadores, donde los jugadores pueden desempeñar cualquier puesto en el equipo. El número de equipos que se pueden conformar es igual a

    a. 59049
    b. 15120
    c. 252
    d. 126

50. En una bolsa hay cinco bolas de colores: roja, verde, amarilla, blanca y negra. Si se extraen de la bolsa tres, el número de formas diferentes en que pueden aparecer es igual a

    a. 20
    b. 30
    c. 60
    d. 10

51. El número de permutaciones diferentes que se pueden formar con todas las letras de la palabra INDIVIDUO es igual a

    a. 120
    b. 30240
    c. 60480
    d. 720

52. El número de permutaciones diferentes que se pueden formar con todos los dígitos del número 314141 es igual a

    a. 60
    b. 120
    c. 30
    d. 720

Índice

1. Recolección de la información **2**

1.1. Tipos de datos 2

2. Algunos estadísticos básicos **5**

2.1. La media o promedio $\overline{x}$ de una muestra 5

2.2. La mediana M_e de una muestra 6

2.3. La moda M_o de una muestra 7

2.4. El rango R de una muestra 8

2.5. La varianza s^2 de una muestra 9

2.6. Cuartiles 11

2.7. La desviación estándar s de una muestra 13

3. Agrupamiento de datos de una muestra aleatoria **19**

3.1. Número de clases N_c para datos cuantitativos 19

3.2. Ancho de las clases A_c para datos cuantitativos 20

3.3. Cálculo de las clases para datos cuantitativos 20

3.3.1. Recuento de los datos 22

3.3.2. Frecuencia simple f_i 23

3.3.3. Frecuencia simple acumulada F_i 24

3.3.4. Frecuencia relativa (f_r) y frecuencia relativa acumulada (F_r) 25

3.3.5. Frecuencia relativa porcentual $(f_r\,\%)$ y frecuencia relativa acumulada porcentual $(F_r\,\%)$ 25

3.4. Agrupamiento de datos de una muestra aleatoria cualitativa 28

4. Actividades de ejercitación **30**

4.1. Apreciación de la calidad de un colegio 30

4.2. Apreciación de los jugadores de la selección Colombia 31

4.2.1. Histograma 32

4.2.2. Diagrama de torta 32

5. Reglas de conteo **36**

5.1. Regla de la multiplicación 36

5.2. Principio aditivo 37

5.3. Factorial de un número 38

5.4. Permutaciones y combinaciones 39

5.4.1. Permutaciones 39

5.4.2. Combinaciones 41

5.4.3. Otros ejemplos 42

6. Breve introducción a la probabilidad **45**

6.1. La probabilidad 47

6.1.1. Propiedades de la probabilidad 48

7. Ejercicios **54**

8. Bibliografía **71**

8. Bibliografía

1. Anderson, Sweeney, Williams. (2008). Estadística para Administración y Economía. Décima edición. Cengage Learning, México.
2. Newbod, P. (1998). Estadística para los negocios y la economía. Cuarta Edición. Prentice Hall, México.
3. García A., Miguel A. (2005). Introducción a la teoría de la probabilidad. Ciencia y Tecnología, Fondo de cultura económica, México.