ESTADÍSTICA DESCRIPTIVA Y FUNDAMENTOS DE PROBABILIDAD CON R, RSTUDIO Y GEOGEBRA

Liliana García Barco
Luisa María Montoya Conde

UNIVERSIDAD COLEGIO
MAYOR DE CUNDINAMARCA
SELLO EDITORIAL

ESTADÍSTICA DESCRIPTIVA
Y FUNDAMENTOS DE PROBABILIDAD
CON R, RSTUDIO Y GEOGEBRA

Liliana García Barco
Luisa María Montoya Conde

ESTADÍSTICA DESCRIPTIVA Y FUNDAMENTOS DE PROBABILIDAD CON R, RSTUDIO Y GEOGEBRA

UNIVERSIDAD COLEGIO
MAYOR DE CUNDINAMARCA

SELLO EDITORIAL

Catalogación en la publicación – Biblioteca Nacional de Colombia

García Barco, Liliana, autora
Estadística descriptiva y fundamentos de probabilidad de R, RStudio y Geogebra / Liliana García Barco, Luisa María Montoya Conde. -- Bogotá: Sello Editorial Universidad Colegio Mayor de Cundinamarca, 2025.

200 páginas.

Incluye datos curriculares de las autoras -- Incluye glosario -- Incluye referencias bibliográficas.

ISBN 978-958-5198-41-8 -- 978-958-5198-42-5 (e-book)

1. Probabilidades - Investigaciones 2. Estadística matemática 3. RStudio (Programa para computador) - Enseñanza - Problemas, ejercicios, etc. 4. RStudio (Programa para computador) - Enseñanza - Problemas, ejercicios, etc. 5. Geogebra (Programa para computador) - Enseñanza - Problemas, ejercicios, etc. I. Montoya Conde, Luisa María, autora

CDD: 519.530785 ed. 23 CO-BoBN– a1155137

© Liliana García Barco
Profesora Asociada, Universidad Colegio Mayor de Cundinamarca

© Luisa María Montoya Conde
Profesora Auxiliar, Universidad Colegio Mayor de Cundinamarca

Sello Editorial Universidad Colegio Mayor de Cundinamarca
Carrera 13 No. 38- 29, Edificio San Juan, noveno piso
selloeditorial@unicolmayor.edu.co
www.unicolmayor.edu.co

ISBN: 978-958-5198-41-8
ISBN e-book: 978-958-5198-42-5

Corrector de Estilo: VJ Romero

Diagramación electrónica: Yaneth Guarín A.

Imagen de portada: Tomada de Freepik
Diseño de portada: Vanessa Peña A.

Impresión: Grupo Editorial Ibáñez

CONTENIDO

Capítulo 3. Conceptos básicos de la probabilidad

Capítulo 4. Variables aleatorias discretas

Capítulo 5. Algunas distribuciones discretas de probabilidad

Lista de figuras

Lista de tablas

Introducción

1. ESTADÍSTICA Y PROBABILIDAD

La estadística y la probabilidad son áreas de la ciencia que tienen amplia aplicabilidad en diferentes campos del conocimiento, como la ingeniería, la economía, la biología, o la sociología, entre otras. Al igual que las matemáticas, la estadística y la probabilidad constituyen hoy uno de los ciclos fundamentales para la formación de un profesional de cualquier disciplina del conocimiento y son esenciales para la explicación y el estudio de la información de fenómenos que involucran procesos de incertidumbre y de variabilidad.

La estadística en términos generales se ocupa de la observación, recolección, organización, síntesis, análisis e interpretación de la información contenida en datos sobre poblaciones por medio del descubrimiento de patrones, tendencias y relaciones entre los datos, para planear y decidir sobre su futuro. De otra parte, la probabilidad pretende cuantificar el grado de verosimilitud de los sucesos o eventos que no podemos controlar y sus relaciones con otros (Velasco & Wisniewski, 2001). La probabilidad se cimienta en la teoría de conjuntos y proporciona la posibilidad de ocurrencia de los eventos futuros, para poder tomar decisiones acerca de estos.

La palabra estadística proviene de la palabra italiana *statista*, que se relaciona con Estado, porque el Estado por lo general se ha encargado de tomar registro y análisis acerca de sus poblaciones. La palabra estadística fue usada por primera vez por Gottfried Achenwall (1719-1772), un profesor de Marlborough (Nueva Zelanda) y desde entonces se popularizó, empero, muchos siglos anteriores a esto ya se hacían registros y análisis de los datos (Badii & Guillen, 2010). Una definición de estadística

La estadística estudia el comportamiento de los fenómenos llamados de colectivo. Está caracterizada por una información acerca de un colectivo o universo, lo que constituye su objeto material; un modo propio de razonamiento, el método estadístico, lo que constituye su objeto formal y unas previsiones de cara al futuro, lo que implica un ambiente de incertidumbre, que constituyen su objeto o causa final (Batanero, 2013) (Cabriá, 1994, p. 22, citado en Batanero, 2013, p. 1).

La probabilidad siempre ha sido relacionada con los juegos de azar y el primer acercamiento matemático a esta se hizo con el filósofo italiano Gerolamo CARDANO (1501-1576), que fue conocido por su trabajo a través de la correspondencia entablada entre los matemáticos Blaise PASCAL (1623-1662) y Pierre DE FERMAT (1601-1665) a través del problema de Chévalier de Meré (BLANCO, 2010). Sin embargo, la probabilidad en el medio académico de los siglos XVII, XVIII y XIX se trataba sólo como una disciplina empírica, aunque hubo aportes significativos de matemáticos como Jacob BERNOULLI (1654-1705), con la ley de los grandes números; Abrahan DEMOIVRE (1667-1754) y PIERRE-SIMON LAPLACE, con el teorema de DeMoivre-Laplace, y Carl Friedrich Gauss, que es conocido por representar datos experimentales por medio de una curva en forma de campana y por emplear el método de mínimos cuadrados. No existía una formalización matemática sobre los conceptos de probabilidad. Sólo en el siglo XX, los aportes de FRËCHET (1878-1973) y CARATHEODORY (1873-1950), con su trabajo sobre la teoría de la medida, labraron el camino para el desarrollo axiomático de la teoría de la probabilidad.

2. FASES DE LA ESTADÍSTICA

En algunos casos, la estadística se puede definir como una herramienta metodológica para la solución de problemas planteados en diferentes áreas de las ciencias empleando el análisis de datos. Es decir, la estadística como parte de la matemática aplicada que usa diferentes técnicas o métodos estadísticos. Estas técnicas se dividen en dos categorías o fases la estadística descriptiva y la estadística inferencial.

La estadística descriptiva sirve como herramienta para hacer representaciones de los datos mediante tablas, gráficos o números (en algunos casos). Es decir, se encarga de organizar y sintetizar la información observable ya sea de variables cualitativas o cuantitativas provenientes de poblaciones (conjunto total de elementos observables que son objeto de estudio) o muestras (un subconjunto de la población). Si se trabaja con muestras, la estadística descriptiva no permite hacer generalizaciones sobre la población de estudio el resumen e interpretación que se obtenga de la muestra sólo puede interpretarse sobre ella (BADII & GUILLEN, 2010)

Por otro lado, la estadística inferencial permite, a partir de muestras, hacer inferencias o dar conclusiones fuertes acerca de las poblaciones. O sea, permite hacer generalizaciones sobre el comportamiento de la población más allá de la muestra y todo esto con la ayuda de la probabilidad, para la toma de decisiones.

Por ejemplo, en un estudio para el lanzamiento de un nuevo producto en el mercado, como una bebida refrescante para Colombia, se pueden determinar u observar algunas variables de preferencia, como sabor, color, precio, tamaño, entre otras. Como no se puede preguntar a todos los colombianos (la población), se escoge una muestra (subconjunto de la población), y cuando se hace el cálculo de proporciones o porcentajes de dichas preferencias se estaría sólo en la fase de la estadística descriptiva, y las conclusiones que se tengan serán válidas sólo para la muestra. Pero cuando tomamos la decisión de lanzar o no este nuevo producto al mercado, porque conocemos la probabilidad de aceptación en la población, estaríamos en la fase de la estadística inferencial.

Capítulo 1. Paquetes matemáticos y estadísticos

El manejo de paquetes matemáticos y estadísticos facilitan la enseñanza-aprendizaje de las matemáticas y de la estadística, porque permiten hacer procedimientos matemáticos o estadísticos de una forma rápida y eficiente (Barahona, Barrera, Vaca, & Hidalgo, 2015).

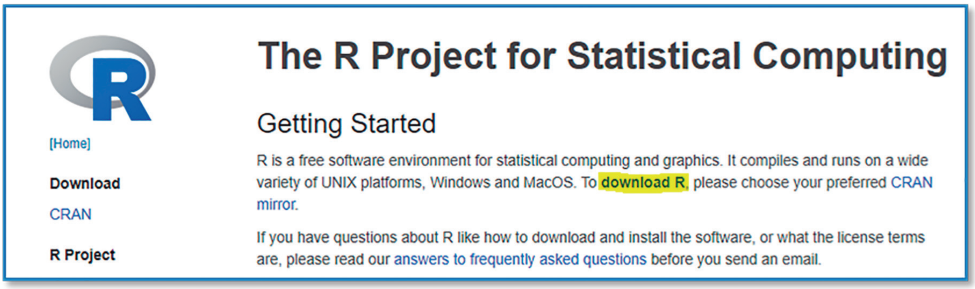
1. R-STUDIO

R es un *software* gratuito para el manejo de datos, y la elaboración cálculos y gráficas. RStudio es un programa editor para manejar R de una forma más sencilla y amigable. Para poder trabajar con RStudio debemos instalar primero el R (Varios, 2023).

A continuación, se presenta los pasos para descargar R y RStudio.

1. Dirigirse a la página principal de R que está en el siguiente enlace https//www.r-project.org/ (Figura 1).

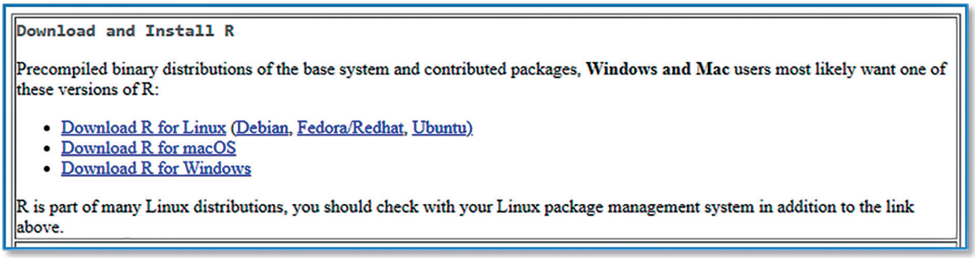
Figura 1. *Página principal de R.*



Fuente: elaboración propia

2. Al dar clic en la página principal encontramos varias opciones, entre ellas descargar (Figura 2).

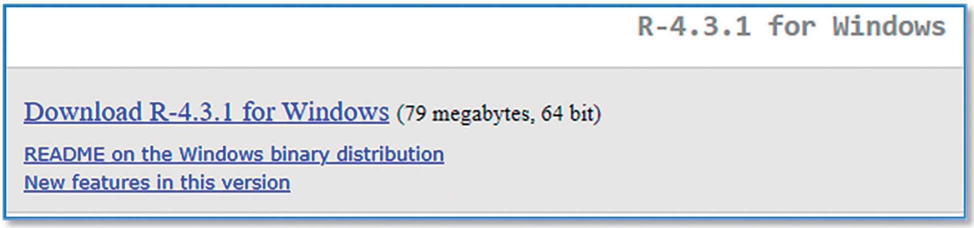
Figura 2. *Opción de descarga de R.*



Fuente: elaboración propia.

3. Buscar el país de origen y luego seleccionar el tipo de sistema operativo del computador para la descarga, teniendo en cuenta que el computador es de 32 o 64 bits (Figura 3).

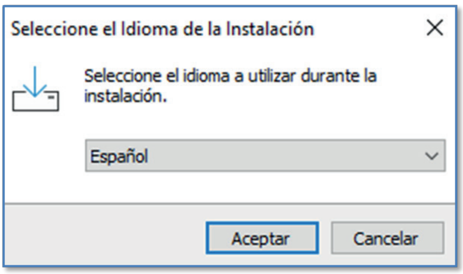
Figura 3. *Opción del sistema operativo y del número de bits.*



Fuente: elaboración propia.

4. Abrir el instalador y dar clic en ejecutar. Luego, seleccione el idioma de preferencia (Figura 4).

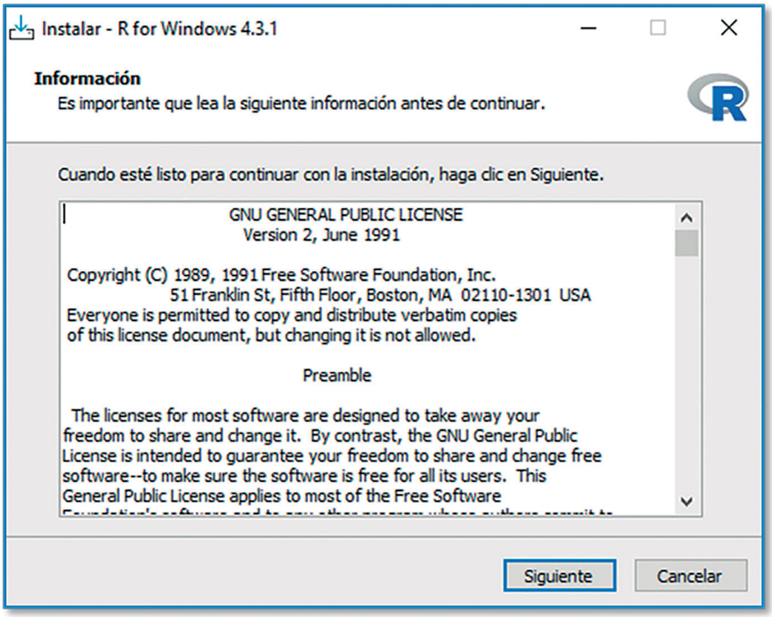
Figura 4. *Opción de idioma.*



Fuente: elaboración propia

5. Aceptar términos y condiciones. Dar clic en siguiente varias veces hasta llegar a la ventana de carga, y finalizar (Figura 5).

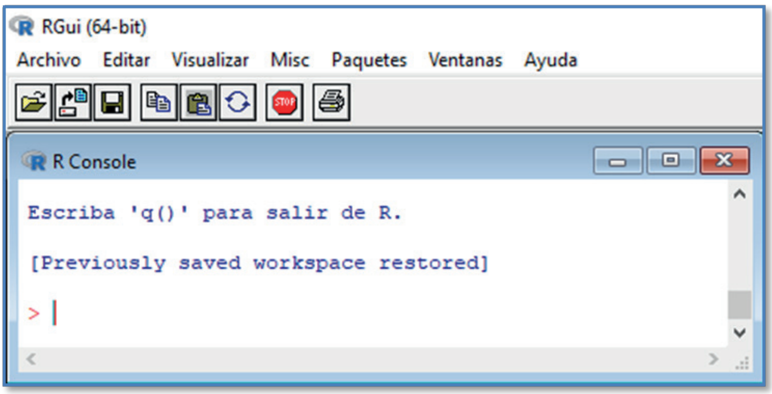
Figura 5. *Términos y condiciones.*



Fuente: elaboración propia.

6. Buscar la aplicación de R e ingresar (Figura 6).

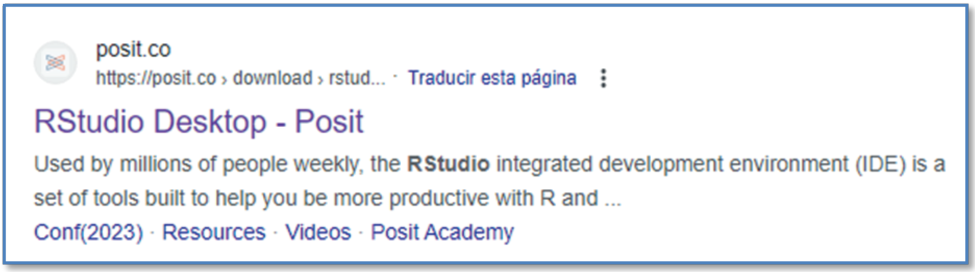
Figura 6. *Interfaz de R.*



Fuente: elaboración propia.

7. Ahora, dirigirse a la página de descarga para RStudio, que está en el siguiente enlace https//posit.co/download/rstudio-desktop/#download (Figura 7).

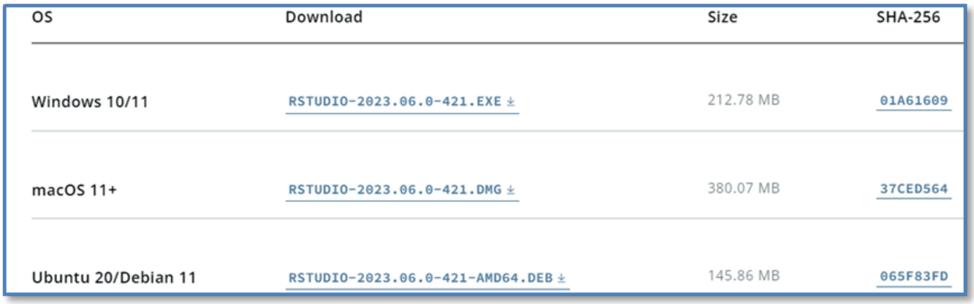
Figura 7. *Página web de descarga de RStudio.*



Fuente: elaboración propia.

8. Seleccionar el sistema operativo donde se instalará RStudio (Figura 8).

Figura 8. *Opciones de descarga de RStudio según sistema operativo.*

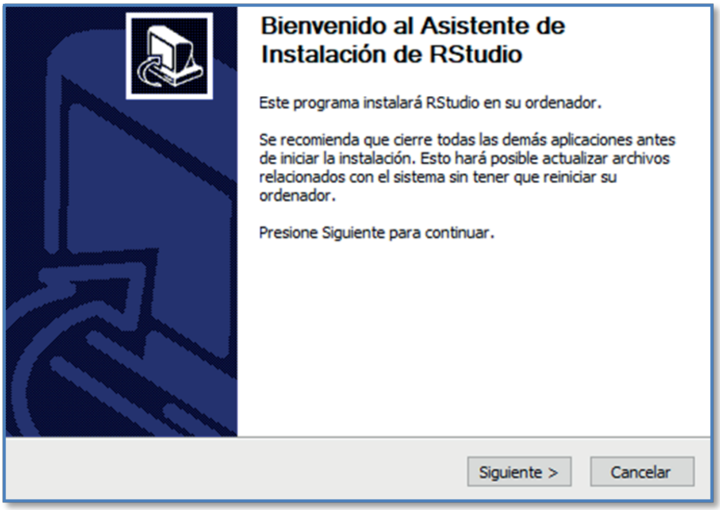


OS	Download	Size	SHA-256
Windows 10/11	RSTUDIO-2023.06.0-421.EXE	212.78 MB	01A61609
macOS 11+	RSTUDIO-2023.06.0-421.DMG	380.07 MB	37CED564
Ubuntu 20/Debian 11	RSTUDIO-2023.06.0-421-AMD64.DEB	145.86 MB	065F83FD

Fuente: elaboración propia.

9. Abrir el instalador, dar clic en ejecutar y luego instalar, dando le clic en siguiente (Figura 9).

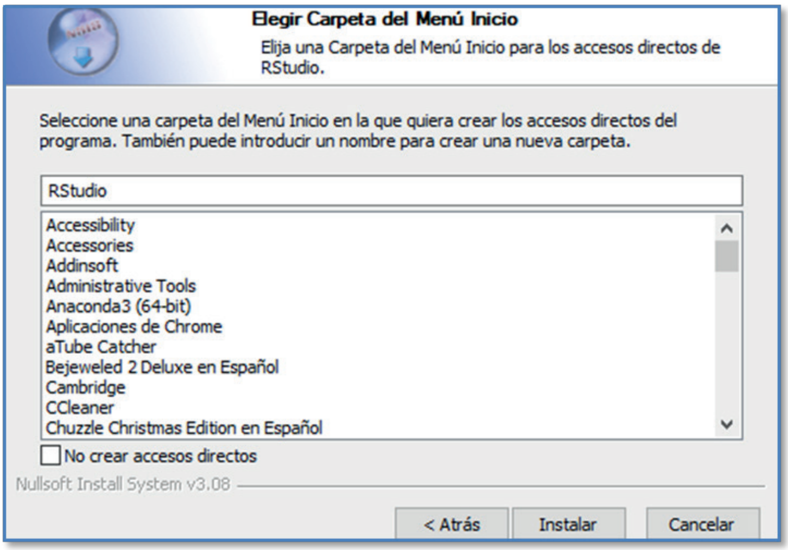
Figura 9. *Ventana de RStudio para la instalación.*



Fuente: elaboración propia.

10. Dar clic en siguiente hasta que se llegue a la opción instalar (Figura 10).

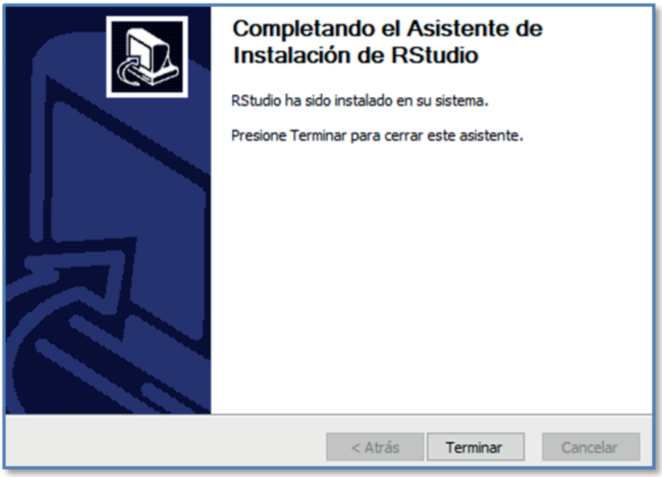
Figura 10. *Ventana de instalación de RStudio.*



Fuente: elaboración propia.

11. Por último, se le da clic en la opción terminar. (Figura 11)

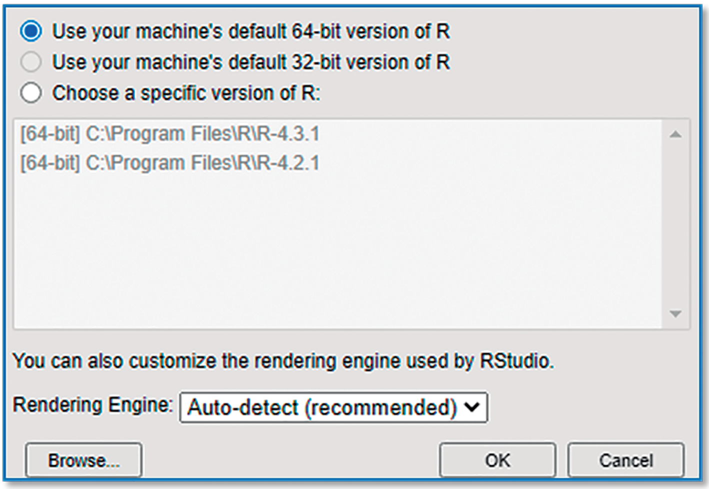
Figura 11. *Ventana de terminar instalación de RStudio.*



Fuente: elaboración propia

12. Buscar la aplicación RStudio, dar clic en la versión y capacidad del computador (Figura 12).

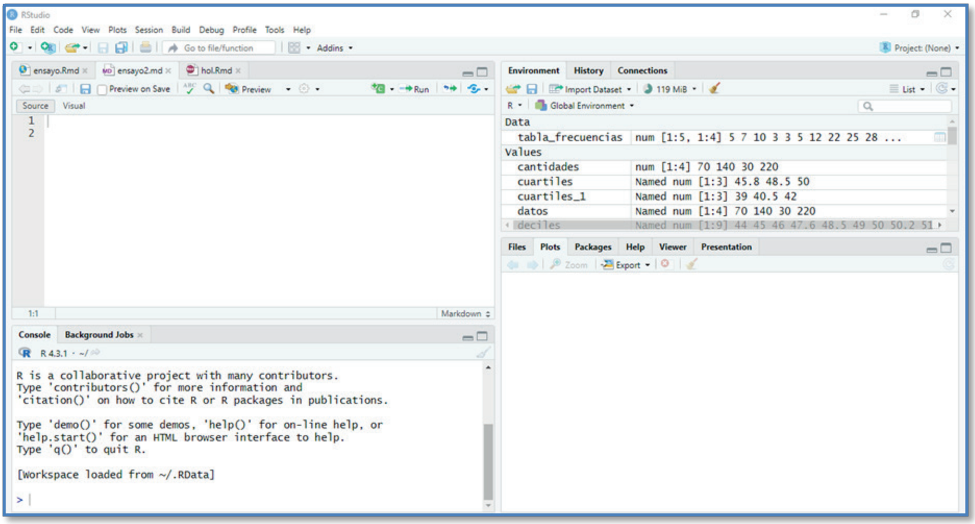
Figura 12. *Ventana de la opción de la versión y capacidad de la computadora para RStudio.*



Fuente: elaboración propia

13. Ventana de interfaz de RStudio (Figura 13).

Figura 13. *Ventana de interfaz de RStudio.*



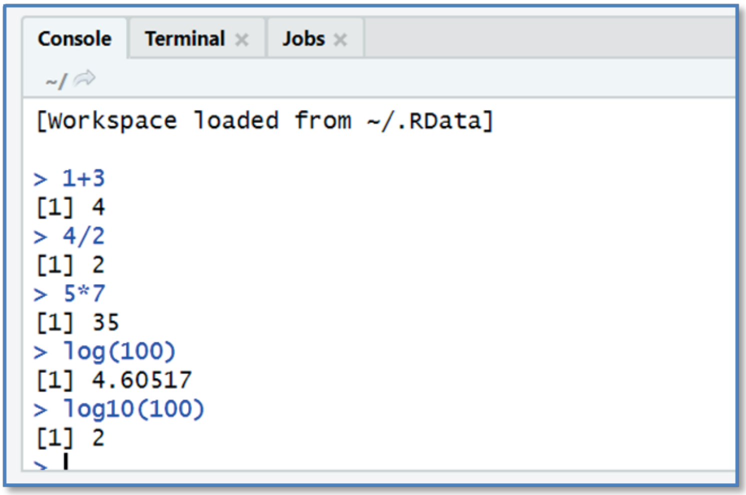
Fuente: elaboración propia.

2. LA ARQUITECTURA O INTERFAZ DE RSTUDIO

En la interfaz de RStudio podemos encontrar en la parte superior un menú de herramientas, como File, Edit, Code, View, entre otras, y cuatro paneles la parte superior izquierda corresponde al editor, la parte inferior izquierda corresponde a la consola, la parte superior derecha corresponde al entorno de variables y la parte inferior derecha corresponde al entorno de utilidades (Figura 13).

En la parte de la consola podemos hacer cálculos u operaciones matemáticas sencillas, como sumas, restas, multiplicaciones o cálculo de funciones, como el logaritmo natural y en base diez (Figura 14).

Figura 14. *Consola de RStudio. Cálculo y operaciones matemáticas.*



Console Terminal Jobs

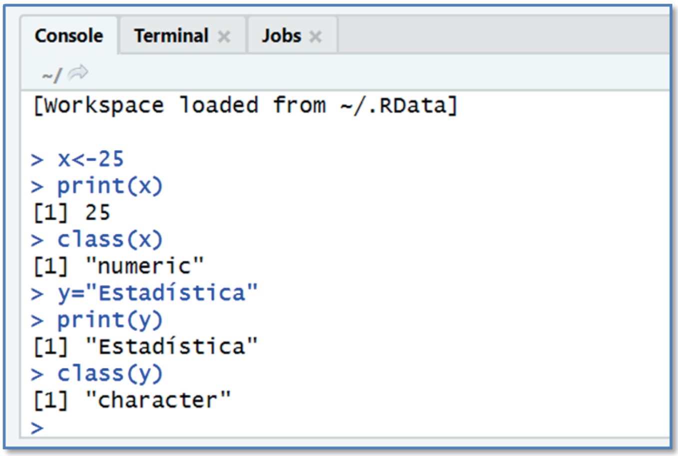
~/

```
[Workspace loaded from ~/.RData]

> 1+3
[1] 4
> 4/2
[1] 2
> 5*7
[1] 35
> log(100)
[1] 4.60517
> log10(100)
[1] 2
>
```

Fuente: elaboración propia.

Una variable en RStudio es un espacio en el que guardamos un objeto en la computadora. Las variables más sencillas pueden ser numéricas o de carácter. El símbolo **<-** o el símbolo = es un operador de asignación de las variables (Figura 15). En este caso, a la variable **x** le asignamos el número **25** y a la variable **y** le asignamos la palabra **"Estadística"** (que debe ir entre comillas por ser de carácter, o también con una sola comilla).

Figura 15. *Asignación de nombres de las variables en RStudio.*



Console Terminal Jobs

~/

```
[Workspace loaded from ~/.RData]

> x<-25
> print(x)
[1] 25
> class(x)
[1] "numeric"
> y="Estadística"
> print(y)
[1] "Estadística"
> class(y)
[1] "character"
>
```

Fuente: elaboración propia.

Las variables también pueden ser vectores donde se almacenan grupos de números o caracteres. El símbolo **c(, , ,)**, es para escribir un vector (Figura 16).

Figura 16. *Vectores numéricos y de carácter en RStudio.*

```
Console   Terminal ×   Jobs ×
~/
[Workspace loaded from ~/.RData]

> x<-c(1,2,3,4,5,6,7,8,9,10)
> print(x)
 [1]  1  2  3  4  5  6  7  8  9 10
> class(x)
[1] "numeric"
> y=c("Amarillo","azul","rojo")
> print(y)
[1] "Amarillo" "azul"     "rojo"
> class(y)
[1] "character"
>
```

Fuente: elaboración propia.

En general, se pueden hacer operaciones con vectores, como se muestra en la Figura 17.

Figura 17. *Funciones para operar vectores numéricos con RStudio.*

```
Console   Terminal ×   Jobs ×
~/
[Workspace loaded from ~/.RData]

> x<-c(1,2,3,4,5,6,7,8,9,10)
> x/2
 [1] 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
> sum(x)
[1] 55
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    3.25    5.50    5.50    7.75   10.00
> mean(x)
[1] 5.5
> |
```

Fuente: elaboración propia

A continuación, se presenta una tabla (Tabla 1) con operadores básicos de programación para RStudio.

Tabla 1. *Operadores básicos en RStudio*

Operación	Descripción
+	Adición
-	Sustracción
*	Multiplicación
/	División
sqrt()	Raíz cuadrada
log()	Logaritmo Natural
log(n,m)	Logaritmo de n de base m
^	Potenciación
.	Punto decimal
exp()	e^x

3. GEOGEBRA

GeoGebra es un paquete matemático dinámico y gratuito, que se encuentra en la red (Figura 18). Este permite la adaptación de sus funciones a los usuarios. GeoGebra es una herramienta tecnológica que facilita el análisis matemático en diferentes niveles educativos, en áreas como geometría, álgebra, cálculo y estadística, entre otras (Barahona, Barrera, Vaca, & Hidalgo, 2015).

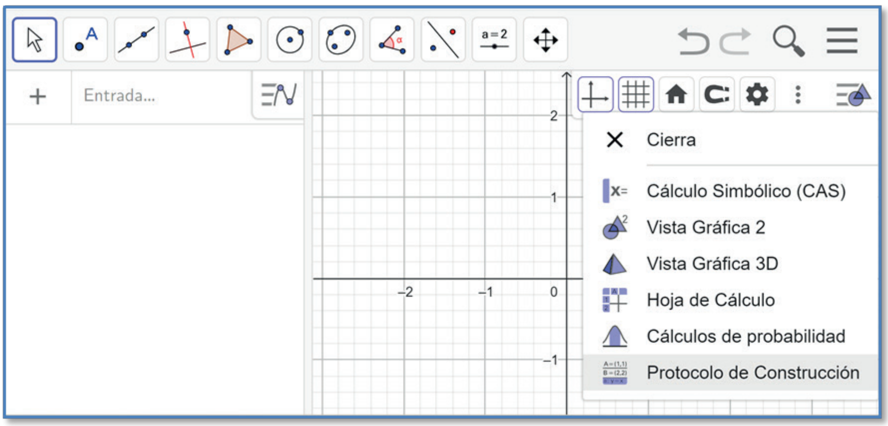
Figura 18. *Búsqueda en Google de la página web de GeoGebra Clásico.*

Fuente: elaboración propia.

4. LA ARQUITECTURA O INTERFAZ DE GEOGEBRA CLÁSICO

GeoGebra Clásico es un paquete integrado que en su interfaz muestra, en su parte superior, una barra de herramientas; un panel de entrada o ventana algebraica, en su parte izquierda; un panel en el centro, para la representación gráfica, y un menú de opciones, en su parte derecha, donde se encuentran, por ejemplo, las opciones de hoja de cálculo y de probabilidad (Figura 19).

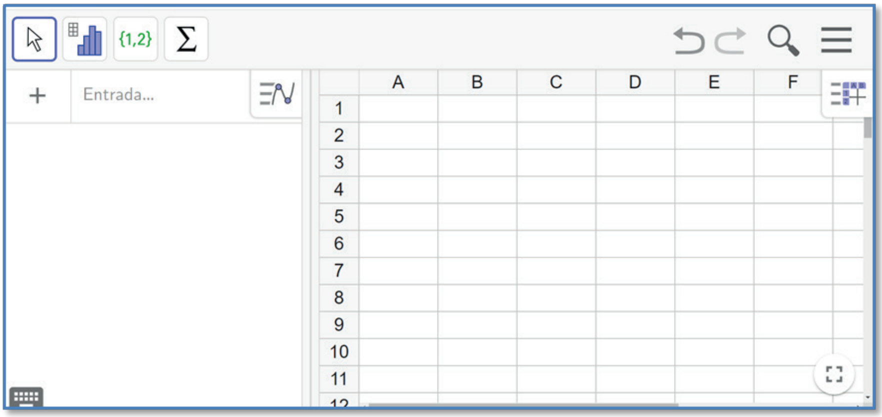
Figura 19. *Interfaz de GeoGebra Clásico.*



Fuente: elaboración propia

Si desplegamos la opción Hoja de Cálculo en GeoGebra nos muestra, en la parte derecha, una matriz donde podemos ingresar los datos (Figura 20).

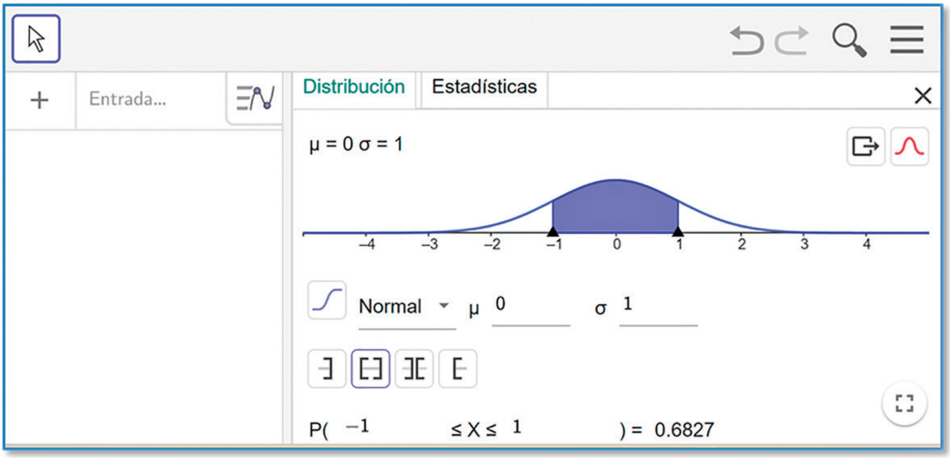
Figura 20. *Hoja de cálculo de GeoGebra.*



Fuente: elaboración propia.

Por otro lado, si desplegamos la opción probabilidad en GeoGebra, muestra en la parte derecha un panel con dos opciones, Distribución y Estadísticas. La primera muestra por defecto el gráfico de la distribución de probabilidad normal estándar (Figura 21).

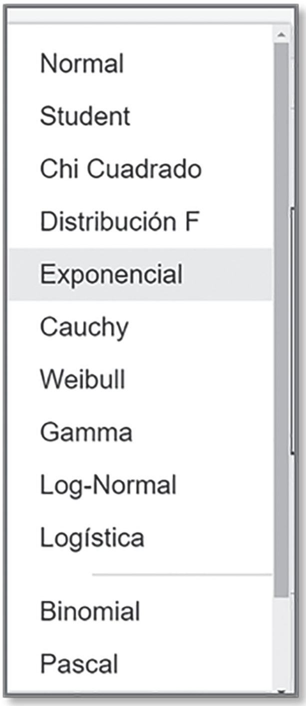
Figura 21. *Opción probabilidad. Distribuciones de probabilidad en GeoGebra.*



Fuente: elaboración propia.

En el menú de distribución Normal podemos desplegar y en la parte superior aparecen las opciones de distribuciones continuas de probabilidad, como Normal, t-Student, Ji-cuadrado, Distribución F, Exponencial, entre otras, y en la parte inferior las opciones de distribuciones discretas de probabilidad Binomial, Pascal, Poisson e Hipergeométrica (Figura 22). Todo este menú permite hacer el cálculo de probabilidades de todas las anteriores distribuciones de probabilidad.

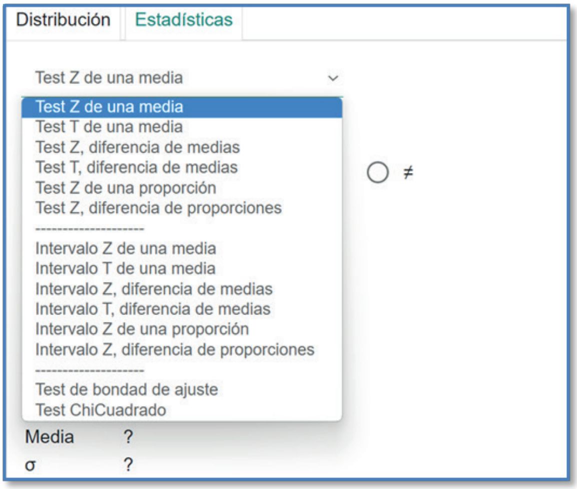
Figura 22. *Menú de distribuciones continuas y discretas de probabilidad en GeoGebra.*



Fuente: elaboración propia.

En la opción de Estadísticas se muestran las elecciones para el cálculo de algunas pruebas de hipótesis y algunos intervalos de confianza (Figura 23).

Figura 23. *Menú estadísticas con pruebas de hipótesis e intervalos de confianza.*



Fuente: elaboración propia.

5. EJERCICIOS

Ejercicio 1. Dados los siguientes datos que representan las edades de un grupo de personas

18	42	54	25	20	27	46	33	25	45	37
20	27	46	33	25	28	48	40	53	27	40

a) Ingresar datos como la variable edad en RStudio.

b) Aplicar las siguientes funciones en RStudio length(edad), max(edad), min(edad), sum(edad), summary(edad), prod(edad), mean(edad), sort(edad) y rev(edad).

Ejercicio 2. Dados los siguientes datos que representan el color preferido de un grupo de personas

Azul	Verde	Amarillo	Azul	Rojo	Dorado
Rojo	Verde	Azul	Verde	Azul	Negro

a) Ingresar datos como la variable color en RStudio.

b) Aplicar las siguientes funciones en RStudio table(color), plot(table(color)), prop.table(table(color)) y prop.table(table(color))*100.

Ejercicio 3. Asigne a la variable x el vector de números de 1 a 20, de la siguiente forma x=120. Realice las siguientes funciones

a) print(x).

b) x/2

c) sum(x)

d) sum(1/(x+1))

Ejercicio 4. Asigne a la variable x el vector de números de 1 a 10, de la siguiente forma x=110. Haga las siguientes funciones

a) x[3]; # Elemento tercero del vector.

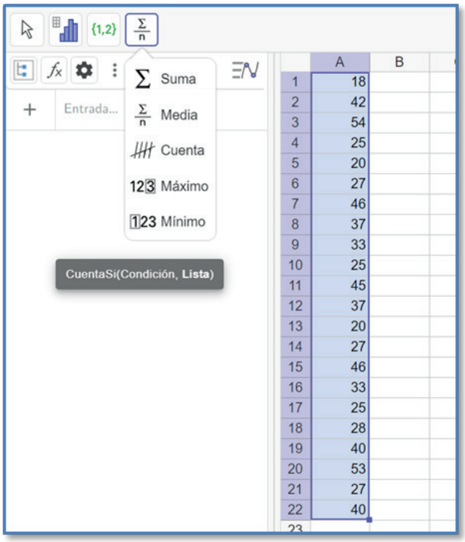
b) x[3:8]; # Elementos del tercero al octavo.

c) x[x>3] # Elementos mayores que 3.

d) x[x>3 & x<=8]; # Elementos mayores que 3 y menores o iguales a 8.

Ejercicio 5. Ingresar los datos del **Ejercicio 1** en GeoGebra en la Hoja de Cálculo (Figura 24) y calcular:

Figura 24. *Hoja de cálculo de GeoGebra con datos del Ejercicio 1*



Fuente: elaboración propia.

a) Suma
b) Media
c) Mínimo
d) Máximo

Capítulo 2. Conceptos básicos de la estdadística descriptiva

1. POBLACIÓN Y MUESTRA

La población o universo es el objeto de estudio. Se puede definir de manera informal como el conjunto total de elementos (Newbold, Carlson, & Thorne, 2010) que tienen una o más características o rasgos comunes posibles de observar y sobre el cual se busca dar conclusiones y tomar decisiones. La población puede ser datos, personas, cosas, acontecimientos y atributos, entre otros.

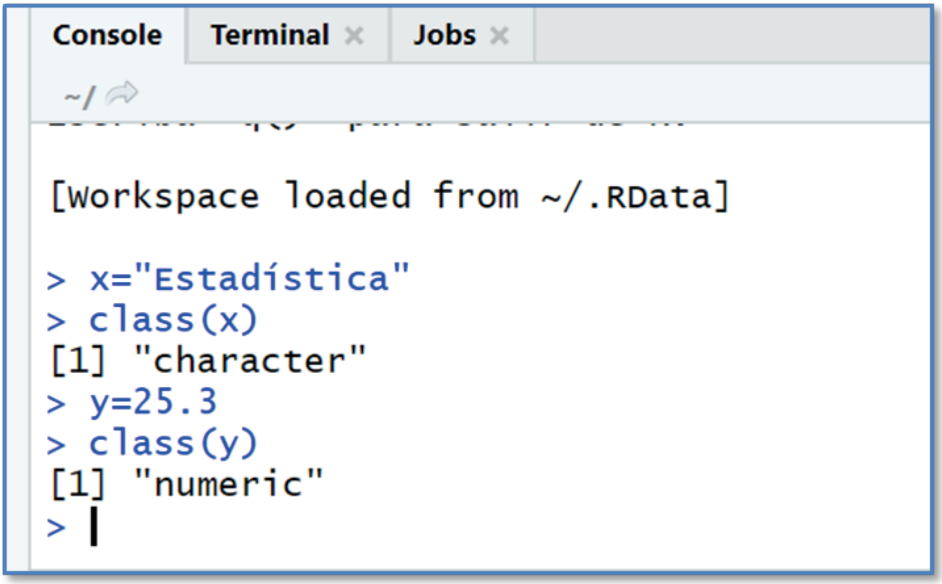
En algunos casos no es posible contar con toda la población u objeto de estudio, esto debido a que el costo, el tiempo o demás recursos que conlleva obtener todos los posibles elementos de la población puede ser muy grande (Devore, 2005). Por ejemplo, si un médico desea hacer un estudio a unos pacientes con cierta enfermedad, como un determinado cáncer, no es posible que él pueda contar para su estudio con todos los pacientes que sufren de esa enfermedad y para esto él debe tomar una *muestra*, es decir, seleccionar un subconjunto de elementos de la población y que de alguna manera esa muestra sea representativa y significativa de la población, para poder inferir acerca de sus propiedades o características. Los procesos en los que se seleccionan muestras se conocen como métodos de muestreo (Pérez, 2005) que en este libro no vamos a tratar.

Por otro lado, la medida de una característica en la población se conoce como *parámetro* (media poblacional, proporciones poblacionales y varianza poblacional, entre otros) y la medida de una característica de la muestra, como *estadístico* (media muestral, proporciones muestrales y varianza muestral, entre otros).

2. VARIABLE ESTADÍSTICA

De manera informal, una variable estadística se considera como una característica o medida que es objeto de estudio. Ejemplos de variables estadísticas son el género, el peso, la estatura, el estado civil, en un conjunto de personas. Existen dos tipos de variables en estadística las *variables cualitativas o categóricas* y las *variables cuantitativas.* Las primeras toman diferentes formas de descripción y se dice que éstas son variables de clasificación mas no de medición; las segundas, las cuantitativas, toman diferentes formas de medición y para esto usan escalas numéricas. Ejemplos de ***variables cualitativas*** puede ser género, estado civil, estrato socioeconómico, lugar de nacimiento y color preferido, entre otras. Ejemplos de ***variables cuantitativas*** puede ser el peso, la estatura, el número de hermanos o ingreso mensual de una persona. En RStudio podemos distinguir las variables cualitativas, porque se reconocen como *caracteres*, mientras que las cuantitativas se reconocen como *numéricas* (Figura 25).

Figura 25. *Tipos de variables cualitativas y cuantitativas en RStudio.*



Fuente: elaboración propia.

Entre las *variables cuantitativas* podemos encontrar las *variables discretas* y las *variables continuas*. Las primeras asumen valores numéricos finitos o infinitos numerables y sus resultados se obtienen del proceso de conteo y están asociadas al conjunto de los números naturales; las segundas asumen cualquier valor numérico en una escala continua y están asociadas a la recta numérica real. Ejemplos de *variables discretas,* cuando preguntamos en una encuesta número de hermanos, número de hijos, etc. (en inglés cuando preguntamos *How many…*). Ejemplos de *variables continuas* edad (tiempo), estatura (longitud), peso, ingresos mensuales, etc. (en inglés cuando preguntamos *How much…*)

3. ESCALAS DE MEDICIÓN

Las escalas de medición permiten hacer clasificaciones de las variables mediante la cantidad de información que contiene cada una de ellas según su descripción (Anderson, Sweeney, & Wlliams, 2009). Cuando se mide, se hacen comparaciones de las magnitudes referidas a un patrón, que en algunos casos tienen unidades de medida, ceros relativos o ceros absolutos. Las escalas de medición se pueden definir de acuerdo con sus patrones de medida de la siguiente manera

Escala nominal. Una variable estadística se puede clasificar en escala nominal porque describe simplemente rótulos, nombres o categorías que permiten identificar un atributo del elemento de la muestra. En este caso, si la variable toma valores numéricos, este número no indica medición u orden con sentido, porque en esta escala las variables carecen de unidad de medida y de orden. Un ejemplo de este tipo de variable es color de preferencia. Aquí las observaciones pueden ser amarillo, azul, rojo, etc. Estas observaciones son simplemente nombres sin orden.

Escala ordinal. Una variable estadística se puede clasificar en escala ordinal cuando la variable tiene propiedades de datos nominales, es decir, rótulos, nombres o categorías, que además permiten el orden entre ellas. En este caso, la variable ordinal permite jerarquizar las observaciones, pero no cuenta con una unidad de medida, es decir, no se puede establecer o estimar la "distancia" o intervalo entre las observaciones. Un ejemplo de este tipo de variable es el nivel educativo de una persona Primaria, Básica, Secundaria, Universitaria, etc. Estas observaciones son rótulos con orden, pero no hay una unidad de medida que identifique el intervalo o "distancia" entre estas observaciones o la diferencia entre ellas.

Escala de intervalo. Una variable estadística se puede clasificar en escala de intervalo cuando la variable permite tener orden entre sus observaciones y además cuenta con una unidad fija de medida, es decir, las observaciones permiten obtener el intervalo o "distancia" entre ellas. En esta escala, el cero es arbitrario o relativo y en este caso entre las observaciones no se pueden establecer razones. Un ejemplo de este tipo de variable es el sistema horario, porque en este caso hay orden (800 a.m., 900 a.m. …) una unidad de medida (horas, minutos, segundos…), además establecer el intervalo o "distancia" entre estas observaciones o la diferencia entre ellas (entre las 800 a.m. y las 1000 a.m. han trascurrido 2 horas), pero no tiene un cero absoluto, es decir, el cero es relativo (las 0 horas no significan ausencia de tiempo) y no puedo establecer razones (1200 m / 600 a.m. no tiene sentido).

Escala de razón. Una variable estadística se puede clasificar en escala de razón cuando la variable tiene propiedades de intervalo y existe un cero real o absoluto que permita considerar cocientes de mediciones o razones. Un ejemplo de este tipo de variable es la distancia de una ciudad a otra. Aquí se observa que la variable tiene orden (10 km, 20 km…), unidad de medida (km, m…), cero absoluto (distancia 0 significa ausencia de distancia) y se pueden establecer razones (20Km/10km=2; lo que significa que 20 km es el doble de distancia que 10 km).

*Aquí se va a hacer un paréntesis para explicar la diferencia entre una fracción y una razón las fracciones comparan entre sí objetos homogéneos, ¼ significa una parte de cuatro partes, es decir objetos de una misma unidad, mientras que las razones comparan entre sí objetos que pueden ser heterogéneos, es decir, objetos que se miden con unidades diferentes. Una fracción es una parte de una unidad, y esta parte puede ser mayor o menor que la unidad y siempre está representada por un número racional. Por el contrario, la razón compara dos cantidades entre sí, ambas constituidas, en general, por números reales y establecen relaciones entre magnitudes que pueden tener diferentes unidades de medida (Batanero, Cid, & Godino, 2003). Ejemplos de fracciones ¼ de manzana, 2/3 de la población, 5/4 de kilogramos de carne, entre otros. Las razones no siempre son números racionales. Por ejemplo, se pueden establecer razones entre los lados de un triángulo y sus longitudes no necesariamente son números enteros.

4. EJERCICIOS

Ejercicio 1. Clasifique cada variable como cuantitativa o cualitativa

a) Tiempo en que tarda ensamblar un rompecabezas.
b) Calificación de un político recién elegido (Excelente, Bueno, Regular, Deficiente).
c) Edad de una persona.
d) El número de llamadas recibidas en una oficina en una semana.
e) El valor del salario mensual de una persona.
f) El número de zapatos vendidos durante una semana en una tienda.
g) La talla de los pantalones de una persona.

Ejercicio 2. Clasifique las siguientes variables como discretas o continuas

a) Producción en kilogramos de trigo en una parcela.
b) El valor de las deudas de una empresa durante un mes.
c) El número de trabajadores que tiene una compañía.

d) El número de puntos obtenidos por un equipo de baloncesto en un partido.
e) El tiempo de espera de un automóvil en un semáforo en rojo.

Ejercicio 3. Clasifique las siguientes variables como cualitativas, discretas o continuas, y además su escala de medición.

a) La edad de una persona.
b) Ubicación geográfica (latitud y longitud).
c) Tipo de almacén en un centro comercial según la característica (ropa, zapatos, etc.).
d) Temperatura de una persona.
e) Altura de una persona.
f) Velocidad de un avión en vuelo.
g) Número telefónico de una persona.
h) Ingreso mensual de una persona.
i) Números de hijos que tiene una persona.
j) Año de nacimiento de una persona.
k) Número de comparendos que hace un policía de tránsito durante un mes.
l) La dureza de los minerales (Escala de Mohs).
m) El estado civil de una persona.
n) El valor de las deudas de una persona en un año fiscal.
o) El número amigos que tiene una persona en la red social como Facebook.
p) La distancia en kilómetros de una ciudad a otra.
q) El rango militar.

5. DATOS CUALITATIVOS

Los datos cualitativos se pueden representar por medio de tablas o gráficos, como el diagrama de barras o el de torta/pastel.

5.1. Tablas de frecuencia o tabla resumen

Para construir una tabla de frecuencias o tabla de resumen para datos cualitativos, se deben tener en cuenta las categorías o características comunes de sus elementos y, en cada una de ellas, hacer el conteo de sus respectivos elementos; este conteo va a representar lo que se conoce como frecuencia absoluta.

Ejemplo 2.1: a continuación, se presenta la variable lugar de nacimiento de un grupo de 40 estudiantes de Estadística I

Bogotá	Medellín	Bogotá	Bogotá	Bogotá	Medellín
Bogotá	Cali	Bogotá	Cali	Medellín	Girardot
Villavicencio	Medellín	Bogotá	Bogotá	Bogotá	Bogotá
Bogotá	Cali	Cali	Bogotá	Girardot	Pereira
Bogotá	Bogotá	Pereira	Bogotá	Pereira	Cali
Medellín	Medellín	Bogotá	Bogotá	Bogotá	Pereira
Bogotá	Armenia	Bogotá	Bogotá		

La Tabla 2 muestra la distribución de las frecuencias absolutas para la variable lugar de nacimiento.

Tabla 2. *Distribuciones de frecuencia de la variable ciudad de nacimiento*

Ciudad	Frecuencia absoluta
Armenia	1
Bogotá	21
Cali	5
Girardot	2
Medellín	6
Pereira	4
Villavicencio	1
Total	40

En RStudio podemos obtener la tabla de frecuencias como se muestra en la Figura 26.

Figura 26. *Tabla de frecuencias absolutas en RStudio.*

```
> x=c("Bogotá", "Medellín", "Bogotá", "Bogotá", "Bogotá", "Medellí
n", "Bogotá", "Cali", "Bogotá", "Cali", "Medellín", "Girardot", "Vil
lavicencio", "Medellín", "Bogotá", "Bogotá", "Bogotá", "Bogotá", "Bo
gotá", "Cali", "Cali", "Bogotá", "Girardot", "Pereira", "Bogotá", "B
ogotá", "Pereira", "Bogotá", "Pereira", "Cali", "Medellín", "Medellí
n", "Bogotá", "Bogotá","Bogotá", "Pereira", "Bogotá", "Armenia", "Bo
gotá", "Bogotá")
> length(x)
[1] 40
> table(x)
x
     Armenia        Bogotá          Cali      Girardot 
           1            21             5             2 
    Medellín       Pereira Villavicencio 
           6             4             1 
> |
```

Fuente: elaboración propia.

De acuerdo con la Figura 26 es importante decir que este tipo de datos son cualitativos, lo que implica que van en comillas o una sola comilla.

Ahora, para hallar la frecuencia absoluta se usará la función **table()**. Esta función genera tablas de frecuencias de un vector, en este caso de las ciudades de nacimientos que se tiene como nombre **x**, el nombre de la tabla en este caso será **table(x)**. Para ver la tabla también se usa la función **print()**, que imprime los resultados de una variable, en este caso, como se nombró la tabla.

Ahora, si se desea, se puede presentar la proporción o frecuencia relativa (división entre el número de elementos que están dentro de la categoría y el número total de elementos) o el porcentaje (la proporción multiplicada por 100), como se muestra en la Tabla 3.

Tabla 3. *Tabla de frecuencias, proporción y porcentaje*

Ciudad	Frecuencia absoluta	Proporción	Porcentaje %
Armenia	1	0,025	2,5
Bogotá	21	0,525	52,5
Cali	5	0,125	12,5
Girardot	2	0,05	5
Medellín	6	0,15	15

Pereira	4	0,1	10
Villavicencio	1	0,025	2,5
Total	40	1	100

En RStudio podemos obtener la tabla de frecuencias, proporciones y porcentajes, como se muestra en la Figura 27.

Figura 27. *Tabla de frecuencias, Proporción y Porcentaje para la variable ciudad de nacimiento en RStudio.*

```
> x=c("Bogotá", "Medellín", "Bogotá", "Bogotá", "Bogotá", "Medellín", "Bogotá",
 "Cali", "Bogotá", "Cali", "Medellín", "Girardot", "Villavicencio", "Medellín",
 "Bogotá", "Bogotá", "Bogotá", "Bogotá", "Bogotá", "Cali", "Cali", "Bogotá", "Gi
rardot", "Pereira", "Bogotá", "Bogotá", "Pereira", "Bogotá", "Pereira", "Cali",
 "Medellín", "Medellín", "Bogotá", "Bogotá","Bogotá", "Pereira", "Bogotá", "Arme
nia", "Bogotá", "Bogotá")
> table(x)
x
      Armenia        Bogotá          Cali      Girardot 
            1            21             5             2 
     Medellín       Pereira Villavicencio 
            6             4             1 
> frecuencia_absoluta=table(x)
> Proporcion=prop.table(table(x))
> Porcentaje=prop.table(table(x))*100
> tabla_frecuencias=cbind(frecuencia_absoluta, Proporcion, Porcentaje)
> colnames(tabla_frecuencias)=c("Frecuencia", "Proporción", "Porcentaje")
> print(tabla_frecuencias)
              Frecuencia Proporción Porcentaje
Armenia                1      0.025        2.5
Bogotá                21      0.525       52.5
Cali                   5      0.125       12.5
Girardot               2      0.050        5.0
Medellín               6      0.150       15.0
Pereira                4      0.100       10.0
Villavicencio          1      0.025        2.5
> |
```

Fuente: elaboración propia.

Según la Figura 27, para hacer la frecuencia relativa se usa la función **prop.table()**, que genera la tabla de la frecuencia relativa. Para hallar los porcentajes se toma la tabla de la frecuencia relativa y se multiplica por 100. Ahora, para construir la tabla completa de frecuencias absolutas, relativas y porcentaje, se usará la función **cbind()**, que concatena o combina los vectores y resultados de otras variables. Para darle los respectivos nombres a cada columna se usa la función **colnames()**. Y finalmente, para ver el resultado, se usa la función **print()**.

5.2. Diagramas circulares

El diagrama circular o de torta es un gráfico circular que muestra la proporción o porcentaje de elementos de cada categoría de una variable cualitativa.

Del **Ejemplo 2.1:** podemos hacer un diagrama circular o de pastel en RStudio, de la siguiente manera, como lo muestra la Figura 28.

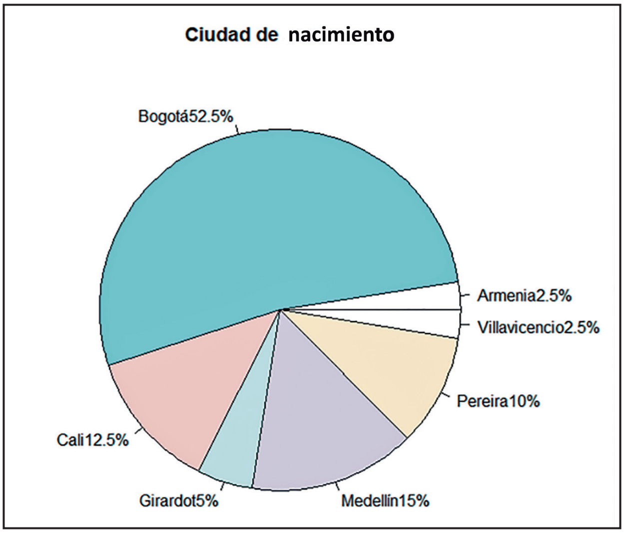
Figura 28. *Procedimiento para hacer diagrama de Pastel en RStudio.*

```
> x=c("Bogotá", "Medellín", "Bogotá", "Bogotá", "Bogotá", "Medellí
n", "Bogotá", "Cali", "Bogotá", "Cali", "Medellín", "Girardot", "Vil
lavicencio", "Medellín", "Bogotá", "Bogotá", "Bogotá", "Bogotá", "Bo
gotá", "Cali", "Cali", "Bogotá", "Girardot", "Pereira", "Bogotá", "B
ogotá", "Pereira", "Bogotá", "Pereira", "Cali", "Medellín", "Medellí
n", "Bogotá", "Bogotá","Bogotá", "Pereira", "Bogotá", "Armenia", "Bo
gotá", "Bogotá")
> table(x)
x
      Armenia         Bogotá          Cali      Girardot 
            1             21             5             2 
     Medellín        Pereira Villavicencio 
            6              4             1 
> Porcentaje=prop.table(table(x))*100
> etiquetas=c("Armenia", "Bogotá", "Cali", "Girardot", "Medellín",
 "Pereira", "Villavicencio")
> pie(Porcentaje, labels = paste0(etiquetas, "", Porcentaje, "%"), m
ain = "Ciudad de naimiento", radius = 1.0, edges = 200)
> 
```

Fuente: elaboración propia.

La Figura 28 muestra que para hacer el diagrama circular se usará la función **pie()**. Se tomará el vector porcentaje como argumento principal, que contiene los tamaños o porcentajes de las diferentes partes del pastel. Además, se usará el argumento **labels=**, que asigna una etiqueta a cada porción del pastel. Aquí se incluye la aplicación **paste()** para concatenar los vectores **etiquetas** y **porcentaje** y convertirlos en caracteres. También el argumento **main=** agregará el título al gráfico circular, en este caso, el título será 'Ciudad de Nacimiento'. Por otro lado, si se desea, se usan los argumentos **radius=** y **edges=**, respectivamente, para el tamaño del radio del pastel y para el contorno de un polígono con esa cantidad de aristas. Y su gráfico será como se observa en la Figura 29.

Figura 29. *Diagrama de Pastel de la variable Ciudad de nacimiento.*



Fuente: elaboración propia.

5.3. Diagrama de barras

El diagrama de barras es un gráfico bidimensional que muestra las categorías de una variable cualitativa en un eje y sus frecuencias absolutas o relativas en el otro, usando barras rectangulares cuya altura es proporcional a la frecuencia de cada categoría. Este gráfico facilita la comparación visual entre las categorías representadas.

Del **Ejemplo 2.1:** podemos hacer un diagrama de barras en RStudio como se muestra en la Figura 30.

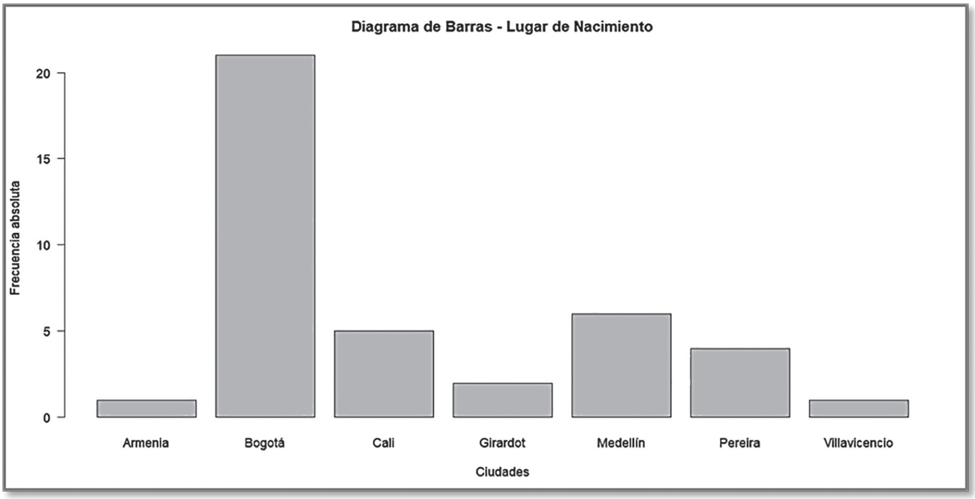
Figura 30. *Procedimiento para hacer un Diagrama de barras en RStudio.*

```
> x=c("Bogotá", "Medellín", "Bogotá", "Bogotá", "Bogotá", "Medellí
n", "Bogotá", "Cali", "Bogotá", "Cali", "Medellín", "Girardot", "Vil
lavicencio", "Medellín", "Bogotá", "Bogotá", "Bogotá", "Bogotá", "Bo
gotá", "Cali", "Cali", "Bogotá", "Girardot", "Pereira", "Bogotá", "B
ogotá", "Pereira", "Bogotá", "Pereira", "Cali", "Medellín", "Medellí
n", "Bogotá", "Bogotá","Bogotá", "Pereira", "Bogotá", "Armenia", "Bo
gotá", "Bogotá")
> barplot(table(x), xlab = "Ciudades", ylab = "Frecuencia absoluta",
 las= 1, main = "Diagrama de Barras - Lugar de Nacimiento")
>
```

Fuente: *elaboración propia.*

Para crear el diagrama de barras, se usará la función **barplot()** (Figura 30). En este caso, los argumentos usados son diferentes a los del diagrama circular. Por ejemplo, se manejan los siguientes argumentos; **xlab=** para agregar una etiqueta al eje x, **ylab=** para agregar una etiqueta al eje y, y **las=** para definir la orientación de las etiquetas en el eje **x,** que las establece como **1** para mostrar las etiquetas de manera horizontal. Estos argumentos permiten personalizar el gráfico de barras según las necesidades. Y su gráfico será como se observa en la Figura 31.

Figura 31. *Diagrama de barras Lugar de nacimiento.*



Fuente: elaboración propia

Ejemplo 2.2: a continuación, se presenta la Tabla 4 con una muestra de 460 personas a las que se les preguntó si fumaban o no y también se las clasificó por el género.

Tabla 4. *Muestra de personas clasificadas por género y por consumo de cigarrillo*

	Hombres	Mujeres	Total
Fuma	70	140	210
No fuma	30	220	250
Total	100	360	460

La tabla anterior se puede representar por medio de un diagrama de barras. El código en RStudio será (Figura 32).

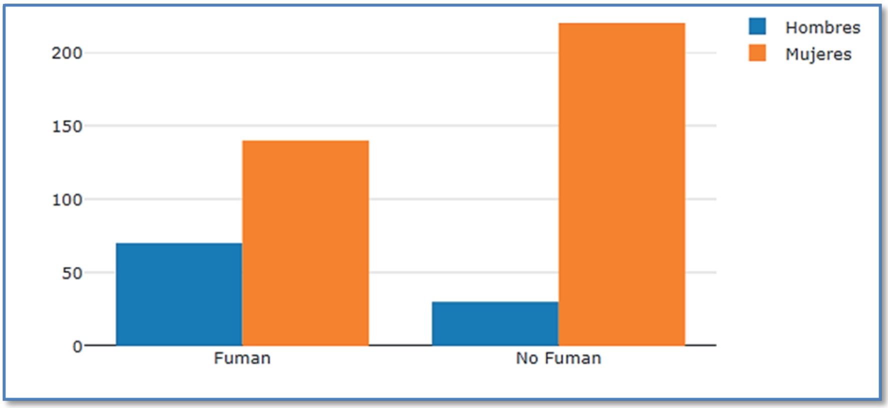
Figura 32. *Código Diagrama de Barras de dos Variables cualitativas.*

```
> variable=c('Fuman','No Fuman')
> hombres=c(70,30)
> mujeres=c(140,220)
> datah=data.frame(variable,hombres,mujeres)
> grafico=plot_ly(datah,x=variable,y=hombres,name='Hombres',type='ba
r')%>%
+   add_trace(y=mujeres,name='Mujeres')
> print(grafico)
```

Fuente: elaboración propia.

El gráfico de barras hecho en RStudio será (Figura 33).

Figura 33. *Diagrama de Barras de Variables Cualitativas de fumadores por Género.*



Fuente: elaboración propia.

- ¿Qué porcentaje de personas en esa muestra fuman? Rta. aproximadamente el 45.7%.
- De las personas que fuman, ¿qué proporción son mujeres? Rta. 2/3=$0.\overline{6}$.
- ¿Qué porcentaje no fuman y son hombres? Rta. aproximadamente el 6.5%.

6. REGLA DE TRES SIMPLE DIRECTA, PROPORCIONES Y PORCENTAJES

6.1. Regla de tres simple directa

Para poder hacer la regla de tres simple directa tenemos que ver que se usa cuando dos magnitudes son directamente proporcionales (dos magnitudes

x y y se dicen que son directamente proporcionales si la razón entre ellas es constante $\frac{y}{x} = k$).

Ejemplo 2.3: el perímetro de un cuadrado P y su lado l son magnitudes directamente proporcionales, la ecuación que representa la relación entre estas magnitudes es $P = 4l$. Se dice que P es directamente proporcional a l con constante de proporcionalidad de 4 porque $\frac{P}{l} = \frac{4l}{l} = 4$, o que l es directamente proporcional a P con constante de proporcionalidad $\frac{1}{4}$ porque $\frac{l}{P} = \frac{l}{4l} = \frac{1}{4}$. Mientras que el área del cuadrado A y su lado l no son magnitudes directamente proporcionales, la ecuación que representa la relación entre estas magnitudes es $A = l^2$, la razón entre estas dos magnitudes no es constante, porque la razón es $\frac{A}{l} = \frac{l^2}{l} = l$, lo cual no es constante, porque depende del lado l. En el primer caso $P = 4l$, podríamos aplicar la regla de tres simple directa, pero para el segundo caso $A = l^2$ no podríamos.

Ejemplo 2.4: en un colegio, el 60% de los estudiantes son de primaria, y el resto, de secundaria. Además, hay 480 estudiantes en primaria. ¿Cuántos estudiantes hay en total en el colegio?

Aquí podemos aplicar regla de tres simple, porque la cantidad de estudiantes es proporcional al porcentaje asignado. Si la variable x corresponde al total de estudiantes en el colegio, que equivale al 100%, entonces:

$$\frac{480}{60\%} = \frac{x}{100\%}$$

Despejando x, tenemos:

$$x = \frac{480}{60\%} * 100\% = 800$$

Entonces el total de estudiantes en el colegio es de 800 estudiantes.

6.2. Proporciones y porcentajes

La proporción es una medida que indica la relación entre una parte y el total de un conjunto. En términos estadísticos, la proporción representa el cociente entre la cantidad de elementos que cumplen con una característica específica y el total de elementos del grupo o muestra en estudio. Se expresa como un valor entre 0 y 1. Además, hay una relación entre la definición de proporción y porcentaje

"Una proporción es una parte o fracción de 1.0 así que la proporción es 1/100 de un porcentaje y un porcentaje es 100 veces una proporción" (GUILFORD & FRUCHTER, 1984, pág. 16) por lo que no se deben confundir estos dos conceptos, que por lo general se usan indistintamente.

Ejemplo 2.5: expresar en notación porcentual la siguiente proporción 0.3.

La notación porcentual será $0.3 * 100 = 30\%$.

Ejemplo 2.6: expresar en notación decimal o proporción el siguiente porcentaje 120%.

La notación decimal será $120/100 = 1.2$.

Ejemplo 2.7: el precio de un computador es de $3'200.000$ sin IVA. ¿Cuánto cuesta el computador, si hay que pagar un IVA que es del 19%?

Aquí podemos usar regla de tres simple directa, donde x es el precio del computador más el IVA:

$$\frac{3'200.000}{100\%} = \frac{x}{119\%}$$

Despejando x, obtenemos:

$$x = \frac{3'200.000}{100\%} * 119\% = 3'808.000$$

También podríamos aplicar:

$$x = 3'200.000 + 3'200.000 * \frac{19\%}{100\%}$$

$$x = 3'200.000 + 3'200.000 * 0.19$$

$$x = 3'200.000 * (1 + 0.19)$$

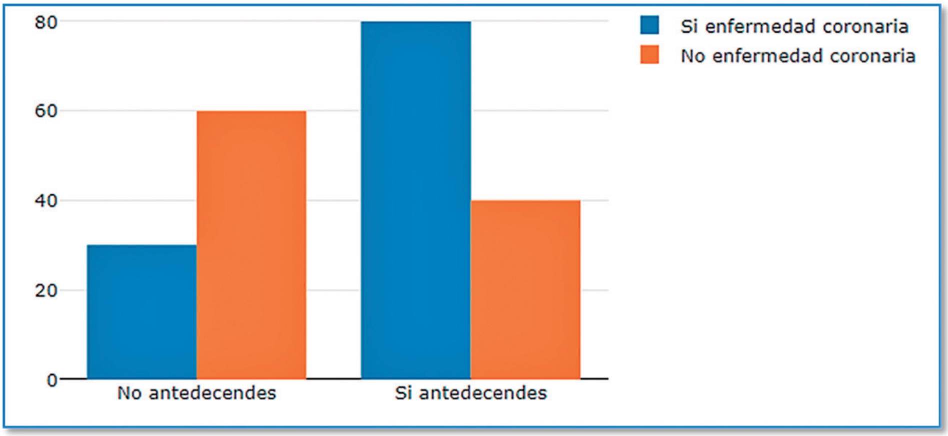
$$x = 3'200.000 * 1.19$$

$$x = 3'808.000$$

7. EJERCICIOS

Ejercicio 1. En un estudio de salud pública, efectuado a un grupo de personas, en las que se establece la relación de los antecedentes cardiacos familiares con enfermedades coronarias, arrojó los siguientes resultados (Figura 34).

Figura 34. *Gráfica de distribución de personas con y sin enfermedad coronaria, con y sin antecedentes cardiacos.*



Fuente: elaboración propia.

a) ¿Qué porcentaje de personas tienen enfermedad coronaria?

b) ¿Qué proporción de personas con enfermedades coronarias tienen antecedentes cardíacos familiares?

c) ¿Qué porcentaje de personas con antecedentes cardiacos familiares no tiene enfermedad coronaria?

d) ¿La proporción de personas no enfermas sin antecedentes familiares es mayor a la proporción de personas enfermas con antecedentes familiares?

Ejercicio 2. A continuación, se muestran los datos de una encuesta sobre las preferencias de consumo de un artículo medidas en una escala de Malo, Regular, Bueno y Excelente, que se hizo a un grupo de hombres y mujeres.

N°	Preferencia de consumo	Género	N°	Preferencia de consumo	Género
1	Excelente	Masculino	11	Bueno	Masculino

2	Bueno	Femenino	12	Regular	Femenino
3	Bueno	Masculino	13	Bueno	Masculino
4	Bueno	Femenino	14	Bueno	Masculino
5	Malo	Femenino	15	Malo	Masculino
6	Bueno	Femenino	16	Malo	Femenino
7	Regular	Masculino	17	Bueno	Femenino
8	Bueno	Masculino	18	Bueno	Masculino
9	Regular	Masculino	19	Excelente	Masculino
10	Excelente	Femenino	20	Bueno	Femenino

a) Construya un diagrama de barras para la variable preferencia de consumo.

b) Construya un diagrama de barras donde aparezcan las variables preferencia de consumo y género.

c) ¿Qué porcentaje de este grupo encuestado son mujeres?

d) ¿Qué porcentaje de este grupo calificaron el producto como excelente y son hombres?

e) Del grupo de los hombres, ¿qué porcentaje calificó el producto como bueno?

f) ¿Qué porcentaje de mujeres calificaron el producto como malo o regular?

Ejercicio 3. Expresar en notación porcentual a) 0.2 b) 0.07 c) 1 d) 1.12 e) 1.5

Ejercicio 4. Exprese en notación decimal *a)* 17% *b)* 180% *c)* 2% *d)* 10% e) 50.5% f) 0.5%.

Ejercicio 5. Un artículo cuesta $ 1000.000 con impuestos. ¿Cuánto cuesta el artículo sin impuestos, si estos son del 19%?

Ejercicio 6. Si pago por una prenda de vestir $85.000 y me dicen que la prenda tenía un 15% de descuento. ¿Cuál es el costo de la prenda sin descuento?

Ejercicio 7. En un almacén, un pantalón cuesta $250.000, y el dueño hace un incremento del 20% en esta prenda. Al mes siguiente, el dueño decide disminuir el precio actual en un 20%. a) ¿Cuál es el precio final de la prenda?, b) ¿Cuál fue su cambio porcentual?

Ejercicio 8. El semestre académico en la Universidad Colegio Mayor de Cundinamarca tiene tres cortes. Los dos primeros con un peso del 30% cada uno y el tercer corte con un peso del 40%, Si un estudiante de estadística obtiene 3.1 y 4.1 en los dos primeros cortes, ¿cuánto debe obtener en el corte final si desea aprobar la asignatura con 4.0?

Ejercicio 9. En un club de fútbol, se juega en campo artificial y campo natural. Se sabe que el 30% de las lesiones se producen en campo artificial y que este valor es 50% más alto de lo que sería en campo natural. ¿Qué porcentaje de lesiones se producen en campo natural?

8. PRESENTACIÓN DE DATOS CUANTITATIVOS

8.1. Diagrama de puntos o frecuencias

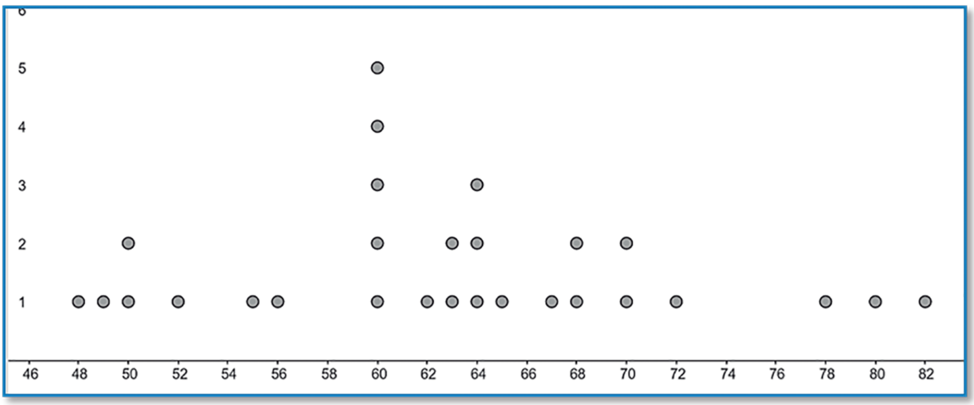
Es un gráfico que representa los valores individuales de una variable cuantitativa mediante puntos. Para hacer el diagrama de puntos se traza un eje horizontal donde se representa la escala de valores de los datos y cada dato se grafica con un punto, apilándose verticalmente si el valor se repite; así, cada punto representará la frecuencia en el que el valor del dato se repite.

Ejemplo 2.8: se tienen el peso en kg de 28 estudiantes:

48	60	80	64	82	68	68
50	60	70	65	60	60	63
50	60	55	70	49	63	78
52	62	64	72	64	56	67

El diagrama de puntos se puede ver como la Figura 35.

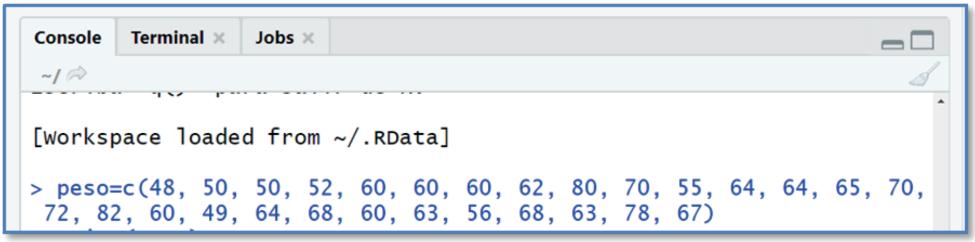
Figura 35. *Diagrama de puntos para los pesos en kg.*



Fuente: elaboración propia.

El diagrama de puntos en RStudio se puede ver de la siguiente manera, ingresando los datos en RStudio, Figura 36.

Figura 36. *Asignando la variable peso al conjunto de datos en RStudio.*

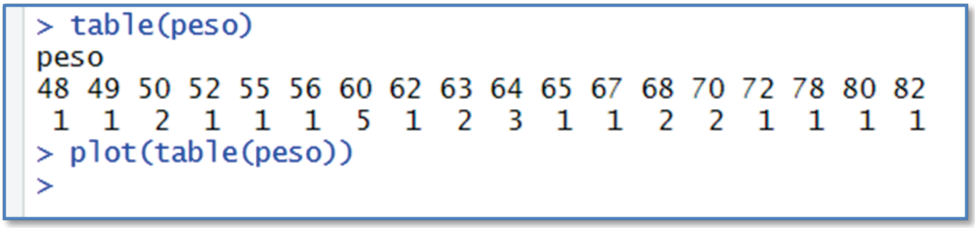



```
[Workspace loaded from ~/.RData]

> peso=c(48, 50, 50, 52, 60, 60, 60, 62, 80, 70, 55, 64, 64, 65, 70,
 72, 82, 60, 49, 64, 68, 60, 63, 56, 68, 63, 78, 67)
```

Fuente: elaboración propia.

Ahora hacemos la tabla de frecuencias y graficamos la tabla de frecuencias (Figura 37), en RStudio.

Figura 37. *Comandos para crear tabla de frecuencias y gráfico de la variable peso en kg.*

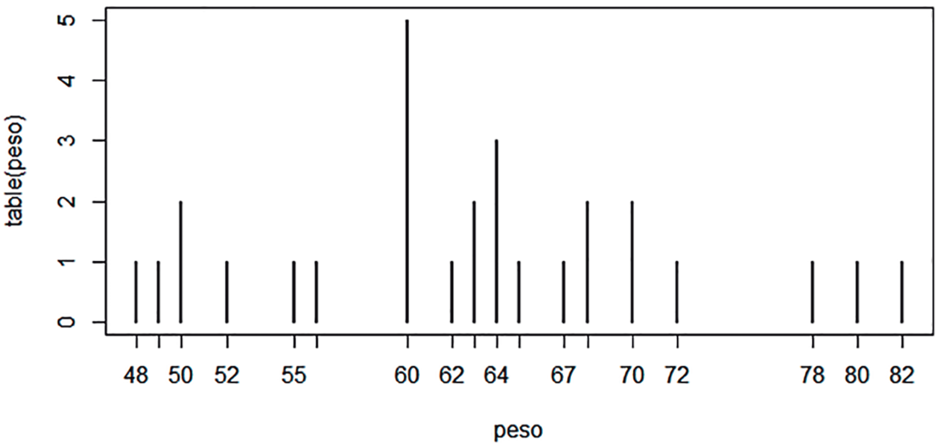


```
> table(peso)
peso
48 49 50 52 55 56 60 62 63 64 65 67 68 70 72 78 80 82
 1  1  2  1  1  1  5  1  2  3  1  1  2  2  1  1  1  1
> plot(table(peso))
>
```

Fuente: elaboración propia.

La gráfica de la tabla de frecuencias será (Figura 38).

Figura 38. *Gráfico de la tabla de frecuencias de la variable peso en Kg.*



Fuente: elaboración propia.

Podemos ver en los anteriores gráficos que la tendencia en los pesos es de 60 kg (es el peso de mayor frecuencia) y que existe cierta dispersión en los pesos. El menor peso es de 48 kg y el mayor de 82 kg.

8.2. Diagrama de tallos y hojas

El diagrama de tallos y hojas es una gráfica visual para organizar un conjunto de datos con valores numéricos de dos o más dígitos, en donde se representan los tallos (generalmente la parte inicial o los primeros dígitos) y las hojas (generalmente la parte final o los últimos dígitos).

Los pasos para construir este diagrama son los siguientes:

1. Se seleccionan los primeros dígitos del valor de cada dato para el valor del tallo. El dígito final de cada uno de los valores de los datos se convierte en la hoja. Se necesitan de al menos 5 tallos para poder hacer la representación.
2. Hacer una lista de los valores del tallo en forma vertical.
3. Registrar las hojas en el valor correspondiente del tallo.
4. Indicar la cantidad de datos que hay en cada tallo.

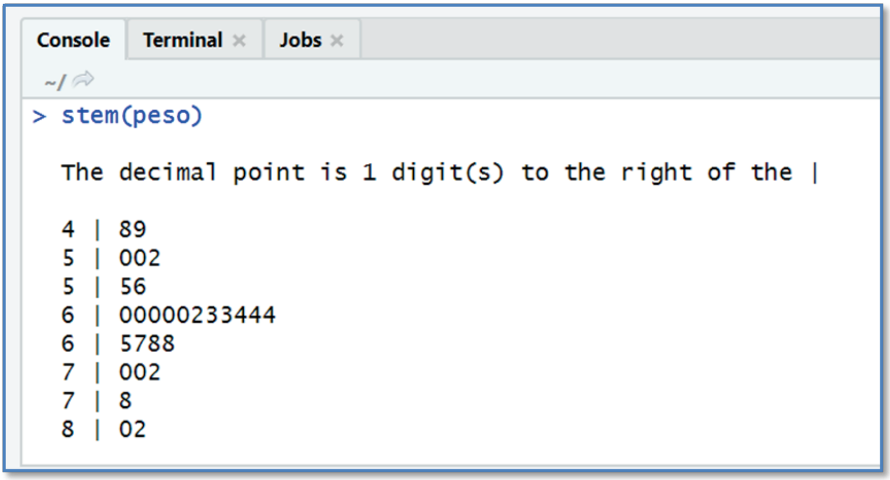
Ejemplo 2.9: tomando los datos del ejemplo 2.8, se tienen el peso en kg de 28 estudiantes:

48	60	80	64	82	68	68
50	60	70	65	60	60	63
50	60	55	70	49	63	78
52	62	64	72	64	56	67

En este caso, como los datos son de dos dígitos, se elige el tallo como el primer dígito y la hoja el segundo dígito. Los valores que toma el tallo son 4, 5, 6, 7, 8.

En RStudio, para hacer un diagrama de tallos y hojas, se escribe el comando **stem()**, cuyo resultado muestra la Figura 39.

Figura 39. *Diagrama de tallos y hojas para la variable peso en kg.*



Fuente: elaboración propia.

En la anterior figura, se ve que los valores del tallo que se repiten son 5, 6 y 7, porque se distribuyen los valores de las hojas de 0 a 4, para el primero, y de 5 a 9, para el segundo. Además, el anterior gráfico permite ordenar los datos y mirar la tendencia (pesos entre 60 kg a 64 kg) y la dispersión (valores de los pesos comprendidos entre 48 kg a 82 kg).

8.3. TABLAS DE FRECUENCIA PARA DATOS CUANTITATIVOS (CONTINUOS)

En una tabla de distribución de frecuencias para datos cuantitativos, por lo general deben aparecer los siguientes elementos intervalos de clase, marca de clase $(\boldsymbol{m_i})$, frecuencia absoluta $(\boldsymbol{f_i})$, frecuencia relativa $(\boldsymbol{h_i})$, frecuencia absoluta acumulada $(\boldsymbol{F_i})$ y frecuencia relativa acumulada $(\boldsymbol{H_i})$. El total de datos de la muestra con los cuales queremos construir esa tabla se denota como $\boldsymbol{n}$. La Tabla 5 muestra las definiciones de cada uno de estos elementos.

Tabla 5. *Elementos de una tabla de frecuencias para datos cuantitativos*

Intervalos de clase	Marca de clase $(\boldsymbol{m_i})$	Frecuencia absoluta $(\boldsymbol{f_i})$	Frecuencia absoluta acumulada $(\boldsymbol{F_i})$	Frecuencia relativa $(\boldsymbol{h_i})$	Frecuencia relativa acumulada $(\boldsymbol{H_i})$
[Límite inferior, límite superior)	Promedio de límite inferior y superior	Cantidad de datos en el intervalo	Cantidad de datos que se van acumulando a través de cada intervalo	Cantidad de datos en el intervalo sobre el total de datos	Frecuencia relativa que se va acumulando a través de cada intervalo

Para construir una tabla de frecuencias (Tabla 5) tenemos

Intervalos de clase es cada uno de los rangos de valores en que se decide agrupar los datos para hacer un resumen de ellos. Se deben tener en cuenta los siguientes pasos para construirlos:

1. Se determina el rango. Se define el rango como la diferencia entre el valor del dato mayor y el menor.
2. Fijar el número de intervalos el número total de intervalos lo denotamos con $\boldsymbol{l}$. El cálculo del número de intervalos $\boldsymbol{l}$ de clase se puede hacer con cualesquiera de las siguientes fórmulas.

   Con la fórmula $\boldsymbol{l} = \sqrt{n}$ o con la regla de Sturges $\boldsymbol{l} = 1 + 3.3\log(n)$, la primera se usa para una cantidad pequeña de datos y la segunda para una cantidad grande de datos, ya que lo recomendable es que la tabla de frecuencias no contenga más de 25 intervalos de clase.
3. Establecer la longitud de cada intervalo, que se calcula como:

$$Longitud\ de\ cada\ intervalo \cong \frac{Rango}{N^{o}\ de\ intervalos}$$

Marca de clase es el punto medio de cada intervalo de clase. Se calcula promediando el valor del límite inferior del intervalo con el valor del límite superior (se denota como m_i, donde i representa el i-esimo intervalo).

Frecuencia absoluta es el número total de datos que se encuentran en ese intervalo de clase (se denota como f_i, donde i representa el i-esimo intervalo).

Frecuencia absoluta acumulada es el número que resulta de sumar la frecuencia absoluta f_i del i-esimo intervalo con todas las anteriores. La siguiente fórmula expresa la frecuencia absoluta acumulada, denota como F_i para el i -esimo intervalo

$$F_i = f_1 + f_2 + f_3 + \cdots + f_i$$

$$F_i = F_{i-1} + f_i$$

Como tenemos un total de l intervalos:

$$F_l = f_1 + f_2 + f_3 + \cdots + f_i + \cdots + f_l$$

Frecuencia relativa es el cociente entre la frecuencia absoluta de clase fi y el número total de datos calculados, como $n = \sum_{i=1}^{l} f_i$. Se denota como h_i, donde i representa el i-esimo intervalo:

$$h_i = \frac{f_i}{\sum_{i=1}^{l} f_i} = \frac{f_i}{n}$$

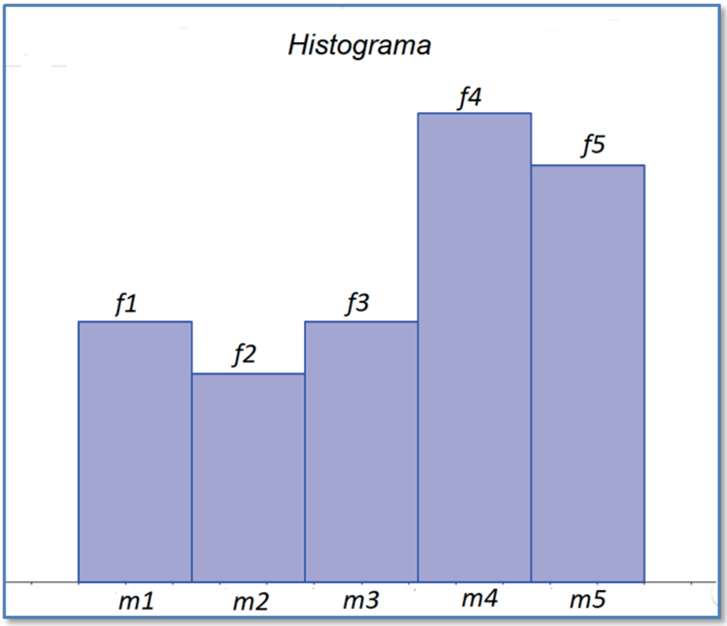
Frecuencia relativa acumulada es el cociente entre la frecuencia absoluta acumulada de clase Fi y el número total de datos calculados como $n = \sum_{i=1}^{l} f_i$. Se denota como H_i, donde i representa el i-esimo intervalo).

$$H_i = \frac{F_i}{\sum_{i=1}^{l} f_i} = \frac{F_i}{n}$$

8.4. HISTOGRAMA

El gráfico de una tabla de frecuencias para datos cuantitativos se conoce como histograma y este representa la distribución de un conjunto de datos. El diagrama muestra en el eje horizontal los intervalos o marcas de clase y en el eje vertical las frecuencias (absolutas o relativas) de cada intervalo mediante barras rectangulares cuya altura corresponde a la frecuencia (absoluta o relativa) de cada clase (Figura 40).

Figura 40. *Representación de un histograma (datos cuantitativos).*



Fuente: elaboración propia.

Ejemplo 2.10: se tienen los pesos (en kg) de los estudiantes del ejemplo 2.8.

48	60	80	64	82	68	68
50	60	70	65	60	60	63
50	60	55	70	49	63	78
52	62	64	72	64	56	67

Para construir una tabla de frecuencias e histograma se halla:

- <u>Rango</u> = 82-48 = 34
- <u>Número de intervalos</u> $\sqrt{28} = 5.29$.También con la fórmula de Sturges = $1 + 3.3\log 28 = 5.77$.
- Podemos trabajar con 5 intervalos.
- La longitud de cada intervalo= $\frac{34}{5} = 6.8 \approx 7$

Y completando la tabla de frecuencias obtenemos (Tabla 6).

Tabla 6. *Tabla de frecuencias para la variable peso en kg*

Intervalos de clase	Marca de clase (m_i)	Frecuencia absoluta (f_i)	Frecuencia absoluta acumulada (F_i)	Frecuencia relativa (h_i)	Frecuencia relativa acumulada (H_i)
[48,55)	51.5	5	5	5/28≈ 0.18	5/28≈ 0.18
[55,62)	58.5	7	12	7/28=0.25	12/28≈ 0.43
[62,69)	65.5	10	22	10/28≈ 0.36	22/28≈ 0.79
[69,76)	72.5	3	25	3/28≈ 0.11	25/28≈ 0.89
[76,83)	79.5	3	28	3/28≈ 0.11	28/28=1

Ahora, para hacer la tabla de frecuencias e histograma del ejemplo 2.10 en RStudio, se usan funciones y argumentos mencionados en la sección 2.8. Además, de construir un vector de datos llamado **peso()** Figura 41.

Figura 41. *Código de los datos peso.*

```
> peso=c(48, 50, 50, 52, 60, 60, 60, 62, 80, 70, 55, 64, 64, 65, 70,
72, 82, 60, 49, 64, 68, 60, 63, 56, 68, 63, 78, 67)
> sort(peso)
 [1] 48 49 50 50 52 55 56 60 60 60 60 60 62 63 63 64 64 64 65 67
[21] 68 68 70 70 72 78 80 82
```

Fuente: elaboración propia.

En la figura anterior se observa que la función **sort()** nos permite organizar los datos de forma ascendente, desde el valor mínimo hasta el máximo.

Para construir la tabla de frecuencias de este tipo de datos se usa la función **table()**. En este caso, se emplea el argumento **cut()** para dividir los datos de la variable 'peso' en intervalos. El argumento **breaks=** nos permite especificar la cantidad de intervalos deseados. En este caso, se parte en cinco intervalos (Figura 42).

Figura 42. *Código para la Frecuencia Absoluta.*

```
> frec_abs=table(cut(peso, breaks = 5))
> print(frec_abs)

 (48,54.8] (54.8,61.6] (61.6,68.4] (68.4,75.2]   (75.2,82]
         5           7          10           3           3
```

Fuente: elaboración propia.

Para calcular la frecuencia absoluta acumulada de estos datos, se usa la función **cumsum()** aplicada sobre la frecuencia absoluta. La función **cumsum()** hace la suma acumulativa de los valores de la frecuencia absoluta (Figura 43).

Figura 43. *Código para la Frecuencia Absoluta Acumulada.*

```
> frec_abs_acum=cumsum(frec_abs)
> print(frec_abs_acum)
 (48,54.8] (54.8,61.6] (61.6,68.4] (68.4,75.2]   (75.2,82]
         5          12          22          25          28
```

Fuente: elaboración propia.

La tabla de frecuencia está casi completa, solo falta calcular la frecuencia relativa y su acumulada. Además, debemos agregar todas las columnas correspondientes con sus respectivos nombres para que la tabla esté completa en su totalidad (Figura 44).

Figura 44. *Código de la tabla de frecuencias.*

```
> frec_rel=prop.table(frec_abs)
> frec_rel_acum=cumsum(frec_rel)
> tabla_frecuencias=cbind(frec_abs, frec_abs_acum, frec_rel, frec_rel
_acum)
> colnames(tabla_frecuencias)=c('Frec. Absoluta', 'Frec. Abs. Acumula
da', 'Frec. Relativa', 'Frec. Rel. Acumulada')
> print(tabla_frecuencias)
            Frec. Absoluta Frec. Abs. Acumulada Frec. Relativa
(48,54.8]                5                    5      0.1785714
(54.8,61.6]              7                   12      0.2500000
(61.6,68.4]             10                   22      0.3571429
(68.4,75.2]              3                   25      0.1071429
(75.2,82]                3                   28      0.1071429
            Frec. Rel. Acumulada
(48,54.8]              0.1785714
(54.8,61.6]            0.4285714
(61.6,68.4]            0.7857143
(68.4,75.2]            0.8928571
(75.2,82]              1.0000000
```

Fuente: elaboración propia.

El gráfico que se usa para representar esta tabla es el histograma. Para construir el histograma es necesario calcular las marcas de clase, que corresponden al punto medio de cada intervalo. Para obtener las marcas de clase, se usa la función **seq()**, que genera una secuencia de números. En este caso, se usa la función **min()** y **max()** para obtener el valor mínimo y máximo de los datos. Luego, se usa la opción **length.out = 5 + 1** en la función **seq()**, para generar una secuencia de **5** marcas de clase completas. La adición de **1** se debe a que las marcas de clase se encuentran en el punto medio de cada intervalo (Figura 45).

Figura 45. *Código para la marca de clase.*

```
> frec_abs

  (48,54.8] (54.8,61.6] (61.6,68.4] (68.4,75.2]   (75.2,82]
          5           7          10           3           3
> etiquetas_1=c('5','7','10','3','3')
```

Fuente: elaboración propia.

Para crear el histograma, se usa la función **hist()**. Al igual que el diagrama de barras, se emplean argumentos similares para las etiquetas, como el título principal (**main**=), etiqueta del eje x (**xlab**=) y etiqueta del eje y (**ylab**=). La diferencia principal al crear un histograma es que debe especificarse el argumento **breaks**, que corresponde al vector de las marcas de clase. Las marcas de clase, obtenidas previamente, representan los puntos medios de cada intervalo en el histograma (Figura 46).

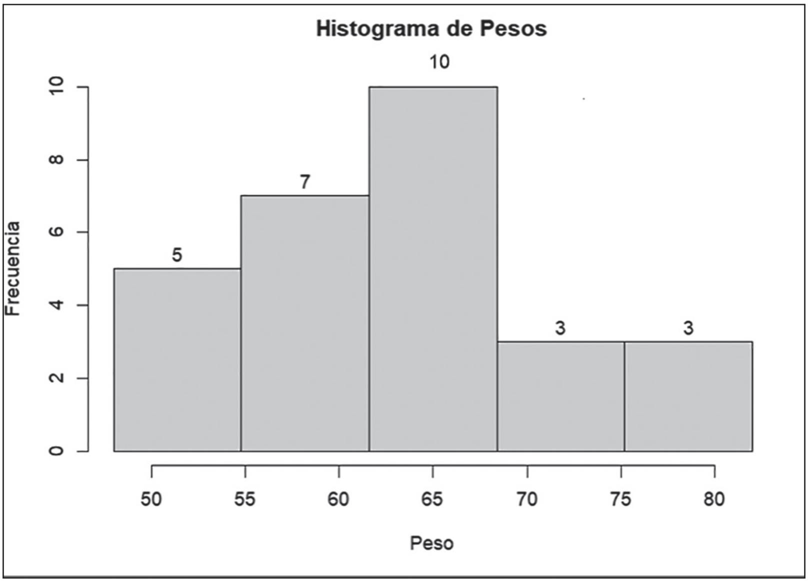
Figura 46. *Código del histograma.*

```
> hist(peso, breaks = seq(min(peso), max(peso), length.out = 5+1 ),la
bels = etiquetas_1, main = 'Histograma de Pesos', xlab = 'Peso', ylab
= 'Frecuencia')
```

Fuente: elaboración propia.

De lo anterior resulta el histograma (Figura 47).

Figura 47. *Histograma de Pesos.*



Fuente: elaboración propia.

9. EJERCICIOS

Ejercicio 1. Considere los datos siguientes que representan el número de kilómetros que recorre una muestra de carros en cierta ciudad, en dos días.

17	62	15	65	28	51	24	65
39	41	35	15	39	32	36	37
40	21	44	37	59	13	44	56
12	54	64	59				

a) Trace un diagrama de puntos.

b) Trace un diagrama de tallos y hojas.

c) Construya una tabla de frecuencias con intervalos de clase (**sugerencia 5 intervalos**), marcas de clase m_i, frecuencias absolutas f_i, frecuencias absolutas acumuladas F_i, frecuencias relativas h_i, frecuencias relativas acumuladas H_i. (especificando los pasos en la construcción).

d) Trace un histograma.

Ejercicio 2. Los datos siguientes son las horas de uso de computadora en el hogar por día en una muestra de 20 personas:

4.1	1.5	10.4	5.9	3.4	5.7	1.6	6.1	3.0	3.7
3.1	4.8	2.0	1.5	5.4	4.2	3.9	4.1	11.5	3.5

a) Construya una tabla de frecuencias con 5 intervalos y muestre intervalos de clase, marcas de clase m_i, frecuencias absolutas f_i, frecuencias absolutas acumuladas F_i, frecuencias relativas h_i, frecuencias relativas acumuladas H_i.

b) ¿Qué proporción de personas usan la computadora en al hogar menos de 5.5 horas por día?

c) Trace el histograma de las frecuencias absolutas.

Ejercicio 3. Se tiene la Tabla 7 que corresponde a los datos de una variable cuantitativa continua.

Tabla 7. *Tabla de frecuencias absolutas, relativas y acumuladas*

Límites	m_i	f_i	h_i	F_i	H_i
[500,550)		5			
[550,600)			0.2		
[600,650)					0.7
[650,700)					
[700,750)		7		50	

a) Complete la Tabla 7.

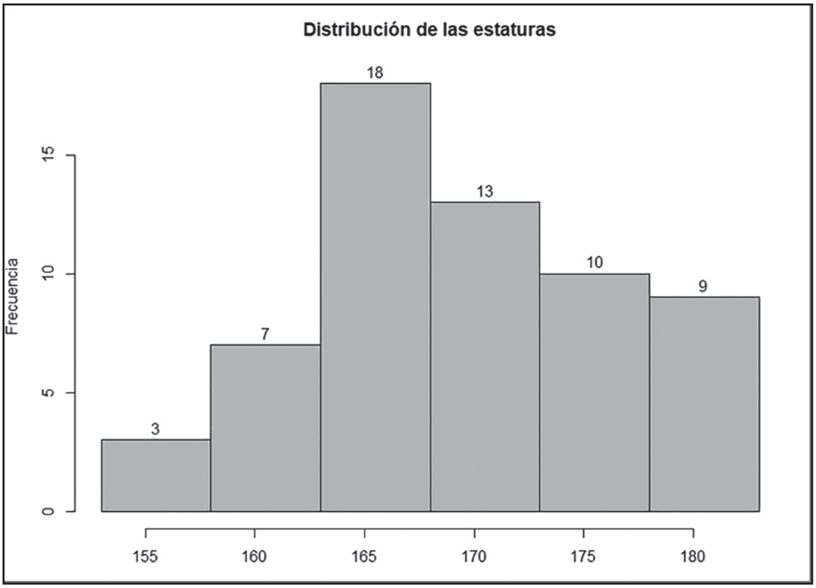
b) ¿Cuál es el intervalo que presenta mayor porcentaje de datos?

c) ¿En qué rango de valores se encuentra el 30% superior de todos los datos?

d) Indique los valores de m_1, f_2, F_i, h_4 y H_3.

Ejercicio 4. A continuación se presenta un histograma con las alturas en cm de un grupo de estudiantes con sus respectivas marcas de clase (Figura 48).

Figura 48. *Distribución de las estaturas en centímetros.*



Fuente: elaboración propia.

a) Construya una tabla de frecuencias y muestre intervalos de clase, marcas de clase m_i, frecuencias absolutas f_i, frecuencias absolutas acumuladas F_i, frecuencias relativas h_i, frecuencias relativas acumuladas H_i.

b) ¿Qué porcentaje de personas miden más de 157.5 cm?

c) ¿Cuántas personas conforman el intervalo de mayor altura?

d) ¿Qué porcentaje de personas en esa muestra tienen una altura menor a 167.5 cm?

10. MEDIDAS DE TENDENCIA CENTRAL PARA DATOS NO AGRUPADOS

10.1. Media aritmética

Muestral

La media aritmética muestral representada con $(\bar{x})$, se calcula sumando todos los valores del conjunto de datos y dividiendo el resultado entre la cantidad total de datos. Si $x_1, x_2, \ldots, x_n$ representan n datos cuantitativos, se define la media aritmética como:

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

Poblacional

La media poblacional μ se define como:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{\sum_{i=1}^{N} x_i}{N}$$

Donde N es el tamaño de la población.

Ejemplo 2.11: para la siguiente muestra calcular la media o promedio 2, 3, 4, 5, 9, y 13.

$$\bar{x} = \frac{2+3+4+5+9+13}{6} = 6$$

En RStudio el código será (Figura 49).

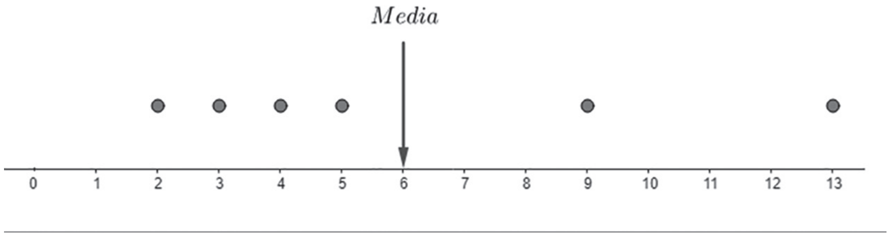
Figura 49. *Código para calcular la media de un conjunto de datos.*

```
> x<-c(2,3,4,5,9,13)
> mean(x)
[1] 6
>
```

Fuente: elaboración propia.

Ahora observemos el diagrama de puntos de los datos del ejemplo 2.11 (Figura 50).

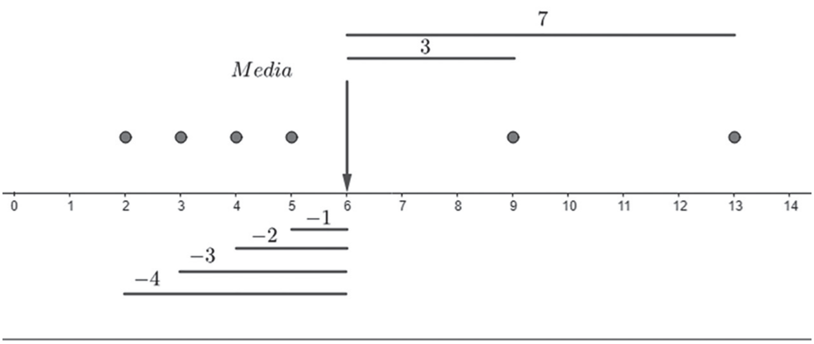
Figura 50. *Diagrama de puntos.*



Fuente: elaboración propia.

Si definimos la desviación= Valor del dato-valor de la media $= x_i - \bar{x}$. (Figura 51).

Figura 51. *Gráfico donde se representan las desviaciones de un conjunto de datos.*



Fuente: elaboración propia.

Se observa en la anterior figura que $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$., la suma de todas las desviaciones (la diferencia entre el dato y la media) es igual a cero.

Ejemplo 2.12: ahora si se tiene la siguiente muestra 2, 3, 4, 5, 9, y 1000. Son algunos datos del **Ejemplo 2.11**, pero en vez del valor de 13, tenemos el valor de 1000 y al calcular la media:

$$\bar{x} = \frac{2+3+4+5+9+1000}{6} = 170.5$$

Vemos que la media se deja afectar por datos inusuales.

En RStudio, el cálculo será (Figura 52).

Figura 52. *Cálculo de la media en RStudio.*

```
> x<-c(2,3,4,5,9,1000)
> mean(x)
[1] 170.5
> |
```

Fuente: elaboración propia.

10.2. Mediana

La mediana muestral, denotada por $\tilde{x}$, es el valor central de un conjunto de datos ordenados en forma ascendente. Si el número de datos en la muestra es impar, la mediana es el valor que ocupa la posición central. Si el número de datos es par, la mediana se calcula como el promedio de los dos valores centrales.

Para calcular la mediana se requiere encontrar su posición (denotad por k), que se calcula de la siguiente manera:

Posición de la mediana $k = (n + 1)(0.5)$. Donde n es el tamaño de la muestra.

Ejemplo 2.13: calcular la mediana para la siguiente muestra (la cantidad de datos es un número impar).

21	35	40	28	24	15	36	23	29	32	23

1. Se ordenan los datos:

15	21	23	23	24	28	29	32	35	36	4

2. Se calcula la posición de la mediana llamada k. Como el tamaño de la muestra es de $n = 11$, entonces la posición de la mediana es $k = (11 + 1)(0.5) = 6$. Al ordenar los datos de forma ascendente, la mediana será el dato de la posición 6.

$$\tilde{x} = x_6.$$

3. Calcular la mediana:

$$\tilde{x} = x_6 = 28$$

En Rstudio, los anteriores pasos se muestran en la Figura 53.

Figura 53. *Código para el cálculo de la mediana cuando la cantidad de datos es impar.*

```
Console   Terminal ×   Jobs ×
~/
> x<-c(21, 35, 40, 28, 24, 15, 36, 23, 29, 32, 23)
> y<-sort(x)
> print(y)
 [1] 15 21 23 23 24 28 29 32 35 36 40
> length(y)
[1] 11
> (11+1)*0.5
[1] 6
> y[6]
[1] 28
> median(x)
[1] 28
```

Fuente: elaboración propia.

Ejemplo 2.14: calcular la mediana para la siguiente muestra (la cantidad de datos es un número par).

201 224 201 219 214 220

1. Se ordenan los datos:

   201 201 214 219 220 224

2. Se calcula la posición de la mediana. Como el tamaño de la muestra es 6, entonces la posición de la mediana es $k = (6 + 1)(0.5) = 3.5$. Al ordenar los datos de forma ascendente, la mediana será el dato de la posición 3 y 4 promediados:

$$\tilde{x} = \frac{x_3 + x_4}{2}$$

3. Calcular la mediana:

$$\tilde{x} = \frac{x_3 + x_4}{2} = \frac{214 + 219}{2} = 216.5$$

En Rstudio, los anteriores pasos se muestran en la Figura 54.

Figura 54. *Código para el cálculo de la mediana cuando la cantidad de datos es par.*

```
[Workspace loaded from ~/.RData]

> x<-c(201, 224, 201, 219, 214, 220)
> y<-sort(x)
> print(y)
[1] 201 201 214 219 220 224
> length(y)
[1] 6
> (6+1)*0.5
[1] 3.5
> y[3:4]
[1] 214 219
> mean(y[3:4])
[1] 216.5
> median(x)
[1] 216.5
```

Fuente: elaboración propia.

Ejemplo 2.15: ahora, si se tienen los mismos datos del ejemplo 2.13, pero en vez de 40 es 1000. Calcular la mediana.

15 21 23 23 24 28 29 32 35 36 1000

Es el mismo procedimiento. La mediana seguirá siendo $\tilde{x} = x_6 = 28$.

En Rstudio, se calcula la mediana con la función **median()** como en la Figura 55.

Figura 55. *Calculo en RStudio de la mediana.*

```
> x<-c(15, 21, 23, 23, 24, 28, 29, 32, 35, 36, 1000)
> median(x)
[1] 28
>
```

Fuente: elaboración propia.

Observamos que la mediana no se deja afectar por datos inusuales.

10.3. Moda

Los valores de los datos que se repiten con mayor frecuencia en un conjunto de observaciones se denomina moda y se denota como Mo.

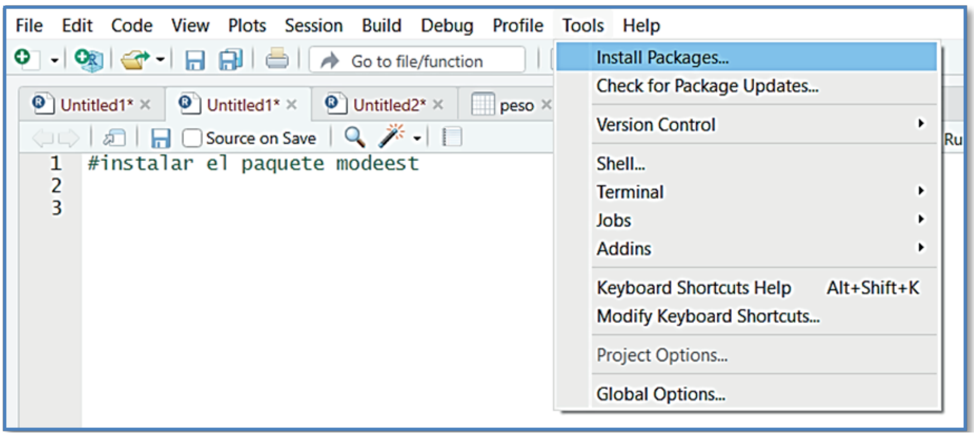
Ejemplo 16: calcular la moda para los datos del ejemplo 2.13.

21 35 40 28 24 15 36 23 29 32 23

La moda $Mo = 23$.

Para calcular la moda en RStudio hay que instalar el paquete **modeest.** Para poder instalar un paquete vamos en la barra del menú superior la opción **Tools->Install Packages** como se muestra en la Figura 56.

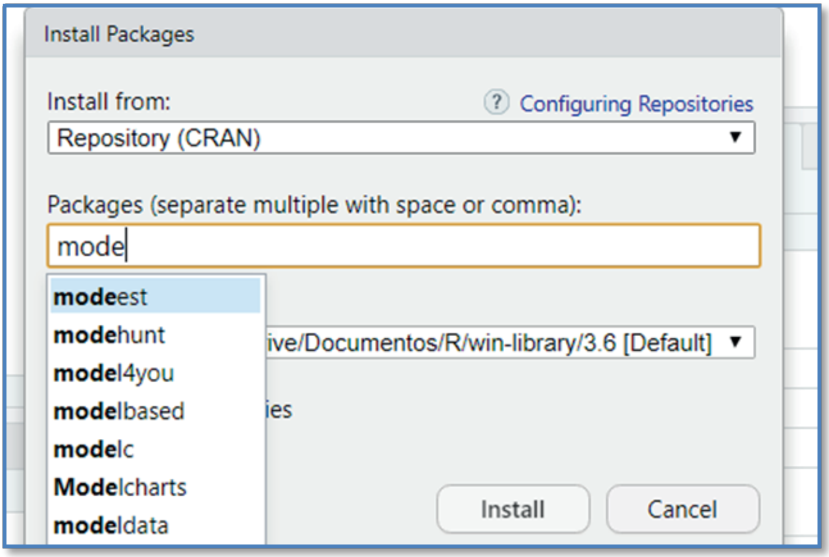
Figura 56. *Herramienta Tools para instalar paquetes en RStudio.*



Fuente: elaboración propia.

Después escribimos **modeest,** seleccionamos e instalamos Figura 57.

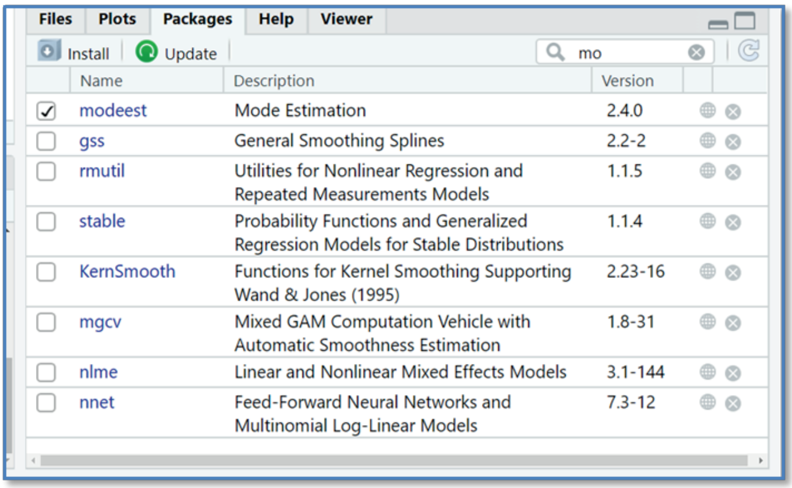
Figura 57. *Ventana para instalación de paquetes en RStudio.*



Fuente: elaboración propia.

Para activar el paquete **modeest,** vamos al panel derecho inferior (llamado entorno de utilidades) y seleccionamos la parte de paquetes y buscamos **modeest** y la activamos dando clic a la opción, Figura 58.

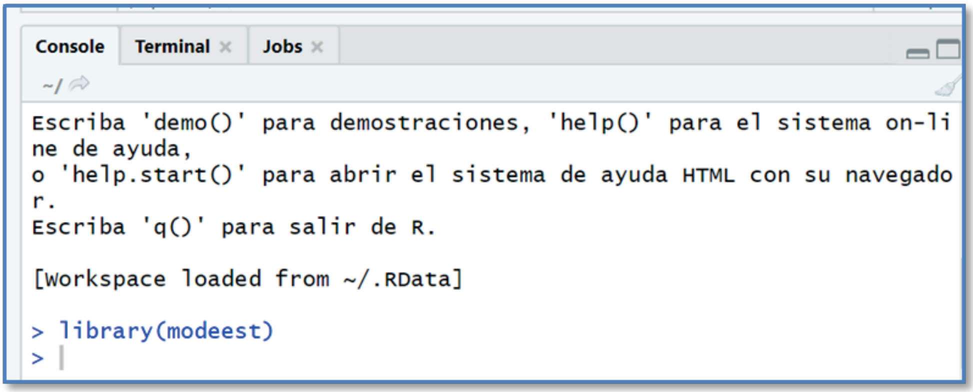
Figura 58. *Ventana de selección de paquetes.*



Fuente: elaboración propia.

En la parte de la consola aparecerá, como se muestra en la Figura 59.

Figura 59. *Función "library" para el llamado de paquetes en RStudio.*



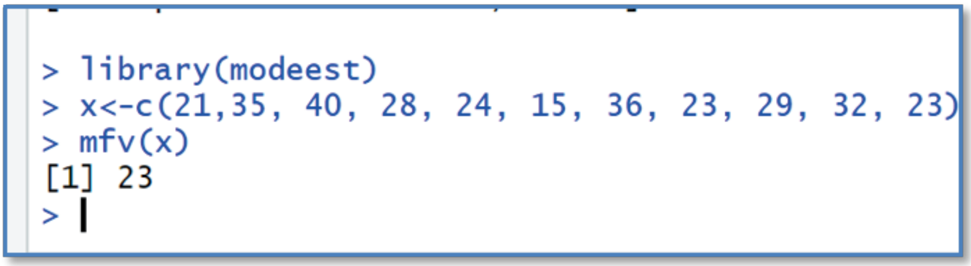
```
Console   Terminal ×   Jobs ×
~/
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-li
ne de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegado
r.
Escriba 'q()' para salir de R.

[Workspace loaded from ~/.RData]

> library(modeest)
> 
```

Fuente: elaboración propia.

Lo que significa que la librería **modeest** ya se encuentra instalada.

Y calculamos la moda (Figura 60) con la función **mfv(),** así:

Figura 60. *Código para el cálculo de la moda.*



```
> library(modeest)
> x<-c(21,35, 40, 28, 24, 15, 36, 23, 29, 32, 23)
> mfv(x)
[1] 23
> 
```

Fuente: elaboración propia.

Ejemplo 2.17: para la siguiente muestra, calcular la moda 2, 3, 4, 5, 9, y 13. En este caso, no hay moda, porque todos tienen la misma frecuencia, que es 1. En RStudio observamos que todos los datos se repiten una sola vez (Figura 61).

Figura 61. *Cálculo de la moda para un conjunto de datos.*

```
> library(modeest)
> x<-c(2, 3, 4, 5, 9, 13)
> mfv(x)
[1]  2  3  4  5  9 13
> 
```

Fuente: elaboración propia.

10.4. Media ponderada

La media ponderada es la misma media aritmética $\bar{x}$, pero en este caso cada uno de los datos x_i, tiene un peso w_i. La media ponderada se define por:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Donde w_i son los pesos de las observaciones x_i.

Ejemplo 2.18: en un ascensor hay 3 hombres y 2 mujeres. Los hombres pesan en promedio 80 kg y las mujeres 60 kg. ¿Cuánto pesan en promedio todas las personas que están en el ascensor?

$$\bar{x} = \frac{3 \times 80 + 2 \times 60}{3 + 2} = 72$$

En RStudio construimos dos matrices con la función **matrix(),** una matriz de pesos **w** de tamaño 2×1 de la forma $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ y otra matriz de valores de **x** de tamaño 1×2 de la forma $\begin{pmatrix} 80 & 60 \end{pmatrix}$ multiplicamos las matrices **x** y **w** de la forma $\begin{pmatrix} 3 \\ 2 \end{pmatrix} * \begin{pmatrix} 80 & 60 \end{pmatrix}$ usando el operador **%*%** (para multiplicar matrices), después se divide por la sumatoria de los pesos **sum(w)** como se muestra en la Figura 62.

Figura 62. *Código para el cálculo de la media ponderada.*

```
> w=matrix(c(3,2), nrow = 2, ncol = 1)
> sum(w)
[1] 5
> x=matrix(c(80,60), nrow = 1, ncol = 2)
> x%*%w
     [,1]
[1,]  360
> x%*%w/sum(w)
     [,1]
[1,]   72
```

Fuente: elaboración propia.

10.5. Media geométrica (MG)

De un conjunto de n números positivos se define como la raíz n-ésima del producto de los n números (Behar & Grima, 2010). La fórmula para la media geométrica está dada por:

$$MG = \sqrt[n]{x_1 x_2 \ldots x_n}$$

Existen dos usos principales de la media geométrica:

1. Se usa para promediar tasas de crecimiento.
2. Cuando los datos en muestras diferentes presentan unidades de medida distintos.

Ejemplo 2.19: el crecimiento de una población de 100.000 habitantes afectada por la inmigración en estos años 2016, 2017 y 2018, respectivamente, ha sido de 25%, 35% y 60%. ¿Cuál ha sido la tasa de crecimiento promedio de la población? Si desarrollamos el ejercicio, especificando la población al finalizar cada año obtenemos la Tabla 8.

Tabla 8. *Tasas de crecimiento anuales de una población*

Año	Población inicial	Tasa de crecimiento	Población final de año
2016	100.000	25%	125.000
2017	125.000	35%	168.750
2018	168.750	60%	270.000

De esta forma:

$$100.000 \times 1.25 \times 1.35 \times 1.60 = 270.000$$

Por otro lado, si calculamos el promedio de las tasas de crecimiento anual con la media nos daría lo siguiente:

$$\bar{x} = \frac{25 + 35 + 60}{3} = 40$$

Estaríamos diciendo que el crecimiento porcentual promedio fue de 40% en cada año, es decir:

$$100.000 \times 1.4 \times 1.4 \times 1.4 = 274.400$$

Pero este resultado no es el mismo que da en la tabla al finalizar el 2018. Ahora, si calculamos el promedio de la tasa de crecimiento de la población con la media armónica:

$$MG = \sqrt[3]{1.25 \times 1.35 \times 1.6} \approx \sqrt[3]{2.7} \approx 1.39247665$$

Si calculamos:

$$100.000 \times 1.39247665 \times 1.39247665 \times 1.39247665 \approx 270.000$$

Este resultado es el mismo que en la tabla al finalizar el 2018. Es decir, la tasa de crecimiento promedio adecuada será la calculada con la media armónica.

10.6. MEDIA ARMÓNICA ($\boldsymbol{MA}$)

La media armónica se calcula como: $MA = \dfrac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right)}$

Este valor se emplea por lo general para promediar variaciones con respecto al tiempo.

Ejemplo 2.20: una persona viaja en su carro de una ciudad a otra, y la distancia entre estas dos ciudades es de 100 km. También se ve que la velocidad promedio de ida fue de 70 km/h, y la de regreso fue de 80 km/h. ¿Cuál fue la velocidad promedio?

Como sabemos:

$$Velocidad = \frac{distancia}{tiempo} = \frac{100+100}{\frac{100}{70}+\frac{100}{80}} = \frac{200}{100\left(\frac{1}{70}+\frac{1}{80}\right)} = \frac{2}{\frac{1}{70}+\frac{1}{80}}$$

O lo que es lo mismo:

$$MA = \frac{1}{\frac{1}{2}\left(\frac{1}{70}+\frac{1}{80}\right)} = 74.\overline{6}$$

11. MEDIDAS DE DISPERSIÓN O VARIABILIDAD PARA DATOS NO AGRUPADOS

Las medidas de dispersión o variabilidad establecen el nivel de proximidad o separación de un conjunto de datos. En algunos casos, este distanciamiento o acercamiento se puede medir frente a su promedio o media aritmética.

Hay dos tipos de medidas de dispersión las absolutas ejemplos de estas son la varianza, la desviación estándar, el rango y el rango intercuartílico; las relativas un ejemplo es el coeficiente de variación. Las primeras se caracterizan porque tienen unidades de medida; las segundas, no.

11.1. Varianza

Muestral

La varianza muestral de un conjunto de datos es el "cuasi promedio de las desviaciones de los datos respecto a la media elevadas al cuadrado. Se calcula como:

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

Las unidades de medida de la varianza están elevadas al cuadrado.

También se puede calcular como:

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

Poblacional

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

También se puede calcular como:

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - N\mu^2}{N} = \frac{1}{N}\left(\sum_{i=1}^{n} x_i^2\right) - \mu^2$$

Demostración:

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

Como:

$$\sum_{i=1}^{n} x_i = n\bar{x} \text{ y } \sum_{i=1}^{n} \bar{x}^2 = n\bar{x}^2$$

Entonces:

$$s^2 = \frac{\sum_{i=1}^{n} x_i{}^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2}{n-1} = \frac{\sum_{i=1}^{n} x_i{}^2 - 2\bar{x}n\bar{x} + n\bar{x}^2}{n-1}$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i{}^2 - 2\bar{x}n\bar{x} + n\bar{x}^2}{n-1} = \frac{\sum_{i=1}^{n} x_i{}^2 - n\bar{x}^2}{n-1}$$

11.2. Desviación estándar (típica)

Muestral (Las unidades de la desviación son las de la variable).

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

Poblacional

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - N\mu^2}{N}} = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{n} x_i^2\right) - \mu^2}$$

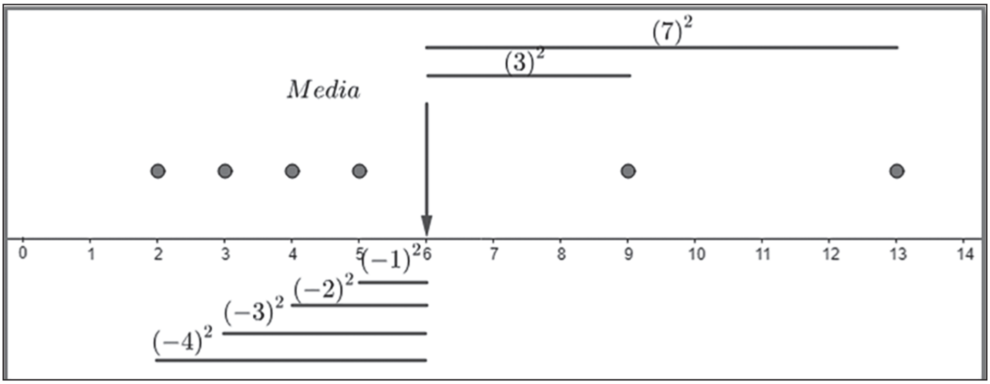
Ejemplo 2.21: para la siguiente muestra calcular la varianza y desviación estándar 2, 3, 4, 5, 9, y 13.

La media es:

$$\bar{x} = \frac{2+3+4+5+9+13}{6} = 6$$

En la Figura 63 se observan la desviación^2=(valor del dato-valor de la media)^2= $(x_i - \bar{x})^2$.

Figura 63. *Desviaciones elevadas al cuadrado.*



Fuente: elaboración propia.

La varianza muestral es:

$$s^2 = \frac{(2-6)^2 + (3-6)^2 + (4-6)^2 + (5-6)^2 + (9-6)^2 + (13-6)^2}{6-1} = \frac{88}{5} = 17.6$$

También la **varianza muestral** se puede calcular como:

$$s^2 = \frac{2^2 + 3^2 + 4^2 + 5^2 + 9^2 + 13^2 - 6(6^2)}{6-1} = \frac{304 - 216}{5} = \frac{88}{5} = 17.6$$

Ahora, calculando la desviación estándar (típica) muestral:

$$s = \sqrt{17.6} \approx 4.195$$

En RStudio la desviación estándar muestral se calcula con la función **sd((),** Figura 64.

Figura 64. *Cálculo de la desviación muestral.*

```
> x=c(2, 3, 4, 5, 9, 13)
> mean(x)
[1] 6
> sd(x)
[1] 4.195235
```

Fuente: elaboración propia.

Ejemplo 2.22: para la siguiente población calcular la varianza y desviación estándar 2, 3, 4, 5, 9, y 13.

La media poblacional es:

$$\mu = \frac{2+3+4+5+9+13}{6} = 6$$

La varianza poblacional será:

$$\sigma^2 = \frac{(2-6)^2 + (3-6)^2 + (4-6)^2 + (5-6)^2 + (9-6)^2 + (13-6)^2}{6} = \frac{88}{6} = 14.\bar{6}$$

También la varianza poblacional se puede calcular como:

$$\sigma^2 = \frac{2^2 + 3^2 + 4^2 + 5^2 + 9^2 + 13^2}{6} - 6^2 = \frac{304}{6} - 36 = 50.\bar{6} - 36 = 14.\bar{6}$$

$$\sigma = \sqrt{14.666\ldots} = 3.8227$$

En RStudio, para hacer el cálculo de la ´desviación poblacional a la desviación muestral la multiplicamos por la expresión $\left(\frac{N-1}{N}\right)^{1/2}$ donde N es el tamaño de la población como se observa en la Figura 65.

Figura 65. *Código para el cálculo de la desviación poblacional.*

```
> x=c(2, 3, 4, 5, 9, 13)
> sd(x)
[1] 4.195235
> sd(x)*((6-1)/6)^0.5
[1] 3.829708
> |
```

Fuente: elaboración propia.

Ejemplo 2.23: ahora, si se tiene la siguiente muestra 2, 3, 4, 5, 9, y 1000. Calcular la varianza muestral.

Al calcular la media:

$$\bar{x} = \frac{2+3+4+5+9+1000}{6} = 170.5$$

La varianza muestral es:

$$s^2 = \frac{(2-6)^2 + (3-6)^2 + (4-6)^2 + (5-6)^2 + (9-6)^2 + (1000-6)^2}{6-1} = \frac{825713.5}{5}$$
$$= 165142.7$$

También la varianza muestral se puede calcular como:

$$s^2 = \frac{2^2 + 3^2 + 4^2 + 5^2 + 9^2 + 1000^2 - 6(170.5^2)}{6-1} = \frac{100135 - 174421.5}{5} =$$
$$\frac{825713.5}{5} = 165142.7$$

Ahora, calculando la desviación estándar (típica) muestral:

$$s = \sqrt{17.6} \approx 406.38$$

11.3. COEFICIENTE DE VARIACIÓN

El coeficiente de variación muestral y poblacional se calcula respectivamente como:

$$CV_{muestral} = \frac{s}{\bar{x}} \times 100\%$$

$$CV_{poblacional} = \frac{\sigma}{\mu} \times 100\%$$

Es una medida de dispersión relativa que permite comparar la variabilidad de dos conjuntos de datos, incluso si tienen diferentes unidades de medida. Se calcula dividiendo la desviación estándar entre la media y multiplicando el resultado por 100 para expresarlo como porcentaje.

Ejemplo 2.24: en un grupo A, el salario promedio mensual es de $30.000.000, con desviación estándar o típica de $400.000. En el grupo B, el salario promedio es de $1´600.000 y la desviación estándar o típica de $400.000. La Tabla 9, muestra los datos del enunciado.

Tabla 9. *Estadísticas descriptivas para el grupo A y B*

Grupo	Media	Desviación estándar o típica
A	$30´000.000	$400.000
B	$1´600.000	$400.000

¿Cuál grupo presenta mayor variabilidad?

Para responder esto, usamos el coeficiente de variación:

$$CV_{grupo\ A} = \frac{\$400.000}{\$30´000.000} \times 100\% = 1.\bar{3}\%$$

$$CV_{grupo\ B} = \frac{\$400.000}{\$1´600.000} \times 100\% = 25\%$$

El grupo que presenta mayor variabilidad es el grupo B, porque el coeficiente de variación es el mayor. Obsérvese que, aunque las muestras presentan la misma desviación estándar, eso no significa que los grupos tienen la misma variabilidad, pues sus medias son diferentes.

12. EJERCICIOS

Ejercicio 1. Conteste falso o verdadero, según el caso:

a) En el siguiente conjunto de datos 2, 4, 3 y 3. La media, la mediana y la moda son iguales.
b) Si se conoce que la media de las notas de un grupo de 40 alumnos es de 3.5 y la de un segundo grupo de 10 alumnos es de 3.0, entonces el promedio aritmético de los 50 alumnos es de 3.25.
c) La mediana es una medida que no se deja influenciar por datos inusuales.
d) La medida de tendencia central que más se usa para promediar tasas de crecimiento es la mediana.
e) El valor de la desviación estándar siempre es menor que la varianza.
f) Si la varianza de un conjunto de datos es cero es porque los valores de todos los datos son iguales.
g) El peso promedio de hombres y mujeres que se encuentran en un ascensor es de 61 kg. Las mujeres pesan en promedio 55 kg y los hombres 70 kg. Hay 6 mujeres. Entonces, el número total de hombres y mujeres en el ascensor es de 10.
h) En una empresa hay dos máquinas de llenado de líquidos. La máquina A presenta un promedio de llenado de 10 litros con una desviación estándar 0.5 litros, y la máquina B presenta un promedio de llenado de 5 litros con una desviación estándar de 0.5 litros. Entonces, al comparar ambas máquinas se encuentra que presentan la misma variabilidad en el llenado de líquido.

Ejercicio 2. Considere los siguientes datos 13, 2, 10, 4, 10, 12, 16 y 12.

a) Calcule la media en RStudio.
b) Calcule la mediana en RStudio.
c) Calcule la moda en RStudio.
d) Calcule la varianza en RStudio.

Ejercicio 3. A continuación, se presentan los tiempos de espera en minutos de un grupo de personas en la parada de un autobús (el menor tiempo de espera en la muestra es de 3.0 minutos).

	Tallo	**Hoja**
(2)	3	01
(3)	4	588
(10)	5	0345789999
(4)	6	1479
(3)	7	369
(3)	8	035

a) Calcular el tiempo promedio y la desviación estándar del tiempo de espera del autobús.

b) ¿En qué rango de valores de los tiempos de espera se encuentran el 50% inferior?

Ejercicio 4. A continuación, se presenta una muestra de 11 localidades de Bogotá (Tabla10), y se indica además su superficie en km^2.

Tabla 10. *Muestra de localidades de Bogotá*

Localidad	Usaquén	Chapinero	Santa Fe	San Cristóbal	Tunjuelito	Usme
Km^2	65	38	45	49	10	215
Localidad	Bosa	Kennedy	Fontibón	Los Mártires	Engativá	
Km^2	24	39	33	7	36	

a) Calcule la superficie promedio en km^2 para esta muestra.

b) ¿Cuál es la localidad con superficie mediana en esa muestra?

c) Calcule la desviación estándar para la superficie.

d) Calcule el rango de la superficie en esa muestra.

Ejercicio 5. A continuación, se presenta una tabla de un grupo de estudiantes a los que se les pregunto el número de hermanos, Tabla 11.

Tabla 11. *Distribución del número de hermanos*

Número de hermanos	Frecuencia absoluta f_i
0	4
1	10
2	6
3	5
4	4
5	2

Calcule el número promedio de hermanos en esa muestra con el promedio ponderado.

Ejercicio 6. A continuación, se presentan las estadísticas descriptivas (Tabla12) de muestras aleatorias sobre las mediciones de los diámetros de unas esferas medidas con los instrumentos 1 (en centímetros) y 2 (en pulgadas).

Tabla 12. *Estadísticas descriptivas para los Instrumentos 1 y 2*

Media	**Instrumento 1**	**Instrumento 2**
Media aritmética	10.8 cm	4.3 in
Desviación estándar	2.7 cm	1.1 in
Tamaño de muestra	30	45

De acuerdo con la información, ¿cuál instrumento presenta mayor variabilidad en su medición?

13. MEDIDAS DE TENDENCIA CENTRAL PARA DATOS AGRUPADOS

13.1. Media

Se calcula de la siguiente forma:

$$\bar{x} = \frac{\sum_{i=1}^{l} m_i f_i}{\sum_{i=1}^{l} f_i} = \frac{\sum_{i=1}^{l} m_i f_i}{n}$$

Es decir, se calcula cómo la media ponderada y en este caso f_i serán los pesos, donde:

m_i =Marca de clase del i-ésimo intervalo.

f_i =Frecuencia absoluta del i-ésimo intervalo.

l =Total de intervalos.

n =Total de datos en la muestra; $n = \sum_{i=1}^{l} f_i$

13.2. Moda

Es la marca de clase (m_i) del intervalo de mayor frecuencia absoluta.

13.3. Mediana

Se calcula de la siguiente forma:

1. Se calcula la posición $(n+1)0.5 = k$. Donde n es el total de datos.
2. Se determina el intervalo donde se encuentra la posición k. El intervalo se representa como:

$$[L_{Ii}, L_{Si})$$

3. Se calcula la mediana con la siguiente fórmula:

$$\tilde{x} = x_k = L_{Ii} + \left(\frac{L_{Si} - L_{Ii}}{f_i}\right)[k - (F_{i-1} + 1)]$$

Donde:

k =Posición de la mediana.

L_{Ii} =Límite inferior del i-ésimo intervalo donde cae la mediana.

L_{Si} =Límite superior del i-ésimo intervalo donde cae la mediana.

F_{i-1} =Frecuencia absoluta acumulada del $(i-1)$-ésimo intervalo, es decir, el intervalo anterior a donde cae la mediana.

f_i =Frecuencia absoluta del i-ésimo intervalo donde cae la mediana.

Ejemplo 2.25: la variable que se presenta en la Tabla 13 (datos agrupados) es el número de días en que una empresa de envíos tarda en entregar 80 paquetes solicitados. Calcular la media, la mediana y la moda de días que tardan los envíos en esa muestra.

Tabla 13. *Tabla de frecuencias absolutas, relativas y acumuladas para la variable días*

Intervalos de clase (días)	Marca de clase (m_i)	Frecuencia absoluta (f_i)	Frecuencia absoluta acumulada (F_i)	Frecuencia relativa (h_i)	Frecuencia relativa acumulada (H_i)
[5,10)	7.5	10	10	0.125	0.125
[10,15)	12.5	12	22	0.150	0.275
[15,20)	17.5	24	46	0.300	0.575
[20,25)	22.5	28	74	0.350	0.925
[25,30)	27.5	6	80	0.075	1

Media $\bar{x} = \frac{(7.5)(10)+(12.5)(12)+(17.5)(24)+(22.5)(28)+(27.5)(6)}{80} = \frac{1440}{80} = 18$ *días.*

Moda $Mo = 22.5$ *días.*

Mediana

1. Encontramos la posición. $k = (80+1)(0.5) = 40.5$.
2. Se determina el intervalo donde se encuentra la posición k. El intervalo es [15,20).
3. Se calcula la mediana con la siguiente fórmula:

$$\tilde{x} = x_k = L_{Ii} + \left(\frac{L_{Si} - L_{Ii}}{f_i}\right)[k - (F_{i-1} + 1)]$$

$$\tilde{x} = x_{40.5} = 15 + \left(\frac{20-15}{24}\right)[40.5 - (22+1)] = 18.6458\bar{3}$$

La mediana es aproximadamente 18.6 días.

14. MEDIDAS DE DISPERSIÓN O VARIABILIDAD PARA DATOS AGRUPADOS

14.1. VARIANZA MUESTRAL

$$s^2 = \frac{\sum_{i=1}^{l}(m_i - \bar{x})^2 f_i}{\sum_i^l f_i - 1} = \frac{\sum_{i=1}^{l}(m_i - \bar{x})^2 f_i}{n-1}$$

m_i =Marca de clase del i-ésimo intervalo.

f_i =Frecuencia absoluta del i-ésimo intervalo.

l =Total de intervalos.

n =Total de datos en la muestra; $n = \sum_{i=1}^{l} f_i$.

$\bar{x}$= Media muestral.

También la varianza muestral se puede calcular como:

$$s^2 = \frac{\left(\sum_{i=1}^{l} m_i^2 f_i\right) - n\bar{x}^2}{n-1} = \frac{1}{n-1}\left(\left(\sum_{i=1}^{l} m_i^2 f_i\right) - n\bar{x}^2\right)$$

14.2. Desviación estándar muestral

$$s = \sqrt{\frac{\sum_{i=1}^{l}(m_i - \bar{x})^2 f_i}{n-1}} = \sqrt{\frac{\left(\sum_{i=1}^{l} m_i^2 f_i\right) - n\bar{x}^2}{n-1}}$$

Donde

$\bar{x}$=Media aritmética de datos agrupados.

f_i =Frecuencia absoluta del i-ésimo intervalo.

m_i =Marca de clase del i-ésimo intervalo.

l=Número de intervalos.

Ejemplo 2.26: se tienen los datos del Ejemplo 2.25, Tabla 13. La variable representa el número de días en que una empresa de envíos tarda en entregar 80 paquetes solicitados. Calcular la varianza y la desviación estándar.

La **media**, calculada en el Ejemplo 2.25 es $\bar{x} = 18$ días.

La varianza muestral será:

$$s^2 = \frac{(7.5-18)^2 10 + (12.5-18)^2 12 + (17.5-18)^2 24 + (22.5-18)^2 28 + (27.5-18)^2 6}{80-1}$$

O también la varianza muestral será:

$$s^2 = \frac{(7.5^2)(10) + (12.5^2)(12) + (17.5^2)(24) + (22.5^2)(28) + (27.5^2)(6) - 80(18^2)}{80 - 1}$$

$$s^2 = \frac{28500 - 25920}{79} = \frac{2580}{79} \approx 32.658 \quad dias^2$$

La desviación estándar es:

$s = \sqrt{32.658} \approx 5.715$ días.

Varianza poblacional:

$$\sigma^2 = \frac{\sum_{i=1}^{l}(m_i - \mu)^2 f_i}{\sum_i^l f_i} = \frac{\sum_{i=1}^{l}(m_i - \mu)^2 f_i}{N}$$

También la varianza poblacional se puede calcular como:

$$\sigma^2 = \frac{\left(\sum_{i=1}^{l} m_i^2 f_i\right) - N\mu^2}{N} = \frac{1}{N}\left(\sum_{i=1}^{l} m_i^2 f_i\right) - \mu^2$$

Desviación estándar muestral:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{l}(m_i - \mu)^2 f_i}{N}} = \sqrt{\frac{\left(\sum_{i=1}^{l} m_i^2 f_i\right)}{N} - \mu^2}$$

m_i =Marca de clase del i-ésimo intervalo.

f_i =Frecuencia absoluta del i-ésimo intervalo.

l =Total de intervalos.

N =Total de datos en la población $N = \sum_{i=1}^{l} f_i$.

μ= Media muestral.

15. PROPIEDADES DE LA MEDIA Y DE LA VARIANZA

15.1. Propiedades de la media

Ejemplo 2.27: se tiene la siguiente muestra 2, 3 y 7. Calcular la media.

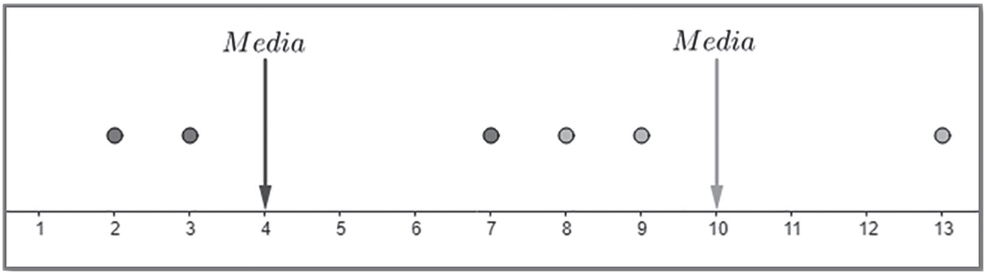
$$\bar{x} = \frac{2+3+7}{3} = \frac{12}{3} = 4$$

- Ahora, si a cada dato le sumo el número real 6, la nueva muestra de datos será 8, 9 y 13. La media para este nuevo conjunto de datos será:

$$\bar{x} = \frac{8+9+13}{3} = \frac{30}{3} = 10$$

- ¿Qué pasó con la media del nuevo conjunto de datos? Como cada uno de los datos se desplazó 6 unidades a la derecha, entonces la nueva media se desplaza 6 unidades a la derecha. Por lo tanto, la media del conjunto nuevo de datos es la media anterior más 6 unidades ($4+6=10$). Esto se puede observar en la Figura 66.

Figura 66. *Nueva media desplazada 6 unidades a la derecha.*



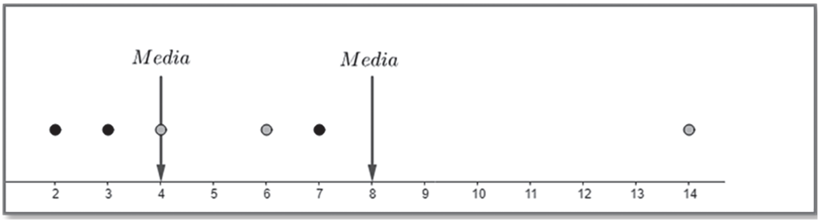
Fuente: elaboración propia.

- Ahora, si a cada dato del ejemplo lo multiplico por el número real 2, la nueva muestra de datos será 4, 6 y 14, la media para este nuevo conjunto de datos será:

$$\bar{x} = \frac{4+6+14}{3} = \frac{24}{3} = 8$$

- ¿Qué pasó con la media del nuevo conjunto de datos? Como cada uno de los datos se multiplicó por 2, entonces la nueva media resulta de multiplicar la media anterior por 2. Por lo tanto, la media del conjunto nuevo de datos es la media anterior multiplicada por 2, es decir, 4x2=8. Esto se puede observar en la Figura 67.

Figura 67. *Nueva media multiplicada por 2 unidades.*



Fuente: elaboración propia.

En general:

Dado un conjunto de datos $x_1, x_2, \dots, x_n$ con media $\bar{x}$.

- Si a cada dato se le suma un número real c, el nuevo conjunto de datos será:

$$x_1 + c,\ x_2 + c, \dots, x_n + c$$

Con media $\bar{x} + c$.

- Si a cada dato se le multiplica por un número real c, el nuevo conjunto de datos será:

$$x_1 \times c,\ x_2 \times c, \dots, x_n \times c$$

Con media $\bar{x} \times c$.

15.2. Propiedades de la varianza

Ejemplo 2.28: se tiene la siguiente muestra 2, 3 y 7. Calcular la varianza.

Primero calculamos la media:

$$\bar{x} = \frac{2 + 3 + 7}{3} = \frac{12}{3} = 4$$

Calculando la varianza muestral:

$$s^2 = \frac{(2-4)^2 + (3-4)^2 + (7-4)^2}{3-1} = \frac{(-2)^2 + (-1)^2 + (3)^2}{2} = \frac{14}{2} = 7$$

- Ahora, si a cada dato le sumo el número real 6, la nueva muestra de datos será 8, 9 y 13. La media para este nuevo conjunto de datos será:

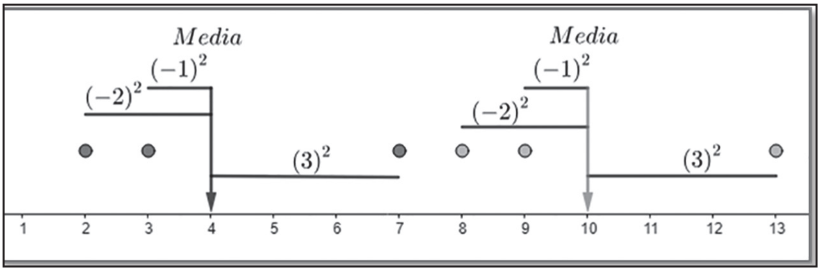
$$\bar{x} = \frac{8+9+13}{3} = \frac{30}{3} = 10$$

Con varianza:

$$s^2 = \frac{(8-10)^2+(9-10)^2+(13-10)^2}{3-1} = \frac{(-2)^2+(-1)^2+(3)^2}{2} = \frac{14}{2} = 7$$

- ¿Qué pasó con la varianza del nuevo conjunto de datos? Como cada uno de los datos se desplazó 6 unidades a la derecha, entonces los datos se desplazan a la derecha, por lo tanto, la varianza se mantiene igual. Esto se puede observar en la Figura 68, que muestra las desviaciones elevadas al cuadrado que son iguales.

Figura 68. *Valores de las desviaciones al cuadrado de los dos grupos son iguales.*



Fuente: elaboración propia.

- Ahora, si a cada dato del ejemplo lo multiplico por el número real 2, la nueva muestra de datos será 4, 6 y 14. La media para este nuevo conjunto de datos será:

$$\bar{x} = \frac{4+6+14}{3} = \frac{24}{3} = 8$$

Con varianza:

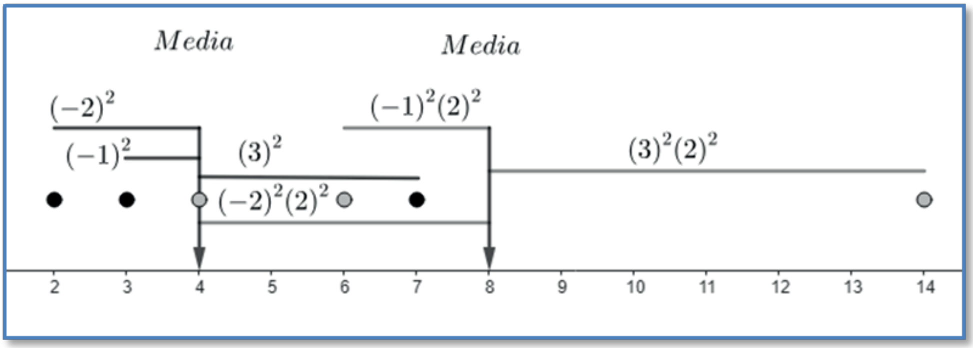
$$s^2 = \frac{(4-8)^2+(6-8)^2+(14-8)^2}{3-1} = \frac{(2)^2(2-4)^2+(2)^2(3-4)^2+(2)^2(7-4)^2}{2}$$

$$= \frac{(2)^2(-2)^2+(2)^2(-1)^2+(2)^2(3)^2}{2} = 2^2 \times \frac{[(-2)^2+(-1)^2+(3)^2]}{2} = 2^2 \times \frac{14}{2}$$

$$= 2^2 \times 7$$

- ¿Qué pasó con la varianza del nuevo conjunto de datos? Como cada uno de los datos se multiplicó por 2, entonces la nueva varianza resulta de multiplicar la varianza anterior por 2^2. Por lo tanto, la varianza aumenta para este caso. Esto se puede ver en la Figura 69, donde las desviaciones al cuadrado aumentan para la muestra donde los datos se han multiplicado por 2.

Figura 69. *Valores de las desviaciones al cuadrado de los dos grupos son diferentes.*



Fuente: elaboración propia.

En general:

Dado un conjunto de datos $x_1, x_2, \dots, x_n$ con varianza s^2.

- Si a cada dato se le suma un número real c, el nuevo conjunto de datos será

$$x_1 + c,\ x_2 + c, \dots, x_n + c$$

Con varianza s^2.

- Si a cada dato se le multiplica por un número real c, el nuevo conjunto de datos será

$$x_1 \times c,\ x_2 \times c, \dots, x_n \times c$$

Con varianza $s^2 c^2$.

- Nota obsérvese que si hubiésemos multiplicado cada uno de los datos por números reales entre -1 y 1, la varianza de los nuevos datos disminuiría.

16. MEDIDAS DE POSICIÓN

16.1. Cuartiles

La mediana $\tilde{x}$ es una medida de posición, porque divide la distribución de datos ordenados en dos partes iguales. Por debajo de la mediana se encuentra el 50% de los datos y por encima el otro 50%. La fórmula para encontrar la posición de la mediana es:

$$k = (n+1) * 0.5 = \frac{(n+1)}{2}$$

Otras medidas de posición son los cuartiles. Éstos dividen la distribución de datos ordenados en cuatro partes iguales, en este caso hay tres cuartiles q_1, q_2 y q_3. Para el cuartil 1 q_1, el porcentaje de datos que se encuentran por debajo de él es el 25% y por encima es el 75%. Para el cuartil 2 q_2 (coincide con la mediana), el porcentaje de datos que se encuentran por debajo de él es el 50% y por encima es el 50%. Para el cuartil 3 q_3, el porcentaje de datos que se encuentran por debajo de él es el 75% y por encima es el 25%. La fórmula para encontrar el $i - ésimo$ cuartil es:

$$k_i = \frac{i * (n+1)}{4}$$

El diagrama de caja y bigotes es una representación visual de los cuartiles y de los valores máximo y mínimo de un conjunto de datos para describir su dispersión y simetría. La caja abarca desde el primer cuartil (q_1) hasta el tercer cuartil (q_3), conteniendo el 50% de los datos centrales y dentro de la caja se visualiza por medio de una línea el cuartil dos (q_2) o mediana. Los bigotes son líneas que se extienden desde los extremos de la caja hasta el valor mínimo y el valor máximo que no se consideran atípicos. Los pasos para hacer un diagrama de caja y bigotes son

1. Encontrar los cuartiles y representarlos en una caja rectangular.
2. Encontrar el rango intercuartílico $RI = q_3 - q_1$.
3. Encontrar intervalos para identificar datos inusuales suaves y extremos.

3.1. Intervalo 1 $(q_1 - 1.5 * RI, q_3 + 1.5RI)$. Si hay datos que están por fuera de este intervalo se dice que es un dato inusual suave.

3.2. Intervalo 2 $(q_1 - 3 * RI, q_3 + 3RI)$. Si hay datos que están por fuera de este intervalo se dice que es un dato inusual extremo.

4. Dibujar el diagrama con los bigotes, que se representan con los valores del dato menor a la caja y de la caja al dato mayor, siempre y cuando estos valores mínimo y máximo no sean datos inusuales extremos. En caso contrario, los bigotes cubrirían el primer intervalo.

Ejemplo 2.29: los datos siguientes, pesos en kilogramos (kg), de dos muestras de estudiantes de colegio el Grupo 1 y el Grupo 2.

Grupo 1	43	44	44	45	45	46	46	47	48	48
	49	49	49	50	50	50	51	51	52	52
Grupo 2	38	38	38	39	39	39	40	40	40	40
	41	41	41	41	42	42	42	42	42	43

Haga en una sola escala el diagrama de cajas para las dos muestras.

Para hacer un diagrama de caja y bigotes encontramos los cuartiles, pero teniendo en cuenta que los datos deben estar ordenados, en nuestro ejemplo los datos están ordenados de menor a mayor.

Encontramos la posición de los cuartiles teniendo en cuenta la cantidad de datos, que en este ejemplo son 20, lo mismo para grupo 1 y grupo 2:

Posición del cuartil 1 $k_1 = \frac{1*(n+1)}{4} = \frac{1*(20+1)}{4} = 5.25$

Posición del cuartil 2 $k_2 = \frac{2*(n+1)}{4} = \frac{2*(20+1)}{4} = 10.5$

Posición del cuartil 3 $k_3 = \frac{3*(n+1)}{4} = \frac{3*(20+1)}{4} = 15.75$

Ahora, encontramos los cuartiles, para el grupo 1 y el grupo 2:

Grupo 1

$q_1 = 45 * 0.75 + 46 * 0.25 = 45.25$;

$q_2 = 48 * 0.5 + 49 * 0.5 = 48.5$;

$q_3 = 50 * 0.25 + 50 * 0.75 = 50$.

Grupo 2

$$q_1 = 39 * 0.75 + 39 * 0.25 = 39;$$

$$q_2 = 40 * 0.5 + 41 * 0.5 = 40.5;$$

$$q_3 = 42 * 0.25 + 42 * 0.75 = 42.$$

En RStudio se ingresan los datos, Figura 70.

Figura 70. *Ingreso de datos Grupo 1 y 2.*

```
> grupo_1=c(43,44,44,45,45,46,46,47,48,48,49,49,49,50,50,50,51,51,52,
52)
> grupo_2=c(38,38,38,39,39,39,40,40,40,40,41,41,41,41,42,42,42,42,42,
43)
```

Fuente: elaboración propia.

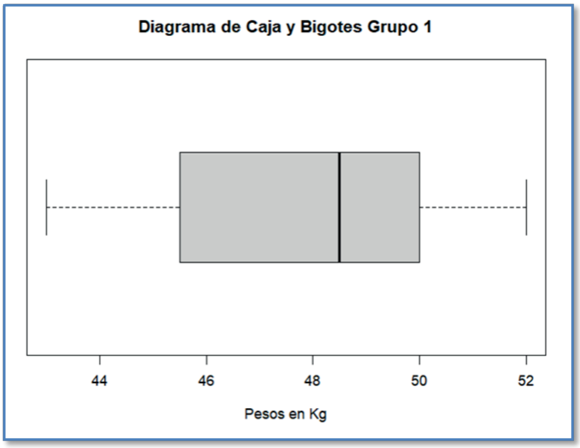
Ahora, usamos la función **boxplot()** para graficar un diagrama de caja y bigotes. Las Figuras 71 y 73 muestran el código en RStudio, y las Figuras 72 y 74 muestran la gráfica para el grupo 1 y grupo 2, respectivamente.

Figura 71. *Código diagrama de caja grupo 1.*

```
> boxplot(grupo_1, horizontal = TRUE,xlab ='Pesos en Kg',main
='Diagrama de Caja y Bigotes Grupo 1')
```

Fuente: elaboración propia.

Figura 72. *Diagrama de caja Grupo 1.*



Fuente: elaboración propia.

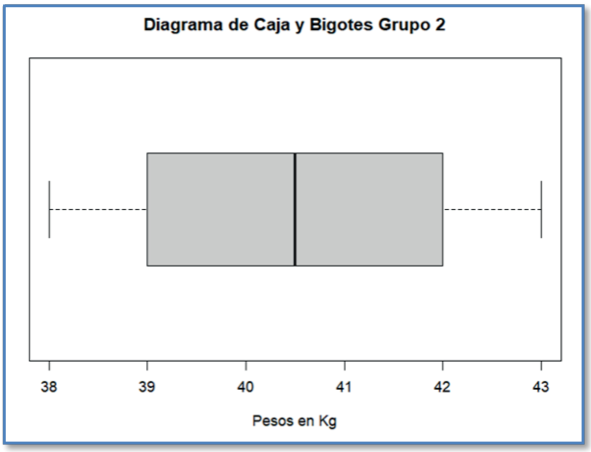
Ahora se tiene para el grupo 2:

Figura 73. *Código diagrama de caja grupo 2.*

```
> boxplot(grupo_1, grupo_2, names = c("Grupo 1", "Grupo 2"),
main = "Diagrama de Caja y Bigotes, Grupo 1 y Grupo 2", xlab
= "Grupos", ylab = " ")
```

Fuente: elaboración propia.

Figura 74. *Diagrama de caja Grupo 2.*



Fuente: elaboración propia.

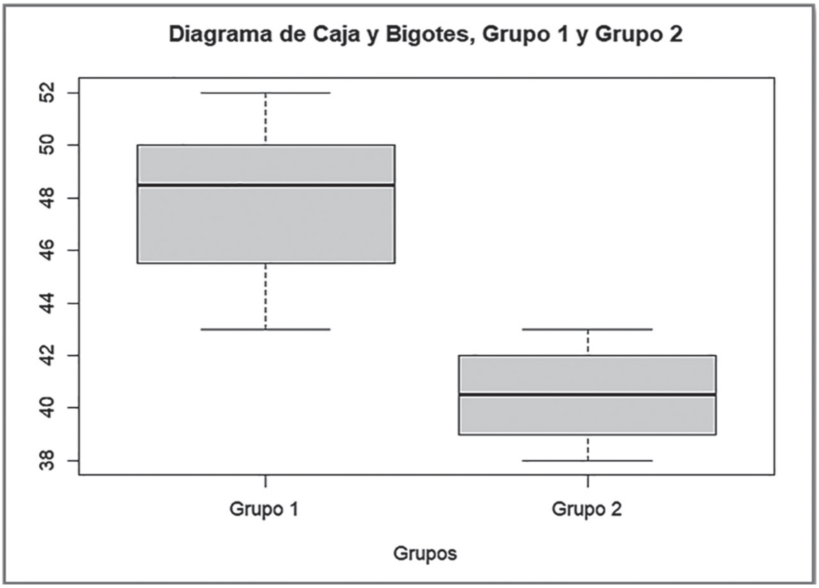
Ahora, se tiene para el grupo 1 y 2 en el mismo gráfico, el código en la Figura 75 y el diagrama en la Figura 76.

Figura 75. *Código del Diagrama de Caja para ambos grupos.*

```
> boxplot(grupo_1, grupo_2, names = c("Grupo 1", "Grupo 2"),
main = "Diagrama de Caja y Bigotes, Grupo 1 y Grupo 2", xlab
= "Grupos", ylab = " ")
```

Fuente: elaboración propia.

Figura 76. *Diagrama de caja Grupo 1 y Grupo 2.*



Fuente: elaboración propia.

1. Encontrar el rango intercuartílico $RI = q_3 - q_1$.

Grupo 1 $RI = q_3 - q_1 = 50 - 45.5 = 4.5$

Grupo 2 $RI = q_3 - q_1 = 42 - 39 = 3$

2. Encontrar intervalos para identificar datos inusuales suaves y extremos.

Grupo 1

- Intervalo 1 $(45.5 - 1.5 * 4.5, 50 + 1.5 * 4.5) = (38.75, 56.75)$. No hay datos que estén por fuera de este intervalo. Entonces, no hay datos inusuales suaves.
- Intervalo 2$(45.5 - 3 * 4.5, 50 + 3 * 4.5) = (32, 63.5)$ No hay datos que estén por fuera de este intervalo. Entonces, no hay datos inusuales extremos.

Grupo 2

- Intervalo 1 $(39 - 1.5 * 3, 42 + 1.5 * 3) = (34.5, 46.5)$. No hay datos que estén por fuera de este intervalo. Entonces, no hay datos inusuales suaves.
- Intervalo 2 $(39 - 3 * 3{,}42 + 3 * 3) = (30, 51)$ No hay datos que estén por fuera de este intervalo. Entonces, no hay datos inusuales extremos.

a) ¿Qué grupo presenta los pesos más altos? Grupo 1, porque esto se puede ver con el cuartil 2, para el grupo 1 es el mayor.

b) ¿En qué grupo hay menos variabilidad? Grupo 2, porque el rango intercuartílico es el mayor que en el Grupo 1.

16.2. Deciles, percentiles entre otros

Otras medidas de posición son los deciles. Éstos dividen la distribución de datos ordenados en diez partes iguales. En este caso hay nueve deciles d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8 y d_9. Para el decil 1 d_1, el porcentaje de datos que se encuentran por debajo de él es el 10%, y por encima el 90%. Así, sucesivamente para los demás deciles. La fórmula para encontrar el $i - ésimo$ decil es:

$$k_i = \frac{i * (n + 1)}{10}$$

Por último, los percentiles dividen la distribución de datos ordenados en cien partes iguales. En este caso, hay noventa y nueve percentiles p_1, p_2, $p_3, \dots, p_{99}$,. Para el percentil 1 p_1, el porcentaje de datos que se encuentran por debajo de él es el 1% y por encima el 99%. Así sucesivamente para los demás percentiles. La fórmula para encontrar el $i - ésimo$ percentil es:

$$k_i = \frac{i * (n+1)}{100}$$

Ejemplo 2.30: para los datos del grupo 1 del Ejemplo 2.29, encontrar el decil 8 y el percentil 65.

Decil 8

Posición decil 8 $k_8 = \frac{8*(n+1)}{10} = \frac{8*(20+1)}{10} = 16.8$

Decil 8 $d_8 = 50 * (1 - 0.8) + 51 * 0.8 = 50.8$

Percentil 65

Posición percentil 65 $k_{65} = \frac{65*(n+1)}{100} = \frac{65*(20+1)}{10} = 13.65$

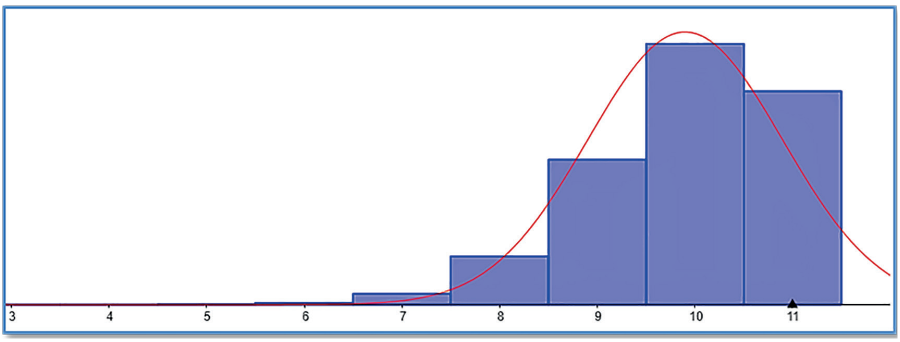
Percentil 65 $p_{65} = 49 * (1 - 0.65) + 50 * 0.65 = 49.65.$

17. MEDIDAS DE ASIMETRÍA Y CURTOSIS

17.1. Medidas de asimetría

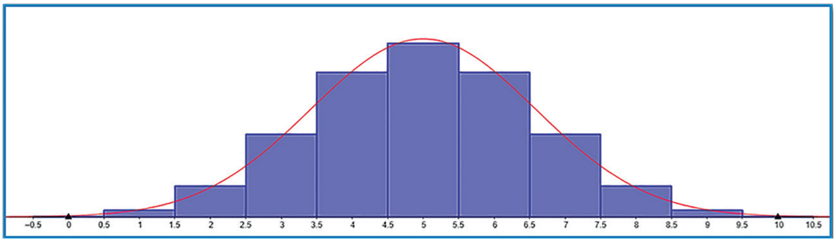
La asimetría es una medida que indica la falta de simetría en la distribución de un conjunto de datos respecto a la media. Hay tres tipos de asimetría **asimetría negativa o sesgo a la izquierda** (la cola de la distribución se alarga para los valores inferiores a la media, Figura 77), **simétrica** (hay un mismo número de elementos a izquierda y a derecha de la media, Figura 78), **asimetría positiva o sesgo a la derecha** (la cola de la distribución se alarga para los valores superiores a la media, Figura 79).

Figura 77. *Distribución con sesgo negativo.*



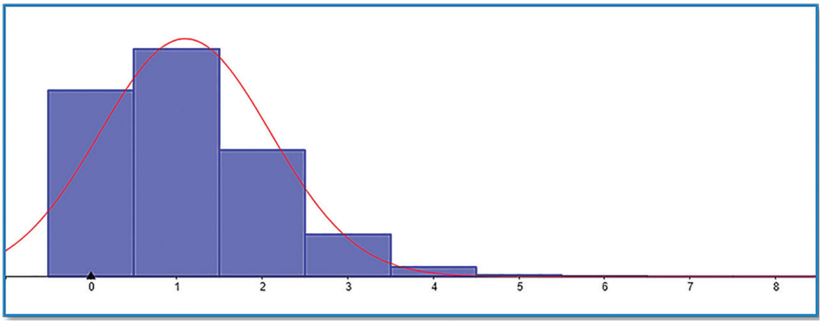
Fuente: elaboración propia.

Figura 78. *Distribución simétrica.*



Fuente: elaboración propia.

Figura 79. *Distribución con sesgo positivo.*



Fuente: elaboración propia.

- **Coeficiente de asimetría de Fisher para una muestra con datos no agrupados**

$$CA = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{ns^3}$$

- **Coeficiente de asimetría de Fisher para una muestra con datos agrupados**

$$CA = \frac{\sum_{i=1}^{l}(m_i - \bar{x})^3 * f_i}{ns^3}$$

- **Criterios**

Si $CA < 0$, la distribución tiene una **asimetría negativa**.

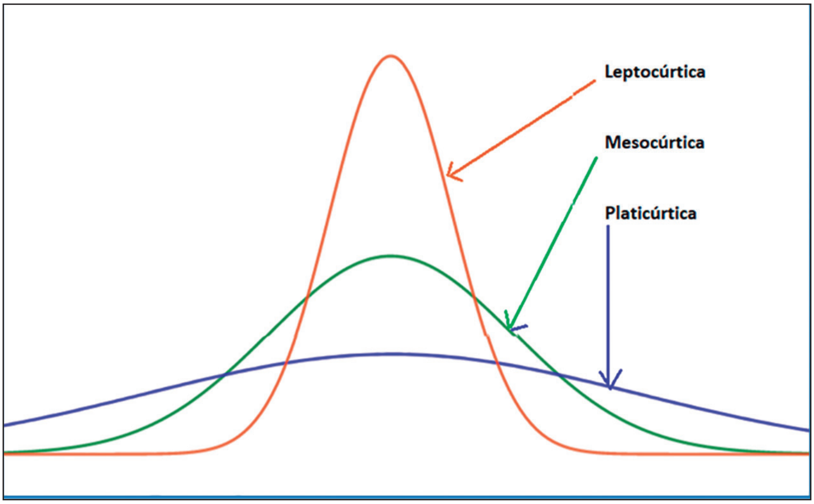
Si $CA = 0$, la distribución tiene es **simétrica.**

Si $CA > 0$, la distribución tiene una **asimetría positiva.**

17.2. Curtosis

La curtosis es una medida que indica cuán aplanada o empinada es la distribución de los datos con respecto a la distribución gaussiana. Hay tres tipos de curtosis, que se muestran en la Figura 80.

Figura 80. *Tipos de curtosis.*



Fuente: elaboración propia.

- **Coeficiente de curtosis para una muestra con datos no agrupados**

$$g = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{ns^4} - 3$$

- **Coeficiente de asimetría de Fisher para una muestra con datos agrupados**

$$g = \frac{\sum_{i=1}^{l}(m_i - \bar{x})^4 * f_i}{ns^4} - 3$$

- **Criterios**

Si $\boldsymbol{g < 0}$, la distribución es una distribución **platicúrtica**.

Si $\boldsymbol{g = 0}$, la distribución es una distribución **mesocúrtica.**

Si $\boldsymbol{g > 0}$, la distribución es una distribución **leptocúrtica.**

18. EJERCICIOS

Ejercicio 1. Conteste Falso o Verdadero

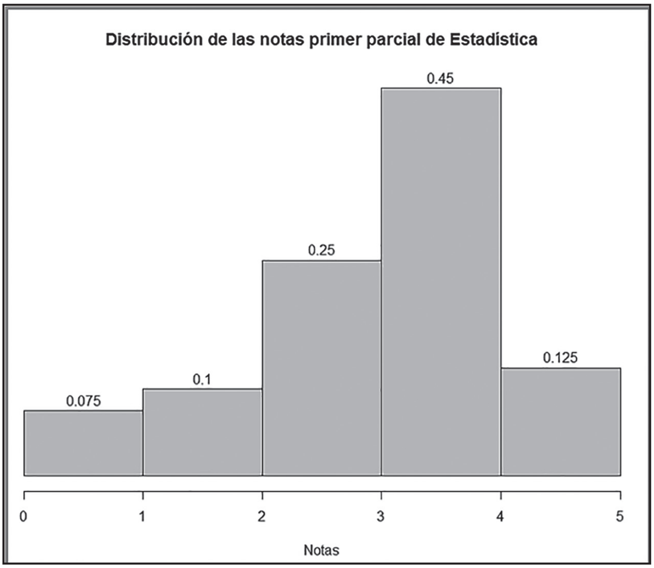
a) El sueldo promedio de una empresa es de $1.800.000, con desviación de $500.000. El dueño decide incrementar para el siguiente año el 4% en los salarios a cada uno de sus empleados. Entonces, la desviación estándar para el siguiente año será la misma de $500.000.

b) Un profesor de estadística tiene un grupo de 30 estudiantes. La nota promedio del primer parcial en este grupo fue de 3.2, y la desviación estándar de 0.7. El profesor decide aumentar 0.2 décimas a cada uno de sus estudiantes en su nota. Entonces, la nueva desviación estándar de las notas se mantiene igual a 0.7.

c) Si el promedio del sueldo de los empleados de una empresa es de $1´320.000 y la desviación de $400.000, si el dueño decide aumentar a cada uno de sus empleados el 5% de su sueldo, entonces el coeficiente de variación es el mismo para antes y después del aumento.

Ejercicio 2. Dado el siguiente histograma, con frecuencias relativas, que representan las notas del primer parcial de un grupo de 160 estudiantes de Estadística I, Figura 81.

Figura 81. *Distribución de las notas.*



Fuente: elaboración propia.

a) Calcule la media de las calificaciones del grupo.

b) ¿Por debajo de qué nota se encuentra el 50% de los estudiantes con menores calificaciones?

c) Calcule la desviación estándar de las calificaciones del curso.

Ejercicio 3. En un estudio que se hizo a una muestra o grupo de personas en una ciudad sobre los precios (en dólares) de los platos o comidas ofrecidos en ciertos restaurantes de un área metropolitana se presentaron los siguientes resultados:

- El precio de cada plato de comida que se paga en esa muestra oscila entre 10 dólares (inclusive) y 50 dólares.
- 76 personas, que era el 38%, contestó que los precios que pagaron por cada plato de comida fueron menores a 20 dólares.
- El 80% de las personas respondieron que cada plato que pagó tenía un precio menor a 30 dólares.
- El 6.5% de las personas manifestaron que consumieron platos de comida cuyo precio estaba comprendido entre 40 dólares (inclusive) y 50 dólares.

  a) Haz una tabla de frecuencias donde aparezcan los intervalos de clase con los precios, marcas de clase m_i, frecuencias absolutas f_i, frecuencias absolutas acumuladas F_i, frecuencias relativas h_i, frecuencias relativas acumuladas H_i. (especificando los pasos en la construcción).

  b) Halle la mediana del valor del plato en la muestra.

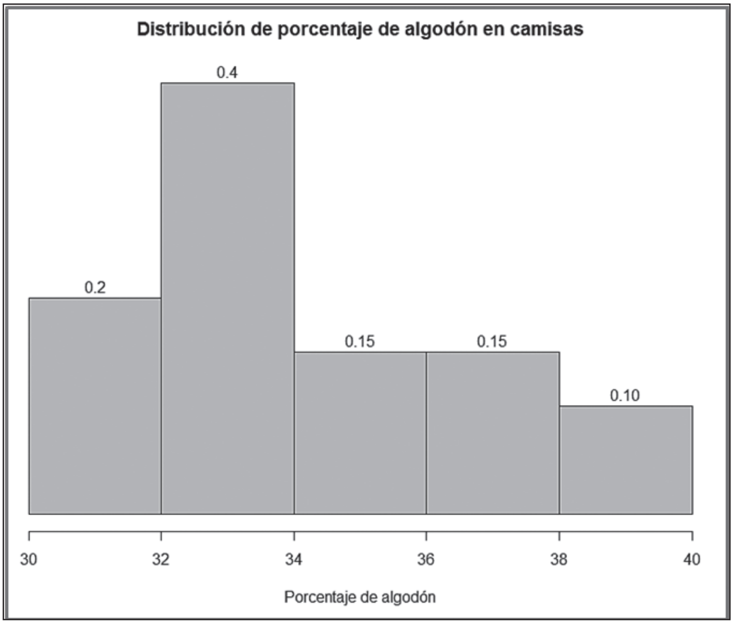
  c) Halle el precio promedio de los platos de la muestra.

  d) Halle la desviación estándar del valor del plato.

  e) ¿Qué porcentaje de estas personas pagó por un plato 20 dólares o más, pero menor a 40 dólares?

Ejercicio 4. A continuación, se presenta la distribución de frecuencias relativas del porcentaje de algodón usado en cada camisa fabricada para caballeros. En total la muestra fue de 200 camisas, Figura 82.

Figura 82. *Distribución de porcentaje de algodón en las camisas.*



Fuente: elaboración propia.

a) ¿En promedio, qué porcentaje de algodón se usa en la fabricación de una camisa?

b) ¿Cuál es el porcentaje de algodón para que la mitad de las camisas en esta muestra estén por debajo de este valor?

c) ¿Cuál es la varianza en el porcentaje de algodón?

Ejercicio 5. A continuación, se presentan la distribución de frecuencias (Tabla 14) relativas acumuladas H_i de los salarios mensuales (en millones) de 400 empleados de una empresa.

Tabla 14. *Distribución de salarios mensuales*

Salario mensual	Frecuencia relativa acumulada $\mathbf{H_i}$
[1.5, 2)	0.40
[2, 2.5)	0.66
[2.5, 3)	0.80

[3, 3.5)	0.90
[3.5, 4)	0.98
[4, 4.5)	1.00

a) Calcular el sueldo promedio de este grupo de empleados.

b) Calcular la varianza de los sueldos.

c) En qué rango de valores se encuentra el salario de los empleados del 60% superior.

Ejercicio 6. Las auditorías contables (en días) de 20 clientes de una organización se presentan en la Tabla 15, de frecuencias relativas.

Tabla 15. *Distribución de días de las auditorías*

Número de días	Frecuencia absoluta f_i	Frecuencia relativa h_i
[10, 15)		0.20
[15, 20)		0.40
[20, 25)		0.25
[25, 30)		0.10
[30, 35)		0.05

a) Complete los valores de la frecuencia absoluta si en total se hicieron 20 auditorías.

b) Calcular el tiempo promedio de una auditoría.

c) ¿Por debajo de qué valor se encuentran el 50% de los menores tiempos en las auditorías?

d) Calcular el tiempo modal de una auditoría.

e) Calcular la desviación estándar del tiempo de las auditorías.

Ejercicio 7. Los datos presentados a continuación representan las ventas de 21 empresas de automóviles durante un año en millones de euros.

8400	1300	1800	8800	2400	11400
600	21000	6400	1800	2800	1300
10400	7400	4000	4300	700	2100
3600	5700	8300			

a) Trace un diagrama de caja.

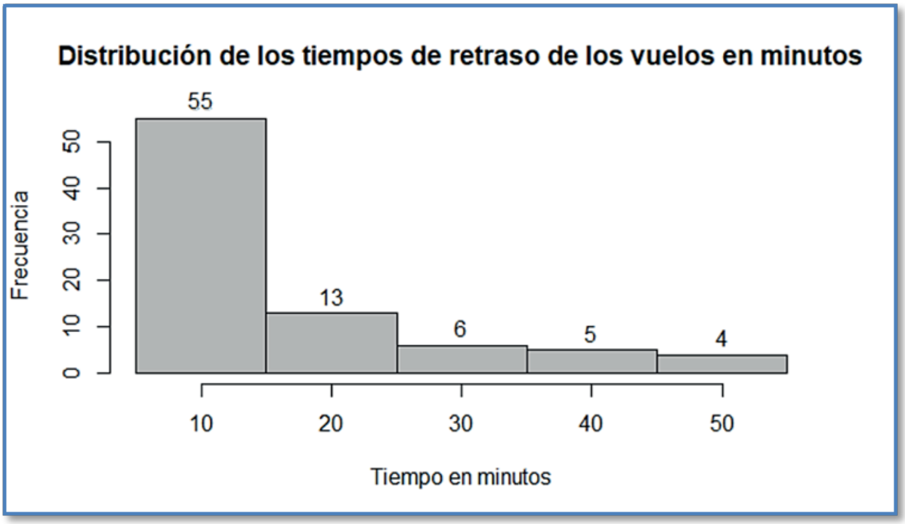
b) ¿Hay algún valor atípico o algún valor atípico suave o extremo en la muestra? Justifique la respuesta.

c) Encuentre el percentil 90.

d) Encuentre el decil 3

Ejercicio 8. Un importante aeropuerto ha contratado recientemente a un consultor para estudiar el problema de los retrasos en el tráfico aéreo, Figura 83.

Figura 83. *Distribución del tiempo de retraso de vuelos en minutos.*



Fuente: elaboración propia.

a) Calcule la media muestral del tiempo de retraso.

b) Calcule la desviación estándar muestral del tiempo de retraso.

c) ¿Qué tiempo corresponde al 50% de los vuelos con menor retraso?

d) ¿Qué tiempo corresponde al 80% de los vuelos con mayor retraso?

e) Halle el coeficiente de asimetría y curtosis, ¿qué concluye?

Capítulo 3. Conceptos básicos de la probabilidad

1. TEORÍA BÁSICA DE CONJUNTOS

Definición informal de un conjunto un conjunto es una colección de elementos u objetos bien definidos.

Los conjuntos se denotan con letras mayúsculas $A, B, C, D.$

Los elementos de un conjunto se denotan con letras minúsculas $x, y, z.$

Un conjunto se puede expresar por comprensión o por extensión. Por comprensión es cuando los elementos se expresan por una característica bien definida, y por extensión es cuando se escriben todos los eventos del conjunto.

Ejemplo 1: Definamos el conjunto de las vocales.

Por extensión $A = \{a, e, i, o, u\}$

Por comprensión $A = \{x/x \text{ es una vocal}\}$ se lee son los $\mathbf{x}$ tales que $\mathbf{x}$ es una vocal)

La notación, $a \in A$, significa que el elemento a pertenece al conjunto A.

La notación, $b \notin A$, significa que el elemento b no pertenece al conjunto A.

1.1. Operaciones entre conjuntos

a) Unión $(\cup)$

Es una operación entre dos o más conjuntos que produce un nuevo conjunto, formado por todos los elementos que están presentes en al menos uno de los conjuntos, sin repetir elementos.

Ejemplo 2: Dados los conjuntos $A = \{a, e, i, o, u\}$ y $B = \{a, b, c, d, e, f\}$. Encuentre $A \cup B$.

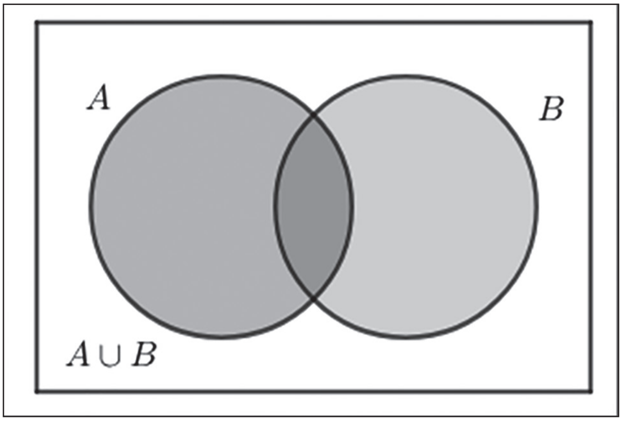
$$A \cup B = \{a, e, i, o, u, b, c, d, f\}$$

Definición de la unión entre dos conjuntos

$A \cup B = \{x / x \in A \vee x \in B\}$(se lee son los x tales que x pertenecen al conjunto A o al conjunto B). Es decir, que están en A o en B, o en ambos.

La unión representada en el diagrama de Venn (Figura84).

Figura 84. *Diagrama Venn de la unión entre conjuntos A, B.*



Fuente: elaboración propia.

b) Intersección $(\cap)$

Es una operación entre dos o más conjuntos que produce un nuevo conjunto, formado por todos los elementos comunes de los conjuntos.

Ejemplo 3: dados los conjuntos $A = \{a, e, i, o, u\}$ y $B = \{a, b, c, d, e, f\}$, encuentre $A \cap B$.

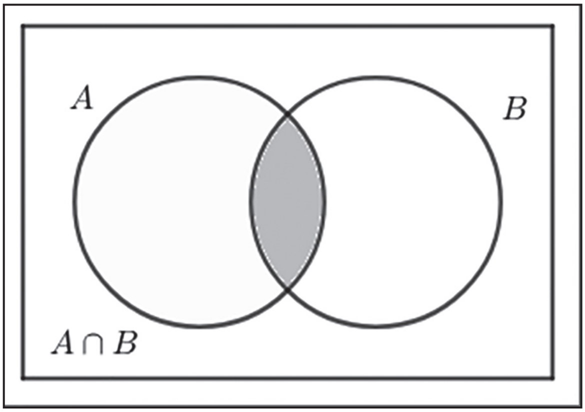
$$A \cap B = \{a, e\}$$

Definición de la intersección entre dos conjuntos

$A \cup B = \{x / x \in A \wedge x \in B\}$ (se lee son los x tales que x pertenecen al conjunto A y al conjunto B). Es decir, que están en A y en B a la vez.

La intersección representada en el diagrama de Venn (Figura 85).

Figura 85. *Diagrama de Venn de la intersección entre los conjuntos A, B.*



Fuente: elaboración propia.

Ejemplo 4: Dados los conjuntos $A = \{a, e, i, o, u\}$ y $B = \{m, n, l\}$, encuentre $A \cap B$.

$$A \cap B = \{\quad\} = \emptyset$$

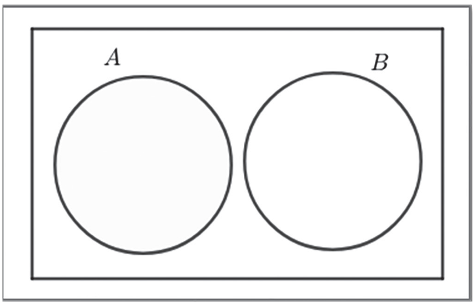
Podemos ver, en este caso, que los conjuntos A y B no tienen elementos en común. Es decir, la intersección entre ellos es el conjunto vacío. Entonces, los conjuntos A y B se dice que son mutuamente excluyentes.

Definición 1: dos conjuntos A y B, se dice que son mutuamente excluyentes si

$$A \cap B = \emptyset$$

Representados en el diagraman de Venn (Figura 86).

Figura 86. *Conjuntos mutuamente excluyentes A, B.*



Fuente: elaboración propia.

Definición 2: tres conjuntos A, B y C, se dice que son mutuamente excluyentes entre sí, si se cumple lo siguiente:

$$A \cap B \cap C = \emptyset$$

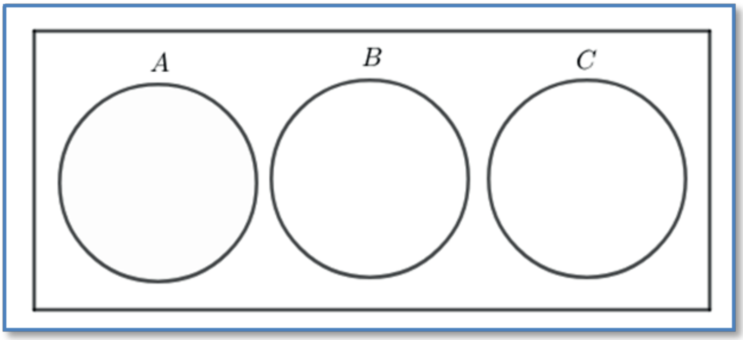
$$A \cap B = \emptyset$$

$$A \cap C = \emptyset$$

$$B \cap C = \emptyset$$

Representados en el diagraman de Venn (Figura 87).

Figura 87. *Conjuntos mutuamente excluyentes A, B, C.*



Fuente: elaboración propia.

Ejemplo 5: dados los conjuntos $A = \{a, e, i, o, u\}$, $B = \{m, n, l\}$ y $C = \{s, t\}$, ¿son estos conjuntos mutuamente excluyentes?

Podemos ver que:

$$A \cap B \cap C = \emptyset$$

$$A \cap B = \emptyset$$

$$A \cap C = \emptyset$$

$$B \cap C = \emptyset$$

Por lo tanto, los conjuntos A, B y C son mutuamente excluyentes entre sí.

c) Complemento (^{c})

Es el conjunto de todos los elementos que pertenecen al universo, pero que no están en el conjunto dado.

Ejemplo 6: dado el conjunto $A = \{a, e, i, o, u\}$ y un conjunto universal $\mathcal{U} = \{xx\ es\ una\ letra\ del\ alfabeto\ inglés\}$, encuentre A^C.

$$A^C = \{xx\ es\ una\ consonante\}$$

O también por extensión:

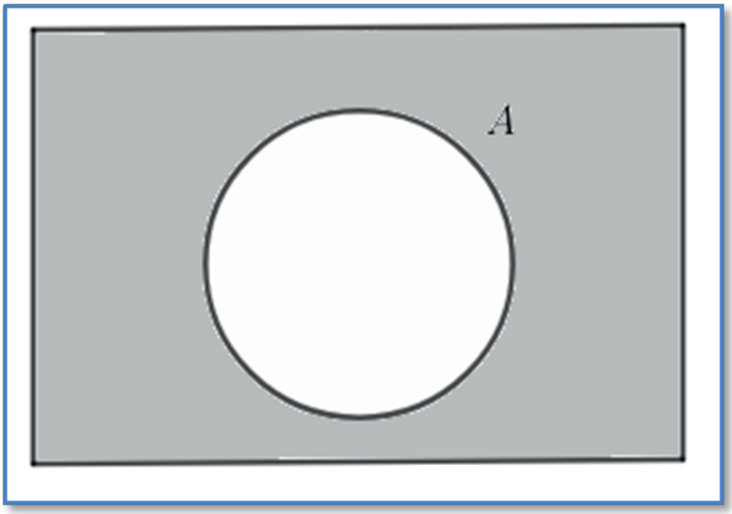
$$A^C = \{b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, y, z\}$$

Definición del complemento de un conjunto

$A^C = \{x/x \notin A\ \wedge x \in \mathcal{U}\}$ (se lee son los x tales que x pertenecen al conjunto A y al conjunto B).

El complemento de un conjunto representado en el diagrama de Venn, Figura 88.

Figura 88. *Diagrama de Veen del complemento del conjunto A.*



Fuente: elaboración propia.

1.2. Propiedades de las operaciones entre conjuntos

a) Ley conmutativa

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

b) Ley asociativa

$$(A \cup B) \cup C = B \cup (A \cup C)$$

$$(A \cap B) \cap C = B \cap (A \cap C)$$

c) Leyes complementarias

$$A \cup \mathcal{U} = \mathcal{U}$$

$$A \cap \mathcal{U} = A$$

$$A \cup \emptyset = A$$

$$A \cap \emptyset = \emptyset$$

$$A \cup A^C = \mathcal{U}$$

$$A \cap A^C = \emptyset$$

d) Leyes distributivas

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

e) Ley involutiva

$$(A^C)^C = A$$

f) Leyes de idempotencia

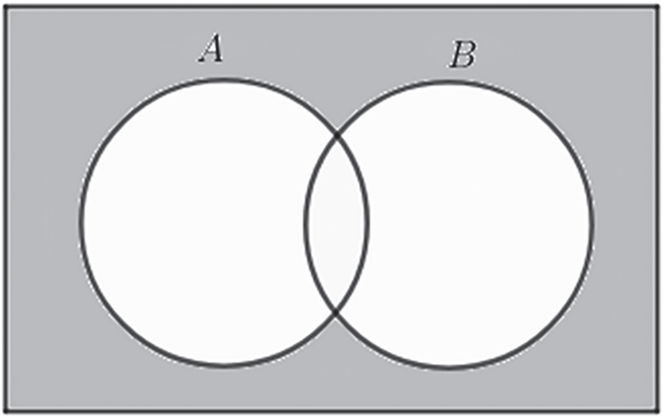
$$A \cup A = A$$

$$A \cap A = A$$

g) Leyes De Morgan

$$(A \cup B)^C = A^C \cap B^C$$

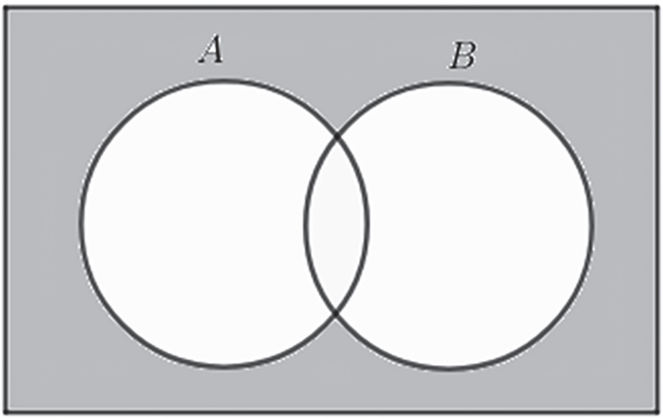
Figura 89. *Complemento de la intersección de A y B.*



$$(A \cap B)^C = A^C \cup B^C$$

Fuente: elaboración propia.

Figura 90. *Diagrama de Venn del complemento de la unión de A, B.*



Fuente: elaboración propia.

1.3. CARDINALIDAD DE CONJUNTOS

Cardinalidad de un conjunto

La cardinalidad de un conjunto A es el número de elementos que hay en el conjunto A. Se denota como $n(A)$.

Ejemplo 7: dado el conjunto $A = \{a, e, i, o, u\}$, encontrar la cardinalidad de A.

Se puede observar que:

$$n(A) = 5$$

Cardinalidad de la unión de dos conjuntos

La cardinalidad de unión de los conjuntos A, B se define como:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

Ejemplo 3: dados los conjuntos $A = \{a, e, i, o, u\}$ y $B = \{a, b, c, d, e, f\}$, encontrar la cardinalidad de $A \cup B$.

Como:

$$A \cup B = \{a, e, i, o, u, b, c, d, f\}$$

$$n(A \cup B) = 9$$

Además, se puede observar que:

$$n(A) = 5$$

$$n(B) = 6$$

También $A \cap B = \{a, e\}$. Entonces:

$$n(A \cap B) = 2$$

Entonces:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

$$9 = 5 + 6 - 2$$

Ejemplo 9: en un grupo de 110 estudiantes se encontró que 52 cursaban inglés, 37 cursaban francés y 24 cursaban las dos materias. Encuentre el número de estudiantes que cursaban al menos una de estas asignaturas.

Según el enunciado:

$$n(I) = 52$$

$$n(F) = 37$$

$$n(I \cap F) = 13$$

Y me pregunta el número de estudiantes que cursaban al menos una de estas asignaturas, es decir, los que están estudiando inglés o francés $n(I \cup F)$.

Como:

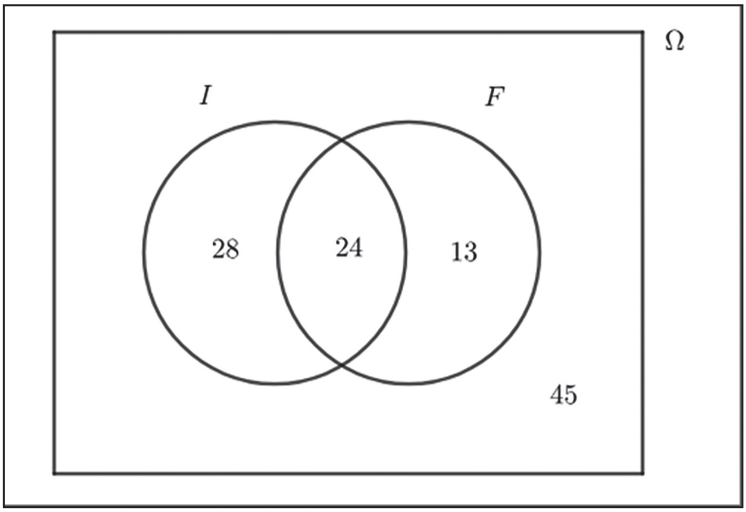
$$n(I \cup F) = n(I) + n(F) - n(I \cap F)$$

$$n(I \cup F) = 52 + 37 - 13$$

$$n(I \cup F) = 76$$

Ahora, resolviendo el ejercicio con el diagrama de Venn, Figura 91.

Figura 91. *Representación de cardinalidad de los que estudian solo inglés, solo francés, ambos o ninguno.*



Fuente: elaboración propia.

El resultado es sumar $28 + 24 + 13 = 76$.

Cardinalidad de la unión de tres conjuntos

La cardinalidad de unión de los conjuntos A, B, C se define como:

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

Ejemplo 10: dado los conjuntos $A = \{a, e, i, o, u\}$, $B = \{a, b, c, d, e, f\}$ *y* $C = \{a, o, c, d, f, m, n, l\}$, encontrar la cardinalidad de $A \cup B \cup C$.

Como:

$$A \cup B \cup C = \{a, e, i, o, u, b, c, d, f, m, n, l\}$$

$$n(A \cup B \cup C) = 12$$

Además, se puede observar que

$$n(A) = 5$$

$$n(B) = 6$$

$$n(C) = 8$$

También:

$$A \cap B = \{a, e\}$$

$$A \cap C = \{a, o\}$$

$$B \cap C = \{a, c, d, f\}$$

$$A \cap B \cap C = \{a\}$$

Entonces:

$$n(A \cap B) = 2$$

$$n(A \cap C) = 2$$

$$n(B \cap C) = 4$$

$$n(A \cap B \cap C) = 1$$

Entonces:

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

$$12 = 5 + 6 + 8 - 2 - 2 - 4 + 1$$

Ejemplo 11: de los 100 estudiantes de segundo semestre de una universidad se tiene que 60 estudiantes asisten a clases de inglés; 40, a clases de francés, y 20, a clases de alemán. 26 estudiantes asisten a las clases de inglés y francés, 10 están asistiendo a las de inglés y alemán y 12 están en las de francés y alemán. Además, 78 asisten por lo menos a alguna de las clases mencionadas anteriormente (es decir, asisten a clases de inglés, francés o alemán). Calcular la cantidad de personas que *a)* solamente asisten a la clase de inglés, *b)* asisten a la clase de inglés y francés, pero no a la de alemán.

Definiendo I inglés. F francés y A alemán.

Según la información del ejercicio, se tiene:

$$n(I) = 60$$

$$n(F) = 40$$

$$n(A) = 20$$

$$n(I \cap F) = 26$$

$$n(I \cap A) = 10$$

$$n(F \cap A) = 12$$

$$n(I \cup F \cup A) = 78$$

Aplicando la fórmula tenemos:

$$n(I \cup F \cup A) = n(I) + n(F) + n(A) - n(I \cap F) - n(I \cap A) - n(F \cap A) + n(I \cap F \cap A)$$

$$78 = 60 + 40 + 20 - 26 - 10 - 12 + n(I \cap F \cap A)$$

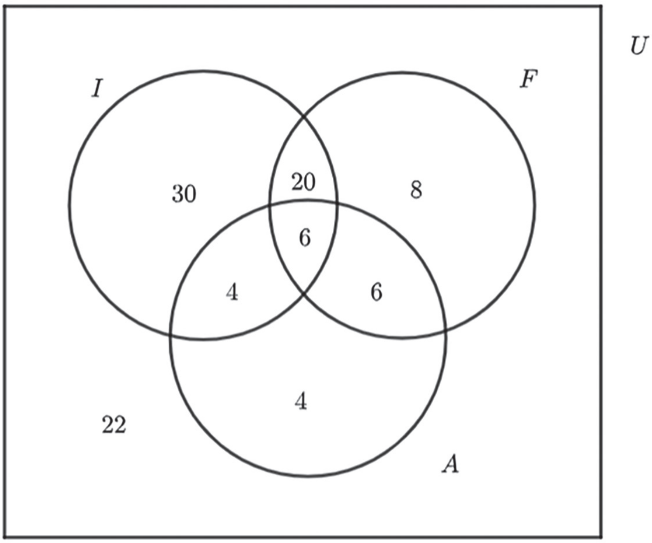
Despejando:

$$n(I \cap F \cap A) = 78 - 60 - 40 - 20 + 26 + 10 + 12$$

$$n(I \cap F \cap A) = 6$$

Ahora, haciendo un diagrama de Venn, Figura 92.

Figura 92. *Representación de la cardinalidad de los que estudian inglés, francés, alemán.*



Fuente: elaboración propia.

Al calcular la cantidad de personas que *a)* solamente asisten a la clase de inglés, es decir, hay que calcular la cantidad de personas que asisten a la clase de inglés y no asisten a las de francés y no asisten a las de alemán:

$$n(I \cap F^C \cap A^C) = 30$$

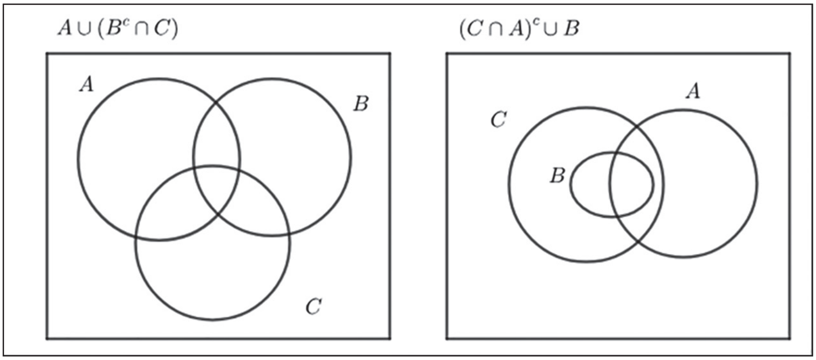
b) asisten a la clase de inglés y francés, pero no a la de alemán.

$$n(I \cap F \cap A^C) = 20$$

2. EJERCICIOS

Ejercicio 1. Sombre la operación indicada en la Figura 93.

Figura 93. *Ejercicio 1*



Fuente: elaboración propia.

Ejercicio 2. En un grupo de 600 estudiantes de una carrera, 80 cursan la asignatura de francés, 100 alemán, 300 inglés. Además, 30 cursan inglés y francés; 40, alemán e inglés; 20, alemán y francés, y 400 cursan por lo menos alguna de estas tres asignaturas. Determine:

a) La cantidad de personas que estudian francés, inglés y alemán.
b) La cantidad de personas que estudian solamente inglés.
c) La cantidad de personas que estudian francés y alemán, pero no inglés.
d) La cantidad de personas que estudian inglés o francés.

Ejercicio 3. De un grupo de 530 personas de segundo semestre, de alguna universidad, 155 están tomando la asignatura A; 137, la asignatura B, y 182, la asignatura C. También se sabe que 39 solo están tomando la asignatura C; 41 solo toman las asignaturas B y C; 67 sólo toman las asignaturas A y C; 64 toman la asignatura B, pero no A, y 219 toman las asignaturas A o B. Si se selecciona un estudiante de este grupo:

a) ¿Cuál es cantidad de personas de este grupo que exactamente está tomando la asignatura A?
b) ¿Cuál es la cantidad de personas de este grupo que no está tomando ninguna de estas asignaturas?

Ejercicio 4. En una ciudad se publican tres periódicos A, B y C. Suponga que 50% de las familias están suscritas al periódico A; 30%, al B, y 20%, al C.

También que 16% de las familias lo están en A y B, pero no en C; 38%, en B o C; 57%, en A o C, y 9%, en los tres. Calcule el porcentaje de familias que:

a) No estén suscritas en el periódico A, pero sí estén suscritas en el B.
b) Estén suscritas en por lo menos un periódico.
c) Estén suscritas en exactamente un periódico.
d) Estén suscritas en el periódico A o B.

Ejercicio 5. En una investigación de mercado se halló que de 500 compradores entrevistados 160 compraban el producto A; 250 compraban el producto B, y 150 no compraban ninguno de los dos.

Complete la Tabla 16 con la información dada anteriormente.

Tabla 16. *Compradores de los productos A, B*

	Compran el producto A	No compran el producto A	Total
Compran el producto B			
No compran el producto B			
Total			

3. CONCEPTOS BÁSICOS DE PROBABILIDAD

3.1. Experimento aleatorio

Es un proceso que cumple con lo siguiente:

- Todos los posibles resultados del experimento se conocen de antemano.
- Cuando se hace el experimento, no se sabe cuál resultado va a ocurrir.
- El experimento se puede repetir bajo idénticas condiciones.

Ejemplo 12: lanzo una moneda.

Ejemplo 13: lanzo un dado.

Ejemplo 14: hago una encuesta donde las preguntas sean cerradas. Por ejemplo, preguntar la preferencia de un producto donde las opciones de respuesta son Bueno, Regular o Malo.

Ejemplo 15: lanzar una moneda hasta obtener cara.

Ejemplo 16: mirar el tiempo de duración de un bombillo.

3.2. Espacio muestral

Es el conjunto de todos los posibles resultados de un experimento aleatorio. Se denota con las letras Ω o S.

Ejemplo 17: lanzo una moneda. Entonces el espacio muestral será**:**

$$\Omega = \{Cara, Sello\}$$

En este ejemplo el espacio muestral es discreto, porque es contable finito.

Ejemplo 18: lanzo un dado. Entonces el espacio muestral será**:**

$$\Omega = \{1,2,3,4,5,6\}$$

En este ejemplo el espacio muestral es discreto, porque es contable finito.

Ejemplo 19: hacer una encuesta donde las preguntas sean cerradas. Por ejemplo, preguntar la preferencia de un producto donde las opciones de respuesta son Bueno, Regular o Malo. Entonces el espacio muestral será**:**

$$\Omega = \{Bueno, Regular, Malo\}$$

En este ejemplo el espacio muestral es discreto, porque es contable finito.

Ejemplo 20: lanzar una moneda hasta obtener cara.

$$\Omega = \{C, SC, SSC, SSSC, SSSSC, \ldots\}$$

En este ejemplo el espacio muestral es discreto, porque es contable infinito.

Ejemplo 21: mirar el tiempo de duración de un bombillo. Entonces el espacio muestral será:

$$\Omega = \{t: es\ el\ tiempo; t \geq 0\}$$

En este ejemplo el espacio muestral es continuo, porque es no contable.

3.3. Evento

Un evento es un subconjunto del espacio muestral. Para los tres primeros ejemplos tenemos:

Ejemplo 22: lanzo una moneda. Entonces el espacio muestral será**:**

$$\Omega = \{Cara, Sello\}$$

Los eventos en este caso son:

$$E_1 = \{Cara\}$$

$$E_2 = \{Sello\}$$

$$E_3 = \{Cara, Sello\}$$

$$E_4 = \{\quad\} = \emptyset$$

Una fórmula para encontrar los subconjuntos de un conjunto A, si el conjunto es finito es $2^{n(A)}$, donde $n(A)$ es la cardinalidad de A. En este ejemplo vemos que el espacio muestral tiene 2 elementos y por lo tanto el número de subconjuntos o eventos para este espacio muestral son $2^2 = 4$.

Ejemplo 23: lanzo un dado de 6 caras. Entonces el espacio muestral será:

$$\Omega = \{1,2,3,4,5,6\}$$

Para este caso, hay 6 elementos en el espacio muestral. Entonces, el número de subconjuntos o eventos será igual a $2^6 = 64$, los eventos se muestran en la Tabla 17.

Tabla 17. *Eventos del experimento aleatorio lanzar un dado*

$E_1 = \{1\}$	$E_{17} = \{3,5\}$	$E_{33} = \{2,3,5\}$	$E_{49} = \{1,3,4,6\}$
$E_2 = \{2\}$	$E_{18} = \{3,6\}$	$E_{34} = \{2,3,6\}$	$E_{50} = \{1,3,5,6\}$

$E_3 = \{3\}$	$E_{19} = \{4,5\}$	$E_{35} = \{2,4,5\}$	$E_{51} = \{1,4,5,6\}$
$E_4 = \{4\}$	$E_{20} = \{4,6\}$	$E_{36} = \{2,4,6\}$	$E_{52} = \{2,3,4,5\}$
$E_5 = \{5\}$	$E_{21} = \{5,6\}$	$E_{37} = \{2,5,6\}$	$E_{53} = \{2,3,4,6\}$
$E_6 = \{6\}$	$E_{22} = \{1,2,3\}$	$E_{38} = \{3,4,5\}$	$E_{54} = \{2,3,5,6\}$
$E_7 = \{1,2\}$	$E_{23} = \{1,2,4\}$	$E_{39} = \{3,4,6\}$	$E_{55} = \{2,4,5,6\}$
$E_8 = \{1,3\}$	$E_{24} = \{1,2,5\}$	$E_{40} = \{3,5,6\}$	$E_{56} = \{3,4,5,6\}$
$E_9 = \{1,4\}$	$E_{25} = \{1,2,6\}$	$E_{41} = \{4,5,6\}$	$E_{57} = \{1,2,3,4,5\}$
$E_{10} = \{1,5\}$	$E_{26} = \{1,3,4\}$	$E_{42} = \{1,2,3,4\}$	$E_{58} = \{1,2,3,4,6\}$
$E_{11} = \{1,6\}$	$E_{27} = \{1,3,5\}$	$E_{43} = \{1,2,3,5\}$	$E_{59} = \{1,2,3,5,6\}$
$E_{12} = \{2,3\}$	$E_{28} = \{1,3,6\}$	$E_{44} = \{1,2,3,6\}$	$E_{60} = \{1,2,4,5,6\}$
$E_{13} = \{2,4\}$	$E_{29} = \{1,4,5\}$	$E_{45} = \{1,2,4,5\}$	$E_{61} = \{1,3,4,5,6\}$
$E_{14} = \{2,5\}$	$E_{30} = \{1,4,6\}$	$E_{46} = \{1,2,4,6\}$	$E_{62} = \{2,3,4,5,6\}$
$E_{15} = \{2,6\}$	$E_{31} = \{1,5,6\}$	$E_{47} = \{1,2,5,6\}$	$E_{63} = \{1,2,3,4,5,6\}$
$E_{16} = \{3,4\}$	$E_{32} = \{2,3,4\}$	$E_{48} = \{1,3,4,5\}$	$E_{64} = \{\ \} = \emptyset$

También podemos expresar, por ejemplo, el evento $E_{27} = \{1,3,5\}$ de la forma E_{27} que salga un número impar.

Ejemplo 24: hacer una encuesta donde las preguntas sean cerradas. Por ejemplo, preguntar la preferencia de un producto donde las opciones de respuesta son Bueno, Regular o Malo. Entonces, el espacio muestral será**:**

$$\Omega = \{Bueno, Regular, Malo\}$$

Para este caso, hay 3 elementos en el espacio muestral. Entonces, el número de subconjuntos o eventos será igual a $2^3 = 8$, los eventos serán:

$$E_1 = \{Bueno\}$$
$$E_2 = \{Regular\}$$
$$E_3 = \{Malo\}$$
$$E_4 = \{Bueno, Regular\}$$
$$E_5 = \{Bueno, Malo\}$$
$$E_6 = \{Regular, Malo\}$$
$$E_7 = \{Bueno, Regular, Malo\}$$
$$E_8 = \{\quad\} = \emptyset$$

4. CONCEPCIONES SOBRE PROBABILIDAD

La probabilidad de un suceso o evento es un concepto muy complejo de precisar, por lo que existen diferentes concepciones acerca de cómo definir la probabilidad de un evento o suceso. En todas estas, lo que sí se ha establecido es que la probabilidad de un evento debe asignar un valor específico (número real) en el intervalo que va de cero a uno. Mientras más se acerca a cero, la probabilidad del evento es más improbable y, por el contrario, entre más se acerca a uno es más probable. Además, si la probabilidad del evento es igual a cero, se dice que es un evento imposible, si la probabilidad es uno, se dice que es un evento seguro, y si la probabilidad es igual a 0.5, el evento es verosímil, porque está en el punto medio de la escala.

4.1. Concepción clásica

En la concepción clásica, la probabilidad de un suceso se define como "cociente entre el número de casos favorables al suceso y el número de casos posibles, siempre que todos sean equiprobables" (Batanero, 2001). Es decir, la probabilidad de un evento E se puede calcular como:

$$P(E) = \frac{n(E)}{n(\Omega)} = \frac{Cardinalidad\ del\ evento}{Cardinalidad\ del\ espacio\ muestral}$$

Esta probabilidad asume que todos los elementos del espacio muestral tienen la misma posibilidad de ocurrir. Además, el cociente de esta expresión da como resultado un número racional, porque tanto el numerador como el denominador son dos números enteros.

Ejemplo 25: lanzo una moneda. Entonces, el espacio muestral será**:**

$$\Omega = \{Cara, Sello\}$$

Los eventos, con sus probabilidades en este caso son**:**

$E_1 = \{Cara\}; P(E_1) = \frac{1}{2} = 0.5$; (evento verosímil)

$E_2 = \{Sello\}; P(E_2) = \frac{1}{2} = 0.5$; (evento verosímil)

$E_3 = \{Cara, Sello\}; P(E_3) = \frac{2}{2} = 1$; (evento seguro)

$E_4 = \{\quad\} = \emptyset; P(E_4) = 0$; (evento imposible)

Ejemplo 26: lanzo un dado. Entonces, el espacio muestral será:

$$\Omega = \{1,2,3,4,5,6\}$$

Definamos el evento E que salga un número mayor que 1, entonces:

$$E = \{2,3,4,5,6\}$$

$$P(E) = \frac{5}{6} = 0.8\overline{3}$$

4.2. Concepción como frecuencia relativa

La definición anterior asumía que todos los resultados de un experimento aleatorio son igualmente posibles de ocurrir y esto no necesariamente se presenta siempre, pues en los fenómenos naturales o en el mundo real hay eventos o sucesos en los que las equiprobabilidades no se presentan. Por ejemplo, la probabilidad de que un vuelo que se hace de cierta ciudad a otra llegue a tiempo puede que no sea la misma de que el vuelo no llegue a tiempo, más bien se puede registrar la proporción de veces en que estos eventos ocurrirán. Por ejemplo, si una aerolínea registra que de cada 100 vuelos, 95 llegaban a tiempo, esta puede afirmar que aproximadamente el 95% de sus vuelos llegan a tiempo. Es decir, la probabilidad de que un vuelo llegue a tiempo sería aproximadamente:

$$P(Vuelo\ llegue\ a\ tiempo) \approx \frac{95}{100}$$

Ahora bien, para el Ejemplo 3.19, donde definíamos el experimento aleatorio, hacer una encuesta con preguntas cerradas para indagar acerca de la preferencia de un producto cuyo espacio muestral es:

$$\Omega = \{Bueno, Regular, Malo\}$$

Vemos que no sólo podemos considerar que todos los resultados de este experimento aleatorio son igualmente posibles, porque la preferencia por una de las anteriores opciones puede ser más que las otras en proporción para un conjunto de personas.

Entonces, habría que definir la probabilidad de un evento E, **como frecuencia relativa,** y para esto asumimos que si un experimento aleatorio se repite n veces bajo las mismas condiciones y n_E son los resultados favorables para el evento E, la probabilidad del evento E será aproximadamente**:**

$$P(E) \approx \frac{n_E}{n}$$

Según observamos en la fórmula anterior, se suministra un valor aproximado a la probabilidad del evento E, ya que la verdadera probabilidad del evento E será:

$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}$$

Cuando n es grande, es decir, cuantos más ensayos se hagan del experimento es más probable que nos acerquemos a la verdadera probabilidad del evento E.

Por otro lado, si miramos Ejemplo 3.20 lanzar una moneda hasta que salga una cara, con espacio muestral infinito:

$$\Omega = \{C, SC, SSC, SSSC, SSSSC, \ldots\}$$

En este caso no podríamos asignar a todos los elementos del espacio muestral la misma probabilidad, porque de ser así la probabilidad de cualquier evento sería cero y tampoco podríamos empíricamente hacer este experimento pues los elementos de este espacio muestral es infinito, es decir, necesitamos de otras formas para definir la probabilidad de un evento o suceso.

4.3. Concepción subjetiva de probabilidad

Hasta el momento se han definido las probabilidades de un evento de manera objetiva como lo es el caso de la concepción clásica y frecuencial de probabilidad (Martinex, 2012). Pero existe también otro tipo de concepción de probabilidad, como lo es el de probabilidad subjetiva, que corresponde a la evaluación personal que se tenga de un evento. Por ejemplo, un médico cuando diagnostica cierta enfermedad en un paciente lo hace con base a su experiencia, sus conocimientos y de un conjunto de exámenes diagnósticos que practica al paciente. Es decir, es una probabilidad *condicionada* a la información con la que se cuenta. En este ejemplo, un paciente que recurre al médico, le da una información sobre su estado de salud. El médico, dada su formación y su experiencia, "cree" que tiene cierta enfermedad, pero debe confirmarla, y le manda ciertos exámenes que espera que ratifiquen la sospecha, o que se rechace, y continuar con otro tipo de exámenes. Esta manera de asignar probabilidades se conoce como probabilidad condicional, que se abordará más adelante.

5. AXIOMAS DE PROBABILIDAD

Axioma 1 si E es un evento cualquiera, entonces $0 \leq P(E) \leq 1$.

Ejemplo 27: se tiene el experimento aleatorio lanzar un dado. Entonces, el espacio muestral será**:**

$$\Omega = \{1,2,3,4,5,6\}$$

Definamos el evento E que salga un número mayor que 1, entonces:

$$E = \{2,3,4,5,6\}$$

$$P(E) = \frac{5}{6} = 0.8\overline{3}$$

$$0 \leq P(E) \leq 1$$

Axioma 2 si A y B son dos eventos mutuamente excluyentes, entonces $P(A \cup B) = P(A) + P(B)$·

Ejemplo 28: experimento aleatorio lanzar un dado. Entonces el espacio muestral será:

$$\Omega = \{1,2,3,4,5,6\}$$

Definamos los eventos E_1 que salga un número par, E_2 salga el número 1. Calcular la probabilidad de que salga el número par o que salga el número 1, entonces:

$$E_1 = \{2,4,6\};\ E_2 = \{1\};\ E_1 \cap E_2 = \emptyset$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) = \frac{3}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3} = 0.\bar{6}$$

Axioma 3 si Ω es el espacio muestral, entonces $P(\Omega) = 1$.

Ejemplo 29: experimento aleatorio lanzar un dado. Entonces, el espacio muestral será:

$$\Omega = \{1,2,3,4,5,6\}$$

Entonces:

$$P(\Omega) = 1$$

Teorema 1: para el conjunto $\emptyset$ (conjunto vacío) su probabilidad es cero, $P(\emptyset) = 0$.

Ejemplo 30: experimento aleatorio lanzar un dado. Entonces, el espacio muestral será:

$$\Omega = \{1,2,3,4,5,6\}$$

Definamos los eventos E_1 que salga el número 7. Calcular la probabilidad del evento E_1.

$$P(E_1) = \emptyset$$

Teorema 2: si A es un evento cualquiera de un espacio muestral, entonces $P(\mathrm{A}) + P(A^C) = 1$.

Ejemplo 31: experimento aleatorio lanzar un dado. Entonces, el espacio muestral será:

$$\Omega = \{1,2,3,4,5,6\}$$

Definamos los eventos E_1 que salga el número par. Entonces ${E_1}^C$no salga un número impar.

$$E_1 = \{2,4,6\}$$

$${E_1}^C = \{1,3,5\}$$

$$P(E_1) + P({E_1}^C) = 1$$

Teorema 3: si A es y B son dos eventos cualesquiera de un espacio muestral, mutuamente excluyentes o no, entonces $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Ejemplo 32: experimento aleatorio lanzar un dado. Entonces, el espacio muestral será:

$$\Omega = \{1,2,3,4,5,6\}$$

Definamos los eventos E_1 que salga un número par, E_2 salga un número mayor que 4. Calcular la probabilidad de que salga el número par o que salga un número mayor que 4, entonces:

$$E_1 = \{2,4,6\}; \; E_2 = \{5,6\}; \; E_1 \cap E_2 = \{6\}$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3} = 0.\overline{6}$$

Ejemplo 33: en un grupo de 110 estudiantes se encontró que 52 cursaban inglés, 37 cursaban francés y 24 cursaban las dos materias. Si se elige un estudiante al azar, encuentre la probabilidad de que estudie al menos una de estas asignaturas.

Según el enunciado:

$$n(I) = 52$$

$$n(F) = 37$$

$$n(I \cap F) = 13$$

Y me pregunta la probabilidad de que estudie al menos una de estas asignaturas, es decir $P(I \cup F)$.

Como:

$$P(I \cup F) = P(I) + P(F) - P(I \cap F)$$

$$P(I \cup F) = \frac{52}{110} + \frac{37}{110} - \frac{13}{110} = \frac{76}{110} \approx 0.69$$

El resultado es sumar $28 + 24 + 13 = 76$ (ver Figura 91 del ejemplo 3.9) y dividirlo por la cantidad total de estudiantes. En este caso 110. $P(I \cup F) = \frac{76}{110}$.

Ejemplo 34: se tiene la Tabla 18 un grupo de personas encuestadas.

Tabla 18. *Tabla Ejemplo 3.34*

	Fuma	No fuma	Total
Hombre	45	125	170
Mujer	28	112	140
Total	73	237	310

Si se elige una de estas personas al azar, ¿cuál es la probabilidad de que sea hombre o fume?

$$P(H \cup F) = P(H) + P(F) - P(H \cap F)$$

$$P(H \cap F) = \frac{170}{310} + \frac{73}{310} - \frac{45}{310} = \frac{198}{310} = \approx 0.64$$

Corolario 1: Si A, B y C son tres eventos cualesquiera de un espacio muestral, mutuamente excluyentes o no, entonces .

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Ejemplo 35: de los 100 estudiantes de segundo semestre de una universidad, se tiene que 60 estudiantes asisten a clases de inglés; 40, a clases de francés, y 20, a clases de alemán. 26 estudiantes asisten a las clases de inglés y francés; 10 están asistiendo a las de inglés y alemán, y 12 están en las de francés y alemán. Además, 78 asisten por lo menos a alguna de las clases (es decir, asisten a clases de inglés, francés o alemán). Si se selecciona al azar una persona de segundo semestre de esta universidad, calcule la probabilidad de que *a)* solamente estudie inglés, *b)* estudie inglés y francés, pero no alemán.

Definiendo I inglés. F francés y A alemán

Según la información del ejercicio, se tiene:

$$n(I) = 60$$
$$n(F) = 40$$
$$n(A) = 20$$
$$n(I \cap F) = 26$$
$$n(I \cap A) = 10$$
$$n(F \cap A) = 12$$
$$n(I \cup F \cup A) = 78$$

Aplicando la fórmula tenemos:

$$P(I \cup F \cup A) = P(I) + P(F) + P(A) - P(I \cap F) - P(I \cap A) - P(F \cap A) + P(I \cap F \cap A)$$

$$78/100 = 60/100 + 40/100 + 20/100 - 26/100 - 10/100 - 12\,1/00 + P(L \cap G \cap H)$$

Despejando:

$$P(I \cap F \cap A) = \frac{78}{100} - \frac{60}{100} - \frac{40}{100} - \frac{20}{100} + \frac{26}{100} + \frac{10}{100} + \frac{12}{100}$$

$$P(I \cap F \cap A) = 6/100$$

Calcular la probabilidad de que a) solamente asista a la clase de inglés (ver Figura 92 del Ejemplo 3.11).

$$P(I \cap F^c \cap A^c) = 30/100$$

Calcular la probabilidad que b) asista a la clase de inglés y francés, pero no a la de alemán (ver Figura 92 del Ejemplo 3.11).

$$P(I \cap F \cap A^C) = \frac{20}{100}$$

6. EJERCICIOS

Ejercicio 1. ¿Cuál es la probabilidad de que el equipo Junior, de Barranquilla, gane el campeonato del fútbol colombiano?

a) Calcular con el enfoque clásico.

b) Calcular con el enfoque frecuencial.

c) Calcular con el enfoque subjetivo.

d) ¿Cuál considera la probabilidad más adecuada para este caso? Explique las razones.

Ejercicio 2. Conteste Falso o Verdadero.

a) Sea el experimento aleatorio lanzar un dado una vez. Los eventos A Salga un número par; B Salga un número menor que dos. Los eventos A y B son mutuamente excluyentes.

b) Para dos eventos A y B de un espacio muestral que son mutuamente excluyentes siempre se tiene que $P(A \cup B) = \mathrm{P(A)} + \mathrm{P(B)}$.

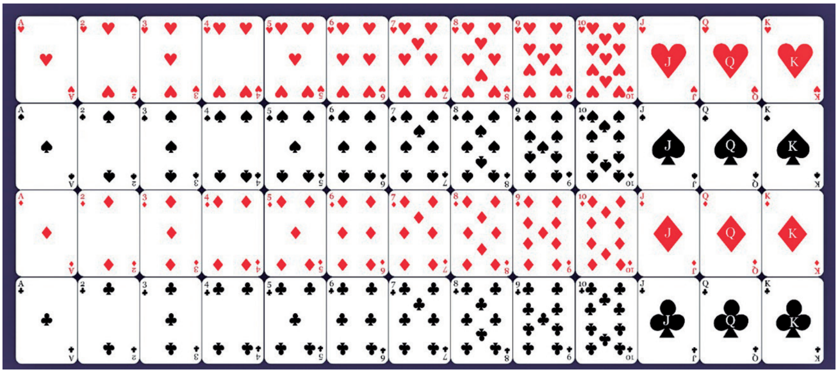
c) Dos eventos A y B de un espacio muestral Ω, son exhaustivos siempre que $A \cup B = \Omega$.

d) Consideremos el lanzamiento de un dado una vez. Usted gana, si el resultado es impar o un número mayor que 4. La probabilidad de ganar es un evento seguro.

Ejercicio 3. Se lanzan dos dados una sola vez, ¿cuál es la probabilidad de que la suma de las caras que muestren sea *¿a)* cinco, *b)* sea doce, *c)* sea mayor a siete y par?, *d)* sea mayor a siete o par?

Ejercicio 4. Se selecciona una carta de un paquete de 52 cartas de póker (Figura 94), a) ¿Cuál es la probabilidad que salga una figura que sea de color negro y una Q?, b) ¿Cuál es la probabilidad que salga una figura que sea de color negro o una Q?

Figura 94. *Cartas de póker.*



Fuente: elaboración: propia.

Ejercicio 5. En un grupo de estudiantes, la probabilidad de que tengan computador es de 0.60; auto, de 0.30, y que tengan ambos de 0.25. *a)* ¿Cuál es la probabilidad de que un estudiante tenga computador o auto?, *b)* ¿Cuál es la probabilidad de que un estudiante tenga solamente computador?, *c)* ¿Cuál es la probabilidad de que un estudiante tenga solamente auto?, *d)* ¿Cuál es la probabilidad de que un estudiante no tenga ni auto ni computador?

Ejercicio 6. De una muestra de 150 personas de cierta ciudad, 100 están suscritos a la plataforma Netflix; 80, a la plataforma Disney; 65, a HBO; 30 están suscritos a estas tres plataformas al tiempo; 10, a las plataformas Disney y HBO, pero no a Netflix; 118 a las plataformas Netflix o HBO, y 26 sólo están suscritas a la plataforma Disney. Si se selecciona aleatoriamente a alguna persona de esa muestra, calcular a) la probabilidad de que esté suscrita a alguna de estas tres plataformas, b) la probabilidad de que este suscrita a la plataforma Netflix, pero no a la de HBO.

Ejercicio 7. La siguiente Tabla 19 muestra el número de estudiantes matriculados en tres especialidades de administración de empresas en una universidad, en los años 2014 y 2015.

Tabla 19. *Especialidades vs año*

Especialidad/ año	2020	2021
Finanzas	160	250
Marketing	140	200
Contabilidad	100	150

Si se escoge un estudiante al azar, calcule la probabilidad de que *a)* estudie Finanzas o Contabilidad, *b)* se haya matriculado en el 2020 o estudie Marketing, *c)* No estudie Contabilidad o se haya matriculado en el 2021, *d)* Estudie finanzas y se haya matriculado en el 2021.

Ejercicio 8. Sean A y B dos eventos asociados a un experimento. A y B son conjuntos mutuamente excluyentes. Suponga que $P(A^C) = 1/3$, $P(B) = 1/5$, y calcule *a)* $P(A \cup B)$, *b)* $P(A^C \cap B^C)$.

Ejercicio 8. Sean A y B dos eventos asociados a un experimento. Suponga que $P(A^C) = 2/5$, $P(A \cup B) = 5/7$; y $P(B^C) = 24/35$. Calcule *a)* $P(A \cap B)$, *b)* $P(A \cap B^C)$.

Ejercicio 9. Sean A y B dos eventos asociados a un experimento. Suponga que $P(A \cup B) = P(B) = 2/3$. $P(A \cap B) = 1/2$. Calcule a) $P(A)$, b) $P(B^C)$.

7. ALGUNAS TÉCNICAS DE CONTEO

Las técnicas de conteo son métodos matemáticos usados para determinar de manera rápida el número de posibles resultados de un espacio muestral, sin necesidad de enumerar cada caso individualmente.

7.1. Regla de multiplicación

Dados k conjuntos con n_1 elementos en el primer conjunto, n_2 en el segundo, hasta n_k el k −ésimo. Si se desea escoger una muestra de k elementos tomando un elemento de cada k conjunto, entonces el número de muestras diferentes que se pueden formar es de:

$$n_1 \times n_2 \times \ldots \times n_k$$

Ejemplo 36: un menú ofrece 2 sopas, 4 platos fuertes, 3 jugos y 6 postres. ¿De cuantas formas diferentes puedo escoger el menú si se escoge una sopa, un plato fuerte, un jugo y un postre?

$$2 \times 4 \times 3 \times 6 = 144$$

Ejemplo 37: se lanza un dado y una moneda. ¿Cuántos elementos hay en el espacio muestral de este experimento aleatorio?

$$2 \times 6 = 12$$

7.2. Regla de combinación

Se tiene un conjunto con N elementos de los cuales se quiere escoger una muestra de n elementos. Entonces, el número de muestras distintas de tamaño n que se obtienen, se puede calcular como:

$$NCn = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

Nota la regla de combinación se usa cuando se desea seleccionar objetos a partir de una colección, sin contar el orden.

Ejemplo 38: se tienen 3 estudiantes María, José y Pedro. Se desea escoger a dos estudiantes para formar un grupo de estudio ¿De cuantas maneras es posible?

$$3C2 = \binom{3}{2} = \frac{3!}{2!\,(3-2)!} = \frac{1 \times 2 \times 3}{(1 \times 2)(1)} = 3$$

Los grupos que se pueden formar son 3 y estos se enuncian a continuación

María, José

María, Pedro

José, Pedro

Ejemplo 39: en una urna hay 4 pelotas negras y 5 blancas, se desea seleccionar 6. ¿Cuántas selecciones son posibles?

$$9C6 = \binom{9}{6} = \frac{9!}{6!\,(9-6)!} = 84$$

Ejemplo 40: en una urna hay 4 pelotas negras y 5 blancas, se desea seleccionar 6, y además dos deben ser negras y cuatro deben ser blancas ¿Cuántas selecciones son posibles?

$$4C2 \times 5C4 = \binom{4}{2} \times \binom{5}{4} = 30$$

Ejemplo 41: en una urna hay 4 pelotas negras y 5 blancas, se desea seleccionar 6, y además deben escogerse por lo menos 2 negras ¿Cuántas selecciones son posibles?

$$4C2 \times 5C4 \ + \ 4C3 \times 5C3 \ + \ 4C4 \times 5C2 \ = \binom{4}{2} \times \binom{5}{4} + \binom{4}{3} \times \binom{5}{3} + \binom{4}{4} \times \binom{5}{2} = 80$$

7.3. Regla de permutación

Se tiene un conjunto con N elementos de los cuales se quiere escoger una muestra de n elementos, donde cada elemento de la muestra debe acomodarse en n posiciones. Entonces, el número de muestras distintas de tamaño n que se obtienen, se puede calcular como:

$$NPn = \frac{N!}{(N-n)!} = N(N-1)(N-2)\ldots(N-n+1)$$

Recuerde que:

$$NPN = N!$$

Nota: la regla de permutación se usa en los casos en que se desea seleccionar objetos a partir de una colección, pero teniendo en cuenta el orden o también se usa cuando es necesario agrupar objetos diferentes en algún orden específico (uno detrás del otro y, por lo tanto, sin repetición).

Ejemplo 42: se tienen tres estudiantes María, José y Pedro. Se desea escoger a dos estudiantes para nombrar presidente y vicepresidente del curso ¿De cuántas maneras es posible?

$$3P2 = \frac{3!}{(3-2)!} = \frac{1 \times 2 \times 3}{(1)} = 6$$

Los grupos que se pueden formar teniendo en cuenta el orden para escoger presidente y vicepresidente son 6. Estos se enuncian a continuación:

Presidente, vicepresidente

María, José

María, Pedro

José, Pedro

José, María

Pedro, María

Pedro, José

Ejemplo 43: en un grupo hay 4 mujeres y 5 hombres. Se desea hacer una fila con este grupo de personas. ¿De cuántas maneras es posible?

$$9! = 362880$$

Ejemplo 44: en un grupo hay 4 mujeres y 5 hombres. Desea hacerse una fila para subir a un autobús teniendo en cuenta que primero se suben las mujeres. ¿De cuantas maneras es posible?

$$4! \times 5! = 2880$$

Ejemplo 45: en un grupo hay 4 mujeres y 5 hombres. Se desea hacer una fila para subir a un autobús teniendo en cuenta que todas las mujeres se suben una detrás de otra. Es decir, juntas. ¿De cuántas maneras es posible?

$$6! \times 4! = 17280$$

8. EJERCICIOS

Ejercicio 1. En un concurso de televisión participan doce personas a) ¿Cuántas posibles clasificaciones se pueden presentar al finalizar el concurso?, b) ¿Cuántas posibles clasificaciones de los tres primeros puestos se pueden presentar al finalizar el concurso?

Ejercicio 2. ¿Cuántos posibles resultados se pueden dar al lanzar un dado cinco veces consecutivas?

Ejercicio 3. En un salón de clases se encuentran 6 hombres y 12 mujeres, a) ¿Cuántos posibles comités de representantes pueden conformarse de 8 miembros, si en cada uno debe haber 2 hombres? b) ¿Cuántos comités de 8 miembros pueden formarse si cada uno de ellos debe contener por mucho 3 mujeres?

Ejercicio 4. ¿De cuántas formas puede contestarse un examen de 20 de preguntas cerradas con cuatro opciones de respuesta?

Ejercicio 5. En un automóvil hay lugar para cinco personas, incluyendo al que conduce. Suponga que cinco amigos salen y se van en ese auto. Determine de cuántas maneras se acomodarán si *a)* cualquiera de ellos puede conducir, b) sólo uno de ellos puede conducir.

Ejercicio 6. Se tiene que formar una clave con cuatro letras (del alfabeto inglés) y tres dígitos (números). ¿De cuantas formas es posible?

Ejercicio 7. Se tiene que formar una clave con cuatro letras (del alfabeto inglés) diferentes y tres números diferentes. ¿De cuántas formas es posible?

Ejercicio 8. Se tienen 3 hombres y 4 mujeres. Desea escogerse un comité de 4 personas. *a)* ¿De cuántas maneras se puede escoger, si debe haber 2 hombres y 2 mujeres?, *b)* ¿De cuántas maneras se puede escoger si debe haber por lo menos 2 hombres?, *c)* ¿De cuántas maneras se puede escoger si al comité no debe pertenecer una mujer y un hombre específicos?

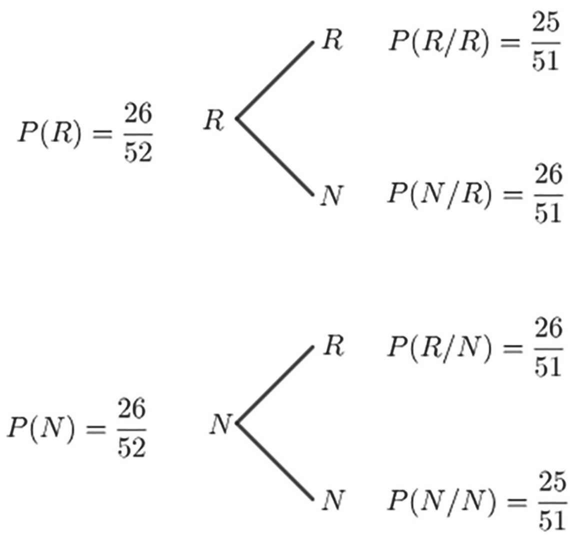
Ejercicio 9. ¿De cuántas maneras diferentes pueden formarse los diferentes arreglos de las letras de la palabra “aleatorio”?

Ejercicio 10. En la fila de un banco se encuentran cuatro jóvenes, tres señores, dos mujeres embarazadas y un adulto mayor. ¿De cuántas formas diferentes pueden pasar si se sabe que siempre deben pasar primero las mujeres embarazadas y el adulto mayor?

9. PROBABILIDAD CONDICIONAL

Ejemplo 46: se tiene un paquete de 52 cartas**.** Se sacan dos cartas (sin reemplazo) y se mira el color de la carta. Nuestro espacio muestral si sólo miramos el color de la carta es de $\Omega = \{RR, RN, NR, NN\}$. Representado este espacio muestral en el diagrama de árbol, con sus respectivas probabilidades.

Figura 95. *Diagrama de árbol cuando se sacan dos cartas (eventos dependientes).*



Fuente: elaboración propia.

Vemos que cuando se saca la primera carta, puede salir una roja o una negra, y las probabilidades, respectivamente, son:

$$P(R_1) = 26/52$$

$$P(N_1) = 26/52$$

Ahora, si se saca la segunda carta, puede salir una roja o negra, pero las probabilidades de sacar una u otra dependen de lo que haya salido anteriormente. Las probabilidades que se obtienen son probabilidades condicionales:

$P(R_2/R_1) = \frac{25}{51}$. Se lee la probabilidad de sacar una carta roja en la segunda, dado que saqué roja en la primera.

$P(N_2/R_1) = \frac{26}{51}$. Se lee la probabilidad de sacar una carta negra en la segunda, dado que saqué roja en la primera.

$P(R_2/N_1) = \frac{26}{51}$. Se lee la probabilidad de sacar una carta roja en la segunda, dado que saqué negra en la primera.

$P(N_2/N_1) = \frac{25}{51}$. Se lee la probabilidad de sacar una carta negra en la segunda, dado que saqué negra en la primera.

a) ¿Cuál es la probabilidad de que salgan dos rojas? En esta pregunta piden encontrar la probabilidad de que salga roja en la primera y roja en la segunda. Y se calcula de la siguiente forma:

$$P(R_1 \cap R_2) = P(R_1) * P(R_2/R_1)$$

$$P(R_1 \cap R_2) = \left(\frac{26}{52}\right) * \left(\frac{25}{51}\right) = \frac{25}{102}$$

b) ¿Cuál es la probabilidad de que la segunda carta que se saque sea negra? En esta pregunta se requiere que la segunda carta sea negra, sin importar cuál haya salido primero. Es decir:

$$P(N_2) = P(R_1 \cap N_2) + P(N_1 \cap N_2)$$

$$P(N_2) = P(R_1) * P(N_2/R_1) + P(N_1) * P(N_2/N_1)$$

$$P(N_2) = \left(\frac{26}{52}\right) * \left(\frac{26}{51}\right) + \left(\frac{26}{52}\right) * \left(\frac{25}{51}\right) = \frac{26}{102} + \frac{25}{102} = \frac{51}{102} = 0.5$$

En este ejemplo, presta atención a que la probabilidad no se puede calcular con la definición clásica, porque el espacio muestral va cambiando a medida que extraigo cartas. En este caso, se dice que los eventos son **dependientes.**

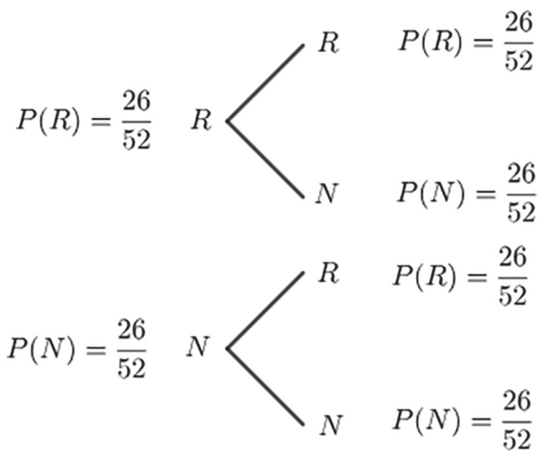
Las probabilidades de cada elemento del espacio muestral para este experimento respectivamente es:

$$\Omega = \{RR, RN, NR, NN\}$$

$$\frac{25}{102}, \frac{26}{102}, \frac{26}{102}, \frac{25}{102}$$

Ejemplo 47: se tiene un paquete de 52 cartas**.** Se sacan dos cartas (con reemplazo) y se observa el color de la carta. Nuestro espacio muestral es $\Omega = \{RR, RN, NR, NN\}$. Representado este espacio muestral en el diagrama de árbol (Figura 96), con sus respectivas probabilidades:

Figura 96. *Diagrama de árbol, cuando se sacan dos cartas (eventos independientes).*



Fuente: elaboración propia.

Observamos que cuando se saca la primera carta, puede salir una roja o una negra y las probabilidades, respectivamente, son:

$$P(R_1) = 26/52$$

$$P(N_1) = 26/52$$

Ahora, si se saca la segunda carta, puede salir una roja o negra, y las probabilidades de sacar una u otra no dependen de lo que haya salido anteriormente. Las probabilidades se mantienen:

$$P(R_2) = 26/52$$
$$P(N2_1) = 26/52$$

a) ¿Cuál es la probabilidad de que salgan dos rojas?, En esta pregunta piden encontrar la probabilidad de que salga roja en la primera y roja en la segunda. Y se calcula de la siguiente forma:

$$P(R_1 \cap R_2) = P(R_1) * P(R_2)$$
$$P(R_1 \cap R_2) = \left(\frac{26}{52}\right) * \left(\frac{26}{52}\right) = \frac{1}{4}$$

b) ¿Cuál es la probabilidad de que la segunda carta que se saque sea negra? En esta pregunta se requiere que la segunda carta sea negra, sin importar cuál haya salido primero. Es decir:

$$P(N_2) = P(R_1 \cap N_2) + P(N_1 \cap N_2)$$
$$P(N_2) = P(R_1) * P(N_2) + P(N_1) * P(N_2)$$
$$P(N_2) = \left(\frac{26}{52}\right) * \left(\frac{26}{52}\right) + \left(\frac{26}{52}\right) * \left(\frac{26}{52}\right) = \frac{1}{4} + \frac{1}{4} = 0.5$$

En este caso, se dice que los eventos son **independientes.**

Las probabilidades de cada elemento del espacio muestral para este experimento, respectivamente, es:

$$\Omega = \{RR, RN, NR, NN\}$$
$$\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$$

Ejemplo 48: con base en la Tabla 18 del Ejemplo 3.34, un grupo de personas encuestadas:

Si se elige una de estas personas al azar y resulta que la persona es fumadora, ¿cuál es la probabilidad de que sea hombre?

$$P(H/F) = \frac{45}{73} \approx 0.62$$

Ejemplo 49: en un grupo de 110 estudiantes se encontró que 52 cursaban inglés, 37 cursaban francés y 24 cursaban las dos materias (Tabla 20). Si se elige un estudiante al azar y resulta que estudia inglés, ¿cuál es la probabilidad de que estudie francés?

Tabla 20. *Ejemplo 3.49*

	Estudia inglés	No estudia inglés	Total
Estudia francés	24	13	37
No estudia francés	28	45	73
Total	52	58	110

Si se elige un estudiante al azar y resulta que estudia inglés, ¿cuál es la probabilidad de que estudie francés?

$$P(F/I) = \frac{24}{52} \approx 0.46$$

Definición 1 dados dos eventos A y B de un espacio muestral Ω. La probabilidad de la intersección se define como:

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

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

Los eventos A y B son **dependientes.**

De lo anterior, las probabilidades condicionales:

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

$P(A/B)$ probabilidad de A dado que ocurrió B.

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

$P(B/A)$ probabilidad de A dado que ocurrió B.

Definición 2 dados dos eventos A y B de un espacio muestral Ω. Los eventos A y B son **independientes** si

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

Se observa que $P(A/B) = A$ o $P(B/A) = P(B)$

Definición 3: dados dos eventos A, B y C de un espacio muestral Ω. La probabilidad de la intersección se define como:

$$P(A \cap B \cap C) = P(A) * P(B/A) * P(C/(A \cap B)$$

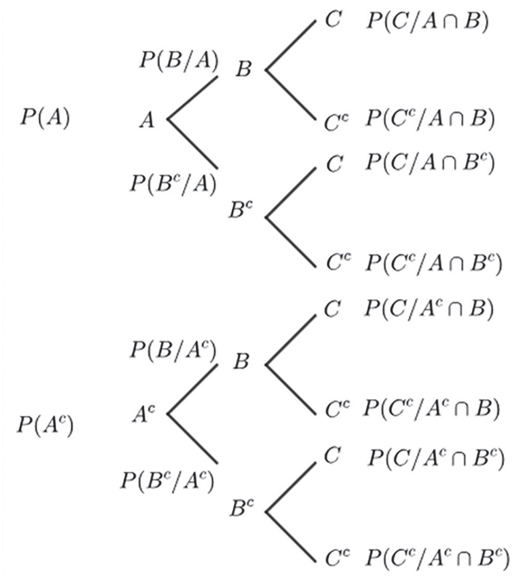
$P(A)$ Probabilidad de A

$P(B/A)$ probabilidad de A dado que ocurrió B

$P(C/(A \cap B)$ probabilidad de C dado que ocurrió A y B.

La Figura 97 muestra, en la primera rama del árbol, el evento compuesto de que ocurra A y B y C $(A \cap B \cap C)$.

Figura 97. *Diagrama de árbol. Definición 3.*

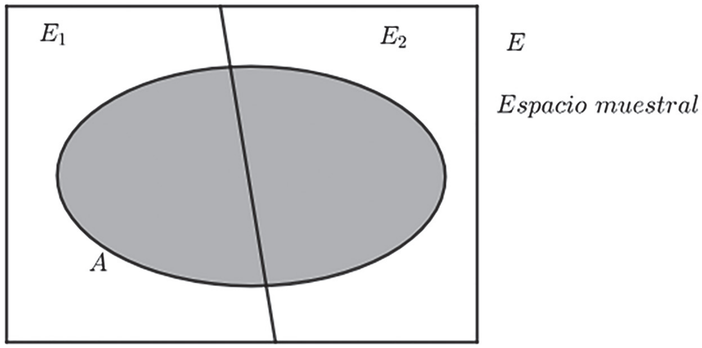


Fuente: elaboración propia.

9.1. Regla de probabilidad total

Dados dos eventos E_1 y E_2 de un espacio muestral S, donde E_1 y E_2 son mutuamente excluyentes $(E_1 \cap E_2 = \emptyset)$ y además exhaustivos ($E_1 \cup E_2 = S$). Y dado un evento A del espacio muestral S, como se muestra en la Figura 98.

Figura 98. *Diagrama de Venn. Probabilidad Total.*



Fuente: elaboración propia.

Entonces, la probabilidad de A se calcula:

$$P(A) = P(A \cap E_1) + P(A \cap E_2) = P(A/E_1)P(E_1) + P(A/E_2)P(E_2)$$

Ejemplo 50: Juan tiene un teléfono celular con dos operadores Movistar y Claro. Normalmente, el 60% de las veces usa el Movistar, y el 40%, el Claro. El primero falla 10% de las veces y el segundo, 5%.

Calcular la probabilidad de que al llamar le falle el teléfono celular.

Los datos del ejercicio son:

$$P(M) = 0.6$$

$$P(C) = 0.4$$

$$P(F/M) = 0.1$$

$$P(F/C) = 0.05$$

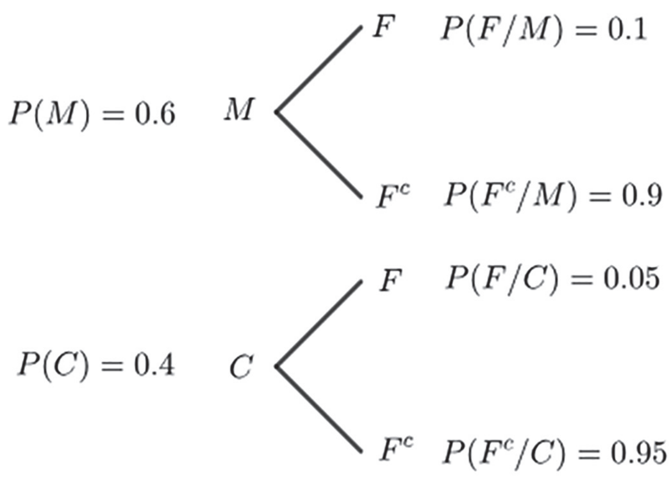
Si Juan hace 100 llamadas, y haciendo una tabla con la anterior información, se tiene la Tabla 21.

Tabla 21. *Ejemplo 50*

Número de llamadas	Movistar	Claro	TOTAL
FALLA	6	2	8
NO FALLA	54	38	92
TOTAL	60	40	100

Con representación de diagrama de árbol (Figura 99).

Figura 99. *Diagrama de árbol. Ejemplo 50.*



Fuente: elaboración propia.

Al calcular la probabilidad de que al llamar le falle el teléfono celular, usando la Tabla 21, se tiene:

$$P(F) = \frac{8}{100} = 0.08$$

Con diagrama de árbol, Figura 99, se tiene:

$$P(F) = P(M \cap F) + P(C \cap F) = P(F/M)P(M) + P(F/C)P(C)$$

$$P(F) = (0.6)(0.1) + (0.4)(0.05) = 0.08$$

9.2. Teorema de Bayes

Dados dos eventos E_1 y E_2 de un espacio muestral S. Donde E_1 y E_2 son mutuamente excluyentes $(E_1 \cap E_2 = \emptyset)$ y además exhaustivos $(E_1 \cup E_2 = S)$. Y dado un evento A del espacio muestral S, entonces:

$$P(E_1/A) = \frac{P(E_1 \cap A)}{P(A)} = \frac{P(A/E_1)P(E_1)}{P(A/E_1)P(E_1) + P(A/E_2)P(E_2)}$$

Ejemplo 51: Juan tiene un teléfono celular con dos operadores Movistar y Claro. Normalmente, el 60% de las veces usa el Movistar y el 40%, el Claro. El primero falla 10% de las veces y el segundo, 5%.

Si Juan hace una llamada y le falla el teléfono celular, ¿cuál es la probabilidad de haya sido por el Movistar?

Los datos del ejercicio son:

$$P(M) = 0.6$$

$$P(C) = 0.4$$

$$P(F/M) = 0.1$$

$$P(F/C) = 0.05$$

Si Juan hace 100 llamadas y haciendo una tabla con la anterior información se tiene:

Si Juan hace una llamada y le falla el teléfono celular, ¿cuál es la probabilidad de que haya sido por el Movistar? Con base a la Tabla 21 se tiene:

$$P(M/F) = \frac{6}{8} = 0.75$$

Con diagrama de árbol (Figura 99).

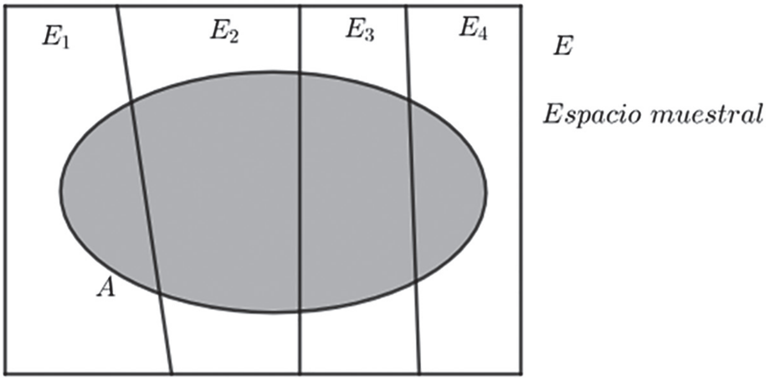
$$P(M/F) = \frac{P(M \cap F)}{P(F)} = \frac{P(F/M)P(M)}{P(F/M)P(M) + P(F/C)P(C)}$$

$$P(M/F) = \frac{P(M \cap F)}{P(F)} = \frac{(0.6)(0.1)}{(0.6)(0.1) + (0.4)(0.05)} = \frac{0.06}{0.08} = 0.75$$

Si generalizamos…

Dados dos eventos E_1, E_2, E_3 y E_4de un espacio muestral S, donde E_1, E_2, E_3 y E_4 son mutuamente excluyentes entre sí y además exhaustivos. Y dado un evento A del espacio muestral S, como se muestra en la Figura 100.

Figura 100. *Diagrama de Venn. Teorema de Bayes.*



Fuente: elaboración propia.

Regla de probabilidad total

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + P(A \cap E_3) + P(A \cap E_4)$$

$$P(A) = P(A/E_1)P(E_1) + P(A/E_2)P(E_2) + P(A/E_3)P(E_3) + P(A/E_4)P(E_4)$$

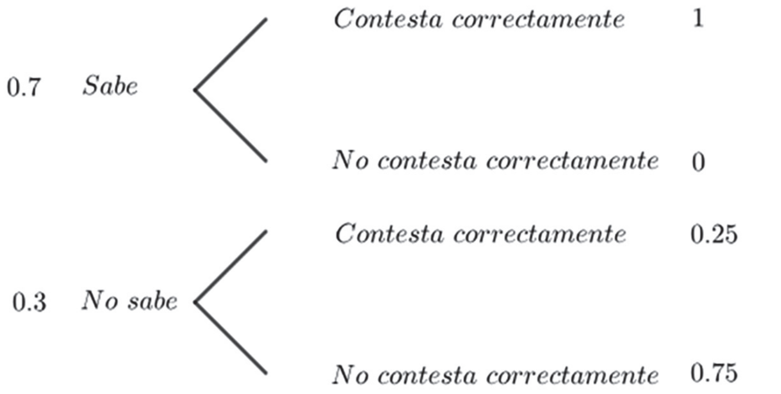
Teorema de Bayes

$$P(E_1/A) = \frac{P(E_1 \cap A)}{P(A)}$$

$$= \frac{P(A/E_1)P(E_1)}{P(A/E_1)P(E_1) + P(A/E_2)P(E_2) + P(A/E_3)P(E_3) + P(A/E_4)P(E_4)}$$

Ejemplo 52: un estudiante presenta un examen de estadística y en una pregunta de selección múltiple con cuatro opciones de respuesta y sólo una correcta tiene la probabilidad de saber la respuesta de 0.7. El diagrama de árbol de este enunciado se puede representar por la Figura 101.

Figura 101. *Diagrama de árbol. Ejemplo 3.52.*



Fuente: elaboración propia.

a) ¿Cuál es la probabilidad de que el estudiante no conteste correctamente a la pregunta?

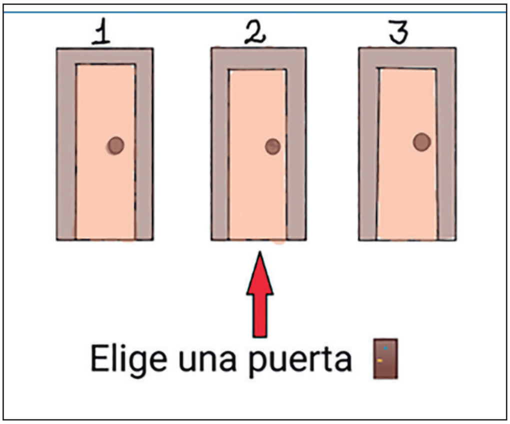
$$P(C^c) = P(S \cap C^c) + P(S^c \cap C^c) = P(S) * P(C^c/S) + P(S^c) * P(C^c/S^c)$$
$$= 0.7 * 0 + 0.3 * 0.75 = 0.225$$

b) Suponga que el estudiante contesta correctamente la pregunta. ¿A qué es igual la probabilidad de que realmente sepa la respuesta correcta?

$$P(S/C) = \frac{P(S \cap C)}{P(C)} = \frac{P(S) * P(C/S)}{P(S) * P(C/S) + P(S^c) * P(C/S^c)} = \frac{0.7 * 1}{0.7 * 1 + 0.3 * 0.25}$$
$$\approx 0.9032$$

Ejemplo 53: el problema de Monty Hall proviene del concurso de un programa de televisión entre los años 60 y mediados de los años 80, cuyo presentador se llamaba Monty Hall. El concurso consiste en que el presentador de televisión enseña 3 puertas cerradas al concursante, detrás de una de ellas hay un premio, que es un carro, y en las otras dos hay un burro en cada una. El presentador sabe cuál puerta tiene detrás el carro, y le dice al concursante que elija una puerta (Figura 102).

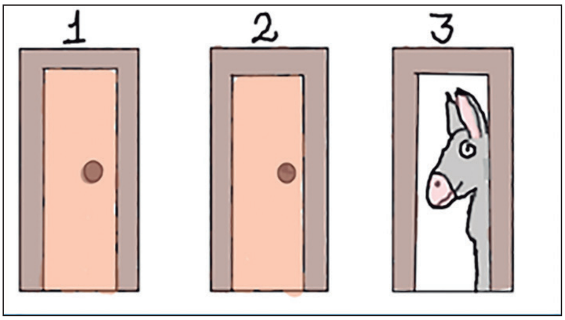
Figura 102. *Elección de puerta. Problema de Monty Hall.*



Fuente: elaboración propia.

Después, sin abrir la puerta seleccionada, el presentador abre alguna de las otras dos puertas y muestra que hay un burro (Figura 103) y, de este modo, el presentador le pregunta al concursante si se queda con la puerta que eligió inicialmente o si cambia la puerta. La pregunta que surge de este problema es si la probabilidad cambia o no si se elige cambiar de puerta.

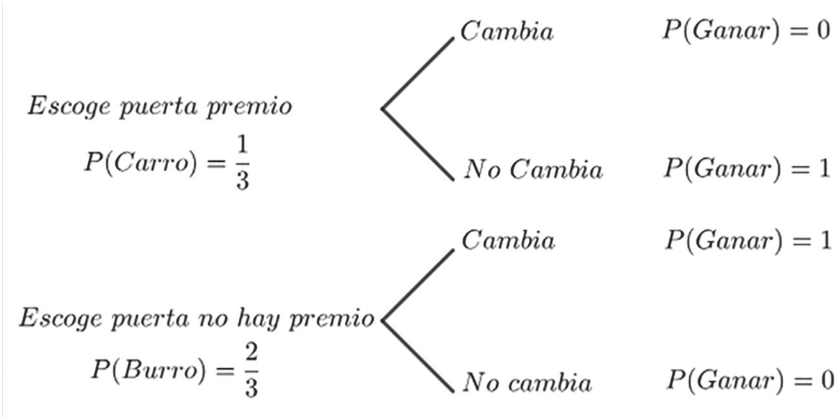
Figura 103. *El presentador abre una puerta. Problema de Monty Hall.*



Fuente: elaboración propia.

El problema se puede resolver por Regla de probabilidad total, hacemos el siguiente gráfico de árbol, que muestra la Figura 104.

Figura 104. *Diagrama de árbol. Problema de Monty Hall.*



Fuente: elaboración propia.

Si observamos la probabilidad de ganar cambiando la puerta es:

P(Ganar cambiando la puerta)=$\frac{1}{3} * 0 + \frac{2}{3} * 1 = 2/3$

Ahora, si observamos la probabilidad de ganar no cambiando la puerta es:

P(Ganar no cambiando la puerta)=$\frac{1}{3} * 1 + \frac{2}{3} * 0 = 1/3$

Entonces, cuando se cambia la puerta hay mayor probabilidad de ganar.

También se puede observar de manera frecuencial estas probabilidades. A continuación, se muestra una tabla en la que participaron 19 estudiantes cada uno jugó 20 veces cambiando la puerta y 20 veces no cambiando la puerta. La Tabla 22 muestra el conteo de ganar que hizo cada estudiante en el juego.

Tabla 22. *Encuesta estudiantes. Problema de Monty Hall*

	Cambia la puerta			**No cambia la puerta**		
Estudiantes	**Gana**	**Pierde**	**Total**	**Gana**	**Pierde**	**Total**
1	11	9	20	6	14	20
2	12	8	20	7	13	20
3	13	7	20	4	16	20
4	14	6	20	3	17	20

5	0	20	20	1	19	20
6	14	6	20	4	16	20
7	15	5	20	9	11	20
8	12	8	20	11	9	20
9	15	5	20	10	10	20
10	15	5	20	8	12	20
11	9	11	20	2	18	20
12	14	6	20	7	13	20
13	13	7	20	7	13	20
14	12	8	20	12	8	20
15	15	5	20	9	11	20
16	13	7	20	9	11	20
17	12	8	20	11	9	20
18	14	6	20	6	14	20
19	12	8	20	3	17	20
Suma	**235**	**145**	**380**	**129**	**251**	**380**
Proporción	**0,61842**	**0,38157**		**0,33947**	**0,660526**	

Al final de la tabla se ve que la proporción de veces que se gana cambiando la puerta es de aproximadamente 2/3 (de 0.618221025) y la proporción de veces que se gana no cambiando la puerta es aproximadamente 1/3 (de 0.33947368).

10. EJERCICIOS

Ejercicio 1. Se seleccionan 2 artículos de manera aleatoria, sin reemplazo, de un contenedor donde hay 400 artículos. De los 400 hay 10 que están defectuosos. *a)* ¿Cuál es la probabilidad de que el segundo artículo seleccionado esté defectuoso si el primero lo fue también?, *b),* ¿cuál es la probabilidad de que los dos artículos seleccionados estén defectuosos?, *c)* ¿Cuál es la probabilidad de que el segundo artículo seleccionado esté defectuoso? *d)* ¿Cuál es la probabilidad de que salga al menos un artículo defectuoso?

Ejercicio 2. En una urna hay 20 bolas 4 azules, 3 rojas, 7 blancas y 6 verdes.

- Se extraen 2 bolas (con sustitución), Determine *a)* la probabilidad de que ambas sean blancas *b)* la probabilidad de que la primera bola sea azul y la segunda roja.

- Ahora se extraen 2 bolas (sin sustitución). Determine *c)* la probabilidad de que la primera sea verde, dado que la primera lo fue *d)* la probabilidad de que ambas sean del mismo color.

Ejercicio 3. En un grupo de 180 estudiantes 118 estudian inglés; 83, francés, y 45, ambos idiomas. Si se selecciona al azar un estudiante de este grupo y resulta que estudia que estudia ingles ¿Cuál es la probabilidad de que estudie francés?

Ejercicio 4. En una fábrica, los obreros trabajan tres turnos distintos. En el último año ocurrieron 200 accidentes. Algunos de estos se pueden atribuir a errores humanos, o en otros casos se relacionan con condiciones de trabajo. La Tabla 23 muestra el porcentaje de accidentes que ocurren en cada categoría en los archivos de la fábrica.

Tabla 23. *Datos del ejercicio 2 Sección 3.10*

Turno		Errores humanos	Condiciones de trabajo
	Diurno	12%	6%
	Vespertino	10%	18%
	Nocturno	14%	40%

Se elige al azar uno de estos casos de los 200. Si este accidente seleccionado resulta que se atribuye a error humano, ¿cuál es la probabilidad de que ocurra en el turno diurno?

Ejercicio 5. Conteste falso (F) o verdadero (V) según el caso:

a) Se lanza un dado una vez, si se definen los eventos A ocurre un número par, B ocurre un número menor que 2. Entonces los eventos A y B son independientes.

b) De un paquete de 52 cartas de póker se extrae una y se dice que es negra, la probabilidad de que sea un 4 es de 1/13.

c) En una urna hay 4 canicas rojas y 3 canicas verdes. Si se seleccionan 2 al azar y con reemplazo, la probabilidad que sean del mismo color es de 25/49.

d) La probabilidad de que llueva hoy es de 0.65, la probabilidad de que haya mucho tráfico dado de que llueva es de 0.74. Entonces, la probabilidad de que en hoy llueva y haya mucho tráfico es de 65/74.

Ejercicio 6. Un grupo de estudiantes presentan exámenes de cálculo y de inglés. La proporción de estudiantes que aprueban inglés es de 0.7 y la proporción de estudiantes que aprueban cálculo es de 0.3. Suponga que las notas en cada una de las asignaturas son independientes. Encuentre la probabilidad de que un estudiante escogido al azar de este grupo haya aprobado ambas asignaturas.

Ejercicio 7. Suponga que el 10% de las personas que hacen su declaración anual de impuestos cometen fraude al tributar, otro 35% cometen algún error al hacerla, y el resto hace bien su declaración. De los que hacen bien su declaración, 8% resultan con saldo a favor, 7% con saldo en cero, y 85% tendrán que pagar. En el caso de los que cometen algún error en hacer su declaración, el 40% tendrá que pagar, el 50% tendrá saldo a favor, y el 10% tendrán saldo en cero. Por otro lado, de los que cometen fraude, el 80% tendrá saldo en cero y el 20% saldo a favor. Si un auditor verifica una de estas declaraciones al azar.

a) ¿Cuál es la probabilidad de que tenga que pagar?
b) Si observa que la persona de esta declaración tiene saldo a favor, ¿cuál es la probabilidad de que le pongan una multa (si comete error o fraude)?

Ejercicio 8. En una ciudad hay tres compañías de gas A, B y C. La compañía A distribuye el 40%; la B, el 25%, y la C, el 35% del gas para la ciudad. Además, se sabe que la compañía A presenta fallas en un 10% de su distribución; la B, en un 8%, y la C, en un 3. Un auditor hace una inspección en una zona de la ciudad y ve que no hay suministro de gas, ¿cuál es la probabilidad de que haya sido la compañía C la responsable?

Ejercicio 9. En una caja hay dos tipos de monedas 2 monedas que por ambos lados es sello y 3 monedas corrientes, es decir, por un lado cara y por el otro, sello. Se selecciona de la caja una moneda al azar y se lanza, y se ve que el resultado de la parte superior de la moneda es sello. ¿Cuál es la probabilidad de que haya sido lanzada una moneda corriente?

Ejercicio 10. Suponga que generalmente, en una asignatura de una universidad, la probabilidad de que un estudiante la cancele, si la está tomando por primera vez, es de 0.31; la probabilidad de cancelarla si la está tomando por segunda vez es de 0.65 y en el caso de que la esté cursando por tercera o más veces es de 0.04. Suponga también que los porcentajes de estudiantes que están viendo en este semestre la asignatura por primera, segunda y tercera o más veces, respectivamente, es de 59%, 25% y el 16%. Si escoge un estudiante al azar de esta asignatura:

a) ¿Cuál es la probabilidad de que no la cancele?
b) Si el estudiante canceló la asignatura, ¿cuál es la probabilidad de que esté cursando por segunda vez?

Ejercicio 11. Una bolsa contiene 3 pelotas negras y 2 blancas. Se sacan de la bolsa dos pelotas, una después de la otra, sin reemplazo. *a)* ¿Cuál es la probabilidad de que sean del mismo color? *b)* ¿Cuál es la probabilidad de que la segunda pelota que saque sea blanca?

Ejercicio 12. Sean A y B dos eventos asociados a un experimento. Suponga que $P(A) = 0.5$, $P(A \cup B) = 0.7$ y $P(B) = w$. ¿Cuál debe ser el valor de w para que los eventos sean independientes?

Ejercicio 13. Dados A y B dos eventos de un experimento aleatorio. Si $P(A^c) = 1/5$, $P(B) = 1/4$, $P(A/B) = 1/3$, calcule $P(A \cup B)$.

Ejercicio 14. Dados A y B dos eventos de un experimento aleatorio. Si $P(A) = 1/3$, $P(B/A) = 1/3$, calcule $P(A \cap B^C)$.

Capítulo 4. Variables aleatorias discretas

1. VARIABLE ALEATORIA DISCRETA Y DISTRIBUCIÓN DE PROBABILIDAD

- **Experimento aleatorio** es un proceso que cumple con lo siguiente:
  1. Todos los posibles resultados del experimento se conocen de antemano.
  2. Cuando se hace el experimento no se sabe cuál resultado va a ocurrir.
  3. El experimento se puede repetir bajo idénticas condiciones.
- **Espacio muestral** es la colección de todos los posibles resultados de un experimento aleatorio. Se denota con las letras Ω o S.
- **Variable aleatoria X**. Una variable aleatoria X es una aplicación que asigna a cada elemento del espacio muestral un único número real.
- **Distribución discreta de probabilidad.** Una distribución discreta de probabilidad es el conjunto de pares de la forma $(x, P(X = x))$. Donde x son los valores que toma la variable aleatoria X y $P(X = x)$ es su respectiva probabilidad.

Ejemplo 4.1: experimento aleatorio lanzar una moneda corriente.

Espacio muestral: $\Omega = \{cara, sello\}$

Variable aleatoria X asignar "1" si es cara, asigne "0" si es sello.

La distribución de probabilidad se muestra en la Tabla 24.

Tabla 24. *Distribución de probabilidad Ejemplo 4.1*

X	0	1
$P(X = x)$	½	½

Ejemplo 4.2: experimento aleatorio lanzar un dado.

Espacio muestral $\boldsymbol{\Omega} = \{\mathbf{1, 2, 3, 4, 5, 6}\}$

Variable aleatoria X asignar el mismo valor.

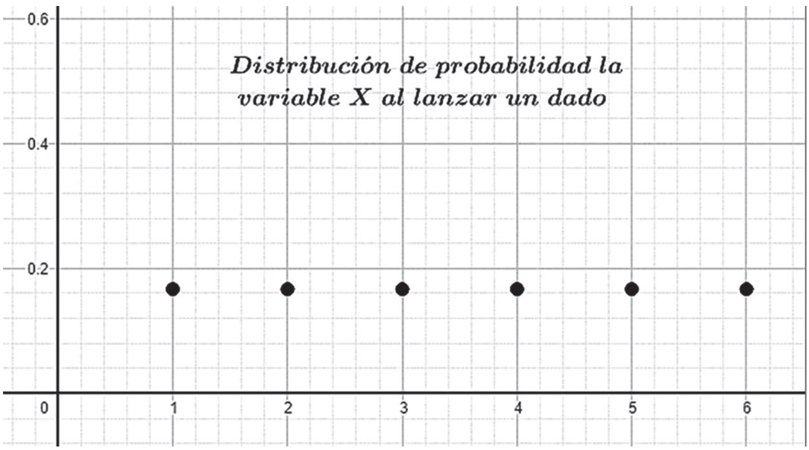
La distribución de probabilidad se muestra en la Tabla 25.

Tabla 25. *Distribución de probabilidad Ejemplo 4.2*

X	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

La Figura 105 muestra la distribución de probabilidad de la variable aleatoria X.

Figura 105. *Variable Aleatoria X. Lanzamiento de un dado.*



Fuente: elaboración propia.

Ejemplo 4.3: experimento aleatorio lanzar un dado.

Espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$

Variable aleatoria X asignar "1" si es 4, asigne "0" en caso contrario.

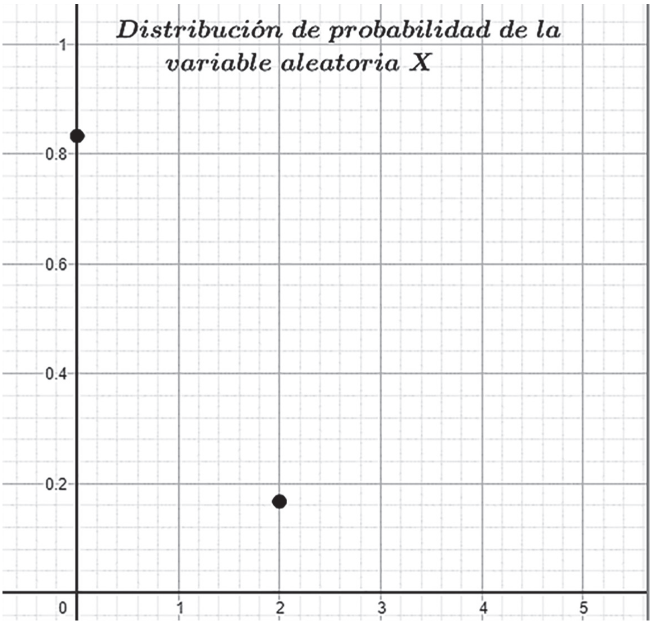
La distribución de probabilidad se muestra en la Tabla 26.

Tabla 26. *Distribución de probabilidad, Ejemplo 4.3*

X	0	1
$P(X = x)$	5/6	1/6

La Figura 106 muestra la distribución de probabilidad de la variable aleatoria X.

Figura 106. *Distribución de probabilidad de la variable X.*



Fuente: elaboración propia.

Ejemplo 4.4: experimento aleatorio lanzar dos dados.

Espacio muestral $\Omega = \begin{Bmatrix} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6) \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6) \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6) \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6) \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{Bmatrix}$

Variable aleatoria X es la suma de los valores de los dos dados.

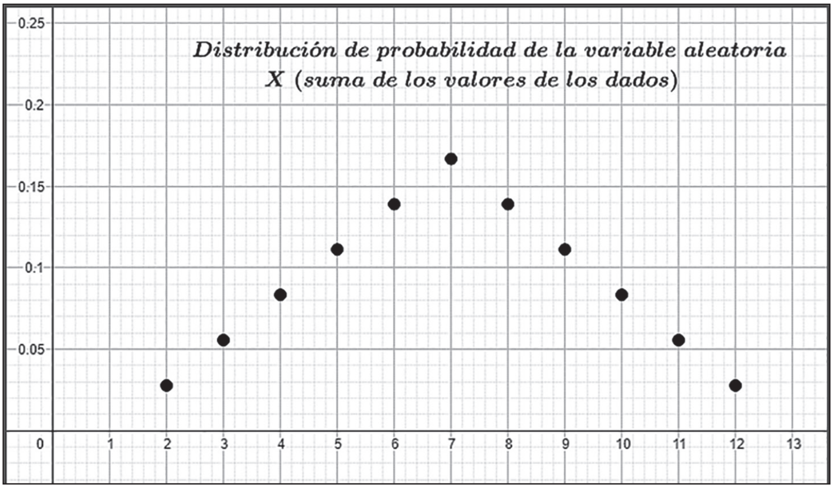
La distribución de probabilidad se muestra en la Tabla 27.

Tabla 27. *Distribución de probabilidad, Ejemplo 4.4*

Espacio muestral	Variable aleatoria X	Probabilidad
(1,1)	2	$P(X=2)=1/36$
(1,2)(2,1)	3	$P(X=3)=2/36$
(1,3)(2,2)(3,1)	4	$P(X=4)=3/36$
(1,4)(2,3)(3,2)(4,1)	5	$P(X=5)=4/36$
(1,5)(2,4)(3,3)(4,2)(5,1)	6	$P(X=6)=5/36$
(1,6)(2,5)(3,4)(4,3)(5,2)(6,1)	7	$P(X=7)=6/36$
(2,6)(3,5)(4,4)(5,3)(6,2)	8	$P(X=8)=5/36$
(3,6)(4,5)(5,4)(6,3)	9	$P(X=9)=4/36$
(4,6)(5,5)(6,4)	10	$P(X=10)=3/36$
(5,6)(6,5)	11	$P(X=11)=2/36$
(6,6)	12	$P(X=12)=1/36$

La Figura 107 muestra la distribución de probabilidad de la variable aleatoria X.

Figura 107. *Variable aleatoria X. Suma de los valores de los dos dados.*



Fuente: elaboración propia.

Ejemplo 4.5: experimento aleatorio lanzar dos dados.

Espacio muestral $\Omega = \begin{Bmatrix} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6) \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6) \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6) \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6) \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{Bmatrix}$

Variable aleatoria X la diferencia en valor absoluto de los valores de los dos dados.

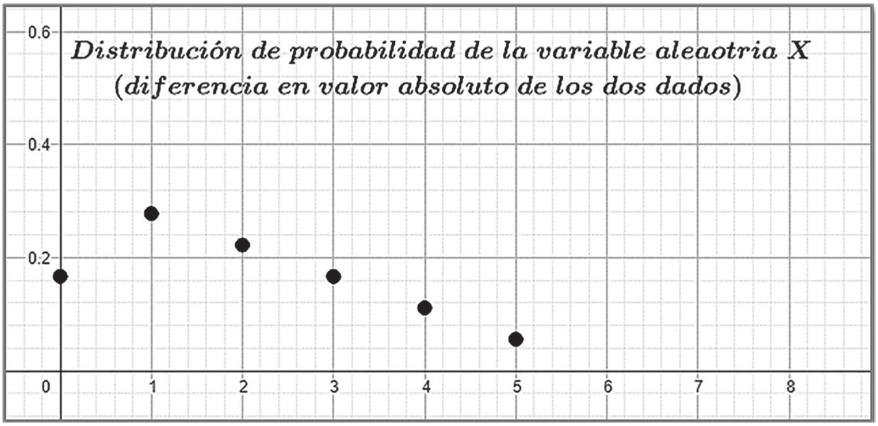
La distribución de probabilidad se muestra en la Tabla 28.

Tabla 28. *Distribución de probabilidad Ejemplo 4.5*

Espacio muestral	Variable aleatoria X	Probabilidad
(1,1)(2,2)(3,3)(4,4)(5,5)(6,6)	0	$P(X=0)=6/36$
(1,2)(2,1)(2,3)(3,2)(3,4)(4,3)(4,5)(5,4)(5,6)(6,5)	1	$P(X=1)=10/36$
(1,3)(3,1)(2,4)(4,2)(3,5)(5,3)(4,6)(6,4)	2	$P(X=2)=8/36$
(1,4)(4,1)(2,5)(5,2)(3,6)(6,3)	3	$P(X=3)=6/36$
(1,5)(5,1)(2,6)(6,2)	4	$P(X=4)=4/36$
(1,6)(6,1)	5	$P(X=5)=2/36$

La Figura 108 muestra la distribución de probabilidad de la variable aleatoria X.

Figura 108. *Variable aleatoria X. Diferencia en valor absoluto.*



Fuente: elaboración propia.

2. PROPIEDADES DE LAS DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

1. $0 \leq P(X = x) \leq 1$

2. $\sum_x f(x) = \sum_x P(X = x) = 1$

3. VALOR ESPERADO

El valor esperado, también conocido como media o esperanza matemática, es el valor promedio que se esperaría obtener si el experimento se repitiera un gran número de veces, para una variable aleatoria X se calcula como $\sum_{i=1}^{n} x_i * P(X = x_i)$, con $x_1, x_2, \dots, x_n$ valores que toma la variable aleatoria X y se denota como μ_X o $E(X)$.

3.1. Varianza

Es una medida de dispersión que indica qué tan lejos están los valores posibles de la variable aleatoria X respecto a su valor esperado (o media), para una variable aleatoria X se calcula como:

$$\sum_{i=1}^{n} (x_i - \mu_X)^2 * P(X = x_i) =$$

$$\left(\sum_{i=1}^{n} x_i^2 * P(X = x_i)\right) - \left(\sum_{i=1}^{n} x_i * P(X = x_i)\right)^2 = E(X^2) - [E(X)]^2$$

Y se denota como $V(X)$ o σ_X^2

Ejemplo 4.6: experimento aleatorio lanzar una moneda corriente.

Espacio muestral $\Omega = \{cara, sello\}$

Variable aleatoria X asignar “1” si es cara, asigne “0” si es sello.

La distribución de probabilidad se muestra en la Tabla 29.

Tabla 29. *Distribución de probabilidad Ejemplo 4.6*

X	0	1
$P(X = x)$	½	½

Con valor esperado $\mu_X = E(X) = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2} = 0.5$

Con varianza $V(X) = 0^2\left(\frac{1}{2}\right) + 1^2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)^2 = 0.25$

Ejemplo 4.7: experimento aleatorio lanzar un dado.

Espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$

Variable aleatoria X asignar el mismo valor.

La distribución de probabilidad se muestra en la Tabla 30.

Tabla 30. *Distribución de probabilidad Ejemplo 4.7*

X	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Con valor esperado $\mu_X = E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{21}{6}$

Con varianza $V(X) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) - \left(\frac{21}{6}\right)^2 = \frac{35}{12}$

Ejemplo 4.8: experimento aleatorio lanzar un dado.

Espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$

Variable aleatoria X asignar "1" si es 4, asigne "0" en caso contrario.

La distribución de probabilidad se muestra en la Tabla 31.

Tabla 31. *Distribución de probabilidad Ejemplo 4.8*

X	0	1
$P(X = x)$	5/6	1/6

Con valor esperado $\mu_X = E(X) = 0\left(\frac{5}{6}\right) + 1\left(\frac{1}{6}\right) = \frac{1}{6}$

Con varianza $V(X) = 0^2\left(\frac{1}{6}\right) + 1^2\left(\frac{1}{6}\right) - \left(\frac{1}{6}\right)^2 = \frac{5}{36}$

Ejemplo 4.9: experimento aleatorio lanzar dos dados.

Variable aleatoria X la suma de los valores de los dos dados.
La distribución de probabilidad se muestra en la Tabla 32.

Tabla 32. *Distribución de probabilidad Ejemplo 4.9*

X	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Con valor esperado

$$\mu_X = E(X) = 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + 4\left(\frac{3}{36}\right) + 5\left(\frac{4}{36}\right) + 6\left(\frac{5}{36}\right) + 7\left(\frac{6}{36}\right) + 8\left(\frac{5}{36}\right) + 9\left(\frac{4}{36}\right) + 10\left(\frac{3}{36}\right) + 11\left(\frac{2}{36}\right) + 12\left(\frac{1}{36}\right) = 7$$

Con varianza

$$V(X) = 2^2\left(\frac{1}{36}\right) + 3^2\left(\frac{2}{36}\right) + 4^2\left(\frac{3}{36}\right) + 5^2\left(\frac{4}{36}\right) + 6^2\left(\frac{5}{36}\right) + 7^2\left(\frac{6}{36}\right) + 8^2\left(\frac{5}{36}\right) + 9^2\left(\frac{4}{36}\right) + 10^2\left(\frac{3}{36}\right) + 11^2\left(\frac{2}{36}\right) + 12^2\left(\frac{1}{36}\right) - (7)^2 = \frac{35}{6}$$

Ejemplo 4.10: experimento aleatorio lanzar dos dados.

Variable aleatoria X la diferencia en valor absoluto de los valores de los dos dados.

La distribución de probabilidad se muestra en la Tabla 33.

Tabla 33. *Distribución de probabilidad Ejemplo 4.10*

X	0	1	2	3	4	5
$P(X = x)$	6/36	10/36	8/36	6/36	4/36	2/36

Con valor esperado

$$\mu_X = E(X) = 0\left(\frac{6}{36}\right) + 1\left(\frac{10}{36}\right) + 2\left(\frac{8}{36}\right) + 3\left(\frac{6}{36}\right) + 4\left(\frac{4}{36}\right) + 5\left(\frac{2}{36}\right) = \frac{35}{18}$$

Con varianza

$$V(X) = 0^2\left(\frac{6}{36}\right) + 1^2\left(\frac{10}{36}\right) + 2^2\left(\frac{8}{36}\right) + 3^2\left(\frac{6}{36}\right) + 4^2\left(\frac{4}{36}\right) + 5^2\left(\frac{2}{36}\right) - \left(\frac{35}{18}\right)^2 = \frac{665}{324}$$

3.2. Combinaciones lineales de variables aleatorias discretas

Ejemplo 4.11: sea X una variable aleatoria con distribución de probabilidad que se muestra en la Tabla 34.

Tabla 34. *Distribución de probabilidad variable X Ejemplo 4.11*

X	1	2	3
$P(X = x)$	0.3	0.5	0.2

Con valor esperado $\mu_X = E(X) = 1(0.3) + 2(0.5) + 3(0.2) = 1.9$

Con varianza $V(X) = 1^2(0.3) + 2^2(0.5) + 3^2(0.2) - (1.9)^2 = 0.49$

Ahora, dada la variable aleatoria Y como función lineal de la variable aleatoria X, expresada como:

$$Y = 2X + 5$$

Entonces, la distribución de probabilidad (Tabla 35) de la variable Y se expresa como:

Tabla 35. *Distribución de probabilidad variable Y, Ejemplo 4.11*

Y	7	9	11
$P(Y=y)$	0.3	0.5	0.2

Con valor esperado $\mu_X = E(X) = 7(0.3) + 9(0.5) + 11(0.2) = 8.8$

Con varianza $V(X) = 7^2(0.3) + 9^2(0.5) + 11^2(0.2) - (8.8)^2 = 1.96$

También se puede calcular de la siguiente forma dada una variable aleatoria X, con valor esperado $E(X)$ y varianza $V(X)$. Dada una variable aleatoria Y, expresada como:

$$Y = aX + b$$

Entonces, el valor esperado de Y es:

$$E(Y) = E(aX + b)$$
$$E(Y) = aE(X) + b$$

Entonces, la varianza de Y es:

$$V(Y) = V(aX + b)$$
$$V(Y) = a^2V(X)$$

Como $E(X) = 1.9$ y $V(X) = 0.49$. Aplicando fórmulas, se tiene:

$$Y = 2X + 5$$
$$E(Y) = 2E(X) + 5 = 2(1.9) + 5 = 8.8$$
$$V(Y) = 2^2V(X) = 4(0.49) = 1.96$$

Dando el mismo resultado.

5. EJERCICIOS

Ejercicio 1. En un paquete de 52 cartas se seleccionan 2, con reemplazo, a) encuentre la distribución de probabilidad para la variable aleatoria X que representa el número de figuras (figuras J, Q y K); b) haga la gráfica de la distribución de probabilidad para la variable aleatoria *X* que representa el número de figuras (figuras J, Q y K).

Ejercicio 2. En una caja hay 2 billetes de \$50.000, 2 de \$20.000 y 3 de \$10.000. Si se seleccionan al azar 2 billetes sin reemplazo, *a)* encuentre la distribución de probabilidad para la variable aleatoria *Y* que representa el valor total de los billetes. *b)* ¿Cuál es la probabilidad de que el valor total de los billetes sea mayor a \$30.000?

Ejercicio 3. En una caja de 10 bombillos hay 2 defectuosos. Un comprador decide seleccionar al azar 3 bombillos, *a)* Encuentre la distribución de probabilidad para la variable aleatoria X que representa el número de bombillos buenos, *b)* ¿Cuál es la probabilidad de que al menos un bombillo de los tres seleccionados resulte defectuoso?

Ejercicio 4. La distribución de probabilidad de una variable aleatoria X está definida por la siguiente función f:

$$f(x) = \binom{3}{x}\left(\frac{1}{6}\right)^{x}\left(\frac{5}{6}\right)^{3-x} \quad x = 0, 1, 2,3$$

Encuentre *a)* el valor esperado de la variable aleatoria X y *b)* la varianza de X.

Ejercicio 5. Una moneda no legal tiene el doble de probabilidad de que salga cara en vez de sello. *a)* Hallar el valor esperado de sellos cuando la moneda se lanza tres veces. *b)* Hallar la varianza para el número de sellos cuando la moneda se lanza tres veces.

Ejercicio 6. Una empresa produce paquetes de clips. El número de clips por paquete varía, como indica la Tabla 36.

Tabla 36. *Datos ejercicio 6 sección 4.6.*

Número de clips	47	48	49	50	51	52	53
Probabilidad	0,04	0,13	0,21	0,29	0,20	0,10	0,03

El costo en pesos de producir un paquete de clips es de $1600 + 100X$, donde X es el número de clips que hay en el paquete. *a)* Halle el valor esperado y la desviación típica para el número de clips. *b)* Halle el valor esperado y la desviación típica del costo de producción de los paquetes de clips.

Ejercicio 7. Una persona participa en un concurso de televisión con las siguientes reglas:

- Si contesta correctamente a una pregunta con 4 opciones de respuestas posibles (sólo una correcta) gana \$ 4´000.000.
- En caso contrario, se le propone una segunda pregunta con 2 respuestas posibles (sólo una correcta). Si acierta gana \$1´000.000 y si falla no gana nada.
- El juego termina cuando la persona acierta, o tras fallar la segunda pregunta.

Si un concursante contesta al azar, calcule el valor de la ganancia esperada.

Ejercicio 8. Suponga que la proporción del número de automóviles que se devuelven en un mes determinado durante el periodo de garantía en una empresa automotriz está dada por la Tabla 37.

Tabla 37. *Datos del ejercicio 8 de la sección 4.6*

X	0	1	2	3	4	5
$P(X = x)$	0.8	0.1	0.04	0.02	0.025	0.015

a) Halle el valor esperado y la varianza para del número de automóviles que se devuelven durante ese mes.

b) Si el costo que cobra un mecánico por la valoración del daño del automóvil está dado por $C(X) = 400.000X + 600.000$, *a)* Encuentre la ganancia esperada del mecánico durante ese mes. *b)* Encuentre la varianza de lo cobrado por el mecánico durante ese mes.

Capítulo 5. Algunas distribuciones discretas de probabilidad

1. DISTRIBUCIÓN UNIFORME

Una variable aleatoria discreta X se distribuye de manera uniforme si todos los valores posibles de X tienen la misma probabilidad de ocurrencia. La variable X con valores $x_1, x_2, x_3, \dots, x_n$ se distribuye de manera uniforme si función de masa de probabilidad puede expresarse como:

$$f(x) = P(X = x) = \frac{1}{n}, x = x_1,\ x_2,\ x_3, \dots, x_n$$

Con valor esperado $\mu_X = E(X) = \frac{\sum_{i=1}^{n} x_i}{n}$

Con varianza $\sigma_X^2 = V(X) = \frac{\sum_{i=1}^{n}(x_i - \mu_X)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \mu_X^2$

Ejemplo 5.1: experimento aleatorio lanzar una moneda corriente.

Espacio muestral $\Omega = \{cara, sello\}$.

Variable aleatoria X asignar “1” si es cara, asigne “0” si es sello.

La distribución de probabilidad se muestra en la Tabla 38.

Tabla 38. *Distribución de probabilidad Ejemplo 5.1*

X	0	1
$P(X = x)$	½	½

Con valor esperado $\mu_X = E(X) = \frac{0+1}{2} = 0.5$.

Con varianza $V(X) = \frac{0^2+1^2}{2} - 0.5^2 = 0.5 - 0.25 = 0.25.$

Ejemplo 5.2: experimento aleatorio lanzar un dado.

Espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$

Variable aleatoria X asignar el mismo valor.

La distribución de probabilidad se muestra en la Tabla 39.

Tabla 39. *Distribución de probabilidad Ejemplo 5.2*

X	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Con valor esperado $\mu_X = E(X) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6}$

Con varianza $V(X) = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} - \left(\frac{21}{6}\right)^2 = \frac{35}{12}$

2. DISTRIBUCIÓN DE BERNOULLI

Es una distribución de probabilidad discreta que describe experimentos aleatorios en los que solo existen dos posibles resultados mutuamente excluyentes generalmente denominados como "éxito" y "fracaso" y cuyas probabilidades son p y q, respectivamente. La variable aleatoria X toma dos valores, cero "0" y uno "1", donde "0" es el fracaso y "1" es el éxito. La función de distribución de masa de probabilidad se expresa en la Tabla 40.

Tabla 40. *Distribución de probabilidad Bernoulli*

X	0	1
$P(X = x)$	q	p

Donde p es la probabilidad del éxito y q es la probabilidad del fracaso. Donde $p + q = 1$.

Con valor esperado $\mu_X = E(X) = 0 * q + 1 * p = p$

Con varianza $\sigma_X^2 = V(X) = 0^2 * q + 1^2 * p - p^2 = p - p^2 = p(1-p) = pq$

Ejemplo 5.3: experimento aleatorio lanzar una moneda corriente.

Espacio muestral $\Omega = \{cara, sello\}$

Variable aleatoria X asignar "1" si es cara, asigne "0" si es sello.

La distribución de probabilidad se muestra en la Tabla 41.

Tabla 41. *Distribución de probabilidad Ejemplo 5.3*

X	0	1
$P(X = x)$	½	½

Con valor esperado $\mu_X = E(X) = p = 0.5$

Con varianza $V(X) = pq = (0.5)(0.5) = 0.25$

Ejemplo 5.4: experimento aleatorio lanzar un dado.

Espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$

Variable aleatoria X asignar "1" si es 4, asigne "0" en caso contrario.

Distribución de probabilidad se muestra en la Tabla 42.

Tabla 42. *Distribución de probabilidad Ejemplo 5.4*

X	0	1
$P(X = x)$	5/6	1/6

Con valor esperado $\mu_X = E(X) = p = 1/6$

Con varianza $V(X) = pq = (1/6)(5/6) = 5/36$

3. DISTRIBUCIÓN BINOMIAL

Una variable aleatoria X se distribuye de manera binomial si X es el número de éxitos en n ensayos de Bernoulli independientes (WALPOLE, *et al.,* 2012). Su distribución de probabilidad se expresa como:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x}; x = 0,1,2, \ldots, n$$

Donde p es la probabilidad del éxito y q es la probabilidad del fracaso. Donde $p + q = 1$.

Valor esperado $E(X) = np$

Varianza $V(X) = np(1-p) = npq$

Ejemplo 5.5: experimento aleatorio lanzar una moneda "cargada" 3 veces. Donde la probabilidad de que salga cara es el doble de la de sello, es decir, $P(C) = 2P(S)$.

Como $P(C) + P(S) = 1$. Reemplazando, $P(C) = 2P(S)$, se tiene:

$$2P(S) + P(S) = 1$$

$$3P(S) = 1$$

$$P(S) = 1/3$$

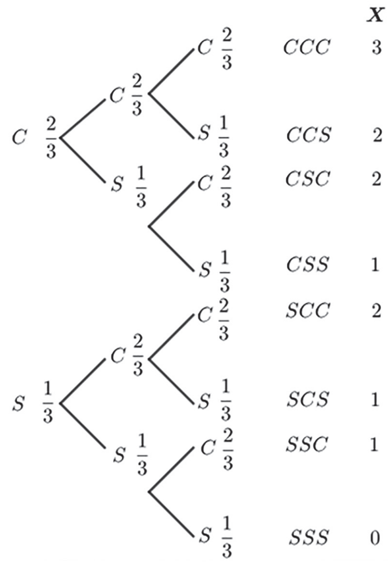
$$P(C) = 2/3$$

Espacio muestral $\Omega = \{CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS\}$

Variable aleatoria X el número de caras que salen al lanzar esa moneda 3 veces.

El diagrama de árbol está representado en la Figura 109.

Figura 109. *Diagrama de árbol lanzamiento de una moneda cargada, 3 veces.*



Fuente: elaboración propia.

La distribución de probabilidad se muestra en la Tabla 43.

Tabla 43. *Distribución de probabilidad Ejemplo 5.5*

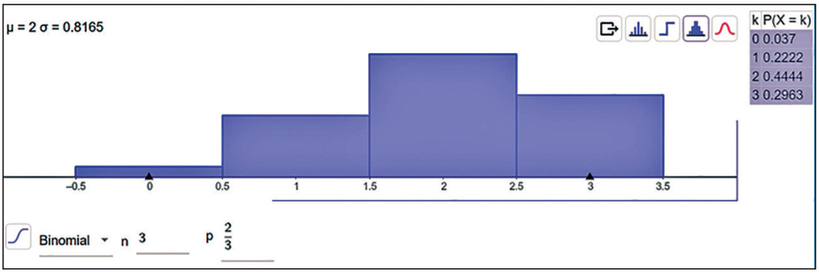
X	0	1	2	3
$P(X=x)$	$\left(\frac{1}{3}\right)^3$	$3\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1$	$3\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2$	$\left(\frac{2}{3}\right)^3$
$P(X=x)$	$\binom{3}{0}\left(\frac{1}{3}\right)^3$	$\binom{3}{1}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1$	$\binom{3}{2}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2$	$\binom{3}{3}\left(\frac{2}{3}\right)^3$
$P(X=x)$	$\binom{3}{0}q^3$	$\binom{3}{1}q^2p$	$\binom{3}{2}qp^2$	$\binom{3}{3}p^3$

Con valor esperado $\mu_X = E(X) = np = 3\left(\frac{2}{3}\right) = 2$

Con varianza $V(X) = npq = 3\left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{2}{3}$

Su gráfica en GeoGebra se muestra en la Figura 110.

Figura 110. *Gráfico de la distribución de probabilidad, Ejemplo 5.5.*



Fuente: elaboración propia.

Ejemplo 5.6: experimento aleatorio lanzar un dado ordinario 5 veces.

Elementos del espacio muestral $6^5 = 7776$

Variable aleatoria X el número de veces que sale el número 4 al lanzar 5 veces el dado.

La distribución de probabilidad se muestra en la Tabla 44.

Tabla 44. *Distribución de probabilidad Ejemplo 5.6*

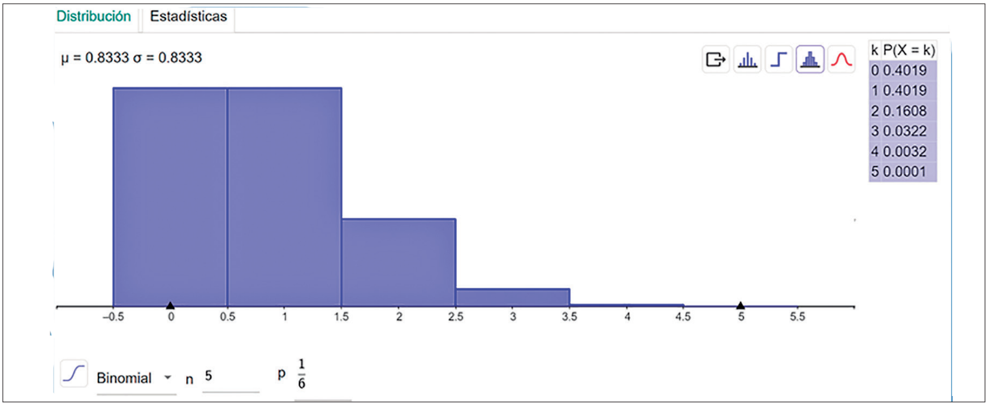
X	0	1	2	3	4	5
$P(X=x)$	$\binom{5}{0}\left(\frac{5}{6}\right)^5$	$\binom{5}{1}\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)^1$	$\binom{5}{2}\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)^2$	$\binom{5}{3}\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)^3$	$\binom{5}{4}\left(\frac{5}{6}\right)^1\left(\frac{1}{6}\right)^4$	$\binom{5}{5}\left(\frac{1}{6}\right)^5$

Con valor esperado $\mu_X = E(X) = np = 5\left(\frac{1}{6}\right) = \frac{5}{6}$

Con varianza $V(X) = npq = 5\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = \frac{25}{36}$

Su gráfica y distribución de probabilidad en GeoGebra se muestra en la Figura 111.

Figura 111. *Gráfico de la distribución de probabilidad Ejemplo 5.6.*



Fuente: elaboración propia.

4. DISTRIBUCIÓN HIPERGEOMÉTRICA

Una variable X se distribuye en forma hipergeométrica si X es el número de éxitos en una muestra de tamaño "n" seleccionada sin reemplazo de un conjunto "N", en donde hay "k" éxitos y "$N-k$" fracasos. La fórmula de la distribución de probabilidad es:

$$f(x) = P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, max\{0, n-(N-k)\} \leq x \leq min\{n, k\}$$

Valor esperado $E(X) = \frac{nk}{N}$

Varianza $V(X) = \left(\frac{N-n}{N-1}\right)\left(\frac{nk}{N}\right)\left(1-\frac{k}{N}\right)$

Ejemplo 5.7: experimento aleatorio se tiene un paquete de 52 cartas. Se sacan dos cartas (sin reemplazo) y se mira el color de la carta.

Espacio muestral es $\Omega = \{RR, RN, NR, NN\}$. Representado este espacio muestral en el diagrama de árbol de la Figura 95.

Variable aleatoria X el número de cartas rojas.

La distribución de probabilidad se muestra en la Tabla 45.

Tabla 45. *Distribución de probabilidad Ejemplo 5.7*

X	0	1	2
$P(X=x)$	$\left(\frac{26}{52}\right)\left(\frac{25}{51}\right)=\frac{25}{102}$	$2\left(\frac{26}{52}\right)\left(\frac{26}{51}\right)=\frac{26}{51}$	$\left(\frac{26}{52}\right)\left(\frac{25}{51}\right)=\frac{25}{102}$
$P(X=x)$	$\frac{\binom{26}{0}\binom{26}{2}}{\binom{52}{2}}=\frac{25}{102}$	$\frac{\binom{26}{1}\binom{26}{1}}{\binom{52}{2}}=\frac{26}{51}$	$\frac{\binom{26}{2}\binom{26}{0}}{\binom{52}{2}}=\frac{25}{102}$

Donde $N=52$, $n=2$, $k=26$ y $N-k=26$.

También podemos ver que los valores que toma X se obtienen:

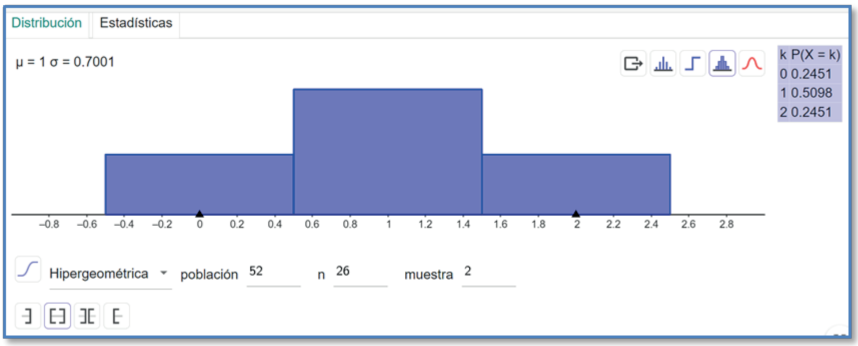
$$max\{0,2-26\} \leq x \leq \min\{2,26\}$$
$$max\{0,-24\} \leq x \leq \min\{2,26\}$$
$$0 \leq x \leq 2$$

Valor esperado $E(X)=\frac{2(26)}{52}=1$

Varianza $V(X)=\left(\frac{52-2}{52-1}\right)\left(\frac{2(26)}{52}\right)\left(1-\frac{26}{52}\right)=\frac{25}{51}$

Su gráfica y distribución de probabilidad en GeoGebra se muestra en la Figura 112.

Figura 112. *Gráfico de la distribución de probabilidad Ejemplo 5.57.*



Fuente: elaboración propia.

Ejemplo 5.8: experimento aleatorio se tiene una urna que contiene 3 bolas blancas y 2 bolas negras. De esta urna se seleccionan 4 bolas (sin reemplazo).

Variable aleatoria X el número de bolas blancas. El éxito es que sea bola blanca.

En este ejercicio $N = 5$, $n = 4$, $k = 3$ y $N - k = 2$.

Podemos observar que los valores que toma X son:

$$max\{0{,}4 - 2\} \leq x \leq \min\{4{,}3\}$$

$$max\{0{,}2\} \leq x \leq \min\{4{,}3\}$$

$$2 \leq x \leq 3$$

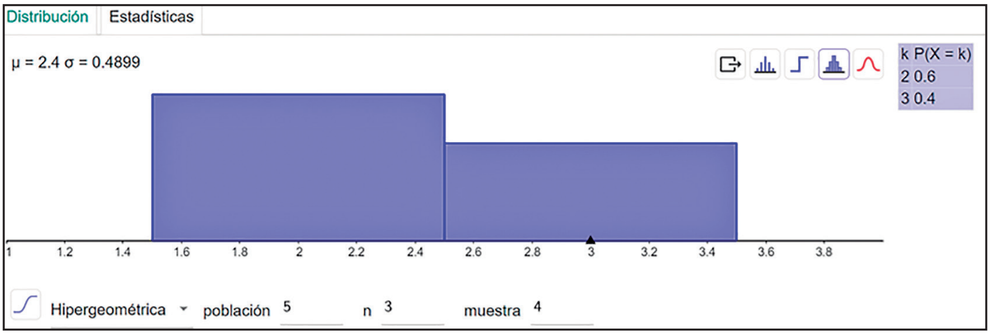
La distribución de probabilidad se muestra en la Tabla 46.

Tabla 46. *Distribución de probabilidad Ejemplo 5.8*

X	2	3
$P(X = x)$	$\frac{\binom{3}{2}\binom{2}{2}}{\binom{5}{4}} = 0.6$	$\frac{\binom{3}{3}\binom{2}{1}}{\binom{5}{4}} = 0.4$

Su gráfica y distribución de probabilidad en GeoGebra se muestra en la Figura 113.

Figura 113. *Gráfico de la distribución de probabilidad Ejemplo 5.8.*



Fuente: elaboración propia.

5. DISTRIBUCIÓN DE POISSON

Una variable aleatoria X se distribuye de forma de Poisson si X es el número de éxitos que ocurren en un intervalo de tiempo, área, volumen, longitud, entre otros. La función de masa de probabilidad se expresa como:

$$f(x) = P(X = x) = \frac{e^{-(\lambda t)}(\lambda t)^x}{x!} \quad x = 0,1,2,\ldots$$

Valor esperado $E(X) = \lambda t$

Varianza $V(X) = \lambda t$

Donde λ es el número de éxitos por unidad de tiempo, área, volumen, longitud, etc.
t es el tiempo, área, volumen, longitud, entre otros.

Ejemplo 5.9: el número promedio de accidentes que ocurren en un mes en cierta intersección es de 0.2 y la variable se distribuye en forma de Poisson.

a) ¿Cuál es la probabilidad de que en dos meses ocurran 3 accidentes?

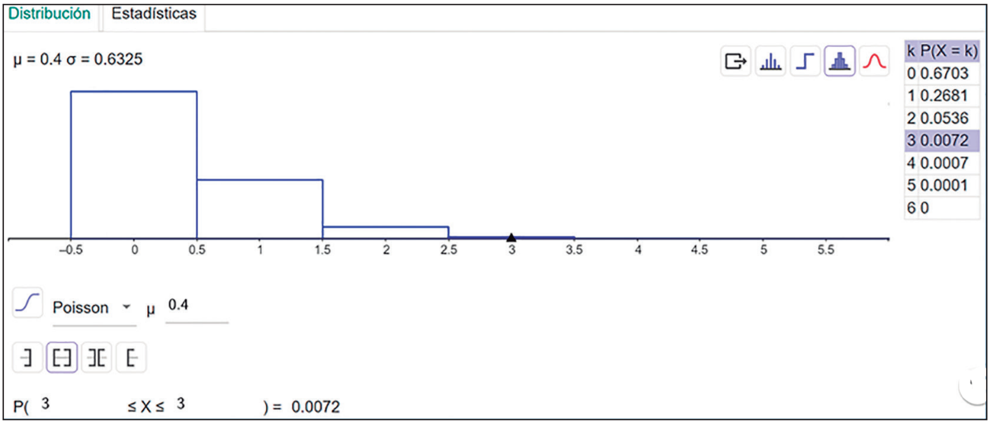
Para este caso $\lambda = \frac{0.2\ accidentes}{1\ mes}$ y $t = 2\ meses$

Entonces $\lambda * t = 0.4\ accidentes$

$$P(X = 3) = \frac{e^{-(0.2*2)}(0.2*2)^3}{3!} = \frac{e^{-(0.4)}(0.4)^3}{3!} \approx 0.00715 \approx 7.2 \times 10^{-3}$$

En GeoGebra (Figura 114).

Figura 114. *Cálculo de probabilidades Ejemplo 5.9. Distribución de Poisson.*



Fuente: elaboración propia.

b) ¿Cuál es la probabilidad de que en dos meses ocurran más de 2 accidentes?

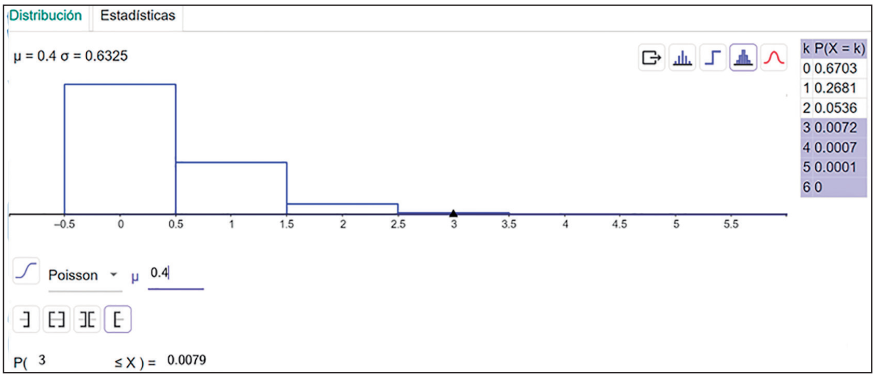
$$P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

$$1 - \left[\frac{e^{-(0.4)}(0.4)^0}{0!} + \frac{e^{-(0.4)}(0.4)^1}{1!} + \frac{e^{-(0.4)}(0.4)^2}{2!}\right]$$

$$1 - e^{-0.4}\left[\frac{0.4^0}{0!} + \frac{0.4^1}{1!} + \frac{0.4^2}{2!}\right] \approx 0.00793 = 7.93 \times 10^{-3}$$

En GeoGebra (Figura 115).

Figura 115. *Cálculo de probabilidades Ejemplo 5.9.*



Fuente: elaboración propia.

6. EJERCICIOS

Ejercicio 1. El 8% de las botellas de una caja de 50 botellas presentan algún defecto. Si un auditor toma una muestra de forma independiente de 40 botellas, *a)* Hallar la probabilidad de que existan 35 botellas sin defectos, *b)* Hallar la probabilidad de que existan por lo menos 3 botellas con defectos.

Ejercicio 2. El 8% de las botellas de una caja de 50 botellas presentan algún defecto. Si un auditor toma una muestra de forma dependiente de 40 botellas, *a)* Hallar la probabilidad de que existan 35 botellas sin defectos *b)* Hallar la probabilidad de que existan por lo menos 3 botellas con defectos.

Ejercicio 3. En la producción de cierto artículo se sabe que el 86% de 50 artículos el terminado es el adecuado. Si se toma una muestra de 12 artículos, la probabilidad a) de que ninguno tenga el terminado no adecuado**,** b) once tengan un terminado adecuado.

Ejercicio 4. De un total de 20 estudiantes que presentaron un examen de estadística, el 75% aprobó el examen. Si se selecciona una muestra de siete exámenes de forma independiente, *a)* ¿cuál es la probabilidad de que todos tengan nota de aprobado?, *b)* ¿cuál es la probabilidad de que más de tres exámenes presenten nota de no aprobado?

Ejercicio 5. En un concurso familiar de televisión, el concursante lanza dos dados grandes y el presentador le dice al concursante que gana si en la suma total resulta un número mayor a 9 y que, en caso contrario, pierde. Si el concursante tiene la oportunidad de lanzar los dos dados 12 veces, *a)* ¿cuántas veces, en promedio, espera ganar?, *b)* ¿cuál es la varianza de la variable aleatoria?, *c)* ¿cuál es la probabilidad de que gane más de una vez?

Ejercicio 6. En 10 registros de pacientes de un hospital hay 2 que indican casos de cierta enfermedad. Si se seleccionan 4 registros de forma independiente, *a)* ¿cuál es la probabilidad de que ninguno presente el caso de enfermedad?, *b)* ¿cuál es la probabilidad de que por mucho uno no presente el caso de enfermedad?

Ejercicio 7. Suponga que una secretaria comete en promedio 1 error por cada 5 páginas que escribe. Para las siguientes 300 páginas, *a)* encuentre la probabilidad de que cometa 10, 11 y 12 errores, *b)* encuentre la media y la varianza de la variable aleatoria X que representa el número de errores cometidos.

Ejercicio 8. Una compañía recibe un envío de un determinado artículo. Se comprueba una muestra aleatoria de 40 de estos de forma independiente y se acepta el envío si resultan defectuosos menos de tres artículos en esta muestra. Calcular la probabilidad de que se acepte el envío que contiene, *a)* un 1% de los artículos defectuosos, *b)* un 10% de artículos defectuosos.

Ejercicio 8. Una empresa que alquila lavadoras cuenta con 50 de ellas. Por información que se tiene, se sabe que seis de estas no se encuentran en buenas condiciones. Si un cliente selecciona en forma aleatoria 5 lavadoras, ¿cuál es la probabilidad, *a)* de que menos de dos de ellas se encuentren en buenas condiciones, *b)* más de tres no se encuentren en buenas condiciones?

Ejercicio 9. Se sacan 5 cartas de un paquete de 52 cartas. Encuentre el valor esperado y la varianza para la variable aleatoria X que representa el número de cartas rojas, si *a)* las cartas se seleccionan con sustitución, *b)* las cartas se seleccionan sin sustitución.

Referencias

Anderson, D., Sweeney, D., & Wlliams, T. (2009). *Estadística para administración y economía.* México Cengage Learning.

Badii, M., & Guillen, A. (2010). Esenciales de la estadística un acercamiento descriptivo. *Daena International Journal of Good Conscience*, 29.

Barahona, F., Barrera, O., Vaca, B., & Hidalgo, B. (2015). GeoGebra para la enseñanza de la matemática y su incidencia en el rendimiento académico estudiantil. *Revista Tecnológiva ESPOR*.

Batanero, C. (2001). *Didactica de la estadística.* Granada, España Universidad de Granada.

Batanero, C. (2013). Sentido estadístico componentes y desarrollo. *Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatoria*, 8.

Batanero, C., Cid, E., & Godino, J. (2003). *Sistemas numéricos y su didáctica para maestros.* Granada Universidad de Granada.

Behar, R., & Grima, P. (2010). *55 respuestas a dudas típicas de estadística.* Cali Ediciones Díaz de Santos.

Blanco, L. (2010). *Probabilidad.* Bogotá D.C Universidad Nacional de Colombia.

Devore, J. (2005). *Probabilidad y estadística para ingeniería y ciencias.* México Thomson.

Estadística para todos. (2023). Obtenido de estadística para todos https//www.estadisticaparatodos.es/taller/montyhall/montyhall.html

Guilford, J., & Fruchter, B. (1984). Estadística aplicada a la psicología y a la educación. En G. J., & B. Fruchter, *Estadística aplicada a la psicología y a la educación* (pág. 16). México Mac Graw Hill.

Martinex, C. (2012). *Estadística y muestreo.* Bogotá ECOE Ediciones.

Newbold, P., Carlson, W., & Thorne, B. (2010). *Estadística para administración y economía.* Madrid Pearson. Prentice Hill.

Pérez, C. (2005). *Muestreo estadístico. Conceptos y problemas resueltos.* Madrid Pearson. Prentice Hill.

Varios (Julio de 2023). *The R Project for Statistical Computing*. Obtenido de https//www.r-project.org/

Velasco, G., & Wisniewski, P. (2001). Probabilidad y estadística para ingeniería y ciencias. México Thomson Learning.

Glosario

Asimetría: Es una medida que indica la falta de simetría en la distribución de un conjunto de datos respecto a la media.

Asimetría negativa: O sesgo a la izquierda, se da cuando la cola de la distribución de un conjunto de datos se alarga para los valores inferiores a la media.

Asimetría positiva: O sesgo a la derecha, se da cuando la cola de la distribución se alarga para los valores superiores a la media.

Cardinalidad de un conjunto: La cardinalidad de un conjunto A es el número de elementos que hay en el conjunto A. Se denota como $n(A)$.

Coeficiente de variación: Es una medida de dispersión relativa que permite comparar la variabilidad de dos conjuntos de datos, incluso si tienen diferentes unidades de medida. Se calcula dividiendo la desviación estándar entre la media y multiplicando el resultado por 100, para expresarlo como porcentaje.

Conjunto: Un conjunto es una colección de elementos u objetos bien definidos.

Conjuntos mutuamente excluyentes: Son conjuntos que no tienen elementos en común. Si los conjuntos A y B no tienen elementos en común, la intersección entre ellos es el conjunto vacío.

Complemento: Es el conjunto de todos los elementos que pertenecen al universo, pero que no están en el conjunto dado. $A^C = \{xx \notin A \wedge x \in \mathcal{U}\}$ (se lee son los x tales que x pertenecen al conjunto A y al conjunto B).

Cuartiles: Es una medida de posición, porque divide la distribución de un conjunto de datos ordenados de menor a mayor en cuatro partes iguales, de modo que cada parte contiene el 25% de los datos.

Curtosis: La curtosis es una medida que indica cuan aplanada o empinada es la distribución de un conjunto de datos con respecto a la distribución gaussiana.

Desviación estándar o típica de una variable estadística: Es una medida de dispersión absoluta que indica cuánto se desvían, en promedio, los valores de un conjunto de datos con respecto a la media. Se calcula como la raíz cuadrada de la varianza. La desviación muestral se denota como s y la poblacional como σ.

Diagramas circulares o de torta: Es un gráfico circular que se usa para representar la proporción o porcentaje de los elementos de cada categoría de una variable cualitativa.

Diagrama de barras: El diagrama de barras es un gráfico bidimensional que muestra las categorías de una variable cualitativa en un eje y sus frecuencias absolutas o relativas en el otro, usando barras rectangulares cuya altura es proporcional a la frecuencia de cada categoría. Este gráfico facilita la comparación visual entre las categorías representadas.

Diagrama de caja y bigotes: Es una representación visual de los cuartiles y de los valores máximo y mínimo de un conjunto de datos para describir su dispersión y simetría. La caja abarca desde el primer cuartil q_1 hasta el tercer cuartil q_3, y contiene el 50% de los datos centrales, y dentro de la caja se ve por medio de una línea el cuartil dos q_2 o mediana. Los bigotes son líneas que se extienden desde los extremos de la caja hasta el valor mínimo y el valor máximo, que no se consideran atípicos.

Diagrama de puntos o frecuencias: Es un gráfico que representa los valores individuales de una variable cuantitativa mediante puntos. El eje horizontal representa la escala de valores de los datos y cada valor del dato se grafica con un punto apilándose verticalmente si el valor del dato se repite.

Diagrama de tallos y hojas: El diagrama de tallos y hojas es una gráfica visual para organizar un conjunto de datos con valores numéricos de dos o más dígitos, en donde se representan los tallos (generalmente la parte inicial o los primeros dígitos) y las hojas (generalmente la parte final o los últimos dígitos).

Distribución binomial: Una variable aleatoria X, se distribuye de manera binomial si X es el número de éxitos en n ensayos de Bernoulli independientes. Su función de masa de probabilidad es

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x}; x = 0,1,2, \ldots, n.$$

Distribución de Bernoulli: Es una distribución de probabilidad discreta que describe experimentos aleatorios en los que solo existen dos posibles resultados mutuamente excluyentes, generalmente denominados "éxito" y "fracaso" y cuyas probabilidades son p y q, respectivamente.

Distribución discreta de probabilidad: Una distribución discreta de probabilidad es el conjunto de pares de la forma $(x, P(X = x))$, representados mediante una tabla o un gráfico. Donde x son los valores que toma la variable aleatoria X y $P(X = x)$ su respectiva probabilidad.

Distribución discreta uniforme: Una variable aleatoria discreta X se distribuye de manera uniforme si todos los valores posibles de X tienen la misma probabilidad de ocurrencia.

Distribución hipergeométrica: Una variable X se distribuye en forma hipergeométrica si X es el número de éxitos en una muestra de tamaño "n" seleccionada sin reemplazo de un conjunto "N", en donde hay "k" éxitos y "$N - k$" fracasos. Su función de masa de probabilidad es

$$f(x) = P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, max\{0, n - (N - k)\} \leq x \leq min\{n, k\}$$

Distribución de Poisson: Una variable aleatoria X se distribuye de forma de Poisson si X es el número de éxitos que ocurren en un intervalo de tiempo, área, volumen o longitud, entre otros. La función de masa de probabilidad se expresa como

$$f(x) = P(X = x) = \frac{e^{-(\lambda t)}(\lambda t)^x}{x!} \quad x = 0,1,2,\ldots$$

Estadística: Ciencia que se ocupa de describir el comportamiento de poblaciones o muestras a través de patrones, tendencias y relaciones.

Estadística descriptiva: Fase de la estadística que se ocupa de organizar y resumir la información observable de variables, tanto cualitativas como cuantitativas, obtenidas de poblaciones o muestras. Cuando se trabaja con muestras, no permite hacer generalizaciones acerca de la población de estudio.

Estadística inferencial: Fase de la estadística que permite hacer generalizaciones sobre el comportamiento de una población a partir de una muestra, usando la probabilidad como herramienta fundamental para apoyar la toma de decisiones.

Escala de intervalo: Cuando la variable estadística permite tener orden entre sus observaciones y además cuenta con una unidad fija de medida. Es decir, las observaciones permiten obtener el intervalo o "distancia" entre ellas, y no posee un cero absoluto, sino relativo.

Escalas de medición: Forma de clasificación de las variables mediante la cantidad de información que contiene cada una de estas según su descripción.

Escala de razón: Cuando la variable estadística tiene propiedades de intervalo y existe un cero real o absoluto que permita considerar cocientes de mediciones o razones.

Escala nominal: Cuando la variable estadística se describe por medio de rótulos, nombres o categorías que permiten identificar un atributo del elemento de la muestra.

Escala ordinal: Cuando la variable estadística tiene propiedades de datos nominales. Es decir, rótulos, nombres o categorías y, además, admite el orden entre estas.

Espacio muestral: Es el conjunto de todos los posibles resultados de un experimento aleatorio. Se denota con las letras Ω o S.

Evento: Es un conjunto de uno o más resultados posibles de un experimento aleatorio. O sea, es un subconjunto del espacio muestral.

Eventos exhaustivos: Son conjuntos de eventos que abarcan todas las posibles ocurrencias en un espacio muestral, de modo que al menos uno de ellos debe suceder cuando se hace un experimento aleatorio.

Evento imposible: Es el evento que no incluye ningún resultado y, por tanto, nunca ocurre.

Eventos mutuamente excluyentes: Son eventos que no pueden ocurrir al mismo tiempo. Si un evento ocurre, excluye automáticamente la posibilidad de que ocurra el otro. La probabilidad de la intersección de estos eventos es vacía.

Evento seguro: Es el evento que abarca todos los resultados posibles, por lo que siempre ocurre.

Experimento aleatorio: Es un proceso que produce un resultado impredecible entre varios posibles y cumple con las siguientes condiciones **1.** Todos los posibles resultados del experimento se conocen de antemano. **2.** Cuando se hace el experimento, no se sabe cuál resultado va a ocurrir. **3.** El experimento se puede repetir bajo idénticas condiciones.

Frecuencia absoluta: Es el número total de datos o elementos f_i que se encuentran en el i-esimo intervalo de clase (variables cuantitativas) o en una categoría (variables cualitativas).

Frecuencia absoluta acumulada: Es el número que resulta de sumar la frecuencia absoluta f_i del i-esimo intervalo de clase con todas las anteriores. Se denota como F_i.

Frecuencia relativa acumulada: Es el cociente entre la frecuencia absoluta acumulada de clase Fi y el número total de datos calculados como $n = \sum_{i=1}^{l} f_i$. Se denota como H_i, donde i representa el i-esimo intervalo).

Frecuencia relativa: Es el cociente entre la frecuencia absoluta de clase fi y el número total de datos calculados como $n = \sum_{i=1}^{l} f_i$. Se denota como h_i, donde i representa el i-esimo intervalo.

GeoGebra: Es un *software* libre interactivo de matemáticas que combina geometría, álgebra, cálculo, estadística y gráficos en una plataforma integrada.

Histograma: Es un gráfico que representa la distribución de un conjunto de datos cuantitativos, mostrando en el eje horizontal los intervalos o marcas de clase y en el eje vertical las frecuencias (absolutas o relativas) de cada intervalo mediante barras rectangulares cuya altura corresponde a la frecuencia (absoluta o relativa) de cada clase.

Intersección: Es una operación entre dos o más conjuntos que produce un nuevo conjunto, formado por todos los elementos comunes de los conjuntos. Si se tiene los conjuntos A y B la unión se denota como $A \cap B = \{xx \in A \wedge x \in B\}$ (se lee son los x tales que x pertenecen al conjunto A y al conjunto B).

Intervalos de clase: Es cada uno de los rangos de valores de una variable cuantitativa continua en el que se decide agrupar los datos para hacer un resumen de ellos.

Marca de clase: Es el punto medio de cada intervalo de clase. Se calcula promediando el valor del límite inferior del intervalo con el valor del límite superior. Se denota como m_i, donde i representa el i-esimo intervalo.

Media aritmética: Es una medida de tendencia central que se obtiene sumando todos los valores de un conjunto de datos y dividiendo el resultado entre la cantidad total de valores. La media aritmética muestral se denota como $\bar{x}$ y la poblacional como μ.

Media armónica: Es una medida de tendencia central que se calcula como el recíproco del promedio de los recíprocos de un conjunto de valores. Es especialmente útil cuando desea encontrarse el promedio de tasas o velocidades, y cuando los valores representan proporciones o se distribuyen de manera inversamente proporcional.

Media geométrica: Es una medida de tendencia central que se calcula multiplicando todos los valores de un conjunto de datos y luego extrayendo la raíz enésima de ese producto, donde n es el número total de elementos del conjunto. La media geométrica es útil para analizar tasas de crecimiento o cambios porcentuales.

Media ponderada: Es la misma media aritmética o promedio, pero a diferencia de la media aritmética simple, en la que todos los valores tienen el mismo peso, en ésta cada valor de un conjunto de datos recibe un peso diferente, y refleja su importancia o frecuencia relativa en el conjunto de datos. Para calcular la media ponderada, se multiplica cada valor por su respectivo peso, se suman estos productos y finalmente se divide el resultado entre la suma total de los pesos.

Mediana: La mediana muestral es una medida de tendencia central denotada por $\tilde{x}$, y es el valor central de un conjunto de datos ordenados en forma ascendente. Si el número de datos en la muestra es impar, la mediana es el valor que ocupa la posición central. Si el número de datos es par, la mediana se calcula como el promedio de los dos valores centrales.

Moda: Es una medida de tendencia central que representa el valor o los valores que aparecen con mayor frecuencia en un conjunto de datos. Un conjunto de datos puede tener una única moda (unimodal), más de una moda (multimodal), o ninguna moda, si todos los valores tienen la misma frecuencia, y se denota como Mo.

Muestra: Es un subconjunto de elementos extraído de una población, que resulta representativo y significativo para poder hacer inferencias sobre las propiedades o características de dicha población.

Población: La población o universo es el objeto de estudio. Se puede definir de manera informal como el conjunto total de elementos que tienen una o más características o rasgos comunes posibles bien definidos.

Porcentaje: El porcentaje es 100 veces la proporción.

Probabilidad: Permite cuantificar el grado de verosimilitud de los sucesos o eventos que no podemos controlar y sus relaciones con otros.

Probabilidad condicional: Es la probabilidad de que ocurra un evento B cuando se sabe que ya ocurrió algún evento A. Se denota como $P(B/A)$.

Probabilidad de un evento: La probabilidad de un suceso o evento es un concepto muy complejo de precisar. Existen diferentes concepciones acerca de cómo definir la probabilidad de un evento o suceso. Entre estas está la concepción clásica o laplaciana, de frecuencia relativa y subjetiva.

Probabilidad total: Es un principio en teoría de probabilidades que permite calcular la probabilidad de un evento en función de sus posibles causas o escenarios que cubren todos los casos posibles (un sistema completo de eventos). Esta regla es útil cuando un evento puede ocurrir como resultado de varios eventos mutuamente excluyentes y exhaustivos. La fórmula general se expresa como $P(A) = \sum_{i=1}^{n} P(A/E_i).* P(E_i)$, donde $E_1, E_2, \dots E_n$ son eventos mutuamente excluyentes y exhaustivos.

Proporción: Es una medida que indica la relación entre una parte y el total de un conjunto. En términos estadísticos, la proporción representa el cociente entre la cantidad de elementos que cumplen con una característica específica y el total de elementos del grupo o muestra en estudio. Se expresa como un valor entre 0 y 1.

Regla de combinación: Es una técnica de conteo usada para determinar el número de formas en que se pueden seleccionar elementos de un conjunto, sin importar el orden de los elementos seleccionados.

Regla de multiplicación: Es una técnica de conteo que permite calcular el número total de formas en que se pueden hacer una serie de eventos independientes. Según esta regla, si un primer evento puede ocurrir de m maneras y, luego, un segundo evento puede ocurrir de n maneras, entonces ambos eventos juntos pueden ocurrir en $m \times n$ maneras.

Regla de permutación: Es una técnica de conteo usada para determinar el número de formas en que se puede ordenar o disponer un conjunto de elementos, por tanto, es una disposición ordenada de todos o algunos de los elementos de un conjunto.

Regla de tres simple directa: Regla de tres simple directa se usa cuando dos magnitudes son directamente proporcionales (dos magnitudes x y y se dice que son directamente proporcionales si la razón entre ellas es constante $y/x = k$)

R-Studio: RStudio es un entorno de desarrollo integrado (IDE, por sus siglas en inglés) diseñado específicamente para trabajar con el lenguaje de programación R, ampliamente usado en análisis estadístico, ciencia de datos y visualización de datos.

Tablas de frecuencia: Tabla de frecuencias o tabla de resumen deben tener en cuenta las categorías o características comunes de sus elementos o intervalos, y en cada una de ellas se hace el conteo de sus respectivos elementos. Este conteo va a representar lo que se conoce como frecuencia absoluta.

Técnicas de conteo: Son métodos matemáticos usados para determinar de manera rápida el número de posibles resultados de un espacio muestral sin necesidad de enumerar cada caso individualmente.

Teorema de Bayes: Es un principio fundamental de la probabilidad que permite actualizar la probabilidad de un evento en función de nueva información. Este teorema describe cómo se puede invertir la probabilidad condicional para calcular la probabilidad de una causa dado el efecto observado. La fórmula general se expresa como $P(A/B) = \frac{P(B/A).*P(A).}{P(B).}$.

Unión: Es una operación entre dos o más conjuntos que produce un nuevo conjunto, formado por todos los elementos que están presentes en al menos uno de los conjuntos, sin repetir elementos. Si se tiene los conjuntos A y B la unión se denota como $A \cup B = \{xx \in A \lor x \in B\}$ (se lee son los x tales que x pertenecen al conjunto A o al conjunto B).

Valor esperado de una variable aleatoria discreta X

También conocido como media o esperanza matemática, es el valor promedio que se esperaría obtener si el experimento se repitiera un gran número de veces. Se calcula como $\sum_{i=1}^{n} x_i * P(X = x_i)$, con $x_1, x_2, \ldots, x_n$ valores que toma la variable aleatoria X y se denota como μ_X o $E(X)$.

Variable aleatoria discreta: Una variable aleatoria discreta X es una aplicación que asigna a cada elemento del espacio muestral un único número real.

Variable continua: Características o atributos medibles que asumen cualquier valor numérico en una escala continua y están asociados a la recta numérica real.

Variables cualitativas o **categóricas:** Son características o atributos que toman diferentes formas de descripción y se dice que éstas son variables de clasificación, pero no de medición.

Variable cuantitativa: Características o atributos que toman diferentes formas de medición y para esto usa escalas numéricas.

Varianza de una variable aleatoria discreta X: Es una medida de dispersión que indica qué tan lejos están los valores posibles de la variable aleatoria X respecto a su valor esperado (o media).

Variable discreta: Características o atributos medibles que toman valores numéricos finitos o infinitos numerables y sus resultados surgen del proceso de conteo.

Varianza de una variable estadística: Es una medida de variabilidad absoluta que para un conjunto de datos es el cuasi promedio de las desviaciones respecto a la media elevadas al cuadrado. La varianza muestral se denota como s^2 y la poblacional como σ^2.

Variable estadística: De manera informal, una variable estadística se considera como una característica, atributo o medida que es objeto de estudio.

Esta obra
se terminó de imprimir
el 20 de septiembre de 2025,
en los talleres gráficos de
Grupo Editorial Ibáñez,
Cra. 69 Bis No. 36-20 Sur
Tels.: 2300731 - 2386035
Bogotá D.C. - Colombia

La estadística y la probabilidad son áreas de la ciencia que tienen amplia aplicabilidad en diferentes campos del conocimiento, como la ingeniería, la economía, la biología o la sociología, entre otras. Al igual que las matemáticas, la estadística y la probabilidad constituyen hoy uno de los ciclos fundamentales para la formación de un profesional de cualquier disciplina del conocimiento y son esenciales para la explicación y el estudio de la información de fenómenos que involucran procesos de incertidumbre y de variabilidad.

La estadística en términos generales se ocupa de la observación, recolección, organización, síntesis, análisis e interpretación de la información contenida en datos sobre poblaciones por medio del descubrimiento de patrones, tendencias y relaciones entre los datos, para planear y decidir sobre su futuro.

ISBN:978-958-5198-41-8

9 789585 198418